Genetica **121:** 165–179, 2004.

The *β-esterase* gene cluster of *Drosophila melanogaster*: is ψ*Est-6* a pseudogene, a functional gene, or both?

Evgeniy S. Balakirev[1,2] & Francisco J. Ayala[1,*]
[1]*Department of Ecology and Evolutionary Biology, 321 Steinhaus Hall, University of California, Irvine, CA 92697-2525, USA;* [2]*Institute of Marine Biology, Vladivostok 690041, Russia and Academy of Ecology, Marine Biology, and Biotechnology, Far Eastern State University, Vladivostok 690600, Russia;* **Author for correspondence (Phone: +1-949-824-8293; Fax: +1-949-824-2474; E-mail: fjayala@uci.edu)*

Received 4 June 2003 Accepted 9 October 2003

Key words: balancing selection, demographic history, directional selection, DNA polymorphism, *Drosophila melanogaster*, epistatic selection, *Est-6*, ψ*Est-6*

Abstract

Pseudogenes have been defined as non-functional sequences of genomic DNA that are originally derived from functional genes, but exhibit degenerative features such as premature stop codons and frameshifts that prevent their expression. However, there is increasing evidence that pseudogenes are often evolutionarily conserved and may have retained some functional role or acquired new ones. Pseudogenes may exhibit non-functional features as well as functional ones. We investigate, as a model case, the *β-esterase* gene cluster of *Drosophila melanogaster* that includes the *Est-6* gene and the ψ*Est-6* putative pseudogene. We study four samples derived from natural populations of east Africa (Zimbabwe), Europe (Spain), North America (California), and South America (Venezuela). The level of nucleotide diversity is higher in Africa than in the non-African populations. There is twice more nucleotide diversity in ψ*Est-6* than in *Est-6*. Linkage disequilibrium within the *β-esterase* gene cluster is strong in non-African samples, but much lower in Africa. The population recombination rate is the same for ψ*Est-6* and *Est-6* in Africa, but significantly different in non-African samples. Intragenic gene conversion events are detected within *Est-6* and, with much higher incidence, within ψ*Est-6*; intergenic gene conversion events are rare. The extensive intragenic gene conversion within ψ*Est-6* can be explained by the invasion of retrotransposons that promote a form of homology-dependent gene conversion upon excision. Tests of neutrality with recombination are significant for the *β-esterase* gene cluster in the non-African populations but not in Africa. The *Est-6* gene sequences exhibit a well-known allozyme dimorphic structure. The sequences of ψ*Est-6* are also dimorphic in North and South America, but they do not correspond at all (South America) or only imperfectly (North America) to the *Est-6* allozyme dimorphism. Sequence dimorphism is less pronounced in the European and African samples. We suggest that demographic history (bottleneck and admixture of genetically differentiated populations) is the major factor shaping the nucleotide pattern in the *β-esterase* gene cluster. However, there are some clear indications of positive selection shaping the distribution of nucleotide polymorphism within the cluster. Intergenic epistatic selection may play an important role in the evolution of the *β-esterase* gene cluster, preserving ψ*Est-6* from degenerative destruction and reflecting its functional interaction with *Est-6*. The *Est-6* gene cluster of *D. melanogaster* represents an example of a functionally interacting complex ('intergene') in which two components (*Est-6* and ψ*Est-6*) or more are required to perform the final function.

Introduction

Pseudogenes have been defined as non-functional sequences of genomic DNA that are originally derived from functional genes, or as 'genes that are no longer expressed but bear sequence similarity to active genes' (Hartl and Clark, 1989, p. 114). Pseudogenes exhibit such degenerative features as premature stop codons

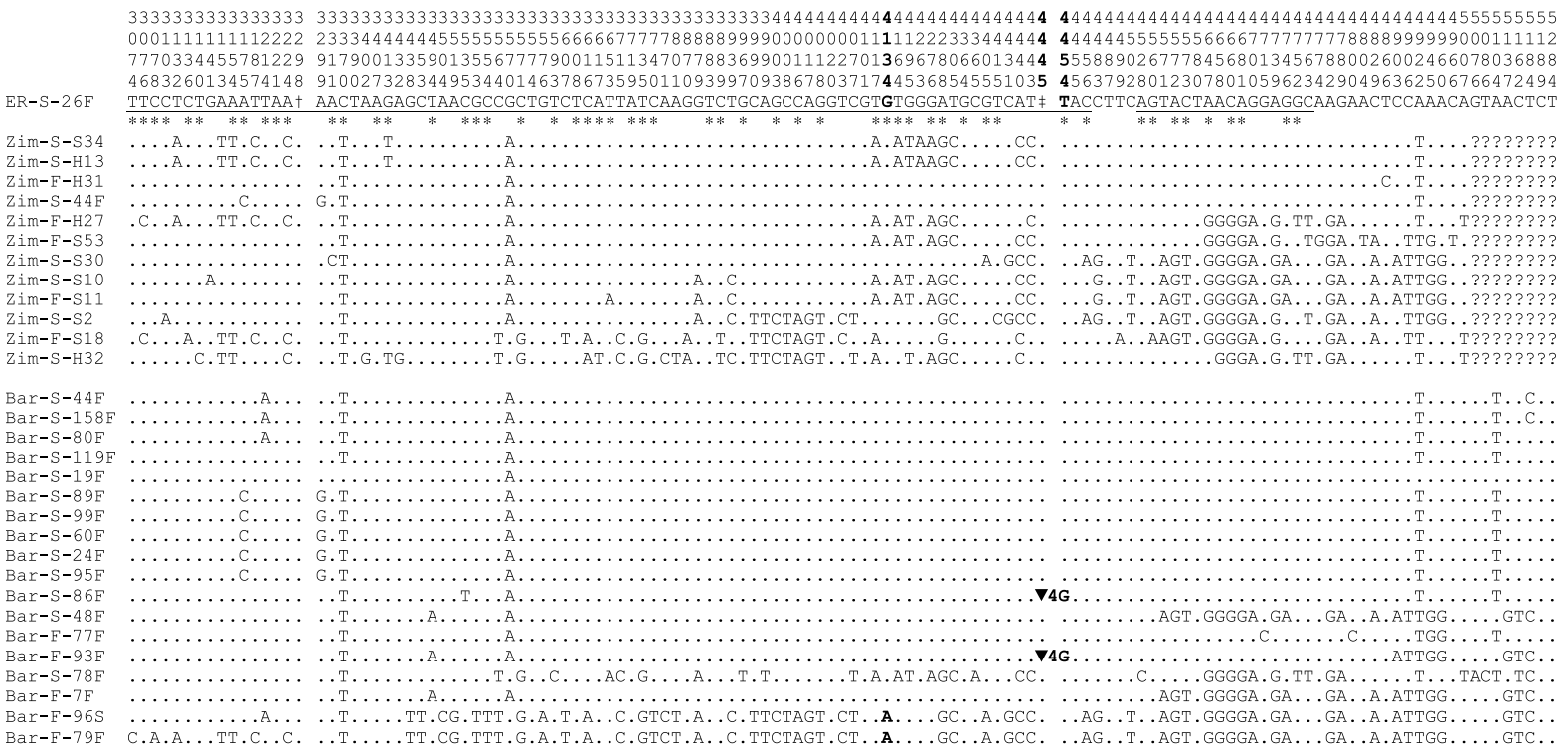

Figure 1. DNA polymorphism of ψ*Est-6* in 78 homozygous lines of *D. melanogaster*. The lines from east Africa, Zimbabwe (Zim), Spain, Barcelona (Bar), USA, California (ER), and Venezuela, Caracas (Ven) are presented sequentially, grouped in accordance with the extent of the genetic similarity based on the full sequence. The S and F after the numbers refer to the allozyme polymorphism at the *Sod* locus; these numbers and letters have been previously used to tag these lines. The S and F letters before the line numbers refer to the EST-6 allozymes, slow and fast. The second letter in the designations of the Zimbabwe lines refers to the locality of collection (Sengwa and Harare; the exception is line Zim-S-44F, which is from Ayala's laboratory). The numbers above the top sequence (ER-S-26F) represent the position of segregating sites and the start of a deletion or insertion. Nucleotides are numbered from the beginning of our sequence (position 32 in Collet et al., 1990). The coding regions (exon I and exon II) of ψ*Est-6* are underlined below the reference sequence. Amino acid replacement polymorphisms are marked with asterisks. Seventeen premature stop codons are due to single nucleotide polymorphisms G → A (site 4134, strains Bar-F-96S, Bar-F-79F, ER-F-357F, ER-F-274F, ER-F-517F, ER-F-517S, and ER-F-1461S) and T → G (site 4454, strains Bar-F-93F, Bar-S-86F, ER-S-501S, ER-S-510S and ER-S-5F) or a 9-bp insertion of ACATTTGAT (position 4445–4453, strains Bar-F-93F, Bar-S-86F, ER-S-501S, ER-S-510S and ER-S-5F). These sites are marked by bold face. The hyphens represent deleted nucleotides. The question marks indicate missing data. ▲ denotes a deletion; † denotes the absence of a deletion; ▼ denotes an insertion; ‡ denotes the absence of an insertion. ▲7 denotes a 3-bp deletion of CAG; ▼4 denotes a 9-bp insertion of ACATTTGAT.

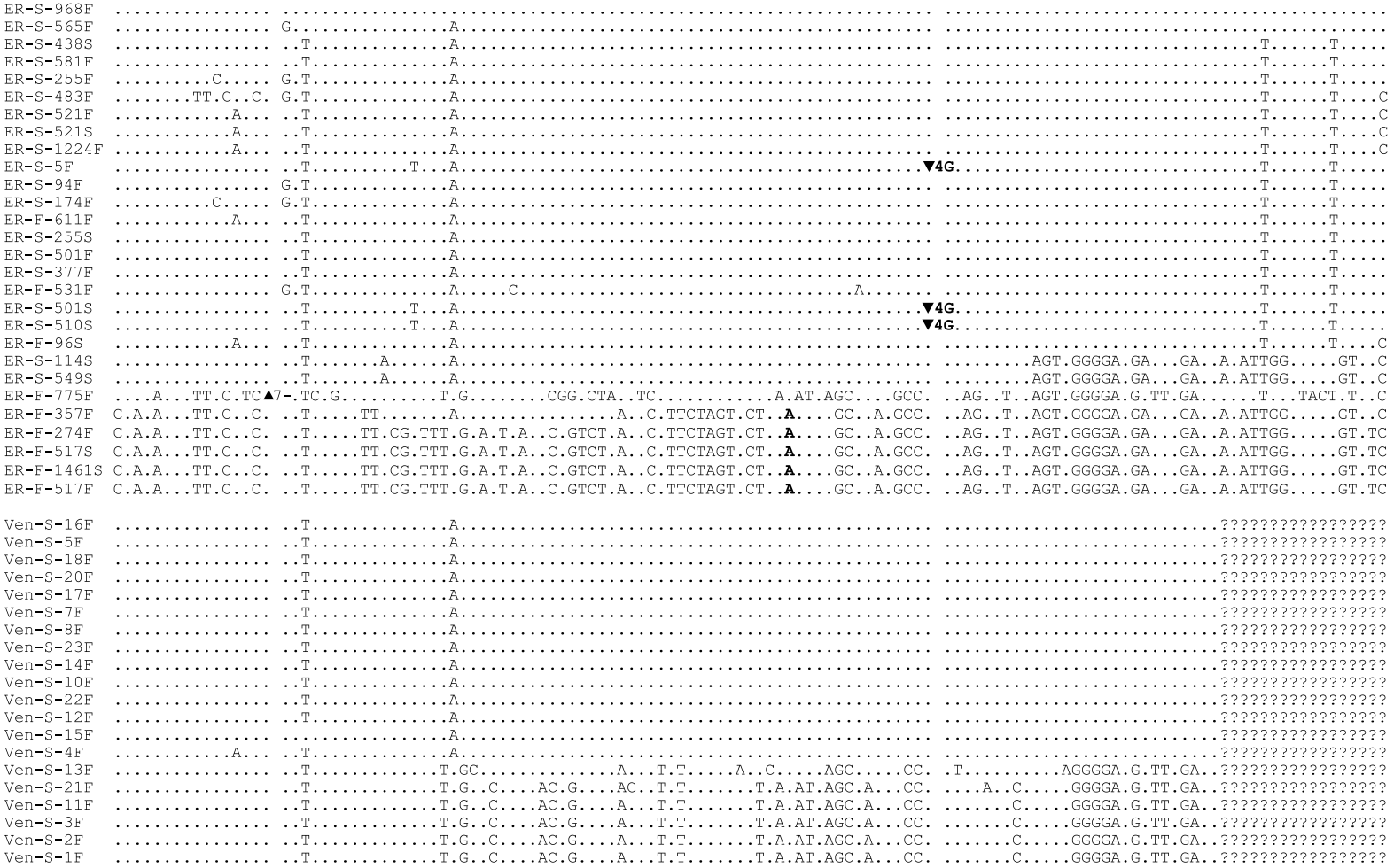

Figure 1. (*continued*)

and frameshift mutations that prevent their expression. Pseudogenes often arise by tandem duplication of genes, with ensuing loss of function as a result of gradual accumulation of disabling mutations. It is assumed that pseudogene mutations (disabling or not) will not be subject to purifying selection and will all have equal probability of becoming fixed in the population (Kimura, 1980; Li, Gojobori and Nei, 1981; Graur and Li, 2000, p. 124). It follows that pseudogenes will generally degenerate, owing to the rapid accumulation of recurrent mutations, and melt into the background of the surrounding DNA (e.g., Graur, Shuali and Li, 1989). However, eukaryotes genomes contain many pseudogenes that appear to have avoided full degeneration (Mighell et al., 2000; Harrison and Gerstein, 2002; Balakirev and Ayala, 2003a).

The *Esterase*-6 (*Est-6*) and the putative pseudogene ψ*Est-6* of *Drosophila melanogaster* are two closely linked genes (separated by 193 bp), which make up the *β-esterase* gene cluster on the left arm of chromosome 3 at 68F7-69A1 in the cytogenetic map (but see Procunier, Smith and Richmond, 1991). The coding regions of these tandemly duplicated genes are 1686 and 1691 bp long, respectively, consisting of two exons (1387 and 248 bp) separated by a small (51 bp in *Est-6* and 56 bp in ψ*Est-6*) intron. The *Est-6* gene encodes the major β-carboxylesterase (EST-6) that is transferred by *D. melanogaster* males to females in the seminal fluid during copulation (Richmond et al., 1980, 1990) and affects the female's consequent behavior and mating proclivity (Gromko, Gilbert and Richmond, 1984). The ψ*Est-6* gene was first named *Est-P* by Collet et al. (1990) who concluded that it was a functional gene, on the evidence of transcriptional activity, intact splicing sites, no premature termination codons, and the presence of initiation and termination codons. However, Balakirev and Ayala (1996) found premature stop codons within the gene's coding region and some other indications suggesting that it might be in fact a pseudogene, which was named ψ*Est-6* (see also Balakirev et al., 2003). Dumancic et al. (1997) showed that some alleles produce a catalytically active esterase corresponding to the previously identified EST-7 isozyme (Healy, Dumancic and Oakeshott, 1991) and proposed to name it *Est-7*. The gene duplication is relatively ancient, since two (three in *D. pseudoobscura*) closely linked genes are also present in the *β-esterase* gene cluster of other *Drosophila* species (Yenikolopov et al., 1989; Brady, Richmond and Oakeshott, 1990; East, Graham and Whitington, 1990; Oakeshott et al., 1993, 1995).

We previously investigated nucleotide variability in *Est-6* and ψ*Est-6*, as well as linkage disequilibrium and other population parameters within each and between the two genes, in a natural population of *D. melanogaster* from California (Balakirev and Ayala, 1996; Balakirev et al., 1999, 2003; Ayala, Balakirev and Sáez, 2002; Balakirev, Balakirev and Ayala, 2002). We have also investigated the nucleotide polymorphism of *Est-6* in four natural populations from different continents: Africa, Europe, North and South America (Balakirev and Ayala, 2003b). We now extend the investigation of these four populations to the *β-esterase* gene cluster. We have sequenced ψ*Est-6* and the intergenic regions in the same chromosomal strains (78 in total) for which we have the *Est-6* information. The full sequence length of 5394 bp includes the 5′-flanking region, *Est-6*, intergenic region, ψ*Est-6*, and 3′-flanking region.

We conclude that ψ*Est-6* exhibits features of both functional and non-functional genes (Balakirev and Ayala, 2003a). We, moreover, propose that the *β-esterase* gene cluster might be considered an 'intergene,' that is, a functionally interacting complex in which two (or more) gene components are jointly required to carry out a functional role.

Materials and methods

The strains are derived from *D. melanogaster* flies, collected in Zimbabwe (east Africa), Barcelona Spain), El Rio (Acampo, California) and Caracas (Venezuela). The procedures of DNA extraction, amplification, and sequencing have been described (Balakirev and Ayala, 1996; Balakirev et al., 1999, 2003; Balakirev, Balakirev and Ayala, 2002). The esterase sequences were assembled using the program SeqMan (Lasergene, DNASTAR, Inc., 1994–1997). The computer programs DnaSP, version 3.4 (Rozas and Rozas, 1999) and PROSEQ, version 2.4 (Filatov and Charlesworth, 1999) were used for the 'sliding window' (Hudson and Kaplan, 1988), and other analyses. Departures from neutral expectations were investigated taking recombination into account (Kelly, 1997; Wall, 1999); and sequence differences between haplotype families were tested with the permutation approach of Hudson, Boos and Kaplan (1992) and Hudson, Slatkin and Maddison (1992). The DnaSP and PROSEQ programs were used for simulations based on the algorithms of the coalescent process with or without recombination (Hudson,

Table 1. Nucleotide diversity and divergence in the *β-esterase* gene cluster of *D. melanogaster*

N	5′-flanking 619	*Est-6*			Intergenic 193	*ψEst-6*			3′-flanking 248	Full sequence		
		Syn 379	Nsyn 1253	Total 1686		Syn 378	Nsyn 1248	Total 1700		Syn 757	Nsyn 2501	Total 4198
Total 78 lines												
π	0.0060	0.0160	0.0024	0.0057	0.0094	0.0244	0.0076	0.0115	–	0.0202	0.0050	0.0083
θ	0.0084	0.0187	0.0032	0.0066	0.0210	0.0231	0.0086	0.0123	–	0.0209	0.0059	0.0098
k	3.589			9.579	1.816			19.446	–			34.431
K	0.0514	0.1474	0.0213	0.0495	0.0707	0.1388	0.0307	0.0540	–	0.1431	0.0260	0.0530
Zim 12 lines												
π	0.0126	0.0194	0.0035	0.0073	0.0148	0.0296	0.0097	0.0145	0.0184	0.0245	0.0066	0.0113
θ	0.0127	0.0236	0.0037	0.0085	0.0189	0.0271	0.0093	0.0135	0.0160	0.0253	0.0065	0.0116
k	7.591			12.303	2.848			24.485	4.561			47.227
K	0.0539	0.1454	0.0224	0.0499	0.0722	0.1360	0.0297	0.0530	0.0693	0.1407	0.0260	0.0529
Bar 18 lines												
π	0.0043	0.0169	0.0023	0.0058	0.0062	0.0208	0.0068	0.0101	0.0134	0.0189	0.0046	0.0073
θ	0.0033		0.0019	0.0048	0.0121	0.0254	0.0088	0.0127	0.0115	0.0200	0.0053	0.0081
k	2.621	0.0146		9.810	1.203			17.033	3.310			30.667
K	0.0504	0.1476	0.0213	0.0494	0.0709	0.1442	0.0314	0.0556	0.0667	0.1459	0.0263	0.0532
El Rio 28 lines												
π	0.0044	0.0152	0.0026	0.0057	0.0141	0.0268	0.0076	0.0122	0.0123	0.0210	0.0051	0.0085
θ	0.0033	0.0156	0.0027	0.0056	0.0186	0.0224	0.0080	0.0114	0.0093	0.0190	0.0053	0.0082
k	2.730			9.542	2.714			20.534	3.061			35.521
K	0.0500	0.1469	0.0215	0.0495	0.0716	0.1393	0.0311	0.0546	0.0673	0.1431	0.0263	0.0528
Ven 20 lines												
π	0.0044	0.0108	0.0013	0.0035	0	0.0171	0.0053	0.0078	–	0.0139	0.0033	0.0052
θ	0.0042	0.0089	0.0016	0.0033	–	0.0134	0.0041	0.0063	–	0.0112	0.0028	0.0045
k	2.668			5.837	–			13.247	–			21.753
K	0.0501	0.1493	0.0206	0.0495	–	0.1419	0.0302	0.0541	–	0.1456	0.0254	0.0525

N is the number of sites. π is the average number of nucleotide differences per site among all pairs of sequences. 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. *k* is the average number of nucleotide differences. *K* is the average proportion of nucleotide differences between *D. melanogaster* and *D. simulans*, corrected according to Jukes and Cantor (1969). Zim (Zimbabwe), Bar (Barcelona), ER (El Rio, California), and Ven (Venezuela) refer to the sampled localities. Syn, synonymous; Nsyn, non-synonymous. The segregating sites associated with indels are excluded from the π, θ, and *K* calculations. The 5′- and 3′-flanking regions are restricted to 619 bp and 248 bp, respectively, to obtain equal lengths for all populations. The full sequence does not include the 3′-flanking region, because this region is not analyzed in the Venezuela population (where it includes only 61 bp).

1990) in order to estimate the probabilities of Kelly's Z_{nS} and Wall's *B* and *Q* statistics and the confidence intervals for the nucleotide diversity values. The program Geneconv version 1.81 (Sawyer, 1999) was used to detect intra- and intergenic conversion events (Sawyer, 1989). The population recombination rate was analyzed by the permutation-based method of McVean, Awadalla and Fearnhead (2002) based on approximate-likelihood coalescent method (Hudson, 2001). The GenBank accessions for the sequences are AF147095-AF147102; AF150809-AF150815; AF217624-AF217645; AF526538-AF526559; AY247664-AY247713; AY247987-AY248036.

Results

Nucleotide polymorphism and recombination

There are 236 polymorphic nucleotide sites among the 78 sequences in the 5394 bp of the *β-esterase* gene cluster: 45 sites in the 5′-flanking region (three sites

are associated with deletions), 49 sites in exon I, two sites in the intron, and five sites in exon II of *Est-6*; 20 sites in the intergenic region; 83 sites in exon I (one site is associated with a deletion), four sites in the intron, 16 sites in exon II of ψ*Est-6*; and 22 sites in the 3′-flanking region. Figure 1 shows the polymorphic sites in ψ*Est-6* and the 3′-flanking region. (For the 5′-flanking region and *Est-6*, see Balakirev and Ayala, 2003b.) There are 11 length polymorphisms in the whole region, but only two in ψ*Est-6*. There are 54 replacement (including one site associated with a deletion) and 45 synonymous polymorphic sites in the coding region of ψ*Est-6*. (There are 20 replacement and 34 synonymous polymorphic sites in the *Est-6* coding region; see Balakirev and Ayala, 2003b.) The ratio of replacement to synonymous polymorphic sites is 0.588 for *Est-6* but 1.2 for ψ*Est-6*. There are 17 premature stop codons (all TGA) within the coding region of ψ*Est-6*: 11 in North America and six in Europe, but none in Africa or South America. The stop codons are generated by single mutations (at positions 4134, 4454) or by a short insertion, ACATTTGAT, at 4445–4453. The *mdg-3* retrotransposon (5.2-kb) is inserted within the intron of ψ*Est-6* in a strain from California, ER-S-438S (data not shown). Game and Oakeshott (1990) found the same insertion in a strain that carried a null allele of ψ*Est-6*.

Table 1 gives estimates of nucleotide diversity for the *β-esterase* gene cluster. The π value for the full sequence is 0.0083, 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 ($\pi = 0.0060$) and *Est-6* regions ($\pi = 0.0057$), but higher in the intergenic region ($\pi = 0.0094$), ψ*Est-6* ($\pi = 0.0115$), and in the 3′-flanking region (0.0123–0.0184). Synonymous polymorphism is $\pi = 0.0160$ in *Est-6*, but 1.5 times higher, $\pi = 0.0244$, in ψ*Est-6*. The difference is more pronounced for non-synonymous variation: 0.0024 in *Est-6* and 3.2 times higher, 0.0076, in ψ*Est-6*. These differences could indicate different degrees of selective constraint. *K* (divergence with *D. simulans*) is similar for *Est-6* (0.1474) and ψ*Est-6* (0.1384), but higher in both than in the 5′- (0.0514), intergenic (0.0707), or 3′-flanking (0.0677; Venezuela not included) regions.

Over the entire β-esterase cluster, π is highest in Africa (0.0113), lowest in South America (0.0052), but intermediate in Europe (0.0073) and North America (0.0085) (Table 1). The greatest heterogeneity of π is in the 5′-flanking region: 0.0126 in Africa, but 0.0044 in the non-African populations. The differences are significant by coalescent simulations.

The recombination rate is nearly three times higher for *Est-6* than for ψ*Est-6* (0.0214 v.s. 0.0082 per site; Table 2). We have conjectured that the different recombination rate between *Est-6* and ψ*Est-6* could be a consequence of non-functionalization of ψ*Est-6*, which may have led to deterioration of specific recognition sequences that promote recombination (Balakirev et al., 2003). However, the rate of recombination is very similar for *Est-6* and ψ*Est-6* in Africa (Table 2). This suggests that the rate of recombination may be the same in both genes, but the influence of demographic history (bottlenecks) and/or directional selection has biased the original nucleotide pattern in derived populations. Similar observations have been made in human African and non-African populations (e.g., Frisse et al., 2001).

Intragenic gene conversion occurs within *Est-6* (except in Venezuela) and ψ*Est-6*, but the number of significant events is considerably higher for ψ*Est-6* (Table 3). This intensive intragenic gene conversion within ψ*Est-6* may be explained by the invasion of retrotransposons (such as *mdg-3*, cited above) that can promote homology-dependent gene conversion upon excision (Engels, 1989; Athma and Peterson, 1991; Lowe, Mathern and Hake, 1992; Preston and Engels, 1996; Svoboda, Robson and Sved, 1996). The number of conversion events is much lower in the protein alignment (Table 3), which suggests that mostly silent sites are involved in the nucleotide alignment. The exception is the North American population (ER), which is the only one where conversion events are detected between *Est-6* gene and ψ*Est-6*, and only for the protein alignment. The low level of intergenic gene conversion may be due to the extensive nucleotide divergence between *Est-6* and ψ*Est-6* (Balakirev et al., 2003). The coding nucleotide sequences show only 64% similarity (Collet et al., 1990), which is probably too low to satisfy the homology requirements for efficient intergenic conversion. Interlocus gene conversion has been detected between *Est-5A* and *Est-5B* of *D. pseudoobscura* (82.5% similarity) (Brady and Richmond, 1992; King, 1998), which are orthologous to *D. melanogaster* *Est-6* and ψ*Est-6*, respectively. Nevertheless, taken together, the overall results demonstrate that intragenic gene conversion has played an important role in the evolution of *D. melanogaster*'s *β-esterase* gene cluster.

Table 2. Recombination estimates (ρ)

	Full sequence			*Est-6*			ψ*Est-6*		
	Per gene	Per site	ρ/θ	Per gene	Per site	ρ/θ	Per gene	Per site	ρ/θ
Zim	82.966	0.0198	1.6923	62.124	0.0368	4.3294	56.513	0.0334	2.4741
Bar	14.629	0.0035	0.4321	7.816	0.0046	0.9583	3.206	0.0019	0.1496
ER	6.814	0.0016	0.1951	10.020	0.0059	1.0536	2.605	0.0015	0.1316
Ven	1.002	0.0002	0.0444	1.002	0.0006	0.1818	0.000	0.0000	0.0000
All	34.068	0.0082	0.8367	36.072	0.0214	3.2424	13.828	0.0082	0.6667

ρ is estimate of the population recombination rate $4N_e r$ (N_e is the effective population size and r is the recombination rate/nucleotide site/generation). All sites are included in the recombination analysis.

Table 3. Gene conversion events

Populations	Within *Est-6*		Within ψ*Est-6*		Between *Est-6* and ψ*Est-6*	
	DNA	Protein	DNA	Protein	DNA	Protein
Zim	1 ($P = 0.0057$)	1 ($P = 0.0428$)	20 ($P = 0.0000$)	N.S.	N.S.	N.S.
Bar	7 ($P = 0.0000$)	N.S.	42 ($P = 0.0000$)	N.S.	N.S.	N. S.
ER	14 ($P = 0.0097$)	N.S.	85 ($P = 0.0000$)	16 ($P = 0.0056$)	N.S.	138 ($P = 0.0102$)
Ven	N.S.	N.S.	60 ($P = 0.0068$)	N.S.	N.S.	N.S.

The table gives the number and significance of fragments.

Haplotype structure

Odgers, Healy and Oakeshott (1995) and Odgers et al. (2002) described two groups of haplotypes for the 5′-flanking region of the *Est-6* gene of *D. melanogaster*. We detected two groups of haplotypes both for the *Est-6* gene (including 5′-flanking region) and ψ*Est-6* putative pseudogene in *D. melanogaster* from the North American (California) population (Balakirev and Ayala 1996, 2003b; Balakirev et al., 1999, 2003; Balakirev, Balakirev and Ayala, 2002) and also for the *Est-6* gene in the European and South American samples (Balakirev and Ayala, 2003b). The divergence between haplotypes is greater for ψ*Est-6* than for *Est-6* (Table 1). Figure 2(A) shows, for the full-length sequence, that the division between haplotypes does not fully correspond to the *Est-6* allozyme variation. The top cluster of 42 very similar sequences includes only S alleles (except for Bar F-77F, ERF-611F, and Zim F-H31), but the other 36, rather heterogeneous lines include F as well as S alleles. The correspondence is somewhat better for *Est-6* (Figure 2(B)) than for ψ*Est-6* (Figure 2(C)). There is not much geographic structure, either: the haplotypes are interspersed irrespective to geographic origin.

In the South American population, only the S *Est-6* allozyme occurs, but there are two significantly divergent sequence types. The average number of nucleotide differences (k) between the two haplotypes is 11.286 (*Est-6* gene) and 28.643 (ψ*Est-6*). This is comparable with the differences between the *Est-6* ($k = 11.809$) and ψ*Est-6* ($k = 20.534$) haplotypes in California (Balakirev, Balakirev and Ayala, 2002; Balakirev et al., 2003). The permutation test (Hudson, Boos and Kaplan, 1992; Hudson, Slatkin and Maddison, 1992) is highly significant for the Venezuelan haplotypes: $K^*_{st} = 0.5867$, $P = 0.0000$, for *Est-6* and $K^*_{st} = 0.5921$, $P = 0.0000$, for ψ*Est-6*. This observation confirms our suggestion (Balakirev, Balakirev and Ayala, 2002; Balakirev et al., 2003) that selection is more intense for the 5′-flanking region than for the *Est-6* coding region and that the latitudinal F/S frequency clines that occur in *Est-6* (e.g., Oakeshott et al., 1993) are due to the interaction of selective processes in the promoter and coding regions, rather than to the coding region alone.

The estimates of population differentiation (F_{st}; Hudson, Boos and Kaplan, 1992) based on the whole length of the β-esterase cluster are fairly similar between the pairs Zim–Bar ($F_{st} = 0.1037$),

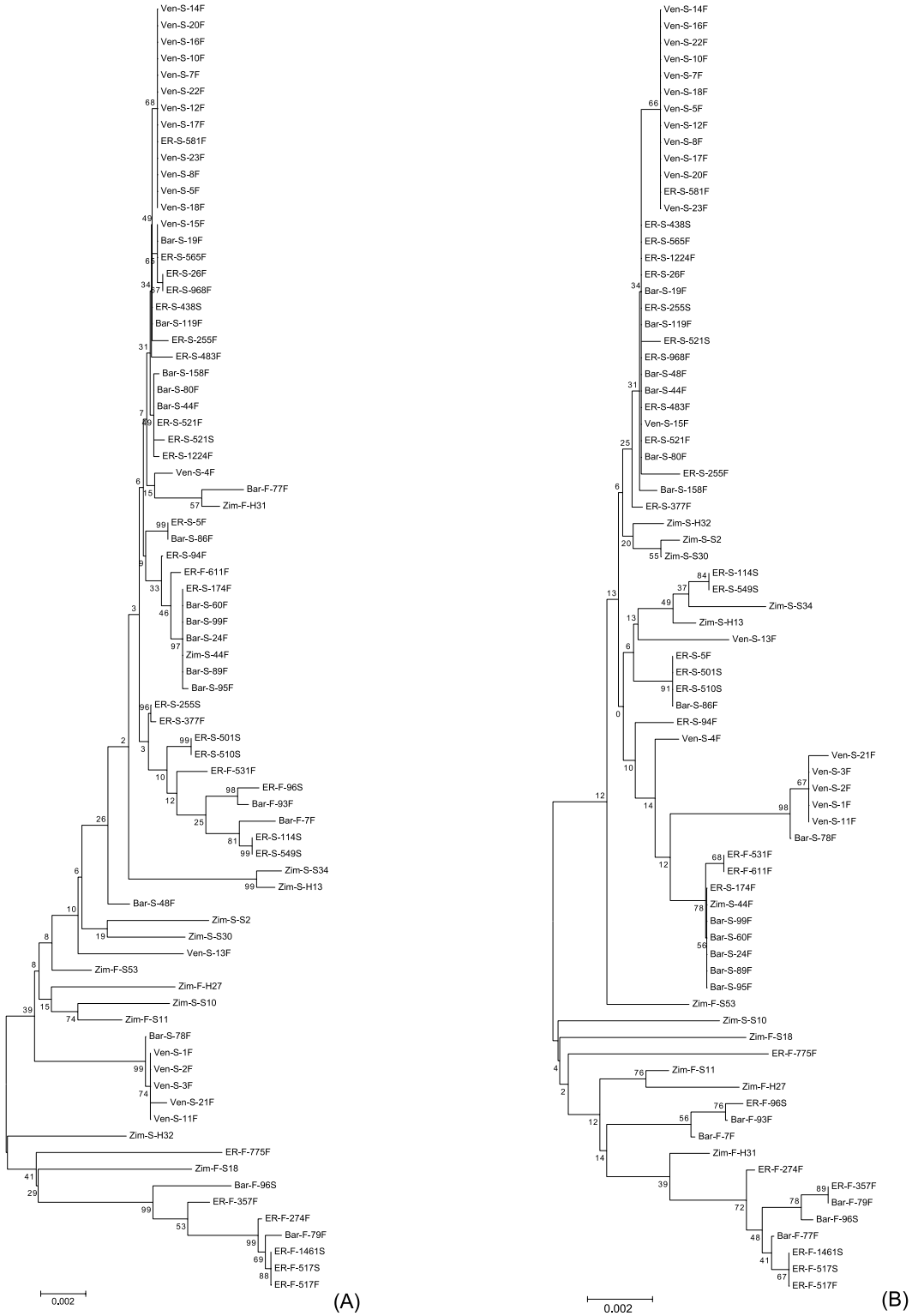

Figure 2. Neighbor-joining trees of the *β-esterase* gene cluster haplotypes of *D. melanogaster* based on the Kimura 2-parameter distances. The trees are based on the complete sequence (A), *Est-6* (B) or ψ*Est-6* (C).

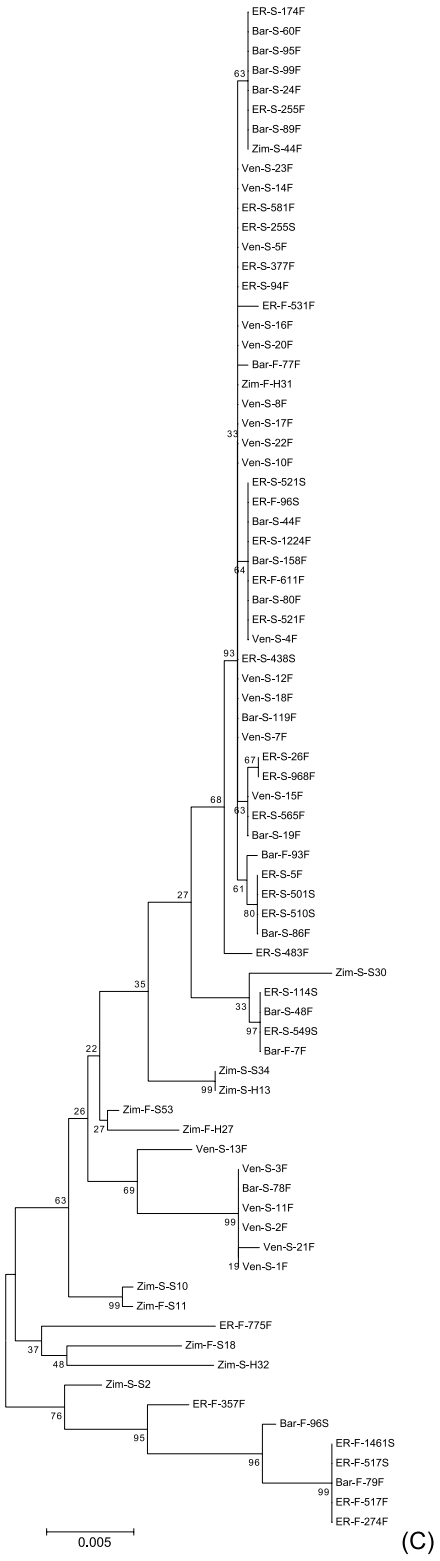

Figure 2. (*continued*)

Zim–ER ($F_{st} = 0.0758$), Bar–Ven ($F_{st} = 0.0702$), and ER–Ven ($F_{st} = 0.0903$) (for locality abbreviations see Figure 1). Maximum and minimum F_{st} values occur, respectively, for the pairs Zim–Ven ($F_{st} = 0.1504$) and Bar–ER ($F_{st} = -0.0143$). The permutation method of (Hudson, Slatkin and Maddison, 1992) detects significant differentiation between Africa and any other population, a result consistent with similar geographic differentiation for other genes (Begun and Aquadro, 1993; Andolfatto, 2001; Aquadro, Bauer and Reed, 2001). The differences between the Europe and the North or South America populations are not significant ($P > 0.05$).

Sliding window analysis

Figure 3 shows the distribution of polymorphism along the whole region. There is a strong peak in the promoter region around the *Rsa*I+/*Rsa*I− polymorphic site. There is also a distinct peak around the *Est-6* F/S site, which may reflect balancing selection (Strobeck, 1983; Hudson and Kaplan, 1988; Balakirev et al., 1999, 2003; Balakirev, Balakirev and Ayala, 2002). There are pronounced peaks along the ψ*Est-6* sequence (with intervals 200–300 bp), which are not centered around the replacement polymorphisms (Figure 1). The tests of Goss and Lewontin (1996) and McDonald (1996,

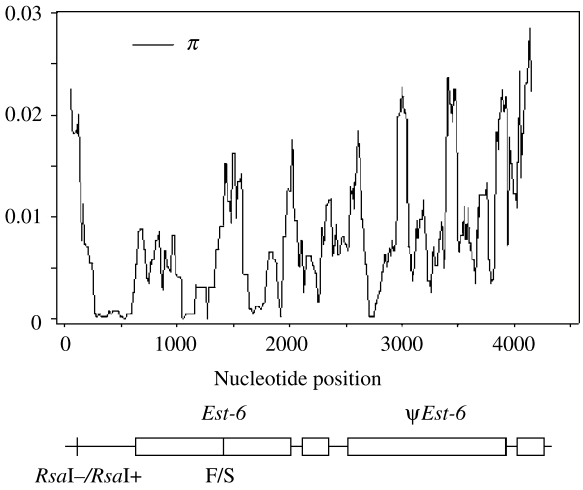

Figure 3. Sliding window plot of nucleotide diversity (π) in the *β-esterase* gene cluster of *D. melanogaster*. A schematic representation of the cluster is displayed at bottom. Exons are indicated by boxes; the intron and the 5′- and 3′-flanking regions are shown by thin lines. Window sizes are 100 nucleotides with 10-nucleotide increments. The location of the *Rsa*I and allozyme polymorphism are marked by the two vertical lines.

1998) do not reveal for ψ*Est-6* any significant heterogeneity in the distribution of polymorphic sites relative to fixed interspecific differences for the African, European, and North American populations, but the tests are significant in South America: maximal $G \geq 18.367$ ($P = 0.048$); $\mathrm{Var} \geq 0.00094$ ($P = 0.012$); modified $\mathrm{Var} \geq 0.00207$ ($P = 0.021$); average $G \geq 6.690$ ($P = 0.019$).

There are valley regions in *Est-6* located around positions 350, 1200, and 1800 (Figure 3). The first valley region includes nearly 400 bp upstream of the *Est-6* coding region. Karotam, Delves and Oakeshott (1993), Karotam, Boyce and Oakeshott (1995) and Odgers, Healy and Oakeshott (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 (Ludwig, Tamarina and Richmond, 1993) that are essential for *Est-6* expression. Another valley region (1100–1300 bp) includes the amino acid residues Arg-159, Asp-181, and Ser-209 positions 1094–1096, 1160–1162, and 1244–1246 in our coordinates), which are highly conserved in different esterases and are likely to be important for esterase enzymatic function (Myers, Richmond and Oakeshott, 1988). A third valley region encompasses the potential N-linked glycosylation site (1877–1879 in our coordinates). The low level of polymorphism around functionally important sites of *Est-6* suggests that the observed valley regions reflect functional constraint.

Figure 4 shows a sliding-window plot of the distribution of nucleotide polymorphism (π) in *D. melanogaster* and interspecific divergence (K) with respect to *D. simulans*. K shows conspicuous valleys in the middle of the 5′-flanking region; the beginning of the *Est-6* gene; the intergenic region and beginning of ψ*Est-6*; and towards the 3′ end of ψ*Est-6*. These valleys of interspecific divergence may be due to either functional constraint and/or gene conversion events. Healy et al. (1996) have shown that 3′ segments within the ψ*Est-6* transcription unit contain elements that modulate the expression of *Est-6*. This observation obviously implies some regulatory function for ψ*Est-6*. Moreover, Brady and Richmond (1992) detected some sequence similarity in the 3′-flanking region between ψ*Est-6* (*Est-P* in Brady and Richmond, 1992) of *D. melanogaster* and *Est-5A* of *D. pseudoobscura*. Also, a 390-bp block within the 609-bp of 3′ flanking sequence of ψ*Est-6* retains similarity with sequences 3′ of *Est-5A* in *D. pseudoobscura* (Collet et al., 1990; Brady and Richmond, 1992). In particular, a segment of 110 bp within this region shows 76% sequence similarity between the two species, which contrasts sharply with 20% or lower similarity in the 5′ regions of *Est-6* (*Est-5B* in *D. pseudoobscura*) or ψ*Est-6* (*Est-5A* in *D. pseudoobscura*) (Brady and Richmond, 1992).

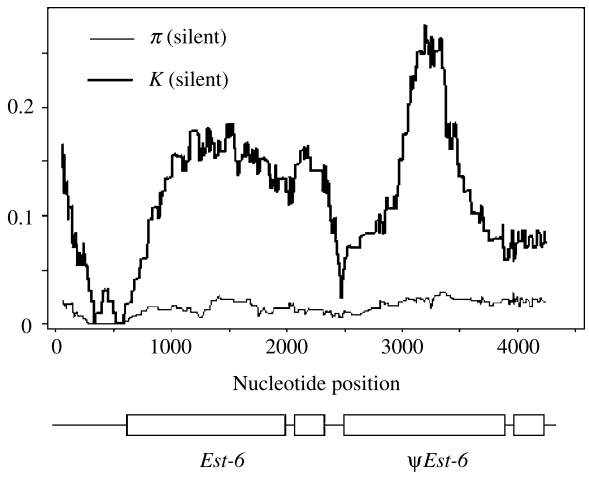

Figure 4. Sliding window plots of nucleotide diversity (π, thin line) and interspecific divergence (K, thick line), along the *β-esterase* gene cluster of *D. melanogaster*. K is the average number of nucleotide substitutions per site between *D. melanogaster* and *D. simulans*. Window sizes are 100 nucleotides with one-nucleotide increments. Notice different scale from Figure 3.

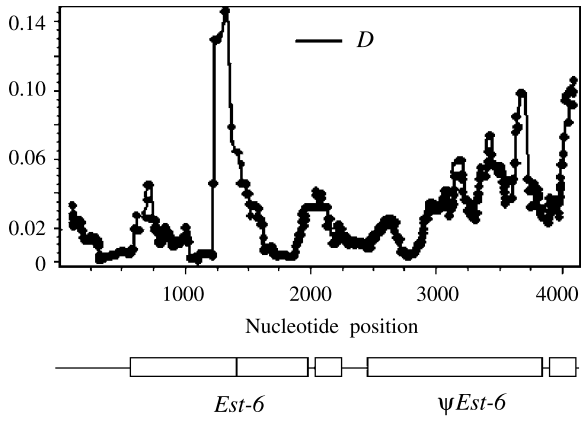

Figure 5. Sliding window plot of linkage disequilibrium (measured by D) along the *β-esterase* gene cluster of *D. melanogaster*. The schematic representation of the cluster at the bottom marks the site of the S/F allozyme polymorphism with a vertical line. Window size of 200 nucleotides with one-nucleotide increments.

Table 4. Correlation between linkage disequilibrium and physical distance between the *Est-6* polymorphic sites

Populations	*Est-6* LD_{Fisher}	*Est-6* Pearson's correlation coefficient for		ψ*Est-6* LD_{Fisher}	ψ*Est-6* Pearson's correlation coefficient for	
		r^2 (P)	D' (P)		r^2 (P)	D' (P)
Zim	0.0199	−0.1510 (0.0000)	−0.0154 (0.4130)	0.0793	−0.3029 (0.0000)	−0.0375 (0.1260)
Bar	0.3333	−0.3012 (0.0000)	−0.2942 (0.0000)	0.3425	−0.2734 (0.0000)	−0.1121 (0.0060)
ER	0.2703	−0.1739 (0.0020)	−0.2298 (0.0010)	0.5596	−0.1093 (0.0070)	−0.0634 (0.0410)
Ven	0.3474	+0.0719 (0.8030)	−0.0031 (0.4490)	0.5775	−0.1180 (0.0710)	n.a.
All	0.1858	−0.1468 (0.0003)	−0.1558 (0.0003)	0.3669	−0.1794 (0.0000)	−0.0723 (0.0070)

LD_{Fisher} is the proportion of significant linkage disequilibrium revealed by Fisher's exact test using all polymorphic sites. All sites types are included in the analysis.

Table 5. Kelly's (1977) test of neutrality for the β-esterase gene cluster

	5′-flanking region (0.6 kb)			*Est-6* region (1.6 kb)			ψ*Est-6* region (1.6 kb)			Entire region (4.1 kb)		
	Z_{nS}	0.05	0.01	Z_{nS}	0.05	0.01	Z_{nS}	0.05	0.01	Z_{nS}	0.05	0.01
All populations	0.106	0.050	N.S.	0.079	0.030	0.035	0.158	0.005	0.010	0.092	0.010	0.010
Zim	0.138	N.S.	N.S.	0.115	N.S.	N.S.	0.176	N.S.	N.S.	0.119	N.S.	N.S.
Bar	0.540	0.005	0.030	0.285	0.010	0.015	0.384	0.005	0.005	0.276	0.005	0.005
ER	0.458	0.005	0.020	0.204	0.010	0.015	0.422	0.005	0.005	0.286	0.005	0.005
Ven	0.453	0.010	0.025	0.386	0.010	0.010	0.564	0.000	0.005	0.482	0.005	0.005

The table presents Kelly's (1977) Z_{nS} values and the recombination rates with which the test is significant at the 5% (column 0.05) or 1% (column 0.01) level of significance. N.S., not significant, even with the laboratory-estimated recombination rate (0.0664) based on the physical and genetic maps of *D. melanogaster* (Comeron, Kreitman and Aguadé, 1999; Balakirev, Balakirev and Ayala, 2002).

Linkage disequilibrium

Figure 5 shows the distribution of D. There is a noticeable peak of high D values around the F/S site of *Est-6* gene and several peaks along ψ*Est-6*. For the whole region, out of 8646 pairwise comparisons, 4092 (47.33%) are significant. With the Bonferroni correction, only 15.07% significant associations. For the 5′-flanking region there are 32.05% (25 out of 78 pairwise comparisons) significant associations (23.08% with the Bonferroni correction). For the *Est-6* gene, 41.48% (219 out of 528) pairwise comparisons are significant (23.11% with the correction). For ψ*Est-6*, 62.65% pairwise comparisons (1833 out of 2926) show statistically significant linkage disequilibrium (32.43% with the correction). There are 19.58% (1.17%) significant associations between the 5′-flanking region and *Est-6* gene and 23.68% (0.80%) between the 5′-flanking region and ψ*Est-6*. Between *Est-6* and ψ*Est-6*, 46.44% (1180 out of 2541) tests are significant (5.86% with the correction). The significant interlocus linkage disequilibria are due to six divergent haplotypes F-517S, F-517F, F-1461S, F-274F, F-357F, and F-775F, which have unique polymorphisms both in *Est-6* and ψ*Est-6*. The intergenic linkage disequilibrium may be due to epistatic selection reflecting possible functional interactions between *Est-6* gene and ψ*Est-6*. Linkage disequilibrium is noticeably low in the African population (only 7.35% significant associations *versus* 41.68%, 62.69%, and 100% in Europe, North, and South America, respectively.

We have analyzed the relation between linkage disequilibrium and nucleotide distance by the method of McVean, Awadalla and Fearnhead, (2002; significance of D' assessed by 10,000 permutations). Linkage disequilibrium significantly decreases with increasing distance (D' in Table 4) in Spain or California, but not in Africa or Venezuela. The strong haplotype structure and the associated high level of linkage disequilibrium in South American population suggest that this population originated from a recent admixture of genetically differentiated populations.

Tests of neutrality

Kelly's Z_{nS} (1997) test detects significant deviations from neutrality for each gene region separately and for the entire segment (4.1 kb), for all populations combined, as well as for each geographic population,

except Africa (Table 5). Wall's (1999) B and Q tests yield similar results (data not shown). For ψ*Est-6* the tests are significant with lower level of recombination than for the *Est-6* gene. For instance, the Z_{nS} statistic for all populations combined is significant ($P = 0.01$) for ψ*Est-6* with recombination $C = 0.010$, but only with $C = 0.035$ for *Est-6* (Table 5, top row). The significance of the tests in the promoter and *Est-6* regions may reflect both the action of selection and the demographic history of *D. melanogaster*, which spread from Africa to other continents. The significance of the tests for ψ*Est-6* may reflect its observed low rate of recombination, as well as the demographic history of *D. melanogaster*.

Discussion

Pseudogenes often arise by tandem duplication of genes (Jacq, Miller and Brownlee, 1977; Proudfoot, 1980; Little, 1982; Wilde, 1986). 'Processed' pseudogenes lack introns, and presumably arise by reverse transcription of processed mRNA, followed by integration into the genome (Nishioka, Leder and Leder, 1980; Vanin et al., 1980; Weiner, Deininger and Efstratiadis, 1986). If the presence of one copy of the gene suffices for the needs of the organism, pseudogene mutations (disabling or not) may not be subject to purifying selection and will all have equal probability of becoming fixed in the population (Kimura, 1980; Li, Gojobori and Nei, 1981; Graur and Li, 2000). Therefore, pseudogenes are expected to degenerate owing to the accumulation of deleterious mutations and become 'junk' or otherwise non-functional DNA.

Pseudogenes are rare in *Drosophila* (Powell, 1997; Harrison et al., 2003) relative to some other animals, especially vertebrates (Mighell et al., 2000). DNA sequence evolution in many of the pseudogenes found in *Drosophila* manifests functional constraints, reflected in lower than expected (if the pseudogenes were not subject to selection) intraspecific variability and interspecific divergence; significant heterogeneity of nucleotide variability and divergence along the sequence; higher rate of substitution at synonymous than non-synonymous nucleotide positions; conservation of important functional regions; transcriptional activity; and codon bias (Jeffs and Ashburner, 1991; Jeffs, Holmes and Ashburner, 1994: Sullivan et al., 1994; Balakirev and Ayala, 1996; Pritchard and Schaeffer, 1997; Ramos-Onsins and Aguadé, 1998; Balakirev et al., 2003).

We have analyzed nucleotide polymorphism in the β-*esterase* gene cluster in 78 strains of *D. melanogaster* from four different populations in four continents seeking to ascertain whether ψ*Est-6* is functional rather than a pseudogene. Some features of ψ*Est-6* indicate that it could be a pseudogene: 17 premature stop codons among 78 sequences are hardly compatible with the functionality of the encoded protein; the number of amino acid replacements is 3 times higher in ψ*Est-6* than in *Est-6*, and some of them are drastic; total nucleotide polymorphism is 2.1 times higher in ψ*Est-6*; moreover, structural entropy analysis reveals significantly lower structural regularity and higher structural divergence for ψ*Est-6*, in accordance with the expectations if it is a pseudogene or non-functional gene (Balakirev et al., 2003).

However, the patterns of nucleotide variation, recombination, and linkage disequilibrium indicate that the evolution of ψ*Est-6* is subject to selective constraints. The Kelly and Wall's tests of neutrality with recombination are significant for ψ*Est-6* (as well as *Est-6*) in the three non-African populations, but not in Africa (Wall's data not shown). In the non-African populations, the recombination rate is at least 2.6 times lower in ψ*Est-6* than in *Est-6* gene, while it is about the same in the African population. This difference might be a consequence of demographic history and natural selection associated with *D. melanogaster*'s colonization of the other continents from its endemic origins in tropical Africa. There is also strong linkage disequilibrium between *Est-6* and ψ*Est-6*. This observation is consistent with the hypothesis that intergenic epistatic selection may play an important role in the evolution of the β-*esterase* gene cluster preserving the ψ*Est-6* putative pseudogene from the degenerative destruction, and reflecting possible functional interaction between *Est-6* and ψ*Est-6* (e.g., regulatory interaction, Healy et al., 1996). We have detected gene conversion both within and (to a much lesser extent) between *Est-6* gene and ψ*Est-6*. Intensive intragenic gene conversion within ψ*Est-6* can be explained by the invasion of retrotransposons (like *mdg-3*, see results) that can promote a form of homology-dependent gene conversion upon excision. Moreover, the haplotype structure of ψ*Est-6* is dimorphic, but so that the pattern of divergence among sequences of ψ*Est-6* does not correspond with the *Est-6* allozyme variation.

We noted earlier that ψ*Est-6* can be expressed (Collet et al., 1990) and some alleles of ψ*Est-6* produce a catalytically active esterase (Dumancic et al., 1997). This has been detected in late larvae and adults

of *both* sexes, whereas the functional *Est-6* genes transcripts are found in all life stages but predominantly in adult males, consistent with the significant role of EST-6 in male mating (Richmond et al., 1980; Gromko, Gilbert and Richmond, 1984). Thus, it seems unlikely that ψ*Est-6* would play a functional role altogether similar to that of *Est-6*.

In conclusion, the evidence is inconsistent with the hypothesis that ψ*Est-6* is a neutrally-evolving, non-functional pseudogene, but it is also inconsistent with the hypothesis that ψ*Est-6* functions as a duplicate of *Est-6*, or even that it is a fully functional gene. A sharp division between genes and pseudogenes may not be appropriate, in this and other instances. A pseudogene may lose some specific function(s) but retain other(s), or acquire new one(s). There are many examples of functional or 'active' pseudogenes. The term 'potogene' may be appropriate for ψ*Est-6*, following Brosius and Gould (1992), who have pointed out that the products of a gene duplication, including those that become pseudogenes, may eventually acquire distinctive functions, and thus might be called *poto*genes to call attention to their *pot*entiality for becoming new genes or acquire new functions.

The data on ψ*Est-6* are in accordance with a general picture observed for many other pseudogenes, including almost all known pseudogenes in *Drosophila* (Balakirev and Ayala, 2003a; Balakirev et al., 2003). This picture is contrary to the traditional view that defines pseudogenes as sequences of genomic DNA that are non-functional ('junk' DNA) and consequently free of selection. Our results help to understand why eukaryote genomes contain many pseudogenes that appear to have avoided full degeneration. Pseudogenes are, rather, an important part of the genome representing a repertoire of sequences often involved in the function of their parental sequences, jointly representing indivisible functionally interacting entities ('intergenic complexes' or 'intergenes') in which each single part cannot successfully accomplish the final functional role. Examples of such intergenes, in addition to *Est-6*-ψ*Est-6*, are *Ste*–[*Su*(*Ste*)] (Hardy, Jokuyasu and Lindsley, 1981; Livak, 1984, 1990) in *D. melanogaster*; *nNOS–pseudo-NOS* in the mollusk *Lymnaea stagnalis* (Korneev, Park and O'Shea, 1999); *Makorin1* gene–*Makorin1-p1* pseudogene in mice (Hirotsune et al., 2003); *cytokeratin 17* gene–*cytokeratin 17* pseudogene (Troyanovsky and Leube, 1994) in humans; immune response and antigenic coding sequences in diverse organisms. We anticipate that the number of intergenes will significantly increase in the near future with the progress of functional genomics.

Acknowledgement

This work is supported by NIH grant GM42397 to F.J. Ayala.

References

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. Bauer & F.A. Reed, 2001. Genome-wide variation in the human and fruitfly: a comparison. Curr. Opinion Genet. Dev. 11: 627–634.

Athma, P. & T. Peterson, 1991. *Ac* induces homologous recombination at the maize *P* locus. Genetics 128: 163–173.

Ayala, F.J., E.S. Balakirev & A.G. Sáez, 2002. Genetic polymorphism at two linked loci, *Sod* and *Est-6*, in *Drosophila melanogaster*. Gene 300: 19–29.

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

Balakirev, E.S. & F.J. Ayala, 2003a. Pseudogenes: are they 'Junk' or functional DNA? Ann. Rev. Genet. 37: 123–151.

Balakirev, E.S. & F.J. Ayala, 2003b. Nucleotide variation of the *Est-6* gene region in natural populations of *Drosophila melanogaster*. Genetics (submitted).

Balakirev, E.S., E.I. Balakirev & 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., E.I. Balakirev, F. Rodriguez-Trelles & F.J. Ayala, 1999. Molecular evolution of two linked genes, *Est-6* and *Sod*, in *Drosophila melanogaster*. Genetics 153: 1357–1369.

Balakirev, E.S., V.R. Chechetkin, V.V. Lobzin & F.J. Ayala, 2003. DNA polymorphism in the *β-esterase* gene cluster of *Drosophila melanogaster*. Genetics 164: 533–544.

Begun, D.J. & C.F. Aquadro, 1993. African and North American populations of *Drosophila melanogaster* are very different at the DNA level. Nature 365: 548–550.

Brady, J.P. & R.C. Richmond, 1992. An evolutionary model for the duplication and divergence of esterase genes in Drosophila. J. Mol. Evol. 34: 506–521.

Brady, J.P., R.C. Richmond & 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.

Brosius, J. & S.J. Gould, 1992. On 'genomenclature': a comprehensive (and respectful) taxonomy for pseudogenes and other 'junk DNA'. Proc. Natl. Acad. Sci. USA 89: 10706–10710.

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

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

Dumancic, M.M., J.G. Oakeshott, R.J. Russell & 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 & G. Whitington, 1990. Molecular isolation and preliminary characterization 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.T. Starmer & R.J. MacInture. Plenum Press, New York.

Engels, W.R., 1989. P elements in *Drosophila melanogaster*, pp. 437–484 in Mobile DNA, edited by D. Berg & M. Howe. American Society for Microbiology, Washington, D.C.

Filatov, D.A. & D. Charlesworth, 1999. DNA polymorphism, haplotype structure and balancing selection in the Leavenworthia PgiC locus. Genetics 153: 1423–1434.

Frisse, L., R.R. Hudson, A. Bartoszewicz, J.D. Wall, J. Donfack & A. Di Rienzo, 2001. Gene conversion and different population histories may explain the contrast between polymorphism and linkage disequilibrium levels. Am. J. Hum. Genet. 69: 831–843.

Game, A.Y. & J.G. Oakeshott, 1990. Associations between restriction site polymorphism and enzyme activity variation for esterase 6 in *Drosophila melanogaster*. Genetics 126: 1021–1031.

Goss, P.J.E. & R.C. Lewontin, 1996. Detecting heterogeneity of substitution along DNA and protein sequences. Genetics 143: 589–602.

Graur, D. & W.-H. Li, 2000. Fundamentals of Molecular Evolution. Sinauer, Sunderland, MA, 2nd edn.

Graur, D., Y. Shuali & W.-H. Li, 1989. Deletions in processed pseudogenes accumulate faster in rodents than in humans. J. Mol. Evol. 28: 279–285.

Gromko, M.H., D.F. Gilbert & 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.

Hardy, R.W., K.T. Tokuyasu & D.L. Lindsley, 1981. Analysis of spermatogenesis in *Drosophila melanogaster* bearing deletions for *Y* chromosome fertility genes. Chromosoma 83: 593–617.

Harrison, P.M. & M. Gerstein, 2002. Studying genomes through the aeons: protein families, pseudogenes and proteome evolution. J. Mol. Biol. 318: 1155–1174.

Harrison, P.M., D. Milburn, Z. Zhang, P. Bertone & M. Gerstein, 2003. Identification of pseudogenes in the *Drosophila melanogaster* genome. Nucl. Acids Res. 31: 1033–1037.

Hartl, D.L. & A.G. Clark, 1989. Principles of Population Genetics. Sinauer, Sunderland, MA, 2nd edn.

Healy, M.J., M.M. Dumancic & J.G. Oakeshott, 1991. Biochemical and physiological studies of soluble esterases from *Drosophila melanogaster*. Biochem. Genet. 29: 365–388.

Healy, M.J., M.M. Dumancic, A. Cao & 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.

Hirotsune, S., N. Yoshida, A. Chen, L. Garrett, F. Sugiyama, S. Takahashi, K. Yagami, A. Wynshaw-Boris & A. Yoshiki, 2003. An expressed pseudogene regulates the messenger-RNA stability of its homologous coding gene. Nature 423: 91–96.

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

Hudson, R.R., D. Boos & N.L. Kaplan, 1992. A statistical test for detecting geographic subdivision. Mol. Biol. Evol. 9: 138–151.

Hudson, R.R., M. Slatkin & W.P. Maddison, 1992. Estimation of levels of gene flow from DNA sequence data. Genetics 132: 583–589.

Jacq, C., J.R. Miller & G.G. Brownlee, 1977. A pseudogene structure in 5S DNA of *Xenopus laevis*. Cell 12: 109–120.

Jeffs, P.S. & M. Ashburner, 1991. Processed pseudogenes in Drosophila. Proc. R. Soc. Lond. B 244: 151–159.

Jeffs, P.S., E.C. Holmes & M. Ashburner, 1994. The molecular evolution of the alcohol dehydrogenase and alcohol dehydrogenase-related genes in the *Drosophila melanogaster* species subgroup. Mol. Biol. Evol. 11: 287–304.

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

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

Kelly, J.K., 1997. A test of neutrality based on interlocus associations. Genetics 146: 1197–1206.

Kimura, M., 1980. A simple method for estimating evolutionary rate of base substitutions through comparative studies of nucleotide sequences. J. Mol. Evol. 16: 111–120.

King, L.M., 1998. The role of gene conversion in determining sequence variation and divergence in the *Est-5* gene family in *Drosophila pseudoobscura*. Genetics 148: 305–315.

Korneev, S.A., J.-H. Park & M. O'Shea, 1999. Neuronal expression of neural nitric oxide synthase (nNOS) protein is suppressed by an antisense RNA transcribed from an NOS pseudogene. J. Neurosci. 19: 7711–7720.

Li, W.-H., T. Gojobori & M. Nei, 1981. Pseudogenes as a paradigm of neutral evolution. Nature 292: 237–239.

Little, P.F., 1982. Globin pseudogenes. Cell 28: 683–684.

Livak, K.J., 1984. Organization and mapping of a sequence on the *Drosophila melanogaster X* and *Y* chromosomes that is transcribed during spermatogenesis. Genetics 107: 611–634.

Livak, K.J., 1990. Detailed structure of the *Drosophila melanogaster Stellate* genes and their transcripts. Genetics 124: 303–316.

Lowe, B., J. Mathern & S. Hake, 1992. Active *Mutator* elements suppress the knotted phenotype and increase recombination at the *Kn1-0* tandem duplication. Genetics 132: 813–822.

Ludwig, M.Z., N.A. Tamarina & 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 & P. Fearnhead, 2002. A coalescent-based method for detecting and estimating recombination from gene sequences. Genetics 160: 1231–1241.

Mighell, A.J., N.R. Smith, P.A. Robinson & A.F. Markham, 2000. Vertebrate pseudogenes. FEBS Letters 468: 109–114.

Moriyama, E.N. & 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 & M. Veuille, 2003. Evidence of a high rate of selective sweeps in African *Drosophila melanogaster*. Genetics 163: 599–609.

Myers, M., R.C. Richmond & J.G. Oakeshott, 1988. On the origins of esterases. Mol. Biol. Evol. 5: 113–119.

Nishioka, Y., A. Leder & P. Leder, 1980. Unusual α-globin-like gene that has cleanly lost both globin intervening sequences. Proc. Natl. Acad. Sci. USA 77: 2806–2809.

Oakeshott, J.G., C. Collet, R. Phillis, K.M. Nielsen, R.J. Russell, G.K. Chambers, V. Ross & R.C. Richmond, 1987. Molecular cloning and characterization of esterase 6, a serine hydrolase from Drosophila. Proc. Natl. Acad. Sci. USA 84: 3359–3363.

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

Oakeshott, J.G., T.M. Boyce, R.J. Russell & 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 & 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 & 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.

Powell, J.R., 1997. Progress and Prospects in Evolutionary Biology. The Drosophila Model. Oxford University Press, Oxford and New York.

Preston, C.R. & W.R. Engels, 1996. *P*-element-induced male recombination and gene conversion in Drosophila. Genetics 144: 1611–1622.

Pritchard, J.K. & S.W. Schaeffer, 1997. Polymorphism and divergence at a *Drosophila* pseudogene locus. Genetics 147: 199–208.

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

Proudfoot, N., 1980. Pseudogenes. Nature 286: 840–841.

Ramos-Onsins, S. & M. Aguadé, 1998. Molecular evolution of the *Cecropin* multigene family in Drosophila: functional genes *vs*. pseudogenes. Genetics 150: 157–171.

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

Richmond, R.C., K.M. Nielsen, J.P. Brady & 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 & R.J. MacIntyre. Plenum Press, New York.

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

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. Distributed by the author. Department of Mathematics, Washington University in St. Louis.

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

Sullivan, D.T., W.T. Starmer, S.W. Curtiss, M. Menotti-Raymond & J. Yum, 1994. Unusual molecular evolution of an *Adh* pseudogene in Drosophila. Mol. Biol. Evol. 11: 443–458.

Svoboda, Y., M. Robson & J.A. Sved, 1996. *P*-element-induced male recombination can be produced in *Drosophila melanogaster* by combining end-deficient elements *in trans*. Genetics 139: 1601–1610.

Troyanovsky, S.M. & R.E. Leube, 1994. Activation of the silent human cytokeratin 17 pseudogene-promoter region by cryptic enhancer elements of the cytokeratin 17 gene. Eur. J. Biochem. 223: 61–69.

Vanin, E.F., G.I. Goldberg, P.W. Tucker & O. Smithies, 1980. A mouse α-globin-related pseudogene lacking intervening sequences. Nature 268: 222–226.

Wall, J.D., 1999. Recombination and the power of statistical tests of neutrality. Genet. Res. 74: 65–79.

Weiner, A.M., P.L. Deininger & A. Efstratiadis, 1986. Nonviral retroposons: gene, pseudogenes, and transposable elements generated by the reverse flow of genetic information. Ann. Rev. Biochem. 55: 631–661.

Wilde, C.D., 1986. Pseudogenes. CRC Crit. Rev. Biochem. 19: 323–352.

Yenikolopov, G.N., O.A. Malevantschuk, N.I. Peunova, P.V. Sergeev & G.P. Georgiev, 1989. *Est* locus of *Drosophila virilis* contains two related genes. Dokl. Acad. Nauk SSSR 306: 1247–1249 (in Russian).