Recombination, Dominance and Selection on Amino Acid Polymorphism in the Drosophila Genome: Contrasting Patterns on the X and Fourth Chromosomes

Lea A. Sheldahl, Daniel M. Weinreich and David M. Rand[1]

Department of Ecology and Evolutionary Biology, Brown University, Providence, Rhode Island 02912

Manuscript received December 14, 2002
Accepted for publication June 27, 2003

ABSTRACT

Surveys of nucleotide polymorphism and divergence indicate that the average selection coefficient on Drosophila proteins is weakly positive. Similar surveys in mitochondrial genomes and in the selfing plant Arabidopsis show that weak negative selection has operated. These differences have been attributed to the low recombination environment of mtDNA and Arabidopsis that has hindered adaptive evolution through the interference effects of linkage. We test this hypothesis with new sequence surveys of proteins lying in low recombination regions of the Drosophila genome. We surveyed >3800 bp across four proteins at the tip of the X chromosome and >3600 bp across four proteins on the fourth chromosome in 24 strains of *D. melanogaster* and 5 strains of *D. simulans*. This design seeks to study the interaction of selection and linkage by comparing silent and replacement variation in semihaploid (X chromosome) and diploid (fourth chromosome) environments lying in regions of low recombination. While the data do indicate very low rates of exchange, all four gametic phases were observed both at the tip of the X and across the fourth chromosome. Silent variation is very low at the tip of the X ($\theta_S = 0.0015$) and on the fourth chromosome ($\theta_S = 0.0002$), but the tip of the X shows a greater proportional loss of variation than the fourth shows relative to normal-recombination regions. In contrast, replacement polymorphism at the tip of the X is not reduced ($\theta_R = 0.00065$, very close to the X chromosome average). MK and HKA tests both indicate a significant excess of amino acid polymorphism at the tip of the X relative to the fourth. Selection is significantly negative at the tip of the X ($N_e s = -1.53$) and nonsignificantly positive on the fourth ($N_e s \sim 2.9$), analogous to the difference between mtDNA (or Arabidopsis) and the Drosophila genome average. Our distal X data are distinct from regions of normal recombination where the X shows a deficiency of amino acid polymorphism relative to the autosomes, suggesting more efficient selection against recessive deleterious replacement mutations. We suggest that the excess amino acid polymorphism on the distal X relative to the fourth chromosome is due to (1) differences in the mutation rate for selected mutations on the distal X or (2) a greater relaxation of selection from stronger linkage-related interference effects on the distal X. This relaxation of selection is presumed to be greater in magnitude than the difference in efficiency of selection between X-linked *vs.* autosomal selection.

SURVEYS of DNA sequence variation in a number of different organisms have established that levels of nucleotide polymorphism show a positive correlation with local rates of recombination across genomes (Begun and Aquadro 1992; Stephan and Langley 1998; Nachman 2001). The lack of a correlation between recombination and rates of nucleotide divergence between species indicates that recombination is not mutagenic (Begun and Aquadro 1992). Consequently, some form of natural selection is thought to be responsible for the reduced variation in regions of low recombination. Two competing hypotheses have been proposed to explain this pattern and have been the focus of much research in theoretical and empirical population genetics. The hitchhiking hypothesis posits that periodic adaptive fixation of new mutations leads to selective sweeps of nucleotide polymorphism in the regions flanking the beneficial mutation (Maynard Smith and Haigh 1974; Kaplan *et al.* 1989). The background selection hypothesis focuses on negative selection against deleterious mutations. Removal of such mutations leads to locally reduced variation among the smaller number of mutation-free chromosomes that remain to leave descendants in the population (Charlesworth *et al.* 1993). These models, and their extensions that invoke both background selection and hitchhiking (Kim and Stephan 2000) or fluctuating selection (Gillespie 1997, 2000), predict a positive correlation between rates of recombination and levels of polymorphism.

Distinguishing how these competing hypotheses account for patterns of nucleotide polymorphism is a major goal of molecular population genetics. Studies that have documented the reduction of polymorphism in

[1]*Corresponding author:* Department of Ecology and Evolutionary Biology, Box G-W, Brown University, Providence, RI 02912.
E-mail: david_rand@brown.edu

regions of low recombination are based on sequence surveys of primarily noncoding, intron sequences or microsatellites (*e.g.*, Aguadé *et al.* 1989; Begun and Aquadro 1991; Aquadro *et al.* 1994; Langley *et al.* 2000; Payseur and Nachman 2000; Jensen *et al.* 2002). Protein-coding exon sequences appear to have been largely avoided, presumably because a higher level of functional constraint should lead to lower levels of variation. One way to distinguish background selection and selective sweeps is to measure site frequencies and estimate the skew of the frequency spectrum (*e.g.*, Langley *et al.* 2000). Given a fixed effort to survey polymorphism, less-constrained noncoding and intron sequences should harbor more polymorphic sites and provide better estimates of Tajima's *D*, a statistic with wide confidence limits (Simonsen *et al.* 1995). One consequence of these efforts is a remarkable lack of polymorphism data from protein-coding sequences in regions of low recombination, even in Drosophila.

The positive correlation between recombination and polymorphism allows one to examine how recombination alters the efficacy of selection across the genome. Low recombination environments should weaken natural selection through effectively reduced population sizes and through interference effects associated with linkage (Hill and Robertson 1966; Felsenstein 1974). For example, selection for preferred codons may be reduced in regions of low recombination due to Hill-Robertson interference (Kliman and Hey 1993; Hey and Kliman 2002; but see Marais *et al.* 2001 for alternative hypotheses). Surveys of silent (synonymous) and replacement (nonsynonymous) variation provide another means of studying this problem, but these surveys are underrepresented in regions of low recombination. Here we seek to fill this gap with new sequence data from protein-coding regions on the tip of the X and the fourth chromosome of *Drosophila melanogaster* and *D. simulans*

There are several motivating factors for this study. First, surveys of polymorphism and divergence at silent and replacement sites allow one to estimate the historical selection coefficient acting on a protein using the McDonald-Kreitman (MK) test and its derivatives (McDonald and Kreitman 1991; Sawyer and Hartl 1992; Akashi 1995; Bustamante *et al.* 2002). Noncoding sequences do not provide the distinct functional classes that can be applied to expected ratios of polymorphism and divergence needed to estimate selection (Kimura 1983). An implicit assumption of studying noncoding regions is that they serve as marker loci of selection operating at some unknown site or sites elsewhere in the low-recombination region. On functional grounds alone, sequencing surveys of coding regions are more likely to survey some of the sites that are responsible for the selective events that drive either background selection or hitchhiking. While it is still virtually impossible to determine whether the selected sites have been surveyed, new data for polymorphism and divergence at silent and replacement sites could provide estimates of historical selection that help distinguish these hypotheses.

A second motivation for this study comes from MK tests of mitochondrial genes. Animal mtDNAs tend to show a nonneutral excess (relative to divergence) of amino acid polymorphism most likely resulting from selection against mildly deleterious mutations that hinder fixation (Nachman 1998; Rand and Kann 1998; Rand 2001). This signature of negative selection in mtDNA is significantly different from the broad distribution of selection coefficients estimated from MK tests of nuclear genes in Drosophila (Weinreich and Rand 2000). The lack of recombination and the haploid mode of inheritance in mtDNA are two factors that are likely to account for this difference in the molecular evolution of nuclear and mitochondrial genomes. In support of this hypothesis, Weinreich and Rand (2000) showed that nuclear genes in Arabidopsis also exhibit a nonneutral excess of amino acid polymorphism indicative of negative selection. They argued that the effectively low-recombination genomic environment of this selfing plant predisposes it to a history of negative selection much like that observed in animal mtDNA. The higher average levels of recombination in outcrossing Drosophila increase the likelihood that beneficial amino acid mutations can rise to fixation, allowing certain genes to exhibit a history of positive selection as measured with MK tests (Weinreich and Rand 2000). The Drosophila data suggest that proteins in regions of low recombination also exhibit an excess of amino acid polymorphism, but more data are needed to examine this relationship.

A third motivation for this study is our wish to examine the possible interaction of recombination and dominance in modulating selection on nucleotide polymorphisms. Aquadro *et al.* (1994) showed that levels of polymorphism were consistently lower across the X chromosome than across the (autosomal) third chromosome in *D. melanogaster*, even after correcting for differences in effective population size. If reduced polymorphism in regions of low recombination is due to hitchhiking of recessive advantageous mutations, the X chromosome should show lower levels of polymorphism since these mutations are more likely to reach fixation on the X *vs.* an autosome. Alternatively, background selection predicts that the purging of deleterious mutations should be more effective on the X, leaving a larger class of mutation-free chromosomes, and hence higher levels of polymorphism relative to low recombination regions of autosomes. These predictions depend on the average dominance coefficient for new mutations, which can blur the distinction between the background selection and hitchhiking hypotheses (Charlesworth 1996; McVean and Charlesworth 2000).

The Drosophila genome provides an ideal resource for testing many of these competing predictions. The

ends of chromosome arms tend to have very reduced levels of recombination (Ashburner 1989), as does most of the small fourth chromosome (Wang *et al.* 2002). The tip of the X chromosome has very low levels of recombination and is haploid in males. Additional data from protein-coding regions should provide some insight into the interaction between recombination and dominance on the efficiency of selection. Here we compare MK tests from proteins at the tip of the X and the fourth chromosome. The data reveal an unexpected excess of amino acid variation at the tip of the X, indicative of negative selection, while the fourth chromosome shows a nonsignificant deficiency of amino acid variation, suggesting positive selection. The relaxation of selection at the tip of the X may outweigh the more efficient selection afforded by partially haploid selection.

MATERIALS AND METHODS

Fly strains: Sequence data were collected from 24 lines of *D. melanogaster*, 5 lines of *D. simulans*, and 1 line of *D. yakuba*. Ten of the *D. melanogaster* lines were collected at a local farm (Four Town Farm, Seekonk, Massachusetts: FTF 1, 2, 5, 6, 14, 20, 23, 26, 28, and 105), 10 were collected by E. Zouros outside Iraklion, Crete, Greece (Crete 8, 24, 26, 30, 31, 35, 40, 42, 43, and 44), and 4 were from Zimbabwe, Africa (Zim 2, 11, 30, and 53, obtained from C.-I Wu). Two of the *D. simulans* lines were from Harare, Zimbabwe (DsimZimH 13 and 48), 2 were from Florida (DsimFl 10 and 13, all from C. F. Aquadro), and 1 was from the Seychelles Islands (DsimSey, provided by C.-I Wu). The Zimbabwe and Florida strains carry the *si*II mtDNA haplotype while the Seychelles stock carries the *si*I mtDNA (Solignac *et al.* 1986; Ballard 2000).

All *D. melanogaster* chromosomes were extracted using either the FM7 balancer for the X chromosome or the $ci^D/eyeless^D$ dominant marked stock for the fourth chromosome. Crosses were carried out to obtain a homozygous stock for a single wild chromosome of interest. All chromosomes were extracted using wild female × balancer male crosses to avoid heterogeneity in the extracted stock due to possible mobility of transposable elements from hybrid dysgenic crosses. Single-pair mating was carried out for three generations to increase homozygosity in the *D. simulans* and *D. yakuba* strains (no heterozygous sequences were observed).

DNA preparation and sequencing: DNA was prepared from a single fly from each homozygous strain following the "squish prep" protocol of Gloor and Engels (1991). A single fly was homogenized with a Kontes pestle and motorized homogenizer in a 1.5-ml microcentrifuge tube in 50 μl of squish buffer [10 mm Tris (pH 8.2), 1 mm EDTA, 25 mm NaCl]. After homogenization, 50 μl of squish buffer with proteinase K at a concentration of 4 μg/ml was added to the tube, and the sample was incubated at 37° for 1 hr and then denatured at 95° for 3 min. Two microliters of this homogenate was used as a template for each PCR amplification.

DNA amplification was carried out in 25-μl reactions using the primers listed in Table 1. The thermal profile was as follows: 1 min denaturation at 95°, followed by 30 cycles of 95° for 30 sec, annealing temperature for 30 sec (see Table 1), and extension at 72° for 1 min. A 10-min extension at 72° followed the 30 cycles of amplification. Amplified DNA was purified using the QIAGEN (Valencia, CA) PCR purification kit (no. 28106) and 3.5 μl of the resulting 20-μl eluate plus 0.32 ml of 10 mm sequencing primer was subjected to cycle sequencing PCR for 30 cycles of 95° for 30 sec, 53° for 1 min, 72° for 1 min, followed by a 72° soak. Sequenced templates were run on an ABI 377 automated sequencer. ABI output was analyzed with Sequencher 3.0 with no alignment ambiguities. All sequence data were confirmed in both directions from a second individual fly from the extracted chromosome stock.

Genes selected for sequencing: In June of 2000 we selected four of the most-distal genes at the tip of the X chromosome in release 1.0 of the Drosophila genome (Adams *et al.* 2000). We chose genes containing large exons (>800 bp) so that a single contiguous sample of coding sequence could be obtained. The four genes on the X chromosome are *RhoGAP1A* (located at 1A; also identified by CG17960 and EG:23E12.2), CG3038 (located at 1A6; also identified by EG:BACR3P7.1), *cinnamon* (located at 1A7; also identified by CG2945), and CG3777 (located at 1B1; also identified by EG:125H10.1). A transcript has been identified for each gene according to annotation available through the GadFly web page of Flybase (www.flybase.org). The transcripts are as follows: for RhoGAP1A, transcript CG17960-RA; for CG3038, transcript CT10170, cDNA LD16783, and expressed sequence tag (EST) RE05756; for *cinnamon*, transcript CT9961 and cDNA GH09380; and for CG3777, transcript CT12604 and EST SD17974. These data indicate that we have sampled exons from expressed genes.

For the fourth chromosome, we selected single large exons from four genes spread widely across the chromosome: *pangolin*, *zinc finger homeodomain 2*, *pleiohomeotic*, and *ATP synthase beta*. Each of these genes is well documented in Flybase as a functional gene with transcripts containing the exons we sequenced. Figure 1 and Table 1 identify the genes and exons chosen for sequencing, plus accession numbers. For the 29 strains (24 *D. melanogaster*, 5 *D. simulans*), 3893 bp of exon sequence were collected from the X chromosome loci and 3629 bp of exon sequence were collected from the fourth chromosome loci, for a total of 218 kb. There are some short gaps in some of the strains where amplification and sequencing were inconsistent.

Statistical analyses: Tabulation of silent and replacement polymorphism and divergence, neutrality tests, and population structure was performed in DNAsp (version 3.51; Rozas and Rozas 1999). MK tests (McDonald and Kreitman 1991) were tabulated by hand on the basis of observed numbers of polymorphic and fixed silent and replacement changes, with significance determined from *G*-tests (Sokal and Rohlf 1981). The significance of the θ_H statistic based on the frequency of derived polymorphisms (Fay and Wu 2000) was calculated with the web tool (http://crimp.lbl.gov/htest.html), using 10,000 simulated genealogies and a probability of back mutation of 0.1.

Selection coefficients were estimated from MK test data using neutrality index (NI) values (Rand and Kann 1996; Weinreich and Rand 2000) and maximum-likelihood estimation methods based on a Poisson random field approach (Sawyer and Hartl 1992; Bustamante *et al.* 2002). NI is defined as (PR/FR)/(PS/FS), where PR and FR refer to *p*olymorphic and *f*ixed *r*eplacement (nonsynonymous) sites, respectively, and PS and FS refer to *p*olymorphic and *f*ixed *s*ilent (synonymous) sites, respectively (Rand and Kann 1996). There is a monotonic relationship between NI and $N_e s$, the historical effective selection coefficient inferred from Kimura's diffusion approximations of rates of divergence and polymorphism for selected and neutral sites (Kimura 1983, p. 44; Sawyer and Hartl 1992; Akashi 1995; Nachman 1998; Weinreich and Rand 2000). NI > 1 derives from an excess of amino acid polymorphism (or a deficiency of amino acid fixation), implying negative selection, while NI < 1 derives from a defi-

TABLE 1

Genes surveyed and primers used for amplification and sequencing

Gene, contig accession no., exon position, accession no. for alleles	Primer positions	Primer (5′ → 3′)	Primer name
RhoGAP1A (EG:23E12.2),	22,520–22,542	atgcagtacaagaaggccatcca	12.2A A+
AL031884; GI:3763961,	23,990–23,971	acttttaactcccaccttaa	12.2A A−
22,520–23,530,	23,135–23,155	gtcatcgcatgtgccaaatac	12.2A B+
AY312787–AY312815	23,075–23,055	aagaattacgcaccatgcgtc	12.2A B−
	22,941–22,960	atgcgcatccggatcaccag	12.2A C+
	23,607–23,588	gcaggttgcatttttgctac	12.2A C−
CG3038 (EG:BACR37P7.1),	2,278–2,297	gccagcccgacaaataggcg	P7.1 B+
AE003417; GI:22831400,	2,842–2,822	ctttgctttaatcaaccattg	P7.1 A−
1,877–2,784,	1,785–1,804	caatatttgtctcttccttc	P7.1 A+
AY312642–AY312670	2,346–2,327	ccaataagtggtacgttagc	P7.1 B−
	1,886–1,905	tcgcaacgcttttatttcag	P7.1 1+
cinnamon,	13,267–13,288	gtgggcttttatccccatagac	cin A+
AE003417; GI:22831400,	13,815–13,795	cttttgggcggatccttggac	cin A−
13,319–14,079,	13,729–13,750	cttttgcctcacgctgttcacc	cin B+
AY312700–312728	14,219–14,200	gccagcggcaaagaattagg	cin B−
CG3777 (EG:125H10.1),	103,879–103,900	atgaagagcatcgaggccaaaa	H10.1 A+
AE003417; GI:22831400,	105,145–105,126	ttctaccagtcttgtttgtt	H10.1 A−
103,879–105,145 complement,	104,862–104,842	ctcatcctcatcagtacttgc	H10.1 C−
AY312671–AY312699	103,798–103,917	gactattcgatgaatctgct	H10.1 C+
pangolin,	127,493–127,512	ctgttctgtaactcttaagg	pan A+
AE003845; GI:28380223,	128,376–128,357	tcgtgggttagcaccaatag	pan B−
127,653–128,742,	128,859–128,840	gatggctttgggctggcgag	pan A−
AY312729–AY312757	127,825–127,844	gaagatgaggactcggaatc	pan B+
	127,673–127,692	gttatatggaagccctgaac	pan 1+
	128,213–128,233	caccagtagttagcacgagca	pan C+
zinc finger homeodomain 2,	63,979–63,999	caatggtgacctcaggcagtg	zfh B+
AE003843; GI:28380216,	64,986–64,965	ccgacaaggactgcatttcatc	zfh B−
63,123–66,020,	64,640–64,621	gaagtcagtaagcgaagaag	zfh D−
AY312816–AY312844			
Pleiohomeotic,	269,264–269,244	gacaaaataaaagcggctgac	pho A+
AE003846; GI:28380228,	268,317–268,338	catttccccacatccagtcgag	pho A−
268,401–269,170,	168,754–268,735	atgcacagacccttgaaatg	pho B+
AY312758–AY312786	269,331–269,312	tgaccgaattattcattcag	pho 1+
ATP synthase beta,	128,795–128,746	tcaatgatagatttgctgac	ATP A−
AE003846; GI:28380228,	127,753–127,772	tgccggtgtgggcaaaactg	ATP B+
127,578–128,719 complement,	128,214–128,233	atgtgcccgctgatgatttg	ATP C+
AY312613–AY312641	128,715–128,696	gcagcttcctttgccaggcg	ATP 1−
	127,485–127,504	gttggtgccgaaacactagg	ATP A+
	128,283–128,264	acagtggtggcatccaaatg	ATP B−

The first four genes are on the X and the second four genes are on the fourth chromosome (see Figure 1).

ciency of amino acid polymorphism (or an excess of amino acid fixation), implying positive selection. The maximum-likelihood methods of Bustamante *et al.* (2002) provide an estimate of $N_e s$ that allows for mutation, selection, and drift at independent sites in evolving DNA sequences and considers each cell of MK test data sets as a Poisson random variable for comparisons to expected values. Estimates of $N_e s$ from Poisson random field (PRF) methods are probably more accurate than those from NI over a wider range of values, but both estimates are correlated.

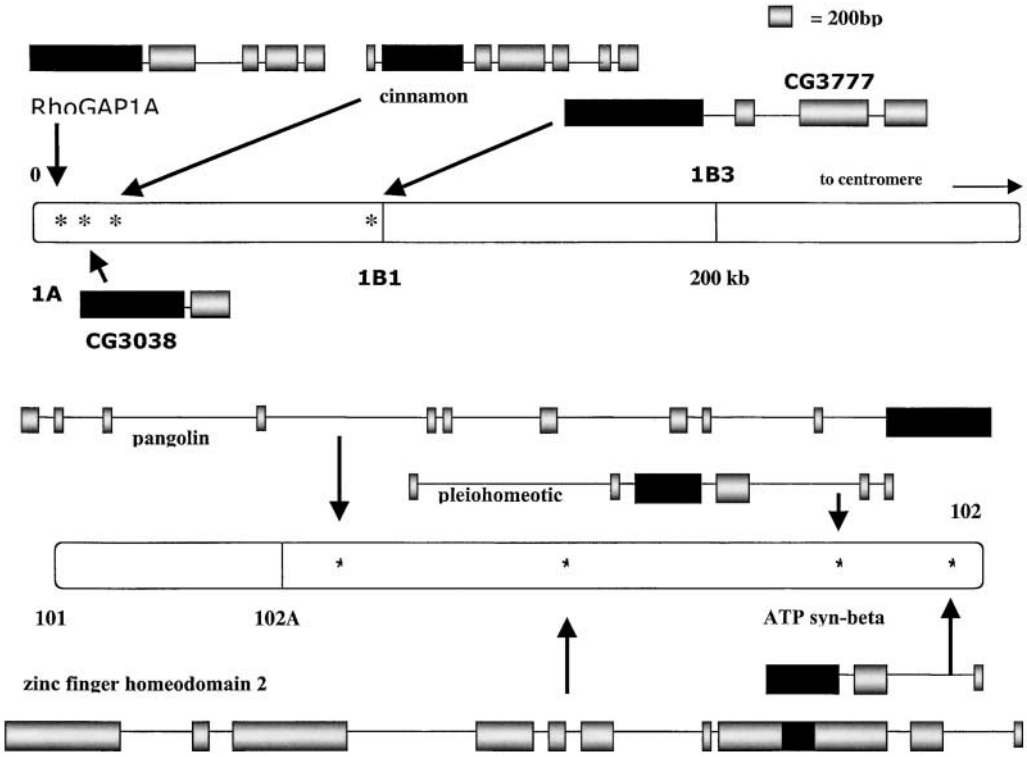

FIGURE 1.—Location of the exons surveyed in this study. The introns (lines) and exons (boxes) show the gene structure for the loci surveyed; the black shading identifies the specific exon regions surveyed for polymorphism and divergence. Top, tip of the X chromosome; bottom, fourth chromosome. Asterisks identify the relative location of the gene on the chromosome, with number plus letter referring to polytene map location.

RESULTS

Polymorphism in *D. melanogaster*: Table 2 shows the estimates of silent and replacement polymorphism for each gene in each sample of chromosomes sequenced. Silent site diversity in the total sample of Massachusetts, Crete, and Zimbabwe chromosomes ($n = 24$) across the tip of the X chromosome ($\theta_{S\text{-distal X}} = 0.00152$) is reduced >14-fold relative to the X chromosome average ($\theta_{S\text{-X}} = 0.0216$ for five genes in regions of normal recombination on the X sampled from both African and non-African localities; ANDOLFATTO 2001). Silent polymorphism in our total sample of four genes on the fourth chromosome ($\theta_{S\text{-fourth}} = 0.00201$) is reduced >8-fold compared to a genome-wide average for autosomes ($\theta_{S\text{-Auto}} = 0.0168$ for 16 genes across the second and third chromosomes sampled from both African and non-African localities; ANDOLFATTO 2001). The same pattern is evident when silent polymorphism is based on pairwise nucleotide diversity (π), although the normal/low recombination ratios are more pronounced. The X chromosome average for silent $\pi = 0.0190$, while the tip of the X $\pi = 0.0.00091$ (average/tip ratio = 20.8); the autosome average silent $\pi = 0.0158$, while the fourth chromosome average is $\pi = 0.0.00132$ (average/fourth ratio = 12.0).

Coalescent simulations run in DNAsp and conditioned on π and θ show that the 95% confidence limits of silent π and θ for the new X and fourth chromosome data do not include the genome-wide averages for π and θ reported in ANDOLFATTO (2001). Moreover, if coalescent simulations are run using input values for π, θ, and sample size for the average X-linked or autosomal data set from Table 3 in ANDOLFATTO (2001), the 95% confidence limits do not include silent π or θ from our data at the tip of the X or the fourth chromosome even using the conservative assumption of no recombination (data not shown). This confirms that protein-coding sequences show reduced polymorphism in regions of low recombination, a pattern that is well documented for primarily noncoding DNA (AGUADÉ *et al.* 1989; BEGUN and AQUADRO 1991; BERRY *et al.* 1991; JENSEN *et al.* 2002; WANG *et al.* 2002).

The greater reduction of polymorphism at the tip of the X compared to the fourth chromosome (relative to genome-wide averages) is due to elevated X/autosome polymorphism ratios in African samples of *D. melanogaster* (ANDOLFATTO 2001). The X/autosome ratio of silent polymorphism (θ) in worldwide samples of *D. melanogaster* is 1.28 (1.60 for African samples and 0.67 for non-African samples). The ratio of polymorphism for our worldwide sample of exons at the tip of the X and fourth chromosome (X/fourth) is 0.76 [0.83 for our African samples and 0.70 in the non-African (Massachusetts + Crete) sample]. Thus, no partitions of our exon data appear to deviate dramatically from the 3/4 ratio of X/autosome polymorphism expected under neutrality and equal sex ratio. Our African sample shows a somewhat higher X/fourth ratio, but it is not significantly greater than the ratio for non-African samples (Fisher's

TABLE 2

Polymorphism in *D. melanogaster*

	Total sample (n = 24)						Massachusetts (n = 10)					
			Silent		Replacement				Silent		Replacement	
Gene (bp)	S	R	π	θ	π	θ	S	R	π	θ	π	θ
					X chromosome							
RhoGAP1A, 937	0	2	0	0	0.00061	0.00076	0	1	0	0	0.00073	0.00048
CG3038, 886	1	1	0.00163	0.00127	0.00075	0.00040	1	1	0.00261	0.00166	0.00029	0.00052
cin, 761	2	2	0.00091	0.00292	0.00042	0.00093	0	0	0	0	0	0
CG3777, 1225	2	2	0.00111	0.00190	0.00078	0.00058	0	2	0	0	0.0008	0.00075
Mean			0.00091	0.00152	0.00064	0.00067			0.00065	0.00042	0.00046	0.00044
X haplotype	5	7	0.00093	0.00151	0.00066	0.00065	1	4	0.00062	0.00039	0.00051	0.00048
					Fourth chromosome							
pan, 1063	3	0	0.00278	0.00352	0	0	1	0	0.00225	0.00143	0	0
zfh 2, 625	1	0	0.00059	0.00189	0	0	1	0	0.00139	0.00245	0	0
ATP, 1142	1	0	0.00138	0.00094	0	0	1	0	0.00137	0.00129	0	0
pho, 766	1	0	0.00052	0.00169	0	0	1	0	0.00126	0.00223	0	0
Mean			0.00132	0.00201	0	0			0.00157	0.00172	0	0
Fourth haplotype	6	0	0.00147	0.00198	0	0	4	0	0.00157	0.00169	0	0
X/fourth ratio of means			0.693	0.757	—	—			0.416	0.241	—	—
X/fourth ratio of haplotypes			0.633	0.763	—	—			0.395	0.231	—	—

	Crete (n = 10)						Zimbabwe (n = 4)					
			Silent		Replacement				Silent		Replacement	
Gene (bp)	S	R	π	θ	π	θ	S	R	π	θ	π	θ
					X chromosome							
RhoGAP1A, 937	0	1	0	0	0.00027	0.00048	0	1	0	0	0.00071	0.00077
CG3038, 886	0	1	0	0	0.00029	0.00052	0	0	0	0	0	0
cin, 761	1	0	0.00109	0.00192	0.0	0.0	1	2	0.00272	0.00297	0.00203	0.00190
CG3777, 1225	2	1	0.00071	0.00126	0.00038	0.00038	1	0	0.00175	0.00191	0	0
Mean			0.00045	0.00080	0.00024	0.00029			0.00112	0.00099	0.00069	0.00067
X haplotype	2	3	0.00045	0.00079	0.00026	0.00036	2	3	0.00113	0.00123	0.00058	0.00057
					Fourth chromosome							
pan, 1063	1	0	0.00081	0.00144	0	0	2	0	0.0051	0.00477	0	0
zfh 2, 625	0	0	0	0	0	0	0	0	0	0	0	0
ATP, 1142	0	0	0	0	0	0	0	0	0	0	0	0
pho, 766	0	0	0	0	0	0	0	0	0	0	0	0
Mean			0.00020	0.00036	0	0			0.00128	0.00119	0	0
Fourth haplotype	1	0	0.00024	0.00042	0	0	2	0	0.00143	0.00134	0	0
X/fourth ratio of means			2.222	2.208	—	—			0.876	0.830	—	—
X/fourth ratio of haplotypes			1.875	1.881	—	—			0.790	0.918	—	—

exact test, $P > 0.5$ based on polymorphism counts). Genome-wide, however, the higher X/autosome ratio in African *vs.* non-African samples is highly significant (Andolfatto 2001). In summary, silent polymorphism in regions of low recombination is reduced at least 10-fold relative to regions of normal recombination and the relative reduction is about two to three times greater at the tip of the X than on the fourth chromosome.

Replacement polymorphism in *D. melanogaster*: Amino acid polymorphism at the tip of the X chromosome is very different from that on the fourth chromosome, and both of these regions differ from the genome-wide

GENE	RhoGAP1		CG3038		CINNAMON					CG3777					
SITE (5'-->3')	721	833	1	677	128	132	266	602	615	27	178	190	285	455	621
Silent/Repl.	R	R	S	R	S	R	S	S	R	S	R	R	S	R	S
CRETE 10	C	C	T	C	A	C	C	G	G	C	G	G	G	A	C
CRETE 16	.	G	.	T	T	.	.	.	.	.	.	.	.	.	.
CRETE 24	.	G	.	T	.	.	.	.	.	.	A	.	.	.	.
CRETE 2	.	G	.	T	.	.	.	.	.	.	A	.	.	.	.
CRETE 33	.	G	.	T	.	.	.	.	.	.	.	.	.	.	.
CRETE 35	.	G	.	T	.	.	.	.	.	.	.	.	.	.	.
CRETE 40	.	G	.	T	.	.	.	.	.	.	.	.	.	.	.
CRETE 43	.	G	.	T	.	.	.	.	.	T	.	.	.	.	.
CRETE 44	.	G	.	T	.	.	.	.	.	.	.	.	.	.	.
CRETE 49	.	G	.	T	.	.	.	.	.	.	.	.	.	.	.
FTF 100	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
FTF 105	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
FTF 14	.	G	.	T	.	.	.	.	.	.	.	.	.	.	.
FTF 15	.	G	C	.	.	.	.	.	.	.	.	.	.	G	.
FTF 17	.	G	C	.	.	.	.	.	.	.	.	.	.	G	.
FTF 26	.	G	C	.	.	.	.	.	.	.	.	.	.	G	.
FTF 28	.	G	C	.	.	.	.	.	.	.	.	.	.	G	.
FTF 29	.	G	C	.	.	.	.	.	.	.	.	.	.	G	.
FTF 5	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
FTF 89	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.
Zmb 11	.	G	.	.	.	.	.	.	T	.	.	.	.	G	.
Zmb 2	.	G	.	.	.	.	T	.	T	.	.	.	.	G	A
Zmb 30	T	G	.	.	.	G	.	.	.	.	.	.	.	G	A
Zmb 53	.	G	.	.	.	.	.	.	.	.	.	.	.	G	A
SIM FL H10	.	.	C	.	.	.	T	C	.	.	.	.	.	G	.
SIM FL H13	.	.	C	.	.	.	T	C	.	.	.	.	.	G	.
SIM ZMB H48	.	.	C	.	.	.	T	.	.	.	.	C	.	G	.
SIM ZMB H13	.	.	C	.	.	.	T	C	.	.	.	C	.	G	.
SIM SEY	.	.	C	.	.	.	T	.	.	.	.	.	A	G	.

FIGURE 2.—Polymorphic sites across the tip of the X. Each row refers to a distinct strain; each column refers to a polymorphic site. Crete, FTF (Four Town Farm, Massachusetts), and Zmb (Zimbabwe) are *D. melanogaster* strains. SIM FL, SIM ZMB, and SIM SEY are *D. simulans* from Florida, Zimbabwe, and the Seychelles, respectively. Numbers at the top refer to sites in an aligned data set for each gene, and S and R refer to silent (synonymous) and replacement (nonsynonymous) changes.

averages in replacement/silent ratios. All four genes at the tip of the X chromosome show amino acid polymorphism, but none of the fourth chromosome genes do (Table 2, Figures 2 and 3). On a per-site basis, there is essentially no reduction of replacement polymorphism at the tip of the X ($\theta_{\text{R-distal X}} = 0.00065$; see Table 2) relative to the X chromosome average ($\theta_{\text{R-X}} = 0.0006$; from ANDOLFATTO 2001). The absence of amino acid polymorphism in our sample of fourth chromosome genes (and in an earlier study of *cubitus interruptus* in *D. melanogaster*; BERRY *et al.* 1991) does indicate levels of replacement polymorphism lower than the autosomal average ($\theta_{\text{R-Auto}} = 0.0014$, from ANDOLFATTO 2001).

Because silent polymorphism has been reduced both at the tip of the X and on the fourth, but replacement polymorphism has been reduced only on the fourth, the ratios of amino acid polymorphisms to silent polymorphisms (A/S ratios) show differences between chromosomes and in comparison to the genome-wide averages. At the tip of the X the A/S ratio is $7/5 = 1.4$, but on the fourth chromosome, $A/S = 0/6 = 0.0$ in our sample (Figures 2 and 3). These ratios are significantly different by Fisher's exact test (two-tailed $P < 0.038$) and indicate an excess of amino acid polymorphism at the tip of the X relative to the fourth chromosome. The A/S ratio for 10 genes across the X chromosome in regions of normal recombination is $17/148 = 0.11$ while the A/S ratio for 35 autosomal genes is $131/269 = 0.49$ (FAY *et al.* 2002). These X *vs.* autosome ratios are highly significantly different ($G = 29.3$, d.f. = 1, two-tailed $P <$ 0.0001) but indicate a *deficiency* of amino acid polymorphism on the X relative to the autosomes. The reduced A/S ratio on the X, relative to the autosome (for average

GENE	pangolin					zfh 2	ATP	pho
SITE (5'-->3')	64	355	472	802	936	145	893	541
Silent/Repl.	S	S	S	S	R	S	S	S
CRETE 24	C	A	A	T	A	G	G	C
CRETE 26	.	.	.	.	.	.	.	.
CRETE 30	.	.	.	C	.	.	.	.
CRETE 31	.	.	.	.	.	.	.	.
CRETE 35	.	.	.	.	.	.	.	.
CRETE 40	.	.	.	.	.	.	.	.
CRETE 42	.	.	.	.	.	.	.	.
CRETE 43	.	.	.	.	.	.	.	.
CRETE 44	.	.	.	.	.	.	.	.
CRETE 8	.	.	.	.	.	.	.	.
FTF 1	.	.	.	C	.	.	.	.
FTF 105	.	.	.	.	.	.	.	T
FTF 26	.	.	.	C	.	.	.	.
FTF 20	.	.	.	.	.	.	.	.
FTF 14	.	.	.	.	.	.	.	.
FTF 28	.	.	.	.	.	.	A	.
FTF 6	.	.	.	C	.	.	.	.
FTF 23	.	.	.	C	.	C	A	.
FTF 5	.	.	.	.	.	.	.	.
FTF 2	.	.	.	C	.	.	.	.
ZIM 11	.	.	.	.	.	.	A	.
ZIM 2	T	.	.	.	.	.	A	.
ZIM 30	.	.	T	.	.	.	A	.
ZIM 53	.	.	T	.	.	.	A	.
SIM FL H10	.	.	.	.	.	.	.	.
SIM FL H13	.	.	.	.	.	.	.	.
SIM ZIM H48	.	.	.	.	.	.	.	.
SIM ZIM H13	.	T	.	.	.	.	.	.
SIM SEY	.	T	.	.	T	.	.	.

FIGURE 3.—Polymorphic sites across the fourth chromosome. Row and column heads are the same as in Figure 2.

TABLE 3

HKA tests

Silent HKA	Polymorphic	Fixed	No. of sites	χ^2	P
X	5 (4.71)	83 (85.8)	883		
Fourth	6 (6.29)	84 (83.7)	813	0.021	0.8856
Replacement HKA	**Polymorphic**	**Divergence**	**No. of sites**	χ^2	P
X	7 (2.9)	33 (37.2)	2876		
Fourth	0 (4.2)	44 (39.9)	2708	8.278	0.0046

Data are for 24 four-gene haplotypes from the X and fourth chromosome (see Table 2, Figures 2 and 3), compared to a single *D. simulans* sequence. Numbers in parentheses are expected values.

levels of recombination) is attributed to more effective elimination of deleterious amino acid polymorphisms on the X chromosome (Begun 1996; Andolfatto 2001; Fay *et al.* 2002). However, this explanation does not hold for our data for low-recombination regions on the X and fourth chromosomes.

Hudson-Kreitman-Aguade tests: Table 3 shows the results of Hudson-Kreitman-Aguade (HKA) tests (Hudson *et al.* 1987), comparing the X and the fourth data done separately for silent and replacement sites. Replacement sites show a much stronger departure from neutral expectations than do silent sites (see Table 3). The fourth chromosome shows no significant reduction of silent polymorphism compared to the tip of the X ($\chi^2 = 0.021$, $P = 0.8856$). Departure from neutral evolution is much more pronounced for replacement sites ($\chi^2 = 8.28$, $P = 0.0046$; see Table 3), due to a relatively higher level of replacement polymorphism at the tip of the X. Since the X and fourth chromosome data derive from the same sample of 24 lines, the different outcomes of these silent and replacement HKA tests are not likely due to sampling problems.

Nucleotide site frequencies in *D. melanogaster*: While Tajima's (1989) *D*-values tend to be negative, none of the individual geographic samples show a significant skew in the frequency spectrum (data not shown). The average values for Tajima's *D* across each gene for pooled samples from all three localities are negative, and if the sequences are concatenated into a single haplotype for each strain, the value of Tajima's *D* become more negative (tip of X, average $D_{taj} = -0.236$, haplotype $D_{taj} = -0.530$; fourth chromosome, average $D_{taj} = -0.518$, haplotype $D_{taj} = -0.781$). These negative *D*-values are somewhat surprising given evidence for population subdivision among sampling localities (see below), which should increase D_{taj} values for the pooled sample. However, neither the X nor the fourth chromosome haplotype analyses show a significant skew in the frequency spectrum by the Tajima test. The same is true for Fu and Li's (1993) statistics ($D^* = -0.72$, $F^* = -0.77$, $P > 0.10$).

Fay and Wu's (2000) test compares the frequencies of derived mutations relative to that expected under a neutral mutation-random drift hypothesis. Derived mutations are inferred by comparing the two segregating nucleotide sites within *D. melanogaster* to the site present in *D. simulans*. Table 4 shows Tajima's *D* and Fay and Wu's θ_H values for four-gene haplotypes in each

TABLE 4

Tajima and Fay and Wu tests

Chromosome	Sample	n	Tajima's D	$p(D)$	θ_H	$p(H)$
Tip of X	Massachusetts	10	0.628	0.778	1.956	0.355
	Crete	10	−1.388	0.108	3.733	0.049
	Zimbabwe	4	−0.212	0.577	2.667	0.262
	Total	24	−0.530	0.333	3.891	0.188
Fourth	Massachusetts	10	−0.279	0.157	0.156	0.767
	Crete	10	−1.112	0.410	0.022	0.703
	Zimbabwe	4	0.592	0.366	0.111	0.874
	Total	24	−0.781	0.465	1.756	0.406

Tajima's *D* and θ_H are based on a concatenated haplotype across four genes surveyed at the tip of the X or the fourth chromosome in the sample locality indicated. $p(D)$ and $p(H)$ refer to the probability that the observed Tajima's *D* or θ_H value differs from zero.

TABLE 5

Population structure in *D. melanogaster*

	Massachusetts	Crete
Tip of X		
Crete	0.16*/1.31	
Zimbabwe	0.45**/0.31	0.63**/0.15
Fourth		
Crete	0.41**/0.36	—
Zimbabwe	0.35*/0.63	0.57**/0.26

Each cell shows F_{st} values/$N_e m$ values calculated in DNAsp on the basis of Hudson *et al.*'s (1992) estimator of F_{st}. *5.0% and **0.01% significance levels for population subdivision are based on 1000 permutation tests of the sequences performed in the Arlequin software package (Excoffier 2002).

geographic sample, plus the associated probabilities. The Crete sample shows an excess of derived sites at high frequency at the tip of the X (due to sites in *RhoGAP1A* and *CG3038*; see Figure 2 and Table 4). The test is just significant at the 5% level and would drop below significance if corrections for multiple tests were made. However, our data from the Crete sample are consistent with the significant excess of derived sites at high frequency at the tip of the X chromosome in European samples of *D. melanogaster* (Fay and Wu 2000). The fourth chromosome data show no hint of nonneutral site frequencies (although the low polymorphism does not permit a powerful test).

Population structure: Table 5 shows F_{st} values among the three geographic samples on the basis of haplotypes from concatenated sequences for each exon sampled on the X or fourth chromosome. In general, the F_{st} values are high as expected for low recombination regions (Begun and Aquadro 1993; Charlesworth 1998). On the fourth chromosome the African sample is more differentiated ($F_{st} > 0.44$) than the Massachusetts and Crete samples ($F_{st} = 0.16$), consistent with other studies showing greater differentiation for African populations (Begun and Aquadro 1993, 1995; Caracristi and Schlötterer 2003). In contrast, the X chromosome shows a relatively high F_{st} value between Massachusetts and Crete (0.41), slightly higher than that of the Massachusetts-Zimbabwe comparison. One explanation for this pattern is a selective sweep(s) on the X in the Crete sample (see Table 4), which would elevate F_{st} values.

The F_{st} values for the X are not generally higher than those on the fourth, which would be expected under both (1) neutrality with lower effective population size for the X relative to the fourth and (2) selective sweeps on the X, which should accentuate population differentiation (Begun and Aquadro 1993; Stephan *et al.* 1998). While hitchhiking can decrease F_{st} between demes with sufficient gene flow (Stephan *et al.* 1998), genome-wide patterns of microsatellite differentiation (Caracristi and Schlötterer 2003) indicate that our Massachusetts-Crete pattern for the X chromosome is the outlier, further implying a sweep. Some of this chromosome-specific difference in F_{st} values may be attributed to sampling artifacts of partitioning low levels of total polymorphism (Charlesworth 1998). However, since the X and fourth chromosomes were extracted from the same 24 isofemale lines, sampling artifacts are reduced in this comparison.

Recombination: All four gametic phases are observed among the polymorphisms at the tip of the X and on the fourth chromosome (see Figures 2 and 3), suggesting that recombination has occurred in the sample of chromosomes we have studied. On the X, three pairs of sites show all four gametic phases: positions 833 in RhoGAP1A and 178 in CG3777(H10.1), positions 677 in CG3038(P7.1) and 178 in CG3777(H10.1), and positions 615 in *cinnamon* and 621 in CG3777(H10.1). The minimum number of recombination events is one, and DNAsp identifies this as between sites 677 in CG3037(P7.1) and 178 in CG3777(H10.1). One site implicated in recombination lies in RhoGAP1A at band 1A, suggesting that nonzero levels of recombination could extend into this distal region of the X chromosome. An accurate estimate of the per-nucleotide rate of recombination cannot be obtained from the current sample because the sites with all four gametic phases lie in separate genes, and thousands of base pairs separate each sampled exon. A lower-bound estimate can be obtained if we assume a minimum of one and a maximum of three recombination events spanning the ~1000 kb from RhoGAP1A to CG3777, giving 0.001–0.003 recombination events per kilobase. This is slightly lower than the lower-bound estimate obtained by Wang *et al.* (2002) for the fourth chromosome (six recombination events spanning 1156 kb = 0.0052/kb in a sample of 10 chromosomes).

Our sample of fourth chromosome exons shows all four gametic phases between site 802 in *pangolin* and 893 in *ATP synthase* (see Figure 3). These sites are separated by more than one-half the length of the fourth chromosome. Assuming 1 recombination event per ~600 kb, a lower bound of 0.0017 recombination events per kilobase is comparable to our lower-bound estimate from the tip of the X, but somewhat lower than the minimum estimates discussed above from Wang *et al.* (2002). Note that the Wang *et al.* (2002) sample included fewer alleles, but from a wider range of geographic locations. Nevertheless, these data provide independent confirmation of two recent reports documenting recombination on the fourth chromosome of Drosophila (Jensen *et al.* 2002; Wang *et al.* 2002). For reference, the Adh locus in a region of normal recombination on the second chromosome has an estimate of 1.84 recombination events per kilobase in a sample of 11 chromosomes (Hudson and Kaplan 1985 cited in Wang *et al.* 2002). This estimate is ~600–1800 times greater than our estimate for the tip of the X, ~1100 times greater than our estimate for the fourth chromosome, and 37–384 times that for the fourth chromosome estimated by Wang *et al.* (2002). These comparisons are rather crude since the sample sizes and geographic distribution of lines differ in each study. Nevertheless, we can assume that recombination on the fourth is reduced by about two orders of magnitude from regions of "normal" recombination and that recombination is lower at the tip of the X than on the fourth chromosome. Since there are few runs of polymorphisms in close proximity, it is difficult to determine the role of gene conversion in this sample, but other surveys near the tip of the X indicate that conversion does contribute to the overall rate of exchange (*cf.* Langley *et al.* 2000).

Polymorphism in *D. simulans*: Very low levels of polymorphism are observed at the tip of the X and on the fourth chromosome in our sample of *D. simulans* (see

TABLE 6

Polymorphism in *D. simulans*

Gene (bp)	S	R	Silent		Replacement	
			π	θ	π	θ
X chromosome						
RhoGAP1A, 936	0	0	0	0	0	0
CG3038, 840	0	0	0	0	0	0
cin, 759	1	0	0.00324	0.00259	0	0
CG3777, 1221	1	1	0.00141	0.00169	0.00064	0.00051
Mean			0.00116	0.00107	0.00016	0.00013
X haplotype	2	1	0.00114	0.00109	0.00021	0.00017
Fourth chromosome						
pan, 1092	1	1	0.00252	0.00202	0.00049	0.00059
zfh 2, 615	0	0	0	0	0	0
ATP, 1137	0	0	0	0	0	0
pho, 766	0	0	0	0	0	0
Mean			0.00063	0.00051	0.00012	0.00015
Fourth haplotype	1	1	0.00073	0.00059	0.00015	0.00018
X/fourth ratio of means			1.85	2.12	1.31	0.86
X/fourth ratio of haplotypes			1.56	1.85	0.00	0.94

Table 6). Silent polymorphisms at the tip of the X and the fourth chromosome are $\theta_{S\text{-tip-of-X}} = 0.00107$ and $\theta_{S\text{-fourth}} = 0.00051$. This represents >20-fold reduction of polymorphism on the X and >50-fold reduction on the fourth compared to the genome-wide average in worldwide samples of *D. simulans* ($\theta_{S\text{-X}} = 0.0234$, $\theta_{S\text{-Auto}} = 0.0276$; see Andolfatto 2001). Silent polymorphism is actually lower in *D. simulans* than in *D. melanogaster* for both low-recombination regions (X chromosome $\theta_{S\text{-mel}} = 0.00152$, $\theta_{S\text{-sim}} = 0.00107$; fourth chromosome, $\theta_{S\text{-mel}} = 0.00201$, $\theta_{S\text{-sim}} = 0.00051$). While these differences are certainly not significant, they are consistent with earlier studies that show greater reduction of polymorphism in the low-recombination regions of *D. simulans* than in *D. melanogaster* (Begun and Aquadro 1991, 1994).

The genome-wide average for the X/autosome ratio of silent polymorphism in worldwide samples of *D. simulans* is $\theta_{S\text{-X}}/\theta_{S\text{-A}} = 0.85$ (Andolfatto 2001). The tip-of-the-X/fourth chromosome ratio from the exon data reported here is $\theta_{S\text{-distal X}}/\theta_{S\text{-fourth}} = 2.12$ based on mean values from the four exons sampled on each chromosome (for concatenated haplotypes from the X and fourth chromosomes this ratio is $\theta_{S\text{-distal X}}/\theta_{S\text{-fourth}} = 1.85$). These ratios and the observations above suggest that the reduction of polymorphism has been more pronounced for the fourth chromosome than for the tip of the X, relative to the genome-wide average in *D. simulans*. The opposite pattern was observed in *D. melanogaster* (see above). This may reflect stronger background selection on the fourth chromosome in *D. simulans* (*cf.* Jensen *et al.* 2002).

Replacement polymorphism is also reduced at the tip of the X and on the fourth chromosome in *D. simulans*, relative to the genome-wide averages reported in Andolfatto (2001) (X average, $\theta_{R\text{-X}} = 0.0009$; tip-of-X, $\theta_{R\text{-distal X}} = 0.00013$; autosome average, $\theta_{R\text{-Auto}} = 0.0024$; fourth average, $\theta_{R\text{-fourth}} = 0.00015$). This represents an ~7-fold reduction at the tip of the X and a 16-fold reduction for the fourth chromosome, which are apparently smaller reductions than those for silent polymorphism (see above).

Tests of selection: The ratio of replacement to silent changes per site between species provides a simple test of deviation from neutrality. K_a/K_s ratios for the four exons at the tip of the X [mean = 0.114, 95% confidence interval (C.I.) = 0.056–0.203] and on the fourth chromosome (mean = 0.148, 95% C.I. = 0.095–0.285) are significantly <1.0, indicating a history of purifying selection ($P = 0.00$; *K*-estimator, Comeron 1999). Notably, these values are not higher than the genome-wide averages for the X and autosomes ($K_a/K_s = 0.186$ and 0.305, respectively; Betancourt *et al.* 2002). Higher K_a/K_s values might be expected for regions where effective purifying selection is reduced (*cf.* Kliman and Hey 1993; Comeron *et al.* 1999; Marais *et al.* 2001; Hey and Kliman 2002). However, as observed in these studies codon bias is reduced in our distal X and fourth chromosome exons [average effective numbers of codons (ENCs) for the distal X and fourth exons are 56.7 and 52.3, respectively]. This suggests that interference effects have relaxed selection on codon usage, but presumably have not weakened selection enough to allow accelerated amino acid evolution.

The main goal of this study was to compare MK tests

TABLE 7

McDonald-Kreitman tests for X and fourth chromosome loci

Gene	No. of sites	FR	FS	PR	PS	NI
		X chromosome				
RhoGAP1A	1011	3	28	2	0	—
CG3038	896	6	16	1	1	2.67
cin	761	13	23	2	3	1.18
CG3777	1225	7	13	3	3	1.86
Total X*	3893	29	80	8	7	3.15
		Fourth chromosome				
pan	1090	17	25	1	4	0.37
zfh 2	625	3	14	0	1	—
ATP	1142	0	23	0	1	—
pho	770	23	19	0	1	—
Total fourth	3627	43	81	1	7	0.27
		Fourth from Wang *et al.* (2002)				
cin	1075	21	23	1	0	—
pan	1090	17	25	1	4	0.37
zfh 2	625	3	14	0	1	0.00
ATP	1142	0	23	0	1	—
pho	770	23	19	0	1	0.00
CG11152	804	9	22	1	0	—
RfaBp	1033	19	33	1	0	—
plexA	828 (1074)	1	17	0	0	—
CG11153	669 (950)	15	14	0	1	0.00
btex	1081	11	22	0	1	0.00
CG1862	288 (965)	7	7	0	0	—
Total fourth	9405	126	219	4	9	0.77

*$P < 0.05$. FR, fixed replacement; FS, fixed silent; PR, polymorphic replacement; PS, polymorphic silent; NI, neutrality index; see MATERIALS AND METHODS.

from the tip of the X and fourth chromosome. These data are presented in Table 7. Using the combined data for polymorphism and divergence in the sample of 24 *D. melanogaster* and 5 *D. simulans* sequences, the tip-of-the-X data reject the neutral model of equal ratios of replacement and silent changes within and between species (G adjusted = 3.97; $P < 0.05$). On the fourth chromosome, the MK test does not reject neutrality (G adjusted = 1.81, $P > 0.5$). Wang *et al.* (2002) presented polymorphism and divergence data across the fourth chromosome, some of which covered protein-coding regions. If these data are added to our counts, the MK test becomes even less significant (G adjusted = 0.176).

The NI is calculated directly from MK test data and provides an estimator of the direction and degree of departure from the neutral expectation (Rand and Kann 1996). For the combined data at the tip of the X, NI = 3.15, implying $N_e s = -2.75$. For the fourth chromosome, NI = 0.27, indicating $N_e s > 5.0$. A 2×4 G-test of heterogeneity comparing the MK data from the tip of the X to the MK data from the fourth shows a significant difference ($G = 7.93$, d.f. = 3, $P = 0.0475$). If the Wang *et al.* (2002) data are included for the fourth, the result is closer to neutrality (NI = 0.77, $N_e s = +0.85$), and the 2×4 test between X and fourth data becomes more significant ($G = 17.4$, d.f. = 3, $P < 0.0006$).

Bustamante *et al.* (2002) developed a maximum-likelihood method to estimate historical selection coefficients from MK test data using a PRF approach. From the data in Table 7, the tip of the X data give $N_e s = -1.27$ ($P = 0.042$), and the fourth chromosome data give $N_e s = +2.88$, $P = 0.163$ (C. Bustamante, personal communication). These estimates of $N_e s$ from the PRF approach are less extreme than those from NI, but the latter index is a poor estimator of positive selection beyond $N_e s \sim 2.0$, where the curve describing the ratio of polymorphism to divergence becomes very flat (Nachman 1998; Weinreich and Rand 2000). In summary, the tip of the X chromosome shows a pattern of excess amino acid polymorphism indicating negative selection, while the fourth chromosome exons indicate positive selection, but cannot reject the neutral model.

NI *vs.* recombination rate: If reduced recombination relaxes effective selection, there should be a negative correlation between NI and recombination. Using available data from the Drosophila genome, the relationship appears to be negative, but is not significant. If our data are combined with those from Fay *et al.* (2002), the correlation between NI and recombination rate is −0.28 $r^2 = 0.08$, $P = 0.104$ for 31 genes, ignoring cases where NI = 0 or is undefined. Our distal X and fourth chromosome data were treated as single loci in this correlation, and recombination rate is the KH93 index cited in Hey and Kliman (2002). The relationship between recombination rate and gamma values ($= N_e s$) from Bustamante *et al.* (2002) plus the new data presented here is also negative but not significant ($n = 19$, $r = -0.21$, $r^2 = 0.04$, $P < 0.45$).

DISCUSSION

This study was motivated by the fact that Drosophila nuclear genes show an average excess of amino acid fixed differences in McDonald-Kreitman tests while mtDNA and Arabidopsis nuclear genes exhibit a consistent excess of amino acid polymorphisms, relative to divergence (Weinreich and Rand 2000). The lack of recombination in mtDNA and selfing in Arabidopsis create linkage conditions that should weaken natural selection and generate signatures of weak negative selection from MK tests (Nachman 1998; Rand and Kann 1998; Weinreich and Rand 2000; Rand 2001). Higher effective recombination in Drosophila nuclear proteins should break up linkage associations between beneficial and deleterious mutations, allowing positive selection to produce a consistent excess of replacement fixation events in MK tests (Weinreich and Rand 2000; Bustamante *et al.* 2002; Fay *et al.* 2002). However, low recombination and the opportunity for selection on recessive mutations are confounding differences in the contrasts between mtDNA, Arabidopsis, and Drosophila. The hap-

loid transmission of mtDNA and selfing in Arabidopsis also result in effectively higher levels of homozygosity than in Drosophila nuclear genes, allowing for recessive mutations to be exposed to selection. The goal of this study was to provide additional coding sequence data from low-recombination regions of the semihaploid X and the diploid fourth chromosomes that bear on the interaction effects of recombination and dominance on the efficiency of selection on amino acid variation.

The data confirm that silent polymorphism is very low both at the tip of the X and on the fourth chromosome, but further show a significant excess of replacement polymorphism at the tip of the X. This is supported both by MK tests and by HKA tests of replacement sites comparing the X and fourth chromosomes. Both the tip of the X and the fourth chromosome show average levels of replacement and silent divergence (and hence K_a/K_s ratios). Thus, the tip of the X shows a significant signature of negative selection like that of mtDNA and Arabidopsis, while the fourth chromosome shows a nonsignificant signature of positive selection, like that of the Drosophila genomic average (Weinreich and Rand 2000; Bustamante *et al.* 2002; Fay *et al.* 2002). To account for these data we need to invoke forces that generally reduce silent polymorphism in regions of low recombination, but allow for the maintenance of amino acid polymorphism at the tip of the X chromosome.

Selective sweeps on the X? Aquadro *et al.* (1994) showed that levels of polymorphism were lower on the X than on the third chromosome in *D. melanogaster*, even after correcting for differences in effective population size. These data were taken as support for hitchhiking under the assumption that recessive advantageous alleles will be more likely to sweep through the population on the hemizygous X chromosome. In *D. simulans*, lower levels of polymorphism are found on the X relative to the third chromosome, consistent with the hitchhiking hypotheses (Begun and Whitley 2000). Other studies of site frequencies near the tip of the X in *D. melanogaster* suggest that hitchhiking is a better fit to the data than background selection (Fay and Wu 2000; Langley *et al.* 2000), as do our data for the Crete sample (Table 5). Can a history of hitchhiking on the X account for our excess of amino acid polymorphism?

It is difficult to attribute the current data to hitchhiking alone. First, in regions of normal recombination on the X there is a deficiency of amino acid polymorphism, relative to the autosomes, suggesting more efficient elimination of recessive deleterious replacement polymorphism (Andolfatto 2001; Fay *et al.* 2002). If hitchhiking of advantageous mutations is occurring on the X, these events are presumably short-lived relative to the persistent elimination of recessive deleterious replacement polymorphisms. These opposing forms of selection should be operating at different time scales if the advantageous mutation rate is considerably lower than the deleterious mutation rate. Under this scenario, an excess of amino acid polymorphism at the tip of the X might be explained by a relaxation of effective selection due to the tight linkage in this region, while occasional hitchhiking events would reduce levels of silent polymorphism. Two predictions that should follow from this scenario are that replacement polymorphisms would be at low frequency as the population recovers from sweeps, and amino acid divergence should be elevated due to the relaxed selection from interference effects (*e.g.*, Comeron and Kreitman 2002). These predictions are not upheld by the data, as some replacement polymorphisms are at intermediate frequency in the sample (Figure 2), and K_a at the tip of the X is slightly lower than the genome-wide average (see results).

Background selection on the fourth? If amino acid polymorphisms are recessive and deleterious, an excess of replacement polymorphism might be expected under background selection on the diploid fourth chromosome. The lack of a significant skew of the frequency spectrum on the fourth chromosome is consistent with background selection (*e.g.*, Jensen *et al.* 2002), which should further reduce effective selection against such replacement polymorphisms. If background selection has been a dominant force reducing variation on the fourth chromosome, it has not distinguished between silent and replacement sites as the reduction in polymorphism has been very similar for these two functional classes (Table 2) relative to autosome averages (*cf.* Andolfatto 2001).

The effects of background selection, or more generally interference selection, can be modulated by differences in gene length and intron position; these in turn interact with coefficients of selection and recombination in affecting levels of neutral and selected polymorphism (Comeron and Kreitman 2002). The exons we surveyed on the X and fourth chromosomes are all uninterrupted exons of similar size (Table 1, Figure 1), suggesting that an exon-length effect does not underlie the difference in replacement polymorphism between the X and fourth chromosome proteins surveyed. Thus, under the simplifying assumptions that coefficients of mutation, selection, recombination, and dominance are similar for new mutations on the distal X and fourth chromosomes, the data presented here are opposite in direction from the predictions of hitchhiking and background selection models. However, there is evidence to suggest that that these evolutionary forces do differ between the distal X and the fourth chromosome (*e.g.*, Wang *et al.* 2002; see also Malcolm *et al.* 2003). Our data may in fact be evidence for such differences in evolutionary forces rather than a poor fit to simplifying assumptions about selection in low-recombination environments.

Variation in recombination, dominance, and the strength of selection: If recombination rates (and other forces reducing effective population size) are suffi-

ciently different between the tip of the X and the fourth chromosome, the difference in effective haploidy could be of limited significance for the expected ratios of silent and replacement variation. Our sequence data do show that recombination (or gene conversion) does occur across the tip of the X and the fourth chromosome, but the rate of exchange appears lower at the tip of the X compared to the fourth (see RESULTS above and WANG *et al.* 2002). The weakening of natural selection due to Hill-Robertson interference and reduced effective population size may be more complete at the tip of the X than on the fourth. This could override any difference in effective selection due to dominance. Under an additive model of selection, the rate of fixation may be increased approximately twofold in a haploid *vs.* a diploid environment (CROW and KIMURA 1970). If the recombination rate is as much as 10 times lower at the tip of the X compared to the fourth, this could render both recessive advantageous and deleterious amino acid mutations effectively neutral, despite the semihaploid environment.

Such a scenario could accommodate the apparently conflicting evidence for (1) hitchhiking on the X and the reduction of silent polymorphism (*e.g.*, AQUADRO *et al.* 1994; BEGUN and WHITLEY 2000), (2) a deficiency of replacement polymorphism in regions of normal recombination on the X, and (3) the excess of replacement polymorphism at the tip of the X. In coding regions ~75% of new mutations will introduce amino acid changes (LI and GRAUR 1991). As interference or effective bottlenecks reduce selection, more amino acid polymorphisms can drift into the population. When occasional selective sweeps move through the population, there would be effectively more "neutral hitchhikers" segregating at the tip of the X to be carried to detectable frequency in sequencing surveys.

The contrasting patterns of replacement polymorphism on the distal X and fourth can also be accommodated by invoking differences in the average selection or dominance coefficients. If amino acid mutations are more deleterious and dominant on the fourth than on the distal X, an excess of amino acid polymorphism in the latter would be expected from this contrast. While deferring to differences in selection and dominance is *ad hoc*, it may be more biologically realistic than assuming that these factors are the same across chromosomes.

The results reported here motivate additional empirical and theoretical studies of the interaction of selection, linkage, and dominance. Surveys of exons at the base and tip of the X and other autosomes could determine whether excess replacement polymorphism is indeed restricted to the semihaploid X. Moreover, it has been shown that background selection and hitchhiking have interacting effects on linked neutral polymorphism when operating together (KIM and STEPHAN 2000). The effects of interference selection on ratios of selected and neutral variation have received less attention (MCVEAN and CHARLESWORTH 2000; COMERON and KREITMAN 2002), and the data reported here suggest that additional interactions with levels of dominance should be explored. The small counts for silent and replacement polymorphisms in these low-recombination regions compromise the power of tests based on site frequencies (FAY and WU 2000) or different functional classes of amino acids (RAND *et al.* 2000; WYCKOFF *et al.* 2000). Larger samples of coding and noncoding sequences spanning the gradients from low to normal recombination across the Drosophila genome could provide a straightforward means of testing the complex problem of how recombination and dominance modulate the strength of natural selection.

We thank C. Aquadro, D. Begun, C. Bustamante, B. Charlesworth, C. Langley, M. Long, K. Thornton, and two anonymous reviewers for comments and suggestions on aspects of this study, which was supported by grants DEB-9981497 and INT-9981452 from the National Science Foundation.

LITERATURE CITED

ADAMS, M. D., S. E. CELNIKER, R. A. HOLT, C. A. EVANS, J. D. GOCAYNE *et al.*, 2000 The genome sequence of Drosophila melanogaster. Science **287:** 2185–2195.

AGUADÉ, M., N. MIYASHITA and C. H. LANGLEY, 1989 Reduced variation in the *yellow-achete-scute* region in natural populations of *Drosophila melanogaster*. Genetics **122:** 607–615.

AKASHI, H., 1995 Inferring weak selection from patterns of polymorphism and divergence at "silent" sites in Drosophila DNA. Genetics **139:** 1067–1076.

ANDOLFATTO, P., 2001 Contrasting patterns of X-linked and autosomal nucleotide variation in Drosophila melanogaster and Drosophila simulans. Mol. Biol. Evol. **18:** 279–290.

AQUADRO, C. F., D. J. BEGUN and E. C. KINDAHL (Editors), 1994 *Selection, Recombination, and DNA Polymorphism in Drosophila*. Chapman & Hall, New York.

ASHBURNER, M., 1989 *Drosophila: A Laboratory Handbook*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

BALLARD, J. W. O., 2000 Comparative genomics of mitochondrial DNA in members of the *Drosophila melanogaster* subgroup. J. Mol. Evol. **51:** 48–63.

BEGUN, D. J., 1996 Population genetics of silent and replacement variation in Drosophila simulans and D. melanogaster: X/autosome differences? Mol. Biol. Evol. **13:** 1405–1407.

BEGUN, D. J., and C. F. AQUADRO, 1991 Molecular population genetics of the distal portion of the *X* chromosome in *Drosophila*: evidence for genetic hitchhiking of the *yellow-achaete* region. Genetics **129:** 1147–1158.

BEGUN, D. J., and C. F. AQUADRO, 1992 Levels of naturally occuring DNA polymorphism correlate with recombination rates in *D. melanogaster*. Nature **356:** 519–520.

BEGUN, D. J., and C. F. AQUADRO, 1993 African and North American populations of *Drosphila melanogaster* are very different at the DNA level. Nature **356:** 519–520.

BEGUN, D. J., and C. F. AQUADRO, 1994 Evolutionary inferences from DNA variation at the 6-phosphogluconate dehydrogenase locus in natural populations of Drosophila: selection and geographic differentiation. Genetics **136:** 155–171.

BEGUN, D. J., and C. F. AQUADRO, 1995 Molecular variation at the vermilion locus in geographically diverse populations of *Drosophila melanogaster* and *D. simulans*. Genetics **140:** 1019–1032.

BEGUN, D. J., and P. WHITLEY, 2000 Reduced X-linked nucleotide polymorphism in Drosophila simulans. Proc. Natl. Acad. Sci. USA **97:** 5960–5965.

BERRY, A. J., J. W. AJIOKA and M. KREITMAN, 1991 Lack of polymorphism on the *Drosophila* fourth chromosome resulting form selection. Genetics **129:** 1111–1117.

Betancourt, A. J., D. C. Presgraves and W. J. Swanson, 2002 A test for faster X evolution in Drosophila. Mol. Biol. Evol. **19:** 1816–1819.

Bustamante, C. D., R. Nielsen, S. A. Sawyer, K. M. Olsen, M. D. Purugganan *et al.*, 2002 The cost of inbreeding in Arabidopsis. Nature **416:** 531–534.

Caracristi, G., and C. Schlötterer, 2003 Genetic differentiation between American and European *Drosophila melanogaster* populations could be attributed to admixture of African alleles. Mol. Biol. Evol. **20:** 792–799.

Charlesworth, B., 1996 Background selection and patterns of genetic diversity in *Drosophila melanogaster*. Genet. Res. **68:** 131–149.

Charlesworth, B., 1998 Measures of divergence between populations and the effect of forces that reduce variability. Mol. Biol. Evol. **15:** 538–543.

Charlesworth, B., M. T. Morgan and D. Charlesworth, 1993 The effect of deleterious mutations on neutral molecular variation. Genetics **134:** 1289–1303.

Comeron, J., 1999 K-estimator: calculation of the number of nucleotide substitutions per site and the confidence intervals. Bioinformatics **15:** 763–764.

Comeron, J., and M. Kreitman, 2002 Population, evolutionary and genomic consequences of interference selection. Genetics **161:** 389–410.

Comeron, J., M. Kreitman and M. Aguade, 1999 Natural selection on synonymous sites is correlated with gene length and recombination in Drosophila. Genetics **151:** 239–249.

Crow, J. F., and M. Kimura, 1970 *An Introduction to Population Genetics Theory*. Harper & Row, New York.

Excoffier, L., 2002 *Arlequin: Software for Population Genetics Analysis*. (http://lgb.unige.ch/arlequin/).

Fay, J. C., and C.-I Wu, 2000 Hitchhiking under positive Darwinian selection. Genetics **155:** 1405–1413.

Fay, J. C., G. J. Wyckoff and C.-I Wu, 2002 Testing the neutral theory of molecular evolution with genomic data from Drosophila. Nature **415:** 1024–1026.

Felsenstein, J., 1974 The evolutionary advantage of recombination. Genetics **78:** 737–756.

Fu, Y.-X., and W.-H. Li, 1993 Statistical tests of neutrality of mutations. Genetics **133:** 693–709.

Gillespie, J. H., 1997 Junk ain't what junk does: neutral alleles in a selected context. Gene **205:** 291–299.

Gillespie, J. H., 2000 Genetic drift in an infinite population: the pseudohitchhiking model. Genetics **155:** 909–919.

Gloor, G., and W. R. Engels, 1991 Single fly DNA preps for PCR. Dros. Inf. Newsl. **1** (flybase.bio.indiana.edu/docs/news/DIN/dinvol1.txt).

Hey, J., and R. M. Kliman, 2002 Interactions between natural selection, recombination and gene density in the genes of Drosophila. Genetics **160:** 595–608.

Hill, W. G., and A. Robertson, 1966 The effect of linkage on limits of artificial selection. Genet. Res. **8:** 269–294.

Hudson, R. R., and N. L. Kaplan, 1985 Statistical properties of the number of recombination events in the history of a sample of DNA sequences. Genetics **111:** 147–164.

Hudson, R. R., M. Kreitman and M. Aguade, 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, 1992 A statistical test for detecting geographic subdivision. Mol. Biol. Evol. **9:** 138–151.

Jensen, M. A., B. Charlesworth and M. Kreitman, 2002 Patterns of genetic variation at a chromosome 4 locus of *Drosophila melanogaster* and *D. simulans*. Genetics **160:** 493–507.

Kaplan, N. L., R. R. Hudson and C. H. Langley, 1989 The "hitchhiking effect" revisited. Genetics **123:** 887–899.

Kim, Y., and W. Stephan, 2000 Joint effects of genetic hitchhiking and background selection on neutral variation. Genetics **155:** 1415–1427.

Kimura, M., 1983 *The Neutral Theory of Molecular Evolution*. Cambridge University Press, Cambridge, UK.

Kliman, R. M., and J. Hey, 1993 Reduced natural selection associated with low recombination in *Drosophila melanogaster*. Mol. Biol. Evol. **10:** 1239–1258.

Langley, C. H., B. P. Lazzaro, W. Phillips, E. Heikkinen and J. M. Braverman, 2000 Linkage disequilibria and the site frequency spectra in the su(s) and su(w(a)) regions of the *Drosophila melanogaster* X chromosome. Genetics **156:** 1837–1852.

Li, W.-H., and D. Graur, 1991 *Fundamentals of Molecular Evolution*. Sinauer Associates, Sunderland, MA.

Malcolm, C. M., G. J. Wyckoff and B. T. Lahn, 2003 Genic mutation rates in mammals: local similarity, chromosomal heterogeneity, and X-versus-autosome disparity. Mol. Biol. Evol. **20:** 1633–1641.

Marais, G., D. Mouchiroud and L. Duret, 2001 Does recombination improve selection on codon usage? Lessons from nematode and fly complete genomes. Proc. Natl. Acad. Sci. USA **98:** 5688–5692.

Maynard Smith, J., and J. Haigh, 1974 The hitch-hiking effect of a favourable gene. Genet. Res. **23:** 23–35.

McDonald, J. H., and M. Kreitman, 1991 Adaptive protein evolution at the *Adh* locus in *Drosophila*. Nature **351:** 652–654.

McVean, G. A., and B. Charlesworth, 2000 The effects of Hill-Robertson interference between weakly selected mutations on patterns of molecular evolution and variation. Genetics **155:** 929–944.

Nachman, M. W., 1998 Deleterious mutations in animal mitochondrial DNA. Genetica **102/103:** 61–69.

Nachman, M. W., 2001 Single nucleotide polymorphisms and recombination rate in humans. Trends Genet. **17:** 481–485.

Payseur, B. A., and M. W. Nachman, 2000 Microsatellite variation and recombination rate in the human genome. Genetics **156:** 1285–1298.

Rand, D. M., 2001 The units of selection on mitochondrial DNA. Annu. Rev. Ecol. Syst. **32:** 415–448.

Rand, D. M., and L. M. Kann, 1996 Excess amino acid polymorphism in mitochondrial DNA: contrasts among genes from *Drosophila*, mice, and humans. Mol. Biol. Evol. **13:** 735–748.

Rand, D. M., and L. M. Kann, 1998 Mutation and selection at silent and replacement sites in the evolution of animal mitochondrial DNA. Genetica **102/103:** 393–407.

Rand, D. M., D. M. Weinreich and B. O. Cezairliyan, 2000 Neutrality tests of conservative-radical amino acid changes in nuclear- and mitochondrially-encoded proteins. Gene **261:** 115–125.

Rozas, J., and R. Rozas, 1999 DnaSP version 3: an integrated program for molecular population genetics and molecular evolution analysis. Bioinformatics **15:** 174–175.

Sawyer, S. A., and D. L. Hartl, 1992 Population genetics of polymorphism and divergence. Genetics **132:** 1161–1176.

Simonsen, K. L., G. A. Churchill and C. F. Aquadro, 1995 Properties of statistical tests of neutrality for DNA polymorphism data. Genetics **141:** 413–429.

Sokal, R. R., and F. J. Rohlf, 1981 *Biometry*, Ed. 2. W. H. Freeman, San Francisco.

Solignac, M., M. Monnerot and J.-C. Mounolou, 1986 Mitochondrial DNA evolution in the *melanogaster* species subgroup of *Drosophila*. J. Mol. Evol. **23:** 31–40.

Stephan, W., and C. H. Langley, 1998 DNA polymorphism in lycopersicon and crossing-over per physical length. Genetics **150:** 1585–1593.

Stephan, W., L. Xing, D. A. Kirby and J. M. Braverman, 1998 A test of the background selection hypothesis based on nucleotide data from Drosophila ananassae. Proc. Natl. Acad. Sci. USA **10:** 5649–5654.

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

Wang, W., K. Thornton, A. Berry and M. Long, 2002 Nucleotide variation along the Drosophila melanogaster fourth chromosome. Science **295:** 134–137.

Weinreich, D. M., and D. M. Rand, 2000 Contrasting patterns of non-neutral evolution in proteins encoded in nuclear and mitochondrial genomes. Genetics **156:** 385–399.

Wyckoff, G. J., W. Wang and C.-I Wu, 2000 Rapid evolution of male reproductive genes in the descent of man. Nature **403:** 304–309.

Communicating editor: W. Stephan