THE EVOLUTIONARY CONSEQUENCES AND MECHANISM OF *SEX-RATIO* MEIOTIC DRIVE IN *DROSOPHILA NEOTESTACEA*

by

KATHLEEN ELIZABETH PIEPER

(Under the Direction of Kelly Dyer)

ABSTRACT

Genetic conflict occurs when one portion of the genome promotes its own transmission at the expense of another portion. Genetic conflict is ubiquitous, and can have significant evolutionary consequences. Meiotic drive is a kind of genetic conflict that occurs when one allele manipulates normal processes during meiosis or gametogenesis to be overrepresented in the gametes. X-chromosomes carrying meiotic drivers are known as *sex-ratio* X-chromosomes (SR) because they manipulate spermatogenesis to sabotage the production of Y-bearing sperm, resulting in the carrier male siring only daughters. The North American fruit fly *Drosophila neotestacea* harbors an SR system found at steady frequencies of up to 30% in some populations. In this system, there are no known suppressors or drive and 99% of offspring produced by SR males are daughters. In this dissertation, I investigate the molecular evolutionary consequences of drive in this system and identify candidate genes for the molecular mechanism. I show that SR carries large inversions and is significantly genetically differentiated from the standard X-chromosome (ST), but there is evidence that gene flow occurs between ST and SR. I find variation on SR is not present in distinct haplotypes,

but is likely maintained through recombination in SR homozygous females. I then identified hundreds of transcripts that are differentially expressed between ST and SR in the testes. Most of these transcripts are located on the X-chromosome, and differential expression begins early in spermatogenesis. I also found dozens of transcripts that show patterns of nucleotide sequence differences between ST and SR that are suggestive of positive selection. I identified candidates for the mechanism of drive as differentially expressed, testis-specific transcripts under putative positive selection. One of these candidates is a fast-evolving X-linked duplicate of the autosomal gene *importin-α2*, a critical part of the nuclear import pathway. At this locus and other candidates, I found molecular population genetic patterns consistent with involvement in the mechanism of drive: relaxed purifying selection in ST and positive selection in SR. Nuclear import is also targeted by a meiotic driver in *D. melanogaster*, suggesting that this process may be particularly susceptible to conflict.

INDEX WORDS: genetic conflict, meiotic drive, gene duplication, *Drosophila neotestacea*, population genetics, evolutionary genetics

THE EVOLUTIONARY CONSEQUENCES AND MECHANISM OF *SEX-RATIO* MEIOTIC DRIVE IN *DROSOPHILA NEOTESTACEA*

by

KATHLEEN ELIZABETH PIEPER

BS, University of Notre Dame, 2012

A Dissertation Submitted to the Graduate Faculty of The University of Georgia in Partial Fulfillment of the Requirements for the Degree

DOCTOR OF PHILOSOPHY

ATHENS, GEORGIA

2017

THE EVOLUTIONARY CONSEQUENCES AND MECHANISM OF *SEX-RATIO* MEIOTIC DRIVE IN *DROSOPHILA NEOTESTACEA*

by

KATHLEEN ELIZABETH PIEPER

Major Professor:	Kelly Dyer
Committee:	Kelly Dawe Dave Hall Jim Leebens-Mack Jessie Kissinger

Electronic Version Approved:

Suzanne Barbour
Dean of the Graduate School
The University of Georgia
May 2017

ACKNOWLEDGEMENTS

There are many people who have contributed to the success of this dissertation. First I would like to thank my advisor, Kelly Dyer, for her steadfast support and guidance over the past five years. I would like to thank the members of my committee for their feedback and positive input, which has greatly improved this dissertation. I would like to thank the past and present members of the Dyer lab, especially Brooke White, Emily Bewick, Cheryl Pinzone, Nick Arthur, Amanda Shaver, and Paul Ginsberg. Thank you for being such wonderful co-workers. I would also like to thank the larger community of grad students at UGA, especially Kerin Bentley and everyone involved in the Athens Science Café and Science Observer. Thanks for five great years of comradery, moral support, and intellectual engagement.

I also want to thank my parents, Steve and Carol Pieper, for always encouraging me to do my best and enabling me to achieve my goals. I hope you are proud of me. I want to thank my sisters, Kim Wolfe and Steph Pieper, for keeping me grounded and connected. And lastly, I want to thank David Higgins for being the best writing partner and life partner I could ever ask for.

TABLE OF CONTENTS

Page

LIST OF TABLES

Page

LIST OF FIGURES

Page

CHAPTER 1

INTRODUCTION

Genetic conflict occurs when one part of the genome promotes its own transmission at the expense of another part of the genome. Conflict is widespread; it is found in many different taxa and occurs in many different forms (Burt & Trivers, 2006). These selfish genetic elements range from gene-sized transposable elements to entire chromosomes, and can have significant evolutionary consequences (Hurst & Werren, 2001; Burt & Trivers, 2006; Rice, 2013). Self-replicating transposable elements are a clear example of how genetic conflict can shape genomes. Many eukaryote genomes are overwhelmingly composed of transposable elements, even though their accumulation is detrimental to the organism (Werren, 2011). Transposable elements can be coopted by the host genome to serve as new regulatory elements (Feschotte, 2008), and epigenetic mechanisms that silence transposable elements are also used to regulate normal host gene expression (Slotkin & Martienssen, 2007). Selection against selfish genetic elements may have led to the evolution of processes as fundamental as crossing over, sex determination, and speciation (Hurst & Werren, 2001; Werren, 2011; Rice, 2013).

Meiotic drive is a specific kind of genetic conflict that involves the subversion of meiosis or post-meiotic gametogenesis by the selfish genetic element to promote its own transmission to the next generation (Burt & Trivers, 2006). Normal meiosis results in the production of equal proportions of gametes carrying each allele. A selfish meiotic drive allele manipulates these processes to ensure it is found in a higher frequency of gametes

than the non-driving allele. Meiotic drive has been found in plants, animals, and fungi, and can occur during female, male, and haploid gametogenesis (Lyttle, 1991; Lindholm *et al.*, 2016). Meiotic drive can have significant evolutionary consequences for organisms, including altering gametogenesis, mating behavior, population sex ratios, and contributing to speciation (Lindholm *et al.*, 2016). Though the most basic evolutionary assumption about meiotic drive is that such an allele would rapidly go to fixation, many meiotic drives are observed to segregate at intermediate frequencies in the wild (Burt & Trivers, 2006; Lindholm *et al.*, 2016).

Artificial meiotic drive elements called "gene drives" have recently been developed with the intention of using these self-propagating elements to transform entire populations of organisms (Champer *et al.*, 2016). Potentially, gene drives could be used to make populations of insect vectors unable to transmit human diseases like malaria (Gantz *et al.*, 2015). The spread of a gene drive through a population is similar to the predicted patterns of a meiotic driver (Lindholm *et al.*, 2016). All identified natural meiotic drive systems are kept from driving to fixation by a variety of different factors, which suggests that using gene drives to rapidly fix desired traits may complicated. Determining how gene drives will act in wild populations is of critical importance before they can be safely used (Esvelt *et al.*, 2014; Oye *et al.*, 2014; DiCarlo *et al.*, 2015). Understanding the mechanisms, dynamics, and evolutionary consequences of meiotic drive thus has important and immediate consequences for this rapidly emerging new technology.

Sex chromosome linked meiotic drive is the most common form of meiotic drive known, with over 20 known systems in the order Diptera alone (Jaenike, 2001). There is

likely an element of discovery bias here because sex-linked meiotic drivers are particularly easy to identify as they result in highly skewed offspring sex ratios. For this reason, meiotic drive systems on the X-chromosome are typically referred to as *sex-ratio* (Jaenike, 2001). The sex chromosomes might also be more prone to conflict than the autosomes, however, as recombination is already suppressed between heteromorphic sex chromosomes and the ideal offspring sex ratio differs for the X and Y-chromosomes (Hurst & Pomiankowski, 1991; Burt & Trivers, 2006). Though selfish sex chromosomes have also been identified in plants and mammals, most research on sex chromosome linked meiotic drive has been done on *sex-ratio* systems in the genus *Drosophila* (Jaenike, 2001). Most of these systems have evolved independently and differ in the strength of drive, their population frequency, the presence of suppressors, and the presence and number of associated inversions (Jaenike, 2001).

There are many outstanding questions about the evolutionary consequences and molecular mechanisms of *sex-ratio* drive that remain to be answered. Because meiotic drive systems have repeatedly evolved independently in many different taxa, one major question is whether the same mechanism has convergent evolved to sabotage of the production of sperm carrying the non-driving allele. If true, this may indicate that certain processes or pathways are particularly susceptible to drive. The evolutionary consequences of meiotic drive, particularly of harboring high frequencies of drive, also remain to clarified as meiotic drivers reduce the effectiveness of natural selection and may carry deleterious alleles with them to high frequencies.

Sex-ratio chromosomes have a broad array of possible evolutionary patterns and influences on populations and behavioral evolution. Inversions are a common feature of

sex-ratio X-chromosomes, and presumably result from selection to keep together multiple loci that are required for drive (Jaenike, 2001; Meiklejohn & Tao, 2010). The effects of the suppression of recombination through inversions on molecular evolution are well documented, especially if there are loci within the inversion that under selection (Guerrero *et al.*, 2012). Meiotic drive will result in patterns similar to that of natural selection, which is clearly shown in two independent *sex-ratio* systems in *D. simulans* that are not associated with inversions (Derome *et al.*, 2008; Kingan *et al.*, 2010). However, there is much evidence that *sex-ratio* chromosomes have negative effects on fitness. The *sex-ratio* chromosome in *D. recens* is highly inverted and highly monomorphic (Dyer *et al.*, 2007). This system is only found at very low frequencies in the wild and females homozygous for *sex-ratio* are sterile, which is probably caused by deleterious alleles uninvolved in the mechanism of drive (Dyer *et al.*, 2007). The *sex-ratio* chromosome in *Teleopsis dalmanni* is also associated with inversions, extremely low polymorphism, and a short stalk-eye length in males that is unattractive to females (Christianson *et al.*, 2011; Cotton *et al.*, 2014). It is possible that female stalk-eyed flies have evolved a preference for long eye stalks to avoid mating with males carrying *sex-ratio* (Cotton *et al.*, 2014). Experimental evolution in *D. pseudoobscura* also suggests that females evolve behaviors to avoid mating with only *sex-ratio* males (Price *et al.*, 2008). The *sex-ratio* chromosome of *D. pseduoobscura* is found at relatively high frequencies in some populations and is associated with inversions, but has very low differentiation with the standard X-chromosome (Sturtevant & Dobzhansky, 1936; Kovacevic & Schaeffer, 2000).

Less work has been done on the molecular mechanisms that cause drive, but they also seem to be quite variable. The genetic basis of *sex-ratio* drive has been identified in only two systems, both of which are found in *D. simulans*. The Winters *sex-ratio* system consists of an X-linked driver and its autosomal suppressor (Tao *et al.*, 2007b). The autosomal suppressor (*Nmy*) appears to be the result of a retrotransposition from the X-chromosome to the autosomes of the distorter (*Dox*). *Dox* itself arose via duplication from another X-linked gene (*MDox*), though the molecular function of both genes is still unknown and they appear to have limited coding potential (Tao *et al.*, 2007a). The suppression of drive likely occurs through a RNA interference mechanism, and it is possible the distorter itself is also a noncoding RNA (Tao *et al.*, 2007a; Tao *et al.*, 2007b). In the independent Paris *sex-ratio* system, an X-linked gene called *HP1D2* causes drive by binding to the heterochromatin of Y-chromatids during meiosis II and preventing them from separating correctly, resulting in the successful production of only X-bearing sperm (Helleu *et al.*, 2016).

The mechanisms of some autosomal male meiotic drivers have also been identified. Though these drivers are located on the autosomes and not the sex chromosomes, they are like *sex-ratio* drive in that development of sperm carrying the non-drive allele is sabotaged. The Segregation Distorter (SD) system of *D. melanogaster* is located near the centromere of the second chromosome and just two loci are required for drive (Larracuente & Presgraves, 2012). The responder locus, *Rsp,* is not a protein coding gene, but a satellite DNA repeat whose size confers sensitivity or resistance to driving (Wu *et al.*, 1988). The distorter locus, *sd,* is a truncated duplicate of the nuclear transport gene RanGAP that mislocalizes to the nucleus from the cytoplasm (Merrill *et*

al., 1999; Kusano *et al.*, 2001) It is unknown how it interacts with *Rsp* and how it causes the death of sperm carrying the sensitive responder allele, but overexpression of wild-type RanGAP also causes drive, indicating that the disruption of nuclear transport is a critical aspect (Kusano *et al.*, 2002). The t-haplotype of *Mus musculus* is another autosomal driver that has four fixed inversions that prevent recombination with the wild-type chromosome and bind together at least four loci (Lyon, 2003; Bauer *et al.*, 2007). The responder locus, *Tcr,* is a hypomorphic duplicate of *smok,* a protein kinase required for sperm motility (Herrmann *et al.*, 1999). *Tcr* is transmitted at a lower rate than the wild-type allele when present without the distorters (Herrmann *et al.*, 1999). The distorters act additively to increase the transmission of *Tcr* to the offspring to up to 99%. Two of the distorters have both been identified as independent upstream regulators of *smok* expression, suggesting that the overexpression of the wild-type protein causes sperm motility defects that are rescued in sperm carrying t-haplotype by the lower expression of *Tcr* (Bauer *et al.*, 2005; Bauer *et al.*, 2007).

In this dissertation, I investigate the evolutionary consequences and molecular mechanisms of *sex-ratio* meiotic drive in the species *Drosophila neotestacea,* a boreal and temperate mushroom feeding North American fruit fly (James & Jaenike, 1990). This is a good system for investigating the consequences of selfish activity as has likely had a significant effect on the sequence evolution of the X-chromosome in *D. neotestacea*. Previous work shows that strong linkage disequilibrium between microsatellites on *sex-ratio* indicates the presence of one or more inversions that likely encompass the entire chromosome, though an excess of microsatellite polymorphism also suggests occasional recombination or gene conversion between the *sex-ratio* and standard chromosomes in

heterozygote females (Dyer, 2012; Dyer *et al.*, 2013). Even a small amount of gene flow with the wild-type chromosome could prevent the degradation of *sex-ratio* from mutation accumulation, which is seen in non-recombining selfish chromosomes containing multiple large inversions in *D. recens* (Dyer *et al.*, 2007). There is no evidence of active *sex-ratio* suppressors in *D. neotestacea* despite the detrimental effects it has on the Y chromosome, and the frequency of *sex-ratio* is as high as 40% in some populations and appears to be stable (Dyer, 2012; Pinzone & Dyer, 2013). Populations with a high frequency of *sex-ratio* also appear to have a female-biased sex ratio (Pinzone & Dyer, 2013). High, stable frequencies of *sex-ratio*, a lack of phenotypic suppressors, and the potential long-term maintenance through recombination strongly suggests that *sex-ratio* has had a large impact on genome evolution in this species.

In the second chapter of my dissertation, I present a review of the important literature on meiotic drive in the format of an Oxford Bibliographies article. I present work on the history, mechanisms, ecology, and evolution of female drive, male autosomal drive, and male sex-linked drive, as well as spore killer drive and synthetic meiotic drive systems. This thorough work serves as a guide to the meiotic drive literature, and outlines the current state of the field and identifies important findings in major systems.

The third chapter of my dissertation is an empirical study of the evolutionary consequences of drive using the *sex-ratio* system of *D. neotestacea.* I use a population genetic approach to compare the molecular evolutionary patterns on the *sex-ratio* (SR) and standard (ST) X-chromosomes. I use roughly 3 KB of nucleotide sequence data from nearly 50 ST and SR males from across the species range to show that at arbitrarily

chosen loci uninvolved in the mechanism of drive, there is significant genetic differentiation between ST and SR. I confirm the presence of multiple, large inversions on SR, but also find that differentiation is variable across loci and no surveyed locus had a fixed sequence difference between ST and SR. I find additional evidence that some gene flow is occurring between ST and SR, and that that there is substantial variation on SR. This variation is not segregating in haplotypes, but appears to be maintained by recombination between SR chromosomes in homozygous females. This suggests that SR may be protected from degradation via mutation accumulation, and that the causal driving loci can be identified by molecular evolutionary patterns consistent with strong selection in SR but not ST.

In the fourth chapter of my dissertation, I identify candidates for the mechanism of drive using transcriptomics and population genetics. I find that there are over 700 significantly differentially expressed transcripts between ST and SR in the testes, and that the majority of them map to the X-chromosome. I also find that differential expression begins in the earliest stages of spermatogenesis. I identify nucleotide sequence differences between ST and SR, and find that there are around 40 transcripts with K_a/K_s values indicative of positive selection. Of these candidates, I find a set of 6 transcripts that are also differentially expressed and testes specific. Two of these transcripts map to a fast-evolving testacea-group specific X-linked duplication of the gene *importin-α2,* which is a critical piece of the nuclear import pathway during spermatogenesis. I sequence this duplicate and other candidate transcripts in a population genetic sample of ST and SR males, and find that the molecular evolutionary patterns are consistent with relaxed purifying selection in ST and positive selection in SR. I propose potential

mechanisms for how this *importin-α2* duplicate may contribute to the mechanism of meiotic drive in *D. neotestacea.* The disruption of nuclear transport also causes drive in the SD system of *D. melanogaster,* suggesting this pathway may be particularly susceptible genetic conflict.

References

Bauer, H., Willert, J.R., Koschorz, B. & Herrmann, B.G. 2005. The t complex-encoded GTPase-activating protein Tagap1 acts as a transmission ratio distorter in mice. *Nature Genetics* **37:** 969-973.

Bauer, H., Véron, N., Willert, J. & Herrmann, B.G. 2007. The t-complex-encoded guanine nucleotide exchange factor Fgd2 reveals that two opposing signaling pathways promote transmission ratio distortion in the mouse. *Genes & Development* **21:** 143-147.

Burt, A. & Trivers, R. 2006. *Genes in Conflict: The Biology of Selfish Genetic Elements*. Belknap Press of Harvard University Press, Cambridge, Mass.

Champer, J., Buchman, A. & Akbari, O.S. 2016. Cheating evolution: engineering gene drives to manipulate the fate of wild populations. *Nature Reviews Genetics* **17:** 146-159.

Christianson, S.J., Brand, C.L. & Wilkinson, G.S. 2011. Reduced polymorphism associated with X chromosome meiotic drive in the stalk-eyed fly *Teleopsis dalmanni*. *PLoS ONE* **6:** 1-7.

Cotton, A.J., Foldvari, M., Cotton, S. & Pomiankowski, A. 2014. Male eyespan size is associated with meiotic drive in wild stalk-eyed flies (*Teleopsis dalmanni*). *Heredity* **112:** 363-369.

Derome, N., Baudry, E., Ogereau, D., Veuille, M. & Montchamp-Moreau, C. 2008. Selective sweeps in a 2-locus model for sex-ratio meiotic drive in *Drosophila simulans*. *Mol. Biol. Evol.* **25:** 409-416.

DiCarlo, J.E., Chavez, A., Dietz, S.L., Esvelt, K.M. & Church, G.M. 2015. Safeguarding CRISPR-Cas9 gene drives in yeast. *Nature Biotechnology* **33:** 1250-+.

Dyer, K.A., Charlesworth, B. & Jaenike, J. 2007. Chromosome-wide linkage disequilibrium as a consequence of meiotic drive. *Proc. Natl. Acad. Sci.* **104:** 1587-1592.

Dyer, K.A. 2012. Local selection underlies the geographic distribution of sex-ratio drive in *Drosophila neotestacea*. *Evolution* **66:** 973-984.

Dyer, K.A., Bray, M.J. & Lopez, S.J. 2013. Genomic conflict drives patterns of X-linked population structure in *Drosophila neotestacea*. *Mol. Ecol.* **22:** 157-169.

Esvelt, K.M., Smidler, A.L., Catteruccia, F. & Church, G.M. 2014. Concerning RNA-guided gene drives for the alteration of wild populations. *Elife* **3**.

Feschotte, C. 2008. Transposable elements and the evolution of regulatory networks. *Nature Reviews Genetics* **9:** 397-405.

Gantz, V.M., Jasinskiene, N., Tatarenkova, O., Fazekas, A., Macias, V.M., Bier, E. & James, A.A. 2015. Highly efficient Cas9-mediated gene drive for population modification of the malaria vector mosquito *Anopheles stephensi*. *Proc. Natl. Acad. Sci.* **112:** E6736-E6743.

Guerrero, R.F., Rousset, F. & Kirkpatrick, M. 2012. Coalescent patterns for chromosomal inversions in divergent populations. *Phil. Trans. R. Soc. B* **367:** 430-438.

Helleu, Q., Gérard, P.R., Dubruille, R., Ogereau, D., Prud'homme, B., Loppin, B. & Montchamp-Moreau, C. 2016. Rapid evolution of a Y-chromosome heterochromatin protein underlies sex chromosome meiotic drive. *Proc. Natl. Acad. Sci.* **113:** 4110-4115.

Herrmann, B.G., Koschorz, B., Wertz, K., McLaughlin, K.J. & Kispert, A. 1999. A protein kinase encoded by the t complex responder gene causes non-mendelian inheritance. *Nature* **402:** 141-146.

Hurst, G.D.D. & Werren, J.H. 2001. The role of selfish genetic elements in eukaryotic evolution. *Nature Reviews Genetics* **2:** 597-606.

Hurst, L.D. & Pomiankowski, A. 1991. Causes of sex-ratio bias may account for unisexual sterility in hybrids - a new explanation of Haldane's rule and related phenomena. *Genetics* **128:** 841-858.

Jaenike, J. 2001. Sex chromosome meiotic drive. *Annu. Rev. Ecol. Syst.* **32:** 25-49.

James, A.C. & Jaenike, J. 1990. Sex-ratio meiotic drive in *Drosophila testacea*. *Genetics* **126:** 651-656.

Kingan, S.B., Garrigan, D. & Hartl, D.L. 2010. Recurrent selection on the Winters sex-ratio genes in *Drosophila simulans*. *Genetics* **184:** 253-265.

Kovacevic, M. & Schaeffer, S.W. 2000. Molecular population genetics of X-linked genes in *Drosophila pseudoobscura*. *Genetics* **156:** 155-172.

Kusano, A., Staber, C. & Ganetzky, B. 2001. Nuclear mislocalization of enzymatically active RanGAP causes segregation distortion in Drosophila. *Developmental cell* **1:** 351-361.

Kusano, A., Staber, C. & Ganetzky, B. 2002. Segregation distortion induced by wild-type RanGAP in Drosophila. *Proc. Natl. Acad. Sci.* **99:** 6866-6870.

Larracuente, A.M. & Presgraves, D.C. 2012. The selfish Segregation Distorter gene complex of *Drosophila melanogaster*. *Genetics* **192:** 33-53.

Lindholm, A.K., Dyer, K.A., Firman, R.C., Fishman, L., Forstmeier, W., Holman, L., Johannesson, H., Knief, U., Kokko, H., Larracuente, A.M., Manser, A., Montchamp-Moreau, C., Petrosyan, V.G., Pomiankowski, A., Presgraves, D.C., Safronova, L.D., Sutter, A., Unckless, R.L., Verspoor, R.L., Wedell, N., Wilkinson, G.S. & Price, T.A.R. 2016. The Ecology and Evolutionary Dynamics of Meiotic Drive. *Trends in Ecology & Evolution* **31:** 315-326.

Lyon, M.F. 2003. Transmission ratio distortion in mice. *Annual Review of Genetics* **37:** 393-408.

Lyttle, T.W. 1991. Segregation distorters. *Annual Review of Genetics* **25:** 511-557.

Meiklejohn, C.D. & Tao, Y. 2010. Genetic conflict and sex chromosome evolution. *Trends. Ecol. Evol.* **25:** 215-223.

Merrill, C., Bayraktaroglu, L., Kusano, A. & Ganetzky, B. 1999. Truncated RanGAP encoded by the Segregation Distorter locus of Drosophila. *Science* **283:** 1742-1745.

Oye, K.A., Esvelt, K., Appleton, E., Catteruccia, F., Church, G., Kuiken, T., Lightfoot, S.B.-Y., McNamara, J., Smidler, A. & Collins, J.P. 2014. Biotechnology: Regulating gene drives. *Science* **345:** 626-628.

Pinzone, C.A. & Dyer, K.A. 2013. Association of polyandry and sex-ratio drive prevalence in natural populations of *Drosophila neotestacea*. *Proc. R. Soc. B Biol. Sci.* **280:** 20131397.

Price, T., Hodgson, D., Lewis, Z., Hurst, G. & Wedell, N. 2008. Selfish genetic elements promote polyandry in a fly. *Science* **322:** 1241-1243.

Rice, W.R. 2013. Nothing in genetics makes sense except in light of genomic conflict. *Annu. Rev. Ecol. Syst.* **44:** 217-237.

Slotkin, R.K. & Martienssen, R. 2007. Transposable elements and the epigenetic regulation of the genome. *Nat Rev Genet* **8:** 272-285.

Sturtevant, A. & Dobzhansky, T. 1936. Geographical distribution and cytology of" sex ratio" in *Drosophila pseudoobscura* and related species. *Genetics* **21:** 473.

Tao, Y., Araripe, L., Kingan, S.B., Ke, Y., Xiao, H. & Hartl, D.L. 2007a. A sex-ratio meiotic drive system in *Drosophila simulans*. II: an X-linked distorter. *PLoS Biol* **5:** e293.

Tao, Y., Masly, J.P., Araripe, L., Ke, Y. & Hartl, D.L. 2007b. A sex-ratio meiotic drive system in *Drosophila simulans*. I: an autosomal suppressor. *PLoS Biol* **5:** e292.

Werren, J.H. 2011. Selfish genetic elements, genetic conflict, and evolutionary innovation. *Proc. Natl. Acad. Sci.* **108:** 10863-10870.

Wu, C.-I., Lyttle, T.W., Wu, M.-L. & Lin, G.-F. 1988. Association between a satellite DNA sequence and the Responder of Segregation Distorter in *D. melanogaster*. *Cell* **54:** 179-189.

CHAPTER 2

OXFORD BIBLIOGRAPHY ON MEIOTIC DRIVE[1]

[1]Pieper, K.E. and Dyer, K.A. Submitted to *Oxford Bibliographies* 18 March 2017

Introduction

Genetic conflict occurs when one part of the genome increases its own fitness at the expense of another part of the genome. Mendel's law of equal segregation states that each member of a pair of alleles will be equally represented in an organism's gametes. Meiotic drive occurs when processes during meiosis or later stages of gametogenesis skew that ratio, such that one allele (the "driving" allele) is overrepresented in the gametes. It is sometimes also called transmission ratio distortion or segregation distortion. Meiotic drive systems are known in plants, animals, and fungi, and can occur in females, males, and during haploid gametogenesis. Female meiosis may be particularly susceptible to meiotic drive since only one of the four products of meiosis develops into the mature gamete, while the other products of meiosis become the polar bodies of the egg. Male meiosis is also susceptible to disruption, usually through sabotaging the development or function of sperm carrying the non-driving allele. While female drive is considered 'true' meiotic drive, the term meiotic drive is also commonly used to describe male-drive systems where interruptions occur in post-meiotic stages of gametogenesis. Far from being a genetic curiosity, meiotic drive and the genetic conflict it can cause may have had a wide-reaching and significant impacts on evolutionary phenomena such as the evolution of sex and recombination, the process of speciation, and genomic structure. Meiotic drive is a rich field of study that includes investigations into the molecular mechanisms that cause drive, the evolutionary and ecological dynamics that maintain drive systems in the wild, and the molecular evolutionary consequences of drive. Synthetic meiotic drive systems are a rapidly growing area of innovation, with potential uses such as controlling populations of insect vectors of disease.

General Overviews

There are some excellent overviews of meiotic drive. Sandler and Novitski (1957) is notable for being the first overview of meiotic drive and the first discussion of its potential evolutionary consequences. Burt & Trivers (2006) is the most thorough overview. This book goes over a wide variety of selfish genetic elements, including many different types of drive. They cover male and female autosomal meiotic drive, sex chromosome drive, spore killer drive, and centromere drive, and for each they give an overview of what is known about both the molecular mechanisms and its ecological and evolutionary impact. Lyttle (1991) is similarly a very thorough overview of different kinds of meiotic drive systems, but this review takes a comparative approach between systems that is quite informative. Lindholm et al. (2016) an excellent overview of the evolutionary and ecological impacts of a wide variety of meiotic drive systems, including very recent developments in synthetic driving systems.

Sandler L., & Novitski, E. (1957). Meiotic drive as an evolutionary force. *American Naturalist,* 91(857), 105-110.

This paper was one of the first to identify meiotic drive as an important process with consequences for evolution. They review the previous work on meiotic drive in various systems, including examples of drive in males and females and involving autosomes and sex chromosomes. They discuss the potential evolutionary implications of drive, including the fixation of linked deleterious alleles.

Burt, A., & Trivers, R. (2006). *Genes in Conflict: The Biology of Selfish Genetic Elements*. Cambridge, Mass.: Belknap Press of Harvard University Press.

This book contains several chapters dedicated to an overview of autosomal killers, sex chromosome drive, and female meiotic drive. These chapters provide a thorough summary of research on both the molecular mechanisms and evolutionary consequences of drive. While perhaps too difficult for a general audience, this book is generally more accessible than the primary literature and is an excellent first introduction to the field.

Lyttle, T. W. (1991). Segregation distorters. *Annual Review of Genetics,* 25(1), 511-581.

This review takes a very informative comparative approach to describe the molecular and evolutionary research into different kinds of meiotic drive, including major study systems like SD, t-haplotype, and *sex-ratio.* More obscure systems are also discussed.

Lindholm, A. K., Dyer, K. A., Firman, R. C., Fishman, L., Forstmeier, W., Holman, L., Johannesson, H., Knief, U., Kokko, H., Larracuente, A.M., Manser, A., Montchamp-Moreau, C., Petrosyan, V.G., Pomiankowski, A., Presgraves, D.C., Safronova, L.D., Sutter, A., Unckless, R.L., Verspoor, R.L., Wedell, N., Wilkinson, G.S. & Price, T.A.R. (2016). The ecology and evolutionary dynamics of meiotic drive. *Trends in Ecology & Evolution, 31*(4), 315-326.

This recent review highlights the ecological and evolutionary effects of a wide swath of meiotic drive elements, including recently developed synthetic meiotic drive systems. They discuss effects on gametogenesis, genome structure, coevolution with suppressors, mating system evolution, and more.

Journals

There are no journals focused primarily on meiotic drive, but papers about drive can be found in a wide variety of journals depending on the angle of investigation. Studies of the molecular genetic mechanisms of driving element can be found in journals like *Genetics* and *PLOS Genetics*. Studies of the molecular evolution of driving elements, the population dynamics of drive, or theoretical studies on driving behavior can be found in journals such as *Molecular Ecology*, *Evolution,* and *American Naturalist.* Papers about meiotic drive can be found in many other journals as well.

Genetics.

This is the journal of the Genetics Society of America, and was first published in 1919. This journal publishes high quality research in many different areas of genetics. Work on the molecular mechanisms and evolutionary genetics of meiotic drive can be found here.

PLOS Genetics.

The Public Library of Science first published this open source journal in 2006. Work dissecting the molecular mechanisms and population genetics of driving systems can be found here.

Molecular Ecology.

This journal was first published in 1992 and focuses on the use of molecular techniques to answer questions about ecology and evolution. Studies on how meiotic drive affects natural variation and ecological processes can be found here.

Evolution.

This is the society journal of the Society for the Study of Evolution and was first published in 1946. Theoretical and empirical work on the evolutionary consequences of meiotic drive can be found published in this journal.

The American Naturalist.

This journal was first published in 1867 and focuses on evolution, ecology, and behavior. Papers describing meiotic drive systems and their evolution and ecology can be found here.

Female Meiotic Drive

Female meiotic drive is the natural result of asymmetry in female meiosis. Only one of the four products of each meiosis will become a mature gamete capable of forming a zygote, and so any gene that can bias its transmission into the egg cell (oocyte) will have an advantage. In female meiotic drive, the physical orientation of the chromatids is manipulated during meiosis such that the chromatid carrying the driver will end up as the oocyte after the second meiotic division. Female meiotic drive typically has no associated reduction in female fertility because every egg still develops, a contrast to what is usually found in male drive systems. Only autosomal female meiotic drivers have been identified.

Overviews of Major Systems

Female meiotic drive is sometimes referred to as "true meiotic drive" since the mechanism of drive occurs during meiosis itself rather than during later stages of gametogenesis as in male meiotic drive. This action is reliant on the inherent asymmetry of female meiosis, where only a single product of meiosis goes on to become a mature gamete (de Villena & Sapienza 2001). The Ab10 system in maize (*Zea mays)* was originally identified in the 1940s, and is the most thoroughly described female meiotic

drive system (Rhoades 1952). Ab10 is an abnormal variant of chromosome 10 that carries a large heterochromatin knob composed of repeat sequences at the end of the chromosome, along with a unique region of euchromatin (Dawe 2009). The knob acts as neocentromere, dragging it toward the spindle pole at a rate much faster than the regular centromere and orienting itself to end up in the mature gamete (Rhoades 1952, Dawe 2009). As a result, Ab10 is found in 70% of the offspring of a heterozygote female (Rhoades 1952, Dawe 2009). More recently, a female meiotic driver has been identified in the monkeyflower *Mimulus guttatus* (Fishman & Willis 2005). This driver, called the *D* locus, is located very close to the centromere of Linkage Group 11. In hybrids produced by crosses between *M. guttatus* and *M. nasutus, D* drives at a rate of 99% (Fishman & Willis 2005). Though other instances of segregation distortion occurring in females have been identified, these two systems are the best studied in terms of mechanism and evolutionary impact (de Villena & Sapienza 2001).

de Villena, F. P. M., & Sapienza, C. (2001). Nonrandom segregation during meiosis: the unfairness of females. *Mammalian Genome, 12*(5), 331-339.

This review takes a very thorough look at what constitutes true meiotic drive in females and focuses on the meiotic properties that lead to driving behavior. It also describes other instances of female meiotic drive than Ab10 and *D* locus.

Rhoades, M. M. (1952). Preferential segregation in maize. In J. W. Gowen (Ed.), *Heterosis* (pp. 66-80 pp.): Iowa State College Press.

This book chapter describes the discovery and cytological phenotype of the Ab10 system in maize and proposes a hypothesis for the mechanism of drive where the neocentromeric activity of the knob orients chromatids carrying Ab10 to preferentially segregate to the region that will become the mature gamete.

Dawe, R. K. (2009). Maize centromeres and knobs (neocentromeres). In J.L. Bennetzen and S. Hake (eds.), *Maize Handbook - Volume II: Genetics and Genomics* (pp 239-250): Springer Science + Business Media.

This book chapter is a comprehensive review of the Ab10 system in maize, including its history, genetic structure, and cellular mechanisms of how it attains preferential segregation. This review also discusses mutant screens to identify the genes that control meiotic drive, how Ab10 promotes the evolution of other knobs in the Maize genome, and how Ab10 may be maintained as a polymorphism in natural populations.

Fishman, L., & Willis, J. H. (2005). A novel meiotic drive locus almost completely distorts segregation in *Mimulus* (monkeyflower) hybrids. *Genetics, 169*(1), 347-353.

This paper describes the discovery of *D* locus in *Mimulus guttatus* and shows that the transmission ratio distortion is due to biased segregation during female meiosis. The authors suggest that *D* is an example of centromeric conflict. Notably, *D* drives at a very high rate in the genetic background the self-fertilizing species *D. nasutus,* which may indicate that selection for selfish centromeres is relaxed in this species.

Mechanisms

Dissecting the molecular mechanisms of meiotic drive is not an easy task. To add to the difficulty in the Ab10 system, the portion of the chromosome that carries the factors responsible for drive does not recombine with the wild type chromosome over a length of roughly 32 cM. Structurally, Ab10 is made up of an inverted region containing genes shared with the wild type chromosome, a large heterochromatic knob, and the distal tip that harbors the genes required for drive (Mroczek et al. 2006). The genes located on the distal tip have no known homologs in the rest of the maize genome (Mroczek et al. 2006). Several distinct steps are required for drive to occur in Ab10, and each has unique genetic controls (Hiatt & Dawe 2003). First, there is a factor on Ab10 that promotes recombination, which increases the efficiency of drive by ensuring the chromatid carrying Ab10 is positioned correctly (Hiatt & Dawe 2003). Next, the neocentromeric activity of the knob is activated by trans-acting factors, and the knob is

then pulled along microtubules towards the spindle pole (Hiatt & Dawe 2003). However, mutant screens show that neocentromere activity can be retained without meiotic drive occurring, indicating that there are multiple trans-acting factors required for drive to occur (Hiatt & Dawe 2003). Ab10 is quite a complex system, and its exact molecular mechanisms and genetic basis remain unknown.

Mroczek, R. J., Melo, J. R., Luce, A. C., Hiatt, E. N., & Dawe, R. K. (2006). The maize Ab10 meiotic drive system maps to supernumerary sequences in a large complex haplotype. *Genetics*, *174*(1), 145-154.

This paper shows that multiple independent arrangements and a large chunk of DNA with no homologs in the maize genome separate Ab10 from the wild type chromosome. This paper contains several excellent figures depicting the structure of Ab10 as well as some of its structural variants.

Hiatt, E. N., & Dawe, R. K. (2003). Four loci on abnormal chromosome 10 contribute to meiotic drive in maize. *Genetics, 164*(2), 699-709.

This paper describes a study that uses a combination of mutant screens and lines carrying large deletions to show that the driving activity of Ab10 involves at least four different loci on the chromosome, and that the genes located on the distal tip are necessary for drive to function.

Ecology and Evolution

Female meiotic drive systems have significant evolutionary consequences for their hosts. The *D* locus in *Mimulus guttatus* has a high male fitness cost – individuals homozygous for *D* have significant pollen viability defects (Fishman & Saunders 2008). *D* also has female fitness costs, including lower seed production (Fishman & Kelly 2015). This is likely due to hitchhiking by deleterious mutations unrelated to the mechanism of drive, and together these fitness costs are sufficient to maintain *D* at the observed intermediate frequencies instead of it driving to fixation (Fishman & Kelly 2015). Three different structural variants of Ab10 are known to segregate at intermediate frequencies within domestic maize strains and populations of its wild progenitor, teosinte (Kanizay et al. 2013). It is likely that the male and female fitness costs of carrying Ab10 have prevented it from going to fixation (Kanizay et al. 2013). Nevertheless, the Ab10 system has likely shaped the maize genome in a very profound way, as heterochromatic knobs are commonly found on all chromosomes (Buckler et al. 1999). These knobs gain neocentromeric function when present with Ab10, and their sizes and locations indicate the meiotic drive has selected for them (Buckler et al. 1999). Female meiotic drive could also have consequences for processes as fundamental as the female recombination rate, since recombination can act as a suppressor of drive (Brandvain & Coop 2012).

Fishman, L., & Saunders, A. (2008). Centromere-associated female meiotic drive entails male fitness costs in Monkeyflowers. *Science, 322*(5907), 1559-1562.

This paper shows that *D* is closely linked to the centromere of linkage group 11, and that there is extremely low polymorphism around the area, consistent with a selective sweep. *D* drives only slightly above 50% within *Mimulus guttatus,* but individuals homozygous for *D* have decreased pollen viability.

Fishman, L., & Kelly, J. K. (2015). Centromere-associated meiotic drive and female fitness variation in *Mimulus*. *Evolution, 69*(5), 1208-1218.

This paper shows that there is also a female fitness cost within *Mimulus guttatus* for carrying *D*. Together with the male fitness cost, these parameters can explain the frequency of *D* found in wild populations using mathematical modeling.

Kanizay, L. B., Pyhajarvi, T., Lowry, E. G., Hufford, M. B., Peterson, D. G., Ross-Ibarra, J., & Dawe, R. K. (2013). Diversity and abundance of the abnormal chromosome 10 meiotic drive complex in *Zea mays*. *Heredity, 110*(6), 570-577.

This paper describes the geographic distribution of Ab10 within maize and teosinte, the wild progenitor of domesticated corn. Ab10 segregates at variable frequencies and up to 75% in some populations of teosinte. It is likely that both intrinsic and extrinsic fitness

costs associated with carrying Ab10 have kept it from driving to fixation. This paper also describes three different Ab10 chromosomes and their evolutionary relationship.

Buckler, E. S., Phelps-Durr, T. L., Buckler, C. S. K., Dawe, R. K., Doebley, J. F., & Holtsford, T. P. (1999). Meiotic drive of chromosomal knobs reshaped the maize genome. *Genetics, 153*(1), 415-426.

This paper uses simulations and phylogenetic analyses to show that knobs of heterochromatin have accumulated in the maize genome in a way that strongly suggests they are maintained by meiotic drive. These knobs are located on every chromosome, but only acquire neocentromeric activity in the presence of Ab10. Without meiotic drive from Ab10, they would likely be selected against.

Brandvain, Y., & Coop, G. (2012). Scrambling eggs: meiotic drive and the evolution of female recombination rates. *Genetics, 190*(2), 709-723.

This paper presents a mathematical model showing how female meiotic drive could impact the evolution of variable female recombination rates. Since recombination can act as a suppressor of meiotic drive, a locus that decreases the recombination rate and is also is linked to a driver could quickly increase in frequency.

Centromere Drive

Centromeres mediate the movement of chromosomes to the poles during cell division. Their cellular function is highly conserved, but their genetic composition varies substantially among species, as does the rate of evolution of the interacting proteins (Rosin & Mellone 2017). To account for this 'centromere paradox', Henikoff et al. (2001) proposed a model where centromeres and centromeric-binding histone proteins evolve in an arms race due to meiotic drive during asymmetric female meiosis. If a centromere can preferentially end up in the oocyte, for instance by expansion of a satellite repeat that attracts more microtubules than the homologous centromere, this stronger centromere would have an evolutionary advantage and is expected to fix in the population. In response, the centromere-binding proteins may evolve to increase the strength of binding to the weaker centromere and thus restore equal segregation. Functional evidence that the strength of a centromere influences the likelihood it will become the egg oocyte was recently demonstrated by Chmatal et al. (2014). A key prediction of the centromere drive model is that it will only occur in taxa with asymmetric meiosis and not in taxa where all four products of meiosis survive. A recent comparative molecular evolutionary analysis of centromeric histone H3 supports this prediction (Zedek & Burez 2016). In addition to the Ab10 and the *Mimulus* systems discussed in the female drive sections, evidence for centromere drive comes from Robertsonian (Rb) fusions, which are chromosomal rearrangements in which two telocentric chromosomes are fused to form a single metacentric chromosome. Across mammals the karyotype of a species is usually biased to include more telocentric or metacentric chromosomes, and the simplest explanation for this pattern is that the

direction of meiotic drive varies across species, whereby one chromosome type or the other is preferentially maintained in the egg (de Villena & Sapienza 2001). Ultimately, rapid centromere evolution and karyotype changes may cause the production of infertile offspring and thus contribute to reproductive incompatibilities and speciation (Henikoff et al. 2001).

Rosin, L. F. & Mellone, B. G. (2017). Centromeres drive a hard bargain. *Trends in Genetics, 33*(2), 101-17.

This comprehensive paper reviews the structure of centromeres and centromere proteins, the evolutionary patterns of centromeric histone proteins, and the evolutionary dynamics of centromere drive. It also presents the functional evidence for centromere drive and discusses outstanding questions about the rapid evolution of centromeres and centromere binding proteins.

Henikoff, S., Ahmad, K., & Malik, H. S. (2001), The centromere paradox: Stable inheritance with rapidly evolving DNA. *Science, 293*(5532), 1098-102.

This paper proposes centromere drive as a mechanism to explain how centromeres can evolve rapidly in the face of strong functional conservation. It describes how centromere drive may occur and the downstream consequences including postzygotic hybrid incompatibilities.

Chmatal, L., Gabriel, S. I., Mitsainas, G. P., Martinez-Vargas, J., Ventura, J., Searle, J. B., Schultz, R. M., & Lampson, M. A. (2014). Centromere strength provides the cell biological basis for meiotic drive and karyotype evolution in mice. *Current Biology, 24*(19), 2295-300.

The authors of this paper use a mouse model to follow the segregation of chromosomes that are heterozygous for a Robertsonian fusion. They show that nonrandom segregation occurs, and that the chromosome type (fused or unfused) that segregates to the polar body has a weaker centromere with reduced levels of centromere proteins than the chromosome type that preferentially segregates to the egg nucleus.

Zedek, F. & Bures, P. (2016). CenH3 evolution reflects meiotic symmetry as predicted by the centromere drive model. *Scientific Reports, 6*, 33308.

This paper conducts a comparative molecular evolutionary analyses of the Centromere histone H3 (CENH3) gene across a variety of eukaryotic taxa. The authors find that clades with asymmetric meiosis have a higher rate of protein evolution at than do clades with symmetric meiosis, consistent with the predictions of the centromere drive hypothesis.

de Villena, F. P. M. & Sapienza, C. (2001). Female meiosis drives karyotypic evolution in mammals. *Genetics, 159*(3), 1179-89.

This paper analyzes karyotype data from different orders of mammals and finds that the distribution of the number of chromosomes with a terminal centromere is nonrandom across species. The authors suggest that asymmetries in female meiosis may lead to directional selection for or against a particular chromosome morphology to result in karotypic differences across species.

Male Meiotic Drive

The mechanisms of male meiotic drive are necessarily different from female drive as all the products of male meiosis typically mature into sperm. As such, male meiotic drive often occurs through a killer system, where the driving gene sabotages the sperm that do not carry it. Male meiotic drive systems generally do not drive in females, and have been found on both the autosomes and the sex chromosomes. Because sex chromosomes control sex determination and have other unique properties, wide spread sex-linked meiotic drive can have evolutionary consequences much different from an autosomal driver.

Autosomal Drive

Overviews of Major Systems

Like most types of genetic conflict, male meiotic drive caused by autosomal genes is likely extremely common, but very difficult to detect. The best-studied

autosomal male drive systems are Segregation Distorter (SD) in *Drosophila melanogaster* (Sandler et al. 1959, Larracuente & Presegraves 2012) and t-haplotype in *Mus musculus* (Lyon 2003, Silver 1993). Both systems were first discovered fortuitously due to a cross with a visible marker. The SD system was first discovered when crossing wild-collected flies to lab flies carrying visible markers and offspring ratios were highly skewed towards the wild-type phenotypes (Sandler et al., 1959). The SD phenotype is only visible during microscopic analysis of spermatogenesis in SD heterozygous males: half of the sperm die during development (Larracuente & Presgraves 2012). The t-haplotype was similarly discovered accidentally; it carries an allele that renders mice completely tail-less when crossed to a particular lab strain (Lyon 2003). Half of the sperm produced by males heterozygous for the t-haplotype have serious motility defects that prevent them from successfully traversing the female reproductive tract (Lyon 2003). Both SD and t-haplotype have the same basic genetic framework of a responder locus linked to one or more distorter loci. In SD the responder and distorter loci are both located near the centromere of chromosome 2, and there are also segregating enhancers of drive, some of which are associated with inversions linking them to the distorter and responder (Larracuente & Presgraves 2012). The t-haplotype is substantially older than SD, and also substantially larger (Silver 1993, Lyon 2003). The t-haplotype makes up roughly 30 – 40 MB of chromosome 17, and consists of a single responder and up to four distorters (Lyon 2003). Multiple inversions cover this area, keeping the distorters and responders linked by suppressing recombination between the t-haplotype and the wild-type chromosome (Lyon 2003). Males that are homozygous for the t-haplotype are completely sterile due to sperm motility defects, indicating that at least one wild-type

gene copy of the distorter loci is required for proper sperm function (Lyon 2003). Like SD, there are multiple versions of t-haplotype segregating in the wild, which are typically differentiated by the specific embryonic recessive lethal alleles they carry (Silver 1993).

Larracuente, A. M., & Presgraves, D. C. (2012). The selfish Segregation Distorter gene complex of *Drosophila melanogaster*. *Genetics, 192*(1), 33-53.

This recent review is an excellent chronological summary of all the work on SD to date, including the molecular mechanisms, population genetics, molecular evolution, and evolutionary history of drive.

Sandler, L., Hiraizumi, Y., & Sandler, I. (1959). Meiotic drive in natural populations of *Drosophila melanogaster*. I. The cytogenetic basis of segregation-distortion. *Genetics, 44*(2), 233.

This paper is the first dedicated to SD, though it is mentioned as a footnote in an earlier study where it was discovered accidentally. This paper describes several structurally and genetically different wild-derived SD chromosomes and how they are inherited in lab crosses. They identify the location of SD near the centromere of the second chromosome and propose a mechanism for its activity.

Lyon, M. F. (2003). Transmission ratio distortion in mice. *Annual Review of Genetics, 37*, 393-408.

This review is a thorough explanation of the genetic work done using partial haplotypes to unravel the structure of the t-haplotype and the sperm dysfunction distortion phenotype. Candidates are presented for the distorter genes, though more recent work has disputed some of this speculation. The significance of embryonic lethal alleles is also discussed.

Silver, L. M. (1993). The peculiar journey of a selfish chromosome - mouse t-haplotypes and meiotic drive. *Trends in Genetics, 9*(7), 250-254.

This review provides a nice summary of the evolutionary research that has been done on the t-haplotype, particularly phylogenetic work estimating the origin of the system, the most recent common ancestor of the modern haplotype, and the selective value of linked embryonic lethal alleles.

Mechanisms

SD and t-haplotype occur in model genetic organisms, and thus both systems have both been the subject of careful molecular genetic dissection for decades. As a result, much is known about their mechanisms and many of the genes directly involved have been identified, though mysteries remain. SD and the t-haplotype both cause meiotic drive through a generally similar responder/distorter mechanism, but the details of two

systems are quite different. In SD, the responder locus, *Rsp,* is not a protein coding gene, but a satellite DNA repeat (Wu et al. 1988). The size of the repeat locus confers sensitivity or resistance to driving (Wu et al. 1988). The distorter locus, *sd,* is a truncated RanGAP duplicate that mislocalizes to the nucleus from the cytoplasm (Merrill et al. 1999; Kusano et al. 2001). It is unknown how it interacts with *Rsp* and how it causes the death of sperm carrying the sensitive responder allele. The t-haplotype has a more complex drive mechanism than SD, with at least three fixed distorter loci contributing to drive (Bauer et al. 2007). The responder, *Tcr,* is a hypomorphic duplicate of *smok,* a protein kinase required for sperm motility (Hermann et al. 1999). Interestingly, *Tcr* is transmitted at a lower rate than the wild-type allele when present without the distorters (Hermann et al. 1999). The distorters act additively to increase the transmission of *Tcr* to up to 99%. The distorters *Tcd1* and *Tcd2* have both been identified as independent upstream regulators of *smok* expression through Rho signaling (Bauer et al. 2005, Bauer et al. 2006). Both *Tcd1* and *Tcd2* upregulate *smok* expression, suggesting that overexpression of the wild type protein causes sperm motility defects that are rescued by the low expression of *Tcr* (Bauer et al. 2007).

Wu, C.-I., Lyttle, T. W., Wu, M.-L., & Lin, G.-F. (1988). Association between a satellite DNA sequence and the responder of Segregation Distorter in *D. melanogaster*. *Cell, 54*(2), 179-189.

This paper shows that *Rsp* corresponds to a satellite DNA sequence, which is found in variable copy numbers. The amount of copies present corresponds to the sensitivity to

drive – chromosomes with many copies of *Rsp* are highly sensitive to drive, while those with few are resistant.

Merrill, C., Bayraktaroglu, L., Kusano, A., & Ganetzky, B. (1999). Truncated RanGAP encoded by the Segregation Distorter locus of Drosophila. *Science, 283*(5408), 1742-1745.

This paper shows that the SD distorter locus is a duplicate copy of RanGAP with the C-terminus truncated. RanGAP is a nuclear transport protein that plays an important regulatory role in allowing transport to occur.

Kusano, A., Staber, C., & Ganetzky, B. (2001). Nuclear mislocalization of enzymatically active RanGAP causes segregation distortion in Drosophila. *Developmental Cell, 1*(3), 351-361.

This paper shows that drive occurs when the duplicate copy of RanGAP is mislocalized to the nucleus rather than remaining in the cytoplasm where the wild type protein is found, suggesting that the disruption of nuclear transport and specifically reduced levels of nuclear RanGTP causes drive.

Herrmann, B. G., Koschorz, B., Wertz, K., McLaughlin, K. J., & Kispert, A. (1999). A protein kinase encoded by the t complex responder gene causes non-mendelian inheritance. *Nature, 402*(6758), 141-146.

This paper shows that the responder in the t-complex is a duplicate copy of a protein kinase called *smok* that plays an important role in sperm flagellar motion. Without distorters present, the mutant responder is expressed at a much lower level than the wild-type protein.

Bauer, H., Willert, J. R., Koschorz, B., & Herrmann, B. G. (2005). The t-complex-encoded GTPase-activating protein *Tagap1* acts as a transmission ratio distorter in mice. *Nature Genetics, 37*(9), 969-973.

This paper shows that the *Tcd1* t-haplotype distorter locus is four duplicate copies of a GTPase-activating protein called *Tagap1*. One duplicate encodes a truncated protein. Together, the components of the *Tcd1* locus upregulate *smok* expression through Rho signaling.

Bauer, H., Veron, N., Willert, J., & Herrmann, B. G. (2007). The t-complex-encoded guanine nucleotide exchange factor *Fgd2* reveals that two opposing signaling pathways promote transmission ratio distortion in the mouse. *Genes & Development, 21*(2), 143-147.

This paper shows that the t-haplotype distorter *Tcd2* encodes a mutant copy of *Fgd2*, which also regulates *smok* expression through Rho signaling, though in an independent pathway from *Tcd1*. A single amino acid change in the t-haplotype copy causes upregulation of *smok* compared to the wild-type.

Molecular Evolution

The differences in the age and size of SD and the t-complex have contributed to their molecular evolutionary differences. Molecular phylogenetics shows that the oldest inversion on the t-haplotype arose around 3 million years ago (Hammer & Silver 1993). The presence of unique recessive lethal alleles shows there are multiple segregating t-haplotypes in the wild, but diversity between haplotypes is very low, suggesting a recent sweep of a particular haplotype to high frequency (Hammer & Silver 1993). The SD chromosome is estimated to be much younger, with the causal RanGAP duplication having occurred around 40,000 years ago (Brand et al. 2015). Though the SD responder and distorter system encompasses a large portion of chromosome 2, there are relatively few nucleotide differences between SD chromosomes and wild-type chromosomes at neutral loci that are physically linked to the driver and responder (Palopoli & Wu 1996).

Hammer, M. F., & Silver, L. M. (1993). Phylogenetic analysis of the alpha-globin pseudogene-4 (Hba-ps4) locus in the house mouse species complex reveals a stepwise evolution of t haplotypes. *Molecular Biology and Evolution, 10*(5), 971-1001.

This paper uses pseudogenes on the t-haplotype to show that the oldest inversion on the t-haplotype occurred about 3 million years ago and that the youngest inversion is at least 1.5 million years old. The authors propose a stepwise model for the evolution of the entire t-complex, and discuss evidence for a recent selective sweep of a newer haplotype.

Brand, C. L., Larracuente, A. M., & Presgraves, D. C. (2015). Origin, evolution, and population genetics of the selfish Segregation Distorter gene duplication in European and African populations of *Drosophila melanogaster. Evolution, 69*(5), 1271-1283. doi:10.1111/evo.12658

This paper examines the molecular evolutionary patterns of the duplicate RanGAP distorter locus to show that multiple recurrent selective sweeps have occurred in the history of SD, keeping the appearance of a steady population frequency as individual SD chromosomes increase and decrease in frequency.

Palopoli, M. F., & Wu, C. I. (1996). Rapid evolution of a coadapted gene complex: Evidence from the Segregation Distorter (Sd) system of meiotic drive in *Drosophila melanogaster*. *Genetics, 143*(4), 1675-1688.

This paper examines variation at loci that are located near SD, but not involved in drive. They show that there is no detectable divergence between SD and the wild type at these sites, and consequently the SD system must be very young.

Evolutionary Theory

The expected population dynamics of autosomal male meiotic drivers has been the subject of much theoretical work, particularly in light of the unexplained low natural frequencies of SD and t-haplotype. In the case of t-haplotype, models incorporating the lethality of homozygous t-haplotype individuals suggest a much higher equilibrium frequency than is observed in nature (Lewontin & Dunn 1960). However, the highly subdivided structure of house mice populations is also predicted to have significant effects on the frequency of t-haplotypes (Lewontin 1962). In contrast, the low equilibrium frequencies of SD can be explained fairly well by a combination of negative fitness effects in homozygotes, reduced fertility of SD males, and the presence of costly segregating suppressors on wild-type chromosomes (Charlesworth & Hartl, 1978). Theory also shows the persistence of drive systems that require multiple loci to drive, such as SD and t-haplotype, can be explained by linkage disequilibrium between the components (Prout et al. 1973).

Lewontin, R. C., & Dunn, L. C. (1960). The evolutionary dynamics of a polymorphism in the house mouse. *Genetics, 45*(6), 705-722.

This paper uses a mathematical model incorporating the strength of drive and the lethality of homozygous individuals to predict a relatively high equilibrium frequency for t-haplotype, which is not seen in nature. Stochastic modeling then suggests that in the small sub-populations house mice favor, t-haplotype will often be lost due to genetic drift.

Lewontin, R. C. (1962). Interdeme selection controlling a polymorphism in the house mouse. *The American Naturalist, 96*(887), 65-78.

This paper uses mathematical modeling to predict the frequency of t-haplotypes that cause male sterility when homozygous. This model also predicts their behavior in small, subdivided populations, showing that small populations may be prone to high levels of t-haplotype and thus extinction once all males are sterile.

Charlesworth, B., & Hartl, D. L. (1978). Population-dynamics of segregation distorter polymorphism of *Drosophila melanogaster*. *Genetics, 89*(1), 171-192.

This model shows that the two-locus SD system can be maintained at a low equilibrium frequency due to fitness effects of the driving and resistant genotypes. Resistant alleles

never go to fixation because they are less fit than the wild type allele when drive is not present.

Prout, T., Bundgaard, J., & Bryant, S. (1973). Population genetics of modifiers of meiotic drive I. The solution of a special case and some general implications. *Theoretical Population Biology, 4*(4), 446-465.

This paper shows that the two-locus SD system requires a certain amount of linkage disequilibrium between the driver and responder to be maintained in the wild, either through physical proximity or genetic suppression of recombination. The paper considers both modifiers and suppressors of drive.

Population Evolution

The simplest expectation for meiotic drive elements is that once they enter a population, they will rapidly sweep to fixation. However, both SD and the t-haplotype have been found at relatively low stable frequencies in wild populations. The question of what factors keep them at these levels is quite complex and has many possible answers; this issue has been best studied for t-haplotypes. The homozygous male sterility and embryonic lethality of the t-haplotype will prevent it from ever reaching fixation, but it is likely that mating behavior also helps keep wild frequencies of t-haplotype low (Ardlie 1998). Polyandry, or female multiple mating, can lessen the selfish effect of t-haplotype as males carrying it produce half the amount of motile sperm that wild-type males do (Maser at al. 2011). The remaining sperm produced by these males are also less fit than

sperm produced by wild type males, and a female mated to both will produce more offspring fathered by the wild type male (Sutter & Lindholm 2015). Together, all of these factors lessen the impact of t-haplotype drive to inhibit its spread.

Ardlie, K. G. (1998). Putting the brake on drive: Meiotic drive of t-haplotypes in natural populations of mice. *Trends in Genetics, 14*(5), 189-193.

This excellent review summarizes research on the numerous factors that could keep the frequency of t-haplotype in check, including fitness effects in homozygotes and heterozygotes, multiple mating and mating systems, and population structure.

Manser, A., Lindholm, A. K., Koenig, B., & Bagheri, H. C. (2011). Polyandry and the decrease of a selfish genetic element in a wild house mouse population. *Evolution, 65*(9), 2435-2447.

This paper uses empirical study and theoretical modeling to show that the fitness effects of t-haplotype cannot explain the observed decreases in its frequency in natural populations, but that polyandry could keep frequencies of the t-haplotype low in natural populations.

Sutter, A., & Lindholm, A. K. (2015). Detrimental effects of an autosomal selfish genetic element on sperm competitiveness in house mice. *Proceedings of the Royal Society B-Biological Sciences, 282*(1811), 20150974.

This paper shows direct experimental evidence that males carrying a t-haplotype are poor sperm competitors. When mated to the same female, these males sire proportionally less offspring than wild type males regardless of the order of mating.

Sex Chromosome Drive

Overviews of Major Systems

Selfish sex chromosomes are much easier to detect than autosomal meiotic drivers because they have an intrinsic, easily visible phenotype: a skewed offspring sex ratio. For this reason, driving X-chromosomes are often referred to as *sex-ratio* X-chromosomes. In XY sex chromosome systems, X-linked drivers prevent the maturation of Y-bearing sperm, thus causing a carrier male to sire a majority of daughters, while Y-linked drivers will result in a majority of sons. Either of these can be highly detrimental to a population if found at high frequencies (Meiklejohn & Tao 2010). There are over 20 independently evolved *sex-ratio* systems identified in the genus *Drosophila* alone, as well as in closely related Dipterans such as mosquitos and stalk-eyed flies (Jaenike 2001). It is unknown if there is some property of Dipteran sex chromosomes that make them particularly susceptible to genetic conflict, but it is likely that there is an element of discovery bias in this well-studied group. Though other taxa, including mammals, are known to carry selfish sex chromosomes, the bulk of research in this area has been done in *Drosophila*

(Jaenike 2001). The majority of these systems arose independently and occur through different molecular mechanisms, as they are scattered through the genus and vary in properties such as inversion content, strength of drive, and presence of suppressors (Jaenike 2001). *Sex-ratio* X-chromosomes were first discovered in the *Drosophila obscura* species group and were among the very first identified meiotic drivers (Sturtevant & Dobzhansky 1936). Because of the way sex chromosomes behave and their unique genetic properties, they may be more prone to conflict than the autosomes, and conflict may have had significant impacts on their evolution (Meiklejohn & Tao 2010).

Meiklejohn, C. D., & Tao, Y. (2010). Genetic conflict and sex chromosome evolution. *Trends in Ecology & Evolution, 25*(4), 215-223.

This review paper describes various ways in which genetic conflict can shape sex chromosome evolution, including the regulation of gene expression from the sex chromosomes, the location of genes in the genome, and speciation via hybrid sterility.

Jaenike, J. (2001). Sex chromosome meiotic drive. *Annual Review of Ecology and Systematics, 32*, 25-49.

This review is an excellently thorough overview of the details of known sex chromosomes meiotic drivers, from the very well-studied to the obscure. The review focuses heavily on Dipterans, but also covers selfish sex chromosomes in mammals and plants. Mechanisms of drive and evolutionary consequences are both discussed. The

supplemental material to this review contains a truly exhaustive summary of each drive system mentioned in the main text.

Sturtevant, A., & Dobzhansky, T. (1936). Geographical distribution and cytology of "sex ratio" in *Drosophila pseudoobscura* and related species. *Genetics, 21*(4), 473.

This paper describes the *sex-ratio* system in *D. pseudoobscura* and several other closely related species, including the sister species *D. persimilis* (referred to in the paper as *D. pseudoobscura* race B). Earlier identification of *sex-ratio* in *D. affinis* and *D. obscura* are also discussed, followed by an investigation into the geographical and taxonomic distribution of drive, its genetic characteristics, and its spermatogenesis phenotype.

Mechanisms

It is no coincidence that the only sex chromosome drive systems with known molecular mechanisms are found in *Drosophila simulans.* This species is closely related to the model *D. melanogaster* and harbors at least three independent sex linked meiotic drivers. The genetic basis of drive has been identified in two of these, called the Winters and Paris systems. Importantly, there are no inversions associated with either. The drive phenotype of the Winters *sex-ratio* system was first discovered when crosses were made with the sister species, *D. sechellia,* separating the X-linked driver from its autosomal suppressor (Tao et al. 2007a). The autosomal suppressor (*Nmy*) appears to be the result of a retrotransposition from the X-chromosome to the autosomes of the distorter (*Dox*). *Dox* itself arose via duplication from another X-linked gene (*MDox*), though the molecular

function of both genes is still unknown and they appear to have limited coding potential (Tao et al. 2007b). The suppression of drive likely occurs through a RNA interference mechanism, and it is possible the distorter itself is also a noncoding RNA (Tao et al. 2007a, Tao et al. 2007b). The Paris *sex-ratio* system of *D. simulans* was discovered segregating in the wild and is independent of the Winters system (Helleu et al. 2016). In this system, an X-linked gene called *HP1D2* causes drive by binding to the heterochromatin of Y-chromatids during meiosis II and preventing them from separating correctly, resulting in the successful production of only X-bearing sperm (Helleu et al. 2016).

Tao, Y., Masly, J. P., Araripe, L., Ke, Y., & Hartl, D. L. (2007a). A *sex-ratio* meiotic drive system in *Drosophila simulans*. I: An autosomal suppressor. *PLOS Biology, 5*(11), e292.

This paper is the first of a pair that describes the molecular mechanism of the Winters sex-ratio system. It describes the sequence and location of the dominant autosomal suppressor, *Nmy,* and its distribution in species closely related to *D. simulans.* The suppressor is a duplicate of the distorter and likely works through RNA interference. The authors also show it only suppresses the Winters distorter.

Tao, Y., Araripe, L., Kingan, S. B., Ke, Y., Xiao, H., & Hartl, D. L. (2007b). A *sex-ratio* meiotic drive system in *Drosophila simulans*. II: An X-linked distorter. *PLOS Biology, 5*(11), e293.

This paper is the second in a pair of papers describing the genetic basis of the Winters *sex-ratio* system. It describes the identification of the X-linked distorter gene, *Dox,* which itself is a duplicate of another X-linked gene, *MDox*. Neither gene has a known function or identified orthologs, and both have limited protein-coding potential. The authors discuss RNA interference mechanisms for the suppression of drive and its evolutionary implications.

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

This paper describes the molecular mechanism of *sex-ratio* drive in the Paris system of *D. simulans.* It is the first study to pinpoint the genes and show the causal molecular mechanism of *sex-ratio* drive. The X-linked distorter is a mutant allele of *HP1D2,* a fast-evolving heterochromatin binding protein. The distorting allele makes a protein that binds to the heterochromatin of sister Y-chromatids and prevents them from separating during meiosis II.

Molecular Evolution

The molecular evolutionary patterns of *sex-ratio* X-chromosomes are highly variable depending on the characteristics of the drive system. The rapid increase in frequency of a meiotic driver will result in evolutionary patterns similar to that of a selective sweep, which is clearly documented in the *sex-ratio* systems of *D. simulans* that lack inversions (Kingan et al. 2010). Chromosomal inversions are common on *sex-ratio* chromosomes and block recombination between the driving and the wild type variant, which can have an important effect on molecular evolution. The *sex-ratio* chromosome of *D. recens* is highly monomorphic, a result of being encompassed by overlapping inversions that suppress recombination with the wild-type chromosome, as well as a recessive mutation that renders females homozygous for *sex-ratio* sterile (Dyer et al. 2007). In contrast, the *sex-ratio* chromosome of *D. pseudoobscura* is found at high frequencies in the wild and only the breakpoints of inversions show substantial differentiation between the *sex-ratio* and wild type chromosomes (Kovacevic & Schaeffer 2000). The *sex-ratio* chromosome in *D. neotestacea* is also found at high frequencies in the wild, but it is highly inverted and there is strong genetic differentiation with the wild type X-chromosome (Pieper & Dyer 2016). An analysis of genes expressed in male stalk eyed flies carrying wild type versus *sex-ratio* X-chromosomes also shows substantial genetic differentiation between the chromosome types on a large scale (Reinhardt et al. 2014).

Kingan, S. B., Garrigan, D., & Hartl, D. L. (2010). Recurrent selection on the Winters *sex-ratio* genes in *Drosophila simulans*. *Genetics, 184*(1), 253-265. *Molecular Biology and Evolution, 25*(2), 409-416.

This paper shows evidence of a selective sweep in the region of the distorter and autosomal suppressor loci in the Winters *sex-ratio* system in *D. simulans*. The authors compare their findings with results from the Paris system in the same species, which shows a more recent selective sweep around the distorter but very similar patterns overall. Neither system contains chromosomal inversions.

Dyer, K. A., Charlesworth, B., & Jaenike, J. (2007). Chromosome-wide linkage disequilibrium as a consequence of meiotic drive. *Proceedings of the National Academy of Sciences of the United States of America, 104*(5), 1587-1592.

This paper shows that the *sex-ratio* chromosome of *D. recens* is highly monomorphic, likely a consequence of its inversions and very small effective population size. It also causes sterility when homozygous in females, and thus the *sex-ratio* chromosome never recombines with the wild-type chromosome. The female sterility is likely the result of hitchhiking deleterious alleles.

Kovacevic, M., & Schaeffer, S. W. (2000). Molecular population genetics of X-linked genes in *Drosophila pseudoobscura*. *Genetics, 156*(1), 155-172.

This paper shows that the only differentiation between the standard and *sex-ratio* X-chromosomes of *D. pseudoobscura* exists near the breakpoints of the inversions, and genes within the inversions that are not involved in drive that are not significantly differentiated between chromosome types.

Pieper, K. E., & Dyer, K. A. (2016). Occasional recombination of a selfish X-chromosome may permit its persistence at high frequencies in the wild. *Journal of Evolutionary Biology, 29*(11), 2229-2241.

This paper shows that despite the presence of multiple overlapping inversions on the *sex-ratio* X-chromosome in *D. neotestacea* and high overall genetic differentiation with the wild type X-chromosome, there is a substantial amount of variation present on the selfish X-chromosome. This diversity is likely maintained by occasional gene conversion between the two chromosome types and recombination in *sex-ratio* homozygous females.

Reinhardt, J. A., Brand, C. L., Paczolt, K. A., Johns, P. M., Baker, R. H., & Wilkinson, G. S. (2014). Meiotic drive impacts expression and evolution of X-Linked genes in stalk-eyed flies. *PLOS Genetics, 10*(5), e1004362.

This paper is the first genome-scale investigation into the gene expression consequences of a *sex-ratio* X-chromosome. It shows that in stalk-eyed flies, the selfish activity of the X-linked meiotic driver has widespread effects on the sequence evolution and the expression of genes located with it on the driving X-chromosome.

Consequences for Speciation

The X-chromosome has long been known to have a disproportionately large impact on the genetics of speciation, referred to as the "large X-effect". This effect is often seen through Haldane's rule, which states that if only one sex is sterile or inviable in an interspecies cross, it is usually the heterogametic sex. Sex chromosome linked meiotic drive has been suggested as an explanation for Haldane's Rule (Hurst & Pomiankowski 1991). If X-linked distorters and Y-linked or autosomal suppressors become fixed within species, when they are separated via hybridization the driver becomes active again and male sterility results (Hurst & Pomiankowski 1991). Indeed, there is clear evidence of sex ratio distortion appearing alongside hybrid sterility in crosses between several different *Drosophila* species pairs. Crosses between *D. simulans* and *D. mauritiana* reveals hybrid male sterility as well as sex ratio distortion, and fine mapping of these two phenotypes reveal their genetic basis lies in the same 20 KB region of the X-chromosome (Tao et al. 2001). Even more convincingly, hybrid male sterility

and sex ratio distortion in crosses of two subspecies of *D. pseudoobscura* are both directly attributable to the same gene, *Overdrive* (Phadnis & Orr 2009). Crosses between stalk-eyed flies from isolated populations in Southeast Asia have also shown an overlap between genetic regions involved in male sterility, sperm length, and progeny sex ratios (Wilkinson et al. 2014).

Hurst, L. D., & Pomiankowski, A. (1991). Causes of sex-ratio bias may account for unisexual sterility in hybrids - a new explanation of Haldane's rule and related phenomena. *Genetics, 128*(4), 841-858.

This paper is one of the first to suggest that meiotic drive may be an important cause of hybrid male sterility and thus Haldane's rule. The authors use mathematical modeling to argue that sex chromosomes are more likely to acquire drivers than autosomes, and point out the overlap between taxa where Haldane's rule is true and where sex-linked meiotic drive is found.

Tao, Y., Hartl, D. L., & Laurie, C. C. (2001). *Sex-ratio* segregation distortion associated with reproductive isolation in *Drosophila*. *Proceedings of the National Academy of Sciences*, *98*(23), 13183-13188.

This paper describes the Durham *sex-ratio* system in *Drosophila simulans,* which has been mapped down to the same small region of the X-chromosome that has been linked to hybrid male sterility. This is a cryptic drive system due to the fixation of an autosomal

suppressor in *D. simulans.* Introgressing the non-suppressing allele from *D. mauritiana* reveals the sex ratio distortion phenotype.

Phadnis, N., & Orr, H. A. (2009). A single gene causes both male sterility and segregation distortion in Drosophila hybrids. *Science, 323*(5912), 376-379.

This paper shows very elegantly that the same gene causes hybrid male sterility and segregation distortion in crosses between *Drosophila pseudoobscura* subspecies from the mainland USA and from Bogota, Columbia. Eight nonsynonymous nucleotide substitutions are identified between the two alleles.

Wilkinson, G. S., Christianson, S. J., Brand, C. L., Ru, G., & Shell, W. (2014). Haldane's Rule Is Linked to extraordinary sex ratios and sperm length in stalk-eyed flies. *Genetics,* 198(3), 1167-1681.

This paper describes a scenario where cryptic drivers and suppressors have fixed between geographically separated populations of stalk-eyed flies. QTL mapping of sperm length, hybrid male sterility, and progeny sex ratio all point to the involvement of the same genetic regions of the X-chromosome.

Theory

Classical evolutionary theory states that the optimal offspring sex ratio for an autosomal locus is 50% males and 50% females. In contrast, a female bias in the

offspring is favored by the X-chromosome because it is present more often in females than in males, while the opposite is true for the Y-chromosome (Hamilton 1967). Early work showed that a driving X chromosome should rapidly spread and thus drive the host population (and itself) to extinction (Hamilton 1967). But as *sex-ratio* chromosomes are usually found at intermediate and low frequencies, there must be other factors that prevent them from driving to fixation. Rigorous mathematical modeling of the dynamics of sex-linked meiotic drivers have attempted to explain their continued persistence in the wild. Fitness costs in homozygous females can keep driving X-chromosomes at stable equilibrium frequencies (Curtsinger & Feldman 1980), as can the presence of Y-linked suppressors of drive (Jaenike 1999). When the population is female biased and females remate frequently, *sex-ratio* carrying males that produce half the sperm of wild type males may become sperm depleted and sire fewer offspring than wild type males, leading to a decrease in driver frequency (Jaenike 1996). Theory also predicts that fitness costs of Y-linked suppressors may lead to wild type, suppressing, and driving sex chromosomes cyclically rising and falling in frequency (Hall 2004).

Hamilton, W. D. (1967). Extraordinary sex ratios. *Science, 156*(3774), 477-488.

This paper models the sex ratios and growth of populations carrying driving X or Y-chromosomes. It shows that though a driving X-chromosome might initially lead to population expansion if there are still enough males to fertilize the females, eventually the population will crash.

Curtsinger, J. W., & Feldman, M. W. (1980). Experimental and theoretical analysis of the "sex-ratio" polymorphism in *Drosophila pseudoobscura*. *Genetics,* 94(2), 445-466.

This paper shows that fitness costs of *sex-ratio* meiotic drive in males will have no effect on driver frequency, but deleterious effects on female fitness are enough to prevent a driving chromosome from reaching fixation. The authors demonstrate the validity of their model in experimental populations of *D. pseudoobscura* carrying *sex-ratio* X-chromosomes.

Jaenike, J. (1999). Suppression of sex-ratio meiotic drive and the maintenance of Y-chromosome polymorphism in *Drosophila*. *Evolution*, 53(1), 164-174.

This paper discusses the presence of Y-linked suppressors of sex-ratio meiotic drive in *Drosophila quinaria,* and presents a model showing that the presence of segregating suppressors could maintain stable frequencies of the driving X-chromosome.

Jaenike, J. (1996). Sex-ratio meiotic drive in the *Drosophila quinaria* group. *American Naturalist, 148*(2), 237-254.

This paper describes two driving X-chromosome systems and presents empirical evidence that males carrying *sex-ratio* are less fertile than wild type males. A mathematical model for how sperm depletion and female remating rate could stabilize

sex-ratio polymorphism without the presence of suppressors or vitality costs is presented. There is also an excellent summary of previous theoretical work predicting drive frequencies in the Introduction.

Hall, D. W. (2004). Meiotic drive and sex chromosome cycling. *Evolution, 58*(5), 925-931.

This paper presents a model predicting the frequencies of four different kinds of sex chromosomes: driving and wild type X-chromosomes, and sensitive and resistant Y-chromosomes. Depending on the fitness costs of each variety, a stable equilibrium with all four present is possible, or populations may experience cycling between the different kinds of X and Y-chromosomes.

Consequences on Population Evolution

Sex-chromosome linked meiotic drive has particularly interesting population dynamics because it has the potential to seriously skew the sex ratio of a population and possibly lead to its extinction. The question of why sex-linked drive persists in the wild is a tricky one with many possible answers depending on the system. In the case of the Paris *sex-ratio* system in *D. simulans,* segregating suppressors of drive likely keep the frequency of driver in check, though the frequency of driving X-chromosomes and suppressors is rapidly changing in some populations (Bastide et al. 2013). In other species, *sex-ratio* chromosomes have been shown to persist at high stable frequencies in the wild without suppressors. The *sex-ratio* X-chromosome in *D. neotestacea* is

distributed in a cline, suggesting ecological fitness costs for carrying it (Dyer 2012). There is also evidence that variation in the frequency of *sex-ratio* in populations of *D. neotestacea* and *D. pseudoobscura* might be maintained by variation in female mating rate: males carrying *sex-ratio* produce fewer sperm than wild type males and as a consequence high remating rates may inhibit the spread of *sex-ratio* (Price et al. 2008; Pinzone & Dyer 2013). Males carrying *sex-ratio* drive have been shown to be less attractive than wild-type males to females in stalk-eyed flies, as the driver is linked to the unpreferred smaller eyestalk size (Cotton et al. 2014). It is possible that female remating rate and female preference are evolving in response to *sex-ratio* drive, as females who mate with wild-type males generally have fitter offspring (Price et al. 2008, Cotton et al. 2014)

Bastide, H., Gerard, P. R., Ogereau, D., Cazemajor, M., & Montchamp-Moreau, C. (2013). Local dynamics of a fast-evolving *sex-ratio* system in *Drosophila simulans*. *Molecular Ecology, 22*(21), 5352-5367.

This paper shows that the Paris *sex-ratio* system of *D. simulans* is rapidly increasing in frequency in some locations, but decreasing in others. These changes in driver frequency are accompanied by changes in frequency of suppressors of drive. The dynamics of this very young system show an evolutionary arms race between the distorter and its suppressors.

Dyer, K. A. (2012). Local selection underlies the geographic distribution of sex-ratio drive in *Drosophila neotestacea*. *Evolution, 66*(4), 973-984.

This paper shows that the frequency of the *sex-ratio* X-chromosome in *D. neotestacea* has remained steady over the course of decades. The prevalence of *sex-ratio* varies among populations in a clinal pattern that is best predicted by overwinter temperature, suggesting that selection from ecological sources plays a role in keeping them steady.

Price, T. A. R., Hodgson, D. J., Lewis, Z., Hurst, G. D. D., & Wedell, N. (2008). Selfish genetic elements promote polyandry in a fly. *Science, 322*(5905), 1241-1243.

This paper presents an experimental evolution study where *D. pseudoobscura* females evolved higher remating rates when exposed to males carrying *sex-ratio* X-chromosomes. This may be because remating generally increases offspring fitness through sperm competition, and the offspring of wild type males are generally more fit.

Pinzone, C. A., & Dyer, K. A. (2013). Association of polyandry and *sex-ratio* drive prevalence in natural populations of *Drosophila neotestacea*. *Proceedings of the Royal Society B-Biological Sciences, 280*(1769).

This paper shows that *D. neotestacea* females from natural populations where *sex-ratio* is at low frequencies remate at a higher rate than females from populations with a high

frequency of drive. They also present evidence of a possible near extinction event due to *sex-ratio* in a semi-isolated population.

Cotton, A. J., Foeldvari, M., Cotton, S., & Pomiankowski, A. (2014). Male eyespan size is associated with meiotic drive in wild stalk-eyed flies (*Teleopsis dalmanni*). *Heredity, 112*(4), 363-369.

This paper shows that X-chromosome drive in stalk-eyed flies is genetically linked with male eye span size, a sexually selected trait. Males with very long stalk eyes are more attractive to females, but the *sex-ratio* driver in this species is associated with shorter eyestalk length. The female preference may have evolved to prevent mating with *sex-ratio* males, and could keep the frequency of the driver in check.

Spore Killer Drive

In fungi, meiotic drive is manifested as spore killing (Sk), in which ascospores (sexual spores) that carry the Sk element survive preferentially (Raju 1994; Turner & Perkins 1991). In a cross between a killer and a sensitive strain, only progeny that contain the killer allele survive. Thus, within each asci half of the haploid nuclei that result from meiosis are inviable. Homozygous crosses between killer or insensitive strains yield normal asci. The best studied Sks are in *Neurospora*, which has three different Sk systems (Turner & Perkins 1991), and in *Podospora anserina*, which harbors at least eight Sk systems (van der Gaag et al. 2000). The molecular basis of several Sk systems has been identified. In *Neurospora*, the Sk-2 system uses a poison-antidote mechanism,

where the antidote protein that is produced by the Sk system provides self-resistance and allows the Sk spores to survive, while spores without it die (Hammond et al. 2012). In one of the Sk systems in *P. anserina*, the Sk driving gene encodes a prion protein that becomes toxic when it changes conformation and causes the abortion of spores that carry the alternate protein form of this allele (Dalstra et al. 2003). Finally, the drive systems of two other Sks in *P. anserina* present a third molecular mechanism, in which the same gene product acts to confer both killing and resistance (Grognet et al. 2014). Studies of the evolutionary dynamics of spore killers using a two-locus model indicate that Sk can only be maintained at a stable equilibrium in a population if there are resistant strains present (Nauta and Hoestra 1993). Similar to other types of meiotic drivers, Sk strains generally occur at low to moderate prevalence in natural populations but their frequency varies across species and geographic populations (Turner and Perkins 1991; van der Gaag et al. 2000).

Raju, N. B. (1979). Cytogenetic behavior of Spore Killer genes in *Neurospora*. *Genetics, 93*(3), 607-23.

This paper reviews the early cytological and genetic analyses of Spore killer (Sk) systems. The author describes the characteristics of Sks and how to detect them, and then discusses the known Sk systems from different species and some of the evolutionary implications of Sks.

Turner, B. C. & Perkins, D. D. (1991). Meiotic drive in *Neurospora* and other fungi. *American Naturalist, 137*(3), 416-29.

This comprehensive review discusses the cytology and genetics of the three *Neurospora* Spore killer systems, their chromosomal basis, and their occurrence in nature.

van der Gaag, M., Debets, A. J. M., Oosterhof, J., Slakhorst, M., Thijssen, J. A. G. M., & Hoekstra, R. F. (2000). Spore-killing meiotic drive factors in a natural population of the fungus *Podospora anserina. Genetics, 156*(2), 593-605.

Van der Gaag and colleagues describe a survey of natural populations of *Podospora anserina* for Spore killer elements. They identify seven types of Sks and characterize the interactions between the different systems. They find that the prevalence of Sk is over 20% in natural populations, and that several of the systems map to a single genomic region.

Hammond, T. M., Rehard, D. G., Xiao, H., & Shiu, P. K. T. (2012). Molecular dissection of *Neurospora* Spore killer meiotic drive elements. *Proceedings of the National Academy of Sciences of the United States of America, 109*(30), 12093-98.

This paper identifies the Sk resistance gene *rsk* in the Sk-2 and Sk-3 systems of *Neurospora*. This gene provides resistance to Sk as the antidote part of the poison-antidote mechanism of meiotic drive, as when it is present neutralizes the killer element

and allows for normal ascospore development. It occurs in Sk-2, Sk-3, and naturally resistant isolates.

Dalstra, H. J. P., Swart, K., Debets, A. J. M., Saupe, S. J., & Hoekstra, R. F. (2003). Sexual transmission of the [Het-s] prion leads to meiotic drive in *Podospora anserina*. *Proceedings of the National Academy of Sciences of the United States of America, 100*(11), 6616-21.

In this paper, the authors dissect the genetic basis of the *het-s* Sk system in *Podospora anserina*. They show that in its prion form the HET-s protein can be sexually transmitted and is responsible for *het-S* spore abortion, thus causing meiotic drive.

Grognet, P., Lalucque, H., Malagnac, F., & Silar, P. (2014). Genes that bias Mendelian segregation. *PLOS Genetics, 10*(5), e1004387.

This paper dissects the molecular basis of two Sk systems in *Podospora anserina*. The authors identify two related Sk genes, *Spok1* and *Spok2*. A single allele of each of these genes acts not only as a spore-killing distorter as well as a spore-protecting responder. These alleles act independent of their genome location and cause Sk when inserted in other fungi.

Nauta, M. J. & Hoekstra, R. F. (1993). Evolutionary dynamics of Spore Killers. *Genetics, 135*(3), 923-30.

This paper presents what may be the only theoretical investigation of the evolutionary dynamics of Spore killers. It assumes a linked two-locus driver-responder genetic system as is found in *Neurospora*. The dynamics of Sks are different from other forms of meiotic drive, as the absolute number of progeny that carry the killer allele is not higher due to Sk and fungi spend most of their life cycle as haploids.

Synthetic Meiotic Drive

Synthetic gene drive systems were initially proposed and modelled by Burt (2003), which suggested that homing endonuclease genes (HEGs) could be engineered to target new host sequences and then used to manipulate natural populations. In cells that are heterozygous for the presence of an HEG, the HEG cuts the DNA strand of the non-HEG allele, which induces the cell to repair the break by copying the HEG gene onto the damaged non-HEG carrying chromosome via homologous recombination. As a result of this 'homing', both chromosomes now contain the HEG and the HEG enjoys increased transmission to the next generation. In addition to HEGs, other types of proposed gene drive systems include reproductive incompatibilities caused by *Wolbachia* endosymbionts, transposable elements, sex-linked meiotic drivers, maternal effect dominant embryonic arrest (*Medea*) elements, and the reduced fitness of individuals that are heterozygous as a locus (Burt 2014, Champer et al. 2016). The development of synthetic gene drive systems is a young and rapidly developing field, with a recent focus

on HEGs, specifically RNA-guided gene drive systems that utilize the clustered regularly interspaced short palindromic repeats (CRISPR) and endonuclease Cas9 (Esvelt et al. 2014). Proof of principle experiments using CRISPR/Cas9 gene drive systems have been completed in both model organisms and disease vectors, including the budding yeast (DeCarlo et al. 2015), the fruit fly *Drosophila melanogaster* (Gantz and Bier 2015), and the mosquitoes *Anopheles gambiae* (Hammond et al. 2016) and *A. stephani* (Gantz et al. 2015). Gene drive systems have been proposed as a mechanism to eradicate insect-borne infectious diseases, assist in the conservation of threatened and endangered species, control invasive species, and develop species-specific pesticides and herbicides. However, there may be unintended consequences once a drive system is released into the wild, for instance horizontal transfer to a nontarget species and the evolution of host resistance (Esvelt et al. 2014). There are several excellent reviews of gene drive that discuss both the genetic mechanisms as well as the biological and political considerations and limitations of its use in natural populations (Sinkins and Gould 2006; Burt 2014; Champer et al. 2016; Esvelt et al. 2014). The safety of gene drive is of broad significance, and in response the National Academies met in 2016 to formulate recommendations for gene drive development and use (National Academies Press 2016).

Burt, A. (2003). Site-specific selfish genes as tools for the control and genetic engineering of natural populations. *Proceedings of the Royal Society B-Biological Sciences, 270*(1518), 921-28.

This paper is the first to suggest that a site-specific selfish gene might be used to drive a foreign gene into a natural population. This paper presents the basic idea of gene drive and presents a theoretical treatment of how gene drive could work. It also discusses the evolution of resistance to the gene drivers and the types of selfish genes that might be used for gene drive systems.

Champer, J., Buchman, A., & Akbari, O. S. (2016). Cheating evolution: engineering gene drives to manipulate the fate of wild populations. *Nature Reviews Genetics, 17*(3), 146-59.

This review discusses the general characteristics of gene drive systems, the current state of the molecular genetic technology to create gene drive systems, and the benefits and drawbacks of the various types of gene drive systems. It also discusses the potential safety issues surrounding the application of gene drive and some of the precautions that can be taken during their application.

Burt, A. (2014). Heritable strategies for controlling insect vectors of disease. *Philosophical Transactions of the Royal Society B-Biological Sciences, 369*(1645), 20130432.

This paper contains a discussion of the application of gene drive systems to control insect disease vectors, with a focus on mosquitoes. It reviews the types of heritable gene drive systems that are being developed in mosquitoes, differentiating between drive systems that are self-limiting versus self-sustaining. Implementation strategies are also discussed, including considerations for the evolution of host resistance and the steps for an effective rollout of a gene drive system.

Esvelt, K. M., Smidler, A. L, Catteruccia, F., & Church, G. M. (2014). Concerning RNA-guided gene drives for the alteration of wild populations. *Elife, 3*, e03401.

This paper reviews RNA-guided gene drive systems with a focus on CRISPR/Cas9-mediated genome engineering. Technical features and challenges of CRISPR/Cas9 systems are discussed in detail. The paper concludes with a description of the potential capabilities, limitations, safeguards, and applications of RNA-guided gene drive systems.

DiCarlo, J. E., Chavez, A., Dietz, S. L., Esvelt, K. M., & Church, G. M. (2015). Safeguarding CRISPR-Cas9 gene drives in yeast. *Nature Biotechnology, 33*(12), 1250-1255.

In this study, the authors construct multiple CRISPR-Cas9 drive systems in the yeast *Saccharomyces cerevisiae*, creating the first synthetic RNA-guided gene drive system in a microorganism. In addition, they develop methods of molecular confinement that prevent the gene drive from escaping from laboratory to natural populations, and they create a gene drive system that is capable of overwriting genetic changes introduced by an earlier gene drive system.

Gantz, V. M. & Bier, E. (2015). The mutagenic chain reaction: A method for converting heterozygous to homozygous mutations. *Science, 348*(6233), 442-44.

This paper presents the first use of RNA-guided endonucleases, specifically CRISPR/Cas9, to create a synthetic gene drive system. This proof of principle study uses the fly *Drosophila melanogaster* to produce loss of function mutations. By creating a mutagenic chain reaction, a mutation of interest can spread efficiently from the chromosome of origin to most of the somatic and germline cells in the progeny.

Gantz, V. M., Jasinskiene, N., Tatarenkova, O., Fazekas, A., Macias, V. M., Bier, E, & James, A. A. (2015). Highly efficient Cas9-mediated gene drive for population modification of the malaria vector mosquito *Anopheles stephensi*. *Proceedings of the National Academy of Sciences of the United States of America, 112*(49), E6736-E6743.

This study is the first application of an RNA guided gene drive system to a major disease vector, the mosquito *Anopheles stephensi*. The authors use CRISPR-Cas9 to drive in a gene that confers resistance to the malaria parasite *Plasmodium falciparum*. Furthermore, they show that the antipathogen genes that are inserted are transcriptionally active following Cas9 mediated movement. The results suggest these methods may be effective for malaria control and eradication.

Hammond, A., Galizi, R., Kyrou, K., Simoni, A., Siniscalchi, C., Katsanos, D., Gribble, M., Baker, D., Marois, E., Russell, S., Burt, A., Windbichler, N., Crisanti, A., & Nolan, T. (2016). A CRISPR-Cas9 gene drive system-targeting female reproduction in the malaria mosquito vector *Anopheles gambiae. Nature Biotechnology, 34*(1), 78-83.

This paper is the first description of a CRISPR-Cas9 gene drive system in the mosquito *Anopheles gambiae*. The authors disrupt somatically expressed loci necessary for female fertility and show that the constructs drive in the germline of both sexes. Furthermore,

they use cage populations to show that these disrupted alleles can spread over successive generations.

Committee on Gene Drive Research in Non-Human Organisms, National Academies of Sciences, Engineering, and Medicine. (2016). "Gene drives on the horizon: Advancing science, navigating uncertainty, and aligning research with public values." National Academies Press. PDFs of the full report (https://www.nap.edu/read/23405/chapter/1) and of the summary (http://nas-sites.org/gene-drives/files/2015/08/Gene-Drives-Brief06.pdf) are freely available.

In this report, a committee of experts stresses that gene drive holds great promise, but before broad scale implementation they call for more research on the population dynamics of gene drive systems and more thorough ecological risk assessments, as well as phased testing so that care can be taken to guard against unintended consequences.

CHAPTER 3

OCCASIONAL RECOMBINATION OF A SELFISH X-CHROMOSOME MAY PERMIT ITS PERSISTENCE AT HIGH FREQUENCIES IN THE WILD[1]

[1]Pieper, K. E., & Dyer, K. A. (2016). *Journal of Evolutionary Biology*. *29*(11), 2229-2241. DOI: 10.1111/jeb.12948
Reprinted here with permission of the publisher.

Abstract

The *sex-ratio* X-chromosome (SR) is a selfish chromosome that promotes its own transmission to the next generation by destroying Y-bearing sperm in the testes of carrier males. In some natural populations of the fly *Drosophila neotestacea,* up to 30% of the X-chromosomes are SR chromosomes. To investigate the molecular evolutionary history and consequences of SR, we sequenced SR and standard (ST) males at 11 X-linked loci that span the ST X-chromosome and at seven arbitrarily chosen autosomal loci from a sample of *D. neotestacea* males from throughout the species range. We found that the evolutionary relationship between ST and SR varies among individual markers, but genetic differentiation between SR and ST is chromosome-wide and likely due to large chromosomal inversions that suppress recombination. However, SR does not consist of a single multi-locus haplotype: we find evidence for gene flow between ST and SR at every locus assayed. Furthermore, we do not find long-distance linkage disequilibrium within SR chromosomes, suggesting that recombination occurs in females homozygous for SR. Finally, polymorphism on SR is reduced compared to ST, and loci displaying signatures of selection on ST do not show similar patterns on SR. Thus, even if selection is less effective on SR, our results suggest that gene flow with ST and recombination between SR chromosomes may prevent the accumulation of deleterious mutations and allow its long-term persistence at relatively high frequencies.

Keywords: genetic conflict, *sex-ratio* drive, meiotic drive, molecular evolution, sex chromosome evolution, *Drosophila neotestacea*

Introduction

Genetic conflict occurs when one part of the genome promotes its own transmission to the detriment of the rest of the genome. Conflict is ubiquitous across the tree of life, and selfish genetic elements can range in size from gene-sized transposable elements to entire chromosomes (Burt & Trivers, 2006; Rice, 2013). Selection against selfish activity is hypothesized to have contributed to the evolution of fundamental processes and structures such as meiotic recombination and centromeres (Werren, 2011; Rice, 2013). Meiotic drivers are selfish elements that cause genetic conflict by manipulating gametogenesis to favor their own transmission to the next generation (Lindholm *et al.*, 2016). One example is *sex-ratio* (SR) drive, which occurs when an X-chromosome kills Y-bearing sperm in carrier males, causing them to sire only daughters that also carry SR (Jaenike, 2001). X-chromosomes may be predisposed to evolve selfish behavior because of the inherent conflict among the autosomes, X-chromosome, and Y-chromosome over the optimal proportion of male and female offspring (Hurst & Pomiankowski, 1991; Jaenike, 2001; Meiklejohn & Tao, 2010). The resulting bias in the offspring sex ratio may also alter the population-level sex ratio, which can favor the evolution of suppressors of drive located in other parts of the genome (Hamilton, 1967; Hall, 2004; Meiklejohn & Tao, 2010).

If the selfish genetic system escapes suppression and restores drive, but now requires two or more components on the driving chromosome to act, selection for genetic linkage may favor suppressed recombination between them (Prout *et al.*, 1973; Charlesworth & Hartl, 1978). Empirically, most drive systems are found in heterochromatic regions or are located within chromosomal inversions, and thus

experience reduced recombination between the driving and wild-type, or Standard (ST), arrangements (Jaenike, 2001; Lindholm *et al.*, 2016). This suppressed recombination can have important consequences for patterns of molecular evolution (Navarro & Barton, 2003; Hoffmann & Rieseberg, 2008). For instance, in addition to high divergence between inverted and noninverted arrangements, selection on variants within an inversion can lead to low polymorphism due to Hill-Robertson effects (e.g. Andolfatto *et al.*, 2001; Cheng *et al.*, 2012; Fabian *et al.*, 2012). The increase in frequency of the driving haplotype due to the selfish behavior of the driving loci can also leave a signature similar to that caused by an increase in frequency due to positive natural selection. For example, on the X-chromosome in *Drosophila mauritiana,* striking polymorphism patterns congruent with strong selective sweeps were detected around loci involved in SR drive in the closely related species *D. simulans* (Nolte *et al.*, 2013).

SR chromosomes are common in Diptera and have evolved independently many times (Jaenike, 2001). Neither of the two well-studied SR systems in *D. simulans* is associated with inversions, but both show evidence of selective-sweep like patterns of low polymorphism at the causal driving loci and at very closely linked loci (Derome *et al.*, 2008; Kingan *et al.*, 2010; Helleu *et al.*, 2016). The SR chromosome in *D. pseudoobscura* is associated with several inversions, but in a study of five X-linked loci, a reduction in polymorphism on SR was only observed at one locus (Babcock & Anderson, 1996; Kovacevic & Schaeffer, 2000). This could be due to the small size of the inversion, the location of this locus near the breakpoint of an inversion, or proximity to a causal driving locus (Kovacevic & Schaeffer, 2000). At the other extreme, the SR chromosome of *D. recens* is completely bound in overlapping inversions and shows

extremely high interlocus linkage disequilibrium (LD) and a dramatic reduction in polymorphism at all sampled loci compared to the standard X-chromosome (Dyer *et al.*, 2007). Similar patterns are observed in the SR chromosome of the stalk-eyed fly, *Teleopsis dalmanni* (Christianson *et al.*, 2011). In both of these cases, the SR inversions carry pleiotropic deleterious alleles; they cause female sterility when homozygous in *D. recens* and decreased eyestalk size in *T. dalmanni* (Dyer *et al.*, 2007; Cotton *et al.*, 2014). Overall, it appears that the rapid increase in frequency of the SR chromosome and restricted recombination with ST tend to result in decreased polymorphism on SR chromosomes. Because SR systems have arisen many times and thus have different evolutionary histories and characteristics, we can compare them to understand how a SR system's age, chromosomal inversions, and pleiotropic effects alter the molecular evolutionary consequences it has for the X-chromosome.

In this study we investigate SR drive in *Drosophila neotestacea*, which is a North American boreal and temperate forest mushroom-feeding fruit fly that harbors an SR chromosome (James & Jaenike, 1990). In some populations of *D. neotestacea*, SR is found at frequencies as high as 30%, which appear to be stable on the order of decades (James & Jaenike, 1990; Dyer, 2012; Pinzone & Dyer, 2013). In *D. neotestacea,* there is also no evidence of active SR suppressors, and SR homozygote females are fully fertile (Dyer, 2012). Previous genetic work is limited to microsatellite repeat markers, which showed strong genetic differentiation between SR and ST suggesting the presence of one or more inversions on SR (Dyer, 2012; Dyer *et al.*, 2013). This work also identified an excess of microsatellite polymorphism on SR that could be due to occasional recombination or gene conversion between SR and ST in heterozygote females (Dyer,

2012; Dyer *et al.*, 2013). High, stable frequencies of SR in some populations, a lack of phenotypic suppressors, and highly differentiated inverted regions potentially protected from mutation accumulation by occasional recombination suggest that SR may have had a large effect on X-chromosome evolution in this species.

We examine the molecular evolutionary consequences of SR and determine the evolutionary relationship between ST and SR. Using a sample of *D. neotestacea* males from throughout the species range, we sequenced SR and ST males at 11 randomly chosen X-linked loci that span the ST X-chromosome. For comparison we also sequenced a subset of these individuals at seven randomly chosen autosomal loci. We analyzed patterns of genetic differentiation between the ST and SR chromosomes, performed a phylogenetic analysis to infer the evolutionary history of SR, and compared levels of recombination, linkage disequilibrium, and nucleotide polymorphism between ST and SR. We confirmed the presence of inversions and decreased recombination between ST and SR. Our results indicate that while SR and ST are strongly differentiated and selection may be acting differently on SR, SR is not segregating as a single extended haplotype. We suggest that SR may be able to persist long-term at relatively high frequencies due to occasional gene flow with ST and among SR chromosomes, which prevents SR from degrading via mutation accumulation.

Methods

Sampling and DNA sequencing

X-chromosomes were sampled from 14 geographic populations that span the *D. neotestacea* species range (Table S1, See Table 1 of Dyer, 2012). These males are the

same as used in Dyer (2012), in which wild-caught males were identified as carrying ST or SR X-chromosomes by the proportion of female offspring they produced. From each population where SR is present, at least three ST and three SR males were sequenced at each of 11 randomly chosen X-linked loci, which included six protein-coding genes (*marf, mof, pgd, rpl, spk,* and *sxl*), and the flanking regions of six microsatellite loci (*neo5261*, *neo6002*, *neo7029*, *neo7040*, *neo8377*, and *neo8385*)(Table S2)(Dyer, 2007). In total, each male was sequenced at 4,510 base pairs on the X-chromosome. In populations without SR present, only ST males were sampled. At least 53 ST and 41 SR males were included for every marker, though the total sample number is variable for each individual locus. From each population, at least three males were chosen randomly with respect to their X-chromosome type using a random number generator (random.org). These males were sequenced at seven arbitrary autosomal protein coding loci located on all of the other Muller elements except F, which included *esc, gl, mago, ntid, sia, tpi,* and *wee* for a total of 2,657 bp per individual (Table S2). All these loci were also sequenced in one or two individuals of *D. orientacea* and *D. putrida*, which are two of the other members of the testacea species group. From the final member of the testacea species group, *D. testacea,* one female from each of 24 isofemale lines collected in Munich, Germany were chosen for sequencing.

DNA extractions were performed with Qiagen Puregene Core Kit A. Fragments were amplified using standard PCR protocols (Table S3) and sequenced on an Applied Biosystems 3730xl DNA Analyzer at the Georgia Genomics Facility. Base calls were confirmed using Geneious (Kearse *et al.*, 2012), and heterozygous SNPs in diploid loci were phased using PHASE (Scheet & Stephens, 2006). Sequences were aligned by hand

in Geneious, and microsatellite repeats were removed manually, leaving only the flanking regions for use in analyses (Kearse *et al.*, 2012). Open reading frames were assigned using annotated *D. melanogaster* orthologs as a guide (flybase.org).

Recombination mapping

We used recombination mapping to determine the order of the loci on the ST chromosome. We performed single pair crosses with flies from isofemale lines originally collected in Seattle, WA (Sea) and Coeur d'Alene, ID (ID-1). Sea females were crossed to ID-1 males, and the F1 females collected and crossed to males from an inbred lab stock. The F2 males (carrying a maternally-derived X-chromosome) were collected and frozen for genotyping. All flies were reared on Instant Drosophila Medium (Carolina Biological Supply, Burlington, NC) supplemented with commercial mushroom (*Agaricus bisporus*) at 20ºC with 60% relative humidity on a 12-hour light/dark cycle. DNA from the parents and 92 F2 males was extracted as described above. Parents were genotyped using repeat number at X-linked microsatellite markers (as described in Pinzone & Dyer, 2013), and sequenced at X-linked protein coding loci to identify SNPs at restriction enzyme cut sites. *BsaWI* (New England Biolabs, Ipswich, MA) was used for restriction fragment analysis of *spk.* F2 males were genotyped at the microsatellite loci and *spk*, and were sequenced at *marf* and *pgd* to identify which parental allele they carried. The remaining two loci (*rpl* and *mof*) contained no polymorphisms in the parental cross and were thus unable to be mapped. The most likely map order and distances were calculated using the Kosambi mapping function in MapDisto with 1000 bootstrap replicates (Kosambi, 1943; Lorieux, 2012).

As SR/SR females are fertile we attempted to use two independently collected and maintained SR lines for recombination mapping of the SR chromosome. These lines were collected in Eugene, OR in 2001 (SR-Par) and in Rochester New York in 1990 (SR-Lab). However, based on sequencing each locus from each line there was not enough variation between them to determine anything except that *neo6002* and *neo7029* are at least 50 cM apart on SR, the same as on ST.

The ST lab stock was also originally collected in New York in 1990, and the SR-Lab stock is maintained in the same genetic background. Every generation, ST/Y males are crossed to SR/SR females to generate SR/Y males, and SR/Y males are also crossed to SR/SR females to produce more SR/SR females. We tested for recombination between ST and SR by crossing ST/SR heterozygous females with ST/Y males, and then genotyping 95 male offspring at the six X-linked microsatellites described above. Parental ST and SR genotypes were identified using four SR/SR females and two ST/Y males from these highly inbred SR and ST stocks.

Phylogenetic analysis

Multi-locus phylogenetic species trees were constructed using *BEAST with the HKY nucleotide substitution model and a chain length of 100 million, with 20% burn-in removed (Heled & Drummond, 2010; Bouckaert *et al.*, 2014). Orthologs from *D. testacea, D. orientacea,* and *D. putrida* were included in the analysis, and based on previous phylogenetic work *D. putrida* was used as an outgroup to root the trees (Perlman & Jaenike, 2003; Dyer *et al.*, 2011). Trees for X-chromosome and autosomal loci were constructed separately. For the X-linked species tree, *D. neotestacea* samples

were divided into ST and SR. The X-linked tree also included the marker *sxl*, which was dropped from further analyses due to extremely low polymorphism. The autosomal tree used all seven autosomal markers. The individual gene trees generated as a part of these *BEAST analyses were used to infer the relationship between *D. neotestacea* ST and SR and *D. testacea* at the individual locus level.

Relationships between individual samples were inferred using a multi-locus phylogenetic tree constructed with all the X-linked makers except *neo5261,* which lacked a sequence from *D. putrida* used to root the tree. This tree was also built in *BEAST using the HKY substitution model and a chain length of 2 billion with 20% burn-in removed.

Patterns of genetic differentiation and recombination

To estimate patterns of genetic differentiation we calculated K_{ST} and S_{nn} (Hudson *et al.*, 1992; Hudson, 2000) in DnaSP (Librado & Rozas, 2009) using both geographic sampling location and ST and SR as groupings (Librado & Rozas, 2009). Significance of individual K_{ST} and S_{nn} values was determined using 1,000 random permutations according to the method of (Hudson *et al.*, 1992; Librado & Rozas, 2009). Statistical analyses were carried out in RStudio (2014).

To estimate patterns of linkage disequilibrium (LD) within and across loci, all X-linked markers were concatenated in the order of the ST genetic map and the pairwise correlation coefficient (R^2) between each pair of parsimony informative sites was calculated according to the method of Hill & Robertson (1968). Significance of each association was calculated in DnaSP using a Fisher's Exact Test with a Bonferroni

correction for multiple testing using $\alpha = 0.05$ (Librado & Rozas, 2009). LD was inferred using all samples (SR and ST) and separately for SR and ST chromosomes. ZnS for each locus was calculated according to the method of Kelly (1997). The population recombination rate (ρ) was estimated for each locus using the composite-likelihood method of Hudson (2001) in the program LDHat (McVean *et al.*, 2004). For the autosomes, $\rho = (1/2)*4N_e r = 2N_e r$ (as males do not recombine in *D. neotesteacea*), where N_e is the effective population size and r is the per-generation recombination rate. For the X-chromosome, $\rho = (2/3)*3N_e r = 2N_e r$. ρ was calculated for ST and SR separately for each marker, and then scaled by the number of sites. Genetic exchange between SR and ST was detected in the concatenated alignment of all X-linked markers using the method of Betran *et al.* (1997) as implemented in DnaSP (Librado & Rozas, 2009).

Patterns of nucleotide polymorphism

For each marker, nucleotide polymorphism was analyzed separately for ST and SR sample groups. In markers that contained open reading frames, segregating sites were split into silent site polymorphisms (synonymous changes and changes outside the open reading frame) and nonsynonymous polymorphisms. Microsatellite flanking regions were assumed to be silent sites. To evaluate patterns of nucleotide polymorphism, average pairwise nucleotide differences (π) (Nei, 1987), Watterson's θ (Watterson, 1975), and Tajima's D (Tajima, 1989) were calculated in DnaSP using only silent sites (Librado & Rozas, 2009). π was also calculated using only nonsynonymous variation for the protein coding genes. Net nucleotide substitutions per site (D_a) between ST and SR was calculated for a combined set of all X-linked markers (Nei, 1987; Librado & Rozas,

2009). Parsimony informative sites were identified for ST and SR individually in DnaSP and used to identify private alleles. Expected Tajima's D values were generated using 10,000 coalescent simulations in the program HKA (https://bio.cst.temple.edu/~hey/software/software.htm#HKA). K_a/K_s and π_a/π_s were calculated for all protein coding loci in DnaSP (Librado & Rozas, 2009), with orthologs from *D. putrida* used to identify substitutions.

Polytene chromosome squashes

Inversions on SR were confirmed using squashes of polytene chromosomes from the salivary glands of *D. neotestacea* third instar larvae. Salivary glands were dissected out in phosphate-buffered saline, fixed in 45% acetic acid, stained with orcein, and then physically squashed on a microscope slide to separate the chromosomes (Sullivan *et al.*, 2000). Chromosomes spreads were examined under 400X magnification using phase contrast. The X-chromosome was identified in ST males from the inbred lab stock because it showed no synapses or chromosomal inversions, and then ST/SR heterozygote females were generated and used to identify inversions between ST and SR.

Results

Genetic map

Recombination mapping of the ST X-chromosome revealed that the nine X-linked loci span a genetic distance of 113 cM (Figure 3.1). This distance likely encompasses the majority of the X-chromosome given the typical size of Muller element A of other species in the subgenus Drosophila (e.g. Gubenko & Evgenev, 1984; Schafer *et al.*, 1993;

Staten *et al.*, 2004; Dyer *et al.*, 2007). However, it is probable that the map order of the markers is rearranged on SR relative to ST. Not a single recombination event was detected between SR and ST chromosomes in the lab crosses; all 95 genotyped sons of ST/SR females inherited either the SR or ST haplotype at all six microsatellite loci.

Differentiation among populations and between sex-ratio carrier status

We find little geographic differentiation among populations, similar to previous studies of *D. neotestacea* (Dyer, 2012; Dyer *et al.*, 2013). K_{ST} and S_{nn} based on geographic origin shows that average differentiation between geographic populations is low for both the autosomal markers (K_{ST} = 0.030, sd = 0.323) and the X-linked markers (0.020, sd = 0.037)(Figure 3.2a, Table S4). There is no significant difference in geographic-based population differentiation between the autosomes and the X-chromosome (t_{14} = 0.617, p = 0.547, two-tailed t-test). Therefore, for the remainder of the analyses we pool geographic samples together and focus on comparisons between SR and ST chromosomes.

There is substantial genetic differentiation between the SR and ST X-chromosomes at every locus we surveyed (Table S4), indicating reduced gene flow between SR and ST on a chromosome-wide scale. The average K_{ST} between ST and SR for the protein coding genes was 0.1044 (sd = 0.1027) and for the microsatellite flanking regions was 0.3314 (sd = 0.1047; Figure 3.2b). In contrast, the average K_{ST} of autosomal markers between ST and SR males is extremely low (mean = 0.0005, sd = 0.0034; Figure 3.2b, Table S4), as expected given independent assortment of the autosomes and X-chromosome. Hudson's S_{nn} supports the results of the K_{ST} analysis (Table S4). An

analysis of variance (ANOVA) comparing K_{ST} of the autosomal loci and the X-linked loci with locus type (microsatellite flanking region or protein coding gene) nested within X-chromosome (ST or SR) showed a significant effect of both chromosome and locus type, with the X-chromosome having significantly higher K_{ST} than the autosomes ($F_{1,15}$ = 34.3, $p < 0.0001$) and microsatellite flanking regions on the X-chromosome having significantly higher K_{ST} than the protein coding genes ($F_{1,15} = 21.72$, $p < 0.0001$) (Figure 3.2b).

The differentiation between ST and SR can be seen in Figure 3.3, which depicts the X-linked haplotype of every individual sample at all parsimony informative segregating sites. With the exception of one locus (*neo8385*), all of the X-linked loci have segregating polymorphisms that are unique to ST or SR (Table S4). At eight of the eleven markers there are shared polymorphisms, and there are no fixed differences between ST and SR in any of the markers sampled (Table S4).

Phylogenetic analysis

The genetic differentiation between ST and SR is also apparent in the multi-locus phylogeny of individual samples (Figure 3.4), where the only clade with any substantial support is the one containing all of the SR samples. It is likely that the very low branch support in the rest of the tree is due to the high amount of free recombination among ST chromosomes. Consistent with this, the species tree made from the X-linked markers suggests the origin of SR occurred after the split with *D. testacea* (Figure S1).

The individual gene trees of the X-linked loci show such a much more variable relationship between ST and SR (Figure S2). Many markers include large polytomies that

involve both ST and SR samples, and many which samples cluster with ST or SR varies from locus to locus (Figure S2). For no single marker is the association between SR status and the SR clade perfect; there are always SR individuals that cluster with ST, or vice versa. At a few markers, including *marf* and *neo8377,* SR is basal to ST (Figure S2). Consistent with the K_{ST} analyses, in the individual gene trees there is no phylogenetic sorting of ST and SR sequences at autosomal loci and no evidence of geographic clustering of samples at any locus (Figure S2).

At a broader taxonomic level, the individual gene trees also show that the relationships among *D. testacea, D. orientacea,* and *D. neotestacea* are variable across markers. This is likely due to incomplete lineage sorting and/or hybridization among the testacea group species. We also note that while the species tree made from the X-linked markers places *D. testacea* sister to *D. neotestacea*, the species tree made from the autosomal markers places *D. orientacea* as most closely related to *D. neotestacea* (Figure S1).

Recombination and linkage disequilibrium

Compared to ST, the population recombination rate (ρ) for each marker on SR is reduced by a degree proportional to its average population frequency. A two-way ANOVA was used to test for significant differences on the X-chromosome, with locus type (silent sites in protein coding genes and microsatellite flanking regions) nested within X-chromosome type (ST or SR). ρ is significantly lower on SR (mean = 0.01, sd = 0.02) than on ST (mean = 0.17, sd = 0.14) ($F_{1,16} = 11.703$, $p < 0.004$), with no effect of marker type ($F_{2,16} = 0.883$, $p = 0.43$). Figure 3.5 depicts the ratio of ρ/(number of sites)

between the X-chromosome and the autosomes; assuming an equal population sex ratio, the expectation for this ratio is 1. We scale this by the average frequency of ST and SR across all populations (0.85 and 0.15, respectively (Dyer, 2012)) to obtain an expectation for the estimate of ρ at each marker on ST and SR. On ST, there was no significant difference from the expectation for either the microsatellite flanking regions (mean = 1.28, sd = 0.97, $t_5 = 1.095$, p = 0.32, two-tailed t-test) or the protein coding genes (mean = 0.77, sd = 0.67, $t_4 = -0.276$, p = 0.80, two-tailed t-test). There was also no significant difference from the expectation on SR for either the microsatellites (mean = 0.07, sd = 0.13, $t_4 = -1.328$, p = 0.25, two-tailed t-test) or the protein-coding genes (mean = 0.05, sd = 0.07, $t_3 = -3.00$, p = 0.06, two-tailed t-test).

Long range LD across the entire X-chromosome is observed between loci when considering SR and ST chromosomes together (Figure 3.1). An R^2 analysis identified 471 statistically significant SNP associations across all X-linked markers from a total of 14,535 comparisons between 171 parsimony informative sites. Associations within a locus were significantly overrepresented when considering their frequency (202 significant associations; 3,710 non-significant) compared to associations between SNPs in different loci (269 significant associations; 10,354 non-significant) ($\chi^2 = 62.30$, p < 0.0001) (Figure 3.1). While most of the LD is within loci, we do identify substantial long-range LD, even between sites that are over 100 cM apart on the ST chromosome. In contrast, within each type of chromosome (SR or ST) nearly all of the significant pairs of associations are within rather than between loci, indicating the presence of recombination within each type (Figure S3). When considering only ST samples, 11,781 comparisons were made between 154 polymorphism informative sites, and 28 significant

associations were identified, all of which were between SNPs located in the same marker. Despite the smaller number of sites considered for only the SR samples (2,080 comparisons between 65 parsimony informative sites), there were more significant SNP associations than on ST (186), and only three of these associations occurred between sites in different markers.

Evidence of gene flow between chromosome types is the detection of gene conversion tracts between ST and SR in both directions and in multiple individuals (Table S5, Figure 3.3). Tracts that begin and end in the same marker likely represent true instances of gene conversion, which typically only involve short stretches of nucleotides. However, many other algorithmically detected tracts crossed multiple markers in the concatenated alignment of all X-linked markers, with the longest tract covering 3,296 of the 4,510 bases in the alignment of concatenated X-linked markers. This tract had an ST donor and an SR recipient, and covered a total of eight markers, or roughly 100cM based on the ST map. Rather than a true gene conversion event, this very large tract and others like it represent ST-typical sequences found in a SR phenotype individual, at least at the endpoints of the identified tract (Figure 3.3). These are likely the result of gene flow between ST and SR due to other mechanisms such as double crossovers. Very long tracts may also represent two separate gene flow events that have been linked together due to a lack of informative sites between them. Out of the 33 total gene conversion events detected, 15 were from ST to SR, and 18 were from SR to ST (Fisher's exact test, $p = 0.8$). However, the ST to SR conversion events are much larger generally, having an average tract length of 1571.4 bp (sd = 1096.07), significantly more than the average tract length of SR to ST conversion events (mean = 183.17, sd = 255.11) ($t_{15} = 4.473$, $p <$

0.001, two-tailed t-test). Of course, these lengths are only within the alignment of the sequence data used in this study, and the actual physical distance between the start and the end of the tract may be much larger. The gene order on SR is also likely highly rearranged relative to ST, so it is possible that loci that are very far apart on ST are actually much closer on SR.

Nucleotide polymorphism

At the autosomal markers, there is no significant difference between silent nucleotide diversity (π_s) in flies that carried an ST (mean = 0.0245, sd = 0.0167) or SR chromosome (mean = 0.0222, sd = 0.0132; $t_6 = 0.744$, p = 0.485, paired two-tailed t-test). This is expected given independent assortment of the autosomes and the X-chromosome during meiosis. Therefore all autosomal samples were grouped together and used as a single comparison for the X-linked markers. A set of general summary statistics for each locus can be found in Table S2.

Silent-site polymorphism on SR is less than half of what is found on ST X-chromosomes: on SR, the mean π_s was 0.005 (sd = 0.003), and on ST it was 0.013 (sd = 0.012) (Figure S4). A two-way ANOVA with marker type (silent sites in protein coding genes and flanking regions of microsatellites) nested within X-chromosome type (ST or SR) was used to compare π_s for the X-linked markers. There was a main effect of X-chromosome type, with SR having a significantly lower π_s than the ST ($F_{1,18} = 5.973$, p = 0.025). There was also an effect of locus type, with the microsatellite markers having a significantly lower π than the silent sites of protein coding regions ($F_{2,18} = 4.000$, p = 0.037).

Nonsynonymous polymorphism (π_a) was very low for every marker (Table S2), and a one-way ANOVA found no significant difference in π_a between ST (mean = 0.001, sd = 0.002), SR (mean = 0.0005, sd = 0.0006), and the autosomes (mean = 0.001, sd = 0.001) ($F_{2,14} = 0.67$, p = 0.528). There was no significant difference between π_a/π_s values on ST (mean = 0.062, sd = 0.079) and SR (mean = 0.141, sd = 0.246) ($t_4 = -1.049$, p = 0.35, two-tailed t-test), or between the X-chromosome (mean = 0.102, sd = 0.178) and the autosomes (mean = 0.057, sd = 0.050) ($t_{10} = -0.753$, p = 0.47, two-tailed t-test) (Table S2). There was also no significant different between K_a/K_s values on ST (mean = 0.116, sd = 0.175) and SR (mean = 0.087, sd = 0.113) ($t_4 = 1.009$, p = 0.37, two-tailed t-test), or between the X-chromosome (mean = 0.102, sd = 0.139) and the autosomes (mean = 0.108, sd = 0.178) ($t_{10} = 0.084$, p = 0.93, two-tailed t-test) (Table S2a).

The ratio of the effective population size (N_e) of X-chromosomes to autosomes in a population with an even sex ratio is expected to be 0.75, and thus the ratio of neutral polymorphism on the X-chromosome and the autosomes should also be 0.75. However, since a substantial proportion of X-chromosomes in this species are SR (15% species wide (Dyer, 2012)), the expected X/A polymorphism ratio for ST is (0.75*0.85) = 0.64, assuming SR frequency is at a long-term equilibrium. Silent polymorphism in protein-coding genes on ST trends higher than this expectation, but does not differ significantly (mean X/A ratio = 0.878, sd =0.641, $t_4 = 0.837$, p = 0.45 two-tailed t-test) (Figure 3.6). Interestingly, the flanking regions of microsatellites on the ST X-chromosome have a significantly lower average level of polymorphism than expected (mean X/A ratio = 0.309, sd = 0.171, $t_5 = -4.696$, p = 0.005, two tailed t-test) (Figure 3.6). The expected amount of polymorphism on SR compared to the autosomes relative to its frequency is

(0.15*0.75) = 0.1125, again assuming that SR frequency is at a long-term equilibrium. In contrast to ST, the observed amount of silent polymorphism at both protein-coding genes and microsatellite loci on SR is higher than this expected value but not significantly different from it (mean X/A ratio: protein coding mean = 0.246, sd = 0.191, $t_4 = 1.564$, p = 0.19, two tailed t-test; microsatellite flanking mean = 0.199, sd = 0.106, $t_5 = 1.991$, p = 0.103 two tailed t-test) (Figure 3.6).

Silent sites at autosomal loci have a generally negative Tajima's D (mean = -1.74, sd = 0.407; Figure S4c and Table S2), with the observed D significantly lower than the neutral expectation for six of the seven autosomal loci ($p < 0.05$ for each). This suggests that this skew in the frequency spectrum is a demographic effect and these values can be taken as the genomic background. The average D is not significantly different between the autosomes and all sequences on the X-chromosome (mean = -1.617, sd = 0.597, t_{15} = -0.614, p = 0.55, two-tailed t-test) (Figure 3.7). On the X-chromosome, a two-way ANOVA with locus type (silent sites in protein coding genes and microsatellite flanking regions) nested within X-chromosome type (ST and SR) was used to determine significant differences between Tajima's D values. There was there was no effect of maker type ($F_{2,18} = 1.079$, p = 0.36), but there was a significant effect of X-chromosome type, with markers on ST having a lower D than on SR ($F_{1,18} = 5.99$, $p < 0.03$) (Figure 3.7). However, this effect appears to be driven primarily by the microsatellite flanking regions, which have a very low D on ST (mean = -2.106, sd = 0.26) compared to SR (mean = -1.37, sd = 0.69). This trend can be observed when comparing D on an individual locus basis (Figure S4c, Table S2a).

The X-chromosome in the sister species *D. testacea* harbors a similar level of polymorphism as the ST X-chromosome in *D. neotestacea* (mean $\pi = 0.008$, sd = 0.005, t_{13} -1.238, p = 0.238, two-tailed t-test)(Table S2b). However, polymorphism in *D. testacea* at the X-linked microsatellites is significantly higher than the expected value of 0.75 of the diversity on the autosomes (mean $\pi = 0.011$, sd = 0.005, $t_5 = 2.667$, p = 0.045, two-tailed t-test), but matches that expectation at the X-linked protein coding loci (mean = 0.005, sd = 0.002, $t_5 = 0.128$, p = 0.903, two-tailed t-test)(Table S2b). Overall, Tajima's D values on the X-chromosome and autosomes in *D. testacea* are much closer to zero than in *D. neotestacea*, with only one locus being significantly different from the simulated expectation (Table S2b), and with no significant differences between the average D values of the X-linked microsatellites (mean = -1.282, sd = 0.631), the X-linked protein coding genes (mean = 1.501, sd = 0.278), and the autosomal loci (mean = -1.260, sd = 0.555)($F_{2,15} = 0.349$, p = 0.711, three-way ANOVA)(Table S2b).

Confirmation of SR inversions

We confirmed chromosomal inversions between ST and SR by examining polytene chromosome squashes from ST/SR larvae (Figure 3.8). *D. neotestacea* has five visible chromosomes: the X-chromosome, one double-length telocentric autosome, two other autosomes, and the dot chromosome (Patterson & Stone, 1952). The X-chromosome was identified in ST males by its lack of inversions compared to the two similarly sized autosomes. Based on banding patterns, we believe that the X-chromosome described in James & Jaenike (1990) is actually one of these other autosomes. We then identified the X-chromosome in ST/SR females using shape and banding pattern. Several

large inversions are clearly visible (Figure 3.8). Despite the use of an inbred lab stock for this analysis, chromosomal inversions occurred on every chromosome except the dot.

Discussion

The effects of genetic conflict encompass the entire SR chromosome of *D. neotestacea* and affect all aspects of its evolution. Our findings of low genetic differentiation among geographic populations and high differentiation between ST and SR are consistent with previous work in this system that analyzed the repeat motifs of microsatellites (Figure 3.2) (Dyer, 2012; Dyer *et al.*, 2013). We also used laboratory crosses and patterns of linkage disequilibrium to show that recombination is restricted between ST and SR along the entire length of the > 110 cM X-chromosome. These results suggest that most or all of SR is tied up in inversions, and we visually confirmed the presence of multiple, large rearrangements on SR. However, the amount of divergence between SR and ST varies from locus to locus, with no individual locus showing a perfect association with SR and no fixed differences between the two types of chromosomes (Figure S2, Table S4). ST individuals with SR haplotypes and vice versa can be seen within individual loci (Figure 3.3). This suggests ongoing gene flow in the form of gene conversion events or double crossovers, as well as incomplete lineage sorting of variation between ST and SR.

Such gene flow results in an imperfect association with SR over long genetic distances, so markers that are tightly linked to the driving locus can be expected to show the strongest differentiation between SR and ST. For instance, in the Winters and Paris SR systems of *D. simulans* only loci near the driver show clear evidence of selective

sweeps (Derome *et al.*, 2008; Kingan *et al.*, 2010; Helleu *et al.*, 2016). Likewise, decreased polymorphism is only found within inversions on the SR chromosome of *D. pseudoobscura* (Kovacevic & Schaeffer, 2000). The location of the driving loci is unknown in *D. neotestacea*, but previous work found microsatellite repeat genotypes at *neo8385* and *neo8377* that predict the SR phenotype at 94% accuracy (Dyer *et al.*, 2013; Pinzone & Dyer, 2013). In our study these loci also have the highest values of K_{ST} between ST and SR (Table S4), and their gene trees show well-supported SR clades (Figure S2). As these markers are physically located only 3.5 cM from each other on the ST chromosome (Figure 3.1), of the loci we surveyed they may be physically closest to the driver(s). So while the presence of inversions precludes mapping of the SR gene(s), a history of occasional gene flow between SR and ST suggests that other approaches such as association mapping may identify loci or regions involved in the drive mechanism, because an SR causal locus would be perfectly associated with SR expression.

The lack of a single SR-associated haplotype in *D. neotestacea* is in contrast to the SR systems in *D. recens* and *T. dalmanni* (Dyer *et al.*, 2007; Christianson *et al.*, 2011), as well as the multiple segregating versions of the selfish autosomes SD in *D. melanogaster* and t-haplotype in mice (Lyon, 2003; Presgraves *et al.*, 2009; Brand *et al.*, 2015). In these systems, however, homozygous carriers of the driving chromosome usually have severe fitness costs, which have not been found in *D. neotestacea*. We suggest that SR has the capacity to purge deleterious mutations through occasional gene flow with ST as well as other SR chromosomes. While ST chromosomes appear to be highly recombining based on overall low levels of LD (Figures 3.5, S3), as a group SR chromosomes have substantial LD within loci but little LD between loci (Figure S3). This

suggests that over longer distances recombination does still occur between SR chromosomes to break up genetic associations, despite reduced intra-locus population recombination rates on SR relative to ST (Figure 3.5). In *D. neotestacea*, females homozygous for SR are fully fertile and comprise about 2-3% of females in the wild (Dyer, 2012; Dyer *et al.*, 2013). Even though these homozygous females are at a low prevalence, *D. neotestacea* is a common species with an overall large N_e, and thus there is likely substantial potential for recombination to occur between different SR chromosomes.

Nevertheless, SR's lower frequency and restricted recombination could lead to a history of reduced N_e compared to ST, and consequently less effective selection. K_a/K_s and π_a/π_s values, which are often used as an estimate of selection in protein coding regions, show no difference between ST and SR. However, other data suggest selection may be acting differently on SR. Specifically, we found that the flanking regions of the microsatellites on ST had both reduced segregating polymorphism and a reduced Tajima's D compared to the expectation based on the autosomes (Figures 3.6, 3.7). This may indicate especially strong purifying selection at these loci. Some of the microsatellite markers may experience reduced recombination due to their location near the end of the ST X-chromosome (Figure 3.1)(Patterson & Stone, 1952), but the pattern remains even when the most distal locus with the lowest polymorphism on ST (*neo5261*) is dropped from the analysis ($p = 0.001$, two-tailed t-test).

The potential for selection in noncoding regions of the genome is well established (e.g., Andolfatto, 2005), though to our knowledge the flanking regions of microsatellites have never been shown to be the target of selection. Alternatively, the microsatellites

themselves may be under selection for a particular repeat length (Haasl & Payseur, 2013). However, previous work using these same loci found repeat heterozygosity on ST is consistent with the expectation based on the autosomes (Dyer *et al.*, 2013). Regardless of the cause of this putative selection on ST, our data indicate that it is not acting on equivalent loci on SR, as the pattern of lower than expected polymorphism and Tajima's D at these loci is not recovered on SR. This suggests that the selfish activity of SR may reduce the efficacy of selection and thus affect the evolution of linked sequences that are otherwise independent of the driving mechanism. More information is required to fully investigate this possibility.

Our species phylogeny combines the coalescent history of all X-linked loci to indicate that ST and SR diverged after the split with sister species *D. testacea* (Figure S2). We obtain a rough age estimate for SR of 330,000 years by using the net genetic divergence between ST and SR at all X-linked markers (D_a = 0.568%) and a nucleotide substitution rate of 1.7% per million years (Table S2a)(Caccone *et al.*, 1988). *D. testacea* is found Europe and Asia (Grimaldi *et al.*, 1992), *D. orientacea* is found in Asia, and the range of *D. neotestacea* is thought to extend into Alaska (Patterson & Stone, 1952); it is possible that the last point of contact between *D. neotestacea* with the other species was across the Bering Land Bridge between 35,000 – 14,000 years ago (Meiri *et al.*, 2014). However, the individual gene trees show that the relationship of SR and ST is variable across markers, and *neo8377* shows a particularly interesting pattern where SR is basal to ST and *D. testacea* (Figure S2). As *neo8377* is one of the markers most differentiated on SR and most closely associated with the selfish phenotype, it is possible that the relationship between SR, ST, and *D. testacea* at this locus represents the true

evolutionary history of the driving loci, and at other sites on the X-chromosome recombination between ST and SR has reduced their genetic divergence since the split with *D. testacea*. Indeed, the age of SR is estimated to be much older when using only divergence between SR and ST at *neo8377* (Da = 0.993%), roughly 580,000 years (Table S2a).

An old origin for parts of SR and variable evolutionary relationships across the X-chromosome could suggest a history of sex chromosome cycling in this species due to conflict, where genetic suppression of drive allowed inactivated SR chromosomes to remain in a population as ST chromosomes (Hall, 2004). A history of coevolution with suppressors would explain the barrier to recombination across the majority of the SR chromosome in *D. neotestacea*. No segregating suppressors have been identified on either the Y-chromosome or the autosomes in *D. neotestacea* (Dyer, 2012); the most recently evolved driving locus on SR may instead be held in check by population dynamic forces (Pinzone & Dyer, 2013). While no SR drive has been identified in North American populations of *D. putrida* or European populations of *D. testacea,* SR drive and suppressors of drive have been found segregating in populations of *D. orientacea* from Japan (K. Dyer, unpublished data). Given this and the estimated age of SR in *D. neotestacea*, it is plausible that the SR drive chromosomes in *D. neotestacea* and *D. orientacea* may share the same origin.

We suggest that even with a potential decrease in the efficacy of selection on SR, the maintenance of variation through recombination with ST and other SR chromosomes has allowed the persistence of SR at high frequencies in natural populations. However, this may not be without negative effects for the ST chromosome, including a reduction in

N_e if SR is at high prevalence and the acquisition of deleterious SR-linked alleles through recombination. The genetic conflict caused by SR drive may therefore affect the evolution of the X-chromosome regardless of SR carrier status. It is possible that this has already been the case. Ultimately, a whole-genome approach will indicate the extent to which recombination is suppressed across the entire SR chromosome, and the resulting consequences for genetic variation and the strength of selection for both types of chromosomes.

Acknowledgements

We are grateful to Brooke White, Jesse Lopez, and Michael Bray for technical assistance and to four anonymous reviewers for useful suggestions. Funding was provided by a Presidential Fellowship from the University of Georgia and by the National Institute of General Medical Sciences (NIGMS) of the National Institutes of Health (NIH) under award T32GM007103 to KEP, and by grants from the National Science Foundation (DEB-1149350), The Ellison Medical Research Foundation, and The University of Georgia Research Foundation to KAD.

References

Andolfatto, P., Depaulis, F. & Navarro, A. 2001. Inversion polymorphisms and nucleotide variability in Drosophila. *Genet. Res.* **77:** 1-8.

Andolfatto, P. 2005. Adaptive evolution of non-coding DNA in Drosophila. *Nature* **437:** 1149-1152.

Babcock, C.S. & Anderson, W.W. 1996. Molecular evolution of the Sex-Ratio inversion complex in *Drosophila pseudoobscura*: analysis of the Esterase-5 gene region. *Mol. Biol. Evol.* **13:** 297-308.

Betran, E., Rozas, J., Navarro, A. & Barbadilla, A. 1997. The estimation of the number and the length distribution of gene conversion tracts from population DNA sequence data. *Genetics* **146:** 89-99.

Bouckaert, R., Heled, J., Kühnert, D., Vaughan, T., Wu, C.-H., Xie, D., Suchard, M.A., Rambaut, A. & Drummond, A.J. 2014. BEAST 2: A software platform for Bayesian evolutionary analysis. *PLoS Comput. Biol.* **10:** e1003537.

Brand, C.L., Larracuente, A.M. & Presgraves, D.C. 2015. Origin, evolution, and population genetics of the selfish Segregation Distorter gene duplication in European and African populations of *Drosophila melanogaster*. *Evolution* **69:** 1271-1283.

Burt, A. & Trivers, R. 2006. *Genes in Conflict: The Biology of Selfish Genetic Elements*. Belknap Press of Harvard University Press, Cambridge, Mass.

Caccone, A., Amato, G.D. & Powell, J.R. 1988. Rates and patterns of scnDNA and mtDNA divergence within the *Drosophila melanogaster* subgroup. *Genetics* **118:** 671-683.

Charlesworth, B. & Hartl, D.L. 1978. Population dynamics of the Segregation Distorter polymorphism of *Drosophila melanogaster*. *Genetics* **89:** 171-192.

Cheng, C., White, B.J., Kamdem, C., Mockaitis, K., Costantini, C., Hahn, M.W. & Besansky, N.J. 2012. Ecological genomics of *Anopheles gambiae* along a latitudinal cline: A population-resequencing approach. *Genetics* **190:** 1417-1432.

Christianson, S.J., Brand, C.L. & Wilkinson, G.S. 2011. Reduced polymorphism associated with X chromosome meiotic drive in the stalk-eyed fly *Teleopsis dalmanni*. *PLoS ONE* **6:** 1-7.

Cotton, A.J., Foldvari, M., Cotton, S. & Pomiankowski, A. 2014. Male eyespan size is associated with meiotic drive in wild stalk-eyed flies (*Teleopsis dalmanni*). *Heredity* **112:** 363-369.

Derome, N., Baudry, E., Ogereau, D., Veuille, M. & Montchamp-Moreau, C. 2008. Selective sweeps in a 2-locus model for sex-ratio meiotic drive in *Drosophila simulans*. *Mol. Biol. Evol.* **25:** 409-416.

Dyer, K.A. 2007. Identification and characterization of 21 polymorphic microsatellite loci from the mycophagous fly *Drosophila neotestacea*. *Mol. Ecol. Notes* **7:** 1120-1122.

Dyer, K.A., Charlesworth, B. & Jaenike, J. 2007. Chromosome-wide linkage disequilibrium as a consequence of meiotic drive. *Proc. Natl. Acad. Sci.* **104:** 1587-1592.

Dyer, K.A., White, B.E., Bray, M.J., Pique, D.G. & Betancourt, A.J. 2011. Molecular evolution of a Y chromosome to autosome gene duplication in Drosophila. *Mol. Biol. Evol.* **28:** 1293-1306.

Dyer, K.A. 2012. Local selection underlies the geographic distribution of sex-ratio drive in *Drosophila neotestacea*. *Evolution* **66:** 973-984.

Dyer, K.A., Bray, M.J. & Lopez, S.J. 2013. Genomic conflict drives patterns of X-linked population structure in *Drosophila neotestacea*. *Mol. Ecol.* **22:** 157-169.

Fabian, D.K., Kapun, M., Nolte, V., Kofler, R., Schmidt, P.S., Schlotterer, C. & Flatt, T. 2012. Genome-wide patterns of latitudinal differentiation among populations of *Drosophila melanogaster* from North America. *Mol. Ecol.* **21:** 4748-4769.

Grimaldi, D., James, A.C. & Jaenike, J. 1992. Systematics and modes of reproductive isolation in the holarctic *Drosophila testacea* species group (Diptera: Drosophilidae). *Ann. Entomol. Soc. Am.* **85:** 671-685.

Gubenko, I.S. & Evgenev, M.B. 1984. Cytological and linkage maps of *Drosophila virilis* chromosomes. *Genetica* **65:** 127-139.

Haasl, R.J. & Payseur, B.A. 2013. Microsatellites as targets of natural selection. *Mol. Biol. Evol.* **30:** 285-298.

Hall, D.W. 2004. Meiotic drive and sex chromosome cycling. *Evolution* **58:** 925-931.

Hamilton, W.D. 1967. Extraordinary sex ratios. A sex-ratio theory for sex linkage and inbreeding has new implications in cytogenetics and entomology. *Science* **156:** 477-488.

Heled, J. & Drummond, A.J. 2010. Bayesian inference of species trees from multilocus data. *Mol. Biol. Evol.* **27:** 570-580.

Helleu, Q., Gérard, P.R., Dubruille, R., Ogereau, D., Prud'homme, B., Loppin, B. & Montchamp-Moreau, C. 2016. Rapid evolution of a Y-chromosome heterochromatin protein underlies sex chromosome meiotic drive. *Proceedings of the National Academy of Sciences* **113:** 4110-4115.

Hill, W.G. & Robertson, A. 1968. Linkage disequilibrium in finite populations. *Theor. Appl. Genet.* **38:** 226-231.

Hoffmann, A.A. & Rieseberg, L.H. 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.

Hudson, R.R., Boos, D.D. & Kaplan, N.L. 1992. A statistical test for detecting geographic subdivision. *Mol. Biol. Evol.* **9:** 138-151.

Hudson, R.R. 2000. A new statistic for detecting genetic differentiation. *Genetics* **155:** 2011-2014.

Hudson, R.R. 2001. Two-locus sampling distributions and their application. *Genetics* **159:** 1805-1817.

Hurst, L.D. & Pomiankowski, A. 1991. Causes of sex-ratio bias may account for unisexual sterility in hybrids - a new explanation of Haldane's rule and related phenomena. *Genetics* **128:** 841-858.

Jaenike, J. 2001. Sex chromosome meiotic drive. *Annu. Rev. Ecol. Syst.* **32:** 25-49.

James, A.C. & Jaenike, J. 1990. Sex-ratio meiotic drive in *Drosophila testacea*. *Genetics* **126:** 651-656.

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., Thierer, T., Ashton, B., Meintjes, P. & Drummond, A. 2012. Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **28:** 1647-1649.

Kelly, J.K. 1997. A test of neutrality based on interlocus associations. *Genetics* **146:** 1197-1206.

Kingan, S.B., Garrigan, D. & Hartl, D.L. 2010. Recurrent selection on the Winters sex-ratio genes in *Drosophila simulans*. *Genetics* **184:** 253-265.

Kosambi, D.D. 1943. The estimation of map distances from recombination values. *Annals of Eugenics* **12:** 172-175.

Kovacevic, M. & Schaeffer, S.W. 2000. Molecular population genetics of X-linked genes in *Drosophila pseudoobscura*. *Genetics* **156:** 155-172.

Librado, P. & Rozas, J. 2009. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* **25:** 1451-1452.

Lindholm, A.K., Dyer, K.A., Firman, R.C., Fishman, L., Forstmeier, W., Holman, L., Johannesson, H., Knief, U., Kokko, H., Larracuente, A.M., Manser, A., Montchamp-Moreau, C., Petrosyan, V.G., Pomiankowski, A., Presgraves, D.C., Safronova, L.D., Sutter, A., Unckless, R.L., Verspoor, R.L., Wedell, N., Wilkinson, G.S. & Price, T.A.R. 2016. The Ecology and Evolutionary Dynamics of Meiotic Drive. *Trends in Ecology & Evolution* **31:** 315-326.

Lorieux, M. 2012. MapDisto: Fast and efficient computation of genetic linkage maps. *Molecular Breeding* **30:** 1231-1235.

Lyon, M.F. 2003. Transmission ratio distortion in mice. *Annual Review of Genetics* **37:** 393-408.

McVean, G.A., Myers, S.R., Hunt, S., Deloukas, P., Bentley, D.R. & Donnelly, P. 2004. The fine-scale structure of recombination rate variation in the human genome. *Science* **304:** 581-584.

Meiklejohn, C.D. & Tao, Y. 2010. Genetic conflict and sex chromosome evolution. *Trends. Ecol. Evol.* **25:** 215-223.

Meiri, M., Lister, A.M., Collins, M.J., Tuross, N., Goebel, T., Blockley, S., Zazula, G.D., van Doorn, N., Dale Guthrie, R., Boeskorov, G.G., Baryshnikov, G.F., Sher, A. & Barnes, I. 2014. Faunal record identifies Bering isthmus conditions as constraint to end-Pleistocene migration to the New World. *Proc. R. Soc. B Biol. Sci.* **281**.

Navarro, A. & Barton, N.H. 2003. Chromosomal speciation and molecular divergence--accelerated evolution in rearranged chromosomes. *Science* **300:** 321-324.

Nei, M. 1987. *Molecular Evolutionary Genetics*. Columbia University Press, New York.

Nolte, V., Pandey, R.V., Kofler, R. & Schlötterer, C. 2013. Genome-wide patterns of natural variation reveal strong selective sweeps and ongoing genomic conflict in *Drosophila mauritiana*. *Genome Research* **23:** 99-110.

Patterson, J.T. & Stone, W.S. 1952. *Evolution in the Genus Drosophila*. The Macmillan Company, New York.

Perlman, S.J. & Jaenike, J. 2003. Infection success in novel hosts: an experimental and phylogenetic study of Drosophila‐parasitic nematodes. *Evolution* **57:** 544-557.

Pinzone, C.A. & Dyer, K.A. 2013. Association of polyandry and sex-ratio drive prevalence in natural populations of *Drosophila neotestacea*. *Proc. R. Soc. B Biol. Sci.* **280:** 20131397.

Presgraves, D.C., Gerard, P.R., Cherukuri, A. & Lyttle, T.W. 2009. Large-scale selective sweep among Segregation Distorter chromosomes in African populations of *Drosophila melanogaster*. *PLoS Genet* **5:** e1000463.

Prout, T., Bundgaard, J. & Bryant, S. 1973. Population genetics of modifiers of meiotic drive I. The solution of a special case and some general implications. *Theor. Popul. Bio.* **4:** 446-465.

RCoreTeam 2014. R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing. Vienna, Austria.

Rice, W.R. 2013. Nothing in genetics makes sense except in light of genomic conflict. *Annu. Rev. Ecol. Syst.* **44:** 217-237.

Schafer, D.J., Fredline, D.K., Knibb, W.R., Green, M.M. & Barker, J.S.F. 1993. Genetics and linkage mapping of *Drosophila buzzatii*. *J. Heredity* **84:** 188-194.

Scheet, P. & Stephens, M. 2006. A fast and flexible statistical model for large-scale population genotype data : Applications to inferring missing genotypes and haplotypic phase. *Am. J. Hum. Genet.* **78:** 629-644.

Staten, R., Schully, S.D. & Noor, M.A.F. 2004. A microsatellite linkage map of *Drosophila mojavensis*. *BMC Genetics* **5:** 8.

Sullivan, W., Ashburner, M. & Hawley, R.S. 2000. *Drosophila Protocols*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.

Tajima, F. 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* **123:** 585-595.

Watterson, G.A. 1975. On the number of segregating sites in genetical models without recombination. *Theor. Popul. Biol.* **7:** 256-276.

Werren, J.H. 2011. Selfish genetic elements, genetic conflict, and evolutionary innovation. *Proc. Natl. Acad. Sci.* **108:** 10863-10870.

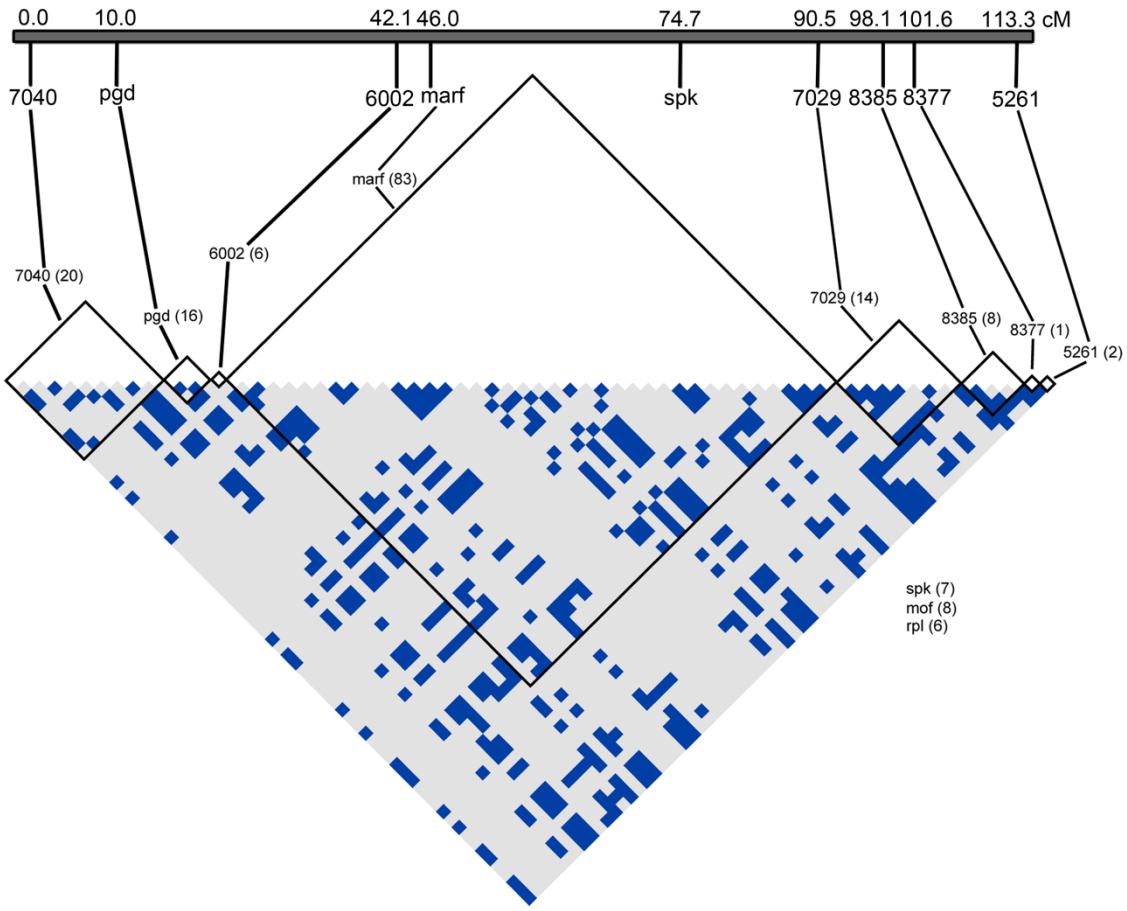


Figure 3.1. Linkage disequilibrium between SR and ST chromosomes. Recombination mapping was performed to determine location of the markers across the standard X-chromosome (ST) in centimorgans (cM). The matrix shows sites with significant R^2 values between pairs of SNPs within and between each marker calculated using only parsimony informative sites. Only sites with significant associations are pictured, with the total numbers of sites considered for each marker noted in parentheses. All ST and SR samples were included in the analysis. Significance was determined using a Fisher's Exact Test with a Bonferroni correction for multiple testing. Light grey squares indicate nonsignificant associations ($p > 0.05$) and blue squares indicate associations significant after correcting for multiple testing. *mof* and *rpl* were unable to be mapped but had no significant associations. See Figure S3 for R^2 values when ST and SR are considered separately.

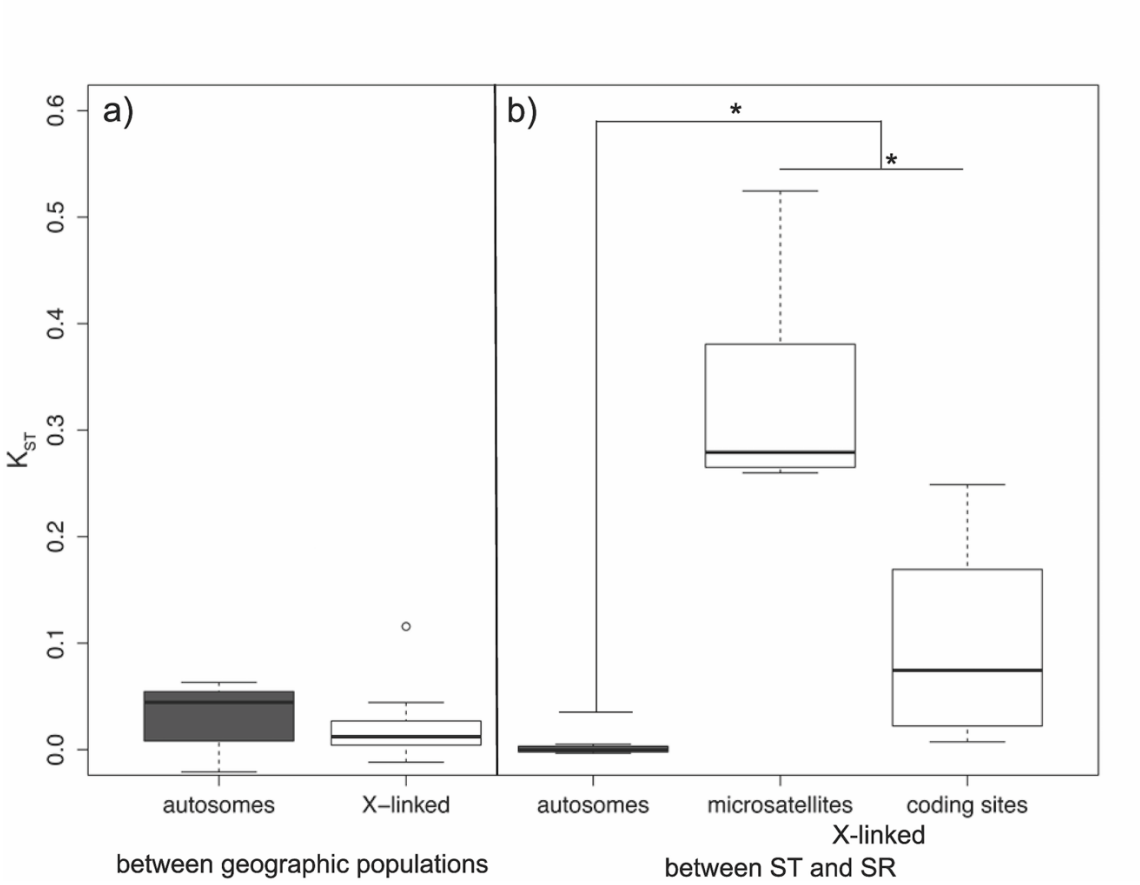


Figure 3.2. Pairwise K_{ST} per marker (a) between geographic populations and (b) between ST and SR males. White boxes are X-linked markers, and grey boxes are autosomal markers. The edges of the boxes are the first and third quartiles, and the dark grey line is the median. The whiskers extend to the last data point before 1.5x the interquartile range.

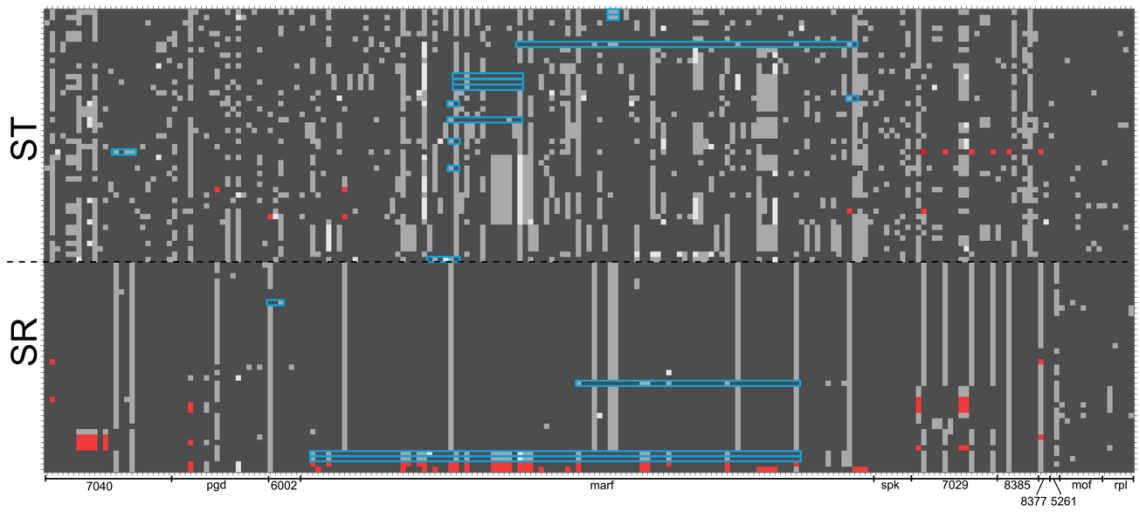


Figure 3.3. X-chromosome haplotype structure of individual samples. Each row is an individual, with ST phenotype males above the dashed black line and SR phenotype males below. Each column is a single parsimony informative site in the concatenated alignment of all the X-linked markers in the order found on the ST map. The boundaries of the markers are noted on the x-axis. Dark grey represents the individual carries the major allele, and light grey is the minor allele. Some sites have a third segregating allele, which is represented by white. Gene conversion tracts that fall within a single marker are outlined in blue. Sites that mark the beginning and end of algorithmically detected gene conversion tracts that span two or more markers are highlighted in red, as are any sites within those tracts that appear characteristic of the opposite chromosome (e.g. SR-common alleles on ST).

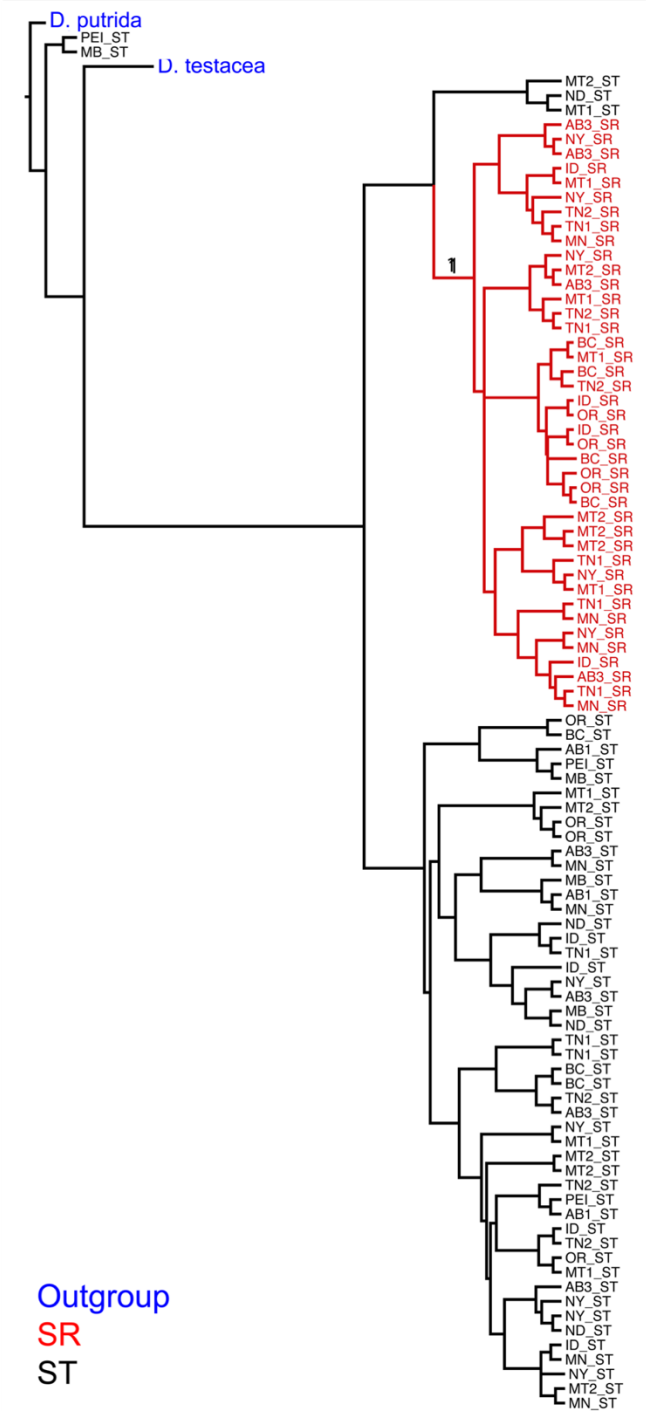


Figure 3.4. Multi-locus phylogeny of individual samples, using only X-linked markers. The tree is rooted with *D. putrida* and includes the sister species *D. testacea.* Branches with a posterior probability support higher than 0.5 are indicated. Branch tips are marked with population of origin as indicated with the state or province abbreviation and X-chromosome type. Blue text denotes outgroups, red is SR, and black is ST.

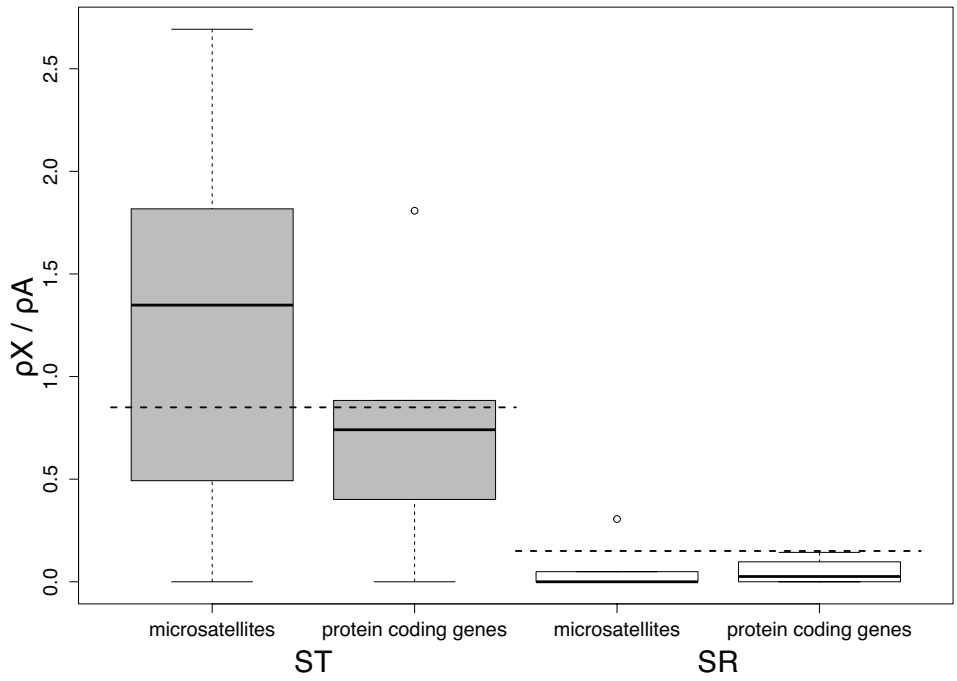


Figure 3.5. Ratio of the population recombination rate scaled by number of sites (ρ/sites) on the X-chromosome to the average ρ/site on the autosomes for ST (grey boxes) and SR (white boxes). Total segregating sites were used to calculate ρ for each marker. The horizontal dashed lines represent the expected X/A ratio of 0.85 for ST and 0.15 for SR. The edges of the boxes are the first and third quartiles, and the dark grey line is the median. The whiskers extend to the last data point before 1.5x the interquartile range.

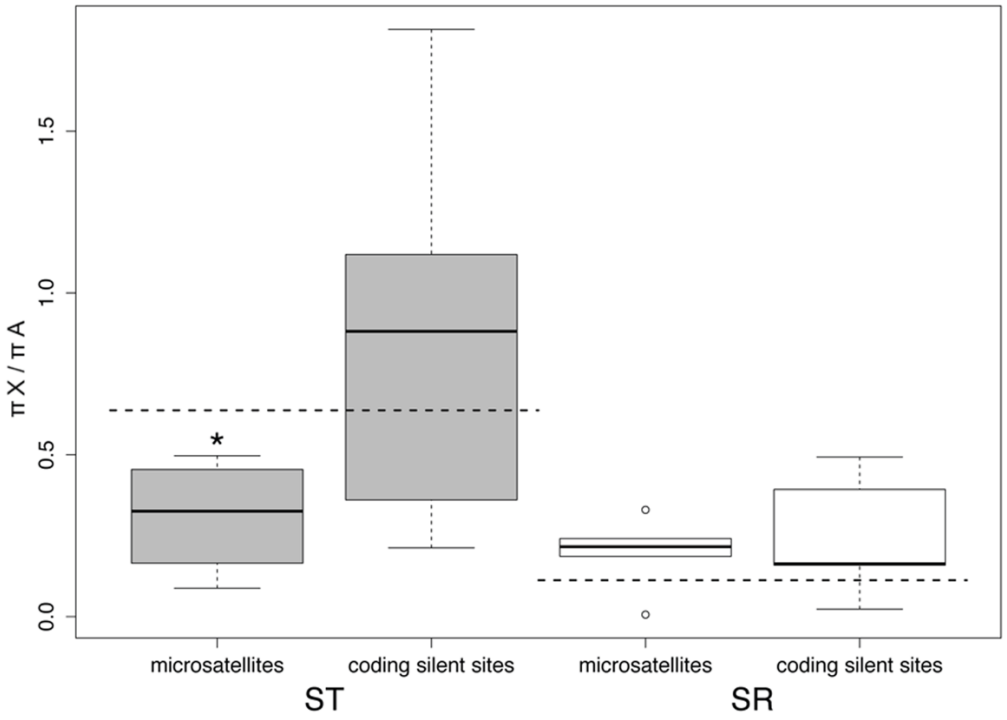


Figure 3.6. Ratio of average π (pairwise nucleotide differences) on the X-chromosome to average π on the autosomes for both ST (grey boxes) and SR (white boxes). Only silent sites were used for the protein coding genes. The horizontal dashed lines represent the expected X/A ratio of (0.75 * 0.85 = 0.64) for ST and (0.75 * 0.15 = 0.11) for SR. The edges of the boxes are the first and third quartiles, and the dark grey line is the median. The whiskers extend to the last data point before 1.5x the interquartile range.

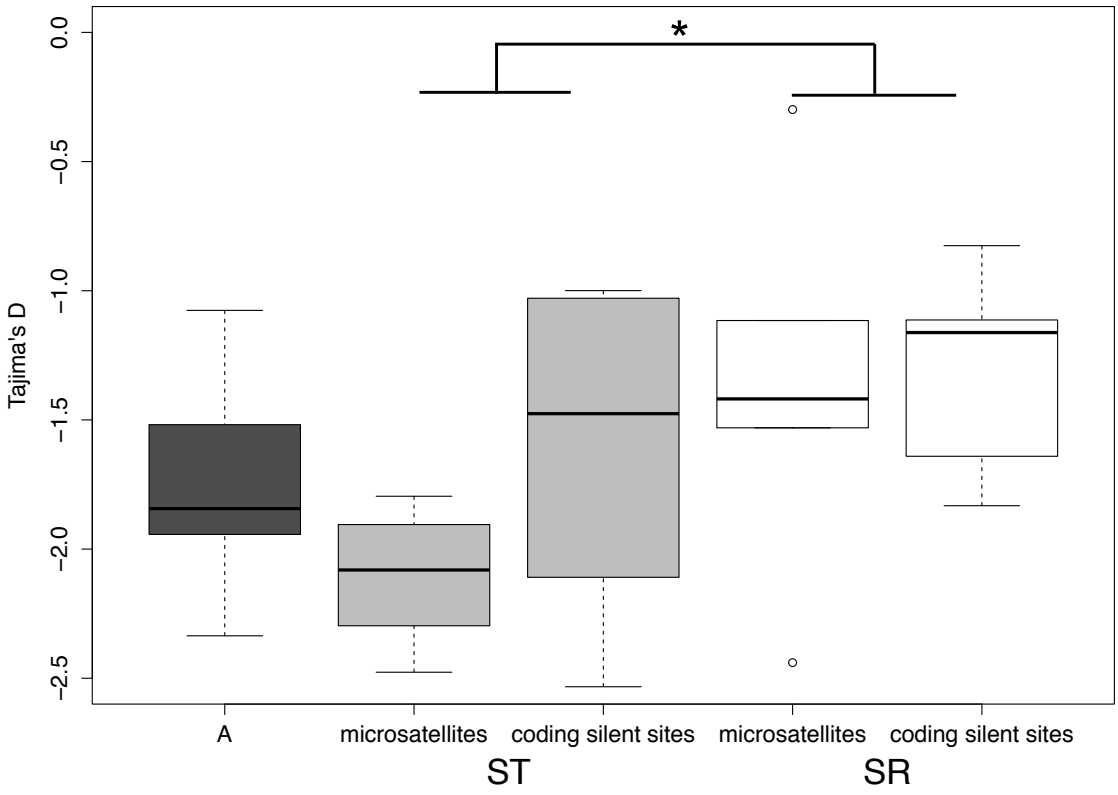


Figure 3.7. Tajima's D values for microsatellite flanking regions on ST and SR and silent sites in protein coding genes on ST (light grey), SR (white), and the autosomes (A, in dark grey). The edges of the boxes are the first and third quartiles, and the dark grey line is the median. The whiskers extend to the last data point before 1.5x the interquartile range.

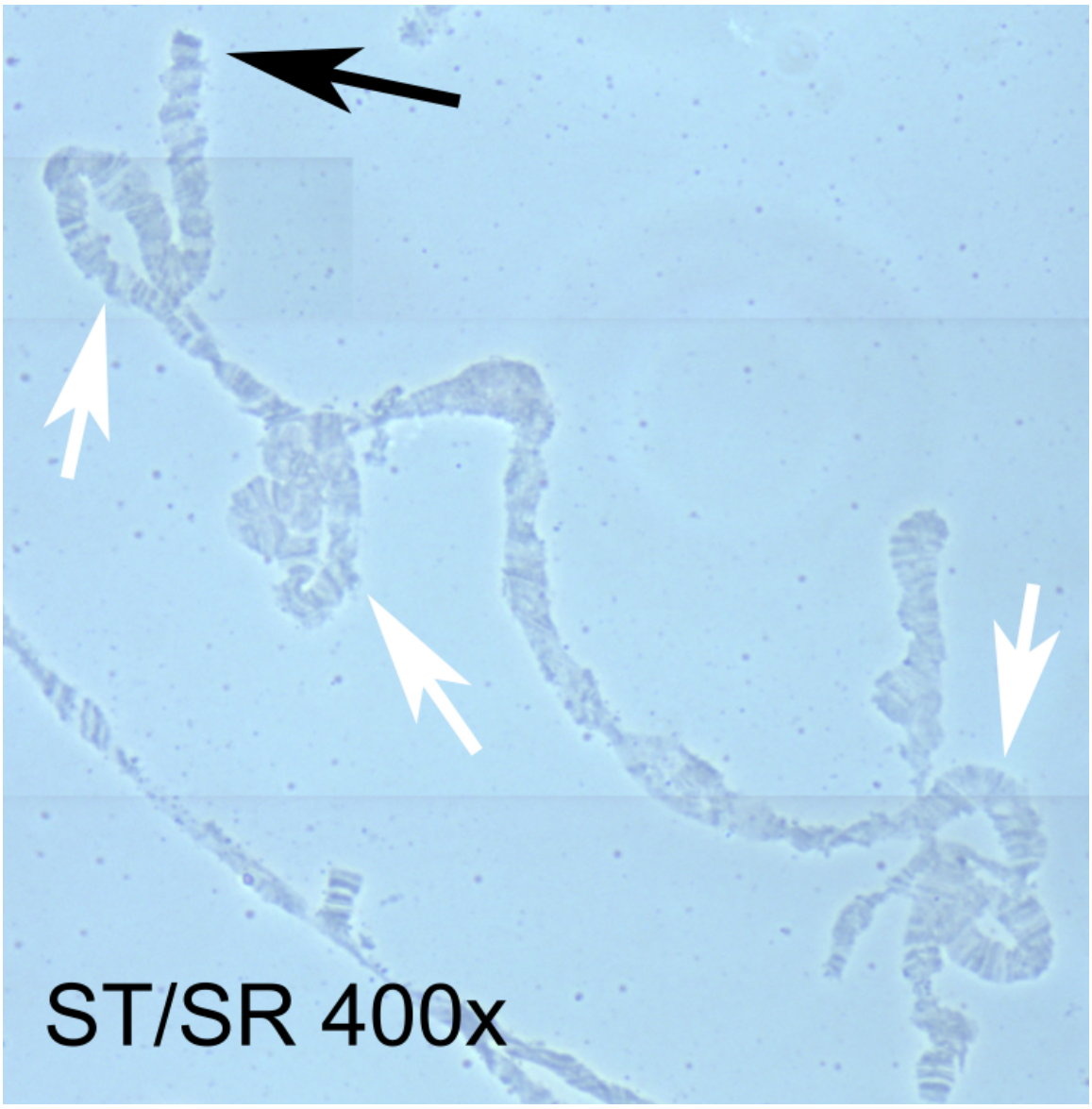


Figure 3.8. Polytene X-chromosomes of an ST/SR female. The black arrow marks the tip of the chromosome, and the white arrows mark inversions or more complex rearrangements.

CHAPTER 4

A FAST EVOLVING X-LINKED DUPLICATE OF *IMPORTIN-α2* IS ASSOCIATED WITH *SEX-RATIO* DRIVE IN *DROSOPHILA NEOTESTACEA*[1]

[1]Pieper, K.E., Unckless, R.L., and Dyer, K.A. To be submitted to *PLOS Genetics.*

Abstract

Selfish genetic elements that manipulate gametogenesis to provide themselves a transmission advantage are known as meiotic drivers. *Sex-ratio* X-chromosomes (SR) are male meiotic drivers that have evolved independently many times, and are frequently associated with large inversions that link together multiple loci required for drive. In this study, we use a combination of transcriptomics and population genetics to identify widespread expression and sequence differentiation between the standard (ST) and *sex-ratio* (SR) X-chromosomes in *Drosophila neotestacea.* We find that the X-chromosome is enriched for differentially expressed transcripts, and that differential expression between ST and SR begins early in spermatogenesis. We found that nearly half of all X-linked transcripts had at least one nucleotide difference between ST and SR, and many transcripts had K_a/K_s values suggestive of positive selection. We identified a set of candidate transcripts, including an X-linked duplicate of the nuclear transport gene *importin-α2* that is overexpressed in SR and is testis-specific. We show that positive selection has occurred on the lineage leading to the duplicate, and that its molecular evolutionary patterns are consistent with relaxed purifying selection in ST. We suggest that this gene is part of the mechanism of drive in this species, and note that nuclear transport may be a target for genetic conflict, as the mechanism of the autosomal *SD* drive system in *D. melanogaster* involves the same pathway.

Introduction

Genetic conflict occurs when one portion of the genome promotes its own transmission to the detriment of another portion of genome. Conflict is frequent,

pervasive, and potentially a major evolutionary force (Burt & Trivers, 2006; Rice, 2013; Lindholm *et al.*, 2016). Meiotic drive is a type of genetic conflict where selfish genes manipulate gametogenesis to subvert Mendel's law of equal segregation and make their way into over 50% of the gametes. Meiotic drive and its potential evolutionary consequences have been known for a long time, but recent developments in engineering synthetic drive systems for various applications gives understanding the dynamics and mechanisms of natural driving systems a new urgency (Sandler & Novitski, 1957; Lindholm *et al.*, 2016). *Sex-ratio* meiotic drive, which involves selfish elements located on the X-chromosome that reduce the transmission of the Y-bearing sperm, is particularly interesting because it can change the sex ratio of the offspring to be extremely female biased, potentially leading to population or species extinction if the driver goes to a high enough frequency (Hamilton, 1967; Carvalho & Vaz, 1999). *Sex-ratio* drive is also the most common form of chromosomal meiotic drive known, having evolved independently dozens of times in Dipterans (Jaenike, 2001). The sex chromosomes may be particularly prone to conflict (Hurst & Pomiankowski, 1991).

Though the first meiotic drive systems were discovered in the 1920s and a variety are known today, only a few have been dissected mechanistically. One reason for this is the inversions that are commonly associated meiotic drive systems (Jaenike, 2001; Burt & Trivers, 2006). Inversions will accumulate on a driving chromosomes to link together interacting elements that are required for drive (Charlesworth & Hartl, 1978). Unfortunately, the suppression of recombination caused by such inversions also makes classical genetic analysis difficult, as it is hard to isolate the genetic elements that cause drive using traditional mapping techniques.

Two *sex-ratio* drive systems have at least part of their genetic and mechanistic basis identified. Both systems are found in *Drosophila simulans,* and neither is associated with inversions (Jaenike, 2001; Tao *et al.*, 2007a; Helleu *et al.*, 2016). In the Paris system, drive has been pinpointed to a protein that binds to the heterochromatin of the Y-chromosome during meiosis and causes nondisjunction events that result in inviable Y-bearing sperm, though there is another locus associated with this system that is still unknown (Helleu *et al.*, 2016). In the independently evolved Winters *sex-ratio* system, the genetic basis of an independent SR distorter and autosomal suppressor has been identified (Tao *et al.*, 2007a; Tao *et al.*, 2007b). Suppression likely occurs through RNA interference, as the suppressor locus is a duplicate of the distorter, but the mechanistic basis of drive itself remains unclear (Tao *et al.*, 2007a; Tao *et al.*, 2007b).

One of the best-studied male meiotic drive systems is *Segregation Distorter (SD) in D. melanogaster* (Larracuente & Presgraves, 2012). Though it is an autosomal driver, it operates in much the same way as *sex-ratio* X-chromosomes where half of the sperm fail to develop properly. There are two loci required for drive in *SD*: the distorter and the responder (Larracuente & Presgraves, 2012). These two elements linked together by proximity to the centromere, and some versions of SD have other modifying loci linked to the complex with inversions (Larracuente & Presgraves, 2012) The distorter (*sd*) is a duplicate of *RanGAP* that only differs from the original in that the C-terminus is truncated (Merrill *et al.*, 1999). Wild-type RanGAP stays in the cytoplasm and powers the nuclear transport cycle, but *sd* mislocalizes to the nucleus, which disrupts nuclear transport to cause spermatogenesis to fail (Kusano *et al.*, 2001; Kusano *et al.*, 2002). Sperm carrying the *SD* chromosome have the resistant allele of the responder (*rsp*),

which is a satellite repeat of variable length where the number of repeats corresponds to sensitivity to drive (Wu *et al.*, 1988). It is unknown how *rsp* and *sd* interact to specifically target sperm carrying long *rsp* alleles, but the mechanism must involve the disruption of nuclear transport (Larracuente & Presgraves, 2012).

The t-haplotype of *Mus musculus* remains the only male meiotic driver with a known mechanistic basis that is heavily associated with inversions. The t-haplotype makes up roughly 30MB of DNA on chromosome 17, and there are four fixed inversions that prevent recombination with the wild-type (Lyon, 2003). However, the phenotype of drive is different than what is observed in most *sex-ratio* systems – sperm that do not carry the driver are produced but have serious motility defects that prevent them from successfully fertilizing females (Lyon, 2003). Drive occurs through the action of a single responder and up to four distorter loci (Lyon, 1991). The distorter loci additively upregulate expression of a sperm motility protein kinase, which causes sperm to be dysfunctional. The responder locus is a mutant version of the kinase with a lower activity than the wild-type that restores the correct expression level in the presence of the distorters (Herrmann *et al.*, 1999; Bauer *et al.*, 2005; Bauer *et al.*, 2007).

A *sex-ratio* chromosome that carries many overlapping inversions likely causes drive though a complicated mechanistic basis like the t-haplotype. One potential solution to the problem caused by these inversions is large scale data analysis such as transcriptomics or genomics. To date, the only genomic level investigation of a meiotic drive element has been a comparison of gene expression between wild-type and *sex-ratio Teleopsis dalmanni* males (Reinhardt *et al.*, 2014). This analysis showed that meiotic drive has had a significant impact on X-linked evolution in this species, as a large number

of X-linked genes were found to be differentially expressed and to carry nucleotide sequences differences between the wild-type and selfish X-chromosome (Reinhardt *et al.*, 2014). Though driving X-chromosomes at found at relatively high frequencies in this species, they appear to be highly monomorphic with no segregating polymorphisms identified in wild collections at arbitrarily chosen nucleotide and microsatellite markers (Christianson *et al.*, 2011). Loci involved in drive or very closely linked to drive are expected to show patterns similar to a selective sweep (Derome *et al.*, 2008; Kingan *et al.*, 2010). The very tight linkage and low polymorphism on the driving X-chromosome in this species suggests that loci across the X-chromosome will show the same molecular evolutionary patterns, regardless of whether they are actually involved in drive.

The *sex-ratio* X-chromosome (SR) of *D. neotestacea* is similarly characterized by large inversions and significant genetic differentiation with the standard X-chromosome (ST), but critically there is substantial variation present on SR (Pieper & Dyer, 2016). In a survey of nucleotide variation on SR and ST at 11 arbitrarily chosen loci uninvolved in drive, there were no fixed differences between the two chromosome types, and on SR linkage disequilibrium was found to vary in patterns consistent with regular recombination between SR chromosomes (Pieper & Dyer, 2016). This species is widely distributed through temporal and boreal forests in North America, and there is high gene flow across geographic regions (Dyer, 2012; Pieper & Dyer, 2016). SR is also found at long-term high frequencies in this species, and has no identified segregating autosomal or Y-linked suppressors (James & Jaenike, 1990; Dyer, 2012; Dyer *et al.*, 2013). Drive is also extremely effective in this species, with SR males producing 98%

daughters and the only sons produced are sterile X0 males that result from nondisjunction (James & Jaenike, 1990).

Altogether, these characteristics make *D. neotestacea* an excellent choice for investigating the mechanistic basis of complex, inversion rich drive systems. We used high throughput sequencing to compare gene expression in the testes of males carrying SR or ST in the same genetic background. We identified where in the genome differentially expressed genes are located, and at what stage of spermatogenesis differential expression between ST and SR begins. We identified the nucleotide differences between transcripts from ST and SR, evaluated their potential function impact, and generated a list of candidates based on putative positive selection. We then carried out a population genetic study of these candidates using wild ST and SR males and identified molecular evolutionary patterns at five loci consistent with direct involvement with the mechanism of drive. One of these top candidates is a novel X-linked duplicate of the autosomal gene *importin-α2*, which is a key part of the nuclear import pathway.

Methods

Samples, sequencing, and transcriptome assembly

The ST and SR *D. neotestacea* males used in this study come from lab stocks maintained by K. Dyer, which were originally collected in New York in 1990. The SR stock was initiated with a single wild-caught SR male and is maintained by crossing to an inbred ST lab stock. Every generation, ST/Y males are crossed to SR/SR females to generate SR/Y males, and SR/Y males are also crossed to SR/SR females to produce

more SR/SR females. Thus, the only genetic difference between ST and SR males is the X-chromosome. All flies were reared on Instant Drosophila Medium (Carolina Biological Supply, Burlington, NC, USA) supplemented with commercial mushroom (*Agaricus bisporus*) at 20 °C with 60% relative humidity on a 12-h light/dark cycle.

All libraries were made from sexually mature adult males. Testes were dissected from 25 ST and SR males at four to seven days after eclosion and flash frozen. The carcass (the body without the testes) was also prepared for RNA extraction. RNA was extracted using a Qiagen RNEasy kit (Qiagen, Germantown, MD, USA) .

Six testes cDNA libraries (three ST and three SR) and four carcass libraries (two ST and two SR) were prepared for high-throughput 75bp single-end sequencing using the Illumina TruSeq kit (Illumina, Inc, San Diego, CA, USA). All libraries were run two independent times on a single Illumina HiSeq lane (San Diego, CA) and demultiplexed at the Cornell Sequencing Center. Base quality was evaluated using FASTX-Toolkit (http://hannonlab.cshl.edu/fastx_toolkit/commandline.html) and sequencing adaptors and low quality bases were trimmed from the ends of reads, and reads with less than 99% of bases with a quality score of more than 20 were discarded.

All filtered sequencing reads from every replicate were combined and used to de novo assemble the transcriptome using the Trinity pipeline (Haas *et al.*, 2013). For transcripts that had multiple isoforms identified, the isoform with the highest quality blastx hit (i.e. lowest e-value) against UniProtKB (http://www.uniprot.org/) was included in further analyses (Camacho *et al.*, 2009; Consortium, 2014). Reads were then split back into their respective libraries and aligned back to the transcriptome using Bowtie2 with the default parameters (Langmead & Salzberg, 2012).

For each library, read abundance per transcript was quantified using RSEM (Li & Dewey, 2011). After analysis showed that there were no differences in the abundance of the technical replicates, they were combined together. The carcass data were separated from the testes data for the rest of the analyses as they were intended to only check for the presence/absence of testes expressed transcripts.

To make the SR only transcriptome assembly, the filtered, trimmed reads from only the SR libraries were used in the same de novo Trinity pipeline (Haas *et al.*, 2013). The same analysis was performed with only the filtered, trimmed ST read libraries to generate an ST-only assembly. These assemblies were used to identify whole transcripts that were only present in fragments in the combined assembly using NCBI's blastn (Camacho *et al.*, 2009).

Differential expression analysis

We used the program RUVg to normalize abundance counts between testes libraries (Risso *et al.*, 2014). First, a differential expression analysis between ST and SR was carried out in DESeq using the unnormalized data (Anders & Huber, 2012). The lowest 30% of total expressed transcripts were filtered out to increase the power, leaving a total of 25,484 transcripts in the differential expression analysis. These transcripts were ranking according to their false discovery rate (FDR), and then every transcript except the 10,000 most differentially expressed were chosen to be the RUVg empirical normalization control. These 15,484 transcripts were used to estimate unwanted variance, since we assume they are not differentially expressed. That estimation of variance was then included in the general linear model of differential expression between ST and SR in

DESeq2 and edgeR (Robinson *et al.*, 2010; Love *et al.*, 2014). The same process was carried out independently for the carcass data.

Significant differential expression between the testes of ST and SR males was evaluated using DESeq2 and edgeR (Robinson *et al.*, 2010; Love *et al.*, 2014). All analyses were performed in the R programming language in RStudio (https://www.rstudio.com/) (RCoreTeam, 2014). In both DESeq2 and edgeR, the RUVg estimate of unwanted variance was included in the GLM model used to calculate differential expression in the testes. Three biological replicates were used for ST and SR each. FDR was calculated to correct for multiple testing, and FDR ≤ 0.01 was required for significance (Benjamini & Hochberg, 1995). Transcripts were only considered significantly differentially expressed in the larger analysis if they had an FDR ≤ 0.01 in both the DESeq2 and edgeR analyses.

The same procedure was carried out for the carcass data, but as there were only two biological replicates each for ST and SR the power of this analysis was very low and no transcripts met the criteria for significance. However, the mean expression estimates from DESeq2 were used to determine the tissue specificity of transcripts differentially expressed in the testes.

Transcript annotation

To identify X-linked transcripts, transcript homology to the genomes of *D. melanogaster* and *D. virilis* were used. The most recent *D. melanogaster* whole genome assembly was obtained from Flybase (http://flybase.org/) and a version of the *D. virilis* genome assembly with scaffolds assigned to Muller elements was obtained from Dr.

Yasir Ahmed-Braimah (personal communication). NCBI's tblastx was used with cut off values of e-value < $1e^{-20}$ and length > 50 bp to find transcript homologs in the *D. melanogaster* and *D. virilis* genomes. If no *D. virilis* homolog was available or the transcript mapped to a scaffold, the location of the *D. melanogaster* homolog was used. Synteny between Muller elements was used to assign genomic location of transcripts in *D. neotestacea* (Schaeffer *et al.*, 2008; Camacho *et al.*, 2009). Though the arrangement of the autosomes in *D. neotestacea* is different than in *D. melanogaster* and *D. virilis,* the X-chromosome is homologous in all three species and unlinked to other Muller elements in *D. neotestacea* (Pieper & Dyer, 2016).

Previous research on gene expression in subsections of the testes has identified sets of *D. melanogaster* genes expressed in the individual mitosis, meiosis, and postmeiosis stages of spermatogenesis (Vibranovski *et al.*, 2009). Any genes expressed during multiple stages were removed, and the translated proteins of the remaining set were downloaded from Flybase (http://flybase.org/). As our RNAseq data was generated from whole testes, genes expressed during every stage of spermatogenesis are included in the dataset. *D. neotestacea* transcripts were assigned to spermatogenesis stages by finding homology with the previously identified sets using NCBI's blastx (e-value > $1e^{-20}$) (Camacho *et al.*, 2009). Any transcripts that appeared in more than one stage were discarded.

GO term enrichment analysis was carried out using the GOseq program in the Trinity package (Haas *et al.*, 2013). Transcript homologs were identified in the SwissProt database using BLASTX (e-value < $1e^{-20}$) (Camacho *et al.*, 2009; Consortium, 2014).

K_a/K_s analyses

Transcripts with nucleotide sequence differences in between ST and SR were found using the mpileup function of samtools to make a vcf file to compile the variant sites between every single library (Li *et al.*, 2009). Vcftools was used to remove any variant sites with a minor allele frequency of less than 50% (Danecek *et al.*, 2011). Further filtering was performed to remove any nucleotide sites with coverage < 100x across all samples and any transcripts with less than 100 sites with at least 100x coverage. Finally, any sites with heterozygotes called in any library were removed. The remaining set of sites contained one allele in every SR sample and a different allele in every ST sample. The percent sequence difference between ST and SR was calculated for each transcript by taking the number of different sites over the total length of the transcript and multiplying by 100. Of the 1,349 transcripts that had at least one sequence difference, 1,067 mapped to the X-chromosome using homology, 32 mapped to the autosomes, and 250 had unknown genomic locations (Figure S1). This confirms that the only genetic differences between ST and SR should be on the X-chromosome, so the 32 transcripts originally assigned to the autosomes were reassigned to the X-chromosome. The 250 transcripts with unknown locations that had sequence differences were also moved to the X-chromosome.

Using the original transcriptome assembly as the reference, GATK was used to create a fasta file of transcript sequences carrying the alternate alleles at each variable site (Van der Auwera *et al.*, 2013). Alternate and reference allele containing transcripts were then carefully partitioned between ST and SR based on the output of the original vcf file. Some SR transcripts contained a mixture of reference and variant alleles at different sites;

these were corrected by hand in Geneious (Kearse *et al.*, 2012). Open reading frames were identified in the transcripts using Transdecoder (Haas *et al.*, 2013). The coding sequences of the SR and ST transcripts were then extracted using bedtools (Quinlan & Hall, 2010). The validity of the open reading frames was confirmed by eye in Geneious (Kearse *et al.*, 2012). K_a/K_s was calculated for the coding region of each transcript using KaKs_Calculator (Zhang *et al.*, 2006).

Molecular evolutionary analyses of candidates

A list of candidate transcripts was compiled considering their K_a/K_s value, differential expression between ST and SR, and testis-specific expression. Two candidate transcripts (TR10603 and TR2814) were found to both be part of the same gene, which we named *X-pendulin;* as the results show, this gene is an X-linked duplication of the autosomal *Pendulin* gene (see Results). The population genetic patterns of *X-pendulin,* two other candidate transcripts (TR23125 and TR24932), and four transcripts with $K_a/K_s > 1$ (TR261, TR11103, TR37304, and TR50351) were investigated. PCR primers were designed with Primer3 in Geneious (Table S1)(Kearse *et al.*, 2012; Untergasser *et al.*, 2012).

A selection of 10 wild-caught ST and 10 wild-caught SR *D. neotestacea* males were chosen for sequencing. These males were randomly chosen from the range-spanning dataset used in (Pieper & Dyer, 2016) (See Table S1 of that paper) using a random number generator (random.org). The males were identified as carrying ST or SR X-chromosomes by the proportion of female offspring they produced (Dyer, 2012). An additional seven ST males were chosen for sequencing *X-pendulin*. All transcripts were

also sequenced in one or two individuals of the outgroups *D. testacea* and *D. orientacea*. Transcripts TR2814 and TR10603 were also sequenced in individuals of *D. putrida* and *D. bizonata* to confirm the phylogenetic distribution of the *X-pendulin* duplication. Fragments were amplified using standard PCR protocols and sequenced on an Applied Biosystems (Foster City, CA, USA) 3730xl DNA Analyzer at the Georgia Genomics Facility. Base calls were confirmed using Geneious and sequences were aligned by hand (Kearse *et al.*, 2012).

Analyses of DNA polymorphism and divergence at the candidate loci were carried out in the program DnaSP (Librado & Rozas, 2009). Population genetic data from a set of five X-linked (*marf, mof, pdg, rpl,* and *spk*) and seven autosomal protein coding loci (*esc, gl, ntid, mago, tpi, sia,* and *wee*) random with respect to sex-ratio were used as a comparison; these data are from (Pieper & Dyer, 2016). Hudson-Kreitman-Agoudé (HKA) tests were carried out in the program MLHKA using an MCMC length of 1,000,000 (Wright & Charlesworth, 2004). Significance was determined using likelihood ratio tests. Coding sequences of the genes *importin-α1,2,* and *3* in *D. melanogaster* and *D. virilis* were obtained from Flybase (http://flybase.org/) and combined with the *D. neotestacea* transcriptome sequences of all three *importin-α* genes and *X-pendulin*, the X-linked duplicate of *importin-α2*, to build an unrooted neighbor-joining tree in Geneious with 1,000 bootstrap samples (Kearse *et al.*, 2012). This tree was then used to analyze branch specific d_N/d_S patterns using the codeml function of PAML (Yang, 2007).

Results

The X-chromosome is enriched for differential expression

In total, the transcriptome contains 63,821 transcripts with an average contig length of 376.66bp. The contig N50 value is 514bp. It is likely that the number of transcripts in this dataset overestimates the actual number of genes expressed in all tissues in this study due to fragmented assembly. Indeed, one of the candidate genes (X-pendulin) was uncovered as three different X-linked candidate transcripts, but using PCR on genomic DNA makes it clear these are all part of the same gene. This is likely not a unique event. Nevertheless, transcript expression levels are a good proxy for gene expression.

In total, 729 transcripts were identified as differentially expressed (DE) in the testes of ST and SR males (FDR ≤ 0.01 via both DESeq2 and edgeR). Of these, 423 had SR-biased expression patterns and 423 were ST-biased (Figure 4.1), and 19 transcripts were expressed in SR but absent in ST, and 149 transcripts were expressed in ST but absent in SR. Though all transcripts mapped relatively uniformly to the large Muller elements (the X-chromosome and autosomes B, C, D, and E), 41.5% (303 transcripts) of DE transcripts mapped to the X-chromosome (Figure 4.1a). Notably, a roughly equal proportion of ST and SR-biased transcripts are found on the X-chromosome. The backcrossing design used to maintain the ST and SR lab stocks ensures that the only genetic difference between ST and SR males is the X-chromosome (see *Methods*), so all transcripts with fixed sequence differences between ST and SR were assumed to belong on the X-chromosome, even if they had initially mapped to the autosomes (32 transcripts) or an unknown location (250 transcripts). In contrast, the other four large autosomes

combined harbor only 18% (131 transcripts) of DE transcripts, despite having a total number of mapped transcripts nearly quadruple that of the X-chromosome. No differentially expressed transcripts were found to map to either the dot chromosome (Muller element F), the Y-chromosome, or the mitochondria (Figure 4.1a). The remaining 40.5% (295 transcripts) of DE transcripts were not able to be mapped to any region of the genome using homology with *D. virilis* and *D. melanogaster* or sequence differences between ST and SR. A chi-squared test comparing DE transcripts with total mapped transcripts for the five large Muller elements was highly significant ($\chi^2 = 868.01$, df = 4, p-value < 2.2e-16), indicating the X-chromosome is enriched for DE transcripts.

Differential expression begins early in spermatogenesis

All transcripts were also evaluated for their involvement in the mitosis, meiosis, or postmeiosis stage of spermatogenesis using a comprehensive dataset of *D. melanogaster* genes known to be primarily expressed at specific stages (Vibranovski *et al.*, 2009). Transcripts were assigned stages based on homology to the *D. melanogaster* genes, and then patterns of differential expression were examined (Figure 4.1b). Most DE transcripts (79.9%, 583 transcripts) were not expressed primarily at a specific stage of spermatogenesis. This majority was also split between ST (378 transcripts) and SR-biased transcripts (205 transcripts). DE transcripts occur at all three stages of spermatogenesis, and a chi-squared test comparing the numbers of DE and non-DE transcripts at each stage was not significant ($\chi^2 = 2.0693$, df = 2, p-value = 0.35), suggesting the effects of SR do not occur predominantly at one stage (Figure 4.1b). However, there is a difference between the proportion of ST-biased and SR-biased

transcripts expressed during each stage. Mitosis (61 SR-biased, 29 ST-biased) and meiosis (19 SR-biased, 0 ST-biased) have an excess of SR-biased transcripts, whereas in post-meiosis SR-biased and ST-biased transcripts are found at similar levels (21 SR-biased, 16 ST-biased). A chi-squared test comparing the number of expression biased transcripts at each stage was significant ($\chi^2 = 11.225$, df = 2, p-value = 0.0037), suggesting that significantly more SR-biased transcripts are expressed during spermatogenesis. This could be because the disruption of spermatogenesis by SR causes the overexpression of some typical spermatogenesis genes, perhaps by increasing the time spent in particular developmental stages.

Nucleotide differences between ST and SR are widespread

Nucleotide differences between ST and SR libraries were identified using stringent criteria. Because the ST and SR samples both came from lab strains that are maintained in the same genetic background, the only genetic differences between them are on the X-chromosome. This is confirmed by the lack of sequence differences on transcripts on the autosomes (Figure S1). Of the 3,030 transcripts that mapped to the X-chromosome, 44.5% (1,349 transcripts) had at least one nucleotide difference between ST and SR (Figure 4.2a). Because of the stringent criteria used to identify sequence differences, this is likely an underestimation of differentiation between ST and SR. The percent difference for these 1,349 transcripts was calculated as the number of sequence differences over the total length of the transcript. The sequence divergence of X-linked transcripts ranged from 0.03 – 4.5% with a mean percent difference of 0.53% (Figure 4.2a). In the coding regions of five genes arbitrarily chosen with respect to SR, the

average percent sequence divergence in a large population genetic sample of ST males ranged from 0.16 to 1.25%, with an average of 0.56% (Figure 4.2a)(Pieper & Dyer, 2016). There was no association between DE and percent sequence difference, with the most highly DE transcripts having no fixed differences between SR and ST (Figure 4.2a).

The potential functional consequences of these nucleotide differences were evaluated by calculating K_a/K_s values for each transcript. Of the 1,349 transcripts with sequence differences, 1,116 transcripts had annotated open reading frames with nucleotide differences within them, allowing K_a/K_s to be estimated. Of these, 674 transcripts had three or more sequence differences (Figure 4.2b). Of these 674 transcripts, 15 had no synonymous substitutions. Of the remaining transcripts, the average K_a/K_s value was 0.23, and there were 31 transcripts with $K_a/K_s > 1$ (Figure 4.2b). Combined with the transcripts that had over three nucleotide differences but no synonymous differences (N = 15), there were a total of 46 potentially positively selected transcripts. Notably, these potentially positively selected transcripts were not the transcripts with the most substitutions (Figure 4.2b).

A series of chi-square tests were carried out to determine if positively selected transcripts were overrepresented in the sets of DE transcripts, spermatogenesis transcripts, or testes-specific transcripts (Table 4.1). Interestingly, there are significantly fewer positively selected transcripts involved in spermatogenesis than transcripts with neutral K_a/K_s values ($\chi^2 = 9.7164$, df = 1, p = 0.002) (Table 4.1). This could be due to purifying selection on transcripts with necessary functions during spermatogenesis. The positively selected transcripts are enriched for testes-specificity ($\chi^2 = 10.383$, df = 1, p =

0.001), but there was no significant difference in the distribution of DE transcripts between positively selected transcripts and not ($\chi^2 = 2.4073$, df = 1, p = 0.121).

Identification of candidate genes

To narrow down our list of candidate transcripts for involvement in the driving mechanism, we focused on those that have $K_a/K_s > 1$ and have testes-specific differential expression between ST and SR (Figure 2b). Six transcripts matching these criteria were identified: TR10603, TR23125, TR24932, TR2814, TR6297, and TR5481 (Table S2). The candidate transcripts were analyzed for homology with *D. melanogaster* and *D. virilis* proteins using BLASTX. Two of them (TR24932 and TR5481) had no identifiable orthologs. One (TR6297) was identified as an ortholog to the *D. melanogaster* gene *CG7366*, which is located on an autosome in *D. melanogaster* and highly expressed in the testes with an unknown function (http://flybase.org/reports/FBgn0035855.html). Another candidate transcript (TR23125) is a homolog of an X-linked DNA polymerase V called *lethal (1) 1Bi*, which is expressed during the mitosis stage of spermatogenesis (http://flybase.org/reports/FBgn0001341.html) The final two candidate transcripts (TR10603 and TR2814) both mapped to an autosomal gene called *Pendulin,* also known as *importin-α2* (http://flybase.org/reports/FBgn0267727.html). Aligning these two transcripts to the *Pendulin* sequence revealed that they both mapped to perfectly sequential sections of the protein, indicating these two transcripts are fragments of a whole open reading frame (Figure S2).

We searched the transcriptome data for other transcripts that mapped to *Pendulin* to identify the other half of the open reading frame. This search revealed the beginning of

the X-linked transcript (TR37105), which overlapped in sequence with the two identified candidate transcripts. TR37105 also has sequence differences between ST and SR, but has a K_a/K_s value of only 0.35 (Table S3). Additionally, it also has SR-biased expression like the other two transcripts, but the adjusted p-value is 0.013, and so it was not counted as significantly DE in the earlier analysis. Together, these three transcripts make up nearly the complete open reading frame of *Pendulin,* with only 126bp before the stop codon missing. To find the complete end of the transcript, we built ST and SR specific assemblies and then queried them with the transcript from the combined assembly using blastn (Camacho *et al.*, 2009). Through this process, we identified the entire open reading frame of the transcript and the 3' UTR.

Searching for transcripts with homology to *Pendulin* also revealed a full-length transcript (TR7043) that covered the entire open reading frame of *Pendulin* and included 5' and 3' UTR sequences. Unlike the three candidate transcripts, this transcript had no sequence differences between ST and SR and was not differentially expressed or testis-specific. This transcript also had much higher homology with the *D. melanogaster* and *D. virilis Pendulin* sequences compared to the candidate transcripts (Figure 4.3, Figure S2). Therefore, TR7043 must represent the homolog of the autosomal copy of *Pendulin*, whereas the three candidate transcripts and the transcripts from the other assembly together must represent an X-linked duplicate copy. These transcripts were concatenated together to form a transcript referred to hereafter as *X-pendulin,* which is 2,106bp long in SR and 1,886bp long in ST. The different lengths are due to the different length of the transcripts containing the end of the open reading frame and the 3' UTR pulled from the ST and SR-specific assemblies. The open reading frame of *X-pendulin* is the same length

in both ST and SR, but is 16 amino acids shorter than the open reading frame of the autosomal gene (Figure S3). Indeed, visual inspection suggests the C-terminal end of *X-pendulin* is more diverged from the autosomal copy than the rest of the open reading frame (Figure S3).

Pendulin, also known as *importin-α2,* is a member of the importin-*α* gene family along with *importin-α1* and *importin-α3*. Sequences of *importin-α1* and *3* from *D. melanogaster* and *D. virilis* were obtained from Flybase (http://flybase.org/). Full length transcripts of *D. neotestacea importin-α1* (TR22571) and *importin-α3* (TR6773) were found in the transcriptome data. Neither of these transcripts were DE or had sequence differences between ST and SR, consistent with their autosomal location. A neighbor-joining unrooted phylogenetic tree was built from all members of the *importin-α* family in *D. melanogaster, D. virilis,* and *D. neotestacea*, including *X-pendulin* (Figure 4.3). Clearly, both the ST and SR *X-pendulin* sequences are most closely related to the *importin-α2* copy in *D. neotestacea,* indicating they are paralogs (Figure 4.3).

An analysis was carried out to estimate the d_N/d_S values for each branch of this tree in PAML. Likelihood ratio tests supported a model where each branch of the tree has an independent d_N/d_S value over a model with one value across the entire phylogeny ($\chi^2 = 485.5$, df = 18, $p < 0.0001$). Within the *importin-α1* and *importin-α3* clades, the estimated d_N/d_S values are uniformly quite low, indicating strong purifying selection (Figure 4.3). The same is true of the *importin-α2* branches for the two outgroups and the autosomal copy in *D. neotestacea* (Figure 4.3). However, the branch leading from the autosomal copy in *D. neotestacea* to the split between the X-linked copies has a K_a/K_s value of 1.03, and the branches leading to the individual ST and SR sequences have

relatively high values of 0.86 and 0.49 respectively (Figure 4.3). Clearly this section of the tree is evolving differently than the rest.

PCR and sanger sequencing of *X-pendulin* in a selection of *D. neotestacea, D. testacea, D. orientacea, D. putrida, and D. bizonata* individuals showed that the X-linked duplication is specific to the testacea-species group, being found in the very closely related *D. neotestacea, D. testacea, and D. orientacea* but not in the more distantly related *D. putrida* or *D. bizonata,* an outgroup to the testacea-species group (Dyer *et al.*, 2011).

Molecular evolutionary patterns of candidates

To examine molecular evolutionary patterns of *X-pendulin* and the other candidate loci, we carried out PCR amplification and sanger sequencing in a selection of roughly 10 ST and SR males from across the species range (Table 4.2, Table S4). Primers were chosen to sequence 1,2777bp of *X-pendulin* that span the first two thirds of the open reading frame and thus all three transcripts identified in the combined assembly. The K_a/K_s candidate transcripts TR23125 and TR24932 were also chosen for population genetic sampling, along with the four other transcripts that had positive K_a/K_s values but did not otherwise meet the candidacy criteria (TR261, TR11103, TR37304, and TR50351) (Table S2). TR261 has 8 nonsynonymous differences and no synonymous differences, and is testis-specific but not DE. It maps to the autosomal *D. melanogaster* gene *CG32371*, which has an unknown function but is highly expressed in the testes (http://flybase.org/reports/FBgn0052371.html). TR11103 has a K_a/K_s value of 2.12, but is not DE or testis-specific. It maps to the X-linked *D. melanogaster* gene *CG2685*, which

binds to mRNA and is involved in mRNA splicing and processing (http://flybase.org/reports/FBgn0024998.html). Notably, this gene interacts with *small bristles (sbr),* which plays a role in exporting mRNA out of the nucleus and itself interacts with other members of the nuclear import/export pathway including *RanGAP* and *importin-ß* (http://flybase.org/reports/FBgn0003321.html). TR37304 has a K_a/K_s value of 1.56 and is primarily expressed during postmeiosis. It is not testis-specific or DE according to strict criteria, but it is ST-biased in expression. It maps to the X-linked *D. melanogaster* gene *CG4198*, which has an unknown function but is highly expressed in the testes (http://flybase.org/reports/FBgn0029753.html). TR50351 has a Ka/Ks value of 1.20, and is not DE or testis-specific. It maps to the X-linked *D. melanogaster CG15452*, which also has an unknown function but is highly expressed in the testes (http://flybase.org/reports/FBgn0031130.html). Nucleotide polymorphism for each of these markers was quantified and compared to a set of five X-linked markers arbitrarily chosen relative to SR and seven autosomal markers (Table S5; see also (Pieper & Dyer, 2016)).

One of the most notable contrasts between the top candidates and the arbitrarily chosen non-candidate X-linked markers is the presence of fixed differences and the general lack of shared mutations in the candidates (Table 4.2). In the non-candidate X-linked markers, including the additional six non-coding markers that are not included in the present analyses, no marker had any fixed differences between ST and SR (Pieper & Dyer, 2016). However, two of the candidates also had no fixed differences. TR23125 was calculated to have a K_a/K_s of 1.29 between the ST and SR transcripts but no fixed differences were present in the population genetic sample. The same was true of

TR50351, which had a Ka/Ks of 1.20 between ST and SR transcripts, but had 18 shared mutations between ST and SR males (Table 4.2). On these grounds, these two loci were excluded from the set of candidates in further analyses, which was amended to include all sequenced loci with fixed differences between ST and SR (Table 4.2). This set of chosen loci is referred to subsequently as the "top candidates". However, it is important to note that even though TR23135 had no fixed differences, it does appear to be evolving non-neutrally as the ST samples have no synonymous polymorphisms, but eight non-synonymous polymorphisms. Considering only the top candidates, the high number of fixed differences and low number of shared mutations between ST and SR is reflected in the K_{ST} values estimating differentiation between ST and SR. In the top candidates, the mean K_{ST} between ST and SR (0.44, sd = 0.27) was significantly higher than in the set of non-candidate X-linked markers (0.10, sd = 0.10) ($t = 2.82$, $df = 6.65$, $p = 0.03$) (Figure S4).

The candidate markers, including *X-pendulin,* showed a pattern characterized by very low diversity on SR, but high diversity, particularly non-synonymous diversity, on ST (Figure 4.4). Because the average population frequency of SR is about 15%, the neutral expectation of polymorphism on SR is much lower than ST. Normally, the X-chromosome is expected to carry 0.75 times the amount of polymorphism on the autosomes. The expected polymorphism on SR is thus 0.75*0.15 times the autosomal polymorphism, and the expected polymorphism on ST is 0.75*0.85 times the autosomal polymorphism. The amount of polymorphism measured with π at the X-linked non-candidate markers falls in line with these expectations (Figure 4.4, Table S5). On SR, the top candidate markers also have an expected amount of silent polymorphism (mean =

0.001, sd = 0.002, t = -1.7, df = 4, p-value = 0.16) and nonsynonymous polymorphism (mean = 0.002, sd = 0.004, t = 1.2, df = 4, p-value = 0.29) (Figure 4.4a,b). On ST, the synonymous π of the top candidates is in line with the expectation (mean = 0.013, sd = 0.009, t = -0.59, df = 4, p-value = 0.59), but the amount nonsynonymous π was significantly higher than the expectation (mean = 0.008, sd = 0.005, t = 2.8796, df = 4, p-value = 0.04503) (Figure 4.4a,b). The amount of non-synonymous variation in ST at *X-pendulin* can be easily viewed in Figure 4.4d, as well as the 4 synonymous and 4 nonsynonymous fixed differences between ST and SR. One ST sample of *X-pendulin* also contains a codon deletion (Figure 4.4b), and one of the ST sequences of TR261 contained a single base insertion that resulted in a premature stop codon (Figure S5).

Examining the polymorphism at *X-pendulin* also reveals the presence of two separate haplotypes at roughly equal frequency in SR (Figure 4d). This distribution of variation contrasts with the many singletons observed in the ST samples. At all the top candidates, the site frequency spectrum in SR was skewed towards common variants (Figure 4.4C). A nested ANOVA comparing Tajima's D of the top candidates and non-candidate X-linked loci in ST and SR males showed a significant effect of top candidate vs. non-candidate locus ($F_{1,14} = 7.505$, p = 0.016) and a significant effect of ST vs SR ($F_{2,14} = 5.008$, p = 0.023). This pattern is driven by the top candidates on SR, which have a much more positive Tajima's D (mean = 0.135, sd = 1.037) than the top candidates on ST (mean = -1.196, sd = 0.286) and any other loci in the genome, including the autosomes (Figure 4.4C). This trend is also evident when examining Tajima's D estimated with only synonymous or nonsynonymous variation (Figure S6).

HKA tests were performed to test if positive selection could explain the polymorphism differences between the top candidates and the non-candidate X-linked loci. An inheritance scalar of 0.6375 (0.75*0.85) was used for all loci on ST, and 0.1125 (0.75*0.15) was used for all loci on SR. When only considering SR samples, a maximum likelihood analysis best supported a model where the five top candidate loci are under positive selection compared to the non-candidate loci (LRT, $\chi^2 = 14.62$, df = 6, p = 0.023). For ST samples, however, the best supported model was one with all of the top candidate and non-candidate loci evolving neutrally (LRT, $\chi^2 = 0.0228$, df = 1, p = 0.88) (Table S6). This is likely due to the high divergence between ST and SR at the candidate loci, combined with the low diversity at the candidates on SR (Table S6).

Discussion

In this study, we identified transcripts that are differentially expressed between the testes of *D. neotestacea* males carrying a wild-type X-chromosome and those carrying the selfish *sex-ratio* X-chromosome. The X-chromosome was enriched for differentially expressed genes, and that differential expression begins in the earliest stages of spermatogenesis. We found that nearly half of transcripts originating from the X-chromosome contain at least one nucleotide difference, though many harbor many more. We found that 46 of the transcripts were potentially under positive selection (i.e., they either had $K_a/K_s > 1$ or three or more nonsynonymous differences with no synonymous differences). We identified a set of candidate transcripts for involvement in the mechanism of drive, and found that two them were part of a rapidly-evolving X-linked duplicate of the nuclear transport gene *importin-α2* called *X-pendulin*. We

sequenced *X-pendulin* and six other candidate transcripts in a population genetic sample and found signatures of positive selection in SR and relaxed purifying selection in ST.

Widespread divergence and differential expression

Several patterns in the RNA sequencing data reveal important qualities of SR. First, there is widespread differentiation between ST and SR, both in differential expression and nucleotide sequence differences (Figure 4.1, Figure 4.2). Nearly half of all transcripts that mapped to the X-chromosome had at least one sequence difference between ST and SR, and some showed signatures suggestive of positive selection. Though this variation is just a sample from two genotypes and most of these differences are probably not fixed, it clearly shows how drive has had a large impact on the X-chromosome. Any two individual X-chromosomes from natural populations would be expected to have many sequence differences between them, and the average percent difference is very close to the average percent difference between a large sample of ST chromosomes (Figure 4.2a). However, this rough comparison is only based on five loci, and so it is difficult to say how many transcripts fall above a neutral expectation. The tail of highly diverged transcripts with high percent differences and particularly transcripts with $K_a/K_s > 1$ likely indicate regions of particularly strong differentiation between SR and ST. Previous work on a small scale showed that differentiation between ST and SR was both widespread and variable across the chromosome, and this genomic scale investigation supports these conclusions (Pieper & Dyer, 2016).

The number of DE transcripts between ST and SR is large, and most of these are found on the X-chromosome (Figure 4.1). Differences in expression on the autosomes

must be activated downstream of DE transcripts on the X-chromosome. Though the presence of inversions on SR suggest the mechanism of drive must involve more than one gene, most of the 729 DE transcripts are likely not directly involved in the drive phenotype. t-haplotype is the most complex drive system that we know of where the genetic basis has been investigated, and it only has up to five mechanistic drive loci (Lyon, 2003). Many of the DE genes in this study likely represent activation of downstream pathways like apoptosis. In other cases, differential expression on the X-chromosome may represent diverging gene expression levels that are due to neutral divergence caused by linkage to the components of SR and thus suppressed from recombining with ST (Harrison *et al.*, 2012). Though nucleotide differences within transcripts is not correlated with expression differences (Figure 4.2a), there may be accumulation of differences in cis-regulatory regions that is not captured in the present analysis. Diverging expression due to drift is also interesting in light of the lack of phenotypic differences between individuals who carry SR and those who carry ST. Similar results were seen in a transcriptomics analysis of X-linked drive in the stalk-eyed fly, though in that system there are clear pleiotropic phenotypic consequences to carrying the driver which seem to be absent in *D. neotestacea* (Cotton *et al.*, 2014; Reinhardt *et al.*, 2014).

This raises the question of whether the differentially expressed transcripts known to be associated with specific stages of spermatogenesis represent actual consequences of the mechanism of drive during gametogenesis, or neutral drift of expression due to divergence between the ST and SR chromosomes. The presence of DE transcripts in the earliest stage of spermatogenesis suggests that the mechanism of *sex-ratio* drive begins

much earlier than the microscopic phenotype appears (Figure S7). However, more detailed cytology is required to determine when the effects of SR are first visible during spermatogenesis.

Molecular evolutionary patterns of mechanistic candidate transcripts

We identified six candidate transcripts that showed evidence of positive selection, were differentially expressed between SR and ST, and had testis-specific expression. Two of these transcripts were revealed to be part of an X-linked duplicate of the autosomal gene *importin-α2* called *X-pendulin.* Sanger sequencing of *X-pendulin* along with a selection of other K_a/K_s candidate transcripts and potentially positively selected transcripts was performed, and the top candidates were identified as the five loci (including *X-pendulin*) with fixed differences between ST and SR.

These top candidates have striking molecular evolutionary patterns that are not consistent with neutral evolution, and suggest that on SR they are under positive selection but on ST are experiencing relaxed purifying selection. Maximum likelihood HKA tests comparing the polymorphism and divergence of the top candidates to non-candidates on the X-chromosome showed that the most likely model was the top candidates are experiencing selection on SR, but for the ST samples the neutral model is more highly supported. This is likely due to the low nucleotide diversity of the top candidates on SR. Notable also is that all five of the top candidates have fixed nucleotide differences between SR and ST, and all but one of them has no shared mutations (Table 4.2, Figure 4.4D). Previous work analyzing the variation of 11 non-candidate X-linked markers on ST and SR found not even one of them had a fixed nucleotide difference between ST and

SR (Pieper & Dyer, 2016). Involvement in drive or tight linkage to the driver is expected to result in molecular evolutionary patterns similar to a selective sweep (Derome *et al.*, 2008; Kingan *et al.*, 2010). However, if the low variation is due to a selective sweep, it is clearly not a recent one, as variation is present on SR. In fact, the average Tajima's D value of the top candidates on SR was significantly higher than at any other locus tested in ST and SR (Figure 4.4C). This is likely due to the presence of multiple segregating haplotypes on SR, as can be noted on *X-pendulin* (Figure 4.4D). The other top candidates have similar patterns, though the samples are not in perfect phase with the haplotypes observed on *X-pendulin* (Figures S8). This contrasts with previous findings that variation on SR is not maintained in as discrete haplotypes, but rather is in low linkage disequilibrium likely through recombination in SR homozygous females and occasional gene flow with ST (Pieper & Dyer, 2016). Combined with the higher divergence between ST and SR at the top candidates than at non-candidate loci (Figure S4), the defined haplotype-like patterns observed on the top candidates on SR suggests that very little recombination is occurring at these loci, though the fact that they are not in phase with one another does suggest recombination may be occurring between them. Cycling haplotypes have been observed in the autosomal *SD* system of *D. melanogaster* (Brand *et al.*, 2015).

A possible reason for this low variation is that these loci are intimately involved in the driving mechanism. The location of the top candidates on the X-chromosome is unknown, but it is likely that the inversions on SR serve to suppress recombination between the driving loci with ST. There are other reasons why recombination might be suppressed at a locus, including proximity to the centromere or inversion breakpoints, but

the fact that the top candidates also share DE between ST and SR and elevated K_a/K_s suggests that the reduction in recombination is not a coincidence of location.

The variation on ST at the top candidates is particularly interesting, as they have a higher than expected level of segregating nonsynonymous variation, which is visibly significantly elevated compared to the non-candidate loci (Figure 4.4B). This pattern suggests relaxed purifying selection on ST at these loci is failing to remove deleterious variation. One top candidate, TR261, had a sample that contained a single base pair insertion that resulted in a premature stop codon, and *X-pendulin* had a sample with an entire codon deleted (Figure 4.4D). If these loci are involved in the mechanism of drive, there is no selection to maintain them on the ST X-chromosome and they are free to accumulate mutations. There are also numerous fixed nonsynonymous differences between ST and SR in the top candidate loci, suggesting that these sequences may have divergent functions or phenotypic effects in SR vs ST (Table 4.2, Figure 4.4D).

<u>*X-pendulin* is a rapidly evolving X-linked duplicate</u>

X-pendulin is an especially strong candidate for involvement in the mechanism of drive. It is overexpressed in *sex-ratio* males, testes-specific, and highly differentiated between ST and SR (Table S3, Figure 4.4D). Our PAML analysis shows that it is clearly evolving differently from the rest of the *importin-α* gene family (Figure 4.3). Overall, the rest of the tree appears to be under quite strong purifying selection, and the d_N/d_S value of 1.03 estimated for the *X-pendulin* branch is comparatively extremely high, suggesting selection for a functional change between the two copies of the gene. PCR and sequencing of *X-pendulin* in a several different taxa shows that the duplication is specific

to the three most closely related members of the testacea species group, *D. neotestacea, D. testacea,* and *D. orientacea*, but is absent from the more distantly related *D. putrida* and the outgroup *D. bizonata* (Dyer et al. unpublished data, Dyer *et al.*, 2011).This would place the origin of the duplication between the split between the lineage leading to *D. neotestacea, D. orientacea,* and *D. testeacea* and that leading to *D. putrida* (Dyer *et al.*, 2011). In *D. melanogaster,* there are three small introns in *importin-α2*, but only one of these in included in *X-pendulin*, suggesting that a retrotransposition event involving a poorly spliced transcript resulted in the duplication (Figure S9). *Sex-ratio* meiotic drive is also present in *D. orientacea* (K. Dyer, unpublished data), suggesting that if *X-pendulin* is involved in SR drive, these two systems may have a shared mechanistic basis. Though there are only three sequences of *X-pedulin* from *D. testacea,* the diversity patterns seem to be like those on ST, with high variation (Table 4.2). The age of SR estimated with percent divergence with ST across *X-pendulin* (D_a = 0.946%) is roughly 556,000 years assuming 1.7% divergence per million years (Caccone *et al.*, 1988). This estimate is in line with previous estimates using sequence divergence in the flanking region of microsatellite markers with haplotypes associated with SR at the 95% level (Pieper & Dyer, 2016). Another of the top candidates, TR24932, has an even higher percent divergence between ST and SR of 2.054%, which would put the estimated age of SR at 1.2 million years old.

Duplication events have been implicated in the origin of multiple other meiotic drive systems. The distorter locus in the *SD* system of *D. melanogaster* is a truncated, duplicated copy of a RanGAP protein (Larracuente & Presgraves, 2012). In the Winters *sex-ratio* system of *D. simulans,* the X-linked distorter is a duplicate copy of another X-

linked gene, and an autosomal suppressor is a transposed duplicate copy of the distorter on an autosome (Tao *et al.*, 2007a; Tao *et al.*, 2007b). One of the loci linked to the Paris *sex-ratio* system of *D. simulans* is also the result of the segmental duplication spanning six genes (Fouvry *et al.*, 2011). Gene duplication is a potent force for genetic innovation as it allows new genes to evolve new functions (including to disrupt spermatogenesis) while still maintaining the original function of the gene (Lynch & Walsh, 2007).

Candidate *X-pendulin* suggests nuclear transport is the target of drive

These molecular evolutionary patterns and its status as a rapidly evolving duplicate make *X-pendulin* a good candidate for involvement in the mechanism of drive, but the function of its parent gene, *importin-α2,* makes it an excellent candidate. All importin-α proteins play a key role in the nuclear transport pathway (Goldfarb *et al.*, 2004; Stewart, 2007). Together with importin-ß*,* they form a complex that binds to cargo macromolecules carrying nuclear localization signals (NLS) and carries them into the nucleus through the nuclear pore complex (NPC). Once inside, importin-α and -ß dissociate and are exported back into the cytoplasm through the NPC (Goldfarb *et al.*, 2004; Stewart, 2007). This cycle of import and export is powered by the GTP gradient maintained between the cytoplasm and the nucleus – in the nucleus, RanGTP binds to importin-ß and triggers dissociation with importin-α, leading to export back out into the cytoplasm where RanGAP hydrolyzes the GTP and dissociates the export complex, freeing importin-α and importin-ß to reassociate with cargo and begin the cycle again (Goldfarb *et al.*, 2004; Stewart, 2007).

Nuclear transport has been previously implicated in genetic conflict in the male germline, most prominently in the autosomal *SD* meiotic driver of *D. melanogaster* (Larracuente & Presgraves, 2012). During spermatogenesis in *Drosophila,* the haploid nuclei of sperm share a cytoplasm until the individualization stage, which is when half of the sperm in *SD* males die (Fuller, 1993; de Cuevas *et al.*, 1997; Larracuente & Presgraves, 2012). In this system, the distorter locus (*sd*) is a truncated, duplicated copy of *RanGAP* that mislocalizes to the nucleus instead of the cytoplasm (Kusano *et al.*, 2001; Larracuente & Presgraves, 2012). Though the exact mechanism is unknown, *sd* is enzymatically active and causes selective failure of sperm carrying the sensitive allele at the responder locus through disruption of the GTP-gradient required for nuclear transport (Kusano *et al.*, 2002; Larracuente & Presgraves, 2012).

Of the three canonical *importin-α* genes in *D. melanogaster, importin-α2* is primarily expressed in the testes and plays a critical role in spermatogenesis (Mason *et al.*, 2002). Male homozygous null *importin-α2* flies are generally completely sterile, and their testes contain no motile sperm (Mason *et al.*, 2002). Though *importin-α2* is expressed primarily in the very early stages of spermatogenesis in the primary and secondary spermatocytes, in null homozygous males the sperm fail at the individualization checkpoint (Giarrè *et al.*, 2002; Mason *et al.*, 2002). As introducing overexpression of *importin-α1* and *3* in the testes can rescue this phenotype, *importin-α2* may bind a specific cargo that is required in the nucleus for spermatogenesis to proceed correctly past the individualization checkpoint (Mason *et al.*, 2002).

The PAML analysis of all members of the *importin-α* family in *D. neotestacea* shows that the autosomal copy of *importin-α2* is under purifying selection and so likely

retains the ancestral function as observed in *D. melanogaster* (Figure 4.3). Our gene expression analysis also shows that the autosomal copy of *importin-α2* is expressed at a high level in both the testes of ST and SR males, with no difference between them. The knockout phenotype of *importin-α2* in *D. melanogaster* is the failure of spermatogenesis during individualization, which is very similar to what is observed in SR males in *D. neotestacea* (Figure S7) (Mason *et al.*, 2002). Therefore, if *X-pendulin* contributes to drive, it must do so in the presence of the wild-type protein.

We can think of several potential models for how this might occur. In the first model, *X-pendulin* has a dominant negative effect by competitively binding with the normal interaction partners of *importin-α2,* but cannot successfully deliver cargo to the nuclei of Y-bearing sperm, causing their death only (Figure 4.5a). In the example pictured in Figure 4.5a, X-pendulin binds to the cargo and importin-ß, but is unable to enter the wild-type NPC present in the nuclear envelope of Y-bearing sperm. The nuclei of X-bearing sperm carry a mutant copy of the NPC, which is encoded by SR, that allows X-pendulin to pass through successfully and deliver the necessary cargo to the nucleus (Figure 4.5a). The key to this model is that X-pendulin prevents importin-α2 from successfully importing cargo to the nucleus of Y-bearing sperm, but there is a haploid specific rescue of X-bearing sperm. A second model for the mechanism by which *X-pendulin* could cause the selective destruction of Y-bearing sperm is the delivery of a "poison" or some factor that causes the sperm to die (Figure 4.5b). In this model, the autosomal copy of importin-α2 successfully delivers its cargo to both the Y- and X-bearing sperm, but X-pendulin specifically targets Y-bearing sperm with some additional

factor. Again, multiple interacting components are required for drive to occur in either of these scenarios.

Because the sperm develop in bundles and share a cytoplasm until the individualization stage, the other part of the SR-drive system that rescues X-bearing sperm from destruction is more likely to remain associated with the nuclear envelope or inside the nucleus (Fuller, 1993; de Cuevas *et al.*, 1997; Burt & Trivers, 2006). However, there are many different candidates for interaction partners with *X-pendulin* and it is impossible to say what other components of the nuclear import/export pathway are involved given the current data. A review of several known interacting partners of *importin-α2* in the current dataset (*importin-ß*, *CAS, Ran, RanGTP, RanGEF, Nup153, Nup50*) found no indications of DE between ST and SR or putatively positive selection (data not shown). Of course, the other factors required for drive and X-linked rescue may not even be proteins, as the responder locus of *SD* is a satellite repeat whose copy number confers sensitivity to drive (Larracuente & Presgraves, 2012).

The duplication of an *importin-α* is not unique to the testacea group. At least three independent duplication events of *importin-α2* and *α3* have been identified in other *Drosophila* lineages, and the loss and gain of noncanonical *importin-α*s seems to be relatively common (Phadnis *et al.*, 2011). Like *X-pendulin*, these duplicates are expressed primarily in the testes, and at least one of them has signatures of positive selection in certain lineages (Phadnis *et al.*, 2011). Genes primarily expressed in the testes and during spermatogenesis are well known to be fast-evolving, and new duplicate genes often have testis-specific expression patterns (Betrán *et al.*, 2002; Emerson *et al.*, 2004; Haerty *et al.*, 2007). This is often attributed to sexual selection, but another hypothesis may be that

spermatogenesis is frequently the site of genetic conflict, and there is rapid co-evolution with driving elements to restore fair meiosis (Kleene, 2005; Haerty *et al.*, 2007). Phadnis et al. suggest that the repeated duplication and evolution of new *importin-α* family members is due to evolutionary pressure to maintain wild-type nuclear import against selfish genetic elements like *SD* (2011). This scenario seems unlikely in the case of *X-pendulin* in *D. neotestacea.* While it is possible that the original duplication onto the X-chromosome became fixed as a guard against autosomal meiotic drivers, *X-pendulin* cannot be a suppressor of *sex-ratio* drive in the current system because there would never be selection for a suppressor on the wild-type X-chromosome, as it never comes into conflict with *sex-ratio.*

Altogether, this study presents a compelling case that the X-linked duplicate of *importin-α2* called *X-pendulin* is part of the causal mechanism of *sex-ratio* drive in *D. neotestacea.* The expression difference combined with molecular evolutionary evidence of positive selection in SR and relaxed purifying selection on ST strongly suggest that *X-pendulin* is active in SR but not ST, and the phenotype of the parent gene suggests several potential mechanisms for how drive could occur (Figure 4.5). Nevertheless, more direct functional evidence is required to confirm that *X-pendulin* is involved in drive. Transformations knocking out *X-pendulin* expression in SR and inducing overexpression in ST should be performed, as well as localization experiments to show whether *X-pendulin* differentially enters the nuclei of X and Y-bearing sperm in the testes of SR males. More work also needs to be done to identify the interacting partners of *X-pendulin;* at least one other element must be required for drive to occur, and complex inversions on SR suggest there may be even more. Transformations have never been done

in *D. neotestacea* nor any closely related species, but the confirmation of *X-pendulin* as a distorter locus requires it. It will both provide a contrast with the independently evolved mechanisms of *sex-ratio* drive in *D. simulans* (Tao *et al.*, 2007a; Tao *et al.*, 2007b; Helleu *et al.*, 2016) and show convergent evolution with SD in *D. melanogaster* of manipulation of the nuclear import/export pathway to cause meiotic drive during spermatogenesis (Larracuente & Presgraves, 2012). Indeed, the rapid sequence evolution of RanGAP and other associated genes in *D. melanogaster* suggests a long history of genetic conflict utilizing nuclear transport (Presgraves, 2007). These experiments in *D. neotestacea* will provide critical evidence that certain pathways and processes are particularly susceptible to genetic conflict.

Acknowledgements

This project was funded by a Rosemary Grant Award from the Society for the Study of Evolution, a Presidential Fellowship from the University of Georgia, and the National Institutes of Health (NIH) under award T32GM007103 to KEP, and National Science Foundation grants 1149350 and 1457707 and funds from the University of Georgia to KAD. We thank Yasir Ahmed-Braimah for providing us with the *D. virilis* genome assembly.

References

Anders, S. & Huber, W. 2012. Differential expression of RNA-Seq data at the gene level–the DESeq package. *Heidelberg, Germany: European Molecular Biology Laboratory (EMBL)*.

Bauer, H., Willert, J.R., Koschorz, B. & Herrmann, B.G. 2005. The t complex-encoded GTPase-activating protein Tagap1 acts as a transmission ratio distorter in mice. *Nature Genetics* **37:** 969-973.

Bauer, H., Véron, N., Willert, J. & Herrmann, B.G. 2007. The t-complex-encoded guanine nucleotide exchange factor Fgd2 reveals that two opposing signaling pathways promote transmission ratio distortion in the mouse. *Genes & Development* **21:** 143-147.

Benjamini, Y. & Hochberg, Y. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Sco. Series B Stat. Methodol.***:** 289-300.

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

Brand, C.L., Larracuente, A.M. & Presgraves, D.C. 2015. Origin, evolution, and population genetics of the selfish Segregation Distorter gene duplication in European and African populations of *Drosophila melanogaster*. *Evolution* **69:** 1271-1283.

Burt, A. & Trivers, R. 2006. *Genes in Conflict: The biology of selfish genetic elements*. Belknap Press of Harvard University Press, Cambridge, Mass.

Caccone, A., Amato, G.D. & Powell, J.R. 1988. Rates and patterns of scnDNA and mtDNA divergence within the *Drosophila melanogaster* subgroup. *Genetics* **118:** 671-683.

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K. & Madden, T.L. 2009. BLAST+: architecture and applications. *BMC Bioinformatics* **10:** 421.

Carvalho, A.B. & Vaz, S.C. 1999. Are Drosophila SR drive chromosomes always balanced? *Heredity* **83:** 221-228.

Charlesworth, B. & Hartl, D.L. 1978. Population-dynamics of segregation distorter polymorphism of *Drosophila melanogaster. Genetics* **89:** 171-192.

Christianson, S.J., Brand, C.L. & Wilkinson, G.S. 2011. Reduced polymorphism associated with X chromosome meiotic drive in the stalk-eyed fly *Teleopsis dalmanni*. *PLoS ONE* **6:** 1-7.

Consortium, T.U. 2014. UniProt: a hub for protein information. *Nucleic Acids Research* **43:** D204-D212.

Cotton, A.J., Foldvari, M., Cotton, S. & Pomiankowski, A. 2014. Male eyespan size is associated with meiotic drive in wild stalk-eyed flies (*Teleopsis dalmanni*). *Heredity* **112:** 363-369.

Danecek, P., Auton, A., Abecasis, G., Albers, C.A., Banks, E., DePristo, M.A., Handsaker, R.E., Lunter, G., Marth, G.T., Sherry, S.T., McVean, G. & Durbin, R. 2011. The variant call format and VCFtools. *Bioinformatics* **27:** 2156-2158.

de Cuevas, M., Lilly, M.A. & Spradling, A.C. 1997. Germline cyst formation in Drosophila. *Annual Review of Genetics* **31:** 405-428.

Derome, N., Baudry, E., Ogereau, D., Veuille, M. & Montchamp-Moreau, C. 2008. Selective sweeps in a 2-locus model for sex-ratio meiotic drive in *Drosophila simulans*. *Mol. Biol. Evol.* **25:** 409-416.

Dyer, K.A., White, B.E., Bray, M.J., Pique, D.G. & Betancourt, A.J. 2011. Molecular evolution of a Y chromosome to autosome gene duplication in Drosophila. *Mol. Biol. Evol.* **28:** 1293-1306.

Dyer, K.A. 2012. Local selection underlies the geographic distribution of sex-ratio drive in *Drosophila neotestacea*. *Evolution* **66:** 973-984.

Dyer, K.A., Bray, M.J. & Lopez, S.J. 2013. Genomic conflict drives patterns of X-linked population structure in *Drosophila neotestacea*. *Mol. Ecol.* **22:** 157-169.

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

Fouvry, L., Ogereau, D., Berger, A., Gavory, F. & Montchamp-Moreau, C. 2011. Sequence Analysis of the Segmental Duplication Responsible for Paris Sex-Ratio Drive in Drosophila simulans. *G3: Genes, Genomes, Genetics* **1:** 401-410.

Fuller, M.T. 1993. *Spermatogenesis*. Cold Spring Harbor Laboratory Press {a}, 10 Skyline Drive, Plainview, New York 11803, USA.

Giarrè, M., Török, I., Schmitt, R., Gorjánácz, M., Kiss, I. & Mechler, B.M. 2002. Patterns of importin-α expression during Drosophila spermatogenesis. *Journal of Structural Biology* **140:** 279-290.

Goldfarb, D.S., Corbett, A.H., Mason, D.A., Harreman, M.T. & Adam, S.A. 2004. Importin α: a multipurpose nuclear-transport receptor. *Trends in Cell Biology* **14:** 505-514.

Haas, B.J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P.D., Bowden, J., Couger, M.B., Eccles, D., Li, B., Lieber, M., Macmanes, M.D., Ott, M., Orvis, J., Pochet, N., Strozzi, F., Weeks, N., Westerman, R., William, T., Dewey, C.N.,

Henschel, R., Leduc, R.D., Friedman, N. & Regev, A. 2013. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat Protoc* **8:** 1494-1512.

Haerty, W., Jagadeeshan, S., Kulathinal, R.J., Wong, A., Ram, K.R., Sirot, L.K., Levesque, L., Artieri, C.G., Wolfner, M.F. & Civetta, A. 2007. Evolution in the fast lane: rapidly evolving sex-related genes in Drosophila. *Genetics* **177:** 1321-1335.

Hamilton, W.D. 1967. Extraordinary sex ratios. A sex-ratio theory for sex linkage and inbreeding has new implications in cytogenetics and entomology. *Science* **156:** 477-488.

Harrison, P.W., Wright, A.E. & Mank, J.E. 2012. The evolution of gene expression and the transcriptome–phenotype relationship. *Seminars in cell & developmental biology* **22:** 222-229. Academic Press.

Helleu, Q., Gérard, P.R., Dubruille, R., Ogereau, D., Prud'homme, B., Loppin, B. & Montchamp-Moreau, C. 2016. Rapid evolution of a Y-chromosome heterochromatin protein underlies sex chromosome meiotic drive. *Proc. Natl. Acad. Sci.* **113:** 4110-4115.

Herrmann, B.G., Koschorz, B., Wertz, K., McLaughlin, K.J. & Kispert, A. 1999. A protein kinase encoded by the t complex responder gene causes non-mendelian inheritance. *Nature* **402:** 141-146.

Hurst, L.D. & Pomiankowski, A. 1991. Causes of sex-ratio bias may account for unisexual sterility in hybrids - a new explanation of Haldane's rule and related phenomena. *Genetics* **128:** 841-858.

Jaenike, J. 2001. Sex chromosome meiotic drive. *Annu. Rev. Ecol. Syst.* **32:** 25-49.

James, A.C. & Jaenike, J. 1990. Sex-ratio meiotic drive in *Drosophila testacea*. *Genetics* **126:** 651-656.

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., Thierer, T., Ashton, B., Meintjes, P. & Drummond, A. 2012. Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **28:** 1647-1649.

Kingan, S.B., Garrigan, D. & Hartl, D.L. 2010. Recurrent selection on the Winters sex-ratio genes in *Drosophila simulans*. *Genetics* **184:** 253-265.

Kleene, K.C. 2005. Sexual selection, genetic conflict, selfish genes, and the atypical patterns of gene expression in spermatogenic cells. *Developmental Biology* **277:** 16-26.

Kusano, A., Staber, C. & Ganetzky, B. 2001. Nuclear mislocalization of enzymatically active RanGAP causes segregation distortion in Drosophila. *Developmental cell* **1:** 351-361.

Kusano, A., Staber, C. & Ganetzky, B. 2002. Segregation distortion induced by wild-type RanGAP in Drosophila. *Proc. Natl. Acad. Sci.* **99:** 6866-6870.

Langmead, B. & Salzberg, S.L. 2012. Fast gapped-read alignment with Bowtie 2. *Nat Meth* **9:** 357-359.

Larracuente, A.M. & Presgraves, D.C. 2012. The selfish Segregation Distorter gene complex of Drosophila melanogaster. *Genetics* **192:** 33-53.

Li, B. & Dewey, C.N. 2011. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* **12:** 323.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G. & Durbin, R. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25:** 2078-2079.

Librado, P. & Rozas, J. 2009. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* **25:** 1451-1452.

Lindholm, A.K., Dyer, K.A., Firman, R.C., Fishman, L., Forstmeier, W., Holman, L., Johannesson, H., Knief, U., Kokko, H., Larracuente, A.M., Manser, A., Montchamp-Moreau, C., Petrosyan, V.G., Pomiankowski, A., Presgraves, D.C., Safronova, L.D., Sutter, A., Unckless, R.L., Verspoor, R.L., Wedell, N., Wilkinson, G.S. & Price, T.A.R. 2016. The Ecology and Evolutionary Dynamics of Meiotic Drive. *Trends in Ecology & Evolution* **31:** 315-326.

Love, M.I., Huber, W. & Anders, S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome biology* **15:** 1.

Lynch, M. & Walsh, B. 2007. *The origins of genome architecture*. Sinauer Associates Sunderland.

Lyon, M.F. 1991. The genetic basis of transmission-ratio distortion and male sterility due to the t complex. *Am. Nat.* **137:** 349-358.

Lyon, M.F. 2003. Transmission ratio distortion in mice. *Annual Review of Genetics* **37:** 393-408.

Mason, D.A., Fleming, R.J. & Goldfarb, D.S. 2002. *Drosophila melanogaster* importin α1 and α3 can replace importin α2 during spermatogenesis but not oogenesis. *Genetics* **161:** 157-170.

Merrill, C., Bayraktaroglu, L., Kusano, A. & Ganetzky, B. 1999. Truncated RanGAP encoded by the Segregation Distorter locus of Drosophila. *Science* **283:** 1742-1745.

Phadnis, N., Hsieh, E. & Malik, H.S. 2011. Birth, death and replacement of karyopherins in Drosophila. *Mol. Biol. Evol.***:** msr306.

Pieper, K.E. & Dyer, K.A. 2016. Occasional recombination of a selfish X-chromosome may permit its persistence at high frequencies in the wild. *Journal of Evolutionary Biology* **29:** 2229-2241.

Presgraves, D.C. 2007. Does genetic conflict drive rapid molecular evolution of nuclear transport genes in Drosophila? *BioEssays* **29:** 386-391.

Quinlan, A.R. & Hall, I.M. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26:** 841-842.

RCoreTeam 2014. R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing. Vienna, Austria.

Reinhardt, J.A., Brand, C.L., Paczolt, K.A., Johns, P.M., Baker, R.H. & Wilkinson, G.S. 2014. Meiotic Drive Impacts Expression and Evolution of X-Linked Genes in Stalk-Eyed Flies. *PLOS Genetics* **10:** 14.

Rice, W.R. 2013. Nothing in genetics makes sense except in light of genomic conflict. *Annu. Rev. Ecol. Syst.* **44:** 217-237.

Risso, D., Ngai, J., Speed, T.P. & Dudoit, S. 2014. Normalization of RNA-seq data using factor analysis of control genes or samples. *Nature biotechnology* **32:** 896-902.

Robinson, M.D., McCarthy, D.J. & Smyth, G.K. 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26:** 139-140.

Sandler, L. & Novitski, E. 1957. Meiotic drive as an evolutionary force. *Am. Nat.* **91:** 105-110.

Schaeffer, S.W., Bhutkar, A.U., McAllister, B.F., Matsuda, M., Matzkin, L.M., O'Grady, P.M., Rohde, C., Valente, V.L.S., Aguadé, M., Anderson, W.W., Edwards, K., Garcia, A.C.L., Goodman, J., Hartigan, J., Kataoka, E., Lapoint, R.T., Lozovsky, E.R., Machado, C.A., Noor, M.A.F., Papaceit, M., Reed, L.K., Richards, S., Rieger, T.T., Russo, S.M., Sato, H., Segarra, C., Smith, D.R., Smith, T.F., Strelets, V., Tobari, Y.N., Tomimura, Y., Wasserman, M., Watts, T., Wilson, R., Yoshida, K., Markow, T.A., Gelbart, W.M. & Kaufman, T.C. 2008. Polytene chromosomal maps of 11 Drosophila species: the order of genomic scaffolds inferred from genetic and physical maps. *Genetics* **179:** 1601-1655.

Stewart, M. 2007. Molecular mechanism of the nuclear protein import cycle. *Nature Reviews Molecular Cell Biology* **8:** 195-208.

Tao, Y., Araripe, L., Kingan, S.B., Ke, Y., Xiao, H. & Hartl, D.L. 2007a. A sex-ratio meiotic drive system in *Drosophila simulans*. II: an X-linked distorter. *PLoS Biol* **5:** e293.

Tao, Y., Masly, J.P., Araripe, L., Ke, Y. & Hartl, D.L. 2007b. A sex-ratio meiotic drive system in *Drosophila simulans*. I: an autosomal suppressor. *PLoS Biol* **5:** e292.

Untergasser, A., Cutcutache, I., Koressaar, T., Ye, J., Faircloth, B.C., Remm, M. & Rozen, S.G. 2012. Primer3—new capabilities and interfaces. *Nucleic Acids Research* **40:** e115-e115.

Van der Auwera, G.A., Carneiro, M.O., Hartl, C., Poplin, R., Del Angel, G., Levy-Moonshine, A., Jordan, T., Shakir, K., Roazen, D., Thibault, J., Banks, E., Garimella, K.V., Altshuler, D., Gabriel, S. & DePristo, M.A. 2013. From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. *Curr Protoc Bioinformatics* **11:** 11 10 11-11 10 33.

Vibranovski, M.D., Lopes, H.F., Karr, T.L. & Long, M. 2009. Stage-Specific Expression Profiling of Drosophila Spermatogenesis Suggests that Meiotic Sex Chromosome Inactivation Drives Genomic Relocation of Testis-Expressed Genes. *PLOS Genetics* **5:** e1000731.

Wright, S.I. & Charlesworth, B. 2004. The HKA test revisited. *Genetics* **168:** 1071-1076.

Wu, C.-I., Lyttle, T.W., Wu, M.-L. & Lin, G.-F. 1988. Association between a satellite DNA sequence and the Responder of Segregation Distorter in *D. melanogaster*. *Cell* **54:** 179-189.

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

Zhang, Z., Li, J., Zhao, X.Q., Wang, J., Wong, G.K. & Yu, J. 2006. KaKs_Calculator: calculating Ka and Ks through model selection and model averaging. *Genomics Proteomics Bioinformatics* **4:** 259-263.

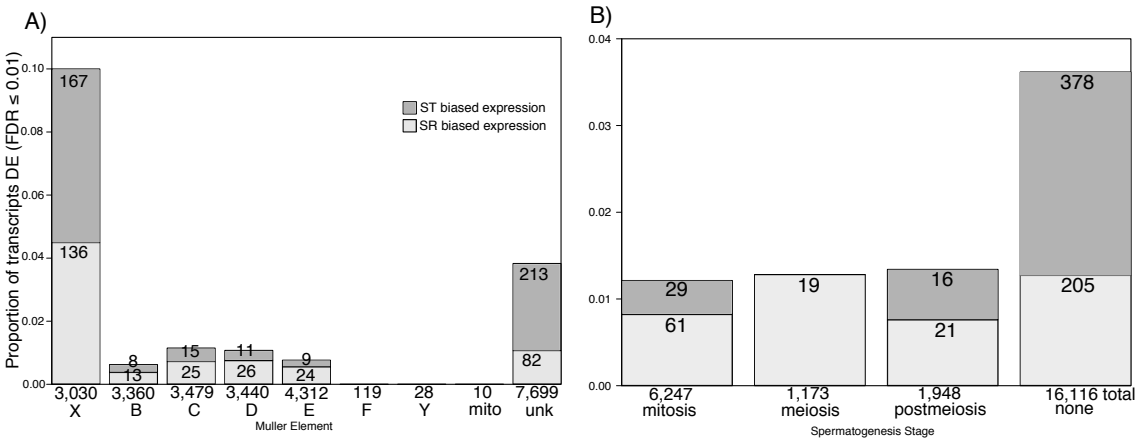


Figure 4.1. Differentially expressed (DE) transcripts are enriched on the X-chromosome and evenly distributed throughout spermatogenesis. A) The distribution of DE transcripts across the chromosomes. The total number of transcripts that mapped to each chromosome is listed beneath the bars. The large Muller elements are X, B, C, D, and E. F is the small, non-recombining dot chromosome, mito is the mitochondria, and unk stands for unknown location. Dark grey is the proportion of ST biased transcripts, and light grey is the proportion of SR biased transcripts. The number of DE transcripts in each category is printed on the bar graph. B) The distribution of DE transcripts annotated as expressed in specific stages of spermatogenesis. The Y-axis is the same as in A. The total number of transcripts mapped to each stage is listed beneath the bars. Dark grey is the proportion of ST biased transcripts, and light grey is the proportion of SR biased transcripts. The exact number of DE transcripts in each category is printed on the bar graph.

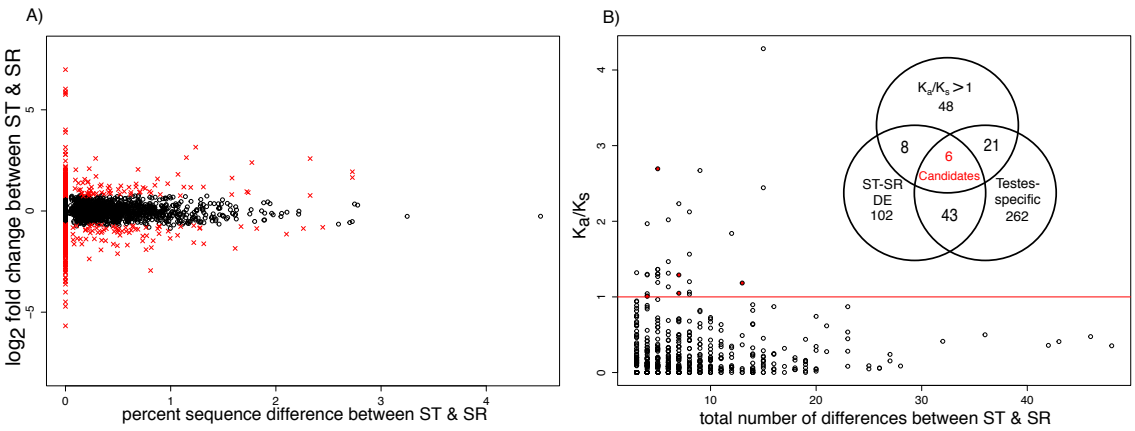


Figure 4.2. Sequence differences between ST and SR in transcripts. A) There is no relationship between differential expression between ST and SR and sequence differences between ST and SR. Percent sequence difference is calculated as the total number of differences divided by the length of the transcript times 100. Transcripts with significant differential expression are marked in red. Only transcripts that met the minimum coverage criteria for detecting sequence differences are included. The dashed line at 0.56% represents the average percent of nucleotides different within a large population genetic sample of ST males at five X-linked genes. The mean percent different of transcripts with nucleotide differences was 0.53%. B) Transcripts with the highest number of differences are not the same as those with the highest K_a/K_s values. Only transcripts with more than three differences and at least one synonymous difference are included in the figure. The inset Venn diagram shows the criteria used to identify K_a/K_s candidates and the number of transcripts in each category. Transcripts with more than three differences but no synonymous differences were also included in the $K_a/K_s > 1$ set. The identified candidates are marked in red; one of these had no synonymous differences and is not pictured.

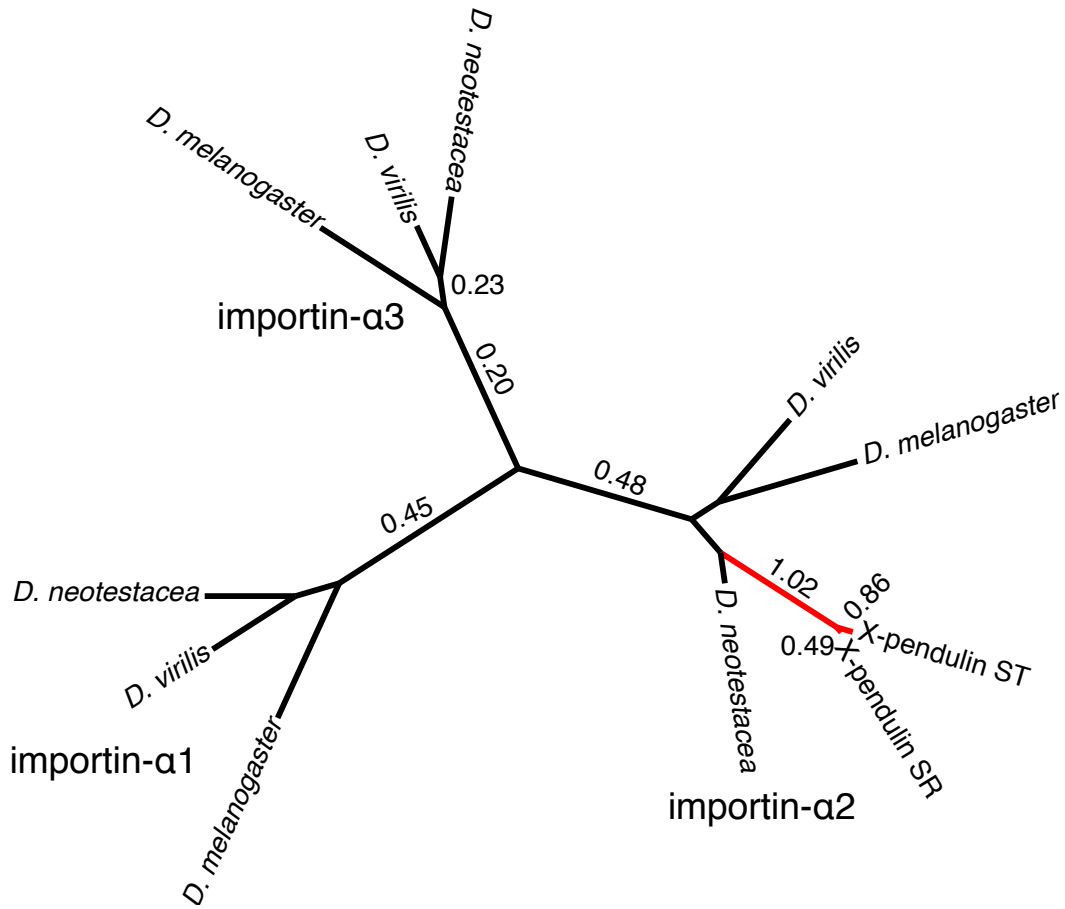


Figure 4.3. *X-pendulin* is a fast-evolving X-linked duplicate of the autosomal gene *importin-α2*. Branch lengths were estimated with a neighbor-joining tree of 1,000 bootstraps. All nodes had bootstrap support > 99. Each of the three *importin-α* clades is labeled. Estimated d_N/d_S values larger than 0.1 are labeled, and branches with d_N/d_S values > 0.60 are marked in red.

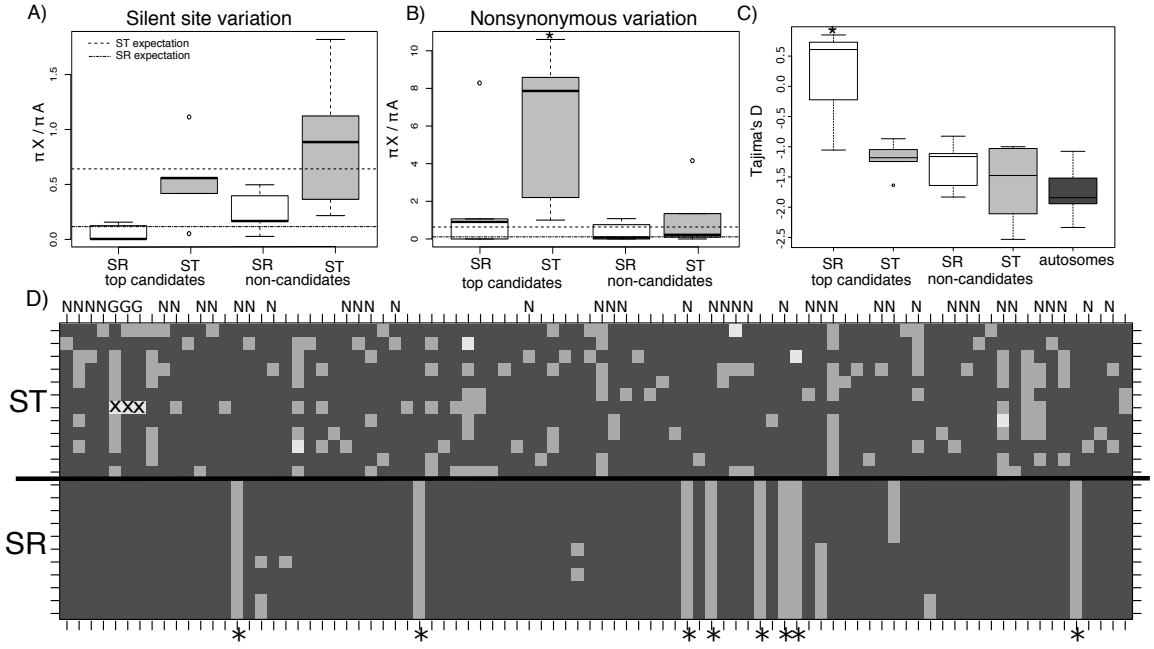


Figure 4.4. Molecular evolutionary patterns of top candidates are consistent with positive selection on SR and relaxed purifying selection on ST. A) The ratio of silent site variation on the X-chromosome to the autosomes on ST and SR is in line with neutral expectations in both the top candidate and non-candidate markers. Silent sites include synonymous sites as well as non-coding sites. SR is represented by white boxes and ST by grey boxes. The horizontal dashed lines represent the expected X/A ratio of 0.85 for ST and 0.15 for SR (see text for details). The edges of the boxes are the first and third quartiles, and the dark grey line is the median. The whiskers extend to the last data point before 1.5x the interquartile range. B) The top candidates have significantly higher non-synonymous variation than the expected ratio with the autosomes. The star denotes statistical significance at $p < 0.05$, two-tailed t-test from the expectation. Colors, boxes, and dashed lines are as described in A. C) Tajima's D is elevated in the top candidates on SR. The star denotes significance at $p < 0.05$, nested ANOVA comparing marker type and location. The darkest grey represents the autosomes, otherwise colors are as described in A. D) Haplotype structure of *X-pendulin.* Each row is a chromosome, with ST phenotype males above the solid black line and SR phenotype males below. Each column is a single segregating site. Dark grey represents the individual carries the major allele, and light grey is the minor allele. Some sites have a third segregating allele, which is represented by white. Sites with a gap are marked with an X. Non-synonymous sites are denoted with an N, deletion polymorphisms are denoted with a G. Unlabeled sites are either synonymous or non-coding. Fixed differences between ST and SR are marked with a star.

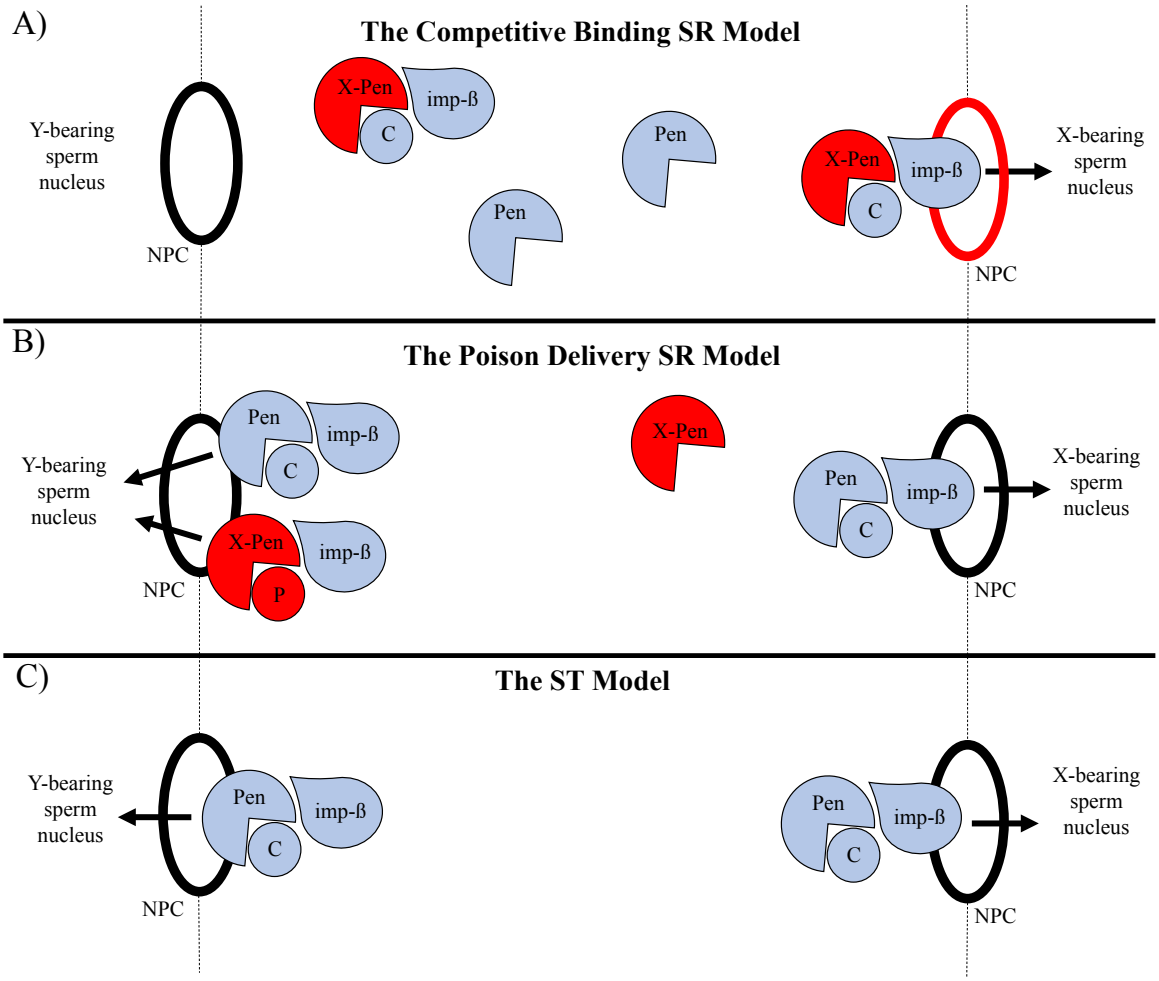


Figure 4.5. A cartoon showing two possible mechanisms by which *X-pendulin* could contribute to meiotic drive during spermatogenesis in SR males. Pictured in each panel is the shared cytoplasm of developing spermatids, with the nucleus and nuclear envelope of a Y-bearing sperm on the left and the nucleus and nuclear envelope of an X-bearing sperm on the right. *Pendulin* binds to its cargo (C) and *importin-ß* to pass through the nuclear pore complex (NPC) and deliver the cargo to the nucleus. A) The competitive binding SR model. In SR males, *X-pendulin* could prevent the transport of necessary cargo into the nuclei of Y-bearing sperm. *X-pendulin* (in red) outcompetes autosomal *Pendulin* in binding cargo, but is unable to enter the NPC of the Y-bearing sperm, which causes their death. Other components of the driving system (depicted here as a variant NPC in red) allow the successful transport of *X-pendulin* and cargo into the nucleus of X-bearing sperm, allowing these sperm to develop successfully. B) The poison delivery SR model. In SR males, *X-pendulin* transports a "poison" (P) only into the nuclei of Y-bearing sperm, which causes spermatogenesis to fail. The autosomal *Pendulin* functions correctly in this scenario, carrying its normal cargo into the nuclei of both kinds of sperm. Lacking the poison, X-bearing sperm develop successfully. C) The ST model. In ST males, *Pendulin* successfully delivers its cargo to the nuclei of both kinds of sperm.

Table 4.1. Chi-squared tests examining enrichment of positively selected transcripts for a) testes-specific expression, b) involvement in spermatogenesis, and c) differential expression between ST and SR. Each test had one degree of freedom. Values bigger than expected are bolded.

a)	$K_a/K_s > 1$	$K_a/K_s < 1$	χ^2	p-value
testes specific	**21**	241	10.383	0.001
not testes specific	27	829		
b)				
spermatogenesis	12	525	9.7164	0.002
not spermatogenesis	**36**	545		
c)				
DE	8	94	2.4073	0.121
not DE	40	976		

Table 4.2. Population genetic summary statistics of sequenced loci. Top candidates for involvement in the driving mechanism are marked. Silent sites include noncoding and synonymous sites. NS stands for nonsynonymous. N is number of samples, S is segregating sites, NC fixed is the number of fixed non-coding sites between ST and SR, S fixed is the number of fixed synonymous sites, and NS fixed is the number of fixed nonsynonymous sites. D_a is the percent divergence between ST and SR. D is Tajima's D. K_{ST} and S_{nn} are calculated between ST and SR.

Top candidate	transcript	Sites	Silent sites	NS sites	X-type	N	S	Silent π	NS π	π_a/π_s	Fixed differences	Shared mutations	D_a	NC fixed	S fixed	NS fixed	Silent D	NS D	K_{ST}	S_{nn}
yes	x-pendulin	1277	339.92	933.08	ST	12	71	0.026	0.010	0.383	8	0	0.946	0	4	4	-1.185	-0.962	0.372	1.000
					SR	11	6	0.004	0.001	0.322							-0.249	1.828		
		1262	345.11	916.89	*D. testacea*	3	43	0.052	0.012	0.223										
yes	TR261	672	269.22	377.78	ST	10	25	0.010	0.014	1.387	2	8	0.944	0	0	2	-0.782	-0.467	0.336	0.972
					SR	8	10	0.003	0.011	3.709							0.069	0.998		
yes	TR11103	826	246.78	573.22	ST	9	16	0.013	0.003	0.215	1	0	0.259	1	0	0	-1.280	0.385	0.330	0.947
					SR	10	0	0.000	0.000	0.000							NA	NA		
yes	TR24932	280	93.08	182.92	ST	8	10	0.013	0.011	0.834	4	0	2.054	0	0	4	-1.595	-1.477	0.674	1.000
					SR	8	1	0.000	0.001	0.000							NA	-1.055		
yes	TR37304	770	175.33	592.67	ST	10	4	0.001	0.001	1.123	4	0	0.600	0	1	3	-1.112	-1.034	0.836	1.000
					SR	10	0	0.000	0.000	0.000							NA	NA		
	TR50351	649	268.02	374.98	ST	10	50	0.031	0.018	0.593	0	18	0.408	0	0	0	-1.053	-0.248	0.091	0.917
					SR	9	37	0.022	0.017	0.753							-0.747	-0.033		
	TR23125	343	67	272	ST	9	8	0.000	0.010	0.000	0	2	0.373	0	0	0	NA	-0.256	0.189	0.763
					SR	10	6	0.008	0.009	1.141							1.303	1.586		

CHAPTER 5

DISCUSSION AND FUTURE DIRECTIONS

In this dissertation, I use the *sex-ratio* X-chromosome (SR) of *Drosophila neotestacea* to investigate the evolutionary consequences and molecular mechanisms of meiotic drive. In this North American fruit fly, SR persists at high stable frequencies in wild populations and drives at an extremely high rate with no known suppressors (James & Jaenike, 1990; Dyer, 2012; Dyer *et al.*, 2013). The suggestion of multiple inversions on SR suggests that the mechanism of drive requires more than one locus, potentially indicating a long history of coevolution with suppressors (Meiklejohn & Tao, 2010; Dyer *et al.*, 2013). Together, these properties suggest that SR may have a significant impact on X-chromosome evolution in this species.

In Chapter 3, I investigated the molecular evolutionary history and consequences of SR using sequencing of a population genetic sample at 11 X-linked loci that span the ST X-chromosome and at seven arbitrarily chosen autosomal loci. The evolutionary relationship between ST and SR varies among individual markers, but genetic differentiation between SR and ST is chromosome-wide. I confirmed the presence of large chromosomal inversions, which likely suppress recombination. However, I did find evidence for gene flow between ST and SR at every locus tested, and the lack of long-distance linkage disequilibrium within SR chromosomes suggests that recombination occurs in SR homozygous females. Finally, polymorphism on SR is reduced compared to ST, and loci displaying signatures of selection on ST do not show similar patterns on SR.

These results suggest that gene flow with ST and recombination between SR chromosomes may prevent the accumulation of deleterious mutations and allow its long-term persistence at relatively high frequencies. Importantly, the shared variation between ST and SR at these arbitrarily chosen loci suggests that loci that are directly involved in drive will display molecular evolutionary patterns consistent with a selective sweep in SR and a complete lack of recombination with ST.

In Chapter 4, I use a combination of transcriptomics and population genetics to identify candidate transcripts for the causal driving loci. In comparing the transcript expression profiles from the testes of ST and SR males, I found over 700 differentially expressed transcripts and that the X-chromosome is enriched for differentially expressed transcripts. I also found differential expression begins early in spermatogenesis. I identified nucleotide sequence differences between ST and SR, and found many transcripts with K_a/K_s values suggestive of positive selection. I identified a set of differentially expressed, testis-specific candidate transcripts under putative positive selection, including an X-linked duplicate of the nuclear transport gene *importin-α2.* This duplicate, called *X-pendulin,* is specific to the testacea species group. I show that positive selection has occurred on the lineage leading to the duplicate. A population genetics analysis of *X-pendulin* and the other candidates show their molecular evolutionary patterns are consistent with relaxed purifying selection in ST and positive selection in SR. I suggest that *X-pendulin* is part of the mechanism of SR drive and present several possible molecular models for its activity.

Overall, my dissertation presents a thorough investigation of the evolutionary properties and potential mechanisms of sex chromosome meiotic drive in *D. neotestacea,*

and sets the stage for more detailed study in this system. Of particular interest is the identification of *X-pendulin* as a strong candidate for the mechanism of drive. There are many independent male meiotic drive systems known, but in only four of them has the mechanistic basis of drive been identified (Herrmann *et al.*, 1999; Bauer *et al.*, 2005; Bauer *et al.*, 2007; Tao *et al.*, 2007a; Tao *et al.*, 2007b; Larracuente & Presgraves, 2012; Helleu *et al.*, 2016). Of these four, the distorter of the SD system of *D. melanogaster* involves the same nuclear import pathway as *importin-α2* (Larracuente & Presgraves, 2012). In SD, the distorter (*sd*) is a duplicated, truncated copy of RanGAP that mislocalizes to the nucleus instead of the cytoplasm (Kusano *et al.*, 2001). Normally, the GTP gradient between the nucleus and the cytoplasm powers the import and export of macromolecules into the nucleus, but the action of *sd* disrupts this process and causes the death of sperm carrying the sensitive allele of the responder (Kusano *et al.*, 2002). Interestingly, the responder (*Rsp*) is an array of satellite repeats, and sensitivity or resistance to drive corresponds to the number of repeats present (Wu *et al.*, 1988). It is still unknown how *sd* and *rsp* interact to cause drive (Larracuente & Presgraves, 2012).

In wild-type testes, *importin-α2* relies on the GTP-GDP gradient to carry cargo from the cytoplasm into the nucleus (Goldfarb *et al.*, 2004). The *importin-α* family of proteins binds to *importin-ß* to carry proteins marked with nuclear localization signals into the nucleus (Goldfarb *et al.*, 2004). *Importin-α2* is primarily expressed in the gonads, and causes sterility in both sexes when knocked out (Mason *et al.*, 2002). In *importin-α2* null males, spermatogenesis fails during the individualization stage and extremely few motile sperm are produced (Mason *et al.*, 2002). It may be the case that *importin-α2*

carries a testes-specific cargo into the haploid nuclei of developing sperm that is required for successful development (Mason *et al.*, 2002).

The convergent evolution of meiotic drive systems that manipulate this process is of great interest, and suggests that nuclear transport may be particularly susceptible to conflict. It is well known that genes involved in spermatogenesis and primarily expressed in the testes are rapidly evolving, though this is often attributed to sexual conflict/sexual selection (Kleene, 2005; Haerty *et al.*, 2007). Previous research has also found that *importin-α* duplications and losses are common in *Drosophila,* perhaps because of rapid evolution to avoid conflict (Phadnis *et al.*, 2011). There is also evidence that some components of the nuclear transport pathway are rapidly evolving, perhaps also to avoid conflict (Presgraves, 2007). The evolutionary and expression patterns of *X-pendulin* combined with this knowledge of its wild-type function strongly suggest it is part of the mechanism of SR drive in *D. neotestacea.* Nuclear transport is a housekeeping process that is required in all cells, and so the independent targeting of this process by selfish genetic elements has important implications for our global understanding of how such pathways evolve (Presgraves, 2007). Experimental confirmation of the functional role of *X-pendulin* in drive is therefore a critical next step.

First, *X-pendulin* must be confirmed to contribute drive in SR males. Ideally, this could be confirmed through genetic knockouts, perhaps using RNA interference or Crispr/cas gene editing. However, no genetic transformations have been carried out in *D. neotestacea* or any other closely related species to date. If *X-pendulin* is part of the mechanism of drive, then eliminating or suppressing its expression in the testes of SR males should lead to a change in the proportion of female offspring these males produce.

Even a small increase in the number of fertile sons produced would be strongly suggestive of *X-pendulin* contributing to drive as normally the only sons sired by SR carrying males are sterile X0 males resulting from chromosomal nondisjunction.

Another important question to ask is whether the effect of *X-pendulin* in SR is due to sequence differences between ST and SR or expression level differences. *X-pendulin* has significantly higher expression in SR than ST, though it is expressed in ST at a low level. There are also multiple fixed nucleotide differences between the ST and SR copies of the gene. Either or both could contribute to drive. One way to test this would be to transform the ST copy of *X-pendulin* into the SR genetic background, and see if there is a change in the level of drive. Another potential experiment is to increase the expression level of *X-pendulin* in ST and see if drive results, or transform the SR copy into the ST genetic background.

However, transforming *X-pendulin* into the ST genetic background may not have any effect because there is strong evidence that multiple loci are required for drive to occur in *D. neotestacea*. The primary evidence for this are the multiple inversions that suppress recombination between ST and SR across most the chromosome (Pieper & Dyer, 2016). Such inversions are only expected to fix if they are binding together multiple distorter loci, since recombination is already completely suppressed between the X and Y chromosomes (Meiklejohn & Tao, 2010). However, there are no identified segregating variants of SR that drive at less than 99%. Combined with evidence of some gene flow between ST and SR (Pieper & Dyer, 2016), this suggests that the distorters act together epistatically and that all the loci are required for drive to occur. In this situation, any disruption of *X-pendulin* function in SR would result in the complete elimination of

drive. Furthermore, the addition of the SR copy of *X-pendulin* to the ST genetic background would have no effect on offspring sex ratio without the other distorter loci. This is the case observed in the SR chromosome in *D. persimilis,* where there are at least three loci that are all required for drive (Wu & Beckenbach, 1983).

It could also be the case that the distorters interact additively, and knocking out *X-pendulin* in SR will only decrease the strength of drive rather than eliminate it completely. This is the situation observed in the autosomal male meiotic driver in *Mus musculus,* t-haplotype (Lyon, 2003). In this system, there is a responder locus and at least three distorters which act additively on the responder to cause drive (Lyon, 2003). In this system, the responder is a protein kinase that affects sperm motility and distorters upregulate its expression to cause drive (Herrmann *et al.*, 1999; Bauer *et al.*, 2005; Bauer *et al.*, 2007). If drive in *D. neotestacea* is caused by the higher expression of *X-pendulin* in SR compared to ST, it could be through a similar mechanism. Overall, more work needs to be done to identify the interacting partners of *X-pendulin* and which are also involved in the mechanism of drive. The transcriptome analysis in Chapter 4 of this dissertation found hundreds of transcripts that are differentially expressed between SR and ST, including transcripts that are present in SR but not ST. Additionally, the other identified top candidates may be a part of the drive mechanism. More study of these candidates is required.

In a similar vein, more work must be done to determine the cellular function of *X-pendulin* and how it differs from that of *importin-α2.* Immunostaining with fluorescent tags can be used to determine if *X-pendulin* localizes to the nucleus and the cytoplasm like *importin-α2*, and whether it does this in both ST and SR. The mechanistic models

presented in Chapter 4 suggest that *X-pendulin* disrupts spermatogenesis in SR by selectively transporting cargo into only the nuclei of X or Y-bearing sperm, depending on the model. (Figure 4.5). The actual localization patterns of *X-pendulin* during spermatogenesis in SR testes can be easily tested and should be done as a first step to confirm these models. Interactions between *X-pendulin* and *importin-α2*'s binding partners including *importin-ß* and *CAS* should also be experimentally tested to determine if the function of *X-pendulin* is like that of *importin-α2* or not. A search of the transcriptome data for several known interacting partners of *importin-α2* did not reveal any unusual expression or sequence differences in those transcripts. However, one of the other candidate transcripts has homology to a *D. melanogaster* gene that is involved in mRNA splicing and processing and interacts with the gene *sbr* that itself interacts with *RanGAP* and *importin-ß.*

Finally, more research needs to be done on the *sex-ratio* drive system in the closely related species *D. orientacea.* In the testacea species group, *D. neotestacea, D. testacea,* and *D. orientacea* are very closely related and show high levels of incomplete lineage sorting (Dyer *et al.*, 2011). Additionally, *D. testacea* and *D. orientacea* are capable of interbreeding, though neither can successfully reproduce with *D. neotestacea* (Dyer *et al.*, 2011). The *X-pendulin* duplication is found only in these three species, and not in the two outgroups, *D. putrida* and *D. bizonata.* This raises the question of whether the drive system in *D. orientacea* has the same origin as the one in *D. neotestacea.* Very little is known about SR in *D. orientacea,* which is found in Asia, except that there are segregating suppressors (K. Dyer, unpublished data). The presence of suppressors is in stark contrast to the *D. neotestacea* system, but it is possible that the two drive systems

have diverged since their common origin. Evidence suggests that SR is maintained as a stable polymorphism in *D. neotestacea* by female remating rate, but there could have been suppressors in the past that have now gone to fixation (Pinzone & Dyer, 2013). Currently there is no evidence of drive in *D. testeacea*, but it may simply have not been discovered yet. More collections and sequencing are needed to test these hypotheses and determine how these two drive systems are related, and whether the evolutionary patterns of *X-pendulin* support a common origin of drive in this species group.

References

Bauer, H., Willert, J.R., Koschorz, B. & Herrmann, B.G. 2005. The t complex-encoded GTPase-activating protein Tagap1 acts as a transmission ratio distorter in mice. *Nature Genetics* **37:** 969-973.

Bauer, H., Véron, N., Willert, J. & Herrmann, B.G. 2007. The t-complex-encoded guanine nucleotide exchange factor Fgd2 reveals that two opposing signaling pathways promote transmission ratio distortion in the mouse. *Genes & development* **21:** 143-147.

Dyer, K.A., White, B.E., Bray, M.J., Pique, D.G. & Betancourt, A.J. 2011. Molecular evolution of a Y chromosome to autosome gene duplication in Drosophila. *Mol. Biol. Evol.* **28:** 1293-1306.

Dyer, K.A. 2012. Local selection underlies the geographic distribution of sex-ratio drive in *Drosophila neotestacea*. *Evolution* **66:** 973-984.

Dyer, K.A., Bray, M.J. & Lopez, S.J. 2013. Genomic conflict drives patterns of X-linked population structure in *Drosophila neotestacea*. *Mol. Ecol.* **22:** 157-169.

Goldfarb, D.S., Corbett, A.H., Mason, D.A., Harreman, M.T. & Adam, S.A. 2004. Importin α: a multipurpose nuclear-transport receptor. *Trends in cell biology* **14:** 505-514.

Haerty, W., Jagadeeshan, S., Kulathinal, R.J., Wong, A., Ram, K.R., Sirot, L.K., Levesque, L., Artieri, C.G., Wolfner, M.F. & Civetta, A. 2007. Evolution in the fast lane: rapidly evolving sex-related genes in Drosophila. *Genetics* **177:** 1321-1335.

Helleu, Q., Gérard, P.R., Dubruille, R., Ogereau, D., Prud'homme, B., Loppin, B. & Montchamp-Moreau, C. 2016. Rapid evolution of a Y-chromosome heterochromatin protein underlies sex chromosome meiotic drive. *Proc. Natl. Acad. Sci.* **113:** 4110-4115.

Herrmann, B.G., Koschorz, B., Wertz, K., McLaughlin, K.J. & Kispert, A. 1999. A protein kinase encoded by the t complex responder gene causes non-mendelian inheritance. *Nature* **402:** 141-146.

James, A.C. & Jaenike, J. 1990. Sex-ratio meiotic drive in *Drosophila testacea*. *Genetics* **126:** 651-656.

Kleene, K.C. 2005. Sexual selection, genetic conflict, selfish genes, and the atypical patterns of gene expression in spermatogenic cells. *Developmental Biology* **277:** 16-26.

Kusano, A., Staber, C. & Ganetzky, B. 2001. Nuclear mislocalization of enzymatically active RanGAP causes segregation distortion in Drosophila. *Developmental cell* **1:** 351-361.

Kusano, A., Staber, C. & Ganetzky, B. 2002. Segregation distortion induced by wild-type RanGAP in Drosophila. *Proc. Natl. Acad. Sci.* **99:** 6866-6870.

Larracuente, A.M. & Presgraves, D.C. 2012. The selfish Segregation Distorter gene complex of *Drosophila melanogaster*. *Genetics* **192:** 33-53.

Lyon, M.F. 2003. Transmission ratio distortion in mice. *Annual Review of Genetics* **37:** 393-408.

Mason, D.A., Fleming, R.J. & Goldfarb, D.S. 2002. *Drosophila melanogaster* importin α1 and α3 can replace importin α2 during spermatogenesis but not oogenesis. *Genetics* **161:** 157-170.

Meiklejohn, C.D. & Tao, Y. 2010. Genetic conflict and sex chromosome evolution. *Trends. Ecol. Evol.* **25:** 215-223.

Phadnis, N., Hsieh, E. & Malik, H.S. 2011. Birth, death and replacement of karyopherins in Drosophila. *Mol. Biol. Evol.* **29:** 1429-40.

Pieper, K.E. & Dyer, K.A. 2016. Occasional recombination of a selfish X-chromosome may permit its persistence at high frequencies in the wild. *Journal of Evolutionary Biology* **29:** 2229-2241.

Pinzone, C.A. & Dyer, K.A. 2013. Association of polyandry and sex-ratio drive prevalence in natural populations of *Drosophila neotestacea*. *Proc. R. Soc. B Biol. Sci.* **280:** 20131397.

Presgraves, D.C. 2007. Does genetic conflict drive rapid molecular evolution of nuclear transport genes in Drosophila? *BioEssays* **29:** 386-391.

Tao, Y., Araripe, L., Kingan, S.B., Ke, Y., Xiao, H. & Hartl, D.L. 2007a. A sex-ratio meiotic drive system in *Drosophila simulans*. II: an X-linked distorter. *PLoS Biol* **5:** e293.

Tao, Y., Masly, J.P., Araripe, L., Ke, Y. & Hartl, D.L. 2007b. A sex-ratio meiotic drive system in *Drosophila simulans*. I: an autosomal suppressor. *PLoS Biol* **5:** e292.

Wu, C.-I. & Beckenbach, A.T. 1983. Evidence for extensive genetic differentiation between the sex-ratio and the standard arrangement of *Drosophila pseudoobscura* and *D. persimilis* and identification of hybrid sterility factors. *Genetics* **105:** 71-86.

Wu, C.-I., Lyttle, T.W., Wu, M.-L. & Lin, G.-F. 1988. Association between a satellite DNA sequence and the Responder of Segregation Distorter in *D. melanogaster*. *Cell* **54:** 179-189.

APPENDIX A

SUPPLEMENTAL MATERIAL FOR CHAPTER 3

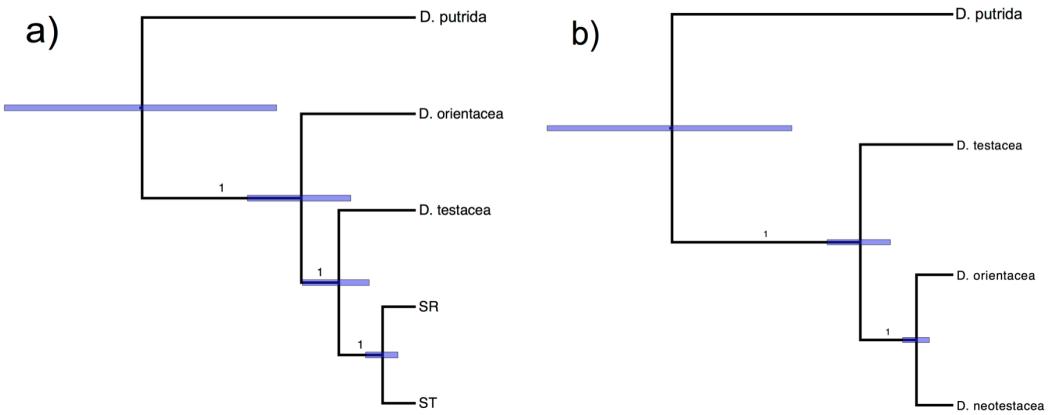
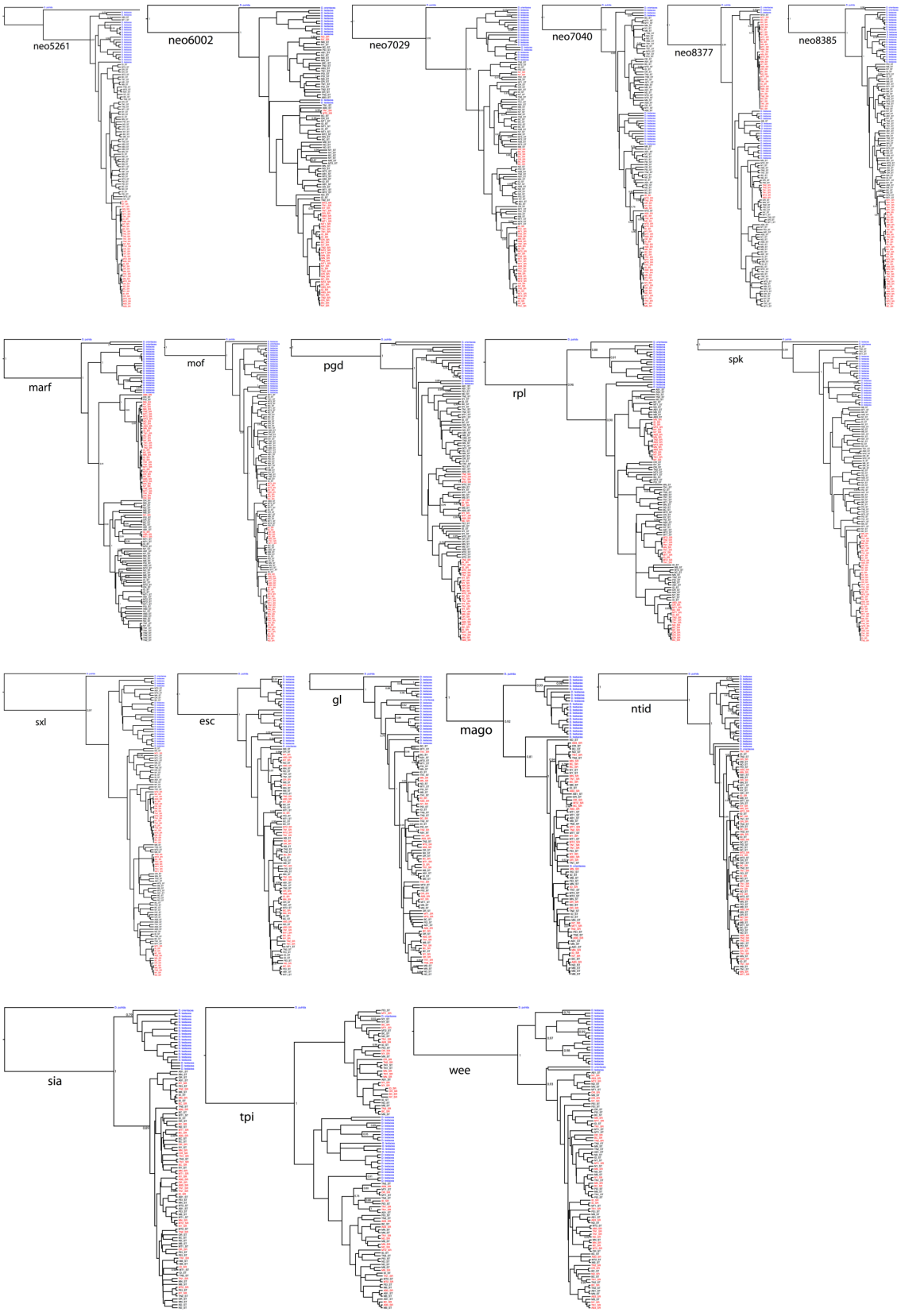


Figure S1. Multi-locus species phylogeny for the a) X-linked loci and b) autosomal loci. Posterior probabilities are mapped onto the branches, and the node bars represent the 95% highest posterior density for the node height. For the X-linked tree, *D. neotestacea* sequences were split into ST and SR samples.

Figures S2. Individual gene trees. Trees were built in *BEAST (see Methods), and the posterior probabilities greater than 0.75 are mapped onto branches. Individual branches are labeled with population or province of origin (Table S1) and X-chromosome type. SR samples are noted in red text, ST samples are black, and outgroup samples are blue.


neo5261
neo6002
neo7029
neo7040
neo8377
neo8385
marf
mof
pgd
rpl
spk
sxl
esc
gl
mago
ntid
sia
tpi
wee

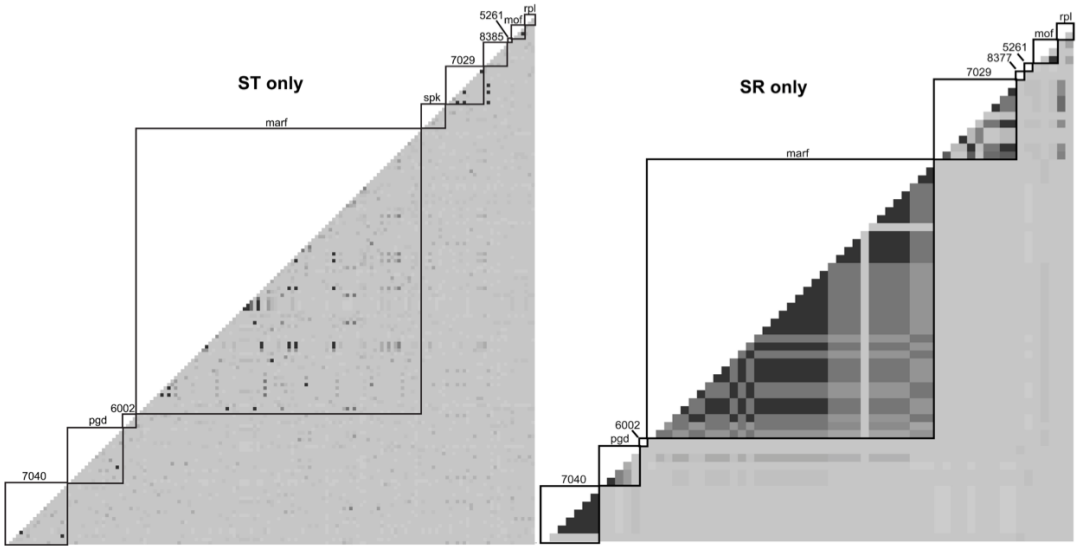


Figure S3. Linkage disequilibrium between SNPs considering ST and ST samples separately. Shown are R^2 values between pairs of SNPs within and between each marker, using parsimony informative sites only. All sites included in the analyses are pictured. R^2 is represented as a heat map, with the lightest grey = 0 and the darkest grey = 1. *mof* and *rpl* were unable to be mapped, have been added to the end of the alignment. Some markers are not represented in the SR-only matrix due to a lack of parsimony informative sites.

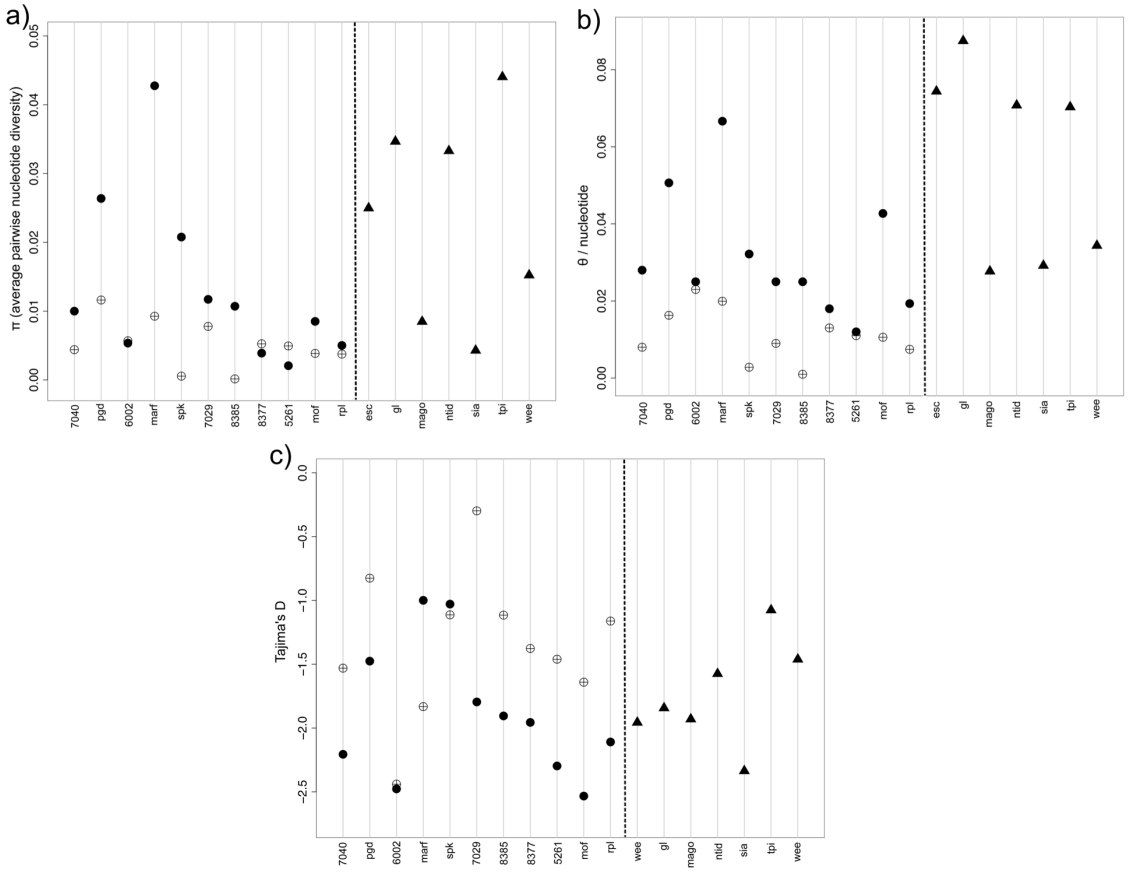


Figure S4. Polymorphism and frequency spectrum estimates for each individual marker, using only silent sites. On the X-chromosome, estimates for ST (filled circles) and SR (open circles) were calculated separately. Estimates for the autosomes (triangles) were calculated using all samples. Panel (a) shows π, panel (b) shows Watterson's θ, and panel (c) shows Tajima's D. Markers are presented in their ST map order. *mof* and *rpl* were unable to be mapped and have been added to the end.

Table S1. All of the populations used in this study, and the number of unique ST and SR males sampled from each. Not all individuals were sequenced at all loci. A subset of individuals random with respect to X-chromosome status was chosen for sequencing at autosomal loci.

Population	Collection location	SR samples	ST samples	Autosomal samples
AB1	Winston Churchill, AB	0	4	2
AB3	Jasper, AB	4	4	3
BC	Vancouver, BC	4	4	4
ID	Coeur d'Alene, ID	6	8	3
MB	The Pas, MB	0	4	3
MN	Bemidji, MN	4	4	4
MT1	Columbia Falls, MT	5	8	3
MT2	St. Regis, MT	4	5	3
ND	Minot, ND	2	4	3
NY	Rochester, NY	6	5	3
OR	MacKenzie Bridge, OR	4	4	4
PEI	Charlottetown, PEI	0	4	3
TN1	Gatlinburg, TN	6	3	4
TN2	Clingmans Dome, TN	3	4	3

Table S2. Molecular population genetic summary statistics for each locus. A) Statistics for *D. neotestacea* samples. S (segregating sites) for total and silent sites is presented. M (mutations), π, θ, and D for silent sites only includes synonymous sites in open reading frames and all noncoding sites. π NS is using nonsynonymous polymorphism only. ρ is population recombination rate ($2N_er$) divided by the number of nucleotides in the marker. For the autosomal markers, all statistics were calculated from the combined set of SR and ST individuals. ρ and ZnS could not be calculated for some markers due to a lack of segregating sites. Bolded D values are less than the expected D in 95% of 10,000 simulations. Da is percent divergence between ST and SR.

	Marker	X-chromosome	N	Total Sites	S	Silent sites	M silent	S silent	π silent	π NS	θ silent	D silent	ρ	ZnS	π_a/π_s	K_a/K_s	Da
X-linked	neo5261	ST	61	183	10	183	10	10	0.002		0.012	**-2.297**	0.000	0.178			0.274
		SR	45		7		7	7	0.005		0.011	-1.461	0.000	0.054			
	neo6002	ST	57	190	19	190	21	19	0.005		0.025	**-2.477**	0.300	0.035			0.388
		SR	44		19		19	19	0.006		0.023	**-2.440**	0.000	0.060			
	neo7029	ST	57	369	39	369	41	39	0.012		0.025	**-1.796**	0.081	0.036			0.722
		SR	42		13		13	13	0.008		0.009	-0.299	0.008	0.317			
	neo7040	ST	57	515	56	515	59	56	0.010		0.028	**-2.206**	0.179	0.029			0.474
		SR	43		18		18	18	0.004		0.008	**-1.531**	0.050	0.106			
	neo8385	ST	55	319	32	319	32	32	0.011		0.025	**-1.905**	0.444	0.029			0.745
		SR	44		1		1	1	0.000		0.001	-1.115	0.000	NA			
	neo8377	ST	56	72	5	72	6	5	0.004		0.018	**-1.956**	0.266	0.000			0.993
		SR	44		4		4	4	0.005		0.013	-1.377	NA	0.502			
	marf	ST	57	1063	186	670.34	206	184	0.043	0.005	0.067	-0.999	0.122	0.037	0.195	0.424	1.581
		SR	48		73	844.49	74	71	0.009	0.001	0.020	**-1.832**	0.008	0.252	0.577	0.282	
	mof	ST	65	576	30	128.36	26	26	0.009	0.000	0.043	**-2.533**	0.000	0.041	0.044	0.065	0.002
		SR	47		7	128.34	6	6	0.004	0.000	0.011	**-1.641**	0.000	0.049	0.038	0.064	
	pgd	ST	56	569	38	128.98	30	28	0.026	0.002	0.051	**-1.476**	0.146	0.023	0.064	0.069	-0.010
		SR	43		18	127.7	9	9	0.012	0.001	0.016	**-0.826**	0.000	0.284	0.092	0.068	
	rpl	ST	57	302	11	123.33	11	11	0.005	0.000	0.019	**-2.109**	0.024	0.051	0.00	0.00	0.008
		SR	44		4	123.33	4	4	0.004	0.000	0.007	-1.162	0.066	0.004	0.00	0.000	
	spk	ST	56	382	13	81.17	12	12	0.021	0.000	0.032	-1.029	0.298	0.015	0.006	0.021	0.041
		SR	45		1	81.17	1	1	0.001	0.000	0.003	-1.113	NA	NA	0.00	0.022	
Autosomal	esc		82	370	29	78.32	29	27	0.025	0.001	0.074	**-1.956**	0.130	0.023	0.020	0.002	
	gl		80	402	52	96.96	42	39	0.035	0.002	0.087	**-1.843**	0.065	0.023	0.060	0.059	
	mago		78	324	14	57.15	7	7	0.010	0.001	0.025	**-1.429**	0.086	0.058	0.109	0.500	
	ntid		88	527	50	123.42	44	41	0.033	0.003	0.071	**-1.575**	0.068	0.027	0.138	0.110	
	sia		82	400	13	96.33	14	13	0.004	0.000	0.029	**-2.336**	0.000	0.015	0.024	0.084	
	tpi		80	347	33	83.28	29	29	0.044	0.001	0.070	-1.011	0.245	0.043	0.00	0.000	
	wee		84	285	14	58.14	10	10	0.015	0.001	0.034	**-1.462**	0.561	0.027	0.048	0.003	

Table S2. Molecular population genetic summary statistics for each locus. B) Statistics for *D. testacea* samples. All analyses included both synonymous and nonsynonymous sites. M is the number of mutations; S is the number of segregating sites. Bolded D values are less than the expected D in 95% of 10,000 simulations. ρ is population recombination rate ($2N_er$) divided by the number of nucleotides in the marker.

	Marker	N	Total sites	M	S	π	θ	D	ZnS	ρ
X-linked	neo5261	19	175	14	12	0.017	0.027	-1.429	0.126	0.218
	neo6002	12	191	6	5	0.007	0.010	-1.225	0.060	9.857
	neo7029	21	352	19	19	0.010	0.017	-1.621	0.056	NA
	neo7040	16	513	28	27	0.012	0.017	-1.264	0.073	3.264
	neo8377	21	72	3	3	0.016	0.017	-0.137	0.061	NA
	neo8385	21	312	9	7	0.004	0.009	**-2.018**	0.004	NA
	marf	20	244	7	7	0.004	0.008	-1.692	0.177	0.002
	mof	23	575	17	15	0.006	0.007	-1.063	0.104	0.114
	pgd	16	560	27	26	0.010	0.014	-1.420	0.192	0.017
	rpl	18	269	9	9	0.005	0.010	-1.561	0.140	0.024
	spk	22	382	11	11	0.004	0.008	-1.768	0.193	0.009
Autosomal	esc	25	369	13	12	0.006	0.009	-1.197	0.072	0.859
	gl	23	402	29	27	0.017	0.018	-0.551	0.103	0.189
	mago	21	326	6	5	0.003	0.005	-1.778	0.003	NA
	ntid	28	407	29	29	0.009	0.009	**-1.823**	0.116	0.012
	sia	20	399	2	2	0.001	0.001	-1.513	0.003	NA
	tpi	23	347	13	13	0.006	0.010	-1.499	0.088	0.140
	wee	21	285	9	9	0.008	0.009	-0.461	0.115	0.186

Table S3. Forward and reverse PCR primers used to amplify sequences for each marker. All primers are presented in 5' to 3' orientation.

Marker	Primer	Sequence
neo5261	F	GAAGCAACAACAAAAGCC
	R	AATGAGGCAAGGTCCCACTG
neo6002	F	TCTAAATGCACAAATCCCAGC
	R	CACGACTGCGTAATACTTCACC
neo7029	F	AGCACATGGCACAGATGTTAG
	R	GAAGGATACAAGAGACGTCAGC
neo7040	F	CAAACAACAATTGCAACGTG
	R	GTGTGCACACATTTCCATACC
neo8377	F	TGGACAATTGTTGTGGACTG
	R	AACAACATCATTCGCATTCG
neo8385	F	AGAGCTTTAATGTGCTGGCA
	R	CCCAACTGAAAGTGAATTG
marf	F	CCCAACATCTTCATCCTGAACAAYMGNTGGGA
	R	GCGGACTGGGAGATGCAYTCYTCRAA
mof	F	CAGAAGCGRCGCTACGA
	R	TAKGCCCAATAGCTGCGATA
pgd	F	ATYGATGGYGGCAACTC
	R	CNCGCATWAGCATRAAKCCYTG
rpl	F	CMRVGSCCACAAGACCWCSAARRTC
	R	CRTGRGTCTGRGCCTTCC
spk	F	AAVATGCCBARYATYAARYTGCARTC
	R	CTTCTCCTCRCACCAYTCRTTC
esc	F	GGCCATCAACGAGCTGAARTTYCAYCC
	R	TTCCAGCACACGATGGCRTTYTCRCA
gl	F	TTTCGATTGCGGCGGNTGYTTYGA
	R	GCCGTGGTGCATGGTCATRTTCAT
mago	F	CCACAAGGGCAAGTTCGGNCAYGARTT
	R	CACTTCAGGTCCTGCACCARRTARTARAA
ntid	F	GGGCCGCATCTTCGARCAYAARTGG
	R	TGGAGGGGTAGGTGTTCCARCARTA
sia	F	TCGAGTGCCCCGTGTGYTTYGAYTA
	R	GAAGTGGAAGCCGAAGCAGSWYTGCATCAT
tpi	F	CAACTGGAAGATGAAYGGIGACC
	R	TTCTTGGCATAGGCGCACATYTG
wee	F	GCCTGGGCCGAGGAYGAYCAYATG
	R	TCACGTGGCCCAGGTCNCCDATYTT

Table S4. Genetic differentiation per marker. Significance for K_{ST} and S_{nn} were calculated in DnaSP using 1,000 randomized permutations. Italics indicate $p < 0.01$, and $p < 0.001$. For the X-linked markers, the number of segregating sites unique to ST/SR and shared between ST and SR are presented.

	Marker	Geographic K_{ST}	Geo S_{nn}	ST-SR K_{ST}	ST-SR S_{nn}	SR only sites	ST only sites	Shared sites
X-linked	neo5261	0.0069	0.06537	***0.286***	***0.7624***	6	10	0
	neo6002	0.01212	0.07205	***0.272***	***0.84451***	10	13	5
	neo7029	**0.11563**	**0.11476**	***0.260***	***0.82035***	4	28	9
	neo7040	-0.02985	0.06235	***0.265***	***0.95568***	7	51	11
	neo8377	0.01795	0.07495	***0.524***	***0.86763***	3	3	1
	neo8385	0.00955	0.07416	***0.381***	***0.95946***	0	31	0
	marf	0.02777	0.08879	***0.249***	***0.97087***	29	155	41
	mof	*0.02616*	**0.10742**	*0.007*	**0.55057**	4	24	2
	pgd	-0.01186	0.09386	***0.169***	***0.93777***	11	29	7
	rpl	**0.04427**	0.08124	**0.022**	***0.5894***	3	10	1
	spk	0.00158	0.06184	***0.074***	***0.73976***	1	11	0
Autosomal	esc	*0.04437*	***0.16071***	-0.003	0.50313			
	gl	-0.02099	0.06884	0.005	*0.59407*			
	mago	*0.0632*	**0.11973**	0.004	0.55051			
	ntid	-0.01788	0.07167	-0.004	0.46653			
	sia	0.01782	0.08351	0.003	0.52953			
	tpi	*0.05581*	*0.11495*	-0.001	0.45123			
	wee	*0.014*	0.08582	-0.001	0.4705			

Table S5. Evdence of gene conversion and/or double crossover events between SR and ST. Start and end positions are the nucleotide positions within the concentrated alignment of all X-linked markers, and the start and end loci are which markers contain these positions. The tract length is also within the concatenated alignment.

Recipient sample	Start position	End position	Tract length (bp)	Source	Recipient	Start marker	End marker
TN2_27|SR	45	3435	3296	ST	SR	neo7040	neo8377
MT2_27|SR	121	3435	3219	ST	SR	neo7040	neo8378
BC_28|SR	45	2885	2753	ST	SR	neo7040	neo7030
OR_6|SR	121	2885	2677	ST	SR	neo7040	neo7031
OR_27|SR	672	2885	2144	ST	SR	pgd	neo7032
BC_8|SR	672	2885	2144	ST	SR	pgd	neo7033
NY_27|SR	672	2224	1490	ST	SR	pgd	marf
NY_29|SR	330	1829	1449	ST	SR	neo7040	marf
MN_5|SR	121	1491	1344	ST	SR	neo7040	marf
AB3_30|SR	672	2047	1318	ST	SR	pgd	marf
ID_1|SR	121	672	534	ST	SR	neo7040	pgd
MT1_30|SR	1491	2047	508	ST	SR	marf	marf
ID_28|SR	1491	2047	507	ST	SR	marf	marf
AB3_27|SR	1851	2047	175	ST	SR	marf	marf
NY_28|SR	1152	1164	13	ST	SR	neo6002	neo6002
MT2_11|ST	831	1605	753	SR	ST	pgd	marf
MT2_8|ST	2761	3435	667	SR	ST	neo7029	neo8379
MN_2|ST	2168	2761	593	SR	ST	marf	neo7029
OR_10|ST	1152	1605	433	SR	ST	neo6003	marf
OR_9|ST	1813	2172	329	SR	ST	marf	marf
MT2_8|ST	141	330	184	SR	ST	neo7040	neo7040
ID_3|ST	1705	1766	62	SR	ST	marf	marf
BC_4|ST	1705	1766	62	SR	ST	marf	marf
PEI_8|ST	1764	1813	50	SR	ST	marf	marf
MT2_10|ST	1766	1813	48	SR	ST	marf	marf
MT2_9|ST	1766	1813	48	SR	ST	marf	marf
MT1_13|ST	1766	1813	48	SR	ST	marf	marf
ND_1|ST	2168	2172	5	SR	ST	marf	marf
OR_7|ST	1764	1766	3	SR	ST	marf	marf
NY_4|ST	1764	1766	3	SR	ST	marf	marf
AB3_1|ST	1764	1766	3	SR	ST	marf	marf
BC_1|ST	1868	1870	3	SR	ST	marf	marf
AB3_4|ST	1868	1870	3	SR	ST	marf	marf

APPENDEX B

SUPPLEMENTAL MATERIAL FOR CHAPTER 4

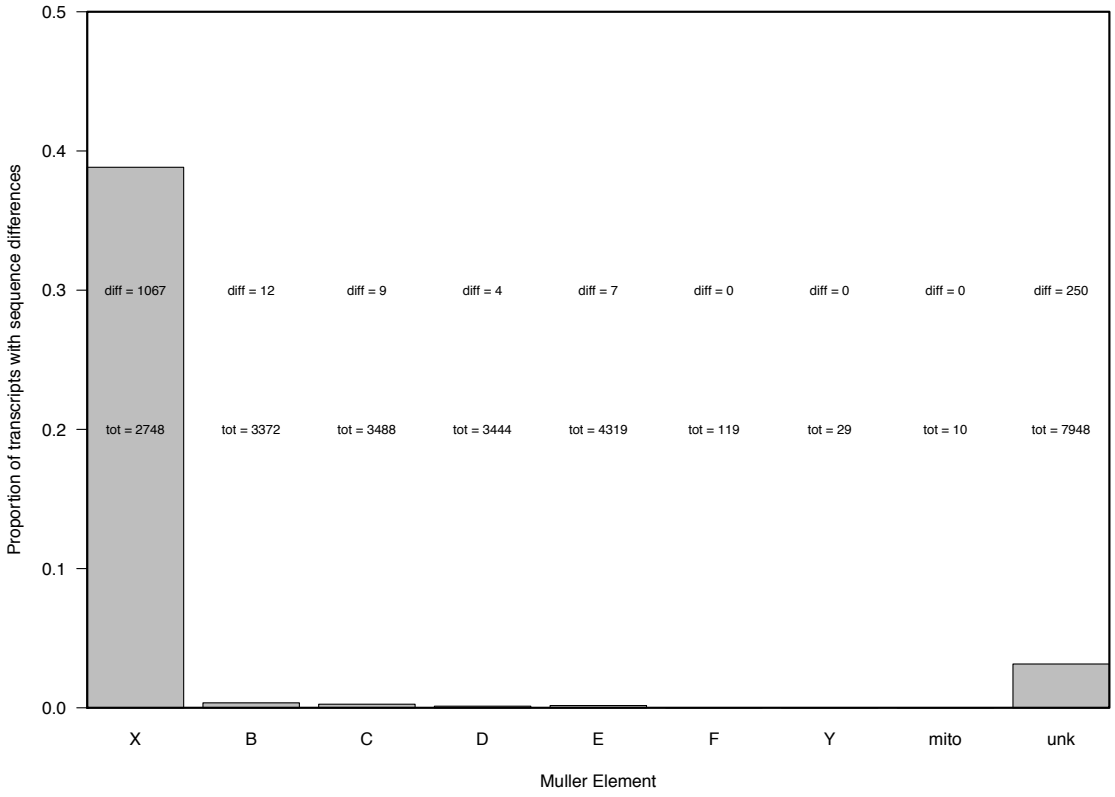


Figure S1. The original genomic locations of the transcripts with sequence differences between ST and SR. The locations were assigned via homology with *D. melanogaster* and *D. virilis.* All transcripts with sequence differences were reassigned to the X-chromosome.

.

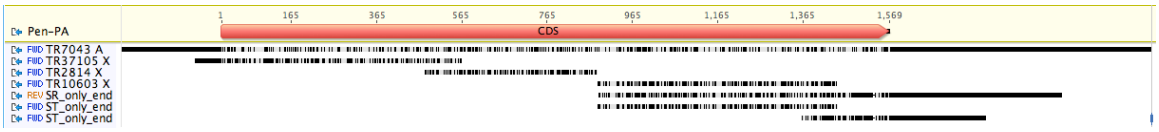


Figure S2. The transcripts that make up *X-pendulin*. The *D. neotestacea* transcripts have been aligned to the *D. melanogaster Pendulin* sequence (Pen-PA), which is covered by the red CDS annotation arrow. TR7043 is the autosomal copy of *Pendulin.* TR37105, TR2814, and TR10603 from the combined transcriptome assembly make up the majority of *X-pendulin.* The end of the transcript has been extracted from both an SR only transcriptome assembly and an ST only transcriptome assembly. A single large transcript covering the end of the transcript and the 3' UTR in the SR only assembly, and two overlapping transcripts covering the same distance were pulled from the ST only assembly.

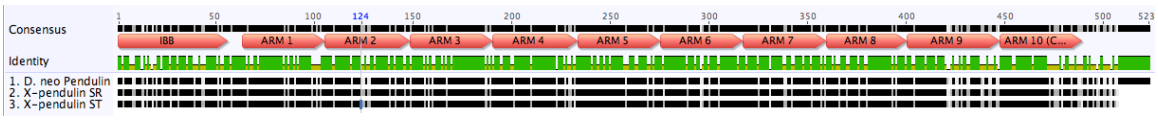


Figure S3a. An alignment of the protein sequences of *Pendulin* and *X-pendulin* in SR and ST. The protein domains are labeled, and the amino acids are coding according to similarity. Black represents identical residues, while white residues are the most different. The N-terminal end of the protein appears to be the most diverged between the autosomal and X-linked copies. The IBB domain binds to *importin-ß*, ARM repeats 1 – 9 bind to the NLS signals of cargo, and ARM repeat 10 binds to the export factor.

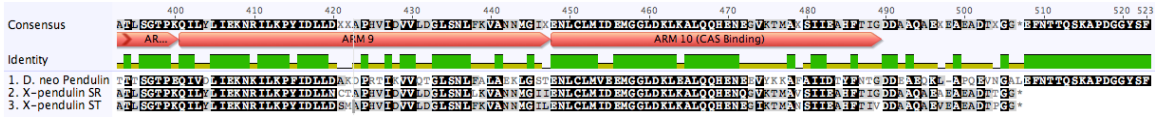


Figure S3b. The protein sequence alignment zoomed in on the diverged N-terminus. The * denotes a stop codon.

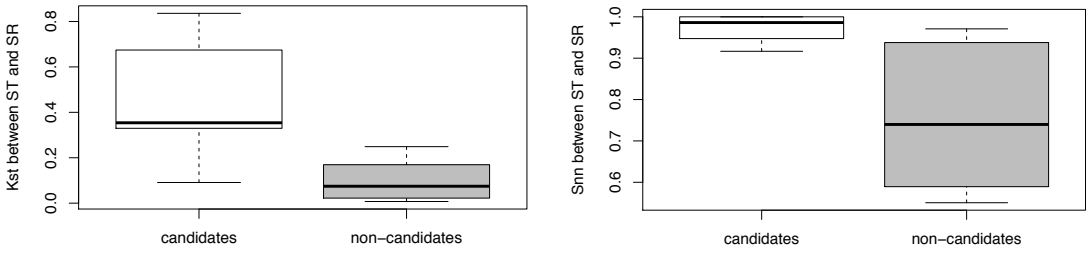


Figure S4. K_{ST} and S_{nn} between the top candidates and the non-candidate X-linked loci.

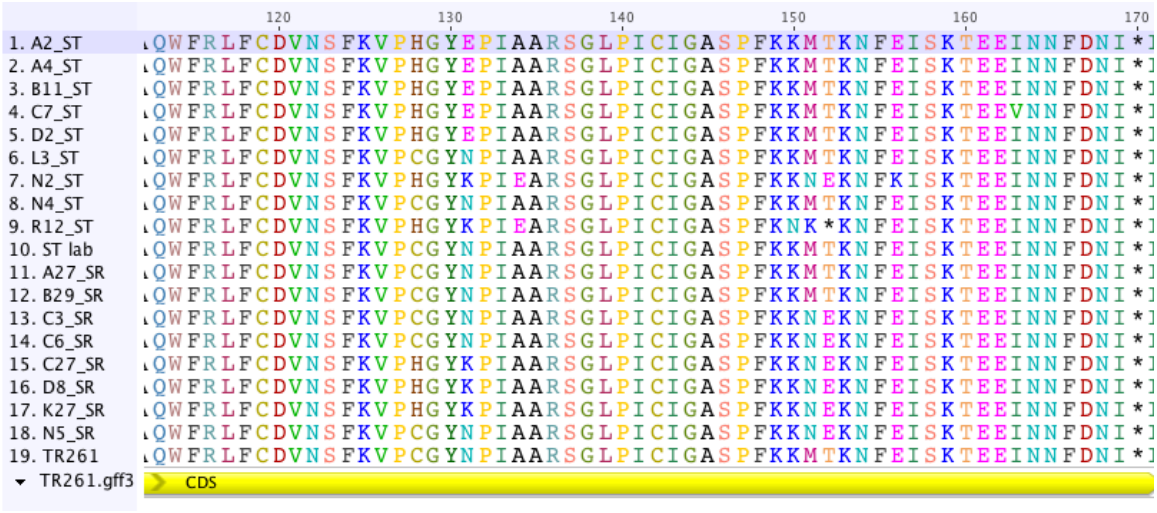


Figure S5. A premature stop codon (*) at residue 152 in TR261 the ST sample R12. The end of the transcript is at 170.

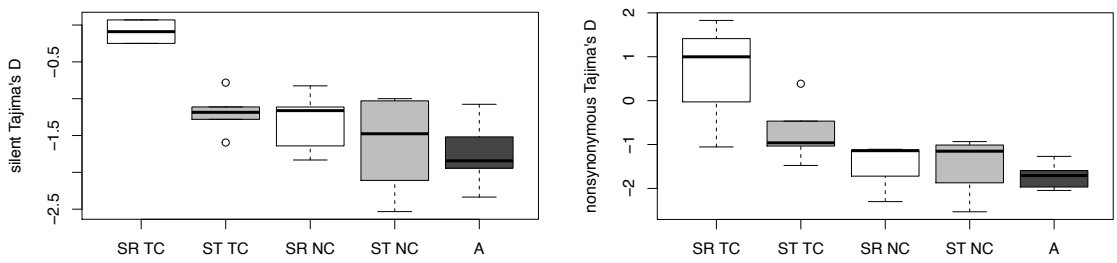


Figure S6. Tajima's D for the top candidates (TC) and non-candidates (NC) on ST and SR and the autosomes (A). The panel on the left was calculated using only non-coding and synonymous variation. The panel on the right was calculated using only nonsynonymous variation.

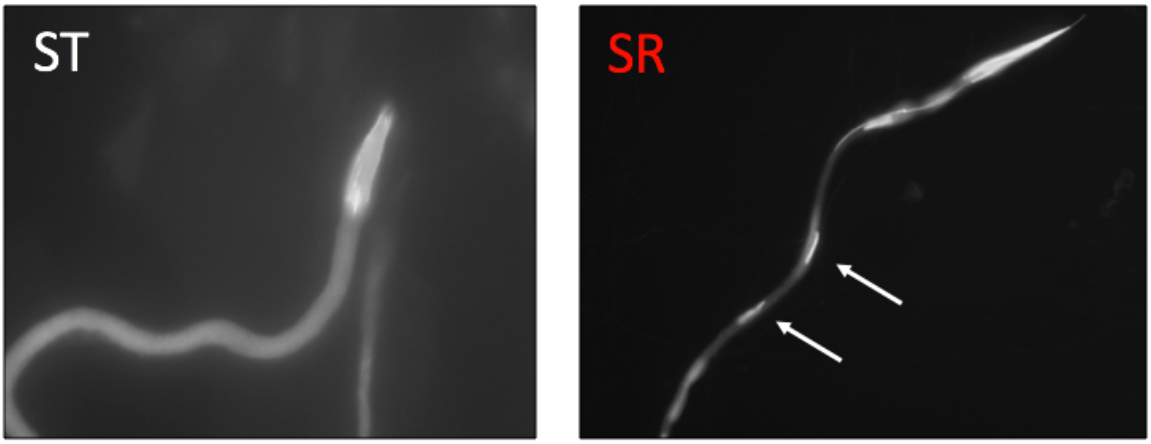


Figure S7. Developing bundle of 64 spermatids in the testes of ST males (left) and SR males (right) at 650X magnification. The DNA is stained with DAPI, revealing the heads of the spermatids. In SR, half of the sperm are not developing properly. Arrows point out the heads of sperm that are not part of the bundle.

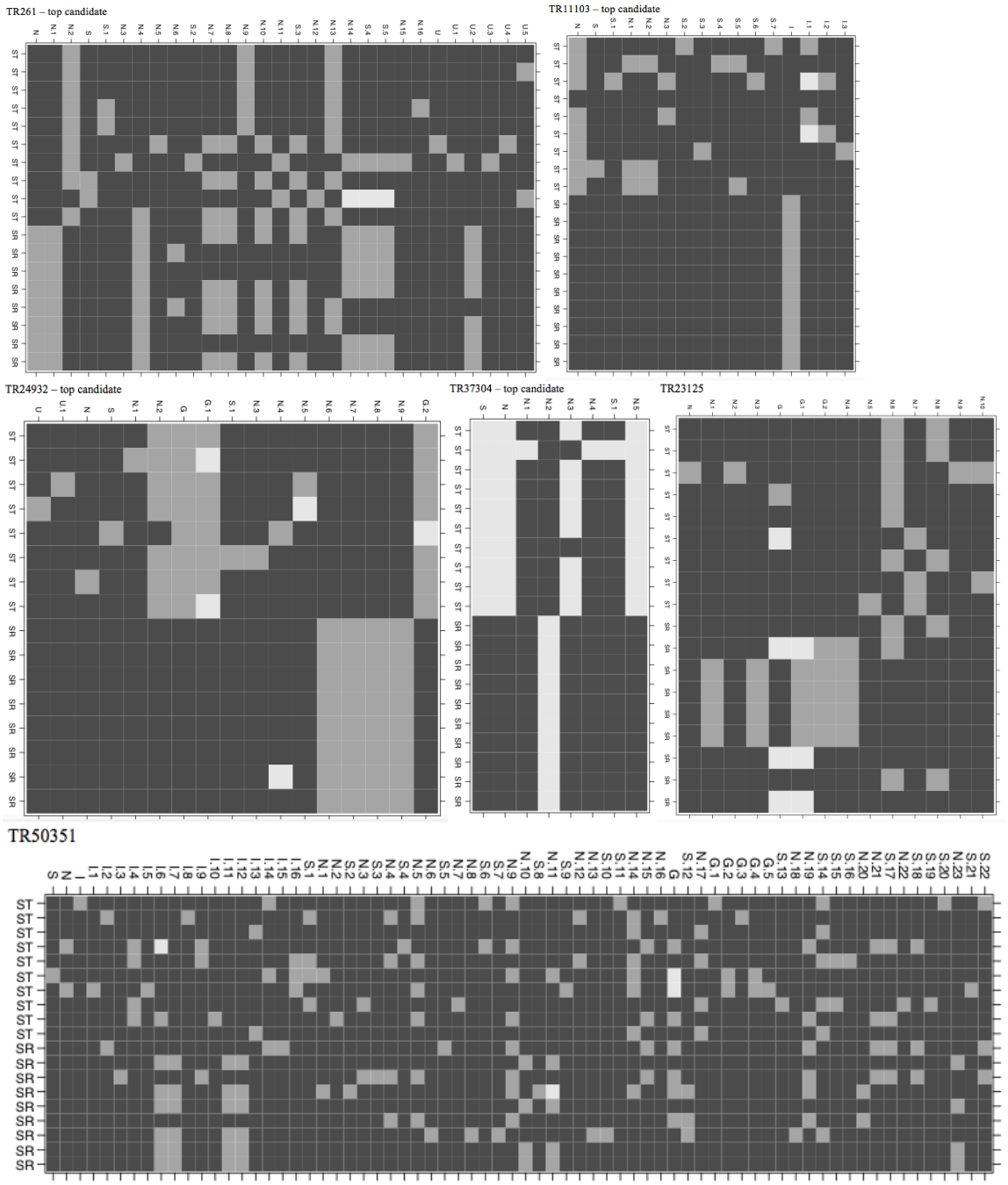


Figure S8. The segregating sites of the sequenced transcripts. Dark grey indicates the major allele, light grey is the minor allele, white is the second minor allele when present. Sites are labeled S for synonymous, N for nonsynonymous, I for intron, U for UTR, and G for gap.

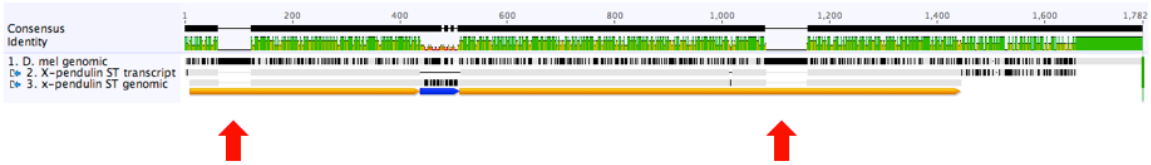


Figure S9. An alignment of the genomic sequence of *Pendulin* in *D. melanogaster,* the transcript sequence of *X-pendulin* in ST *D. neotestacea*, and the genomic sequence of *X-pendulin* in *D. neotestacea.* The *D. melanogaster* sequence contains the entire open reading frame, from start to stop. Note the *X-pendulin* transcript ends short of the *D. melanogaster* transcript. The ORFs of the genomic sequence of *X-pendulin* are annotated as yellow arrows along the transcript, and the intron is marked in blue. There are two introns in *D. melanogaster* genomic sequence that are not present in the *D. neotestacea X-pendulin* genomic sequence, which are marked with red arrows.

Table S1. Forward and reverse PCR primers.

Transcript	**Forward 5' – 3'**	**Reverse 5' – 3'**
TR261	GCAAATACACGAGCACCAACA	TGTCAGGAGAAGGTGAGATGG
TR11103	CGATCCCCAGCATCATTCGA	CACCAGTCCAAAGTTGCAGC
TR23125	GACAAGGAAGGCGAACAGGA	TGCTATTGATGCGCTCCGAT
TR24932	CCTTGTCGCCGTGAGTGTTA	TGGGGAACAATTGAAAGCGC
TR37304	CCGGGTGCAATAATTGAGCG	GCGGAAAATGGCACTGGTTT
TR50351	TTGAAGTGGCAATCCCGGAA	GATCAACGTTACTGCGTCGC
X-pendulin	AACATCAGGAAAGAAAAGCATAACT	GCACCACATCGATGACATGC

Table S2. The six K_a/K_s candidate transcripts and four other transcripts with K_a/K_s values suggestive of positive selection. Mean expression level, $\log_2$ fold change between ST and SR, and FDR corrected p-value was calculated in DESeq2.

	transcript	K_a	K_s	K_a/K_s	Length	Subs	S Subs	N Subs	Mean Expression	log2 Fold Change	padj	% difference	Mean Expression	stage	annotation
K_a/K_s candidate	TR10603	0.025	0.021	1.183	558	13	2.789	10.211	464.581	0.767	0	2.326	0.468	none	pendulin
K_a/K_s candidate	TR23125	0.012	0.01	1.29	609	7	1.527	5.473	420.433	2.638	0	1.149	1.046	mitosis	l(1)1Bi
K_a/K_s candidate	TR24932	0.019	0.007	2.694	357	5	0.792	4.208	208.266	-1.22	0	0.772	1.578	none	none
K_a/K_s candidate	TR2814	0.018	0.017	1.048	399	7	2.072	4.928	585.588	0.911	0	1.737	1.014	none	pendulin
K_a/K_s candidate	TR6297	0.008	0.008	1.012	510	4	0.998	3.002	514.679	1.188	0	0.67	0.718	none	CG7366
K_a/K_s candidate	TR5481	0.013	0.002	8.283	1137	10	0.521	9.479	426.12	-0.873	0	0.876	0.718	none	none
positive	TR261	0.02	0	50	522	8	0.05	7.95	911.2	-0.613	0	0.936	6.488	none	CG32371
positive	TR11103	0.008	0.004	2.124	1224	8	1.306	6.694	508.192	-0.036	0.949	0.671	11.376	mitosis	CG2685
positive	TR37304	0.006	0.004	1.565	1371	8	1.44	6.56	3140.84	-0.427	0.001	0.593	16.864	postmeiosis	CG4198
positive	TR50351	0.013	0.011	1.201	678	8	1.699	6.301	742.04	-0.346	0.1	1.261	2.655	none	CG15452

Table S3. The transcripts from canonical *importin-α* genes and *X-pendulin* in *D. neotestacea.* Mean expression level, $\log_2$ fold change between ST and SR, and FDR corrected p-value was calculated in DESeq2.

Gene	Transcript	Mean Expression	$\log_2$ Fold Change	Adjusted P-value	Chromosome	Carcass Mean Expression	% difference	K_{ST}/K_S	Substitutions	Length
importin-α1	TR22571	5769.290	0.177	0.534	D	29.387	0.000	0.000	0.000	1655
importin-α3	TR6773	2018.999	0.134	0.482	E	49.905	0.000	0.000	0.000	1542
Pendulin	TR7043	52604.270	-0.081	0.668	B	136.289	0.000	0.000	0.000	1569
X-pendulin	TR37105	1149.714	0.450	0.013	X	1.514	1.923	0.349	12.000	624
X-pendulin	TR10603	464.581	0.767	0.000	X	0.468	2.326	1.183	13.000	559
X-pendulin	TR2814	585.588	0.911	0.000	X	1.014	1.737	1.048	7.000	403

Table S4. The population origin of males used in for the population genetic analysis. Not all individual samples were sequenced at every locus.

Population	Collection location	SR samples	ST samples
AB3	Jasper, AB	2	0
BC	Vancouver, BC	1	2
ID	Coeur d'Alene, ID	1	4
MB	The Pas, MB	0	1
MN	Bemidji, MN	1	2
MT1	Columbia Falls, MT	0	2
MT2	St. Regis, MT	2	3
ND	Minot, ND	0	1
OR	MacKenzie Bridge, OR	3	1
PEI	Charlottetown, PEI	0	1

Table S5. Molecular population genetic summary statistics for each non-candidate locus. Statistics for *D. neotestacea* samples. S (segregating sites) for total and silent sites is presented. M (mutations), π, θ, and D for silent sites only includes synonymous sites in open reading frames and all noncoding sites. π NS is using nonsynonymous polymorphism only. ρ is population recombination rate ($2N_er$) divided by the number of nucleotides in the marker. For the autosomal markers, all statistics were calculated from the combined set of SR and ST individuals. ρ and ZnS could not be calculated for some markers due to a lack of segregating sites. Bolded D values are less than the expected D in 95% of 10,000 simulations. Da is percent divergence between ST and SR.

	Marker	X-chromosome	N	Total Sites	S	Silent sites	M silent	S silent	π silent	π NS	θ silent	D silent	ρ	ZnS	π_a/π_s	K_a/K_s	Da
X-linked	marf	ST	57	1063	186	670.34	206	184	0.043	0.005	0.067	-0.999	0.122	0.037	0.195	0.424	1.581
		SR	48		73	844.49	74	71	0.009	0.001	0.020	**-1.832**	0.008	0.252	0.577	0.282	
	mof	ST	65	576	30	128.36	26	26	0.009	0.000	0.043	**-2.533**	0.000	0.041	0.044	0.065	0.002
		SR	47		7	128.34	6	6	0.004	0.000	0.011	**-1.641**	0.000	0.049	0.038	0.064	
	pgd	ST	56	569	38	128.98	30	28	0.026	0.002	0.051	**-1.476**	0.146	0.023	0.064	0.069	-0.010
		SR	43		18	127.7	9	9	0.012	0.001	0.016	**-0.826**	0.000	0.284	0.092	0.068	
	rpl	ST	57	302	11	123.33	11	11	0.005	0.000	0.019	**-2.109**	0.024	0.051	0.00	0.00	0.008
		SR	44		4	123.33	4	4	0.004	0.000	0.007	-1.162	0.066	0.004	0.00	0.000	
	spk	ST	56	382	13	81.17	12	12	0.021	0.000	0.032	-1.029	0.298	0.015	0.006	0.021	0.041
		SR	45		1	81.17	1	1	0.001	0.000	0.003	-1.113	NA	NA	0.00	0.022	
Autosomal	esc		82	370	29	78.32	29	27	0.025	0.001	0.074	**-1.956**	0.130	0.023	0.020	0.002	
	gl		80	402	52	96.96	42	39	0.035	0.002	0.087	**-1.843**	0.065	0.023	0.060	0.059	
	mago		78	324	14	57.15	7	7	0.010	0.001	0.025	**-1.429**	0.086	0.058	0.109	0.500	
	ntid		88	527	50	123.42	44	41	0.033	0.003	0.071	**-1.575**	0.068	0.027	0.138	0.110	
	sia		82	400	13	96.33	14	13	0.004	0.000	0.029	**-2.336**	0.000	0.015	0.024	0.084	
	tpi		80	347	33	83.28	29	29	0.044	0.001	0.070	-1.011	0.245	0.043	0.00	0.000	
	wee		84	285	14	58.14	10	10	0.015	0.001	0.034	**-1.462**	0.561	0.027	0.048	0.003	

Table S6. Data used for the HKA tests. The first set of data is for SR, and the second set is for ST. Divergence is measured between ST and SR. S is the number of segregating sites, N is the number of samples, and divergence is measured between ST and SR. The scalar was calculated as 0.75*0.15 for SR and 0.75*0.85 for ST.

	Marker	length	S	N	divergence	theta	scalar
SR	x-pen	1283	6	11	22.013	0.00158	0.1125
	TR261	627	10	8	12.25	0.00631	0.1125
	TR11103	826	0	10	4.556	0.00001	0.1125
	TR23125	343	6	10	4.167	0.00618	0.1125
	TR24932	280	1	8	7.234	0.00138	0.1125
	TR37304	770	0	10	5.1	0.00001	0.1125
	TR50351	649	37	9	16.522	0.0207	0.1125
	marf	788	54	46	29.996	0.01779	0.1125
	mof	576	7	47	0.889	0.00275	0.1125
	pgd	565	18	42	4.356	0.0074	0.1125
	rpl	302	4	44	0.565	0.00304	0.1125
	spk	382	14	45	1.04	0.0006	0.1125
ST	x-pen	1283	48	7	22.013	0.01622	0.6375
	TR261	627	25	10	12.25	0.01527	0.6375
	TR11103	826	10	10	4.556	0.00713	0.6375
	TR23125	343	8	9	4.167	0.00843	0.6375
	TR24932	280	10	8	7.234	0.01717	0.6375
	TR37304	770	4	10	5.1	0.00184	0.6375
	TR50351	649	50	19	16.522	0.02832	0.6375
	marf	788	186	57	29.996	0.05808	0.6375
	mof	576	30	65	0.889	0.01098	0.6375
	pgd	565	38	56	4.356	0.01541	0.6375
	rpl	302	11	57	0.565	0.0079	0.6375
	spk	382	13	56	1.04	0.00741	0.6375