J Mol Evol (1988) 28:43–54

Journal of Molecular Evolution

Nucleotide Sequence of the Genomic Region Encoding Alcohol Dehydrogenase in *Drosophila affinidisjuncta*

Robert G. Rowan and W.J. Dickinson

Department of Biology, University of Utah, Salt Lake City, Utah 84112, USA

Summary. The DNA sequence of a 3886-bp genomic region containing the alcohol dehydrogenase (*Adh*) gene from *Drosophila affinidisjuncta,* and the RNA sequences of the *D. affinidisjuncta Adh* transcripts, are presented. These data support the conclusion that two *Adh* promoters generate distinct, developmentally regulated *Adh* transcripts. Correlations between these sequences and the transcription map are discussed. Comparisons between these and equivalent data from *D. melanogaster* are also presented. We note the following observations: (1) Except at the extreme 3′ end, the two genes are identically organized. (2) *Drosophila Adh* protein accumulates amino acid replacements at the rate of approximately 0.5 per million years. (3) Among the non-protein-coding DNA sequences, putative homologies occur in the two promoter regions.

Key words: Alcohol dehydrogenase — Hawaiian picture-winged *Drosophila* — Gene promoters — DNA sequence evolution

Introduction

Drosophila alcohol dehydrogenase (ADH) provides an interesting model system for the analysis of gene regulation (e.g., Dickinson and Carson 1979; Maroni et al. 1982; Batterham et al. 1983; Goldberg et al. 1983; Fischer and Maniatis 1986) and also for studies on gene organization (Benyajati et al. 1983; Fischer and Maniatis 1985; Rowan et al. 1986), gene evolution (Kreitman 1983; Bodmer and Ashburner 1984), and ecological genetics (e.g., McKenzie and Parsons 1972; McKenzie 1980; David and Van Herrewege 1983). In particular, the developmental expression (Ursprung et al. 1970; Savakis et al. 1985) and the molecular organization (Benyajati et al. 1981, 1983) of the *Drosophila melanogaster* ADH gene (*Adh*) have been well characterized. Previous work from this laboratory has analyzed the expression of ADH activity in many members of the Hawaiian picture-winged group of *Drosophila* (Dickinson 1980a), and the *Adh* gene from one of these, *Drosophila affinidisjuncta,* has also been studied at the molecular level (Brennan et al. 1984a; Rowan et al. 1986). Although they are distantly related, *D. affinidisjuncta* and *D. melanogaster* exhibit similar patterns of ADH expression (Rowan et al. 1986). In contrast, ADH is expressed with different developmental specificities in some other Hawaiian picture-winged species (Dickinson 1980a; Rowan and Dickinson 1986).

Several data imply that DNA sequences near the *Adh* structural gene are important for the developmental regulation of ADH activity. Interspecific differences in ADH expression patterns within the Hawaiian *Drosophila* are *cis*-dominant in species hybrids, suggesting that regulatory elements are closely linked to the enzyme locus (Dickinson and Carson 1979; Dickinson 1980b). In *D. melanogaster,* gene transformation experiments have show that correct expression can be obtained with an 11,800-bp fragment of cloned *Adh* genomic DNA (Goldberg et al. 1983). Also, similar experiments show that a 5400-bp *D. affinidisjuncta* genomic *Adh* clone is expressed with very nearly correct accuracy in transgenic *D. melanogaster* (Brennan and Dickinson 1988). Here we present the DNA sequence of the *D. affinidisjuncta Adh* genomic region and the corresponding mRNA and ADH amino acid se-

Offprint requests to: W.J. Dickinson

quences. We then compare these sequences with those of *D. melanogaster*. These data demonstrate the similarity of these *Drosophila* genes and identify genomic sequences that may regulate *Adh* expression. The evolution of the *Adh* locus is also discussed.

Materials and Methods

DNA Sequencing. The molecular cloning (Brennan et al. 1984a) and characterization (Rowan et al. 1986) of the *D. affinidisjuncta Adh* genomic region have been reported. A total of 3886 contiguous base pairs of DNA from the *Adh*-containing bacteriophage lambda clone S36G1-11B (Brennan et al. 1984a) was subcloned as fragments from restriction enzyme digestions into the vectors m13mp8 and/or m13mp9 (Messing and Vieira 1982). These clones were sequenced by the dideoxynucleoside chain termination procedure (Sanger et al. 1977) according to the map presented in Fig. 1. In all but two locations, both DNA strands were sequenced. For the two exceptions, the data obtained by the chain termination method were confirmed by chemical cleavage sequencing (Maxam and Gilbert 1980) of DNA subcloned into the vector pUC8 (Vieira and Messing 1982). Enzymes and oligonucleotide primers were purchased from New England Biolabs. Deoxynucleoside triphosphates and dideoxynucleoside triphosphates were purchased from Boehringer Mannheim Biochemicals and from P.L. Biochemicals, respectively. Radiolabeled nucleoside triphosphates were purchased from New England Nuclear.

Denaturing polyacrylamide gels for DNA sequencing were prepared and used according to Maxam and Gilbert (1980), except that gels were bound to one glass plate of the gel former and processed for autoradiography according to Garoff and Ansorge (1981).

RNA Sequencing. RNA sequencing by chain termination (Hamlyn et al. 1978) was performed as follows. ^{32}P-end-labeled DNA fragments (primers) were obtained from genomic *Adh* clones and annealed to *Adh* RNA at 49°C as previously described (Rowan et al. 1986). Positions of the DNA primers are given in Fig. 1 (open arrows) and in the Results section. Each annealing contained either 20 μg of poly(A)-selected RNA from adult flies (for sequencing around intron 1) or 10 μg of poly(A)-selected RNA from larvae (for sequencing around introns 2 and 3). Annealed nucleic acids were purified by ethanol precipitation and resuspended in 125 mM NaCl, 25 mM Tris-HCl (pH 8.3), 10 mM $MgCl_2$, and 5 mM dithiothreitol, which contained 100 μM each of dATP, dGTP, dTTP, and dCTP (120 μl total volume). The four corresponding dideoxynucleoside triphosphates were added separately to four 30-μl aliquots of the above reaction mixture, to a final concentration of 10 μM (Fig. 3A) or 40 μM (Fig. 3B and C). Reactions with 10 units of reverse transcriptase (Seikagaku America) were incubated at 42°C for 10 min and terminated by phenol extraction. After recovery by ethanol precipitation, the nucleic acids were resuspended in formamide, heated to 65°C for 10 min, and analyzed on sequencing polyacrylamide gels.

The isolation of *D. affinidisjuncta* RNA from whole organisms (Brennan et al. 1984a) and from larval or fly tissues (Rowan and Dickinson 1986) has been described. Feeding third-instar larvae or adults aged 1–3 weeks posteclosion were used in this study.

Primer Extension Analysis. DNA fragments (primers) were ^{32}P-end-labeled with DNA polymerase large fragment (Maniatis et al. 1982) using α-[^{32}P]dATP or α-[^{32}P]dCTP (New England Nuclear) and purified by polyacrylamide gel electrophoresis. Primers were extended on RNA templates with reverse transcriptase as described above, with the omission of dideoxynucleoside triphosphates. Polyadenylated RNA (0.5 μg) from larvae (Fig. 4A) or adults (Fig. 4B) and the total RNA from the equivalent of two adult abdominal fat body samples, two adult heads, and the Malpighian tubules from 25 adults (Fig. 4C) were analyzed. The same primers were also extended on complementary DNA (cDNA) templates (m13 genomic subclones) in typical DNA sequencing reactions to yield markers for the relevant genomic regions. For the analysis of larval *Adh* transcripts, a 608-bp *Eco*RI–*Hin*dIII fragment of cloned genomic DNA (coordinates 1986–2594 in Fig. 2) was used. Synthesis of adult *Adh* cDNA was primed with a 1522-bp *Sal*I–*Hin*dIII genomic DNA fragment (coordinates 1432–2594 in Fig. 2). These large primers were more easily prepared and more efficiently extended than smaller DNA fragments (unpublished observations), but are too large for accurate sizing by gel electrophoresis. Therefore, the labeled extension products were cleaved from the primer sequences by restriction enzyme digestion as described by Rowan et al. (1986). Only extended primers contain a regenerated restriction enzyme recognition site at their 3′ end, and the radiolabel is cleaved from the primer and retained by the extension as a result of digestion. In the present instance, cleavage of adult *Adh* cDNA with *Sal*I and of larval *Adh* cDNA with *Eco*RI results in appropriately sized ^{32}P-labeled cDNAs (Fig. 4). DNA template-directed sequencing reactions were phenol extracted, ethanol precipitated, and digested with either *Sal*I or *Eco*RI. Digested DNAs were ethanol precipitated, resuspended in formamide, heated to 65°C for 10 min, and analyzed by electrophoresis on sequencing gels.

Data Analysis. An unpublished sequence of the *D. melanogaster Adh* locus was provided by M. Ashburner. This is a composite sequence derived from two sources: 2540 bp, inclusive from 473 bp 5′ to the distal *Adh* mRNA sequence to 211 bp 3′ to the polyadenylation site, was determined from the Canton-S clone psAC1 (Goldberg 1980) by H. Haymerle and M. Ashburner (unpublished). Contiguous with this sequence is an additional 4293 bp of 5′-flanking and 590 bp of 3′-flanking sequence (M. Aguade and M. Kreitman, unpublished; Kreitman 1983) of an African isolate contained in the clone Af-S (Kreitman 1983). All of this composite sequence was compared to the *D. affinidisjuncta* sequence presented in Fig. 2. The similarity of the *D. affinidisjuncta* protein coding region to the corresponding *D. melanogaster* sequence was obvious (see Results). Flanking DNA sequences were analyzed for similarity as follows. A computer program was used to locate many short blocks of intersequence similarity, which were then evaluated manually. In general, the acceptable alignments were those that, when taken together, produced a larger alignment with the individual members occurring in the same relative order in each species' genome (i.e., only apparent deletions/insertions, and not transpositions, were tolerated; no relative inversions were indicated). This parsimonious approach might have neglected some biologically meaningful similarities, but did remove many similarities that are probably fortuitous.

Results

The *D. affinidisjuncta Adh* transcription map is stage-specific (Fig. 1). Most of the *Adh* RNA from adult flies contains a small 5′ exon (exon 1) that is separated from the rest of the transcript by a relatively large intron (intron 1). Larval *Adh* transcripts lack exon 1 sequences, and intron 1 sequences contribute to the 5′ end of larval *Adh* RNA (see Fig. 1). These two RNA 5′ end structures account for all of the

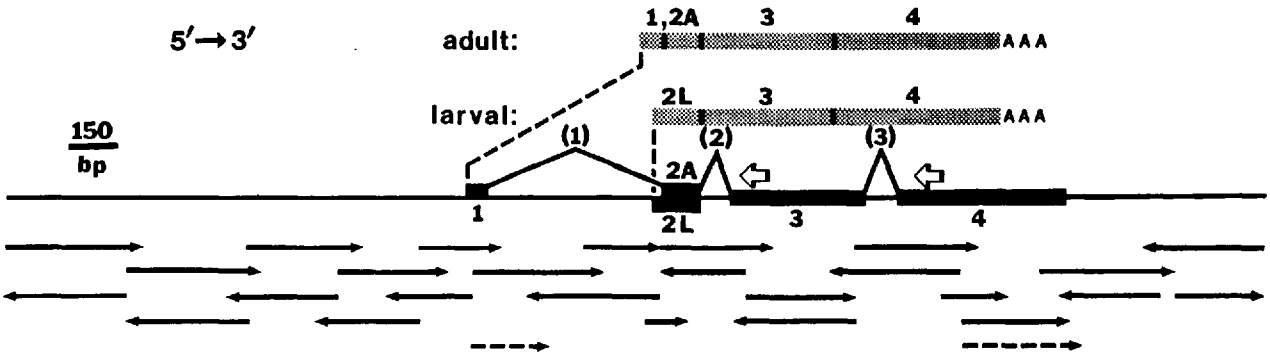

Fig. 1. Transcription map of the *D. affinidisjuncta Adh* gene (Rowan et al. 1986) and DNA sequencing strategy. The polarity with reference to *Adh* mRNA (5′-3′) and size scale (bar = 150 bp) are shown. Solid bars represent the genomic locations of the exons in adult (1, 2A, 3, 4) and in larval (2L, 3, 4) *Adh* mRNA. Shaded bars show the adult and larval *Adh* mRNA structures, referred to as the "distal" and "proximal," respectively, transcripts. Dashed lines locate the 5′ end of each mRNA to its genomic position. The positions of introns [(1), (2), and (3)] are indicated. Open arrows mark the priming sites used in *Adh* RNA sequencing (see Fig. 3). For simplicity, only one of several polyadenylation sites (AAA) is shown. Below the transcript map, solid arrows designate individual sequences obtained by the Sanger method from m13 genomic subclones. Dashed arrows designate sequences obtained by the Maxam and Gilbert method.

detectable postembryonic *Adh* transcripts in *D. affinidisjuncta* (Rowan et al. 1986). Introns 2 and 3 (Fig. 1) are common to all detectable *Adh* transcripts in both larvae and adults. For brevity, the transcripts are referred to as "distal" (containing the distal exon 1) or "proximal" (lacking exon 1) in this report.

The nucleotide sequence of the *D. affinidisjuncta Adh* region is presented in Fig. 2. *Adh* RNA transcripts, previously located relative to restriction enzyme recognition sites (Rowan et al. 1986; Fig. 1), are more accurately mapped to the genomic DNA sequence by the following data. *Adh* RNA sequences for the regions interrupted by introns 1 and 2 were obtained using a 121-bp *Hin*cII–*Taq*I DNA fragment (coordinates 2231–2352 in Fig. 2; open arrow over exon 3 in Fig. 1) as the sequencing primer. The alignment of these sequences (Fig. 3A and B) with the genomic DNA sequence is obscured by a dinucleotide reiteration in the donor and acceptor splice sequences of both intron 1 (GT reiterated) and intron 2 (AG reiterated). In each case, however, only one of the two possible alignments will yield donor and acceptor sequences that obey the well-established consensus rules (Mount 1982). These define coordinates for intron 1 (1473–2003) and for intron 2 (2126–2206) as indicated in Fig. 2. Intron 3 was located by sequencing *Adh* RNA using a 132-bp *Ava*I–*Bgl*II DNA fragment (coordinates 2728–2860 in Fig. 2; open arrow over exon 4 in Fig. 1). The implied sequence of intron 3, indicated by the RNA sequence (Fig. 3C), is unambiguous (coordinates 2612–2685 in Fig. 2) and shows donor and acceptor splice sequences that obey consensus rules.

The 5′ ends of *D. affinidisjuncta Adh* RNAs were located by primer extension experiments as shown in Fig. 4. Both transcript types map as expected (Rowan et al. 1986), but appear heterogeneous over several nucleotides. The 5′ terminal nucleotides of proximal transcripts (Fig. 4A; "p"s in Fig. 2) are the same in each larval tissue that expresses *Adh* RNA (data not shown). Distal *Adh* transcripts occur in the adult head, fat body, and Malphigian tubules (Rowan and Dickinson 1986). There are four apparent 5′ distal cap sites (Fig. 4B; "d"s in Fig. 2), the most distal of which is not utilized in the adult head (Fig. 4C).

Features of the D. affinidisjuncta Adh Region

Available data indicate that distal and proximal *Adh* transcripts are produced from a single copy gene (Brennan et al. 1984a; Rowan et al. 1986). Transcription of this gene has not been studied directly, but aspects of the *Adh* DNA sequence argue that these two RNAs arise from different transcriptional events, rather than from differential processing of a single primary transcript. First, genomic "TATA box" sequences, centered at 1387 and at 1945 (Fig. 2, "box"), are located just upstream from the apparent 5′ end of each transcript. Also, donor and acceptor intron splicing sequences (Mount 1982) in the predicted distal RNA primary transcript accurately predict the 5′ end structure of the observed transcript. Thus, like the *D. melanogaster Adh* gene (Benyajati et al. 1983), the *D. affinidisjuncta Adh* gene appears to be expressed from two distinct promoters.

Drosophila affinidisjuncta Adh RNA 3′ ends are polyadenylated and map to several locations within a 250-bp genomic region. The observed 3′ ends in larval RNA (Fig. 2, "lv." and "+") and in adult RNA (Fig. 2, "ad." and "+") differ from each other at some positions (Rowan et al. 1986). The more proximal of these termini are not correlated with the presence of generally recognized polyadenyla-

```
AAGCTTCAAGAAGATGACGTTGAAATGTTGCAATCGTTGATCCAAGCCTCCACAGTTTACTTCACTTTTA   70
TTGCACACCTCGTCTCCTCTCCCATCAACTTACAACAATAACTATTGTAATATGTATAGGTTTCTGTGCA  140
TGTATTGTGAGAGAGTGAGGTGTGTCAAGTGTCGAGTGTGTGTGAGCTGCCACGTTTTCAAAATCAAATT  210
GAAACTTAATTTTCACCTGAGTGCTCTTTTATAGAGCACAAAACAGAAATTAGGAATTATTTAAAACTGG  280
GGCCAAGTGGATTTTTATGACCAGATAAAGATAAGGTACAATATAAATCGGCTCATATTTTTGTAGTTTC  350
TTGATCAAGCTAATTTAATACAATTAGAATTCTAGAATTTATAGATTAATAAATTGGACTTTCTTTGTTG  420
TTGTTAAATAGCTGTAATGTAATCTATTGATAATTATAAAACATTGCTAAAGGCAAACAGCAAACATTTT  490
AAGAACCCTAACTTGTCGTCGGATTAAGAAGCGTGGCGACTTATATAATAGATTGACTACAAAGTGTGTT  560
TTATTTCGAGTTTCGTTGACAATCAAATTCAATATGTGTGGCAATAGCGCAGAATAGGTCAGTAATTGTC  630
TGCGCCTCGCTCTCACGTCACAATTCATGGTGATTATGTATAATCAGCAACTCGTCTGATTGCGGTTAAG  700
AAGCGACCCATTCCTAACAAGAGTGATAACCCGGACTAGTGTGACATGCCTGATATCAGGCCATATTGGC  770
GTCAGACGTCAGAGGTATTTTGTGGTATATTGTATATAAGTCAGGTTCTGCTAATCATTTCATAAATGAA  840
ATGCATTTATTCTTATTCTACTTGAATCTGTTATTATACCCATATTAATATTATATTTATGTTGCCTATG  910
TTTTTTATAGTAACAACGTAGAATCATATATAATGATTACTCTTTCACTTCCTGTTCTAATCAAGATACT  980
TGTCAAAACTCAATTGAATAATCTGAATGAATTCTTAATGTATATGTATAAATTACATAACTTGTATCAA 1050
TATGGTATACTACCTAGTGAAAAATGTACGTAAATGCCATATACTTGTAATACCACTCCCCCACTTCTCG 1120
CATTCTCGGCGCCGTCAGTCAATTTGAATAATAATGCAAATAATGCAATCTGCGAACATAACAACAGCAA 1190
CAACAACAATATAAGCCGACTAATATTGCTTTGACATTCAGTGAAGGTCGAGGCTTCAAGTGTGTTCCGC 1260
TCTCTGCAGCTCAGTTTCATCTCTATCTTATTAATAACCTCCCCTATTTGTTGTGTTATATTCAAAGTAG 1330
                                                    ( box )
AGGTTTACATCTCTTAAGCATTCACGATTTCAACATCGACTCGTGTGTGTAAGTATAAAAGGTGTTACTC 1400
            (ddd  d>)
GTATTGAGAGCACTGATTATTGCCATAGCAGTCGACGCACGCAGATCGGACCAGTTACAATATTTTCAAA 1470
  (intron 1
AAGTAAGTGCAGTTCAGACAGTCCAAACCAGAACTGGCGCGTCGGGCAAAAATAATTTAATCTGAATTTT 1540
CTGGCAAAAGTTTCCACGGGCAAAGGCTGATAAACAAATGAATAAAAGATTAAAATAAGCAAAATATACA 1610
CACGTACATAAACATAAAAAATATGTCTGTATGTGCATGTACATCCAAGAATATATATTCTAGGTTTTAT 1680
TATTTGTTTGGTGATATGACCTTCAATATGTTAAATAATGCTGGACAAAATTAGAATACCCCAATTGAAG 1750
AGCATATAAAAAATACACATATGTACCTAAATTAAGCAAGTAGAATGAGAAGCGCCGGCAATATTTGACT 1820
GCAATAACCAAAACAACAAATCAAAATAAAAAAAAGTCAAGACCACGAGCAGTGCTGATGATCGCCAACA 1890
                                                    ( box )
TTACTGATAAACAGCTGCGAGGAAATACTCAAACTTAGCAGACTGTTCGCCTATAAATAGAAGGTATCAA 1960
        (pppp>)                    intron 1)
CCGGCCAAACTCATATAGTGTTTTGAATTCTCTGTCTGAACAGGTGAACAGAGTTGAGGGTATCGCTGAA 2030
 (*** start)
AAATGGTTATCGCTAACAGTAACGTCATCTTTGTCGCTGGTCTGGGTGGCATTGGCCTGGACACCAGTCG 2100
                    (intron 2
CGAGATTGTCAAGAGTGGCCCCAAGGTAGGTTACAGCCTGTCAATCTCAGGAAAAAGTATGGAAGAATAA 2170
                      intron 2)
AATAAAATTATTATTATTTTTTTTTTTTCGTTTAGAACCTGGTGGTGCTTGACCGCGTTGACAACCCCG 2240
CTGCCATTGCCGAGTTGAAAGCCCTTAATCCCAAGGTGACCGTTACATTCTATCCGTATGATGTAACCGT 2310
GCCGTTGGCTGAAACCAAAAAGCTATTGAAGACAATCTTCGACAAGCTGAAGACAGTTGATCTGCTCATC 2380
AATGGCGCTGGTATCCTGGACGATAATCAGATTGAGCGCACAATTGCGGTGAATTTTACAGGTACAGTGA 2450
ACACCACAACTGCGATCATGGATTTCTGGGACAAGCGTAAGGGCGGCCCAGGTGGTGTTGTTGCCAACAT 2520
TTGCTCAGTAACTGGATTCAACTCCATCTATCAGGTGCCCGTTTACTCTGCCTCAAAAGCGGCTGCTCTA 2590
                    (intron 3
AGCTTCACCACTTCCATTGCGGTAAGTAAATATACATATATATATATATATATATACTTATACTTCATCA 2660
                intron 3)
GTGACTACCTATTGAATCTTAACAGAAATTGGCGCATATTACTGGCGTTACGGCATATTCCATCAACCCG 2730
GGCATTACCAAGACTGTTCTGGTGCATAAATTCAACTCCTGGCTGAGTGTTGAGCCACGCGTTGCCGAGC 2800
TACTGCTTGAGCACCCCACACAGACAACATTGCAGTGCGCACAGAACTTTGTGAAGGCCATTGAGGCCAA 2870
CCAGAATGGTGCCATCTGGAAACTGGATCTTGGCCGTCTGGATGCAATCGAATGGACCAAGCACTGGGAC 2940
   (stop ***)
TCAGGCATCTAAACTGTGCATGAAATCCGTACAAGGAGTGAAACTGCATTTAACATACCATTGTCAAAAT 3010
                                                  polyA?          polyA?
```

Fig. 2. Nucleotide sequence of the mRNA-like strand in the 5′ to 3′ orientation of the *D. affinidisjuncta Adh* genomic region. A *Hin*dIII recognition sequence (AAGCTT) at position 2590 identifies the center (position 0) of the published restriction enzyme map (Brennan et al. 1984a; Rowan et al. 1986). The sequence of the distal *Adh* mRNA, extending from the most distal 5′ cap site to the most highly expressed polyadenylation site (see Rowan et al. 1986), is underlined. Notations above the sequence: "box" indicates TATA sequences; "d" and "p" identify the cap sites of distal and proximal, respectively, *Adh* mRNAs; the locations of introns 1, 2, and 3; asterisks identify the translational start and stop codons; "lv." and "ad." identify polyadenylation sites unique to larvae and adults, respectively; "+" indicates a polyadenylation site common to both larvae and adults. Below the sequence, "polyA?" and "polyA" indicate possible and known, respectively, polyadenylation signal sequences; two overlapping signals are indicated by "polyAs." The sequence from 3127 to 3652 is presented in an alignment with the sequence from 1501 to 2031, the latter in lowercase letters. Within this alignment, a blank space in either sequence indicates the absence of a nucleotide at that position, sequence identity is indicated by the dashed line, and sequence differences are explicitly indicated.

tion signal sequences (reviewed by Birnstiel et al. 1985). Other A/T-containing hexanucleotide sequences (Fig. 2, "polyA?") do occur at appropriate positions, and might therefore be functional signals for this gene. Distal 3′ termini follow recognizable polyadenylation signals (Fig. 2, "polyA") by 5–41 nucleotides. For comparison, the consensus value for this distance has been given as a range of 10–30 nucleotides (Birnstiel et al. 1985).

Interestingly, 525 bp of DNA on the 3′ side of the ADH coding region is apparently a direct, imperfect duplication of the sequence immediately 5′ to the coding region. Most of intron 1 and all of the (proximal) mRNA leader sequence, but none of the ADH codons, are duplicated. Blot hybridization data (Brennan et al. 1984a and unpublished) also indicate that the entire *Adh* region is not duplicated. This duplication is displayed in Fig. 2 by aligning the

```
                lv.                    ad.          lv.  +  +
TTGATAAGAAGTATATATTCATTCAAAACACCGTTATGTACGAAAGCATAATTTCTTTTCTTCTGTTTAT 3080
           polyA?
        lv.       ad.lv.  ad.
AATTAAAAATAGCATGGCCTCTAATTAATGAAATTAACACAACAAAAAAGTGGCGCGTAGGGCTAAACAA 3150
polyAs                  polyA    polyA    (1501 g--c--------c----a---a--
     lv. ad.                +
ATAATAAATCTGAATTTTCTGTCAAAC TTTCC CGAGCAAAGGCTGATAAAACAAATAAACATTTAAGC 3218
ta-t-----------------g----ag-----a--g--------------- -----g--t-aaag-
polyAs                                                  polyA
                    +                          +
ATACAAATTTGCAAAATAATACAAATCTTTTCATCTAATAGTGGCAACTAATGTCTGTATGTGTATTTGC 3288
 -ta----aa-------- ----c-cgtacat-aac-ta-aaa     at-------------c--g-a-
              polyA
ATCCAAGAATATCTATTCCAGGTTTTATTATTAGATTGGTGATATGACCTTCAATATGTTTAAT       3352
------------a-----t-------------t-t-------------------------a---aatgct

    AAATTAGAATACCCCAATAGCAGAGTATATAAATAATATAAATATGTATGTATGCATCTAAATTAA 3418
ggac-----------------t-a----c-------a----c-c-------       c---------

GCAAGTAGAATGAGAAGAGCCGGTAAAATTTGACTGGAGTAGCAAAAACAACACATTAAAAGAAAAAAAA 3488
----------------c-----c--t---------c-a--a-c---------a--c----t--------

 CAAATACCACGAGCAGTGCTGATGATCACCAACGTGAGCGATAAACAGCAGCGAGTAAATACGCAA    3554
gtc--g---------------------g-----a-t-ct----------t-----g------t---act

       TACCTTCTAAAAAATAGAAGGTATCAGCAGGCCAAACTTATATAGTGTTTTGAATTCTCTGT 3616
tagcagac-gttcg-ct-t---------------a-c---------c----------------------

CCGAACAGGTGAACAGAGTTGAATGCATCACTGAAACTATATATAAAGTCATACATTTTAAATAAAAGTA 3686
-t--------------------gg-t---g------ 2031)
TATTTAAAGTAGTATATAGTAAATAATATAATGAAATTTAGCATATTTCATGGTATTTAAAAAAGAAGAC 3756
ACGGTATATTAGACTCGCAAGATTACATACCCTTGTATATGGAGGATAATATGTCAACATTTAACAACAT 3826
TTAATTTTGGTTCGGCTTGCGTTTTCTGTTTAACGCCTGGCTGTAGTACGTAAAATTCGA           3886
```

Fig. 2. Continued.

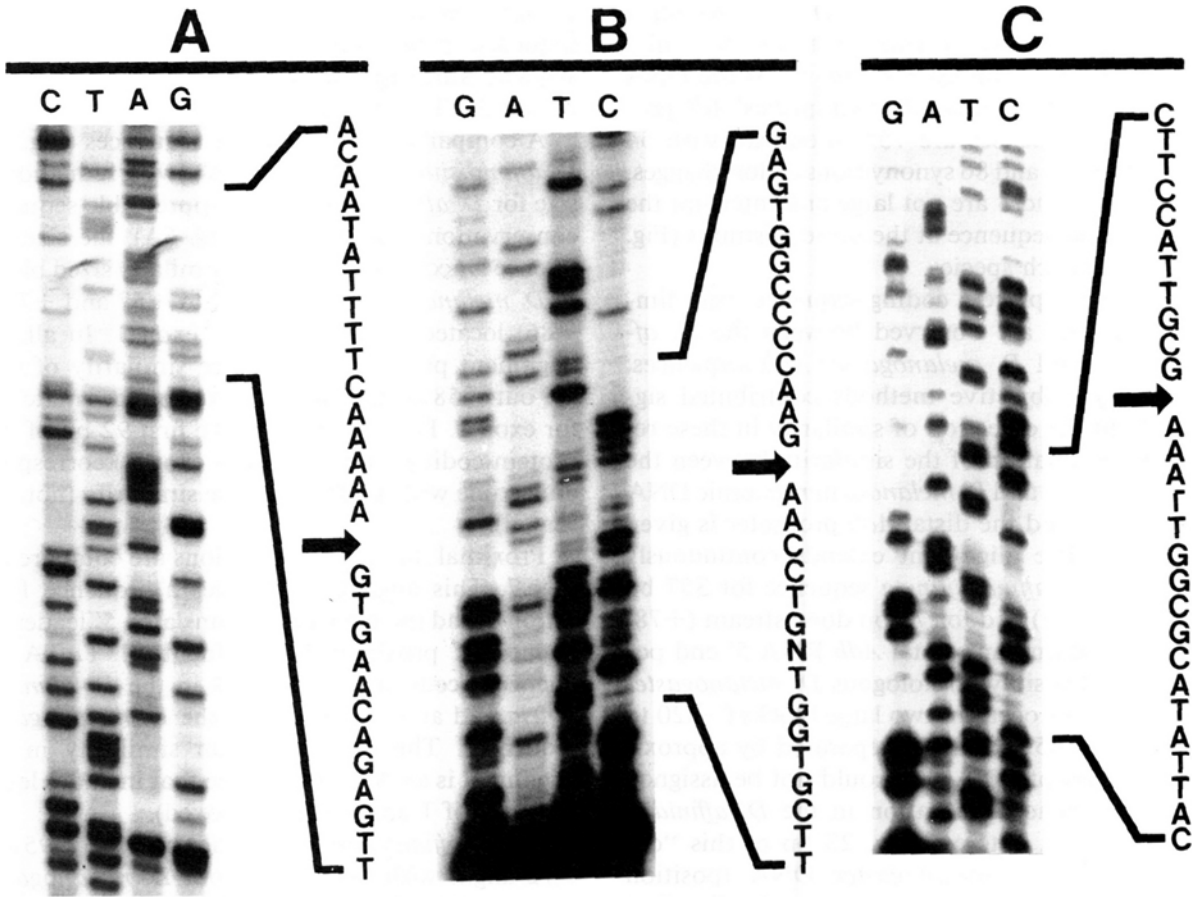

Fig. 3. *Drosophila affinidisjuncta Adh* mRNA sequences. Sequences covering the regions interrupted by introns 1, 2, and 3 are given in **A**, **B**, and **C**, respectively. See text for methods. Arrows mark the location of intervening sequences.

sequence in the 3′ region (numbered 3127–3652) with the 5′ sequence (numbered 1501–2031). The divergence between these sequences is similar to the divergence of the 3′ *Adh* region within the picture-winged species group (R.G. Rowan and J.A. Hunt, unpublished), suggesting that the duplication is about as ancient as this taxonomic group. Several of the observed *Adh* mRNA polyadenylation sites, including the most abundant site at position 3238 (Fig. 2), are in the duplicated region. Consequently, it is tempting to speculate that the unusually large number of 3′ ends in *D. affinidisjuncta Adh* mRNA is a result of the disruption, by the duplication event, of an ancestral gene organization that was more conventional.

There is no suggestion that different *D. affinidisjuncta Adh* transcripts encode different proteins. Proximal and distal RNAs utilize the same putative initiation (position 2033, Fig. 2) and termination (position 2950) codons, and no data imply heterogeneity in the central portion of the *Adh* transcription map (Rowan et al. 1986).

Comparison of the D. affinidisjuncta and D. melanogaster Adh regions

Conceptual translation of the DNA sequence encoding the ADH protein from *D. affinidisjuncta* yields an amino acid sequence that is very similar to that from *D. melanogaster* (Fig. 5). At the DNA level (data not presented), the two species' *Adh* protein coding sequences are 75% identical, with 54 nonsynonymous and 86 synonymous codon changes. *Adh* introns 2 and 3 are not large and interrupt the protein coding sequence at the same positions (Fig. 5, arrows) in each species.

Flanking the protein coding sequence, very limited similarities are observed between the *D. affinidisjuncta* and *D. melanogaster Adh* sequences. Accordingly, subjective methods contributed significantly to the detection of similarity in these regions. An evaluation of the similarity between the *D. affinidisjuncta* and *D. melanogaster* genomic DNA sequences around the distal *Adh* promoter is given in Fig. 6A. The alignment extends continuously through the *D. affinidisjuncta* sequence for 357 bp upstream (−357) and for 78 bp downstream (+78) from the predominant distal *Adh* RNA 5′ end position (+1). Possibly homologous *D. melanogaster* DNA sequences occur in two large blocks (−620 to −466 and −185 to +108), separated by approximately 280 bp of DNA that could not be assigned to the same genomic location in the *D. affinidisjuncta* sequence. Interestingly, 23 bp of this "deleted/inserted" *D. melanogaster* DNA (position −302 to −280) is very similar to a 28-bp *D. affinidisjuncta* sequence that occurs just outside of the long alignment (position −357 to −329; see Fig. 6A). The TATA sequence occupies position −31 to −25 in the DNA sequence of both species.

The *D. melanogaster* DNA region −86 to −46 (region D1) contributes to the activity of the distal *Adh* promoter in vitro, and a putative *Adh* transcription factor can be specifically bound to this region in vitro (Heberlein et al. 1985). This region is not well conserved in *D. affinidisjuncta,* and over one-half of region D1 has been excluded from the alignment in Fig. 6A. However, *D. affinidisjuncta* sequences with similarity to parts of this *D. melanogaster* region do occur at their homologous location (−60 to −53, Fig. 6D and E) and also in relatively upstream (Fig. 6B and C) locations. The upstream similarity contains one copy of the sequence CTGC, which is repeated four times, and also twice in reverse, in the *D. melanogaster* D1 region. Overlapping the *D. affinidisjuncta* CTGC is GCAGC (Fig. 6B); the related sequence GCTGC overlaps the *D. melanogaster* CTGC repeats. The single *D. affinidisjuncta* CTGC aligns with one other of the four *D. melanogaster* CTGC repeats (Fig. 6C). Relatively downstream similarities to region D1 emphasize a different sequence repeat, TCGAC (Fig. 6D and E). This sequence and the related TCAAC in the *D. affinidisjuncta* sequence form a conserved tandem repeat that aligns with the *D. melanogaster* sequence in two ways. The two species' promoter sequences are slightly displaced, relative to each other and the TATA sequences, in these instances.

A comparison of the exon 1 sequences (Fig. 6A, *D. affinidisjuncta* +1 to +58) shows a reduction in size for *D. affinidisjuncta,* but appreciable sequence conservation. As presented (Fig. 6A), the size difference is accounted for in two similarly sized blocks of *D. melanogaster* DNA (+42 to +57 and +72 to +86) located in the 3′ half of exon 1. In all, the alignment produces a sequence similarity of 67% (39 out of 58 *D. affinidisjuncta* nucleotides matched) for exon 1. For comparison, the first 92 bp of *Adh* protein-coding sequence (from exon 2) correspond one to one with a 68% sequence similarity (not presented).

Proximal *Adh* promoter regions are compared in Fig. 7. This alignment includes a sequence from intron 1 and includes the untranslated 5′ leader sequence of proximal RNA. Identical TATA sequences occur at −31 to −24 in the *D. affinidisjuncta* and at −32 to −25 in the *D. melanogaster* sequences. The most upstream similarity in the alignment is an A-rich sequence that includes lesser amounts of T and C (Fig. 7, boxed).

The *D. affinidisjuncta* sequence between −95 and −72 aligns with the homologous *D. melanogaster* sequence (Fig. 7). The latter (designated P0) binds a possible transcription factor in vitro (Heberlein et

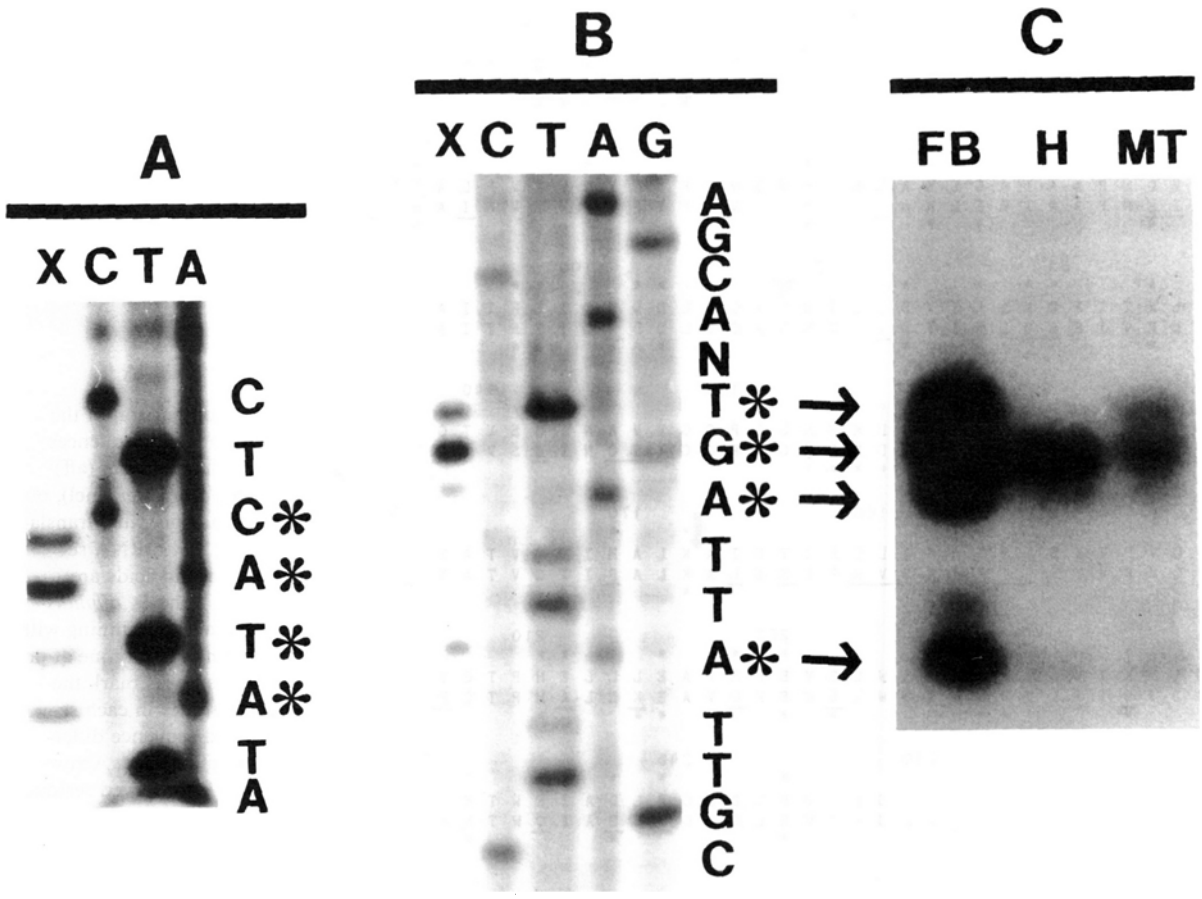

Fig. 4. Primer extension mapping of the 5′ ends of proximal (**A**) and distal (**B** and **C**) *Adh* transcripts from *D. affinidisjuncta*. See text for methods. Asterisks mark the alignment of cDNA termini (lane X in **A** and **B**) with the genomic DNA sequence (**A** and **B**, lanes G, A, T, C). The 5′ ends of distal *Adh* transcripts from separate adult tissues (FB, abdominal fat body; H, head; MT, Malpighian tubules) are identified in **C**.

al. 1985). A second upstream region of the *D. melanogaster* sequence (−150 to −104, designated P1) possesses a binding activity that is related to that of the distal promoter region D1 (Heberlein et al. 1985; see above). Recall that the *D. melanogaster* D1 sequence is not well represented in *D. affinidisjuncta* (Fig. 6A), and note from Fig. 7 (see asterisks) that the related region P1 is also not similar to the *D. affinidisjuncta* sequence. A third protein-binding region of the *D. melanogaster* proximal *Adh* promoter (P2; Heberlein et al. 1985) is not apparent in the *D. affinidisjuncta* sequence, and is also not present in the *D. melanogaster* sequence that is used in the present analysis.

The two species' proximal *Adh* RNA 5′ leaders are of similar size, and the conserved DNA sequence is clustered around the intron 1 splice acceptor sequence (+28 to +42 in *D. affinidisjuncta*; Fig. 7). In all, the proximal 5′ leader sequences are 45% similar (28 of 62 *D. affinidisjuncta* nucleotides matched). For comparison, the distal leader similarity is 55%. A fraction of both of these values can be attributed to the conservation of intron 1 splicing sequences.

Downstream (3′) to the *Adh* protein-coding region, *D. affinidisjuncta* and *D. melanogaster* DNA sequences are not similar (not presented). Only two polyadenylation signal sequences occur in the *D. melanogaster* sequence, and only one of these (122 nucleotides from the translation termination codon) is used in vivo (Benyajati et al. 1983). This organization contrasts sharply with the multiple occurrences of apparently functional polyadenylation signals in the *D. affinidisjuncta Adh* gene sequence (Fig. 2).

Discussion

Previous reports have stressed the apparent similarity between *D. affinidisjuncta* and *D. melanogaster Adh*. Distal and proximal *Adh* mRNAs occur in both species, and the two transcipt types show the same developmental restriction in both species (Rowan et al. 1986). Also, the two species' ADH activities are biochemically similar (Dickinson and Carson 1979; Dickinson 1980a). The present com-

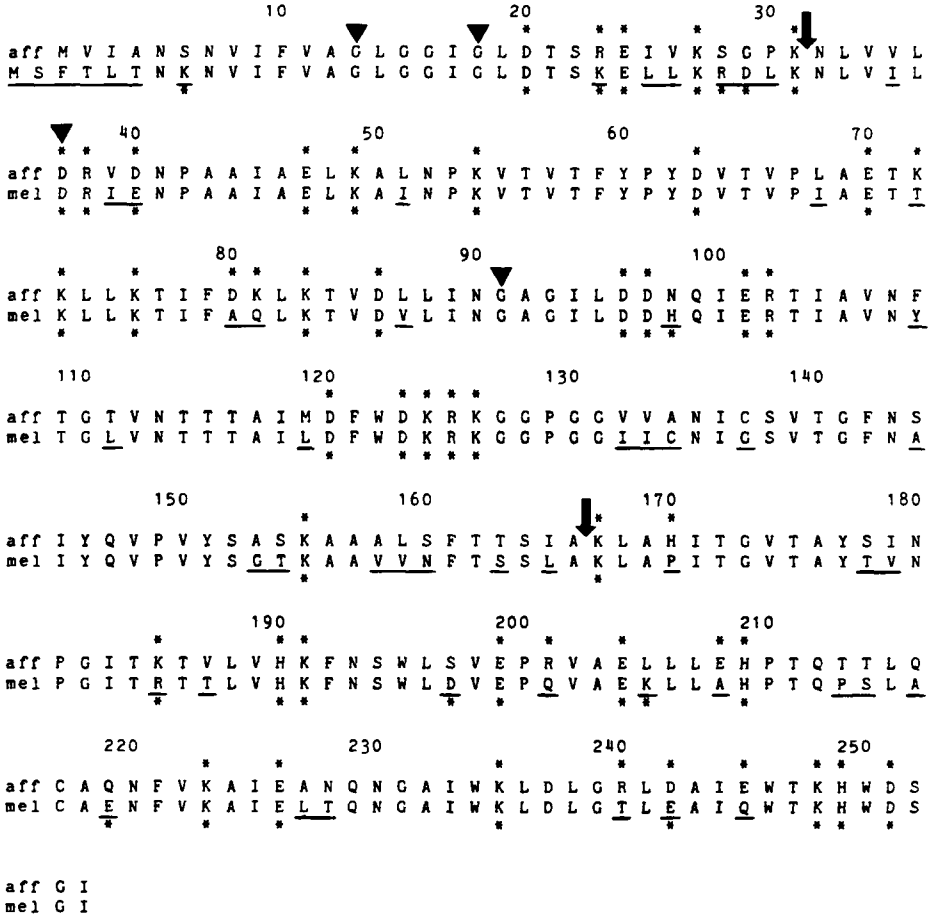

Fig. 5. Comparison of the ADH amino acid sequences from *D. affinidisjuncta* (aff) and *D. melanogaster* (mel), obtained by a conceptual translation of the *Adh* mRNA sequences. Amino acids are numbered in the *D. affinidisjuncta* sequence beginning with the initiating methionine at position 1. Asterisks mark the charged residues in each sequence, and sequence differences are underlined. Arrowpoints mark conserved residues implicated in cofactor binding (Benyajati et al. 1981). Arrows show the locations of *Adh* introns 2 (position 31/32) and 3 (position 166/167).

parison shows the correlates of this similarity at the DNA sequence level.

Drosophila melanogaster and *D. affinidisjuncta* are not closely related. One measure of the evolutionary distance between these species is the disparity of their ADH amino acid sequences (Fig. 5): of the 254 positions in the *D. affinidisjuncta* sequence, 54 are different in *D. melanogaster*. Historical and biogeographical data (Throckmorton 1975) and immunological distance measurements (Beverley and Wilson 1982, 1985) indicate that the subgenera *Drosophila* (which includes *D. affinidisjuncta*) and *Sophophora* (which includes *D. melanogaster*) diverged approximately 40–60 million years (Myr) ago. Accordingly, the amino acid replacement rate for *Drosophila* ADH appears to be approximately 0.5/Myr. Amino acid replacement rates in *Adh* genes from closely related members of the *melanogaster* species group, e.g., 2 per approximately 2 Myr in *D. simulans/D. melanogaster* and 3 per approximately 1 Myr in *D. simulans/D. mauritiana* (Bodmer and Ashburner 1984), are similar to this value.

It is commonly observed that a given protein evolves at a uniform rate in separate lineages (Dickerson 1971; Kimura 1983). For *Drosophila* ADH, a relative rate test (Sarich and Wilson 1967) may be performed using the published *D. mulleri* ADH-1 sequence (Fischer and Maniatis 1985). *Drosophila mulleri* and *D. affinidisjuncta* are members of different radiations within the subgenus *Drosophila* (Throckmorton 1975), and are therefore equally removed from the sophophoran species *D. melanogaster*. The pairwise comparisons *D. affinidisjuncta/D. melanogaster* (54 amino acid replacements; see above), *D. mulleri/D. melanogaster* (47 replacements; not presented), and *D. affinidisjuncta/D. mulleri* (27 replacements; not presented) indicate a uniform rate of ADH protein evolution, and provide an estimate of the time of divergence between members of the *virilis–repleta* (*D. mulleri*) and the *immigrans–Hirtodrosophila* (*D. affinidisjuncta*) radiations (see Throckmorton 1975).

Benyajati et al. (1981) previously discussed physical properties of the *D. melanogaster* ADH enzyme as inferred from its amino acid sequence; their conclusions are germane to the present analysis. Briefly, the first 140 amino acid residues of *D. melanogaster* ADH probably include the cofactor (NAD+)-binding domain, and catalytic activity therefore resides in the remaining sequence (Benyajati et al. 1981). Four *Drosophila* ADH amino acids, which are in-

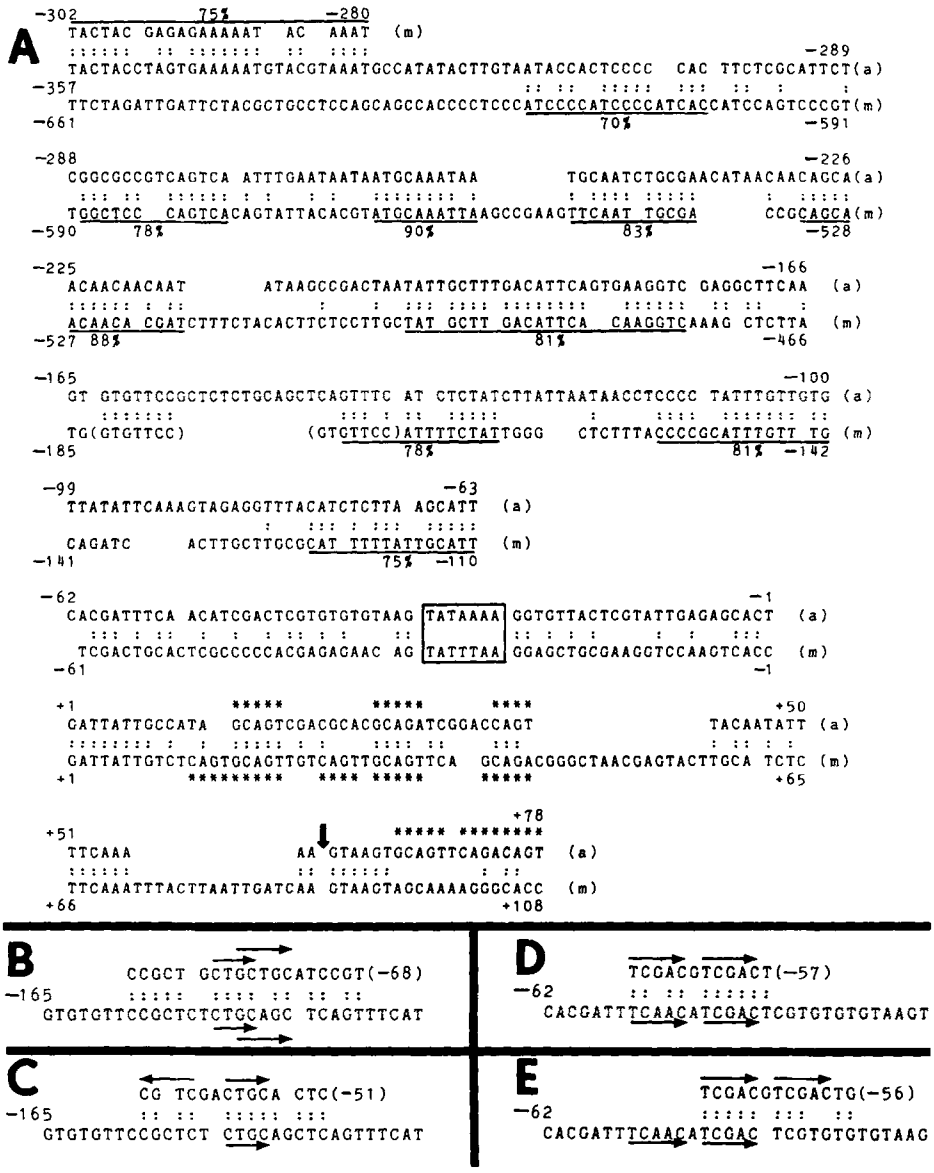

Fig. 6. Comparison of the distal *Adh* promoters from *D. affinidisjuncta* (a) and *D. melanogaster* (m). Sequences are negatively numbered proceeding upstream, and positively numbered proceeding downstream, from the 5′ end of *Adh* mRNA at +1. Only the most abundant of four *D. affinidisjuncta* RNA 5′ ends (see Fig. 4B) is considered. The TATA sequences at −25 are boxed. Underlined sequences exhibit the indicated identity (percentage of nucleotide positions matched), and these occur in the same relative order in each species' genome. An overlined sequence occurs in a different position in the two species. The *D. affinidisjuncta* sequence GTGTTCC at −137 is given in parentheses in two alternate alignments with the *D. melanogaster* sequence. Asterisks indicate conserved repeats of (G)CAG(A/T). The vertical arrow at *D. affinidisjuncta* position +58/+59 marks the boundary between exon 1 and intron 1 (see Fig. 1). **B**, **C**, **D**, **E** Alignment of distal *Adh* promoter sequences from *D. affinidisjuncta* (lower sequence) with those from *D. melanogaster* region D1 (see text). Numbering of sequences as in **A**.

variantly positioned in the coenzyme binding domains of all dehydrogenases (Benyajati et al. 1981), are indicated by arrowpoints in Fig. 5. Also, amino acid replacements that appear in *D. affinidisjuncta* ADH are "conservative" (Zuckerkandl and Pauling 1965; Miyata et al. 1979), and are equally partitioned between both putative domains (28 changes in the first 138 residues and 26 changes in the remaining 116 residues).

Functions other than protein coding are not easily recognized from DNA or RNA sequence data alone, but molecular genetic analyses do suggest consensus properties for sequences that regulate gene expression (Davidson et al. 1983). Because some regulatory sequences appear as similarities in the DNA sequences of homologous or similarly regulated genes (e.g., Pelham 1982; McKnight et al. 1984; Parslow et al. 1984), we have compared the *Adh* promoter sequences from *D. affinidisjuncta* to those from *D. melanogaster*.

As aligned in Fig. 6A, distal *Adh* promoter sequences contain 10 areas of limited interspecific similarity (underlined; average size 17 bp, average identity 79%). Proximal promoter sequences contain three such areas (Fig. 7, underlined; average size 13 bp, average identity 80%). The biological significance of these similarities remains untested, but more thorough studies of other genes (e.g., Davidson et al. 1983; Donahue et al. 1983; Dynan and Tjian 1985; Karin et al. 1984; McKnight et al. 1984; Stuart et al. 1984; Johnson and Herskowitz 1985; Osborne et al. 1985; Simon et al. 1985) show that the sizes, percentage similarities, and organization of these *Adh* sequences are appropriate for sequences that encode regulatory function(s). P element-mediated gene transfer methods (Rubin and

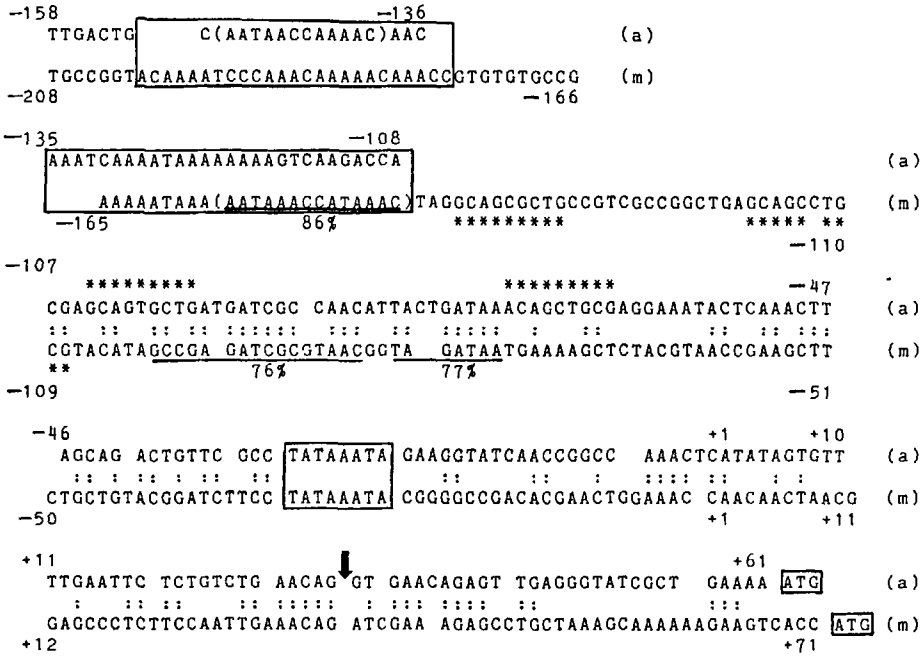

Fig. 7. Comparison of the proximal *Adh* promoters from *D. affinidisjuncta* (a) and *D. melanogaster* (m). The sequence is numbered relative to the 5′ end of the proximal *Adh* transcript at +1. TATA sequences, conserved A-rich upstream regions, and the translation initiation codon (ATG) are boxed. Underlined sequences exhibit the indicated identity (percentage of nucleotide positions matched). Sequences in parentheses, *D. affinidisjuncta* −139 and *D. melanogaster* −144, show an 86% identity, but are differently located within the A-rich region. Asterisks indicate sequences within the *D. melanogaster* region P1 (Heberlein et al. 1985) and similar sequences in *D. affinidisjuncta*. The vertical arrow marks the intron 1–exon 2A boundary (see Fig. 1).

Spradling 1982) are being developed to test this hypothesis directly (Brennan et al. 1984b; Brennan and Dickinson 1988).

Nontranslated leader sequences (in exon 1) of the two species' distal *Adh* transcripts are similar to each other (Fig. 6A, sequence +1 to +56, and asterisks). In contrast, similarities between the nontranslated leaders of the proximal *Adh* transcripts are limited, and these probably reflect the conservation of intron 1 splice recognition sequences (Fig. 7, sequence +1 to +61). Distal, more than proximal, mRNA leader sequences might be important for *Adh* translational control (e.g., Klemenz et al. 1985; see Kozak 1983), or perhaps exon 1 sequences regulate the abundance of *Adh* transcripts. In principle, these too are testable hypotheses.

Functional tests of the *D. affinidisjuncta Adh* gene, placed into the *D. melanogaster* genome by P element-mediated transformation, show that *D. affinidisjuncta Adh* does indeed contain regulatory information that is accurately expressed in the *D. melanogaster* milieu (Brennan and Dickinson 1988). With the reservation that the introduced *D. affinidisjuncta Adh* clone was larger (by about 900 bp at the 5′ and about 600 bp at the 3′ end) than the *Adh* sequence reported here, these data do strengthen the inference that observed DNA sequence similarities are functionally significant.

The present study found a partial duplication of an *Adh* sequence immediately following the complete *Adh* gene. This duplication is in turn followed closely by a sequence that is repeated several times within the *D. affinidisjuncta* genome. This repetitive element has not been characterized, but blot hybridizations show that at least part of its sequence is included in Fig. 2 (Brennan et al. 1984a and unpublished). These and other data distinguish the *D. affinidisjuncta Adh* duplication from duplications of the *Adh* region in other *Drosophila* species.

A probable tandem duplication of *Adh* has been preserved in *D. melanogaster* and in *Drosophila pseudoobscura* as a gene encoding a protein now only vaguely similar to ADH (Schaeffer and Aquadro 1987). This putative gene is directly 3′ to the single *Adh* gene in these two species; related sequences have been cloned from *D. affinidisjuncta* and mapped (by in situ hybridization) to a chromosome location unlinked to *Adh* (Brennan et al. 1984a and unpublished).

The *Drosophila mulleri* genome contains three tandem copies of *Adh*, two of which encode ADH (Fischer and Maniatis 1985). Unlike the duplication of *Adh* intron 1 in *D. affinidisjuncta*, sequences that flank the duplicated ADH coding regions in *D. mulleri* have diverged greatly from each other and are not recognizably similar (Fischer and Maniatis 1985).

Thus, three unique tandem duplications of *Adh* DNA have now been described. At the least, these kinds of data should be valuable for phylogenetic analyses.

Acknowledgments. Our colleague Mark Brennan provided valuable insights, encouragement, and technical advice during this study. M. Ashburner provided an unpublished *Adh* sequence from *D. melanogaster*. We are grateful to L. Bossi and D. Smith for technical advice on m13 cloning and sequencing, and to P. Keim for help with Maxam and Gilbert sequencing. E. Kofoid made computer programs available, and patiently taught computer methods for DNA sequence analysis to R.G.R. Glenn Herrick critically reviewed an earlier draft of the manuscript. This work was supported by grant HD 10723 from the National In-

stitutes of Health (NIH). R.G.R. was supported by an NIH predoctoral service award trainee grant (5-T32-GM7464).

References

Batterham P, Lovett JA, Starmer WT, Sullivan DT (1983) Differential regulation of duplicate alcohol dehydrogenase genes in *Drosophila mojavensis*. Dev Biol 96:346–354

Benyajati C, Place AR, Powers DA, Sofer W (1981) Alcohol dehydrogenase gene of *Drosophila melanogaster*: relationships of intervening sequences to functional domains in the protein. Proc Natl Acad Sci USA 78:2717–2721

Benyajati C, Spoerel N, Haymerle H, Ashburner M (1983) The messenger RNA for alcohol dehydrogenase in *Drosophila melanogaster* differs in its 5′ end in different developmental stages. Cell 33:125–133

Beverley SM, Wilson AC (1982) Molecular evolution in *Drosophila* and higher Diptera. I. Micro-complement fixation studies of a larval hemolymph protein. J Mol Evol 18:251–264

Beverley SM, Wilson AC (1985) Ancient origin for Hawaiian *Drosophilinae* inferred from protein comparisons. Proc Natl Acad Sci USA 82:4753–4757

Birnstiel ML, Busslinger M, Strub K (1985) Transcription termination and 3′ processing: the end is in sight! Cell 41:349–359

Bodmer M, Ashburner M (1984) Conservation and change in the DNA sequences coding for alcohol dehydrogenase in sibling species of *Drosophila*. Nature 309:425–430

Brennan MD, Dickinson WJ (1988) Complex developmental regulation of the *Drosophila affinidisjuncta* alcohol dehydrogenase gene in *Drosophila melanogaster*. Dev Biol 125:64–74

Brennan MD, Rowan RG, Rabinow L, Dickinson WJ (1984a) Isolation and initial characterization of the alcohol dehydrogenase gene from *Drosophila affinidisjuncta*. J Mol Appl Genet 2:436–446

Brennan MD, Rowan RG, Dickinson WJ (1984b) Introduction of a functional P element into the germ-line of *D. hawaiiensis*. Cell 38:147–151

David JR, Van Herrewege J (1983) Adaptation to alcoholic fermentation in *Drosophila* species: relationship between alcohol tolerance and larval habitat. Comp Biochem Physiol 74A:283–288

Davidson EH, Jacobs HT, Britten RJ (1983) Very short repeats and coordinate induction of genes. Nature 301:468–470

Dickerson RE (1971) Structure of cytochrome *c* and the rates of molecular evolution. J Mol Evol 1:26–45

Dickinson WJ (1980a) Evolution of patterns of gene expression in Hawaiian picture-winged *Drosophila*. J Mol Evol 16: 73–94

Dickinson WJ (1980b) Complex *cis*-acting regulatory genes demonstrated in *Drosophila* hybrids. Dev Genet 1:229–240

Dickinson WJ, Carson HL (1979) Regulation of the tissue specificity of an enzyme by a *cis*-acting genetic element: evidence from interspecific *Drosophila* hybrids. Proc Natl Acad Sci USA 76:4559–4562

Donahue TF, Daves RS, Lucchini G, Fink GR (1983) A short nucleotide sequence required for regulation of *HIS4* by the general control system of yeast. Cell 32:89–98

Dynan WS, Tjian R (1985) Control of eukaryotic messenger RNA synthesis by sequence-specific DNA-binding proteins. Nature 316:774–778

Fischer JA, Maniatis T (1985) Structure and transcription of the *Drosophila mulleri* alcohol dehydrogenase genes. Nucleic Acids Res 13:6899–6917

Fischer JA, Maniatis T (1986) Regulatory elements involved in *Drosophila Adh* gene expression are conserved in divergent species and separate elements mediate expression in different tissues. EMBO J 5:1275–1289

Garoff H, Ansorge W (1981) Improvements of DNA sequencing gels. Anal Biochem 115:450–457

Goldberg DA (1980) Isolation and partial characterization of the *Drosophila* alcohol dehydrogenase gene. Proc Natl Acad Sci USA 77:5794–5798

Goldberg DA, Posakony JW, Maniatis T (1983) Correct developmental expression of a cloned alcohol dehydrogenase gene transduced into the *Drosophila* germ line. Cell 34:59–73

Hamlyn PH, Brownlee GG, Cheng C, Gait MJ, Milstein C (1978) Complete sequence of constant and 3′ noncoding regions of an immunoglobin mRNA using the dideoxynucleotide method of RNA sequencing. Cell 15:1067–1075

Heberlein U, England B, Tjian R (1985) Characterization of *Drosophila* transcription factors that activate the tandem promoters of the alcohol dehydrogenase gene. Cell 41:965–977

Johnson AD, Herskowitz I (1985) A repressor (MAT α2 product) and its operator control expression of a set of cell type specific genes in yeast. Cell 42:237–247

Karin M, Haslinger A, Holtgreve H, Richards RI, Krauter P, Westphal HM, Beato M (1984) Characterization of DNA sequences through which cadmium and glucocorticoid hormones induce the human metallothionein-II_A gene. Nature 308:513–519

Kimura M (1983) The neutral theory of molecular evolution. Cambridge University Press, Cambridge

Klemenz R, Hultmark D, Gehring WJ (1985) Selective translation of heat shock mRNA in *Drosophila melanogaster* depends on sequence information in the leader. EMBO J 4: 2053–2060

Kozak M (1983) Comparison of initiation of protein synthesis in prokaryotes, eukaryotes and organelles. Microbiol Rev 47: 1–45

Kreitman M (1983) Nucleotide polymorphism at the alcohol dehydrogenase locus of *Drosophila melanogaster*. Nature 304: 412–417

Maniatis T, Fritsch EF, Sambrook J (1982) Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor NY

Maroni G, Laurie-Ahlberg CC, Adams DA, Wilton AN (1982) Genetic variation in the expression of *Adh* in *Drosophila melanogaster*. Genetics 101:431–446

Maxam AM, Gilbert W (1980) Sequencing end-labeled DNA with base-specific chemical cleavages. In: Grossman L, Moldave K (eds) Methods in enzymology. Academic Press, New York, pp 499–560

McKenzie JA (1980) An ecological study of the alcohol dehydrogenase polymorphism of *Drosophila melanogaster*. Aust J Zool 28:709–716

McKenzie JA, Parsons PA (1972) Alcohol tolerance: an ecological parameter in the relative success of *Drosophila melanogaster* and *Drosophila simulans*. Oecologia 10:373–388

McKnight SL, Kingsbury RC, Spence A, Smith M (1984) The distal transcription signals of the herpes virus *tk* gene share a common hexanucleotide control sequence. Cell 37:253–262

Messing J, Vieira J (1982) A new pair of M13 vectors for selecting either DNA strand of double-digest restriction fragments. Gene 19:269–276

Miyata T, Miyazawa S, Yasunaga T (1979) Two types of amino acid substitutions in protein evolution. J Mol Evol 12:219–236

Mount S (1982) A catalogue of splice junction sequences. Nucleic Acids Res 10:459–472

Osborne TF, Goldstein JL, Brown MS (1985) 5′ end of HMG CoA reductase gene contains sequences responsible for cho-

lesterol-mediated inhibition of transcription. Cell 42:203–212

Parslow TG, Blair DL, Murphy WJ, Granner DK (1984) Structure of the 5′ ends of immunoglobulin genes: a novel conserved sequence. Proc Natl Acad Sci USA 81:2650–2654

Pelham HRB (1982) A regulatory upstream promoter element in the *Drosophila* hsp 70 heat-shock gene. Cell 30:517–528

Rowan RG, Brennan MD, Dickinson WJ (1986) Developmentally regulated RNA transcripts coding for alcohol dehydrogenase in *Drosophila affinidisjuncta*. Genetics 114:405–433

Rowan RG, Dickinson WJ (1986) Two alternate transcripts coding for alcohol dehydrogenase accumulate with different developmental specificities in different species of Hawaiian picture-winged *Drosophila*. Genetics 114:435–452

Rubin GM, Spradling AC (1982) Genetic transformation of *Drosophila* with transposable element vectors. Science 218: 348–353

Sanger F, Nicklen S, Coulson AR (1977) DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci USA 74: 5463–5467

Sarich VM, Wilson AC (1967) Immunological time scale for hominoid evolution. Science 158:1200–1203

Savakis C, Ashburner M, Willis JH (1985) The expression of the gene coding for alcohol dehydrogenase during the development of *Drosophila melanogaster*. Dev Biol 114:194–207

Schaeffer SW, Aquadro CF (1987) Nucleotide sequence of the *Adh* gene region of *Drosophila pseudoobscura*: evolutionary change and evidence for an ancient gene duplication. Genetics 117:61–73

Simon JA, Sutton CA, Lobell RB, Glaser RL, Lis JT (1985) Determinants of heat shock-induced chromosome puffing. Cell 40:805–817

Stuart G, Searle PF, Chen HY, Brinster RL, Palmiter RD (1984) A 12-base-pair-DNA motif that is repeated several times in metallothionein gene promoters confers metal regulation to a heterologous gene. Proc Natl Acad Sci USA 81:7318–7322

Throckmorton LH (1975) The phylogeny, ecology and geography of *Drosophila*. In: King RC (ed) Handbook of genetics. Plenum, New York, pp 421–469

Ursprung H, Sofer WH, Burroughs N (1970) Ontogeny and tissue distribution of alcohol dehydrogenase in *Drosophila melanogaster*. Wilhelm Roux's Arch Dev Biol 164:201–208

Vieira J, Messing J (1982) The pUC plasmids, an M13mp7-derived system for insertion mutagenesis and sequencing with synthetic universal primers. Gene 19:259–268

Zuckerkandl E, Pauling L (1965) Evolutionary divergence and convergence in proteins. In: Bryson V, Vogel HJ (eds) Evolving genes and proteins. Academic Press, New York, pp 97–166

Received October 15, 1986/Revised and accepted February 20, 1988