Genetic innovation through gene duplication in *Drosophila*

by

Seyedeh Ayda Mirsalehi

DISSERTATION

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

at

The University of Texas at Arlington

August 2021

Arlington, TX

Supervising Committee:

Esther Betrán, Supervising Professor

Michael Buszczak

Shawn Christensen

Jeff Demuth

Mark Pellegrino

ABSTRACT

Genetic innovation through gene duplication in *Drosophila*

Seyedeh Ayda Mirsalehi

The University of Texas at Arlington, 2021

Supervising Professor: Esther Betrán

Among several mechanisms, whole gene duplication is one of the major sources for the generation of new genes. A large portion of young genes that emerged through whole gene duplication in *Drosophila* are specifically expressed in the testes and fast evolving, i.e., their proteins are changing very fast between species under positive selection. In this dissertation, we have studied the rates of duplications and functions of newly duplicated genes in three different projects to understand the evolutionary pressures/drivers of this gene innovation.

Chapter 1 is an introductory section. In Chapter 2, by using a comprehensive, detailed phylogenomic study, we examined the duplications of nuclear transport genes (131 genes) in 22 species of *Drosophila* because best well-known meiotic drive systems involve duplication of a nuclear transport gene and explored if those duplications exist in 29 non-*Drosophila* insects. This broad examination revealed the components of nuclear transport that might be under selection in testes, that the selective pressures exist

mainly in Drosophila, and that RNA-mediated duplications are the major contributors to these patterns.

In Chapter 3 of this dissertation, we investigated the function of two nuclear transport duplicated genes (*Ntf-2r* & *Ran-like*) in *D. melanogaster*. In Chapter 4, the function of a nuclearly encoded mitochondrial gene duplicate (*COX4L*) is studied. These duplicates' parental genes (*Ntf-2*, *Ran*, and *COX*4) are essential genes with a broad expression, while the duplicates are testis specific. We utilized different techniques to study the function of those genes. Some of the approaches include generating knock-out mutants of the duplicated genes using CRISPR-Cas9 technology to understand how the absence of these genes affects fertility in male, studying gene expression in these knock-out flies by RNA-seq analysis or tagging the proteins produced by those genes to study their localization.

These functional studies of testis-biased duplicated genes in *Drosophila* has expanded our understanding of the processes they affect and the selective advantages they may provide for the organism and contribute to explaining their retention in the male germline.

Copyright © by Seyedeh Ayda Mirsalehi, 2021

All Rights Reserved

ACKNOWLEDGEMENTS

I want to express my sincere and deepest gratitude to my advisor Dr. Esther Betrán for her mentorship, great support, and guidance. Her dedication, scholarly advice, and keen interest have helped me to a great extent to finish this task. I would also like to thank my committee members, Dr. Michael Buszczak, Dr. Shawn Christensen, Dr. Jeff Demuth, and Dr. Mark Pellegrino, for their valuable suggestions and recommendations.

I also would like to thank all the past and current members of the Betrán lab. Javier Rio for providing the best laboratory management and technical assistance that supported my research. Mehdi Eslamieh, Diwash Jangam, Fatema Ruma, Mira Markova, and Anya Williford for their time in brainstorming ideas, valuable input in my experimental designs, and for being great friends. I would also thank Victor Palacios and Courtney Goldstein from Buszczak lab for their input in my projects. This research was funded by NIH grant (R01GM071813) and UTA RIGS and IDC Return funds to Esther Betrán and Phi Sigma research and travel grants, UTA COS Maverick Science Graduate Research Fellowship, and UTA Dissertation Fellowship to Ayda Mirsalehi.

This acknowledgment would not be complete without expressing my appreciation to my parents, brothers, and husband for their invaluable love, motivation, and encouragement. This journey would not have been possible without their patience and support.

Date: July 27, 2021

Table of contents:

Chapter 1: Introduction

Studying the origin of new functional genes is of great importance due to their significant role in adaptation. Different molecular mechanisms can lead to the formation of new genes. Novel genes can arise from pre-existing genes in an organism via whole or partial gene duplication, retrotransposition, *de novo* gene origination from non-coding sequences or from sources that are foreign to the host genome such as horizontal gene transfer or domestication of transposable elements proteins or viral proteins (Long, et al. 2003; Zhou, et al. 2008; Betrán 2015). Among all these mechanisms, gene duplication is believed to be the most prolific (Kaessmann, et al. 2009; Mendivil Ramos and Ferrier 2012). Gene duplication is a kind of mutation and like any other mutation can have neutral, advantageous or deleterious effects, thus; the evolutionary trajectory of new duplicates is dependent on these effects. However, most researchers are interested in duplicated genes because they can acquire new functions while the parental gene maintains its original function (Long, et al. 2003).

Gene duplication can be mediated by DNA or RNA. RNA-mediated gene duplication which is also called retrotransposition is the mechanism that gives rise to duplicate genes through the reverse transcription of the mRNA of the parental gene and insertion into a new genomic locus. Duplicates originated from this mechanism are usually intronless, contain only some parts of ancestral genes, and they rarely inherit the parental gene's regulatory regions (Kaessmann, et al. 2009). Due to the lack of regulatory elements and presence of mutations such as premature stop codons observed in the retroduplicates of mammalian and fly genomes, it was often assumed that retrogenes were processed pseudogenes (Jeffs and Ashburner 1991; Mighell, et al. 2000). This idea has changed

since a plethora of functional retrogenes have been discovered since the late 1980s (Betrán, et al. 2002; Emerson, et al. 2004; Vinckenbosch, et al. 2006; Bai, et al. 2007; Potrzebowski, et al. 2008; Ding, et al. 2012). Analysis of retrogenes and their tissue expression in human and *Drosophila* showed a general pattern in which the ancestral gene is often located on the X chromosome and the retroduplicate has relocated to an autosomal chromosome and has acquired testis expression. This has been termed, out of the X pattern of duplication (Betrán, et al. 2002). These studies show that the majority of the autosomal retrogenes derived from broadly expressed genes on the X chromosomes have evolved a male and tissue-specific pattern of expression in testis (Betrán, et al. 2002; Vinckenbosch, et al. 2006; Bai, et al. 2007). The origination rate for functional retrogenes during the last 63 million year of *Drosophila* evolution has also been observed to be constant and estimated to be 0.5 retrogene per My per lineage (Bai, et al. 2007).

Several hypotheses have been proposed to explain the evolution of testis-specific duplicated genes. They might evolve to express genes higher in male germline, i.e., by having an extra dose (Innan and Kondrashov 2010). They might be retained to express during the meiotic sex chromosome inactivation (MSCI). MSCI happens during meiosis in the germline of the heterogametic sex and results in inactivation of sex chromosome and a copy of the gene on an autosome could replace the X-linked copy (Emerson, et al. 2004). It has also been proposed that gene duplication occurs to resolve intralocus sexual antagonism. When male and female select for different alleles for a broadly expressed locus, a gene duplication might be the solution to this conflict by giving rise to a sex-specific or sex-biased gene (Gallach and Betrán 2011). The antagonism might take place if there is strong selection for specialization in a tissue (e.g., testis). Duplicates could also

have been retained because they offer a specialized function in the suppression of genomic conflicts such as meiotic drive systems (Presgraves 2007).

To understand the evolutionary pressures that give rise to the new retroduplicates and retention of them in the genome we have studied the rates of duplications and functions of newly duplicated genes in three different projects. These are duplicated genes belonging to nuclear transport and nuclearly-encoded mitochondrial gene ontologies. These are gene ontologies that are overrepresented in the genome of *Drosophila* for having retroduplicates with germline related functions (Bai, et al. 2007).

References

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

Betrán E. 2015. The “Life Histories” of Genes. Journal of Molecular Evolution 80:186-188.

Betrán E, Thornton K, Long M. 2002. Retroposed new genes out of the X in Drosophila. Genome Res 12:1854-1859.

Ding Y, Zhou Q, Wang W. 2012. Origins of New Genes and Evolution of Their Novel Functions. Annual Review of Ecology, Evolution, and Systematics 43:345-363.

Emerson JJ, Kaessmann H, Betrán E, Long M. 2004. Extensive gene traffic on the mammalian X chromosome. Science 303:537-540.

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

Innan H, Kondrashov F. 2010. The evolution of gene duplications: classifying and distinguishing between models. Nature Reviews Genetics 11:97-108.

Jeffs P, Ashburner M. 1991. Processed pseudogenes in Drosophila. Proceedings of the Royal Society of London. Series B: Biological Sciences 244:151-159.

Kaessmann H, Vinckenbosch N, Long M. 2009. RNA-based gene duplication: mechanistic and evolutionary insights. Nat Rev Genet 10:19-31.

Long M, Betrán E, Thornton K, Wang W. 2003. The origin of new genes: glimpses from the young and old. Nature Reviews Genetics 4:865-875.

Mendivil Ramos O, Ferrier DEK. 2012. Mechanisms of Gene Duplication and Translocation and Progress towards Understanding Their Relative Contributions to Animal Genome Evolution. International Journal of Evolutionary Biology 2012:846421.

Mighell AJ, Smith NR, Robinson PA, Markham AF. 2000. Vertebrate pseudogenes. FEBS Lett 468:109-114.

Potrzebowski L, Vinckenbosch N, Marques AC, Chalmel F, Jégou B, Kaessmann H. 2008. Chromosomal gene movements reflect the recent origin and biology of therian sex chromosomes. PLoS Biol 6:e80.

Presgraves DC. 2007. Does genetic conflict drive rapid molecular evolution of nuclear transport genes in Drosophila? Bioessays 29:386-391.

Vinckenbosch N, Dupanloup I, Kaessmann H. 2006. Evolutionary fate of retroposed gene copies in the human genome. Proc Natl Acad Sci U S A 103:3220-3225.

Zhou Q, Zhang G, Zhang Y, Xu S, Zhao R, Zhan Z, Li X, Ding Y, Yang S, Wang W. 2008. On the origin of new genes in Drosophila. Genome research 18:1446-1455.

Chapter 2

BMC Genomics

Nuclear transport genes recurrently duplicate by means of RNA intermediates in *Drosophila* but not in other insects

Ayda Mirsalehi[1], Dragomira N. Markova[1], Mohammadmehdi Eslamieh [1], and Esther Betrán[1*]

[1] Department of Biology, the University of Texas at Arlington, Arlington, TX, USA.

* Corresponding author

Esther Betrán

Biology Department, Box 19498

University of Texas at Arlington, Arlington, TX 76019, USA

Phone (817) 272 1446

E-mail: betran@uta.edu

Abstract

Background: The nuclear transport machinery is involved in a well-known male meiotic drive system in *Drosophila*. Fast gene evolution and gene duplications have been major underlying mechanisms in the evolution of meiotic drive systems, and this might include some nuclear transport genes in *Drosophila*. So, using a comprehensive, detailed phylogenomic study, we examine 51 insect genomes for the duplication of the same nuclear transport genes.

Results: We find that most of the nuclear transport duplications in *Drosophila* are of a few nuclear transport genes, RNA mediated and fast evolving. We also retrieve many pseudogenes. Some of the duplicates are very young and likely contributing to the turnover expected for genes under strong but changing selective pressures. These are duplications of a few classes of nuclear transport genes, potentially revealing what features of nuclear transport are under selection. However, we find only a few duplications of the same nuclear transport genes in Diptera species outside of *Drosophila* and none in other insects.

Conclusions: These findings strengthen the hypothesis that nuclear transport gene duplicates in *Drosophila* evolve either as drivers or suppressors of meiotic drive systems or as other male-specific adaptations circumscribed to flies and involving a handful of nuclear transport functions.

Keywords: nuclear transport, recurrent gene duplication, gene turnover, genetic conflict, *Drosophila*.

Running title: Nuclear transport genes recurrently retroduplicate in *Drosophila*

Background

In eukaryotes, the nucleus is separated from the cytoplasm by a double membrane nuclear envelope. The nuclear envelope prevents the free flow of macromolecules between the nucleus and cytoplasm. Selective nucleocytoplasmic transport of proteins and RNAs occurs through the nuclear transport system [1]. The conventional nuclear transport system consists of several components that fall into three main categories: 1) Nuclear pore complexes (NPCs) that are huge protein complexes residing on the nuclear envelope and consist of several copies of approximately 30 different nucleoporins (Nups) that assemble to make the NPCs [2, 3]. 2) Nuclear transport receptors/carriers called karyopherins that consist of importins and exportins. The Importin superfamily consists of importin-α and importin-β sub-groups [4]. The carrier proteins recognize and translocate the cargo across the nuclear envelope through interactions with the NPCs. 3) Factors that assist the process and directionality of nuclear transport such as Ran, RCC1, RanGAP, and Ntf-2 [2, 5, 6]. These conventional mechanisms of receptor-mediated nuclear import and export in the case of proteins are depicted in Figure 1A [6, 7]. Karyopherins are not only involved in protein import and export but are also involved in the transport of RNAs [6]. The type of cargo indicates which karyopherin will be used and if additional adaptors are needed [8]. Export of small RNAs such as tRNAs and microRNAs follow the general pattern of exportin-mediated protein export. In this case, tRNA/miRNA-specific exportins directly bind to tRNA/miRNA and RanGTP to mediate the export [8-10]. Export of large RNAs such as ribosomal RNAs (rRNAs) and mRNAs requires export receptors as well as additional export adaptors assembled into complicated ribonucleoprotein (RNP) particles. The general mRNA export receptor complex in metazoans (Nxf1–Nxt1) is not a karyopherin family member [11]. Similar to the export of rRNAs and mRNAs, the export of snRNAs

requires adaptor proteins that recruit the export receptor. However, assembly into an RNP is not needed for the nuclear export of snRNAs [8].

Aside from the conventional nuclear transport pathways, an increasing number of alternative non-conventional mechanisms have been described during the past years that mediate the nuclear transport independent of the karyopherins [12, 13]. These karyopherin-independent pathways include transport by utilizing alternative carriers such as the calcium-binding proteins calmodulin for nuclear entry and calreticulin for nuclear export. Also, some proteins can translocate through direct interaction with nucleoporins of the NPCs and be able to piggyback other cargoes with themselves across the nuclear membrane. Many proteins are transported by both conventional and alternative pathways. Exploiting multiple pathways for transport of the same cargo has been suggested to assure cellular functionality in the situations that one mechanism is inhibited or adjust the nuclear transport based on the cellular demand [12, 13].

Although nuclear transport mechanisms are quite conserved and needed in all eukaryotic cells [6], duplications of some of their components have been observed. One well characterized protein gene family is the importin family of adapter proteins [14]. Karyopherin family members have a moderate sequence identity, and their highest similarity is for a binding domain for the small Ras-like GTPase Ran [5]. The diversity between karyopherins seems to have evolved related to their specific cargoes. However, although it was originally thought that each karyopherin can function as either import or export factor but not both, there is evidence that some karyopherins have import and export functions at the same time but for different cargo [15]. An N-terminal importin-β binding domain and a tandem array of Armadillo (ARM) repeats are characteristics of importin-αs, while importin-βs consist mainly of HEAT repeats. Previous studies showed

shared ancestry between ARM and HEAT repeats and importin-β being the progenitor of the importin-α karyopherins [16]. The importin family has diversified in higher eukaryotes. In which, *Saccharomyces cerevisiae* encodes for a single importin-α, while most metazoans encode for three importin-αs [17]. However, this is not the case for most other components of nuclear transport, e.g., *Ran* and *Ntf2*, that are present as a single copy in most genomes.

Interestingly, previous studies have shown recurrent and convergent duplication of some of these single copy nuclear transport genes and expansions of the importin gene families in different *Drosophila* lineages. Losses of gene duplicates (i.e., high turnover), fast evolution, and particular patterns of gene evolution and expression have also been observed in some *Drosophila* lineages [18-21].

Initially, *Ntf-2* and *Ran*, two nuclear transport genes, were reported to have given rise to retroposed copies at least three independent times within the same *Drosophila* lineages [18, 19, 21], but DNA mediated duplications were not explored. These duplicates have acquired testis-specific expression and have shown signatures of positive selection in certain lineages [19, 21, 22]. Similarly, *importin-α* genes were later shown to have undergone six duplication events detected in the initially 12 *Drosophila* sequenced species genomes available in public databases. Some members of the *importin-α* gene family have been found to be repeatedly gained and lost during different stages of *Drosophila* evolution, with the retained copies gaining testis-specific pattern of expression. On average, *Drosophila* genomes have been reported to contain between four to five importins [20].

Another member of the nuclear transport GO, *e(y)2b* is retroposed and was generated from *e(y)2*. This retroposition event is an example whereby a retrogene takes

over the functions of the parental gene throughout evolution and gain ubiquitous expression while the parental gene, *e(y)2b* has acquired a testis-specific pattern of expression. E(y)2 was shown to be a component of a protein complex involved in transcription coupled mRNA export in *D. melanogaster* [23]. Another reported instance of gene duplication involving nuclear transport is the evolution of a pair of *D. melanogaster*-specific tandem duplicated genes, *Artemis* (*Arts*) and *Apollo* (*Apl*) from a 7.7kb region on chromosome arm 3L of *D. melanogaster*. The ancestral gene which has characteristics of *Importin-β* was duplicated to produce two identical gene copies. However, after duplication, *Apl* and *Arts* evolved to acquire male-biased and female-biased expression, respectively. Investigations showed that *Apl* is male-biased and affects development and male fertility and *Arts* affects female fertility. These tissue-specific patterns of expression might have evolved to resolve the sexually antagonistic effects of these genes [24].

So far, based on previous findings, mostly RNA-mediated duplication events contributed to the formation of new *importin-α*s, *Dntf-2, Ran* [18, 20], and *e(y)2* [23]. Recurrent duplication of *Dntf-2*, *Ran* and *importins* and positive selection acting on some components of nuclear transport such as *Dntf-2* and *Ran* duplicates*, RanGAP* and six nucleoporins (specifically Nup107), have been proposed to be involved in suppressing male germline conflicts potentially meiotic drive systems [19, 21, 25, 26]. Meiotic drive is a phenomenon in which heterozygous individuals favor transmission of one allele (selfish genetic element) to the gametes at the expense of the alternative allele and causes deviations from Mendel's first law. Meiotic drive systems in *Drosophila* are often generated and evolve through gene duplication events [27]. Examples of such drivers are segregation distorter (SD) system of *D. melanogaster* [28], Winters Sex-Ratio (SR) system of *D. simulans* [29], and the SR system in *D. neotestacea* [30]. Segregation Distorter (*SD*) as one of the

well-characterized meiotic drive systems described in *D. melanogaster* is a multi-locus gene complex comprised of three main interacting loci clustered near the centromere of chromosome 2. The three loci are, 1) The driver, segregation distorter (*Sd*), which is the main distorting locus and a truncated tandem duplication of the *RanGAP* gene, 2) Enhancer of *SD* (*E(SD)*), and 3) The target of drive, *Responder* (*Rsp*), which is a large array of 120 bp pericentromeric satellite repeats. The copy number of these repeats define the sensitivity of the *Rsp*. The *Rsp* alleles with <200 copies are insensitive, and those alleles with >700 copies are sensitive. Molecular characterization of the SD components showed that SD chromosomes carry insensitive alleles of *Rsp* (Rsp^{i}), while the wildtype alleles mainly carry the sensitive alleles (Rsp^{s}). With the three defined loci, the genotypes of the *SD* and wildtype chromosomes will be *Sd E(SD)* Rsp^{i} and $Sd^{+}E(SD)^{+}$ Rsp^{s} respectively. SD causes meiotic drive in which although heterozygous SD/SD^{+} females pass *SD* and SD^{+} equally to their progenies, 99% of heterozygous *SD*/ SD^{+} male progenies inherit *SD* bearing chromosomes. *Sd* performs the same enzymatic activity as the wild-type *RanGAP*. However, since it is truncated it lacks its nuclear export signal (NES), and it mislocalizes to the nucleus resulting in accumulation of RanGDP molecules in the nucleus [25]. Disruption of the RanGTP gradient might cause segregation distortion. The reason why only *SD*-bearing sperms survive spermatogenesis might be the disruption of chromatin remodeling in SD^{+}-bearing spermatid nuclei and the failure of the transition between histones and protamines and maturation into functional sperm, resulting in their accumulation in the waste bags [28].

Recent studies have suggested the involvement of Piwi-interacting RNA (piRNA) pathway in silencing the *Rsp* satellite repeats needed for maturation of normal spermatids. piRNA pathway is a germline specific small RNA-based silencing system. In *Drosophila*, most of piRNAs derive from piRNA clusters that are large loci with high proportion of TE

sequences [31, 32]. piRNA precursors which are piRNA cluster transcripts are exported from the nucleus. In the cytoplasm several factors contribute to the piRNA biogenesis and loading of piRNA into Piwi proteins. After formation of piRNA-induced silencing complexes (pi-RISCs), they are imported back into the nucleus, where they guide transcriptional silencing of repetitive elements by inducing heterochromatin formation at the target loci (Figure 2) [33]. Piwi is the nuclear member of the Argonaute protein family in *Drosophila*. The two other family members, Aubergine (Aub) and Argonaut3 (AGO3), remain in the cytoplasm and function in the ping-pong cycle. These proteins, Piwi, Aubergine (Aub), and Argonaut3 (AGO3), are specifically expressed in germline, and maintain the integrity of the genome during gametogenesis by silencing of the transposons and other repetitive elements [32]. The need of proper import and export mechanisms for the functionality of piRNA pathway in silencing repetitive elements highlights the importance of nuclear transport components in suppressing meiotic drive systems. Previous studies in *Drosophila* ovaries have shown that *importin-α* have a crucial role in the localization of Piwi to the nucleus. In the loss of function mutants of *importin-α*2 and *importin-α*3, Piwi is not imported into the nucleus, but this can be rescued by overexpression of any of the Importin-*α* member [34]. Investigations of whether the Responder array of satellite repeats could be a target of the piRNA pathway has shown that mutations to both *Aubergine* and *Piwi* act as enhancers of distortion [32].

Here, we present the results of a comprehensive study of nuclear transport duplications in insects. Nuclear transport genes (131 genes) were studied in 12 species of *Drosophila*. The duplicated genes in these species were studied in 10 additional *Drosophila* species and 29 species of non-*Drosophila* insects. These species encompass seven orders of insects representing divergence time of 583 Mya from *Drosophila* (Timetree.org).

We explored if the same components of nuclear transport have been recurrently duplicated in *Drosophila* and in other insects. We also studied the duplication mechanisms and mode of evolution of these duplicates. Broad examination of the nuclear transport system duplications can reveal additional components that might have been under selection in testes, if the selective pressures exist only in *Drosophila* or are broader, i.e., the same trends are observed in other insects. This study also reveals the contribution of RNA-mediated and DNA-mediated duplications to the observed patterns.

Results and Discussion

Most nuclear transport duplicates are retrogenes in *Drosophila*

We analyzed duplications for 131 genes assigned to nuclear transport gene ontology (Supplementary material 1) initially in 12 *Drosophila* species genomes obtained from Ensembl Metazoa (species are listed in Supplementary material 2). In other *Drosophila* species (Supplementary material 2), we performed tBLASTn searches using the *Drosophila melanogaster* parental protein of each gene found to be duplicated in the initial search (Supplementary material 3). To find *importin-α-5* duplicates in *Drosophila*, we used *Drosophila eugracilis* parental protein as a query. We analyzed the synteny across species for all recovered duplicated sequences to be certain of our classification as orthologous copies or independent duplications (Supplementary material 4). Our phylogenetic results support the synteny based conclusions (See below).

Drosophila duplication events detected for each gene are depicted in Figure 3. Among the 131 nuclear transport genes that we explored, eight classes of genes were detected to have undergone duplication events: *Ran*, *Ntf-2*, *Importin-α*s*, Nucleoporin-93 (Nup93)*, *NTF2-related export protein-1 (Nxt1), enhancer of yellow 2* (*e(y)2), Transportin*

(Tnpo) and *Importin-β,* with *Ran, Importin-α, and Ntf-2* being the three components experiencing the highest numbers within genomes (Figures 1B and 3 and Supplementary material 4).

Using *Drosophila melanogaster*'s Ntf-2 as a query we identified a total of six independent duplication events for *Ntf-2* in the *Drosophila* species. Three of these duplicates have not been reported before [18, 19] and include only one DNA-mediated duplication event (*Ntf-2r(5*)). All duplicates possess the characteristics of functional copies. *Ntf-2* has four exons and four transcripts are annotated for this gene in FlyBase. All detected retroduplications are derived from *Ntf-2* transcript RA and that is the transcript used in the phylogeny in Figure 4A. Transcript RA is the isoform that is expressed highest among all the transcripts [35, 36]. The only detected DNA duplicate (*Ntf-2r(5*)), also lacks the specific exon of Ntf-2 transcript B, indicating that this duplicate has lost this exon after duplication. *Ntf-2* retroduplicates are complete retrocopies of the parental gene in which all four exons are retained and were thus detected as a single hit in the Blast searches. As mentioned above, *Ntf-2r(5)* has duplicated through a DNA intermediate, a mode of duplication that has not been previously reported for this gene but it is not a tandem duplication.

Interestingly, *Ran* has been the gene with the highest turnover in terms of gain and loss events. By using *D. melanogaster*'s Ran as a query, we detected a total of 13 instances of independent duplication events of *Ran* with only one of these being a DNA-mediated duplication (Figure 3 and Figure 4B). Three of these duplication events have been reported in previous studies [18, 19]. *Ran* is a gene with two exons and only one transcript that has given rise to these duplicates. Unlike the *Ntf-2* duplicates that were all identified to be functional, at least seven pseudogenes were identified for *Ran* in different

Drosophila species and some of these pseudogenes are very young duplicates (>80% protein identity to the parental gene) that are disabled. Identification of several gain and loss events through gene duplication and pseudogenizations indicates the high turnover of these genes across *Drosophila* species. This high number of *Ran* pseudogenes (7/22=0.31 pseudogene per gene and per genome) is interesting since it is statistically significantly higher than expected number of *Drosophila* pseudogenes per gene in general (P<0.0001 in Z-test), which is estimated to be around 110 total pseudogenes in the *Drosophila melanogaster* genome, which will be about one pseudogene for every 130 proteins encoded or 0.0077 pseudogenes per gene per genome [37]. High number of detected pseudogenes is specific to *Ran* and we have not observed as many pseudogenes and turnover for other nuclear transport genes.

Using Importin-α1 as a query, we identified a new *Importin-α* gene which we named as *αKap7* (*α-Karyopherin-7*) following the nomenclature from previous work [20] for consistency (Figure 3 and Supplementary material 4). αKap7 possess both an N-terminal IBB domain as well as Armadillo repeats which are the characteristics of the Importin-αs. *αKap7* which is present in *D. rhopaloa* and *D. elegans* is evolutionary close related to *αKap1* confirmed with both identity scores from Blast searches and phylogenetic relatedness (Figure 4C and Supplementary material 3). We have also detected one copy of *αKap7* in *D. willistoni* which has been pseudogenized (Supplementary material 4).

Blast searches with *Importin-α2* revealed that this gene has given rise to several (8) independent duplications. Three independent duplications from *αKap2* (*αKap2A*, *αKap2B* and *αKap2C*) were reported before [20]. We found five more independent duplications from *αKap2* in the Blast searches of 22 *Drosophila* genomes (Supplementary material 4). We identified a new *αKap2* duplicate in *D. eugracilis* (*αKap2D*), three independent duplicates in

D. willistoni (*αKap2E*, *αKap2F* and *αKap2G*) and an additional duplicate in *D. grimshawi* (*αKap2H*) (Figure 3). We follow previous nomenclature to indicate that they are duplicates from *αKap2*. All of these newly detected *αKap2* duplicates contain the Armadillo repeats, characteristic of the canonical *importin-α*s and they all lack the IBB domain at their N-terminal. *αKap2D*, *αKap2E*, *αKap2F* and *αKap2G* are retrogenes as they do not contain any of the introns of *αKap2*. *αKap2H* is a partially processed retroduplicate and contains one of the introns of the *αKap2* gene.

Using Importin-α3 as a query, we identified presence of *αKap4* in two species outside of the *melanogaster* subgroup (Figure 3). Previously the presence of *αKap4* was reported to be only in *melanogaster* species subgroup [20]. These αKap4s lack the IBB domain and contain one of the introns of *αKap3* as described before for this importin.

Exploring *Drosophila* species with αKap3 duplicates we identified another independent duplication event which we named as *αKap6*. This duplicate that is present in six *Drosophila* species (Figure 3 and Supplementary materials 3 and 4), also lacks the IBB domain at the N-terminal and appears to be partially processed retroduplicated by having a similar structure to αKap4s and containing one intron from the parental sequence.

We used *D. eugracilis* αKap5 to identify further duplications of this gene. No new independent duplications were found for *αKap5*. However, we detected the presence of *αKap5* in four species that have not been explored in the previous studies, two of these species are close to *D. melanogaster* group. We have found that *αKap5* has been pseudogenized in *D. mojavensis* as it possesses several premature stop codons. Finding additional species containing *αKap4* and *αKap5* provides a more precise dating of *αKap5* loss and origination of *αKap4* than previously reported.

Interestingly, the results from Importin duplications shows that the majority of independent duplication events occurred for *importin-α2*, one importin that was characterized in previous studies to have specialized in *Drosophila* for gametogenesis functions [17, 20, 38, 39].

Our conserved domain analysis of newly detected *αKap4s*, *αKap5s, αKap6* and every new duplication of *αKap2* (*αKap2D-2H)* shows absence of the IBB domain in these duplicates. Similarly, Phadnis et al. showed absence of IBB domain in *αKap4, αKap5 and αKap2A,* highlighting presence of at least one testis-specific *importin-α* that lacks an IBB domain in all 22 *Drosophila* species that was studied with the exception of *D. kikkawai*. IBB-less *importin-as* were shown to be functional in nuclear transport in *S. cerevisiae* and *Drosophila* [40, 41]. Origination of several *importin-αs* that lack the IBB domain might be an adaptation for *Importin-β* independent and specialized nuclear transport function in the male germline.

Among other nuclear transport-related genes examined here for *Drosophila*, there are few others that have undergone duplication events. Of all 30 nucleoporins we examined, *Nup93* is the only nucleoporin that has been duplicated. We have found an old DNA duplication event which is present in all 22 *Drosophila* species we analyzed. In addition, two other independent duplicates were found. One retroduplicate present is *D. kikkawai* and a second DNA-mediated duplicate with the least identity to its parental gene in *D. albomicans* (Supplementary material 4). We could not find any traces of pseudogenization in these duplicates and despite being relatively old, all detected *Nup93* duplicates seem to be functional.

Nxt-1 is another nuclear transport component found to have undergone duplications. We discovered three independent duplications for this gene and all copies seem to be

functional. Two out of the three duplications events of *Nxt-1* are DNA-mediated but not in tandem (Figure 3 and Supplementary material 4).

An additional detected gene family of nuclear transports is *e(y)2b* which has retroposed and generated *e(y)2*. This previously described retroposition event is an exceptional example in which the retrogene (*e(y)2*) takes over the functions of the parental gene (*e(y)2b*) during evolution and gain ubiquitous expression while the parental gene has acquired testis-specific pattern of expression [23]. In the previous study of this gene family, [23] presence of *e(y)2* and *e(y)2b* in eight species of *Drosophila* was reported. Our analysis shows the existence of both parental and duplicate genes in the 22 *Drosophila* species studied here indicative of one old retroduplication event (Figure 3 and Supplementary materials 3 and 4). All the retrieved sequences seem to be functional.

Tnpo (*βKap2*) is another gene that has a duplication present in seven species of *Drosophila*. The shared synteny shows that all these duplicates are orthologous and derive from a single DNA-mediated duplication event (Figure 3 and Supplementary material 3).

Our results show a greater number of duplication events and losses for nuclear transport genes in *Drosophila* than was previously reported. This study has also revealed an excess of RNA-mediated duplications (Figure 1B) that cannot be explained by a lower detection probability of DNA duplicates because using single exons as queries for our Blast searches did not change the results (See Methods).

The high turnover of nuclear transport genes is circumscribed to *Drosophila*

We explored duplication of nuclear transport components in 29 non-*Drosophila* insect species representing seven orders of insects. For these searches outside of *Drosophila*, we used each species' specific parental protein as a query. Despite the efforts

that were put to detect duplications of the genes observed to be duplicated in *Drosophila* by blasting the parental sequence of each lineage, no functional gene duplicate for those genes was found in species outside of Diptera. Our analysis shows that nuclear transport duplication events are not only limited to *Drosophila* lineages. However, the duplication events are observed in fewer genes and are limited to three dipteran species close to *Drosophila* and not any species outside of Diptera contain this kind of functional nuclear transport duplicates (Supplementary materials 3 and 4). Diptera duplicates detected out of *Drosophila* are limited to *Glossina morsitans morsitans* (Tsetse fly), *Aedes aegypti* (Yellow fever mosquito) and *Anopheles gambiae* (Malaria mosquito). We observed retroduplicates of *Ntf-2* in *Glossina morsitans morsitans* and *Anopheles gambiae*. *Glossina morsitans morsitans Ntf-2* duplicate has been pseudogenized. Similarly, *Ran* has given rise to duplicates outside of *Drosophila* species. We observed retroduplicates of *Ran* in *Glossina morsitans morsitans* and *Aedes aegypti*. *Aedes aegypti* duplicate has been pseudogenized. *Nxt-1* has also one retroduplicate in *Aedes aegypti*, and *Nup93* has a retroduplicate in *Glossina morsitans morsitans.* Our analysis show presence of *Importin-α1*, *Importin-α2* and *Importin-α3* in all 29 outside *Drosophila* insect species, however unlike *Drosophila*, we did not observe any additional *Importin-α* duplicates in non-*Drosophila* species. Likewise, no duplication of *e(y)2b* and *Tnpo (βKap2)* is detected in non-*Drosophila* species (Supplementary materials 3 and 4). Scarcity of duplicates out of *Drosophila* species cannot be explained by the quality of the genomes (Supplementary material 2).

Nuclear transport gene duplicates have testis-specific expression

Expression data from *D. melanogaster* modENCODE RNA-Seq data for different tissues [35, 36] (Supplementary material 5) and several previous studies in additional

Drosophila species showed that nuclear transport genes follow a general pattern in which the ancestral genes actively express in almost every tissue, while the duplicate genes evolve to have a tissue specific expression in testis. RT-PCR results from *D. melanogaster* and *D. ananassae* showed that *Ntf-2r* and *Ran-like* are strongly testis-biased [19, 21, 42], while the parental genes, *Ntf-2* and *Ran* are expressed in every tissue with a higher expression in the ovaries.

Profiled expression pattern studied by PCR analyses from cDNA collected from various species of adult *Drosophila* male and female tissues showed that *aKap4*, *aKap2B*, *aKap2C*, *aKap5A* and *aKap5B* had gained a highly testis-specific pattern of expression in contrast to the *aKap1* and *aKap3* that have a ubiquitous expression and *aKap2* which is enriched in both testes and ovaries [20]. This is in agreement with published profiles of gene expression for those genes [42].

Tnpo, the ancestral gene of *CG8219* has ubiquitous expression with high expression in ovary and the duplicate gene has acquired high testis-specific expression. Similarly, while the expression of *e(y)2* was detected in all tissues at the same level, the mRNA of *e(y)2b* was detected only in testis [23]. modENCODE RNA-Seq data shows very low to moderate expression of *Nup-93* (*CG11092*) in every tissue and high expression of this gene in the ovary, while *Nup-93-like* (*CG7262*) has acquired moderately high expression in in testis.

Therefore, accumulated expression data available for the duplicated genes and parental genes supports presence of testis-biased expression pattern for all duplicated genes for which expression has been studied (Supplementary material 5).

Mode of evolution and phylogenetic analyses of the duplicates

To examine the evolutionary relationships between parental genes and duplicates, we performed phylogenetic analysis using the protein sequences and a maximum likelihood (ML) approach in 22 *Drosophila* species. We found that parental and duplicate genes are grouped into distinct clades in which parental sequences are associated with short branch lengths consistent with a slower rate of evolution for parental sequences than duplicates suggesting a high degree of evolutionary constraints for parental genes. In contrast, duplicate genes have long branches indicating that the duplicate genes are evolving at a significantly faster rate than their respective parental counterparts (See below).

To identify the selection pressures acting on parental and duplicated genes newly discovered here, first, the ratio of nonsynonymous substitutions per nonsynonymous sites to synonymous substitutions per synonymous sites (dN/dS) was calculated using the codeml algorithm (Yang, 2007) in EasyCodeML (www.github.io/bioeasy/EasyCodeml; Supplementary material 5). The dN/dS ratio was smaller than 1 for all genes tested, indicating that purifying selection is the major evolutionary force at the protein level. However, the mean ω values for the parental genes in the species compared ($\omega_{Ran(1)}$ = 0.0001, $\omega_{Ran(3)}$ = 0.0147, $\omega_{Nxt\text{-}1}$ = 0.06286, $\omega_{Nup\text{-}93}$ = 0.1281, $\omega_{\alpha Kap3}$ = 0.0224) were lower than the ω values for the retroduplicates, ($\omega_{Ran\text{-}like(1)}$ = 0.47031, $\omega_{Ran\text{-}like(3)}$ = 0.5177, $\omega_{Nxt\text{-}1\text{-}like(1)}$ = 0.1022, $\omega_{\alpha Kap6}$ = 0.354) (Supplementary material 5) except for Nup-93-like ($\omega_{Nup\text{-}93\text{-}like(1)}$ = 0.0827). All ratios were statistically significantly smaller than 1 except *Ran-like(1)* (Supplementary material 5). In *Ran-like(1)*, we see the duplicate only in three very close

related species (i.e., where few synonymous changes have occurred) and that might rend this estimate less reliable.

Second, we performed two-ratio branch model analysis further confirming that parental sequences are subjected to highly significantly stronger purifying selection than the duplicates ($p<0.001$) (Supplementary material 5), with the exception of *Nup-93* in which a higher rate is observed for the parent instead of the duplicate. This is in agreement with results from the phylogenetic analysis of each set of genes (Figure 4A-G). Two-ratio branch model can be used to test whether there are significant ω differences among branches of the tree by assuming that specific branches can have an ω that differs from the rest of the tree [43-46].

Previous studies showed that novel *αKaps* (*αKap2B* and *αKap5)* evolve under purifying selection except for ARM repeats of *αKap4* which showed to be evolving under positive selection [20]. Also, while both *Dntf-2* and *Ran* were shown to be evolving under purifying selection in *Drosophila* with *Ran* being under stronger purifying selection (dN/dS=0.0188 and 0.0065 respectively), *Ran-like* and *Ntf-2r* were shown to be evolving under positive selection in certain *Drosophila* lineages [19, 21]. The statistically significant higher rate of evolution for the duplicated genes is at least partially explained by positive selection as supported by the McDonald and Kreitman test and site specific likelihood models [19-21].

Conclusions

The majority of gene duplicates (~80%) that are limited to single species are tandem gene duplications in *Drosophila* [47] revealing the mutational biases. However, in this study we confirm and expand on the finding that the vast majority of nuclear

transport duplications in *Drosophila* are relocations (copies to a different location than the parental gene) (Supplementary material 3), fast evolving and many are RNA mediated (77.77%) (Supplementary material 5). The mode of evolution of these nuclear transport-derived gene duplicates is related to the function retroposed copies are selected for in male germline. Relocation or retroposition might allow these copies to acquire testis-specific expression and the lack of introns might be beneficial for the processing and nuclear export of the transcripts during meiosis in male germline [48-50].

Although we find a few duplications of nuclear transport genes in Diptera species outside of *Drosophila*, most of them are in *Drosophila* species and duplicates for those genes are not found in other insect genomes. Total number of duplicate genes per species is shown in Supplementary material 5 indicating presence of at least 3 gene duplicates in *Drosophila* species. *D. rhopaloa* is the species with the highest number of nuclear transport duplicates (10 gene duplicates) while *Glossina morsitans morsitans* with only two gene duplicates has the highest number of nuclear transport duplicates among other dipteran species. The other two species, *Aedes aegypti* and *Anopheles gambiae* have only one gene duplicate.

The number of independent duplication events per gene (Figure 3, Supplementary materials 3 and 4) shows greatest number of duplications for *Ran*, *Importin-α* and *Ntf-2-RA* (with 13, 13 and 6 duplication events respectively) and in particular *Importin-α2* (8 duplication events). These few gene families of nuclear transport with a high number of gene duplications reveal what features of nuclear transport are under selection in *Drosophila*. These are all genes or transcripts that are highly or specifically expressed in *Drosophila* testis. Significant number of *Ran*, *importin-α* and *Ntf-2* duplications can be an adaptation for the suppression of male meiotic drive systems similar to Segregation

Distortion (SD) of *D. melanogaster*. In the presence of SD/SD^{+} males which may suffer disruption of the RanGTP gradient, the canonical import of piRNAs loaded into Piwi proteins may be perturbed. Thus, the piRNAs cannot efficiently be imported back to the nucleus for transcriptional silencing of the *Rsp* repeats (Figure 2). In this process spermatids with *Rsp*^{S} alleles which have higher copy number of repeats than *Rsp*^{i} disproportionally get affected and fail to mature to normal spermatids [28]. Two other genes that show duplications (*Nxt1* and *Nup-93*) have been described to be part of the piRNA pathway in ovaries [33]. The positive selection observed for some of these duplicates [19-21] or the effects of the extra dose of genes like *Ran* [51] on the suppression of SD would support this.

Alternatively, the many duplicates of *Ntf-2*, *Ran* and *importin-α* could have other male-specific functions. In any case, the presence of these duplicates mostly in *Drosophila* would support that these conflicts or other testis-specific selective pressures are restricted to fruit flies. The genes in which this is found are *Ntf-2* and *Ran*. Average number of *Ran* young duplicates (i.e., >80% identity to their parents) per species is significantly higher in *Drosophila* than in other insects. (i.e., 30% (7/22) in *Drosophila* and only 3% (1/29) in other insects including Diptera). Two proportions were tested with a Z-test. (Z=2.7591; *P*= 0.0058). Likewise, average number of *Ntf-2* young duplicates per species is also significantly higher in *Drosophila* than in other insects (i.e., 23% (5/22) in *Drosophila* and only 3% (1/29) in other insects; Z=2.1164; *P*= 0.034). *Ntf-2* and *Ran* are the only genes with more than one new duplicate indicating high turnover. This number of young genes and pseudogenes speaks of strong and changing selective pressures specially for these two genes and restricted to *Drosophila*.

Methods

Identification of orthologous and paralogous genes of nuclear transport components in 12 *Drosophila* species

We analyzed genes assigned to Gene Ontology ID, GO: 0051169 (Nuclear transport; 131 genes) to detect the duplicates of nuclear transport components. BlastP searches [52] considering a cutoff level of >=50% identity were performed for 12 *Drosophila* species genomes obtained from Ensembl Metazoa for every protein included in the above Gene Ontology. Using the Markov Cluster Algorithm (MCL Algorithm) [53], genes were grouped into gene families. Gene families with more than one member were considered to have duplications and were manually analyzed in order to understand their detailed structure. We used previously published scripts for these analyses [54].

Identification of analogous duplications of nuclear transport components in additional insect lineages

We expanded our search of the candidate nuclear transport genes' duplications in more *Drosophila* species as well as outside of *Drosophila* to see if the same selective pressures exist in all analyzed species. Searches for the duplication of *Ntf-2*, *Ran*, *Importin-α, Nup93*, *Nxt-1*, *e(y)2b*, *Tnpo* and *Importin-β* were performed for a total of 22 *Drosophila* species and 29 non-*Drosophila* insect species listed in Supplementary material 1; for which full genome sequences are available on FlyBase. We used tBLASTn implemented in FlyBase [55] or NCBI genome databases. *Drosophila melanogaster* parental protein of each gene was used as a query for searches within the *Drosophila* genus. For finding additional *importin-α5* duplicates we used *Drosophila eugracilis* as a query. For searches outside of *Drosophila*, we used each species' specific parental protein

as a query. tBLASTn hits with identity >=40% were retained. We analyzed all tBLASTn results to understand the mode of evolution of the duplicated gene copies. In cases of retroduplication, the gene copy appears in BLAST searches as a solid hit with no introns in between. While in a DNA-mediated duplication event, the exons appear as individual hits on the same chromosome or scaffold. In order not to miss a DNA duplication event, we also used single exons as queries of our Blast searches, which did not change the results. For annotated genomes, we used the FlyBase genome browser track option to obtain all sequences. For genomes that were sequenced but not annotated, we used the Graphics track option of tBLASTn in the NCBI genome database to collect the complete sequence of the hits. We analyzed and annotated all gathered parental and duplicate sequences manually using the ExPASy bioinformatics resource portal [56], thus allowing us to confirm their functionality and detect pseudogenes. If one or more premature stop codons were observed in the translated proteins or the transcript was truncated, the copy was considered a pseudogene. The results of all BLAST searches can be found in Supplementary material 3.

Using the NCBI conserved domain database [57, 58], we compared the regions flanking each duplicate for synteny conservation, thus allowing us to confirm if detected sequences are orthologs or independent duplications.

Phylogenetic analyses

To examine the evolutionary relationships between parental genes and duplicates, we built phylogenetic trees using the maximum likelihood (ML) approach [59]. Multiple alignments of protein sequences were performed using ClustalW [60] implemented in Geneious software (Version 2020.1) [61]. Maximum likelihood phylogenetic trees of *Ntf-2*,

Ran, *Importin-α, Nup93, Nxt-1, e(y)2b* and *Tnpo* parental and duplicate sequences among 22 *Drosophila* species were constructed by using the BlOSUM62 substitution model with 100 bootstrap branch support in PhyML [62] applied in the Geneious software. We used FigTree (Version 1.4.4) (http://tree.bio.ed.ac.uk/software/figtree) and Inkscape software to modify visual features of the phylogenies.

Mode of evolution analyses

To detect the signatures of selection on different genes, the ratio of nonsynonymous substitutions per nonsynonymous sites to synonymous substitutions per synonymous sites (*d*N/*d*S) was estimated and compared between newly described parental and new gene pairs using the CodeML algorithm (Supplementary material 5)[59] implemented in EasyCodeML [46]. Accordingly, under the assumption of neutral evolution, *d*N/*d*S ratios are expected to have a value of 1, ratios less than 1 indicate negative or purifying selection, corresponding to high selective constraints, and values greater than 1 indicate positive selection, suggesting adaptive evolution [63]. First, the branch model was used with a null model assuming that each respective group of sequences is evolving at the same rate (one-ratio model) and an alternative model in which the *d*N/*d*S ratio was fixed to *d*N/*d*S =1. Second, to test whether the parental and the retroduplicate genes evolve under different evolutionary constraints, additional branch model specifying different rates for the different gene branches were compared to a single rate. The likelihood ratio test (LRT; [64]) was conducted to perform pairwise comparisons of both models for all comparisons. Only a *P*-value of 0.05 or less in the LRTs was considered to be significant, indicating that the rates between parental gene and duplicate genes were significantly different (Supplementary material 5).

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and materials

All sequences extracted from the genomes or analyzed during this study are provided in Supplementary material 6.

Competing interests

Authors declare that they have no competing interest.

Funding

EB would like to acknowledge the support from the National Institute of General Medical Sciences of the National Institutes of Health (R01GM071813). The content of this work is solely the responsibility of the authors and does not necessarily represent the

official views of the National Institutes of Health. AM would like to acknowledge UTA COS Maverick Science Graduate Research Fellowship and UTA Dissertation Fellowship.

Authors' contributions

AM, DNM, ME and EB contributed to the idea and study design. AM conducted all the comparative genomics analyses and further annotation, processing of the hits and figures and tables preparation. AM wrote the manuscript. DNM contributed to processing of the hits and phylogenetic tree analyses and figures and tables preparation. DNM performed the rate of evolution analyses and provided editorial feedback on the manuscript. ME helped run the gene duplication searches in *Drosophila* species genomes obtained from Ensembl and in the initial comparative genomics analyses. EB was a major contributor to data processing and analyses and the writing of the manuscript. All authors read and approved the final manuscript.

Acknowledgements

The authors would like to thank two former members of the Betrán Lab, Susana Domingues and Anya Williford, that contributed to initial explorations of these patterns. Figure 1 and Figure 2 were drawn using BioRender.

References

1. Knockenhauer KE, Schwartz TU: **The Nuclear Pore Complex as a Flexible and Dynamic Gate**. *Cell* 2016, **164**(6):1162-1171.

2. Miyamoto Y, Yoneda Y, Oka M: **17 - Protein Transport Between the Nucleus and Cytoplasm**. In: *Nuclear Architecture and Dynamics.* Edited by Lavelle C, Victor J-M, vol. 2. Boston: Academic Press; 2018: 387-403.

3. Hampoelz B, Andres-Pons A, Kastritis P, Beck M: **Structure and Assembly of the Nuclear Pore Complex**. *Annu Rev Biophys* 2019, **48**:515-536.

4. Kosyna FK, Depping R: **Controlling the Gatekeeper: Therapeutic Targeting of Nuclear Transport**. *Cells* 2018, **7**(11):221.

5. Macara IG: **Transport into and out of the nucleus**. *Microbiol Mol Biol Rev* 2001, **65**(4):570-594.

6. Wente SR, Rout MP: **The nuclear pore complex and nuclear transport**. *Cold Spring Harbor perspectives in biology* 2010, **2**(10):a000562-a000562.

7. Isgro TA, Schulten K: **Association of Nuclear Pore FG-repeat Domains to NTF2 Import and Export Complexes**. *Journal of Molecular Biology* 2007, **366**(1):330-345.

8. Köhler A, Hurt E: **Exporting RNA from the nucleus to the cytoplasm**. *Nature Reviews Molecular Cell Biology* 2007, **8**(10):761-773.

9. Arts GJ, Fornerod M, Mattaj IW: **Identification of a nuclear export receptor for tRNA**. *Curr Biol* 1998, **8**(6):305-314.

10. Kutay U, Lipowsky G, Izaurralde E, Bischoff FR, Schwarzmaier P, Hartmann E, Görlich D: **Identification of a tRNA-specific nuclear export receptor**. *Mol Cell* 1998, **1**(3):359-369.

11. Erkmann JA, Kutay U: **Nuclear export of mRNA: from the site of transcription to the cytoplasm**. *Exp Cell Res* 2004, **296**(1):12-20.

12. Wagstaff KM, Jans DA: **Importins and beyond: non-conventional nuclear transport mechanisms**. *Traffic* 2009, **10**(9):1188-1198.

13. Cautain B, Hill R, de Pedro N, Link W: **Components and regulation of nuclear transport processes**. *Febs j* 2015, **282**(3):445-462.

14. Pemberton LF, Paschal BM: **Mechanisms of receptor-mediated nuclear import and nuclear export**. *Traffic* 2005, **6**(3):187-198.

15. Yoshida K, Blobel G: **The karyopherin Kap142p/Msn5p mediates nuclear import and nuclear export of different cargo proteins**. *J Cell Biol* 2001, **152**(4):729-740.

16. Malik HS, Eickbush TH, Goldfarb DS: **Evolutionary specialization of the nuclear targeting apparatus**. *Proc Natl Acad Sci U S A* 1997, **94**(25):13738-13742.

17. Mason DA, Stage DE, Goldfarb DS: **Evolution of the metazoan-specific importin alpha gene family**. *J Mol Evol* 2009, **68**(4):351-365.

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

19. Tracy C, Río J, Motiwale M, Christensen SM, Betrán E: **Convergently Recruited Nuclear Transport Retrogenes Are Male Biased in Expression and Evolving Under Positive Selection in Drosophila**. *Genetics* 2010, **184**(4):1067.

20. Phadnis N, Hsieh E, Malik HS: **Birth, death, and replacement of karyopherins in Drosophila**. *Mol Biol Evol* 2012, **29**(5):1429-1440.

21. Betrán E, Long M: **Dntf-2r, a Young Drosophila Retroposed Gene With Specific Male Expression Under Positive Darwinian Selection**. *Genetics* 2003, **164**(3):977.

22. Bai Y, Casola C, Betrán E: **Evolutionary origin of regulatory regions of retrogenes in Drosophila**. *BMC Genomics* 2008, **9**(1):241.

23. Krasnov AN, Kurshakova MM, Ramensky VE, Mardanov PV, Nabirochkina EN, Georgieva SG: **A retrocopy of a gene can functionally displace the source gene in evolution**. *Nucleic Acids Research* 2005, **33**(20):6654-6661.

24. VanKuren NW, Long M: **Gene duplicates resolving sexual conflict rapidly evolved essential gametogenesis functions**. *Nat Ecol Evol* 2018, **2**(4):705-712.

25. Presgraves DC: **Does genetic conflict drive rapid molecular evolution of nuclear transport genes in Drosophila?** *Bioessays* 2007, **29**(4):386-391.

26. Presgraves DC, Stephan W: **Pervasive Adaptive Evolution among Interactors of the Drosophila Hybrid Inviability Gene, Nup96**. *Molecular Biology and Evolution* 2007, **24**(1):306-314.

27. Courret C, Chang CH, Wei KH, Montchamp-Moreau C, Larracuente AM: **Meiotic drive mechanisms: lessons from Drosophila**. *Proc Biol Sci* 2019, **286**(1913):20191430.

28. Larracuente AM, Presgraves DC: **The Selfish Segregation Distorter Gene Complex of Drosophila melanogaster**. *Genetics* 2012, **192**(1):33.

29. Tao Y, Araripe L, Kingan SB, Ke Y, Xiao H, Hartl DL: **A sex-ratio Meiotic Drive System in Drosophila simulans. II: An X-linked Distorter**. *PLOS Biology* 2007, **5**(11):e293.

30. Pieper KE, Unckless RL, Dyer KA: **A fast-evolving X-linked duplicate of importin-α2 is overexpressed in sex-ratio drive in Drosophila neotestacea**. *Molecular Ecology* 2018, **27**(24):5165-5179.

31. Brennecke J, Aravin AA, Stark A, Dus M, Kellis M, Sachidanandam R, Hannon GJ: **Discrete small RNA-generating loci as master regulators of transposon activity in Drosophila**. *Cell* 2007, **128**(6):1089-1103.

32. Gell SL, Reenan RA: **Mutations to the piRNA pathway component aubergine enhance meiotic drive of segregation distorter in Drosophila melanogaster**. *Genetics* 2013, **193**(3):771-784.

33. Handler D, Meixner K, Pizka M, Lauss K, Schmied C, Gruber FS, Brennecke J: **The genetic makeup of the Drosophila piRNA pathway**. *Mol Cell* 2013, **50**(5):762-777.

34. Yashiro R, Murota Y, Nishida KM, Yamashiro H, Fujii K, Ogai A, Yamanaka S, Negishi L, Siomi H, Siomi MC: **Piwi Nuclear Localization and Its Regulatory Mechanism in Drosophila Ovarian Somatic Cells**. *Cell Rep* 2018, **23**(12):3647-3657.

35. Graveley BR, Brooks AN, Carlson JW, Duff MO, Landolin JM, Yang L, Artieri CG, van Baren MJ, Boley N, Booth BW *et al*: **The developmental transcriptome of Drosophila melanogaster**. *Nature* 2011, **471**(7339):473-479.

36. Brown JB, Boley N, Eisman R, May GE, Stoiber MH, Duff MO, Booth BW, Wen J, Park S, Suzuki AM *et al*: **Diversity and dynamics of the Drosophila transcriptome**. *Nature* 2014, **512**(7515):393-399.

37. Harrison PM, Milburn D, Zhang Z, Bertone P, Gerstein M: **Identification of pseudogenes in the Drosophila melanogaster genome**. *Nucleic Acids Res* 2003, **31**(3):1033-1037.

38. Mason DA, Fleming RJ, Goldfarb DS: **Drosophila melanogaster importin alpha1 and alpha3 can replace importin alpha2 during spermatogenesis but not oogenesis**. *Genetics* 2002, **161**(1):157-170.

39. Ratan R, Mason DA, Sinnot B, Goldfarb DS, Fleming RJ: **Drosophila importin alpha1 performs paralog-specific functions essential for gametogenesis**. *Genetics* 2008, **178**(2):839-850.

40. Miyamoto Y, Hieda M, Harreman MT, Fukumoto M, Saiwaki T, Hodel AE, Corbett AH, Yoneda Y: **Importin alpha can migrate into the nucleus in an importin beta- and Ran-independent manner**. *Embo j* 2002, **21**(21):5833-5842.

41. Gorjánácz M, Török I, Pomozi I, Garab G, Szlanka T, Kiss I, Mechler BM: **Domains of Importin-alpha2 required for ring canal assembly during Drosophila oogenesis**. *J Struct Biol* 2006, **154**(1):27-41.

42. Chintapalli VR, Wang J, Dow JAT: **Using FlyAtlas to identify better Drosophila melanogaster models of human disease**. *Nature Genetics* 2007, **39**(6):715-720.

43. Yang Z, Nielsen R: **Codon-substitution models for detecting molecular adaptation at individual sites along specific lineages**. *Mol Biol Evol* 2002, **19**(6):908-917.

44. Yang Z: **Likelihood ratio tests for detecting positive selection and application to primate lysozyme evolution**. *Mol Biol Evol* 1998, **15**(5):568-573.

45. Yang Z, Nielsen R: **Synonymous and nonsynonymous rate variation in nuclear genes of mammals**. *J Mol Evol* 1998, **46**(4):409-418.

46. Gao F, Chen C, Arab DA, Du Z, He Y, Ho SYW: **EasyCodeML: A visual tool for analysis of selection using CodeML**. *Ecology and evolution* 2019, **9**(7):3891-3898.

47. Zhou Q, Zhang G, Zhang Y, Xu S, Zhao R, Zhan Z, Li X, Ding Y, Yang S, Wang W: **On the origin of new genes in Drosophila**. *Genome Res* 2008, **18**(9):1446-1455.

48. Betrán E, Thornton K, Long M: **Retroposed new genes out of the X in Drosophila**. *Genome Res* 2002, **12**(12):1854-1859.

49. Kaessmann H, Vinckenbosch N, Long M: **RNA-based gene duplication: mechanistic and evolutionary insights**. *Nat Rev Genet* 2009, **10**(1):19-31.

50. Caporilli S, Yu Y, Jiang J, White-Cooper H: **The RNA export factor, Nxt1, is required for tissue specific transcriptional regulation**. *PLoS Genet* 2013, **9**(6):e1003526.

51. Kusano A, Staber C, Ganetzky B: **Nuclear Mislocalization of Enzymatically Active RanGAP Causes Segregation Distortion in Drosophila**. *Developmental Cell* 2001, **1**(3):351-361.

52. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ: **Gapped BLAST and PSI-BLAST: a new generation of protein database search programs**. *Nucleic Acids Res* 1997, **25**(17):3389-3402.

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

54. Eslamieh M, Williford A, Betrán E: **Few Nuclear-Encoded Mitochondrial Gene Duplicates Contribute to Male Germline-Specific Functions in Humans**. *Genome Biol Evol* 2017, **9**(10):2782-2790.

55. Attrill H, Falls K, Goodman JL, Millburn GH, Antonazzo G, Rey AJ, Marygold SJ: **FlyBase: establishing a Gene Group resource for Drosophila melanogaster**. *Nucleic Acids Res* 2016, **44**(D1):D786-792.

56. Artimo P, Jonnalagedda M, Arnold K, Baratin D, Csardi G, de Castro E, Duvaud S, Flegel V, Fortier A, Gasteiger E *et al*: **ExPASy: SIB bioinformatics resource portal**. *Nucleic Acids Research* 2012, **40**(W1):W597-W603.

57. Lu S, Wang J, Chitsaz F, Derbyshire MK, Geer RC, Gonzales NR, Gwadz M, Hurwitz DI, Marchler GH, Song JS *et al*: **CDD/SPARCLE: the conserved domain database in 2020**. *Nucleic Acids Res* 2020, **48**(D1):D265-d268.

58. Marchler-Bauer A, Bo Y, Han L, He J, Lanczycki CJ, Lu S, Chitsaz F, Derbyshire MK, Geer RC, Gonzales NR *et al*: **CDD/SPARCLE: functional classification of proteins via subfamily domain architectures**. *Nucleic Acids Res* 2017, **45**(D1):D200-d203.

59. Yang Z: **PAML 4: Phylogenetic Analysis by Maximum Likelihood**. *Molecular Biology and Evolution* 2007, **24**(8):1586-1591.

60. Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, Lopez R, McWilliam H, Remmert M, Söding J *et al*: **Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega**. *Mol Syst Biol* 2011, **7**:539.

61. Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C *et al*: **Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data**. *Bioinformatics* 2012, **28**(12):1647-1649.

62. Guindon S, Dufayard J-F, Lefort V, Anisimova M, Hordijk W, Gascuel O: **New Algorithms and Methods to Estimate Maximum-Likelihood Phylogenies: Assessing the Performance of PhyML 3.0**. *Systematic Biology* 2010, **59**(3):307-321.

63. Nei M, Gojobori T: **Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions**. *Mol Biol Evol* 1986, **3**(5):418-426.

64. Anisimova M, Bielawski JP, Yang Z: **Accuracy and power of the likelihood ratio test in detecting adaptive molecular evolution**. *Mol Biol Evol* 2001, **18**(8):1585-1592.

A.

B.

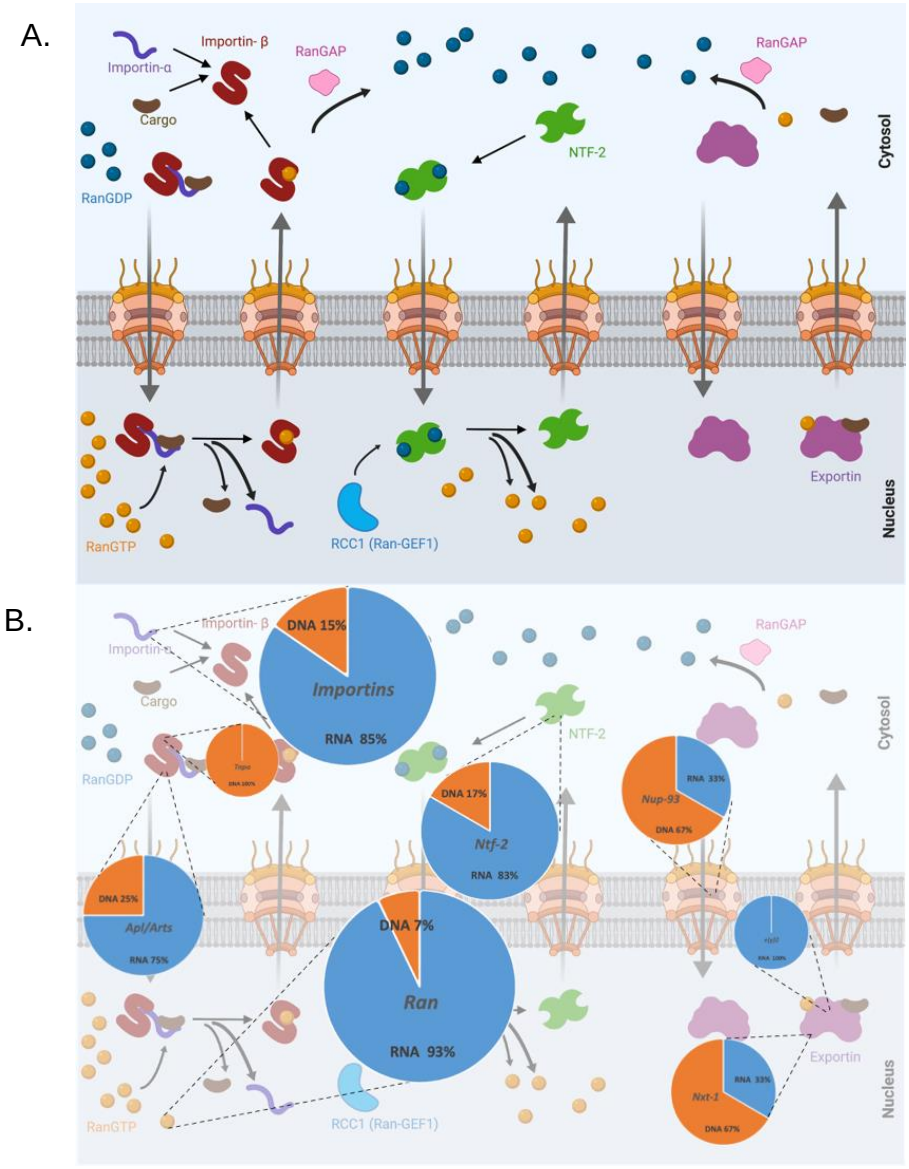


Figure 1. A. Schematic view of conventional nuclear transport (Redrawn from [6, 7] with modifications). B. Percentages of RNA- *vs.* DNA-mediated duplications for each gene. Size of the pie charts are relative to the total number of independent duplication events for that gene.

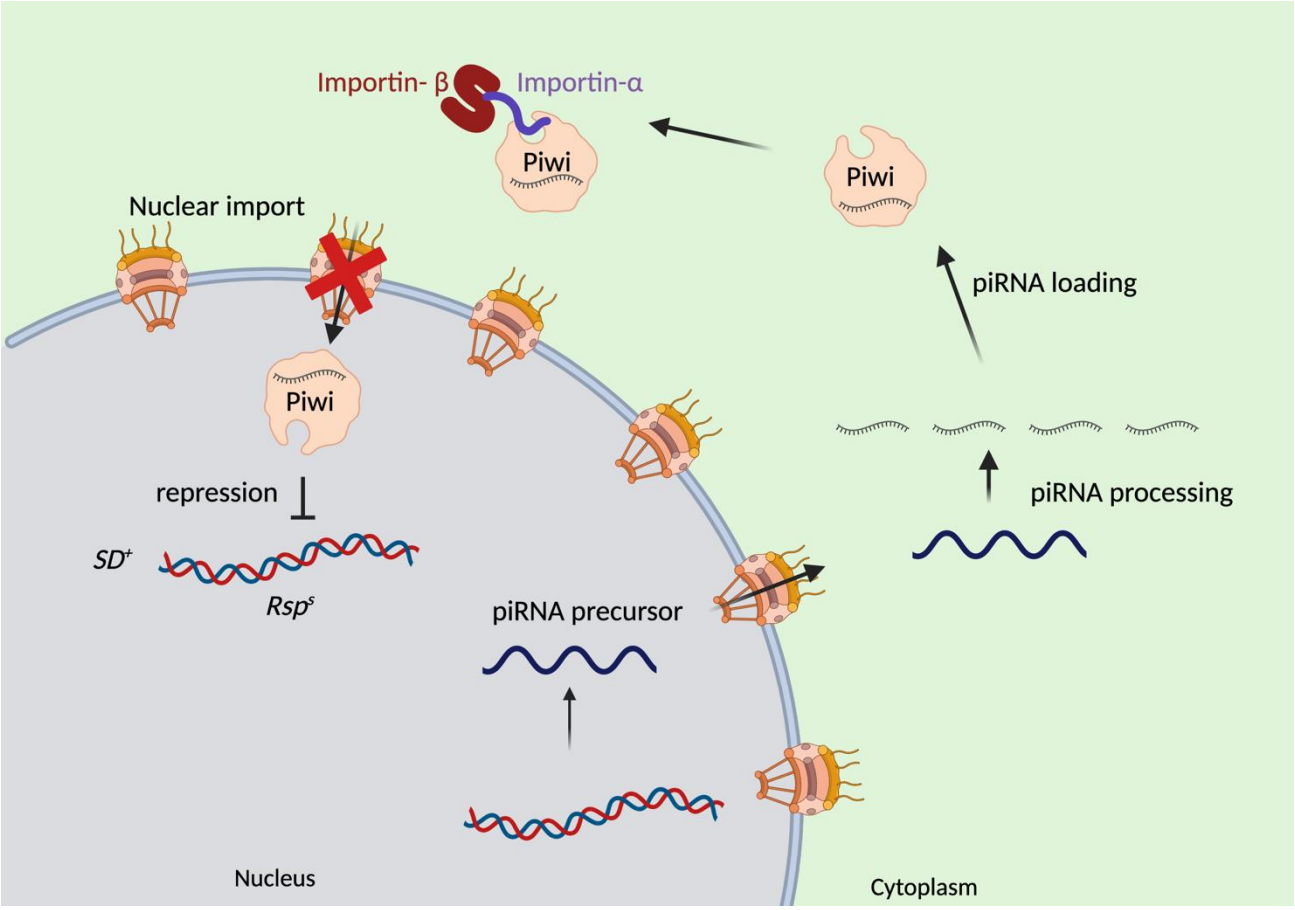
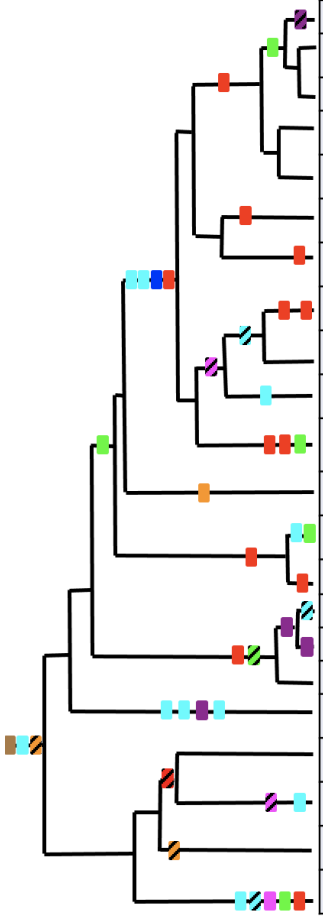


Figure 2. Proposed model of how the piRNA pathway might be involved in the SD system in *Drosophila*. piRNA precursors are exported from the nucleus to the cytoplasm, where piRNA biogenesis and loading of the RNA to the Piwi protein happens. Import of this complex to the nucleus has been proposed to be needed for chromatin condensation of the *Rsp* repeats and in SD it might not occur due to disruption of Ran-GTP/Ran-GDP gradient [28, 33, 34].



Gene family	Ntf-2	Ran	Importin-α	Nup93	Nxt-1	e(y)2b	Tnpo	Importin-β
D. melanogaster	Ntf-2r(1)	Ran-like(2)	aKap4	**Nup93-like**	-	e(y)2	**CG8219**	**Apl / Arts**
D. simulans	Ntf-2r(1)	Ran-like(2) / Ran-like(3)	aKap4	**Nup93-like**	-	e(y)2	**CG8219**	-
D. sechellia	Ntf-2r(1)	Ran-like(2) / Ran-like(3)	aKap4	**Nup93-like**	-	e(y)2	-	-
D. yakuba	-	-	aKap4	**Nup93-like**	-	e(y)2	**CG8219**	-
D. erecta	-	Ran-like(2) / Ran-like(3)	aKap4	**Nup93-like**	-	e(y)2	**CG8219**	-
D. takahashi	-	Ran-like(4)	aKap5/ aKap6	**Nup93-like**	-	e(y)2	-	-
D. biarmipes	Ntf-2r(2)	Ran-like(5)	aKap5/ aKap6	**Nup93-like**	-	e(y)2	**CG8219**	-
D. rhopaloa	Ntf-2r(2)	Ran-like(3)/ Ran-like(6)/ Ran-like(7)	aKap4/aKap6/ **aKap7**	**Nup93-like**	**Nxt1-like(1)**	e(y)2	-	**-**
D. elegans	-	-	aKap6/**aKap7**	**Nup93-like**	-	e(y)2	**CG8219**	-
D. eugracilis	-	-	aKap2D/ aKap6	**Nup93-like**	-	e(y)2	**CG8219**	-
D. ficusphila	Ntf-2r(3)	Ran-like(8)/Ran-like(9)	aKap6	**Nup93-like**	-	e(y)2	-	-
D. kikkawai	-	-	-	**Nup93-like/** Nup93-like(2)	-	e(y)2	-	-
D. ananassae	Ntf-2r(2)/ Ntf-2r(4)	Ran-like(10)	aKap2B/ aKap5	**Nup93-like**	-	e(y)2	-	-
D. bipectinata	Ntf-2r(2)	Ran-like(11)	-	**Nup93-like**	-	e(y)2	-	-
D. pseudoobscura	**Ntf-2r(5)**	Ran-like(1)	aKap5/ **aKap5B**	**Nup93-like**	-	e(y)2	-	Apl-like(2)
D. persimilis	**Ntf-2r(5)**	Ran-like(1)	aKap5	**Nup93-like**	-	e(y)2	-	Apl-like(2)/(3)
D. miranda	**Ntf-2r(5)**	Ran-like(1)	aKap5	**Nup93-like**	-	e(y)2	-	-
D. willistoni	-	-	aKap5/aKap2E/ aKap2F/aKap2G	**Nup93-like**	-	e(y)2	-	Apl-like(4)
D. mojavensis	-	**Ran-like(12)**	-	**Nup93-like**	-	e(y)2	-	-
D. virilis	-	**Ran-like(12)**	aKap5/aKap2C	**Nup93-like**	**Nxt1-like(2)**	e(y)2	-	-
D. albomicans	-	-	aKap5	**Nup93-like/ Nup93-like(3)**	Nxt1-like(3)	e(y)2	-	-
D. grimshawi	Ntf-2r(6)	Ran-like(13)	aKap2A/**aKap2H**	**Nup93-like**	-	e(y)2	-	-

Figure 3. Presence and absence of gene duplicates across 22 species of *Drosophila* is shown. Summary of nuclear transport genes duplication events across the *Drosophila* species is shown in the phylogeny. Each rectangle represents a duplication event. Duplication events are shown at an approximate distance from the tips based on the percent identity to the parental gene protein in that species (Supplementary material 4). DNA-mediated duplications are shown in bold in the table and with striped boxes in the phylogeny.

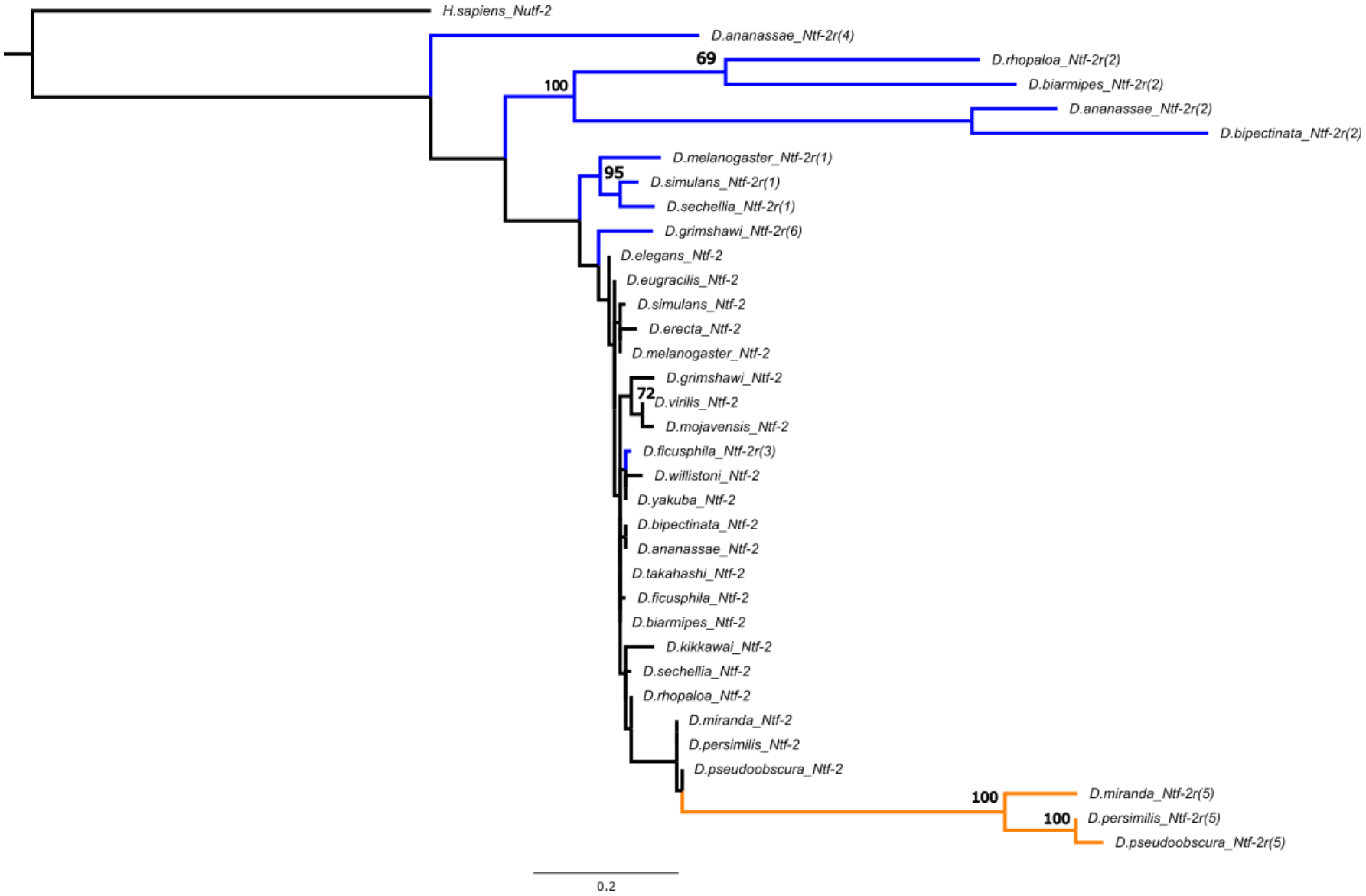


Figure 4. A. Maximum-likelihood tree constructed using PhyML showing the phylogenetic relationships between Ntf-2 parental and duplicates based on amino acid sequences. DNA mediated duplication are shown in orange and RNA-mediated duplications are shown in blue.

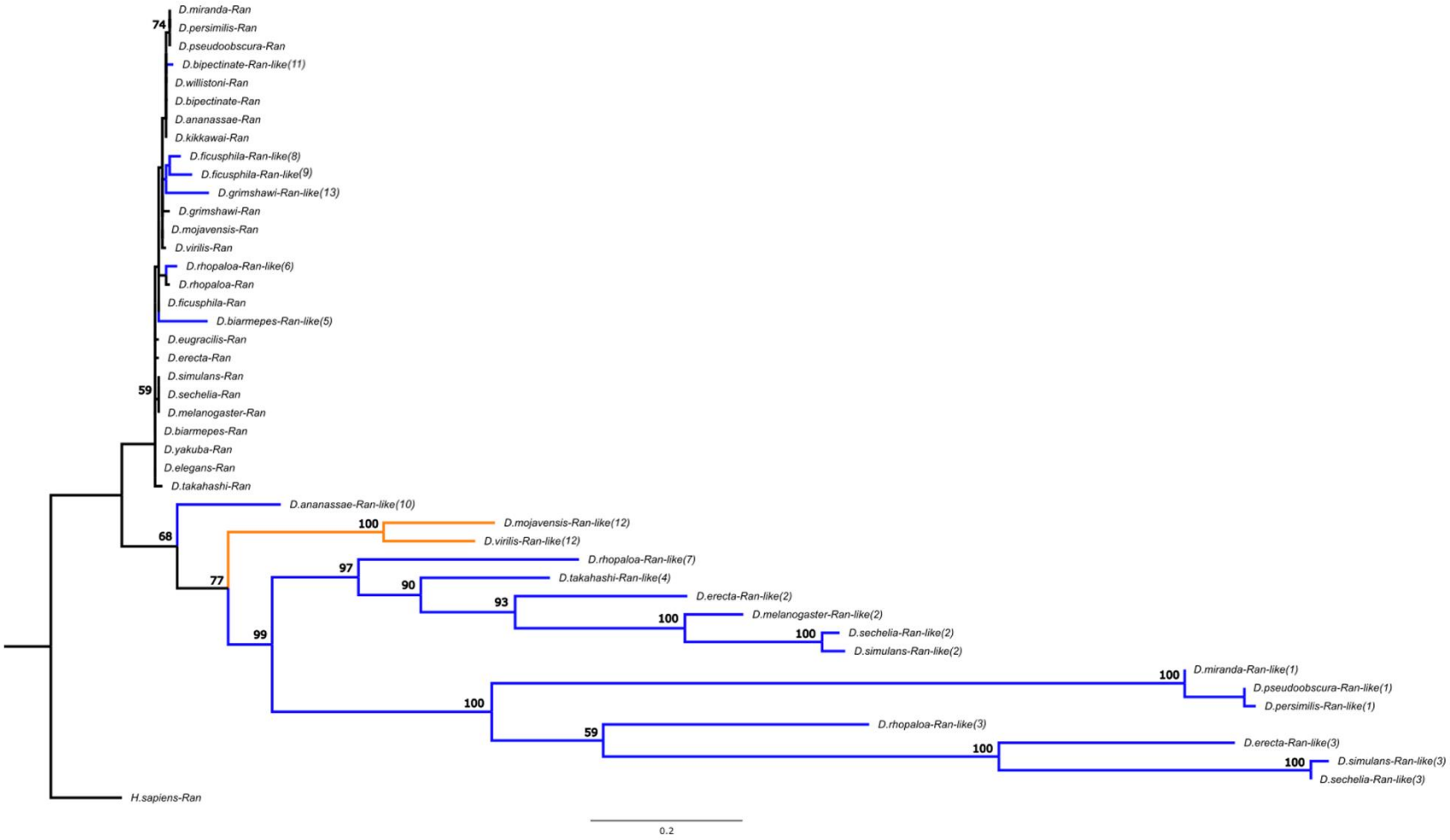


Figure 4. B. Maximum-likelihood tree constructed using PhyML showing the phylogenetic relationships between Ran parental and duplicates based on amino acid sequences. DNA mediated duplication are shown in orange and RNA-mediated duplications are shown in blue.

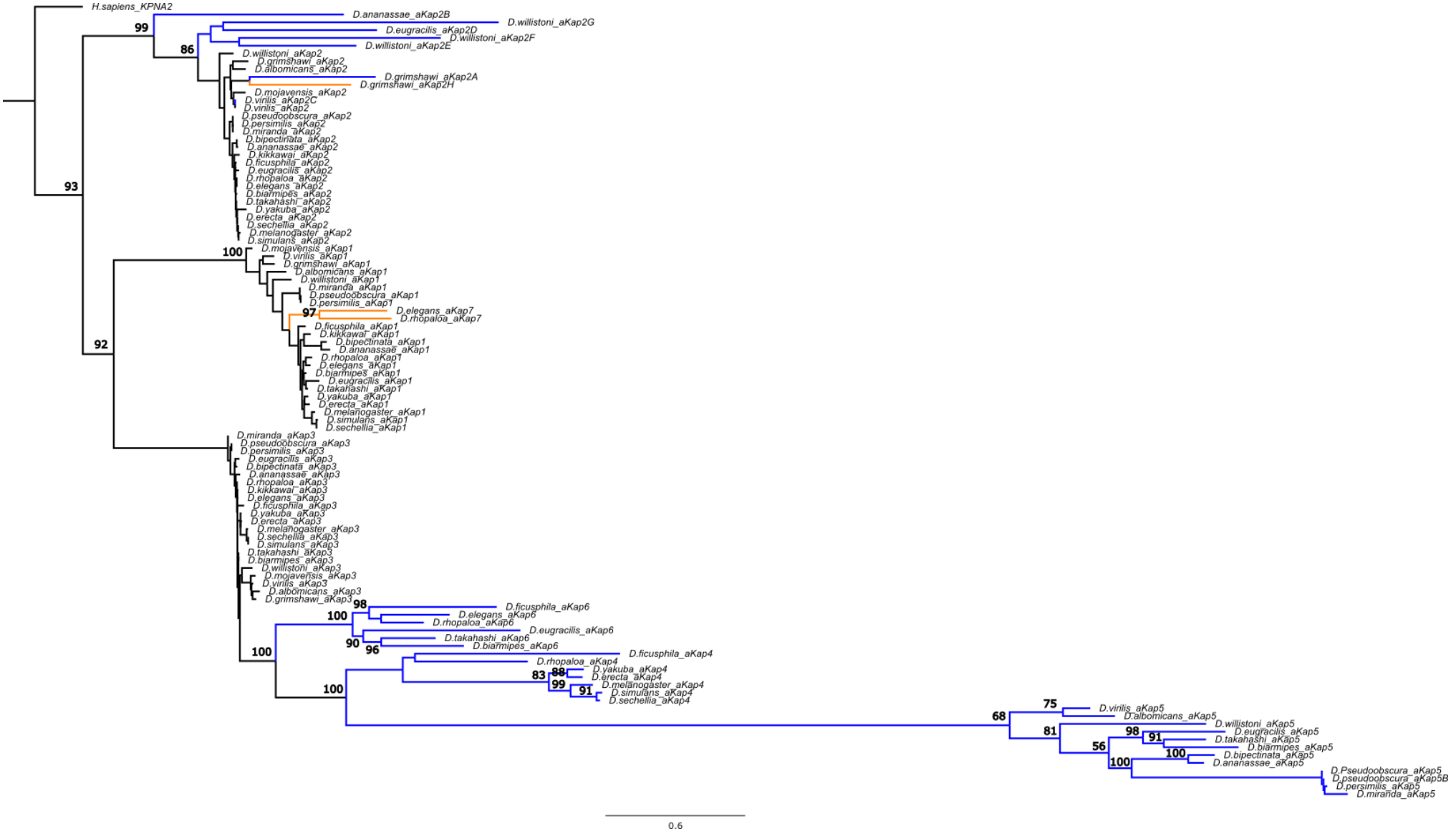


Figure 4. C. Maximum-likelihood tree constructed using PhyML showing the phylogenetic relationships between *Importin-α* parental and duplicates based on amino acid sequences. DNA mediated duplication are shown in orange and RNA-mediated duplications are shown in blue.

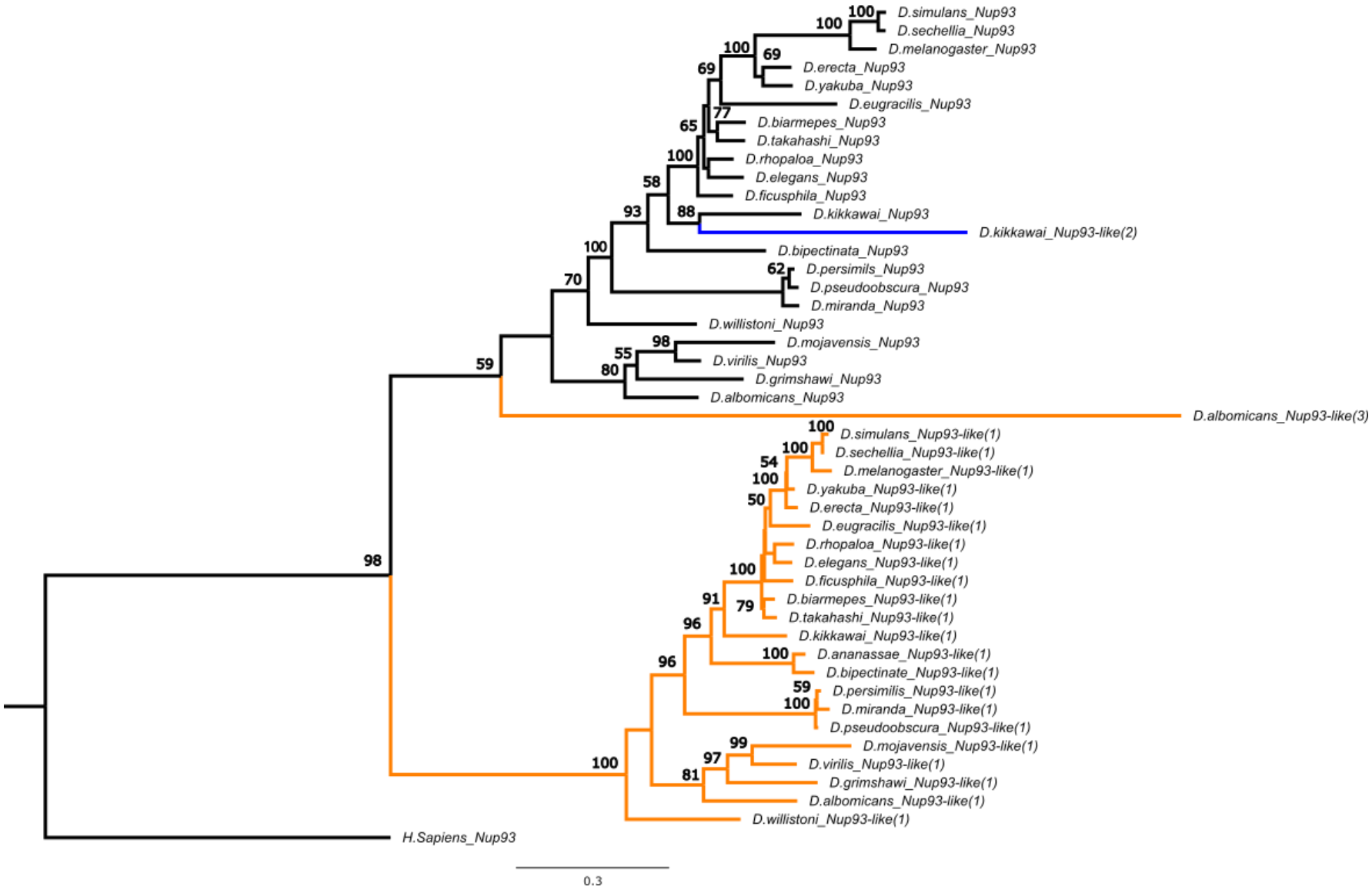


Figure 4. D. Maximum-likelihood tree constructed using PhyML showing the phylogenetic relationships between Nup-93 parental and duplicates based on amino acid sequences. DNA mediated duplication are shown in orange and RNA-mediated duplications are shown in blue.

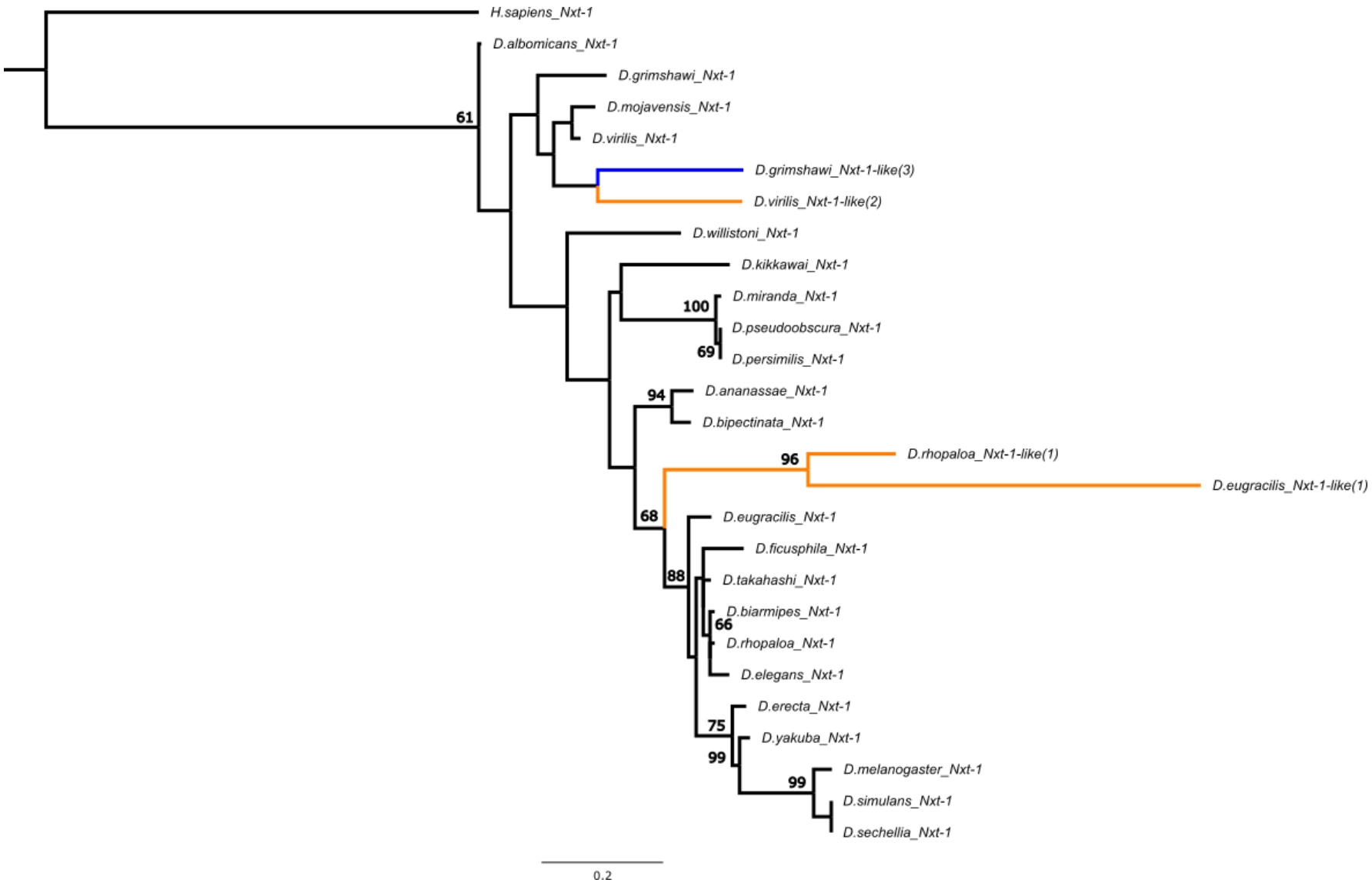


Figure 4. E. Maximum-likelihood tree constructed using PhyML showing the phylogenetic relationships between Nxt-1 parental and duplicates based on amino acid sequences. DNA mediated duplication are shown in orange and RNA-mediated duplications are shown in blue.

Figure 4. F. Maximum-likelihood tree constructed using PhyML showing the phylogenetic relationships between Tnpo parental and duplicates based on amino acid sequences. DNA mediated duplication are shown in orange.

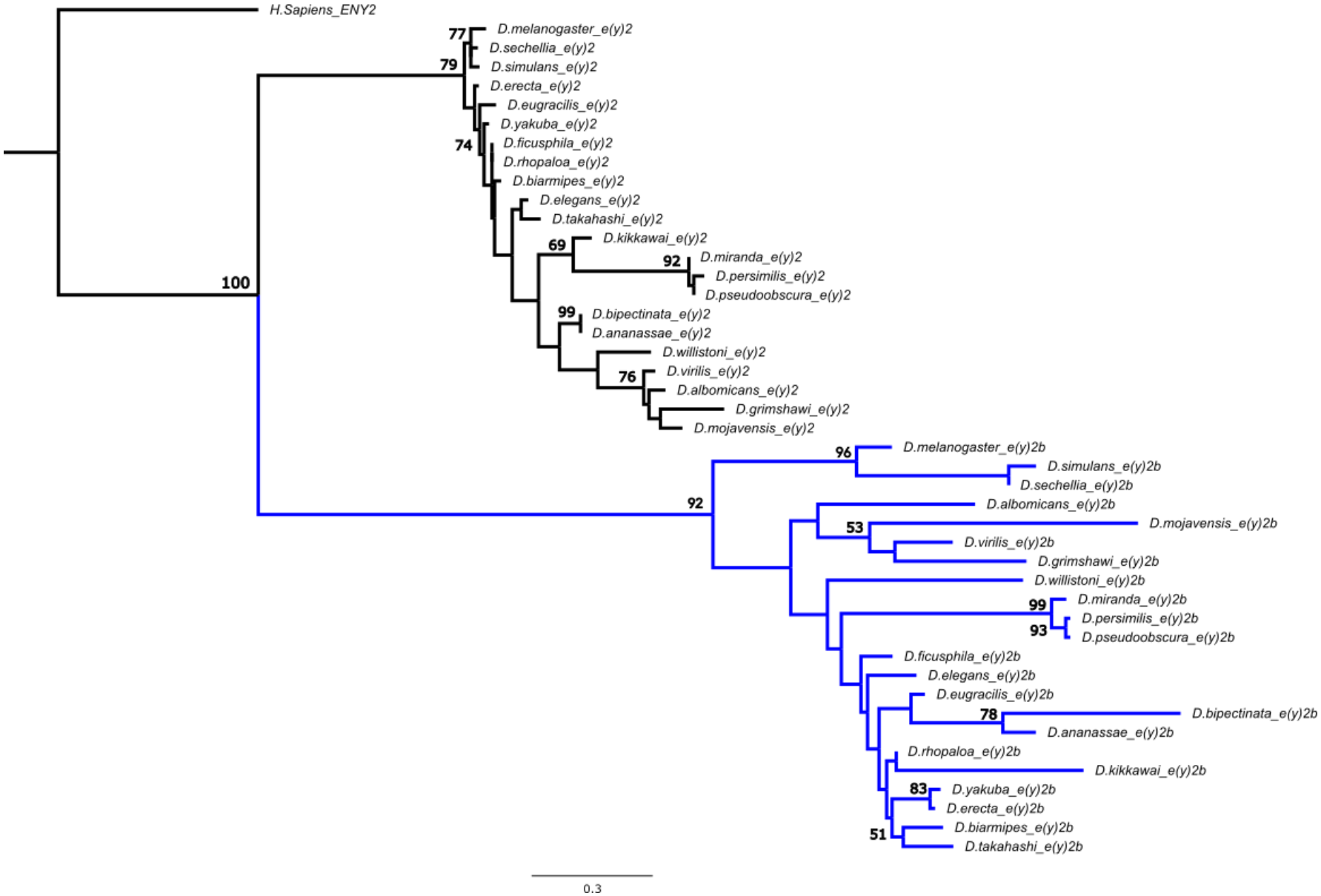


Figure 4. G. Maximum-likelihood tree constructed using PhyML showing the phylogenetic relationships between e(y)2 parental and duplicates based on amino acid sequences. RNA-mediated duplications are shown in blue.

Chapter 3

PLoS Genetics

Functional study of testis-specific nuclear transport retrogenes, *Ntf-2r* and *Ran-like*, in *Drosophila melanogaster*

Ayda Mirsalehi[1#], Susana Domingues[1#], Dragomira N. Markova[1], Diwash Jangam[1], and Esther Betrán[1*]

[1] Department of Biology, the University of Texas at Arlington, Arlington, TX, USA.

# These authors made equal contribution to this work,

* Corresponding author

Esther Betrán

Biology Department, Box 19498

University of Texas at Arlington, Arlington, TX 76019, USA

Phone (817) 272 1446

E-mail: betran@uta.edu

Abstract

Components of the nuclear transport machinery have been involved in genetic innovation in several instances in *Drosophila*. Phylogenetic analyses show that some nuclear transport genes, i.e., *Ran*, *Ntf-2* and some *importin-α* genes, have been recurrently duplicated and the new genes have acquired a highly testis-biased expression. Supposedly, these genes duplicated for a specific function during spermatogenesis, although they have sometimes been lost in some species. Molecular evolution studies have shown that several nucleoporins, *RanGAP* and the duplicates of *Ran* and *Ntf-2* have evolved under positive selection, i.e., their proteins have been changing faster than the neutral expectations at those sites. Because of the high turnover and fast evolution of these testis-biased nuclear transport genes, it is hypothesized that they are involved in suppressing germline conflicts like segregation distortion, i.e., in an arms race with selfish chromosomes. Here we study the function of *Ran* and *Ntf-2* retroduplicates named *Ran-like* and *Ntf-2r*, in *D. melanogaster*, to shed light on the evolutionary pressures that have resulted in their strong selection and expression in testis. *Ran-like* and *Ntf-2r* proteins could be involved in many functions during meiosis and sperm development judging from their localization. However, the null mutants of *Ntf-2r* and *Ran-like* show no fertility effects in standard laboratory conditions. RNAseq analyses of null mutant lines show no upregulation of the parental genes in the retroduplicate mutant backgrounds. Despite very high expression of *Ran-like* compared to *Ntf-2r* in the testis, the loss of *Ntf-2r* appears to affect expression of more genes associated with spermatogenesis, translation, microtubule and mitochondrial gene ontologies. Functional differences between Ran and Ran-like that fit this hypothesis were revealed by the failure of Ran-like to rescue the lethality phenotype of the Ran null mutant. Given the essential cellular function of these kinds of nuclear transport

genes but the lack of strong fertility phenotype and the relatively minor gene expression changes in the *Ran-like* null mutant, we favor the hypothesize that the parental gene (*Ran*) that is highly expressed in testis might still have a function during spermatogenesis and the new gene might be differentiated and expressed to prevent deleterious interactions (meiotic drive systems interactions) with the parental gene. In the case of *Ntf-2r*, the parental gene is not that highly transcribed in testes and there are extensive changes in gene expression in *Ntf-2r* null revealing that it could be a duplicate involve in those functions.

Running title: Function of testis-specific nuclear transport retrogenes

Introduction

Ntf-2 and *Ran* are highly conserved housekeeping genes with essential roles in the transport of cargo proteins between the nucleus and cytoplasm and are required in all eukaryotes (Quimby, et al. 2000). Involvement of *Ntf-2* and *Ran* in nuclear transport has been studied in the budding yeast, *Saccharomyces cerevisiae* as well as in the higher eukaryotes such as *Caenorhabditis elegans* and *Xenopus* (Ribbeck, et al. 1998; Quimby, et al. 2000). The proposed model for the role of these genes in import involves several interactions: dimerization of Ntf-2, formation of Ntf-2-RanGDP complex in the cytoplasm, followed by translocation of this complex to the nucleus, conversion of RanGDP to RanGTP by means of a catalytic enzyme RanGEF (RanGTPase Exchange Factor), also known as RCC1, in the nucleus, and binding of RanGTP to importin-β transport receptors and induction of conformational changes that lead to the dissociation of importin α/β heterodimer and release of the cargo protein. This occurs because RanGTP has no detectable affinity to Ntf-2, and the conversion of RanGDP to RanGTP will result in the RanGTP interaction with importins and the release of the cargo protein in the nucleus. RanGTP bound to importin-β is then transported out of the nucleus where RanGAP transforms RanGTP into RanGDP (Isgro and Schulten 2007). So, the gradient of higher RanGTP in the nucleus leads to the assembly of the cargo protein with importins in the cytoplasm and its release in the nucleus.

Besides the nuclear transport functions, *Ran* has been shown to have crucial roles during mitosis in *Drosophila*, *Xenopus* and *Caenorhabditis elegans*. *Ran* is responsible for microtubule assembly and synchronizing nuclear and chromosomal functions all through the cell-division cycle. The Ran pathway directs chromosome alignment and assembly of the mitotic spindles before metaphase. Also, Ran pathway coordinates post-metaphase

events such as chromosome segregation, nuclear envelope dynamics and timing of cell-cycle transitions (Moore 1998; Peter, et al. 2002; Kaláb, et al. 2006; Silverman-Gavrila and Wilde 2006; Clarke and Zhang 2008).

Unexpectedly, *Ntf-2* and *Ran* that are usually single copy genes in eukaryotes have been recurrently duplicated in *Drosophila*. This has occurred recurrently as well for *importins* in different *Drosophila* lineages. Losses of gene duplicates (i.e., high turnover), fast evolution, and particular patterns of gene evolution and testis-biased expression have also been observed (Betrán and Long 2003; Bai, et al. 2007; Tracy, et al. 2010; Phadnis, et al. 2012). These gene duplications as well as the fast evolution of additional components of the nuclear transport such as nucleoporins and RanGAP have been hypothesized to occur to defend the male germline against meiotic drive systems similar to the Segregation Distorter (*SD*) system of *Drosophila melanogaster* (Presgraves 2007; Presgraves and Stephan 2007; Larracuente and Presgraves 2012; Phadnis, et al. 2012).

SD system of *Drosophila melanogaster* is a young and well-described meiotic drive system and involves the duplication of a nuclear transport gene (Presgraves, et al. 2009). SD is comprised of three interacting loci linked to the centromere of chromosome 2, 1) Segregation distorter (*Sd*), this is the main distorting locus and is a truncated tandem duplication of the *RanGAP* gene, 2) Enhancer of *SD* (*E(SD)*), 3) Responder (*Rsp*), which is a cis-acting target locus. *SD* is a meiotic drive system causing deviations from Mendel's first law in which although heterozygous *SD/SD*$^{+}$ females pass SD and *SD*$^{+}$ equally to their progenies according to the Mendelian ratio of 1:1, 99% of heterozygous *SD*/ *SD*$^{+}$ male progenies inherit *SD* bearing chromosomes. *Sd* performs the same enzymatic activity as the wild-type RanGAP. However, since it is truncated it lacks its nuclear export signal (NES), and it mislocalizes to the nucleus resulting in accumulation of RanGDP molecules

in the nucleus (Presgraves 2007). Disruption of the RanGTP gradient might cause segregation distortion. The reason why only *SD*-bearing sperms survive spermatogenesis is disruption of chromatin remodelling in SD^+-bearing spermatid nuclei and failing to transition between histones and protamines to mature into functional sperm, resulting in their accumulation in the waste bags (Larracuente and Presgraves 2012).

To clarify the evolutionary pressures that give rise to the recurrent emergence of new nuclear transport genes with testis-specific expression, in particular, *Ran-like* and *Ntf-2r* in *Drosophila melanogaster,* we have studied function of these retroduplicates during spermatogenesis.

Results

Ntf-2r and Ran-like cellular localization and co-localization during spermatogenesis

Localization and co-localization of Ntf-2r and Ran-like during spermatogenesis was studied using *Ntf-2r-EGFP* and *Ran-like-DsRed.T4 P-element* gene fusion transgenes. In these transformed strains, proteins are tagged with fluorescent proteins and contain the upstream endogenous regulatory regions for the gene (See Materials and Methods). Fluorescence for both proteins was observed starting at the 16-cell stage in the primary spermatocytes that will start meiosis for both genes (Figure 1). See confirmation of this below. During the 16-cell stage in the primary spermatocytes, Ntf-2r is present in the cytoplasm, nuclear membrane and inside the nucleus. Ran-like co-localizes with Ntf-2r in the nuclear membrane and nucleus but is not present in the other cellular compartments in these undividing meiotic cells (Figure 2.A). Throughout meiosis, in the meiotic spermatocytes, Ran-like was observed localizing with the cell spindles and at the cell poles in the dividing cells. Ntf-2r localization is more disperse. Ntf-2r is present in the same

structures as Ran-like, but also in the cell cytoplasm (Figure 2.B). The localization of the Ran-like is similar to the known localization of the parental gene (Ran) localizing around the spindles in mitosis (Trieselmann and Wilde 2002) which would suggest at least few similar interactions for these genes. In round spermatids, i.e., at the onion stage of spermatogenesis when the mitochondria have fused into a round multilayered dense structure, bundles of 64 cells can be found along the testis just before the mitochondria start elongating. At this stage, both retrogenes are in a structure named dense body that is surrounding the nucleus where the nuclear pores are accumulating, and microtubules are assembling (Fabian and Brill 2012). They are also present at the beginning and at the end of the mitochondria where microtubules are organizing for tail axoneme and mitochondria elongation. Ntf-2r can also be observed overlaying with the cell nucleus (Figure 2.C). When the 64 cell bundles are formed, the mitochondria and the nucleus will start elongating to make the sperm bundles and give rise to mobile sperms. At this stage Ran-like and Ntf-2r show co-localization in the dense body and at the mitochondrial poles at the ends of two elongating fused mitochondria (Figure 2.D). After sperm elongation, Ntf-2r is present all along the sperm bundles, however Ran-like seems to have a very precise localization along the sperm bundle tails (Figure 2.E). From the localization, in addition to the nuclear transport functions, both genes appear to have many interactions from spermatocytes onwards related to nuclear membrane assembly, microtubule organization during chromosomal segregation, and sperm head and tail elongation.

Ran-like co-localization with known spermatogenesis proteins

Anti lamin protein antibody was used to label the first dividing cells in spermatogenesis (See Materials and Methods). This antibody stains the nuclear

membrane of 2, 4, 8 and 16 cells stages (Shevelyov, et al. 2009). Co-localization between Ran-like and lamin was observed in 16-cell stage (i.e., primary spermatocytes; Figure 3. A and B). Ran-like interaction with microtubule nucleation was studied using anti alpha-tubulin antibody (Dorogova, et al. 2008). Ran-like co-localizes with alpha-tubulin (one of the two components that make up microtubules) in the cell spindles during meiosis close to the DNA (Figure 3.C) as well as at the microtubules. Using the same antibody, colocalization of alpha-tubulin with Ran-like during sperm head elongation is observed in the dense body (Figure 3.D). Alpha-tubulin also appears to localize at the end of microtubules during tail elongation and close to the DNA in the dense body of elongating sperm bundles heads (Figure 3.D). Phalloidin and DAPI were used to stain DNA and F-actin cones during sperm individualization. At this stage of spermatogenesis, F-actin cones are responsible for the progress of individualization and removal of undesired proteins from the sperm bundles, breaking the bridges between the haploid cells and leading to individual mobile sperm. Ran-like is close and at the distal end of the F-actin cones (Figure 3.E) at the head of sperm bundles (Noguchi, et al. 2012). Ran-like localization is very similar to the localization of the proteasomes, protein complexes that degrade proteins as individualization proceeds (Zhong and Belote 2007). However, proteasomes are not directly involved in Ran-like degradation as I observe that Ran-like remains undigested in the waste bags (Figure 3.F). Proteins that are no longer required and have not been degraded by proteasomes, accumulate in the waste bags (Ghosh-Roy, et al. 2005).

Ntf-2r **and** ***Ran-like*** **are not essential for male fertility**

Taking advantage of the CRISPR-Cas9 technique combined with homologous recombination, knockout (KO) mutants were generated for *Ntf-2r* and *Ran-like* where the

genes have been replaced by a Ds-Red protein driven in the eye (See Materials and Methods). A double mutant of the KOs of *Ntf-2r* and *Ran-like* was also generated. The knockout strains were confirmed by doing PCR as well as the observation of glowing in the eyes of the flies. Given the strong expression of *Ntf-2r* and *Ran-like* in the male germline and their broad and precise protein localization during spermatogenesis, we hypothesize that removing the function of either *Ntf-2r* or *Ran-like* would lead to fertility effects. However, we have not observed any major sterility phenotype in the KOs of these genes (Figures 4 and 5). No fertility effect is observed for the knock outs in the first four days. In the second four days, *Ntf-2r*-KO males were having significantly lower number of progeny compared to the line of control (P=0.0095 in a t-test) (Figure 4). From this comparison it seems that *Ntf-2r*-KO males are aging faster in respect to sperm production but there is an interaction because this effect is not present in the double mutant.

We also performed a male sperm exhaustion assay modified from Sun et al. 2004 and Flores et al. 2015. See Meterials and Methods for details. Sperm exhaustion assay is a more stringent fertility assay to detect subtle differences in sperm production or recovery in KOs compared to the strain of control (w^{1118}) (Sun, et al. 2004; Flores, et al. 2015). No significant fertility reduction was observed for the knockout strains compared to the control strain in the sperm exhaustion assay (results for the first 5 days are shown in the figure). However, we see that in day five, *Ntf-2r*-KO males were having significantly higher number of progeny compared to the line of control (P=0.02) (Figure 5).

Ntf-2-PB cellular localization during spermatogenesis

To study the protein localization of the parental genes and compare it to the duplicates, we attempted the tagging of the parental genes with hemagglutinin (HA) at their

endogenous sites using CRISPR-Cas9 technology. We were able to produce Ntf-2-PB-HA, i.e., a tag at the C-terminus of the Ntf-2-PB isoform, while Ntf-2-PA-HA was detected only in the females as heterozygote and could not be fixed or observed in males. Ran-HA is only detected in the pooled first set of offspring of the injected individuals but gets lost in the generation after. Tagging these genes at the N-terminus showed similar results. We conclude that there is susceptibility of these genes to homozygous modification affecting hemizygous males and preventing us from studying their expression during spermatogenesis. We also infer a different function for Ntf-2-PA and Ntf-2-PB from these observations.

Immunostaining of Ntf-2-PB-HA shows presence of its protein starting at the tip of the testis. The distribution does not resemble that of Ntf-2r. It is not diffuse in the cytoplasm, at the nuclear membrane and at structures where microtubules are organized. It appears in foci in the sperm tails of sperm bundles mimicking Ran-like localization but with higher density and overlapping DNA in some cell types. See Figures 6 and 7.

Driving somatic expression of Ran and Ran-like

UAS constructs for *Ran-EGFP* and *Ran-like-EGFP* were made to be able to drive these proteins broadly or in specific tissues (See Materials and Methods). Driving *Ran-EGFP* generates healthy progeny without any viability phenotype regardless of the driver used (arm-GAL4, Act5C-GAL4 and tubP-GAL4). However, driving *Ran-like*-EGFP in somatic tissues caused developmental arrest and lethality (arm-GAL4 and tubP-GAL4 and Act5C-GAL4 during embryogenesis but with arm-GAL4 pupae arrest was observed). It can be inferred from these results and the lines used above (*Ran-like-DsRed.T4*) that

overexpressing *Ran-like* in tissues where it is already expressed has no phenotypic effects while overexpression of *Ran-like* appears to be toxic when it is present in ectopic places. This is inferred to be due to functional differences between *Ran* and *Ran-like.*

Ran knockout is rescued by _Ran_ but not _Ran-like_ expression

From the fast evolution and positive selection of *Ntf-2r* and *Ran-like*, it can be inferred that the function of these genes might have partially changed from the parentals (Betrán and Long 2003; Tracy, et al. 2010). We have used a *P-element* insertion null mutant strains of the *Ran* gene (BL 11800) that shows a lethality phenotype and try to rescue the mutant with UASt-Ran-EGFP and UASt-Ran-like-EGFP. *P-element* insertion null mutant strains of *Ran* was previously rescued by driving Ran (Cesario and McKim 2011) . *Ran* and *Ran-like* constructs for rescue were made using the Gateway system. See more details in Materials and Methods. These constructs contain the UASt upstream of the coding region of the gene and in-frame EGFP downstream of the gene.

Crosses were performed using Act-5C Gal4 driver and UASt-Ran-EGFP and UASt-Ran-like-EGFP to test if the parental strain and/or retroduplicate can rescue the mutant. Young males from the crosses between Act-5C-GAL4 and UASt-Ran-EGFP/UASt-Ran-like-EGFP were selected and crossed with female Ran^{-}/FM7c flies. From the progeny of this cross, the number of males with no Bar phenotype (i.e., non FM7C) and the number of Bar males were recorded and compared to the 1:4 expected ratio under full rescue. The *Ran* mutant lethality phenotype was fully rescued when parental gene was driven (Supporting information S1 Table1). However, the duplicate gene could not rescue the *Ran* mutant phenotype and zero non-Bar males were observed from the second cross.

Differential gene expression analyses in *Ntf-2r*-KO and *Ran-like*-KO

To study differential gene expression between *Ntf-2r*-KO and *Ran-like*-KO and the strain of control (w^{1118}) during spermatogenesis, RNA-Seq analyses were performed. RNA was extracted and RNA-seq was carried out for testes of *Ntf-2r*-KO and *Ran-like*-KO flies from each of the 2 independently backcrossed lines to w^{1118} and compared to two replicates of the control line. See Materials and Methods for details.

Results from FastQ showed that all raw read data passed quality control test to be used for downstream analyses. From our read mapping, we confirmed that no reads from *Ntf-2r* or *Ran-like* CRISPR-Cas9 excised regions were detected for the knockout mutants (Supporting Information S2 Figures 1 & 2) further supporting that we have obtained complete KOs for the genes under study. Correlation of the gene expression levels (FPKM) between samples are shown with Pearson correlation coefficient (Supporting Information S2 Figure 3). The square of the Pearson correlation coefficient values (R^2) between all the samples are greater than 0.8 indicative of a high similarity between all the samples with the highest correlation between replicates but not always and revealing that we were able to introduce the w^{1118} background and the knockouts do not produce large-scale major effects. The samples cluster according to their genotype (Supporting Information S2 Figure 3).

Differential gene expression level analyses were performed to understand if any particular pathways suffer changes that could illuminate gene function. See Materials and Methods. Supporting information S3 Table and S4 Table show all the results related to differentially expressed genes. First, we checked if parental genes (*Ran* or *Ntf-2*) are upregulated in the knockout mutants and might compensate for the loss of those genes in the male germline given that we did not see infertility in the KO males. Neither parental is in

the list of differentially expressed genes. FPKM values (Supporting information S5 Table) show very high expression of *Ran* in the wild type testis (168.78 ± 2.646) and it is not changing significantly in the *Ran-like*-KO (169.75 ± 0.773). FPKM values for *Ntf-2* show low expression of this gene in the wild type testis (4.66 ± 0.355) and stays low in the *Ntf-2r*-KO (4.57 ± 0.077). Also, from this analysis we confirm that parental genes express significantly lower than the retroduplicates in the wild type testis (*P*-values for Ntf-2 compared to Ntf-2r FPKM and Ran compared to Ran-like FPKM are 0.000416 and 0.01510164 respectively. (Supporting information S5 Table). Despite *Ran-like* been expressed much higher than *Ntf-2r* (279.72 ± 0.076 *vs.* 30.46 ± 0.378), the loss of *Ntf-2r* appears to affect expression of ten times more genes (Figure 9). There are 119 DE genes in *Ran-like*-KO testes with 87 upregulated and 32 downregulated genes (Figure 9). FBgn0036410 CG8100, FBgn0052475 mthl8, FBgn0259952 Sfp24Bb are three most upregulated genes in *Ran-like*-KO testes. FBgn0036497(Ran-like), FBgn0036497(lncRNA:CR43950) and FBgn0036879 (Cpr76Bb) are three most downregulated genes in *Ran-like*-KO testes. *Dntf-2r*-KO testes have 1003 DE genes with 449 upregulated and 559 downregulated genes (Figure 9). FBgn0264391 (long non-coding RNA:CR43839), FBgn0264436 (long non-coding RNA:CR43854) and FBgn0028872 are three most upregulated genes in *Ntf-2r*-KO testes. FBgn0261349 (*Mst36Fb*), FBgn0032680 (*Ntf-2r*) and FBgn0263333 (long non-coding RNA:CR43414) are three most downregulated genes. The first two most upregulated and the three most downregulated genes in *Ntf-2r*-KO are testes specific genes.

Functional enrichment analyses were performed for differentially expressed genes to understand which biological functions or pathways are significantly associated with the loss of *Ntf-2r* or *Ran-like*. Gene ontologies (GO) associated with DE genes including three

main branches of GO; cellular component, molecular function and biological process are shown histograms (Figure 10). Many spermatogenesis-specific genes and microtubule associated genes are going down for *Ntf-2r-KO* and many genes with translation and mitochondria functions are going up. For *Ran-like-KO*, there are more than an order of magnitude fewer DE genes associated with GOs compared to *Ntf-2r*-KO. The upregulated genes in the *Ran-like* -KO are associated with mating, insemination, copulation and sperm competition GOs (Figure 10).

Discussion

The loss of *Ntf-2r* or *Ran-like* does not affect fertility

Ntf-2r and *Ran-like* are testis-specific genes unlike their parental genes but while localization studies of Ntf-2r-EGFP and Ran-like-RFP show precise patterns of localization for these retroduplicate proteins in many cell types in testis suggesting a major role for those genes during spermatogenesis, no major sterility effects for the null mutant of these genes or double mutant were detected.

Ran-like might not be needed for normal spermatogenesis

RNA-seq results reveal minor and diverse changes in expression in *Ran-like-KO*. So, we initially hypothesize that the parental gene (*Ran*) might increase in expression in *Ran-like-KO* to compensate for the function of this retroduplicate and this would explain that we observe only these minor effects. However, we observe no changes in the level of expression of *Ran* in *Ran-like-KO*. Our RNA-seq analysis and data from single-cell RNA-seq analysis show that *Ran* is expressed very high in testes in general and comparably to

Ran-like in some meiotic spermatocytes. While *Ntf-2* has high expression in the spermatogonia cells where *Ntf-2r* expression is low, *Ntf-2* expression decreases in meiotic spermatocytes but *Ntf-2r* expression is high in this stage. High transcription level of *Ran* from the early stages of spermatogenesis through meiotic stages can explain lower number of DE genes in the absence of *Ran-like* while Ntf-2r absence affects more genes (Supporting information S5 Table and Supporting information S6 Table) (Witt, et al. 2019).

Ran-like is a more than 12 My old duplicate (Bai, et al. 2007) that has diverged a lot from *Ran* (Tracy, et al. 2010). Here, we show that Ran-like has a different function than Ran. The functional differences between Ran and Ran-like were revealed by the failure of Ran-like to rescue the lethality phenotype of the Ran null mutant. In addition, Ran-like appears to be specialized for male germline and its ectopic expression in the soma is shown to be lethal unlike the overexpression of the parental gene. Analyses of the domains conserved between Ran and Ran-like (Tracy, et al. 2010) appeared to indicate that Ran-like has retained Ntf-2 and RanGap interactions but possibly no other interactions.

Given the essential cellular function of these kinds of nuclear transport genes but the lack of strong fertility phenotype and the relatively minor gene expression changes in the *Ran-like* null mutant, we favor the hypothesize that the parental gene (*Ran*) might still have a major function during spermatogenesis as it is highly transcribed in some meiotic cells (Supporting information S6 Table) and the new gene might be differentiated and expressed to prevent deleterious interactions (meiotic drive systems interactions) with the parental gene as has been previously suggested (Tracy, et al. 2010; Larracuente and Presgraves 2012) and might not be needed for spermatogenesis unless selfish elements are present. *Ran-like* was observed to be lost in the *D. yakuba* lineage (Tracy, et al. 2010) and this would be consistent with this interpretation.

Ntf-2r might be needed for normal spermatogenesis

Ntf-2r is a younger duplicate than *Ran-like* (Bai, et al. 2007) and between 5 to 12 My old duplicate and not very differentiated from *Ntf-2* except for their expression patterns. While we do not observe major fertility effects for the null mutant (except from a mild but significant lower fertility as males age), we see major transcription upregulation for genes related to nuclear transport function in spermatogenesis including microtubule organization. We observe no change in the level of expression of *Ntf2* in *Ntf-2r-KO* and *Ntf-2* is expressed quite low in testes and mainly in mitotic cells (Supporting information S5 Table and Supporting information S6 Table) and does not change in any of the KOs. So, we conclude that *Ntf-2r* might be needed during spermatogenesis and the effect of losing it might potentially be mild to be detected cytologically and with the assays performed.

We attempted hemagglutinin (HA) tagging of the parental genes at the native site at both C-terminal and N-terminal of the coding regions using CRISPR-Cas9 technology, to investigate the protein localization of Ran, Ntf-2-PA and Ntf-2-PB in the wildtype testis and KOs. We observed that only HA tag of the Ntf-2-PB at the C-terminal was functional, and the other parental genes did not tolerate this gene modification at any of the protein ends. This observation and the fact that retroduplicates of *Ntf-2* are always duplications of the Ntf-2-RA transcript (Tracy, et al. 2010) confirm a different function for the two isoforms of *Ntf-2*, Ntf-2-PA and Ntf-2-PB. We still do not fully understand why we are able to generate EFGP and RFP tags for these genes in transgenes but not HA tags at the native site. For the Ntf-2-PB HA-tagged, we observe a localization that is broad in testes, high in elongated sperm tails, and not overlapping with the localization we observe for *Ntf-2r*.

UASt-*Ntf-2r-GFP* transgenic flies can be made for further investigation of the function

of this gene to understand if *Ntf-2r-GFP* can rescue the parental mutant (Bhattacharya and Steward 2002) but our hypothesis is that it will likely rescue.

Concluding remarks

Given the lack of strong fertility effects of knocking out *Ntf-2r* and *Ran-like,* involvement of these retroduplicates in genomic conflicts, in particular, segregation distortion system remains a possibility. One approach we are considering is to investigate if the expression of the responder satellite or piRNA production from this locus that might accompany distortion (Larracuente and Presgraves 2012) are perturbed in these mutants. This investigation can be performed to understand if *Ntf-2r* and/or *Ran-like* could have an effect in this young segregation distortion system.

Even, if involvement of *Ntf-2r* or *Ran-like* as modifiers of Rsp silencing is confirmed, it should not be assumed that this is their function or the only meiotic drive system in which those genes might have a role, as this system is too young (i.e., originated in *D. melanogaster*) (Presgraves 2007) to explain the origin of these genes. It will not be surprising if *Ntf-2r* or *Ran-like* have an effect on SD (this has been observed for extra doses of Ran, RanGAP and RCC1 (Kusano, et al. 2002) but showing that they can be players in this type of system will support the hypothesis that they are evolving fast to press segregation distortion systems and are now part of this arms race. However, more sensitive or long term selection experiments that include male-male competition in a population box could be performed including *Ntf-2r-KO*, *Ran-like-KO*, double KO and w^{1118} in different combinations and in a number of replicates to further explore potential benefits of these genes. The prediction would be that loosing *Ntf-2r-KO* might affect fertility/competitiveness of older males and it would be good to set up the cages in a way that includes those effects.

Materials and Methods

Generating *Ntf-2r* and *Ran-like* fusion proteins under native regulatory regions

Ntf-2r-EGFP constructs had been previously produced in the lab and were already available for analysis (Sorourian, et al. 2014). The complete *Ntf-2r* coding region and variable lengths of the upstream regions had been amplified from genomic DNA and cloned into the plasmid pEGFP1 (U55761; Clontech, Mountain View, CA) to put the *Ntf-2r* in frame with the EGFP gene and generate a fluorescent fusion protein. These regions containing different lengths of the putative regulatory region, 5'UTR of *Ntf-2r*, *Ntf-2r-EGFP* fused coding regions and the SV40 polyadenylation site were then further cloned into the P-element *Drosophila* transformation vector – pCaSpeR 4 (X81645) and used for fly transformation.

Similar to *Ntf-2r-EGFP* fusion and regulatory regions cloned, the complete *Ran-like* coding region and variable lengths of the upstream region were amplified from genomic DNA and cloned into a destination plasmid. pRed H-Pelican plasmid (provided by Dr. Barolo Lab, La Jolla, CA) was used to put *Ran-like* in frame with the Ds.RedT4 gene (red fluorescent protein; DsRed.T4) and generate a red fluorescent fusion protein. These clones contain different lengths of the putative regulatory region, 5'UTR of Ran-like, and Ran-like-DsRed.T4 fused coding regions. Before *Ran-like region* was inserted the TATA box was removed from the original pRed H-Pelican construct using AgeI and XhoI (Promega, Madison, WI) restriction enzymes. The three different inserts containing *Ran-like* and the diverse upstream regions were also digested using the same restriction enzymes and cloned in frame with DsRed.T4 into the P element *Drosophila* transformation vector (pCaSpeR 4; X81645) and used for fly transformation. To clone the fragments into the plasmid. P-element plasmids were sent for injection to Genetic Services, Inc. (Cambridge,

MA). The white mutant stock w^{1118} was used for injection. A helper plasmid containing the P-element transposase gene was also injected along with the desired plasmid in order to provide the transposase function to excise the P-element region from the plasmid. The flies carrying the longest upstream region for *Ntf-2r-EGFP* and *Ran-like-DsRed.T4* were used here in the protein expression and localization analyses.

Generating the *Ntf-2r*-KO and *Ran-like*-KO lines

Null mutants of *Ntf-2r* and *Ran-like* were generated using the CRISPR-Cas9 technology. Two guide RNAs (gRNAs) for each gene were designed using the online platform https://flycrispr.org/target-finder (Gratz, et al. 2014) and synthesized by IDT, Inc. as 5' unphosphorylated oligonucleotides. Oligonucleotides were annealed, phosphorylated, and ligated into the BbsI sites of the pU6-BbsI-chiRNA plasmid (Addgene plasmid # 45946) separately producing two plasmids to express the gRNAs in germline upon embryo injection (Gratz, et al. 2013). In addition, two homologous arms for each gene were designed and synthesized by IDT, Inc. The pHD-DsRed-attP plasmid (Addgene # 80898) was cut using EcoRI and XhoI restriction enzymes (Promega Corporation) and the homologous arms were cloned into the cut plasmid (donor vector)(Gratz, et al. 2014) flanking the eye driven DsRed Cassette designed to replace *Ntf-2r* or *Ran-like* (Addgene # 80898; (Gratz, et al. 2014)). NEBuilder® HiFi DNA Assembly Master Mix (NEB, Inc.) was used to assemble the homologous arms flanking the DsRed cassette following the protocol provided in the kit. gRNA clones and the donor vector were co-injected into nos-Cas9 attp2 and nos-Cas9 attP40 strains embryos (Kondo and Ueda 2013) at the following concentrations: 250 ng/ml PhD-DsRed-attP donor vector and 20 ng/ml of each of the pU6-BbsI-chiRNA plasmids containing the guide RNAs by Rainbow Transgenic Flies, Inc.

(Camarillo, CA). The gRNAs and homologous arm sequences are provided in Supporting information S1 Table 2 and Table 3. Injected flies were collected from injected embryos and crossed with w^{1118} flies. The progeny of these crosses was screened for fluorescent glowing eyes meant to confirm the replacement of the desired gene by the eye-driven DsRed gene. The mutant allele was fixed using balancer chromosomes. Absence of the genes was tested in the homozygote mutant individuals by PCR and sequencing. In addition, absence of any transcript from the knocked-out genes were confirmed by RNA-sequencing. To control for the background effects, flies from a KO line were backcrossed with the w^{1118} for six generations as outlined in previous publications (Slawson, et al. 2011; Chandler, et al. 2013). Two homozygote backcrossed replicates were produced for each of *Ntf-2r*-KO and *Ran-like*-KO strains. Double-mutant strain of *Ntf-2r*-KO and *Ran-like*-KO was also produced using the backcrossed confirmed single knock out strains. All the downstream fertility analyses were performed using confirmed backcrossed strains.

Fertility assays

We measured male fertility using two separate assays. In the first assay following Dyer et al., a single male and single female (2 days old) were placed in a vial, transferred to fresh food after 4 days, and then removed after an additional 4 days (Dyer, et al. 2010). All offspring were counted after 20 days from the first cross. For each cross 4 replicates were used and crosses were kept at 25°C. This assay was done for *Ntf-2r*-KO, *Ran-like*-KO and double knock-out strain of these genes and were compared to the w^{1118}.

We also used a male sperm exhaustion assay. Individual 2-day-old virgin male was set up to mate with two 3- to 4-day-old virgin females for 24 hours (Sun, et al. 2004). The peak of egg laying in the female flies is between 4–15 days. The male was transferred into

a fresh vial on each of the following 10 days to mate with two fresh virgin females. All the inseminated females were allowed to oviposit and were transferred to fresh vials every 2 days for 8 days for a total of 4 transfers. The offspring were collected and counted till the 15th day after the initial mating. For each cross 5 replicates were used. All crosses were kept at 25°C.

Generation of parental genes hemagglutinin tags

To study the protein localization of the parental genes and compare it to the duplicates, we attempted the generation of parental genes hemagglutinin (HA) tags at endogenous sites using CRISPR-Cas9 technology. We made constructs to tag Ntf-2-PA, Ntf-2-PB, and Ran at their C-terminus and also tried to tag Ntf-2 and Ran at their N-terminus following the same approach described above for the CRISPR knock-outs with the modification of using a single gRNA. In addition, we did not use DsRed reporter in the donor plasmids for tagging these genes and screened the replacement of the genes with the genes tagged by PCR. We were only successful at tagging Ntf-2-PB at its C-terminus. The gRNA and homologous arm sequence used are provided in Supporting information (S1 Tables 4-8).

UASt transformants for Ran mutant rescue

Using the gateway system *Ran* and *Ran-like* constructs were produced containing the UASt upstream of the coding region of the gene and an in-frame EFGP downstream of the gene. The rescue crosses were carried out according to previous work. Each UASt fusion gene was expressed under the control of the Act5C-GAL4 (stock number 4414) and checked for their ability to rescue *Ran* knockout phenotype (Cesario and McKim 2011).

Immunofluorescent staining of Ntf-2-RB-HA testes

The testes of one-day-old virgin males were dissected in 1X PBS within 20 minutes, fixed on slides and stained following the formaldehyde fixation protocol described in (White-Cooper 2004). In brief, the testes were fixed in 4% formaldehyde in PBS plus 0.1% Triton X-100 (PBS-T) for 7 minutes and then washed two times in PBS for 5 minutes at room temperature. Slides containing testes were immersed in PBS-T for 30 minutes at room temperature to permeabilize cell membranes and washed two times in PBS. Forty µl of Primary antibody (Anti-HA; Cat # C29F4, Cell Signaling Technology, Inc.) was added to the fixed tissues in a concentration of 1:100 in PBS-T and stored at room temperature for two hours or at 4°C overnight. After testes were thoroughly washed three times in PBS, 40 µl of the secondary antibody (Anti-Rabbit; Cat # A11008, Invitrogen, Thermo Fisher Scientific) was added at a concentration of 1:200 in PBS and stored at room temperature for two hours. The testes were washed in 1X PBS three times. MitoTracker™ Deep Red FM in the concentration of 500 nM (Cat # M22426, Invitrogen, Thermo Fisher Scientific) was used to stain mitochondria (Gilmore and Wilson 1999). The slides were washed two times in 1% PBS and then NucBlue™ Fixed Cell ReadyProbes™ Reagent (DAPI); Catalog # R37606, (Invitrogen, Thermo Fisher Scientific) was used to stain the nuclear DNA. The confocal microscope at UT Arlington (Nikon Eclipse Ti2 laser scanning confocal microscope) was used for imaging. NIS-Elements imaging software (Version 5.20.00) was used for image processing.

Immunostaining of known spermatogenesis proteins

To study the co-localization of Ran-like-DsRed.T4 with other proteins in different stages of spermatogenesis, antibody staining of genes with known localization during

spermatogenesis was performed. Anti-lamin and anti-alpha-tubulin (tub) were ordered from the Developmental Studies Hybridoma Bank, University of Iowa. Lamin antibody stains the internal part of the nucleus in the mitotic cells before the start of meiosis. Alpha tubulin antibody labels microtubules during cell division and can be used to check for the presence or absence of Ran-like in the waste bags. Testes were dissected from pupae and 0-days-old males from Ran-like-DsRed.T4 transformed strains and from w^{1118}. Some testes were squashed using the cover slip and froze for 1 minutes at -80°C. The cover slip was then removed, and the tissues were dipped in methanol for 10 minutes, followed by 30 seconds in acetone and then 5 minutes in PTW. In the last step the samples were washed in PBS 2 times for 10 minutes and allowed to prehybridize for 2 hours in 1% PBS with BSA. Different antibodies were used one at a time in a concentration of 1:50 overnight in a wet chamber. The next day, samples were again washed in 1% PBS and secondary antibody was applied. For secondary antibody Alexa Fluor 488 Goat Anti-Rabbit and Alexa Fluor 488 Goat Anti-mouse was used (Molecular Probes, Life Technologies, Carlsbad, CA). Before microscopy, samples were washed one more time for 30 minutes. Whole testes were also dissected and fixed with PFA and washed 2 times with 1% PBS for 10 minutes followed by staining with primary antibody at 1:100 overnight at 4°C. After overnight hybridization the samples were washed with 1% PBS and followed by staining with secondary antibody at 1:500. Finally, the testes were washed and mounted in FluorGlo mounting solution (Valley Scientific, Mayville, NY) for examination by fluorescence microscopy (Olympus BX51TRF florescent microscope) and/or confocal microscopy (Zeiss LSM 150). To visualize F-actin cones in testes, phalloidin staining was used. Testes were dissected from 0-days-old males from Ran-like-DsRed.T4 transformed strains and w^{1118} and fixed following the protocol describes above. Tissues were then washed with 1% PBS for 10 minutes and 0.5µl of

10μM phalloidin (Biotium Inc., Hayward, CA) was added to 750μl of 1% PBS and left 1 hour in the dark at room temperature. After one hour the testes were washed twice with 1% PBS and mounted in FluorGlo mounting solution (Valley Scientific, Mayville, NY) for examination by fluorescence microscopy (Olympus BX51TRF fluorescence microscope) and/or confocal microscopy (Zeiss LSM).

RNA sequencing

Forty pairs of testes were dissected from 0-1 day old *Ntf-2r*-KO, *Ran-like*-KO and control (w^{1118}) male flies. There were two replicates produced from independent backcrosses of the KO line into w^{1118} background for each gene and two replicates for the control. RNA was extracted using Direct-zol (Zymo Research) kit and stored at -70°C. RNA was sent for quality control, library preparation and mRNA sequencing at Novogene Co., Ltd. Part of the bioinformatic analysis were also performed by Novogene Co., Ltd. After RNA quantification and qualification, sequencing libraries were generated and mRNA was purified from total RNA using poly-T oligo-attached magnetic beads and were sequenced on Illumina NovoSeq 6000 and paired-end reads were generated. Raw data from sequencing (FASTQ format) were checked for quality by processing through fastp. All the downstream analyses were performed using clean data with high quality. Paired-end clean reads were mapped to the *Drosophila* reference genome (genome assembly BDGP6.88) using HISAT2 software. Differential expression analysis was performed using DESeq2 R package (Anders and Huber 2010). Genes with resulting adjusted P value < 0.05 found by DESeq2 were assigned as differentially expressed. ClusterProfiler R package was used for GO enrichment analysis (Kanehisa and Goto 2000; Yu, et al. 2012).

Acknowledgments

The authors would like to thank Victor Palacios and Michael Buszczak for their help with the construct design of the HA tags of the parental genes and their insightful suggestions and discussions. Authors would like to also thank Yan Chang for her assistance with confocal microscopy.

Authors' contributions

AM, SD, DNM, DJ, and EB contributed to the idea and study design. AM and DJ generated the null mutants. AM and DNM generated the HA tags. AM conducted the fertility assays and rescue crosses. AM conducted the immunostaining and confocal microscopy of the HA tag. SD conducted localization and colocalization studies of Ntf-2r and Ran-like. AM and DNM prepared samples for RNA-sequencing. AM and DJ contributed to the bioinformatic analysis of the RNA-seq data. AM and SD wrote the manuscript. EB was a major contributor to data processing and analyses and the writing of the manuscript. All authors read and approved the final manuscript.

Funding

The Nikon Nikon Eclipse Ti2 laser scanning confocal microscope was acquired with the NIH S10 OD025230 award. EB would like to acknowledge the support from the National Institutes of Health (R01GM071813). AM would like to acknowledge Phi-Sigma Research Grant, UTA COS Maverick Science Graduate Research Fellowship and UTA Dissertation Fellowship. The content of this work is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

References

Anders S, Huber W. 2010. Differential expression analysis for sequence count data. Genome Biology 11:R106.

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

Betrán E, Long M. 2003. Dntf-2r, a Young Drosophila Retroposed Gene With Specific Male Expression Under Positive Darwinian Selection. Genetics 164:977.

Bhattacharya A, Steward R. 2002. The Drosophila homolog of NTF-2, the nuclear transport factor-2, is essential for immune response. EMBO Rep 3:378-383.

Cesario J, McKim KS. 2011. RanGTP is required for meiotic spindle organization and the initiation of embryonic development in Drosophila. J Cell Sci 124:3797-3810.

Chandler CH, Chari S, Dworkin I. 2013. Does your gene need a background check? How genetic background impacts the analysis of mutations, genes, and evolution. Trends Genet 29:358-366.

Clarke PR, Zhang C. 2008. Spatial and temporal coordination of mitosis by Ran GTPase. Nat Rev Mol Cell Biol 9:464-477.

Dorogova NV, Akhmametyeva EM, Kopyl SA, Gubanova NV, Yudina OS, Omelyanchuk LV, Chang LS. 2008. The role of Drosophila Merlin in spermatogenesis. BMC Cell Biol 9:1.

Dyer KA, White BE, Bray MJ, Piqué DG, Betancourt AJ. 2010. Molecular Evolution of a Y Chromosome to Autosome Gene Duplication in Drosophila. Molecular Biology and Evolution 28:1293-1306.

Fabian L, Brill JA. 2012. Drosophila spermiogenesis: Big things come from little packages. Spermatogenesis 2:197-212.

Flores HA, Bubnell JE, Aquadro CF, Barbash DA. 2015. The Drosophila bag of marbles Gene Interacts Genetically with Wolbachia and Shows Female-Specific Effects of Divergence. PLOS Genetics 11:e1005453.

Ghosh-Roy A, Desai BS, Ray K. 2005. Dynein light chain 1 regulates dynamin-mediated F-actin assembly during sperm individualization in Drosophila. Mol Biol Cell 16:3107-3116.

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.

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.

Isgro TA, Schulten K. 2007. Association of Nuclear Pore FG-repeat Domains to NTF2 Import and Export Complexes. Journal of Molecular Biology 366:330-345.

Kaláb P, Pralle A, Isacoff EY, Heald R, Weis K. 2006. Analysis of a RanGTP-regulated gradient in mitotic somatic cells. Nature 440:697-701.

Kanehisa M, Goto S. 2000. KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res 28:27-30.

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

Kusano A, Staber C, Ganetzky B. 2002. Segregation distortion induced by wild-type RanGAP in Drosophila. Proceedings of the National Academy of Sciences 99:6866.

Larracuente AM, Presgraves DC. 2012. The Selfish Segregation Distorter Gene Complex of Drosophila melanogaster. Genetics 192:33.

Moore MS. 1998. Ran and nuclear transport. J Biol Chem 273:22857-22860.

Noguchi T, Koizumi M, Hayashi S. 2012. Mitochondria-driven cell elongation mechanism for competing sperms. Fly (Austin) 6:113-116.

Peter A, Schöttler P, Werner M, Beinert N, Dowe G, Burkert P, Mourkioti F, Dentzer L, He Y, Deak P, et al. 2002. Mapping and identification of essential gene functions on the X chromosome of Drosophila. EMBO Rep 3:34-38.

Phadnis N, Hsieh E, Malik HS. 2012. Birth, death, and replacement of karyopherins in Drosophila. Mol Biol Evol 29:1429-1440.

Presgraves DC. 2007. Does genetic conflict drive rapid molecular evolution of nuclear transport genes in Drosophila? Bioessays 29:386-391.

Presgraves DC, Gérard PR, Cherukuri A, Lyttle TW. 2009. Large-Scale Selective Sweep among Segregation Distorter Chromosomes in African Populations of Drosophila melanogaster. PLOS Genetics 5:e1000463.

Presgraves DC, Stephan W. 2007. Pervasive adaptive evolution among interactors of the Drosophila hybrid inviability gene, Nup96. Mol Biol Evol 24:306-314.

Quimby BB, Lamitina T, L'Hernault SW, Corbett AH. 2000. The mechanism of ran import into the nucleus by nuclear transport factor 2. J Biol Chem 275:28575-28582.

Ribbeck K, Lipowsky G, Kent HM, Stewart M, Görlich D. 1998. NTF2 mediates nuclear import of Ran. The EMBO Journal 17:6587-6598.

Shevelyov YY, Lavrov SA, Mikhaylova LM, Nurminsky ID, Kulathinal RJ, Egorova KS, Rozovsky YM, Nurminsky DI. 2009. The B-type lamin is required for somatic repression of testis-specific gene clusters. Proceedings of the National Academy of Sciences 106:3282.

Silverman-Gavrila RV, Wilde A. 2006. Ran is required before metaphase for spindle assembly and chromosome alignment and after metaphase for chromosome segregation and spindle midbody organization. Mol Biol Cell 17:2069-2080.

Slawson JB, Kuklin EA, Ejima A, Mukherjee K, Ostrovsky L, Griffith LC. 2011. Central regulation of locomotor behavior of Drosophila melanogaster depends on a CASK isoform containing CaMK-like and L27 domains. Genetics 187:171-184.

Sorourian M, Kunte MM, Domingues S, Gallach M, Özdil F, Río J, Betrán E. 2014. Relocation facilitates the acquisition of short cis-regulatory regions that drive the expression of retrogenes during spermatogenesis in Drosophila. Mol Biol Evol 31:2170-2180.

Sun S, Ting CT, Wu CI. 2004. The normal function of a speciation gene, Odysseus, and its hybrid sterility effect. Science 305:81-83.

Tracy C, Río J, Motiwale M, Christensen SM, Betrán E. 2010. Convergently Recruited Nuclear Transport Retrogenes Are Male Biased in Expression and Evolving Under Positive Selection in Drosophila. Genetics 184:1067.

Trieselmann N, Wilde A. 2002. Ran localizes around the microtubule spindle in vivo during mitosis in Drosophila embryos. Curr Biol 12:1124-1129.

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

Witt E, Benjamin S, Svetec N, Zhao L. 2019. Testis single-cell RNA-seq reveals the dynamics of de novo gene transcription and germline mutational bias in Drosophila. Elife 8:e47138.

Yu G, Wang LG, Han Y, He QY. 2012. clusterProfiler: an R package for comparing biological themes among gene clusters. Omics 16:284-287.

Zhong L, Belote JM. 2007. The testis-specific proteasome subunit Prosalpha6T of D. melanogaster is required for individualization and nuclear maturation during spermatogenesis. Development 134:3517-3525.

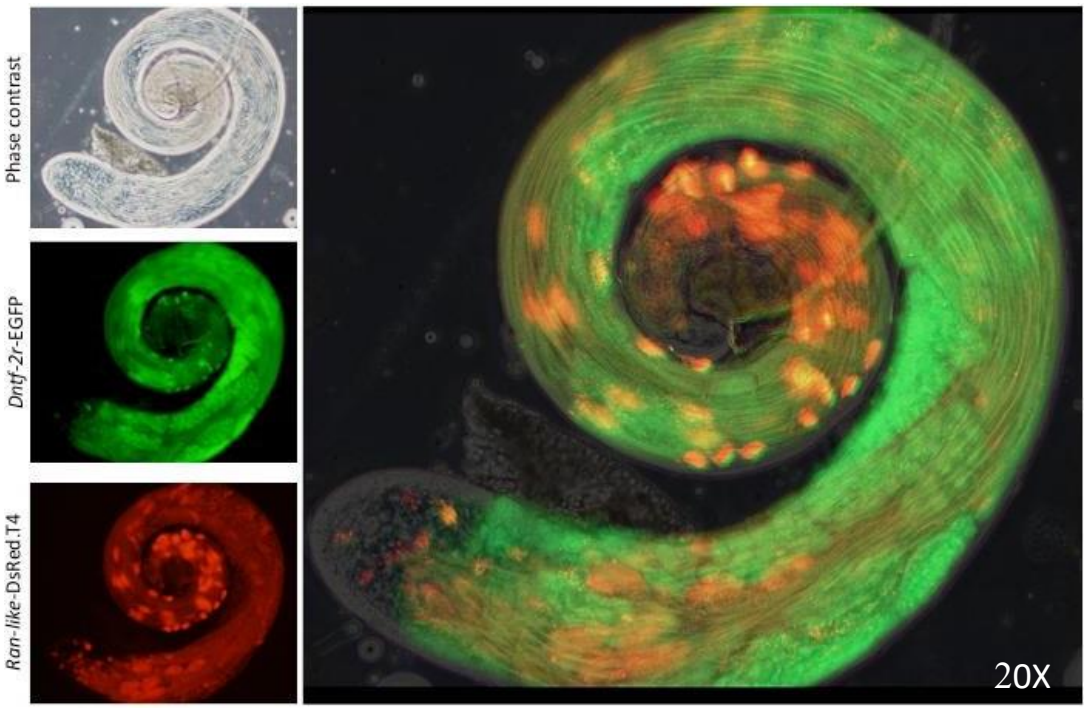


Figure 1. Whole testis showing Dntf-2r-EGFP and Ran-like-RFP localization in the different cell stages. Dntf-2r in green and Ran-like in red.

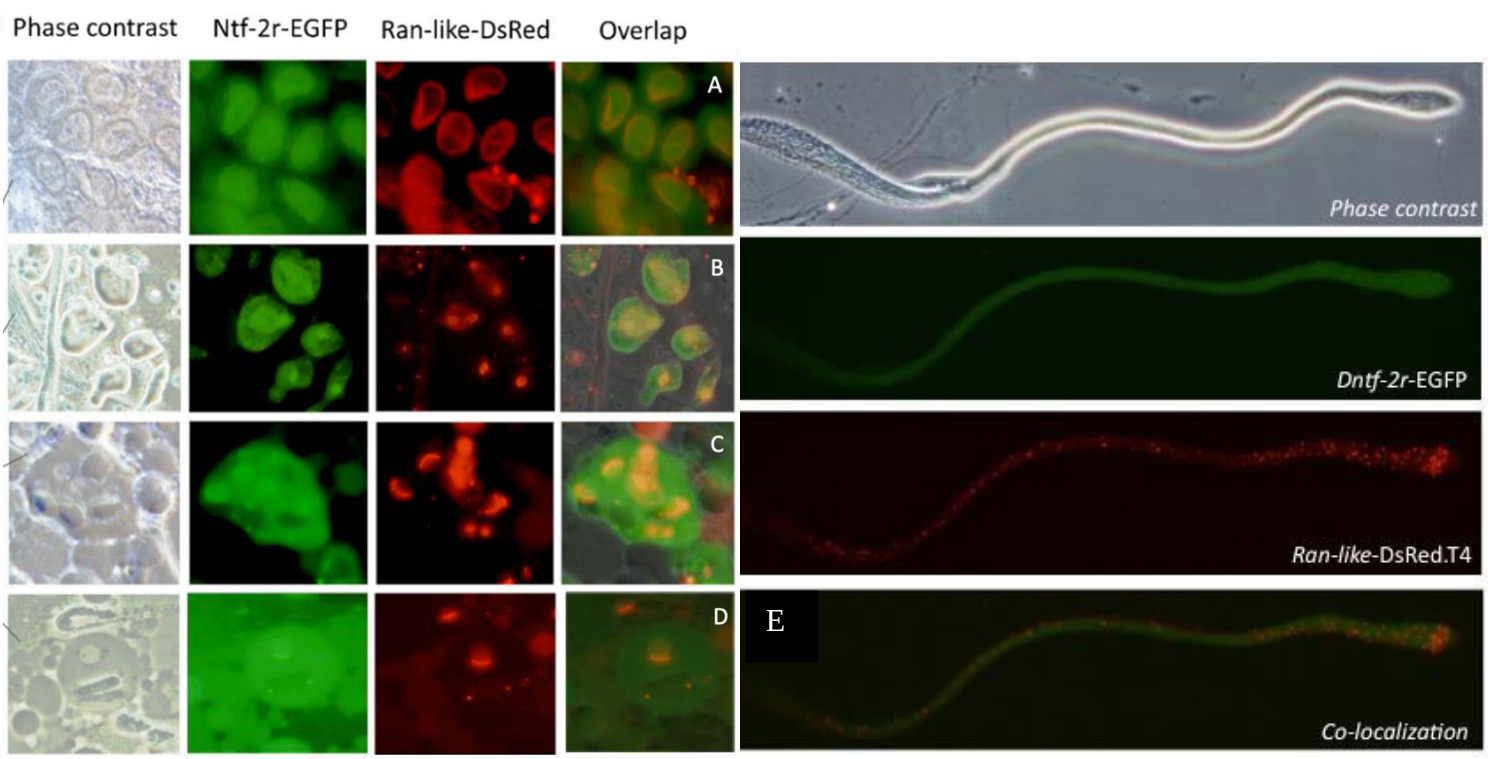


Figure 2. Dntf-2r-EGFP and Ran-like-RFP localization in different cell stages of spermatogenesis. In undividing meiotic cells, Ntf-2r localizes in the cytoplasm, nuclear membrane and inside the nucleus. Ran-like co-localizes with Ntf-2r in the nuclear membrane and nucleus but no other cellular compartments (A). In dividing spermatocytes, Ran-like and Ntf-2 localize with the cell spindles and at the cell poles. Ntf-2r localizes in the cytoplasm as well (B). In round spermatids (onion stage), both retrogenes localize in dense body surrounding the nucleus and at the beginning and at the end of the mitochondria with microtubules organizing for mitochondria elongation, Ntf-2r also localizes with the cell nucleus (C). In the 64-cell stage when the mitochondria and the nucleus are elongating, both Ran-like and Ntf-2r show co-localization in the dense body and at the mitochondrial poles at the ends of two elongating fused mitochondria (D). After sperm elongation, Ntf-2r is present all along the sperm bundles, however, Ran-like seems to have a very precise localization along the sperm bundle tails (E).

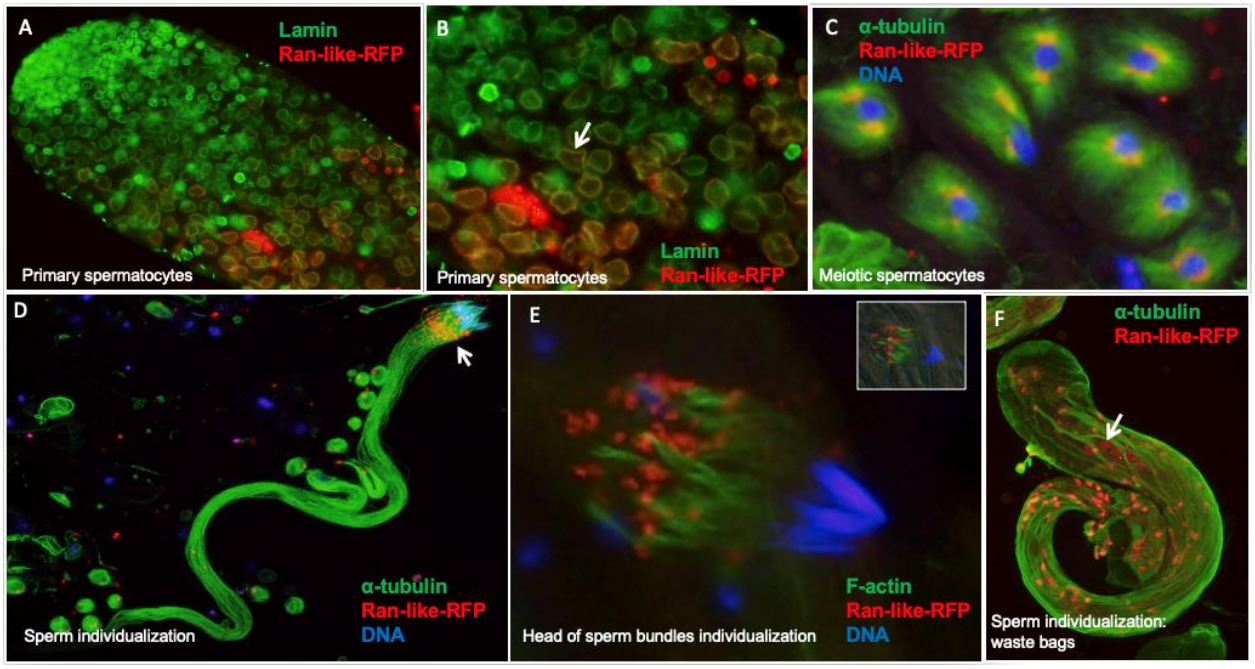


Figure 3. Ran-like-RFP co-localization with known spermatogenesis proteins. Overlap in expression between Ran-like and Lamin antibody in the internal part of the nucleus in primary spermatocytes (A and B). Ran-like co-localizes with microtubule nucleation sites and microtubules, labeled with alpha-tubulin antibody (C). During sperm individualization, Ran-like continues to localize at nucleation sites of dynamic microtubules (white arrow) (D). During individualization of sperm bundle heads, Ran-like localizes distally in respect to F-actin cones (E). Ran-like accumulate in the waste bags at the end of sperm individualization (F).

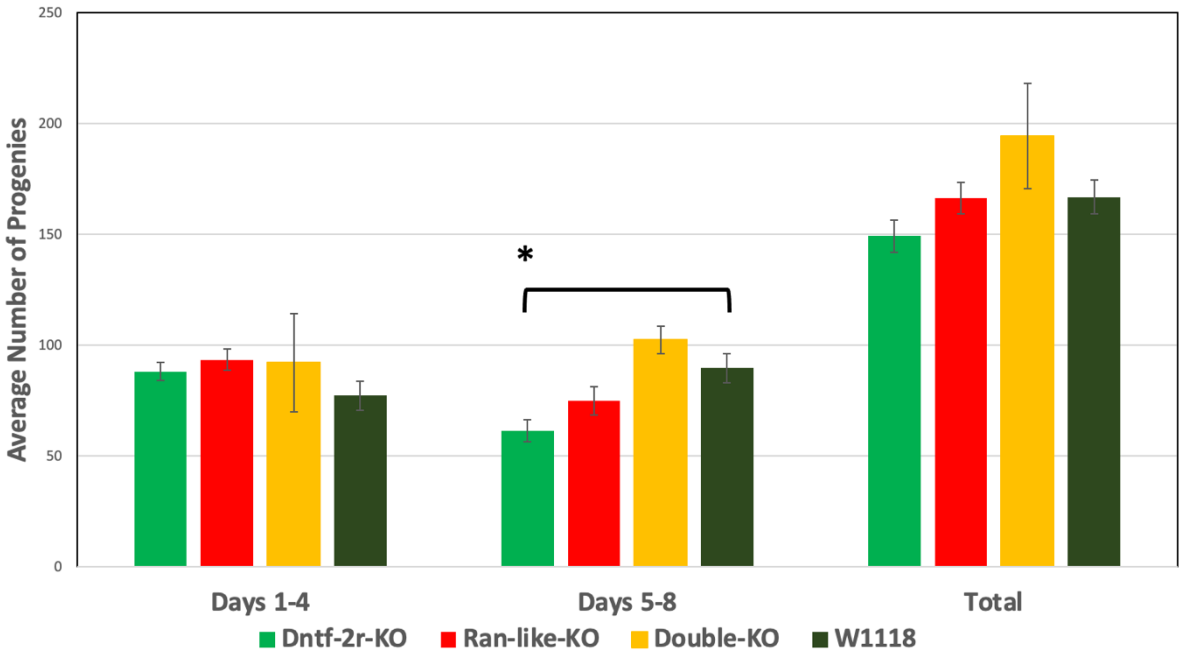


Figure 4. Fertility assay of backcrossed *Ntf-2r*-KO, *Ran-like*-KO and double knockout compared to the control strain (w^{1118}). Error bars represent the SE of the mean (n=5). No fertility effect is observed for the knockouts in the first four days. In the second four days *Ntf-2r*-KO males were having significantly lower number of progeny (shown with *) compared to the line of control (P<0.05).

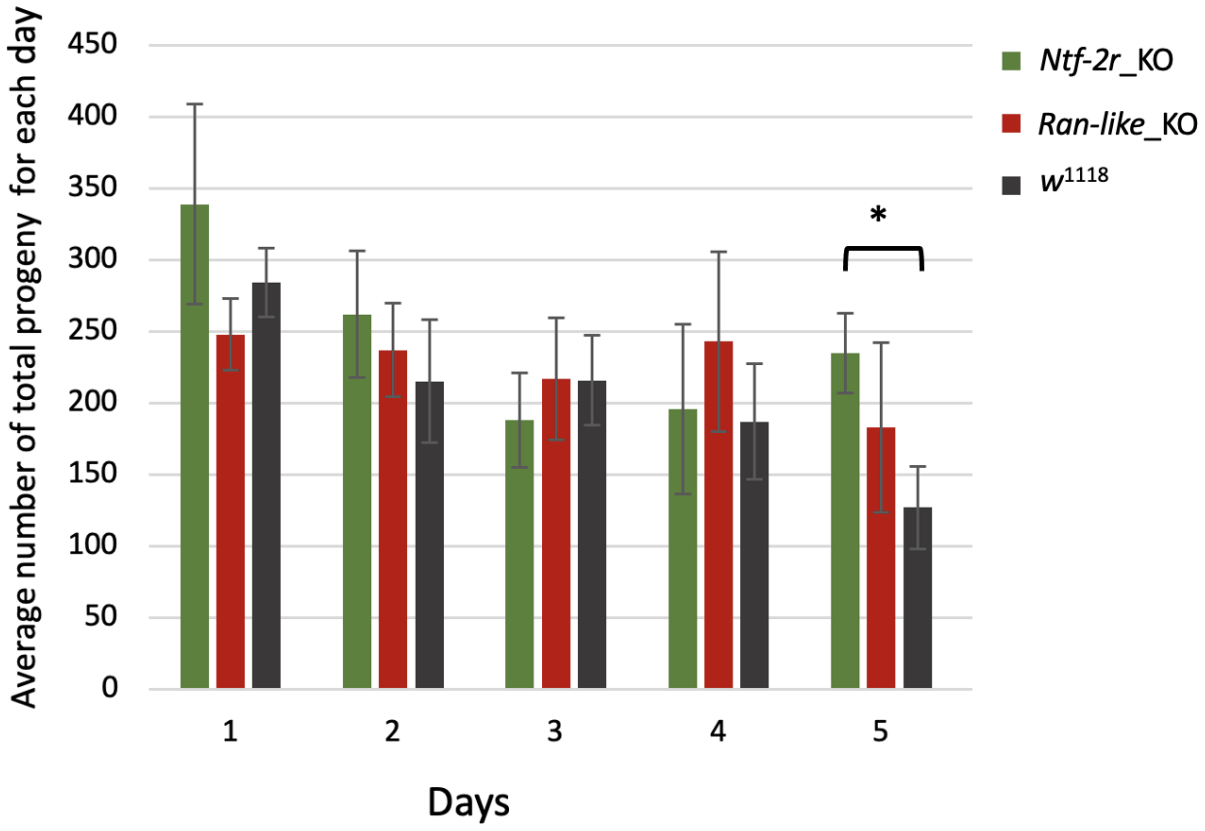


Figure 5. Sperm exhaustion assay of backcrossed *Ntf-2r* and *Ran-like* knockouts. Average number of progeny after 15 days from each cross or transfer is shown. Error bars represent the SE of the mean (n=5). No significant fertility reduction was observed for the knockout strains compared to the control strain in this assay. Only in day five, *Ntf-2r*-KO males were having significantly higher number of progeny (shown with *) compared to the line of control (P<0.05).

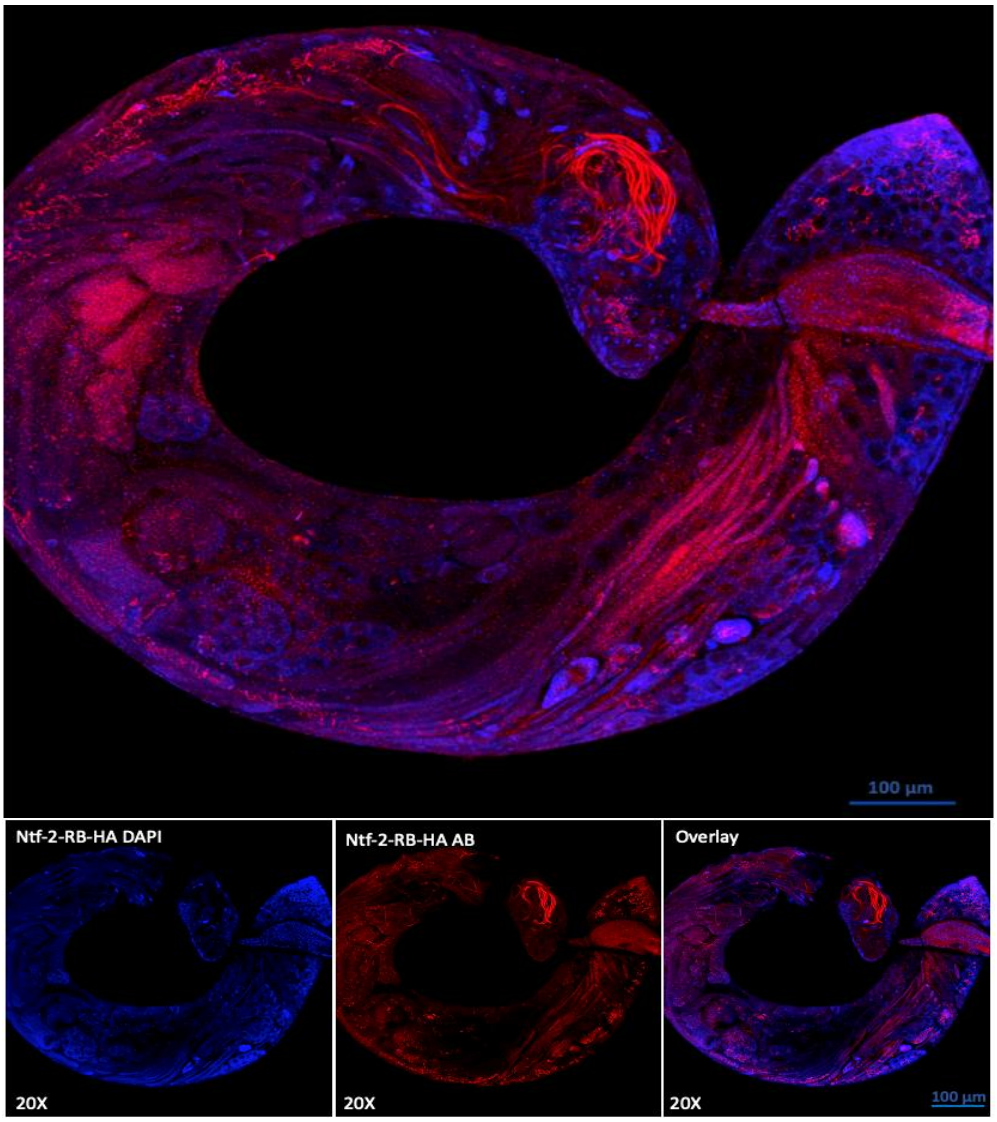


Figure 6. Ntf-2-PB-HA immunostaining in whole testis. DNA is stained with DAPI in blue and Ntf-2-PB-HA is stained with Anti-Rabbit secondary antibody observed in red. Ntf-2-PB-HA protein is observed starting at the tip of the testis. Accumulation of Ntf-2-PB is observed at the sperm bundles at precise localization.

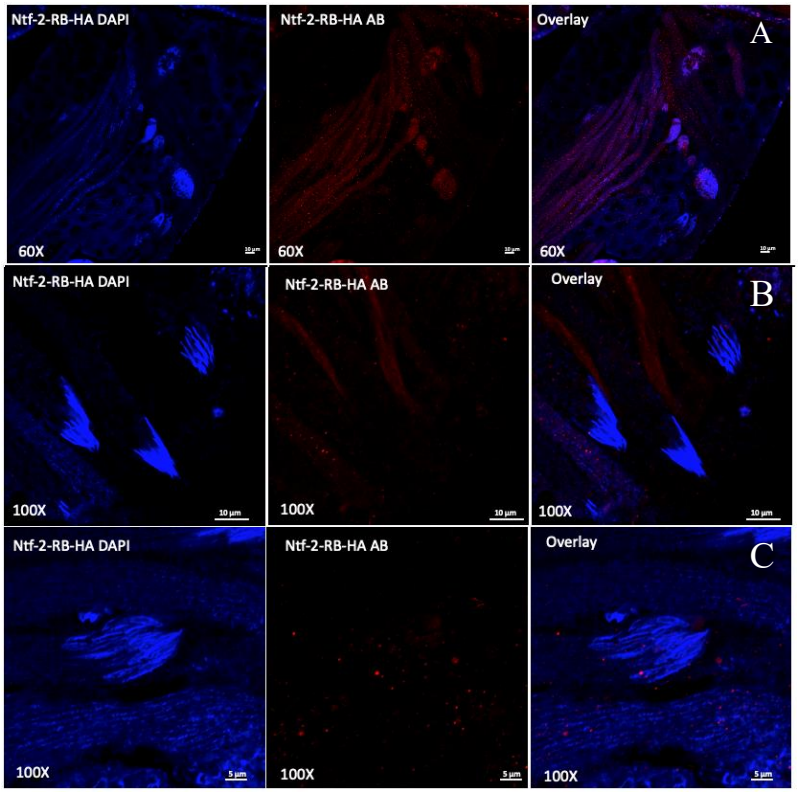


Figure 7. Ntf-2-PB-HA immunostaining in in different stages of spermatogenesis. Distribution of Ntf-2-PB-HA localization does not resemble that of Ntf-2r. Ntf-2-PB-HA is not diffused in the cytoplasm, at the nuclear membrane and at structures where microtubules are organized (A). Ntf-2-PB-HA appears to be accumulated at the sperm tails of sperm bundles mimicking Ran-like localization but with higher density and overlapping DNA in some cell types (A, B and C).

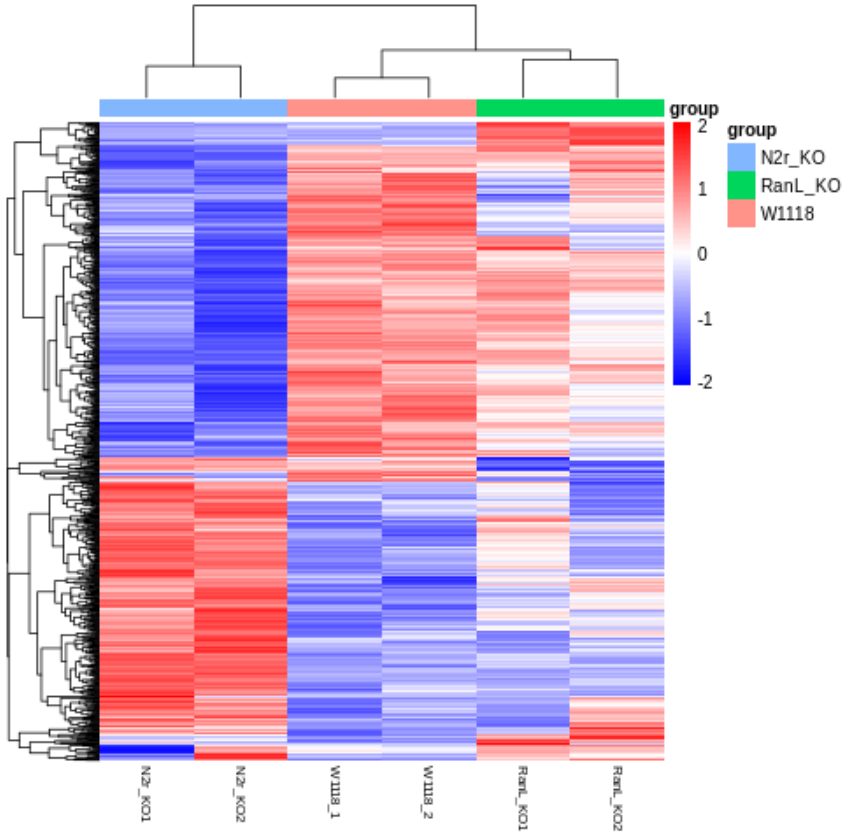


Figure 8. Hierarchical Clustering Heatmap. The overall results of FPKM cluster analysis, clustered using the log2(FPKM+1) value. Red color indicates genes with high expression levels, and blue color indicates genes with low expression levels. Samples cluster according to their phenotype.

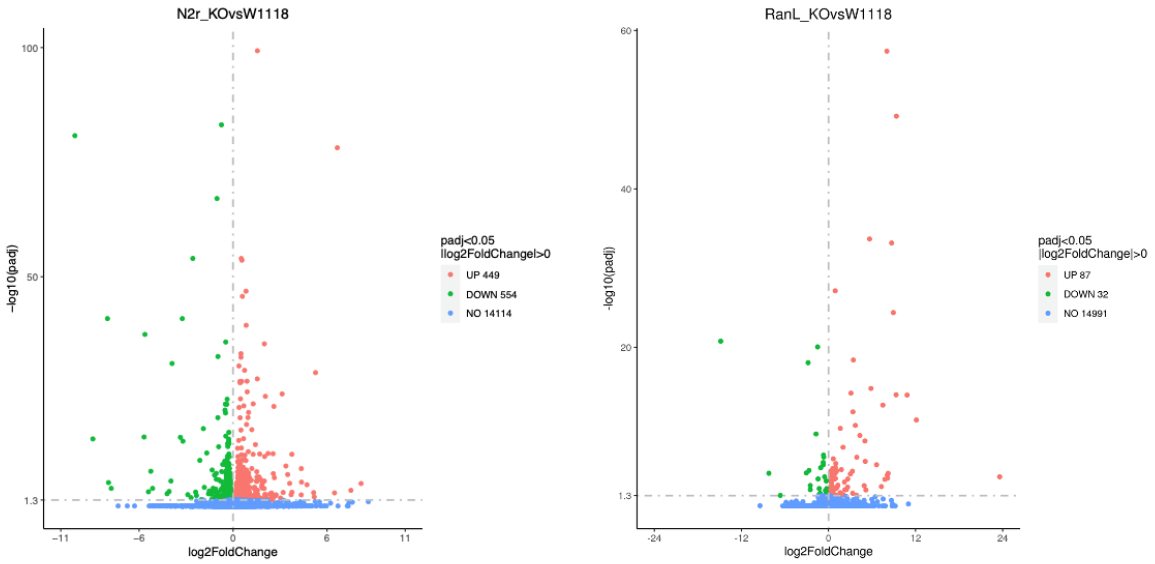


Figure 9. Volcano plot. Horizontal axis for the fold change of genes in Sample Ntf-2r-KO (A) and Ran-like-KO (B). Vertical axis for statistically significant degree of changes in gene expression levels, the smaller the corrected *P*-value, the bigger -log10(corrected *P*-value), the more significant the difference. The points represent genes, blue dots indicate no significant difference in genes, red dots indicate upregulated differential expression genes, green dots indicate downregulated differential expression genes.

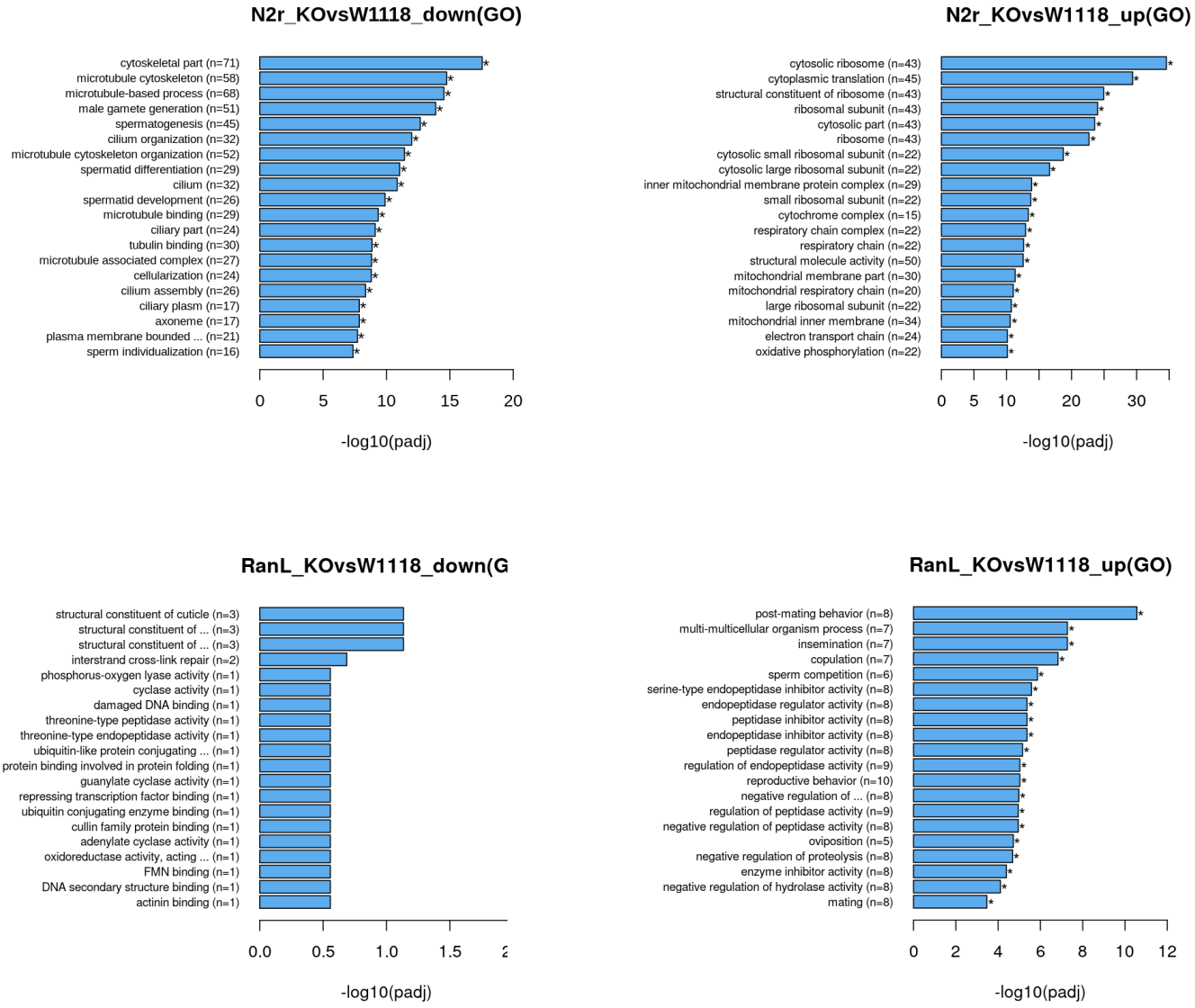


Figure 10. GO enrichment histogram. Significantly enriched terms in the GO enrichment analysis are displayed. GO terms with *P*adj < 0.05 are significant enrichment and marked with a star.

Chapter 4

Genes

COX4-like*, a nuclear-encoded mitochondrial gene duplicate, is essential for male fertility in *Drosophila melanogaster

Mohammadmehdi Eslamieh[1#], Ayda Mirsalehi[1#], Dragomira Markova and Esther Betrán[1*]

[1] Department of Biology, the University of Texas at Arlington, Arlington, TX, USA.

# These authors made equal contribution to this work.

* Corresponding author

Esther Betrán

Biology Department, Box 19498

University of Texas at Arlington, Arlington, TX 76019, USA

Phone (817) 272 1446

E-mail: betran@uta.edu

Keywords: nuclear-encoded mitochondrial gene, gene duplication, CRISPR knock-out, *COX4L*, spermatogenesis, *Drosophila melanogaste*r

Running title: *COX4-like* is essential for male fertility in *Drosophila*

Abstract

Recent studies of nuclear-encoded mitochondrial genes (N-mt genes) in *Drosophila melanogaster* have shown a unique pattern of expression for newly duplicated N-mt genes with many duplicates having a testis-biased expression and an essential role in spermatogenesis. In this study, we investigated a newly duplicated N-mt gene, *Cytochrome c oxidase 4-like* (*COX4L*), in order to understand its function and consequently the reason behind its retention in the *Drosophila melanogaster* genome. *COX4L* is a duplicate of *Cytochrome c oxidase 4* (*COX4*) of OXPHOS complex IV. While the parental *COX4* gene has been found in all eukaryotes including single-cell eukaryotes like yeast, we show that *COX4L* is only present in the *Brachycera* suborder of Diptera. Thus, both genes are present in all *Drosophila* species but have a significantly different pattern of expression: *COX4* is highly expressed in all tissues, while *COX4L* has a testis-specific expression. To understand the function of this new gene, we first knocked down its expression in the *D. melanogaster* germline using two different RNAi lines driven by the *bam-Gal4* driver; and second, we created a knockout strain for this gene using CRISPR-Cas9 technology. Our results showed that knockdown and knockout lines of *COX4L* produce partial sterility and complete sterility in males, respectively, where a lack of sperm individualization was observed in both cases. Male infertility was rescued by driving *COX4L*-HA in the germline but not when driving *COX4*-HA. In addition, ectopic expression of *COX4L* in soma caused embryonic lethality and overexpression in the germline led to a reduction in male fertility. *COX4L-KO* mitochondria show reduced membrane potential and seem to be not functional, providing a plausible explanation for the male sterility observed in these flies. This prominent loss-of-function phenotype along with its testis-biased expression, and its presence in the *Drosophila* sperm

proteome suggests that *COX4L* is a paralogous, specialized gene that is assembled in the OXPHOS complex IV of male germline cells and/or sperm mitochondria.

Introduction

Mitochondria not only produce a big fraction of the cellular energy but are also involved in a diverse set of cellular functions, such as metabolism (Wai and Langer 2016), immune regulation (Weinberg, et al. 2015), and cell death (Wallace 2005). Alphaproteobacteria have long been considered as the mitochondria ancestor (Gray 2012), however, a recent phylogenetic study has challenged this view by showing that mitochondria might have evolved even earlier from a proteobacterial lineage before the divergence of alphaproteobacteria (Martijn, et al. 2018). Mitochondria are organelles in the eukaryotic cells with their own DNA (mtDNA). During its approximately 1.5 billion years of evolution, the mitochondrial genome has experienced many changes. Some of the ancestral proteobacterial genes have been lost while others that encode mitochondrial proteins have been transferred to the nucleus (i.e., nuclear-encoded mitochondrial genes or N-mt genes). Consequently, few genes have remained in the mitochondrial genome, which in most metazoans consists of 13 protein-coding genes, two rRNA genes, and 22 tRNA genes (Gray 2012). The *Drosophila melanogaster* (*D. melanogaster*) mitochondrial genome is a circular molecule estimated to be only ~19.5 Kb in size (Lewis, et al. 1995). As a consequence of this severe reduction in gene content, mitochondria import most of their proteins (N-mt proteins) from the cytoplasm. So, the N-mt genes encompass genes that had mitochondrial functions and relocated to the nucleus or evolved to gain mitochondrial functions. They are transcribed in the nucleus, translated in the cytoplasm and ultimately the resulting proteins enter the mitochondria by five distinct transport pathways (Wiedemann and Pfanner 2017). Because the mitochondrial proteins are encoded by two different genomes, the successful interactions between N-mt proteins and the 13 proteins encoded by the mitochondrial DNA are critical for all mitochondrial functions (Bar-Yaacov, et al. 2012; Friedman and Nunnari

2014). These are all protein-protein interactions between subunits encoded by these two genomes in most of the mitochondrial OXPHOS complexes (Complex I, III-V), which are essential for the adenosine triphosphate (ATP) production.

In *D. melanogaster*, there are 786 N-mt protein-coding genes (http://www.biomart.org; Ensemble Gene96, BDGP 6.22; (Smedley, et al. 2009)). It is estimated that 24% of the N-mt genes belong to gene families and many (54%) of the duplicated genes have acquired tissue-specific expression (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 testis is not the only tissue with high energy demands. 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-related functions and considered or even observed to replace the parental gene during spermatogenesis or in mature sperm (Wasbrough, et al. 2010).

The unique expression pattern and high enrichment for energy-related functions of these newly duplicated genes has led to several nonexclusive hypotheses about their retention in the fly genome. They might have evolved in response to male mtDNA harming mutations, to resolve intralocus sexual conflict at the parental gene, to partition the pattern of expression or to have more of the same protein (Gallach and Betran 2011; Rogell, et al. 2014). Some of the proteins encoded by these new genes are present in the *Drosophila* Sperm Proteome (DSP; (Wasbrough, et al. 2010)) and 61% (17/28 of the N-mt duplicated genes with testis-biased expression are present in DSP while their parental counterparts are not {Eslamieh, 2017 #129}. Some N-mt duplicated 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) or under positive selection, indicating persistent selection on their functions (Proschel, et al. 2006). Studies have shown that the parental gene is capable of rescuing the phenotype of the duplicated gene revealing that the two genes might have a similar function (Aoidi, et al. 2016) but fails in some other instances revealing that 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 mature sperm functions.

In *D. melanogaster*, there are more N-mt duplicated genes with testis-specific expression for OXPHOS complexes subunits than for any other mitochondrial compartment (12/39; 31%) (Eslamieh, 2019). 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 that form this complex are encoded by both the mitochondrial and the nuclear 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 remaining 10 subunits, including some other cytochrome c oxidase-specific regulatory proteins are encoded by the nuclear genome (N-mt genes which only exist in eukaryotes; (Saraste 1990; Tsukihara, et al. 1996; Barrientos, et al. 2002)). 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*) in *Drosophila* is one of the complex IV N-mt genes which has been duplicated through RNA-mediated duplication. The new copy, called *Cytochrome c Oxidase 4-Like* (*COX4L*, *CG10396*), is believed to still encode a subunit in that complex. *COX4L* is an old duplicate (at least 63 My

old; time of *Drosophila* genus diversification) and it is present in all *Drosophila* species (Zhang, et al. 2010). *COX4* is a highly conserved gene found in all eukaryotes, but so far it has not been found in bacteria that also lack this subunit. In *D. melanogaster*, *COX4* has a high expression in every tissue and is considered to be a non-tissue-biased/housekeeping gene (Sex Biased Ratio (SBR) = 0.58; (Campos, et al. 2018). *COX4L*, however, is highly expressed in testis and is considered to be a male-biased gene (SBR = 9). Since only *COX4L* is present in the sperm proteome but not the parental gene (Wasbrough, et al. 2010), *COX4L* might replace the parental gene function in sperm mitochondria. Transcriptional studies of testes using GeneChip and RNA-Seq analyses have shown that the maximum expression of *COX4L* occurs at the proximal region of testis where the expression is significantly higher in meiosis than post-meiosis or mitosis during spermatogenesis (Vibranovski, et al. 2009; Vedelek, et al. 2018). *COX4* gene has been the subject of multiple studies in a variety of organisms. For example, reduction in COX activity, impaired ATP production, and elevated ROS production have been reported in human patients with *COX4I1* disabling mutations (Abu-Libdeh, et al. 2017). Similarly, knockdown of *COX4* expression in *D. melanogaster* showed a reduction in the rate of mitochondrial respiration, walking speed when driven with *arm-Gal4* driver (drives ubiquitous expression in embryos and larvae) and complete lethality with either of *da-Gal4* or *Tub-Gal4* (drive ubiquitous expression at all developmental stages) drivers (Klichko, et al. 2014).

Here, we study the phylogenetic distribution of *COX4L* and its function in *D. melanogaster* to understand the evolutionary pressures that have led to the retention of this duplicated gene after its origination. Results from 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 in which *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 and or mature sperm than in the female germline and the soma. Through phylogenetic analyses, we also show that the *COX4L* is older than previously thought and it is present in the *Brachycera* suborder of Diptera.

Materials and Methods

COX4L RNAi and viability and fertility tests in knockdown flies

Flies were raised on standard cornmeal/malt medium at room temperature (25°C). All crosses were performed at room temperature except crosses set up to express a UAS transgene under a Gal4 driver for RNAi which were run at either 25°C or 27°C. These temperatures were chosen because although it has been previously reported that the optimal temperature for Gal4-UAS function is 29°C (Duffy 2002), this temperature has been shown to be detrimental to spermatogenesis, thus affecting both viability and fertility results (Ben-David, et al. 2015). 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 lines used in this study 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). This driver was obtained from the Bloomington *Drosophila* Stock Center (Stock #4414). The *bam-Gal4* driver (a germline driver) was crossed with the RNAi lines to study the knockdown in male and female

germlines (fertility test; (Chen and McKearin 2003)) and was obtained from the 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 (See Supplementary table 1). Reciprocal crosses with at least three replicates were performed for all experiments. For the viability test, virgin males and females were collected from either strains 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 was counted on day 15. All viability crosses were made at two different temperatures: 25°C and 27°C. For the fertility test, one virgin male and two virgin females with either the male or the female being an individual where RNAi was driven were kept in the vial for five days and then the number of progeny was counted on day 15. All fertility crosses were performed at two different temperatures: 25°C and 27°C. For both tests, data was analyzed with the R Stats package (http://www.r-project.org; (RCoreTeam 2017).

Generating *COX4* and *COX4L* knockouts and performing viability and fertility tests

We used CRISPR-Cas9 technology to generate *COX4* and *COX4L* knockout flies. 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 two plasmids to express the guide RNAs in germline upon embryo injection (Gratz, et al. 2013). In addition, two homologous arms were designed with the same tool to be assembled in the donor vector

cloned into pHD-DsRed-attP plasmid flanking the eye driven DsRed cassette, designed to replace COX4 and *COX4L* (Addgene # 80898; (Gratz, et al. 2014)). NEBuilder HiFi DNA Assembly kit (NEB, Inc.) was used to assemble the homologous arms flanking the DsRed cassette. The two gRNA 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 homologous arm sequences are provided in Supplementary tables 2 and 3. Flies were collected from injected embryos and crossed with w^{1118} flies. The progeny of these crosses were screened for fluorescent glowing eyes meant to confirm the replacement of the desired gene by the eye-driven DsRed gene. The mutant allele was fixed using balancer chromosomes. Absence of the genes was tested in the homozygote mutant individuals by PCR and sequencing. Two primers inside *COX4* and *COX4L* genes were designed for this purpose (Supplementary tables 2 and 3). Our results confirmed that, in the case of *COX4L*, the gene was excised from the genome (Supplementary figure 1). In the case of *COX4*, PCR and sequencing analysis of the mutant line showed the presence of the entire *COX4* gene adjacent to the DsRed reporter highlighting the essential function of this gene in *D. melanogaster* and suggesting that *COX4* has a strong dose effect and not even one copy of *COX4* can be removed.

A viability test was performed for *COX4L-KO* flies in which the heterozygote virgin males and females were mated for 5 days at 25°C. Flies were discarded from the vial at day five. The number of homozygous offspring was counted on day 15 and made relative to 1/3. For the fertility test of *COX4L-KO*, one homozygote virgin male and two control virgin females were kept in the vial for five days at 25°C, and then the progeny was counted on day 15. T-

tests were performed with the R Stats package (http://www.r-project.org; (RCoreTeam 2017)).

Rescue of the *COX4L-KO* with *COX4L* and *COX4* transgenes

FlyORF stocks of *COX4L* (Fly Line ID: F002652) and *COX4* (Fly Line ID: F005047) were obtained from Zurich ORFeom Project Center 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. UAS-ORF lines are a valuable stock that can be used either for ectopic expression, overexpression, or expression of a gene in the knock-out mutant to rescue the loss of function effects. We drove *COX4L* UAS-ORF with both *bam*- and *nos-Gal4* drivers and *COX4* UAS-ORF with bam-Gal4 to rescue *COX4L-KO* infertility phenotype.

Ectopic expression and overexpression of *COX4L*

The *COX4L* UAS-ORF line was used to study the effect of ectopic expression of *COX4L* in soma and overexpression of *COX4L* in the germline. *Actin5c-Gal4* ubiquitous driver was used for ectopic expression of this gene in the soma and *bam-Gal4* driver was used for overexpression in the germline. Three replicates were performed at 25°C for all crosses. To pinpoint the developmental stage affected by 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 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 adult flies were counted after 15 days.

Mitotracker staining in testes

The testes of one-day-old virgin males were dissected in 1X PBS within 20 minutes, fixed on slides and stained following a formaldehyde fixation protocol (Sitaram, et al. 2014). In brief, the testes were fixed in 4% formaldehyde in PBS plus 0.1% Triton X-100 (PBS-T) for 7 minutes and then washed two times in PBS for 5 minutes at room temperature. Slides containing testes were immersed in PBS-T for 30 min at room temperature to permeabilize cell membranes and washed two times in PBS. MitoTracker™ Deep Red FM in the concentration of 500 nM (Cat # M22426, Invitrogen, Thermo Fisher Scientific) was used to stain mitochondria (Gilmore and Wilson 1999). The slides were washed two times in 1% PBS and then NucBlue™ Fixed Cell ReadyProbes™ Reagent (DAPI); Catalog # R37606, (Invitrogen, Thermo Fisher Scientific) was used to stain the nuclear DNA. The confocal microscope at UT Arlington (Nikon Eclipse Ti2 laser scanning confocal microscope) was used for imaging. NIS-Elements imaging software (Version 5.20.00) was used for image visualization. The fluorescence observed at 60X magnification on the sperm bundles was quantified by the NIS-Elements imaging software by choosing 15 random equally sized regions of interest and the values were compared between *COX4L-KO* and the line of control (w^{1118}). T-test was performed to evaluate if the observed difference in glowing is statistically significant between these two lines.

Computational analyses

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% in *D. melanogaster*). To explore if both genes have the same function, we analyzed both genes for the presence of mitochondrial localization signal (MLS). 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 the protein structure of both genes, domains of both proteins were predicted by Conserved Domain Database (CDD) (Marchler-Bauer, et al. 2017) and Phyre was used to characterize the tertiary structure of predicted domains (Kelley, et al. 2015). STRING v.11 (https://string-db.org) was used to predict protein-protein interactions for both *COX4* and *COX4L* in *D. melanogaster*. In addition, we evaluated 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 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 analyses

The *COX4* and *COX4L* sequences were retrieved from NCBI database. Protein sequences were aligned with MUSCLE (Edgar 2004). Maximum likelihood gene trees (Chor and Tuller 2005) were constructed (Guindon, et al. 2010) using the BlOSUM62 substitution model with 100 bootstrap branch support in PhyML implemented in Geneious

software. We used FigTree (Version 1.4.4) (http://tree.bio.ed.ac.uk/software/figtree) to visualize all protein phylogenies.

Coding sequences of *COX4* and *COX4L* were aligned following protein alignments using Geneious software and the ratio of nonsynonymous to synonymous substitution per site (*d*N/*d*S) was estimated using the CODEML algorithm (Yang 2007) implemented in EasyCodeML (Gao, et al. 2019a). The branch model was used with a null model assuming that each respective group of sequences is evolving at the same rate (one-ratio model) and an alternative model in which the *d*N/*d*S ratio was fixed to *d*N/*d*S =1. We also performed two-ratio branch model analysis to test whether the parental and the retroduplicate genes evolve under different evolutionary constraints. Two-ratio branch model can be used to test whether there are significant ω differences among branches of the tree by assuming that specific branches can have an ω that differs from the rest of the tree (Yang 1998; Yang and Nielsen 1998, 2002; Gao, et al. 2019b). In both approaches, The likelihood ratio test (LRT;(Anisimova, et al. 2001)) was conducted to perform pairwise comparisons of both models for each set of parental and duplicated genes. Only a *P*-value of 0.05 or less in the LRTs was considered to be significant.

Results

COX4L is a well-conserved sperm protein differentiated from *COX4*

COX4L is present in all *Drosophila* species, *Musca domestica*, and *Lucilia cuprina* but not in mosquito and more distantly related genomes. So, *COX4L* originated approximately 126 million years ago through an RNA-mediated duplication (Supplementary figure 1) from *COX4* (divergence time between *Bactrocera oleae* and *Drosophila melanogaster* is 126 My according to timetree.org). A phylogenetic tree using the Maximum Likelihood (ML) model

(Figure 1) shows that the two genes cluster into two distinct clades suggesting that they have been evolving separately since their origin. Contrary to *COX4*, only *COX4L* protein is present 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 counterpart. In addition, the two genes are evolving under different evolutionary rates (dN/dS_{COX4} = 0.06719; dN/dS_{COX4L} = 0.10732), showing that different selective pressures are acting on *COX4* and *COX4L* and *COX4L* is evolving faster than *COX4*. Two-ratio branch model analysis further confirmed that parental sequences are subjected to highly significantly stronger purifying selection than the duplicates ($P<0.001$). However, the dN/dS ratio for both genes is <1 indicating that purifying selection is the main evolutionary force acting on them.

According to the conserved domain analysis (data not shown), 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). This is also reflected in the high degree of similarity between the tertiary structures of both proteins (Supplementary figure 1C).

Mitochondrial localization signal analysis (see Materials and Methods) predicts that both genes have a probability of over 0.9 of being imported into the mitochondria. *COX4* has one cleavage site (mitochondrial processing peptidase or MPP cleavage site) important for cleaving off the presequences once the protein is inside the matrix (Hawlitschek, et al. 1988). *COX4L* on the other hand appears to have an extra cleavage site, i.e., Intermediate cleaving peptidase of 55 kDa (Icp55) cleavage site, 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). A physical protein interaction analysis with STRING v.11, an online biological interaction repository is shown in Supplementary table 4. Both COX4 and

COX4L interact with many common OXPHOS complexes subunits. However, there is a specific partner for COX4L that does not interact with COX4 notably COX5BL. *COX5BL* is a partial DNA-mediated tandem duplicate of the *COX5B* gene on chromosome 2L that in contrary to its parental counterpart that has a broad expression in every tissue, has acquired a testis-specific pattern of expression similarly to *COX4L*.

The evolutionary rate covariance (ERC) can be measured across a phylogeny to find genes that directly interact and coevolve (i.e., genes that have similar evolutionary histories). Typically, a high ERC values between two genes suggests that they are working in a common pathway or have related functions (Clark, et al. 2012, 2013). Therefore, the ERC value can be used to discover previously unknown functional connection between genes (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 male-biased 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 results suggest that *COX4* and *COX4L* have different evolutionary rate covariance with different members of the gene families in agreement with the interaction analysis run above. New N-mt genes might work together in a specialized role during spermatogenesis. Altogether, our data shows that *COX4L* is a well-conserved duplicated sperm protein among *Brachycera* suborder which is also likely functionally differentiated from its parent *COX4*. Further support for the functional differentiation between these proteins is presented below.

Knockdown of *COX4L*

Two different strains from two 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, these UAS lines were crossed with *Actin5c-Gal4* line and the number of progeny was counted (See Materials and Methods for more details). The results between different UAS libraries and temperatures (25°C and 27°C) were consistent with each other, and no significant differences were observed between UAS-Gal4 crosses and controls (P>0.05 in all t-test comparisons; Supplementary table 5) for any of the temperatures we used. These results suggest that *COX4L* is not needed for viability.

To study the effect of *COX4L* on fertility, the RNAi lines (GD and KK libraries) 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 virgin knockdown males and females were collected and crossed with virgin females and males w^{1118} flies, respectively. The progeny of the last crosses was counted for the experimental and control crosses (see Materials and Methods). Knockdown of *COX4L* in the germline causes semi-sterility in males (P-value= 0.043; Figure 2A and Supplementary table 6). However, knockdown of *COX4L* causes an increase in female fertility. The fertility crosses were performed at 25°C and 27°C, but not at 29°C because this temperature has been shown to have a detrimental effect on male fertility (Ben-David, et al. 2015). Our 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; Figure 2B and Supplementary table 6). The results of *COX4L* knockdowns

in soma and germline were consistent with the unique expression pattern of this gene, in which only male infertility was expected.

Knockout of *COX4L* results in male sterility

To confirm the results from the knockdown experiments and to study the function of *COX4L* in more details, we generated a *COX4L* null mutant. Taking advantages of CRISPR-Cas9 technology the entire coding region of *COX4L* was removed from the genome (*COX4L-KO* mutants), and this region was replaced by an eye-driven DsRed gene (See Materials and Methods for more details). Results are shown for one line only. No significant change in viability was observed between homozygote *COX4L-KO* individuals and the controls ($P > 0.05$; Data not shown). However, when we performed the fertility assay on *COX4L-KO* flies, males were completely sterile (Figure 2A). This is a recessive phenotype as *COX4L-KO* heterozygous males are fertile. However, compared to the w^{1118}, fertility of homozygous *COX4L-KO* females increased consistent with *COX4L* knock down results in females (Figure 2B). The male complete infertility was observed on six independent knockout lines (Data not shown).

Male sterility in *COX4L-KO* mutant is due to an individualization defect

We dissected the testes of homozygote knock-out flies to study the sterility phenotype and observed empty seminal vesicles and sperm bundles that fail to individualize and produce mobile sperms (Figure 3A and 3B). These observations suggest that a defect in the sperm individualization step occurs in *COX4L* knockouts resulting in non-obstructive azoospermia in these flies.

Mitochondria membrane potential is reduced in *COX4L-KO*

When stained with Mitotracker, cells with reduced mitochondrial membrane potential will fluoresce less. MitoTracker® dyes passively diffuse across the plasma membrane and accumulate in active mitochondria. Similarly, quantification of fluorescence in *COX4L-KO* testes mitochondria after staining with MitoTracker™ Deep Red FM show a significantly weaker i.e., in average three times less fluorescence signal in the sperm bundles of *COX4L-KO* compared to the control (w^{1118}) (*P*= 1.49E-10; Figure 4) confirming that loss of *COX4L* results in the reduction of mitochondria membrane potential. The electron leakage in the absence of *COX4L* may reduce the membrane potential. Staining of *COX4L-KO* testes mitochondria also show thinner sperm bundles indicative of thinner sperm tail (Figure 4).

Loss of function mutation of *COX4* is not tolerated in *Drosophila melanogaster*

In this study, two gRNAs (Supplementary table 3) were designed to target the entire reading frame of the *COX4* gene. gRNAs were designed so that there were no mismatches between the gRNA target site and the gRNA sequence on *COX4* in the strain used for injection. PCR screening of resulting lines revealed that even the deletion of a single copy of *COX4* is lethal and thus modification of this essential gene is not tolerated in *D. melanogaster*. Sequencing analyses showed that in lines showing the eye DsRed reporter, *COX4* DNA region was not excised out but the gene and regulatory region were intact and immediately followed by the DsRed gene. This reveals that we recover only transformants where the downstream gRNA was successfully used and a rearrangement occurred that incorporated the DsRed gene and supports that even the loss of a single copy of *COX4* is not tolerated in *Drosophila melanogaster*.

Rescue of the *COX4L-KO* phenotype

To confirm that the lack of *COX4L* gene is responsible for the observed infertility phenotype in *COX4L-KO* males and if the phenotype can be rescued by overexpression of the parental gene, we drove the *COX4L* FlyORF and *COX4* FlyORF lines under *bam-Gal4* and *nos-Gal4* drivers (Figure 2A and 2B). The cross scheme is presented in Supplementary figure 2. In all crosses, male fertility was completely rescued by driving *COX4L* and no increased fertility effects were observed for females compared to controls. However, when *COX4* was driven, the fertility phenotype was not rescued further validating a different function between the parental and duplicated gene (Figure 2A).

Overexpression of *COX4L* in soma and germline

COX4L FlyORF line was also used to study the overexpression of *COX4L* in soma and germline. Overexpression of *COX4L* in soma caused complete lethality when driven with *Actin5C-Gal4,* a ubiquitous driver. In which, in the crosses between *Act5C-Gal4*/CyO and *COX4L*-ORF, no progeny without balancer chromosome was observed (Figure 2C). To find at which stage lethality has occurred, we collected and examined all embryos and larvae from the overexpression crosses. The viability effect appears to have happened before larvae stage as no significant difference in larvae viability was observed between overexpression crosses and controls (Figure 2C). More precisely, our results show that lethality occurs at early embryonic stage. The overexpression of *COX4L* with *bam-Gal4* germline driver, in the *COX4L*-ORF X Bam/TM3 crosses, did not show any viability effect. However, overexpression of *COX4L* in germline showed significant fertility reduction in males compared to the control group (Figure 2D). Interestingly, female fertility did not

increase when we overexpressed *COX4L* in the germline. These results suggest that a fine-tuned expression of *COX4L* is necessary for male fertility.

Discussion

Here, we studied *COX4L*, a duplicate of *COX4*, that has testis-specific expression in *D. melanogaster*. We revealed that, in addition to *Drosophila*, this duplication is present in other flies including *Musca domestica* and *Lucilia cuprina*. This observation suggests that *COX4L* is much older than previously estimated, but its origin does not appear to coincide with the advent of giant mitochondria along the sperm tail in these flies, as this is likely an old trait present in many insects and arthropods (Noguchi, et al. 2011).

COX4L is evolving at a faster rate than *COX4* and has different inferred protein interactions, however 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 genes show a high probability of targeting the mitochondria. In addition, COX4L shows an extra cleavage site known to have evolved to increase protein stability in mitochondria (Vogtle, et al. 2009). More importantly, knockdowns of *COX4L* in germline and its complete knockout cause partial and complete infertility in males, respectively. The partial infertility of the knockdown line could be explained by the efficiency of the RNAi or the UAS-Gal4 system which leads to the presence of enough mRNA to show some fertility effect. The complete rescue of the *COX4L-KO* with *COX4L*-ORF but not with *COX4*-ORF line confirms that the male infertility is due to the absence of the duplicate and that parental and duplicated genes have evolved different functions. This phenotype is consistent with the absence of COX4 in sperm where only COX4L is present. Similarly autosomal recessive mutation in *COX4I1* gene in human has been reported to be associated with decreased COX activity in patient's fibroblasts, impaired

ATP production, and increased ROS production (Abu-Libdeh, et al. 2017). The fruit fly spermatogenesis, an energy-demanding process, might proceed without COX4L until the individualization step but might fail afterward because of ATP reduction. Our results show 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 production in the mitochondria and also decrease the mitochondrial membrane potential which depolarizes the mitochondrial membrane and makes it less negative and potentially nonfunctional. Further analyses should reveal if the sperm elongation process is completed in *COX4L-KO* as it seems an energy-demanding step and the unindividualized bundles appear shorter. It is also of interest at what point COX4L replaces COX4 in the mitochondria during spermatogenesis and precisely why.

We know there is at least a 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, potentially higher energy-producing mitochondria in male germline than in female germline and soma if there is a cost, e.g. ROS production and mutations (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 of *Drosophila* and zebrafish (de Paula, Agip, et al. 2013). Mitochondria is metabolically different in male and female gametes. 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) are metabolically active in which transcribe mitochondrial genes for respiratory electron and also produce free radicals (de Paula, Agip, et al. 2013; de Paula, Lucas, et al. 2013) that might cause mtDNA damage but not be selected against because

they are not passing mitochondria to their offspring. All together, these results support the hypothesis that *COX4L* plays a role in higher/specialized energy production at least for sperm function and fertilization.

Acknowledgements

We would like to thank the Betrán lab members for helpful discussions and Victor Palacios and Michael Buszczak for their help producing the *COX4L-KO*. AM would like to acknowledge UTA COS Maverick Science Graduate Research Fellowship and UTA Dissertation Fellowship. This work was supported by the National Institutes of Health under award number R01GM071813 to E.B. and Phi-Sigma Research Grant to M.E. The Nikon Nikon Eclipse Ti2 laser scanning confocal microscope was acquired with the NIH S10 OD025230 award. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

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.

Anisimova M, Bielawski JP, Yang Z. 2001. Accuracy and power of the likelihood ratio test in detecting adaptive molecular evolution. Mol Biol Evol 18:1585-1592.

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. Sci Signal 9:ra9.

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

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

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

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.

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

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

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.

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

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.

de Paula WB, Agip AN, 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 Biol Evol 5:1969-1977.

de Paula WB, Lucas CH, Agip AN, Vizcay-Barrena G, Allen JF. 2013. Energy, ageing, fidelity and sex: oocyte mitochondrial DNA as a protected genetic template. Philos Trans R Soc Lond B Biol Sci 368:20120263.

Dietzl G, Chen D, Schnorrer F, Su KC, 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.

Duffy JB. 2002. 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.

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

Eslamieh M. 2019. Functional Analysis of Nuclear-Encoded Mitochondrial Gene Duplicates with Testis-Biased Expression in *Drosophila melanogaster*. PhD Dissertation. The University of Texas at Arlington.

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.

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 428:363-374.

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. 2011. 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 sexually antagonistic conflict in Drosophila. Genome Biol Evol 2:835-850.

Gao F, Chen C, Arab DA, Du Z, He Y, Ho SYW. 2019a. EasyCodeML: A visual tool for analysis of selection using CodeML. Ecol Evol 9:3891-3898.

Gao F, Chen C, Arab DA, Du Z, He Y, Ho SYW. 2019b. EasyCodeML: A visual tool for analysis of selection using CodeML. Ecology and evolution 9:3891-3898.

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.

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.

Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O. 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst Biol 59:307-321.

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.

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.

Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C, et al. 2012. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics 28:1647-1649.

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.

Klichko V, Sohal BH, Radyuk SN, Orr WC, Sohal RS. 2014. Decrease in cytochrome c oxidase reserve capacity diminishes robustness of Drosophila melanogaster and shortens lifespan. Biochem J 459:127-135.

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

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

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.

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

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

RCoreTeam. 2017. R: A Language and Environment for Statistical Computing. https://www.R-project.org.

Rogell B, Dean R, Lemos B, Dowling DK. 2014. Mito-nuclear interactions as drivers of gene movement on and off the X-chromosome. BMC Genomics 15:330.

Saraste M. 1990. Structural features of cytochrome oxidase. Q Rev Biophys 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.

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

Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, Lopez R, McWilliam H, Remmert M, Soding J, et al. 2011. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol Syst Biol 7:539.

Sitaram P, Hainline SG, Lee LA. 2014. Cytological analysis of spermatogenesis: live and fixed preparations of Drosophila testes. J Vis Exp:e51058.

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

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

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, Grezal G, Kovacs B, Boros IM, Laurinyecz B, Sinka R. 2018. Analysis of Drosophila melanogaster testis transcriptome. BMC Genomics 19:697.

Venken KJ, 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 Genet 5:e1000731.

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.

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.

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

Yang Z. 1998. Likelihood ratio tests for detecting positive selection and application to primate lysozyme evolution. Mol Biol Evol 15:568-573.

Yang Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. Mol Biol Evol 24:1586-1591.

Yang Z, Nielsen R. 2002. Codon-substitution models for detecting molecular adaptation at individual sites along specific lineages. Mol Biol Evol 19:908-917.

Yang Z, Nielsen R. 1998. Synonymous and nonsynonymous rate variation in nuclear genes of mammals. J Mol Evol 46:409-418.

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. J Clin Invest 117:2962-2973.

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

Figures

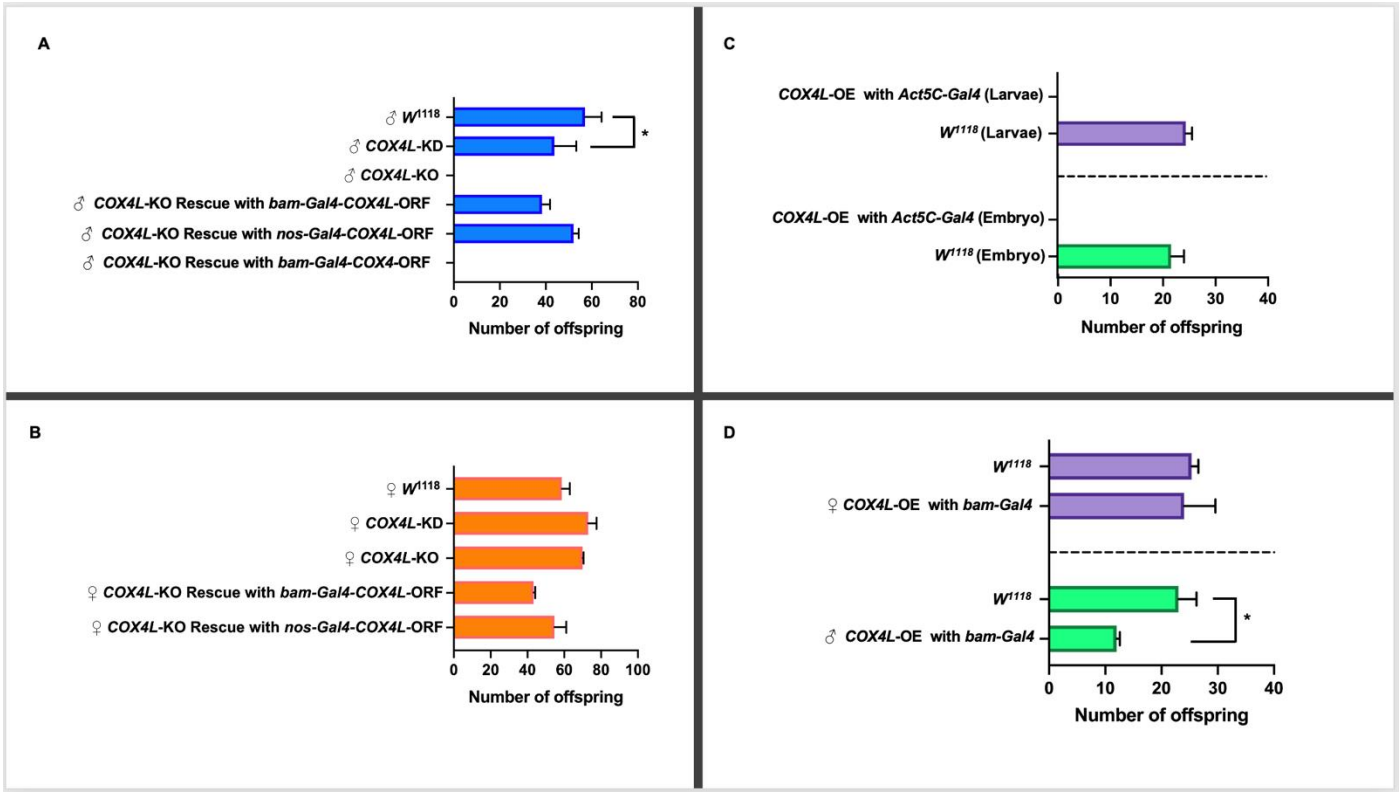


Figure 1. (A) Fertility study of COX4L in males. The knock down of *COX4L* with *bam-Gal4* was semi-fertile while the *COX4-KO* male are complete sterile. The *COX4L-KO* fertility phenotype was completely rescued by *bam-Gal4-COX4L*-ORF and *nos-Gal4-COX4L*-ORF but not with *bam-Gal4-COX4*-ORF. (B) Fertility study of COX4L in females. The knock down of *COX4L* with *bam-Gal4* and *COX4L-KO* females performed better than controls. The overexpression of *COX4L* in soma did not show any viability effect in females. (C) Study of the overexpression of *COX4L* in soma at larvae and embryo stages. (D) Study of the overexpression of *COX4L* in the germline.

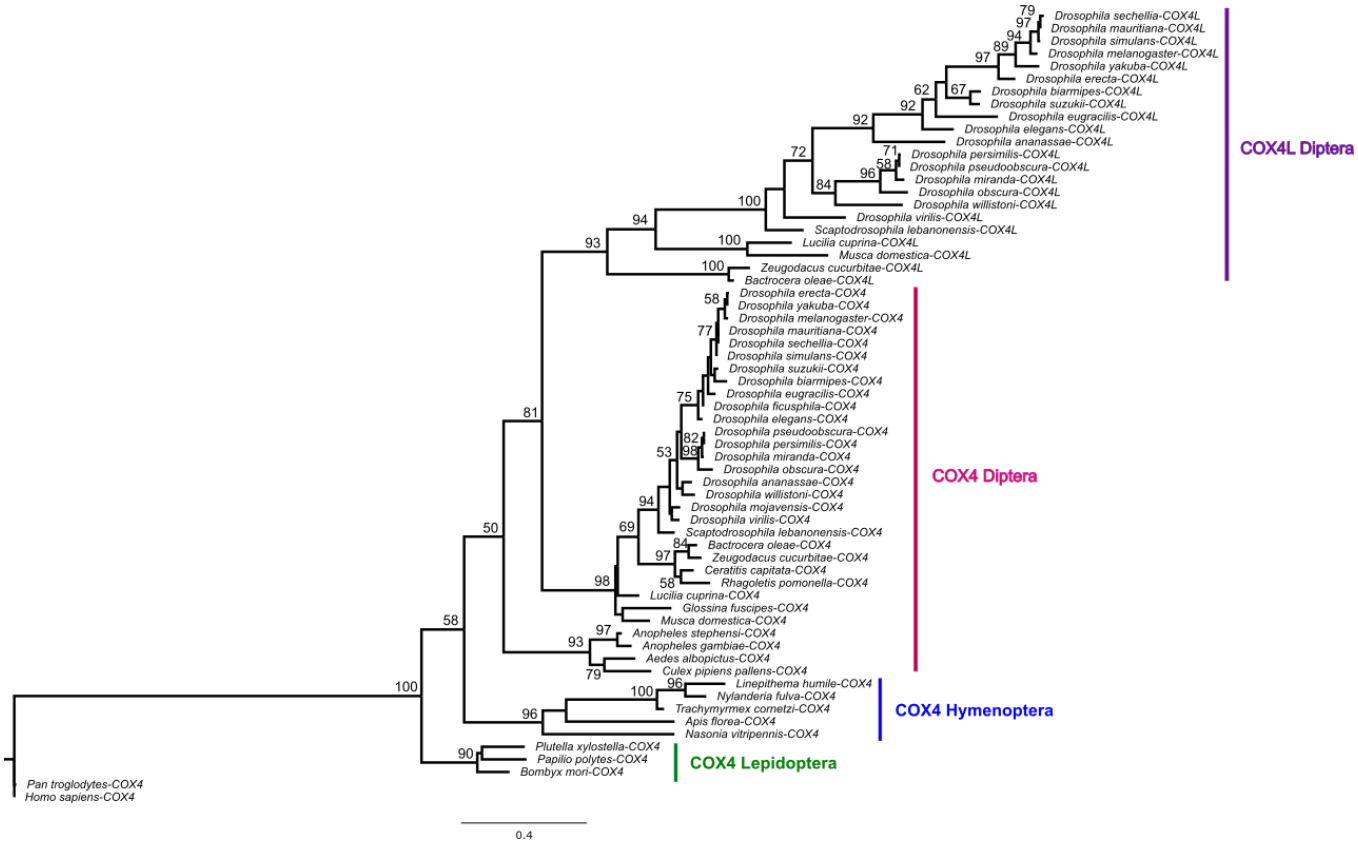


Figure 2. Maximum likelihood tree of COX4 and COX4L using amino acid sequences. Multiple alignments of protein sequences were performed using ClustalW (Sievers, et al. 2011) implemented in Geneious (Version 2020.1) (Kearse, et al. 2012). Maximum likelihood phylogenetic trees of COX4 and COX4L were constructed by using the BlOSUM62 substitution model with 100 bootstrap branch support in PhyML (Guindon, et al. 2010) applied in Geneious. Only bootstrap support values ≥50 are shown. We used FigTree (Version 1.4.4) (http://tree.bio.ed.ac.uk/software/figtree) to visualize all phylogenies.

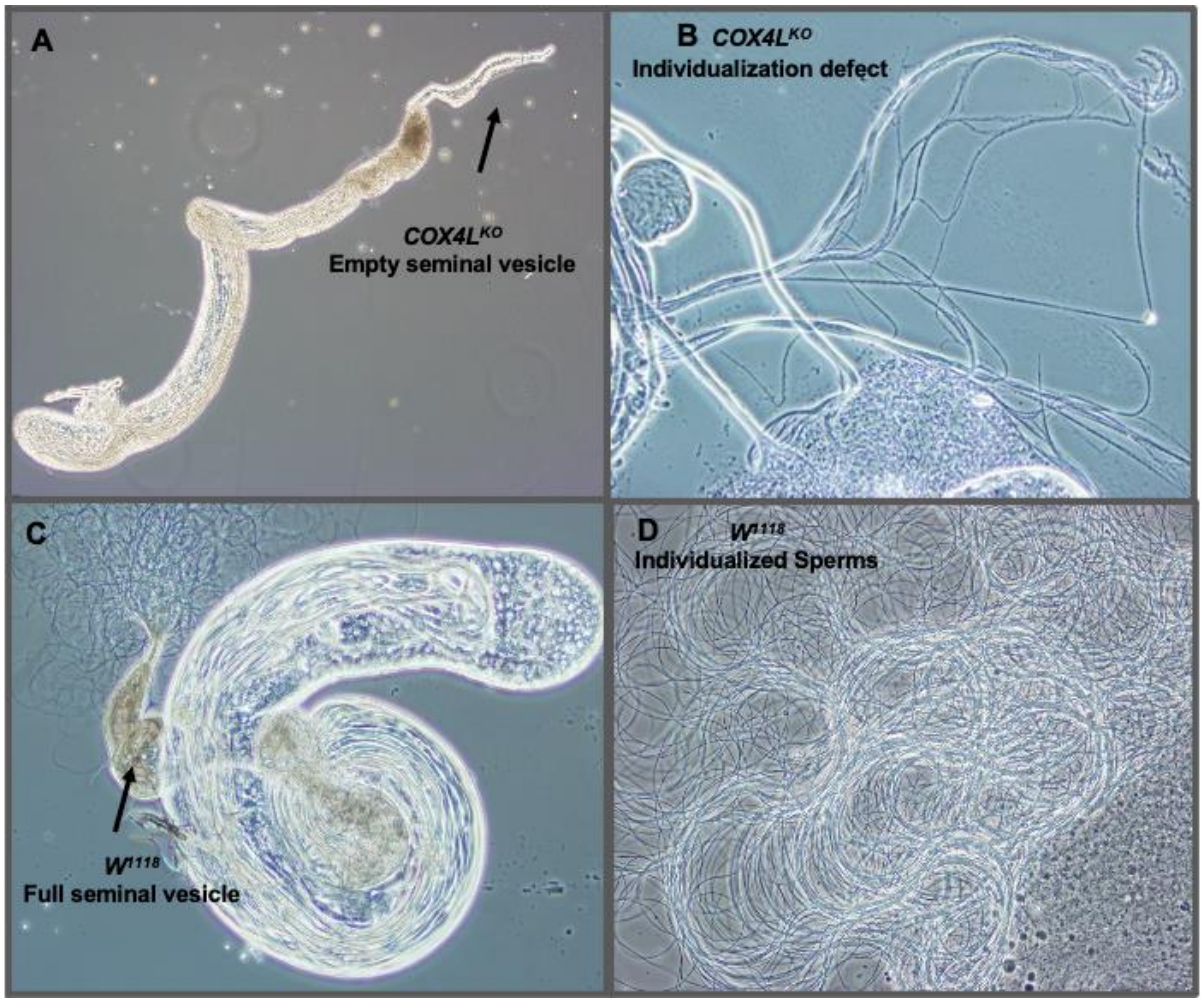


Figure 3. Dissected testes of *COX4L-KO*. (A) Dissected testes of *COX4L-KO* with an empty seminal vesicle are shown. (B) Sperm bundle with individualization defect. (C) w^{1118} control testis with normal spermatogenesis stages and full seminal vesicle. Also shown, mature sperms moving around the raptured seminal vesicle. (D) wildtype mobile individualized sperm.

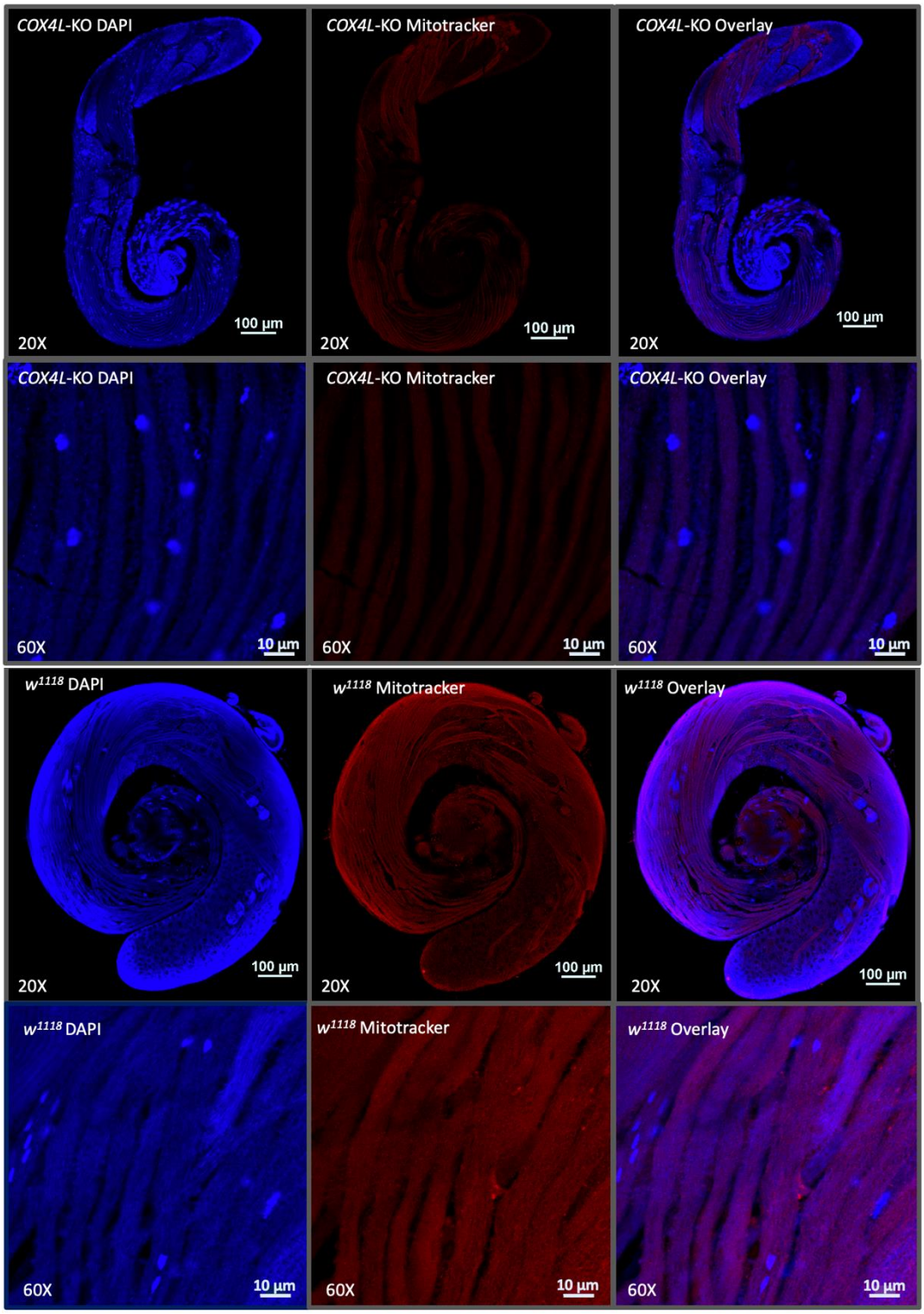


Figure 4. One day old male testes stained with MitoTracker™ Deep Red FM and DAPI. Sperm bundles stained with Mitotracker in *COX4L-KO* (A) show less fluorescence compared to w^{1118} (B) due to reduced mitochondrial membrane potential.

Chapter 2

Supplementary Materials

Supplementary material 1. Genes included in GO: 0051169 (Nuclear transport) in *D. melanogaster* (AmiGO 2 version 2.5.15)

Gene/product	Gene/product name	PANTHER family	Type	Source	Synonyms
eag	ether a go-go	voltage and ligand gated potassium channel pthr10217	protein	FB	CG10952 CIKE Ether-a-go-go ether-a go-go ether-a-gogo
Hph	HIF prolyl hydroxylase	egl nine homolog-related pthr12907	protein	FB	0303/04 31543 more...
Sem1	Suppressor of exocyst mutations 1	26 proteasome complex subunit dss1 pthr16771	protein	FB	BcDNA:RH34416 CG13779 DSS1 Sem1 ortholog Sem1p dmdss1
sqd	squid	rna-binding (rrm/rbd/rnp motifs) family protein pthr48033	protein	FB	BcDNA:GM03761 CG16901 more...
sgg	shaggy	glycogen synthase kinase-3 alpha pthr24057	protein	FB	CG2621 DMSGG3 more...
Karybeta3	Karyopherin beta 3	importin beta pthr10527	protein	FB	CG1059 Karbeta3 more...
edl	ETS-domain lacking	ets pthr11849	protein	FB	CG15085 Ets domain lacking more...
Nxf3	Nuclear export factor 3, isoform B	nuclear rna export factor pthr10662	protein	FB	UniProtKB:Q8IQK4 PTN000070536
wts	warts	serine/threonine-protein kinase pthr24356	protein	FB	CG12072 Dlats more...
Mys45A	Mystery 45A	hsda/sda1-related pthr12730	protein	FB	CG8070 Mys-45A Mystery anon-WO0118547.170
Gle1	Gle1 RNA export mediator	gle-1-related pthr12960	protein	FB	CG14749
Moe	Moesin	merlin/moesin/ezrin/radixin pthr23281	protein	FB	CG10701 D17 more...
SF2	Splicing factor 2	serine/arginine rich splicing factor pthr23147	protein	FB	154911_at ASF ASF/SF2 CG6987 SRSF1 SRp30 dASF dSF2/ASF dSRp30 dmSF2/p28
Dmel\CG11123	RH42110p	nucleolar protein 9 pthr13102	protein	FB	UniProtKB:Q7JX95 PTN000318485
Nup154	Nucleoporin 154kD	nuclear pore complex protein nup155 pthr10350	protein	FB	BcDNA:LD21772 CG4579 more...
Nxt1	NTF2-related export protein 1	nuclear transport factor 2 pthr12612	protein	FB	CG12752 DmNXT1 Nxt dNxt dNxt1 dmNXT dmP15 p15
fu	fused	protein kinase related pthr22983	protein	FB	CG6551 Dm fu Fused d Fu dFu fu[mel] l(1)fu
cdm	cadmus	transportin 3 and importin 13 pthr12363	protein	FB	3R23 CG 7212 CG7212 Cadmus Imp13 Importin 13 Importin-13 complementation group 5 sd-5
Nup93-2	Nucleoporin 93kD-2	complex protein nup93 nucleoporin nup93 dead eye prote	protein	FB	CG7262 Nucleoporin 93 Nup93
Ranbp16	Ranbp16	exportin 4,7-related pthr12596	protein	FB	CG33180 CG9126 CG9136
Mat89Ba	Maternal transcript 89Ba	nucleolar rna-associated protein pthr17972	protein	FB	CG12785
Tnpo-SR	Transportin-Serine/Arginine rich	transportin 3 and importin 13 pthr12363	protein	FB	CG2848 TNPO3 Transportin-serine/arginine rich Trn-SR anon-EST:Posey294 dTRN-SR transportin
Hpr1	Hpr1	tho complex subunit 1 pthr13265	protein	FB	CG2031 Thoc1
tho2	tho2	tho2 protein pthr21597	protein	FB	CG31671 CG4263 Thoc2 macadamia
Dmel\CG9915	Uncharacterized protein, isoform B	protein iws1 homolog pthr46010	protein	FB	UniProtKB:A8JV07 PTN000781462
xmas-2	Protein xmas-2	80 kda mcm3-associated protein pthr12436	protein	FB	UniProtKB:Q9U3V9 PTN000277509

Gene/product	Gene/product name	PANTHER family	Type	Source	Synonyms
Rcc1	Regulator of chromosome condensation 1	family not named pthr45982	protein	FB	BJ1 Bj1 protein more...
chinmo	Chronologically inappropriate morphogenesis	btb domain transcription factor pthr23110	protein	FB	CG10871 CG17156 more...
pix	pixie	atp-binding transport protein-related pthr19248	protein	FB	ABCE1 BEST:GH18088 CG5651 DmCG5651 E3-c Pixie
Nup75	Nucleoporin 75kD	frount protein-related pthr13373	protein	FB	CG5733 Nucleoporin 75 Nup 75
Mapmodulin	Mapmodulin	acidic leucine-rich nuclear phosphoprotein 32 pthr11375	protein	FB	CG5784 CT18148 CT42180 anon-EST:fe3A2 anon-fast-evolving-3A2
Dmel\CG10950	Uncharacterized protein	transportin 3 and importin 13 pthr12363	protein	FB	UniProtKB:A1ZAR6 PTN002285938
S6kII	Ribosomal protein S6 kinase II	ribosomal protein s6 kinase pthr24351	protein	FB	CG17596 D-RSK more...
thoc5	thoc5	fms interacting protein pthr13375	protein	FB	CG2980 garm garmcho
eIF6	eukaryotic translation initiation factor 6	eukaryotic translation initiation factor 6 pthr10784	protein	FB	CG17611 DeIF6 anon-WO0172774.153 anon-WO0172774.155 eIF6 l(2)k13214
Nup153	Nucleoporin 153kD	nuclear pore complex protein nup pthr23193	protein	FB	CG4453 Nucleoporin 153 Nup 153 dNup153 dmNup153
tim	timeless	timeout/timeless-2 pthr22940	protein	FB	Bruchpilot CG3234 Ritsu TIMELESS Tim-1 dTIM dtimeless mel_tim rit s-tim tim1 timeless1
CG12782		mrna export factor and bub3 pthr10971	protein	FB	
Sec13	Secretory 13	ar pore complex protein sec13 / seh1 family member pthr	protein	FB	BEST:CK00043 BcDNA:LD03471 CG6773 CK00043 Dsec13 Sec13 ortholog l(3)01031
Pen	Importin subunit alpha	importin alpha pthr23316	protein	FB	UniProtKB:P52295 PTN002359632
rdx	roadkill	speckle-type poz protein pthr24413	protein	FB	0869/09 BEST:GM07940 more...
Hel25E	Helicase at 25E	atp-dependent rna helicase dbp3 pthr47958	protein	FB	CG7269 Dbp25F more...
Rae1	Rae1	mrna export factor and bub3 pthr10971	protein	FB	CG9862 DRae1 dmrae1
Gp210	Glycoprotein 210 kDa	clear pore membrane glycoprotein gp210-related pthr230	protein	FB	CG14467 CG7897 Dm gp210 Gp188 Gp210 ortholog gp210[[D]] lyadi
ebo	ellipsoid body open	exportin-6 pthr21452	protein	FB	CG3923 EG:165H7.3 Exp1 Exp6 Exportin 6 Exportin-6 ellipsoid-body-open
Ntf-2	Nuclear transport factor-2	nuclear transport factor 2 pthr12612	protein	FB	BcDNA:GM08921 CG1740 more...
Nup35	Nucleoporin 35kDa	nucleoporin nup35 pthr21527	protein	FB	CG6540 Nup53
Pka-C1	rotein kinase, cAMP-dependent, catalytic subunit	cyclic nucleotide-dependent protein kinase pthr24353	protein	FB	6353 C more...
nxf2	nuclear RNA export factor 2	nuclear rna export factor pthr10662	protein	FB	CG4118 dNxf2
Nup205	Nucleoporin 205kD	nuclear pore complex protein nup205 pthr31344	protein	FB	CG11943 Nucleoporin 205 Ov4 anon-EST:Posey281
					CG12234 Exp-5 Exp5

Gene/product	Gene/product name	PANTHER family	Type	Source	Synonyms
Ranbp21	Ranbp21	exportin 1/5 pthr11223	protein	FB	Exportin 5
					Exportin-5
					Exportin5
					dmExp5
Nup93-1	Nucleoporin 93kD-1	complex protein nup93 nucleoporin nup93 dead eye prote	protein	FB	BcDNA:LD21129
					CG11092
					Nup93
xmas	xmas		protein	FB	CG32561
					CG32562 more...
Ran-like	Ran-like	ran gtpase pthr24071	protein	FB	BcDNA:GH25818
					CG7815
					anon-WO0140519.150
Ns3	Nucleostemin 3	large subunit gtpase 1 homolog-related pthr45709	protein	FB	A60
					BACN32G11.5 more...
p53	p53	cellular tumor antigen p53 pthr11447	protein	FB	CG10873
					CG31325 more...
Tnpo	Transportin	importin beta pthr10527	protein	FB	CG7398
					IMPbeta2
					Kapbeta2
					Karyopherin-beta2
					TRN
					Tm
					Transportin-1
					Trn1
					dTNPO
					dTRN
					transportin
Kap-alpha1	karyopherin alpha1	importin alpha pthr23316	protein	FB	CG8548
					Dalpha1 more...
thoc6	thoc6	tho complex subunit 6 homolog pthr44411	protein	FB	CG5632
Klp64D	Kinesin-like protein at 64D	kinesin-related pthr24115	protein	FB	CG10642
					DmKlp64D
					KIF 3A
					KIF3A
					KLP4
					KLP64Ddm
					Kinesin-2
					Kinesin2B
					Klp 64D
Nup107	Nucleoporin 107kD	nup107-related pthr13003	protein	FB	BcDNA:LD18761
					CG6743
					Nucleoporin 107
					Nup
					Nup 107
					Nup170
					dmNup107
Dbp80	Dead box protein 80	atp-dependent rna helicase dbp3 pthr47958	protein	FB	BcDNA:RE44177
					CG17023
					DmDbp80
					HEL80
					Hel40
					Helicase 40
					Helicase 80
					dbp-80
					hel-40
					mdcds_36124
PCID2	PCI domain-containing protein 2	acterized proteasome component region pci-containing pth	protein	FB	CG7351
					Drosophila melanogaster PCID2
					dmPCID2
CG8219		importin beta pthr10527	protein	FB	148274_at
					EP(3)3072
alphaKap4	alpha Karyopherin-4	importin alpha pthr23316	protein	FB	CG10478
					CG11040
					Importin alpha4
					g7295403
Nup58	Nucleoporin 58kD	nucleoporin p58/p45 nucleoporin-like protein 1 pthr13437	protein	FB	CG7360
					Dm NUP58
					Nucleoporin 58
					RE52572p
Mtor	Megator	nucleoprotein tpr-related pthr18898	protein	FB	Bx34
					CG8274
					TPR
					l(2)k03905
					lethal (2) k03905
Nup188	Uncharacterized protein, isoform C	nucleoporin nup188 homolog pthr31431	protein	UniProtKB	UniProtKB:A0A0B4K859
					PTN002418051
e(y)2	enhancer of yellow 2	enhancer of yellow 2 transcription factor pthr12514	protein	FB	CG15191
					ENY2
					LBC
					Sus1
					d
					dE(y)2
					dENY2
					e(y)[2]
					e(y2)
					late boundary complex
Cse1	Chromosome segregation 1	importin-7, 8, 11 pthr10997	protein	FB	BcDNA:LD14270
					CAS
					CAS/CSE1 segregation protein
					CG13281
					Dcas
					Importin-alpha re-exporter
					cse
					l(2)k03902

Gene/product	Gene/product name	PANTHER family	Type	Source	Synonyms
Ndc1	Nuclear division cycle 1	uncharacterized pthr13269	protein	FB	CG5857
					Ndc1 ortholog
Nup98-96	Nucleoporin 98-96kD	nucleoporin pthr23198	protein	FB	CG10198
					CG10201 more...
RanBP3	Ran binding protein 3	ran binding protein pthr23138	protein	FB	CG10225
					GC10225
Dmel\CG14712	LD43047p	nuclear pore complex protein nup pthr23193	protein	FB	UniProtKB:Q9VGL0
					PTN000574083
Nup133	Nucleoporin 133kD	nuclear pore complex protein nup133 pthr13405	protein	FB	CG6958
					Nucleoporin 133
					Nup 133
Ranbp9	Ranbp9	importin-7, 8, 11 pthr10997	protein	FB	CG5252
Nup214	Nucleoporin 214kD	nuclear pore complex protein nup pthr23193	protein	FB	CAN
					CAN/Nup214
					CG3820
					DCAN
					DNup214
					Nucleoporin 214
					Nup
					Nup 214
					l(2)10444
rempA	reduced mechanoreceptor potential A	ift140/172-related pthr15722	protein	FB	CG11838
					CHE-11 more...
LTV1	LTV1 ribosome biogenesis factor	low-temperature viability protein ltv1-related pthr21531	protein	FB	BcDNA:LD21529
					CG7686
					Low Temperature Viability Protein 1
CG8149		sap domain-containing ribonucleoprotein pthr46551	protein	FB	
Ripalpha	RPA-interacting protein alpha	rpa-interacting protein rpain pthr31742	protein	FB	CG18145
					Tes83
					dRIPalpha
CG13137		aladin/adracalin/aaas pthr14494	protein	FB	
nxf4	Nuclear export factor 4	nuclear rna export factor pthr10662	protein	FB	UniProtKB:Q8INS6
					PTN002240678
Kap-alpha3	karyopherin alpha3	importin alpha pthr23316	protein	FB	0335/13
					CG9423 more...
tex	tex	tho complex subunit 3 tho3 pthr22839	protein	FB	CG9615
					TEX1
					Thoc3
Nup62	Nucleoporin 62kD	nuclear pore glycoprotein p62-related pthr12084	protein	FB	CG6251
					Nucleoporin
					Nucleoporin 62
					p62
Fs(2)Ket	Female sterile (2) Ketel	importin beta pthr10527	protein	FB	38E.11
					CG2637 more...
Snup	Snurportin	snurportin1 rnut1 protein rna, u transporter 1 pthr13403	protein	FB	CG1247
					CG32297
					CG42303
					dSNUP
Clbn	Caliban	family not named pthr15239	protein	FB	CG11847
Nup160	Nucleoporin 160kD	nuclear pore complex protein nup160 pthr21286	protein	FB	CG4738
					Nucleoporin 160
					Nup
					Nup160[mel]
					l(2)SH2 2055
					l(2)SH2055
					lethal (2) SH2055
					nuclear pore protein 160
Impbeta11	Importin beta11	importin-7, 8, 11 pthr10997	protein	FB	CG33139
					CG8212 more...
swm	second mitotic wave missing	rna recognition rrm/rnp domain pthr14398	protein	FB	CG10084
					E1
					EC2-10
					S(Rux)2B
					l(2)37Dh
					l(2)E1
					lethal (2) 37Dh
Ntf-2r	CG10174 protein	nuclear transport factor 2 pthr12612	protein	FB	UniProtKB:Q9VJ85
					PTN000289086
Nup358	Nucleoporin 358kD	ran binding protein pthr23138	protein	FB	CG11856
					Nucleoporin 358
					Q9VBU7_DROME
					RanBP2
Nbp70	Nuclear localization sequence-binding protein 70		gene_produc	FB	
e(y)2b	enhancer of yellow 2b	enhancer of yellow 2 transcription factor pthr12514	protein	FB	CG14612
ZC3H3	ZC3H3	ccch zingc finger pthr46156	protein	FB	CG6694
					dZC3H3
Torsin	Torsin	torsin pthr10760	protein	FB	CG3024
					DTor
					DmTorp4a
					EG:84H4.1
					TORP4A
					dtorsin
Phax	Phosphorylated adaptor for RNA export	cytosolic resiniferatoxin binding protein rbp-26 pthr13135	protein	FB	BEST:GH22533
					CG8069
					PHosphorylated Adaptor for RNA Export
					Similar to PHAX
					dPHAX
Aladin	Aladin	aladin/adracalin/aaas pthr14494	protein	FB	CG16892
Apl	Apollo	importin beta pthr10527	protein	FB	CG18218
					CG18861
					CG32165
					Imp4/5
					anon-73Bb
					CG18219

Gene/product	Gene/product name	PANTHER family	Type	Source	Synonyms
Arts	Artemis	importin beta pthr10527	protein	FB	CG18860
					CG32164
Pen	Pendulin		protein	FB	2.1
					CG4799 more...
CG10286		uncharacterized pthr13347	protein	FB	
emb	embargoed	exportin 1/5 pthr11223	protein	FB	CG13387
					CRM1 more...
U2af50	U2 small nuclear riboprotein auxiliary factor 50	rna-binding protein pthr23139	protein	FB	21-Sep
					CG9998 more...
Su(fu)	Suppressor of fused	suppressor of fused pthr10928	protein	FB	CG6054
					SUFU
					Suppressor of Fused
					d Sufu
					dSufu
Nup50	Nucleoporin 50kD	ran binding protein pthr23138	protein	FB	CG2158
					Nuclear pore protein 50
					Nucleoporin 50
					Nup
Ran	Ran	ran gtpase pthr24071	protein	FB	AAF30287
					CG1404
					Q9VZ23
					Ran GTPase
					RanGTP
					dran
					l(1)G0075
					ran10A
oho31	overgrown hematopoietic organs-31		gene_product	FB	l(2)144/1
					l(2)k14401
					l(2)oho31
					oho-31
Dmel\CG32409	Ribosome biogenesis regulatory protein	ribosome biogenesis regulatory protein pthr17602	protein	FB	UniProtKB:Q8I937
					PTN000426715
Abl	Abl tyrosine kinase	tyrosine-protein kinase pthr24418	protein	FB	4674
					Abelson more...
msk	moleskin	importin-7, 8, 11 pthr10997	protein	FB	61
					CG7935 more...
mbo	members only	nucleoporin nup84-related pthr13257	protein	FB	1465/07
					1465/7
					CG6819
					DNup88
					Dnup 88
					Members only
					NUP88
					anon-WO0118547.263
					l(3)05043
					l(3)S146507
Vhl	von Hippel-Lindau	von hippel-lindau protein pthr15160	protein	FB	BcDNA:RH61560
					CG13221
					DVhl
					Von Hippel Lindau
					d-VHL
					dmVHL
					von Hippel Lindau protein
CG13926	Protein OPI10 homolog	hikeshi family member pthr12925	protein	FB	UniProtKB:Q9W0C7
					PTN000309101
RanGAP	Ran GTPase activating protein	ran gtpase-activating protein 1 pthr24113	protein	FB	CG9999
					CT28175 more...
sbr	small bristles	nuclear rna export factor pthr10662	protein	FB	CG1664
					CG17335 more...
thoc7	thoc7	maintenance of killer 16 mak16 protein-related pthr23405	protein	FB	CG17143
Rpb4	RNA polymerase II subunit Rpb4	dna-directed rna polymerase ii pthr21297	protein	FB	CG31237
					CG33520
					CG7150
					DmRPB4
					RNA polymerase II
					dRPB4
Nmd3	60S ribosomal export protein NMD3	nonsense-mediated mrna decay protein 3 pthr12746	protein	FB	UniProtKB:O46050
					PTN000297378
Nup188	Nucleoporin 188kDa		protein	FB	CG8771
Nup54	Nucleoporin 54kD	nucleoporin p54 pthr13000	protein	FB	CG8831
					Nucleoporin 54
mgr	merry-go-round	prefoldin subunit 3 pthr12409	protein	FB	BcDNA:LD34406
					CG6719
					Merry-go-round
					PAC10
					Pfdn3
					Prefoldin 3
mtgo	miles to go	e3 ubiquitin-protein ligase trim36-related pthr24099	protein	FB	CG13260
					CG13261
					CG31738
					CG42389
					CG4668
					CT15043
					CT15049
					anon-EST:Liang-1.16
					clone 1.16
Gld	Glucose dehydrogenase	glucose-methanol-choline gmc oxidoreductase pthr11552	protein	FB	CG1152
					GDH
					Glucose oxidase
					Go
					glucose dehydrogenase
					l(3)84Cg
					CG3182
					Derg

Gene/product	Gene/product name	PANTHER family	Type	Source	Synonyms
sei	seizure	voltage and ligand gated potassium channel pthr10217	protein	FB	Ether-a-go-go-related Seizure Seizure/HERG eag related gene erg seit seizure potassium channel
CG12290	G protein alpha o subunit	gtp-binding protein alpha subunit pthr10218	protein	FB	Bkh CG2204 more...
Dop2R	Dopamine 2-like receptor	energic receptor-related g-protein coupled receptor pthr24	protein	FB	CG17004 CG33517 more...
CG12290		energic receptor-related g-protein coupled receptor pthr24	protein	FB	anon-WO0170980.49 anon-WO0170980.50 lincRNA.S7704
pelo	pelota	pelota pthr10853	protein	FB	CG3959 DmPelota Pelo Hbs1 Pelota ms(2)01559

Supplementary material 2. List of studied species and their genome quality scores

FlyBase insect species	Ensembl Drosophila species	RefSeq assembly accession:	Scaffold N50	Contig N50
Sophophora (subgenus)				
Drosophila melanogaster[1,2,3,4]	Drosophila melanogaster[1,2,3,4]	GCF_000001215.4	25,286,936	21,485,538
Drosophila simulans[22]	Drosophila simulans (w501)	GCF_016746395.1	23,399,919	22,319,038
Drosophila sechellia[6,7]	Drosophila sechellia (Rob3c)	GCF_004382195.1	24,956,976	19,907,079
Drosophila yakuba[6,7]	Drosophila willistoni (TSC#14030-0811.24)	GCF_016746365.1	25,180,726	18,744,349
Drosophila erecta[6,7]	Drosophila erecta (TSC#14021-0224.01)	GCF_003286155.1	-	22,146,549
Drosophila ficusphila[21]	Drosophila ananassae (TSC#14024-0371.13)	GCF_000220665.1	1,050,541	275,894
Drosophila eugracilis[21]	Drosophila persimilis (MSH-3)	GCF_000220665.1	1,050,541	275,894
Drosophila biarmipes[21]	Drosophila pseudoobscura pseudoobscura (MV2-25)	GCF_000233415.1	3,386,121	474,639
Drosophila takahashii[21]	Drosophila willistoni (TSC#14030-0811.24)	GCF_000224235.1	387,676	126,259
Drosophila elegans[21]	Drosophila mojavensis (TSC#15081-1352.22)	GCF_000224195.1	1,714,184	212,818
Drosophila rhopaloa[21]	Drosophila virilis (TSC#15010-1051.87)	GCF_000236305.1	45,514	19,484
Drosophila kikkawai[21]	Drosophila grimshawi (TSC#15287-2541.00)	GCF_000224215.1	903,682	209,056
Drosophila ananassae[6,7]		GCF_003285975.2	-	6,212,830
Drosophila bipectinata[21]		GCF_000236285.1	663,995	149,088
Drosophila pseudoobscura pseudoobscura[5,6,7]		GCF_009870125.1	32,422,566	30,706,867
Drosophila persimilis[6,7]		GCF_003286085.1	-	5,212,974
Drosophila miranda[17]		GCF_003369915.1	35,263,383	11,978,448
Drosophila willistoni[6,7]		GCF_000005925.1	4,511,350	180,217
Drosophila (subgenus)				
Drosophila mojavensis[6,7]		GCF_000005175.2	24,764,193	121,517
Drosophila virilis[6,7]		GCF_003285735.1	-	8,697,263
Drosophila albomicans[18]		GCF_009650485.1	33,427,555	33,427,555
Drosophila grimshawi[6,7]		GCF_000005155.2	8,399,593	91,192
Musca domestica (House fly)[19]		GCF_000371365.1	226,573	11,807
Glossina morsitans morsitans (Tsetse fly)[20]		GCA_001077435.1 (GenBank)	-	49,769
Culex quinquefasciatus (Southern house mosquito)[6,7]		GCF_015732765.1	201,550,677	2,875,282
Aedes aegypti (Yellow fever mosquito)[13]		GCF_002204515.2	409,777,670	11,758,062
Anopheles darlingi (American malaria mosquito)[16]		GCA_000211455.3 (GenBank)	-	115,072
Anopheles gambiae (Malaria mosquito)[11,12]		GCA_001542645.1 (GenBank)	-	101,465
Mayetiola destructor (Hessian fly)[16]		GCA_000149185.1 (GenBank)	756,041	14,032
Lepidoptera				
Bombyx mori (silkworm)[9,10]		GCF_014905235.1	16,796,068	12,201,325
Danaus plexippus (Monarch butterfly)[16]		GCF_009731565.1	9,209,872	108,026
Hymenoptera				
Nasonia giraulti (Parasitic wasp)[16]		GCA_016647725.1 (GenBank)	545,346	34,917
Nasonia longicornis (Parasitic wasp)[16]		GCA_000004795.1 (GenBank)	758,407	1,876
Nasonia vitripennis (Parasitic wasp)[16]		GCF_009193385.2	7,180,486	7,180,486
Apis mellifera (Western honey bee)[8]		GCF_003254395.2	13,619,445	5,382,476
Apis florea (Dwarf honey bee)[16]		GCF_000184785.3	2,863,240	24,915
Bombus impatiens (Common eastern bumblebee)[16]		GCF_000188095.3	1,399,493	59,072
Bombus terrestris (Buff-tailed bumblebee)[16]		GCF_000214255.1	3,506,793	76,043
Megachile rotundata (Alfalfa leafcutter bee)[16]		GCF_000220905.1	1,699,680	64,153
Acromyrmex echinatior (Panamanian leafcutter ant)[16]		GCF_000204515.1	1,110,580	80,630
Atta cephalotes (Leafcutter ant)[16]		GCF_000143395.1	5,154,485	14,798
Camponotus floridanus (Florida carpenter ant)[16]		GCF_003227725.1	1,585,631	1,278,439
Harpegnathos saltator (Jerdon's jumping ant)[16]		GCF_003227715.1	1,078,644	911,506
Linepithema humile (Argentine ant)[16]		GCF_000217595.1	1,402,257	35,858
Pogonomyrmex barbatus (Red harvester ant)[16]		GCF_000187915.1	819,605	11,605
Solenopsis invicta (Red fire ant)[16]		GCF_016802725.1	26,227,205	9,421,447
Hemiptera				
Acyrthosiphon pisum (Pea aphid)[16]		GCF_005508785.1	132,544,852	25,858
Rhodnius prolixus (Kissing bug)[16]		GCA_000181055.3 (GenBank)	1,088,772	35,751
Phthiraptera				
Pediculus humanus corporis (Human body louse)[16]		GCF_000006295.1	497,057	34,097
Ixodida				
Ixodes scapularis (Deer tick)[16]		GCF_016920785.1	-	1,735,392
Rhipicephalus microplus (Southern cattle tick)[16]		GCF_013339725.1	183,350,851	1,791,079

Genome Reference

1. Celniker SE et al., Genome Biol. 2002;3(12):RESEARCH0079. Epub 2002 Dec 23. PMID: 12537568
2. Misra et al., Genome Biol. 2002;3(12):RESEARCH0083. Epub 2002 Dec 31. PMID: 12537572
3. Kaminker et al., Genome Biol. 2002;3(12):RESEARCH0084. Epub 2002 Dec 23. PMID: 12537573
4. Hoskins et al., Science. 2007 Jun 15; 316(5831):1625-1628. PMID: 17569867
5. Richards et al., Genome Research 2005 Jan; 15(1):1-18 PMID: 15632085
6. Drosophila 12 Genomes Consortium, Nature 2007 Nov 8; 450:203-218 PMID: 17994087
7. Stark, A, et al., Nature 2007 Nov 8; 450:219-232 PMID: 17994088
8. The Honeybee Genome Sequencing Consortium, Nature 2006 Oct 26; 443(7114):931-949 PMID: 17073008
9. Mita, K, et al., DNA Res. 2004 Feb 29; 11(1):27-35 PMID: 15141943
10. Xia, Q, et al., Science 2004 Dec 10; 306(5703):1937-40 PMID: 15591204
11. Holt et al., Science 2002 Oct 4; 298(5591):129-149 PMID: 12364791
12. Zdobnov, EM, et al., Science 2002 Oct 4; 298(5591):149-159 PMID: 12364792
13. Nene et al., Science 2007 Jun 22; 316(5832):1718-1723 PMID: 17510324
14. Tribolium Genome Sequencing Consortium, Nature 2008 Mar 23; doi:10.1038/nature06784 PMID: 18362917
15. modENCODE Comparative Genomics WhitePaper PDF
16. GenBank Whole Genome Shotgun sequencing project list
17. Zhou Q, Bachtrog D. Science. 2012 Jul 20;337(6092):341-5. PMID: 22822149
18. Zhou Q, et al. BMC Genomics 2012, 13(109) doi:10.1186/1471-2164-13-109
19. Scott JG, et al. Genome Biol 2014 Oct 14;15(10):466 PMID: 25315136
20. Lawson, D. et al. International Glossina Genome Initiative W.H.O. VectorBase
21. Chen ZX et al., Genome Res. 2014 Jul;24(7):1209-23 PMID: 24985915
22. Hu T, et al., Genome Res. 2013 Jan; 23(1): 89-98. PMID: 22936249

Supplementary material 3.1. *Ntf-2* Blast search results

Query	Database	Species|Chr	Location	Expect	Identities	Positives	Query beg.	Query end	Subj. beg.	Subj. end	Release
Dmel Ntf-2 (CG1740)- PA	FlyBase	dmel|2L	2L	4.44E-63	114 / 130 (87.7%)	120 / 130 (92.3%)	1	130	18454767	18455156	r6.22
		dsim|Scf_2L	Scf_2L:1..23539531	1.26E-64	114 / 130 (87.7%)	122 / 130 (93.8%)	1	130	17976692	17977081	r2.02
		dsec|scaffold_7	scaffold_7:1..3727775	6.43E-63	112 / 130 (86.2%)	118 / 130 (90.8%)	1	130	2076161	2076550	r1.3
		Dyak	-	-	-	-	-	-	-	-	-
		dere	-	-	-	-	-	-	-	-	-
		Dtak	-	-	-	-	-	-	-	-	-
		Dbiarm|Unplaced genomicScaffold	scf7180000302428	3.18625e-35	67 / 130 (51.5%)	92 / 130 (70.8%)	1	130	4810704	4810315	r1.0
		Drhoplaloa|Unplaced genomic scafold	scf7180000779428	7.27E-39	71 / 124 (57.3%)	92/124(74%)	1	124	95585	95214	r1.0
		Delegans	-	-	-	-	-	-	-	-	-
		Deugracilis	-	-	-	-	-	-	-	-	-
		Dficusphila|Unplaced genomic scafold	scf7180000453932	2.23E-71	127 / 130 (97.7%)	129 / 130 (99.2%)	1	130	418717	418328	r1.0
		Dkikkawai	-	-	-	-	-	-	-	-	-
		dana|scaffold_13337	scaffold_13337:1..23293914	6.94E-42	78 / 131 (59.5%)	99 / 131 (75.6%)	1	129	13348900	13348508	r1.05
		dana|scaffold_13337	scaffold_13337:1..23293914	9.02E-31	63 / 126 (50%)	85 / 126 (67.5%)	1	126	17919135	17918761	r1.05
		Dbipectinata|Unplaced genomic scafold	scf7180000396741	1.05E-27	54/120(45%)	82/120(68%)	7	126	1097871	1097515	r1.0
		dpse|4_group3	4_group3:1..11685562	4.98E-11	34 / 76 (44.7%)	47 / 76 (61.8%)	35	93	4176807	4176580	r3.04
			4_group3:1..11685562	8.10E-06	22 / 46 (47.8%)	28 / 46 (60.9%)	1	46	4177155	4177024	r3.04
			4_group3:1..11685562	0.000235599	20 / 34 (58.8%)	25 / 34 (73.5%)	92	125	4176375	4176274	r3.04
		dper|scaffold_1	scaffold_1:1..10282868	4.04E-11	34 / 76 (44.7%)	47 / 76 (61.8%)	35	93	5665290	5665063	r1.3
			scaffold_1:1..10282868	1.79E-05	22 / 34 (64.7%)	26 / 34 (76.5%)	92	125	5664782	5664681	r1.3
			scaffold_1:1..10282868	1.54E-01	15 / 34 (44.1%)	20 / 34 (58.8%)	13	46	5665602	5665507	r1.3
			scaffold_1:1..10282868	3.08112	11 / 21 (52.4%)	15 / 21 (71.4%)	78	98	7463897	7463835	r1.3
		Dmir| strain MSH22	chromosome 4	1.98E-09	35 / 76 (46.1%)	43 / 76 (56.6%)	35	93	20130529	20130302	r1.0
			chromosome 4	3.81E-07	24 / 46 (52.2%)	31 / 46 (67.4%)	1	46	20130870	20130739	r1.0
			chromosome 4	9.62E-04	22 / 34 (64.7%)	25 / 34 (73.5%)	92	125	20130168	20130070	r1.0
		dwil|scf2_1100000004909	-	-	-	-	-	-	-	-	-
		dmoj|scaffold_6473	-	-	-	-	-	-	-	-	-
		dvir|scaffold_12970	-	-	-	-	-	-	-	-	-
		Dalbomicans	-	-	-	-	-	-	-	-	-
		dgri|scaffold_15110	scaffold_15110:1..24565398	1.13E-67	117 / 130 (90%)	126 / 130 (96.9%)	1	130	16614828	16615217	r1.05
Glossina morsitans		gi|594167218|emb|CCAG010007352.1		1.04E-49	96 / 131 (73.3%)	105 / 131 (80.2%)	2	129	18656	18264	r1.0
Anopheles gambiae			chromosome 2L	7.05E-70	122 / 130 (93.8%)	128 / 130 (98.5%)	1	130	41026040	41026429	r1.0

Supplementary material 3.2. *Ran* Blast search results

Query	Database	Species|Chr	Location	Genome annotated	Expect	Identities	Positives	Query beg.	Query end	Subj. beg.	Subj. end	Release
Dmel Ran (CG1404)- PA	FlyBase	dmel|3L	3L	Yes	2.28E-71	123 / 200 (61.5%)	152 / 200 (76%)	11	210	15485328	15484729	r6.32
	FlyBase	dsim|Scf_3L	3L	Yes	1.02794E-65	117 / 201 (58.2%)	147 / 201 (73.1%)	10	210	15111039	15110437	r2.02
	FlyBase	dsim|Scf_2L	2L	Yes	1.02794E-65	89 / 176 (50.6%)	114 / 176 (64.8%)	11	186	2318645	2318145	r2.02
	FlyBase	dsec|scaffold_0	scaffold_0:1..21120651	Yes	6.85E-65	116 / 201 (57.7%)	147 / 201 (73.1%)	10	210	7587101	7586499	r1.3
	FlyBase	dsec|scaffold_5	scaffold_5:1..5866729	Yes	2.27E-44	89 / 176 (50.6%)	113 / 176 (64.2%)	11	186	599003	598503	r1.3
	FlyBase	**dyak|v2_chr3L_random_081**	v2_chr3L_random_081:1..146109	Yes	3.08E-77	106 / 141 (75.2%)	124 / 141 (87.9%)	6	146	98875	98459	r1.05
	FlyBase		v2_chr3L_random_081:1..146110	Yes	3.08E-77	36 / 70 (51.4%)	48 / 70 (68.6%)	149	216	98457	98248	r1.05
	FlyBase	dyak|3L	3L:1..24197627	Yes	1.15E-72	93 / 145 (64.1%)	112 / 145 (77.2%)	74	216	17992831	17992397	r1.05
	FlyBase	dyak|2L	2L:1..22324452	Yes	7.61E-44	87 / 191 (45.5%)	119 / 191 (62.3%)	11	200	2414112	2413555	r1.05
	FlyBase	dere|scaffold_4784	scaffold_4784:1..25762168	Yes	8.91E-72	127 / 210 (60.5%)	157 / 210 (74.8%)	1	210	17732891	17732271	r1.05
	FlyBase	dere|scaffold_4929	scaffold_4929:1..26641161	Yes	8.69E-45	85 / 173 (49.1%)	112 / 173 (64.7%)	11	180	2467653	2467141	r1.06
	FlyBase	Dtak|Unplaced genomicScaffold KB461111.1	scf7180000415168	No	1.86E-85	138/215(64%)	175 / 215 (81.4%)	1	215	20568	21212	r1.0
	FlyBase	**Dtak|Unplaced genomicScaffold**	scf7180000415409	No	2.79E-72	127 / 145 (87.6%)	133 / 145 (91.7%)	1	145	410293	409859	r1.0
	FlyBase		scf7180000415409	No	2.37E-29	61 / 70 (87.1%)	65 / 70 (92.9%)	143	212	409350	409147	r1.0
	FlyBase	Dtak|Unplaced genomicScaffold|lcl|KB460641.1	scf7180000413208	No	6.21173e-48	90 / 191 (47.1%)	122 / 191 (63.9%)	10	198	137477	136908	r1.0
	FlyBase	Dbiarm|	scf7180000396427	No	2.00E-157	201/216(93%)	213/216(98%)	1	216	1354454	1353807	r1.0
	FlyBase	Dbiarm	scf7180000302402	No	2.06E-115	129 / 145 (89%)	139 / 145 (95.9%)	1	145	5007194	5007628	r1.0
	FlyBase			No	2.06E-115	71 / 76 (93.4%)	ositives = 75 / 76 (98.7	141	216	5007613	5007840	r1.0
	FlyBase	Drhop|Unplaced genomicScaffold	scf7180000761302	No	4.15806e-121	204 / 210 (97.1%)	207 / 210 (98.6%)	1	210	127747	127118	r1.0
	FlyBase	Drhop|Unplaced genomicScaffold	scf7180000779279	No	4.17E-81	136 / 211 (64.5%)	170 / 211 (80.6%)	7	216	9538	8906	r1.0
	FlyBase	Drhop|Unplaced genomicScaffold	scf7180000779970	No	2.70E-59	108 / 206 (52.4%)	147 / 206 (71.4%)	11	215	94326	93724	r1.0
	FlyBase	Dele|Unplaced genomicScaffold	scf7180000491273	No	3.96E-62	112 / 201 (55.7%)	149 / 201 (74.1%)	11	210	807346	807942	r1.0
	FlyBase	Deug|Unplaced genomicScaffold	scf7180000409554	No	9.43E-55	100 / 181 (55.2%)	129 / 181 (71.3%)	8	187	3703755	3704291	r1.0
	FlyBase	Deug|Unplaced genomicScaffold	scf7180000408844	No	6.76E-25	50 / 82 (61%)	63 / 82 (76.8%)	6	87	202042	201797	r1.0
	FlyBase	Deug|Unplaced genomicScaffold		No	1.95E-19	52 / 121 (43%)	72 / 121 (59.5%	85	201	201849	201496	r1.0
	FlyBase	Dfic|Unplaced genomicScaffold	scf7180000453842	No	1.20E-124	209 / 216 (96.8%)	215 / 216 (99.5%)	1	216	1183468	1184115	r1.0
	FlyBase	Dfic|Unplaced genomicScaffold	scf7180000454044	No	2.00E-158	206 / 216 (95.4%)	212 / 216 (98.1%)	1	216	398663	399310	r1.0
	FlyBase	Dkik|Unplaced genomicScaffold	scf7180000302634	No	2.80E-35	70 / 128 (54.7%)	81 / 128 (63.3%)	19	145	529460	529807	r1.0
	FlyBase	Dkik|Unplaced genomicScaffold		No	2.80E-35	23 / 41 (56.1%)	28 / 41 (68.3%)	170	209	529917	530039	r1.0
	NCBI	Sequence ID: XM_017177386.1	LOC108082090	No	3.00E-46	83/209(40%)	121/209(57%)	8	215	28	597	
	NCBI	Sequence ID: XM_017173662.1	LOC108079355	No	1.00E-28	65/176(37%)	96/176(54%)	6	179	428	904	
	FlyBase	Dana|scaffold_12943	scaffold_12943:1..5039921	Yes	9.14E-105	174 / 210 (82.9%)	194 / 210 (92.4%)	1	210	3460108	3460737	r1.05
	FlyBase	Dbip|Unplaced genomicScaffold	scf7180000396427	No	1.38E-125	212 / 216 (98.1%)	214 / 216 (99.1%)	1	216	1354454	1353807	r1.0
	FlyBase	Dbip|Unplaced genomicScaffold	scf7180000396433	No	9.42E-64	109 / 148 (73.6%)	127 / 148 (85.8%)	67	214	352964	353407	r1.0
	FlyBase	dpse|4_group4	4_group4:1..6594820	Yes	8.99E-46	96 / 193 (49.7%)	127 / 193 (65.8%)	10	201	3164146	3163586	r3.04
	FlyBase	dper|scaffold_10	scaffold_10:1..3432795	Yes	1.54E-45	95 / 194 (49%)	127 / 194 (65.5%)	10	202	2174557	2173994	r1.3
	FlyBase	Dmir| strain MSH22 chromosome 4	Chr 4	No	5.77E-48	102 / 207 (49.3%)	133 / 207 (64.3%)	10	215	3524849	3525451	r1.0
	FlyBase	dwil|scf2_1100000004585	scf2_1100000004585:1..8906247	Yes	7.38E-99	166 / 210 (79%)	190 / 210 (90.5%)	7	216	7110884	7111513	r1.05
	FlyBase	dwil|scf2_1100000004521	scf2_1100000004521:1..12563649	Yes	3.59E-51	91 / 175 (52%)	121 / 175 (69.1%)	12	186	8602852	8603367	r1.05
	FlyBase	dmoj|scaffold_6540	scaffold_6540:1..34148556	Yes	9.93E-80	150 / 262 (57.3%)	179 / 262 (68.3%)	1	204	7014889	7015671	r1.04
	FlyBase	dvir|scaffold_12855	scaffold_12855:1..10161210	Yes	2.67E-64	110 / 152 (72.4%)	133 / 152 (87.5%)	5	156	6639654	6640109	r1.06
	FlyBase	dvir|scaffold_12855			6.18E-20	46 / 71 (64.8%)	54 / 71 (76.1%)	146	216	6640488	6640697	r1.06
	FlyBase	Dalb|		No	-	-	-	-	-	-	-	-
	FlyBase	dgri|scaffold_15110	scaffold_15110:1..24565398	Yes	3.49E-118	198 / 216 (91.7%)	209 / 216 (96.8%)	1	216	7262918	7263565	r1.05
Glossina morsitans (Tsetse fly)	FlyBase	emb|CCAG010005519.1	contig ctg10005519		4.97E-112	182 / 215 (84.7%)	203 / 215 (94.4%)	1	215	33546	32902	r1.0
Aedes aegypti	FlyBase	AAGE02022265.1	supercont1.654	No	7.92E-34	59 / 124 (47.6%)	67 / 124 (54%)	9	121	567036	566734	r1.0
Aedes aegypti	FlyBase				7.92E-34	50 / 90 (55.6%)	57 / 90 (63.3%)	123	212	566731	566480	r1.0

Supplementary material 3.3. *Nup-93* Blast search results

Query	Database	Species|Chr	Location	Genome annotated	Expect	Identities	Positives	Query beg.	Query end	Subj. beg.	Subj. end	Release
Dmel 11092- Nup93-1-PA	FlyBase	dmel|3R	3R:1..32079331	Yes	1.52E-180	210 / 471 (44.6%)	296 / 471 (62.8%)	115		14897834	14896488	r6.33
	FlyBase					85 / 233 (36.5%)	140 / 233 (60.1%)	591	823	14896395	14895727	
	FlyBase					43 / 62 (69.4%)	56 / 62 (90.3%)	54	115	14898070	14897885	
	FlyBase					30 / 54 (55.6%)	39 / 54 (72.2%)	1	54	14898287	14898126	
	FlyBase	dsim|Scf_3R	Scf_3R:1..27160941	Yes	0.00E+00	211 / 473 (44.6%)	296 / 473 (62.6%)	115	581	10486145	10487491	r2.02
	FlyBase					90 / 233 (38.6%)	140 / 233 (60.1%)	591	823	10487583	10488251	
	FlyBase					44 / 69 (63.8%)	59 / 69 (85.5%)	54	122	10485909	10486115	
	FlyBase					31 / 54 (57.4%)	40 / 54 (74.1%)	1	54	10485692	10485853	
	FlyBase	dsec|scaffold_0	scaffold_0:1..21120651	Yes	0.00E+00	211 / 471 (44.8%)	295 / 471 (62.6%)	115	581	11211650	11212996	r1.3
	FlyBase					88 / 233 (37.8%)	139 / 233 (59.7%)	591	823	11213088	11213750	
	FlyBase					43 / 62 (69.4%)	57 / 62 (91.9%)	54	115	11211414	11211599	
	FlyBase					31 / 54 (57.4%)	40 / 54 (74.1%)	1	54	11211197	11211358	
	FlyBase	Dyak|3R	3R:1..28832112	Yes	1.66E-151	299 / 734 (40.7%)	442 / 734 (60.2%)	115	823	15704264	15706369	r1.05
	FlyBase				6.29E-36	50 / 77 (64.9%)	65 / 77 (84.4%)	54	127	15704028	15704258	
	FlyBase				6.29E-36	33 / 54 (61.1%)	40 / 54 (74.1%)	1	54	15703811	15703972	
	FlyBase	dere	scaffold_4770:1..17746568	Yes	2.99E-152	210 / 471 (44.6%)	298 / 471 (63.3%)	115	581	10259660	10258314	r1.05
	FlyBase				2.99E-152	90 / 233 (38.6%)	143 / 233 (61.4%)	591	823	10258222	10257554	
	FlyBase				6.74E-35	49 / 77 (63.6%)	63 / 77 (81.8%)	54	127	10259896	10259666	
	FlyBase				6.74E-35	33 / 54 (61.1%)	40 / 54 (74.1%)	1	54	10260117	10259956	
	FlyBase	Dtak| unplaced genomic scaffold	scf7180000415397	Yes	1.95E-117	262 / 654 (40.1%)	374 / 654 (57.2%)	1	602	182033	183922	r1.0
	FlyBase		scf7180000415397		5.03E-41	88 / 231 (38.1%)	138 / 231 (59.7%)	593	823	184583	185245	r1.0
	FlyBase	Dbiarm|Unplaced genomicScaffold	scf7180000302402	Yes	0	212 / 475 (44.6%)	297 / 475 (62.5%)	111	581	3689841	3688483	r1.0
	FlyBase					90 / 235 (38.3%)	138 / 235 (58.7%)	589	823	3688396	3687722	
	FlyBase					77 / 135 (57%)	96 / 135 (71.1%)	1	115	3690292	3689888	
	FlyBase	Drhoplaloa	Contig26693	Yes	0.00E+00	380/888(43%)	546/888(61%)	1	823	31680	29116	-
	FlyBase	Delegans|unplaced genomic scaffold	scf7180000486474	Yes	0	303 / 736 (41.2%)	445 / 736 (60.5%)	115	823	1275004	1276428	r1.0
	FlyBase					76 / 135 (56.3%)	95 / 135 (70.4%)	1	115	1273859	1274263	
	FlyBase	Deugracilis| unplaced genomic scaffold	scf7180000409794	Yes	0	253 / 550 (46%)	355 / 550 (64.5%)	55	581	300487	302067	r1.0
	FlyBase					90 / 233 (38.6%)	138 / 233 (59.2%)	591	823	302159	302827	
	FlyBase					33 / 54 (61.1%)	41 / 54 (75.9%)	1	54	300266	300427	
	FlyBase	Dficusphila|Unplaced genomic scafold	scf7180000454104	Yes	0.00E+00	286 / 623 (45.9%)	398 / 623 (63.9%)	1	581	418717	418328	r1.0
	FlyBase					89 / 233 (38.2%)	140 / 233 (60.1%)	591	823	646802	647470	
	FlyBase	Dkikkawai |unplaced genomic scaffold		Yes	0.00E+00	445/820(54%)	601/820(73%)	1	819	31611	29194	
	FlyBase	Dkikkawai |unplaced genomic scaffold	scf7180000302707	Yes	3.77E-142	214 / 472 (45.3%)	295 / 472 (62.5%)	115	582	83452	84795	r1.0
	FlyBase					45 / 64 (70.3%)	57 / 64 (89.1%)	55	117	83210	83401	
	FlyBase					33 / 54 (61.1%)	40 / 54 (74.1%)	1	54	82989	83150	
	FlyBase				1.70E-41	89 / 234 (38%)	147 / 234 (62.8%)	591	823	85104	85775	
	FlyBase	dana| scaffold_12911	scaffold_12911:1..5364042	Yes	2.05E-173	302 / 763 (39.6%)	441 / 763 (57.8%)	115	823	2134355	2136532	1.05
	FlyBase		scaffold_12911:1..5364042			44 / 69 (63.8%)	58 / 69 (84.1%)	48	115	2134096	2134302	
	FlyBase		scaffold_12911:1..5364042			30 / 56 (53.6%)	39 / 56 (69.6%)	1	54	2133897	2134064	
	FlyBase	Dbipectinata|Unplaced genomic scafold	scf7180000396714	Yes	3.08E-135	291 / 793 (36.7%)	437 / 793 (55.1%)	101	823	945379	943103	r1.0
	FlyBase				4.25E-31	73 / 134 (54.5%)	94 / 134 (70.1%)	1	115	945774	945373	
	FlyBase	dpse|2	2:1..30819483	Yes	1.43899e-174	340 / 824 (41.3%)	496 / 824 (60.2%)	55	823	25899207	25896844	r3.04
	FlyBase					22 / 46 (47.8%)	28 / 46 (60.9%)	1	54	25899424	25899263	
	FlyBase	dper|scaffold_6	scaffold_6:1..6141320	Yes	8.10E-172	337 / 824 (40.9%)	492 / 824 (59.7%)	55	823	1054829	1057192	r1.3
	FlyBase					32 / 54 (59.3%)	39 / 54 (72.2%)	1	54	1054612	1054773	
	FlyBase	Dmir| chromosome 2	chromosome 2	Yes	2.88448e-107	162 / 398 (40.7%)	236 / 398 (59.3%)	255	624	5181588	5180461	r1.0
	FlyBase					24 / 46 (52.2%)	31 / 46 (67.4%)	631	823	5180444	5179890	
	FlyBase				3.05E-63	22 / 34 (64.7%)	25 / 34 (73.5%)	115	328	5182261	5181650	
	FlyBase					44 / 62 (71%)	56 / 62 (90.3%)	55	115	5182506	5182321	
	FlyBase					32 / 54 (59.3%)	37 / 54 (68.5%)	1	54	5182724	5182563	
	FlyBase	dwil|scf2_1100000004921	scf2_1100000004921:1..4707319	Yes	0	303 / 648 (46.8%)	405 / 648 (62.5%)	1	597	3630500	3632362	r1.05
	FlyBase					89 / 232 (38.4%)	136 / 232 (58.6%)	591	822	3632425	3633084	
	FlyBase	dmoj|scaffold_6540	scaffold_6540:1..34148556	Yes	1.25E-132	244 / 552 (44.2%)	336 / 552 (60.9%)	52		12963636	12965228	r1.04
	FlyBase				1.25E-132	28 / 54 (51.9%)	42 / 54 (77.8%)	1	54	12963424	12963582	
	FlyBase				2.08E-32	81 / 233 (34.8%)	136 / 233 (58.4%)	593	823	12965496	12966158	
	FlyBase	dvir|scaffold_13047	scaffold_13047:1..19223366	Yes	6.83E-178	215 / 477 (45.1%)	296 / 477 (62.1%)	115	587	16656319	16657683	r1.06
	FlyBase					78 / 225 (34.7%)	129 / 225 (57.3%)	587	805	16657733	16658383	
	FlyBase					41 / 62 (66.1%)	53 / 62 (85.5%)	55	115	16656074	16656259	
	FlyBase					34 / 57 (59.6%)	44 / 57 (77.2%)	1	57	16655856	16656023	
	FlyBase	Dalbomicans| unplaced genomic scaffold	Dalb_scaffold_53912	Yes	0.00E+00	221 / 480 (46%)	306 / 480 (63.7%)	115	589	28673	30037	r1.0
	FlyBase					93 / 233 (39.9%)	138 / 233 (59.2%)	591	823	30114	30779	
	FlyBase					39 / 64 (60.9%)	52 / 64 (81.2%)	55	117	28425	28616	
	FlyBase					30 / 54 (55.6%)	41 / 54 (75.9%)	1	54	28211	28369	
		Dalbomicans	LOC117565303	Yes	306/832(37%)	306/832(37%)	484/832(58%)	3	823	124	2493	
	FlyBase	dgri| scaffold_14906	scaffold_14906:1..14172833	Yes	3.70E-131	247 / 551 (44.8%)	333 / 551 (60.4%)	55	581	3073895	3075487	r1.05
	FlyBase					32 / 58 (55.2%)	42 / 58 (72.4%)	1	58	3073677	3073847	
	FlyBase				1.60E-39	92 / 233 (39.5%)	139 / 233 (59.7%)	593	823	3075722	3076372	
Glossina morsitans morsitans (Tsetse fly)	FlyBase	gi|594150724|emb|CCAG010023846.1	contig ctg10023846	No	0	375 / 827 (45.3%)	538 / 827 (65.1%)	1	823	28469	26043	r1.0

Supplementary material 3.4. *Nxt-1* Blast search results

Query	Database	Species|Chr	Location	Genome annotated	Gene name	Expect	Identities	Positives	Query beg.	Query end	Subj. beg.	Subj. end	Release
Dmel Nxt-1(CG12752) PA	FlyBase	Drhop|Unplaced genomicScaffold	scf7180000779514	Yes	Nxt-like-1	1.08E-45	86 / 149 (57.7%)	105 / 149 (70.5%)	4	132	98194	97748	r1.0
	FlyBase	Deug|Unplaced genomicScaffold	scf7180000408744	Yes	Nxt-like-1	2.78E-33	60 / 101 (59.4%)	72 / 101 (71.3%)	32	132	107570	107872	r1.0
							12 / 29 (41.4%)	16 / 29 (55.2%)	4	32	107427	107513	
	FlyBase	Dvir|scaffold_12928	scaffold_12928:1..7717345	Yes	Nxt-like-2	7.18E-43	59 / 100 (59%)	82 / 100 (82%)	33	132	6414545	6414844	r1.06
							21 / 29 (72.4%)	26 / 29 (89.7%)	4	32	6414384	6414470	
	FlyBase	Dgri|scaffold_15110	scaffold_15110:1..24565398	Yes	Nxt-like-3	5.14E-52	87 / 132 (65.9%)	110 / 132 (83.3%)	1	132	3806350	3805955	r1.05
Aedes aegypti	FlyBase		supercont1.654			7.92E-34	99/119(83%)	107/119(89%)	21	139	7	363	r1.0

Supplementary material 3.5. *αKap1* Blast search results

Query	Database	Species|Chr	Location	Genome annotated	Gene name by NCBI	Gene name	Expect	Identities	Positives	Query beg.	Query end	Subj. beg.	Subj. end	Mode of Duplication
Dmel_aKap1	NCBI	Drhop| XM_017132964.1	LOC108051023	Yes	Importin-α5-like	αKap7	0	355/544(65%)	424/544(77%)	1	543	202	1788	DNA Duplication
	NCBI	Dele| XM_017261023.1	LOC108138676	Yes	Importin-α5-like	αKap7	0	352/545(65%)	410/545(75%)	1	543	70	1605	DNA Duplication
	NCBI	Dwill| XM_023178550.1	LOC111519174	Yes	Importin-α5-like	αKap7	1.00E-79	111/200(56%)	153/200(76%)	344	543	3	602	Pseudogenes (Premature stop codon)
Aedes aegypti	NCBI	Aedes aegypti	LOC5574064	Yes	Importin-α6	αKap1	0	345/547(63%)	410/547(74%)	2	543	146	1699	DNA Duplication
Anopheles gambiae	FlyBase	Anopheles gambiae	2L	Yes		αKap1	6.64E-87	116 / 204 (56.9%)	138 / 204 (67.6%)					DNA Duplicate

Supplementary material 3.6. *αKap2* Blast search results

Query	Database	Species|Chr	Location	Genome annotated	Gene name by NCBI	Gene name	Expect	Identities	Positives	Query beg.	Query end	Subj. beg.	Subj. end	Mode of Duplication
Dmel_aKap2	NCBI	Deug| XM_017221465.1	LOC108111855	Yes	Importin-α-like	αKap2D	1.00E-142	218/392(56%)	288/392(73%)	1	391	646	1821	Retroduplicate
	NCBI	Dana|XM_032452203.1	LOC6504407	Yes	Importin-α	αKap2B	8e-151	247/603(41%)	360/603(59%)	1	522	311	2116	Retroduplicate
	NCBI	Dwill| XM_002074696.3	LOC6652440	Yes	Importin-α	αKap2E	0	271/471(58%)	349/471(74%)	19	486	71	1465	Retroduplicate
	NCBI	Dwill| XM_002074828.3	LOC6652143	Yes	Importin-α	αKap2F	5.00E-115	202/467(43%)	290/467(62%)	27	490	78	1370	Retroduplicate
	NCBI	Dwill| XM_015178626.2	LOC26529034	Yes	Importin-α	αKap2G	6.00E-92	174/432(40%)	262/432(60%)	56	486	43	1329	Retroduplicate
	NCBI	Dvir| XM_032433312.1	LOC6636616	Yes	Importin-α	αKap2C	0	479/524(91%)	504/524(96%)	1	522	97	1665	Partial Retroduplicate
	NCBI	Dgrim| XM_001986922.2	LOC6559418	Yes	Importin-α	αKap2A	0	282/449(63%)	343/449(76%)	74	522	274	1599	Retroduplicate
	NCBI	Dgrim| XM_001988272.2	LOC6562044	Yes	Importin-α	αKap2H	1.00E-174	275/417(66%)	328/417(78%)	57	472	24	1271	DNA Duplicate
Aedes aegypti	FlyBase	Aedes aegypti	LOC5577048	Yes	Importin-α	αKap2	0	328/519(63%)	415/519(79%)	7	522	293	1837	Retroduplicate
Anopheles gambiae	FlyBase	Anopheles gambiae	3R	Yes		αKap2	4.45E-175	297 / 487 (61%)	374 / 487 (76.8%)	24	508	43352073	43353527	Retroduplicate

Supplementary material 3.7. *αKap3* Blast search results

Query	Database	Species|Chr	Location	Genome annotated	Gene name by NCBI	Gene name	Expect	Identities	Positives	Query beg.	Query end	Subj. beg.	Subj. end	Mode of Duplication	
Dmel_aKap3	NCBI	Drhop| XM_017125897.1	LOC108046290	Yes	Importin-α4	αKap4	9.00E-53	101/210(48%)	135/210(64%)	274	482	689	1318	Partial retroduplication	
							2.00E-33	95/323(29%)	162/323(50%)	10	329	203	1102		
	NCBI	Dfic| XM_017193168.1	LOC108093205	Yes	Importin-α4	αKap4	2.00E-75	169/494(34%)	267/494(54%)	7	500	68	1477	Partial retroduplication	
	NCBI	Dmel| NM_139750.2		Yes	Importin-α4	αKap4	4.00E-82	159/390(41%)	238/390(61%)	92	480	203	1366	Partial retroduplication	(Phadnis, et al. 2012)
	NCBI	Dsim| JQ173086.1		Yes	Importin-α4	αKap4	4.00E-75	157/393(40%)	231/393(58%)	92	483	133	1305	Partial retroduplication	(Phadnis, et al. 2012)
	NCBI	Dsec| 3R	LOC6610947	Yes	Importin-α4	αKap4	2e-73	157/393(40%)	230/393(58%)	92	483	286	1458	Partial retroduplication	(Phadnis, et al. 2012)
	NCBI	Dyak| 3R	GE20518	Yes	Importin-α4	αKap4	3.00E-86	157/391(40%)	236/391(60%)	92	482	133	1299	Partial retroduplication	(Phadnis, et al. 2012)
	NCBI	Dere| JQ173093.1	LOC6545099	Yes	Importin-α4	αKap4	6.00E-88	163/409(40%)	250/409(61%)					Partial retroduplication	(Phadnis, et al. 2012)

Query	Database	Species|Chr	Location	Genome annotated	Gene name by NCBI	Gene name	Expect	Identities	Positives	Query beg.	Query end	Subj. beg.	Subj. end	Mode of Duplication
Dmel_aKap3	NCBI	Dtak| XM_017140526.1	LOC108056655	Yes	Importin-α3-like	αKap6	1e-145	224/405(55%)	285/405(70%)	78	482	135	1346	Partial retroduplication
	NCBI	Dbiarm| XM_017109587.1	LOC108034648	Yes	Importin-α3-like	αKap6	4.00E-142	232/431(54%)	303/431(70%)	54	483	109	1395	Partial retroduplication
	NCBI	Drhop| XM_017134021.1	LOC108051802	Yes	Importin-α3-like	αKap6	7.00E-162	271/479(57%)	326/479(68%)	7	483	95	1492	Partial retroduplication
	NCBI	Dele| XM_017278131.1	LOC108150161	Yes	Importin-α3-like	αKap6	6.00E-153	227/406(56%)	295/406(72%)	78	483	135	1346	Partial retroduplication
	NCBI	Deug| XM_017224687.1	LOC108113971	Yes	Importin-α3-like	αKap6	3.00E-111	189/404(47%)	255/404(63%)	80	483	219	1358	Partial retroduplication
	NCBI	Dfic| XM_017188037.1	LOC108089690	Yes	Importin-α3-like	αKap6	5.00E-134	215/405(53%)	276/405(68%)	81	485	274	1476	Partial retroduplication
Aedes aegypti	FlyBase	Aedes aegypti	LOC5575103	Yes	Importin-α3	αKap3	0	383/517(74%)	450/517(87%)	1	514	378	1925	Partial retroduplication
Anopheles gambiae	FlyBase	Anopheles gambiae	2R	Yes		αKap3	0	363 / 523 (69.4%)	404 / 523 (77.2%)	21	483	2177805	2179370	Partial retroduplication

Supplementary material 3.8. *αKap5* Blast search results

Query	Database	Species|Chr	Location	Genome annotated	Gene name	Expect	Identities	Positives	Query beg.	Query end	Subj. beg.	Subj. end	Mode of Duplication	
Deug- αKap5	NCBI	Dtak| XM_017148512.1	LOC108062019	Yes	αKap5	0.00E+00	280/443(63%)	350/443(79%)	1	443	70	1395	Retroduplicate	
	NCBI	Dbiarm| XM_017100945.1	LOC108028926	Yes	αKap5	1.00E-173	260/432(60%)	324/432(75%)	1	432	111	1406	Retroduplicate	
	NCBI	Dmir| XM_033392804.1	LOC108163341	Yes	αKap5	7.00E-96	172/446(39%)	269/446(60%)	1	443	236	1570	Retroduplicate	
	NCBI	Dalb| 2L	LOC117565775	Yes	αKap5	8.00E-97	190/423(45%)	267/423(63%)	12	431	90	1349	Retroduplicate	
	NCBI	Dana| XM_001962595.3	LOC6497170	Yes	αKap5	5e-160	247/443(56%)	321/443(72%)	2	443	273	1598	Retroduplicate	(Phadnis, et al. 2012)
	NCBI	Dbipec| XM_017243369.1	LOC108126697	Yes	αKap5	1.00E-158	241/443(54%)	318/443(71%)	2	443	146	1471	Retroduplicate	(Phadnis, et al. 2012)
	NCBI	Dpseudo| XM_015179925.1		Yes	αKap5	1.00E-104	182/446(41%)	278/446(62%)	1	443	140	1474	Retroduplicate	(Phadnis, et al. 2012)
	NCBI	Dper| XM_002018827.2	LOC6593409	Yes	αKap5	5.00E-103	180/446(40%)	277/446(62%)	1	443	198	1532	Retroduplicate	(Phadnis, et al. 2012)
	NCBI	Dwill| XM_002066465.3	LOC6643478	Yes	αKap5	2.00E-116	197/436(45%)	282/436(64%)	9	442	31	1323	Retroduplicate	(Phadnis, et al. 2012)
	NCBI	Dmoj| XM_032729177.1	LOC6576335	Yes	αKap5	2e-116	198/435(46%)	280/435(64%)	1	431	113	1411	Pseudogene	(Phadnis, et al. 2012)
	NCBI	Dvir| XM_002051452.3	LOC6628493	Yes	αKap5	2.00E-108	199/435(46%)	283/435(65%)	1	431	198	1496	Retroduplicate	(Phadnis, et al. 2012)

Query	Database	Species|Chr	Location	Genome annotated	Gene name	Expect	Identities	Positives	Query beg.	Query end	Subj. beg.	Subj. end	Mode of Duplication	
Deug- αKap5	NCBI	Dpseudo| XM_015179925.1	LOC6903455	Yes	αKap5B	2.00E-103	180/446(40%)	278/446(62%)	1	443	168	1502	Tandem-Retroduplicate	(Phadnis, et al. 2012)

Supplementary material 3.9. *e(y)2* Blast search results

Query	Database	Species|Chr	Location	Genome annotated	Gene name	Expect	Identities	Positives	Query beg.	Query end	Subj. beg.	Subj. end
Dmel e(y)2 (CG15191)- PA	NCBI	Dmel	3R	Yes	e(y)2b	1.00E-17	32/77(42%)	55/77(71%)	11	87	138	365
	NCBI	Dsim	LOC6727039	Yes		3.00E-15	33/77(43%)	53/77(68%)	11	87	170	397
	NCBI	Dsec	LOC6614196	Yes		1.00E-16	32/77(42%)	53/77(68%)	11	87	248	475
	NCBI	Dyak	LOC6536098	Yes		3.00E-18	37/81(46%)	51/81(62%)	7	87	161	400
	NCBI	Dere	LOC6553056	Yes		7.00E-18	38/81(47%)	51/81(62%)	7	87	211	450
	NCBI	Dtak	LOC108063276	Yes		8.00E-18	35/81(43%)	54/81(66%)	7	87	167	406
	NCBI	Dbiarm	LOC108027183	Yes		1.00E-18	36/80(45%)	53/80(66%)	8	87	156	392
	NCBI	Drhoplaloa	LOC108041220	Yes		1.00E-18	37/86(43%)	54/86(62%)	2	87	167	421
	NCBI	Delegans	LOC108141035	Yes		5.00E-16	32/86(37%)	52/86(60%)	2	87	120	374
	NCBI	Deugracilis	LOC108116202	Yes		4.00E-17	34/73(47%)	48/73(65%)	15	87	180	395
	NCBI	Dficusphila	LOC108098439	Yes		3.00E-16	33/86(38%)	54/86(62%)	2	87	122	376
	NCBI	Dkikkawai	LOC108081021	Yes		2.00E-13	28/72(39%)	45/72(62%)	16	87	205	417
	NCBI	Dana		Yes								
	NCBI	Dbipectinata		Yes								
	NCBI	Dpse	LOC4802914	Yes		3.00E-14	32/69(46%)	47/69(68%)	20	87	209	415
	NCBI	Dper	LOC6594630	Yes		3.00E-13	30/69(43%)	47/69(68%)	20	87	114	320
	NCBI	Dmir	LOC108155083	Yes		4.00E-13	30/69(43%)	46/69(66%)	20	87	228	434
	NCBI	Dwil	LOC6647810	Yes		1.00E-14	33/81(41%)	50/81(61%)	7	87	205	444
	NCBI	Dmoj	LOC6573376	Yes		7.00E-13	27/74(36%)	48/74(64%)	14	87	161	379
	NCBI	Dvir	LOC6636707	Yes		1.00E-13	30/58(52%)	41/58(70%)	30	87	189	359
	NCBI	Dalbomicans	LOC117575499	Yes		8.00E-16	35/78(45%)	50/78(64%)	10	87	186	416
	NCBI	Dgri	LOC6563983	Yes		8.00E-17	35/84(42%)	53/84(63%)	4	87	65	313

No duplication outside Drosophila is found

Supplementary material 3.10. *Tnpo* Blast search results

Query	Database	Species|Chr	Location	Genome annotated	Gene name	Expect	Identities	Positives	Query beg.	Query end	Subj. beg.	Subj. end	Mode of Duplication
Dmel Tnpo (CG7398)- PA	NCBI	Dmel	3L	Yes	CG8219	0.0	611/893(68%)	708/893(79%)	1	893	68	2620	DNA Duplication
	NCBI	Dsim	LOC6737001	Yes		0.0	626/893(70%)	722/893(80%)	1	893	118	2733	DNA Duplication
	NCBI	Dsec	-	Yes		-	-	-	-	-	-	-	
	NCBI	Dyak	LOC6533186	Yes		0.0	685/893(77%)	756/893(84%)	1	893	770	3400	DNA Duplication
	NCBI	Dere	LOC6545211	Yes		0.0	696/893(78%)	768/893(86%)	1	893	334	2964	DNA Duplication
	NCBI	Dtak	-	Yes		-	-	-	-	-	-	-	DNA Duplication
	NCBI	Dbiarm	LOC108028277	Yes		0.0	671/893(75%)	764/893(85%)	1	893	87	2720	DNA Duplication
	NCBI	Drhoplaloa	-	Yes		-	-	-	-	-	-	-	DNA Duplication
	NCBI	Dele	LOC108142179	Yes		0.0	369/448(82%)	397/448(88%)	1	448	160	1488	DNA Duplication
						9e-165	232/266(87%)	246/266(92%)	628	893	1483	2280	
	NCBI	Deug	LOC108107533	Yes		0.0	452/545(83%)	501/545(91%)	1	545	34	1659	DNA Duplication

No duplication outside Drosophila is found

Supplementary material 4.1. Duplication summary of *Ntf-2*

	Ntf-2 (CG1740)	Ntf-2r(1) (CG10174)	Ntf-2r(2)	Ntf-2r(3)	Ntf-2r(4)	Ntf-2r(5)	Ntf-2r(6)
D. melanogaster	Present	+ (87.69%)	Absent	-	-	-	-
D. simulans	Present	+ (88.4%)	Absent	-	-	-	-
D. sechelia	Present	+ (86.15%)	Absent	-	-	-	-
D. yakuba	Present	-	Absent	-	-	-	-
D. erecta	Present	-	Absent	-	-	-	-
D. takahashii	Present	-	Absent	-	-	-	-
D. biarmipes	Present	-	+ (52.34%)	-	-	-	-
D. rhopaloa	Present	-	+ (54.61%)	-	-	-	-
D. elegans	Present	-	Absent	-	-	-	-
D. eugracilis	Present	-	Absent	-	-	-	-
D. ficusphila	Present	-	Absent	+ (97.69%)	-	-	-
D. kikkawai	Present	-	Absent	-	-	-	-
D. ananassae	Present	-	+ (50.39%)	-	+ (59.09%)	-	-
D. bipectinata	Present	-	+ (44.09%)	-	-	-	-
D. pseudoobscura	Present	-	-	-	-	+ (58.46%)	-
D. persimilis	Present	-	-	-	-	+ (60%)	-
D. miranda	Present	-	-	-	-	+ (60.76%)	-
D. willistoni	Present	-	-	-	-	-	-
D. mojavensis	Present	-	-	-	-	-	-
D. virilis	Present	-	-	-	-	-	-
D. albomicans	Present	-	-	-	-	-	-
D. grimshawi	Present	-	-	-	-	-	+ (87.69%)

Absent means not found in the current assembly
Percentages indicate protein identity to the parental gene
Duplications shown in blue were identified in previous studiess
Orange colour indicates DNA-mediated duplications.

Supplementary material 4.2. Duplication summary of *Ran*

	Ran (CG1404)	Ran-like(1) (CG7815)	Ran-like(2)	Ran-like(3)	Ran-like(4)	Ran-like(5)	Ran-like(6)	Ran-like(7)	Ran-like(8)	Ran-like(9)	Ran-like(10)	Ran-like(11)	Ran-like(12)	Ran-like(13)	Pseudogenes	
D. melanogaster	Present	-	+ (57.40%)	X	-	-	-	-	-	-	-	-	-	-		
D. simulans	Present	-	+ (54.62%)	+ (41.90%)	-	-	-	-	-	-	-	-	-	-		
D. sechelia	Present	-	+ (54.16%)	+ (41.42%)	-	-	-	-	-	-	-	-	-	-		
D. yakuba	Present	-	ψ(47.43%)	ψ(43.37%)	-	-	-	-	-	-	-	-	-	-		
D. erecta	Present	-	+(58.79%)	+(45.36%)	-	-	-	-	-	-	-	-	-	-		
D. takahashii	Present	-	-	ψ (44.1%)	+ (63.88%)	-	-	-	-	-	-	-	-	-	ψ(86.85%)	
D. biarmipes	Present	-	-	-	-	+ (93.51%)	-	-	-	-	-	-	-	-	ψ(93.05%)	
D. rhopaloa	Present	-	-	+ (50.23%)	-	-	+ (98.4%)	+ (62.44%)	-	-	-	-	-	-		
D. elegans	Present	-	-	ψ(54.22%)	-	-	-	-	-	-	-	-	-	-		
D. eugracilis	Present	-	-	ψ(49.55%)	-	-	-	-	-	-	-	-	-	-	ψ (46.33%)	
D. ficusphila	Present	-	-	Absent	-	-	-	-	+ (97.68%)	+ (96.29%)	-	-	-	-		
D. kikkawai	Present	-	-	-	-	-	-	-	-	-	-	-	-	-	ψ (35.61%)	ψ (32.64%)
D. ananassae	Present	-	-	-	-	-	-	-	-	-	+ (81.48%)	-	-	-		
D. bipectinata	Present	-	-	-	-	-	-	-	-	-	ψ (72.37%)	+ (99.07%)	-	-		
D. pseudoobscura	Present	+ (41.96%)	-	-	-	-	-	-	-	-	-	-	-	-		
D. persimilis	Present	+ (41.51%)	-	-	-	-	-	-	-	-	-	-	-	-		
D. miranda	Present	+ (43.24%)	-	-	-	-	-	-	-	-	-	-	-	-		
D. willistoni	Present	-	-	-	-	-	-	-	-	-	-	-	-	-	ψ (50.54%)	ψ(78.42%)
D. mojavensis	Present	-	-	-	-	-	-	-	-	-	-	-	+ (67.43%)	-		
D. virilis	Present	-	-	-	-	-	-	-	-	-	-	-	+ (70.83%)	-		
D. albomicans	Present	-	-	-	-	-	-	-	-	-	-	-	-	-		
D. grimshawi	Present	-	-	-	-	-	-	-	-	-	-	-	-	+ (93.89%)		

Absent means not found in the current assembly
Percentages indicate protein identity to the parental gene
Duplications shown in blue were identified in previous studiess
Orange colour indicates DNA-mediated duplications.

Supplementary material 4.3. Duplication summary of *Nup93*

	Nup93 (CG11092)	Nup93-like(1) (CG7262)	Nup93-like (2)	Nup93-like (3)
D. melanogaster	Present	+ (43.75%)		
D. simulans	Present	+ (43.20%)		
D. sechelia	Present	+ (43.08%)		
D. yakuba	Present	+ (45.84%)		
D. erecta	Present	+ (46.32%)		
D. takahashii	Present	+ (44.93%)		
D. biarmipes	Present	+ (45.41%)		
D. rhopaloa	Present	+ (45.26%)		
D. elegans	Present	+ (44.092%)		
D. eugracilis	Present	+ (42.71%)		
D. ficusphila	Present	+ (46.81%)		
D. kikkawai	Present	+ (45.34)	+ (53.66%)	
D. ananassae	Present	+(43.76%)		
D. bipectinata	Present	+ (45.25%)		
D. pseudoobscura	Present	+ (44.66%)		
D. persimilis	Present	+ (42.41%)		
D. miranda	Present	+ (42.58%)		
D. willistoni	Present	+ (45.26%)		
D. mojavensis	Present	+ (41.12%)		
D. virilis	Present	+ (43.48%)		
D. albomicans	Present	+ (47.12%)		+ (36.54%)
D. grimshawi	Present	+ (42.79%)		

Percentages indicate protein identity to the parental gene
Orange colour indicates DNA-mediated duplications.

Supplementary material 4.4. Duplication summary of *Nxt-1*

	Nxt1 (CG12752)	Nxt1-like(1)	Nxt1-like(2)	Nxt1-like(3)
D. melanogaster	Present	-	-	-
D. simulans	Present	-	-	-
D. sechelia	Present	-	-	-
D. yakuba	Present	-	-	-
D. erecta	Present	-	-	-
D. takahashii	Present	-	-	-
D. biarmipes	Present	-	-	-
D. rhopaloa	Present	+ (72.72%)	-	-
D. elegans	Present	-	-	-
D. eugracilis	Present	+ (56.81%)	-	-
D. ficusphila	Present	-	-	-
D. kikkawai	Present	-	-	-
D. ananassae	Present	-	-	-
D. bipectinata	Present	-	-	-
D. pseudoobscura	Present	-	-	-
D. persimilis	Present	-	-	-
D. miranda	Present	-	-	-
D. willistoni	Present	-	-	-
D. mojavensis	Present	-	-	-
D. virilis	Present	-	+ (79.85%)	-
D. albomicans	Present	-	-	-
D. grimshawi	Present	-	-	+ (74.62%)

Percentages indicate protein identity to the parental gene
Orange colour indicates DNA-mediated duplications.

Supplementary material 4.5. Duplication summary of *αKaps*

	αKap1 (CG8548)	αKap2 (CG4799)	αKap3 (CG9423)	αKap4 (CG10478)	αKap5	αKap5B	αKap2A	αKap2B	αKap2C	αKap2D	αKap2E	αKap2F	αKap2G	αKap2H	αKap6	αKap7
D. melanogaster	Present	Present	Present	+ (33.95%)												
D. simulans	Present	Present	Present	+ (34.5%)												
D. sechelia	Present	Present	Present	+ (34.93%)												
D. yakuba	Present	Present	Present	+ (34.72%)												
D. erecta	Present	Present	Present	+ (34.5%)												
D. takahashii	Present	Present	Present		+ (63.20%)										+ (53.98%)	
D. biarmipes	Present	Present	Present		+ (59.45%)										+ (53.31%)	
D. rhopaloa	Present	Present	Present	+ (34.15%)											+ (55.57%)	+ (64%)
D. elegans	Present	Present	Present												+ (49.06%)	+ (63.03%)
D. eugracilis	Present	Present	Present							+(54.50%) Missing IBB					+ (41.7%)	
D. ficusphila	Present	Present	Present	+ (34.14%)											+ (45.45%)	
D. kikkawai	Present	Present	Present													
D. ananassae	Present	Present	Present		+ (55.63%)			+ (41%)(IBB-Nter)								
D. bipectinata	Present	Present	Present		+ (54.27%)											
D. pseudoobscura	Present	Present	Present		+ (39.59%)	+ (39.37%)										
D. persimilis	Present	Present	Present		+ (39.37%)											
D. miranda	Present	Present	Present		+ (38.25%)											
D. willistoni	Present	Present	Present		+ (43.91%)						+(53.07%) Missing IBB	+(40.08%) Missing IBB	+(36.97%) Missing IBB			ψ (56%)
D. mojavensis	Present	Present	Present		ψ											
D. virilis	Present	Present	Present		+ (44.07%)				+ (91%) (IBB-Nter)							
D. albomicans	Present	Present	Present		+ (42.05%)											
D. grimshawi	Present	Present	Present				+ (63%)(missing IBB)							+(64.66%) Missing IBB		

missing IBB missing IBB missing IBB missing IBB IBB

Percentages indicate protein identity to the parental gene
Duplications shown in blue were identified in previous studiess
Orange colour indicates DNA-mediated duplications.

Supplementary material 4.6. Duplication summary of *e(y)2*

	e(y)2 (CG15191)	e(y)2b (CG14612)
D. melanogaster	Present	34.44%
D. simulans	Present	34.04%
D. sechelia	Present	34.04%
D. yakuba	Present	36.08%
D. erecta	Present	38.54%
D. takahashii	Present	35.05%
D. biarmipes	Present	36.08%
D. rhopaloa	Present	37.10%
D. elegans	Present	32.99%
D. eugracilis	Present	32.99%
D. ficusphila	Present	32.99%
D. kikkawai	Present	29.59%
D. ananassae	Present	31.95%
D. bipectinata	Present	29.89%
D. pseudoobscura	Present	34.88%
D. persimilis	Present	31.95%
D. miranda	Present	34.88%
D. willistoni	Present	32.38%
D. mojavensis	Present	30.52%
D. virilis	Present	41.05%
D. albomicans	Present	33.67%
D. grimshawi	Present	32.63%

Percentages indicate protein identity to the parental gene

Supplementary material 4.7. Duplication summary of *Tnpo*

	Tnpo (CG7398)	CG8219
D. melanogaster	Present	68.75%
D. simulans	Present	69.98%
D. sechelia	Present	-
D. yakuba	Present	77.04%
D. erecta	Present	78.72%
D. takahashii	Present	-
D. biarmipes	Present	76.14%
D. rhopaloa	Present	-
D. elegans	Present	74.43%
D. eugracilis	Present	80.69%
D. ficusphila	Present	-
D. kikkawai	Present	-
D. ananassae	Present	-
D. bipectinata	Present	-
D. pseudoobscura	Present	-
D. persimilis	Present	-
D. miranda	Present	-
D. willistoni	Present	-
D. mojavensis	Present	-
D. virilis	Present	-
D. albomicans	Present	-
D. grimshawi	Present	-

Percentages indicate protein identity to the parental gene
Orange colour indicates DNA-mediated duplications.

Supplementary material 4.8. Duplication summary of *Importin-β*

	Importin-β	Apollo (CG32165)	Artemis (CG32164)	Apl-like (2)	Apl-like (3)	Apl-like (4)		
D. melanogaster	Present	Present	Present				Tandem duplication	(VanKuren and Long 2018)
D. simulans	Present							
D. sechelia	Present							
D. yakuba	Present							
D. erecta	Present							
D. takahashii	Present							
D. biarmipes	Present							
D. rhopaloa	Present							
D. elegans	Present							
D. eugracilis	Present							
D. ficusphila	Present							
D. kikkawai	Present							
D. ananassae	Present							
D. bipectinata	Present							
D. pseudoobscura	Present			92.22%	56.85%		Retrotransposition	(VanKuren and Long 2018)
D. persimilis	Present			93.33%			Retrotransposition	(VanKuren and Long 2018)
D. miranda	Present							
D. willistoni	Present					40.90%	Retrotransposition	(VanKuren and Long 2018)
D. mojavensis	Present							
D. virilis	Present							
D. albomicans	Present							
D. grimshawi	Present							

Percentages indicate protein identity to the parental gene
Duplications shown in blue were identified in previous studiess
Orange colour indicates DNA-mediated duplications.

Supplementary material 4.9. Duplication summary of nuclear transport genes in non-*Drosophila Dipterans*

	Ntf-2	Ran	Importin-α	Nup93	Nxt-1	e(y)2b	Tnpo
Glossina morsitans morsitans (Tsetse fly)	ψ	85%	-	45%	-	-	-
Aedes aegypti (Yellow fever mosquito)	-	ψ	-	-	83%	-	-
Anopheles gambiae (Malaria mosquito)	94%	-	-	-	-	-	-

Percentages indicate protein identity to the parental gene

Supplementary material 5.1. Branch model, One ratio analysis

RL1	Branch model- One Ratio	omega (dN/dS)	0.47031
		lnL	-1011.5397
		np	5
	omega =1	omega (dN/dS)	1
		lnL	-1013.01502
		np	4
		Pvalue:	0.085843558

Ran1	Branch model- One Ratio	omega (dN/dS)	0.0001
		lnL	-947.167811
		np	5
	omega =1	omega (dN/dS)	1
		lnL	-973.883783
		np	4
		Pvalue:	0

RL3	Branch model- One Ratio	omega (dN/dS)	0.5177
		lnL	-1133.362837
		np	5
	omega =1	omega (dN/dS)	1
		lnL	-1137.044282
		np	4
		Pvalue:	0.006658369

Ran3	Branch model- One Ratio	omega (dN/dS)	0.0147
		lnL	-927.840111
		np	5
	omega =1	omega (dN/dS)	1
		lnL	-962.35491
		np	4
		Pvalue:	0

Ntf-2r(5)	Branch model- One Ratio	omega (dN/dS)	999
		lnL	-662.432263
		np	5
	omega =1	omega (dN/dS)	1
		lnL	-670.993459
		np	4
		Pvalue:	0.000035047

Ntf-2(5)	Branch model- One Ratio	omega (dN/dS)	0.0345
		lnL	-583.010769
		np	5
	omega =1	omega (dN/dS)	1
		lnL	-592.924887
		np	4
		Pvalue:	0.000008472

Nup93_like	Branch model- One Ratio	omega (dN/dS)	0.0827
		lnL	-23260.79888
		np	43
	omega =1	omega (dN/dS)	1
		lnL	-26126.57031
		np	42
		Pvalue:	0

Nup93	Branch model- One Ratio	omega (dN/dS)	0.1281
		lnL	-24877.57724
		np	43
	omega =1	omega (dN/dS)	1
		lnL	-26937.29375
		np	42
		Pvalue:	0

Impa6	**Branch model- One Ratio**	**omega (dN/dS)**	0.354
		lnL	-6524.455684
		np	11
	omega =1	omega (dN/dS)	1
		lnL	-6641.052351
		np	10
		Pvalue:	0

Impa3	Branch model- One Ratio	omega (dN/dS)	0.0224
		lnL	-3691.120428
		np	11
	omega =1	omega (dN/dS)	1
		lnL	-4125.035472
		np	10
		Pvalue:	0

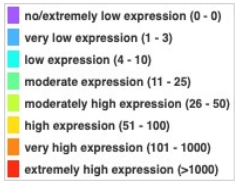
Supplementary material 5.2. Branch model, two ratio analysis

Duplicate	1 rate	Rate	2 rates	Parent	New gene	2deltaL~X2	d.f.	*P*
Nup93-2	-30557.5186	0.096	-30482.3273	0.1338	0.0622	150.382602	1	0
Ntf2r-2	-2202.05772	0.0685	-2163.22154	0.0061	0.2761	77.67235	1	0
Ntf2r-5	-1045.75973	0.967	-1026.12376	0.0306	999	39.271924	1	0
Ran-like1	-1824.42476	0.158	-1812.41869	0.0001	0.4244	24.012148	1	0
Ran-like3	-1860.41979	0.2738	-1841.84081	0.0089	0.4993	37.15797	1	0
Ran-like13	-1830.5318	0.0506	-1822.94361	0.005	0.0782	15.176386	1	0.0001
Impa6	-10555.8772	0.1963	-10417.8442	0.0242	0.3418	276.065912	1	0

Gene with signifacntly higher rate shown in red

Supplementary material 5.3. modENCODE tissue expression data

modENCODE Tissue Expression Data	RPKM Expression Value (*D. melanogaster*)														
Tissues	**Ntf-2**	**Ntf-2r**	**Ran**	**Ran-like**	**Nup-93**	**Nup93-like**	**Nxt1**	**αKap1**	**αKap2**	**αKap3**	**αKap4**	**e(y)2**	**e(y)2b**	**Tnpo**	**CG8219**
imaginal disc, larvae L3 wandering	32	69	453	20	33	21	31	45	549	115	8	30	18	23	32
central nervous system, larvae L3	43	1	640	0	31	15	34	11	115	90	0	40	1	34	0
central nervous system, pupae P8	18	1	188	0	7	4	11	10	10	51	0	16	1	23	0
head, virgin 1-day female	22	1	156	0	3	2	11	9	1	60	0	10	1	16	0
head, virgin 4-day female	18	0	144	0	4	3	7	9	2	67	0	12	0	14	0
head, virgin 20-day female	14	0	86	0	5	5	8	15	0	68	0	12	0	12	0
head, mated 1-day female	19	1	149	0	2	2	5	6	4	52	0	7	1	11	0
head, mated 4-day female	14	0	155	0	3	3	6	11	0	48	0	8	0	11	0
head, mated 20-day female	11	0	82	0	3	3	7	11	2	53	0	10	0	11	0
head, mated 1-day male	19	0	112	0	3	3	10	10	2	67	0	13	1	17	0
head, mated 4-day male	19	0	141	0	5	4	9	15	1	58	0	9	0	18	0
head, mated 20-day male	23	0	142	0	8	7	14	19	3	63	0	17	1	20	0
salivary gland, larvae L3 wandering	19	0	87	0	1	1	5	2	3	44	0	14	0	11	0
salivary gland, white prepupae	15	3	97	2	3	2	4	7	18	143	0	22	3	22	2
digestive system, larvae L3 wandering	27	1	197	1	10	5	19	9	17	109	0	22	0	18	0
digestive system, 1-day adult	14	1	165	1	5	3	14	8	19	59	0	10	0	8	0
digestive system, 4-day adult	18	1	234	0	7	4	12	10	31	61	0	10	0	13	0
digestive system, 20-day adult	15	0	156	1	8	5	12	7	20	58	0	11	0	10	0
fat body, larvae L3 wandering	14	1	107	0	2	2	3	3	5	37	0	6	0	4	0
fat body, white prepupae	17	17	155	6	6	4	7	10	97	79	2	13	5	12	6
fat body, pupae P8	30	8	256	8	6	8	3	6	75	44	4	26	5	9	1
carcass, larvae L3 wandering	32	1	344	0	25	12	26	12	92	96	0	28	0	29	0
carcass, 1-day adult	17	0	134	0	4	3	7	7	12	52	0	8	0	12	0
carcass, 4-day adult	17	1	207	0	6	4	7	10	30	63	0	11	0	12	0
carcass, 20-day adult	16	1	142	0	7	5	12	11	25	69	0	12	0	11	0
ovary, virgin 4-day female	59	0	985	0	85	25	88	41	572	217	0	33	0	94	0
ovary, mated 4-day female	74	1	1172	0	63	19	74	30	489	180	0	34	0	74	0
testis, mated 4-day male	11	101	419	49	11	29	5	50	1067	93	31	33	47	5	52
accessory gland, mated 4-day male	15	29	313	12	7	6	9	12	261	65	6	30	8	8	1



Supplementary material 5.4. total number of duplications per gens and per species

Species	Total	*Ntf-2*	*Ran*	*Importin-α*	*Nup93*	*Nxt-1*	*e(y)2*	*Tnpo*	*Importin-β*
D. melanogaster	7	1	1	1	1	0	1	1	1
D. simulans	7	1	2	1	1	0	1	1	0
D. sechellia	6	1	2	1	1	0	1	0	0
D. yakuba	4	0	0	1	1	0	1	1	0
D. erecta	6	0	2	1	1	0	1	1	0
D. takahashi	5	0	1	2	1	0	1	0	0
D. biarmipes	7	1	1	2	1	0	1	1	0
D. rhopaloa	10	1	3	3	1	1	1	0	0
D. elegans	5	0	0	2	1	0	1	1	0
D. eugracilis	5	0	0	2	1	0	1	1	0
D. ficusphila	6	1	2	1	1	0	1	0	0
D. kikkawai	3	0	0	0	2	0	1	0	0
D. ananassae	6	2	1	2	1	0	1	0	0
D. bipectinata	3	1	1	0	1	0	1	0	0
D. pseudoobscura	7	1	1	2	1	0	1	0	1
D. persimilis	6	1	1	1	1	0	1	0	1
D. miranda	5	1	1	1	1	0	1	0	0
D. willistoni	7	0	0	4	1	0	1	0	1
D. mojavensis	3	0	1	0	1	0	1	0	0
D. virilis	6	0	1	2	1	1	1	0	0
D. albomicans	5	0	0	1	2	1	1	0	0
D. grimshawi	6	1	1	2	1	0	1	0	0
Glossina morsitans morsitans	2	0	1	0	1	0	0	0	0
Aedes aegypti	1	0	0	0	0	1	0	0	0
Anopheles gambiae	1	1	0	0	0	0	0	0	0

Supplementary material 5.5. Average number of young genes duplicates

Ntf-2	*Ran*	*Importin-α*	*Nup93*	*Nxt-1*	*e(y)2*	*Tnpo*
5/22=23% vs. 1/29=3%	7/22=31% vs. 1/29=3%	1/22=4% vs. 0	0	0 vs. 1/29=3%	0	1/22=4% vs. 0

Chapter 3

Supplementary Materials

Supplementary Table 1. Non-bar/Bar ratio of males from Act-5C-GAL4/UASt-Ran-EGFP X female Ran^{-}/FM7c shows complete rescue (Non-Bar/Bar > 25%) of the *Ran* mutant lethality phenotype.

Replicate			non-Bar/Bar
1	non-Bar ♂	16	533.33%
	Bar ♂	3	
2	non-Bar ♂	12	300%
	Bar ♂	4	
3	non-Bar ♂	8	88.88%
	Bar ♂	9	

Supplementary Table 2. Oligos and gene blocks (5'-3') designed to produce the *Dntf-2r* knockout.

Dntf-2r guide 1 sense	CTTCGATTAAATCGGTTAAGCTGAC
Dntf-2r guide 1 antisense	AAACGTCAGCTTAACCGATTTAATC
Dntf-2r homologous arm 1 *Eco*RI	GGGTGTCGCCCTTCGCTGAAGCAGGTGGAATcgattgcgagctggtggc cgtcctgtgtgagcggatctcatcgagcacctcctccagatcgcgtctggagatgcgcccaatg cgacgaacatcctggtccagccggcgaatctgtttatccaaactcaactcagctttggccgccg cattcgtggcaaaaccactaggaagataattaaaaataaacttaattaatatataaagatatat gaattaacaaaatcctaacgtccttatagattattttccggctgaccttaatgagaaactgagttgt tataagataatcgtgttgctttttgaataaaaataagatatttagcttctttctacatatataactaga ctttctacaagaaaaacgctcgactaaggccaaaattggatgagcacatagtaggcgctcca aatccggatgaggttatgtaaatatgtgagacatcagtcgaatgcaactgcgatgatcccaagt ggcggcacgtgcgcctcttttcgggattcaatcgctggaaagtggacgattccccactaatcac gtaaatacatgagcacattgcgtaatgccggcgatttggggccagattttactcacttttggaact ggccgcacatgcagggcgcattgttcagcagattgctcgtctcgttctcgcgaaccgagttcac cagcacattgcgtgtgaagccggcgaagtagcgcagcagtttgcccgtcctcaggatggatg ccatggcggaatcgttcggttttaattaataaatgagcttattattaaaatgacgttcgcacgagtg cggctgaaaccgatgtctcttcgcgcctatcgatgtttacctgcgacaatcgattgttttcggtggg accaaaacgttcaaatcaaatgtatcagcttagcggtgacacaaacgagaggatatttgaac agtgcctaagcctgtcAATTCTTGCATGCTAGCGGCCGCGGACATAT
Dntf-2r guide 2 sense	CTTCGTGTGGTGTACATATGGGTC
Dntf-2r guide 2 antisense	AAACGACCCATATGTACACCACAC
Dntf-2r homologous arm 2 *Xho*I	TGCATAAGGCGCGCCTAGGCCTTCTGCAGCccatatgtacaccacacat aatcgacatccaaagacgcccagcgcctaatagtgataacatgataacaacagcacggca gtgggactcagaaaaaagcaaaacaaagccagccaacggtctcaagatcgtcagcgaata caaaagttaatacaagacaaacaaaaaaataattaaaaaaagcttaggacttttattttgaaa acttttcgagaaataaaatagaaacaaatcttttcccaattcacaaataatgggtcacattgga

	attttgaagcaatagtctttaatatttgtagtcttcatacgatgcatgttttattctagcttaatttttcgga gctgtctttgtgtttgcaatgccaaaagtgcgctgtaaatgcaccgcaaagaaagagaacaat ccacctgactaagagaaactgcatgctccctattcatttgctgctattgcggagcgtcatttctgca gcagacgccgccgttgacgtcggcgtcgacgtagctttcttttgtggcgaatttggaactggaaa ttgcatttcatttgcaaacgtatcgttgcacacggcggttcgcgaatcgcgcttcctcttcccgctg cagctgtgtgcgtgttgtggtgtactcgtattggcagcgcgtatgtgtgtgcgtaactacactgca aattgtgacggattcttgtgttcctcaggggggaaataacttgcgtttgacaggtgcgatctagtgtg ccgtctcaaaaatgaaaaaaaacgatgtctggttgcattctctatttgcgccactggtaagcgag aaaattgattaaaatgcacattacaaaggggGCTCGAGGCTCTTCCGTCAAT CGAGTTCAAG

Supplementary Table 3. Oligos and gene blocks (5'-3') designed to produce the *Ran-like* knockout.

Ran-like guide 1 sense	CTTCGATTCAATGGTGTAAAATTGC
Ran-like guide 1 antisense	AAACGCAATTTTACACCATTGAATC
Ran-like homologous arm 1 *Eco*RI	GGGTGTCGCCCTTCGCTGAAGCAGGTGGAATtaaaatacatagaaattatc aaattgcctaagagcttacctttttcgaagattttgtacaaataaaaatcaaaattgtggtaaagctgt gatgcattcaatatttcgaaatatatttactttcagtgaggtccttgctgaagtccttgtgctgtccccctttt tcaggtgctgttgctgtaattgttgctgccgttgctgtgtgcaggtgcttaagcattttacgagcaactgt attcaatttgactgccagcctcccgctttctcgctctggcagcacattacgcatacgccgtgtgttaaa tgcgcttaaaagtcaggccaataagttgttgttgttgctgctgccactgctgctgcgcctgctgttgttg ctgctcttgttgccggtaaattgtagttgctgccactgctgctgtttgtgtaatgaaaacttttacaacgc catatgcaggctattaaaaaacagctcgcacacacagaagcacaaggatgggtccgccagcg ccttggaatatgcaatcaagctcatttatactcattaagcgcttatgtgcactggtatgggtaatgtag atggctacgaagcgagcgtggactccccgggcttaaaatgaatccccactgaaggagcggtttat ttggagcccccgaaattggccaggatccttcaatttgcagcagaacaatgcgcagaacttaatga aatcgaatcgatttttctgcatttttttaacatatgggtcaaaaacttatgaagccgtttaaattataaaa gtttatttaaatttaaatgtagtatctattacatctatctatatatttaaataacagatatagtgggatacat tagtcttttagaaatataaatccgagcctgcaAATTCTTGCATGCTAGCGGCCGCG GACATAT
Ran-like guide 2 sense	CTTCGGTTGTTGTATGATGCCAAG
Ran-like guide 2 antisense	AAACCTTGGCATCATACAACAACC
Ran-like homologous arm 2 *Xho*I	TGCATAAGGCGCGCCTAGGCCTTCTGCAGCggcatcatacaacaacccatt ttcggtacacagttcgtgttgggcaaaagctgtcgcatcaaatgagaatcataaataatttcattttca acttcaatccaaccccaaaaacccccttcgaccaaagattttggcttttaaaattgacacacacacc ccgcccgaaatccttaaacctgtaatttacgagaggcaactgttacacgcacaattttatatgtacat atgtatgtatatatatatatatatttacaatcgcaaacaaaaggaaattgcattcaacaatagttaca

	attgctaggggccgagaatgccataaatcatacgcccagttgtacacgaaggcatttgctgggg cgttaattgtaaataaataacatattaaaatttccagtcaataggtttaagccgtaattaaaagtaata atgcgaaggcgagtgtttgtccttcgccgcttgccatttccatcgtctttgcccatttcgtcctgaaacg ccagcagcagacggtgcaaaagtttttggccccccaaaaaggtcaattaaactgcattaagttta gtaactttcagtgcccggaacagttgcgaagcactgccgattggggtgggcgaaagctgtgggtt cctcggcaggaccaaacaggatcagccttggcgagcggattatgtagaaaacagacatacga atagaacagattgtgcggcaggccacatgcaacgtcaacaacaaacaacacctgtttatggttttt cgttagcagaaattttacaactgccaggacgaaaaggacgaggaatgcgatagagcacacac aGCTCGAGGCTCTTCCGTCAATCGAGTTCAAG

Supplementary Table 4. Oligos, primers and gene block (5'-3') designed to produce the *Ntf-2 (RA) HA Tag* at the C-terminus.

Ntf-2_RA_Guide_Sense	CTTCGCAACGCAGGCACCTTCTTTG
Ntf-2_RA_Guide_Antisense	AAACCAAAGAAGGTGCCTGCGTTGC
Ntf-2_RA_HA2_OL_F	CTACGTATGCACACCACTCAG
Ntf-2_RA_HA2_R_XhoI	CTTGAACTCGATTGACGGAAGAGCCTCGAGCgatccgctggagcttctttc
Ntf-2_RA_HA1_F_EcoRI	GGGTGTCGCCCTTCGCTGAAGCAGGTGGAATgaaaattaatcaaatcaaaatgt tgcgcatttcatttcgtgcatggttttcgttgattttgagttggatttgctctttctttcggttggtcgctggggacct ggaacctcctaataatgcaattttaatgtaacacaaacaatttctgtaacctctattgctttacatttttaacta aacggtttatatttatattatatttttagatgctttcaattgatattgcacttataaaaattctgtttttcatttgttaca aaaccaaataccaattgactagtcttaacttgcatattgttttgttcaatagctcgaagaagagaactgta aatatcccaaattgttaagcgcaccctaatatgccttaaaaaaaaaactttcaaatgttgttcaattttgaat ggaactgaggaaagttttgagaagttcttatatcagtgaataataattttgctcgcacatttggtataaagtt tattctgctcgtgaaaatgttttggagttatgttcaatttatcgagcgagccctgtacagtaaacacgcacct tacacgatcaatcggctgccagttgtacacccttccctctgaccctggcttcaaactgtgtgtactccattc atatctaactcaatacgttgaaattggtttttcgcacgcttgtgttattgactttaaatgtgactggtttgcctgg catctgctaacgtatttatgatttcattcctttgcagtgcgatgacgatcccccacatgccttctcgcaggtctt tttcctgaaggccaacgcaggcaccttctttgtcgcccacgacatcttccgtctcaacatccacaactctg cccaccggtatacaagtttgtacaaaaaagcaggctccgcggccgccccttcaccatggatctccac cgcggtggaggccgcatcttttacccatacgatgttcctgactatgcgggctatccctatgacgtcccgga ctatgcaggatcctatccatatgacgttccagattacgctgctcatggcggataggagcactccacttac ctacgtatgcacaccactcagcaccacaca

Supplementary Table 5. Oligos, primers and gene block (5'-3') designed to produce the *Ntf-2 (RB) HA Tag* at the C-terminus.

Ntf-2_RB_Guide_Sense	CTTCGTCAGACCTTCGTATTGAAGC
Ntf-2_RB_Guide_Antisense	AAACGCTTCAATACGAAGGTCTGAC
Ntf-2_RB_HA2_OL_F	GTCTCGTCGTCGTCCAATTTC
Ntf-2_RB_HA2_R_XhoI	CTTGAACTCGATTGACGGAAGAGCCTCGAGCgatcgtcatcgcactgcaaag
Ntf-2_RB_HA1_F_EcoRI	GGGTGTCGCCCTTCGCTGAAGCAGGTGGAATgaacattatttattttgcactgtttccc atttatttgttatcatgtatgggctttattaagcctgccactttattccattaagtgttcttcattctaagatacattat acccgttactagtgcagttaaagagtatactatattcgttaaaaagtatgtaacatgtagaagcaatcgtttt cgacaaaatgtagtatatatattcttgatcaggatcaatagccgagtcgatctggccctctccatatgaacg tcgagatctcaggaactataaaagctagaaggttgatattaagcagaccgatcctagagacaccgacg cagcgcaagtttgttgatccatgttgccacgcccactcatattgttcaatatgaaatttcgcgttctcacttcca gtcggaaaagtcgaccatagcattctctcttgttttgaatataatatccattgacttattgatgcagttatagta ggtagataaatgaaatgaattttaaaaagagtaactgcctcaatgtattttacgataccttattgttcaaagt ggatatatgtagagacgtagaagcaaatttcaatcgcccatcacttaagatatgaaatcattcatcattcttc cgtaaagtcgtaagggcagtgctatttgccgaccgccgctgcatctattattccatataagtcacctgaatct gaatcgaatatctcaattctggttattacaaaacgaaacacaaatgccaaaacaaaccaatgcagacc gacgaggatcagccgcatgcctatattcagaccttcgtattgaagcccgtgggcggcagtttctttgtgca gcacgatatattccgactctcgctgcacgatgtgcaccggtatacaagtttgtacaaaaaagcaggctcc gcggccgccccttcaccatggatctccaccgcggtggaggccgcatcttttacccatacgatgttcctga ctatgcgggctatccctatgacgtcccggactatgcaggatcctatccatatgacgttccagattacgctgc tcatggcggatagcaccagcgagcccgacctgccacgcccagtccacagactccagcgcccggctat atgtgtgtaaatcgagatcgagatcagcagccaaacgatcagccagcagacacgagacgagagcca gttaaaaaacgaaagaaatccatatctaagcgataagatagttctgctgcttgtcgaccttatgccgccca gccatgcccactccgcccacatatacaaaaatatttatccttggcgtctcgtcgtcgtccaatttcttctaatta

Supplementary Table 6. Oligos, and gene block (5'-3') designed to produce the *Ran HA Tag* at the C-terminus.

*Ran*_Guide_Sense	CTTCGGCGACCGCACTGCCCGACG
*Ran*_Guide_Antisense	AAACCGTCGGGCAGTGCGGTCGCC
Ran_ HA_Gene Block	GGGTGTCGCCCTTCGCTGAAGCAGGTGGAATcaaacgcaattgttccatct gcagctcaccgttccagaaattttgcattaacgaagataccgagtgcttagaaggatggctcagg aaggtcaggatatacccacattcaagtgcgtgctagtcggcgatggcggcaccggcaagacca cctttgtcaagcggcacatgaccggtgagttcgagaagaagtacgtggccacactgggcgtgga ggtgcatccattgatcttccataccaatcgtggcgccatccgtttcaatgtgtgggataccgctggcc aggagaagttcggcgggctgcgcgatggctactacattcagggccagtgtgccgtcatcatgttcg atgttacctcgcgtgtcacgtacaagaacgtgcccaactggcatcgcgatctagtccgcgtctgcg agaacataccaatcgtcctctgtggcaacaaagtcgatatcaaggatcgcaaggtgaaagcga agagcatcgtcttccaccgaaagaagaacttgcaggtaaataaagaaaagctgaacttttatttga aattttgttaacgttattgtgcttgtgccttctttttgtcaaacagtattacgatatttctgccaaatcgaac tacaacttcgagaaaccattcctttggctggcgcgcaagctggttggtgatcccaacctggagtttgt cgccatgccagccctgctgccgcccgaggttaagatggataaggattggcaggcgcagatcga gcgtgacttgcaggaggcccaggcgaccgcactgcccgacgaagacgaggagctacaccgg tatacaagtttgtacaaaaaagcaggctccgcggccgccccttcaccatggatctccaccgcgg tggaggccgcatcttttacccatacgatgttcctgactatgcgggctatccctatgacgtcccggact atgcaggatcctatccatatgacgttccagattacgctgctcatggcggataagcctatttgtagtta acttagagcgaaatccgacttgcgcgaaaacacaaacacacacacactcaaacacaaaacc acaacacacaacaccacaccgaacatagaacactgaatcacaggcaccgtatgaaaaggcg tggcaagagggaatgaaggattacggattagaggaagtagcattagcagtagtgtactatccgta cactattggatttgccagcctatgttcttcgatcgatcgttgggaaaatctacaaacaagcccttctta atacgtcgattgtatttaaattttagatccccaattcattaataagtcggttataattgatttttttttgtttgtt tccttatacgttggctcttggaaaaaaaggagtaagtaagtagtattgtcattaaagtgctaaagca gcaggcaacagcaacttaatcgaaaaccaagagaaataaatagtcgcagaacgtatttttttaaa

	acttgcaacaccaatgttcttttccacccaaaaaaaaaaaggaagaaaaaaccaagttacttca acaaaagtgtgaatgttagaaacgagcgagacggcatgctcggagcgGCTCGAGGCTC TTCCGTCAATCGAGTTCAAG

Supplementary Table 7. Oligos, and gene block (5'-3') designed to produce the *Ntf-2 HA Tag* at the N-terminus.

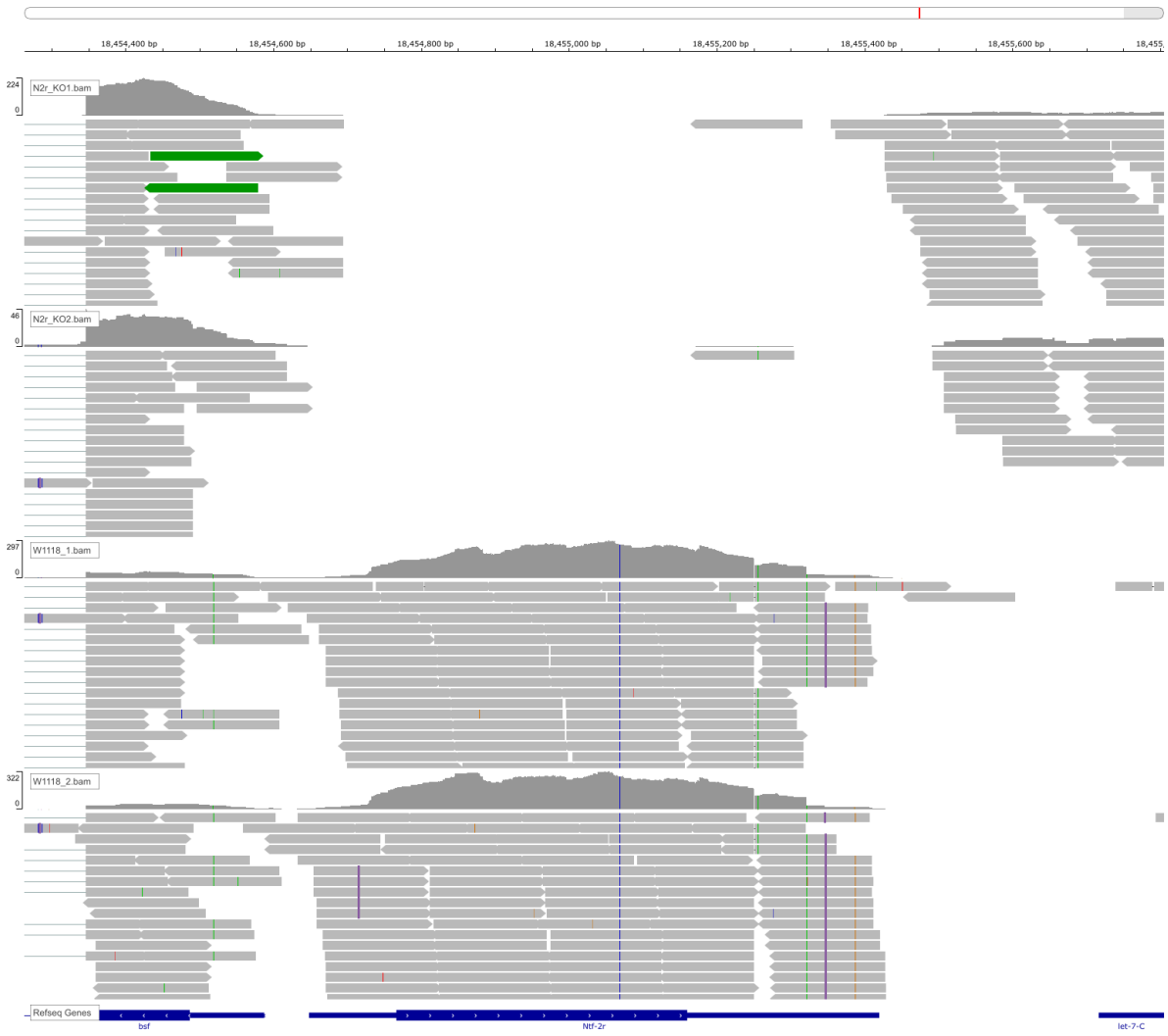
HA_Ntf2_Guide_Sense	CTTCGCCCGGCGAATCGGGCGAACG
HA_Ntf2_Guide_Antisense	AAACCGTTCGCCCGATTCGCCGGGC
HA_Ntf-2_ Gene Block	GGGTGTCGCCCTTCGCTGAAGCAGGTGGAATcccgtctgaaaattgtgaaaatacaagcgattccacattatgaataaacggatctaaaaatcaactttagataccttacaatggattgtaagttgcgttagcttaagtatgacctgcaattaaacaacaaatcaaagaaaagtcgtcccgagcgaggaaaaaaaatatatatgaaacacaaataattgttttccaaatgaccaagtattacgcctgacagaaatgcaataatttgttcttacatttatgtgaccctatcatcatcgaatcatatttcaaattccaattccgatttcgttattaggacatcgggctgtagccctgggtatcgtaatcgatagttgcgaaatagcagtcagcgatagtgtctgacaaaatcgacgttgttccacccctaacaaagagtggtgatctgagcgcagtcggttgatttcatttggttttttttttaattattcgtgtcgccgcgatcggatcggatttcccataatctctgagcgttccgcctatccttcaagtgaaatggatctccaccgcggtggaggccgcatcttttacccatacgatgttcctgactatgcgggctatccctatgacgtcccggactatgcaggatcctatccatatgacgttccagattacgctgctcatggcggacaccggtatacaagtttgtacaaaaaagcaggctccgcggccgccccttcaccatgtcgctgaatccgcagtacgaggacattggcaagggatttgtgcagcagtactatgcgatattcgatgacccggcgaatcgggcgaacgtcgttaatttctatagcgtaagtgttgttgcttcccgcttctgctcgtcgcctttcatctctcctttattcccgtctccgctttcgcctgcacatttttcgcttgccatccccatcacacgcacaggtgacgcgcggcgcgcccttgaaaaattccgcctcctttcagcagaaaatgaataaagaaactgttgcaatgcgaaaaagagcgcgaaaaagagcagctggagatatacggtcggcgtatccgtgatatcgtgggaataggatgaatgttgataggggcgggggcagcgttcgattctcgaagtttcggggtgacacattacgcccatgctcgcataaataaatacacacatacgaaatatatggccaacatacaactgcggttagcacaatagggcccaatatacgatataaattcgtgaaaaaattgtattaatatcttgaaaaatatggcataattgtgtttcattagtttcaatcaattcagttttttgtcacataaaagagtaaatgaaagaataaaaaagtctataatgataaagtattgaaaataaaactatctgtttgatatgtttttaatttaaatattaacagttttataaccaattttgttcttaataatacgaaatatcattactaagtaaaacatcatcataacacttatatcattatatattttatacttccaggctaccgactcattcatgacctttgaaggccaccaaatacaggggcacccaagattctggaa

	aaagttcaggtaagtcctgattatatacagtcatgggcatGCTCGAGGCTCTTCCGTCAATC GAGTTCAAG

Supplementary Table 8. Oligos, and gene block (5'-3') designed to produce the *Ran HA Tag* at the N-terminus.

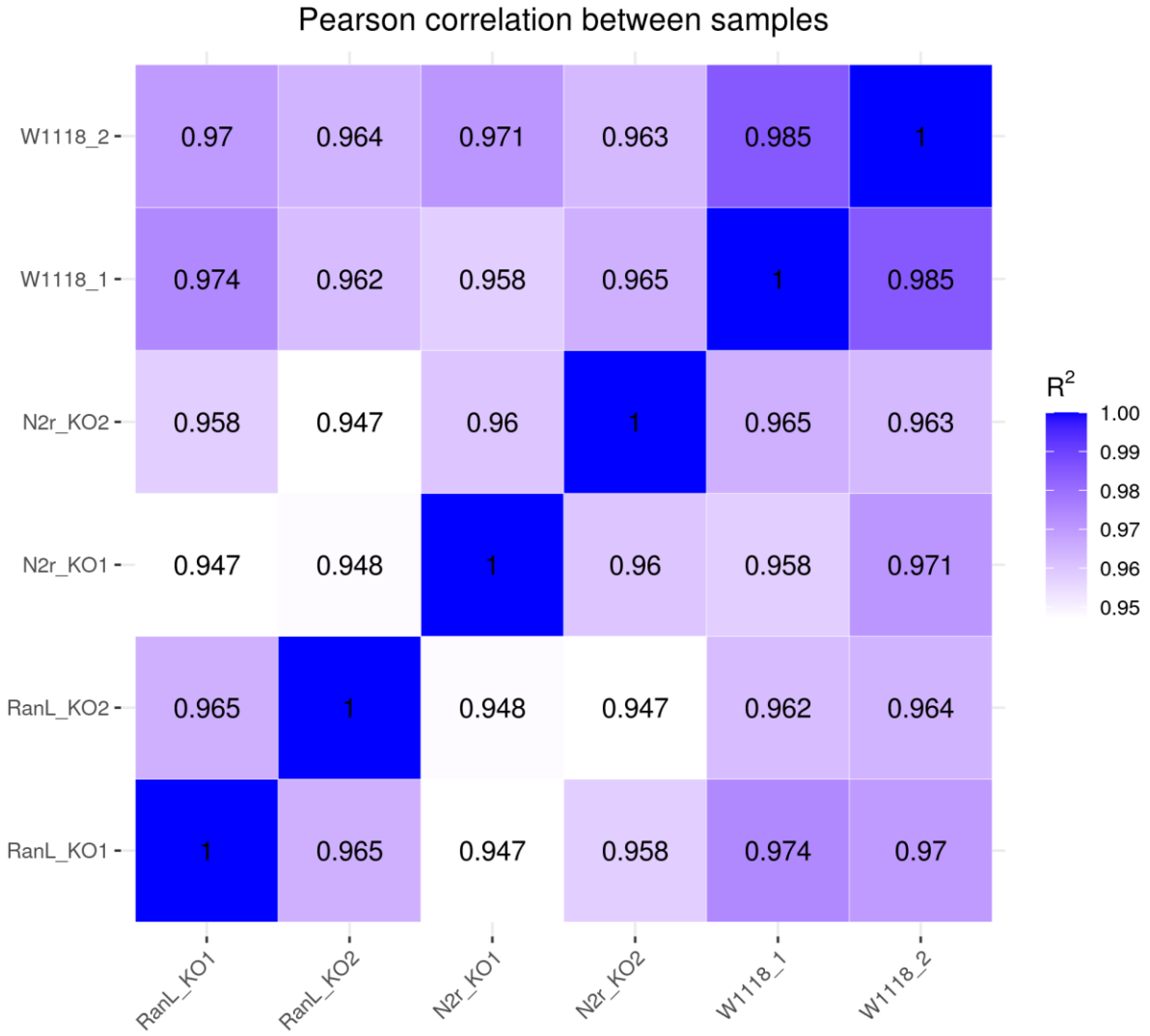
HA_Ran_Guide_Sense	CTTCGCTTGACAAAGGTGGTCTTGC
HA_Ran_Guide_Antisense	AAACGCAAGACCACCTTTGTCAAGC
Ran_ HA_Gene Block	GGGTGTCGCCCTTCGCTGAAGCAGGTGGAATgggctgcgcataaacgtggtggg aatgtatgtgtgcgagtatgcagagtatatccttccgcctccctgccgccttcttgttgttgttgtttggtgcg cctggcatttctgcttacctgcacacacagccagccataccctgtggccatctcgctctggctcggcatctc ggccttcttatccttgcgggctcagcgggagcggcagctgcagtgtcctttgtcgttatcctgctcgctcctcg ccgaacagctggtcgccgtccttggcgggctttgacgcatccattgtgcacgtcacattttggacggctcg caggacgggcagtcagtcgcctggcgctcgccgaaccaacaacatcctagccaacacattccgctcct gttcctgttcctgctcaagctccgctatatatatatatatttaaatatataaagtagcatagtagttgcaagttg ctaaccagccaaacgcaattgttccatctgcagctcaccgttccagaaattttgcattaacgaagataccg agtgcttagaaggatggatctccaccgcggtggaggccgcatcttttacccatacgatgttcctgactatgc gggctatccctatgacgtcccggactatgcaggatcctatccatatgacgttccagattacgctgctcatgg cggacaccggtatacaagtttgtacaaaaaagcaggctccgcggccgccccccttcaccatggctcagg aaggtcaggatatacccacattcaagtgcgtgctagtcggcgatggcggcacgggcaagaccacctttt gtcaagcggcacatgaccggtgagttcgagaagaagtacgtggccacactgggcgtggaggtgcatc cattgatcttccataccaatcgtggcgccatccgtttcaatgtgtgggataccgctggccaggagaagttcg gcgggctgcgcgatggctactacattcagggccagtgtgccgtcatcatgttcgatgttacctcgcgtgtca cgtacaagaacgtgcccaactggcatcgcgatctagtccgcgtctgcgagaacataccaatcgtcctct gtggcaacaaagtcgatatcaaggatcgcaaggtgaaagcgaagagcatcgtcttccaccgaaagaa gaacttgcaggtaaataaagaaaagctgaacttttatttgaaattttgttaacgttattgtgcttgtgccttcttttt tgtcaaacagtattacgatatttctgccaaatcgaactacaacttcgagaaaccattcctttggctggcgcg caagctggttggtgatcccaacctggagtttgtcgccatgccagccctgctgccgcccgaggttaagatg gataaggattggcaggcgcagatcgagcgtgacttgcaggaggcccaggcgaccgcactgcccgac gaggacgaggagctataagcctatttgtagttaacttagagcgaaatccgacttgcgcgaaaacacaa

	acacacacacactcaaacacaaaaccacaacacacaacaccacaccgaacatagaacactgaatc acaggcaccgtatgaaaaggcgtGCTCGAGGCTCTTCCGTCAATCGAGTTCAAG



Supporting Figure 1. Mapping results in BAM format are visualized by the Integrative Genomics Viewer (IGV) software and shown for the *Ntf-2r* locus and compared to strain of control w^{1118}. the visualization of mapping results shows absence of *Ntf-2r* transcripts in *Ntf-2r* knock out strains.

Supporting Figure 2. Mapping results in BAM format are visualized by the Integrative Genomics Viewer (IGV) software and shown for the *Ran-like* locus and compared to strain of control w^{1118}. the visualization of mapping results shows absence of *Ran-like* transcripts in the *Ran-like* knockout strains.



Supporting Figure 3. Correlation coefficient matrix between samples.

Values represent R^2 (Square of Pearson correlation coefficient(R)).

Chapter 4

Supplementary Materials

Supplementary table 1. Detailed information about the strains used in this study.

Strain name	Genotype	Stock Center
COX4L-RNAi / CG10396-RNAi (ID # 1482)	*w^{1118}; P[GD414]v1482*	VDRC
COX4L-RNAi / CG10396-RNAi (ID # 106700)	*w^{1118}; P[KK102531]v106700*	VDRC
Nos Cas9 attp2	*y,sc,v; +/+; nos-Cas9*	Rainbow Transgenic Flies, Inc.
Actin5C-Gal4 (ID # 4414)	*y[1] w[*]; P{w[+mC]=Act5C-GAL4}25FO1/CyO, y[+]*	BDSC
bam-Gal4	*y[1] w[*] P{w[+mC]=bam-GAL4:VP16}1*	BDSC
Tub-Gal4	*y[1] w[*]; P{w[+mC]=tubP-GAL4}LL7/TM3, Sb[1] Ser[1]*	BDSC
w^{1118} / GD control (ID # 60000)	*w[1118]*	VDRC
w^{1118}/ KK control (ID # 60100)	*y,w[1118];P{attP,y[+],w[3`]}*	VDRC

Supplementary table 2. gRNA primers and homologous arm gene blocks (5'-3') designed to produce the *COX4L* knockout and primers to confirm the *COX4L* knockout by PCR.

COX4L guide 1 sense	CTTCGTGTAATTATGCGCAAGCACT
COX4L guide 1 antisense	AAACAGTGCTTGCGCATAATTACAC
COX4L homologous arm 1 *Eco*RI	GGGTGTCGCCCTTCGCTGAAGCAGGTGGAATgcttttcgct ggccatagctctcgtggcccttgcagccactctggttcgtgccaatatagatgata acctgccaattaacactgagggacggattgattttgatttcggcttcgaagaggca atttggggaaaaaaccgtaaaaataggaaatatctaaagaatagctcgacaatt tttcacaaattacaaatttaattattaaataattatttggaaagtttttaaaatttttgtttt cggaattgttttatttttgtgtttttttttttcatgatatttacttttaagagattggcaaatgc ttcattctctaagtagagcgagattgtctttaatgtcttatattttctaaagtatagcttttt ttaaaattcttaagggtgggccaacaatgttattgcgatttaaaaatttttgaaaaa agtcaactagttgattcttaaactttatcaaaatttcagatattgaaaactggacgtg ggcaaaaaaaataattattgggcaaacagttctagatttcaaaaattcgatttttcc gaacccagcttctttgagctgacatgacagccatttttaaaaatgtttgttttttttttgt gacaaaaaatttgatcttcataatttttgccacgccttaaacaatttttaagaagaa gtaaaattttcagactatcttagtgctcaacgaagagtgcaattcagaacttaaaa agtacatctagtttgtagataaggaaactgtcatatttttttttgtattcaacaaacag actagagaaatttcattttcattcgacacgagcaacacaactgtcgaatttccgga tgaagtaaaaaacaaaaaattgaaaagcgagtataaaataaaatacactcaa ggtacagttacgaccaagtAATTCTTGCATGCTAGCGGCCGCG GACATAT

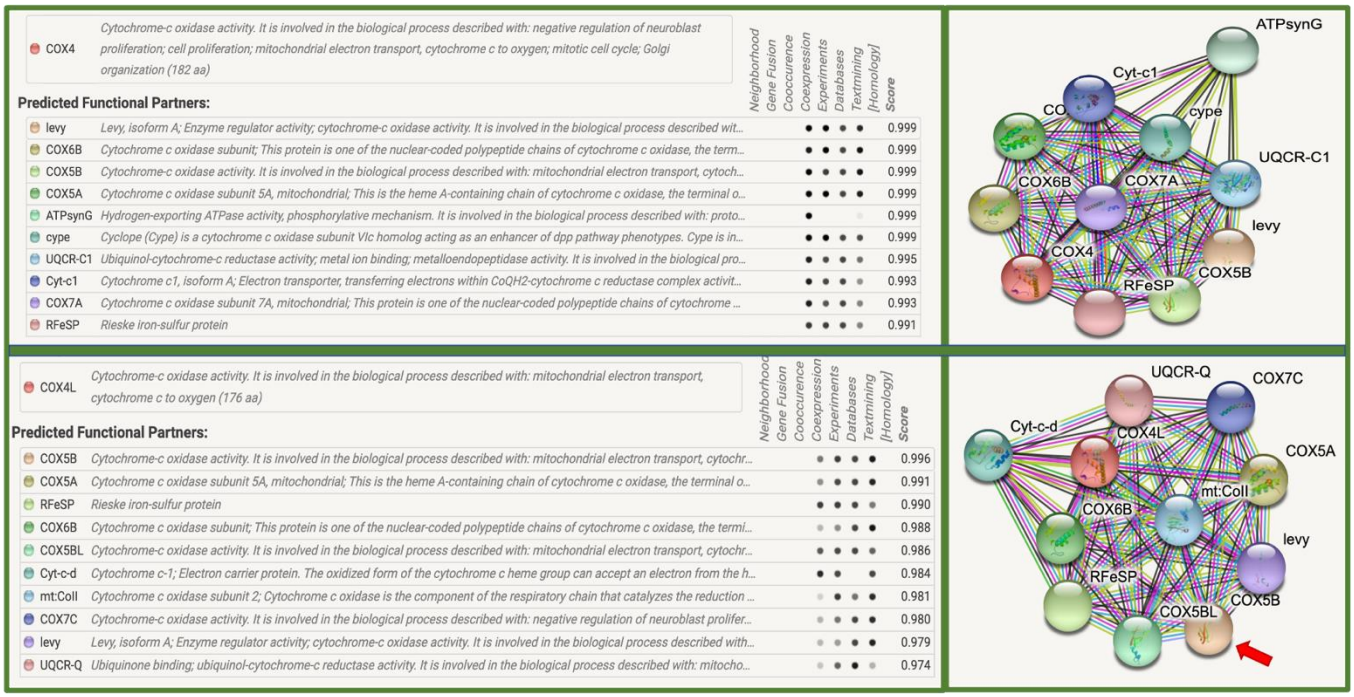
COX4L guide 2 sense	CTTCGTTCTCTCGGTAGCACCATT
COX4L guide 2 antisense	AAACAATGGTGCTACCGAGAGAAC
COX4L homologous arm 2 *Xho*I	TGCATAAGGCGCGCCTAGGCCTTCTGCAGCggtgctaccg agagaacaagtggaagtagcacatcaacacgatgattttccgtaaactatgtac agaaacgtaactagcaaaatacaattcaacagcaaagtcgcttgccatttgatc gttacgtgctgaatcgggcaaatacccaatatctatagatttttgtgtctctagctgtg taactcgactatagcatttcctcccgtttgaaattagggtgtgtatgtaaattctcaga cacaacttaatttagtgtaattttagtccacgatagatatgttaagcattgaaatcgt gtcctgtgttcctttgactagtacacgtacactgcgcgtcatcagattagcgcctcc ctgtatgccaccgtttcatcttatgatctgtattttccattgcacgaaaatctatcaatg ttattgttttttgttcactgatattccctctctctttgagaataaaaaagaggttgagag aagaacagttatctcttttattctgctttgtgttaactttggcgcaaaattgaacacgt gttttgcatcatcagattagcgcccccagttttgaatatcgtccataatttgaaaggt aggacaacaaaattttattaaaaacaaggaatcttatagaaaaaactataattgt ggcaataaccgtgtcgttaccgcggcagattgcaggtccatacttcgaattgcttc caattcccacgactcctccgccaaaattcgagaaaaaagtggtataagcgcga gcaaatcgacgattccaagggtgacttagaaagctaaacatttaaagcacaac gaactaacGCTCGAGGCTCTTCCGTCAATCGAGTTCAAG
Forward primer to confirm the *COX4L* knock-out	GTACACTGCCGTCGAAATGAG
Reverse primer to confirm the *COX4L* knock-out	GGGTTCATTTGCAGCTGGATG

Supplementary table 3. gRNA primers and homologous arm gene blocks (5'-3') designed to produce the *COX4* knock-out and primers to confirm the *COX4* knock-out by PCR.

COX4 guide 1 sense	CTTCGTATATTCAATCGCTAGCAAT
COX4 guide 1 antisense	AAACATTGCTAGCGATTGAATATAC
COX4 homologous arm 1 *Eco*RI	GGGTGTCGCCCTTCGCTGAAGCAGGTGGAATgcgcacttg ttaaaaatgcaaattgtcctagccacaaactgaacctgataacgaagcactacc atcccaaaaagcaaggaccccggctgcgtttccgaaaaacccaaatcttgaga aaaggaaacttcactaacatgaaatgttacagagataagacgagtgaattcattt ttgtacatatagatgaaaactaaccccaagtccccctatgcttaagtgtgaatct gaccgacggtctagacttttataccctttatcccagcccagagtcggcggatcat atgaattaaatcgcaactaatacctaactctactgtaaatgtagctataagttttgc tactaccaatttacaagatgtatagttgaaggaaaaagctttcactttagtttaattat ttgcgtaattatcattaaatattgttacaaacatattgcatatttataaagcccaataa agcaaaataccattaacatatttattatgccaagcggcgcttaaacacaggaata cgcgaaagtgaatacgtaatgccccatcctcgataaaatagctaaatttccgaa ctccagatccctatcgaaagataatcttatgtaatacaaactcaagcagcaaact atatgatacgataaagcaaattatgtgtctctgtctaatttaaatgtgtgcatttatga gatatgacatcaaatcaaaataaaatataatatcgaacaataattgttttgttgttgt ttctcttctcgcggacagactacaaaaagaatgtgaaataccaaaccaaatattt caccatgctgcgaaataaaataaaacataatagttgcccttatttgctgtttctttattt taatatatttagatttaaataccccgccgaattccgattAATTCTTGCATGC TAGCGGCCGCGGACATAT

COX4 guide 2 sense	CTTCGCTCTTAAAACGGATCATTG
COX4 guide 2 antisense	AAACCAATGATCCGTTTTAAGAGC
COX4 homologous arm 2 *Xho*I	TGCATAAGGCGCGCCTAGGCCTTCTGCAGCtgatccgtttta agagcaatgattgttattcgatttattaacagacctttagtgtgtagattacaaacttt tcgttactaaaatcgccacaggcatagccaatttaattaccacttagtgagaccttt tagcactcttatgtctcatgttccggctgttcgatttttgaacctcatgaacaaggatc tgcaaggcttccaaggcatttgtgttcttggcaaattccctttcaaacatgcggccg aatcttaagatcaatggtttcagttccggcgttatcattgctttctctttttcggtcctaa aaaggcggatttcaaaggcattgccagacatataggcgaacactacttacaca gctccctttttttatcacctctttcaggatctggacaaccgtttgaatgtccccaaataa tctgtacttgtctttggtagaaatttcaccgtcaaagaataacttcagttcgttgtcagt tatattcttctttcgaccccattcctccagctctgccgtctctttttggtaggcgctccga gctctgctgaaaagaagtcaggcattccttgatctcagagcattcatacttgatgc agatgatgcacctacctccagttcgaaatatccctagctgtctgcttcatctttttaa atgtcctgatctgtgtttgaaagtcttctcttctttctcctttgcaactgtcgcctttggttc ggcatctcggttaaagctattgaaagcgaaaggttggtgccattcgaaagtgtat atatccttcagctcgatcgaatctgtatcttactgtttgtgccaatcgtatccatcgtat ggaggcctgtgcagattgtgtgctcggagatgacgctgtctttccatctgggtgtgg ttctggttaccGCTCGAGGCTCTTCCGTCAATCGAGTTCAAG
Forward primer to confirm the *COX4* knock-out	ACTACTCAACAGTGCTGTGCT
Reverse primer to confirm the *COX4* knock-out	GTGTAGTCCTCGTTGTGGGA

Supplementary table 4. Physical protein interaction analysis of *COX4L* with STRING (V 11.0)



		Neighborhood	Gene Fusion	Cooccurence	Coexpression	Experiments	Databases	Textmining	[Homology]	Score
COX4	*Cytochrome-c oxidase activity. It is involved in the biological process described with: negative regulation of neuroblast proliferation; cell proliferation; mitochondrial electron transport, cytochrome c to oxygen; mitotic cell cycle; Golgi organization (182 aa)*									
Predicted Functional Partners:										
levy	*Levy, isoform A; Enzyme regulator activity; cytochrome-c oxidase activity. It is involved in the biological process described wit...*				●	●	●	●		0.999
COX6B	*Cytochrome c oxidase subunit; This protein is one of the nuclear-coded polypeptide chains of cytochrome c oxidase, the term...*				●	●	●	●		0.999
COX5B	*Cytochrome-c oxidase activity. It is involved in the biological process described with: mitochondrial electron transport, cytoch...*				●	●	●	●		0.999
COX5A	*Cytochrome c oxidase subunit 5A, mitochondrial; This is the heme A-containing chain of cytochrome c oxidase, the terminal o...*				●	●	●	●		0.999
ATPsynG	*Hydrogen-exporting ATPase activity, phosphorylative mechanism. It is involved in the biological process described with: proto...*				●					0.999
cype	*Cyclope (Cype) is a cytochrome c oxidase subunit VIc homolog acting as an enhancer of dpp pathway phenotypes. Cype is in...*				●	●	●	●		0.999
UQCR-C1	*Ubiquinol-cytochrome-c reductase activity; metal ion binding; metalloendopeptidase activity. It is involved in the biological pro...*				●	●	●	●		0.995
Cyt-c1	*Cytochrome c1, isoform A; Electron transporter, transferring electrons within CoQH2-cytochrome c reductase complex activit...*				●	●	●	●		0.993
COX7A	*Cytochrome c oxidase subunit 7A, mitochondrial; This protein is one of the nuclear-coded polypeptide chains of cytochrome ...*				●	●	●	●		0.993
RFeSP	*Rieske iron-sulfur protein*				●	●	●	●		0.991

		Neighborhood	Gene Fusion	Cooccurence	Coexpression	Experiments	Databases	Textmining	[Homology]	Score
COX4L	*Cytochrome-c oxidase activity. It is involved in the biological process described with: mitochondrial electron transport, cytochrome c to oxygen (176 aa)*									
Predicted Functional Partners:										
COX5B	*Cytochrome-c oxidase activity. It is involved in the biological process described with: mitochondrial electron transport, cytochr...*				●	●	●	●		0.996
COX5A	*Cytochrome c oxidase subunit 5A, mitochondrial; This is the heme A-containing chain of cytochrome c oxidase, the terminal o...*				●	●	●	●		0.991
RFeSP	*Rieske iron-sulfur protein*				●	●	●	●		0.990
COX6B	*Cytochrome c oxidase subunit; This protein is one of the nuclear-coded polypeptide chains of cytochrome c oxidase, the termi...*				●	●	●	●		0.988
COX5BL	*Cytochrome-c oxidase activity. It is involved in the biological process described with: mitochondrial electron transport, cytochr...*				●	●	●	●		0.986
Cyt-c-d	*Cytochrome c-1; Electron carrier protein. The oxidized form of the cytochrome c heme group can accept an electron from the h...*				●	●		●		0.984
mt:CoII	*Cytochrome c oxidase subunit 2; Cytochrome c oxidase is the component of the respiratory chain that catalyzes the reduction ...*				●	●	●	●		0.981
COX7C	*Cytochrome-c oxidase activity. It is involved in the biological process described with: negative regulation of neuroblast prolifer...*				●	●	●	●		0.980
levy	*Levy, isoform A; Enzyme regulator activity; cytochrome-c oxidase activity. It is involved in the biological process described with...*				●	●	●	●		0.979
UQCR-Q	*Ubiquinone binding; ubiquinol-cytochrome-c reductase activity. It is involved in the biological process described with: mitocho...*				●	●	●	●		0.974

Supplementary table 5. Average values [± SE] of viability results of *COX4L* knockdown in soma with *Actin5c-Gal4* driver at 25°C and 27°C.

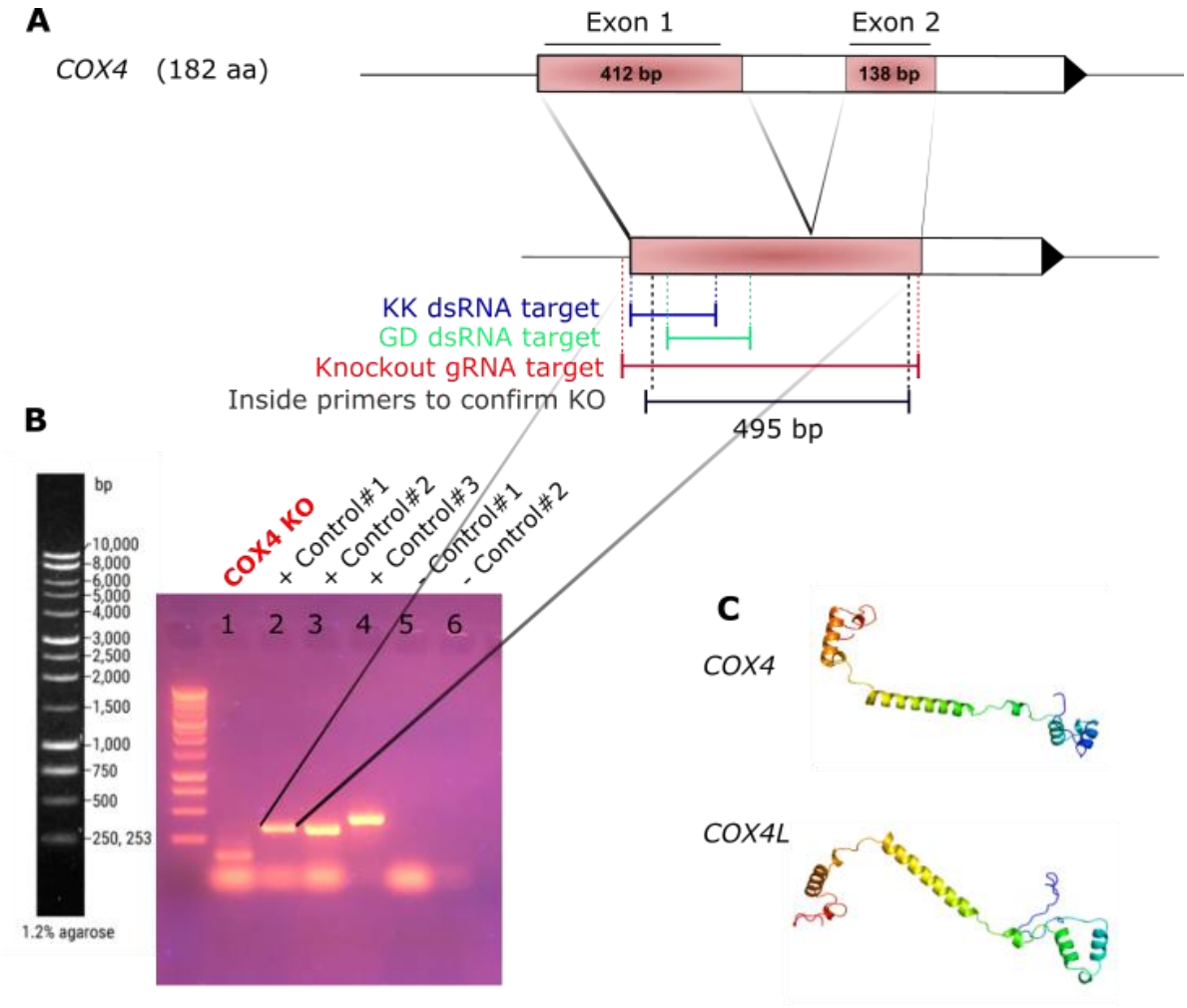
Crosses	Average number of progeny at 25 °C	Average number of progeny at 27 °C
♀COX4L-KK x ♂Actin5c-Gal4	91.3 ± 9.9	101.3 ± 1.7
♂COX4L-KK x ♀Actin5c-Gal4	86.7 ± 2.4	100.7 ± 6.1
♀COX4L-KK x ♂w^{1118}	80.7 ± 11.7	85.7 ± 17.8
♂COX4L-KK x ♀w^{1118}	65.0 ± 23.6	87.3 ± 15.2
♀Actin5c-Gal4 x ♂w^{1118}	78.7 ± 2.9	79.3 ± 5.0
♂Actin5c-Gal4 x ♀w^{1118}	80.0 ± 7.4	82.0 ± 7.9
♀COX4L-GD x ♂Actin5c-Gal4	80.0 ± 8.3	64.7 ± 3.5
♂COX4L-GD x ♀Actin5c-Gal4	59.3 ± 14.8	72.7 ± 2.2
♀COX4L-GD x ♂w^{1118}	85.7 ± 12.4	66.3 ± 5.7
♂COX4L-GD x ♀w^{1118}	70.7 ± 1.9	58.3 ± 9.3
♀Actin5c-Gal4 x ♂w^{1118}	67.3 ± 2.6	62.0 ± 9.3
♂Actin5c-Gal4 x ♀w^{1118}	72.7 ± 1.5	67.3 ± 6.3

Supplementary table 6. Average values [± SE] of fertility results of *COX4L* knockdown in germline with *bam-Gal4* driver at 25°C, and 27°C.

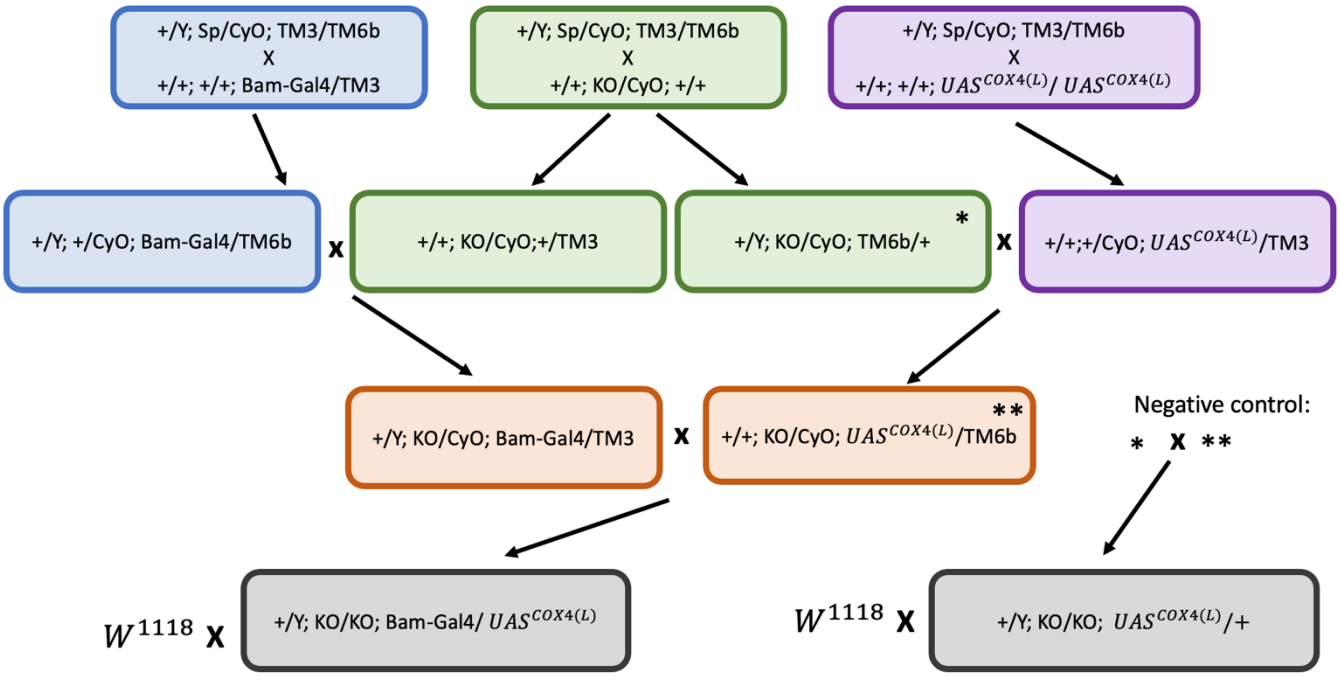
Crosses	Average number of progeny at 25 °C	Average number of progeny at 27 °C
♀ [♀KK x ♂Gal4](27c) x ♂W1118(25c)	69.3± 4.9	80.3± 1.2
♀[♀KK x ♂W1118] (27c) x♂W1118(25c)	57.0 ± 3.7	68.0± 1.4
♂ [♀KK x ♂Gal4](27c) x ♀W1118(25c)	44.7± 8.6	51.7± 2.9
♂[♀KK x ♂W1118] (27c) x♀W1118(25c)	72.7± 4.2	71.0± 8.1
♀[♂GD x ♀Gal4] (27c) X ♂W1118(25c)	69.0± 4.7	80.3± 1.6
♀[♂GD x ♀W1118](27c) X ♂W1118(25c)	60.3± 3.5	65.7± 2.4
♂[♂GD x ♀Gal4] (27c) X ♀W1118(25c)	47.0± 2.1	59.7± 4.1
♂[♂GD x ♀W1118] (27c) X ♀W1118(25c)	62.3± 3.3	74.0± 1.9

Supplementary table 7. Average values [± SE] of viability results of *COX4L* knock-out strain at 25°C

Crosses	**Average number of KO/KO progeny at 25 °C**	**Average number of KO/CyO progeny at 25 °C**
♀KO/CyO x ♂KO/CyO(25c)	23.3± 1.8	39.7± 1.2
	Average number of +/+ progeny at 25 °C	**Average number of +/CyO progeny at 25 °C**
♀+/CyO x ♂+/CyO(25c)	24.3 ± 1.2	40.0± 1.0



Supplementary figure 1. (A) *COX4* and *COX4L* gene structures. *COX4L* is an RNA-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 *COXL*-KO target region is shown. The location of the primers for PCR within *COX4L* are depicted. (B) The lack of a PCR amplification for *COX4L*-KO homozygotes is shown confirming the removal of *COX4L* from the genome. (C) Tertiary structures of *COX4* and *COX4L* generated with Phyre.



Supplementary figure 2. Cross scheme used in this study for the rescue of the *COX4L-KO* phenotype with driving COX4L-ORF and COX4-ORF with *bam*- and *nos*-*Gal4* drivers.