

A Survey of Chromosomal and Nucleotide Sequence Variation in *Drosophila miranda*

Soojin Yi,^{*,1} Doris Bachtrog[†] and Brian Charlesworth[†]

^{*}Department of Ecology and Evolution, University of Chicago, Chicago, Illinois 60637-1573 and [†]Institute of Cell, Animal and Population Biology, University of Edinburgh, King's Buildings, Edinburgh EH9 3JT, United Kingdom

Manuscript received December 31, 2002

Accepted for publication March 5, 2003

ABSTRACT

There have recently been several studies of the evolution of *Y* chromosome degeneration and dosage compensation using the neo-sex chromosomes of *Drosophila miranda* as a model system. To understand these evolutionary processes more fully, it is necessary to document the general pattern of genetic variation in this species. Here we report a survey of chromosomal variation, as well as polymorphism and divergence data, for 12 nuclear genes of *D. miranda*. These genes exhibit varying levels of DNA sequence polymorphism. Compared to its well-studied sibling species *D. pseudoobscura*, *D. miranda* has much less nucleotide sequence variation, and the effective population size of this species is inferred to be several-fold lower. Nevertheless, it harbors a few inversion polymorphisms, one of which involves the neo-*X* chromosome. There is no convincing evidence for a recent population expansion in *D. miranda*, in contrast to *D. pseudoobscura*. The pattern of population subdivision previously observed for the *X*-linked gene *period* is not seen for the other loci, suggesting that there is no general population subdivision in *D. miranda*. However, data on an additional region of *period* confirm population subdivision for this gene, suggesting that local selection is operating at or near *period* to promote differentiation between populations.

DROSOPHILA miranda is a sibling species of *D. pseudoobscura* and *D. persimilis*, first described by DOBZHANSKY (1935). Except for a few slight differences observed in the laboratory, *D. miranda* is phenotypically indistinguishable from its sibling species (DOBZHANSKY 1935). Nevertheless, it is almost completely reproductively isolated from them (DOBZHANSKY and TAN 1936; DOBZHANSKY 1937). In addition, *D. miranda* possesses an unusual karyotype. DOBZHANSKY (1935) was puzzled by the fact that males and females of *D. miranda* each had a different number of chromosomes at meiosis. Later it was shown that the autosome that is monosomic in females, designated X_2 (or the neo-*X*), corresponds to chromosome 3 in *D. pseudoobscura* and *D. persimilis* (DOBZHANSKY and TAN 1936; MACKNIGHT 1939). The missing homolog of X_2 was found to have fused to the *Y* chromosome and is now termed the neo-*Y* chromosome (MACKNIGHT 1939; STEINEMANN and STEINEMANN 1998).

Since then, a series of investigations have shown that the neo-sex chromosomes of *D. miranda* represent a remarkable intermediate stage in the evolution of sex chromosomes. More specifically, the neo-*Y* chromosome of *D. miranda* is partially degenerate, and parts of the

counterpart neo-*X* chromosome have evolved dosage compensation (STEINEMANN and STEINEMANN 1998). Examination of the pattern of molecular evolution of neo-sex-linked genes in *D. miranda* provides a means for investigating the early phases of sex chromosome evolution. It has been proposed that *Y* chromosomes degenerate as a result of the greatly reduced effective population size (N_e) of the nonrecombining *Y* chromosome, compared with that of its freely recombining counterpart, the *X* chromosome (CHARLESWORTH 1996; CHARLESWORTH and CHARLESWORTH 2000). Polymorphism data show that N_e for the neo-*Y* is ~30-fold less than that for the neo-*X* (BACHTROG and CHARLESWORTH 2000, 2002; YI and CHARLESWORTH 2000). On the other hand, data on variation in *D. miranda* from genes other than neo-sex-linked loci are scarce, and little is known about the population biology of this species. Anecdotal evidence suggests that it was relatively abundant just a few decades ago in northern California and in the Canadian province of British Columbia (J. COYNE and A. BECKENBACH, personal communications). In recent years, however, many attempts to collect this species from its natural habitat have failed, and currently the last time *D. miranda* was caught in the wild was in 1997 in Mather, California (M. NOOR, personal communication).

Some authors have suggested that there is a substantial amount of genetic heterogeneity between populations of *D. miranda*. DOBZHANSKY and KOLLER (1938) reported some incipient reproductive isolation between strains collected from two different locations. GODDARD

Sequence data from this article have been deposited with the EMBL/GenBank Data Libraries under the accession nos. AY238758–AY238879.

¹Corresponding author: Department of Ecology and Evolution, University of Chicago, 1101 E. 57th St., Chicago, IL 60637.
E-mail: soojinyi@midway.uchicago.edu

et al. (1990) reported that the DNA composition within the species of *D. miranda*, as measured by DNA-DNA hybridization, varied considerably. However, no subsequent study has followed up these observations. Recently, a survey of nucleotide polymorphism at the X-linked locus *period* (*per*) revealed an unusual pattern of sequence diversity, suggesting the existence of population subdivision (YI and CHARLESWORTH 2000). This is the only study of nucleotide variation to show such a pattern so far.

Another possible type of genetic variation in *D. miranda* is chromosomal variation. Both of its sibling species, *D. pseudoobscura* and *D. persimilis*, have numerous inversion polymorphisms, which are particularly abundant on chromosome 3 (POWELL 1992, 1997). The extent of inversion polymorphism in *D. miranda*, however, has not been systematically documented, even though there are references to such polymorphism in the literature (DOBZHANSKY and TAN 1936; MACKNIGHT 1939; DOBZHANSKY and POWELL 1975).

It is important to characterize the extent of genetic and chromosomal variation in *D. miranda* and to compare it with its sibling species. This information is essential for inferring the selective forces underlying the evolution of the neo-sex chromosomes in this species, since it provides a benchmark against which the unusual patterns of DNA sequence polymorphism and evolution on the neo-sex chromosomes (BACHTROG and CHARLESWORTH 2000, 2002; YI and CHARLESWORTH 2000) can be compared. We have accordingly studied several aspects of the population genetics of *D. miranda*. First, we have investigated the amount of chromosomal variation in *D. miranda*. This is the first such documentation from this species and is important for our purposes, since inversions can distort the patterns of neutral DNA sequence variability (ANDOLFATTO *et al.* 2001). Second, we have surveyed the amount of genetic diversity at 12 nuclear genes and their divergence from *D. pseudoobscura*. This provides estimates of the typical extent of neutral variation and the effective population size of this species, which are critical parameters for inferring the operation of various evolutionary forces. Third, using nucleotide sequence data from our additional loci, we have asked if the pattern of population subdivision observed at the *per* locus (YI and CHARLESWORTH 2000) is a general phenomenon in this species. Such subdivision would complicate the interpretation of standard tests for departure from neutrality, which assume panmixia. Finally, we have used the pattern of nucleotide diversity from many loci to examine the population history of this species, *i.e.*, whether there has been an episode of population expansion, as proposed for its sibling species *D. pseudoobscura* (KOVACEVIC and SCHAEFFER 2000; MACHADO *et al.* 2002). Again, such an effect would complicate tests of neutrality.

MATERIALS AND METHODS

Strains used and genomic DNA preparation for polymorphism analyses: The strains of *D. miranda* that were used in YI and CHARLESWORTH (2000) were also used in this work. For all genes except *Adh*, genomic DNA (gDNA) was extracted from a single male fly from each strain of *D. miranda* by means of the Puregene gDNA extraction kit (Gentra, Research Triangle Park, NC). Genomic DNA from a strain of *D. pseudoobscura* from Mather, California, used by YI and CHARLESWORTH (2000), was also extracted from single male flies using the same method and used to obtain outgroup sequences for *cyp1*, *elav*, *Gapdh2*, *sc*, and *sesB* (gene nomenclature according to FlyBase; see below).

The extracted gDNA from each strain was rehydrated in 50 μ l TE buffer and then electrophoresed in 1% agarose gels for the estimation of quality and quantity. For one 25- μ l PCR reaction, <20 ng of gDNA sufficed for successful amplification of the final product. Among the 12 loci newly amplified in this study, 10 are expected to be on the X chromosome of the obscura group of *Drosophila* (see below). Therefore, except for the 2 autosomal loci *Adh* and *sry- α* , use of a single male fly should ensure homozygosity of the sequencing templates.

Genomic DNA used for the amplification of *Adh* was obtained in the following way, which is similar to the method used by McALLISTER and CHARLESWORTH (1999) for the study of *Adh* in *D. americana*. Interspecific crosses between a single *D. pseudoobscura* male and a female from *D. miranda* were performed. This was done separately for each of the 12 strains of *D. miranda*. Single progeny from these crosses should contain one copy of a *D. pseudoobscura Adh* allele and one copy of a *D. miranda Adh* allele. Genomic DNA was extracted from single progeny from each cross according to the same protocol as above. The *Adh* allele specific for *D. miranda* could thus be amplified by using *D. miranda* specific primers. For the amplification of *sry- α* , the PCR product from a single male of *D. miranda* was used for sequencing. If we found any potentially ambiguous sites from a single PCR product, we chose gDNA from another single male fly to repeat the process until we were certain of the homozygosity of the sequences.

Polytene chromosome preparation and inversion analyses: Larvae were reared at 18° in low-density banana medium cultures (YI and CHARLESWORTH 2000). Larvae were cleaned in Shen's solution (NICOLETTO 1959), and salivary glands of third instar larvae were dissected out in 45% acetic acid. The glands were transferred to a drop of 1 part concentrated lactic acid:2 parts water:3 parts glacial acetic acid on a siliconized coverslip and fixed for 5 min. A drop of a solution of Gurr's orcein (Bio/medical Specialties, Santa Monica, CA), prepared according to NICOLETTO (1959), was added, and the glands were stained for 2 min. A siliconized slide was gently lowered onto the coverslip and moved from side to side, to spread the glands. The slide was then examined at $\times 100$; if the nuclei remained clustered and intact, the coverslip was tapped with the end of a mounted needle to burst them. The slide was then pressed gently between the folds of a paper towel, and the edges of the coverslip were sealed with nail varnish. Slides were examined with brightfield illumination at $\times 400$ or $\times 630$ (oil immersion) with a Zeiss Standard microscope.

To identify polymorphic inversions, one line (MSH 38; see YI and CHARLESWORTH 2000) was arbitrarily selected as a tester strain. Crosses of females from each of the other 11 strains to this tester were made, and a minimum of one female larva per strain was scored for the presence of inversion loops (means of 1.5 female and 0.67 male larvae per line were scored). The breakpoints of inversions were established with reference to the photographic and drawn polytene chromo-

some maps of DAS *et al.* (1982); the standard arrangements were taken to be the band orders represented in their Figure 2. Once the inversions had been identified (see RESULTS), further characterization of the lines was carried out by examining squashes of larvae obtained directly from the lines.

Primers, PCR amplification, and sequencing for polymorphism analyses: We report new DNA sequence polymorphism data at 12 loci from *D. miranda*. We chose 10 genes that were expected to be located on the left arm of the X chromosome (XL), which corresponds to the ancestral X chromosome of the genus *Drosophila* (Muller's element A; POWELL 1997) and the right arm of the X chromosome (XR). These are: *cyclophylin* (*cyp1*), *elav*, an intergenic region between *esterase 5C* and *esterase 5B* (*Est-5*; BABCOCK and ANDERSON 1996), *Gapdh2*, a segment of *period* (*per-new*; see below), *runt* (*run*), *scute* (*sc*), *ADP/ATP translocase* (*sesB*), *sisterless A* (*sisA*), and *swallow* (*swa*). This selection of genes partly depended upon the availability of orthologous sequences from other species. In addition, we also chose two autosomal loci, *Alcohol dehydrogenase* (*Adh*) and *serendipity alpha* (*sry- α*). The *Adh* region sequenced includes the intergenic region between *Adh* and *Adh-related* (*Adhr*) loci and also the first exon and intron of the *Adhr* locus. We treat these two loci as a single unit throughout the analyses, because they are amplified as a single consecutive segment. The structure of each locus and the positions of the sequenced regions are depicted in Figure 1.

PCR primers were designed for each gene using conserved sequences inferred from the alignments of available orthologous sequences (Table 1). All sequenced regions except for *Est-5* include coding regions. Usually one 25- μ l PCR reaction was enough to provide templates for the subsequent sequencing reactions. For 7 of the 10 X-linked genes (excluding *Est-5*, *run*, and *swa* loci), the same male genomic DNA was used to generate the polymorphism data. We used Taq polymerase from Roche Molecular and Biochemicals (Indianapolis) for PCR and the BIG Dye-termination cycle sequencing kit from Perkin Elmer (Foster City, CA) for sequencing reactions. All sequences were run on an ABI 377 sequencer for both strands, using overlapping internal primers when necessary. All sequences are stored under GenBank accession nos. AY238758–AY238879.

Localization of gene probes to *D. miranda* chromosomes: PCR products from genomic DNA were extracted from 1% agarose gel using the QIAquick gel extraction kit (QIAGEN, Chatsworth, CA) and then labeled with Biotin High Prime (Roche Molecular Biochemicals) for use as a probe. Salivary glands from third instar larvae were dissected and hybridized with each probe. The hybridization procedure largely followed the protocol of MASIDE *et al.* (2001). Sites of hybridization were detected by staining with diaminobenzidine and peroxidase (Vector Laboratories, Burlingame, CA). Slides were examined under a $\times 40$ phase-contrast objective and stained with 5% pH 7.1 Giemsa solution (Gurr) when chromosome bands were faint or when there was a lot of background. The banding patterns of polytene chromosomes of *D. miranda* reported by DAS *et al.* (1982) were used as the reference point when determining the sites of hybridization.

Polymorphism and divergence data analyses: DNA sequences were aligned using the Clustal option in the MegAlign program (part of the DNA STAR package). The highly variable region from the 3' region of the *sisA* gene was manually aligned using the interface of the SeqEd program, minimizing the number of insertion/deletions. Further editing was performed as appropriate for various analyses. For most of the within-population variation studies, the SITES program (HEY and WAKELEY 1997) and DnaSP program v. 3.50 (ROZAS and ROZAS 1999) were used. A coalescent simulation to investigate

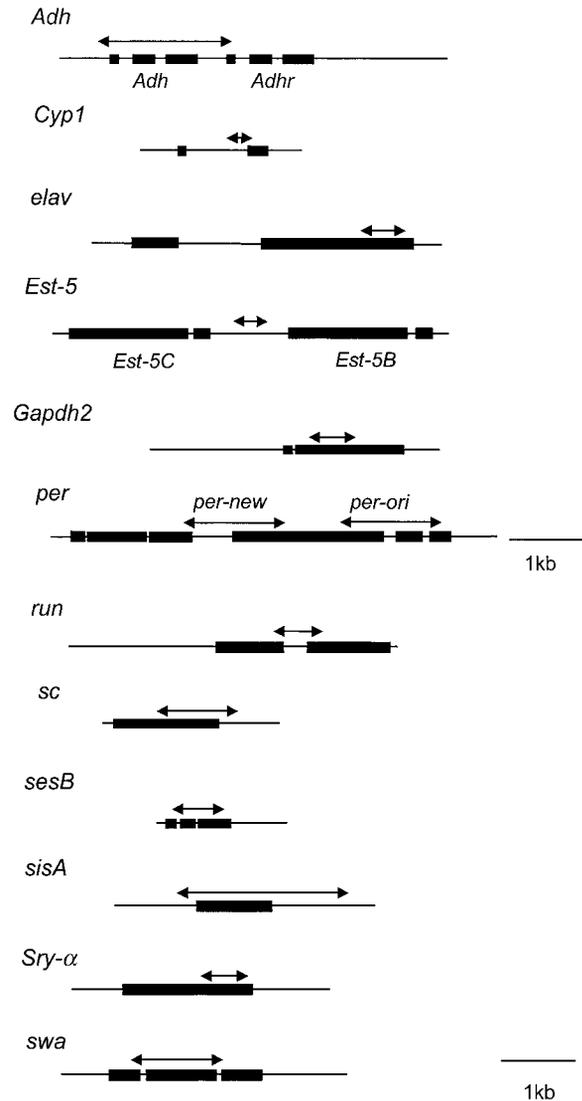


FIGURE 1.—Structure of gene segments surveyed in this study. Exons are shown as solid boxes, drawn to approximate scales. The double-headed arrows indicate the segments of DNA sequenced for each locus.

the variance of F_{ST} measures was obtained from Molly Przeworski and Eli Stahl. Multilocus Hudson-Kreitman-Aguadé (HKA) tests were performed using J. Hey's program using coalescent simulations (MACHADO *et al.* 2002).

RESULTS

Locations of genes surveyed in this study: The chromosomal locations of the 12 genes whose polymorphism pattern have been reported here are shown in Figure 2. All genes were found on the chromosomes expected from the conserved chromosome arm homologies within the genus *Drosophila* (Table 9.4, p. 307 of POWELL 1997). All the genes that are X linked in *D. melanogaster* are on XL in *D. miranda*. *Est-5* (on chromosome arm 3L of *D. melanogaster*) was found to be on XR, as

TABLE 1
Primer sequences used for PCR amplification

Gene	Primer	Sequence
<i>Adh</i>	adhmr-f	AGT TCC AGT TGT AAA AAT GCG
	adhmr-r	ATG AAG CAC ATT CCT TGG TTT
<i>cyp1</i>	cyp-f	ATG AGT ACT TTG CCT AGG
	cyp-r	GAT CTT CTT GGA GGT CTT
<i>elav</i>	elav-f	GGC AAC GAT ACG CAG AC
	elav-r	TTG TTG GTC TTG AAG CTG A
<i>Est-5</i>	est5-f	CGA TAA GTC GAG CCT CTC TCT ATG
	est5-r	AAC CAG TCT CAG GGG GAT AGC TCT
<i>Gapdh2</i>	gapdh2-f	GCC TCC GTT GTT GCC GTC
	gapdh2-r	ATC GCT AAG GAA ATC AGT
<i>per-new</i>	per1719f	GTC GGC CAT CAT CGT GTG TT
	per2965r	GGT GTT GGC GTC CAC GTT GC
<i>run</i>	run-f	GTG GCC AAG TTC AAC GAT CT
	run-r	ATG ATG GTA GTC GGC AGG AC
<i>sc</i>	sc-f	GTG AAG CAG GTG AAC AAC A
	sc-r	CAC CTT CAT TAT AAT GCC
	sc-ps-r	GTG TGC AGT CCT CCT CGT C
<i>sesB</i>	sesB-f	GAT GCT ATT GGA TTC GTC
	sesB-r	CAC CGC CAG TGC CTC TGA
<i>sisA</i>	sisA-f	CTA ATC CCT TTG TTT ACA T
	sisA-r	ATG AAA TCA TTT GTA ACC TG
<i>sry-α</i>	sry-f	CTG GAG CAC CAC TTC AAC G
	sry-r	TTT ATT CTG AGA ATT TCC G
<i>swa</i>	swallow-f	GCC AAC AAG AGC GCC AAG AC
	swallow-r	ATG GCA GAG TCG TTG CTG CT

expected from the fact that, in the pseudoobscura and affinis subgroups of the subgenus *Sophophora* of *Drosophila*, a fusion between elements A and D created a new X chromosome arm, XR (p. 309 of POWELL 1997). *Adh*, which is on 2L in *D. melanogaster*, is found on the homologous chromosome, chromosome 4 of *D. miranda*. The location of *sry-α* also follows the expected chromosomal homologies (chromosome 2, homologous to 3R of *D. melanogaster*).

The chromosomal assignments all correspond to those in the closely related species, *D. pseudoobscura* (BABCOCK and ANDERSON 1996; SEGARRA *et al.* 1996; KOVACEVIC and SCHAEFFER 2000), where these have been determined. However, the physical locations of the genes within the chromosomes are likely to differ. One significant difference when compared to the genetic map of XL of *D. pseudoobscura* (KOVACEVIC and SCHAEFFER 2000) is that the positions of *sisA* and *run* are inverted in *D. miranda*. This probably corresponds to the major medial inversion distinguishing the XL chromosome arms of these two species (DOBZHANSKY and TAN 1936; ANDERSON *et al.* 1977).

Polymorphic inversions: Table 2 shows the lines for which inversion loops were detected in crosses with MSH38 (see MATERIALS AND METHODS). A total of four inversions were detected, on chromosome arms XR, X₂, and 2, respectively; no inversions were found on XL or chromosome 4 (Figure 2). The large inversion 2-2 was

found only in association with the included inversion 2-1. Further examination of the lines themselves showed that MSH38 is segregating for inversion XR-I; out of six female larvae examined, four were heterozygous for this inversion. In good preparations, it was possible to identify the inversions XR-I, X₂-I, and 2-2 in homo- or hemizygotes, using the orders of certain landmarks (prominent banding patterns and puffs). No lines other than MSH38 showed any signs of the presence of XR-I, so that it is apparently a low-frequency inversion. All six MSH38 female larvae examined for X₂-I were found to be homozygous for the standard arrangement. Similar examinations of the other strains suggests that this inversion is probably fixed in the seven lines that had been found to contain it in the test crosses with MSH38 (Table 2). This indicates that it has a frequency of ~50% in the *D. miranda* population. Inversion 2-1 was found to segregate in the cross of line 0101.3 with MSH38, but all five larvae examined from 0101.7 (from both the testcross and the line itself) carried the inversion. It is interesting to note that MACKNIGHT (1939) described an inversion that was almost certainly X₂-I; he found it in a stock that had been treated with X rays and inferred that it had been induced by this treatment. Line 0101.4 showed segregation for the compound inversion 2-1 + 2-2 in the testcrosses with MSH38; neither inversion was seen in any of the other lines. Both 2-1 and 2-1 + 2-2 are therefore low-frequency inversions.

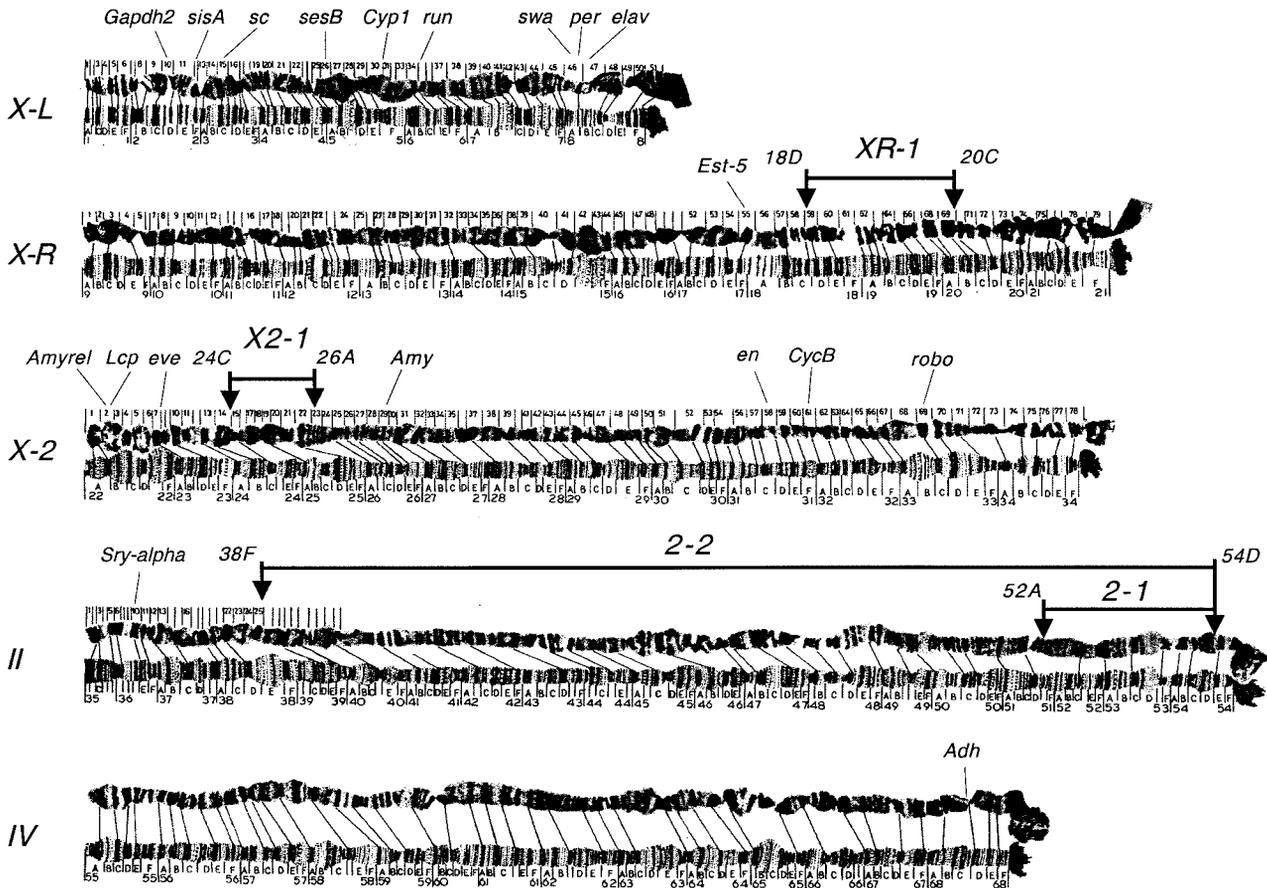


FIGURE 2.—Gene locations and the newly discovered inversions from this study. Drawn on the polytene chromosome map of *D. miranda* from Figure 2 of DAS *et al.* (1982). The proximal region for each chromosome arm is on the right.

Summary of data on DNA sequence polymorphisms, divergence, and functional constraints: The loci surveyed in this study showed a wide range of levels of genetic diversity at the DNA sequence level (Table 3; see supplementary material at <http://www.genetics.org/supplemental/> for full details of haplotypes). Both indel variants and single nucleotide polymorphisms (SNPs) were present. There are a total of 23 indel variants from seven loci, all from introns or untranscribed regions. Five loci, *elav*, *Gapdh2*, *sesB*, *sry-α*, and *swa*, did not show any indels. The 3' noncoding region of *sisA*

had a region that was highly variable for complex indel variants. This corresponds to the position 1091–1157 of the alignment. At the *run* locus, all the indel variants occurred in a region surrounding a core indel of 12 nucleotides. Note that the indels are determined using the most parsimonious alignments, so that the real numbers may slightly differ from our estimates.

We observed more than five times as many SNPs as indels. In three cases two or three adjacent SNPs were in complete linkage disequilibrium (two cases at the *run* locus, one at the *per-new* locus: see supplementary

TABLE 2
Polymorphic inversions detected in *D. miranda*

Chromosome inversions in test arm MSH38	Breakpoints of inversions	Lines showing inversion loops in crosses with MSH38
XL	—	—
XR	18D-20C (<i>XR-1</i>)	MA28, 0101.4
X ₂	24C-26A (<i>X₂-1</i>)	SP235, SP295, 0101.3, 0101.4, 0101.5, 0101.7, 0101.9
2	52A-54D/E (<i>2-1</i>) 38F-54D/E (<i>2-2</i>)	0101.3, 0101.4, 0101.7 0101.4
4	—	—

TABLE 3
Summary of polymorphism and divergence statistics

Gene	No. of haplotypes	Length ^a	Nucleotide diversity ^b			K(silent) (%) ^c	K _A /K _S ^d
			rep	syn	sil		
<i>Adh</i>	11	2042 (10)	0	4.07	2.94	3.39	0.207
<i>cyp1</i>	8	581 (11)	—	—	4.87	2.11	—
			—	—	6.26		
<i>elav</i>	1	543 (0)	0	0	0	2.97	0
<i>Est-5</i>	9	532 (13)	—	—	5.36	3.74	—
			—	—	8.06		
<i>Gapdh2</i>	1	638 (0)	0	0	0	4.23	0.046
<i>per-ori</i>	12	1480 (18)	3.58	5.73	5.18	5.90	0.225
<i>per-new</i>	8	1408 (13)	3.41	5.88	5.00	6.02	0.288
			0.34	3.03	3.76		
<i>run</i>	11	604 (19)	0.67	4.27	4.85	4.96	0.029
			0.66	12.62	15.16		
<i>sc</i>	9	994 (7)	1.31	14.91	16.66	1.70	0.176
			1.12	3.62	2.52		
<i>sesB</i>	4	869 (4)	1.29	4.16	3.44	4.16	0.151
			0	0.98	3.23		
<i>sisA</i>	12	1818 (35)	0	1.95	4.10	2.45	0.368
			4.01	8.32	6.39		
<i>sry-α</i>	8	495 (7)	4.25	6.49	7.22	3.65	0
			3.52	4.96	4.96		
<i>sza</i>	3	1109 (3)	4.48	5.56	5.56	4.44	0.271
			0.45	1.81	1.11		
			0.89	1.46	0.90		

^a The actual number of sites compared for polymorphism analyses, excluding sites with alignment gaps or missing data. The number of segregating sites is shown in parentheses.

^b Two types of nucleotide diversity measures are presented. The upper number in each row is the pairwise nucleotide diversity, π (NEI 1987); the lower number is θ_w , the measure of nucleotide site variability based on the number of segregating sites (WATTERSON 1975). All values are shown after multiplication by 10^3 .

^c Per-site silent (synonymous + noncoding sites) site divergence from *D. pseudoobscura* with Jukes and Cantor correction for multiple hits. Except for *Gapdh2* and *elav*, all silent site diversities are computed as the average of all comparisons between sequences.

^d The ratio of nonsynonymous to synonymous site divergence from *D. pseudoobscura* estimated using the method of NEI and GOJOBORI (1986).

^e Data from YI and CHARLESWORTH (2000).

material at <http://www.genetics.org/supplemental/>). The mutation events that created these clusters of SNPs are unlikely to be independent events. This is inconsistent with the infinite sites model (KIMURA 1971), which is the basis of the methods used to analyze the pattern of SNPs. Therefore, in these three cases we counted each cluster as one SNP. This is reflected in all the SNP analyses described below.

Summary statistics on SNPs at each locus are presented in Table 3, together with the estimated numbers of nucleotide substitutions per site from *D. pseudoobscura* for silent sites (K_S). Previous data from the *per* locus (YI and CHARLESWORTH 2000) are also compiled in Table 3 for the purpose of comparison. These are denoted as “*per-ori*.” Newly obtained results from another region of the *per* locus (see MATERIALS AND METHODS) are shown

as “*per-new*.” A simple average of the silent site pairwise nucleotide diversity from Table 3 is 0.43%. The mean K_S (weighted by the numbers of synonymous sites per gene) is 3.4%; a contingency χ^2 test shows highly significant variation in K_S among loci ($\chi^2 = 36.4$, $P < 0.0003$). The variance in observed K_S values is 1.75×10^{-4} , ~ 1.78 times that expected purely on the basis of sampling variation with a constant underlying K_S .

We observed reasonable correlations between nucleotide site diversities and the presumed functional roles of the corresponding region. For example, the 3' non-coding region of *Adh* is a short intergenic region between the *Adh* and the *Adhr* locus (see Figure 1), which may include several motifs related to the expression of both *Adh* and *Adhr* (MARFANY and GONZALES-DUARTE 1992) and/or sequences required for appropriate splic-

TABLE 4

The distribution of the Tajima's *D* statistic and the estimated recombination parameter

Gene	Total no. of sites	S^e (total)	D (total)	S^e (silent)	D (silent)	$\hat{\rho}/\text{site}^b$	$\hat{\rho}/\hat{\theta}^c$	95% C.I. of D assuming no recombination	95% C.I. of D assuming free recombination	Critical value of C (per site)
<i>Adh</i>	2042	10	0.99	10	0.99	0.121	51.05	[-1.74, 1.70]	[-1.10, 1.11]	—
<i>cyp1</i>	584	11	-0.92	11	-0.92	0.023	3.67	[-1.80, 1.68]	[-1.03, 1.03]	—
<i>Est-5</i>	534	13	-1.42	13	-1.42	0.030	3.72	[-1.73, 1.74]	[-1.10, 1.12]	$C > 0.022$
<i>per-ori</i>	1480	18	0.18	9	0.15	0.037	7.40	[-1.83, 1.66]	[-0.85, 0.93]	—
<i>per-new</i>	1408	13	-1.04	12	-0.94	0.012	2.47	[-1.74, 1.72]	[-0.99, 1.01]	—
<i>run</i>	641	18	-0.39	17	-0.39	0.422	25.33	[-1.81, 1.66]	[-0.86, 0.88]	—
<i>sc</i>	994	7	-0.90	5	-0.99	>1	>291	[-1.83, 1.81]	[-1.29, 1.40]	—
<i>sesB</i>	869	4	-0.74	4	-0.75	0.019	4.63	[-1.75, 1.92]	[-1.38, 1.47]	—
<i>sisA</i>	1927	35	-0.52	29	-0.52	0.477	66.07	[-1.76, 1.72]	[-0.63, 0.63]	—
<i>sry-α</i>	495	7	-0.72	2	-0.32	0.273	48.75	[-1.94, 1.79]	[-1.25, 1.30]	—
<i>swa</i>	1109	3	-0.83	1	0.53	<10 ⁻³	<10 ⁻¹	[-1.75, 1.67]	[-1.16, 1.24]	—

^a The number of segregating sites.^b The estimate of $C = 4N_e r$ (HUDSON 2001).^c For silent sites only.

ing. This is consistent with its low level of variation (one variant from 282 sites surveyed). The sequenced 3' noncoding region of *sc* is highly conserved between *D. melanogaster* and *D. subobscura* and known to contain the polyadenylation signal and a conserved region with an unknown function, as well as putative *Sxl*-binding sites (BOTELLA *et al.* 1996). We found that the 3' region of *sc* harbors very little variation. Divergence from *D. pseudoobscura* is also very low ($K_S = 1.79\%$). The 5' flanking region of the *sisA* locus showed a much lower level of nucleotide diversity ($\pi = 0.07\%$) than the 3' region ($\pi = 0.76\%$). When compared with the homolog from *D. pseudoobscura*, the 5' region was much more conserved than the 3' region, indicating either that the two regions have very different mutation rates or that the 5' region is under strong selective constraint in both *D. pseudoobscura* and *D. miranda*. The 5' region of the *sisA* gene surveyed in this study encodes two putative numerator sequences for the X chromosome/autosome counting mechanism (ERICKSON and CLINE 1998) as well as a TATA box. In contrast to this, the 3' region of this gene is not known to include any highly conserved regions when compared with those of *D. melanogaster*, *D. pseudoobscura*, and *D. virilis* (ERICKSON and CLINE 1998).

The most polymorphic locus was *run*, with a value of silent site pairwise diversity (π) of 1.52%, more than twice as much as the next most variable locus, *sisA*. This is mainly due to its very polymorphic intron, which harbored 14 polymorphic sites within 261 bp. The pattern of SNPs within the introns of the *run* locus shows tight associations between many closely linked variants (see supplementary material at <http://www.genetics.org/supplemental/>), although the level of recombination estimated from the SNPs is fairly large (Table 4). Fu's *F*-statistic (Fu 1997) suggests that the number of haplotypes is smaller than expected under the neutral

model ($F = -4.28$, $P < 0.05$). It is possible that a site under balancing selection is closely linked to these variants. Interestingly, the synonymous sites at the *run* locus showed a much lower level of divergence from *D. pseudoobscura* than that of the intron sites ($K_S = 1.13\%$, compared to K_I for the intron = 6.32%). The numbers of conserved and substituted synonymous and intronic sites show a significantly heterogeneous pattern (two-tailed Fisher's exact test, $P < 0.05$).

Patterns of SNP variation, neutrality tests, and the effect of recombination: Using several different methods, we tested whether the observed patterns of SNP variation and divergence from all loci are compatible with the simple neutral model. First, we performed the HKA test (HUDSON *et al.* 1987). This was done for silent sites only, to avoid the effects of purifying selection on replacement sites (see the previous section). We excluded the *per-ori* and *per-new* region from the following analyses, to avoid the possible confounding effect of local selection suggested by their high F_{ST} values (see below). We used the multilocus form of the HKA test, using 10,000 coalescent simulations to test the significance of the deviation (see MACHADO *et al.* 2002).

When all the loci in Table 3 (excluding *per-ori* and *per-new*) were considered, the sum of the deviations was 17.03 (d.f. = 10, $P < 0.03$). The largest deviation came from the lower-than-expected number of segregating sites for the *swa* locus (1 observed and 5.56 expected) and to a lesser degree the higher-than-expected divergence (15.04 observed and 11.65 expected). When we excluded the *swa* locus from the multilocus HKA test, the sum of the deviations decreased to 12.96 and was not significant (d.f. = 9, $P < 0.09$). This strongly suggests that the observed deviation from the neutral model was caused mostly by the low polymorphism relative to divergence of the *swa* locus. The same conclusion was

reached using a likelihood-ratio alternative to the HKA test (S. WRIGHT and B. CHARLESWORTH, unpublished results).

The McDonald-Kreitman test (MCDONALD and KREITMAN 1991) examines the pattern of polymorphism and divergence between two different types of sites along the same sequence. Under the neutral model, the ratio of replacement to silent variation should be similar to the ratio of replacement to silent divergence. Only the combined data from *per-ori* and *per-new* showed a significant deviation (see DISCUSSION). However, the numbers of polymorphic sites within *D. miranda* and the fixed differences between the *D. miranda* and *D. pseudoobscura* in the exons are small in all the comparisons, due to both the small effective population size of *D. miranda* (see DISCUSSION) and the short divergence time between these two species. The power of this test is thus likely to be low.

Another class of neutrality tests focuses on the frequency distribution of mutations. We will describe the results from the Tajima's *D* statistic (TAJIMA 1989). Other tests yielded similar conclusions (results not shown). A significantly negative value of Tajima's *D* suggests that more low-frequency variants than expected are under the equilibrium neutral model of molecular variation. A significantly positive Tajima's *D* indicates a higher-than-expected fraction of high-frequency variants.

Table 4 shows the Tajima's *D* statistics for each locus for all types of sites combined and for silent sites alone. *Est-5* had the most negatively skewed value, with 10 singletons out of 13 SNPs. In contrast, *Adh* had mostly high-frequency variants, giving the most positive Tajima's *D*. We assessed whether the observed frequency distributions of SNPs were compatible with neutral expectation by examining the distributions of the Tajima's *D* statistics on the null hypothesis of neutrality and equilibrium. The variance of Tajima's *D* decreases with increasing recombination level among the sites in the sample (WALL 1999). The significance of Tajima's *D* should therefore be assessed using the recombination rates for the sampled genes. There are, however, no independent estimates of recombination rates for *D. miranda*. We employed the information from the SNP data to infer the recombination parameters for each locus (Table 4). We used ρ , obtained by the composite likelihood method of HUDSON (2001) as an estimate of the population recombination rate $C = 4N_e r$ (where N_e is the effective population size; r is the evolutionarily effective recombination rate, which is one-half or two-thirds the recombination rate in females for autosomal and X-linked loci, respectively).

In Table 4, we show the 95% confidence intervals (C.I.s) of Tajima's *D* for the two most extreme cases, no recombination *vs.* free recombination. The estimated population recombination estimate per nucleotide site for each locus and its value relative to the estimate of θ from the number of segregating sites, $\hat{\theta}_w$ (WATTERSON

1975), are also shown. Under the neutral model, this ratio reflects the ratio of recombination rate to mutation. Tajima's *D* for all but one locus, *Est-5*, in Table 4 falls within the observed 95% C.I., assuming free recombination. To investigate *Est-5* further, we performed coalescent simulations to find the critical values of C for which the model cannot explain the observed pattern. The observed negative Tajima's *D* for *Est-5* falls outside the lower bound of the 95% C.I. when the recombination parameter becomes >12 for the whole gene. The estimated recombination parameter for the gene was 16.0. The probability of a Tajima's *D* less than -1.42 , conditioned upon the estimated level of recombination, was 0.013. Other test statistics for the frequency distribution (FU and LI 1993) also showed a significant departure from neutrality for *Est-5* when the estimated recombination rate was used (Fu and Li's $D^* = -1.68$, $P < 0.008$; Fu and Li's $F^* = -1.83$, $P < 0.015$). However, this is not significant after a sequential Bonferroni correction (RICE 1989), so that further data are required to determine whether there really is a departure from neutrality at *Est-5*.

In the sibling species to *D. miranda*, *D. pseudoobscura*, and *D. persimilis*, most of the genes surveyed exhibit a negative Tajima's *D*, which has been attributed to a recent population expansion in these species (MACHADO *et al.* 2002). We also tested, using coalescent simulations, whether the prevalence of negative Tajima's *D* values in *D. miranda* provides any evidence for such an expansion. When we used only X-linked genes, the observed mean Tajima's *D* (-0.64) was significantly lower than the simulated mean value (-0.58 , $P = 0.03$). However, this result was not robust: when we included the two autosomal loci, the deviation from neutrality was not significant ($P < 0.1$). There is thus no firm evidence for a recent genome-wide population expansion in *D. miranda*.

DISCUSSION

Inversion polymorphism in *D. miranda*: The data reported here apparently represent the first systematic survey of chromosomal polymorphism in *D. miranda*. However, the dearth of wild-caught strains of *D. miranda* has limited the scope of this survey. The pattern of inversion polymorphism is different from that of the sibling species of *D. miranda*, *D. pseudoobscura*, and *D. persimilis*, in which autosomal inversion polymorphisms are largely confined to the highly polymorphic chromosome 3 (POWELL 1992). This is homologous to X_2 of *D. miranda*, which we have found to carry only a single small distal inversion at intermediate frequency. Owing to the difficulty in aligning the banding patterns of *D. miranda* and its sibling species for this arm (DOBZHANSKY and TAN 1936), it is not possible to be sure whether there is any homology between the *D. miranda* arrangements on X_2 and the arrangements in its siblings.

In both *D. pseudoobscura* and *D. persimilis*, chromosome

arm XR shows inversion polymorphisms associated with the SR meiotic drive system; we have found only a single medium-size, proximal inversion at low frequency on this chromosome arm in *D. miranda*. The standard arrangement of XR in *D. miranda*, however, shows high homology to the standard XR arrangement in *D. pseudoobscura*, except for a distal inversion (DOBZHANSKY and TAN 1936), as we have confirmed in our own preparations of hybrids between *D. miranda* and *D. pseudoobscura*. Together with the fact that XL of *D. miranda* is more similar to XL of *D. pseudoobscura* than is XL of *D. persimilis*, this suggests that the ancestor of *D. persimilis* split from that of *D. pseudoobscura* after their common ancestor split from the ancestor of *D. miranda*.

In contrast to the situation for *D. pseudoobscura* and *D. persimilis*, *D. miranda* has inversion polymorphisms on chromosome 2. As is often found in the obscura group (KRIMBAS and POWELL 1992), the two inversions we have found (Table 2) share a common breakpoint at 54D/E, and the larger inversion includes the smaller one. The chromosome 2 inversions provide yet another example of inversions that are specific to a given species (KRIMBAS and POWELL 1992), indicating that the lifetime of inversion polymorphism is usually less than that of a species.

The presence of these inversion polymorphisms raises the question of what effect they might have in distorting the pattern of nucleotide sequence polymorphism at loci that are either inside the region covered by the breakpoints or sufficiently close to the breakpoints to experience suppressed recombination in heterokaryotypes (ANDOLFATTO *et al.* 2001). This is a serious issue only for chromosome X₂, the neo-X chromosome, since there are either no inversions or only low-frequency inversions on the other chromosomes. The locations of genes reported by STEINEMANN and STEINEMANN (1990) and BACHTROG and CHARLESWORTH (2002) indicate that, among X₂ genes for which DNA sequence polymorphism data have so far been obtained, *eve* is the closest to the breakpoints of X₂-I. Its position at 23A is just over one polytene chromosome division from the distal breakpoint of the inversion. In *D. pseudoobscura*, heterozygosity for a single inversion can reduce the frequency of recombination in the distal section of the chromosome by 50-fold but has a much smaller effect on recombination in the proximal section of the chromosome (DOBZHANSKY and EPLING 1948). There could thus be a very low frequency of recombination between *eve* and the distal breakpoint of the inversion, but reduced recombination is less likely for the other loci.

Indeed, some effects of the X₂-I inversion on patterns of nucleotide sequence polymorphism at *eve* can be detected. Silent site diversity π at *eve* is about twice as high for individuals carrying the inverted chromosome compared to the standard arrangement ($\pi = 1.39\%$ vs. $\pi = 0.68\%$). Of the 27 segregating sites at *eve*, 1 is in significant linkage disequilibrium with the inversion by

Fisher's exact test, and 3 by a χ^2 test at the 5% significance level (none of these associations, however, remains significant after a sequential Bonferroni correction). Nevertheless, there are no fixed differences between the inverted and the standard class at *eve*, and the polymorphism data show evidence for recombination between the standard and the inverted chromosomes. Tests of population subdivision indicate some differentiation between the standard and inverted arrangement (significance levels were assessed using a permutation test as implemented in the DNAsp program). Hudson's S_{nn} statistic (HUDSON 2000), which estimates the proportion of nearest neighbors in sequence space that are found in the same population, indicates a highly significant differentiation between the two classes of chromosomes ($P < 0.01$). But other statistics, which measure the ratio of the mean pairwise differences within populations to the mean total pairwise differences, are not significant: for K_{ST} or K_{ST}^* (HUDSON 1992a), $P < 0.1$. No other loci studied so far are likely to be seriously affected by the inversion polymorphisms we have described.

The effective population size of *D. miranda* compared to other *Drosophila* species: The level of silent site diversity based upon the number of segregating sites ($\hat{\theta}_w$) is an unbiased estimator of $4N_e\mu$ under the infinite sites neutral model (WATTERSON 1975). If we assume that the mutation rate (μ) is constant across the genome, the effective population size (N_e) can be inferred from the $\hat{\theta}_w$ values, given an estimate of μ . This is of importance, since some models of the degeneration of the neo-Y chromosome are very sensitive to the value of N_e (GORDO and CHARLESWORTH 2001). The arithmetic average of $\hat{\theta}_w$ for silent sites for the set of X-linked genes in Table 3 is 0.51%. WRIGHT *et al.* (2003) have developed an alternative maximum-likelihood method for obtaining a multilocus estimate of θ , using Equation 12 of HUDSON (1990) for the probability of observing S segregating sites in a sample of n chromosomes, conditioned on θ . Using this method (which assumes the same value of θ for all loci), we obtained a maximum-likelihood estimate (MLE) of θ for silent sites of 0.53% for the 11 X-linked genes that we have investigated (95% C.I. 0.38–0.74%); this agrees closely with the other estimate.

How, then, does the N_e of *D. miranda* compare to that for other *Drosophila* species? If we assume a constant rate of neutral mutation across the genus (although this is rather questionable), the relative values of $\hat{\theta}_w$ for silent sites from different species indicate the corresponding relative N_e values. Table 5 summarizes the results for several *Drosophila* species. Clearly, the overall level of genetic variation in *D. miranda* is much lower than that of its sibling species *D. pseudoobscura*. On the basis of relative $\hat{\theta}_w$ values, we conclude that there is a four- to sixfold reduction of N_e in *D. miranda* compared to *D. pseudoobscura*. The difficulty in collecting wild *D. miranda* (see Introduction) is consistent with a low population

TABLE 5
Comparison of estimates of θ_w for silent sites from different *Drosophila* species

	<i>D. miranda</i> ^a	<i>D. pseudoobscura</i> ^b	<i>D. melanogaster</i> ^c		<i>D. simulans</i> ^c	
			African	Non-African	African	Non-African
From synonymous sites						
Autosomal	5.1 (2)	30.1 (7)	16.1 (16)	15.5 (16)	32.9 (7)	25.0 (7)
X chromosomal	4.2 (8)	27.0 (1)	25.7 (5)	10.2 (5)	29.8 (5)	19.8 (5)
From noncoding sites						
Autosomal	2.7 (1)	15.6 (16)				
X chromosomal	6.2 (9)	18.9 (9)				

All numbers are shown after multiplication by 10^3 . The number of loci on which the estimation is based is shown in parentheses.

^a From current work.

^b From the following references: MORIYAMA and POWELL (1996), HAMBLIN and AQUADRO (1999), KOVACEVIC and SCHAEFFER (2000), and MACHADO *et al.* (2002).

^c From the compilation of ANDOLFATTO (2001).

size for *D. miranda*. Genetic diversity is also lower in *D. miranda* than in *D. melanogaster* and *D. simulans* (synonymous diversity in *D. pseudoobscura* is similar to that of African *D. simulans*; Table 5). When we compared the genetic diversity of X-linked genes, for which a fair amount of data is from *D. miranda*, N_e seems to be approximately one-half that of non-African *D. melanogaster*, while it is several-fold lower than that of *D. simulans*. Considering that N_e for *D. simulans* is estimated to be in the millions (WALL *et al.* 2002), N_e for *D. miranda* is likely to be in the vicinity of 1 million or less.

Interestingly, Table 5 shows little difference in variability levels between X-linked and autosomal loci in both *D. miranda* and *D. pseudoobscura*, as is also the case for African populations of *D. simulans* and *D. melanogaster* (ANDOLFATTO 2001). Application of the likelihood version of the HKA test mentioned above indicates that the X-linked loci in *D. miranda* depart significantly from the level of diversity expected on the basis of their having three-quarters the N_e of the autosomal loci, the value for a Poisson offspring number distribution (S. WRIGHT and B. CHARLESWORTH, unpublished results). These findings suggest that sexual selection may be inflating the value of N_e for X-linked loci over the Poisson expectation in all of these species (CHARLESWORTH 2001).

The effective population size of the neo-X chromosome and its implications for the evolution of dosage compensation: Interestingly, the MLE of θ for six neo-X-linked loci (YI and CHARLESWORTH 2000; BACHTROG and CHARLESWORTH 2002) is similar (MLE 0.55%, 95% C.I. 0.35–0.84%) to the above estimates of θ for X-linked loci. Microsatellite data also show very similar levels of variability on the X chromosome and the neo-X chromosome (BACHTROG and CHARLESWORTH 2000). The neo-X chromosome has evolved from an autosome homologous to chromosome 3 in the sibling species of *D. miranda*, following the fusion of the neo-Y chromosome to the true Y chromosome. The current level of

genetic diversity on this chromosome is dependent upon several factors, including the level of genetic diversity before its differentiation as a sex chromosome, the time since the formation of the neo-sex chromosomes, the number of “founder” neo-X chromosomes, and the intensity and nature of selection on the genes on the neo-X chromosome. The lack of a general reduction in variability on the neo-X chromosome compared with other chromosomes suggests that the last two factors have not had a significant effect.

This is particularly relevant to the fact that many loci on the neo-Y chromosome are degenerating (STEINEMANN and STEINEMANN 1998); in response, the neo-X chromosome has been shown to have evolved dosage compensation in some regions (BONE and KURODA 1996; MARÍN *et al.* 1996; STEINEMANN *et al.* 1996). The exact nature of the molecular mechanism of dosage compensation is not known but it is believed that *cis*-acting initiation sites on the X chromosome allow the spread of chromatin-modifying activities along the X (PARK and KURODA 2001). Thus, one might expect to see the signature of selective sweeps for *cis*-acting sites required for dosage compensation on the neo-X chromosome, *i.e.*, reduced variation compared to the X chromosome. However, the observed mean value of $\hat{\theta}_w$ on the neo-X chromosome is not smaller, nor is the variance of $\hat{\theta}_w$ among loci on the neo-X larger, than that on the X. This does not support multiple selective sweeps on the neo-X chromosome. Indeed, recent evidence suggests that dosage compensation in *Drosophila* does not evolve on a gene-by-gene basis, as previously believed, but rather involves relatively large blocks of genes (PARK and KURODA 2001). Only few *cis*-acting binding sites may thus be necessary to evolve chromosome-wide dosage compensation.

Population subdivision in *D. miranda* and the behavior of F_{ST} statistics: Data from a 1480-bp region from the *per* locus provided the first DNA sequence variation

TABLE 6
The degree of population subdivision estimated
from each locus

Gene	F_{ST}^a
<i>Adh</i>	-0.05
<i>cyp1</i>	0.14
<i>Est-5</i>	0.00
<i>per-ori</i>	0.37
<i>per-new</i>	0.33
<i>run</i>	0.17
<i>sc</i>	0.12
<i>sesB</i>	0.14
<i>sisA</i>	0.01
<i>sry-α</i>	0.08
<i>swa</i>	0.04

^a F_{ST} between British Columbia population and the rest (see DISCUSSION).

data for *D. miranda* (YI and CHARLESWORTH 2000). The authors noted a pattern that suggested the existence of population subdivision, namely a distinction between the British Columbia sample and the U.S. samples. The authors also noted that, as the PER protein functions as an essential component of the biological clock in many different *Drosophila* species and shows evidence of geographic variation in *D. melanogaster* that is probably due to selection (ROSATO and KYRIACOU 2001), the observed differentiation between alleles from two different geographic localities might correspond to adaptation to different light cycles. However, various tests did not detect a departure from neutrality (YI and CHARLESWORTH 2000). Here, we revisit this issue with more sequence data from other loci and additional data from the *per* locus.

First, to determine whether the observed differentiation between the British Columbian population and the other Northern American populations is a pattern common to all loci, we estimated the F_{ST} statistics for the difference between these samples for all the loci with available SNP data (using the method of HUDSON *et al.* 1992b), including data from the originally sequenced *per-ori* region and the newly sequenced *per-new* region. The results are shown in Table 6. The two *per* regions show by far the highest values of F_{ST} . If we use the concatenated SNP data set for the whole *per* sequence, $F_{ST} = 0.36$. Including this value, the arithmetic mean of the F_{ST} statistics from the 10 loci studied here is 0.10. If we exclude *per*, the mean F_{ST} is 0.07. A permutation test (HUDSON *et al.* 1992a) and the S_{nn} test (HUDSON 2000) did not yield any significant values for measures of population differentiation except for the two regions from the *per* locus ($P < 0.05$ for *per-ori* and *per-new*, $P < 0.01$ for the concatenated *per* sequences). When the samples were divided into within-British Columbia and within-U.S. populations, no significant differentiation was de-

tected (results not shown). Microsatellite loci from *D. miranda* do not show any pattern of population subdivision (BACHTROG and CHARLESWORTH 2000). We conclude that there is no genome-wide pattern of population subdivision in *D. miranda*.

This raises the question of whether the pattern of between-population differentiation for the *per* locus reflects the effects of local selection at *per* on sites linked to the targets of selection (CHARLESWORTH *et al.* 1997) or is simply a stochastic result of neutral evolution. We emphasize that the newly sequenced *per-new* region is 468 bp upstream of the original region, on the basis of the annotation of *per* in *D. pseudoobscura* (Figure 1), but shows a pattern that is very similar to that for the downstream sequence (see above). Any effect of local selection must therefore extend over a region of several kilobases, given the size of *per*. This is compatible with the predictions of models of the effects of local selection, if the intensity of selection is comparable in size with the amount of recombination over the region in question (CHARLESWORTH *et al.* 1997).

The *per* region seems to be experiencing a considerable amount of recombination. The scaled recombination parameters for the two regions differ by threefold (Table 4) when estimated by Hudson's composite-likelihood method (HUDSON 2001). If we take these differences literally, this suggests either a gradient of recombination frequencies increasing toward the 3' direction or a stronger selective force causing higher linkage disequilibrium in the 5' region of *per*. However, as the confidence intervals for this measure of recombination tend to be large (HUDSON 2001) and there are no independent data on recombination or selection at the *per* locus, these remain speculations.

For simplicity, we used the concatenated data set to obtain an approximate estimate of the recombination parameter for the whole region, which encompasses 2888 nucleotides, excluding gaps in the alignment [note that the composite-likelihood method (HUDSON 2001) does not assume contiguity in the sampled region]. When the whole region was used, the estimated scaled recombination rate, C , for the *per* region was 41.6 (0.014 per nucleotide site). If there is selection inducing linkage disequilibrium among polymorphic sites in this region, this will be an underestimate. The recombination parameter is thus likely to be at least 20–100.

To investigate the behavior of F_{ST} in relation to the *per* data, we first took the average F_{ST} value of 0.07 for loci excluding *per* (see above) and computed an estimator of $Nm = 3.22$ from the formula for F_{ST} for a pair of populations of size N exchanging genes at rate m (HUDSON *et al.* 1992b). This is a reasonable proxy for the scaled migration parameter for exchange between the British Columbia population and the other populations for *D. miranda*, given the limited amount of data. We then used coalescent simulations to generate samples under this two-population model, conditioning on the esti-

mated migration parameter, assuming sample sizes of 4 and 8 for the two populations, with 31 segregating sites in 2888 bp (HUDSON 2002). The 95% C.I. for F_{ST} from 1000 simulated genealogies, assuming no recombination, is $[-0.27, 0.48]$. The observed F_{ST} for the *per* region, 0.36, is within this interval. However, the distribution of F_{ST} changes with increasing recombination: with increasing recombination, the variance of the F_{ST} statistic decreases (HUDSON *et al.* 1992b). When we incorporated a moderate amount of recombination, the distribution of F_{ST} shrank. For example, with a scaled recombination parameter of $C = 10$, the 95% C.I. was $[-0.21, 0.35]$. With $C = 50$, the 95% C.I. was $[-0.15, 0.28]$. Therefore, with the proposed recombination rate parameter for the *per* locus, the observed F_{ST} falls outside the 95% C.I. expected under the neutral model (this was also found when we used the estimate of Nm from the set of loci including *per*).

The above results suggest that the *per* locus is under some form of selection, probably involving selection in different directions in the two localities (see above). Tests of neutrality such as Tajima's D , however, did not detect such selection. Given that the extent of recombination at this locus is not completely understood, this does not reject the presence of selection. One relevant piece of information is the result of the McDonald-Kreitman test (MCDONALD and KREITMAN 1991). When we use the combined data set for *per*, the McDonald-Kreitman test yields a marginally significant result ($P < 0.05$, one-tailed Fisher's exact test), due to either increased divergence at the silent sites or larger-than-expected nonsynonymous within-species polymorphism. The latter possibility is consistent with the action of diversifying selection on the PER protein. Together with the incompatibility of the observed large F_{ST} statistics with the neutral model, this suggests that the peculiar pattern of polymorphism at the *per* locus is likely to be the result of local selective effects that differ between the two populations.

We thank Molly Przeworski and Eli Stahl for providing the codes to compute the F_{ST} from a simulated data set, Stephen Wright for running the likelihood-ratio tests, and Dick Hudson for discussions. S.Y. was funded by the National Science Foundation Doctoral Dissertation Improvement grant DEB-9701098 during the course of this research. B.C. is supported by the Royal Society and D.B. by a European Molecular Biology Organization postdoctoral fellowship.

LITERATURE CITED

- ANDERSON, W. W., F. J. AYALA and R. E. MICHOD, 1977 Chromosomal and allozymic diagnosis of three species of *Drosophila*. *J. Hered.* **68**: 71–74.
- ANDOLFATTO, P., 2001 Contrasting patterns of X-linked and autosomal nucleotide variation in *Drosophila melanogaster* and *Drosophila simulans*. *Mol. Biol. Evol.* **18**: 279–290.
- ANDOLFATTO, P., F. DEPAULIS and A. NAVARRO, 2001 Inversion polymorphisms and nucleotide variability in *Drosophila*. *Genet. Res.* **77**: 1–8.
- BABCOCK, C. S., and W. W. ANDERSON, 1996 Molecular evolution of the sex-ratio inversion complex in *Drosophila pseudoobscura*: analysis of the *esterase-5* gene region. *Mol. Biol. Evol.* **13**: 297–308.
- BACHTROG, D., and B. CHARLESWORTH, 2000 Reduced levels of microsatellite variability on the neo-Y chromosome of *Drosophila miranda*. *Curr. Biol.* **10**: 1025–1031.
- BACHTROG, D., and B. CHARLESWORTH, 2002 Reduced adaptation of a non-recombining neo-Y chromosome. *Nature* **416**: 323–326.
- BONE, J. R., and M. I. KURODA, 1996 Dosage compensation regulatory proteins and the evolution of sex chromosomes in *Drosophila*. *Genetics* **144**: 705–713.
- BOTELLA, L. M., C. DOÑORO, L. SÁNCHEZ, C. SEGARRA and B. GRANADINO, 1996 Cloning and characterization of the *scute* (*sc*) gene of *Drosophila subobscura*. *Genetics* **144**: 1043–1051.
- CHARLESWORTH, B., 1996 The evolution of Y chromosomes, dosage compensation and X/autosome sex determination. *Curr. Biol.* **6**: 149–162.
- CHARLESWORTH, B., 2001 The effect of life-history and mode of inheritance on neutral genetic variability. *Genet. Res.* **77**: 153–166.
- CHARLESWORTH, B., and D. CHARLESWORTH, 2000 The degeneration of Y chromosomes. *Philos. Trans. R. Soc. Lond.* **355**: 1563–1572.
- CHARLESWORTH, B., M. NORDBORG and D. CHARLESWORTH, 1997 The effects of local selection, balanced polymorphism and background selection on equilibrium patterns of genetic diversity in subdivided populations. *Genet. Res.* **70**: 155–174.
- DAS, M., D. MUTSUDDI, A. K. DUTTAGUPTA and A. S. MUKHERJEE, 1982 Segmental heterogeneity in replication and transcription of the X₂ chromosome of *Drosophila miranda* and conservatism in the evolution of dosage compensation. *Chromosoma* **87**: 373–388.
- DOBZHANSKY, T., 1935 *Drosophila miranda*, a new species. *Genetics* **20**: 377–391.
- DOBZHANSKY, T., 1937 Further data on *Drosophila miranda* and its hybrids with *Drosophila pseudoobscura*. *J. Genet.* **34**: 135–151.
- DOBZHANSKY, T., and C. EPLING, 1948 The suppression of crossing over in inversion heterozygotes of *Drosophila pseudoobscura*. *Proc. Natl. Acad. Sci. USA* **34**: 137–141.
- DOBZHANSKY, T., and P. C. KOLLER, 1938 An experimental study of sexual isolation in *Drosophila*. *Biol. Z.* **58**: 589–607.
- DOBZHANSKY, T., and J. R. POWELL, 1975 *Drosophila pseudoobscura* and its American relatives, *Drosophila persimilis* and *Drosophila miranda*, pp. 537–587 in *Handbook of Genetics: Invertebrates of Genetic Interest*, Vol. 3, edited by R. C. KING. Plenum Press, New York.
- DOBZHANSKY, T., and C. C. TAN, 1936 Studies on hybrid sterility. III. A comparison of the gene arrangement in two species, *Drosophila pseudoobscura* and *Drosophila miranda*. *Z. Indukt. Abstammungs. Vererbungs.* **72**: 88–114.
- ERICKSON, J. W., and T. W. CLINE, 1998 Key aspects of the primary sex determination mechanism are conserved across the genus *Drosophila*. *Development* **125**: 3259–3268.
- FU, Y.-X., 1997 Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics* **147**: 915–925.
- FU, Y.-X., and W.-H. LI, 1993 Statistical tests of neutrality of mutations. *Genetics* **133**: 693–709.
- GODDARD, K., A. CACCONE and J. R. POWELL, 1990 Evolutionary implications of DNA divergence in the *Drosophila obscura* group. *Evolution* **44**: 1656–1670.
- GORDO, I., and B. CHARLESWORTH, 2001 The speed of Muller's ratchet with background selection, and the degeneration of Y chromosomes. *Genet. Res.* **78**: 149–161.
- HAMBLIN, M. T., and C. F. AQUADRO, 1999 DNA sequence variation and the recombinational landscape in *Drosophila pseudoobscura*: a study of the second chromosome. *Genetics* **153**: 859–869.
- HEY, J., and J. WAKELEY, 1997 A coalescent estimator of the population recombination rate. *Genetics* **145**: 833–846.
- HUDSON, R. R., 1990 Gene genealogies and the coalescent process. *Oxf. Surv. Evol. Biol.* **7**: 1–44.
- HUDSON, R. R., 2000 A new statistic for detecting genetic differentiation. *Genetics* **155**: 2011–2014.
- HUDSON, R. R., 2001 Two-locus sampling distributions and their application. *Genetics* **159**: 1805–1817.
- HUDSON, R. R., 2002 Generating samples under a Wright-Fisher neutral model of genetic variation. *Bioinformatics* **18**: 337–338.
- HUDSON, R. R., M. KREITMAN and M. AGUADÉ, 1987 A test of neutral

- molecular evolution based on nucleotide data. *Genetics* **116**: 153–159.
- HUDSON, R. R., D. D. BOOS and N. L. KAPLAN, 1992a A statistical test for detecting geographic subdivision. *Mol. Biol. Evol.* **9**: 138–151.
- HUDSON, R. R., M. SLATKIN and W. P. MADDISON, 1992b Estimation of levels of gene flow from DNA sequence data. *Genetics* **132**: 583–589.
- KIMURA, M., 1971 Theoretical foundation of population genetics at the molecular level. *Theor. Popul. Biol.* **2**: 174–208.
- KOVACEVIC, M., and S. W. SCHAEFFER, 2000 Molecular population genetics of X-linked genes in *Drosophila pseudoobscura*. *Genetics* **156**: 155–172.
- KRIMBAS, C. B., and J. R. POWELL, 1992 *Drosophila Inversion Polymorphism*. CRC Press, Boca Raton, FL.
- MACHADO, C. A., R. M. KLIMAN, J. A. MARKERT and J. HEY, 2002 Inferring the history of speciation from multilocus DNA sequence data: the case of *Drosophila pseudoobscura* and close relatives. *Mol. Biol. Evol.* **19**: 472–488.
- MACKNIGHT, R. H., 1939 The sex-determining mechanism of *Drosophila miranda*. *Genetics* **24**: 180–201.
- MARFANY, G., and R. GONZALES-DUARTE, 1992 The *Drosophila subobscura Adh* genomic region contains valuable evolutionary markers. *Mol. Biol. Evol.* **9**: 261–277.
- MARÍN, I., A. FRANKE, G. I. BRADSHAW and B. S. BAKER, 1996 The dosage compensation system of *Drosophila* is co-opted by newly evolved X chromosomes. *Nature* **383**: 160–163.
- MASIDE, X., C. BARTOLOME, S. ASSIMACOPOULOS and B. CHARLESWORTH, 2001 Rates of movement and distribution of transposable elements in *Drosophila melanogaster*: *in situ* hybridization vs. Southern blotting data. *Genet. Res.* **78**: 121–136.
- MCALLISTER, B. F., and B. CHARLESWORTH, 1999 Reduced sequence variability on the neo-Y chromosome of *Drosophila americana americana*. *Genetics* **153**: 221–233.
- MCDONALD, J. H., and M. KREITMAN, 1991 Adaptive protein evolution at the *Adh* locus in *Drosophila*. *Nature* **351**: 652–654.
- MORIYAMA, E. N., and J. R. POWELL, 1996 Intraspecific nuclear DNA variation in *Drosophila*. *Mol. Biol. Evol.* **13**: 261–277.
- NEI, M., 1987 *Molecular Evolutionary Genetics*. Columbia University Press, New York.
- NICOLETTO, B., 1959 An efficient method for salivary gland preparations. *Dros. Inf. Serv.* **33**: 181–182.
- PARK, Y., and M. I. KURODA, 2001 Epigenetic aspects of X-chromosome dosage compensation. *Science* **293**: 1083–1085.
- POWELL, J. R., 1992 Inversion polymorphisms in *Drosophila pseudoobscura* and *Drosophila persimilis*, pp. 73–126 in *Drosophila Inversion Polymorphism*, edited by C. B. KRIMBAS and J. R. POWELL. CRC Press, Boca Raton, FL.
- POWELL, J. R., 1997 *Progress and Prospects in Evolutionary Biology: The Drosophila Model*. Oxford University Press, New York.
- RICE, W. R., 1989 Analyzing tables of statistical tests. *Evolution* **43**: 223–225.
- ROSATO, E., and C. P. KYRIACOU, 2001 Flies, clocks and evolution. *Philos. Trans. R. Soc. Lond. B* **356**: 1769–1778.
- ROZAS, J., and R. ROZAS, 1999 DnaSP version 3: an integrated program for molecular population genetics and molecular evolution analysis. *Bioinformatics* **15**: 174–175.
- SEGARRA, C., G. RIBO and M. AGUADE, 1996 Differentiation of Muller's chromosomal elements D and E in the *obscura* group of *Drosophila*. *Genetics* **144**: 139–146.
- STEINEMANN, M., and S. STEINEMANN, 1990 Evolutionary changes in the organization of the major LCP gene cluster during sex chromosomal differentiation in the sibling species *Drosophila persimilis*, *D. pseudoobscura*, and *D. miranda*. *Chromosoma* **99**: 424–431.
- STEINEMANN, M., and S. STEINEMANN, 1998 Enigma of Y chromosome degeneration: neo-Y and neo-X chromosomes of *Drosophila miranda*, a model for sex chromosome evolution. *Genetica* **102/103**: 409–420.
- STEINEMANN, M., S. STEINEMANN and B. M. TURNER, 1996 Evolution of dosage compensation. *Chromosome Res.* **4**: 1–6.
- TAJIMA, F., 1989 Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* **123**: 585–595.
- WALL, J. D., 1999 Recombination and the power of statistical tests of neutrality. *Genet. Res.* **74**: 65–79.
- WALL, J. D., P. ANDOLFATTO and M. PRZEWSKI, 2002 Testing models of selection and demography in *Drosophila simulans*. *Genetics* **162**: 203–216.
- WATTERSON, G. A., 1975 On the number of segregating sites in genetical models without recombination. *Theor. Popul. Biol.* **7**: 256–276.
- WRIGHT, S. I., B. LAUGA and D. CHARLESWORTH, 2003 Subdivision and haplotype structure in natural populations of *Arabidopsis lyrata*. *Mol. Ecol.* **12**: 1247–1263.
- YI, S., and B. CHARLESWORTH, 2000 Contrasting patterns of molecular evolution of genes on the new and old sex chromosomes of *Drosophila miranda*. *Mol. Biol. Evol.* **17**: 703–717.

Communicating editor: S. W. SCHAEFFER

