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, 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 or QTLs map 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 that contribute to the establishment of the rearrangement. 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 mapped. This study uses the inversion polymorphism of *Drosophila pseudoobscura* as a model system to show that gene conversion analysis is a useful tool for mapping selected 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 of five *D. pseudoobscura* gene arrangements identified outlier genes within inverted regions based on local clusters of fixed SNP differences. These outlier genes also had potentially heritable 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 or 7.2 x coverage of the 19.7 Mb chromosome. We estimated gene conversion tract coverage in the 2,668 genes on Muller C and tested whether the number of genes with significantly low coverage was the same for outlier versus non-outlier loci. . Genes with low gene conversion tract coverage were more frequent in the outlier class than the non-outlier class suggesting that selection removes exchanged DNA from the outlier genes more often than non-outlier genes. These data support the hypothesis that the pattern and organization of genetic diversity on the third chromosome in *D. pseudoobscura* is consistent with the capture of locally adapted combinations of alleles prior to inversion mutation events.

Introduction

Chromosomal inversions were first observed indirectly as recombination modifiers in *Drosophila melanogaster* (STURTEVANT 1917) and later demonstrated to be the result of the reordering of genes along chromosomes (STURTEVANT 1921). An inversion results from two breaks along a chromosome and the rejoining of the central segment in reverse order. The direct examination of polytene chromosomes from Dipteran larval salivary glands revealed that many species are segregating for inversion polymorphisms on one or more chromosomal arms (DOBZHANSKY AND STURTEVANT 1938; WASSERMAN 1982; SPERLICH AND PFRIEM 1986; CARSON 1992; COLUZZI *et al.* 2002). An outstanding question in evolutionary genetics has been whether neutral or selective forces are responsible for establishing inversions in populations (BERDAN *et al.* 2023). Quantitative trait loci (QTL) and genes that underlie ecotypes have been found to map within inversions or supergenes in a wide variety of species (LOWRY AND WILLIS 2010; HUYNH *et al.* 2011; KÜPPER *et al.* 2015; MÉROT *et al.* 2018; SINCLAIR-WATERS *et al.* 2018; AROSTEGUI *et al.* 2019; AYALA *et al.* 2019; FARIA *et al.* 2019; LUCEK *et al.* 2019; RENNISON *et al.* 2019; HAN *et al.* 2020; HUANG *et al.* 2020; YAN *et al.* 2020; FUNK *et al.* 2021; HAGER *et al.* 2022; HARRINGMEYER AND HOEKSTRA 2022). Indirect evidence such as clines, seasonal cycling, and population cage experiments have inferred that inversion polymorphisms have been maintained by selection (DOBZHANSKY 1944; WRIGHT AND DOBZHANSKY 1946; DOBZHANSKY 1948; DOBZHANSKY 1950; COLUZZI *et al.* 1979; KAPUN *et al.* 2016; WESTRAM *et al.* 2018). In Drosophila, a number of phenotypic characters have been associated with inversion polymorphisms including mating speed, body size, temperature resistance, wing length, and thorax length (SPIESS AND LANGER 1964; HOFFMANN AND WEEKS 2007; REGO *et al.* 2010). In addition, different karyotypes have differentiation survival under different environmental conditions such as humidity, temperature, food sources (HEUTS 1947; HEUTS 1948; CUNHA 1951; LEVINE 1952; BATTAGLIA AND SMITH 1961; DRUGER 1962; DOLGOVA *et al.* 2010). What is not clear is what genes lead to the establishment of new chromosomal inversions. Identifying candidate selected genes underlying QTLs and phenotypes within different arrangements have remained elusive because inversions suppress recombination leading to extensive linkage disequilibrium within inverted regions that can mask signatures of selection.

Genomic re-sequencing data have provided insights into the mechanisms for the origin, establishment, and maintenance of chromosomal inversions in natural populations (For a recent review see, BERDAN *et al.* 2023). Comparison of sequences between different gene arrangements have 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; HARRINGMEYER AND HOEKSTRA 2022). Ectopic exchange between repeat sequences at breakpoints provide one model for the facilitation of inversion mutations (FULLER *et al.* 2017; HARRINGMEYER AND HOEKSTRA 2022), while staggered cuts that duplicate genes at breakpoints is an alternative mechanism observed to generate new gene arrangements (RANZ *et al.* 2007; PUERMA *et al.* 2016).

Once an inversion mutation arises in the population, it is usually lost because its initial frequency (1/2N) is quite low. On the other hand, a 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 and over time inter-arrangement divergence will be greatest at the breakpoints because the homogenizing effect of genetic flux among arrangements is greatest in the central regions of the inverted regions (NAVARRO *et al.* 2000; GUERRERO *et al.* 2012). There are several possible mechanisms that can explain how an inversion is established by selection. First, direct or position effects (SPERLICH

AND PFRIEM 1986) may create beneficial variation when the breakpoints either disrupt genes, create new hybrid genes, or alter gene expression of boundary genes at or near the lesions. Empirical studies of inversion breakpoints have shown that gene duplications can occur at breakpoints (RANZ *et al.* 2007), but have failed to find evidence for gene disruptions within genes (FULLER *et al.* 2017). Inversion breakpoints can altered gene expression of boundary genes (PUIG *et al.* 2002). Recent analyses have shown that chromatin states and architecture may influence where breakpoints occur (MCBROOME *et al.* 2020; WRIGHT AND SCHAEFFER 2022).

A second mechanism proposes that indirect effects associated with the suppression of recombination can maintain advantageous haplotypes within the inverted region leading to elevated linkage disequilibrium among multiple genes within the inverted segment (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). In the case of the local adaptation model, inversions arise to counter the formation of maladaptive genotypes formed through recombination of alternatively selected chromosomes from different habitats. This typically occurs when the migration parameter (*Nm*) is > 1 and populations exist in a heterogeneous environment (DOBZHANSKY 1944; KENNINGTON *et al.* 2006; SCHAEFFER 2008; CHENG *et al.* 2012; KENNINGTON AND HOFFMANN 2013; KAPUN *et al.* 2016). The size of an inversion may also influence whether it is established based on the number of genes involved in local adaptation (KIRKPATRICK AND BARTON 2006; FULLER *et al.* 2017). The size of an inversion and its selective role will be a trade off between the number of selected genes versus the ability for double cross overs to reduce associations among selected genes (CACERES *et al.* 1999). Finally, a new gene arrangement may rise to intermediate frequency because one beneficial allele is trapped within the inverted region (MAYNARD SMITH AND HAIGH 1974) and arrangements can later accumulate additional selected genes .

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 if two cross overs occur within the inverted region, but the probability of double cross over events is small unless the size of the inverted segment is large (NAVARRO *et al.* 1997). Gene conversion events, on the other hand, are genetic exchanges at double strand breaks that are resolved as non-cross over events 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) (Figure 1). Conversion happens when the DNA strand from one homologous chromosome (pink strand in Figure 1C) invades the other homolog creating heteroduplexes everywhere where there are SNP differences. Repair of the heteroduplexes in favor of the invading DNA strand leads to the gene conversion event (Figure 1 D). Gene conversion events occur uniformly across the inverted segment, but the homogenizing effect is limited because of the small DNA segment 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

155 have a greater homogenizing effect than gene conversion because larger genomic segments
156 are exchanged, but crossing over is less likely to occur relative to gene conversion events.
157 Fuller *et al*. (2017) analyzed complete genome sequences of the third chromosome
158 inversion polymorphism of *D. pseudoobscura* as a model system to determine mechanisms
159 responsible for the establishment of inversions in populations. The third chromosome or Muller
160 C (MULLER 1940) of *D. pseudoobscura* has a wealth of gene arrangements that were generated
161 through a series of overlapping inversion events (DOBZHANSKY AND STURTEVANT 1938). These
162 arrangements have been suggested to be targets of selection based on stable clines across
163 heterogenous environments (DOBZHANSKY 1944; ANDERSON *et al.* 1991), stable altitudinal
164 gradients (DOBZHANSKY 1948), and seasonal cycling (DOBZHANSKY 1948). Aquadro *et al*.
165 (1991) used phylogenetic analysis of restriction fragment length polymorphisms to support the
166 unique origin hypothesis of the *D. pseudoobscura* gene arrangements. Fuller *et al*. (2017)
167 tested 2,668 Muller C genes within each of six arrangements for evidence of selection. Loci
168 with significant elevated frequencies of derived nucleotide variants and with significantly long
169 population specific branch lengths (SHRIVER *et al.* 2004; YI *et al.* 2010) were identified as outlier
170 genes. In addition, outlier genes that were differentially expressed or had fixed inversion
171 specific amino acids among arrangements were considered potential targets of selection.
172 These outlier loci may reflect regions enriched for nucleotide variation originally associated with
173 the new inversion event that has yet to be decoupled by genetic flux. Alternatively, genetic flux
174 in the form of gene conversion may be sufficient to homogenize genetic diversity, but purifying
175 selection has removed deleterious conversion tracts.
176 Direct empirical estimates of gene conversion rates in inverted and colinear regions of *D.*
177 *pseudoobscura* have been estimated to be 3.1×10^{-5} and 7.7×10^{-6} converted nucleotides per
178 base per generation, respectively (KORUNES AND NOOR 2019). Korunes and Noor (2019)
179 generated genome wide SNP heterozygotes and used computational methods to infer gene
180 conversion events on four of the five major chromosomal arms. In the cross, the third
181 chromosome was heterozygous for the ST and PP arrangements. Ten F_2 offspring were
182 sequenced for this cross and 11 and 23 gene conversion events were scored within inverted
183 and colinear regions respectively. Identical events were scored in the cross and were
184 presumed to occur during mitosis prior to meiosis (KORUNES AND NOOR 2019). If only unique
185 events are scored, then three and eight gene conversion tracts were observed in inverted and
186 colinear regions, respectively, which reduces the estimates of gene conversion rates in inverted
187 and colinear regions to 8.5×10^{-6} and 2.4×10^{-6} converted nucleotides per base per generation,
188 respectively. These estimates of gene conversion rates are two orders of magnitude greater
189 than the mutation rate, which would allow gene conversion to homogenize genetic diversity
190 among arrangements. Korunes and Noor (2019) show that gene conversion events can be
191 detected after a single generation, but do not provide insights about the selective effects of
192 these events, which might be inferred over longer evolutionary time.
193 A general computational method to detect gene conversion events indirectly was
194 developed based on the configuration of polymorphic nucleotides (phylogenetic partitions) at
195 multiple sites across a sample of nucleotide sequences (STEPHENS 1985). This method was
196 refined to detect gene conversion events in inversion systems when differentiated DNA
197 segments are exchanged between one arrangement and another (BETRAN *et al.* 1997). Figure
198 2 shows an example of two detected gene conversion events on Muller C of *D. pseudoobscura*.
199 The Betran *et al*. (1997) method assumes that the frequency of an informative variant is greater
200 in the source arrangement than in the recipient. Two probabilities are used to infer gene
201 conversion tracts from samples of gene arrangement sequence data. The first is the probability

202 that a SNP is informative of a conversion event based on the frequencies of a variant within the
203 source and recipient subpopulations or ψ. The second is the probability that a gene conversion
204 event is extended to an additional nucleotide or ϕ. This method was applied by Schaeffer and
205 Anderson (2005) to gene arrangement samples of short aligned sequences (339 to 517 bp)
206 from seven gene regions of *D. pseudoobscura*. They estimated the mean conversion tract
207 length to be 205 bp and mean population conversion rate to be 3.4×10^{-6}, which is comparable
208 to the rates inferred by Korunes and Noor (2019).
209 Under a neutral model and an infinite amount of time, maximal nucleotide divergence will
210 be found at inversion breakpoints with the central inverted region showing less divergence
211 among different arrangements (NAVARRO *et al.* 2000; NOOR *et al.* 2007; GUERRERO *et al.* 2012).
212 When an inversion captures locally selected alleles, selection opposes the homogenizing effect
213 of genetic flux leading to local peaks of differentiation (GUERRERO *et al.* 2012). The observation
214 of peaks of differentiation will become more apparent as time progresses and gene conversion
215 events homogenize variation among arrangements except at selected loci.
216 Five *D. pseudoobscura* arrangements had sufficient sample sizes to detect multiple
217 outlier loci. These gene arrangements had between 36 and 149 outlier genes that were
218 distributed within and outside the boundaries of the derived inversion (Figure 1). Outlier genes
219 had decreased estimates of the recombination parameter rho (ρ) (CHAN *et al.* 2012) compared
220 to intervening non-outlier loci (see Table 3 in FULLER *et al.* 2017) suggesting that genetic flux is
221 limiting differentiation at a fine scale. The predominant force homogenizing local regions among
222 arrangements, however, remains unclear. The current study infers gene conversion events
223 over the accumulated history of the Muller C inversion polymorphism in *D. pseudoobscura* to
224 test the hypothesis that gene conversion events occur uniformly across all genes in colinear and
225 inverted regions of Muller C in *D. pseudoobscura*. With genome re-sequencing data, we used
226 the Bertran *et al*. (1997) method to do a comprehensive analysis of gene conversion tracts
227 across Muller C of *D. pseudoobscura*. We show here that gene conversion has been sufficient
228 to homogenize genetic variation among arrangements within inverted regions. Previously
229 identified outlier genes tended to have lower coverage from gene conversion events than non-
230 outlier genes. These data provide evidence for the indirect effect of recombination maintaining
231 associations of selected genes originally captured by inversion events. Outlier genes that have
232 low gene conversion coverage represent loci where purifying selection has removed exchanged
233 DNA.
234

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 Arrowhead 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 19.8 Mb third chromosome (Muller C) were used in the gene conversion analysis. The mapping of Illumina reads might be expected to decrease with increasing inversion distance between arrangements, however, the different inverted chromosomes have similar levels of inferred gaps or missing data (2.0 to 2.2 % sites of the 19.8 Mb sequence) and have similar distributions of gaps or missing data across the 73 regions examined in the gene conversion analysis (See Supplemental Figure S1).

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 at 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, but will have lower frequency in the recipient strain. 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 between SNPs be preserved in the data set. The seven pairs of *D. pseudoobscura* inversion breakpoints used in this study divide the third chromosome into 14 regions where genes within the region are in the same order in all arrangements (Figure 3). While conversion events may occur near breakpoints (CROWN *et al.* 2018), they are unlikely to straddle the breakpoints.

The Gene Conversion analysis module 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 181 to 440 kb each, with most 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-PP, AR-CH, AR-TL, AR-CU, ST-PP, ST-CH, ST-TL, ST-CU, PP-CH, PP-TL, PP-CU, CH-TL, CH-CU, and TL-CU). Each analysis excluded nucleotide sites with alignment gaps and the output provided lists of conversion tracts within strain, the beginning and end coordinates of the tract, the tract length excluding gaps, mean value of ψ, number of informative sites, and a list of sites with its associated value of ψ. The detected events are unidirectional in the pairwise analysis. The polarity of events is based on the informative sites being fixed or nearly fixed in the source arrangement and at low or intermediate frequency in the recipient arrangement. For instance, in the comparison of AR and CH, an event detected in the AR recipient strain will have CH as the source arrangement. Results of each of the 1,095 pairwise analyses were stored in separate text files and were also combined into a single file for subsequent summary analyses.

A gene conversion tract can be detected multiple times in this analysis. There are two sources of redundancy. First, a gene conversion tract can be inferred from multiple pairwise comparisons of gene arrangements. For instance, a gene conversion tract in an AR strain might be inferred from a comparison of AR with CH or with TL. Only one instance of this detected event is used. For the analysis of event polarity, only events detected with a single pairwise comparison is used so that we can unambiguously infer the source and recipient. Second, a gene conversion tract can be inferred in multiple strains of the same gene arrangement. It is assumed that gene conversion tracts result from unique events and finding the same event in multiple strains is a result of the tract increasing in frequency within the arrangement (BETRAN *et al.* 1997). Only one tract from a single strain is used when the conversion event is found in multiple strains. The effect of removing redundant gene conversion tracts from the data set is a conservative adjustment that lowers the overall gene conversion tract coverage per site in both outlier and non-outlier genes.

A histogram of gene conversion coverage in each of 2,668 genes was constructed from the set of non-redundant gene conversion tracts (Figure 4). Outlier and non-outlier transcripts were mapped to Muller C (Figure 4A). Next, gene conversion tracts were mapped to Muller C (Figure 4B). Finally, a histogram of gene conversion coverage was constructed by counting the number of tracts that cover each nucleotide across the chromosome (Figure 4C).

The mean coverage per nucleotide was estimated for outlier and non-outlier genes from the gene conversion histogram. The sum of gene conversion tract coverage was estimated across the exons of each gene and divided by the total number of nucleotides in the gene. The mean gene conversion coverage was determined separately for outliers and non-outlier genes.

To determine if the donor and recipient conversion events were independent across the 14 syntenic block regions and overall, we permuted the donor and recipient classifications across the 14 regions. We constrained the permutations so that only non-self-arrangement events occurred. A total of 10,000 random permutations were used. For each permutation of the donors and recipients, the number of gene conversion events was re-estimated and compared with the observed data to determine the probability of obtaining more extreme values as well as the minimum and maximum values found across all permutations.

Maximum Likelihood Estimates of ϕ and Expected Number of Conversion Tracts. We used the maximum likelihood approach of Betran *et al*. (1997) to infer ϕ within each of the 14 regions and for each donor-recipient pair using the observed list of conversion tracts for each pair (See equations 4, 6, and 8). The second derivative of the likelihood equation was used to estimate the asymptotic variance, which is used to infer a lower and upper bound for tract

329 length. We did separate analyses for each region and donor-recipient pair because we
330 assumed that the breakpoints that defined the 14 regions might interfere with the progression of
331 conversion tracts. This may be a conservative assumption because we define the syntenic
332 block regions based on the set of 13 breakpoints of gene arrangements used in this study.
333 **Tests for Differences in Gene Conversion Tract Coverage between Outlier and**
334 **Non-Outlier Genes.** We tested the hypothesis that outlier and non-outlier genes had
335 equivalent gene conversion coverage using several methods. This assumes that gene
336 conversion can occur uniformly across the chromosome.

337 1. ***Random Permutation of Gene Class.*** We used a permutation test where we
338 randomly assigned outlier and non-outlier designations to the genes within the
339 arrangement. A total of 10,000 replicates were used to determine the frequency of
340 random outlier assignments that had a mean tract coverage less than the observed
341 value.
342 2. ***Random Permutation of Observed Gene Conversion Tracts.*** We randomly
343 assigned the observed gene conversion tracts to Muller C without replacement
344 10,000 times and for each random mapping we re-constructed the conversion tract
345 histogram for the outlier and non-outlier genes. The average coverage in outlier and
346 non-outlier genes was determined for each replicate and the frequency of replicates
347 that had a value less than the observed was estimated.
348 3. ***Random Gene Conversion Tracts Based on Maximum Likelihood Parameters.***
349 We used maximum likelihood estimates of region-specific ϕ and the expected
350 number of gene conversion tracts ($E(k_T)$) in the 14 different syntenic block regions to
351 generate a set of conversion tracts. The parameter ϕ is the probability of extending
352 the conversion tract to an additional nucleotide and the process of generating tract
353 lengths is geometric random variable. For each gene conversion tract within a
354 recipient arrangement, we sampled a tract length as a geometric random deviate. If
355 the tract length was greater than the region size, then a new tract length was chosen.
356 Next, a random nucleotide site was chosen from a uniform distribution to begin the
357 gene conversion tract. If the mapped tract spanned two syntenic blocks, a new
358 starting coordinate was chosen. A total of $E(k_T)$ tracts are generated for each region
359 for each of the five donor arrangements, where k_T is the expected number of
360 conversion tracts (See page 95 in BETRAN *et al.* 1997). A total of 10,000 replicates
361 was generated and the mean tract coverage in outlier genes was re-estimated.

362 It is possible that the original classification of Muller C genes into outlier and non-outlier
363 categories may reflect ascertainment bias because they were originally identified based on
364 higher levels of divergence. To overcome this possibility, we classified each gene into one of
365 three categories based on the random permutation minimum and maximum coverage
366 estimates. The three categories are: (1) below the permuted minimum coverage; (2) between
367 the permuted minimum and maximum coverage; and (3) above the permuted maximum
368 coverage. We used a Chi-square test of homogeneity for each arrangement to determine if the
369 distribution of outlier and non-outlier genes into gene coverage classes is similar.

370 **Results**

371 **Numbers and Lengths of Conversion Tracts in 14 Syntenic Block Regions.** There
372 were 98,443 non-redundant gene conversion tracts observed across Muller C. The mean
373 number of tracts per strain is 1930.2. Sample sizes varied among arrangements, which could
374 bias the number of gene conversion tracts detected. The distribution of gene conversion tracts
375 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 (Figure 5). 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 6). 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 7). 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 regional 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.

Of the 98,443 observed gene conversion tracts in recipient arrangements, the source arrangement could be inferred unambiguously in 74,419 events. We observed gene conversion events between all pairs of arrangements across Muller C ranging between 1,270 conversion events from ST to PP and 5,266 conversion events from CH to AR. Figure 8 shows the polarity of the observed gene conversion events (Donor to Recipient) in the 14 syntenic block regions and across all regions. For all gene conversion events across all regions, three pairs showed a deficiency of events in both directions (AR-TL, CH-PP, and ST-TL), one pair showed an excess of events in both directions (AR-ST), one pair showed a deficiency in one direction and an excess in the other direction (PP-ST).. The rest showed either a deficiency or excess in only one 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., the ST to CH transition had a significant overall deficiency (2,195) and ten of 14 regions 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 PP tended to be a donor to AR.

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

425 donor-recipient pairs in each region are in the Supplemental Materials (Supplemental Table S1).
426 These values are used below in tests of gene conversion coverage in outlier genes.
427 **Gene Conversion Tract Coverage in Outlier and Non-Outlier Genes.** The 98,443
428 observed conversion tracts were mapped to Muller C and the mean coverage per nucleotide
429 was estimated for the exons of each transcript. The expected gene conversion tract coverage
430 per nucleotide across Muller C is 1.2 per nucleotide for all arrangements except for PP, which
431 had an expected coverage rate of 0.8 tracts per nucleotide. Thus, at least one conversion tract
432 should occur in each gene across the chromosome (Figure 5).
433 Figure 9 shows the gene conversion tract coverage for outlier and non-outlier genes
434 across Muller C in five gene arrangements of *D. pseudoobscura*. Proximal and distal regions of
435 all arrangements have higher conversion tract coverage than central inverted regions. Outlier
436 genes tend to have lower levels of coverage than non-outliers. The maximum coverage values
437 are sample size dependent. Standard, Chiricahua, Pikes Peak, and Tree Line have sample
438 sizes between 8 and 10 with maximum coverage of 10 tracts/nucleotide. Arrowhead, on the
439 other hand, has a sample size of 15 with a maximum coverage of 18.
440 The mean gene conversion tract coverage estimated for the 2,668 genes on Muller C
441 shows that non-outlier genes have higher mean coverage than outlier genes (Table 1). Gene
442 conversion tract coverage in non-outlier genes matches the expected value for the entire
443 chromosome, while outlier genes have coverage values < 0.1.
444 We used three tests to determine if the lower gene conversion coverage is expected by
445 chance (Table 1). The Random Permutation of Gene Class test found that the observed mean
446 gene conversion coverage in outlier genes was always less than random permutations in all
447 gene arrangements. One factor to consider is that we filtered the conversion tracts to remove
448 redundant tracts. If the redundant tracts were to be included, we would likely see higher
449 coverage both in outlier and non-outlier genes, but still show a significant difference. The
450 Random Permutation of Observed Gene Conversion Tracts test showed that the observed
451 mean coverage in outliers was always less than the replicate permutations. The Random Gene
452 Conversion Tracts Based on Maximum Likelihood Parameters test found that the observed
453 mean conversion tract coverage in outlier genes was less than was found with random
454 conversion tracts. The results of this test should be viewed with caution because the expected
455 number of true tracts is greater than the observed number of tracts. Thus, it is not surprising
456 that we observe higher gene conversion tract coverage in outlier genes than in our replicates.
457 The important result from these three tests is that outlier genes have lower gene conversion
458 coverage than expected.
459 It is possible that the previous classification of genes into outlier and non-outliers may
460 bias our analysis of gene conversion tract coverage because these genes were identified based
461 on high levels of fixed nucleotide mutations, which would have low levels of shared
462 polymorphism that would result from gene conversion. We examined the relationship between
463 the Population Specific Branch Length (PSBL) statistic and gene conversion tract coverage in
464 outlier and non-outlier genes in the five gene arrangements (Supplement Figure S2). Non-
465 outlier genes had gene conversion tract coverage estimates from 0 to 18 while non-outlier
466 genes had tract coverage estimates from 0 to 4. The random permutation analysis provides an
467 overall estimate for the minimum and maximum coverage across all genes. We divided genes
468 up into three classes, (1) below the permuted minimum value, (2) between the permuted
469 minimum and maximum value, and (3) above the permuted maximum value. We did this
470 separately for the outlier and non-outlier genes and used a chi-square test of homogeneity to
471 determine if the distribution into the three classes is the same for outlier and non-outlier genes
472 (See Supplemental Figure S2 A-E and Tables S2 to S6). The Chi-square tests show that the
473 distributions reject homogeneity in the five arrangements (Supplemental Tables S2-S6). For the

five gene arrangements, the outlier class has an excess of genes with gene conversion tract coverage below the permutation minimum (Percent AR = 94.4%; CH=87.5%; PP=92.6%; ST=95.1%; TL=88.9%). Each arrangement had classified outlier genes with coverages between the permuted minimum and maximum coverages (Numbers, AR, 2; CH, 5; PP, 10; ST, 4; TL, 5) and two arrangements had classified outlier genes above the permuted maximum coverage (Numbers, PP, 1; TL, 1), which suggest that some previously identified outliers do experience expected levels of gene conversion.

Gene Conversion and the Site Frequency Spectrum. We examined whether the site frequency spectrum differs in regions with different levels of gene conversion. Gene conversion events would likely introduce maladaptive genetic variants and our analysis suggests that these variants are removed by purifying selection, however, purifying selection does not have as profound effect on the site frequency spectrum as beneficial mutations do. Beneficial mutations will sweep out nucleotide variation likely leading to an excess of rare variants and a significant negative Tajima's D. Purifying selection on deleterious variants will only remove or decrease the frequency of chromosomes that carry the changes. Jackson et al. (2015) found that purifying selection alters the site frequency spectrum of nonsynonymous nucleotides, but did not alter the site frequency spectrum of synonymous sites. Supplemental Figure S3 A-E shows volcano plots of Tajima's D for all genes of the five arrangements with significant and non-significant D indicated in pink or blue for outlier genes. The five arrangements have both significant positive and negative Tajima's *D*s with only the CH having more positive than negative D values.

496 **Discussion**

497 The work presented here shows that gene conversion events among arrangements are
498 readily detectable and serve as a source of genetic flux among the gene arrangements of *D.*
499 *pseudoobscura*. The number of observed conversion tracts tends to be uniformly distributed
500 across Muller C (Figure 6) except for the proximal two syntenic blocks of the chromosome
501 consistent with theory on genetic flux (NAVARRO *et al.* 1997). Fewer conversion tracts may
502 result from the initiation of fewer double stranded breaks in the proximal regions. While regions
503 1 and 2 are closer to the centromere where crossing over is reduced (LINDSLEY AND SANDLER
504 1977), these regions are in euchromatic DNA, which is unlikely to be affected by the
505 centromeric effect. More likely is that the Betran *et al*. (1997) method has lower power to detect
506 gene conversion events in regions 1 and 2. This method relies on informative nucleotide
507 polymorphisms to detect an exchange between two gene arrangements. Because regions 1
508 and 2 are outside most inverted regions, genetic exchange is likely to reduce differentiation and
509 prevent conversion tract detection.

510 The observed median gene conversion tract lengths are largest in the proximal and
511 distal regions of Muller C and are reduced in size in the inverted regions (Figure 7). The
512 detected events in the proximal and distal regions may be the result of cross overs rather than
513 gene conversion tracts. The Betran *et al*. (1997) method uses informative nucleotide sites to
514 infer exchanges between arrangements. The accumulation of genetic differences among gene
515 arrangements can extend beyond the boundaries of the inversion breakpoints providing
516 information about gene exchanges. In this case, the exchange events are longer than what is
517 observed within the inverted regions. The most proximal region (1) has the largest median
518 conversion track length compared to the most distal region (14). This may reflect that region 1
519 is further from the influence of interference caused by inversion breakpoints (GONG *et al.* 2005).
520 Most common arrangement pairs are homosequential across regions 1 to 3 with the exception
521 of the PP arrangement, which has its proximal breakpoint at the boundary of regions 2 and 3.
522 As a result, there is likely to be less interference of inversion breakpoints in the more proximal
523 regions leading to resolution of exchanges as cross overs and longer sequences being
524 exchanged.

525 We expected exchanges among all pairs of gene arrangements because
526 heterokaryotypes for these pairs have frequencies between 0.054 and 0.380 in at least one
527 population of *D. pseudoobscura*. The number of gene conversion events are not positively
528 correlated with heterokaryotypic frequency. We find larger numbers of events between
529 arrangements with lower observed heterokaryotypic frequencies than higher heterokaryotypic
530 frequencies.(ANDERSON *et al.* 1991). A total of 4,888 tracts were observed between CH and
531 PP, which has an observed maximum heterokaryotypic frequency of 0.054 (ANDERSON *et al.*
532 1991), while ST and CH had 4409 observed tracts with an observed heterokaryotypic frequency
533 of 0.280. We also found that in 15 of 20 cases, the number of events from donor to recipients
534 are either lower or greater than expected in a permutation test (Figure 8). For most pairs of
535 arrangements, there is a bias in which arrangement is the donor versus the recipient. For
536 instance, AR is the recipient of more tracts from CH than expected while there is neither a
537 deficiency nor excess of events in the reverse direction. (see heterokaryotype frequencies in
538 Figure 8). The CH and TL arrangements have two and three cases of deficiencies of being gene
539 conversion recipients. The AR, PP, and ST arrangements have both deficiencies and excesses
540 of being gene conversion recipients. The significant deficiencies may reflect selection against
541 exchanged segments of DNA from one background to another especially in regions within
542 inverted regions (Figure 8).

543 Arrangements that receive significant excesses of gene conversion events may reflect
544 positively selected changes. For example, the PP arrangement received an excess of events
545 from the TL arrangement especially in regions 8 and 9, which are cytological regions 74B-70D

546 and 70C-70B (Figure 2 and Figure 8). These two regions are unusual because phylogenetic
547 analysis of SNPs of the 14 regions of Muller C shows that PP and TL share a more recent
548 common ancestor than either does with their respective cytogenetic ancestor (ST for PP and SC
549 for TL) (see Figure 4 in FULLER *et al.* 2017). This suggests that gene conversion may have
550 introduced sequences from the TL to the PP arrangement that were important for the
551 proliferation of the PP arrangement. Further examination of the exact nature of how the TL
552 donated conversion tracts alter genes in the PP arrangement are needed to understand if
553 introduced variation is the target of selection.
554 The rate of gene conversion is sufficient to homogenize the sequences among the gene
555 arrangements, yet one observes differentiation of some genes (outliers) and not others across
556 the inverted segments. Our analysis of gene conversion coverage tracts shows that gene
557 conversion occurs less frequently in outlier versus non-outlier genes. Tests that randomly
558 shuffle the gene conversion tracts show that there are enough events to homogenize the
559 sequences. Each gene arrangement has fewer than 12% of outlier genes without a significant
560 reduction in gene conversion events (see Supplemental Tables S2-S6). The reason we can
561 detect these outlier loci is that we observe differentiation in the face of the homogenizing effect
562 of gene flux, which works in a similar manner as gene flow. This requires that the various gene
563 arrangements accumulate genetic differences and that there is sufficient time for gene
564 conversion events to accumulate. The Muller C inversion polymorphism in *D. pseudoobscura* is
565 estimated to be at least 1 million years old (AQUADRO *et al.* 1991; WALLACE *et al.* 2011) with the
566 AR arrangement being the newest arrangement of the set originating 350,000 years ago.
567 Despite being a young arrangement, our sample of AR chromosomes has over 30 thousand
568 gene conversion events with a mean coverage of 2.33 events per nucleotide and 36 detected
569 outlier genes with a mean coverage of 0.11. A similar pattern of reduced coverage in outlier
570 genes is observed in the other four older arrangements.
571 Essentially what is observed here is an example of survivorship bias. Gene conversion
572 events survive over time in genes within the inverted regions that can tolerate additional genetic
573 diversity, while gene conversion events in selected outlier genes from different arrangements do
574 not survive because they damage the important features of selected alleles within the inversion
575 haplotype. We assume that these outlier genes are in fact genes essential for the particular
576 gene arrangement's adaptive phenotype. Thus, the lack of homogenization of alleles by gene
577 conversion allows one to detect a signal of selected genes within gene arrangements despite
578 cross over suppression in heterokaryotypes.
579 Each of the major gene arrangements examined in *D. pseudoobscura* have evidence for
580 multiple outlier genes that show arrangement specific variation. We show here that the
581 arrangement specific variation in outlier genes is observed despite the homogenizing effect of
582 extensive genetic flux among arrangements driven by gene conversion events through the
583 history of these chromosomes. These data are consistent with a model where the inversions
584 captured sets of genes involved in local adaptation (KIRKPATRICK AND BARTON 2006). We
585 propose that the local populations of *D. pseudoobscura* in different physiographic provinces
586 (DOBZHANSKY 1944; LOBECK 1948) are selected for different locally adapted phenotypes
587 encoded by multiple genes within the arrangements. These polygenic traits were generated
588 through shuffling of allelic variation to generate adaptive combinations in different environments.
589 Adaptive combinations in one environment may not be beneficial in another habitat and are
590 likely to generate maladaptive types through recombination. This creates a setting for
591 chromosomal inversions to establish if they capture combinations of selected genes
592 (CHARLESWORTH AND CHARLESWORTH 1973). The set of selected genes is not likely to be
593 recognized initially because there will be extensive linkage disequilibrium of SNPs across the
594 inverted region from the ancestral arrangement. As the ancestral and derived arrangements
595 diverge in sequence, conditions to recognize selected loci emerge. Gene conversion begins to

homogenize the unselected genes within the inversion and with sufficient time differentiated selected 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 distributed across the inverted regions, which 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 the accumulation of gene conversion events.

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 that add multiple adaptive alleles 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 selected loci within inversions. Selected genes 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-selected 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. The identification of selected genes within inversions is an important step in understanding the molecular basis for what led to the establishment of inversions in populations. In this case, the outlier genes within the *D. pseudoobscura* inversions fall into two broad categories, perception and detoxification genes (FULLER *et al.* 2017). Mapping gene conversion events provides an important tool for discriminating selected from non-selected genes within inverted regions of the genome.

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.

641

642 Table 1. Gene Conversion Tract Coverage of Outlier and Non-Outlier Genes on Muller C of
643 *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

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

650 **Figure Legends**

651 Figure 1. Gene conversion events generated in a heterozygote for a paracentric inversion. A.
652 Two different gene arrangements with a paracentric inversion between the F and P
653 genes. B. Pairing of these two inversions during meiosis showing the initiation of
654 exchange at a double strand break (DSB within the Box). C. Exchange of strands
655 between the two homologs via the exchange of the pink strand into the double helix of
656 the blue strand, the so-called Holliday junction. Nucleotide mismatches are formed in
657 this structure as DNA strands are exchanged between the homologs. The Holliday
658 junction can be resolved either by breaking strands indicated by the green arrows
659 leading to cross over events or by the orange arrows leading to non-cross over events
660 and potential gene conversion tracts. D. The resolution of heteroduplex sequences
661 because of repair. If the blue strand in the heteroduplex is used as the template for
662 repair, then gene conversion does not occur. In this case, the pink strand was used as
663 the template for repair leading to gene conversion event.
664 Figure 2 Detection of gene conversion. A map of SNP polymorphism is shown for a short
665 region of the third chromosome of *D. pseudoobscura*. Each column represents a SNP
666 with gray boxes matching the reference genome in the first row. A nucleotide that differs
667 from the reference genome is represented in one of four colors (see legend on the right-
668 hand side). Multi-site genotypes are shown for six gene arrangements along with their
669 sample sizes: Arrowhead (AR), Standard (ST), Pikes Peak (PP), Chiricahua (CH), Tree
670 Line (TL), Cuernavaca (CU), and the outgroup *D. miranda*. Two inferred gene
671 conversion events are shown with the black outline box indicating the recipient sequence
672 and the light blue outline boxes shown in light blue representing the source sequence.
673 The length of the conversion tract is determined from the first and last nucleotide
674 coordinate of the detected tract.
675 Figure 3. Network of Drosophila pseudoobscura gene arrangements. The network of gene
676 arrangements (HY, ST, AR, PP, SC, CU, TL, and CH) and the inversion events (colored
677 arrows) are shown to the left where the HY arrangement is the ancestral arrangement.
678 One the right side are the cytogenetic maps for the gene arrangements along with the
679 inverted segments indicated by the colored brackets. The color-coded lines below each
680 arrangement show the locations of the different inversion breakpoints. Outlier genes in
681 each of the gene arrangements are indicated by a black dot. An outlier gene had a
682 significant elevation of derived alleles or population specific branch lengths and had
683 either fixed amino acid differences or differential expression. Below the AR
684 arrangement, 14 regions where genes are in the same order among all chromosomes in
685 the set. The dotted line below the CH, ST, PP, and AR arrangements indicates the
686 regions that would be inverted in ST/AR, ST/CH, AR/CH, or AR/PP heterozygotes
687 whose frequency is at least 10% at least one locality in the Southwestern United States
688 (Dobzhansky 1944).
689 Figure 4. Example of a gene conversion map from the AR chromosome Coordinates 8,900,000
690 – 8,993,000. A. The locations of non-outlier (black) and outlier genes (blue) are shown in
691 this section of the Muller C nucleotide sequence. B. The positions of mapped gene
692 conversion tracts are shown on the Muller C nucleotide sequence. C. The gene
693 conversion tracts are converted to a histogram based on the number of tracts that cover
694 each nucleotide of the Muller C sequence.
695 Figure 5. Gene conversion tract information in five Muller C gene arrangements in D.
696 pseudoobscura. A. Observed and expected numbers of gene conversion tracts where

the expected number = Number of Observed/[1-Probability(Unobserved Gene Conserved Events)]. B. Observed and expected mean number of gene conversion tracts per strain where the number of strains is shown in parentheses next to the gene arrangement name on the x-axis. C. Gene conversion tract coverage per nucleotide for the five gene arrangements.

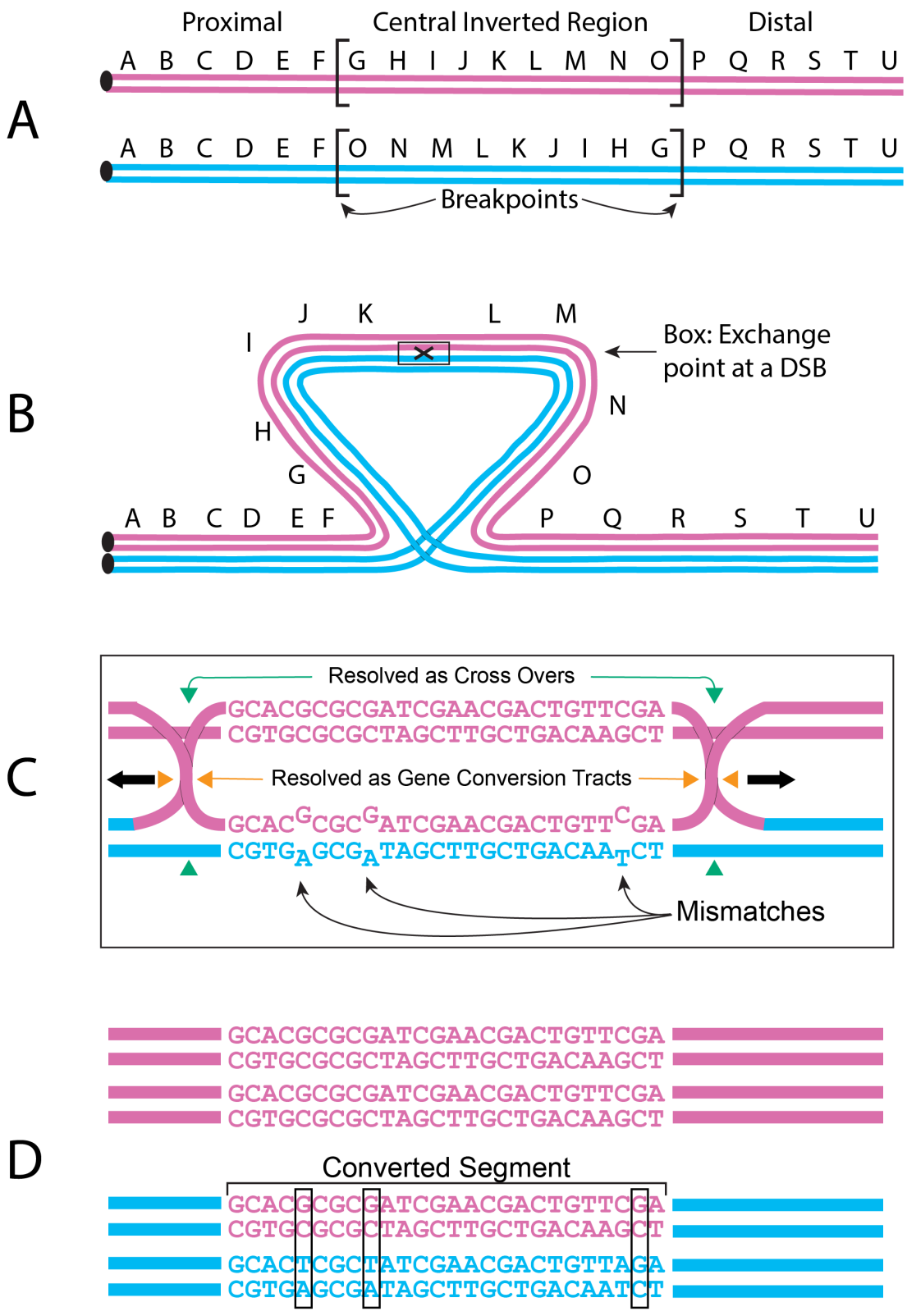
Figure 6. Distribution of the number of conversion tracts within recipient gene arrangement for 14 regions along Muller C of D. pseudoobscura. Dots underneath the x-axis indicate regions within gene arrangement with a deficiency (blue) or excess of gene conversion tracts based on residuals from a Chi Square test of homogeneity. The bar underneath the x-axis shows the relative sizes of the different regions.

Figure 7. Distribution of conversion tract lengths within recipient gene arrangement for 14 regions along Muller C of *D. pseudoobscura*. Dots underneath the x-axis indicate regions within gene arrangement where gene conversion tract lengths are shorter (blue) or longer (pink) than expected based on a random permutation test. The bar underneath the x-axis shows the relative sizes of the different regions.

Figure 8. Number of gene conversion events per strain between pairs of gene arrangements in *D. pseudoobscura* in 14 different syntenic blocks. The five gene arrangements are abbreviated with their two letter codes: AR, Arrowhead; CH, Chiricahua; PP, Pikes Peak; ST, Standard; TL, Tree Line. Histograms show the number of gene conversion events in the 14 syntenic blocks with the maximum value on all scales being 2,500. Blue and pink boxes below each bar indicate regions with a significant deficiency (blue) or excess (pink) of events based on a random permutation test (see the text). The number above each histogram indicates the total number of events across all 14 regions with a blue number indicating a significant deficiency, a pink number indicating a significant excess of events, and a black number indicating neither a significant deficiency or excess. The double headed arrows connect pairs of arrangements with the arrows pointing to the number of events in the recipient arrangement from the donor arrangement. For instance, 1654 conversion events were from ST to TL and 1973 conversion events were from TL to ST. The boxes with frequencies on the arrows are the maximum observed heterokaryotype frequencies in the 1980 collections of *D. pseudoobscura* (ANDERSON *et al.* 1991).

Figure 9. Mean gene conversion tract coverage in 2,668 transcripts across Muller C in *D. pseudoobscura* in five gene arrangements. Coverage is the number of times that a nucleotide is covered by a gene conversion tract. Coverage was estimated for each exon within a transcript and the mean coverage for the entire transcript was averaged over all exons. Outlier genes are shown with an orange marker while non-outliers are shown with a gray marker. The location of the derived inversion giving rise to each gene arrangement are shown with a black bar.

735 **Figure 1**



736

Figure 2

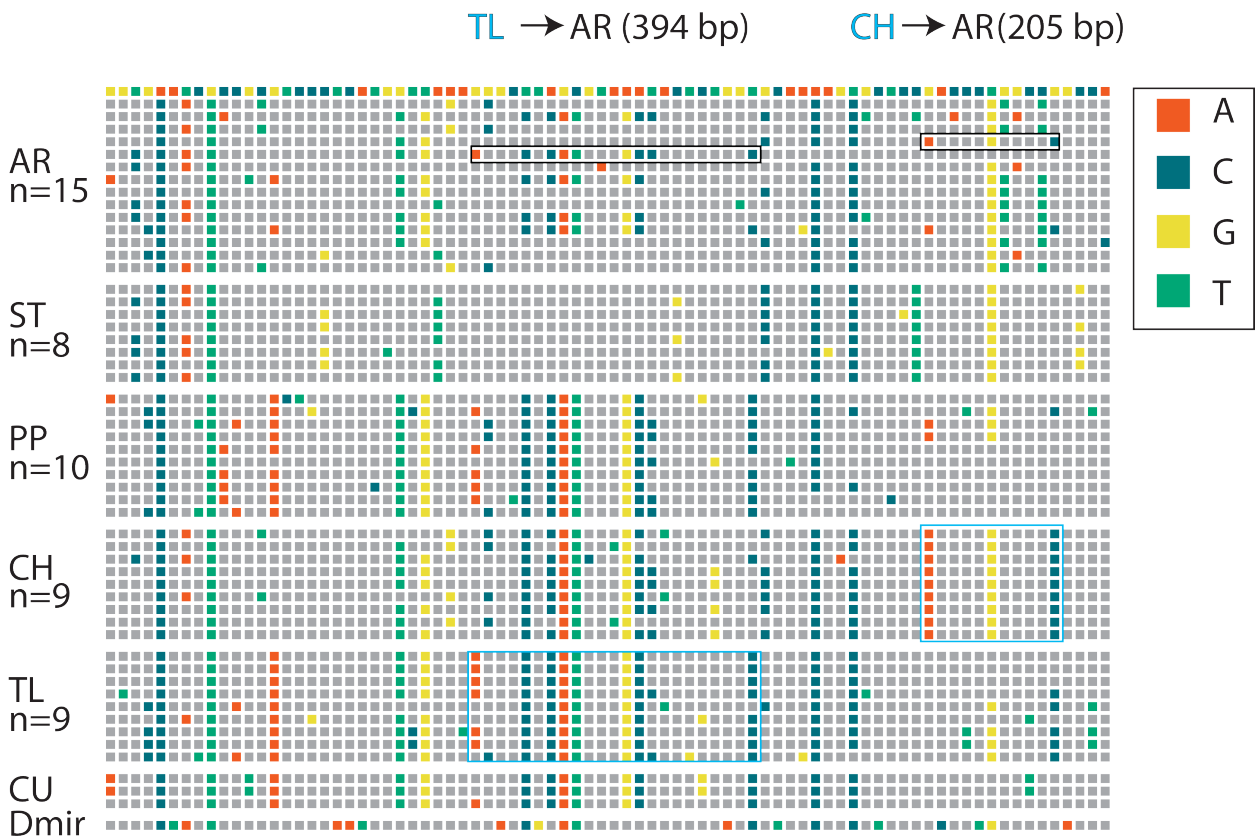
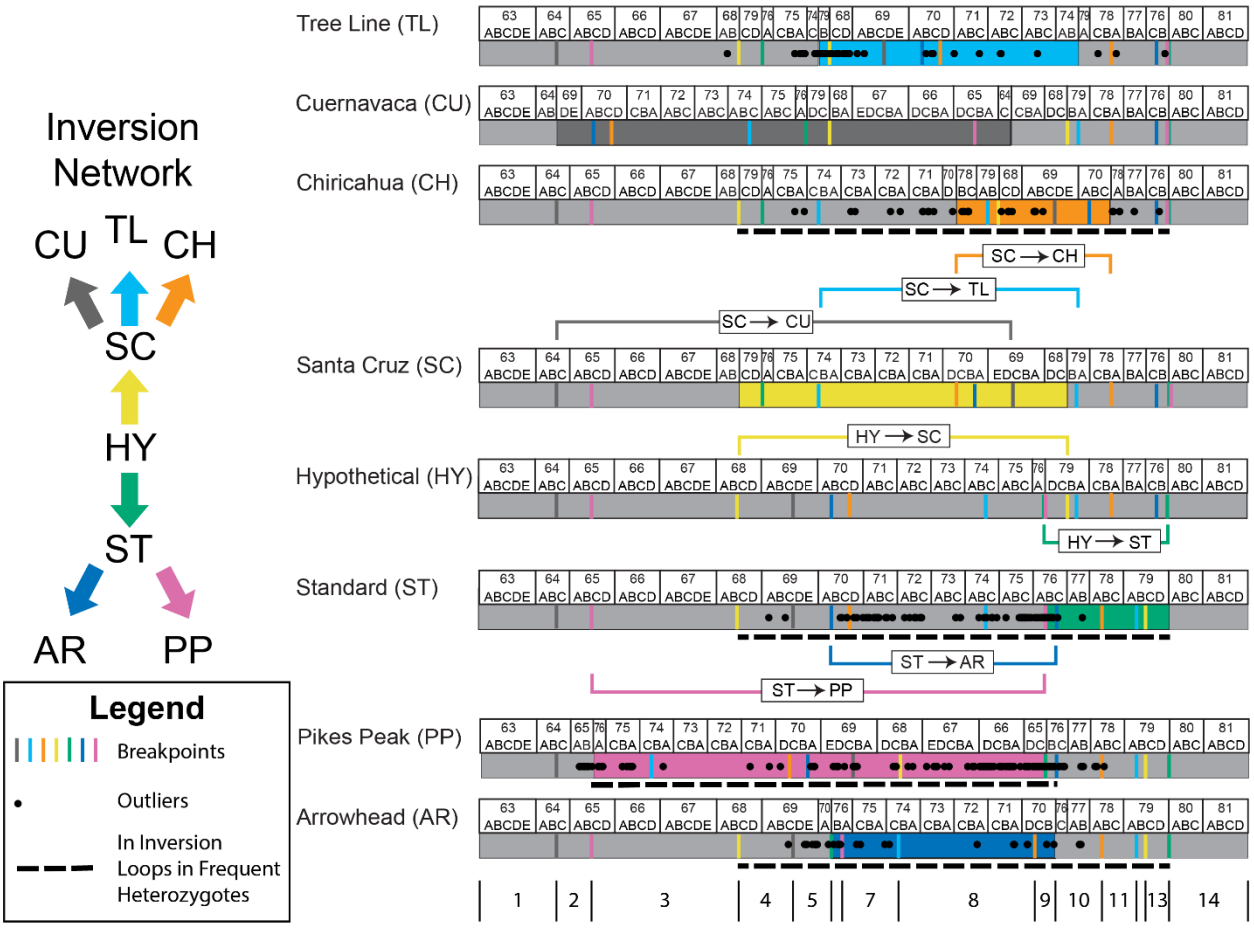
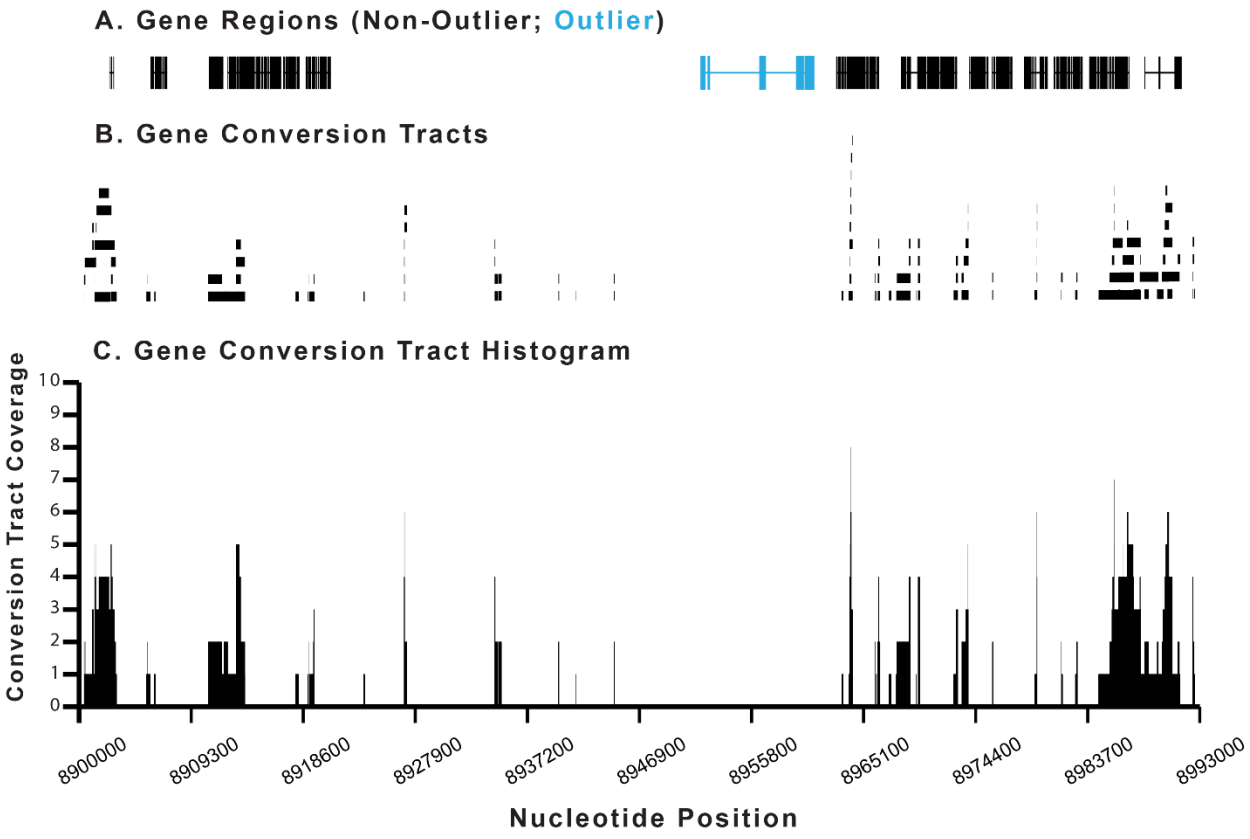

TL → AR (394 bp)
CH → AR(205 bp)
AR
n=15
ST
n=8
PP
n=10
CH
n=9
TL
n=9
CU
Dmir
A
C
G
T

Figure 3**Figure 2**

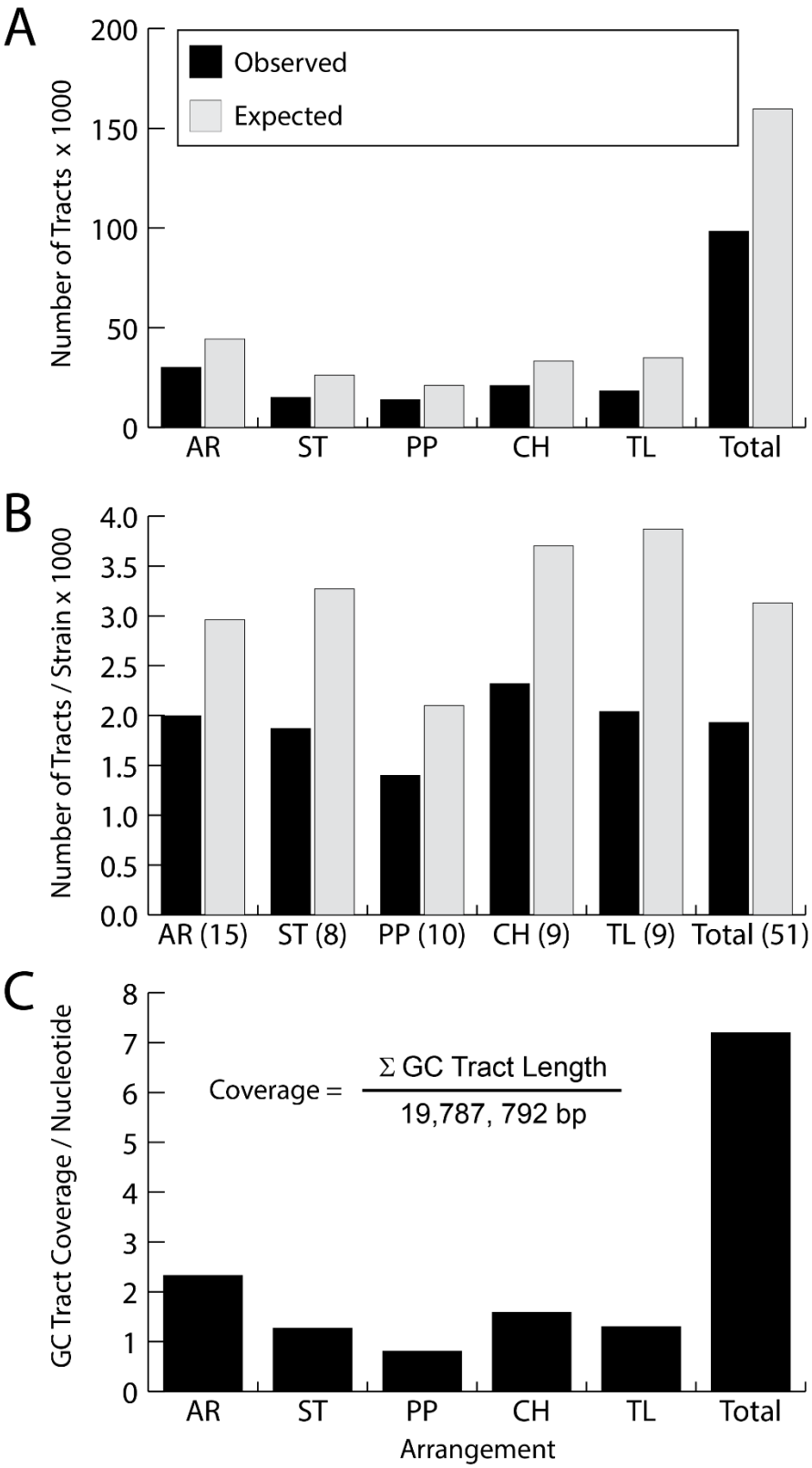


742 Figure 4



743

744 **Figure 5**



745

Figure 6**Figure 6**

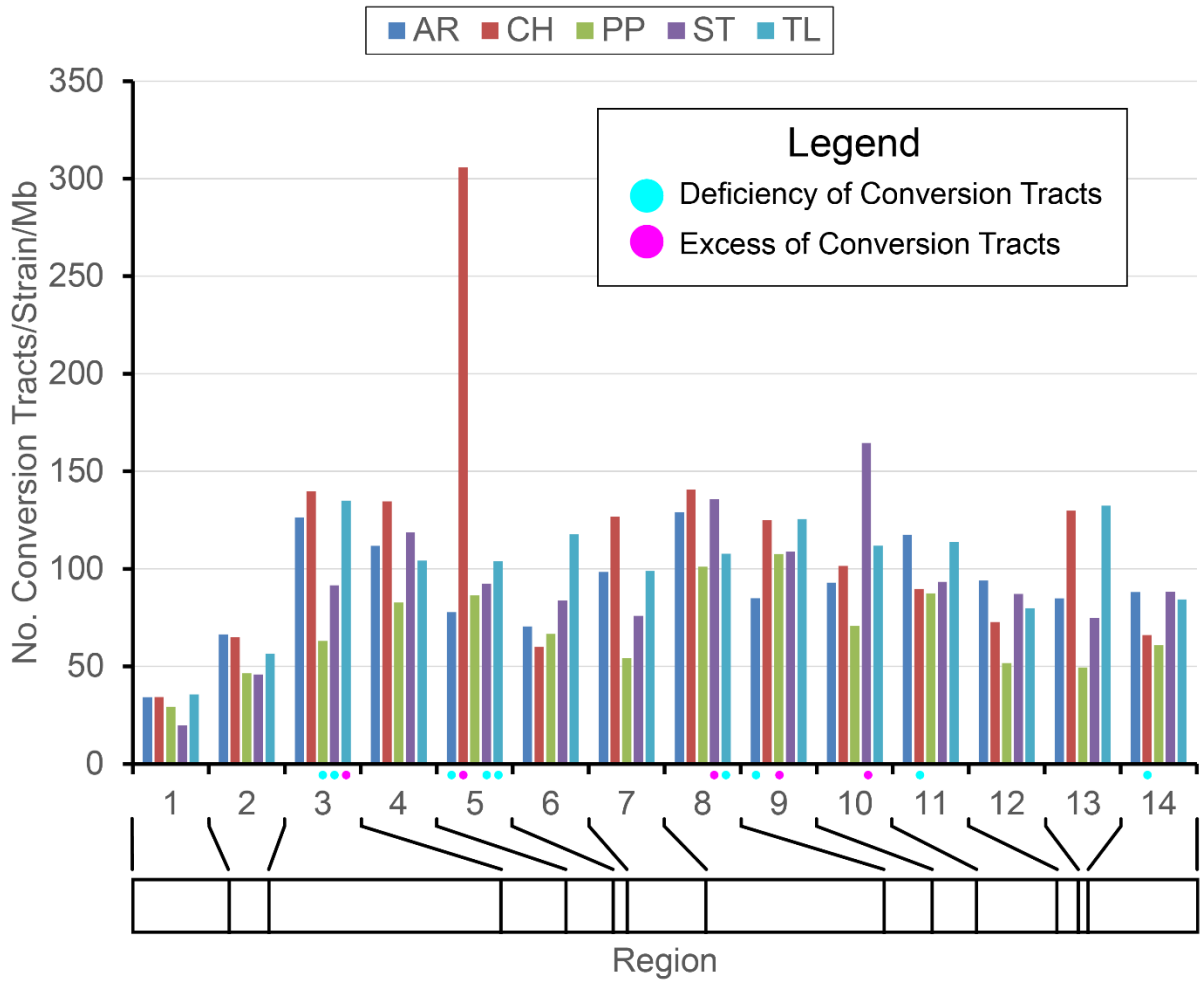


Figure 7**Figure**

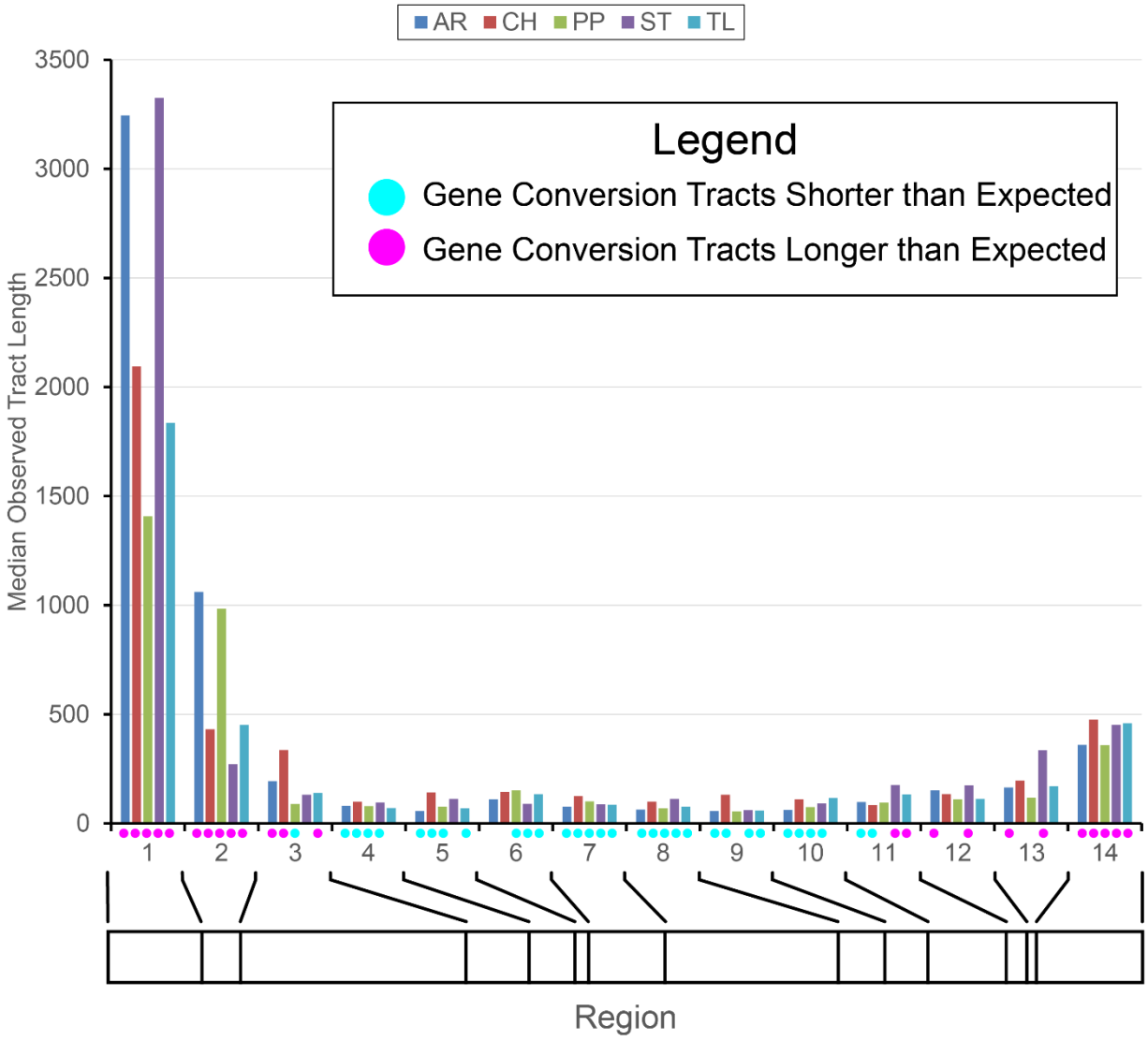


Figure 8

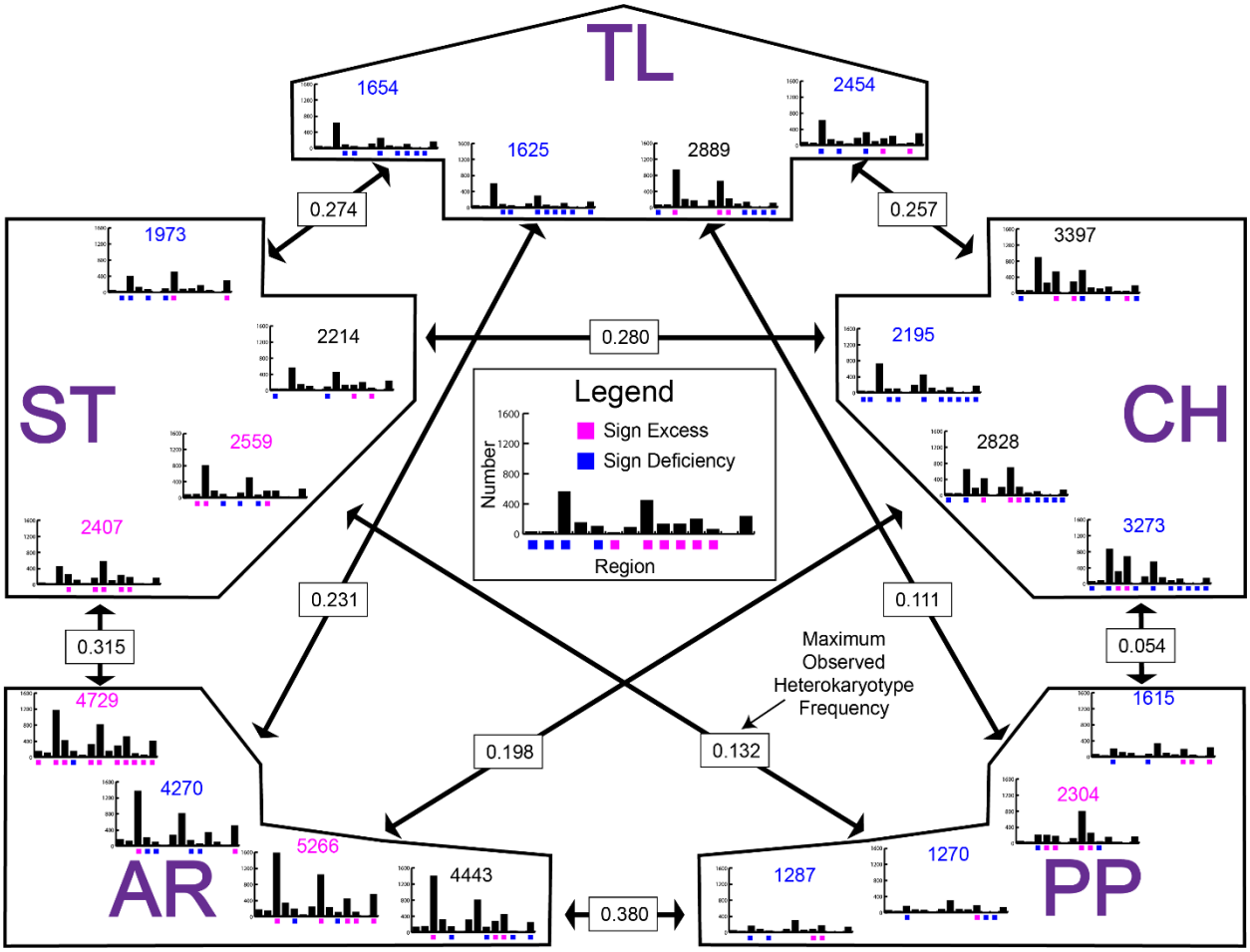
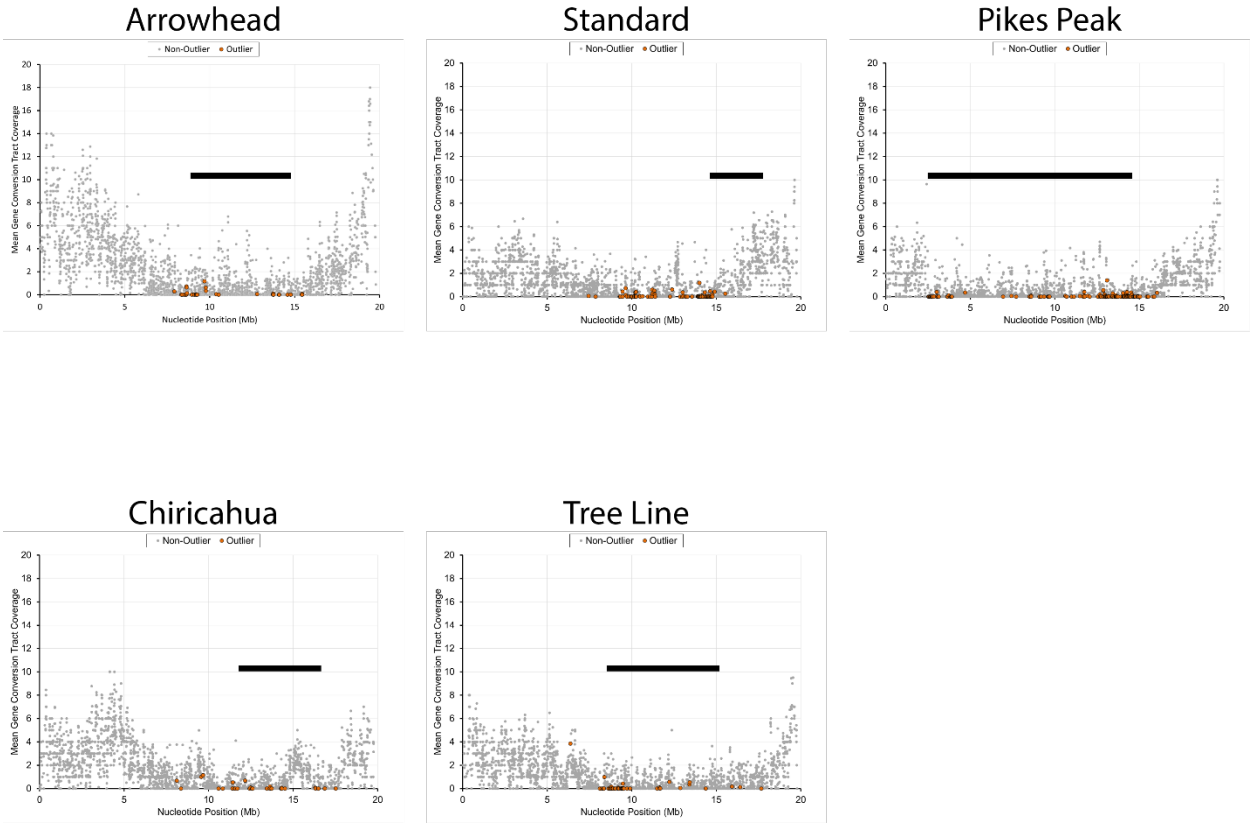

TL
1654
1625
2889
2454
0.274
0.257
ST
1973
2214
2559
2407
CH
3397
2195
2828
3273
0.280
Legend
Sign Excess
Sign Deficiency
Number
Region
0.231
0.111
0.315
0.054
Maximum
Observed
Heterokaryotype
Frequency
0.198
0.132
AR
4729
4270
5266
4443
PP
1615
2304
1270
1287
0.380

Figure 9


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.

Arostegui, M. C., T. P. Quinn, L. W. Seeb, J. E. Seeb and G. J. McKinney, 2019 Retention of a chromosomal inversion from an anadromous ancestor provides the genetic basis for alternative freshwater ecotypes in rainbow trout. Mol Ecol 28**:** 1412-1427.

Ayala, D., S. Zhang, M. Chateau, C. Fouet, I. Morlais *et al.*, 2019 Association mapping desiccation resistance within chromosomal inversions in the African malaria vector Anopheles gambiae. Molecular Ecology 28**:** 1333-1342.

Battaglia, B., and H. Smith, 1961 The Darwinian fitness of polymorphic and monomorphic populations of *Drosophila pseudoobscura* at 16° C. Heredity 16**:** 475-484.

Berdan, E. L., N. H. Barton, R. Butlin, B. Charlesworth, R. Faria *et al.*, 2023 How chromosomal inversions reorient the evolutionary process. J Evol Biol 36**:** 1761-1782.

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.

Caceres, M., A. Barbadilla and A. Ruiz, 1999 Recombination rate predicts inversion size in Diptera. Genetics 153**:** 251-259.

Carson, H. L., 1992 Inversions in Hawaiian *Drosophila*, pp. 407-439 in *Drosophila Inversion Polymorphism*, edited by C. Krimbas and J. R. Powell. CRC Press, Boca Raton, FL.

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.

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

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

Coluzzi, M., A. Sabatini, A. della Torre, M. A. Di Deco and V. Petrarca, 2002 A polytene chromosome analysis of the *Anopheles gambiae* species complex. Science 298**:** 1415-1418.

Coluzzi, M., A. Sabatini, V. Petrarca and M. A. Di Deco, 1979 Chromosomal differentiation and adaptation to human environments in the *dob* complex. Transactions of The Royal Society of Tropical Medicine and Hygiene 73**:** 483-497.

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.

Cunha, A. B. D., 1951 Modification of the Adaptive Values of Chromosomal Types in Drosophila pseudoobscura by Nutritional Variables. Evolution 5**:** 395-404.

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., 1950 The genetics of natural populations. XIX. Origin of heterosis through natural selection in populations of *Drosophila pseudoobscura*. Genetics 35**:** 288-302.

806 Dobzhansky, T., and M. L. Queal, 1938 Genetics of natural populations. II. Genic variation in
807 populations of *Drosophila pseudoobscura* inhabiting isolated mountain ranges. Genetics
808 23**:** 463-484.
809 Dobzhansky, T., and A. H. Sturtevant, 1938 Inversions in the chromosomes of *Drosophila*
810 *pseudoobscura*. Genetics 23**:** 28-64.
811 Dolgova, O., C. Rego, G. Calabria, J. Balanya, M. Pascual *et al.*, 2010 Genetic constraints for
812 thermal coadaptation in Drosophila subobscura. BMC Evol Biol 10**:** 363.
813 Druger, M., 1962 Selection and body size in *Drosophila pseudoobscura* at different
814 temperatures. Genetics 47**:** 209-222.
815 Faria, R., P. Chaube, H. E. Morales, T. Larsson, A. R. Lemmon *et al.*, 2019 Multiple
816 chromosomal rearrangements in a hybrid zone between Littorina saxatilis ecotypes. Mol
817 Ecol 28**:** 1375-1393.
818 Fuller, Z. L., G. D. Haynes, S. Richards and S. W. Schaeffer, 2017 Genomics of natural
819 populations: Evolutionary forces that establish and maintain gene arrangements in
820 *Drosophila pseudoobscura*. Molecular Ecology 26**:** 6539-6562.
821 Fuller, Z. L., S. A. Koury, N. Phadnis and S. W. Schaeffer, 2019 How chromosomal
822 rearrangements shape adaptation and speciation: Case studies in *Drosophila*
823 *pseudoobscura* and its sibling species *Drosophila persimilis*. Molecular Ecology 28**:**
824 1283-1301.
825 Funk, E. R., N. A. Mason, S. Palsson, T. Albrecht, J. A. Johnson *et al.*, 2021 A supergene
826 underlies linked variation in color and morphology in a Holarctic songbird. Nat Commun
827 12**:** 6833.
828 Gong, W. J., K. S. McKim and R. S. Hawley, 2005 All Paired Up with No Place to Go: Pairing,
829 Synapsis, and DSB Formation in a Balancer Heterozygote. PLOS Genetics 1**:** e67.
830 Guerrero, R. F., and M. Kirkpatrick, 2014 Local adaptation and the evolution of chromosome
831 fusions. Evolution 68**:** 2747-2756.
832 Guerrero, R. F., F. Rousset and M. Kirkpatrick, 2012 Coalescent patterns for chromosomal
833 inversions in divergent populations. Philosophical Transactions of the Royal Society B:
834 Biological Sciences 367**:** 430-438.
835 Hager, E. R., O. S. Harringmeyer, T. B. Wooldridge, S. Theingi, J. T. Gable *et al.*, 2022 A
836 chromosomal inversion contributes to divergence in multiple traits between deer mouse
837 ecotypes. Science 377**:** 399-405.
838 Han, F., M. Jamsandekar, M. E. Pettersson, L. Su, A. P. Fuentes-Pardo *et al.*, 2020 Ecological
839 adaptation in Atlantic herring is associated with large shifts in allele frequencies at
840 hundreds of loci. Elife 9.
841 Harringmeyer, O. S., and H. E. Hoekstra, 2022 Chromosomal inversion polymorphisms shape
842 the genomic landscape of deer mice. Nature Ecology & Evolution 6**:** 1965-1979.
843 Heuts, M. J., 1947 Influence of Humidity on the Survival of Different Chromosomal Types in
844 *Drosophila pseudoobscura*. Proc Natl Acad Sci U S A 33**:** 210-213.
845 Heuts, M. J., 1948 Adaptive properties of carriers of certain gene arrangements in *Drosophila*
846 *pseudoobscura*. Heredity 2**:** 63-75.
847 Hoffmann, A. A., and A. R. Weeks, 2007 Climatic selection on genes and traits after a 100 year-
848 old invasion: a critical look at the temperate-tropical clines in Drosophila melanogaster
849 from eastern Australia. Genetica 129**:** 133-147.
850 Huang, K., R. L. Andrew, G. L. Owens, K. L. Ostevik and L. H. Rieseberg, 2020 Multiple
851 chromosomal inversions contribute to adaptive divergence of a dune sunflower ecotype.
852 Molecular Ecology 29**:** 2535-2549.
853 Huynh, L. Y., D. L. Maney and J. W. Thomas, 2011 Chromosome-wide linkage disequilibrium
854 caused by an inversion polymorphism in the white-throated sparrow (Zonotrichia
855 albicollis). Heredity 106**:** 537-546.

Jackson, B. C., J. L. Campos and K. Zeng, 2015 The effects of purifying selection on patterns of genetic differentiation between Drosophila melanogaster populations. Heredity 114: 163-174.

Kapun, M., D. K. Fabian, J. Goudet and T. Flatt, 2016 Genomic evidence for adaptive inversion clines in *Drosophila melanogaster*. Molecular Biology and Evolution 33: 1317-1336.

Kennington, W., and A. Hoffmann, 2013 Patterns of genetic variation across inversions: geographic variation in the In(2L)t inversion in populations of Drosophila melanogaster from eastern Australia. BMC Evolutionary Biology 13: 100.

Kennington, W. J., L. Partridge and A. A. Hoffmann, 2006 Patterns of diversity and linkage disequilibrium within the cosmopolitan inversion *In*(*3R*)*Payne* in *Drosophila melanogaster* are indicative of coadaptation. Genetics 172: 1655-1663.

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.

Küpper, C., M. Stocks, J. E. Risse, N. dos Remedios, L. L. Farrell *et al.*, 2015 A supergene determines highly divergent male reproductive morphs in the ruff. Nature Genetics 48: 79.

Levine, R. P., 1952 Adaptive responses of some third chromosome types of *Drosophila pseudoobscura*. Evolution 6: 216-233.

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.

Lowry, D. B., and J. H. Willis, 2010 A Widespread Chromosomal Inversion Polymorphism Contributes to a Major Life-History Transition, Local Adaptation, and Reproductive Isolation. PLOS Biology 8: e1000500.

Lucek, K., Z. Gompert and P. Nosil, 2019 The role of structural genomic variants in population differentiation and ecotype formation in Timema cristinae walking sticks. Molecular Ecology 28: 1224-1237.

Maynard Smith, J., and J. Haigh, 1974 The hitch-hiking effect of a favorable gene. Genetical Research 23: 23-35.

McBroome, J., D. Liang and R. Corbett-Detig, 2020 Fine-Scale Position Effects Shape the Distribution of Inversion Breakpoints in Drosophila melanogaster. Genome Biol Evol 12: 1378-1391.

Mérot, C., E. L. Berdan, C. Babin, E. Normandeau, M. Wellenreuther *et al.*, 2018 Intercontinental karyotype–environment parallelism supports a role for a chromosomal inversion in local adaptation in a seaweed fly. Proceedings of the Royal Society B: Biological Sciences 285.

Miller, D. E., C. B. Smith, N. Y. Kazemi, A. J. Cockrell, A. V. Arvanitakas *et al.*, 2016 Whole-genome analysis of individual meiotic events in *Drosophila melanogaster* reveals That noncrossover gene conversions are insensitive to interference and the centromere effect. Genetics 203: 159-171.

Miller, D. E., S. Takeo, K. Nandanan, A. Paulson, M. M. Gogol *et al.*, 2012 A Whole-Chromosome Analysis of Meiotic Recombination in Drosophila melanogaster. G3 Genes|Genomes|Genetics 2: 249-260.

Muller, H. J., 1940 Bearings of the 'Drosophila' work on systematics., pp. 185-268 in *The New Systematics*, edited by J. Huxley. Clarendon Press, Oxford.

Munte, A., J. Rozas, M. Aguade and C. Segarra, 2005 Chromosomal inversion polymorphism leads to extensive genetic structure: a multilocus survey in *Drosophila subobscura*. Genetics 169**:** 1573-1581.

Navarro, A., A. Barbadilla and A. Ruiz, 2000 Effect of inversion polymorphism on the neutral nucleotide variability of linked chromosomal regions in *Drosophila*. Genetics 155**:** 685-698.

Navarro, A., E. Betrán, A. Barbadilla and A. Ruiz, 1997 Recombination and gene flux caused by gene conversion and crossing over in inversion heterokaryotypes. Genetics 146**:** 695-709.

Nei, M., K. I. Kojima and H. E. Schaffer, 1967 Frequency changes of new inversions in populations under mutation-selection equilibria. Genetics 57**:** 741-750.

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

Puerma, E., D. J. Orengo and M. Aguadé, 2016 The origin of chromosomal inversions as a source of segmental duplications in the Sophophora subgenus of Drosophila. Scientific Reports 6**:** 30715.

Puig, M., M. Cáceres and A. Ruiz, 2002 Effect of the polymorphic inversion 2j of *Drosophila buzzatii* on the genes adjacent to its breakpoints, pp. 328A in *Annual Drosophila Research Conference*, San Diego, CA.

Ranz, J. M., D. Maurin, Y. S. Chan, M. v. Grotthuss, L. W. Hillier *et al.*, 2007 Principles of genome evolution in the *Drosophila melanogaster* species group Public Library of Science Biology 5**:** 1366-1381.

Rego, C., J. Balanya, I. Fragata, M. Matos, E. L. Rezende *et al.*, 2010 Clinal patterns of chromosomal inversion polymorphisms in Drosophila subobscura are partly associated with thermal preferences and heat stress resistance. Evolution 64**:** 385-397.

Rennison, D. J., Y. E. Stuart, D. I. Bolnick and C. L. Peichel, 2019 Ecological factors and morphological traits are associated with repeated genomic differentiation between lake and stream stickleback. Philos Trans R Soc Lond B Biol Sci 374**:** 20180241.

Richards, S., Y. Liu, B. R. Bettencourt, P. Hradecky, S. Letovsky *et al.*, 2005 Comparative genome sequencing of *Drosophila pseudoobscura*: Chromosomal, gene and *cis*-element evolution. Genome Research 15**:** 1-18.

Rozas, J., and M. Aguadé, 1993 Transfer of genetic information in the *rp49* region of *Drosophila subobscura* between different gene arrangements. Proceedings of the National Academy of Sciences USA 90**:** 8083-8087.

Rozas, J., and M. Aguadé, 1994 Gene conversion is involved in the transfer of genetic information between naturally occurring inversions of *Drosophila*. Proceedings of the National Academy of Sciences USA 91**:** 11517-11521.

Rozas, J., A. Ferrer-Mata, J. C. Sánchez-DelBarrio, S. Guirao-Rico, P. Librado *et al.*, 2017 DnaSP 6: DNA Sequence Polymorphism Analysis of Large Data Sets. Molecular Biology and Evolution 34**:** 3299-3302.

Schaeffer, S. W., 2008 Selection in heterogeneous environments maintains the gene arrangement polymorphism of *Drosophila pseudoobscura*. Evolution 62**:** 3082-3099.

Schaeffer, S. W., and W. W. Anderson, 2005 Mechanisms of genetic exchange within the chromosomal inversions of *Drosophila pseudoobscura*. Genetics 171**:** 1729-1739.

Shriver, M., G. Kennedy, E. Parra, H. Lawson, V. Sonpar *et al.*, 2004 The genomic distribution of population substructure in four populations using 8,525 autosomal SNPs. Human Genomics 1**:** 274 - 286.

Sinclair-Waters, M., I. R. Bradbury, C. J. Morris, S. Lien, M. P. Kent *et al.*, 2018 Ancient chromosomal rearrangement associated with local adaptation of a postglacially

955 colonized population of Atlantic Cod in the northwest Atlantic. Molecular Ecology 27**:**
956 339-351.
957 Sperlich, D., and P. Pfriem, 1986 Chromosomal polymorphism in natural and experimental
958 populations, pp. 257-309 in *The Genetics and Biology of Drosophila 3e*, edited by M.
959 Ashburner, H. L. Carson and J. N. Thomson. Academic Press, New York, NY.
960 Spiess, E. B., and B. Langer, 1964 Mating speed control by gene arrangements in *Drosophila*
961 *pseudoobscura* homokaryotypes. Proceedings of the National Academy of Sciences
962 USA 51**:** 1015-1019.
963 Stephens, J. C., 1985 Statistical methods of DNA sequence analysis: Detection of intragenic
964 recombination or gene conversion. Molecular Biology and Evolution 2**:** 539-556.
965 Sturtevant, A. H., 1917 Genetic factors affecting the strength of genetic linkage in *Drosophila*.
966 Proceedings of the National Academy of Sciences USA 3**:** 555-558.
967 Sturtevant, A. H., 1921 A case of rearrangement of genes in Drosophila. Proceedings of the
968 National Academy of Sciences USA 7**:** 235-237.
969 Sturtevant, A. H., and G. W. Beadle, 1936 The relations of inversions in the X chromosome of
970 *Drosophila melanogaster* to crossing over and disjunction. Genetics 21**:** 544-604.
971 Wallace, A. G., D. Detweiler and S. W. Schaeffer, 2011 Evolutionary history of the third
972 chromosome gene arrangements of *Drosophila pseudoobscura* inferred from inversion
973 breakpoints. Molecular Biology and Evolution 28**:** 2219-2229.
974 Wasserman, M., 1982 Evolution of the *repleta* group, pp. 61-139 in *The Genetics and Biology of*
975 *Drosophila 3b*, edited by M. Ashburner, H. L. Carson and J. N. Thompson. Academic
976 Press, New York.
977 Westram, A. M., M. Rafajlovic, P. Chaube, R. Faria, T. Larsson *et al.*, 2018 Clines on the
978 seashore: The genomic architecture underlying rapid divergence in the face of gene
979 flow. Evol Lett 2**:** 297-309.
980 Wright, D., and S. W. Schaeffer, 2022 The relevance of chromatin architecture to genome
981 rearrangements in *Drosophila*. Philosophical Transactions of the Royal Society B:
982 Biological Sciences 377**:** 20210206.
983 Wright, S., and T. Dobzhansky, 1946 Genetics of natural populations. XII. Experimental
984 reproduction of some of the changes caused by natural selection in certain populations
985 of *Drosophila pseudoobscura*. Genetics 31**:** 125-156.
986 Yan, Z., S. H. Martin, D. Gotzek, S. V. Arsenault, P. Duchen *et al.*, 2020 Evolution of a
987 supergene that regulates a trans-species social polymorphism. Nat Ecol Evol 4**:** 240-
988 249.
989 Yi, X., Y. Liang, E. Huerta-Sanchez, X. Jin, Z. X. P. Cuo *et al.*, 2010 Sequencing of 50 human
990 exomes reveals adaptation to high altitude. Science 329**:** 75-78.

991