Molecular Population Genetics of X-Linked Genes in *Drosophila pseudoobscura*

Miro Kovacevic and Stephen W. Schaeffer

*Department of Biology and Institute of Molecular Evolutionary Genetics, The Pennsylvania State University, University Park, Pennsylvania 16802-5301*

Manuscript received January 28, 2000
Accepted for publication May 1, 2000

**ABSTRACT**

This article presents a nucleotide sequence analysis of 500 bp determined in each of five X-linked genes, *runt*, *sisterlessA*, *period*, esterase 5, and Heat-shock protein 83, in 40 *Drosophila pseudoobscura* strains collected from two populations. Estimates of the neutral migration parameter for the five loci show that gene flow among *D. pseudoobscura* populations is sufficient to homogenize inversion frequencies across the range of the species. Nucleotide diversity at each locus fails to reject a neutral model of molecular evolution. The sample of 40 chromosomes included six *Sex-ratio* inversions, a series of three nonoverlapping inversions that are associated with a strong meiotic drive phenotype. The selection driven by the *Sex-ratio* meiotic drive element has not fixed variation across the X-chromosome of *D. pseudoobscura* because, while significant linkage disequilibrium was observed within the *sisterlessA*, *period*, and esterase 5 genes, we did not find evidence for nonrandom association among loci. The *Sex-ratio* chromosome was estimated to be 25,000 years old based on the decomposition of linkage disequilibrium between esterase 5 and Heat-shock protein 83 or 1 million years old based on the net divergence of esterase 5 between Standard and *Sex-ratio* chromosomes. Genetic diversity was depressed within esterase 5 within *Sex-ratio* chromosomes, while the four other genes failed to show a reduction in heterozygosity in the *Sex-ratio* background. The reduced heterogeneity in esterase 5 is due either to its location near one of the *Sex-ratio* inversion breakpoints or that it is closely linked to a gene or genes responsible for the *Sex-ratio* meiotic drive system.

THE amount of gene flow that occurs in *Drosophila pseudoobscura* is critical to understanding the evolutionary forces that generate clines in inversion frequencies across the species distribution. Two chromosomes within *D. pseudoobscura*, the X and the third chromosome, show clinal variation for gene arrangements in natural populations. The X-chromosome is polymorphic for two gene arrangements and the frequencies of the two inversion chromosomes form a north/south cline in the North American range of *D. pseudoobscura* (STURTEVANT and DOBZHANSKY 1936). The third chromosome is polymorphic for >30 different gene arrangements that relate via an unrooted network of single inversion events (DOBZHANSKY 1944). The frequencies of the third chromosomal inversions collected from populations across the Great Basin region form classical clines that have been stable for the last 60 years (DOBZHANSKY and EPLING 1944; ANDERSON et al. 1991).

Gene frequency clines can occur in species with low or high levels of gene flow; however, the underlying evolutionary forces that generate the gradient in allele frequencies are quite different. When gene flow is low, clines are likely to be generated from mutation and genetic drift where a new gene arrangement emerges in one location and slowly disperses to new populations over time to generate the observed gradients (WRIGHT 1931). When gene flow is extensive, clines are likely to be caused by selection in local environments where gene flow acts to homogenize gene frequencies among populations, but natural selection alters the gene arrangement frequencies (SLATKIN 1987; SIMMONS et al. 1989; BERRY and KREITMAN 1993). The observation of parallel clines on different continents has been used as strong evidence that natural selection generates the gradients in gene frequency (DAVID 1982; PREVOSTI et al. 1988); however, no evidence for parallel clines exists for *D. pseudoobscura* because this species is endemic only to North America.

Direct measurements of dispersal in *D. pseudoobscura* have suggested that gene flow is limited, suggesting that genetic drift and mutation have generated the inversion clines (CRUMPACKER and WILLIAMS 1973; DOBZHANSKY and POWELL 1974; POWELL et al. 1976). On the other hand, indirect estimates of gene flow based on protein diversity have indicated that gene flow is extensive, suggesting that selection in local environments alters inversion frequencies among populations (PRAKASH et al. 1969; SINGH et al. 1976; COYNE and Felton 1977; COYNE et al. 1978; JONES et al. 1981; KEITH 1983; KEITH et al. 1985; SCHAEFFER and MILLER 1992a). Protein data, however, have limited utility for estimating gene flow because proteins can be the targets of selection and allozymes with similar electrophoretic mobility may not be identical by descent. With this in mind, SCHAEFFER and...
Miller (1992a) used nucleotide sequence analysis at the alcohol dehydrogenase (Adh) locus to infer levels of gene flow in *D. pseudoobscura* and found that gene flow was sufficient to homogenize gene frequencies throughout the range of the species. The Adh locus is a good marker to infer levels of gene flow because there is no evidence that adaptive selection has acted in the recent history of the region (Schaeffer and Miller 1992b); however, this locus provides only a single estimate of gene flow for the species. In this article, we extend the work of Schaeffer and Miller (1992a) by estimating gene flow from five loci that are distributed across both arms of the X chromosome of *D. pseudoobscura*.

The X chromosome of *D. pseudoobscura* is metacentric and is composed of a left arm (XL) that is syntenic with the *D. melanogaster* X chromosome and a right arm (XR) that is homologous with the left arm of the *D. melanogaster* third chromosome. XR is segregating for two gene arrangements, Standard and Sex-ratio, where the two chromosomal arms differ by a set of three nonoverlapping inversions (Dobzhansky 1939). The Sex-ratio (SR) inversion is associated with a strong meiotic drive phenotype where males that carry the SR chromosome transmit their X chromosome preferentially so that they sire 95–100% female offspring (Novitski et al. 1965; Anderson et al. 1967; Policansky and Ellison 1970). The exact number of genes responsible for the Sex-ratio phenotype is unknown in *D. pseudoobscura*, but at least four gene regions control Sex-ratio in the sibling species *D. persimilis* (Wu and Beckenbach 1983). The two gene arrangements on XR vary in frequency among *D. pseudoobscura* populations, where SR is not detected in individuals collected from Oregon, Washington, and British Columbia, but is found at frequencies as high as 25% in populations collected from the southwestern United States and Mexico (Sturtevant and Dobzhansky 1936; Dobzhansky 1944). Our study estimates gene flow between two populations in the southwestern United States to avoid the known gene frequency differences of Sex-ratio in the north/south cline.

Nucleotide sequence analysis of the esterase-5 locus in *D. pseudoobscura* shows that the Sex-ratio chromosome is monophyletic and is derived from the Standard gene arrangement (Babcock and Anderson 1996). The rapid increase of the initial Sex-ratio chromosome in response to meiotic drive was likely to create large levels of linkage disequilibrium across the X chromosome as has been observed in other inversion systems (Aquadro et al. 1991; Bénassi et al. 1993; Babcock and Anderson 1996; Hasson and Eanes 1996; Andolfatto et al. 1999; Rozas et al. 1999; Andolfatto and Kreitman 2000). Gene arrangements are likely to accumulate genetic differences because of the reduction in single crossover events within inversion heterozygotes (Sturtevant and Beadle 1936). Some genetic exchanges have been observed between inversions either through double crossover or gene conversion events (Rozas and Aguadé 1990, 1993, 1994; Popadic et al. 1995). Extensive genetic differentiation between the Standard and SR chromosomes has been documented with allozyme electrophoresis (Prakash 1972, 1974; Keith 1983), but as was mentioned above, proteins may not give an accurate picture of the past evolutionary history of this meiotic drive system.

We present here a nucleotide sequence analysis of five loci on the X chromosome of *D. pseudoobscura*. The Heat-shock protein 83 (*Hsp83*), esterase-5 (*Est-5*), period (*per*), sisterlessA (*sisa*), and runt (*run*) genes were sequenced from 40 strains of *D. pseudoobscura* collected from Arizona and Texas. Three genes map to the XL of *D. pseudoobscura* (*run* (Pepling and Gergen 1995), *sisa* (Erickson and Cline 1998), and *per* (Colot et al. 1988). Two loci map to XR in *D. pseudoobscura* and are located within or distal to the subbursal inversion of the Sex-ratio gene arrangement: Est-5 (Brady et al. 1990; Babcock and Anderson 1996) and *Hsp83* (Blackman and Meselson 1986). Nucleotide data at these five loci were used to address three major questions: (1) How much gene flow occurs in *D. pseudoobscura*? (2) How much linkage disequilibrium occurs across an X chromosome subject to meiotic drive? And (3) how much linkage disequilibrium has decayed across the X chromosome since the origin of the Sex-ratio chromosome?

**MATERIALS AND METHODS**

**Fly strains and DNA extraction:** Twenty isofemale lines of *D. pseudoobscura* were collected from the Kaibab National Forest of Arizona (KB; 36° 42′ N latitude, 112° 12′ W longitude) and 20 isofemale lines were collected from the Davis Mountains of Texas (DM; 30° 35′ N latitude, 103° 53′ W longitude) during the late summer of 1998, respectively. The two populations are separated by ~600 miles. Single F₁ males were isolated from each of 20 KB strains numbered PSU 596–633 and 20 DM strains numbered PSU 994–1029. The *D. miranda* strain, SP 295, was kindly provided by Soojin Yi (University of Chicago). Genomic DNA was extracted from single F₁ males with the procedure of Gloor and Engels (1992). These DNAs were used to amplify and sequence five X-linked genes with PCR-mediated nucleotide sequencing (Saiki et al. 1988).

**DNA sequencing and analysis:** A noncoding region of ~500 bp in length was chosen for four of the five genes to maximize the number of segregating sites obtained for our study. The *run* locus marker included 106 bases of coding sequence and 300 nucleotides of intron sequence in the GenBank reference sequence. Figure 1 shows the fine structure of the five gene regions examined for this study and the DNA fragment that was amplified and sequenced. Table 1 shows information about the PCR primers used to amplify the five genes of the X chromosome, including the GenBank accession number of the sequence used to design oligonucleotides for amplification reactions. Our standard PCR reaction used 2 μl DNA, 5 μl 10× PCR buffer (100 mm Tris-HCl pH 8.3, 500 mm KCl, 15 mm MgCl₂, 1% Triton; PGC Scientific), 8 μl 1.25 mm dNTPs (Pharmacia, Piscataway, NJ), 1 μl primer 1 (100 ng/μl), 1 μl primer 2 (100 ng/μl), 32.75 μl sterile deionized H₂O, and 0.25 μl display Taq FL DNA polymerase (PGC Scientific). The standard PCR profile for the five loci was as follows: denaturation for 5 min at 94° for 1 cycle followed by 30 cycles of denaturation for 1 minute at 94°, anneal for 2 min at the appropriate
of PCR primers, f indicates the forward primer, r indicates the reverse primer, and mf indicates the forward primer used for D.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>run</td>
<td>2450</td>
<td>f 5’ CTG ACC ATC ATC ACG ATT GCC AC 3’</td>
<td>65°</td>
<td>U22357</td>
<td>AF196134–AF196174</td>
</tr>
<tr>
<td></td>
<td>2937</td>
<td>r 5’ AAG TAG TCC GCG TAG GCG TA 3’</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>sixA</td>
<td>1794</td>
<td>f 5’ TGT TCG CCA TTC TCC ATA CG 3’</td>
<td>50°</td>
<td>AF045586</td>
<td>AF196093–AF196133</td>
</tr>
<tr>
<td></td>
<td>2322</td>
<td>r 5’ TGT CGG CTG GTC GGA GAT AA 3’</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>per</td>
<td>1719</td>
<td>mf 5’ GTC GGC CAT CAT CGT GTG TT 3’</td>
<td>55°</td>
<td>X13878</td>
<td>AF196052–AF196092</td>
</tr>
<tr>
<td></td>
<td>1776</td>
<td>f 5’ CCT AGG AAG TTC CAA GCA GT 3’</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2330</td>
<td>r 5’ GCG ATC TTC TCC GGT GAG TT 3’</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Est-5</td>
<td>2038</td>
<td>f 5’ CGA TAA GTG GAG GCT CCT CTC ATG 3’</td>
<td>65°</td>
<td>M55907</td>
<td>AF196011–AF196051</td>
</tr>
<tr>
<td></td>
<td>2625</td>
<td>r 5’ AAC CAG TCT CAG CAG GAT AGC CCT TCT 3’</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hsp83</td>
<td>1401</td>
<td>f 5’ CAT GCA CAC AGG ACC ACA TA 3’</td>
<td>60°</td>
<td>X03812</td>
<td>AF195970–AF196010</td>
</tr>
<tr>
<td></td>
<td>1936</td>
<td>r 5’ AAT GGC AGA CTT CTC ATG CT 3’</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Size, length in nucleotides of the PCR amplification product; Nucleotide, 5’ position of the forward or reverse PCR primer; PCR primers, f indicates the forward primer, r indicates the reverse primer, and mf indicates the forward primer used for D. miranda; Temp., annealing temperature for the PCR cycle; GenBank no. primer, GenBank accession number for the sequence that was used to design the PCR primers for the gene; GenBank no. this study, GenBank accession numbers for the 40 strains of D. pseudoobscura and 1 strain of D. miranda.

Physical and genetic mapping: The five genes examined in this study were placed on the genetic map by scoring nucleotide polymorphisms in F1 males derived from females that were heterozygous at two or more marker loci. We used the polymorphism data from each locus to pick restriction site or sequence length variation that could be used to genetically map the five loci. Single females were placed in a vial and allowed to lay eggs. After 1 wk, females were killed and genomic DNA was purified from each female with a single fly DNA isolation procedure (Gloor and Engels 1992). PCR was used to amplify the five marker genes from each female and an appropriate restriction fragment length polymorphism (RFLP) analysis was used to score each locus as homo- or heterozygous (Table 2). F1 males were collected from mothers who were heterozygous for two or more loci and genetic map distances were estimated from the offspring haplotypes. A sixth X-linked marker, X006, was used to confirm the position of the five genes in this study with the X map of Noor et al. (2000). The presence of inversions in the genetic background can increase the genetic distance estimates on unlinked chromosomes (Lucchesi 1976). These mapping crosses were not likely to be affected by interchromosomal effect because all crosses were done with strains collected from the Kaibab population where the Arrowhead chromosome had an estimated frequency of 94%. The Konami (1944) mapping function was used to correct the percentage of crossing-over values for multiple recombinational events that can occur between distant markers.

RESULTS

Genetic map location of the five X-linked loci: Figure 2 shows the locations of the run, sixA, per, Est-5, and Hsp83 loci placed on the genetic map of the D. pseudoobscura X chromosome. Our map is based on the X chro-
mosomal map of the *D. pseudoobscura* revised by Orr (1995). Orr’s revision incorporated the *scalloped* locus onto the map and corrected the original map positions (Lancefield 1922; Donald 1936; Anderson 1993) with the Kosambi (1944) mapping function that adjusts distances of loosely linked loci for multiple crossover events. We have included 7 of the 10 microsatellite loci mapped by Noor et al. (2000) on the genetic map.

We used three-point mapping data to determine the order of the five genes on the X chromosome of *D. pseudoobscura*. In addition, the five genes were mapped relative to the *X006* microsatellite locus that maps near the centromere of the X (Mohamed Noor, personal communication; Noor et al. 2000). The central gene of each trio of genes was inferred from the frequencies of the parental and double recombinant genotypes. Genotyped male offspring from females heterozygous at the following combinations of loci were used to infer the gene order: (1) *run*, *sisA*, *per*; (2) *sisA*, *per*, *X006*; (3) *sisA*, *per*, *Hsp83*; (4) *per*, *X006*, *Hsp83*; and (5) *X006*, *Est-5*, *Hsp83*. We combined data from all mapping crosses of adjacent loci to further refine the genetic distances among the six loci.

The gene order for the five loci proceeding from XL to XR is *run*, *sisA*, *per*, *Est-5*, and *Hsp83*. We mapped the positions of *sisA*, *per*, *Est-5*, and *Hsp83* relative to the *run* locus, which is located at position 50.8 on the *D. pseudoobscura* X chromosome (Noor et al. 2000). We estimated the following genetic distances among the six genes using the Kosambi mapping function: *run-sisA* = 30.8 ± 3.8 cM (*n* = 146), *sisA-per* = 13.7 ± 2.2 cM (*n* = 240), *per-X006* = 11.5 ± 2.3 cM (*n* = 186), *X006-Est-5* = 9.7 ± 2.4 cM (*n* = 146), and *Est-5-Hsp83* = 2.5 ± 1.0 cM (*n* = 243).

The X chromosomal map that we developed agrees with previous physical and genetic mapping data from *D. pseudoobscura*. The locations of *run*, *sisA*, and *per* on XL and *Est-5* and *Hsp83* on XR agree with the syntenic relationships between the *D. pseudoobscura* X and the *D. melanogaster* X and XL (Sturtevant and Tan 1937; Sturtevant and Novitski 1941). The positions of *Est-5* and *Hsp83* on the genetic map also correspond with the physical locations of the two loci observed with *in situ* hybridization of DNA probes to polytene chromosomes of the salivary glands (Babcock and Anderson 1996; Segarra et al. 1996).

We have made some adjustments to the genetic map of the *D. pseudoobscura* X chromosome on the basis of the addition of the five new loci. The intergene distances for *sisA* to *per*, *per* to *X006*, *X006* to *Est-5*, and *Est-5* to *Hsp83* are all <14 cM. Thus, the likelihood that multiple crossovers would lead to underestimates of genetic distance for these gene pairs is low (Ashburner 1989). These distances help to refine the genetic map in the central region of the X chromosome. On the new map, *yellow* and *Est-5* are separated by 42.0 cM, which is similar to the 43.4 cM estimated by Beckenbach (1981). Orr’s revised map suggests that *yellow* and *Est-5* are separated by 53.4 cM, assuming that *Est-5* is 6.1 cM to the right.

### TABLE 2

<table>
<thead>
<tr>
<th>Gene</th>
<th>Polymorphism</th>
<th>Fragment pattern 1 (bp)</th>
<th>Fragment pattern 2 (bp)</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>run</em></td>
<td>Indel</td>
<td>250</td>
<td>193</td>
</tr>
<tr>
<td><em>sisA</em></td>
<td>RFLP-Pdf</td>
<td>529</td>
<td>311</td>
</tr>
<tr>
<td><em>per</em></td>
<td>Indel</td>
<td>760</td>
<td>181</td>
</tr>
<tr>
<td><em>X006</em></td>
<td>Indel</td>
<td>190</td>
<td>182</td>
</tr>
<tr>
<td><em>Est-5</em></td>
<td>RFLP-HindI</td>
<td>499, 96</td>
<td>193, 147, 133, 64</td>
</tr>
<tr>
<td><em>Hsp83</em></td>
<td>RFLP-Xbal</td>
<td>256, 147, 133</td>
<td>193, 147, 133, 64</td>
</tr>
</tbody>
</table>

Indel, insertion or deletion polymorphism; RFLP, restriction fragment length polymorphism with the enzyme used to generate the pattern.

![Figure 2](image-url)---Genetic map of the *D. pseudoobscura* X chromosome. The seven microsatellite loci *X001, X002, X003, X004, X006, X008*, and *X009* were previously mapped by Noor et al. (2000). The abbreviations of the genes are as follows: *Pt*, Pointed; *ct*, cut; *sd*, scalloped; *y*, yellow; *w*, white; *v*, vermilion; *sn*, singed; *mg*, magenta; *co*, compressed; *se*, sepia; *ll*, lanceolate; *sh*, short.
of compressed. We suggest that the positions of singed, vermilion, and compressed should be adjusted from positions 81.7, 84.3, and 121.8 to positions 86.3, 88.3, and 110.3, respectively, to reflect the reduction in size of the central region of the X chromosome. The positions of magenta, sepa, lanceolate, and short were also adjusted to reflect this change in the map. These adjustments to the map increase the length of the X chromosome by 12.6 cM based on the ANDERSON (1993) map and reduce the length of the chromosome by 11.4 cM on the ORR (1995) map.

Estimates of nucleotide diversity: Figure 3 presents a graphical view of segregating sites in the five genes sequenced from the X chromosome. The figure is designed to give a qualitative view of the diversity in the five genes and the arrangement of haplotypes among the five loci distributed across the chromosome. For more detailed views of the data, the aligned sequences that were used in all analyses can be obtained from the POPSET database at GenBank. Several observations can be made from the polymorphism data. First, there is no obvious pattern of association among segregating sites or haplotypes among the five genes with the exception of chromosomes designated Sex-ratio. Second, the five loci differ in the number of segregating sites found within each region. Third, two strains, PSU 633 and PSU 996, appear to have quite different sequences from the consensus in the 5′ end of the per gene. Initially, we thought that this might be due to an alignment problem, but on closer examination, the 5′ end of these two alleles is quite divergent from the consensus, but is also quite similar to the 5′ region of the outgroup species D. miranda. It appears that the 5′ segment of these sequences is a vestige of a recombination event in an ancestor of D. miranda and D. pseudoobscura. This phenomenon has been observed in the Adh locus, where a small segment of the adult intron was shared between D. pseudoobscura and D. miranda (SCHAEFFER and MILLER 1993), and in the period locus downstream of exon VII, where an allele of D. pseudoobscura and an allele of D. persimilis share considerable sequence similarity (WANG and HEY 1996).

The estimates of nucleotide diversity for the five X-linked genes within the two populations are reported in Table 3. In all cases, the variation estimates are given on a variation per silent site basis, i.e., synonymous sites within the coding region, intron sites, and flanking regions. In addition, these estimates were corrected for being on the X chromosome. Nucleotide diversity (\(\pi\)) based on average number of pairwise differences varies from a low of 0.004 ± 0.003 in the sisA locus to a high of 0.038 ± 0.006 in the per locus in the Kaibab population. The average nucleotide diversity for the five loci is 0.016 ± 0.002 in the Davis Mountains population. The estimates of genetic diversity within the per locus are slightly less when the two sequences with the diverged 5′ regions (PSU 633 and PSU 996) are removed from the estimates of \(\pi\) and \(\theta\), but the reductions are not significantly less than the total data set (results not shown). Nucleotide heterozygosities based on the number of segregating sites (\(\Theta\)) were \(\geq \pi\) for all of the genes or the concatenated set of genes.

The per, Est-5, and Hsp83 genes have been the focus of study.
<table>
<thead>
<tr>
<th></th>
<th>XL</th>
<th>per</th>
<th>XR</th>
<th>Est5</th>
<th>Hsp83</th>
<th>All loci</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>run</td>
<td>sisA</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Variation within Kaibab NF (n = 20)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$S_s$ ($S_t$)</td>
<td>30 (20)</td>
<td>7 (3)</td>
<td>103 (68)</td>
<td>32 (13)</td>
<td>18 (12)</td>
<td>190 (116)</td>
</tr>
<tr>
<td>$\pi$</td>
<td>0.013 ± 0.001</td>
<td>0.004 ± 0.001</td>
<td>0.038 ± 0.008</td>
<td>0.015 ± 0.001</td>
<td>0.007 ± 0.001</td>
<td>0.016 ± 0.002</td>
</tr>
<tr>
<td>$\Theta$</td>
<td>0.020 ± 0.004</td>
<td>0.004 ± 0.002</td>
<td>0.060 ± 0.006</td>
<td>0.017 ± 0.003</td>
<td>0.011 ± 0.003</td>
<td>0.023 ± 0.002</td>
</tr>
<tr>
<td>Variation within Davis Mtn (n = 20)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$S_s$ ($S_t$)</td>
<td>33 (24)</td>
<td>8 (3)</td>
<td>100 (48)</td>
<td>29 (12)</td>
<td>13 (8)</td>
<td>183 (95)</td>
</tr>
<tr>
<td>$\pi$</td>
<td>0.014 ± 0.001</td>
<td>0.004 ± 0.001</td>
<td>0.045 ± 0.009</td>
<td>0.012 ± 0.001</td>
<td>0.006 ± 0.001</td>
<td>0.016 ± 0.002</td>
</tr>
<tr>
<td>$\Theta$</td>
<td>0.022 ± 0.004</td>
<td>0.005 ± 0.002</td>
<td>0.063 ± 0.006</td>
<td>0.016 ± 0.003</td>
<td>0.008 ± 0.002</td>
<td>0.023 ± 0.002</td>
</tr>
<tr>
<td>Net divergence between populations</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$D_a$</td>
<td>-0.0001 ± 0.001</td>
<td>-0.0001 ± 0.001</td>
<td>-0.0003 ± 0.008</td>
<td>-0.0001 ± 0.001</td>
<td>-0.0002 ± 0.001</td>
<td>-0.0002 ± 0.002</td>
</tr>
<tr>
<td>Variation in combined populations (n = 40)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$S_s$ ($S_t$)</td>
<td>46 (27)</td>
<td>10 (4)</td>
<td>127 (43)</td>
<td>40 (15)</td>
<td>21 (11)</td>
<td>244 (100)</td>
</tr>
<tr>
<td>$\pi$</td>
<td>0.013 ± 0.001</td>
<td>0.004 ± 0.0004</td>
<td>0.041 ± 0.006</td>
<td>0.013 ± 0.001</td>
<td>0.006 ± 0.001</td>
<td>0.016 ± 0.001</td>
</tr>
<tr>
<td>$\Theta$</td>
<td>0.026 ± 0.004</td>
<td>0.005 ± 0.002</td>
<td>0.067 ± 0.006</td>
<td>0.019 ± 0.003</td>
<td>0.011 ± 0.002</td>
<td>0.026 ± 0.002</td>
</tr>
<tr>
<td>$D_f$</td>
<td>-1.871*</td>
<td>-0.722</td>
<td>-1.571</td>
<td>-1.151</td>
<td>-1.573</td>
<td>-1.579</td>
</tr>
<tr>
<td>$D_{FL}$</td>
<td>-3.011*</td>
<td>-1.035</td>
<td>-1.883</td>
<td>-1.528</td>
<td>-2.253</td>
<td>-1.829</td>
</tr>
<tr>
<td>$3N_c$</td>
<td>&gt;10,000</td>
<td>0.070</td>
<td>0.005</td>
<td>0.107</td>
<td>0.070</td>
<td></td>
</tr>
<tr>
<td>$c/\mu$</td>
<td>$\infty$</td>
<td>17.5</td>
<td>0.122</td>
<td>8.23</td>
<td>11.7</td>
<td></td>
</tr>
<tr>
<td>$Nm$</td>
<td>16.2</td>
<td>25.4</td>
<td>14.6</td>
<td>16.0</td>
<td>29.3</td>
<td>16.1</td>
</tr>
<tr>
<td>Average divergence between $D. \text{pseudoobscura}$ and $D. \text{miranda}$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$Div$</td>
<td>0.041 ± 0.006</td>
<td>0.010 ± 0.002</td>
<td>0.142 ± 0.022</td>
<td>0.034 ± 0.005</td>
<td>0.024 ± 0.004</td>
<td>0.049 ± 0.008</td>
</tr>
</tbody>
</table>

$S_s$, number of segregating sites; $S_t$, number of segregating sites where the rare variant had a frequency of 1; $\pi$, nucleotide heterozygosity, $3N_s$, based on the number of pairwise differences; $\Theta$, nucleotide heterozygosity, $3N_s$, based on the number of segregating sites; $3N_c$, recombination parameter estimated between adjacent sites with the method of Hudson (1987); $D_f$, net divergence between populations Nei (1987); $D_{FL}$, Tajima’s (1989) $D$; $D_{FL}$, Fu and Li’s (1993) $D$ that used $D. \text{miranda}$ as an outgroup; $N$, effective population size; $\mu$, neutral mutation rate; $c$, neutral recombination rate; $m$, migration rate; $Div$, average divergence per site with a Jukes and Cantor (1969) correction. *, an observed $D_f$ or $D_{FL}$ value with a probability <0.05.
of previous nucleotide sequence analysis studies. Our estimate of nucleotide diversity in the fourth intron of per is higher ($\pi = 0.041$) than that observed by Wang and Hey (1996) in their study of 1.5 kb of the per transcript from exon 5 to exon 8 ($\pi = 0.008$). This difference is likely to be due to variation in functional constraints because the ratio of polymorphism to the divergence between D. pseudoobscura and D. miranda is similar for the two regions. Our estimate of variation within the intergenic region between Est-5B and Est-5C is similar to that estimated for the same region previously examined by Babcock and Anderson (1996) and within the range of values estimated for noncoding regions of all Est-5 genes ($\pi = 0.004$–0.0148) by King (1998).

Finally, our estimates of variation within the Hsp83 intron are not significantly different from that observed in 2-kb transcript sequenced by Wang et al. (1997). We can determine if the number of segregating sites is too high or low given our average value of $\Theta$. Kreitman and Hudson (1991) proposed that one could estimate the probability of observing a given number of polymorphic sites based on $\Theta$. The average value of $\Theta$ estimated from the five loci was $0.026 \pm 0.002$. This test shows that the numbers of segregating sites at Hsp83 and sisA are too small and the number of polymorphic nucleotides at per is too high based on the $\Theta$ value we observed (Equation 3 and see the test described on p. 571 of Kreitman and Hudson 1991). We test whether selective forces are responsible for the significant reduction and elevation in the numbers of segregating sites in sisA, Hsp83, and per below.

Gene flow estimates between the KB and DM populations: We tested whether the estimates of $\pi$ differed significantly between the two populations by examining a 95% confidence interval of $D_\theta$, the average net divergence between two populations (Equations 10.21 and 10.25 in Nei 1987). We found no evidence for genetic differentiation between populations at any of the five loci or for the concatenated data set (Table 3). The neutral migration parameter $N m$ was estimated with the method of Nei (1982). This method estimates $\gamma_0$, the average inbreeding coefficient among individuals between subpopulations. The neutral migration parameter assuming an island model of migration is estimated by

$$N m = \frac{1 - \gamma_0}{3\gamma_0}.$$

The neutral migration parameter was estimated for the five loci separately and for a concatenated set of the five loci. Table 3 shows estimates of the neutral migration parameter for the five genes separately and for the concatenated set of five genes. The five genes from the X chromosome estimate gene flow to be 14.6 migrants per generation or greater. In addition, the concatenated set of genes estimates that 16.1 migrants are exchanged between the KB and DM populations.

We also used the maximum-likelihood approach of Beerli and Felsenstein (1999) that simultaneously estimates the effective population size and migration rates scaled by the mutation rate. Table 4 shows the results of the Beerli and Felsenstein analysis. The maximum-likelihood method shows extensive migration between the two populations with the Hsp83, per, and sisA markers. The Est-5 and run loci show that extensive migration occurs from the Kaibab National Forest to the Davis Mountains; however, migration is limited in the reverse direction.

Given the lack of significant genetic differentiation between the KB and DM populations, the statistical tests of the neutral mutation hypothesis and the analyses of linkage disequilibrium and recombination were performed on the combined sample of 40 chromosomes from the two populations.

Tests of the neutral mutation hypothesis: The Tajima (1989) and Fu and Li (1993) statistics test frequency spectra for excesses of either rare or intermediate frequency variants. Tajima’s test statistic $D_\theta$ tests the null hypothesis that two estimates of the neutral mutation parameter ($\pi$) and ($\Theta$) are equal. Fu and Li’s (1993) test statistic $D_{st}$ detects selection by comparing the observed numbers of internal and external mutations to their expectations under selective neutrality. When the statistics for both tests are significantly negative, there is an excess of rare variants consistent with purifying or directional selection. On the other hand, a significantly positive test statistic indicates balancing selection due to an excess of intermediate frequency variants. Both tests were performed on each X-linked locus and on the concatenated set of the five loci.

The Hsp83, Est-5, per, and sisA loci all failed to reject the selectively neutral hypothesis with both frequency spectrum tests (Table 3). The Fu and Li and Tajima test statistics all had negative values indicating that the frequency spectra were biased toward rare variants. The run locus, on the other hand, showed a significant excess of rare variants with the Fu and Li and the Tajima tests. The Tajima and Fu and Li tests failed to reject an equilibrium neutral model for the five concatenated loci within each population and for the total data set. The frequency spectra tests for run could depart from neutral expectations if D. pseudoobscura populations were subdivided, but our analysis of gene flow finds no evidence of population subdivision at any of the five loci. In addition, there were no fixed differences observed between the KB and DM populations.

The Hudson-Kreitman-Aguadé (HKA) test (Hudson et al. 1987) tests the neutral mutation hypothesis with intraspecific and interspecific nucleotide sequence data. This test utilizes a goodness-of-fit test that determines whether the levels of intraspecific polymorphism and interspecific divergence are equivalent in two unlinked loci. A deficiency of polymorphic sites within species indicates directional selection, while an excess of segre-
gating sites within species suggests the action of balancing selection. The HKA test was performed on all pairs of five X-linked loci and each X-linked locus against Adh (Schaeffer and Miller 1992a, b; 1993), an autosomal gene located on the fourth chromosome. We used the HKA correction for comparisons of X-linked and autosomal loci described by Begun and Aquadro (1991) for all comparisons of Adh with the five genes from the X chromosome.

None of the HKA tests found a significant departure from an equilibrium neutral model either in tests that compared pairs of the X-linked loci or in comparisons of Adh with each X-linked locus (results not shown). Thus, the level of polymorphism within species does not significantly differ from the divergence level between species when corrected for the average coalescence time. A multiple comparison correction would not change these results given that none of the tests reject the null hypothesis. Figure 4 shows the relationship between polymorphism and divergence for 13 genes from the four major chromosomes of D. pseudoobscura. There is a strong linear relationship between levels of polymorphism and divergence, where divergence for a genetic locus is twice the value of polymorphism.

In summary, the sixA, per, Est-5, and Hsp83 markers should give unbiased estimates of gene flow because they each fail to reject an equilibrium neutral model with three tests of selective neutrality. On the other hand, the run locus might underestimate the neutral migration parameter if the significant excess of rare variants revealed with the Tajima and Fu and Li tests was discovered to occur preferentially in one population vs. the other. This is not the case for the run locus where the frequency of rare variants is not significantly different between the two populations (Fisher’s exact test, P = 0.609).

Sex-ratio inversions: We have evidence that the chromosomal sequences of four strains from the Kaibab population (PSU 603, PSU 627, PSU 629, and PSU 630) and two strains from the Davis Mountains population (PSU 1006 and PSU 1015) carry the Sex-ratio inversion (Figure 3). First, Babcock and Anderson (1996) showed that the Est-5 gene is within the subbasal SR inversion and that the Est-5 sequence data indicate that SR is monophyletic. We tested for the presence of SR chromosomes in our sample with a phylogenetic analysis of the Est-5 gene using the neighbor-joining method (Saitou and Nei 1987), using Kimura’s two-parameter model and total elimination of insertions and deletions. The phylogenetic tree of the Est-5 sequences from this study and the Est-5 Sex-ratio alleles of Babcock and Anderson (1996) are shown in Figure 5. The six strains PSU 603, PSU 627, PSU 629, PSU 630, PSU 1006, and PSU 1015 form a monophyletic group with the Est-5 genes sequenced from Sex-ratio chromosomes by Babcock and Anderson (1996). The monophyletic group of Sex-ratio chromosomes clusters together in 99% of the bootstrap replicates we performed in the phylogenetic analysis.

Second, the Est-5 genes from Sex-ratio chromosomes

<table>
<thead>
<tr>
<th>Gene</th>
<th>(\theta_{ha})</th>
<th>(\theta_{otd})</th>
<th>(3N_m(\text{KB} \rightarrow \text{DM}))</th>
<th>(3N_m(\text{DM} \rightarrow \text{KB}))</th>
</tr>
</thead>
<tbody>
<tr>
<td>run</td>
<td>0.356</td>
<td>0.055</td>
<td>194.3</td>
<td>5.5 \times 10^{-12}</td>
</tr>
<tr>
<td>sixA</td>
<td>0.002</td>
<td>0.006</td>
<td>54.0</td>
<td>11.8</td>
</tr>
<tr>
<td>per</td>
<td>0.509</td>
<td>0.090</td>
<td>6.50</td>
<td>127</td>
</tr>
<tr>
<td>Est-5</td>
<td>0.033</td>
<td>0.399</td>
<td>216</td>
<td>0.717</td>
</tr>
<tr>
<td>Hsp83</td>
<td>0.008</td>
<td>0.010</td>
<td>17.9</td>
<td>45.4</td>
</tr>
<tr>
<td>All loci</td>
<td>0.085</td>
<td>0.492</td>
<td>153</td>
<td>1.32</td>
</tr>
</tbody>
</table>
Figure 5.—Neighbor-joining phylogenetic tree of the Est-5 genes sequenced in this study and the Sex-ratio chromosomes sequenced in BABCOCK and ANDERSON (1996). This tree used the Kimura two-parameter model, complete elimination of gaps in the alignment, and 1000 bootstrap replicates. Sequences with a PSU prefix were determined in this study while the SR prefixes were determined in the BABCOCK and ANDERSON (1996) study. Kaibab National Forest strains are numbered PSU 596–633 and the Davis Mountains strains are numbered PSU 994–1029.
sequenced by Babcock and Anderson (1996) were fixed for a 5-bp deletion not shared with Est-5 genes in Standard chromosomes. The six strains PSU 603, PSU 627, PSU 629, PSU 630, PSU 1006, and PSU 1015 all carried this 5-bp deletion (M. Kovacevic and S. W. Schaeffer, data not shown).

Third, we asked if the observed number of Sex-ratio chromosomes is consistent with the population frequency of SR given our sampling strategy. A sample of 20 males was drawn from each of 328 and 71 strains collected from the Kaibab and Davis Mountains populations, respectively, for our study of X-linked genes. We used a Monte Carlo simulation to generate 328 and 71 female genotypes [Standard (ST)/ST, ST/ST, and SR/ST] from the KB and DM populations, respectively, assuming the populations were in Hardy-Weinberg equilibrium and the SR frequencies were 0.166 and 0.149 in the KB and DM populations, respectively (Anderson et al. 1967). Twenty random females from each population were sampled and an X chromosome was chosen to simulate the random collection of males for our sequencing study. We counted the number of Sex-ratio chromosomes in this random sample of 20 X chromosomes. We repeated the simulation 1000 times and derived a mean and a 95% confidence interval from the 1000 replicate runs. The frequency of Sex-ratio chromosomes found in our samples is consistent with an average Sex-ratio frequency of 15% in each population (KB, expected number of SR chromosomes = 3.3, 95% confidence interval = 0–7; DM, expected number of SR chromosomes = 3.0, 95% confidence interval = 0–7). On the basis of the phylogenetic evidence and the observed frequencies of the Sex-ratio chromosomes within our samples of males in the two populations, we conclude that these 6 males collected from the two populations carried the Sex-ratio inversion on XR.

Genetic diversity within Sex-ratio and Standard chromosomes: We can explore the effect that the Sex-ratio meiotic drive system had on linked variation by estimating levels of diversity in the five X-linked genes within each gene arrangement. We partitioned the 40 chromosomes from the two populations into 6 Sex-ratio and 34 Standard chromosomes and estimated population parameters for the five loci individually and for the concatenated set of genes (Table 5). As expected, levels of heterozygosity (\( \pi \)) within the Standard chromosomes (\( n = 34 \)) did not differ significantly from the total data set of 40 chromosomes. Nucleotide diversity within Sex-ratio chromosomes was equivalent to that in Standard chromosomes at all loci except Est-5 where SR chromosomes had sixfold less diversity than ST chromosomes. The ratio of \( \pi \) in Sex-ratio to \( \pi \) in Standard for the five loci shows that the Est-5 locus has the greatest reduction in genetic variation associated with the Sex-ratio phenotype (Table 5).

We used a random permutation method (Erron and Tibshirani 1993) to determine (1) if the ratio of \( \pi \) in Sex-ratio to \( \pi \) in Standard in Est-5 is significantly <1.0; and (2) if the the ratio of \( \pi \) in Sex-ratio to \( \pi \) in Standard in Est-5 is significantly less than the same ratio in the four other genes. The \( \pi \) ratio is expected to be 1.0 if the variation in Sex-ratio chromosomes is a subset of that found in Standard chromosomes. The random permutation test was performed by randomly assigning 6 of the 40 chromosomes sequenced in this study to the Sex-ratio chromosome without replacement. The remaining 34 chromosomes were assigned to the Standard arrangement group. This random data set was used to estimate the ratio of \( \pi \) in Sex-ratio to \( \pi \) in Standard separately for the five genes. We repeated the resampling procedure 1000 times and estimated a mean and two-tailed 95% confidence interval from the rank-ordered random distribution of values (Table 5). The random permutation test showed that each gene had a mean ratio of ~1.0 for the five genes. The observed ratio of \( \pi \) in Sex-ratio to \( \pi \) in Standard in Est-5 is significantly less than expected when chromosomes are randomly assigned to the Sex-ratio chromosome. In contrast, the observed \( \pi \) ratios in run, sisA, per, and Hsp83 were not significantly different from 1.0, suggesting that these loci are now independent of the Sex-ratio meiotic drive system. The observed reduction in variation in Est-5 Sex-ratio chromosomes is also significantly less than that observed in the four other loci (Table 5).

**Sex-ratio and recombination across the X chromosome:** We can determine the impact that the Sex-ratio meiotic drive system has had on variation in linked genes by examining levels of recombination and linkage disequilibrium within and between genes on the X chromosome. Table 3 shows the estimates of the neutral recombination parameter \( (3Nc) \) of Hudson (1987) within each of the five gene segments. Estimates of the recombination parameter vary among the five loci with the per locus having the lowest estimate of \( 3Nc \) and the run locus having the highest estimate of \( 3Nc \). The recombination parameter can be standardized relative to the neutral mutation parameter by taking a ratio of the two estimates \( (3Nc/3Nmu \text{ or } \mu/c) \) to determine how many recombination events occur for each mutation (Table 3). At run, sisA, Est-5, and Hsp83, at least seven recombination events occur for each mutation event. The opposite is true at the per locus where eight mutations have occurred for each recombination event. The two divergent alleles PSU 633 and PSU 996 are responsible for this low recombination to mutation ratio because when these two alleles are removed from the calculation, one recombination event is observed for each mutation. The sisA and Hsp83 genes have the highest \( \mu/c \) ratios compared to other loci, but these values should be viewed with caution due to the limited amount of nucleotide diversity found in the two regions.

We tested pairs of segregating sites within and between loci for departures from random association or linkage equilibrium with Fisher’s exact test (Sokal and
**TABLE 5**

Estimates of the neutral mutation and recombination parameters within *Sex-ratio* and Standard chromosomes for five X-linked loci in *Drosophila pseudoobscura*

<table>
<thead>
<tr>
<th></th>
<th><em>run</em></th>
<th><em>sisA</em></th>
<th><em>per</em></th>
<th><em>Est-5</em></th>
<th><em>Hsp83</em></th>
<th>All loci</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>π</em></td>
<td>0.016 ± 0.002</td>
<td>0.004 ± 0.001</td>
<td>0.035 ± 0.006</td>
<td>0.002 ± 0.001</td>
<td>0.005 ± 0.001</td>
<td>0.012 ± 0.002</td>
</tr>
<tr>
<td>Θ</td>
<td>0.018 ± 0.004</td>
<td>0.004 ± 0.002</td>
<td>0.035 ± 0.006</td>
<td>0.003 ± 0.001</td>
<td>0.004 ± 0.002</td>
<td>0.012 ± 0.002</td>
</tr>
<tr>
<td>3<em>Nc</em></td>
<td>&gt;10,000</td>
<td>0.068</td>
<td>0.136</td>
<td>&gt;10,000</td>
<td>0.026</td>
<td></td>
</tr>
<tr>
<td><em>c/μ</em></td>
<td>∞</td>
<td>17.0</td>
<td>3.89</td>
<td>∞</td>
<td>5.20</td>
<td></td>
</tr>
</tbody>
</table>

**Sex-ratio chromosomes (n = 6)**

<table>
<thead>
<tr>
<th></th>
<th><em>run</em></th>
<th><em>sisA</em></th>
<th><em>per</em></th>
<th><em>Est-5</em></th>
<th><em>Hsp83</em></th>
<th>All loci</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>π</em></td>
<td>0.013 ± 0.001</td>
<td>0.004 ± 0.0004</td>
<td>0.044 ± 0.007</td>
<td>0.012 ± 0.001</td>
<td>0.006 ± 0.001</td>
<td>0.016 ± 0.002</td>
</tr>
<tr>
<td>Θ</td>
<td>0.024 ± 0.004</td>
<td>0.005 ± 0.002</td>
<td>0.069 ± 0.006</td>
<td>0.019 ± 0.003</td>
<td>0.012 ± 0.003</td>
<td>0.026 ± 0.002</td>
</tr>
<tr>
<td>3<em>Nc</em></td>
<td>&gt;10,000</td>
<td>0.089</td>
<td>0.004</td>
<td>0.178</td>
<td>0.096</td>
<td></td>
</tr>
<tr>
<td><em>c/μ</em></td>
<td>∞</td>
<td>22.2</td>
<td>0.091</td>
<td>14.8</td>
<td>16.0</td>
<td></td>
</tr>
</tbody>
</table>

**Standard chromosomes (n = 34)**

**Comparison of *Sex-ratio* and Standard chromosomes**

<table>
<thead>
<tr>
<th></th>
<th><em>run</em></th>
<th><em>sisA</em></th>
<th><em>per</em></th>
<th><em>Est-5</em></th>
<th><em>Hsp83</em></th>
<th>All loci</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>π</em></td>
<td>1.231</td>
<td>1.000</td>
<td>0.795</td>
<td>0.167</td>
<td>0.833</td>
<td></td>
</tr>
<tr>
<td>95% confidence interval</td>
<td>0.661–1.412</td>
<td>0.547–1.657</td>
<td>0.411–2.213</td>
<td>0.590–1.413</td>
<td>0.327–1.811</td>
<td></td>
</tr>
</tbody>
</table>

**Net divergence between *Sex-ratio* and Standard chromosomes**

<table>
<thead>
<tr>
<th></th>
<th><em>run</em></th>
<th><em>sisA</em></th>
<th><em>per</em></th>
<th><em>Est-5</em></th>
<th><em>Hsp83</em></th>
<th>All loci</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>D_s</em></td>
<td>-0.0001 ± 0.002</td>
<td>-0.0002 ± 0.001</td>
<td>0.0005 ± 0.007</td>
<td>0.014 ± 0.002</td>
<td>0.001 ± 0.001</td>
<td>0.003 ± 0.002</td>
</tr>
</tbody>
</table>

**Net divergence between *D. pseudoobscura* Standard chromosomes and *D. miranda***

<table>
<thead>
<tr>
<th></th>
<th><em>run</em></th>
<th><em>sisA</em></th>
<th><em>per</em></th>
<th><em>Est-5</em></th>
<th><em>Hsp83</em></th>
<th>All loci</th>
</tr>
</thead>
<tbody>
<tr>
<td>Div</td>
<td>0.035 ± 0.007</td>
<td>0.008 ± 0.002</td>
<td>0.119 ± 0.025</td>
<td>0.028 ± 0.006</td>
<td>0.022 ± 0.004</td>
<td>0.041 ± 0.008</td>
</tr>
</tbody>
</table>

*π*, nucleotide heterozygosity, 4*Nₜ*, based on the number of pairwise differences; Θ, nucleotide heterozygosity, 4*Nₜ*, based on the number of segregating sites; 3*Nₑ*, recombination parameter estimated between adjacent sites with the method of Hudson (1987); Νₑ, effective population size; μₑ, neutral mutation rate; cₑ, neutral recombination rate; *Dₐ*, net divergence between populations Nei (1987); Div, average divergence per site with a Jukes and Cantor (1969) correction.
A total of 33,654 pairwise comparisons are possible for the 244 segregating sites located in the five X-linked genes (25,878 comparisons of 228 segregating sites with two nucleotides + 7296 comparisons of 228 segregating sites with two nucleotides and 16 segregating sites with three nucleotides + 480 comparisons of 16 segregating sites with three nucleotides). Lewontin (1995) demonstrated that it is impossible to reject the null hypothesis of random association with Fisher’s exact test for some comparisons of two sites where the linkage disequilibrium value is negative. Of the 33,654 possible pairwise comparisons of segregating sites on the X chromosome, only 6834 pairwise tests are capable of rejecting the null hypothesis of independence given the marginal frequencies observed at the polymorphic sites and the sign of linkage disequilibrium. Thirty-six of the 6834 (0.527%) of the valid pairwise comparisons of segregating sites were nonrandomly associated when Fisher’s exact test was used with the sequential Bonferroni correction.

**The age of the Sex-ratio inversion:** We can estimate the age of the Sex-ratio inversion in two ways. First, we can use the net genetic divergence between the Sex-ratio and Standard chromosomes and derive an age for Sex-ratio from the substitution rate estimated by Caccione et al. (1988; 1.7% sequence divergence per million years). The net divergence between Sex-ratio and Standard chromosomes is not significantly different from zero at run, sisA, per, and Hsp83 (Table 5), but is significantly different from zero in Est-5. On the basis of 1.4% net divergence in Est-5, we estimate Sex-ratio inversion to be 800,000 years old. Alternatively, we can use the ratio of net divergence between Sex-ratio and Standard vs. net divergence between Standard chromosomes and D. miranda (approach suggested in Hasson and Eanes 1996). The divergence between Sex-ratio and Standard is half that observed between Standard and D. miranda. This suggests that the Sex-ratio inversion emerged half of the 2.1 million years that separates D. miranda and D. pseudoobscura, or one million years ago. These two estimates are consistent with those of Babcock and Anderson (1996), who estimated Sex-ratio to be 700,000 to 1.3 million years old.

The second method to estimate the age of the Sex-ratio inversion uses the decomposition of linkage disequilibrium between Est-5 and Hsp83 to get a rough estimate of the minimum number of generations since the Sex-ratio chromosome originated. The two observations that motivate this approach are that Hsp83 is clearly linked to Est-5 either within the subbasal inversion or in between the subbasal and central inversion (Figure 2) and that three haplotypes exist at Hsp83 on Sex-ratio chromosomes (Figure 3). We can estimate the time of origin of Sex-ratio by asking how long it would take for genetic exchange, either through double crossovers or gene conversion, to move Hsp83 haplotypes found on a Standard background to the Sex-ratio background.

If we assume that a single haplotype at the Hsp83 locus was associated with the Sex-ratio trait at the time of its origin, then we can derive an estimate of the age of Sex-ratio from the current linkage disequilibrium at time \(t\) from the relationship

\[
D_t = D_0(1 - r)^t,
\]

where \(r\) is the recombination rate between the two loci and \(D_0\) is the initial level of linkage disequilibrium at the time 0 or time of origin. Solving for \(t\),

\[
t = \frac{\ln D_t - \ln D_0}{\ln(1 - r)},
\]

where \(t\) is the number of generations since the origin of the Sex-ratio inversion. The likely common ancestral...
haplotype at the \( \text{Hsp83} \) locus is that shown in Figure 7A because two of the three segregating sites are shared with \( \text{D. miranda} \). In addition, this haplotype is not observed in our sample of Standard chromosomes, while one of the two haplotypes is present within the Standard gene arrangements and the other haplotype appears to have picked up a segregating nucleotide through a gene conversion event (data not shown). The initial linkage disequilibrium in the population would depend on the frequency of the ancestral \( \text{Hsp83} \) haplotype on the Standard chromosomes. We consider two cases, one where the frequency of the ancestral \( \text{Hsp83} \) haplotype was at an intermediate frequency in the original population (0.5), or a second where the frequency of the ancestral \( \text{Hsp83} \) haplotype was low (0.029; Figure 7A).

To compare the current and past states of linkage disequilibrium, the sample size was assumed to be 40 and the number of \( \text{Sex-ratio} \) chromosomes to be six. Thus, \( D_0 \) between \( \text{Est-5} \) and \( \text{Hsp83} \) was 0.128 or 0.064 at the origin of \( \text{Sex-ratio} \), assuming frequencies of the \( \text{Hsp83} \) ancestral haplotype to be 0.5 and 0.029, respectively. The current level of linkage disequilibrium, \( D_r \), is 0.042 (Figure 7A).

We used the ratio of genetic distances estimated between \( \text{magenta} \) and \( \text{short} \) in Standard homozygotes and \( \text{Sex-ratio/Standard} \) heterozygotes to derive a rough estimate for genetic exchange between \( \text{Est-5} \) and \( \text{Hsp83} \) in \( \text{Sex-ratio/Standard} \) heterozygotes. The current map of the \( \text{D. pseudoobscura} \) X chromosome shows that \( \text{magenta} \) and \( \text{short} \) are separated by a distance of 110.9 cM based on genetic mapping data from Standard homozygotes (Sturtevant and Dobzhansky 1936). Sturtevant and Dobzhansky (1936) estimated the distance from \( \text{magenta} \) to \( \text{short} \) to be 0.5 cM in \( \text{Sex-ratio/Standard} \) heterozygotes (Figure 7B). Our estimate of genetic distance between \( \text{Est-5} \) and \( \text{Hsp83} \) is 2.5 cM in a Standard background. Based on these data, a maximum percentage crossing over between \( \text{Est-5} \) and \( \text{Hsp83} \) in \( \text{Sex-ratio/Standard} \) heterozygotes is \( r = 1.1 \times 10^{-4} \). Genetic exchange between \( \text{Sex-ratio} \) and Standard chromosomes only occurs in females; thus, the observed crossing-over rate must be reduced by one-half or \( r = 5.6 \times 10^{-5} \). The average frequency of \( \text{Sex-ratio/Standard} \) heterozygotes is 0.25 based on the estimates of the frequency of \( \text{Sex-ratio} \) chromosomes in the desert southwest [frequency \( (\text{Sex-ratio}) = 0.15 \); Anderson et al. 1967]; thus the observed crossing-over rate must be reduced by an additional one-fourth or \( r = 1.4 \times 10^{-5} \). From this recombination rate per generation, we estimate the age of the \( \text{Sex-ratio} \) inversion to be 29,000 to 76,000 generations old. If we assume that there are three generations per year, then \( \text{Sex-ratio} \) is at least 10,000–25,000 years old.

**DISCUSSION**

Extensive gene flow between the Kaibab and Davis Mountain populations of \( \text{D. pseudoobscura} \): The estimates of the neutral migration parameter based on five \( \text{X-linked} \) loci are consistent with previous estimates of gene flow in \( \text{D. pseudoobscura} \) (Schaeffer and Miller 1992a). Our estimate of \( N_m \) suggests that at least 14 individuals are exchanged between the two populations per generation. This level of dispersal is sufficient to homogenize gene frequencies at all loci in the genome. Allele frequencies at enzyme encoding loci have been shown to be homogenized on all \( \text{D. pseudoobscura} \) chromosomes except for the third chromosome, where the paracentric inversions are segregating (Lewontin 1974). The extensive gene flow observed in \( \text{D. pseudoobscura} \) suggests that the clines in gene arrangement frequency result from selection in heterogeneous environments. The frequencies of \( \text{Sex-ratio} \) in natural populations of \( \text{D. pseudoobscura} \) form a north to south gradient. The \( \text{Sex-ratio} \) chromosome is not detected in the northwestern United States and Canada, but is present in the southwestern United States and Mexico at average frequencies of 15% (Dobzhansky 1944; Anderson et al. 1967).
Powell (1992) suggested that the clinal pattern in Sex-ratio frequencies observed in both D. pseudoobscura and D. persimilis hints that natural selection could be responsible for the geographic differentiation. This study and that done by Schaeffer and Miller (1992a) strongly support Powell’s hypothesis because the migration parameter estimates in D. pseudoobscura are sufficient to homogenize gene frequencies throughout the genome, including the Sex-ratio trait.

The frequencies of the paracentric inversions on the third chromosome of D. pseudoobscura also form a clinal pattern in the western United States. The Kaibab population is dominated by the Arrowhead inversion, where frequencies reach over 90%, and the Pikes Peak inversion is virtually absent (Anderson et al. 1991). The Davis Mountain population, on the other hand, is dominated by the Pikes Peak inversion, whose frequency has ranged from 56 to 82% over the last 60 years. The frequency of the Arrowhead inversion varies from 15 to 33% in the Davis Mountains (Anderson et al. 1991). The levels of gene flow estimated for the X-linked genes in this study predict that the gene arrangement frequencies in the two populations will be similar. These data reinforce the conclusion of Schaeffer and Miller (1992a), who suggested that selection must be quite strong to offset the homogenizing effects of migration. These data predict that homosequential inversions on the third chromosome collected from different populations will not be genetically differentiated, which is counter to the selection experiments done by Dobzhansky (1948) that suggested that homosequential inversions from different populations contain different genetic information.

Levels of variation and selection on the X chromosome of D. pseudoobscura: The five loci on the X chromosome of D. pseudoobscura differed in the level of genetic diversity found within each of the gene segments even though the sequences that we examined were in non-coding regions. The Hsp83 intron and the sixA 3’ flanking region had fewer segregating sites than expected given the average heterozygosity level for the five X-linked genes, while the per intron had more segregating sites than expected. This difference could be due to differences in selective constraint among the regions, to differences in the background selection process in the different gene regions, or to differences in genetic hitchhiking events in the regions. The HKA test rules out the last two alternatives because we do not see a significant reduction or increase in the levels of polymorphism within species as a function of between-species divergence (Table 4 and Figure 4). Therefore, the likely explanation for the observed differences in diversity levels is that the five regions have different selective constraints on the sequences.

Hamblin and Aquadro (1999) noted that 9 of 10 polymorphism studies in D. pseudoobscura have shown negative values of Tajima’s D. In other words, most loci have more rare variants than intermediate frequency mutations. More rare variants in genes are expected either when a population is rapidly expanding or when purifying selection is removing diversity. Our data add five genes to the total number of loci that have negative Tajima’s D’s. Analysis of the mismatch distribution (Rogers and Harpending 1992) strongly supports Hamblin and Aquadro’s argument that D. pseudoobscura has experienced a rapid population expansion. The mismatch distributions for the five loci are all consistent with a recent expansion of the D. pseudoobscura population (M. Kovacevic and S. W. Schaeffer, data not shown).

Evolution of the Sex-ratio chromosome: Babcock and Anderson (1996) used a nucleotide sequence analysis of the intergenic region between the Est-5B and Est-5C genes to follow the evolutionary history of the Sex-ratio chromosome. The Est-5 region is located within the subbasal inversion on XR where the Sex-ratio trait maps (Wallace 1948; Beckenbach 1996). Babcock and Anderson found that the Est-5 sequences within Sex-ratio chromosomes had low levels of diversity compared to Est-5 sequences within Standard chromosomes and that the gene arrangement is monophyletic. We also observed a significant reduction in nucleotide diversity within the Est-5 locus and estimated the inversion to be 800,000 to 1 million years old with genetic divergence data.

The reduction of nucleotide diversity at Est-5 is consistent with studies of allozyme diversity on XR that showed fewer allozyme alleles within Sex-ratio chromosomes than within the Standard gene arrangements (Praakash and Merritt 1972; Praakash 1974). Previous electrophoretic studies, however, did not examine the frequencies of allozyme alleles for enzymes located on the left arm of the X chromosome (Praakash and Merritt 1972; Praakash 1974). This study shows that genes on XL are exchanged between Standard and Sex-ratio chromosomes in spite of the meiotic drive system operating on XR within D. pseudoobscura (Table 5).

Our data fail to show the strong associations of Sex-ratio with linked genes that were observed by Praakash and Merritt (1972). Seven nucleotide sites within Est-5 are in significant linkage disequilibrium with the Sex-ratio chromosome, but the polymorphic sites within run, sixA, per, and Hsp83 fail to show strong associations with Est-5 or SR. The Est-5 gene had a significant reduction of nucleotide diversity in Sex-ratio chromosomes, but none of the other loci showed the same decrease in genetic variation (Table 5). This observation is quite profound because the Hsp83 locus is tightly linked to Est-5 in Sex-ratio/Standard heterozygotes. One might expect to see a correlated history of these two loci; however, the genetic exchange between Est-5 and Hsp83, either by double crossovers or gene conversion, provided an excellent opportunity to estimate the age of Sex-ratio.
The age estimate of 25,000–76,000 generations for the Sex-ratio chromosome based on the decomposition of linkage disequilibrium is lower than the estimate based on net divergence (Prakash and Merritt 1972; Prakash 1974; Babcock and Anderson 1996). This estimate should be viewed as a minimum estimate of the age of the Sex-ratio inversion because we did not estimate the recombination rate between Est-5 and Hsp83 in inversion heterozygotes directly. If the actual recombination rate between these two genes is $<1.4 \times 10^{-3}$, then we would estimate an even greater age of the Sex-ratio inversion. In addition, the frequency of Sex-ratio/Standard heterozygotes in females is not likely to have been constant over the course of linkage disequilibrium decay. Little recombination between Sex-ratio and Standard chromosomes would occur during the early times of the expansion because Sex-ratio would have been in low frequency. As a consequence, we would expect our estimate of the age of Sex-ratio to be even greater if we took the historical expansion into account.

The meiotic drive that acts on Sex-ratio chromosomes is a powerful form of selection that is expected to fix the SR chromosome in natural populations. SR/SR females have reduced fecundity compared to ST/ST and ST/SR genotypes, but this lower component of fitness is insufficient to explain the balanced polymorphism of ST and SR (Beckenbach 1996). Wu (1983) suggested that the Sex-ratio males have a slight disadvantage when they mate with a female that has mated previously. These experiments, however, do not test whether Sex-ratio males also have a disadvantage when they mate first with a female who will multiply mate (Prout and Clark 1996). If the Est-5 gene is near the Sex-ratio gene, then the Est-5 locus does not show an elevated level of synonymous substitution in any of the pairwise HKA tests, which is the characteristic signature of an old balanced polymorphism (Hudson and Kaplan 1988; Kreitman and Hudson 1991). However, the pattern that is observed at Est-5 is more like that observed at the superoxide dismutase (Sod) locus in D. melanogaster (Hudson et al. 1994), which appears to be a relatively recent expansion of an Sod-Slow allosyme from a closely related Sod-Fast ancestor.

The depression of nucleotide diversity in Est-5 could be due to the recent origin of the chromosome, to Est-5 being located near the breakpoint of the subbasal inversion, or to the selection associated with the meiotic drive system. Two pieces of evidence would argue against a recent origin of the Sex-ratio chromosome. First, we observed two fixed derived mutations in Est-5 within Sex-ratio chromosomes that were fixed for the ancestral state in the Standard chromosome. In addition, three derived mutations are still segregating in Est-5 for the sample of Sex-ratio chromosomes that are not found in the Standard chromosomes that were sampled. These sites are responsible for the large net divergence between Standard and Sex-ratio chromosomes that is nearly half the net divergence between Standard chromosomes and D. miranda. Second, sufficient time has passed for genetic exchange between Est-5 and Hsp83 due either to gene conversion or crossing over. It is not likely to be due to single crossovers because recombinants between the three nonoverlapping inversions that comprise the Sex-ratio inversion system are rarely observed (Beckenbach 1996). Thus, it seems unlikely that the recent origin hypothesis explains the lack of diversity within Est-5.

The subbasal inversion of the Sex-ratio arrangement is quite small and Est-5 has been localized near one of the breakpoints, although Wyatt Anderson (University of Georgia, personal communication) has placed the cytogenetic location of Est-5 at least one band away from the inversion breakpoint. Hasson and Eanes (1996) have suggested that inversions are analogous to subdivided populations where genetic exchange is similar to gene flow. Central regions within the Sex-ratio inversion that exchange information at least once every other generation will tend to share polymorphisms with the Standard arrangement. Regions near the breakpoints of the Sex-ratio inversion will tend to diversify (Wesley and Eanes 1994; Andolfatto et al. 1999; Rozas et al. 1999; Andolfatto and Kreitman 2000). In the case of the subbasal inversion of Sex-ratio, genetic exchanges are likely to be limited to gene conversion events rather than crossing over because the inversion is small (Navarro et al. 1997). It is clear that the pattern of genetic variation within Est-5 is consistent with the pattern expected for a gene near an inversion breakpoint that experiences little genetic exchange with the Standard chromosome.

The last explanation for the lack of diversity within the Est-5 gene is that it is closely linked to the gene or genes that cause the meiotic drive responsible for the Sex-ratio phenotype. Est-5 is the only locus in this study to show a significant reduction in genetic diversity on Sex-ratio chromosomes. Genes on either side of Est-5 did not show a significant decrease in nucleotide heterozygosity on Sex-ratio chromosomes even though one gene, Hsp83, is 2.5 cM to the right of Est-5. The Sex-ratio chromosome was in significant linkage disequilibrium with polymorphic nucleotides only within Est-5 and not with the variable sites in the four other genes. The lack of significant nonrandom associations with Hsp83 should be viewed with caution because the frequencies of most variable nucleotides within Hsp83 were low, preventing significance tests from being done (Lewontin 1995).

Are there other genes on X that influence the Sex-ratio phenotype? Wu and Beckenbach (1983) used reciprocal introgressions of Sex-ratio and Standard chromosomes between D. pseudoobscura and D. persimilis to map genes that influence the D. persimilis meiotic drive system. The Sex-ratio chromosome in D. persimilis is homosequential with the Standard chromosome of D. pseudoobscura, allowing the reciprocal translocations to oc-
cur. Wu and Beckenbach (1983) found that when any one of four modifier locus regions was introgressed onto the \textit{D. persimilis} \textit{Sex-ratio} chromosome, they lost the \textit{Sex-ratio} phenotype. These data suggest that as many as four genes could control the \textit{Sex-ratio} phenotype. The caveat here is that only the \textit{Sex-ratio} chromosome of \textit{D. persimilis} was tested and not that of \textit{D. pseudoobscura}. At this time, we cannot say whether other regions of XR in \textit{D. pseudoobscura} have similar reductions in nucleotide heterozygosity as \textit{Est-5}. Our study of XR was limited because we considered only two genes in our analyses, \textit{Est-5} and \textit{HspB3}. Examination of other genes within each of the nonoverlapping inversions is needed to better understand the genetics of the \textit{Sex-ratio} tract and how the inversion system alters the pattern and organization of genetic diversity.

This article benefited from discussions with Wyatt W. Anderson, Charles F. Aquadro, A. Bernardo Carvalho, Andrew G. Clark, Stephane Hourdez, Brian P. Lazzaro, Erin R. McMullin, Mohamed A. Noor, H. Allen Orr, Sheila A. Plock, Chung-j Wu, and Soojin Yi. Two anonymous reviewers significantly improved the quality of the article with their extremely helpful comments. We especially thank Robert K. Selander for the use of his ABI 373 automated sequencer that was purchased with funds from the National Institutes of Health. This work was supported by a National Science Foundation grant to S.W.S. (DEB 9726825).

**LITERATURE CITED**


Andolfatto, P., and M. Kreitman, 2000 Molecular variation at the \textit{In(2L)it} proximal breakpoint site in natural populations of \textit{Drosophila melanogaster} and \textit{D. simulans}. Genetics 154: 1681-1691.

Andolfatto, P., J. D. Wall and M. Kreitman, 1999 Unusual haplotype structure at the proximal breakpoint of \textit{In(2L)it} in a natural population of \textit{Drosophila melanogaster}. Genetics 153: 1297-1311.


Sturtevant, A. H., and G. W. Beadle, 1936 The relations of inver-
sions in the X chromosome of *Drosophila melanogaster* to crossing over and disjunction. Genetics **21**: 544–604.


Wang, R. L., and J. Hey, 1996  The speciation history of *Drosophila pseudoobscura* and close relatives: inferences from DNA sequence variation at the period locus. Genetics **144**: 1113–1126.


Wright, S., 1931  Evolution in mendelian populations. Genetics **16**: 97–159.


Communicating editor: J. Hey