Molecular Evolution of Two Linked Genes, *Est-6* and *Sod*, in *Drosophila melanogaster*

Evgeniy S. Balakirev,*,†,‡ Elena I. Balakirev,* Francisco Rodríguez-Trelles*,§ and Francisco J. Ayala*

**Department of Ecology and Evolutionary Biology, University of California, Irvine, California 92697-2525, †Institute of Marine Biology, Vladivostok 690041, Russia, ‡Department of General Biology, Ecology and Soils, Far Eastern State University, Vladivostok 690600, Russia and §Departament de Genética, Universitat Autónoma de Barcelona, 08193 Bellaterra (Barcelona), Spain*

Manuscript received June 7, 1999
Accepted for publication July 22, 1999

ABSTRACT

We have obtained 15 sequences of *Est-6* from a natural population of *Drosophila melanogaster* to test whether linkage disequilibrium exists between *Est-6* and the closely linked *Sod*, and whether natural selection may be involved. An early experiment with allozymes had shown linkage disequilibrium between these two loci, while none was detected between other gene pairs. The *Sod* sequences for the same 15 haplotypes were obtained previously. The two genes exhibit similar levels of nucleotide polymorphism, but the patterns are different. In *Est-6*, there are nine amino acid replacement polymorphisms, one of which accounts for the S-F allozyme polymorphism. In *Sod*, there is only one replacement polymorphism, which corresponds to the S-F allozyme polymorphism. The transversion/transition ratio is more than five times larger in *Sod* than in *Est-6*. At the nucleotide level, the S and F alleles of *Est-6* make up two allele families that are quite different from each other, while there is relatively little variation within each of them. There are also two families of alleles in *Sod*, one consisting of a subset of F alleles, and the other consisting of another subset of F alleles, designed F(A), plus all the S alleles. The *Sod* F(A) and S alleles are completely or nearly identical in nucleotide sequence, except for the replacement mutation that accounts for the allozyme difference. The two allele families have independent evolutionary histories in the two genes. There are traces of statistically significant linkage disequilibrium between the two genes that, we suggest, may have arisen as a consequence of selection favoring one particular sequence at each locus.

THE understanding of the genome as an aggregate of relatively independent genes (bean-bag genetics) was a feature of the classical period of genetics. Gene interaction (epistasis) played a primary role in the theory of evolution starting in the 1920s (Wright 1931; Dobzhansky 1937; Schmalhausen 1946; Mather 1953; Waddington 1957; Mayr 1963). Linkage disequilibrium and nonrandom associations between alleles or groups of nucleotides may indicate epistatic relationships, and much empirical work has been devoted over several decades to demonstrate that linkage disequilibrium occurs between nonallelic genes. The issue of gene interaction is also important in connection with the long-lasting neutralist-selectionist controversy that has generated much research in population genetics for nearly 30 years (Kimura 1968, 1983; Kimura and Ohta 1971; Ayala *et al.* 1971, 1972a,b; Lewontin 1974; and many others). Linkage disequilibrium is often considered strong evidence supporting the selectionist position, especially if its pattern is consistent between populations (Lewontin 1974).

Corresponding author: Francisco J. Ayala, Department of Ecology and Evolutionary Biology, 321 Steinhaus Hall, University of California, Irvine, CA 92697-2525. E-mail: fjayala@uci.edu

The evidence for linkage disequilibrium between individual loci remains scarce, except when genes are very closely linked or associated with chromosomal inversions (reviewed by Langley 1977; Hedrick *et al.* 1978; Barker 1979; Krimbas and Powell 1993). In the cases when significant associations have been detected, it is often far from clear whether they are caused by nonrandom haplotype sampling, random genetic drift, or natural selection (Mukai *et al.* 1974; Mukai and Voelker 1977). Significant disequilibrium can indeed arise without epistasis as a result of random genetic drift within a given population (Hill and Robertson 1968; Ohta and Kimura 1969a,b; Hill 1975, 1976), in subdivided populations (Nei and Li 1973; Li and Nei 1974; Feldman and Christiansen 1975; Ohta 1982a,b), and by founder effects (Avery and Hill 1979).

Numerous examples of significant linkage disequilibrium have been discovered in Drosophila between specific allozymes and chromosomal inversions, which have been interpreted as reflecting selection for favored multilocus allele combinations (Prakash and Lewontin 1968; Prakash 1974; Zouros 1976; Voelker *et al.* 1978). Ishii and Charlesworth (1977) and Nei and Li (1980) have, however, shown that nonrandom associations between allozymes and inversions can be ex-

plained by absence or limited recombination without selection (as a consequence of chance and insufficient time for associations generated by mutation to decay by recombination and gene conversion). The inference from the allozyme studies is, therefore, that linkage disequilibrium is mostly associated with closely linked genes, but may involve distantly linked genes when special cytological mechanisms (polymorphic inversions) allow it to exist. Allozyme loci that can recombine freely exhibit little, if any, linkage disequilibrium. Failure to detect disequilibrium may, of course, be a consequence of the limited statistical power of the tests to detect it (Brown 1975), which might be overcome by using larger sample sizes and combining probabilities from independent tests (Brown 1975; Zapata and Alvarez 1992, 1993).

The introduction of DNA sequencing and other molecular techniques in population studies makes it possible to gain considerable information about linkage disequilibrium. Nonrandom associations have been detected between polymorphic sites of *Adh, Adh-Dup* (Schaeffer and Miller 1993), and *Est-5B* (Veuille and King 1995) in *Drosophila pseudoobscura; vermilion* (Begun and Aquadro 1995) and *G6pd* (Eanes *et al.* 1996) in *D. simulans*; and the following in *D. melanogaster*: *Adh* and *Adh-Dup* (Kreitman and Hudson 1991), *vermilion* (Begun and Aquadro 1995), *Pgd* (Begun and Aquadro 1994), *white* (Kirby and Stephan 1995, 1996), *G6pd* (Eanes *et al.* 1996), *dpp* (Richter *et al.* 1997), *Acp70A* (Cirera and Aguadé 1997), *period* (Rosato *et al.* 1997), and *Acp26Aa* and *Acp26Ab* (Aguadé 1998). But, as in the case with allozymes, a positive correlation between the presence of chromosome inversions and the extent of linkage disequilibrium is often the case (Aquadro 1993, and references therein). Nevertheless, DNA linkage disequilibrium between nucleotide sites is well established in some cases, as well as the fact that it reflects epistatic relationships (Kirby *et al.* 1995; Kirby and Stephan 1996), although the nature of the epistatic interactions between genes remains enigmatic.

In this article, we investigate the nucleotide polymorphisms in the *Est-6* and *Sod* genes of *D. melanogaster* and compare them in the two genes, seeking to identify processes that contribute to the polymorphisms. We test whether linkage disequilibrium may occur between these two fairly closely linked genes, as has been intimated by the results of Smit-McBride *et al.* (1988), who investigated natural and laboratory populations and detected linkage disequilibrium between the allozyme polymorphisms of *Sod* and *Est-6*, but not between any other gene pairs. Hudson *et al.* (1994, 1997) have, moreover, shown that a selective sweep has recently occurred involving a many-kilobases-long region that includes the *Sod* gene. Est-*6* and *Sod* are closely linked on the left arm of chromosome 3 of *D. melanogaster*, ~1000 kb apart (Procunier *et al.* 1991; Heino *et al.* 1994; Hartl and Lozovskaya 1995). The two genes are examined in a set of 15 haplotypes from a natural population in California.

MATERIALS AND METHODS

Drosophila strains: The 15 *D. melanogaster* strains were derived from wild flies collected by F. J. Ayala (October 1991) in El Rio Vineyard (Acampo, CA). The strains were made fully homozygous for the third chromosome by means of crosses with balancer stocks, as described by Seager and Ayala (1982). The homozygous strains were named in accordance with the Cu,Zn superoxide dismutase (SOD) electrophoretic allele they carry, Fast (F) or Slow (S), as follows: 255S, 438S, 510S, 521S, 94F, 174F, 377F, 483F, 498F, 521F, 565F, 581F, 968F, 517S, and 357F. The *Sod* gene sequence of these strains has been investigated previously in our laboratory (Hudson *et al.* 1994, 1997).

Allozyme electrophoresis: Twenty flies from each *D. melanogaster* strain were homogenized in 20 μl 0.1 m Tris-HCl buffer, pH 8.0. The homogenates were electrophoresed for 8–9 hr using a Tris-borate-EDTA continuous buffer system, pH 8.6 (Markert and Faulhaber 1965) at 100 V in a 11% starch gel. The gels were stained for esterase-6 (EST-6) activity according to standard procedures (Manchenko 1994). See Ayala *et al.* (1972b) for additional details.

DNA extraction, amplification, and sequencing: Total genomic DNA was extracted using the procedure described by Palumbi *et al.* (1991).

We used the *Est-6* sequence, previously published by Collet *et al.* (1990), for designing PCR and sequence primers. The amplified fragments (2017 bp long) included 56 bp of the 5′-flanking region, the *Est-6* gene, the intergenic region, and 82 bp of the ψ*Est-6* gene (Figure 1). The ψ*Est-6* gene has been referred to in the literature as *Est-P*, but the evidence indicates that it is a pseudogene (Balakirev and Ayala 1996).

The two primers used for the PCR amplification reactions (1 and 2, Figure 1) were 5′-gcaattgccgcatctcaagatagt-3′ (forward primer) and 5′-caacaatcaagggatcagcttcag-3′ (reverse primer). All PCR reactions were carried out, as described by Kwiatowski *et al.* (1991), in final volumes of 100 μl containing 40 μm each of dNTP, 2.5 units of AmpliTaq DNA Polymerase (Perkin Elmer, Norwalk, CT), 0.2 μm each of forward and reverse primers, buffer (Perkin Elmer) at a final concentration of 10 mm Tris-HCl (pH 8.3), 50 mm KCl, 1.5 mm $MgCl_2$, and ~50 ng of template (total genomic) DNA. The mixtures were overlaid with mineral oil, placed in a DNA thermal cycler (Perkin Elmer), incubated 5 min at 95°, and subjected to multiple cycles of denaturation, annealing, and extension under the following conditions: 95° for 1 min, 56° for 1 min, and 72° for 2 min (for the first cycle and progressively adding 3 sec at 72° for every subsequent cycle). After 30 cycles, a final 7-min extension period at 72° concluded each amplification reaction. Samples were stored at 4° for several hours or at −20° for up to 2 wk.

One-tenth of each reaction volume was assayed on a 0.8% agarose gel. If the desired PCR product was detected, the remainder volume of the reaction was purified with the Wizard PCR preps DNA purification system (Promega, Madison, WI). The purified PCR product was directly sequenced by the dideoxy chain termination technique (Sanger *et al.* 1977) using Dye Terminator chemistry and separated with the ABI PRISM 377 automated DNA sequencer (Perkin Elmer). For each line, the sequences of both strands were determined. Eight internal primers spaced on average 350 nucleotides (a–d for the 5′-to-3′ direction and a′–d′ for the 3′-to-5′ direction; Figure 1) were used for sequencing the *Est-6*: (a) 5′-ccaacaaattggtagga gagg-3′, (b) 5′-aactatggactgaaagatcaacg-3′, (c) 5′-ctgtattggagc

catcggatg-3′, (d) 5′-gatcttcatcgcaaatatgg-3′, (a′) 5′-tcacgcat cacgttctcgtgtcc-3′, (b′) 5′-gcgtctggagcatccgatggctcc-3′, (c′) 5′-tcgaatatcaaaaagtagtcgtc-3′, and (d′) 5′-aaattccacatgctgcctattct-3′. The *Est-6* sequences reported in this article have been deposited in the GenBank sequence database library under accession nos. AF147095–AF147102.

For the *Sod* gene, we analyzed a region (1408 bp long) that included 43 bp of exon I, the intron (725 bp), exon II (396 bp), and 244 bp of the 3′ flanking region. The DNA preparation, amplification, cloning, and sequencing procedures for the *Sod* gene are described by Hudson *et al.* (1994, 1997).

DNA sequence analysis: All primers were designed using the computer program DNASIS for Windows (1994, Hitachi Software Engineering), which allows one to check the secondary structure of primers. Multiple alignment was carried out manually, using the program DARWIN (elaborated by Robert Tyler from our laboratory), and automatically, using the program CLUSTAL W (Thompson *et al.* 1994). Two DNA sequences were obtained from the GenBank database: *D. melanogaster*, *Est-6* (accession no. M33780), and *D. simulans*, *Sod* (accession no. X15685). The maximum parsimony analysis of the *Est-6* and *Sod* sequences was performed with the computer program PAUP (Swofford 1993). The computer program DNASP (Rozas and Rozas 1997) was used to analyze the data by means of a "sliding window" (Hudson and Kaplan 1988) and was used for most intraspecific analyses.

Linkage disequilibrium analysis: Linkage disequilibrium within and between *Est-6* and *Sod* was evaluated using Fisher's exact test and chi-square test for independence between sites. Singleton polymorphisms (mutations appearing in only one sampled allele) were omitted. Lewontin's sign test, elaborated especially for molecular sequence data (Lewontin 1995), was also used to test the significance of linkage disequilibrium between *Est-6* and *Sod*, including singletons. Lewontin (1995) has shown that Fisher's exact test is not sensitive in cases with very asymmetrical distribution of allele frequencies, a situation that is commonly observed in molecular sequence data, including our data. In such a situation, there is, for numerical reasons, very low probability of obtaining a significant test by means of the usual two-by-two contingency tables (Fisher's exact test or chi-square test). To overcome this problem, Lewontin's (1995) test is based on the distribution of the disequilibrium sign ("sign" test), which allows the analysis of asymmetrical allele frequency data to make inferences about overall linkage disequilibrium. The procedure involves examining the number of positive and negative *D* values for each polymorphic site within and between all types of pairwise comparisons (singletons vs. singletons, singletons *vs.* doublets, doublets *vs.* doublets, and so on; Lewontin 1995). The observed value, summed for a particular type of pairwise comparison, is compared with the expected value using a goodness-of-fit test (the likelihood ratio statistic, *G*-test).

RESULTS

The genes *Est-6* and *Sod* are on the left arm of chromosome 3 of *D. melanogaster*, genetically mapped at 35.9 and 32.5 and located at 69A1-A5 and 68A8-A9 on the polytene chromosomes, respectively (Heino *et al.* 1994).

Allozyme polymorphism: We have analyzed *Est-6* in 15 *D. melanogaster* strains, fully homozygous for the third chromosome, derived from flies collected in the El Rio Vineyard in California. The strains were chosen because their *Sod* gene sequences have been previously investigated in our laboratory (Hudson *et al.* 1994, 1997) and may be considered as randomly selected with respect to their *Est-6* alleles. Electrophoretic analysis reveals two common EST-6 allozymes: Fast (357F and 517S strains) and Slow (255S, 438S, 510S, 521S, 94F, 174F, 377F, 483F, 498F, 521F, 565F, 581F, and 968F strains). (The S or F in each strain's designation refers to the SOD allele it carries; see materials and methods.)

Nucleotide polymorphism: The organization of the *Est-6* gene is outlined in Figure 1. The sequenced region is 1879 bp long, comprising the *Est-6* gene and the 193 bp of the intergenic region between *Est-6* and ψ*Est-6*. Table 1 displays the *Est-6* polymorphisms observed in the 15 *D. melanogaster* strains. There are 35 nucleotide polymorphic sites, 9 of which yield amino acid replacements.

Table 2 gives the values of π and θ for the two genes, *Est-6* and *Sod*. The data for *Sod* are obtained from Hudson *et al.* (1997). The overall nucleotide diversity is somewhat larger for *Sod*. However, the diversity is about the same for the coding or the noncoding regions considered separately (*Sod* has greater overall diversity because the noncoding region is larger). The diversity, as expected, is much greater in the noncoding than in the coding regions, reflecting the usual higher level of functional constraint in the coding regions of genes. The distribution of polymorphic sites among the *Sod* haplotypes is extremely nonuniform: 69.6% of the *Sod* polymorphic sites are introduced by just two haplotypes (498F and 968F; see Hudson *et al.* 1994, 1997). It is also quite polarized in the case of *Est-6*, with 62.9% of the polymorphic sites resulting from two haplotypes, 517S and 357F, which are the only two *Est-6 F* alleles in the sample (shown at the bottom of Table 1). When these two pairs of sequences are removed, the π and θ values considerably decrease for both genes; but in the coding region, the decrease is somewhat greater for *Est-6* (Table 2). This difference may be due to different selective constraints in the two genes, but it may also reflect the different histories of the particular sets of alleles included in our analysis (see discussion).

There are 26 polymorphic nucleotide sites (1.6%) in the coding region of *Est-6*, but only 5 (1.1%) in *Sod*. In *Est-6*, the transversion/transition (Tv/Ts) ratio is $3/23 = 0.130$, much lower than expected from random mutation, reflecting the usual selection effect against Tv. The Tv/Ts ratio for *Sod* is significantly higher than for *Est-6* ($2/3 = 0.667$). The ratio of replacement to synonymous segregating sites is lower for *Sod* ($1/4 = 0.250$) than for *Est-6* ($9/17 = 0.529$). The *Sod* data of our study do not come from a random sample, since the number of S alleles was made intentionally larger (33%) in the sample than would be expected in a random sample (frequency of S in the natural population is ~5–15%; see Hudson *et al.* 1994, 1997). Thus, in our sample, there are more replacement substitutions in *Sod* than would be expected and fewer synonymous substitutions, since the S alleles are all completely (or very

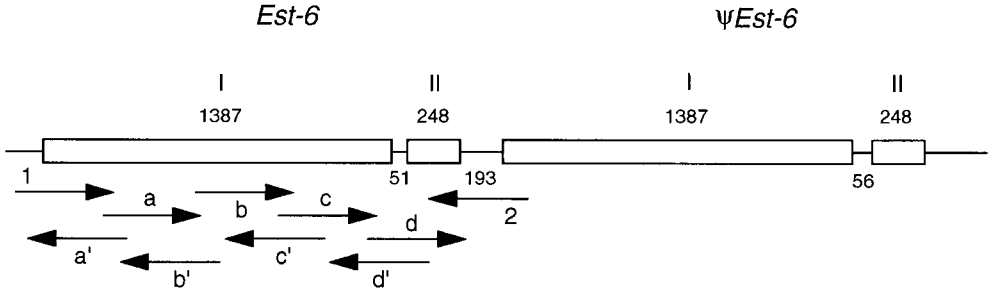

Figure 1.—Schematic representation of the *D. melanogaster* β-esterase gene cluster. This cluster consists of functional gene *Est-6* and pseudogene ψ*Est-6* (often called *Est-P*, but see Balakirev and Ayala 1996). Exons are indicated by open boxes marked by Roman numerals (I and II). The 5′ and 3′ untranslated regions, introns, and intergenic region are shown by thin lines. Lengths (in base pairs) are given above the figure for exons, and below the figure for introns and the intergenic region. The horizontal arrows indicate the location and direction of the PCR amplification primers (1 and 2) and the sequencing primers (a–d and a′–d′).

nearly) identical in sequence, reflecting their recent origin (Hudson *et al.* 1994, 1997). In the case of *Est-6*, Karotam *et al.* (1993, 1995) observed a relatively high Rep:Syn ratio in *D. melanogaster*, *D. simulans*, and *D. mauritiana*. Moriyama and Powell (1996) indicate that *D. melanogaster* tends to have the highest incidence of replacement polymorphisms (26.4%) when compared with two other Drosophila species, *D. simulans* (11.6%) and *D. pseudoobscura* (16.9%).

TABLE 1

DNA polymorphism in the *Est-6* gene of *D. melanogaster*

	1111111111111111
	1123335778889999233445566777778 8
	2577831579070450016267042729000593 4
Strain	9413118981522986906584566301347819 3
Reference	TTCCCCAGCTTAACCTATTTTCTAGTGCCAATTTC
	* ** * ** * **
255S	.....
438S	.....
510S	.....T.....AG..A.....
521S	.....C.....A
94F	.....T.....AGA...G.....
174F	.....GT....C...AGA...G.....
377F	..T.....
483F	.....
498F	.....
521F	.....
565F	.....
581F	.C.....
968F	.....
517S	C.TTTTG.TCCGGT.CGC.G.....G.A.CTGAC.
357F	C.TTTTG.TCCGGT.CGC.GCTAG.G.A.CTGAC.
simulans	.C.T..GC..C.G...GC..C.AG.G....TG...

The reference sequence is from Collet *et al.* (1990). The 15 *D. melanogaster* haplotypes are designed according to the *Sod* strain from which they originate. The numbers above the sequences represent the positions of segregating sites. Nucleotides are numbered from the initiation codon. The coding regions (exon I and exon II) are underlined below the reference sequence. Amino acid replacement polymorphisms are marked with asterisks. The S-F allozyme polymorphism is determined by site 772, where S has A (asparagine) and F has G (aspartic acid). The nucleotides in the *D. simulans* sequence are shown at the bottom; only those sites that are polymorphic in *D. melanogaster* are shown.

For *Est-6*, the A → G replacement substitution (nucleotide site 772, Table 1) results in a charge-altering amino acid replacement (Asn → Asp, amino acid position 258), which was first detected by Cooke and Oakeshott (1989). Cooke and Oakeshott (1989) have suggested that another replacement substitution (G → A, at nucleotide site 802, resulting in Ala → Thr at amino acid site 268) might also contribute to the selective differences observed between the two *Est-6* allozymes. They have proposed that one or both of the polymorphisms at 258 and 268 are the primary target for the selection underlying the F-S latitudinal clines. We have found, however, that the Ala → Thr replacement at 268 is not diagnostic for the observed F/S *Est-6* polymorphism but rather occurs in both F and S *Est-6* strains (see Table 1, site 802). Hasson and Eanes (1996) found, like us, that the F/S allozyme difference can be attributed to the single-amino-acid polymorphism at site 258 (nucleotide site 772), which is diagnostic for the *Est-6* allozyme lineages.

Haplotype structure: Figure 2 represents the maximum parsimony tree of the *Est-6* haplotypes. The distribution of the pairwise differences is bimodal, owing to haplotypes 517S and 357F, which are largely different from the rest, although quite similar to each other. These two haplotypes code for the *Est-6* Fast allozyme and will be denoted as the *Est-6* F allelic lineage, whereas the other 13 haplotypes will be denoted as the *Est-6* S allelic lineage. The great divergence between the two lineages indicates that this S-F enzyme polymorphism is ancient, at least relative to the allelic diversity within each of the two lineages.

In *D. melanogaster*, the *Est-6* S → F (asparagine → aspartic acid) replacement is associated with site 772 (A → G). *D. simulans* has an A at position 772, which suggests that in *D. melanogaster*, the S allozyme may have been the ancient condition from which the F allelic lineage derived. However, Cooke and Oakeshott (1989) and Hasson and Eanes (1996) suggest that the F lineage is ancestral on the grounds that the overall level of polymorphism is significantly greater among the

TABLE 2

Nucleotide diversity of *Sod* and *Est-6* in 15 strains of *D. melanogaster*

	Full sequence			Coding regions			Noncoding regions		
Gene	*n*	π	θ	*n*	π	θ	*n*	π	θ
All 15 haplotypes									
Est-6	1879	4.50	5.73	1635	3.77	4.89	244	9.37	11.34
		(1.63)	(0.97)		(1.38)	(0.96)		(3.32)	(3.78)
Sod	1408	6.67	9.45	439	3.34	3.50	969	8.20	12.16
		(2.58)	(1.44)		(0.76)	(1.57)		(3.48)	(1.97)
Excluding the 2 F haplotypes									
Est-6	1879	1.51	2.23	1635	1.24	1.97	244	3.36	3.96
		(0.46)	(0.62)		(0.38)	(0.62)		(1.11)	(2.29)
Sod	1408	2.17	3.43	439	2.16	2.20	969	2.17	3.99
		(0.69)	(0.89)		(0.45)	(1.27)		(1.02)	(1.15)

π is the average number of nucleotide differences per site among all pairs of sequences (Nei 1987, p. 256). θ is the average number of segregating nucleotide sites among all sequences, based on the expected distribution of neutral variants in a panmictic population at equilibrium (Watterson 1975). Standard deviations (SD) are given in parentheses below the parameter values. All π, θ, and SD values are multiplied by 10^3. Values for all 15 haplotypes are on top. Below are the values obtained after excluding the two F alleles for *Est-6* (strains 357F and 517S) and the 498F and 968F strains for *Sod*.

F than among S alleles. The two *Est-6* allelic lineages (F and S) are about equally different from *D. simulans* (90–98 nucleotide substitutions per sequence), much more so than they are from each other (22–29 substitutions per sequence), which indicates that the F-S polymorphism originated well after the divergence of the two species. The haplotypes have accumulated a number of substitutions after the divergence between the F and S allelic lineages: the two F strains differ by 4 synonymous substitutions, but no replacements, whereas there are 4 polymorphic replacement sites among the 13 S haplotypes, 1 of which (at 1526) is shared by 2 haplotypes (94F and 174F), for a total of 5 replacement polymorphisms among the 13 S haplotypes.

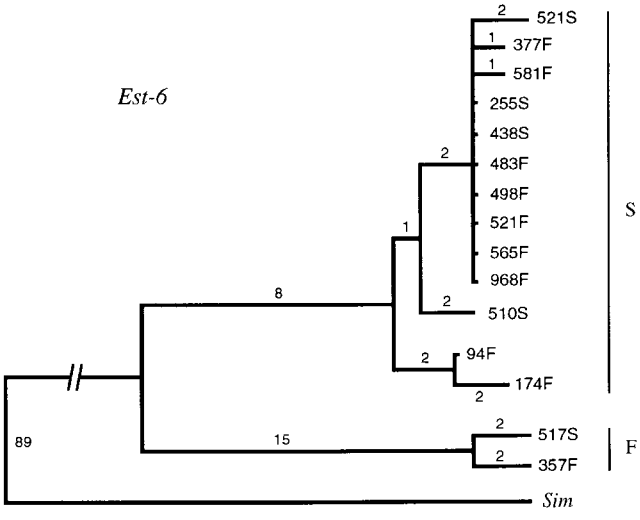

Figure 2.—Unrooted maximum-parsimony tree of the *Est-6* haplotypes of *D. melanogaster*. The haplotypes are designated according to the *Sod* strain from which they originate. Along the branches are the numbers of mutational steps. The 15 *Est-6* haplotypes group into 2 lineages, designated S and F on the right. The analysis is based on 1879 bp comprising the whole sequenced region of *Est-6*. *D. simulans* is used as an outgroup.

Figure 3 represents the maximum parsimony tree for the *Sod* haplotypes. The contrasts with *Est-6* are notable. There is only one replacement polymorphism (F → S allozyme) in *Sod*. Two F haplotypes (498F and 968F) are very different from all others, whereas the eight other F haplotypes are much more similar to the S haplotypes than they are to 498F and 968F.

Linkage disequilibrium: Within *Est-6*, 262 out of 351 pairwise comparisons (74.6%) between nonsingleton pairs of polymorphisms show statistically significant linkage disequilibrium by the chi-square test; with the Bonferroni correction for multiple comparisons, there are 192 (54.7%) significant associations. The distribution of significant associations is fairly uniform across the *Est-6* sequence; linkage disequilibrium does not decline as distance between polymorphic sites increases.

We have also found an excess of nonrandom associations within *Sod*: 211 out of 325 pairwise comparisons (64.9%) are significant, and 191 (58.8%) are significant with the Bonferroni correction. The significant associations do not form any obvious cluster, nor is the strength of linkage disequilibrium related to the distance between polymorphic sites.

We have first evaluated linkage disequilibrium between the *Sod* and *Est-6* genes using Fisher's exact test and the chi-square test, which fail to detect any significant interlocus association, as might be expected owing to asymmetrical allelic frequencies (Lewontin 1995;

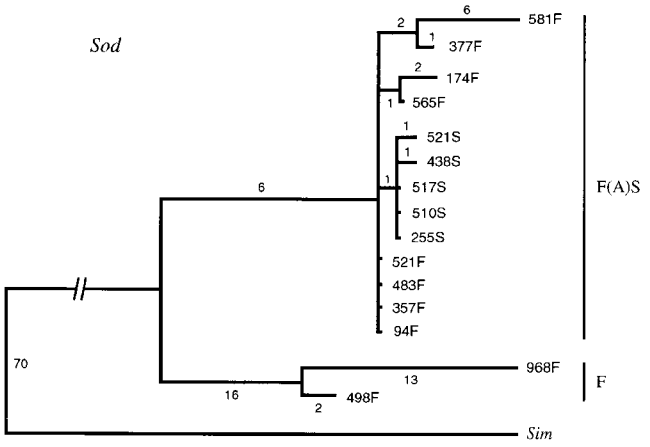

Figure 3.—Unrooted maximum-parsimony tree of the *Sod* haplotypes of *D. melanogaster*, based on 1408 bp that include the whole sequenced region of *Sod*. The 15 *Sod* haplotypes group into two lineages designated F and F(A)S. *D. simulans* is used as an outgroup.

see materials and methods). We have also used the "sign" method (Lewontin 1995), which is based on the distribution of the disequilibrium sign, which is sensitive to asymmetrical allele frequencies and efficiently operates with singleton polymorphisms, which are not informative when Fisher's exact test is used.

The sign method involves examining the number of positive and negative *D* values for each polymorphic site within and between all types of pairwise comparisons (singletons *vs.* singletons, singletons *vs.* doublets, doublets *vs.* doublets, and so on). The observed negative value summed for a particular type of pairwise comparison is compared with the expected value using a goodness-of-fit test (the likelihood ratio statistic, *G*-test, was recommended by Lewontin 1995). We have calculated the *D* values for all pairwise comparisons between sites in *Sod* and *Est-6* (Table 3). The observed total number of negative associations is 1551 *vs.* the expected number of 1248.27 ($G = 459.60$, d.f. $= 1$, $P < 0.001$; Sokal and Rohlf 1981, pp. 695–707), manifesting a significant excess of negative associations; *i.e.*, less frequent alleles at different loci are predominantly in the repulsion phase (the result remains the same after applying Williams' correction: $G^* = 459.45$, d.f. $= 1$, $P < 0.001$).

Seeking to localize the nonrandom associations, we have applied the Lewontin test separately to different regions. There are very significant associations between the *Est-6* coding region and the *Sod* intron ($G = 277.36$, d.f. $= 1$, $P < 0.001$) and between the *Est-6* coding region and the 3′-flanking region of *Sod* ($G = 53.11$, d.f. $= 1$, $P < 0.001$). There are associations between the coding regions of the two genes that are less pronounced, but still statistically significant ($G = 6.12$, d.f. $= 1$, $P < 0.05$). It is interesting that in this last comparison, the significant disequilibrium occurs between the regions of *Est-6* and *Sod* that include the sites responsible for the F/S polymorphism at each locus (*i.e.*, exon I of *Est-6* and exon II of *Sod*: $G = 6.52$, d.f. $= 1$, $P < 0.05$); no association can be detected between exon II of *Est-6* and exon II of *Sod* ($G = 1.41$, d.f. $= 1$, $P > 0.05$). The observed pattern of linkage disequilibrium between the two genes remains unchanged when singletons are excluded.

Test of neutrality: We have applied the HKA (Hudson *et al.* 1987), Tajima (1989), McDonald and Kreitman (1991), and Fu and Li (1993) tests of neutral equilibrium to our sequence data. We have first examined the intergenic region between *Est-6* and ψ*Est-6* *vs.* the coding sequence of *Est-6* without observing any significant departures from neutral expectations within *D. melanogaster* relative to the differences between *D. melanogaster* and *D. simulans* ($\chi^2 = 1.24$, $P > 0.05$). The same result was obtained by Karotam *et al.* (1995), who used various upstream sequences of *Est-6* and *Adh* as neutral reference sequences and various pairs of *D. melanogaster* and *D. simulans Est-6* alleles. Moriyama and Powell (1996) found significant deviation from neutral expectations in the comparison of the intraspecific polymorphism of *Est-6* with the interspecific divergence between *Est-6* and *Pgd*. However, nucleotide variation at *Pgd* is highly unusual because most HKA tests that include this locus (whether as the reference or test locus) show significant deviations from neutrality (Moriyama and Powell 1996). We have applied the Tajima (1989) and Fu and Li (1993) tests separately to different parts of *Est-6* and to the whole gene. These tests do not reveal any significant deviation from neutrality. However, the Tajima (1989) test applied to our data (exon I), combined with previously published *Est-6* sequences (Cooke and Oakeshott 1989; Hasson and Eanes 1996), reveals significant deviation from neutrality expectations for the S alleles ($D = -1.864$, $P < 0.05$), but not for the F alleles ($D = -0.181$, $P > 0.10$). Also, the McDonald and Kreitman (1991) test applied to four *D. simulans* (from Karotam *et al.* 1995) and 15 *D. melanogaster Est-6* coding sequences reveals significant deviation from neutrality ($G = 7.07$, d.f. $= 1$, $P < 0.01$, Table 4). The ratio of replacement to silent polymorphism within species is lower than the ratio of fixed replacement to silent differences between them.

We have applied all the neutrality tests mentioned above to the *Sod* data from Hudson *et al.* (1997) and, using the Tajima test ($D = -1.873$, $P < 0.05$), found significant departure from neutrality only for the 3′-flanking region.

Sliding window analysis: We have analyzed separately different regions of *Est-6* and *Sod* with the sliding window method (Hudson and Kaplan 1988). Figure 4 shows the sliding window plots of exon I of *Est-6* for all sequences (Figure 4A) and for the S allelic lineage only (Figure 4B). In both cases, there is a distinct peak of increased variation in the region surrounding the replacement site (772) responsible for the *Est-6* F-S allo-

TABLE 3

Linkage associations for pairs of biallelic sites at the *Est-6* gene

		D+		*D*−			
k	*m*	Observed	Expected	Observed	Expected	*G*	*G**
1	1	6	8.576	122	119.424	0.933	0.929
1	2	6	74.480	554	485.520	115.968	115.865
1	3	5	11.200	51	44.800	5.157	5.111
1	4	3	12.816	45	35.184	13.434	13.296
1	5	4	2.664	4	5.336	0.947	0.891
2	2	2	147.004	570	424.996	317.465	317.188
2	3	2	43.778	116	74.222	91.251	90.866
2	4	3	37.128	75	40.872	75.962	75.478
2	5	20	12.562	2	9.438	12.396	12.121
3	3	3	3.096	3	2.904	0.007	0.006
3	4	3	5.733	6	3.267	3.409	3.230
3	5	2	1.472	0	0.528	1.226	0.981
4	5	0	1.221	3	1.779	3.135	2.687

k is the number of copies of the rarer allele at a biallelic site; *m* is the number of copies of the rarer allele at another biallelic site for $m \geq k$. *D*+ and *D*− refer to positive and negative associations, respectively.

zyme polymorphism; the peak becomes more apparent when only the S allelic lineage is considered (Figure 4B). The same peak is also clearly distinguishable for the data of Cooke and Oakeshott (1989) and Hasson and Eanes (1996) (Figure 5, A and B). For the *Sod* gene, a distinct peak appears within the intron (Figure 6A) and toward the end of the 3′-flanking region (Figure 6B). These two peaks coincide with areas of significantly high values, according to Tajima's *D*-test.

Overall, the sliding window analysis and the neutrality tests (Tajima's and McDonald and Kreitman's) suggest that the polymorphism distribution in the *Est-6* and *Sod* genes significantly deviates from the expectations of neutrality.

Interspecific comparisons and divergence time: The *Est-6* average distance (nucleotide differences) between *D. simulans* and *D. melanogaster* is 91.9 (see Table 4). The average distance between the two main (F and S) allelic lineages of *D. melanogaster* is 25.5. If we assume that the divergence between the species occurred 2.3–2.5 mya (Powell and DeSalle 1995; Russo *et al.* 1995), the divergence of the two *D. melanogaster* lineages would have occurred ~666,000 years ago—assuming that *Est-6* is a good molecular clock. The divergence between the two F alleles occurred ~105,000 years ago and between S alleles ~73,000 years ago.

The *Sod* average distance between *D. melanogaster* and *D. simulans* is 73.1, while between the two main allelic lineages [F *vs.* F(A)S] it is 29.3, which corresponds to ~962,000 years, while the divergence between the two F alleles occurred ~427,000 years ago. The average distance among all F(A) and S alleles (excluding 581F, which is probably a recombinant) is 2.4, corresponding to 79,000 years ago; the average distance among the S alleles is 0.8 or ~26,000 years ago.

We have calculated the time of divergence between and within the allelic lineages for both genes following the Hudson *et al.* (1997) approach, using all sequences or their homogeneous subsets [see Hudson *et al.* (1997) for details]. For the *Est-6*, the expected time of divergence between the F and S allelic lineages is ~223,000 years [$\mu \times t \times 15\ (1879 - 0.75 \times 1635) = 35$; $t = 223413.8$], where μ is the neutral mutation rate at noncoding and silent sites, assumed to be 16×10^{-9} per

TABLE 4

McDonald-Kreitman test of neutrality for the *Est-6* gene

		Polymorphisms		
	Fixed differences	*D. melanogaster*	*D. simulans*[a]	Total[b]
Replacement	19	9	11	18
Silent	25	17	54	69

[a] Calculated from Karotam *et al.* (1995).

[b] Sites that are polymorphic in both species are counted only once. For the two-tailed Fisher's exact test, $P = 0.013$; *G*-test without correction = 7.070, $P = 0.008$; *G*-test with Williams' correction = 6.949, $P = 0.008$; *G*-test with Yates' correction = 6.042, $P = 0.014$.

site per year. For the *Sod* F(A)S and F allelic lineages, the divergence time is ~178,000 years [$\mu \times t \times 15$ $(1408 - 0.75 \times 439) = 46$; $t = 177633.6$]. The divergence time within the *Est-6* S allelic lineage is 96,000 years [$\mu \times t \times 13$ $(1879 - 0.75 \times 1635) = 13$; $t = 95748.8$], considering all polymorphic sites, and 59,000 years [$\mu \times t \times 13$ $(1879 - 0.75 \times 1635) = 8$; $t = 58922.3$], if we exclude putative recombinant sites. The expected time of divergence within the *Sod* F(A)S allelic lineage is ~45,000 years [$\mu \times t \times 13$ $(1408 - 0.75 \times 439) = 10$; $t = 44556.9$] and 27,000 years [$\mu \times t \times 13$ $(1408 - 0.75 \times 439) = 6$; $t = 26734.1$], excluding the putative recombinant sites.

DISCUSSION

The *Est-6* and *Sod* genes of *D. melanogaster* are closely linked on the left arm of chromosome 3, separated by 3.4 cM, or 1 Mb. Both genes are characterized in natural populations by a polymorphism with two common allozymes, S and F, which differ by a single-nucleotide substitution and a corresponding amino acid replacement. In *Sod*, the S-F replacement is the only amino acid polymorphism commonly found in natural populations, whereas in *Est-6*, additional rare amino acid replacements are found.

Natural selection: The Cu,Zn SOD is involved in protecting the cell against the toxicity of oxygen radicals by scavenging superoxide radicals and dismutating them to hydrogen peroxide and molecular oxygen (Fridovich 1986). The *Sod* F-S polymorphism is of recent evolutionary origin, as shown by the virtually complete absence of silent variation among S alleles collected in populations that are geographically very distant, from China and Europe to the United States (Hudson *et al.* 1994, 1997). S alleles have not been found in South America, nor in Africa, whence *D. melanogaster* spread throughout the world, probably in recent millennia. In the United States, the frequency of S has been found to range from 0 to 0.15 in different populations, and can vary from year to year in a given population from 0.05 to 0.15 (Hudson *et al.* 1994, 1997). The S alleles are completely or very nearly identical in nucleotide sequence to a set of F alleles denominated F(A). The F(A) alleles have not been found in Africa, but are present in Europe, the United States, and South America, where they often account for ~50% of all F alleles in a population (Hudson *et al.* 1994, 1997).

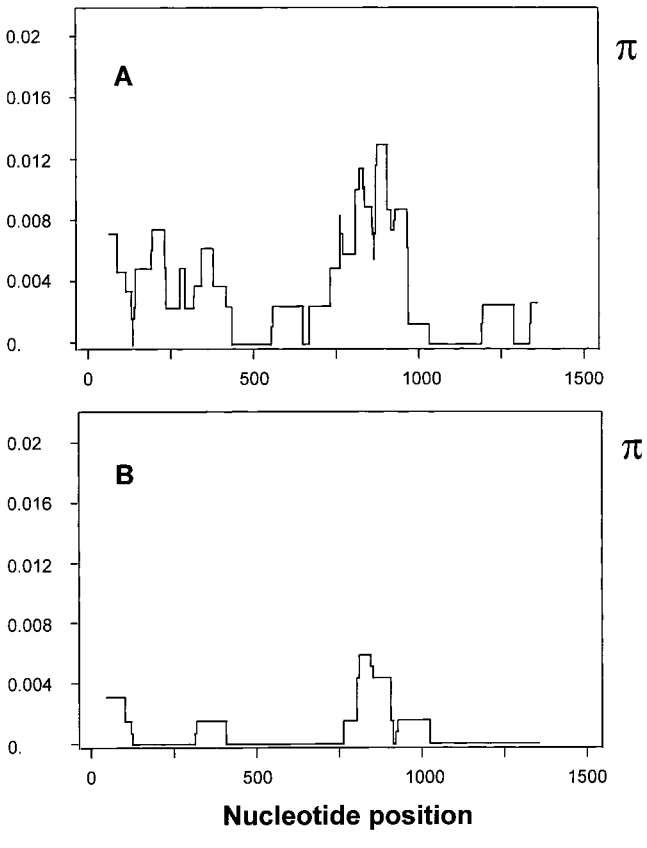

Figure 4.—Sliding window plot of exon I polymorphism (π) in the *Est-6* gene of *D. melanogaster*. (A) All data. (B) Excluding the F allelic lineage. Window size is 100 nucleotides, with one-nucleotide increments.

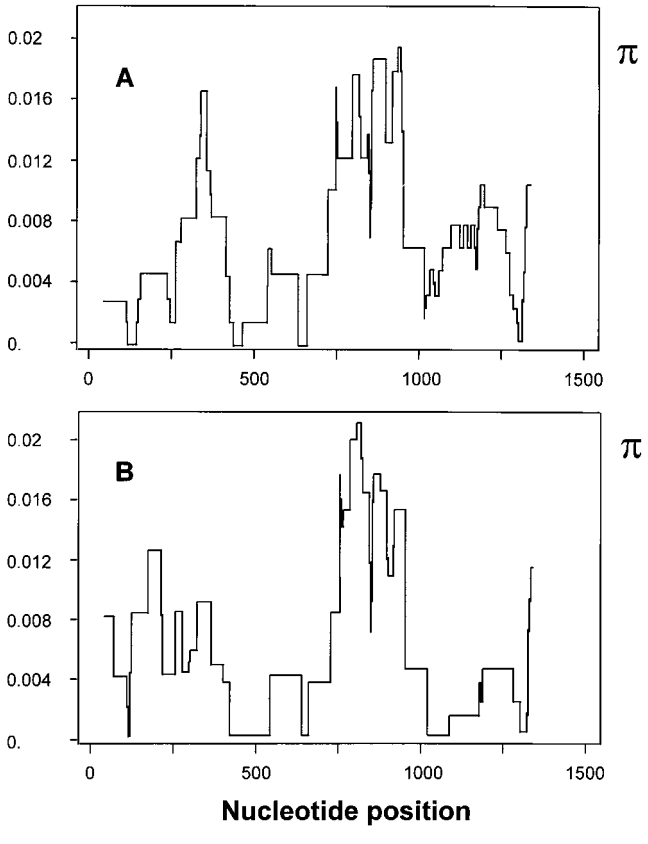

Figure 5.—Sliding window plot of exon I polymorphism (π) in the *Est-6* gene of *D. melanogaster*. (A) Data of Cooke and Oakeshott (1989). (B) Data of Hasson and Eanes (1996). Window size is 100 nucleotides, with one-nucleotide increments.

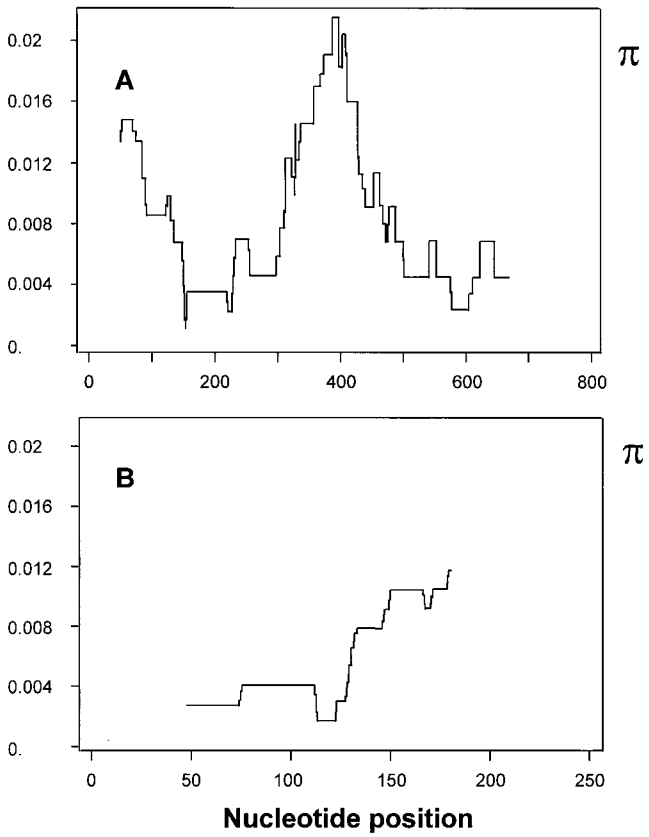

Figure 6.—Sliding window plot of noncoding polymorphism (π) in the *Sod* gene of *D. melanogaster*. (A) Intron. (B) The 3′-flanking region. Window size is 100 nucleotides, with one-nucleotide increments.

A hypothesis of the geographic evolution of the *Sod* alleles consistent with the information just summarized is as follows. The F(A) mutation arose in a *D. melanogaster* population outside Africa ~5000 years ago (Hudson *et al.* 1997) and rapidly spread throughout other continents, impelled by natural selection (Hudson *et al.* 1994). The rapidity of the F(A) world expansion is evidenced by the virtually complete absence of silent substitutions throughout a fragment >10 kb that includes *Sod* (Hudson *et al.* 1997). The S allele arose by a single-nucleotide replacement from F(A), either in Europe or the United States, and spread from one to the other continent, but never reached South America or Africa. The nucleotide identity of S alleles from widely distant continents (and the nucleotide identity of a fragment >10 kb that includes *Sod* S) favors the interpretation that the F(A) → S mutation occurred only once. The rapid expansion of S in Europe and the United States must have been impelled by natural selection. It is uncertain, however, whether the selective advantage is the same favoring F(A) over other F alleles, or whether the S replacement is also favored. The S and F enzymes differ in such biochemical properties as thermostability and specific activity (Lee *et al.* 1981). Moreover, the S allele is at an advantage relative to F in heavily irradiated populations (Peng *et al.* 1991) or in those selected for reproduction at an advanced age (Tyler *et al.* 1993).

Detecting selection is handicapped in our sample because we have studied only 15 sequences (Hanfstingl *et al.* 1994; Simonsen *et al.* 1995; Hasson and Eanes 1996; Richter *et al.* 1997). Nevertheless, we have found evidence of natural selection within the 3′-flanking region of *Sod* with Tajima's (1989) test. Moreover, the sliding window method of Hudson and Kaplan (1988) manifests a distinct peak in the 3′-flanking region as well as within the intron of *Sod*; these peaks are expected as a consequence of balancing selection (Strobeck 1983; Hudson and Kaplan 1988). Selective effects within introns have been recognized previously. Berry and Kreitman (1993) have shown that an intron polymorphism at *Adh* in *D. melanogaster* exhibits a more pronounced cline than the F-S allozyme polymorphism that is generally attributed to selection at that locus. Kirby and Stephan (1995) have proposed that positive selection accounts for intron polymorphism also at the *white* locus of *D. melanogaster*. This may be because introns may include regulatory sequences (Bingham *et al.* 1988; Aronow *et al.* 1989; Gasch *et al.* 1989; Huang *et al.* 1993; Pogulis and Freytag 1993). There is also other evidence for selective constraints and epistatic selection on the nucleotide sequence evolution of introns (Learn *et al.* 1992; Leicht *et al.* 1993, 1995; Stephan and Kirby 1993; Kirby *et al.* 1995).

The *Est-6* protein is transferred by males to females in the semen fluid during copulation in *D. melanogaster* (Richmond *et al.* 1980; Richmond and Senior 1981), and it affects the female's consequent behavior and mating proclivity (Gromko *et al.* 1984; Scott 1986). The *Est-6* coding sequence is selectively constrained (Cooke and Oakeshott 1989; Karotam *et al.* 1993, 1995), although presumably to a lesser extent than *Sod*, which exhibits only one polymorphic amino acid site in our sample *vs.* the nine polymorphic sites present in *Est-6*. The *Est-6* 5′-flanking region contains positive *cis*-regulatory elements controlling the expression of *Est-6* and may contain binding sites for regulatory protein factors involved in tissue-specific transcription control. The evidence indicates that the 5′-flanking region is evolving under greater selective constraints than the coding region (Karotam *et al.* 1993, 1995; Odgers *et al.* 1995). The occurrence of parallel latitudinal clines of the *Est-6* F-S polymorphism in Europe, North America, and Australia, the pattern of temporal variation, and other lines of evidence all indicate that the *Est-6* F-S polymorphism is subject to balancing selection (Oakeshott *et al.* 1989, 1993, 1995; Richmond *et al.* 1990). The peak around the S-F amino acid polymorphism we have observed (Figure 4) is also evidence of balancing selection impacting the *Est-6* S-F polymorphism in our sample. The detected tendency cannot be explained by chance, because similar peaks are observed in the *Est-6* data of other authors (Cooke and Oake-

shott 1989; Hasson and Eanes 1996). The McDonald and Kreitman (1991) test applied to the *Est-6* coding region also reveals significant deviation from neutrality.

The *Est-6* and *Sod* loci are enclosed within the cosmopolitan inversion *In(3L)Payne*, which ranges globally from 0 to 40% (Voelker *et al.* 1978; Lemeunier and Aulard 1992). Selection at the two loci might simply be a consequence of distinctive associations between alleles and chromosomal arrangements. Hasson and Eanes (1996) have shown, however, that the *In(3L)Payne* inversion suppresses recombination in regions proximal to the chromosome breakpoints, but does not affect the central region, where *Est-6* and *Sod* are located. Moreover, there is no association between *Est-6* sequence variation and arrangement type, consistent with the extensive genetic exchange observed between standard (*ST*) and inverted [*In(3L)Payne*] third chromosomes of *D. melanogaster* (Hasson and Eanes 1996). In any case, there are no third chromosome inversions segregating in the El Rio population (Smit-McBride *et al.* 1988).

History of the allelic lineages: Figures 2 and 3 are maximum-parsimony trees of the *Est-6* and *Sod* alleles. It is apparent that the phylogeny of the electrophoretic alleles is different in the two genes. The two *Est-6* F alleles are quite similar to each other, but they are in haplotypes that carry an F(A) *Sod* allele in one case and an S *Sod* allele in the other case. The 13 *Est-6* S alleles are associated with S or F *Sod* alleles, and the *Sod* F alleles include closely related F(A) alleles as well as the other distant F alleles. The F(A) and S alleles of *Sod* have diverged very recently (Hudson *et al.* 1994, 1997), which is also apparent in Figure 3 and our extensive unpublished data (we reckon that the exceptional *Sod* allele 581F represents a case of intragenic recombination). We did not, however, expect that the two *Est-6* F alleles would be closely related to one another, and even less so that the 13 *Est-6* S alleles would have quite similar (and even identical) nucleotide sequences. Rather, we would have expected that the *Est-6* S alleles, which may represent the ancestral electrophoretic state, would consist of alleles quite heterogeneous in nucleotide sequence. Such is, indeed, the case for the *Sod* F alleles, which, if we exclude the F(A) set, are extremely heterogeneous in nucleotide sequence (Hudson *et al.* 1994; F. J. Ayala, unpublished data). A reasonable account of the observations is to assume that the *Sod* F(A) mutation occurred in a haplotype carrying an *Est-6* S allele, and that the rapid world expansion of the *Sod* F(A) and S alleles greatly enhanced the frequency of that particular *Est-6* S allele. The *Sod* F(A) mutation is estimated to have occurred ~5000 years ago or somewhat earlier (Hudson *et al.* 1997). The 5000 years of the world expansion of the *Sod* F(A) and S alleles would allow for only limited recombination between *Sod* and *Est-6*. This scenario is consistent with the interpretation that the selective pressure favoring *Sod* F(A) and S alleles has been strong. The two strains 517S and 357F [which have F(A) and S alleles, respectively, at *Sod*, but the *Est-6* F allele] would have arisen by a recombination event between *Sod* and *Est-6*. The fact that the two *Sod* F alleles that are not F(A) occur in haplotypes with *Est-6* S alleles is not surprising because the *Est-6* S allozyme is the most ancient and common.

An alternative historical scenario that might account for the presence of the two largely divergent sets of alleles that we observe at each locus is population subdivision with subsequent admixture. *D. melanogaster* may have been geographically split into two populations that remained separate for a time long enough to accumulate a number of nucleotide substitutions within each gene. The substitutions would have accumulated independently in the two populations. Recent admixture of the two populations would have brought together the two sets of alleles, as we find them in the El Rio population, where our strains were collected. According to this scenario, however, the two sets of alleles at the two loci would be associated in the same haplotypes. This is not the case. As shown in Figures 2 and 3, the strains with the two *Est-6* F alleles (517S and 357F) are different from the strains carrying the two *Sod* F alleles (377F and 581F).

Linkage disequilibrium: *Sod* and *Est-6* are 3.4 cM apart in the recombination map (3–32.5 and 3–35.9; Heino *et al.* 1994), separated by 983 kb in the genomic map (A. Long, personal communication). Yet the nucleotide polymorphisms are negatively associated between the two loci (Table 3, $P \ll 0.001$); that is, minority nucleotide substitutions (present in one to five strains) at one locus are negatively associated with minority nucleotide substitutions at the other locus. A selection account of this observation implies that mutations occurring at one locus are selected against if nucleotide substitutions are present in the other locus, but not if the other locus exhibits the consensus sequence.

There is an extensive genetic literature advancing the notion that the unit of evolution is not the single gene, but rather, interacting gene complexes (Wright 1931; Dobzhansky 1937; Schmalhausen 1946; Mather 1953; Waddington 1957; Mayr 1963; Franklin and Lewontin 1970). There also are traditional arguments propounding that sets of genes for quantitative traits are built up in an alternated arrangement of plus and minus alleles on chromosomes, so that selection minimizes actual variation for such traits, but maximizes potential variability, which is released by recombination. This + / − alternating arrangement explains why populations rapidly respond to varying environmental challenges, as well as the success of artificial selection favoring extreme values of the traits (Mather 1953; Thoday *et al.* 1964). However, even if we accept such theoretical constructs and their evidential support, it remains obscure why the plus/minus compensatory arrangement would occur between two genes that are

unlikely to share distinctive functional interactions, and between nucleotide substitutions that are in most cases synonymous.

It seems more likely that linkage disequilibrium has arisen as a consequence of selection strongly favoring one particular sequence at a locus, such as seems to have occurred at *Sod*, where the F(A) alleles (and the derivative S alleles) rapidly increased in frequency. Rare substitutions present in *Est-6* would have hitchhiked along with *Sod* F(A) without allowing enough time for their elimination by purifying selection with or without recombination between the two loci. The reciprocal situation would have also occurred, in which common *Est-6* alleles are favored by selection and low-frequency *Sod* substitutions are hitchhiking along. As noted earlier, there is evidence of positive selection in favor of *Sod* F(A); in the case of *Est-6*, the evidence favors balancing selection between the two common allozymes, S and F. In a study of two wild samples and four experimental populations, Smit-McBride *et al.* (1988) did not observe linkage disequilibrium between *Est-6* and *Sod* in the large wild samples, but linkage disequilibrium appeared after 8–30 generations in three of the four experimental populations derived from the wild samples. Population bottlenecks and other factors were excluded. Smit-McBride *et al.* (1988) noted that natural selection acting on the allozymes, or on loci very tightly linked to them, was the most likely explanation for the disequilibrium.

Hudson *et al.* (1997) have estimated that the time since the selective sweep for *Sod* F(A)S is ~5000 years or more, but not likely more than 30,000–40,000 years. Assuming a molecular (neutral) clock, the age of this lineage is estimated at 79,000 years. Similarly, the time of divergence between the two main *Sod* allelic lineages is 178,000 years with the method of Hudson *et al.* (1997), but five times greater, 962,000 years, under the molecular clock assumption. These discrepancies can be accounted for by natural selection accelerating the evolution of the selectively favored F(A)S alleles. Hudson *et al.* (1997) argue that the selection is strong ($s > 0.01$). Parallel differences occur for *Est-6*, but the differences between the estimates obtained by the two methods are smaller. Thus, the age of the *Est-6* S allelic lineage is 60,000–95,000 *vs.* 73,000 years, and the divergence between the S-F lineages is 220,000 *vs.* 666,000 years, corresponding to the method of Hudson *et al.* (1997) *vs.* the clock method. This is consistent with the hypothesis that the divergence between the two *Est-6* allelic lineages has been impelled by natural selection, but with lesser strength than in the case of *Sod*.

We thank several members of our laboratory: Kevin Bailey, Heather Carstens, Victor DeFilippis, Alberto García Sáez, Jan Kwiatowski, Carlos Machado, Stephen Rich, Douglas Skarecky, and Andrei Tatarenkov for encouragement and help. Walter M. Fitch, Brandon Gaut, Richard R. Hudson, and Anthony Long read the manuscript and offered valuable comments. This work is supported by National Institutes of Health grant GM-42397 to F.J.A.

LITERATURE CITED

Aguadé, M., 1998 Different forces drive the evolution of the *Acp26Aa* and *Acp26Ab* accessory gland genes in the *Drosophila melanogaster* species complex. Genetics **150:** 1079–1089.

Aquadro, C. F., 1993 Molecular population genetics of Drosophila, pp. 222–266 in *Molecular Approaches to Fundamental and Applied Entomology*, edited by J. Oakeshott and M. J. Whitten. Springer-Verlag, New York.

Aronow, B., D. Lattler, R. Silbiger, M. Dusing, J. Hutton *et al.*, 1989 Evidence for a complex regulatory array in the first intron of the human adenosine deaminase gene. Genes Dev. **3:** 1384–1400.

Avery, P. J., and W. G. Hill, 1979 Distribution of linkage disequilibrium with selection and finite population size. Genet. Res. **33:** 29–48.

Ayala, F. J., J. R. Powell and Th. Dobzhansky, 1971 Enzyme variability in the *Drosophila willistoni* group. II. Polymorphisms in continental and island populations of *Drosophila willistoni*. Proc. Natl. Acad. Sci. USA **68:** 2480–2483.

Ayala, F. J., J. R. Powell, M. L. Tracey, C. A. Mourão and S. Pérez-Salas, 1972a Enzyme variability in the *Drosophila willistoni* group. IV. Genic variation in natural populations of *Drosophila willistoni*. Genetics **70:** 113–139.

Ayala, F. J., J. R. Powell and M. L. Tracey, 1972b Enzyme variability in the *Drosophila willistoni* group. V. Genetic variability in natural populations of *Drosophila equinoxialis*. Genet. Res. **20:** 19–42.

Balakirev, E. S., and F. J. Ayala, 1996 Is esterase-P encoded by a cryptic pseudogene in *Drosophila melanogaster*? Genetics **144:** 1511–1518.

Barker, J. A. F., 1979 Inter-locus interactions: a review of experimental evidence. Theor. Popul. Biol. **16:** 323–346.

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.

Berry, A., and M. Kreitman, 1993 Molecular analysis of an allozyme cline: alcohol dehydrogenase in *Drosophila melanogaster* on the east coast of North America. Genetics **134:** 869–893.

Bingham, P. M., T.-B. Chou, I. Mims and Z. Zachar, 1988 On-off regulation of gene expression at the level of splicing. Trends Genet. **4:** 134–138.

Brown, A. D. H., 1975 Sample sizes required to detect linkage disequilibrium between two or three loci. Theor. Popul. Biol. **8:** 184–201.

Cirera, S., and M. Aguadé, 1997 Evolutionary history of the sex-peptide (*Acp70A*) gene region in *Drosophila melanogaster*. Genetics **147:** 189–197.

Collet, C., K. M. Nielsen, R. J. Russell, M. Karl, J. G. Oakeshott *et al.*, 1990 Molecular analysis of duplicated esterase genes in *Drosophila melanogaster*. Mol. Biol. Evol. **7:** 9–28.

Cooke, P. H., and J. G. Oakeshott, 1989 Amino acid polymorphisms for esterase-6 in *Drosophila melanogaster*. Proc. Natl. Acad. Sci. USA **86:** 1426–1430.

Dobzhansky, Th., 1937 *Genetics and the Origin of Species*. Columbia University Press, New York.

Eanes, W. F., M. Kirchner, J. Yoon, C. Ch. Biermann, I.-N. Wang *et al.*, 1996 Historical selection, amino acid polymorphism and lineage-specific divergence at the *G6pd* locus in *Drosophila melanogaster* and *D. simulans*. Genetics **144:** 1027–1041.

Feldman, M. W., and F. B. Christiansen, 1975 The effect of population subdivision on two loci without selection. Genet. Res. **24:** 151–162.

Franklin, I., and R. C. Lewontin, 1970 Is the gene the unit of selection? Genetics **65:** 707–734.

Fridovich, I., 1986 Superoxide dismutases. Adv. Enzymol. **58:** 62–97.

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

Gasch, A., U. Hinz and R. Renkawitz-Pohl, 1989 Intron and upstream sequences regulate expression of the Drosophila beta 3-tubulin gene in the visceral and somatic musculature, respectively. Proc. Natl. Acad. Sci. USA **86:** 3215–3218.

Gromko, M. H., D. F. Gilbert and R. C. Richmond, 1984 Sperm transfer and use in the multiple mating system of Drosophila, pp. 371–426 in *Sperm Competition and the Evolution of Animal Mating Systems*, edited by R. L. Smith. Academic Press, New York.

Hanfstingl, U., A. Berry, E. A. Kellogg, J. T. Costa III, W. Rüdiger *et al.*, 1994 Haplotypic divergence coupled with lack of diversity at the *Arabidopsis thaliana* alcohol dehydrogenase locus: roles for both balancing and directional selection? Genetics **138:** 811–828.

Hartl, D. L., and E. R. Lozovskaya, 1995 *The Drosophila Genome Map: A Practical Guide*. R. G. Landes, Austin, TX.

Hasson, E., and W. F. Eanes, 1996 Contrasting histories of three gene regions associated with *In(3L)Payne* of *Drosophila melanogaster*. Genetics **144:** 1565–1575.

Hedrick, P. W., S. Jain and L. Holden, 1978 Multilocus systems in evolution. Evol. Biol. **11:** 101–182.

Heino, T. I., A. O. Saura and V. Sorsa, 1994 Maps of the salivary gland chromosomes of *Drosophila melanogaster*. Dros. Inf. Serv. **73:** 621–738.

Hill, W. G., 1975 Linkage disequilibrium among multiple neutral alleles produced by mutation in a finite population. Theor. Popul. Biol. **8:** 117–126.

Hill, W. G., 1976 Non-random association of neutral linked genes in finite populations, pp. 339–376 in *Population Genetics and Ecology*, edited by S. Karlin and E. Nevo. Academic Press, New York.

Hill, W. G., and A. Robertson, 1968 Linkage disequilibrium in finite populations. Theor. Appl. Genet. **38:** 226–231.

Huang, J. D., D. H. Schwyter, J. M. Shirokawa and A. J. Courey, 1993 The interplay between multiple enhancer and silencer elements defines the pattern of decapentaplegic expression. Genes Dev. **7:** 694–704.

Hudson, R. R., and N. Kaplan, 1988 The coalescent process in models with selection and recombination. Genetics **120:** 831–840.

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., K. Bailey, D. Skarecky, J. Kwiatowski and F. J. Ayala, 1994 Evidence for positive selection in the superoxide dismutase (*Sod*) region of *Drosophila melanogaster*. Genetics **136:** 1329–1340.

Hudson, R. R., A. G. Sáez and F. J. Ayala, 1997 DNA variation at the *Sod* locus of *Drosophila melanogaster*: an unfolding story of natural selection. Proc. Natl. Acad. Sci. USA **94:** 7725–7729.

Ishii, K., and B. Charlesworth, 1977 Associations between allozyme loci and gene arrangements due to hitch-hiking effects of new inversions. Genet. Res. **30:** 93–106.

Karotam, J., A. C. Delves and J. G. Oakeshott, 1993 Conservation and change in structural and 5′ flanking sequences of esterase 6 in sibling Drosophila species. Genetics **88:** 11–28.

Karotam, J., T. M. Boyce and J. G. Oakeshott, 1995 Nucleotide variation at the hypervariable esterase 6 isozyme locus of *Drosophila simulans*. Mol. Biol. Evol. **12:** 113–122.

Kimura, M., 1968 Evolutionary rate at the molecular level. Nature **217:** 624–626.

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

Kimura, M., and T. Ohta, 1971 Protein polymorphism as a phase of molecular evolution. Nature **229:** 467–469.

Kirby, D. A., and W. Stephan, 1995 Haplotype test reveals departure from neutrality in a segment of the *white* gene of *Drosophila melanogaster*. Genetics **141:** 1483–1490.

Kirby, D. A., and W. Stephan, 1996 Multi-locus selection and the structure of variation at the *white* gene of *Drosophila melanogaster*. Genetics **144:** 635–645.

Kirby, D. A., S. V. Muse and W. Stephan, 1995 Maintenance of pre-mRNA secondary structure by epistatic selection. Proc. Natl. Acad. Sci. USA **92:** 9047–9051.

Kreitman, M., and R. R. Hudson, 1991 Inferring the evolutionary histories of the *Adh* and *Adh-dup* loci in *Drosophila melanogaster* from patterns of polymorphism and divergence. Genetics **127:** 565–582.

Krimbas, C. B., and J. R. Powell, 1993 Introduction, pp. 1–52 in *Inversion Polymorphism in Drosophila*, edited by C. B. Krimbas and J. R. Powell. CRC Press, Boca Raton, FL.

Kwiatowski, J., D. Skarecky, S. Hernandez, D. Pham, F. Quijas *et al.*, 1991 High fidelity of the polymerase chain reaction. Mol. Biol. Evol. **8:** 884–887.

Langley, C., 1977 Nonrandom associations between allozymes in natural populations of *Drosophila melanogaster*, pp. 265–273 in *Measuring Selection in Natural Populations*, Vol. 19, edited by F. B. Christiansen and T. M. Fenchel. Lect. Notes Biomath., Springer-Verlag, New York.

Learn, G. H., J. S. Shore, G. R. Furnier, G. Zurawski and M. T. Clegg, 1992 Constraints on the evolution of plastid introns: the group II intron in the gene encoding tRNA-val (UAC). Mol. Biol. Evol. **9:** 856–871.

Lee, Y. M., H. P. Misra and F. J. Ayala, 1981 Superoxide dismutase in *Drosophila melanogaster*: biochemical and structural characterization of allozyme variants. Proc. Natl. Acad. Sci. USA **78:** 7052–7055.

Leicht, B. G., E. M. S. Lyckegaard, C. M. Benedict and A. G. Clark, 1993 Conservation of alternative splicing and genomic organization of the myosin alkali light chain (*Mlc1*) gene among Drosophila species. Mol. Biol. Evol. **10:** 769–790.

Leicht, B. G., S. V. Muse, M. Hanczyc and A. G. Clark, 1995 Constraints on intron evolution in the gene encoding the myosin alkali light chain in Drosophila. Genetics **139:** 299–308.

Lemeunier, F., and S. Aulard, 1992 Inversion polymorphism in *Drosophila melanogaster*, pp. 339–405 in *Drosophila Inversion Polymorphism*, edited by C. B. Krimbas and J. R. Powell. CRC Press, Cleveland.

Lewontin, R. C., 1974 *The Genetic Basis of Evolutionary Change*. Columbia University Press, New York.

Lewontin, R. C., 1995 The detection of linkage disequilibrium in molecular sequence data. Genetics **140:** 377–388.

Li, W.-H., and M. Nei, 1974 Stable linkage disequilibrium without epistasis in subdivided populations. Theor. Popul. Biol. **6:** 173–183.

Manchenko, G. P., 1994 *Handbook of Detection of Enzymes on Electrophoretic Gels*. CRC Press, Boca Raton, FL.

Markert, C. L., and I. Faulhaber, 1965 Lactate dehydrogenase isozyme patterns in fish. J. Exp. Zool. **159:** 319–325.

Mather, K., 1953 The genetical structure of populations. Symp. Soc. Exp. Biol. **7:** 66–95.

Mayr, E., 1963 *Animal Species and Evolution*. Belknap, Cambridge.

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.

Mukai, T., and R. A. Voelker, 1977 The genetic structure of natural populations of *Drosophila melanogaster*. XIII. Further studies on linkage disequilibrium. Genetics **86:** 175–185.

Mukai, T., T. K. Watanabe and O. Yamaguchi, 1974 The genetic structure of natural populations of *Drosophila melanogaster*. XII. Linkage disequilibrium in a large local population. Genetics **77:** 771–793.

Nei, M., 1987 *Molecular Evolutionary Genetics*. Columbia University Press, New York.

Nei, M., and W.-H. Li, 1973 Linkage disequilibrium in subdivided populations. Genetics **75:** 213–219.

Nei, M., and W.-H. Li, 1980 Non-random association between electromorphs and inversion chromosomes in finite populations. Genet. Res. **35:** 65–83.

Oakeshott, J. G., P. H. Cooke, R. C. Richmond, A. Bortoli, A. Y. Game *et al.*, 1989 Molecular population genetics of structural variance of esterase-6 in *Drosophila melanogaster*. Genome **31:** 788–796.

Oakeshott, J. G., E. A. van Papenrecht, T. M. Boyce, M. J. Healy and R. J. Russell, 1993 Evolutionary genetics of Drosophila esterases. Genetica **90:** 239–268.

Oakeshott, J. G., T. M. Boyce, R. J. Russell and M. J. Healy, 1995 Molecular insights into the evolution of an enzyme: esterase 6 in *Drosophila*. TREE **10:** 103–110.

Odgers, W. A., M. J. Healy and J. G. Oakeshott, 1995 Nucleotide polymorphism in the 5′ promoter region of esterase 6 in *Drosophila melanogaster* and its relationship to enzyme activity variation. Genetics **141:** 215–222.

Ohta, T., 1982a Linkage disequilibrium due to random genetic drift in finite subdivided populations. Proc. Natl. Acad. Sci. USA **79:** 1940–1944.

Ohta, T., 1982b Linkage disequilibrium with the island model. Genetics **101:** 139–155.

Ohta, T., and M. Kimura, 1969a Linkage disequilibrium due to random genetic drift. Genet. Res. **13:** 47–55.

Ohta, T., and M. Kimura, 1969b Linkage disequilibrium at steady state determined by random genetic drift and recurrent mutation. Genetics **63:** 229–238.

Palumbi, S., A. Martin, S. Romano, W. O. McMillan, L. Stice *et al.*, 1991 *The Simple Fool's Guide to PCR*, Version 2.0. Department of Zoology and Kewalo Marine Laboratory, University of Hawaii, Honolulu.

Peng, T. X., A. Moya and F. J. Ayala, 1991 Two modes of balancing selection in *Drosophila melanogaster*: overcompensation and overdominance. Genetics **128:** 381–391.

Pogulis, R. J., and S. O. Freytag, 1993 Contribution of specific *cis*-acting elements to activity of the mouse pro-alpha 2(1) collagen enhancer. J. Biol. Chem. **268:** 2493–2499.

Powell, J. R., and R. DeSalle, 1995 *Drosophila* molecular phylogenies and their uses, pp. 87–138 in *Evolutionary Biology*, edited by M. K. Hecht, R. J. MacIntyre and M. T. Clegg. Plenum Press, New York.

Prakash, S., 1974 Gene differences between the sex ratio and standard gene arrangements of the X chromosome and linkage disequilibrium between loci in the standard gene arrangement of the X chromosome in *Drosophila pseudoobscura*. Genetics **77:** 795–804.

Prakash, S., and R. C. Lewontin, 1968 A molecular approach to the study of genic heterozygosity in natural populations. III. Direct evidence of coadaptation in gene arrangements of *Drosophila*. Proc. Natl. Acad. Sci. USA **59:** 398–405.

Procunier, W. S., J. J. Smith and R. C. Richmond, 1991 Physical mapping of the *Esterase-6* locus of *Drosophila melanogaster*. Genetica **84:** 203–208.

Richmond, R. C., and A. Senior, 1981 Esterase 6 of *Drosophila melanogaster*: kinetics of transfer to female, decay in females and male recovery. J. Insect Physiol. **27:** 849–853.

Richmond, R. C., D. G. Gilbert, K. B. Sheehan, M. H. Gromko and F. J. Butterworth, 1980 Esterase 6 and reproduction in *Drosophila melanogaster*. Science **207:** 1483–1485.

Richmond, R. C., K. M. Nielsen, J. P. Brady and E. M. Snella, 1990 Physiology, biochemistry and molecular biology of the *Est-6* locus in *Drosophila melanogaster*, pp. 273–292 in *Ecological and Evolutionary Genetics of Drosophila*, edited by J. S. F. Barker, W. T. Starmer and R. J. MacIntyre. Plenum, New York.

Richter, B., M. Long, R. C. Lewontin and E. Nitasaka, 1997 Nucleotide variation and conservation at the *dpp* locus, a gene controlling early development in Drosophila. Genetics **145:** 311–323.

Rosato, E., A. A. Peixoto, R. Costa and C. P. Kyriacou, 1997 Linkage disequilibrium, mutational analysis and natural selection in the repetitive region of the clock gene, *period*, in *Drosophila melanogaster*. Genet. Res. **69:** 89–99.

Rozas, J., and R. Rozas, 1997 DnaSP, Version 2.0: a novel software package for extensive molecular population genetics analysis. Comput. Appl. Biosci. **13:** 307–311.

Russo, C. A. M., N. Takezaki and M. Nei, 1995 Molecular phylogeny and divergence times of drosophilid species. Mol. Biol. Evol. **12:** 391–404.

Sanger, F., S. Nicklen and A. R. Coulson, 1977 DNA sequencing with chain terminating inhibitors. Proc. Natl. Acad. Sci. USA **75:** 5463–5467.

Schaeffer, S. W., and E. L. Miller, 1993 Estimates of linkage disequilibrium and the recombination parameter determined from segregating nucleotide sites in the alcohol dehydrogenase region of *Drosophila pseudoobscura*. Genetics **135:** 541–552.

Schmalhausen, I. I., 1946 *Factors of Evolution: The Theory of Stabilizing Selection*. The University of Chicago Press, Chicago (reprinted 1986).

Scott, D., 1986 Inhibition of female *Drosophila melanogaster* remating by a seminal fluid protein (esterase 6). Evolution **40:** 1084–1091.

Seager, R. D., and F. J. Ayala, 1982 Chromosome interactions in *Drosophila melanogaster*. I. Viability studies. Genetics **102:** 467–483.

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.

Smit-McBride, Z., A. Moya and F. J. Ayala, 1988 Linkage disequilibrium in natural and experimental populations of *Drosophila melanogaster*. Genetics **120:** 1043–1051.

Sokal, R. R., and F. J. Rohlf, 1981 *Biometry: The Principles and Practice of Statistics in Biological Research*, Ed. 2. W. H. Freeman, San Francisco.

Stephan, W., and D. Kirby, 1993 RNA folding in Drosophila shows a distance effect for compensatory fitness interactions. Genetics **135:** 97–103.

Strobeck, C., 1983 Expected linkage disequilibrium for a neutral locus linked to a chromosomal arrangement. Genetics **103:** 545–555.

Swofford, D. L., 1993 *PAUP: Phylogenetic Analysis Using Parsimony*, Version 3.1.1. Illinois Natural History Survey, Champaign, IL.

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

Thoday, J. M., J. B. Gibson and S. G. Spickett, 1964 Regular responses to selection. 2. Recombination and accelerated response. Genet. Res. **5:** 1–19.

Thompson, J. D., D. G. Higgins and T. J. Gibson, 1994 CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. **22:** 4673–4680.

Tyler, R. H., H. Brar, M. Singh, A. Latorre, J. L. Graves *et al.*, 1993 The effect of superoxide dismutase alleles on aging in Drosophila. Genetica **91:** 143–149.

Veuille, M., and L. M. King, 1995 Molecular basis of polymorphism at the Esterase-5B locus in *Drosophila pseudoobscura*. Genetics **141:** 255–262.

Voelker, R. A., C. C. Cockerham, F. M. Johnson, H. E. Schaffer, T. Mukai *et al.*, 1978 Inversions fail to account for allozyme clines. Genetics **88:** 515–527.

Waddington, C. H., 1957 *The Strategy of the Genes*. Allen & Unwin, London.

Watterson, G. A., 1975 On the number of segregating sites in genetical models without recombination. Theor. Popul. Biol. **10:** 256–276.

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

Zapata, C., and G. Alvarez, 1992 The detection of gametic disequilibrium between allozyme loci in natural populations of *Drosophila*. Evolution **48:** 1900–1917.

Zapata, C., and G. Alvarez, 1993 On the detection of nonrandom association between DNA polymorphisms in natural populations of *Drosophila*. Mol. Biol. Evol. **10:** 823–841.

Zouros, E., 1976 The distribution of enzyme and inversion polymorphism over the genome of *Drosophila*: evidence against balancing selection. Genetics **83:** 169–179.

Communicating editor: M. Slatkin