Nucleotide Sequence of the *Adh* Gene Region of *Drosophila pseudoobscura*: Evolutionary Change and Evidence for an Ancient Gene Duplication

Stephen W. Schaeffer,^{*,†,1} and Charles F. Aquadro^{*,2}

^{*}Laboratory of Genetics, National Institute of Environmental Health Sciences, Research Triangle Park, North Carolina 27709, and

[†]Department of Genetics, University of Georgia, Athens, Georgia 30602

Manuscript received March 13, 1987

Revised copy accepted June 8, 1987

ABSTRACT

The alcohol dehydrogenase (*Adh*) locus (ADH; alcohol: NAD⁺ oxidoreductase, EC 1.1.1.1) of *Drosophila pseudoobscura* was cloned and sequenced. Forty-five percent of the "effectively silent sites" have changed between *Adh* in *D. pseudoobscura* of the *obscura* species group and the homologous DNA sequence in *D. mauritiana*, the latter representing the *melanogaster* species group. The untranslated leader sequence of the adult transcript of *D. pseudoobscura* has two deletions relative to the *D. mauritiana* message. The ADH protein sequences of *D. pseudoobscura* is missing the third and fourth amino acids at the N-terminus relative to the *D. mauritiana* enzyme. Of the remaining 254 amino acid positions, 27 (10.64%) differ between the two species. Amino acid replacements are randomly distributed into hydrophilic and hydrophobic domains of ADH. However, replacement substitutions are distributed nonrandomly across the three exons among *D. pseudoobscura* and members of the *melanogaster* subgroup, suggesting that functional constraints across the exons are different. Surprisingly, silent substitutions are also nonrandomly distributed with the third exon being the most divergent. This pattern suggests possible selective constraints on supposedly neutral silent substitutions and/or variation in underlying mutation rates across the gene. The presence of transcriptional and translational signals at the beginning and end of conserved sequences 3' to *Adh* implies the existence of a previously undescribed gene. Codon usage and patterns of nucleotide divergence are consistent with a protein coding function for this gene. In addition, conservation of nucleotide and amino acid sequence and similarity in hydropathy plots suggests that the gene 3' to *Adh* represents an ancient duplication of the *Adh* gene.

THE alcohol dehydrogenase (*Adh*) locus of *Drosophila* is a model system for studying the processes of molecular evolution. The DNA sequence of *Adh* can be partitioned into nucleotide sites that alter amino acid sequence (replacement sites) or those that do not (synonymous, intron and flanking sites). Comparing the rates of substitution in replacement versus silent sites in closely related taxa, we can often distinguish between the effects of some forms of natural selection and random genetic drift (KREITMAN 1983; LEWONTIN 1985). Synonymous substitutions in the coding region outweighed replacement changes 13 to 1 when 11 sequences of *Adh* within *D. melanogaster* were compared (KREITMAN 1983). A comparison of *Adh* sequences among sibling species of *D. melanogaster* shows a similar trend (BODMER and ASHBURNER 1984; COHN, THOMPSON and MOORE 1984); all studies suggesting strong purifying selection removing deleterious amino acid changes. KREITMAN (1983) also found that sequences 3' to *Adh* were more conserved

than either silent sites in exons or intron sites. The assumption that intergenic regions are not selectively constrained appears to be contradicted by the latter results. Additional DNA sequence comparisons of the *Adh* region with more divergent taxa may provide more information about the evolution of this locus and the possible significance of the flanking sequence conservation.

In this paper, we present the DNA sequence of the *Adh* region of *D. pseudoobscura*, part of the *obscura* species group, thought to have diverged from the *melanogaster* species group during the mid-Oligocene 20–25 mya (THROCKMORTON 1975). Our DNA sequence comparison between *D. pseudoobscura* and *D. mauritiana* highlights strongly conserved regions that correlate with functional domains. *Drosophila mauritiana* was chosen as the representative of the *melanogaster* species group over *D. melanogaster* because substantially more sequence 3' to *Adh* was available for the former species. As with comparisons among species in the *melanogaster* species group, silent substitutions were more numerous than replacement changes in our *D. pseudoobscura* and *D. mauritiana* comparison, but the absolute number of changes reflects the increased divergence time. The present comparison also

¹ Present address and to whom correspondence should be addressed: Museum of Comparative Zoology, Harvard University, 26 Oxford Avenue, Cambridge, Massachusetts 02139.

² Present address: Section of Genetics and Development, Emerson Hall, Cornell University, Ithaca, New York 14853.

The sequence data presented in this article have been submitted to the EMBL/GenBank Data Libraries under the accession number Y00602.

shows the conservation of sequence 3' to *Adh* in a pattern consistent with the presence of a gene similar to *Adh* which may have resulted from an ancient gene duplication.

MATERIALS AND METHODS

Total genomic DNA was isolated following BINGHAM, LEVIS and RUBIN (1981) from a strain of *D. pseudoobscura* homozygous for the standard third chromosomal arrangement and the Esterase-5 "100" allozyme (from F. J. AYALA). The library (provided by C. H. LANGLEY) was constructed by ligating genomic DNA, partially cleaved with *Mbo*I then treated with calf intestine alkaline phosphatase, into the *Bam*HI site of an EMBL4 phage vector (FRISCHAUF *et al.* 1983). A total of 50,000 recombinant plaques from the library was screened (BENTON and DAVIS 1977) for sequences homologous to the *D. melanogaster* sAC1 probe which contains a 4.75-kb insert that includes the entire *Adh* transcriptional unit (GOLDBERG 1980). *Adh* in *D. melanogaster* corresponds to the electrophoretically monomorphic *Adh*-1 locus of *D. pseudoobscura* (CHAMBERS *et al.* 1978). The sAC1 probe labeled with [α -³²P]dCTP (RIGBY *et al.* 1977) was hybridized to phage DNA on nitrocellulose filters in 50% formamide at 37° overnight. Nonspecifically hybridized probe was removed with three washes in 2 × SSC/0.1% NaDodSO₄ at room temperature for five min each, then two washes in 0.1 × SSC/0.1% NaDodSO₄ at 42° for 15 min each (1 × SSC is 0.15 M NaCl/0.015 M sodium citrate, pH 7.5). Autoradiography followed the wash steps. DNA from recombinant phages was isolated according to MANIATIS, FRITSCH and SAMBROOK (1982).

Characterization of *D. pseudoobscura* clones: The restriction sites of *Bam*HI, *Eco*RI, *Hind*III, *Sal*I, *Xba*I and *Xho*I were located in each of the recombinant clones using single and double digests (McDONELL, SIMON and STUDIER 1977; MANIATIS, FRITSCH and SAMBROOK 1982). *Adh* was localized on the phage restriction map by transferring digested and size fractionated DNA to nylon filters (Zetabind from AMF Cuno) using the transfer method of SOUTHERN (1975) with modifications of SMITH and SUMMER (1980). The labeling of sAC1, hybridization of the probe, washing of the filters and autoradiography, was the same as for plaque hybridization. The chromosomal location of *Adh* in *D. pseudoobscura* was determined by *in situ* hybridization of biotinylated recombinant clones to salivary chromosomes (PARDEE and GALL 1975; LANGER, WALDROP and WARD 1981; E. MONTGOMERY, personal communication) (biotinylated dUTP was obtained commercially from Bethesda Research Laboratories). Total poly(A)⁺ RNA from *D. pseudoobscura* and *D. melanogaster* was isolated according to MANIATIS, FRITSCH and SAMBROOK (1982). Northern analysis of poly(A)⁺ RNA from both species used the method of GOLDBERG (1980). The Northern filters were hybridized with a *D. pseudoobscura* *Adh* probe (*Adh*6, see below) labeled with [α -³²P]dCTP (RIGBY *et al.* 1977) in 50% formamide at 42° overnight. Nonspecifically hybridized probe was removed with two washes in 2 × SSC/0.1% NaDodSO₄ at room temperature for 10 min each, followed by two washes in 0.1 × SSC/0.1% NaDodSO₄ at 50° for 15 min each. Autoradiography followed the wash steps. Labeled *D. melanogaster* sAC1 clone was hybridized, washed and autoradiographed to the Northern filter as above after the *D. pseudoobscura* clone was removed with two 10 min washes of 0.1 × SSC/0.1% NaDodSO₄ at 95°. The *D. melanogaster* copia and amylase transcripts were also probed on the northern filters as size standards (FLAVELL *et al.* 1980; GEMMILL, LEVY and DOANE 1985).

DNA sequence analysis: The 2.7-kb *Hind*III, 3.9-kb *Eco*RI, 0.2-kb *Sal*I/*Hind*III and 0.6-kb *Sal*I fragments of recombinant phage *Adh*6 were subcloned into M13 strains mp18 and mp19 (NORRANDER, KEMPE and MESSING 1983) in both orientations. The 3.9-kb *Eco*RI fragment was used to sequence across the *Hind*III and *Sal*I sites. A series of nested deletions was generated for the *Hind*III 2.7-kb cloned fragment using the method of HENIKOFF (1984). These clones were sequenced according to the methods of SANGER, NICKLEN and COULSON (1977) on TBE buffer gradient gels (BIGGINS, GIBSON and HONG 1983). A single *Adh* region sequence was constructed from the M13 deletion clones using the database programs of R. STADEN.

The *Adh* sequence of *D. pseudoobscura* was aligned with the 4596-kb sequence of *D. mauritiana*, a representative of the *melanogaster* species group (COHN 1985), using the NUCALN program (WILBUR and LIPMAN 1983). The NUCALN program uses the NEEDLEMAN and WUNSCH (1970) algorithm of alignment, assigning a score of +1 for a matched base pair, -1 for a mismatched base pair, and a penalty of 7 for the introduction of a gap in the sequence. The *D. mauritiana* sequence was chosen for comparison since more sequence was available 3' to *Adh* for this species than for *D. melanogaster*. The numerous insertions and deletions generated by the alignment of the two sequences were not included in the tabulation of nucleotide substitution frequencies. The effective number of silent sites for the *Adh* coding sequence in *D. pseudoobscura* was determined according to HOLMQUIST, CANTOR and JUKES (1972). All silent and replacement changes in the *Adh* sequence were tabulated for *D. pseudoobscura*, *D. melanogaster*, *D. simulans*, *D. mauritiana*, *D. sechellia* and *D. oreana* (KREITMAN 1983; BODMER and ASHBURNER 1984; COHN, THOMPSON and MOORE 1984; COHN 1985; COYNE and KREITMAN 1986) to examine the distribution of variable sites into exons and protein domains. Protein domains in the predicted ADH protein were determined by hydropathy plots using the algorithm of HOPP and WOODS (1981).

RESULTS

Clone characterization: The library screen yielded five overlapping phage clones homologous to the *D. melanogaster* sAC1 probe: *Adh*2, *Adh*3, *Adh*4, *Adh*6 and *Adh*7. *Adh*6 has a 15.2-kb insert that includes the sequences found in the other four clones, thus *Adh*6 is the only *Adh* clone in *D. pseudoobscura* that is used in subsequent analyses. A partial restriction map of *Adh*6 is given in Figure 1B. The complete restriction map can be found in SCHAEFFER, AQUADRO and ANDERSON (1987). *Adh*6 hybridizes to chromosomal section 88 on chromosome IV of *D. pseudoobscura* (Figure 2). Chromosome IV in *D. pseudoobscura* is homologous to Chromosome IIL of *D. melanogaster* (STURTEVANT and NOVITSKI 1941) where *Adh* has been previously mapped (GRELL, JACOBSEN and MURPHY 1965).

The *Adh*6 clone hybridizes to both *D. pseudoobscura* and *D. melanogaster* poly(A)⁺ RNA, but hybridization is stronger to *D. pseudoobscura* RNA. The *D. pseudoobscura* adult transcript is approximately 1040 bp long, about 30 bp shorter than the *D. melanogaster* message (Figure 3A). Hybridization of the *D. melanogaster* sAC1 probe (Figure 3B) shows the same mes-

FIGURE 1.—Fine structure and restriction map of the *Adh* region in *D. pseudoobscura*. (A) Fine structure of the *Adh* locus showing the adult (P_A) and larval (P_L) promoters, the adult (L_A) and larval (L_L) leader sequences, and the *Adh* coding region (shaded blocks). An open reading frame corresponding to a newly discovered gene 3' to *Adh* is also shown (see text). (B) Restriction map of the *Adh* region. Location of restriction sites for *Bam*HI, *Eco*RI, *Hind*III, *Sal*I and *Xho*I are indicated by the letters B, E, H, S and X, respectively. (C) Nucleotide sequence similarity of homologous *D. pseudoobscura* and *D. mauritiana* *Adh* regions. Each block represents the number of base pairs of 30 that matched. Sequence length variation in *D. mauritiana* relative to *D. pseudoobscura* is shown above the plot as triangles pointing down (insertions) or up (deletions).

FIGURE 2.—An *in situ* hybridization of the *Adh6* clone to *D. pseudoobscura* polytene chromosomes. Hybridization (arrow) is in section 88 on the fourth chromosome on the cytogenetic map.

sages being probed as in Figure 3A, but as expected the *D. melanogaster* RNA shows the strongest hybridization.

Sequence comparisons of *Adh* coding regions: The *D. pseudoobscura* *Adh* gene region was sequenced for a total of 3535 bp and aligned with the *D. mauritiana* sequence (Figure 4). The coding and leader sequences had the highest homology between the two species (Figure 1C). The third and fourth codons have been

deleted from the translated *D. pseudoobscura* sequence; all other codon positions align exactly between *D. pseudoobscura* and *D. mauritiana*.

The *Adh* coding region has accumulated 114 nucleotide substitutions out of 762 bp since the divergence of *D. pseudoobscura* and *D. mauritiana*. Substitutions have not occurred equally in the three codon positions ($\chi^2 = 69.5$; d.f. = 2; $P < 0.0001$); most occurred in the third position (Figure 4). The number

FIGURE 3.—Northern analysis of poly(A) mRNA from *D. pseudoobscura* (lane 2) and *D. melanogaster* (lane 1). (A) Total poly(A) mRNA probed with the *Adh* probe of *D. pseudoobscura*, *Adh6*. (B) The same filter in (A) probed with the *Adh* probe from *D. melanogaster*, *sAC1*.

of transitions and transversions were equivalent, thus deviating significantly from the expected ratio of 2:1, transversions to transitions ($\chi^2 = 18.0$; d.f. = 1; $P < 0.001$). Additionally, within the transition class of substitutions, the numbers of the two types of transitions deviated significantly from the expected ratio of 1:1 ($\chi^2 = 14.26$; d.f. = 1, $P < 0.0001$); T-C transitions were more frequent than A-G transitions. Within the transversion class, G-C transversions are overrepresented while A-T and G-T transversions are less frequent in comparison to A-C transversions (a test for equal frequencies of the four kinds of transversion gave $\chi^2 = 16.96$; d.f. = 3; $P < 0.01$).

Only a fraction of nucleotide sites (termed silent or synonymous sites) in structural genes can change without altering the amino acid sequence of the gene product. The effective number of silent sites for a gene is the total number of coding nucleotide sites weighted by the proportion of potentially silent sites in the gene. This latter proportion was calculated as the average, over all codons in the *D. pseudoobscura* *Adh* coding sequence, of the fraction of the nine possible single base pair substitutions that would not lead to an amino acid replacement. Twenty-five percent of all sites (190.4 of 762) in the *Adh* coding sequence of *D. pseudoobscura* are "effectively silent" (synonymous). We observed 87 nucleotide differences in silent sites between the *D. pseudoobscura* and *D.*

mauritiana *Adh* sequences. Thus 45.6% of silent sites differ between these two species. Silent substitutions were nonrandomly distributed across the three exons (Table 1; $\chi^2 = 8.99$; d.f. = 2; $P < 0.05$); the second exon had fewer changes than expected while the third exon had an excess of changes (Table 1). This trend is supported further by the significant nonrandom distribution of silent site changes summed over the *melanogaster* species subgroup relative to *D. pseudoobscura* (Table 2).

Twenty-seven replacement changes have accumulated between the two *Adh* sequences. The effective number of replacement sites is equal to the total number of coding sites minus the number of effectively silent sites ($762 - 190.4 = 571.6$). Thus, an average of 4.7% of the replacement sites have changed between the *D. pseudoobscura* and *D. mauritiana* *Adh* sequences. Replacement differences between these two species were randomly distributed among the three exons (Table 1; $\chi^2 = 4.67$; d.f. = 2; $P > 0.05$). Most amino acid replacements were conservative in that amino acids of similar charge were substituted. The four amino acid differences that do result in a charge change seem to leave the net charge of the enzyme intact (Figure 4). Codon usage was equivalent to that observed in other *Drosophila* genes (ASHBURNER, BODMER and LEMENIER 1984) ($\chi^2 = 25.18$; d.f. = 18; $P > 0.10$).

The hydropathy plot of ADH presented in Figure 5 predicts the hydrophilic and hydrophobic domains of the protein based on the hydropathy values of the constituent amino acids (HOPP and WOODS 1981). We find that most amino acid replacements among the *melanogaster* species subgroup of *D. pseudoobscura* are distributed randomly into hydrophilic and hydrophobic domains ($\chi^2 = 2.36$; d.f. = 1; $P > 0.10$; amino acid replacements among the *melanogaster* species group and *D. pseudoobscura* can be found in Figure 5 under the hydropathy plot of *D. melanogaster* ADH⁶).

The adult and larval promoters given in *D. mauritiana* (by analogy to *D. melanogaster*) (BENYAJATI *et al.* 1983) can be found in the same relative position in the *D. pseudoobscura* *Adh* sequence (Figure 4). The proposed polyadenylation signal in *D. pseudoobscura* differs from the *D. mauritiana* by a single nucleotide change to give one of the infrequent polyadenylation signals 'AATACA' (NEVINS 1983; WICKENS and STEPHENSON 1984). In addition to the numerous nucleotide changes in the leader sequences and coding regions, two small deletions are observed in the *D. pseudoobscura* nontranslated adult message (Figure 4). Ten base pairs have been deleted at position +40 in the adult leader sequence of *D. pseudoobscura*. A deletion of 24 bp is observed in the larval leader/adult leader sequence of *D. pseudoobscura* at position +887 of the adult message.

The positions of the introns in *Adh* are identical in

FIGURE 4. Sequence of the *Adh* region from *D. pseudoobscura*. The +1 position is the first nucleotide of the adult leader sequence. The promoter and polyadenylation recognition sequences are boxed. Adult and larval promoters of *Adh* are at position -31 and +806, respectively; the polyadenylation signal is at +1901. The amino acid sequence of the *Adh* gene is presented below the nucleotide sequence, beginning at +888 and ending at +1773. Gaps in the amino acid sequence show intron positions. The promoter of the proposed 3' gene is at position +1958. The polyadenylation recognition sequence of the proposed 3' gene is at position +3320. The amino acid sequence of this gene begins at position +2094 and ends at position +3298. Silent changes in coding regions are shown above the nucleotide sequence. Replacement changes are also shown above the nucleotide sequence but are underlined, as are the amino acids which are different between *D. pseudoobscura* and *D. mauritiana*.

D. pseudoobscura and *D. mauritiana* despite changes in length. The adult intron is larger in *D. pseudoobscura*, 794 vs. 654 bp. Both introns in the coding sequence are shorter in *D. pseudoobscura*, 63 vs. 65 bp in the first intron and 60 vs. 70 bp in the second intron. The consensus splice junction for the introns in *Adh* are 'GTAAG' at the donor site and 'AG' at the acceptor site. These consensus sequences for *Adh* agree with those previously published (NEVINS 1983).

Evidence for the presence of a gene 3' to *Adh*: KREITMAN (1983) found more conservation in sequences 3' to *Adh* than in intron sites or silent sites in exons. Similarly, we observed that 31.4% of the sites downstream from *Adh* differed between *D. pseudoobscura* and *D. mauritiana* which is much less than differences in intron sites, 50.7%, or silent sites, 45.7%, of *Adh* ($\chi^2 = 76.38$; d.f. = 2; $P < 0.0001$) (Table 1).

The presence of a protein-coding gene appears to be responsible for the strong conservation of nucleotide sequence in the 3' region. We inferred the position of 3 exons in this gene based on the three discontinuous blocks of strong sequence similarity broken up by two sections of reduced similarity (Figure 1C). Each hypothesized exon is an open reading frame (Figure 4). Locations of these exons is further supported by the presence of typical *Drosophila* codon bias shown by the Protein Coding Region Locator program written by JAMES PUSTELL (PUSTELL and KAFATOS 1986) and distributed by International Biotechnologies Inc. Transcriptional and translational signals are located at the appropriate locations relative to these blocks of similarity (i.e., 'AUG' start codon, 'UGA' stop codon, and 'GT...AG' intron splice junction sequences).

Transcriptional signals upstream of the gene were

```

      *   G G *       *       *       *T T C T T   C*       *
1281 CTGGTCAACACCAACAGCCATTCTGGACTTCTGGGCAAGCGCGGCCAGCGGCATCATTGCAACATTGG
    LeuValAsnThrThrThrAlaIleLeuAspPheTrpAspLysArgLysGlyGlyProGlyGlyIleIleCysAsnIleG1

      A       T       *       *       C       A       C *C   C * C   AG *
1361 CTCGTCACCGGTTTCAATGCCATCTACAGGTGCCGTTTACTCGGGCTCCAAGGCGCGGTGTTAACTTTACCTCCT
    ySerValThrGlyPheAsnAlaIleTyrGlnValProValTyrSerGlySerLysAlaAlaValValAsnPheThrSerS

      *       *       *       *       *       *       *       C *
1441 CCCTGGCGGTAAAGAAATCCTACAATCTACTCCCAAACTAAATATTTCTTTGAAATCCCATAGAACTGGCTCCC
    erLeuAla                                     LysLeuAlaPro

      C C* G C   *   C G C C   * CCGC   *C   G * G   * A   *
1521 ATTACTGGTGTCACTGCTTACACTGTCAATCTGGCATCACTAAGACCACTCTGGTCCACAAATCAACTCGTGGCTGGA
    IleThrGlyValThrAlaTyrThrValAsnProGlyIleThrLysThrThrLeuValHisLysPheAsnSerTrpLeuAs

      T       * AG T C*       C *G CT   *       C * ATT GCC *       * C C *
1601 TGTGGAGCCCCGTGTGGCGGAGAAGCTGCTCGAGCATCCACCCAGACCTCTCAGCAGTGCGCCGAGAAGCTTTGTGAAGG
    pValGluProArgValAlaGluLysLeuLeuGluHisProThrGlnThrSerGlnGlnCysAlaGluAsnPheValLysA

      T C   *   C   *   A   *   AC   *CT A   *C   G G   CAG   * A   *
1681 CCATTGAGCTGAACAAGAACGGTGCCATCTGGAAGTTGGATCTGGGCACCTTGGAGCCCATCACATGGACCCAGCACTGG
    laIleGluLeuAsnLysAsnGlyAlaIleTrpLysLeuAspLeuGlyThrLeuGluProIleThrTrpThrGlnHisTrp

      C C   *       *       *       *       *       *       *       *
1761 GATTCCGGCATCTAAACCATCCAGAGACTCTATGGGACATGGCGTTTTAGCTTTTGTCTTTTTCCTCAATTG
    AspSerGlyIle

      *       *       *       *       *       *       *       *
1841 TTACGTATATACATACATATGGCAATAAGGCTGATTGAACCCCGTTTGAATATGATATAACAATAATTATATTGAG

      T       *       *       *       *       *       *       *
1921 AAATTTCAACAAATCGATAAGAGCTATAGCTCTCTTTTAATTAAAAAAAGCAATGGTCAGTGGTG

      *       *       *       *       *       *       *       *
2001 GTGGGCAGTGGTGTGGTGTGGCTCTATCGATTCCACACAAAACTTTACTGTTTACTAATAGAAAAGAACTGAGAAA

      *       T * T   *C       *       * G T   *A A   *       *
2081 GGCAGCCAAAAGATGTACGATCTGACGGTAAGCATGTCTGTCTATGTGGCTGACTGCGGTGGCATGCACTGGAGACCA
    MetTyrAspLeuThrGlyLysHisValCysTyrValAlaAspCysGlyGlyIleAlaLeuGluThrS

      *       * C   *       *       *       *       *       *
2161 GCAAGGTTCTCATGACCAAGAATATAGCGGTGAGTGGTAGAGTGGTGTAGAGAGTGAAGTGGTTCTGGTTCTTAA
    erLysValLeuMetThrLysAsnIleAla

      *       *       *       *       *       *       *       *
2241 CAAGTCTCTAGTCTCTAGTCTCTAGTCTCTGGCTGCTGCTGGTAGCTGTCGAACTAATTAAGTTAATGACTTTTTCAT

      *       *       *       *       *       *       *       *
2321 GTTTTGTTTTGTCCGCTGACTTTTGTCTGTTCAAAGCTACAATTAAGTTTATGTTTGTAGTTTGATCTACTTTTGCTTAA

      *       *       *       *       *       *       *       *
2401 GGGAAGTACTTTTCGATTACGGATCGAATTGCTGGCATATAAAATAGATACCAAGGAATGTGCATCATATCATATTGATT

```

FIGURE 4B.

inferred from sequences conserved among *D. pseudoobscura*, *D. melanogaster*, *D. simulans* and *D. mauritiana*. A putative promoter, 'TTAATTAAAA,' beginning at position +1958 is conserved in all the species (COHN 1985) and is located 29 bp upstream from another completely conserved sequence 'AATGG,' the latter possibly representing the start site of transcription. A 16-bp sequence, 'GCAATAAGGCTGATT,' at position +1863 is conserved in all the species and may represent the 'CAAT' box for the gene downstream of *Adh*. A consensus polyadenylation recognition sequence 'AATAAA' is 22 bp downstream from the 'UGA' stop codon for the 3' gene at position +3317. Translation of the proposed message would begin at position +2094 and end at +3298 transcribed in the same orientation as *Adh* (Figure 4). The predicted 3' gene thus has exons of 96, 405 and 333 bp interrupted by introns of 312

and 58 bp. The protein would contain 278 amino acids and have a subunit molecular weight of 30,875 in *D. pseudoobscura* (Figure 4). A 34-bp segment of conserved sequence in the first putative intron is not consistent with coding function since the 'AG' acceptor site is missing and the reading frames show no codon bias. Whether the high degree of conservation of this segment indicates an important regulatory region will be of interest to examine.

Partitioning the downstream nucleotide sequence into a transcriptional unit with coding and noncoding regions reveals a pattern of nucleotide substitution similar to that seen at *Adh* (Table 1). Nucleotide differences occurred 24 times more frequently in silent than in replacement sites in the 3' gene between *D. pseudoobscura* and *D. mauritiana* (interestingly, *Adh* silent differences occur only 9 times as often as replacement substitutions). Substitutions in introns and

FIGURE 4C.

Taken together, these data strongly support the presence of a protein-coding gene immediately downstream from *Adh* in the *melanogaster* and *obscura* subgroup species examined. The identity and function of this putative gene is of particular interest given the lack of previous genetic or molecular evidence for a locus in this region (e.g., GOLDBERG 1980; CHIA *et al.* 1985). Several lines of evidence argue that it may represent an ancestral duplication of *Adh*. The nucleotide sequence of the 3' gene coding exons is 52% homologous to the *Adh* gene, significantly greater than the random homology expectation of 25% ($\chi^2 = 304.27$; d.f. = 1; $P < 0.0001$). Seventy-nine percent of silent sites and 37% of replacement sites differ between *Adh* and the 3' gene (Table 3). The intron placement is identical between the two genes, however the size of introns and exons differs to some extent between the genes.

TABLE 1

Changes in the *Adh* region nucleotide sequence between
D. pseudoobscura and *D. mauritiana*

Region	Length (bp) ^a	No. of changes	Percent differences	Corrected percent divergence ^b
ADH TRANSCRIPTIONAL UNIT				
5' noncoding	39.0	14	35.9	48.9
Adult leader	71.0	16	22.5	26.8
Adult intron	564.0	286	50.7	84.5
Adult/larval leader	50.0	21	42.0	61.6
Exon 1				
Silent	23.2	12	51.6	87.4
Replacement	69.8	4	5.7	5.9
Intron 1	55.0	27	49.1	79.7
Exon 2				
Silent	101.2	36	35.6	48.3
Replacement	303.8	9	3.0	3.1
Intron 2	59.0	31	52.5	90.3
Exon 3				
Silent	66.0	39	59.1	>100.0
Replacement	198.0	14	5.7	7.5
INTERGENIC REGION	268.0	145	54.1	95.8
3' GENE				
Exon 1				
Silent	24.0	7	29.2	37.0
Replacement	72.0	1	1.4	1.4
Intron 1	271.0	126	46.5	72.6
Exon 2				
Silent	101.2	65	64.2	>100.0
Replacement	303.8	7	2.3	2.3
Intron 2	52.0	26	50.0	82.4
Exon 3				
Silent	79.5	57	71.7	>100.0
Replacement	238.5	9	3.8	3.9
SUMMARY				
Noncoding	1308.0	655	50.1	82.7
Adh				
Silent sites	190.4	87	45.6	70.2
Replacement sites	571.6	27	4.7	4.9
Total coding sequence	762.0	114	14.9	16.7
3' gene				
Silent sites	204.7	129	63.0	>100.0
Replacement sites	614.3	17	2.7	2.8
Total coding sequence	819.0	145	17.7	20.2

^a The effective number of silent sites in exons is 25% of the number of nucleotides in the exon (see text for calculation).

^b Corrected percent divergence estimated as $d = 3/4 \ln(1 - 4/3p)$, where p is the proportion of nucleotide sites that differ between the two sequences (JUKES and CANTOR 1969). Estimates over 100% corrected divergence are indicated only as >100%.

The amino acid sequences of the two genes are 38% homologous, also significantly greater than the random expectation of 6.1% ($\chi^2 = 458.71$; d.f. = 1; $P < 0.0001$). We also used the RDF program of LIPMAN and PEARSON (1985) to evaluate the statistical significance of the similarity of ADH and the proposed 3' gene's protein. The RDF program shuffled the amino acid sequence of the 3' protein 100 times and scored each alignment of each random sequence to ADH. Statistical significance was evaluated based on the number of standard deviations the actual alignment

TABLE 2

Nucleotide differences among the coding sequence of *Adh* genes
from *D. melanogaster*, *D. simulans*, *D. orena*, *D. mauritiana*,
D. sechellia and *D. pseudoobscura*

Region	Differences/effect and no. sites	
	Silent sites	Replacement sites
Exon 1	15/23.2 = 0.65	6/69.8 = 0.09
Exon 2	46/101.2 = 0.45	10/303.8 = 0.03
Exon 3	48/66.0 = 0.73	17/198.0 = 0.09
Silent sites over exons: $\chi^2 = 12.73$, d.f. = 2, $P < 0.005$		
Replacement sites over exons: $\chi^2 = 7.34$, d.f. = 2, $P < 0.05$		

Sequences were aligned as in BODMER and ASHBURNER (1984), as corrected by STEPHENS and NEI (1985), with the addition of the sequences from KREITMAN (1983), COHN (1985), COYNE and KREITMAN (1986) and this paper. The effective number of silent and replacement sites were calculated for each exon for the *D. pseudoobscura* and *D. mauritiana* sequences as described in the text.

was from the mean of the shuffled alignments. The alignment score of ADH vs. the 3' gene's protein was 63 standard deviations greater than the mean of the score for the randomly shuffled sequences. LIPMAN and PEARSON (1985) suggest that 10 standard deviations greater than the mean is significant.

Despite extensive divergence of amino acid sequence, the two proteins have retained a surprising similarity of conformation as indicated by the location of hydrophobic and hydrophilic domains (Figure 5). These data leave open the possibility of a dehydrogenase function for the product of the 3' gene.

DISCUSSION

The pattern of nucleotide substitution among the *Adh* genes of *D. pseudoobscura* and members of the *melanogaster* species group is consistent with that seen within *D. melanogaster* and among its close relatives (KREITMAN 1983; BODMER and ASHBURNER 1984; COHN, THOMPSON and MOORE 1984; COHN 1985; COYNE and KREITMAN 1986). Silent substitutions in exon and intron sites are more frequent than amino acid replacements, with the absolute level of change greater between *D. pseudoobscura* and *D. mauritiana* reflecting the increased time of divergence between the two species groups. As with the earlier studies, strong purifying selection seems to have constrained replacement substitutions.

Some sections of the ADH protein appear to be more selectively constrained than others since amino acid replacements are distributed nonrandomly across the three exons among *D. pseudoobscura* and species of the *melanogaster* subgroup. More changes are observed in the third exon than expected at random while fewer changes are observed in the second exon suggesting that purifying selection is limiting replacement substitutions in the NAD⁺ binding and catalytic domains thought to be present in the first and second

FIGURE 5. Hydropathy plots of the various ADH enzymes and the proposed product of the 3' gene in *Drosophila* using an algorithm of HOPP and WOODS (1981). Hydrophilic domains are shown above the zero baseline while hydrophobic domains are below the zero baseline. Amino acid replacements between *D. pseudoobscura* and members of the *melanogaster* species group are mapped on the amino acid sequence. The substrate and NAD⁺ binding domains are predicted to be in the first two exons (32, 39).

exons (BENYAJATI *et al.* 1983; DUESTER, JORNCALL and HATFIELD 1986).

Notably, silent changes are also distributed nonrandomly across the three exons among *D. pseudoobscura* and the *melanogaster* species group representatives (Tables 1 and 2). The silent changes show the same pattern that the replacement changes do, namely the

second exon is deficient in changes and the third exon seems to have an excess of change. This observation seems incompatible with the notion that silent changes are selectively neutral and/or that the rates of mutation are equivalent across the gene.

The nonrandom distribution of silent substitutions across *Adh* seen in our interspecific comparison is

TABLE 3

Nucleotide sequence comparison between *Adh* and the 3' gene of *D. pseudoobscura*

Region	Length (bp) ^a	No. of changes	Percent differences	Corrected percent divergence ^b
Exon 1				
Silent	23.25	22	94.6	>100.0
Replacement	69.75	19	27.2	33.8
Exon 2				
Silent	101.25	79	78.0	>100.0
Replacement	303.75	105	34.6	46.4
Exon 3				
Silent	66.0	49	72.7	>100.0
Replacement	198.0	89	44.9	68.5
Total exons				
Silent	190.5	150	78.7	>100.0
Replacement	571.5	213	37.3	51.6
Total	762	363	47.6	75.5

^a The effective number of silent sites in exons is 25% of the number of nucleotides in the exon (see text for calculation).

^b Corrected percent divergence estimated as $d = 3/4 \ln(1 - 4/3p)$, where p is the proportion of nucleotide sites that differ between the two sequences (JUKES and CANTOR 1969). Estimates over 100% corrected divergence are indicated only as >100%.

unlikely to be a result of hitchhiking of these mutations with selected amino acid replacements. Directional selection will tend to reduce linked neutral polymorphism within species (MAYNARD SMITH and HAIGH 1974; OHTA and KIMURA 1975) while balancing selection will tend to increase it (STROBECK 1980, 1983). However, neither process seems likely to significantly influence the rate of divergence between species since the ultimate probability of fixation of a newly arisen neutral mutation is simply equal to its mutation rate (KIMURA 1983).

The identical relative ranking of exons with respect to intraspecific polymorphism (KREITMAN 1983) and interspecific divergence presented here could indicate that some silent substitutions in *Adh* are not selectively neutral, but are screened by natural selection to differing degrees across the three exons. A silent change could be slightly deleterious if it altered mRNA processing or translation. Analysis of β -thalassaemia mutants in humans has revealed examples of silent and intron substitutions that alter primary transcript processing through the creation of alternative splice sites (e.g., ORKIN *et al.* 1982; GOLDSMITH *et al.* 1983). These variants result in the reduced steady-state mRNA and β -globin levels characteristic of β -thalassaemia. Mutation to (or from) rare codons could also conceivably lead to alterations of *Adh* translation that might be disfavored by selection.

An alternative, though not mutually exclusive, hypothesis to selection to explain the variation in both silent and replacement substitutions across exons of *Adh* is that the underlying rate of mutation varies significantly across the gene. While too little data exist to allow us to comment further on this interesting

idea, variation in the number of silent site substitutions between *Adh* and the 3' gene (detailed below) suggest further exploration of this hypothesis is warranted.

Previous comparisons of *Adh* sequence among members of the closely related *melanogaster* species group had revealed that the frequency of synonymous substitutions in exons was greater than the frequency of changes in introns (KREITMAN 1983; BODMER and ASHBURNER 1984; COHN, THOMPSON and MOORE 1984; COHN 1985). However, we observe the frequency of changes in introns and synonymous sites in *Adh* to be equivalent in our comparison of *D. pseudoobscura* and *D. mauritiana*. The latter result could be due to the saturation of intron and silent sites in this more divergent comparison. In addition, the frequency of changes in introns and noncoding regions may be an under estimate because gaps are assumed in the alignment which may inflate the number of similarities in the two sequences. Despite numerous insertions and deletions, the size of introns has remained roughly equivalent across species groups. The number of insertions, deletions, and nucleotide substitutions make sequence alignments extremely difficult in noncoding regions but clearly serve to highlight regions of putative functional constraint.

The conservation within *D. melanogaster* of sequences 3' to *Adh* described by KREITMAN (1983) can best be explained by the presence of a protein coding gene in this region. Support for the transcriptional activity of this 3' gene includes indirect evidence provided by our DNA sequence comparison of *D. pseudoobscura* and *melanogaster* species group representatives. Silent sites in the putative mRNA differ 24 times more frequently than replacement sites between *D. pseudoobscura* and *D. mauritiana*, for example, a pattern not expected for a pseudogene. In addition, the fine structure of the 3' gene is conserved between *D. pseudoobscura* and members of the *melanogaster* species group. Direct proof of the expression of this gene requires the demonstration of the presence of a transcript. At present, we know only that it is not abundantly transcribed in adult flies. Further studies are underway.

The discovery of a gene just downstream from *Adh* was a considerable surprise given the previous lack of genetic or molecular evidence from *D. melanogaster* for its presence. Inquiry into its origin and possible function led to the similarly surprising result that the 3' gene appears to be an ancient duplication of *Adh*. This conclusion is supported by significant, though low, levels of DNA and amino acid similarity. Perhaps most striking is the similarity of the hydrophobicity pattern of *Adh* and the proposed protein product of the 3' gene (Figure 5) which leaves open the possibility of an enzymatic function similar to ADH.

A rough estimate of the age of the duplication can

be obtained from the nucleotide divergence between the two genes. A relative estimate of the age is that the duplicate is 4.6 times as old as the split between *D. mauritiana* and *D. pseudoobscura* because the corrected percent sequence divergence of *Adh* and the 3' gene is 4.6 times greater than the corrected percent sequence divergence of *Adh* between *D. mauritiana* and *D. pseudoobscura*. The calibration of the time of divergence is difficult due to a general uncertainty of nucleotide substitution rates of *Drosophila*. Estimates range from 0.66% change per million years (ZWIEBEL *et al.* 1982) to 1.3% change per million years (POWELL *et al.* 1986). These give estimates of the time of divergence of the duplication event which range from 114.4 to 58.1 million years ago (mya), respectively. Such a date would predict the presence of both genes in most and perhaps all members of the genus *Drosophila*. These estimates assume, of course, that there has been no concerted evolution since the original duplication of these genes. (We do not consider the silent rate alone since these sites are essentially saturated, being 79% different.)

The frequency of substitutions in *Adh* and the 3' gene are not significantly different. However, the changes are partitioned into silent and replacement substitutions differently between the two loci. This suggests that evolutionary forces (including selection and/or mutation) are not acting on these two genes in an equivalent manner. The proportion of silent substitution in the 3' gene is 2.7 times greater than *Adh* (24 *vs.* 9 times more frequent than replacement substitutions, respectively; Table 1). These data underscore the difficulty of estimating time of divergence from the rate of synonymous substitution in structural genes under assumption of constant and uniform rates of silent substitution. For example, estimation of the time since divergence of *D. pseudoobscura* and *D. mauritiana* vary from 62 to 126 mya based on corrected silent site differences for *Adh* and the 3' gene, respectively [silent rate constant from HAYASHIDA and MIYATA (1983)]. Intron site divergence gives 80 and 66 mya for these comparisons. Comparison of the transcriptional units as a whole yields estimates of 25 and 31 million years for *Adh* and the 3' gene, respectively [rate constant of ZWIEBEL *et al.* (1982)] or 12.6 mya and 15.6 mya [rate constant of POWELL *et al.* (1986)]. While the ZWIEBEL *et al.* (1982) estimates are relatively consistent with the prediction from biogeographical data that the *melanogaster* and *obscura* species groups diverged in the mid-Oligocene (20–25 mya; (THROCKMORTON 1975), the large range on the estimates indicates that they should be viewed with caution. These results also raise some doubt as to the constancy of the rates of substitution, and possibly mutation, between adjacent genes or even regions within the same gene.

The close proximity of *Adh* and the 3' gene suggest

that regulatory sequences for the two genes may overlap. A 16-bp block conserved between *D. pseudoobscura* and *D. mauritiana* at position +1863 is located upstream of the termination of the *Adh* transcript. BENYAJATI and DRAY (1984) found multiple termination sites of the *D. melanogaster Adh* message in tissue culture. This particular conserved sequence has no resemblance to the consensus polyadenylation signal reported in *D. melanogaster* (BENYAJATI and DRAY 1984). It seems possible that this sequence is a regulatory sequence for the downstream gene. Indeed, the 16-bp sequence is located in the appropriate position for a 'CAAT' box.

We can only speculate at present why the 3' gene was not previously uncovered by extensive genetic and molecular studies of this region. Our preliminary transcript studies have not detected an abundant transcript in adults. However, previous molecular work has focused on *Adh*, an extremely abundant mRNA (*e.g.*, GOLDBERG 1980; BENYAJATI *et al.* 1983). A rarer transcript of similar size might have been obscured. In addition, many studies have used very specific, small probes containing only *Adh* sequence. No lethal complementation groups have been localized to this region, either (CHIA *et al.* 1985). The locus may not be lethal mutable by nature of its function, or because its function, when eliminated, is covered by other proteins. These results suggest it will be worth examining other well characterized regions of the *Drosophila* genome for other "hidden" genes using the same strategy of comparing naturally occurring variation within and between species. *Drosophila pseudoobscura* appears to be an ideal choice for comparison with *D. melanogaster* species group members for studies of this type.

We thank the following people for their advice, criticism and discussion: W. ANDERSON, H. HARRIS, R. HUDSON, M. KREITMAN, C. LANGLEY, R. LEWONTIN, E. MONTGOMERY, W. NOON, J. R. POWELL, J. C. STEPHENS and an anonymous reviewer.

LITERATURE CITED

- AQUADRO, C. F., S. F. DEESE, M. M. BLAND, C. H. LANGLEY and C. C. LAURIE-AHLBERG, 1986 Molecular population genetics of the alcohol dehydrogenase gene region of *Drosophila melanogaster*. *Genetics* **114**: 1165–1190.
- ASHBURNER, M., M. BODMER and F. LEMENIER, 1984 On the evolutionary relationships of *Drosophila melanogaster*. *Dev. Genet.* **4**: 295–312.
- BENTON, W. D. and R. W. DAVIS, 1977 Screening lambda-gt recombinant clones by hybridization to single plaques *in situ*. *Science* **196**: 180–182.
- BENYAJATI, C. and J. F. DRAY, 1984 Cloned *Drosophila* alcohol dehydrogenase genes are correctly expressed after transfection into *Drosophila* cells in culture. *Proc. Natl. Acad. Sci. USA* **81**: 1701–1705.
- BENYAJATI, C., N. SPOEREL, H. HAYMERLYE and M. ASHBURNER, 1983 The messenger RNA for alcohol dehydrogenase in *Drosophila melanogaster* differs in its 5' end in different developmental stages. *Cell* **33**: 125–133.
- BIGGINS, M. D., T. J. GIBSON and G. F. HONG, 1983 Buffer

- gradient gels and ^{35}S label as an aid to rapid DNA sequence determination. *Proc. Natl. Acad. Sci. USA* **80**: 3963–3965.
- BINGHAM, P. M., R. LEVIS and G. M. RUBIN, 1981 Cloning of DNA sequences from the white locus of *D. melanogaster* by a novel and general method. *Cell* **25**: 693–704.
- BODMER, M. and M. ASHBURNER, 1984 Conservation and change in the DNA sequences coding for alcohol dehydrogenase in sibling species of *Drosophila*. *Nature* **309**: 425–430.
- CHAMBERS, G. K., J. F. McDONALD, M. McELFRESH and F. J. AYALA, 1978 Alcohol-oxidizing enzymes in 13 *Drosophila* species. *Biochem. Genet.* **16**: 757–767.
- CHIA, W., R. KARP, S. MCGILL and M. ASHBURNER, 1985 Molecular analysis of the *Adh* region of the genome of *Drosophila melanogaster*. *J. Mol. Biol.* **186**: 689–706.
- COHN, V. H., 1985 Organization and evolution of the alcohol dehydrogenase gene in *Drosophila*. Ph.D. dissertation, University Microfilms, Ann Arbor, Mich.
- COHN, V. H., M. A. THOMPSON and G. P. MOORE, 1984 Nucleotide sequence comparison of the *Adh* gene in three *Drosophilids*. *J. Mol. Evol.* **20**: 31–37.
- COYNE, J. A. and M. KREITMAN, 1986 Evolutionary genetics of two sibling species, *Drosophila simulans* and *D. sechellia*. *Evolution* **40**: 673–691.
- DUESTER, G., H. JORNVALLE and G. W. HATFIELD, 1986 Intron-dependent evolution of the nucleotide-binding domains within alcohol dehydrogenase and related enzymes. *Nucleic Acids Res.* **14**: 1931–1941.
- FLAVELL, A. J., S. W. RUBY, J. J. TOOLE, B. E. ROBERTS and G. M. RUBIN, 1980 Translation and developmental regulation of RNA encoded by the eukaryotic transposable element *copia*. *Proc. Natl. Acad. Sci. USA* **77**: 7107–7111.
- FRISCHAUF, A., H. LEHRACH, A. POUSTKA and N. MURRAY, 1983 Lambda replacement vectors carrying polylinker sequences. *J. Mol. Biol.* **170**: 827–842.
- GEMMILL, R. M., J. N. LEVY and W. W. DOANE, 1985 Molecular cloning of α -amylase genes from *Drosophila melanogaster*. I. Clone isolation by use of a mouse probe. *Genetics* **110**: 299–312.
- GOLDBERG, D. A., 1980 Isolation and partial characterization of the *Drosophila* alcohol dehydrogenase gene. *Proc. Natl. Acad. Sci. USA* **77**: 5794–5798.
- GOLDSMITH, M. E., R. K. HUMPHRIES, T. LEY, A. CLINE, J. A. KANTOR and A. W. NIENHUIS, 1983 "Silent" nucleotide substitution in a beta-thalassemia. *Proc. Natl. Acad. Sci. USA* **80**: 2318–2322.
- GRELL, E. H., K. B. JACOBSEN and J. B. MURPHY, 1965 *Adh* in *D. melanogaster*: isozymes and genetic variants. *Science* **149**: 80–82.
- HAYASHIDA, H. and T. MIYATA, 1983 Unusual evolutionary conservation and frequent DNA segment exchange in class I genes of the major histocompatibility complex. *Proc. Natl. Acad. Sci. USA* **80**: 2671–2675.
- HENIKOFF, S., 1984 Unidirectional digestion with exonuclease III creates targeted breakpoints for DNA sequencing. *Gene* **28**: 351–359.
- HOLMQUIST, R., C. CANTOR and T. JUKES, 1972 Improved procedures for comparing homologous sequences in molecules of proteins and nucleic acids. *J. Mol. Biol.* **64**: 145–161.
- HOPP, T. P. and K. R. WOODS, 1981 Prediction of protein antigenic determinants from amino acid sequences. *Proc. Natl. Acad. Sci. USA* **78**: 3824–3828.
- JUKES, T. H. and C. R. CANTOR, 1969 Evolution of protein molecules. pp. 21–132. In: *Mammalian Protein Metabolism*, Vol. 3, Edited by H. N. MUNRO. Academic Press, New York.
- KIMURA, M., 1983 *The Neutral Theory of Molecular Evolution*. Cambridge University Press, New York.
- KREITMAN, M., 1983 Nucleotide polymorphism at the alcohol dehydrogenase locus of *Drosophila melanogaster*. *Nature* **304**: 412–417.
- LANGER, P. R., A. A. WALDROP and D. C. WARD, 1981 Enzymatic synthesis of biotin-labeled polynucleotides: novel nucleic acid affinity probes. *Proc. Natl. Acad. Sci. USA* **78**: 6633–6637.
- LANGLEY, C. H., E. A. MONTGOMERY and W. F. QUATTLEBAUM, 1982 Restriction map variation in the *Adh* region of *Drosophila*. *Proc. Natl. Acad. Sci. USA* **79**: 5631–5635.
- LEWONTIN, R. C., 1985 Population genetics. *Annu. Rev. Genet.* **19**: 81–102.
- LIPMAN, D. J. and W. R. PEARSON, 1985 Rapid and sensitive protein similarity searches. *Science* **227**: 1435–1441.
- MANIATIS, T., D. F. FRITSCH and J. SAMBROOK, 1982 *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
- MAYNARD SMITH, J. and J. HAIGH, 1974 The hitchhiking effect of a favorable gene. *Genet. Res.* **23**: 23–35.
- McDONELL, M. W., M. N. SIMON and F. W. STUDIER, 1977 Analysis of restriction fragments of T7 DNA and determination of molecular weights by electrophoresis in neutral and alkaline gels. *J. Mol. Biol.* **110**: 119–146.
- NEEDLEMAN, S. B. and C. D. WUNSCH, 1970 A general method applicable to the search for similarities in amino acid sequence of two proteins. *J. Mol. Biol.* **48**: 443–453.
- NEVINS, J. R., 1983 The pathway of eucaryotic mRNA formation. *Annu. Rev. Biochem.* **52**: 441–466.
- NORRANDER, J., T. KEMPE and J. MESSING, 1983 Construction of improved M13 vectors using oligonucleotide-directed mutagenesis. *Gene* **26**: 101–106.
- OHTA, T. and M. KIMURA, 1975 The effect of a selected linked locus on heterozygosity of neutral alleles (the hitch-hiking effect). *Genet. Res.* **25**: 313–326.
- ORKIN, S. H., H. H. KAZAZIAN, JR., S. E. ANTONARAKIS, H. OSTRER, S. C. GOFF and J. P. SECTON, 1982 Abnormal RNA processing due to the exon mutation of beta E-globin gene. *Nature* **300**: 768–769.
- PARDUE, M. L. and J. G. GALL, 1975 Nucleic acid hybridization to the DNA of cytological preparations. *Methods Cell Biol.* **10**: 1–16.
- POWELL, J. R., A. CACCONE, G. D. AMATO and C. YOON, 1986 Rates of nucleotide substitution in *Drosophila* mitochondrial DNA and nuclear DNA are similar. *Proc. Natl. Acad. Sci. USA* **83**: 9090–9093.
- PUSTELL, J. and F. C. KAFATOS, 1986 A convenient and adaptable microcomputer environment for DNA and protein sequence manipulation and analysis. *Nucleic Acids Res.* **14**: 479–488.
- RIGBY, P. W. J., M. DIECKMAN, C. RHODES and P. BERG, 1977 Labeling deoxyribonucleic acid to high specific activity in vitro by nick translation with DNA polymerase I. *J. Mol. Biol.* **113**: 237–251.
- SANGER, F., S. NICKLEN and A. R. COULSON, 1977 DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA* **74**: 5463–5467.
- SCHAEFFER, S. W., C. F. AQUADRO and W. W. ANDERSON, 1987 Restriction map variation in the alcohol dehydrogenase region of *Drosophila pseudoobscura*. *Mol. Biol. Evol.* **4**: 254–265.
- SMITH, G. E. and M. D. SUMMERS, 1980 The bidirectional transfer of DNA and RNA to nitrocellulose or diazobenzylomethyl paper. *Anal. Biochem.* **109**: 123–129.
- SOUTHERN, E. M., 1975 Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J. Mol. Biol.* **98**: 503–517.
- STEPHENS, J. C. and M. NEI, 1985 Phylogenetic analyses of polymorphic DNA sequences at the *Adh* locus in *Drosophila melanogaster* and its sibling species. *J. Mol. Evol.* **22**: 289–300.
- STROBECK, C., 1980 Heterozygosity of a neutral locus linked to a self-incompatibility locus or balanced lethal. *Evolution* **34**: 779–788.
- STROBECK, C., 1983 Expected linkage disequilibrium for a neutral

- locus linked to a chromosomal arrangement. *Genetics* **103**: 545–555.
- STURTEVANT, A. H. and E. NOVITSKI, 1941 The homologies of the chromosome elements in the genus *Drosophila*. *Genetics* **26**: 517–541.
- THROCKMORTON, L. H., 1975 The phylogeny, ecology and geography of *Drosophila*. pp. 421–469. In: *Handbook of Genetics*, Vol. 3, Edited by R. C. KING. Plenum Press, New York.
- WICKENS, M. and P. STEPHENSON, 1984 Role of the conserved AAUAAA sequence: four AAUAAA point mutants prevent messenger RNA 3' end formation. *Science* **226**: 1045–1051.
- WILBUR, W. J. and D. J. LIPMAN, 1983 Rapid similarity searches of nucleic acid and protein data banks. *Proc. Natl. Acad. Sci. USA* **80**: 726–730.
- ZWIEBEL, L. J., V. H. COHN, D. R. WRIGHT and G. P. MOORE, 1982 Evolution of single-copy DNA and the *Adh* gene in seven *Drosophilids*. *J. Mol. Evol.* **19**: 62–71.

Communicating editor: J. R. POWELL