The Evolution of Recombination Landscapes and Mechanisms in *Drosophila* in Light of Intragenomic Conflict

By

Lucas W. Hemmer

Submitted to the graduate degree program in Ecology and Evolutionary Biology and the Graduate Faculty of the University of Kansas in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

Date Defended: 13 June 2018

The dissertation committee for Lucas Hemmer certifies that this is the approved version of the following dissertation:

The Evolution of Recombination Landscapes and Mechanisms in *Drosophila* in Light of Intragenomic Conflict

Chair: Justin P. Blumenstiel

Date Approved: 9 July 2018

Abstract

Sex and recombination are ubiquitous across the vast majority of life on earth. In eukaryotes, recombination during meiosis yields new variation that selection acts upon and, thus, facilitates evolution. However, meiosis provides an arena for manipulation and exploitation by selfish genetic elements. Selfish elements can increase in abundance independently of their host organism and frequently at a cost to host fitness. Several types of selfish elements act during meiosis and therefore it is possible for recombination rates and mechanisms to evolve to counteract and ameliorate their negative effects. However, few studies have investigated the interaction between recombination and selfish genetic elements. I conducted three studies on the evolution of recombination mechanisms in light of the impact of selfish elements. I begin my thesis with an introduction on selfish elements, recombination, and their possible interactions in Chapter 1. In Chapter 2, I found evidence that the synaptonemal complex (SC), a protein complex necessary for proper meiotic recombination, is evolving rapidly in *Drosophila* due to positive selection. I proposed several hypotheses to explain the rapid evolution of the SC including the interaction between the SC and centromere-mediated meiotic drive. In the next two experiments, I utilized advances in DNA sequencing to genotype hundreds of *Drosophila* progeny to quantify recombination events. In Chapter 3, I demonstrate that recombination frequency and distribution is robust to transposable element activity in *D. virilis*. The only effect of increased transposable element activity that I discovered was found in rare cases of aberrant recombination events that occur prior to meiosis. In Chapter 4, I constructed the first complete genetic map for *D. yakuba*, a close relative of *D. melanogaster*. The genetic map and previous studies of recombination in species within the *melanogaster* subgroup suggest rapid evolution of recombination, especially in regards to the suppression of recombination near the centromere. My findings support theoretical work that suggests that centromere-mediated meiotic drive can result in the rapid evolution of recombination rates near centromeres. Further studies are needed to definitively prove the link between selfish genetic element behavior and meiotic recombination and how their interaction impacts the evolution of genes, genomes, and species.

Acknowledgements

This dissertation is the culmination of six years of hard work including reading, studying, writing, editing, coding, analyzing, laboratory bench work, failed projects, dead ends, redoing the same experiments over and over again, rebooting the same projects twice or more, late nights, early mornings, coffee, beer, but most of all the dedication and stubbornness to finish. I could not have gotten far without a number of people who have positively impacted my life in a number of ways. They include supporters, mentors, experts, influencers, study buddies, pushers, thinkers, doers, procrastination encouragers, or just other graduate students in the same stressful boat.

During childhood, I was a huge nerd for living things and the natural world both past and present in a small community that was, at best, indifferent to science. Early on, animal and dinosaur books, especially Zoobooks, and PBS shows such as *The Magic School Bus* and *Bill Nye the Science Guy* captured my young imagination. Later in junior high and high school, Russ Schumacher taught difficult science classes which not only prepared me for college but also increased my understanding of the complexity in all scientific fields. Becky Lindgren taught me the importance of math and introduced me to a biology research program at the University of Nebraska-Lincoln that I participated in the summer before college. This was my first experience with science and programming and ultimately influenced my decisions in the years to come.

My four years at Nebraska Welseyan University impacted my life in so many ways. I would like to thank the entire Biology Department at Nebraska Wesleyan for not only teaching biology but also instilling curiousity about the world around me in all aspects of life. I would like to specifically thank my academic advisor, Dr. Jerry Bricker, for encouraging me to go to graduate school, giving me a research project that ended up as my undergraduate thesis, and teaching me to be hypocritical and not take things for face

value. I also want to thank my unofficial advisor and mentor, Dr. Garry Duncan, for encouraging me to stay curious, recommending research programs to me, and always being happy to see me when I showed up at your door. Without both of these professors, I would not have been selected to participate in a Research Experience for Undergraduates (REU) program at Iowa State University. In this program, I worked other scientists, graduate students, and science-focused undergraduates. This was also my first experience working with *Drosophila* as a study organism. My participation in the REU program opened the doors for me get into graduate school immediately after graduating from college.

I want to especially thank Justin Blumenstiel who has impacted my life during graduate school in more ways I can count. He took a chance on me, a small town student attending a small university with few research experiences and no idea what was in store in graduate school. Justin is my biggest supporter but still challenged me to improve and be the best scientist and thinker I can be. A lot of this dissertation would have been impossible without his help. My committee members, John Kelly, Maria Orive, Jamie Walters, and Rob Ward have been a tremendous help along the way as well. Many of taught me in the classroom and under their tutelage as a teaching assistant as well as giving me ideas and insights into my own work and development throughout graduate school.

I could not have completed the research for this dissertation without funding. The Ecology and Evolutionary Biology Department at the University of Kansas provided funding and other support. The amazing staff, especially Aagje Ashe and Dorothy Johanning, is always behind the scenes keeping the department running and sending me emails reminding me to complete one thing or another. Thank you to the Entomology Department at KU for providing money for several summers as well.

I would also like to thank members of the Blumenstiel lab, both past and present, for help with projects, general support, friendship, or just being there during my time at KU. This includes Michelle Wickersheim, Mauricio Galdos, Xi Chen, Kendra Marr, Kelley Van Vaerenberghe, Emily Grantham,

Fiona Wood, Jaimie Johnson, Michael Long, Aquib Jamil, Griffin Bins, Marilyn Barragan, Angel Li, and most of all Alex Erwin. Since joining the lab at the same time, Alex was there going through the same graduate student hardships, giving encouragement, and celebrating our accomplishments. Without her, graduate school would have been more difficult and less fun.

There are several friends to thank that made life enjoyable when not working. My college friends, too numerous to list, were always there when I needed a couch to crash on and let me hang around them even after I left for KU. A large number of graduate students, including Sally Chang, Andrew Mongue, Kaila Colyott, Aniket Sengupta, Kaylee Herzog, Tori Pocius, Desiree Small, and many others made the experience fun, rewarding, and worthwhile. Friends outside of graduate school including Michael Pope, Kelly Kluthe, Jon Lane, and others also helped to forget about graduate school life for a while.

Finally, I want to thank my family for being so supportive during my time in graduate school and loving me no matter what I chose to do. My parents, Dale and Mary, are so proud of me even though they have no understanding of my research and they supported me in a thousand different ways. My siblings, Alex, Zach, and Emily, are always there to support me and even come down to visit me every once in awhile. But most of all, I want to thank my future wife, Ellen Glock, for loving me, tolerating me during my most stressful times, and encouraging me throughout the course of my dissertation. The first two years in graduate school were the most difficult without you there. You have made my life infinitely better and I would not have been able to do this without you. Thank you and love you all.

Table of Contents

List of Tables

List of Figures

List of Appendices

Chapter 1

Introduction to Recombination, Selfish Genetic Elements, and Their Interactions

Sex and recombination, in one form or another, are ubiquitous processes occurring in the vast majority of life on Earth. In theoretical work, asexually reproducing organisms should have a competitive advantage over sexually reproducing organisms (Maynard Smith 1978). However, purely asexual reproduction in eukaryotes is exceedly rare. The generation of genetic variation via recombination is likely one advantage of sexual reproduction. Recombination facilitates evolution by producing genetic variation for natural selection to act upon. It also increases the efficacy of natural selection through the decoupling of beneficial mutations from deleterious mutations (Hill and Robertson 1966). Finally, recombination reduces genetic load by limiting the accumulation of deleterious mutations (Muller 1964). Just as recombination impacts evolution, selection can also direct the evolution of recombination. Recombination rates vary from several crossovers per chromosome arm to being completely absent in one sex (White 1977; Wilfert et al. 2007). Recombination is highly plastic and variable as recombination rates can vary due to genotype, stress, temperature, sequence motifs, heterochromatin formation, chromosome structure, and changing environments (Hill and Robertson 1966; Charlesworth and Charlesworth 1985; True et al. 1996; Burt 2000; Lenormand and Otto 2000; Pardo-Manuel de Villena et al. 2001b; Wilfert et al. 2007; Marand et al. 2017). Studies of recombination require hundreds, if not thousands, of samples to detect these types of fine scale differences. Classic studies of recombination have relied on a handful of phenotypic or genotypic markers to detect recombination events. In contrast, next-generation sequencing technology allows rapid, high-throughput genotyping of thousands of markers to detect recombination events. This now allows a level of fine-scale resolution that is unprecedented (Andolfatto et al. 2011). These advances are opening the door to allow us to learn how recombination changes in different conditions and evolves in different species.

Despite the benefits of sexual reproduction, it also creates an arena for genomic conflict. Reproduction is necessary for the transmission of genetic information to the next generation and survival. Selfish genetic elements gain a reproductive advantage by competing against another allele, gene, or the host organism to increase in frequency during reproduction. Under Mendelian segregation, two independent alleles will

segregate so that each is transmitted to the next generation in roughly equal frequency. Selfish elements can manipulate the processes of meiosis during reproduction to favor their own transmission, distorting equal segregation. This is called meiotic drive and it can happen in one of two ways. The meiotic products containing non-biased allele may be targeted for destruction resulting in overrepresentation of the selfish allele; this is often referred to as segregation distortion. Secondly, meiotic drive may distort Mendelian segregation during meiosis to preferentially increase transmission of one allele over another to the next generation. Asymmetric meiosis in females provides an opportunity for biased transmission as only one meiotic product becomes the female pronucleus; the remaining products become polar bodies and do not transmit their genetic material to the next generation. A selfish element that distorts meiosis to increase its chances of inclusion in the pronucleus will have a significant reproductive advantage over an opposing element. Meiotic drive often creates conflict because drive can infer a cost to the host organism either by reducing fertility or reducing viability due to linkage to deleterious mutations (reviewed in Sarkar 2016). Meiotic drive of sex chromosomes will distort sex ratios within populations and may result in extinction (Hamilton 1967). Several meiotic drivers have been discovered in *Mus musculus*, *Mimulus guttatus*, the genus *Drosophila*, several species of fungi, and other organisms (Raju 1994; Merrill et al. 1999; Bauer et al. 2005; Dyer et al. 2007; Phadnis and Orr 2009; Didion et al. 2015; Finseth et al. 2015; Unckless et al. 2015). Female meiotic drive by modified chromosome structures such as the centromere are thought to explain the rapid evolution of centromere sequences and centromere-interacting proteins (Malik 2009).

Transposable elements (TEs), another type of selfish gene, are mobile elements that increase in number within a given genome. TEs can reduce fitness of an organism via mutation, misregulation of gene expression, chromosomal rearrangement, and aberrant recombination events (Kidwell et al. 1977; Kazazian et al. 1988; Bennetzen 2000; Slotkin et al. 2007; Zhang et al. 2011). Reproduction provides an opportunity for TEs to increase in abundance by invading a new genome after fertilization (Hickey 1982). Organisms have evolved several mechanisms to suppress the negative impacts of TE activity in the germline (reviewed in Ku and Lin 2014; Haig 2016). Additionally, TEs are mostly active within the

germline of eukaryotes with differentiation between germline and somatic cells. Because the germline is responsible for producing haploid products to be transmitted to the next generation it provides an opportunity for TEs to increase in abundance and invade new genomes.

It is advantageous to the organism to evolve a response to the behavior of selfish genetic elements to suppress their deleterious effects. In fact, many have observed rapid evolution of genes hypothesized to be evolving in response to selfish genetic elements (Malik and Henikoff 2001; Anderson et al. 2009; Reinhardt et al. 2014; Blumenstiel et al. 2016; Helleu et al. 2016). Many studies attribute rapid evolution of genes in response to selfish genetic elements to the Red Queen Hypothesis. The Red Queen Hypothesis, originally proposed to explain species interactions (Valen 1973) and popularized to explain parasite and host coevolution (Bell 1982), states that a host evolves to oppose the action of a parasite and the parasite evolves in response to gain an advantage over the host. This creates a cycle of evolution arising from host-parasite conflict while neither maintains a steady advantage. This cycle is reminiscent of the line from *Through the Looking Glass* (Carroll 1872), "it takes all the running you can do to keep in the same place." The parasite and host must constantly evolve as to not be overcome by the other and risk extinction. The same concept can be applied to intragenomic conflict between selfish genetic elements (a genetic parasite) and host genomes; both must keep evolving to maintain an advantage. Losing in an intragenomic conflict results in genetic extinction of the selfish element or extinction of the host.

Recombination and the action of some selfish genetic elements both occur during meiosis. Therefore, it is reasonable for recombination to affect the behavior of selfish genetic elements and vice versa. A theoretical study suggests recombination modifiers can evolve in response to meiotic drivers, such as driving centromeres, resulting in rapid evolution of recombination rates (Brandvain and Coop 2012). Many genes involved in meiosis, including mechanisms of recombination, and transposon suppression are also rapidly evolving under positive selection (Sawyer and Malik 2006; Anderson et al. 2009; Oliver et al. 2009; Demogines et al. 2010; Lou et al. 2014). Several researchers have proposed that the evolution of

these genes is the result of Red Queen dynamics arising from genetic conflict to ameliorate the action of selfish genetic elements. However, there are few studies investigating a definitive link between selfish genetic element behavior and meiotic drive. Additionally, the evolution of recombination between species is little explored. The most extensive studies on differences in recombination between species are in mice (Dumont and Payseur 2011), great apes (Stevison et al. 2015), and the *D. pseudoobscura* subgroup (Stevison and Noor 2010; Smukowski Heil et al. 2015). Additional studies on recombination in different species are needed to provide a better understanding of how recombination changes and evolves and what determines those differences.

The *Drosophila* genus is an ideal system for investigating the interactions between recombination evolution and selfish genetic elements. The phenomenon of recombination was discovered in *D. melanogaster* to create the first genetic map of any organism (Sturtevant 1913). *Drosophila* species are also capable of producing hundreds of progeny in a short time frame necessary to construct genetic maps. Our knowledge of recombination in *D. melanogaster* is arguably the most extensive of any organism (Fiston-Lavier et al. 2010; Comeron et al. 2012; Hunter et al. 2016; Miller et al. 2016; Hughes et al. 2018). Finally, there are major differences in the recombination rates of different species of *Drosophila* in the species investigated so far. The genetic map length *of D. pseudoobscura* is twice the length of *D. melanogaster* (Anderson 1993) and *D. virilis* has a genetic map four times as long (Gubenko and Evgen'ev 1984). In species closely related to *D. melanogaster* there are major differences in recombination rates and distributions (True et al. 1996).

Drosophila is also a key system in our knowledge of selfish genetic elements. Hybrid dysgenesis is a syndrome induced by transposable element activity resulting in reduced fertility, mutation, chromosomal breakage, and aberrant recombination (Kidwell et al. 1977; Kazazian et al. 1988; Bennetzen 2000; Slotkin et al. 2007; Zhang et al. 2011). This syndrome unique to *Drosophila* was integral to studies of how TEs function, their effects on the genome, and the mechanisms of their suppression (McCarron et al. 1994;

Sved et al. 1995; Gray et al. 1996; Preston et al. 1996; Preston and Engels 1996; Blumenstiel and Hartl 2005; Aravin et al. 2007; Brennecke et al. 2008). Several meiotic drive systems are also known in the genus *Drosophila* (Cazemajor et al. 1997; Merrill et al. 1999; Unckless et al. 2015; Pieper and Dyer 2016). The centromere-drive hypothesis in which centromeres act as meiotic drivers was based on rapid evolution of centromere histones in *Drosophila* (Malik and Henikoff 2001; Henikoff and Malik 2002). While the mechanisms of centromere drive in *Drosophila* are elusive, centromere drive has been observed in other organisms (Fishman and Saunders 2008; Chmátal et al. 2014). With a considerable background of investigation in the fields of recombination and genetic conflict, *Drosophila* is an excellent system to investigate how recombination frequency, distribution, and genetic mechanisms can evolve and how intragenomic conflict can influence their evolution.

In this thesis, I seek to understand the influence of selfish genetic element behavior on the evolution of recombination. The genus *Drosophila* is useful for my studies due to the large amount of prior work on recombination and selfish elements, genomic sequences, and reproductive capacity necessary for recombination studies. I conducted three studies focusing on the evolution of proteins necessary for recombination, the effects of transposable element activity on recombination, and the evolution of recombination rates among different species of *Drosophila*. In Chapter 2, I utilized the genomes of over 20 species of *Drosophila* and hundreds of genomes of *D. melanogaster* from two populations to study the molecular evolution of the synaptonemal complex (SC). The proteins comprising the SC are one part of a large number of proteins necessary for meiotic recombination. The SC also interacts with the centromere, hypothesized to be a meiotic drive element in female meiosis. In Chapter 3, I used the hybrid dysgenesis syndrome in *D. virilis* as a model to investigate the effects of transposable element activity on the frequency and distribution of recombination. In Chapter 4, I constructed the first complete genetic map for *D. yakuba* to map changes in recombination rates and distributions among four closely related species of *Drosophila* to understand how recombination changes over evolutionary time.

Chapter 2 *

Holding it Together: Rapid Evolution and Positive Selection in the Synaptonemal Complex of *Drosophila*

* Hemmer LW, Blumenstiel JP. 2016. Holding it together: rapid evolution and positive selection in the synaptonemal complex of *Drosophila*. BMC Evol Biol 16: 91.

Abstract

Background

The synaptonemal complex (SC) is a highly conserved meiotic structure that functions to pair homologs and facilitate meiotic recombination in most eukaryotes. Five *Drosophila* SC proteins have been identified and localized within the complex: C(3)G, C(2)M, CONA, ORD, and the newly identified Corolla. The SC is required for meiotic recombination in *Drosophila* and absence of these proteins leads to reduced crossing over and chromosomal nondisjunction. Despite the conserved nature of the SC and the key role that these five proteins have in meiosis in *D. melanogaster*, they display little apparent sequence conservation outside the genus. To identify factors that explain this lack of apparent conservation, I performed a molecular evolutionary analysis of these genes across the *Drosophila* genus.

Results

For the five SC components, gene sequence similarity declines rapidly with increasing phylogenetic distance and only ORD and C(2)M are identifiable outside of the *Drosophila* genus. SC gene sequences have a higher *dN/dS* (ω) rate ratio than the genome wide average and this can in part be explained by the action of positive selection in almost every SC component. Across the genus, there is significant variation in ω for each protein. It further appears that ω estimates for the five SC components are in accordance with their physical position within the SC. Components interacting with chromatin evolve slowest and components comprising the central elements evolve the most rapidly. Finally, using population genetic approaches, I demonstrate that positive selection on SC components is ongoing.

Conclusions

SC components within *Drosophila* show little apparent sequence homology to those identified in other model organisms due to their rapid evolution. I propose that the *Drosophila* SC is evolving rapidly due to two combined effects. First, I propose that a high rate of evolution can be partly explained by low purifying selection on protein components whose function is to simply hold chromosomes together. I also propose that positive selection in the SC is driven by its sex-specificity combined with its role in facilitating both recombination and centromere clustering in the face of recurrent bouts of drive in female meiosis.

Background

In sexually reproducing eukaryotes, successful meiosis ensures faithful transmission of a haploid set of chromosomes to the next generation. Problems arising during meiosis can lead to meiotic arrest, chromosomal nondisjunction, and infertility. A key step in meiosis is the close alignment of homologous chromosomes, a process known as synapsis. Synapsis is typically essential for establishing meiotic crossovers and a specialized, tripartite protein structure known as the synaptonemal complex (SC) forms the foundation for synapsis (Lake and Hawley 2012; Page and Hawley 2004; von Wettstein et al. 1984).

The SC has been cytologically observed across eukaryotes and the molecular components have been characterized in a range of model organisms including *Arabidopsis thaliana*, *Caenorhabditis elegans*, *Drosophila melanogaster*, *Saccharomyces cerevisiae*, *Mus musculus*, and several species of *Hydra* (Costa and Cooke 2007; Fraune et al. 2012; Page and Hawley 2004). Across this diverse group of eukaryotes the SC maintains, with some exceptions, evolutionary conservation in both structure as a tripartite complex and function in meiotic recombination and synapsis (Page and Hawley 2004). The SC consists of three main parts in most eukaryotes: the lateral elements (LEs), the transverse filaments (TFs) and the central element (CE) (Carpenter 1975; Fawcett 1956; Moses 1956). Two LEs runs along the length of each pair of sister chromatids and directly interact with the meiotic cohesin complex. The TFs extend out from the LEs, resembling rungs of a ladder connecting the juxtaposed chromosomes. The CE is a solid visible element in the center of the TFs and secures them in place. Some eukaryotes lack an observable SC including *Schizosaccharomyces pombe* and *Aspergillus nidulans* (Egel et al. 1982; Olson et al. 1978; Rasmusse 1973). In the case of *S. pombe*, the SC may have been replaced by thin thread-like structures known as the linear elements (Lorenz et al. 2004). *D. melanogaster* males also lack the SC. This coincides with the fact that *D. melanogaster* males also have no meiotic recombination. These

observations indicate that other mechanisms can ensure proper chromosome segregation in the absence of the SC.

Despite the strong structural conservation across eukaryotes, the proteins that comprise the SC are strikingly varied (Grishaeva and Bogdanov 2014). Based on the fact that several eukaryote lineages lack the SC (Loidl 2006; Loidl and Scherthan 2004; Zickler 1999), several authors have theorized that the SC has evolved independently multiple times (Costa and Cooke 2007; Kouznetsova et al. 2011; Page and Hawley 2004). However, a recent analysis (Fraune et al. 2012; Fraune et al. 2013) found that *M. musculus* SC proteins formed monophyletic groups with orthologs in metazoans ranging from cnidarians to humans. This supports a hypothesis of a single SC origin in at least all metazoans. The SC of the Ecdysozoa (which includes molting animals such as crustaceans, *D. melanogaster*, and *C. elegans*) appears substantially different from the SC in other metazoans. SC components from species such as *D. melanogaster* and *C. elegans* show low conservation outside arthropods and nematodes, respectively. The reason for such lack of conservation of SC components is unknown (Fraune et al. 2012; Fraune et al. 2013).

Several SC proteins have been identified and characterized in *D. melanogaster*. Five such proteins are included in this study. EM studies in *D. melanogaster* females indicate the SC is similar in structure to other eukaryotes (Carpenter 1975; von Wettstein et al. 1984) and all five proteins are contained within the tripartite structure (Collins et al. 2014; Page and Hawley 2001; Webber et al. 2004) (Figure 2.1). ORD and C(2)M have been identified as two of the LE proteins in *Drosophila* (Bickel et al. 1996; Manheim and McKim 2003; Webber et al. 2004). ORD localizes to the chromosome arms during early prophase I and is necessary for chromosome segregation, loading of the cohesin complex on the chromosomal axis, normal levels of meiotic recombination, and SC stability (Bickel et al. 1996; Bickel et al. 1997; Khetani and Bickel 2007; Webber et al. 2004). Its role in crossing over is not entirely understood as recombination is not completely eliminated in *ord* mutants and there is an increased amount of DSB repair via the sister

chromatid. This suggests that ORD suppresses sister chromatid exchange (Webber et al. 2004). C(2)M is also a component of the LE and is responsible for chromosome core formation (Khetani and Bickel 2007), SC-dependent meiotic DSB repair, and assembling a continuous CE (Anderson et al. 2005; Manheim and McKim 2003; Page and Hawley 2004). The N-terminus of C(2)M lies within the inner region of the LE and the C-terminus is assumed to face the central region (Anderson et al. 2005). So far, C(3)G is the only known *Drosophila* TF protein (Lake and Hawley 2012). Like other TF proteins, it has globular N- and C-terminal domains and an internal coiled-coil central domain (Page and Hawley 2004). C(3)G forms into parallel dimers with the N-terminal globular domains extending into the CE and the C-terminal domains are anchored to the LE (Anderson et al. 2005). C(3)G is necessary for synapsis, conversion of DSBs into crossovers (Hall 1972; Page and Hawley 2001) and perhaps gene conversion (Carlson 1972). Finally, the CE is comprised of two other proteins along with the C(3)G N-termini, Corona and Corolla. Corona, commonly referred to as CONA, is a pillar-like protein that aligns outside of the dense CE (Lake and Hawley 2012). CONA promotes DSB maturation into crossovers and synapsis does not occur in *cona* mutants (Page et al. 2008). Additionally, CONA both co-localizes with C(3)G and stabilizes C(3)G polycomplexes (Page et al. 2008). Corolla is also localized within the CE and interacts with CONA (Collins et al. 2014). Thought to be comprised of coiled-coil domains much like C(3)G, it is also essential for SC function and recombination. All of these proteins have roles exclusive to female meiosis except for ORD, which also functions in sister-chromatid cohesion in Meiosis I and II and is necessary for gametogenesis in both *Drosophila* sexes (Mason 1976; Miyazaki and Orr-Weaver 1992).

Two hypotheses have been proposed to explain the lack of conservation of the SC: genetic drift and positive selection. A high rate of evolutionary drift in protein evolution in *Caenorhabditis* and *Drosophila* has been proposed to explain the evolution of the lamin proteins (Erber et al. 1999; Peter and Reimer 2012) and ribosomal proteins in Ecdysozoa (Aguinaldo et al. 1997) as well as olfactory genes in *Drosophila* (Nozawa and Nei 2007). Low levels of purifying selection on *Drosophila* SC components would allow it to diverge at a high rate resulting in little conservation. Low levels of purifying selection

might be expected if the major function of the SC was simply to hold homologs together at a proper distance. Under this scenario, there may be few selective constraints on the particular amino acids that function primarily as structural spacers within the SC.

Alternatively, positive selection may contribute to the rapid evolution of SC components. Many studies have demonstrated that reproductive proteins evolve rapidly (Anderson et al. 2009; Civetta et al. 2006; Jagadeeshan and Singh 2005; Nielsen et al. 2005; Swanson and Vacquier 2002; Swanson et al. 2004; Torgerson et al. 2002). In fact, population genetic analyses in *D. melanogaster* and close relatives have previously revealed that *ord* shows a significant deviation from neutrality in *D. simulans*, with more non-synonymous fixations than expected (Anderson et al. 2009). Recurrent meiotic drive and selection to ameliorate this conflict has been proposed to drive positive selection in meiosis genes (Anderson et al. 2009; Malik and Henikoff 2002; Thomas et al. 2009).

I aimed to perform a molecular evolutionary analysis of the SC proteins in *Drosophila* to determine what forces may be driving the high rate of evolution of these proteins. Using the genomic sequence data available for different *Drosophila* species and *D. melanogaster* population data, I aimed to test the null hypothesis that divergence in SC proteins is effectively neutral. In addition, I sought to test the hypothesis that patterns of molecular evolution in SC components are uniform across the genus. Finally, I examined available *D. melanogaster* population data to determine if any deviations from neutrality have occurred in recent time, which would be consistent with ongoing positive selection.

Methods

Ortholog Search

The amino acid sequences of *c(2)M* (CG8249; FBgn0028525), *c(3)G* (CG17604; FBgn0000246), *cona* (CG7676; FBgn0038612) , *corolla* (CG8316; FBgn0030852) and *ord* (CG3134; FBgn0003009) in *D. melanogaster* were acquired from FlyBase 5.57 (Marygold et al. 2013). An additional SC component, SOLO, was not examined due to the fact that it is an alternative splice variant of *vasa*, which is known to play a role in piRNA biogenesis (Yan et al. 2010). These were used in a tBLASTn (Gertz et al. 2006) homolog search in 21 available genomes of *Drosophila* species with a liberal cutoff of E=0.1. This liberal cutoff was chosen to ensure detection of highly divergent orthologs that were subjected to further validation. *D. melanogaster*, *D. sechellia*, *D. yakuba*, *D. erecta*, *D. ananassae*, *D. pseudoobscura*, *D. willistoni*, *D. virilis*, *D. mojavensis*, and *D. grimshawi* genomes were obtained from FlyBase (Marygold et al. 2013). The genomes for *D. ficusphila* [GenBank: AFFG00000000.1], *D. eugracilis* [GenBank: AFPQ00000000.1], *D. biarmipes* [GenBank: AFPQ00000000.1], *D. takahashii* [GenBank: AFFD00000000.1], *D. elegans* [GenBank: AFFI00000000.1], *D. bipectinata* [GenBank: AFFF00000000.1], and *D. miranda* [GenBank: AJMI00000000.1] were obtained from NCBI. The genome of *D. simulans* was obtained from the Andolfatto lab server (Hu et al. 2013) and the *D. mauritiana* genome was obtained from the Schlötterer lab server (Nolte et al. 2013). To identify highly divergent orthologs, an additional tBLASTn search was performed using the most diverged protein sequence captured in the original tBLASTn search. These results were combined with results from BLASTp searches of annotated proteins using the *D. melanogaster* protein sequence. Finally, I included additional ortholog searches with HMMER 3.1b2 (Finn et al. 2011) and PhylomeDB v3 (Huerta-Cepas et al. 2014) as well as orthologs listed in OrthoDB v7 (Waterhouse et al. 2011). This combined approach allowed us to obtain a broad list of candidate orthologs for each of the five SC components. Orthology was then evaluated for candidates by using a reciprocal best BLAST hits approach with tBLASTn. In all cases where orthology was determined the second reciprocal BLAST hit E-value was substantially worse than the ortholog E-value. In addition, synteny for orthologs was evaluated (Appendix 1), though it should be noted that there is substantial gene shuffling within Muller elements across the genus (*Drosophila* 12 Genomes et al. 2007).

Sequence Retrieval

Upon identification of orthologs, sequences from annotated and un-annotated genomes were extracted using identical approaches to limit biases that might arise from using gene annotations only from annotated genomes. DNA sequences 3000 bp upstream and downstream of identified orthologous sequences were first extracted. These were analyzed with FGENESH+, a Hidden Markov Model protein-based gene predictor used to identify the open reading frames in un-annotated DNA sequence using a known protein sequence as a guide (Solovyev et al. 2006). I included 3000 nucleotides of upstream and downstream flanking sequence to ensure that parts of the open reading frame not originally identified in tBLASTn were included. The *D. melanogaster* amino acid sequence was used as the guide.

Sequence Alignments and Drosophila Phylogeny

Sequence alignments were generated using coding sequences (when identified) obtained with FGENESH+ from each species using both translational MAFFT (Katoh et al. 2002) and translational MUSCLE (Edgar 2004) in Geneious v5.6 (Kearse et al. 2012) with default parameters. Sequence alignments were also generated using codon-based PRANK (Loytynoja and Goldman 2005) based on a pre-determined phylogenetic tree (see below) with the "-F" option allowing insertions. These three alignment programs were used to evaluate sensitivity of results to alignment procedure. Concatenated alignments of SC sequences (obtained either by MUSCLE or MAFFT) were used to generate phylogenetic trees required for PRANK alignment and other analyses. Phylogenetic analysis was performed using the Cipres Science Gateway v3.0 with RAxML-HPC Blackbox using default parameters and a GTR model with 100 bootstrap iterations (Miller et al. 2010). The tree topologies produced by concatenated MAFFT and MUSCLE alignments were identical to each other. The SC gene tree topology

also matched the known phylogeny for the *Drosophila* species used in this analysis (*Drosophila* 12 Genomes et al. 2007).

Molecular Evolutionary Analysis

The global omega (ω) value, often referred to as the global *dN/dS* estimate, is a measure of the average selective pressure acting on a gene across an entire phylogeny (Yang and Bielawski 2000). Global ω for each alignment was calculated using HyPhy with a GTR model (Pond et al. 2005) and also with the one-ratio model F3x4 codon model (M0) in the codeml program of PAML v4.4 (Yang 2007). Both analyses made use of the tree topology obtained from phylogenetic analysis described above. Global ω estimates were obtained using all available orthologs, a smaller subset of 12 species within the *melanogaster* group (*D. melanogaster*, *D. sechellia*, *D. simulans*, *D. mauritiana*, *D. yakuba*, *D. erecta*, *D. ficusphila*, *D. eugracilis*, *D. biarmipes*, *D. takahashii*, *D. elegans*, and *D. bipectinata*), and an even smaller subset of six species within the *melanogaster* subgroup (*D. melanogaster*, *D. sechellia*, *D. simulans*, *D. mauritiana*, *D. yakuba*, and *D. erecta*). Estimates were obtained at different levels of divergence to account for potential problems that might occur in the alignment of highly diverged protein sequence.

To quantify heterogeneity in selection pressure, alignments were analyzed with GA Branch using a GTR model of nucleotide substitution (Pond and Frost 2005b) and the previously described phylogenetic tree. Analysis was performed using Datamonkey, the HyPhy web server (Pond and Frost 2005a). GA Branch uses a genetic algorithm and the Akaike Information Criterion to identify the best fitting model for the number of branch ω classes. This allows one to evaluate evidence for heterogeneity in ω across the tree. A model-averaged probability of positive selection ($\omega > 1$) on any of these branches is used to test whether positive selection has occurred.

An analysis of ω was also performed in PAML (Yang 2007) by comparing two different codon based models of evolution. A likelihood ratio-test was performed to compare a model allowing a beta-distributed value of global ω ranging from zero to one (M7) to a model that also included an additional class of codons with ω greater than one (M8). Both of these models were run with the F3xF4 codon model using the nucleotide frequencies at each codon separately and the phylogenetic tree constructed above.

Tests of Neutrality Using Polymorphism and Divergence

While codon models of molecular evolution provide insight into long-term patterns of selection acting on protein coding sequence, population genetic analyses allow for tests of neutrality in more recent time. McDonald-Kreitman (MK) tests of neutrality were performed using polymorphism data from two *D. melanogaster* populations and *D. simulans* and *D. yakuba* reference genomes served as outgroups. The *Drosophila* Genetic Reference Panel v1 (DGRP) (Mackay et al. 2012) provided DNA sequences from 162 *D. melanogaster* isofemale lines collected from a population in Raleigh, North Carolina. In addition, 139 genomes from the *Drosophila* Population Genomics Project v3 (DPGP) (Pool et al. 2012) from 20 separate populations in Sub-Sahara Africa were used. SC gene sequences were collected using BLAST with *D. melanogaster* reference genes as the query. BLAST was performed locally in Geneious. Gaps in the alignment were removed and MK tests were performed online with the standardized and generalized MK test website (Egea et al. 2008). Polarized MK tests were also performed using *D. yakuba* sequences to polarize lineage-specific substitutions. In addition, GammaMap (Wilson et al. 2011) was used to identify particular codons within the SC genes of *D. melanogaster* that have likely been fixed by positive selection. A challenge of the MK test is that polymorphic sites are treated equally and allele frequencies are not taken into account. In contrast, GammaMap utilizes population and divergence data fully. Under a codon model of evolution, polymorphism and divergence data are used to estimate the distribution of fitness effects (DFE) for new mutations and substitutions. GammaMap estimates the γ parameter for each

codon along the length of the gene. γ is the population-scaled selection coefficient, $\gamma = 2PN_e s$, where P is the ploidy level, N_e is the effective population size, and s is the fitness advantage of a derived allele relative to the ancestral allele if the derived amino acid differs from the ancestral allele. Evidence for positive selection driving an amino acid substitution in *D. melanogaster* was deemed significant if the probability of γ greater than 0 was greater than 0.5 in *D. melanogaster* following Wilson et al. (2011). In addition, DnaSP 5.10.1 (Rozas 2009) was used to estimate average pairwise differences within each gene (π) and I compared these to the average pairwise site differences for other meiosis genes previously measured (Anderson et al. 2009). Tajima's D was also calculated in DnaSP (Tajima 1989). Haplotype structure was illustrated with phylogenetic trees built using UPGMA, a hierarchal clustering method (Day and Edelsbrunner 1984), in Geneious 5.6.5 (Kearse et al. 2012).

Results

Distant Orthologs of Drosophila SC Components are Elusive Using Diverse Search Methods

I assembled a list of candidate orthologs of SC components in *D. melanogaster* using BLAST, the HMMER search tool (Finn et al. 2011), and by consulting databases of listed orthologous genes including PhylomeDB and OrthoDB (Appendices 2-7). Orthologs were validated using the reciprocal best BLAST hit approach and hits were consistent with prior ortholog annotations. Only *c(2)M* and *ord* orthologs could be identified in all *Drosophila* species and further outside the genus (Appendices 2-7). The LE gene sequences were identified in every *Drosophila* species by tBLASTn and in several closely related Diptera species using BLASTp against annotated proteins (Appendix 3, Appendix 4). These include *Bactrocera cucurbitae* (melon fly), *B. dorsilas* (oriental fruit fly), *Ceratitis capitata* (Mediterranean fruit fly), *Musca domestica* (housefly) and *Glossina morsitans morsitans* (Tsetse fly). The remaining three SC components, *c(3)G*, *corolla*, and *cona*, could be identified in all species of *Drosophila* with annotated

genomes using BLASTp (Appendices 5-7). The one exception is that *cona* could not be identified within *D. willistoni* (Appendix 7). None of the TF and CE gene sequences could be identified outside of the *Drosophila* genus. These results suggest that the TF and CE proteins are less conserved than those comprising the LE.

SC Genes are Evolving Quickly and According to Position within the SC

HyPhy and PAML were used to calculate global ω with sequences obtained from the tBLASTn search. Orthologs that were only identified with BLASTp could not be reasonably aligned. Thus, the orthologs of *c(3)G* in *D. willistoni* and the *Drosophila* subgenus and orthologs of *cona* in *D. ananassae*, *D. bipectinata*, *D. willistoni*, and the *Drosophila* subgenus were not included in the molecular evolutionary analyses (Appendix 5, Appendx 7). To account for possible issues with alignment quality for divergent sequences, I generated alignments with MAFFT, MUSCLE, and PRANK. The global ω estimates were robust to the three alignment methods (Figure 2.2, Appendix 8). To account for long divergence times between many of the *Drosophila* species, global ω was also estimated across three different scales of divergence. I selected a subset of 12 species within the *melanogaster* group (*D. melanogaster*, *D. sechellia*, *D. simulans*, *D. mauritiana*, *D. yakuba*, *D. erecta*, *D. ficusphila*, *D. eugracilis*, *D. biarmipes*, *D. takahashii*, *D. elegans*, and *D. bipectinata*) and an even smaller set of 6 species within the *melanogaster* subgroup (*D. melanogaster*, *D. sechellia*, *D. simulans*, *D. mauritiana*, *D. yakuba*, and *D. erecta*). Global ω estimates were similar across different scales of divergence and different alignment methods (Figure 2.2). The global ω of each SC component was higher than the median ω for each Gene Ontology (GO) category in *Drosophila* (*Drosophila 12* Genomes et al. 2007). The majority of genes within *Drosophila* have ω estimates less than 0.1 (Haerty et al. 2007) and only two GO categories have a median ω greater than 0.1 (response to biotic stimulus and odorant binding) (*Drosophila 12* Genomes et al. 2007). *ord* has the lowest ω amongst all the SC genes at ~ 0.24 which is twice as high as the median ω for odorant

binding genes and greater than the reported value for seminal fluid proteins (0.17) in the *D. melanogaster* species group (Haerty et al. 2007).

There is an apparent relationship between position within the SC and ω. Although the LE component *ord* is evolving at more than twice the average genome-wide rate ratio, it has the lowest value of ω in the SC (ω: ~ 0.240 PAML, Figure 2.2, ~ 0.265 HyPhy, Appendix 8). *cona* is evolving with the highest rate ratio (ω: ~ 0.500 PAML, Figure 2.2, ~ 0.520 HyPhy, Appendix 8) and the global ω estimate is even higher within the species of the *melanogaster* subgroup (~ 0.600, Figure 2.2, Appendix 8). The estimates of ω increase as a function of position within the SC: lateral element components evolve the slowest, central element components evolve the fastest, and *c(3)G*, which functions as a transverse filament, evolves at an intermediate rate. Because I have only characterized five proteins, there is little power in a test for significance in this relationship. However, it is worth noting that this result is robust to different times scales of analysis

Evolutionary Rate Ratio Variation and Signatures of Positive Selection

I further tested for heterogeneity in ω estimates across the genus. GA Branch (Pond and Frost 2005b) uses a genetic algorithm to estimate and evaluate evidence for multiple classes of ω within a phylogenetic context using the Akaike Information Criteria. It further tests a model for averaged probability for $\omega > 1$ for each branch. Results from GA Branch indicate that the evolutionary rate of SC components has varied considerably (Figure 2.3A, Appendix 9, Appendix 10). *c(3)G* and *corolla* have the fewest evolutionary rate ratio classes (three), *ord* had the most (five), and *c(2)M* and *cona* both have four rate ratio classes (Figure 2.3A). There was support for positive selection ($\omega > 1$) on at least one branch in every SC-coding gene except *c(2)M*. *corolla* had the highest ω estimate in any of the GA Branch analyses. *corolla* also demonstrated a strong signature of positive selection on the branch containing *D. biarmipes* and *D. takahashii* and also the branch prior to the split between *D. eugracilis* and the *melanogaster* subgroup

(Figure 2.3A). *cona* shows the most branches with signatures of positive selection (six). The LE protein *ord* has the lowest global ω but shows multiple branches with high probabilities of positive selection within the *obscura* group and prior to the *D. eugracilis* and *melanogaster* subgroup divergence. Along with the fact that *ord* had the most ω rate classes, this suggests that the evolution of *ord* is highly variable even amongst SC components. It should be noted that since alignment of divergent sequences can be challenging, ω estimates on deep internal branches might not be precise. However, rate ratio variation and significant evidence for positive selection are clearly evident on terminal branches. In particular, for each gene, support for the highest ω class on the phylogeny is evident on at least one terminal or near terminal branch.

Given this rate ratio heterogeneity, I sought to evaluate whether changes in ω estimates tended to co-occur among SC components. This would be the case if structural changes in one SC component drove structural changes in other SC components. A simple test for a correlation between branch ω estimates of different SC components must control for shared demographic changes that influence all proteins in the genome. Therefore, I employed the method of Evolutionary Rate Covariation (Clark et al. 2012; Wolfe and Clark 2015). Clear, alignable orthologs of *cona* are found in the fewest number of species and *cona* was not included in this analysis, limiting this analysis to four SC components. I find significant evidence that ω estimates are correlated between *ord* and *corolla* and also *ord* and *c(2)M* (Figure 2.3B). *c(3)G* shows no significant evidence of evolutionary rate co-variation with any other component, even though it interacts with both the lateral element and the central element.

Evidence for positive selection across the genus was evaluated using the M7 vs. M8 test in PAML. Two models of evolution were compared using a likelihood ratio test; a model with beta-distributed ω values less than one (M7) and the same model with an additional class of codons with ω values greater than one (M8) (Yang 2007). A significant likelihood test indicates a signature of positive selection. Positive selection is evident in *corolla* and this result is robust to both alignment procedure and sampling across

different levels of divergence (Table 2.1). GA Branch also identified at least one branch with evidence of positive selection within each of the three levels of divergence. *c(3)G* also demonstrated evidence for positive selection within the *Drosophila* genus and *melanogaster* group but none was detected within the six species in the *melanogaster* subgroup. This is consistent with results from GA Branch that only identified branches with ω estimates near one outside of this clade. In contrast, *ord* showed significant evidence for positive selection in the *melanogaster* subgroup and nowhere else. The likelihood ratio tests and GA Branch both suggest that while *ord* is the most conserved of the SC components, positive selection intermittently contributes to its divergence. No signatures of positive selection were detected in *c(2)M* and *cona*. For *c(2)M*, this is consistent with results from GA Branch. However, the failure to reject a model of neutral evolution in *cona* stands in contrast to the positive selection detected on multiple branches by GA Branch. This may be explained by the fact that the *cona* coding sequence is much shorter and multiple branches were identified to be very conserved in GA branch. Under these circumstances, global PAML analysis of *cona* may have reduced power to detect a class of codons with ω greater than one.

The results of GA Branch and PAML complement each other and detect positive selection in most of the SC components. Both agree that *c(2)M* shows no sign of positive selection anywhere in the phylogeny or across different divergence times. The TF protein *c(3)G* does show signatures of positive selection outside of the *melanogaster* subgroup in both tests. Likewise, *corolla* shows evidence of positive selection throughout the *Drosophila* phylogeny across different time scales of divergence. Despite having the lowest calculated ω, *ord* shows strong a signature of positive selection within the *melanogaster* subgroup.

Polymorphism and Divergence in the D. melanogaster Subgroup

To characterize the forces that have shaped the evolution of SC components in more recent time, I turn to readily available population data for *D. melanogaster*. I used the second *Drosophila* Population Genomics Project African survey of 139 genomes from 20 African *D. melanogaster* populations (Pool et al. 2012) as well as 162 genomes made available by the *Drosophila* Genetic Reference Panel, a sampling of inbred lines from Raleigh, North Carolina (Mackay et al. 2012). I performed a series of McDonald-Kreitman (MK) tests (McDonald and Kreitman 1991) using *D. simulans* sequences as an outgroup to test neutrality in divergence of SC components. To account for deleterious recessive polymorphisms that are retained in low frequencies, I removed singletons, doubletons, and tripletons. Additionally, the MK test can be used to calculate an alpha parameter – the proportion of substitutions that are positively selected (Smith and Eyre-Walker 2002). A negative alpha value indicates the fixation or segregation of deleterious mutations within the gene. Polarized MK tests were also performed with the *D. yakuba* sequence as an outgroup.

The MK test revealed evidence for deviation from neutrality in some, but not all, SC components. Using population genetic data from *D. melanogaster* and *D. simulans* as an outgroup, an unpolarized MK test does not localize signatures of deviation from neutrality to a certain branch. Polarizing fixations on the *D. melanogaster* branch with *D. yakuba* as an additional outgroup allows one to determine whether the deviation from neutrality occurred on the *D. melanogaster* lineage. Across all tests, I find no evidence for recent selection in *c(2)M* and *cona* (Table 2.2), consistent with molecular evolutionary analyses. In contrast to its overall slowest ω estimate, but consistent with PAML results in the *D. melanogaster* subgroup (Table 2.1), *ord* is the only gene found to deviate from neutrality in both the polarized and unpolarized MK tests (Table 2.2), supporting previous results (Anderson et al. 2009). Positive alpha values from the polarized MK test indicate recent positive selection in *D. melanogaster*. Evidence for positive selection was found for *c(3)G* and *cona* in the unpolarized test using African populations only. However, polarized tests that examine fixations on the *D. melanogaster* lineage fail to reject neutrality for *c(3)G* and *cona*. Thus, the signature of positive selection in *c(3)G* and *cona* can be attributed to changes on the *D. simulans* lineage.

Further investigation revealed *D. simulans* was more highly divergent when compared to both *D. melanogaster* in four SC components (Table 2.3), with *ord* being the exception. *c(3)G* and *cona* both show an excess of non-synonymous divergence within *D. simulans* (Table 2.3). Thus, the results of the MK tests for *c(3)G* and *cona* can be explained by an excess level of non-synonymous divergence on the *D. simulans* lineage. This observation is also made in the GA branch analysis (Figure 2.3A). Though not significant, both *c(2)M* and *cona* show a similar pattern of increased non-synonymous divergence in *D. simulans*. Pooling polarized fixations in every SC gene revealed significantly more non-synonymous fixations in *D. simulans* than *D. melanogaster* (2x2 χ^2 test, N. Carolina $p = 0.004$, Africa $p = 0.003$).

The MK test is inadequate for identifying the codons that have been fixed positive selection. I therefore complemented the MK approach using GammaMap (Wilson et al. 2011) to estimate the γ selection coefficient for each codon. Similar to the MK test, GammaMap utilizes both polymorphism and divergence data. However, it also makes use of frequency data to estimate the strength of selection that has acted individual codons. The selection coefficient is expressed in terms of γ, which is equal to $2PN_e s$, twice the product of the effect population size multiplied by the ploidy level and the selection coefficient. In accordance to Wilson et al. 2011(Wilson et al. 2011), I used the probability of $\gamma > 0$ being 50% or greater as a cutoff for a significant signature of positive selection (Wilson et al. 2011). Since I was using polymorphism data from *D. melanogaster*, I did not perform estimation of γ in *D. simulans*.

Overall, signatures of positive selection on the *D. melanogaster* lineage are demonstrated for all SC proteins across the entire length, with the exception of *cona*. The distribution of putative selection effects were similar using data from two subpopulations of *D. melanogaster* (Figure 2.4, Appendix 11), though more codon variants were deemed significant for evidence of positively selection using data from the North American subpopulations compared to African populations. For example, results for *corolla* using African data provide no significant evidence for recent positive selection at the 50% threshold, in contrast

to results using North American data. This is likely an effect of recent demographic history in North America (Andolfatto 2001; Baudry et al. 2004; Begun and Aquadro 1993; Duchen et al. 2013). Additionally, *corolla* sequences contain many low-frequency segregating alleles that are potentially deleterious. Using DGRP data, no codons in *c(2)M* were identified to be under significant positive selection while there were six noted in using population data from Africa (Figure 2.4, Appendix 11). Overall, many of the same codons estimated to be putatively positively selected using data from one population were also were also found using data from the other population. *ord* and *corolla* show evidence positive selection in specific regions, specifically between codons 50 and 200 in *ord* and between codons 300 and 500 in *corolla* (Figure 1.4). Evidence for selection was also concentrated in *c(2)M* between codons 350 and 500, but using data from Africa, these sites were not above the threshold of 50% probability of $\gamma > 0$. While there were many codons identified to be under significant positive selection in *c(3)G* (16 using African populations, 36 using North American populations), codons under positive selection appeared dispersed along the length of the coding sequence. *cona* showed no particular codons under selection in both *D. melanogaster* samples despite having the highest calculated global ω. This coincides with the failure to detect deviation from neutrality in the polarized MK test (Table 2.2) and a drastic reduction of ω in *D. melanogaster* according to GA Branch (Figure 2.3A).

Finally, pairwise nucleotide polymorphism (π) was calculated for each SC gene. Overall, there is a similar level of nucleotide diversity in every SC component when compared to π genome-wide and mean π for meiosis genes reported in Anderson et al. (2009). The one exception was for *corolla* (Appendix 12). *corolla* estimates of synonymous π are considerably lower in both North America and Africa. Considering Tajima's D, only *corolla* demonstrated a strong negative value (N. Carolina D = -2.055, Africa D = -2.443, Appendix 12), possibly an indication of ongoing positive selection within *corolla*. A sliding window analysis of π and Tajima's D reveal that the central region of *corolla*, 1000 to 1200 nucleotides downstream of the start codon, is almost entirely lacking polymorphism save one doubleton in the African populations (Figure 2.5A) and two singletons within North Carolina (Appendix 13). In

North Carolina populations, 250 bp sliding windows within this region reveal gene regions where $\pi = 0$ (Appendix 13). Flanking this central region, polymorphism increases and Tajima's D is negative as many of the site-wise differences can be attributed to singletons, doubletons, and tripletons. Haplotype structure within *corolla* is illustrated with dendrograms constructed using UPGMA (Day and Edelsbrunner 1984). A region of possible recurrent selection shows a higher proportion of individuals carrying a single haplotype with no diversity (Figure 2.5C). Crucially, within this span, there are 178 base pairs that are completely monomorphic in both Africa and North Carolina. Flanking this region, there is an increase of diversity and fewer individuals carry the haplotype with no diversity (Figure 1.5B, D). This pattern was also observed in the North Carolina population (Appendix 13, B-D). Strikingly, within the 178 bp monomorphic span, there are 8 non-synonymous substitutions and 1 synonymous substitution between *D. melanogaster* and *D. simulans* with ω estimated to be 3.40. This also corresponds to the region identified with GammaMap with the highest density of codons characterized by the highest probability that $\gamma > 0$ (Figure 1.4). This suggests that ongoing positive selection has driven rapid and recurrent change in the protein coding sequence of *corolla*. The low levels of nucleotide diversity within *corolla* in *D. melanogaster* can not be attributed to strong purifying selection since K_a/K_s values between *D. melanogaster* and *D. simulans* are high (Appendix 14). In the African populations, the genomic region including *corolla* has reduced polymorphism compared to flanking regions (Appendix 15A). However, the signature is less clear within the North Carolina population (Appendix 15B) possibly due to overall less nucleotide diversity in the DGRP sequences in comparison to the DGPG sequences. This pattern of reduced polymorphism in a 3 kb region is weaker than other signatures of recent positive selection in *D. melanogaster* (Beisswanger et al. 2006; Benassi et al. 1999; Nurminsky et al. 2001; Rogers et al. 2010). This may indicate that this pattern of reduced polymorphism in *corolla* may be a remnant of positive selection that is not as recent or as strong as other examples of recent positive selection.

Discussion

The SC has been identified across diverse eukaryotes with only a few rare exceptions (Egelmitani et al. 1982; Olson et al. 1978; Page and Hawley 2004; Rasmusse 1973). Homologous protein components of the SC can be found in metazoans ranging from mammals to hydra, indicating that the SC is very likely present at the origin of animals. However, these metazoan SC components are very difficult to detect in Ecdysozoa, including *D. melanogaster* and *C. elegans*, despite the fact that EM studies identify the SC to be structurally similar. Two hypotheses exist for the presence of the SC in the Ecdysozoans: either there has been non-homologous replacement of the SC or an extreme amount of divergence in SC homologs from other lineages.

In support of the hypothesis that a high rate of divergence explains lack of apparent SC protein homology between Ecdysozoa and other metazoans, I presented evidence that the SC is evolving very rapidly within the *Drosophila* genus. Importantly, there is a relationship between the estimated global ω estimates for each protein and the ability to identify orthologs in divergent taxa. Only two genes, *ord* and *c(2)M*, were identified outside of the *Drosophila* genus. These both comprise the lateral element, interact with chromatin, and their ω estimates are the lowest. In contrast *c(3)G*, *corolla*, and *cona* have higher ω estimates and ortholog identification was more difficult in divergent taxa. Therefore, it is reasonable to conclude that the failure to identify orthologs for SC components outside of the *Drosophila* genus is due to their fast rate of evolution, not necessarily by de novo origination within *Drosophila* (Fraune et al. 2013). Such rapid sequence divergence between orthologs may also suggest that sequence identity is not essential for structural integrity of the SC, despite many *Drosophila*-specific SC components sharing remarkable functional homology with SC components in other eukaryotes. Further resolution of this question may require additional approaches to orthology detection that incorporate structural information and ancestral state reconstruction. Alternatively, proteomic analysis of the SC in species outside the *Drosophila* genus may also identify orthologs that this analysis did not.

I further demonstrate that rapid divergence of sequence identity is not effectively neutral and can in part be explained by prevalent and recurrent positive selection within the *Drosophila* species examined. Using GA Branch, I find that SC evolution is not uniform as originally hypothesized. I provide evidence for a range of ω estimates that have significantly fluctuated across time. GA Branch analysis indicated that *cona*, a component of the CE, had the greatest number of branches with evidence of positive selection. Across the full phylogeny and also the *melanogaster* group, a comparison of M7 and M8 models in PAML identified the strongest signatures of positive selection in *corolla*, also a component of the CE (Table 2.1); this same gene also posed a challenge for ortholog detection outside of the genus. In contrast, *ord*, a component of the LE, was estimated to have the lowest global ω across the genus and a strong signature of positive selection was observed only when examining the six species within the *melanogaster* group. I found an increased ω for SC components that do not directly interact with chromatin: components of the CE have the highest ω estimates, components of the LE have the lowest and *c(3)G*, which comprises the transverse filament, has an intermediate estimate. A higher rate of evolution for CE proteins in *Drosophila* is concordant with the observation that CE components are more dynamic across metazoans compared to other components (Fraune et al. 2013). From a structural perspective, the chromatin interaction required of the LE may constrain the rate of evolution. However, CE proteins likely interact with a variety of other meiotic proteins. Therefore, a higher rate of evolution in CE proteins may be partly driven by changes in these interactions.

As the SC is so conserved across eukaryotes, what can explain recurrent positive selection of the SC in *Drosophila*? As previously mentioned, SC components are highly divergent in both *Drosophila* and *Caenorhabditis*. Since both of these genera are in the Ecdysozoa, there may be a shared cause of rapid SC divergence within these two lineages. One shared cause may be the fact that both *D. melanogaster* and *C. elegans* have DSB-independent synapsis. This may lead to reduced constraint on SC components, though it is hard to see how this would lead to recurrent positive selection.

Alternatively, there may be different underlying causes for rapid divergence in these two lineages. There are several features of meiosis that make these lineages unique. *Caenorhabditis* species have holocentric chromosomes with complete crossover interference. *Drosophila* males lack both the SC and meiotic recombination. Thus, multiple forces may independently contribute to the high rate of SC protein evolution in these two lineages.

One possibility is that the rapid evolution and positive selection in SC proteins of *Drosophila* is driven by an interaction between the sex-specific nature of the SC and the rapid turnover of centromeric sequences caused by recurrent bouts of meiotic drive. Previous studies have suggested that sex-specific function can relax selective constraint on a gene and allow it to diverge more freely. This has been proposed to explain the higher divergence of maternally expressed genes such as *bicoid* (Barker et al. 2005; Cruickshank and Wade 2008; Demuth and Wade 2007). All of the SC proteins studied have no phenotypic effect in males when mutant, with the exception of ORD which also plays a role in sister chromatid cohesion in the first and second division of meiosis in both sexes (Bickel et al. 1996; Bickel et al. 1997; Khetani and Bickel 2007; Webber et al. 2004). This additional burden of constraint required by being functional in both sexes may explain why *ord* has the lowest ω value among the SC genes examined.

Because the SC is expressed only in females, it may be particularly influenced by rapid evolution of centromeric sequences driven by meiotic drive. In contrast to male meiosis where all four meiotic products become functional gametes, only one of four meiotic products becomes the egg pronucleus, with the remaining three forming polar bodies. Therefore, strong selection in female meiosis can favor a centromere that is biased to enter the pronucleus over an opposing centromere. A centromeric variant that strongly distorts meiosis in its favor will sweep through the population even though it may convey deleterious effects such as interfering in male spermatogenesis (Hamilton 1967; Henikoff and Malik 2002; Zwick et al. 1999b). This form of competition has been proposed drive rapid evolution of centromeric sequences (Charlesworth et al. 1994; Chmatal et al. 2014; Fishman and Saunders 2008; Haaf

and Willard 1997; Malik 2009; Samonte et al. 1997). Rapid evolution of centromeric sequences arising from centromere drive has also been proposed to explain signatures of positive selection on centromere-associated proteins such as the centromeric variant of histone H3 (Chmatal et al. 2014; Fishman and Willis 2005; Malik and Henikoff 2001).

SC components also have specialized functions at centromeres. Across diverse organisms, early centromeric associations are mediated by components of the SC (Kurdzo and Dawson 2015). For example, in budding yeast, the TF protein Zip1 is required for early centromere coupling (Tsubouchi and Roeder 2005), though not through formation of the SC *per se* (Obeso et al. 2014). In *Drosophila*, SC components have the unique property of mediating centromere pairing in mitotically dividing germ cells (Christophorou et al. 2013; Joyce et al. 2013). Additionally, the *Drosophila* SC is essential for centromere synapsis within the chromocenter (Takeo et al. 2011; Tanneti et al. 2011) where the SC is first assembled prior to assembling along the length of the chromosome arms. Finally, across diverse organisms, the SC persists in centromeric regions long after SC disassembly from the euchromatin (Kurdzo and Dawson 2015). This persistence likely facilitates proper chromosomal segregation (Obeso et al. 2014).

Due to these multiple functions at the centromere, and as has been proposed for centromeric histones (Malik 2009; Malik and Henikoff 2002), positive selection in SC components may be driven by the need to accommodate rapid turnover of centromere sequences driven by bouts of centromere drive in female meiosis. This signal may be enhanced by the sex-specific nature of the SC in *Drosophila*. Additional support for this hypothesis lies in the conservation of *c(2)M* when compared to other SC components. My analyses showed few signs of positive selection in *c(2)M* beyond its high global ω, which was higher than *ord*. In the studies of SC centromere association, *c(2)M* mutants either showed partially reduced centromere clustering (Takeo et al. 2011) or no effect (Tanneti et al. 2011). *c(2)M* may show a weaker signature of positive selection compared to other SC components because it has a limited role in centromeric clustering.

It is also worth noting that the SC plays a crucial role in establishing the landscape of recombination in meiosis. Recent studies have shown that selection may act to modify recombination landscapes as a means to reduce the cost of female meiotic drive, particularly by modulating recombination rates near centromeres (Brandvain and Coop 2012). Previous studies have also shown that the centromere can vary significantly in its effects on local recombination in closely related species of the *D. melanogaster* group (True et al. 1996). Overall, I propose that positive selection may jointly arise from the role that SC components have at rapidly evolving centromeres and modulation of recombination rates in these regions. A combination of these forces, along with sex-specificity, may play an important role in driving rapid evolution of this highly conserved structure in *Drosophila*.

Conclusions

The SC shows little sequence conservation across eukaryotes despite its conserved function in meiotic segregation and recombination. The genes comprising the *Drosophila* SC show almost no apparent homology when compared to SC components in other model organisms. I have determined that the SC components in *Drosophila* are evolving rapidly and their ω estimates are higher than observed for most genes. I conclude that this can be partly explained by positive selection detected in nearly every SC gene. This contrasts to our understanding of the SC as a conserved structure necessary for fertility. I propose that the combination of the female-exclusive function of the SC within *Drosophila*, its role in meiotic recombination, and its interaction with centromeres is driving the rapid evolution of the SC within *Drosophila*.

Tables

Table 2.1: *P*-values of a likelihood ratio between a model of variable selection pressures with no positive selection (M7) and the same model with positive selection (M8) for each SC component.*

Gene & Alignment	Full Phylogeny	*mel* Group	*mel* Subgroup
ORD MAFFT	0.98	0.77	**7.9E-03**
ORD MUSCLE	0.38	1.00	**5.5E-03**
ORD PRANK	0.36	0.81	**4.9E-02**
C(2)M MAFFT	0.27	0.10	0.07
C(2)M MUSCLE	0.02	0.11	0.32
C(2)M PRANK	0.12	0.10	0.45
C(3)G MAFFT	**7.3E-07**	**2.5E-03**	1.00
C(3)G MUSCLE	**6.0E-16**	**1.7E-03**	1.00
C(3)G PRANK	**7.6E-03**	**0.02**	1.00
Corolla MAFFT	**1.2E-07**	**5.3E-08**	**1.3E-03**
Corolla MUSCLE	**6.5E-09**	**1.1E-07**	**0.03**
Corolla PRANK	**1.3E-04**	**5.3E-03**	0.05
CONA MAFFT	0.08	0.05	0.26
CONA MUSCLE	0.06	0.11	0.32
CONA PRANK	0.13	0.11	0.34

*Significant *p*-values in bold.

Table 2.2: McDonald-Kreitman tests (MKT) detecting deviation from neutrality within two population samples of *D. melanogaster* for all SC components.*

		Unpolarized MKT[a]		**Polarized MKT[b]**	
		N. Carolina	Africa	N. Carolina	Africa
ord	α	**0.685**	**0.570**	**0.692**	0.581
	p	**0.021**	**0.042**	**0.034**	0.070
c(2)M	α	0.232	0.733	-0.007	0.657
	p	0.515	0.088	0.988	0.186
c(3)G	α	0.301	**0.709**	-0.111	0.511
	p	0.453	**0.005**	0.834	0.139
corolla	α	0.119	0.471	-0.036	0.371
	p	0.867	0.294	0.964	0.465
cona	α	0.450	**0.833**	-0.406	0.532
	p	0.429	**0.006**	0.680	0.321

* Significant *p*-values in bold.
[a] Detects deviation within *D. melanogaster* and *D. simulans*
[b] Detects deviation within *D. melanogaster* exclusively

Table 2.3: Fisher's Exact tests reveal an increase of non-synonymous substitutions on the *D. simulans* lineage.*

		N. Carolina			**Africa**		
	Substitution	D. mel	D. sim	*p*-value	D. mel	D. sim	*p*-value
ord	Non-syn	15	15	1	15	15	1
	Syn	20	21		20	21	
c(2)M	Non-syn	27	32	0.192	30	32	0.202
	Syn	33	23		36	23	
c(3)G	Non-syn	36	47	**0.030**	31	47	**0.025**
	Syn	40	24		36	24	
corolla	Non-syn	37	35	0.544	35	33	0.539
	Syn	23	15		22	15	
cona	Non-syn	8	17	**0.032**	7	17	**0.014**
	Syn	9	3		9	3	
Total	Non-syn	123	146	**0.004**	118	144	**0.003**
	Syn	125	86		123	86	

* Significant *p*-values in bold.

Figures

Figure 2.1: A model of the synaptonemal complex in *Drosophila*. This model is adapted from Lake & Hawley, 2012 and Collins et al. 2014. So far five genes have been found coding SC components and their proteins localized within the structure: *ord*, *c(2)M*, *c(3)G*, *corolla*, and *cona*.

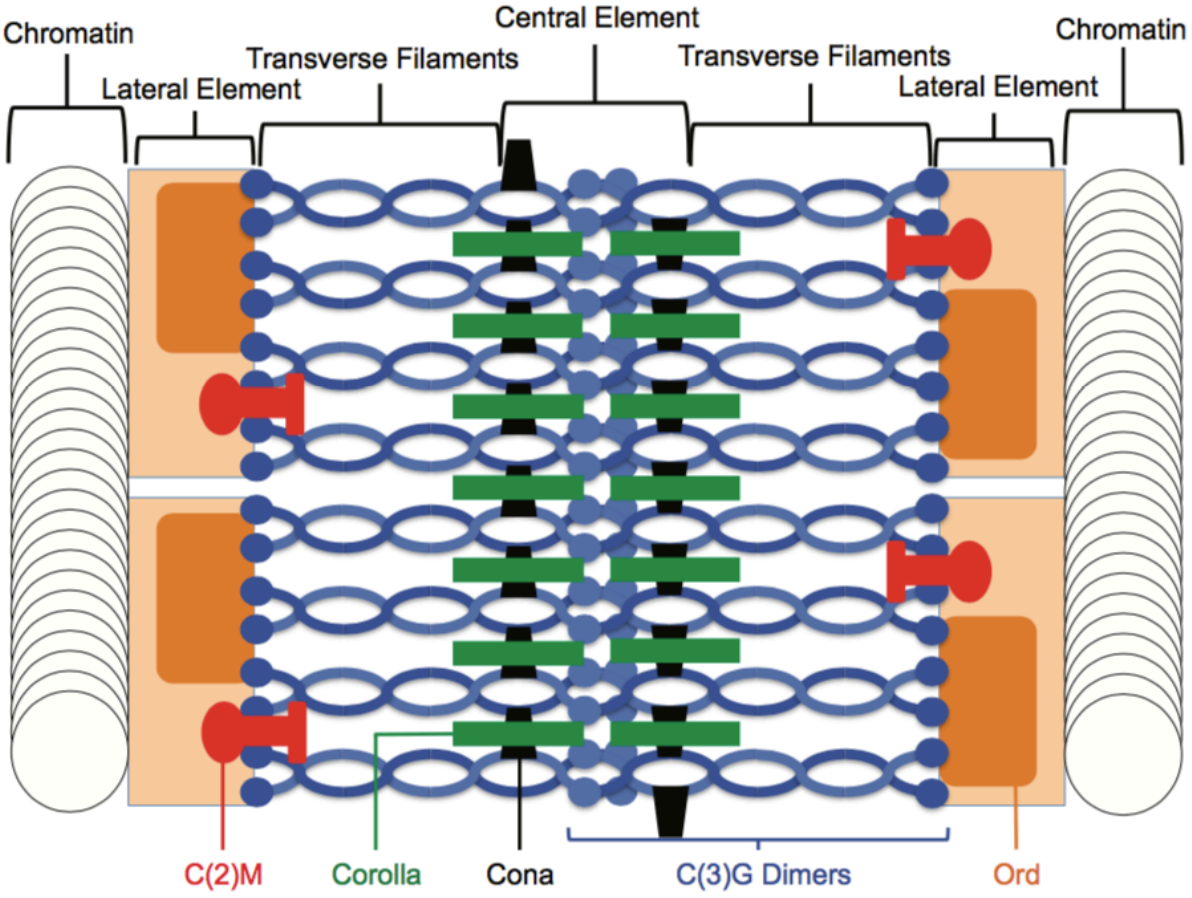


Figure 2.2: The ω ratio of each SC gene is high in each gene and robust to alignment and divergence level. ω was estimated in PAML using a GTR nucleotide substitution model with 95% confidence intervals. The ratio remains relatively consistent for each alignment program used (MAFFT, MUSCLE, and PRANK) and the sampling from the phylogeny. ω estimates are reflected in spatial orientation of the protein within the complex respective to the chromosome; ω increases as distance from the chromosome increases.

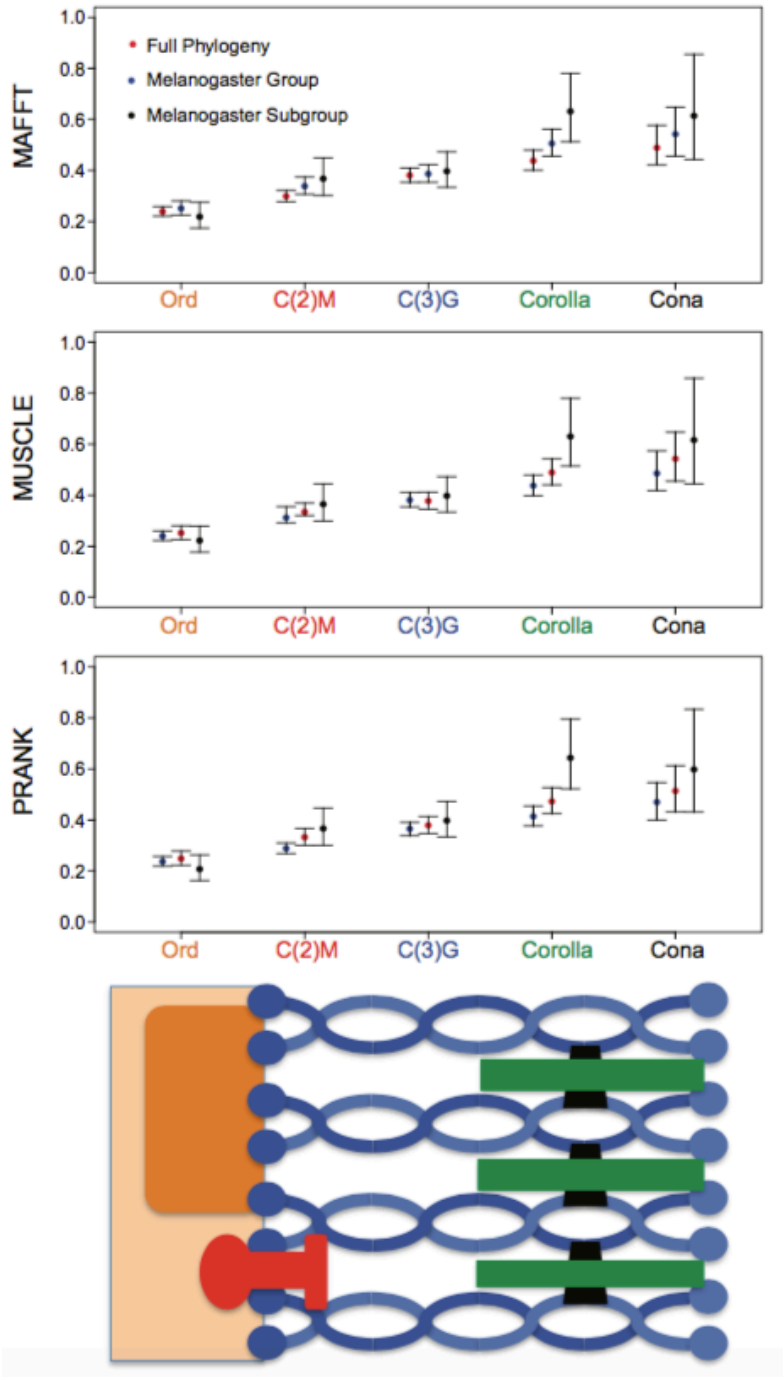


Figure 2.3: GA Branch analysis of the *Drosophila* phylogenetic tree reveals heterogeneity of evolutionary rates for each SC gene. A) Supported rate ratio classes correspond to branch colors. Numbers on the branches present the posterior probability that a gene has evolved under positive selection along that particular branch. The phylogenetic trees correspond to the sequences of *ord*, *c(2)M*, *c(3)G*, *corolla*, and *cona* used in the molecular evolution analyses. B) Evolutionary Rate Covariation analysis. Evolutionary Rate Covariation values (ERC) are above the diagonal. Values closer to 1 indicate higher levels of covariation. *p*-values are below the diagonal.

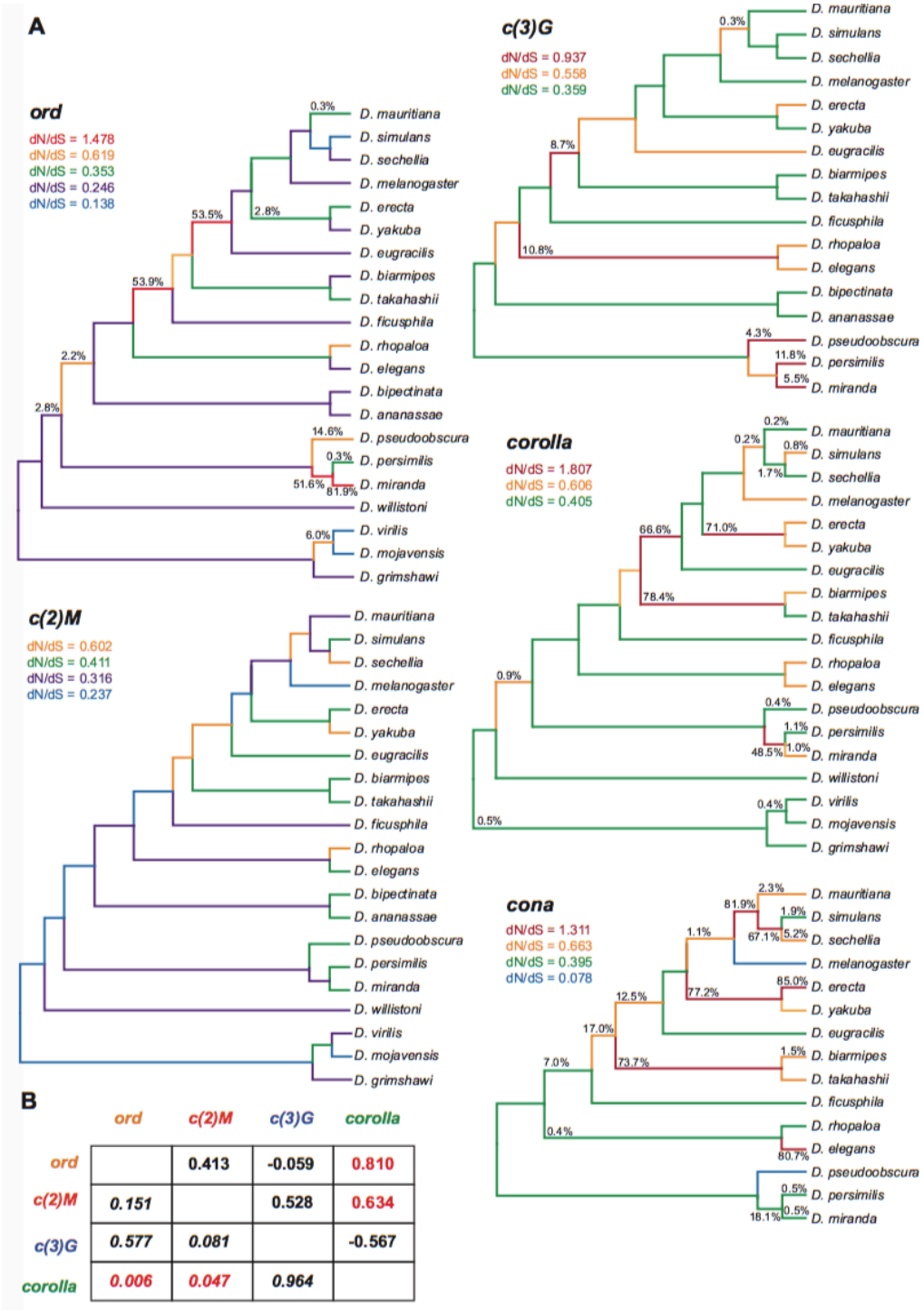


Figure 2.4: GammaMap reveals the posterior probability for positive selection coefficient at each codon using population data from the DPGP. In concordance with Wilson et al. (2011), a codon is under significant signature of selection when the posterior probabilities of selection (lines) are greater than 0.5 (primary Y-axis). Vertical bars illustrate minor allele frequencies in *D. melanogaster* (secondary Y-axis) and the substitutions are the circular dots. The colors correspond to *D. melanogaster* non-synonymous (red) and synonymous (dark green) variants as well as *D. simulans* non-synonymous (orange) and synonymous (light green) variants. Estimated number of selected codons is indicated in the upper right of each plot.

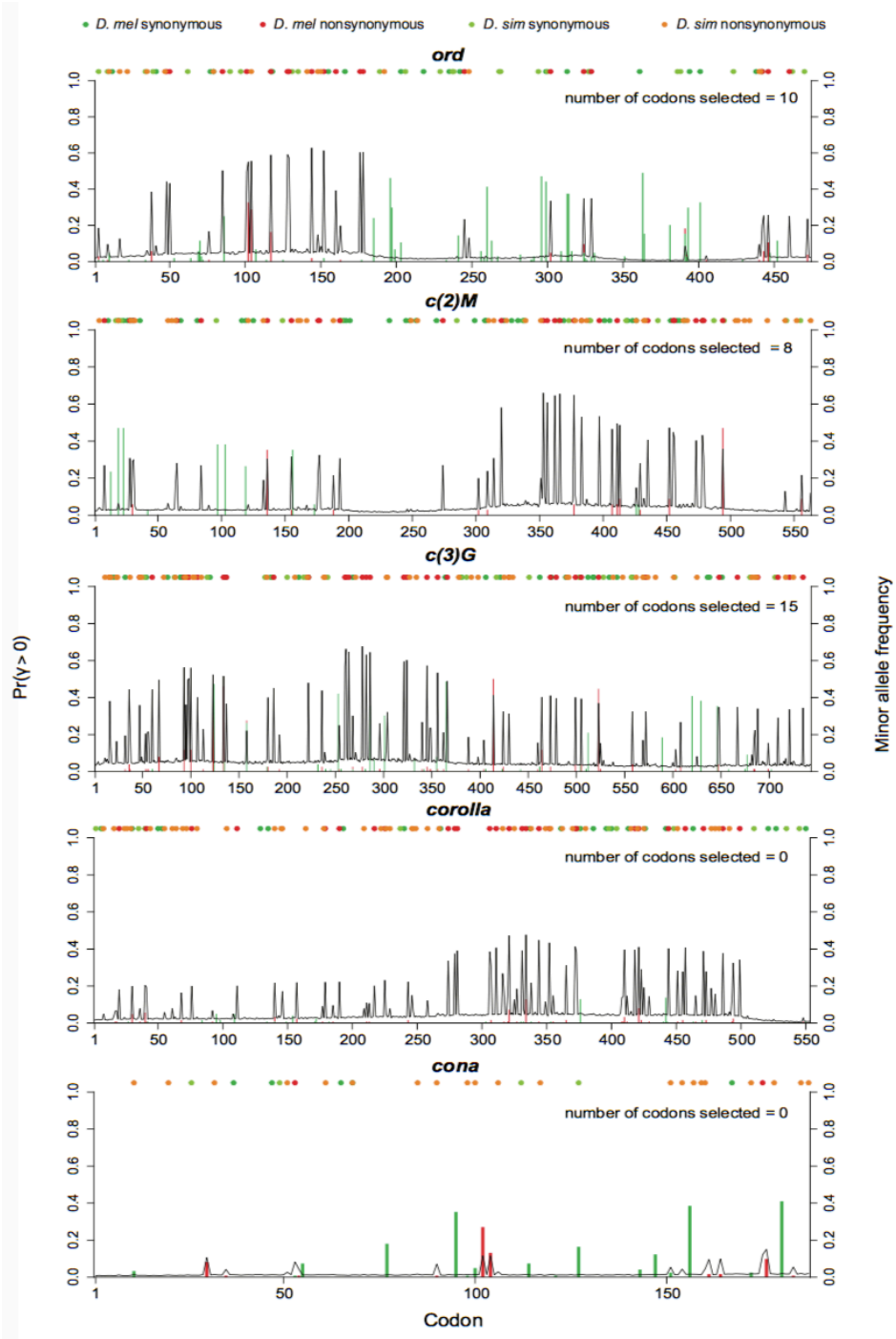
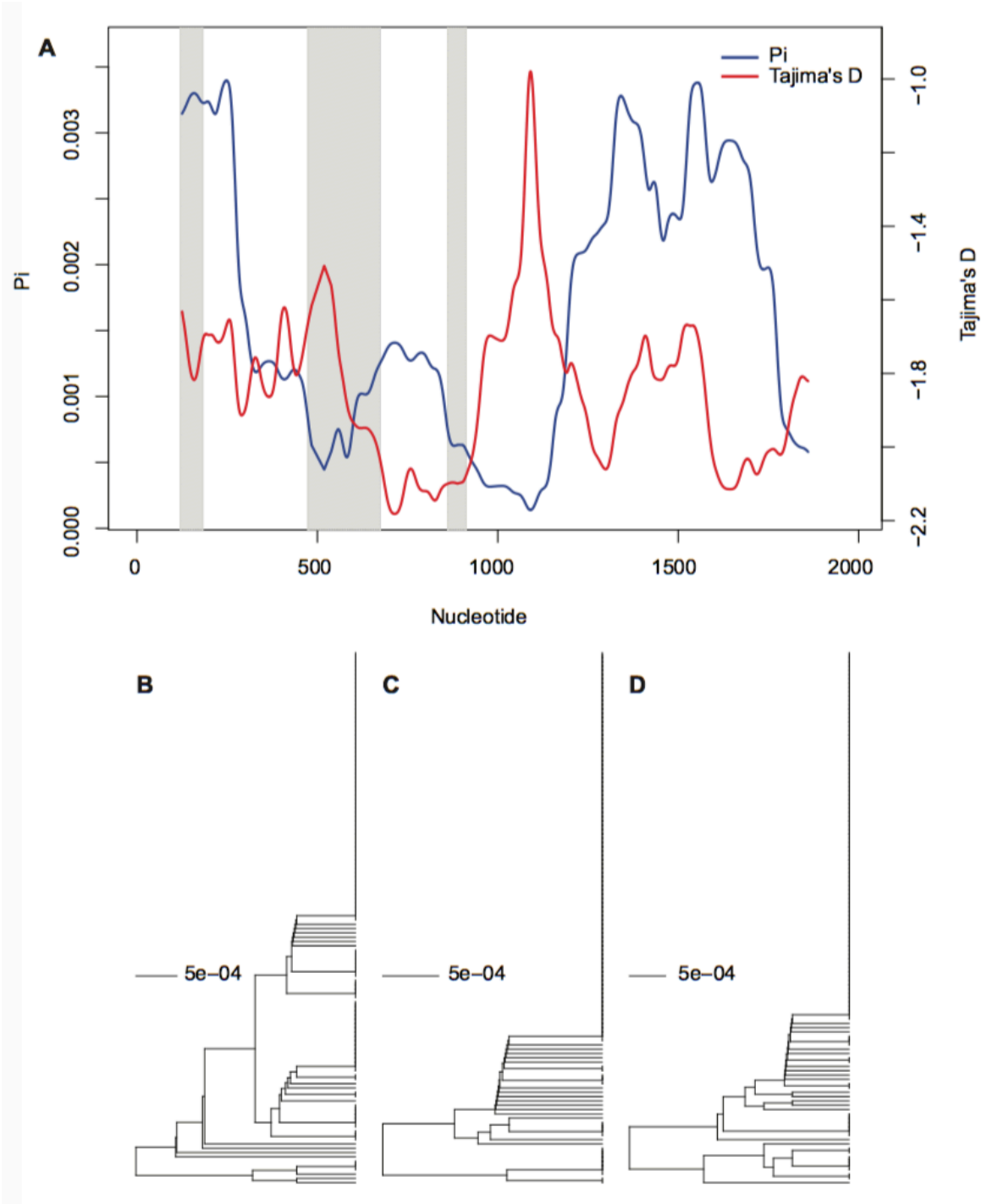


Figure 2.5: Sliding window estimates of pairwise diversity and Tajima's D reveal potential recent signature of positive selection resulting in loss of haplotype diversity. (A) Pairwise diversity (π) and Tajima's D measured in 250 bp windows along the length of *corolla* within the DPGP sequences. Introns are indicated in gray bars. Black lines indicate portions of the gene used in the dendrograms for parts B, C, and D. (B-D) Dendrograms constructed using a HKY model of UPGMA between nucleotides 1-700 (B), 701-1300 (C), and 1301-1938 (D) downstream of the translation start site.



Chapter 3

The Recombination Landscape of *Drosophila virilis* is Robust to Transposon Activation in Hybrid Dysgenesis

Abstract

DNA damage in the germline is a double-edged sword. Induced double-strand breaks establish the foundation for meiotic recombination and proper chromosome segregation but can also pose a significant challenge for genome stability. Within the germline, transposable elements are powerful agents of double-strand break formation. How different types of DNA damage are resolved within the germline is poorly understood. For example, little is known about the relationship between the frequency of double-stranded breaks, both endogenous and exogenous, and the decision to repair DNA through one of the many pathways, including crossing over and gene conversion. I aimed to use the *Drosophila virilis* hybrid dysgenesis model to determine how recombination landscapes change under transposable element activation. In this system, a cross between two strains of *D. virilis* with divergent transposable element profiles results in the hybrid dysgenesis phenotype, which includes the germline activation of diverse transposable elements, reduced fertility, and male recombination. However, only one direction of the cross results in hybrid dysgenesis. This allows the study of recombination in genetically identical F1 females; those with baseline levels of programmed DNA damage and those with an increased level of DNA damage resulting from transposable element proliferation. I used multiplexed shotgun genotyping to map crossover events to compare the recombination landscapes of hybrid dysgenic and non-hybrid dysgenic individuals. The frequency and distribution of meiotic recombination appears to be robust during hybrid dysgenesis. However, I have shown that hybrid dysgenesis is associated with occasional clusters of recombination derived from single dysgenic F1 mothers. The clusters of recombination are hypothesized to be the result of mitotic crossovers during early germline development. Overall, these results show that meiotic recombination in *D. virilis* is robust to the damage caused by transposable elements during early development.

Introduction

Meiosis produces haploid gametes and meiotic recombination plays a critical role in ensuring proper chromosome segregation. However, recombination also serves a benefit to populations since it creates new genotypic combinations that can facilitate adaptation to changing environments (Burt 2000). While meiosis is a conserved process important for fitness, it can be exploited by selfish genetic elements. Transposable elements (TEs) are one selfish genetic element that can increase in abundance within the genome, but at the detriment of the host. TEs can exploit sexual reproduction because fertilization provides an opportunity for TEs to invade new genomes (Hickey 1982). Upon genome invasion, TE proliferation can result in mutation, misregulation of gene expression, chromosomal rearrangement, and improper recombination (Kidwell et al. 1977; Kazazian et al. 1988; Bennetzen 2000; Slotkin et al. 2007; Zhang et al. 2011). Moreover, TEs can activate the DNA damage response within developing germline stem cells resulting in apoptosis and sterility (Levine et al. 2016).

The harmful effects of TEs are especially evident in syndromes of hybrid dysgenesis, when sterility can arise through intraspecific crosses of males carrying TEs absent in females (Bingham et al. 1982; Bucheton et al. 1984; Yannopoulos et al. 1987; Lozovskaya et al. 1990). Hybrid dysgenesis is the result of TE activation in the absence of maternal repression by PIWI-interacting RNAs or piRNAs (Aravin et al. 2007; Brennecke et al. 2008). The piRNA system of genome defense requires recognition of native TEs and maternal deposition of piRNA to successfully silence TEs within the progeny germline. The combination of unrecognized TEs introduced to a naive genome without the accompanying piRNA deposition result in TE activation and hybrid dysgenesis (Brennecke et al. 2007; Brennecke et al. 2008). A hybrid dysgenesis system is known within *D. virilis* in intraspecific crosses between two strains, the inducing strain 160 and the reactive strain 9 (Lozovskaya et al. 1990). The primary TE family responsible for inducing dysgenesis remains unknown but sterility appears to be due to the mass activation of several

TE families abundant in strain 160 but not strain 9. Up to four elements are proposed to contribute significantly to dysgenesis; *Penelope*, *Helena*, *Paris*, and *Polyphemus* (Petrov et al. 1995; Evgen'ev et al. 1997; Vieira et al. 1998; Blumenstiel 2014).

The relationship between TEs and recombination rates is complicated and varies between organisms and TE families. Typically, TEs accumulate in low-recombining regions of the genome, including heterochromatic regions and centromeres (Bartolomé et al. 2002; Rizzon et al. 2002; Jensen-Seaman et al. 2004; Kent et al. 2017). However, some TEs, such as *Alu* elements in humans and DNA transposons in wheat and potato, accumulate in gene rich areas and their density can be positively correlated with recombination rates (Witherspoon et al. 2009; Daron et al. 2014; Marand et al. 2017). The differences might lie in the coevolution between TEs and their host genomes over evolutionary time to drive TEs into low-recombining regions in contrast to other forces such as insertion bias associated with certain TE families (Kent et al. 2017).

TEs may also modulate recombination rates directly through transposition. Transposition itself can induce illegitimate recombination. For example, in the *P-M* system, some males experiencing hybrid dysgenesis undergo recombination (Hiraizumi 1971; Kidwell and Kidwell 1975; Kidwell et al. 1977). This is abnormal since meiotic recombination is absent in *D. melanogaster* males. However, those rates of recombination are typically low, at approximately 2-3% of meiotic events (Preston and Engels 1996). These have been usually attributed to an increased rate of mitotic exchange induced by DNA damage (Isackson et al. 1981). Likewise, many of these recombination events occur closely near the locations of *P* elements, require transposase, and are likely the product of transposition (McCarron et al. 1994; Sved et al. 1995; Gray et al. 1996; Preston et al. 1996; Preston and Engels 1996).

The effects of hybrid dysgenesis on female recombination rates in *D. melanogaster* are less clear. Changes in recombination rates were not initially observed in the *P* element system (Hiraizumi 1971;

Chaboissier et al. 1995) but later studies found a slight increase in female recombination rates during hybrid dysgenesis (Broadhead et al. 1977; Kidwell 1977; Sved et al. 1991). These increases in recombination are often within pericentric heterochromatin. This could indicate loss of control in recombination mechanisms since meiotic recombination within the centromere may impair the centromere functions role in segregation during meiosis (Koehler et al. 1996; Hassold and Hunt 2001; Rockmill et al. 2006). Slightly increased rates were also observed within the *I-R* element system but these were not localized specifically to the centromere (Chaboissier et al. 1995). However, others have identified no increase in recombination rates but rather a redistribution of crossovers towards lower recombining and heterochromatic regions near the centromere (Slatko 1978; Hiraizumi 1981). Since studies in male *Drosophila* indicate that aberrant recombination occurs at low frequencies, differences in female recombination levels are often difficult to detect. Genotyping high numbers of progeny or constructing fine-scale genetic maps are needed to detect changes in the recombination landscape in dysgenic females.

How hybrid dysgenesis impacts the recombination landscape during meiosis remains unclear. Several non-mutually exclusive mechanisms could impact meiotic recombination upon transposable element activation and transposition. These include the following: 1) Increased rates of double-stranded break (DSBs) formation arising transposition, 2) feedback between TE activation of the DNA damage response and meiotic recombination, and 3) modulation of epigenetic marks leading to heterochromatin formation.

The transposition of mobile elements frequently produces DSBs at insertion and excision sites. Coincidentally, DSBs are also the necessary substrate for the initiation of meiotic recombination. In meiosis, this is a controlled, programmed process. Spo11 protein initiations DSBs that become the substrate for repair through homologous recombination. Additional DSBs produced by TEs during meiosis may similarly be repaired by the same homologous recombination machinery resulting in crossovers at the sites of TE insertion and excision. However, TE-induced DSBs that occur in meiosis

may also be repaired by homologous recombination without subsequent crossing. This may occur during synthesis-dependent strand annealing or single-strand annealing. Alternately, non-homologous end joining would also preclude crossover generation. In the *P* element system, crossing over is rare (< 1%), occurring at rates indistinguishable from non-dysgenic crosses (Preston et al. 2006; Johnson-Schlitz et al. 2007).

Transposon activity and DSBs often activate a suite of genes regulating the DNA damage response pathway (Joyce et al. 2011; Shim et al. 2014; Wylie et al. 2016; Ma et al. 2017; Tasnim and Kelleher 2018). Key regulators of the DNA damage response pathway that become activated include *p53*, *chk2*, and ATM (*tefu* in *D. melanogaster*). Additionally, *p53* and *chk2* are activated during hybrid dysgenesis to determine the fate of germline stem cells and the severity of the dysgenic phenotype (Ma et al. 2017; Tasnim and Kelleher 2018). Many of the DNA damage-response genes, including *brca2* and *p53*, are also necessary for regulation of meiotic recombination (Klovstad et al. 2008; Lu et al. 2010; Joyce et al. 2011). Meiotic recombination begins with programmed DSBs created by Spo11 to activate the DNA damage response pathway in a similar manner to non-programmed DSBs or TE activity. Induction of the DNA damage response pathway via transposon activation and hybrid dysgenesis could create feedback between the pathway and typical meiotic recombination. This could lead to global effects on meiotic recombination frequency and distribution in dysgenic progeny.

Transposon activity is often suppressed through heterochromatin formation (Josse et al. 2007; Phalke et al. 2009) and many piRNA genes responsible for directly suppressing TE activity modify heterochromatin formation in progeny by maternal transmission of piRNA (Brower-Toland et al. 2007; Sienski et al. 2012; Le Thomas et al. 2014). Likewise, loss of epigenetic silencing marks upon disruption of piRNA genes leads to de-repression of TEs via reduced heterochromatin formation (Klenov et al. 2007; Sienski et al. 2012; Le Thomas et al. 2013). Differences in maternally-provisioned piRNA profiles from *Drosophila* strains in hybrid dysgenesis crosses could lead to differences in the establishment of

heterochromatin at TE sites. Likewise, chromatin accessibility is a predictor for crossover locations as recombination hotspots are associated with open heterochromatin marks in different species (Berchowitz et al. 2009; Getun et al. 2010; Wang and Elgin 2011; Choi et al. 2013; Shilo et al. 2015; Marand et al. 2017). Even though *Drosophila* do not have recombination hotspots, motifs associated with open chromatin are still reliable predictors for recombination distribution in *D. melanogaster* (Adrian et al. 2016). Thus, differences in the establishment of heterochromatin associated with dysgenesis or TE profiles between strains could lead to changes in the frequency and distribution of recombination along the length of chromosomes. In *Drosophila*, TEs suppressed by heterochromatin-associated epigenetic marks exhibit epigenetic suppression of genes up to 20 kb from the site of the TE (Lee and Karpen 2017). Meanwhile, loss of epigenetic silencing pathways in *Arabidopsis* increases recombination rates in pericentromeric regions resembling the changes in recombination associated with the loss of position-effect variegation genes in *D. melanogaster* (Westphal and Reuter 2002; Underwood et al. 2017).

Previous studies on recombination during hybrid dysgenesis were all in either the *P-M* or *I-R* systems in *D. melanogaster* using phenotypic markers to detect crossovers (Broadhead et al. 1977; Kidwell 1977; Chaboissier et al. 1995). The influence of hybrid dysgenesis on recombination within *D. virilis* is still unknown. Likewise, little is known about recombination in *D. virilis*. Previous studies constructed genetic maps with a limited number of phenotypic or genotypic markers and the estimated recombination rates are highly varied between studies (Weinstein 1920; Gubenko and Evgen'ev 1984; Huttunen et al. 2004) (Table 3.1). This study produces the first fine-scaled genetic map for *D. virilis* using thousands of genotypic markers for further recombination studies. It is also the first to investigate differences in crossover frequency and distribution in the hybrid dysgenesis syndrome of *D. virilis*. I find no detectable differences in recombination between dysgenic and non-dysgenic progeny except in two cases of mitotic recombination produced during dysgenesis, leading to the conclusion that there is no effect of TE activity on meiotic recombination landscapes.

Materials and Methods

Fly Stocks and Crosses

Each strain and all subsequent crosses were maintained on standard media at 25°C. Strain 9 and strain 160 were previously inbred for 10 generations from sibling crosses to form two highly inbred lines for accurate genotyping. Approximately 20 virgin females of one strain and 20 younger males 2-10 days old of the other strain were crossed for six days. Strain 9 females crossed to strain 160 males induced dysgenesis in the F1 generation while the cross in the other direction produced non-dysgenic F1 flies. Individual F1 females four days post-emergence were backcrossed to two or three 2-10 day old strain 9 males in vials for six days. Non-dysgenic females were often allowed to lay embryos for 4-5 days because of their high fertility. Some dysgenic F1 females were transferred to another vial after ten days and allowed to mate for an additional four days to obtain greater numbers of progeny and test whether fertility may be restored with age in hybrid dysgenesis. F2 females were collected once per day and immediately frozen at -20°C. Only 12-20 of the early emerging flies from non-dysgenic F1 backcrosses were collected while all progeny of the dysgenic F1 backcrosses were collected to keep record of the dysgenic F1 parents' fertility.

DNA Extraction and Library Preparation

I extracted DNA with the Agencourt DNAdvance Genomic DNA Isolation Kit (Beckman Coulter) following the Insect Tissue Protocol and stored at -20°C. Prior to DNA extraction, flies were homogenized by 3.5 mm glass grinding balls (BioSpec) placed into a U-bottom polypropylene 96-well plate with lysis buffer from the kit placed into a MiniBeadBeater-96 at 2,100 rpm for 45 seconds. DNA extraction yields varied from <0.1 ng/µl to 5 ng/µl. DNA was quantified using a Qubit fluorometer

(Invitrogen). DNA quantification was not performed on the majority of samples but was assumed to average 1-2 ng/µl necessary for library preparation based on the measured samples.

Library preparations for 192 samples were performed following the protocol outlined in Andolfatto et al. (2011) with some minor changes. Assuming a DNA concentration around 1 ng/µl, 10 µl of genomic DNA were added to a clean 96-well PCR plate and digested with 3.3 U of MseI in 20 µl of reaction volume for 4 h at 37°C with heat inactivation at 65°C for 20 min. Bar-coded adapters (5 µM) were attached to the digested DNA with 1 U of T4 DNA ligase (New England Biolabs) in 50 µl of reaction volume at 16°C for 5 h and inactivated at 65°C for 10 minutes. The samples were pooled and concentrated using isopropanol precipitation (1/10 vol NaOAc at pH 5.2, 1 vol of 100% isopropanol, and 1 µl glycogen). The library was resuspended in 125 µl of 1X Tris-EDTA (pH 8). Adapter dimers were removed 1.5X vol AMPure XP Beads (Agencourt) and resuspended in 32 µl of 1X Tris-EDTA (pH 8). I selected for 200-400 bp DNA fragments using a BluePippin (Sage Science). Size-selected fragments were cleaned using 2X vol of AMPure XP beads and resuspended 20 µl of 1X elution buffer (10 µM Tris, pH 8). The libraries were quantified using a Qubit fluorometer before an 18-cycle PCR amplification on bar-coded fragments with Phusion high-fidelity PCR Kit (New England Biolabs). The adaptors used were the FC1 and FC2 adaptors specifically noted in Andolfatto et al. (2011). The PCR products were cleaned using 1X vol of AMPure XP Beads.

Library preparations for 768 samples were performed using in-house produced Tn5 transposase produced following the procedure outlined in Picelli et al. (2014) following a similar tagmentation protocol. I extracted DNA using the same Agencourt DNAdvance Genomic DNA isolation kit and protocols. Assuming an average of 1-2 ng/µl DNA concentration per sample in a 96-well plate, 1 µl of DNA was tagmented with the in-house Tn5 transposase at a concentration of 1.6 mg/ml with pre-annealed oligonucleotides in a 20 µl reaction volume for 55°C for 7 min and stopped by holding at 10°C. The reaction volume also contained 2 µl of 5X TAPS-DMF buffer (50 mM TAPS-NaOH, 25 mM MgCl2 (pH

8.5), 50% v/v DMF) and 2 µl of 5x TAPS-PEG buffer (50 mM TAPS-NaOH, 25 mM MgCl2 (pH 8.5), 60% v/v PEG 8000) for the desired DNA fragment lengths. The in-house Tn5 transposase was inactivated with an addition of 5 µl of 0.2% SDS and heating the total reaction to 55°C for 7 min. Only 2.5 µl of tagmentation reaction is needed for the PCR amplification with KAPA HiFi HotStart ReadyMix PCR Kit (Thermo Fisher Scientific), 1 µl of 4 µM Index 1 (i7) primers (Appendix 16), and 1 µl of 4 µM Index 2 (i5) primers (Appendix 17) in 9 µl of reaction volume. The PCR occurred as follows: 3 min at 72°C, 2 min 45 sec at 98°C, and then 14 cycles of 98°C for 15 sec, 62°C for 30 sec, 72°C for 1 min 30 sec. The PCR-amplified samples were pooled and cleaned using 0.8 X vol AMPure XP Beads. I size-selected DNA fragments 250-400 bp on a BluePippin and cleaned using 1X vol of AMPure XP Beads.

Sequencing and Crossover Quantification

All libraries were sequenced at the University of Kansas Genomics Core on an Illumina HiSeq 2500 Sequencer with 100 bp single-end sequencing. The first 192 samples were sequenced on two lanes using the Rapid-Run Mode resulting in 120 million reads per lane while the Tn5-produced libraries were sequenced on two lanes using the High-Output Mode producing 180 million reads per lane. FASTQ files were parsed before using the multiplex shotgun genotyping (MSG) bioinformatic pipeline for identifying reliable genotype markers and determining ancestry at those markers using a Hidden Markov Model. Samples with a minimum of 10,000 reads provided reliable genotype calls with high confidence; all samples with less were discarded. Crossover events identified by this pipeline were manually curated for errors. Double crossovers less than 750 kb apart were discarded as these events are unlikely to occur within 1 Mb and most were due to mapping errors. Due to low quality sequences leading to erroneous mapping calls on the ends of chromosomes, crossovers located within 500 kb of the telomere end for the X and 4th chromosome and crossovers within 700 kb on the 2nd, 3rd, and 5th chromosomes were removed. Crossovers near the centromere ends of each chromosome were removed due to erroneous mapping as follows: within 3.5 Mb on the X chromosome, within 1.1 Mb on the 2nd chromosome, within

1.5 Mb on the 3rd chromosome, within 2.4 Mb on the 4th chromosome, and 2.3 Mb on the 5th chromosome.

Data Analysis

Crossover data was parsed and analyzed within the R Version 3.4.2 programming environment (R Core Team 2017). The following packages were also used in genetic map construction, model testing, and general data analysis: R/qtl (Broman et al. 2003), lme4 (Bates et al. 2015), lsmeans (Lenth 2016), and Biostrings (Pagès et al. 2017). Figures were produced using ggplot2 (Wickham 2016). Maker (Cantarel et al. 2008) was used to annotate the genomes of strains 9 and 160 previously assembled using Illumina short-read sequence data in the Blumenstiel lab (unpublished) with the most up-to-date gff file for *D. virilis* (1.6) downloaded from Flybase (Gramates et al. 2017). The genome was annotated for TEs annotations using Repeatmasker (Tarailo-Graovac and Chen 2009) with the current catalog of TE sequences in *D. virilis* from Repbase (Bao et al. 2015). Repeatmasker was also used to identify intact *Polyphemus*, *Penelope*, and *Helena* sites that were less than 5% diverged from the annotated sequence and within 100 bp of the annotated length in a genome of strain 160 assembled using long-read PacBio data available in the Blumenstiel lab (Blumenstiel and Bergman, unpublished).

Results

Crosses and Genotyping

The hybrid dysgenesis syndrome in *D. virilis* is induced in crosses between strain 9 females and strain 160 males but the severity of dysgenesis varies in the resulting progeny (Lozovskaya et al. 1990; Erwin et al. 2015). The F1 females from the dysgenic and reciprocal crosses were backcrossed to strain 9 males. The F2 progeny were sequenced to quantify crossover events produced within F1 female germline. F1

females collected in the non-dysgenic reciprocal crosses were highly fertile and capable of producing large numbers of progeny. By the nature of dysgenesis, most dysgenic F1 females have reduced fertility with many producing few or no offspring. Therefore, all the progeny produced by the dysgenic flies termed "low-fecund" were sequenced and the first 12-20 F2 flies produced from non-dysgenic females were sequenced to balance the number of progeny from a single mother. As previously mentioned, up to 30-50% of the F1 females produced in dysgenic crosses show no outward signs of dysgenesis. The F1 females termed "high-fecund" are capable of producing as many progeny as the non-dysgenic flies. Approximately 40 F2 progeny from several highly fecund dysgenic females were sequenced to obtain greater resolution on the effects of transposition on recombination within a single maternal germline. The variation in dysgenesis provides an additional comparison in the analysis of recombination landscapes between two outcomes of TE activation: TE activation with deleterious effects to fertility and TE activation with no observable negative effects.

F2 female progeny were sequenced on an Illumina HiSeq 2500 and genotypes were called following multiplexed shotgun genotyping (MSG) protocol for indexing (Andolfatto et al. 2011). In total, 828 F2 female flies were sequenced to map recombination breakpoints. Out of the total, 275 F2 flies were sequenced from 20 F1 non-dysgenic females, 311 F2 flies collected from 66 lowly fecund F1 dysgenic females, and 242 F2 flies were collected from seven highly fecund F1 dysgenic females. The MSG pipeline identified a total of 1,150,592 quality-filtered SNPs between the two parental genomes with an average of distance of 136 bp between SNPs. The median crossover localization interval for identified crossovers was approximately 18 kb with 84% of all crossovers localized within 50 kb or less. There were 17 crossovers with interval range at ~ 1 Mb but those were in samples with low read counts near the cutoff for reliable crossover detection (10,000-20,000 reads).

A High-Resolution Genetic Map of D. virilis

The few previous studies on recombination within *D. virilis* indicate much higher rates of recombination than *D. melanogaster*. This has been shown in a genetic map approximately three times the size of the genetic map of *D. melanogaster* (Gubenko and Evgen'ev 1984; Huttunen et al. 2004). Critically, the genetic map lengths estimated between these two studies is highly variable due to the limited number of physical and genetic markers available (Table 3.1). In grouping my samples, I provide the first high-resolution recombination map for *D. virilis*. Of particular interest, the dot chromosome of *D. virilis* occasionally undergoes recombination unlike the dot chromosome in *D. melanogaster*. The rate of recombination on the *D. virilis* dot chromosome is low but detectable occurring at a frequency of 1% (Chino and Kikkawa 1933; Fujii 1940; Gubenko and Evgen'ev 1984). In contrast, I was unable to detect any crossovers on the dot chromosome, likely due to its repetitive nature. There were a limited number of markers for the MSG pipeline on the dot chromosome with 4,013 quality SNPs within the 2 Mb chromosome identified and only 21 genotypic makers shared between all samples. Crossovers identified on the dot chromosome in the MSG pipeline were primarily on the distal ends of the chromosome. Crossover events on the dot chromosome failed to pass inspection because few high quality genotypic markers were located on the chromosome ends resulting in false recombination events inferred by the Hidden Markov Model. The MSG pipeline may not be adequate to detect crossovers in dimunitive chromosomes with a high density of repetitive sequences.

In support of previous studies, the total genetic map length of *D. virilis* in my reciprocal crosses between strain 9 and 160 is 732 cM (centiMorgans) or 2.5 times longer than the genetic map length of *D. melanogaster* (Comeron et al. 2012) (Table 3.1). The total genetic map length in this study is over 100 cM shorter than the first detailed genetic map of *D. virilis* (Table 3.1). This may be due to my stringent criteria of excluding crossovers less than 700 kb in effort to reduce genotyping errors from the MSG pipeline. The differences in recombination rates between *D. virilis* and *D. melanogaster* may appear to be comparable to their difference in genome size. The estimated genome size of *D. virilis* is roughly twice the size of the *D. melanogaster* genome, 404 Mb to 201 Mb respectively (Bosco et al. 2007). Thus, across

the entire genome, the average rate of recombination in *D. virilis* is 1.8 cM/Mb and is similar to the average recombination rate of 1.4 cM/Mb in *D. melanogaster*. However, close to half of the *D. virilis* genome is composed of satellite DNA where little or no recombination takes place (Bosco et al. 2007). Thus, the *D. virilis* euchromatic assembly, where most crossovers take place, is 206 Mb in length and is only half of the total genome size. Accounting for satellite DNA in both species, the average rate of euchromatic recombination in *D. virilis* is twice as high as *D. melanogaster* based on euchromatic assembly genome size (3.5 cM/Mb to 1.8 cM/Mb respectively). The genetic map length of each chromosome correlates with physical length of the chromosome in *D. virilis* ($R^2 = 0.851$, $p = 0.025$).

Crossover interference reduces the probability of an additional crossover in proximity to other crossovers. I calculated interference in *D. virilis* using the Housworth-Stahl model to calculate *nu*, a unitless measure of interference, with a maximum likelihood function based on intercrossover distances (Housworth and Stahl 2003). If crossovers are not subject to interference, intercrossover distances are Poisson distributed and the collection of distances from many crosses resembles a Poisson distribution with a $nu = 1$ (Broman and Weber 2000). Each chromosome in *D. virilis* has detectable interference between crossovers with an average *nu* of ~ 3 (Table 3.2). The Houseworth-Stahl model also calculates the percentage of crossovers produced through an alternative pathway not subject to interference as the escape parameter *P* (de los Santos et al. 2003). Less than 1% of the crossovers in my study are estimated to be produced through the alternative pathway (Table 3.2). In contrast to interference, crossover assurance is the recombination control mechanism to ensure the minimal number of crossovers on each chromosome for proper chromosome segregation during meiosis. In the absence of crossover assurance and interference, the distribution of crossovers within a collection of progeny should resemble a Poisson distribution in that the variance in crossover number is equal to the mean. I find the crossover mean and variance are not equal; the mean crossover number amongst all F2 progeny is 7.3 and variance is 4.9. The distribution of crossovers is significantly different from a Poisson distribution ($\chi^2 = 53.6$, $p = 5.74E\text{-}08$). Moreover, *D. virilis* has a higher than expected frequency of individuals with crossovers numbers close to the mean (5-8

crossovers) with a lower than expected frequency of extreme crossover numbers (2-3 on the low end, 12-15 on the high end) (Appendix 18). This indicates that the collective action of crossover assurance and interference ensures an appropriate number of crossovers on average produced during meiosis.

In many organisms, the total number of crossovers created in a single tetrad is unavailable because crossovers are detected only in a single chromatid transmitted to the progeny. This is known as random spore analysis. Tetrad analysis uses the number of crossovers for each chromosome to calculate the frequency of non-exchange tetrads (E_0), single-exchange tetrads (E_1), or multiple-exchange tetrads (E_n) (Weinstein 1918). I used the classic Weinstein method to determine the frequency of E_0 tetrads for each chromosome in *D. virilis*. The X chromosome and third chromosome E_0 tetrad frequencies were 1.2% and 2.1% respectively (Appendix 19). The calculated E_0 tetrad frequencies in *D. virilis* are lower in comparison to *D. melanogaster*; previously estimates of E_0 tetrad frequencies in *D. melanogaster* range between 5-10% (Zwick et al. 1999a; Hughes et al. 2018). However, the second, fourth, and fifth chromosomes each had biologically meaningless negative E_0 tetrad frequencies ranging from -2.6% to -3.9%. There were also several negative tetrad frequencies estimated for single-exchange and multiple-exchange tetrads as well (Appendix 19). Negative tetrad frequencies are a drawback to using the classic Weinstein method (Zwick et al. 1999a). Nonetheless, these results indicate there are fewer non-exchange tetrads in *D. virilis* compared to *D. melanogaster*.

Recombination rates are often correlated with certain sequence features, such as GC content, simple motifs, and nucleotide polymorphism (Begun and Aquadro 1992; Kong et al. 2002; Comeron et al. 2012). In *D. virilis*, recombination rates appear to be weakly correlated with GC content and gene density as not all chromosomes show significant correlations to either sequence parameter (Table 3.3). This may be due to unusual patterns of recombination along the length of the chromosome (discussed later). Simple repeats and SNP density show strong positive correlations amongst all chromosomes even after removal of non-recombining regions. Nucleotide diversity is frequently correlated with recombination rates (Begun and

Aquadro 1992; Kong et al. 2002) and the strong correlation between SNP density and recombination in my data is consistent with this pattern (Appendix 20). TE density shows a strong negative correlation until non-recombining regions are removed ($p = 0.037$ after Bonferroni correction). The similar pattern of weak or no correlation between TE density and recombination is also seen in *D. melanogaster* because the majority of TEs are found in the non-recombining heterochromatin regions near centromeres (Kofler et al. 2012; Adrion et al. 2017).

No Differences in Recombination Rates nor Frequency in Hybrid Dysgenesis

To compare and contrast the sum of crossovers in the F2 progeny produced by dysgenic and non-dysgenic females, I constructed a full mixed-effects likelihood model using the lme4 R package for the data (Bates et al. 2015). The state of dysgenesis and brood were treated as fixed effects in a Poisson link model. The mother of origin and the fecundity of the mother were treated as random effects. In the full model, I find no difference in the total number of crossovers between the two broods (Type III Wald χ^2, $p = 0.171$). The effect of dysgenesis and the interaction between dysgenesis and brood were nearly significant (Type III Wald χ^2, $p = 0.060$, 0.075 respectively) despite nearly identical mean crossover number between the dysgenic and non-dysgenic flies (7.3 and 7.2 crossovers respectively, Kruskal-Wallis χ^2, $p = 0.622$, Figure 3.1A); random effects of F1 mother and fecundity of the F1 did not have a significant contribution to variance within groups. The progeny of a single highly fecund dysgenic mother termed 701 had a significantly larger mean crossover number in the than the F2 progeny of the other high fecund dysgenic mothers (8.3 crossovers, least squares mean contrast, $p = 0.021$, Figure 3.1B). Without the F2 progeny produced by the 701 mother, the difference in recombination between dysgenic and non-dysgenic, broods, and their interaction is negligible (Type III Wald χ^2, p= 0.874, 0.515, and 0.803 respectively). Likewise, there were no major differences in E_0 tetrad frequency and interference between dysgenic and non-dysgenic mothers (data not shown).

The higher recombination rates in *D. virilis* in comparison to *D. melanogaster* are due to a higher number of crossovers on any given chromosome. In *D. melanogaster*, chromosome arms typically have zero, one, or two crossovers with three crossovers on a single chromosome arm being incredibly rare (Miller et al. 2016). A chromosome with three or more crossovers is a common observation in *D. virilis* in both dysgenic and non-dysgenic directions of the cross with as many as five crossovers on a single chromosome (Figure 3.2). The proportion of chromosomes with zero, one, two, three, or more crossovers is not different between the progeny of dysgenic and non-dysgenic mothers ($\chi^2 = 0.529$, $p = 0.97$). There was also no difference between the progeny non-dysgenic mothers and dysgenic mothers if they were highly fecund ($\chi^2 = 3.70$, $p = 0.45$) nor if the mothers were low fecund ($\chi^2 = 3.45$, $p = 0.49$). Additionally, neither the X chromosome nor any of the autosomes differed in crossover number between the progeny of non-dysgenic mothers, low-fecund dysgenic mothers, and high-fecund dysgenic mothers (Figure 3.2).

I also examined the distribution of recombination along the length of each chromosome between non-dysgenic flies, high fecundity dysgenic flies, and low fecundity dysgenic flies. There were no major changes in the distribution of recombination along the length of the chromosomes among any of the chromosomes (Figure 3.3). The recombination rates between all three groups are strongly correlated amongst all chromosomes (Appendix 21). Overall, I find no differences in the recombination landscape between dysgenic and non-dysgenic F1 mothers in *D. virilis* at the global level. It appears there is no feedback between activation of the DNA damage response by transposition and the modulation of meiotic recombination. I also examined differences in crossover number at the fine scale near intact transposons. Crossover counts examined are near TEs hypothesized to be the main inducers of dysgenesis including the retrotransposons *Helena* and *Penelope* as well as *Polyphemus*, a DNA transposon. There were no apparent differences in recombination rates near the retrotransposons (data not shown). On first examination, there was a significant increase in the crossover number in close proximity to *Polyphemus* sites in the progeny of dysgenic F1 mothers. However, the signal is largely attributed by increased crossing over near a single *Polyphemus* site in the progeny of a single F1 mother (reviewed in the next

section). This suggests that there is no global influence of heterochromatin on recombination near the sites of transposons.

Evidence for Mitotic Crossing Over in Dysgenic Progeny

A single *Polyphemus* site exhibited a much higher number of crossovers among the dysgenic progeny, specifically from progeny of the highly fecund F1 mothers. This cluster of recombination is located near a *Polyphemus* site 9.7 MB away from the telomere on the third chromosome. The 500 kb interval containing the cluster has an increased recombination rate of 26 cM/Mb, nearly twice as high as any other interval within the genome (Figure 3.3C). The region contained 32 crossovers in the progeny of the F1 dysgenic mothers compared to a single crossover in the progeny of non-dysgenic mothers within a 250 kb interval. Of those dysgenic mothers, 21 crossovers were derived from a single F1 mother labeled 5011. Within a 500 kb window, there are 28 crossovers in proximity to this *Polyphemus* site attributed to the progeny of F1 mother 5011. Reciprocal products of the crossover were observed with equal frequency with no transmission distortion (Binomial test, all markers $p > 0.05$, Figure 3.4B). In addition to the crossovers near the *Polyphemus* site, there were additional crossovers detected along the entire length of the third chromosome in the progeny (Figure 3.4B). Only four progeny out of 32 did not have a crossover within the interval on the third chromosome.

An additional cluster of recombination was identified on the X chromosome approximately 21.7 Mb from the telomere with an effective recombination rate of 15.7 cM/Mb (Figure 3.2A). While the rate of recombination within this interval was not as extreme as the interval on the third chromosome, the vast majority of crossovers at this site are attributed to progeny from a single highly fecund dysgenic F1 female labeled 4029. Half of the progeny of the 4029 female had the crossover and no crossovers were detected on the X chromosome distal to the cluster of recombination in all progeny (Figure 3.4C). All progeny obtained were heterozygous for all markers distal to the cluster of recombination. Since F1

females were heterozygous for all SNPs, only half of the progeny are expected to be heterozygous for a given SNP. The extreme excess of heterozygosity in the F2s indicates an extreme transmission distortion of the strain 160 genome from the 4029 mother for this chromosomal region (227 markers between 0.5 - 21.4 MB, Binomial test, $p < 1.6E-08$). However, the proximal region of the chromosome from the cluster of recombination shows no transmission distortion (86 markers between 21.5 - 29.0 Mb Binomial test, $p > 0.5$). Any additional crossovers on the X chromosome in the 4029 progeny were within this proximal region regardless of the presence or absence of crossovers within the cluster of recombination.

I propose both clusters of recombination originated from two separate mitotic crossover events during the early development of germline stem cells (Figure 3.5). Transposons can become mobile in the germline during early development (Engels and Preston 1979; Sokolova et al. 2010). DSBs produced as an outcome of transposition are repaired by one of several mechanisms including homologous recombination via mitotic crossing over. The mass action of TEs from dysgenesis, specifically a *Polyphemus* DNA transposon on the third chromosome in F1 mother 5011, likely produced a DSB to be repaired through homologous recombination in the mitotic germline. One possibility is that mitotic crossover may have occurred prior to DNA duplication, within an early developing germline stem cell in the 5011 mother (Fig 3.5A). The crossover would appear in any daughter cells derived from this germline stem cell and reciprocal products would be observed in equal frequency on average. Alternately, a mitotic crossing over may have occurred after DNA replication prior to mitosis in the 5011 mother (Figure 3.5B). During mitosis, the chromatids segregated resulting in both crossover products in one daughter cell while the other daughter cell receives the non-crossover chromatids. Other germline stem cells must have been present within the 5011 mother because there are several progeny without the common crossover product. The limited number of progeny without a crossover within the cluster of recombination indicates a severe depletion of other intact germline stem cells without the miotic crossover due to hybrid dysgeneis in the F1 mother. Germline stem cells with the mitotic crossover were able to recover and replace other germline stem cells to rescue fertility after hybrid dysgeneis.

Another mitotic crossover event may have occurred in an early developing germline stem cell on the X chromosome of the 4029 mother (Figure 3.5C). A mitotic crossover event occured after DNA replication and the chromatids segregated during metaphase resulting in each daughter cell receiving one chromatid with the crossover and one without. On average, the mitotic crossover would be transmitted to half of the progeny as seen in the data. The pattern of segregation of crossover and noncrossover chromatids results in a loss-of-heterozygosity in the region of the chromosome distal to the crossover. The loss of heterozygosity is responsible for failure to detect additional meiotic crossover derived from the homozygous distal region as seen in the data. The complete transmission distortion of one genome in the distal region is again the result of the depletion of germline stem cells containing the reciprocal mitotic CO products attributed to hybrid dysgenesis. The single lineage of germline stem cells with the mitotic crossover and loss of heterozygosity in favor of the 160 genome recovered and replaced germline stem cells with the reciprocal crossover products or no mitotic crossovers to rescue fertility after dysgenesis. However, the cluster of recombination in the 4029 mother is not in proximity to the three transposons initially investigated nor any other intact TEs. The mitotic crossover may have been the product of a TE insertion but it is unclear from the present data.

Discussion

A High-Resolution Genetic Map of D. virilis

There were no major differences in global recombination between non-dysgenic and dysgenic mothers with the exception of mothers 4029 and 5011, as outlined above. The recombination data was combined together to produce the first high-resolution genetic map for *D. virilis*. The genus *Drosophila* contributes significantly to our understanding of meiotic recombination. Recombination was first discovered in *D.*

melanogaster over 100 years ago and continues to serve as a model for understanding the mechanisms and consequences of recombination. There is also significant work on recombination in *D. pseudoobscura* and genetic maps or recombination studies in *D. simulans*, *D. mauritiana*, *D. yakuba*, *D. persimilis*, *D. miranda*, *D. serrata*, *D. mojavensis*, and others (True et al. 1996; Takano-Shimizu 2001; Staten et al. 2004; Kulathinal et al. 2008; Stevison and Noor 2010; Stocker et al. 2012; Smukowski Heil et al. 2015). A high-resolution genetic map of *D. virilis* will continue to add to the growing number of genetic maps of species of *Drosophila* for future studies of recombination. Of note is the high rate of recombination in *D. virilis* in comparison to other species, especially *D. melanogaster*. Typically in *Drosophila*, recombination rates frequently peak in the middle of the chromosome arm and decrease towards the centromere and telomere to resemble a bell curve (True et al. 1996). However, the distribution of recombination on the second, third, and fourth chromosomes in *D. virilis* resembles a bimodal distribution (Figure 3.3). The high rate of recombination in *D. virilis* may explain why some chromosomes have higher recombination rates near the chromosome ends. Even though a single chromosome may have two or more crossovers, interference prevents crossovers from forming too close together. Interference would favor crossover formation on opposite ends of the chromosome when recombination rates are high. Additionally, *D. virilis* has the highest amount of satellite DNA amongst *Drosophila* species (Bosco et al. 2007) and most of the satellite sequences are not in the current genome assembly of *D. virilis*. The distribution of satellite DNA, associated with low rates of recombination, may be shaping patterns of recombination along the lengths of the chromosomes of *D. virilis*.

Meiotic Recombination in Light of Hybrid Dysgenesis in D. virilis

The vast majority of studies on the association between TEs and meiotic recombination focus on sites where TEs accumulate on an evolutionary time frame. There are also a number of studies in the field of molecular biology elucidating the mechanisms of TE activity and its association with inducing aberrant recombination through a variety of mechanisms including non-homologous end joining and single strand

annealing. This study is one of few to examine differences in the meiotic recombination landscape upon TE activation by using the hybrid dysgenesis syndrome in *D. virilis*. Previous studies of hybrid dysgenesis in *D. melanogaster* either are conflicting as some found no effect (Hiraizumi 1971; Chaboissier et al. 1995), increases in recombination rates (Broadhead et al. 1977; Kidwell 1977; Sved et al. 1991), or changes in the distribution of recombination (Slatko 1978; Hiraizumi 1981). While my findings are in the syndrome of a different species, it is the first study to investigate recombination differences using high-throughput genotyping rather than phenotypic markers. This allows a greater insight in the fine-scale changes in recombination rates and distribution that may have escaped unnoticed before.

I found no major differences in the distribution and frequency of recombination in *D. virilis* overall under hybrid dysgenesis. There is no evidence for feedback between the activation of the DNA damage response to TE mobilization and the response to meiotic recombination. DNA damage response regulators such as p53 may determine the fates of germline stem cells to either undergo atrophy or tolerate TE mobilization (Tasnim and Kelleher 2018). The incomplete penetrance of hybrid dysgenesis in *D. virilis* may be due to undiscovered differences in the modulation of the DNA damage response and germline fate. Differences in the DNA damage response may lead to differences in meiotic recombination between high and low fecund dysgenic mothers. However, the null effect of fecundity on recombination is further evidence the DNA damage pathway activated by dysgenesis does not feedback into meiotic recombination. This is presumably due to the early effects of hybrid dysgenesis in *D. virilis*.

DNA Transposon Inducing Mitotic Rather than Meiotic Recombination

Upon initial inspection, there appeared to be increases in crossover number in proximity to the DNA transposon *Polyphemus*. The increase in recombination was originally attributed to modulation of heterochromatin at *Polyphemus* sites or the mechanism of transposition. No difference in crossover

number at retrotransposons suggests the latter is more likely. Why would heterochromatin formation at *Polyphemus* be any different? *Polyphemus*, *Helena*, and *Penelope* are all more highly abundant in strain 160, are provisioned with more piRNA in the strain 160 germline, and mobilize during dysgenesis (Petrov et al. 1995; Blumenstiel 2014; Erwin et al. 2015). *Helena* is interesting in that it continues to be highly expressed in the germlines of the dysgenic progeny in comparison to non-dysgenic progeny (Erwin et al. 2015) but that does not result in higher crossover numbers near it. Additionally, there is evidence that while maternally-provisioned piRNA profiles responsible for heterochromatin formation at TE sites differ between strains 9 and 160, this does not translate to major differences in the heterochromatin modulation between dysgenic and non-dysgenic progeny (Evgen'ev, personal communication). It appeared that *Polyphemus*, a DNA transposon, increases crossover number by producing a DSB upon activation via excision from the site. However, initial observation of an increase in recombination rates near *Polyphemus* is attributed to a single case of a mitotic crossover in nearly all off the progeny of mother 5011. A mitotic crossover may be the result of a DSB produced by a *Polyphemus* excision in the developing germline of the F1 mother. Mother 5011 was capable of producing high numbers of progeny, providing better power to detect the mitotic crossover event. Other mitotic crossovers, especially in low fecundity dysgenic flies, can occur without detection in my data analysis because of the reduced power to detect those events in a limited sample of progeny. Previous studies indicate TEs mobilize during hybrid dysgenesis in the early developing germline within embryos (Engels and Preston 1979; Sokolova et al. 2010). In *D. virilis*, TE suppression resumes by adulthood in dysgenic progeny via production of piRNAs and the negative impacts of dysgenesis disappear in the following generations (Erwin et al. 2015). This indicates TEs rarely produce DSBs when germline cells are undergoing meiosis and homologous recombination to repair transposon-induced DNA damage occurs prior to meiosis. Meiotic recombination appears robust to TE activity and recombination is observed only in rare cases when homologous recombination produces a cluster of recombination or loss of heterozygosity as a result of a crossover in the early developing germline.

Tables

Table 3.1: Genetic map lengths in centiMorgans of *D. virilis* chromosomes reported in previous studies and this study.

Source	Chromosome					
	X	2	3	4	5	6
Gubenko & Evgen'ev 1984	170	257	145	108	203	1
Huttunen et al. 2004	-	118	125	147	60	-
This study	143.5	160.9	139.6	148.3	140.0	-

Table 3.2: Interference values (*nu*) and frequency of crossovers created in the non-interference pathway (*P*) for all chromosomes. Both values were estimated with the Housworth-Stahl model for the entire dataset.

Variable	**Chromosome**				
	X	2	3	4	5
nu	3.22	3.17	2.68	3.09	3.37
P	4.81E-02	1.05E-02	2.73E-08	6.91E-03	9.75E-03

Table 3.3: Correlations between rates of recombination and sequence parameters in 250 kb intervals along each chromosome in *D. virilis*. Pearson's correlation coefficients (R) and significance (*p*-value) are listed. Significant values are bolded.

Sequence Parameter		Chromosome					Total	Total minus zero recomb
		X	2	3	4	5		
GC Content	R	0.08	**0.35**	0.11	0.18	**0.33**	**0.23**	**0.15**
	p	0.372	**1.29E-05**	0.263	0.079	**5.03E-04**	**1.08E-08**	**03.68E-04**
Gene density	R	**0.31**	**0.21**	0.12	**0.32**	**0.33**	**0.19**	0.03
	p	**2.89E-04**	**1.16E-02**	0.222	**3.27E-04**	**5.443E-04**	**1.88E-06**	0.506
Simple repeats	R	**0.44**	**0.43**	**0.31**	**0.32**	**0.54**	**0.39**	**0.177**
	p	**2.04E-07**	**3.11E-08**	**1.31E-03**	**03.57E-04**	**1.14E-09**	**3.03E-24**	**2.91E-05**
SNP Density	R	**0.64**	**0.553**	**0.60**	**0.67**	**0.65**	**0.62**	**0.49**
	p	**4.73E-16**	**2.49E-13**	**7.31E-12**	**4.70E-17**	**1.69E-14**	**5.34E-66**	**6.42E-34**
TE Density	R	**-0.47**	**-0.47**	**-0.33**	**-0.44**	**-0.49**	**-0.44**	**-0.14**
	p	**1.58E-08**	**1.27E-09**	**6.11E-04**	**4.08E-07**	**5.18E-08**	**4.05E-30**	**1.07E-03**

Figures

Figure 3.1: The distribution of the total crossover (CO) count per F2 progeny with the mean and standard deviation. The mean for each group is represented with a black dot and the standard deviation is the black line. A) The distribution of the total CO count per F2 progeny of low fecund dysgenic, high fecund and non-dysgenic F1 mothers. B) The distribution of CO count per F2 progeny of each high fecund dysgenic mother with mean and standard deviation. Asterisks denotes statistical significance by least square means ($p < 0.05$). Progeny from mother 701 had a higher average CO count than progeny from other mothers while progeny from mother 4029 exhibited a lower average CO count.

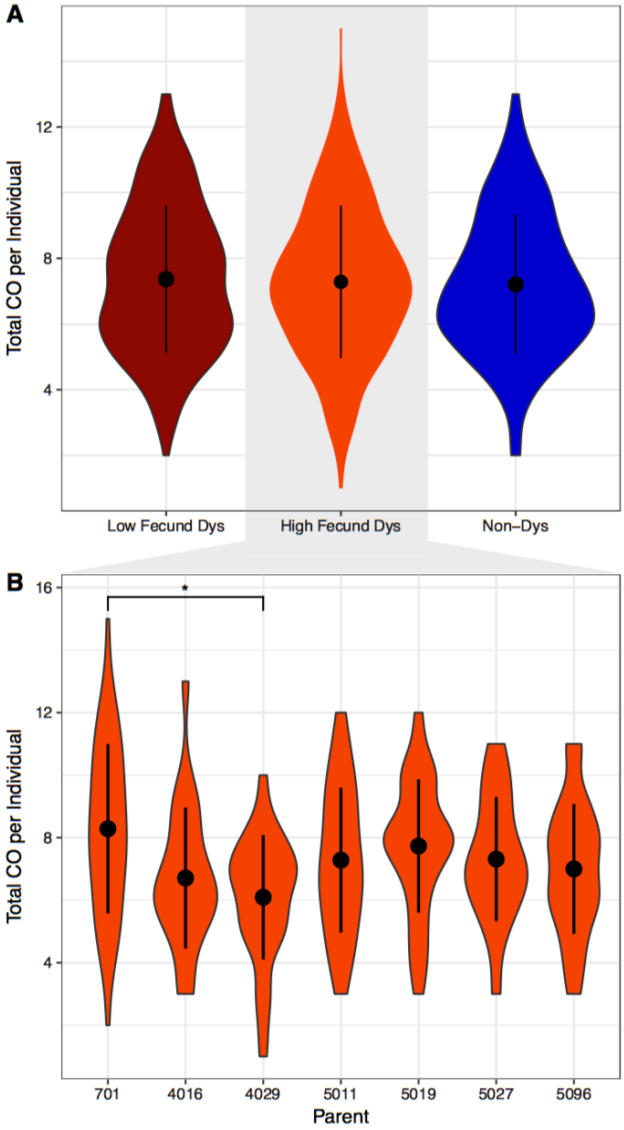


Figure 3.2: The proportion of chromosomes grouped by crossover (CO) count in F2 progeny of high fecund dysgenic, low fecund dysgenic, and non-dysgenic F1 mothers. 95% confidence intervals were calculated by bootstrap sampling F2 progeny 1000 times.

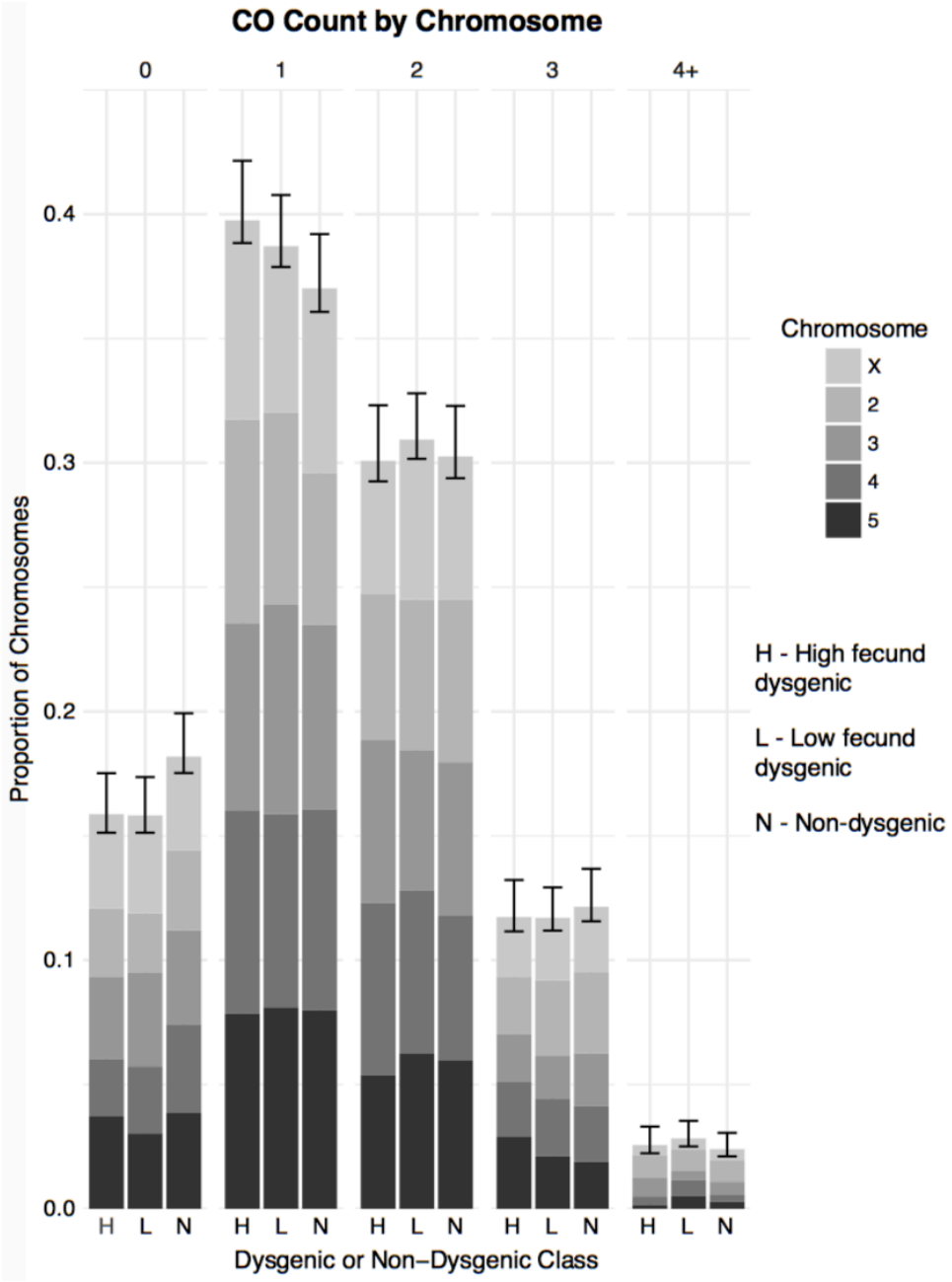


Figure 3.3: Loess smoothed splines of the recombination rate along the length of each chromosome in D. virilis from the telomere (left) to the centromere (right) with standard error. The dotted line represents the centromere effect of recombination suppression as recombination = 0 from the line to the end of the sequence. The rate of recombination was calculated in 500 kb intervals in F2 progeny of low fecund dysgenic, high fecund and non-dysgenic F1 mothers for the A) X chromosome, B) 2nd chromosome, C) 3rd chromosome, D) 4th chromosome, and E) 5th chromosome.

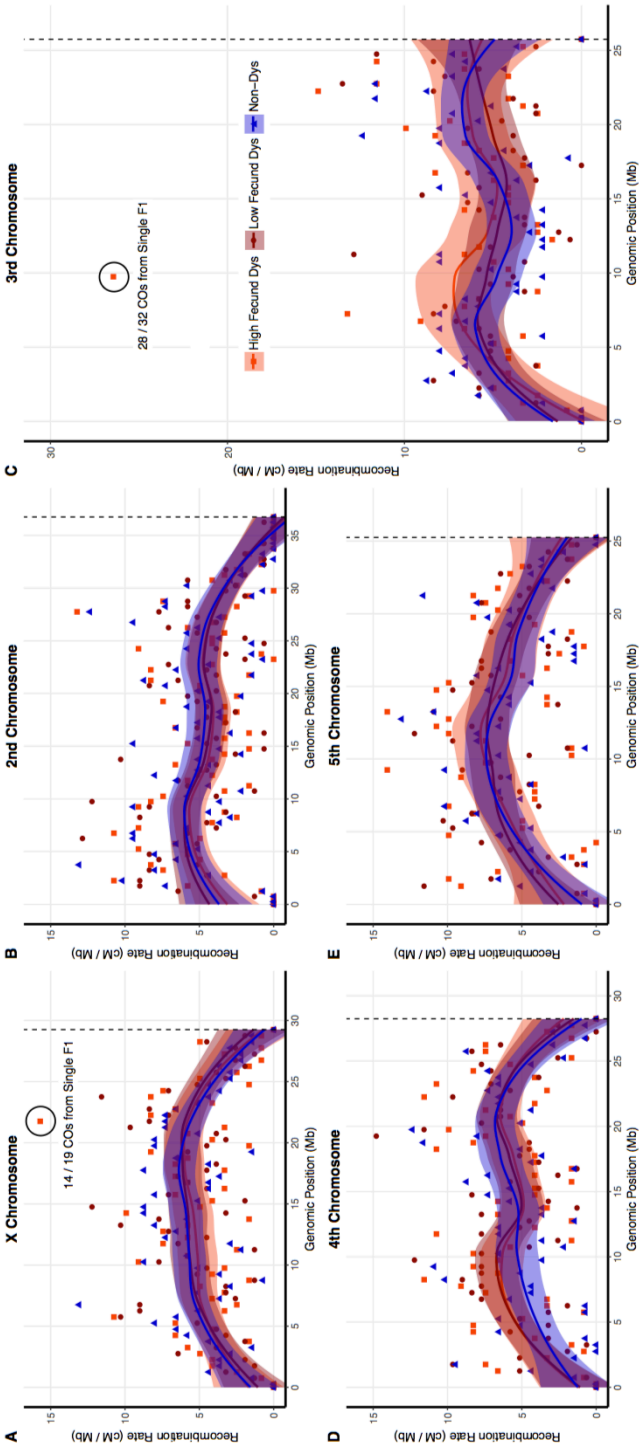


Figure 3.4: Haplotypes of F2 progeny from a single high-fecund dysgenic mother. A) Haplotypes of the third chromosome in progeny of the 4029 F1 mother is typical of most chromosomes with no cluster of recombination. B) Haplotypes of the third chromosome in progeny of the 5011 F1 mother identify a cluster of recombination in most of the progeny and reciprocal products of recombination in equal frequency (Binomial test, $p > 0.05$). C) Haplotypes of the X chromosome in progeny of the 4029 F1 mother indicate a cluster of recombination in half of the progeny and extreme segregation distortion of the distal portion of the chromosome (227 markers 0.5 - 21.4 MB, Binomial test, $p < 1.6E-08$). The proximal region of the chromosome shows no segregation distortion (86 markers 21.5 - 29.0 Mb Binomial test, $p > 0.5$).

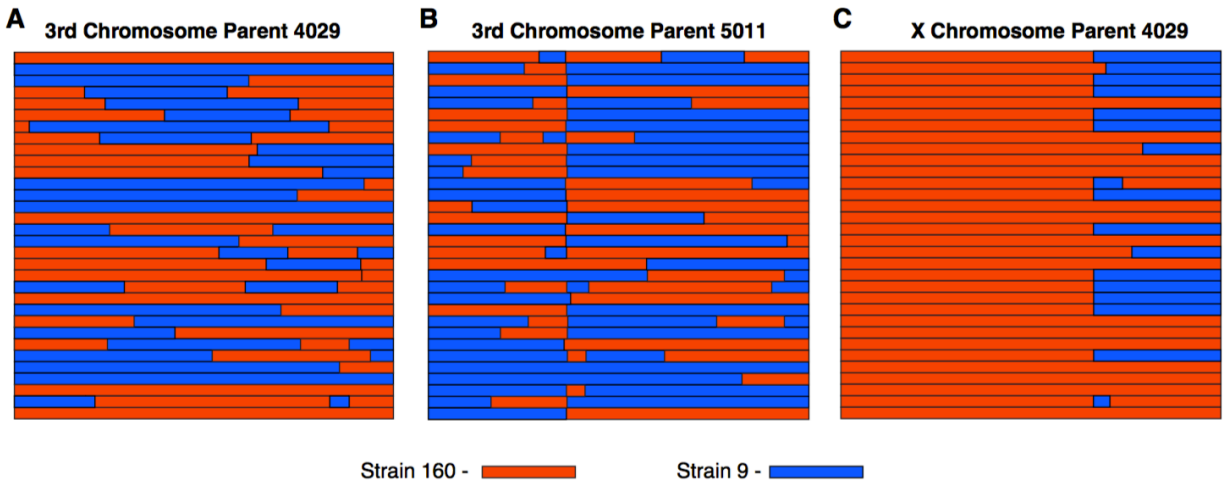
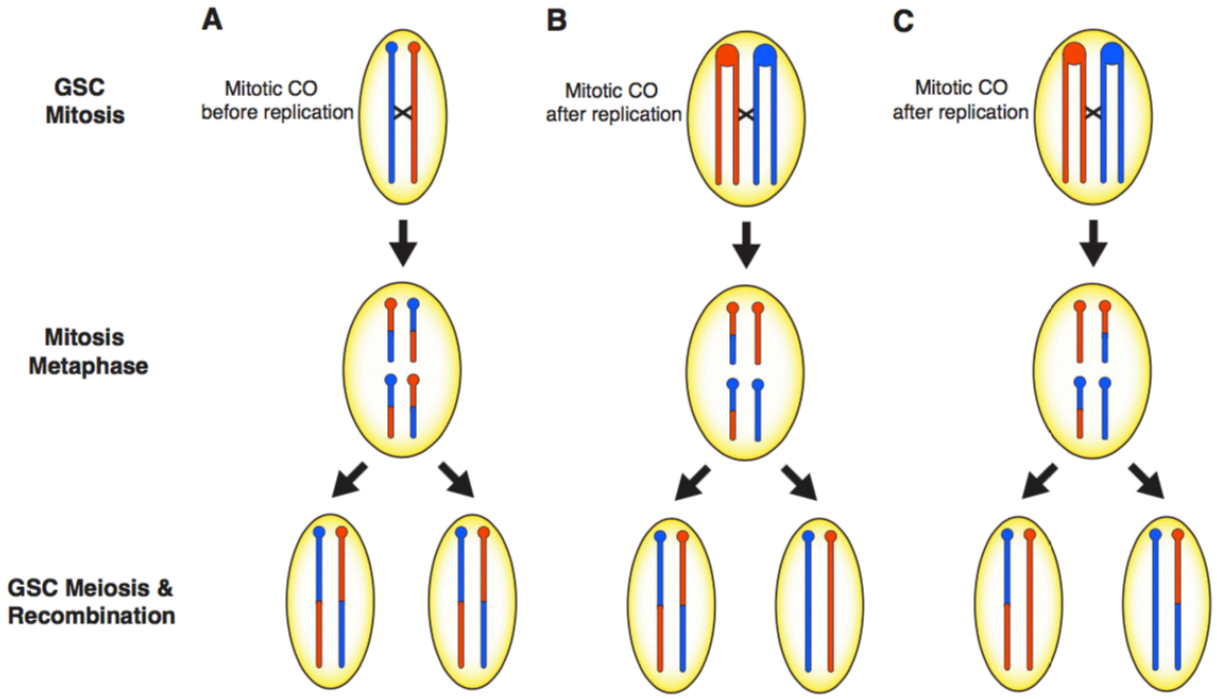


Figure 3.5: A model to explain the clusters of recombination on the third and X chromosomes in the progeny of two highly fecund dysgenic mothers. In the 5011 F1 female, a mitotic crossover (CO) either occured A) prior to DNA replication in the early developing germline resulting in two daughter cells with the CO or B) after DNA replication and segregating so that one daughter cell has both CO chromatids. Oocytes produced by these germline stem cells will transmit the CO and the reciprocal products. C) A mitotic CO in the 4029 F1 female occurred after DNA replication in the developing germline and each daughter cell received one CO chromatid and one non-crossover. This results in a loss of heterozygosity in the distal portion of the chromosome to resemble segregation distortion and recombination events are not detectable.



Chapter 4

The Recombination Landscape of *Drosophila yakuba* Reveals Dynamic Evolution of the Centromere Effect within the *Drosophila melanogaster* Species Subgroup

Abstract

Meiotic recombination is the exchange of DNA between homologous chromosomes during the first prophase division of meiosis. Recombination facilitates evolution while at the same time is subject to selection and can vary among individuals, populations, and species. The mechanisms of fine- and broad-scale control of recombination as well as how recombination evolves in species are still largely unknown. Reduced crossing over near the centromere, termed the centromere effect, is prevalent in eukaryotes and the strength of suppression can be variable in different species. The centromere effect in *Drosophila melanogaster* is considerably stronger than in its sister species, *D. simulans* and *D. mauritiana*. It is unknown if the strong centromere effect is ancestral to the clade or derived within *D. melanogaster*. The genetic map of *D. yakuba*, an outgroup species to the *melanogaster-simulans* clade, is largely unknown. I utilized low-coverage sequencing to genotype hundreds of progeny from a controlled cross to construct the first high-resolution genetic map of *D. yakuba* and determine the phylogenetic origin of the strong centromere effect observed in *D. melanogaster*. Strong suppression of recombination in the pericentric regions of both autosomes in *D. yakuba* suggests the strong centromere effect is ancestral to the *melanogaster-simulans* clade. The genetic map length of *D. yakuba* is longer than the genetic map of *D. melanogaster* with higher recombination rates on the X chromosome. The changes in both recombination rates and the centromere effect in the four species of *Drosophila* compared suggest dynamic evolution of recombination in the *melanogaster* subgroup.

Introduction

Recombination is one of the most rapidly changing aspects of meiosis (Coop and Przeworski 2007; Smukowski and Noor 2011; Ritz et al. 2017). Recombination is frequently heralded as one advantage of sexual reproduction because it decouples deleterious and beneficial alleles, thereby limiting how Hill-Robertson interference constrains adaptation (Hill and Robertson 1966). Recombination also limits the accumulation of deleterious mutations under Muller's Ratchet (Muller 1964). As such, recombination is an evolutionary force that is also subject to selection, albeit indirectly, through the genetic variation it produces (Felsenstein 1974). In theoretical studies recombination is favored in novel or rapidly changing environments with fluctuating optima in fitness (Charlesworth 1993; Burt 2000; Lenormand and Otto 2000). In addition to its evolutionary role, recombination (and specifically meiotic crossing over) plays a critical role in ensuring homologous chromosome segregation. The absence of crossovers typically results in aneuploid gametes and inviable progeny (Page and Hawley 2003) and, at least in mammals, the number of chromosome arms is positively associated with global recombination rates (Pardo-Manuel de Villena et al. 2001a). This is due to the fact that at least one crossover is typically needed per chromosome arm to ensure proper segregation. However, crossing over is not completely necessary as some species do not undergo recombination in the heterogametic sex, most noticeably in male *Drosophila* (White 1977).

Recombination occurs when native enzymes induce DNA damage in the form of double-stranded breaks (DSBs) (Cao et al. 1990; McKim and Hayashi-Hagihara 1998). These DSBs proceed to be repaired by either a crossover, involving exchange between homologous chromosomes, or by non-crossovers, where repair of a DSB is accomplished utilizing the homologous chromosome without exchange (reviewed in Gray and Cohen 2016). Thus, variation in the meiotic recombination rate is fundamentally determined by variation in DSB formation and the decision to crossover or not. What factors influence these processes in populations? *Cis*- and *trans*-acting elements (Baker et al. 1976), environment, demography (Wilfert et al.

2007), differences in chromosome structure (True et al. 1996; Pardo-Manuel de Villena et al. 2001a), genetic influences with local effects such as hotspots (Winckler et al. 2005), DNA sequence motifs (Comeron et al. 2012), and epigenetic effects (Borde et al. 2009) all contribute to variation in the recombination landscape within species. But the mechanism by which these factors cause difference in the rates of meiotic recombination between species is not well understood. Much of the difference in the recombination landscape between species of mice and between species of great apes can be attributed to *Prdm9*, a determinant of recombination hotspot evolution (Baudat et al. 2010; Parvanov et al. 2010). These results do not apply to other species, such as *D. melanogaster*, which have neither recombination hotspots nor *Prdm9* (Heil and Noor 2012). Many of the genes involved in recombination and meiosis are found to be rapidly evolving (Swanson and Vacquier 2002; Sawyer and Malik 2006; Anderson et al. 2009; Oliver et al. 2009). Moreover, differences in the recombination landscape within and across species can have a genetic basis (Chinnici 1971; Kidwell 1972; Charlesworth and Charlesworth 1985; Hunter et al. 2016; Brand et al. 2018). Species within the genus *Drosophila* provide ample opportunities to discover these genetic components controlling recombination and how their evolution affects variation of recombination in the absence of hotspots. Prior studies suffer from a lack of a standard in discussing differences in recombination in terms of scale and how they define divergence (Smukowski and Noor 2011). Advances in DNA sequencing technology and the growing number of sequenced genomes are making it possible to genotype at fine scales with diminishing cost. Evolutionary analysis of recombination needs to incorporate new and improving methods for accurate genotyping at finer scales to uncover local and genome-wide controls of recombination and explain observed variation among populations and species.

Reduced crossing over near the centromere, termed the centromere effect, is prevalent in eukaryotes (reviewed in Choo 1998; Talbert and Henikoff 2010). Crossover events near the centromere are often harmful and can result in nondisjunction and aneuploidy. This may be due to the centromeres' role in kinetochore attachment and proper segregation of chromosomes. True and colleagues (1996) compared

the genetic map of *D. mauritiana* to those of *D. melanogaster* and *D. simulans*. *D. melanogaster* diverged from the simulans clade, which includes *D. simulans*, *D.sechellia*, and *D. mauritiana*, around 3-4 million years ago (Obbard et al. 2012) (Figure 4.1). *D. mauritiana* has a genetic map nearly twice the length of *D. melanogaster* with a severely reduced centromere effect; *D. simulans* is intermediate between the two species in genetic map length and centromere effect (True et al. 1996). The centromere effect extends out further into euchromatin in *D. melanogaster* than the other sibling species of *Drosophila*. The mechanisms governing the intensity of centromere suppression of crossovers and how they can evolve over time has only been investigated recently (Brady et al. 2017; Hatkevich et al. 2017; Brand et al. 2018). This is one example of our lack of understanding of interspecific differences in recombination; if one crossover per arm is necessary for proper segregation in *D. melanogaster* then why does *D. mauritiana* experience nearly two crossovers per chromosome arm despite nearly identifical karyotypes? Are changes in the strength of the centromere effect associated with changes in genetic map length?

Changes in the centromere effect may arise from genetic conflict during female meiosis. Female meiosis typically results in only one in four meiotic products becoming the gamete pronucleus. The remaining products become polar bodies. A selfish allele distorting Mendelian segregation in meiosis to favor its own transmission into the single pronucleus produces female-specific meiotic drive. The selfish meiotic driver can fix within the population despite associated deleterious consequences such as increased nondisjunction, decreased fertility in males and sex-ratio distortion (Hamilton 1967; Zwick et al. 1999b; Henikoff and Malik 2002). Deleterious mutations hitchhiking with a driving allele can also impose a fitness cost on the host organism (Chevin and Hospital 2006). Genes unlinked to the driver that ameliorate meiotic drive will be favored by selection if meiotic drive imposes a cost (Sandler and Novitski 1957; Hartl 1975). Previous studies propose the centromere is one such meiotic driver and this may explain the rapid evolution of centromeres and centromere-related proteins (Malik and Henikoff 2001; Henikoff and Malik 2002; Fishman and Saunders 2008; Malik 2009; Chmátal et al. 2014). Centromere meiotic drive may also be responsible for differences in recombination rates observed in

males and females and rapid evolution of recombination near centromeres (Brandvain and Coop 2012). If a centromere is capable of distorting meiosis in its favor, a recombination modifier increasing rates of recombination near the centromere will be favored to uncouple it from the rest of the genome. This will decrease the strength of the driving centromere and slow its increase in frequency within a population. Recombination may also increase fitness by unlinking deleterious mutations increasing in frequency because they are linked to the meiotic driver. There may be a significant trade-off between limiting recombination near the centromere, to avoid the harmful effects of nondisjunction, and maintaining a sufficient recombination rate to limit the effects of driving centromeres linked to deleterious alleles. This trade-off could contribute to rapid fluctuation in recombination rates near the centromere.

How recombination rates change and evolve between species is largely unexplored. The most extensive comparisons of recombination rates between closely related species include mice (Dumont and Payseur 2011), great apes (Stevison et al. 2015), and the *D. pseudoobscura* subgroup (Stevison and Noor 2010; Smukowski Heil et al. 2015). There are few studies on recombination in species within the *melanogaster* subgroup despite the tremendous genomic, cytological, and other resources (True et al. 1996; Brand et al. 2018). I have elected to determine how changes in the centromere influence the total amount of recombination across the genome. True et al. (1996) first noted drastic differences between *D. melanogaster*, *D. simulans*, and *D. mauritiana* in genetic map length and strength of centromere suppression of recombination. Because *D. simulans* and *D. mauritiana* are sibling species to *D. melanogaster* (Figure 5.1), it is unclear if the strong centromere effect is ancestral to the *melanogaster-simulans* clade divergence or derived within *D. melanogaster*. To shed some light on the phylogenetic origin of the strong centromere effect within the melanogaster subgroup, I generated the first high-resolution genetic map for *D. yakuba*. *D. yakuba* is an outgroup species to the *D. simulans* and *D. melanogaster* frequently used to polarize comparisons between them (McDonald and Kreitman 1991). I found *D. yakuba* experiences higher rates of recombination than *D. melanogaster* while maintaining strong autosomal centromere suppression on recombination. A strong centromere effect in the outgroup

species suggests the strong centromere effect is ancestral to the divergence of the *D. melanogaster* and simulans species complex. The unique aspects of the recombination landscape in *D. yakuba* is evidence that the evolution of recombination rates is dynamic within the *melanogaster* subgroup and further exploration could provide valuable insights to how recombination evolves in a short evolutionary timespan.

Materials and Methods

Fly stocks and Crosses

Two *D. yakuba* strains labeled ST and Jess were previously inbred by full-sib mating in the Peter Andolfatto laboratory for 15 and 20 generations, respectively. The Andolfatto lab quantified polymorphism and divergence between the two strains and the *D. yakuba* reference strain, TAI18E2. There were no signs of inversions in either of the parental strains based on the polymorphism and divergence data (Appendix 22). Individuals heterozygous for an inversion experience reduced crossing over within the inverted region; minimizing inversions is necessary for the construction of a full genetic map. Approximately 20 3-day-old virgin females three days post-emergence of one strain and 20 younger males 2-10 days post-emergence of the other strain were crossed for three days. Each strain and all subsequent crosses were maintained on standard media at 25°C. Reciprocal crosses were performed to account for maternal effects on meiotic recombination. Individual virgin F1 females three days post-emergence were backcrossed to two or three 2-10 day post-emergence ST males in vials for two days to minimize the effect of maternal age on recombination. All F2 females from an F1 cross were collected upon emergence and immediately frozen at -20°C.

DNA Extraction and Library Preparation

DNA was extracted using the Quick-DNA (Zymo) extraction kit for 96-well plates following kit instructions. Prior to DNA extraction, flies were homogenized by 3.5 mm glass grinding balls (BioSpec) placed into a U-bottom polypropylene 96-well plate with lysis buffer from the Quick-DNA kit placed into a MiniBeadBeater-96 at 2,100 rpm for 45 seconds. DNA extraction yields varied from <0.1 ng/µl to 5 ng/µl. DNA was quantified using a Qubit fluorometer (Invitrogen). DNA quantification was not performed on the majority of samples but was assumed to average 1-2 ng/µl necessary for library preparation based on the measured samples.

Library preparations for 960 samples were performed using in-house produced Tn5 transposase (in collaboration with Stuart Macdonald and Brittny Smith) following the procedure outlined in Picelli et al. (2014) with a similar tagmentation protocol. Assuming an average of 1-2 ng/µl DNA concentration per sample in a 96-well plate, 1 µl of DNA was tagmented with the in-house Tn5 transposase at a concentration of 1.6 mg/ml with pre-annealed oligonucleotides in a 20 µl reaction volume for 55°C for 7 min and stopped by holding at 10°C. The reaction volume also contained 2 µl of 5X TAPS-DMF buffer (50 mM TAPS-NaOH, 25 mM MgCl2 (pH 8.5), 50% v/v DMF) and 2 µl of 5x TAPS-PEG buffer (50 mM TAPS-NaOH, 25 mM MgCl2 (pH 8.5), 60% v/v PEG 8000) for the desired DNA fragment lengths. The in-house Tn5 transposase was inactivated with an addition of 5 µl of 0.2% SDS and heating the total reaction to 55°C for 7 min. Only 2.5 µl of tagmentation reaction is needed for the PCR amplification with KAPA HiFi HotStart ReadyMix PCR Kit (Thermo Fisher Scientific), 1 µl of 4 µM Index 1 (i7) primers, and 1 µl of 4 µM Index 2 (i5) primers in 9 µl of reaction volume. The PCR occurred as follows: 3 min at 72°C, 2 min 45 sec at 98°C, and then 14 cycles of 98°C for 15 sec, 62°C for 30 sec, 72°C for 1 min 30 sec. The PCR-amplified samples were pooled and cleaned using 0.8 X vol AMPure XP Beads. I size-selected DNA fragments 250-400 bp on a BluePippin and cleaned using 1X vol of AMPure XP Beads.

Sequencing and Crossover Quantification

All libraries were sequenced at the University of Kansas Genomics Core. The first 192 samples were sequenced on an Illumina MiSeq Sequencer v3 with 150 bp single-end sequencing resulting in ~25 million reads. The remaining 768 samples were sequenced on an Illumina NextSeq550 with 150 bp single-end sequencing to produce around ~400 million reads. FASTQ files were parsed before using the multiplex shotgun genotyping (MSG) bioinformatic pipeline for identifying reliable genotype markers and determining ancestry at those markers using a Hidden Markov Model. The genomes of strain Jess and ST were assembled with FASTQ files supplied by Peter Andolfatto and aligned using BWA 0.7.15 mem algorithm (Li and Durbin 2009) and Samtools 1.4.1 (Li 2011). The reads were aligned to the *D. yakuba* reference genome, TAI18E2 release 1.06 (Gramates et al. 2017a), except the 2R chromosome arm was replaced with the sequence from isofemale strain NY73 collected, inbred, and sequenced by the Andolfatto lab (Rogers et al. 2015). The 2R chromosome arm was replaced because the arm in the reference genome contains an inversion, *2Rn*, which covers 40% of the length of the chromosome arm (Lemeunier and Ashburner 1976). The MSG pipeline mapped reads to each parental genome based on SNP differences between the two genomes. Recombination breakpoints were detected based on changes in genotyping, whether reads mapped to both genotypes at the same site (heterozygous) or just one genotype (homozygous). Samples with a minimum of 10,000 reads provided sufficient read coverage to make reliably genotype calls with high confidence in the MSG pipeline; all samples with fewer reads were discarded. Crossover events identified by this pipeline were manually curated for errors. Double crossovers less than 750 kb apart were discarded as these events are unlikely to occur within 1 Mb and most were due to mapping errors. Due to low quality sequences leading to erroneous mapping calls on the ends of the chromosome arms as follows: for the X, 2L, 2R, 3L, and 3R chromosome arms, crossovers were removed within 650 kb, 700 kb, 430 kb, 3.3 Mb, and 400 kb of the telomere ends and within 3.3 Mb, 10 Mb, 8 Mb, 2.8 Mb, and 7.5 Mb of the centromere ends, respectively.

Data analysis

Crossover data was parsed and analyzed within the R Version 3.4.2 programming environment (R Core Team 2017). The following packages were also used in genetic map construction, model testing, and general data analysis: R/qtl (Broman et al. 2003), lme4 (Bates et al. 2015), and Biostrings (Pagès et al. 2017). Figures were produced using ggplot2 (Wickham 2016). Maker (Cantarel et al. 2008) was used to annotate the genome assemblies of strains ST and Jess with the GFF file for the latest *D. yakuba* genome (1.5) downloaded from Flybase (Gramates et al. 2017). The genome was annotated for TEs and simple motifs with Repeatmasker (Tarailo-Graovac and Chen 2009) using the current catalog of TE sequences in *D. yakuba* from Repbase (Bao et al. 2015). The Classic Weinstein method was used to calculate frequency of non-exchange tetrads (E_0), single exchange tetrads (E_1), and multiple-exchange tetrads (E_n) produced during meiosis in *D. yakuba*. The following equations from Weinstein (1918) were used to calculate tetrad frequencies from the frequency of chromosomes experiencing zero crossovers (a_0), one crossover (a_1), two crossovers (a_2), and three or greater crossovers (a_3):

$$E_0 = a_0 - a_1 + a_2 - a_3$$

$$E_1 = 2a_1 - 4a_2 + 6a_3$$

$$E_2 = 4a_2 - 12a_3$$

$$E_3 = 8a_3$$

Results

Crosses and Genotyping

In total, 478 F2 female flies produced by 50 F1 mothers were successfully sequenced to quantify crossovers. The MSG pipeline identified a total of 974,871 high-quality SNPs between the two parental genomes with an average distance of 136 bp and a median distance of 65 bp between SNPs. The median interval for identification of a crossover was 6990 bp and 95.3% of all crossovers were localized within a

100 kb interval. Only six crossovers were localized within intervals greater than 1 Mb but these crossovers were in individuals with a number of reads close to the cutoff for reliable detection of recombination events (10,000-20,000 reads). There were 3,107 genetic markers shared among all samples. However, there were no common genetic markers on the fourth chromosome and relatively few high-quality SNPs identified compared to the other chromosomes (1,994). Just as in *D. melanogaster*, no crossovers were detected on the fourth chromosome of *D. yakuba* after close inspection. Hypotheses for the lack of recombination on the fourth chromosome, commonly referred as the dot chromosome, include its small size, heterochromatic structure, and repetitive sequences (reviewed in Hartmann and Sekelsky 2017). Thus, the fourth chromosome will be ignored when referring to the autosomes for the remainder of the study. Prior to grouping all recombination breakpoints to create a genetic map, I constructed a mixed-effects likelihood model to test for differences in total crossover number between direction of the parental cross and brood. I used a Poisson link model with F1 mother as a random variable. In the full model, I detected no differences between reciprocal crosses or broods (Wald χ^2 test $p = 0.89, 0.97$ respectively). This allowed crossovers quantified in all samples to be grouped together to create a single genetic map for *D. yakuba*.

High-Resolution Genetic Map of D. yakuba

The only previous study of recombination in *D. yakuba* measured recombination between seven pairs of genetic markers with all but one pair on the X chromosome (Takano-Shimizu 2001). In contrast, my study utilized 974,871 genetic makers with 3,107 genetic markers shared amongst all samples across every chromosome minus the fourth chromosome. The total genetic map length of *D. yakuba* in this study is 353 cM with an average recombination rate of 2.7 cM/Mb. The X chromosome experiences higher rates of recombination on average than the autosome chromosome arms even though it is the second smallest in terms of physical length. There are approximately three crossovers detected on average in a single F2 fly or roughly one per chromosome. Among all samples, crossover number ranged from zero to eight in a

single F2 fly. The most crossovers observed on a single chromosome arm was four on a single X chromosome and several X chromosomes experienced three crossover events. For the autosomes, three crossovers on a single arm is rare in close resemblance to crossover counts in *D. melanogaster* (Figure 4.2). The autosome arms are more likely to have zero crossovers while the X chromosome is overrepresented in higher crossover counts (Figure 4.2A). When crossover counts are grouped by entire chromosome instead, the proportion of crossovers each chromosome contributes is more evenly distributed between the X chromosome and the autosomes (Figure 4.2B). This may indicate a single crossover per chromosome rather than per chromosome arm is necessary for proper meiotic function in *D. yakuba*.

It is not possible to observe the total number of crossovers generated in a single bivalent during meiosis as only one chromatid is transmitted as a gamete pronucleus. Tetrad analysis uses the number of crossovers for each chromosome to calculate the frequency of non-exchange tetrads (E_0), single-exchange tetrads (E_1), or multiple-exchange tetrads (E_n) generated during meiosis (Weinstein 1918). I used the Classic Weinstein method to calculate tetrad frequencies for zero to triple-exchange tetrads for each chromosome arm in our *D. yakuba* data (Table 4.1). In *D. melanogaster*, the frequencies for E_0, E_1, and E_2 are 5-10%, 60-70%, and 30-35% respectively (Hughes et al. 2018). The E_0 tetrad frequency for the 3L chromosome is estimated to be negative and therefore biologically meaningless; negative tetrad frequencies are a drawback of the Classic Weinstein method (Zwick et al. 1999a). The E_0 tetrad frequency for the other chromosomes ranged from 5% to over 20% in *D. yakuba*, similar to estimates of non-exchange tetrad frequencies in *D. melanogaster* (Table 4.1). The frequencies of E_1 and E_2 tetrads for the autosomes are variable within *D. yakuba* ranging between 30% and 70%.

Crossover interference reduces the probability of an additional crossover in proximity to other crossovers. I quantified interference in *D. yakuba* using the Housworth-Stahl model to calculate *nu*, a unitless measure of interference, with a maximum likelihood function based on intercrossover distances

(Housworth and Stahl 2003). If COs are not subject to interference, intercrossover distances are Poisson distributed and the collection of distances from many crosses resembles a Poisson distribution with a *nu* = 1; *nu* > 1 is indicative of positive interference (Broman and Weber 2000). Each chromosome in *D. yakuba* has detectable interference between crossovers as ranges from *nu* = 2.5 to ~7 (Table 4.2). The chromosomes with the strongest interference or the highest values of *nu* also have the strongest centromere effects (reviewed later). The Houseworth-Stahl model also calculates the percentage of crossovers produced through an alternative pathway not subject to interference as the escape parameter *P* (de los Santos et al. 2003). Around 4-5% of crossovers are estimated to be produced through the alternative pathway when detected (Table 4.2). In contrast to interference, crossover assurance is the recombination control mechanism to ensure the minimal number of crossovers on each chromosome for proper chromosome segregation during meiosis. In the absence of crossover assurance and interference, the distribution of crossovers within a collection of progeny should resemble a Poisson distribution in that the variance in crossover number is equal to the mean. The crossover mean and variance are not equal (3.2 and 1.9, respectively) and the distribution of crossovers is significantly different from a Poisson distribution (χ^2 test p = 7.92E-10). *D. yakuba* has a higher than expected frequency of individuals with CO numbers close to the mean (2-4) with a lower than expected frequency of crossover numbers closer to the outliers, 0-1 on the low end and 6-8 on the high end (Appendix 23). The collective action of crossover assurance and interference observed in *D. yakuba* ensures an appropriate number of crossovers on average for proper meiotic function.

Evidence for a Strong Centromere Effect in D. yakuba

The distribution of recombination in *D. yakuba* indicates a strong centromere effect on both autosomes (Figure 4.3). The X chromosome shows the highest rates of recombination among any of the chromosomes and a much higher amount of recombination near the centromere (Figure 4.3A). Chromosomes 2 and 3 have large regions of zero recombination in the proximal portion of their

chromosomes (Figure 4.3B-E). The majority of crossovers occur near the telomeres on the autosomes rather than the middle of the chromosome similar to the distribution of crossovers in *D. melanogaster* (True et al. 1996; Fiston-Lavier et al. 2010). Large segments of the genome with no detectable recombination influenced the correlations between recombination and genomic sequence features. Recombination rates can correlate with sequence features including GC content, motifs, and nucleotide polymorphism (Begun and Aquadro 1992; Kong et al. 2002; Comeron et al. 2012). In *D. yakuba*, recombination shows no correlations with GC content except for the 3L chromosome (Table 4.3). Recombination is significantly correlated with gene density, simple motifs, and TE density when all of the data is summed together. However, the correlations appear weak as not all chromosomes show a significant correlation with the sequence parameters. Only SNP density is strongly correlated with recombination rates for all chromosomes. The weak correlations may be due to the exclusion of large segments of the genome with no recombination detected in our study or the small number of crossover events in our data. When non-recombining regions are included in the correlation calculations, many of the parameters are significantly correlated (Appendix 24). The correlations are inflated due to a large number of intervals within the genome with zero recombination.

In comparison to other species in the *melanogaster* subgroup, the genetic map length of *D. yakuba* most resembles the genetic map length of *D. simulans* (Table 4.4). A previous study of recombination in *D. yakuba* found the genetic map of the X chromosome and one far distal interval on the 2L chromosome to be ~1.5x the size of the *D. melanogaster* genetic map (Takano-Shimizu 2001). I found this to be consistent for my high-resolution genetic map of *D. yakuba* as the genetic map length length of the X chromosome is 1.4x longer than the X chromosome in *D. melanogaster* (Table 4.4). The distal 2.5 MB portion of the 2L chromosome also experiences about 1.5x more recombination in *D. yakuba* with ~6 cM/MB to compared to ~4 cM/MB in *D. melanogaster* (Figure 4.3B) (Fiston-Lavier et al. 2010; Comeron et al. 2012). There is more recombination on the X and 3rd chromosomes in *D. yakuba* in comparison to *D. melanogaster* resulting in a longer total genetic map length. It appears *D. melanogaster* has a

drastically reduced genetic map for the third chromosome in comparison to all other species (Table 4.4). However, the distribution of recombination in *D. yakuba* more closely resembles the *D. melanogaster* distribution (True et al. 1996; Fiston-Lavier et al. 2010). In both species large portions of the autosome arms experience almost no recombination and crossovers are localized to the distal ends of the chromosomes. Recombination frequently occurs in the middle of the chromosome arms in *D. mauritiana* and *D. simulans* with weak suppression of recombination near the centromere (True et al. 1996).

To quantify the strength of the centromere effect on recombination, I utilized the *CE* metric from Hatkevich et al. (2017) with some modifications. *CE* is analogous to interference as it is equal to 1 – (*observed* / *expected*) where *observed* is equal to the number of crossovers within an interval near the centromere. The *expected* value is the number of crossovers expected to fall within the same interval if crossovers are uniformly distributed along the length of the chromosome (Hatkevich et al. 2017). When *CE* = 0, there is no suppression of recombination near the centromere and *CE* = 1 is complete suppression of recombination. To compare between different species and datasets, I defined the pericentric interval as the physical proximal third of each chromosome. The *observed* is the genetic map length of pericentric region of the chromosome and the *expected* is equal to one third of the total genetic map length of the chromosome assuming recombination rates are uniform across the entire length of the chromosome. I used previous studies for genetic map and physical chromosome length data for *D. melanogaster* (Gramates et al. 2017), *D. simulans* (True et al. 1996; Barker and Moth 2001), and *D. mauritiana* (True et al. 1996). The centromere effect of the X chromosome is strongest in *D. melanogaster* and very weak or nonexistent in the other species (Table 4.5). *CE* values on the autosomes are high in both *D. melanogaster* and *D. yakuba* but very weak in *D. mauritiana*; *D. simulans* is intermediate.

Because I did not quantify crossovers in the other species of *Drosophila*, it is difficult to make direct comparisons in *CE*. However, it is possible to use the genetic map length of the other species to calculate the expected number and distribution of crossovers to occur for a given number of meiotic events. I

compared the observed ratio of pericentric to distal crossovers in *D. yakuba* to the expected ratio of pericentric crossovers to distal crossovers according to genetic maps in other species using a Chi-Squared Test. The expected number of crossovers in the pericentric region based on the genetic maps of the other species is equal to the number of total crossovers observed on a single chromosome in *D. yakuba* multiplied by the ratio of pericentric genetic map length to total genetic map length. The number of distal crossovers is the number of crossovers remaining or the total crossover number multipled by the ratio of distal map length to total map length. The X chromosome *CE* is significantly stronger in *D. melanogaster* than *D. yakuba* but not different in comparison to *D. simulans* and *D. mauritiana* (Table 4.6). The *CE* of the second chromosome in *D. yakuba* is stronger than all other species including *D. melanogaster* where the centromere effect was first noted. The third chromosome also has a larger *CE* in *D. yakuba* compared to *D. melanogaster*, although this is marginally non-significant when accounting for multiple tests (χ^2 test p = 0.00557, 0.0501 after Bonferroni correction). The signal is derived from the significantly stronger *CE* on the right arm of the third chromosome in *D. yakuba*, the left arm shows no difference in centromere effect between the two species (χ^2 test p = 1.26E-5, 0.43 for 3L, 3R respectively).

There are numerous differences in both frequency and distribution of recombination in the *melanogaster* subgroup observed so far. Assuming parsimony, there are still several changes to recombination over the course of evolution in these four species (Figure 4.5). The strong centromere effect on the autosomes present in *D. yakuba* suggests it is ancestral to the *melanogaster-simulans* clade divergence. Changes in the distribution of recombination are expected to accompany changes to the frequency of recombination (Zhang et al. 2014). Thus, increases in recombination rate should accompany decreases in centromere effect as interference forces crossovers to separate out more evenly along the length of a chromosome. True et al. (1996) originally proposed this to explain the differences in distribution of recombination in *D. melanogaster* and *D. mauritiana*. However, this is not always the case upon further examination with additional data from this study. *D. yakuba* has a much stronger centromere effect on the second chromosome than *D. simulans* (χ^2 test p = 1.59E-39) despite a nearly identical genetic map length (Table

4.4). Similarly, the genetic map length of the 3L chromosome arm in *D. yakuba* is ~1.6x longer than the same chromosome arm in *D. melanogaster* but the centromere strength is not significantly different between *D. melanogaster* and *D. yakuba* (χ^2 test $p = 0.42$).

Discussion

Strong Autosome Centromere Effect Appears Ancestral to melanogaster-simulans Divergence

The recombination landscape of *D. melanogaster* has been studied for over 100 years and continues to be a landmark in recombination studies today. However, *D. melanogaster* has the smallest genetic map in the genus *Drosophila* at ~280 cM (Gramates et al. 2017). The genetic map of *D. pseudoobscura* is twice the size of *D. melanogaster* (~450 cM) (Anderson 1993) and *D. virilis* has a genetic map four times the size (~950 cM) (Gubenko and Evgen'ev 1984). Within the *melanogaster* subgroup, the genetic map length of *D. mauritiana* is ~1.7x the size of the genetic map of *D. melanogaster* despite both lineages diverging less than 3 million years ago (True et al. 1996). The strong centromere effect observed in *D. melanogaster* was thought to be unique to this species. This study demonstrates for the first time another species of *Drosophila* with a strong centromere effect. Because *D. yakuba* is an outgroup to the *melanogaster-simulans* clade (Figure 4.1), the strong centromere effect in *D. yakuba* suggests that this phenomenon is ancestral to the *melanogaster-simulans* divergence (Figure 4.5). The centromere effect subsequently became weaker within the *simulans* clade accompanied by higher rates of recombination. However, observing the strong centromere effect in a single outgroup is not conclusive evidence for inferring the ancestral state within the *melanogaster* subgroup. The strong centromere effect may have arisen twice, once each in *D. melanogaster* and *D. yakuba*. Roughly 8-10 million years of evolution separate the two species and recombination rates may evolve more rapidly than the timescale for speciation within *Drosophila*. For example, a recent study suggests the evolution of a single gene is responsible for higher

rates of recombination and the weak centromere effect in *D. mauritiana* (Brand et al. 2018). The increase of *CE* on the X chromosome in *D. melanogaster*, stronger 2nd chromosome *CE* in *D. yakuba*, and increase in recombination rates in *D. mauritiana* also suggest the evolution of recombination and the centromere effect is more labile than previously considered. More recombination studies in additional species of *Drosophila* within the *melanogaster* subgroup will be needed in order to definitively locate the origin of the strong centromere effect.

One major assumption of this study is that I can compare recombination rates between different species of *Drosophila* while ignoring possible impacts of genome and chromosome evolution. Differences in karyotype can influence recombination rates because crossovers ensure proper disjunction during meiosis (Mather 1938). For example, the number of chromosome arms in mammals is a strong predictor for crossover rates (Pardo-Manuel de Villena et al. 2001b; Segura et al. 2013; Dumont 2017). There are no differences in karyotype of species within the *melanogaster* subgroup and the polytene chromosomes structure is extremely similar between species (Lemeunier and Ashburner 1976; Schaeffer et al. 2008). It is currently unknown if there are differences in the amount of heterochromatin, specifically in the pericentric regions, between the species. Differences in the amount of heterochromatin could lead to changes in the distribution and frequency of crossover events. The large amount of heterochromatin on the X chromosome is thought to be responsible for its reduced centromere effect in comparison to the autosomes in *D. melanogaster* (Yamamoto and Miklos 1978). Satellite DNA within heterochromatic regions is also associated with reduced recombination rates. While the composition of satellite DNA is rapidly diverging among species in the *melanogaster* subgroup (Jagannathan et al. 2017), the proportion of the genome comprised of satellite DNA is similar between species (Bosco et al. 2007). It is currently unknown if changes in the heterochromatin landscape and satellite DNA have an impact on recombination rates in the pericentric regions within the *melanogaster* subgroup. The suppression of recombination is also an intrinsic property unique to centromeres; the range of the centromere effect is maintained during the removal of pericentric heterochromatin in *D. melanogaster* (Yamamoto and Miklos

1978). The exact mechanism of the suppression of recombination by the centromere is currently unknown. The effect may be the consequence of the kinetochore attachment and subsequent blocking of DSB formation (Vincenten et al. 2015) but the relationship remains unclear.

The Relationship Between Recombination and Genome Evolution

All of the sequence parameters with the exception of SNP diversity were weakly correlated or uncorrelated with recombination rates in *D. yakuba* in this study (Table 4.3). Only recombining regions of chromosomes were included in the calculation of correlation coefficients. Large proportions of each chromosome with zero recombination inflated correlation coefficients and their significance (Appendix 24). Most of the non-recombining regions of the genome were the result of the strong centromere effect (Figure 4.3). Other non-recombining intervals within the genome are likely the result of the low number of individuals and fewer recombination events characterized in the study compared to other recombination studies. More samples will be needed to test the strength of correlations between recombination rates and different sequence parameters. The regions of the genome with reduced recombination rates due to the centromere effect should also experience a reduction in nucleotide diversity (Begun and Aquadro 1992). Nucleotide diversity is noticeably reduced in pericentric regions in *D. melanogaster* in comparison to *D. simulans* (Barghi et al. 2017). Differences in recombination influences the efficacy of natural selection within an organism and thus influences the molecular evolution of genes located in the pericentric regions of different species of *Drosophila* (Hill and Robertson 1966).

Inversions can also influence genome evolution through the suppression of recombination within the boundaries of the inversion (Hoffmann and Rieseberg 2008; Kirkpatrick 2010). *D. yakuba* is highly polymorphic in inversions with nine common inversions segregating in different populations, all of which on the autosomes (Lemeunier and Ashburner 1976). The *D. yakuba* reference genome is derived from an

isofemale line fixed for the *2Rn* inversion covering 40% of the right arm of the second chromosome (Llopart et al. 2002). The segregating inversions affect heterozygosity and natural selection in *D. yakuba* population sequences (Rogers et al. 2015). The influence of segregating inversions may be one reason for weak correlations between recombination rate and different sequence parameters in *D. yakuba* (Table 4.3). The two strains of *D. yakuba* in this study were picked because there was little evidence for fixed differences in polymorphic inversions based on genomic divergence between the two strains (Appendix 22). Segregating inversions could influence recombination rates within populations of *D. yakuba*. Additional recombination experiments in different genetic backgrounds, with and without inversions, are needed to investigate the effects of polymorphic inversions on recombination landscapes in *D. yabuba*.

Recombination and Centromere Meiotic Drive

Meiotic drive is the non-Mendelian transmission of a gene or chromosome through the manipulation of meiosis to favor its own transmission even at the cost of the organism. This is accomplished via the asymmetric transmission of an allele to the single pronucleus product of female meiosis. Centromeres have long been hypothesized to be meiotic drivers, explaining the rapid evolution of centromere sequences and positive selection detected in centromere-interacting proteins (Malik 2009). Theoretical simulations support rapid evolution of recombination rates near centromeres as a result of centromere-mediated meiotic drive (Brandvain and Coop 2012).

I find that the recombination rates near the centromere in four species of *Drosophila* within the *melanogaster* subgroup are changing between species in a chromosome-specific manner as predicted (Figure 4.5). However, the study provides no direct evidence to support the link between centromere drive and changing recombination rates. One way to provide direct evidence for the effect of meiotic drive on recombination is to identify recombination modifiers in the species studied. One recombination modifier

was recently proposed to be responsible for the higher recombination rates and weak centromere effect in *D. mauritiana* (Brand et al. 2018) but our knowledge on the mechanisms of recombination modification in *Drosophila* is still in its infancy.

Centromeres are complex chromosome structures comprised of centromere-specific heterochromatin and satellite DNA that must interact faithfully with kinetochores and microtubules during meiosis and mitosis. There may be any number of other factors involved in determining recombination rates in pericentric regions of the chromosome. Centromere-interacting proteins, recombination machinery, and other meiotic genes evolve rapidly as well (Malik et al. 2002; Anderson et al. 2009; Myers et al. 2010; Hemmer and Blumenstiel 2016). The evolution of meiosis and recombination genes may account for the rapid evolution of recombination rates. Once again, many meiosis genes are hypothesized to be evolving rapidly due to genetic conflict including meiotic drive. Further work to identify selfish genetic elements is necessary to identify genetic conflict as a determinant in recombination rate evolution.

Tables

Table 4.1: Exchange tetrad frequencies calculated using the Classic Weinstein method for each chromosome arm in *D. yakuba*.

Chromosome	**Exchange Tetrad**			
	E_0	E_1	E_2	E_3
X	0.040	0.365	0.260	0.335
2L	0.136	0.704	0.143	0.017
2R	0.241	0.449	0.294	0.017
3L	0.119	0.327	0.503	0.050
3R	-0.023	0.528	0.444	0.050

Table 4.2: Interference values (*nu*) and frequency of crossovers created in the non-interference pathway (*P*) in *D. yakuba* as estimated with the Housworth-Stahl model. Values of *nu* > than 1 signify positive interference.

Variable	**Chromosome**				
	X	2L	2R	3L	3R
nu	2.55	6.96	6.90	3.83	6.56
P	0.00	0.042	0.043	0.00	0.052

Table 4.3: Correlations between sequence parameters and recombination without non-recombining regions in *D. yakuba*. Pearson's correlation coefficients (R) and significance (*p*-value) are listed. Measurements were recorded in 500 kb intervals along each chromosome in *D. yakuba*. Significant values are bolded.

Sequence Parameter		Chromosome					Total
		X	2	3	4	5	
GC Content	R	-0.24	0.12	0.055	**0.48**	0.25	0.058
	p	0.150	0.530	0.777	**1.81E-03**	0.145	0.498
Gene density	R	0.15	**0.56**	0.12	0.28	**0.35**	**0.15**
	p	0.392	**8.69E-04**	0.522	0.083	**0.037**	**0.010**
Simple repeats	R	**0.32**	0.13	0.26	**0.42**	0.13	**0.37**
	p	**3.77E-04**	0.468	0.171	**8.06E-03**	0.457	**3.27E-07**
SNP Density	R	**0.40**	**0.50**	**0.53**	**0.49**	**0.70**	**0.63**
	p	**3.57E-03**	**3.57E-03**	**2.92E-03**	**1.59E-03**	**1.94E-06**	**2.79E-08**
TE Density	R	-0.0024	-0.25	-0.21	**-0.28**	-0.034	**-0.34**
	p	0.989	0.167	0.147	**9.68E-03**	0.42	**1.24E-02**

Table 4.4: The genetic map lengths different species of *Drosophila* in the *melanogaster* subgroup in centiMorgans (cM). Genetic map data for species other than *D. yakuba* (this study) were collected from the following sources: *D. mauritiana* from True et al. (1996), *D. simulans* from Barker and Moth (2001), and *D. melanogaster* from Flybase (Grametes et al. 2017). Genetic map lengths in *D. simulans* are listed for the entire chromosome as reported in Barker and Moth (2001).

Species	**Chromosome Arm**					
	X	2L	2R	3L	3R	Total
D. mauritana	111	69	71	95	116	462
D. simulans	75	144		157		376
D. melanogaster	66	55	52	47	56	276
D. yakuba	94	53	55	73	78	353

Table 4.5: The centromere strength (*CE*) of different species of *Drosophila* in the *melanogaster* subgroup. Higher values indicate stronger centromere suppression of recombination. Genetic map data for species other than *D. yakuba* (this study) were collected from the following sources: *D. mauritiana* from True et al. (1996), *D. simulans* chromosomes X and 3 from True et al. (1996), *D. simulans* chromosome 2 from Barker and Moth (2001), and *D. melanogaster* from Flybase (Grametes et al. 2017). The genetic map length of the entire second chromosome in *D. simulans* was used to calculate *CE* listed as reported in Barker and Moth (2001).

Species	Chromosome Arm				
	X	2L	2R	3L	3R
D. mauritiana	-0.01	0.04	0.01	0.24	0.35
D. simulans	0.05	0.22		0.42	0.58
D. melanogaster	0.36	0.78	0.82	0.87	0.83
D. yakuba	0.09	0.99	0.99	0.80	0.99

Table 4.6: Comparisons between the observed ratio of pericentric crossovers to distal crossovers in *D. yakuba* to the expected ratio based on the genetic maps of other species. The expected numbers are calculated as follows: Pericentric = Total Crossover Number in *D. yakuba* × (Pericentric Genetic Map Length ÷ Total Genetic Map Length), Remainder = Total Crossover Number in *D. yakuba* × (Distal Genetic Map Length ÷ Total Genetic Map Length). Observed and expected ratios are compared with a Chi-Squared Test; *p*-values are listed and significant *p*-values are bolded.

Comparison	Chromosome					
	X		2		3	
	Pericentric	Remainder	Pericentric	Remainder	Pericentric	Remainder
D. yakuba observed	130	308	2	505	19	707
D. melanogaster expected	93	345	33	474	35	691
	***p* = 7.24E-07**		***p* = 2.39E-08**		***p* = 5.57E-03**	
D. yakuba observed	130	308	2	505	19	707
D. simulans expected	138	300	132	376	175	551
	p = 0.419		***p* = 1.59E-39**		***p* = 9.56E-42**	
D. yakuba observed	130	308	2	505	19	707
D. mauritiana expected	148	290	164	343	196	530
	p = 0.103		***p* = 2.23E-53**		***p* = 1.53E-49**	

Figures

Figure 4.1: The phylogeny of the *Drosophila melanogaster* subgroup with divergence times (MYA). Phylogeny and divergence times recreated from Ko et al. (2003) and Obbard et al. (2012).

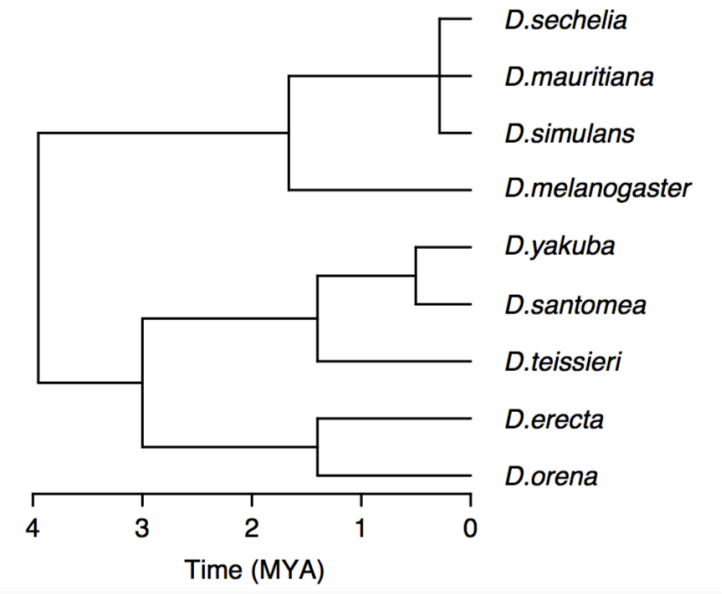


Figure 4.2: The proportion of chromosome arms (A) or whole chromosomes (B) grouped by crossover (CO) count in F2 progeny of *D. yakuba*. 95% confidence intervals were calculated by sampling F2 progeny for bootstrapping 1000 times.

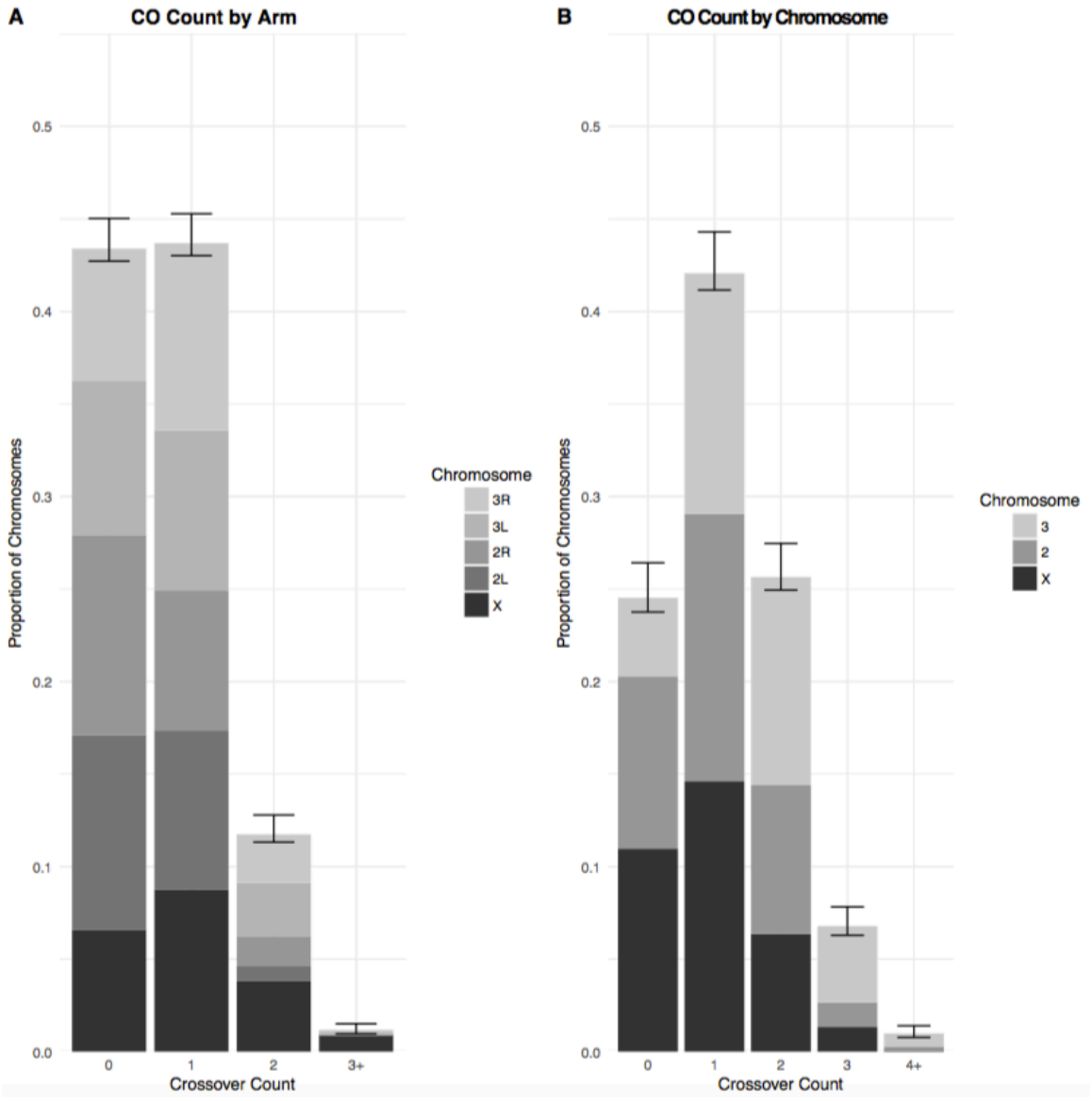


Figure 4.3: Loess smoothed splines of the recombination rate along the length of each chromosome in D. yakuba with standard error. The dotted line separates the proximal third of each chromosome from the remainder region used to calculate the strength of the centromere effect. The rate of recombination was calculated in 500 KB intervals in F2 progeny for the A) X chromosome, B) 2L chromosome arm, C) 2R chromosome arm, D) 3L chromosome arm, and E) 3R chromosome arm.

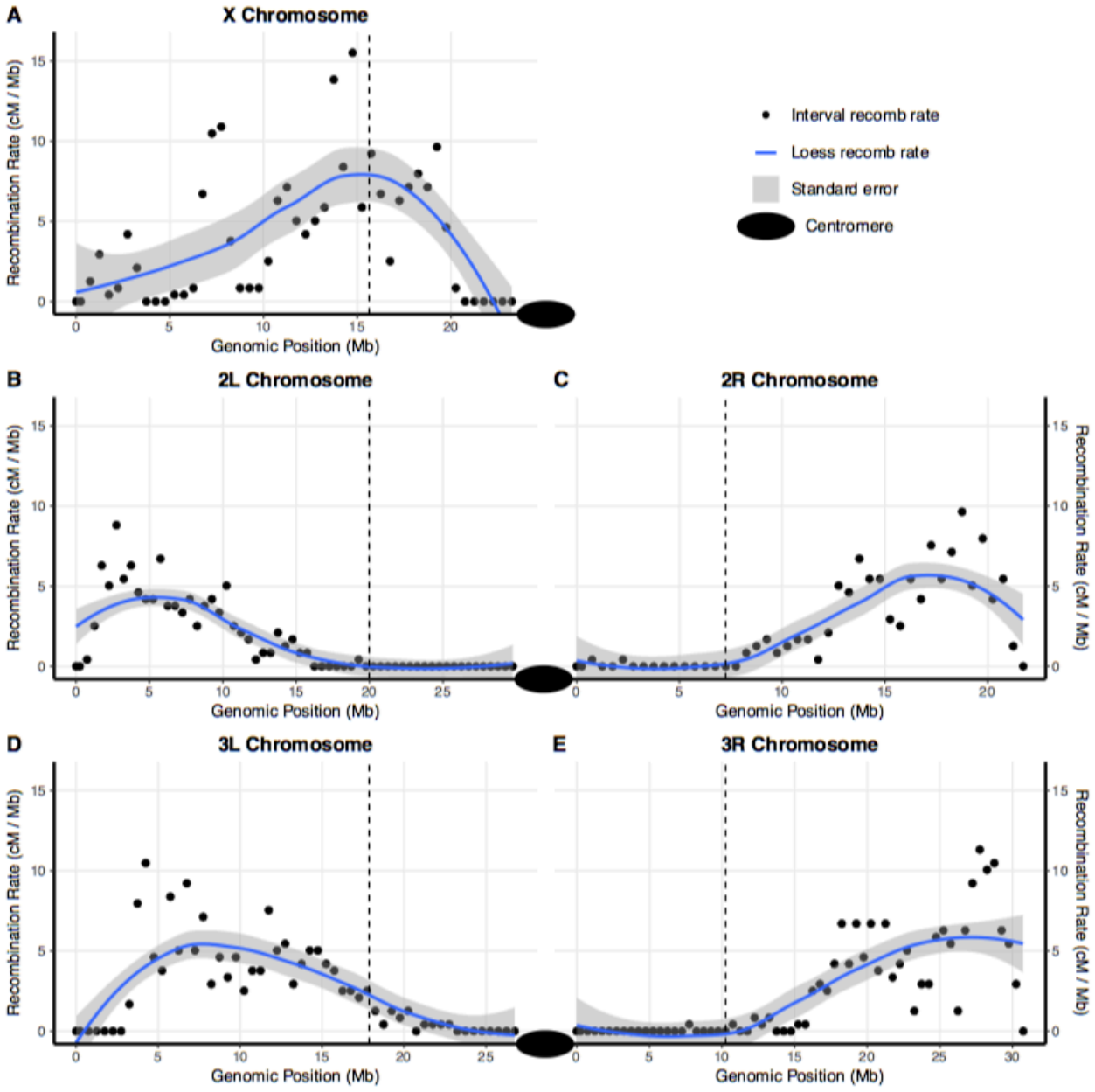
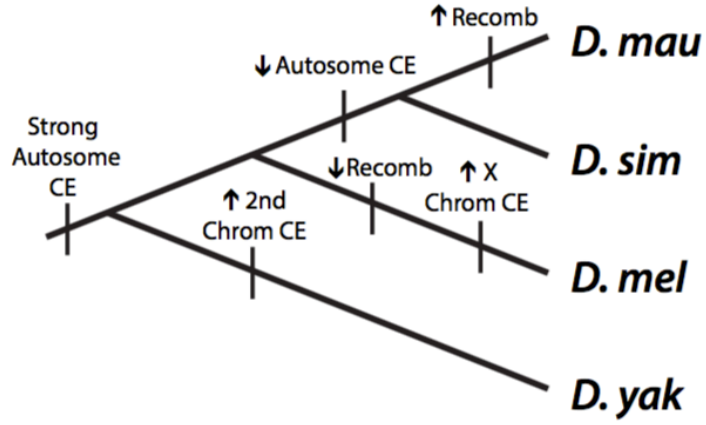


Figure 4.4: The specific changes in recombination rates and centromere effect mapped to a cladogram of four species of *Drosophila* in the *melanogaster* subgroup. The proposed changes are mapped based on parsimony principle.



References

Adrian AB, Corchado JC, Comeron JM. 2016. Predictive models of recombination rate variation across the *Drosophila melanogaster* genome. Genome Biol Evol 8: 2597–2612.

Adrion JR, Song MJ, Schrider DR, Hahn MW, Schaack S. 2017. Genome-wide estimates of transposable element insertion and deletion rates in *Drosophila melanogaster*. Genome Biol Evol 9: 1329–1340.

Aguinaldo AMA, Turbeville JM, Linford LS, Rivera MC, Garey JR, Raff RA, Lake JA. 1997. Evidence for a clade of nematodes, arthropods and other moulting animals. Nature 387: 489-493.

Anderson WW. 1993. Linkage map of *Drosophila pseudoobscura*. Genetic Maps: Locus Maps of Complex Genomes 252-253. Cold Spring Harbor Laboratory Press, Cold Spring Harbor NY.

Anderson JA, Gilliland WD, Langley CH. 2009. Molecular population genetics and evolution of *Drosophila* meiosis genes. Genetics 181: 177-185.

Anderson LK, Royer SM, Page SL, McKim KS, Lai A, Lilly MA, Hawley RS. 2005. Juxtaposition of C(2)M and the transverse filament protein C(3)G within the central region of *Drosophila* synaptonemal complex. Proc Natl Acad Sci USA 102: 4482-4487.

Andolfatto P. 2001. Contrasting patterns of X-linked and autosomal nucleotide variation in *Drosophila melanogaster* and *Drosophila simulans*. Mol Biol Evol 18: 279-290.

Andolfatto P, Davison D, Erezyilmaz D, Hu TT, Mast J, Sunayama-Morita T, Stern DL. 2011. Multiplexed shotgun genotyping for rapid and efficient genetic mapping. Genome Res 21: 610–617.

Aravin AA, Hannon GJ, Brennecke J. 2007. The Piwi-piRNA pathway provides an adaptive defense in the transposon arms race. Science 318: 761–764.

Baker BS, Carpenter AT, Esposito MS, Esposito RE, Sandler L. 1976. The genetic control of meiosis. Annu Rev Genet 10: 53–134.

Bao W, Kojima KK, Kohany O. 2015. Repbase Update, a database of repetitive elements in eukaryotic genomes. Mob DNA 6: 11.

Barghi N, Tobler R, Nolte V, Schlötterer C. 2017. *Drosophila simulans*: A species with improved resolution in evolve and resequence studies. G3 7: 2337–2343.

Barker MS, Demuth JP, Wade MJ. 2005. Maternal expression relaxes constraint on innovation of the anterior determinant, *bicoid*. PLoS Genet 1: e57.

Barker JSF, Moth JJ. 2001. Linkage maps of *D. simulans*: An update of Sturtevant (1929) with additional loci. *Dros Inf Serv* 84: 204-206.

Bartolomé C, Maside X, Charlesworth B. 2002. On the abundance and distribution of transposable elements in the genome of *Drosophila melanogaster*. Mol. Biol. Evol. 19: 926–937.

Bates D, Mächler M, Bolker B, Walker S. 2015. Fitting linear mixed-effects models using lme4. J Stat Softw 67: 1-48.

Baudat F, Buard J, Grey C, Fledel-Alon A, Ober C, Przeworski M, Coop G, de Massy B. 2010. PRDM9 is a major determinant of meiotic recombination hotspots in humans and mice. Science 327: 836–840.

Baudry E, Viginier B, Veuille M. 2004. Non-African populations of *Drosophila melanogaster* have a unique origin. Mol Biol Evol 21: 1482-1491.

Bauer H, Willert J, Koschorz B, Herrmann BG. 2005. The t complex–encoded GTPase-activating protein Tagap1 acts as a transmission ratio distorter in mice. Nat Genet 37: 969–973.

Begun DJ, Aquadro CF. 1992. Levels of naturally occurring DNA polymorphism correlate with recombination rates in *D. melanogaster*. Nature 356: 519–520.

Begun DJ, Aquadro CF. 1993. African and North American populations of *Drosophila melanogaster* are very different at the DNA level. Nature 365: 548-550.

Beisswanger S, Stephan W, De Lorenzo D. 2006. Evidence for a selective sweep in the *wapl* region of *Drosophila melanogaster*. Genetics 172: 265-274.

Bell G. 1982. The Masterpiece of Nature: The evolution and Genetics of Sexuality. University of California Press, Berkeley CA.
Benassi V, Depaulis F, Meghlaoui GK, Veuille M. 1999. Partial sweeping of variation at the Fbp2 locus in a west African population of *Drosophila melanogaster*. Mol Biol Evol 16: 347-353.
Bennetzen JL. 2000. Transposable element contributions to plant gene and genome evolution. Plant Mol Biol 42: 251–269.
Berchowitz LE, Hanlon SE, Lieb JD, Copenhaver GP. 2009. A positive but complex association between meiotic double-strand break hotspots and open chromatin in *Saccharomyces cerevisiae*. Genome Res 19: 2245–2257.
Bickel SE, Wyman DW, Miyazaki WY, Moore DP, Orr-Weaver TL. 1996. Identification of ORD, a *Drosophila* protein essential for sister chromatid cohesion. EMBO J 15: 1451-1459.
Bickel SE, Wyman DW, Orr-Weaver TL. 1997. Mutational analysis of the Drosophila sister-chromatid cohesion protein ORD and its role in the maintenance of centromeric cohesion. Genetics 146: 1319-1331.
Bingham PM, Kidwell MG, Rubin GM. 1982. The molecular basis of *P-M* hybrid dysgenesis: the role of the *P* element, a *P*-strain-specific transposon family. Cell 29: 995–1004.
Blumenstiel JP. 2014. Whole genome sequencing in *Drosophila virilis* identifies *Polyphemus*, a recently activated Tc1-like transposon with a possible role in hybrid dysgenesis. Mob DNA 5: 6.
Blumenstiel JP, Erwin AA, Hemmer LW. 2016. What drives positive selection in the piRNA machinery? The genomic autoimmunity hypothesis. Yale J Biol Med 89: 499–512.
Blumenstiel JP, Hartl DL. 2005. Evidence for maternally transmitted small interfering RNA in the repression of transposition in *Drosophila virilis*. Proc Natl Acad Sci USA 102: 15965–15970.
Borde V, Robine N, Lin W, Bonfils S, Géli V, Nicolas A. 2009. Histone H3 lysine 4 trimethylation marks meiotic recombination initiation sites. EMBO J 28: 99–111.
Bosco G, Campbell P, Leiva-Neto JT, Markow TA. 2007. Analysis of *Drosophila* species genome size and satellite DNA content reveals significant differences among strains as well as between species. Genetics 177: 1277–1290.
Brady MM, McMahan S, Sekelsky J. 2017. Loss of *Drosophila* Mei-41/ATR alters meiotic crossover patterning. Genetics 208: 579–588.
Brand CL, Cattani MV, Kingan SB, Landeen EL, Presgraves DC. 2018. Molecular evolution at a meiosis gene mediates species differences in the rate and patterning of recombination. Curr Biol 28: 1289–1295.
Brandvain Y, Coop G. 2012. Scrambling eggs: meiotic drive and the evolution of female recombination rates. Genetics 190: 709-723.
Brennecke J, Aravin AA, Stark A, Dus M, Kellis M, Sachidanandam R, Hannon GJ. 2007. Discrete small RNA-generating loci as master regulators of transposon activity in *Drosophila*. Cell 128: 1089–1103.
Brennecke J, Malone CD, Aravin AA, Sachidanandam R, Stark A, Hannon GJ. 2008. An epigenetic role for maternally inherited piRNAs in transposon silencing. Science 322: 1387–1392.
Broadhead RS, Kidwell JF, Kidwell MG. 1977. Variation of the recombination fraction in *Drosophila* melanogaster females. J Hered 68: 323–326.
Broman KW, Weber JL. 2000. Characterization of human crossover interference. Am J Hum Genet. 66: 1911–1926.
Broman KW, Wu H, Sen S, Churchill GA. 2003. R/qtl: QTL mapping in experimental crosses. Bioinformatics 19: 889–890.
Brower-Toland B, Findley SD, Jiang L, Liu L, Yin H, Dus M, Zhou P, Elgin SCR, Lin H. 2007. *Drosophila* PIWI associates with chromatin and interacts directly with HP1a. Genes Dev. 21: 2300–2311.
Bucheton A, Paro R, Sang HM, Pelisson A, Finnegan DJ. 1984. The molecular basis of *I-R* hybrid dysgenesis in *Drosophila melanogaster*: identification, cloning, and properties of the *I* factor. Cell 38: 153–163.

Burt A. 2000. Perspective: sex, recombination, and the efficacy of selection--was Weismann right? Evolution 54: 337–351.
Cantarel BL, Korf I, Robb SMC, Parra G, Ross E, Moore B, Holt C, Sánchez Alvarado A, Yandell M. 2008. MAKER: an easy-to-use annotation pipeline designed for emerging model organism genomes. Genome Res. 18: 188–196.
Cao L, Alani E, Kleckner N. 1990. A pathway for generation and processing of double-strand breaks during meiotic recombination in *S. cerevisiae*. Cell 61: 1089–1101.
Carlson PS. 1972. The effects of inversions and the C(3)G mutation on intragenic recombination in *Drosophila*. Genet Res 19: 129-132.
Carpenter AT. 1975. Electron microscopy of meiosis in *Drosophila melanogaster* females. I. Structure, arrangement, and temporal change of the synaptonemal complex in wild-type. Chromosoma 51: 157-182.
Carrol L. 1872. Through the looking glass and what Alice found there. Macmillan, London.
Cazemajor M, Landré C, Montchamp-Moreau C. 1997. The sex-ratio trait in *Drosophila simulans*: genetic analysis of distortion and suppression. Genetics 147: 635–642.
Chaboissier MC, Lemeunier F, Bucheton A. 1995. *IR* hybrid dysgenesis increases the frequency of recombination in *Drosophila melanogaster*. Genet Res 65: 167–174.
Charlesworth B. 1993. Directional selection and the evolution of sex and recombination. Genet Res 61: 205–224.
Charlesworth B, Charlesworth D. 1985. Genetic variation in recombination in *Drosophila*. II. Genetic analysis of a high recombination stock. Heredity 54: 85–98.
Charlesworth B, Sniegowski P, Stephan W. 1994. The evolutionary dynamics of repetitive DNA in eukaryotes. Nature 371: 215-220.
Chevin L-M, Hospital F. 2006. The hitchhiking effect of an autosomal meiotic drive gene. Genetics 173: 1829–1832.
Chinnici JP. 1971. Modification of recombination frequency in *Drosophila*. II. The polygenic control of crossing over. Genetics 69: 85–96.
Chino M, Kikkawa H. 1933. Mutants and crossing over in the dot-like chromosome of *Drosophila virilis*. Genetics 18: 111–116.
Chmátal L, Gabriel SI, Mitsainas GP, Martinez-Vargas J, Ventura J, Searle JB, Schultz RM, Lampson MA. 2014. Centromere strength provides the cell biological basis for meiotic drive and karyotype evolution in mice. Curr Biol 24: 2295-2300.
Choi K, Zhao X, Kelly KA, Venn O, Higgins JD, Yelina NE, Hardcastle TJ, Ziolkowski PA, Copenhaver GP, Franklin FCH, et al. 2013. *Arabidopsis* meiotic crossover hot spots overlap with H2A.Z nucleosomes at gene promoters. Nat Genet 45: 1327–1336.
Choo KH. 1998. Why is the centromere so cold? Genome Res 8: 81–82.
Christophorou N, Rubin T, Huynh JR. 2013. Synaptonemal complex components promote centromere pairing in pre-meiotic germ cells. PLoS Genet 9: e1004012.
Civetta A, Rajakumar SA, Brouwers B, Bacik JP. 2006. Rapid evolution and gene-specific patterns of selection for three genes of spermatogenesis in *Drosophila*. Mol Biol Evol 23: 655-662.
Clark NL, Alani E, Aquadro CF. 2012. Evolutionary rate covariation reveals shared functionality and coexpression of genes. Genome Res 22: 714-720.
Collins KA, Unruh JR, Slaughter BD, Yu Z, Lake CM, Nielsen RJ, Box KS, Miller DE, Blumenstiel JP, Perera AG, et al. 2014. Corolla is a novel protein that contributes to the architecture of the synaptonemal complex of *Drosophila*. Genetics 198: 219-228.
Comeron JM, Ratnappan R, Bailin S. 2012. The many landscapes of recombination in *Drosophila melanogaster*. PLoS Genet 8: e1002905.
Coop G, Przeworski M. 2007. An evolutionary view of human recombination. Nat Rev Genet 8: 23–34.
Costa Y, Cooke HJ. 2007. Dissecting the mammalian synaptonemal complex using targeted mutations. Chromosome Res 15: 579-589.
Cruickshank T, Wade MJ. 2008. Microevolutionary support for a developmental hourglass: gene

expression patterns shape sequence variation and divergence in *Drosophila*. Evol Dev 10: 583-590.
Daron J, Glover N, Pingault L, Theil S, Jamilloux V, Paux E, Barbe V, Mangenot S, Alberti A, Wincker P, et al. 2014. Organization and evolution of transposable elements along the bread wheat chromosome 3B. Genome Biol 15: 546.
Day WHE, Edelsbrunner H. 1984. Efficient algorithms for agglomerative hierarchical-clustering methods. J Classif 1: 7-24.
Demogines A, East AM, Lee J-H, Grossman SR, Sabeti PC, Paull TT, Sawyer SL. 2010. Ancient and recent adaptive evolution of primate non-homologous end joining genes. PLoS Genet 6: e1001169.
Demuth JP, Wade MJ. 2007. Maternal expression increases the rate of *bicoid* evolution by relaxing selective constraint. Genetica 129: 37-43.
Didion JP, Morgan AP, Clayshulte AM-F, Mcmullan RC, Yadgary L, Petkov PM, Bell TA, Gatti DM, Crowley JJ, Hua K, et al. 2015. A multi-megabase copy number gain causes maternal transmission ratio distortion on mouse chromosome 2. PLoS Genet 11: e1004850.
Drosophila 12 Genomes C, Clark AG, Eisen MB, Smith DR, Bergman CM, Oliver B, Markow TA, Kaufman TC, Kellis M, Gelbart W, et al. 2007. Evolution of genes and genomes on the Drosophila phylogeny. Nature 450: 203-218.
Duchen P, Zivkovic D, Hutter S, Stephan W, Laurent S. 2013. Demographic inference reveals African and European admixture in the North American *Drosophila melanogaster* population. Genetics 193: 291-301.
Dumont BL. 2017. Variation and evolution of the meiotic requirement for crossing over in mammals. Genetics 205: 155–168.
Dumont BL, Payseur BA. 2011. Evolution of the genomic recombination rate in murid rodents. Genetics 187: 643–657.
Dyer KA, Charlesworth B, Jaenike J. 2007. Chromosome-wide linkage disequilibrium as a consequence of meiotic drive. Proc Natl Acad Sci USA 104: 1587–1592.
Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res 32: 1792-1797.
Egea R, Casillas S, Barbadilla A. 2008. Standard and generalized McDonald-Kreitman test: a website to detect selection by comparing different classes of DNA sites. Nucleic Acids Res 36: W157-162.
Egel R, Egelmitani M, Olson LW. 1982. Meiosis in *Schizosaccharomyces pombe* and *Aspergillus nidulans* - 2 examples lacking synaptonemal complexes in the absence of crossover interference. Hereditas 97: 316-316.
Egelmitani M, Olson LW, Egel R. 1982. Meiosis in *Aspergillus nidulans* - another example for lacking synaptonemal complexes in the absence of crossover interference. Hereditas 97: 179-187.
Engels WR, Preston CR. 1979. Hybrid dysgenesis in *Drosophila melanogaster*: the biology of female and male sterility. Genetics 92: 161–174.
Erber A, Riemer D, Hofemeister H, Bovenschulte M, Stick R, Panopoulou G, Lehrach H, Weber K. 1999. Characterization of the *Hydra* lamin and its gene: A molecular phylogeny of metazoan lamins. J Mol Evol 49: 260-271.
Erwin AA, Galdos MA, Wickersheim ML, Harrison CC, Marr KD, Colicchio JM, Blumenstiel JP. 2015. piRNAs are associated with diverse transgenerational effects on gene and transposon expression in a hybrid dysgenic syndrome of *D. virilis*. PLoS Genet 11: e1005332.
Evgen'ev MB, Zelentsova H, Shostak N, Kozitsina M, Barskyi V, Lankenau DH, Corces VG. 1997. *Penelope*, a new family of transposable elements and its possible role in hybrid dysgenesis in *Drosophila virilis*. Proc Natl Acad Sci USA 94: 196–201.
Fawcett DW. 1956. The fine structure of chromosomes in the meiotic prophase of vertebrate spermatocytes. J Biophys Biochem Cytol 2: 403-406.
Felsenstein J. 1974. The evolutionary advantage of recombination. Genetics 78: 737–756.
Finn RD, Clements J, Eddy SR. 2011. HMMER web server: interactive sequence similarity searching.

Nucleic Acids Res 39: W29-37.
Finseth FR, Dong Y, Saunders A, Fishman L. 2015. Duplication and adaptive evolution of a key centromeric protein in *Mimulus*, a genus with female meiotic drive. Mol Biol Evol 32: 2694–2706.
Fishman L, Saunders A. 2008. Centromere-associated female meiotic drive entails male fitness costs in monkeyflowers. Science 322: 1559-1562.
Fishman L, Willis JH. 2005. A novel meiotic drive locus almost completely distorts segregation in *Mimulus* (Monkeyflower) hybrids. Genetics 169: 347-353.
Fiston-Lavier A-S, Singh ND, Lipatov M, Petrov DA. 2010. *Drosophila melanogaster* recombination rate calculator. Gene 463: 18–20.
Fraune J, Alsheimer M, Volff JN, Busch K, Fraune S, Bosch TC, Benavente R. 2012. *Hydra* meiosis reveals unexpected conservation of structural synaptonemal complex proteins across metazoans. Proc Natl Acad Sci USA 109: 16588-16593.
Fraune J, Brochier-Armanet C, Alsheimer M, Benavente R. 2013. Phylogenies of central element proteins reveal the dynamic evolutionary history of the mammalian synaptonemal complex: ancient and recent components. Genetics 195: 781-793.
Fujii S. 1940. Crossing over in the sixth chromosome of *Drosophila virilis*. Jpn J Genet. 16: 290–293.
Gertz EM, Yu YK, Agarwala R, Schaffer AA, Altschul SF. 2006. Composition-based statistics and translated nucleotide searches: improving the TBLASTN module of BLAST. BMC Biol 4: 41.
Getun IV, Wu ZK, Khalil AM, Bois PRJ. 2010. Nucleosome occupancy landscape and dynamics at mouse recombination hotspots. EMBO Rep 11: 555–560.
Gramates LS, Marygold SJ, Santos GD, Urbano J-M, Antonazzo G, Matthews BB, Rey AJ, Tabone CJ, Crosby MA, Emmert DB, et al. 2017. FlyBase at 25: looking to the future. Nucleic Acids Res. 45: D663–D671.
Gray S, Cohen PE. 2016. Control of meiotic crossovers: From double-strand break formation to designation. Annu Rev Genet 50: 175–210.
Gray YH, Tanaka MM, Sved JA. 1996. P-element-induced recombination in *Drosophila melanogaster*: hybrid element insertion. Genetics 144: 1601–1610.
Grishaeva TM, Bogdanov YF. 2014. Conservation and variability of synaptonemal complex proteins in phylogenesis of eukaryotes. Int J Evol Biol 2014: 856230.
Gubenko IS, Evgen'ev MB. 1984. Cytological and linkage maps of *Drosophila virilis* chromosomes. Genetica 65: 127–139.
Haaf T, Willard HF. 1997. Chromosome-specific alpha-satellite DNA from the centromere of chimpanzee chromosome 4. Chromosoma 106: 226-232.
Haerty W, Jagadeeshan S, Kulathinal RJ, Wong A, Ravi Ram K, Sirot LK, Levesque L, Artieri CG, Wolfner MF, Civetta A, et al. 2007. Evolution in the fast lane: rapidly evolving sex-related genes in *Drosophila*. Genetics 177: 1321-1335.
Haig D. 2016. Transposable elements: Self-seekers of the germline, team-players of the soma. Bioessays 38: 1158–1166.
Hall JC. 1972. Chromosome segregation influenced by two alleles of the meiotic mutant *c(3)G* in *Drosophila melanogaster*. Genetics 71: 367-400.
Hamilton WD. 1967. Extraordinary sex ratios. A sex-ratio theory for sex linkage and inbreeding has new implications in cytogenetics and entomology. Science 156: 477-488.
Hartl DL. 1975. Modifier theory and meiotic drive. Theor Popul Biol 7: 168–174.
Hartmann MA, Sekelsky J. 2017. The absence of crossovers on chromosome 4 in *Drosophila melanogaster*: Imperfection or interesting exception? Fly 11: 253–259.
Hassold T, Hunt P. 2001. To err (meiotically) is human: the genesis of human aneuploidy. Nat Rev Genet. 2: 280–291.
Hatkevich T, Kohl KP, McMahan S, Hartmann MA, Williams AM, Sekelsky J. 2017. Bloom syndrome helicase promotes meiotic crossover patterning and homolog disjunction. Curr Biol 27: 96–102.
Heil CSS, Noor MAF. 2012. Zinc finger binding motifs do not Explain recombination rate variation

within or between species of *Drosophila*. PLoS One 7: e45055.
Helleu Q, Gérard PR, Dubruille R, Ogereau D, Prud'homme B, Loppin B, Montchamp-Moreau C. 2016. Rapid evolution of a Y-chromosome heterochromatin protein underlies sex chromosome meiotic drive. Proc Natl Acad Sci USA 113: 4110–4115.
Hemmer LW, Blumenstiel JP. 2016. Holding it together: rapid evolution and positive selection in the synaptonemal complex of *Drosophila*. BMC Evol Biol 16: 91.
Henikoff S, Malik HS. 2002. Centromeres: selfish drivers. Nature 417: 227.
Hickey DA. 1982. Selfish DNA: a sexually-transmitted nuclear parasite. Genetics 101: 519–531.
Hill WG, Robertson A. 1966. The effect of linkage on limits to artificial selection. Genet. Res. 8: 269–294.
Hiraizumi Y. 1971. Spontaneous recombination in *Drosophila melanogaster* males. Proc Natl Acad Sci USA 68: 268–270.
Hiraizumi Y. 1981. Heterochromatic recombination in germ cells of *Drosophila melanogaster* females. Genetics 98: 105–114.
Hoffmann AA, Rieseberg LH. 2008. Revisiting the impact of inversions in evolution: From population genetic markers to drivers of adaptive shifts and speciation? Annu Rev Ecol Evol Syst 39: 21–42.
Housworth EA, Stahl FW. 2003. Crossover interference in humans. Am J Hum Genet 73: 188–197.
Hu TT, Eisen MB, Thornton KR, Andolfatto P. 2013. A second-generation assembly of the *Drosophila simulans* genome provides new insights into patterns of lineage-specific divergence. Genome Res 23: 89-98.
Huerta-Cepas J, Capella-Gutierrez S, Pryszcz LP, Marcet-Houben M, Gabaldon T. 2014. PhylomeDB v4: zooming into the plurality of evolutionary histories of a genome. Nucleic Acids Res 42: D897-902.
Hughes SE, Miller DE, Miller AL, Hawley RS. 2018. Female meiosis: Synapsis, recombination, and segregation in *Drosophila melanogaster*. Genetics 208: 875–908.
Hunter CM, Huang W, Mackay TFC, Singh ND. 2016. The Genetic architecture of natural variation in recombination rate in *Drosophila melanogaster*. PLoS Genet.12: e1005951.
Huttunen S, Aspi J, Hoikkala A, Schlötterer C. 2004. QTL analysis of variation in male courtship song characters in *Drosophila virilis*. Heredity 92: 263–269.
Isackson DR, Johnson TK, Denell RE. 1981. Hybrid dysgenesis in *Drosophila*: the mechanism of T-007-induced male recombination. Mol Gen Genet. 184: 539–543.
Jagadeeshan S, Singh RS. 2005. Rapidly evolving genes of *Drosophila*: differing levels of selective pressure in testis, ovary, and head tissues between sibling species. Mol Biol Evol 22: 1793-1801.
Jagannathan M, Warsinger-Pepe N, Watase GJ, Yamashita YM. 2017. Comparative analysis of satellite DNA in the species complex. G3 7: 693–704.
Jensen-Seaman MI, Furey TS, Payseur BA, Lu Y, Roskin KM, Chen C-F, Thomas MA, Haussler D, Jacob HJ. 2004. Comparative recombination rates in the rat, mouse, and human genomes. Genome Res 14: 528–538.
Johnson-Schlitz DM, Flores C, Engels WR. 2007. Multiple-pathway analysis of double-strand break repair mutations in *Drosophila*. PLoS Genet 3: e50.
Josse T, Teysset L, Todeschini A-L, Sidor CM, Anxolabéhère D, Ronsseray S. 2007. Telomeric trans-silencing: an epigenetic repression combining RNA silencing and heterochromatin formation. PLoS Genet 3: 1633–1643.
Joyce EF, Apostolopoulos N, Beliveau BJ, Wu CT. 2013. Germline progenitors escape the widespread phenomenon of homolog pairing during *Drosophila* development. PLoS Genet 9: e1004013.
Joyce EF, Pedersen M, Tiong S, White-Brown SK, Paul A, Campbell SD, McKim KS. 2011. *Drosophila* ATM and ATR have distinct activities in the regulation of meiotic DNA damage and repair. J Cell Biol 195: 359–367.
Katoh K, Misawa K, Kuma K, Miyata T. 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res 30: 3059-3066.
Kazazian HH Jr, Wong C, Youssoufian H, Scott AF, Phillips DG, Antonarakis SE. 1988. Haemophilia A

resulting from *de novo* insertion of *L1* sequences represents a novel mechanism for mutation in man. Nature 332: 164–166.
Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C, et al. 2012. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics 28: 1647-1649.
Kent TV, Uzunović J, Wright SI. 2017. Coevolution between transposable elements and recombination. Phil Trans R Soc B 372: 20160458
Khetani RS, Bickel SE. 2007. Regulation of meiotic cohesion and chromosome core morphogenesis during pachytene in *Drosophila* oocytes. J Cell Sci 120: 3123-3137.
Kidwell MG. 1972. Genetic change of recombination value in *Drosophila melanogaster*. I. Artificial selection for high and low recombination and some properties of recombination-modifying genes.
Genetics 70: 419–432.
Kidwell MG. 1977. Reciprocal differences in female recombination associated with hybrid dysgenesis in *Drosophila melanogaster*. Genet Res 30: 77–88.
Kidwell MG, Kidwell JF. 1975. Spontaneous male recombination and mutation in isogenic-derived chromosomes of *Drosophila melanogaster*. J Hered 66: 367–375.
Kidwell MG, Kidwell JF, Sved JA. 1977. Hybrid dysgenesis in *Drosophila melanogaster*: A syndrome of aberrant traits including mutation, sterility and male recombination. Genetics 86: 813–833.
Kirkpatrick M. 2010. How and why chromosome inversions evolve. PLoS Biol 8: e1000501.
Klenov MS, Lavrov SA, Stolyarenko AD, Ryazansky SS, Aravin AA, Tuschl T, Gvozdev VA. 2007. Repeat-associated siRNAs cause chromatin silencing of retrotransposons in the *Drosophila melanogaster* germline. Nucleic Acids Res 35:v5430–5438.
Klovstad M, Abdu U, Schüpbach T. 2008. *Drosophila brca2* is required for mitotic and meiotic DNA repair and efficient activation of the meiotic recombination checkpoint. PLoS Genet 4: e31.
Ko W-Y, David RM, Akashi H. 2003. Molecular phylogeny of the *Drosophila melanogaster* species subgroup. J Mol Evol 57: 562–573.
Koehler KE, Hawley RS, Sherman S, Hassold T. 1996. Recombination and nondisjunction in humans and flies. Hum Mol Genet 5: 1495–1504.
Kofler R, Betancourt AJ, Schlötterer C. 2012. Sequencing of pooled DNA samples (Pool-Seq) uncovers complex dynamics of transposable element insertions in *Drosophila melanogaster*. PLoS Genet 8: e1002487.
Kong A, Gudbjartsson DF, Sainz J, Jonsdottir GM, Gudjonsson SA, Richardsson B, Sigurdardottir S, Barnard J, Hallbeck B, Masson G, et al. 2002. A high-resolution recombination map of the human genome. Nat Genet 31: 241–247.
Kouznetsova A, Benavente R, Pastink A, Hoog C. 2011. Meiosis in mice without a synaptonemal complex. PLoS One 6: e28255.
Ku H-Y, Lin H. 2014. PIWI proteins and their interactors in piRNA biogenesis, germline development and gene expression. Natl Sci Rev 1: 205–218.
Kulathinal RJ, Bennett SM, Fitzpatrick CL, Noor MAF. 2008. Fine-scale mapping of recombination rate in *Drosophila* refines its correlation to diversity and divergence. Proc Natl Acad Sci USA 105: 10051–10056.
Kurdzo EL, Dawson DS. 2015. Centromere pairing--tethering partner chromosomes in meiosis I. FEBS J 282: 2458-2470.
Lake CM, Hawley RS. 2012. The molecular control of meiotic chromosomal behavior: events in early meiotic prophase in *Drosophila* oocytes. Annu Rev Physiol 74: 425-451.
Lasko D. 1991. Loss of constitutional heterozygosity In human cancer. Annu Rev Genet 25: 281–314.
Lee YCG, Karpen GH. 2017. Pervasive epigenetic effects of euchromatic transposable elements impact their evolution. Elife 6:e25762
Lenth RV. 2016. Least-squares means: The R package lsmeans. J Stat Softw 69: 1-33
Lemeunier F, Ashburner M. 1976. Relationships within the *melanogaster* species subgroup of the genus

Drosophila (*Sophophora*). II. Phylogenetic relationships between six species based upon polytene chromosome banding sequences. P Roy Soc B-Biol Sci 193: 275–294.
Lenormand T, Otto SP. 2000. The evolution of recombination in a heterogeneous environment. Genetics 156: 423–438.
Le Thomas A, Marinov GK, Aravin AA. 2014. A transgenerational process defines piRNA biogenesis in *Drosophila virilis*. Cell Rep 8: 1617–1623.
Le Thomas A, Rogers AK, Webster A, Marinov GK, Liao SE, Perkins EM, Hur JK, Aravin AA, Tóth KF. 2013. Piwi induces piRNA-guided transcriptional silencing and establishment of a repressive chromatin state. Genes Dev 27: 390–399.
Levine AJ, Ting DT, Greenbaum BD. 2016. P53 and the defenses against genome instability caused by transposons and repetitive elements. Bioessays 38: 508–513.
Li H. 2011. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. Bioinformatics 27: 2987–2993.
Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25: 1754–1760.
Llopart A, Elwyn S, Lachaise D, Coyne JA. 2002. Genetics of a difference in pigmentation between Drosophila yakuba and Drosophila santomea. Evolution 56: 2262–2277.
Loidl J. 2006. *S. pombe* linear elements: the modest cousins of synaptonemal complexes. Chromosoma 115: 260-271.
Loidl J, Scherthan H. 2004. Organization and pairing of meiotic chromosomes in the ciliate *Tetrahymena* thermophila. J Cell Sci 117: 5791-5801.
Lorenz A, Wells JL, Pryce DW, Novatchkova M, Eisenhaber F, McFarlane RJ, Loidl J. 2004. *S. pombe* meiotic linear elements contain proteins related to synaptonemal complex components. J Cell Sci 117: 3343-3351.
Lou DI, McBee RM, Le UQ, Stone AC, Wilkerson GK, Demogines AM, Sawyer SL. 2014. Rapid evolution of *BRCA1* and *BRCA2* in humans and other primates. BMC Evol Biol 14: 155.
Loytynoja A, Goldman N. 2005. An algorithm for progressive multiple alignment of sequences with insertions. Proc Natl Acad Sci USA 102: 10557-10562.
Lozovskaya ER, Scheinker VS, Evgen'ev MB. 1990. A hybrid dysgenesis syndrome in *Drosophila virilis*. Genetics 126: 619–623.
Lu W-J, Chapo J, Roig I, Abrams JM. 2010. Meiotic recombination provokes functional activation of the p53 regulatory network. Science 328: 1278–1281.
Ma X, Zhu X, Han Y, Story B, Do T, Song X, Wang S, Zhang Y, Blanchette M, Gogol M, et al. 2017. Aubergine controls germline stem cell self-renewal and progeny differentiation via distinct mechanisms. Dev Cell 41: 157–169.
Mackay TF, Richards S, Stone EA, Barbadilla A, Ayroles JF, Zhu D, Casillas S, Han Y, Magwire MM, Cridland JM, et al. 2012. The *Drosophila melanogaster* Genetic Reference Panel. Nature 482: 173-178.
Marand AP, Jansky SH, Zhao H, Leisner CP, Zhu X, Zeng Z, Crisovan E, Newton L, Hamernik AJ, Veilleux RE, et al. 2017. Meiotic crossovers are associated with open chromatin and enriched with *Stowaway* transposons in potato. Genome Biol 18: 203.
Malik HS. 2009. The centromere-drive hypothesis: a simple basis for centromere complexity. Prog Mol Subcell Biol 48: 33-52.
Malik HS, Henikoff S. 2001. Adaptive evolution of Cid, a centromere-specific histone in *Drosophila*. Genetics 157: 1293-1298.
Malik HS, Henikoff S. 2002. Conflict begets complexity: the evolution of centromeres. Curr Opin Genet Dev 12: 711-718.
Malik HS, Vermaak D, Henikoff S. 2002. Recurrent evolution of DNA-binding motifs in the *Drosophila* centromeric histone. Proc Natl Acad Sci USA 99: 1449–1454.
Manheim EA, McKim KS. 2003. The Synaptonemal complex component C(2)M regulates meiotic crossing over in *Drosophila*. Curr Biol 13: 276-285.

Marygold SJ, Leyland PC, Seal RL, Goodman JL, Thurmond J, Strelets VB, Wilson RJ, FlyBase c. 2013. FlyBase: improvements to the bibliography. Nucleic Acids Res 41: D751-757.
Mason JM. 1976. Orientation disruptor (*ord*): a recombination-defective and disjunction-defective meiotic mutant in *Drosophila melanogaster*. Genetics 84: 545-572.
Maynard Smith J. 1978. The evolution of sex. Cambridge Press, Cambridge.
McCarron M, Duttaroy A, Doughty G, Chovnick A. 1994. *Drosophila P* element transposase induces male recombination additively and without a requirement for *P* element excision or insertion. Genetics 136: 1013–1023.
McDonald JH, Kreitman M. 1991. Adaptive protein evolution at the Adh locus in *Drosophila*. Nature 351: 652-654.
McKim KS, Hayashi-Hagihara A. 1998. *mei-W68* in *Drosophila melanogaster* encodes a Spo11 homolog: evidence that the mechanism for initiating meiotic recombination is conserved. Genes Dev 12: 2932–2942.
Merrill C, Bayraktaroglu L, Kusano A, Ganetzky B. 1999. Truncated RanGAP encoded by the *Segregation Distorter* locus of *Drosophila*. Science 283: 1742–1745.
Miller DE, Smith CB, Kazemi NY, Cockrell AJ, Arvanitakas AV, Blumenstiel JP, Jaspersen SL, Hawley RS. 2016. Whole-genome analysis of individual meiotic events in *Drosophila melanogaster* reveals that noncrossover gene conversions are insensitive to interference and the centromere effect. Genetics 203: 159–171.
Miller MA, Pfeiffer W, Schwartz T editors. Proceedings of the Gateway Computing Environments Workshop (GCE). 2010 November 14: New Orleans, LA.
Miyazaki WY, Orr-Weaver TL. 1992. Sister-chromatid misbehavior in *Drosophila ord* mutants. Genetics 132: 1047-1061.
Moses MJ. 1956. Chromosomal structures in crayfish spermatocytes. J Biophys Biochem Cytol 2: 215-218.
Muller HJ. 1964. The relation of recombination to mutational advance. Mutat Res 106: 2–9.
Myers S, Bowden R, Tumian A, Bontrop RE, Freeman C, MacFie TS, McVean G, Donnelly P. 2010. Drive against hotspot motifs in primates implicates the PRDM9 gene in meiotic recombination. Science 327: 876–879.
Nielsen R, Bustamante C, Clark AG, Glanowski S, Sackton TB, Hubisz MJ, Fledel-Alon A, Tanenbaum DM, Civello D, White TJ, et al. 2005. A scan for positively selected genes in the genomes of humans and chimpanzees. PLoS Biol 3: e170.
Nolte V, Pandey RV, Kofler R, Schlotterer C. 2013. Genome-wide patterns of natural variation reveal strong selective sweeps and ongoing genomic conflict in Drosophila *mauritiana*. Genome Res 23: 99-110.
Nozawa M, Nei M. 2007. Evolutionary dynamics of olfactory receptor genes in *Drosophila* species. Proc Natl Acad Sci USA 104: 7122-7127.
Nurminsky D, Aguiar DD, Bustamante CD, Hartl DL. 2001. Chromosomal effects of rapid gene evolution in *Drosophila melanogaster*. Science 291: 128-130.
Obbard DJ, Maclennan J, Kim K-W, Rambaut A, O'Grady PM, Jiggins FM. 2012. Estimating divergence dates and substitution rates in the *Drosophila* phylogeny. Mol Biol Evol 29: 3459–3473.
Obeso D, Pezza RJ, Dawson D. 2014. Couples, pairs, and clusters: mechanisms and implications of centromere associations in meiosis. Chromosoma 123: 43-55.
Oliver PL, Goodstadt L, Bayes JJ, Birtle Z, Roach KC, Phadnis N, Beatson SA, Lunter G, Malik HS, Ponting CP. 2009. Accelerated evolution of the Prdm9 speciation gene across diverse metazoan taxa. PLoS Genet 5: e1000753.
Olson LW, Eden U, Egelmitani M, Egel R. 1978. Asynaptic meiosis in fission yeast. Hereditas 89: 189-199.
Page SL, Hawley RS. 2001. *c(3)G* encodes a *Drosophila* synaptonemal complex protein. Genes Dev 15: 3130-3143.
Page SL, Hawley RS. 2003. Chromosome choreography: the meiotic ballet. Science 301: 785–789.

Page SL, Hawley RS. 2004. The genetics and molecular biology of the synaptonemal complex. Annu Rev Cell Dev Biol 20: 525-558.
Page SL, Khetani RS, Lake CM, Nielsen RJ, Jeffress JK, Warren WD, Bickel SE, Hawley RS. 2008. Corona is required for higher-order assembly of transverse filaments into full-length synaptonemal complex in *Drosophila* oocytes. PLoS Genet 4: e1000194.
Pagès H, Aboyoun P, Gentleman RGentleman, DebRoy S. 2017. Biostrings: Efficient manipulation of biological strings. R package version 2.46.0.
Pardo-Manuel de Villena F, de Villena FP-M, Sapienza C. 2001a. Nonrandom segregation during meiosis: the unfairness of females. Mamm Genome 12: 331–339.
Pardo-Manuel de Villena F, de Villena FP-M, Sapienza C. 2001b. Recombination is proportional to the number of chromosome arms in mammals. Mamm Genome 12: 318–322.
Parvanov ED, Petkov PM, Paigen K. 2010. Prdm9 controls activation of mammalian recombination hotspots. Science 327: 835.
Peter A, Reimer S. 2012. Evolution of the lamin protein family: what introns can tell. Nucleus 3: 44-59.
Petrov DA, Schutzman JL, Hartl DL, Lozovskaya ER. 1995. Diverse transposable elements are mobilized in hybrid dysgenesis in *Drosophila virilis*. Proc Natl Acad Sci USA 92: 8050–8054.
Phadnis N, Orr HA. 2009. A single gene causes both male sterility and segregation distortion in *Drosophila* hybrids. Science 323: 376–379.
Phalke S, Nickel O, Walluscheck D, Hortig F, Onorati MC, Reuter G. 2009. Retrotransposon silencing and telomere integrity in somatic cells of *Drosophila* depends on the cytosine-5 methyltransferase DNMT2. Nat Genet 41: 696–702.
Picelli S, Björklund AK, Reinius B, Sagasser S, Winberg G, Sandberg R. 2014. Tn5 transposase and tagmentation procedures for massively scaled sequencing projects. Genome Res 24: 2033–2040.
Pieper KE, Dyer KA. 2016. Occasional recombination of a selfish X-chromosome may permit its persistence at high frequencies in the wild. J Evol Biol 29: 2229–2241.
Pond SL, Frost SD. 2005a. Datamonkey: rapid detection of selective pressure on individual sites of codon alignments. Bioinformatics 21: 2531-2533.
Pond SL, Frost SD. 2005b. A genetic algorithm approach to detecting lineage-specific variation in selection pressure. Mol Biol Evol 22: 478-485.
Pond SL, Frost SD, Muse SV. 2005. HyPhy: hypothesis testing using phylogenies. Bioinformatics 21: 676-679.
Pool JE, Corbett-Detig RB, Sugino RP, Stevens KA, Cardeno CM, Crepeau MW, Duchen P, Emerson JJ, Saelao P, Begun DJ, et al. 2012. Population Genomics of Sub-Saharan *Drosophila melanogaster*: African diversity and non-African admixture. PLoS Genet 8: e1003080.
Preston CR, Engels WR. 1996. P-element-induced male recombination and gene conversion in *Drosophila*. Genetics 144: 1611–1622.
Preston CR, Flores CC, Engels WR. 2006. Differential usage of alternative pathways of double-strand break repair in *Drosophila*. Genetics 172: 1055–1068.
Preston CR, Sved JA, Engels WR. 1996. Flanking duplications and deletions associated with P-induced male recombination in *Drosophila*. Genetics 144: 1623–1638.
R Core Team. 2017. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL: https://www.R-project.org/.
Raju NB. 1994. Ascomycete spore killers: Chromosomal elements that distort genetic ratios among products of meiosis. Mycologia 86: 461-473.
Rasmusse SW. 1973. Ultrastructural studies of spermatogenesis in *Drosophila melanogaster*. Z Zellforsch Und Mik Ana 140: 125-144.
Reinhardt JA, Brand CL, Paczolt KA, Johns PM, Baker RH, Wilkinson GS. 2014. Meiotic drive impacts expression and evolution of X-linked genes in stalk-eyed flies. PLoS Genet 10: e1004362.
Ritz KR, Noor MAF, Singh ND. 2017. Variation in recombination rate: Adaptive or not? Trends Genet 33: 364–374.
Rizzon C, Marais G, Gouy M, Biémont C. 2002. Recombination rate and the distribution of transposable

elements in the *Drosophila melanogaster* genome. Genome Res 12: 400–407.
Rockmill B, Voelkel-Meiman K, Roeder GS. 2006. Centromere-proximal crossovers are associated with precocious separation of sister chromatids during meiosis in *Saccharomyces cerevisiae*. Genetics 174: 1745–1754.
Rogers RL, Bedford T, Lyons AM, Hartl DL. 2010. Adaptive impact of the chimeric gene *Quetzalcoatl* in *Drosophila melanogaster*. Proc Natl Acad Sci USA 107: 10943-10948.
Rogers RL, Cridland JM, Shao L, Hu TT, Andolfatto P, Thornton KR. 2015. Tandem duplications and the limits of natural selection in *Drosophila yakuba* and *Drosophila simulans*. PLoS One 10: e0132184.
Rozas J. 2009. DNA sequence polymorphism analysis using DnaSP. Methods Mol Biol 537: 337-350.
Samonte RV, Ramesh KH, Verma RS. 1997. Comparative mapping of human alphoid satellite DNA repeat sequences in the great apes. Genetica 101: 97-104.
Sandler L, Novitski E. 1957. Meiotic drive as an evolutionary force. Am Nat 91: 105–110.
de los Santos T, Hunter N, Lee C, Larkin B, Loidl J, Hollingsworth NM. 2003. The Mus81/Mms4 endonuclease acts independently of double-Holliday junction resolution to promote a distinct subset of crossovers during meiosis in budding yeast. Genetics 164: 81–94.
Sarkar B. 2016. Random and non-random mating populations: Evolutionary dynamics in meiotic drive. Math Biosci 271: 29–41.
Sawyer SL, Malik HS. 2006. Positive selection of yeast nonhomologous end-joining genes and a retrotransposon conflict hypothesis. Proc Natl Acad Sci USA 103: 17614–17619.
Schaeffer SW, Bhutkar A, McAllister BF, Matsuda M, Matzkin LM, O'Grady PM, Rohde C, Valente VLS, Aguadé M, Anderson WW, et al. 2008. Polytene chromosomal maps of 11 *Drosophila* species: the order of genomic scaffolds inferred from genetic and physical maps. Genetics 179: 1601–1655.
Segura J, Ferretti L, Ramos-Onsins S, Capilla L, Farré M, Reis F, Oliver-Bonet M, Fernández-Bellón H, Garcia F, Garcia-Caldés M, et al. 2013. Evolution of recombination in eutherian mammals: insights into mechanisms that affect recombination rates and crossover interference. Proc Biol Sci 280: 20131945.
Shilo S, Melamed-Bessudo C, Dorone Y, Barkai N, Levy AA. 2015. DNA crossover motifs associated with epigenetic modifications delineate open chromatin regions in *Arabidopsis*. Plant Cell 27:2427–2436.
Shim HJ, Lee E-M, Nguyen LD, Shim J, Song Y-H. 2014. High-dose irradiation induces cell cycle arrest, apoptosis, and developmental defects during *Drosophila* oogenesis. PLoS One 9: e89009.
Sienski G, Dönertas D, Brennecke J. 2012. Transcriptional silencing of transposons by Piwi and maelstrom and its impact on chromatin state and gene expression. Cell 151: 964–980.
Slatko BE. 1978. Parameters of male and female recombination influenced by the T-007 second chromosome in *Drosophila melanogaster*. Genetics 90: 257–276.
Slotkin RK, Keith Slotkin R, Martienssen R. 2007. Transposable elements and the epigenetic regulation of the genome. Nat Rev Genet 8: 272–285.
Smith NG, Eyre-Walker A. 2002. Adaptive protein evolution in *Drosophila*. Nature 415: 1022-1024.
Smukowski CS, Noor MAF. 2011. Recombination rate variation in closely related species. Heredity 107: 496–508.
Smukowski Heil CS, Ellison C, Dubin M, Noor MAF. 2015. Recombining without hotspots: A comprehensive evolutionary portrait of recombination in two closely related species of *Drosophila*. Genome Biol Evol 7: 2829–2842.
Sokolova MI, Zelentsova ES, Rozhkov NV, Evgeniev MB. 2010. Morphologic and molecular manifestations of hybrid dysgenesis in ontogenesis of *Drosophila virilis*. Russ J Dev Biol 41: 391–393.
Solovyev V, Kosarev P, Seledsov I, Vorobyev D. 2006. Automatic annotation of eukaryotic genes, pseudogenes and promoters. Genome Biol 7 Suppl 1: S10 11-12.
Staten R, Schully SD, Noor MAF. 2004. A microsatellite linkage map of *Drosophila mojavensis*. BMC

Genet 5: 12.
Stevison LS, Noor MAF. 2010. Genetic and evolutionary correlates of fine-scale recombination rate variation in *Drosophila persimilis*. J Mol Evol 71: 332–345.
Stevison LS, Woerner AE, Kidd JM, Kelley JL, Veeramah KR, McManus KF, Bustamante CD, Hammer MF, Wall JD. 2016. The time-scale of recombination rate evolution in great apes. Mol Bio Evol 33: 928-945.
Stocker AJ, Rusuwa BB, Blacket MJ, Frentiu FD, Sullivan M, Foley BR, Beatson S, Hoffmann AA, Chenoweth SF. 2012. Physical and linkage maps for *Drosophila serrata*, a model species for studies of clinal adaptation and sexual selection. G3 2: 287–297.
Sturtevant AH. 1913. The linear arrangement of six sex-linked factors in *Drosophila*, as shown by their mode of association. J Exp Zool.14: 43–59.
Sved JA, Blackman LM, Gilchrist AS, Engels WR. 1991. High levels of recombination induced by homologous *P* elements in *Drosophila melanogaster*. Mol Gen Genet 225: 443–447.
Sved JA, Blackman LM, Svoboda Y, Colless R. 1995. Male recombination with single and homologous *P* elements in *Drosophila melanogaster*. Mol Gen Genet 246: 381–386.
Swanson WJ, Vacquier VD. 2002. The rapid evolution of reproductive proteins. Nat Rev Genet 3: 137-144.
Swanson WJ, Wong A, Wolfner MF, Aquadro CF. 2004. Evolutionary expressed sequence tag analysis of *Drosophila* female reproductive tracts identifies genes subjected to positive selection. Genetics 168: 1457-1465.
Symington LS, Rothstein R, Lisby M. 2014. Mechanisms and regulation of mitotic recombination in *Saccharomyces cerevisiae*. Genetics 198: 795–835.
Tajima F. 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics 123: 585-595.
Takano-Shimizu T. 2001. Local changes in GC/AT substitution biases and in crossover frequencies on *Drosophila* chromosomes. Mol Biol Evol 18: 606–619.
Takeo S, Lake CM, Morais-de-Sa E, Sunkel CE, Hawley RS. 2011. Synaptonemal complex-dependent centromeric clustering and the initiation of synapsis in *Drosophila* oocytes. Curr Biol 21: 1845-1851.
Talbert PB, Henikoff S. 2010. Centromeres convert but don't cross. PLoS Biol 8: e1000326.
Tanneti NS, Landy K, Joyce EF, McKim KS. 2011. A pathway for synapsis initiation during zygotene in *Drosophila* oocytes. Curr Biol 21: 1852-1857.
Tarailo-Graovac M, Chen N. 2009. Using RepeatMasker to identify repetitive elements in genomic sequences. Curr Protoc in Bioinformatics 4: 1-14
Tasnim S, Kelleher ES. 2018. p53 is required for female germline stem cell maintenance in *P*-element hybrid dysgenesis. Dev Biol 434: 215–220.
Thomas JH, Emerson RO, Shendure J. 2009. Extraordinary molecular evolution in the PRDM9 fertility gene. PLoS One 4: e8505.
Torgerson DG, Kulathinal RJ, Singh RS. 2002. Mammalian sperm proteins are rapidly evolving: evidence of positive selection in functionally diverse genes. Mol Biol Evol 19: 1973-1980.
True JR, Mercer JM, Laurie CC. 1996. Differences in crossover frequency and distribution among three sibling species of *Drosophila*. Genetics 142: 507-523.
Tsubouchi T, Roeder GS. 2005. A synaptonemal complex protein promotes homology-independent centromere coupling. Science 308: 870-873.
Unckless RL, Larracuente AM, Clark AG. 2015. Sex-ratio meiotic drive and Y-linked resistance in *Drosophila affinis*. Genetics 199: 831–840.
Underwood CJ, Choi K, Lambing C, Zhao X, Serra H, Borges F, Simorowski J, Ernst E, Jacob Y, Henderson IR, et al. 2017. Epigenetic activation of meiotic recombination in *Arabidopsis thaliana* centromeres via loss of H3K9me2 and non-CG DNA methylation. Genome Res 28: 519-531
Valen V. 1973. A new evolutionary law. Evol Theory 1: 1-30

Vieira J, Vieira CP, Hartl DL, Lozovskaya ER. 1998. Factors contributing to the hybrid dysgenesis syndrome in *Drosophila virilis.* Genet Res 71: 109–117.
Vincenten N, Kuhl L-M, Lam I, Oke A, Kerr AR, Hochwagen A, Fung J, Keeney S, Vader G, Marston AL. 2015. The kinetochore prevents centromere-proximal crossover recombination during meiosis. Elife 4: e10850.
Wang SH, Elgin SCR. 2011. *Drosophila* Piwi functions downstream of piRNA production mediating a chromatin-based transposon silencing mechanism in female germ line. Proc Natl Acad Sci USA 108: 21164–21169.
Waterhouse RM, Zdobnov EM, Tegenfeldt F, Li J, Kriventseva EV. 2011. OrthoDB: the hierarchical catalog of eukaryotic orthologs in 2011. Nucleic Acids Res 39: D283-288.
Webber HA, Howard L, Bickel SE. 2004. The cohesion protein ORD is required for homologue bias during meiotic recombination. J Cell Biol 164: 819-829.
Weinstein A. 1918. Coincidence of crossing over in Drosophila melanogaster (Ampelophila). Genetics 3: 135–172.
Weinstein A. 1920. Homologous genes and linear linkage in *Drosophila virilis*. Proc Natl Acad Sci USA 6: 625–639.
Westphal T, Reuter G. 2002. Recombinogenic effects of suppressors of position-effect variegation in *Drosophila*. Genetics 160: 609–621.
von Wettstein D, Rasmussen SW, Holm PB. 1984. The synaptonemal complex in genetic segregation. Annu Rev Genet 18: 331-413.
White MJD. 1977. Animal cytology and evolution. Macmillan, New York NY.
Wickham H. 2009. ggplot2: elegant graphics for data analysis. Springer.
Wilfert L, Gadau J, Schmid-Hempel P. 2007. Variation in genomic recombination rates among animal taxa and the case of social insects. Heredity 98: 189–197.
Wilson DJ, Hernandez RD, Andolfatto P, Przeworski M. 2011. A population genetics-phylogenetics approach to inferring natural selection in coding sequences. PLoS Genet 7: e1002395.
Winckler W, Myers SR, Richter DJ, Onofrio RC, McDonald GJ, Bontrop RE, McVean GAT, Gabriel SB, Reich D, Donnelly P, et al. 2005. Comparison of fine-scale recombination rates in humans and chimpanzees. Science 308: 107–111.
Witherspoon DJ, Watkins WS, Zhang Y, Xing J, Tolpinrud WL, Hedges DJ, Batzer MA, Jorde LB. 2009. *Alu* repeats increase local recombination rates. BMC Genomics 10: 530.
Wolfe NW, Clark NL. 2015. ERC analysis: web-based inference of gene function via evolutionary rate covariation. Bioinformatics 31: 3835-3837.
Wylie A, Jones AE, D'Brot A, Lu W-J, Kurtz P, Moran JV, Rakheja D, Chen KS, Hammer RE, Comerford SA, et al. 2016. p53 genes function to restrain mobile elements. Genes Dev 30: 64–77.
Yamamoto M, Miklos GL. 1978. Genetic studies on heterochromatin in *Drosophila melanogaster* and their implications for the functions of satellite DNA. Chromosoma 66: 71–98.
Yan R, Thomas SE, Tsai JH, Yamada Y, McKee BD. 2010. SOLO: a meiotic protein required for centromere cohesion, coorientation, and SMC1 localization in *Drosophila melanogaster*. J Cell Biol 188: 335-349.
Yang Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. Mol Biol Evol 24: 1586-1591.
Yang Z, Bielawski JP. 2000. Statistical methods for detecting molecular adaptation. Trends Ecol Evol 15: 496-503.
Yannopoulos G, Stamatis N, Monastirioti M, Hatzopoulos P, Louis C. 1987. *hobo* is responsible for the induction of hybrid dysgenesis by strains of *Drosophila melanogaster* bearing the male recombination factor 23.5MRF. Cell 49: 487–495.
Zhang J, Yu C, Krishnaswamy L, Peterson T. 2011. Transposable elements as catalysts for chromosome rearrangements. Methods Mol Biol 701: 315–326.
Zhang L, Liang Z, Hutchinson J, Kleckner N. 2014. Crossover patterning by the beam-film model: Analysis and implications. PLoS Genet 10: e1004042.
Zickler D. 1999. The synaptonemal complex: a structure necessary for pairing, recombination or

organization of the meiotic chromosome? J Soc Biol 193: 17-22.
Zwick ME, Cutler DJ, Langley CH. 1999. Classic Weinstein: tetrad analysis, genetic variation and achiasmate segregation in *Drosophila* and humans. Genetics 152: 1615–1629.
Zwick ME, Salstrom JL, Langley CH. 1999. Genetic variation in rates of nondisjunction: association of two naturally occurring polymorphisms in the chromokinesin *nod* with increased rates of nondisjunction in *Drosophila melanogaster*. Genetics 152: 1605-1614.

Appendices

Appendix 1: Syntenic checks for all confirmed orthologs. “X” marks the presence of one of the four nearby genes (2 on the Left and 2 on the Right). “Y” marks certainty of synteny while “N” is not syntenic. A “?” marks an ambiguous identified ortholog, either due to the absence of nearby syntenic genes or being located on the wrong Muller Element.

Gene	Species	Contig/ Chromosome	2L Gene	1L Gene	1R Gene	2R Gene	Muller Element	Synt -enic
			CG12782	*CG13540*	*CG3124*	*CG13541*		
Ord	*D. ananassae*	scaffold_13266	X	-	X	-	C	Y
	D. erecta	scaffold_4845	X	X	X	X	BC	Y
	D. grimshawi	scaffold_15112	-	-	X	-	C	Y
	D. melanogaster	2R	X	X	X	X	C	Y
	D. mojavensis	scaffold_6496	-	-	X	-	C	Y
	D. persimilis	scaffold_2	-	-	X	-	C	Y
	D. pseudoobscura	3	-	-	X	-	C	Y
	D. sechilia	scaffold_9	X	X	X	X	C	Y
	D. simulans	2R	X	X	X	X	C	Y
	D. virilis	scaffold_12875	-	-	X	-	C	Y
	D. willistoni	scf2_1100000004512	-	-	-	-	C	?
	D. yakuba	2R	X	X	X	X	BC	Y
			CG17328	*CG5869*	*CG5861*	*Syx5*		
C(2)M	*D. ananassae*	scaffold_12916	-	-	-	-	B	Y?
	D. erecta	scaffold_4929	X	X	X	X	BC	Y
	D. grimshawi	scaffold_15252	X	X	X	X	B	Y
	D. melanogaster	2L	X	X	X	X	B	Y
	D. mojavensis	scaffold_6500	X	X	X	X	B	Y
	D. persimilis	scaffold_8	X	X	X	X	B	Y
	D. pseudoobscura	4_group2	X	X	X	X	B	Y
	D. sechilia	scaffold_5	X	X	X	X	B	Y
		scaffold_6278	-	-	-	-	?	N
	D. simulans	2L	X	X	X	X	B	Y
	D. virilis	scaffold_12963	X	X	X	X	B	Y
		scaffold_12970	-	-	-	-	A	N
	D. willistoni	scf_1100000004516	-	-	X	X	B	Y
	D. yakuba	2L	X	X	X	X	BC	Y
			CG9590	*CG9589*	*Acyp2*	*wah*		
C(3)G	*D. ananassae*	scaffold_13266	-	-	-	-	C	?
	D. erecta	scaffold_4770	X	X	X	X	E	Y

		scaffold_4770	X	X	X	X	E	Y
	D. grimshawi	-	-	-	-	-	-	-
	D. melanogaster	3R	X	X	X	X	E	Y
	D. mojavensis	-	-	-	-	-	-	-
	D. persimilis	scaffold_0	-	-	-	-	E	?
	D. pseudoobscura	2	-	-	-	-	E	?
	D. sechilia	scaffold_0	X	X	X	X	E	Y
	D. simulans	3R	X	X		X	E	Y
	D. virilis	-	-	-	-	-	-	-
	D. willistoni	-	-	-	-	-	-	-
	D. yakuba	3R	X	X	X	X	E	Y
			stas	*CG8326*	*CG5703*	*CG8289*		
	D. ananassae		-	-	-	-	-	
	D. erecta	scaffold_4690	X	X	X	X	A	Y
	D. grimshawi	scaffold_15074	-	-	-	-	E	?
	D. melanogaster	X	X	X	X	X	A	Y
	D. mojavensis	X	X	-	-	-	A	Y
Corolla	*D. persimilis*	scaffold_17	X	X	X	X	A	Y
	D. pseudoobscura	XL_group1e	X	X	X	X	A	Y
	D. sechilia	scaffold_17	X	X	X	X	A	Y
	D. simulans	X	X	X	X	X	A	Y
	D. virilis	scaffold_12970	X	-	-	-	A	Y
	D. willistoni	-	-	-	-	-	-	-
	D. yakuba	X	X	X	X	X	A	Y
			CG7675	*CG14309*	*Vha100 -4*	*Vha100-2*		
	D. ananassae	-	-	-	-	-	-	-
	D. erecta	scaffold_4770	X	X	X	X	E	Y
	D. grimshawi	-	-	-	-	-	-	-
	D. melanogaster	3R	X	X	X	X	E	Y
	D. mojavensis	-	-	-	-	-	-	-
	D. persimilis	scaffold_17	-	-	-	-	A	N
Cona	*D. pseudoobscura*	XL_group1e	-	-	-	-	A	N
	D. sechilia	scaffold_5	X	X	X	X	E	
	D. simulans	3R	X	X	X	X	E	
	D. virilis	-	-	-	-	-	-	-
	D. willistoni	scf2_1100000004963	-	-	-	-	A	N
	D. yakuba	3R	X	X	X	X	E	Y
		v2_chrUn_2917	-	-	-	-	?	?

Appendix 2: Orthology searches of all 5 genes amongst PhylomeDB, OrthoDB, and HMMER.

Gene	Species	PhylomeDB	OrthoDB	HMMER	HMMER E-value
Ord	*D. ananassae*	FBgn0090337	FBgn0090337	FBgn0090337	3.3E-134
	D. erecta	FBgn0112252	FBgn0112252	FBgn0112252	7.6E-270
	D. grimshawi	FBgn0130509	FBgn0130509	FBgn0130509	9.4E-104
	D. melanogaster	FBgn0003009	FBgn0003009	FBgn0003009	0.0E+00
	D. mojavensis	FBgn0143855	FBgn0143855	FBgn0143855	1.5E-116
	D. persimilis	FBgn0148894	FBgn0148894	FBgn0148894	9.6E-127
	D. pseudoobscura	FBgn0076207	FBgn0076207	FBgn0076207	1.4E-126
	D. sechilia	FBgn0170485	FBgn0170485	FBgn0170485	1.3E-236
	D. simulans	FBgn0196381	FBgn0196381	FBgn0196381	1.5E-302
	D. virilis	FBgn0208105	FBgn0208105	FBgn0208105	2.3E-116
	D. willistoni	FBgn0221554	-	FBgn0221554	4.4E-60
	D. yakuba	FBgn0229389	FBgn0229389	FBgn0229389	1.5E-266
C(2)M	*D. ananassae*	FBgn0091606	FBgn0091606	FBgn0091606	1.7E-104
	D. erecta	FBgn0117325	FBgn0117325	FBgn0117325	3.0E-295
	D. grimshawi	FBgn0117794	FBgn0117794	FBgn0117794	2.6E-66
	D. melanogaster	FBgn0028525	FBgn0028525	FBgn0028525	0.0E+00
	D. mojavensis	FBgn0139819	FBgn0139819	FBgn0139819	2.8E-76
	D. persimilis	FBgn0153925	FBgn0153925	FBgn0153925	2.3E-110
	D. pseudoobscura	FBgn0078066	FBgn0078066	FBgn0078066	1.0E-115
	D. sechilia	FBgn0173571	FBgn0173571	FBgn0173571	0.0E+00
	D. simulans	FBgn0195408	FBgn0195408	FBgn0195408	1.3E-272
	D. virilis	FBgn0197897	FBgn0197897	FBgn0197897	2.3E-116
	D. willistoni	-	FBgn0206237	FBgn0206237	4.4E-60
	D. yakuba	FBgn0238700	FBgn0238700	FBgn0238700	1.5E-266
C(3)G	*D. ananassae*	-	-	-	-
	D. erecta	FBgn0112578	FBgn0112578	FBgn0112578	0.0E+00
		FBgn0112568	FBgn0112568	FBgn0112568	5.0E-185
	D. grimshawi	-	-	-	-
	D. melanogaster	FBgn0000246	FBgn0000246	FBgn0000246	0.0E+00
	D. mojavensis	-	-	-	-
	D. persimilis	FBgn0161282	FBgn0161282	FBgn0161282	1.1E-25
	D. pseudoobscura	FBgn0248078	FBgn0248078	FBgn0248078	3.8E-08
	D. sechilia	FBgn0180610	FBgn0180610	FBgn0180610	0.0E+00
	D. simulans	FBgn0191803	FBgn0191803	FBgn0191803	0.0E+00
	D. virilis			FBgn0202545	6.9E-04
	D. willistoni	FBgn0212362	FBgn0212362	FBgn0212362	2.6E-10
	D. yakuba	FBgn0243389	FBgn0243389	FBgn0243389	2.5E-305
Corolla	*D. ananassae*	-	FBgn0099545	FBgn0099545	3.6E-08

	D. erecta	FBgn0110389	FBgn0110389	FBgn0110389	2.8E-200
	D. grimshawi	FBgn0120438	FBgn0120438	FBgn0120438	5.3E-17
	D. melanogaster	FBgn0030852	FBgn0030852	FBgn0030852	0.0E+00
	D. mojavensis	FBgn0137382	FBgn0137382	FBgn0137382	1.8E-13
	D. persimilis	FBgn0157987	FBgn0157987	FBgn0157987	2.6E-17
	D. pseudoobscura	FBgn0250506	FBgn0250506	FBgn0250506	7.0E-16
	D. sechilia	FBgn0168236	FBgn0168236	FBgn0168236	4.8E-291
	D. simulans	FBgn0188911	FBgn0188911	FBgn0188911	3.9E-292
	D. virilis	FBgn0206426	FBgn0206426	FBgn0206426	1.4E-17
	D. willistoni	FBgn0227567	FBgn0227567	FBgn0227567	2.5E-08
	D. yakuba	FBgn0233155	FBgn0233155	FBgn0233155	1.3E-217
Cona	*D. ananassae*	-	FBgn0097212	FBgn0097212	2.9E-03
	D. erecta	FBgn0115037	FBgn0115037	FBgn0115037	9.6E-67
	D. grimshawi	-	-	FBgn0132111	1.9E-03
	D. melanogaster	-	-	-	-
	D. mojavensis	-	-	FBgn0138867	1.8E-02
	D. persimilis	FBgn0157804	FBgn0157804	FBgn0157804	6.4E-10
	D. pseudoobscura	FBgn0248220	FBgn0248220	FBgn0248220	9.5E-11
	D. sechilia	FBgn0172774	FBgn0172774	FBgn0172774	2.1E-104
	D. simulans	FBgn0190734	FBgn0190734	FBgn0190734	7.8E-105
	D. virilis	-	-	FBgn0203581	1.3E-02
	D. willistoni	-	-	-	-
	D. yakuba	FBgn0242600	FBgn0242600	FBgn0242600	2.2E-69
		FBgn0232288	FBgn0232288	FBgn0232288	1.5E-65

Appendix 3: Ord orthologs confirmed by reciprocal BLAST back to the *D. melanogaster* genome.

Species	Accession Number	Method	Seed	E-value
D. melanogaster[a]	CG3134	tBLASTn		0.00[b]
D. simulans[a]	GD25609	tBLASTn	*D. melanogaster*	0.00[b]
D. mauritiana[a]	NA	tBLASTn	*D. melanogaster*	0.00[b]
D. sechellia[a]	GM15567	tBLASTn	*D. melanogaster*	0.00[b]
D. yakuba[a]	GE11589	tBLASTn	*D. melanogaster*	0.00[b]
D. erecta[a]	GG20053	tBLASTn	*D. melanogaster*	0.00[b]
D. eugracilis[a]	KB465221.1	tBLASTn	*D. melanogaster*	0.00[b]
D. takahashii[a]	KB461151.1	tBLASTn	*D. melanogaster*	0.00[b]
D. biarmipies[a]	KB462460.1	tBLASTn	*D. melanogaster*	0.00[b]
D. ficusphila[a]	KB457516.1	tBLASTn	*D. melanogaster*	0.00[b]
D. elegans[a]	KB458548.1	tBLASTn	*D. melanogaster*	0.00[b]
D. rhopaloa[a]	KB450382.1	tBLASTn	*D. melanogaster*	6.27E-196
D. bipectinata[a]	KB464224.1	tBLASTn	*D. melanogaster*	1.99E-154
D. ananassae[a]	GF13305	tBLASTn	*D. melanogaster*	4.28E-182
D. persimilis[a]	GL11285	tBLASTn	*D. melanogaster*	1.43E-157
D. miranda[a]	CM001519.2	tBLASTn	*D. melanogaster*	4.16E-92
D. pseudoobscura[a]	GA16191	tBLASTn	*D. melanogaster*	6.51E-160
D. willistoni[a]	GK19556	tBLASTn	*D. melanogaster*	2.72E-150
D. mojavensis[a]	GI21123	tBLASTn	*D. melanogaster*	2.30E-141
D. virilis[a]	GJ20970	tBLASTn	*D. melanogaster*	9.56E-133
D. grimshawi[a]	GH23052	tBLASTn	*D. melanogaster*	2.25E-148
B. cucurbitae	XP_011176821.1	BLASTp	*D. melanogaster*	1E-43
B. dorsalis	XP_011211900.1	BLASTp	*D. melanogaster*	3E-37
C. capitata	XP_004523724.1	BLASTp	*D. melanogaster*	2E-34
M. domestica	XP_011296419.1	BLASTp	*D. melanogaster*	9E-21
G. morsitans morsitans	CCAG010006519.1	BLASTp	*M. domestica*	1.08E-46

[a] The ortholog sequence used in the molecular evolutionary analyses
[b] E-values < E-200 were considered zero

Appendix 4: C(2)M orthologs confirmed by reciprocal BLAST back to the *D. melanogaster* genome.

Species	Accession Number	Method	Seed	E-value
D. melanogaster[a]	CG4249	tBLASTn		0.00[b]
D. simulans[a]	GD24050	tBLASTn	*D. melanogaster*	0.00[b]
D. mauritiana[a]	NA	tBLASTn	*D. melanogaster*	0.00[b]
D. sechellia[a]	GM18665	tBLASTn	*D. melanogaster*	0.00[b]
D. yakuba[a]	GE21442	tBLASTn	*D. melanogaster*	0.00[b]
D. erecta[a]	GG25201	tBLASTn	*D. melanogaster*	0.00[b]
D. eugracilis[a]	KB464450.1	tBLASTn	*D. melanogaster*	0.00[b]
D. takahashii[a]	KB461686.1	tBLASTn	*D. melanogaster*	0.00[b]
D. biarmipies[a]	KB462833.1	tBLASTn	*D. melanogaster*	0.00[b]
D. ficusphila[a]	KB457528.1	tBLASTn	*D. melanogaster*	5.44E-180
D. elegans[a]	KB458274.1	tBLASTn	*D. melanogaster*	0.00[b]
D. rhopaloa[a]	KB451894.1	tBLASTn	*D. melanogaster*	0.00[b]
D. bipectinata[a]	KB464241.1	tBLASTn	*D. melanogaster*	4.70E-97
D. ananassae[a]	GF14579	tBLASTn	*D. melanogaster*	9.62E-91
D. persimilis[a]	GL16321	tBLASTn	*D. melanogaster*	7.09E-94
D. miranda[a]	CM001520.2	tBLASTn	*D. melanogaster*	7.84E-96
D. pseudoobscura[a]	GA18058	tBLASTn	*D. melanogaster*	6.29E-59
D. willistoni[a]	GK23985/partial	tBLASTn	*D. melanogaster*	4.25E-20
D. mojavensis[a]	GI17074	tBLASTn	*D. melanogaster*	2.80E-39
D. virilis[a]	GJ16321	tBLASTn	*D. melanogaster*	4.49E-23
D. grimshawi[a]	GH10313	tBLASTn	*D. melanogaster*	2.60E-23
B. cucurbitae	XP_011190022.1	BLASTp	*D. melanogaster*	6E-03
M. domestica	XP_011292775.1	BLASTp	*D. melanogaster*	1E-04
B. dorsalis	XP_011207064.1	BLASTp	*D. melanogaster*	0.034
C. capitata	XP_004523724.1	BLASTp	*D. melanogaster*	2E-34
M. domestica	XP_011292775.1	BLASTp	*D. melanogaster*	9E-21
G. morsitans morsitans	CCAG010009633.1	BLASTp	*M. domestica*	1.08E-46

[a] The ortholog sequence used in the molecular evolutionary analyses
[b] E-values < E-200 were considered zero

Appendix 5: C(3)G orthologs confirmed by reciprocal BLAST back to the *D. melanogaster* genome.

Species	Accession Number	Method	Seed	E-value
D. melanogaster[a]	CG17604	tBLASTn		0.00[b]
D. simulans[a]	GD20329	tBLASTn	*D. melanogaster*	0.00[b]
D. mauritiana[a]	NA	tBLASTn	*D. melanogaster*	0.00[b]
D. sechellia[a]	GM25754	tBLASTn	*D. melanogaster*	0.00[b]
D. yakuba[a]	GE26360	tBLASTn	*D. melanogaster*	0.00[b]
D. erecta[a]	GG20377	tBLASTn	*D. melanogaster*	0.00[b]
	GG20388	tBLASTn	*D. melanogaster*	0.00[b]
D. eugracilis[a]	KB465333.1	tBLASTn	*D. melanogaster*	0.00[b]
D. takahashii[a]	KB461113.1	tBLASTn	*D. melanogaster*	0.00[b]
D. biarmipies[a]	KB462598.1	tBLASTn	*D. melanogaster*	0.00[b]
D. ficusphila[a]	KB457030.1	tBLASTn	*D. melanogaster*	7.06E-170
D. elegans[a]	KB458458.1	tBLASTn	*D. melanogaster*	3.04E-166
D. rhopaloa[a]	KB448329.1	tBLASTn	*D. melanogaster*	4.97E-192
D. bipectinata[a]	KB464131.1	tBLASTn	*D. melanogaster*	1.29E-02
D. ananassae[a]	GF26923	tBLASTn	*D. melanogaster*	4.44E-22
D. persimilis[a]	GL23692	tBLASTn	*D. melanogaster*	1.32E-02
D. miranda[a]	CM001528.2	tBLASTn	*D. melanogaster*	6.59E-03
D. pseudoobscura[a]	GA26705	tBLASTn	*D. melanogaster*	3.51E-02
D. willistoni	GK10347	BLASTp	*D. melanogaster*	3.95E-05
D. mojavensis	GI14995	BLASTp	*D. virilis*	1.95E-77
D. virilis	GJ15351	BLASTp	*D. melanogaster*	4.15E-05
D. grimshawi	GH12738	BLASTp	*D. virilis*	2.04E-64
B. cucurbitae	-	-	-	-
M. domestica	-	-	-	-
B. dorsalis	-	-	-	-
C. capitata	-	-	-	-
M. domestica	-	-	-	-
G. morsitans morsitans	-	-	-	-

[a] The ortholog sequence used in the molecular evolutionary analyses
[b] E-values < E-200 were considered zero

Appendix 6: Corolla orthologs confirmed by reciprocal BLAST back to the *D. melanogaster* genome.

Species	Accession Number	Method	Seed	E-value
D. melanogaster[a]	CG8316	tBLASTn		0.00[b]
D. simulans[a]	GD17351	tBLASTn	*D. melanogaster*	0.00[b]
D. mauritiana[a]	NA	tBLASTn	*D. melanogaster*	0.00[b]
D. sechellia[a]	GM13305	tBLASTn	*D. melanogaster*	0.00[b]
D. yakuba[a]	GE15583	tBLASTn	*D. melanogaster*	2.43E-144
D. erecta[a]	GG18173	tBLASTn	*D. melanogaster*	1.36E-129
D. eugracilis[a]	AFPQ02005309.1	tBLASTn	*D. melanogaster*	3.70E-150
D. takahashii[a]	KB461135.1	tBLASTn	*D. melanogaster*	3.56E-42
D. biarmipies[a]	KB462463.1	tBLASTn	*D. melanogaster*	1.09E-45
D. ficusphila[a]	KB457527.1	tBLASTn	*D. melanogaster*	2.19E-37
D. elegans[a]	KB458387.1	tBLASTn	*D. melanogaster*	2.13E-31
D. rhopaloa[a]	KB451800.1	tBLASTn	*D. melanogaster*	3.98E-62
D. bipectinata	AFFE01006525.1	tBLASTn	*D. ananassae*	0.00[b]
D. ananassae	GF22551	BLASTp	*D. melanogaster*	6.36E-07
D. persimilis[a]	GL20392	tBLASTn	*D. miranda*	0.00[b]
D. miranda[a]	CM001516.2	tBLASTn	*D. melanogaster*	1.56E-04
D. pseudoobscura[a]	GA29148	tBLASTn	*D. miranda*	0.00
D. willistoni[a]	GK25608	BLASTp	*D. melanogaster*	1.46E-08
D. mojavensis[a]	GI14631	tBLASTn	*D. melanogaster*	4.12E-16
D. virilis[a]	GJ19282	tBLASTn	*D. melanogaster*	1.59E-08
D. grimshawi[a]	GH12960	tBLASTn	*D. melanogaster*	1.94E-06
B. cucurbitae	-	-	-	-
M. domestica	-	-	-	-
B. dorsalis	-	-	-	-
C. capitata	-	-	-	-
M. domestica	-	-	-	-
G. morsitans morsitans	-	-	-	-

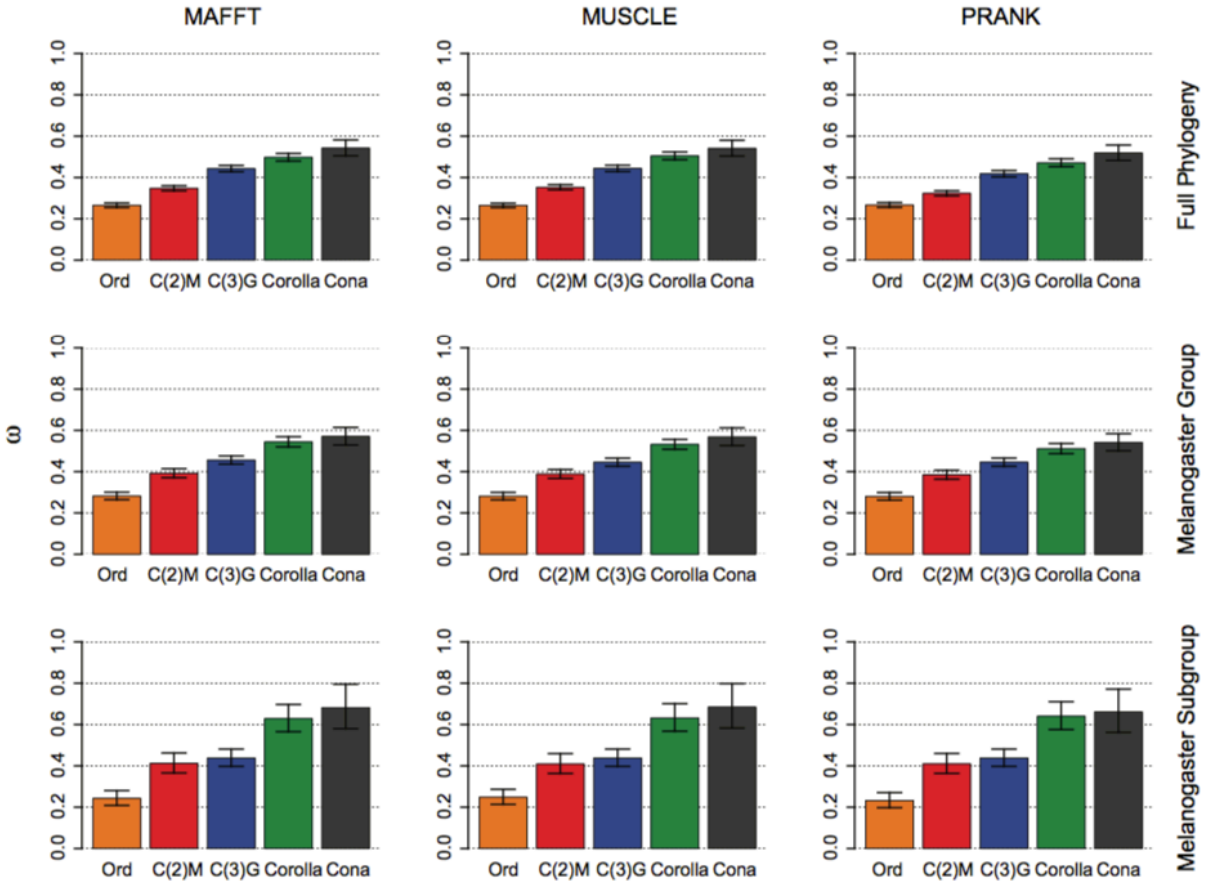
[a] The ortholog sequence used in the molecular evolutionary analyses
[b] E-values < E-200 were considered zero

Appendix 7: Cona orthologs confirmed by reciprocal BLAST back to the *D. melanogaster* genome.

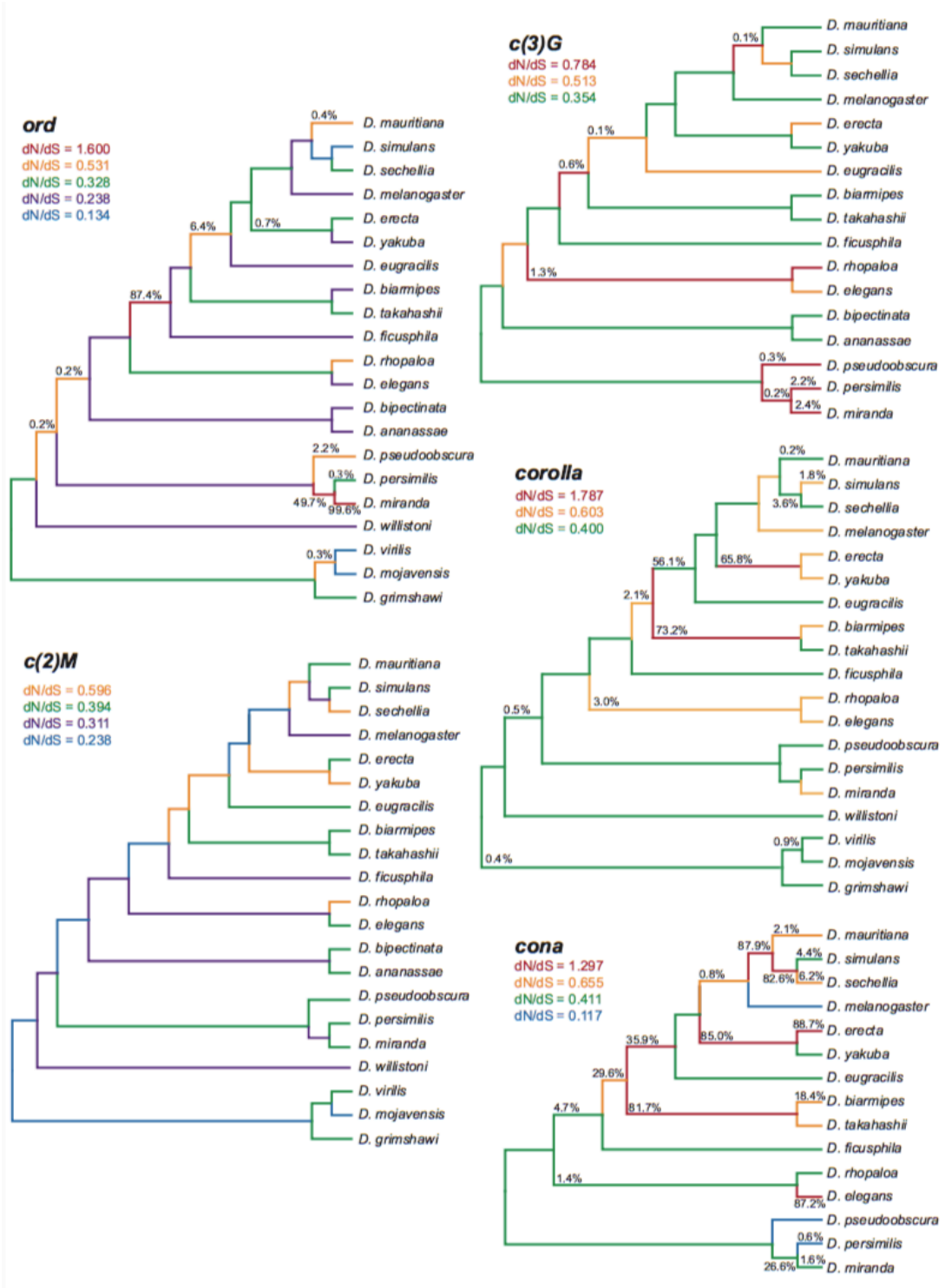
Species	Accession Number	Method	Seed	E-value
D. melanogaster[a]	CG7676	tBLASTn		0.00[b]
D. simulans[a]	GD19229	tBLASTn	*D. melanogaster*	4.18E-76
D. mauritiana[a]	NA	tBLASTn	*D. melanogaster*	1.27E-77
D. sechellia[a]	GM17867	tBLASTn	*D. melanogaster*	3.47E-77
D. yakuba[a]	GE25535	tBLASTn	*D. melanogaster*	1.08E-102
	GE14694	tBLASTn	*D. melanogaster*	6.46E-101
D. erecta[a]	GG22880	tBLASTn	*D. melanogaster*	1.99E-92
D. eugracilis[a]	KB465338.1	tBLASTn	*D. melanogaster*	1.59E-38
D. takahashii[a]	KB461676.1	tBLASTn	*D. melanogaster*	1.70E-09
D. biarmipies[a]	KB462068.1	tBLASTn	*D. melanogaster*	1.40E-11
D. ficusphila[a]	KB457292.1	tBLASTn	*D. melanogaster*	2.04E-12
D. elegans[a]	KB458397.1	tBLASTn	*D. melanogaster*	6.37E-12
D. rhopaloa[a]	KB452427.1	tBLASTn	*D. melanogaster*	8.61E-15
D. bipectinata	AFFE01007043.1	tBLASTn	*D. ananassae*	7.37E-82
D. ananassae	GF20205	BLASTp	*D. melanogaster*	3.00E-03
D. persimilis[a]	GL20209	tBLASTn	*D. ficusphila*	4.91E-02
D. miranda[a]	CM001516.2	tBLASTn	*D. ficusphila*	8.50E-02
D. pseudoobscura[a]	GA26847	tBLASTn	*D. ficusphila*	4.91E-02
D. willistoni	-	-	-	-
D. mojavensis	GI16118	BLASTp	*D. melanogaster*	6.57E-02
D. virilis	GJ20698	BLASTp	*D. melanogaster*	3.18E-02
D. grimshawi	GH24655	BLASTp	*D. melanogaster*	6.09E-03
B. cucurbitae	-	-	-	-
M. domestica	-	-	-	-
B. dorsalis	-	-	-	-
C. capitata	-	-	-	-
M. domestica	-	-	-	-
G. morsitans morsitans	-	-	-	-

[a] The ortholog sequence used in the molecular evolutionary analyses
[b] E-values < E-200 were considered zero

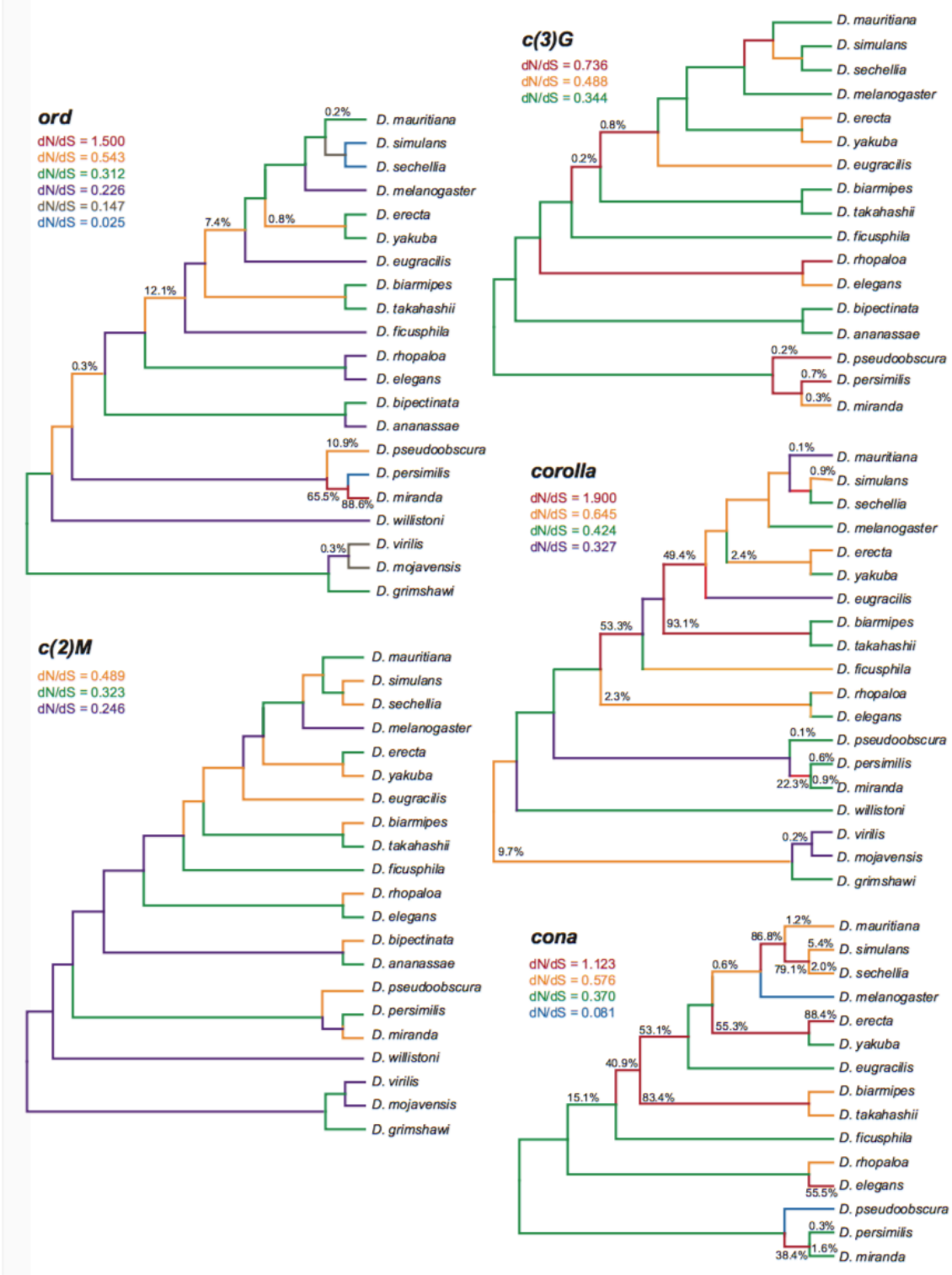
Appendix 8: The global ω of each SC gene calculated in HyPhy using a GTR nucleotide substitution model with 95% confidence intervals. The ratio remains relatively consistent for each alignment program used (MAFFT, MUSCLE, and PRANK) and divergence times. It is also consistent with the PAML-derived results albeit slightly higher ω estimates.



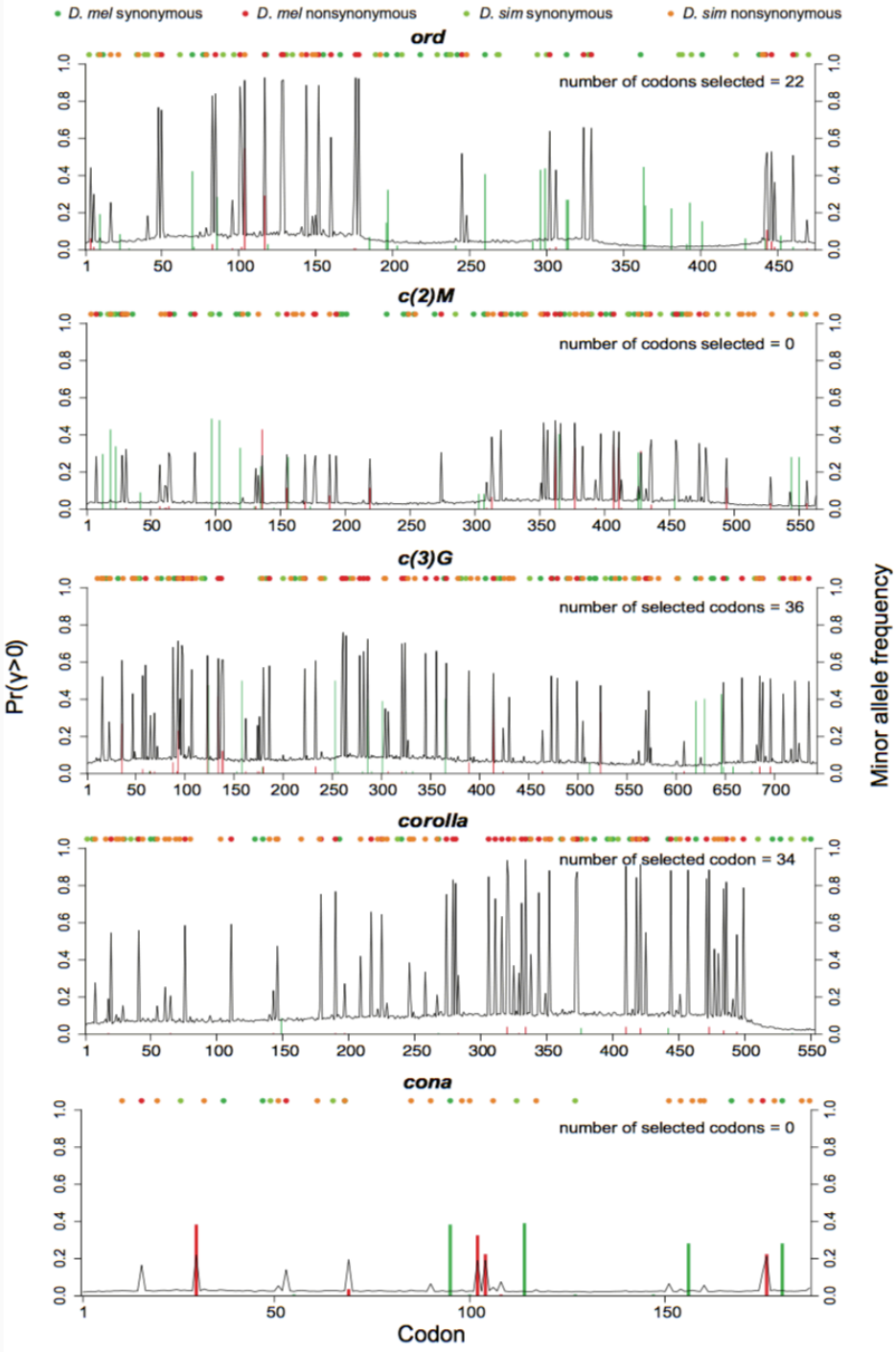
Appendix 9: MUSCLE-aligned GA Branch diagrams of A) Ord, B) C(2)M, C) C(3)G, D) Corolla, and E) Cona. Branch colors correspond with the associated ω ratio and posterior probabilities of positive selection are listed as a percent on each branch.



Appendix 10: PRANK-aligned GA Branch diagrams of A) Ord, B) C(2)M, C) C(3)G, D) Corolla, and E) Cona. Branch colors correspond with the associated ω ratio and posterior probabilities of positive selection are listed as a percent on each branch.



Appendix 11: GammaMap figures of the DGRP (North Carolina) *D. melanogaster* population sequences. In concordance with Wilson *et al.* 2011, a codon is under significant signature of selection when the posterior probabilities of selection (lines) are greater than 0.5. Vertical bars illustrate polymorphisms in D. melanogaster and the substitutions are the circular dots. The colors correspond to *D. melanogaster* non-synonymous (red) and synonymous (dark green) variants as well as *D. simulans* non-synonymous (orange) and synonymous (light green) variants. Estimated number of selected codons is indicated in the upper right of each plot. Additionally, a red bar indicates the region of the hypothesized selective sweep in *corolla*.



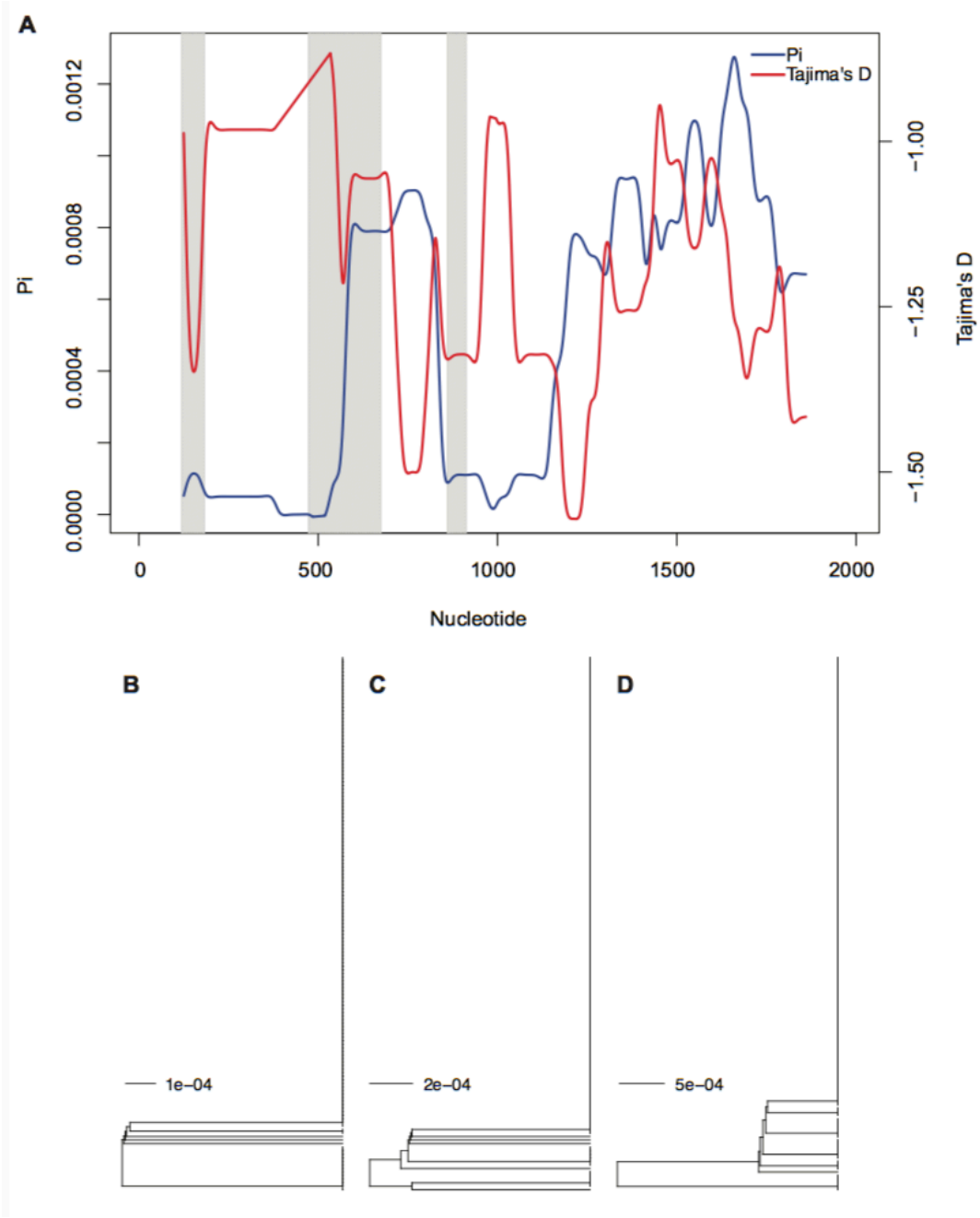
Appendix 12: Population parameters of the DGRP and DGPG samples including non-synonymous pairwise diversity (π_N), synonymous pairwise diversity (π_S), and Tajima's D test of neutrality.

		π_N	π_S	Tajima's D	*p*-value
Ord	N. Carolina	0.00169	0.02240	0.275	0.784
	Africa	0.00193	0.02943	-0.585	0.559
C(2)M	N. Carolina	0.00283	0.01718	0.653	0.514
	Africa	0.00170	0.00973	-0.136	0.892
C(3)G	N. Carolina	0.00224	0.01100	-0.293	0.770
	Africa	0.00230	0.01358	-1.169	0.242
Corolla	N. Carolina	0.00041	0.00091	-2.055	**0.040**
	Africa	0.00114	0.00328	-2.443	**0.015**
Cona	N. Carolina	0.00383	0.01579	1.023	0.306
	Africa	0.00244	0.02489	-0.527	0.598
Meiosis Means[a]	N. America	0.001	0.013	-	-
	Africa	0.002	0.020	-	-

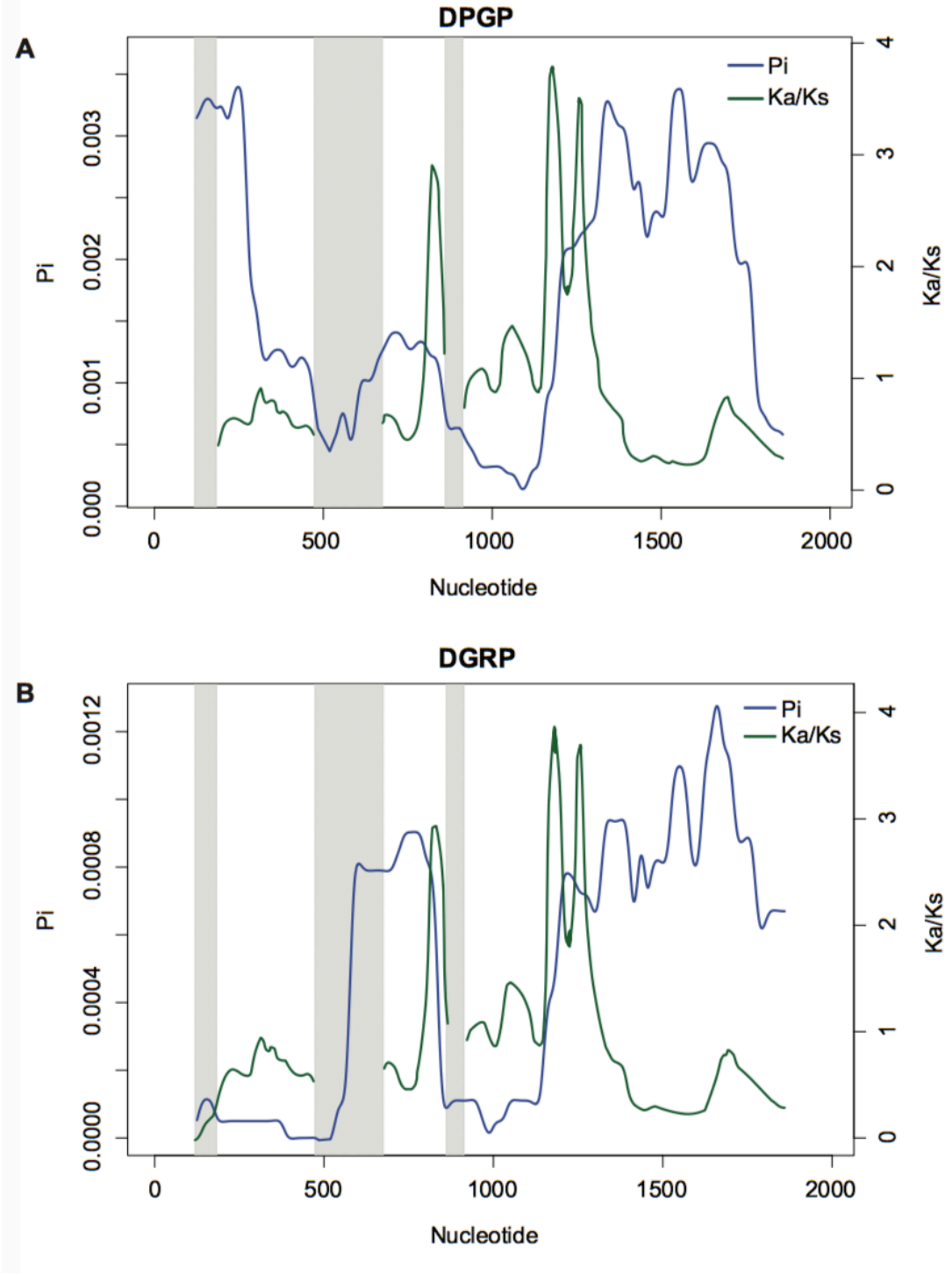
***** Significant *p* values in bold

[a] Mean nucleotide diversity values for all meiotic genes in Anderson *et al*. 2009 included for comparison.

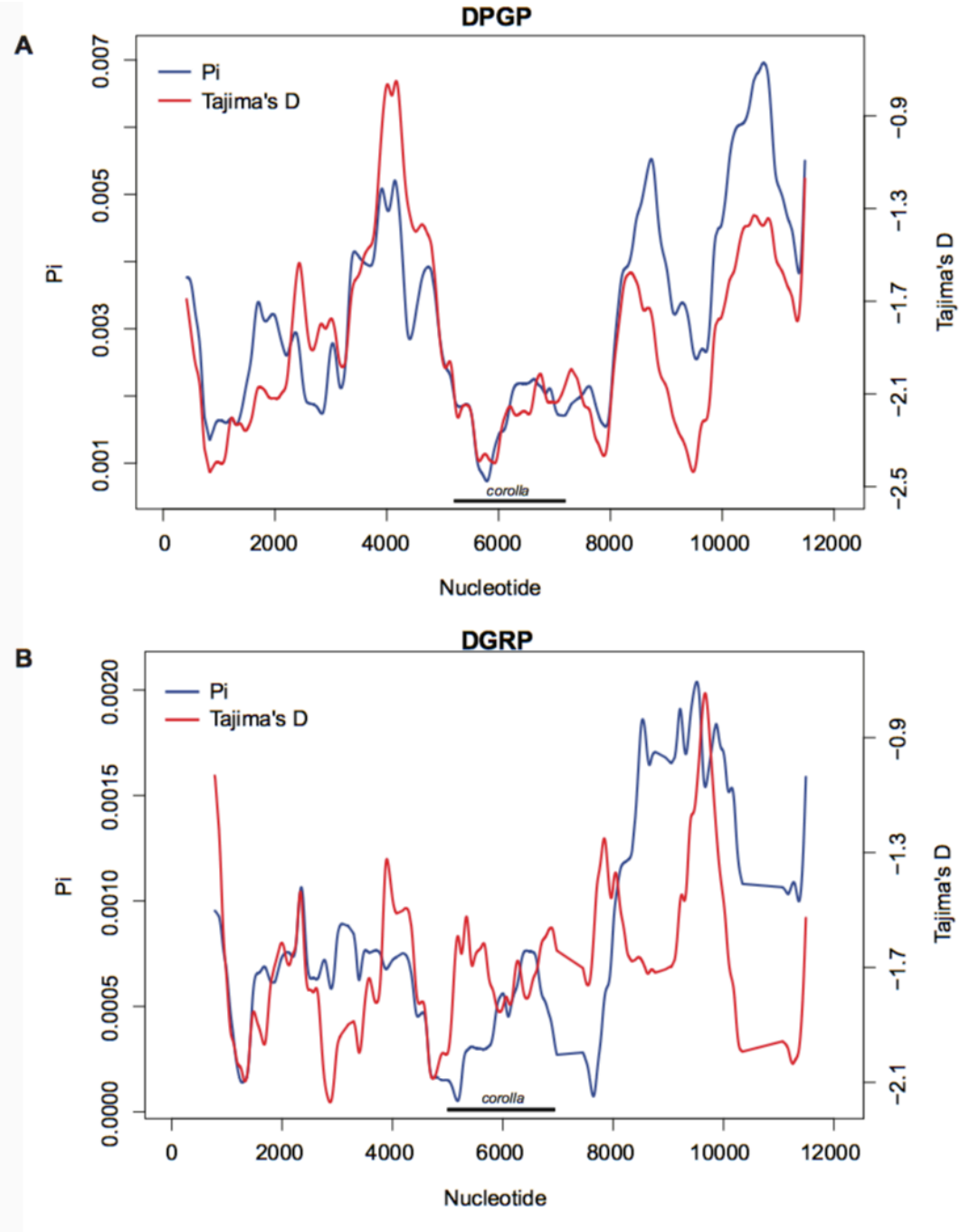
Appendix 13: Sliding window estimates of pairwise divergence and Tajima's D reveal recent positive selection resulting in loss of haplotype diversity. (A) Pairwise differences (π) and Tajima's D measured in 250 bp windows along the length of *corolla* within the DGRP sequences. Introns are indicated in gray bars. Black lines indicate portions of the gene used in the dendrograms for parts B, C, and D. (B-D) Dendrograms constructed using a HKY model of UPGMA between nucleotides 1-700 (B), 701-1300 (C), and 1301-1938 (D) downstream of the translation start site.



Appendix 14: Sliding window estimates of pairwise divergence and K_a/K_s show little correlation in selection between *D. melanogaster* and *D. simulans*. Pairwise differences (π) and divergence (K_a/K_s) measured in 250 bp windows along the length of *corolla* in the DGPG sequences (A) and the DGRP sequences (B). Introns are indicated in gray bars.



Appendix 15: Sliding window estimates of pairwise divergence and Tajima's D reveal reduced polymorphism surrounding *corolla* in Africa but not North Carolina. Pairwise differences (π) and Tajima's D measured in 800 bp windows along the genomic region containing *corolla* and 5 kb upstream and downstream of the gene in the DGPG sequences (A) and DGRP sequences (B). Introns are indicated in gray bars and the length of the *corolla* gene from the starting codon to the stop codon is indicated by the black bar.



Appendix 16: i7 primers used for Tn5 tagging and PCR amplification for multiplex shotgun sequencing.

NAME	i7 Index Barcode	PRIMER (5'-3')	i7 Demultiplex Orientation
jpbhb701	AACGTGAT	CAAGCAGAAGACGGCATACGAGAT AACGTGATGTCTCGTGGGCTCGG	ATCACGTT
jpbhb702	AAACATCG	CAAGCAGAAGACGGCATACGAGAT AAACATCGGTCTCGTGGGCTCGG	CGATGTTT
jpbhb703	ATGCCTAA	CAAGCAGAAGACGGCATACGAGAT ATGCCTAAGTCTCGTGGGCTCGG	TTAGGCAT
jpbhb704	AGTGGTCA	CAAGCAGAAGACGGCATACGAGAT AGTGGTCAGTCTCGTGGGCTCGG	TGACCACT
jpbhb705	ACCACTGT	CAAGCAGAAGACGGCATACGAGAT ACCACTGTGTCTCGTGGGCTCGG	ACAGTGGT
jpbhb706	ACATTGGC	CAAGCAGAAGACGGCATACGAGAT ACATTGGCGTCTCGTGGGCTCGG	GCCAATGT
jpbhb707	CAGATCTG	CAAGCAGAAGACGGCATACGAGAT CAGATCTGGTCTCGTGGGCTCGG	CAGATCTG
jpbhb708	CATCAAGT	CAAGCAGAAGACGGCATACGAGAT CATCAAGTGTCTCGTGGGCTCGG	ACTTGATG
jpbhb709	CGCTGATC	CAAGCAGAAGACGGCATACGAGAT CGCTGATCGTCTCGTGGGCTCGG	GATCAGCG
jpbhb710	ACAAGCTA	CAAGCAGAAGACGGCATACGAGAT ACAAGCTAGTCTCGTGGGCTCGG	TAGCTTGT
jpbhb711	CTGTAGCC	CAAGCAGAAGACGGCATACGAGAT CTGTAGCCGTCTCGTGGGCTCGG	GGCTACAG
jpbhb712	AGTACAAG	CAAGCAGAAGACGGCATACGAGAT AGTACAAGGTCTCGTGGGCTCGG	CTTGTACT
jpbhb713	AACAACCA	CAAGCAGAAGACGGCATACGAGAT AACAACCAGTCTCGTGGGCTCGG	TGGTTGTT
jpbhb714	AACCGAGA	CAAGCAGAAGACGGCATACGAGAT AACCGAGAGTCTCGTGGGCTCGG	TCTCGGTT
jpbhb715	AACGCTTA	CAAGCAGAAGACGGCATACGAGAT AACGCTTAGTCTCGTGGGCTCGG	TAAGCGTT
jpbhb716	AAGACGGA	CAAGCAGAAGACGGCATACGAGAT AAGACGGAGTCTCGTGGGCTCGG	TCCGTCTT
jpbhb717	AAGGTACA	CAAGCAGAAGACGGCATACGAGAT AAGGTACAGTCTCGTGGGCTCGG	TGTACCTT
jpbhb718	ACACAGAA	CAAGCAGAAGACGGCATACGAGAT ACACAGAAGTCTCGTGGGCTCGG	TTCTGTGT
jpbhb719	ACAGCAGA	CAAGCAGAAGACGGCATACGAGAT ACAGCAGAGTCTCGTGGGCTCGG	TCTGCTGT
jpbhb720	ACCTCCAA	CAAGCAGAAGACGGCATACGAGAT ACCTCCAAGTCTCGTGGGCTCGG	TTGGAGGT
jpbhb721	ACGCTCGA	CAAGCAGAAGACGGCATACGAGAT ACGCTCGAGTCTCGTGGGCTCGG	TCGAGCGT
jpbhb722	ACGTATCA	CAAGCAGAAGACGGCATACGAGAT ACGTATCAGTCTCGTGGGCTCGG	TGATACGT
jpbhb723	ACTATGCA	CAAGCAGAAGACGGCATACGAGAT ACTATGCAGTCTCGTGGGCTCGG	TGCATAGT
jpbhb724	AGAGTCAA	CAAGCAGAAGACGGCATACGAGAT AGAGTCAAGTCTCGTGGGCTCGG	TTGACTCT
jpbhb725	AGATCGCA	CAAGCAGAAGACGGCATACGAGAT AGATCGCAGTCTCGTGGGCTCGG	TGCGATCT
jpbhb726	AGCAGGAA	CAAGCAGAAGACGGCATACGAGAT	TTCCTGCT

		AGCAGGAAGTCTCGTGGGCTCGG	
jpbhb727	AGTCACTA	CAAGCAGAAGACGGCATACGAGAT AGTCACTAGTCTCGTGGGCTCGG	TAGTGACT
jpbhb728	ATCCTGTA	CAAGCAGAAGACGGCATACGAGAT ATCCTGTAGTCTCGTGGGCTCGG	TACAGGAT
jpbhb729	ATTGAGGA	CAAGCAGAAGACGGCATACGAGAT ATTGAGGAGTCTCGTGGGCTCGG	TCCTCAAT
jpbhb730	CAACCACA	CAAGCAGAAGACGGCATACGAGAT CAACCACAGTCTCGTGGGCTCGG	TGTGGTTG
jpbhb731	GACTAGTA	CAAGCAGAAGACGGCATACGAGAT GACTAGTAGTCTCGTGGGCTCGG	TACTAGTC
jpbhb732	CAATGGAA	CAAGCAGAAGACGGCATACGAGAT CAATGGAAGTCTCGTGGGCTCGG	TTCCATTG
jpbhb733	CACTTCGA	CAAGCAGAAGACGGCATACGAGAT CACTTCGAGTCTCGTGGGCTCGG	TCGAAGTG
jpbhb734	CAGCGTTA	CAAGCAGAAGACGGCATACGAGAT CAGCGTTAGTCTCGTGGGCTCGG	TAACGCTG
jpbhb735	CATACCAA	CAAGCAGAAGACGGCATACGAGAT CATACCAAGTCTCGTGGGCTCGG	TTGGTATG
jpbhb736	CCAGTTCA	CAAGCAGAAGACGGCATACGAGAT CCAGTTCAGTCTCGTGGGCTCGG	TGAACTGG
jpbhb737	CCGAAGTA	CAAGCAGAAGACGGCATACGAGAT CCGAAGTAGTCTCGTGGGCTCGG	TACTTCGG
jpbhb738	CCGTGAGA	CAAGCAGAAGACGGCATACGAGAT CCGTGAGAGTCTCGTGGGCTCGG	TCTCACGG
jpbhb739	CCTCCTGA	CAAGCAGAAGACGGCATACGAGAT CCTCCTGAGTCTCGTGGGCTCGG	TCAGGAGG
jpbhb740	CGAACTTA	CAAGCAGAAGACGGCATACGAGAT CGAACTTAGTCTCGTGGGCTCGG	TAAGTTCG
jpbhb741	CGACTGGA	CAAGCAGAAGACGGCATACGAGAT CGACTGGAGTCTCGTGGGCTCGG	TCCAGTCG
jpbhb742	CGCATACA	CAAGCAGAAGACGGCATACGAGAT CGCATACAGTCTCGTGGGCTCGG	TGTATGCG
jpbhb743	CTCAATGA	CAAGCAGAAGACGGCATACGAGAT CTCAATGAGTCTCGTGGGCTCGG	TCATTGAG
jpbhb744	CTGAGCCA	CAAGCAGAAGACGGCATACGAGAT CTGAGCCAGTCTCGTGGGCTCGG	TGGCTCAG
jpbhb745	CTGGCATA	CAAGCAGAAGACGGCATACGAGAT CTGGCATAGTCTCGTGGGCTCGG	TATGCCAG
jpbhb746	GAATCTGA	CAAGCAGAAGACGGCATACGAGAT GAATCTGAGTCTCGTGGGCTCGG	TCAGATTC
jpbhb747	CAAGACTA	CAAGCAGAAGACGGCATACGAGAT CAAGACTAGTCTCGTGGGCTCGG	TAGTCTTG
jpbhb748	GAGCTGAA	CAAGCAGAAGACGGCATACGAGAT GAGCTGAAGTCTCGTGGGCTCGG	TTCAGCTC
jpbhb749	GATAGACA	CAAGCAGAAGACGGCATACGAGAT GATAGACAGTCTCGTGGGCTCGG	TGTCTATC
jpbhb750	GCCACATA	CAAGCAGAAGACGGCATACGAGAT GCCACATAGTCTCGTGGGCTCGG	TATGTGGC
jpbhb751	GCGAGTAA	CAAGCAGAAGACGGCATACGAGAT GCGAGTAAGTCTCGTGGGCTCGG	TTACTCGC
jpbhb752	GCTAACGA	CAAGCAGAAGACGGCATACGAGAT GCTAACGAGTCTCGTGGGCTCGG	TCGTTAGC
jpbhb753	GCTCGGTA	CAAGCAGAAGACGGCATACGAGAT GCTCGGTAGTCTCGTGGGCTCGG	TACCGAGC

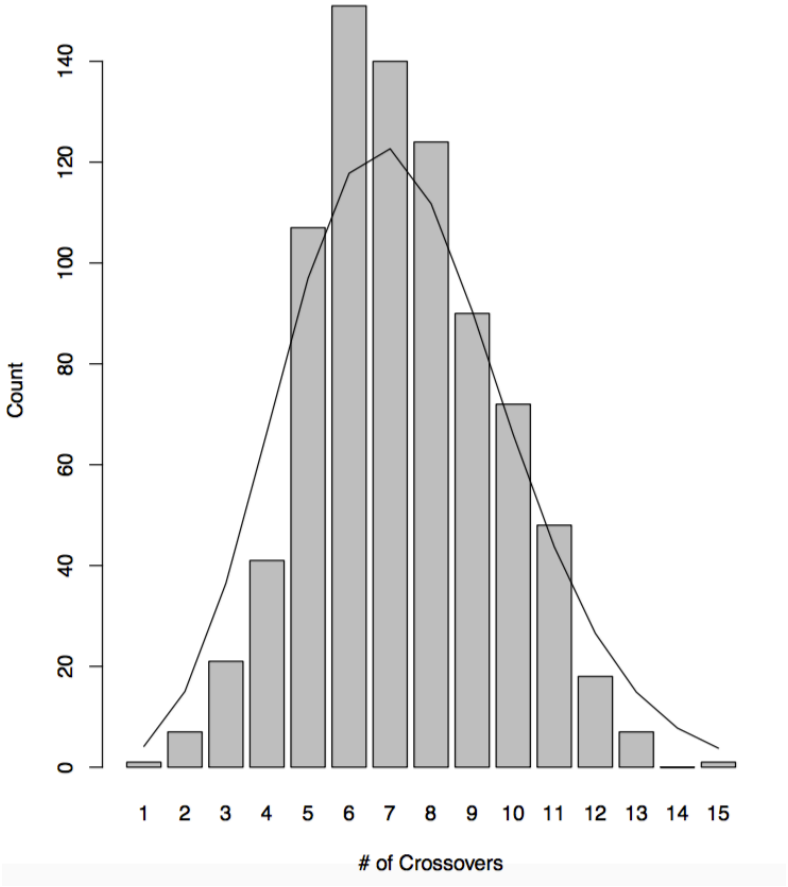
jpbhb754	GGAGAACA	CAAGCAGAAGACGGCATACGAGAT GGAGAACAGTCTCGTGGGCTCGG	TGTTCTCC
jpbhb755	GGTGCGAA	CAAGCAGAAGACGGCATACGAGAT GGTGCGAAGTCTCGTGGGCTCGG	TTCGCACC
jpbhb756	GTACGCAA	CAAGCAGAAGACGGCATACGAGAT GTACGCAAGTCTCGTGGGCTCGG	TTGCGTAC
jpbhb757	GTCGTAGA	CAAGCAGAAGACGGCATACGAGAT GTCGTAGAGTCTCGTGGGCTCGG	TCTACGAC
jpbhb758	GTCTGTCA	CAAGCAGAAGACGGCATACGAGAT GTCTGTCAGTCTCGTGGGCTCGG	TGACAGAC
jpbhb759	GTGTTCTA	CAAGCAGAAGACGGCATACGAGAT GTGTTCTAGTCTCGTGGGCTCGG	TAGAACAC
jpbhb760	TAGGATGA	CAAGCAGAAGACGGCATACGAGAT TAGGATGAGTCTCGTGGGCTCGG	TCATCCTA
jpbhb761	TATCAGCA	CAAGCAGAAGACGGCATACGAGAT TATCAGCAGTCTCGTGGGCTCGG	TGCTGATA
jpbhb762	TCCGTCTA	CAAGCAGAAGACGGCATACGAGAT TCCGTCTAGTCTCGTGGGCTCGG	TAGACGGA
jpbhb763	TCTTCACA	CAAGCAGAAGACGGCATACGAGAT TCTTCACAGTCTCGTGGGCTCGG	TGTGAAGA
jpbhb764	TGAAGAGA	CAAGCAGAAGACGGCATACGAGAT TGAAGAGAGTCTCGTGGGCTCGG	TCTCTTCA
jpbhb765	TGGAACAA	CAAGCAGAAGACGGCATACGAGAT TGGAACAAGTCTCGTGGGCTCGG	TTGTTCCA
jpbhb766	TGGCTTCA	CAAGCAGAAGACGGCATACGAGAT TGGCTTCAGTCTCGTGGGCTCGG	TGAAGCCA
jpbhb767	TGGTGGTA	CAAGCAGAAGACGGCATACGAGAT TGGTGGTAGTCTCGTGGGCTCGG	TACCACCA
jpbhb768	TTCACGCA	CAAGCAGAAGACGGCATACGAGAT TTCACGCAGTCTCGTGGGCTCGG	TGCGTGAA
jpbhb769	AACTCACC	CAAGCAGAAGACGGCATACGAGAT AACTCACCGTCTCGTGGGCTCGG	GGTGAGTT
jpbhb770	AAGAGATC	CAAGCAGAAGACGGCATACGAGAT AAGAGATCGTCTCGTGGGCTCGG	GATCTCTT
jpbhb771	AAGGACAC	CAAGCAGAAGACGGCATACGAGAT AAGGACACGTCTCGTGGGCTCGG	GTGTCCTT
jpbhb772	AATCCGTC	CAAGCAGAAGACGGCATACGAGAT AATCCGTCGTCTCGTGGGCTCGG	GACGGATT
jpbhb773	AATGTTGC	CAAGCAGAAGACGGCATACGAGAT AATGTTGCGTCTCGTGGGCTCGG	GCAACATT
jpbhb774	ACACGACC	CAAGCAGAAGACGGCATACGAGAT ACACGACCGTCTCGTGGGCTCGG	GGTCGTGT
jpbhb775	ACAGATTC	CAAGCAGAAGACGGCATACGAGAT ACAGATTCGTCTCGTGGGCTCGG	GAATCTGT
jpbhb776	AGATGTAC	CAAGCAGAAGACGGCATACGAGAT AGATGTACGTCTCGTGGGCTCGG	GTACATCT
jpbhb777	AGCACCTC	CAAGCAGAAGACGGCATACGAGAT AGCACCTCGTCTCGTGGGCTCGG	GAGGTGCT
jpbhb778	AGCCATGC	CAAGCAGAAGACGGCATACGAGAT AGCCATGCGTCTCGTGGGCTCGG	GCATGGCT
jpbhb779	AGGCTAAC	CAAGCAGAAGACGGCATACGAGAT AGGCTAACGTCTCGTGGGCTCGG	GTTAGCCT
jpbhb780	ATAGCGAC	CAAGCAGAAGACGGCATACGAGAT ATAGCGACGTCTCGTGGGCTCGG	GTCGCTAT
jpbhb781	ATCATTCC	CAAGCAGAAGACGGCATACGAGAT ATCATTCCGTCTCGTGGGCTCGG	GGAATGAT

jpbhb782	ATTGGCTC	CAAGCAGAAGACGGCATACGAGAT ATTGGCTCGTCTCGTGGGCTCGG	GAGCCAAT
jpbhb783	CAAGGAGC	CAAGCAGAAGACGGCATACGAGAT CAAGGAGCGTCTCGTGGGCTCGG	GCTCCTTG
jpbhb784	CACCTTAC	CAAGCAGAAGACGGCATACGAGAT CACCTTACGTCTCGTGGGCTCGG	GTAAGGTG
jpbhb785	CCATCCTC	CAAGCAGAAGACGGCATACGAGAT CCATCCTCGTCTCGTGGGCTCGG	GAGGATGG
jpbhb786	CCGACAAC	CAAGCAGAAGACGGCATACGAGAT CCGACAACGTCTCGTGGGCTCGG	GTTGTCGG
jpbhb787	CCTAATCC	CAAGCAGAAGACGGCATACGAGAT CCTAATCCGTCTCGTGGGCTCGG	GGATTAGG
jpbhb788	CCTCTATC	CAAGCAGAAGACGGCATACGAGAT CCTCTATCGTCTCGTGGGCTCGG	GATAGAGG
jpbhb789	CGACACAC	CAAGCAGAAGACGGCATACGAGAT CGACACACGTCTCGTGGGCTCGG	GTGTGTCG
jpbhb790	CGGATTGC	CAAGCAGAAGACGGCATACGAGAT CGGATTGCGTCTCGTGGGCTCGG	GCAATCCG
jpbhb791	CTAAGGTC	CAAGCAGAAGACGGCATACGAGAT CTAAGGTCGTCTCGTGGGCTCGG	GACCTTAG
jpbhb792	GAACAGGC	CAAGCAGAAGACGGCATACGAGAT GAACAGGCGTCTCGTGGGCTCGG	GCCTGTTC
jpbhb793	GACAGTGC	CAAGCAGAAGACGGCATACGAGAT GACAGTGCGTCTCGTGGGCTCGG	GCACTGTC
jpbhb794	GAGTTAGC	CAAGCAGAAGACGGCATACGAGAT GAGTTAGCGTCTCGTGGGCTCGG	GCTAACTC
jpbhb795	GATGAATC	CAAGCAGAAGACGGCATACGAGAT GATGAATCGTCTCGTGGGCTCGG	GATTCATC
jpbhb796	GCCAAGAC	CAAGCAGAAGACGGCATACGAGAT GCCAAGACGTCTCGTGGGCTCGG	GTCTTGGC

Appendix 17: i5 primers used for Tn5 tagging and PCR amplification for multiplex shotgun sequencing.

NAME	i5 Index Barcode	FINAL PRIMER (5'-3')	i5 Demultiplex Orientation
sjmhb501	TAGATCGC	AATGATACGGCGACCACCGAGATCTA CACTAGATCGCTCGTCGGCAGCGTC	TAGATCGC
sjmhb502	CTCTCTAT	AATGATACGGCGACCACCGAGATCTA CACCTCTCTATTCGTCGGCAGCGTC	CTCTCTAT
sjmhb503	TATCCTCT	AATGATACGGCGACCACCGAGATCTA CACTATCCTCTTCGTCGGCAGCGTC	TATCCTCT
sjmhb504	AGAGTAGA	AATGATACGGCGACCACCGAGATCTA CACAGAGTAGATCGTCGGCAGCGTC	AGAGTAGA
sjmhb505	GTAAGGAG	AATGATACGGCGACCACCGAGATCTA CACGTAAGGAGTCGTCGGCAGCGTC	GTAAGGAG
sjmhb506	ACTGCATA	AATGATACGGCGACCACCGAGATCTA CACACTGCATATCGTCGGCAGCGTC	ACTGCATA
sjmhb507	AAGGAGTA	AATGATACGGCGACCACCGAGATCTA CACAAGGAGTATCGTCGGCAGCGTC	AAGGAGTA
sjmhb508	CTAAGCCT	AATGATACGGCGACCACCGAGATCTA CACCTAAGCCTTCGTCGGCAGCGTC	CTAAGCCT

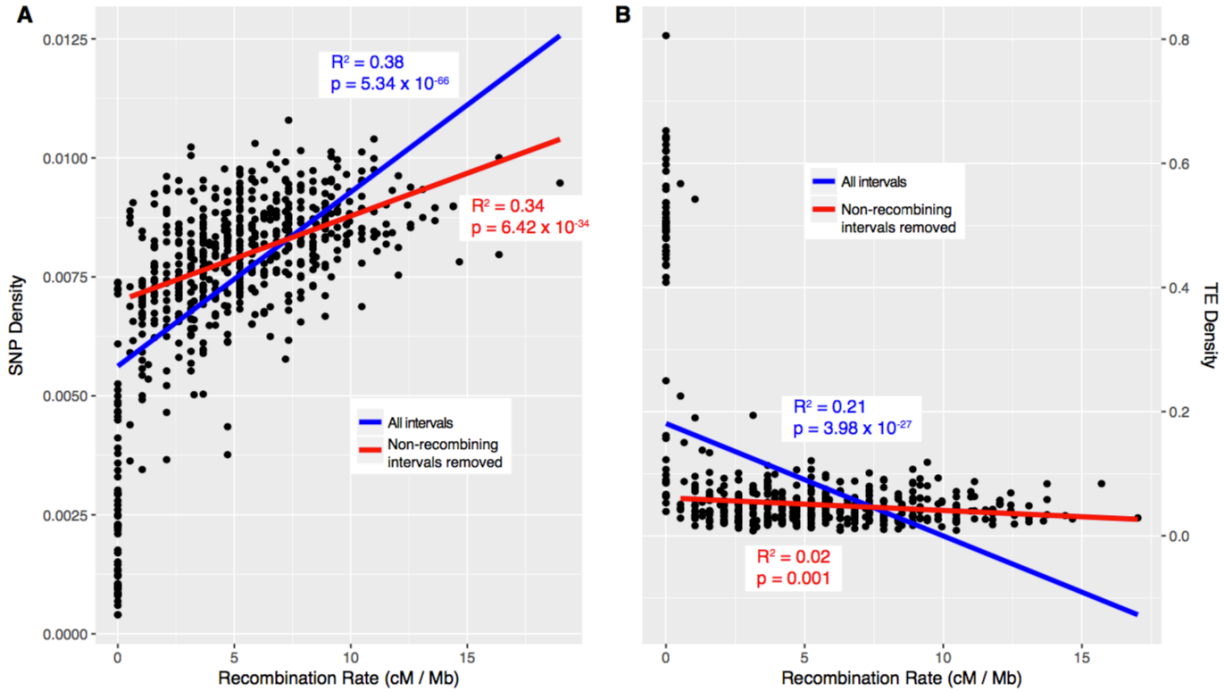
Appendix 18: The distribution of total crossover counts in *D. virilis* F2 progeny. The line is the expected number of crossovers under a Poisson distribution given the mean number of crossovers among all samples.



Appendix 19: Tetrad frequencies for each chromosome of the *D. virilis* F2 progeny. The tetrad frequencies were estimated using the Classic Weinstein method from Weinstein (1918). Negative tetrad frequencies are biologically meaningless and a drawback to using this method.

Chromosome	**N-Exchange (E_n) Tetrad Frequency**					
	E_0	E_1	E_2	E_3	E_4	E_5
X	0.012	0.176	0.0386	0.580	0.116	0.0773
2	-0.0399	0.184	-0.0338	0.754	-0.135	0.271
3	0.0217	-0.0314	0.556	0.184	0.193	0.0773
4	-0.0338	0.0870	0.213	0.618	0.000	0.116
5	-0.0266	0.167	0.222	0.367	0.271	0.000

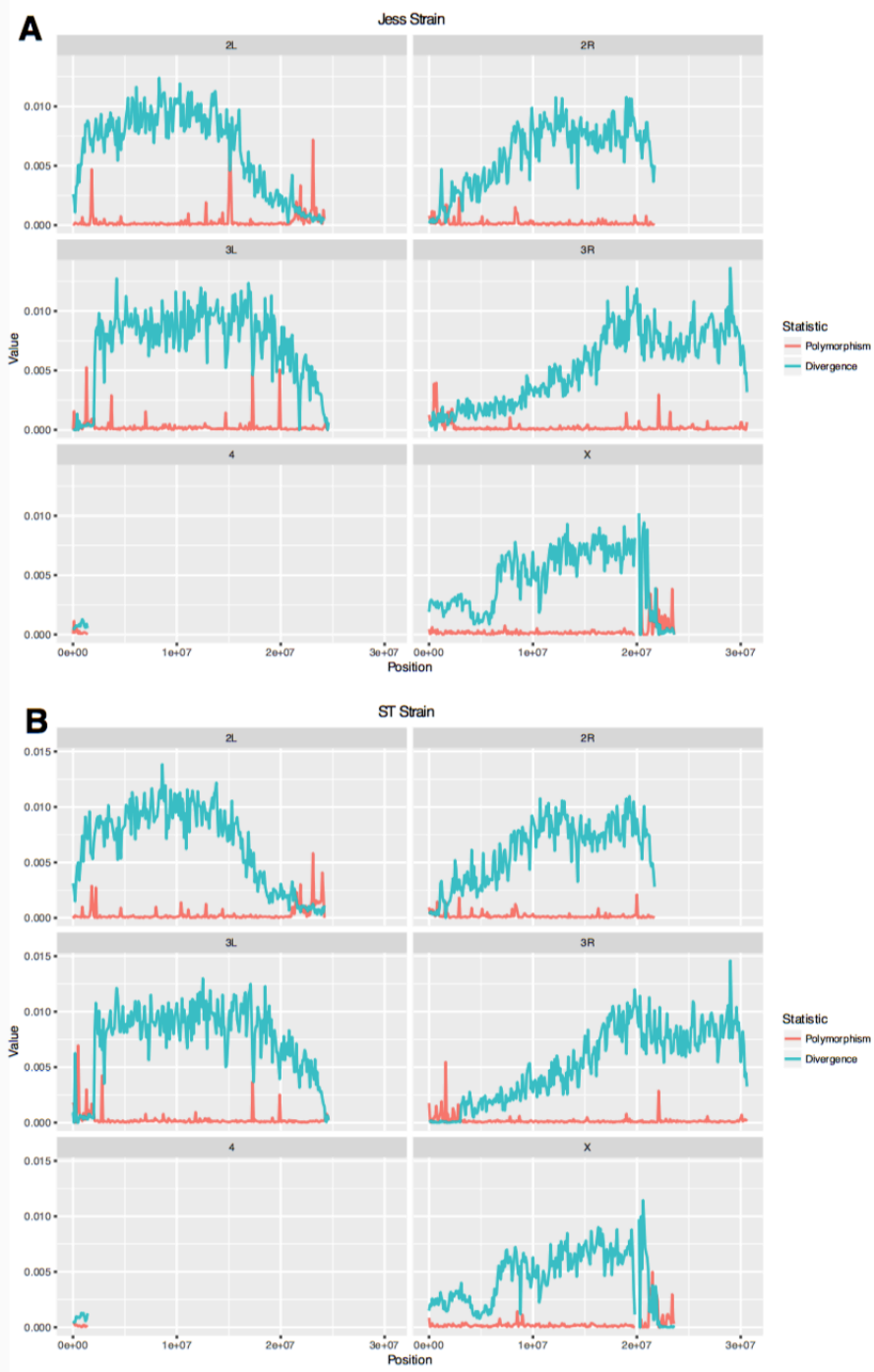
Appendix 20: Correlations between recombination rate and A) SNP Density and B) TE density with and without non-recombining regions. Density is the percentage of a 250 kb interval made up of either SNPs or TEs. The data is the summation of intervals across all chromosomes.



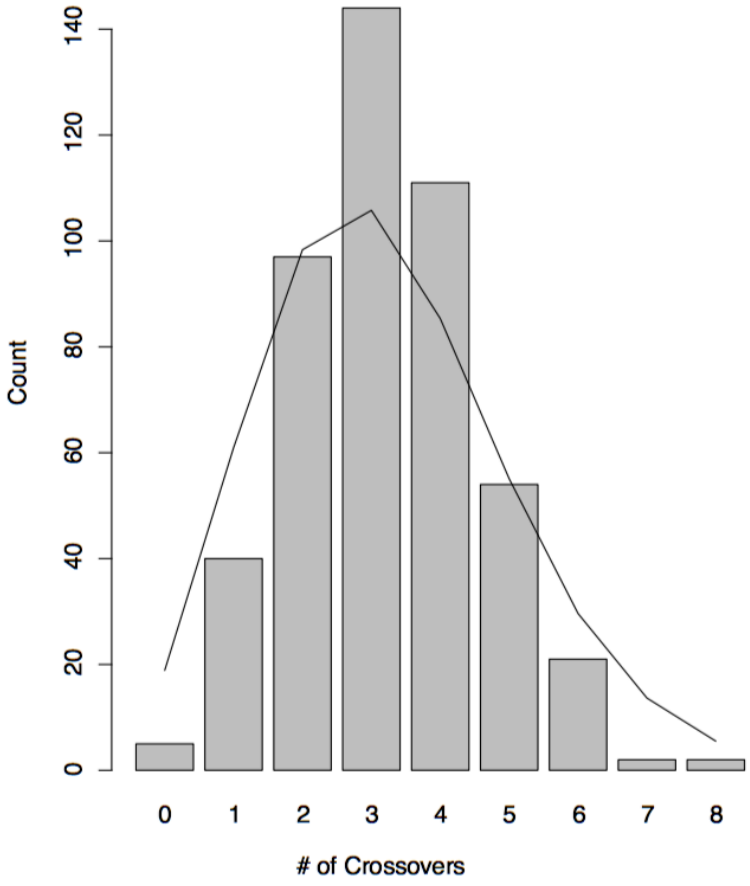
Appendix 21: Correlations between recombination rates of dysgenic and non-dysgenic flies and high fecund and low fecund dysgenic flies in 250 kb intervals. Pearson's correlation coefficients (R) and significance (*p*-value) are listed. All correlations are significant.

Comparison		**Chromosome**					**Total**	**Total minus zero recomb**
		X	2	3	4	5		
Dysgenic vs Non-Dys	R	0.49	0.68	0.58	0.67	0.70	0.63	0.43
	p	3.24E-09	1.11E-18	3.91E-13	2.64E-18	2.41E-20	7.16E-72	2.59E-24
High fecund vs low fecund	R	0.51	0.44	0.55	0.68	0.67	0.57	0.41
	p	6.97E-10	1.76E-07	1.83E-11	9.60E-19	2.27E-18	9.25E-57	2.70E-20

Appendix 22: Polymorphism within *D. yakuba* strains A) Jess and B) ST and divergence from the reference genome, Tai18E2. The low polymorphism indicates the strains were successfully inbred to reduce heterozygosity necessary for accurate genotype calling and recombination detection. An inversion would increase divergence within the boundaries of the inversion followed by a sharp decrease in divergence outside of those boundaries. Because the Andolfatto lab did not observe these patterns in either strain, they concluded there were no inversions in Jess or ST. Information and figure are courtesy of the Andolfatto lab.



Appendix 23: The distribution of total crossover counts in *D. yakuba* F2 progeny. The line is the expected number of crossovers under a Poisson distribution given the mean number of crossovers among all samples.



Appendix 24: Correlations between sequence parameters and recombination with all intervals included in *D. yakuba*. Pearson's correlation coefficients (R) and significance (*p*-value) are listed. Significant values are bolded. Measurements were recorded in 500 kb intervals along each chromosome in *D. yakuba*. Removing non-recombining regions reduces the correlation coefficients for all sequence parameters but SNP density still appears strongly correlated. Significant correlations are bolded.

Sequence Parameter		Chromosome					
		X	2L	2R	3L	3R	Total
GC Content	R	-0.012	**0.28**	**0.41**	**0.47**	**0.48**	**0.27**
	p	0.935	**3.02E-02**	**6.52E-03**	**4.31E-04**	**1.03E-04**	**8.51E-05**
Gene Density	R	**0.37**	**0.57**	**0.43**	**0.36**	0.15	**0.35**
	p	**1.05E-02**	**3.40E-06**	**4.44E-03**	**8.67E-03**	0.257	**4.59E-09**
Simple repeats	R	**0.65**	**0.54**	0.26	**0.64**	**0.48**	**0.55**
	p	**1.15E-06**	**8.11E-06**	0.171	**2.19E-07**	**9.21E-05**	**1.53E-22**
SNP Density	R	**0.66**	**0.75**	**0.70**	**0.61**	**0.78**	**0.63**
	p	**5.12E-07**	**8.92E-12**	**2.09E-07**	**1.03E-06**	**2.15E-13**	**2.93E-24**
TE Density	R	**-0.31**	**-0.54**	**-0.32**	**-0.46**	**-0.33**	**-0.44**
	p	**3.53E-02**	**1.06E-05**	**3.56E-02**	**5.58E-04**	**9.58E-03**	**2.60E-11**