The Evolution of Sex Determination Mechanisms and Sex Chromosomes in House fly, *Musca domestica* L.

A Dissertation Presented to

the Faculty of the Department of Biology and Biochemistry

University of Houston

In Partial Fulfillment

of the Requirements for the Degree

Doctor of Philosophy

By

Jae Hak Son

May 2019

The Evolution of Sex Determination Mechanisms and Sex Chromosomes in House fly, *Musca domestica* L.

_____

Jae Hak Son

APPROVED:

_____

Dr. Richard P. Meisel, Chairman

_____

Dr. Brigitte Dauwalder

_____

Dr. Diane Wiernasz

_____

Dr. Rebecca Zufall

_____

Dr. Julia Saltz,
Rice University

_____

Dr. Dan Wells, Dean,
College of Natural Science and Mathematics

Acknowledgements

My special thanks go to Dr. Richard P. Meisel, Ph.D., for his guidance, encouragement, patience, and support throughout this endeavor. Without him, I could not reach to this point and have passion pursuing my career in academia.

My sincerest appreciation goes to my committee: Dr. Rebecca Zufall, Dr. Diane Wiernasz, Dr. Brigittte Dauwalder, and Dr. Julia Saltz. Thanks to their efforts, feedback, and criticisms for my research works over years, I could know how to become a scientist in this world.

I like to thank my collaborators Dr. Daniel Bopp from University of Zurich and Dr. Leo Beukeboom from University of Groningen. I am particularly grateful to Dr. Bopp for allowing me to conduct research in his laboratory and for his invaluable contribution. I also thank Tea Kohlbrenner, Svenia Heinze, and Claudia Brunner for their assistance during my stay in the Bopp lab.

I'd like to thank my lab members and the Kelleher lab members for their constructive criticism and comments for my works: Kiran Adhikari, Danial Asgari, Dr. Erin Kelleher, Shuo Zhang, Luyang Wang, and Jyoti Lama. Special thanks goes to a former lab member as technician, Christopher Gonzales, for spending his time on teaching me a lot of molecular works at the beginning of my PhD.

Lastly, I would like to thank my wife and son. My PhD and dissertation would not have been possible without her continuous support and encouragement

over many years. My seven months old son have enabled me to complete my PhD through his smile and happiness.

The Evolution of Sex Determination Mechanisms and Sex Chromosomes in House fly, *Musca domestica* L.

An Abstract of a Dissertation

Presented to

the Faculty of the Department of Biology and Biochemistry

University of Houston

In Partial Fulfillment

of the Requirements for the Degree

Doctor of Philosophy

By

Jae Hak Son

May 2019

Abstract

Sex determination (SD) evolves fast, often due to changes in master regulatory SD genes. SD genes are found on sex chromosomes, and evolution of SD can also drive evolutionary turnover of sex chromosomes. As a consequence, young (proto) sex chromosomes are generated, in which the X and Y are minimally diverged from each other. Multiple young sex chromosomes can coexist in species with multiple (polygenic) SD master regulators. Species with polygenic SD are informative of the factors driving the early evolution of sex chromosomes. The house fly, *Musca domestica*, is a well-suited model to those ends because natural populations have polygenic SD and young sex chromosomes. Natural selection appears to maintain polygenic SD, but the targets of selection are elusive. To address this, I examined the effects of two house fly proto-Y chromosomes (Y^M and III^M) on gene expression. I find that the proto-Y chromosomes have minor effects on gene expression, which is paradoxical given that natural selection likely maintains polygenic SD in house fly. I identified evidence for disproportionate effects of the proto-Y chromosomes on sex-biased expression. These few expression differences could be targets upon which natural selection acts to maintain polygenic SD. The frequencies of Y^M and III^M vary along latitudinal clines, and I tested whether temperature could explain these clines by examining the expression of SD pathway genes at different developmental temperatures. I did not find differences in expression of the sex-determining genes between Y^M and III^M males consistent with the clines. I also found that one house fly proto-Y chromosome is differentiated from its homologous proto-X in the sequences of many genes, but gene expression divergence between the proto-X and

proto-Y is limited to a subset of genes. This suggests that subtle gene expression differentiation constitutes the earliest stages of X-Y differentiation. Lastly, I tried constructing a stable line that has females carrying the dominant female-determining *Md-tra*^{D} allele. In this line, all females and males would carry two copies of III^{M}. However, I could not create a stable line because males homozygous for III^{M} could have low fitness or deleterious effects.

Table of Contents

Chapter 1

General Introduction

1.1 Fast-evolving master regulators of sex determination

Sex determination is the process by which an embryo develops into either a female or male, caused by genetic or environmental cues (Bull 1983; Graves 2008; Janousek and Mrackova 2010). The genetic signals discovered to date include a single sex determining chromosome/gene, multiple interacting sex determining genes, ploidy/heterozygosity, and genomic imprinting. The documented environmental cues include temperature, photoperiod, and population density. Although many organisms use sex determination to produce separate sexes, the sex determination mechanisms are not well conserved across taxa (Beukeboom and Perrin 2014). The fast evolution of sex determination is often due to changes in the top regulators of sex determination pathways (Wilkins 1995). The primary signal at the top of sex determination pathways can even be variable within a species, in what is known as polygenic or multifactorial sex determination (Bull 1983; Moore and Roberts 2013; Beukeboom and Perrin 2014). In addition, transitions between environmental and genetic sex determination can also occur. The diverged upstream components of sex determination pathways can be contrasted with the conserved downstream components across distantly related taxa, suggesting the bottom-up evolution in sex determination mechanisms (Wilkins 1995).

1.2 Models for transition of sex determination mechanisms

Several hypotheses have been proposed to explain evolutionary transitions at the top of sex determination pathways. First, a new sex determiner can invade a population if it is itself beneficial or linked to a beneficial allele (van Doorn and Kirkpatrick 2007, 2010). Sexually antagonistic selection, which acts on traits or mutations beneficial to one sex but

deleterious to the other sex, might be an important driving force for transitions at the top of sex determination pathways—the new sex determining variant could resolve the sexual conflict by allowing a sexually antagonistic allele to be inherited only in the sex in which it is beneficial (van Doorn and Kirkpatrick 2007, 2010; Roberts *et al.* 2009). Second, sex-ratio selection may drive the invasion of a new sex-determining factor if a population has a non-equilibrium sex ratio (Bulmer and Bull 1982) or sex determination mechanisms are out of equilibrium (van Doorn 2014a). The rarer sex has a selective advantage because the rarer sex is likely to have a larger number of available mates. The new-sex determining factor can equalize the biased sex ratio. Lastly, the evolutionary turnover of sex determination can occur through a set of neutral intermediate states in which polygenic sex determination exists (Bull and Charnov 1977; Saunders *et al.* 2018).

1.3 Polygenic sex determination

In species with genetic sex determination, the master sex determining factors can be polygenic within species or monogenic (Moore and Roberts 2013). In monogenic sex determination, a single genetic locus acts as a master regulator of the sex determination pathway. In polygenic sex determination, there are multiple loci or alleles involved in the master regulation of the sex determination pathway. Polygenic sex determination can occur through functional variation at a single sex-determining locus, or if autosomal loci gain new abilities to regulate sexual development (Moore and Roberts 2013). Although polygenic sex determination was previously considered to happen very rarely or dismissed as odd exceptions, polygenic sex determination has been recently found in numerous species (Orzack *et al.* 1980; Moore and Roberts 2013; Bachtrog *et al.* 2014).

However, population genetic models predict that polygenic sex determination should be transient between monogenic sex determination systems (Rice 1986; van Doorn 2014b). Despite this prediction, polygenic sex determination systems in some species appear to be stable, with the frequencies of sex determiners unchanged over many generations (Kozielska *et al.* 2006; Meisel *et al.* 2016). It remains elusive what genetic or environmental factors affect selection pressures to maintain stable polygenic sex determination. Characterizing the selection pressures that maintain polygenic sex determination would illuminate our understanding of the evolution of sex determination.

1.4 Turnover and differentiation of sex chromosomes

Sex determining genes are often found on sex chromosomes. The most commonly known sex chromosomes are X and Y chromosomes in XX/XY male heterogametic system, and Z and W chromosomes in ZZ/ZW female heterogametic system. Sex chromosome transitions can occur between XX/XY and ZZ/ZW systems or within XX/XY and ZZ/ZW systems, creating new sex chromosomes (Bachtrog *et al.* 2014). The turnover of sex determination enables autosomes to obtain sex-determining loci, which means that autosomes can evolve into new (very young) sex chromosomes (D. Charlesworth *et al.* 2005). When they first arise, this new pair of sex chromosomes (for example, one X and one Y chromosome in an XY system or one Z and one W chromosomes in a ZW system) are nearly identical in sequence and gene content, and they diverge from each other through evolutionary time (Bull 1983).

Despite the independent origins of sex chromosomes among different taxa, independently derived sex chromosomes have similar features, implying that the

independent sex chromosomes experience similar evolutionary trajectories (B. Charlesworth 1996; Bachtrog 2013). For instance, in XY sex determination systems, "masculinization" of Y chromosome gene content occurs because male-limited inheritance of the genes on the Y chromosome favors the fixation of male-beneficial genetic variation (Rice 1996). Recombination of a Y (or W) chromosome with its homologous X (or Z) chromosomes is often suppressed around the sex-determining locus to maintain genetic linkage between the sex determiner and sexually antagonistic alleles (Bull 1983; Bachtrog and Charlesworth 2002; Bachtrog *et al.* 2014). This is predicted to lead to "degeneration" of the Y chromosome because suppressed recombination inhibits the purging of deleterious mutations in Y genes (Bachtrog 2013). A similar phenomenon is expected in a ZW system as well. Sex chromosomes of a variety of ages (from young to old) have been studied to characterize these features. However, the very first stages of sex chromosome evolution are poorly understood because of a lack of extremely young sex chromosome systems.

1.5 House fly sex determination and sex chromosomes.

The house fly, *Musca domestica* L., has polygenic sex determination and very young sex chromosomes, making it a good model system for studying the early evolution of a new sex determining mechanism and sex chromosomes (Dübendorfer *et al.* 2003). A dominant male-determining factor has been found on all five autosomal chromosomes (A^M; I-V), the Y chromosome (Y^M), and even on the X chromosome (X^M) in natural populations (Hamm *et al.* 2015). The male-determining factor was recently identified and named *Mdmd* (*Musca domestica male determiner*). *Mdmd* is a paralog of the generic

splice factor gene *CWC22*, also known as *ncm* (Sharma *et al.* 2017). *Mdmd* regulates splicing of the house fly ortholog of *transformer* (*Md-tra*), located on autosome IV, by preventing males from producing a functional female-determining isoform of *Md-tra* (Hediger *et al.* 2010; Sharma *et al.* 2017). A dominant (female-determining) allele, *Md-tra*^{D}, that is not sensitive to *Mdmd* negative regulation also segregates in natural populations, allowing females to carry *Mdmd* (Mcdonald *et al.* 1978; Kozielska *et al.* 2006; Hediger *et al.* 2010).

Md-tra plays an important role as the key switch in the house fly sex determination pathway because this gene is alternatively spliced between males and females in response to regulation by *Mdmd* (Hediger *et al.* 2010). The maternal deposition of functional Md-TRA proteins or *Md-tra* mRNA is required to activate zygotic *Md-tra*, which initiates an auto-regulatory loop to maintain female-determining splicing of *Md-tra*, promoting female development (Dübendorfer *et al.* 2003; Bopp 2010). *Md-tra* is known to regulate two downstream genes in the sex determination pathway: the house fly ortholog of *doublesex* (*Md-dsx*; Hediger et al. 2004, 2010) and the ortholog of *fruitless* (*Md-fru*; Meier et al. 2013). *Md-dsx* is involved in morphological sexual dimorphism and *Md-fru* regulates male sexual behavior. Orthologs of *transformer* are the most upstream component of the sex determination pathway conserved across holometabolous insects, including Diptera, Hymenoptera, and Coleoptera (Verhulst *et al.* 2010).

There are minimal morphological and sequence differences, as well as similar gene content, between X and Y^{M} chromosomes in house fly (Boyes *et al.* 1964; Schmidt

et al. 1997; Meisel *et al.* 2017), suggesting that the house fly Y^M chromosome is minimally degenerated and has little divergence from its homologous X chromosome. The third chromosome with *Mdmd* (III^M) is also recently derived from the standard third chromosomes (Meisel *et al.* 2017). I, therefore, refer to any chromosomes bearing *Mdmd* as a "proto-Y" chromosome. The most frequently found proto-Y chromosomes in natural populations are Y^M and III^M, which form stable latitudinal clines over multiple continents (Franco *et al.* 1982; Tomita and Wada 1989; Hamm *et al.* 2005; Kozielska *et al.* 2008; Hamm *et al.* 2015). The clines of Y^M and III^M are predicted by seasonality of temperature, suggesting that there might be phenotypic effects depending on proto-Y chromosomes and temperature (Feldmeyer *et al.* 2008). Although populations have different frequencies of *Mdmd*-bearing proto-Y chromosomes (e.g. Y^M vs III^M), the frequencies of the proto-Y chromosomes within natural populations have not changed for decades (Kozielska *et al.* 2008; Meisel *et al.* 2016), suggesting that house fly has stable polygenic sex determination across natural populations.

1.6 Dissertation outline

The research in this doctoral dissertation addresses unresolved questions about the evolution of sex determination and sex chromosomes using house fly as a model system. Understanding the evolutionary forces that allow for the stable maintenance of polygenic sex determination would help shed light on the rapid evolution of sex determination. Identifying how young proto-Y and X chromosomes have differentiated at the very earliest stages would contribute to understanding why sex chromosomes have similar features across taxa.

In Chapter 2, I investigate the expression profiles of males carrying different proto-Y chromosomes that could be responsible for phenotypic differences between Y^M and III^M males. To measure gene expression, I generated and analyzed RNA-seq data in two experiments. The first data set comes from four strains with males isogenic except for the proto-Y chromosomes. I use these data to test the effects of the different proto-Y chromosomes on gene expression. The second experiment uses one III^M strain subjected to an *Md-tra* RNA interference (RNAi) treatment that causes genotypic females to be sex-reversed into phenotypic males. I use RNA-seq to examine the effects of *Mdmd* and the proto-Y chromosomes on gene expression in these flies. My finding in this chapter is that the proto-Y chromosomes have a minor effect on male expression profiles.

In Chapter 3, I examine the effects of different proto-Y chromosomes and developmental temperatures on the splicing of *Md-tra*. This work tests the hypothesis that temperature-dependent phenotypic effects of *Mdmd* located on different chromosomes (e.g., Y^M and III^M males) explain stable Y^M-III^M north-south clines found in natural populations (Hamm *et al.* 2005; Kozielska *et al.* 2008). Mechanistically this could be possible if the expressional abundance of *Mdmd* varies across different developmental temperatures and Y^M/III^M genotypes, and if the abundance of MDMD protein affects the fidelity of *Md-tra* splicing. To quantify splicing isoforms of *Md-tra* and expression of *Mdmd*, I used quantitative reverse transcription PCR (qRT-PCR) in samples with different genotype-by-temperature combinations. I find that *Md-tra* splicing and *Mdmd* expression do not show a consistent interaction between genotype (Y^M and III^M) and developmental temperature (18 °C and 27 °C), suggesting that temperature-dependent

expression of *Mdmd* or splicing of *Md-tra* cannot explain the clinal distribution or temperature-dependent fitness effects of Y^M and III^M.

In Chapter 4, I characterize the very early evolution of sex chromosomes by identifying heterozygosity on house fly sex chromosomes and quantifying allele-specific expression. I used RNA-seq data produced in the *Md-tra* RNAi experiment in Chapter 2. This allowed me to compare sex-reversed males with no proto-Y chromosome and normal males with the III^M proto-Y chromosome. I use this approach to test for differentiation of the proto-Y chromosome from its homologous proto-X chromosome in terms of DNA sequence and gene expression. I find that males with the III^M proto-Y chromosome exhibit elevated heterozygosity on the third chromosomes relative to sex-reversed males with no proto-Y. I also find a higher fraction of genes with allele-specific expression (ASE) in the genotypic males and no ASE in the sex-reversed males on the third chromosome, which does not appear in the other chromosomes. These results suggest that the early divergence of very young X and Y chromosomes is achieved through the evolution of DNA sequence and gene expression.

The fifth and final chapter of my PhD dissertation describes my attempts to establish a stable *Md-tra*^{D} line of house fly using a crossing scheme that involves multiple house fly strains. The goal was to create a line in which all females carry *Md-tra*^{D} and all females and males have two copies of the III^M chromosome. However, I was not able to create a stable *Md-tra*^{D} line because it was too difficult to create males that are homozygous for III^M. In addition, a red eye allele, used as a phenotypic marker in the cross because it is genetically linked to *Md-tra*^{D}, frequently recombined away from *Md-*

tra^{D}, making it difficult to keep track of the *Md-tra*^{D} allele based on the red eye phenotype.

Chapter 2

Minimal effects of proto-Y chromosomes on house fly gene expression in spite of evidence that selection maintains stable polygenic sex determination

Jae Hak Son[1], Tea Kohlbrenner[2], Svenia Heinze[2], Leo Beukeboom[3], Daniel Bopp[2], Richard P. Meisel[1]

1. Department of Biology and Biochemistry, University of Houston, USA

2. Institute of Molecular Life Sciences, University of Zurich, Switzerland

3. Groningen Institute for Evolutionary Life Sciences, University of Groningen, The Netherlands

This chapter is submitted to *Genetics*, and this manuscript is currently in review.

(Submission date: Feb 14, 2019)

A pre-print is available at: https://doi.org/10.1101/545178

2.1 Introduction

Sex determination is the process by which genetic or environmental cues cause an individual to develop into either a female or male. Sex determination evolves rapidly, often due to changes in the master regulatory genes at the top of sex determining pathways (Beukeboom and Perrin 2014). Sex determining pathways can also be variable (polygenic or multifactorial) within species (Moore and Roberts 2013). Many population genetic models predict that polygenic sex determination should be a transient state between monogenic equilibria (Rice 1986; van Doorn and Kirkpatrick 2007). It is therefore surprising that polygenic sex determination has been found in numerous species (Orzack *et al.* 1980; Moore and Roberts 2013; Bachtrog *et al.* 2014). Understanding how polygenic sex determining systems are maintained will help shed light on the forces driving the rapid evolution in sex determination pathways.

The house fly, *Musca domestica*, is a model species to study polygenic sex determination because it has a well characterized and highly variable sex determination system (Dübendorfer *et al.* 2003; Hamm *et al.* 2015). The male-determining gene, *Mdmd*, appears to be recently derived in house fly as it is absent in its close relative *Stomoxys calcitrans*, and it cannot be found in other related dipterans (Sharma *et al.* 2017). *Mdmd* regulates the splicing of the house fly ortholog of *transformer* (*Md-tra*), preventing males from producing a functional female-determining isoform of *Md-tra* (Hediger *et al.* 2010; Sharma *et al.* 2017). A dominant female-determining allele ($Md\text{-}tra^{D}$) that is not sensitive to *Mdmd* regulation also segregates in natural populations, allowing females to carry *Mdmd* (Mcdonald *et al.* 1978; Kozielska *et al.* 2008; Hediger *et al.* 2010).

Mdmd can be found on multiple different chromosomes in natural populations of house fly (Sharma *et al.* 2017). The two most common locations of *Mdmd* in natural populations are the Y chromosome (Y^M) and third chromosome (III^M) (Hamm *et al.* 2015). Historically, the chromosomes carrying the male determiner were designated as the Y (Y^M), X (X^M), and any of the five autosomes (e.g., III^M). However, recent work showed that the Y^M chromosome is highly similar in gene content to the X chromosome, and therefore Y^M is a very young proto-Y chromosome (Meisel *et al.* 2017). These findings align with the independent discovery that *Mdmd* is of a recent origin (Sharma *et al.* 2017). Moreover, previous work observed minimal morphological and sequence differences between the X and Y chromosomes (Boyes *et al.* 1964). The third chromosome carrying *Mdmd* is also very recently derived from the standard third chromosome (Meisel *et al.* 2017). I therefore refer to any chromosome carrying *Mdmd* (including Y^M and III^M) as a proto-Y chromosome.

There are multiple lines of evidence that natural selection maintains polygenic sex determination in house fly. First, Y^M and III^M form stable latitudinal clines on multiple continents (Franco *et al.* 1982; Tomita and Wada 1989; Hamm *et al.* 2005; Kozielska *et al.* 2008). Y^M is most frequent in northern populations, and III^M predominates in southern populations. The distributions of the proto-Y chromosomes correlate with seasonality in temperature (Feldmeyer *et al.* 2008), suggesting that temperature modulates the fitness of males carrying different proto-Y chromosomes. Second, males carrying Y^M or III^M differ in their success courting female mates, and the frequency of the III^M chromosome reproducibly increased over generations in laboratory population cages kept at a warm

temperature (Hamm *et al.* 2009). Third, in some populations, individual males carry multiple proto-Y chromosomes, which would cause them to produce male-biased broods with their mates (Kozielska *et al.* 2006; Hamm *et al.* 2015). The frequency of males that carry multiple proto-Y chromosomes is positively correlated with the frequency of *Md-tra*^{D} across populations (Meisel *et al.* 2016). This suggests that *Md-tra*^{D} invaded to balance the sex ratio or *Md-tra*^{D} allows for the increase in frequency of proto-Y chromosomes.

If selection maintains polygenic sex determination in house fly, Y^{M} and III^{M} must have different phenotypic and fitness effects for selection to act upon. However, a recent analysis of Y^{M} and III^{M} sequences revealed very few differences from their homologous X and III chromosomes, respectively (Meisel *et al.* 2017). To examine this paradox of evolutionarily important phenotypic effects of proto-Y chromosomes yet minimal sequence divergence from their homologs, I used high throughput mRNA sequencing (RNA-seq) to measure gene expression in house flies with different Y^{M} and III^{M} genotypes. This included testing the effects of multiple different naturally occurring Y^{M} and III^{M} chromosomes on a common genetic background. I also used RNA interference (RNAi) to knock down *Md-tra* and create sex-reversed males that do not carry any proto-Y chromosomes (Hediger *et al.* 2010), which I compared to males with the same genetic background carrying III^{M}. My experiments therefore allow us to determine the phenotypic effects of the Y^{M} and III^{M} chromosome on common genetic backgrounds to test the hypothesis that natural selection acts on phenotypic differences between males carrying different proto-Y chromosomes.

2.2 Materials and Methods

2.2.1 Strains with naturally occurring proto-Y chromosomes

I examined gene expression in four house fly strains that each have a different naturally occurring proto-Y chromosome (either Y^M or III^M) on a common genetic background (Figure 2.1). To put different Y^M and III^M chromosomes on a common genetic background, I used a previously described backcrossing method (Meisel *et al.* 2015). The common background was from the Cornell Susceptible (CS) strain, an inbred III^M strain produced by mixing strains collected from throughout the United States (Scott *et al.* 2014). My first proto-Y chromosome is the CS III^M on its native background. The second strain (CSrab) was created by backcrossing the III^M chromosome from the rspin strain collected in New York (Shono and Scott 2003) onto the CS background, replacing the CS III^M chromosome. The third strain (IsoCS) is a Y^M strain that was previously created by introducing the Y^M chromosome from a strain collected in Maine onto the CS background without III^M (Hamm *et al.* 2009). The fourth strain was created to test the effect of a non-*Mdmd*-bearing third chromosome on gene expression. To that end, I introduced the third chromosome carrying the recessive *brown body* mutation (*bwb*) and the Y^M chromosome from the genome reference strain (aabys) onto the CS background to create the bwbCS strain (III^{bwb}/ III^{bwb}; X/Y^M). I then crossed bwbCS males with CS females (bwbCS×CS) to create males that carry the aabys Y^M chromosome and are heterozygous for the non-*Mdmd* third chromosomes from CS and aabys on a CS background (III^{CS}/ III^{bwb}; X^{CS}/Y^M). I therefore have two III^M strains (CS and CSrab) with different origins of the III^M chromosome and two Y^M strains (IsoCS and bwbCS×CS)

with different origins of the Y chromosome. In three of the strains (CS, CSrab, and IsoCS), females are isogenic for the CS background and males are isogenic except for their *Mdmd*-bearing proto-Y chromosomes.

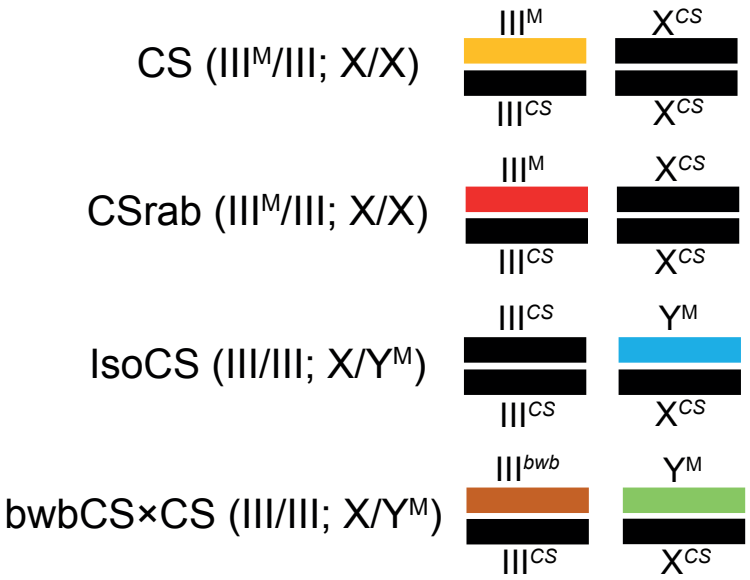


Figure 2.1. Four strains that have different naturally occurring Y^M or III^M proto-Y chromosomes on a common genetic background. Black bars represent chromosomes used as a common genetic background and colored bars are chromosomes that are replaced on that background. Different colors of chromosomes indicate the chromosomal origins from different strains. Chromosomes in the rest of genome (not shown), are from the common genetic background as well.

2.2.2 RNAi knockdown to create sex-reversed flies

I used RNAi targeting *Md-tra* to create sex-reversed males that do not carry a male-determining proto-Y chromosome. For the RNAi experiments, I used a house fly strain that allows easy identification of sex-reversed individuals that are genotypic females but phenotypic males (Hediger *et al.* 2010). Females of this strain are homozygous for a third chromosome containing recessive alleles for *pointed wing* (*pw*) and *bwb*. Males carry one copy of the third chromosome with *pw* and *bwb*, and one copy of a III^M chromosome

with wild-type alleles (*Mdmd* pw^+ bwb^+/*pw bwb*). Females, therefore, have pointed wings and brown bodies, as do sex-reversed males, whereas normal males have wild-type wings and wild-type bodies.

Double-stranded RNA (dsRNA) targeting *Md-tra* (*Md-tra*-RNAi) and GFP (GFP-RNAi) was generated and injected into early blastoderm embryos of the *pw bwb* strain following established protocols (Hediger *et al.* 2001, 2004). The fragment of dsRNA targeting *Md-tra* ranges from exon 1 to exon 5, and it was generated by amplifying cDNA from female house flies (Hediger *et al.* 2010). The sequences of the T7 extended primers used to produce dsRNA targeting *Md-tra* are 5'-gtaatacgatcactatagggTGGTGTAATATGGCTCTATCG-3' and 5'-gtaatacgatcactatagggGCTGCCATACAAACGTGTC-3' (sequences in lower case are the T7 region and sequences in upper case anneal to *Md-tra*). The sequences of the T7 extended primers used to produce dsRNA targeting GFP are 5'-gtaatacgatcactatagggATGTGAGCAAGGGC-3' and 5'-gtaatacgatcactatagggCTTGTACAGCTCGTC-3'.

The larvae that hatched from embryos injected with either *Md-tra*-RNAi or GFP-RNAi were raised on porcine feces because the small number of injected larvae are less likely to develop into adult flies on standard rearing media (Schmidt *et al.* 1997). Under the injection scheme (Table 2.1), I could collect four types of flies: (A) genotypic females with the GFP-RNAi treatment (phenotypic females), (B) genotypic females with the *Md-tra*-RNAi treatment (sex-reversed males), (C) genotypic males with GFP-RNAi treatment (III^M males #1), and (D) genotypic males with the *Md-tra*-RNAi treatment (III^M males

#2). Both types of genotypic males (III^M males #1 and #2) are also phenotypic males, and the GFP-RNAi treated genotypic females are phenotypic females. Sex reversal to a phenotypic male occurs in genotypic females under the *Md-tra*-RNAi treatment (Hediger *et al.* 2010).

Table 2.1. Injection scheme for RNAi treatments in both sexes.

Genotypic Sex	**RNAi treatment**	
	GFP-RNAi	*Md-tra*-RNAi
Genotypic Female (III/III)	(A) Phenotypic Female (III/III) "females"	(B) Phenotypic Male (III/III) "sex-reversed males"
Genotypic Male (III^M/III)	(C) Phenotypic Male (III^M/III) "III^M males #1"	(D) Phenotypic Male (III^M/III) "III^M males #2"

Genotypic females with the *Md-tra*-RNAi treatment are sex-reversed to phenotypic males (B). The other genotypic females and males are not sex-reversed (A, C, D); their phenotypic sexes are congruent with their genotypic sexes (i.e. normal males or females).

After emergence from pupa, each injected single phenotypic male was kept in a small cage with three or four females from the *pw bwb* strain that did not have any injection treatments. Only phenotypic males that successfully sired offspring with those females were retained for the RNA-seq experiment. All three types of phenotypic males produced offspring, but the sex-reversed males sired only female offspring (because they do not carry *Mdmd*). To measure gene expression in females, I collected virgin GFP-RNAi treated genotypic females. Those females were collected within eight hours of emergence and kept separate from males to ensure they were virgin. The females were aged for five days, and I selected three females to dissect for RNA-seq experiments. I measured gene expression in virgin females to exclude mating effects on female gene expression.

2.2.3 RNA-seq experiments

I used RNA-seq to measure gene expression in heads and abdomens from individual males of the four strains carrying either the Y^M or III^M proto-Y chromosomes (Figure 2.1). The larvae were raised at 25 ˚C on a standard diet of wheat bran, calf manna, yeast, reptile litter, and water, as described previously (Hamm *et al.* 2009; Meisel *et al.* 2015). Unmated adult males and females were sorted within eight hours of emergence, kept separately at 22 ˚C, and provided water, sugar, and powdered milk *ad libitum*. Heads and abdomens from adult flies at five days post emergence were dissected and frozen at -80 ˚C. The heads and abdomens from individual males were homogenized in TRIzol Reagent, and then RNA was extracted using the Zymo Direct-zol kit following the manufacturer's protocol including DNA digestion steps. Three biological replicates (i.e., three individual male heads and abdomens) were prepared from males of each of the four strains. Because females of three strains (CS, CSrab, and IsoCS) are isogenic, I sampled only one female from each of the strains. However, the RNA-seq library preparation for CS female abdomen failed, so that the female abdomen had only two biological replicates.

I also performed RNA-seq on heads and abdomens from the four types of flies injected with dsRNA (Table 2.1). Individual four to five days old adult flies (described above) were frozen in liquid nitrogen, and RNA was extracted from the individual flies with the NucleoSpin RNAII kit (Macherey-Nagel, Düren, Germany) following the protocol of the manufacturer, which includes DNA digestion steps. Three biological replicates (i.e. three individual flies) from each of the four genotype-by-treatment

combinations were collected.

RNA-seq libraries were prepared with the Illumina TruSeq Stranded mRNA Sample Preparation Kit following the protocols of the manufacturer. The libraries were run in six lanes for 75 cycles (i.e. 75 nucleotide reads) on an Illumina NextSeq500 machine at University of Houston Seq-N-Edit Core. For the strains with different naturally occurring proto-Y chromosomes (Figure 2.1), two of three lanes contained ten libraries comprised of one replicate from each strain, sex, and body part (four strains of males plus one of the strains of females by two tissues): CS male head and abdomen, CSrab male head and abdomen, IsoCS male head and abdomen, bwbCS male head and abdomen, and female head and abdomen. The third lane contained nine of the samples described above, but no CS female abdomen because that library preparation failed. For the RNAi experiment I ran three lanes, and each lane contained eight library samples, one replicate from each genotype-by-treatment combination and body part: *Md-tra*-RNAi genotypic female head and abdomen (sex-reversed male), *Md-tra*-RNAi genotypic male head and abdomen, GFP-RNAi genotypic female head and abdomen, and GFP-RNAi genotypic male head and abdomen.

2.2.4 Data analysis

Illumina RNA-seq reads were aligned to house fly genome assembly v2.0.2 and annotation release 102 (Scott *et al.* 2014) using HISAT2 v2.0.1 (Kim *et al.* 2015). First, read coverage across the sex determining genes *Md-tra*, *doublesex* (*Md-dsx*), and *fruitless* (*Md-fru*) was determined with the 'mpileup' function in SAMtools (Li *et al.* 2009). Second, the aligned reads were assigned to all annotated genes with htseq-count in

HTSeq v0.9.1 (Anders *et al.* 2015), with the --stranded=reverse option because I generated stranded RNA-seq libraries.

The HTSeq output was used as input into DESeq2 v1.16.1 to identify differentially expressed genes (Love *et al.* 2014). For the DESeq2 analysis of the four strains with different naturally occurring Y^M and III^M chromosomes, I performed pair-wise comparisons between males of each strain. I also performed pair-wise comparisons of males from each strain against females. For the RNAi experiment, I created a model in DESeq2 in which gene expression is predicted by genotypic sex, RNAi treatment (GFP-RNAi or *Md-tra*-RNAi), and the interaction between genotypic sex and RNAi treatment. The model allows for pair-wise comparisons between individuals with either the same genotypic sex or RNAi treatment. From the pair-wise comparisons, $\log_2$ fold-changes ($\log_2$FC) were extracted for each gene with false discovery rate corrected *P* values (Benjamini and Hochberg 1995). I also extracted $\log_2$FC for III^M males #2 over females using the equation: $\log_2(III^M \text{ males \#2 / females}) = \log_2(III^M \text{ males \#1 / females}) + \log_2(III^M \text{ males \#2 / } III^M \text{ males \#1})$. I cannot calculate a *P* value for a test of whether $\log_2(III^M \text{ males \#2 / females})$ is different from zero because it is not a pair-wise comparison performed by the model I created in DESeq2. Only genes with adjusted *P* values reported by DESeq2 are presented and used for downstream analyses. In other words, I considered a gene to be expressed if there was enough data to compare gene expression levels, and I ignored genes where a statistical test was not performed because expression was too low.

I performed a principal component (PC) analysis and used a grade of membership

model implemented in the R package 'CountClust' (Dey *et al.* 2017) to analyze the normalized expression count data from DESeq2. For the PC analysis, the normalized count data were transformed using the 'rlog' function in DESeq2 (Love *et al.* 2014). Because genes with low counts show the highest relative differences among samples and create large variances, these low count genes dominate the results of the PC analysis. The function 'rlog' stabilizes the variance of the data, making it homoscedastic. Gene Ontology (GO) terms were analyzed with DAVID v.6.8 (Huang *et al.* 2008).

I assigned house fly genes to chromosomes using the conservation of Muller elements across flies (Foster *et al.* 1981; Weller and Foster 1993), as done previously (Meisel *et al.* 2015; Meisel and Scott 2018). Briefly, the house fly and *Drosophila* genomes are organized into six chromosome arms (Muller elements A-F). Elements A-E correspond to the house fly chromosomes that were historically considered the autosomes. Element F is the historical house fly X chromosome (Vicoso and Bachtrog 2013). One-to-one orthologs between house fly and *Drosophila melanogaster* genes were identified as part of the house fly genome annotation (Scott *et al.* 2014). I assigned house fly scaffolds to Muller elements using a "majority rules" approach—if the majority of genes on a scaffold were orthologous to *D. melanogaster* genes on a single Muller element, then the house fly scaffold was assigned to that Muller element. In turn, all genes on that scaffold are assigned to the same Muller element.

2.3 Results

2.3.1 The III^M chromosome has a minor effect on gene expression

It is previously observed that hundreds of genes are differentially expressed between males carrying III^M and males carrying Y^M (Meisel *et al.* 2015). However, it is not clear from that work if the expression differences were specific to introducing the Y^M or III^M proto-Y chromosomes on a genetic background, or if changing any single chromosome can induce similar expression effects. To address this question, I used RNA-seq to measure gene expression in males from four nearly isogenic strains carrying either Y^M or III^M chromosomes (Figure 2.1). Two strains with "III^M males" have different III^M chromosomes on a common genetic background. A third strain with "Y^M males" has a Y^M chromosome, instead of III^M, on the same genetic background. The fourth strain carries a different Y^M chromosome and a single copy of a different standard third chromosome (without *Mdmd*) on the same genetic background as the other three strains. If the III^M chromosome has a disproportionate effect on gene expression, I expect to observe more genes differentially expressed between III^M and Y^M males than between Y^M males that differ from each other by a single copy of a standard third chromosome.

To compare gene expression profiles across the strains, I used both a PC analysis and a grade of membership model (Dey *et al.* 2017). I excluded one of three male replicates from each of the four strains in both abdomen and head because they had outlier expression profiles (Appendix Figure 1A, B), likely as a result of faulty sample preparation or extreme batch effects in sequencing. In abdomen, the first PC (PC1) and second PC (PC2) explain 84% and 7% of variance in gene expression across samples,

respectively (Figure 2.2A). In head, PC1 and PC2 explain 43% and 28% of variance, respectively (Figure 2.2B). In both body parts, all males from the four strains are separated from females along PC1. Notably, the two Y^M strains (that differ from each other by a single copy of a standard third chromosome) have the greatest separation of any pair of male samples along PC2 in the abdomen data. I observed similar results with a grade of membership model: males from all four strains show different membership from females in abdomen and head, and males from the two different Y^M strains have the most different membership composition (Appendix Figure 1C, D). In head, one of the III^M genotypes is separated from the other males along PC2 (Figure 2.2B). In neither abdomen nor head is the greatest separation between Y^M and III^M males, suggesting that the Y^M and III^M chromosomes affect gene expression to a similar extent as a non-*Mdmd*-bearing third chromosome.

I also identified individual genes with significant differential expression between strains using DESeq2 (Love *et al.* 2014). Previous work found that an excess of third chromosome genes is differentially expressed between Y^M and III^M males (Meisel *et al.* 2015), as expected based on the differences in their genotypes. I similarly find that excesses of genes on the third chromosome are differentially expressed in 5/8 comparisons between males with different III^M chromosomes, between Y^M males that differ by a standard third chromosome, and between Y^M and III^M males (Appendix Figure 2). Only one other chromosome has a significant excess of differentially expressed genes in a single comparison. Notably, there are more differentially expressed genes

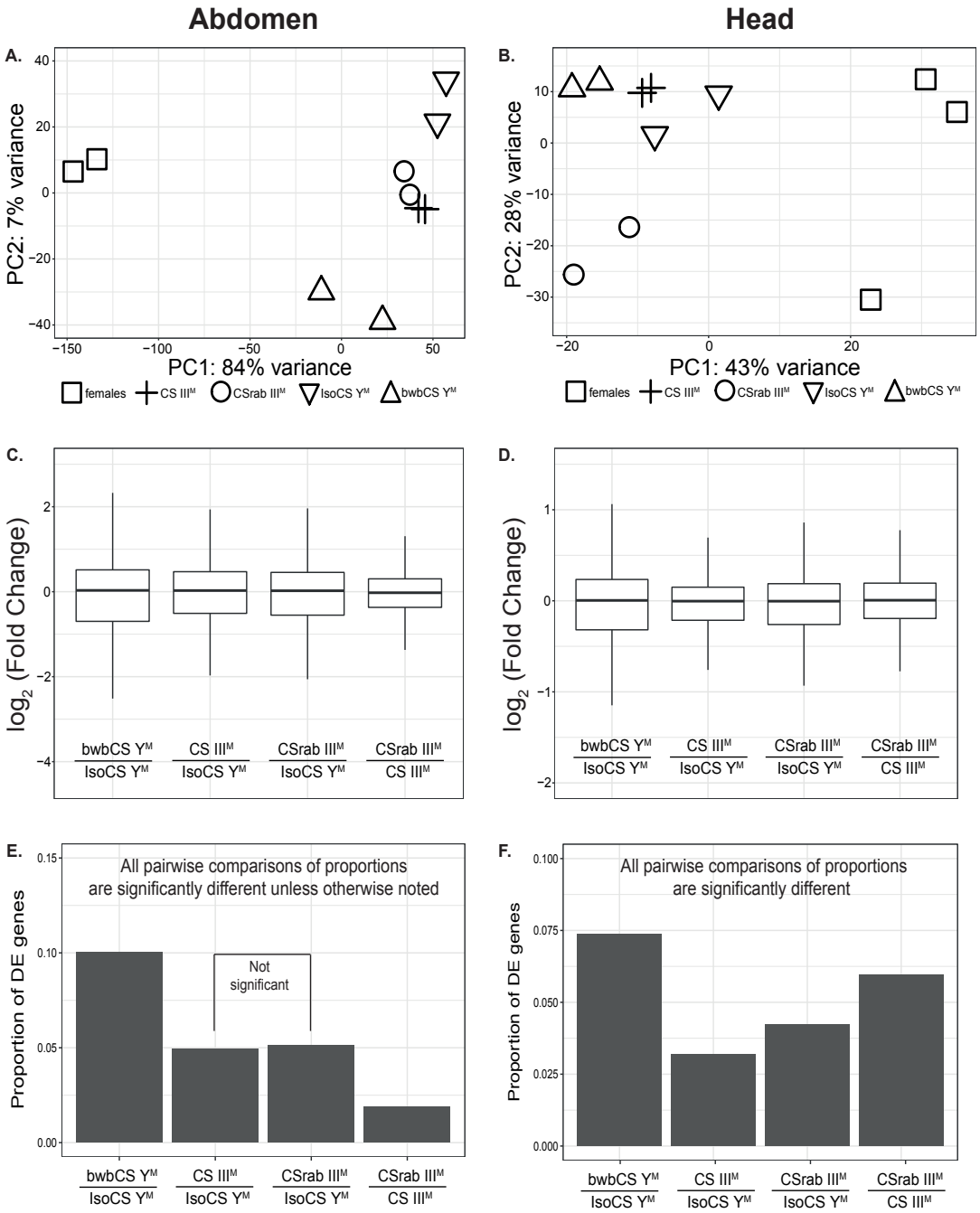


Figure 2.2. PC analysis of global expression in males with different Y^M or III^M proto-Y chromosomes in abdomens (A) and heads (B). Boxplot shows fold changes of gene expression between males with different *Mdmd*-bearing chromosomes in abdomens (C) and heads (D). Bar graphs show the proportions of differentially expressed (DE) genes between males with different Y^M or III^M proto-Y chromosomes in abdomens (E) and heads (F). bwbCS Y^M stands for the strain bwbCS×CS. Asterisks indicate significant differences.

across the entire genome in the pair-wise comparison between Y^M males with different standard third chromosomes than in any other pair-wise comparison between males, including between Y^M and III^M males (Figure 2.2C-F, Appendix Figure 3, Appendix Table 1). The PCA, grade of membership, and differential expression analyses therefore all suggest that the non-*Mdmd* bearing standard third chromosome has an equal or greater effect on male gene expression than the III^M chromosome.

2.3.2 Expression of genes in the house fly sex determination pathway following *Md-tra* knockdown

To further examine the effect of the III^M chromosome on gene expression, I used RNAi targeting *Md-tra* to create sex-reversed males that have a male phenotype and female genotype without any male-determining proto-Y chromosome. I compared gene expression in these sex-reversed males with genotypic males carrying a III^M chromosome. My 2×2 experimental design consisted of injecting dsRNA targeting either *Md-tra* (to sex-reverse genotypic females) or GFP (sham treatment) into genotypic males and females (Table 2.1). The *Md-tra*-RNAi treatment mimics the effect of the male-determining *Mdmd* gene that disrupts the splicing of *Md-tra* and the positive autoregulatory function of *Md-tra* in the early embryo (Hediger *et al.* 2010).

To confirm that the *Md-tra*-RNAi treatment knocks down *Md-tra* expression, I examined the expression of *Md-tra* using RNA-seq coverage data collected from the abdomens of each of my four sample types (Figure 2.3A). I expected the expression of *Md-tra* in females to be higher than in males because males produce a splice variant with a premature stop codon that is likely to be processed by the nonsense-mediated decay

(NMD) pathway (Hediger *et al.* 2010; Kervestin and Jacobson 2012). In addition, the ovaries are expected to produce large amounts of *Md-tra* transcripts because *Md-tra* activity is necessary for maternal establishment of zygotic splicing of *Md-tra* (Dübendorfer and Hediger 1998). In abdomen, normal females (GFP-RNAi treated genotypic females) did indeed express *Md-tra* approximately three times higher than normal males (genotypic males with either the GFP-RNAi or *Md-tra*-RNAi treatment). This high *Md-tra* expression in female abdomens might reflect the outcomes of strong ovarian expression. Importantly, *Md-tra* expression in sex-reversed males (genotypic females that are phenotypic males because of *Md-tra*-RNAi) was comparable to that of the genotypic males, not the normal females (Figure 2.3A). This is likely because knock down of *Md-tra* by RNAi produces sex-reversed males that have functioning testes instead of ovaries. The *Md-tra* exons that are included in the functional, female-determining transcript were the highest expressed exons in phenotypic females (Figure 2.3A), consistent with the production of the female-determining transcript in female ovaries (Hediger *et al.* 2010).

I found that *Md-tra* was also differentially expressed between females and males in head, but the difference was much smaller than in abdomen (Figure 2.3B). Notably, when I analyzed the read mapping to *Md-tra* using DESeq2, expression was significantly higher in normal females than in genotypic (III^M #1) males. However, there was not a significant difference in *Md-tra* expression between sex-reversed males and either normal males or normal females. These results were observed after I excluded a sex-reversed male head sample that had an outlier expression profile (see below). The lack of sexually

dimorphic expression of *Md-tra* in head was consistent with minimal sex-biased expression in *Drosophila* and house fly heads (Goldman and Arbeitman 2007; Meisel *et al.* 2015). In addition, most somatic cells in *Drosophila* were sexually monomorphic as a result of cell autonomous sex determination in *Drosophila* (Robinett *et al.* 2010), suggesting that the same may be true for most cells in house fly heads.

Md-TRA protein regulates the splicing of at least two downstream genes, *Md-dsx* and *Md-fru*, which are both differentially spliced between females and males (Hediger *et al.* 2004, 2010; Meier *et al.* 2013). The expression of *Md-dsx* and *Md-fru* in sex-reversed males was more similar to that of normal (genotypic) males (especially *Md-tra*-RNAi treated III^{M} males #2) than phenotypic females (Appendix Figure 4), confirming that *Md-tra* knock down affects the downstream genes in the sex determination pathway (Hediger *et al.* 2010; Meier *et al.* 2013). For example, *Md-dsx* expression in phenotypic males was higher than in phenotypic females, especially across male-specific exons (Appendix Figure 4A, B), consistent with the expected effect of Md-TRA on *Md-dsx* splicing in females (Hediger *et al.* 2004, 2010).

The expression of *Md-fru* was higher in head than in abdomen (Appendix Figure 4C, D), consistent with its role as a behavioral regulator (Meier *et al.* 2013). Md-TRA regulates the splicing of *Md-fru* by promoting the production of splice variants with premature stop codons in females (Meier *et al.* 2013). Sex-specific splicing of *Md-fru* occurs at the 5' end of the transcript (Meier *et al.* 2013), but the 5' end of *Md-fru* was not completely assembled and annotated in the reference genome. I therefore cannot test for differential splicing of *Md-fru* between males and females.

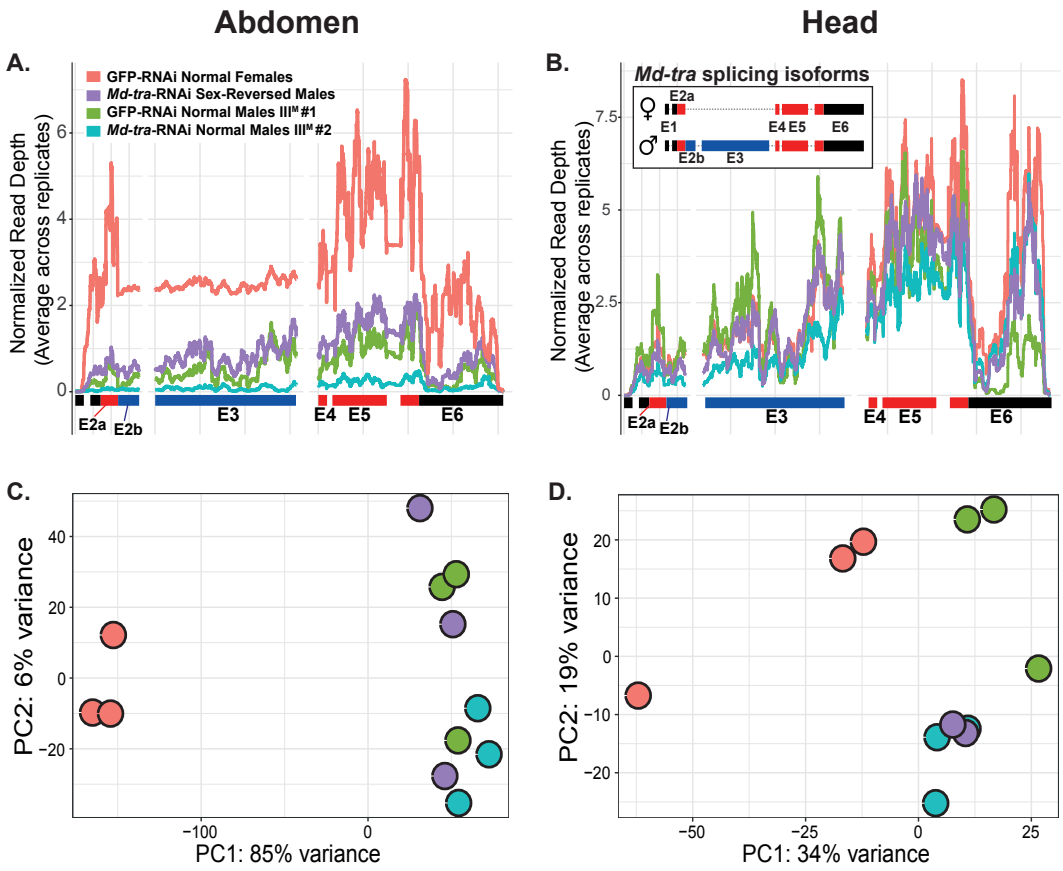


Figure 2.3. *Md-tra* expression (A, B) and PC analysis of global expression (C, D) of GFP-RNAi and *Md-tra*-RNAi treated genotypic females and males in abdomens (A, C) and heads (B, D). An inset in (B) shows female and male isoforms of *Md-tra*. (A, B) Blue exons (E2b, E3) that contain premature stop codons are included in the male isoforms of *Md-tra* but excluded from the female isoforms. Read coverage in the long introns between E2b-E3 and E3-E4 is not shown to better visualize *Md-tra* expression within the exons.

However, I expected expression of *Md-fru* to be higher in males than females because the female splice variants were removed by the NMD pathway. Indeed, I observed that *Md-fru* expression was much higher in the heads of GFP-RNAi treated genotypic males (III^M males #1) than GFP-RNAi treated normal females (Appendix Figure 4C, D). However, in *Md-tra*-RNAi treated genotypic males (III^M males #2) and sex-reversed males, the expression of *Md-fru* is intermediate between females and GFP-RNAi treated genotypic

males (Appendix Figure 4D). A possible explanation is that RNAi knock down of *Md-tra* affects the expression or splicing of *Md-fru* in these flies (sex-reversed males and III^M males #2), but testing this hypothesis is beyond the scope of the work presented here.

2.3.3 Expression profiles of sex-reversed males are similar to genotypic males, not phenotypic females

I next examined how the III^M chromosome affects the global gene expression profiles in males using the RNA-seq data from the four genotype-by-RNAi-treatment combinations. I first used a PC analysis on the regularized log-transformed normalized expression count data for each gene in each replicate (Love *et al.* 2014). In the abdomen expression data, PC1 explains 85% of the variance in expression levels across samples. PC1 clusters all types of phenotypic males together, including the sex-reversed males, separately from normal females (Figure 2.3C). In the head data, I found that one of the sex-reversed males had elevated *Md-tra* expression and an RNA-seq profile that did not cluster with normal females or genotypic males (Appendix Figure 5), suggesting incomplete knock down of *Md-tra* in that sex-reversed animal's head. After excluding that sample, PC1 and PC2 explain 34% and 19% of the variance in gene expression in head, respectively. PC1 for the head expression data separates normal females and GFP-RNAi treated genotypic males (Figure 2.3D). Curiously, *Md-tra*-RNAi-treated phenotypic males (which includes both sex-reversed males and III^M males #2) were intermediate between GFP-RNAi-treated normal females and males along head PC1 (Figure 2.3D) and separated from GFP-RNAi-treated normal females and males along head PC2. The *Md-tra*-RNAi-treated phenotypic male heads (sex-reversed males and III^M males #2) also had reduced

expression of *Md-fru* (Appendix Figure 4D). Therefore, *Md-tra* knock down might influence overall gene expression as well as *Md-fru* expression or splicing in heads.

My PC analysis demonstrates that sex-reversed males have similar abdominal gene expression profiles as genotypic males (III^M males #1 and #2), which are clearly distinguishable from phenotypic females. However, the gene expression of the sex-reversed males (and genotypic males treated with *Md-tra*-RNAi) in heads is not as sexually dimorphic. To validate this result, I also used a grade of membership model to compare gene expression patterns among samples (Dey *et al.* 2017). These results were consistent with the above PC analysis, showing that the sex-reversed males have similar expression profiles as genotypic males (III^M males #1 and #2) and different from normal females in abdomen (Appendix Figure 6A). The sexual dimorphism in head, however, is more ambiguous (Appendix Figure 6B), consistent with the PC analysis.

2.3.4 Sex-reversed and genotypic males have similar sex-biased gene expression

Sexual dimorphism is achieved through differential (sex-biased) gene expression between males and females (Ellegren and Parsch 2007). I therefore compared sex-biased expression in sex-reversed and genotypic males. I used genotypic males treated with GFP-RNAi (III^M males #1) as my normal male reference because the model in DESeq2 I used for RNA-seq analysis allows for pair-wise comparisons between individuals with either the same genotypic sex or treatment. Normal females and III^M males #1 both were exposed to the GFP-RNAi treatment, which allows us to make the pairwise comparison. I first quantified the degree of sex-biased expression by the distribution of the $\log_2$ fold-change between male and female expression levels ($\log_2$M/F). In the abdominal samples,

the distributions of $\log_2$M/F for sex-reversed and genotypic males, when compared to normal females, are quite similar (Figure 2.4A; Appendix Figure 7A).

I defined genes with sex-biased expression as those with a $\log_2$M/F significantly different from 0 using DESeq2 (Love *et al.* 2014). Similar fractions of genes have sex-biased expression in abdomen for sex-reversed and genotypic males:11,005/14,686 (74.9%) of genes are significantly sex-biased in the comparison between sex-reversed males and females, and 11,030/14,993 (73.6%) of genes have sex-biased expression when comparing genotypic males (III^M males #1) and females (Figure 2.4C; Appendix Table 2). The distributions of $\log_2$M/F are symmetrical, with similar fractions of genes with male- and female-biased expression for both sex-reversed and genotypic males (Figure 2.4A). In contrast, the magnitude of differential gene expression is much smaller in comparisons between genotypic males than male-female comparisons (Figure 2.4A, C). Notably, the proportion of differentially expressed genes is similar between sex-reversed and genotypic males as between the two types of genotypic males (III^M males #1 and #2), providing additional evidence that sex-reversed males have similar gene expression profiles as normal (genotypic) males (Figure 2.4C).

Sex-biased expression in fly heads is reduced relative to whole fly or gonad tissue (Goldman and Arbeitman 2007; Lebo *et al.* 2009; Meisel *et al.* 2015). In house fly heads, I only detect 5,077 sex-biased genes between genotypic males (III^M males #1) and normal females out of 13,558 expressed genes (Figure 2.4B, D; Appendix Figure 7B; Appendix Table 2). Similarly, there are only 735 sex-biased genes between sex-reversed males and normal females out of 12,360 expressed genes (Figure 2.4B, D; Appendix

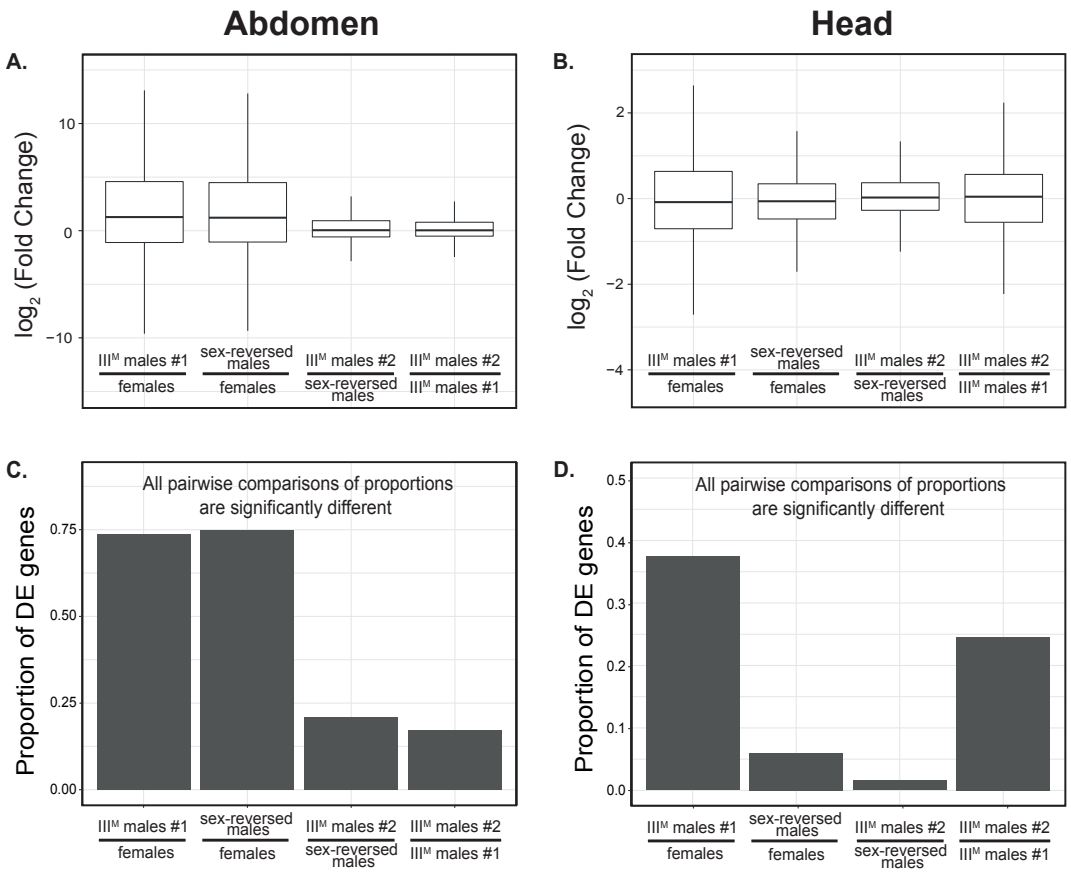


Figure 2.4. Boxplot showing fold changes of gene expression among comparisons in abdomens (A) and heads (B). Bar graphs show the proportions of differentially expressed (DE) genes between different types of individuals in abdomens (C) and heads (D). "Females" refers to GFP-RNAi treated normal females. Asterisks indicate significant differences.

Figure 7B; Appendix Table 2). The lower number of sex-biased genes between sex-reversed males and normal females could be a result of decreased power because of a smaller sample size—only two replicate sex-reversed male heads were included because the third replicate had an outlier expression profile (see above). Alternatively, sex-reversed male heads could be less sexual dimorphic than genotypic male heads. In addition, there are fewer genes differentially expressed in head between sex-reversed males and *Md-tra*-RNAi treated genotypic males (III^M males #2) than between the two

types of genotypic males (Figure 2.4D). This result is consistent with the clustering of sex-reversed males and III^M males #2 in the PC analysis of global expression in heads (Figure 2.3D), suggesting that gene expression in head is more affected by *Md-tra*-RNAi than by the III^M chromosome.

I next tested if the same genes have sex-biased expression in sex-reversed males and genotypic males (III^M males #1). In both abdomen and head, the majority of male-biased genes in genotypic males are also male-biased in sex-reversed males (Figure 2.5A, B). The same is true for female-biased genes. I tested if the sex-biased genes in common between sex-reversed and genotypic males is greater than expected by chance with a permutation test. I determined a null distribution assuming that sex-biased genes in the sex-reversed and genotypic males are independent of each other from 1,000 random permutations of my data. The actual number of sex-biased genes in both abdomen and head in common between sex-reversed males and genotypic males is much greater than all values in the null distribution (Figure 2.5C, D). This result implies that sexual dimorphism is achieved by similar means in both sex-reversed males and genotypic males: silencing of *Md-tra*, independent of alleles on the III^M chromosome.

2.3.5 Disproportionate differential expression of third chromosome genes

Although all phenotypic males, regardless of genotypic sex, showed very similar gene expression profiles (Figure 2.3), I identified some genes that are differentially expressed between genotypic males and sex-reversed males (Figure 2.4). These differentially expressed genes could reveal important phenotypic effects of the III^M proto-Y chromosome, which may be important for the maintenance of both Y^M and III^M across

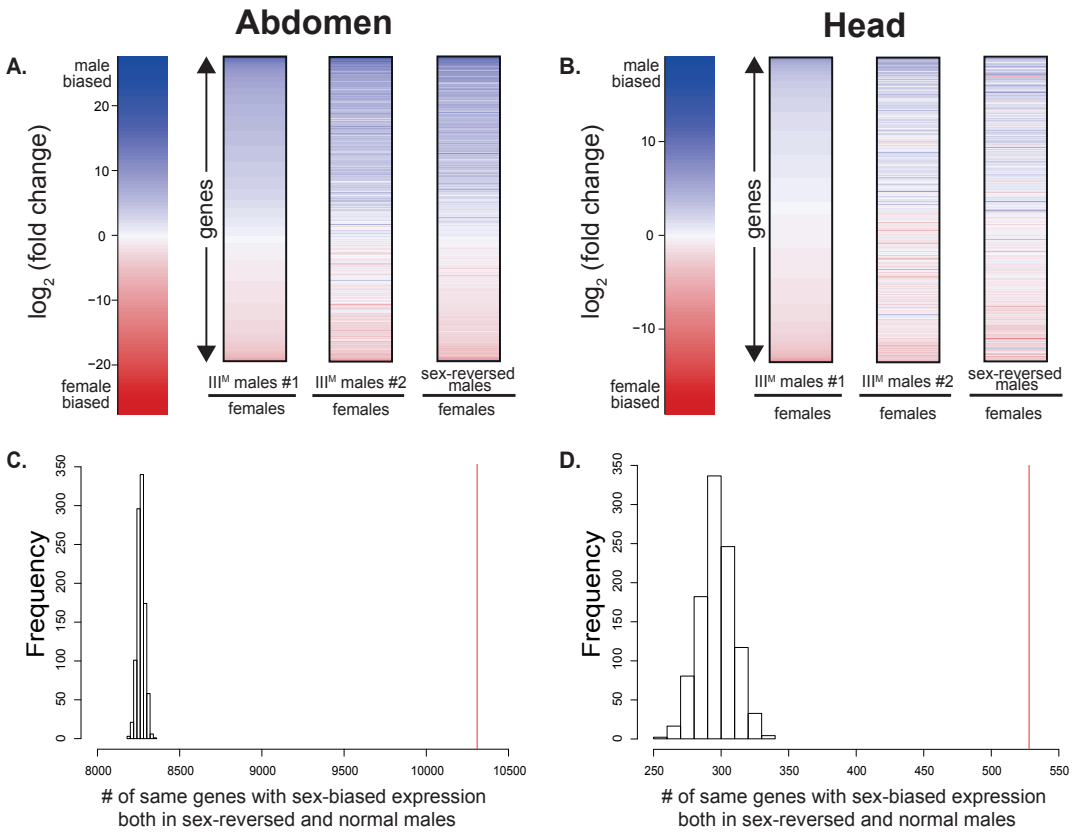


Figure 2.5. Heat maps showing expression differences between each type of male and females in abdomens (A) and heads (B). Permutation tests for whether the same genes have sex-biased expression both in sex-reversed males and normal males (III^M males #1) in abdomens (C) and heads (D). Histograms represent null distribution and red lines indicate the observed number of genes with the same sex-biased expression both in sex-reversed and normal males.

populations. I therefore further examined differential expression between sex-reversed and genotypic males to determine the effect of the III^M chromosome. As expected based on their genotypic differences, there are significant excesses of third chromosome genes differentially expressed between genotypic III^M males and sex-reversed males in abdomen and head (Figure 2.6). There is also a significant excess of third chromosome genes differentially expressed between genotypic males and normal females in head (Figure 2.6B). In contrast, there is not an excess of third chromosome genes differentially

expressed between normal females and sex-reversed males (Figure 2.6), who share the same genotype. These patterns are consistent with my previous work (Meisel *et al.* 2015) and other results presented here (Appendix Figure 2) in which the third chromosome has an excess of differentially expressed genes between flies that differ in their third chromosome genotype. However, I surprisingly find that there are excesses of differentially expressed genes on the third chromosome in comparisons between III^M males with the *Md-tra*-RNAi and GFP-RNAi treatments (Figure 2.6). Therefore, in addition to the expected genotypic effects, dsRNA targeting *Md-tra* and/or GFP disproportionately affects the expression of genes on the house fly third chromosome.

Sex reversed males and genotypic males also have very similar sex-biased expression relative to females (Figure 2.5). In spite of these similarities, I identified some "discordant sex-biased genes" that have sex-biased expression in either sex-reversed or genotypic males, but not both. To further examine the effect of the III^M chromosome on gene expression, I divided the discordant sex-biased genes into two groups: "sex-reversed-up-discordant" and "normal-up-discordant". I considered a gene to be sex-reversed-up-discordant if it belongs to one of two categories: 1) male-biased expression in sex-reversed males and not male-biased in genotypic III^M males #1 ($\log_2 M/F<0$ but not necessarily significant), or 2) female-biased expression in genotypic males and not female-biased in sex-reversed males ($\log_2 M/F>0$ but not necessarily significant). I identified 49 sex-reversed-up-discordant genes in abdomens and 170 in heads (Appendix Table 3). Likewise, I classified genes as normal-up-discordant if they are in one of two categories: 1) male-biased expression in genotypic (normal) males and not male-biased in

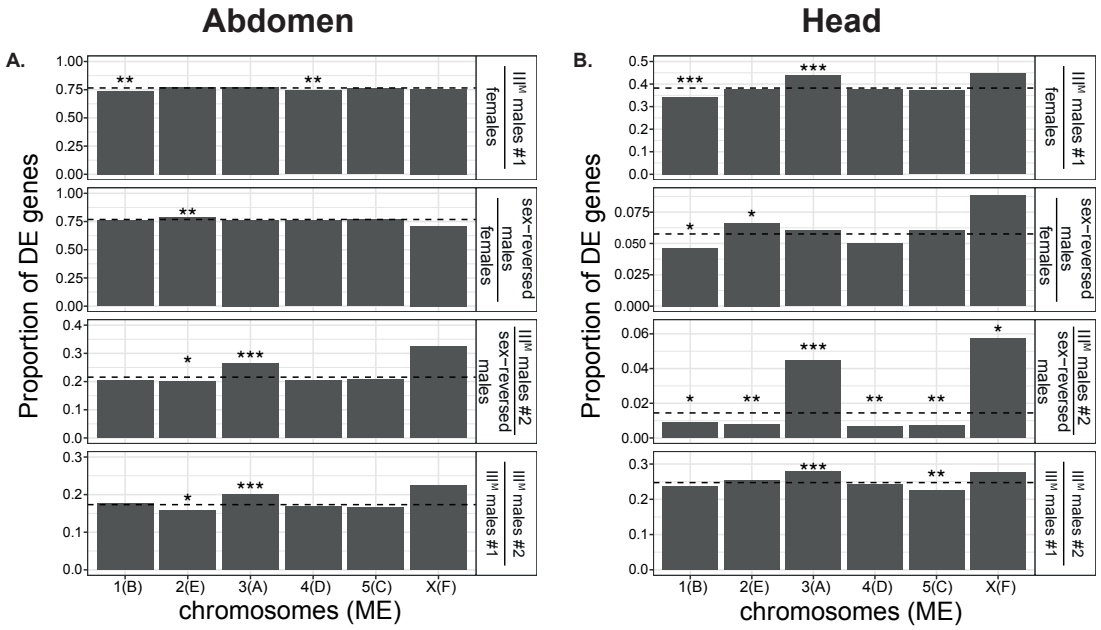


Figure 2.6. Bar graphs indicate the proportions of genes on each chromosome (*Drosophila* Muller element in parentheses) that are differentially expressed (DE) between different genotype and treatment combinations in abdomens (A) and heads (B). Asterisks indicate significant differences based on Fisher's exact test (**P*<0.05, ***P*<0.01, ****P* < 0.001).

sex-reversed males ($\log_2$M/F<0 but not necessarily significant), or 2) female-biased expression in sex-reversed males and not female-biased expression in normal males ($\log_2$M/F>0 but not necessarily significant). I identified 25 normal-up-discordant genes in abdomens and 418 in heads (Appendix Table 3). There are no GO (gene ontology) terms significantly enriched in either the sex-reversed-up-discordant or normal-up-discordant genes. However, both sex-reversed-up-discordant and normal-up-discordant genes in both abdomen and head are significantly enriched on the third chromosome (Tables 2.2 and 2.3). Therefore, in comparisons between males with and without a III^M chromosome, the third chromosome is enriched for differentially expressed genes and genes with discordant sex-biased expression.

Table 2.2. Chromosomal distribution of discordant sex-biased genes in abdomen

		Abdomen					
		normal-up-discordant (n-u-d)			sex-reversed-up-discordant (sr-u-d)		
Chromosomes (Muller elements)	# genes on chr	# genes	Odds ratio	95% CI	# genes	Odds ratio	95% CI
1(B)	2000	1	0.270	0.006 - 1.709	2	0.270	0.031 - 1.048
2(E)	2910	4	0.806	0.194 - 2.532	11	1.231	0.550 - 2.568
3(A)	**2094**	**9**	**4.129**	**1.482 - 11.322**	**13**	**2.386**	**1.119 - 4.853**
4(D)	2184	3	0.817	0.152 - 2.58	5	0.660	0.201 - 1.705
5(C)	2469	2	0.440	0.049 - 1.855	7	0.844	0.313 - 1.958
X(F)	45	0	0	0 - 57.539	0	0	0 - 27.459
Total	11702	19			38		

The chromosomal distribution of discordant genes was compared to all genes in the genome. Genes that were not assigned to a chromosome were excluded. A Fisher's exact test was performed to test for an excess of normal-up-discordant genes on each chromosome relative to the number of total genes on each chromosome.

Table 2.3. Chromosomal distribution of discordant sex-biased genes in head

		Head					
		normal-up-discordant (n-u-d)			sex-reversed-up-discordant (sr-u-d)		
Chromosomes (Muller elements)	# genes on chr	# genes	Odds ratio	95% CI	# genes	Odds ratio	95% CI
1(B)	1750	67	1.145	0.859 - 1.507	18	0.862	0.490 - 1.436
2(E)	2550	73	0.792	0.601 - 1.032	24	0.759	0.463 - 1.201
3(A)	**1842**	**108**	**2.028**	**1.592 - 2.569**	**44**	**2.665**	**1.787 - 3.933**
4(D)	1888	50	0.735	0.531 - 0.999	12	0.494	0.247 - 0.902
5(C)	2147	51	0.641	0.465 - 0.868	21	0.805	0.476 - 1.303
X(F)	34	1	0.858	0.021 - 5.144	0	0	0 - 9.947
Total	10211	350			119		

The chromosomal distribution of discordant genes was compared to all genes in the genome. Genes that were not assigned to a chromosome were excluded. A Fisher's exact test was performed to test for an excess of normal-up-discordant genes on each chromosome relative to the number of total genes on each chromosome.

2.4 Discussion

The house fly Y^M and III^M proto-Y chromosomes are geographically distributed in a way that suggests ecological factors favor different proto-Y chromosomes across different habitats (Franco *et al.* 1982; Tomita and Wada 1989; Hamm *et al.* 2005; Feldmeyer *et al.* 2008; Kozielska *et al.* 2008). This predicts that there will be sequence differences between the proto-Y chromosomes and their homologous (proto-X) chromosomes that confer ecologically dependent phenotypic and fitness effects. These differences could be in transcribed sequences (e.g. protein coding genes) or in regulatory regions that control their expression. Paradoxically, however, both the Y^M and III^M chromosomes have minimal sequence differences relative to their homologous chromosomes (Meisel *et al.* 2017).

I tested if minimal sequence differences between the proto-Y chromosomes and their homologs could be responsible for phenotypic effects by investigating gene expression differences between males carrying different proto-Y chromosomes. To those ends, I first compared gene expression in four house fly strains carrying either a Y^M or III^M chromosome on a common genetic background (Figure 2.1). The biggest differences in gene expression were observed between two Y^M strains carrying different standard third chromosomes (not carrying *Mdmd*), and not between III^M and Y^M males (Figure 2.2). My results therefore suggest that the magnitude of gene expression differences between III^M and Y^M males can be explained by replacing a chromosome on a common genetic background, and they are not specific to the effect of the III^M or Y^M chromosomes.

Second, I examined the effects of the III^M chromosome on male gene expression using an RNAi experiment. I chose to knock down *Md-tra* because it allows us to create sex-reversed fertile males that do not carry any proto-Y chromosomes, as opposed to knock down/out of *Mdmd*, which creates sex-reversed fertile females carrying a proto-Y (Sharma *et al.* 2017). I found that gene expression profiles of sex-reversed and normal (genotypic) males are very similar (Figures 2.3, 2.4, and 2.5), with only a few genes exhibiting different sex-biased expression between the genotypic and sex-reversed males (Tables 2.2 and 2.3; Appendix Table 3). I therefore conclude that the III^M chromosome has a minor effect on male gene expression in a constant environment as assayed in my experiments.

2.4.1 Gene expression effects of the proto-Y chromosomes

A previous experiment identified many genes whose expression differs between Y^M and III^M males, but that experiment did not compare the effect of the proto-Y chromosomes with the effects of equivalent autosomes (Meisel *et al.* 2015). I observe more expression differences between Y^M males that carry different copies of standard (non-*Mdmd*-bearing or autosomal) third chromosomes than between Y^M males and III^M males (Figure 2.2). This minimal effect of the III^M proto-Y chromosome on expression, relative to a standard third chromosome, suggests that III^M is essentially a normal third chromosome that happens to carry *Mdmd*, as opposed to a "masculinized" proto-Y chromosome (Rice 1996). In addition, it also suggests that III^M is not differentiated enough from the standard third chromosome to require dosage compensation in heterogametic males. Alternatively,

III^M males may have a dosage compensation mechanism (i.e. through preferred expression of genes on the standard third chromosomes), which could mask the effects of the III^M chromosome on gene expression.

It is curious that the Y^M males with different standard third chromosomes have more expression differences than between Y^M and III^M males (Figure 2.2). One explanation for the amount of expression differences between the Y^M males is that the standard third chromosome in my experiment has a greater effect on gene expression than the III^M chromosome. Alternatively, the different origins of the Y^M chromosomes in the two Y^M genotypes could have a large effect on gene expression. Unfortunately, my experimental design prevents us from differentiating between the effects of the Y^M chromosomes and standard third chromosome on the expression differences between these Y^M males. However, if differences between Y^M chromosomes were responsible for the elevated differential expression between the two Y^M male genotypes, this would suggest that variation amongst the effects of Y^M chromosomes in my experiment exceeds differences between Y^M and X chromosomes. Non-recombining Y chromosomes are expected to have low levels of polymorphism (Clark 1987, 1988). However, variation across *D. melanogaster* Y chromosomes has been shown to affect gene expression across the genome and may be involved in the resolution of sexual conflicts (Lemos *et al.* 2008, 2010). In addition, human Y chromosomes harbor high levels of copy number variation of ampliconic genes (Ye *et al.* 2018). Intriguingly, the house fly Y^M chromosome carries recently duplicated genes that differentiate it from the homologous X chromosome (Meisel *et al.* 2017). If these Y^M duplications vary in their copy number or if there are

chromatin-level differences across Y^M chromosomes, this could explain a possible effect of the Y^M chromosome on global gene expression. Additional work is necessary to test these hypotheses.

2.4.2 The III^M chromosome, *cis*-regulation, and sexual conflicts

Despite the minimal effects of the III^M chromosome on gene expression, I do identify two notable patterns across all types of males. First, higher proportions of genes on the third chromosome, relative to other chromosomes, are differentially expressed in many of my comparisons between males with different genotypes (Figure 2.6; Appendix Figure 2). Second, genes with discordant sex-biased expression between genotypic males and sex-reversed males are also over-represented on the third chromosome (Tables 2.2 and 2.3). This is consistent with my previous results showing that the III^M chromosome disproportionately promotes male-biased expression (Meisel *et al.* 2015). These results are contingent on inference of the chromosomal assignment of house fly genes, which I have confirmed is accurate by comparing with an independent mapping approach (Meisel and Scott 2018).

A high fraction of genes on the third chromosomes differentially expressed between males with and without the III^M chromosome might be indicative of divergence of *cis*-regulatory alleles between the III^M and standard third chromosomes. These expression differences of third chromosome genes could have important phenotypic effects that could partially be responsible for fitness differences between males with and without the III^M chromosome. Those fitness differences could in turn explain the maintenance of both the Y^M and III^M proto-Y chromosomes in natural populations.

Additional work is necessary to connect gene expression differences to fitness effects of the III^M chromosome.

The enrichment of genes with discordant sex-biased expression on the third chromosome between males with and without the III^M chromosome may be consistent with population genetics theory that predicts that sexually antagonistic selection is a major driver of the evolution of sex determination and the maintenance of polygenic sex determination (Orzack *et al.* 1980; van Doorn and Kirkpatrick 2007, 2010; Roberts *et al.* 2009; Ser *et al.* 2010; Parnell and Streelman 2013; Meisel *et al.* 2016). For example, sexual conflicts could be resolved if sexually antagonistic alleles are inherited in a sex-limited manner through the origination of a tightly linked sex-determining factor (van Doorn and Kirkpatrick 2007). In addition, male-beneficial alleles are expected to accumulate on proto-Y chromosomes once they have acquired male-limited inheritance (Rice 1992). The excess of discordant sex-biased genes on the third chromosome may be consistent with these theoretical predictions if the up- or down-regulation of these genes on the III^M chromosome is beneficial to males and deleterious to females. In this case, the male-beneficial (and female-detrimental) alleles would be *cis*-regulatory elements that affect the expression of the discordant sex-biased genes on the III^M chromosome (Tables 2.2 and 2.3). A similar phenomenon was observed in Lake Malawi cichlids, where an allele underlying a sexually antagonistic pigmentation phenotype is a *cis*-regulatory variant that up-regulates the expression of a gene linked to a new sex determiner (Roberts *et al.* 2009). Although the house fly male determiner (*Mdmd*) is molecularly characterized (Sharma *et al.* 2017), its location on the III^M chromosome is not known,

which prevents us from testing if the discordant sex-biased genes are nearby and genetically linked to the male determiner.

There are two considerations, however, that may be important limitations of these interpretations. First, the fitness effects of the proto-Y chromosomes appear to be environmentally dependent. Y^M is most frequent at northern latitudes and III^M predominates in the south, suggesting that temperature-dependent fitness differences could be responsible for north-south clines (Franco *et al.* 1982; Tomita and Wada 1989; Hamm *et al.* 2005; Feldmeyer *et al.* 2008; Kozielska *et al.* 2008). I did not test for temperature-dependent effects of the proto-Y chromosomes in my experiment, which may have prevented us from identifying key fitness-related gene expression differences between Y^M and III^M males. These temperature-dependent effects could be the result of temperature-dependent expression of genes on Y^M and III^M, differences in temperature-dependent activity of the copies of *Mdmd* across proto-Y chromosomes, or some other temperature-dependent genotype-by-environment interaction. Second, Y^M and III^M can be carried by females who also carry the epistatic *Md-tra*^{D} allele (Mcdonald *et al.* 1978; Hediger *et al.* 2010). The fitness differences between Y^M and III^M could therefore be mediated through the effects of the proto-Y chromosomes on female phenotypes, which I did not assay in my experiments.

2.4.3 The effect of *Md-tra* on gene expression

My results suggest that *Md-tra* has effects on gene expression beyond the direct regulation of *Md-dsx* and *Md-fru* splicing. Previous results, as well as my experiments here, demonstrate that knock down of *Md-tra* in blastoderm embryos causes complete

sex-reversal of genotypic females into fertile phenotypic males (Hediger *et al.* 2010). My results suggest that this sex-reversal does not affect all adult tissues equally—I observed one fertile sex-reversed male with higher *Md-tra* expression than normal females in head and a head gene expression profile that does not cluster with phenotypic females or males (Appendix Figure 4). Curiously, the outlier sex-reversed male in my experiment does not have a gene expression profile intermediate between genotypic males and females (Appendix Figure 4), as I would expect from partial masculinization. This suggests *Md-tra*-RNAi treatment in blastoderm embryo can have effects on adult somatic gene expression that does not act in the expected direction of sex-reversal.

I find additional evidence that *Md-tra* knockdown can affect adult gene expression independently of genotype. For example, the two genotypic males in my RNAi knockdown experiment have the same genotypic and phenotypic sex, yet their head gene expression profiles do not cluster together in my PC analysis; instead, genotypic males and females with *Md-tra*-RNAi treatment cluster together (Figure 2.3D). There are also more genes differentially expressed between III^M males with and without *Md-tra*-RNAi treatment than between genotypic males and sex-reversed males (Figure 2.4D). These results suggest that *Md-tra* affects head gene expression independent of genotypic sex. The effects of *Md-tra*-RNAi on head expression are likely mediated either through direct effects of *Md-tra* on the splicing of transcripts other than *Md-dsx* and *Md-fru*, downstream effects of *Md-dsx* and *Md-fru* alternative splicing, or off-target effects of dsRNA targeting *Md-tra*. In contrast, I do not observe a disproportionate effect of *Md-tra*-RNAi on abdominal gene expression—knocking down *Md-tra* converts genotypic

females into phenotypic (sex-reversed) males with expression profiles that nearly perfectly mimic genotypic (normal) males (Figures 2.3A, 2.4C, and 2.5A).

Notably, the expression of *Md-tra* does not differ across the heads of genotypic males or females with either RNAi treatment (Figure 2.3B). This suggests that the expression effects of knocking down *Md-tra* in adult heads is not through direct effects on *Md-tra*, but instead is caused by off-target effects or downstream effects of the direct targets of *Md-tra*. It is therefore possible that silencing *Md-tra* in early blastoderm embryos affects regulatory pathways that modulate head gene expression independently of the activity *Md-tra* in adult heads. Sex determination in flies is cell autonomous, and many cells in *Drosophila* somatic tissues do not express sex-determining genes downstream of *tra* (Robinett *et al.* 2010). My results suggest that even if somatic tissues do not differentially express sex-determining genes, they carry the memory of regulation of the sex determination pathway from their progenitor cells.

Curiously, there is an excess of third chromosome genes differentially expressed between III^M males with *Md-tra*-RNAi treatment and III^M males with *GFP*-RNAi treatment (Figure 2.6). The III^M chromosome is a proto-Y, and the standard third chromosome is a proto-X. Therefore, knockdown of *Md-tra* could be disproportionately affecting proto-Y genes or proto-X genes. Unfortunately, my data lack the resolution to determine if the expression changes between III^M males with different RNAi treatments is the result of changes in expression of genes on the III^M chromosome, standard third chromosome, or both. Regardless of which homolog is changing in expression, one explanation for the biased effect of *Md-tra* knockdown on third chromosome genes is that

there is an excess of third chromosome targets regulated by *Md-tra* or the sex determination pathway. For example, the house fly sex determination pathway could regulate gene expression specifically on the proto-X chromosome, analogous to how *Drosophila* X chromosome dosage compensation is controlled in a sex-specific manner by a gene in the sex determination pathway (Salz and Erickson 2010). Intriguingly, knockdown of *transformer* in red flour beetle, *Tribolium castaneum*, females causes them to produce nearly all male progeny, possibly as a result of misregulation of the diploid X chromosome in the female progeny (Shukla and Palli 2012). *Md-tra* in house fly may have a similar role regulating X chromosome expression. Additional work is necessary to evaluate why *Md-tra* knockdown disproportionately affects third chromosome expression.

2.4.4 Conclusions

I have performed multiple RNA-seq experiments in an attempt to resolve the paradox of ecologically relevant fitness effects of the house fly Y^M and III^M proto-Y chromosomes despite minimal sequence divergence between proto-Y and proto-X chromosomes. I identified some effects of the Y^M and III^M chromosomes on gene expression, but the number of differentially expressed genes and their effect sizes are small relative to the effect of a standard third chromosome or knockdown of the key sex determining gene *Md-tra*. Therefore, gene expression in house flies depends more on phenotypic sex (mediated by the sex determination pathway) than sex chromosome genotype. This is consistent with a recent study in *Rana temporaria* frogs that have polygenic sex determination, which found that sex-biased gene expression depends more on phenotypic

sex than genotypic sex (Ma *et al.* 2018). Thus, I hypothesize that the geographic distribution of the Y^M and III^M chromosomes arises primarily from selection on environmentally sensitive phenotypes that I did not assay in my experiments. Because seasonality of temperature is predictive of the frequencies of Y^M and III^M in natural populations (Feldmeyer *et al.* 2008), a fitness or phenotypic assay across temperatures may be needed to identify ecologically relevant differences between Y^M and III^M males.

Chapter 3

The expression in sex determination pathway genes depending on temperature could not explain the latitudinal Y^M-III^M clines for house fly polygenic sex determination

3.1 Introduction

Sex determination can be triggered by heritable genetic variation (e.g. a male- or female-determining locus on the Y or W chromosome) or environmental cues (e.g. developmental temperature) (Bull 1983; Beukeboom and Perrin 2014; Bachtrog *et al.* 2014). The initial signal is then transmitted to downstream parts in the sex determination pathway. The master regulators that trigger sex determination often differ across species (Beukeboom and Perrin 2014; Bachtrog *et al.* 2014), and they can even be variable within species, which is known as polygenic sex determination (Moore and Roberts 2013). Species with the polygenic sex determination segregate for multiple male- and/or female-determining loci on different chromosomes (Moore and Roberts 2013). Polygenic sex determination has been observed in many species (Orzack *et al.* 1980; Moore and Roberts 2013; Bachtrog *et al.* 2014) even though population genetic models predict that polygenic sex determination exists only transiently during the transition between two monogenic sex determination systems (van Doorn 2014a, 2014b). This suggests that polygenic sex determination can be maintained stably, in contrast to the population genetics theory. However, the specific selection pressures responsible for the maintenance of polygenic sex determination remain elusive (Bull and Charnov 1977; Rice 1986; van Doorn 2014a, 2014b).

House fly, *Musca domestica*, is a model system for studying polygenic sex determination because it has a male-determining gene, *Mdmd*, on multiple chromosomes (Hamm *et al.* 2015; Sharma *et al.* 2017). *Mdmd* causes male development by preventing the house fly ortholog of *transformer* (*Md-tra*) from being spliced into a female-

determining isoform (Hediger *et al.* 2010). *Mdmd* has been found most frequently on the Y (Y^M) and third (III^M) chromosomes. Y^M and III^M are both considered to be proto-Y chromosomes because they have few morphological and sequence differences from their homologous X chromosomes (Boyes *et al.* 1964; Hediger *et al.* 1998; Meisel *et al.* 2017). These proto-Y chromosomes form stable latitudinal clines; males with the Y^M chromosome (Y^M males) are most common in northern populations where temperatures are colder, and males with III^M chromosomes (III^M males) dominate the low latitudes where temperature is warmer (Franco *et al.* 1982; Tomita and Wada 1989; Hamm *et al.* 2005; Kozielska *et al.* 2008; Hamm *et al.* 2015). The frequencies of the proto-Y chromosomes across house fly populations have been almost unchanged for decades (Kozielska *et al.* 2008; Meisel *et al.* 2016). These observations suggest that spatially heterogeneous, temperature-dependent selection pressures maintain the Y^M and III^M chromosomes across house fly populations (Feldmeyer *et al.* 2008). This is similar to how theoretical and empirical studies have shown that sex-specific selection pressures can be variable across environments and along a cline (Delph *et al.* 2011; Punzalan *et al.* 2014; Connallon 2015), and the interactions between genotype and environment in sex-specific selection can be responsible for maintaining polymorphism (Ingleby *et al.* 2010).

Mdmd is found in a tandemly repeated array on each proto-Y chromosome (Sharma *et al.* 2017). Most of the copies of *Mdmd* within each array have a truncated open reading frame, and there are different copy numbers of *Mdmd* on each proto-Y chromosome (e.g. III^M vs Y^M). Gene conversion between the *Mdmd* elements of these arrays could rescue loss of function mutations (i.e. prevent Muller's ratchet) in the

absence of X-Y recombination, similar to what has been proposed for Y chromosome duplications in other taxa (Connallon and Clark 2010). Importantly, differences between *Mdmd* alleles on each proto-Y chromosome, as well as copy number differences, might influence the expression of *Mdmd*, which could differentially affect the splicing of *Md-tra*, possibly in temperature-dependent way.

If temperature-dependent selection pressures cause the stable maintenance of Y^M-III^M latitudinal clines, then I expect that the phenotype and fitness of Y^M and III^M males would vary across a temperature gradient. I tested this hypothesis in this study by evaluating if the expression and splicing of *Mdmd* and *Md-tra* depend on temperature in a way that are consistent with temperature-dependent phenotypic differences between Y^M and III^M. *Mdmd* promotes male development by negatively regulating *Md-tra*, causing *Md-tra* to be spliced into non-functional isoforms with premature stop codons that cannot be translated into a functional female-determining protein (Hediger *et al.* 2010). Embryos without *Mdmd* develop into females because *Md-tra* is spliced into an isoform that is translated into a functional female-determining Md-TRA^F protein (Figure 3.1). There is evidence that alternative splicing in other flies varies depending on genotype-by-temperature interactions (Jakšić and Schlötterer 2016). I tested if temperature affects the abundance of *Mdmd* transcripts and the alternative splicing of *Md-tra* differently between Y^M and III^M males. To test for a genotype-by-temperature (G×T) interaction affecting *Mdmd* expression and the splicing of *Md-tra*, I used RT-qPCR (quantitative reverse transcription PCR) to measure the expression of *Mdmd* and abundance of *Md-tra* isoforms in abdomens and heads of Y^M and III^M males that developed at low (18 °C) or

high (27°C) temperatures.

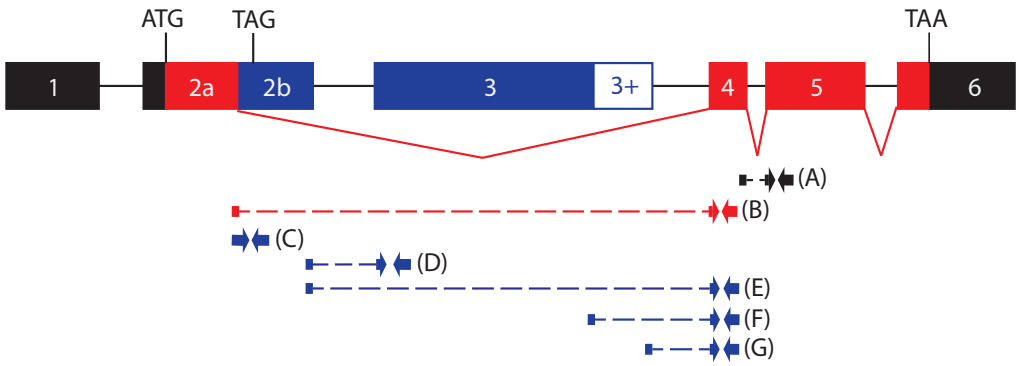


Figure 3.1. Schematic representation of the *Md-tra* locus based on DNA sequencing, cDNA clones, and RNA-Seq data (Hediger et al. 2010, Scott et al. 2014). Exons are colored and numbered: black exons contain untranslated sequences, red exons are included in the female-determining transcript encoding Md-Tra^{F}, and blue exons are specific to the male-specific transcript. Splicing of the female-determining transcript is illustrated by the red diagonal lines connected exons. The start and stop codon locations are shown. Locations of forward and reverse qPCR primers for splice junctions found in (A) all transcripts, (B) female transcripts, and (C-G) male transcripts are indicated by inward facing arrows.

3.2 Materials and Methods

3.2.1 House fly strains and rearing

I measured expression and splicing of *Md-tra* and *Mdmd* in two III^{M} strains and two Y^{M} strains that each has a different naturally occurring proto-Y chromosome on a common genetic background. One III^{M} strain (CS) is used as the common background. CS is an inbred and lab-adapted strain that was produced by mixing strains collected throughout United States (Scott *et al.* 1996). The other III^{M} strain (CSkab) was created by backcrossing the III^{M} chromosome from the KS8S3 strain collected in Florida (Kaufman *et al.* 2010) onto the CS genetic background. A Y^{M} strain (CSaY) has the same genetic background as CS, with a Y^{M} chromosome that came from the reference genome strain,

aabys, instead of the CS III^M chromosome (Meisel *et al.* 2015). The last Y^M strain (IsoCS) was created by introducing the Y^M chromosome from a strain collected in Maine onto the CS background without III^M (Hamm *et al.* 2009). The CS and CSaY strains were used to quantify the temperature-dependent splicing of *Md-tra* at seven exon-exon junctions. CSkab and IsoCS were assayed to quantify the expression of *Mdmd* as well as the splicing of *Md-tra* at two exon-exon junctions (2a-4 and 4-5 in Figure 3.1). Colonies of each strain were raised at 18 °C and 27 °C for two generations. House fly strains raised at 18 °C were transferred to 22 °C for collecting eggs because I was not able to recover progeny at 18°C. Male progeny were collected from offspring produced by females of the first generation kept a particular temperature, and males were aged for five days separately from females at the appropriate developmental temperatures (18 °C or 27 °C). A colony of another strain collected from Chino, CA in 2014 (Meisel *et al.* 2016) was reared at 25 °C to provide RNA for making standard curves for the qPCR assays.

3.2.2 Tissue samples, RNA extraction, and cDNA

Abdomens and heads were dissected from five-day old adult males that were anesthetized with carbon dioxide (CO_2). Dissections of each tissue sample from 5-7 males were pooled to create three biological replicates for each genotype-by-temperature (GxT) combination. The collected abdomens and heads were homogenized in TRIzol reagent (Life Technologies) with a motorized grinder, and a Direct-zol miniprep (Zymo Research) was used to extract RNA. The isolated RNA was reverse transcribed into cDNA with M-MLV RT (Promega), following the manufacturer's protocol.

3.2.3 qPCR assays

I designed seven pairs of primers to amplify *Md-tra* exon-exon junctions, with one primer spanning the corresponding exon-exon junction (Appendix Table 4). I used a pair of primers residing in the second exon of *Mdmd* to amplify *Mdmd* transcripts. The primer pair is specific to *Mdmd*, and it does not amplify the paralog of *Mdmd*, called *Md-ncm* (Sharma *et al.* 2017). A different primer pair was used to specifically amplify *Md-ncm* (Sharma *et al.* 2017). A primer pair on a transcript shown in previous study (Meisel *et al.* 2015) not to be differentially expressed between Y^M and III^M males (LOC101888902) was used for internal control for cDNA content in each biological replicate. I conducted qPCR of cDNA on a StepOnePlus machine using PowerUp SYBR Green Master Mix (Applied Biosystems). I measured the expression of the transcript regions corresponding to the primer pairs above in three technical replicates of the three biological replicates for each GxT combination. With the same primer pairs, I also measured the expression of serial dilutions (1/1, 1/5, 1/25, 1/125, and 1/625) of cDNA from house flies collected from Chino to establish standard curves.

Samples and primer pairs combinations were interspersed across 15 96-well plates to minimize batch effects. Standard curves were constructed for each primer pair by calculating the linear relationship between CT values and $\log_{10}$(concentration in the serial dilution) using the lm() function in the R statistical programming package (R CoreTeam 2013). I used the equations of the standard curves to calculate the concentration of transcripts (cDNA) from each *Md-tra* exon-exon junction and *Mdmd* in each technical replicate. I then determined a normalized expression level of each exon junction in each

technical replicate by dividing the concentration of the technical replicate by the mean concentration of the control transcript across the three technical replicates from the same biological replicate.

3.2.4 Statistical analysis

I used the lmer() function in lme4 package (Bates *et al.* 2017) in R to model the effect of genotype (*G*), temperature (*T*), and the interaction term as fixed effect factors, as well as biological replicate (*R*) as a random effect, on expression level:

$$\text{Expression} \sim G + T + G \times T + R$$

I also created a model without the interaction term, and then I performed a likelihood ratio test with the anova() function in R to determine if the model with the interaction term fits significantly better than the model without. Because the common exon junction is used in both the male-specific isoforms and female-specific isoforms it can be used as an indicator the total amount of *Md-tra* transcripts. I constructed a model that included the expression of both the female-specific isoform and the common exon junction (*X*) as fixed effect, as well as a three-way interaction.

$$\text{Expression} \sim G + T + X + G \times T + G \times T \times X + R$$

This full model was compared to a model without the three-way interaction using a likelihood ratio test with the anova() function in R. If the full model fits significantly better than the model without the interaction term, then we conclude that the interaction term has a significant effect on the total amount of *Md-tra* transcripts. If there is not a significant difference in model fit, then the three-way interaction term does not have a significant effect on *Md-tra* expression.

3.3 Results

The geographical distribution of Y^M and III^M are predicted by seasonality of temperature (Feldmeyer *et al.* 2008), suggesting that temperature-dependent selection pressures maintain polygenic sex determination in house fly. Y^M is found in the north where temperatures are colder, and III^M is found in the south where temperatures are warmer (Franco *et al.* 1982; Tomita and Wada 1989; Hamm *et al.* 2005; Kozielska *et al.* 2008). I tested if temperature-dependent selection pressures on Y^M and III^M are the result of GxT interactions that affect the abundance of the *Mdmd* transcripts or the splicing of *Md-tra*. For example, males may have maximal fitness if *Md-tra* transcripts are spliced into isoforms other than the female determining isoform (Md-TRA^F) with higher fidelity. If temperature-dependent selection pressures on Y^M and III^M result from effects of temperature on the fidelity of *Md-tra* splicing, then I predict that the abundance of the Md-TRA^F isoform will be greatest in Y^M (III^M) males raised at high (low) temperatures because these temperatures are incongruent with the natural distribution of Y^M and III^M. Moreover, *Mdmd* negatively regulates splicing of *Md-tra* into the Md-TRA^F isoforms, which suggests that greater *Mdmd* expression could result in reduced production of Md-TRA^F. I therefore also predict that *Mdmd* expression will be greatest in Y^M males raised at low temperatures and III^M males raised at high temperatures.

3.3.1 Temperature-dependent expression of *Mdmd*

I first used RT-qPCR to quantify the expression of *Mdmd* mRNA in the abdomens of Y^M and III^M males that developed at low (18 °C) or high (27 °C) temperature. I also examined the expression of the *Mdmd* paralog *Md-ncm* (Sharma et al. 2017). This

experiment tests the hypothesis Y^M and III^M are maintained across populations because of temperature-dependent differences in the expression of *Mdmd*. I performed analysis of variance (ANOVA) to examine the interaction effect between genotype and temperature on *Mdmd* expression measured using RT-qPCR. The GxT interaction effect on the expression of *Mdmd* was not significant ($P = 0.1021$) (Figure 3.2A). Moreover, *Mdmd* expression is not highest in III^M (Y^M) males at 27 °C (18 °C), which would be the expected pattern if *Mdmd* expression explains the distribution of III^M and Y^M. There is also not a significant GxT interaction for the expression of *Md-ncm* ($P = 0.3233$) (Figure 3.2B). These results suggest that temperature-dependent expression of *Mdmd* cannot explain the geographic distribution of Y^M and III^M (Franco *et al.* 1982; Tomita and Wada 1989; Hamm *et al.* 2005; Kozielska *et al.* 2008).

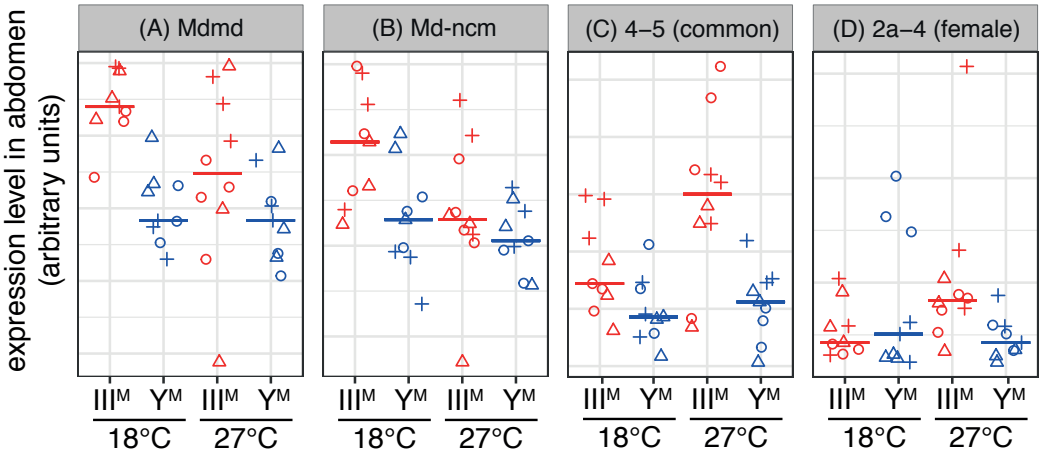


Figure 3.2. Expression levels of *Mdmd* (A), and *Md-ncm* (B), and two *Md-tra* exon junctions (C, D) in the abdomens of CSkab (III^M) and IsoCS (Y^M) males raised at 18°C and 27°C. Each point is a technical replicate, and points with the same shape are from the same biological replicate. The horizontal lines indicate the median across all replicates.

3.3.2 Temperature-dependent splicing of *Md-tra*

I failed to find evidence of a GxT interaction for *Mdmd* expression (Figure 3.2A-B), suggesting that *Mdmd* regulation of *Md-tra* splicing is unlikely to be a phenotype under temperature-dependent selection to maintain polygenic sex determination in house fly. However, that does not rule out the possibility that there is temperature-dependent splicing of *Md-tra* that depends on Y^M/III^M genotype independently of *Mdmd* expression level. To test that hypothesis, I measured the expression of the *Md-tra* female-determining isoform (Figure 3.1B), a splice junction shared by all predicted *Md-tra* transcripts (Figure 3.1A), and *Md-tra* transcripts carrying "male" splice junctions (found in isoforms with premature stop codons; Figure 3.1C-G). Four of the five exon junctions that are found in male isoforms (Figure 3.1C-G) and the one junction that is common to all isoforms (Figure 3.1A) were detected in all male genotypes that I assayed. Surprisingly, the female splice junction (Figure 3.1B) was detected in all but one technical replicate of males.

I used the expression level at the female splice junction of *Md-tra* (Figure 3.1B) as an indicator of the fidelity of *Md-tra* splicing in Y^M and III^M males. To that end, I conducted ANOVA to determine if there is a significant effect of the interaction between genotype (III^M vs Y^M) and developmental temperature (18 °C vs 27 °C) on the expression at each splice junction. I performed experiments in two pairs of Y^M and III^M males using the splice junction in common to all isoforms and the splice junction specific to the female-determining isoform. In one Y^M-III^M pair there is not a significant GxT interaction effect in the expression of the *Md-tra* female splice junction (Figure 3.2D), but there is a

significant GxT effect in the other pair of strains (Figure 3.3B). In the latter pair of strains, expression of the exon-exon junction unique to the female isoform (2a-4) is higher in III^M males than Y^M male abdomen at 18 °C, and expression of the female splice junction is higher in Y^M male abdomens at 27 °C (Figure 3.3B). This result provides mixed support for my hypothesis that GxT interactions for the fidelity of *Md-tra* splicing are responsible for the maintenance of Y^M-III^M clines.

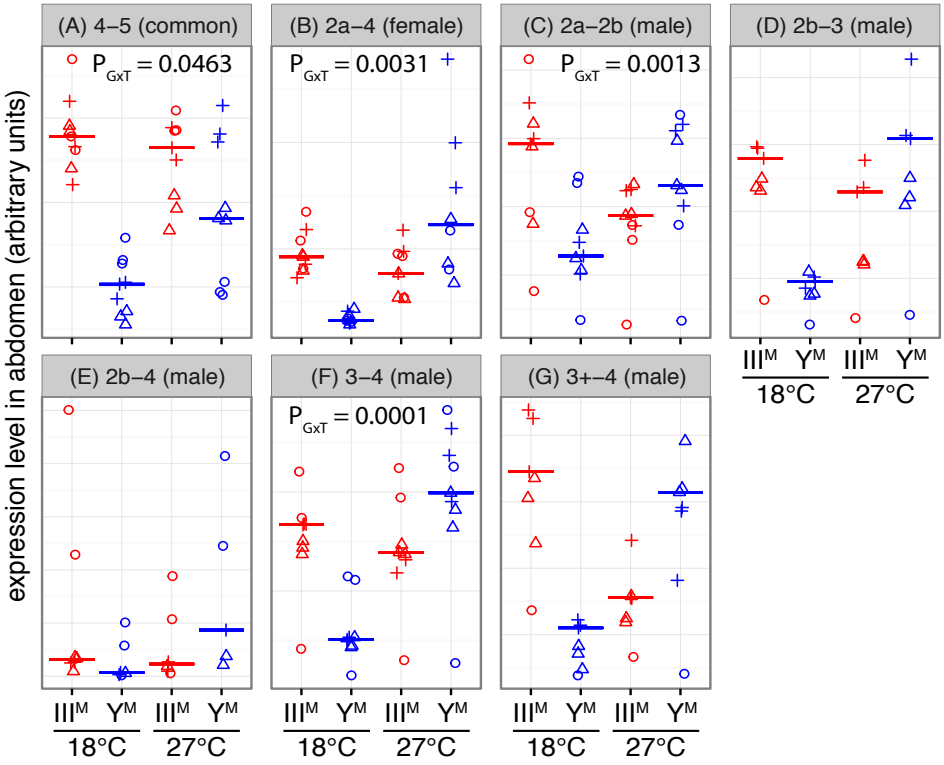


Figure 3.3. Expression levels of each *Md-tra* exon junction in the abdomens of CS (III^M) and CSaY (Y^M) males raised at 18 °C and 27 °C. Each point is a technical replicate, and points with the same shape are from the same biological replicate. Extreme outliers are excluded from this plot. The horizontal lines indicate the median across all technical replicates, including outliers. P values indicate a significantly better fit in a model with a G × T interaction than a model without.

To further examine the effect of GxT interactions on *Md-tra* splicing, I considered the expression of other exon-exon junctions. I first examined the exon junction in

common to all splice variants (Figure 3.1A). A model that includes the G×T interaction term does not fit significantly better that a model without the interaction term for one pair of Y^M-III^M strains (Figure 3.2C), but a model with the interaction term does fit significantly better ($P < 0.05$) for a different pair of strains in abdomen (Figure 3.3A). In addition, a model with an interaction term fits significantly better than one without for two exon junctions unique to male transcripts (Figure 3.3C, F), but not for three other male junctions (Figure 3.3D, E, G). I therefore find a similar amount of evidence for G×T interactions affecting the expression of non-functional isoforms as I do for production of the female isoform of *Md-tra*. These results suggest that temperature-dependent *Md-tra* splicing, especially mis-expression of Md-TRA^F in males does not explain the predominance of Y^M at northern latitudes and III^M at southern latitudes (Franco *et al.* 1982; Tomita and Wada 1989; Hamm *et al.* 2005; Kozielska *et al.* 2008; Hamm *et al.* 2015).

I performed one final analysis of my data to control for the overall *Md-tra* expression level in my measurement of *Md-tra* splicing. I analyzed a linear model that contains the expression of both the female splice junction (as a measuring of splicing fidelity) and the common junction (as a measure of *Md-tra* expression level), and I included a three-way interaction between genotype, temperature, and exon junction in the model. I analyzed both pairs of Y^M and III^M strains separately. For the CS (III^M) and CSaY (Y^M) strains (Figure 3.3A, B), the full model including the three-way interaction fits significantly better than the model excluding the three-way interaction ($\chi_3^2 = 10.5$, $P = 0.0145$). However, the model with the three-way interaction effect does not fit

significantly better ($\chi^2_3 = 5.33$, $P = 0.1494$) using *Md-tra* splicing from CSkab (III^M) and CSaY (Y^M) (Figure 3.2C, D). These results are consistent with my analysis using two-way interactions to assess the fidelity of *Md-tra* splicing. They similarly provide mixed support for the hypothesis that genotype-by-temperature interactions influence the splicing of the *Md-tra* female isoform. The inconsistent patterns in temperature-dependent mis-expression of the *Md-tra* female isoform suggests that the female isoform is not mis-expressed in a temperature-dependent way across III^M and Y^M genotypes.

3.4 Discussion

This study revealed that *Mdmd* expression and *Md-tra* splicing likely do not depend on GxT interactions in a way that can explain the maintenance of stable Y^M-III^M clines in house fly. Therefore, temperature-dependent expression or splicing of sex determining genes cannot explain the maintenance of polygenic sex determination in house fly. In addition, I found that the female isoform of *Md-tra* is expressed in house fly males, even though the expression of the female isoform in males does not depend on a genotype-by-temperature interaction. It remains unclear whether the level of mis-expression of Md-TRA^F in males has any deleterious phenotypic or fitness consequences in males.

My results suggest that the latitudinal cline of Y^M and III^M males are not explained by differences in the male-determining *Mdmd* gene on the Y^M III^M chromosomes. This suggests that the cline could be better explained by selection acting on alleles at other loci on the Y^M or III^M chromosomes. Both Y^M and III^M are new proto-Y chromosomes (Meisel *et al.* 2017). These proto-Y chromosomes might influence the expression of genes throughout the genome, resulting in phenotypic and fitness

differences between Y^M and III^M chromosomes. For example, heterochromatinization and structural variation in Y chromosomes of *Drosophila* species affect the expression of many genes (Francisco and Lemos 2014; Brown and Bachtrog 2017). In addition, *Drosophila* Y chromosomes can affect splicing and sex-biased intron retention (Wang *et al.* 2018). The house fly Y^M and X chromosomes show high level of heterochromatinization and low gene content, which is very different from the euchromatic autosomes (Boyes *et al.* 1964; Meisel *et al.* 2017). It is therefore plausible that X/Y^M genotype could affect phenotypic differences between Y^M and III^M males via effects of X and Y^M chromosomes on gene expression or splicing. Alternative splicing patterns in *D. melanogaster* also vary across genotypes and temperatures (Jakšić and Schlötterer 2016). Therefore, the interaction of genotype and temperature affecting phenotypic variation could reasonably be mediated through gene expression or alternative splicing independently of *Mdmd*, which may be responsible for the geographic distribution of house fly proto-Y chromosomes.

3.4.1 Conclusions

I tested the hypothesis that temperature-dependent selection could explain the latitudinal clines of Y^M and III^M males in natural populations (Franco *et al.* 1982; Tomita and Wada 1989; Hamm *et al.* 2005; Feldmeyer *et al.* 2008; Kozielska *et al.* 2008; Hamm *et al.* 2015). Specifically, I performed RT-qPCR to investigate if there is a G×T interaction for the expression of *Mdmd* and the splicing of *Md-tra*. I found that *Mdmd* expression does not show temperature-dependent expression and there is mixed support for a G×T interaction for *Md-tra* splicing. Therefore temperature-dependent expression of *Mdmd* or

splicing of *Md-tra* are unlikely to explain stable maintenance of Y^M-III^M clines. I suggest alternative hypotheses that the phenotypic consequences of different proto-Y chromosomes that result in latitudinal clines come from the temperature-dependent effects of alleles of other genes on Y^M and III^M or heterochromatin on Y^M that affects global gene expression.

Chapter 4

A very young Y chromosome is cryptically differentiated from a homologous X chromosome at gene sequence and expression levels

4.1 Introduction

In many organisms with two separate sexes, a gene on a sex chromosome determines whether an individual become a male or female. In XX/XY sex chromosome systems, males are the heterogametic sex (XY genotype), and females have the XX genotype (Bull 1983). X and Y chromosomes are derived from a pair of ancestral autosomes when one homolog obtains a sex-determining locus, and the X-Y pair diverge from each other over time (Bull 1983; D. Charlesworth *et al.* 2005) Sex chromosomes have originated and diverged from each other in multiple independent evolutionary lineages.

Despite their independent origins, independently evolved Y chromosomes share many common features across species (Figure 4.1). First, "masculinization" occurs because male-limited inheritance of the Y chromosome favors the fixation of male-beneficial genetic variation (Rice 1996). Second, "degeneration" occurs in nonrecombining regions; functional genes that were present on ancestral chromosomes are pseudogenized because suppressed recombination between the X and Y inhibits the purging of deleterious mutations in Y-linked genes (Muller's ratchet) and enhances the effects of hitchhiking (B. Charlesworth and Charlesworth 2000; Bachtrog 2013). Other common features of Y chromosomes are repetitive sequences and enlarged heterochromatic regions due to reduced effectiveness of purifying selection caused by suppressed recombination (Skaletsky *et al.* 2003). Recombination suppression has been considered to come after the emergence of a new sex-determining locus on a Y chromosome, and it is hypothesized to favor the co-inheritance of the sex-determining locus and male-beneficial/female-detrimental sexually antagonistic alleles. As additional

sexually antagonistic alleles accumulate on a Y chromosome, this is predicted to trigger progressive spread of the nonrecombining region along the Y chromosome (Rice 1987b; van Doorn and Kirkpatrick 2007). Although these features have been characterized in sex chromosomes of various ages and degeneration levels (Bachtrog 2013; Zhou *et al.* 2014), the very first stages of Y chromosome evolution are poorly understood because of a lack of extremely young sex chromosome systems. This study addresses this shortcoming by determining how a young "proto-Y" chromosome has been differentiated from its homologous proto-X chromosome shorty after its emergence.

I used the house fly, *Musca domestica*, as a model system to study the early evolution of sex chromosomes because it has very young sex chromosomes that are still segregating as polymorphisms within natural populations (Hamm *et al.* 2015). The house fly has a male-determining gene (*Mdmd*) on the Y chromosome (Y^M) and on all five autosomes (A^M; I-V) (Sharma *et al.* 2017). Each chromosome carrying *Mdmd*, including Y^M, is a recently derived proto-Y chromosome (Meisel *et al.* 2017). The proto-Y and X chromosomes show minimal sequence and morphological divergence, as well as similar gene content (Boyes *et al.* 1964; Hediger *et al.* 1998; Meisel *et al.* 2017), consistent with their recent origin. It is not clear the extent to which the proto-Y chromosomes are masculinized or degenerated. Studies in Chapter 2 revealed a small, but significant, effect of the proto-Y chromosomes on gene expression. However, it could not resolve if the expression differences are the result higher or lower expression of the proto-Y or X.

In this chapter, I tested if the III^M proto-Y chromosome has evidence of differentiation from its homologous proto-X chromosomes by evaluating gene expression

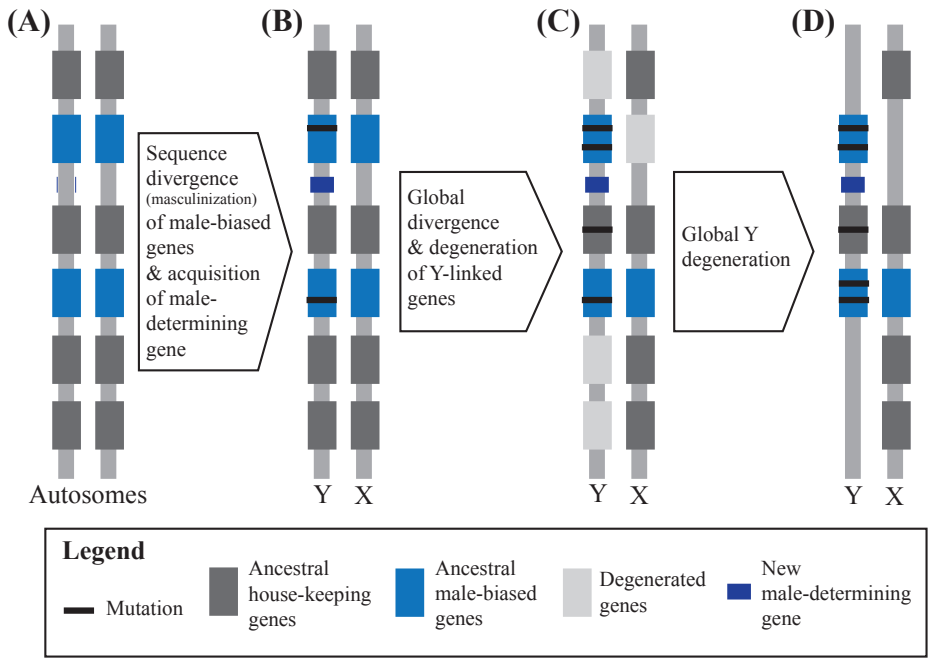


Figure 4.1. Model for sex chromosome evolution. (A) X and Y chromosomes are derived from a pair of autosomes with the same gene content. (B) A new proto-Y chromosome obtains a male-determining gene and male-biased genes start to diverge between the new Y and X chromosomes at the sequence level. (C) Over evolutionary time, most of genes on the Y chromosome accumulate deleterious mutations and many of genes begin to degenerate. (D) The Y chromosome continues to accumulate male-biased genes and degenerate by losing many genes.

and DNA sequence differences between proto-Y genes and their proto-X counterparts. I compared normal (genotypic) males with a III^M proto-Y chromosome to sex-reversed males with no proto-Y chromosome using the RNA-seq data created in the *Md-tra* RNAi experiment in Chapter 2. The genotypic males contain one copy of each of the proto-Y and proto-X, while the sex-reversed males carry only two copies of the proto-X chromosome. This approach represents an improvement over a previous study that compared III^M males with females (Meisel *et al.* 2017) because my experiment uses sex reversed males that are phenotypically nearly identical to normal (genotypic) males (see Chapter 2). This allows me to control for the effect of sexually dimorphic gene

expression on the inference of divergence between the proto-Y (III^M) and proto-X (standard third chromosome). If the proto-Y chromosomes are at the early stage of divergence from the proto-X chromosomes, then I expect that the genotypic (III^M/III) males would exhibit increased heterozygosity on the third chromosome compared to sex-reversed (III/III) males (Meisel *et al.* 2017). As the proto-Y and proto-X chromosomes diverge, it is also expected that alleles on the proto-Y chromosomes are up- or down-regulated because of *cis*-regulatory sequence differences that contribute to proto-Y gene expression. I therefore expect a higher proportion of allele-specific expression (ASE) from third chromosome genes in genotypic males than sex-reversed males because the genotypic males are heterozygous for their third chromosome (III^M/III) while the sex-reversed males are homozygous for a standard third chromosome carrying no *Mdmd*.

4.2 Methods

To compare gene expression between the standard third chromosome (III) and the III^M proto-Y chromosome, I analyzed RNA-seq data from the RNAi experiment described in Chapter 2, which includes normal III^M males and sex-reversed males (genotypic females that do not carry a proto-Y chromosome but are phenotypically male). Therefore, the two males only differ in whether they have a III^M proto-Y chromosome. The sex-reversed males are homozygous for a standard third chromosomes, whereas the normal males are heterozygous for a standard third chromosome and a III^M proto-Y. Because two copies of a chromosome could show heterozygosity at the sequence level even if they come from an inbred strain, I used the heterozygosity of the sex-reversed males on each chromosome as baselines in order to examine the levels of heterozygosity in normal males.

4.2.1 Heterozygous SNPs

I used RNA-seq data to identify genetic variants that differentiate the III^M proto-Y chromosome from the standard third chromosome, and then I tested if III^M males have elevated heterozygosity on the third chromosome (Vicoso and Bachtrog 2015; Meisel *et al.* 2017). I used the Genome Analysis Toolkit (GATK) pipeline for calling variants in the RNA-seq data from the *Md-tra* RNAi experiment in Chapter 2, following the best practices for SNP and indel calling on RNA-seq data (McKenna *et al.* 2010; Meisel *et al.* 2017). I used STAR (Dobin *et al.* 2013) to align reads from three genotypic (III^M/III) male libraries and three sex-reversed (III/III) male libraries to reference assembly v2.0.2 and annotation release 102 of the house fly genome (Scott *et al.* 2014). The aligned reads were used to generate a new reference genome index from the detected splice junctions in the first alignment run, and then a second alignment was performed with the new reference. I next marked duplicate reads from the same RNA molecule and used the GATK tool 'SplitNCigarReads' to reassign mapping qualities to 60 with the 'ReassignOneMappingQuality' read filter for alignments with a mapping quality of 255. Indels were detected and realigned with 'RealignerTargetCreator' and 'IndelRealigner'. The realigned reads were used for base recalibration with 'BaseRecalibrator' and 'PrintReads'. The base recalibration was performed in three sequential iterations in which recalibrated and filtered reads were used to train the next round of base recalibration, at which point there were no beneficial effects of the base calibration as verified by 'AnalyzeCovariates'. I next used the recalibrated reads from all three replicates of genotypic and sex-reversed males to call variants using 'HaplotypeCaller' with emission

and calling confidence thresholds of 20. I applied 'genotypeGVCFs' to the variant calls from the two types of males for joint genotyping between the two, and then I filtered the variants using 'VariantFiltration' with a cluster window size of 35 bp, cluster size of three SNPs, FS > 20, and QD < 2. The final variant calls were used to identify heterozygous SNPs within genes using the coordinates from the genome sequencing project (Scott et al. 2014). I measured relative heterozygosity within each gene in genotypic (III^M/III) and sex-reversed (III/III) males as the percent of heterozygous SNPs in genotypic males over total heterozygous SNPs in both genotypic and sex-reversed males.

4.2.2 Allele-specific expression

Diploid species can have two alleles at a locus, one of which was inherited maternally and the other paternally. The maternal and paternal alleles can be expressed unequally in the diploid, which is called allele-specific expression (ASE). I investigated if there is elevated ASE on the third chromosome in males carrying one III^M proto-Y and one proto-X chromosome compared to sex-reversed males with two proto-X chromosomes. To do this, I implemented the IDP-ASE tool at the gene level, following the developers' recommended analysis steps (Deonovic *et al.* 2016). I first prepared information on the number and locations of variants within each gene, as well as read counts at each variant location, which were supplied to the IDP-ASE software as raw reads, aligned reads, and variant calls created by GATK. The prepared data from each gene was next run in an MCMC (Markov chain Monte Carlo) sampling simulation to estimate the haplotype within each gene with a Metropolis-Hastings sampler (Bansal *et al.* 2008). Next, the

software estimates the proportion of each estimated haplotype that contributes to the total expression of the gene (ρ) from each iteration using slice sampling (Neal and others 2003). A value of ρ=0.5 indicates equal expression between two alleles, whereas ρ<0.5 or ρ>0.5 indicates ASE. The MCMC sampling was run with a 1000 iteration burn-in followed by at least 500 iterations where data were recorded. The actual number of iterations was automatically adjusted during the simulation to produce the best simulation output for quantifying ASE within a gene; this usually involved increased iterations for my RNA-seq data. The IDP-ASE simulation generated a distribution of ρ for each gene across all post-burn-in iterations, and then it calculated the proportion of iterations with ρ > 0.5. This proportion was used to estimate the extent of ASE for each gene. For example, if all iterations for a gene have ρ > 0.5, then the proportion is 1 and the gene has strong evidence for ASE. Similarly, if all iterations for a gene have ρ < 0.5, then the proportion is 0 and the gene also has strong evidence for ASE. In contrast, if half of the iterations have ρ > 0.5 and the other half have ρ < 0.5, then the proportion is 0.5 and there is not any evidence for ASE.

IDP-ASE only quantifies ASE within bi-allelic loci, so I only included genes with heterozygous sites within transcripts in genotypic (III^M/III) or sex-reversed (III/III) males. In addition, I removed heterozygous SNPs with the same genotype in genotypic and sex-reversed males because they do not allow me to discriminate between the proto-Y and proto-X alleles. Removing these SNPs may have also sped up the simulation times, but this was not rigorously investigated.

4.3 Results

4.3.1 DNA sequence divergence between the proto-Y and proto-X chromosomes

There are few sequence differences between the house fly proto-X and proto-Y chromosomes, suggesting that they are very young sex chromosomes (Meisel *et al.* 2017). Although the proto-X and proto-Y chromosomes are very similar in gene density and gene content, early X-Y divergence at the sequence level could lead to elevated heterozygosity in heterogametic (XY) males (Vicoso and Bachtrog 2015). Therefore, I tested for elevated heterozygosity by investigating SNPs within genes in the genotypic (III^M/III) and sex-reversed (III/III) males described in Chapter 2.

If the III^M chromosome is a very young proto-Y, I expect that the III^M males will have more heterozygous SNPs on the third chromosome compared to sex-reversed males as a consequence of divergence between III^M and the standard third chromosome. I calculated the proportion of heterozygous SNPs within genes in the genotypic (III^M/III) males relative to the sex-reversed (III/III) males. As expected, the genotypic (III^M/III) males have an excess of heterozygous SNPs on the third chromosome, relative to the sex-reversed (III/III) males (Figure 4.2; $P < 10^{-16}$ in a Wilcoxon rank sum test comparing genes on the third chromosome with genes on the other chromosomes). These results suggest that the sequences of genes on the III^M proto-Y chromosome are differentiated from the copies on the proto-X chromosome.

Although I predicted that there would be a similar level of heterozygosity on the X chromosome in the genotypic (III^M/III) males and sex-reversed (III/III) males due to the presence of two copies of the X chromosome in both males, the III^M males have

elevated heterozygosity on the X chromosome (Figure 4.2; $P = 8.32\times10^{-13}$ in a Wilcoxon rank sum test comparing the X chromosome with Chromosomes I, II, IV, and V). This was also observed in a comparison between females and III^M males (Meisel *et al.* 2017), and the cause of the elevated X chromosome heterozygosity in III^M males remains unresolved.

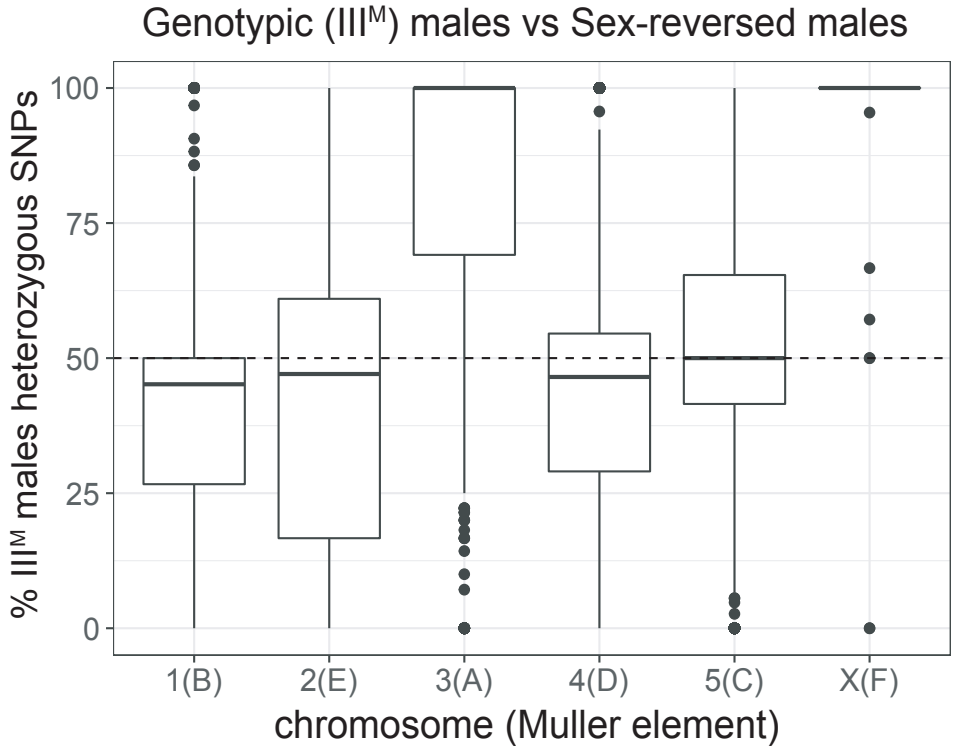


Figure 4.2. The elevated heterozygosity is shown on the third and X chromosomes in the genotypic (III^M/III) males relative to the sex-reversed (III/III) males. The boxplot shows the distribution of percentage of heterozygous SNPs within genes on each chromosome in the genotypic males relative to the sex-reversed males. Values more than 50% indicate the increased heterozygosity in genotypic (III^M/III) males. The expected median across all autosomes as null are represented by a dashed line.

4.3.2 Expression divergence between the proto-Y and proto-X chromosomes.

Elevated heterozygosity on the proto-sex chromosome in genotypic (III^M/III) males relative to sex-reversed (III/III) males suggests that the DNA sequences of the house fly

III^M proto-Y chromosomes have been differentiated from the standard third (proto-X) chromosome even if the proto-Y and proto-X chromosomes are almost similar morphologically and in terms of gene content (Boyes *et al.* 1964; Hediger *et al.* 1998; Meisel *et al.* 2017). I hypothesized that these X-Y sequence differences contribute to expression differentiation between the proto-Y and the proto-X chromosomes. To test this hypothesis, I used IDP-ASE to quantify differential gene expression between the proto-X and proto-Y chromosome copies (Deonovic *et al.* 2016) using the RNA-seq data from the *Md-tra* RNAi experiment described in Chapter 2.

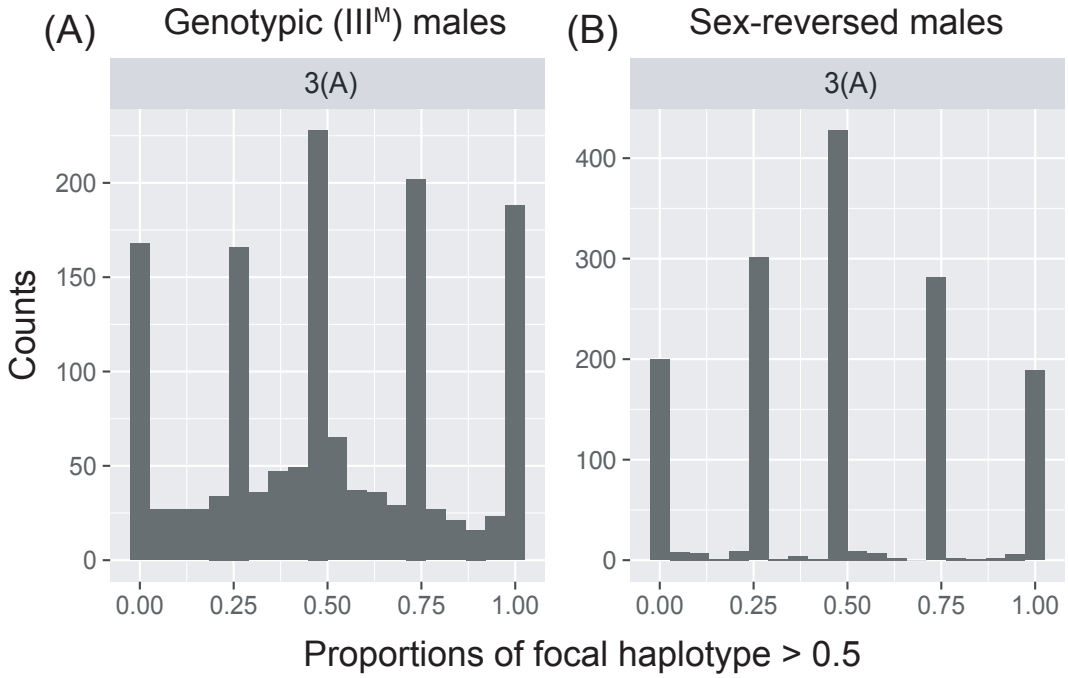


Figure 4.3. Histograms of ASE for third chromosome genes are shown in the genotypic (III^M/III) males and the sex-reversed (III/III) males. If a gene is expressed equally between two alleles, the proportion of focal haplotype is 0.5; otherwise, the proportion is greater or less than 0.5.

I used IDP-ASE to quantify ASE of genes in genotypic (III^M/III) and sex-reversed (III/III) males on the third (proto-sex) chromosome (Figure 4.3) and the other

chromosomes (Appendix Figure 8). I measured ASE as the proportion of iterations in an MCMC simulation with expression of a focal haplotype estimated as > 0.5. This give a measure of ASE ranging from 0 (extreme ASE in favor of one allele) to 1 (extreme ASE in favor of another allele), with 0.5 indicating equal expression of both alleles. I divided my measures of ASE into five sections: 1) extreme ASE with a value between 0 and 0.125, 2) moderate ASE with a value between 0.125 and 0.375, 3) no ASE with a value between 0.375 and 0.625, 4) moderate ASE with a value between 0.625 and 0.875, and 5) extreme ASE with a value between 0.875 and 1. Following these criteria, I identified 109 genes with extreme ASE in the genotypic (III^M/III) males and no ASE in the sex-reversed (III/III) males on the third chromosome (Table 4.1). These genes could contain proto-Y alleles in *cis* regulator sequences that are differentiated from their homologous proto-X alleles, affecting gene expression levels. There is a significant excess of genes with ASE on the third chromosome in genotypic (III^M/III) males and no ASE in sex-reversed (III/III) males compared to all of the other chromosomes (Table 4.1; Fisher's exact test, $P = 0.00715$). To confirm that this excess of third chromosome genes is caused by excess ASE in genotypic (III^M/III) males and not an artifact of my analysis approach, I next identified genes with no ASE in genotypic males and ASE in sex-reversed males (i.e., the opposite of what I did first). The third chromosome has a marginal deficiency of genes with no ASE in genotypic males and ASE in sex-reversed males (Table 4.1; Fisher's exact test, $P = 0.05268$). Therefore, there is indeed an excess of expression divergence between the III^M proto-Y and its homologous proto-X.

Table 4.1. Counts of genes with ASE on each chromosome in genotypic (G) males and sex-reversed (SR) males

		ASE in G males and non-ASE in SR males			non-ASE in G males and ASE in SR males		
Chromosome (Muller element)	# genes	# genes	Odds ratio compared with 3(A)	95% CI	# genes	Odds ratio compared with 3(A)	95% CI
3(A)	1417	109			76		
genome except 3(A)	4184	**236**	**1.394**	**1.091 - 1.773**	286	0.7725	0.587 - 1.006
1(B)	807	51	1.235	0.867 - 1.780	55	0.7750	0.534 - 1.131
2(E)	1232	**64**	**1.521**	**1.095 - 2.126**	**101**	**0.6348**	**0.460 - 0.873**
4(D)	975	62	1.227	0.879 - 1.724	69	0.7443	0.524 - 1.058
5(C)	1144	**58**	**1.560**	**1.112 - 2.207**	61	1.0062	0.702 - 1.448
X(F)	26	1	2.082	0.334 - 86.289	0	Inf	0.363 - Inf

genes in the second column indicates the total number of genes on each chromosome, the third column is the number of genes with ASE in genotypic males and no ASE in sex-reversed males, and the sixth column is the number of genes with no ASE in genotypic males and ASE in sex-reversed males.

The analysis described above demonstrates that some of the III^M proto-Y alleles are expressed differentially from their homologous proto-X copies, but it does not reveal global expression differences between the III^M proto-Y and proto-X chromosomes. If the III^M proto-Y chromosome is differentiated in overall gene expression from its homologous X chromosome, then I expect a higher fraction of genes with ASE on the third chromosome in the genotypic (III^M/III) males than in the sex-reversed (III/III) males. I therefore tested if there is an excess of ASE in genes on the third chromosome in the genotypic males relative to the sex-reversed males. Surprisingly, I did not find an excess of genes with ASE on the third chromosome in genotypic males (Table 4.2). This result suggests that the III^M proto-Y chromosome is not fully differentiated at the

expression level from the standard third (proto-X) chromosome, providing an additional evidence that the house fly III^M chromosome is a very young Y chromosome that is recently derived.

Table 4.2. Chromosomal distribution of allele-specific expression (ASE) and no allele-specific expression (non-ASE) in genotypic (G) males and sex-reversed (SR) males

Chromosome (Muller element)	# genes with ASE in G males	# genes with non-ASE in G males	# genes with ASE in SR males	# genes with non-ASE in SR males	Odds ratio	95% CI
3(A)	445	415	412	447	1.163	0.958 - 1.412
1(B)	315	189	328	206	1.047	0.808 - 1.356
2(E)	450	339	503	327	0.863	0.704 - 1.057
4(D)	391	238	397	223	0.923	0.728 - 1.169
5(C)	471	258	456	277	1.109	0.891 - 1.380
X(F)	4	4	7	8	1.136	0.147 - 8.813

Male-beneficial/female-detrimental sexually antagonistic alleles are expected to accumulate on a Y chromosome (Rice 1984). These sexually antagonistic polymorphisms can cause X-Y divergence and promote recombination inhibition in the vicinity of the sex determination gene (van Doorn and Kirkpatrick 2007, 2010). The sexually antagonistic alleles could also be differentiated in expression from the homologous alleles on the X chromosome. I therefore investigated if genes on the third chromosome with ASE in the genotypic (III^M/III) males and no ASE in the sex-reversed (III/III) males are differentiated in expression between genotypic and sex-reversed males.

In Chapter 2, I showed that genes with discordant sex-biased expression between the genotypic and sex-reversed males are over-represented on the third chromosome, suggesting divergence of *cis*-regulatory alleles between the III^M (proto-Y) and standard

third (proto-X) chromosomes. I found two genes (LOC101899975; XM_011293910.2 XM_011293909.2 and LOC101894537; XM_005179940.3) with discordant sex-biased gene expression out of the 109 genes on the third chromosome with ASE in the genotypic (III^M/III) males and no ASE in the sex-reversed (III/III) males. These genes are homologous to *dynein assembly factor 5, axonemal* (human gene *DNAAF5* and *Drosophila* gene *HEATR2*) and *mitochondrial ornithine transporter 1* (human gene *SLC25A15*), respectively. The two genes are expected to encode proteins that function in flagellated sperm motility and mitochondrial L-ornithine transmembrane transport in the urea cycle. These two genes were male-biased in the genotypic males and female-biased in the sex-reversed males, suggesting that expression differences between the III^M proto-Y and the standard third chromosome cause the male-biased expression of the two genes in the genotypic males, not in the sex-reversed males.

With haplotypes estimated by the IDP-ASE, I could identify three and one diagnostic variant sites for ASE in the two genes (Figure 4.4). The genotypic (III^M/III) males are heterozygous and the sex-reversed (III/III) males are homozygous at all diagnostic sites. I inferred the allele on the standard third chromosome as the one in common between genotypic and sex-reversed males, and the III^M allele as the one unique to genotypic (III^M/III) males at each diagnostic variant site. The two genes are expressed higher in III^M genotypic males than in sex-reversed males (Figure 4.4). In the III^M genotypic males, the III^M (Y-linked) alleles are expressed higher than the X-linked alleles, indicating that the Y-linked alleles are associated with the up-regulation of the two genes in III^M genotypic males relative to sex-reversed males (Figure 4.4). This

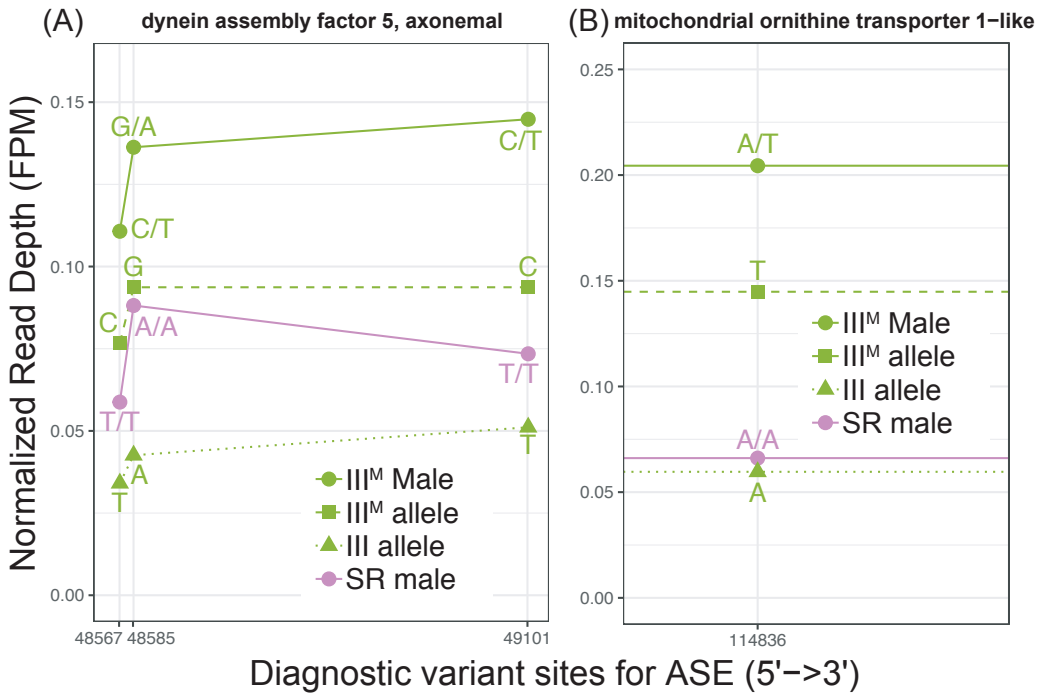


Figure 4.4. Diagnostic variant sites for ASE of two genes, (A) *dynein assembly factor 5, axonemal* and (B) *mitochondrial ornithine transporter 1*, based on haplotypes estimated in IDP-ASE. The *dynein assembly factor 5, axonemal* gene has three diagnostic variant sties and the *mitochondrial ornithine transporter 1-like* gene has one diagnostic variant site.

suggests that these two genes on the proto-sex chromosome are diverged at sequence and expression levels between the proto-Y and proto-X copies.

4.4 Discussion

X and Y chromosomes become differentiated from each other as they evolve on different trajectories once recombination between the two chromosomes is suppressed (Bull 1983). The X-Y differentiation creates notable hallmarks on the Y chromosome (Figure 4.1), such as degeneration, extensive heterochromatinization, loss of the majority of genes, and accumulation of repetitive sequences (Skaletsky *et al.* 2003; Bachtrog 2013). These characteristics of old Y chromosomes have been intensively studied. On the other hand, very young Y chromosomes are poorly understood because they are hard to study—

young Y chromosomes have almost same gene content as their homologous X chromosomes, and most of the genes are not differentiated sufficiently to be detected as Y-linked. In this study, I investigated DNA sequence and gene expression differentiation of the very young III^M proto-Y chromosome and its homologous proto-X in house fly to address how Y chromosomes evolve at the very earliest stages shortly after the Y is formed.

I first determined the extent to which the sequence of the III^M proto-Y chromosome has differentiated from the standard third chromosome by comparing transcript sequences from genotypic (III^M/III) and sex-reversed (III/III) males. I found that the genotypic males have elevated heterozygosity on the third chromosome relative to the sex-reversed males (Figure 4.2). This result suggests that the III^M proto-Y chromosome has accumulated mutations as a consequence of X-Y sequence divergence. Next, I examined whether the III^M proto-Y genes are differentially expressed from their homologous proto-X copies by quantifying ASE. To do this, I compared ASE between genotypic and sex-reversed males. Genotypic males do not have an excess of genes on the third (proto-sex) chromosome with ASE compared to sex-reversed males (Table 4.2). However, I found an enrichment of genes that have ASE in the genotypic males and no ASE in the sex-reversed males on the third (proto-sex) chromosome (Table 4.1). This implies that the III^M proto-Y chromosome is not globally differentiated at the expression level from the standard third chromosome, but some genes on the third chromosome have differentiated in expression.

In Chapter 2, I showed that there is a high fraction of discordant sex-biased expression on the third chromosome between the genotypic (III^M/III) and sex-reversed (III/III) males, implying divergence of *cis*-regulatory alleles on the third (proto-sex) chromosome. However, I did not previously determine which alleles (proto-Y or proto-X copies) are responsible for the expression differences between genotypic and sex-reversed males. In this chapter, I identified two genes on the third chromosome with male-biased expression in genotypic males, female-biased expression in sex-reversed males, ASE in genotypic males, and no ASE in sex-reversed males (Figure 4.4). These two genes provide additional evidence that the proto-Y alleles have been differentiated in gene sequence and expression from the proto-X copies. Their pattern of expression suggests that this expression divergence results in up-regulation of the genes in flies carrying the III^M proto-Y. This could constitute an early stage of X-Y differentiation before full X-Y differentiation occurs (Bachtrog 2013).

Most of the genes on the third chromosome with discordant sex-biased gene expression do not have evidence of ASE in genotypic males and no ASE in sex-reversed males. This is likely because of a limitation of the IDP-ASE tool and expression data that I used to quantify ASE: I am only able to identify ASE within genes with bi-allelic sites in exonic regions. This means I cannot detect ASE if the genes do not have polymorphic sites segregating within exons. Thus, this study could not investigate the effects of sequence divergence in *cis*-regulatory regions that might control gene expression on the III^M chromosome for many genes.

If alleles have sexually antagonistic effects (e.g., beneficial to males and deleterious to females), then selection on these alleles could drive sex chromosome turnover if they are tightly linked to a new sex-determining gene (Orzack *et al.* 1980; van Doorn and Kirkpatrick 2007, 2010; Roberts *et al.* 2009). The genetic linkage between sexually antagonistic alleles and the new sex-determining locus could favor restricted or suppressed recombination between the proto-Y and proto-X chromosomes in that linked region, triggering additional X-Y differentiations (Bachtrog 2013). I identified two genes (homologs of *dynein assembly factor 5, axonemal* and *mitochondrial ornithine transporter 1*) with discordant sex-biased expression in genotypic males (III^M/III) and sex reversed males (III/III) as well as differential expression of the III^M and standard third chromosome alleles. The expression of these two genes could be under sexually antagonistic selection because they function in flagellated sperm motility and mitochondrial L-ornithine transmembrane transport, respectively. These genes have male-biased expression in abdominal tissue, consistent with a function in spermatogenesis. Axonemal dynein is important for male fertility by affecting sperm motility in *Drosophila* (Kurek *et al.* 1998; Carvalho *et al.* 2000), suggesting that it may be beneficial to male fitness to have higher expression of *dynein assembly factor 5, axonemal*. In addition, investment in expressing the gene in females could be costly. Furthermore, many nuclear-encoded mitochondrial genes are expressed in testis or male-specifically in *Drosophila* species. Some of the testis-expressed nuclear-encoded mitochondrial *Drosophila* genes were duplicated from broadly expressed ancestral paralogs, and the testis-expressed copies have been hypothesized to resolve intra-locus

sexually antagonistic conflicts over male germline functions and function in other tissues (Bai *et al.* 2007; Gallach *et al.* 2010). One previous study also found that mitochondrial-nuclear genes show sexually antagonistic effects in *Drosophila* species (Rand *et al.* 2001). This is consistent with the gene I identified in this study because alleles of this mitochondrial-nuclear gene could have been under sexually antagonistic selection evidenced by discordant expression between the genotypic (III^M/III) and sex-reversed (III/III) males, which might have been resolved by sex-specific expression as a male-biased gene.

4.4.1 Conclusions

I investigated gene sequence and expression differences between the III^M proto-Y and homologous proto-X chromosomes to determine how a very young Y chromosome has been differentiated from its homologous X chromosome shorty after it is formed. To address this, I used genotypic (III^M/III) and sex-reversed (III/III) males because they are phenotypically (in terms of gene expression and morphology) almost the same but differ in the proto-sex chromosomes they carry. I found the increased heterozygosity in genotypic males relative to sex-reversed males, consistent with divergence between the proto-Y and proto-X. I also found an excess of genes with ASE in the genotypic males and no ASE in the sex-reversed males on the third (proto-sex) chromosome. Two of the genes (ASE in genotypic males and with no ASE in sex-reversed males on the proto-sex chromosome) showed discordant sex-biased expression between genotypic and sex-reversed males, suggesting that the expressional differentiation could contribute to X-Y

differentiation. Therefore, the house fly III^M proto-Y chromosome is differentiated at the DNA sequence and expression levels from its homologous proto-X chromosome.

Chapter 5

Attempts to establish a stable *Md-tra*^{D} line

5.1 Background

A male-determining gene (*Mdmd*) regulates splicing of the house fly ortholog of *transformer* (*Md-tra*), located on autosome IV, by preventing males from producing a functional female-determining isoform of *Md-tra* (Hediger *et al.* 2010; Sharma *et al.* 2017). A dominant female-determining allele of *Md-tra* (*Md-tra*^{D}) that is not affected by *Mdmd* negative regulation also segregates in natural populations, allowing females to carry *Mdmd* in their genome (Mcdonald *et al.* 1978; Kozielska *et al.* 2008; Hediger *et al.* 2010). *Md-tra*^{D} may have invaded populations because it balances sex ratios (Meisel *et al.* 2016). For example, a population that has males with multiple proto-Y chromosomes carrying *Mdmd* could have an excess of males and be out of its equilibrium sex ratio. This could favor a mutation that generates females independent of *Mdmd*. The positive correlation between *Md-tra*^{D} allele frequencies and males carrying multiple proto-Y chromosomes supports the hypothesis that *Md-tra*^{D} invaded populations to balance the sex-ratio (Meisel *et al.* 2016). Alternatively, *Md-tra*^{D} could allow for the increase in frequency of proto-Y chromosomes that have beneficial fitness effects. Testing these hypotheses requires doing experiments with house flies carrying different *Md-tra* alleles and proto-Y chromosomes. The final chapter of my thesis reports my attempt to establish a stable *Md-tra*^{D} line of house fly in which all females carry one copy of *Md-tra*^{D} and two proto-Y chromosomes, and all males carry two proto-Y chromosomes. Constructing a stable line would allow us to manipulate the *Md-tra*^{D} and proto-Y copy number to test hypotheses about what affects the frequencies of the sex determining factors and proto-sex chromosomes in natural populations of house fly.

5.2 Results and Discussion

The house fly genome consists of five major chromosomes (numbered I-V) and a small, heterochromatic chromosome (names X or Y, depending on whether it carries *Mdmd*). To construct a stable line in which all females carry *Md-tra*^{D}, I used a crossing scheme that took advantage of a house fly strain (aabys) with a different recessive phenotypic marker on each of the five major chromosomes (I-V). There are no markers for the X chromosome. The markers on the third and fourth chromosomes are important for my crossing scheme because I am trying to create a strain homozygous for the III^{M} proto-Y chromosome, and the *Md-tra* gene is located on chromosome IV. The third chromosome marker is *bwb*, which is homologous to *Drosophila yellow* (Heinze *et al.* 2017). Flies homozygous for the *bwb* mutation have brown bodies, whereas wild-type flies have black bodies. The fourth chromosome marker is *ye*, with homozygous mutants having yellow eyes as opposed to the wild-type red eyes.

My goal was to create a strain in which males carry two copies of the III^{M} chromosome, and females have the same genotype with the addition of the *Md-tra*^{D} allele (Figure 5.1). I collected females from a wild-derived colony in which some, but not all, of the females carry the *Md-tra*^{D} allele. I mated single females from this colony with multiple males from the aabys strain with recessive markers on the five major chromosomes. My objective was to use a backcrossing approach to move the fourth chromosome carrying the *Md-tra*^{D} allele from the females onto the aabys background (Figure 5.1A). After collecting progeny from these single female crosses, I PCR

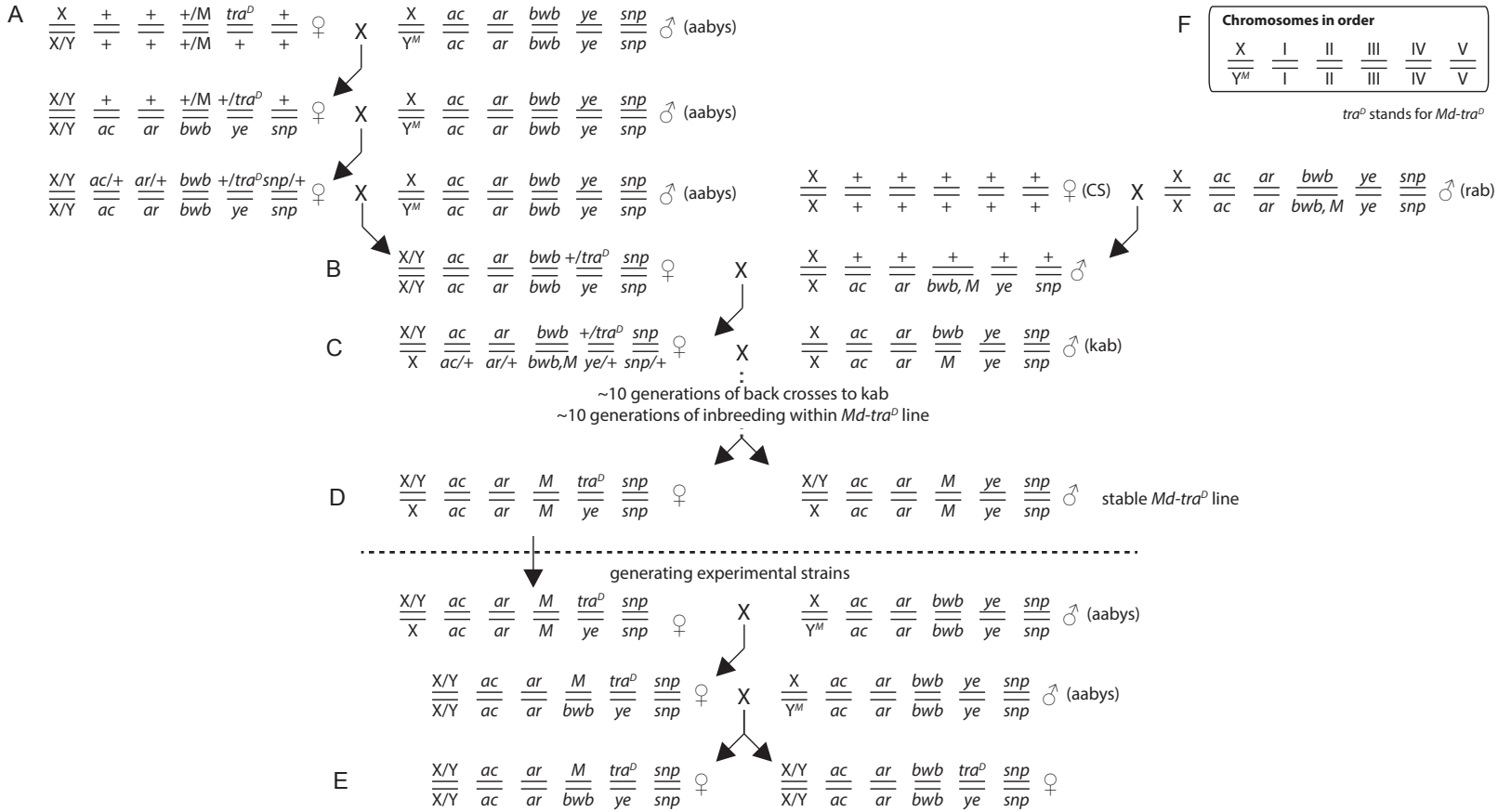


Figure 5.1. Crossing scheme. (A) Crossing between wild *Md-tra*^{D} females and aabys strain. (B) Crossing between *Md-tra*^{D} females on aabys background and males heterozygous for CS and rab, carrying a rab III^M chromosome. (C) Crossing between *Md-tra*^{D} females with a rab III^M chromosome and males of kab strain. (D) Expected genotypes of a stable *Md-tra*^{D} line females and males. (E) Expected experimental strain genotypes of two females; one female has one copy of *Mdmd*, and the other with no *Mdmd* on the single genome. (F) House fly chromosome composition; there is one pair of sex chromosomes (X and Y) and five autosomes. Each pair of parallel line represents homologous chromosomes.

genotyped the mothers to identify those that carried the *Md-tra*^{D} allele (Figure 5.1A, Figure 5.2A), and then I only kept the progeny from mothers carrying the *Md-tra*^{D} allele.

I then backcrossed the female progeny of *Md-tra*^{D} mothers to aabys males. I selected female offspring from this cross with brown bodies (because they are unlikely to have the III^{M} chromosome) and red eyes (because they are likely to have the *Md-tra*^{D} allele). These selected females were then mated to males that were heterozygous for phenotypic markers on all autosomes and have *Mdmd* linked to the brown body recessive allele on the third chromosome (Figure 5.1B). The males heterozygous for these phenotypic markers were created by crossing CS strain females and rab strain males. From the crosses between the females potentially carrying *Md-tra*^{D} and the heterozygous males, I selected female progeny with brown bodies and red eyes because they are likely to carry III^{M} and *Md-tra*^{D}, respectively (Figure 5.1C). Single female offspring were mated with multiple males of the kab strain (Figure 5.1C). After collecting progeny from this cross, I PCR genotyped the mothers for *Md-tra*^{D}, and I kept the offspring from the crosses with *Md-tra*^{D}-bearing females (Figure 5.2B). At this stage, I obtained two lines (line 3 and 8) from two *Md-tra*^{D}-bearing females (indicated by arrows in Figure 5.2B). From these lines, I collected female progeny with black bodies because they are likely to carry a III^{M} chromosome with a wild-type copy of the *bwb* gene inherited from the kab male. All black-bodied female progeny collected from the two lines (five and seven flies from line 3 and 8, respectively) were also PCR genotyped for the *Md-tra*^{D} allele after mating with kab males and producing progeny, and all but one was confirmed to carry *Md-tra*^{D} on their genome (Figure 5.2C). The DNA concentration from one female in line

3 was too low to use in PCR, causing no band in the result (Figure 5.2C). I expect that the female with no PCR amplification carries *Md-tra*^{D} allele on the genome because this fly has black body phenotype as a phenotypic marker of carrying *Mdmd*.

I took only female progeny from lines 3 and 8 and backcrossed them with kab males for ten generations to introgress the III^{M} chromosome into the genotype. However, I could not obtain *Md-tra*^{D} females with two copies of III^{M}. If females carry one copy of *Md-tra*^{D} and two copies of III^{M}, males in the same line are expected to also have two copies III^{M}. To evaluate whether males have two copies of III^{M}, I mated multiple single males with aabys females. Crosses between aabys females and males with two copies of III^{M} are expected to produce all male progeny because the male parent passes along III^{M} to all progeny. However, the crosses I performed resulted in a mixture of females and males, indicating that males did not all carry two III^{M} chromosomes. This suggests that the females in the same line also did not have two III^{M} chromosomes.

In a final effort to create a stable line with females all carrying *Md-tra*^{D} and two copies of III^{M}, I next tried inbreeding black-bodied and red-eyed females and males within lines 3 and 8 for another ten generations because black body and red eye are phenotypic markers of chromosomes III and IV carrying *Mdmd* and *Md-tra*^{D}, respectively. However, ten-generations of inbreeding could not create a stable *Md-tra*^{D} line carrying all females and males with two copies of III^{M} (evaluated using crosses with aabys females as described above).

There are at least two possible reasons why I could not construct a stable *Md-tra*^{D} line. First, recessive deleterious alleles could be accumulated on the III^{M} chromosome.

Deleterious alleles are expected to accumulate on a Y chromosome, such as III^M, because of the combined effects of Muller's ratchet and hitchhiking (Rice 1987a; B. Charlesworth and Charlesworth 2000; Bachtrog 2013). Therefore, two copies of the III^M chromosomes in a single genome might produce lethal or deleterious phenotypic effects, resulting in males that are not viable or have reduced fitness relative to individuals with one copy. Second, the red eye allele, used as a phenotypic marker in the cross because it is genetically linked to *Md-tra*^{D} on the fourth chromosome, frequently recombined away from *Md-tra*^{D}, making it difficult to keep track of the *Md-tra*^{D} allele based on the red eye phenotype. To check whether recombination occurs between the red eye allele and *Md-tra*^{D} allele, I collected red-eyed and yellow-eyed females and PCR genotyped both. I confirmed that both red-eyed and yellow-eyed females carried the *Md-tra*^{D} allele (Figure

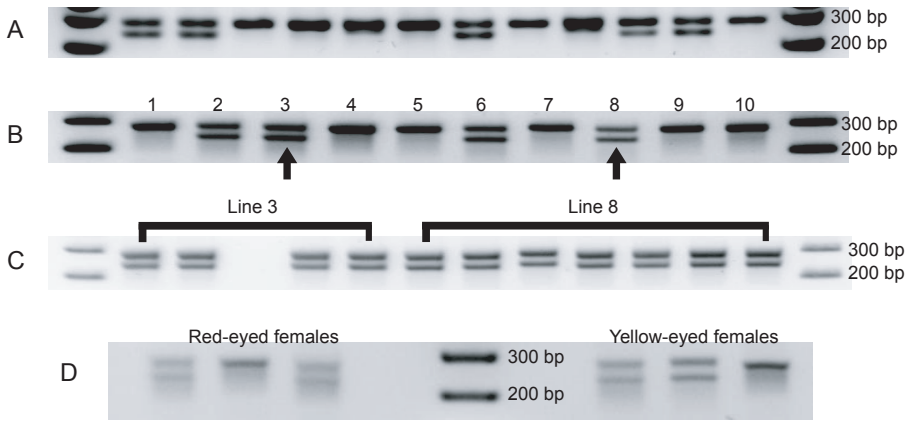


Figure 5.2. Female are genotyped for their *Md-tra* alleles using PCR. The PCR product from the "normal" *Md-tra* allele is slightly larger than the product from the *Md-tra*^{D} allele. Females carrying *Md-tra*^{D} (which must be heterozygous) have two bands, whereas females with two copies of the normal allele have only one band. (A) PCR products used to genotype females used in the first cross from Figure 5.1A are shown. Only progeny from females with two bands were retained. (B) PCR products used to genotype females used in the cross from Figure 5.1C are shown. Only progeny from females with two bands were retained. (C) PCR products used to genotype five and seven females from line 3 and line 8 are shown. (D) PCR products used to genotype red-eyed and yellow-eyed females are shown.

5.2D), indicating that the recombination between red eye and *Md-tra*^{D} allele has occurred.

In conclusion, I was unable to establish a stable *Md-tra*^{D} line. If a stable *Md-tra*^{D} line is established (Figure 5.1D), future work could compare phenotypes of *Md-tra*^{D} females with 0-2 copies of III^{M} (Figure 5.1E). This would allow investigation of the phenotypic effects of the III^{M} proto-Y chromosome on female phenotypes and fitness.

5.3 Materials and Methods

5.3.1 House fly strains

The flies used for the source of the *Md-tra*^{D} allele were collected in College Station, TX. I PCR genotyped females from this strain to confirm that some carry *Md-tra*^{D}. The genome reference strain (aabys) was used in the crossing scheme because it carries visible recessive phenotypic markers *ali-curve* (*ac*), *aristapedia* (*ar*), *brown body* (*bwb*), *yellow eyes* (*ye*), and *snip wings* (*snp*) on autosomes I, II, III, IV, and V, respectively (Figure 5.1). The rab strain was created by backcrossing the III^{M} chromosome from the rspin strain collected in New York (Shono and Scott 2003) onto the aabys background. The kab strain was created by backcrossing the III^{M} chromosome from the KS8S3 strain collected in Florida (Kaufman *et al.* 2010) onto the aabys background. Lastly, the Cornell Susceptible (CS) strain is an inbred III^{M} strain produced by mixing strains collected from throughout the United States (Scott *et al.* 1996).

5.3.2 Genotyping for *Md-tra*

To genotype females for the *Md-tra*^{D} allele, DNA was extracted from individual female house flies. Individual females were homogenized in 2 mL plastic tubes containing 0.1

mL of buffer (0.1 M Tris-HCL pH 7.5, 0.1M EDTA, 0.1M NaCl, and 0.5% sodium dodecyl sulfate), ground by a hand-held motorized pestle. Samples were next incubated on 65 °C for 30 min followed by addition of 0.18 mL of LiCl (4.3 M) plus potassium acetate (1.4 M), and then ice-incubated for 10 min after mixing well. The reactions were centrifuged at 15000 × g for 15 min, and then the supernatant was transferred to new 2 mL tubes. After adding 0.12 mL isopropanol to the new tubes with supernatant, mixing, and 15000 × g for 15 min, the DNA was recovered as a pellet. The pellet was washed in 0.5 mL of 70% EtOH by vortex, centrifugated at 15000 × g for 3 min, the supernatant was removed, and the pellet was air dried. The pellet was resuspended in 10 μL of nuclease-free water (Promega, Madison, WI).

I genotyped individual females using a diagnostic 38 bp deletion present only in the *Md-tra*^{D} allele (Hediger *et al.* 2010; Scott *et al.* 2014). I performed PCR with a primer pair that amplifies a region of exon 3 of the *Md-tra* gene containing the 38 bp deletion at *Md-tra*^{D} allele (forward primer: GCAACGCAAGACGTATTAACCAG; reverse primer: CCTATTGGTTTTGGCTGTCC). PCR products were run on a 2% agarose gel along with Invitrogen 1 Kb Plus Ladder (Life Technologies, Grand Island, NY), stained with GelRed (Biotium, Hayward, CA), and visualized on a NucleoTech ultraviolet gel imager. Individual females with a single band at 277 bp were counted as homozygous for the *Md-tra* wild type allele, and females with bands at both 277 bp and 239 bp were counted as *Md-tra*/*Md-tra*^{D} heterozygotes.

References

Anders, S., P.T. Pyl, and W. Huber. 2015. HTSeq---a Python Framework to Work with High-Throughput Sequencing Data. *Bioinformatics* **31** (2): 166–69.

Bachtrog, D. 2013. Y-Chromosome Evolution: Emerging Insights into Processes of Y-Chromosome Degeneration. *Nature Reviews Genetics* **14** (2): 113.

Bachtrog, D. and B. Charlesworth. 2002. Reduced Adaptation of a Non-Recombining Neo-Y Chromosome. *Nature*. https://doi.org/10.1038/416323a.

Bachtrog, D., J.E. Mank, C.L. Peichel, M. Kirkpatrick, S.P. Otto, T.-L. Ashman, M.W. Hahn, *et al.* 2014. Sex Determination: Why so Many Ways of Doing It? *PLoS Biology* **12** (7): e1001899.

Bai, Y., C. Casola, C. Feschotte, and E. Betrán. 2007. Comparative Genomics Reveals a Constant Rate of Origination and Convergent Acquisition of Functional Retrogenes in Drosophila. *Genome Biology* **8** (1): R11.

Bansal, V., A.L. Halpern, N. Axelrod, and V. Bafna. 2008. An MCMC Algorithm for Haplotype Assembly from Whole-Genome Sequence Data. *Genome Research* **18** (8): 1336–46.

Bates, D., M. Maechler, B. Bolker, S. Walker, R.H.B. Christensen, H. Singman, B. Dai, F. Scheipl, G. Grothendieck, and P. Green. 2017. Linear Mixed-Effects Models Using "Eigen" and S4. *R Package Version 1.7*.

Benjamini, Y. and Y. Hochberg. 1995. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, 289–300.

Beukeboom, L.W. and N. Perrin. 2014. *The Evolution of Sex Determination*. Oxford University Press, USA.

Bopp, D. 2010. About Females and Males: Continuity and Discontinuity in Flies. *Journal of Genetics* **89** (3): 315–23.

Boyes, J.W., M.J. Corey, and H.E. Paterson. 1964. Somatic Chromosomes of Higher Diptera: IX. Karyotypes of Some Muscid Species. *Canadian Journal of Zoology* **42** (6): 1025–36.

Brown, E. and D. Bachtrog. 2017. The Drosophila Y Chromosome Affects Heterochromatin Integrity Genome-Wide. *BioRxiv*, 156000.

Bull, J.J. 1983. *Evolution of Sex Determining Mechanisms.* The Benjamin/Cummings Publishing Company, Inc.

Bull, J.J. and E.L. Charnov. 1977. Changes in the Heterogametic Mechanism of Sex Determination. *Heredity* **39** (1): 1.

Bulmer, M.G. and J.J. Bull. 1982. Models of Polygenic Sex Determination and Sex Ratio Control. *Evolution*, **36** (1), 13–26.

Carvalho, A.B., B.P. Lazzaro, and A.G. Clark. 2000. Y Chromosomal Fertility Factors Kl-2 and Kl-3 of Drosophila Melanogaster Encode Dynein Heavy Chain Polypeptides. *Proceedings of the National Academy of Sciences* **97** (24): 13239–44.

Charlesworth, B. 1996. The Evolution of Chromosomal Sex Determination and Dosage Compensation. *Current Biology* **6** (2): 149–62.

Charlesworth, B. and D. Charlesworth. 2000. The Degeneration of Y Chromosomes. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences* **355** (1403): 1563–72.

Charlesworth, D., B. Charlesworth, and G. Marais. 2005. Steps in the Evolution of Heteromorphic Sex Chromosomes. *Heredity* **95** (2): 118–28.

Clark, A.G. 1987. Natural Selection and Y-Linked Polymorphism. *Genetics* **115** (3): 569–77.

Clark, A.G. 1988. The Evolution of the Y Chromosome with XY Recombination. *Genetics* **119** (3): 711–20.

Connallon, T. 2015. The Geography of Sex-Specific Selection, Local Adaptation, and Sexual Dimorphism. *Evolution*. https://doi.org/10.1111/evo.12737.

Connallon, T. and A.G. Clark. 2010. Gene Duplication, Gene Conversion and the Evolution of the Y Chromosome. *Genetics* **186** (1): 277–86.

Delph, L.F., J. Andicoechea, J.C. Steven, C.R. Herlihy, S. V. Scarpino, and D.L. Bell. 2011. Environment-Dependent Intralocus Sexual Conflict in a Dioecious Plant. *New Phytologist*. https://doi.org/10.1111/j.1469-8137.2011.03811.x.

Deonovic, B., Y. Wang, J. Weirather, X.-J. Wang, and K.F. Au. 2016. IDP-ASE: Haplotyping and Quantifying Allele-Specific Expression at the Gene and Gene Isoform Level by Hybrid Sequencing. *Nucleic Acids Research* **45** (5): e32--e32.

Dey, K.K., C.J. Hsiao, and M. Stephens. 2017. Visualizing the Structure of RNA-Seq

Expression Data Using Grade of Membership Models. *PLoS Genetics* **13** (3): e1006599.

Dobin, A., C.A. Davis, F. Schlesinger, J. Drenkow, C. Zaleski, S. Jha, P. Batut, M. Chaisson, and T.R. Gingeras. 2013. STAR: Ultrafast Universal RNA-Seq Aligner. *Bioinformatics* **29** (1): 15–21.

Doorn, G.S. van. 2014a. Evolutionary Transitions between Sex-Determining Mechanisms: A Review of Theory. *Sexual Development* **8** (1–3): 7–19.

Doorn, G.S. van. 2014b. Patterns and Mechanisms of Evolutionary Transitions between Genetic Sex-Determining Systems. *Cold Spring Harbor Perspectives in Biology* **6** (8): a017681.

Doorn, G.S. van and M. Kirkpatrick. 2007. Turnover of Sex Chromosomes Induced by Sexual Conflict. *Nature* **449** (7164): 909.

Doorn, G.S. van and M. Kirkpatrick. 2010. Transitions between Male and Female Heterogamety Caused by Sex-Antagonistic Selection. *Genetics* **186** (2): 629–45.

Dübendorfer, A. and M. Hediger. 1998. The Female-Determining Gene F of the Housefly, Musca Domestica, Acts Maternally to Regulate Its Own Zygotic Activity. *Genetics* **150** (1): 221–26.

Dübendorfer, A., M. Hediger, G. Burghardt, and D. Bopp. 2003. Musca Domestica, a Window on the Evolution of Sex-Determining Mechanisms in Insects. *International Journal of Developmental Biology* **46** (1): 75–79.

Ellegren, H. and J. Parsch. 2007. The Evolution of Sex-Biased Genes and Sex-Biased Gene Expression. *Nature Reviews Genetics* **8** (9): 689.

Feldmeyer, B., M. Kozielska, B. Kuijper, F.J. Weissing, L.W. Beukeboom, and I. Pen. 2008. Climatic Variation and the Geographical Distribution of Sex-Determining Mechanisms in the Housefly. *Evolutionary Ecology Research* **10** (6): 797–809.

Foster, G.G., M.J. Whitten, C. Konovalov, J.T.A. Arnold, and G. Maffi. 1981. Autosomal Genetic Maps of the Australian Sheep Blowfly, Lucilia Cuprina Dorsalis R.-D.(Diptera: Calliphoridae), and Possible Correlations with the Linkage Maps of Musca Domestica L. and Drosophila Melanogaster (Mg.). *Genetics Research* **37** (1): 55–69.

Francisco, F.O. and B. Lemos. 2014. How Do Y-Chromosomes Modulate Genome-Wide Epigenetic States: Genome Folding, Chromatin Sinks, and Gene Expression. *Journal of Genomics* **2**: 94–103.

Franco, M.G., P.G. Rubini, and M. Vecchi. 1982. Sex-Determinants and Their Distribution in Various Populations of Musca Domestica L. of Western Europe. *Genetics Research* **40** (3): 279–93.

Gallach, M., C. Chandrasekaran, and E. Betran. 2010. Analyses of Nuclearly Encoded Mitochondrial Genes Suggest Gene Duplication as a Mechanism for Resolving Intralocus Sexually Antagonistic Conflict in Drosophila. *Genome Biology and Evolution* **2**: 835–50.

Goldman, T.D. and M.N. Arbeitman. 2007. Genomic and Functional Studies of Drosophila Sex Hierarchy Regulated Gene Expression in Adult Head and Nervous System Tissues. *PLoS Genetics* **3** (11): e216.

Graves, J.A.M. 2008. Weird Animal Genomes and the Evolution of Vertebrate Sex and Sex Chromosomes. *Annual Review of Genetics* **42**: 565–86.

Hamm, R.L., J.-R. Gao, G.G.-H. Lin, and J.G. Scott. 2009. Selective Advantage for IIIM Males over YM Males in Cage Competition, Mating Competition, and Pupal Emergence in Musca Domestica L.(Diptera: Muscidae). *Environmental Entomology* **38** (2): 499–504.

Hamm, R.L., R.P. Meisel, and J.G. Scott. 2015. The Evolving Puzzle of Autosomal versus Y-Linked Male Determination in Musca Domestica. *G3: Genes, Genomes, Genetics* **5** (3): 371–84.

Hamm, R.L., T. Shono, and J.G. Scott. 2005. A Cline in Frequency of Autosomal Males Is Not Associated with Insecticide Resistance in House Fly (Diptera: Muscidae). *Journal of Economic Entomology* **98** (1): 171–76.

Hediger, M., G. Burghardt, C. Siegenthaler, N. Buser, D. Hilfiker-Kleiner, A. Dübendorfer, and D. Bopp. 2004. Sex Determination in Drosophila Melanogaster and Musca Domestica Converges at the Level of the Terminal Regulator Doublesex. *Development Genes and Evolution* **214** (1): 29–42.

Hediger, M., C. Henggeler, N. Meier, R. Perez, G. Saccone, and D. Bopp. 2010. Molecular Characterization of the Key Switch F Provides a Basis for Understanding the Rapid Divergence of the Sex-Determining Pathway in the Housefly. *Genetics* **184** (1): 155–70.

Hediger, M., A.D. Minet, M. Niessen, R. Schmidt, D. Hilfiker-Kleiner, Ş. Çakir, R. Nöthiger, and A. Dübendorfer. 1998. The Male-Determining Activity on the Y Chromosome of the Housefly (Musca Domestica L.) Consists of Separable Elements. *Genetics* **150** (2): 651–61.

Hediger, M., M. Niessen, E.A. Wimmer, A. Dübendorfer, and D. Bopp. 2001. Genetic Transformation of the Housefly Musca Domestica with the Lepidopteran Derived Transposon PiggyBac. *Insect Molecular Biology* **10** (2): 113–19.

Heinze, S.D., T. Kohlbrenner, D. Ippolito, A. Meccariello, A. Burger, C. Mosimann, G. Saccone, and D. Bopp. 2017. CRISPR-Cas9 Targeted Disruption of the Yellow Ortholog in the Housefly Identifies the Brown Body Locus. *Scientific Reports* **7** (1): 4582.

Huang, D.W., B.T. Sherman, and R.A. Lempicki. 2008. Systematic and Integrative Analysis of Large Gene Lists Using DAVID Bioinformatics Resources. *Nature Protocols* **4** (1): 44.

Ingleby, F.C., J. Hunt, and D.J. Hosken. 2010. The Role of Genotype-by-Environment Interactions in Sexual Selection. *Journal of Evolutionary Biology* **23** (10): 2031–45.

Jakšić, A.M. and C. Schlötterer. 2016. The Interplay of Temperature and Genotype on Patterns of Alternative Splicing in Drosophila Melanogaster. *Genetics* **204** (1): 315–25.

Janousek, B. and M. Mrackova. 2010. Sex Chromosomes and Sex Determination Pathway Dynamics in Plant and Animal Models. *Biological Journal of the Linnean Society* **100** (4): 737–52.

Kaufman, P.E., S.C. Nunez, C.J. Geden, and M.E. Scharf. 2010. Selection for Resistance to Imidacloprid in the House Fly (Diptera: Muscidae). *Journal of Economic Entomology* **103** (5): 1937–42.

Kervestin, S. and A. Jacobson. 2012. NMD: A Multifaceted Response to Premature Translational Termination. *Nature Reviews Molecular Cell Biology* **13** (11): 700–12.

Kim, D., B. Langmead, and S.L. Salzberg. 2015. HISAT: A Fast Spliced Aligner with Low Memory Requirements. *Nature Methods* **12** (4): 357–60.

Kozielska, M., B. Feldmeyer, I. Pen, F.J. Weissing, and L.W. Beukeboom. 2008. Are Autosomal Sex-Determining Factors of the Housefly (Musca Domestica) Spreading North? *Genetics Research* **90** (2): 157–65.

Kozielska, M., I. Pen, L.W. Beukeboom, and F.J. Weissing. 2006. Sex Ratio Selection and Multi-Factorial Sex Determination in the Housefly: A Dynamic Model. *Journal of Evolutionary Biology* **19** (3): 879–88.

Kurek, R., A.M. Reugels, K.H. Glaätzer, and H. Bünemann. 1998. The Y Chromosomal Fertility Factor Threads in Drosophila Hydei Harbors a Functional Gene Encoding

an Axonemal Dynein β Heavy Chain Protein. *Genetics* **149** (3): 1363–76.

Lebo, M.S., L.E. Sanders, F. Sun, and M.N. Arbeitman. 2009. Somatic, Germline and Sex Hierarchy Regulated Gene Expression during Drosophila Metamorphosis. *BMC Genomics* **10** (1): 80.

Lemos, B., L.O. Araripe, and D.L. Hartl. 2008. Polymorphic Y Chromosomes Harbor Cryptic Variation with Manifold Functional Consequences. *Science* **319** (5859): 91–93.

Lemos, B., A.T. Branco, and D.L. Hartl. 2010. Epigenetic Effects of Polymorphic Y Chromosomes Modulate Chromatin Components, Immune Response, and Sexual Conflict. *Proceedings of the National Academy of Sciences* **107** (36): 15826–31.

Li, H., B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, and R. Durbin. 2009. The Sequence Alignment/Map Format and SAMtools. *Bioinformatics* **25** (16): 2078–79.

Love, M.I., W. Huber, and S. Anders. 2014. Moderated Estimation of Fold Change and Dispersion for RNA-Seq Data with DESeq2. *Genome Biology* **15** (12): 550–70.

Ma, W.J., P. Veltsos, R. Sermier, D.J. Parker, and N. Perrin. 2018. Evolutionary and Developmental Dynamics of Sex-Biased Gene Expression in Common Frogs with Proto-Y Chromosomes. *Genome Biology* **19** (1): 156–72.

Mcdonald, I. c, P. Evenson, C.A. Nickel, and O.A. Johnson. 1978. House Fly Genetics: Isolation of a Female Determining Factor on Chromosome 4. *Annals of the Entomological Society of America* **71** (5): 692–94.

McKenna, A., M. Hanna, E. Banks, A. Sivachenko, K. Cibulskis, A. Kernytsky, K. Garimella, *et al.* 2010. The Genome Analysis Toolkit: A MapReduce Framework for Analyzing next-Generation DNA Sequencing Data. *Genome Research* **20** (9): 1297–1303.

Meier, N., S.C. Käppeli, M.H. Niessen, J.-C. Billeter, S.F. Goodwin, and D. Bopp. 2013. Genetic Control of Courtship Behavior in the Housefly: Evidence for a Conserved Bifurcation of the Sex-Determining Pathway. *PLoS One* **8** (4): e62476.

Meisel, R.P., T. Davey, J.H. Son, A.C. Gerry, T. Shono, and J.G. Scott. 2016. Is Multifactorial Sex Determination in the House Fly, Musca Domestica L., Stable Over Time? *J Hered* **107** (7): 615–25.

Meisel, R.P., C.A. Gonzales, and H. Luu. 2017. The House Fly Y Chromosome Is Young and Minimally Differentiated from Its Ancient X Chromosome Partner. *Genome*

Research **27** (8): 1417–26.

Meisel, R.P. and J.G. Scott. 2018. Using Genomic Data to Study Insecticide Resistance in the House Fly, Musca Domestica. *Pesticide Biochemistry and Physiology* **151**: 76–81.

Meisel, R.P., J.G. Scott, and A.G. Clark. 2015. Transcriptome Differences between Alternative Sex Determining Genotypes in the House Fly, Musca Domestica. *Genome Biology and Evolution* **7** (7): 2051–61.

Moore, E.C. and R.B. Roberts. 2013. Polygenic Sex Determination. *Current Biology* **23** (12): R510--R512.

Neal, R.M. and others. 2003. Slice Sampling. *The Annals of Statistics* **31** (3): 705–67.

Orzack, S.H., J.J. Sohn, K.D. Kallman, S.A. Levin, and R. Johnston. 1980. Maintenance of the Three Sex Chromosome Polymorphism in the Platyfish, Xiphophorus Maculatus. *Evolution* **34** (4): 663–72.

Parnell, N.F. and J.T. Streelman. 2013. Genetic Interactions Controlling Sex and Color Establish the Potential for Sexual Conflict in Lake Malawi Cichlid Fishes. *Heredity* **110** (3): 239.

Punzalan, D., M. Delcourt, and H.D. Rundle. 2014. Comparing the Intersex Genetic Correlation for Fitness across Novel Environments in the Fruit Fly, Drosophila Serrata. *Heredity* **112** (2): 143.

R CoreTeam. 2013. R: A Language and Environment for Statistical Computing. *R Foundation for statistical computing*, *Vienna*

Rand, D.M., A.G. Clark, and L.M. Kann. 2001. Sexually Antagonistic Cytonuclear Fitness Interactions in Drosophila Melanogaster. *Genetics* **159** (1): 173–87.

Rice, W.R. 1984. Sex Chromosomes and the Evolution of Sexual Dimorphism. *Evolution* **38** (4): 735–42.

Rice, W.R. 1986. On the Instability of Polygenic Sex Determination: The Effect of Sex-Specific Selection. *Evolution* **40** (3): 633–39.

Rice, W.R. 1987a. Genetic Hitchhiking and the Evolution of Reduced Genetic Activity of the Y Sex Chromosome. *Genetics* **116** (1): 161–67.

Rice, W.R. 1987b. The Accumulation of Sexually Antagonistic Genes as a Selective Agent Promoting the Evolution of Reduced Recombination between Primitive Sex

Chromosomes. *Evolution* **41** (4): 911–14.

Rice, W.R. 1992. Sexually Antagonistic Genes: Experimental Evidence. *Science* **256** (5062): 1436–39.

Rice, W.R. 1996. Evolution of the Y Sex Chromosome in Animals. *Bioscience* **46** (5): 331–43.

Roberts, R.B., J.R. Ser, and T.D. Kocher. 2009. Sexual Conflict Resolved by Invasion of a Novel Sex Determiner in Lake Malawi Cichlid Fishes. *Science* **326** (5955): 998–1001.

Robinett, C.C., A.G. Vaughan, J.-M. Knapp, and B.S. Baker. 2010. Sex and the Single Cell. II. There Is a Time and Place for Sex. *PLoS Biology* **8** (5): e1000365.

Salz, H.K. and J.W. Erickson. 2010. Sex Determination in Drosophila: The View from the Top. *Fly* **4** (1): 60–70.

Saunders, P.A., S. Neuenschwander, and N. Perrin. 2018. Sex Chromosome Turnovers and Genetic Drift: A Simulation Study. *Journal of Evolutionary Biology* **31** (9): 1413–19.

Schmidt, R., M. Hediger, R. Nöthiger, and A. Dübendorfer. 1997. The Mutation Masculinizer (Man) Defines a Sex-Determining Gene with Maternal and Zygotic Functions in Musca Domestica L. *Genetics* **145** (1): 173–83.

Scott, J.G., P. Sridhar, and N. Liu. 1996. Adult Specific Expression and Induction of Cytochrome P450tpr in House Flies. *Archives of Insect Biochemistry and Physiology: Published in Collaboration with the Entomological Society of America* **31** (3): 313–23.

Scott, J.G., W.C. Warren, L.W. Beukeboom, D. Bopp, A.G. Clark, S.D. Giers, M. Hediger, *et al.* 2014. Genome of the House Fly, Musca Domestica L., a Global Vector of Diseases with Adaptations to a Septic Environment. *Genome Biology* **15** (10): 466–81.

Ser, J.R., R.B. Roberts, and T.D. Kocher. 2010. Multiple Interacting Loci Control Sex Determination in Lake Malawi Cichlid Fish. *Evolution* **64** (2): 486–501.

Sharma, A., S.D. Heinze, Y. Wu, T. Kohlbrenner, I. Morilla, C. Brunner, E.A. Wimmer, *et al.* 2017. Male Sex in Houseflies Is Determined by Mdmd, a Paralog of the Generic Splice Factor Gene CWC22. *Science* **356** (6338): 642–45.

Shono, T. and J.G. Scott. 2003. Spinosad Resistance in the Housefly, Musca Domestica,

Is Due to a Recessive Factor on Autosome 1. *Pesticide Biochemistry and Physiology* **75** (1–2): 1–7.

Shukla, J.N. and S.R. Palli. 2012. Sex Determination in Beetles: Production of All Male Progeny by Parental RNAi Knockdown of Transformer. *Scientific Reports* **2**: 602–10.

Skaletsky, H., T. Kuroda-Kawaguchi, P.J. Minx, H.S. Cordum, L. Hillier, L.G. Brown, S. Repping, *et al.* 2003. The Male-Specific Region of the Human Y Chromosome Is a Mosaic of Discrete Sequence Classes. *Nature* **423** (6942): 825.

Tomita, T. and Y. Wada. 1989. Multifactorial Sex Determination in Natural Populations of the Housefly (Musca Domestica) in Japan. *The Japanese Journal of Genetics* **64** (5): 373–82.

Verhulst, E.C., L. van de Zande, and L.W. Beukeboom. 2010. Insect Sex Determination: It All Evolves Around Transformer. *Current Opinion in Genetics & Development* **20** (4): 376–83.

Vicoso, B. and D. Bachtrog. 2013. Reversal of an Ancient Sex Chromosome to an Autosome in Drosophila. *Nature* **499** (7458): 332.

Vicoso, B. and D. Bachtrog. 2015. Numerous Transitions of Sex Chromosomes in Diptera. *PLoS Biology* **13** (4): e1002078.

Wang, M., A.T. Branco, and B. Lemos. 2018. The Y Chromosome Modulates Splicing and Sex-Biased Intron Retention Rates in Drosophila. *Genetics* **208** (3): 1057–67.

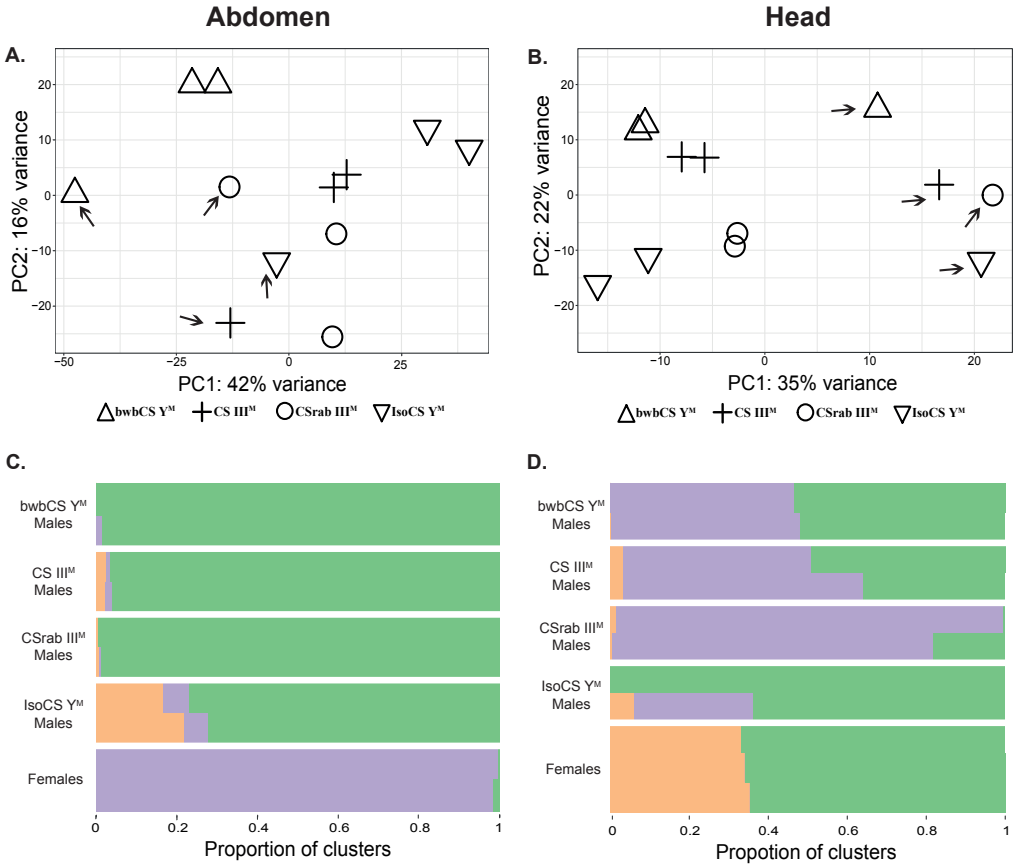
Weller, G.L. and G.G. Foster. 1993. Genetic Maps of the Sheep Blowfly Lucilia Cuprina: Linkage-Group Correlations with Other Dipteran Genera. *Genome* **36** (3): 495–506.

Wilkins, A.S. 1995. Moving up the Hierarchy: A Hypothesis on the Evolution of a Genetic Sex Determination Pathway. *Bioessays* **17** (1): 71–77.

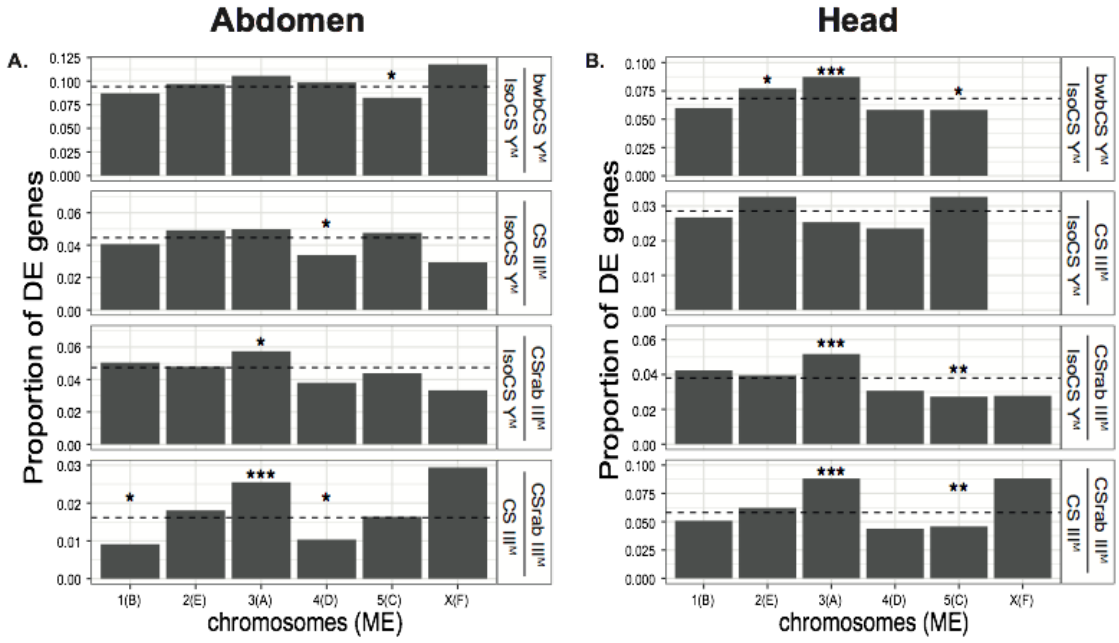
Ye, D., A.A. Zaidi, M. Tomaszkiewicz, K. Anthony, C. Liebowitz, M. DeGiorgio, M.D. Shriver, and K.D. Makova. 2018. High Levels of Copy Number Variation of Ampliconic Genes across Major Human Y Haplogroups. *Genome Biology and Evolution* **10** (5): 1333–50.

Zhou, Q., J. Zhang, D. Bachtrog, N. An, Q. Huang, E.D. Jarvis, M.T.P. Gilbert, and G. Zhang. 2014. Complex Evolutionary Trajectories of Sex Chromosomes across Bird Taxa. *Science* **346** (6215): 1246338.

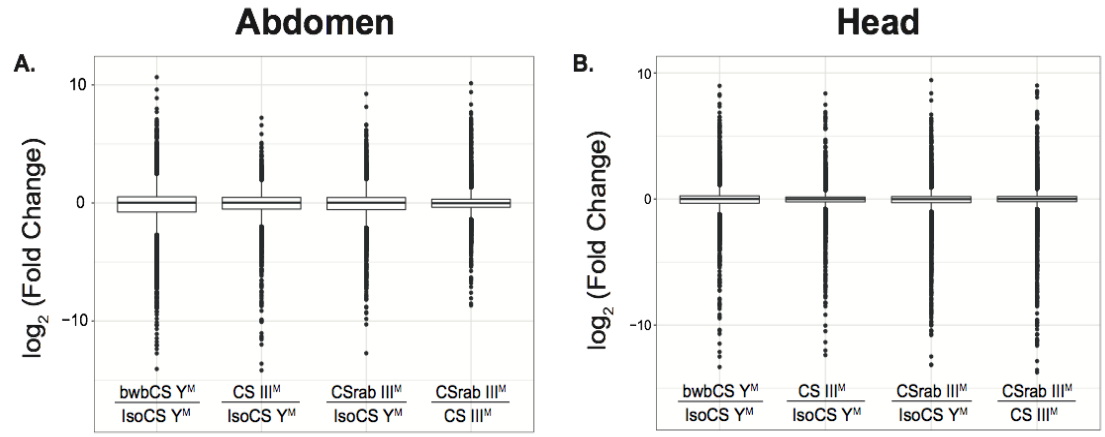
Appendices



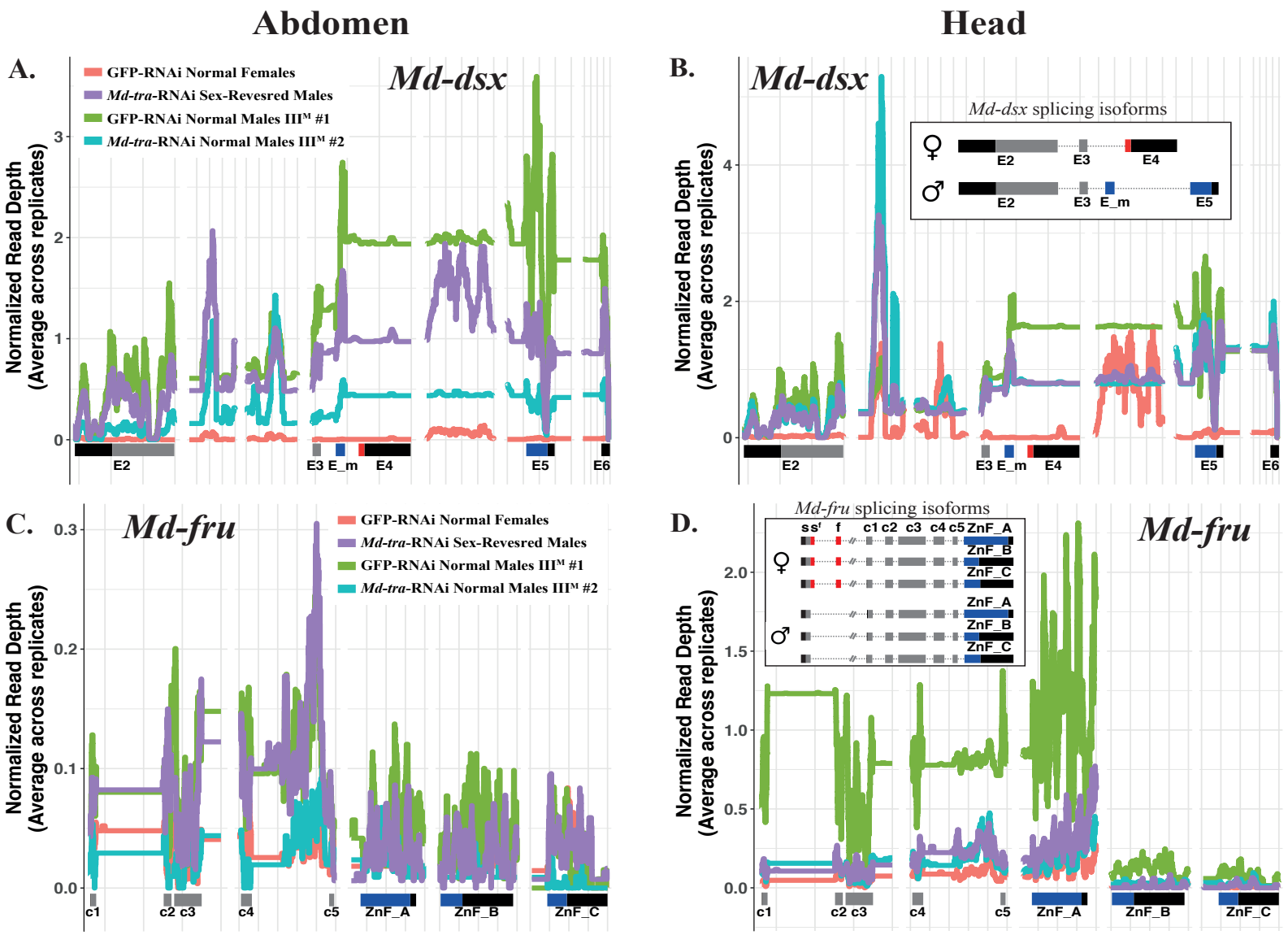
Appendix Figure 1. (A, B) Principal component (PC) analysis of four strains that have different naturally occurring neo-Y chromosomes on a common genetic background in abdomens (A) and heads (B). Arrows point to outlier samples, one for each of the four strains. Female abdomens are excluded from the PC analysis (A) to show the outliers. (C, D) Grade of membership model ($K = 3$) for gene expression patterns of four strains that have different naturally occurring proto-Y chromosomes on a common genetic background in abdomens (C) and heads (D). Each row represents one replicate of a genotype, with the outliers excluded. Each color represents the proportion of each replicate assigned to each of the three clusters.



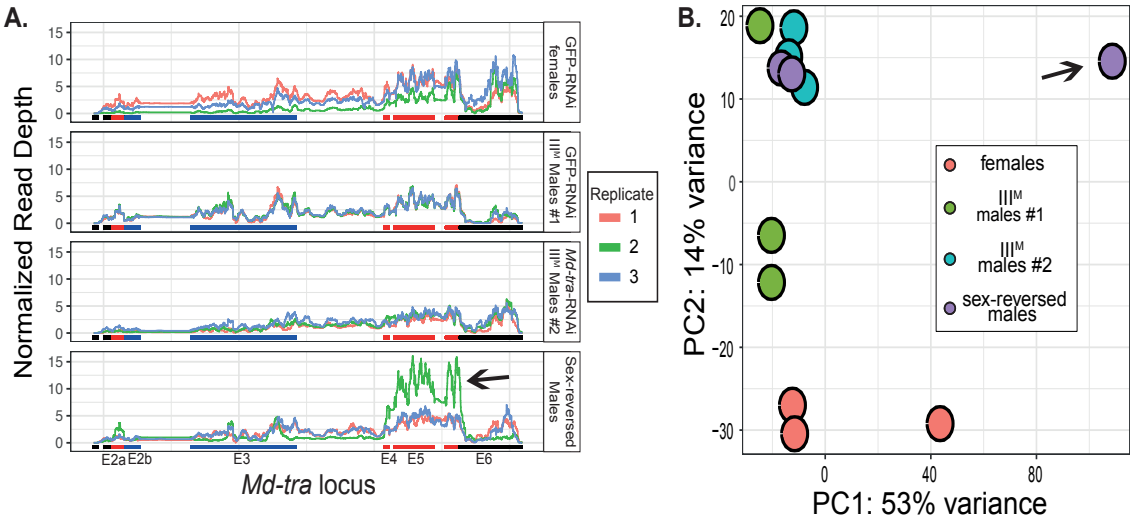
Appendix Figure 2. Bar graphs indicate the proportions of genes on each chromosome (*Drosophila* Muller element in parentheses) that are differentially expressed (DE) between different male genotypes in abdomens (A) and heads (B). Asterisks indicate significant differences based on Fisher's exact test (*P<0.05, **P<0.01, ***P < 0.001).



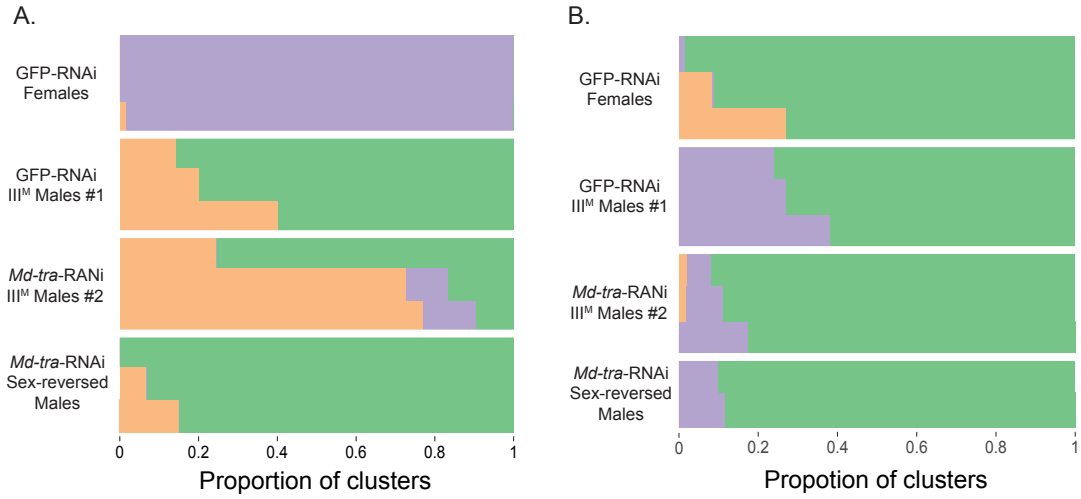
Appendix Figure 3. Boxplots show fold changes of gene expression between males with different *Mdmd*-bearing chromosomes in abdomens (A) and heads (B). Outliers are included as points.


Abdomen
Head
A.
B.
C.
D.
Md-dsx
Md-fru
GFP-RNAi Normal Females
Md-tra-RNAi Sex-Revesred Males
GFP-RNAi Normal Males IIIM #1
Md-tra-RNAi Normal Males IIIM #2
Normalized Read Depth
(Average across replicates)
Md-dsx splicing isoforms
Md-fru splicing isoforms
E2
E3
E_m
E4
E5
E6
c1
c2
c3
c4
c5
ZnF_A
ZnF_B
ZnF_C

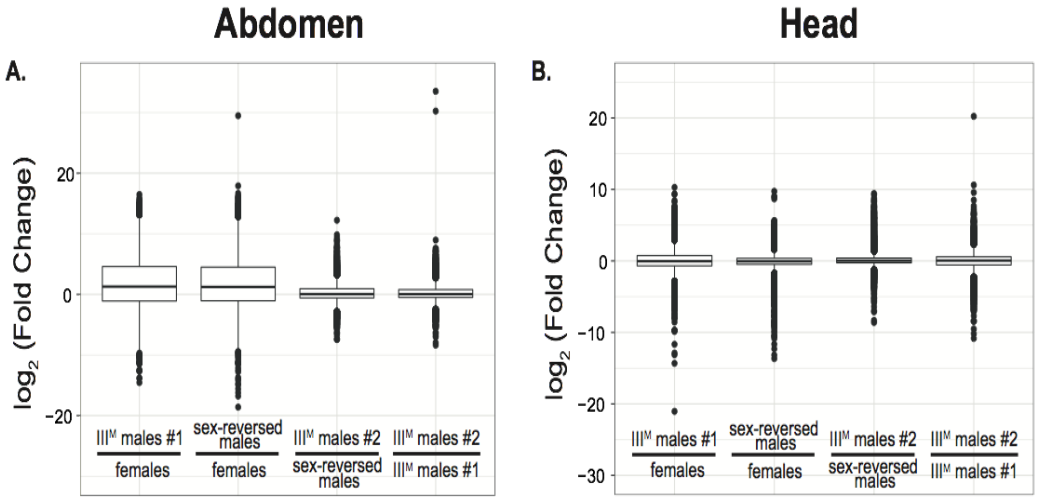
Appendix Figure 4. *Md-tra* regulates the splicing of at least two downstream genes, *Md-dsx* and *Md-fru*, which are both differentially spliced between females and males (Hediger *et al.* 2004, 2010; Meier *et al.* 2013). Only the female isoform of *Md-tra* is translated into a functional protein. In the presence of Md-Tra, *Md-dsx* is spliced into an isoform that promotes female morphological development. *Md-dsx* is spliced into an isoform that initiates male morphological development in the absence of Md-Tra (Hediger *et al.* 2004, 2010). *Md-fru* is spliced into a male-specific behavioral regulator in the absence of Md-Tra (Meier *et al.* 2013). Read depth coverage of *Md-dsx* (A, B) and *Md-fru* (C, D) in abdomens (A, C) and heads (B, D) of flies with different RNAi treatments. Exons of *Md-dsx* and *Md-fru* are presented along the X-axis. The names of the *Md-dsx* and *Md-fru* exons follow published nomenclature (Hediger *et al.* 2004; Meier *et al.* 2013). Insets in (B) and (D) show female and male isoforms of *Md-dsx* and *Md-fru*, respectively. In *Md-fru*, red exons (s^f and f) that are contained in female isoforms have premature stop codons, but are excluded from the male isoforms. Exons (s, s^f, f) upstream from an exon 'c1' of *Md-fru* are not included in the read depth coverage because they are not on the same scaffold in the genome assembly.



Appendix Figure 5. *Md-tra* expression (A) and PC analysis of global expression (B) of GPF-RNAi and *Md-tra*-RNAi individuals in heads. Arrows indicate the sex-reversed male head sample that we excluded from our analysis because of its outlier expression profile. Females are GFP-RNAi Normal Females; III^M males #1 are GFP-RNAi Normal Males; III^M males #2 are *Md-tra*-RNAi Normal Males; sex-reversed males are *Md-tra*-RNAi Sex-Reversed Males. SR stands for sex-reversed.



Appendix Figure 6. Grade of membership model ($K = 3$) for gene expression patterns of four types of dsRNA injected flies in abdomens (A) and heads (B). Each row is one replicate of each genotype-by-treatment combination. Each color represents the proportion of each replicate assigned to each of the three clusters.



Appendix Figure 7. Boxplots show fold changes of gene expression among comparisons in abdomens (A) and heads (B). Outliers are included as points.

Appendix Figure 8. Histograms of ASE for all chromosome genes are shown in the genotypic (III^M/III) males and the sex-reversed (III/III) males. If a gene is expressed equally between two alleles, the proportion of focal haplotype is 0.5; otherwise, the proportion is greater or less than 0.5.

Appendix Table 1. Differential expression between males with different genotypes

Tissue	Comparison	#Diff	#Genes	Freq Diff
Abdomen	Y^M vs Y^M with new chr III	1159	11533	0.100
	Y^M vs III^M (CS)	511	10344	0.049
	Y^M vs III^M (CSrab)	479	9346	0.051
	III^M (CS) vs III^M (CSrab)	196	10460	0.19
Head	Y^M vs Y^M with new chr III	878	11909	0.074
	Y^M vs III^M (CS)	377	11845	0.032
	Y^M vs III^M (CSrab)	525	12390	0.042
	III^M (CS) vs III^M (CSrab)	739	12409	0.060

Counts of the number of genes that are expressed differentially (# Diff) and total genes expressed (#Genes) are shown, as well as the frequency of genes that are expressed differentially (Freq Diff). Y^M males are from the IsoCS strain; Y^M with new chr III are bwbCS Y^M males with a with standard chromosome III from CS (bwbCS×CS males); III^M males from either the CS or CSrab.

Appendix Table 2. Differential expression between genotypic males and females with different RNAi treatments

Tissue	Comparison	#Diff	#Genes	Freq Diff
Abdomen	III^M males #1 vs females	11030	14993	0.736
	sex-reversed males vs females	11005	14686	0.749
	III^M males #2 vs sex-reversed males	2867	13769	0.208
	III^M males #2 vs III^M males #1	2243	13162	0.170
Head	III^M males #1 vs females	5077	13558	0.374
	sex-reversed males vs females	735	12360	0.059
	III^M males #2 vs sex-reversed males	204	12959	0.016
	III^M males #2 vs III^M males #1	3260	13258	0.246

Counts of the number of genes that are expressed differentially (# Diff) and total genes expressed (#Genes) are shown, as well as the frequency of genes that are expressed differentially (Freq Diff). Females are GFP-RNAi treated normal females; sex-reversed males are *Md-tra*-RNAi treated sex-reversed males; III^M males #1 are GFP-RNAi treated normal males; III^M males #2 are *Md-tra*-RNAi treated normal males.

Appendix Table 3. Genes with sex-biased expression in sex-reversed or genotypic males

Abdomen (# genes)		sex-reversed males vs females			
		male-biased	not female-biased	not male-biased	female-biased
genotypic male (III^M males #1) vs females	male-biased	6136	407	19	1
	not female-biased	483	1574	271	5
	not male-biased	26	274	841	182
	female-biased	5	18	277	4167
Head (# genes)		sex-reversed males vs females			
		male-biased	not female-biased	not male-biased	female-biased
genotypic male (III^M males #1) vs females	male-biased	254	1897	380	4
	not female-biased	50	2221	1177	34
	not male-biased	13	1091	2660	110
	female-biased	3	154	2045	267

Counts are the number of genes that belong to each column and row combination. Columns compare sex-reversed males and normal females. Rows compare genotypic (normal) males and normal females. Genes with male-biased (female-biased) expression are expressed at significantly different levels between the sexes. Genes with not female-biased (not male-biased) have $\log_2$M/F not greater (less) than zero.

Appendix Table 4. Primers for qPCR

Gene	Junction	Sequence Source	Forward Primer	Reverse Primer	Anneal
Md-tra	(A) 4-5	XM 005190699.2	AAAGTCACAAG GCATCTAGTCC	CGTAACGTCTGG GGAGTTTTG	55°C
	(B) 2a-4	XM 005190699.2	AAGCGGAGATTT GGTGAAGG	ACTTTCACCATCG GCAACTG	55°C
	(C) 2a-2b	XM 011297189.2	AAGCGGAGATTT GGTAATTTCCT	TATAGCTTGATAC GAAACTAGCTGA	60°C
	(D) 2b-3	XM 011297189.2	TCTCTTTGGCAG CTTGGATTGG	TTTGTAGATTCCG CTGGCCAAG	60°C
	(E) 2b-4	XM 011297189.2	TGGCCCTGTTTA GTGAAGGTAC	ACTTTCACCATCG GCAACTG	60°C
	(F) 3-4	GU070694.1	TGATGAGGTGAA GGTACCAAGC	ACTTTCACCATCG GCAACTG	60°C
	(G) 3+-4	XM 011297189.2	ACGACAATAATG GTGAAGGTACC	CACCATCGGCAA CTGTATTCAC	60°C
Mdmd		22793_c0	TGGTGCGCCCTT CTTTAAAC	GTTGACGCGGAC AATCAACG	55°C
Md-ncm			TTCCGACTCTGA ATCATCTGAC	GCACTCCTCATA ATCCAAACTG	55°C