© Copyright 2019

Bhavatharini Kasinathan

Phylogenomic and functional analyses of ZAD-ZNF transcription factors in *Drosophila*

Bhavatharini Kasinathan

A dissertation

submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

University of Washington

2019

Reading Committee:

Harmit S. Malik, Chair

Celeste Berg

Barbara Wakimoto

Program Authorized to Offer Degree:

Molecular and Cellular Biology

University of Washington

Abstract

Phylogenomic and functional analyses of ZAD-ZNF transcription factors in *Drosophila*

Bhavatharini Kasinathan

Chair of the Supervisory Committee:
Harmit S. Malik, Affiliate Professor
Department of Genome Sciences

Evolutionarily dynamic, essential genes challenge the long-held dogma that essential genes are highly conserved. Although these genes are scientifically compelling, only a handful of studies have identified evolutionarily young, rapidly evolving genes and determined their function. Examining the process by which a young gene becomes essential can shed light on how gene origin and evolution drive biological diversity. I focused my studies on *ZAD-ZNF* genes, which encode the most abundant yet dynamic class of transcription factors in *Drosophila melanogaster* and related species. I found significant evidence of genetic innovation via both gene turnover and positive selection in *ZAD-ZNF* genes. Although many ancient, conserved ZAD-ZNFs are essential for viability or fertility, I unexpectedly found that genetically innovating ZAD-ZNFs are more likely to be essential. I took a candidate locus approach to understand the cause of ZAD-ZNF dynamism, focusing on one cluster of five ZAD-ZNFs. Through cytological analyses, I found that three of the five ZAD-ZNFs localize to heterochromatin, a rapidly evolving compartment of the nucleus. I focused my functional characterization on Nick Nack, a rapidly evolving, evolutionarily young, essential ZAD-ZNF within the cluster. I found that Nick Nack is necessary for larval development and for maintenance of heterochromatin. This work identifies the seemingly paradoxical finding that rapidly evolving ZAD-ZNF genes can be essential for development in Drosophila and suggests

that rapidly changing heterochromatin functions may underlie the diversification of this gene family.

TABLE OF CONTENTS

List of Figures iv

List of Tables vi

Chapter 1. Introduction 1

1.1 Mechanism of gene birth 2

1.2 How to identify new genes 6

1.3 *Umbrea:* Stepwise evolution of a young, essential gene 7

1.4 Dissertation overview 10

Chapter 2. Phylogenomics of *Drosophila* ZAD-ZNFs 12

2.1 Introduction 12

2.2 Results 16

2.2.1 ZAD-ZNFs are a dynamic and diverse protein family in Drosophila 16

2.2.2 Rapidly evolving ZAD-ZNFs are frequently essential in D. melanogaster 17

2.2.3 The Odj-Nnk cluster of ZAD-ZNF genes is dynamically evolving in Drosophila 22

2.2.4 Four Oddjob cluster ZAD-ZNFs localize to heterochromatin 29

2.2.5 Oddjob and Nick Nack are necessary for viability and fertility 32

2.2.6 Nnk's recent adaptive evolution is not linked to its essential function 34

2.3 Discussion 38

2.4 Methods 42

2.4.1 *Bioinformatic analyses* 42

2.4.2 *Defining gene essentiality* 42

2.4.3 *Defining orthologs* 42

2.4.4 *Analyses of positive selection* 43

2.4.5 *Cloning Oddjob cluster genes into expression vectors* 44

2.4.6 *Tissue culture and transfection* 44

2.2.7 *Immunocytochemistry and imaging* 44

2.2.8 *Viability studies* 45

2.2.9 *Fertility studies* 46

2.2.10 *Transgene design* 46

2.2.9 *Transgene rescue cross* 47

Chapter 3. Nick Nack is necessary for larval progression in *D. melanogaster* 48

3.1 Introduction 48

3.2 Results 51

3.2.1 *Nick Nack is an essential gene in D. melanogaster* 51

3.2.2 *Nick Nack is necessary for larval progression* 54

3.2.3 *Examining the transcriptome of Nnk-null larvae* 55

3.2.4 *Nick Nack is an essential gene in D. melanogaster* 59

3.2.5 *Nick Nack is a suppressor of position-effect variegation* 65

3.3 Discussion 67

3.4 Methods 69

3.4.1 *Determining Lethal Phase of Nnk-null mutants* 69

3.4.2 *Larvae collection for RNA-sequencing* 69

3.4.3 *RNA extractions* 70

3.4.4 *Transcriptome Data Analysis* 70

3.4.5 *Dietary Sterol Supplementation* 71

3.4.6 *Defining Heterochromatic Genes* 71

3.4.7 *Sterol Absorption Assay* 72
3.4.6 *Position Effect Variegation Assay* 72
Chapter 4. Conclusions and Future Perspectives 73
Bibliography 79

LIST OF FIGURES

1.1 Mechanisms for the origins of new genes 2

1.2 Fates of duplicated genes 4

1.3 Syntenic alignments facilitate the identification of new genes........ 6

1.4 Gene duplication and diversification produced *Umbrea* 9

2.1 ZAD-ZNF Structure 14

2.2 Total number of ZAD-ZNFs across Drosophila 16

2.3 Evolutionary relationship among all Drosophila *Oddjob* paralogs........ 25

2.4 *Odj-Nnk* cluster genes are rapidly gained and lost in Drosophila 27

2.5 Recombination and positive selection analyses. 28

2.6 Oddjob cluster ZAD-ZNFs localize to distinct regions of heterochromatin........ 31

2.7 Oddjob and Nick Nack are necessary for viability and fertility........ 32

2.8 Nick Nack transgene rescue 34

2.9 Nick Nack *D. simulans* ortholog localizes to heterochromatin in *D. melanogaster* 36

2.10 Nick Nack *D. simulans* rescues *Nnk* null ubiquitous knockdown in *D. melanogaster*.. 37

3.1 Fluctuations in ecdysone levels trigger developmental transitions........ 49

3.2 Nutrient flux during ecdysone biosynthesis 50

3.3 *Nick Nack* alleles used in this study 53

3.4 *Nick Nack* null larvae arrest in early larval development 55

3.5 *Nick Nack-null* larvae fail to turn on a significant number of genes 48 hours AEL 57

3.6 *Nick Nack*-null larval transcriptome 48 hours AEL is similar to the *Nick Nack* null and control larval transcriptomes 24 hours AEL. 58

3.7 *Nick Nack* null midguts show sterols within midgut upon filipin staining 59

3.8 Addition of dietary sterols has no effect on viability of *Nnk*-null larvae 62

3.9 Comparison of expressed heterochromatin genes in *Nnk*-null and control larvae across two timepoints 64

3.10 Progressive derepression of heterochromatin-embedded elements in the *Nnk*-null larvae 65

3.11 *Nick Nack* is a suppressor of variegation 66

LIST OF TABLES

2.1 Thirteen rapidly evolving ZAD-ZNFs in *D. melanogaster.* 20

2.2 Oddjob cluster ZAD-ZNFs evolve under different pressures 29

ACKNOWLEDGEMENTS

This work would not be possible without the scientific guidance, contributions and insights from colleagues, collaborators and Harmit Malik, my advisor. The work in chapter 2 was assisted by Hannah McConnell and the work in chapter 3 was assisted by Janet Young and Isabelle Hwang, with fly lines generated in the labs of Leo Pallanck, Ryusuke Niwa, and Eric Lai and input from Leo Pallanck, Serafin Colmenares and Gary Karpen. This work was financially supported by the NIH individual predoctoral fellowship (F30 CA225077), the Chromosome Metabolism and Cancer Training Grant (T32 CA009657) and the Julie Tall ARCS endowment from the Seattle Chapter of the ARCS Foundation.

To the entire Malik lab, past and present, thank you for making the lab an amazing place to learn and for helping me become a better scientist. I specifically want to thank Mia Levine, Antoine Molaro, Aida de la Cruz, Janet Young, and Michelle Hays for their guidance and support. I am grateful to the Henikoff lab, especially to Steve Henikoff, Paul Talbert, Derek Janssens, Tayler Hentges and Terri Bryson, for including me in lab meetings, providing technical advice and experimental support. Thank you to the amazing scientists, trainees and staff that make the Fred Hutch such an incredible place to learn. Specifically, thanks to Susan Parkhurst, Sue Biggins, Ava Brent and Julio Vasquez for their experimental advice; to Pat Heath and Luna Yu who have solved all of my IT problems; to Elizabeth Jensen for the sequencing support; to Laura Masserman for keeping tabs on all Fred Hutch graduate students and for the well-stocked candy bowl.

Thank you to my thesis committee, Sue Biggins, Celeste Berg, Stephen Tapscott and Barbara Wakimoto for challenging me to think critically about my work and how it fits into the larger context and to Celeste and Barbara for reading and providing feedback on my thesis and making sure that I am welcome in the fly community. Thanks especially to Barbara for helping me connect with insect development experts, Celeste for her enthusiasm and encouragement on the project, and to Stephen for his pragmatic advice at each stage of this path to becoming a clinician scientist.

Thank you to my mentors during my undergraduate studies, Alexey Soshnev, Lacy Barton, Sarit Smolikove, and Laurie Snider, who showed me how fun science can be and encouraged me to pursue this path.

I am incredibly grateful for the UW MSTP leadership, Marshall Horowitz, Mary-Claire King, and Stephen Tapscott, their mentorship and support over the years. Thank you to the wonderful administrators that have kept me on track, especially to Marcie Buckner, Maia Low, Alex Moreno, and Laura Masserman.

I am so grateful for all my MSTP classmates who have turned from colleagues to dear friends over the years, especially to Elise Cai, Alec Gibson, Ted Gobillot, Juhye Lee and Mark Ragheb. Thank you for raising the bar and for helping me meet those standards. A huge thank you to my friends and family outside the program for their patience, love and support. I'm especially indebted to Britney Keller for her steady friendship and insistence to celebrate the milestones in life, to Katie Priske for reminding me to read broadly, take breaks, and never pass up on a good dessert, to Zakir Durumeric for listening and not listening to me, and to Emily Hsieh for being such great company to navigate graduate school, a menu, a recipe or a trail.

I am so thankful for the Marshaks for (officially) welcoming me into their clan and for their boundless generosity, talent and love; you all have been an anchor through the highs and lows of graduate school. Thank you to Josh, Laurie, Jerome, and Rita taking care of everything while I set the groundwork for my dissertation—I can't believe my luck in having such a wonderful family. To Siva, my brother, I'm grateful for the trails you've blazed and especially for your encouragement to find my own path. To Josh, my husband, I'm so grateful for your steadfast support, love and company through this all.

Finally, I want to acknowledge my parents without whom I would not be where I am today. I cannot thank you enough for everything you have done to make this possible. And to my grandparents – thank you helping make sure that Siva and I had a great start in life. I am forever grateful for the sacrifices my family made to make this possible.

Chapter 1

INTRODUCTION

There is great diversity in organisms on the planet. This variation can be observed in the vast phenotypic diversity of organisms, the extreme environments they inhabit and the ways in which they reproduce. Underlying this variation are differences within genomes. Genomic variation can be driven by large changes in karyotype by whole genome or chromosome duplications, losses and translations of chromosomes or by large changes in the non-protein coding regions of the genome. Variation in genomes can also result in changes to preexisting non-coding (regulatory) and coding (open reading frame) sequences as well as through the origin of new genes. New genes originate as mutations within the genome of a germ-cell and once they arise, are subject to evolutionary forces such as mutation, genetic drift, and natural selection. Through these evolutionary processes, new genes can acquire beneficial properties, sweep through a population, and become fixed within a species, or, they can offer only transient benefit and eventually become inactive pseudogenes.

1.1 Mechanisms of gene birth

There are three major mechanisms by which genes can arise: (1) *de novo* origin from non-coding sequences (2) horizontal gene transfer (3) gene duplication and diversification (**figure 1.1**).

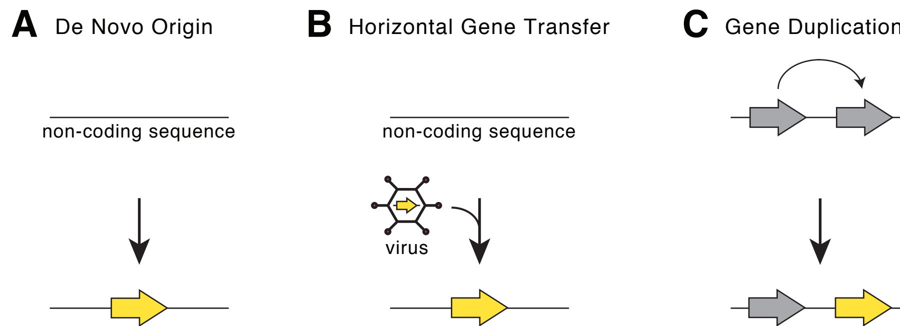


Figure 1.1 | Mechanisms for the origin of new genes. (A) A new gene (yellow) originates from previously non-coding sequence within genome (black line). **(B)** A pre-existing gene (yellow) within a virus is integrated into the host genome through horizontal gene transfer. **(C)** A new gene (yellow) arises from a segmental duplication of preexisting gene (gray).

A striking example of *de novo* origin of genes enabling survival in a new environment is observed in the northern gadid (codfish), which express an antifreeze glycoprotein (AFGP) that is crucial for their survival in frigid arctic waters. There are no homologs to AFGPs outside of northern gadids, strongly suggesting a *de novo* origin of these genes (Zhuang *et al.* 2019). Recently, Zhuang *et al.* compared the genomes of closely related gadids that contain and lack the AFGP to trace the origin of the AFGP coding sequence. They found that the northern gadid AFGP arose *de novo* from a non-coding element that underwent microsatellite duplications. AFGPs need to be secreted into the extracellular fluids in order to prevent ice crystal formation and in order to do so, they have to have a signal peptide. The AFGP in northern gadids acquired a signal peptide from the noncoding sequence directly upstream of the gene through a small

deletion of the intervening sequence. Finally, AFGP in northern gadids gained a *cis*-acting promoter (containing a TATA sequence) necessary for transcriptional initiation through a translocation event. In summary, AGFP provides a clear example of an adaptive protein that arose from previously non-coding DNA.

Some genes are not made from scratch but instead, come from other species. One mechanism by which genes are exchanged from one species to another is through horizontal gene transfer (HGT). HGT is the transfer of genetic material from one organism to another by a mechanism distinct from the transmission of genetic material from parent to offspring through reproduction (vertical transmission). In retroviruses, HGT between retroviruses and eukaryotes is facilitated by the fact that one of the most challenging parts of HGT, genome integration, is a part of the retroviral lifecycle. If a retrovirus infects a germ cell and integrates into the host genome, it can be vertically transmitted to the offspring. Such integrated retroviruses, if they lose their ability to infect other cells, become endogenous retroviruses.

Endogenous retroviruses are not always a burden on host genomes. Sometimes, their genetic material serves as substrates for evolution. For example, retroviral envelope proteins are key mediators of cell receptor recognition and binding, cellular fusion between the virus and host cell, and suppression of the host immune system. A notable example of host co-option of viral proteins occurred in the evolution of the placenta, a tissue that makes it possible for many species to give birth to live offspring. Host co-option of envelope proteins has leveraged envelope protein properties of cell fusion and immunosuppression for placental growth and development (reviewed in Denner 2016, Lavialle *et al.*, 2013). In fact, multiple independent HGTs of retroviral envelope proteins and placental co-option events occurred during mammalian evolution (Cornelis *et al.*, 2017).

Finally, some genes arise from preexisting genes through the process of gene duplication and diversification. Gene duplication can result through whole-genome or chromosome duplication, segmental DNA duplications, or through an RNA intermediate (retrotransposition). Whole-genome duplication (WGD) events lead to hundreds to thousands of duplicate genes. What is the fate of duplicate genes?

At the gene level, a duplication event of an essential gene relieves selective pressure on the gene so that one of the daughter genes can maintain the ancestral, essential function while the other daughter gene can sample new mutations. Through this process, daughter genes can evolve to have novel, essential functions distinct from the ancestral gene (neofunctionalization) (**figure 1.2A**). An example of a neofunctionalization event is described in depth later on in this chapter.

Alternatively, daughter genes can subfunctionalize (**figure 1.2B)**, a process by which they acquire mutations such that both daughters are required to perform the ancestral gene's function.

If the duplicate serves no adaptive function, it acquires mutations such that it can no longer produce a protein (pseudogenization **figure 1.2C**).

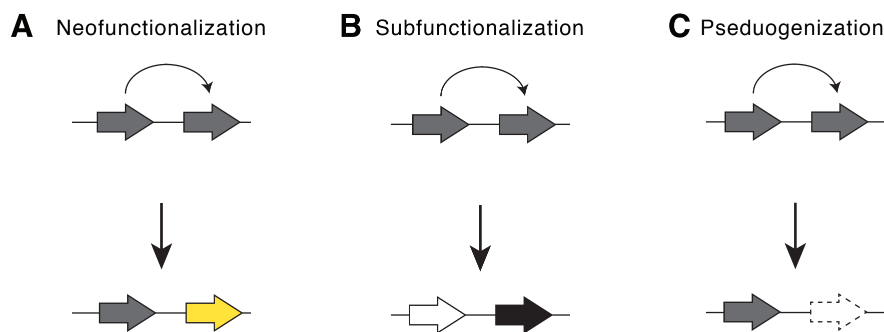


Figure 1.2 | Fates of duplicated genes. (A) Neofunctionalization occurs when a duplicated gene (gray) acquires a novel function (yellow) **(B)** When a gene subfunctionalizes after duplication, daughter genes (gray) acquire mutations in both daughter genes (black and white arrows) so that they both are required to perform ancestral function of gene **(C)** After duplication (gray genes), if a gene does not serve an adaptive advantage, it is lost through pseudogenization (white gene with dashed line).

Approximately 12% of WGD genes are retained in yeast after 80 million years (Kellis, Birren, & Lander, 2004), 15% of WGD genes are maintained in teleost fishes after 350 million years (Brunet *et al*., 2006), and 30% of WGD genes are preserved after 80 million years in *Arabidopsis* (Bowers, Chapman, Rong, & Paterson, 2003), suggesting that many duplicated genes can serve an adaptive advantage since they are not lost after duplication.

Studies in humans and *Drosophila* reveal that gene duplication is the primary mechanism by which most new genes arise (Zhang *et al*., 2010). Observations of gene duplications leading to morphological variation in *Drosophila* began to emerge in the early 1900s through careful genetic crosses by Sturtevant and Bridges on *Bar* mutants (Bridges, 1936; Sturtevant, 1925). The *Bar* mutation is X-linked so that hemizygous males and homozygous females carrying the mutant allele have slit-like eyes. Yet *Bar* mutant females sometimes produce progeny with wild-type eyes at a low frequency (52/85008) or progeny with a more extreme *Bar* eye phenotype (*ultra-Bar)* at a very low frequency of 3/85008 (Zeleny 1921). How are these altered phenotypes possible? Through careful genetic mapping experiments using adjacent markers, Sturtevant concluded that the *Bar* mutation was the result of a duplication event from unequal crossing over of the *Bar* gene during meiosis (Sturtevant, 1925). This led to the variability in phenotypes of *Bar* mutant female progeny. The work on *Bar* served as the basis for Muller's theory in 1936 that gene duplication can serve as the basis for new gene evolution; this concept was subsequently expanded by Ohno (Muller, 1936; Ohno, 1970).

Remarkably, the early work on gene duplication was elucidated without knowing the structure and sequence of genes but instead depended on careful cytogenetic

analyses and a bit of luck that morphological variation can be mapped to duplications of a single gene.

1.2 How to identify new genes

The advent and ease of whole genome sequencing as well as the accessibility of the resulting data has facilitated the identification and characterization of new genes. New genes are defined as genes that arose less than 30 million years ago in eukaryotes, and have consequently not undergone many mutations that obscure their early history (reviewed in Long *et al.*, 2013). Putative new genes can be identified by comparing syntenic regions of genomes of closely related species and looking for genes that are present in some species but not others (**figure 1.3**). What if the species lacking the putative new gene (C and D in **figure 1.3**) have the gene in another location? To investigate this possibility, one can search genomes for genes that are similar in

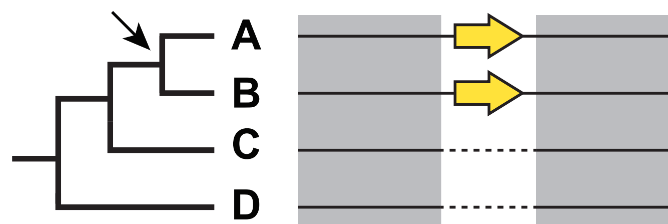


Figure 1.3 | Syntenic alignments facilitate the identification of new genes. A cladogram on the left shows relatedness of species A-D. An alignment of a locus across these species is represented by the black lines to the right of the cladogram, with the syntenic regions highlighted by the gray boxes. The dashed lines in species C and D represent regions of the alignment missing from these genomes. This syntenic alignment allows us to pinpoint the origin (black arrow) of a novel gene (yellow arrow) in species A and B.

sequence (putative orthologs) to the gene of interest across an entire genome. A phylogenetic tree can be generated from the multiple sequence alignment of putative orthologs to distinguish true orthologs, which will cluster in a monophyletic clade with gene of interest, from paralogs. Using a combination of phylogenetic and synteny analyses, one can pinpoint the origin of a new gene.

Before conducting time-intensive functional characterization of new genes, it is possible to predict whether these genes are functional based on transcriptomic and genomic information. For example, gene expression data can be analyzed to determine if a new gene is expressed; for a new gene to be functional, it must be expressed. In addition, comparing rates of polymorphisms/substitutions and divergence of new genes between closely related species can be indicative of natural selection acting on the retention or pseudogenization of a new gene. Subsequent characterization can help pinpoint the function of a new gene.

1.3 *Umbrea:* The stepwise evolution of a young, essential gene

Only a handful of evolutionarily young genes have been functionally characterized (Lee *et al.*, 2017; Long *et al.*, 2013; Ross *et al.*, 2013). The expectation is that evolutionarily young genes would serve novel, non-essential functions, since organisms are able to carry out highly conserved, essential processes (such as cell division) before the origin of the new gene. Surprisingly, many of these evolutionarily young genes become integrated in preexisting, essential processes (Chen *et al.*, 2010). How do these young genes become essential?

To answer this question, Ross *et al.* dissected Umbrea's stepwise acquisition of an essential role in cell division in *Drosophila* (Ross *et al.*, 2013). They found that *Umbrea* arose 10-15 million years ago in the ancestor of the melanogaster subgroup through a DNA-duplication of *heterochromatin protein 1B* (*HP1B*) (**figure 1.4a**). HP1B is preserved across *Drosophila*, enriched in heterochromatin, and regulates gene expression, but it is dispensable for viability (Zhang *et al.*, 2011) (**figure 1.4b)**. In contrast, Umbrea arose 10-15 mya and has been independently lost in at least three different *Drosophila* lineages (**figure 1.4c**); it localizes to centromeres and is essential for cell division and viability in *D. melanogaster* (Ross *et al.*, 2013) (**figure 1.4b)**. It is

estimated that *Umbrea* acquired its essential function 5-7 million years ago (after the most recent gene loss event). Ross *et al.* hypothesized that *Umbrea* became essential through the loss of the chromodomain and acquisition of centromere localization through non-synonymous changes within the chromoshadow domain. Ross *et al.* proposed that recurrent changes at centromeres led to the rise of a lineage-specific gene necessary for the conserved, essential process of cell division.

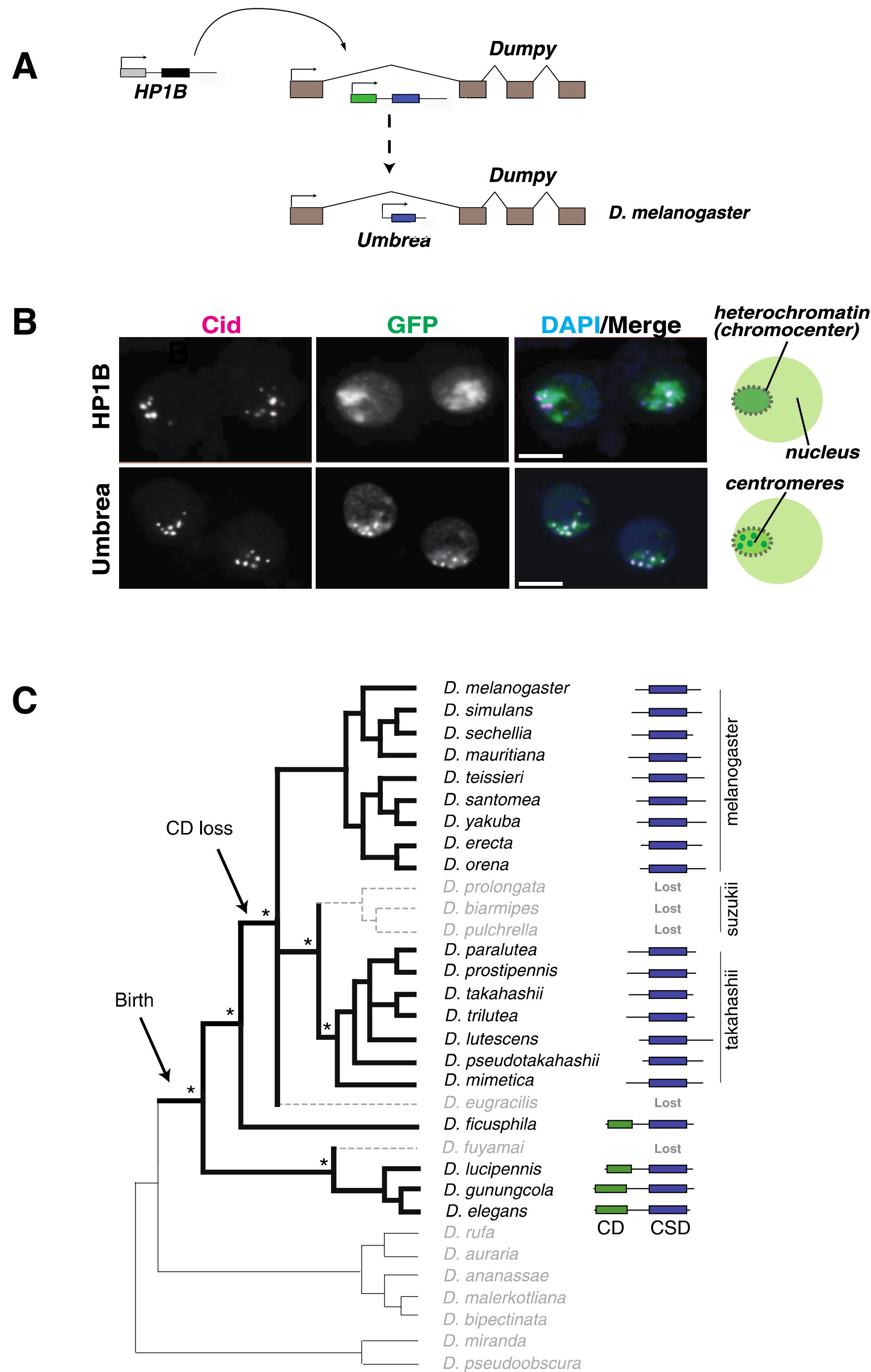


Figure 1.4 | Gene duplication and diversification produced *Umbrea.* (a) *HP1B* gene duplication into the *Dumpy* intron, loss of chromodomain (in green) and diversification of the shadow domain (in blue) led to the birth of *Umbrea.* **(b)** In *D. melanogaster* Kc cells, GFP-tagged HP1B localizes to the chromocenter, the heterochromatic compartment of the nucleus (centromeres, as defined by the centromeric histone H3 variant, CenH3 (centromere identifier or Cid) is marked in magenta and blue is 4',6-diamidino-2-phenylindole (DAPI) which marks DNA) while GFP-tagged Umbrea localizes to centromeres. **(c)** presence and structure of Umbrea genes summarized from polymerase chain reaction (PCR) of syntenic locus with chromodomain (CD, green) and chromoshadow domain (CSD) in blue. Asterisks denote branchpoints with significant bootstrap support and species lacking *Umbrea* are grayed out. *Umbrea* was lost at least three independent times since its origin. Adapted from figures 1 and 2 in (Ross *et al.* 2013).

1.4 Dissertation overview

For my thesis work, I focused on the origin and evolution of Zinc-finger Associated Domain Zinc finger proteins (ZAD-ZNFs), a large, dynamic class of transcription factors in *Drosophila*. These transcription factors, which arose in the ancestor of vertebrates and arthropods, have undergone numerous segmental DNA-duplications and losses within insect genomes (Chung *et al.*, 2002; Chung, *et al.*, 2007b). ZAD-ZNF family members can be lineage-specific and essential for development and fertility (Chen *et al.*, 2000; Gaszner *et al.*, 1999; Lake *et al.*, 2011; Niwa & Niwa, 2016; Page *et al.*, 2005; Uryu *et al.*, 2018). I reasoned that this gene family would provide an example to study how evolutionarily young genes can become essential.

In **chapter 2**, I examine the evolutionary dynamics of this transcription factor family in *D. melanogaster*. I found that surprisingly, preservation of *ZAD-ZNF* genes is not predictive of their essentiality. Instead, ZAD-ZNFs that are rapidly evolving are frequently essential in *D. melanogaster.* What could be driving the rapid evolution of essential ZAD-ZNFs? To answer this question, I conducted detailed evolutionary and functional characterizations of one locus containing five ZAD-ZNFs in *D. melanogaster,* two of which are rapidly evolving and essential. I found that these ZAD-ZNFs are 1) dynamically evolving across *Drosophila*, 2) show evidence for recombination and positive selection, and 3) localize to specific regions of heterochromatin.

In **chapter 3**, I conducted a detailed functional analysis of one evolutionarily young, essential ZAD-ZNF gene in this locus. I named this ZAD-ZNF Nick Nack. I found that *Nick Nack* is necessary for larval progression in *D. melanogaster,* and null larvae fail to activate expression of numerous genes required for L2 progression. In addition, arrested larvae express many heterochromatin-embedded elements. From my evolutionary analysis and functional characterization of ZAD-ZNFs, I propose that the

expansion and diversification of ZAD-ZNFs may be driven by their interaction with heterochromatin, a rapidly evolving compartment of the nucleus. Finally, in **chapter 4,** I discuss future perspectives on ZAD-ZNF research, with an emphasis on the possible roles of these genes in development and evolution.

Chapter 2

PHYLOGENOMICS OF *DROSOPHILA* ZAD-ZNFs

2.1 Introduction to ZAD-ZNFs

Although biological organisms display enormous phenotypic diversity, their cellular organization and early development are frequently conserved across broad taxonomic ranges. Such widespread conservation led to a commonly-held idea that fundamental biological functions are encoded by a conserved, ancient genetic architecture. However, a recent study of 185 new genes in Drosophilids demonstrated that 30% of these young genes (Chen *et al*., 2010) have acquired roles in development, cell biology, and reproduction that render them essential for viability or fertility (Lee *et al*., 2017; Long *et al*., 2013; Ross *et al*., 2013). Sometimes, wholesale evolutionary turnover of genes underlying essential cellular processes, such as kinetochore function, can occur (Drinnenberg *et al*., 2016). Furthermore, even when they are retained over large evolutionary periods, such as the foundational histone variant (CenH3) at centromeres, genes encoding essential functions can evolve unexpectedly rapidly across plants and animals (Malik & Henikoff 2001, Talbert *et al*., 2002). Thus, genes that are either rapidly evolving or subject to high genetic turnover encode at least a subset of essential functions.

Rapidly evolving genes can participate in cellular processes that require constant adaptation for essential functions. For example, genes subject to intragenomic conflicts require constant evolutionary adaptation to maintain cellular

processes essential for viability or fertility. Heterochromatin functions are an example of such essential functions that require constant adaptation. Although heterochromatin is a gene-poor component of most eukaryotic genomes, its establishment and maintenance is nevertheless essential for many cellular processes including chromosome condensation and segregation, repression of transposable elements (TEs) embedded in heterochromatin or euchromatin, and for genome stability (Abe *et al.*, 2016; Azzaz *et al.*, 2014; Brennecke *et al.*, 2007; Goriaux *et al.*, 2014a; Goriaux *et al.*, 2014b; Grézy *et al.*, 2016; Inoue *et al.*, 2008; Levine *et al.*, 2015; Liu *et al.*, 2014; Nambiar & Smith, 2018; Okita *et al.*, 2019; Ruiz-Estévez *et al.*, 2014; Senti & Brennecke, 2010; Vernì & Cenci, 2015; Verschure *et al.*, 2005). Indeed, there are many genes that encode protein components enriched in heterochromatin and are essential for viability and/or fertility in organisms including *Drosophila melanogaster,* yet such genes are quite variable even among closely-related species (Klattenhoff *et al.*, 2009; Levine *et al.*, 2015; Levine *et al.*, 2016; Parhad *et al.*, 2017; Ross *et al.*, 2013; Vermaak & Malik, 2009; Vermaak *et al.*, 2005). It has been suggested that this variability reflects both lineage-specific mechanisms to package and silence heterochromatic DNA, as well as constant adaptation to maintain this silencing, triggered by, for example, TE escape mutations.

A variable, large family of DNA binding elements in insects is the ZAD-ZNFs. ZAD-ZNF proteins contain a conserved N-terminal ZAD (Zinc-finger associated domain), a linker, and a C-terminal domain that includes tandem C_2H_2 zinc fingers (Chung *et al.*, 2002; Lespinet *et al.*, 2002) (**figure 2.1)**. The ZAD facilitates protein-protein interactions and does not have DNA-binding ability while the C_2H_2 zinc fingers mediate sequence specific DNA binding (Jauch *et al.*, 2003). ZAD-ZNF genes arose in the ancestor of vertebrates and arthropods, and dramatically expanded within insect

lineages (Chung *et al.*, 2007b), becoming the most abundant class of TFs (by number of genes) in many genomes including in *D. melanogaster* (Chung *et al.*, 2002). Intriguingly, ZAD-ZNF gene repertoires can vary quite extensively across insect lineages (Chung *et al.*, 2002; 2007b), as characterized by whole gene gains and losses from the genome, but the cause of this dynamism is not understood.

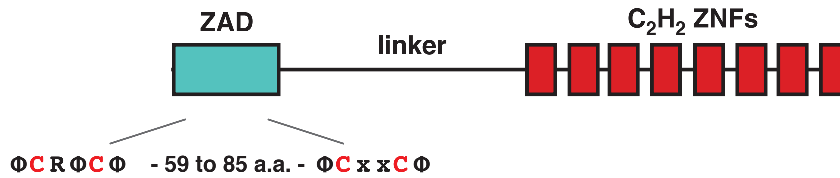


Figure 2.1 | ZAD-ZNF Structure. ZAD domain (in blue) is characterized by conserved cysteines ('C' in red) flanked by hydrophobic residues (denoted by '**φ**'). An unstructured linker connects the C_2H_2 ZNF domains to the ZAD domain.

Despite the abundance of ZAD-ZNFs in insect genomes, only a handful of studies have characterized the function of these proteins. These studies reveal that approximately half of *ZAD-ZNF* genes are highly expressed in the ovary and early embryo and play essential roles in fertility or development by promoting the expression of other essential genes (Chung *et al.*, 2002). Biochemical characterization of 21 ZAD-ZNF proteins in *D. melanogaster* using electromobility shift assays (EMSAs) show that they all bind to distinct DNA consensus sequences, with putative targets in the regulatory regions of specific target genes (Krystel & Ayyanathan, 2013). For example, the ZAD-ZNF Motif 1 Binding Protein (M1BP) binds core promoters and promotes expression of numerous housekeeping genes (Baumann *et al.*, 2017; Li *et al.*, 2013) while the ZAD-ZNFs Molting Defective, Ouija Board, and Séance help promote the transcription of heterochromatin-embedded genes, *Spookier* and *Neverland,* which are required for larval progression (Uryu *et al.*, 2018). Similarly, ZAD-ZNF Grauzone promotes *cortex* expression and is necessary for meiotic progression during *D.*

melanogaster oogenesis (Chen *et al.*, 2000; Harms *et al.*, 2000; Page & Orr-Weaver, 1996).

ZAD-ZNFs can also play roles in genome architecture that are distinct from transcription. For example, the ZAD-ZNFs ZIPIC, Zw5, and Pita play a role in chromatin structure in the *D. melanogaster* genome (Gaszner *et al.*, 1999; Zolotarev *et al.*, 2016) whereas the ZAD-ZNF Trade Embargo binds chromatin and initiates meiotic recombination during *D. melanogaster* oogenesis (Lake *et al.*, 2011). Furthermore, the ZAD-ZNF Oddjob directly interacts with HP1A, dynamically interacts with heterochromatin, and is a suppressor of variegation (Swenson *et al.*, 2016) while its closely related paralog, CG17801, helps repress *HetA* and *Blood* transposable elements in the ovary (Czech *et al.*, 2013). Finally, some ZAD-ZNFs might not function in the nucleus at all. For example, even though the Weckle ZAD-ZNF protein possesses C_2H_2 zinc finger domains, it localizes to the plasma membrane instead of nuclear chromatin, where it interacts with the Toll-DmMy88 complex to help establish the anterior-posterior axis of the developing embryo (Chen *et al.*, 2006).

Here, I took advantage of extensive opportunities in *D. melanogaster* and related species for phylogenomic and population genetics studies, genome-wide functional screens, and tools for cytological and genetic analyses to investigate the relationship between genetic innovation and essentiality within the ZAD-ZNF gene family. Unexpectedly, I find that even though a subset of ZAD-ZNF genes have undergone positive selection, they are more likely to be required for viability or fertility in *D. melanogaster* than genes that are not rapidly evolving. Focusing on one cluster of ZAD-ZNF genes that includes *Oddjob*, I find evidence for changes in ZAD-ZNF repertoires across *Drosophila* species. From my functional and cytological analyses, I suggest that

changing requirements for heterochromatin function likely drove essential innovation of some ZAD-ZNF proteins in *Drosophila*.

2.2 Results

2.2.1 ZAD-ZNFs are a dynamic and diverse protein family in Drosophila

I searched the Pfam database (Pfam.org) to identify all genes in *D. melanogaster* that had a ZAD domain (PF07776). I found 90 ZAD-ZNFs distributed across Chromosomes 2, 3 and X in the *D. melanogaster* genome. 31 of 90 *D. melanogaster* ZAD-ZNFs occur in 13 gene clusters, which I defined as containing two or more tandemly arrayed ZAD-ZNF genes. Many evolutionarily young ZAD-ZNF genes are present within these clusters. Of these, many share intron/exon structures with their neighbors and presumed evolutionary ancestors; it is likely that these genes arose via segmental duplication. In contrast, seven ZAD-ZNFs (*CG3032, CG4318, CG9215, CG44002, CG17361, CG7963, CG17359*) lack introns that are found in their closest relatives and, therefore, either arose via retrotransposition or via gene duplication from intron-less ancestors.

Further analysis of ZAD-ZNFs using NCBI's Conserved Domain Database revealed that some ZAD-containing proteins have no C_2H_2 domains (*dlip, dbr, CG15435, CG31109, CG31457*) while others contain up to 23 C_2H_2 domains (*CG11902*). The average ZAD-containing protein has 6.2 C_2H_2 domains. In addition to the ZAD in the N-terminus, 13 ZAD-ZNF

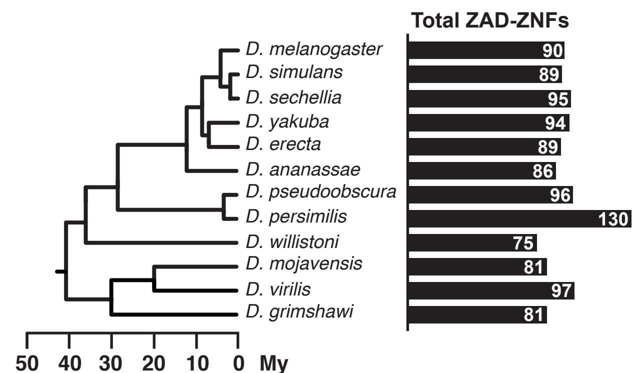


Figure 2.2 | Total number of ZAD-ZNFs across Drosophila. Phylogeny of 12 Drosophila genomes with a scale bar approximating the divergence times. Number of ZAD-containing genes within each genome indicated with the black bars.

proteins possess another N-terminal domain such as a Sodium-Calcium exchanger domain (*CG12391*) or ASF1 histone chaperone domain (*CG10321*).

I next expanded the Pfam survey to all 12 previously sequenced and annotated *Drosophila* genomes. These species represent a range of evolutionary divergences from *D. melanogaster*, from just a few million years (*e.g., D. simulans*) to more than 40 million years (*e.g., D. virilis*) (**figure 2.2**) (*Drosophila* 12 Genomes Consortium *et al.*, 2007). This analysis revealed a surprisingly wide range in the size of the ZAD-ZNF repertoire across different *Drosophila* species (**figure 2.2**). For example, I found that the *D. persimilis* genome encodes 130 ZAD-ZNF genes whereas the *D. willistoni* genome encodes only 75 ZAD-ZNF genes. Such analyses are dependent on the state of completion and annotation of individual *Drosophila* species' genomes, so this number may be an underestimate of the total number of ZAD-ZNFs in each genome. Therefore, to complement these analyses, I assessed the apparent age of each of the *D. melanogaster* ZAD-ZNF genes by examining which of the 12 annotated genomes of *Drosophila* species have orthologs (flybase.org). I found that 73 of 90 *D. melanogaster* ZAD-ZNF genes arose in the ancestor of *D. melanogaster* and *D. virilis* and estimate that they are at least 40 million years old. Since these 73 ZAD-ZNFs first arose, 12 have been lost in at least one lineage. Finally, I inferred that at least 3 ZAD-ZNF genes found in *D. melanogaster (CG4318, CG17612, neu2)* originated via gene duplication less than 10 million years ago. My findings corroborate previous large-scale surveys that identified rapid changes in ZAD-ZNF gene repertoires within insect genomes (Chung *et al.*, 2002; 2007b; Lespinet *et al.*, 2002).

2.2.2 Rapidly evolving ZAD-ZNFs are frequently essential in D. melanogaster

The rapid changes in the ZAD-ZNF repertoire among *Drosophila* (and other insect) species suggested that selection might favor their genetic innovation. To

further explore this possibility, I investigated the *D. melanogaster* ZAD-ZNF genes for signatures of recent positive selection. For this analysis, I performed McDonald-Kreitman tests of all ZAD-ZNF genes that are present in both *D. melanogaster* and the closely related species, *D. simulans*. One of the 90 *D. melanogaster ZAD-ZNF* genes, *CG2202* is absent in *D. simulans*, so the analyses were performed on 89 ZAD-ZNF pairs of orthologs as defined by FlyBase. The McDonald-Kreitman test (McDonald and Kreitman 1991) compares the ratio of non-synonymous (amino-acid altering, or replacement) to synonymous substitutions fixed during the divergence of the two species (Dn: Ds), to that of non-synonymous to synonymous polymorphisms within a species (Pn: Ps). Since Pn: Ps is a proxy for functional constraint acting on a gene, this ratio is expected to be similar within and between species, i.e., the Dn: Ds ratio should be similar to the Pn: Ps ratio. A higher than expected number of fixed non-synonymous changes would indicate the action of adaptive evolution during the species divergence (McDonald & Kreitman, 1991).

I took advantage of previous efforts that sequenced the genomes of hundreds of *D. melanogaster* strains and a few *D. simulans* strains (Begun *et al.*, 2007; Lack *et al.* 2016; Langley *et al.*, 2012) to perform the McDonald-Kreitman test using the Popfly server (popfly.uab.cat) (Hervas *et al.*, 2017). I found that 13 ZAD-ZNFs show evidence for recent adaptive evolution, *i.e.,* have an excess of fixed non-synonymous changes (**table 2.1**). In 5 of these cases, due to inadequate number of polymorphisms, I was unable to delineate whether the positive selection occurred in the ZAD, the linker, or the C_2H_2 zinc finger domains. In 6 of the remaining nine genes, I found that the linker domain, which varies both in size and sequence between ZAD-ZNF paralogs, is evolving under positive selection. One gene (*CG7386*) showed signatures of adaptive evolution in its ZAD domain whereas three genes (*CG2712, CG7386, CG10321*) showed evidence

of adaptive evolution in the Zinc-finger domain. The findings implicate the poorly characterized linker region as being primarily responsible for the signatures of positive selection and highlight that the linker region may be important to the essential function of ZAD-ZNFs.

Gene	McDonald Kreitman Test					Neutrality Index (NI)	Orthologs	Mutant Phenotype
	p-value	D_N	D_S	P_N	P_S			
Nnk (CG17802)	**0.007***	52	24	14	20	0.32	mel, sim, sec, ere, yak, pse, per	lethal[1]
ZAD	0.12	6	2	2	4	0.83		
linker	**0.019**	34	10	9	10	0.26		
C_2H_2	0.090	6	11	0	6	0		
Odj (CG7357)	**0***	40	32	5	25	0.16	mel, sim, sec, ere, yak, ana, pse, per, wil, vir, gri	lethal[1]
ZAD	0.20	4	4	0	2	0		
linker	0.052	18	9	4	8	0.25		
C_2H_2	0.10	2	7	0	11	0		
Trem (CG4413)	**0.014***	30	29	3	14	0.21	mel, sim, sec, ere, ana, pse, per, wil, vir, moj, gri	female sterile[2]
ZAD	0.57	2	6	0	1	0		
linker	**0.036***	18	13	2	8	0.18		
C_2H_2	0.24	3	10	0	5	0		
Zw5 (CG2711)	**0.003***	55	51	6	22	0.25	mel, sim, sec, ere, yak, ana, pse, per, wil, vir, moj, gri	lethal[3]
ZAD	0.124	3	4	0	4	0		
linker	0.251	32	17	3	4	0.40		
C_2H_2	0.059	12	25	0	8	0		
CG2712	**0.004***	94	51	16	24	0.36	mel, sim, sec, ere, yak, ana, pse, per, vir, moj, gri	viable[4]
ZAD	0.161	10	6	3	6	0.30		
linker	0.088	45	14	10	8	0.39		
C_2H_2	**0.038***	15	14	2	10	0.19		
D19B (CG10270)	**0.004***	11	42	3	66	0.17	mel, sim, sec, yak, ana, pse, wil	viable[4]
ZAD	0.121	2	4	0	6	0		
linker	0.181	6	12	2	13	0.31		
C_2H_2-1	null	0	6	0	25	null		
C_2H_2-2	null	0	2	0	7	null		
CG7386	**0.002***	78	38	28	36	0.38	mel, sim, sec, ere, yak, ana, pse, wil, vir, moj, gri	early pupal lethal[4]
ZAD	**0.032***	3	6	3	5	1.2		
linker	0.118	12	5	13	1	5.4		
C_2H_2-1	null	0	0	0	1	n/a		
C_2H_2-2	**0.005***	14	7	5	16	0.16		
C_2H_2-2	0.097	2	6	3	1	9		
CG17359	**0***	43	17	3	14	0.084	mel, sim, sec, ere, yak, ana	lethal[4]
ZAD	0.157	3	1	1	3	0.11		
linker	**0***	28	4	1	5	0.028		
C_2H_2	0.064	10	10	0	4	0		
wek (CG4148)	**0.001***	29	42	3	30	0.144	mel, sim, sec, ere, yak, ana, pse, per, wil, vir, moj, gri	lethal, maternal effect lethal[5]
ZAD	0.205	2	3	0	3	0		
linker	**0.011***	22	19	3	14	0.185		
C_2H_2	0.073	5	17	0	12	0		
CG10321	**0***	28	50	9	63	0.255	mel, sim, sec, ere, yak, ana, pse, per, wil, vir, moj, gri	viable[4]
ZAD	null	0	2	0	4	null		
linker	**0.008***	24	38	8	42	0.301		
C_2H_2	**0.023***	3	7	0	15	0		
mld (CG34100)	**0.020***	84	70	37	57	0.54	mel, sim, sec, ere, yak, ana, pse, per, wil, vir, moj, gri	lethal[6]
ZAD	null	0	0	0	0	n/a		
linker	0.28	63	51	30	34	0.71		
linker-2	0.075	6	8	2	13	0.21		
C_2H_2-1	0.24	4	5	0	2	0		
C_2H_2-2	0.43	1	1	1	4	0.25		
CG4282	**0.029***	13	32	2	25	0.20	mel, sim, sec, ere, yak, ana, pse, per, wil vir, moj, gri	viabile[4]
ZAD	0.053	2	0	1	4	0		
linker	0.25	9	18	1	7	0.29		
C_2H_2	null	0	11	0	14	null		
CG8159	**0***	17	35	13	2	13.4	mel, sim, sec, ere, yak, ana, pse, per, wil, vir, moj, gri	viable[4]
ZAD	0.37	7	3	2	0	null		
linker	**0.001***	5	12	7	0	null		
C_2H_2	null	0	18	0	1	null		

Table 2.1 | 13 rapidly evolving ZAD-ZNFs in *D. melanogaster*. Summary statistics for the McDonald-Kreitman (MK) test with statistically significant values ($p<0.05$) denoted with an asterisk. Comparisons within *D. melanogaster* for synonymous (Ps) and non-synonymous (Pn) are made to the rate of fixed synonymous (Ds) and non-synonymous (Dn) changes between species. The neutrality index (N.I.) <1 is suggestive of excesses fixed non-synonymous changes between species. Orthologs column indicates species that contain an ortholog as determined by OrthoDB. The phenotype column reports phenotypes as follows: [1]this study, [2]Lake et al., 2011, [3]Gaszner et al., 1999, [4]flybase, [5]Chen et al., 2006, [6]Neubueser et al., 2005.

Using this signature of positive selection in a subset of ZAD-ZNF genes, I tested whether evolutionarily constrained genes are more likely to encode essential functions. I took advantage of the fact that knockdown or knockout phenotypes have been phenotypically characterized in almost genome-wide scale in *D. melanogaster,* revealing phenotypic outcomes in the case of 84 ZAD-ZNF genes. Of these, knockdown or knockout of 30 ZAD-ZNF genes showed complete lethality or sterility in *D. melanogaster*, whereas the other 54 were not essential. Intriguingly, I found that 9 of 13 genes that have evolved under positive selection are essential, whereas only 22 of the remaining 71 genes are essential. Thus, remarkably, ZAD-ZNF genes that are subject to positive selection are more likely to encode proteins with functions essential for viability or fertility (9:5 versus 22:49, Chi-squared test, $p<0.05$).

I next addressed whether evolutionary retention is also a predictor of essentiality. Again, I restricted my analyses only to those ZAD-ZNF genes for which I had reliable phenotypic information in *D. melanogaster*. I found that 17 of the 61 ZAD-ZNF genes conserved in all 12 annotated *Drosophila* species' genomes are essential (three of these genes have no phenotypic data available). In comparison, I found that 11 of the 29 genes not universally conserved in *Drosophila* species are essential (one of these genes has no phenotypic data available). Thus, genes not globally retained over *Drosophila* evolution are just as likely to encode a necessary function in *D. melanogaster* as genes that have been strictly maintained over 40 million years of *Drosophila* evolution (11:18 versus 17:44, $p>0.2$). The findings further support the idea that many crucial roles in organisms like *D. melanogaster* are encoded by dynamic, rapidly evolving genes, rather than by evolutionarily constrained genes.

2.2.3 The Odj-Nnk cluster of ZAD-ZNF genes is dynamically evolving in Drosophila

To gain further insight into the biological basis for this unexpected correlation, I focused on the evolution and function of a cluster of five ZAD-ZNF genes on *D. melanogaster* chromosome 3. This cluster includes two of the 13 positively-selected, essential ZAD-ZNF genes – *Oddjob (CG7357),* which Swenson *et al.* (2016) named for a James Bond henchman who silenced people, and *CG17802*, which I name *Nick Nack,* or *Nnk,* in keeping with the 'James Bond henchmen' theme. The cluster also includes three genes that do not evolve under positive selection *(CG17801, CG17806, and CG17803)* (**figure 2.3A**). Among these five genes, *Oddjob* has been best characterized functionally. Oddjob protein interacts with heterochromatin protein HP1A, localizes to heterochromatin dynamically, and acts as a suppressor of position effect variegation (Swenson *et al.*, 2016). *Odj* knockdown or overexpression is lethal in *D. melanogaster* (Schertel *et al.*, 2015). *CG17801* germline knockdown is associated with de-repression of *HetA* and *Blood* transposable elements in the ovary but does not affect fertility or viability (Czech *et al.*, 2013). Although a genome-wide screen suggested that *Nick Nack* is essential for viability (Chen *et al.*, 2010), subsequent elucidation of problems associated with the VDRC RNAi KK lines (Green *et al.*, 2014; Vissers *et al.*, 2016), which had a second site mutation associated with lethality. Further analyses (Kondo *et al.*, 2017) using CRISPR mutants have called the initial finding, that *Nick Nack* is essential, into question. Outside of genome-wide screens, the roles of the other two ZAD-ZNF genes in this cluster (*CG17801, CG17803*) have not been not studied.

The proximity of these genes in the *D. melanogaster* genome prompted us to examine their phylogenetic relationships to infer the evolutionary dynamics of the *Odj-Nnk* cluster in *Drosophila* species. I first identified orthologs to *Oddjob*-cluster members in other *Drosophila* species using reciprocal tBLASTn searches with

individual *D. melanogaster Oddjob*-cluster members as queries. I searched both the originally-sequenced, well-annotated 12 *Drosophila* genomes (*Drosophila* 12 Genomes Consortium *et al*., 2007), as well as eight additional genomes that were subsequently sequenced, to sample the melanogaster group within the Sophopora subgenus more densely (Chen *et al*., 2014). In several cases, technical issues prevented confident assignment of the tBLASTn hits to orthologous groups: the gene matched closely to more than one *D. melanogaster* gene; there were several putative hits within a single genome; or the hit contained either the ZAD or zinc-finger domain only.

I next generated a multiple alignment of the amino acid sequence of all *Odj-Nnk* cluster members that contained both ZAD and Zinc-finger domains, ignoring orthologs that were missing either of these two domains. I found that there is considerably less conservation in the rapidly evolving linker region of *Odj-Nnk* cluster orthologs and therefore trimmed poorly aligned segments of the alignment for all the genes in the alignment. I then performed phylogenetic analyses using maximum likelihood methods. Through this analysis, I was able to conclude that all *Odj-Nnk* cluster genes belong to one of five monophyletic groups that each contains one ortholog present in *D. melanogaster* (**figure 2.3B**). I was able to subsequently assign each of the single-domain 'hits' I had previously identified to each of these orthology groups via phylogenies based on single-domains, although these analyses lacked the same phylogenetic resolution as the full alignment. Furthermore, upon evaluating the multiple alignments of each orthologous group for signatures of recombination using GARD analysis on the Datamonkey server (Kosakovsky *et al*., 2006), I found that *CG17806, Oddjob,* and *Nick-nack* showed evidence for recombination within the linker region whereas *CG17803* and *CG17801* did not. The ability to discern orthology groups

in spite of this recombination stems from the fact that the predominant phylogenetic signal derives from the longer, C-terminal Zinc-finger domains.

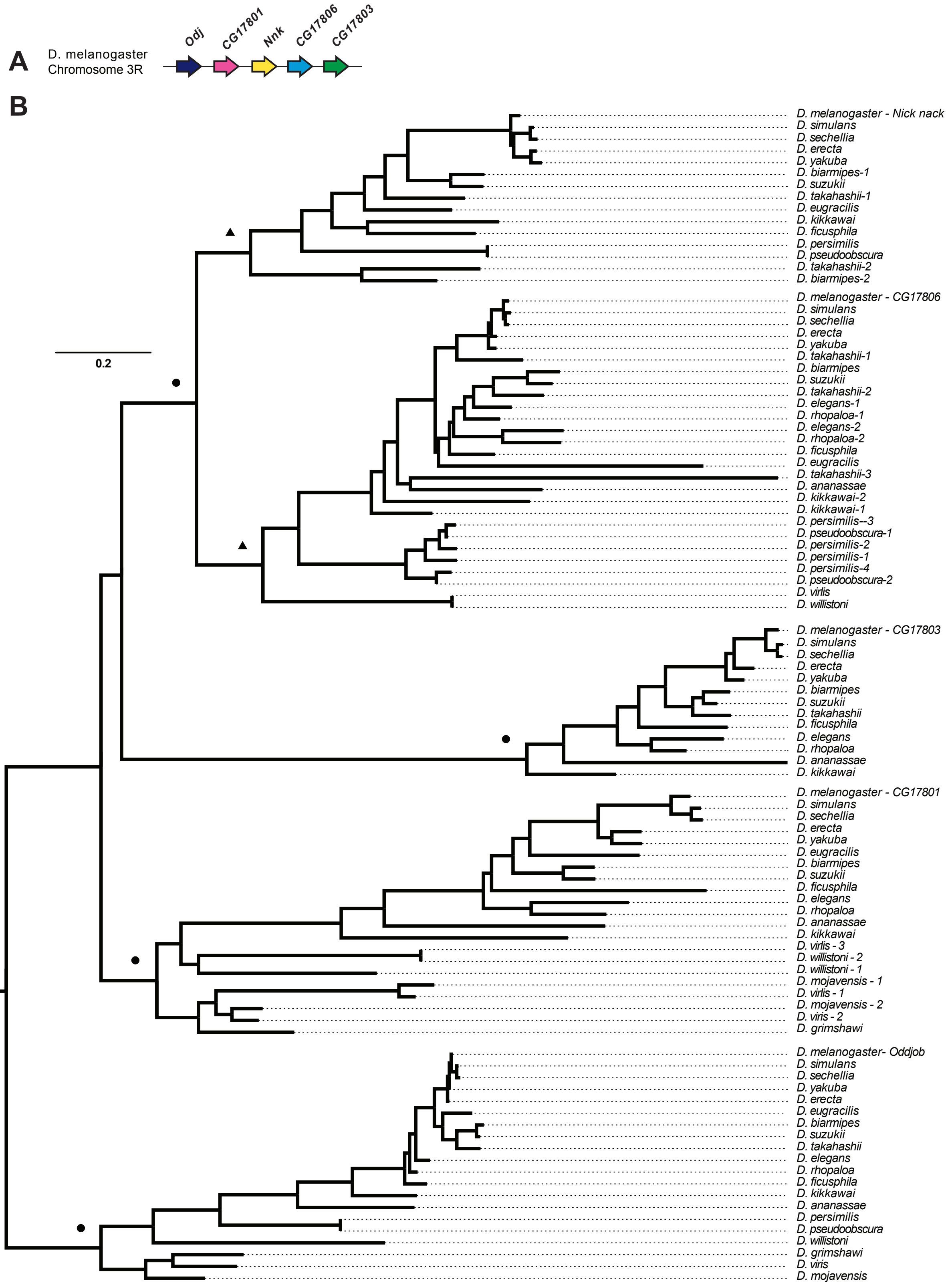


Figure 2.3 | Evolutionary relationship among all Drosophila *Oddjob* paralogs. (A) The *Oddjob* cluster in *D. melanogaster*. **(B)** We used PhyML to perform maximum likelihood phylogenetic analyses with a protein alignment of the ZAD and zinc finger domains. We find that all paralogs form a well-supported monophyletic clade, suggesting a single origin for each of these *Oddjob* locus gene. Bootstrap support at > 70 shown in black circles and >40 in black triangles at major nodes. This tree is rooted on the *Oddjob* clade. The scale bar represents the number of substitutions per residue.

The phylogenetic analyses reveal that *Odj-Nnk* cluster evolution has been highly dynamic during the evolution of the *Drosophila* genus (**figure 2.5**). I found no orthologs to the *Odj-Nnk* cluster genes outside *Drosophila* (queried genomes were *Aedes aegypti, Armadillidium vulgare, Bactrocera dorsalis, ceratitis capitate)* using the same reciprocal tBLASTn and phylogenetics approach described above. While *Oddjob* orthologs are present in all fly genomes analyzed, representing 40 million years of *Drosophila* evolution, no other genes in the cluster are universally conserved. For example, although *CG17801* dates back to the origin of *Drosophila*, it has been lost in the obscura clade and in *D. takahashii*. Similarly, *CG17806* arose before the birth of the *Drosophila* genus but subsequently underwent multiple independent duplication and loss events. *Nnk* appears to have emerged more recently in the ancestor of *D. melanogaster* and *D. pseudoobscura* (~30 million years ago) but also experienced multiple independent duplications and losses; I note that the age estimates for *Nnk* are higher than those reported previously based on fewer sequenced species (Chen, *et al.* 2010). *CG17803* arose in the ancestor of *D. melanogaster* and *D. ananassae* but underwent at least one independent loss, in *D. eugracilis.* Based on the evidence for recent adaptive evolution of genes within the cluster and the recurrent gene gains and losses, I find that the *Odj-Nnk* cluster is dynamically evolving in *Drosophila*.

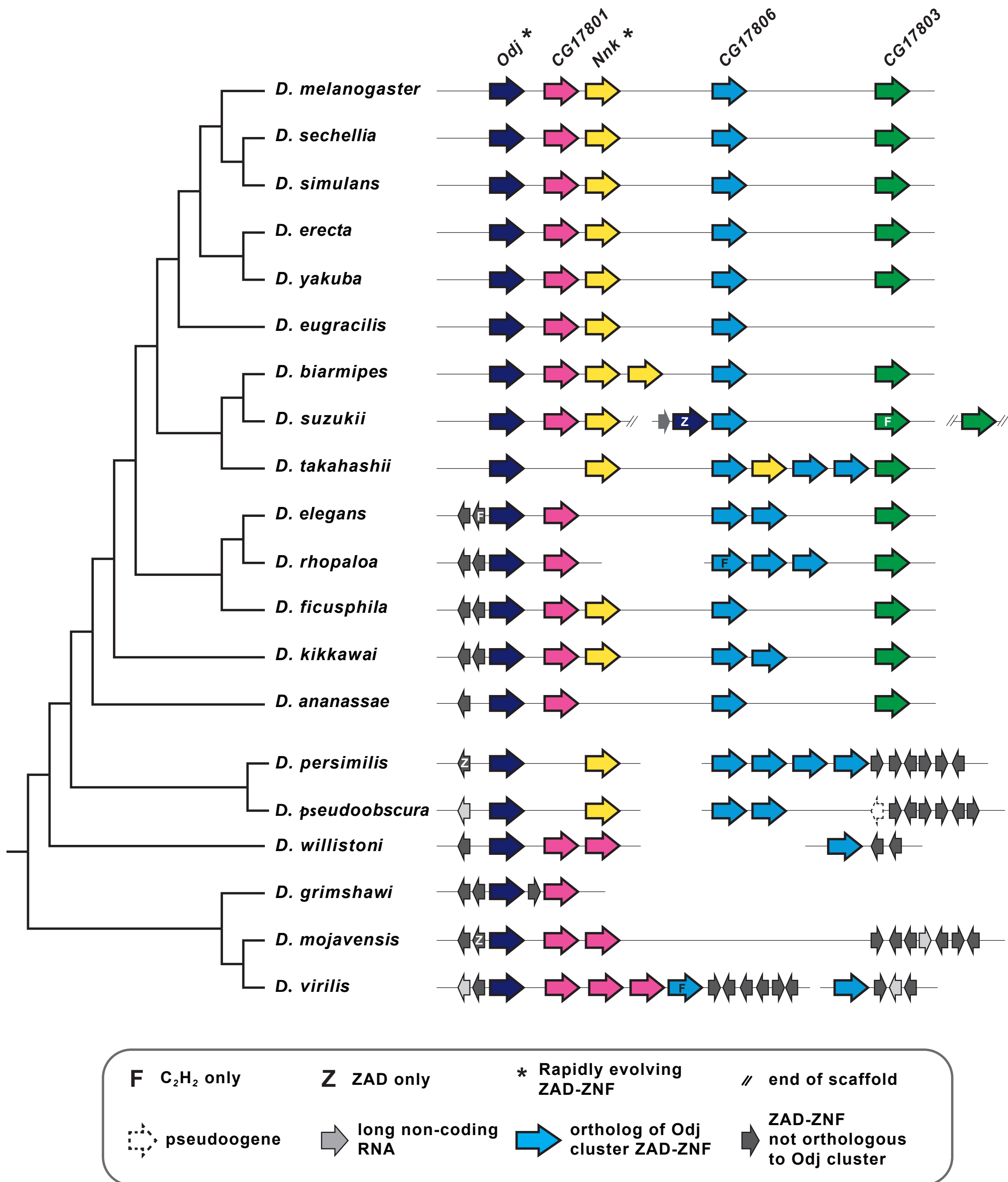


Figure 2.4 | *Odj-Nnk* cluster genes are rapidly gained and lost in Drosophila. *Oddjob* is the only gene within the cluster found in all queried Drosophila species. Both *CG17801* and *CG17806* dates back to the origin of Drosophila but has been lost within the obscura clade and in *D. takahashii*, while *CG17806* has undergone numerous gains and losses since it first arose. *Nick Nack* arose in the ancestor of *D. melanogaster* and *D. pseudoobscura* and has been lost in *D. ananassae* and the *D. elegans/D. rhopaloa* lineage as well as a duplication event in the genomes of *D. biarmipes* and *D. takahashii*. *CG17803* arose in the ancestor of *D. melanogaster* and *D. ananassae* and has been lost once in *D. eugracilis.*

I next evaluated whether we could identify signatures of recurrent positive selection in any members of the *Odj-Nnk* gene cluster. I used maximum likelihood analyses using the PAML and Datamonkey suite of programs to infer positive selection. I accounted for recombination within a locus, which can conflate positive selection studies. In cases where recombination was detected (*CG17806, Odj,* and *Nnk),* I performed analyses on each segment separately (**figure 2.5A**). The analyses revealed recurrent positive selection only in one case- *Nnk*- at position 114 in the linker region (**figure 2.5B**), which I also identified as having evolved under positive selection between *D. melanogaster* and *D. simulans* (**table 2.1**).

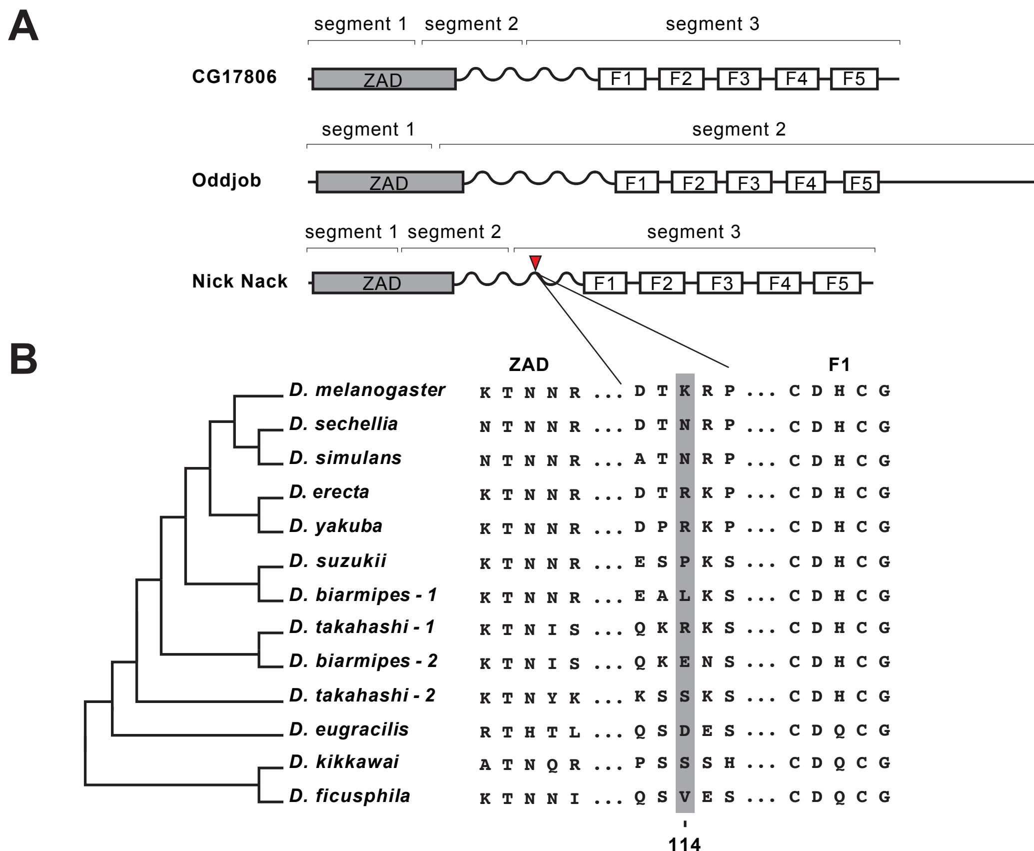


Figure 2.5 | Recombination and positive selection analyses. (A) schematic of CG17806, Oddjob and Nick Nack (where 'F1" is the first zinc finger domain) with significant evidence for breakpoint(s) and the most likely location of these breakpoints. Each segment represents a region that did not undergo recombination **(B)** A gene tree of Nnk and the corresponding protein alignment showing well-aligned regions (ZAD and F1) as well as rapidly evolving linker region with site showing recurrent, site specific positive selection, position 114 highlighted in gray.

Gene	MK Test		GARD	PAML				
	p value	N.I.	number of breakpoints	Segment	Alignment Length	M7 vs M8 p-value	M8 vs M8a p-value	M8 sites (BEB PP > 0.95)
Oddjob	**0***	0.169	1	1	186	1	1	-
				2	912	0.24	1	-
CG17801	0.522	0.774	0	-	741	0.71	1	-
Nnk	**0.008***	0.332	2	1	138	0.81	0.52	-
				2	156	1	1	-
				3	564	**9.8E-04***	0.053	**11 R**
CG17806	0.248	1.209	2	1	165	0.36	0.16	-
				2	156	0.48	0.36	-
				3	564	0.43	0.11	-
CG17803	0.269	0.656	0	-	1335	0.38	0.24	-

Table 2.2 | Oddjob cluster ZAD-ZNFs evolve under different pressures. Summary of positive selection tests conducted on each *Oddjob* cluster gene. Statistically significant test results (P<0.05) are marked with an asterisk. For the McDonald-Kreitman (MK) test, Neutrality Index (N.I.) <1 is suggestive of positive selection as it indicates an excess of non-synonymous fixed differences between species. Sites with a Bayes Empirical Bayes (BEB) posterior probability > 0.95 are listed in the last column as sites we have high confidence to be rapidly evolving.

2.2.4 Four Oddjob cluster ZAD-ZNFs localize to heterochromatin

Given the genomic and phylogenetic proximity of the *Odj-Nnk* cluster members, I next examined whether they share functional properties, starting with their sub-cellular localization. In *D. melanogaster* cells, Oddjob localizes to heterochromatin in *D. melanogaster* Schneider 2 (S2) cells under the control of a *copia* promoter (Swenson *et al*., 2016). I investigated whether the heterochromatic localization is a shared feature of proteins encoded by the *Odj-Nnk* ZAD-ZNF gene cluster. Using transient transfection of epitope-tagged *Odj-Nnk* cluster members in *D. melanogaster* Schneider 2 (S2) cells, I found that 4 of the 5 ZAD-ZNFs in the cluster encode proteins that localize specifically to heterochromatin. Consistent with a previous study by Swenson *et al*., I found that Oddjob has a broad heterochromatin localization pattern. Nick Nack and CG17803 proteins localize to foci within heterochromatin, which are distinct from centromeres (identified by the centromeric histone Cid) and dual-strand piRNA clusters (characterized by a piRNA-binding protein Rhino) (**figure 2.6**). Intriguingly, while

CG17801 localization is distinct from Rhino-bound regions, there is a correlation between Cid localization and CG17801 localization. In contrast, CG17806 localizes broadly to chromatin. The localization patterns in S2 cells were robust to the type of tag (Venus or Flag) the location (amino-terminus or carboxy-terminus) of the tag. It is possible that heat shock, which can affect chromatin itself, affects localization of the tagged peptides. In addition, S2 ceels do not normally express *Nnk, CG17801, CG17803* or *CG17806,* so it is possible that their localization is different in endogenous tissues.

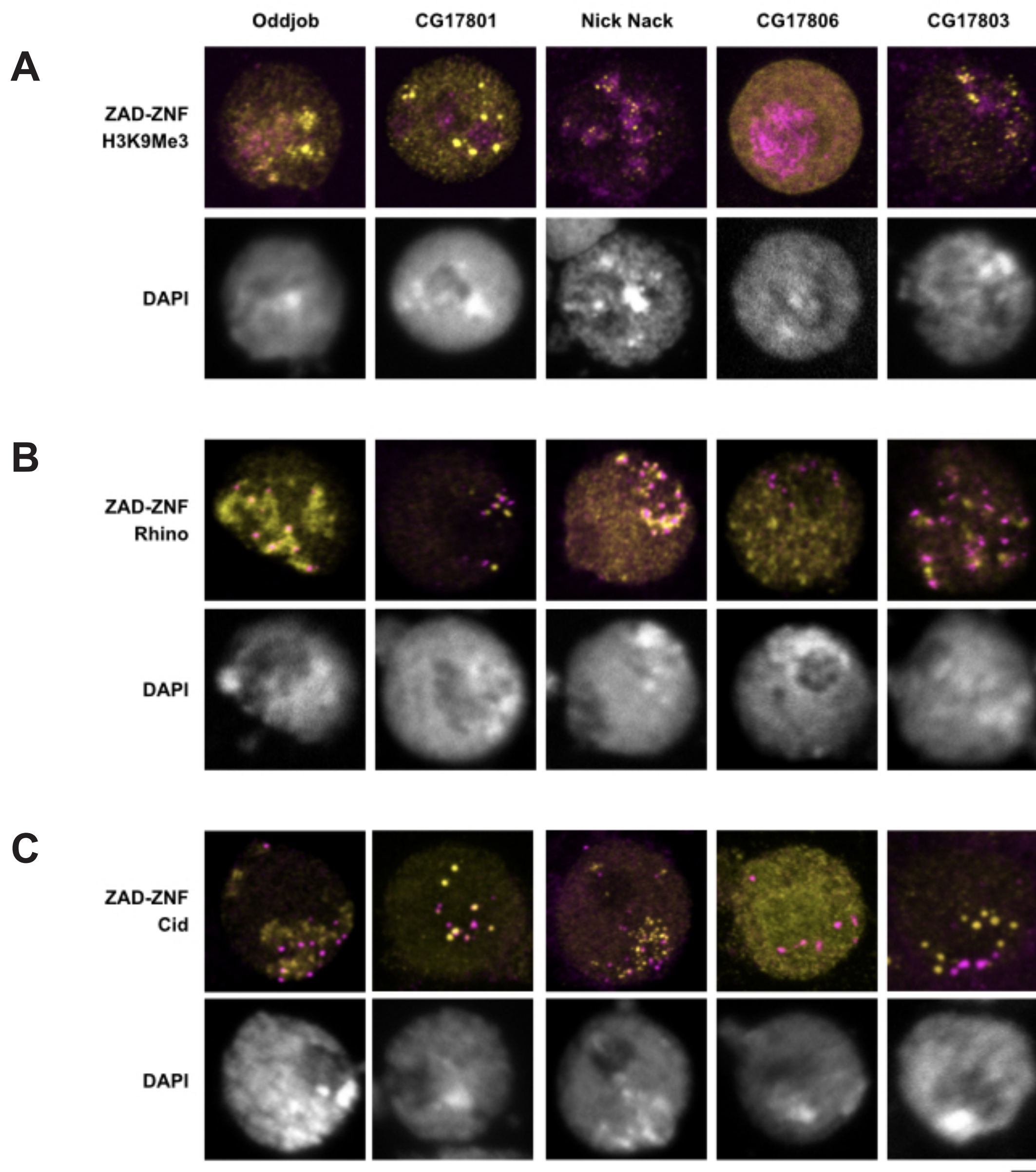


Figure 2.6 | Oddjob cluster ZAD-ZNFs localize to distinct regions of heterochromatin. All images are of a representative S2 cell nucleus transfected with a Venus-tagged ZAD-ZNF (yellow). DAPI marks DNA in each nucleus. **(A)** shows that most ZAD-ZNFs localize within and around heterochromatin (marked by pink for regions containing H3K9Me3), with the exception of CG17806, which localizes broadly across the nucleus. **(B)** we co-transfected cells with Venus-tagged ZAD-ZNF and Flag-tagged Rhino and found that Rhino and CG17801, Nick Nack and CG17803 localize to regions within the nucleus that are adjacent but not-overlapping Rhino. Oddjob localization extends beyond Rhino-bound regions and CG17806 localizes broadly across the entire nucleus. **(C)** Oddjob localizes to regions around centromeres while Nick Nack, CG17801 and CG17803 form puncta that are sometimes adjacent to but not overlapping centromeres. CG17806 localizes broadly throughout nucleus. Scale bar = 20 μm.

2.2.5 Oddjob and Nick Nack are necessary for viability and fertility

I next examined the phenotypic roles of *Odj-Nnk* cluster genes. Previous large-scale genome-wide studies analyzed the *Odj-Nnk* cluster of genes for defects in viability (Chen *et al.*, 2010; Schertel *et al.*, 2015). Some of these results were subsequently called into question because of phenotypic effects associated with secondary insertions in the KK knockdown lines that were used (Green *et al.*, 2014; Vissers *et al.*, 2016). To re-evaluate the role of all *Odj-Nnk* cluster genes in development and fertility, I knocked down each of these five ZAD-ZNF genes using ubiquitous or germline-specific GAL4 driver lines and VDRC RNAi knockdown lines that do not have the second site insertions associated with lethality (Green *et al.*, 2014; Vissers *et al.*, 2016) with the help of a research technician in the lab, Hannah McConnell. Confirming previous results from a genome-wide transcription factor screen,

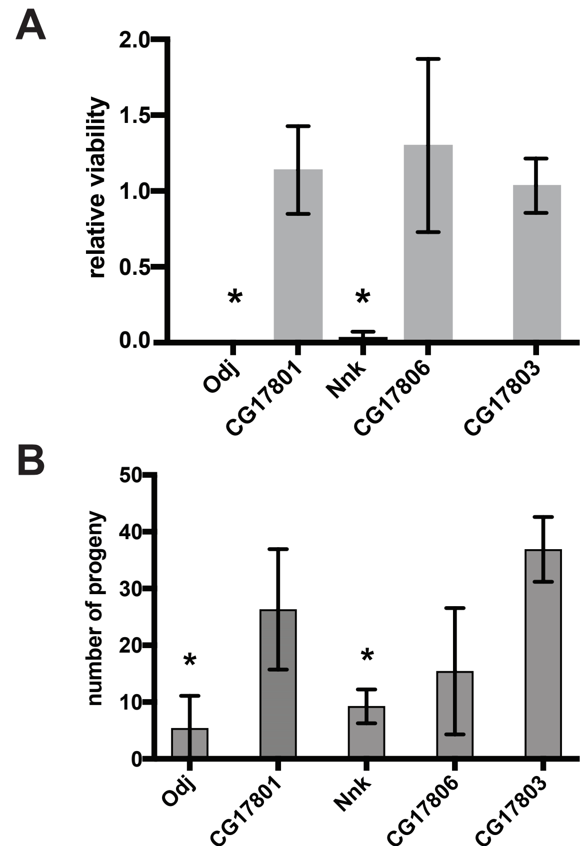


Figure 2.7 | Oddjob and Nick Nack are necessary for viability and fertility. (A) Quantification of the viability of adult flies that ubiquitously knockdown each of the Oddjob cluster members shown on the horizontal axis. The vertical axis shows the viability (relative to the number of knockdown progeny expected in each cross). Knockdowns were one in triplicate and standard deviation is shown as error bars. N >100 per cross. **(B)** Quantification of the number of progeny produced per single pair mating of a germline knockdown virgin female to a control (w1118) male. At least 10 crosses per knockdown were evaluated and standard deviation is shown as error bars on graph. Asterisk denotes p-value < 0.001, Student's t-test comparing data between control and ZAD-ZNF knockdowns.

we found that ubiquitous knockdown of *Odj* and *Nnk* are lethal whereas *CG17806*,

CG17801, and *CG17803* are dispensable for viability (**figure 2.7A**) (Schertel *et al.*, 2015).

RNA *in situ* hybridization suggests that transcripts of *Odj-Nnk* cluster genes are loaded maternally into the embryo (Fisher *et al.*, 2012); I therefore evaluated the effect of maternal depletion of these transcripts on progeny. In collaboration with Hannah McConnell, a research technician in the lab, I used the maternal-alpha-tubulin GAL4 driver line in conjunction with the RNAi lines, to attempt to knock down expression of *Odj-Nnk* cluster genes in mid-to-late oogenesis and presumably deplete maternal loading of the targeted transcripts. We found mothers carrying the driver and either the *Odj* or *Nnk* RNAi transgene produced significantly reduced the number of viable adult progeny, suggesting that Nnk and Odj likely function during oogenesis or embryogenesis (**figure 2.7B**).

In contrast to my RNAi results, a recent study reported that *D. melanogaster* flies hemizygous for a CRISPR-Cas9-mediated frameshift mutation of *Nnk* are viable (Kondo *et al.*, 2017). To test whether my knockdown approach actually produced lethality due to off-target effects (e.g., against other *Odj-Nnk* cluster genes that might be redundant with *Nnk*), and to re-confirm the phenotypic consequences of *Nnk* knockdown, I created a *Nnk* "rescue construct." This transgene contained *D. melanogaster Nnk* coding sequence that was 're-coded' using synonymous changes in the RNAi hairpin-targeted segment; this *Nnk* transgene is predicted to be resistant to RNAi inhibition. I also flanked this RNAi-resistant *Nnk* coding regions with one kilobase regions from upstream and downstream of the *Nnk* gene to capture the putative endogenous regulatory regions (**figure 2.8A**). Using PhiC31-mediated transgenesis, I inserted this 'rescue' *Nnk* locus into an *attP* site on the *D. melanogaster* X chromosome. I found that the recoded *Nnk* transgene completely rescued the

lethality associated with the ubiquitous knockdown of endogenous *Nnk* (**figure 2.8B**). While Kondo *et al.* observed that the CRISPR line was homozygous lethal, they found the line was viable over a deficiency and attributed the homozygous inviability to second site mutations. However, it is possible the deficiency spared Nnk, which may explain why the CRISPR/deficiency is viable. Thus, I conclude that *Nnk* is essential for viability in *D. melanogaster.*

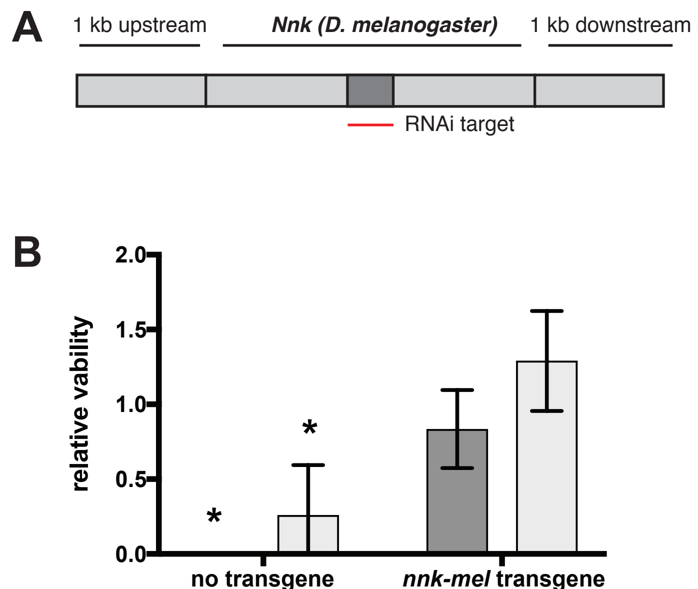


Figure 2.8 | Nick Nack transgene rescue. (**A**) schematic of transgene design, which contains the genomic region 1 kb upstream of the *Nnk* start codon and one kb downstream of the stop codon. The region targeted by the RNAi hairpin is highlighted by the red line and we recoded this region (marked in dark gray) so that it is RNAi resistant. **(B)** When Nick Nack is ubiquitously knocked down using Act5c-GAL4 RNAi driver, there is a significant drop in viability with a few female flies (gray bars) that escape lethality. Viability of males (dark gray bars) and females (light gray bars) is restored with the RNAi-resistant transgene. Error bars represent standard deviation between replicates of single pair matings and for each cross (>5 pair matings per cross), >20 total flies were counted per cross. Asterisks indicate $p<0.001$, Chi-squared test.

2.2.6 Nnk's recent adaptive evolution is not linked to its essential function

The analyses revealed several striking features about *Nnk*. It is evolutionarily young (~30 million years old) and dynamic in its evolution. It has been subject to gene duplication (in *D. takahashii*) and gene loss (in *D. elegans/ D. rhopaloa*, and *D. ananassae*). It has also evolved under positive selection with 52 fixed non-synonymous differences between *D. melanogaster* and *D. simulans* orthologs in the 439 aa protein-coding region (**Table 2.1**). In spite of its young age and evolutionary dynamism, *Nnk* is essential in *D. melanogaster.* I asked whether observed differences in *Nnk* amino acid sequences would affect Nnk subcellular localization.

To assay this, I used transient transfections of epitope-tagged *Nnk* orthologs from *D. melanogaster* and *D. simulans* into *D. melanogaster* S2 cells and found that both orthologs localize to heterochromatin (**figure 2.9A**). I found that when the orthologs were co-transfected into S2 cells, both proteins co-localize to foci (**figure 2.9B**). Thus, the rapid evolution of *Nnk* during the *D. melanogaster- D. simulans* divergence has not visibly affected its subcellular localization. *D. simulans* Nnk appears to have a more diffuse localization pattern in S2 cells when co-transfected with *D. melanogaster* Nnk, perhaps due to subtle differences in the affinity or stability of the proteins' interaction with DNA or other protein partners. However, the cytological assay may be unable to reveal subtle differences in DNA-binding. Thus, it remains possible that positive selection has affected Nnk DNA-binding specificity even if it has not affected its gross heterochromatic localization.

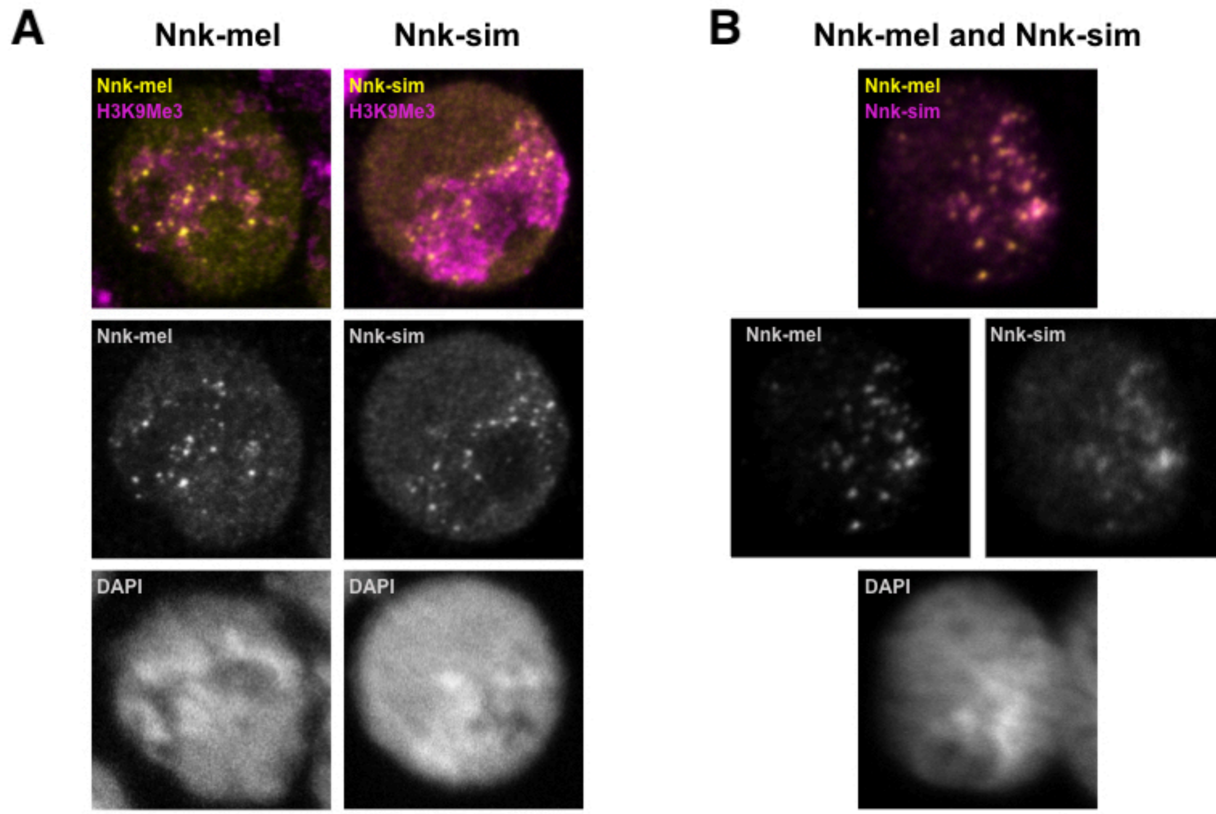


Figure 2.9 | Nick Nack *D. simulans* ortholog localizes to heterochromatin in *D. melanogaster*. Each image is of a representative S2 cell nucleus with DAPI marking DNA. (**A)** Venus-tagged *D. melanogaster* Nnk (Nnk-mel) and Venus-tagged *D. simulans* Nnk (Nnk-sim) are shown in yellow with H3K9Me3 in pink in the merged image. Both *Nnk* orthologs localize to foci within heterochromatin. **(B)** co-transfection of Flag-tagged Nnk-mel (yellow in merge) and Venus-tagged Nnk-sim (pink in merge) show that both orthologs colocalize to foci within the S2 cell nucleus. Scale bar = 20 μm.

Next, I examined the genetic consequences of *Nnk* positive selection on viability in *D. melanogaster*. If positive selection were a result of species-specific adaptation in the essential function of Nick Nack, I predict that the *D. simulans Nnk* gene would not be able to complement a *D. melanogaster Nnk* knockdown. Alternatively, positive selection might affect Nnk properties that are distinct from its requirements for viability. To distinguish between these possibilities, I created a 'rescue' transgene (**figure 2.10A**) containing the *D. simulans Nnk* coding sequence with 1 kb upstream/downstream of the gene from *D. melanogaster*. I then introduced this *D.*

simulans Nnk 'rescue' transgene into the same *attP* site on *D. melanogaster* X chromosome as the *D. melanogaster Nnk* 'rescue' transgene via PhiC31-mediated transgenesis (**figure 2.10A**). This comparison allowed for a powerful near-isogenic comparison of *D. simulans* and *D. melanogaster Nnk* transgenes in *D. melanogaster* strains. I then assessed whether the *D. simulans Nnk* transgene can rescue the lethality caused by knockdown of endogenous *D. melanogaster Nnk* with the assistance of Hannah McConnell, a research technician in the lab. We found that, despite being more than 10% divergent at the amino acid level, *D. simulans Nnk* can robustly rescue viability just as well as the *D. melanogaster Nnk* transgene in females (**figure 2.10B**). However, the *D. simulans* transgene has a significant fitness cost to males (**figure 2.10B**).

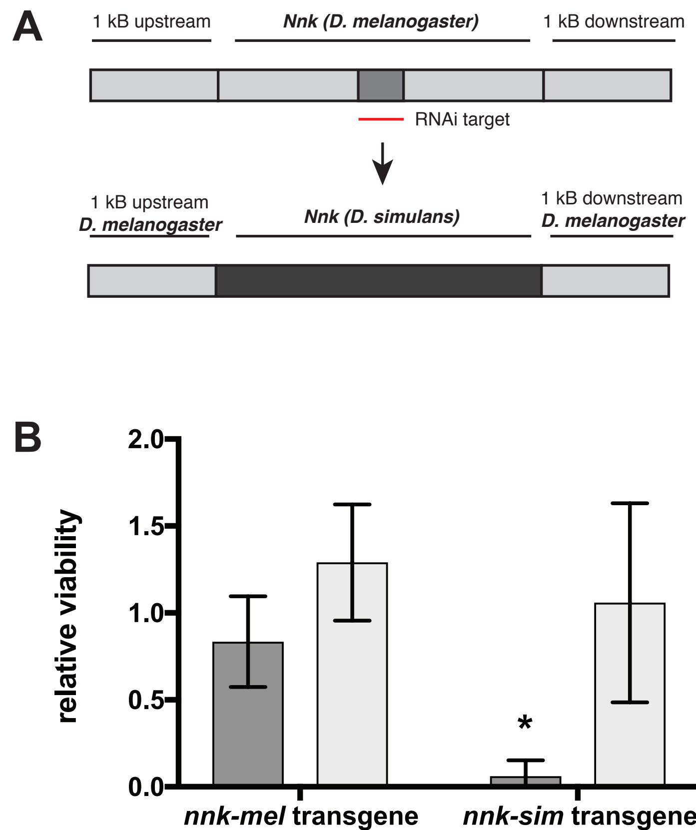


Figure 2.10 | Nick Nack *D. simulans* rescues *Nnk* null ubiquitous knockdown in *D. melanogaster*. (A) Schematic showing that we took the genomic region containing *Nnk* from *D. simulans* (from the Nnk start codon to stop codon) and put it within the *D. melanogaster* genomic region (1 kb upstream/downstream) of *Nnk*. **(B)** The lethal phenotype associated with ubiquitous knockdown of Nick Nack can be restored in males (dark gray bars) and females (light gray bars) with a RNAi-resistant Nick Nack *D. melanogaster* transgene. Although female viability is restored the Nick Nack *D. simulans* transgene, male viability is significantly affected even with the transgene. . Error bars represent standard deviation between replicates of single pair matings and for each cross (>5 pair matings per cross), >20 total flies were counted per cross. Asterisk indicate $p<0.001$, Chi-squared test.

I hypothesize that this striking effect on male viability is the consequence of the largely heterochromatic Y chromosome. I predict that *D. simulans Nnk* is not optimized to *D. melanogaster* genome and consequently, is less efficient at formation of *D. melanogaster* heterochromatin. This leads to increased mortality when there is additional demand for formation and maintenance of heterochromatin in the genome (for example, through the presence of the Y chromosome).

2.3 Discussion

Lineage-specific amplification, variable expression patterns and differences in protein structure have arisen in the family of *Drosophila* ZAD-ZNFs (Chung *et al.*, 2002; 2007b). I provide significant evidence for evolutionary dynamism in *Drosophila* ZAD-ZNFs. For example, I find that 61 of 90 *D. melanogaster* ZAD-ZNFs are at least 40 million years old (originated in the ancestor of *D. melanogaster* and *D. virilis).* Furthermore, 13 ZAD-ZNFs have evolved under positive selection since *D. melanogaster- D. simulans* divergence. Positive selection in ZAD-ZNFs is not primarily focused on their C_2H_2 domains but rather on the poorly characterized linker domains that connect the ZAD and C_2H_2 domains. Surprisingly, I find that ZAD-ZNF genes that evolve under positive selection are more likely to encode an essential function. Indeed, 9 of 13 positively-selected ZAD-ZNF genes in *D. melanogaster* are essential for functions such as embryonic axial patterning, larval progression, and meiotic progression (Chen *et al.*, 2000; Chen *et al.*, 2006; Harms *et al.*, 2000; Lake *et al.*, 2011; Page *et al.*, 2005; Uryu *et al.*, 2018)

One possibility is that the conflict that may drive the recurrent genetic innovation of some *ZAD-ZNF* genes is likely to stem from the localization and function of ZAD-ZNF proteins in the heterochromatic compartment of *Drosophila* genomes. Indeed, six characterized ZAD-ZNFs have roles at heterochromatin. For example,

Oddjob directly binds to HP1A (Swenson, *et al.* 2016) while ZIPIC, Pita, and Zw5 are found at heterochromatin–euchromatin boundaries (Zolotarev *et al.*, 2016). Moreover, Séance, Molting Defective and Ouija Board control the expression of heterochromatin-embedded genes necessary for larval development (Uryu *et al.*, 2018). Finally, my analyses of the Odj-Nnk cluster of ZAD-ZNF genes highlight that four of the five genes in this evolutionarily dynamic cluster encode heterochromatin-localizing proteins. Based on these observations, I propose that ZAD-ZNF diversification (marked by gene turnover, recent adaptive evolution, and recurrent site-specific rapid evolution) is driven by the high turnover of sequences embedded within heterochromatin. My findings are reminiscent of previous work that finds rapid diversification of many *Drosophila* heterochromatin-binding proteins to maintain their interaction with this rapidly evolving compartment of the nucleus (Bayes & Malik, 2009; Helleu *et al.,* 2016; Levine *et al.*, 2016; 2012).

My study of *Drosophila* ZAD-ZNF genes revealed an intriguing correlation between positive selection and essential function. In the *Odj* cluster, I find a potential cause of turnover within the locus: the propensity of the encoded proteins to localize to heterochromatin. Are positive selection and heterochromatin-localization causally linked? I was able to test this link by comparing the cytological localization and genetic function of the Nnk orthologs from *D. melanogaster* and *D. simulans*, which have acquired more than 10% divergence at the amino acid level despite being only 2-million-years diverged from each other. If the amino acid changes driven by positive selection were directly affecting heterochromatin localization, I would have predicted that the *D. simulans* Nnk cytological localization would be dramatically different from *D. melanogaster* Nnk. Contrary to this expectation, I found that *D. simulans* Nnk localizes similarly to *D. melanogaster* Nnk in heterochromatic foci in *D. melanogaster*

S2 cells. Moreover, I find that *D. simulans Nnk* can rescue the inviability caused by *D. melanogaster Nnk* knockdown. Intriguingly, male viability is significantly reduced in the *D. simulans* rescue. I hypothesize that the recent rapid evolution of Nnk has resulted in a decrease in efficiency of *D. simulans* Nnk to perform the essential function of Nnk in *D. melanogaster.* Together, these results suggest that positive selection during the *D. melanogaster-D. simulans* divergence does not affect heterochromatic localization but has an effect on its essential function. Furthermore, the ability for the divergent *D. simulans* ortholog to rescue *Nnk*-dependent essentiality suggests that its heterochromatin-localization may be tied to its essential function.

Why would this heterochromatin binding be essential? Although the bulk of heterochromatin is made up of highly repetitive elements such as satellite DNAs and TEs, heterochromatin also harbors many genes that are deeply embedded in heterochromatin (Devlin *et al.*, 1990; Eberl *et al.*, 1993; Schulze *et al.*, 2006; Schulze *et al.*,2005; Yasuhara & Wakimoto, 2006; Yasuhara *et al.*, 2005). These genes, many of which encode essential functions (Sinclair *et al.,* 2000), are sensitive to the amount of heterochromatin in the cell and even require a heterochromatic environment to ensure their correct expression and regulation (Wakimoto & Hearn, 1990). The sensitivity to the amount of heterochromatin is a potential explaination as to why males (which carry the increased burden of heterochromatin on the Y chromosome) are significantly less fit than females in the *D. simulans* rescue experiment.

I posit that the constant turnover of flanking and embedded sequence elements such as TEs and satellite DNAs may require a constant adaptation of transcription factors required for the proper expression of genes embedded in heterochromatin. Consistent with this hypothesis, a recent study found that three ZAD-ZNFs that are conserved across *Drosophila* regulate heterochromatin-embedded Cyp450 (Cytochrome

p450) genes in *D. melanogaster* (Uryu *et al.*, 2018); these genes are necessary for biosynthesis of the molting hormone, ecdysone in arthropods (Rewitz & Gilbert, 2008). Intriguingly, there is an expansion of Cyp450 genes in insect genomes, an expansion that parallels the dramatic rise in the number of insect ZAD-ZNF genes (Bergé *et al.*, 1998; Feyereisen, 1999; Li *et al.*, 2007; J. G. Scott & Wen, 2001).

Cyp450 genes also mediate detoxification and are therefore a recurrent target of insecticide-mediated adaptation as well as adaptation to counteract plant defenses against insect herbivory (Bergé *et al.*, 1998; Feyereisen, 1999; Li *et al.*, 2007; Scott & Wen, 2001). For example, insertion of an *Accord* transposable element upstream of the *Cyp6g1* gene leads to tissue-specific increased expression of the gene and DDT resistance (Chung *et al.*, 2007b). It is possible that the constant adaptation propels positive selection of TFs such as ZAD-ZNFs that help mediate proper gene regulation of Cyp450 genes. In this way, rapid diversification of ZAD-ZNFs can adapt to a changing landscape of Cyp450 gene regulation, while ensuring that the essential genetic functions of Cyp450 genes in early development are robustly maintained.

2.4 Methods

2.4.1 Bioinformatic analyses

I used the Pfam database (https://pfam.xfam.org) to identify all ZAD-containing proteins (PF07776) in 12 sequenced and annotated *Drosophila* species (El-Gebali *et al.*, 2019). Using NCBI's Conserved Domains search, I identified other domains found in the ZAD-containing proteins (Marchler-Bauer *et al.*, 2017). To estimate the evolutionary age of these ZAD-containing genes, I used OrthoDB to identify orthologs of the 90 ZAD-containing peptides across *Drosophila* (Zdobnov *et al.*, 2017).

2.4.2 Defining gene essentiality

I used FlyBase gene summaries (FlyBase.org) and published studies (when available) to define gene essentiality. My criteria for essentiality were broad: if there was a lethal allele reported, I counted the gene as essential for viability in *D. melanogaster.*

2.4.3 Defining orthologs

Since *Oddjob* cluster genes experienced numerous independent segmental duplications, it was not possible to determine orthologs by synteny. Instead, I took a tBLASTn approach to identify candidate orthologs of *Oddjob* cluster genes, using the genes in *D. melanogaster* as queries. I used a reciprocal blast search strategy to identify potential orthologs and validated these orthologs by making a maximum likelihood phylogenetic tree (LG substitution model in PhyML with 100 bootstrap replicates) of a manually trimmed protein alignment of all the sequences I collected. I called orthologs based on genes that formed a monophyletic clade with the each of the *D. melanogaster Oddjob* cluster genes. I mapped these *Oddjob* cluster orthologs back to each genome assembly to determine the composition of the *Oddjob* locus in other *Drosophila* species. In cases where there were other ZAD-ZNFs present, I blasted them against the *D. melanogaster* genome to examine if there were any orthologs present in

D. melanogaster. If the top hit was not a member of the Oddjob locus, I did not include it in the tree. In the case where there are partial ZAD-ZNFs (containing just the ZAD domain or the zinc finger domains), I performed a blast search against the *D. melanogaster* genome. If the top hit was a member of the *Oddjob* cluster, I assigned orthology by making a phylogenetic tree with all other *Oddjob* cluster orthologs.

2.4.4 Analyses of positive selection

For the McDonald-Kreitman test, I extracted the gene of interest from *D. melanogaster* population genetic datasets available through Popfly (www.popfly.com) and removed low frequency (<0.05) variants from the dataset to minimize the effects of false positives (Hervas *et al.*, 2017). I used manually trimmed alignment of the *D. melanogaster* filtered dataset and the reference *D. simulans* sequence for the McDonald-Kreitman test (http://mkt.uab.es/mkt/) (McDonald & Kreitman, 1991; Egea *et al.*, 2008).

To look for evidence of recombination across *Oddjob* cluster members, I utilized GARD (www.datamonkey.org) (Delport *et al.*, 2010; Kosakovsky *et al.*, 2006; Pond & Frost, 2005; Weaver *et al.*, 2018) on trimmed codon alignments using the ClustalW "translation align" function (Larkin *et al.*, 2007) of the Geneious software package (version 9) (Kearse *et al.*, 2012). I utilized the codeml algorithm from the PAML suite of programs to look for recurrent site-specific positive selection on each *Oddjob* cluster gene (Yang, 1997). I ran M7 vs M8 and M8 vs M8a; starting omega = 0.7 and 1.5 with codon frequency of 1/61 and F3x4 and found that the results were robust to changes in these parameters. I used manually trimmed alignments of each gene and gene trees (generated from the manually trimmed alignments using PhyML with HKY85 substitution model) as inputs for PAML analysis. Sites under positive selection are defined as having a Bayes Empirical Bayes posterior probability > 0.95.

2.4.5 Cloning Oddjob cluster genes into expression vectors

Oddjob cluster genes were amplified from genomic DNA from *D. melanogaster* and *D. simulans* and directionally cloned into pENTER/D-TOPO (ThermoFisher) according to the manufacturer's instructions. I sequence verified that the clones had the appropriate gDNA sequence. I used LR clonase II (ThermoFisher) to get each *Oddjob* cluster gene into the *Drosophila* Gateway Vector destination vector to express the gene of interest with an N-terminal Venus tag under the control of the *D. melanogaster Hsp70* promoter (pHVW). For Rhino localization in S2 cells, the *Rhino* gene was cloned into the *Drosophila* Gateway Vector destination vector to enable expression of Rhino with an N-terminal 3xFlag tag under the control of the *D. melanogaster Hsp70* promoter (pHFW).

2.4.6 Tissue culture and transfection

Schneider 2 cells were obtained from the *Drosophila* Genomics Resource Center (Bloomington, IN, USA) and grown at 25°C in M3+BPYE+10%FCS. For the transfections, one million cells were seeded, and one day later 2 micrograms of plasmid DNA was transfected into cells using Xtremegene HP transfection reagent (Roche) according to the manufacturer's specifications. For the Rhino localization experiment, cells were co-transfected with 1 ug of GFP-tagged ZAD-ZNF vector and 1 ug of Flag-tagged Rhino vector. Cells were allowed to recover for 24 hours post transfection, heat shocked for 1 hour at 37°C and recovered for 3 hours at 25°C prior to fixation.

2.4.7 Immunocytochemistry and imaging

Cells were transferred to coverslips for 30-45 minutes prior to starting the immunohistochemistry protocol. 0.5% sodium citrate hypotonic solution was added to

the coverslip for 10 minutes to swell cells, which was then spun at 1900 rpm for 1 minute in a Cytospin III to remove the cytoplasm from the cells. The sodium citrate was immediately removed and cells were subsequently fixed. For fixation, 4% PFA + PBST (PBS + 0.1% Triton) was added to the cells for 10 minutes. Coverslips were then washed in PBST and blocked for 30 minutes in PBST + 3% BSA. Cells were incubated with primary antibody overnight at 4 ^{o} C in a humid chamber. For immunolocalization, the following dilutions of primary antibodies were used: GFP (Abcam AB13970) 1:1000, H3K9Me3 (Abcam AB8898) 1:500 M2 FLAG (Sigma-Aldrich F4042) 1:1000. After washing with PBST three times for 10 minutes per wash, the following fluorescent secondary antibodies were used at 1:1000 dilution: goat anti-chicken (Invitrogen Alexa Fluor 488, A-11039), goat anti-rabbit (Invitrogen Alexa Fluor 568, A-11011) and goat anti-mouse (Invitrogen Alexa Fluor 568, A-11031). The cells were incubated with 1x DAPI in the final wash and mounted in SlowFade Gold Mounting Medium (ThermoFisher). I imaged cells on a Leica TCS SP5 II confocal microscope with LASAF software and images were processed using ImageJ and were representative of the cell population.

2.4.8 Viability studies

I used an *Act-GAL4/Cy-GFP* for ubiquitous knockdown. The RNAi lines used to specifically target *Oddjob* cluster genes are the VDRC KK or GD lines: *Oddjob* (27971), *CG17803* (38869), *CG17801* (29501), *CG17806* (101592) and *Nick Nack* (102311). RNAi controls used for the experiment were *Cid* (43856) and *HP1B* (26097) for the ubiquitous knockdown. Ubiquitous knockdown of *Cid* produced no viable progeny. With the help of Hannah McConnell, a technician in the lab, we crossed 5 virgin females carrying *Act-GAL4/Cy-GFP* to 3 males of each RNAi line. We allowed the females to lay eggs for 3 days and flipped the flies into fresh vials three times. Each

cross was performed in triplicate. Progeny were counted 10-15 days after each cross was set up. A minimum of 20 Cy-GFP males and 20 Cy-GFP females were required for us to quantify the crosses.

2.4.9 Fertility studies

To deplete targeted transcripts in mid-to-late oogenesis and maternal loading, I used the *P{matalpha4-GAL-VP16}V37* (BL7063) driver line. The RNAi lines used to specifically target *Oddjob* cluster genes are the VDRC KK or GD lines: *Oddjob* (27971), *CG17803* (38869), *CG17801* (29501), *CG17806* (101592) and *Nick Nack* (102311). RNAi controls used for the experiment were *HP1B* (26097) and *shy (y[1] sc[*] v[1]; P{y[+t7.7] v[+t1.8]=TRiP.GL00646}attP40* (BL38207)) for the germline knockdown. These experiments were set up and quantified with the help of Hannah McConnell, a research technician in the lab. Germline knockdown of *shy* produced no viable progeny. For the germline knockdown experiments, we crossed 5 virgin female *P{matalpha4-GAL-VP16}V37* to 3 males from each of the RNAi lines. We allowed the females to lay eggs for 3 days and flipped the files into fresh vials three times. Each cross was performed in triplicate. Germline-knockdown females (i.e. flies bearing both the *GAL4* driver allele and the RNAi allele) were collected and crossed to w^{1118} males in order to test their fertility. Crosses were performed in triplicate and set up with a single virgin female (3-5 days old) and a single male (5-7 days old); females were allowed to lay eggs for 3 days and the parents flipped three times. We counted progeny from each cross 10-15 days after the cross was set up. Minimum number of progeny counted were 20 per replicate. Each cross had at least three replicates. All flies were raised at 25˚C.

2.4.10 Transgene design

I designed *D. melanogaster* recoded transgene by GENEWIZ Co. Ltd. (Suzhou, China), which comprised a 3.3 kB fragment containing the genomic region of *Nick-nack* plus 1

kB upstream and downstream of the start and stop codons, based on *D. melanogaster* reference sequence. I recoded the sequence targeted by the VDRC RNAi KK line (103211) by making synonymous changes at each codon. The resulting fragment was cloned into a plasmid I generated which contains a 3xP3-DsRed attP, which produces florescent red eyes in the adult to mark the presence of the transgene. To generate the *D. simulans* transgenic allele, I codon optimized the *D. simulans Nnk* coding sequence for the *D. melanogaster* genome using IDT's codon optimization tool and synthesized by GENEWIZ Co. Ltd. (Suzhou, China). The resulting fragment was swapped for the *D.melanogaster Nnk* coding sequence in the plasmid described above using the NEBuilder kit (New England Biolabs). I submitted transgenic constructs to The BestGene Inc. (Chino Hills, CA) for injection into the X-chromosome *attP18* line (BL 32107) using *PhiC31* site specific integration (Groth *et al*., 2004).

2.4.11 Transgene rescue cross

For the transgene rescue cross, Hannah McConnell and I used *Nnk* transgene bearing virgin females were crossed to the *RNAi* line. We crossed 3 male flies bearing the *Nnk* transgene and the RNAi allele to 5 *Act5C-GAL4/CyO-GFP* virgin females. We allowed the females to lay eggs for 3 days and flipped the cross three times. We set up each cross in triplicate and counted progeny were counted 10-15 days. Minimum number of progeny counted were 20 per replicate. Each cross had at least three replicates. All flies were raised at 25˚C.

Chapter 3

THE ZAD-ZNF NICK-NACK IS NECESSARY FOR LARVAL PROGRESSION IN *D. MELANOGASTER*

3.1 Introduction

The steroid hormone ecdysone is a master regulator of insect development. Emerging studies in *D. melanogaster* reveal that *Drosophila*-specific transcription factors are necessary for promoting the transcription of genes embedded in heterochromatin that encode ecdysone biosynthetic enzymes (Neubueser *et al.*, 2005; Komura-Kawa *et al.*, 2015; Uryu *et al.*, 2018). These proteins belong to a large family of transcription factors, the Zinc Finger Associated-Domain Zinc Finger Proteins (ZAD-ZNFs). ZAD-ZNFs arose in the ancestor of arthropods and vertebrates and are the largest class of transcription factors within insect genomes (Chung *et al.*, 2007a). ZAD-ZNFs (Séance, Ouija Board and Molting Defective) necessary for ecdsyone biosynthetic gene expression arrest in larval development and die (Neubueser *et al.*, 2005; Komura-Kawa *et al.*, 2015; Uryu *et al.*, 2018). This lethality can be rescued by supplementing the diet with ecdysone or by with overexpression of ecdysone biosynthetic enzymes. In chapter 2, I identified a ZAD-ZNF, *Nick Nack (Nnk, CG17802),* which is necessary for larval development in *D. melanogaster* despite having a young and evolutionary dynamic history in *Drosophila*. In this chapter, I investigate whether the essential function of *Nnk* is tied to its role in ecdysone biosynthesis or regulation.

Progression of insect development requires an integration of temporal and environmental signals. Holometabolous insects undergo metamorphosis and undergo periods of robust growth during larval stages (instars). In response to this growth and to circadian, temperature, and other cues, ecdysone levels fluctuate and trigger developmental transitions in flies (**figure 3.1**). At the end of each instar stage, insects shed their exoskeleton and replace it with a newly synthesized, larger one. *D. melanogaster* have three larval instar stages, which together last for four days at 25˚C on nutrient rich food. At the end of this period, the larvae pupate, undergo metamorphosis and become a reproductively competent adult flies.

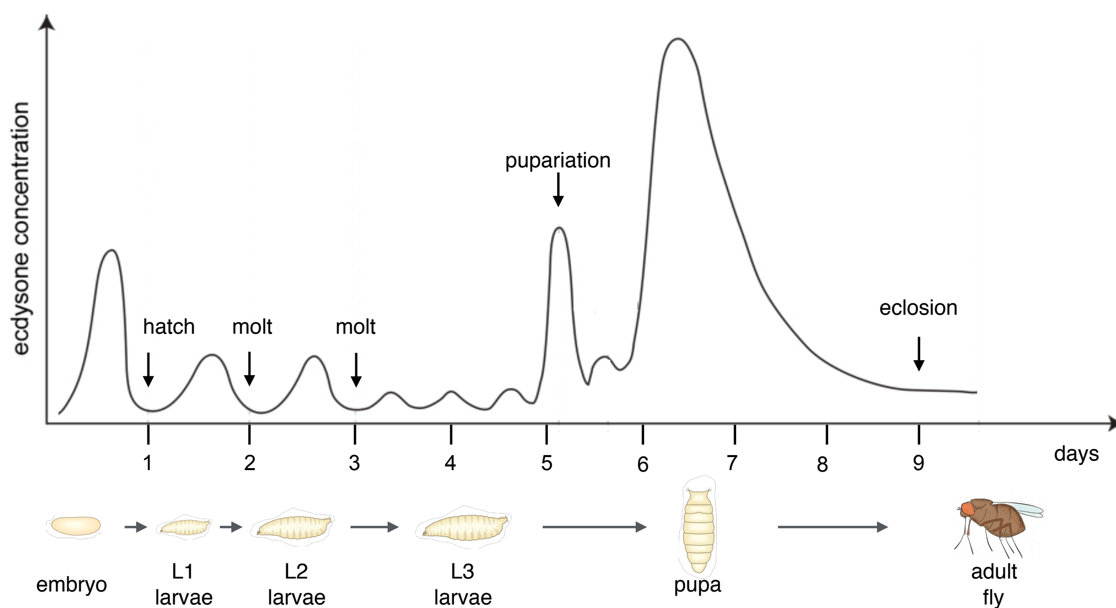


Figure 3.1 | Fluctuations in ecdysone levels trigger developmental transitions. Changes in peripheral ecdysone concentration trigger developmental transitions in a time dependent manner at 25˚C with a nutrient-rich diet (adapted from Mirth Lab webpage).

Ecdysone is synthesized in the prothoracic gland. Dietary sterols are modified through a series of biosynthetic steps (nutrient flow is summarized in **figure 3.2**) and release of the steroid hormone product is triggered by a combination of physiological

and environmental cues. Wieschaus and Nusslein-Volhard first identified mutants in the *Drosophila* ecdysone enzymatic pathway in their saturation screen of embryonic mutants (Jürgens *et al*., 1984; Wieschaus *et al*., 1984). In addition to identifying the key genes involved in embryonic body patterning for which they were awarded the 1995 Nobel Prize, the team also identified a handful of genes that affected cuticular development. Based on the ghostly pattern of the cuticle preparation used in the screen, they named these mutant genes *disembodied, shade, shadow* and *phantom.* Decades later, O'Connor and Gilbert began to study these genes, and they found that these mutants result from defects in ecdysone biosynthesis enzymes (Chávez *et al*., 2000). Intriguingly, although ecdysone is a regulator of molting in insects and arthropods, species have their own repertoires of not only ecdysone biosynthesis

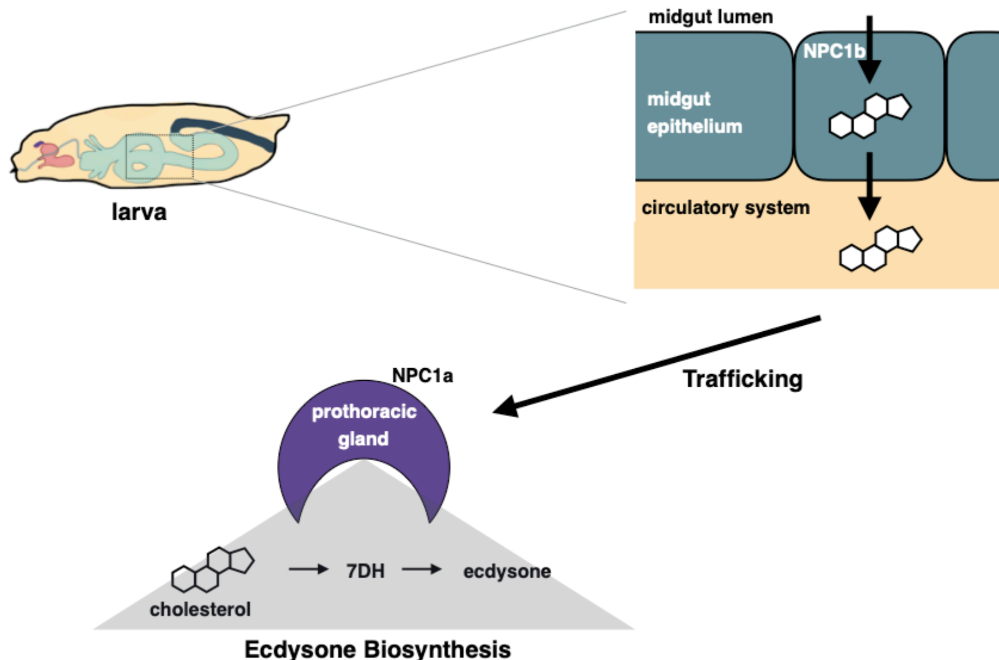


Figure 3.2 | Nutrient flux during ecdysone biosynthesis. Dietary sterols are absorbed in the midgut lumen through Npc1b receptors, transported to the circulatory system and absorbed into the prothoracic gland through the Npc1a receptor. Within the prothoracic gland, dietary sterols are converted into ecdysone.

genes (Rewitz & Gilbert, 2008; Schumann *et al.*, 2018) but also genes that regulate these genes (Uryu *et al.*, 2018).

Based on these prior discoveries of an evolutionary dynamic means of ecdysone biosynthesis and regulation, and the role of three ZAD-ZNF transcription factors in this process, we investigated whether *Nnk* also participates in ecdysone biosynthesis or regulation. First, we confirmed the essential function of *Nnk* using new transposon-insertion and CRISPR-mediated gene deletions. We found that *Nnk* mutants display a characteristic phenotype of developmental arrest in the L1 larval stage. Whole larvae RNA-sequencing revealed that, compared to wild type, *Nnk* mutants are unable to activate the L2 transcriptional program. Specifically, we find that many components of the steroidogenesis and transport pathways are misregulated in the *Nnk*-null larvae. In contrast to previously characterized ZAD-ZNFs that are necessary for larval development, we found that *Nnk* mutants cannot be rescued by dietary supplementation with ecdysone or its biosynthetic intermediates. We also found that *Nnk*-null larvae had no defect in steroid absorption. Instead, we find that the *Nnk*-null larvae have a general defect in maintaining regulation of euchromatic genes expression and genetic elements embedded within heterochromatin through the larval transition. We hypothesize that the lethality due to loss of *Nnk* occurs, in part, due to loss of expression or regulation of crucial genes embedded in heterochromatin.

3.2 Results

3.2.1 Nick Nack is an essential gene in D. melanogaster

Nick Nack was initially characterized as an evolutionarily young, essential gene in a *D. melanogaster* screen for evolutionarily young genes (Chen *et al.*, 2010). However, the screen utilized the KK RNA-interference (RNAi) collection by the Vienna *Drosophila* Stock Center (VDRC), which was later found to harbor a second-site lethal

mutation that caused lethality when crossed to *GAL4*-driver lines (Green *et al.*, 2014; Vissers *et al.*, 2016). This lethality is the result of ectopic *tiptop* expression and is independent of the hairpin insertion, calling into question the essentiality of the genes identified in the Chen *et al.* study.

To determine if *Nnk* is essential for viability in *D. melanogaster,* we began by validating the RNAi line that was used in the Chen *et al.* 2010 study, which first identified *Nnk* as a young, essential gene. We found that the VDRC RNAi line used in this study did not have an insertion upstream of the *tiptop* gene, which is associated with lethality. We also confirmed that ubiquitous knockdown of *Nnk* is lethal and rescued this lethality with a *Nnk* transgene (see chapter 2).

We also evaluated previously generated mutant *Nnk* alleles to further interrogate its essentiality. Schuldiner *et al.* generated a putative *Nick Nack* null allele as part of a genome-wide gene disruption study (Schuldiner *et al.*, 2008). This allele has a *piggyBac* transposon insertion in the 5' UTR of *Nnk* two bp upstream of the start codon. This insertion also contains a DsRed reporter driven by an eye-specific promoter and is flanked by stop codons in all three reading frames; these features prevent translation downstream of the insertion (**figure 3.3**). We found that this allele is homozygous lethal. However, viability can be rescued after precise mobilization of the *piggyBac* element, which restores an intact *Nnk* locus. This result suggests that the lethality from the allele is caused by the *piggyBac* insertion upstream of *Nnk,* which we therefore refer to as "*Nnk* null allele."

An independent *Nnk* mutant allele was generated by CRISPR mutagenesis by Kondo *et al.* as a follow up to the Chen *et al.* study (Kondo *et al.,* 2017); we will refer to this allele as "*Nnk* CRISPR allele." This allele contains a four-base-pair deletion within the coding sequence of the gene that creates a frameshift and a premature stop codon

within the linker region of the encoded Nick Nack protein (**figure 3.3**). The CRISPR allele is predicted to produce a truncated protein. Kondo *et al.* reported that although this allele is homozygous lethal, it was viable over a deficiency and concluded that the lethality was likely the result of a second-site mutation on the chromosome. However, we found that the *Nnk* CRISPR allele is indeed lethal over deficiency (BL9207) (three independent crosses were set up, all 72 resulting progeny were heterozygous for either the deficiency or *Nnk* CRISPR allele) and fails to complement the *Nnk*-null (*piggyBac*-containing) allele, strongly suggesting that the lethality is the result of the mutation within the *Nnk* coding sequence.

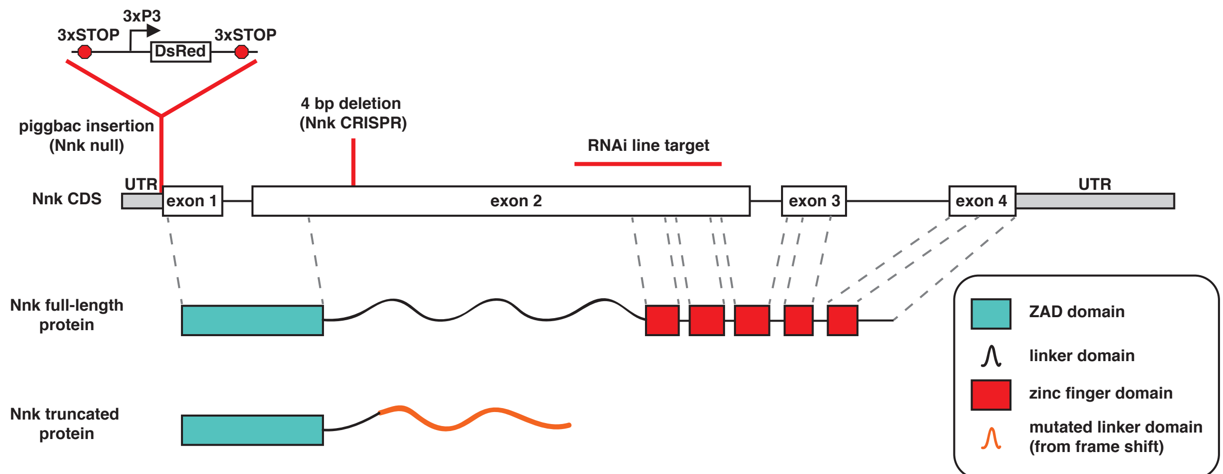


Figure 3.3 | *Nick Nack* alleles used in this study. *Nnk*-null is a *piggyBac* insertion in the 5' UTR of the gene that stops transcription through the gene. *Nnk* CRISPR is a 4 bp deletion within exon 2; this deletion creates a frameshift and a premature stop codon. The resulting truncated protein is shown below the full-length protein and contains a mutated linker region. The VDRC RNAi line targets the distal part of exon 2 (ZNF 1 and 2).

To confirm my findings, we utilized the rescue *Nnk* transgene (described in detail in Chapter 2) to rescue both the *Nnk pBac*-null as well as the *Nnk* CRISPR-null allele. The rescue transgene contains *Nnk* coding sequence plus 1 kilobase upstream and downstream of the gene to capture *Nnk* regulatory sequences. We found that the rescue transgene was able to rescue the viability of both the *Nnk*-null and CRISPR null alleles. These observations support the hypothesis that the lethality observed with

both the *Nnk pBac* null and CRISPR null alleles are the direct result of disruptions in producing a full-length *Nick Nack* transcript.

3.2.2 Nick Nack is necessary for larval progression

From the characterization and validation of the null alleles, we determined that *Nick Nack* was necessary for viability. To pinpoint when *Nnk*-null organisms were dying, we balanced the mutant *Nnk* alleles over a TM3G balancer chromosome expressing GFP. We crossed heterozygous *Nnk*-null flies and tracked the stage in development when *Nnk-null* homozygous progeny arrest. We could distinguish between heterozygous or homozygous *Nnk*-null larvae by focusing only on larvae lacking GFP (GFP-). We found that GFP-expressing larvae progress through embryogenesis at the same rate as their heterozygous siblings and are morphologically indistinguishable from the heterozygous siblings at the L1 stage (**figure 3.4**). The null larvae are able to move towards yeast paste and consume the yeast paste much like their heterozygous siblings. When the heterozygous siblings molt into the L2 stage 48 hours after egg laying (AEL), the *Nnk*-null larvae fail to achieve this developmental transition. From 48 hours AEL to the time of death (within 60 hours AEL), the *Nnk*-null larvae progressively become unable to move or eat and die. The *Nnk*-null larvae never form the L2 mouthhooks or attempt to molt.

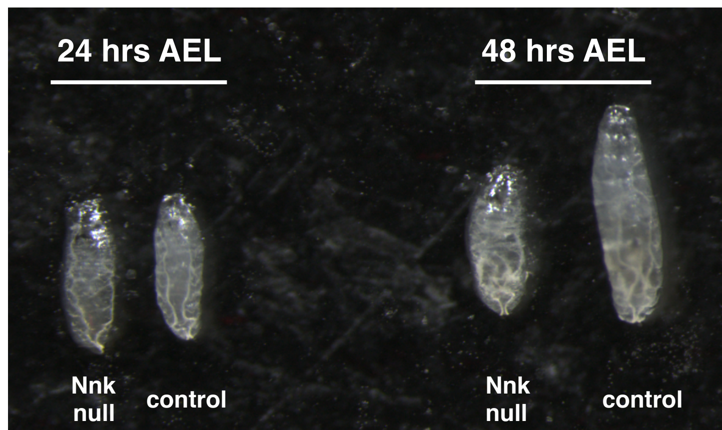


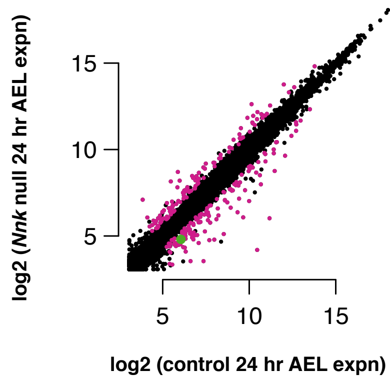
Figure 3.4 | *Nick Nack* null larvae arrest in early larval development. 24 hours after egg laying (AEL) the *Nnk*-null larva is morphologically indistinguishable from the control larva. 48 hours AEL, the *Nnk*-null larva is significantly smaller than the age matched control and also fails to undergo the first larval molt.

Intriguingly, we find that although the Nnk CRISPR larvae (homozygous for the CRISPR allele) are also unable to progress through larval development; however, they differ from the *Nnk*-null homozygous larvae in that they stay arrested at this early developmental timepoint for over one week. Furthermore, trans-heterozygous larvae, which carry one Nnk CRISPR allele and one Nnk-null allele, behave like the homozygous Nnk CRISPR organisms, suggesting a dominant effect of the truncated Nnk produced by the Nnk CRISPR allele. The findings suggest that the CRISPR allele is capable of producing a truncated Nick Nack protein and is slightly less deleterious than the Nnk-null allele, which appears to be much more penetrant and early acting in its lethality.

3.2.3 Examining the transcriptome of Nnk-null larvae

To understand the basis of this profound stage of developmental arrest in *Nnk*-null larvae, we compared the transcriptome of the *Nnk*-null larvae to their age-matched heterozygous siblings (controls) 24 hours AEL, when the two different genotypes are

morphologically indistinguishable, and 48 hours AEL, when the null larvae are significantly smaller than their age matched controls. Each time point was collected in triplicate. The RNA sequencing data analysis was done with the help of Janet Young, a staff scientist in the lab. To examine the early effects of *Nnk* loss, we initially focused exclusively on euchromatic genes. 24 hours AEL, find that the *Nnk*-null larvae have 249 genes (2.4% of all expressed genes) that are significantly differentially expressed between the *Nnk*-null and control larvae. Specifically, 116 genes are expressed at least two-fold higher in the *Nnk*-null larvae compared to the control larvae whereas 113 genes are expressed at least two-fold lower (**figure 3.5A).** However, this relatively subtle pattern of mis-expression is dramatically altered 48 hours AEL (**figure 3.5B).** We find that 3,027 (28.1% of all expressed genes) genes are differentially expressed in *Nnk*-null compared to age-matched control larvae, with 1,301 genes at least two-fold up-regulated and 1,726 genes at least two-fold down-regulated in the *Nnk*-null mutants. Overall, we find that genes are more significantly down-regulated than up-regulated in the *Nnk*-null larvae (**figure 3.5B**).



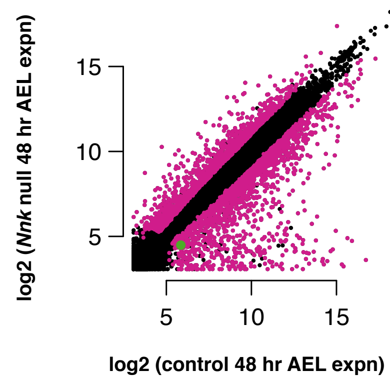


Figure 3.5 | *Nick Nack-null* larvae fail to turn on a significant number of genes 48 hours AEL. Comparing all expressed genes between *Nnk*-null larvae and control larvae, genes that are more than two-fold up- or down-regulated are shown in magenta. The green dot is *Nick Nack.* **(A)** 24 hours AEL, approximately an equal number of genes are upregulated and downregulated. **(B)** At 48 hours AEL, there are significantly more genes affected by *Nnk* loss and there are a significant number of genes that down-regulated in the *Nnk*-null larvae compared to control larvae.

Upon comparing all the expressed genes in each of the four categories, we found that the *Nnk* larvae 24 hours and 48 hours AEL are more similar to the control larvae 24 hours AEL. From this analysis, I conclude that not only are the null larvae 48 hours AEL morphologically similar to the larvae 24 hours AEL, but also their transcriptional profiles are more similar to more juvenile stages of development (**figure 3.6).**

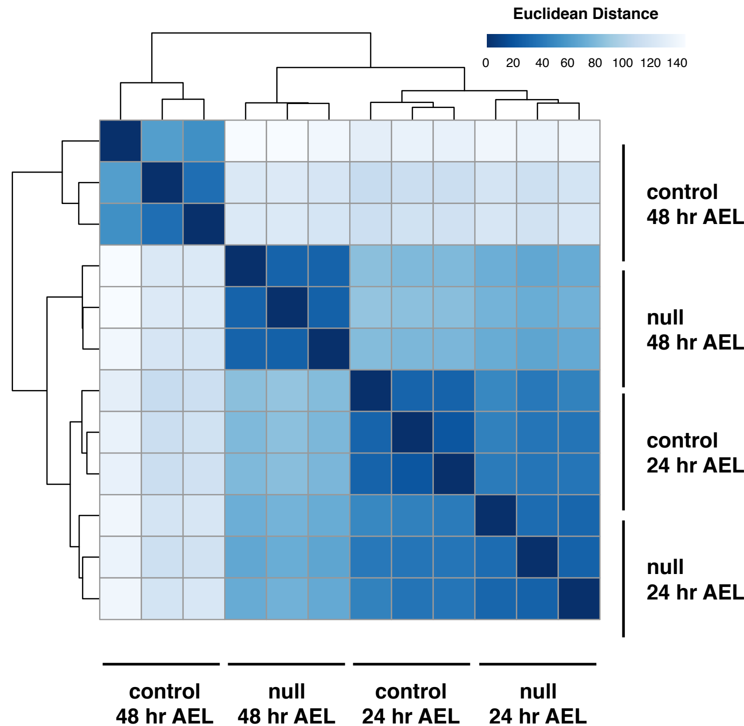


Figure 3.6 | *Nick Nack*-null larval transcriptome 48 hours AEL is similar to the *Nick Nack* null and control larval transcriptomes 24 hours AEL. Relationship of transcriptome profiles among control (*Nnk*-null heterozygotes) and null samples (*Nnk*-null homozygotes) reveals that null transcriptomes 48 hours AEL are more similar to transcriptomes of larvae 24 hours AEL than the age-matched controls. Sample-to-sample Euclidean distance matrix with hierarchical clustering using raw read counts of 11,428 genes.

I hypothesized that if *Nnk* regulates euchromatic genes, it may act as a transcriptional activator like some previously characterized ZAD-ZNFs (Harmes *et al.,* 2000; Komura-Kawa *et al.,* 2015; Li & Gilmour, 2013; Uryu *et al.,* 2018). Therefore, if *Nnk*-null larvae fail to express crucial gene(s) early in larval development, this failure will lead to more dysfunction later in development. To capture the early effects of *Nnk* loss, we examined the 24-hours AEL transcriptome for genes that are most significantly decreased in expression in the *Nnk* larvae compared to the controls. We found that the most significantly down-regulated genes between the *Nnk*-null larvae and the control larvae are involved in digestion and absorption.

3.2.4 Steroid Absorption is not visibly affected in the Nnk-null organisms

From the RNAseq analysis, I was struck by how many euchromatic steroid pathway genes were differentially expressed in the *Nnk*-null larvae compared to their control siblings. Specifically, a significant number of genes associated with sterol transport (*Niemann-Pick Type C*), sterol metabolism (*cytochrome P450*) and sterol sensing (*Hormone receptor-like in 96*) were affected (Bujold *et al.*, 2010; Fluegel *et al.*, 2006; Gilbert, 2004; Huang *et al.,* 2005; Voght *et al.*, 2007). In light of the analysis of the RNAseq dataset, I hypothesized that the *Nnk*-null lethal phenotype could be the result of a failure of sterol absorption, transport, or the synthesis of ecdysone itself.

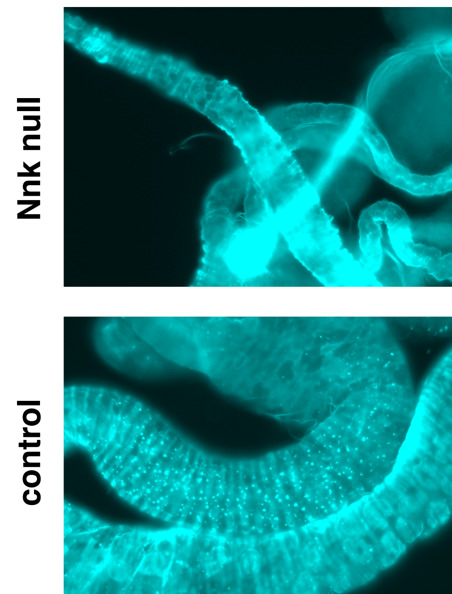


Figure 3.7 | *Nick Nack* null midguts show sterols within midgut upon filipin staining. Total sterol levels were assessed by staining age-matched w^{1118} (control) and Nnk-null larval midguts with filipin. Both tissues show similar levels of fluorescent intensity. Clusters of puncta in control are undigested yeast in lumen. Images were taken at the same fluorescence intensity. Scale bar = 25 μm.

Dietary sterols are absorbed by midgut epithelial cells by the Niemann-Pick Type C 1b (NPC1b) transmembrane receptor (Voght *et al.*, 2007). They are then shuttled to the prothoracic gland (PG) via the hemolymph and absorbed by the PG through the NPC1a receptor (Fluegel *et al.*, 2006; Huang *et al.,* 2005) (**figure 3.2).** In the PG, sterols are converted to ecdysone through a series of biosynthetic enzymes expressed specifically in the tissue.

To examine whether *Nnk*-null larvae are able to absorb steroids, I dissected the digestive system of *Nnk*-null L1 larvae and stained the tissues with Filipin. Filipin is a florescent fungal compound that binds to sterols. If the midgut were not fluorescent, it would imply that the *Nnk*-null larvae are unable to absorb sterols from the lumen into the midgut epithelial cells. However, contrary to this expectation, I find that the *Nnk*-null larval epithelial cells do fluoresce at levels comparable to control larvae (**figure 3.7).** I, therefore, conclude that they do not have a defect in steroid absorption.

Since the *Nnk*-null larvae are able to absorb sterols, I hypothesized that the defect could be in the transport of sterols to the prothoracic gland (the site of ecdysone biosynthesis) or in the generation of active ecdysone. There are still large gaps in our understanding of the mechanism by which sterols are transported to the prothoracic gland. However, what is known is that supplementing larval diets with cholesterol, ecdysone or its intermediates can extend the life of larvae or completely rescue the larval arrest phenotype (Huang *et al.,* 2005).

In order to evaluate the ability of ecdysone and its intermediates to extend the lifespan of *Nnk*-null mutants or rescue the Nnk-null phenotype, I dissolved these sterols in yeast paste and fed the paste to first-instar *Nnk*-null larvae. I evaluated the ability of the supplemented diet to rescue the lethal phase of the larvae. I found that supplementing dietary cholesterol, 7DH (7-Dehydrocholesterol) and 20E (20-

Hydroxyecdysone) failed to rescue the *Nnk*-null larvae to adulthood (**Figure 3. 8).** I found that supplementation also failed to extend the lifespan or induce a molt in the *Nnk*-null larvae. Npc1a mutants were used as a positive control for the ecdysone feeding assay. Npc1a is required for sterol trafficking and the lethal phase of the mutants is significantly extended with the addition of supplemental dietary sterols or ecdysone (Fluegel *et al.*, 2006).

It is possible that the cause of *Nnk*-null lethality is the buildup of a toxic intermediate in the ecdysone biosynthesis pathway. While I cannot rule this possibility out, the fact that supplemental cholesterol, 7DH, and 20E do not shorten the *Nick Nack* null lifespan, as might be expected by mass action, suggest that this possibility is unlikely. Taken together, I concluded that the *Nnk*-null larvae arrest in development, and subsequent death is not due to defects in known steps in the ecdysone pathway**.**

I next turned our attention to the transcriptomic analyses of 261 heterochromatic

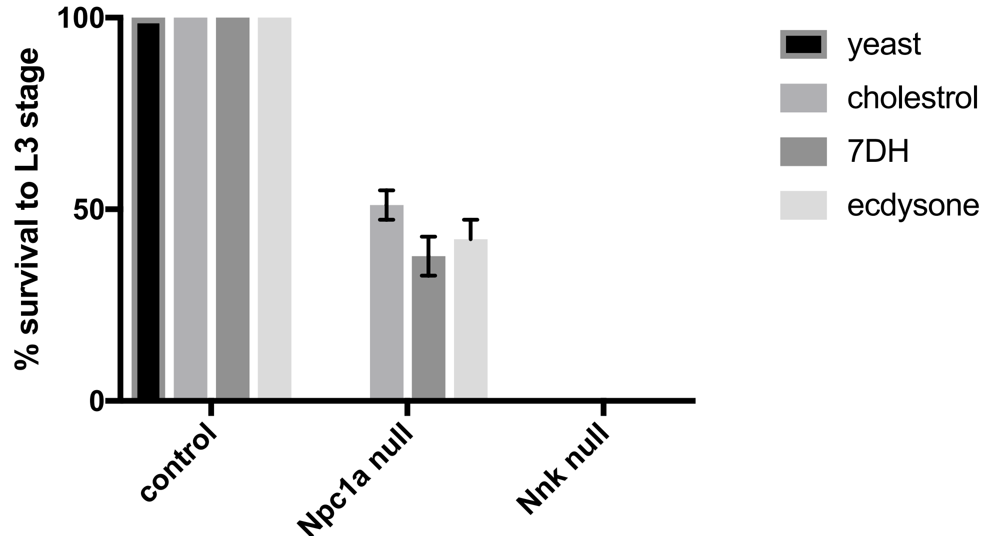


Figure 3.8 | Addition of dietary sterols has no effect on viability of *Nnk*-null larvae. Supplementing dietary sterols (cholesterol, 7DH or 20E) to the diets of control (w^{1118}) larvae did not alter their development to L3 larvae within 80 hours AEL. Npc1a null larvae as (*Npc1*^{a57}/ *Npc1*^{a57}) can be partially rescued (>40% molting into L3 larvae) with the addition of cholesterol, 7DH and 20E. In contrast, Nnk-null mutants fail to progress through development with the addition of dietary sterols. n >100 for all genotypes and treatments and graphs show mean of three replicates with error bars showing standard deviation.

genetic elements. These analyses were done with the help of Janet Young, a staff scientist in the lab. Briefly, we defined heterochromatic genes by collecting all genes that fall into heterochromatin as defined by the UCSC Table Browser (Karolchik *et al.*, 2004), FlyBase (http://flybase.org) and the *Drosophila* Heterochromatin Genome Project ((http://flybase.org/reports/FBrf0173307.html and http://flybase.org/reports/FBrf0199188.html). We first examined whether heterochromatic genes are globally mis-regulated in the *Nnk*-null larvae. We found that 24 hours AEL, there is a slight decrease in the expression of heterochromatin-embedded genes compared to euchromatic genes in *Nnk*-null larvae versus the control larvae (**figure 3.9a).** In contrast, at 48 hours AEL, there is a modest increase in the abundance of transcripts from heterochromatin-embedded genes compared to euchromatic genes in the *Nnk*-null larvae versus the controls (**figure 3.9b).** Indeed,

when we compared the levels of expression of heterochromatin-embedded genes in control versus *Nnk*-null larvae across the two time points, we found that there is a slightly decreased expression of heterochromatin-embedded genes compared to euchromatic genes in control larvae over time (**figure 3.9c).** In contrast, we found that *Nnk*-null larvae show no change in expression of heterochromatin embedded genes compared to euchromatic genes (**figure 3.9d).** We note that although there is a change in heterochromatic gene expression in the *Nnk*-null larvae, the difference is not statistically significant. There is a possibility that *Nnk*-null larvae may not be able to appropriately shut off a few, crucial heterochromatin-embedded genes during development.

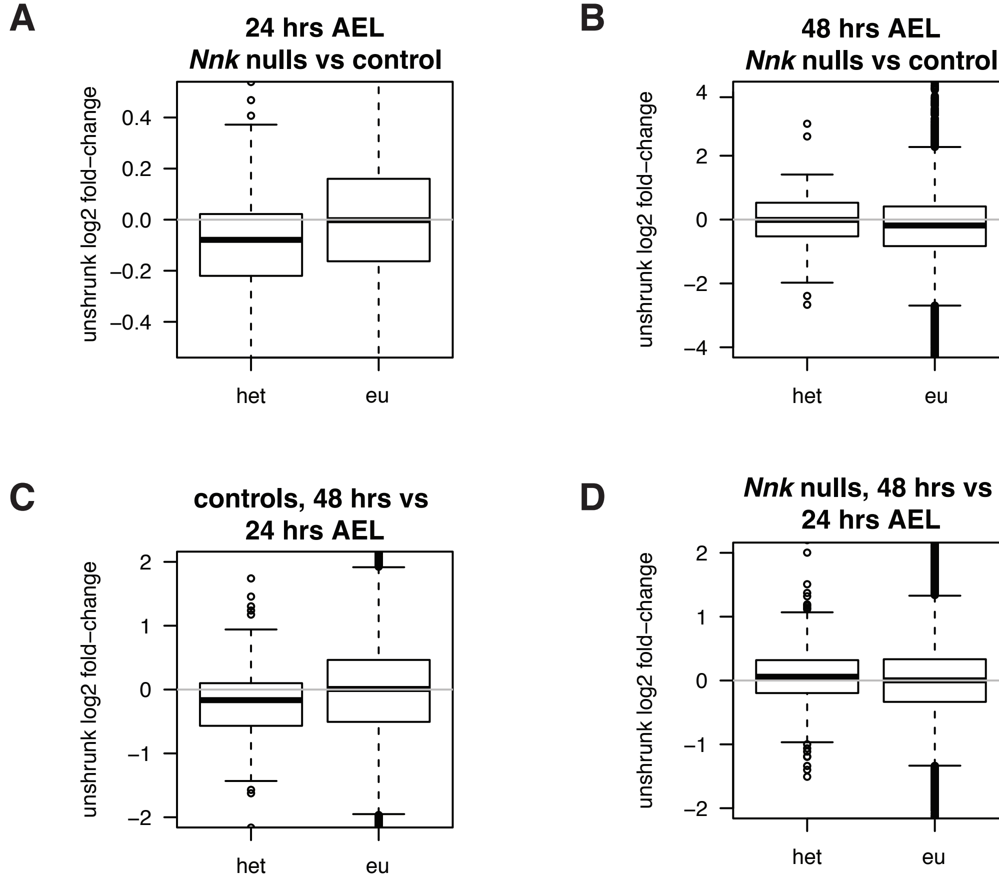


Figure 3.9 | Comparison of expressed heterochromatin genes in *Nnk*-null and control larvae across two timepoints. (a) 24 hours AEL, there is a modest decrease in the expression of the heterochromatin embedded genes (het) in the *Nnk*-null larvae compared to control larvae while there is no difference in euchromatic gene expression **(b)** 48 hours AEL there is an increase in the expression of heterochromatin embedded genes in the *Nnk*-null larvae compared to controls. **(c)** In control larvae, there is a decrease in expression of heterochromatin embedded genes over time. **(d)** In the *Nnk*-null larvae, there is a modest increase in the expression of heterochromatin embedded genes over time.

Based on the loss of repression of heterochromatin-embedded genes in *Nnk*-null larvae, I hypothesized that other heterochromatin-embedded elements may be incorrectly regulated in the *Nnk*-null mutants. In collaboration with Janet Young, we examined our transcriptomic dataset for elements enriched in heterochromatin, including repetitive non-coding (satellite) DNA and transposable elements. We found that the *Nnk*-null larvae mis-express various heterochromatin-embedded elements from 24 hours AEL to 48 hours AEL over time (**figure 3.10**), confirming that the null larvae mis-express not only heterochromatin-embedded genes but also other genetic elements embedded within heterochromatin.

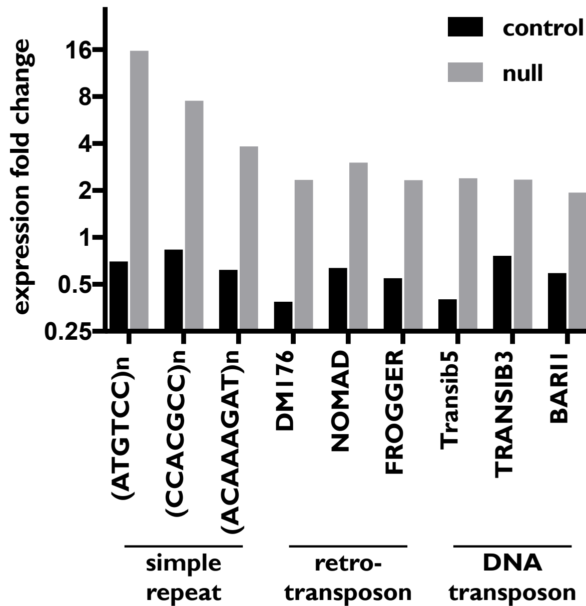


Figure 3.10 | Progressive derepression of heterochromatin-embedded elements in the *Nnk*-null larvae. Nnk-null larvae begin to express various elements embedded within heterochromatin over time including satellite DNA, DNA transposons and retrotransposons.

3.2.5 Nick Nack is a suppressor of position-effect variegation

The cytological analysis revealed Nick Nack protein localizes to specific foci within heterochromatin (see chapter 2). Furthermore, the transcriptomic analyses had suggested that Nnk-null larvae have a significant defect in maintaining silencing in the heterochromatic compartment. Therefore, I hypothesized that the cause of *Nnk*-null lethality is the result of heterochromatin dysfunction. I investigated whether *Nnk* has a defect in heterochromatin formation or maintenance by leveraging the position effect variegation (PEV) assay using a *brown dominant (bw^D)* allele (**figure 3.11a**)

(Schneiderman *et al.*, 2010), with the help of an undergraduate intern, Isabelle Hwang. In wild-type flies with the *bw^D* allele, the resulting eye color is mostly white with brown speckles (**figure 3.11b**); this variegated pattern reflects stochastic expression of the *brown* gene in a few cells while most cells are white because the *brown* gene is silenced due to being packaged into heterochromatin that resides near the mutant *brown* gene (Figure 3.5b). In fly eye cells lacking key heterochromatin proteins such as HP1A, there is an inability to form and maintain heterochromatin; therefore, *brown* is expressed in a majority of the cells, resulting in an overall darkly pigmented eye (**figure 3.11b**).

If *Nnk* were playing a silencing role in heterochromatin, *Nnk* knockdown in the eye would lead to loss of silencing of the heterochromatin-embedded *brown* reporter gene, resulting in a dark brown eye. If *Nnk* were necessary for euchromatin formation or for preventing the spread of heterochromatin, we would expect the opposite result- there would be an enhancement of the PEV phenotype, resulting in a whiter eye.

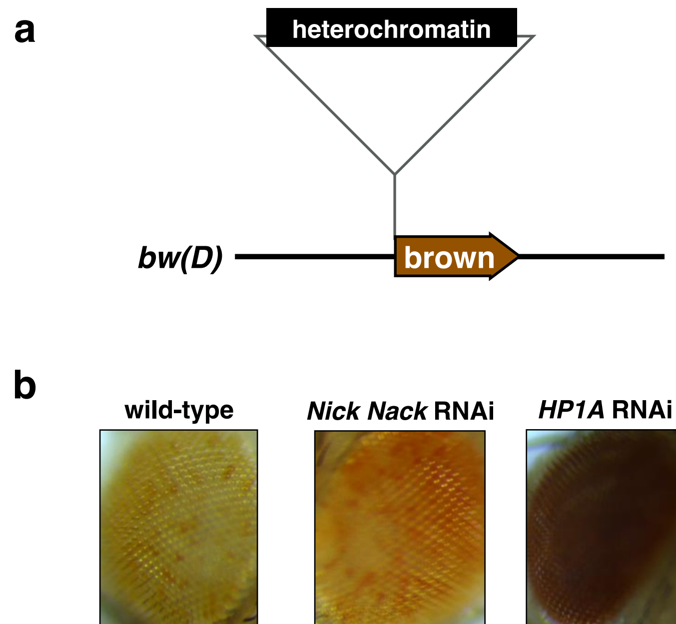


Figure 3.11 | *Nick Nack* is a suppressor of variegation. **(a)** schematic of *bw^D* allele showing heterochromatic insertion near *brown* gene. **(b)** *Nick Nack* knockdown in the eye (middle panel) leads to expression of the *brown* reporter gene in more cells than the wildtype eye (left panel). The right panel reflects expression of *brown* reporter in when *HP1A* is knocked down in the eye.

We find that when *Nnk* is depleted, the brown pigment is increased relative to wild type, suggesting a role for *Nnk* in formation or maintenance of heterochromatin

and supporting its role in heterochromatin-mediated silencing. We note this effect is subtler than the comparable loss of variegation observed in HP1A knockdown.

3.3 Discussion

ZAD-ZNFs are a large class of insect transcription factors that have undergone lineage specific expansions (Chung *et al.*, 2007a; Chung *et al.,* 2002). Previous studies reveal that three of these lineage-specific ZAD-ZNFs, Séance, Molting Defective and Ouija Board, are essential for *Drosophila* larval development and act through the transcriptional regulation of ecdysone biosynthetic enzymes (Harmes *et al.,* 2000; Komura-Kawa *et al.,* 2015; Uryu *et al.,* 2018). Larvae that cannot produce these specific ZAD-ZNFs arrest during development and can be rescued with dietary sterols or by overexpression of ecdysone biosynthetic genes that ZAD-ZNFs regulate.

I identified a ZAD-ZNF that is also necessary for larval development, a gene that I named *Nick Nack. Nick Nack* null organisms arrest in early larval development and die within three days of hatching without growing or molting. I found that *Nick Nack,* unlike the previously characterized ZAD-ZNFs, cannot be rescued by dietary ecdysone or any of its intermediates.

Some larval development mutants arrest because of an inability to absorb sterols (Huang *et al.*, 2005; Fluegel *et al.*, 2006). I found that the *Nick Nack*-null larvae are able to effectively absorb sterols; therefore, the cause of lethality in these mutants is not due to an inability to take up sterols from their diet. I also find that although *Nick Nack* null larvae are able to absorb sterols, they cannot be rescued by supplemental dietary sterols alone. This result suggests that inability to transcribe ecdysone biosynthetic enzymes is not the primary cause of the *Nick Nack* null lethality, a finding that distinguishes *Nick Nack* from other ZAD-ZNF transcription factors that are essential to larval development.

Instead, I find that *Nick Nack* null larvae are unable to turn on the transcriptional program of wild-type larvae 48 hours AEL, or appropriately turn off the previous transcriptional program. Strikingly, they fail to express numerous genes that are turned on 48 hours AEL and continue to express genes that are normally expressed at the 24 hours AEL stage. In addition, these *Nick Nack* null organisms transcribe various elements embedded within heterochromatin such as satellite repeats and transposable elements. The expression of these elements increases from 24 hours AEL to 48 hours AEL, suggesting that *Nick Nack's* primary or secondary target is in establishing or maintaining silencing within heterochromatin. Supporting this hypothesis, I found that Nnk has a role in establishing or maintaining heterochromatin as it acts as a suppressor of variegation.

In light of these results, I hypothesize that Nick Nack is directly or indirectly responsible for heterochromatin regulation in larval development. Loss of Nick Nack leads to global heterochromatin dysfunction, as exhibited by the PEV assay and by the progressive loss of repression of various elements embedded within heterochromatin. However, it is still unclear whether the heterochromatin defects I deduce from the transcriptomic analyses are causal to the larval lethality, or whether inappropriate regulation of one of the few hundred genes that *Nnk* controls are more likely to be responsible.

3.4 Methods

3.4.1 Determining Lethal Phase of Nnk-null mutants

I placed 50-75 flies heterozygous for the *Nnk*-null allele (*Nnk pBac null/TM3G)* or for the *Nnk* CRISPR allele (*Nnk CRISPR /TM3G)* into a small embryo collection cage containing a grape-juice plate with a thin strip of yeast paste and collected embryos for 3 hours at 25˚C. I transferred the larvae to fresh grape-juice plates containing yeast paste daily and scored developmental stage by mouth hook morphology. I used fluorescence to distinguish between heterozygotes (GFP-positive larvae) and homozygotes (GFP-negative larvae). For the transheterozygote evaluation, I crossed 30-40 virgin female *Nnk CRISPR/TM3G* to 10 *Nnk pBac null/TM3G* males. Crosses were done in triplicate and at least 100 progeny were counted per cross.

I made the grape-juice plates with 15g of agar (3% w/v), 125 mL grape juice, 1.5 g sucrose (0.3% w/v), and distilled water up to 500 mL. I combined the agar, grape juice, sucrose and water and heated the mixture in the microwave until the solution dissolved. The solution was allowed to cool to approximately 60˚C and tegosept (*para*-hydroxybenzoate) was added from a 10% stock solution to a final concentration of 0.03%. I poured the solution into 35 mm petri dishes and cooled before use.

3.4.2 Larvae collection for RNA-sequencing

I placed 50-75 flies heterozygous for the *Nnk*-null allele (*Nnk pBac/TM3G)* into a small embryo collection cage containing a grape-juice plate with a thin strip of yeast paste and collected embryos for 3 hours at 25˚C. The first time point was collected 24 hours AEL and the second 48 AEL. I transferred the larvae to fresh grape-juice plates containing yeast paste daily.

3.4.3 RNA extractions

Whole larvae (~30 animals at 24 hrs AEL and ~20 animals at 48 hrs AEL for each sample; RNA from each time point and genotype was prepared in triplicate) were ground in a 1.5 mL Eppendorf tube containing 50 uL of TRIzol reagent using a DNase, RNase and DNA free 1.5 mL pestle. 450 uL of TRIzol reagent was added after the grinding. Immediately, I added 500 uL of chloroform and the tube was inverted gently 2-3 times. I removed the aqueous phase into a fresh tube containing 1 mL of 200 proof EtOH and mixed by inversion. The mixture was then bound to Zymo-spin column according to the manufacturer's instructions (Zymo Research). I followed the DNase extraction and purification protocol outlined in the RNA Clean and Concentrator kit (Zymo Research). I eluted the RNA in 15 uL of DNase/RNase-free water and immediately placed the samples at -80°C. I checked the quality of the samples with 2200 Tapestation (Agilent Technologies) and selected samples that had an RNA integrity number > 9.0 for library preparation. Library construction and the 150-bp paired-end RNA-sequencing was conducted at Novogene Bioinformatics Technology Co., Ltd (Beijing, China).

3.4.4 Transcriptome Data Analysis

These analyses were done with the help of Janet Young, a staff scientist in the lab. We used Kallisto (Bray *et al.*, 2016) to quantify transcript abundances against the UCSC *D. melanogaster* reference transcriptome (refMrna.fa, which contains 34,114 transcripts). For each transcript, we acquired the gene name using R (org.Dm.eg.db). Kallisto counts were read into R and alternative splice-form counts were summarized into a single count per gene. We made comparisons between all samples (controls 24 hrs AEL, controls 48 hrs AEL, *Nick Nack*-null 24 hrs AEL and *Nick Nack*-null 48 hrs AEL). We did not include unexpressed genes (which we defined as <100 counts total across all

samples) in the analysis. We used DESeq2 to identify differentially expressed genes with adjusted p-value <=0.05 and the absolute log2fold change >=1 (Anders & Huber, 2010). Janet Young wrote a number of scripts in R for this analysis.

3.4.5 Dietary Sterol Supplementation

I followed the sterol absorption assay outlined by Uryu *et al.* 2018. To evaluate the ability for dietary sterols to rescue *Nnk*-null phenotype, we mixed together 20 mg of dry yeast in 38 uL of water. To this yeast paste, we added either 2 uL of EtOH (negative control) or 2 uL of EtOH plus cholesterol (Sigma), 7-dehydrocholestrol (Sigma) or 20-hydroxyecdysone. Stocks used for these experiments were balanced over GFP-balancers (*Npc1a57/CyO-GFP* and *Nnk-null/TM3G*). The control used for these experiments was the w^{1118} stock. Eggs were laid on yeasted grape-juice plates for 3 hours at 25˚C. 24 hours AEL, GFP-negative larvae were transferred onto grape plates containing fresh yeast paste at 25˚C. For 72 hours AEL, larvae were transferred to fresh grape-juice plates containing yeast paste daily and scored for their developmental stage based on the morphology of their mouth hooks. *Npc1a57/CyO-GFP* stocks were obtained from Leo Pallanck.

3.4.6 Defining Heterochromatic Genes

We used a broad definition of heterochromatic genes for our analysis. We downloaded the table of Dm3 genome through UCSC Table Browser (Karolchik *et al.,* 2004) and 190 heterochromatic genes fall within the UCSC genome browser definition, as denoted by falling within genes on "chromosomes" containing "Het" in the name (i.e.-chr2LHet). We downloaded lists of FlyBase gene IDs from FlyBase (http://flybase.org) from the 3.2 and 5.2 release of the *Drosophila* Heterochromatin Genome Project (http://flybase.org/reports/FBrf0173307.html and http://flybase.org/reports/FBrf0199188.html) and obtained 227 genes from release 3.2

and 463 genes from release 5.2. We defined a heterochromatic gene as a gene belonging to any of these three lists (261 genes in total).

3.4.7 Sterol Absorption Assay

I followed the method outlined by Voght *et al.* 2007. L1 larvae were collected from grape-juice plates containing yeast 24-30 hours AEL. Larval digestive tracts were dissected in PBS and fixed in 4% paraformaldehyde in PBS for 30 minutes. Tissues were subsequently washed three times in PBS. A freshly made 250 ug/mL stock solution of filipin (Sigma) in DMSO was diluted to 50 ug/mL. Tissues were stained in the dark for 30 minutes in the 50 ug/mL solution of Filipin then subsequently washed three times in PBS. Tissues were then placed in a 50% glycerol/PBS solution overnight at 4˚C. Then slides were mounted with SlowFade Gold Mounting Medium (ThermoFisher). I imaged the cells on a Leica TCS SP5 II confocal microscope with LASAF software and images were processed using ImageJ.

3.4.8 Position Effect Variegation Assay

The PEV fly line (*GMRGAL* bw^D*/CyO; st*) used in this assay and the experimental protocol was followed as described in Schneiderman *et al.* 2010. Briefly, *GMRGAL* bw^D*/CyO; st* virgin females were crossed to males homozygous for the RNAi hairpin specific to *Nick Nack* (VDRC 103-211). Progeny carrying both the *GMRGAL* bw^D and *Nick Nack* RNAi allele were collected and their eyes were photographed 3-4 days after eclosion. This assay was conducted with the assistance of Isabelle Hwang, a summer undergraduate intern.

Chapter 4

CONCLUSIONS & FUTURE PERSPECTIVES

The previous chapters examine how evolutionarily young genes can become necessary for development. In **chapter 2**, I took an evolutionary approach to identify rapidly evolving members of a large family of transcription factors, Zinc Finger Associated Domain Zinc Finger proteins (ZAD-ZNFs). In insects, the ZAD-ZNFs transcription factors have undergone multiple lineage-specific gene amplifications and losses (Chung *et al.*, 2002; Chung *et al.,* 2007). The cause of this ZAD-ZNF dynamism is unknown. Studies on individual ZAD-ZNFs have revealed that some members of this protein family are necessary for development or fertility (Chen *et al.,* 2000; Gaszner *et al.,* 1999; Lake *et al.*, 2011; Niwa & Niwa, 2016; Page *et al.*, 2005; Uryu *et al.,* 2018), but none have linked the function and evolution of these transcription factors. The expectation is that genes that encode essential functions are highly conserved and do not change much over time because mutations acquired in essential genes would be deleterious. Contrary to this expectation, I found that rapidly evolving ZAD-ZNFs are more likely to be essential than highly conserved ZAD-ZNFs.

To understand what factors could be driving the diversification of essential, rapidly evolving ZAD-ZNFs, we took a candidate locus approach and examined one cluster containing five ZAD-ZNFs in *D. melanogaster,* the *Odj-Nnk* cluster. The cluster contains two rapidly evolving, essential ZAD-ZNFs, *Oddjob (Odj)* and a gene we named *Nick Nack (Nnk).* We find that all the genes within the cluster are dynamically evolving across

Drosophila, with numerous gene gains and losses, recombination, and positive selection. Most *Odj-Nnk* cluster ZAD-ZNFs localize to heterochromatin, which is enriched for highly mobile and dynamic sequences, specifically transposable elements and satellite DNA. In addition, the two rapidly evolving members of this cluster, Oddjob and Nick Nack, are necessary for viability and fertility or early embryonic development. We hypothesize that the diversification of ZAD-ZNFs may be driven, in part, by their interaction with heterochromatin, which is a rapidly evolving part of the nucleus.

To examine the consequence of the rapid evolution of ZAD-ZNFs, we focused on one member of the *Odj-Nnk* cluster, *Nick Nack.* We found that *Nick Nack* arose an estimated 30 million years ago in the ancestor of *D. melanogaster* and *D. pseudoobscura,* has undergone multiple independent gains and losses since it first arose, and shows evidence for both recombination and recurrent-site specific rapid evolution. We find that over 10% of the protein is different between *D. melanogaster* and *D. simulans.* We hypothesized that the significant number of changes between the *Nick Nack* orthologs would make *D. simulans Nick Nack* non-functional in *D. melanogaster.* To test this hypothesis, we began by examining the localization of *D. simulans Nick Nack* in *D. melanogaster* cells. We found that, surprisingly, *D. simulans Nick Nack* is able to localize to heterochromatin in *D. melanogaster* cells, and that when both orthologs are co-transfected, they co-localize on heterochromatin. To determine whether this heterochromatin localization is associated with function, we evaluated the ability for the *D. simulans Nick Nack* to rescue the lethality caused by knocking down *Nick Nack* in *D. melanogaster.* We found that strikingly, *D. simulans Nick Nack* can rescue the lethality caused by *Nick Nack* ubiquitous RNAi depletion in females but male viability is significantly affected. We hypothesize that Nnk rapid

evolution results in Nnk specializing to the heterochromatin in its native genome. In the *D. simulans* Nnk rescue experiment, *D. simulans* Nnk is less efficient at performing its essential function. Consequently, males, which have an extra burden of heterochromatin carried on the Y chromosome, are significantly affected in this *D. simulans* rescue experiment.

Ongoing work in the lab is to pinpoint structural features of *Nick Nack* that facilitate its localization to heterochromatin and that link this localization with function. We hypothesize that heterochromatin localization is an essential property of Nick Nack. My preliminary studies show that *Nick Nack* does not localize to heterochromatin in *D. pseduoobscura,* suggesting that Nick Nack more recently acquired the ability to localize to heterochromatin. I hypothesize that *Nick Nack* began to localize to heterochromatin after the last loss event (in *D. elegans,* ~15-20 million years ago). Examining localization of *Nick Nack* before and after the most recent loss event can help pinpoint when *Nick Nack* neofunctionalized to localize to heterochromatin. Conducting domain swaps (ZAD, linker and zinc finger domains) between orthologs that are able to localize to heterochromatin and orthologs that cannot localize to heterochromatin will narrow down the region of the protein necessary for heterochromatin localization. Based on the positive selection analysis, I predict that the linker domain plays a role in the heterochromatin localization of Nick Nack. Evaluating the ability for these Nick Nack orthologs with differential heterochromatin/euchromatin localization to rescue lethality caused by Nick Nack knockdown in *D. melanogaster* can help clarify the role of heterochromatin localization in Nick Nack's essential function.

More broadly, future work examining the correlation between ZAD-ZNF rapid evolution and heterochromatin localization can help identify whether this is a common

theme for this dynamic protein family. For example, *Trade embargo (Trem)* is necessary for initiation of DNA double strand breaks and we find that it shows signatures of recent adaptive evolution across the gene as well as its linker region (Lake *et al.,* 2011). Mutagenesis of the ZAD and the C_2H_2 domains reveal that they are both necessary for function but the role or effect of the rapid evolution of the linker region of *Trade embargo* remains unknown. Evaluating the ability for *D. simulans Trem* to localize to meiotic chromosomes and rescue the fertility in *D. melanogaster Trem* mutants will reveal whether the rapid evolution of *Trem* is necessary for maintaining its interactions with chromatin and initiation of meiotic double strand breaks.

In **chapter 3**, I examined the function of *Nick Nack.* I found that *Nick Nack* is necessary for larval progression in *D. melanogaster.* I found that *Nick Nack* nulls arrest at the L1 stage and die without growing or molting within three days after they hatch. We find that *Nick Nack* null larvae fail to express a small set of euchromatic genes soon after hatching and that within 48 hours (when control larvae are at the L2 larval stage), *Nick Nack* null larvae continue to be transcriptionally more similar to the control L1 larvae than the control L2 larvae. I found that the null larvae are significantly enriched for mis-expressing genes involved with digestion, nutrient absorption and transport.

I evaluated the *Nick Nack* null organisms for their ability to absorb sterols from their diet. Specifically, dietary sterols are necessary for the synthesis of ecdysone, the master regulator of insect development. I found that *Nick Nack* null organisms are able to absorb sterols from their midgut so I hypothesized that they may have a defect in ecdysone biosynthesis. Indeed, some ZAD-ZNFs have been linked to regulation of ecdysone biosynthetic enzymes (Komura-Kawa *et al.,* 2015, Uryu *et al.*, 2018). These

ZAD-ZNF mutants can be rescued by supplementing ecdysone intermediates and active ecdysone, an approach that bypasses the need for the mutants to synthesize the hormone themselves. I found that in contrast to previously characterized ZAD-ZNF mutants, *Nick Nack* null organisms cannot be rescued with dietary ecdysone or its intermediates. In addition, supplemental dietary sterols had no effect on lifespan of these larvae.

I found that instead, *Nick Nack* null larvae express elements embedded within heterochromatin suggesting heterochromatin dysfunction. Corroborating this finding is that *Nick Nack* acts as a suppressor of variegation in a position-effect variegation assay, which suggests that the ZAD-ZNF has a role in establishing or maintaining heterochromatin.

The mutant analysis of the Nick Nack CRISPR allele, which contains a premature stop codon in the linker region, leads to an extension in the lifespan but not the development of these larvae. This result suggests that the zinc finger domains are essential for development of an adult fly. Serafin Colmeras' ongoing work in the Karpen lab is aimed at determining the Nnk-bound sites genome wide. Combining the *Nnk*-null transcriptomic dataset with the Nnk binding sites will help us determine the genes that Nick Nack directly regulates and can potentially help narrow down the mechanism by which Nick Nack helps with larval progression. In addition, by examining the conservation of Nick Nack bound sites/genes across *Drosophila* evolution, we can examine the co-evolution of this dynamic ZAD-ZNF and its regulatory sequences and target genes. I predict that species lacking Nick Nack will have altered regulatory sequences to control target genes or have lost these target genes all together.

Future work examining the co-evolution of ZAD-ZNFs and the sequences/genes they regulate can help shed light onto how transcription factor networks are rewired. For example, the ZAD-ZNFs involved in regulating ecdysone biosynthetic genes are *Drosophila* specific, and their targets are embedded within heterochromatin (Uryu *et al.,* 2018). It will be exciting to examine how other ecdysone-synthesizing species regulate these critical enzymes and whether they find similar solutions to transcriptional control—specifically, if they also utilize ZAD-ZNFs to regulate heterochromatin-embedded ecdysone biosynthetic enzymes.

Collectively, the work presented in this thesis use phylogenomic and genetic tools to uncover rapid evolving ZAD-ZNF transcription factors that are essential for viability and fertility. These studies also highlight the potential for using ZAD-ZNFs to examine the functional consequences of rapid evolution on developmental gene regulatory networks and genome architecture.

BIBLIOGRAPHY

Abe, Y., Sako, K., Takagaki, K., Hirayama, Y., Uchida, K. S. K., Herman, J. A., *et al.* (2016). HP1-Assisted Aurora B Kinase Activity Prevents Chromosome Segregation Errors. *Developmental Cell*, *36*(5), 487–497. http://doi.org/10.1016/j.devcel.2016.02.008

Anders, S., & Huber, W. (2010). Differential expression analysis for sequence count data. *Genome Biology*, *11*(10), R106. http://doi.org/10.1186/gb-2010-11-10-r106

Azzaz, A. M., Vitalini, M. W., Thomas, A. S., Price, J. P., Blacketer, M. J., Cryderman, D. E., *et al.* (2014). Human heterochromatin protein 1α promotes nucleosome associations that drive chromatin condensation. *The Journal of Biological Chemistry*, *289*(10), 6850–6861. http://doi.org/10.1074/jbc.M113.512137

Baumann, D. G., Dai, M.-S., Lu, H., & Gilmour, D. S. (2017). GFZF, a glutathione S-transferase protein implicated in cell cycle regulation and hybrid inviability, is a transcriptional co-activator. *Molecular and Cellular Biology*, *38*(4), e00476–17. http://doi.org/10.1128/MCB.00476-17

Bayes, J. J., & Malik, H. S. (2009). Altered heterochromatin binding by a hybrid sterility protein in Drosophila sibling species. *Science (New York, N.Y.)*, *326*(5959), 1538–1541. http://doi.org/10.1126/science.1181756

Begun, D. J., Holloway, A. K., Stevens, K., Hillier, L. W., Poh, Y.-P., Hahn, M. W., *et al.* (2007). Population genomics: whole-genome analysis of polymorphism and divergence in Drosophila simulans. *PLoS Biology*, *5*(11), e310. http://doi.org/10.1371/journal.pbio.0050310

Bergé, J. B., Feyereisen, R., & Amichot, M. (1998). Cytochrome P450 monooxygenases and insecticide resistance in insects. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *353*(1376), 1701–1705. http://doi.org/10.1098/rstb.1998.0321

Bowers, J. E., Chapman, B. A., Rong, J., & Paterson, A. H. (2003). Unravelling angiosperm genome evolution by phylogenetic analysis of chromosomal duplication events. *Nature*, *422*(6930), 433–438. http://doi.org/10.1038/nature01521

Bray, N. L., Pimentel, H., Melsted, P., & Pachter, L. (2016). Near-optimal probabilistic RNA-seq quantification. *Nature Biotechnology*, *34*(5), 525–527. http://doi.org/10.1038/nbt.3519

Brennecke, J., Aravin, A. A., Stark, A., Dus, M., Kellis, M., Sachidanandam, R., & Hannon,

G. J. (2007). Discrete small RNA-generating loci as master regulators of transposon activity in Drosophila. *Cell*, *128*(6), 1089–1103. http://doi.org/10.1016/j.cell.2007.01.043

Bridges, C. B. (1936). THE BAR “GENE” A DUPLICATION. *Science (New York, N.Y.)*, *83*(2148), 210–211. http://doi.org/10.1126/science.83.2148.210

Brunet, F. G., Roest Crollius, H., Paris, M., Aury, J.-M., Gibert, P., Jaillon, O., *et al.* (2006). Gene loss and evolutionary rates following whole-genome duplication in teleost fishes. *Molecular Biology and Evolution*, *23*(9), 1808–1816. http://doi.org/10.1093/molbev/msl049

Bujold, M., Gopalakrishnan, A., Nally, E., & King-Jones, K. (2010). Nuclear receptor DHR96 acts as a sentinel for low cholesterol concentrations in Drosophila melanogaster. *Molecular and Cellular Biology*, *30*(3), 793–805. http://doi.org/10.1128/MCB.01327-09

Chávez, V. M., Marqués, G., Delbecque, J. P., Kobayashi, K., Hollingsworth, M., Burr, J., *et al.* (2000). The Drosophila disembodied gene controls late embryonic morphogenesis and codes for a cytochrome P450 enzyme that regulates embryonic ecdysone levels. *Development (Cambridge, England)*, *127*(19), 4115–4126.

Chen, B., Harms, E., Chu, T., Henrion, G., & Strickland, S. (2000). Completion of meiosis in Drosophila oocytes requires transcriptional control by grauzone, a new zinc finger protein. *Development (Cambridge, England)*, *127*(6), 1243–1251.

Chen, L.-Y., Wang, J.-C., Hyvert, Y., Lin, H.-P., Perrimon, N., Imler, J.-L., & Hsu, J.-C. (2006). Weckle is a zinc finger adaptor of the toll pathway in dorsoventral patterning of the Drosophila embryo. *Current Biology*, *16*(12), 1183–1193. http://doi.org/10.1016/j.cub.2006.05.050

Chen, S., Zhang, Y. E., & Long, M. (2010). New genes in Drosophila quickly become essential., *330*(6011), 1682–1685. http://doi.org/10.1126/science.1196380

Chen, Z.-X., Sturgill, D., Qu, J., Jiang, H., Park, S., Boley, N., *et al.* (2014). Comparative validation of the D. melanogaster modENCODE transcriptome annotation. *Genome Research*, *24*(7), 1209–1223. http://doi.org/10.1101/gr.159384.113

Chung, H., Bogwitz, M. R., McCart, C., Andrianopoulos, A., ffrench-Constant, R. H., Batterham, P., & Daborn, P. J. (2007a). Cis-regulatory elements in the Accord retrotransposon result in tissue-specific expression of the Drosophila melanogaster insecticide resistance gene Cyp6g1. *Genetics*, *175*(3), 1071–1077. http://doi.org/10.1534/genetics.106.066597

Chung, H.-R., Löhr, U., & Jäckle, H. (2007b). Lineage-specific expansion of the zinc finger associated domain ZAD. *Molecular Biology and Evolution*, *24*(9), 1934–1943. http://doi.org/10.1093/molbev/msm121

Chung, H.-R., Schäfer, U., Jäckle, H., & Böhm, S. (2002). Genomic expansion and clustering of ZAD-containing C2H2 zinc-finger genes in Drosophila. *EMBO Reports*,

3(12), 1158–1162. http://doi.org/10.1093/embo-reports/kvf243

Cornelis, G., Funk, M., Vernochet, C., Leal, F., Tarazona, O. A., Meurice, G., *et al.* (2017). An endogenous retroviral envelope syncytin and its cognate receptor identified in the viviparous placental Mabuya lizard. *Proceedings of the National Academy of Sciences of the United States of America*, *114*(51), E10991–E11000. http://doi.org/10.1073/pnas.1714590114

Czech, B., Preall, J. B., McGinn, J., & Hannon, G. J. (2013). A transcriptome-wide RNAi screen in the Drosophila ovary reveals factors of the germline piRNA pathway. *Molecular Cell*, *50*(5), 749–761. http://doi.org/10.1016/j.molcel.2013.04.007

Delport, W., Poon, A. F. Y., Frost, S. D. W., & Kosakovsky Pond, S. L. (2010). Datamonkey 2010: a suite of phylogenetic analysis tools for evolutionary biology. *Bioinformatics (Oxford, England)*, *26*(19), 2455–2457. http://doi.org/10.1093/bioinformatics/btq429

Denner, J. (2016). Expression and function of endogenous retroviruses in the placenta. *APMIS : Acta Pathologica, Microbiologica, Et Immunologica Scandinavica*, *124*(1-2), 31–43. http://doi.org/10.1111/apm.12474

Devlin, R. H., Bingham, B., & Wakimoto, B. T. (1990). The organization and expression of the light gene, a heterochromatic gene of Drosophila melanogaster. *Genetics*, *125*(1), 129–140.

Distinct molecular evolutionary mechanisms underlie the functional diversification of the Wnt and TGFbeta signaling pathways. (2010). Distinct molecular evolutionary mechanisms underlie the functional diversification of the Wnt and TGFbeta signaling pathways. *Journal of Molecular Evolution*, *70*(4), 303–312. http://doi.org/10.1007/s00239-010-9337-z

Drinnenberg, I. A., Henikoff, S., & Malik, H. S. (2016). Evolutionary Turnover of Kinetochore Proteins: A Ship of Theseus? *Trends in Cell Biology*, *26*(7), 498–510. http://doi.org/10.1016/j.tcb.2016.01.005

Drosophila 12 Genomes Consortium, Clark, A. G., Eisen, M. B., Smith, D. R., Bergman, C. M., Oliver, B., *et al.* (2007). Evolution of genes and genomes on the Drosophila phylogeny. *Nature*, *450*(7167), 203–218. http://doi.org/10.1038/nature06341

Eberl, D. F., Duyf, B. J., & Hilliker, A. J. (1993). The role of heterochromatin in the expression of a heterochromatic gene, the rolled locus of Drosophila melanogaster. *Genetics*, *134*(1), 277–292.

El-Gebali, S., Mistry, J., Bateman, A., Eddy, S. R., Luciani, A., Potter, S. C., *et al.* (2019). The Pfam protein families database in 2019. *Nucleic Acids Research*, *47*(D1), D427–D432. http://doi.org/10.1093/nar/gky995

Feyereisen, R. (1999). Insect P450 enzymes. *Annual Review of Entomology*, *44*(1), 507–533. http://doi.org/10.1146/annurev.ento.44.1.507

Fluegel, M. L., Parker, T. J., & Pallanck, L. J. (2006). Mutations of a Drosophila NPC1 gene

confer sterol and ecdysone metabolic defects. *Genetics*, *172*(1), 185–196. http://doi.org/10.1534/genetics.105.046565

Gaszner, M., Vazquez, J., & Schedl, P. (1999). The Zw5 protein, a component of the scs chromatin domain boundary, is able to block enhancer-promoter interaction., *13*(16), 2098–2107.

Gilbert, L. I. (2004). Halloween genes encode P450 enzymes that mediate steroid hormone biosynthesis in Drosophila melanogaster. *Molecular and Cellular Endocrinology*, *215*(1-2), 1–10. http://doi.org/10.1016/j.mce.2003.11.003

Goriaux, C., Desset, S., Renaud, Y., Vaury, C., & Brasset, E. (2014a). Transcriptional properties and splicing of the flamenco piRNA cluster. *EMBO Reports*, *15*(4), 411–418. http://doi.org/10.1002/embr.201337898

Goriaux, C., Théron, E., Brasset, E., & Vaury, C. (2014b). History of the discovery of a master locus producing piRNAs: the flamenco/COM locus in Drosophila melanogaster. *Frontiers in Genetics*, *5*(R112), 257. http://doi.org/10.3389/fgene.2014.00257

Green, E. W., Fedele, G., Giorgini, F., & Kyriacou, C. P. (2014). A Drosophila RNAi collection is subject to dominant phenotypic effects. *Nature Methods*, *11*(3), 222–223. http://doi.org/10.1038/nmeth.2856

Grézy, A., Chevillard-Briet, M., Trouche, D., & Escaffit, F. (2016). Control of genetic stability by a new heterochromatin compaction pathway involving the Tip60 histone acetyltransferase. *Molecular Biology of the Cell*, *27*(4), 599–607. http://doi.org/10.1091/mbc.E15-05-0316

Groth, A. C., Fish, M., Nusse, R., & Calos, M. P. (2004). Construction of transgenic Drosophila by using the site-specific integrase from phage phiC31. *Genetics*, *166*(4), 1775–1782.

Harms, E., Chu, T., Henrion, G., & Strickland, S. (2000). The only function of Grauzone required for Drosophila oocyte meiosis is transcriptional activation of the cortex gene. *Genetics*, *155*(4), 1831–1839.

Helleu, Q., Gérard, P. R., Dubruille, R., Ogereau, D., Prud'homme, B., Loppin, B., & Montchamp-Moreau, C. (2016). Rapid evolution of a Y-chromosome heterochromatin protein underlies sex chromosome meiotic drive. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(15), 4110–4115. http://doi.org/10.1073/pnas.1519332113

Hervas, S., Sanz, E., Casillas, S., Pool, J. E., & Barbadilla, A. (2017). PopFly: the Drosophila population genomics browser. *Bioinformatics (Oxford, England)*, *33*(17), 2779–2780. http://doi.org/10.1093/bioinformatics/btx301

Huang, X., Suyama, K., Buchanan, J., Zhu, A. J., & Scott, M. P. (2005). A Drosophila

model of the Niemann-Pick type C lysosome storage disease: dnpc1a is required for molting and sterol homeostasis. *Development (Cambridge, England)*, *132*(22), 5115–5124. http://doi.org/10.1242/dev.02079

Inoue, A., Hyle, J., Lechner, M. S., & Lahti, J. M. (2008). Perturbation of HP1 localization and chromatin binding ability causes defects in sister-chromatid cohesion. *Mutation Research*, *657*(1), 48–55. http://doi.org/10.1016/j.mrgentox.2008.08.010

Jauch, R., Bourenkov, G. P., Chung, H.-R., Urlaub, H., Reidt, U., Jäckle, H., & Wahl, M. C. (2003). The zinc finger-associated domain of the Drosophila transcription factor grauzone is a novel zinc-coordinating protein-protein interaction module. *Structure (London, England : 1993)*, *11*(11), 1393–1402.

Jürgens, G., Wieschaus, E., Nüsslein-Volhard, C., & Kluding, H. (1984). Mutations affecting the pattern of the larval cuticle inDrosophila melanogaster : II. Zygotic loci on the third chromosome. *Wilhelm Roux's Archives of Developmental Biology*, *193*(5), 283–295. http://doi.org/10.1007/BF00848157

Karolchik, D., Hinrichs, A. S., Furey, T. S., Roskin, K. M., Sugnet, C. W., Haussler, D., & Kent, W. J. (2004). The UCSC Table Browser data retrieval tool. *Nucleic Acids Research*, *32*(Database issue), D493–6. http://doi.org/10.1093/nar/gkh103

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., *et al.* (2012). Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics (Oxford, England)*, *28*(12), 1647–1649. http://doi.org/10.1093/bioinformatics/bts199

Kellis, M., Birren, B. W., & Lander, E. S. (2004). Proof and evolutionary analysis of ancient genome duplication in the yeast Saccharomyces cerevisiae. *Nature*, *428*(6983), 617–624. http://doi.org/10.1038/nature02424

Klattenhoff, C., Xi, H., Li, C., Lee, S., Xu, J., Khurana, J. S., *et al.* (2009). The Drosophila HP1 homolog Rhino is required for transposon silencing and piRNA production by dual-strand clusters. *Cell*, *138*(6), 1137–1149. http://doi.org/10.1016/j.cell.2009.07.014

Komura-Kawa, T., Hirota, K., Shimada-Niwa, Y., Yamauchi, R., Shimell, M., Shinoda, T., *et al.* (2015). The Drosophila Zinc Finger Transcription Factor Ouija Board Controls Ecdysteroid Biosynthesis through Specific Regulation of spookier. *PLoS Genetics*, *11*(12), e1005712. http://doi.org/10.1371/journal.pgen.1005712

Kondo, S., Vedanayagam, J., Mohammed, J., Eizadshenass, S., Kan, L., Pang, N., *et al.* (2017). New genes often acquire male-specific functions but rarely become essential in Drosophila. *Genes & Development*, *31*(18), 1841–1846. http://doi.org/10.1101/gad.303131.117

Kosakovsky Pond, S. L., Posada, D., Gravenor, M. B., Woelk, C. H., & Frost, S. D. W. (2006). GARD: a genetic algorithm for recombination detection. *Bioinformatics (Oxford, England)*, *22*(24), 3096–3098. http://doi.org/10.1093/bioinformatics/btl474

Krystel, J., & Ayyanathan, K. (2013). Global analysis of target genes of 21 members of the ZAD transcription factor family in Drosophila melanogaster., *512*(2), 373–382. http://doi.org/10.1016/j.gene.2012.09.114

Lack, J. B., Lange, J. D., Tang, A. D., Corbett-Detig, R. B., & Pool, J. E. (2016). A Thousand Fly Genomes: An Expanded Drosophila Genome Nexus. *Molecular Biology and Evolution*, *33*(12), 3308–3313. http://doi.org/10.1093/molbev/msw195

Lake, C. M., Nielsen, R. J., & Hawley, R. S. (2011). The Drosophila zinc finger protein trade embargo is required for double strand break formation in meiosis. *PLoS Genetics*, *7*(2), e1002005. http://doi.org/10.1371/journal.pgen.1002005

Langley, C. H., Stevens, K., Cardeno, C., Lee, Y. C. G., Schrider, D. R., Pool, J. E., *et al.* (2012). Genomic variation in natural populations of Drosophila melanogaster. *Genetics*, *192*(2), 533–598. http://doi.org/10.1534/genetics.112.142018

Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H., *et al.* (2007). Clustal W and Clustal X version 2.0. *Bioinformatics (Oxford, England)*, *23*(21), 2947–2948. http://doi.org/10.1093/bioinformatics/btm404

Lavialle, C., Cornelis, G., Dupressoir, A., Esnault, C., Heidmann, O., Vernochet, C., & Heidmann, T. (2013). Paleovirology of "syncytins," retroviral env genes exapted for a role in placentation. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *368*(1626), 20120507–20120507. http://doi.org/10.1098/rstb.2012.0507

Lee, Y. C. G., Yang, Q., Chi, W., Turkson, S. A., Du, W. A., Kemkemer, C., *et al.* (2017). Genetic Architecture of Natural Variation Underlying Adult Foraging Behavior That Is Essential for Survival of Drosophila melanogaster. *Genome Biology and Evolution*, *9*(5), 1357–1369. http://doi.org/10.1093/gbe/evx089

Lespinet, O., Wolf, Y. I., Koonin, E. V., & Aravind, L. (2002). The role of lineage-specific gene family expansion in the evolution of eukaryotes., *12*(7), 1048–1059. http://doi.org/10.1101/gr.174302

Levine, M. T., McCoy, C., Vermaak, D., Lee, Y. C. G., Hiatt, M. A., Matsen, F. A., & Malik, H. S. (2012). Phylogenomic analysis reveals dynamic evolutionary history of the Drosophila heterochromatin protein 1 (HP1) gene family. *PLoS Genetics*, *8*(6), e1002729. http://doi.org/10.1371/journal.pgen.1002729

Levine, M. T., Vander Wende, H. M., & Malik, H. S. (2015). Mitotic fidelity requires transgenerational action of a testis-restricted HP1. *eLife*, *4*, e07378. http://doi.org/10.7554/eLife.07378

Levine, M. T., Vander Wende, H. M., Hsieh, E., Baker, E. P., & Malik, H. S. (2016). Recurrent Gene Duplication Diversifies Genome Defense Repertoire in Drosophila. *Molecular Biology and Evolution*, *33*(7), 1641–1653. http://doi.org/10.1093/molbev/msw053

Li, J., (null), (null), & Gilmour, D. S. (2013). Distinct mechanisms of transcriptional pausing orchestrated by GAGA factor and M1BP, a novel transcription factor. *The EMBO Journal*, *32*(13), 1829–1841. http://doi.org/10.1038/emboj.2013.111

Li, X., Schuler, M. A., & Berenbaum, M. R. (2007). Molecular mechanisms of metabolic resistance to synthetic and natural xenobiotics. *Annual Review of Entomology*, *52*(1), 231–253. http://doi.org/10.1146/annurev.ento.51.110104.151104

Liu, X., Song, Z., Huo, Y., Zhang, J., Zhu, T., Wang, J., *et al.* (2014). Chromatin protein HP1 interacts with the mitotic regulator borealin protein and specifies the centromere localization of the chromosomal passenger complex. *The Journal of Biological Chemistry*, *289*(30), 20638–20649. http://doi.org/10.1074/jbc.M114.572842

Long, M., VanKuren, N. W., Chen, S., & Vibranovski, M. D. (2013). New gene evolution: little did we know. *Annual Review of Genetics*, *47*(1), 307–333. http://doi.org/10.1146/annurev-genet-111212-133301

Marchler-Bauer, A., Bo, Y., Han, L., He, J., Lanczycki, C. J., Lu, S., *et al.* (2017). CDD/SPARCLE: functional classification of proteins via subfamily domain architectures. *Nucleic Acids Research*, *45*(D1), D200–D203. http://doi.org/10.1093/nar/gkw1129

McDonald, J. H., & Kreitman, M. (1991). Adaptive protein evolution at the Adh locus in Drosophila. *Nature*, *351*(6328), 652–654. http://doi.org/10.1038/351652a0

Muller, H. J. (1936). BAR DUPLICATION. *Science (New York, N.Y.)*, *83*(2161), 528–530. http://doi.org/10.1126/science.83.2161.528-a

Nambiar, M., & Smith, G. R. (2018). Pericentromere-Specific Cohesin Complex Prevents Meiotic Pericentric DNA Double-Strand Breaks and Lethal Crossovers. *Molecular Cell*, *71*(4), 540–553.e4. http://doi.org/10.1016/j.molcel.2018.06.035

Neubueser, D., Warren, J. T., Gilbert, L. I., & Cohen, S. M. (2005). molting defective is required for ecdysone biosynthesis. *Developmental Biology*, *280*(2), 362–372. http://doi.org/10.1016/j.ydbio.2005.01.023

Niwa, Y. S., & Niwa, R. (2016). Ouija board: A transcription factor evolved for only one target in steroid hormone biosynthesis in the fruit fly Drosophila melanogaster. *Transcription*, *7*(5), 196–202. http://doi.org/10.1080/21541264.2016.1210370

Okita, A. K., Zafar, F., Su, J., Weerasekara, D., Kajitani, T., Takahashi, T. S., *et al.* (2019). Heterochromatin suppresses gross chromosomal rearrangements at centromeres by repressing Tfs1/TFIIS-dependent transcription. *Communications Biology*, *2*(1), 17. http://doi.org/10.1038/s42003-018-0251-z

Ohno S. 1970. *Evolution by Gene Duplication*. New York: Springer-Verlag. 160 pp.

Page, A. R., Kovacs, A., Deak, P., Torok, T., Kiss, I., Dario, P., *et al.* (2005). Spotted-dick, a zinc-finger protein of Drosophila required for expression of Orc4 and S phase. *The EMBO Journal*, *24*(24), 4304–4315. http://doi.org/10.1038/sj.emboj.7600890

Page, A. W., & Orr-Weaver, T. L. (1996). The Drosophila genes grauzone and cortex are necessary for proper female meiosis. *Journal of Cell Science*, *109 (Pt 7)*, 1707–1715.

Parhad, S. S., Tu, S., Weng, Z., & Theurkauf, W. E. (2017). Adaptive Evolution Leads to Cross-Species Incompatibility in the piRNA Transposon Silencing Machinery. *Developmental Cell*. http://doi.org/10.1016/j.devcel.2017.08.012

Pond, S. L. K., & Frost, S. D. W. (2005). Datamonkey: rapid detection of selective pressure on individual sites of codon alignments. *Bioinformatics (Oxford, England)*, *21*(10), 2531–2533. http://doi.org/10.1093/bioinformatics/bti320

Rewitz, K. F., & Gilbert, L. I. (2008). Daphnia Halloween genes that encode cytochrome P450s mediating the synthesis of the arthropod molting hormone: evolutionary implications. *BMC Evolutionary Biology*, *8*(1), 60. http://doi.org/10.1186/1471-2148-8-60

Ross, B. D., Rosin, L., Thomae, A. W., Hiatt, M. A., Vermaak, D., la Cruz, de, A. F. A., *et al.* (2013). Stepwise evolution of essential centromere function in a Drosophila neogene., *340*(6137), 1211–1214. http://doi.org/10.1126/science.1234393

Ruiz-Estévez, M., Bakkali, M., Cabrero, J., Camacho, J. P. M., & López-León, M. D. (2014). HP1 knockdown is associated with abnormal condensation of almost all chromatin types in a grasshopper (Eyprepocnemis plorans). *Chromosome Research : an International Journal on the Molecular, Supramolecular and Evolutionary Aspects of Chromosome Biology*, *22*(3), 253–266. http://doi.org/10.1007/s10577-013-9399-z

Schertel, C., Albarca, M., Rockel-Bauer, C., Kelley, N. W., Bischof, J., Hens, K., *et al.* (2015). A large-scale, in vivo transcription factor screen defines bivalent chromatin as a key property of regulatory factors mediating Drosophila wing development., *25*(4), 514–523. http://doi.org/10.1101/gr.181305.114

Schneiderman, J. I., Goldstein, S., & Ahmad, K. (2010). Perturbation analysis of heterochromatin-mediated gene silencing and somatic inheritance. *PLoS Genetics*, *6*(9), e1001095. http://doi.org/10.1371/journal.pgen.1001095

Schuldiner, O., Berdnik, D., Levy, J. M., Wu, J. S., Luginbuhl, D., Gontang, A. C., & Luo, L. (2008). piggyBac-based mosaic screen identifies a postmitotic function for cohesin in regulating developmental axon pruning. *Developmental Cell*, *14*(2), 227–238. http://doi.org/10.1016/j.devcel.2007.11.001

Schulze, S. R., McAllister, B. F., Sinclair, D. A. R., Fitzpatrick, K. A., Marchetti, M., Pimpinelli, S., & Honda, B. M. (2006). Heterochromatic genes in Drosophila: a comparative analysis of two genes. *Genetics*, *173*(3), 1433–1445. http://doi.org/10.1534/genetics.106.056069

Schulze, S. R., Sinclair, D. A. R., Fitzpatrick, K. A., & Honda, B. M. (2005). A genetic and molecular characterization of two proximal heterochromatic genes on chromosome 3 of Drosophila melanogaster. *Genetics*, *169*(4), 2165–2177.

http://doi.org/10.1534/genetics.103.023341

Schumann, I., Kenny, N., Hui, J., Hering, L., & Mayer, G. (2018). Halloween genes in panarthropods and the evolution of the early moulting pathway in Ecdysozoa. *Royal Society Open Science*, *5*(9), 180888. http://doi.org/10.1098/rsos.180888

Scott, J. G., & Wen, Z. (2001). Cytochromes P450 of insects: the tip of the iceberg. *Pest Management Science*, *57*(10), 958–967. http://doi.org/10.1002/ps.354

Senti, K.-A., & Brennecke, J. (2010). The piRNA pathway: a fly's perspective on the guardian of the genome. *Trends in Genetics : TIG*, *26*(12), 499–509. http://doi.org/10.1016/j.tig.2010.08.007

Sinclair, D. A., Schulze, S., Silva, E., Fitzpatrick, K. A., & Honda, B. M. (2000). Essential genes in autosomal heterochromatin of Drosophila melanogaster. *Genetica*, *109*(1-2), 9–18.

Standard and generalized McDonald-Kreitman test: a website to detect selection by comparing different classes of DNA sites. (2008). Standard and generalized McDonald-Kreitman test: a website to detect selection by comparing different classes of DNA sites. *Nucleic Acids Research*, *36*(Web Server issue), W157–62. http://doi.org/10.1093/nar/gkn337

Sturtevant, A. H. (1925). The Effects of Unequal Crossing over at the Bar Locus in Drosophila. *Genetics*, *10*(2), 117–147.

Swenson, J. M., Colmenares, S. U., Strom, A. R., Costes, S. V., & Karpen, G. H. (2016). The composition and organization of Drosophila heterochromatin are heterogeneous and dynamic. *eLife*, *5*, e16096. http://doi.org/10.7554/eLife.16096

Uryu, O., Ou, Q., Komura-Kawa, T., Kamiyama, T., Iga, M., Syrzycka, M., *et al.* (2018). Cooperative Control of Ecdysone Biosynthesis in Drosophila by Transcription Factors Séance, Ouija Board, and Molting Defective. *Genetics*, *208*(2), 605–622. http://doi.org/10.1534/genetics.117.300268

Vermaak, D., & Malik, H. S. (2009). Multiple roles for heterochromatin protein 1 genes in Drosophila. *Annual Review of Genetics*, *43*(1), 467–492. http://doi.org/10.1146/annurev-genet-102108-134802

Vermaak, D., Henikoff, S., & Malik, H. S. (2005). Positive Selection Drives the Evolution of rhino, a Member of the Heterochromatin Protein 1 Family in Drosophila, *1*(1), e9–13. http://doi.org/10.1371/journal.pgen.0010009

Vernì, F., & Cenci, G. (2015). The Drosophila histone variant H2A.V works in concert with HP1 to promote kinetochore-driven microtubule formation. *Cell Cycle (Georgetown, Tex.)*, *14*(4), 577–588. http://doi.org/10.4161/15384101.2014.991176

Verschure, P. J., van der Kraan, I., de Leeuw, W., van der Vlag, J., Carpenter, A. E., Belmont, A. S., & van Driel, R. (2005). In vivo HP1 targeting causes large-scale chromatin condensation and enhanced histone lysine methylation. *Molecular and*

Cellular Biology, *25*(11), 4552–4564. http://doi.org/10.1128/MCB.25.11.4552-4564.2005

Vissers, J. H. A., Manning, S. A., Kulkarni, A., & Harvey, K. F. (2016). A Drosophila RNAi library modulates Hippo pathway-dependent tissue growth. *Nature Communications*, *7*(1), 10368. http://doi.org/10.1038/ncomms10368

Voght, S. P., Fluegel, M. L., Andrews, L. A., & Pallanck, L. J. (2007). Drosophila NPC1b promotes an early step in sterol absorption from the midgut epithelium. *Cell Metabolism*, *5*(3), 195–205. http://doi.org/10.1016/j.cmet.2007.01.011

Wakimoto, B. T., & Hearn, M. G. (1990). The effects of chromosome rearrangements on the expression of heterochromatic genes in chromosome 2L of Drosophila melanogaster. *Genetics*, *125*(1), 141–154.

Weaver, S., Shank, S. D., Spielman, S. J., Li, M., Muse, S. V., & Kosakovsky Pond, S. L. (2018). Datamonkey 2.0: a modern web application for characterizing selective and other evolutionary processes. *Molecular Biology and Evolution*, *35*(3), 773–777. http://doi.org/10.1093/molbev/msx335

Wieschaus, E., Nüsslein-Volhard, C., & Jürgens, G. (1984). Mutations affecting the pattern of the larval cuticle inDrosophila melanogaster : III. Zygotic loci on the X-chromosome and fourth chromosome. *Wilhelm Roux's Archives of Developmental Biology*, *193*(5), 296–307. http://doi.org/10.1007/BF00848158

Yang, Z. (1997). PAML: a program package for phylogenetic analysis by maximum likelihood. *Computer Applications in the Biosciences : CABIOS*, *13*(5), 555–556.

Yasuhara, J. C., & Wakimoto, B. T. (2006). Oxymoron no more: the expanding world of heterochromatic genes. *Trends in Genetics : TIG*, *22*(6), 330–338. http://doi.org/10.1016/j.tig.2006.04.008

Yasuhara, J. C., DeCrease, C. H., & Wakimoto, B. T. (2005). Evolution of heterochromatic genes of Drosophila. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(31), 10958–10963. http://doi.org/10.1073/pnas.0503424102

Zdobnov, E. M., Tegenfeldt, F., Kuznetsov, D., Waterhouse, R. M., Simão, F. A., Ioannidis, P., *et al.* (2017). OrthoDB v9.1: cataloging evolutionary and functional annotations for animal, fungal, plant, archaeal, bacterial and viral orthologs. *Nucleic Acids Research*, *45*(D1), D744–D749. http://doi.org/10.1093/nar/gkw1119

Zhang, D., Wang, D., & Sun, F. (2011). Drosophila melanogaster heterochromatin protein HP1b plays important roles in transcriptional activation and development. *Chromosoma*, *120*(1), 97–108. http://doi.org/10.1007/s00412-010-0294-5

Zhang, Y. E., Vibranovski, M. D., Landback, P., Marais, G. A. B., & Long, M. (2010). Chromosomal redistribution of male-biased genes in mammalian evolution with two bursts of gene gain on the X chromosome. *PLoS Biology*, *8*(10), e1000494. http://doi.org/10.1371/journal.pbio.1000494

Zhuang, X., Yang, C., Murphy, K. R., & Cheng, C.-H. C. (2019). Molecular mechanism and history of non-sense to sense evolution of antifreeze glycoprotein gene in northern gadids. *Proceedings of the National Academy of Sciences of the United States of America*, *116*(10), 4400–4405. http://doi.org/10.1073/pnas.1817138116

Zolotarev, N., Fedotova, A., Kyrchanova, O., Bonchuk, A., Penin, A. A., Lando, A. S., *et al.* (2016). Architectural proteins Pita, Zw5,and ZIPIC contain homodimerization domain and support specific long-range interactions in Drosophila. *Nucleic Acids Research*, *44*(15), 7228–7241. http://doi.org/10.1093/nar/gkw371