

Nucleotide Variation of the *Est-6* Gene Region in Natural Populations of *Drosophila melanogaster*

Evgeniy S. Balakirev^{*,†,‡} and Francisco J. Ayala^{*,1}

^{*}Department of Ecology and Evolutionary Biology, University of California, Irvine, California 92697-2525, [†]Institute of Marine Biology, Vladivostok 690041, Russia and [‡]Academy of Ecology, Marine Biology and Biotechnology, Far Eastern State University, Vladivostok 690600, Russia

Manuscript received February 27, 2003
Accepted for publication August 20, 2003

ABSTRACT

We have investigated nucleotide polymorphism in the *Est-6* gene region in four samples of *Drosophila melanogaster* derived from natural populations of East Africa (Zimbabwe), Europe (Spain), North America (California), and South America (Venezuela). There are two divergent sequence types in the North and South American samples, which are not perfectly (North America) or not at all (South America) associated with the *Est-6* allozyme variation. Less pronounced or no sequence dimorphism occurs in the European and African samples, respectively. The level of nucleotide diversity is highest in the African sample, lower (and similar to each other) in the samples from Europe and North America, and lowest in the sample from South America. The extent of linkage disequilibrium is low in Africa (1.23% significant associations), but much higher in non-African populations (22.59, 21.45, and 37.68% in Europe, North America, and South America, respectively). Tests of neutrality with recombination are significant in non-African samples but not significant in the African sample. We propose that demographic history (bottleneck and admixture of genetically different populations) is the major factor shaping the nucleotide patterns in the *Est-6* gene region. However, positive selection modifies the pattern: balanced selection creates elevated levels of nucleotide variation around functionally important (target) polymorphic sites (*RsaI*−/*RsaI*+ in the promoter region and F/S in the coding region) in both African and non-African samples; and directional selection, acting during the geographic expansion phase of *D. melanogaster*, creates an excess of very similar sequences (*RsaI*− and S allelic lineages, in the promoter and coding regions, respectively) in the non-African samples.

FROM the very beginning of the “allozyme era,” esterase 6 (*Est-6*) has been one of the most investigated and informative molecular markers in *Drosophila* population, evolutionary, and development genetics (reviewed by OAKESHOTT *et al.* 1989, 1993, 1995; KOROCHKIN *et al.* 1990; RICHMOND *et al.* 1990). WRIGHT (1961, 1963) described two main allozymes (Fast and Slow) of EST-6, showed their Mendelian inheritance, found a differential response to the organophosphate inhibitor, and raised questions about the adaptive significance of the polymorphism. The main allozymes show large-scale latitudinal clines (OAKESHOTT *et al.* 1981), with the Slow allozyme more common at higher latitudes. This, together with other data on the temporal and geographic allozyme variation in natural populations and results of laboratory experiments, suggests that the EST-6 polymorphism is maintained by some form of positive selection (reviewed by OAKESHOTT *et al.* 1989, 1993, 1995; RICHMOND *et al.* 1990).

The *Est-6* gene is on the left arm of chromosome 3 of *Drosophila melanogaster*, at cytogenetic map position

69A1–A3 (PROCUNIER *et al.* 1991). OAKESHOTT *et al.* (1987) first obtained the nucleotide sequence and characterized the exon-intron structure of the *Est-6* gene. Using available information on nine other eukaryotic esterases, they identified the active site and other functionally important regions of the gene. The coding region of *Est-6* is 1686 bp long and consists of two exons (1387 and 248 bp) and a small (51-bp) intron. The gene encodes the major β -carboxyl esterase (EST-6) that is transferred by *D. melanogaster* males to females in the seminal fluid during copulation (RICHMOND *et al.* 1980) and affects the female's consequent behavior and mating proclivity (GROMKO *et al.* 1984). The *Est-6* gene is duplicated (COLLET *et al.* 1990) but there is evidence that the adjacently located duplicate, referred to as *Est-P* (COLLET *et al.* 1990) or *Est-7* (DUMANCIC *et al.* 1997), may be a pseudogene (ψ *Est-6*, BALAKIREV and AYALA 1996; BALAKIREV *et al.* 2003; but see DUMANCIC *et al.* 1997). The β -esterase gene cluster in other *Drosophila* species also includes two (three in *D. pseudoobscura*) closely linked genes (YENIKOLOPOV *et al.* 1989; BRADY *et al.* 1990; EAST *et al.* 1990; OAKESHOTT *et al.* 1993, 1995).

The expression of *Est-6* in *D. melanogaster* has been investigated using *P*-element-mediated transformation

¹Corresponding author: Department of Ecology and Evolutionary Biology, 321 Steinhaus Hall, University of California, Irvine, CA 92697-2525. E-mail: fjayala@uci.edu

(LUDWIG *et al.* 1993; HEALY *et al.* 1996; TAMARINA *et al.* 1997). Within the ~ 1.2 kb of the 5'-flanking region, several independently acting *cis*-regulatory promoter elements that control the expression of the gene in different tissues have been identified. GAME and OAKESHOTT (1990) investigated restriction site polymorphism and its association with functional variation within a 21.5-kb region including the *Est-6* gene and found that a restriction polymorphism at an *RsaI* site in the 5'-flanking region of *Est-6* shows a significant association with male amount and activity of EST-6. Given other evidence showing that differences in male EST-6 activity affect the reproductive success of their mates (RICHMOND *et al.* 1990), GAME and OAKESHOTT (1990) concluded that *Est-6 cis*-acting regulatory polymorphisms may be important contributors to adaptive variation. Indeed, OAKESHOTT *et al.* (1994) and SAAD *et al.* (1994) have detected significant associations between the fitness components (preadult viability, development time, time to mating, remating frequency, egg production, and fertility) of *D. melanogaster* and the EST-6 activity level.

COOKE and OAKESHOTT (1989) sequenced the complete coding region of *Est-6* in 13 *D. melanogaster* lines in an Australian population (chosen so as to include all allozyme variants known in the population). They suggested that the main Fast and Slow allozymes differ by two amino acids (Asn/Asp at position 237 and Thr/Ala at position 247; but see HASSON and EANES 1996 and BALAKIREV *et al.* 1999) and considered these two amino acid replacements as the most likely targets for selection underlying the previously detected latitudinal clines (OAKESHOTT *et al.* 1981). ODGERS *et al.* (1995) sequenced 974 bp of the *Est-6* 5'-flanking region in *D. melanogaster* and identified a nucleotide substitution responsible for the *RsaI* polymorphism (T \rightarrow G at -531). They also revealed the presence of two highly diverged haplotype groups and a peak of polymorphism around the *RsaI* site. ODGERS *et al.* (1995) showed that the *RsaI*+ haplotype group yields $\sim 25\%$ more EST-6 enzyme activity in adult males than does the *RsaI*- haplotype and detected weak disequilibrium between the promoter polymorphism and the Fast/Slow allozyme polymorphism. Later, ODGERS *et al.* (2002) carried out *P*-element-mediated germ-line transformation, fusing representative promoter alleles to an identical *Est-6* coding region. They found a twofold difference in EST-6 activity in the male anterior sperm ejaculatory duct. ODGERS *et al.* (2002) also conducted restriction fragment length polymorphism (RFLP) and sequencing of the promoter region in populations from Africa, America, Asia, and Australia and detected significant deviation from neutral expectations in the non-African samples but not in the African one. HASSON and EANES (1996) investigated the nucleotide polymorphism of the *Est-6* coding region in 16 lines from disparate parts of the world, selected on the basis of the presence/absence of the cosmopolitan inversion *In (3L) Payne*, and de-

tected shared polymorphisms between *St* and *In (3L) Payne* chromosomes, indicating extensive genetic exchange between arrangements. BALAKIREV *et al.* (1999) sequenced 15 alleles of the *Est-6* coding region from a Californian population and found two highly differentiated haplotypes, one encompassing the Fast alleles and the other consisting of Slow alleles. They also detected a distinct peak of increased variation in the region surrounding the replacement site responsible for the EST-6 Fast/Slow allozyme polymorphism and suggested that balancing selection might be involved in the polymorphism. All these studies involve samples that are too small (COOKE and OAKESHOTT 1989; HASSON and EANES 1996; BALAKIREV *et al.* 1999), nonrandom, or both (COOKE and OAKESHOTT 1989; HASSON and EANES 1996) and thus unsuitable for certain population genetic tests (HUDSON *et al.* 1994; SIMONSEN *et al.* 1995).

We (BALAKIREV *et al.* 2002) increased the sample size and the length of the region sequenced to carry out significant tests of neutrality and to analyze the possible association between the regulatory and structural nucleotide polymorphism, seeking also to test for linkage disequilibrium within the gene region, a possibility suggested by the patterns observed in our previous study (BALAKIREV *et al.* 1999). We investigated the 5'-flanking, coding, and 3'-flanking regions of the *Est-6* gene (3062 bp total) in a random sample of 30 lines (and thus large enough for the population genetic tests; see HUDSON *et al.* 1994; SIMONSEN *et al.* 1995) of *D. melanogaster* from a natural population of California. We detected a highly structured pattern of variability, with distinctive features in the coding and 5'-flanking regions. We discovered two distinct allelic lineages for the promoter and coding region of the *Est-6* gene. The pattern of variability was complex and differed between the coding and the 5'-flanking regions, although the level of nucleotide diversity was very similar in the two regions. We detected strong linkage disequilibrium within the 5'-flanking region and *Est-6* coding region separately but it was much less pronounced between these two functional regions of the gene. The neutrality tests of KELLY (1997) and WALL (1999) incorporating recombination were highly significant for the studied regions. We suggested that the *Est-6* nucleotide polymorphism is shaped by a combination of directional and balancing selection acting on the promoter and coding region polymorphisms and by interactions between the two regions due to different degrees of hitchhiking (BALAKIREV *et al.* 2002).

We now present the analysis of nucleotide variation of the *Est-6* gene region in three additional samples of *D. melanogaster* derived from the natural populations of East Africa (Zimbabwe), Europe (Spain), and South America (Venezuela). The motivation for examining this gene in different populations is to analyze the pattern of nucleotide variation in the ancestral (African) and derived (European and American) *D. melanogaster* populations; we attempt further to clarify the question

concerning the evolutionary forces shaping the regulatory (*RsaI*+/*RsaI*-) and structural [Fast/Slow (F/S)] nucleotide polymorphisms. ODGERS *et al.* (1995, 2002) could not analyze the association between the regulatory and structural nucleotide polymorphisms, because they did not sequence the *Est-6* coding region in the same lines of *D. melanogaster* for which they obtained the promoter region sequences.

MATERIALS AND METHODS

Drosophila strains: *D. melanogaster* strains were derived from random samples of wild flies collected in Europe (Spain), North America (California), and South America (Venezuela). The strains were made fully homozygous for the third chromosome by crosses with balancer stocks, as described by SEAGER and AYALA (1982). The strains were named, in accordance with the electrophoretic alleles they carry for esterase-6 (the letter before the hyphen) and superoxide dismutase (the letter after the number), Ultra Slow (US), Slow (S), and Fast (F) (Figure 1). Chung-I Wu kindly provided the *D. melanogaster* strains from East Africa (Sengwa and Harare, Zimbabwe). The strain Zim S-44F (Zimbabwe) is from F. J. Ayala's laboratory.

DNA extraction, amplification, and sequencing: Methods are as previously described (BALAKIREV *et al.* 2003). The sequences of both strands were determined for each line, using 12 overlapping internal primers spaced, on average, 350 nucleotides. (See GenBank accessions AF526538–AF526559, AF150809–AF150815, AF147095–147102, and AF217624–AF217645). At least two independent PCR amplifications were sequenced for each polymorphic site in all *D. melanogaster* strains to prevent possible PCR and sequencing errors.

DNA sequence analysis: The *Est-6* sequences were assembled using the program SeqMan (Lasergene, 1994–1997; DNASTAR, Madison, WI). The computer programs DnaSP, version 3.4 (ROZAS and ROZAS 1999), and PROSEQ, version 2.4 (FILATOV and CHARLESWORTH 1999), were used for most intraspecific analyses. Departures from neutral expectations were investigated using KELLY's (1997) and WALL's (1999) tests on the basis of linkage disequilibrium between segregating sites and incorporating recombination. The permutation approach of HUDSON *et al.* (1992a,b) was used to estimate the significance of sequence differences between populations and haplotype families. Simulations based on the algorithms of the coalescent process with recombination (HUDSON 1990) were performed with the PROSEQ program to estimate the probabilities of the observed values of Kelly's Z_{ns} and Wall's B and Q statistics. The coalescent approach was also used to estimate confidence intervals for the nucleotide diversity values. The program Geneconv version 1.81 (SAWYER 1999) was used to detect gene conversion events. The population recombination rate was analyzed with the permutation-based approach (MCVEAN *et al.* 2002) on the basis of the approximate-likelihood coalescent method of HUDSON (2001).

RESULTS

Nucleotide polymorphism and recombination: The sequenced region consists of 3066 bp (2498 bp in the African sample). Figure 1 shows a total of 121 polymorphic sites (124 mutations because of three different nucleotides at each of positions 763, 1391, and 2396) in a sample of 78 sequences of the *Est-6* gene from four

populations of *D. melanogaster*: 45 sites (46 mutations) in the 5'-flanking region (3 sites, positions 329, 405, and 424, are associated with deletions), 49 sites (51 mutations) in exon I, 2 sites in the intron, 5 sites in exon II, and 20 sites in the intergenic region. Within the *Est-6* exons we detected 20 replacement and 34 synonymous polymorphic sites. Nine length polymorphisms, six deletions ($\blacktriangle 1$ – $\blacktriangle 6$), and three insertions ($\blacktriangledown 1$ – $\blacktriangledown 3$) occur within the whole sequenced region (Figure 1).

The length of the 5'-flanking region sequenced in the East-African sample is 619 bp but 1183 bp in the other samples. To obtain comparable estimates of nucleotide variation in all samples, we restrict the analysis of the 5'-flanking region to the 619 bp ("standard length") sequenced in all. Table 1 shows estimates of nucleotide diversity for the standard length of the *Est-6* gene and flanking regions. The π value for the full sequence is 0.0060 ± 0.0005 , which is within the range of values observed in other highly recombining gene regions of *D. melanogaster* (MORIYAMA and POWELL 1996). The π value is very similar in the 5'-flanking (0.0060 ± 0.0007) and *Est-6* regions (0.0057 ± 0.0005), but higher in the intergenic region (0.0094 ± 0.0018). The synonymous variation (0.0160) is 6.7 times higher than the nonsynonymous variation (0.0024) in the *Est-6* coding region. This sort of difference is expected if there is selective constraint on the *Est-6* nonsynonymous substitution rate. The level of silent divergence is at least 2.0 times higher for the *Est-6* gene than for the 5'-flanking or intergenic region (Table 1). The level of nucleotide diversity is highest in the African sample ($\pi = 0.0092 \pm 0.0008$) and lowest in the sample from South America ($\pi = 0.0034 \pm 0.0007$). Intermediate (and very similar) values of nucleotide diversity are observed in the European ($\pi = 0.0055 \pm 0.0008$) and North American ($\pi = 0.0060 \pm 0.0008$) samples (Table 1).

Previously, we detected in the California population lower polymorphism in the coding region of the S haplotypes than in that of the F haplotypes and lower polymorphism in the promoter region of the *RsaI*- haplotypes than in that of the *RsaI*+ haplotypes. We also noted that the "double sweep" (*RsaI*-/S) haplotypes (the haplotypes that have the more common mutations in both the promoter and coding region) were least variable (BALAKIREV *et al.* 2002). A similar tendency is observed in the East-African and European samples but not in the South American sample (Table 2). The South American sample is unique in the sense that it has no F allelic lineage at *Est-6* (see Figure 1). [We note that this population also lacks the S allele at the *Sod* locus (HUDSON *et al.* 1994).]

The method of HUDSON and KAPLAN (1985) reveals a minimum of 20 recombination events in the whole region analyzed: 3 for the 5'-flanking region, 16 for the *Est-6* gene, and 1 between them. The population recombination rate (MCVEAN *et al.* 2002) is 0.0216 for

B

ER S-5FA1-A. A.	+--C.....T.....A.G...A.
ER S-94FT.....A.GA...G.
ER S-174FG.T.....CA.GA...G
ER F-611FA2.....	G-T.....GG-T.....CA.GA...G
ER S-255S	TAT.A...C...A3
ER S-377F	TAT.A...C...A3
ER F-531FA2.....GG-T.....CA.GA...G
ER S-501SA1-A. A.	+--C.CT.GGT.G.TTG.CT.....A.G...A.
ER S-510SA1-A. A.	+--C.CT.GGT.G.TTG.CT.....A.G...A.
ER F-96SA1-A. A.	+--C.CT.GGT.G.TTG.CG.T...CCGG.T.CGCG...CTA.G.GT.A.
ER S-114SA1-A. A.	+--C.CT.GT.G.TTG.CG.....CTA.G
ER S-549SA1-A. A.	+--C.CT.GT.G.TTG.CG.....CTA.G
ER S-288SA1-A. A.	+--C.CT.GT.G.TTG.CG.....CTA.G
ER F-75FA1-A. AG	+--C.CT.GT.G.TTG.CT...T.....T.GG.CGC.CCTCTT...AA.CGC.G.A.C
ER F-357FC.T.TT.T...G.T...CCGG.T.CGCG...CTA.G.G.A.C.T...G.A...C
ER F-274FT.....G...CTA.G.G.A.C.T...G.A...C
ER F-517ST.....G...CTA.G.G.A.C.T...G.A...C
ER F-1461ST.....G...CTA.G.G.A.C.T...G.A...C
ER F-517FT.....G...CTA.G.G.A.C.T...G.A...C
Ven S-16FG.C.....
Ven S-5FGC.....
Ven S-18FG.C.....
Ven S-20FG.C.....
Ven S-17FG.C.....
Ven S-7FG.C.....
Ven S-8FG.C.....
Ven S-23FG.C.....
Ven S-14FG.C.....
Ven S-10FG.C.....
Ven S-22FG.C.....
Ven S-12FG.C.....
Ven S-15FT.....
Ven S-4FG.T.....CG.....
Ven S-13F	AT.A...A1-A. A.	+--C.C...C.....A6T.....C.A.....A.G.....
Ven S-21FA1-A. A.	+--C.CT.GT.G.TTG.CT...G.T...G.T...C...T.AG...G.GT.AG...G.G
Ven S-11FA1-A. A.	+--C.CT.GGT.G.TTG.CT...G.T...G.T...C...T.AG...G.GT.AG...G.G
Ven S-3FA1-A. A.	+--C.CT.GGT.G.TTG.CT...G.T...G.T...C...T.AG...G.GT.AG...G.G
Ven S-2FA1-A. A.	+--C.CT.GGT.G.TTG.CT...G.T...G.T...C...T.AG...G.GT.AG...G.G
Ven S-1FA1-A. A.	+--C.CT.GGT.G.TTG.CT...G.T...G.T...C...T.AG...G.GT.AG...G.G

FIGURE 1.—Continued.

TABLE 1
Nucleotide diversity and divergence in the *Est-6* gene region of *D. melanogaster*

	5'-flanking region (619 bp)	<i>Est-6</i>			Intergenic region (193 bp)	Full sequence	
		Syn (416 bp)	Nsyn (1274 bp)	Total (1686 bp)		Silent sites (1228 bp)	All sites (2498 bp)
Total (78 lines)							
π	0.0060	0.0160	0.0024	0.0057	0.0094	0.0097	0.0060
θ	0.0084	0.0187	0.0032	0.0066	0.0210	0.0137	0.0082
<i>K</i>	0.0514	0.1474	0.0213	0.0495	0.0707	0.0847	0.0519
Zim (12 lines)							
π	0.0126	0.0194	0.0035	0.0073	0.0148	0.0150	0.0092
θ	0.0127	0.0236	0.0037	0.0085	0.0189	0.0173	0.0103
<i>K</i>	0.0539	0.1454	0.0224	0.0499	0.0722	0.0856	0.0529
Bar (18 lines)							
π	0.0043	0.0169	0.0023	0.0058	0.0062	0.0087	0.0055
θ	0.0033	0.0146	0.0019	0.0048	0.0121	0.0082	0.0050
<i>K</i>	0.0504	0.1476	0.0213	0.0494	0.0709	0.0838	0.0516
ER (28 lines)							
π	0.0044	0.0152	0.0026	0.0057	0.0141	0.0094	0.0060
θ	0.0033	0.0156	0.0027	0.0056	0.0186	0.0095	0.0061
<i>K</i>	0.0500	0.1469	0.0215	0.0495	0.0716	0.0836	0.0516
Ven (20 lines)							
π	0.0044	0.0108	0.0013	0.0035	0	0.0056	0.0034
θ	0.0042	0.0089	0.0016	0.0033	0	0.0050	0.0033
<i>K</i>	0.0501	0.1493	0.0206	0.0495	0.0682	0.0841	0.0514

bp, the number of sites; π , the average number of nucleotide differences per site among all pairs of sequences (NEI 1987, p. 256); θ , 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); *K*, the average proportion of nucleotide differences between *D. melanogaster* and *D. simulans*, corrected according to JUKES and CANTOR (1969); Syn, synonymous; Nsyn, nonsynonymous. The segregating sites associated with indels are excluded from the π , θ , and *K* calculations. The 5'-flanking region is restricted to the 619 bp sequenced in all samples. Zim (Zimbabwe), Bar (Barcelona), ER (El Rio, California), and Ven (Venezuela) refer to the sampled localities.

the combined data set (Table 3), which is about three times less than the laboratory estimate of recombination rate (0.0664) based on the physical and genetic maps of *D. melanogaster* (J. M. COMERON, personal communication; COMERON *et al.* 1999; BALAKIREV *et al.* 2002). The rate of recombination is several times greater in the African than in the non-African samples (Table 3). The lowest recombination occurs in the South American sample. There is a positive correlation between nucleotide variation and recombination rate, as observed elsewhere (*e.g.*, BEGUN and AQUADRO 1992; see Tables 1 and 3).

The method of SAWYER (1989, 1999) detects gene conversion events within the *Est-6* gene in all samples except Venezuela. The number of significant fragments varies from 1 (Africa) to 14 (North America). The average length of fragments is 636 bp (range 314–1183 bp). The conversion events are less pronounced in the protein alignment (only 2 significant fragments, 1 in Africa and 1 in North America), which suggests the involve-

ment mostly of silent sites in significant fragments of the nucleotide alignment.

Haplotype structure and differentiation of populations: Maximum haplotype diversity occurs in East Africa ($H_{\text{div}} = 1.000$; no identical sequence pairs); less occurs in Europe ($H_{\text{div}} = 0.895$; 16 identical sequence pairs) and North America ($H_{\text{div}} = 0.947$; 20 identical sequence pairs); and the minimum occurs in South America ($H_{\text{div}} = 0.621$; 72 identical sequence pairs).

Figure 2 shows a neighbor-joining tree of the *Est-6* sequences (standard length). Due to recombination and gene conversion, this tree is not a good reflection of the genealogical process, but it serves to show the genetic structure of the data. The tree shows a relative absence of geographic structure: the sequences from a given population do not all group together. However, recombination has not completely erased all information, since there are two clusters of haplotypes related to *RsaI* polymorphism (data not shown). The first cluster includes the sequences with the *RsaI*- haplotypes (all

TABLE 2
Nucleotide diversity in different allelic lineages of the *Est-6* gene

Population	<i>RsaI</i> -	<i>RsaI</i> +	S	F	Double-sweep haplotypes	Other haplotypes
Zim	$N = 4, S = 4$ $\pi = 0.0033^{***\dagger}$ $\theta = 0.0036^{***\dagger}$ $D = -0.7801$	$N = 8, S = 19$ $\pi = 0.0123^{***\dagger}$ $\theta = 0.0122^{***\dagger}$ $D = -0.2164$	$N = 7, S = 25$ $\pi = 0.0053$ $\theta = 0.0061$ $D = -0.6953$	$N = 5, S = 28$ $\pi = 0.0077$ $\theta = 0.0080$ $D = -0.2444$	$N = 3, S = 16$ $\pi = 0.0043^{**}$ $\theta = 0.0043^{**}$ $D = \text{NA}$	$N = 9, S = 73$ $\pi = 0.0099^{***}$ $\theta = 0.0108^{***}$ $D = -0.4834$
Bar	$N = 14, S = 10$ $\pi = 0.0022$ $\theta = 0.0028$ $D = -0.7232$	$N = 4, S = 11$ $\pi = 0.0048$ $\theta = 0.0053$ $D = -0.8374$	$N = 13, S = 14$ $\pi = 0.0027$ $\theta = 0.0027$ $D = 0.0484$	$N = 5, S = 12$ $\pi = 0.0036$ $\theta = 0.0034$ $D = 0.3014$	$N = 12, S = 11$ $\pi = 0.0020^{***\dagger\dagger}$ $\theta = 0.0015^{***\dagger\dagger}$ $D = 1.4468^{**}$	$N = 6, S = 33$ $\pi = 0.0060^*$ $\theta = 0.0058^{***\dagger}$ $D = 0.2113$
ER	$N = 19, S = 10$ $\pi = 0.0012$ $\theta = 0.0026^*$ $D = -1.8648^{***\dagger}$	$N = 9, S = 14$ $\pi = 0.0050^{**}$ $\theta = 0.0050$ $D = 0.3769$	$N = 19, S = 16$ $\pi = 0.0021^{***\dagger}$ $\theta = 0.0027^{***\dagger}$ $D = -0.9273$	$N = 9, S = 30$ $\pi = 0.0070^*$ $\theta = 0.0066^*$ $D = 0.3613$	$N = 13, S = 16$ $\pi = 0.0013^{***\dagger\dagger}$ $\theta = 0.0021^{***\dagger\dagger}$ $D = -1.5061^{**}$	$N = 15, S = 52$ $\pi = 0.0079^{***\dagger\dagger}$ $\theta = 0.0064^{***\dagger}$ $D = 1.0034^*$
Ven	$N = 15, S = 11$ $\pi = 0.0016^{***\dagger}$ $\theta = 0.0030^{***\dagger\dagger}$ $D = -1.9446^{***\dagger}$	$N = 5, S = 1$ $\pi = 0.0004^{***\dagger}$ $\theta = 0.0004^{***\dagger}$ $D = -0.8165$	$N = 20, S = 20$ $\pi = 0.0035$ $\theta = 0.0033$ $D = 0.1344$	$N = 0$	$N = 15, S = 15$ $\pi = 0.0009^{**}$ $\theta = 0.0019^{**\dagger}$ $D = -2.0801^{***\dagger\dagger}$	$N = 5, S = 1$ $\pi = 0.0002$ $\theta = 0.0002^{**\dagger}$ $D = -0.8165$
Bar + ER	$N = 33, S = 15$ $\pi = 0.0017$ $\theta = 0.0033$ $D = -1.591^{**\dagger}$	$N = 13, S = 14$ $\pi = 0.0046^*$ $\theta = 0.0040$ $D = 0.5673$	$N = 32, S = 22$ $\pi = 0.0024^*$ $\theta = 0.0032^*$ $D = -0.8897$	$N = 14, S = 30$ $\pi = 0.0059^*$ $\theta = 0.0056$ $D = 0.2476$	$N = 25, S = 17$ $\pi = 0.0017^{***\dagger\dagger}$ $\theta = 0.0018^{***\dagger\dagger}$ $D = -0.1648$	$N = 21, S = 55$ $\pi = 0.0075^{***\dagger\dagger}$ $\theta = 0.0061^{***\dagger\dagger}$ $D = 0.8988$
Zim + Bar + ER	$N = 37, S = 7$ $\pi = 0.0011^{***\dagger}$ $\theta = 0.0027^{***\dagger\dagger}$ $D = -1.7418^{***\dagger}$	$N = 21, S = 19$ $\pi = 0.0067^{***\dagger\dagger}$ $\theta = 0.0051$ $D = -0.8864$	$N = 39, S = 36$ $\pi = 0.0030$ $\theta = 0.0088^{***\dagger}$ $D = -1.4082^*$	$N = 19, S = 41$ $\pi = 0.0070^*$ $\theta = 0.0070$ $D = -0.0039$	$N = 28, S = 26$ $\pi = 0.0020^{***\dagger\dagger}$ $\theta = 0.0027^{***\dagger\dagger}$ $D = -0.9292^*$	$N = 30, S = 88$ $\pi = 0.0086^{***\dagger\dagger}$ $\theta = 0.0090^{***\dagger\dagger}$ $D = -0.1673$

N , number of haplotypes; S , number of polymorphic sites. The data for the 5'-flanking region are based on 1183 bp, except for Zimbabwe and Zim + Bar + ER, which include only 619 bp. The haplotype analyses are based on the "standard" length (see RESULTS). The π and θ values are significantly lower or higher than the corresponding values in the counterpart group of haplotypes with the population recombination rate equal to 0.0216 (* $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$) or without recombination ($^\dagger P < 0.05$; $^\dagger\dagger P < 0.01$; $^\dagger\dagger\dagger P < 0.001$). TAJIMA's (1989) D is significant by the coalescent simulations with the population recombination rate equal to 0.0216 (* $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$) or without recombination ($^\dagger P < 0.05$; $^\dagger\dagger P < 0.01$; $^\dagger\dagger\dagger P < 0.001$).

strains from Ven S-10F at the top to ER F-1461S at the bottom); the second cluster contains the *RsaI*+ haplotypes (all strains from ER S-255S down to Ven S-2F). The *RsaI*-/*RsaI*+ clusters are even more apparent in the tree for the promoter region only (data not shown). If we restrict the analysis only to the coding region, the two clusters obtained differ to some extent (but not exclusively) with respect to the S and F haplotypes (data not shown).

ODGERS *et al.* (1995) described two groups of haplotypes for the 5'-flanking region of the *Est-6* gene of *D. melanogaster* from Australia. We detected two groups of haplotypes both for the *Est-6* gene (including the 5'-flanking region) and for the ψ *Est-6* putative pseudogene from North America (California; BALAKIREV and AYALA 1996; BALAKIREV *et al.* 1999, 2002, 2003). Two significantly divergent sequence types are also detected in South America (Figure 3A), where only the Slow *Est-6* allozyme occurs. The average number of nucleotide differences

(K) between the two haplotypes is 11.286. This is comparable with the differences between *RsaI*+/*RsaI*- ($K = 6.720$) and F/S ($K = 11.809$) allelic lineages in California (BALAKIREV *et al.* 2002). The permutation test (HUDSON *et al.* 1992a) of the Venezuelan haplotypes is highly significant, $K_{st}^* = 0.5867$ ($P = 0.0000$). Sequence dimorphism is less pronounced in the European sample (Figure 3B). The two divergent sequence types are not associated with *Est-6* allozyme variation (South America) or imperfectly associated (Europe, North America). The African sample (Figure 3C) has no clear sequence dimorphism (although all S haplotypes but one cluster together).

The estimates of population differentiation (HUDSON *et al.* 1992a) are fairly similar between the pairs Zim-Bar ($F_{st} = 0.0653$), Zim-ER ($F_{st} = 0.0398$), Bar-Ven ($F_{st} = 0.1093$), and ER-Ven ($F_{st} = 0.0920$) (for locality abbreviations see Figure 1). The maximal and minimal F_{st} values are obtained, respectively, for the pairs Zim-Ven ($F_{st} =$

TABLE 3
Recombination estimate

Population	R_m	Recombination (ρ)		ρ/θ
		Per gene	Per site	
Zim	15	64.329	0.0259	2.5146
Bar	5	13.226	0.0053	1.0606
ER	8	14.830	0.0059	0.9672
Ven	1	0.601	0.0002	0.0731
All populations	20	53.507	0.0216	2.6341

The population recombination rate, ρ , is $4N_e r$ (N_e is the effective population size and r is the recombination rate per nucleotide site per generation) obtained by the method of McVEAN *et al.* (2002). R_m is the minimum number of recombination events (HUDSON and KAPLAN 1985). All sites are included in the recombination analysis. The sequence analyzed includes 619 bp of the 5'-flanking region, the whole *Est-6* gene, and the intergenic region.

0.1508) and Bar-ER ($F_{st} = -0.0059$). We assess the statistical significance of the F_{st} values with the permutation method of HUDSON *et al.* (1992b), with 10,000 permutations. The differences are significant ($P < 0.05$) between Africa and all other samples (Europe, North America, and South America), a result consistent with other data (BEGUN and AQUADRO 1993, 1995). The differences between European and the North or South American samples are not significant ($P > 0.05$).

Sliding-window analysis: Figure 4 shows the distribution of polymorphism along the *Est-6* sequences. There is a distinct peak in the 5'-flanking region, which includes the *RsaI*+/*RsaI*- site (position 653 in Figure 1). ODGERS *et al.* (1995) detected this peak of variation in an Australian population of *D. melanogaster*. We also detected this peak (BALAKIREV *et al.* 1999, 2002, 2003) in the Californian population of *D. melanogaster*. Another distinct peak of variation occurs around the F/S site except in Venezuela. We detected this peak (BALAKIREV *et al.* 1999, 2002, 2003) in our Californian data and also in data of HASSON and EANES (1996) and COOKE and OAKESHOTT (1989) and suggested that it may reflect the effect of balancing selection (STROBECK 1983; HUDSON and KAPLAN 1988) between the F and S haplotypes, rather than within them (BALAKIREV *et al.* 2002). The absence of the peak in Venezuela may be a consequence of the absence of F haplotypes in this sample (Figure 1). The strong presence of both the promoter (*RsaI*+/*RsaI*-) and coding region (F/S) peaks in the African sample (Figure 4) suggests that these polymorphic sites were targets of balancing selection already in the African population (from which the others derive by migration).

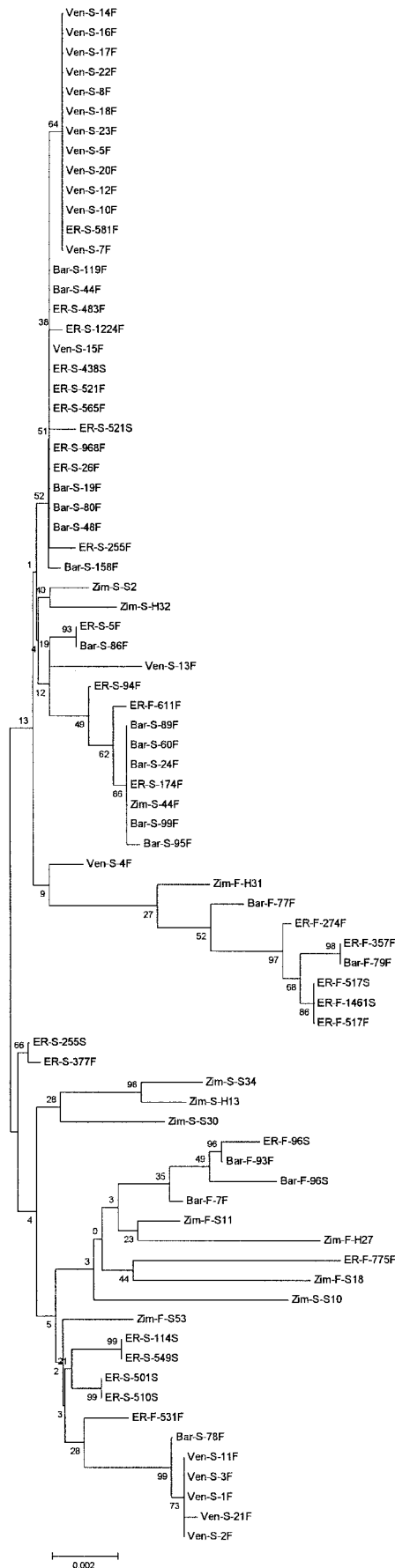
The valley regions located between the peaks of nucleotide variation are centered around positions 350, 1200, and 1800 (Figure 4). The first valley region includes nearly 400 bp upstream of the *Est-6* coding region. KAR-

OTAM *et al.* (1993, 1995) and ODGERS *et al.* (1995) detected strong conservation and low nucleotide variation of this region in *D. melanogaster*, *D. simulans*, and *D. mauritiana*. The region is under strong functional constraint because it contains several regulatory elements that are essential for *Est-6* expression (LUDWIG *et al.* 1993). Another valley region (1100–1300 bp) corresponds to amino acid residues Arg-159, Asp-181, and Ser-209 (codons at nucleotide sites 475–477, 541–543, and 625–627; positions 1094–1096, 1160–1162, and 1244–1246 in our coordinates). These residues (along with the surrounding sequences) are highly conserved in different esterases and are likely to be important for esterase enzymatic function (MYERS *et al.* 1988). A third valley region encompasses the potential N-linked glycosylation site, corresponding to codon position 1258–1260 (1877–1879 in our coordinates). The correspondence between the level of polymorphism and localities of functionally important sites implicated in the catalytic mechanism suggests that the observed valley regions reflect functional constraint.

We have measured heterogeneity in the distribution of silent polymorphic sites along the *Est-6* sequence and discordance between the level of within-*melanogaster* polymorphism and the *melanogaster-simulans* divergence by means of GOSS and LEWONTIN's (1996) and McDONALD's (1996, 1998) statistics and have assessed their significance by Monte Carlo simulations of the coalescent model incorporating recombination (McDONALD 1996, 1998). On the basis of 10,000 simulations, with the recombination parameters varying from 1 to 64, the tests are not significant for any of the separate samples or for the combined data set (data not shown).

Linkage disequilibrium: Linkage disequilibrium (LD) is measured by calculating the P value of Fisher's exact test in all pairwise comparisons between polymorphic sites. For the whole standard region (2498 bp) there are 1485 pairwise comparisons and 467 (31.45%) of them are significant. (With the Bonferroni correction, 11.92% remain significant; Bonferroni-corrected values are italicized in the ensuing sentences.) For the 5'-flanking region 25 of 78 (32.05%; 23.08%) pairwise comparisons are significant. For the *Est-6* coding region (including the intron) 219 of 528 (41.48%; 23.11%) comparisons are significant. There are 19.58% (1.17%) significant associations between the 5'-flanking region and the *Est-6* gene coding region. The proportion of pairs of sites with LD values significantly different from zero, at the 5% level, is much higher within the 5'-flanking region and *Est-6* coding region (244 of 606 pairwise comparisons) than between them (84 of 429, Fisher's exact test, $P < 0.001$; Fisher's criterion $F = 52.919$; $P < 0.001$). This observation corroborates our hypothesis (BALAKIREV *et al.* 2002) that the promoter and coded regions are subject to separate selection processes.

Linkage disequilibrium is notably low in the African sample: only 1.23% significant associations are in this



sample, but 22.59, 21.45, and 37.68% are in the European, North American, and South American samples, respectively. Figure 5 shows the distribution of *D* values along the whole region studied. A notable peak is around the F/S site and a less pronounced peak is around the *RsaI*-/*RsaI*+ site.

The significance of Pearson's correlation coefficient between LD and physical distance between sites is estimated by 10,000 permutations (McVEAN *et al.* 2002). For all samples, except South America, there is significant decline in LD with increasing distance (Table 4). The strong haplotype structure and pattern of linkage disequilibrium suggest that the South American population originated from a recent admixture of genetically differentiated populations.

Tests of neutrality: The tests of HUDSON *et al.* (1987), TAJIMA (1989), and DEPAULIS and VEUILLE (1998) do not reveal any significant deviation from neutrality for the *Est-6* gene region in any of the four populations of *D. melanogaster* (see also BALAKIREV *et al.* 2002). However, KELLY's (1997) *Z*_{ns} and WALL's (1999) *B* and *Q* tests detect significant deviations from neutrality in the non-African samples, with the population recombination rate ranging from 0.005 to 0.010 (Table 5; data for *B* and *Q* are not shown). The tests fail to detect any significant deviation from neutrality for the African sample, even when using 0.0664 as the recombination rate (laboratory estimate and based on the physical and genetic maps of *D. melanogaster*; J. M. COMERON, personal communication; COMERON *et al.* 1999; BALAKIREV *et al.* 2002), which is at least 2.5 times higher than the value of recombination obtained by the method of McVEAN *et al.* (2002) (Table 3). The significant values of Kelly's and Wall's statistics are grouped around the peaks of linkage disequilibrium and centered around the functionally important sites within both the 5'-flanking region (*RsaI* site) and the coding region (F/S polymorphism) of the *Est-6* gene (data not shown), which has been interpreted as evidence that these sites are targets of balancing selection (AYALA *et al.* 2002; BALAKIREV *et al.* 2002, 2003).

DISCUSSION

We have investigated nucleotide polymorphism in the *Est-6* gene region in four populations of *D. melanogaster* from Zimbabwe, Spain, California, and Venezuela. A dimorphic haplotype structure exists in the North American sample, which is not perfectly associated with the *Est-6* allozyme variation (S/F) and in South America, where there are no *Est-6* F haplotypes. The presence

FIGURE 2.—Neighbor-joining tree of the *Est-6* haplotypes of *D. melanogaster*, based on Kimura's two-parameter distance. The numbers are bootstrap probability values based on 10,000 replications. The trees are based on the *Est-6* standard length.

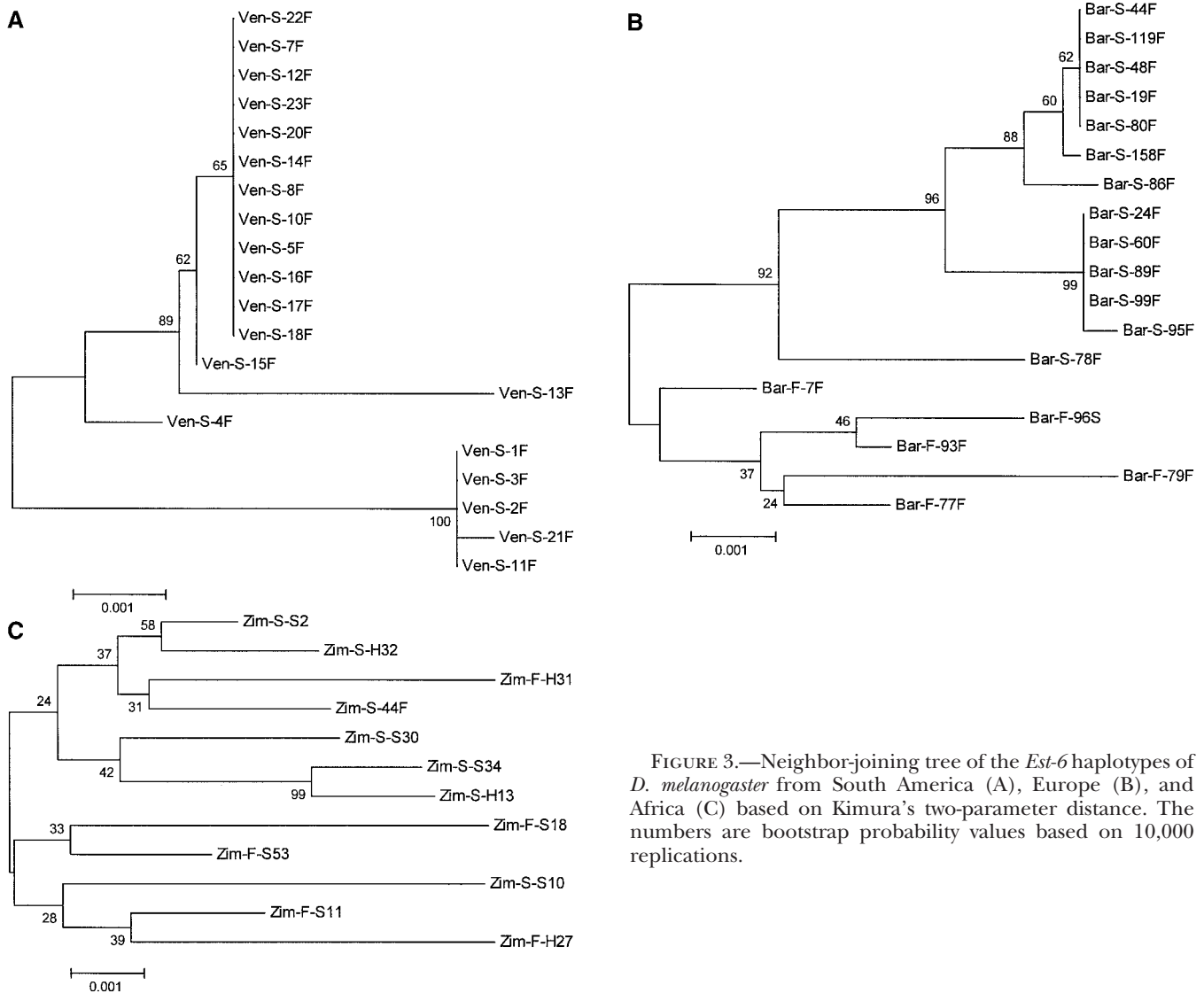


FIGURE 3.—Neighbor-joining tree of the *Est-6* haplotypes of *D. melanogaster* from South America (A), Europe (B), and Africa (C) based on Kimura's two-parameter distance. The numbers are bootstrap probability values based on 10,000 replications.

of two or more highly diverged haplotypes has been interpreted as a result of positive selection in *D. melanogaster* (see, e.g., HUDSON *et al.* 1994, 1997; BÉNASSI *et al.* 1999; LABATE *et al.* 1999). TEETER *et al.* (2000) investigated single-nucleotide polymorphism in 66 sequences of *D. melanogaster* spaced at 5- to 20-cM intervals and generated a map with no gaps greater than one-half of a chromosome arm (TEETER *et al.* 2000). Two-thirds of all sequences were dimorphic. If the dimorphism results from positive selection, TEETER *et al.* (2000) estimate that one site for every few kilobases would be subject to strong positive selection, which seems improbable. TEETER *et al.* (2000) suggest that admixture between two differentiated populations of *D. melanogaster* would account for and be a more appropriate explanation of the dimorphism. Suggestions of admixture have also been made on the basis of nucleotide sequencing, RFLP, and allozyme analyses of *D. melanogaster* populations

(e.g., DAVID and CAPY 1988; SINGH and LONG 1992; RICHTER *et al.* 1997; HASSON *et al.* 1998).

Our *Est-6* data are compatible with this proposal. We have found a strong dimorphic haplotype structure in three other *D. melanogaster* genes on the third chromosome, *Sod* (HUDSON *et al.* 1997), *tinman*, and *bagpipe* (E. S. BALAKIREV and F. J. AYALA, unpublished data), which may also have resulted from population admixture. Nevertheless, the *Est-6* data suggest that positive selection may also contribute to the observed patterns: balanced selection would account for the elevated nucleotide variation and linkage disequilibrium around the target polymorphic sites (*RsaI*⁻/*RsaI*⁺ in the promoter region and F/S in the coding region), while directional selection would yield an excess of very similar sequences exhibiting a very low level of variability (*RsaI*⁻ and S allelic lineages, in the promoter and coding region, respectively).

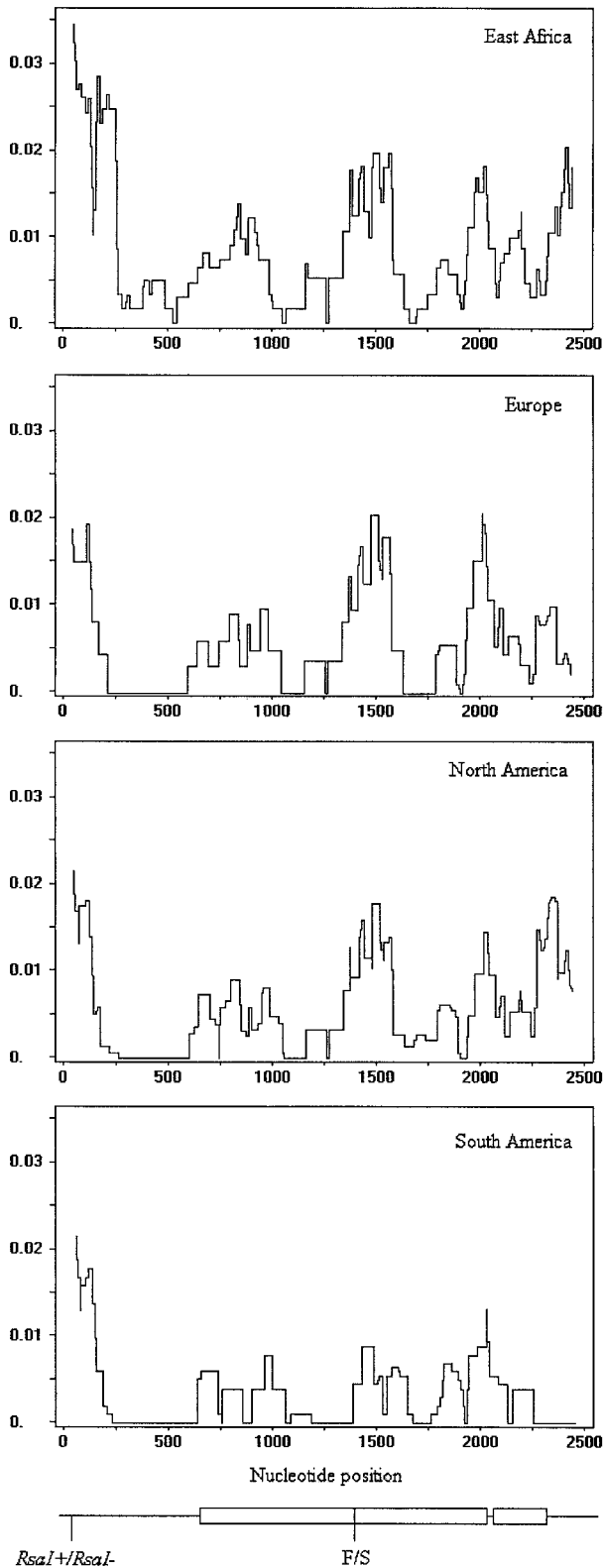


FIGURE 4.—Sliding-window plots of nucleotide diversity (π) along the *Est-6* gene region of *D. melanogaster*. A schematic of the *Est-6* gene is displayed at bottom. Exons are indicated by open boxes; the intron and the 5'- and 3'-flanking regions are shown by thin lines. Window sizes are 100 nucleotides with 1-nucleotide increments. The locations of the *RsaI* and allozyme polymorphisms are marked.

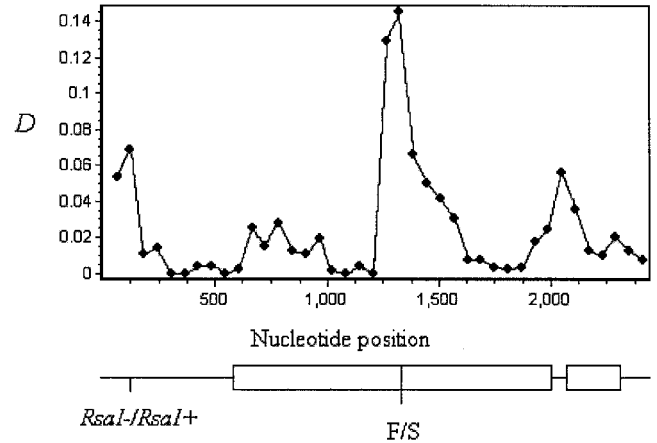


FIGURE 5.—Sliding-window plot of linkage disequilibrium (measured by D) along the *Est-6* gene region of *D. melanogaster*. A schematic of the *Est-6* putative pseudogene is displayed at bottom. Window sizes are 130 nucleotides with 60-nucleotide increments.

The African sample has the highest level of nucleotide diversity and the lowest level of linkage disequilibrium. The non-African samples show a pattern of haplotype distribution consistent with selective sweep hypotheses in the history of the species. The distribution of haplotype frequency in non-African samples is highly asymmetric: from a total of 66 sequences, 52 belong to the S haplotype and 14 belong to *RsaI*- haplotype. The haplotype test (HUDSON *et al.* 1994) is significant for the North and South American (excluding the recombinant strain Ven S-13F) samples, but not significant for the European sample. We conclude that bottlenecks have been an important evolutionary factor changing the genetic composition of colonizing *D. melanogaster* populations. The haplotype structure and polymorphism of the *Est-6* gene region are in accordance with the general pattern of relationships between the African and non-African populations of *D. melanogaster* (ANDOLFATTO 2001; AQUADRO *et al.* 2001). However, the peaks of nucleotide variation in the African sample, centered on functionally important sites (Figure 4), suggest that this population is not in mutation-drift equilibrium. The footprints of directional selection have been previously shown in African populations (*e.g.*, MOUSSET *et al.* 2003).

We found lower polymorphism in the S than in the F haplotypes (coding region) and lower polymorphism in the *RsaI*- than in the *RsaI*+ haplotypes (promoter region) in the California population (BALAKIREV *et al.* 2002). The same pattern occurs in the other populations (excluding Venezuela, where no F haplotypes occur), as well as in the total data set encompassing all four populations (Table 2): π is six times higher for the *RsaI*+ than for the *RsaI*- haplotypes; for the coding region, π is twice as large for the F as for the S haplotypes but double (0.00695) for the F haplotypes. Thus the

TABLE 4
Correlation between linkage disequilibrium and physical distance between the *Est-6* (full-sequence) polymorphic sites

Population	LD _{Fisher}	Pearson's correlation coefficient for	
		r^2	D'
Zim	0.0123	-0.0928 ($P = 0.0000$)	-0.0507 ($P = 0.0383$)
Bar	0.2259	-0.2918 ($P = 0.0000$)	-0.1328 ($P = 0.0108$)
ER	0.2145	-0.1525 ($P = 0.0004$)	-0.1963 ($P = 0.0000$)
Ven	0.3768	+0.0626 ($P = 0.9333$)	+0.0443 ($P = 0.8203$)
All	0.1293	-0.1419 ($P = 0.0000$)	-0.0603 ($P = 0.0091$)

LD_{Fisher} is the proportion of significant linkage disequilibrium revealed by Fisher's exact test using all polymorphic sites (including singletons).

lower variability among *RsaI*- and S haplotypes is not limited to the California population. But the differences are smaller in the African sample, which could indicate that the *RsaI*- and S haplotypes increased in frequency in Europe and America after their colonization.

We propose that the *RsaI*+ /F (zero-sweep) haplotypes may represent the ancestral condition (BALAKIREV *et al.* 2002). The frequency of these haplotypes is higher in Africa (0.333) than elsewhere (0.091). We also suggest that the *RsaI*- /S (double-sweep) haplotypes have evolved under directional selection, since they are less variable but more frequent in non-African samples (0.606) than in African (0.250). Directional selection, however, does not lead toward fixation of the double-sweep haplotypes in the derived populations because of balancing selection maintaining both divergent haplotypes (*RsaI*- /*RsaI*+ and F/S) in the promoter and coding regions (BALAKIREV *et al.* 2002).

The population data available suggest two different migrations of *D. melanogaster* during the expansion period from the African continent: (1) Africa → Europe → North America and (2) Africa → South America (see also DAVID and CAPY 1988; SINGH and LONG 1992).

The second migration is supported by the fact that the East-African and South American samples share a deletion (▲6, Figure 1) that is absent in other samples. This deletion is present in 5 of 12 East-African strains but absent in Europe and North America (Figure 1). Gaps constitute a valuable source of phylogenetic information (GIRIBET and WHEELER 1999). The absence of the F *Est-6* allele (and of the S *Sod* allele; HUDSON *et al.* 1994) also suggests that the South American population does not derive from Europe or America. The South American population might represent an admixture of migrants from North America and Africa. The most common haplotype (*RsaI*- /S) is from North America, while the haplotype *RsaI*+ /S clusters with most of the African strains (Figure 2). The admixture would have been recent, since the strong haplotype structure has not been eroded by recombination (linkage disequilibrium is highest in the South America sample).

We are grateful to G. McVean, D. A. Filatov, J. K. Kelly, J. H. McDonald, J. D. Wall, J. M. Comeron, F. Depaulis, and J. Rozas for useful advice on analyses and for providing computer programs. We thank Elena Balakireva, Andrei Tatarenkov, Victor DeFilippis, Martina Zurovkova, and Carlos Márquez for encouragement and help; and

TABLE 5
Kelly's (1977) test of neutrality for the *Est-6* gene region

	5'-flanking region (0.6 kb)			<i>Est-6</i> region (1.6 kb)			Entire region (2.4 kb)		
	Z _{ns}	0.05	0.01	Z _{ns}	0.05	0.01	Z _{ns}	0.05	0.01
All populations	0.106	0.05	NS	0.079	0.030	0.035	0.062	0.025	0.035
Zim	0.138	NS	NS	0.115	NS	NS	0.111	NS	NS
Bar	0.540	0.005	0.030	0.285	0.010	0.015	0.227	0.010	0.015
ER	0.458	0.005	0.020	0.204	0.010	0.015	0.181	0.010	0.015
Ven	0.453	0.010	0.025	0.386	0.010	0.010	0.417	0.005	0.005

The columns give KELLY's (1977) Z_{ns} values and the recombination rates at which the test becomes significant at the 5% (0.05) or 1% (0.01) level of significance. NS, not significant, even with the laboratory-estimated recombination rate (0.0664). Zim (Zimbabwe), Bar (Barcelona), ER (El Rio, California), and Ven (Venezuela) refer to the sampled localities.

W. M. Fitch, B. Gaut, R. R. Hudson, A. Long, and two anonymous reviewers for detailed and valuable comments. This work is supported by National Institutes of Health grant GM42397 to F. J. Ayala.

LITERATURE CITED

- 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., V. B. DUMONT and F. A. REED, 2001 Genome-wide variation in the human and fruitfly: a comparison. *Curr. Opin. Genet. Dev.* **11**: 627–634.
- AYALA, F. J., E. S. BALAKIREV and A. G. SÁEZ, 2002 Genetic polymorphism at two linked loci, *Sod* and *Est-6*, in *Drosophila melanogaster*. *Gene* **300**: 19–29.
- BALAKIREV, E. S., and F. J. AYALA, 1996 Is esterase-P encoded by a cryptic pseudogene in *Drosophila melanogaster*? *Genetics* **144**: 1511–1518.
- BALAKIREV, E. S., E. I. BALAKIREV, F. RODRIGUEZ-TRELLES and F. J. AYALA, 1999 Molecular evolution of two linked genes, *Est-6* and *Sod*, in *Drosophila melanogaster*. *Genetics* **153**: 1357–1369.
- BALAKIREV, E. S., E. I. BALAKIREV and F. J. AYALA, 2002 Molecular evolution of the *Est-6* gene in *Drosophila melanogaster*: contrasting patterns of DNA variability in adjacent functional regions. *Gene* **288**: 167–177.
- BALAKIREV, E. S., V. R. CHECHETKIN, V. V. LOBZIN and F. J. AYALA, 2003 DNA polymorphism in the β -*esterase* gene cluster of *Drosophila melanogaster*. *Genetics* **164**: 533–544.
- BEGUN, D. J., and C. F. AQUADRO, 1992 Levels of naturally occurring DNA polymorphism correlate with recombination rates in *Drosophila melanogaster*. *Nature* **356**: 519–520.
- BEGUN, D. J., and C. F. AQUADRO, 1993 African and North American populations of *Drosophila melanogaster* are very different at the DNA level. *Nature* **365**: 548–550.
- BEGUN, D. J., and C. F. AQUADRO, 1995 Evolution at the tip and base of the X chromosome in an African population of *Drosophila melanogaster*. *Mol. Biol. Evol.* **12**: 382–390.
- BÉNASSI, V., F. DEPAULIS, G. K. MEGHIAOUI and M. VEUILLE, 1999 Partial sweeping of variation at the *Fbp2* locus in a West African population of *Drosophila melanogaster*. *Mol. Biol. Evol.* **16**: 347–353.
- BRADY, J. P., R. C. RICHMOND and J. G. OAKESHOTT, 1990 Cloning of the esterase-5 locus from *Drosophila pseudoobscura* and comparison with its homologue in *D. melanogaster*. *Mol. Biol. Evol.* **7**: 525–546.
- 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.
- COMERON, J. M., M. KREITMAN and M. AGUADÉ, 1999 Natural selection on synonymous sites is correlated with gene length and recombination in *Drosophila*. *Genetics* **151**: 239–249.
- 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.
- DAVID, J. R., and P. CAPY, 1988 Genetic variation of *Drosophila melanogaster* natural populations. *Trends Genet.* **4**: 106–111.
- DEPAULIS, F., and M. VEUILLE, 1998 Neutrality tests based on the distribution of haplotypes under an infinite-site model. *Mol. Biol. Evol.* **15**: 1788–1790.
- DUMANCIC, M. M., J. G. OAKESHOTT, R. J. RUSSELL and M. J. HEALY, 1997 Characterization of the *EstP* protein in *Drosophila melanogaster* and its conservation in Drosophilids. *Biochem. Genet.* **35**: 251–271.
- EAST, P. D., A. GRAHAM and G. WHITTINGTON, 1990 Molecular isolation and preliminary characterisation of a duplicated esterase locus in *Drosophila buzzatii*, pp. 389–406 in *Ecological and Evolutionary Genetics of Drosophila*, edited by J. S. F. BARKER, W. STARMER and R. J. MACINTYRE. Plenum Press, New York.
- FILATOV, D. A., and D. CHARLESWORTH, 1999 DNA polymorphism, haplotype structure and balancing selection in the Leavenworthia PgiC locus. *Genetics* **153**: 1423–1434.
- GAME, A. Y., and J. G. OAKESHOTT, 1990 Associations between restriction site polymorphism and enzyme activity variation for esterase 6 in *Drosophila melanogaster*. *Genetics* **126**: 1021–1031.
- GIRIBET, G., and W. C. WHEELER, 1999 On gaps. *Mol. Phylogenet. Evol.* **13**: 132–143.
- GOSS, P. J. E., and R. C. LEWONTIN, 1996 Detecting heterogeneity of substitution along DNA and protein sequences. *Genetics* **143**: 589–602.
- 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.
- 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.
- HASSON, E., I. N. WANG, L. W. ZENG, M. KREITMAN and W. EANES, 1998 Nucleotide variation in the Triosephosphate isomerase (*Tpi*) locus of *Drosophila melanogaster* and *D. simulans*. *Mol. Biol. Evol.* **15**: 756–769.
- HEALY, M. J., M. M. DUMANCIC, A. CAO and J. G. OAKESHOTT, 1996 Localization of sequences regulating ancestral and acquired sites of esterase 6 activity in *Drosophila melanogaster*. *Mol. Biol. Evol.* **13**: 784–797.
- HUDSON, R. R., 1990 Gene genealogies and the coalescent process. *Oxf. Surv. Biol.* **7**: 1–44.
- HUDSON, R. R., 2001 Two-locus sampling distributions and their application. *Genetics* **159**: 1805–1817.
- HUDSON, R. R., and N. 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., 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., D. BOOS and N. L. KAPLAN, 1992a A statistical test for detecting geographic subdivision. *Mol. Biol. Evol.* **9**: 138–151.
- HUDSON, R. R., M. SLATKIN and W. P. MADDISON, 1992b Estimation of levels of gene flow from DNA sequence data. *Genetics* **132**: 583–589.
- HUDSON, R. R., K. BAILEY, D. SKARECKY, J. KWIAKOWSKI 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.
- JUKES, T. H., and C. R. CANTOR, 1969 Evolution of protein molecules, pp. 21–120 in *Mammalian Protein Metabolism*, edited by H. M. MUNRO. Academic Press, New York.
- 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. *Genetica* **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.
- KELLY, J. K., 1997 A test of neutrality based on interlocus associations. *Genetics* **146**: 1197–1206.
- KOROCHKIN, L., M. Z. LUDWIG, N. A. TAMARINA, I. USPENSKY, G. YENIKOLOPOV *et al.*, 1990 Molecular genetic mechanisms of tissue-specific esterase isozymes and protein expression in *Drosophila*, pp. 399–440 in *Isozymes: Structure, Function, and Use in Biology and Medicine*, edited by C. MARKERT and J. SCANDALIOS. Wiley-Liss, New York.
- LABATE, J. A., C. H. BIEMANN and W. F. EANES, 1999 Nucleotide variation at the *run1* locus in *Drosophila melanogaster* and *Drosophila simulans*. *Mol. Biol. Evol.* **16**: 724–731.
- LUDWIG, M. Z., N. A. TAMARINA and R. C. RICHMOND, 1993 Localization of sequences controlling the spatial, temporal, and sex-specific expression of the esterase 6 locus in *Drosophila melanogaster* adults. *Proc. Natl. Acad. Sci. USA* **90**: 6233–6237.
- MCDONALD, J. H., 1996 Detecting non-neutral heterogeneity across a region of DNA sequence in the ratio of polymorphism to divergence. *Mol. Biol. Evol.* **13**: 253–260.
- MCDONALD, J. H., 1998 Improved tests for heterogeneity across a region of DNA sequence in the ratio of polymorphism to divergence. *Mol. Biol. Evol.* **15**: 377–384.
- MCVEAN, G., P. AWADALLA and P. FEARNHEAD, 2002 A coalescent-based method for detecting and estimating recombination from gene sequences. *Genetics* **160**: 1231–1241.

- MORIYAMA, E. N., and J. R. POWELL, 1996 Intraspecific nuclear DNA variation in *Drosophila*. *Mol. Biol. Evol.* **13**: 261–277.
- MOUSSET, S., L. BRAZIER, M.-L. CARIOU, F. CHARTOIS, F. DEPAULIS *et al.*, 2003 Evidence of a high rate of selective sweeps in African *Drosophila melanogaster*. *Genetics* **163**: 599–609.
- MYERS, M., R. C. RICHMOND and J. G. OAKESHOTT, 1988 On the origins of esterases. *Mol. Biol. Evol.* **5**: 113–119.
- NEI, M., 1987 *Molecular Evolutionary Genetics*. Columbia University Press, New York.
- OAKESHOTT, J. G., G. K. CHAMBERS, J. B. GIBSON and D. A. WILLCOCKS, 1981 Latitudinal relationships of esterase-6 and phosphoglucosmutase gene frequencies in *Drosophila melanogaster*. *Heredity* **47**: 385–396.
- OAKESHOTT, J. G., C. COLLET, R. PHILLIS, K. M. NIELSEN, R. J. RUSSELL *et al.*, 1987 Molecular cloning and characterization of esterase 6, a serine hydrolase from *Drosophila*. *Proc. Natl. Acad. Sci. USA* **84**: 3359–3363.
- 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 PAPANRECHT, T. M. BOYCE, M. J. HEALY and R. J. RUSSELL, 1993 Evolutionary genetics of *Drosophila* esterases. *Genetica* **90**: 239–268.
- OAKESHOTT, J. G., M. SAAD, A. Y. GAME and M. J. HEALY, 1994 Causes and consequences of esterase 6 enzyme activity variation in pre-adult *Drosophila melanogaster*. *Heredity* **73**: 160–169.
- 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.
- ODGERS, W. A., C. F. AQUADRO, C. W. COPPIN, M. J. HEALY and J. G. OAKESHOTT, 2002 Nucleotide polymorphism in the *Est6* promoter, which is widespread in derived populations of *Drosophila melanogaster*, changes the level of Esterase 6 expressed in the male ejaculatory duct. *Genetics* **162**: 785–797.
- 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., D. G. GILBERT, K. B. SHEEHAN, M. H. GROMKO and F. M. 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 Press, New York.
- RICHTER, B., M. LONG, R. C. LEWONTIN and E. NITASAKA, 1997 Nucleotide variation and conservation of the *dpp* locus, a gene controlling early development in *Drosophila*. *Genetics* **145**: 311–323.
- ROZAS, J., and R. ROZAS, 1999 DnaSP version 3: an integrated program for molecular population genetics and molecular evolution analysis. *Bioinformatics* **15**: 174–175.
- SAAD, M., A. Y. GAME, M. J. HEALY and J. G. OAKESHOTT, 1994 Associations of esterase 6 allozyme and activity variation with reproductive fitness in *Drosophila melanogaster*. *Genetica* **94**: 43–56.
- SAWYER, S. A., 1989 Statistical tests for detecting gene conversion. *Mol. Biol. Evol.* **6**: 526–538.
- SAWYER, S. A., 1999 *GENECONV: A Computer Package for the Statistical Detection of Gene Conversion*. Department of Mathematics, Washington University, St. Louis.
- 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. J. AQUADRO, 1995 Properties of statistical tests of neutrality for DNA polymorphism data. *Genetics* **141**: 413–429.
- SINGH, R. S., and A. D. LONG, 1992 Geographic variation in *Drosophila*: from molecules to morphology and back. *TREE* **7**: 340–345.
- STROBECK, C., 1983 Expected linkage disequilibrium for a neutral locus linked to a chromosomal arrangement. *Genetics* **103**: 545–555.
- TAJIMA, F., 1989 Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* **123**: 585–595.
- TAMARINA, N. A., M. Z. LUDWIG and R. C. RICHMOND, 1997 Divergent and conserved features in the spatial expression of the *Drosophila pseudoobscura esterase-5B* gene and the *esterase-6* gene of *Drosophila melanogaster*. *Proc. Natl. Acad. Sci. USA* **94**: 7735–7741.
- TEETER, K., M. NAEEMUDDIN, R. GASPERINI, E. ZIMMERMAN, K. P. WHITE *et al.*, 2000 Haplotype dimorphism in a SNP collection from *Drosophila melanogaster*. *J. Exp. Zool.* **288**: 63–75.
- WALL, J. D., 1999 Recombination and the power of statistical tests of neutrality. *Genet. Res.* **74**: 65–79.
- WATTERSON, G. A., 1975 On the number of segregating sites in genetical models without recombination. *Theor. Popul. Biol.* **10**: 256–276.
- WRIGHT, T. R. F., 1961 The genetic control of an esterase in *Drosophila melanogaster*. *Am. Zool.* **1**: 476.
- WRIGHT, T. R. F., 1963 The genetics of an esterase in *Drosophila melanogaster*. *Genetics* **48**: 787–801.
- YENIKOLOPOV, G. N., O. A. MALEVANTSCHUK, N. I. PEUNOVA, P. V. SERGEEV and G. P. GEORGIEV, 1989 *Est* locus of *Drosophila virilis* contains two related genes. *Dokl. Acad. Nauk SSSR* **306**: 1247–1249 (in Russian).

Communicating editor: M. ASMUSSEN