CHROMOSOMAL INVERSIONS FACILITATE THE ACCUMULATION OF DIVERGENCE AND HYBRID INCOMPATIBILITIES BETWEEN SPECIES

by

Christopher Leonard

A thesis submitted to the faculty of
The University of Utah
in partial fulfillment of the requirements for the degree of

Master of Science

Department of Biology

The University of Utah

December 2017

The University of Utah Graduate School

STATEMENT OF THESIS APPROVAL

The thesis of **Christopher Leonard**

has been approved by the following supervisory committee members:

Nitin Phadnis	, Chair	**09/28/2017** Date Approved
Michael Shapiro	, Member	**10/02/2017** Date Approved
Kent Golic	, Member	**09/29/2017** Date Approved

and by **M. Denise Dearing** , Chair/Dean of

the Department/College/School of **Biology**

and by David B. Kieda, Dean of The Graduate School.

ABSTRACT

The chromosomal inversions of *D. persimilis* and *D. pseudoobscura* have deeply influenced our understanding of the evolutionary forces that shape natural variation, speciation, and selfish chromosome dynamics. Here, we perform a comprehensive reconstruction of the evolutionary histories of the chromosomal inversions in these species and provide a solution to the puzzling origins of the selfish *Sex-Ratio* chromosome in *D. persimilis*. We show that this *Sex-Ratio* chromosome directly descends from an ancestrally-arranged chromosome, suggesting that unsuppressed selfish chromosomes may remain polymorphic within populations for long periods of time. We further show that all fixed inversions between *D. persimilis* and *D. pseudoobscura* were segregating in the ancestral population long before speciation, and that the genes contributing to reproductive barriers must have evolved within them afterwards. We propose a new model for the role of chromosomal inversions in speciation and suggest that higher levels of divergence and an over-abundance of hybrid incompatibilities are emergent properties of ancestrally segregating inversions. Our findings force a reconsideration of the role of chromosomal inversions in speciation, not as a protector of existing hybrid incompatibility alleles, but as fertile ground for their formation.

TABLE OF CONTENTS

LIST OF TABLES

Tables

LIST OF FIGURES

Figures

ACKNOWLEDGMENTS

It’s hard to do science without bright thinkers around you. I first want to thank our collaborators, without whom this work would not have happened. I want to thank the people in my lab who have really enabled me to think critically and broadly, especially Jacob, Nora, and Jim. I want to thank my friends and family for their continued support and for dealing with my antics. I have had the blessing of having many mentors who have taught me to be a capable scientist, a professional employee, and a more decent human being: I cannot thank you enough. Lastly, I want to thank my advisor Nitin Phadnis, who believed in me from the beginning and who has built a highly creative and integrative lab where I have been able to grow. I’ve been so lucky to spend the last few years there.

CHAPTER 1

INTRODUCTION

Dobzhansky and Sturtevant's key insight that chromosomal inversions can be used as a direct readout of natural genetic variation revolutionized evolutionary biology, and played a critical role in the unification of genetics and evolutionary theory.[1,2] This insight powered a highly influential body of work centered on the polymorphic, fixed, and *Sex-Ratio* inversions of *Drosophila persimilis* and *D. pseudoobscura*.[3]

The inversions of *D. persimilis* and *D. pseudoobscura*

First, studies on the polymorphic inversions on the third chromosomes of *D. persimilis* and *D. pseudoobscura* led to the development of the first genetics-based phylogenetic reconstruction of the evolutionary history of the different inversion polymorphisms, and pioneered investigations of genetic variation across geographical distributions and timescales.[2] Ongoing genomic analyses of these polymorphisms continue to reveal how the interplay between mechanisms of local adaptation, selection, recombination, and epistasis govern the patterns of genetic variation in natural populations.[4–6] Together, these studies have transformed our understanding of the general principles that govern the patterns of natural genetic variation.

Second, studies on the genetic basis of reproductive isolation between *D. persimilis* and *D. pseudoobscura* led to the development of the Dobzhansky-Muller

model of the evolution of hybrid incompatibilities, and provided a framework for understanding speciation.[7–11] Recent studies have led to the development of two new models for the role of chromosomal inversions in speciation.[12–17] These models attempt to explain two empirical patterns observed in this hybridization: i) nearly all genes that contribute to reproductive isolation are located among fixed chromosomal inversions between these species, and ii) these fixed chromosomal inversions display higher genetic divergence than collinear regions of the genome. Both models, which rely on a persistent history of gene flow either during secondary contact after speciation (Noor-Reiseberg model) or during speciation (Navarro-Barton model), can sufficiently explain both empirical patterns. As a result, the idea that chromosomal inversions may contribute to speciation has seen a dramatic resurgence.[18–21]

Third, studies on the *Sex-Ratio* inversions in *D. persimilis* and *D. pseudoobscura* have provided insights into the mechanisms of meiotic drive, the population dynamics of selfish chromosomes, the fitness components that lead to protected polymorphisms, and the role of polyandry in countering selfish genetic elements.[22–25] Males that carry *Sex-Ratio* inversions on the X-chromosome eliminate nearly all *Y*-bearing sperm,[26] produce nearly all female offspring, and heavily distort progeny sex ratios (therefore called *Sex-Ratio* chromosomes). By tipping the balance of segregation in their favor in excess of Mendelian expectations, these selfish *X*-chromosomes can rapidly spread through populations even if they reduce the fitness of the individuals that carry them.[27,28] In the absence of opposing forces such as the evolution of suppressor alleles, *Sex-Ratio* chromosomes may even drive populations extinct.[29] The *Sex-Ratio* chromosomes of *D. persimilis* and *D. pseudoobscura*, however, represent enigmatic cases of unsuppressed distorting systems that are stably maintained within populations. These studies on *Sex-*

Ratio chromosomes played a critical role in bringing forth the role of meiotic drive as a potent evolutionary force.[30,31]

The strange collinearity between *D. persimilis SR* and *D. pseudoobscura*

The *Sex-Ratio* chromosome of *D. persimilis* presents an enigmatic case. *Sex-Ratio* (*SR*) chromosomes have been identified in many Dipteran species, and are almost always associated with one or more chromosomal inversions relative to the wild type or *Standard* (*ST*) chromosomes. When a new inversion generates tight linkage between a segregation distorter and alleles that enhance distortion (or alleles that neutralize suppressors-of-distortion), this produces a stronger driving chromosome that can supplant its weaker versions.[32] This explains why most *Sex-Ratio* chromosomes are associated with one or more derived inversions. The *D. persimilis SR* chromosome is also inverted with respect to the *D. persimilis ST* chromosome on the right arm of the *X* chromosome (*XR*). The *Standard D. persimilis* also differs from *D. pseudoobscura* by a single derived inversion. Surprisingly, however, the *D. persimilis SR* inversion appears to have reversed the same derived *D. persimilis ST* inversion, such that *D. persimilis SR* now appears collinear with *D. pseudoobscura* (Figure 1.1). This strange collinearity is thought to be the result of a second inversion on the background of *D. persimilis ST* at approximately the same breakpoints as the original *D. persimilis* inversion. However, previous molecular evolutionary studies have yielded conflicting results, and the origin of the *D. persimilis Sex-Ratio* inversion remains the subject of speculation.[33–35]

Figure 1.1: A phylogeny of the *X* chromosomes of *D. persimilis SR*, *ST*, and *D. pseudoobscura*. The *D. persimilis SR* chromosome is collinear across species boundaries with *D. pseudoobscura*.

The evolutionary history of the inversions of *D. pseudoobscura* and *D. persimilis*

The inversions of *D. persimilis* and *D. pseudoobscura* have played an outsized influence on current evolutionary thought, and reconstructing their correct evolutionary history has important consequences for understanding the dynamics of selfish genes and speciation. Here, we sequenced the full genomes of *ST* and *SR* strains in *D. persimilis* and several strains of *D. pseudoobscura* to investigate the evolutionary history of chromosomal inversions in these two closely related species. First, we identify the inversion breakpoints on *XR* and find that the *D. persimilis SR* chromosome is precisely collinear with *D. pseudoobscura ST*. Moreover, the regions flanking the *D. persimilis SR* inversion breakpoints display phylogenetic discordance in the form of being more closely related to the *D. pseudoobscura*, rather than to the *D. persimilis ST* chromosome. Our results demonstrate that the *D. persimilis SR* chromosome arose, not from a second inversion event, but directly from the ancestrally-arranged chromosome. Second, we show that this phylogenetic discordance is not due to the result of gene flow between the two species and demonstrate that the patterns of discordance at the *SR* chromosome breakpoints are the result of incomplete lineage sorting of the derived *D. persimilis ST* inversion from the ancestor of *D. pseudoobscura* and *D. persimilis*. Third, by estimating the absolute divergence between the two collinear chromosomes, we demonstrate that the *D. persimilis SR* chromosome is a long-term segregating polymorphism that predates the species divergence between *D. persimilis* and *D. pseudoobscura*. Lastly, by estimating divergence in regions surrounding inversion breakpoints, we show that all of the fixed rearrangements between *D. persimilis* and *D. pseudoobscura* arose in the ancestor of the two species, but were passed exclusively to *D. persimilis*. Together, our results challenge the current understanding of the evolutionary history of these inversions, present

evidence of the long-term maintenance of *Sex-Ratio* chromosome polymorphisms, and suggest a new model for the role of chromosomal inversions in speciation.

CHAPTER 2

MATERIALS AND METHODS

Isolation and maintenance of *Sex-Ratio* chromosome strains

D. persimilis strains were provided as a generous gift by Dean Castillo, collected in the Sierra Nevada mountain range and Mt. St. Helena, CA. We tested individuals from these strains for the presence of *Sex-Ratio* chromosomes by crossing males to standard *D. persimilis* females. We isolated two individual *D. persimilis Sex-Ratio* strains and generated stable stocks through eight to twelve generations of inbreeding. All stocks were raised on standard cornmeal media at 18 degrees C.

Polytene chromosome squashes

We used two crosses of *D. persimilis SR*/*ST* heterozygotes to compare the *D. persimilis SR* chromosome with *D. pseudoobscura* and *D. persimilis ST* chromosomes. In the first cross, a *D. persimilis SR*/*ST sepia* heterozygous female was crossed to a *D. pseudoobscura ST se* male. Of the two *XL*/*XR* karyotypes possible from this cross, we examined females heterozygous for *XL* and homozygous for *XR* inversions. These females allow us to evaluate whether the *D. persimilis SR* and *D. pseudoobscura ST* chromosomes are homosequential. In a second cross, a *D. persimilis SR*/*ST* sepia heterozygous female was crossed to a *D. persimilis ST se* male. Of the two *XL*/*XR*

karyotypes possible from this cross, we examined females homozygous for *XL* and heterozygous for *XR* inversions. These females allow us to examine the *D. persimilis SR* and *D. persimilis ST* heterozygotes. We prepared salivary squashes from larvae from these two crosses using standard techniques, with modifications described by Harshman (1977) and Ballard and Bedo (1991).[36–38]

DNA extractions and sequencing

To generate whole genome shotgun sequencing libraries for *D. persimilis* strains, we pooled one male each from two *SR* strains and two *ST* strains (from Sierra Nevada and Mt. St. Helena collections). We extracted DNA from these flies using the 5 Prime Archive Pure DNA extraction kit according to the manufacturer's protocol (ThermoFisher, Waltham, MA). All libraries were generated with the Illumina TruSeq Nano kit (Epicentre, Illumina Inc, CA) using the manufacturers protocol, and sequenced as 500bp paired end reads on an Illumina HiSeq 2000 instrument.

Sequence alignment and SNP identification

Low-quality bases were removed from the ends of the raw paired end reads contained in FASTQ files using *seqtk* (https://github.com/lh3/seqtk) with an error threshold of 0.05. Illumina adapter sequences and polyA tails were trimmed from the reads using Trimmomatic version 0.30.[39] The read quality was then manually inspected using FastQC. Following initial preprocessing and quality control, the reads from each pool were aligned to the *D. pseudoobscura* reference genome (v 3.2) using *bwa* version 0.7.8 with default parameters.[40] Genome wide, the average fold coverage was ~180x and

~133x for the *D. persimilis ST* and *SR* pools, respectively. For reads mapping to X chromosome scaffolds, the average fold coverage was ~97x and ~74x for *D. persimilis ST* and *SR*, respectively.

After the binary alignments were sorted and indexed with SAMtools,[41] single nucleotide polymorphisms (SNPs) were called using *freebayes* (v. 0.9.21)[42] with the expected pairwise nucleotide diversity parameter set to 0.01, based on a previous genome-wide estimate from *D. pseudoobscura*.[43] The samples were modeled as discrete genotypes across pools by using the "–J" option and the ploidy was set separately for *X* chromosome scaffolds (1*N*) and autosomes (2*N*). SNPs with a genotype quality score less than 30 were filtered from the dataset. We restricted all downstream analyses to sites that had coverage greater than *1N* and less than 3 standard deviations away from the genome-wide mean for all samples (Table S1). Across the genome, we identified a total of 3,598,524 polymorphic sites, 703,908 and 844,043 of which were located on chromosomes *XR* and *XL*, respectively.

The *D. pseudoobscura* reference assembly does not contain complete sequences for either of the arms of the *X* or *4*th chromosomes. Instead, each is composed of a series of scaffold groups that differ both in size and orientation relative to one another.[44] Schaeffer et al. (2008) previously determined the approximate locations and ordering of each of these scaffolds.[44] We used their map to convert the scaffold-specific coordinates of each site to the appropriate location on the corresponding chromosome to construct a continuous sequence.

Estimating the phylogenetic relationship of *Sex-Ratio* chromosomes

We estimated the genetic distance between each pairwise grouping in 10 kb windows using Nei's D_A distance, which has been shown to accurately recover the topology of phylogenetic trees from allele frequency data.[45,46] To root the tree with an outgroup, we aligned publically available short reads of *D. miranda* (SRX965461; strain SP138) to the *D. pseudoobscura* reference genome. In each window, we constructed neighbor-joining trees [47] using distance matrices constructed from the estimated genetic distances (D_A) and classified the phylogeny based on the topology it supported. If a window contained fewer than 10 segregating sites, we did not construct a tree or estimate the genetic distance. For each tree, we performed 10,000 bootstrap replicates and only included those windows with a support value of 0.75 or higher.

Divergence estimates

We estimated absolute divergence with Nei's d_{xy}, a measure of the average number of pairwise nucleotide substitutions per site.[48] d_{xy} was measured between each population grouping in 10 Kb, nonoverlapping windows across the genome. To convert estimates of absolute divergence into divergence times, the d_{xy} values were scaled to a 2 My species split between *D. pseudoosbcura* and *D. miranda* in each window.

Identification and verification of inversion breakpoints

The proximal and distal breakpoints have both been characterized previously, and the regions in *D. pseudoobscura* contain unique sequence flanking a series of 302-bp repeats known as Leviathan repeats, present throughout the genomes of both *D.*

pseudoobscura and *D. persimilis*. We designed primers to capture both the array of repeats as well as portions of unique sequence. We extracted DNA from all three genotypes and amplified the proximal breakpoint region using primers designed to anneal to the *D. pseudoobscura* genomic sequence flanking the Leviathan repeats (F5'-GATCTAATCCAGAAAGTTCGCTTGCG -3', R5'-AGTGTGACCCATTTTAAGCGG-3'). These primers amplified a single, approximately 1500bp, product in *D. pseudoobscura* and *D. persimilis SR*, but not *D. persimilis ST* (Figure 2.1). PCR products were Sanger sequenced using the forward and reverse PCR primers at the DNA Sequencing Core Facility, University of Utah. The reads were aligned both to one another and to sequence from the *D. pseudoobscura* genome assembly around the proximal breakpoint. The sequenced PCR product was confirmed to contain both the repeats and sections of unique sequence flanking the repeat region at the proximal breakpoint.

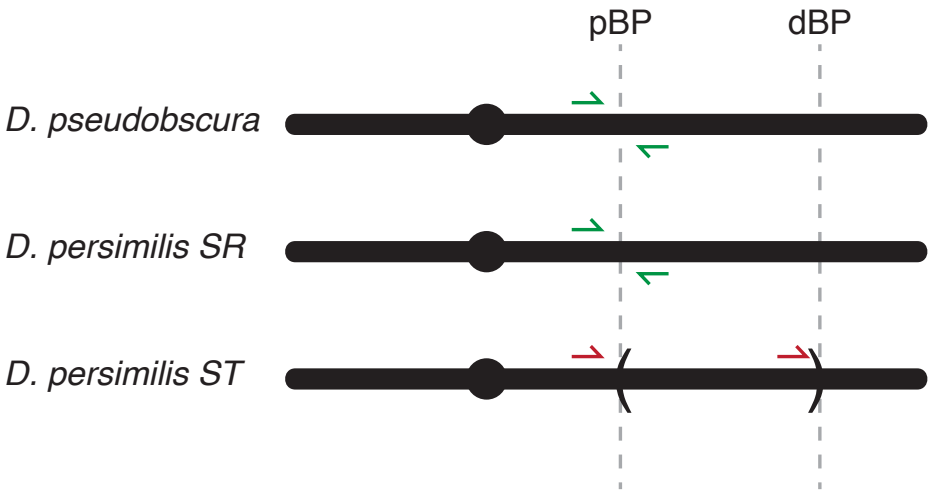

Figure 2.1: Primer design strategy for amplifying the proximal breakpoint. Primers that bridge the breakpoint can amplify the breakpoint only if the chromosome is in the ancestral orientation, as in *D. pseudoobscura* and *D. persimilis SR*. The primers are not in the correct orientation to produce an amplicon in *D. persimilis ST*.

CHAPTER 3

RESULTS

A high-resolution examination of polytene chromosomes confirms the apparent collinearity of *D. persimilis Sex-Ratio* with *D. pseudoobscura*

To uncover the evolutionary origins of the *D. persimilis SR* chromosome, we screened for the *Sex-Ratio* trait in wild caught *D. persimilis* flies. We isolated two independent *D. persimilis SR* strains that produce >90% female progeny, and generated high-quality mosaic images of polytene chromosomes with squashes of larval salivary glands. Consistent with previous reports,[30] the *D. persimilis SR* chromosomes in the strains that we isolated differ by one major inversion on *XR* with respect to *D. persimilis ST*, but appear collinear with *D. pseudoobscura* (Figure 3.1). If *D. persimilis SR* was derived from *D. persimilis ST* through a somewhat imprecise reversion to the ancestral arrangement, the banding patterns of polytene chromosomes in *D. persimilis SR*/*D. pseudoobscura* female hybrids may reveal slight imperfections near the inversion breakpoints. Even under close examination, we did not observe any disruption of chromosome pairing near the inversion breakpoints in *D. persimilis SR/D. pseudoobscura* heterozygotes, suggesting that any secondary inversion event may have been in close vicinity of the original breakpoints of the *D. persimilis ST* inversion.

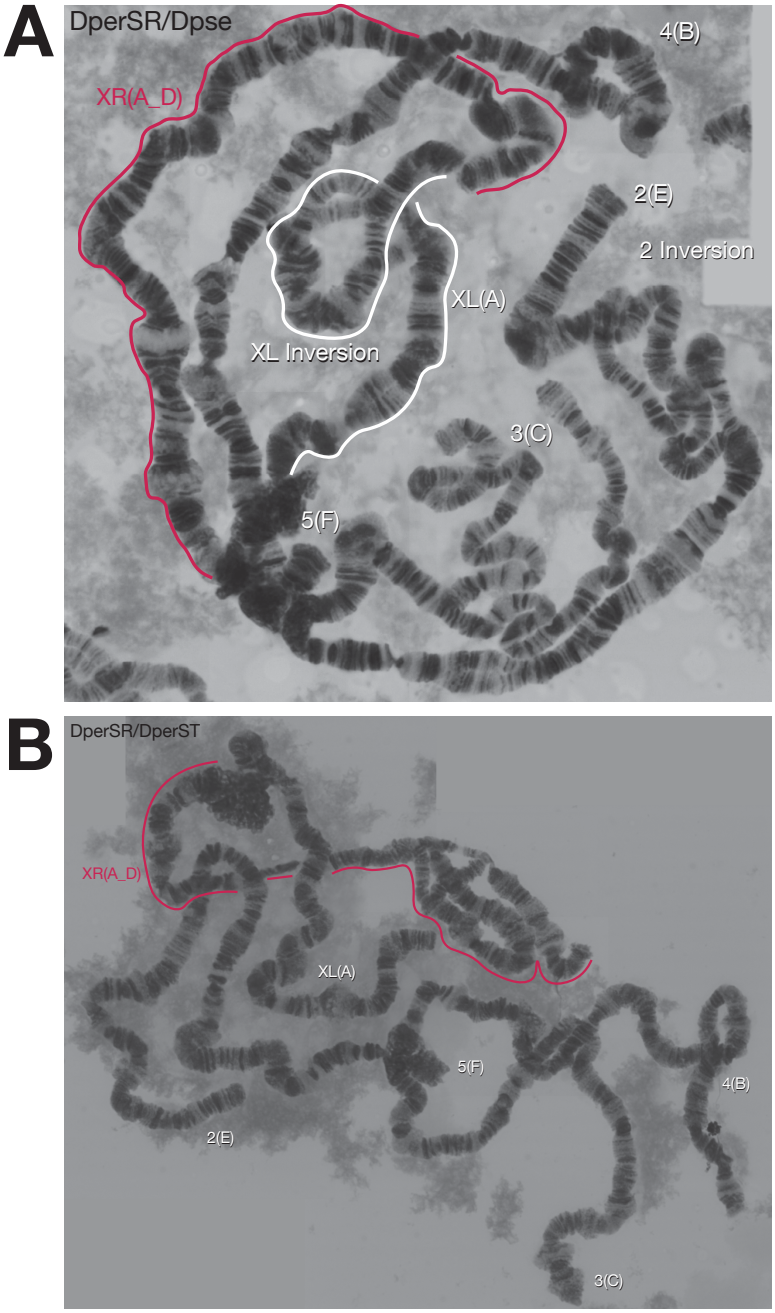

Figure 3.1: Polytene chromosome squashes show *D. persimilis SR* is collinear with *D. pseudoobscura* but not *D. persimilis ST*. While *D. persimilis SR* is collinear on the right arm of the *X* chromosome, the characteristic fixed inversions can be seen on the *XL* and *2*nd chromosomes in *(A)*, but these chromosomes are collinear in *(B)*.

D. persimilis Sex-Ratio and *D. pseudoobscura* are precisely collinear at a single base pair resolution

While our polytene analyses showed no visible aberrations at the breakpoints of the *D. persimilis* inversion, such analyses provide only a coarse view of chromosome structure. Previously, the *D. persimilis ST* inversions breakpoints were mapped at a resolution of 30kb.[15] To precisely identify the inversion breakpoints on the *D. persimilis SR* chromosome, we first performed whole genome sequencing of males pooled from two *D. persimilis SR* strains, as well as males pooled from two *D. persimilis ST* strains. Using the approximate genomic coordinates of the inversion breakpoints, we designed multiple primer pairs that span the proximal and distal inversion breakpoint sequences from *D. persimilis SR* and *D. pseudoobscura*. We performed PCR with these primers to successfully amplify single products using *D. persimilis SR* and *D. pseudoobscura* genomic DNA as templates. We were able to amplify sequences corresponding to the proximal breakpoint, but not the distal breakpoint. We identified the precise molecular breakpoints of this inversion by Sanger sequencing the proximal breakpoint PCR products, which revealed the presence of four 319bp *Leviathan* repeats [49] at the breakpoint. More importantly, *D. persimilis SR* and *D. pseudoobscura* sequences that flank the *Leviathan* repeats are precisely collinear to a single base pair resolution (Figure 3.2). These results show that a slightly staggered second inversion event is not the basis for the collinearity between the *D. persimilis SR* and *D. pseudoobscura* chromosomes.

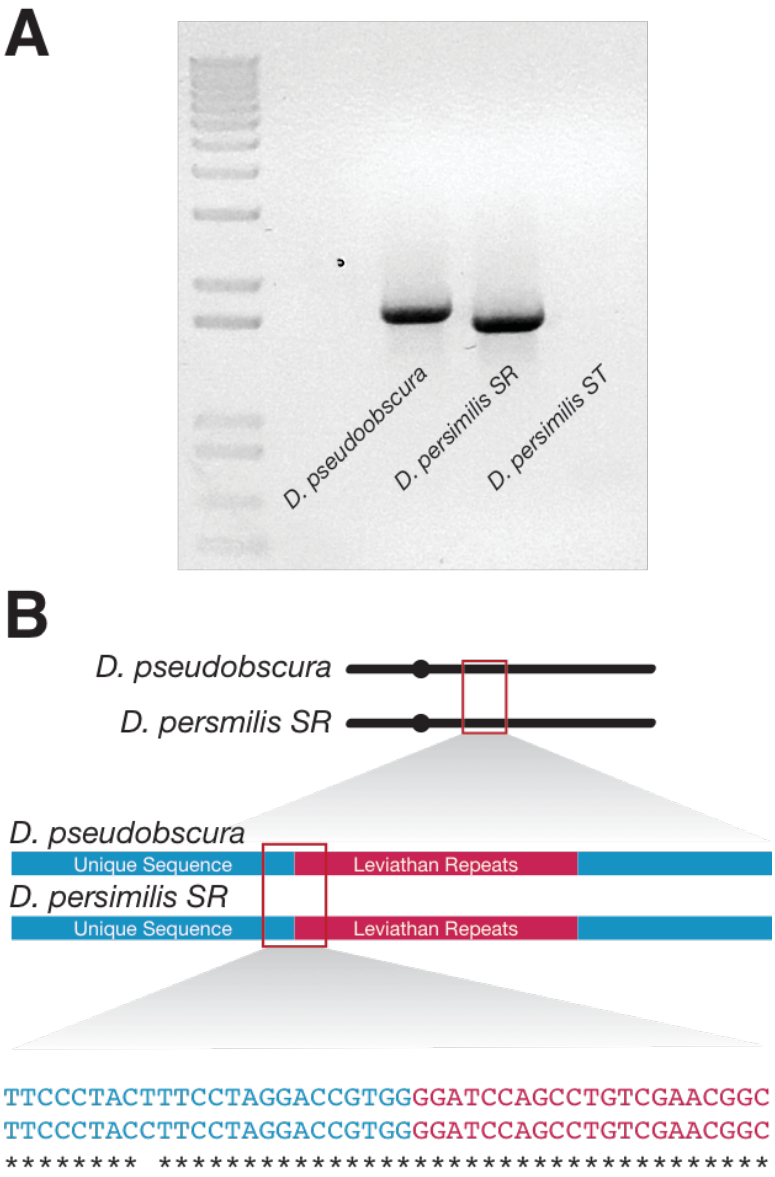

Figure 3.2: *D. persimilis SR* is collinear with *D. pseudoobscura* at the base pair level. This primer pair produced amplicons from *D. pseudoobscura* and *D. persimilis SR* genomic template, but not *D. persimilis ST* (A). We sequenced this amplicon and find that the sequence at the inversion breakpoint region is collinear between *D. persimilis SR* and *D. pseudoobscura*.

The *D. persimilis Sex-Ratio* chromosome is more closely related to *D. pseudoobscura* than to *D. persimilis* at the inversion breakpoints

If the *D. persimilis SR* inversion originated through a recombination event within *Leviathan* sequences at the inversion breakpoints, such an event can generate the same pattern of perfect collinearity of the flanking sequences. Such repetitive sequences are known to be hotspots for inversion breakpoints.[49,50] While *Leviathan* repeats are unique to *D. persimilis* and *D. pseudoobscura*, there are more than 850 of these repeats spread across their genomes. Because *XR* alone harbors more than 650 Leviathan repeats spread across the chromosome arm, the probability of a second inversion event on *D. persimilis SR* at the same two *Leviathan* repeats as the original breakpoints appears vanishingly small. However, to directly test whether *D. persimilis SR* is recently derived from *D. persimilis ST* through a secondary inversion event, we constructed phylogenies in sliding windows across the chromosome using *D. miranda* as an outgroup. As expected, *D. persimilis SR* sequences cluster with those from *D. persimilis ST* across nearly the entire genome (Figure 3.3). Surprisingly, we find two large blocks of phylogenetic discordance concentrated at the inversion breakpoints on *XR*. In these recombination-limited regions of phylogenetic discordance, *D. persimilis SR* sequences are more closely related to *D. pseudoobscura* rather than to *D. persimilis ST*, with several regions within the inversion also showing the same discordant pattern (Figure 3.3). Together with the precise collinearity of *D. persimilis SR* and *D. pseudoobscura,* these results support a single origin of the arrangements of these two chromosomes.

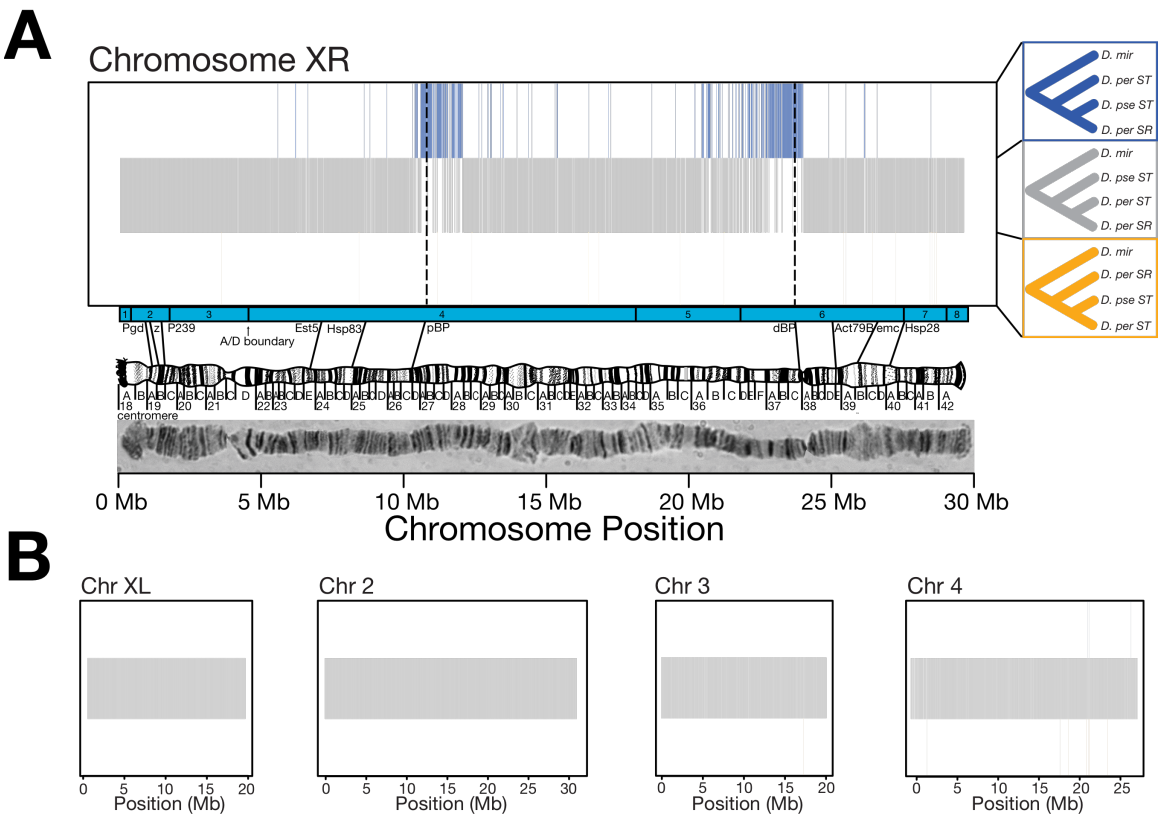

Figure 3.3: The inversion breakpoints on *XR* show extensive phylogenetic discordance. (A) Sliding window phylogeny classification on *XR*. Blue, grey, and orange vertical lines represent the tree topology supported by neighbor-joining trees. Grey trees represent no phylogenetic discordance. Blue trees represent regions where the two collinear chromosomes appear more similar. Large regions centered on the proximal and distal breakpoints (dashed lines) of the *XR* inversion show discordant clustering of *D. persimilis SR* with *D. pseudoobscura* rather than *D. persimilis ST*. (B) Large regions of phylogenetic discordance are not observed in the remainder of the genome.

Phylogenetic discordance is a specific property of the *D. persimilis Sex-Ratio* chromosome

We next asked whether the phylogenetic discordance that we observed with the *D. persimilis SR* chromosome is a general property of chromosomal inversions between *D. persimilis* and *D. pseudoobscura*. There are two other fixed inversion differences on the *XL* and *2nd* chromosomes between the two species. Similar to the *XR* inversion, the un-inverted *XL* and *2nd* chromosomes of *D. pseudoobscura* represent the ancestral state and the inverted *D. persimilis* chromosomes represent the derived state. Our sliding window phylogenetic analyses show that the sequences at the breakpoints of these two fixed inversions show no phylogenetic discordance and recapitulate the species tree. Although these analyses also revealed small regions of phylogenetic discordance in other regions of the genome, there is no clustering of consecutive windows showing this discordant pattern and discordant windows are not associated with inversions. These results demonstrate that the pattern of phylogenetic discordance is not a general feature of inversion differences between these species, but instead is a specific property of the *D. persimilis SR* chromosome (Figure 3.3).

D. persimilis and *D. pseudoobscura* are known to rarely, but successfully, produce hybrids in nature.[51] Gene flow across species may generate the same pattern of phylogenetic discordance we observe around the *D. persimilis XR* inversion if the species share a chromosomal arrangement that is also polymorphic within each species. We were able to test this idea because, like the *Standard* arrangement of *XR*, the *Standard* arrangement on the *3rd* chromosome (*3ST*) is both shared across *D. persimilis* and *D. pseudoobscura*, and is polymorphic within each species.[52] Using *3ST* from both

species and the *Arrowhead* (3^{AR}) arrangement of *D. pseudoobscura*, we performed the same phylogenetic analysis across the 3^{rd} chromosome. Sequences at the breakpoints of this shared polymorphic inversion recapitulate the correct species tree, again indicating that the large blocks of phylogenetic discordance at the inversions breakpoints on *XR* is a unique property of the *D. persimilis SR* chromosome (Figure 3.4).

Together with the precisely-shared breakpoints, the relatedness between *D. persimilis SR* and *D. pseudoobscura* at the inversion breakpoints rejects the currently accepted secondary-inversion hypothesis for the origin of the *D. persimilis SR* arrangement, and suggests a single origin for these chromosomes. Our results raise the surprising possibility that *D. persimilis SR* was derived either through a recent introgression event from *D. pseudoobscura*, or from the common ancestor of *D. persimilis* and *D. pseudoobscura*.

The *D. persimilis SR* inversion is derived from the ancestor of *D. persimilis* and *D. pseudoobscura*

Because *D. persimilis* and *D. pseudoobscura* can hybridize in nature,[51] our results raise the possibility that *D. persimilis SR* originated as a recent introgression of *D. pseudoobscura XR*. Under the introgression scenario, repeated back-crossing to *D. persimilis* after the initial hybridization event gradually removes *D. pseudoobscura* material through single crossovers outside the inversion, and through double crossovers or gene conversion events inside the inversion. These recombination events homogenize *D. persimilis SR* and *ST*, largely wiping out any hints of a potential cross-species origin of *D. persimilis SR* from *D. pseudoobscura*. However, this history of introgression

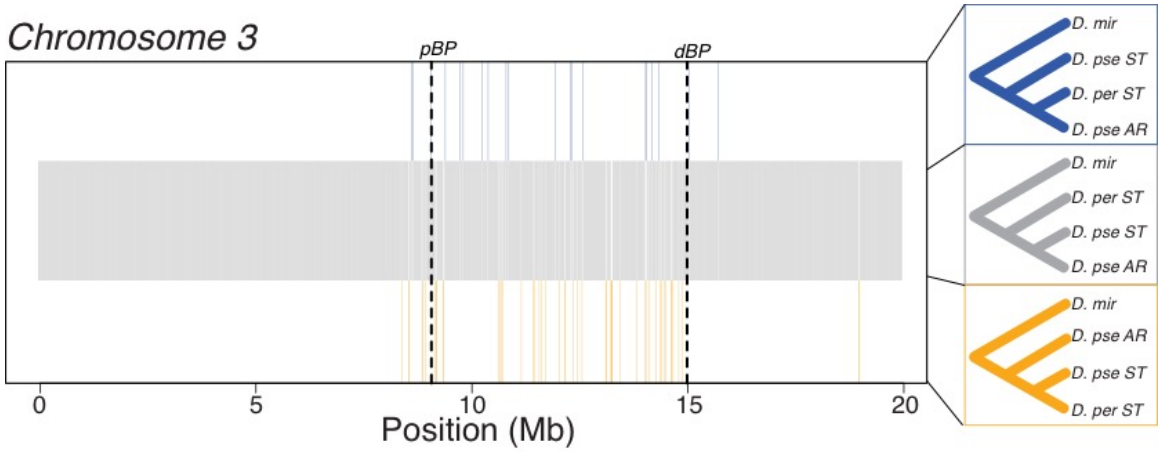

Figure 3.4: Species clustering within inversion polymorphisms on chromosome 3. The *D. pseudoobscura 3*rd chromosome arrangements Standard (*ST*) and Arrowhead (*AR*) lack the large breakpoint-specific phylogenetic discordance observed at the inversion break points of the inversion between *D. pseudoobscura* and *D. persimilis SR* on chromosome *XR*. While some windows demonstrate phylogenetic discordance, these windows are independent of the arrangement of the chromosome forms and, unlike the *XR* inversion, do not cluster at the inversion breakpoints.

would be best preserved at the breakpoints of the inversion where suppression of crossovers is greatest.[53,54] The preservation of *D. pseudoobscura* material at the inversion breakpoints would then generate the blocks of phylogenetic discordance we observe on *D. persimilis SR*. The modified f_d statistic is a broadly used and effective test to discriminate between introgression versus incomplete lineage sorting (ILS), similar to related "ABBA-BABA" measures.[55] We analyzed our genomic data from *D. persimilis SR* and *ST*, along with *D. pseudoobscura* and *D. miranda* sequences, to estimate the modified f_d across the entire genome. Indeed, we observed significant f_d between *D. pseudoobscura* and *D. persimilis SR* at the same chromosomal inversion breakpoint regions that show phylogenetic discordance (Figure 3.5).

Interpreting significant values of f_d and related statistics as introgression involves an implicit assumption of free recombination in the ancestral population. However, in regions of limited recombination, such as when inversions segregate in the ancestral population, it is incorrect to conclude introgression from the results of these statistical tests. Because the *D. persimilis SR* chromosome involves a chromosomal inversion that was potentially segregating in the ancestral population, this violates the assumptions required to reliably conclude introgression from the f_d statistic alone.[56] The interpretation of introgression based on the modified f_d statistic in this context may, therefore, be premature and instead may be the result of incomplete lineage sorting (ILS). Indeed, it is not clear if any of the existing statistical approaches can effectively discriminate between introgression and ILS to determine the ancestry of chromosomal inversions.

Moreover, an alternative model that involves the inheritance of the *D. persimilis*

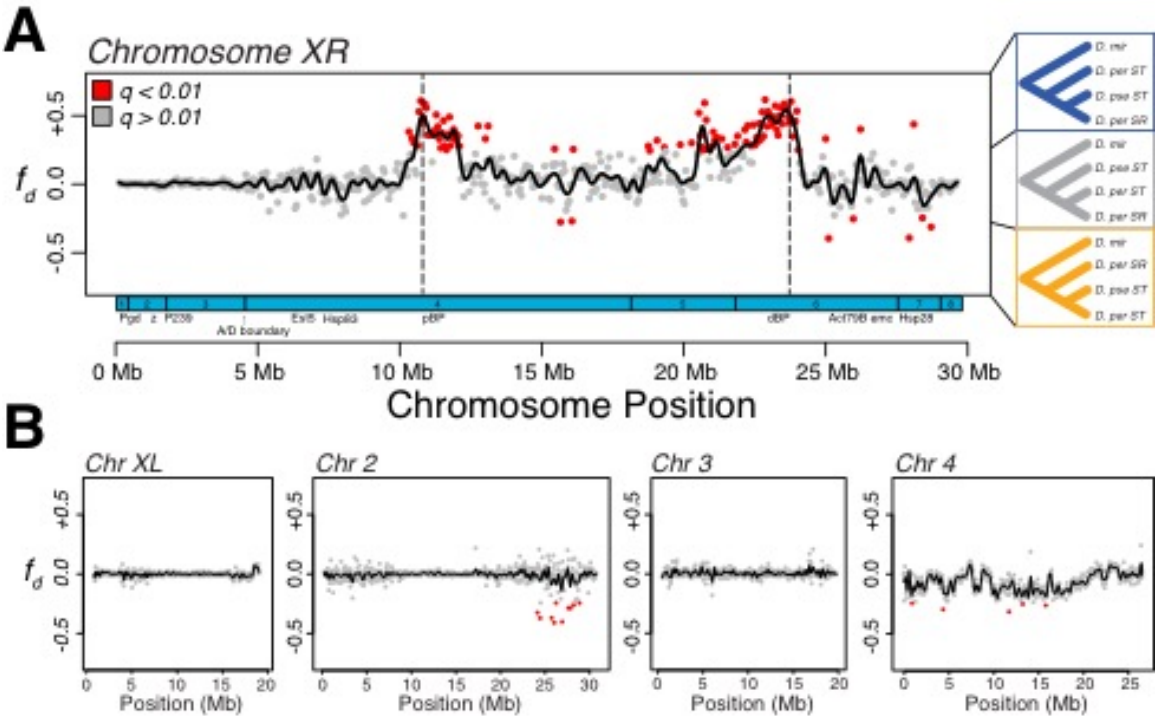

Figure 3.5: Significant f_d signature at phylogenetically discordant breakpoints. (A) We observe large regions with significant f_d at both of the inversion breakpoints on the *XR* chromosome. (B) In contrast, we observe fewer and less concentrated windows with significant f_d through the remainder of the genome.

SR and *D. pseudoobscura ST* arrangements from the common ancestor of both species can adequately explain the observed patterns. In particular, the phylogenetic discordance that we observe can be explained by the inheritance of the *D. persimilis SR* arrangement from the ancestor of *D. persimilis* and *D. pseudoobscura*, in combination with ILS of the *D. persimilis ST* chromosome (Figure 3.6). Under the ILS scenario, the *D. persimilis ST* inversion originates as a freely segregating polymorphic chromosome in the ancestral population of *D. persimilis* and *D. pseudoobscura*. The recombination-suppressed regions at the breakpoints of the *D. persimilis ST* inversion begin diverging from the ancestrally-arranged chromosomes long before speciation. During this time, the ancestor of *D. persimilis SR* and *D. pseudoobscura* ST chromosomes continue to freely recombine until the time of speciation, but diverge from the derived inverted *D. persimilis* ST chromosome. Similar to the introgression scenario, recombination events homogenize *D. persimilis SR* and *ST* after speciation, except at the breakpoints of the inversion, thus leading to the patterns of phylogenetic discordance.

We reasoned that the same recombination-suppressing properties of chromosomal inversions that thwart the application of standard statistical approaches may also preserve the information necessary to discriminate between introgression and ILS. In particular, because recombinants at the sequences in the regions near the inversion breakpoints are less frequent, the divergence of the chromosomes can be reliably estimated using these sequences. The introgression and ILS hypotheses make distinct and testable predictions about the relative divergence times of each chromosomal arrangement. Under the introgression scenario, we expect the *D. persimilis SR* chromosome to appear much younger than the species divergence time

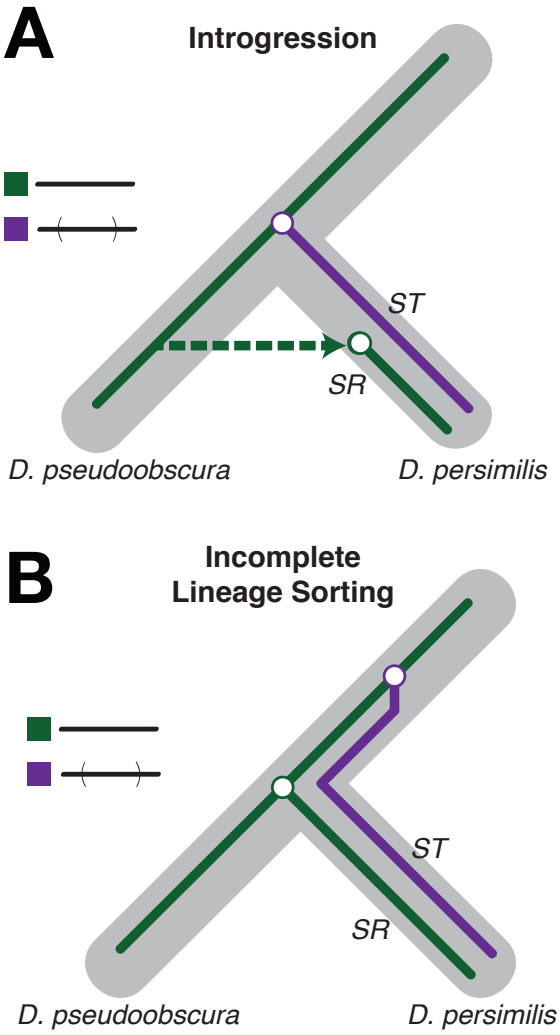

Figure 3.6: Discordance may be produced by introgression or incomplete lineage sorting of the *XR* arrangements**.** Under model *(A)*, the *D. persimilis ST* inversion segregates in the ancestral population of the species. Later divergence between *D. persimilis SR* and *D. pseudoobscura* chromosomes and recombination restriction between the two *D. persimilis* chromosomes leads to phylogenetic discordance at the inversion breakpoints. *(B)* An introgression model again predicts discordance if the *D. persimilis SR* chromosome introgressed from *D. pseudoobscura* after species divergence. Recombination between the introgressed chromosome *and D. persimilis ST* will gradually homogenize the two chromosomes excluding the inversion breakpoints.

due to a more recent coalescence at introgressed loci. In contrast, the ILS scenario makes two distinct predictions. First, we expect the *D. persimilis SR* chromosome to be as old or older than the species divergence time. Second, we expect the *D. persimilis ST* chromosome to be much older than the species divergence time and more diverged from the *D. persimilis SR* chromosome near inversion breakpoints. To test these predictions, we estimated the absolute divergence (d_{xy}) in 10 kb windows between *D. persimilis* and *D. pseudoobscura* in all collinear regions across the genome and observed a mean d_{xy} of 2.42×10^{-3} (95% CI: $2.37 - 2.47 \times 10^{-3}$). When standardized to a *D. miranda* divergence set to 2 million years in each window, this corresponds to a speciation time between *D. persimilis* and *D. pseudoobscura* of approximately 500,000 years ago. We used the sequences flanking the inversion breakpoints (± 250 kb) to estimate d_{xy} between *D. persimilis SR* and *D. pseudoobscura* and observe a significantly higher ($p < 2.2 \times 10^{-16}$) mean divergence (4.55×10^{-3}; 95% CI: $4.22 - 4.89 \times 10^{-3}$) than estimated between species in collinear regions, indicating the *D. persimilis SR* chromosome is older than the speciation time (Table 3.1). When similarly standardized to the *D. miranda* divergence in each window flanking the breakpoints, we estimate the *D. persimilis SR* chromosome to be ~1.09 million years old. This is inconsistent with the introgression scenario, and suggests that ILS may better describe the evolutionary origins of the *D. persimilis SR* chromosome. Moreover, the ILS hypothesis makes a second prediction that the *D. persimilis ST* chromosome should be older than the divergence time between the two species. Indeed, we estimate d_{xy} between *D. persimilis ST* and *D.pseudoobscura* in the same sequences flanking the breakpoint regions (d_{xy}: 4.94×10^{-3}; 95% CI: $4.67 - 5.21 \times 10^{-3}$) to be significantly greater ($p < 0.038$) corresponding to an older standardized

Table 3.1: The fixed inversions between *D. persimilis* and *D. pseudoobscura* were segregating prior to speciation. The order of the inversions between the two species is concordant with previous estimates. The *XL* inversion is the oldest, followed by the *2nd* inversion, and lastly the *XR* inversion, which still predates species divergence by 730,000 years.

Divergence Time Estimates

	Species divergence	*DperSR* divergence	*Chr XR* inversion	*Chr 2* inversion	*Chr XL* inversion	*Chr 3 PP* inversion	*Chr 3 AR* inversion	*Chr 3 AR* divergence
Divergence time (mya)	*0.50*	*1.09*	*1.23*	*1.48*	*1.66*	*0.75*	*0.42*	*0.59*
Time prior to species divergence	*0*	*0.59*	*0.73*	*0.98*	*1.16*	*0.25*	*-0.08*	*0.09*

divergence time of ~1.23 million years old.

It is important to note three points. First, accurately estimating absolute divergence time in years is known to be fraught with several sources of error and relies on an accurate calibration point in the absence of an estimate of the mutation rate in each species.[57] We instead rely on the relative comparison between d_{xy} estimates, which are sufficient to resolve the questions that we seek to address here. Second, our conclusion that the *D. persimilis ST* inversion existed as a segregating polymorphism in the ancestor of *D. persimilis* and *D. pseudoobscura* is robust to various methods of inferring the divergence between the two species. For example, our results are not significantly different (p<0.70) if we use whole genome data or only the collinear regions to estimate the absolute divergence between *D. persimilis* and *D. pseudoobscura* (genome-wide mean d_{xy}: 2.41 x10^{-3}; 95% CI: 2.36 – 2.46 x10^{-3}). Third, our data only address the order of origins of the various chromosomal inversions, but do not allow us to estimate when the *Sex-Ratio* distortion alleles evolved in these populations. Identifying the causal segregation distortion genes may allow us to address this aspect in the future. Despite these important caveats of our analysis, the two observations of *D. persimilis SR* being older than the species divergence and the *D. persimilis ST* chromosome appearing significantly more diverged than both the collinear regions and the *D. persimilis SR* chromosome reject the introgression scenario and support the ILS explanation.

All fixed inversions in *D. persimilis* originated as segregating polymorphisms in the ancestral population of *D. persimilis* and *D. pseudoobscura*

Because the *XR* inversion exists only in *D. persimilis* and not in *D. pseudoobscura*, it is often immediately assumed that this inversion must have originated in the *D. persimilis* lineage after speciation. The idea that the *XR* inversion on the *Standard* chromosome of *D. persimilis* originated as a segregating polymorphic inversion in the ancestral population prior to speciation goes against this widely-accepted notion. The two other fixed inversions on the *XL* and *2nd* chromosomes in *D. persimilis* are thought to be even older than the *XR* inversion.[12] We used the same approach that utilizes the sequences flanking inversion breakpoints to also estimate the divergence of the inversions on the *XL* and *2nd* chromosomes. Consistent with the idea that the *XL* and *2nd* chromosome inversions are older than the *XR* inversion, we observed greater mean levels of d_{xy} for both fixed inversions (*XL*: 1.08×10^{-2}; *2*: 6.82×10^{-3}) than for *XR* (d_{xy}: 4.94×10^{-3}). Likewise, standardizing to the speciation time with *D. miranda*, we estimate that the inversions on *XL* and the *2nd* chromosomes originated approximately 1.66 and 1.48 million years ago, respectively (Table 3.1). Our results show that all of these fixed inversions originated in the ancestral population long before the speciation event that separated *D. persimilis* and *D. pseudoobscura*.

Because it may be argued that genome-wide divergence estimates may be a poor proxy for speciation time, we calculated an estimate using another method. *D. persimilis* and *D. pseudoobscura* harbor numerous inversion polymorphisms on the *3rd* chromosome that are exclusive to each species. Structural analyses of the relationship between these inversion polymorphisms show that many of these have been derived

from the *Standard* arrangement in *D. persimilis* (3^{ST}).[52,58] This ancestral 3^{ST} arrangement continues to segregate in both *D. persimilis* and *D. pseudoobscura* populations, and is the only 3^{rd} chromosome arrangement that is shared between the two species. Because the 3^{ST} arrangement was present in the ancestral population, and was inherited by both *D. persimilis* and *D. pseudoobscura*, the amount of divergence between the 3^{ST} arrangements of the two species provides an opportunity to determine an upper bound to the estimates of speciation time. We estimated absolute divergence in inversion-associated sequences from the 3^{ST} strains of *D. persimilis* and *D. pseudoobscura*, and observed a mean d_{xy} of 2.49×10^{-3} (95% CI: $2.34 - 2.65 \times 10^{-3}$). Standardizing to the speciation time with *D. miranda* in these regions, we estimate the 3^{ST} inversion between *D. persimilis* and *D. pseudoobscura* to have diverged approximately 590,000 years ago. This estimate is consistent with those using genome-wide sequences, and is far younger than any of the fixed inversion differences. This difference in the age of the fixed inversions and the time of speciation is not subtle: while the speciation time is estimated at around 500,000 years, the *XL*, *XR*, and 2^{nd} chromosome inversions are at least twice as old as this estimate (Figure 3.7). These results suggest that all of the fixed, derived inversions in *D. persimilis* must have freely segregated in the ancestral population for a substantial period of time before speciation.

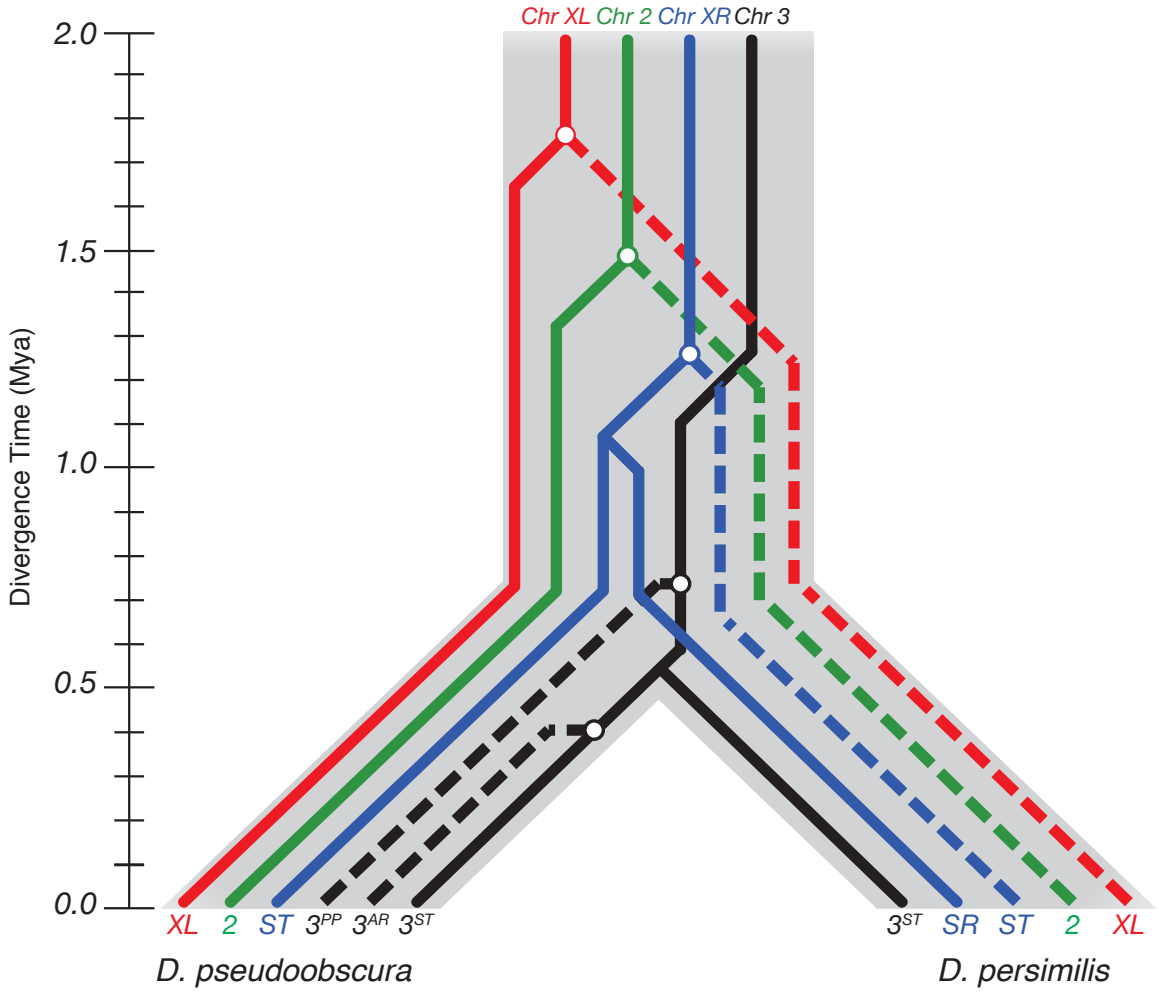

Figure 3.7: Incomplete lineage sorting of the inversions of *D. persimilis* and *D. pseudoobscura*. The fixed inversions on the *XL* and *2nd* chromosomes, as well as the polymorphic inversions on *XR* and the Pikes Peak (3^{PP}) inversion arose before species divergence. Incomplete lineage sorting produced the observed inversion patterns in the species present today.

CHAPTER 4

DISCUSSION

The study of chromosomal inversions in the classic systems of *D. pseudoobscura* and *D. persimilis* has deeply informed our understanding of the evolutionary forces that shape natural variation, the evolution of new species, and selfish chromosome dynamics. Our results have several important implications.

First, our results provide a solution to the strange collinearity of the *D. persimilis SR* and *D. pseudoobscura ST* chromosomes first observed by Dobzhansky. We show that this collinearity is a consequence of the direct descent of these chromosomes from one of the ancestrally segregating arrangements, and not due to two independent inversions at the same breakpoints. Segregation distorters are often associated with inversions because new inversions that tightly link a segregation distorter gene with existing enhancer alleles enjoy a selective advantage. In contrast to most other *Sex-Ratio* systems associated with derived inversions, the *D. persimilis* SR system evolved on the background of an ancestral arrangement. This indicates that segregation distorters may not only become associated with new inversions, but can utilize existing chromosome inversion polymorphisms.

Second, a new *Sex-Ratio* distorting allele may follow three potential fates over time: 1) it may get stochastically lost; 2) it may drive populations to extinction through

biased sex ratios; 3) it may get suppressed and become cryptic.[59] When we observe extant *Sex-Ratio* systems, this represents a snapshot in the life of a distorting chromosome on its way to one of these three fates. Because *Sex-Ratio* chromosomes have such a large effect on phenotypes tightly linked to fitness traits, one may expect this dwell time to be short. Our findings with the *D. persimilis SR* chromosome along with a growing number of other studies suggest that, contrary to these expectations, some unsuppressed *Sex-Ratio* chromosomes may remain polymorphic for long periods of time.[60–63] This can occur if the evolutionary advantage conferred by the selfish behavior of a distorter is balanced by a fitness cost associated with the chromosome.[64] These results may also have implications for artificial gene drive systems.[65,66] While evolutionary analyses of the fate of artificial gene drive systems have focused on the rise of resistant targets of distortion, our data suggest that even unsuppressed drive systems may linger in populations for a long time.

Third, our results demonstrate the difficulty in interpreting statistics such as the modified f_d when inversions segregate in the ancestral population. When we used such an approach in this study, we observed significant f_d at the inversion breakpoints of *D. persimilis SR*. Such signals are often interpreted as evidence for past introgression events. In the context of our study, these signals could not distinguish between introgression and ILS because of restricted recombination in the ancestral population. Nonetheless, many analyses of this class are applied to systems where the history of recombination landscapes is unknown. Such signals should be interpreted with caution, particularly when inversions potentially segregate in ancestral populations. Our results show that, in the case of inversions, the same recombination block that complicates the interpretation

of these statistics also preserves information that can be used to reconstruct evolutionary histories of these inversions.

Fourth, the inversion differences in *D. pseudoobscura* and *D. persimilis* have led to the development of models for the role of chromosomal inversions in the evolution of hybrid incompatibilities. Any model exploring this role must explain at least two empirical patterns: a) the fixed inversions between *D. persimilis* and *D. pseudoobscura* have higher divergence as compared to collinear regions of the genome, and b) most genes that underlie reproductive isolation between *D. persimilis* and *D. pseudoobscura* reside within these inversion differences. Our results show that these inversions were freely segregating in the ancestral population long before speciation, and that the genes contributing to reproductive barriers must have evolved within them afterwards.

Here, we propose a simple model to explain the above two empirical patterns (Figure 4.1). 1) Chromosomal inversions can arise and persist in ancestral populations for long periods of time. During this period, the genomic regions spanning the inversions and the corresponding regions on the un-inverted chromosomes can accumulate genetic divergence aided by the suppression of recombination in heterozygotes.[4,5,43] 2) These chromosomal inversions may undergo incomplete lineage sorting when the ancestral population is split into two allopatric populations.[67] At the initial time of separation, the genes within the chromosomal inversions are already highly diverged, whereas the genes within the collinear regions are nearly identical, with little or no divergence. The highly diverged genes associated with chromosomal inversions are fewer mutational steps away from reaching an incompatible state and are, therefore, likely to evolve to an incompatible state more quickly than those in the collinear regions of the genome. This

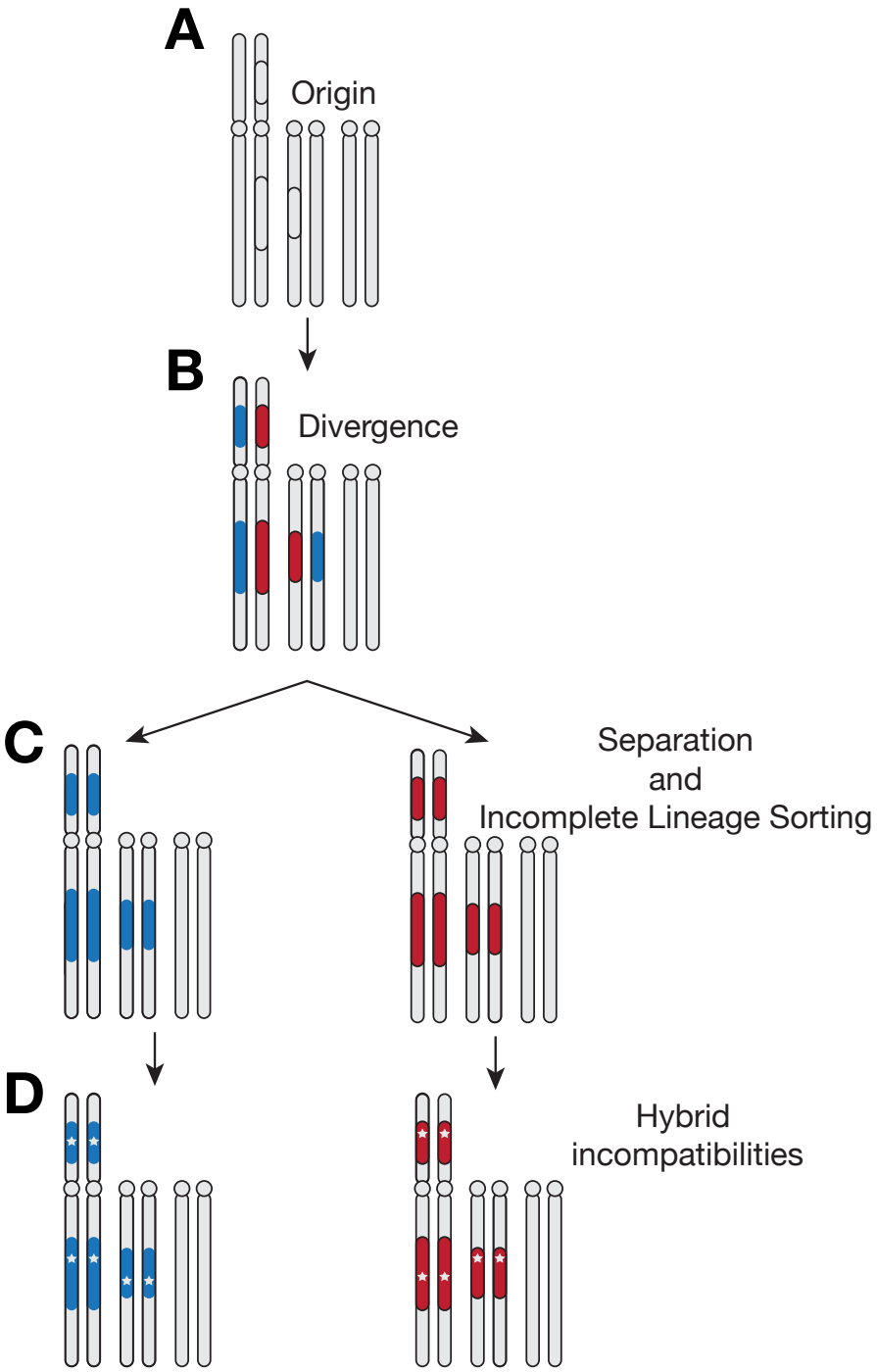

Figure 4.1: Inversions accelerate the formation of hybrid incompatibilities. *(A)* Polymorphic inversions arise in the ancestor of the two species. *(B)* Restricted recombination between the inversions leads to accumulating divergence (red, blue) distinct from collinear regions of the genome (grey). *(C)* Incomplete sorting of the inversions between two isolated populations generates immediate divergence between the two populations. *(D)* Preexisting divergence increases the chance of hybrid incompatibilities forming in the inverted regions as compared to the collinear regions.

accumulation of hybrid incompatibilities occurs in isolation, unopposed by selective cost of producing unfit offspring, in a manner consistent with the Dobzhansky-Muller model.[11,68] This simple model is consistent with all of our findings, and sufficient to explain both patterns. Under our model, the heterogeneity in divergence across the genome caused by ancestrally segregating inversions makes the evolution of alleles that cause reproductive isolation more likely in the regions encompassed by these inversions rather than in the collinear regions of the genome.

Previously, two other models have attempted to explain these patterns. Work centered on the inversions between *D. persimilis* and *D. pseudoobscura,* and those between species in the *Helianthus* sunflower genus, led to the development of the Noor-Rieseberg model.[16,17] According to the Noor-Rieseberg model, if populations diverge in isolation and later re-hybridize on secondary contact, then any incompatible alleles that may have evolved will carry a fitness cost and be selected against. Because inversions suppress recombination, this generates a large block of tightly linked loci. If an incompatible allele is associated with an inversion, then its linkage to other beneficial alleles within the inversion may preserve it in the face of gene flow. In contrast, any incompatible allele that is contained within collinear regions will be replaced by a compatible allele. Thus, fixed inversions associated with hybrid incompatibility genes will appear more diverged, while collinear regions will be homogenized by pervasive gene flow after speciation.

In contrast to the Noor-Rieseberg model, which relies on gene flow after speciation, the Navarro-Barton model invokes gene flow during speciation to explain the above patterns.[69] Normally, due to the cost of producing unfit hybrids, an incompatible

allele is not expected to spread within populations connected by migration. This model considers a scenario where an incompatible allele is located within a chromosomal inversion that carries beneficial alleles. According to this model, the fitness cost incurred by an incompatible allele due to producing unfit hybrids can be offset by the fitness advantage conferred by its linkage to beneficial alleles. Because an incompatible allele located in a collinear region is less likely to be tightly linked to a beneficial allele, inversions rather than collinear regions will contain hybrid incompatibility genes. In addition, inversions associated with alleles that are beneficial in one population but not the other may persist for a long time and accumulate genetic divergence. Under this model, ancestrally segregating inversions explain both patterns described above.

Both the Navarro-Barton and Noor-Rieseberg models rely on gene flow during or after speciation to account for the higher divergence of fixed inversions, and their association with hybrid incompatibility genes. We, however, find little evidence for extensive gene flow that would be required to explain the phylogenetic discordance that we observed on the *D. persimilis XR* chromosome. In the localities where *D. pseudoobscura* and *D. persimilis* overlap, there are frequent arrangements in both species that were derived from the ST karyotype.[52] *D. pseudoobscura* alone harbors more than 30 different polymorphic inversions.[2] If any of the *3rd* chromosome inversion polymorphisms that arose after speciation in *D. pseudoobscura* were to be found in *D. persimilis*, this would provide strong evidence for gene flow during secondary contact between these species. For example, if *D. persimilis* harbored an Arrowhead *3rd* chromosome arrangement, which arose in *D. pseudoobscura* after speciation,[70] this would provide indisputable evidence of large-scale gene flow on secondary contact as

necessitated by the Noor-Rieseberg model. Instead, *D. persimilis* and *D. pseudoobscura* each have their own exclusive series of third chromosomes inversions. Similarly, strong evidence for gene flow during the accumulation of hybrid incompatibilities as necessitated by the Navarro-Barton model is also lacking.

Our model can also account for a third empirical pattern that sympatric species are more likely to harbor fixed chromosomal inversion differences as compared to allopatric species.[71] Under our model, when two temporarily isolated populations inherit ancestrally segregating chromosomal inversions through incomplete lineage sorting, these populations start with highly diverged genomic regions at birth (i.e., associated with inversions) where genes underlying isolating mechanisms may evolve quickly. In contrast, populations that inherit fully collinear genomes have little or no divergence between them at birth and may, therefore, require more time to evolve isolating mechanisms.

If single species often fragment into temporarily isolated populations and merge again, then the populations that inherit ancestrally segregating inversion differences through ILS are more likely to survive as separate species even if they later become fully sympatric. In contrast, those with collinear genomes are less likely to evolve hybrid incompatibilities during temporary allopatry and collapse back into single species on secondary contact. Such cases will not be observed, unless allopatry is maintained long enough to allow the evolution of hybrid incompatibilities. Together, this process is predicted to generate a pattern of sympatric species pairs that are enriched for inversion differences, and allopatric species pairs that have collinear genomes.

Our model also makes a distinct prediction regarding young allopatric species that

inherit ancestrally segregating inversions through ILS: if such cases are found, hybrid incompatibility genes must be enriched in regions spanned by the inversion differences despite no gene flow between these populations. Because hybrid incompatibilities may accumulate across the genome over time through the snowball effect, this enrichment of hybrid incompatibility genes at inversions may decay over time in older species. This prediction is not expected under the Noor-Rieseberg or Navarro-Barton models.

We propose that higher divergence and hybrid incompatibilities are emergent properties of ancestrally segregating inversions that are inherited through incomplete lineage sorting. Our model explains previously observed empirical patterns without invoking gene flow across populations during or after speciation. These findings force a reconsideration of the role of inversion polymorphisms in speciation, perhaps not as a protector of existing hybrid incompatibility alleles, but as fertile ground for their formation.

REFERENCES

1. Dobzhansky, T. & Dobzhansky, T. G. *Genetics and the Origin of Species*. (Columbia University Press, 1937).

2. Dobzhansky, T. & Sturtevant, A. H. Inversions in the chromosomes of Drosophila pseudoobscura. *Genetics* **23,** 28–64 (1938).

3. Lewontin, R. C., Moore, J. A., Provine, W. B. & Wallace, B. *Dobzhansky's Genetics of Natural Populations I-XLIII*. (Columbia University Press, 1981).

4. Schaeffer, S. W. *et al.* Evolutionary genomics of inversions in Drosophila pseudoobscura: Evidence for epistasis. *Proc. Natl. Acad. Sci. U. S. A.* **100,** 8319–8324 (2003).

5. Fuller, Z. L., Haynes, G. D., Richards, S. & Schaeffer, S. W. Genomics of natural populations: How differentially expressed genes shape the evolution of chromosomal inversions in Drosophila pseudoobscura. *Genetics* genetics.116.191429 (2016). doi:10.1534/genetics.116.191429

6. Richards, S. *et al.* Comparative genome sequencing of Drosophila pseudoobscura: Chromosomal, gene, and cis-element evolution. *Genome Res.* **15,** 1–18 (2005).

7. Orr, H. A. Genetics of male and female sterility in hybrids of Drosophila pseudoobscura and D. persimilis. *Genetics* **116,** 555–563 (1987).

8. Noor, M. A. F. Genetics of sexual isolation and courtship dysfunction in male hybrids of Drosophila pseudoobscura and Drosophila persimilis. *Evolution* **51,** 809–815 (1997).

9. Dobzhansky, T. On the sterility of the interracial hybrids in Drosophila pseudoobscura. *Proc. Natl. Acad. Sci. U. S. A.* **19,** 397–403 (1933).

10. Dobzhansky, T. Role of the autosomes in the Drosophila pseudoobscura hybrids. *Proc. Natl. Acad. Sci. U. S. A.* **19,** 950–953 (1933).

11. Dobzhansky, T. Studies on Hybrid Sterility. II. Localization of sterility factors in Drosophila pseudoobscura hybrids. *Genetics* **21,** 113–135 (1936).

12. Noor, M. A. F., Garfield, D. A., Schaeffer, S. W. & Machado, C. A. Divergence between the Drosophila pseudoobscura and D. persimilis genome sequences in relation to chromosomal inversions. *Genetics* **177,** 1417–1428 (2007).

13. Kulathinal, R. J., Stevison, L. S. & Noor, M. A. F. The genomics of speciation in Drosophila: Diversity, divergence, and introgression estimated using low-coverage genome sequencing. *PLOS Genet.* **5,** e1000550 (2009).

14. Stevison, L. S., Hoehn, K. B. & Noor, M. A. F. Effects of inversions on within- and between-species recombination and divergence. *Genome Biol. Evol.* **3,** 830–841 (2011).

15. Machado, C. A., Haselkorn, T. S. & Noor, M. A. F. Evaluation of the genomic extent of effects of fixed inversion differences on intraspecific variation and interspecific gene flow in Drosophila pseudoobscura and D. persimilis. *Genetics* **175,** 1289–1306 (2007).

16. Noor, M. A. F., Grams, K. L., Bertucci, L. A. & Reiland, J. Chromosomal inversions and the reproductive isolation of species. *Proc. Natl. Acad. Sci.* **98,** 12084–12088 (2001).

17. Rieseberg, L. H. Chromosomal rearrangements and speciation. *Trends Ecol. Evol.* **16,** 351–358 (2001).

18. Lee, Y. *et al.* Chromosome inversions, genomic differentiation and speciation in the African malaria mosquito Anopheles gambiae. *PLOS ONE* **8,** e57887 (2013).

19. Kirkpatrick, M. & Barton, N. Chromosome inversions, local adaptation and speciation. *Genetics* **173,** 419–434 (2006).

20. Sanchez-Flores, A. *et al.* Genome evolution in three species of cactophilic Drosophila. *G3 Genes Genomes Genet.* **6,** 3097–3105 (2016).

21. Manoukis, N. C. *et al.* A test of the chromosomal theory of ecotypic speciation in Anopheles gambiae. *Proc. Natl. Acad. Sci.* **105,** 2940–2945 (2008).

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

23. Price, T. A. R. *et al.* Does polyandry control population sex ratio via regulation of a selfish gene? *Proc. R. Soc. B Biol. Sci.* **281,** (2014).

24. Wedell, N. The dynamic relationship between polyandry and selfish genetic elements. *Philos. Trans. R. Soc. B Biol. Sci.* **368,** (2013).

25. Curtsinger, J. W. & Feldman, M. W. Experimental and theoretical analysis of the 'Sex-Ratio' polymorphism in Drosophila pseudoobscura. *Genetics* **94,** 445–466 (1980).

26. Policansky, D. & Ellison, J. 'Sex ratio' in Drosophila pseudoobscura: Spermiogenic failure. *Science* **169,** 888–889 (1970).

27. Bastide, H., Gérard, P. R., Ogereau, D., Cazemajor, M. & Montchamp-Moreau, C. Local dynamics of a fast-evolving sex-ratio system in Drosophila simulans. *Mol. Ecol.* **22,** 5352–5367 (2013).

28. Jaenike, J. Sex Chromosome Meiotic Drive. *Annu. Rev. Ecol. Syst.* **32,** 25–49 (2001).

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

30. Sturtevant, A. H. & Dobzhansky, T. Geographical distribution and cytology of 'Sex Ratio' in Drosophila pseudoobscura and related species. *Genetics* **21,** 473–490 (1936).

31. Sandler, L. & Novitski, E. Meiotic drive as an evolutionary force. *Am. Nat.* **91,** 105–110 (1957).

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

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

34. Wu, C. I. & Beckenbach, A. T. 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 (1983).

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

36. Painter, T. S. A new method for the study of chromosome aberrations and the plotting of chromosome maps in Drosophila melanogaster. *Genetics* **19,** 175–188 (1934).

37. Harshman, L. G. A technique for the preparation of Drosophila salivary gland chromosomes. *Drosoph. Inf. Serv.* **52,** (1977).

38. Ballard, J. W. O. & Bedo, D. G. Population cytogenetics of Austrosimulium bancrofti (Diptera: Simuliidae) in eastern Australia. *Genome* **34,** 338–353 (1991).

39. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* btu170 (2014). doi:10.1093/bioinformatics/btu170

40. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinforma. Oxf. Engl.* **25,** 1754–1760 (2009).

41. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25,** 2078–2079 (2009).

42. Garrison, E. & Marth, G. Haplotype-based variant detection from short-read sequencing. *ArXiv12073907 Q-Bio* (2012).

43. Fuller, Z. L. *et al.* Evidence for stabilizing selection on codon usage in chromosomal rearrangements of Drosophila pseudoobscura. *G3 GenesGenomesGenetics* g3.114.014860 (2014). doi:10.1534/g3.114.014860

44. Schaeffer, S. W. *et al.* Polytene chromosomal maps of 11 Drosophila species: The order of genomic scaffolds inferred from genetic and physical maps. *Genetics* **179,** 1601–1655 (2008).

45. Nei, M., Tajima, F. & Tateno, Y. Accuracy of estimated phylogenetic trees from molecular data. II. Gene frequency data. *J. Mol. Evol.* **19,** 153–170 (1983).

46. Kalinowski, S. T. Evolutionary and statistical properties of three genetic distances. *Mol. Ecol.* **11,** 1263–1273 (2002).

47. Saitou, N. & Nei, M. The neighbor-joining method: A new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* **4,** 406–425 (1987).

48. Nei, M. *Molecular Evolutionary Genetics*. (Columbia University Press, 1987).

49. Garfield, D. A. & Noor, M. A. Characterization of novel repetitive element Leviathan in Drosophila pseudoobscura. *Drosoph. Inf. Serv.* **90,** 1–9 (2007).

50. Aguado, C. *et al.* Validation and genotyping of multiple human polymorphic inversions mediated by inverted repeats reveals a high degree of recurrence. *PLOS Genet.* **10,** e1004208 (2014).

51. Dobzhansky, T. Is there gene exchange between Drosophila pseudoobsura and Drosophila persimilis in their natural habitats? *Am. Nat.* **107,** 312–314 (1973).

52. Dobzhansky, T. *Chromosomal races in Drosophila pseudoobscura and Drosophila persimilis*. **554,** (Washington Publ., 1944).

53. Navarro, A., Barbadilla, A. & Ruiz, A. Effect of inversion polymorphism on the neutral nucleotide variability of linked chromosomal regions in Drosophila. *Genetics* **155,** 685–698 (2000).

54. Navarro, A., Betrán, E., Barbadilla, A. & Ruiz, A. Recombination and gene flux caused by gene conversion and crossing over in inversion heterokaryotypes. *Genetics* **146,** 695–709 (1997).

55. Durand, E. Y., Patterson, N., Reich, D. & Slatkin, M. Testing for ancient admixture between closely related populations. *Mol. Biol. Evol.* **28,** 2239–2252 (2011).

56. Martin, S. H., Davey, J. W. & Jiggins, C. D. Evaluating the use of ABBA–BABA statistics to locate introgressed loci. *Mol. Biol. Evol.* **32,** 244–257 (2015).

57. Graur, D. & Martin, W. Reading the entrails of chickens: Molecular timescales of evolution and the illusion of precision. *Trends Genet. TIG* **20,** 80–86 (2004).

58. Moore, B. C. & Taylor, C. E. Drosophila of Southern California: III. Gene arrangements of Drosophila persimilis. *J. Hered.* **77,** 313–323 (1986).

59. Carvalho, A. B. & Vaz, S. C. Are Drosophila SR drive chromosomes always balanced? *Heredity* **83,** 221–228 (1999).

60. Pieper, K. E. & Dyer, K. A. Occasional recombination of a selfish X-chromosome may permit its persistence at high frequencies in the wild. *J. Evol. Biol.* **29,** 2229–2241 (2016).

61. Lindholm, A. K. *et al.* The ecology and evolutionary dynamics of meiotic drive. *Trends Ecol. Evol.* **31,** 315–326 (2016).

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

63. Jaenike, J. X chromosome drive. *Curr. Biol.* **18,** R508–R511 (2008).

64. Wu, C.-I., True, J. R. & Johnson, N. Fitness reduction associated with the deletion of a satellite DNA array. *Nature* **341,** 248–251 (1989).

65. Eckhoff, P. A., Wenger, E. A., Godfray, H. C. J. & Burt, A. Impact of mosquito gene drive on malaria elimination in a computational model with explicit spatial and temporal dynamics. *Proc. Natl. Acad. Sci.* **114,** E255–E264 (2017).

66. Burt, A. Heritable strategies for controlling insect vectors of disease. *Phil Trans R Soc B* **369,** 20130432 (2014).

67. Guerrero, R. F. & Hahn, M. W. Speciation as a sieve for ancestral polymorphism. *Mol. Ecol.* n/a–n/a doi:10.1111/mec.14290

68. Muller, H. J. Isolating mechanisms, evolution and temperature. *Biology Symposium* 71–125 (1942)

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

70. Wallace, A. G., Detweiler, D. & Schaeffer, S. W. Evolutionary history of the third chromosome gene arrangements of Drosophila pseudoobscura inferred from inversion breakpoints. *Mol. Biol. Evol.* **28,** 2219–2229 (2011).

71. Brown, K. M., Burk, L. M., Henagan, L. M. & Noor, M. A. F. A test of the chromosomal rearrangement model of speciation in Drosophila pseudoobscura. *Evol. Int. J. Org. Evol.* **58,** 1856–1860 (2004).