The Role of Gene Conversion in Determining Sequence Variation and Divergence in the *Est-5* Gene Family in *Drosophila pseudoobscura*

Lynn Mertens King

Department of Biology, University of Miami, Coral Gables, Florida 33124

Manuscript received January 10, 1997
Accepted for publication September 26, 1997

ABSTRACT

Nucleotide sequences of eight *Est-5A* and *Est-5C* genes corresponding to previously sequenced *Est-5B* genes in *Drosophila pseudoobscura* were determined to compare patterns of polymorphism and divergence among members of this small gene family. The three esterase genes were also sequenced from *D. persimilis* and *D. miranda* for interspecific comparisons. The data provide evidence that gene conversion between loci contributes to polymorphism and to the homogenization of the *Est-5* genes. For *Est-5B*, which encodes one of the most highly polymorphic proteins in Drosophila, 12% of the segregating amino acid variants appear to have been introduced via gene conversion from other members of the gene family. Interlocus gene conversion can also explain high sequence similarity, especially at synonymous sites, between *Est-5B* and *Est-5A*. Tests of neutrality using interspecific comparisons show that levels of polymorphism conform to neutral expectations at each *Est-5* locus. However, McDonald-Kreitman tests based on intraspecific gene comparisons indicate that positive selection on amino acids has accompanied *Est-5* gene duplication and divergence in *D. pseudoobscura*.

THE *X*-linked *Esterase-5* locus in *Drosophila pseudoobscura* is one of the most polymorphic allozyme loci in Drosophila (Lewontin and Hubby 1966; Coyne *et al.* 1978; Keith 1983). As such, there has been considerable interest in addressing the adaptive significance of *Est-5* variation, especially in reference to the high frequency allozyme variants (Yamazaki 1971; Árnason 1982, 1991; Keith 1983). However, studies of allozyme variants are not likely to be appropriate for *Est-5*, because members of a single protein electrophoretic class may be heterogeneous in amino acid composition and thus an assemblage of "true" alleles (Veuille and King 1995).

A molecular characterization of the *Est-5* gene region revealed three closely linked genes called *Est-5C*, *Est-5B*, and *Est-5A* (arranged 5′ to 3′; Brady *et al.* 1990). An analysis of gene expression showed that *Est-5A* is transcribed in the third instar larvae, *Est-5B* is expressed in adults of both sexes and is the structural locus for the major adult EST5 protein (hereafter called EST5B), and *Est-5C* expression was not detected (Brady *et al.* 1990). Brady and Richmond (1990) proposed an evolutionary history of the *Est-5* gene duplications in *D. pseudoobscura*, with reference to *Est-6* and *Est-P* in *Drosophila melanogaster*, based on comparisons of nucleotide sequences, patterns of gene expression, and properties of the enzymes. They propose that the first gene duplication predated the divergence of *D. pseudoobscura* and *D. melanogaster*, and gave rise to the *Est-5A–Est-P* lineage and to the *Est-5B/C–Est-6* lineage. A second duplication in the *D. pseudoobscura* lineage gave rise to *Est-5B* and *Est-5C*. Based on this scenario, there is a lower than expected level of sequence divergence between *Est-5A* and *Est-5B* (17.5% nucleotide and 18.6% amino acid differences), compared with the orthologous *Est-P* and *Est-6* loci (32.7% nucleotide and 35.8% amino acid differences), which is attributed to gene conversion (or reciprocal recombination) between *Est-5A* and *Est-5B* (Brady and Richmond 1990).

Many studies have now shown that members of multigene families do not evolve independently, and various mechanisms of homogenization, including unequal crossing over and gene conversion, have been proposed to explain the concerted evolution of the gene family members (Arnheim 1983). Although gene conversion is considered to be a homogenizing mechanism, there is evidence that gene conversion can also generate variability among members of multigene families (Xiong *et al.* 1988; Kuhner *et al.* 1991; Wines *et al.* 1991; Ohta 1992, 1995).

Although the evolutionary history of the *Est-5/6* gene family shows evidence of gene conversion (and/or reciprocal recombination) and concerted evolution, it is unclear if interlocus gene conversion generates genetic variability in the gene family members, and especially if this mechanism generates *Est-5B* sequence variation and amino acid polymorphism in *D. pseudoobscura*. Previous work indicates that sequence variation

Address for correspondence: Lynn Mertens King, Medical Genetics Branch, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD 20892.

among members of different EST5B protein electrophoretic classes does not deviate from neutral expectations, suggesting that the considerable amino acid polymorphism is selectively neutral (Veuille and King 1995). Thus, an analysis of sequence variation including *Est-5A* and *Est-5C* is likely to provide a more complete picture of the evolutionary forces and molecular mechanisms influencing polymorphism and divergence at *Est-5B*.

In this study, eight *Est-5A* and *Est-5C* alleles corresponding to previously sequenced *Est-5B* alleles were sequenced in *D. pseudoobscura*, and the three genes were sequenced in *Drosophila persimilis* and *Drosophila miranda* for interspecific comparisons. The goals of this study were to describe patterns of polymorphism and divergence in this gene family and to examine if gene conversion contributes to sequence variation, especially to the highly polymorphic *Est-5B* locus in *D. pseudoobscura*. The data also allow examination of amino acid divergence, which may accompany functional divergence of the duplicated genes. The interspecific comparisons allow tests of neutrality and examination of putative gene conversion tracts within a phylogenetic context.

MATERIALS AND METHODS

Sampling: *D. pseudoobscura* isofemale lines were established by Keith (1983) from collections made in 1979 in the James Reserve in the San Jacinto Mountains in southern California and near the Gundlach-Bundschu Winery in the Sonoma Valley of northern California, and they were maintained since then in the laboratory. Nucleotide sequences of *Est-5A* and *Est-5C* were determined from lines representing eight different EST5B protein electrophoretic classes: three lines from the James Reserve (J3, J5, and J10) and five lines from the Gundlach-Bundschu (G2–G6) populations. Line G3 represents the most common EST5B electrophoretic class. The *Est-5B* lines were chosen originally to characterize the nature of electrophoretic classes, and they are a nonrandom population sample (Veuille and King 1995).

Cloning and sequencing: The *Est-5* genes in *D. pseudoobscura* were isolated from λZAPII (Stratagene, La Jolla, CA) subgenomic libraries and were constructed by cloning 8- or 3-kb *Eco*RI restriction fragments that include the *Est-5C* and *Est-5B*, and *Est-5A* gene regions, respectively. A *D. persimilis* genomic library in λEMBL3 and a *D. miranda* genomic library in λEMBL4 were provided by R. Norman. Clones were isolated by plaque hybridization using *D. pseudoobscura Est-5* clones provided by J. Brady. The clones were purified, and the three gene regions were individually subcloned into either pUC19 or pBSKS- (Stratagene) using standard procedures (Sambrook *et al.* 1989).

Either plasmids or PCR-amplified templates were sequenced using oligonucleotide primers designed from published sequences (Brady and Richmond 1992). Plasmid templates were manually sequenced using Sequenase version 2.0 (United States Biochemical, Cleveland, OH). PCR templates were amplified using a thermal cycler (MJ Research, Watertown, MA); the reaction components 1× rTth buffer, 240 μM dNTPs, 5U rTth polymerase (Perkin Elmer, Norwalk, CT), 50 nM primers, 625 ng genomic DNA, 1.25 mm $MgCl_2$; and the reaction profile 30 cycles of 94° 30 sec, 54° 1 min, 72° 2 min, followed by 72° 5 min, and 4° hold. The PCR products were purified with spin columns (Centricon 100; Amicon, Beverly, MA), and ∼300 ng of template was used in the automated sequencing dye termination reaction (model 373A; Applied Biosystems, Foster City, CA). Complete and overlapping coverage was obtained in both directions for all sequences.

Sequence analysis: Sequences were assembled using the Gap and Pretty programs of the University of Wisconsin Genetics Computer Group (Devereux *et al.* 1984) or the Genetic Data Environment programs (Smith *et al.* 1994). The sequences have the following accession numbers in the GenBank database: *D. pseudoobscura Est-5A*, AF016135–AF016142; *D. pseudoobscura Est-5C*, AF016143–AF016160; *D. persimilis Est-5C* and *Est-5B*, AF016110; *D. persimilis Est-5A*, AF016111; *D. miranda Est-5C* and *Est-5B*, AF016109; and *D. miranda Est-5A*, AF016108. Sequence alignments were made using ClustalW (Thompson *et al.* 1994), followed by manual adjustments based on amino acid alignments. The alignments included *Est-6* and *Est-P* of *D. melanogaster* (GenBank accession numbers M33780 and M33781, respectively). Nucleotide diversity and the number of net nucleotide substitutions per site between populations (loci) were estimated by the method of Nei (1987) using the computer program DnaSP, version 2.0 (Rozas and Rozas 1995). Estimates of nucleotide substitution were made using the Jukes-Cantor correction, and numbers of synonymous and nonsynonymous sites were estimated by the method of Nei and Gojobori (1986) using the computer program MEGA (Kumar *et al.* 1993). Alignment gaps were excluded in pairwise comparisons.

Phylogenetic analysis: The genealogical relationships of genes and alleles were estimated using maximum parsimony (PAUP; Swofford 1992) and by the neighbor-joining method using the Jukes-Cantor distance estimation in MEGA (Kumar *et al.* 1993). For the parsimony analysis, heuristic searches with 100 random addition replicates, TBR branch swapping, and MULPARS options were invoked. Strict consensus trees were constructed from the multiple equally parsimonious trees. The tree topologies were evaluated by 100 bootstrap replicates. Three data sets were analyzed: coding regions, 5′ flanking regions, and 3′ flanking regions.

RESULTS

Nucleotide sequence variation: Figure 1 shows the location of the polymorphic nucleotide sites and the interspecific differences in the total region sequenced of each *Est-5* gene. *Est-5A* is the only one of three genes that shows length variation in the coding region both within and between species. In *D. pseudoobscura*, *Est-5A* is polymorphic for a CTA (Leu) deletion from 73 to 75 bp (Figure 1A), relative to *Est-5A* in *D. persimilis* and *D. miranda*. CTA is duplicated in this region, and the

Figure 1.—Polymorphic sites in *D. pseudoobscura Est-5*, including sequences of *D. persimilis* (per5) and *D. miranda* (mir5). Dots indicate sequence identity, and dashes indicate deletions compared with line J5. The numbering of sites above the sequence is relative to the initiation codon, which begins with +1. Domains of the gene regions are noted above the numbered nucleotide sites. (A) *Est-5A*, (B) *Est-5B*, and (C) *Est-5C*. I, intron regions.

A

```
      |--------- 5' ------------||-------------- EXON I -------------------------------|| I||-- EXON II --||-----------3'---------------|
               (-)                                                                      11111111111111111111111111111111111111111111111111111
      33222221111111111111            112233333344566677777888888899999990000001111223333344444555555556666777777777777777888888888889999
      2099221776333310000099966111477729250335671153781557990044691578881268895788090245645778011223485899011233366667789900133455590127
      5262320866765425432109875383933453582603768459972567857562331913057870349826589305838005848573601422231595023813485716716213656956389
   J5 CA-GCCTCTC----GTGGTGGAAT-ATCTGCTATACACCGGTAAGTGGTGGGGAAGTGTGGTAAATCCGGTGCCCTGGGTCTTTGGCTCCCCTCCGCGGGTGTCATGGCGTTACTATGCCGCATAAATGTCAT
   J8 ..........................AC..............C.CG...TG..TA.A...T..A....T.........A.AC...G.......T.T.........--..C..A..AA...........C.GC
  J10 ...G.....................TC.AC.........TA..CAC.CG...TG.GT.................A...A.A.....T.........G.....--..........AA...........C.GC
   G2 ...G.....A....A...........ACA---.GT.........C..G.A.T.GC.T...............C.....A.A..A....A.....TA.........--.....A..AA...........C.G.
   G3 ....TT...A................ACA---.G..........C.CG....G..T.............T..C.....A.A................A.......GC......................TGC
   G4 ...G.....................C...A---.G..T.T.....C.CG....G..T..........A.A.T..C.....A.A.........C.....A.T.....--.....A.AAA.CAG.G...A..G.
   G5 ...G.....................C.........G.........C..G....G..T...A......A...T.A.....A.A......................--.....A..AA.......G..C.GC
   G6 ....TT...A................ACA---.G..........C.CG....G..T.............T..C.....A.A................A.......GC......................TGC
per5A T.TT..C..................GACA........T....CAC..GT....G..T....TC........T......C.A.A................A.T.....--.....A..AA....C.....CT**
mir5A .A.....GC.TGTA.----------...ACA...C..G...ACCCACT.G..A..G....A.....GA..A.T.T.CA.A.AAGAAT.C.TA..AGC.A....CATAC--CC.GTAG.AAG...C.C.G.****
                                                                                                                |----3'region of mir5A-----|
```

B

```
      |----------- 5' -------------||------------- EXON II ----------------------------------------------------||--||---EXON II--||----- 3' -----|
                 (-)                                                                                       111111111111111111111111111111111111111111111112
      444444444433333333222222221111                  222222233333444445556666677788888999999999999900000000111123333334444444555555566666777777788899990
      665211000099888732755442333393112555677934447782248804880692789912623557133444577890123334785678001369900012378226677890556623455577589955660
      543442964097743690987905876517011347806616785854983454018411370676281231838289189168105895096768321296825720179464759852130350514834852809894
   J5 ACAGGCGAAGCCCATTTGACAGCCCTCGCAACACCGCGCTCTCTAGTAGCGTCCGTGGGCTATTCCCAGCCGTCGATGGCCAGGGGGGCATCTCGTCAGCAACGGGAACACTAAGATAATGGGCGACTACACATCAACTTG
   J8 .......G..A.AC..................G....TT....GAC..........AC......CC...................A....A....G.............................C...A....T...C.C.....
  J10 ...........-AC..................G..............A.......TAC......CC...........................G...............G......CTG.CAAT.G...A.....C.....
   G2 .............C.C.....A.A----..G.....TT.........TAG...TA......CCC.G..A....TC.AAC.A....C......G......A.........G......CTG.C........T..G..C.....
   G3 ......A......C.C...................G.C.....A...T....TAG.GA...A..TG.CC........T...................GT...........................C........T..........
   G4 .....A.......C.C...............-.G.C................TAG...T..A.....CC......A...................G...A.............T..G......CT.......A--...C.....
   G5 .....G.......CCC..G.............G....TT.....T..........A.AC.A...G.CCAGT...G...................A....G.............A..T..............C.......GT...........
   G6 ............AC...A............CG.CTT....C......CTAG.G.T.........CC.................G.A........G....................G..C....C.A......T.....C.....
per5B C.........A.AC.................G.C.........TCG.........AC..AAG.CC..G.G.......GA.C.......T.TGG.G....G..C..T.......C..A.....C...............C...T
mir5B .TCTA...T-..AC..AC.TT.T......GAC.......T.........GTG.TAC....G.CC......A..A..GAACA....A.T....TG.ACA.A..CT.T.GC.......C....AC.....T-..TG..A.CGCGT
```

C

```
      |-- 5'-||--------------------- EXON I -----------------------------||I|EXON II|-------- 3' --------------|
          (-)                                                 1111111111111111111111111111111111111111111111111122
      211            1111111112223444455566677788888899999900122222233333333444555566777777777778888888999999900
      0208541 1133562222347899179900470022393491116770122676711444581135678890024478372233455777791345667047999911
      006736382703463452449289630287719810317734672309149531725891117866716613228016089077362378781272960103679 12
   J5 CCGGTAG-CTGTCCTTTCCATCGCTTCAATTAGACGCAACCTATCGAGATGTCGCGGTATAACTCGAGTGTGTTGGTTTAAAGGTTAA--GTTCGCGAATGTGCACA
   J8 .T..C..-T..........CGT.C.TGG..TAC....CTTC............T.A..A...........TC.....A....A....--.....AC.G.TC.....
  J10 ....C..-..........GTCG..CC.GG.C.........TC..GC....................A.T............C...AA.....AC..........
   G2 .......A......CAAT.....TC..GG.C...G...................A..AA....G.CG..A...T......A..........--.....AC.....A....
   G3 ....C..-...GT...........TC..GG.C.......................A...A..A...........T......A..G..C...--.....AC.G......A.
   G4 ....C..-..........GTCG..C..GG.C.........TC............................AC.GA.C...--.C...AC.G........
   G5 .T-.C..-..T.TT...........CC.GG.C...................TA......T................T......A.....C...--................T.T
   G6 .......-.G....CAAT..C...CC..G..TA.......T.........C............................A.........--.....A..........
per5C ....CG.-T...TT...........CC.GG..T.C.C...........T..A..A.T.A...GT...TT....T..C...A.C.A.C...--.....A..G.......
mir5C A..ACGC-.....T...........C.GGCC.....AG.T..GG..T..AAC.AT..CG..GT...T.C.GT.CCCCCA.C.A.CCTT--A.CGTACG.C...G...
```

```
A
        1222222222333333334444444555555555
      123223566688012223564555899144555555
      559689306918775790522016702069012345
   J5 VLGSLAAVNVSRSQSKSQSVFSDQRDEDC.......
   J8 ...TV...S.I....M.K..Y.EH............
  J10 ..STV...SGI.........Y.E.....G.......
   G2 I-..V..LSLI.........Y.E.S...........
   G3 I-.TV...S.I.........Y.E.............
   G4 I-.TV...S.I.......T.Y.E.....F.......
   G5 ....V...S.I.T.......Y.E.............
   G6 I-.TV...S.I.........Y.E.............
per5A I.S.VS..S.I..HR.....Y.E.....F.......
mir5A I.S.V.T.S..Q....A..LYAE..EQN.QRCLIFF
```

```
B
              1112223333333444445555
       12228991272890112458457990113
      483463230805516377976474017463
   J5 TSADLLQGRGDLRADSVQKLYDKIIINTVR
   J8 A.V..T..............D.........
  J10 A......E............D..M...A.Q
   G2 A.VY....K......TIK..DE.M...A..
   G3 A...M...K.NV.S......D.........
   G4 A.......K.N.........D........L
   G5 A.V........VG.......D.........
   G6 AF......K...........D....VT...
per5B A....SR....V......EWD.Q.T.....
mir5B E........V.V..E.I...D.......D.
```

```
C
           12223444555
       145974690156023
      6128804617761467
   J5 LSFEVGQTAAIAVVFV
   J8 .....D.M.......E
  J10 ...DA...G.......
   G2 ..N.........I..E
   G3 .R.............E
   G4 ...D...........E
   G5 ....A....T.....E
   G6 R.N.AD.M.......E
per5C ....A......S...E
mir5C ....A.RM..V..LLE
```

Figure 2.—EST5 amino acid polymorphism in *D. pseudoobscura*, including *D. persimilis* (per) and *D. miranda* (mir). Dots indicate sequence identity, and dashes indicate deletions compared to line J5. Numbering is relative to the initial Met. (A) EST5A, (B) EST5B, and (C) EST5C.

TABLE 1

Nucleotide diversity among eight *D. pseudoobscura* haplotypes

Locus	Region	N (bp)	S	k	π
Est-5A	5′ Noncoding	460	6	2.75	0.0060 ± 0.0010
	3′ Noncoding	377	15	5.43	0.0145 ± 0.0030
	Coding	1647	43	13.86	0.0084 ± 0.0013
	Total	2540	64	22.29	0.0088 ± 0.0009
Est-5B	5′ Noncoding	404	10	3.11	0.0078 ± 0.0012
	3′ Noncoding	314	7	2.39	0.0083 ± 0.0013
	Coding	1638	62	20.89	0.0128 ± 0.0013
	Total	2411	81	26.89	0.0113 ± 0.0011
Est-5C	5′ Noncoding	231	2	0.96	0.0042 ± 0.0008
	3′ Noncoding	404	18	5.96	0.0148 ± 0.0027
	Coding	1638	45	15.96	0.0097 ± 0.0012
	Total	2333	68	24.11	0.0104 ± 0.0010

Total bp sequenced includes the intron regions: *Est-5A* (56 bp), *Est-5B* (55 bp), and *Est-5C* (60 bp). Estimates were made using DnaSP (Rozas and Rozas 1995). Estimates are of $3N\mu$. S, number of polymorphic sites; k average number of nucleotide differences; π, nucleotide diversity ± SE (Nei 1987).

polymorphism involves the presence or absence of one of the duplications. The polymorphism is in intermediate frequency, with half of the lines having the deletion. Based on an alignment of the three *Est-5* genes, the CTA duplication is present in *Est-5A* but not *Est-5B* in these species. *Est-5A* in *D. pseudoobscura* is presumably functional because no stop codons occur in the coding regions, and putative regulatory sequences are conserved in these eight lines; although no EST5A proteins have been identified, there is evidence that the gene is transcribed (Brady *et al.* 1990).

In *D. miranda, Est-5A* encodes a protein seven amino acids longer than in *D. pseudoobscura*, *D. persimilis*, and the putatively homologous *Est-P* gene in *D. melanogaster.* Thus, assuming that the shorter *Est-5A* gene is ancestral, the increase in gene length results from a T to C substitution at position 1701 that changes the UAG stop codon to a CAG (Gln) sense codon (Figure 1A). A UAG stop codon is present 18 nucleotides downstream from the CAG codon [from 1722 to 1724 base pairs (bp)], and the EST5A protein in *D. miranda* is extended by seven amino acids (Figure 2A).

Tables 1 and 2 summarize *Est-5* variation in *D. pseudoobscura* by gene region and class of site. The complete intergenic region between *Est-5C* and *Est-5B* was sequenced and was divided into two regions of equal length to compare 5′ and 3′ flanking regions. Comparisons of the level of polymorphism across genes at functionally different classes of sites show several significant differences in the patterns of variation.

For all three *Est-5* genes, noncoding sites are significantly less polymorphic than synonymous sites (*Est-5A*, $G = 17.70$, 2 d.f., $P < 0.001$; *Est-5B*, $G = 31.09$, 2 d.f., $P < 0.001$; *Est-5C*, $G = 21.49$, 2 d.f., $P < 0.001$). This may be a general pattern in Drosophila (Moriyama and Powell 1996). They show an average nucleotide diversity (π) at synonymous sites = 0.028 and noncoding sites $\pi = 0.017$ for five nuclear genes in *D. pseudoobscura*. The following estimates of *Est-5* variation are multiplied 4/3 to compare the *X*-linked *Est-5* genes to autosomal genes. Averaging over the three *Est-5* genes, $\pi = 0.039$ for synonymous sites, and $\pi = 0.013$ for noncoding sites. It appears that the nonrandom sample of *Est-5B* sequences does not inflate these estimates. Based on a random sample of 16 sequences, $\pi = 0.016$ for a

TABLE 2

Nucleotide diversity at synonymous and nonsynonymous sites in *D. pseudoobscura*

Locus	N (bp)	π
Est-5A		
Nonsynonymous sites	1250.52	0.0043 ± 0.0010
Synonymous sites	390.48	0.0217 ± 0.0042
Est-5B		
Nonsynonymous sites	1250.44	0.0059 ± 0.0012
Synonymous sites	384.56	0.0351 ± 0.0053
Est-5C		
Nonsynonymous sites	1253.00	0.0035 ± 0.0009
Synonymous sites	382.00	0.0310 ± 0.0051

Nucleotide diversity ± SE was estimated by the method of Nei (1987) using MEGA (Kumar *et al.* 1993). Estimates are of $3N\mu$.

504-bp intergenic region between *Est-5B* and *Est-5C* in *D. pseudoobscura* (Babcock and Anderson 1996). This compares with $\pi = 0.015$ for the same (808 bp) intergenic region in this study. Thus, the *Est-5* genes show more variation at synonymous sites than other genes in *D. pseudoobscura*.

The 3′ flanking regions are significantly more polymorphic than the 5′ flanking regions summing across all three genes ($G = 8.80$, 1 d.f., $P = 0.003$), but if the regions are compared by gene, *Est-5A* and *Est-5C* show significant differences, but not *Est-5B* ($G = 0.046$, 1 d.f., $P = 0.830$). Considering the two intergenic regions, the 808-bp region between *Est-5C* and *Est-5B* is significantly more polymorphic than the 774-bp region (of the ~1100-bp region in total) sequenced between *Est-5B* and *Est-5A* ($G = 5.12$, 1 d.f., $P = 0.024$).

Each gene shows significantly different levels of polymorphism at synonymous and nonsynonymous sites (*Est-5A*, $G = 27.10$, 1 d.f., $P < 10^{-6}$; *Est-5C*, $G = 49.17$, 1 d.f., $P < 10^{-6}$; *Est-5C*, $G = 51.95$, 1 d.f., $P < 10^{-6}$); however, the three genes have similar levels of polymorphism at synonymous sites ($G = 3.47$, 2 d.f., $P = 0.176$) and nonsynonymous sites ($G = 3.05$, 2 d.f., $P = 0.218$). Although the genes show similar levels of polymorphism, estimates of nucleotide diversity, the average pairwise number of differences per nucleotide site, are lowest at *Est-5A* for synonymous sites and lowest at *Est-5C* for nonsynonymous sites, and both classes of sites show the highest nucleotide diversity at *Est-5B* (Table 2). The original nonrandom sample of *Est-5B* sequences will cause an upward bias in estimates of variation at nonsynonymous sites at this locus, but this is not expected to influence variation at synonymous sites.

The distribution of nucleotide polymorphism was tested for heterogeneity by the variance test of Goss and Lewontin (1996). This method measures the distance between polymorphic sites and compares the observed variance with the expected values. The test was applied to the following: (1) the coding plus intron regions of each gene, (2) the intergenic region between *Est-5C* and *Est-5B*, and (3) the intergenic region between *Est-5B* and *Est-5A*. All tests show a highly significant nonrandom spatial distribution of polymorphism. The observed variances of interval length for these regions are *Est-5A* = 0.00567, $P < 0.001$; *Est-5B* = 0.00106, $P < 0.001$; *Est-5C* = 0.00545, $P < 0.001$; intergenic *Est-5C* and *Est-5B* = 0.01277, $P < 0.001$; and intergenic *Est-5B* and *Est-5A* = 0.05363, $P < 0.001$.

Amino acid variation: Figure 2 shows the amino acid polymorphisms in the EST5 proteins, which were determined from the nucleotide sequences. EST5A has 3.1% amino acid polymorphism, and the proteins differ by an average of 5.8 amino acids (Figure 2A). EST5B has 4% amino acid polymorphism, and the proteins differ by an average of 8.9 amino acids (Figure 2B). All 16 EST5B amino acid sequences show 6.1% polymorphism, and they differ by an average of 7.7 amino acids (Veuille and King 1995). EST5C has 2% amino acid polymorphism, and the average number of amino acid differences among the sequences is 3.8 (Figure 2C).

Tests of gene conversion: The method of Betrán *et al.* (1997) was used to detect gene conversion events between the *Est-5* loci. Their method uses the relative frequency of a nucleotide at a site to determine if the site is informative of a conversion event between two groups of sequences. A segregating nucleotide is informative if its relative frequency in a group of "converted" sequences is 20% or less and its relative frequency in the group of "converting" sequences is three or more times higher than in the group of "converted" sequences. The two outermost informative sites determine the length of the observed conversion tract. Conversion tracts of 1 bp in length are not considered because they cannot be distinguished from parallel mutation events.

This method detected six interlocus gene conversion events (Table 3). In addition, visual inspection of the data showed that nucleotide sites 132–143 and 942–960 in *Est-5C* and *Est-5B*, respectively, have segregating nucleotides in higher frequency (25%) than considered by the method of Betrán *et al.* (which is based on a minimum of an informative nucleotide pair) that are shared with another locus. These nucleotides may also be interpreted as resulting from gene conversion rather than from parallel mutation events, and they are included in Table 3.

Considering information from interspecific comparisons, there are two ways to explain the *Est-5A* AG/CA haplotype variation at nucleotide sites 414–415. In *D. persimilis* and *D. miranda*, these sites are CA at all three loci, suggesting that these nucleotides are the ancestral state. Therefore, the low-frequency CA polymorphism in *Est-5A* can be explained either by unique mutations (AG) that have increased in population frequency plus the maintenance of ancestral variation (CA), or by the conversion of AG to CA by either *Est-5B* or *Est-5C*.

TABLE 3

Interlocus *Est-5* gene conversion in *D. pseudoobscura*

	Nucleotide position								
								1111	11111111
	1111	222	44	66	677	99	90	0000	33333333
	3334	555	11	34	900	44	56	2344	22222344
	2341	567	45	91	925	27	80	9824	03569517
Est-5A									
J5	CCAC	GCT	AG	**GC**	**TAT**	**TC**	**AC**	**CCCT**	ACTTTGCT
J8	....	...	..	..	...	..	..	....	..A.....
J10	....	...	CA*	..	...	..	..	....	..A.....
G2	....	...	..	..	...	..	..	....	..A.....
G3	....	...	..	..	...	..	..	....	..A.....
G4	....	...	..	..	...	..	..	....	..A.....
G5	....	...	..	..	...	..	..	....	..A.....
G6	....	...	..	..	...	..	..	....	..A.....
per-5A	....	GAT	CA	..	...	..	..	....	..A.....
mir-5A	....	...	CA	TC	...	..	..	....	..A.....
Est-5B									
J4	**CAAT**	TCT	**CA**	TG	CAC	CG	GG	GCCG	**GTACACCC**
J1	....	...	..	..	...	TC*	AC*	....	....
J2	....	...	..	GC*	...	..	..	....	....
J3	....	GAC*	..	..	...	TC*	AC*	....	....
J6	....	.T.	A.	..	...	..	..	....	....
J7	....	...	..	..	...	..	..	CCCT	....
J9	....	...	..	..	...	..	..	....	....
J5	....	...	..	..	TAT	..	..	....	....
J8	....	GAC*	..	..	...	..	..	....	....
J10	....	.T.	..	..	...	..	..	....	....
G2	....	...	..	..	...	TC*	AC*	CCC.	......A.
G3	....	...	A.	..	...	..	..	....	....
G4	....	...	..	..	...	..	..	....	....
G5	....	...	A.	..	...	..	..	....	....
G6	....	...	..	..	...	..	..	....	....
G1	....	...	..	..	...	TC*	AC*	....	....
per-5B	....	.TC	..	..	...	..	.C	...T	......A.
mir-5B	....	...	..	..	...	..	AC	...T	A.....A.
Est-5C									
J5	TTTC	**GAC**	**CA**	**GC**	TCC	CG	GG	CAGG	ATACACCT
J8	....	...	..	..	...	..	..	....	....
J10	....	...	..	..	...	..	..	....	....
G2	CAAT*	...	..	..	...	..	..	....	GTACACCC
G3	....	...	..	..	...	..	..	....	....
G4	....	...	..	..	...	..	..	....	....
G5	....	...	..	..	...	..	..	....	....
G6	CAAT*	...	..	..	...	..	..	....	....
per-5C	....	...	..	CC	...	..	..	....	T......T
mir-5C	....	...	..	..	.T.	..	..	....	...T...T

Sites associated with an amino acid change are indicated with an asterisk. The sites of the converting locus are in bold, and the converted nucleotides are in shaded boxes. Nucleotides identical to the first sequence are indicated by a dot. per, *D. persimilis*; mir, *D. miranda*.

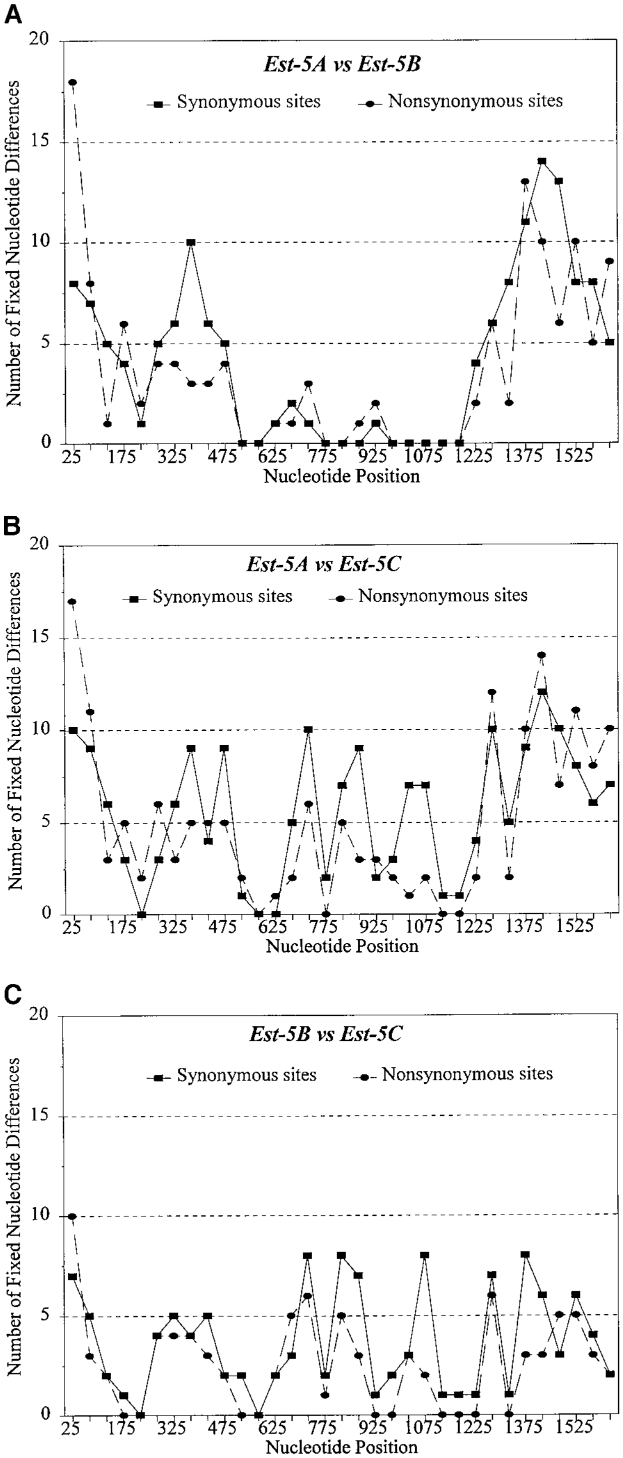

Figure 3.—The number of fixed differences between *Est-5* genes at synonymous and nonsynonymous sites in 50-bp intervals along the coding sequence. (A) *Est-5A vs. Est-5B* (B) *Est-5A vs. Est-5C*, and (C) *Est-5B vs. Est-5C*.

The observed gene conversion tracts between nucleotides 639 and 1044, where *Est-5A* putatively converted *Est-5B*, are coincident with a region of few fixed nucleotide differences between these two genes (Figure 3A). For example, in the 500–1200-bp region, there are only 13 fixed differences (5 at synonymous sites, 8 at nonsynonymous sites) between *Est-5A* and *Est-5B*. This contrasts with 110 fixed differences in the first 500 nucleotides (57 at synonymous sites, 53 at nonsynonymous sites) and 140 fixed differences (77 at synonymous sites, 63 at nonsynonymous sites) in the last 447 bp of the coding region. This pattern of fixed nucleotide differences does not occur between *Est-5A* and *Est-5C* (Figure 3B) or between *Est-5B* and *Est-5C* (Figure 3C). The region of few fixed differences between *Est-5A* and *Est-5B* does not correspond to a region of low polymorphism in either gene, so it does not seem likely that constraint on sequence divergence is maintaining the similarity.

The patterns of fixed differences between the *Est-5* genes in *D. pseudoobscura* are similar to patterns of divergence between the *Est-5* genes in *D. miranda* and *D. persimilis* (not shown). However, *Est-6* and *Est-P* in *D. melanogaster* do not show the same pattern of divergence as their putative homologs (*Est-5B* and *Est-5A*, respectively; Brady and Richmond 1992). Three regions where there are no fixed differences between the *Est-5* genes in all pairwise comparisons, centered on intervals 575, 1125, and 1175 bp, suggests that the lack of divergence is related to functional constraint. Two of these regions are near but not entirely coincident with amino acid residues Ser-210 and Glu-340 (the positions are based on alignment of the three EST5 proteins), corresponding to codons at nucleotide sites 628–630 and 1028–1030, respectively, which are thought to be involved in the catalytic function of the enzyme (Karotam *et al.* 1993). The region encompassing the third residue of a proposed catalytic triad, His-470, corresponding to nucleotide sites 1408–1410, is not conserved.

Thus, the length of the region of similarity between *Est-5A* and *Est-5B* may be explained partly by functional constraint but perhaps mostly by a single gene conversion (and/or reciprocal recombination) event that predates the divergence of *D. pseudoobscura*, *D. persimilis*, and *D. miranda*. The accumulation of unique polymorphisms and fixed differences in this region is also evidence that the event was not recent. The putative converted *Est-5B* regions in *D. pseudoobscura* may then be remnants of one old conversion event that have been reshuffled by interallelic recombination.

The lengths of the observed converted gene regions in Table 3 range from 2 to 28 bp, or up to 405 bp if the region between nucleotide sites 639 and 1044 is considered to result from a single event between *Est-5A* and *Est-5B*. The lengths of the true gene conversion tracts are difficult to estimate. From the model of Betrán *et al.* (1997), the estimates of true tract length are 10 bp for *Est-5A* and *Est-5B* and 16 bp for *Est-5B* and *Est-5C* (A. Barbadilla, personal communication). These estimates, however, are based on the assumption that conversion events have not been broken up by subsequent recombination events. This assumption does not appear to be valid for *Est-5*, which shows considerable intragenic recombination (Figure 1), so these estimates are not likely to be meaningful. In *D. melanogaster*, estimates of the mean length of gene conversion tracts within the *rosy* locus, based on intragenic recombination using strains with known molecular markers, is estimated to be 352 bp (Hilliker *et al.* 1994). Thus, it seems plausible that a single gene conversion even may account for the 405-bp region of *Est-5B* haloptype variation in *D. pseudoobscura*. Longer tract lengths may not be observed in *Est-5* because intragenic recombination would quickly reshuffle the polymorphisms, except for those that are very close together.

Tests of neutral molecular evolution: Tests of neutral molecular evolution applied to the *Est-5* data fail to reject the neutral model. Tajima's *D*-statistic (*Est-5A*, $D = -0.60$; *Est-5B*, $D = -0.73$; *Est-5C*, $D = -0.42$) is not significant for any locus, although the values of *D* are negative and suggest purifying selection (Tajima 1993). The *HKA*-test (Hudson *et al.* 1987), using the intergenic region between *Est-5C* and *Est-5B* as the reference locus and *D. miranda* for the interspecific comparison, yields nonsignificant χ^2 values: *Est-5A*, $\chi^2 = 0.269$; *Est-5B*, $\chi^2 = 1.094$; *Est-5C*, $\chi^2 = 0.375$, $P < 0.10$ (1 d.f.). Application of the test to all sites in the *Est-5* coding regions also failed to reject the neutral model, as did pairwise comparisons of synonymous sites between *Est-5* genes.

Finally, the ratios of nonsynonymous to synonymous polymorphisms in *D. pseudoobscura* (0.65, *Est-5A*; 0.51, *Est-5B*; 0.36, *Est-5C*) are not significantly different from the ratios of nonsynonymous to synonymous fixed differences between *D. pseudoobscura* and *D. miranda* (0.53, *Est-5A*; 0.40, *Est-5B*; 0.29, *Est-5C*; $P < 0.07$ for each gene comparison), based on McDonald and Kreitman's (1991) test. In contrast, the results of this test applied to between gene comparisons are significant for *Est-5A vs. Est-5B* and *Est-5A vs. Est-5C* (Table 4). There is an excess of fixed differences at nonsynonymous sites in these comparisons, which indicates selective amino acid divergence between *Est-5A* and *Est-5B* and between *Est-5A* and *Est-5C*.

Genealogical inference: The phylogenetic analyses were based on an alignment of 1695 nucleotide sites of the coding regions of the *Est-5* genes in *D. pseudoobscura*, *D. miranda*, and *D. persimilis*, and of the *Est-6* and *Est-P* genes in *D. melanogaster*. The maximum parsimony analysis showed that the genes of the obscura group species (*D. pseudoobscura*, *D. miranda*, and *D. persimilis*) clustered within each locus (Figure 4). In this species group, separate analyses of the 5′ flanking region (268 nucleotide sites) and the 3′ flanking region (376 nucleotide sites) showed similar relationships between genes and loci, although the nearest neighbors within each *Est-5* gene cluster differed, most likely as a result of recombination (not shown).

The clustering of species within genes indicates that the *Est-5* gene duplications predate the divergence of the three sibling species and that mechanisms of concerted evolution have not homogenized the genes,

TABLE 4

Polymorphism and fixed differences at nonsynonymous and synonymous sites between *Est-5* genes in *D. pseudoobscura*

	Est-5A vs. Est-5B		*Est-5A vs. Est-5C*		*Est-5B vs. Est-5C*	
	Polymorphism	Fixed	Polymorphism	Fixed	Polymorphism	Fixed
Nonsynonymous	51	125	29	164	46	86
Synonymous	93	136	59	186	100	121
	$P = 0.016$		$P = 0.023$		$P = 0.058$	

For the two-sided Fisher's exact tests, 8 *Est-5A*, 8 *Est-5C*, and 16 *Est-5B* sequences were used.

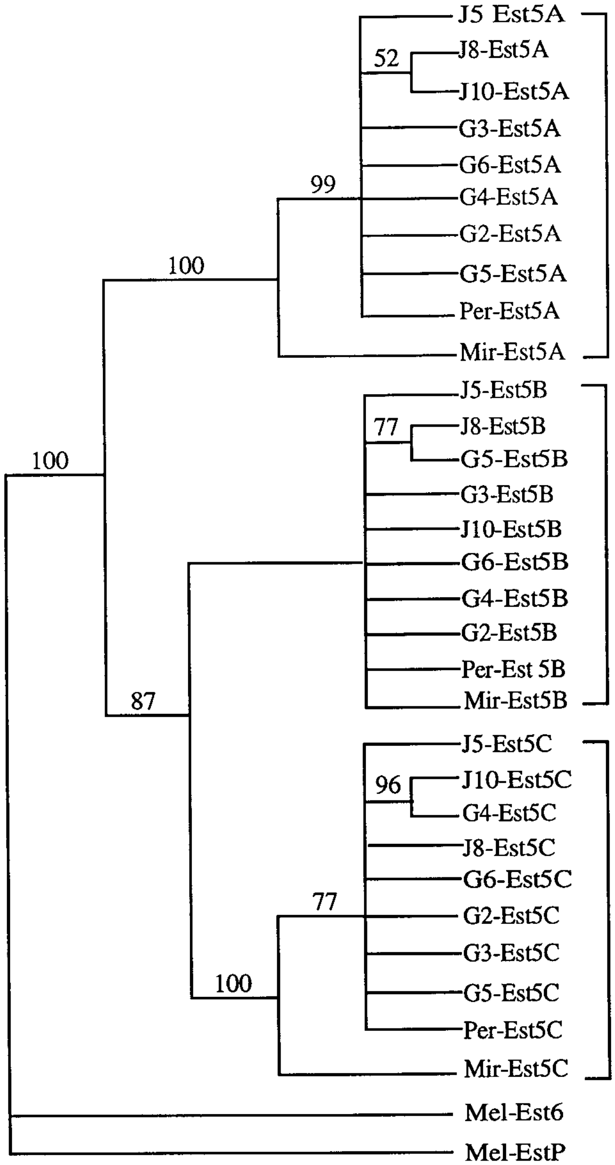

Figure 4.—A 50% majority rule bootstrap consensus tree. Numerals adjacent to each branch refer to bootstrap support from 100 replicates. Forty-two equally parsimonious trees were found. The tree length of the strict consensus tree is 1799, with a consistency index of 0.638.

since the species diverged from a common ancestor. This is in contrast to the relationship between the *Est-5* genes in *D. pseudoobscura* and the *Est-6* and *Est-P* in *D. melanogaster*: *Est-6* and *Est-P* genes cluster with one another and not with the putatively orthologous *Est-5B* and *Est-5A* genes, respectively, so the esterase gene family members have been homogenized within the *melanogaster* and *obscura* species groups since the time of their divergence (*e.g.*, ~25 mya; Russo *et al.* 1995). The relationships among clusters of esterase genes were the same for the neighbor-joining tree; only the relationships of the *D. miranda* and the *D. persimilis* genes with respect to the *D. pseudoobscura* alleles differed (not shown).

DISCUSSION

The data on *Est-5A* and *Est-5C* polymorphism and haplotype variation, in addition to previous data on *Est-5B* nucleotide sequence polymorphism, contribute to understanding the factors that influence *Est-5B* polymorphism and to understanding the evolution of this small multigene family. The polymorphism data provide statistical support for the hypothesis that interlocus gene conversion contributes to amino acid polymorphism, and may partly explain why *Est-5B* is a highly polymorphic allozyme locus. Gene conversion was detected in the coding regions between *Est-5A* and *Est-5B*, and between *Est-5B* and *Est-5C*, but not between the two outer loci, *Est-5A* and *Est-5C*. The flanking regions were not examined for evidence of gene conversion (between loci) because they are not alignable much beyond 250–350 bp. The interlocus conversion events can explain at least 4 of the 33 (12.1%) polymorphic amino acid positions in EST5B (16 sequences), 1 of the 17 (5.9%) polymorphic amino acid positions in EST5A, and 1 of the 12 (8.3%) polymorphic amino acid positions in EST5C. Interlocus gene conversion can also explain the following proportions of polymorphic synonymous sites: 1/26 (3.8%) in *Est-5A*, 12/67 (17.9%) in *Est-5B* (16 sequences), and 5/34 (14.7%) in *Est-5C*.

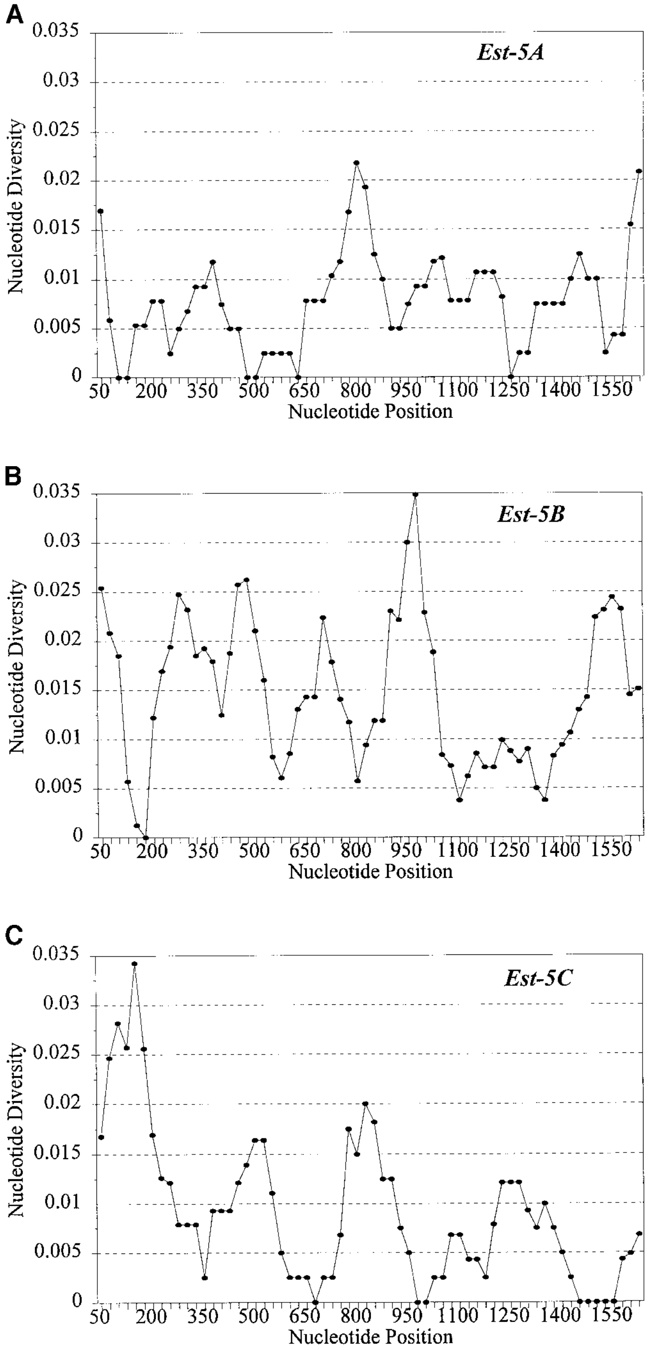

Figure 5.—Nucleotide diversity in the coding regions of the *Est-5* genes in sliding windows of 100 bp, step size 25 bp, with centers on 50-bp intervals. (A) *Est-5A*, (B) *Est-5B*, and (C) *Est-5C*.

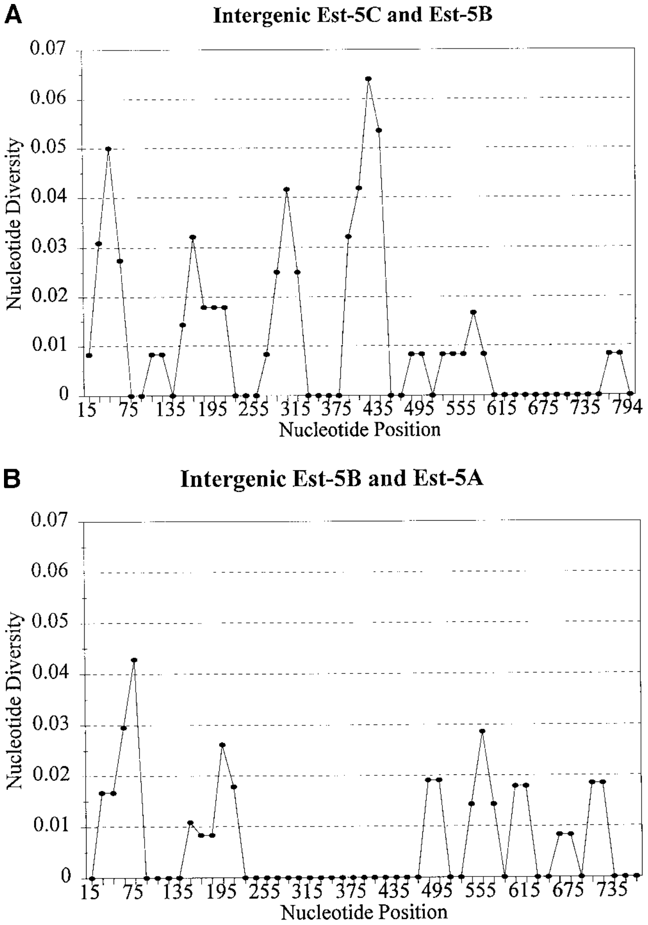

Figure 6.—Nucleotide diversity in the noncoding regions between the *Est-5* genes in sliding windows of 30 bp, step size 15 bp, with centers on 15-bp intervals. (A) Intergenic region between *Est-5C* and *Est-5B*. (B) Intergenic region between *Est-5B* and *Est-5A*.

The levels of polymorphism in the coding regions are similar for all three genes and fit neutral theory expectations. However, the polymorphic sites have a significantly heterogeneous distribution in the coding and intron regions of each gene. Figure 5 shows nucleotide diversity in sliding window intervals across the coding region of each gene. The magnitude of variation does not always correspond to the same location in the three genes, and comparisons of the location of conversion tracts (Table 3) and peaks of nucleotide diversity show that they are related. In *Est-5B*, at least three peaks of nucleotide diversity, at intervals entered at 250, 700, and 950 bp (Figure 5B), correspond to gene conversion tracts at sites 255–257, 699–705, and 942–947. In *Est-5C*, the intervals with the highest nucleotide diversity at 100–150 bp (Figure 5C) correspond to the conversion tract at nucleotide sites 132–134. In *Est-5A*, nucleotide diversity at the 400-bp interval corresponds to the tract at sites 414–415. The heterogeneity is also likely to be influenced by regions of functional constraint, for example, at residues putatively involved in the catalytic mechanism of esterases (noted above) and at six cysteine residues involved in disulfide bridges (Brady *et al.* 1990) that are conserved in the three genes and three *obscura* group species studied here, as well as in *Est-6* of *D. melanogaster*, *D. simulans*, and *D. mauritiana* (Karotam *et al.* 1993).

The maintenance of functional regulatory sequences may explain the low level and pattern of variation in the intergenic regions (Figure 6). One regulatory motif, ACTGGT, identified in *D. pseudoobscura* (Healy *et al.* 1996), corresponds to sites 693–698 bp in the intergenic *Est-5C*/*Est-5B* region, where there is no sequence variation (Figure 6A). This motif is also conserved in *D. persimilis* and *D. miranda*. The motif is present in *Est-5A* (sites 667–673 bp in Figure 6B) in *D. pseudoobscura* and *D. persimilis*, but not in *D. miranda*, where a 9-bp deletion is located. *Est-5C* shows an imperfect motif, ATTGGT, at sites −89 to −90 bp from the translation start site in all three sibling species. In the 3′ flanking regions, polyadenylation signal sequences (Brady and Richmond 1992) are conserved in the three genes and species. These are located beginning at 263 bp in *Est-5C* (Figure 6A) and at either 110 or 119 bp in *Est-5B* (Figure 6B; the 3′ region of *Est-5A* is not shown).

Evidence of positive selection on *Est-5* genes: Ohta (1994) suggests that an acceleration of amino acid changes between duplicated genes in conjunction with functional differentiation is evidence of positive selection. Although the total number of fixed differences at nonsynonymous sites is not greater than the total number of fixed differences at synonymous sites in pairwise comparisons of the *Est-5* genes, the ratios of nonsynonymous to synonymous variation for both fixed differences and polymorphism show evidence of adaptive amino acid divergence between *Est-5A* and *Est-5B*/*Est-5C*. The gene duplication resulting in *Est-5B* and *Est-5C* is putatively the most recent event in the evolution of the *Est-5* gene family (Brady and Richmond 1992). The number of net nucleotide substitutions per site is lowest between *Est-5B* and *Est-5C* (0.16450 ± 0.0210) compared with *Est-5A* and *Est-5B* (0.1989 ± 0.0251) and with *Est-5A* and *Est-5C* (0.1989 ± 0.0251), and supports this hypothesis. Therefore, most of the amino acid divergence between *Est-5A* and *Est-5B*/*C* may have predated the *Est-5B*/*C* duplication. The evidence of differential gene expression of *Est-5A* and *Est-5B* (Brady *et al.* 1990) is consistent with the interpretation of positive selection on functional divergence, although it is unknown if the difference in amino acid composition is associated with a difference in enzyme function.

Variation at *Est-5B* is higher than variation at *Est-5A* and *Est-5C*, and is relatively high compared to other *D. pseudoobscura* genes (Moriyama and Powell 1996), although there is so far no evidence of selective mechanisms operating on amino acid or total nucleotide sequence variation. Perhaps what makes this gene most unusual is that it lies between *Est-5A* and *Est-5C*, so that in the short term, gene conversion involving two different loci contributes to *Est-5B* sequence variation and amino acid polymorphism. Mutation and intragenic recombination also contribute to haplotype diversity, so these three factors may explain the considerable *Est-5B* allozyme variation. Over longer periods of time, gene conversion and/or reciprocal recombination has homogenized the *Est-5* genes, based on a phylogenetic analysis of esterase genes (Figure 4), although there appears to be selection for amino acid divergence between the EST5 proteins.

I thank B. Richter for assistance in sequencing and A. Barbadilla for assistance in the analysis of gene conversion tract lengths. A. Babadilla, M. P. Cummings, and R. C. Lewontin, and anonymous reviewers provided useful comments on the manuscript. This work was supported by National Institutes of Health grant GM-21179 to R. C. Lewontin, and National Science Foundation grant DEB 95-24595 and University of Miami General Research Awards to L. M. King.

LITERATURE CITED

Arnhein, N., 1983 Concerted evolution of multigene families, pp. 38–61 in *Evolution of Genes and Proteins*, edited by M. Nei and R. K. Koehn. Sinauer, Sunderland, MA.

Árnason, E., 1991 Perturbation-reperturbation test of selection vs. hitchhiking of the two major alleles of *esterase-5* in *Drosophila pseudoobscura*. Genetics **129:** 145–168.

Árnason, E., 1982 An experimental study of neutrality at the *Malic Dehydrogenase* and *Esterase-5* loci in *Drosophila pseudoobscura*. Hereditas **96:** 13–27.

Babcock, C. S., and W. W. Anderson, 1996 Molecular evolution of the sex-ratio inversion complex in *Drosophila pseudoobscura*: analysis of the esterase-5 gene region. Mol. Biol. Evol. **13:** 287–308.

Betrán, E., J. Rozas, A. Navarro and A. Barbadilla, 1997 The estimation of the number and the length distribution of gene conversion tracts from population DNA sequence data. Genetics **146:** 89–99.

Brady, J. P., and R. C. Richmond, 1990 Molecular analysis of evolutionary changes in the expression of *Drosophila* esterases. Proc. Natl. Acad. Sci. USA **87:** 8217–8221.

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

Brady, J. P., R. C. Richmond and J. G. Oakeshott, 1990 Cloning of the esterase-5 locus from *Drosophila pseudoobscura* and comparison with its homologue in *D. melanogaster*. Mol. Biol. Evol. **7:** 525–546.

Coyne, J. A., A. A. Felton and R. C. Lewontin, 1978 Extent of genetic variation at a highly polymorphic esterase locus in *Drosophila pseudoobscura*. **75:** 5090–5093.

Devereux, J., P. Haeberli and O. Smithies, 1984 A comprehensive set of sequence analysis programs for the VAX. Nucleic Acids Res. **14:** 623–633.

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

Healy, M. J., M. M. Dumancic, A. Cao and J. G. Oakeshott, 1996 Localization of sequences regulating ancestral and acquired sites of esterase 6 activity in *Drosophila melanogaster*. Mol. Biol. Evol. **13:** 784–797.

Hilliker, A. J., G. Harauz, A. G. Reanume, M. Gray, S. H. Clark *et al.*, 1994 Meiotic gene conversion tract length distribution within the *rosy* locus of *Drosophila melanogaster*. Genetics **137:** 1019–1026.

Hudson, R. R., M. Kreitman and M. Aguadé, 1987 A test of neutral molecular evolution based on nucleotide data. Genetics **116:** 153–159.

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

Keith, T. P., 1983 Frequency distribution of esterase-5 alleles in two populations of *Drosophila pseudoobscura*. Genetics **105:** 135–155.

Kuhner, M. K., D. A. Lawlor, P. D. Ennis and P. Parham, 1991 Gene conversation in the evolution of the human and chimpanzee MHC class I loci. Tissue Antigens **38:** 152–164.

Kumar, S., K. Tamura and M. Nei, 1993 MEGA Molecular Evolutionary Genetics Analysis. Version 1.01. The Pennsylvania State University, University Park, PA.

Lewontin, R. C., and J. L. Hubby, 1966 A molecular approach to the study of genic heterozygosity in natural populations. II. Amount of variation and degree of heterozygosity in natural populations of *Drosophila pseudoobscura*. Genetics **54:** 595–609.

McDonald, J., and M. Kreitman, 1991 Adaptive protein evolution at the *Adh* locus in *Drosophila*. Nature **351:** 652–654.

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

Nei, M., 1987 *Molecular Evolutionary Genetics*. Columbia University Press, New York.

Nei, M., and T. Gojobori, 1986 Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. Mol. Biol. Evol. **3:** 418–426.

Ohta, T., 1994 Further evidence of evolution by gene duplication revealed through DNA sequence comparisons. Genetics **138:** 1331–1337.

Rozas, J., and R. Rozas, 1995 DnaSP: DNA sequence polymorphism—an interactive program for estimating population genetics parameters from DNA sequence data. Comput. Appl. Biosci. **11:** 621–625.

Russo, C. A. M., N. Takezaki and M. Nei, 1995 Molecular phylogeny and divergence times of Drosophilid species. Mol. Biol. Evol. **12:** 391–404.

Sambrook, J., E. F. Fritsch and T. Maniatis, 1989 *Molecular Cloning: A Laboratory Manual*, Ed. 2, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.

Smith, S. W., R. Overbeek, C. R. Woese, W. Gilbert and P. M. Gillevet, 1994 The genetic data environment an expandable GUI for multiple sequence analysis. Comput. Appl. Biosci. **10:** 6715.

Swofford, D. L., 1992 PAUP: Phylogenetic analysis using parsimony, portable version (Unix) 3.0r+4 (pre-release 0.4). Illinois Natural History Survey, Champaign.

Tajima, F., 1993 Measurement of DNA polymorphism, pp. 37–59 in *Mechanisms of Molecular Evolution*, edited by N. Takahata and A. G. Clark. Sinauer, Sunderland, MA.

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

Veuille, M., and L. M. King, 1995 Molecular basis of polymorphism at the *esterase-5B* locus in *Drosophila pseudoobscura*. Genetics **141:** 255–262.

Wines, D. R., J. M. Brady, E. M. Southard and R. J. MacDonald, 1991 Evolution of the rat kallikrein gene family: gene conversion leads to functional diversity. J. Mol. Evol. **32:** 476–492.

Xiong, Y., B. Sakaguchi and T. H. Eickbush, 1988 Gene conversion can generate sequence variants in the late chorion multigene families of *Bombyx mori*. Genetics **120:** 221–231.

Yamazaki, T., 1971 Measurement of fitness at the esterase-5 locus in *Drosophila pseudoobscura*. Genetics **67:** 579–603.

Communicating editor: A. G. Clark