Nucleotide Variation at the Hypervariable Esterase 6 Isozyme Locus of Drosophila simulans

Jill Karotam, *† Thomas M. Boyce, * and John G. Oakeshott*

*CSIRO Division of Entomology, and †Division of Biochemistry and Molecular Biology, Australian National University

Esterase 6 (Est-6/EST6) is polymorphic in both Drosophila melanogaster and D. simulans for two common allozyme forms, as well as for several other less common variants. Parallel latitudinal clines in the frequencies of the common EST6-F and EST6-S allozymes in these species have previously been interpreted in terms of a shared amino acid polymorphism that distinguishes the two variants and is subject to selection. Here we compare the sequences of four D. simulans Est-6 isolates and show that overall estimates of nucleotide heterozygosity in both coding and 5' flanking regions are more than threefold higher than those obtained previously for this gene in D. melanogaster. Nevertheless, the ratio of replacement to exon silent-site polymorphism in D. simulans is less than the ratio of replacement to silent divergence between D. simulans and D. melanogaster, which could be the result of increased efficiency of selection against replacement polymorphisms in D. simulans or to divergent selection between the two species. We also find that the amino acid polymorphisms separating EST6-F and EST6-S in D. simulans are not the same as those that separate these allozymes in D. melanogaster, implying that the shared clines do not reflect shared molecular targets for selection. All comparisons within and between the two species reveal a remarkable paucity of variation in a stretch of nearly 400 bp immediately 5' of the gene, indicative of strong selective constraint to retain essential aspects of Est-6 promoter function.

Introduction

Nucleotide sequence analyses of several Drosophila genes have recently revealed some striking examples of how natural selection may affect the extent and distribution of nucleotide variation. Comparisons of amino acid replacement and silent-site polymorphisms within species have revealed clear examples of balancing selection (e.g., Adh in Drosophila melanogaster; Kreitman and Hudson 1991), purifying selection or “selective sweep” (e.g., Adh in D. pseudoobscura and Sod and tra in D. melanogaster: Schaeffer and Miller 1992; Hudson et al. 1994; Walthour and Schaeffer 1994), and agreement with neutral expectations (e.g., Xdh in D. pseudoobscura; Riley et al. 1992). Comparisons of the ratios of replacement to silent-site differences within and across species have also suggested two cases of adaptive amino acid divergence between sibling species (Adh and G6pd in D. melanogaster and D. simulans; McDonald and Kreitman 1991; Eanes et al. 1993).

In this article we extend our molecular analyses of variation for the esterase 6 enzyme in D. simulans and D. melanogaster. EST6 is highly polymorphic at the protein level in both species and contains not only a major “classical” polymorphism involving two common allozymes (EST6-F and EST6-S), but also many less frequent but nevertheless polymorphic variants (Albuquerque and Napp 1981; Cooke et al. 1987; Labate et al. 1989). Under standard starch gel techniques, the relative mobilities of the common EST6-F allozymes of each species are identical, as are those of their common EST6-S allozymes (see, e.g., Hyyata et al. 1985). The two species show parallel latitudinal clines in the frequencies of these shared major allozyme variants (Anderson and Oakeshott 1984), from which it has been inferred that selection may be targeted at shared amino acid polymorphism(s) responsible for the EST6-F/EST6-S electrophoretic difference (see Oakeshott et al. 1989, 1990 for reviews). No information is available on whether the many minor variants are subject to selection in the field although, as with EST6-F and EST6-S, various experimental studies of D. melanogaster suggest that selection does occur in laboratory populations (Oakeshott et al. 1989, 1990 and references therein).

To determine the amino acid polymorphisms separating the various EST6 allozymes, and in particular...
EST6-F and EST6-S, Cooke and Oakeshott (1989) sequenced Est-6 genes from 13 isozymic lines, covering 10 electrophoretic variants of D. melanogaster. Of a total of 16 polymorphisms, two only 10 residues apart were found to be highly associated with the EST6-F/EST6-S difference (Cooke and Oakeshott 1989). One of these, an Asp/Asn polymorphism at residue 237 involves a charge change and is presumably the cause of the electrophoretic mobility difference. However, given their close linkage, either this or the second polymorphism, an Ala/Thr difference at residue 247, could have been the target for the clinal selection (Cooke and Oakeshott 1989). Karotam et al. (1993) subsequently published the sequence of a single EST6-S allele of D. simulans and showed that it shares Asn-237 but not Thr-247 with the D. melanogaster EST6-S. This result is consistent with the hypothesis that the common Asp/Asn polymorphism at 237 is the target for shared clinal selection on EST6-F and EST6-S in these species. However, the sequence of a D. simulans EST6-F form has not been available to confirm this hypothesis.

We have sequenced three more D. simulans Est-6 alleles for comparison with the published D. simulans sequence and have tested the distribution of nucleotide polymorphisms among the four against their divergence from the D. melanogaster sequences cited above for any departures from neutrality. We have also used the high-resolution electrophoretic procedures of Cooke et al. (1987) to determine the electrophoretic mobilities of the EST6 allozymes encoded by the four sequenced D. simulans alleles and compared their mobilities with those of the various EST6 allozymes in D. melanogaster. By these means we are able to determine whether there is a common molecular basis for the parallel latitudinal clines for the shared EST6-F and EST6-S variants and test how selection on the loci may have structured the underlying nucleotide sequence variation.

Material and Methods
DNA Cloning and Sequencing

Three isofemale lines (sim-1, sim-2, and sim-3) from a natural population of Drosophila simulans at Coff’s Harbour, Australia, were made homozygous for EST6 by repeated sib matings (Karotam and Oakeshott 1993). Libraries of EcoRI digests of genomic DNA were prepared from each line using the vector λgt10 (Promega). The libraries were screened at high stringency with the genomic clone of Est-6 from D. melanogaster (Collet et al. 1990) using the hybridization conditions of Reed and Mann (1985). DNA was prepared from hybridizing clones by the methods of Cooke and Oakeshott (1989) and digested with restriction enzymes under conditions specified by the manufacturers. Two different classes of clones from each line hybridized to the probe, one containing a 1.6-kb and the other a 6.4-kb EcoRI fragment, because of the presence of an EcoRI site just 5’ of the start of Est-6 translation (fig. 1). Both of these fragments were then subcloned from each of the three libraries into the plasmid vectors pTZ18U and pTZ19U (BioRad) using standard procedures (Sambrook et al. 1989, pp. F.6–F.7). Single-stranded plasmid DNA was prepared from these subclones for sequencing by the method of Vieira and Messing (1987). Both strands of a 2.95-kb region including the Est-6 gene (1.68 kb, including a single 50–51-bp intron), plus 1.1 kb of 5’ and 0.2 kb of 3’ flanking regions were sequenced using the dideoxy chain termination method of Sanger et al. (1977) and various oligonucleotide primers homologous to appropriately spaced sequences along the region. The next 0.46 bp of 5’ DNA was also subcloned and sequenced from the sim-1 and sim-3 1.6-kb genomic clones, as well as from homologous regions of two previously characterized clones (sim-B from an American stock of D. simulans, and Dm145, from an American stock of D. melanogaster: Collet et al. 1990; Karotam et al. 1993).

To confirm the continuity of the 1.6-kb 5’ and 6.4-kb coding fragments of the sim-1, sim-2, and sim-3 genes characterized above, complementary PCR products spanning the common EcoRI site were generated from genomic DNA of each stock using the methods of Saiki et al. (1988) and two additional oligonucleotide primers starting at −1,288 bp and +368 bp in the sim-1 sequence (fig. 1). Direct sequencing of these PCR products (Higuchi and Ochman 1989) indeed confirmed a single EcoRI site and the continuity of the DNA sequence from the 1.6-kb and 6.4-kb cloned fragments. Nucleotide sequences of sim-1, sim-2, and sim-3 have been submitted to GenBank under accession numbers L34263, L34264, and L34265, respectively. Sequences were analyzed using version 6.0 of the University of Wisconsin Sequence Analysis Package (Devereux et al. 1984).

Protein Electrophoresis

The D. simulans stocks analyzed by high-resolution electrophoresis were the four stocks from which Est-6 alleles have been cloned and sequenced and a further set of 31 isofemale lines isolated from an American population (Raleigh, N.C.) by C. F. Aquadro. Native polyacrylamide gel electrophoresis was carried out on homogenates of whole virgin adult males by the methods of Healy et al. (1991). EST6 activity was visualized by the method of Wright (1963) using fast garnet GBC salt and β-naphthyl acetate as a substrate. Staining was carried out in the presence of 10−4 M eserine and 5 × 10−4 M p-chloromercuribenzoate, which inhibit the activities of most other Drosophila esterases (Healy et al. 1991). Repeated pairwise comparisons (method of Cooke et al.
1987) among the *D. simulans* stocks and representatives of the *D. melanogaster* stocks characterized by Cooke et al. (1987) and Labate et al. (1989) were used to distinguish between allozymes showing very small mobility differences.

**Results**

Nucleotide Polymorphism in *Drosophila simulans*

There are 124 nucleotide site polymorphisms and six small (≤6 bp) insertion/deletion differences in the 2.95-kb region sequenced in all four *D. simulans* isolates, and a further 19 site polymorphisms and one complex insertion/deletion in the adjacent 0.46 kb upstream, which was only sequenced in sim-1, sim-3, and sim-B (fig. 1). The sim-B sequence from the United States is no more different from the three Australian lines than they are to each other. There are 22 differences unique to the sim-B sequence in the 2.95 kb sequenced in all four isolates, compared with 13, 35, and 23 unique differences in the sim-1, sim-2, and sim-3 sequences, respectively. It is intriguing that sim-B and sim-1 are identical at all 53 polymorphic sites in a block of about 800 bp at the 5’ end of the sequenced region, although they are no more similar to one another than to the other sequences in the remainder of the region. Such blocks of sequence similarity between alleles of otherwise divergent sequence have been noted in some previous sequence surveys and are generally attributed to the proliferation of particular recombinants (e.g., Cooke and Oakeshott 1989).

Levels of nucleotide polymorphism are heterogeneous across the sequenced region (table 1). Within the transcribed region estimates of heterozygosity (θ; eq. [1] Kreitman and Hudson 1991) are highest in exon silent sites and the intron; θ is about fourfold lower in the 3’ untranslated region and about 17-fold lower in exon replacement sites, while the 41 bp of 5’ untranslated sequence are invariable. Following Karotam et al. (1993) we have divided the 1.1 kb of 5’ untranscribed DNA sequenced in all four *D. simulans* lines into proximal (−42 to −391-bp 5’ of the start of translation), central (−392 to −758 bp), and distal (−759 to −1,138 bp)
Table 1
Estimates of Nucleotide Polymorphism among the Four Est-6 Sequences from Drosophila simulans

| Region                  | Number of Sites Compared | Number of Polymorphic Sites | Heterozygosity (θ) | Expected Number of Polymorphic Sites | Χ²
|-------------------------|--------------------------|----------------------------|-------------------|-------------------------------------|-----
| 5' far distal (n = 3)*  | 459                      | 19                         | 2.76              |                                     |     |
| 5' untranscribed        | 1,096                    | 49                         | 2.44              | 46.1                                | 0.01|
| Distal                  | 379                      | 35                         | 5.04              | 15.9                                | 3.05|
| Central                 | 367                      | 11                         | 1.63              | 15.4                                | 0.18|
| Proximal                | 350                      | 3                          | 0.47              | 14.7                                | 1.34|
| Transcribed             | 1,853                    | 75                         | 2.21              | 77.9                                | 0.00|
| 5' untranslated         | 41                       | 0                          | 0                 | 1.7                                 | 1.02|
| Exon silent             | 358                      | 51.5                       | 7.85              | 15.1                                | 12.4|
| Exon replacement        | 1,268                    | 11.5                       | 0.49              | 53.3                                | 1.45|
| Intron                  | 46                       | 6                          | 7.11              | 1.9                                 | 4.79|
| 3' untranslated         | 140                      | 6                          | 2.34              | 5.9                                 | 0.00|
| Overall                 | 2,949                    | 124                        | 2.29              | 124                                  | Σ Χ² - 24.3** |

NOTE.—Heterozygosity (θ) is calculated using eq. (1) of Kreitman and Hudson (1991). Expected number of segregating sites and Χ² are calculated using eq. (4) of Kreitman and Hudson (1991). The Χ² statistic tests for heterogeneity in levels of polymorphism across regions by testing the goodness of fit of the observed to expected number of polymorphic sites for each region and summing those statistics (Kreitman and Hudson 1991).

In all cases n = 4, except for the 5' far distal region, where only three lines were sequenced. Expected number of polymorphic sites and Χ² values were not calculated for this region across the three sequences.

** P < 0.01.

segments (fig. 1). The fourth segment, termed the far distal segment, is the next 459-bp 5' that was only sequenced in three of the lines. The level of polymorphism is highest in the distal and far distal segments (albeit still less than the exon silent-site level above), lower in the central segment, and lower still in the proximal segment. The level of polymorphism in the proximal segment is similar to that in the exon replacement sites above. The proximal segment also abuts the invariable 41-bp 5' untranslated element of the transcribed region, forming a contiguous stretch of nearly 400 bp of highly conserved sequence.

Interspecific Comparisons

The level of exon silent-site polymorphism (θ, eq. [1–4] in Nei and Jin 1989) across the four Est-6 alleles in D. simulans (8.42±1.13 per 100 nucleotide sites) is higher than that previously obtained from 13 fully sequenced isolates of Est-6 from D. melanogaster (2.39±0.43; Cooke and Oakeshott 1989 and references therein). The same trend is evident, albeit not as pronounced, in the levels of replacement-site polymorphism between the samples from the two species (0.56±0.15 vs. 0.32±0.08). The smaller difference between the species in replacement polymorphism could reflect a bias in the sample of D. melanogaster lines sequenced. These lines were chosen to maximize the number of different allozymes represented (10), so we expect an inflated estimate of replacement-site polymorphism in that sample.

Application of the McDonald and Kreitman (1991) test to the four D. simulans and the 13 D. melanogaster coding-region sequences (table 2) shows that the ratio of replacement to exon silent-site polymorphism within the two species (0.25) is lower than the ratio of fixed replacement to silent-site differences between them (0.44; G = 4.84, df = 1, P < 0.05). This difference occurs

Table 2
Comparisons of the Proportions of Est-6 Replacement and Silent Sites Fixed to Those That Are Polymorphic in Drosophila simulans and D. melanogaster

<table>
<thead>
<tr>
<th>POLYMORPHISMS</th>
<th>FIXED DIFFERENCES</th>
<th>sim and mel Combined</th>
<th>sim Only</th>
<th>mel Only</th>
</tr>
</thead>
<tbody>
<tr>
<td>Replacement</td>
<td>16</td>
<td>27</td>
<td>11</td>
<td>16</td>
</tr>
<tr>
<td>Silent</td>
<td>20</td>
<td>83</td>
<td>54</td>
<td>29</td>
</tr>
<tr>
<td>G test</td>
<td>4.84*</td>
<td>8.52**</td>
<td>0.65</td>
<td></td>
</tr>
</tbody>
</table>

NOTE.—All values were derived following the counting conventions of McDonald and Kreitman (1991) and using all four D. simulans (sim) and all 13 D. melanogaster (mel) sequences (Cooke and Oakeshott 1989 and references therein). G tests use the Williams correction for continuity (Sokal and Rohlf 1981, p. 738).

* P < 0.05.

** P < 0.01.
Table 3
Estimates of Est-6 Nucleotide Diversity and Divergence Within and Among Drosophila simulans (sim), D. melanogaster (mel), and D. mauritiana (mau)

<table>
<thead>
<tr>
<th>Region</th>
<th>sim ((\hat{\pi}))</th>
<th>mel ((\hat{\pi}))</th>
<th>sim:mel ((d_{xy}))</th>
<th>sim:mel ((d_x))</th>
<th>sim:mau ((d_{xy}))</th>
</tr>
</thead>
<tbody>
<tr>
<td>5' far distal</td>
<td>2.81 ± 0.59</td>
<td>...</td>
<td>11.90 ± 1.76</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>5' untranscribed</td>
<td>2.39 ± 0.36</td>
<td>...</td>
<td>7.11 ± 0.78</td>
<td>...</td>
<td>3.78 ± 0.51</td>
</tr>
<tr>
<td>Distal</td>
<td>5.54 ± 0.95</td>
<td>...</td>
<td>9.12 ± 1.45</td>
<td>...</td>
<td>9.37 ± 1.45</td>
</tr>
<tr>
<td>Central</td>
<td>1.70 ± 0.48</td>
<td>...</td>
<td>10.14 ± 1.70</td>
<td>...</td>
<td>1.72 ± 0.57</td>
</tr>
<tr>
<td>Proximal</td>
<td>0.47 ± 0.28</td>
<td>...</td>
<td>2.32 ± 0.79</td>
<td>...</td>
<td>0.28 ± 0.18</td>
</tr>
<tr>
<td>Coding sequences</td>
<td>2.59 ± 0.99</td>
<td>0.81 ± 0.36</td>
<td>7.56 ± 2.83</td>
<td>5.86 ± 2.76</td>
<td>2.37 ± 1.02</td>
</tr>
<tr>
<td>Exon silent</td>
<td>8.42 ± 1.13</td>
<td>2.39 ± 0.43</td>
<td>15.52 ± 1.88</td>
<td>10.11 ± 1.74</td>
<td>6.42 ± 0.97</td>
</tr>
<tr>
<td>Replacement</td>
<td>0.56 ± 0.15</td>
<td>0.32 ± 0.08</td>
<td>2.16 ± 0.38</td>
<td>1.73 ± 0.36</td>
<td>0.87 ± 0.23</td>
</tr>
</tbody>
</table>

**Note.**—Nucleotide diversity is calculated as \(\hat{\pi} \pm SE\) (eqq. [1–4] in Nei and Jin 1989) and divergence as \(d_x\) or \(d_{xy} \pm SE\) (eqq. [16–21] in Nei and Jin 1989). All estimates are corrected for multiple substitutions by the Jukes-Cantor method, and the standard errors include both sampling and intralocus stochastic errors.

despite the sampling bias in favor of replacement-site polymorphisms in the D. melanogaster sample noted above. The difference is indeed mainly due to the D. simulans data. The ratio of replacement to silent polymorphism in D. simulans alone (0.17) is significantly different from the ratio for the fixed differences between the species (\(G = 8.52, df = 1, P < 0.01\)), whereas the corresponding ratio for D. melanogaster alone (0.36) is not (\(G = 0.65, df = 1, P > 0.05\)).

Following Whittam and Nei (1991) we have also estimated the divergence between species (\(d_x\)), after correcting for multiple hits by the Jukes-Cantor method and including the intralocus stochastic variance (eqq. [16–20] in Nei and Jin 1989). Estimates of polymorphism (\(\hat{\pi}\)) in the various sequenced regions of Est-6 in D. simulans and D. melanogaster and divergence between D. simulans and two of its sibling species (\(d_{xy}\) or \(d_x\)) are shown in table 3. Consistent with the McDonald and Kreitman test, we find that the ratio of values for replacement and silent sites for D. simulans versus D. melanogaster (\(d_R/d_S = 0.17 \pm 0.04\) per 100 nucleotide sites) is indeed higher than the corresponding ratio of averaged \(\hat{\pi}\) values (\(\hat{\pi}_R/\hat{\pi}_S = 0.08 \pm 0.01\)), although in this case the difference between the two ratios is only statistically significant at the 10% level (\(Z = 1.72, P = 0.085\)).

We have also applied the HKA test of Hudson et al. (1987) and Kreitman and Hudson (1991) to the four D. simulans sequences, using various upstream sequences of Est-6 and Adh as our neutral reference sequence and various pairs of D. melanogaster and D. simulans Est-6 alleles to calculate reference divergence statistics. None of the test statistics calculated were significant (data not shown). However, it is doubtful whether the HKA test statistic would approximate a \(\chi^2\) distribution for data from a sample of only four lines (Kreitman and Hudson 1991), and several of our test statistic values were indeed substantially smaller than expected under \(\chi^2\). Because they do not represent a random sample of EST6 alleles, the 13 D. melanogaster sequences of Cooke and Oakeshott (1989) are also unsuitable for use in calculations of polymorphism in a HKA test.

Promoter region sequence is available for only one isolate of Est 6 from D. melanogaster (Collet et al. 1990; Karotam et al. 1993), so we cannot compare levels of polymorphism in this region across the two species. However, we can at least compare the D. simulans polymorphism data with the pattern of divergence between the two species in this region. The major features of both datasets are the remarkably low levels of variation in the first 400-bp 5' of the Est-6 coding region and the rapid rise in variation either side of this highly conserved region (fig. 2). This distribution of divergence is closely mirrored by the distribution of polymorphism among the D. simulans sequences (fig. 2).
Amino Acid Sequence Comparisons

None of the 11 amino acid replacements segregating among the four *Drosophila simulans* alleles disrupt any known structural features of the protein (Collet et al. 1990; Karotam et al. 1993). These conserved structural features include the 19 residue signal peptide (residues -19 to -1), three recognition sequences for N-linked glycosylation (starting at residues 21, 399, and 435), six cysteine residues involved in disulfide bridges (at 65 and 84, 240 and 253, 493 and 514) and three noncontiguous residues (Ser-188, Asp-319, and His-445) that comparisons with other esterases suggest are involved in the catalytic mechanism (Cygler et al. 1993). The majority of the 11 amino acid differences occurring within *D. simulans* are also conservative for at least three of the four physicochemical properties of charge, molecular volume, polarity, and hydrophobicity (fig. 3). Moreover, all changes but one (Ile/Val at 270) are predicted to be in hydrophilic regions likely to be on the surface of the protein (Kyte and Doolittle 1982), where a substitution would be less likely to affect its higher-order structure (fig. 3). Amino acid replacements differing between *D. simulans*, *D. melanogaster*, and *D. mauritiana* also apparently preserve secondary and most tertiary structures of EST6 (Karotam et al. 1993; Oakeshott et al. 1994).

Which of the 11 amino acid polymorphisms in *D. simulans* identified here underlie the allozymic variation for EST6 in this species? Our electrophoretic analyses show that each of the four *D. simulans* alleles sequenced encodes a different EST6 allozyme. These allozymes, with their relative mobilities in parentheses, are EST6-F (107) for sim-1, EST6-vF (113) for sim-2, EST6-vF (115) for sim-3, and EST6-S (98) for sim-B (mobilities are calculated relative to the common EST6-S variant in *D. melanogaster*, EST6-8, whose Rf is taken as 100; Labate et al. 1989). Two amino acid polymorphisms separate the EST6-S sequence from all three fast variants in *D. simulans*; Thr/Asn at 237 and Asp/Val at 487. Only the latter involves a charge difference (negatively charged Asp in all the fast variants versus uncharged Val in EST6-S) and is presumably responsible for the more anodal mobility of the EST6-S variant (see table 4). Another polymorphism involving a charge difference, Asn/Asp at 377, is likely to be responsible for the major difference in mobility between EST6-F (uncharged Asn) and the two very fast variants, EST6-vF and EST6-vF (negatively charged Asp). The cause of the minor mobility difference between the two very fast variants is less clear, as all four amino acid differences between the two are charge conservative.

Our larger electrophoretic analyses of 31 *D. simulans* lines from the Raleigh, North Carolina, population reveal that two allozymes of similar but not identical mobilities segregate within both EST6-F (with Rf values of 107 and 108, comprising 36% and 31% of the sample, respectively) and EST6-S (with Rf values of 98 and 99, comprising 32% and 2% of the sample). These data suggest that the *D. simulans* EST6-F and EST6-S variants sequenced here (Rf's of 107 and 98, respectively) are representative of common EST6-F and EST6-S forms in this population, and they concur with the observation of Albuquerque and Napp (1981) that the major F and S forms of EST6 in *D. simulans* each include minor mobility variants.

Curiously, two of the three Australian lines sequenced herein are of EST6-vF allozymes that are not represented among the 31 American lines electrophoresed here and that previous surveys have shown are not

### Table 4

<table>
<thead>
<tr>
<th>Amino Acid Substitutions</th>
<th><em>Drosophila simulans</em> (F/S)</th>
<th><em>Drosophila melanogaster</em> (F/S)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Charge altering</td>
<td>Asp/Val (487)</td>
<td>Asp/Asn (237)</td>
</tr>
<tr>
<td></td>
<td>(-0.5/0)</td>
<td>(-0.5/0)</td>
</tr>
<tr>
<td>Charge conservative</td>
<td>Thr/Asn (237)</td>
<td>Ala/Thr (247)</td>
</tr>
</tbody>
</table>

**Note:** The positions of each difference in the primary sequence of the mature protein and the charge state of the residues involved in the charge-altering polymorphisms are indicated in parentheses.
generally common in other Australian, European, or African populations either (Anderson and Oakeshott 1984; Hyyata et al. 1985). We suspect that some selection among Est 6 variants may have occurred over the several generations of inbreeding originally carried out to homozygose the D. simulans alleles for sequencing. Such selection is difficult to avoid, given the high likelihood of nonrandom associations developing between Est-6 alleles and linked viability genes during the inbreeding. In so much as the result has been the inclusion of two relatively uncommon allozymes in the sample sequenced, the sample may be biased upward in respect of amino acid polymorphisms. Some caution is therefore required in, for example, comparing the absolute levels of polymorphism with other loci, although it makes the relative deficit of replacement polymorphism with respect to divergence even more noteworthy.

Table 5 compares the various D. simulans allozymes we have sequenced or electrophoresed to those of all 14 EST6 allozymes that have been reported in D. melanogaster (Cooke and Oakeshott 1989; Labate et al. 1989). As noted above, the D. melanogaster variants can be grouped into five major mobility classes, the most common of which are EST6-F (which includes five allozymes) and EST6-S (which includes six allozymes). Only two of the four D. simulans allozymes sequenced have Rf values the same as any of the fourteen D. melanogaster allozymes so far identified. One is the D. simulans fast variant (sim-1), which has identical mobility to one of the common EST6-F variants (EST6-4) of D. melanogaster. The other is the most common EST6-S variant of D. simulans (sim-B) the mobility of which is the same as a rare D. melanogaster allozyme (EST6-10), both having a more anodal mobility than the most common EST6-S of D. melanogaster (EST6-8). This latter discrepancy is contrary to the findings of previous studies that have classified the EST6-F and EST6-S forms of the two species as identical, at least on relatively insensitive starch gels (e.g., Hyyata et al. 1985).

There are no shared amino acid polymorphisms between the species although there are two instances of a different polymorphism occurring at the same site. One is at position 487, where a charge-altering Asp/Val polymorphism separates EST6-F and EST6-S in D. simulans...
*simulans*. but a charge conservative Ser/Ala polymorphism occurs in both fast and slow electrophoretic variants in *D. melanogaster* (Cooke and Oakeshott 1989). The other is at residue 237, where a charge conservative Thr/Asn occurs in *D. simulans* but a charge-altering Asp/Asn occurs in *D. melanogaster*. The latter difference, along with a charge conservative Ala/Thr polymorphism at 247, separates the major EST6-F and EST6-S electrophoretic classes in *D. melanogaster* (Cooke and Oakeshott 1989). These results indicate that the EST6-F/EST6-S electrophoretic differences in the two species are due to different polymorphisms at different sites.

**Discussion**

The level of replacement polymorphism ($\theta = 0.47$) for *Est-6* in *Drosophila simulans* is higher than the highest values previously reported for isozymes in *Drosophila* (0.41 for *Est-6* in *D. melanogaster* and 0.36 for *Xdh* in *D. pseudoobscura*; Cooke and Oakeshott 1989; Riley et al. 1992). Although our estimate of $\theta$ may be biased upward by an unintentional oversampling of uncommon allozymes, the surveys of *Est-6* in *D. melanogaster* and *Xdh* in *D. pseudoobscura* were also based on overdispersed samples (deliberately so in those studies). The value for *Est-6* in *D. simulans* corresponds to an average pairwise difference of 6.5 residues between sequences, which is nearly twice as high as for any other enzyme-encoding locus reported to date.

Although the level of replacement-site polymorphism for *Est-6* is high in comparison to other loci, it is still less than what might be expected from comparison to levels of *Est-6* divergence between *D. simulans* and *D. melanogaster*. The proportion of replacement to exon silent polymorphism in both *D. simulans* and *D. melanogaster* combined is less than the proportion of replacement to silent divergence between the species. The difference is statistically significant ($P < 0.05$) in the form of comparison advocated by McDonald and Kreitman (1991) and nearly so ($P = 0.085$) in the form developed by Whittam and Nei (1991). Moreover, the difference is likely to be underestimated because both the *D. simulans* and *D. melanogaster* data on levels of polymorphism are derived from overdispersed samples of allozymes. The difference is in fact largely due to the *D. simulans* data; use of the polymorphism data from this species alone extends the significance of the McDonald and Kreitman test to $P < 0.01$ (no valid form of the Whittam and Nei test can be generated using the polymorphism data from one species alone). The differences observed could be due to divergent selection on EST6 function between the two species or to selection reducing replacement polymorphism within *D. simulans*. We have no direct evidence to distinguish between these alternatives.

Our sequence data also indicate that the amino acid differences underlying the clinal selection of the EST6-F and EST6-S allozymes of *D. simulans* are not the same as those underlying the parallel clines for the EST6-F and EST6-S alleles in *D. melanogaster*. Indeed, we can now see that the respective EST6-F/EST6-S allele polymorphisms in the two species differ in three critical respects. Firstly the $R_f$ values for the common EST6-S allozymes are 100 in *D. melanogaster* and 98 in *D. simulans* and thus are not identical in the first place. Secondly, although $R_f$ values for one of the common EST6-F allozymes is 107 in each species, in each case other minor forms can also rise to high frequencies. Finally, the EST6-F/EST6-S difference is associated with a Thr/Asn polymorphism at position 237 and an Asp/Val at 487 in *D. simulans* but with an Asp/Asn polymorphism at 237 and a Thr/Ala at 247 in *D. melanogaster*. Whatever the nature of the selective processes, if any, underlying the clines in each species, we must now conclude that the targets for the selection are not identically shared amino acid polymorphisms in the two.

The most parsimonious explanation is that the clines have completely independent molecular bases, namely, the polymorphisms assumed to cause the EST6-F/EST6-S mobility differences in the two species at position 237 in *D. melanogaster* and 487 in *D. simulans*. An alternative explanation that does invoke some similarity in molecular basis would be that the clines in both species reflect selection on the one site among the four polymorphisms involved where an amino acid residue is shared by only one allozyme class of each species, the Asn residue at position 237 in both EST6-S classes. Such selection would require that Asn-237 differ in its fitness effects in the same way from the charged Asp in *D. melanogaster* and the uncharged Thr in *D. simulans*. This explanation also assumes that the Asp/Val polymorphism at 487, associated with the EST6-F/EST6-S difference in the *D. simulans* alleles sequenced to date, is characteristic for the two allozyme classes in that species. Further sampling of *D. simulans* allele sequences, particularly from the slow allozyme class, may reveal additional relevant polymorphisms.

In addition to the possible influence of selection on amino acid polymorphism in the coding region of *Est-6*, its regulatory sequences appear to be subject to a different type of selection. The 41 bp of untranslated sequence immediately 5' of the coding region is invariant among the four *D. simulans* alleles, and the adjacent proximal 5' untranscribed region is nearly so (table 1). Strong directional selection or functional constraint would be necessary to retain such a low level of poly-
morphism. Levels of sequence divergence from *D. melanogaster* and *D. mauritiana* are also extremely low in the same regions (table 3), indicating the operation of strong selective constraint. This strong constraint of proximal regulatory sequences stands in sharp contrast to levels of polymorphism and divergence within the coding region and further 5' (fig. 2) and suggests that appropriate expression of EST6 is essential in these species. Consistent with this interpretation, the similarity between the promoter regions of *D. melanogaster* EST-6 and its homolog in *D. pseudoobscura* (EstSB) is essentially confined to the most proximal 174 bp of 5' sequences (Brady et al. 1990).

Moreover, evidence from functional studies indicates that the most proximal 150 bp of 5' sequence is necessary for basal levels of Est-6 expression in *D. melanogaster* (M. J. Healy, M. M. Dumancic, and J. G. Oakeshott, unpublished data) and no EST6-null individuals have ever been recovered from natural populations (Oakeshott et al. 1990).

Acknowledgments

We thank Chip Aquadro for fly stocks and valuable advice, Marion Healy and Mira Dumancic for sharing unpublished results, and Peter Cooke, Anne Game, Mark Myers, Wendy Odgers, Lis van Papenrecht, Robyn Russell, and the referees of early versions of this manuscript for valuable advice.

LITERATURE CITED


RICHARD HARRISON, reviewing editor

Received July 15, 1994
Accepted August 2, 1994