Cloning of the Esterase-5 Locus from *Drosophila pseudoobscura* and Comparison with Its Homologue in *D. melanogaster*

**J. P. Brady,** * R. C. Richmond, * and J. G. Oakeshott†

*Department of Biology and Institute for Molecular and Cellular Biology, Indiana University, and †CSIRO Division of Entomology, Australia

A clone of the esterase-5 (Est-5) gene from *Drosophila pseudoobscura* has been isolated by hybridization to the cloned Est-6 gene of *D. melanogaster*. Southern analysis and sequencing of the cloned DNA revealed three regions of similarity to Est-6 that have been tentatively identified as genes, Est-5A, Est-5B, and Est-5C. Introduction of each of the three genes separately into *D. melanogaster* by P-element transformation has demonstrated that Est-5B encodes an enzyme with the same physical properties as EST 5 in *D. pseudoobscura*. Sequence analysis indicates that Est-5B encodes a 545-amino-acid protein and is composed of two exons separated by a 55-bp intron in the same position as the 51-bp intron in Est-6. Comparison of the Est-5B coding region with that of Est-6 reveals an overall similarity (73% at both the nucleotide and amino acid levels) that is substantially lower than that for other genes sequenced in both of these species. Total nucleotide and nonsynonymous site differences between Est-6 and Est-5B are more abundant in the second exon than in the first, suggesting differential effects of selection or mutation on these two exons. Comparisons of the 5'-flanking DNA of Est-5B and Est-6 reveal four short conserved sequence elements, but the remaining upstream sequences show no significant similarity. Conservation in the 3'-flanking DNA is limited to the presence of two polyadenylation sites that may correlate with the existence of two transcripts from both Est-5B and Est-6. The patterns of nucleotide substitutions and amino acid replacements between Est-5B and Est-6 are consistent with the hypothesis that mutation and genetic drift are responsible for the differences between these two genes.

**Introduction**

The relative roles of natural selection and mutation/genetic drift in producing evolutionary change have been widely contested (for a thorough discussion, see Nei 1987). Kreitman and Aguadt (1986) and Hudson et al. (1987) have proposed a test of the neutral theory based on the prediction that the degree of variation among parts of a gene (e.g., exons, introns, and flanking regions) should be the same for inter- and intraspecific comparisons if mutation and genetic drift are the primary forces affecting this variation, as the neutral theory of molecular evolution posits (Kimura 1983). Any comprehensive theory of molecular evolution must account for sequence data from many loci that encompass a range of substitution levels. We have undertaken a
molecular analysis of the highly polymorphic esterase-6 (Est-6 = gene, EST 6 = enzyme) gene of *Drosophila melanogaster* and its homologues in several related species in order to compare the patterns and levels of substitution at this locus with those of other genes and to obtain data that will allow us to test the neutral theory in the manner of Kreitman and Aguadé (1986).

Natural populations of *Drosophila* contain two predominant electrophoretic variants of the EST 6 enzyme as well as many minor allozymes (Cooke et al. 1987). We have cloned the Est-6 gene from *D. melanogaster* and have sequenced 13 isolates of the locus that are representative of variation found in natural populations (Oakeshott et al. 1987; Cooke and Oakeshott 1989). These studies suggest that purifying natural selection is acting on the locus, as predicted by the neutral theory, but there is also evidence that balancing selection maintains at least the major electrophoretic classes of alleles. In the present paper, we provide data on interspecific evolutionary changes at the Est-6 locus, through a comparative molecular analysis of the Est-6 homologue in *D. pseudoobscura*, Est-5.

We have chosen to study the Est-5 locus for several reasons. First, EST 5 allozyme variation segregating within natural populations of *D. pseudoobscura* has been extensively quantified (Lewontin and Hubby 1966; Coyne et al. 1978; Keith 1983). A molecular clone of Est-5 will allow the assessment of nucleotide variation within and between electrophoretic classes, through restriction mapping and DNA sequence analysis (Langley et al. 1982; Kreitman 1983). A second reason for studying Est-5 is that a substantial number of other genes have been sequenced in both *D. melanogaster* and *D. pseudoobscura*. These sequences will provide a basis for comparing substitution rates, testing models of evolution, and constructing phylogenies. Finally, EST 5 and EST 6 are significantly different from one another in their expression patterns and quaternary structures. EST 5 is expressed at high levels throughout development (Berger and Canter 1973) and is found in the eyes and hemolymph of adult *D. pseudoobscura* flies of both sexes (Lunday and Farmer 1983). EST 6, on the other hand, is expressed primarily in adult *D. melanogaster* males, and its activity is mainly in the anterior ejaculatory duct (Sheehan et al. 1979). Moreover, EST 5 occurs primarily as a dimer in *D. pseudoobscura* (Hubby and Lewontin 1966), with an estimated molecular weight of 100 kD (Narise and Hubby 1966), whereas EST 6 in *D. melanogaster* is a monomer of approximately 62 kD (Mane et al. 1983). These differences suggest a high rate of divergence at this locus, and therefore sequence comparisons between Est-5 and Est-6 are likely to reveal patterns of evolutionary change that are not apparent in more highly conserved genes.

Our efforts to clone Est-5 were based on the presumption that it is homologous to Est-6 in *D. melanogaster*. Evidence for homology between these two genes is provided by genetic (Wright 1963; Abraham and Lucchessi 1974) as well as biochemical (Arnason and Chambers 1984) data. Overlapping bacteriophage clones that hybridize to Est-6 were isolated from two *D. pseudoobscura* libraries. Our data suggest that three closely linked genes, all with sequence identity to Est-6, are contained in the cloned interval. Here we present and analyze the sequence of one of these genes, which we propose is the structural locus for the EST 5 protein. The nucleotide and amino acid sequences of Est-5 and Est-6 are analyzed in terms of possible selective constraints and are compared with findings from similar studies on other *Drosophila* genes, to assess the nature of the evolutionary changes at the Est-5/6 locus.
Cloning of the Esterase 5 Gene

Material and Methods

Drosophila Stocks

*Drosophila melanogaster* flies homozygous for both the ry<sup>206</sup> mutation and the slow allele of EST 6 were a gift from M. Brennan. Flies homozygous for the null allele of EST 6 have been described elsewhere (Sheehan et al. 1979). The following strains of *D. pseudoobscura* were obtained from Dr. R. C. Lewontin’s laboratory at Harvard University: *Est-5<sup>Null</sup>*, *Est-5<sup>0.85</sup>* and *Est-5<sup>1.00</sup>* (Hubby and Lewontin 1966).

Genomic Library Screening

Two *D. pseudoobscura* libraries were screened with a cDNA clone of *Est-6* (Oakeshott et al. 1987). One library contained completely digested *EcoRI* fragments of *D. pseudoobscura* genomic DNA in the *EcoRI* site of lambda gt10 bacteriophage vectors (Huynh et al. 1984). The second library, made by C. Langley and obtained from D. Cavener, contained partially digested *Sau3A* fragments of genomic DNA from the *Est-5<sup>1.00</sup>* strain (Narise and Hubby 1966) that were cloned into the *BamHI* site of lambda EMBL4. Approximately 50,000 plaques from the lambda gt10 library and approximately 80,000 plaques from the lambda EMBL4 library were screened using standard techniques outlined below.

Nucleic Acid Analysis

Bacteriophage lambda and plasmid DNA were prepared using standard procedures (Maniatis et al. 1982). Both genomic DNA and total RNA were isolated from *D. pseudoobscura* flies isoallelic for *Est-5* and from *D. melanogaster* flies by using the RNA extraction protocol of Collet et al. (1990). Transfer of DNA and RNA from agarose gels to nylon membranes (Amersham) was performed according to the method of Maniatis et al. (1982).

Radioactive probes for Southern blot and plaque hybridizations were made by random priming of isolated restriction fragments or plasmid subclones (Maniatis et al. 1982). Southern and plaque hybridizations were performed at 55°C in 3 × SSC (20 × SSC = 3 M NaCl, 0.3 M Na<sub>2</sub> citrate 2 H<sub>2</sub>O, pH 7.0) followed by two washes at 55°C in 2 × SSC. Single-strand RNA probes for Northern blot hybridizations were transcribed from promoters on the plasmid vectors pGEM-3 (Promega Biotec) and Bluescribe M13( +) (Stratagene) with T7 RNA polymerase. Northern hybridizations were performed at 65°C in 5 × SSC and 1 × PE [5 × PE = 250 mM Tris-HCl, pH 7.5, 5% sodium dodecyl sulfate (SDS), 1% polyvinylpyrrolidone, 1% Ficoll, 25 mM ethylene diaminetetraacetate]. Northern blots were washed twice at 65°C in 2 × SSC, 0.1% SDS for 30 min, followed by two more 30 min washes in 0.1 × SSC, 0.1% SDS, also at 65°C.

Characterization of Clones and DNA Sequencing

DNA from lambda clones was digested with *EcoRI*, separated on agarose gels, blotted, and probed with the *Est-6* cDNA clone. Restriction fragments hybridizing to *Est-6* were subcloned into plasmid vectors. The resulting plasmids were digested with a series of restriction enzymes, ligated into the corresponding sites of M13mp18 and M13mp19 (Norrander et al. 1983), and sequenced in both directions by using the dideoxy chain-termination method of Sanger et al. (1977). Sequence analysis was performed using version 5.0 of the software assembled by the University of Wisconsin Genetics Computer Group (Devereux et al. 1984).
P-Element Transformation

Transformations were carried out essentially according to the method of Rubin and Spradling (1982). The EcoRI/SalI, HindIII/EcoRI, and EcoRI/XbaI fragments indicated in figure 1 were ligated into a Carnegie 20 derivative, pDM30 (Rubin and Spradling 1983; Mismer and Rubin 1987). *Drosophila melanogaster ry<sup>506</sup>* embryos were injected with the cloned DNA at concentrations from 100 µg/ml to 500 µg/ml, along with 100 µg of the helper plasmid pPi25.7 wc/ml (Karess and Rubin 1984). Homogenates of transformed adult flies were run on 10% polyacrylamide native gels that were subsequently stained for esterase activity according to the methods of Mane et al. (1983) by using alpha-naphthyl acetate and beta-naphthyl acetate as substrates. Heterodimers were formed in vitro by jointly homogenizing flies from different lines in 0.1 M phosphate buffer, pH 6.8, and allowing the homogenates to sit at room temperature for 1 h before loading the gels.

Results

Characterization of Cloned DNA

Several overlapping bacteriophage clones that hybridized to an *Est-6* cDNA probe from *Drosophila melanogaster* were isolated from *D. pseudoobscura* genomic libraries in lambda gt10 and lambda EMBL4. All of the lambda gt10 clones contained a 3.2-kb EcoRI fragment that hybridized to *Est-6*, and the lambda EMBL4 clones contained both a 7.9-kb and a 3.2-kb EcoRI fragment, or portions thereof, that also hybridized with the probe. The 3.2-kb EcoRI fragments from both libraries have matching restriction maps. The failure to isolate the 7.9-kb EcoRI fragment from the lambda gt10 library is likely a result of the small cloning capacity (<8 kb) of this vector (Huynh et al. 1984).

A restriction map of the cloned DNA is shown in figure 1, along with a Southern blot of *D. pseudoobscura* genomic DNA probed with a cDNA clone of *Est-6* under the same conditions used to screen the library. The blot shows a pattern of bands that corresponds to the restriction map and thus suggests that we detected all classes of clones with potential homology to *Est-6* in the library screens. We also have evidence that the cloned DNA hybridizes in situ to a portion of the right arm of the X chromosome in *D. pseudoobscura* that is included within the “sex-ratio” inversion (Wallace 1948; W. Anderson, personal communication). This region has been identified as the cytological location of *Est-5* (Beckenbach 1981). These results suggest that the clones isolated from the *D. pseudoobscura* libraries are good candidates for the locus homologous to *Est-6* in *D. pseudoobscura* and are likely to contain the *Est-5* gene.

Southern hybridization of the *Est-6* cDNA clone to restriction fragments of the cloned *D. pseudoobscura* sequences identified three separate regions, indicated in figure 1, that hybridized to *Est-6* DNA. We have sequenced almost the entire *D. pseudoobscura* genomic region that contains the three segments that hybridize to *Est-6*. All three regions contain open reading frames that can be aligned with the open reading frame of *Est-6*. The *D. pseudoobscura* open reading frames all have considerable similarity to one another (78%–85% nucleotide, 76%–85% amino acid) as well as to the open reading frame of *Est-6* (68%–73% nucleotide, 70%–74% amino acid). Moreover, each of the three open reading frames is interrupted by a sequence of approximately the same length (55–60 bp) that is in the same position as the intron in *Est-6* (Collet et al. 1990). The ends of all three introns correspond to consensus *Drosophila* splice sites (Keller and Noon 1985). Therefore, we believe that these regions represent three
FIG. 1.—A, Restriction map of cloned DNA isolated from *Drosophila pseudoobscura* genomic libraries probed with *Est-6* of *D. melanogaster*. C, B, and A indicate the locations of three putative genes, based on hybridization of the *Est-6* cDNA probe to cloned restriction fragments. Cross-hatched regions below the map represent regions that were sequenced; the heavy bars indicate restriction fragments that were subcloned and transformed into *D. melanogaster* flies. R = *EcoR*I; B = *BamHI*; H = *HindIII*; P = *Pst*I; S = *SalI*; X = *XbaI*. B, Southern blot of *D. pseudoobscura* genomic DNA probed with *Est-6*. Bands on the blot correspond to the restriction map of the cloned DNA in panel A.
separate genes that we refer to as Est-SA, Est-SB, and Est-SC. In a future report, we will present a detailed analysis of sequence data for all three D. pseudoobscura genes (J. P. Brady and R. C. Richmond, unpublished data). We present here the sequence of the Est-5B gene, which we show encodes the EST 5 protein.

EST 5 is a very polymorphic enzyme (Coyne et al. 1978; Keith 1983). If this protein were encoded by more than one gene, then multiple bands of EST 5 activity should frequently be present in native gel analyses of individual male flies, because it is unlikely that the same allele would be present at all three loci. That only one EST 5 band is observed in such analyses (Hubby and Lewontin 1966) suggests that EST 5 is encoded by a single structural locus. We next determined which of the genes—Est-5A, Est-5B, or Est-5C—encodes the EST 5 protein.

P-Element Transformation Experiments

Three restriction fragments, indicated in figure 1A, were introduced, individually, into the chromosomes of D. melanogaster flies by using a derivative of the Carnegie 20 vector (see Material and Methods). These fragments contained all of the coding sequences for the three genes, as well as 450–2,700 bp of 5' and 125–1,100 bp of 3'-flanking DNA. At least two different transformed lines were obtained for each construct. Confirmation of successful transformations was obtained by genomic Southern analysis (data not shown). Adult transformants were analyzed by running homogenates on native polyacrylamide gels that were subsequently stained for esterase activity. No new esterases were seen in flies transformed with genes Est-5A and Est-5C. However, in the four lines transformed with Est-5B a new beta esterase (i.e., an esterase that preferentially cleaves beta-naphthyl acetate) was observed (fig. 2A) that comigrated with the EST 5^1.00 allozyme of D. pseudoobscura, the same line from which the EMBL4 library was made and from which Est-5B was isolated.

Hubby and Narise (1967) demonstrated that mixed homogenates from D. pseudoobscura strains homozygous for different electromorphs of EST 5 form heterodimers in vitro with an intermediate electrophoretic phenotype. Figure 2B shows that the new esterase in our transformants is also capable of forming heterodimers with a slower-migrating electromorph of EST 5 and that these heterodimers do not form between EST 6 and EST 5 in combined homogenates of D. melanogaster and D. pseudoobscura. The different relative mobilities of EST 5 and EST 6 in figures 2A and 2B are the result of the gels being run at different temperatures. At room temperature, EST 5 behaves as a monomer; at colder temperatures it forms dimers and, therefore, migrates more slowly (Arnason and Chambers 1987).

Figure 3A shows a northern blot of total RNA from D. pseudoobscura and from D. melanogaster flies that is transformed with Est-5A, Est-5B, and Est-5C. The blot was hybridized with a single-stranded probe corresponding to the HindIII/EcoRI fragment containing Est-5B (fig. 1A). Two transcripts that are nearly the same lengths as the transcripts produced by the Est-6 gene in D. melanogaster (fig. 3B) are seen in D. pseudoobscura flies and in D. melanogaster flies transformed with Est-5B. The two Est-6 transcripts likely result from the alternative usage of two polyadenylation signals that are separated by 132 nucleotides (Collet et al. 1990). Likewise, we propose that the two Est-5B transcripts in D. pseudoobscura result from the use of different polyadenylation sites, as described below. No transcripts were detected in either D. pseudoobscura or the D. melanogaster transformants when the same blot was hybridized with probes corresponding to the EcoRI/EcoRI and HindIII/HindIII restriction fragments that contain Est-5A and Est-5C, respectively (fig. 1). However, similar analyses
of other life-cycle stages revealed that the Est-5A probe hybridizes to a single 1.8-kb transcript in third-instar larvae of D. pseudoobscura and D. melanogaster transformed with Est-5A (fig. 4). Est-5C transcripts could not be detected either in any developmental stage of D. pseudoobscura or in the appropriate transformants (data not shown).

The similarity of transcripts produced by Est-5B and Est-6, the ability of the new esterase in flies transformed with Est-5B to form heterodimers with EST 5, the parallel electrophoretic mobilities of this esterase and the EST 5.100 allozyme, and the similar substrate preferences (β-naphthyl acetate is preferentially cleaved relative to α-naphthyl acetate) of this enzyme with EST 5 lead us to conclude that Est-5B encodes the EST 5 protein.

A line of D. pseudoobscura (Est-5Null) that does not make any detectable EST 5 protein still produces the two Est-5B transcripts (fig. 3A, lane 4). Genomic Southern analysis of DNA from these null flies (data not shown) shows no detectable differences
Fig. 3.—Northern analysis of total RNA from untransformed *Drosophila melanogaster* homozygous for both the slow allele of EST 6 and the *ry*\(^{506}\) allele (lane 1); *D. melanogaster*, EST 6 null (lane 2); *D. pseudoobscura*, EST-5\(^{1,00}\) (lane 3); *D. pseudoobscura*, EST 5 null (lane 4); *D. melanogaster* (EST-6\(^{8}\), *ry*\(^{506}\)) transformed with EST-5A (lane 5); *D. melanogaster* (EST-6\(^{5}\), *ry*\(^{506}\)) transformed with EST-5B (lane 6); *D. melanogaster* (EST-6\(^{5}\), *ry*\(^{506}\)) transformed with EST-5C (lane 7); and *D. pseudoobscura*, EST-5\(^{1,00}\) (lane 8). Each lane contains RNA from 10 adult male flies. A, Filter probed with EST-5B. B, Same filter probed with EST-6.

from wild-type *D. pseudoobscura* DNA within the 11.1-kb region indicated in figure 1, suggesting that a point mutation or a small deletion or insertion is responsible for the loss of EST 5 activity in this strain. EST 6 null flies do not produce any EST-6 transcripts because of a transposable element insertion within the EST-6 locus (R. W. Phillis, K. M. Nielsen, and R. C. Richmond, personal communication).

Sequence Analysis of EST-5B

The complete nucleotide sequence of EST-5B and its flanking regions is presented in figure 5. Alignment of the coding region with the coding region of the genomic
DNA sequence of the Est-6, Dm145 allele (Collet et al. 1990), reveals a high degree of overall similarity (73% nucleotide, 73% amino acid) between the two genes (table 1). The levels of divergence for these genes, however, are substantially higher (by a factor of 1.7–9) than those for other genes whose amino acid sequences have been compared between the melanogaster and obscura species groups (table 2). Est-5B contains a single 55-bp intron in the same position as the 51-bp intron in Est-6 (Collet et al. 1990). Although the ends of the introns in both genes conform to consensus Drosophila intron splice sites (Keller and Noon 1985), there is little similarity between the two introns. The consensus polyadenylation signal (Birnstiel et al. 1985) at position 1692 (AATAAA) or the nonconsensus signal (AATACA) at 1726 is likely to produce the smaller Est-5B transcript (fig. 3), whereas one of the nonconsensus polyadenylation signals at positions 1803 (AACAAA) or 1812 (AAAATA) may be responsible for the larger transcript from this gene.

A conceptual translation of the coding region of Est-5B produces a protein of 545 amino acids, compared with 544 amino acids for EST 6. The first 19 amino acids of the Est-5B product are predominantly hydrophobic and match other criteria for a signal peptide (von Heine 1984). EST 6 also appears to have a signal peptide (Oakeshott
FIG. 5.—Nucleotide and inferred amino acid sequence of the Est-5B gene of Drosophila pseudoobscura.

This sequence was obtained from DNA isolated from an EMBL4 genomic library of D. pseudoobscura prepared from a strain homozygous for the Est-5B allele. Numbering begins with the start of translation. The putative TATA box and potential polyadenylation signals are underlined. The intron sequence is in lowercase letters. Nucleotide differences between Est-5B and Est-5A are indicated below the Est-5B sequence. The coding region of Est-5B was aligned with the coding region of Est-5A by using the algorithm of Needleman and Wunsch (1970). Separate alignments were made for each exon. Dashes represent gaps introduced into the sequences.
Table 1
Sequence Differences between Est-5B in Drosophila pseudoobscura and Est-6 in D. melanogaster

<table>
<thead>
<tr>
<th>Region</th>
<th>Length (bp)</th>
<th>No. (%) of Nucleotide Differences&lt;sup&gt;a,b&lt;/sup&gt;</th>
<th>No. (%) of Amino Acid Differences&lt;sup&gt;a&lt;/sup&gt;</th>
<th>%S/%R&lt;sup&gt;c&lt;/sup&gt;</th>
<th>Transitions/Transversions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Signal sequence</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>total</td>
<td>57</td>
<td>21 (36.8)</td>
<td>11 (57.9)</td>
<td>ND</td>
<td>ND</td>
</tr>
<tr>
<td>Exon I&lt;sup&gt;d&lt;/sup&gt;</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>1,333</td>
<td>350 (26.4)</td>
<td>108 (24.5)</td>
<td>4.9</td>
<td>1.0</td>
</tr>
<tr>
<td>Silent</td>
<td>302</td>
<td>206.5 (69.2)</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>Replacement</td>
<td>1,031</td>
<td>143.5 (14.0)</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>Intron total</td>
<td>55</td>
<td>32 (62.7)</td>
<td>...</td>
<td>...</td>
<td></td>
</tr>
<tr>
<td>Exon II&lt;sup&gt;d&lt;/sup&gt;</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>248</td>
<td>81 (32.7)</td>
<td>31 (37.8)</td>
<td>3.2</td>
<td>0.91</td>
</tr>
<tr>
<td>Silent</td>
<td>55.7</td>
<td>37 (71.4)</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>Replacement</td>
<td>192.3</td>
<td>44 (22.3)</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>Exon 1 + Exon II:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>1,581</td>
<td>431 (27.4)</td>
<td>139 (26.6)</td>
<td>4.6</td>
<td>0.98</td>
</tr>
<tr>
<td>Silent</td>
<td>357.7</td>
<td>243.5 (69.6)</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>Replacement</td>
<td>1,223.3</td>
<td>187.5 (15.3)</td>
<td>...</td>
<td>...</td>
<td></td>
</tr>
</tbody>
</table>

NOTE.—ND = not determined.

<sup>a</sup> Insertions/deletions were not included in the numbers and percentages of differences.

<sup>b</sup> Percentages of silent and replacement substitutions are based on the average number of silent and replacement sites in both sequences.

<sup>c</sup> Ratio of percentages of silent and replacement substitutions.

<sup>d</sup> Exons I, II coding regions only.

et al. 1987), but the degree of similarity between the two signal sequences is much lower than that for the rest of the two proteins (table 1). Alignment of the EST 6 and EST 5 amino acid sequences (fig. 6) indicates that the putative EST 6 signal peptide is two amino acids longer than the proposed EST 5 signal sequence, but the mature EST 6 protein is likely to be three amino acids shorter than EST 5, because of two insertions/deletions in exon I.

Figure 7 shows the distribution of nucleotide substitutions and amino acid replacements in nonoverlapping segments of 30 bp across the length of Est-5B/6. The numbers of these changes in each exon, as well as synonymous and nonsynonymous site differences, are listed in table 1. Synonymous site changes are evenly distributed between the two exons ($\chi^2 = 0.1, df = 1, P = 0.75$). However, total nucleotide and nonsynonymous site differences are significantly higher in the second exon than in the first ($\chi^2 = 4.1, df = 1, P = 0.04$; and $\chi^2 = 8.8, df = 1, P = 0.003$). Table 1 also indicates that nucleotide substitutions in effectively synonymous sites have occurred at a 4.6-fold higher rate than nonsynonymous site changes. This suggests that purifying selection has acted to remove unfavorable amino acid replacements.

The transition/transversion ratio in both exons is $\sim 1:1$—conflicting with the expected transition/transversion ratio of 1:2 (Nei 1987). Est-6 and Est-5B show a similar codon bias ($\chi^2 = 46.62, df = 45, P > 0.40$), but both differ from the codon bias for other Drosophila genes ($\chi^2 = 172.5, df = 45, P < 0.001$; and $\chi^2 = 101.0, df = 45, P < 0.001$) (M. Ashburner, personal communication).

Comparison of amino acid replacements between Est-5B and Est-6 reveals that...
Table 2
Amino Acid and Silent-Site Divergence of Homologous Genes Sequenced in *melanogaster* and *obscura* Species Groups

<table>
<thead>
<tr>
<th>Gene</th>
<th>Species Compared</th>
<th>Amino Acid Divergence (%)</th>
<th>Silent Site Divergence (%)</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Est5B</em>/6</td>
<td>Dm/Dps</td>
<td>27</td>
<td>70</td>
<td>Present paper</td>
</tr>
<tr>
<td><em>Adh</em></td>
<td>Dm/Dps</td>
<td>9*</td>
<td>29*</td>
<td>Schaeffer and Aquadro 1987</td>
</tr>
<tr>
<td><em>Adh</em></td>
<td>D. maur./Dps</td>
<td>10</td>
<td>46</td>
<td>Schaeffer and Aquadro 1987</td>
</tr>
<tr>
<td><em>Adh 3</em> gene</td>
<td>D. maur./Dps</td>
<td>5</td>
<td>63</td>
<td>Schaeffer and Aquadro 1987</td>
</tr>
<tr>
<td><em>Gart</em></td>
<td>Dm/Dps</td>
<td>14</td>
<td>43*</td>
<td>Henikoff and Eghtedarzadeh 1987</td>
</tr>
<tr>
<td><em>Gld</em></td>
<td>Dm/Dps</td>
<td>9</td>
<td>39</td>
<td>Krasney 1988</td>
</tr>
<tr>
<td><em>Hsp82</em></td>
<td>Dm/Dps</td>
<td>3</td>
<td>24*</td>
<td>Blackman and Meselson 1986</td>
</tr>
<tr>
<td><em>l(3)s12</em></td>
<td>Dm/Dps</td>
<td>13</td>
<td>48</td>
<td>Riley 1989</td>
</tr>
<tr>
<td><em>Per</em></td>
<td>Dm/Dps</td>
<td>16</td>
<td>36</td>
<td>Colot et al. 1988</td>
</tr>
<tr>
<td><em>RP 49</em></td>
<td>Dm/D. sub.</td>
<td>6</td>
<td>43</td>
<td>Aguadé 1988</td>
</tr>
<tr>
<td><em>Ubx</em></td>
<td>Dm/Dps</td>
<td>9*</td>
<td>32*</td>
<td>Wilde and Akam 1987</td>
</tr>
<tr>
<td><em>Xdh</em></td>
<td>Dm/Dps</td>
<td>11</td>
<td>41</td>
<td>Riley 1989</td>
</tr>
</tbody>
</table>

* Dm = *D. melanogaster*; Dps = *D. pseudoobscura*; D. maur. = *D. mauritiana*; D. sub. = *D. subobscura*.
* Figures were obtained from Riley (1989).

the majority of these changes are conservative according to the criteria of Taylor (1986): 67% for hydrophobicity, 67% for polarity, and 61% for size. Cooke and Oakeshott (1989) also found that most of the amino acid replacements among alleles of *Est-6* in *D. melanogaster* were conservative. Eight of the 16 amino acid polymorphisms that they recorded among 13 alleles of *Est-6* correspond to sites of amino acid replacement between *Est-5B* and *Est-6* (fig. 6).

Higher-Order Protein Structure

The secondary and tertiary structures of EST 5 and EST 6 would be expected to show a degree of conservation paralleling the conservation of primary structure. Hydrophathy plots of EST 5 and EST 6 are shown in figure 8. The overall similarity between these plots is high, suggesting that the secondary structure of these two proteins has indeed been conserved. The six cysteine residues found in EST 6 (Oakeshott et al. 1987) are present in the same positions in EST 5 (fig. 6). The first four cysteines have been implicated in the formation of disulfide bridges in EST 6, by implied homology with cysteine residues that have been shown empirically to form disulfide bridges in acetylcholinesterase (Chatonnet and Lockridge 1989; Cooke and Oakeshott 1989). The last two cysteines in EST 6 are also believed to form a disulfide bridge because a naturally occurring polymorphism that eliminates one of these cysteines causes a " smeared" electrophoretic phenotype (Cooke and Oakeshott 1989). Thus, conservation of the six cysteine residues suggests that selection has also acted to maintain a specific tertiary structure in both enzymes.

Upstream Sequences

There is a dramatically altered tissue localization of EST 5 and EST 6. By analogy with other systems, it is likely that the unique expression profiles of EST 5 and EST 6 are due to specific enhancer or promoter elements in the 5' regions of their respective genes (Maniatis et al. 1987). We have sequenced >450 bp between the *HindIII* site...
5' to Est-5 and the start of translation. This stretch of DNA appears to be adequate to promote expression of EST 5 in our transformed D. melanogaster flies in the same tissues where it is seen in D. pseudoobscura (J. P. Brady and R. C. Richmond, unpublished data). The putative TATA box (Bucher and Trivonov 1986) for Est-5 is underlined in figure 5. The start of transcription probably lies between nucleotides −38 and −32, because primer extension studies in D. melanogaster have implicated a corresponding region as being the CAP site for Est-6 (Collet et al. 1990).

Comparison of the 5' sequences of Est-5 and Est-6 shows very little overall similarity. Conservation is confined to four separate regions ranging in length from 12 to 40 bp (fig. 9). Closer examination of the regions −258 to −230 in Est-5B and −541 to −512 in Est-6 (fig. 9) revealed the presence of a perfect palindrome (ATATGT-ACATAT) that is identical in both species. Further evidence for the importance of this palindrome comes from the finding that D. melanogaster lines homozygous for mutations in this sequence appear to have decreased levels of EST 6 enzyme activity (Oakeshott et al. 1990). Data-base searches have revealed that this palindrome is also present in the first intron of the per locus (Jackson et al. 1986). The significance of this observation is unclear, but the deletion of the above palindrome along with surrounding sequences from the per gene greatly reduces its expression in certain adult tissues (J. Hall, personal communication).

Discussion

We have cloned from the Drosophila pseudoobscura X chromosome an 11.1-kb segment that contains three regions (Est-5A, Est-5B, and Est-5C) having high sequence similarity to Est-6 in D. melanogaster. We propose that one of these regions (Est-5B) is the Est-5 locus of D. pseudoobscura and that the other two regions are related genes. These three loci very likely arose by the duplication of an ancestral esterase-encoding locus. Analyses of D. melanogaster flies transformed with each of the three genes demonstrates that Est-5B encodes the EST 5 protein.

Comparison of the Est-5B sequence with that of its homologue in D. melanogaster, Est-6, reveals that these loci have less similarity than obtains for other genes that have been compared between these species groups. The wide range of amino acid replacement values (3%-27%) seen in table 2 could be a reflection of different mutation rates, selective pressures, or some combination of both effects. If mutation rates are the main determinant of the replacement rates, then a comparison of synonymous site variation among the genes in table 2 should produce a range similar to that observed for amino acid replacements. $\chi^2$ Analysis of synonymous site substitutions indicates significant differences in these values among the 12 genes listed in table 2 ($\chi^2 = 227.1$, df = 11; P < 0.001). Therefore, if one assumes that synonymous site divergence reflects mutation rates at these loci, the range of amino acid divergences in table 2 may be at least partly explained by different mutation rates for each of these genes. However, if different mutation rates, as measured by synonymous site changes, drive the evolution of amino acid sequences, we would expect to find a correlation between synonymous and nonsynonymous substitutions among the loci in table 2. At best, the correlation is weak. Furthermore, differences in substitution rates between the exons of individual loci, as well as unusual codon biases in certain genes, may frustrate any attempts to correlate synonymous and nonsynonymous substitution rates between different loci (Riley 1989; Sharp and Li 1989). Thrcforc, it sccms likcly that variation in selective pressures, as well as differences in mutation rates, account for the wide range of sequence divergences seen in table 2.
**Fig. 6.**—An alignment of the EST 6 and EST 5B amino acid sequences. Amino acids which are identical in the two sequences are surrounded by a light-lined box. Insertions/deletions are the same as those shown in fig. 5. The putative N-terminus signal sequences are surrounded by a heavy-lined rectangle, and potential N-linked glycosylation sites are indicated by asterisks. Conserved cysteines are indicated by black dots, and the three conserved residues believed to play a role in catalysis are circled and are in boldface letters. The 16 amino acids that are polymorphic among alleles of EST 6 in *Drosophila melanogaster* are enclosed by squares. None are in the signal sequence. The triangle (Δ; at position 466) shows the location of the intron in the glycine codon.

The figures in table 2 probably underrepresent the actual substitution levels in those genes, because the *melanogaster* and *obscura* species groups are so distantly related. Many of the interspecific differences as well as some of the similarities between the homologous genes may actually represent multiple hits at those nucleotide positions. The high level of nonsynonymous site substitution between *Est-SB* and *Estd* (table 1) makes it fruitless to try to estimate the actual number of substitutions between these two genes by using a multiple-substitution correction formula (Lewontin 1989). Thus, the range of substitution values in table 2 underscores the importance of using sequence data from multiple genes when trying to estimate species divergence times and construct phylogenies.

The high nucleotide substitution and amino acid replacement levels in *Est-5B/*
6 may reflect low selective constraints acting on both synonymous and nonsynonymous sites in this locus in one or both species. Although the similar hydropathy profiles, conserved cysteine residues, and prevalence of conservative amino acid changes argue that there has been an overall conservation of protein structure between EST 5 and EST 6, the presence of duplicated genes in both species could mean that functions formerly fulfilled by a single enzyme are now being shared by two proteins in D. melanogaster and by three proteins in D. pseudoobscura. This implies that any one of these proteins is now under less functional constraint than before the duplication or triplication took place and can, therefore, tolerate higher levels of amino acid replacements as well as synonymous substitutions. Alternatively, rather than becoming functionally divergent, the multiple esterase genes in each species might still retain their original functions. Mutations that would have been eliminated by purifying selection in the ancestral unduplicated gene are now tolerated because the other proteins encoded by these loci can compensate for the reduced enzymatic efficiency of the altered esterase. This hypothesis seems unlikely, because both our analysis of D. melanogaster flies transformed with each of the D. pseudoobscura genes and studies on D. melanogaster flies that are null for Est-6 but not for Est-P, the paralogous neighbor
Fig. 7. Nucleotide and amino acid differences between Est-5B and Est-6 in nonoverlapping segments of 30 bp. Black bars are nucleotide differences; white bars are amino acid differences. Shaded regions below the histogram represent exons of the mature protein; the lighter region represents the signal peptide.

The presence of multiple related genes in each species also raises the possibility that the genes compared in the present study may actually be paralogous—rather than orthologous—homologues. Such a possibility could explain the high level of divergence of Est-6 (Collet et al. 1990; R. W. Phillis, K. M. Nicolsn, and R. C. Richmond, personal communication), indicate that the duplicated genes in both species do not appear to encode redundant products. Furthermore, the divergent expression patterns of EST 5 and EST 6 also suggest a functional diversification.

Fig. 8.—Hydropathy profiles of EST 5B in Drosophila pseudoobscura and of EST 6 in D. melanogaster, calculated using the Kyte and Doolittle (1982) algorithm with a window of nine amino acids. Hydrophobic regions are above the baseline; hydrophilic regions are below the baseline.
Cloning of the Esterase-5 Gene

FIG. 9.—Similar sequences in the 5′-flanking DNA of Est-5B and Est-6. Regions of 5′-flanking sequence similarity identified by a dot-matrix comparison (data not shown) were manually aligned. Numbers in parentheses indicate the beginning and end of each sequence, relative to the start of translation for Est-5B or Est-6.

between Est-5B and Est-6. However, sequence data for the other two genes in D. pseudoobscura and for Est-P in D. melanogaster reveal comparable levels of interspecific divergence in all pairwise comparisons of the D. pseudoobscura and D. melanogaster genes (data not shown). It is also conceivable that the orthologue for Est-5B or Est-5B has been deleted from one of these species. Although we cannot definitively disprove this hypothesis, similarities in the protein products of Est-5B and Est-6 and the fact that Est-B and Est-6 are the only genes in this family that are abundantly expressed in adult flies suggest that these two genes are orthologous.

The different substitution levels seen across regions of Est-5B and Est-6 could be attributed to regional differences in functional constraints. Although, the 5′-flanking sequences contain blocks of conservation that may represent regulatory regions required for basal gene expression, they also display a high level of overall diversity that could easily account for the dissimilar expression profiles of EST 5 and EST 6. While the Est-5B and Est-6 5′-flanking sequence data cannot be utilized to identify those substitutions that underlie interspecific differences in gene expression, similarities in the two 5′ sequences may be useful for identifying regulatory elements that control common features of Est-5B and Est-6 expression (Blackman and Meselson 1986; Wilde and Akam 1987).

The region of the first exon represented in the mature enzyme shows the highest level of conservation among the DNA regions that we examined (table 1). It has been proposed that functionally more important parts of a protein accumulate substitutions more slowly than do less important regions (Kimura and Ohta 1974). Therefore, the higher conservation seen in the first exon most likely stems from the presence of the active site residues. Myers et al. (1988) have identified, in conserved regions of both EST 6 and several other esterases from different organisms, three noncontiguous residues (arginine, aspartate, and serine) that they suggest may be involved in substrate cleavage. The aspartate and serine residues are also a common feature of serine proteases, which, like esterases, are members of a serine hydrolase superfamily of enzymes (Augustinsson 1968; Price and Stevens 1982, pp. 172–175). However, most serine
proteases utilize a histidine rather than an arginine as the third member of the catalytic triad. With this caveat in mind, we can identify three residues (arginine 159, aspartate 181, and serine 209) in the first exon of EST 5 that are conserved relative to EST 6 (fig. 6) and that may play a role in substrate binding or cleavage.

The neutral theory predicts that regions of the genome that evolve at high rates as shown by interspecific sequence comparisons should also be highly variable within a species (Kimura 1983; Hudson et al. 1987). Therefore, one can test whether observed patterns of variation differ from neutral expectations by comparing substitution levels in two separate regions of a gene within and between species and by looking for statistical heterogeneities in a contingency table, as suggested by Kreitman and Aguadé (1986). Table 3 shows the numbers of total and nonsynonymous site nucleotide substitutions in exons I and II of Est-5B and Est-6 and compares those figures with nucleotide polymorphisms among alleles of Est-6 in D. melanogaster. We chose to look at nonsynonymous site differences as well as at total substitutions because saturation effects in synonymous sites might obscure any deviations, in the total substitution levels, between the intra- and interspecific comparisons. There are no statistical heterogeneities in either total or nonsynonymous site changes ($\chi^2 = 0.28, \text{df} = 1, P = 0.96$; and $\chi^2 = 0.018, \text{df} = 1, P = 1.0$). These results are somewhat surprising in light of our prior observation that, in the interspecific comparisons between Est-5B and Est-6 but not in the intraspecific analyses among alleles of Est-6 (Cooke and Oakeshott 1989), there are significantly more nonsynonymous site substitutions per site in exon II than in exon I. Therefore, the results of the Kreitman and Aguadé test do not allow us to reject the null hypothesis that the majority of the differences between EST 5 and EST 6 are due to mutation and drift. That is, the divergence of these loci can be explained by the neutral theory of evolution. An alternative interpretation of this result is that selection is acting on EST 5 and EST 6 in similar ways in the two species. Support for such a contention is provided by similarities in the hydropathy profiles of EST 5 and EST 6 (fig. 8) and by the observation that half of the intraspecific amino acid polymorphisms among alleles of Est-6 correspond to sites of interspecific amino acid replacements between Est-5B and Est-6 (fig. 6). However, this interpre-

Table 3

<table>
<thead>
<tr>
<th></th>
<th>Exon I</th>
<th>Exon II</th>
</tr>
</thead>
<tbody>
<tr>
<td>Interspecific:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>350/1,324</td>
<td>81/248</td>
</tr>
<tr>
<td>Replacement</td>
<td>143.5/1,031</td>
<td>44/192.3</td>
</tr>
<tr>
<td>Intraspecific:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>38/1,324</td>
<td>7/248</td>
</tr>
<tr>
<td>Replacement</td>
<td>12/1,025.7</td>
<td>4/200.2</td>
</tr>
</tbody>
</table>

* Positions containing insertions/deletions were not counted in interspecific calculations.

* Data are from Cooke and Oakeshott (1989).
Cloning of the Esterase-5 Gene 543

tation conflicts with the hypothesis that EST 5 and EST 6 have different functions and would thus be expected to experience different selection pressures.

No similarity could be detected in the 3' flanking DNA of Est-5B and Est-6. This region in Est-6 probably contains transcriptional regulatory sites for the adjacent gene, Est-P, whose reading frame begins 193 bp after the stop codon in Est-6 (Collet et al. 1990). With Est-5B, on the other hand, the downstream gene, Est-5A, is much farther away (>1 kb). Therefore, other than providing sequences necessary for polyadenylation and transcription termination (Birnstiel et al. 1985) or, possibly, for RNA stability (Brawerman 1989), the constraints on the 3' flanking region in Est-5B probably do not coincide with those on the homologous region in Est-6. The apparent maintenance of two polyadenylation signals despite the lack of extensive sequence identity in these regions suggests that the synthesis of two alternate transcripts may be functionally significant.

In summary, our data show that Est-6 and Est-5B are diverging largely in accord with neutral expectations. However, the different expression patterns and quaternary structures of Est-5B and Est-6 may well indicate unique selective forces acting on one or both of these homologues. Evidence for these selective forces is not seen in the sequence data, probably because of the high levels of neutral substitutions that may obscure any selectively maintained differences between these genes. Analyses of Est-6 and Est-5B homologues from species more closely related to D. melanogaster (e.g., D. simulans and D. tesseria) and to D. pseudoobscura (e.g., D. persimilis and D. miranda) will allow tests of this hypothesis.

Acknowledgments

The authors wish to thank W. Anderson and J. Hall for sharing unpublished data, M. Ashburner for providing information on Drosophila codon usage, M. Brennan for donating the ry506 fly stock, D. Cavener and C. Langley for providing the D. pseudoobscura EMBL4 library, R. Hudson for advice on statistical calculations, and R. C. Lewontin for providing the several lines of D. pseudoobscura. We are also grateful to K. Nielsen and E. Snell for helpful advice and discussion. Two anonymous reviewers and W. Fitch made numerous suggestions that substantially improved the manuscript. This work was supported by a National Institutes of Health predoctoral genetics training fellowship to J.P.B. and by grants from the National Institutes of Health and the National Science Foundation to R.C.R.

LITERATURE CITED


Schaeffer, S., and C. F. Aquadro. 1987. Nucleotide sequence of the Adh gene region of


WALTER M. FITCH, reviewing editor

Received June 6, 1990; revision received July 2, 1990

Accepted July 13, 1990