

RESEARCH ARTICLE

Ancestral polymorphisms explain the role of chromosomal inversions in speciation

Zachary L. Fuller^{1☯^{‡a}}, Christopher J. Leonard^{2☯}, Randee E. Young^{2^{‡b}}, Stephen W. Schaeffer¹, Nitin Phadnis^{2*}

1 Department of Biology, Erwin W. Mueller Laboratories, The Pennsylvania State University, University Park, PA, United States of America, **2** Department of Biology, University of Utah, Salt Lake City, UT, United States of America

☯ These authors contributed equally to this work.

^{‡a} Current address: Fairchild Center, Department of Biological Sciences, Columbia University, New York, NY, United States of America

^{‡b} Current address: Laboratory of Genetics, University of Wisconsin, Madison, WI, United States of America.

* nitin.phadnis@utah.edu



 OPEN ACCESS

Citation: Fuller ZL, Leonard CJ, Young RE, Schaeffer SW, Phadnis N (2018) Ancestral polymorphisms explain the role of chromosomal inversions in speciation. *PLoS Genet* 14(7): e1007526. <https://doi.org/10.1371/journal.pgen.1007526>

Editor: Patricia Wittkopp, University of Michigan, UNITED STATES

Received: May 17, 2018

Accepted: June 29, 2018

Published: July 30, 2018

Copyright: © 2018 Fuller et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: Data are available from NCBI: <https://www.ncbi.nlm.nih.gov/bioproject/PRJNA420372/>, under identifiers 11040X1(DperST): SAMN08110862, 11040X2 (DperSR): SAMN08110863, 10700X1(DpserST): SAMN08110864 and 10700X2(DpseSR): SAMN08110865.

Funding: This work was supported by the National Institutes of Health (Genetics Training Grant 5T32GM007464-40 (GJL), R01 GM115914 (NP), R01 GM 098478 (SWS), a Mario Capecchi

Abstract

Understanding the role of chromosomal inversions in speciation is a fundamental problem in evolutionary genetics. Here, we perform a comprehensive reconstruction of the evolutionary histories of the chromosomal inversions in *Drosophila persimilis* and *D. pseudoobscura*. We provide a solution to the puzzling origins of the selfish *Sex-Ratio* arrangement in *D. persimilis* and uncover surprising patterns of phylogenetic discordance on this chromosome. These patterns show that, contrary to widely held views, all fixed chromosomal inversions between *D. persimilis* and *D. pseudoobscura* were already present in their ancestral population long before the species split. Our results suggest that patterns of higher genomic divergence and an association of reproductive isolation genes with chromosomal inversions may be a direct consequence of incomplete lineage sorting of ancestral polymorphisms. These findings force a reconsideration of the role of chromosomal inversions in speciation, not as protectors of existing hybrid incompatibilities, but as fertile grounds for their formation.

Author summary

Studies on chromosomal inversions and reproductive isolation between *Drosophila persimilis* and *D. pseudoobscura* have played a profound role in shaping our understanding of inversions, speciation and selfish chromosomes. In this study, we reconstruct the evolutionary histories of chromosomal inversions in *D. persimilis* and *D. pseudoobscura* to show that, contrary to widely accepted ideas, these inversions existed as polymorphisms in the ancestor of both species before their initial split. These findings force a reconsideration of the role of chromosomal inversions in speciation and raise the possibility that the higher genetic divergence of sequences spanning these chromosomal inversions and an association with hybrid incompatibility genes may be an emergent property of the long-term segregation of these inversions.

endowed assistant professorship (NP), and the Pew Biomedical Scholars Program. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

Introduction

Chromosomal inversions are structural rearrangements where the linear gene order is reversed. In crosses between two species that differ by one or more inversions, the resulting hybrids can experience meiotic chromosome pairing problems and may, therefore, become sterile. Chromosomal inversions can, thus, potentially play an important role in the evolution of intrinsic postzygotic barriers between species. Understanding the extent to which such chromosomal rearrangements play a role in speciation is a longstanding and fundamental problem in evolutionary genetics [1–3]. In a number of plant species, direct experimental evidence has cemented the role of chromosomal rearrangements in the evolution of reproductive isolation through the reduced fertility in heterokaryotic hybrids [1,4–6]. In contrast, classic studies in hybrids between *Drosophila persimilis* and *D. pseudoobscura* have shown that chromosomal inversions do not play a direct role in causing hybrid sterility in animal species [2,4]. There is now clear evidence for genic incompatibilities as the cause hybrid sterility in many cases, and the idea that chromosomal inversions may play a role in animal speciation fell out of favor [2,4]. Recent studies in *D. persimilis* and *D. pseudoobscura*—the same species that helped lead to the demise of the idea of a direct role of chromosomal inversions in hybrid sterility—, however, have led to a dramatic resurgence of a modified version for the role of chromosomal inversions in speciation. Two new empirical observations regarding the patterns of reproductive isolation and genetic divergence in *D. persimilis* and *D. pseudoobscura* are key to these developments: i) the fixed chromosomal inversions between these species display higher genetic divergence than collinear regions of the genome, and ii) nearly all genes that contribute to reproductive isolation between these species are located among the fixed chromosomal inversion differences [7–12].

These two empirical patterns are explained by new versions of the chromosomal theory of speciation, which may be explained as follows. Consider a single species that has recently separated into two isolated populations [9,13]. These populations evolve independently, and genes that contribute to reproductive isolation initially evolve uniformly across the genome in both populations. If these populations were to later re-hybridize on secondary contact, any incompatible alleles will be selected against because such alleles suffer a fitness cost in the form of unfit hybrid progeny. However, in populations that have evolved fixed inversion differences, incompatible alleles may become locked together with beneficial alleles in large blocks of tightly linked loci generated by the recombination suppressing properties of inversions. In such a situation, linked beneficial alleles may prevent selection from eliminating the incompatible alleles on secondary contact and, thus, help maintain reproductive isolation between these incipient species during secondary contact.

In contrast, collinear regions of the genome may continue to exchange genes, leading to the elimination of incompatible alleles in these regions. Any gene flow between species is, thus, prevented within genomic regions spanning chromosomal inversions through the maintenance of hybrid incompatibilities, but continues across collinear regions of the genomes. Due to this heterogeneous pattern of gene flow across the genome after the initial evolution of reproductive isolating barriers, hybrid incompatibility alleles may become disproportionately associated with chromosomal inversion differences between species, and genomic regions spanning inversions may appear more genetically diverged as compared to collinear regions. This ‘speciation with gene flow’ process can, thus, explain both empirical patterns found in *Drosophila persimilis* and *D. pseudoobscura*, which represent one of the most thoroughly studied hybridizations in speciation genetics. Consistent with this idea, Dobzhansky (1973) observed a single hybrid female between *D. pseudoobscura* and *D. persimilis* from nature and multiple studies have detected genomic signatures of recent gene flow between these species,

suggesting that these species may continue to exchange genes at a detectable level [12,14–17]. Moreover, the empirical patterns described above appear difficult to explain without invoking a major role for gene flow after the initial evolution of reproductive isolation. Together, these results support the ‘speciation with gene flow’ idea, and have led to the widespread acceptance of the role of recombination suppression by chromosomal inversions in the maintenance of animal species [11,18–23].

Here, we comprehensively dissect the evolutionary histories of the chromosomal inversions in *D. persimilis* and *D. pseudoobscura* to show that, contrary to the currently accepted view, all fixed chromosomal inversions between these species segregated in their common ancestral population, and pre-dated the divergence between these species by a remarkable length of time. Our key insights into deciphering the evolutionary histories of these chromosomal inversions came from resolving the origins of the chromosomal arrangement associated with the *D. persimilis* Sex-Ratio phenotype and from uncovering strong patterns of phylogenetic discordance along the Sex-Ratio chromosome. We, therefore, explain our resolution of the evolutionary history of this Sex-Ratio chromosome before proceeding to reconstruct the evolutionary history of the fixed chromosomal inversions differences between *D. persimilis* and *D. pseudoobscura*.

Sex-Ratio chromosomes are variants of X-chromosomes that are often found at high frequencies within natural populations [24]. Males that carry a Sex-Ratio chromosome eliminate nearly all Y-bearing sperm [25], and produce nearly all female offspring (*i.e.*, heavily distorted progeny sex-ratios). By distorting the balance of segregation in their favor in excess of Mendelian expectations, these Sex-Ratio chromosomes can rapidly spread through populations even if they reduce the fitness of the individuals that carry them [26,27]. When a new chromosomal inversion generates tight linkage between an existing segregation distorter allele and other alleles that enhance distortion (or alleles that neutralize suppressors-of-distortion), this produces a stronger driving chromosome that can supplant its weaker versions [28]. This process sets up an expected order for the evolution of Sex-Ratio chromosomes: distorter alleles arise first, enhancers of distortion appear next, and chromosomal inversions that tie these together arrive last. This framework explains why most Sex-Ratio chromosomes are associated with derived inversions relative to the wild type, or Standard (ST) chromosomes [27]. Consistent with this pattern, the *D. persimilis* SR chromosome is inverted with respect to the *D. persimilis* ST chromosome on the right arm of the X chromosome (XR). However, the Standard *D. persimilis* XR differs from *D. pseudoobscura* XR by a single derived inversion. Curiously, the *D. persimilis* SR inversion appears to have reversed the same derived *D. persimilis* ST inversion, such that *D. persimilis* SR appears collinear with *D. pseudoobscura* (Fig 1A). It is not clear whether this unexpected collinearity of the *D. persimilis* SR chromosome with the ST chromosome of its sister species is the result of a second inversion event on the background of *D. persimilis* ST at approximately the same breakpoints as the original *D. persimilis* XR inversion, or whether a single chromosomal arrangement was inherited from the ancestor of the two species [29,30]. Previous molecular evolutionary studies on the origins of this chromosome have yielded conflicting results, and the origin of the *D. persimilis* Sex-Ratio inversion remains the subject of speculation [7,31,32].

Here, we show that the *D. persimilis* SR chromosome did not arise from a second inversion event, but is the ancestrally-arranged chromosome. Surprisingly, we also discovered large blocks of phylogenetic discordance in the regions flanking the *D. persimilis* SR inversion breakpoints, such that they are more closely related to the *D. pseudoobscura*, rather than to the *D. persimilis* ST chromosome. These patterns provide evidence that, contrary to the currently held view, fixed rearrangement differences between *D. persimilis* and *D. pseudoobscura* arose in the ancestor of the two species before being passed exclusively to *D. persimilis*. Using whole-genome data in this same model system, Kulathinal et al. (2009) concluded that the similarly observed patterns of increased divergence associated with inverted regions was the result of *D.*

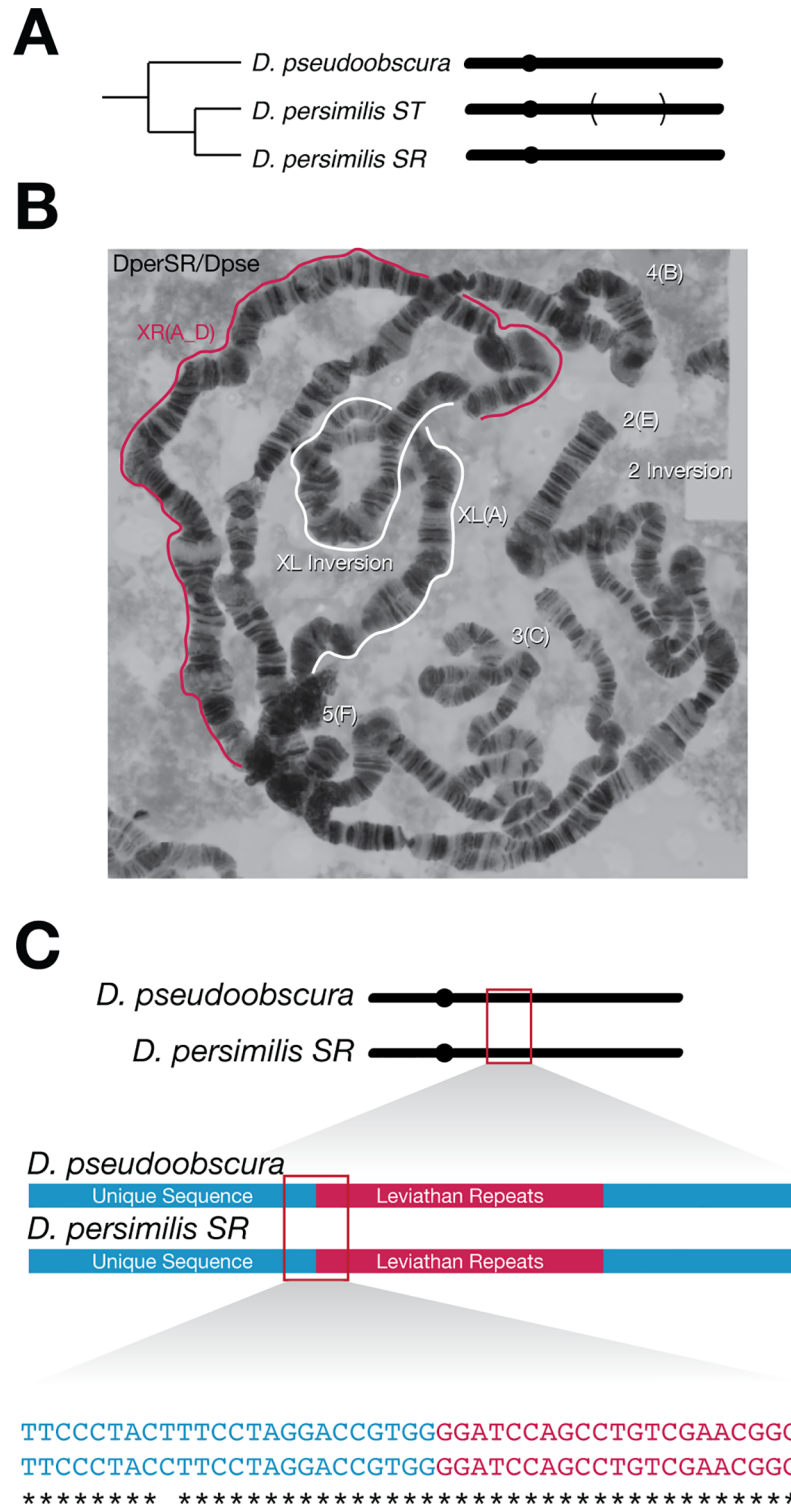


Fig 1. The *D. persimilis* Sex-Ratio (SR) chromosome is precisely collinear with *D. pseudoobscura*. (A) The right arm of the X chromosome (XR) of *D. persimilis* is normally inverted as compared to its sister species, *D. pseudoobscura*, but the *D. persimilis* Sex-Ratio chromosome is collinear with its sister species. (B) Polytene chromosome squash of a *D. persimilis* SR/*D. pseudoobscura* hybrid female demonstrating perfect interspecies collinearity on XR. (C) Amplification and sequencing of the proximal breakpoint of the *D. persimilis* inversion reveals that the breakpoints are collinear at the base-pair level.

<https://doi.org/10.1371/journal.pgen.1007526.g001>

persimilis having acquired all three inversions after speciation and the homogenizing effect of post-secondary contact. In our study, with higher resolution sequences, multiple statistical approaches, and the inclusion of the *D. persimilis* SR arrangement, we instead show that all fixed inversion differences are the result of ancestrally segregating polymorphisms and offer a model which does not rely on post-speciation gene flow or ongoing hybridization to explain the observed patterns of divergence. Together, our results challenge our current understanding of the evolutionary history of the inversions in *D. persimilis* and *D. pseudoobscura*, and suggest that ancestrally segregating polymorphisms may play a critical role in establishing the patterns of divergence and an association of reproductive isolation genes with chromosomal inversion differences between species.

Results

D. pseudoobscura XR and *D. persimilis* Sex-Ratio are precisely collinear

We isolated two independent *D. persimilis* SR strains that produce >90% female progeny, and generated high quality mosaic images of polytene chromosomes from squashes of larval salivary glands. Consistent with previous reports [24], the *D. persimilis* SR chromosome differs by one major inversion on XR with respect to *D. persimilis* ST, but appears collinear with *D. pseudoobscura* (Fig 1B, S1 Fig). 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 hybrid *D. persimilis* SR/*D. pseudoobscura* females may reveal slight imperfections near the inversion breakpoints. 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 proximity to the original breakpoints of the *D. persimilis* ST inversion.

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 [10]. 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 were able to successfully amplify sequences corresponding to the proximal breakpoint (S2 Fig) and Sanger sequencing of these products revealed the presence of four 319bp *Leviathan* repeats [33]. More importantly, *D. persimilis* SR and *D. pseudoobscura* sequences that flank the *Leviathan* repeats are precisely collinear to a single base pair resolution (Fig 1C). We were unable to amplify the sequences across the distal breakpoint, likely because of the presence of a large block of repetitive sequences accumulated at this breakpoint after the initial inversion event. Nevertheless, information about the proximal inversion breakpoint accurately provides the position of the distal breakpoint, which is sufficient to answer the questions that we seek to address here. In particular, our results from the proximal breakpoint show that a slightly staggered second inversion event is not the basis for the collinearity between the *D. persimilis* SR and *D. pseudoobscura* chromosomes.

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

Repetitive elements, such as *Leviathan* sequences, are known to be hotspots for inversion breakpoints [33,34]. While *Leviathan* repeats are unique to *D. persimilis* and *D. pseudoobscura*,

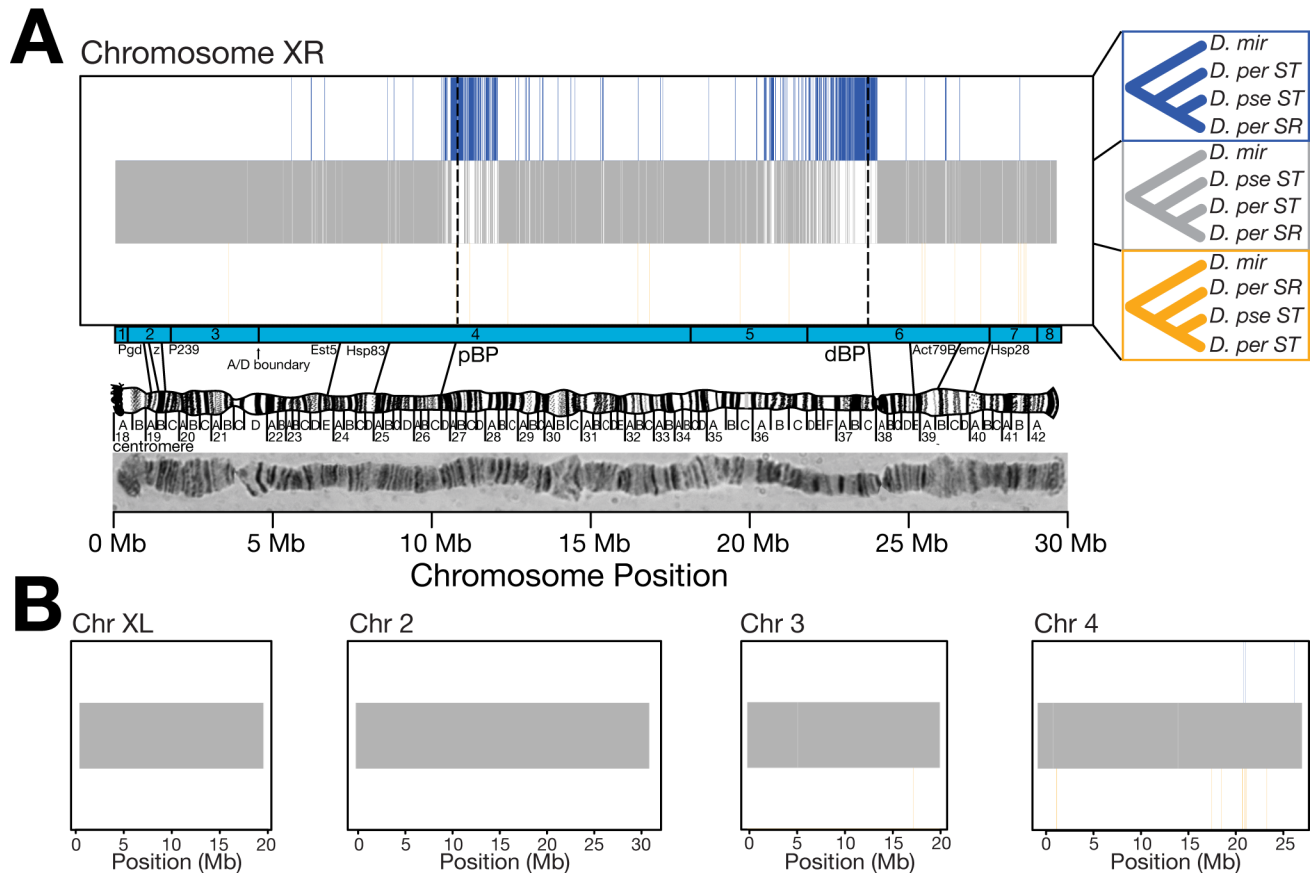


Fig 2. The inversion breakpoints on XR show extensive phylogenetic discordance. (A) Sliding window phylogeny classification on XR. Blue, grey, and orange vertical lines represent the tree topology supported by neighbor-joining trees. Grey trees represent no phylogenetic discordance. Blue trees represent 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.

<https://doi.org/10.1371/journal.pgen.1007526.g002>

XR alone harbors more than 650 of these repeats spread across the chromosome arm. Given this large number, the probability of a second inversion event (e.g. [35–37]) on *D. persimilis* SR at the same two *Leviathan* repeats as the original breakpoints appears vanishingly small. To directly test whether *D. persimilis* SR is recently derived from *D. persimilis* ST through a secondary inversion event, we inferred phylogenetic relationships in 10kb non-overlapping 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 (Fig 2A). Surprisingly, we find two large blocks of phylogenetic discordance concentrated at the inversion breakpoints on XR. In these regions of phylogenetic discordance that span a few megabases of sequences, *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 (Fig 2B).

We next asked whether the phylogenetic discordance observed on the *D. persimilis* SR chromosome is found anywhere else in the genome. Our sliding window phylogenetic analyses based on the XR classification (DpseST, DperST, and DperSR) show that there are no other large blocks of phylogenetic discordance anywhere else in the genome (Fig 2B). Although these analyses revealed small regions of phylogenetic discordance in other regions of the

genome, there is no clustering of consecutive discordant windows, and the discordant windows are not associated with other fixed inversions. We also separately analyzed the *Standard* arrangement on the 3rd chromosome (3^{ST}) which, like *D. persimilis* SR, is both shared across *D. persimilis* and *D. pseudoobscura* and is polymorphic within each species, and the *Arrowhead* arrangement (3^{AR}) which is unique to *D. pseudoobscura*. 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 are a unique property of the *D. persimilis* SR chromosome (S1 Text; S3 Fig). Together with the precisely-shared breakpoints, the relatedness between *D. persimilis* SR and *D. pseudoobscura* at the inversion breakpoints rejects the 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 possibilities that *D. persimilis* SR was derived either through a recent introgression event from *D. pseudoobscura*, or from incomplete lineage sorting of the polymorphism from the common ancestor of *D. persimilis* and *D. pseudoobscura* (Fig 3).

Regions of phylogenetic discordance allow a dating of free gene exchange between the *D. persimilis* SR and *D. pseudoobscura* ST arrangements

Because *D. persimilis* and *D. pseudoobscura* can potentially hybridize in nature [17], our results raise the possibility that the *D. persimilis* SR arrangement originated as a recent introgression of *D. pseudoobscura* XR (Fig 3A). 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 would be best preserved at the breakpoints of the inversion where suppression of crossovers is greatest [38,39]. The preservation of *D. pseudoobscura* material at the inversion breakpoints would then generate the blocks of phylogenetic discordance that we observe on *D. persimilis* SR.

An alternative explanation involving the inheritance of the *D. persimilis* SR and *D. pseudoobscura* ST arrangements from the common ancestor of both species is also consistent with 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 the loss of one arrangement from *D. pseudoobscura* (Fig 3B). Under this scenario of incomplete lineage sorting (ILS) in *D. persimilis*, the ST inversion originates as a 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 the initial evolution of reproductive isolation. During this time, the ancestor of *D. persimilis* SR and *D. pseudoobscura* ST chromosomes (which are collinear) continue to freely recombine until the splitting of the two species, but diverge from the ancestor of the *D. persimilis* ST chromosome. Similar to the introgression scenario, recombination events homogenize the central regions of the *D. persimilis* SR and ST arrangements after speciation, except at the breakpoints of the inversion, thus leading to the patterns of phylogenetic discordance.

Common approaches to distinguish introgression from ILS, such as *f*-statistics and related “ABBA-BABA” methods, involve an implicit assumption of free recombination in the ancestral population. However, in the case of inversions and other recombination limited regions of the genome this assumption is violated and these measures cannot reliably distinguish between

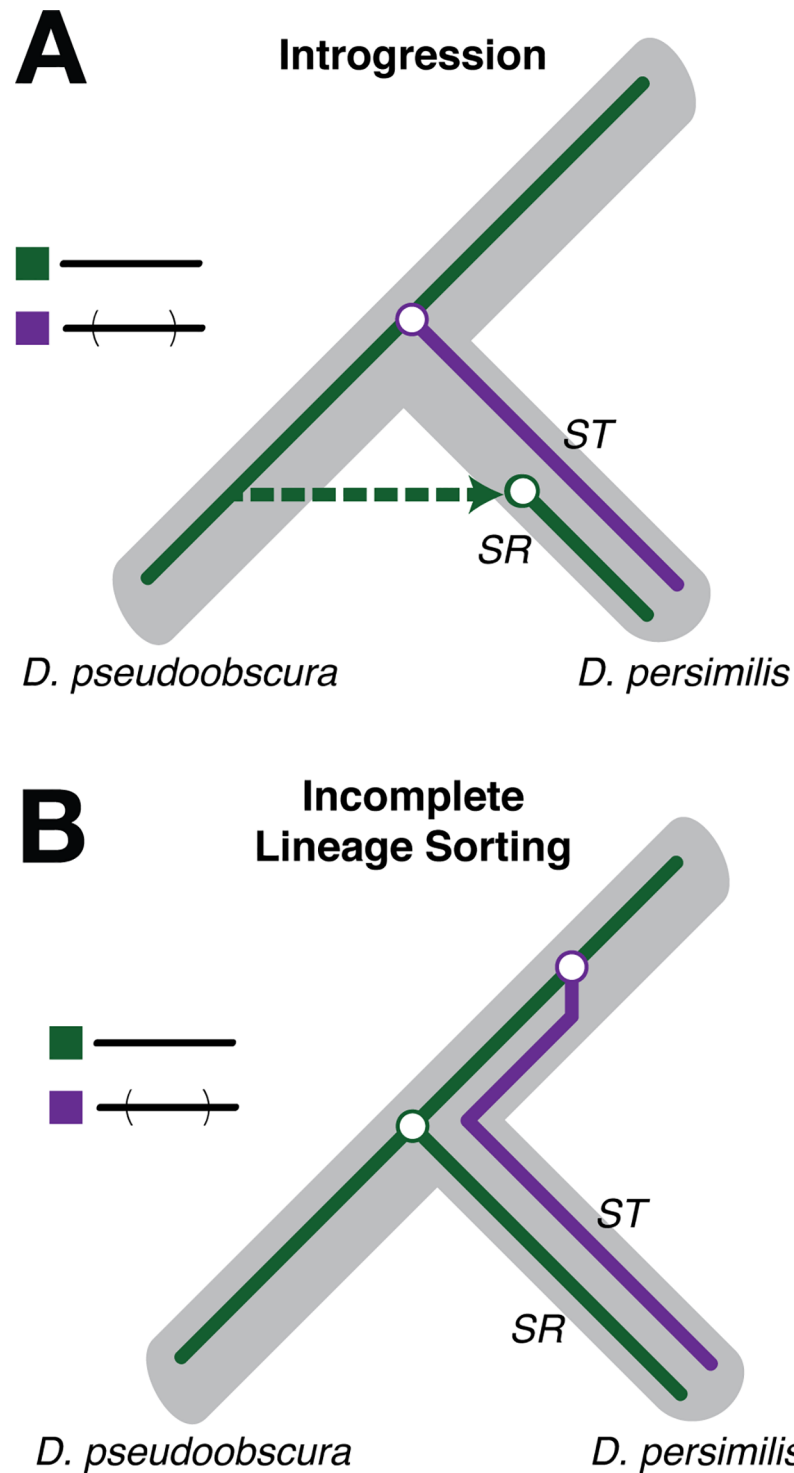


Fig 3. Discordance may be produced by introgression or incomplete lineage sorting of the XR arrangements. Under model (A), the *D. persimilis* ST inversion segregates in the ancestral population of the species. Later divergence between *D. persimilis* SR and *D. pseudoobscura* chromosomes and recombination 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.

<https://doi.org/10.1371/journal.pgen.1007526.g003>

the two hypotheses. Alternatively, we can discriminate between these scenarios by determining whether the exchange occurred after the spitting of the two species (introgression) or in the ancestor of both species before the evolution of reproductive isolation (ILS). To estimate the date of exchange relative to reproductive isolation, we first estimated absolute divergence (d_{xy}) in 10 kb windows for different regions of the genome. We then normalized d_{xy} in each window relative to the divergence with the *D. miranda* outgroup, a measure known as the “relative node depth” (RND), to adjust for regional variation in the mutation rate [40]. It is important to note that accurately converting absolute divergence to units of years is known to be fraught with several sources of error and requires an accurate calibration point in the absence of an estimate of the mutation rate in each species [41]. For the sake of interpretability, we scale the genetic differentiation in each window to the widely used *D. pseudoobscura*-*D. miranda* divergence time of 2 million years [42]. However, we rely on the relative comparison between distributions of d_{xy} and RND which are sufficient to resolve the questions we seek to address here.

D. persimilis and *D. pseudoobscura* are thought to have diverged approximately 500,000 years ago [15,42]. Indeed, in our data the average RND between *D. persimilis* and *D. pseudoobscura* in all collinear regions across the genome is 0.528 (95% CI: 0.521–0.535; Median: 0.513) and the mean divergence time based on genetic differentiation is estimated as 452,806 years ago (95% CI: 445,713–459,890). To determine the timing of chromosome exchange of the *D. persimilis* SR/*D. pseudoobscura* ST arrangements, we used the sequences flanking the inversion breakpoints (\pm 250 kb) to estimate divergence between *D. persimilis* SR and *D. pseudoobscura* and observe a mean RND of 0.662 (95% CI: 0.639–0.685; Median: 0.659). In these regions, we estimate the *D. persimilis* SR chromosome to have shared a common ancestor with *D. pseudoobscura* ST ~1 million years ago (95% CI: 0.95–1.05 Mya; Table 1). The estimated distribution of RND in these flanking regions is significantly greater ($P < 2.2 \times 10^{-16}$, Wilcoxon rank-sum test) than the distribution of RND in collinear regions of the genome. Because the free exchange of the *D. persimilis* SR/ *D. pseudoobscura* ST arrangement appears to have occurred long before the time of species divergence, these results argue against a recent introgression event, and are consistent with incomplete lineage sorting of an ancestral chromosomal arrangement of the *D. persimilis* SR/ *D. pseudoobscura* ST arrangement in the ancestor of both species.

The inference that the *D. persimilis* SR and *D. pseudoobscura* ST chromosomes were freely segregating before the evolution of reproductive isolation between the two species suffers from two potential caveats. First, although some reproductive isolating mechanisms such as hybrid male sterility can potentially evolve quickly, speciation may be considered as a gradual process. Under this scenario, an estimate for the range of time rather than a point estimate for the evolution of reproductive isolation between *D. persimilis* and *D. pseudoobscura* may be more appropriate. Second, recent gene flow between the two species may lead to some degree of homogenization of the two genomes and a reduction in genomic divergence between the two species. This scenario may lead to an underestimate of the species divergence time. Nonetheless, in the absence of information regarding the genes that contribute to reproductive

Table 1. Estimates of the relative ages of chromosomal inversions in *D. persimilis* and *D. pseudoobscura* relative to species divergence time. The fixed inversions on the XL and 2nd chromosomes, as well as the polymorphic inversions on XR and the Pikes Peak (^{3PP}) inversion arose before species divergence.

	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)	0.45	1.00	1.14	1.55	1.65	0.75	0.42	0.59
Time prior to species divergence	0	0.55	0.69	1.10	1.20	0.30	-0.03	0.14

<https://doi.org/10.1371/journal.pgen.1007526.t001>

isolation between the species, there is little guidance for the degree to which the genomic divergence estimates must be adjusted to take into account gene flow after the evolution of reproductive isolation.

We, therefore, pursued a second independent line of enquiry that does not depend on inferences from sequence divergence or differentiation to test whether the *D. persimilis* SR/*D. pseudoobscura* ST chromosomes freely segregated in the ancestor of both species before the evolution of reproductive isolation. Hybrid F1 males between *D. persimilis* and *D. pseudoobscura* are sterile in both directions of the cross, whereas all hybrid females are fully fertile. We determined whether the current day *D. pseudoobscura* ST can be transferred to *D. persimilis* through introgression to yield fertile hybrid males. We used marker assisted backcrossing to transfer the *D. pseudoobscura* ST chromosome into an otherwise *D. persimilis* genetic background. If these hybrid males are fertile, then this may provide strong evidence that introgression of the *D. pseudoobscura* ST arrangement into *D. persimilis* is potentially possible. Despite backcrossing for 15 generations and repeated testing of the fertility of the males from these crosses, all resulting hybrid males were sterile (S4 Fig). Consistent with previous studies, these results indicate the presence of strong hybrid male sterility genes on *D. pseudoobscura* XR [9,43–45]. These results further contradict the recent introgression scenario, and indicate that hybrid male sterility loci on XR must have evolved after these chromosomes were exchanged in the ancestor of both species. Together with the divergence estimates, these results are consistent with the idea that *D. persimilis* SR and *D. pseudoobscura* may have freely segregated in the ancestor of both species prior to the evolution of reproductive isolating loci on XR. More importantly, these results also allow us to provide a range estimate for speciation with a lower bound of approximately 450,000 years based on allelic divergence estimates in collinear regions, and an upper bound of approximately 1 million years ago.

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

Because the XR inversion polymorphism exists only in *D. persimilis* and not in *D. pseudoobscura*, it is often assumed that this inversion must have originated in the *D. persimilis* lineage after the splitting of the two species [31,46]. 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 what is widely-accepted, although this scenario has been hypothesized previously [11]. The two other fixed inversions on the XL and 2nd chromosomes in *D. persimilis* are thought to be even older than the XR inversion [11,46,47]. We estimated divergence between *D. pseudoobscura* and *D. persimilis* ST in sequences flanking the XL and 2nd chromosome inversion breakpoints, and, consistent with previous studies [11,46,47], observed greater levels of divergence for both fixed inversions (RND_{XL}: 0.962, 95% CI: 0.941–0.983; RND₂: 0.941, 95% CI: 0.923–0.959) than for XR (RND_{XR}: 0.808, 95% CI: 0.776–0.840) as the distribution of RND was significantly increased for each ($P < 2.2 \times 10^{-16}$, Wilcoxon rank-sum test; Fig 4). A similar pattern is observed for the median levels of RND in each inversion (RND_{XL}: 0.958; RND₂: 0.937, RND_{XR}: 0.780). The median *D. pseudoobscura*—*D. persimilis* ST RND for each inversion is more than double the genome-wide median RND (RND_{Genome}: 0.259). Likewise, scaling genetic differentiation to the speciation time with *D. miranda*, we estimate that the inversions on XL and the 2nd chromosomes originated approximately 1.64 ± 0.41 and 1.55 ± 0.24 million years ago, respectively (Table 1; Fig 5). From the analysis of *D. pseudoobscura* and *D. persimilis* ST divergence in 10kb sliding windows, we observe a significant overrepresentation of RND estimates in the top 1% genome-wide across all three inversions

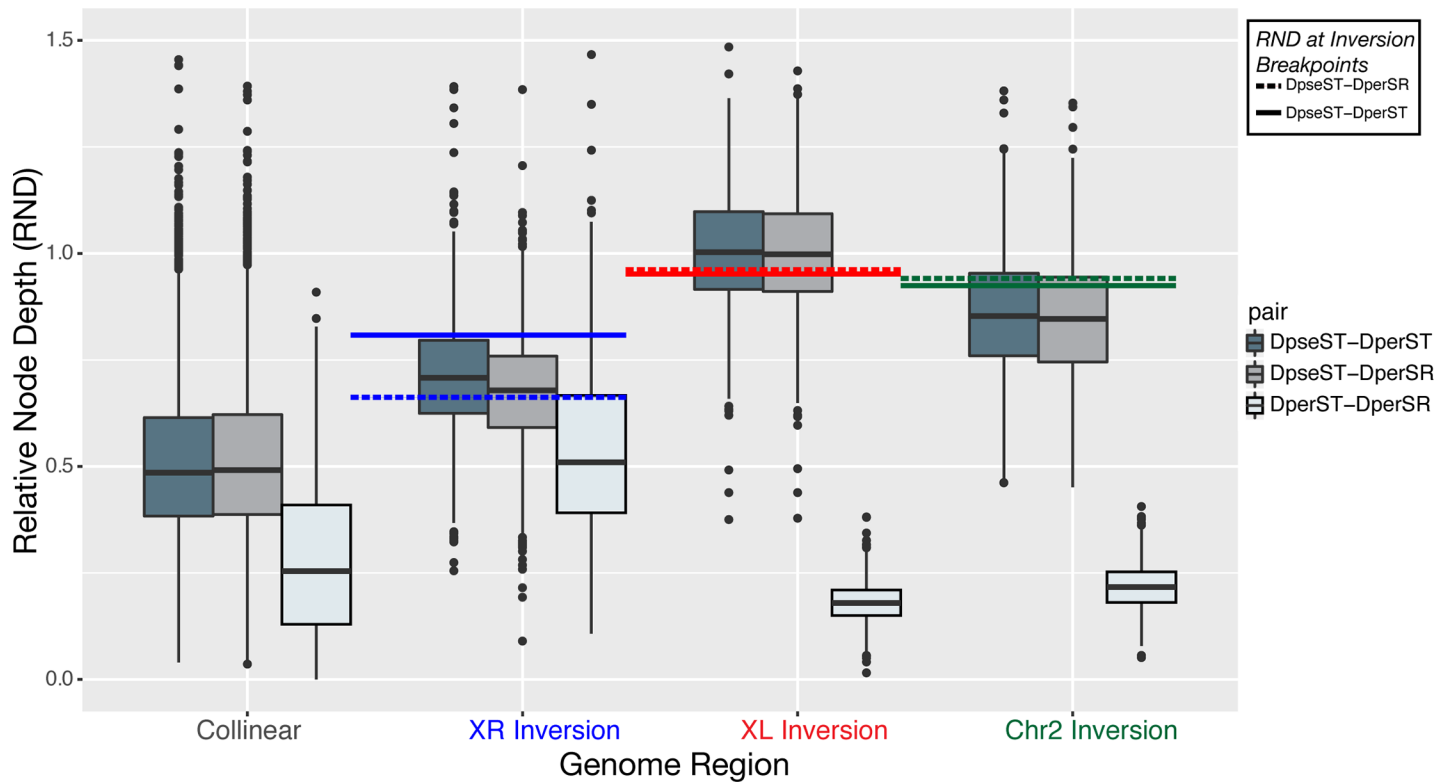


Fig 4. The distribution of divergence estimated across genomic regions. Divergence was estimated in 10 kb windows as the Relative Node Depth (RND; d_{xy} normalized to the outgroup) across the genome. The boxplots show the distribution of RND for each comparison in all collinear regions, and across the XR, XL and 2nd chromosome inversions. The horizontal lines depicted in the three fixed inversions indicate the mean RND estimated in the regions flanking the inversion breakpoints (± 250 kb) for *D. pseudoobscura*-*D. persimilis* ST (solid) and *D. pseudoobscura*-*D. persimilis* SR (dashed).

<https://doi.org/10.1371/journal.pgen.1007526.g004>

relative to collinear regions ($\chi^2 = 208.3, P < 2 \times 10^{-16}$; S5 Fig). The proportion of RND windows in the top 1% is greatest on the XL inversion, followed by the 2nd chromosome inversion, with the fewest across the XR inversion (S5 Fig). Our results suggest that all of these fixed inversions originated in the ancestral population before the speciation event that separated *D. persimilis* and *D. pseudoobscura*. Furthermore, the relative divergence and differentiation pattern of $XL > 2 > XR$ that we infer is consistent with findings from previous studies [10,47].

The difference in divergence and differentiation of the fixed inversions and collinear regions is not subtle (Fig 4): the XL, XR and 2nd chromosome inversions are nearly twice as old as the estimates for collinear regions between the two species and the distributions of RND are significantly greater for each (Fig 4, S5 Fig). The increased divergence we observe in the fixed XL, XR, and 2nd inversions is not a novel finding and has been well documented by others [11,46,47]. Although the possibility of these inversions arising in the ancestral population has previously been raised, all studies to date have concluded that the reduced divergence in collinear regions is the result of gene flow upon secondary contact and that all inversions must have originated after speciation [11,12,14,15,46,48]. To test if the fixed inversions originated as segregating polymorphisms in the ancestral species as our results suggest, we modeled divergence and gene flow under alternative evolutionary scenarios of speciation.

Using loci sampled from intergenic regions across inverted and collinear regions of the genome, we fit our data to models of strict divergence in isolation, isolation-with-migration (IM), and isolation-with-initial-migration (IIM) with maximum-likelihood estimation [49]. In collinear regions the IIM model gave a significantly better fit than the IM model or a null

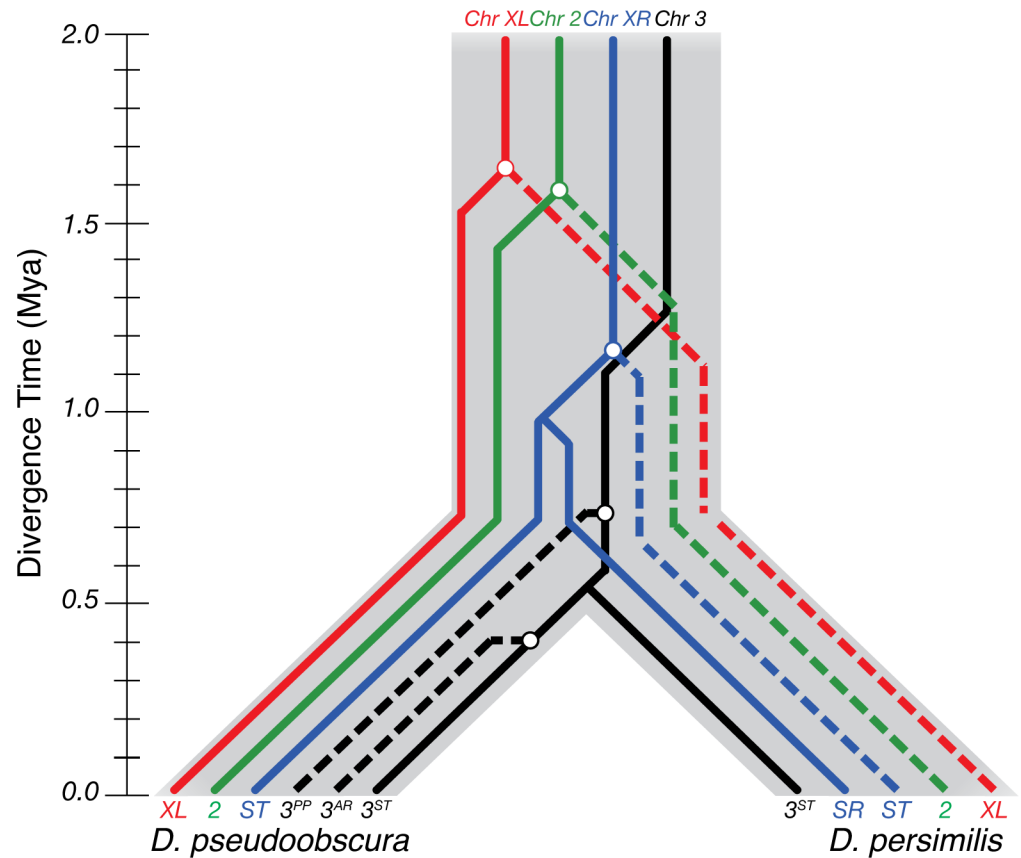


Fig 5. Incomplete lineage sorting of the inversions of *D. persimilis* and *D. pseudoobscura*. The fixed inversions on the XL and 2nd chromosomes, as well as the polymorphic inversions on XR and the Pikes Peak (3^{PP}) inversion arose before species divergence. Incomplete lineage sorting produced the observed inversion patterns in the species present today.

<https://doi.org/10.1371/journal.pgen.1007526.g005>

model of strict divergence (Table 2), providing further evidence of post-speciation gene flow as supported by several previous studies [11,12,14,15,46,48]. Under the IIM model, the estimated time of population divergence in inverted regions should represent its origin [50]. To test if the inversions are associated with an older population divergence time than collinear regions and therefore predate the species split, we allowed the parameters of the IIM model to vary independently between each inversion and collinear regions and compared the results to a fully constrained model where the parameters are fixed between regions [50]. The model allowing for individual parameters to differ between regions fit the data significantly better

Table 2. Maximum likelihood support and likelihood ratio tests for gene flow under models of speciation. The log-likelihoods are displayed for isolation (Iso), isolation-with-migration (IM), and isolation-with-initial-migration (IIM) models. The estimates in bold correspond to the maximum likelihoods for each genomic region. In each case, the IIM model has the best support. The columns labeled Iso and IM show the likelihood ratio test statistics for each model relative to the IIM model.

	$\ln L_{Iso}$	$\ln L_{IM}$	$\ln L_{IIM}$	Iso	IM
Full Genome	-19000.18	-18899.42	-18698.87	77.81	51.79
Collinear	-13848.79	-13689.17	-13591.07	66.55	25.34
Chr XL Inversion	-1390.10	-1390.10	-1372.71	4.49	4.49
Chr 2 Inversion	-1429.95	-1429.95	-1408.24	18.52	18.52
Chr XR Inversion	-2251.10	-2241.11	-2225.54	6.61	4.02

<https://doi.org/10.1371/journal.pgen.1007526.t002>

($\chi^2 = 26.2$, $P < 8.6 \times 10^{-6}$), indicating that the *XL*, *XR* and 2^{nd} inversions arose prior to the population divergence in collinear regions and further supporting the idea that they existed as ancestral polymorphisms (S6 Fig). For each inversion, the parameter estimate for the population divergence time t_0 is greater than in collinear regions, although we note the confidence intervals overlap for the case of *XR*. However, we find evidence to support that t_0 is different between the *XR* inversion and collinear regions, as a model where we allow parameters to vary in each fits the data significantly better than a constrained model where divergence parameters are held constant ($2\Delta\ln L = -6.76$; $P < 3.4 \times 10^{-2}$). In each region, we estimate one-way gene flow from *D. pseudoobscura* to *D. persimilis* and no migration in the other direction (S6 Fig). Although we find evidence for gene flow from *D. pseudoobscura* to *D. persimilis* after speciation in agreement with several previous studies [11,12,14,15,46,48], we do not conclude this is solely responsible for the pattern of increased divergence observed across fixed inversion differences. Instead, these results indicate 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 the reproductive barriers were complete.

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 important implications for all of these fields. We provide a resolution to the strange collinearity of the *D. persimilis* *SR* and *D. pseudoobscura* *ST* chromosomes first observed by Dobzhansky [24,51]. 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. Our results also provide evidence that pervasive gene flow after the initial evolution of reproductive isolation is not necessarily required to explain the observed phylogenetic discordance. A similar maintenance of chromosomal arrangements across species resulting from an ancient inversion polymorphism has also been demonstrated in *Anopheles* mosquitoes [52]. 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 [27]. In contrast to most other *Sex-Ratio* systems associated with derived inversions, our results suggest that the *D. persimilis* *SR* system evolved on the background of an ancestral arrangement. Similarly, recent studies of the *t*-haplotype in *M. musculus* also support an ancient origin of inversions associated with segregation distortion [53]. These results indicate that segregation distorters may not only become associated with new inversions, as is traditionally thought, but can also arise on the genetic backgrounds of existing chromosome inversion polymorphisms.

In addition to clarifying the evolutionary history of *Sex-Ratio* chromosome in *D. persimilis*, the age estimates of the fixed chromosomal inversion differences in *D. pseudoobscura* and *D. persimilis* suggest a new role of chromosomal inversions in the evolution of reproductive isolation genes. 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 [7–9]. Previous work in this species pair reconciled these empirical observations with a model where inversions arise after speciation and secondary contact between taxa homogenizes collinear regions [11]. Thus, previous models explained the role of chromosomal inversions in speciation as protectors of hybrid incompatibly alleles from the homogenizing force of

extensive hybridization [9,11]. Instead, we show that these inversions were freely segregating in the ancestral population long before the complete isolation of *D. pseudoobscura* and *D. persimilis*, and that genes contributing to reproductive barriers must have evolved within them afterwards.

Here, we propose a simple model under which ancestrally segregating inversions that undergo incomplete lineage sorting can lead to high allelic divergence at these inversions, which may in turn accelerate the formation of hybrid incompatibilities (Fig 6). Chromosomal inversions can arise and persist in ancestral populations [18]. 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 [18,54–57]. Populations with ancient segregating inversions diverge within inverted regions, but stay genetically similar in collinear regions [54,57]. These chromosomal inversions may undergo incomplete lineage sorting if the ancestral population is split into two allopatric populations [58]. At the initial time of separation, all loci across collinear and inverted backgrounds start as equally compatible, the genes in collinear regions are nearly identical, while genes within the chromosomal inversions are already highly diverged. This accumulation of hybrid incompatibilities occurs in isolation, unopposed by the selective cost of producing unfit offspring, and in a manner consistent with the Dobzhansky-Muller model [59,60]. The collinear regions will retain their low divergence signature from the ancestral population until speciation is complete. Under this 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.

Our reasoning that highly diverged genes may evolve to an incompatible state more quickly than those with little divergence rests on the implicit assumption that the evolution of hybrid incompatibilities requires multiple genetic changes. This view, although somewhat speculative, is supported by three lines of evidence. First, theory shows that changes at a minimum of two genes are required to produce a hybrid incompatibility, and that it may be easier to evolve more complex incompatibilities that involve changes at multiple genes [61]. These ideas have strong empirical support [1]. For example, the genetic architecture of hybrid sterility between *D. pseudoobscura pseudoobscura* and *D. pseudoobscura bogotana*—one of the youngest hybridizations to be studied—involves a single hybrid incompatible interaction between *at least* six genes [45]. Second, nearly all hybrid incompatibility genes that have been identified so far show the rapid accumulation of many amino acid changes, and represent some of the most highly diverged genes in the genome [62,63]. Ultra-fine scale mapping studies that dissect how many of these changes within these genes contribute to hybrid sterility or hybrid inviability have not yet been performed. However, there are no known cases of hybrid incompatibility genes that involve one or only a few amino acid changes. Third, both theory and empirical data show that hybrid incompatibilities accumulate faster than linearly with divergence between populations [64–66]. Populations that display higher genomic divergence are, therefore, more likely to have evolved hybrid incompatibilities as compared to those that have little or no genomic divergence [67]. Together, these lines of evidence support the idea that the evolution of hybrid incompatibilities is a multi-step process. By accumulating genetic divergence even before the initial population split, the genes associated with ancestrally segregating chromosomal inversions may be fewer steps away from reaching an incompatible state. In contrast, genes in collinear regions of the genome show little or no divergence between recently split populations and must start accumulating changes from scratch if they are to eventually reach an incompatible state.

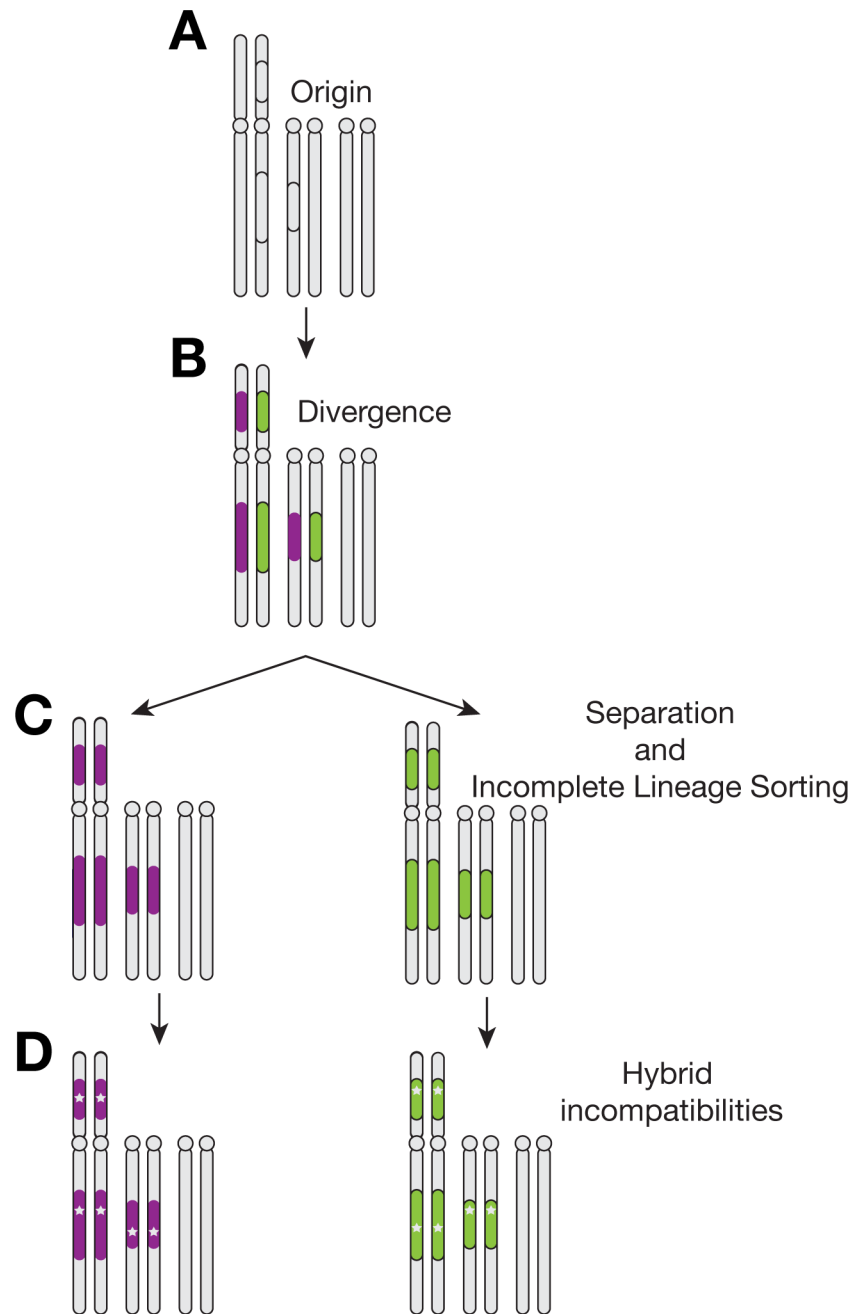


Fig 6. Inversions accelerate the formation of hybrid incompatibilities. (A) Polymorphic inversions arise in the ancestor of the two species. (B) Restricted recombination between the inversions leads to accumulating divergence (red, blue) distinct from collinear regions of the genome (grey). (C) Incomplete 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.

<https://doi.org/10.1371/journal.pgen.1007526.g006>

The idea that chromosomal inversions are often associated with hybrid incompatibility genes is a widely-held view among evolutionary geneticists [18,68]. There are four lines of evidence for the widespread acceptance of this association. First, direct genetic mapping of loci that underlie reproductive barriers may show these genes to be located in genomic regions that harbor fixed chromosomal inversions [9,43,69,70]. Such genetic studies provide the most

direct line of evidence for a potential association of reproductive isolation genes with chromosomal inversions. Second, genomic regions spanning chromosomal inversions often show signatures of higher divergence or reduced introgression [11,19,46]. As our results show, this line of evidence may be susceptible to erroneous interpretations when the evolutionary histories and the ages of these inversions are unknown. Third, sympatric species show higher incidence of fixed inversions than allopatric species. While there are limited data supporting such a pattern [9,47,71,72], this line of evidence for the association of hybrid incompatibility genes with chromosomal inversions is indirect and prone to observational biases. Fourth, theoretical studies show that it may be possible for hybrid incompatibility genes to evolve and persist despite gene flow during or after speciation [20,73]. These theoretical results, however, are not a good substitute for direct empirical evidence. We, therefore, consider direct genetic mapping studies that localize reproductive isolation genes to regions spanning chromosomal inversions as the most reliable line of evidence supporting the association of chromosomal inversions with reproductive isolation genes. Such genetic studies that map loci that contribute to reproductive isolating barriers, and overlay those loci on the locations of chromosomal inversions are surprisingly rare. To our knowledge, the only direct study of this nature in animal taxa involves the *D. pseudoobscura*-*D. persimilis* hybridization, where genetic mapping studies have shown that loci that contribute to reproductive isolation are enriched, but not exclusively located, on chromosomes that also carry fixed inversion differences between these species [9,47,60,74,75]. In the absence of other such studies, it is not clear whether this pattern is specific to this particular species pair, or is a broadly held pattern. We, therefore, find that the amount of evidence for the association of hybrid incompatibility genes with fixed chromosomal inversions is not proportionate to how widely this pattern is believed to be true.

This paucity of genetic mapping studies to determine the locations of hybrid incompatibility genes relative to chromosomal inversions is not entirely surprising. A necessary step in understanding the molecular basis of speciation involves the identification of the genes that contribute to reproductive barriers. Most speciation geneticists who aim to identify such genes may either focus on studying species pairs that lack chromosomal inversion differences, or abandon such studies when these genes map to chromosomal inversions because there is little hope of precisely identifying the causal genes. Fortunately, uncovering evidence for an association of reproductive isolation genes with chromosomal inversions requires neither the precise identification of the genes nor determining the precise breakpoints of chromosomal inversions. Coarse mapping of quantitative trait loci that underlie reproductive isolation across several species pairs, and overlaying these loci with the approximate locations of chromosomal inversion differences between these species may prove sufficient to establish the generality of this pattern [69].

In summary, we propose that incomplete lineage sorting of ancestrally segregating polymorphisms can establish patterns of higher divergence within chromosomal inversions, and may potentially promote the evolution of hybrid incompatibilities in these highly diverged regions. Our model can explain previously observed empirical patterns even in cases where there is no evidence for gene flow across populations during or after speciation. Together, these ideas force a reconsideration of the role of chromosomal inversions in speciation, perhaps not as protectors of existing hybrid incompatibility alleles, but as fertile grounds for their formation.

Materials and methods

Isolation and maintenance of Sex-Ratio chromosome strains

Wild caught *D. persimilis* strains were provided as a generous gift by Dean Castillo, collected in the Sierra Nevada mountain range and near 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 analyses

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 sepi* (*se*) 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 se* 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) [76–78].

DNA extraction 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 manufacturer's 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 [79]. 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 [80]. Genome wide, the average fold coverage was ~180x and ~133x for the *D. persimilis ST* and *SR* pools, respectively (S1 Table). For reads mapping to X chromosome scaffolds, the average fold coverage was ~97x and ~74x for *D. persimilis ST* and *SR*, respectively (S2 Table).

After the binary alignments were sorted and indexed with SAMtools [81], single nucleotide polymorphisms (SNPs) were called using *freebayes* (v. 0.9.21; [82] with the expected pairwise nucleotide diversity parameter set to 0.01, based on a previous genome-wide estimate from *D. pseudoobscura* [55]. The samples were modeled as discrete genotypes across pools by using the “-j” option and the ploidy was set separately for X chromosome scaffolds (1N) and autosomes (2N). 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 (S1 Table). 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 4th chromosomes. Instead, each is composed of a series of scaffold groups that differ both in size and orientation relative to one another [83]. Schaeffer et al. (2008) previously determined the approximate locations and ordering of each of these scaffolds [83]. 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 [84,85]. To root the tree with an outgroup, we aligned publicly available short reads of *D. miranda* (SRX965461; strain SP138) to the *D. pseudoobscura* reference genome. In each window, we constructed neighbor-joining trees [86] 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 allelic divergence with Nei's d_{xy} , a measure of the average number of pairwise nucleotide substitutions per site [87,88]. d_{xy} was measured between each population grouping in 10 Kb, nonoverlapping windows across the genome. Each comparison was then normalized to the divergence with the outgroup *D. miranda* in the same window to account for regional mutational differences, a measure known as the "relative node depth" [40]. Confidence intervals were determined from 1000 bootstrap replicates of windows in each region under consideration. Divergence time estimates were obtained with the Cavalli-Sforza transformation of F_{ST} as

$$T = -\log(1 - F_{ST})$$

and then multiplied by a scaling factor in each window so that the divergence time between *D. pseudoobscura* and *D. miranda* was equal to 2 Mya [42,89–91].

Modeling gene flow

To test for evidence of post-speciation gene flow we considered three different models: (i) strict divergence in isolation (Iso) with an instantaneous split of an ancestral population at time t_0 without any gene flow after, (ii) isolation-with-migration (IM) where an ancestral population split into two subpopulations at time t_0 with constant migration rates M_1 and M_2 between them afterwards, and (iii) isolation-with-initial-migration (IIM) where gene flow is restricted to occur over a time V after the initial split, ceasing at time t_1 . We used the methods derived by Costa and Wilkinson-Herbots (2017) to obtain maximum-likelihood estimates for the parameters under each model. Becquet and Przeworski (2009) and Strasburg and Rieseberg (2010), among others, have shown that parameter estimation with IM models can be unreliable if assumptions concerning population structure and recombination are broken [92,93]. While the maximum-likelihood method of Costa and Wilkinson-Herbots has been demonstrated to be robust to demographic misspecification, we nonetheless do not rely on this analysis to provide accurate parameter estimates of divergence times and instead use the approach to test for the relative support among speciation models. Some previous studies have suggested that gene flow between *D. pseudoobscura* and *D. persimilis* has occurred upon

secondary contact more recently after initial isolation, however the IIM model has been shown to approximate the dynamics of this scenario reasonably well [49]. To remove potential confounding effects of selection, we restricted our analysis to intergenic noncoding regions of each chromosome. We then randomly sampled 500 bp segments that were separated by a minimum of at least 10 kb to create a set of loci for each region, similar to the multilocus dataset of Wang and Hey (2010). The coalescent models of Costa and Wilkinson-Herbots (2017) require separate estimates of pairwise differences in loci (i) within *D. pseudoobscura*, (ii) within *D. persimilis*, and (iii) between *D. pseudoobscura* and *D. persimilis*. Therefore, we randomly divided the loci for each analysis into three nonoverlapping datasets. Relative mutation rates are also required for each locus. Here, as recommended by Costa and Wilkinson-Herbots (2017), we used the divergence (i.e. d_{xy}) to the outgroup *D. miranda* to estimate these relative mutation rates [94,95].

We used likelihood ratio tests to determine the relative support for each model, where the difference in log-likelihood between models $2\Delta\ln L$ is assumed to follow a χ^2 distribution with the number of degrees of freedom equal to the difference between the dimensions of parameter space of the two models. The maximum-likelihood estimates for each model can be computed rapidly because linkage is assumed to be negligible between loci. Thus, to correct for the statistical effect of LD between loci, we scaled the difference in $\ln L$ between models by a factor of $1/x$ as in Lohse et al. (2015), where x is the average number of loci sampled in each 100 kb region ($x = 7.75$). To test if the fixed inversions are older than the species split we allowed individual parameters of the IIM model to vary between collinear regions, and each of the XL , XR and 2^{nd} inversions. We then compared this complex model to a constrained model, where each parameter was fixed across the genome, similar to the hierarchical model testing in Lohse et al. (2015). The confidence intervals reported for each parameter are the Wald confidence intervals computed from the inverted Hessian matrix of the maximum-likelihood estimators [49].

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'- GATCTAATCCAGAAAGTTTCGCTTGCG -3', R5'- AGTGTGACCCATTTAAGCGG-3'). These primers amplified a single, approximately 1500bp, product in *D. pseudoobscura* and *D. persimilis* SR, but not *D. persimilis* ST. 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 the unique sequence flanking the repeat region at the proximal breakpoint.

Supporting information

S1 Text. Supplementary methods for phylogenetic and divergence analyses in *D. pseudoobscura*. This text details the methods used to analyze phylogenetic discordance on the third chromosome of *D. pseudoobscura* and *D. persimilis*. Further, this text contains the methods used to determine the relative age of the Arrowhead (3^{AR}) and Pikes Peak (3^{PP})

arrangements in *D. pseudoobscura*.
(PDF)

S1 Fig. Polytene squash of a *D. persimilis* ST/SR female heterozygote. The XR chromosome is contains a single inversion as observed by a characteristic inversion loop. The remainder of the genome is homosequential.

(PDF)

S2 Fig. PCR amplification of the proximal breakpoint. Genomic template from *D. pseudoobscura* and *D. persimilis* SR, but not *D. persimilis* ST, generated an approximately 1.5kb amplicon of the proximal breakpoint with primers specific for the ancestral orientation of the XR chromosome.

(PDF)

S3 Fig. Species clustering within inversion polymorphisms on chromosome 3. The *D. pseudoobscura* 3rd 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.

(PDF)

S4 Fig. Introgression of the *D. pseudoobscura* ST arrangement into a *D. persimilis* genetic background. Despite 15 generations of marker-assisted backcrossing, all hybrid males that carry the *D. pseudoobscura* XR material in an otherwise *D. persimilis* genetic background are sterile. These results indicate that the chromosome-level gene exchange must have happened before the evolution of hybrid incompatibilities on this chromosome arm.

(PDF)

S5 Fig. Divergence in sliding windows across chromosomes. Smoothing splines are shown for divergence measured as relative node depth (RND) in 10kb windows across chromosomes XR (A), XL (B), and 2 (C). The different colors for each line indicate the taxa pair RND is estimated for, with the key in the legend. Colored dots represent individual windows that are in the top 1% of RND values genome-wide and are considered outliers. Black vertical lines indicate the locations of inversion breakpoints on each chromosome. The insets on XR show a close-up view of RND estimated around the proximal and distal inversion breakpoints ± 250 kb.

(PDF)

S6 Fig. Isolation with initial migration model. The width of the bars are proportional to the population sizes and the heights of bars indicate time using the maximum likelihood approach of Costa and Wilkinson-Herbots (2017). The ancestral population for each set of data is indicated by a single colored bar that splits into two subpopulations at time t_0 . From t_0 to t_1 (V) the populations diverge in allopatry with the estimated levels of gene flow (M ; in units of number of migrants per generation). At time t_1 , the populations no longer exchange genes among the subpopulations. The vertical white bars are the confidence intervals for time t_0 and t_1 . The col-linear region represents species divergence, while XR, 2, and XL represent the divergence of fixed inversion differences between *D. pseudoobscura* and *D. persimilis*.

(PDF)

S1 Table. *D. pseudoobscura* and *D. persimilis* reference alignment statistics. Statistics are presented for the total number of reads mapped to the *D. pseudoobscura* reference genome for each sample and the *D. miranda* outgroup.

(PDF)

S2 Table. Reference alignment statistics for each sample across scaffolds. Each scaffold in the *D. pseudoobscura* reference genome is listed, with corresponding coverage and mapping statistics for each sample and the *D. miranda* outgroup.

(PDF)

Acknowledgments

We thank Dean Castillo for generously providing wild-caught *D. persimilis* flies. We are particularly grateful to Molly Schumer, and Matthew Hahn for his third reviewer services (@3rdreviewer) and for originally asking us to consider an incomplete lineage sorting hypothesis.

Author Contributions

Conceptualization: Zachary L. Fuller, Christopher J. Leonard, Stephen W. Schaeffer, Nitin Phadnis.

Data curation: Zachary L. Fuller, Christopher J. Leonard, Stephen W. Schaeffer, Nitin Phadnis.

Formal analysis: Zachary L. Fuller, Christopher J. Leonard, Randee E. Young, Stephen W. Schaeffer.

Funding acquisition: Stephen W. Schaeffer, Nitin Phadnis.

Investigation: Zachary L. Fuller, Christopher J. Leonard, Randee E. Young, Stephen W. Schaeffer, Nitin Phadnis.

Methodology: Zachary L. Fuller, Christopher J. Leonard, Randee E. Young, Stephen W. Schaeffer, Nitin Phadnis.

Project administration: Stephen W. Schaeffer, Nitin Phadnis.

Resources: Randee E. Young, Stephen W. Schaeffer, Nitin Phadnis.

Software: Zachary L. Fuller, Christopher J. Leonard, Stephen W. Schaeffer.

Supervision: Stephen W. Schaeffer, Nitin Phadnis.

Validation: Zachary L. Fuller, Stephen W. Schaeffer.

Visualization: Zachary L. Fuller, Christopher J. Leonard, Stephen W. Schaeffer, Nitin Phadnis.

Writing – original draft: Zachary L. Fuller, Christopher J. Leonard, Stephen W. Schaeffer, Nitin Phadnis.

Writing – review & editing: Zachary L. Fuller, Stephen W. Schaeffer, Nitin Phadnis.

References

1. Coyne JA, Orr HA. Speciation. Sinauer; 2004.
2. Dobzhansky T, Dobzhansky TG. Genetics and the Origin of Species. Columbia University Press; 1937.
3. White MJD. Modes of speciation. San Francisco: W.H. Freeman; 1978.

4. Dobzhansky T. On the Sterility of the Interracial Hybrids in *Drosophila Pseudoobscura*. *Proc Natl Acad Sci USA*. 1933; 19: 397–403. PMID: [16577530](#)
5. Stebbins GL. The inviability, weakness, and sterility of interspecific hybrids. *Adv Genet*. 1958; 9: 147–215. PMID: [13520442](#)
6. Stebbins GL. Variation and Evolution in Plants: Progress During the Past Twenty Years. *Essays in Evolution and Genetics in Honor of Theodosius Dobzhansky*. Springer, Boston, MA; 1970. pp. 173–208. Available: https://link.springer.com/chapter/10.1007/978-1-4615-9585-4_6
7. Wu CI, Beckenbach AT. Evidence for Extensive Genetic Differentiation between the Sex-Ratio and the Standard Arrangement of *DROSOPHILA PSEUDOBSCURA* and *D. PERSIMILIS* and Identification of Hybrid Sterility Factors. *Genetics*. 1983; 105: 71–86. PMID: [17246158](#)
8. Brown KM, Burk LM, Henagan LM, Noor MAF. A test of the chromosomal rearrangement model of speciation in *Drosophila pseudoobscura*. *Evolution*. 2004; 58: 1856–1860. PMID: [15446438](#)
9. Noor MAF, Grams KL, Bertucci LA, Reiland J. Chromosomal inversions and the reproductive isolation of species. *PNAS*. 2001; 98: 12084–12088. <https://doi.org/10.1073/pnas.221274498> PMID: [11593019](#)
10. Machado CA, Haselkorn TS, Noor MAF. 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*. 2007; 175: 1289–1306. <https://doi.org/10.1534/genetics.106.064758> PMID: [17179068](#)
11. Kulathinal RJ, Stevison LS, Noor MAF. The Genomics of Speciation in *Drosophila*: Diversity, Divergence, and Introgression Estimated Using Low-Coverage Genome Sequencing. *PLOS Genetics*. 2009; 5: e1000550. <https://doi.org/10.1371/journal.pgen.1000550> PMID: [19578407](#)
12. Machado CA, Kliman RM, Markert JA, Hey J. Inferring the history of speciation from multilocus DNA sequence data: the case of *Drosophila pseudoobscura* and close relatives. *Mol Biol Evol*. 2002; 19: 472–488. <https://doi.org/10.1093/oxfordjournals.molbev.a004103> PMID: [11919289](#)
13. Rieseberg LH. Chromosomal rearrangements and speciation. *Trends Ecol Evol (Amst)*. 2001; 16: 351–358.
14. Wang RL, Wakeley J, Hey J. Gene flow and natural selection in the origin of *Drosophila pseudoobscura* and close relatives. *Genetics*. 1997; 147: 1091–1106. PMID: [9383055](#)
15. Hey J, Nielsen R. Multilocus methods for estimating population sizes, migration rates and divergence time, with applications to the divergence of *Drosophila pseudoobscura* and *D. persimilis*. *Genetics*. 2004; 167: 747–760. <https://doi.org/10.1534/genetics.103.024182> PMID: [15238526](#)
16. Powell JR. Interspecific cytoplasmic gene flow in the absence of nuclear gene flow: evidence from *Drosophila*. *Proc Natl Acad Sci USA*. 1983; 80: 492–495. PMID: [6300849](#)
17. Dobzhansky T. Is there Gene Exchange between *Drosophila pseudoobscura* and *Drosophila persimilis* in Their Natural Habitats? *The American Naturalist*. 1973; 107: 312–314. <https://doi.org/10.2307/2459801>
18. Kirkpatrick M, Barton N. Chromosome Inversions, Local Adaptation and Speciation. *Genetics*. 2006; 173: 419–434. <https://doi.org/10.1534/genetics.105.047985> PMID: [16204214](#)
19. Stevison LS, Hoehn KB, Noor MAF. Effects of Inversions on Within- and Between-Species Recombination and Divergence. *Genome Biol Evol*. 2011; 3: 830–841. <https://doi.org/10.1093/gbe/evr081> PMID: [21828374](#)
20. Feder JL, Nosil P. Chromosomal Inversions and Species Differences: When Are Genes Affecting Adaptive Divergence and Reproductive Isolation Expected to Reside Within Inversions? *Evolution*. 2009; 63: 3061–3075. <https://doi.org/10.1111/j.1558-5646.2009.00786.x> PMID: [19656182](#)
21. Wadsworth CB, Li X, Dopman EB. A recombination suppressor contributes to ecological speciation in *OSTRINIA* moths. *Heredity (Edinb)*. 2015; 114: 593–600. <https://doi.org/10.1038/hdy.2014.128> PMID: [25626887](#)
22. Navarro A, Barton NH. Chromosomal speciation and molecular divergence—accelerated evolution in rearranged chromosomes. *Science*. 2003; 300: 321–324. <https://doi.org/10.1126/science.1080600> PMID: [12690198](#)
23. Ayala FJ, Coluzzi M. Chromosome speciation: humans, *Drosophila*, and mosquitoes. *Proc Natl Acad Sci USA*. 2005; 102 Suppl 1: 6535–6542. <https://doi.org/10.1073/pnas.0501847102> PMID: [15851677](#)
24. Sturtevant AH, Dobzhansky T. Geographical Distribution and Cytology of “Sex Ratio” in *Drosophila Pseudoobscura* and Related Species. *Genetics*. 1936; 21: 473–490. PMID: [17246805](#)
25. Policansky D, Ellison J. “Sex ratio” in *Drosophila pseudoobscura*: spermiogenic failure. *Science*. 1970; 169: 888–889. PMID: [5432586](#)

26. Bastide H, Gérard PR, Ogereau D, Cazemajor M, Montchamp-Moreau C. Local dynamics of a fast-evolving sex-ratio system in *Drosophila simulans*. *Mol Ecol*. 2013; 22: 5352–5367. <https://doi.org/10.1111/mec.12492> PMID: 24118375
27. Jaenike J. Sex Chromosome Meiotic Drive. *Annual Review of Ecology and Systematics*. 2001; 32: 25–49. <https://doi.org/10.1146/annurev.ecolsys.32.081501.113958>
28. Presgraves DC, Gérard PR, Cherukuri A, Lyttle TW. Large-Scale Selective Sweep among Segregation Distorter Chromosomes in African Populations of *Drosophila melanogaster*. *PLOS Genetics*. 2009; 5: e1000463. <https://doi.org/10.1371/journal.pgen.1000463> PMID: 19412335
29. Lyttle TW. Segregation Distorters. *Annual Review of Genetics*. 1991; 25: 511–581. <https://doi.org/10.1146/annurev.ge.25.120191.002455> PMID: 1812815
30. Policansky D, Zouros E. Gene Differences between the Sex Ratio and Standard Gene Arrangements of the X Chromosome in *DROSOPHILA PERSIMILIS*. *Genetics*. 1977; 85: 507–511. PMID: 17248742
31. Babcock CS, Anderson WW. Molecular evolution of the Sex-Ratio inversion complex in *Drosophila pseudoobscura*: analysis of the Esterase-5 gene region. *Mol Biol Evol*. 1996; 13: 297–308. <https://doi.org/10.1093/oxfordjournals.molbev.a025589> PMID: 8587496
32. Kovacevic M, Schaeffer SW. Molecular population genetics of X-linked genes in *Drosophila pseudoobscura*. *Genetics*. 2000; 156: 155–172. PMID: 10978282
33. Garfield DA, Noor MA. Characterization of novel repetitive element Leviathan in *Drosophila pseudoobscura*. *Drosophila Information Service*. 2007; 90: 1–9.
34. Aguado C, Gayà-Vidal M, Villatoro S, Oliva M, Izquierdo D, Giner-Delgado C, et al. Validation and Genotyping of Multiple Human Polymorphic Inversions Mediated by Inverted Repeats Reveals a High Degree of Recurrence. *PLOS Genetics*. 2014; 10: e1004208. <https://doi.org/10.1371/journal.pgen.1004208> PMID: 24651690
35. Puerma E, Orengo DJ, Salguero D, Papaceit M, Segarra C, Aguadé M. Characterization of the breakpoints of a polymorphic inversion complex detects strict and broad breakpoint reuse at the molecular level. *Mol Biol Evol*. 2014; 31: 2331–2341. <https://doi.org/10.1093/molbev/msu177> PMID: 24881049
36. Puerma E, Orengo DJ, Aguadé M. Multiple and diverse structural changes affect the breakpoint regions of polymorphic inversions across the *Drosophila* genus. *Scientific Reports*. 2016; 6: 36248. <https://doi.org/10.1038/srep36248> PMID: 27782210
37. González J, Casals F, Ruiz A. Testing Chromosomal Phylogenies and Inversion Breakpoint Reuse in *Drosophila*. *Genetics*. 2007; 175: 167–177. <https://doi.org/10.1534/genetics.106.062612> PMID: 17028333
38. Navarro A, Barbadilla A, Ruiz A. Effect of Inversion Polymorphism on the Neutral Nucleotide Variability of Linked Chromosomal Regions in *Drosophila*. *Genetics*. 2000; 155: 685–698. PMID: 10835391
39. Navarro A, Betrán E, Barbadilla A, Ruiz A. Recombination and Gene Flux Caused by Gene Conversion and Crossing Over in Inversion Heterokaryotypes. *Genetics*. 1997; 146: 695–709. PMID: 9178017
40. Feder JL, Xie X, Rull J, Velez S, Forbes A, Leung B, et al. Mayr, Dobzhansky, and Bush and the complexities of sympatric speciation in *Rhagoletis*. *PNAS*. 2005; 102: 6573–6580. <https://doi.org/10.1073/pnas.0502099102> PMID: 15851672
41. Graur D, Martin W. Reading the entrails of chickens: molecular timescales of evolution and the illusion of precision. *Trends Genet*. 2004; 20: 80–86. <https://doi.org/10.1016/j.tig.2003.12.003> PMID: 14746989
42. Wang RL, Hey J. The Speciation History of *Drosophila Pseudoobscura* and Close Relatives: Inferences from DNA Sequence Variation at the Period Locus. *Genetics*. 1996; 144: 1113–1126. PMID: 8913754
43. McDermott SR, Noor MAF. Mapping of within-species segregation distortion in *D. persimilis* and hybrid sterility between *D. persimilis* and *D. pseudoobscura*. *J Evol Biol*. 2012; 25: 2023–2032. <https://doi.org/10.1111/j.1420-9101.2012.02581.x> PMID: 22966762
44. Phadnis N, Orr HA. A single gene causes both male sterility and segregation distortion in *Drosophila* hybrids. *Science*. 2009; 323: 376–379. <https://doi.org/10.1126/science.1163934> PMID: 19074311
45. Phadnis N. Genetic Architecture of Male Sterility and Segregation Distortion in *Drosophila pseudoobscura* Bogota–USA Hybrids. *Genetics*. 2011; 189: 1001–1009. <https://doi.org/10.1534/genetics.111.132324> PMID: 21900263
46. Noor MAF, Garfield DA, Schaeffer SW, Machado CA. Divergence Between the *Drosophila pseudoobscura* and *D. persimilis* Genome Sequences in Relation to Chromosomal Inversions. *Genetics*. 2007; 177: 1417–1428. <https://doi.org/10.1534/genetics.107.070672> PMID: 18039875
47. McGaugh SE, Noor MAF. Genomic impacts of chromosomal inversions in parapatric *Drosophila* species. *Philos Trans R Soc Lond B Biol Sci*. 2012; 367: 422–429. <https://doi.org/10.1098/rstb.2011.0250> PMID: 22201171

48. Noor MA, Johnson NA, Hey J. Gene flow between *Drosophila pseudoobscura* and *D. persimilis*. *Evolution*. 2000; 54: 2174–2175; discussion 2176–2177. PMID: [11209795](https://pubmed.ncbi.nlm.nih.gov/11209795/)
49. Costa RJ, Wilkinson-Herbots H. Inference of Gene Flow in the Process of Speciation: An Efficient Maximum-Likelihood Method for the Isolation-with-Initial-Migration Model. *Genetics*. 2017; 205: 1597–1618. <https://doi.org/10.1534/genetics.116.188060> PMID: [28193727](https://pubmed.ncbi.nlm.nih.gov/28193727/)
50. Lohse K, Clarke M, Ritchie MG, Etges WJ. Genome-wide tests for introgression between cactophilic *Drosophila* implicate a role of inversions during speciation. *Evolution*. 2015; 69: 1178–1190. <https://doi.org/10.1111/evo.12650> PMID: [25824653](https://pubmed.ncbi.nlm.nih.gov/25824653/)
51. Dobzhansky T. Chromosomal races in *Drosophila pseudoobscura* and *Drosophila persimilis*. Carnegie Inst.: Washington Publ.; 1944.
52. Fontaine MC, Pease JB, Steele A, Waterhouse RM, Neafsey DE, Sharakhov IV, et al. Extensive introgression in a malaria vector species complex revealed by phylogenomics. *Science*. 2015; 347: 1258524. <https://doi.org/10.1126/science.1258524> PMID: [25431491](https://pubmed.ncbi.nlm.nih.gov/25431491/)
53. Kelemen RK, Vicoso B. Complex History and Differentiation Patterns of the t-Haplotype, a Mouse Meiotic Driver. *Genetics*. 2017; genetics.300513.2017. <https://doi.org/10.1534/genetics.117.300513> PMID: [29138255](https://pubmed.ncbi.nlm.nih.gov/29138255/)
54. Fuller ZL, Haynes GD, Richards S, Schaeffer SW. Genomics of Natural Populations: How Differentially Expressed Genes Shape the Evolution of Chromosomal Inversions in *Drosophila pseudoobscura*. *Genetics*. 2016; genetics.116.191429. <https://doi.org/10.1534/genetics.116.191429> PMID: [27401754](https://pubmed.ncbi.nlm.nih.gov/27401754/)
55. Fuller ZL, Haynes GD, Zhu D, Batterton M, Chao H, Dugan S, et al. Evidence for Stabilizing Selection on Codon Usage in Chromosomal Rearrangements of *Drosophila pseudoobscura*. *G3*. 2014; g3.114.014860. <https://doi.org/10.1534/g3.114.014860> PMID: [25326424](https://pubmed.ncbi.nlm.nih.gov/25326424/)
56. Schaeffer SW, Goetting-Minesky MP, Kovacevic M, Peoples JR, Graybill JL, Miller JM, et al. Evolutionary genomics of inversions in *Drosophila pseudoobscura*: Evidence for epistasis. *Proc Natl Acad Sci U S A*. 2003; 100: 8319–8324. <https://doi.org/10.1073/pnas.1432900100> PMID: [12824467](https://pubmed.ncbi.nlm.nih.gov/12824467/)
57. Fuller ZL, Haynes GD, Richards S, Schaeffer SW. Genomics of Natural Populations: Evolutionary Forces that Establish and Maintain Gene Arrangements in *Drosophila pseudoobscura*. *Mol Ecol*. 2017; (20)5362–5368. <https://doi.org/10.1111/mec.14290>
58. Guerrero RF, Hahn MW. Speciation as a Sieve for Ancestral Polymorphism. *Mol Ecol*.: n/a–n/a. <https://doi.org/10.1111/mec.14290> PMID: [28792649](https://pubmed.ncbi.nlm.nih.gov/28792649/)
59. Muller HJ. Isolating mechanisms, evolution and temperature. *Biological Symposia*. Lancaster, PA; 1942. pp. 71–125.
60. Dobzhansky T. Studies on Hybrid Sterility. II. Localization of Sterility Factors in *Drosophila Pseudoobscura* Hybrids. *Genetics*. 1936; 21: 113–135. PMID: [17246786](https://pubmed.ncbi.nlm.nih.gov/17246786/)
61. Orr HA. The population genetics of speciation: the evolution of hybrid incompatibilities. *Genetics*. 1995; 139: 1805–1813. PMID: [7789779](https://pubmed.ncbi.nlm.nih.gov/7789779/)
62. Presgraves DC. The molecular evolutionary basis of species formation. *Nat Rev Genet*. 2010; 11: 175–180. <https://doi.org/10.1038/nrg2718> PMID: [20051985](https://pubmed.ncbi.nlm.nih.gov/20051985/)
63. Maheshwari S, Barbash DA. The genetics of hybrid incompatibilities. *Annu Rev Genet*. 2011; 45: 331–355. <https://doi.org/10.1146/annurev-genet-110410-132514> PMID: [21910629](https://pubmed.ncbi.nlm.nih.gov/21910629/)
64. Matute DR, Butler IA, Turissini DA, Coyne JA. A test of the snowball theory for the rate of evolution of hybrid incompatibilities. *Science*. 2010; 329: 1518–1521. <https://doi.org/10.1126/science.1193440> PMID: [20847270](https://pubmed.ncbi.nlm.nih.gov/20847270/)
65. Moyle LC, Nakazato T. Hybrid incompatibility “snowballs” between *Solanum* species. *Science*. 2010; 329: 1521–1523. <https://doi.org/10.1126/science.1193063> PMID: [20847271](https://pubmed.ncbi.nlm.nih.gov/20847271/)
66. Wang RJ, White MA, Payseur BA. The Pace of Hybrid Incompatibility Evolution in House Mice. *Genetics*. 2015; 201: 229–242. <https://doi.org/10.1534/genetics.115.179499> PMID: [26199234](https://pubmed.ncbi.nlm.nih.gov/26199234/)
67. Roux C, Fraïsse C, Romiguier J, Anciaux Y, Galtier N, Bierne N. Shedding Light on the Grey Zone of Speciation along a Continuum of Genomic Divergence. *PLoS Biol*. 2016; 14: e2000234. <https://doi.org/10.1371/journal.pbio.2000234> PMID: [28027292](https://pubmed.ncbi.nlm.nih.gov/28027292/)
68. Hoffmann AA, Rieseberg LH. Revisiting the Impact of Inversions in Evolution: From Population Genetic Markers to Drivers of Adaptive Shifts and Speciation? *Annual Review of Ecology, Evolution, and Systematics*. 2008; 39: 21–42. <https://doi.org/10.1146/annurev.ecolsys.39.110707.173532> PMID: [20419035](https://pubmed.ncbi.nlm.nih.gov/20419035/)
69. Fishman L, Stathos A, Beardsley PM, Williams CF, Hill JP. Chromosomal rearrangements and the genetics of reproductive barriers in *mimulus* (monkey flowers). *Evolution*. 2013; 67: 2547–2560. <https://doi.org/10.1111/evo.12154> PMID: [24033166](https://pubmed.ncbi.nlm.nih.gov/24033166/)

70. Lowry DB, Willis JH. A Widespread Chromosomal Inversion Polymorphism Contributes to a Major Life-History Transition, Local Adaptation, and Reproductive Isolation. *PLOS Biology*. 2010; 8: e1000500. <https://doi.org/10.1371/journal.pbio.1000500> PMID: 20927411
71. Castiglia R. Sympatric sister species in rodents are more chromosomally differentiated than allopatric ones: implications for the role of chromosomal rearrangements in speciation. *Mammal Review*. 2014; 44: 1–4. <https://doi.org/10.1111/mam.12009>
72. Davey JW, Barker SL, Rastas PM, Pinharanda A, Martin SH, Durbin R, et al. No evidence for maintenance of a sympatric *Heliconius* species barrier by chromosomal inversions. *Evolution Letters*. 2017; 1: 138–154. <https://doi.org/10.1002/evl3.12>
73. Pinho C, Hey J. Divergence with Gene Flow: Models and Data. *Annual Review of Ecology, Evolution, and Systematics*. 2010; 41: 215–230. <https://doi.org/10.1146/annurev-ecolsys-102209-144644>
74. Orr HA. Genetics of Male and Female Sterility in Hybrids of *Drosophila pseudoobscura* and *D. persimilis*. *Genetics*. 1987; 116: 555–563. PMID: 3623079
75. Noor MA, Grams KL, Bertucci LA, Almendarez Y, Reiland J, Smith KR. The genetics of reproductive isolation and the potential for gene exchange between *Drosophila pseudoobscura* and *D. persimilis* via backcross hybrid males. *Evolution*. 2001; 55: 512–521. PMID: 11327159
76. Painter TS. A New Method for the Study of Chromosome Aberrations and the Plotting of Chromosome Maps in *Drosophila Melanogaster*. *Genetics*. 1934; 19: 175–188. PMID: 17246718
77. Harshman LG. A technique for the preparation of *Drosophila* salivary gland chromosomes. *Drosophila Information Service*. 1977;52.
78. Ballard JWO, Bedo DG. Population cytogenetics of *Austrosimulium bancrofti* (Diptera: Simuliidae) in eastern Australia. *Genome*. 1991; 34: 338–353. <https://doi.org/10.1139/g91-056>
79. Bolger AM, Lohse M, Usadel B. Trimmomatic: A flexible trimmer for Illumina Sequence Data. *Bioinformatics*. 2014; btu170. <https://doi.org/10.1093/bioinformatics/btu170> PMID: 24695404
80. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*. 2009; 25: 1754–1760. <https://doi.org/10.1093/bioinformatics/btp324> PMID: 19451168
81. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics*. 2009; 25: 2078–2079. <https://doi.org/10.1093/bioinformatics/btp352> PMID: 19505943
82. Garrison E, Marth G. Haplotype-based variant detection from short-read sequencing. arXiv:12073907 [q-bio]. 2012; Available: <http://arxiv.org/abs/1207.3907>
83. Schaeffer SW, Bhutkar A, McAllister BF, Matsuda M, Matzkin LM, O'Grady PM, et al. Polytene Chromosomal Maps of 11 *Drosophila* Species: The Order of Genomic Scaffolds Inferred From Genetic and Physical Maps. *Genetics*. 2008; 179: 1601–1655. <https://doi.org/10.1534/genetics.107.086074> PMID: 18622037
84. Nei M, Tajima F, Tateno Y. Accuracy of estimated phylogenetic trees from molecular data. II. Gene frequency data. *J Mol Evol*. 1983; 19: 153–170. PMID: 6571220
85. Kalinowski ST. Evolutionary and statistical properties of three genetic distances. *Mol Ecol*. 2002; 11: 1263–1273. PMID: 12144649
86. Saitou N, Nei M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol*. 1987; 4: 406–425. <https://doi.org/10.1093/oxfordjournals.molbev.a040454> PMID: 3447015
87. Nei M. *Molecular Evolutionary Genetics*. Columbia University Press; 1987.
88. Nei M, Li WH. Mathematical model for studying genetic variation in terms of restriction endonucleases. *PNAS*. 1979; 76: 5269–5273. PMID: 291943
89. Cavalli-Sforza LL, Cavalli-Sforza L, Menozzi P, Piazza A. *The History and Geography of Human Genes*. Princeton University Press; 1994.
90. Yi X, Liang Y, Huerta-Sanchez E, Jin X, Cuo ZXP, Pool JE, et al. Sequencing of 50 human exomes reveals adaptation to high altitude. *Science*. 2010; 329: 75–78. <https://doi.org/10.1126/science.1190371> PMID: 20595611
91. Holsinger KE, Weir BS. Genetics in geographically structured populations: defining, estimating and interpreting F_{ST} . *Nature Reviews Genetics*. 2009; 10: 639–650. <https://doi.org/10.1038/nrg2611> PMID: 19687804
92. Becquet C, Przeworski M. A new approach to estimate parameters of speciation models with application to apes. *Genome Res*. 2007; 17: 1505–1519. <https://doi.org/10.1101/gr.6409707> PMID: 17712021
93. Strasburg JL, Rieseberg LH. How robust are “isolation with migration” analyses to violations of the im model? A simulation study. *Mol Biol Evol*. 2010; 27: 297–310. <https://doi.org/10.1093/molbev/msp233> PMID: 19793831

94. Yang Z. Likelihood and Bayes estimation of ancestral population sizes in hominoids using data from multiple loci. *Genetics*. 2002; 162: 1811–1823. PMID: [12524351](https://pubmed.ncbi.nlm.nih.gov/12524351/)
95. Wang Y, Hey J. Estimating Divergence Parameters With Small Samples From a Large Number of Loci. *Genetics*. 2010; 184: 363–379. <https://doi.org/10.1534/genetics.109.110528> PMID: [19917765](https://pubmed.ncbi.nlm.nih.gov/19917765/)