Inversions shape the divergence of *Drosophila pseudoobscura* and *D. persimilis* on multiple timescales

Katharine L Korunes[1*], Carlos A Machado[2], Mohamed AF Noor[3]

1. Department of Evolutionary Anthropology, Duke University, Durham, NC 27708
2. Department of Biology, University of Maryland, College Park, Maryland 20742
3. Biology Department, Duke University, Durham, NC 27708

* Corresponding author: kkorunes@gmail.com

Abstract

By shaping meiotic recombination, chromosomal inversions can influence genetic exchange between hybridizing species. Despite the recognized importance of inversions in evolutionary processes such as divergence and speciation, teasing apart the effects of inversions over time remains challenging. Many recent studies have examined gene exchange between hybridizing species and observed higher divergence associated with inversion polymorphisms in hybrids. Studies have attributed such patterns either to interspecies gene flow homogenizing collinear regions or to pre-speciation segregation of inversion polymorphisms in the ancestral population. We provide a comprehensive look into the influence of chromosomal inversions on gene flow throughout the evolutionary history of a classic system: *Drosophila pseudoobscura* and *D. persimilis*. We acquired extensive whole-genome sequence data to report patterns of introgression and divergence with respect to chromosomal arrangement. Curiously, divergence measures to *D. persimilis* are higher for allopatric than sympatric *D. pseudoobscura* populations in both inverted and collinear regions, which could indicate recent gene exchange that affects even inverted regions. However, consideration of ancestral inversion polymorphisms and different evolutionary rates in these populations reveals that recent gene exchange is a major but not the sole driver of variation in overall sequence divergence. Overall, we show that inversions have contributed to divergence patterns between *Drosophila pseudoobscura* and *D. persimilis* over three distinct timescales: 1) pre-speciation segregation of ancestral polymorphism, 2) post-speciation ancient gene flow, and 3) recent introgression. We discuss these results in terms of our understanding of evolution in this classic system and provide cautions for interpreting divergence measures in similar datasets in other systems.

Keywords: inversions, introgression, divergence, recombination, speciation

Introduction

Divergence and speciation sometimes occur in the presence of gene exchange between taxa. Estimates suggest that over 10% of animal species hybridize and exchange genes with related species (Mallet 2005). Analyses in the genomic era have provided further evidence of the widespread prevalence of hybridization and revealed many previously unanticipated instances of hybridization (Payseur & Rieseberg 2016; Taylor & Larson 2019). Understanding genetic exchange between species gives us insights into the genetic processes underlying later stages of the speciation continuum. Many approaches can examine evidence for introgression, including comparing sympatric vs. allopatric populations to test for differences in nucleotide divergence. Other available methods for characterizing gene flow include model-based frameworks and examinations of differences in divergence reflected in coalescence. Differences in coalescence are often observed between species in regions where recombination is limited in hybrids, such as fixed chromosomal inversion differences (Guerrero *et al.* 2012). When species differing by inversions hybridize, the collinear genomic regions can freely recombine, while inverted regions experience severely limited genetic exchange in hybrids and often accumulate greater sequence differentiation over generations. This process can lead to locally adapted traits and reproductive isolating barriers mapping disproportionately to inverted regions (reviewed in Ayala & Coluzzi 2005; Butlin 2005; Jackson 2011).

Many studies examine the timing and frequency of gene exchange between hybridizing species, with emphasis on the implications of patterns of divergence in allopatric vs sympatric pairs and in regions of reduced recombination in hybrids. However, different approaches sometimes yield distinct interpretations regarding the presence or extent of introgression. Model-based approaches yield important insights but are also limited in the scenarios that they consider and the assumptions they make about population histories and evolutionary rates (reviewed in Payseur & Rieseberg 2016). Further, shared patterns of variation are often interpreted as evidence of ongoing gene flow, but segregating ancestral polymorphism could also be primary, or even the sole, driver of these patterns (Fuller *et al.* 2018). Fuller *et al.* (2018) recently discussed the possibility that ancestrally segregating inversions that sort between species may provide a "head-start" in molecular divergence, possibly predisposing them to harbor a disproportionate fraction of alleles associated with species differences. Unlike models assuming homogenization of collinear regions via post-speciation gene flow, this model predicts that young species that diverged in allopatry may also exhibit higher divergence in inverted regions than collinear regions. These models are not mutually exclusive: dynamics of the ancestral population as well as post-speciation gene flow can shape patterns of variation between species.

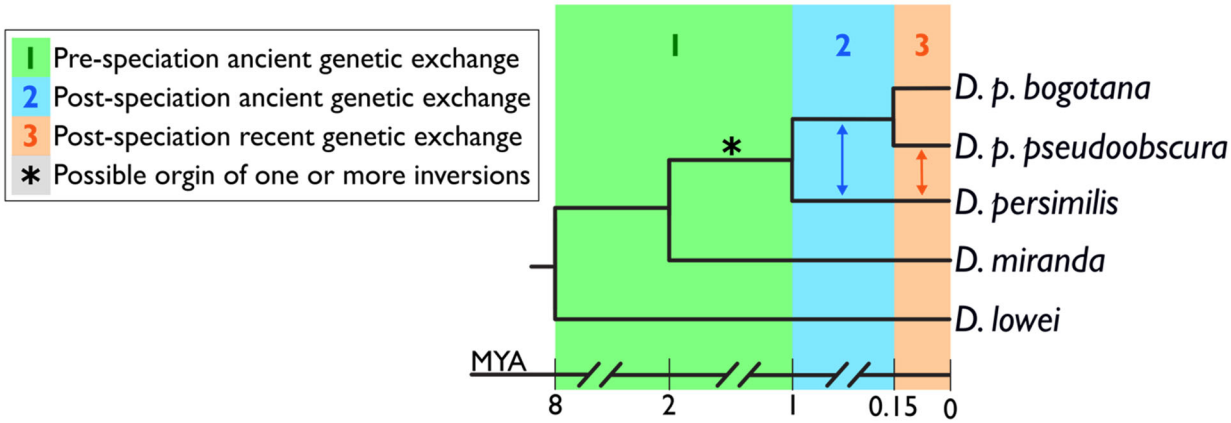
Disentangling the effects of ancestral polymorphism from the effects of post-speciation gene flow is a fundamental puzzle in understanding speciation. To achieve a cohesive picture of how hybridization influences divergence and speciation, we need to consider the approaches outlined above in a model system with extensive whole-genome sequence data to assess models and reconcile interpretations of possible signals of introgression. The sister species pair *Drosophila pseudoobscura* and *D. persimilis* present an ideal opportunity to dissect an evolutionary history of divergence nuanced by multiple inversions, lineage sorting, and gene flow. Despite the rich history of work on understanding speciation and divergence in *D. pseudoobscura* and *D. persimilis*, there are unresolved questions about the rates and timing of introgression between these species. A

few F_1 hybrids of these species have been collected in the wild (Powell 1983) and many previous studies have documented molecular evidence of introgression, detectable in both nuclear and mitochondrial loci (e.g., Machado *et al.* 2002; Machado & Hey 2003; Hey & Nielsen 2004; Fuller *et al.* 2018). Inverted regions between these species exhibit greater sequence differences than collinear regions, and this pattern was previously inferred to result from introgression post-speciation. In the largest scale study, McGaugh and Noor (2012) used multiple genome sequences of both species and an outgroup, and reinforced previous studies (e.g., Noor *et al.* 2007) showing that the three chromosomal inversions differ in divergence time. They inferred a "mixed mode geographic model" (Feder *et al.* 2011) with sporadic periods of introgression during and after the times that the inversions spread. However, in addition to confirming evidence for gene flow between *D. pseudoobscura* and *D. persimilis* after speciation, Fuller *et al.* (2018) recently argued the inversions arose within a single ancestor species, differentially sorted into one of the descendant species (*D. persimilis*), and this sorting of ancestral polymorphisms may explain observed patterns of nucleotide variation. To fully understand the role of hybridization in the speciation process, the contrasting models must be reconciled.

We acquired extensive whole-genome sequence data to re-explore patterns of introgression and divergence in the *Drosophila pseudoobscura* / *D. persimilis* system. We leverage the allopatric *D. pseudoobscura* subspecies, *D. pseudoobscura bogotana* (*D. p. bogotana)* and two outgroup species (*D. miranda* and *D. lowei*) to distinguish recent from ancient effects of inversions on gene flow. To help clarify the role of inversions over the evolutionary history of these species, we consider three distinct time scales: 1) pre-speciation segregation of ancestral polymorphism, 2) post-speciation ancient gene flow, and 3) recent introgression (Figure 1). Patterns of divergence between *D. persimilis* and allopatric *D. p. bogotana* can be explained by the effects of segregating ancestral polymorphism and by gene flow prior to the split of *D. p. bogotana* (Figure 1, green and blue regions). In comparing the sympatric species, *D. persimilis* and *D. p. pseudoobscura,* the same forces factor into patterns of divergence, with the added effects of recent or ongoing gene flow (Figure 1, orange arrows). We leverage these two comparisons to weigh the relative contributions of recent genetic exchange.

We first examine patterns of divergence in inverted regions compared to collinear regions, and we discuss evidence for early, pre-speciation exchange. We next examine evidence of post-speciation gene flow to test whether some of this genetic exchange predates the split of *D. p. bogotana,* and we discuss signals of possible introgression in the past 150,000 years since the split of the allopatric *D. p. bogotana,* from North American *D. pseudoobscura* (*D. p. pseudoobscura*) (Schaeffer & Miller 1991). We implement Patterson's D-statistic to contrast the sympatric (*D. p. pseudoobscura*) and allopatric (*D. p. bogotana)* subspecies in their similarity to *D. persimilis*. Kulathinal *et al.* (2009) previously argued that recent post-speciation gene flow contributes to the difference in coalescence time between inverted and collinear regions, observable in the higher genetic similarity in collinear regions between *D. persimilis* and sympatric *D. p. pseudoobscura* compared to similarity between *D. persimilis* and allopatric *D. p. bogotana*. That study also tested for an excess of shared, derived bases between *D. persimilis* and *D. p. pseudoobscura* compared to *D. persimilis* and allopatric *D. p. bogotana*. Their application of this D-statistic precursor suggested a borderline statistically significant signature of gene flow, but this test was limited by low sequencing coverage. Nonetheless, consistent with that and other previous studies, we observe higher divergence between these species in inverted than collinear regions, and our implementation of Patterson's D-statistic indicates very recent gene exchange in collinear regions. Notably, divergence measures

104 to *D. persimilis* are also higher for allopatric than sympatric *D. pseudoobscura* populations in *both* inverted and
105 collinear regions. One possible explanation for this pattern is extensive recent gene exchange throughout
106 much of the genome, even in inverted regions. To determine the relative effect of introgression, we correct
107 for different evolutionary rates across taxa, and we consider the role of segregating inversions in ancestral
108 populations. We discuss these results in the context of the extensive past work towards understanding
109 divergence and speciation in this classic system, and we provide cautions for interpreting divergence measures
110 in similar datasets in other systems.



111

112 **Figure 1** | Gene flow in the context of the evolutionary history of *D. pseudoobscura* and *D. persimilis*. We
113 consider how inversions differing between *D. pseudoobscura* and *D. persimilis* might shape patterns of
114 divergence by affecting gene flow at 3 timescales: 1) pre-speciation recombination in ancestral populations
115 with segregating inversion polymorphisms, 2) post-speciation ancient gene flow, and 3) recent introgression.
116 Here, we show the evolutionary context and approximate divergence times of the taxa considered in the
117 present study, with arrows indicating post-speciation gene flow between *D. pseudoobscura* and *D. persimilis*.

118 ## *Methods*

119 ### Genomic datasets

120 Whole-genome sequence data was analyzed from 19 *D. p. pseudoobscura* and 8 *D. persimilis* individuals, and we
121 used 4 *D. p. bogotana* genomes for an allopatric point of comparison (sequence data is described in
122 Supplementary Table 1). We used *D. lowei* as an outgroup. *D. lowei* and *D. pseudoobscura* likely diverged 5-11
123 MYA (Beckenbach *et al.* 1993), and hybrids between these two species are sterile (Heed *et al.* 1969). Scripts
124 used for genome alignment, SNP calling, and analyses are available on GitHub
125 (https://github.com/kkorunes/Dpseudoobscura_Introgression). To avoid biasing identification of variants
126 towards *D. pseudoobscura*, the *D. miranda* reference genome assembly (DroMir2.2; GenBank assembly accession
127 GCA_000269505.2) was chosen as the reference for all subsequent alignments. *D. miranda* diverged from *D.*
128 *pseudoobscura* only within the past ~2 million years (Wang & Hey 1996), facilitating alignment of *D.*
129 *pseudoobscura* and *D. persimilis* genomes to the *D. miranda* genome assembly. Further, the arrangement of the
130 assembled *D. miranda* chromosomes matches the arrangement of the *D. pseudoobscura* contig order and

131 arrangement described by Schaeffer *et al.* (2008), but with the advantage of being assembled into 6 continuous
132 chromosome arms: chromosomes XL, XR, 2, 3, 4, and 5. For all present analyses, we focus on chromosomes
133 XL, XR, 2, and 4, which together make up 83% of the assembled genome. We excluded chromosome 3 from
134 these analyses because of confounding factors from its inversion polymorphisms within species (Dobzhansky
135 & Epling 1944; Powell 1992), and we excluded the <2 Mb, largely nonrecombining "dot" chromosome
136 (chromosome 5).

137 **Alignments and variant calling**

138 To confirm the arrangement of *D. pseudoobscura* contigs with respect to the *D. miranda* reference, each *D.*
139 *pseudoobscura* chromosome was split into lengths of 1 Mb, and these segments were aligned to the *D. miranda*
140 reference using BWA-0.7.5a (Li & Durbin 2009). We then extracted the 2 kb regions surrounding published
141 inversion breakpoints to obtain the breakpoint locations in the coordinates of the *D. miranda* reference (see
142 Supplementary Table 2). After confirming that the arrangement of the assembled *D. miranda* chromosomes
143 matched the arrangement of the *D. pseudoobscura* contig order and arrangement described by Schaeffer *et al.*
144 (2008), all sequencing data were aligned to the reference genome of *D. miranda* using BWA-0.7.5a (Li &
145 Durbin 2009). Variants were called used GATK v4 (McKenna *et al.* 2010; Van der Auwera *et al.* 2013) after
146 using Picard to mark adapters and duplicates (http://broadinstitute.github.io/picard). SNPs were filtered
147 based on quality, excluding sites with QualByDepth (QD) < 2.0, FisherStrand (FS) > 60, StrandOddsRatio
148 (SOR) > 3.0, MQ < 40, MQRankSum < -12.5, ReadPosRankSum < -8.

149 **Patterns of divergence**

150 The resulting VCF files were then processed using PLINK (Purcell *et al.* 2007). VCFs were converted to
151 PLINK's bed/bim format, keeping only sites that passed the filters described above. SNPs were pruned for
152 linkage disequilibrium using the --indep-pairwise function of PLINK ("--indep-pairwise 50 50 0.5") before
153 performing principal components analysis (PCA) using PLINK's --pca function to confirm the grouping of
154 individuals within their respective species (Supplementary Figure 1). Admixtools was used to implement D-
155 statistics (Patterson *et al.* 2012) using *D. lowei* as on outgroup to polarize ancestral vs derived alleles. For input
156 into Admixtools, we used the *convertf* program of Admixtools to convert each PLINK ped file to Eigenstrat
157 format which includes a *genotype* file, a *snp* file, and an *indiv* file. Per recommendations in the Admixtools
158 documentation, we defined the physical positions of each SNP in the *snp* file to be 10 kb apart from each
159 adjacent SNP to allow Admixtools to interpret every 100 SNPs as 1 Mb or 1cM, since this software uses
160 centiMorgans as the unit for block size during jackknifing and assumes that 1 Mb = 1 cM. We set the block
161 size parameter to 0.01 cM, which in this case is interpreted as blocks of 100 SNPs. Next, *qpDstat* was used to
162 obtain D-statistics for each chromosome. These four-population tests were of the form (((A,B), C), D), where
163 A = *D. p. bogotana,* B = *D. p. pseudoobscura,* C = *D. persimilis,* and D = *D. lowei.* Absolute divergence, D_{xy}, was
164 calculated using custom scripts over fixed window sizes of 500 kb. D_{xy} was calculated from variant and
165 invariant sites after subjecting SNPs to the filters described above and filtering invariant sites based on depth
166 (depth >= 10). Per-site depths for all sites were acquired from BAM files using Samtools ("samtools depths -
167 a <in>") (Li *et al.* 2009).
168

Models of gene flow

To test for evidence of gene flow after the split of *D. pseudoobscura* and *D. persimilis,* but before the split of *D. p. bogotana,* we used the maximum-likelihood methods derived by Costa & Wilkinson-Herbots (2017) to compare three scenarios of divergence between *D. persimilis* and *D. p. bogotana* (Figure 2): (A) divergence in isolation (Iso) without gene flow following the split of an ancestral population; (B) divergence in isolation-with-migration (IM) with constant (but potentially asymmetric) gene flow since the split of an ancestral population until the present; and (C) divergence in isolation-with-initial-migration (IIM) with gene flow until some timepoint in the past and divergence in isolation since that timepoint. Under the IIM model, we tested four scenarios: IIM1 estimates parameters under divergence with potentially asymmetric bidirectional gene flow until some timepoint in the past, assuming constant population sizes; IIM2 is the same scenario as IIM1, but allows for changes in population sizes; IIM3 and IIM4 are the same as IIM2, but assume unidirectional gene flow from population 2 to population 1 (IIM3) or from population 1 to population 2 (IIM4).

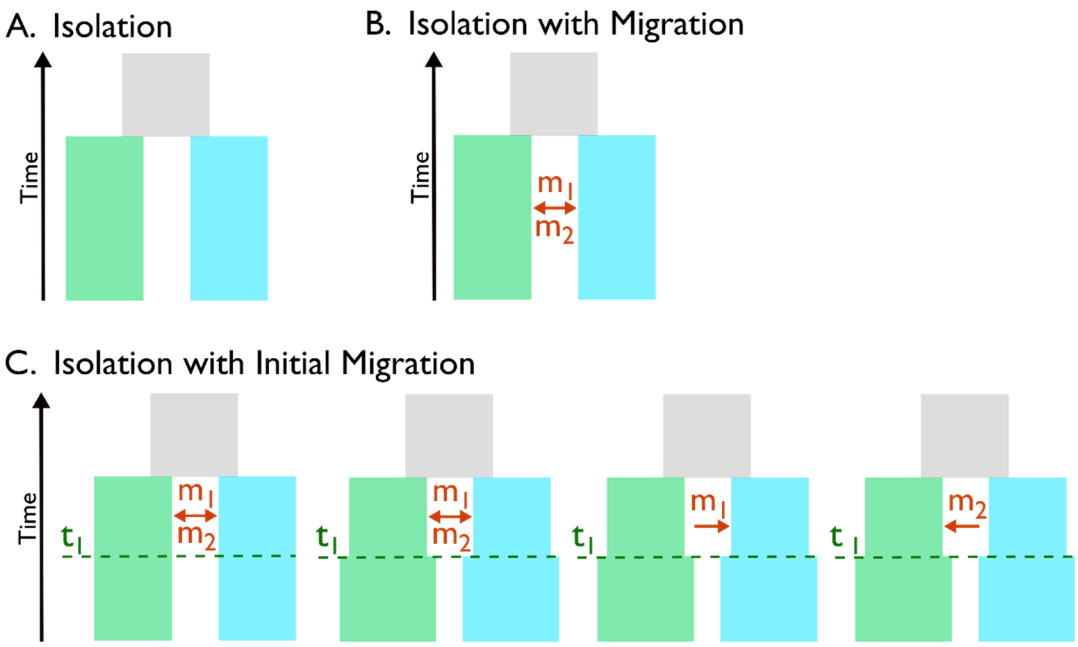


Figure 2 | Models of Divergence. We considered the following coalescent models described by Costa & Wilkinson-Herbots (2017) to consider scenarios of divergence of *D. persimilis* and *D. p. bogotana* since the split of the ancestral population (gray box): (A) divergence in isolation without gene flow; (B) divergence in isolation-with-migration (IM) with constant (but potentially asymmetric) gene flow; and (C) divergence in isolation-with-initial-migration (IIM) with gene flow until some timepoint (t_1) in the past. Under the IIM model, we tested the four scenarios shown from left to right: the first scenario assumes constant population sizes, the second allows for changes in population sizes, and the third and fourth allow for changes in population size but assume unidirectional gene flow.

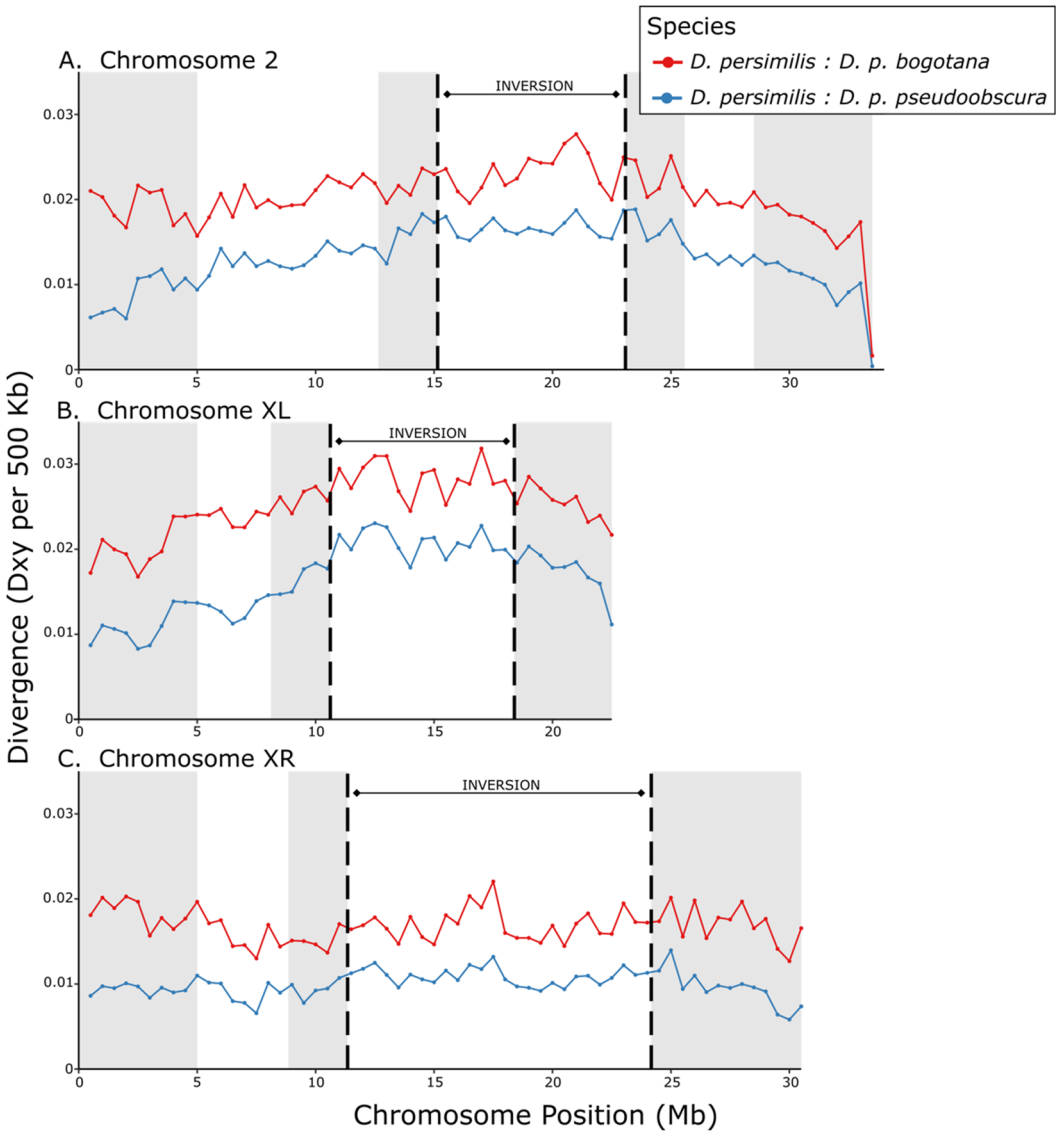
190 We computed the likelihood of our *D. persimilis* and *D. p. bogotana* sequence data under each of the six
191 scenarios described above and in Figure 2 (Iso, IM, IIM1, IIM2, IIM3, and IIM4). To reduce potential effects
192 of selection, we used intergenic loci spaced at least 2 kb apart, similar to the strategy of Wang & Hey (2010).
193 The spacing between loci accounts for the models' assumption of free recombination between loci. To
194 identify intergenic regions in the *D. miranda* genome, we used the set of all *D. pseudoobscura* gene annotations
195 published by Flybase (http://flybase.org, Full Annotation Release 3.04), and we used BLAST to identify
196 genomic regions with significant similarly to the *D. pseudoobscura* gene annotations, using cutoffs of evalue =
197 1e-6 and percent identity = 80 (Altschul *et al.* 1990). From the remaining regions, we then randomly sampled
198 500 bp segments separated by at least 2 kb to create a set of ~15,000 intergenic loci. We then randomly
199 divided these loci into three nonoverlapping subsets to satisfy the models' requirement of independent
200 estimates of pairwise differences and mutation rates in loci (1) within *D. persimilis*, (2) within *D. p. bogotana*, and
201 (3) between *D. persimilis* and *D. p. bogotana*. Costa & Wilkinson-Herbots (2017) recommends using per-locus
202 relative mutation rates, which we calculated using the average distance to the outgroup *D. lowei*, following the
203 equation from Yang (2002), which gives the relative mutation rate at a locus as the outgroup distance at that
204 locus divided by the average outgroup distance along all loci. To select the model that best fits the data, we
205 then tested the relative support among the divergence models using likelihood-ratio tests following the
206 sequence of pairwise comparisons shown in Table 1, where the degrees of freedom in each test is the
207 difference in the dimensions of parameter space (Costa & Wilkinson-Herbots 2017). We repeated these
208 analyses using only collinear loci (~11,000 intergenic loci) to ensure that our results were robust against the
209 effects of linkage within inverted regions.
210

211 ## *Results*

212 ### Patterns of divergence in inverted vs collinear regions

213 The suppression of crossing over within inversions leads to distinct signatures of nucleotide divergence
214 within and near inversions. Figure 3 presents windowed (500 kb windows) estimates of divergence between
215 *D. persimilis* and *D. p. bogotana* and divergence between *D. persimilis* vs *D. p. pseudoobscura* for the three
216 chromosome arms that contain fixed (chromosome 2, XL) or nearly-fixed (chromosome XR) inversion
217 differences between *D. persimilis* and *D. pseudoobscura*. As previously observed, divergence is low in regions
218 near the centromere (Noor *et al.* 2007; Kulathinal *et al.* 2009). As per Stevison & Noor (2010), we excluded
219 regions within 5 mb of chromosome ends from our "collinear regions" to obtain divergence estimates not
220 affected by chromosome-end effects (Stevison & Noor 2010). Similarly, we excluded collinear regions within
221 2.5 mb of inversion breakpoints, based on previous studies which have reported suppression of crossing over
222 extending 1-2 mb beyond inversion breakpoints (Machado *et al.* 2007; Kulathinal *et al.* 2009; Stevison *et al.*
223 2011).
224
225 Confirming many previous studies (Machado *et al.* 2007; Noor *et al.* 2007; Kulathinal *et al.* 2009; Stevison *et al.*
226 2011; McGaugh & Noor 2012), we observed that *D. persimilis* vs *D. p. pseudoobscura* divergence is higher in
227 inverted than collinear windows ($p < 0.01$ on each chromosome, Mann Whitney U test). *D. persimilis* vs
228 allopatric *D. p. bogotana* divergence is also higher in inverted than collinear windows ($p < 0.01$, Mann-Whitney

229 U test) on chromosomes 2 and XL, with a nonsignificant difference on chromosome XR (p = 0.138, Mann-
230 Whitney U test). The higher divergence in inverted vs collinear regions could be due to pre-speciation
231 segregation of inversion polymorphisms in the ancestral population or to interspecies gene flow
232 homogenizing collinear regions. We next examined evidence of pre-speciation genetic exchange.



233

234 **Figure 3** | Genome-wide divergence between species. Each of the 3 inversion-bearing chromosome arms
235 are plotted from centromere (0) to telomere, inversion boundaries are shown with vertical black lines. D_{xy} per
236 500 kb window is plotted to show divergence between *D. persimilis* and *D. p. bogotana* (red) and divergence
237 between *D. persimilis* vs *D. p. pseudoobscura* (blue). The grayed-out portions indicate the regions that were
238 excluded from estimates of collinear D_{xy} due to their proximity to the inversion breakpoints or to
239 chromosome ends.

Confirming evidence for early, pre-speciation exchange

Previous studies identified differences in sequence divergence between *D. persimilis* and *D. p. pseudoobscura* among the three inverted regions (Noor *et al.* 2007; McGaugh & Noor 2012; Fuller *et al.* 2018). They interpreted this finding as evidence that the derived inversions arose during, or interspersed by, periods where gene exchange was occurring in other regions of the genome. For both *D. persimilis* vs *D. p. pseudoobscura* and *D. persimilis* vs *D. p. bogotana*, we compared measures of windowed divergence among the inverted regions. Each pairwise comparison between the inversions yielded a significant difference wherein XL > 2 > XR ($p < 0.0001$, Supplementary Table 3). Our observed average divergence within the inversions confirms previous accounts of the relative divergence and apparent relative age of the inversions (Noor *et al.* 2007; McGaugh & Noor 2012; Fuller *et al.* 2018), supporting the previous suggestions of ancestral exchange prior to the completion of speciation. We next explored the possibility of exchange after speciation but before the split of *D. p. pseudoobscura* and *D. p. bogotana* (region 2 in Figure 1).

Evidence for early post-speciation exchange

Much of the previous support for post-speciation gene flow between these species has focused on comparisons of *D. persimilis* and *D. pseudoobscura* (Wang *et al.* 1997; Machado *et al.* 2002; Hey & Nielsen 2004; Kulathinal *et al.* 2009; Fuller *et al.* 2018). To distinguish between post-speciation ancient gene flow and recent introgression (Figure 1), we leveraged the allopatric *D. p. bogotana*. Some past studies suggesting evidence of introgression between *D. pseudoobscura* and *D. persimilis* also leveraged this allopatric subspecies (Machado & Hey 2003; Brown *et al.* 2004; Chang & Noor 2007; Kulathinal *et al.* 2009). We sampled intergenic loci from *D. persimilis* and *D. p. bogotana,* and we fit the observed patterns of nucleotide variation to models of divergence in isolation, isolation-with-migration (IM), and isolation-with-initial-migration (IIM) using maximum-likelihood estimation of parameters under these models (Figure 2; Costa & Wilkinson-Herbots 2017). In traditional IM models applied to infer gene flow, parameter estimates can be skewed by the underlying assumption that gene flow is constant. IIM specifically addresses this assumption by operating on the assumption of an initial period of gene flow followed by isolation. Maximizing parameter likelihoods under an IIM framework is appropriate for the *D. persimilis* and *D. p. bogotana* comparison, given our knowledge that these taxa have been evolving in allopatry for the past 150,000 years. Indeed, nested model comparison to test the relative support among the models rejects the null hypothesis of divergence in isolation and suggest that IIM models best fit the data. These results were consistent between the full set of loci and the subset sampled from only collinear regions. All models allowing for migration gave a significantly better fit than a model of strict divergence in isolation, and the log-likelihood of the data under the tested models was maximized in the IIM2 scenario (Table 1 and Supplementary Table 4). The IIM2 model estimates parameters under divergence with potentially asymmetric bidirectional gene flow until some timepoint in the past and, unlike the IIM1 model, does not assume constant population sizes (Figure 2). We also considered models similar to IIM2, but assuming unidirectional gene flow from *D. p. bogotana* to *D. persimilis* (IIM3) or from *D. persimilis* to *D. p. bogotana* (IIM4). Nested model comparison supports the choice of any of the three models with varying population sizes (IIM2, IIM3, or IIM4) over IIM1, and the likelihood of IIM2 supports bidirectional gene flow (Table 1). These results provide additional support for gene flow between the *D. persimilis* and *D.*

278 *pseudoobscura* lineages after speciation and demonstrate that a significant amount of this exchange was
279 relatively ancient (region 2 in Figure 1), occurring prior to the split of *D. p. bogotana*.

280 **Table 1** | Forward selection of the best model of *D. persimilis - D. p. bogotana* divergence using the maximized
281 log-likelihood (LogL) under each model in likelihood-ratio tests. See Figure 2 for illustration of the different
282 models.

H_0	H_1	Deg. of Freedom	LogL H_0	LogL H_1	LRT Statistic	P-value
Loci sampled genome-wide						
Iso	IM	2	-33702.73	-33659.39	86.68	1.505e-19
IM	IIM1	1	-33659.39	-33659.39	0	-
IM	IIM2	3	-33659.39	-33538.92	240.94	5.960e-52
IIM1	IIM2	2	-33659.39	-33538.92	240.94	4.792e-53
IIM1	IIM3	1	-33659.39	-33543.76	231.26	3.166e-52
IIM1	IIM4	1	-33659.39	-33542.89	233	1.322e-52
Loci sampled from collinear regions only						
Iso	IM	2	-25511.23	-25492.84	36.78	1.031e-08
IM	IIM1	1	-25492.84	-25492.84	0	-
IM	IIM2	3	-25492.84	-25442.42	100.84	1.025e-21
IIM1	IIM2	2	-25492.84	-25442.42	100.84	1.267e-22
IIM1	IIM3	1	-25492.84	-25447.73	90.22	2.131e-21
IIM1	IIM4	1	-25492.84	-25444.03	97.62	5.069e-23

283

284 D-statistics suggest recent genetic exchange

285 Given the evidence for gene flow between *D. persimilis* and *D. pseudoobscura*, we next examined the timing of
286 this gene flow. To contrast sympatric and allopatric subspecies of *D. pseudoobscura* in their similarity to *D.*
287 *persimilis*, we implemented Patterson's D-statistic using the tree: (((*D. p. bogotana*, *D. p. pseudoobscura), D.*
288 *persimilis), D. lowei))).* Patterson's D-statistic is an implementation of ABBA-BABA, which uses parsimony
289 informative sites to test whether derived alleles ("B") in *D. persimilis* are shared with *D. p. bogotana* or with *D. p.*
290 *pseudoobscura* at equal frequencies. Derived alleles in *D. persimilis* may be shared with *D. p. pseudoobscura* due to
291 ancestral polymorphism, ancient gene flow (prior to the split of the two *D. pseudoobscura* subspecies), recent
292 gene flow (since the split of the two *D. pseudoobscura* subspecies), or a combination of these factors. The null
293 expectation is that the two phylogeny-discordant patterns, ABBA and BABA, should be present equally if
294 ancestral polymorphism and ancient gene flow are the sole drivers of patterns of divergence. Gene flow
295 between *D. p. pseudoobscura* and *D. persimilis* since the split of the two *D. pseudoobscura* subspecies (estimated at
296 150,000 years ago: Schaeffer & Miller 1991) would promote an excess of ABBA over BABA patterns,
297 particularly on freely recombining chromosomes. Indeed, ABBA sites exceed BABA sites on all
298 chromosomes (Table 2), and chromosome 4 shows an unambiguously significant excess of ABBA (|Z-score|
299 >= 5), suggesting that the phylogenetic relationship between these four taxa does not fully explain the
300 observed patterns of divergence and some very recent gene exchange has occurred between the North

301 American species. Furthermore, the genome wide z-score for collinear regions is significant (|Z-score| =
302 7.215).

303 **Table 2** | D-statistics. For each examined chromosome[1], the number of sites with BABA and ABBA
304 patterns are provided along with the total number of SNPs at sites where all four populations have data.

Region	D-Statistic	Standard Error	Z-score	BABA Sites	ABBA Sites	SNPs
Whole-Genome						
Inverted Regions	-0.0111	0.0046	-2.401	10,269	10,500	2,063,794
Collinear Regions	-0.0170	0.0024	-7.215**	30,749	31,813	3,750,444
Chromosome 4						
Whole Chromosome	-0.0133	0.0026	-5.071**	22,716	23,328	2,527,672
Chromosome 2						
Whole Chromosome	-0.0151	0.0031	-4.811*	18,769	19,343	2,835,488
Inverted Region	-0.003	0.0082	-0.361	3404	3424	695,888
Collinear Region	-0.0176	0.0048	-3.686*	7,061	7,314	927,665
Chromosome XL						
Whole Chromosome	-0.0160	0.0053	-3.043*	7,937	8,195	1,661,022
Inverted Region	-0.0011	0.0097	-0.109	2,347	2,352	607,297
Collinear Region	-0.0388	0.0125	-3.114*	1513	1635	245,552
Chromosome XR						
Whole Chromosome	-0.0165	0.0044	-3.736*	11,147	11,522	1,979,554
Inverted Region	-0.0170	0.0064	-2.643	5252	5434	893,307
Collinear Region	-0.0224	0.0133	-1.688	1577	1649	265,259

305
306 [1] Note that the collinear regions exclude sequence within 5 mb of chromosome ends or within 2.5 mb of inversion breakpoints.
307 * Significant at |z| >= 3.
308 ** Significant at |z| >= 5.

309

310 Is recent gene flow responsible for patterns of higher divergence in allopatry vs 311 sympatry?

312 All three inversion-bearing chromosomes exhibit lower divergence in the *D. persimilis:D. p. pseudoobscura*
313 comparison vs the *D. persimilis:D. p. bogotana* comparison (Figure 3). Notably, this lower divergence in the
314 sympatric comparison is statistically significant for both the collinear and inverted regions ($p < 0.001$ on each
315 chromosome, Mann-Whitney U test). Average divergence between species pairs in both inverted and collinear
316 regions shows the magnitude of this difference (Figure 4A). Indeed, differentiation in the inverted regions
317 between *D. persimilis:D. p. bogotana* is higher than differentiation between *D. persimilis* and any of the North
318 American *D. p. pseudoobscura* genomes (Figure 4B).

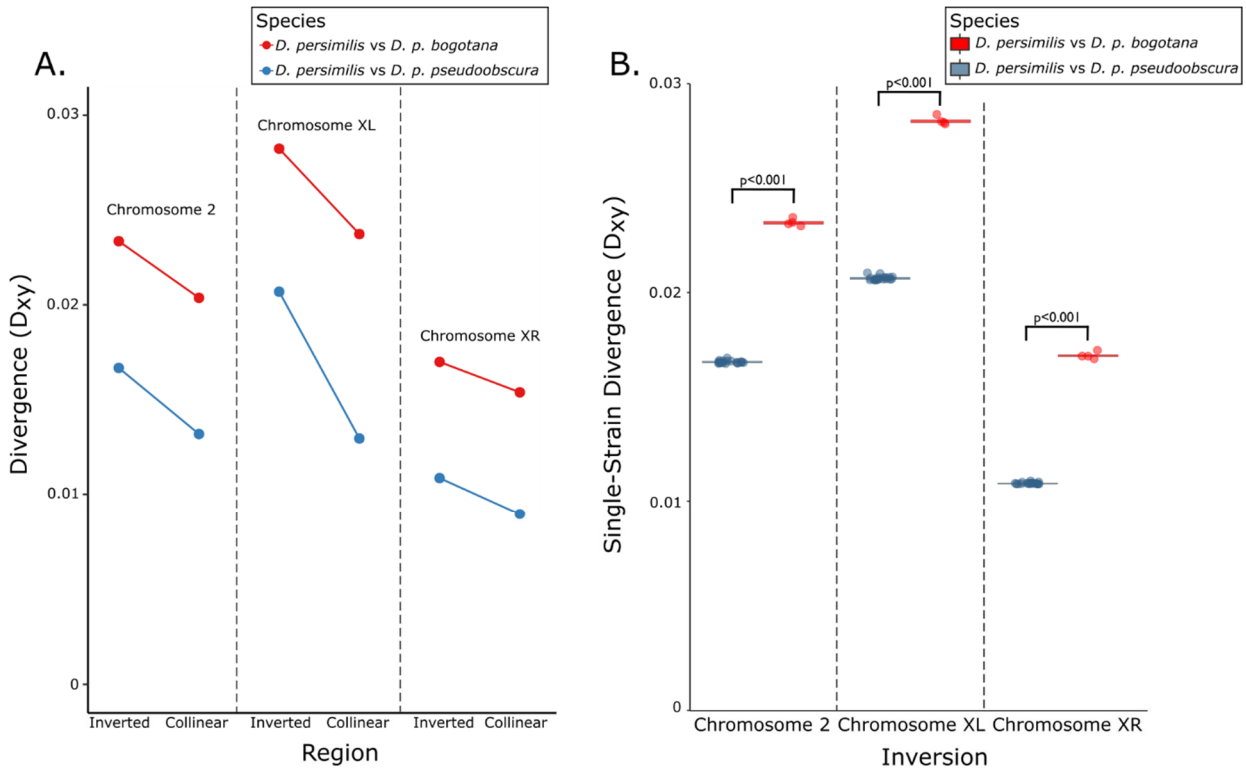


Figure 4 | Divergence in inverted vs collinear regions. (A) For the three inversion-bearing chromosomes, average divergence between species pairs is shown with inverted and collinear regions separated. (B) For each individual *D. pseudoobscura* genome, divergence from *D. persimilis* is shown for inverted regions.

The observation that divergence is lower in sympatry compared to allopatry even in the inverted regions is somewhat surprising. A possible explanation for the lower divergence in the sympatric species pair is that gene flow is homogenizing these species even in inverted regions. Though there is evidence that double crossovers and gene conversions occur within inversions (Schaeffer & Anderson 2005; Stevison *et al.* 2011; Crown *et al.* 2018; Korunes & Noor 2019), an alternative explanation is that *D. p. bogotana* has experienced more substitutions per site, possibly due to a population bottleneck upon colonization of South America leading to a subsequently small effective population size (Schaeffer & Miller 1991; Wang & Hey 1996; Machado *et al.* 2002).

To correct for differences in evolutionary rates in these populations, we considered the effects of introgression on divergence after adjusting for an elevated rate of fixation leading to a longer branch length in *D. p. bogotana*. We compared the divergence of each *D. pseudoobscura* subspecies from *D. persimilis* to the divergence of each *D. pseudoobscura* subspecies from *D. lowei* using the following equation to define the "introgression effect": (D_{xy} [*D.persimilis*:*D.p.bogotana*] - D_{xy} [*D.persimilis*:*D.p.pseudo.*]) – (D_{xy} [*D.lowei*:*D.p.bogotana*] - D_{xy} [*D.lowei*:*D.p.pseudo.*]). The first half of this equation should include the effects of branch length in *D. p. bogotana* and the effects of any introgression between *D. pseudoobscura* and *D. persimilis* (Figure 5A). Since *D. lowei* does not hybridize with any of these species, the second half of the equation

should reflect only the effects of branch length in *D. p. bogotana*. Thus, the difference between these terms should subtract the effects of evolutionary rate, leaving the effects of recent introgression. After subtracting the effects of branch length in *D. p. bogotana*, there is still evidence that gene flow has statistically significantly influenced patterns of divergence between *D. pseudoobscura* and *D. persimilis* (Figure 5B).

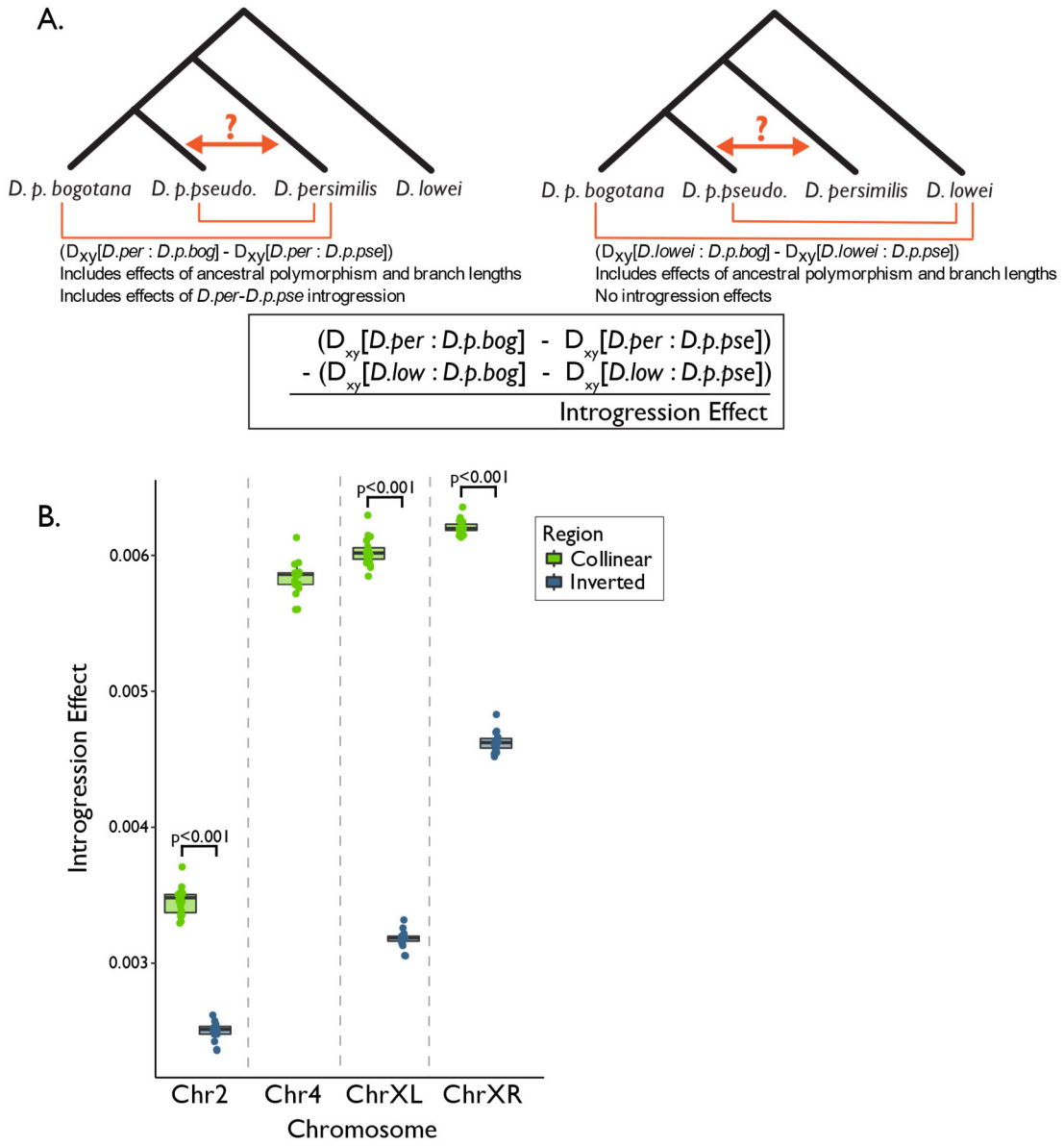


Figure 5 | The effects of introgression of divergence after correcting for evolutionary rate. (A) Shows the strategy for examining the "introgression effect" of gene flow on divergence. The first half of the equation compares the divergence of each *D. pseudoobscura* subspecies from *D. persimilis,* and the second half compares the divergence of each *D. pseudoobscura* subspecies from *D. lowei.* The difference between these terms, plotted in (B), should reflect the effects of introgression after correcting for branch length. The differences in the distributions of the introgression effect values between inverted and collinear regions are significant at $p < 0.001$, Mann-Whitney U test.

Discussion

Our model-based examination of gene flow and our implementation of Patterson's D-statistic to examine shared variation found between *D. persimilis* and sympatric and allopatric subspecies of *D. pseudoobscura* indicates post-speciation gene exchange, including both early post-speciation gene flow and gene flow within the past ~150,000 years. Specifically, we interpret the excess of ABBA > BABA sites as evidence that *D. pseudoobscura* and *D. persimilis* have exchanged genes in collinear regions since the split of *D. p. bogotana*. This evidence that introgression is driving patterns of divergence between *D. pseudoobscura* and *D. persimilis* is in agreement with previous reports of ongoing hybridization in these species (Dobzhansky 1973; Powell 1983; Machado & Hey 2003; Hey & Nielsen 2004). Despite our evidence for recent gene exchange, it appears that introgression is not the sole driver of patterns of divergence between these species overall. Here, we explore other important factors that might underlie the observed patterns of divergence, with particular consideration of how these factors might confound signals of recent introgression.

As seen in Figure 3 and in previous studies (Kulathinal *et al.* 2009), divergence is higher in inverted vs collinear regions in this system. This difference holds for divergence between *D. persimilis* and either *D. p. bogotana* or *D. p. pseudoobscura*. The lower observed divergence in sympatry compared to allopatry even in the inverted regions is somewhat surprising given the expectation that recombination in hybrids will be restricted in these inverted regions. This observation led us to consider the possibility that the allopatric subspecies, *D. p. bogotana*, might have experienced more nucleotide substitutions per site than the other taxa. Thus, we considered four factors that might contribute to the observed patterns of divergence with respect to chromosomal arrangement: 1) the segregation of ancestral polymorphism (as advocated by Fuller *et al.* (2018)), 2) increased branch length in the allopatric *D. p. bogotana,* 3) gene flow prior to the split of *D. p. bogotana*, and 4) recent/ongoing gene flow (the latter two discussed in Powell 1983; Wang & Hey 1996; Wang *et al.* 1997; Noor *et al.* 2001, 2007; Machado & Hey 2003; Hey & Nielsen 2004; Machado *et al.* 2007; Kulathinal *et al.* 2009; McGaugh & Noor 2012).

For the sympatric pair *D. persimilis* vs *D. p. pseudoobscura,* any difference in D_{xy} in inverted regions compared to collinear regions could be due to the segregation of ancestral inversion polymorphism or to post-speciation genetic exchange. In contrast, any difference in D_{xy} in inverted regions compared to collinear regions in the allopatric *D. persimilis* vs *D. p. bogotana* could be driven by the segregation of ancestral inversion polymorphism or by post-speciation gene flow prior to the split of *D. p. bogotana*. This comparison will not reflect any recent gene flow, since *D. p. bogotana* has evolved in allopatry for the past 150,000 years. We leveraged this contrast to isolate the effects of recent introgression on divergence, and our findings here suggest that any contribution from recent gene flow explains only a proportion of observed divergence patterns.

The observed "introgression effect" can be thought of as an indicator of the relative influence of recent introgression on the reduction in *D. persimilis* vs *D. pseudoobscura* divergence in sympatry vs allopatry. Our method does not account for all evolutionary forces shaping nucleotide divergence, so we do not consider it a quantitative measurement of introgression. However, it provides a useful way to consider the relative contribution of recent introgression compared to ancestral polymorphism and branch length. The second half

392 of the introgression effect equation (Figure 5A; D_{xy}[*D.lowei*:*D.p.bogotana*] - D_{xy}[*D.lowei*:*D.p.pseudo.*]) includes the
393 effects of branch length in *D. p. bogotana* and the effects of ancestral polymorphism without including any
394 effects of recent introgression. We can subtract this term from D_{xy}[*D.persimilis*:*D.p.bogotana*] to obtain a
395 hypothetical D_{xy} reflecting what divergence between *D. persimilis* and *D. p. pseudoobscura* might look like
396 without the effects of recent introgression, ancestral polymorphism, or longer branch length in *D. p. bogotana*.
397 Dividing the introgression effect by this hypothetical D_{xy} gives the hypothetical change in D_{xy} in allopatry vs
398 sympatry due specifically to recent introgression. Given the observed average introgression effect (0.0045)
399 and the hypothetical D_{xy}[*D. persimilis*:*D. p. pseudoobscura*] of 0.01), recent introgression might explain roughly
400 half of the reduction in D_{xy} observed in sympatry vs allopatry. As such, the homogenization effect of recent
401 introgression on sequence divergence appears far from trivial.
402
403 These results suggest that patterns of divergence between *D. persimilis* and *D. pseudoobscura* are explained by a
404 combination of segregating ancestral polymorphism and post-speciation gene flow. We applied a model-
405 based approach to investigate the timing of gene flow between *D. persimilis* and *D. pseudoobscura*. Our results
406 suggest that an isolation-with-initial-migration model best explains the divergence of *D. persimilis* and *D. p.*
407 *bogotana* when compared to a model of strict isolation. This result provides further evidence for gene flow
408 between *D. persimilis* and *D. pseudoobscura*, and it suggests that some of this gene flow occurred prior to the
409 split of *D. p. bogotana* and remains detectable in observed genetic patterns.
410
411 Our results question interpretations from earlier studies of this system. Given that *D. p. bogotana* can be
412 reasonably assumed to not be currently exchanging genes with either *D. persimilis* or *D. p. pseudoobscura*
413 (Schaeffer & Miller 1991; Wang *et al.* 1997), *D. persimilis*:*D. p. bogotana* divergence was argued to be a suitable
414 "negative control" for examining the effect of recent hybridization between *D. persimilis* and *D. p. pseudoobscura*
415 (Brown *et al.* 2004). By this argument, the effect of recent gene flow can be estimated by an allopatric vs
416 sympatric comparison of the difference in divergence (whether in DNA sequence or in phenotype) in
417 inverted regions to divergence in collinear regions. Specifically, Brown *et al.* (2004) and Chang and Noor
418 (2007) inferred multiple hybrid sterility factors between *D. p. bogotana* and *D. persimilis* that did not distinguish
419 North American *D. p. pseudoobscura* and *D. persimilis* (Brown *et al.* 2004; Chang & Noor 2007). Similarly,
420 Kulathinal *et al.* (2009) observed significantly greater sequence difference between *D. p. bogotana* and *D.*
421 *persimilis* than between North American *D. p. pseudoobscura* and *D. persimilis*. In both cases, the authors
422 interpreted the difference to result from recent homogenization of the collinear regions in the latter pair.
423 Based on our findings, we suggest this difference may result at least in part from the accelerated rate of
424 divergence in *D. p. bogotana*.
425
426 Overall, we caution that simple allopatry-sympatry comparisons can easily be misleading, and the population
427 histories and rates of evolution of the examined species should be carefully considered. Variation in N_e due to
428 events such as recent population bottlenecks in one of the taxa can dramatically influence evolutionary rates
429 (reviewed in Charlesworth 2009; Lanfear *et al.* 2014). Additionally, the process of genetic divergence that
430 shapes alleles responsible for local adaptation and hybrid incompatibility can extend deep into the history of
431 the species. In fact, the influence of inversions on the divergence of a species pair can predate the split of the
432 species. Inversion polymorphisms in the ancestral population of a species pair can contribute to patterns of

433 higher sequence differentiation between species in those inverted regions (Fuller *et al.* 2018). Separating these
434 effects requires an understanding of the timing and extent of introgression, which can only be understood
435 with an appreciation for the evolutionary processes occurring in each of the taxa at hand. Though there are
436 many remaining questions about how inversions shape divergence, we present evidence that inversions have
437 contributed to the divergence of *D. pseudoobscura* and *D. persimilis* over multiple distinct periods during their
438 speciation: 1) pre-speciation segregation of inversions in the ancestral population, 2) post-speciation gene
439 flow prior to the split of *D. p. bogotana*, and 3) recent introgression.
440

441 ## *Acknowledgements*

442 We thank all members of the Noor lab for helpful discussions. KLK was supported by the National Science
443 Foundation Graduate Research Fellowship Program under Grant No. DGE 1644868, and this work was
444 additionally supported by National Science Foundation grants DEB-1754022 and DEB-1754439 to MAFN,
445 and MCB-1716532 and DEB-1754572 to CAM.

Supplementary Figures and Tables

Supplementary Table 1 | Genomic data for introgression analysis.

Strains	Runs/Details
Drosophila lowei	
Lab3Lowei	See SRA Experiment SRX091467: SRR330416, SRR330418
Drosophila persimilis	
8 strains	SRA accessions will be provided by the time of final publication.
Drosophila pseudoobscura bogotana	
4 strains	SRA accessions will be provided by the time of final publication.
Drosophila pseudoobscura pseudoobscura	
19 strains	SRA accessions will be provided by the time of final publication.

Supplementary Table 2 | Inversion breakpoints relative to the *D. miranda* reference genome. Breakpoints (Chr 2, XL, XR) were originally described in Machado *et al.* (2007) and provided here in coordinates of DroMir2.2.

Chromosome	Approximate Proximal Breakpoint	Approximate Distal Breakpoint
Chromosome 2	15,154,864	23,081,317
X Chromosome – XL	10,619,782	18,389,958
X Chromosome – XR	24,174,893	11,348,640

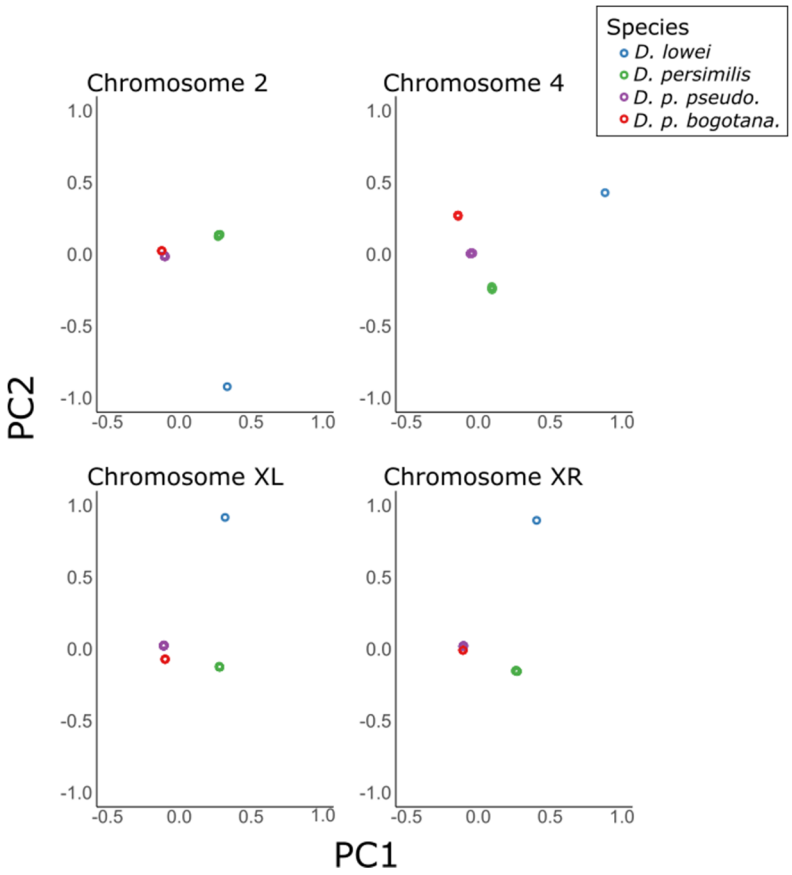
Supplementary Table 3 | Average divergence in inverted regions[1].

Chromosome	Average D_{xy}: *D. persimilis* vs *D. p. pseudoobscura*	Average D_{xy}: *D. persimilis* vs *D. p. bogotana*
Chromosome 2	0.017	0.023
X Chromosome – XL	0.021	0.028
X Chromosome – XR	0.011	0.017

[1] Statistical comparisons of these estimates were made using Mann-Whitney U tests to compare the divergence estimates in each region, split into 500 kb windows.

Supplementary Table 4 | Maximum-likelihood estimates under models of *D. persimilis - D. pseudoobscura bogotana* divergence and values of the maximized log-likelihood (LogL).

Model	θ_1	θ_2	θ_A	θ_{C1}	θ_{C2}	V	T1	M1	M2	logL
Loci sampled genome-wide										
Divergence in Isolation										
ISO	2.756	1.992	3.468	-	-	3.626	-	-	-	-33702.73
Divergence in isolation with migration										
IM	2.599	1.721	3.391	-	-	4.083	-	0	0.251	-33659.39
Divergence in isolation with initial migration										
IIM1	2.599	1.721	3.391	-	-	4.082	1e-06	0	0.251	-33659.39
IIM2	2.631	1.950	1.825	9.587	1.830	3.380	1.089	0.187	0.164	-33538.92
IIM3	2.661	1.715	2.303	10.453	1.839	3.413	0.909	-	0.353	-33543.76
IIM4	2.654	2.288	1.543	9.735	1.807	3.215	1.158	0.300	-	-33542.89
Loci sampled from collinear regions only										
Divergence in Isolation										
ISO	2.857	2.026	4.120	-	-	3.162	-	-	-	-25511.23
Divergence in isolation with migration										
IM	2.724	1.738	4.082	-	-	3.538	-	0	0.368	-25492.84
Divergence in isolation with initial migration										
IIM1	2.724	1.739	4.082	-	-	3.533	0.005	0	0.369	-25492.84
IIM2	2.710	1.988	1.995	8.328	1.903	3.938	1.202	0.424	0.237	-25442.42
IIM3	2.751	1.586	2.835	8.878	1.935	2.930	0.936	-	0.694	-25447.73
IIM4	2.732	2.371	1.718	8.616	1.873	2.820	1.234	0.569	-	-25444.03



Supplementary Figure 1 | Principal components analysis. The first two principal components are plotted for the four examined chromosome arms.

References

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. *Journal of Molecular Biology*, **215**, 403–410.

Van der Auwera GA, Carneiro MO, Hartl C *et al.* (2013) From FastQ data to high-confidence variant calls: The Genome Analysis Toolkit best practices pipeline. In: *Current Protocols in Bioinformatics*, pp. 11.10.1-11.10.33. John Wiley & Sons, Inc., Hoboken, NJ, USA.

Ayala FJ, Coluzzi M (2005) Chromosome speciation: humans, Drosophila, and mosquitoes. *Proceedings of the National Academy of Sciences*, **102 Suppl**, 6535–42.

Beckenbach AT, Wei YW, Liu H (1993) Relationships in the Drosophila obscura species group, inferred from mitochondrial cytochrome oxidase II sequences. *Molecular Biology and Evolution*, **10**, 619–634.

Brown KM, Burk LM, Henagan LM, Noor MAF (2004) A test of the chromosomal rearrangement model of speciation in Drosophila pseudoobscura. *Evolution*, **58**, 1856–60.

Butlin RK (2005) Recombination and speciation. *Molecular Ecology*, **14**, 2621–2635.

Chang AS, Noor MAF (2007) The genetics of hybrid male sterility between the allopatric species pair Drosophila persimilis and D. pseudoobscura bogotana: dominant sterility alleles in collinear autosomal regions. *Genetics*, **176**, 343–9.

Charlesworth B (2009) Effective population size and patterns of molecular evolution and variation. *Nature Reviews Genetics*, **10**, 195–205.

Costa RJ, Wilkinson-Herbots H (2017) Inference of gene flow in the process of speciation: An efficient maximum-likelihood method for the isolation-with-initial-migration model. *Genetics*, **205**, 1597–1618.

Crown KN, Miller DE, Sekelsky J, Hawley RS (2018) Local inversion heterozygosity alters recombination throughout the genome. *Current Biology*, **28**, 2984–2990.

Dobzhansky T (1973) Is there gene exchange between Drosophila pseudoobsura and Drosophila persimilis in their natural habitats? *The American Naturalist*, Vol. 107, No. 954, pp. 312–314.

Dobzhansky T, Epling C (1944) Contributions to the genetics, taxonomy, and ecology of Drosophila pseudoobscura and its relatives. *Carnegie Institute of Washington*, **544**, 1–46.

Feder JL, Gejji R, Powell THQ, Nosil P (2011) Adaptive chromosomal divergence driven by mixed geographic mode of evolution. *Evolution*, **65**, 2157–2170.

Fuller ZL, Leonard CJ, Young RE, Schaeffer SW, Phadnis N (2018) Ancestral polymorphisms explain the role of chromosomal inversions in speciation. *PLoS Genetics*, **14**, e1007526.

Guerrero RF, Rousset F, Kirkpatrick M (2012) Coalescent patterns for chromosomal inversions in divergent populations. *Philosophical transactions of the Royal Society B*, **367**, 430–8.

Heed WB, Crumpacker DW, Ehrman L (1969) Drosophila lowei, a new American member of the Obscura species group1,2. *Annals of the Entomological Society of America*, **62**, 388–393.

Hey J, Nielsen R (2004) Multilocus methods for estimating population sizes, migration rates and divergence time, with applications to the divergence of Drosophila pseudoobscura and D. persimilis. *Genetics*, **167**, 747–60.

Jackson BC (2011) Recombination-suppression: how many mechanisms for chromosomal speciation? *Genetica*, **139**, 393–402.

Korunes KL, Noor MAF (2019) Pervasive gene conversion in chromosomal inversion heterozygotes. *Molecular Ecology*, **28**, 1302–1315.

Kulathinal RJ, Stevison LS, Noor MAF (2009) The genomics of speciation in Drosophila: diversity, divergence, and introgression estimated using low-coverage genome sequencing. *PLoS Genetics*, **5**, e1000550.

Lanfear R, Kokko H, Eyre-Walker A (2014) Population size and the rate of evolution. *Trends in Ecology and Evolution*, **29**, 33–41.

Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler Transform. *Bioinformatics*, **25**, 1754–60.

Li H, Handsaker B, Wysoker A *et al.* (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, **25**, 2078–9.

Machado CA, Haselkorn TS, Noor MAF (2007) 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**, 1289–306.

Machado CA, Hey J (2003) The causes of phylogenetic conflict in a classic Drosophila species group. *Proceedings of the Royal Society B: Biological Sciences*, **270**, 1193–202.

Machado CA, Kliman RM, Markert JA, Hey J (2002) Inferring the history of speciation from multilocus DNA sequence data: The case of Drosophila pseudoobscura and close relatives. *Molecular Biology and Evolution*, **19**, 472–488.

Mallet J (2005) Hybridization as an invasion of the genome. *Trends in ecology & evolution*, **20**, 229–37.

McGaugh SE, Noor MAF (2012) Genomic impacts of chromosomal inversions in parapatric Drosophila species. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **367**, 422–9.

McKenna A, Hanna M, Banks E *et al.* (2010) The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research*, **20**, 1297–303.

Noor MAF, Garfield DA, Schaeffer SW, Machado CA (2007) Divergence between the Drosophila pseudoobscura and D. persimilis genome sequences in relation to chromosomal inversions. *Genetics*, **177**, 1417–28.

Noor MA, Grams KL, Bertucci LA, Reiland J (2001) Chromosomal inversions and the reproductive isolation of species. *Proceedings of the National Academy of Sciences*, **98**, 12084–8.

Patterson N, Moorjani P, Luo Y *et al.* (2012) Ancient admixture in human history. *Genetics*, **192**, 1065–93.

Payseur BA, Rieseberg LH (2016) A genomic perspective on hybridization and speciation. *Molecular ecology*, **25**, 2337–60.

Powell JR (1983) Interspecific cytoplasmic gene flow in the absence of nuclear gene flow: Evidence from Drosophila. *Proceedings of the National Academy of Sciences*, **80**, 492–5.

Powell JR (1992) Inversion polymorphisms in Drosophila pseudoobscura and Drosophila persimilis. *Drosophila Inversion Polymorphism*, 73–126.

Purcell S, Neale B, Todd-Brown K *et al.* (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. *American journal of human genetics*, **81**, 559–75.

Schaeffer SW, Anderson WW (2005) Mechanisms of genetic exchange within the chromosomal inversions of Drosophila pseudoobscura. *Genetics*, **171**, 1729–39.

Schaeffer SW, Miller EL (1991) Nucleotide sequence analysis of Adh genes estimates the time of geographic isolation of the Bogota population of Drosophila pseudoobscura. *Proceedings of the National Academy of Sciences*, **88**, 6097–101.

Stevison LS, Hoehn KB, Noor MAF (2011) Effects of inversions on within- and between-species

recombination and divergence. *Genome Biology and Evolution*, **3**, 830–41.

Stevison LS, Noor MAF (2010) Genetic and evolutionary correlates of fine-scale recombination rate variation in Drosophila persimilis. *Journal of Molecular Evolution*, **71**, 332–345.

Taylor SA, Larson EL (2019) Insights from genomes into the evolutionary importance and prevalence of hybridization in nature. *Nature Ecology and Evolution*, **3**, 170–177.

Wang RL, Hey J (1996) The speciation history of Drosophila pseudoobscura and close relatives: inferences from DNA sequence variation at the period locus. *Genetics*, **144**, 1113–26.

Wang Y, Hey J (2010) Estimating divergence parameters with small samples from a large number of loci. *Genetics*, **184**, 363–379.

Wang RL, Wakeley J, Hey J (1997) Gene flow and natural selection in the origin of Drosophila pseudoobscura and close relatives. *Genetics*, **147**, 1091–1106.

Yang Z (2002) Likelihood and Bayes estimation of ancestral population sizes in hominoids using data from multiple loci. *Genetics*, **162**, 1811–23.