Title
Molecular population genetics of the β-esterase gene cluster of Drosophila melanogaster

Permalink
https://escholarship.org/uc/item/95g059dv

Journal
Journal of Genetics, 82(3)

ISSN
0022-1333

Authors
Balakirev, ES
Ayala, FJ

Publication Date
2003

DOI
10.1007/BF02715813

Copyright Information

Peer reviewed

Molecular population genetics of the *b-esterase* gene cluster of *Drosophila melanogaster*

EVGENIY S. BALAKIREV[1,2] and FRANCISCO J. AYALA[1]*

[1]*Department of Ecology and Evolutionary Biology, 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*

Abstract

We have investigated nucleotide polymorphism at the *b–esterase* gene cluster including the *Est-6* gene and *yEst-6* putative pseudogene in four samples of *Drosophila melanogaster* derived from natural populations of southern Africa (Zimbabwe), Europe (Spain), North America (USA: California), and South America (Venezuela). A complex haplotype structure is revealed in both *Est-6* and *yEst-6*. Total nucleotide diversity is twice in *yEst-6* as in *Est-6*; diversity is higher in the African sample than in the non-African ones. Strong linkage disequilibrium occurs within the *b–esterase* gene cluster in non-African samples, but not in the African one. Intragenic gene conversion events are detected within *Est-6* and, to a much greater extent, within *yEst-6*; intergenic gene conversion events are rare. Tests of neutrality with recombination are significant for the *b–esterase* gene cluster in the non-African samples but not significant in the African one. We suggest that the demographic history (bottleneck and admixture of genetically differentiated populations) is the major factor shaping the pattern of nucleotide polymorphism in the *b–esterase* gene cluster. However there are some 'footprints' of directional and balancing selection shaping specific distribution of nucleotide polymorphism within the cluster. Intergenic epistatic selection between *Est-6* and *yEst-6* may play an important role in the evolution of the *b–esterase* gene cluster preserving the putative pseudogene from degenerative destruction and reflecting possible functional interaction between the functional gene and the putative pseudogene. *Est-6* and *yEst-6* may represent an indivisible intergenic complex ('intergene') in which each single component (*Est-6* or *yEst-6*) cannot separately carry out the full functional role.

Introduction

The *b-esterase* gene cluster (Korochkin *et al.* 1987, 1990) in *Drosophila melanogaster* is on the left arm of chromosome 3 at bands 68F7-69A1 in the cytogenetic map (but see Procunier *et al.* 1991). The cluster comprises two tandemly duplicated genes, *Est-6* and *Est-P*, with the same 5′ to 3′ orientation, separated by only 193 bp (Collet *et al.* 1990). The coding region of the two genes is 1686 and 1691 bp long, respectively, and consists of two exons

(1387 bp and 248 bp) and a small (51 bp in *Est-6* and 56 bp in *Est-P*) intron (Oakeshott *et al.* 1987; Collet *et al.* 1990). The two genes show 64% and 60% similarity in DNA and protein sequence, respectively (Collet *et al.* 1990). The *Est-6* gene is well characterized (reviewed by Richmond *et al.* 1990; Oakeshott *et al.* 1993, 1995). The gene encodes the major *b*-carboxylesterase (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 subsequent behaviour and mating proclivity (Gromko *et al.* 1984). The function of *Est-P* is not known. Collet *et al.* (1990) first described *Est-P* and detected its transcripts in *D. melanogaster* late larvae and

*For correspondence. E-mail: fjayala@uci.edu.

adults of both sexes, whereas the *Est-6* transcripts are found mainly in adult males; accordingly they proposed distinct physiological functions for the products of the two genes. Balakirev and Ayala (1996) and Balakirev *et al.* (2003) found premature stop codons within the *Est-P* coding region and some other indications suggesting that *Est-P* might in fact be a pseudogene, which was named *ψEst-6*. Dumancic *et al.* (1997) showed that some alleles of *Est-P* produce a catalytically active esterase, corresponding to the previously identified EST-7 isozyme (Healy *et al.* 1991) and renamed the gene, correspondingly, *Est-7*. In other *Drosophila* species the *β-esterase* gene cluster also includes two (or three, only in *D. pseudoobscura*) closely linked genes that have the same direction of transcription and similar exon/intron structure (Yenikolopov *et al.* 1989; Brady *et al.* 1990; East *et al.* 1990; for review see Oakeshott *et al.* 1993, 1995).

We have investigated nucleotide variability separately in *Est-6* and *ψEst-6*, first in a single *D. melanogaster* sample derived from a natural population of California in the US (Balakirev and Ayala 1996; Balakirev *et al.* 1999, 2002, 2003; Ayala *et al.* 2002). We also investigated the nucleotide variation of *Est-6* in three additional samples derived from natural populations of southern Africa (Zimbabwe), Europe (Spain), and South America (Venezuela) (Balakirev and Ayala 2003a). We now present the analysis of nucleotide variation in the complete *β-esterase* gene cluster (including *Est-6* and *ψEst-6*) in four natural populations of *D. melanogaster*. Totally, we have investigated 78 *D. melanogaster* lines. The full sequence length is 5394 bp and includes the 5′-flanking region, complete *Est-6*, intergenic region, complete *ψEst-6*, and 3′-flanking region.

Materials and methods

D. melanogaster strains were derived from random samples of wild flies collected in Africa (Zim: Zimbabwe), Europe (Bar: Barcelona, Spain), North America (ER: El Rio, Acampo, California, USA), and South America (Ven: 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 strain number): Ultra Slow (US), Slow (S), and Fast (F) (see figure 1). Chung-I Wu kindly provided the *D. melanogaster* strains from southern Africa (Sengwa and Harare in Zimbabwe). The strain Zim S-44F (Zimbabwe) is from Ayala's laboratory.

DNA extraction, amplification and sequencing were described earlier (Balakirev and Ayala 1996, 2003a; Balakirev *et al.* 1999, 2002, 2003). For each line, the sequences of both strands were determined, using 24 overlapping internal primers spaced, on average, 350 nucleotides apart (see GenBank accessions AF526538–AF526559; AF150809–

AF150815; AF147095–147102; AF217624–AF217645; AY247664–AY247713; AY247987–AY248036). At least two independent PCR amplifications were sequenced in both directions for each polymorphic site in all *D. melanogaster* strains to prevent possible PCR or sequencing errors.

The esterase sequences were assembled using the program SeqMan (Lasergene, DNASTAR, Inc., 1994–1997). Multiple alignment was carried out manually and by using the program CLUSTAL W (Thompson *et al.* 1994). Linkage disequilibrium between polymorphic sites was evaluated using Fisher's exact test of independence. The computer programs DnaSP, version 3.4 (Rozas and Rozas 1999), and PROSEQ, version 2.4 (Filatov and Charlesworth 1999), were used to analyse the data by the 'sliding window' method (Hudson and Kaplan 1988), and for most intraspecific analyses. Departures from neutral expectations were investigated using the tests of Kelly (1997) and Wall (1999) incorporating recombination. The permutation approach of Hudson *et al.* (1992a,b) was used to estimate the sequence differences between haplotype families, treating them as geographical populations. Simulations based on the algorithms of the coalescent process with or without recombination (Hudson 1983, 1990) were performed with the DnaSP and PROSEQ programs to estimate the probabilities of the observed values of Kelly's Z_{nS} and Wall's B and Q statistics and to estimate confidence intervals for the nucleotide diversity values. The simulations were conditional on the value of θ ($\theta = 4Nu$, where N is the effective population size and u is the per gene mutation rate; Watterson 1975). The simulations were based on 10,000 independent replicates. The program Geneconv version 1.81 (Sawyer 1999) was used to detect intragenic and intergenic conversion events. The population recombination rate was analysed by the permutation-based method of McVean *et al.* (2002), based on the approximate-likelihood coalescent method of Hudson (2001).

Results

Nucleotide polymorphism and recombination

Figure 1 shows the polymorphic sites detected in the *β-esterase* gene cluster. Totally, there are 236 polymorphic sites in a sample of 78 sequences of the *β-esterase* gene cluster (5394 bp): 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 deletion), four sites in the intron and 16 sites in exon II of *ψEst-6*; and 22 sites in the 3′-flanking region. Eleven length polymorphisms were detected for the whole region but only two in *ψEst-6* (figure 1). In the coding region of *Est-6* there are 20 replacement and 34 synonymous polymorphic sites; in the coding region

of $\boldsymbol{y}$*Est-6* there are 54 replacement (including one site associated with a deletion) and 45 synonymous polymorphic sites. Thus, the ratio of replacement to synonymous polymorphic sites is 0.588 for *Est-6* and 1.2 for $\boldsymbol{y}$*Est-6*. We detected 17 premature stop codons (all TGA) within the coding region of $\boldsymbol{y}$*Est-6* (11 in North American lines and six in European lines). The stop codons are generated either by single mutations (positions 4134, 4454) or by a short insertion, ACATTTGAT, at positions 4445–4453 (figure 1). The *mdg-3* retrotransposon insertion (5.2 kb) was detected within the intron of $\boldsymbol{y}$*Est-6* in the ER-S-438S strain (North American sample, data not shown). Game and Oakeshott (1990) detected the same insertion previously in strain 12I-11.2 of *D. melanogaster*, which carries a null allele of $\boldsymbol{y}$*Est-6*.

Table 1 shows estimates of nucleotide diversity for the $\boldsymbol{b}$*-esterase* gene cluster. For the full sequence, $\boldsymbol{p} = 0.0083$, which is within the range of values observed in other highly recombining gene regions in *D. melanogaster* (Moriyama and Powell 1996). The $\boldsymbol{p}$ value is very similar in the 5′-flanking region ($\boldsymbol{p} = 0.0060$) and *Est-6* ($\boldsymbol{p} = 0.0057$), and higher in the intergenic region ($\boldsymbol{p} = 0.0094$), $\boldsymbol{y}$*Est-6* ($\boldsymbol{p} = 0.0115$) and 3′-flanking region ($\boldsymbol{p} = 0.0086$, for the North American and European samples). Synonymous variation is 0.0160 in the *Est-6* coding region but 1.5 times higher, 0.0244, in $\boldsymbol{y}$*Est-6*. The difference is more pronounced for nonsynonymous variation: 0.0024 in *Est-6* and 3.2 times higher, 0.0076, in $\boldsymbol{y}$*Est-6*. This last difference could reflect different degrees of selective constraint in *Est-6* and $\boldsymbol{y}$*Est-6*. The level of silent divergence between *D. melanogaster* and *D. simulans* is similar for *Est-6* ($K = 0.1474$) and $\boldsymbol{y}$*Est-6* (0.1384), but lower in the 5′-flanking (0.0807) and 3′-flanking (0.0418) regions (the data for the 5′-flanking and 3′-flanking regions are limited to the European, North and South American, and European and North American samples, respectively).

The level of nucleotide diversity, over the entire length of the $\boldsymbol{b}$*-esterase* cluster, is highest in Africa ($\boldsymbol{p} = 0.0113$) and lowest in South America ($\boldsymbol{p} = 0.0052$); intermediate values are observed in Europe ($\boldsymbol{p} = 0.0073$) and North America ($\boldsymbol{p} = 0.0085$) (table 1). The most extreme difference of $\boldsymbol{p}$ is in the 5′-flanking region: 0.0043 (average) in non-African samples versus 0.0126 in Africa. The differences are significant by coalescent simulations.

The method of Hudson and Kaplan (1985) reveals a minimum of 35 recombination events in the whole region; 12 for *Est-6* and 15 for $\boldsymbol{y}$*Est-6*. The rate of recombination (table 2) is highest in Africa, lower in Europe and North America, and lowest in South America, which may be a consequence of the age of the populations, since *D. melanogaster* originates from Africa, whence Europe and the Americas were colonized, so that it reached South America most recently. In all derived populations, the rate of recombination is higher for *Est-6* than for $\boldsymbol{y}$*Est-6*. We have previously conjectured that this difference may be a

consequence of the nonfunctionalization of $\boldsymbol{y}$*Est-6*, involving deterioration of specific recognition sequences that promote recombination (Balakirev *et al.* 2003). The difference between derived and ancestral populations may be due to the joint action of demographic factors and natural selection, as has been also proposed for human populations (Frisse *et al.* 2001).

Gene conversion events are detected within *Est-6* and $\boldsymbol{y}$*Est-6* (except in Venezuela, where they are not significant; table 3, method of Sawyer (1989, 1999)). The number of significant fragments is considerably higher for $\boldsymbol{y}$*Est-6* in all cases, which may be due to the invasion of retrotransposons (like *mdg-3*) that can promote a form of homology-dependent gene conversion upon excision (Engels 1989; Athma and Peterson 1991; Lowe *et al.* 1992; Preston and Engels 1996; Svoboda *et al.* 1996). Transposable elements are flanked by short repeated sequences (see, for example, Sherratt 1995) that serve as targets for the recombination machinery (see, for example, Jeffreys *et al.* 1985; Treco and Arnheim 1986; Collick and Jeffreys 1990). The conversion events are less pronounced and rarely significant in the protein alignment (table 3), which suggests the involvement mostly of silent sites.

Intergenic gene conversion is significant only in the North American sample by use of protein alignment (the implicated fragment is located between amino acids 41 and 55). The nucleotide sequences of the coding regions of *Est-6* and $\boldsymbol{y}$*Est-6* show 64% similarity (Collet *et al.* 1990), which may not satisfy the homology requirements for efficient intergenic conversion. The recombination machinery is sensitive even to a single nucleotide mismatch; individual nucleotide substitutions have been shown to affect recombination in yeast (Borts and Haber 1987; Selva *et al.* 1995), bacteria (Shen and Huang 1986, 1989), and mammalian cells (Lukacsovich and Waldman 1999). Small sequence divergence strongly inhibits intrachromosomal recombination, often by 100-fold to 1000-fold (Claverys and Lacks 1986; Waldman and Liskay 1987; Rayssiguier *et al.* 1989; Bailis and Rothstein 1990; Harris *et al.* 1993; De Wind *et al.* 1995; Matic *et al.* 1995; Selva *et al.* 1995; Chambers *et al.* 1996; Datta *et al.* 1996, 1997; Porter *et al.* 1996; Yang and Waldman 1997; Elliott *et al.* 1998; Chen and Jinks-Robertson 1999). Interlocus gene conversion has been detected between *Est-5A* and *Est-5B* in *D. pseudoobscura* (82.5% similarity) (Brady and Richmond 1992; King 1998), two genes that are orthologous to *Est-6* and $\boldsymbol{y}$*Est-6*.

Haplotype structure

Odgers *et al.* (1995, 2002) have described two groups of haplotypes for the 5′-flanking region of *Est-6* of *D. melanogaster*. We detected two groups of haplotypes both for *Est-6* (including the 5′-flanking region) and $\boldsymbol{y}$*Est-6* in *D. melanogaster* in a population from California (Balakirev
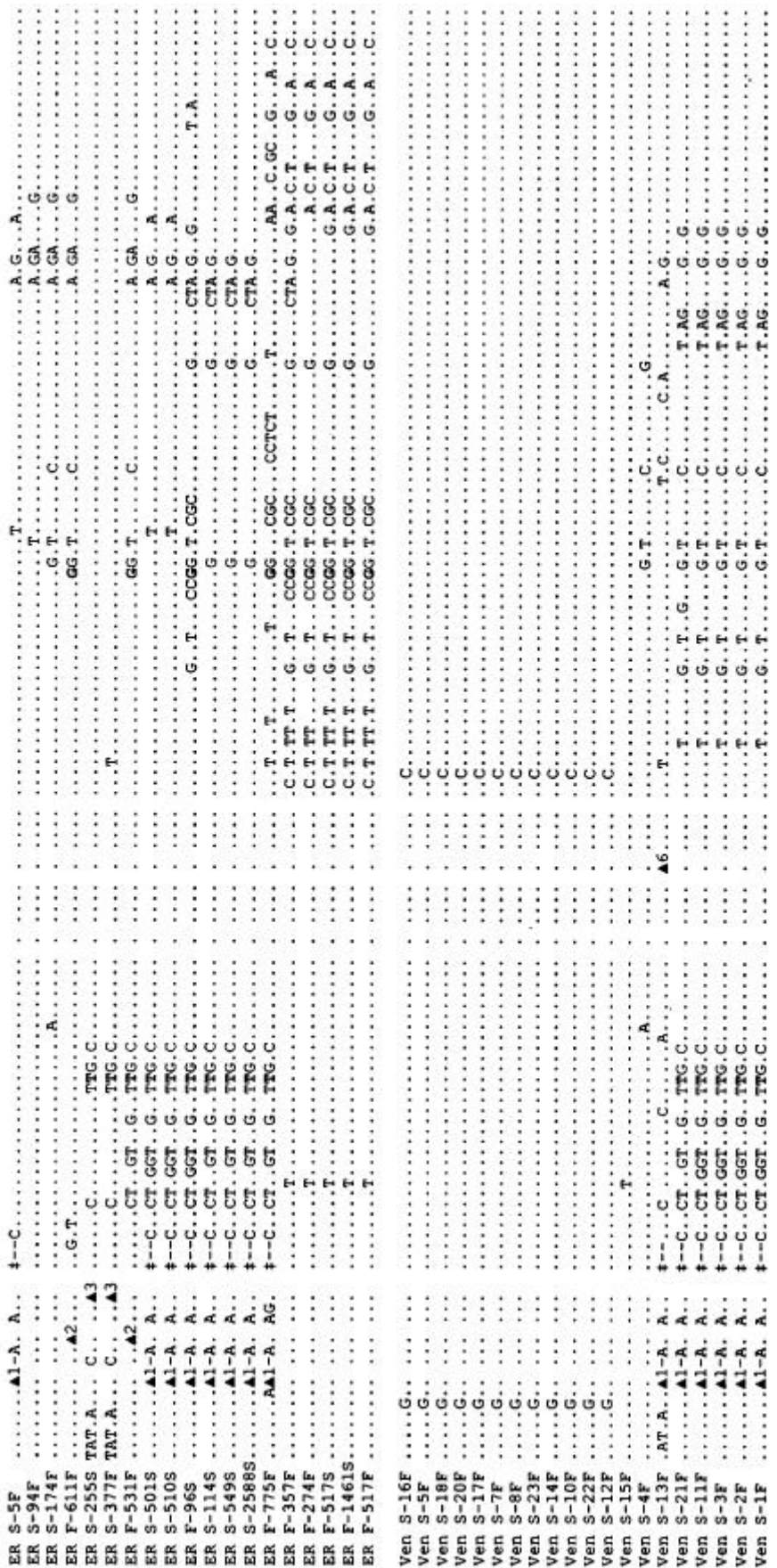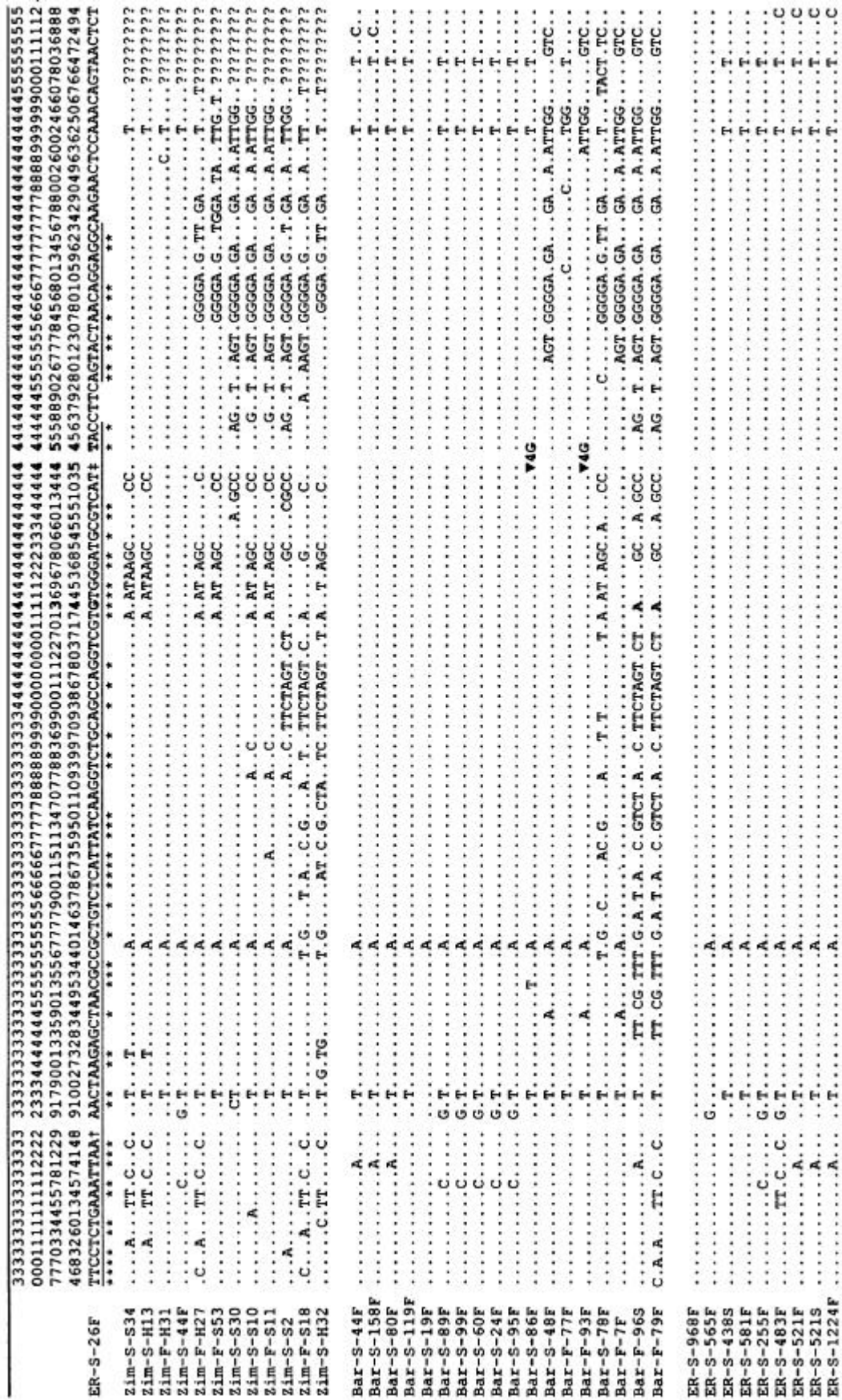
Figure 1. (Caption on p. 121).

Figure 1. (Caption on p. 121).

Figure 1. (Caption on p. 121).

Figure 1. DNA polymorphism in the **b**-esterase gene cluster of *D. melanogaster*. The 78 lines of *D. melanogaster* from southern Africa (Zimbabwe, Zim), Spain (Barcelona, Bar), USA (El Rio, California, ER) and Venezuela (Caracas, Ven) are presented sequentially. The lines within each population are grouped according to their genetic similarity over the full sequence length. The letters S, US and F before the line numbers refer to the EST-6 allozymes Slow, Ultra-Slow, and Fast. The S and F after the number refer to the allozyme polymorphism at the *Sod* locus and have been previously used to tag these lines. The letter after the hyphen in most Zimbabwe lines (except Zim S-44F) refers to the locality of collection (Sengwa or Harare). The line Zim S-44F is from Ayala's laboratory. The numbers above the top sequence 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 *Est-6* and **y***Est-6* are underlined below the reference sequence (ER-S-26F). Amino acid replacement polymorphisms are marked with asterisks. Seventeen premature stop codons are due to single nucleotide polymorphisms G → A (site 4134, strains ER-F-357F, ER-F-274F, ER-F-517F, ER-F-517S, ER-F-1461S, Bar-F-96S and Bar-F-79F) and T → G (site 4454, strains ER-S-501S, ER-S-510S, ER-S-5F, Bar-F-93F and Bar-S-86F), as well as due to a 9-bp insertion of ACATTTGAT (positions 4445–4453, strains ER-S-501S, ER-S-510S, ER-S-5F, Bar-F-93F and Bar-S-86F). These sites are marked by bold lettering. The hyphens represent deleted nucleotides. The question marks indicate missing data, ▲ denotes a deletion, † denotes the absence of a deletion, ‡ denotes the absence of an insertion, ▲ denotes an insertion, ▲7 denotes a 3-bp deletion of CAG, ▼4 denotes the 9-bp insertion of ACATTTGAT.

Line labels (top to bottom):

ER-S-5F
ER-S-94F
ER-S-174F
ER-F-611F
ER-S-255S
ER-S-501F
ER-S-377F
ER-S-531F
ER-S-501S
ER-S-510S
ER-F-96S
ER-S-114S
ER-S-549S
ER-F-775F
ER-F-357F
ER-F-274F
ER-F-517S
ER-F-1461S
ER-F-517F
Ven-S-16F
Ven-S-5F
Ven-S-18F
Ven-S-20F
Ven-S-17F
Ven-S-7F
Ven-S-8F
Ven-S-23F
Ven-S-14F
Ven-S-10F
Ven-S-22F
Ven-S-12F
Ven-S-15F
Ven-S-4F
Ven-S-13F
Ven-S-21F
Ven-S-11F
Ven-S-3F
Ven-S-2F
Ven-S-1F

and Ayala 1996, 2003a; Balakirev *et al*. 1999, 2002, 2003); and also for *Est-6* in European and South American samples (Balakirev and Ayala 2003a). Table 1 shows that the divergence between sequences is greater for **y***Est-6* than for *Est-6* (k is about two times higher, table 1). Figure 2 shows the phylogenetic relationships among the full sequences. It is obvious that the division between haplotypes is not associated with the *Est-6* S/F allozyme variation. Moreover, there is no geographic structure: the haplotypes are interspersed irrespective of geographic origin.

Two significantly divergent sequence types are detected in the South American population, where only the Slow *Est-6* allozyme occurs. The average number of nucleotide differences (k) between the two haplotypes is 11.286 (*Est-6*) and 28.643 (**y***Est-6*). These values are comparable to those between the *Est-6* (k = 11.809) and **y***Est-6* (k = 20.534) allelic lineages of California (Balakirev *et al*. 2002, 2003). The permutation test (Hudson *et al*. 1992a) 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 **y***Est-6*.

The estimates of population differentiation (F_{ST}; Hudson *et al*. 1992a) based on the whole length of the **b**-*esterase* cluster are fairly similar between the pairs Zim–Bar ($F_{ST} = $

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

Sites	5'-Flanking region[a] 619	*Est-6* Syn 379	Nsyn 1253	Total 1686	Intergenic region[a] 193	*y*Est-6 Syn 378	Nsyn 1248	Total 1700	3'-Flanking region[a] 248	Full sequence[b] Syn 757	Nsyn 2501	Total 4198
Total	78 lines											
p	0.0060	0.0160	0.0024	0.0057	0.0094	0.0244	0.0076	0.0115	–	0.0202	0.0050	0.0083
q	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											
p	0.0126	0.0194	0.0035	0.0073	0.0148	0.0296	0.0097	0.0145	0.0184	0.0245	0.0066	0.0113
q	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											
p	0.0043	0.0169	0.0023	0.0058	0.0062	0.0208	0.0068	0.0101	0.0134	0.0189	0.0046	0.0073
q	0.0033	0.0146	0.0019	0.0048	0.0121	0.0254	0.0088	0.0127	0.0115	0.0200	0.0053	0.0081
k	2.621			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											
p	0.0044	0.0152	0.0026	0.0057	0.0141	0.0268	0.0076	0.0122	0.0123	0.0210	0.0051	0.0085
q	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											
p	0.0044	0.0108	0.0013	0.0035	0	0.0171	0.0053	0.0078	–[c]	0.0139	0.0033	0.0052
q	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

p is the average number of nucleotide differences per site among all pairs of sequences, **q** 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* the average number of nucleotide differences, and *K* the average proportion of nucleotide differences between *D. melanogaster* and *D. simulans*.

Syn, synonymous; Nsyn, nonsynonymous. The segregating sites associated with indels are excluded from the **p**, **q** and *K* calculations.

[a]The 5'-flanking and 3'-flanking regions are restricted to the 619 bp and 248 bp, respectively, shared by the Zim, Bar and ER populations.

[b]The full sequence includes 619 bp of the 5'-flanking region, *Est-6*, intergenic region and **y***Est-6* (coordinates 569–4766).

[c]The 3'-flanking region in the Ven population has not been analysed because it includes only 61 bp; it has also not been analysed in the 'total' sample of sequences (Total).

Table 2. Rate of recombination.

	Full sequence			*Est-6*			y*Est-6*		
	Per gene	Per site	**r/q**	Per gene	Per site	**r/θ**	Per gene	Per site	**r/q**
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

The table gives the value of **r**, an estimate of the population recombination rate, $4N_{e}r$ (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). All site types are included in the recombination analysis.

Table 3. Gene conversion.

	Within *Est-6*		Within y*Est-6*		Between *Est-6* and y*Est-6*	
Populations	Nucleotide	Protein	Nucleotide	Protein	Nucleotide	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 involved in the conversion events.

0.1037), Zim–ER ($F_{ST} = 0.0758$), Bar–Ven ($F_{ST} = 0.0702$), and ER–Ven ($F_{ST} = 0.0903$). The maximal and minimal F_{ST} values are obtained, respectively, for the pairs Zim–Ven ($F_{ST} = 0.1504$) and Bar–ER ($F_{ST} = -0.0143$). The permutation method of Hudson *et al*. (1992b) detects significant differentiation between African and all other samples, a result consistent with other data (Begun and Aquadro 1993). The differences between the European and the North or South American samples are not significant ($P > 0.05$).

Sliding-window analysis

Figure 3 shows the distribution of polymorphism along the whole region studied. There is a strong peak in the promoter region around *Rsa*I+/*Rsa*I– polymorphic site and a distinct peak centred on the F/S site of *Est-6*. We previously detected similar peaks in *Est-6* (Balakirev *et al*. 1999, 2002, 2003), in our data as well as in the data of Hasson and Eanes (1996) and Cooke and Oakeshott (1989), and suggested that they may reflect balancing selection (Strobeck 1983; Hudson and Kaplan 1988). There are several y*Est-6* peaks, which repeat with some regularity along the sequence (interval of 200–300 bp), without any consistent relationship to replacement polymorphisms (figure 1).

There are valley regions of nucleotide variation in *Est-6* centred around positions 350, 1200 and 1800 (figure 3). The first valley region includes nearly 400 bp upstream of the *Est-6* coding region. Karotam *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 (Ludwig *et al*. 1993) that are essential for *Est-6* expression. 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 functionally important sites in the catalytic function suggests that the observed valley regions reflect functional constraint.

Figure 4 shows a sliding-window plot of the distribution of nucleotide polymorphism (**p**) in *D. melanogaster* and divergence (*K*) between *D. melanogaster* and *D. simulans*. *K* is significantly low at the end of the 5′-flanking region, the beginning of the *Est-6* gene, and at the beginning and end of y*Est-6*. These low values may be due to functional constraint (the ends of the 5′-flanking region and 3′-flanking region) and gene conversion events (near the beginning of y*Est-6*). Healy *et al*. (1996) have shown that 3′ sequences that lie within the y*Est-6* transcription unit contain elements that modulate the expression of
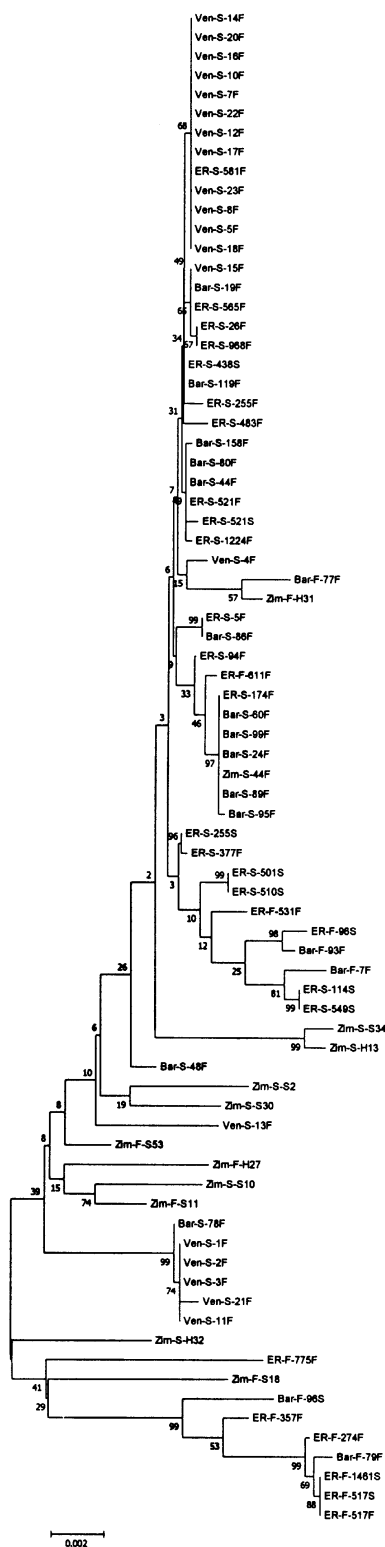
Figure 2. Neighbour-joining tree of the *b-esterase* gene cluster haplotypes of *D. melanogaster*, based on Kimura 2-parameter distances. The tree is based on the complete sequence studied for all samples. The numbers at the nodes are bootstrap per cent probability values based on 10,000 replications.

Est-6, which implies some regulatory function for **y***Est-6*. Moreover, Brady and Richmond (1992) detected some sequence similarity in the 3′-flanking region between **y***Est-6* (*D. melanogaster*) and its orthologue *Est-5A* (*D. pseudo-obscura*). Also, a 390-bp block within the 609-bp flanking sequence 3′ of **y***Est-6* retains similarity with sequences 3′ of *Est-5A* in *D. pseudoobscura* (Collet *et al.* 1990; Brady and Richmond 1992). In particular, in a segment of 110 bp within this region, the two species show 76% sequence similarity, which contrasts sharply with levels of similarity of 20% or lower in the 5′ region between either **y***Est-6* and *Est-5A* or between the orthologues *Est-6* and *Est-5B* (Brady and Richmond 1992). There is an obvious decrease in interspecific divergence (K) in this region (figure 4).

The Goss and Lewontin's (1996) and McDonald's (1996, 1998) tests, with the recombination parameters varying from 1 to 64, do not reveal any significant heterogeneity in the distribution of polymorphic sites relative to fixed interspecific differences for the African, European and North American samples. However, the tests are significant for the South American sample: maximal $G \geq 18.367$ ($P = 0.048$); Var ≥ 0.00094 ($P = 0.012$); modified Var ≥ 0.00207 ($P = 0.021$); average $G \geq 6.690$ ($P = 0.019$).

Linkage disequilibrium

Linkage disequilibrium (LD) was studied by calculating the P value of Fisher's exact test in all pairwise comparisons of polymorphic sites. Figure 5 shows the D distribution along the whole region studied. There is a noticeable peak of high D values around the F/S site of *Est-6* and several peaks along **y***Est-6*. For the whole region, there are 8646 pairwise comparisons; 4092 (47.33%) of them are significant. With the Bonferroni correction, there are *15.07%* significant associations (Bonferroni-corrected values henceforward are given in *italics*). For the 5′-flanking region there are 32.05% (25 out of 78 pairwise comparisons) significant associations (*23.08%*). For the *Est-6* gene 41.48% (219 out of 528) pairwise comparisons are significant (*23.11%*). For the putative pseudogene 62.65% pairwise comparisons (1833 out of 2926) show statistically significant LD (*32.43%*). There are 19.58% (*1.17%*) significant associations between the 5′-flanking region and *Est-6* and 23.68% (*0.80%*) between the 5′-flanking region and **y***Est-6*. Between *Est-6* and **y***Est-6* 46.44% (1180 out of 2541) tests are significant (*5.86%*), which are due to six divergent haplotypes F-517S, F-517F, F-1461S, F-274F, F-357F and F-775F that exhibit unique polymorphisms both in *Est-6* and in **y***Est-6*. Intergenic LD is as pronounced as intragenic LD, which may be caused by epistatic selection reflecting possible functional interaction between *Est-6* and **y***Est-6*. The extent of the LD is low in the African sample (there are only 7.35% significant associations in this sample, but 41.68%,

62.69% and 100% in the European, and North and South American samples, respectively.

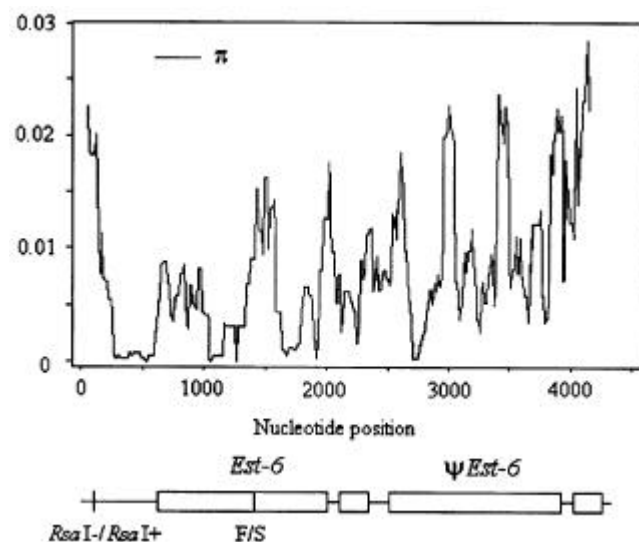We have analysed the relationship between LD and physical distance between sites by the method of McVean *et al.* (2002), with the significance of Pearson's correlation coefficient estimated by 10,000 permutations. For all samples, except the South American, there is a significant decline in LD with increasing distance (table 4); this is not significant in the African sample. The strong haplotype structure and the associated high level of linkage disequilibrium suggest that the South American population originated from a recent admixture of genetically differentiated populations.

Tests of neutrality

Kelly's (1997) Z_{nS} test with recombination (based on linkage disequilibrium between segregating sites) detects significant deviations from neutrality for all populations combined, for the entire region as well as for the separate regions (table 5). Wall's (1999) B and Q tests show similar results (not shown). For ψEst-6 the tests are significant with lower level of recombination than for Est-6. For instance, the Z_{nS} statistic obtained for ψEst-6 is significant ($P = 0.01$) with recombination rate $C = 0.010$, while Est-6 requires $C = 0.035$ (table 5). All non-African samples show significant Z_{nS} with $C \geq 0.005$ ($P = 0.01$). The tests are also significant for the different partitions of the β-esterase gene cluster (5′-flanking region, Est-6 and ψEst-6) in these samples. However, for the African sample Z_{nS} is not significant for any partition or for the entire region, even with the laboratory estimate of recombination rate ($C_{lab} = 0.0664$) based on the physical and genetic maps of *D. melanogaster* (Josep M. Comeron, personal communication; Comeron *et al.* 1999; Balakirev *et al.* 2002). We suggest that the significance of the tests in the promoter region and Est-6 could reflect the action of selection combined with the colonization history of *D. melanogaster*, which originated from Africa, whence it
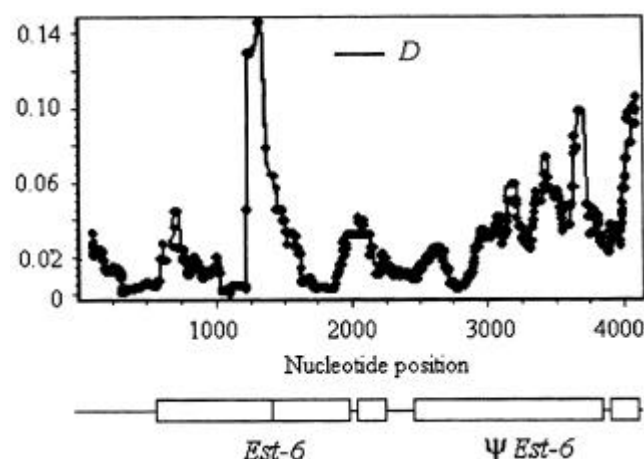

Figure 3. Sliding-window plot of nucleotide diversity (π) in the β-esterase gene cluster of *D. melanogaster*. A schematic representation of the cluster is displayed at bottom. Exons are indicated by boxes; the intron and the 5′-flanking and 3′-flanking regions are shown by thin lines. Window sizes are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

Figure 4. Sliding-window plots of silent nucleotide diversity (π, thin line) and nucleotide divergence (K, thick line) across 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.

Figure 5. Sliding-window plot of linkage disequilibrium (measured by D) along the β-esterase gene cluster of *D. melanogaster*. A schematic representation of the cluster is displayed at bottom. Window sizes are 200 nucleotides with one-nucleotide increments.

migrated in relatively recent times to the rest of the world (Balakirev *et al.* 1999, 2002, 2003; Ayala *et al.* 2002; Balakirev and Ayala 2003a). Balanced selection has also recently been suggested for *D. buzzatii*'s *Est-A* gene (Gomez and Hasson 2003), which is a presumptive homologue of *Est-6*. The significance of the tests for *yEst-6* may reflect the specific character (rare recombination) of the evolution of this putative pseudogene, as well as the demographic history of *D. melanogaster*.

Discussion

1. We have investigated nucleotide polymorphism in the *b*-esterase gene cluster in four populations of *Drosophila melanogaster*: from southern Africa (Zimbabwe), Europe (Spain), North America (California in the US) and South America (Venezuela). A dimorphic haplotype structure exists in the Californian sample, which is not perfectly associated with the *Est-6* allozyme variation (S/F), and in the Venezuelan, where there are no *Est-6* F haplotypes. Less pronounced dimorphism is detected in the Spanish sample and none in the African sample.

The presence of two or more highly diverged haplotypes has been interpreted as a result of positive selection (see, for example, Hudson *et al.* 1994, 1997; Bénassi *et al.* 1999; Labate *et al.* 1999). Teeter *et al.* (2000) investigated single nucleotide polymorphism (SNP) in 66 sequences of *D. melanogaster* spaced at 5 to 20 centimorgan

intervals and generated a map that should not leave any gaps greater than one half of a chromosome arm (Teeter *et al.* 2000). Two-thirds of all sequences were dimorphic. If this result is extrapolated to the whole *D. melanogaster* genome, one site for every few kilobases would need to be subject to strong positive selection (assuming that the dimorphism results from positive selection), which seems highly unlikely (Teeter *et al.* 2000). Alternatively, the admixture of two differentiated populations could account for, and be a more appropriate explanation of, the dimorphism (Teeter *et al.* 2000). A similar suggestion has been made on the basis of nucleotide sequencing, RFLP and allozyme analyses of *D. melanogaster* natural populations (e.g. David and Capy 1988; Singh and Long 1992; Richter *et al.* 1997; Hasson *et al.* 1998).

Our data are compatible with the proposal of Teeter *et al.* (2000) and others. Moreover, we have detected strong dimorphic haplotype structure in three other *D. melanogaster* genes, *Sod* (Hudson *et al.* 1997), *tinman* and *bagpipe* (Balakirev and Ayala 2004), which are also located on the third chromosome. Nevertheless, the *Est-6* data suggest that positive selection may contribute to the observed patterns: balancing selection creates an elevated level of nucleotide variation and linkage disequilibrium around target polymorphic sites (*RsaI*–/*RsaI*+ in the promoter region, and F/S in the coding region), while directional selection creates an excess of very similar sequences exhibiting a very low level of variability

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

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

LD$_{Fisher}$ is the proportion of significant linkage disequilibrium revealed by Fisher's exact test using all polymorphic sites.
All site types are included in the analysis. D' is a measure of linkage disequilibrium, a normalized equivalent of D (see Lewontin 1964).
n.a., Not applicable.

Table 5. Kelly's (1997) test of neutrality for the *b*-esterase gene cluster.

	5′-flanking region (0.6 kb)			*Est-6* region (1.6 kb)			*yEst-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 shows Kelly's (1997) Z_{nS} values and the recombination rates at which the test is significant at the 5% (column 0.05) or 1% (column 0.01) level of significance.
n.s., The value of the test statistic is not significant even with the laboratory-estimated recombination rate (0.0664).

(*Rsa*I– and S allelic lineages in the promoter and coding region, respectively).

The pattern of variability in ***y****Est-6* is more polarized (the dimorphic structure is more marked) than in *Est-6*. This could be a consequence of the superposition effects of balancing selection, recombination and demographic history in *Est-6*. In ***y****Est-6* there is no influence of balancing selection, recombination is limited, and the pattern of variability reflects more adequately the demographic history of the species.

2. The African sample has the highest level of nucleotide diversity and the lowest level of linkage disequilibrium. The Kelly and Wall tests of neutrality with recombination are significant in the non-African samples, but not significant in the African sample. The non-African samples show a pattern of haplotype distribution consistent with a series of selective sweeps 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 48 belong to *Rsa*I–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. It seems likely that bottlenecks associated with colonization events significantly changed the genetic composition of expanding *D. melanogaster* populations (the Barcelona sample has lower variability and higher linkage disequilibrium than the African sample).

The haplotype structure and level of variation in *Est-6* 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, we have detected in the African sample obvious peaks of nucleotide variation centred on functionally important sites in the promoter (*Rsa*I+/*Rsa*I–) and coding region (F/S) (Balakirev *et al.* 1999, 2002, 2003). This observation suggests that the African population is not in mutation–drift equilibrium. The footprints of directional selection have been shown previously for African samples (Mousset *et al.* 2003).

3. The population data available suggest two different migrations of *D. melanogaster* during the expansion period from the African continent: (i) Africa → Europe → North America and (ii) Africa → South America. The second migration is supported by the fact that the southern African and South American samples share a deletion (▲6, figure 1) that is absent in other samples. This deletion is present in five out of 12 African strains, but absent in European and North American samples (figure 1). Indels constitute a valuable source of phylogenetic information (Giribet and Wheeler 1999). That the South American population does not derive from Europe or America is also supported by the absence of the F *Est-6* allele (and of the S *Sod* allele; Hudson *et al.* 1994). The outlined

scheme for expansion of *D. melanogaster* agrees with other data on the evolutionary history of the species (David and Capy 1988; Singh and Long 1992). The South American population of *D. melanogaster* might, nevertheless, represent an admixture of migrants from North America and Africa. The most common haplotype (*Rsa*I-/S) is from North America, while the haplotype *Rsa*I+/S clusters with most of the African samples (figure 2). The admixture would have been recent, since the strong haplotype structure has not been eroded by recombination (the level of linkage disequilibrium is highest in this sample; see above).

4. We have detected some contrasting characteristics of nucleotide variation in *Est-6* and ***y****Est-6*. The total nucleotide variation is 2.1 times higher in ***y****Est-6* than in *Est-6*. The ratio of replacement to synonymous polymorphic sites is 0.588 for *Est-6*, but 1.2 for ***y****Est-6*. There is strong linkage disequilibrium between *Est-6* and ***y****Est-6*. We suggest that intergenic epistatic selection may play a significant role in the evolution of the ***b****-esterase* gene cluster, preserving ***y****Est-6* from degenerative destruction and reflecting possible functional interactions between *Est-6* and ***y****Est-6* (e.g. regulatory interaction; Healy *et al.* 1996). Gene conversion is detected both within and (to a much lesser extent) between *Est-6* and ***y****Est-6*. Extensive intragenic gene conversion within ***y****Est-6* can be explained by the invasion of retrotransposons (like *mdg-3*, see above), which can promote a form of homology-dependent gene conversion upon excision; but also by relaxation of negative selection against intensive intragenic conversion within *Est-6*. The haplotype structure of ***y****Est-6* is dimorphic. However, the divergent sequences of ***y****Est-6* are not perfectly associated with the *Est-6* allozyme variation.

Some of the features of ***y****Est-6* indicate that it could be a pseudogene: 17 premature stop codons are hardly compatible with functionality of the encoded protein; the level of nonsynonymous variation is 3.0 times higher in ***y****Est-6* than in *Est-6*; the results of structural entropy analysis (Balakirev *et al.* 2003) reveal lower structural regularity and higher structural divergence for ***y****Est-6*, in accordance with the expectations if it is a pseudogene or nonfunctional gene. On the other hand, the gene is expressed (Collet *et al.* 1990) and some alleles of ***y****Est-6* produce a catalytically active esterase (Dumancic *et al.* 1997). These contradictory observations, some consistent, but not others, with the hypothesis that ***y****Est-6* is a pseudogene, also apply to other putative pseudogenes in *Drosophila* as well as in a variety of organisms (Balakirev and Ayala 2003b,c). A pseudogene may lose the initial specific coding function but retain or acquire others, which may not be simply recognizable. Brosius and Gould (1992) 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 potential for becoming new genes or acquire new functions. Pseu-

dogenes are surprisingly often extremely conserved and transcriptionally active. Moreover there are direct indications that pseudogenes can be functional by being involved in regulation of gene expression and genetic diversity generation (Balakirev and Ayala 2003b,c; Balakirev *et al.* 2003). The data for *ψEst-6* are in accordance with this general picture (Balakirev and Ayala 1996; Balakirev *et al.* 2003). The observed patterns, however, are contrary to the traditional view that defines pseudogenes as nonfunctional, neutrally evolving sequences of genomic DNA ('junk' DNA). Our observations help to understand why eukaryote genomes contain many pseudogenes that appear to have avoided full degeneration. Pseudogenes may be an important part of the genome, representing a repertoire of sequences evolving towards the acquisition of new or changing functions. Pseudogenes along with their parental sequences may constitute indivisible, functionally interacting entities ('intergenic complexes' or 'intergenes') in which a single component cannot itself successfully accomplish the final functional role. The *Est-6–ψEst-6* complex in *D. melanogaster* may represent such an intergene (for other examples see Balakirev and Ayala 2003b,c). The *Est-6* gene may play the structural role (coding for the EST-6 enzyme) in this complex while *ψEst-6* may enhance genetic variation in the *Est-6* gene and contribute to regulation of its expression.

The population dynamics of pseudogenes is largely unexplored. It is, however, unwarranted to assume that 'all mutations occurring in pseudogenes are selectively neutral and become fixed in the population with equal probability' (Graur and Li 2000, p. 124). Once a pseudogene appears in a population, presumably in a single genome, by duplication or retroposition, it may at first follow neutral population dynamics and may or may not become extinct. However, a pseudogene or gene duplication may sometimes be immediately advantageous and maintained by selection, for example if it (or its flanking regions) contains regulatory elements that interact positively with its homologue ancestor. A pseudogene may rapidly become fixed if it is closely linked to a gene experiencing a selective sweep, or slowly increase in frequency by neutral drift. However, even if a pseudogene is at first not subject to natural selection, any new mutant that provides a functional role for the pseudogene may be favoured by selection and, thus, have a higher probability of becoming fixed than its nonfunctional alleles (Lynch and Conery 2000; Lynch *et al.* 2001; Van de Peer *et al.* 2001; Kondrashov *et al.* 2002; Mazet and Shimeld 2002; Prince and Pickett 2002; Thornton and Long 2002; Wagner 2002).

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

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

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

Bailis A. M. and Rothstein R. 1990 A defect in mismatch repair in *Saccharomyces cerevisiae* stimulates ectopic recombination between homeologous genes by an excision repair dependent process. *Genetics* **126**, 535–547.

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

Balakirev E. S. and Ayala F. J. 2003a Nucleotide variation of the *Est-6* gene region in natural populations of *Drosophila melanogaster. Genetics* **165**, 1901–1914.

Balakirev E. S. and Ayala F. J. 2003b Pseudogenes: are they "junk" or functional DNA? *Annu. Rev. Genet.* **37**, 123–151.

Balakirev E. S. and Ayala F. J. 2003c Pseudogenes are not junk DNA. In *Evolutionary theory and processes: modern horizons* (ed. S. P. Wasser), pp. 1–17. Kluwer, Amsterdam.

Balakirev E. S. and Ayala F. J. 2004 Nucleotide variation in the *tinman* and *bagpipe* homeobox genes of *Drosophila melanogaster. Genetics* (in press).

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

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

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

Bénassi V., Depaulis F., Meghlaoui G. K. and Veuille M. 1999 Partial sweeping of variation at the *Fbp2* locus in a West African population of *Drosophila melanogaster. Mol. Biol. Evol.* **16**, 347–353.

Borts R. H. and Haber J. E. 1987 Meiotic recombination in yeast: alteration by multiple heterozygosities. *Science* **237**, 1459–1465.

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

Brady J. P., Richmond R. C. and Oakeshott J. G. 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. and Gould S. J. 1992 On "genomenclature": A comprehensive (and respectful) taxonomy for pseudogenes and other "junk DNA". *Proc. Natl. Acad. Sci. USA* **89**, 10706–10710.

Chambers S. R., Hunter N., Louis E. J. and Borts R. H. 1996 The mismatch repair system reduces meiotic homeologous recombination and stimulates recombination-dependent chromosome loss. *Mol. Cell. Biol.* **16**, 6110–6120.

Chen W. and Jinks-Robertson S. 1999 The role of mismatch repair machinery in regulating mitotic and meiotic recombination between diverged sequences in yeast. *Genetics* **151**, 1299–1313.

Claverys J. P. and Lacks S. A. 1986 Heteroduplex deoxyribonucleic acid base mismatch repair in bacteria. *Microbiol. Rev.* **50**, 133–165.

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

Collick A. and Jeffreys A. J. 1990 Detection of a novel minisatellite-specific DNA-binding protein. *Nucl. Acids Res.* **18**, 625–629.

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

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

Datta A., Adjiri A., New L., Grouse G. F. and Jinks-Robertson S. 1996 Mitotic crossovers between diverged sequences are regulated by mismatch repair proteins in *Saccharomyces cerevisiae. Mol. Cell. Biol.* **176**, 1085–1093.

Datta A., Hendrix M., Lipsitch M. and Jinks-Robertson S. 1997 Dual roles for DNA sequence identity and the mismatch repair system in the regulation of mitotic crossing-over in yeast. *Proc. Natl. Acad. Sci. USA* **94**, 9757–9762.

David J. R. and Capy P. 1988 Genetic variation of *Drosophila melanogaster* natural populations. *Trends Genet.* **4**, 106–111.

De Wind N., Dekker M., Berns A., Radman M. and Riele H. T. 1995 Inactivation of the mouse *Msh2* gene results in mismatch repair deficiency, methylation tolerance, hyperrecombination and predisposition to cancer. *Cell* **82**, 321–330.

Dumancic M. M., Oakeshott J. G., Russell R. J. and Healy M. J. 1997 Characterization of the *EstP* protein in *Drosophila melanogaster* and its conservation in Drosophilids. *Biochem. Genet.* **35**, 251–271.

East P. D., Graham A. and Whitington G. 1990 Molecular isolation and preliminary characterization of a duplicated esterase locus in *Drosophila buzzatii.* In *Ecological and evolutionary genetics of Drosophila* (ed. J. S. F. Barker, W. T. Starmer and R. J. MacIntyre), pp. 389–406. Plenum, New York.

Elliott B., Richardson C., Winderbaum J., Nickoloff J. A. and Jasin M. 1998 Gene conversion tracts in mammalian cells from double-strand break repair. *Mol. Cell. Biol.* **18**, 93–101.

Engels W. R. 1989 P elements in *Drosophila melanogaster.* In *Mobile DNA* (ed. D. Berg and M. Howe), pp. 437–484. American Society for Microbiology, Washington, D. C.

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

Frisse L., Hudson R. R., Bartoszewicz A., Wall J. D., Donfack J. and Di Rienzo A. 2001 Gene conversion and different population histories may explain the contrast between polymor-

phism and linkage disequilibrium levels. *Am. J. Hum. Genet.* **69**, 831–843.

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

Giribet G. and Wheeler W. C. 1999 On gaps. *Mol. Phylogenet. Evol.* **13**, 132–143.

Gomez G. A. and Hasson E. 2003 Transpecific polymorphisms in an inversion linked esterase locus in *Drosophila buzzatii. Mol. Biol. Evol.* **20**, 410–423.

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

Graur D. and Li W.-H. 2000 *Fundamentals of molecular evolution,* 2nd edition. Sinauer, Sunderland.

Gromko M. H., Gilbert D. F. and Richmond R. C. 1984 Sperm transfer and use in the multiple mating system of *Drosophila.* In *Sperm competition and the evolution of animal mating systems* (ed. R. L. Smith), pp. 371–426. Academic Press, New York.

Harris S., Rudnicki K. S. and Haber J. E. 1993 Gene conversions and crossing over during homologous and homeologous ectopic recombination in *Saccharomyces cerevisiae. Genetics* **135**, 5–16.

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

Hasson E., Wang I.-N., Zeng L.-W., Kreitman M. and Eanes W. F. 1998 Nucleotide variation in the Triosephosphate isomerase (*Tpi*) locus of *Drosophila melanogaster* and *Drosophila simulans. Mol. Biol. Evol.* **15**, 756–769.

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

Healy M. J., Dumancic M. M., Cao A. and Oakeshott J. G. 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. 1983 Properties of a neutral allele model with intragenic recombination. *Theor. Popul. Biol.* **23**, 183–201.

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

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

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

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

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

Jeffreys A. J., Wilson V. and Thein S. L. 1985 Hypervariable minisatellite regions in human DNA. *Nature* **314**, 67–73.

Karotam J., Delves A. C. and Oakeshott J. G. 1993 Conservation and change in structural and 5′ flanking sequences of esterase 6 in sibling *Drosophila* species. *Genetica* **88**, 1–28.

Karotam J., Boyce T. M. and Oakeshott J. G. 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.

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.

Kondrashov F. A., Rogozin I. B., Wolf Y. I. and Koonin E. V. 2002 Selection in the evolution of gene duplication. *Genome Biol.* **3** (2), research0008.1-0008.9.

Korochkin L. I., Ludwig M. Z., Poliakova E. V. and Philinova M. R. 1987 Some molecular genetic aspects of cellular differentiation in *Drosophila*. *Sov. Sci. Rev. F.* **1**, 411–466.

Korochkin L., Ludwig M., Tamarina N., Uspensky I., Yeni-kolopov G., Khechumijan R. *et al.* 1990 Molecular genetic mechanisms of tissue-specific esterase isozymes and protein expression in *Drosophila*. In *Isozymes: structure, function, and use in biology and medicine* (ed. C. Markert and J. Scandalios) pp. 399–440. Wiley-Liss, New York.

Labate J. A., Biermann C. H. and Eanes W. F. 1999 Nucleotide variation at the *runt* locus in *Drosophila melanogaster* and *Drosophila simulans*. *Mol. Biol. Evol.* **16**, 724–731.

Lewontin R. C. 1964 The interaction of selection and linkage. I. General considerations; heterotic models. *Genetics* **49**, 49–67.

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

Ludwig M. Z., Tamarina N. A. and Richmond R. C. 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.

Lukacsovich T. and Waldman A. S. 1999 Suppression of intrachromosomal gene conversion in mammalian cells by small degrees of sequence divergence. *Genetics* **151**, 1559–1568.

Lynch M. and Conery J. S. 2000 The evolutionary fate and consequences of duplicate genes. *Science* **290**, 1151–1155.

Lynch M., O'Hely M., Walsh B. and Force A. 2001 The probability of preservation of a newly arisen gene duplicate. *Genetics* **159**, 1789–1804.

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. A. T. 2001 What do patterns of genetic variability reveal about mitochondrial recombination? *Heredity* **87**, 613–620.

McVean G., Awadalla P. and Fearnhead P. 2002 A coalescent-based method for detecting and estimating recombination from gene sequences. *Genetics* **160**, 1231–1241.

Matic I., Rayssiguier C. and Radman M. 1995 Interspecies gene exchange in bacteria: the role of SOS and mismatch repair systems in evolution of species. *Cell* **80**, 507–515.

Mazet F. and Shimeld S. M. 2002 Gene duplication and divergence in the early evolution of vertebrates. *Curr. Opin. Genet. Dev.* **12**, 393–396.

Moriyama E. N. and Powell J. R. 1996 Intraspecific nuclear DNA variation in *Drosophila. Mol. Biol. Evol.* **13**, 261–277.

Mousset S., Brazier L., Cariou M.-L., Chartois F., Depaulis F. and Veuille M. 2003 Evidence of a high rate of selective sweeps in African *Drosophila melanogaster*. *Genetics* **163**, 599–609.

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

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

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

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

Odgers W. A., Healy M. J. and Oakeshott J. G. 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., Aquadro C. F., Coppin C. W., Healy M. J. and Oakeshott J. G. 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.

Porter G., Westmoreland J., Priebe S. and Resnick M. A. 1996 Homologous and homeologous intermolecular gene conversion are not differentially affected by mutations in the DNA damage or mismatch repair genes *RAD1*, *RAD50*, *RAD52*, *RAD54*, *PMS1*, or *MSH2*. *Genetics* **143**, 755–767.

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

Prince V. E. and Pickett F. B. 2002 Splitting pairs: the diverging fates of duplicated genes. *Nat. Rev. Genet.* **3**, 827–837.

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

Rayssiguier C., Thaler D. S. and Radman M. 1989 The barrier to recombination between *Escherichia coli* and *Salmonella typhimurium* is disrupted in mismatch repair mutants. *Nature* **342**, 396–401.

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

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

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

Rozas J. and Rozas R. 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, St Louis, USA.

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

Selva E. M., New L., Crouse G. F. and Lahue R. S. 1995 Mismatch correction acts as a barrier to homologous recombination in *Saccharomyces cerevisiae*. *Genetics* **139**, 1175–1188.

Shen P. and Huang H. V. 1986 Homologous recombination in *Escherichia coli*: dependence on substrate length and homology. *Genetics* **112**, 441–457.

Shen P. and Huang H. V. 1989 Effect of base pair mismatches on recombination via the RecBCD pathway. *Mol. Gen. Genet.* **218**, 358–360.

Sherratt D. J. (ed.) 1995 *Mobile genetic elements*. Oxford University Press. Oxford.

Singh R. S. and Long A. D. 1992 Geographic variation in *Drosophila*: from molecules to morphology and back. *Trends Ecol. Evol.* **7**, 340–345.

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

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

Teeter K., Naeemuddin M., Gasperini R., Zimmerman E., White K. P., Hoskins R. and Gibson G. 2000 Haplotype dimorphism in a SNP collection from *Drosophila melanogaster*. *J. Exp. Zool.* **288**, 63–75.

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

Thornton K. and Long M. 2002 Rapid divergence of gene duplicates on the *Drosophila melanogaster* X chromosome. *Mol. Biol. Evol.* **19**, 918–925.

Treco D. and Arnheim N. 1986 The evolutionarily conserved repetitive sequence d(TG AG)ₙ promotes reciprocal exchange and generates unusual recombinant tetrads during yeast meiosis. *Mol. Cell. Biol.* **6**, 3934–3947.

Van de Peer Y., Taylor J. S., Braasch I. and Meyer A. 2001 The ghost of selection past: rates of evolution and functional divergence of anciently duplicated genes. *J. Mol. Evol.* **53**, 436–446.

Wagner A. 2002 Asymmetric functional divergence of duplicate genes in yeast. *Mol. Biol. Evol.* **19**, 1760–1768.

Waldman A. S. and Liskay R. M. 1987 Differential effects of base-pair mismatch on intrachromosomal versus extrachromosomal recombination in mouse cells. *Proc. Natl. Acad. Sci. USA* **84**, 5340–5344.

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.

Yang D. and Waldman A. S. 1997 Fine-resolution analysis of products of intrachromosomal homeologous recombination in mammalian cells. *Mol. Cell. Biol.* **17**, 3614–3628.

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