Genomics of Natural Populations: Evolutionary Forces that Establish and Maintain Gene Arrangements in *Drosophila pseudoobscura*

Zachary L. Fuller[*,1], Gwilym D. Haynes[*,2], Stephen Richards[†], and Stephen W. Schaeffer[*]

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

† Human Genome Sequencing Center, Baylor College of Medicine, One Baylor Plaza, Houston TX 77030Current Address:

1. Current address: 606 Fairchild Center, Department of Biological Sciences, Columbia University, New York, NY 10027
2. Current address: BC Cancer Agency, 675 W. 10th Avenue, Vancouver, BC V5Z 1L3

Key Words: Drosophila, Chromosomal Inversions, Population Genomics, Recombination, Linkage Disequilibrium

Corresponding Author: Zachary L. Fuller

606 Fairchild Center

Department of Biological Sciences

Columbia University,

New York, NY 10027

Telephone: (212) 854-4581

Email: zlf2101@columbia.edu

Running Title: Population Genomics of Inversions

Abstract

The evolution of complex traits in heterogeneous environments may shape the order of genes within chromosomes. *Drosophila pseudoobscura* has a rich gene arrangement polymorphism that allows one to test evolutionary genetic hypotheses about how chromosomal inversions are established in populations. *D. pseudoobscura* has >30 gene arrangements on a single chromosome that were generated through a series of overlapping inversion mutations with > 10 inversions with appreciable frequencies and wide geographic distributions. This study analyzes the genomic sequences of 54 strains of *Drosophila pseudoobscura* that carry one of six different chromosomal arrangements to test whether (1) genetic drift, (2) hitchhiking with an adaptive allele, (3) direct effects of inversions to create gene disruptions caused by breakpoints, or (4) indirect effects of inversions in limiting the formation of recombinant gametes are responsible for the establishment of new gene arrangements. We found that the inversion events do not disrupt the structure of protein coding genes at the breakpoints. Population genetic analyses of 2,669 protein coding genes identified 277 outlier loci harboring elevated frequencies of arrangement-specific derived alleles. Significant linkage disequilibrium occurs among distant loci interspersed between regions with low levels of association indicating that distant allelic combinations are held together despite shared polymorphism among arrangements. Outlier genes showing evidence of genetic differentiation between arrangements are enriched for sensory perception and detoxification genes. The data presented here support the indirect effect of inversion hypothesis where chromosomal inversions are favored because they maintain linked associations among multi-locus allelic combinations among different arrangements.

Introduction

Chromosome structure and genome organization have been fluid over evolutionary history. An important force that may shape how genes are distributed across the genome and their order on chromosomes may be the evolution of complex traits. For single loci, natural selection is more efficient and effective in the presence of recombination because beneficial mutations can be dissociated from linked deleterious alleles (Becks & Agrawal, 2010, 2012; McDonald, Rice, & Desai, 2016; Otto & Lenormand, 2002) On the other hand, for traits controlled by multiple loci, recombination can break apart favorable combinations of alleles just as easily as it can bring them together. Therefore, the architecture of the genome may be shaped by mechanisms that modulate recombination rates such as fusing chromosomes together (Dobigny, Britton-Davidian, & Robinson, 2015; McAllister, Sheeley, Mena, Evans, & Schlotterer, 2008) or generating karyotypic variation among closely related species (Carbone et al., 2014; Nachman, Boyer, Searle, & Aquadro, 1994; O'Neill, Eldridge, & Metcalfe, 2004; O'Neill et al., 1999).

Genes within chromosome can be shuffled through meiotic recombination, which can be reduced by the presence of chromosomal inversions (Sturtevant, 1926). Inversions reduce recombination in rearrangement heterozygotes because unbalanced gametes form as a result crossing over within the inverted segments. In humans, inversions are discovered in patients with disease symptoms and are found to have chromosomal breaks in vital genes (Pettenati et al., 1995). With the advent of genomic sequencing, non-disease causing inversions have been discovered to be segregating in human populations (Stefansson et al., 2005).

Inversions are segregating in a wide variety of taxa (Coluzzi, Sabatini, Petrarca, & Di Deco, 1979; Engelbrecht, Taylor, Daniels, & Rambau, 2011; Kupper et al., 2016; Lee, Fishman,

Kelly, & Willis, 2016; Pyhäjärvi, Hufford, Mezmouk, & Ross-Ibarra, 2013; Silva et al., 2014; Diether Sperlich & Pfriem, 1986; Stefansson et al., 2005; Zinzow-Kramer et al., 2015) and are assumed to lead to the wealth of fixed chromosomal rearrangements observed between species (W. W. Anderson, Ayala, & Michod, 1977; Carson, 1992; Coluzzi, Sabatini, della Torre, Di Deco, & Petrarca, 2002; Fontaine et al., 2015; Levitan, 1992; McGraw, Davis, Young, & Thomas, 2011; Nishikawa et al., 2015; Wasserman, 1982). Indirect evidence has suggested that natural selection modulates the frequencies of gene arrangements because many show clinal variation across environmental gradients (Balanya et al., 2003; Cheng et al., 2012; Dobzhansky, 1944; Martin Kapun, Fabian, Goudet, & Flatt, 2016; M. Kapun, Schmidt, Durmaz, Schmidt, & Flatt, 2016; M. Kapun, van Schalkwyk, McAllister, Flatt, & Schlotterer, 2014; Knibb, 1982; Knibb, Oakeshott, & Gibson, 1981; Mettler, Voelker, & Mukai, 1977) including independent parallel clines observed in northern and southern hemispheres, as well across different continents (A. R. Anderson, Hoffmann, McKechnie, Umina, & Weeks, 2005; Calboli, Kennington, & Partridge, 2003; Knibb, 1982; Kolaczkowski, Kern, Holloway, & Begun, 2011; M. Santos et al., 2005). For example, natural selection acting on inversion polymorphism in *D. buzzattii* is thought to maintain the genetic architecture underlying thermal adaptation across an environmental gradient (Soto et al., 2010). Furthermore, environmental selection is hypothesized to shape the distribution of chromosomal arrangements in the malaria vector *Anopheles funestus* (D. Ayala et al., 2011) and inversions are implicated in clinal adaptation in other *Anopheles* species (Diego Ayala et al., 2017). Despite their pervasiveness both as polymorphisms within species and fixed differences between species, the mechanisms that establish and maintain different gene arrangements in populations are unclear.

Four broad classes of hypotheses have been proposed to explain how inversions are established in populations: (1) mutation and genetic drift, (2) hitchhiking with an adaptive mutation, (3) direct effects of the inversion mutation, and (4) indirect effects of suppressed recombination (Hoffmann & Rieseberg, 2008; Kirkpatrick & Barton, 2006). A new gene arrangement can be established by drift in small isolated populations subject to high rates of extinction and recolonization (Lande, 1984). An inversion could also rise to high frequency if it captures an adaptive allele that is sweeping through the population (Maynard Smith & Haigh, 1974). Sperlich (1959) suggested that inversions could directly create selectable variation by altering structure or expression of genes flanking breakpoints (Fuller, Haynes, Richards, & Schaeffer, 2016). Another direct effect of inversions might be in their potential to form segregation distorters (Novitski, 1951). Inversions may also be established through the maintenance of linked associations of alleles because of their indirect effect of reducing recombinant gametes (Sturtevant & Beadle, 1936) either because the chromosome is free of deleterious recessive alleles (Nei, Kojima, & Schaffer, 1967; Ohta, 1971), the chromosome has epistatic combinations of alleles (Charlesworth & Charlesworth, 1973; Dobzhansky, 1950) or the chromosome has locally adapted alleles (Kirkpatrick & Barton, 2006).

Drosophila pseudoobscura has been used as a model system to investigate inversions in natural populations with > 30 different gene arrangements generated through single inversion events (Dobzhansky & Sturtevant, 1938; Powell, 1992). The polymorphism is estimated to be 1.4-1.7 million years old (Aquadro, Weaver, Schaeffer, & Anderson, 1991; A. G. Wallace, Detweiler, & Schaeffer, 2011). Several lines of direct and indirect evidence suggest that selection helps establish and maintain the different gene arrangements. Five gene arrangements are frequent, widely distributed, and form clines across the southwestern United States where

frequency shifts coincide with changes in major physiographic provinces (Dobzhansky, 1944; Lobeck, 1948; Schaeffer, 2008) (Figure 1).

The geographic clines have been stable since their initial description in the 1940s (Wyatt W. Anderson et al., 1991; Dobzhansky, 1944) despite the homogenizing effect of extensive gene flow among populations (Coyne, Bryant, & Turelli, 1987; Kovacevic & Schaeffer, 2000; Riley, Hallas, & Lewontin, 1989; Schaeffer & Miller, 1992). Altitudinal clines have also been observed (Dobzhansky, 1948a) and frequencies of the gene arrangements show seasonal cycling (Dobzhansky, 1943). In some cases, inversion frequencies in population cage experiments formed stable equilibria mimicking patterns in natural populations (Dobzhansky, 1948b, 1950; Wright & Dobzhansky, 1946). Inversion polymorphism tended to be eliminated from population cages initiated with *D. pseudoobscura* (W. W. Anderson, Dobzhansky, & Kastritsis, 1967). The problem is that karyotypic fitness can vary among different climatic zones (Schaeffer, 2008) and the Anderson et al. (1967) population cage studies do not replicate natural climates.

Allozyme and nucleotide sequence polymorphism data support the indirect effect hypothesis because different arrangements are found to have unique allelic combinations (Prakash & Lewontin, 1968, 1971; Schaeffer et al., 2003), however, these results should be viewed with caution due to the small sample sizes and limited number of loci examined. Recent transcriptomic analyses have shown that inversions capture multiple differentially expressed genes among the arrangements further supporting the indirect effect of recombination suppression hypothesis (Fuller et al., 2016).

Schaeffer (2008) used numerical analyses of a selection-migration balance model to infer the fitnesses of gene arrangement karyotypes in six niches across the east-west inversion cline (Figure 1). The fitness estimates revealed over- and under-dominance operating on the gene

arrangement karyotypes in the different niches. The selection in heterogeneous environments model completely recapitulated the inversion cline from an ancestral arrangement where the frequencies of some arrangements both increased and decreased over time. The AR, PP, and CH arrangements show a steady increase to intermediate frequency from their origins suggesting that nucleotide variation at selected loci may show signatures of hard sweeps. The model also showed that ST had both increased and decreased depending on the frequency of other arrangements in the niche suggesting that patterns of ST nucleotide diversity may show a soft sweep signature. The cline observed across the southwestern United States only represents a portion of the ecological and karyotypic diversity seen in *D. pseudoobscura* with the species range extending into Mexico where the TL arrangement is found at much higher frequencies.

Inexpensive high-throughput sequencing methods now allow us to evaluate inversion establishment and maintenance hypotheses in more detail. Here, we present an analysis of 54 third chromosomes (Muller C) from six *D. pseudoobscura* gene arrangements. We used large insert mate-pair libraries to map the inversion breakpoints to test one aspect of the position effect hypothesis, i.e., whether inversion breakpoints disrupt the coding regions of genes. To test a monophyletic origin of the arrangements, phylogenetic analyses across different syntenic regions of the third chromosome were carried out. Molecular population genetic analyses tested 2,669 gene regions on the third chromosome for signatures of adaptive evolution and the structure of linkage disequilibrium was investigated across the chromosome to test for significant non-random associations as an indirect effect of suppressed recombination.

Materials and Methods

158 ***Drosophila pseudoobscura* Strains.** Genome sequences of 54 *D. pseudoobscura* strains
159 were analyzed in this study. The genomes of 47 strains were described in the previous study of
160 codon usage bias by Fuller et al. (2014). The strains were collected from the seven localities:
161 Mount Saint Helena, CA (MSH), Santa Cruz Island, CA (SCI) (collected by Luciano Matzkin,
162 University of Arizona), James Reserve, CA (JR) (collected by Wyatt W. Anderson, University of
163 Georgia), Kaibab National Forest, AZ (KB), Bosque del Apache Wild Life Refuge (BdA,
164 collected by Sara Sheeley, Upper Iowa University and provided by Bryant McAllister,
165 University of Iowa), Davis Mountains, TX (DM), and San Pablo Etla, Oaxaca, Mexico (SPE123)
166 (collected by Theresa A. Markow, UC San Diego). The genome sequences for seven additional
167 strains were generated for this study. The strains carried one of six different gene arrangements:
168 Standard (ST), Arrowhead (AR), Pikes Peak (PP), Chiricahua (CH), Cuernavaca (CU), or Tree
169 Line (TL) (Dobzhansky & Sturtevant, 1938) (Figure 2). The strains were made homozygous for
170 the third chromosome (Muller C, See Muller, 1940) by inbreeding, brother sister mating in a
171 heterozygous strain, or balancer crosses (see Supporting Information).

172 **Illumina Library Construction and Sequencing.** Genomic DNA samples from single
173 male flies of each strain were purified using Qiagen DNAeasy Blood and Tissue Kit following
174 manufacturer recommendations, including an RNAse digestion step. High molecular weight
175 double stranded genomic DNA samples were constructed into Illumina paired end libraries
176 according to the manufacturer's protocol (Illumina Inc.) with modifications as described in the
177 Supporting Information. Sequencing analysis was first done with Illumina analysis pipeline.
178 Sequencing image files were processed to generate base calls and phred-like base quality scores
179 and to remove low-quality reads.

Analysis of High Throughput Sequencing Reads and Final SNP Dataset. The paired end sequence reads (101 bp) were aligned to the *D. pseudoobscura* reference strain FlyBase version 3.02 (http://flybase.org) using bwa-mem (v. 0.7.8 Li & Durbin, 2009) with default parameters. GATK (v. 3.1.1 McKenna et al., 2010) software was used to remove duplicate sequence reads, recalibrate base quality scores, and locally realign regions around indels for BWA alignments (DePristo et al., 2011). Pileup files generated with SAMtools (v. 0.1.19 Li et al., 2009) were used to determine the coverage distributions for each strain. The reference genome carries the AR arrangement (Richards et al., 2005). We called single nucleotide polymorphisms (SNPs) using the population haplotype-based software FreeBayes (v. 0.9.21 Garrison & Marth, 2012). Sites with a phred-quality score less than 30 or coverage less than two were filtered from the data. On average, 96.066 % of nucleotide sites on the third chromosome were retained for each individual strain. Because our crossing scheme produced individuals isogenic for the third chromosome, we filtered sites called as heterozygous as they are likely to be the result of read misalignment, repetitive sequence regions, or poor read mapping. Of the 19.8 Mb on the third chromosome, 0.03-0.06 % of the nucleotide sites were called as heterozygous with genotype Phred score >30. As described in Fuller et al.(2014), we found no evidence for clusters of such sites that might suggest particular regions that resisted becoming isogenic. A SNP table for the third chromosome is available through Scholarsphere (https://scholarsphere.psu.edu).

Mapping Gene Arrangement Breakpoints. The breakpoints of the 5.9 Mb inversion that converted the Standard arrangement into Arrowhead were previously mapped using PCR (Richards et al., 2005). Large insert (3 kb) mate-pair libraries were constructed for the PP_DM1020_B, CH_JR4_L, CH_JR32_B, TL_MSH76_B, and CU_SPE123_5-2_B strains to

map the locations of derived breakpoints for the inversions that converted ST to PP, HY to ST, HY to SC, SC to CH, SC to CU, and SC to TL (see Figure 2 for the schematic maps and Figure S1 in the Supporting Information for the karyotypic maps). DNA from each strain was fragmented to an average size of 3 kb using the Nextera protocol from Illumina. The 3 kb fragments were circularized and sonicated to produce 400 bp fragments. The resulting DNAs were sequenced from the 5' and 3' ends. The sequenced fragments included both mate pairs where the ends are separated by 3 kb and paired ends separated by 400 bp. The reads from the mate-pair library were mapped to the *D. pseudoobscura* AR reference genome (MV 2-25 version 3.02) using bwa-mem (v. 0.7.8 Li & Durbin, 2009) under default parameters. The resulting SAM file was used to locate the breakpoints (Corbett-Detig, Cardeno, & Langley, 2012). For more details about breakpoint mapping, see the Supporting Information.

Phylogenetic Analysis of Gene Arrangements. Previous molecular evolutionary analysis of a limited number of genetic markers across the third chromosome supported a unique origin of the different gene arrangements of *D. pseudoobscura* (Aquadro et al., 1991; A. G. Wallace et al., 2011). We performed phylogenetic analysis on the more complete SNP data from across the third chromosome to test whether inversions were of unique origin. In addition, we used phylogenetic analysis of SNPs from 14 syntenic blocks across the chromosome to determine if all regions have a similar history. We used neighbor-joining as implemented in MEGA6 (Koichiro Tamura, Stecher, Peterson, Filipski, & Kumar, 2013) to infer the relationships among the 54 *D. pseudoobscura* strains using the outgroup strain, *D. miranda* (Saitou & Nei, 1987). The evolutionary distances were computed using the Maximum Composite Likelihood method (K. Tamura, Nei, & Kumar, 2004) assuming either a uniform or heterogeneous rate of evolution. All positions containing gaps and missing data were eliminated.

The percentage of replicate trees in which the associated taxa clustered together were determined with 500 bootstrap replicates (Felsenstein, 1985).

Classification of SNP sites. The inversions represent different subpopulations and we can test for genes with significant elevated frequencies of unique derived mutations, which are candidate genes targeted by adaptive evolution. The segregating sites discovered in this sample are a composite of unique and shared mutations. A unique derived mutation was one that occurred on the branch immediately before or after the inversion mutation. Shared mutations either predate two or more inversion mutations or are alleles transferred among arrangements via gene conversion or double cross overs (Arcadio Navarro, Betrán, Barbadilla, & Ruiz, 1997). Shared polymorphisms could also result from the balancer crosses that were used to generate the isochromosomal strains (Miller et al., 2016). See the Supporting Information for more details on how polymorphic sites were classified.

Estimates of Site Frequency Spectrum and Mean Derived Allele Frequency within Protein Coding Genes within Chromosomal Arrangements. A total of 2,669 protein coding gene models are annotated for the third chromosome in release Dpse 3.02 FB14_04 in FlyBase (http://flybase.org). We examined the site frequency spectrum and mean derived allele frequency (DAF) for each gene including 1,000 bps upstream and downstream of the annotated transcriptional start and end sites. We used the largest of multiple encoded transcripts for each gene. The site frequency spectrum for each gene was summarized with the ratio of Tajima's D (1989) to the absolute value of its theoretical minimum (D_{min}), which occurs when all segregating sites have a frequency of $1/n$, where n is the sample size. Tajima's D is sensitive to sample size and numbers of segregating sites, but $D/|D_{min}|$ controls for arrangements with different sample sizes (Schaeffer, 2002). The $D/|D_{min}|$ has a minimum value of -1 and its mean

value will reflect the demographic history of the population, which should be similar for all loci in the genome (Hahn, Rausher, & Cunningham, 2002; Schaeffer, 2002). We used a random permutation test to detect outlier loci that have clusters of low or high frequency variants. Shuffling the position of sites will break up these clusters such that the observed values of $D/|D_{min}|$ will be more extreme than the values based on random permutations for a given number of segregating sites and the mean local site frequency spectrum in proximal, inverted, or distal regions. The segregating sites within each region were shuffled without replacement 100,000 times and we estimated how often the permuted Tajima's $D/|D_{min}|$ was more extreme than the observed value in the gene region. This approach maintained the linkage relationships of variable nucleotides within the chromosome region. We used the *q*-value method with a false discovery rate (FDR) of 0.01 to correct for multiple tests (Storey, 2002, 2003). Shared and unique SNPs were analyzed separately for the five gene arrangements with sample sizes greater than eight, Arrowhead (15), Standard (8), Pikes Peak (10), Chiricahua (9), and Tree Line (9). A similar approach was used to detect significant clusters of high frequency derived unique mutations per segregating site (DAF) in the 2,669 genes, which includes polymorphic and fixed mutations.

Population Specific Branch Length Tests. A second method to detect outlier genes that accumulated large numbers of arrangement specific mutations is similar to other statistics such as the "Population Branch Statistic" (PBS; Yi et al., 2010) and "Locus Specific Branch Length" (LSBL; Shriver et al., 2004). This test has been used to identify loci with evidence of adaptive evolution (Huerta-Sanchez et al., 2013). Prior use of PBS and LSBL require exactly three subpopulations. Here, we extend these methods to an unrooted phylogeny of known topology

containing any number of branches and nodes and details of the analysis can be found in Supporting Information.

We estimated the Population Specific Branch Length (PSBL) for each gene on the third chromosome including 1 kb upstream and downstream. We used similar bootstrap approaches as our analyses of $D/|D_{min}|$ and DAF, to detect significantly large PSBL estimates with a false discovery rate (FDR) of 0.01 (Storey, 2002, 2003).

The detection of statistical outliers could be performed with coalescent simulations which specify the demographic history of each arrangement. We know the branching order of the different gene arrangements, but we have insufficient information about the demography of each arrangement to carefully specify a precise model for the system of inversions. For this reason, we used random permutation tests that shuffled SNP positions and looked for genes that were statistical outliers rather than searching for parameter values that could support any particular model or hypothesis.

Detection of Putative Functional Variation in Protein Coding Genes Within Gene Arrangements. We extracted and translated the 2,669 third chromosome coding sequences from the *D. pseudoobscura* reference annotation version 3.02 available from FlyBase, (dos Santos et al., 2015). We used the same scoring scheme that was used to classify nucleotide polymorphisms (Figure S2 in the Supporting Information). Variation in stop codons was counted in the data set. We inferred the amino acid transition matrix counting the number of events from the ancestral to derived amino acids. Additionally, we examined the frequency spectrum of the amino acid transitions for all sites.

Detection of Significant Linkage Disequilibrium. Linkage disequilibrium was characterized for all polymorphic sites across the third chromosome using the correlation based

294 approach of Zaykin *et al.* (2008), similar to r^2 (Hill & Robertson, 1968). A heat map image
295 matrix was generated by taking the average significance values at 100 adjacent sites. At each
296 polymorphic site, we also performed a Fisher's exact test to assess the significance of allele
297 association within chromosomal arrangements. All significance values were corrected for
298 multiple testing by controlling the FDR (Storey, 2003).
299

Results

Map of Inversion Breakpoints. Thirty eight genetic and physical markers from Muller C were used to bracket the locations of each breakpoint in the AR reference sequence (see Supplemental Table 21 Muller C tab, Schaeffer et al., 2008). Seven pairs of inversion breakpoints were mapped with the mate pair data similar to previous approaches (Corbett-Detig et al., 2012; Cridland & Thornton, 2010)(See Supporting Information). Mate pair reads that mapped at megabase distances apart are candidates for breakpoint positions if the two ends coincide with the approximate locations of breakpoints on the cytological map (Dobzhansky, 1944; Dobzhansky & Sturtevant, 1938).

All breakpoints mapped within intergenic regions and not within the transcripts of the boundary genes (Table 1). An average of ~4.5 breaks (95% CI: 1-8) are expected in intergenic regions if 14 breakpoints are placed randomly, using a uniform distribution, on the third chromosome map that includes 6.3 Mb of intergenic regions and 13.4 Mb of coding nucleotides. Four of the thirteen breakpoint regions had nucleotide sites with elevated coverage (pHYSC, pSTAR, dSTAR, and pSCCH). We failed to observe elevated read coverage in any boundary gene indicating that genes adjacent to inversion breakpoints are not duplicated as has been observed in other Drosophila breakpoints (Calvete, Gonzalez, Betran, & Ruiz, 2012; Guillen & Ruiz, 2012; Papaceit, Segarra, & Aguadé, 2013; Puerma et al., 2014; Ranz et al., 2007).

Phylogenetic Relationships of Arrangements 14 Syntenic Regions Show Diverse Histories across the Third Chromosome. We used paired-end reads generated from each strain to identify single nucleotide polymorphisms (SNPs). The median coverage across all strains was 35 (Table S4 and Figure S24). From these aligned reads, we identified 1,028,037 SNPs on the

third chromosome and estimated error rates to vary from 9.7×10^{-4} in AR to 3.4×10^{-3} in PP, CH, and TL using a set of previously sequenced markers (Tables S5 and S6).

The overall phylogenetic tree supports the unique origin hypothesis based on the monophyletic relationship of each of the six arrangements (Figure 3) consistent with RFLP and short nucleotide marker data (Aquadro et al., 1991; A. G. Wallace et al., 2011). The root of the tree, HY, was also confirmed with gene adjacency information at the breakpoints (Bhutkar et al., 2008)(See Identification of Inversion Breakpoints in Supporting Information). The branching order is consistent with the cytogenetic phylogeny, except that PP is inferred to be the sister clade to the AR, ST, CH, CU, and TL gene arrangements rather than forming a monophyletic group with ST and AR (Dobzhansky & Sturtevant, 1938).

We tested whether PP showed this unexpected relationship in all regions of the third chromosome (Figure 2). We used neighbor-joining to infer the relationships among the six arrangements in the 14 syntenic regions (Figure 4, Figures S25-S38 in the Supporting Information). The six gene arrangements form monophyletic clusters in the central syntenic blocks where recombination is reduced. Not all syntenic blocks, however, are concordant with the established cytogenetic phylogeny, e.g., see regions 68C-69C, 69D-70A, 76C-78A, and 78B-79A in Figure 4 compared with the cytogenetic phylogeny in (Figure 2). PP consistently diverges before the split of the AR and ST lineages within the ST phylad in these four regions. The branching order of the CH, CU, and TL arrangements within the Santa Cruz phylad is also not consistent among these four syntenic regions. While CU is always the most derived member of the Santa Cruz phylad, it is not clear whether CH or TL is the more ancestral member.

In syntenic regions that are discordant with the cytogenetic phylogeny, the PP clade is the most inconsistent group. PP clusters with the Santa Cruz phylad in four regions (76B, 76A-74C,

74B-70D, 70C-70B) and is basal to the phylad in two of the four regions (76B, 76A-74C). PP is sister to TL in the other two blocks (74B-70D and 70C-70B). Block 74B-70D is also unusual because the CH and CU clades cluster with the Standard phylad. In the three remaining syntenic blocks (79B, 79C-79D, 80A-81D), PP strains fail to form monophyletic groups and clusters among the AR and ST strains.

The Site Frequency Spectrum of Inversion-Specific Mutations. We examined the site frequency spectra of inversion-specific mutations with Tajima's $D/|D_{min}|$ (Figure 5). The frequency spectra shows more high frequency variants near and within inverted regions and an excess of rare variants in the proximal and distal regions. Proximal and distal regions have more shared than unique polymorphisms and the frequency of derived mutations is higher in shared versus unique polymorphisms with random permutation tests in five arrangements (All arrangements and regions $P < 1.0 \times 10^{-3}$)(Table 3) and Chi square tests of homogeneity (AR, X^2=12,688.9, df=2, $P< 1x10^{-6}$; PP, X^2=14,562.0, df=2, $P< 1x10^{-6}$; CH, X^2=8,923.5, df=2, $P< 1x10^{-6}$; TL, X^2=12,495.1, df=2, $P< 1x10^{-6}$; CU, X^2=457.8, df=2, $P< 1x10^{-6}$). ST has an excess of unique SNPs in the proximal region compared to the inverted and distal regions (ST, X^2=1,345.1, df=2, $P< 1x10^{-6}$).

We tested for outlier genes that either have significant excesses of low ($D/|D_{min}|<0$) or intermediate ($D/|D_{min}|>0$) frequency SNPs. The majority of extremely low values of $D/|D_{min}|$ are found in the proximal and distal regions where shared mutations with higher frequencies were removed leaving younger low frequency mutations. Significant elevations in $D/|D_{min}|$ occurred for genes within the inversion where fewer shared polymorphisms were removed. For ST and CH, the elevation of Tajima's $D/|D_{min}|$ extends approximately five and seven Mb upstream of the proximal breakpoint, respectively.

We estimated $D/|D_{min}|$ for the non-inverted second chromosome (McGaugh et al., 2012). Tajima's $D/|D_{min}|$ varies uniformly across chromosome two with a mean $D/|D_{min}|$ of -0.55, which is consistent with a population expansion parameter (Nr) of 30 based on coalescent simulations (McGaugh et al., 2012; Schaeffer, 2002) (See Supporting Information and Figure S39). Fifteen regions show a significantly excess of intermediate frequency variants and thirty regions have a significant excess of rare variants (Figure S39). This indicates the pattern of Tajima's $D/|D_{min}|$ across the inverted third chromosome is influenced by the presence of inversions.

The mean values of Tajima's $D/|D_{min}|$ for the five arrangements are AR =-0.72, ST = -0.53, PP= -0.33, CH= -0.41, and TL = -0.49. The $D/|D_{min}|$ values for ST, CH, and TL are similar to that of the second chromosome consistent with a similar demographic history. The distribution of $D/|D_{min}|$ for AR is more negative than the other arrangements suggesting a greater rate of expansion for this chromosome.

Detection of Protein Coding Regions with Elevated Frequencies of Derived Mutations. The site frequency spectrum as summarized by Tajima's $D|D_{min}|$ does not include nucleotide sites that are fixed within a gene arrangement background in our sample. We estimated the mean derived allele frequencies (DAF) of unique polymorphisms in the 2,669 genes and tested loci for significant clusters of high frequency alleles using random permutation tests (Figure 6). The AR, ST, PP, CH, and TL arrangements had 138, 229, 233, 161, and 173 genes, respectively, with a significant elevation of mean DAF per segregating site. The mean DAF tends to be at its lowest near the centromere and telomere and increases near the inversion breakpoints with the highest values observed within the inverted regions and up to 1 Mb upstream or downstream. The exceptions are ST and CH where the mean DAF is elevated within five to seven Mb of the proximal breakpoint, which overlaps with the inverted region of AR.

The mean DAF frequency is uniformly low across chromosome two (~0.2) not reaching the high values seen for any arrangement on the third chromosome (See Figure S39 in Supporting Information).

Detection of Arrangement Specific Allele Frequency Changes with Population Specific Branch Length Analysis. Population specific branch length (PSBL) analysis allowed us to determine which gene regions have a significantly high proportion of arrangement specific allele frequency changes (see Figure 7 and Supplemental Spreadsheet 1). A significantly long branch will occur if a gene has an excess of arrangement specific mutations relative to the other four arrangements. For each lineage on the phylogeny, intervals with the longest branch lengths are located within or near inversion breakpoints. The mean branch length for genes in inverted regions is significantly greater than for genes located outside of the breakpoints in every case (Wilcoxon rank-sum test, $P<0.05$). After correcting for multiple testing with a FDR of 0.01, we determined that 317, 380, 396, 350, and 340 genes had a significantly long branch length in the AR, ST, PP, and CH arrangements, respectively. Of these, 71, 130, 192, 74, and 75 genes had long branches only within AR, ST, PP, CH, and TL, respectively. Some genes had longer branches in multiple arrangements with 249, 149, 54, and 16 genes in two, three, four, or five arrangements, respectively. A total of 1,659 genes did not have a significantly long branch length in any of the five arrangements, while 1,010 genes had a significantly long branch length in at least one arrangement.

Elevated Derived Allele Frequencies and Fixed Amino Acid Changes within Inverted Gene Regions. Here, we ask how many of the DAF and PSBL outlier gene regions also harbor at least one fixed amino acid change within a particular arrangement in our sample. A total of 277 gene regions DAF and PSBL outlier genes contained at least one fixed amino acid

414 difference in at least one gene arrangement. There are 28, 74, 144, 31, and 47 outlier genes with
415 at least one fixed amino acid in AR, ST, PP, CH, or TL, respectively. Chi-square tests of
416 homogeneity support a heterogeneous distribution of candidate selected genes with more outliers
417 in inverted regions for all arrangements except ST (AR, X^2=24.1, df=2, P=5.9x10^{-6}; ST, X^2=9.2,
418 df=2, P=0.010; PP, X^2=36.1, df=2, P=1.4x10^{-8}; CH, X^2=28.7, df=2, P=5.7x10^{-7}; TL, X^2=42.2,
419 df=2, P=6.9x10^{-10}).

420 The outlier genes are found across the length of the inverted regions, but are not
421 uniformly distributed based on Kolmogorov-Smirnov tests (AR D_{max}=-0.258 P=0.005; ST,
422 D_{max}=0.387 P=0.001; PP D_{max}=0.387 P<1 x 10^{-4}; CH D_{max}=-0.285 P=0.001; TL D_{max}=-0.555 P<1
423 x 10^{-4}).The distance of the closest outlier gene to the center of the inverted region in AR, ST, PP,
424 CH, and TL is 0.92, 0.62, 0.04, 0.08, and 0.11 Mb, or 15.5, 20.0, 0.3, 1.7, and 1.7 % of the total
425 inversion size, respectively. Genes with the maximum DAF frequency or PSBL are more
426 centrally located within the inverted region and are not adjacent to either the proximal or distal
427 breakpoints in each case.

428 We tested for a linear association between the size of each chromosomal region
429 (proximal, inverted, and distal) and the proportion of genes showing evidence of adaptive
430 evolution (Table 4), observing a significant positive relationship (P=0.003, R^2=0.96). While there
431 is a positive correlation in proximal and distal regions, the relationship is not significant
432 (proximal: P=0.14, R^2=0.57; distal: P=0.37, R^2=0.27). In four gene arrangements (AR, PP, TL,
433 CH), there is an excess of candidate adaptive genes within inverted regions.

434 We asked if recombination rates varies within and outside of outlier gene regions using
435 the fine-scale maps of the population scaled recombination rate ρ across all arrangements
436 estimated by Fuller et al. (2014). Outlier genes within inverted regions have significantly lower

mean estimates of ρ than non-outlier genes (Table 5 and Figure 8). These data show strongly differentiated genes interspersed among regions that experience higher levels of recombination within the inversion especially in the middle of the inversion which should be sufficient to decrease LD (Schaeffer & Anderson, 2005).

Each arrangement has at least one gene with multiple fixed amino acid changes in our sample. Of these, the GA16823 gene in CH has the smallest number of fixed amino acid differences (9) while GA24454 in PP is the most extreme case with 47 fixed changes. In *D. melanogaster*, the ortholog of GA24454 is CG33017 and polymorphisms in the gene are significantly associated with olfactory responses to 2-phenyl ethyl alcohol (Arya et al., 2015). GA24454 is within inverted regions when PP and CU are paired with AR, ST, CH, or TL, however, the gene is outside inverted regions in heterozygotes in all other heterokaryotypes. Even though 207 amino acid polymorphisms are segregating in GA24454 across all arrangements, none of the other arrangements has a fixed amino acid difference. Coalescent simulations show that 47 fixed amino acid changes is greater than expected with a model of nested subsamples given 207 segregating amino acids (P=0.028; Hudson & Kaplan, 1986; Schaeffer et al., 2003). Furthermore, PP has only three shared amino acid polymorphisms while non-PP arrangements have 12 to 26 shared amino acid polymorphisms. The significantly large number of fixed amino acid differences in PP is evidence for adaptive evolution in GA24454.

Lack of Extended Haplotype Homozygosity Surrounding Fixed Unique Nucleotide Alleles. At each fixed, derived site within an arrangement, we estimated the integrated haplotype score (*iHS*) to test for extended haplotype structure (See Figure S41 in the Supporting Information). Long stretches of extended homozygosity within an arrangement are likely to be associated with a strong or recent selection event. In each arrangement, we detected at least one

site with an extreme positive or extreme negative (±3) *iHS* value. Non-neutral forces are expected to generate clusters of consecutive sites with extreme *iHS* values. We find evidence for only one such interval in any arrangement, an ~18-kb stretch from 9,271,032 to 9,289,710 in AR. The interval intersects the coding regions of two overlapping genes (GA12653 and GA24326) neither of which showed a significantly elevated mean DAF or PSBL. For all other arrangements, no significant intervals of consecutive extreme *iHS* scores were observed.

Third Chromosome Linkage Disequilibrium Patterns. Now we ask whether significant associations between sites decay with distance across the third chromosome (Figure 9). In *D. melanogaster*, LD tends to decay rapidly within the genome through recombination (Langley et al., 2012; Mackay et al., 2012), although recombination appears to be suppressed across the length of the inverted regions here (Fuller et al., 2016). In proximal and distal regions of the chromosome, we observe few pairwise comparisons that are in significant LD (3.97% of all comparisons). However, we find extensive LD generated within inverted regions (20.29% of all comparisons) and observe a significantly elevated proportion of significant pairwise comparisons present relative to non-inverted segments (Wilcoxon rank-sum test, $P<0.05$). Furthermore, multiple arrangements show significant associations between SNPs and arrangements with Fisher's Exact Test across the majority of the chromosome (Figure 9) indicating that the pattern of LD is not being driven by a single arrangement. The exception is in the region where only PP differs from the other arrangements. Nucleotide variation around the proximal and distal breakpoints did not always show significant LD (Figure S42 in Supporting Information).

Evidence for Differentiation among Inversions, but not Populations. The six arrangements sampled were collected from five different geographic locations. We found

minimal evidence for geographic differentiation (Table 2). Population-specific singletons account for 9.8 to 19.8% of the total unique polymorphism and the majority of derived polymorphisms (70.9 to 89.3 %) are shared among populations. The Mexican (SPE) population did have 56 (0.01%) unique fixed differences, but the other four populations did not have fixed unique derived polymorphisms. The extensive shared polymorphisms among geographic locations and lack of fixed differences are consistent with extensive gene flow among populations.

Gene arrangements, on the other hand, are differentiated from each other. Inversion specific polymorphism with a frequency of 1 in our sample account for 13.5 to 23.6 % of the observed unique polymorphism. Despite suppressed recombination in inverted regions, the majority of derived polymorphisms (56.3 to 78.7 %) are shared among arrangements. All arrangements have a greater proportion of fixed unique polymorphisms (0.8 to 7.4%) than the geographic populations. Gene arrangements harbor a substantial proportion of fixed unique sites, even in the presence of extensive shared polymorphism.

Gene Ontology. To test for enrichment of common biological functions in the genes displaying evidence of adaptive evolution, we performed a Gene Ontology (GO) analysis using DAVID (v6.8) software (Huang, Sherman, & Lempicki, 2009a, 2009b). We note that direct experimental evidence is needed to confirm any of our results, however our analyses provide insight into genes with signals of adaptive evolution involved in similar biological functions that may underlie the targets of selection acting on the inversion polymorphism. For genes across all arrangements with at least one fixed amino acid change and significant mean DAF and PSBL, after correcting for multiple testing there is a significant enrichment of categories including odorant binding ($q < 8.75 \times 10^{-4}$; see Table 5) and neuroactive ligand-receptor interaction ($q < 1.34 \times 10^{-3}$). Several other categories of potential biological interest, including starch and

sucrose metabolism and limonene and pinene degradation, contain genes which show evidence of adaptive evolution, although they are not significantly enriched.

Limonene and pinene are monoterpenes that are produced by a variety of plant species, including several coniferous pines, which have fungicidal properties and act as repellants to insects such as the mountain pine beetle (*Dendroctonus ponderosae*) (Wang et al., 2014). The composition of limonene and pinene in the resin of the Ponderosa pine (*Pinus ponderosa*) varies clinally across the Southwestern United States and in the overlapping species range of *D. pseudoobscura* (Smith, 1977). Furthermore, *D. ponderosae* has been shown to differentially colonize pines depending on the composition of monoterpenes, indicating that the species has evolved the ability to preferentially recognize the relative abundance of limonene and pinene (Thoss & Byers, 2006). A number of significant differentially expressed genes between *D. pseudoobscura* arrangements are also involved in limonene and pinene degradation (Fuller et al. 2016). Despite the wealth of genetic information supported by nearly a century of research in *D. pseudoobscura*, little is known regarding the species' ecology and life history in nature. Our results here may suggest a link between the environment of *D. pseudoobscura* and genes showing evidence of adaptive evolution involved in sensory perception, metabolism and limonene and pinene degradation.

Discussion

Test of the Position Effect Hypothesis. The breakpoints examined here reject the position effect hypothesis in a narrow sense because none of the breaks disrupted the coding sequences of genes. Breakpoints may also generate position effects by altering the expression of boundary genes. Fuller et al. (2016) found only one case where a gene immediately adjacent to a

breakpoint was differentially expressed. The gene GA22082, which is adjacent to the nearly coincident pHYST and dSTPP breakpoints, is expressed higher in ST than in PP chromosomes in larvae. Inversion events could also disrupt gene expression beyond breakpoints by altering topologically associated domains (TADs) (Hou, Li, Qin, & Corces, 2012). TADs partition the genome into structural domains within the nucleus that may be important for the coordination of gene expression. We inferred TADs in *D. pseudoobscura* based on synteny with TADs in *D. melanogaster* (Hou et al., 2012) (unpublished data). The 14 inversion breakpoints involve 20 *D. melanogaster* TADs. Five breakpoints (pHYSC, dSCCU, pSTAR, pSCTL, and dSTAR) disrupt a TAD and contain one or two differentially expressed genes (max: 2). The ST to AR event is particularly interesting because it splits a 79 kb TAD in half and leads to differential expression of two genes within the TAD on either side of the break. These results are intriguing, but should be viewed with caution because we do not have experimental evidence to suggest that TAD structure has been conserved between *D. melanogaster* and *D. pseudoobscura*.

There are other possible mechanisms for inversions to directly generate variation in the population. Novitski (1967) suggested that crossing over in inversion heterozygotes could set up conditions for meiotic drive to operate. Certain cross over events could lead to chromatids that differ in length such that shorter chromatids are transmitted at higher frequencies. Finally, Corbett-Detig (2016) showed that large inversions are often in close proximity to “sensitive sites” in the *D. melanogaster* genome. Sensitive sites normally promote crossing over, however, an inversion can lead reduce levels of genetic exchange when near the sensitive sites. Corbett-Detig (2016) concluded this can lead to a selective advantage for the inversion. Here, we conclude that breakpoint lesions have had minimal effects in generating phenotypic variation through the structural or expression alterations of boundary genes for selection to act on,

552 however, further studies that map *D. pseudoobscura* TADs in different arrangements, test for
553 meiotic drive in heterozygotes, and map sensitive sites in *D. pseudoobscura* are needed to fully
554 reject the position effect hypothesis.

555 **Gene Phylogenies Support the Unique Origin of the Gene Arrangements.** The
556 phylogeny based on SNP variation across the third chromosome mirrors the history of the
557 inversion mutations. Phylogenetic analysis supports the unique origin of the different
558 arrangements except for regions near the centromere and telomere where recombination is
559 prevalent. The expected phylogeny of the different arrangements is largely consistent with the
560 cytogenetic phylogeny (Dobzhansky & Sturtevant, 1938), although some syntenic regions show
561 an aberrant clustering of PP with TL.

562 One possible explanation for aberrant cluster of PP and TL is that there was an ancestral
563 genetic exchange event. PP and TL have 739 SNPs that are shared between the arrangements in
564 our samples. Although these regions are located within inverted regions of PP/TL heterozygotes,
565 there is a 6.3 Mb region that can pair in the heterokaryotypes including regions (68C-69D, 69D-
566 70A, 70B-70C, and 70D-74B). PP and TL do occur in the same populations and
567 heterokaryotypes can form at appreciable frequencies allowing for recombinants. Indeed, PP/TL
568 heterokaryotypes frequencies have been recorded as high as 32% in Mexico (Salceda, Guzman-
569 Rincon, Rosa, & Olvera, 2015). The size of the region affected suggests that perhaps a double
570 cross over event generated shared variation between the two arrangements (Arcadio Navarro et
571 al., 1997) and the exchanged segment may have transferred adaptive genes based on a number of
572 outlier genes that map to this segment.

573 **Evolutionary Forces for the Establishment and Maintenance of New Inversions.**
574 Here, we provide novel insights into the evolutionary forces that establish new inversion

mutations in populations. If a new inversion increases by strictly neutral forces, as the arrangement increases in frequency, it will carry the allelic variants initially captured from the ancestral arrangement along with it and lead to extensive LD across the inverted region. Over time, LD of SNPs associated with the gene arrangement will decrease through genetic exchange except nearest to the breakpoints (Arcadio Navarro, Barbadilla, & Ruiz, 2000; Arcadio Navarro et al., 1997; Peischl, Koch, Guerrero, & Kirkpatrick, 2013). How much decay depends on the size of the inversion with larger inversions showing a greater reduction in LD in central regions than smaller inversions (Arcadio Navarro et al., 1997). The ages of the *D. pseudoobscura* arrangements are 0.5 to 1.4 million years old, which should be sufficient time for LD of variation to decay except near the breakpoints.

We can rule out genetic drift as an explanation for the inversion polymorphism in *D. pseudoobscura*. If genetic drift is responsible for the establishment and maintenance of the arrangements, their frequency is expected to be related to their relative ages (Kimura & Ohta, 1973), leading to higher levels of variation (Arcadio Navarro et al., 2000) and lower levels of LD (Hartl & Clark, 1997; Toomajian, Ajioka, Jorde, Kushner, & Kreitman, 2003). The strongest evidence against genetic drift is the pattern of variation in the AR chromosome. AR is the most frequent arrangement in the Southwestern United States suggesting that it is one of the oldest chromosomes, yet AR is relatively young because it was derived from the ST arrangement. This implies that AR emerged from the ST chromosome and has rapidly spread from California to Texas. The average value of Tajima's $D|D_{min}|$ across AR is more negative than any other arrangement, which is consistent with a recent expansion. A similar observation is found in the closely related species *D. subobscura*, where allele frequencies of microsatellite markers have been shown to increase more than expected under genetic drift alone likely due to a recent

598 expansion (J. Santos et al., 2016). The lack of significant windows of extended homozygosity in
599 AR suggests that even though this arrangement is relatively young, there has been sufficient time
600 for gene conversion break up LD. Additionally, genetic drift can be ruled out because each
601 arrangement has multiple outlier genes which are spread across the inverted region. One would
602 not expect outlier genes in the central region of the inversion because there has been sufficient
603 time for gene conversion and crossovers to degrade LD.

604 Next, we can rule out a single beneficial sweeping allele because we find evidence for not
605 one, but multiple outlier genes within the gene arrangements showing evidence of adaptive
606 evolution. We find evidence of multiple genes with long branch lengths or elevated frequencies
607 of derived mutations broken up by regions with high frequencies of shared polymorphisms. This
608 is consistent with previous results showing multiple significantly differentially expressed genes
609 between third chromosome arrangements maintained by the inversions (Fuller et al., 2016).
610 Multiple loci are also thought to be functionally important for adaptive inversion clines in *D.*
611 *melanogaster* (Kennington, Partridge, & Hoffmann, 2006).

612 The most likely explanation for the establishment and maintenance of chromosomal
613 inversions in *D. pseudoobscura* is that inversions act as negative modifiers of recombination
614 holding adaptive haplotypes together. By limiting genetic exchange, the original inversion
615 mutation captures sets of linked alleles and maintains their associations. This is supported by the
616 numbers of outlier genes that show evidence of adaptive evolution and the pattern of LD across
617 the chromosome. Our surprising finding was that regions of the chromosome with high levels of
618 LD are interspersed with regions of low LD or high values of ρ. In addition, we observed few
619 regions of unusual haplotype structure or extended homozygosity within arrangements. We
620 hypothesize that gene conversion and double crossovers can break up associations across an

arrangement, but purifying selection removes any maladaptive combinations, consistent with inversion establishment models (Charlesworth & Charlesworth, 1973). These results should be interpreted with caution because some of the shared polymorphism we observe may be due to gene conversion events from balancer chromosomes used to create the isogenic strains (Miller et al., 2016).

Two models are consistent with the establishment of inversions to reduce recombination, the local adaptation and epistasis models, which may not be mutually exclusive. The distinction between these two hypotheses is whether the gene products interact at the molecular level or have a synergistic effect on fitness. Under the local adaptation model, there is no need for the gene products to interact or act synergistically either in a genetic pathway or as transcription factors regulating a common gene (Kirkpatrick & Barton, 2006). The geographic range of *D. pseudoobscura* straddles several physiographic provinces (Lobeck, 1948) and adult flies are capable of dispersing among these ecological niches. Numerical analysis has shown that geographic inversion frequencies of *D. pseudoobscura* are consistent with a model of selection in heterogeneous environments (Schaeffer, 2008). However, it remains unclear why multiple arrangements exist in single populations under the local adaptation model and further work is needed to characterize the local fitness environments in ecological niches

The data presented here supports the hypothesis that the size of the inversion may be an indirectly selected character (Caceres, Barbadilla, & Ruiz, 1997, 1999). The ST arrangement is an exceptional case because the majority of the candidate selected genes are outside of the original derived inversion (HY to ST) and instead found in the proximal region, which overlaps with the AR inversion. Because ST gave rise to AR and AR remained in populations with ST, it appears that new adaptive genes in ST have become fixed from standing variation after the HY

644 to ST inversion event. We hypothesize this resulted from selection on standing variation within
645 ST (Wyatt W. Anderson et al., 1991; Dobzhansky, 1944) and suppression of recombination due
646 to the presence of AR and CH in populations with ST. Thus, our data suggest that an initial
647 inversion event can lock up adaptive combinations of genes within the inverted segment, but
648 additional multi-locus combinations can be established outside of the initial inversion
649 breakpoints based on what other arrangements are present in the population.

650 If double cross overs are responsible for breaking up LD, one would expect larger regions
651 of sharing between arrangements (Arcadio Navarro et al., 1997), especially between ST and PP.
652 The ST to PP inversion corresponds to 52% of the length of the third chromosome and one
653 would expect seven percent of gametes from a ST/PP heterozygote to produce double cross overs
654 leading to more similarity within their inverted regions. Homogenization, however, is unlikely
655 because ST and PP do not co-occur at appreciable frequencies in nature. Wallace (1953)
656 proposed the triad model where coadapted gene complexes could be eroded when sets of
657 arrangements separated by single inversion steps such as AR-ST-PP are present in the same
658 population. Therefore, it is reasonable to speculate that all members of a triad do not co-occur
659 for this reason. ST and PP support the triad model because of the non-trivial potential to
660 generate maladaptive recombinant haplotypes.

661 There is a significant enrichment of genes involved in sensory perception and odorant
662 binding showing evidence of adaptive evolution. Analyses of insect genomes have found that
663 odorant perception and detoxification proteins are typically found to be amplified in copy
664 numbers and show evidence of recent positive selection (Chen et al., 2015; Nene et al., 2007;
665 Scott et al., 2014; Smadja, Shi, Butlin, & Robertson, 2009; The Honeybee Genome Sequencing
666 Consortium, 2006; The International Aphid Genomics Consortium, 2010; Tribolium Genome

667 Sequencing Consortium, 2008; You et al., 2013). Our study further suggests these genes may be
668 targets of selection in insect species because they can play a role in adapting to complex
669 environments. Furthermore, several of the genes showing evidence of adaptive evolution are
670 members of the limonene and pinene degradation pathway. It is intriguing to speculate that
671 ponderosa pine is part of the ecology of *D. pseudoobscura* either because the adults are
672 associated with bark of the tree or larvae feed on the pine nuts.

673 The establishment of inversion clines may play a role in the speciation process. In such a
674 system, the emergence of Dobzhansky-Muller incompatibility genes within one arrangement
675 background can easily spread within the inversion type, but not to other chromosomal
676 backgrounds (A. Navarro & Barton, 2003). This may have been how *D. persimilis* was formed
677 (Noor, Grams, Bertucci, & Reiland, 2001). The distribution of *D. persimilis* is sympatric with *D.*
678 *pseudoobscura* and the two species differ by several fixed inversion differences that include
679 reproductive isolation genes. Thus, *D. pseudoobscura* may be ripe for additional speciation
680 events.

681 The evolution of different gene arrangements in *D. pseudoobscura* may be in response to
682 life history and environmental challenges that require many genes of small effect to adapt to a
683 heterogeneous environment. Independent assortment and recombination play a vital role in
684 helping an organism to generate multiple combinations for selection to sift through. While this is
685 an advantage, it is also a curse because adaptive combinations of genes can be split apart as
686 quickly as they are made. In this system, it seems that chromosomal inversions help to hold
687 adaptive combinations together.

Acknowledgements

Ian S. Leopold for preparation of polytene chromosome squashes, digital photography of the chromosomes, and assembly of the mosaic images of the chromosomes. Andre G. Wallace for sequencing the 18 regions used to validate the Illumina sequences. Richard Kliman at Cedar Crest College who kindly provided the SNP data for the *D. pseudoobscura* second chromosome. The authors would like to thank the staff at the Baylor College of Medicine Human Genome Sequencing Center for their technical and computational contributions. This work was supported by a grant from the National Institute for General Medical Sciences at the National Institutes of Health R01 GM098478 to S.W.S. and the Penn State-National Institutes of Health funded Computation, Bioinformatics and Statistics (CBIOS) Predoctoral Training Program to Z.L.F.

References

Anderson, A. R., Hoffmann, A. A., McKechnie, S. W., Umina, P. A., & Weeks, A. R. (2005). The latitudinal cline in the *In*(*3R*)*Payne* inversion polymorphism has shifted in the last 20 years in Australian *Drosophila melanogaster* populations. *Mol Ecol, 14*(3), 851-858.

Anderson, W. W., Arnold, J., Baldwin, D. G., Beckenbach, A. T., Brown, C. J., Bryant, S. H., . . . Moore, J. A. (1991). Four decades of inversion polymorphism in *Drosophila pseudoobscura*. *Proceedings of the National Academy of Sciences USA, 88*, 10367-10371.

Anderson, W. W., Ayala, F. J., & Michod, R. E. (1977). Chromosomal and allozymic diagnosis of three species of *Drosophila*, *Drosophila pseudoobscura*, *Drosophila persimilis*, and *Drosophila miranda*. *Journal of Heredity, 68*, 71-74.

Anderson, W. W., Dobzhansky, T., & Kastritsis, C. D. (1967). Selection and inversion polymorphism in experimental populations of *Drosophila pseudoobscura* initiated with the chromosomal constitutions of natural populations. *Evolution, 21*, 664-671.

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

Arya, G. H., Magwire, M. M., Huang, W., Serrano-Negron, Y. L., Mackay, T. F. C., & Anholt, R. R. H. (2015). The genetic basis for variation in olfactory behavior in *Drosophila melanogaster*. *Chemical Senses, 40*(4), 233-243.

Ayala, D., Acevedo, P., Pombi, M., Dia, I., Boccolini, D., Costantini, C., . . . Fontenille, D. (2017). Chromosome inversions and ecological plasticity in the main African malaria mosquitoes. *Evolution, 71*(3), 686-701.

Ayala, D., Fontaine, M. C., Cohuet, A., Fontenille, D., Vitalis, R., & Simard, F. (2011). Chromosomal inversions, natural selection and adaptation in the malaria vector *Anopheles funestus*. *Mol Biol Evol, 28*(1), 745-758.

Balanya, J., Serra, L., Gilchrist, G. W., Huey, R. B., Pascual, M., Mestres, F., & Sole, E. (2003). Evolutionary pace of chromosomal polymorphism in colonizing populations of Drosophila subobscura: an evolutionary time series. *Evolution, 57*(8), 1837-1845.

728 Becks, L., & Agrawal, A. F. (2010). Higher rates of sex evolve in spatially heterogeneous
729 environments. *Nature, 468*(7320), 89-92.
730 Becks, L., & Agrawal, A. F. (2012). The evolution of sex Is favoured during adaptation to new
731 environments. *PLoS Biol, 10*(5), e1001317.
732 Bhutkar, A., Schaeffer, S. W., Russo, S., Xu, M., Smith, T. F., & Gelbart, W. M. (2008).
733 Chromosomal rearrangement inferred from comparisons of twelve Drosophila genomes.
734 *Genetics, 179*, 1657-1680.
735 Caceres, M., Barbadilla, A., & Ruiz, A. (1997). Inversion length and breakpoint distribution in
736 the *Drosophila buzzatii* species complex: Is inversion length a selected trait? *Evolution,*
737 *51*(4), 1149-1155.
738 Caceres, M., Barbadilla, A., & Ruiz, A. (1999). Recombination rate predicts inversion size in
739 Diptera. *Genetics, 153*(1), 251-259.
740 Calboli, F. C., Kennington, W. J., & Partridge, L. (2003). QTL mapping reveals a striking
741 coincidence in the positions of genomic regions associated with adaptive variation in
742 body size in parallel clines of *Drosophila melanogaster* on different continents.
743 *Evolution, 57*(11), 2653-2658.
744 Calvete, O., Gonzalez, J., Betran, E., & Ruiz, A. (2012). Segmental duplication, microinversion,
745 and gene loss associated with a complex inversion breakpoint region in Drosophila. *Mol*
746 *Biol Evol, 29*(7), 1875-1889.
747 Carbone, L., Harris, R. A., Gnerre, S., Veeramah, K. R., Lorente-Galdos, B., Huddleston, J., . . .
748 Gibbs, R. A. (2014). Gibbon genome and the fast karyotype evolution of small apes.
749 *Nature, 513*(7517), 195-201.
750 Carson, H. L. (1992). Inversions in Hawaiian *Drosophila*. In C. Krimbas & J. R. Powell (Eds.),
751 *Drosophila Inversion Polymorphism* (pp. 407-439). Boca Raton, FL: CRC Press.
752 Charlesworth, B., & Charlesworth, D. (1973). Selection of new inversion in multi-locus genetic
753 systems. *Genetical Research, 21*(2), 167-183.
754 Chen, X.-G., Jiang, X., Gu, J., Xu, M., Wu, Y., Deng, Y., . . . James, A. A. (2015). Genome
755 sequence of the Asian tiger mosquito, *Aedes albopictus*, reveals insights into its biology,
756 genetics, and evolution. *Proceedings of the National Academy of Sciences, 112*(44),
757 E5907-E5915.
758 Cheng, C., White, B. J., Kamdem, C., Mockaitis, K., Costantini, C., Hahn, M. W., & Besansky,
759 N. J. (2012). Ecological genomics of *Anopheles gambiae* along a latitudinal cline: A
760 population-resequencing approach. *Genetics, 190*(4), 1417-1432.
761 Coluzzi, M., Sabatini, A., della Torre, A., Di Deco, M. A., & Petrarca, V. (2002). A polytene
762 chromosome analysis of the *Anopheles gambiae* species complex. *Science, 298*(5597),
763 1415-1418.
764 Coluzzi, M., Sabatini, A., Petrarca, V., & Di Deco, M. A. (1979). Chromosomal differentiation
765 and adaptation to human environments in the *dob* complex. *Transactions of The Royal*
766 *Society of Tropical Medicine and Hygiene, 73*(5), 483-497.
767 Corbett-Detig, R. B. (2016). Selection on inversion breakpoints favors proximity to pairing
768 sensitive sites in *Drosophila melanogaster*. *Genetics, 204*(1), 259-265.
769 Corbett-Detig, R. B., Cardeno, C., & Langley, C. H. (2012). Sequence-based detection and
770 breakpoint assembly of polymorphic inversions. *Genetics, 192*(1), 131-137.
771 Coyne, J. A., Bryant, S. H., & Turelli, M. (1987). Long-distance migration of *Drosophila*. 2.
772 Presence in desolate sites and dispersal near a desert oasis. *American Naturalist, 129*,
773 847-861.

774 Cridland, J. M., & Thornton, K. R. (2010). Validation of rearrangement break points identified
775 by paired-end sequencing in natural populations of *Drosophila melanogaster*. *Genome*
776 *Biol Evol, 2*, 83-101.
777 DePristo, M. A., Banks, E., Poplin, R., Garimella, K. V., Maguire, J. R., Hartl, C., . . . Daly, M.
778 J. (2011). A framework for variation discovery and genotyping using next-generation
779 DNA sequencing data. *Nat Genet, 43*(5), 491-498.
780 Dobigny, G., Britton-Davidian, J., & Robinson, T. J. (2015). Chromosomal polymorphism in
781 mammals: an evolutionary perspective. *Biological Reviews, 92*, 1-21.
782 Dobzhansky, T. (1943). Genetics of natural populations. IX. Temporal changes in the
783 composition of populations of *Drosophila pseudoobscura*. *Genetics, 28*, 162-186.
784 Dobzhansky, T. (1944). Chromosomal races in *Drosophila pseudoobscura* and *Drosophila*
785 *persimilis*. *Carnegie Inst. Washington Publ., 554*, 47-144.
786 Dobzhansky, T. (1948a). Genetics of natural populations. XVI. Altitudinal and seasonal changes
787 in certain populations of *Drosophila pseudoobscura* and *Drosophila persimilis*. *Genetics,*
788 *33*, 158-176.
789 Dobzhansky, T. (1948b). Genetics of natural populations. XVIII. Experiments on chromosomes
790 of *Drosophila pseudoobscura* from different geographic regions. *Genetics, 33*, 588-602.
791 Dobzhansky, T. (1950). The genetics of natural populations. XIX. Origin of heterosis through
792 natural selection in populations of *Drosophila pseudoobscura*. *Genetics, 35*, 288-302.
793 Dobzhansky, T., & Sturtevant, A. H. (1938). Inversions in the chromosomes of *Drosophila*
794 *pseudoobscura*. *Genetics, 23*, 28-64.
795 dos Santos, G., Schroeder, A. J., Goodman, J. L., Strelets, V. B., Crosby, M. A., Thurmond, J., . .
796 . Consortium, t. F. (2015). FlyBase: Introduction of the *Drosophila melanogaster* Release
797 6 reference genome assembly and large-scale migration of genome annotations. *Nucleic*
798 *Acids Research, 43*(D1), D690-D697.
799 Engelbrecht, A., Taylor, P. J., Daniels, S. R., & Rambau, R. V. (2011). Chromosomal
800 polymorphisms in African Vlei Rats, *Otomys irroratus* (Muridae: Otomyini), detected by
801 banding techniques and chromosome painting: inversions, centromeric shifts and diploid
802 number variation. *Cytogenetic and Genome Research, 133*(1), 8-15.
803 Felsenstein, J. (1985). Confidence limits on phylogenies: An approach using the bootstrap.
804 *Evolution, 39*, 783-791.
805 Fontaine, M. C., Pease, J. B., Steele, A., Waterhouse, R. M., Neafsey, D. E., Sharakhov, I. V., . .
806 . Besansky, N. J. (2015). Extensive introgression in a malaria vector species complex
807 revealed by phylogenomics. *Science, 347*(6217).
808 Fuller, Z. L., Haynes, G. D., Richards, S., & Schaeffer, S. W. (2016). Genomics of natural
809 populations: How differentially expressed genes shape the evolution of chromosomal
810 inversions in *Drosophila pseudoobscura*. *Genetics, 204*, 287-301.
811 Fuller, Z. L., Haynes, G. D., Zhu, D., Batterton, M., Chao, H., Dugan, S., . . . Schaeffer, S. W.
812 (2014). Evidence for stabilizing selection on codon usage in chromosomal
813 rearrangements of *Drosophila pseudoobscura*. *G3: Genes|Genomes|Genetics, 4*(12),
814 2433-2449.
815 Garrison, E., & Marth, G. (2012). Haplotype-based variant detection from short-read sequencing.
816 *arXiv:1207.3907v2*.
817 Guillen, Y., & Ruiz, A. (2012). Gene alterations at Drosophila inversion breakpoints provide
818 prima facie evidence for natural selection as an explanation for rapid chromosomal
819 evolution. *BMC Genomics, 13*, 53.

Hahn, M. W., Rausher, M. D., & Cunningham, C. W. (2002). Distinguishing between selection and population expansion in an experimental lineage of bacteriophage T7. *Genetics, 161*(1), 11-20.

Hartl, D. L., & Clark, A. G. (1997). *Principles of Population Genetics*. Sunderland, MA: Sinauer Associates, Inc.

Hill, W. G., & Robertson, A. (1968). Linkage disequilibrium in finite populations. *Theoretical and Applied Genetics, 38*, 226-231.

Hintze, J. L., & Nelson, R. D. (1998). Violin plots: A box plot-density trace synergism. *The American Statistician, 52*(2), 181-184.

Hoffmann, A. A., & Rieseberg, L. H. (2008). Revisiting the impact of inversions in evolution: From population genetic markers to drivers of adaptive shifts and speciation? *Annual Review of Ecology and Systematics, 39*, 21-42.

Hou, C., Li, L., Qin, Zhaohui S., & Corces, Victor G. (2012). Gene density, transcription, and insulators contribute to the partition of the Drosophila genome into physical domains. *Molecular Cell, 48*(3), 471-484.

Huang, D. W., Sherman, B. T., & Lempicki, R. A. (2009a). Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Research, 37*(1), 1-13.

Huang, D. W., Sherman, B. T., & Lempicki, R. A. (2009b). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protocols, 4*(1), 44-57.

Hudson, R. R., & Kaplan, N. L. (1986). On the divergence of alleles in nested subsamples from finite populations. *Genetics, 113*, 1057-1076.

Huerta-Sanchez, E., Degiorgio, M., Pagani, L., Tarekegn, A., Ekong, R., Antao, T., . . . Nielsen, R. (2013). Genetic signatures reveal high-altitude adaptation in a set of ethiopian populations. *Mol Biol Evol, 30*(8), 1877-1888.

Kapun, M., Fabian, D. K., Goudet, J., & Flatt, T. (2016). Genomic evidence for adaptive inversion clines in *Drosophila melanogaster*. *Molecular Biology and Evolution, 33*(5), 1317-1336.

Kapun, M., Schmidt, C., Durmaz, E., Schmidt, P. S., & Flatt, T. (2016). Parallel effects of the inversion *In*(*3R*)*Payne* on body size across the North American and Australian clines in *Drosophila melanogaster*. *Journal of Evolutionary Biology, 29*, 1059-1072.

Kapun, M., van Schalkwyk, H., McAllister, B., Flatt, T., & Schlotterer, C. (2014). Inference of chromosomal inversion dynamics from Pool-Seq data in natural and laboratory populations of *Drosophila melanogaster*. *Mol Ecol, 23*(7), 1813-1827.

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

Kimura, M., & Ohta, T. (1973). The age of a neutral mutant persisting in a finite population. *Genetics, 75*, 199-212.

Kirkpatrick, M., & Barton, N. (2006). Chromosome inversions, local adaptation and speciation. *Genetics, 173*(1), 419-434.

Knibb, W. R. (1982). Chromosome inversion polymorphisms in *Drosophila melanogaster* II. Geographic clines and climatic associations in Australasia, North America and Asia. *Genetica, 58*(3), 213-221.

Knibb, W. R., Oakeshott, J. G., & Gibson, J. B. (1981). Chromosome inversion polymorphisms in *Drosophila melanogaster*. I. Latitudinal clines and associations between inversions in Australasian populations. *Genetics, 98*(4), 833-847.

Kolaczkowski, B., Kern, A. D., Holloway, A. K., & Begun, D. J. (2011). Genomic differentiation between temperate and tropical Australian populations of *Drosophila melanogaster*. *Genetics, 187*(1), 245-260.

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

Kupper, C., Stocks, M., Risse, J. E., dos Remedios, N., Farrell, L. L., McRae, S. B., . . . Burke, T. (2016). A supergene determines highly divergent male reproductive morphs in the ruff. *Nat Genet, 48*(1), 79-83.

Lande, R. (1984). The expected fixation rate of chromosomal inversions. *Evolution, 38*, 743-752.

Langley, C. H., Stevens, K., Cardeno, C., Lee, Y. C. G., Schrider, D. R., Pool, J. E., . . . Begun, D. J. (2012). Genomic variation in natural populations of *Drosophila melanogaster*. *Genetics, 192*(2), 533-598.

Lee, Y. W., Fishman, L., Kelly, J. K., & Willis, J. H. (2016). Fitness Variation Is Generated by a Segregating Inversion in Yellow Monkeyflower (*Mimulus guttatus*). *Genetics*.

Levitan, M. (1992). The inversion polymorphism of *Drosophila robusta*. In C. B. Krimbas (Ed.), *Drosophila Inversion Polymorphism* (pp. 221-338). Boca Raton: CRC Press.

Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics, 25*(14), 1754-1760.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., . . . Subgroup, G. P. D. P. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics, 25*(16), 2078-2079.

Lobeck, A. K. (Cartographer). (1948). Physiographic diagram of North America

Mackay, T. F. C., Richards, S., Stone, E. A., Barbadilla, A., Ayroles, J. F., Zhu, D., . . . Gibbs, R. A. (2012). The *Drosophila melanogaster* genetic reference panel. *Nature, 482*(7384), 173-178.

Maynard Smith, J., & Haigh, J. (1974). The hitch-hiking effect of a favorable gene. *Genetical Research, 23*, 23-35.

McAllister, B. F., Sheeley, S. L., Mena, P. A., Evans, A. L., & Schlotterer, C. (2008). Clinal distribution of a chromosomal rearrangement: A precursor to chromosomal speciation? *Evolution, 62*(8), 1852-1865.

McDonald, M. J., Rice, D. P., & Desai, M. M. (2016). Sex speeds adaptation by altering the dynamics of molecular evolution. *Nature, 531*(7593), 233-236.

McGaugh, S. E., Heil, C. S. S., Manzano-Winkler, B., Loewe, L., Goldstein, S., Himmel, T. L., & Noor, M. A. F. (2012). Recombination modulates how selection affects linked sites in *Drosophila*. *PLoS Biol, 10*(11), e1001422.

McGraw, L. A., Davis, J. K., Young, L. J., & Thomas, J. W. (2011). A genetic linkage map and comparative mapping of the prairie vole (*Microtus ochrogaster*) genome. *BMC Genetics, 12*(1), 1-10.

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., . . . DePristo, M. A. (2010). The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res, 20*(9), 1297-1303.

Mettler, L. E., Voelker, R. A., & Mukai, T. (1977). Inversion clines in populations of *Drosophila melanogaster*. *Genetics, 87*, 169-176.

Miller, D. E., Cook, K. R., Yeganeh Kazemi, N., Smith, C. B., Cockrell, A. J., Hawley, R. S., & Bergman, C. M. (2016). Rare recombination events generate sequence diversity among balancer chromosomes in *Drosophila melanogaster*. *Proceedings of the National Academy of Sciences, 113*(10), E1352-E1361.

Muller, H. J. (1940). Bearings of the 'Drosophila' work on systematics. In J. Huxley (Ed.), *The New Systematics* (pp. 185-268). Oxford: Clarendon Press.

Nachman, M. W., Boyer, S. N., Searle, J. B., & Aquadro, C. F. (1994). Mitochondrial DNA variation and the evolution of Robertsonian chromosomal races of house mice, *Mus domesticus*. *Genetics, 136*(3), 1105-1120.

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

Navarro, A., & Barton, N. H. (2003). Accumulating postzygotic isolation genes in parapatry: a new twist on chromosomal speciation. *Evolution, 57*(3), 447-459.

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

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

Nene, V., Wortman, J. R., Lawson, D., Haas, B., Kodira, C., Tu, Z. J., . . . Severson, D. W. (2007). Genome sequence of *Aedes aegypti*, a major arbovirus vector. *Science, 316*(5832), 1718-1723.

Nishikawa, H., Iijima, T., Kajitani, R., Yamaguchi, J., Ando, T., Suzuki, Y., . . . Fujiwara, H. (2015). A genetic mechanism for female-limited Batesian mimicry in Papilio butterfly. *Nat Genet, 47*(4), 405-409.

Noor, M. A., Grams, K. L., Bertucci, L. A., & Reiland, J. (2001). Chromosomal inversions and the reproductive isolation of species. *Proceedings of the National Academy of Sciences USA, 98*(21), 12084-12088.

Novitski, E. (1951). Non-random disjunction in Drosophila. *Genetics, 36*, 267-280.

Novitski, E. (1967). Nonrandom disjunction in Drosophila. *Annual Review of Genetics, 1*, 71-86.

O'Neill, R. J., Eldridge, M. D., & Metcalfe, C. J. (2004). Centromere dynamics and chromosome evolution in marsupials. *J Hered, 95*(5), 375-381.

O'Neill, R. J., Eldridge, M. D., Toder, R., Ferguson-Smith, M. A., O'Brien, P. C., & Graves, J. A. (1999). Chromosome evolution in kangaroos (Marsupialia: Macropodidae): cross species chromosome painting between the tammar wallaby and rock wallaby spp. with the 2n = 22 ancestral macropodid karyotype. *Genome, 42*(3), 525-530.

Ohta, T. (1971). Associative overdominance caused by linked deleterious mutations. *Genetical Research, 18*, 227-286.

Otto, S. P., & Lenormand, T. (2002). Resolving the paradox of sex and recombination. *Nat Rev Genet, 3*(4), 252-261.

Papaceit, M., Segarra, C., & Aguadé, M. (2013). Structure and population genetics of the breakpoints of a polymorphic inversion in *Drosophila subobscura*. *Evolution, 67*(1), 66-79.

Peischl, S., Koch, E., Guerrero, R. F., & Kirkpatrick, M. (2013). A sequential coalescent algorithm for chromosomal inversions. *Heredity*.

Pettenati, M. J., Rao, P. N., Phelan, M. C., Grass, F., Rao, K. W., Cosper, P., . . . Palmer, C. G. (1995). Paracentric inversions in humans: a review of 446 paracentric inversions with presentation of 120 new cases. *American Journal of Medical Genetics, 55*(2), 171-187.

Powell, J. R. (1992). Inversion polymorphisms in *Drosophila pseudoobscura* and *Drosophila persimilis*. In C. B. Krimbas & J. R. Powell (Eds.), *Drosophila Inversion Polymorphism* (pp. 73-126). Ann Arbor, MI: CRC Press.

Prakash, S., & Lewontin, R. C. (1968). A molecular approach to the study of genic heterozygosity in natural populations, III. Direct evidence of coadaptation in gene arrangements of Drosophila. *Proceedings of the National Academy of Sciences USA, 59*, 398-405.

Prakash, S., & Lewontin, R. C. (1971). A molecular approach to the study of genic heterozygosity in natural populations. V. Further direct evidence of coadaptation in inversions of Drosophila. *Genetics, 69*, 405-408.

Puerma, E., Orengo, D. J., Salguero, D., Papaceit, M., Segarra, C., & Aguadé, M. (2014). Characterization of the breakpoints of a polymorphic inversion complex detects strict and broad breakpoint reuse at the molecular level. *Molecular Biology and Evolution, 31*(9), 2331-2341.

Pyhäjärvi, T., Hufford, M. B., Mezmouk, S., & Ross-Ibarra, J. (2013). Complex patterns of local adaptation in Teosinte. *Genome Biology and Evolution, 5*(9), 1594-1609.

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

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

Riley, M. A., Hallas, M. E., & Lewontin, R. C. (1989). Distinguishing the forces controlling genetic variation at the *Xdh* locus in *Drosophila pseudoobscura*. *Genetics, 123*, 359-369.

Saitou, N., & Nei, M. (1987). The neighbor-joining method: A new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution, 4*, 406-425.

Salceda, V. M., Guzman-Rincon, J., Rosa, M. E. d. l., & Olvera, O. (2015). Four decades of inversion polymorphism in *Drosophila pseudoobscura* from Mexico. *Genetika, 47*(3), 959-966.

Santos, J., Pascual, M., Fragata, I., Simões, P., Santos, M. A., Lima, M., . . . Matos, M. (2016). Tracking changes in chromosomal arrangements and their genetic content during adaptation. *Journal of Evolutionary Biology, 29*, 1151-1167.

Santos, M., Cespedes, W., Balanya, J., Trotta, V., Calboli, F. C., Fontdevila, A., & Serra, L. (2005). Temperature-related genetic changes in laboratory populations of *Drosophila subobscura*: evidence against simple climatic-based explanations for latitudinal clines. *Am Nat, 165*(2), 258-273.

Schaeffer, S. W. (2002). Molecular population genetics of sequence length diversity in the *Adh* region of *Drosophila pseudoobscura*. *Genetical Research, 80*, 163-175.

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

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

Schaeffer, S. W., Bhutkar, A., McAllister, B. F., Matsuda, M., Matzkin, L. M., Grady, P. M. O., . . . Kaufman, T. C. (2008). Polytene chromosomal maps of 11 Drosophila species: The order of genomic scaffolds inferred from genetic and physical maps. *Genetics, 179*, 1601-1655.

Schaeffer, S. W., Goetting-Minesky, P., Kovacevic, M., Peoples, J., Graybill, J. L., Miller, J. M., . . . Anderson, W. W. (2003). Evolutionary genomics of inversions in *Drosophila pseudoobscura*: Evidence for epistasis. *Proceedings of the National Academy of Sciences USA, 100*, 8319-8324.

Schaeffer, S. W., & Miller, E. L. (1992). Estimates of gene flow in *Drosophila pseudoobscura* determined from nucleotide sequence analysis of the alcohol dehydrogenase region. *Genetics, 132*, 471-480.

Scott, J. G., Warren, W. C., Beukeboom, L. W., Bopp, D., Clark, A. G., Giers, S. D., . . . Liu, N. (2014). Genome of the house fly, *Musca domestica* L., a global vector of diseases with adaptations to a septic environment. *Genome Biology, 15*(10), 1-17.

Shriver, M., Kennedy, G., Parra, E., Lawson, H., Sonpar, V., Huang, J., . . . Jones, K. (2004). The genomic distribution of population substructure in four populations using 8,525 autosomal SNPs. *Human Genomics, 1*(4), 274 - 286.

Silva, M., Ribeiro, E. D., Matoso, D. A., Sousa, L. M., Hrbek, T., Py-Daniel, L. R., & Feldberg, E. (2014). Chromosomal polymorphism in two species of Hypancistrus (Siluriformes: Loricariidae): an integrative approach for understanding their biodiversity. *Genetica, 142*(2), 127-139.

Smadja, C., Shi, P., Butlin, R. K., & Robertson, H. M. (2009). Large gene family expansions and adaptive evolution for odorant and gustatory receptors in the pea aphid, *Acyrthosiphon pisum*. *Mol Biol Evol, 26*(9), 2073-2086.

Smith, R. H. (1977). *Monoterpenes of Ponderosa Pine Xylem Resin in Western United States*: USDA Forest Service Technical Bulletin 1532.

Soto, I. M., Soto, E. M., Carreira, V. P., Hurtado, J., Fanara, J. J., & Hasson, E. (2010). Geographic patterns of inversion polymorphism in the second chromosome of the cactophilic *Drosophila buzzatii* from northeastern Argentina. *J Insect Sci, 10*, 181.

Sperlich, D. (1959). [Experimental contributions to the problem of the positive heterosis effect in the structural polymorphous species *Drosophila subobscura*]. *Z Vererbungsl, 90*, 273-287.

Sperlich, D., & Pfriem, P. (1986). Chromosomal polymorphism in natural and experimental populations. In M. Ashburner, H. L. Carson, & J. N. Thomson (Eds.), *The Genetics and Biology of Drosophila 3e* (Vol. 3e, pp. 257-309). New York, NY: Academic Press.

Stefansson, H., Helgason, A., Thorleifsson, G., Steinthorsdottir, V., Masson, G., Barnard, J., . . . Stefansson, K. (2005). A common inversion under selection in Europeans. *Nature Genetics, 37*(2), 129-137.

Storey, J. D. (2002). A direct approach to false discovery rates. *Journal of the Royal Statistical Society, Series B, 64*, 479-498.

Storey, J. D. (2003). The positive false discovery rate: A Bayesian interpretation and the q-value. *Annals of Statistics, 31*, 2013-2035.

Sturtevant, A. H. (1926). A crossover reducer in Drosophila melanogaster due to inversion of a section of the third chromosome. *Biologische Zentralblatt, 46*, 697-702.

Sturtevant, A. H., & Beadle, G. W. (1936). The relations of inversions in the X chromosome of *Drosophila melanogaster* to crossing over and disjunction. *Genetics, 21*, 544-604.

Tajima, F. (1989). Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics, 123*, 585-595.

Tamura, K., Nei, M., & Kumar, S. (2004). Prospects for inferring very large phylogenies by using the neighbor-joining method. *Proc Natl Acad Sci U S A, 101*(30), 11030-11035.

Tamura, K., Stecher, G., Peterson, D., Filipski, A., & Kumar, S. (2013). MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0. *Molecular Biology and Evolution, 30*(12), 2725-2729.

The Honeybee Genome Sequencing Consortium. (2006). Insights into social insects from the genome of the honeybee *Apis mellifera*. *Nature, 443*(7114), 931-949.

The International Aphid Genomics Consortium. (2010). Genome Sequence of the Pea Aphid *Acyrthosiphon pisum*. *PLoS Biol, 8*(2), e1000313.

Thoss, V., & Byers, J. A. (2006). Monoterpene chemodiversity of ponderosa pine in relation to herbivory and bark beetle colonization. *CHEMOECOLOGY, 16*(1), 51-58.

Toomajian, C., Ajioka, R. S., Jorde, L. B., Kushner, J. P., & Kreitman, M. (2003). A method for detecting recent selection in the human genome from allele age estimates. *Genetics, 165*(1), 287-297.

Tribolium Genome Sequencing Consortium. (2008). The genome of the model beetle and pest *Tribolium castaneum*. *Nature, 452*(7190), 949-955.

Wallace, A. G., Detweiler, D., & Schaeffer, S. W. (2011). Evolutionary history of the third chromosome gene arrangements of *Drosophila pseudoobscura* inferred from inversion breakpoints. *Molecular Biology and Evolution, 28*, 2219-2229.

Wallace, B. (1953). On coadaptation in *Drosophila*. *American Naturalist, 87*, 343-358.

Wang, Y., Lim, L., Madilao, L., Lah, L., Bohlmann, J., & Breuil, C. (2014). Gene discovery for enzymes involved in limonene modification or utilization by the mountain pine beetle-associated pathogen *Grosmannia clavigera*. *Appl Environ Microbiol, 80*(15), 4566-4576.

Wasserman, M. (1982). Evolution of the *repleta* group. In M. Ashburner, H. L. Carson, & J. N. Thompson (Eds.), *The Genetics and Biology of Drosophila 3b* (Vol. 3b, pp. 61-139). New York: Academic Press.

Wright, S., & Dobzhansky, T. (1946). Genetics of natural populations. XII. Experimental reproduction of some of the changes caused by natural selection in certain populations of *Drosophila pseudoobscura*. *Genetics, 31*, 125-156.

Yi, X., Liang, Y., Huerta-Sanchez, E., Jin, X., Cuo, Z. X. P., Pool, J. E., . . . Wang, J. (2010). Sequencing of 50 human exomes reveals adaptation to high altitude. *Science, 329*(5987), 75-78.

You, M., Yue, Z., He, W., Yang, X., Yang, G., Xie, M., . . . Wang, J. (2013). A heterozygous moth genome provides insights into herbivory and detoxification. *Nat Genet, 45*(2), 220-225.

Zaykin, D. V., Pudovkin, A., & Weir, B. S. (2008). Correlation-based inference for linkage disequilibrium with multiple alleles. *Genetics, 180*(1), 533-545.

Zinzow-Kramer, W. M., Horton, B. M., McKee, C. D., Michaud, J. M., Tharp, G. K., Thomas, J. W., . . . Maney, D. L. (2015). Genes located in a chromosomal inversion are correlated with territorial song in white-throated sparrows. *Genes, Brain and Behavior, 14*(8), 641-654.

Data Accessibility

DNA Sequences: NCBI SRA: SRX204748-SRX204792, SRX091323, SRX091311, SRX815755, SRX2484948, SRX2484950, SRX2484953, SRX2484955, SRX2484969.

SNP table for the third chromosome containing the sites used for all subsequent analyses is available through Scholarsphere (https://scholarsphere.psu.edu).

Computer Code is available: (https://scholarsphere.psu.edu)

Author Contributions

S.W.S conceived the project. S.R., generated the sequence data. Z.L.F., G.D.H., S.R.and S.W.S, analyzed the data. Z.L.F., G.D.H., S.R., and S.W.S. wrote the manuscript.

Table 1. Breakpoint map locations for seven inversion events in the MV 2-25 reference genome of *D. pseudoobscura.*

Breakpoint	Cyto_Ref	5'_Gene	5'_Coor	3'_Gene	3'_Coor	BP_Interval
pSCCU	64B|64C	GA17643	1,766,685	GA24634	1,769,869	1,769,503..1769,545
pSTPP	65B|65C	GA13487	2,500,7218	GA13475	2,505,628	2,504,717..2,505,525
pHYSC	68B|68C	GA21475	6,814,856	GA19795	6,830,821	6,817,305..6,819,934
dSCCU	69C|69D	GA18750	8,021,189	GA24813	8,027,622	8,026,967..8,027,480
pSTAR	70A|76B	GA24334	8,902,967	GA19155	8,905,717	8,902,967..8,905,362
dSTPP	76B|76A	GA22082	9,162,365	GA20777	9,164,720	9,163,072..9,164,670
pHYST	76B|76A	GA22082	9,162,375	GA20777	9,164,720	9,163,072..9,164,057
pSCTL	74C|74B	GA19582	10,623,204	GA14679	10,628,103	10,626,931..10,627,448
pSCCH	70D|70C	GA20939	13,940,052	GA12507	13,941,493	13,940,125..13,941,042
dSTAR	70B|76C	GA17716	14,813,423	GA17559	14,833,672	14,831,322..14,833,071
dSCCH	78A|78B	GA10355	15,653,239	GA13882	15,658,703	15,657,098..15,657,203
dSCTL	79B	GA13539	17,149,432	GA24132	17,153,461	17,151,554..17152,967
dHYSC	79B|79C	GA21720	17,545,780	GA24111	17,567,890	17,545,438..17,545,671
dHYST	79D|80A	GA30270	17,725,786	GA10842	17,729,964	17,727,100..17,727,669

Breakpoint, is the first letter indicates if the breakpoint is proximal (p) or distal (d), the second two letters indicates the ancestral gene arrangement and the last two letters indicates the derived gene arrangement. Cyto, is the location of the breakpoint in the AR reference genome. The location of the dSTPP breakpoints differ from the original location mapped in Dobzhansky and Sturtevant (1938). The original map location of dSTPP was 75C|76A. 5'_Gene, the FlyBase annotation symbol of the gene 5' to the breakpoint. 5'_Coor, the nucleotide coordinate of the last base of the 5'_Gene. 3'_Gene, the FlyBase annotation symbol of the gene 3' to the breakpoint. 3'_Coor, the nucleotide coordinate of the first base of the 3'_Gene. BP_Interval, the coordinates of the breakpoint region in the AR reference sequence defined by mate pair reads spanning the inversion breakpoint. All coordinates are those of the AR reference genome (MV 2-25).

Table 2. Site frequency spectrum analysis based on gene arrangement and population categorical variables.

Arr	Freq	Number (%)	Pop	Freq	Number(%)
AR	1	90,653 (23.56)	MSH	1	44,949 (9.80)
	2-14	43,707 (11.36)		2-6	4,221 (0.92)
	15 (Fixed)	3,104 (0.81)		7 (Fixed)	0 (0.00)
	Shared	247,283 (64.27)		Shared	409,607 (89.28)
	Total	384,747		Total	458,777
ST	1	36,586 (13.47)	JR	1	72,751 (13.61)
	2-7	24,037 (8.85)		2-13	26,021 (4.87)
	8 (Fixed)	5,078 (1.87)		14 (Fixed)	0 (0.00)
	Shared	205,925 (75.81)		Shared	435,818 (81.52)
	Total	271,626		Total	534,590
PP	1	55,027 (16.87)	KB	1	48,658 (12.87)
	2-9	63,241 (19.39)		2-7	5,572 (1.47)
	10 (Fixed)	24,248 (7.43)		8 (Fixed)	0 (0.00)
	Shared	183,618 (56.30)		Shared	323,797 (85.65)
	Total	326,134		Total	378,027
CH	1	48,840 (14.63)	DM	1	73,274 (16.04)
	2-8	42,354 (12.43)		2-11	36,166 (7.91)
	9 (Fixed)	4,613 (1.35)		12 (Fixed)	0 (0.00)
	Shared	243,856 (71.58)		Shared	347,499 (76.05)
	Total	340,663		Total	456,939
TL	1	64,346 (17.94)	SPE	1	81,973 (19.77)
	2-8	47,670 (13.29)		2-7	38,535 (9.29)
	9 (Fixed)	8,856 (2.47)		8 (Fixed)	56 (0.01)
	Shared	237,703 (66.29)		Shared	294,144 (70.93)
	Total	358,575		Total	414,708
CU	1	31,504 (13.67)			
	2	9,554 (4.15)			
	3 (Fixed)	7,995 (3.47)			
	Shared	181,351 (78.71)			
	Total	230,404			

Arr, gene arrangement, AR, Arrowhead, ST, Standard, PP, Pikes Peak, CH, Chiricahua, TL, Tree Line, and CU, Cuernavaca; Freq, Frequency of the derived mutation within the arrangement or population; Pop, population; Number, number of SNP sites in the different frequency classes.

Table 3. Unique and Shared polymorphisms in proximal, inverted, and distal regions in six gene arrangements.

Arrangement	Type	Proximal Obs (Exp)	Inverted Obs (Exp)	Distal Obs (Exp)
Arrowhead*	Shared	114,700 (104,926.2)	52,305 (67,343.1)	58,155 (52,890.6)
	Unique	53,908 (63,681.8)	55,910 (40,871.9)	26,836 (32,100.4)
Standard*	Shared	141,750 (143,519.3)	28,261 (28,761.2)	18,563 (16,293.5)
	Unique	51,594 (49,824.7)	10,485 (9,984.8)	3,387 (5,656.5)
Pikes Peak*	Shared	14,064 (11,154.7)	96,036 (111,515.7)	53,828 (41,257.5)
	Unique	6,696 (9,605.3)	111,505 (96,025.3)	22,956 (35,526.5)
Chiricahua*	Shared	141,777 (131,368.2)	49,533 (60,142.2)	31,078 (30,877.6)
	Unique	46,509 (56,917.8)	36,667 (26,057.8)	13,178 (13,378.4)
Tree Line*	Shared	104,174 (92,840.4)	63,446 (78,297.3)	46,432 (42,914.3)
	Unique	40,719 (52,052.6)	58,750 (43,898.7)	20,543 (24,060.7)
Cuernavaca*	Shared	6,360 (5,626.2)	105,495 (106,453.1)	51,849 (51,624.6)
	Unique	947 (1,680.8)	32,760 (31,801.9)	15,198 (15,422.4)

*, $P<0.0001$ in a Chi Square test of homogeneity.

Table 4. Relationship between length of three chromosomal regions and the percent of selected protein coding genes within the regions.

Region	N	Intercept (SE)	Slope (SE)	*F*	*P*	R^2
Proximal	5	-1.08 (1.11)	0.22 (0.11)	4.00	0.14	0.57
Inverted	5	0.77 (0.79)	1.55 (0.19)	66.56	0.003	0.96
Distal	5	3.48 (0.77)	0.53 (0.50)	1.11	0.36	0.27

Table 5. Mean estimates of the recombination parameter *Rho* in outlier and non-outlier genes within inverted regions.

Arr	n Outlier	Rho Outlier	n Non-Outlier	Rho Non-Outlier	*P*
AR	20	0.037	765	0.047	2.5×10^{-2}
ST	8	0.028	393	0.045	3.9×10^{-2}
PP	123	0.041	1547	0.053	1.0×10^{-4}
CH	21	0.025	673	0.044	2.0×10^{-5}
TL	38	0.028	915	0.047	$<1.0 \times 10^{-5}$

Arr, arrangement; P, probability of observing a greater difference between outlier and non-outlier genes.

Table 6. GO analysis for genes in arrangements with significantly high mean DAF and PSBL and at least one fixed amino acid difference.

Term	Function	Count	*P* Value	Fold Enrichment	Benjamini (*q*)
GO:0016021	Integral component of membrane	69	6.51E-06	1.47	9.77E-05
GO:0050909	Odorant binding	11	1.09E-05	6.04	8.75E-04
dpo04080	Neuroactive ligand-receptor interaction	6	3.15E-04	9.33	1.34E-03

Figure Legends

Figure 1. *Drosophila pseudoobscura* inversion cline and ages. The inversion abbreviations are AR, Arrowhead; CH, Chiricahua; CU, Cuernavaca; HY, Hypothetical; PP, Pikes Peak; SC, Santa Cruz; ST, Standard; and TL, Tree Line. (Left) Gene arrangement cline in *D. pseudoobscura* across the southwestern United States. The six regions delineated on the map are different climatic zones or niches inferred in Schaeffer (2008) with pie charts indicating the frequencies of the major arrangements found in each zone from 1940 inversion samples(Dobzhansky, 1944). (Right) Ages of the inversions estimated from 18 nucleotide markers sampled from the third chromosome (A. G. Wallace et al., 2011). Dmir is the outgroup *D. miranda*. The Hypothetical arrangement is the common ancestor of all arrangements within *D. pseudoobscura*. The CU arrangement is connected to the SC arrangement via a dotted line because CU is derived from SC, but was not sampled in the Wallace et al. (2011) study

Figure 2. Phylogenetic network of the *Drosophila pseudoobscura* third chromosome gene arrangements inferred from polytene chromosomes isolated from larval salivary glands (Dobzhansky, 1944; Dobzhansky & Sturtevant, 1938). The box on the left shows the phylogenetic network of the different gene arrangements. The hypothetical arrangement is the ancestral arrangement (A. G. Wallace et al., 2011) and the arrows indicate the derivation of each arrangement from its ancestor. The strains with the circled arrangements were sequenced in this study. The chromosomes depicted to the right represent the numbered sections on the cytological map of *D. pseudoobscura* (Schaeffer et al., 2008). Beginning with the Hypothetical arrangement, the colored brackets indicate the segment of the ancestral arrangement that inverted and the corresponding segment is indicated with the same colored bar in the derived arrangement. These colors also match the arrows in the network in the box on the right. The colored sections under the maps indicate syntenic regions that are conserved among all gene arrangements.

Figure 3. Phylogeny inferred with the neighbor-joining method (Saitou & Nei, 1987). A total of 1,028,037 SNPs were used in each of the regions to construct the gene arrangement phylogeny. A total of 500 bootstrap replicated were used to determine the confidence in the nodes of the trees where only nodes with > 90% support shown on the tree.

Figure 4. Top Row: Phylogenies inferred with the neighbor-joining method (Saitou & Nei, 1987). Phylogenies were inferred in each of 14 syntenic blocks defined by seven pairs of inversion breakpoints. Below each phylogeny is a colored block with a label that corresponds to the 14 syntenic blocks on the cytogenetic map shown at the bottom. Indels were removed from the analysis. The following numbers of SNPs were used in each of the regions to construct the phylogenies: 63A-64B, 36,372 SNPs; 64C-65B, 24,361 SNPs; 65C-68B, 239,432 SNPs; 68C-69C, 77,386 SNPs; 69D-70A, 57,145 SNPs; 76B, 16,556 SNPs; 76A-74C, 91893 SNPs; 74B-70D, 203,463 SNPs; 70C-70B, 59,084 SNPs; 76C-78A, 51,476 SNPs; 78B-79A, 75,395 SNPs; 79B, 18,542 SNPs; 79C-79D, 9,324 SNPs; and 80A-81D, 67,608 SNPs. A total of 500 bootstrap replicated were used to determine the confidence in the nodes of the trees where only nodes with > 90% support shown on the tree. Middle Row: The 14 matrices indicate whether a particular region is outside (open box) or within (filled box) the inverted region in a heterozygote for a particular pair of arrangements (shown on the horizontal and vertical axis). The fill colors match the colors for the syntenic blocks of the reference sequence. Bottom Row: A

diagram of the cytogenetic map of the Arrowhead reference genome with colored segments representing the 14 syntenic blocks.

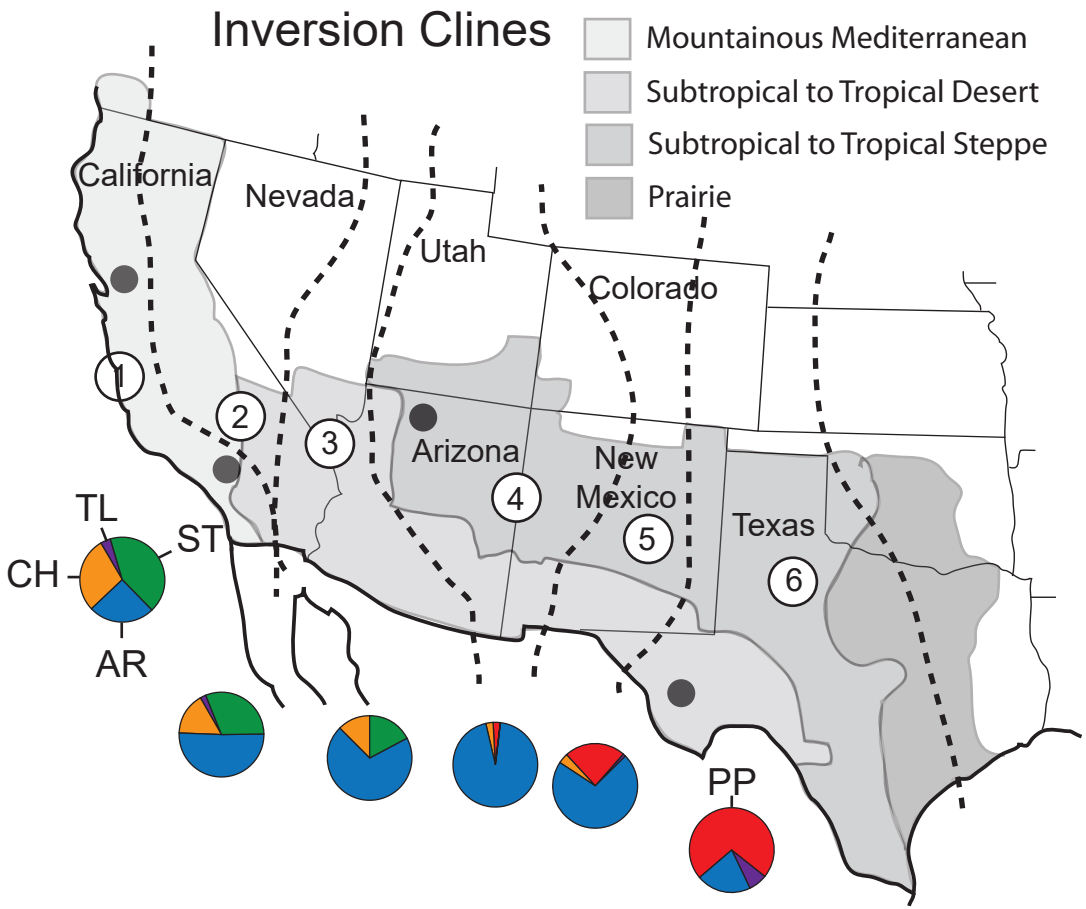
Figure 5. Estimates of Tajima's $D/|D_{min}|$ estimated from unique mutations in 2,669 gene regions across the third chromosome (Muller C) of *D. pseudoobscura* in five gene arrangements, Arrowhead, Standard, Pikes Peak, Chiricahua, and Tree Line. Outlier gene regions that showed a significant negative or positive value of Tajima's $D/|D_{min}|$ are indicated with a red point. Gene regions that did not have an extreme value of Tajima's $D/|D_{min}|$ are shown with a gray dot. The order of genes is specific to the particular gene arrangement background. The locations of inversion breakpoints are shown with a dotted line. Box plots for all $D/|D_{min}|$ values across the third chromosome are shown in the lower right hand corner for each of the arrangements (Hintze & Nelson, 1998).

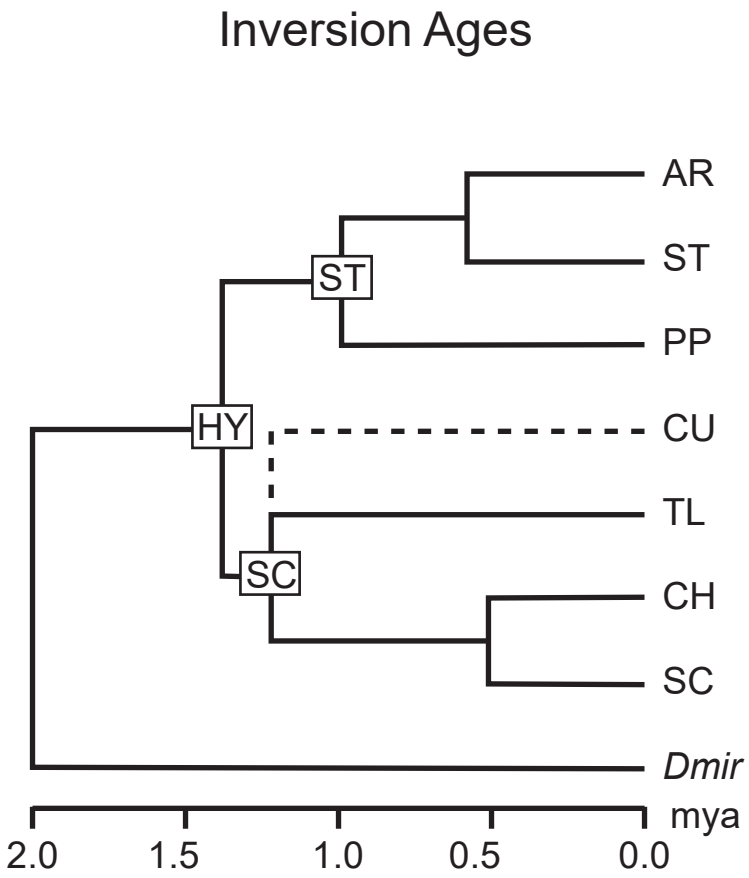
Figure 6. Estimates of the mean derived allele frequency (DAF) per segregating site estimated from unique mutations in the five gene arrangements, Arrowhead, Standard, Pikes Peak, Chiricahua, and Tree Line. Regions that showed a significantly large mean given the number of SNPs are indicated with a red point. Gene regions that did not have an extreme mean DAF value are shown with a gray dot. The locations of inversion breakpoints are shown with a solid or dotted line. The order of genes is specific to the particular gene arrangement background. Box plots for all mean DAF values across the third chromosome are shown in the lower right hand corner for each of the arrangements (Hintze & Nelson, 1998).

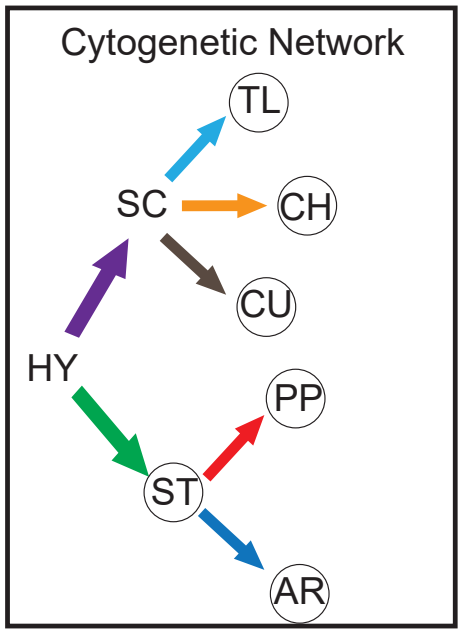
Figure 7. Population specific branch lengths estimated for five of the six arrangements. The plots to the left show the estimates of PSBL for the 2,669 genes across the third chromosome. Each gene is represented by a point, with those colored in red classified as statistical outliers after correcting for multiple testing. The histograms to the right show the distribution of PSBL for each arrangement with the branch highlighted in red.

Figure 8. Estimates of the recombination parameter ρ within 2,669 gene regions across the third chromosome. Regions that showed a significantly large mean DAF or PSBL given the number of SNPs are indicated with a red circle. Regions that did not have an extreme mean DAF value are shown with a gray dot. The locations of inversion breakpoints are shown with a solid or dotted line. The order of genes is specific to the particular gene arrangement background.

Figure 9. Triangular heat map showing the significance of linkage disequilibrium (LD) for all polymorphic sites on the third chromosome using the correlation-based procedure of Zaykin et al. (2008). Each point is calculated as the average of 100 adjacent sites. Red indicates greater LD and blue represent associations that are not significant. The top bars depict the results of a Fisher's Exact Test for significant associations of alleles with a particular arrangement. The colors represent the proportions of significant tests ($p < 1 \times 10^{-6}$) in blocks of 100 sites.


Inversion Clines
Mountainous Mediterranean
Subtropical to Tropical Desert
Subtropical to Tropical Steppe
Prairie
California
Nevada
Utah
Colorado
Arizona
New Mexico
Texas
1
2
3
4
5
6
TL
ST
CH
AR
PP


Inversion Ages
AR
ST
PP
CU
TL
CH
SC
Dmir
HY
ST
SC
mya
2.0
1.5
1.0
0.5
0.0



Tree Line (TL)

63 ABCDE | 64 ABC | 65 ABCD | 66 ABCD | 67 ABCDE | AB | 79 CD | 76 A | 75 CBA | 74 C | 79 B | 68 CD | 69 ABCDE | 70 ABCD | 71 ABC | 72 ABC | 73 ABC | 74 AB | 79 A | 78 CBA | 77 BA | 76 CB | 80 ABC | 81 ABCD

Cuernavaca (CU)

63 ABCDE | 64 AB | 69 DE | 70 ABCD | 71 CBA | 72 ABC | 73 ABC | 74 ABC | 75 ABC | 76 A | 79 DC | 68 BA | 67 EDCBA | 66 DCBA | 65 DCBA | 64 C | 69 CBA | 68 DC | 79 BA | 78 CBA | 77 BA | 76 CB | 80 ABC | 81 ABCD

Chiricahua (CH)

63 ABCDE | 64 ABC | 65 ABCD | 66 ABCD | 67 ABCDE | 68 AB | 79 CD | 76 A | 75 CBA | 74 CBA | 73 CBA | 72 CBA | 71 CBA | 70 D | 78 BC | 79 AB | 68 CD | 69 ABCDE | 70 ABC | 78 A | 77 BA | 76 CB | 80 ABC | 81 ABCD

SC → CH

SC → TL

SC → CU

Santa Cruz (SC)

63 ABCDE | 64 ABC | 65 ABCD | 66 ABCD | 67 ABCDE | 68 AB | 79 CD | 76 A | 75 CBA | 74 CBA | 73 CBA | 72 CBA | 71 CBA | 70 DCBA | 69 EDCBA | 68 DC | 79 BA | 78 CBA | 77 BA | 76 CB | 80 ABC | 81 ABCD

HY → SC

Hypothetical (HY)

63 ABCDE | 64 ABC | 65 ABCD | 66 ABCD | 67 ABCDE | 68 ABCD | 69 ABCDE | 70 ABCD | 71 ABC | 72 ABC | 73 ABC | 74 ABC | 75 ABC | A | 79 DCBA | 78 CBA | 77 BA | 76 CB | 80 ABC | 81 ABCD

HY → ST

Standard (ST)

63 ABCDE | 64 ABC | 65 ABCD | 66 ABCD | 67 ABCDE | 68 ABCD | 69 ABCDE | 70 ABCD | 71 ABC | 72 ABC | 73 ABC | 74 ABC | 75 ABC | 76 ABC | 77 AB | 78 ABC | 79 ABCD | 80 ABC | 81 ABCD

ST → AR

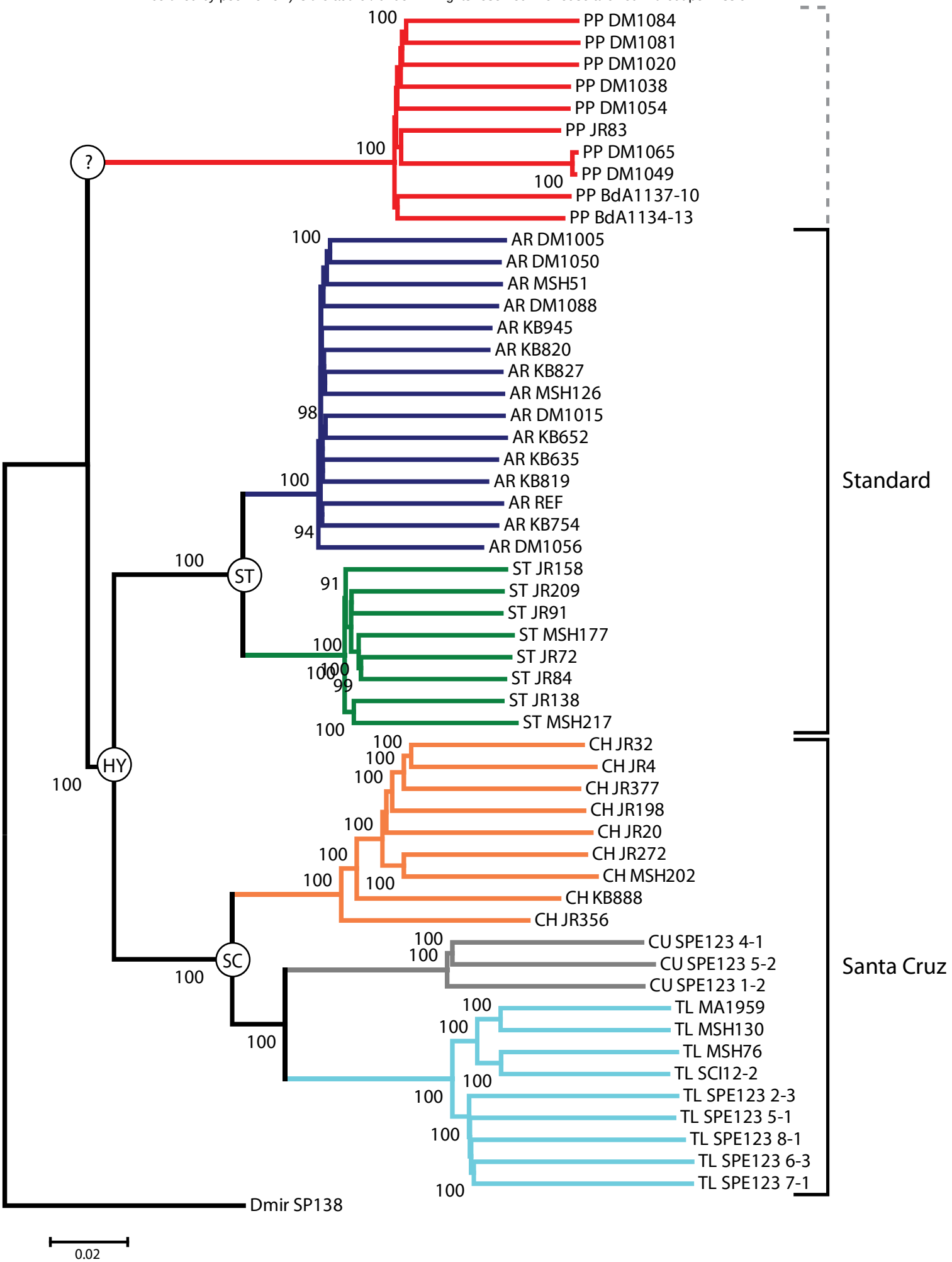
ST → PP

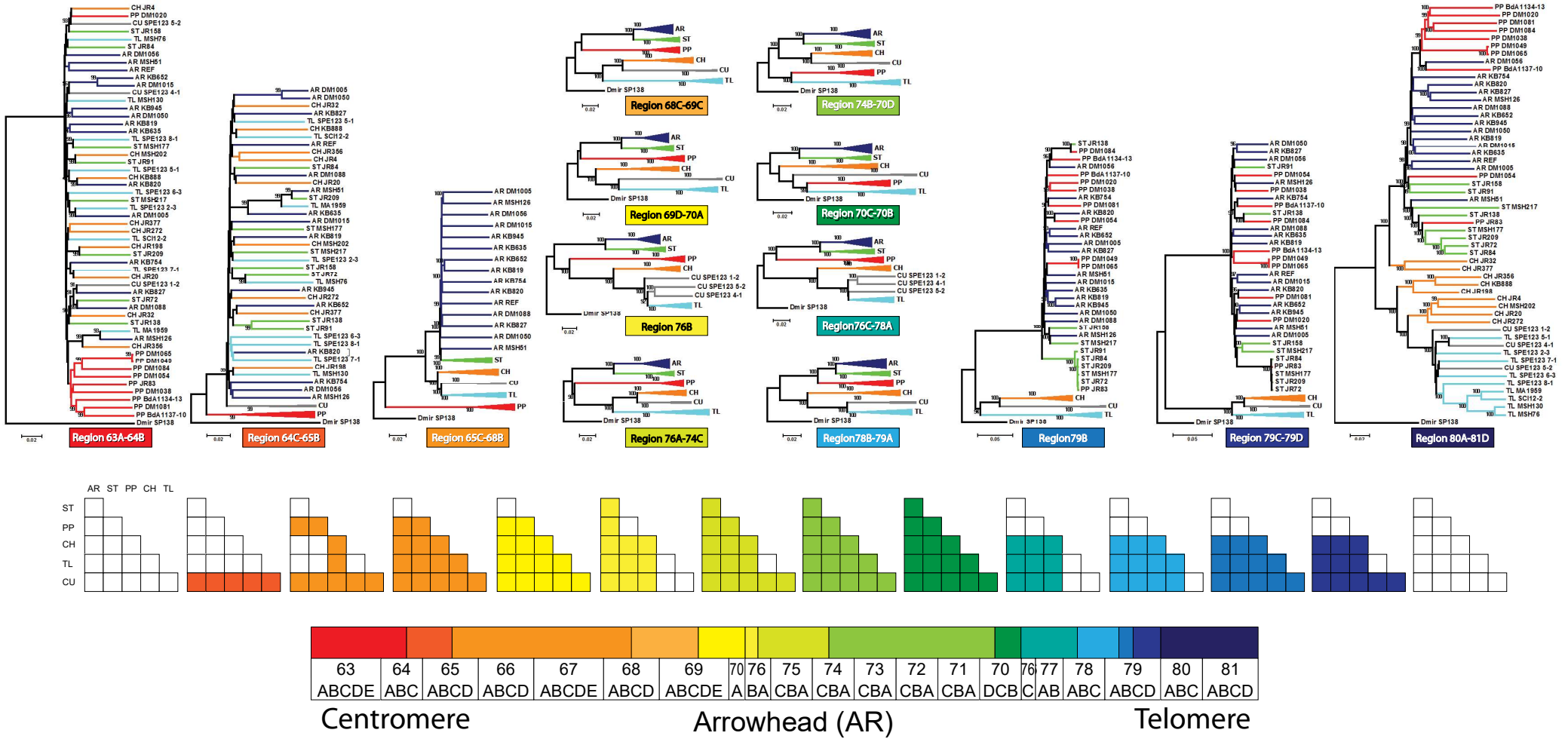
Pikes Peak (PP)

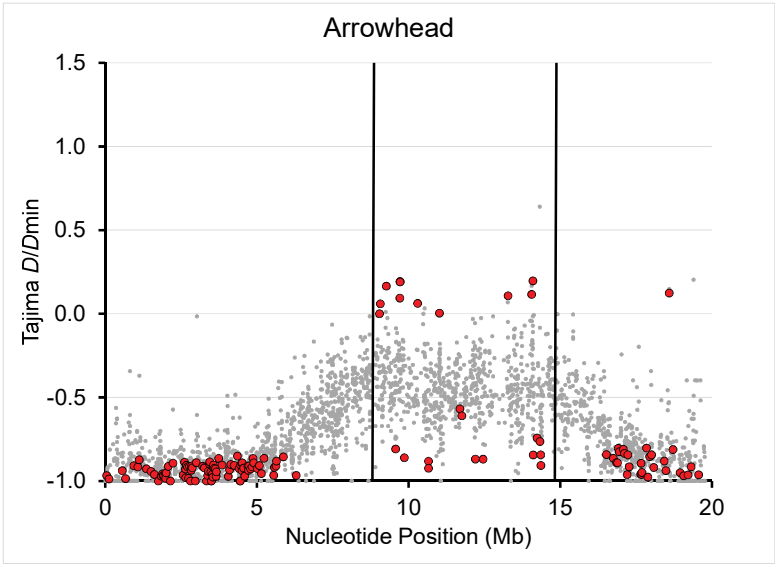
63 ABCDE | 64 ABC | 65 AB | 76 A | 75 CBA | 74 CBA | 73 CBA | 72 CBA | 71 CBA | 70 DCBA | 69 EDCBA | 68 DCBA | 67 EDCBA | 66 DCBA | 65 DC | 76 BC | 77 AB | 78 ABC | 79 ABCD | 80 ABC | 81 ABCD

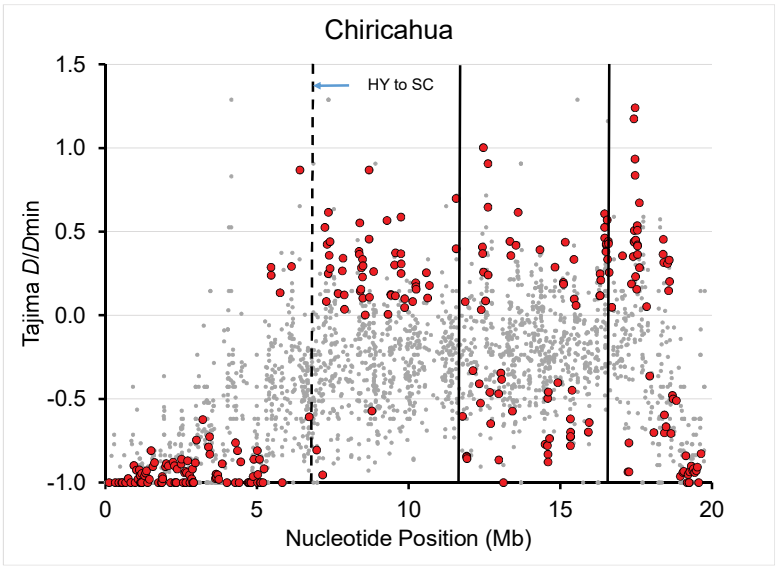
Arrowhead (AR)

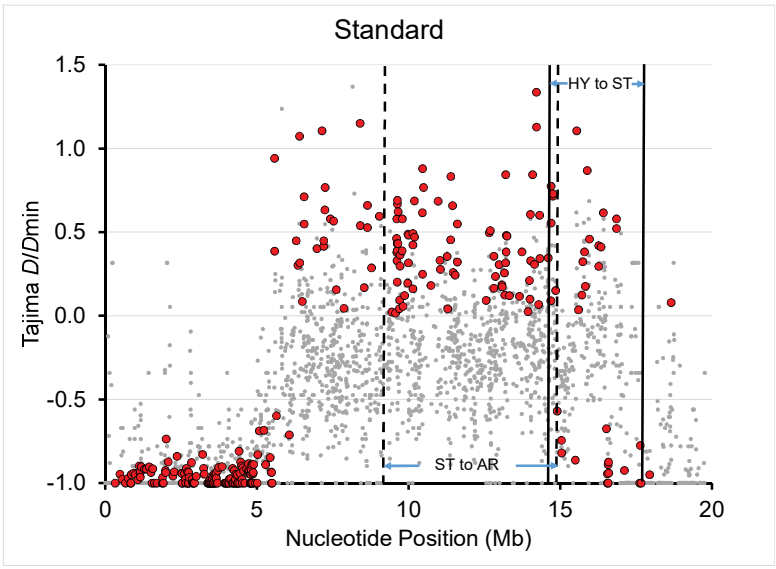
63 ABCDE | 64 ABC | 65 ABCD | 66 ABCD | 67 ABCDE | 68 ABCD | 69 ABCDE | 70 A | 76 BA | 75 CBA | 74 CBA | 73 CBA | 72 CBA | 71 CBA | 70 DCB | 76 C | 77 AB | 78 ABC | 79 ABCD | 80 ABC | 81 ABCD

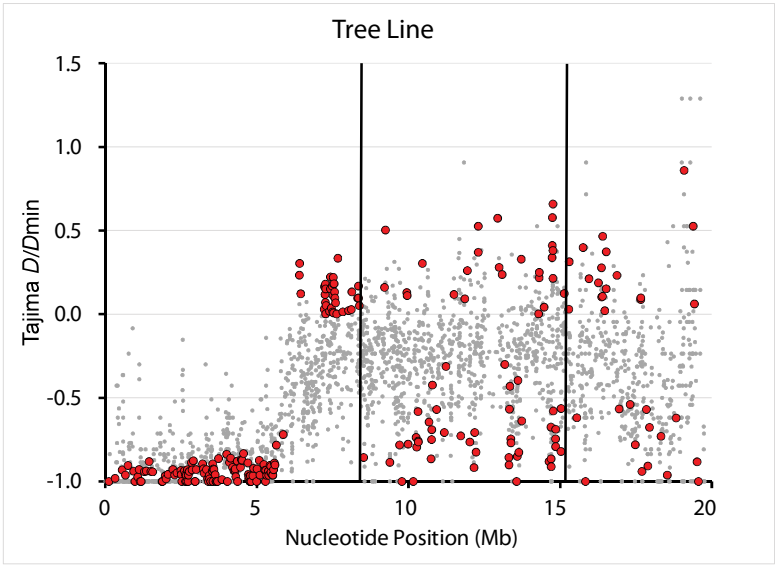


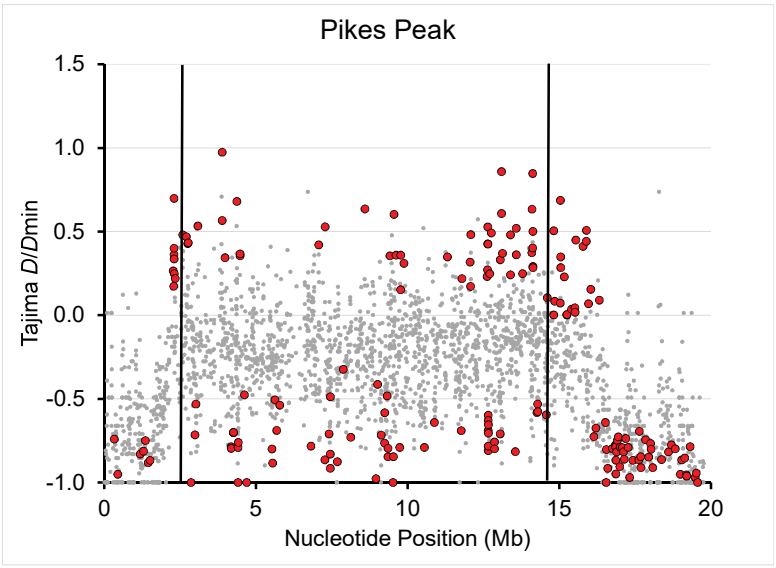

Region 63A-64B
Region 64C-65B
Region 65C-68B
Region 68C-69C
Region 69D-70A
Region 76B
Region 76A-74C
Region 74B-70D
Region 70C-70B
Region76C-78A
Region78B-79A
Region79B
Region 79C-79D
Region 80A-81D
AR ST PP CH TL
ST
PP
CH
TL
CU
63 ABCDE
64 ABC
65 ABCD
66 ABCD
67 ABCDE
68 ABCD
69 ABCDE
70 A
76 BA
75 CBA
74 CBA
73 CBA
72 CBA
71 CBA
70 DCB
76 C
77 AB
78 ABC
79 ABCD
80 ABC
81 ABCD
Centromere
Arrowhead (AR)
Telomere

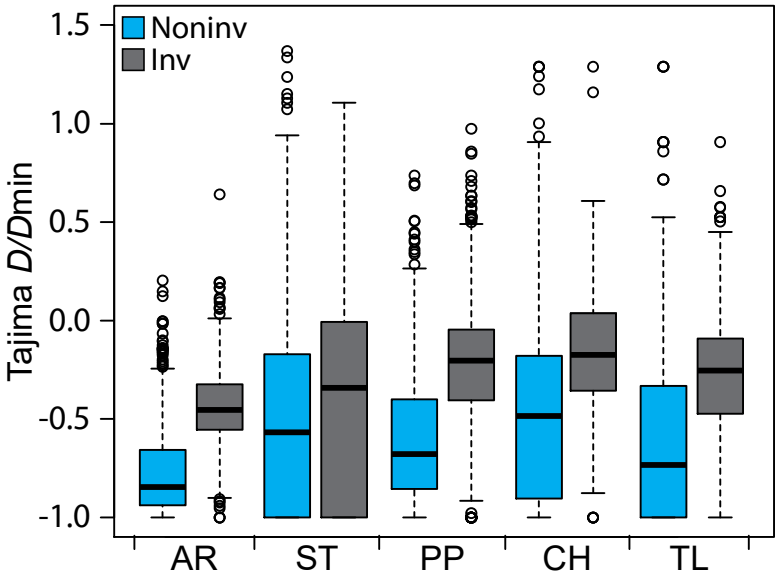


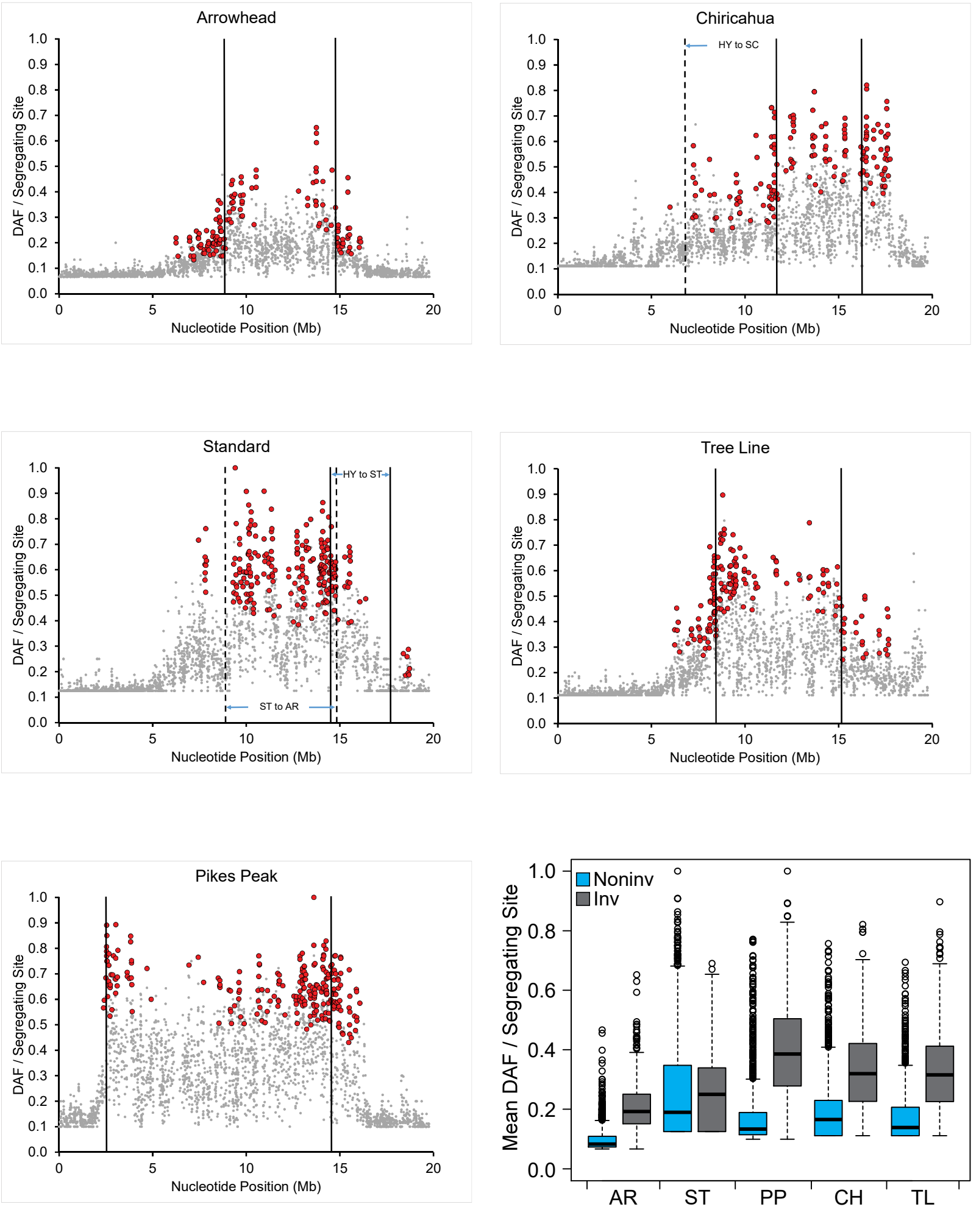


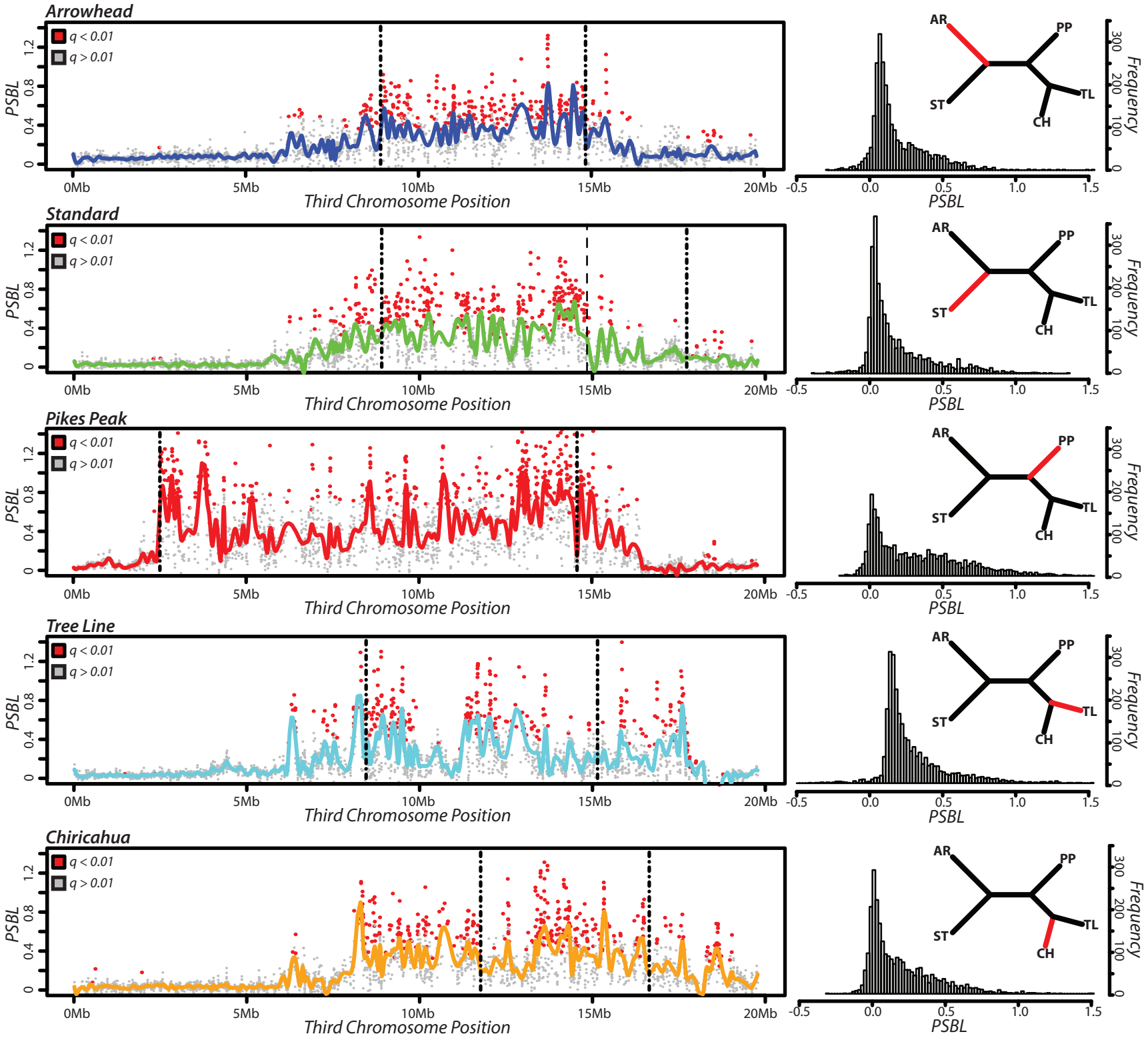











Arrowhead
Standard
Pikes Peak
Tree Line
Chiricahua
q < 0.01
q > 0.01
PSBL
Third Chromosome Position
0Mb
5Mb
10Mb
15Mb
20Mb
AR
PP
ST
TL
CH
Frequency