Genomics of Natural Populations: Gene Conversion Events Reveal Selected Genes within the Inversions of *Drosophila pseudoobscura*

Stephen W. Schaeffer*, Stephen Richards[1], and Zachary L. Fuller[2]

*Department of Biology
The Pennsylvania State University
208 Erwin W. Mueller Laboratory
University Park, PA 16802-5301

[1]Current address:
Human Genome Sequencing Center, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030

[2]Current address:
23andMe, Inc., Sunnyvale, CA 94086

Data Availability: Sequence data in FASTA format and aligned NEXUS files are available at https://scholarsphere.psu.edu/resources/720f454e-ceb2-4139-a3a9-c59a5b7fdfaf . Additional data and computer code are available at https://github.com/SWSchaeffer/InversionGC

Running title: Gene Conversion within Inversions

Key Words: Gene Conversion, Inversion, Epistatic Selection, Local Adaptation, *Drosophila pseudoobscura*

Corresponding Author: Stephen W. Schaeffer
Department of Biology
The Pennsylvania State University
208 Erwin W. Mueller Laboratory
University Park, PA 16802-5301
(814) 865-3269
sws4@psu.edu

Abstract

When adaptive phenotypic variation maps within an inverted segment of a chromosome, researchers often despair because it is thought that the suppression of crossing over will prevent the discovery of selective target genes through fine scale genetic mapping. If an inversion polymorphism is old enough, then the accumulation of gene conversion tracts offers the promise that QTLs or selected loci within inversions can be identified. This study uses the inversion polymorphism of *Drosophila pseudoobscura* as a model system to determine if gene conversion analysis can be a useful tool for mapping adaptive loci within inversions. *D. pseudoobscura* has over 30 different chromosomal arrangements on the third chromosome (Muller C) in natural populations and their frequencies vary with changes in environmental habitats. Statistical tests identified outlier genes within inverted regions based on local clusters of fixed SNP differences. These outlier genes also had potentially selectable variation, either fixed amino acid differences or differential expression patterns among arrangements. Here, we use genome sequences of the inverted third chromosome (Muller C) to infer 98,443 gene conversion tracts for a total coverage of 142 Mb. We tested whether the frequency of gene conversion tracts in outlier versus non-outlier genes were equal. Gene conversion tracts occur less frequently in outlier versus non-outlier genes suggesting that selection removes exchanged DNA from multiple outlier genes. These data support the hypothesis that these inversions captured adaptive alleles and that strong epistatic selection removes recombinants generated through gene conversion. The pattern and organization of genetic diversity in the third chromosome in *D. pseudoobscura* is consistent with the capture of locally adapted combinations of alleles prior to inversion mutation events.

Introduction

Genomic sequence data have provided insights into the mechanisms for the origin, establishment, and maintenance of chromosomal inversions in natural populations. Comparison of sequences between different gene arrangements has allowed the identification of regions of chromosomes where breakage and rejoining events led to the origin of new gene orders (RICHARDS *et al.* 2005; RANZ *et al.* 2007). Ectopic exchange between repeat sequences at breakpoints provide one model for the facilitation of inversion mutations, while staggered cuts that duplicate genes is another alternative mechanism observed to generate new gene arrangements.

Once an inversion mutation arises in the population, it can be rapidly lost because its initial frequency (1/2N) is quite low. On the other hand, the new inversion can be established in the population either due to neutral or selective forces. Like point mutations, a selectively neutral inversion can drift to higher frequency through random genetic drift. There are several possible ways for an inversion to be established if selection is involved in the process. First, direct or position effects (SPERLICH AND PFRIEM 1986) may create positively selectable variation when the breakpoints either disrupt genes, create new hybrid genes, or alter gene expression of boundary genes at or near the lesions. Recent analyses have shown that breakpoints can modify chromatin architecture and disrupt expression of nearby genes (WRIGHT AND SCHAEFFER 2022). Second, indirect effects associated with the suppression of recombination can maintain advantageous haplotypes within the inverted region, which are held in linkage disequilibrium due to the reduced genetic exchange among gene arrangements in heterozygous individuals (KIRKPATRICK AND BARTON 2006). These models assume that the new inversion captures sets of beneficial alleles (CHARLESWORTH AND CHARLESWORTH 1973). Sets of alleles may be beneficial if they are free of deleterious recessive alleles (NEI *et al.* 1967) or are alleles that contribute to local adaptation (KIRKPATRICK AND BARTON 2006). Third, a new gene arrangement may rise to intermediate frequency because one beneficial allele is trapped within the inverted region (MAYNARD SMITH AND HAIGH 1974).

Once established, the new inversion will act as an isolated subpopulation and accumulate genetic differences from its ancestral arrangement (NAVARRO *et al.* 2000). Recombination or genetic flux among different gene arrangements will oppose genetic differentiation keeping sequences homogeneous among inversion types (NAVARRO *et al.* 1997). Genetic flux is initiated by double strand breaks between inverted chromosomes and are either resolved as genetic cross overs or as gene conversion events. Single cross overs within an inverted region of a Drosophila heterokaryotype will generate unbalanced inviable gametes that are selected against leaving only parental gametes (STURTEVANT AND BEADLE 1936). Cross overs are possible when two cross overs occur within the inverted region, but the probability of double cross over events depends on the size of the inverted segment in the heterokaryotypes (NAVARRO *et al.* 1997). Gene conversion events, on the other hand, are genetic exchanges that are resolved as non-cross overs leading to the exchange of small segments of 200 to 300 nucleotides among arrangements without the formation of unbalanced gametes (CHOVNICK 1973; ROZAS AND AGUADÉ 1993; ROZAS AND AGUADÉ 1994; SCHAEFFER AND ANDERSON 2005). Gene conversion events occur uniformly across the inverted segment, but the homogenizing effect is reduced because of the short segments exchanged among chromosomes (NAVARRO *et al.* 1997; MILLER *et al.* 2012; MILLER *et al.* 2016; KORUNES AND NOOR 2019). Overall, crossing over is expected to have a greater homogenizing effect than gene conversion because larger genomic segments are exchanged, but crossing over is less likely to occur relative to gene conversion events.

Under a neutral model and an infinite amount of time, maximal nucleotide divergence will be found at the breakpoints with the central inverted region showing less divergence among different arrangements (NAVARRO *et al.* 2000; NOOR *et al.* 2007; GUERRERO *et al.* 2012). When an inversion captures locally adapted alleles, selection opposes the homogenizing effect of genetic flux leading to local peaks of differentiation (GUERRERO *et al.* 2012). The observation of peaks of differentiation will become more apparent as time progresses and gene conversion events homogenize variation among arrangements except at selected loci.

Fuller *et al*. (2017) analyzed complete genome sequences of the third chromosome inversion polymorphism of *D. pseudoobscura* as a model system to determine mechanisms responsible for the establishment of inversions in populations. The third chromosome or Muller C (MULLER 1940) of *D. pseudoobscura* has a wealth of gene arrangements that were generated through a series of overlapping inversion events (DOBZHANSKY AND STURTEVANT 1938). These arrangements have been suggested to be targets of selection based on stable clines across heterogenous environments (DOBZHANSKY 1944; ANDERSON *et al.* 1991), stable altitudinal gradients (DOBZHANSKY 1948), and seasonal cycling (DOBZHANSKY 1948). Aquadro *et al*. (1991) used phylogenetic analysis of restriction fragment length polymorphisms to support the unique origin hypothesis of the *D. pseudoobscura* gene arrangements. Fuller *et al*. (2017) tested the 2,668 Muller C genes within each of six arrangements for evidence of adaptive evolution. Loci with significant elevated frequencies of derived nucleotide variants and with significantly long population specific branch lengths (SHRIVER *et al.* 2004; YI *et al.* 2010) were identified as outlier genes. In addition, outlier genes that were differentially expressed or had fixed inversion specific amino acids among arrangements were considered potential targets of selection.

Five of the six *D. pseudoobscura* arrangements had sufficient sample sizes to detect multiple outlier loci. These gene arrangements had between 36 and 149 outlier genes that were distributed within and outside the boundaries of the derived inversion (Figure 1). Outlier genes had decreased estimates of the recombination parameter rho (ρ) (CHAN *et al.* 2012) compared to intervening non-outlier loci (see Table 3 in FULLER *et al.* 2017) suggesting that genetic flux is limiting differentiation at a fine scale. Because genetic flux due to crossing over is expected to homogenize the central regions of inversions rather than local areas, we hypothesize that gene conversion is the predominant force homogenizing local regions among arrangements and that adaptive outlier genes will have fewer detected gene conversion events than non-outlier loci because natural selection is predicted to remove the gene conversion tracts that bring maladaptive variants to loci responsible for the increase of inversions.

Direct empirical estimates of gene conversion rates in inverted and colinear regions of *D. pseudoobscura* were estimated to be 3.1×10^{-5} and 7.7×10^{-6} converted nucleotides per base per generation, respectively (KORUNES AND NOOR 2019). In this experiment, Korunes and Noor (2019) generated genome wide SNP heterozygotes and used computational methods to infer gene conversion events on four of the five major chromosomal arms. In the cross between the PP1137 and MSH177 strains, the third chromosome was heterozygous for the ST and PP arrangements. Ten F_2 offspring were sequenced for this cross and 11 and 23 gene conversion events were scored within inverted and colinear regions respectively. Identical events were scored in the cross and that were presumed to occur during mitosis prior to meiosis (KORUNES AND NOOR 2019). If unique events are scored, then three and eight gene conversion tracts were observed in inverted and colinear regions, respectively, which reduces the estimates of gene conversion rates in inverted and colinear regions to be 8.5×10^{-6} and 2.4×10^{-6},converted nucleotides per base per generation, respectively. These estimates of gene conversion rates

are two orders of magnitude greater than the mutation rate, which would allow gene conversion to homogenize genetic diversity among arrangements. The Korunes and Noor (2019) show that gene conversion events can be detected after a single generation, but do not provide insights about the selective effects of these events. The current study will infer gene conversion events over the accumulated history of the third chromosome inversion polymorphism to determine if gene conversion events are retained.

A general computational method to detect gene conversion indirectly was developed based on the configuration of polymorphic nucleotides (phylogenetic partitions) at multiple sites across a sample of nucleotide sequences (STEPHENS 1985). This method was refined to detect gene conversion events in inversion systems when differentiated DNA segments are exchanged between one arrangement and another (BETRAN *et al.* 1997). Figure 2 shows an example of two detected gene conversion events on Muller C of *D. pseudoobscura*. The Betran *et al*. (1997) method assumes that the frequency of an informative variant is greater in the source arrangement than in the recipient. Two probabilities are used to infer gene conversion tracts from samples of gene arrangement sequence data. The first is the probability that a SNP is informative of a conversion event based on the frequencies of a variant within the source and recipient subpopulations or ψ. The second is the probability that a gene conversion event is extended to an additional nucleotide or φ. This method was applied by Schaeffer and Anderson (2005) to gene arrangement samples of short aligned sequences (339 to 517 bp) from seven gene regions of *D. pseudoobscura* They estimated the mean conversion tract length to be 205 bp and mean population conversion rate to be 3.4×10^{-6}, which is comparable to the rates inferred by Korunes and Noor (2019). With genome re-sequencing data, we now use the Bertran *et al*. (1997) method to do a comprehensive analysis of gene conversion tracts across Muller C of *D. pseudoobscura* to test whether outlier genes have fewer gene conversion tracts than non-outlier loci.

Materials and Methods

Strains. The genome sequencing of 54 strains of *D. pseudoobscura* that carry one of six different gene arrangements on Muller C have been previously described (FULLER *et al.* 2017). In brief, isofemale strains were collected from seven localities: Mount St. Helena, CA (MSH), Santa Cruz Island, CA (SCI); James Reserve, CA (JR); Mather, CA (MA); Kaibab National Forest, AZ (KB); Davis Mountains, TX (DM), San Pablo Etla, Oaxaca, Mexico (SPE). Crosses with either the Blade or Lobe third chromosome balancer strains were used to generate isochromosomal strains (Drosophila Species Stock Center Strain ID 14011-0121.173 is the Blade balancer for non-AR chromosomes, Genotype: or Bl cv AR/lethal CU; Strain ID 14011-0121.171 is the Lobe balancer for non-SC chromosomes, Genotype: or L SC/ or + lethal ST; and Strain ID 14011-0121.172 is an additional strain used in the balancer crosses, Genotype or pr) (see Figure 1 in DOBZHANSKY AND QUEAL 1938 for the general crossing scheme). The 54 strains used in this study carry one of six Muller C gene arrangements where the names are based on the locality where the chromosome was first collected (DOBZHANSKY AND STURTEVANT 1938): Arrowhead (AR, n=15); Standard (ST, n=8); Pikes Peak (PP, n=10); Chiricahua (CH, n=9); Tree Line (TL, n=9); and Cuernavaca (CH,n=3).

Genome Sequences. Each of the 54 strains were sequenced with Illumina HiSeq short reads (for full details of the sequencing, read mapping, and variant calling, see FULLER *et al.* 2017). Reads were mapped to the *D. pseudoobscura* v3.2 reference sequence. The Arrowhead MV2-25 reference genome is available at NCBI in BioProject PRJNA18793, the re-sequencing genomes are available in BioProject PRJNA358242. Two additional Pikes Peak genomes are available at NCBI as BioSamples SAMN00709017 and SAMN00709004. The set of 54 aligned sequences for the 19 Mb third chromosome were used in the gene conversion analysis.

Gene Conversion Analysis. We estimated gene conversion tract coverage in outlier and non-outlier genes with the following analysis. Coverage refers to how often conversion tracts occur a particular nucleotide in a gene region. As gene arrangements age, they accumulate arrangement specific mutations. These SNPs can be exchanged between arrangements in heterokaryotypes via crossing over or gene conversion. We used the approach of Betran *et al*. (1997) to infer gene conversion tracts across Muller C of *D. pseudoobscura*. This approach compares sets of sequences among different chromosomal arrangements to identify two or more SNPs in linkage disequilibrium that are found in two or more arrangements.

The gene conversion events detected by the Betran *et al*. (1997) approach are based on exchange between gene arrangement heterokaryotypes. The gene conversion analysis requires that the SNPs use a common gene order and that the distance among SNPs be preserved in the data set. The seven pairs of inversion breakpoints divide the third chromosome up into 14 regions where genes within the region are in the same order in all arrangements used in this study (Figure 1). While conversion events may occur near breakpoints (CROWN *et al.* 2018), they are unlikely to straddle the breakpoints.

The Gene Conversion analysis within DNASP v 6 (ROZAS *et al.* 2017) was used to infer gene conversion tracts across the 14 cytogenetic regions of Muller C. Each of the 14 regions were subdivided into smaller subregions of 250 kb each. The beginning or end of subregions were adjusted to accommodate breakpoint positions that resulted in subregions smaller than 100 kb. The final set of 73 subregions varied in size from 181 to 440 kb with 49 subregions being 250 kb. This was done to accommodate DNASP memory constraints and to maximize the speed of analysis. All pairs of gene arrangements were compared in each of the 73 subregions with the Gene Conversion analysis in DNASP (15 total comparisons: AR-ST, AR-

227 PP, AR-CH, AR-TL, AR-CU, ST-PP, ST-CH, ST-TL, ST-CU, PP-CH, PP-TL, PP-CU, CH-TL,
228 CH-CU, and TL-CU). Each analysis excluded nucleotide sites with alignment gaps and the
229 output provided lists of conversion tracts within strain, the beginning and end coordinates of the
230 tract, the tract length excluding gaps, mean value of ψ, number of informatic sites, and a list of
231 sites with its associated value of ψ. Results of each of the 1,095 pairwise analyses were stored
232 in separate text files and were also combined into a single file for subsequent analyses.
233 A gene conversion tract can be detected multiple times in this analysis. There are two
234 sources of redundancy. First, a gene conversion tract can be inferred from multiple pairwise
235 comparisons of gene arrangements. For instance, a gene conversion tract in an AR strain might
236 be inferred from a comparison of AR with CH or with TL. Second, a gene conversion tract can
237 be inferred in multiple strains of the same gene arrangement. It is assumed that gene
238 conversion tracts result from unique events and finding the same event in multiple strains is a
239 result of the tract increasing in frequency (BETRAN *et al.* 1997). Only one tract from a single
240 strain is used when the conversion event is found in multiple strains. The effect of removing
241 redundant gene conversion tracts from the data set is a conservative adjustment that lowers the
242 overall gene conversion tract coverage per site in both outlier and non-outlier genes.
243 A histogram of gene conversion coverage relative to outlier and non-outlier genes was
244 constructed from the set of non-redundant gene conversion tracts (Figure 3). Outlier and Non-
245 outlier transcripts were mapped to Muller C (Figure 3A). Next, gene conversion tracts were
246 mapped to Muller C (Figure 3B). Finally, a histogram of gene conversion coverage was
247 constructed by counting the number of tracts that cover each nucleotide across the
248 chromosome (Figure 3C).
249 The mean coverage per nucleotide was estimated for outlier and non-outlier genes from
250 the gene conversion histogram. The sum of gene conversion tract coverage was estimated
251 across the exons of each gene and divided by the total number of nucleotides in the gene. The
252 mean gene conversion coverage was determined separately for outliers and non-outlier genes.
253 To determine if the donor and recipient conversion events were independent across the
254 14 syntenic block regions and overall, we permuted the donor and recipient classifications
255 across the 14 regions. We constrained the permutations so that only non-self-arrangement
256 events occurred. A total of 10,000 random permutations were used. For each permutation of
257 the donors and recipients, the number of gene conversion events was re-estimated and
258 compared with the observed data to determine the probability of obtaining more extreme values
259 as well as the minimum and maximum values found across all permutations.
260 **Maximum Likelihood Estimates of ϕ and Expected Number of Conversion Tracts.**
261 We used the maximum likelihood approach of Betran *et al*. (BETRAN *et al.* 1997) to infer ϕ within
262 each of the 14 regions and for each donor-recipient pair using the observed list of conversion
263 tracts for each pair (See equations 4, 6, and 8). The second derivative of the likelihood
264 equation was used to estimate the asymptotic variance, which is used to infer a lower and upper
265 limit of tract length. We did separate analyses for each region and donor-recipient pair because
266 we assumed that the breakpoints that defined the 14 regions might interfere with the
267 progression of conversion tracts. This may be a conservative assumption because we define
268 the syntenic block regions based on the set of 13 breakpoints of gene arrangements used in this
269 study.
270 **Tests for Differences in Gene Conversion Tract Coverage between Outlier and**
271 **Non-Outlier Genes.** We tested the hypothesis that outlier and non-outlier genes had
272 equivalent gene conversion coverage using several methods. This assumes that gene
273 conversion can occur uniformly across the chromosome. First, we used a permutation test

where we randomly assigned outlier and non-outlier designations to the genes within the arrangement. A total of 10,000 replicates were used to determine the frequency of random outlier assignments that had a mean tract coverage less than the observed value. Second, we randomly assigned the observed gene conversion tracts to Muller C without replacement 10,000 times and for each random map we re-constructed the conversion tract histogram. The average coverage in outlier and non-outlier genes was determined for each replicate and the frequency of replicates that had a value less than the observed was estimated. Third, we used the maximum likelihood estimate of ϕ and the expected number of conversion tracts in each region to generate a random set of conversion events across the different regions of Muller C. We used 10,000 replicates of gene conversion tracts generated with maximum likelihood parameters to test whether the coverage differed between outlier and non-outlier genes.

Results

Numbers and Lengths of Conversion Tracts in 14 Syntenic Block Regions. There were 98,443 non-redundant gene conversion tracts observed across Muller C. The mean number of tracts per strain is 1930.2. Sample sizes varied among arrangements, which could bias the number of gene conversion tracts detected. The distribution of gene conversion tracts rejects the null hypothesis that conversion tracts are proportional to the number of strains sampled within each gene arrangement using a Chi-Square Goodness of Fit test (X^2=2281.0, df=4, P<0.0001). The AR, CH, and TL arrangements have more gene conversion tracts than expected while the PP and ST arrangements have fewer tracts than expected (Table 1). Gene conversion tract numbers are not similar in the 14 syntenic block regions in the five different arrangements based on a Chi-Square test for Homogeneity (X^2=366.6, df=52, P<0.0001) (Figure 4). Examination of residuals shows that the AR arrangement has a deficiency of tracts in regions 5 and 9, the CH arrangement has an excess of gene conversion tracts in region 5 and a deficiency of tracts in regions 11 and 14, the PP arrangement has a deficiency of tracts in region 5 and an excess of tracts in region 9, the ST arrangement has deficiencies of tracts in regions 3 and 5 and an excess of tracts in regions 8 and 10, while the TL arrangement has an excess of tracts in region 3 and deficiencies of tracts in regions 5 and 8.

Median observed gene conversion tract lengths are greatest in the proximal and distal regions of Muller C compared to the central inverted regions (Figure *5*). The central areas of the chromosome which are often within inverted regions in heterokaryotypes have an observed median tract length of 112, 160, 97, 127, and 115 bps in the AR, CH. PP, ST, and TL arrangements, respectively. We used a random permutation test to determine if the median gene conversion tract lengths are lower or greater than expected given the distribution of all observed tract lengths. The test was performed by shuffling the list of tracts without replacement, which randomly assigns tracts to the 14 different regions keeping the number of tracts per region the same as the observed. For each of the 10,000 permutation replicates, the region median was re-estimated and the probability of obtaining a median greater than or less than the observed value within each region was determined. Figure *5* shows the regions that had median gene conversion tract lengths that were significantly less than or greater than expected given the distribution of all tract lengths. Proximal regions 1 and 2 and distal regions 12 through 14 had median tract lengths that were greater than expected in two or more gene arrangements while central regions 4 through 10 had median tract lengths that were less than expected in four or more karyotypes. Regions 3 and 11 had arrangements with median tract lengths less than or greater than expected reflecting boundary regions that the transition between largely collinear or inverted segments of the chromosome.

We observe gene conversion events between all pairs of arrangements across Muller C ranging between 1,270 conversion events from ST to PP and 9,281 conversion events from CH to AR. Figure *6* shows the polarity of the observed gene conversion events in the 14 syntenic block regions and across all regions. For all gene conversion events across all regions, two pairs showed a deficiency of events in both directions (CH-PP and ST-TL), one pair showed an excess of events in both directions (PP-TL), and the other seven pairs showed a deficiency in one direction and an excess in the other direction. The test of independence of donor to recipient events for the 14 different regions found subsets of regions with significant deficiencies or excesses of conversion events with a tendency toward what was found for the sum of all events. For instance, if the overall trend was toward a deficiency of events, then most regions with a significant departure tended to also show a deficiency, e.g., ST to CH transition had a significant deficiency (2,195) and nine of 14 regions also showed a significant deficiency. For the arrangement pairs likely to form heterozygotes at frequencies of 10% or more in natural

populations (CH/ST,CH/PP, AR/ST, and AR/PP) CH tended to be a donor to AR and ST, ST tended to be a donor to AR, and AR tended to be a donor to PP.

Maximum Likelihood Estimates of ϕ and Expected Number of Conversion Tracts. The estimates of ϕ and the expected number of conversion tracts along with other values for all donor-recipient pairs in each region are in the Supplemental Materials. These values are used below in tests of gene conversion coverage in outlier genes.

Gene Conversion Tract Coverage in Outlier and Non-Outlier Genes. The 98,443 observed conversion tracts were mapped to Muller C and the mean coverage per nucleotide was estimated for the exons of each transcript. The expected gene conversion tract coverage for each nucleotide across Muller C is at least 1.2 per nucleotide for all arrangements except for PP, which had an expected coverage rate of 0.8 tracts per nucleotide. Thus, at least one conversion tract should occur in each gene across the chromosome (Table 1).

Figure 7 shows the gene conversion tract coverage for outlier and non-outlier genes across Muller C in five gene arrangements of *D. pseudoobscura*. Proximal and distal regions of all arrangements have higher conversion tract coverage than central inverted regions. Outlier genes tend to have lower levels of coverage than non-outliers. The maximum coverage values are sample size dependent. Standard, Chiricahua, Pikes Peak, and Tree Line have sample sizes between 8 and 10 with maximum coverage of 10 tracts/nucleotide. Arrowhead, on the other hand, has a sample size of 15 with a maximum coverage of 18.

The mean gene conversion tract coverage estimated for the 2,668 genes on Muller C shows that non-outlier genes have higher mean coverage than outlier genes (Table 2). Gne conversion tract overage in non-outlier genes matches the expected value for the entire chromosome, while outlier genes have coverage values < 0.1. We used three tests to determine if the lower gene conversion coverage is expected by chance. First, we used a permutation test to determine whether outlier and non-outlier genes are expected to have similar gene conversion tract coverage per nucleotide. For each of 10,000 replicates, genes were randomly assigned to either the outlier or non-outlier class. The average gene conversion tract coverage was then re-estimated for the outlier class. For all gene arrangements, the observed mean gene conversion coverage was always less than random permutations. One factor to consider is that we filtered the conversion tracts to remove redundant tracts. If the redundant tracts were to be included, we would likely see higher coverage both in outlier and non-outlier genes, but still show a significant difference.

Second, we shuffled the observed gene conversion tracts along the Muller C chromosomal map. This test removes potential biases about where conversion tracts are in the genome. For each of 10,000 replicates, all conversion tracts were randomly placed on the chromosome. For each tract, a nucleotide position was drawn from a uniform distribution and a tract of the given length was mapped beginning at that location. The newly mapped conversion tracts were used to re-estimate the mean coverage for outlier genes. All arrangements showed that the observed mean coverage in outliers was always less than the replicates where gene conversion tracts were shuffled.

Third, we used maximum likelihood estimates of region-specific ϕ and the expected number of gene conversion tracts ($E(k_T)$ in the 14 different syntenic block regions to generate a set of conversion tracts. The parameter ϕ is the probability of extending the conversion tract to an additional nucleotide and the process of generating tract lengths is geometric random variable. For each gene conversion tract within a recipient arrangement, we sampled a tract length as a geometric random deviate. If the tract length was greater than the region size, then a new tract length is chosen. Next, a random nucleotide site was chosen from a uniform distribution to begin the gene conversion tract. If the mapped tract spanned two regions, a new starting coordinate was chosen. A total of $E(k_T)$ tracts are generated for each region for each of the five donor arrangements, where k_T is the expected number of conversion tracts (See page

95 in BETRAN *et al.* 1997). A total of 10,000 replicates was generated and the mean tract coverage in outlier genes was re-estimated. This test found that the observed mean conversion tract coverage in outlier genes was less than was found with random conversion tracts. The results of this test should be viewed with caution because the expected number of true tracts is greater than the observed number of tracts. Thus, it is not surprising that we observe higher gene conversion tract coverage in outlier genes than in our replicates. The important result from these tests is that outlier genes should have higher levels of coverage than was observed.

Discussion

The work presented here shows that gene conversion events among arrangements are readily detectable and serve as a source of genetic flux among the gene arrangements of *D. pseudoobscura*. The number of observed conversion tracts tends to be uniformly distributed across Muller C (Figure 4) except for the proximal two regions of the chromosome consistent with theory on genetic flux (NAVARRO *et al.* 1997). Fewer conversion tracts may result from the initiation of fewer double stranded breaks in the proximal regions. While regions 1 and 2 are closer to the centromere where crossing over is reduced (LINDSLEY AND SANDLER 1977), these regions are in euchromatic DNA, which is unlikely to be affected by the centromeric effect. More likely is that the Betran *et al*. (1997) method has lower power to detect gene conversion events in regions 1 and 2. This method relies on informative nucleotide polymorphisms to detect an exchange between two gene arrangements. Because regions 1 and 2 are outside most inverted regions, genetic exchange is likely to reduce differentiation and prevent conversion tract detection.

The observed median gene conversion tract lengths are largest in the proximal and distal regions of Muller C and are reduced in size in the inverted regions (Figure *5*). The detected events in the proximal and distal regions may be the result of cross overs rather than gene conversion tracts. The Betran *et al*. (1997) method uses informative nucleotide sites to infer exchanges between arrangements. The accumulation of genetic differences among gene arrangements can extend beyond the boundaries of the inversion breakpoints providing information about gene exchanges. In this case, the exchange events are longer than what is observed within the inverted regions. The most proximal region (1) has the largest median conversion track length compared to the most distal region (14). This may reflect that region 1 is further from the influence of interference caused by inversion breakpoints (GONG *et al.* 2005). Most common arrangement pairs are homosequential across regions 1 to 3 with the exception of the PP arrangement, which has its proximal breakpoint at the boundary of regions 2 and 3. As a result, there is likely to be less interference of inversion breakpoints in the more proximal regions leading to resolution of exchanges as cross overs and longer sequences being exchanged.

We observed exchanges among all pairs of gene arrangements, although the number of events from donor to recipients are either lower or greater than expected in a permutation test (Figure *6*). For most pairs of arrangements, there is a bias in which arrangement is the donor versus the recipient. For instance, AR is the recipient of more tracts from CH than expected while there is a deficiency of events in the reverse direction. There is no clear pattern that might suggest the nature of the bias. The expected frequency of heterozygotes, common versus rare, does not explain the biases (pink versus black arrows, respectively in Figure *6*). No arrangement tends to be biased for being a donor or recipient. For instance, four arrangements (AR, PP, ST, and TL) have excess conversions with two arrangements and a deficient of exchanges with two arrangements. CH is the only arrangement that shows a bias against being a recipient of conversion tracts with deficiencies with three arrangements and one case of showing an excess of being a donor with AR. Thus, we failed to find a plausible explanation for the deficiencies and excesses of being donors or recipients.

The rate of gene conversion is sufficient to homogenize the sequences among the gene arrangements, yet one observes differentiation of some genes and not others across the inverted segments. Our analysis of gene conversion coverage tracts shows that gene conversion occurs less frequently in outlier versus non-outlier genes. Tests that random shuffle the gene conversion tracts shows that there are enough events to homogenize the sequences. The reason we can detect these outlier loci is that we observe differentiation in the face of the homogenizing effect of gene flux, which works in a similar manner as gene flow. This requires that the various gene arrangements accumulate genetic differences and that there is sufficient

time for gene conversion events to accumulate. The Muller C inversion polymorphism in *D. pseudoobscura* is estimated to be at least 1 million years old (AQUADRO *et al.* 1991; WALLACE *et al.* 2011) with the AR arrangement being the newest arrangement of the set originating 510,000 years ago. Despite being a young arrangement, our sample of AR chromosomes has over 30 thousand gene conversion events with a mean coverage of 2.33 events per nucleotide and 36 detected outlier genes with a mean coverage of 0.11. A similar pattern of reduced coverage in outlier genes is observed in the other four older arrangements.

Essentially what is observed here is an example of survivorship bias. A famous example of this survivorship bias is World War II aircraft and the question of how much armor to add to better protect the plane from bullets from anti-aircraft guns. Too much armor makes the plane too heavy and not enough armor results in the plane being shot down due to bullets from anti-aircraft guns. Examination of where bullet holes were in the aircraft were used as data to determine where to add armor. Initially, it was thought that one should add armor to cover regions that had been hit with excessive bullets, but it was pointed out by the Statistician Abraham Wald that the returning planes had survived so one should armor regions without bullet holes. In this study of inversions, genes within inversions are the airplanes and gene conversion events are the bullets being fired from anti-aircraft guns. The outlier genes are analogous to the parts of returning airplanes without bullet holes. Gene conversion events survive over time in less consequential genes within the inverted regions, while gene conversion events in outlier genes from different arrangements do not survive because they damage the important features of the inversion. We assume that these outlier genes are in fact adaptive genes essential for the gene arrangements survival. Thus, gene conversion allows one to detect a signal of genes involved in adaptive evolution within gene arrangements despite cross over suppression in heterokaryotypes.

Each of the major gene arrangements examined in *D. pseudoobscura* have evidence for multiple outlier genes that show arrangement specific variation. We show here that the arrangement specific variation in outlier genes is observed despite the homogenizing effect of extensive genetic flux among arrangements driven by gene conversion events through the history of these chromosomes. These data are consistent with a model where the inversions captured sets of genes involved in local adaptation (KIRKPATRICK AND BARTON 2006). We propose that the local populations of *D. pseudoobscura* in different physiographic provinces (DOBZHANSKY 1944; LOBECK 1948) are selected for different locally adapted phenotypes encoded by multiple genes. These polygenic traits were generated through shuffling of allelic to generate adaptive combinations for different environments. Adaptive combinations in one environment may not be beneficial in another habitat and are likely to generate maladaptive types through recombination. This creates a setting for chromosomal inversions to establish if they capture adaptive combinations of genes (CHARLESWORTH AND CHARLESWORTH 1973). The set of adaptive genes is not likely to be recognized as selective targets initially because there will be extensive linkage disequilibrium of SNPs across the chromosome. As the ancestral and derived arrangements diverge in sequence, this sets up conditions to recognize adaptive loci. Gene conversion begins to homogenize the less important regions within the inversion and with sufficient time differentiated adaptive genes can be recognized.

If genes within the inverted region are neutral, then we would expect to see the highest levels of genetic differentiation at or near the inversion breakpoints and not within the central regions of the inversion (NAVARRO *et al.* 2000; MUNTE *et al.* 2005; NOOR *et al.* 2007; GUERRERO AND KIRKPATRICK 2014). We observe outlier genes across the inverted regions that are not restricted to the breakpoint regions even in the largest inversion event considered in this study, Pikes Peak. The inversion that gave rise to the PP arrangement accounts for 60% of Muller C. Double cross overs would be expected to homogenize the central region of the inverted region, yet 129 outlier loci are distributed across the rearranged segment. This differentiation pattern is the so-called suspension bridge pattern where the towers of the bridge are the regions of high

differentiation or outlier genes and the cables are the regions of reduced differentiation caused by gene conversion.

An alternative model to the capture model is one where new inversions gain beneficial alleles within the arrangement over time. This model does not seem plausible because in this scenario, one would predict that the repeated selected sweeps would continue to remove segregating variation from the inverted region making it impossible to identify outlier genes. A majority of nucleotide sites within the arrangement would be in near absolute linkage disequilibrium at all times due to the repeated sweeps (FULLER *et al.* 2019).

In summary, we show here that gene conversion is a powerful force for genetic flux among different gene arrangements that can allow population geneticists to identify adaptive loci within the inversion. Adaptive loci are discovered as differentiated loci surrounded by regions of relative homogeneity among arrangements. This approach will not work for recent inversions because sufficient time must pass for differentiation among arrangements and for gene conversion to homogenize non-consequential genes. The inversion polymorphism of *D. pseudoobscura* is sufficiently old that we have identified multiple loci within each arrangement are identified as selective targets. We think this supports the model that each arrangement captured sets of locally adapted alleles. Locally adapted inversions were favored because they prevented the formation of maladaptive combinations from other locally adapted migrant chromosomes (KIRKPATRICK AND BARTON 2006). When an adaptive QTL is discovered to map within an inversion, researchers often have a look of despair on their faces. This study suggests there is hope to identify an adaptive QTL within the confines of an inversion.

Acknowledgements

The data presented in this work was generated through support from the National Institutes of Health R01 GM 098478 to (S.W.S.). We thank the sequencing team at the Baylor College of Medicine Human Genome Sequencing Center for their help in sequencing data generation. The funders had no role in study design, data collection, data analysis, the decision to publish, or preparation of the manuscript.

521 Table 1. Gene conversion tracts detected in five recipient gene arrangements in *D. pseudoobscura*.

Gene Arrangement	Strains	Observed Tracts	Expected Tracts	Tracts/Strain	Σ Tract Length	GC Coverage
Arrowhead (AR)	15	30,129	44,342	2,008.6	46,173,391	2.33
Standard (ST)	8	14,982	26,185	1,872.8	25,122,088	1.27
Pikes Peak PP)	10	14,043	20,971	1,404.3	15,948,639	0.81
Chiricahua (CH)	9	20,935	33,259	2,326.1	31,529,250	1.59
Tree Line (TL)	9	18,354	34,861	2,039.3	23,830,343	1.20
Total	51	98,443	159,618	1,930.2	142,603,711	7.20

522 Strains, number of strains within arrangement; Observed Tracts, observed number of gene conversion tracts detected; Expected
523 Tracts, expected number of gene conversion tracts = Observed Tracts/(1-Pr(unobserved tracts)), where Pr(unobserved tracts) is
524 from equation 9 in Betran *et al*. (1997); Tracts/Strain, mean number of gene conversion tracts per strain; Tracts, Σ Tract Length, sum
525 of nucleotides in all gene conversion tracts; GC Coverage, Σ Tract Length divided by nucleotides in Muller C (19,787,792).

526 Table 2. Gene Conversion Tract Coverage of Outlier and Non-Outlier Genes on Muller C of
527 *Drosophila pseudoobscura.*

Gene Arrangement	Outlier Obs Coverage (SD)	Non-Outlier Obs Coverage (SD)	Min – Max
Arrowhead (AR)			
Transcripts	36	2,632	
Nucleotides	78,468 bp	4,099,525 bp	
Permute Out	0.112 (0.503)*	2.585 (3.064)	1.406 – 4.037
Shuffle CT			0.679 – 2.652
ML Parameter			2.018 – 4.066
Standard (ST)			
Transcripts	82	2,596	
Nucleotides	179,724 bp	3,998,269 bp	
Permute Out	0.067 (0.356)*	1.415 (1.662)	0.974 – 1.926
Shuffle CT			0.563 – 1.758
ML Parameter			1.415 – 2.565
Pikes Peak (PP)			
Transcripts	149	2,519	
Nucleotides	365,249 bp	3,812,744 bp	
Permute Out	0.048 (0.295)*	0.922 (1.488)	0.609 – 1.156
Shuffle CT			0.233 – 0.922
ML Parameter			0.922 – 2.242
Chiricahua (CH)			
Transcripts	40	2,628	
Nucleotides	95,262 bp	4,082,731 bp	
Permute Out	0.055 (0.322)*	1.735 (1.873)	1.087 – 2.709
Shuffle CT			0.483 – 1.735
ML Parameter			1.595 – 3.199
Tree Line (TL)			
Transcripts	54	2,614	
Nucleotides	113,586 bp	4,064,407 bp	
Permute Out	0.101 (0.444)*	1.290 (1.656)	0.832 – 1.917
Shuffle CT			0.216 – 1.290
ML Parameter			0.658 – 1.660

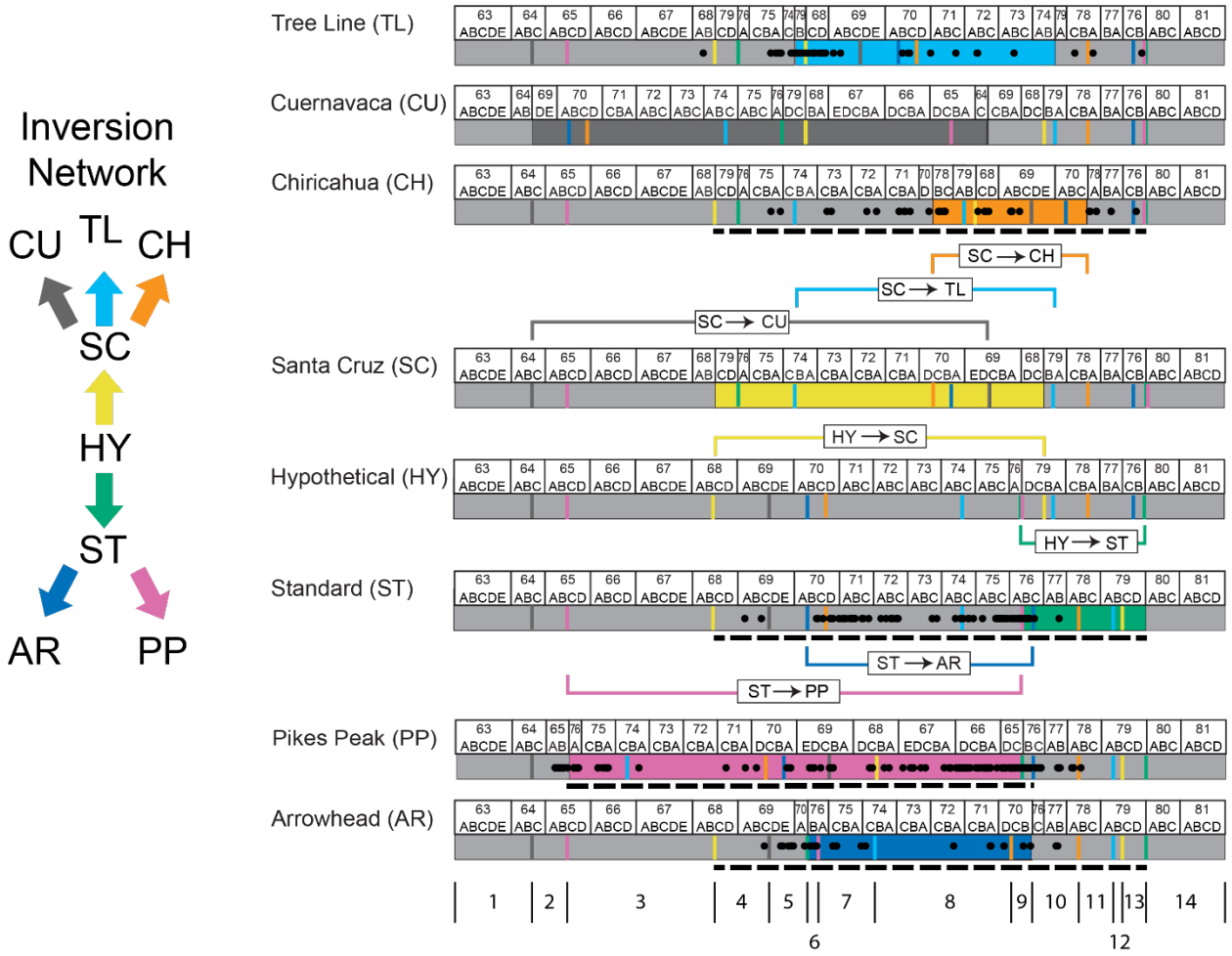
528 *, $P<0.0001$ the observed coverage value for the outlier genes is less than that observed in the
529 randomly permuted data. Min – Max, minimum and maximum values from 10,000 replicates;
530 Permute Out, permutation test where outlier label was permuted; Shuffle CT, permutation test
531 where the observed conversion tracts were shuffled; ML Parameter, test where the expected
532 number of conversion tracts with random tract length with maximum likelihood parameter ϕ is
533 used.

534 **Figure Legends**

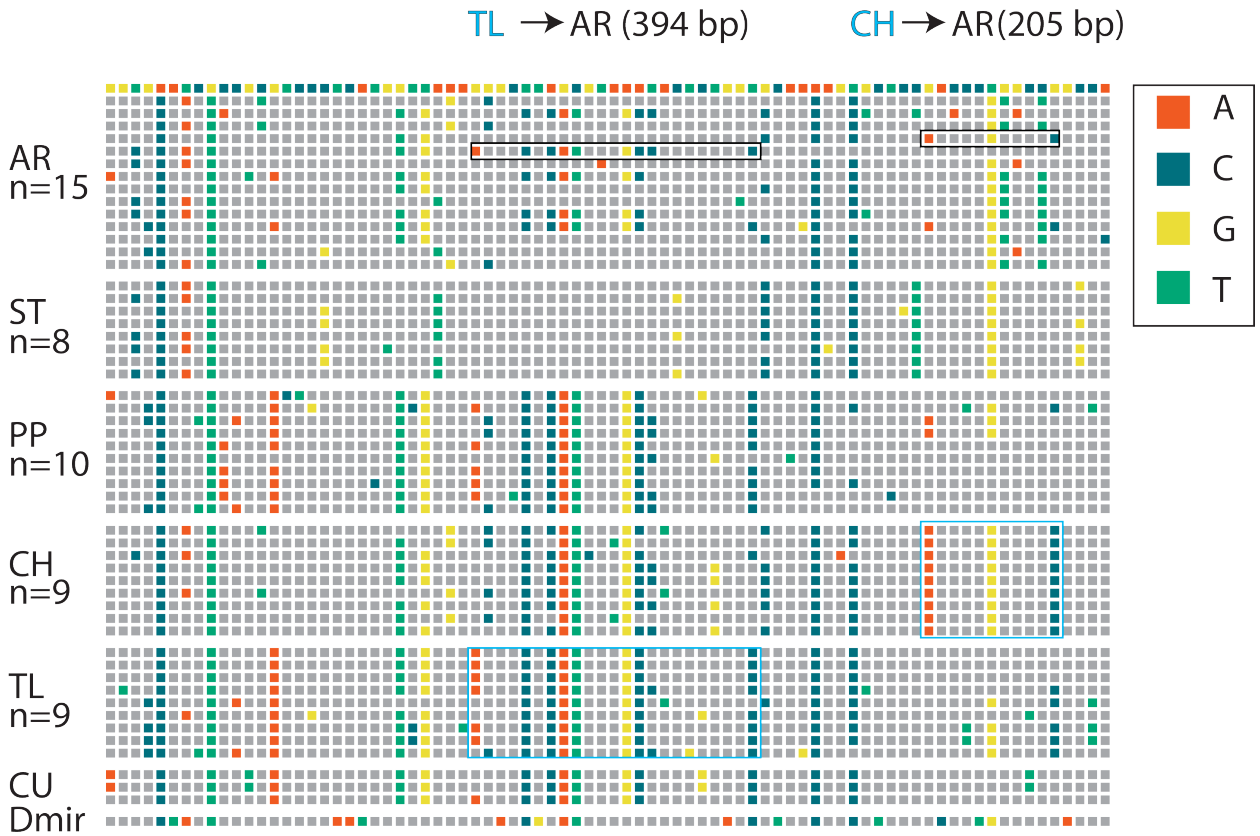
535 Figure 1. Network of *Drosophila pseudoobscura* gene arrangements. The network of gene
536 arrangements (HY, ST, AR, PP, SC, CU, TL, and CH) and the inversion events (colored
537 arrows) are shown to the left where the HY arrangement is the ancestral arrangement.
538 One the right side are the cytogenetic maps for the gene arrangements along with the
539 inverted segments indicated by the colored brackets. The color-coded lines below each
540 arrangement show the locations of the different inversion breakpoints. Outlier genes in
541 each of the gene arrangements are indicated by a black dot. An outlier gene had a
542 significant elevation of derived alleles or population specific branch lengths and had
543 either fixed amino acid differences or differential expression. Below the AR
544 arrangement, 14 regions where genes are in the same order among all chromosomes in
545 the set. The dotted line below the CH, ST, PP, and AR arrangements indicates the
546 regions that would be inverted in ST/AR, ST/CH, AR/CH, or AR/PP heterozygotes
547 whose frequency is at least 10% at least one locality in the Southwestern United States
548 (DOBZHANSKY 1944).
549 Figure 2. Detection of gene conversion. A map of SNP polymorphism is shown for a short
550 region of the third chromosome of *D. pseudoobscura*. Each column represents a SNP
551 with gray boxes matching the reference genome in the first row. A nucleotide that differs
552 from the reference genome is represented in one of four colors (see legend on the right-
553 hand side). Multi-site genotypes are shown for six gene arrangements along with their
554 sample sizes: Arrowhead (AR), Standard (ST), Pikes Peak (PP), Chiricahua (CH), Tree
555 Line (TL), Cuernavaca (CU), and the outgroup *D. miranda*. Two inferred gene
556 conversion events are shown with the black outline box indicating the recipient sequence
557 and the light blue outline boxes shown in light blue representing the source sequence.
558 The length of the conversion tract is determined from the first and last nucleotide
559 coordinate of the detected tract.
560 Figure 3. Example of a gene conversion map. A. The locations of non-outlier (black) and outlier
561 genes (blue) are shown in this section of the Muller C nucleotide sequence. B. The
562 positions of mapped gene conversion tracts are shown on the Muller C nucleotide
563 sequence. C. The gene conversion tracts are converted to a histogram based on the
564 number of tracts that cover each nucleotide of the Muller C sequence.
565 Figure 4. Distribution of the number of conversion tracts within recipient gene arrangement for
566 14 regions along Muller C of *D. pseudoobscura*. Dots underneath the x-axis indicate
567 regions within gene arrangement with a deficiency (blue) or excess of gene conversion
568 tracts based on residuals from a Chi Square test of homogeneity.
569 Figure 5. Distribution of conversion tract lengths within recipient gene arrangement for 14
570 regions along Muller C of *D. pseudoobscura*. Dots underneath the x-axis indicate
571 regions within gene arrangement with a deficiency (blue) or excess of gene conversion
572 tract lengths based on a random permutation test.
573 Figure 6. Number of gene conversion events per strain between pairs of gene arrangements in
574 *D. pseudoobscura* in 14 different syntenic blocks. The five gene arrangements are
575 abbreviated with their two letter codes: AR, Arrowhead; CH, Chiricahua; PP, Pikes Peak;
576 ST, Standard; TL, Tree Line. Histograms show the number of gene conversion events in
577 the 14 syntenic blocks with the maximum value on all scales being 2,500. Blue and pink
578 boxes below each bar indicate regions with a significant deficiency (blue) or excess
579 (pink) of events based on a random permutation test (see the text). The number above
580 each histogram indicate the total number of events across all 14 regions with a blue
581 number indicating a deficiency and a pink number indicating an excess of events. The

582 double headed arrows connect pairs of arrangements with the arrows pointing to the
583 number of events in the recipient arrangement from the donor arrangement. For
584 instance, 1802 conversion events were from ST to TL and 1973 conversion events were
585 from TL to ST. The pink arrows indicate pairs of arrangements that are expected to
586 have frequencies greater than 10% in populations in the Southwestern United States.
587 Gray shading is provided to delineate recipient data for each arrangement.
588 Figure 7. Mean gene conversion tract coverage in 2,668 transcripts across Muller C in *D.*
589 *pseudoobscura* in five gene arrangements. Coverage is the number of times that a
590 nucleotide is covered by a gene conversion tract. Coverage was estimated for each
591 exon within a transcript and the mean coverage for the entire transcript was averaged
592 over all exons. Outlier genes are shown with an orange marker while non-outliers are
593 shown with a gray marker. The location of the derived inversion giving rise to each gene
594 arrangement are shown with a black bar.

Figure 1


Inversion Network
CU
TL
CH
SC
HY
ST
AR
PP
Tree Line (TL)
Cuernavaca (CU)
Chiricahua (CH)
Santa Cruz (SC)
Hypothetical (HY)
Standard (ST)
Pikes Peak (PP)
Arrowhead (AR)
SC → CH
SC → TL
SC → CU
HY → SC
HY → ST
ST → AR
ST → PP
1
2
3
4
5
6
7
8
9
10
11
12
13
14

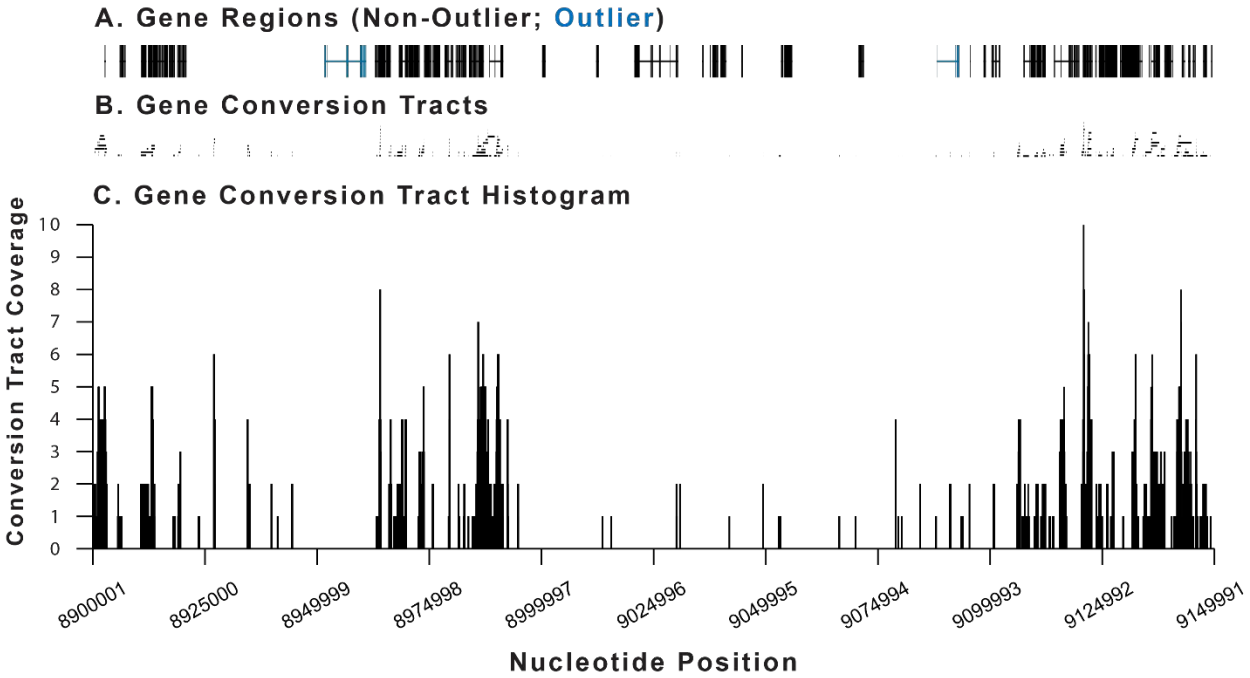
598 **Figure 2.**



599

600

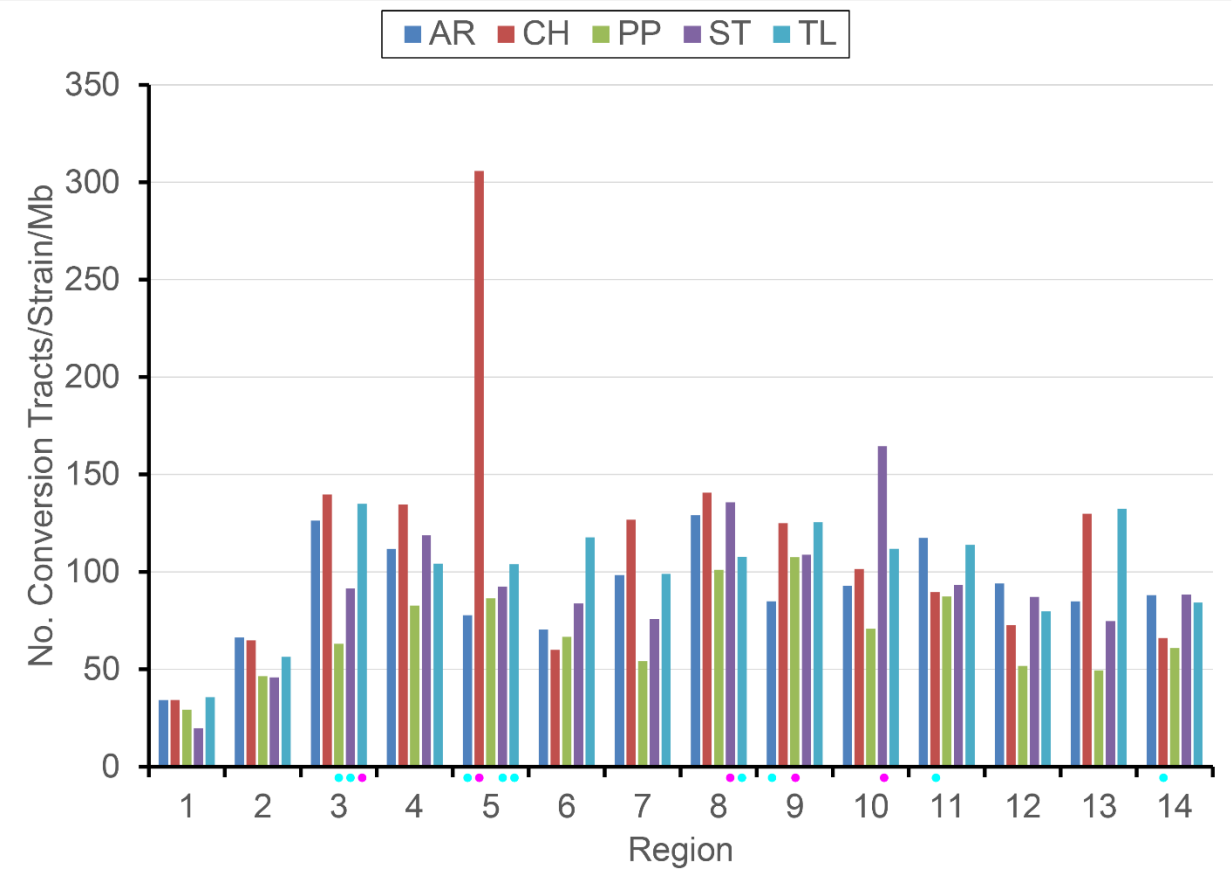
601 **Figure 3**



602

603

604 **Figure 4**



605

606

Figure 5

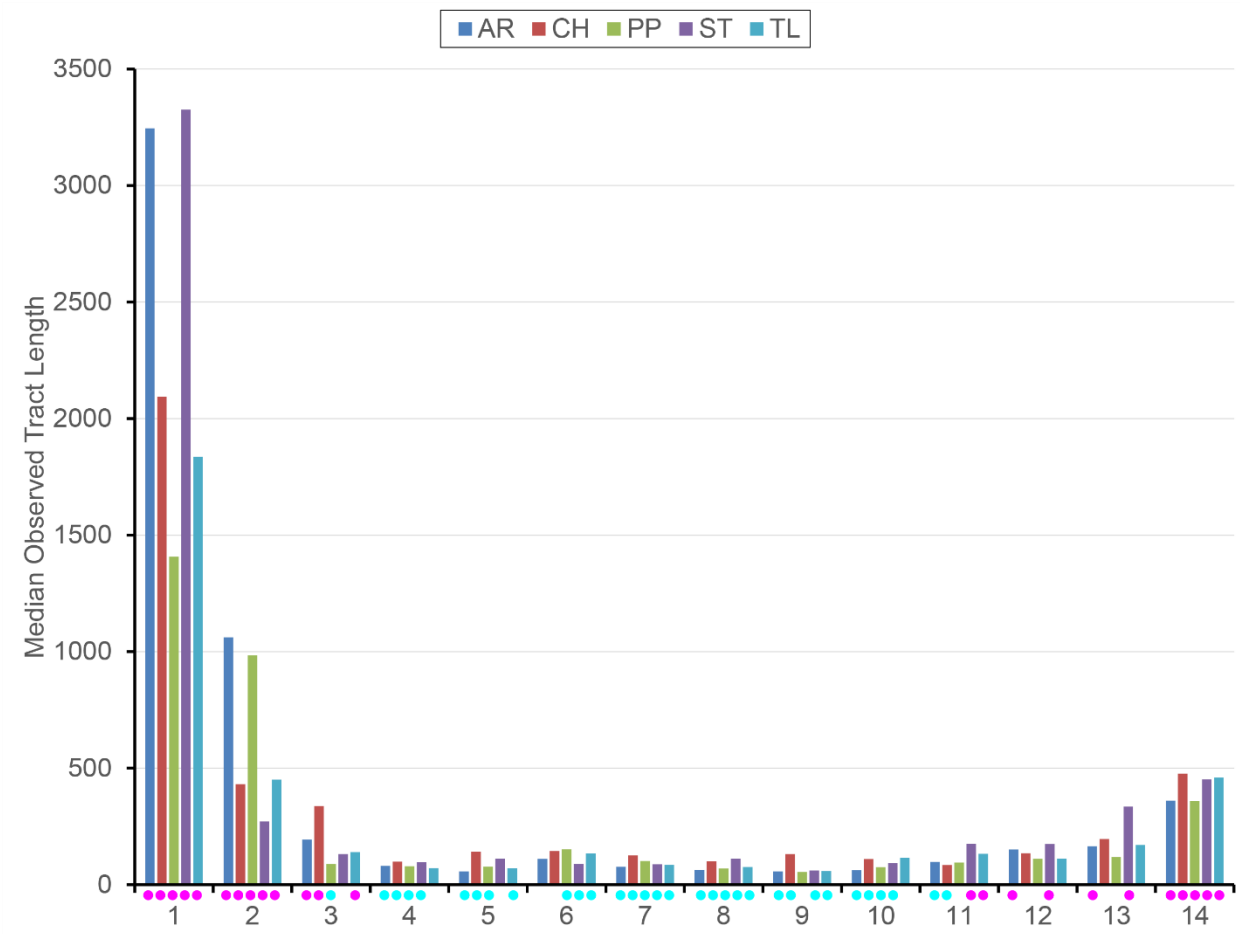

AR
CH
PP
ST
TL
Median Observed Tract Length
3500
3000
2500
2000
1500
1000
500
0
1
2
3
4
5
6
7
8
9
10
11
12
13
14

Figure 6

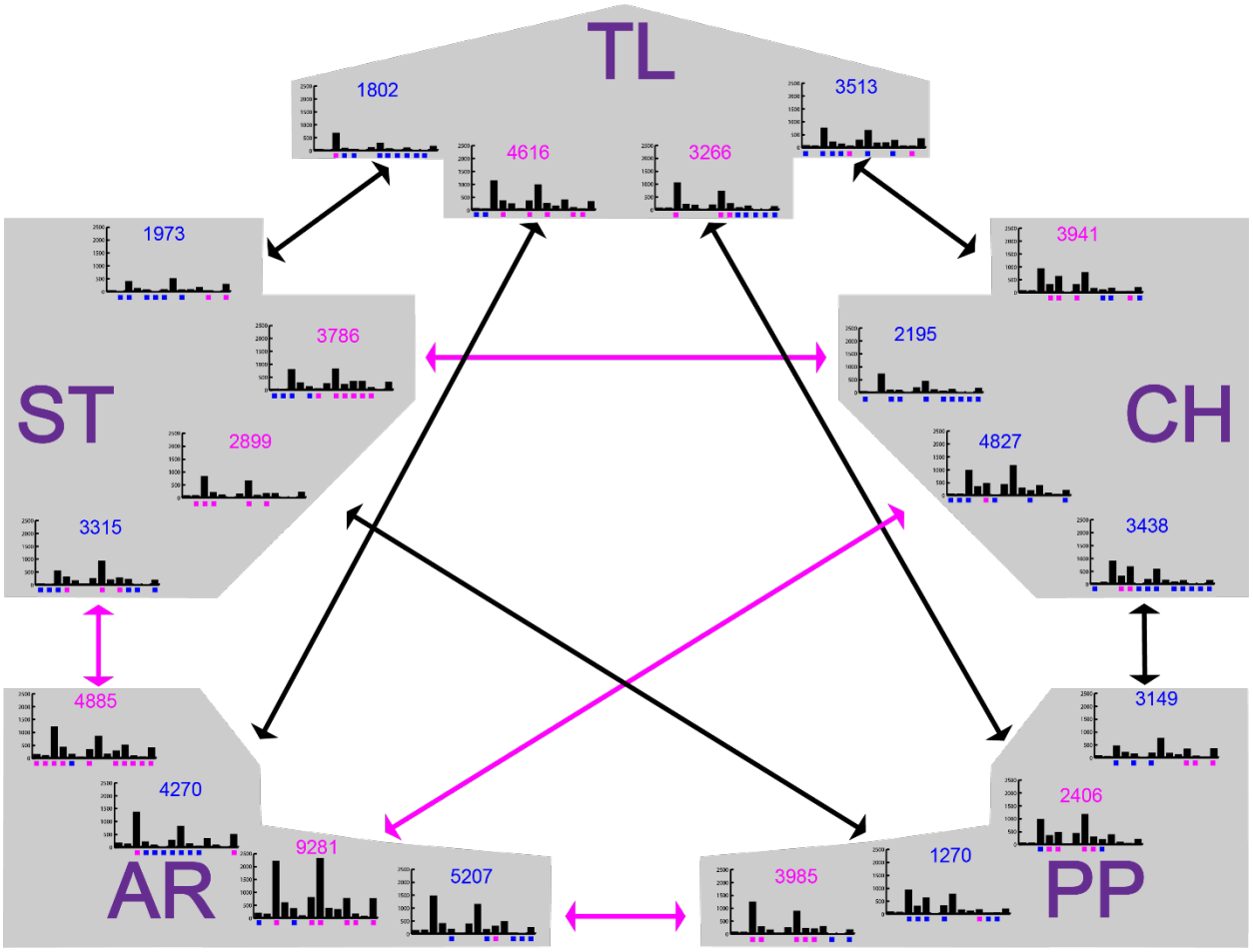
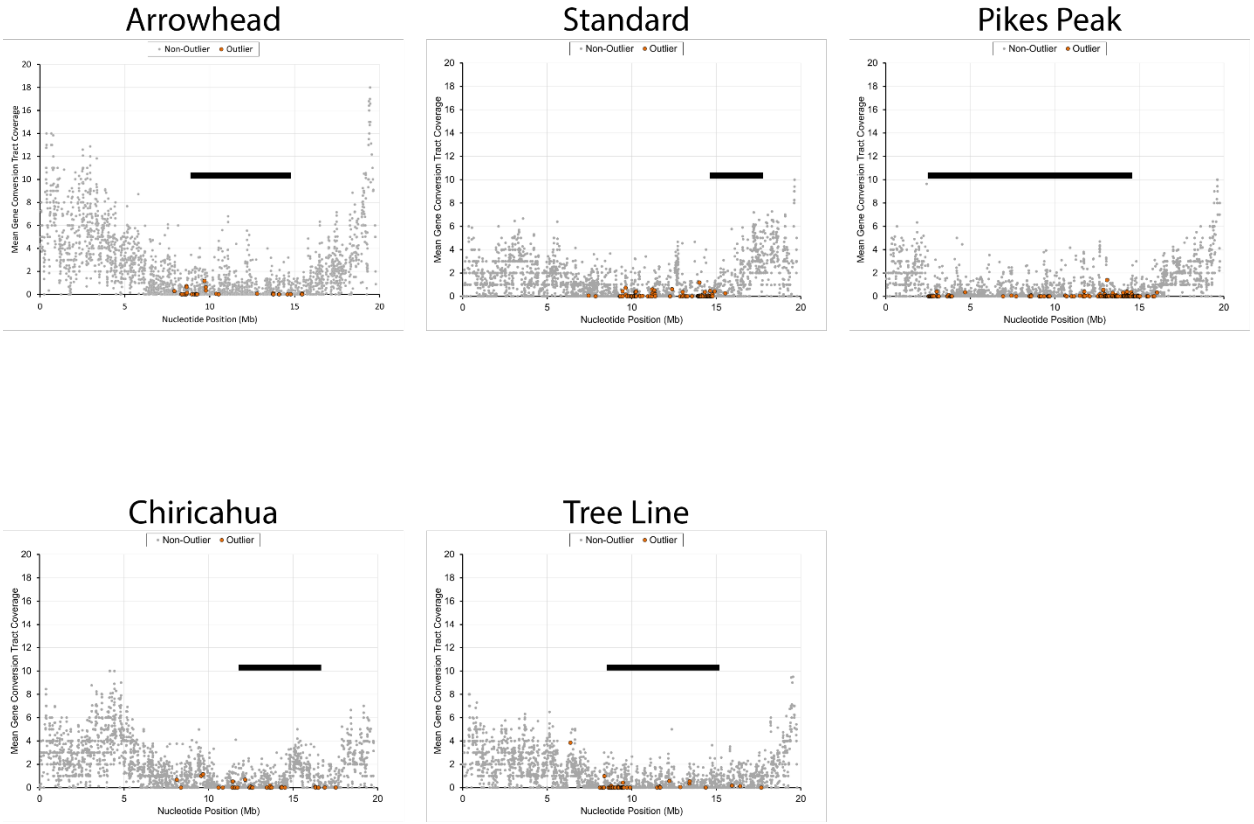

TL
1802
4616
3266
3513
ST
1973
3786
2899
3315
CH
3941
2195
4827
3438
AR
4885
4270
9281
5207
PP
3149
2406
1270
3985

Figure 7


Arrowhead
Standard
Pikes Peak
Chiricahua
Tree Line
Non-Outlier
Outlier
Mean Gene Conversion Tract Coverage
Nucleotide Position (Mb)

Literature Cited

Anderson, W. W., J. Arnold, D. G. Baldwin, A. T. Beckenbach, C. J. Brown *et al.*, 1991 Four decades of inversion polymorphism in *Drosophila pseudoobscura*. Proceedings of the National Academy of Sciences USA 88**:** 10367-10371.

Aquadro, C. F., A. L. Weaver, S. W. Schaeffer and W. W. Anderson, 1991 Molecular evolution of inversions in *Drosophila pseudoobscura* : The amylase gene region. Proceedings of the National Academy of Sciences USA 88**:** 305-309.

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

Chan, A. H., P. A. Jenkins and Y. S. Song, 2012 Genome-Wide Fine-Scale Recombination Rate Variation in Drosophila melanogaster. PLOS Genetics 8**:** e1003090.

Charlesworth, B., and D. Charlesworth, 1973 Selection of new inversion in multi-locus genetic systems. Genetical Research 21**:** 167-183.

Chovnick, A., 1973 Gene conversion and transfer of genetic information within the inverted region of inversion heterozygotes. Genetics 75**:** 123-131.

Crown, K. N., D. E. Miller, J. Sekelsky and R. S. Hawley, 2018 Local Inversion Heterozygosity Alters Recombination throughout the Genome. Current Biology 28**:** 2984-2990.

Dobzhansky, T., 1944 Chromosomal races in *Drosophila pseudoobscura* and *Drosophila persimilis*. Carnegie Inst. Washington Publ. 554**:** 47-144.

Dobzhansky, T., 1948 Genetics of natural populations. XVI. Altitudinal and seasonal changes in certain populations of *Drosophila pseudoobscura* and *Drosophila persimilis*. Genetics 33**:** 158-176.

Dobzhansky, T., and M. L. Queal, 1938 Genetics of natural populations. II. Genic variation in populations of *Drosophila pseudoobscura* inhabiting isolated mountain ranges. Genetics 23**:** 463-484.

Dobzhansky, T., and A. H. Sturtevant, 1938 Inversions in the chromosomes of *Drosophila pseudoobscura*. Genetics 23**:** 28-64.

Fuller, Z. L., G. D. Haynes, S. Richards and S. W. Schaeffer, 2017 Genomics of natural populations: Evolutionary forces that establish and maintain gene arrangements in *Drosophila pseudoobscura*. Molecular Ecology 26**:** 6539-6562.

Fuller, Z. L., S. A. Koury, N. Phadnis and S. W. Schaeffer, 2019 How chromosomal rearrangements shape adaptation and speciation: Case studies in *Drosophila pseudoobscura* and its sibling species *Drosophila persimilis*. Molecular Ecology 28**:** 1283-1301.

Gong, W. J., K. S. McKim and R. S. Hawley, 2005 All Paired Up with No Place to Go: Pairing, Synapsis, and DSB Formation in a Balancer Heterozygote. PLOS Genetics 1**:** e67.

Guerrero, R. F., and M. Kirkpatrick, 2014 Local adaptation and the evolution of chromosome fusions. Evolution 68**:** 2747-2756.

Guerrero, R. F., F. Rousset and M. Kirkpatrick, 2012 Coalescent patterns for chromosomal inversions in divergent populations. Philosophical Transactions of the Royal Society B: Biological Sciences 367**:** 430-438.

Kirkpatrick, M., and N. Barton, 2006 Chromosome inversions, local adaptation and speciation. Genetics 173**:** 419-434.

Korunes, K. L., and M. A. F. Noor, 2019 Pervasive gene conversion in chromosomal inversion heterozygotes. Mol Ecol 28**:** 1302-1315.

Lindsley, D. L., and L. Sandler, 1977 The genetic analysis of meiosis in female *Drosophila melanogaster*. Philosophical Transactions of the Royal Society of London Biology 277B**:** 295-312.

Lobeck, A. K., 1948 Physiographic diagram of North America, pp. C. S. Hammond & Co., Maplewood, NJ.

666 Maynard Smith, J., and J. Haigh, 1974 The hitch-hiking effect of a favorable gene. Genetical
667 Research 23**:** 23-35.
668 Miller, D. E., C. B. Smith, N. Y. Kazemi, A. J. Cockrell, A. V. Arvanitakas *et al.*, 2016 Whole-
669 genome analysis of individual meiotic events in *Drosophila melanogaster* reveals That
670 noncrossover gene conversions are insensitive to interference and the centromere
671 effect. Genetics 203**:** 159-171.
672 Miller, D. E., S. Takeo, K. Nandanan, A. Paulson, M. M. Gogol *et al.*, 2012 A Whole-
673 Chromosome Analysis of Meiotic Recombination in Drosophila melanogaster. G3
674 Genes|Genomes|Genetics 2**:** 249-260.
675 Muller, H. J., 1940 Bearings of the 'Drosophila' work on systematics., pp. 185-268 in *The New*
676 *Systematics*, edited by J. Huxley. Clarendon Press, Oxford.
677 Munte, A., J. Rozas, M. Aguade and C. Segarra, 2005 Chromosomal inversion polymorphism
678 leads to extensive genetic structure: a multilocus survey in *Drosophila subobscura*.
679 Genetics 169**:** 1573-1581.
680 Navarro, A., A. Barbadilla and A. Ruiz, 2000 Effect of inversion polymorphism on the neutral
681 nucleotide variability of linked chromosomal regions in *Drosophila*. Genetics 155**:** 685-
682 698.
683 Navarro, A., E. Betrán, A. Barbadilla and A. Ruiz, 1997 Recombination and gene flux caused by
684 gene conversion and crossing over in inversion heterokaryotypes. Genetics 146**:** 695-
685 709.
686 Nei, M., K. I. Kojima and H. E. Schaffer, 1967 Frequency changes of new inversions in
687 populations under mutation-selection equilibria. Genetics 57**:** 741-750.
688 Noor, M. A., D. A. Garfield, S. W. Schaeffer and C. A. Machado, 2007 Divergence between the
689 *Drosophila pseudoobscura* and *D. persimilis* genome sequences in relation to
690 chromosomal inversions. Genetics 177**:** 1417–1428.
691 Ranz, J. M., D. Maurin, Y. S. Chan, M. v. Grotthuss, L. W. Hillier *et al.*, 2007 Principles of
692 genome evolution in the *Drosophila melanogaster* species group Public Library of
693 Science Biology 5**:** 1366-1381.
694 Richards, S., Y. Liu, B. R. Bettencourt, P. Hradecky, S. Letovsky *et al.*, 2005 Comparative
695 genome sequencing of *Drosophila pseudoobscura*: Chromosomal, gene and *cis*-
696 element evolution. Genome Research 15**:** 1-18.
697 Rozas, J., and M. Aguadé, 1993 Transfer of genetic information in the *rp49* region of *Drosophila*
698 *subobscura* between different gene arrangements. Proceedings of the National
699 Academy of Sciences USA 90**:** 8083-8087.
700 Rozas, J., and M. Aguadé, 1994 Gene conversion is involved in the transfer of genetic
701 information between naturally occurring inversions of *Drosophila*. Proceedings of the
702 National Academy of Sciences USA 91**:** 11517-11521.
703 Rozas, J., A. Ferrer-Mata, J. C. Sánchez-DelBarrio, S. Guirao-Rico, P. Librado *et al.*, 2017
704 DnaSP 6: DNA Sequence Polymorphism Analysis of Large Data Sets. Molecular Biology
705 and Evolution 34**:** 3299-3302.
706 Schaeffer, S. W., and W. W. Anderson, 2005 Mechanisms of genetic exchange within the
707 chromosomal inversions of *Drosophila pseudoobscura*. Genetics 171**:** 1729-1739.
708 Shriver, M., G. Kennedy, E. Parra, H. Lawson, V. Sonpar *et al.*, 2004 The genomic distribution
709 of population substructure in four populations using 8,525 autosomal SNPs. Human
710 Genomics 1**:** 274 - 286.
711 Sperlich, D., and P. Pfriem, 1986 Chromosomal polymorphism in natural and experimental
712 populations, pp. 257-309 in *The Genetics and Biology of Drosophila 3e*, edited by M.
713 Ashburner, H. L. Carson and J. N. Thomson. Academic Press, New York, NY.
714 Stephens, J. C., 1985 Statistical methods of DNA sequence analysis: Detection of intragenic
715 recombination or gene conversion. Molecular Biology and Evolution 2**:** 539-556.

716 Sturtevant, A. H., and G. W. Beadle, 1936 The relations of inversions in the X chromosome of
717 *Drosophila melanogaster* to crossing over and disjunction. Genetics 21**:** 544-604.
718 Wallace, A. G., D. Detweiler and S. W. Schaeffer, 2011 Evolutionary history of the third
719 chromosome gene arrangements of *Drosophila pseudoobscura* inferred from inversion
720 breakpoints. Molecular Biology and Evolution 28**:** 2219-2229.
721 Wright, D., and S. W. Schaeffer, 2022 The relevance of chromatin architecture to genome
722 rearrangements in *Drosophila*. Philosophical Transactions of the Royal Society B:
723 Biological Sciences 377**:** 20210206.
724 Yi, X., Y. Liang, E. Huerta-Sanchez, X. Jin, Z. X. P. Cuo *et al.*, 2010 Sequencing of 50 human
725 exomes reveals adaptation to high altitude. Science 329**:** 75-78.

726