

UC Irvine

UC Irvine Previously Published Works

Title

Genetic polymorphism at two linked loci, Sod and Est-6, in *Drosophila melanogaster*

Permalink

<https://escholarship.org/uc/item/6tj0z4bs>

Journal

Gene, 300(1-2)

ISSN

0378-1119

Authors

Ayala, Francisco J
Balakirev, Evgeniy S
Sáez, Alberto G

Publication Date

2002-10-01

DOI

10.1016/s0378-1119(02)00852-1

Copyright Information

This work is made available under the terms of a Creative Commons Attribution License, available at <https://creativecommons.org/licenses/by/4.0/>

Peer reviewed



ELSEVIER

Gene 300 (2002) 19–29

GENE
AN INTERNATIONAL JOURNAL ON
GENES AND GENOMES

www.elsevier.com/locate/gene

Genetic polymorphism at two linked loci, *Sod* and *Est-6*, in *Drosophila melanogaster*[☆]

Francisco J. Ayala^{a,*}, Evgeniy S. Balakirev^{a,b}, Alberto G. Sáez^{a,c}

^aDepartment of Ecology and Evolutionary Biology, University of California, 321 Steinhaus Hall, Irvine, CA 92697-2525, USA

^bInstitute of Marine Biology, Vladivostok 690041 and Academy of Ecology, Marine Biology, and Biotechnology, Far Eastern State University, Vladivostok 690600, Russia

^cAlfred Wegener Institute, Am Handelshafen 12, D-27570 Bremerhaven, Germany

Received 20 December 2001; received in revised form 6 June 2002; accepted 17 July 2002

Abstract

We have examined the patterns of polymorphism at two linked loci, *Sod* and *Est-6*, separated by nearly 1000 kb on the left arm of chromosome 3 of *Drosophila melanogaster*. The evidence suggests that natural selection has been involved in shaping the polymorphisms. At the *Sod* locus, a fairly strong ($s > 0.01$) selective sweep, started ≥ 2600 years ago, increased the frequency of a rare haplotype, F(A), to about 50% frequency in populations of Europe, Asia, and the Americas. More recently, an F(A) allele mutated to an S allele, which has increased to frequencies 5–15% in populations of Europe, Asia and North America. All S alleles are identical (or very nearly) in sequence and differ by one nucleotide substitution (which accounts for the F \rightarrow S electrophoretic difference) from F(A) alleles. At the *Est-6* locus, the evidence indicates both directional and balancing selection impacting separately the promoter and the coding regions of the gene, with linkage disequilibrium occurring within each region. Some linkage disequilibrium also exists between the two genes. © 2002 Elsevier Science B.V. All rights reserved.

Keywords: Natural selection; Directional selection; Selective sweep; Linkage disequilibrium; Balanced polymorphism

1. Introduction

The neutrality theory of evolution recognizes that the morphological, functional and behavioral features of organisms evolve by natural selection operating through adaptive changes in the DNA. Proponents of the theory argue that, nevertheless, much evolutionary change in the DNA (and therefore in the proteins encoded in the DNA) occurs by the random process of sampling errors through the generations (Kimura, 1968, 1983 and references therein; King and Jukes, 1969; Kimura and Ohta, 1971).

The neutrality theory assumes that a large fraction of all newly arising mutations are unconditionally deleterious. These harmful genetic variants are eliminated or kept at very low frequencies by natural selection. Proponents of the theory argue that at many, perhaps most, gene loci there are a large number of mutants that are effectively equivalent

with respect to adaptation. These are functional mutants, any one of which is favorably selected relative to the deleterious ones, but the carriers of alternative adaptive genotypes do not differ in their adaptedness to the environment; their frequencies in populations are, therefore, not affected by natural selection. Since natural populations consist of finite numbers of individuals, the frequencies of neutral mutants will change from generation to generation due to sampling drift. It is therefore predicted that DNA or protein sequence differences between species are for the most part due to random processes of change, not to natural selection. Similarly, the pervasive DNA and protein polymorphisms observed in natural populations represent, according to the neutrality theory, transient conditions in populations going from fixation for one allele towards fixation for another, adaptively equivalent allele.

It should also be clear from the outset that the theory of evolution by natural selection, on its part, recognizes that genetic frequencies are affected by the stochastic process of sampling errors. The question between the two theories is whether protein differences between species and protein polymorphisms within populations are due to stochastic processes alone (or mostly), as claimed by the neutrality

Abbreviations: *Sod*, superoxide dismutase; *Est-6*, Esterase-6.

[☆] Presented at the workshop 'Natural Selection and the Neutral Theory,' Ischia (Naples), October 28–30, 2001.

* Corresponding author. Tel.: +1-949-824-8293; fax: +1-949-824-2474.

E-mail address: fjayala@uci.edu (F.J. Ayala).

theory; or whether nonrandom processes must be postulated, as claimed by the theory of natural selection.

The neutrality theory of evolution is rich in empirical content, according to Popper's criterion of falsifiability. The theory makes precise predictions about the patterns of genetic polymorphisms in populations and of differences between populations. It can, therefore, be subject to critical tests by examining the congruence, or lack thereof, between the predictions derived from the theory and the results of relevant observations and experiments. As Kimura and Ohta (1971) stated, "The neutral...theory allows us to make a number of definite quantitative as well as qualitative predictions by which the theory can be tested. We hope that through this process (of testing) we will be able to gain deeper understanding of the mechanisms of evolution at the molecular level". The neutrality theory of evolution has, indeed, been subject to a variety of empirical tests.

One family of tests concerns the expectation of rate constancy in the molecular evolution of genes and proteins. Kimura and Ohta (1971; see Kimura, 1983; Gillespie, 1991) predicted that the number of (nucleotide or amino acid) substitutions (with mean $M = kt$, where k is the rate of neutral mutation and t is the time elapsed in years) has a Poisson distribution, which has the property that the variance, V , is equal to the mean, so that expected value of the variance-to-mean ratio, $R = V/M = 1$.

A common observation, however, is that genes and proteins evolve more erratically than allowed by the neutral theory (the so-called overdispersed molecular clock; Gillespie, 1989, 1991), which casts doubts on the validity of the molecular clock model (Gillespie, 1991; Ayala, 1999). Several subsidiary hypotheses have been proposed that modify the predictions of the neutrality theory, allowing for greater variance in evolutionary rates (Ayala, 1999). It has been shown, however, that the predictions obtained from anyone of the neutrality-derived hypotheses do not hold. Detailed studies manifest that the rates of molecular evolution vary from one lineage to another, and from one time span to another, in disparate patterns when different genes are investigated (Ayala, 1997, 1999; Rodriguez-Trelles et al., 2001).

In the present paper, we present results from a very different kind of test, namely the patterns of polymorphism observed in populations of a given species. We examine the

distribution of DNA sequence variance in each of two genes, and between them, in a natural population of *Drosophila melanogaster*.

2. Polymorphism at the Cu,Zn superoxide dismutase (*Sod*) locus

The superoxide dismutases are abundant enzymes in aerobic organisms, with highly specific superoxide dismutation activity that protects the cell against the harmfulness of free oxygen radicals (Fridovich, 1986). These enzymes are neither parts of structural proteins nor involved in intermediate metabolism, providing a distinctive situation to be studied by population geneticists. These enzymes have active centers that contain either iron or manganese, or both copper and zinc (Fridovich, 1986). The Cu,Zn superoxide dismutase (SOD) is a well-studied protein, found in eukaryotes but also in some bacteria (Steinman, 1988).

The SOD of *D. melanogaster* is a dimer molecule consisting of two identical polypeptide subunits associated with two Cu^{2+} and two Zn^{2+} per molecule. Each subunit has a molecular weight of 15,750 and consists of 151 amino acids, the same as in other fruit fly species. The *Drosophila Sod* gene consists of two exons, separated by an intron 300–700 bp in length, located between codons 22 and 23. The *Sod* locus is on the left arm of chromosome 3, located at bands 68A7 in the cytological map, and at 35.9 in the genetic map (Fig. 1).

Many natural populations are segregating for two electrophoretically distinguishable alleles (*Sod*^S and *Sod*^F; hence forward the Slow or S, and Fast or F alleles) (Singh et al., 1982; Peng et al., 1986; Hudson et al., 1994, 1997; see Table 1). The S and F alleles differ by a single amino acid residue, lysine 96 in S but asparagine 96 in F (Lee and Ayala, 1985). The S and F enzymes differ in properties such as specific activity and thermostability (Lee et al., 1981; Graf and Ayala, 1986). Laboratory experiments have shown that the Slow and Fast alleles, or variation in linkage disequilibrium with them, have large fitness effects in the presence of ionizing radiation (Peng et al., 1986), as well as under different conditions of temperature and larval crowding (Peng et al., 1991). Experiments with laboratory populations of *D. melanogaster* suggest that variation at the

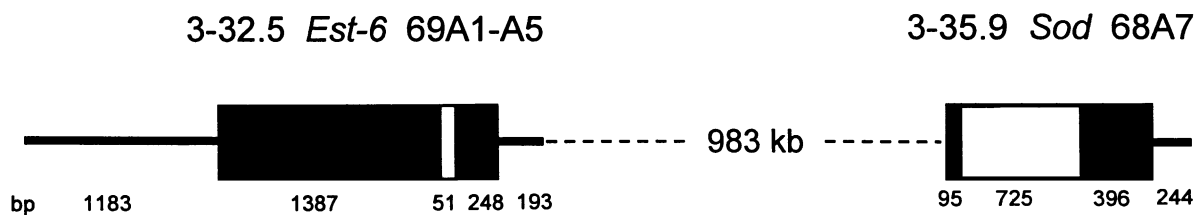


Fig. 1. Schematic representation of a left-arm segment of the third chromosome of *Drosophila melanogaster* that includes the *Est-6* and *Sod* genes. The coding regions are represented as boxes, with the exons in black. Exon I is to the left in both genes. The 5' and 3' untranslated regions are shown as lines. The number of base pairs sequenced are shown below. The locus of each gene in the genetic map and in the polytene chromosome is shown above.

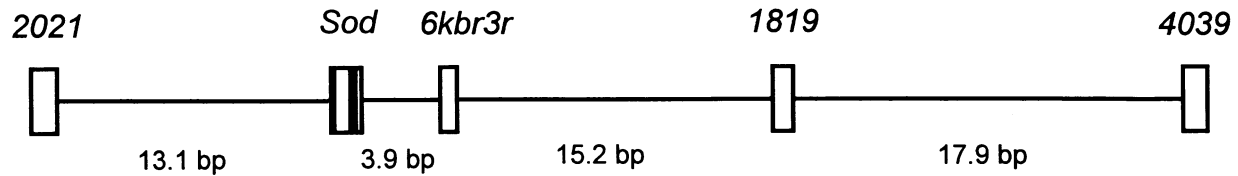


Fig. 2. Schematic representation of a 50 kb-long segment of the third-chromosome left arm of *Drosophila melanogaster* that includes the *Sod* gene. This and four other sequenced fragments are shown as boxes. The number of bases sequenced are 1185, 1410 (*Sod* gene), 759, 899, and 1130, respectively, for the segments from left to right. The 4039 segment was not sequenced at first, but only in a subsequent experiment (see text).

Fifteen lines of *D. melanogaster* from El Rio vineyard (Lockeford, San Joaquin County, CA) and the Canton S strain of *D. melanogaster* were sequenced at the *Sod* locus and three neighboring regions (Fig. 2). The three neighboring regions, denoted, 2021, *6kbr3r*, and 1819, are located approximately 12.7 kb upstream of *Sod*, 3.7 kb downstream of *Sod*, and 19.2 kb downstream of *Sod*, respectively. (The 4039 segment, which is 38.0 kb downstream of *Sod*, was sequenced in a subsequent experiment.)

The patterns of variation observed in the linked regions are as follows. Both the *6kbr3r* and the 1819 region show patterns of variation that are very similar to the pattern observed at *Sod*. That is, most of the sequences are very similar to each other, forming a very homogeneous subset, whereas the other sequences (between 2 and 5, of 16 lines) are relatively diverged from the homogeneous subset, and show some divergence among themselves. Thus, the region showing the *Sod* pattern of variation encompasses both *6Kbr3r* and 1819, i.e. it extends at least 20 kb downstream from *Sod*. As indicated earlier, if this pattern of variation is due to a partial selective sweep, quite strong selection is required (selection coefficient on the order of 0.01).

A second important feature of the data from these downstream regions is the pattern of recombination. We note that some of the lines, which form part of the homogeneous subset at the *Sod* locus, are lines that constitute part of the heterogeneous subset at the *6kbr3r* region (Hudson et al., 1997). The lack of complete linkage disequilibrium between *Sod* and *6kbr3r* suggests that the selective sweep is not extremely recent. To be somewhat more precise, since linkage disequilibrium decays approximately as $\exp(-rt)$, where r is the recombination rate per generation and t is the time measured in generations, we can estimate t . Because linkage disequilibria between sites in the *Sod* locus and sites in the *6kbr3r* region are substantially decayed, rt is unlikely to be less than one. The recombination rate in *D. melanogaster* females at *Sod* is estimated to be about 3.8×10^{-5} /kb per generation (based on Recomb-Rate v1.0; Comeron et al., 1999). Taking into account that recombination does not occur in males, we estimate that the recombination rate between *Sod* and *6kbr3r*, which are approximately 4 kb apart, is approximately 7.6×10^{-5} . This suggests that the time since the selective sweep is roughly 13,000 generations ($= 1.0 / \{7.6 \times 10^{-5}\}$) or longer. (13,000 generations correspond to 2600 years, assuming an average of five generations per year.)

The time since the putative selective event can also be inferred from the amount of variation that has accumulated within the relatively homogeneous subsets as follows (Hudson et al., 1997). We start by assuming provisionally that the sampled lines that constitute the homogeneous subset are related by a star genealogy (i.e. all lineages of the sampled regions remain distinct back to a time near the time of the selective event). This is a reasonable assumption if the effective population size is large and the selective event recent. The low observed frequency of most variants in the homogeneous set is consistent with this assumption. In the *Sod* region 3 lines (581F, 498F, and 968F) constitute a heterogeneous and diverged subset, while the other 13 lines constitute the homogeneous subset. There are nine polymorphic sites in this homogeneous subset. Similarly, in the *6kbr3r* region there are ten sequences in the homogeneous subset, and there are two sites polymorphic; and finally in the 1819 region there are 14 lines in the homogeneous subset, with 12 polymorphic sites. The sequence in the *Sod* region is 1408 bp long, of which 439 bp are protein coding. Since in protein coding sequences about 25% of changes are synonymous, the sequenced *Sod* is approximately equivalent to 1079 ($= 1408 - 0.75 \times 439$) bp of noncoding sequence. Assuming that the other sequenced regions are noncoding and denoting the neutral mutation rate at noncoding and silent sites by μ (assumed to be 16×10^{-9} per site per year; see Sharp and Li, 1989; Rowan and Hunt, 1991), we find that the expected number of polymorphic sites in the homogeneous subset is μt ($13 \times 1079 + 10 \times 764 + 14 \times 937$) $= \mu t \times 34,785$, where t is the time back to the selection event (in years). If we set $\mu t \times 34,785$ equal to 23, which is the observed number of polymorphisms in the homogeneous subset, and solve for t , we find $t \approx 41,000$ years. Seven polymorphisms within the homogeneous subset could be the result of conversion from haplotypes in the heterogeneous subset (Hudson et al., 1997), which leaves only 17 mutations and leads to an estimate of t of 31,000 years. Other polymorphisms in the homogeneous subset could have resulted from conversion or recombination and, in addition, the neutral mutation rate that we have used is based on substitution rates at silent sites in coding regions and may underestimate the neutral mutation rate in noncoding regions. Hence, our estimate may be biased upward, but is consistent with our conclusion from the pattern of recombination, which is that the selective event is about or older than 2600 years.

The pattern of variation in the, 2021 region, which is 12.7 kb upstream from the *Sod* locus, is completely different. There is no homogeneous subset of any appreciable size. There is a high level of polymorphism and no hint of the partial selective sweep evident in the other regions. The 2021 region is apparently outside the region of the partial sweep, indicating that the upstream boundary of the selective sweep is somewhere between the *Sod* locus and the 2021 region.

In conclusion, the Slow/Fast polymorphism of *Sod* is clearly not an old balanced polymorphism. The data suggest that natural selection has acted recently and strongly on variation in the neighborhood of *Sod*. The observations indicate that the swept region is greater than 20 kb in length, which implies that surprisingly strong selection has acted on the selected site (selection coefficient on the order of 0.01 or higher).

We have recently sequenced a 4039 bp-long segment, about 38 kb downstream from the *Sod* locus (see Fig. 2) in the same 15 strains of *D. melanogaster*. This segment exhibits a polymorphism intermediate between the pattern of the *Sod* locus (as well as *6kbr3r* and *1819*) and that of segment 2021, indicating that this new segment has been only partially affected by the selective sweep. The total length of DNA sequence encompassed between segments 2021 and 4039 is 55,513 bp. Thus, the length of the chromosomal segment impacted by the selective sweep can be estimated between 41 and 54 kb. This leads to estimating the age of the selective sweep between 2600 and 22,000 years, and a selective advantage of $0.020 < s < 0.027$. The selective sweep may have been downstream from *Sod*, some 12–18 kb from this gene if we assume that the boundaries of the selective sweep are approximately equidistant from the selective site.

3. Polymorphism at the Esterase-6 (*Est-6*) locus

The *Est-6* gene is on the left arm of chromosome 3 of *D. melanogaster*, mapped at bands 69A1–A5 in the cytological map and at 32.5 in the genetic map (Fig. 1). The coding region is 1686 bp long and consists of two exons (1387 bp and 248 bp) and a small (51 bp) intron (Oakeshott et al., 1987). The gene is duplicated but there is evidence that the adjacently located duplicate may be a pseudogene (Balakirev and Ayala, 1996; but see Dumancic et al., 1997). The EST-6 protein is transferred by *D. melanogaster* males to females in the semen fluid during copulation (Richmond et al., 1980; Richmond and Senior, 1981) and affects the female's consequent behavior and mating proclivity (Gromko et al., 1984; Scott, 1986).

Two main allozymes are known (Fast or F and Slow or S) that exhibit large-scale repeatable latitudinal clines (Oakeshott et al., 1981), with the Slow allozyme more common at higher latitudes. This and results from laboratory experiments suggests that the EST-6 polymorphism is maintained

by some form of selection (reviewed by Oakeshott et al., 1989, 1993, 1995; Richmond et al., 1990).

Cooke and Oakeshott (1989) suggested that the main Fast and Slow allozymes differ by two amino acids (Asn/Asp at position 237 and Thr/Ala at position 247) (but see Hasson and Eanes, 1996; Balakirev et al., 1999) and that these two amino acid polymorphisms are the most likely targets for selection underlying the latitudinal clines (Oakeshott et al., 1981).

Several independently acting *cis*-regulatory promoter elements that control the expression of the gene in different tissues have been identified within ~ 1.2 kb of the 5'-flanking region (Ludwig et al., 1993; Healy et al., 1996; Tamarina et al., 1997). Game and Oakeshott (1990) found that a polymorphism at an *Rsa* I site in the 5'-flanking region of *Est-6* shows significant association with the amount and activity of EST-6 in males. Given the evidence from other studies that differences in male EST-6 activity affect the reproductive success of their mates (Richmond et al., 1990), Game and Oakeshott (1990) have proposed that *Est-6 cis*-acting regulatory polymorphisms may be important contributors to adaptive variation (see also Oakeshott et al., 1994; and Saad et al., 1994).

Oggers et al. (1995) identified a nucleotide substitution responsible for the *Rsa* I polymorphism (T \rightarrow G at -531) and a peak of polymorphism around the *Rsa* I site. By comparing their data with the results of Game and Oakeshott (1990); Oggers et al. (1995) showed that the *Rsa* I + haplotype group yields $\sim 25\%$ more EST-6 enzyme activity in adult males than the *Rsa* I – one, and detected weak disequilibrium between the promoter polymorphism and the Fast/Slow allozyme polymorphism. However, Oggers et al. (1995) did not investigate the *Est-6* coding region in the same lines of *D. melanogaster* for which they obtained the promoter region sequences, which would have allowed them to analyse the pattern and extent of the association between the regulatory and structural nucleotide polymorphism.

We sequenced the *Est-6* gene in 15 lines of *D. melanogaster* and found departures from random polymorphism (Balakirev et al., 1999), but the size of the sample was too small for ascertaining the role of natural selection. More recently, we have sequenced a longer DNA fragment in a larger sample to test the hypothesis that the polymorphism is neutral. We have investigated the 5'-flanking, coding, and 3'-flanking regions of the *Est-6* gene (3062 bp total) in a random sample of 30 lines (and thus large enough for the population genetic tests; see Hudson et al., 1994; Simonsen et al., 1995) of *D. melanogaster* derived from a natural population of California. The detected pattern of variability is highly structured with distinctive features in the coding and 5'-flanking regions. We suggest that the *Est-6* nucleotide polymorphism is shaped by a combination of directional and balancing selection acting on the promoter and coding region polymorphisms, and by the interactions between the two regions due to different degrees of hitchhiking.

We have calculated various measures of nucleotide diversity for the entire data set and for different haplotype families separately (Table 3). In the pooled sample, total nucleotide diversity is very similar in the promoter and coding regions, but higher in the 3'-flanking region. The level of silent variation in the coding region is higher than in the promoter region, but similar to the variation in the 3'-flanking region, which could indicate different degrees of selective constraint in the 5'- and 3'-flanking regions. Polymorphism is lower in the S than in the F haplotypes (coding region and full sequence) and lower in the *RsaI* - than in the *RsaI* + (promoter region and full sequence). The difference is significant by coalescent simulations for the coding region S and F haplotypes ($P < 0.05$), but not for the promoter haplotypes. The level of divergence (K) between *D. melanogaster* and *D. simulans* is similar in different haplotype groups within the same functional region.

The *RsaI* restriction site difference is due to a T (*RsaI* +) \rightarrow G (*RsaI* -) transversion at position 653. The average number of nucleotide differences (K) between the *RsaI* + and *RsaI* - haplotypes is 6.720. The *RsaI* - haplotypes are fairly homogeneous ($K = 1.584$); the

RsaI + haplotypes are more heterogeneous ($K = 4.756$). The *RsaI* - haplotypes are most frequent in our data set (20 out of 30) and also in the data of Game and Oakeshott (1990) (20 out of 29). Odgers et al. (1995) suggested that *RsaI* + is the ancestral state, which is consistent with the higher polymorphism of the *RsaI* + haplotypes and is supported by comparison with *D. simulans*.

The average number of nucleotide differences between the two coding-region haplotypes is $K = 11.809$. The S group includes most haplotypes (21 out of 30), which are fairly homogeneous ($K = 3.810$); the nine F haplotypes are significantly more heterogeneous ($K = 16.722$). This suggests that the F lineage may be ancestral, a conclusion also reached by Cooke and Oakeshott (1989) and Hasson and Eanes (1996). However, *D. simulans* has an A at position 1985, the same as the S lineage, which would support the inference that the S lineage may have been the ancient condition from which the F allelic lineage derived. (This hypothesis was earlier favored by Balakirev et al., 1999 on that basis alone; they had in their sample only two F alleles, quite similar to one another.)

We have assessed the differences between haplotype

Table 3
Nucleotide diversity and divergence of *Est-6* and flanking regions in 30 strains of *D. melanogaster*^a

	Promoter region	Coding region			3'-flanking region	Full sequence	Silent sites
		Syn	Nsyn	Total			
No. of sites	1183	379	1256	1635	193	3062	1746
<i>Full sample</i>							
π	0.0059	0.0147	0.0025	0.0054	0.0133	0.0062	0.0088
θ	0.0058	0.0153	0.0026	0.0056	0.0183	0.0065	0.0093
K	0.0833	0.1499	0.0214	0.0492	0.0955	0.0643	0.0995
<i>Slow haplotypes</i>							
π	0.0057	0.0053	0.0008	0.0019	0.0014	0.0034	0.0053
θ	0.0054	0.0066	0.0013	0.0026	0.0029	0.0037	0.0054
K	0.0828	0.1504	0.0208	0.0489	0.0933	0.0639	0.0989
<i>Fast haplotypes</i>							
π	0.0063	0.0199	0.0030	0.0069	0.0253	0.0079	0.0115
θ	0.0056	0.0194	0.0227	0.0065	0.0248	0.0074	0.0107
K	0.0831	0.1487	0.0227	0.0499	0.1007	0.0650	0.0997
<i>RsaI</i> - haplotypes							
π	0.0012	0.0137	0.0026	0.0052	0.0142	0.0043	0.0055
θ	0.0025	0.0134	0.0022	0.0048	0.0117	0.0044	0.0060
K	0.0832	0.1499	0.0215	0.0494	0.0967	0.0645	0.0997
<i>RsaI</i> + haplotypes							
π	0.0046	0.0152	0.0023	0.0053	0.0104	0.0054	0.0077
θ	0.0044	0.0177	0.0025	0.0061	0.0183	0.0062	0.0089
K	0.0832	0.1499	0.0210	0.0490	0.0930	0.0638	0.0990

^a π is the average number of nucleotide differences per site among all pairs of sequences. θ is the expected number of segregating nucleotide sites among all sequences. K is the average proportion of nucleotide differences between *D. melanogaster* and *D. simulans*, corrected according to Jukes and Cantor (1969). Syn, synonymous; nsyn, nonsynonymous; silent sites include noncoding regions and synonymous sites in coding region. The coding region includes exons I and II. Statistics are calculated for the whole data set, as well as for the coding (Slow and Fast haplotypes) and promoter (*RsaI* - and *RsaI* +) regions separately. Polymorphic sites are not homogeneously distributed along the promoter region (Odgers et al., 1995); we use the average over the 1183 bp for comparisons with other regions.

families by the permutation test of Hudson et al. (1992). The tests are highly significant for the promoter haplotypes $K_{st}^* = 0.4175$ ($K_{st}^{*0.999} = 0.0966$, $P < 0.001$); as well as for the coding region haplotypes $K_{st}^* = 0.2846$ ($K_{st}^{*0.999} = 0.0701$, $P < 0.001$). Two F haplotypes (strains F-531F and F-611F) may have arisen by recombination between S and F coding region sequences (see Fig. 3); the difference between them and the other F haplotypes is significant ($K_{st}^* = 0.3352$, $K_{st}^{*0.95} = 0.1290$, $P < 0.05$).

Sliding-window analysis of the distribution of divergence within and between the sets of haplotypes manifests conspicuous peaks of variation, one around the *RsaI* – / *RsaI* + site (Fig. 4A) and another around the S/F site (Fig. 4B). These peaks may reflect the effect of balancing selection (Strobeck, 1983; Hudson and Kaplan, 1988; see our earlier discussion of *Sod* polymorphism).

We have examined linkage disequilibrium within the *Est-6* gene region as follows. We have first eliminated all singleton polymorphisms (i.e. mutations present in only one sequence). Then, we have made all pairwise comparisons between the remaining polymorphic sites, using Fisher's exact test for linkage disequilibrium. 36.3% (445 out of 1225) comparisons are statistically significant. With the

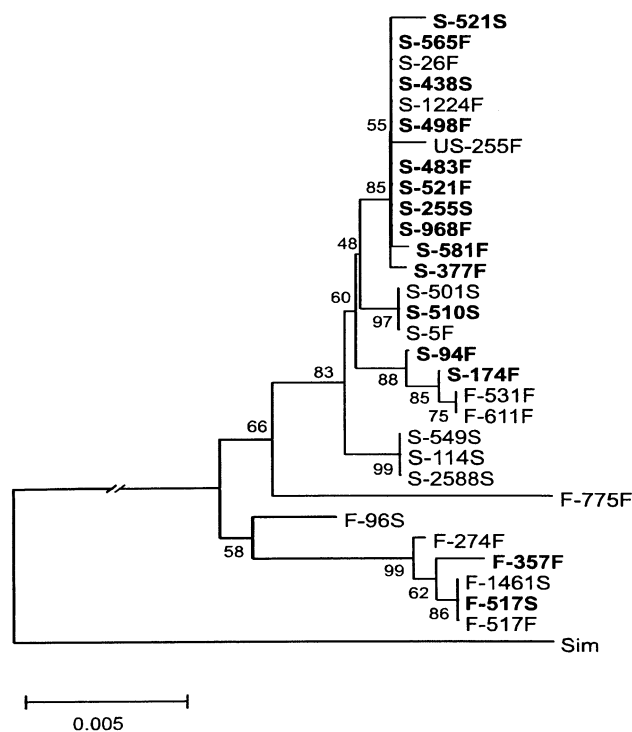


Fig. 3. Unrooted neighbor-joining tree (Kimura's 2-parameter model) of 30 *Est-6* haplotypes of *D. melanogaster*. S or F to the left of the number refers to the *Est-6* locus. (Notice US-255F, a strain showing an 'ultra-slow' *Est-6* electromorph.) The number and the S or F on the right refer to the *Sod* strain designation. The 15 haplotypes in bold are those used for investigating linkage disequilibrium between the two genes. Notice that all *Est-6* S haplotypes group as a sister clade to the F-775F haplotype. The sequences of F-531F and F-611F indicate that they are recombinant haplotypes. Numbers at the nodes are percent bootstrap values, based on 500 replications.

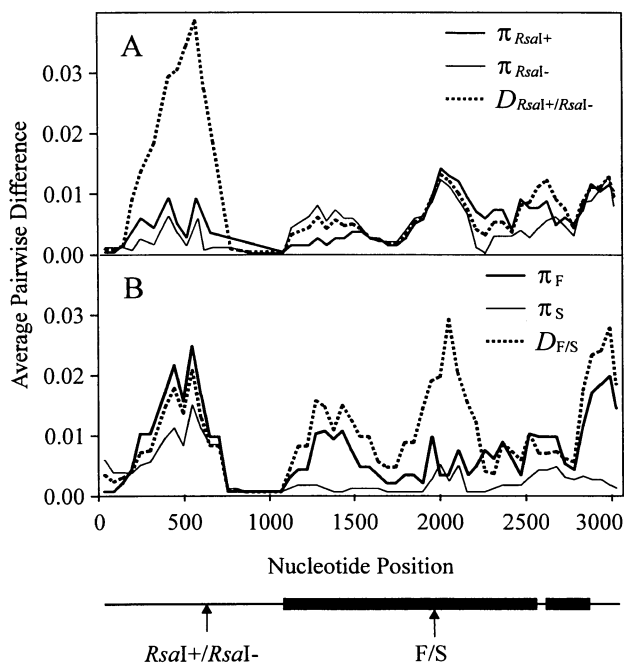


Fig. 4. Sliding-window analysis of the *Est-6* polymorphisms for two pairs of haplotype sets: *RsaI* + versus *RsaI* – (top) and *F* versus *S* (bottom). Nucleotide diversity within sets (π) is compared to nucleotide diversity between sets (D). Window size is 200 bp with 50 bp increments. The gene's structure is represented below, with the coding regions as black boxes; the arrows indicate the *RsaI* + /*RsaI* – and *S/F* sites. The peaks around these two sites suggest balanced polymorphisms. There is a suggestion of a balanced polymorphism (involving primarily the F haplotypes) in the 3'-flanking region; this is a short segment that separates *Est-6* from a duplicated gene (represented by various authors as *Est-P*, *Est-7*, or ψ *Est-6*; see Balakirev and Ayala, 1996), and thus corresponds to the putative promoter region of the duplicated gene.

Bonferroni correction for multiple comparisons, the significant associations reduce to 9.3%. The pattern of disequilibrium is quite distinct. Two clearly defined regions occur, one of which encompasses the promoter region; the other region comprises the rest of the gene. There is strong disequilibrium within each region, but virtually none between regions: of the pairwise comparisons between sites, 52.9 and 55.2% are significant within each region, respectively (33.3 and 11.5%, with the Bonferroni correction), but only 15.1% between the regions (1.0% with the Bonferroni correction). That is, with respect to linkage disequilibrium, the promoter and the coding segments behave as if they were each tightly linked, but evolving independently from one another.

The McDonald and Kreitman (1991) test of neutrality is statistically significant when applied to the 30 strains in our study (Table 4). However, the Hudson et al. (1987), Tajima (1989), Fu and Li (1993) and Depaulis and Veuille (1998) tests do not reveal any significant deviation from neutrality, although the Tajima (1989) test applied to our data (exon I) combined with previously published *Est-6* sequences (Cooke and Oakeshott, 1989; Hasson and Eanes, 1996) reveals significant deviation from neutrality expectations for

Table 4
McDonald-Kreitman test of neutrality for the *Est-6* gene

	Fixed differences	Polymorphisms		
		<i>D. melanogaster</i>	<i>D. simulans</i> ^a	Total ^b
Replacement	18	13	11	22
Silent	24	23	54	74

^a Calculated using data from Karotam et al. (1995).

^b Sites that are polymorphic in both species are counted only once. For the two-tailed Fisher's exact test $P = 0.025$.

the S alleles ($D = -1.864$, $P < 0.05$; not so for the F alleles, $D = -0.181$, $P > 0.10$); this test also detects significant deviation from neutrality in the promoter region for the *RsaI* – haplotypes ($D = -2.065$, $P < 0.05$). We have also used Kelly's Z_{nS} (Kelly, 1997) and Wall's B and Q tests (Wall, 1999), which are based on linkage disequilibrium between segregating sites. For the entire *Est-6* region both tests are highly significant ($Z_{nS} = 0.154$, $P = 0.004$; $B = 0.270$, $P = 0$; $Q = 0.427$, $P = 0$) with $C = 0.015$, equal to the C_{\min} based on the inferred minimum number of the recombination events (Hudson and Kaplan, 1985). The tests are also significant separately for the coding and promoter regions, with $C \geq 0.020$ (promoter region) and $C \geq 0.010$ (coding region) for Kelly's test and $C \geq 0.010$ for Wall's tests. The areas of significant values of Kelly's and Wall's statistics are centered around the *RsaI* site and the F/S polymorphisms.

The pattern of variability in the promoter region of the *Est-6* gene suggests involvement of both directional and balancing selection. We have observed a pronounced area of highly significant linkage disequilibrium around the *RsaI* site (Fig. 4). Thus, the *RsaI* site might be a target of selection in the promoter region. However this site is in linkage disequilibrium with 11 other polymorphic sites in the promoter region, which makes uncertain the precise site that impacts EST-6 activity. Conceivably, selection could act on any site that is in linkage disequilibrium with the *RsaI* site, or on the whole stretch of linked sites.

The present analysis of the *Est-6* coding region confirms our previous suggestion that the pattern of nucleotide variability of the *Est-6* coding region is shaped by the influence of both directional and balancing selection (Balakirev et al., 1999; see also Oakeshott et al., 1989, 1993, 1995; Richmond et al., 1990). Window analysis of DNA sequence variation reveals an excess of polymorphism surrounding the site that determines the F/S allozyme polymorphism (Fig. 4). This is consistent with a history of balancing selection impacting the allozyme polymorphism site. Our interpretation is that the pattern of nucleotide variation in *Est-6* is shaped by the superposition of the effects of directional and balancing selection in the promoter region; and by analogous superposition effects in the coding region.

4. Linkage disequilibrium between *Sod* and *Est-6*

Linkage disequilibrium and nonrandom associations between alleles or groups of nucleotides may indicate epistatic relationships, and much empirical work has been devoted over several decades to ascertain whether linkage disequilibrium occurs between gene loci. The issue of gene interaction is also important in connection with the long-lasting neutralist-selectionist controversy. Linkage disequilibrium is often considered strong evidence of selection, especially if its pattern is consistent between populations (Lewontin, 1974).

The evidence for linkage disequilibrium between individual loci in *Drosophila* remains scarce, except when genes are very closely linked or associated with chromosomal inversions. In the cases when significant associations have been detected, it is often far from clear whether they are due to non-random haplotype sampling, random genetic drift, or natural selection. Significant disequilibrium can indeed arise without epistasis as a result of random genetic drift within a given population in subdivided populations and by gene migration or founder effects (Balakirev et al., 1999).

Numerous examples of significant linkage disequilibrium have been discovered in *Drosophila* between specific allozymes and chromosomal inversions, which have been interpreted as reflecting selection for favored multilocus allele combinations. The general inference from these studies is, however, that linkage disequilibrium is mostly associated with closely linked genes, but may involve distantly linked genes when special cytological mechanisms (polymorphic inversions) allow it to exist. Gene loci that can recombine freely exhibit little, if any, linkage disequilibrium. Failure to detect disequilibrium may, of course, be a consequence of the limited statistical power of the tests to detect it (Brown, 1975). Nevertheless, DNA linkage disequilibrium between nucleotide sites in different loci is well established in some cases, and also that it reflects epistatic relationships (Kirby et al., 1995; Kirby and Stephan, 1996), although the nature of the epistatic interactions between genes remains enigmatic.

Sod and *Est-6* are closely linked on the left arm of chromosome 3 of *D. melanogaster*, ~938 kb apart. We have examined whether linkage disequilibrium occurs between the DNA sequences of these two genes. This possibility has been intimated by the results of Smit-McBride et al. (1988), who investigated natural and laboratory populations and detected linkage disequilibrium between the allozyme polymorphisms of *Sod* and *Est-6*, but not between other gene pairs. Moreover, as we have shown, a selective sweep has recently occurred involving a many-kilobases-long region that includes the *Sod* gene (see also Hudson et al., 1994, 1997). The pattern of the *Est-6* polymorphism may also have arisen as a consequence of an S haplotype sweep (see Fig. 3).

We have investigated jointly the *Est-6* and *Sod* sequence polymorphisms in a set of 15 strains from the El Rio

population of *D. melanogaster*. (These are the strains labeled in bold type in Fig. 3.) We include in this analysis a segment 1879 bp long of *Est-6* that encompasses the two exons and intron plus 193 bp of the 3'-flanking region. For *Sod* we analyse a segment 1408 bp long that encompasses the two exons and intron plus 244 bp of the 3'-flanking region.

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

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

We have evaluated linkage disequilibrium between the *Sod* and *Est-6* genes, first, using Fisher's exact test and the chi-square test, which fail to detect any significant interlocus association, as might be expected owing to asymmetrical allelic frequencies (Lewontin, 1995). We have also used the 'sign' method (Lewontin, 1995), based on the distribution of the disequilibrium sign, which is sensitive to asymmetrical allele frequencies and efficiently operates with singleton polymorphisms, which are not informative when Fisher's exact test is used. The sign method involves examining the number of positive and negative *D* values for each polymorphic site within and between all types of pairwise comparisons (singletons vs.

Table 5
Linkage associations between *Est-6* and *Sod* for pairs of biallelic sites^a

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

^a *k* is the number of copies of the rarer allele at a biallelic site; *m* is the number of copies of the rarer allele at another biallelic site for *m* ≥ *k*. *D* + and *D* - refer to positive and negative associations.

Table 6
Linkage disequilibrium between *Est-6* and *Sod* in *Drosophila melanogaster*^a

<i>Est-6</i>	<i>Sod</i>	<i>G</i>	<i>P</i>
Full sequence	Full sequence	459.6	<0.001
Coding	Intron	277.4	<0.001
Coding	3'-flanking	53.1	<0.001
Coding	Coding	6.1	<0.05
Exon I (F/S) site	Exon II (F/S) site	6.5	<0.05
Exon II	Exon II	1.4	n.s.

^a *G* is a likelihood ratio statistic (Sokal and Rohlf, 1981, pp. 695–707; see Lewontin, 1995). The values of *G* and *P* are not materially significant when using the Williams correction (*G**, see Sokal and Rohlf, 1981). n.s., not significant.

singletons, singletons vs. doublets, doublets vs. doublets, and so on). The results are summarized in Table 5.

In order to identify the gene regions that might be involved in the disequilibrium between the two genes, we have applied the Lewontin test separately to different regions. The results are summarized in Table 6. The most significant associations occur between the *Est-6* coding region and the *Sod* intron (row 2 in Table 6) and between the *Est-6* coding region and the 3'-flanking region of *Sod* (row 3 in Table 6). It is interesting that the disequilibrium between exon I of *Est-6* and exon II of *Sod* (row 5 in Table 6), which include the sites of the S/F polymorphisms, is less pronounced than between the regions just noted. This may likely be due, at least in part, to the selective sweep that has considerably reduced variability in *Sod*, and has, consequently, erased all or most preexisting linkage disequilibrium. The pattern of linkage disequilibrium between the two genes remains unchanged when singletons are excluded (Balakirev et al., 1999).

5. Concluding remarks

The analysis of DNA sequence variation in two loci of *D. melanogaster* suggest that natural selection is involved in modulating the patterns of the polymorphisms. In the case of *Sod* there is evidence of positive directional selection and perhaps some form of balanced selection; directional selection as well as balance polymorphism appear to be involved in *Est-6*.

A hypothesis of the evolutionary history of *Sod* would be as follows. The F(A) haplotype arose ≥ 5000 years ago and increased in frequency under strong natural selection (*s* > 0.01) reaching a frequency of about 50% in Europe, Asia, North America and South America (Hudson et al., 1994 and our unpublished data). We have not found the F(A) allele in African populations (data not shown), which suggests that the mutation responsible for the selective sweep arose after *D. melanogaster* had colonized other continents from its African origins. The S mutation arose in an F(A) haplotype and may have been swept together with

other F(A) haplotypes. The S allele does not occur in Africa or in South America, consistent with its recent origin in Europe, Asia, or North America. The locus responsible for the selective sweep of the F(A) haplotypes may not be the *Sod* locus itself, but rather it is probably a site several kb downstream from that locus. Whether natural selection is also impacting the *Sod* locus itself is uncertain.

There is evidence that a recent selective sweep may also have occurred encompassing an S haplotype of the *Est-6* locus. There is very little sequence divergence among the *Est-6* S alleles; in fact a majority of these are identical in sequence or differ by only one or two nucleotides from most others (see Fig. 3). The site of the selective sweep may be within the *Est-6* locus itself (or very closely linked to it) given that all S alleles seemed to have been carried by the sweep, without enough time for recombination with non-sweep haplotypes.

The occurrence of similar S/F *Est-6* latitudinal clines in different continents (North America and Australia) suggests that balanced selection is also impacting the *Est-6* polymorphism (Oakeshott et al., 1981, 1993 and references therein; Healy et al., 1996). The role of *Est-6* in *D. melanogaster* mating propensity (Richmond et al., 1980, 1990; Ludwig et al., 1993); its association with increased reproductive fitness, pre-adult viability, tissue expression pattern, and levels of enzyme activity strongly indicate that natural selection is involved in shaping the S/F and *RsaI* – / *RsaI* + polymorphisms (Gromko et al., 1984; Game and Oakeshott, 1990; Richmond et al., 1990; Ludwig et al., 1993; Oakeshott et al., 1994, 1995; Odgers et al., 1995; Tamarina et al., 1997). Our window analysis also supports that these are balanced polymorphisms (Fig. 4).

The selective sweeps involving *Est-6* S haplotypes and *Sod* F(A) and S haplotypes may have been different. This is supported by the observation that haplotypes S-498F and S-968F have identical *Est-6* sequences (see Fig. 3) and are not *Sod* F(A) haplotypes, but rather are very different in *Sod* sequence from all F(A) alleles (see Fig. 3 in Balakirev et al., 1999). However, this discrepancy can be accounted by genetic recombination between the two loci, even if the selective sweep were very recent (~2600 years old), given that *Est-6* and *Sod* are nearly 1000 kb apart and in a region of high-recombination frequency (see Fig. 1). A single selective sweep would account for the linkage disequilibrium we have observed between the two loci. Alternatively, this disequilibrium might have arisen as a consequence of epistatic interactions between the two regions. The functional grounds of such epistatic interactions, if they exist, are unknown at present, since the sweep site in the *Sod* region remains to be genetically characterized.

The pattern of two highly diverged sets of haplotypes that occurs in both loci (the F versus the S haplotypes in *Est-6* and the F(A) and S versus the other F haplotypes in *Sod*) might be accounted for by merging of previously geographically isolated populations. Another possibility is that one set of diverged haplotypes is or was part of an inversion.

The common and widespread *In(3L)P* inversion embraces *Est-6* and *Sod*, but no third-chromosome inversions have been found segregating in the El Rio population (Smit-McBride et al., 1988, and unpublished data from our laboratory). It is not known whether the one set of diverged haplotypes may represent sequences that have ‘escaped’ from an inversion as it has been suggested for sequences at the *Est-6* locus (Odgers et al., 1995). In conclusion, the intriguing similarities in patterns of DNA sequence polymorphism may have arisen by similar independent selection events, but we cannot completely exclude the possibility of some event or process affecting simultaneously *Est-6* and *Sod* within a relatively large segment of chromosome 3 (Hudson et al., 1997).

References

- Ayala, F.J., 1997. Vagaries of the molecular clock. Proc. Natl. Acad. Sci. USA 94, 7776–7783.
- Ayala, F.J., 1999. Molecular clock mirages. BioEssays 21, 71–75.
- Balakirev, E.S., Ayala, F.J., 1996. Is esterase-P encoded by a cryptic pseudogene in *Drosophila melanogaster*? Genetics 144, 1511–1518.
- Balakirev, E.S., Balakirev, E.I., Rodríguez-Trelles, F., Ayala, F.J., 1999. Molecular evolution of two linked genes, *Est-6* and *Sod*, in *Drosophila melanogaster*. Genetics 153, 1357–1369.
- Brown, A.D.H., 1975. Sample sizes required to detect linkage disequilibrium between two or three loci. Theor. Popul. Biol. 8, 184–201.
- Cooke, P.H., Oakeshott, J.G., 1989. Amino acid polymorphisms for esterase-6 in *Drosophila melanogaster*. Proc. Natl. Acad. Sci. USA 86, 1426–1430.
- Comeron, J.M., Kreitman, M., Aguadé, M., 1999. Natural selection on synonymous sites is correlated with gene length and recombination in *Drosophila*. Genetics 151, 239–249.
- Depaulis, F., Veuille, M., 1998. Neutrality tests based on the distribution of haplotypes under an infinite-site model. Mol. Biol. Evol. 15, 1788–1790.
- Dumancic, M.M., Oakeshott, J.G., Russell, R.J., Healy, M.J., 1997. Characterization of the EstP protein in *Drosophila melanogaster* and its conservation in drosophilids. Biochem. Genet. 35, 251–271.
- Fridovich, I., 1986. Superoxide dismutases. Adv. Enzymol. 58, 62–97.
- Fu, Y.-X., Li, W.-H., 1993. Statistical tests of neutrality of mutations. Genetics 133, 693–709.
- Game, A.Y., Oakeshott, J.G., 1990. Associations between restriction site polymorphism and enzyme activity variation for esterase 6 in *Drosophila melanogaster*. Genetics 126, 1021–1031.
- Gillespie, J.H., 1989. Lineage effects and the index of dispersion of molecular evolution. Mol. Biol. Evol. 6, 636–647.
- Gillespie, J.H., 1991. The Causes of Molecular Evolution, Oxford University Press, New York.
- Graf, J.D., Ayala, F.J., 1986. Genetic variation for superoxide dismutase level in *Drosophila melanogaster*. Biochem. Genet. 24, 153–168.
- Gromko, M.H., Gilbert, D.F., Richmond, R.C., 1984. Sperm transfer and use in the multiple mating system of *Drosophila*. In: Smith, R.L., (Ed.), Sperm Competition and the Evolution of Animal Mating Systems, Academic Press, New York, pp. 371–426.
- Hasson, E., Eanes, W.F., 1996. Contrasting histories of three gene regions associated with *In(3L)Payne* of *Drosophila melanogaster*. Genetics 144, 1565–1575.
- Healy, M.J., Dumancic, M.M., Cao, A., 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., 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., Kaplan, N., 1988. The coalescent process in models with selection and recombination. *Genetics* 120, 831–840.
- Hudson, R.R., Kreitman, M., Aguadé, M., 1987. A test of neutral molecular evolution based on nucleotide data. *Genetics* 116, 153–159.
- Hudson, R.R., Boos, D., Kaplan, N.L., 1992. A statistical test for detecting geographic subdivision. *Mol. Biol. Evol.* 9, 138–151.
- Hudson, R.R., Bailey, K., Skarecky, D., Kwiatowski, J., Ayala, F.J., 1994. Evidence for positive selection in the superoxide dismutase (*Sod*) region of *Drosophila melanogaster*. *Genetics* 136, 1329–1340.
- Hudson, R.R., Sáez, A.G., 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.
- Jukes, T.H., Cantor, C.R., 1969. Evolution of protein molecules. In: Munro, H.M., (Ed.), *Mammalian Protein Metabolism*, Academic Press, New York, pp. 21–120.
- Kaplan, N.L., Hudson, R.R., Langley, C.H., 1989. The ‘hitchhiking effect’ revisited. *Genetics* 123, 887–899.
- Karotam, J., Boyce, T.M., 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.
- Kimura, M., 1968. Evolutionary rate at the molecular level. *Nature* 217, 624–626.
- Kimura, M., 1983. *The Neutral Theory of Molecular Evolution*, Cambridge University Press, Cambridge.
- Kimura, M., Ohta, T., 1971. Protein polymorphism as a phase of molecular evolution. *Nature* 229, 467–469.
- King, J.L., Jukes, T.H., 1969. Non-Darwinian evolution. *Science* 164, 788–798.
- Kirby, D.A., Stephan, W., 1996. Multi-locus selection and the structure of variation at the *white* gene of *Drosophila melanogaster*. *Genetics* 144, 635–645.
- Kirby, D.A., Muse, S.V., Stephan, W., 1995. Maintenance of pre-mRNA secondary structure by epistatic selection. *Proc. Natl. Acad. Sci. USA* 92, 9047–9051.
- Lee, Y.M., Ayala, F.J., 1985. Superoxide dismutase in *Drosophila melanogaster*: Mutation site difference between two electromorphs. *FEBS Lett.* 179, 115–119.
- Lee, Y.M., Misra, H.P., Ayala, F.J., 1981. Superoxide dismutase in *Drosophila melanogaster*: Biochemical and structural characteristics of allozyme variants. *Proc. Natl. Acad. Sci. USA* 78, 7052–7055.
- Lewontin, R.C., 1974. *The Genetic Basis of Evolutionary Change*, Columbia University Press, New York.
- Lewontin, R.C., 1995. The detection of linkage disequilibrium in molecular sequence data. *Genetics* 140, 377–388.
- Ludwig, M.Z., Tamarina, N.A., 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, 6233–6273.
- McDonald, J.H., Kreitman, M., 1991. Adaptive protein evolution at the *Adh* locus in *Drosophila*. *Nature* 351, 652–654.
- Oakeshott, J.G., Chambers, G.K., Gibson, J.B., Willcocks, D.A., 1981. Latitudinal relationships of esterase-6 and phosphoglucomutase gene frequencies in *Drosophila melanogaster*. *Heredity* 47, 385–396.
- Oakeshott, J.G., Collet, C., Phillis, R., Nielsen, K.M., Russell, R.J., Chambers, G.K., Ross, V., 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., Cooke, P.H., Richmond, R.C., Bortoli, A., Game, A.Y., Labate, J., 1989. Molecular population genetics of structural variance of esterase-6 in *Drosophila melanogaster*. *Genome* 31, 788–796.
- Oakeshott, J.G., van Papenrecht, E.A., Boyce, T.M., Healy, M.J., Russell, R.J., 1993. Evolutionary genetics of *Drosophila* esterases. *Genetica* 90, 239–268.
- Oakeshott, J.G., Saad, M., Game, A.Y., Healy, M.J., 1994. Causes and consequences of esterase 6 enzyme activity variation in pre-adult *Drosophila melanogaster*. *Heredity* 73, 160–169.
- Oakeshott, J.G., Boyce, T.M., Russell, R.J., Healy, M.J., 1995. Molecular insights into the evolution of an enzyme; esterase 6 in *Drosophila*. *Trends Ecol. Evol.* 10, 103–110.
- Ogders, W.A., Healy, M.J., 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.
- Peng, T., Moya, A., Ayala, F.J., 1986. Irradiation-resistance conferred by superoxide dismutase: Possible adaptive role of a natural polymorphism in *Drosophila melanogaster*. *Proc. Natl. Acad. Sci. USA* 83, 684–687.
- Peng, T.-X., Moya, A., Ayala, F.J., 1991. Two modes of balancing selection in *D. melanogaster*: Overcompensation and overdominance. *Genetics* 128, 381–391.
- Richmond, R.C., Gilbert, D.G., Sheehan, K.B., Gromko, M.H., Butterworth, F.M., 1980. Esterase 6 and reproduction in *Drosophila melanogaster*. *Science* 207, 1483–1485.
- Richmond, R.C., Senior, A., 1981. Esterase 6 of *Drosophila melanogaster*: kinetics of transfer to female, decay in females and male recovery. *J. Insect Physiol.* 27, 849–853.
- Richmond, R.C., Nielsen, K.M., Brady, J.P., Snella, E.M., 1990. Physiology, biochemistry and molecular biology of the *Est-6* locus in *Drosophila melanogaster*. In: Barker, J.S.F., Starmer, W.T., MacIntyre, R.J. (Eds.), *Ecological and Evolutionary Genetics of Drosophila*, Plenum, New York, pp. 273–292.
- Rodriguez-Trelles, F., Tarrio, R., Ayala, F.J., 2001. Erratic overdispersion of three molecular clocks: GPDH, SOD, and XDH. *Proc. Natl. Acad. Sci. USA* 98, 11405–11410.
- Rowan, R.G., Hunt, J.A., 1991. Rates of DNA change and phylogeny from the DNA-sequences of the alcohol-dehydrogenase gene for 5 closely related species of Hawaiian *Drosophila*. *Mol. Biol. Evol.* 8, 49–70.
- Saad, M., Game, A.Y., Healy, M.J., Oakeshott, J.G., 1994. Associations of esterase 6 allozyme and activity variation with reproductive fitness in *Drosophila melanogaster*. *Genetica* 94, 43–56.
- Scott, D., 1986. Inhibition of female *Drosophila melanogaster* remating by a seminal fluid protein (esterase 6). *Evolution* 40, 1084–1091.
- Sharp, P.M., Li, W.-H., 1989. On the rate of DNA-sequence evolution in *Drosophila*. *J. Mol. Evol.* 28, 398–402.
- Simonsen, K.L., Churchill, G.A., Aquadro, C.J., 1995. Properties of statistical tests of neutrality for DNA polymorphism data. *Genetics* 141, 413–429.
- Singh, R.S., Hickey, D.A., David, J., 1982. Genetic differentiation between geographically distant populations of *Drosophila melanogaster*. *Genetics* 101, 235–256.
- Smit-McBride, Z., Moya, A., Ayala, F.J., 1988. Linkage disequilibrium in natural and experimental populations of *Drosophila melanogaster*. *Genetics* 120, 1043–1051.
- Sokal, R.R., Rohlf, F.J., 1981. *Biometry. The Principles and Practice of Statistics in Biological Research*, 2nd Edition, Freeman, San Francisco.
- Steinman, H.M., 1988. Bacterial superoxide dismutases. *Basic Life Sci.* 49, 641–646.
- Strobeck, C., 1983. Expected linkage disequilibrium for a neutral locus linked to a chromosomal arrangement. *Genetics* 103, 545–555.
- Tajima, F., 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* 123, 585–595.
- Tamarina, N.A., Ludwig, M.Z., Richmond, R.C., 1997. Divergent and conserved features in the spatial expression of the *Drosophila pseudoobscura esterase-5B* gene and the *esterase-6* gene of *Drosophila melanogaster*. *Proc. Natl. Acad. Sci. USA* 94, 7735–7741.
- Tyler, R.H., Brar, H., Singh, M., Latorre, A., Graves, J.L., Mueller, L.D., Rose, M.R., Ayala, F.J., 1993. The effect of superoxide dismutase alleles on aging in *Drosophila*. *Genetica* 91, 143–149.
- Wall, J.D., 1999. Recombination and the power of statistical tests of neutrality. *Genet. Res.* 74, 65–79.