Structure and Evolution of the *Adh* Genes of *Drosophila mojavensis*

Peter W. Atkinson, Leslie E. Mills, William T. Starmer and David T. Sullivan

Department of Biology, Syracuse University, Syracuse, New York 13244

Manuscript received April 4, 1988
Revised copy accepted July 7, 1988

ABSTRACT

The nucleotide sequence of the *Adh* region of *Drosophila mojavensis* has been completed and the region found to contain a pseudogene, *Adh-2* and *Adh-1* arranged in that order. Comparison of the sequence divergence of these genes to one another and to the *Adh* region of *Drosophila mulleri* and other species has allowed the development of a model for the evolution of the duplication of the *Adh* genes. There have been two major events. An initial duplication of an *Adh* gene whose dual promoter structure was similar to *Drosophila melanogaster*, resulted in a species with two *Adh* genes, one of which may have had only a proximal promoter. A second duplication of this gene generated an *Adh* region containing three genes. It is proposed that one of these is the ancestral gene having dual promoters, while the other two possess only proximal promoters. Subsequent events have resulted in both a change in the regulation of *Adh-2* such that it is expressed as if it had a "distal" type promoter and the mutational inactivation of the most upstream gene resulting in the creation of a pseudogene. The sequence of the *D. mojavensis Adh* region has also revealed the presence of an element which is composed of juxtaposed inverted imperfectly repeated elements. There is a surprising and not fully explainable strong similarity of the nucleotide sequence of the 5′ flanking region of the pseudogene in *D. mojavensis* and *D. mulleri*.

TWO species subgroups of the repleta group of the genus *Drosophila* have been found to have a duplication for the gene which encodes the enzyme alcohol dehydrogenase (ADH) (OAKESHOTT *et al.* 1982; BATTERHAM *et al.* 1983b). This duplication is likely to be a relatively recent event since it is only found in the closely related mulleri and hydei subgroups. Events leading to the duplicated *Adh* genes probably occurred on the order of 20 million years ago (see DISCUSSION). Studies in our laboratory using *Drosophila mojavensis* and related species (BATTERHAM *et al.* 1983b, 1984) and studies on *Drosophila mulleri* (FISCHER and MANIATIS 1986) have demonstrated that the two genes, referred to as *Adh-1* and *Adh-2*, are not coordinately controlled. We have pursued the analysis of these genes for this reason. It seems likely that differences in expression of *Adh-1* and *Adh-2* resulted from changes in nucleotide sequence which occurred at or following the duplication event. This provides the opportunity for analyzing changes in DNA sequences involved in the specific regulation of each *Adh* gene by comparing the genes and their flanking sequence both within and between species. In addition this gene duplication may be a good case with which to elucidate the evolution of *cis*-acting regulatory sequences and the role such changes may have during speciation. Thus, it may be possible to determine whether the changes in regulation of a particular *Adh* gene occurred as a direct result of the duplication or, alternatively, arose through a secondary event or events which occurred later during the divergence of the two *Adh* genes.

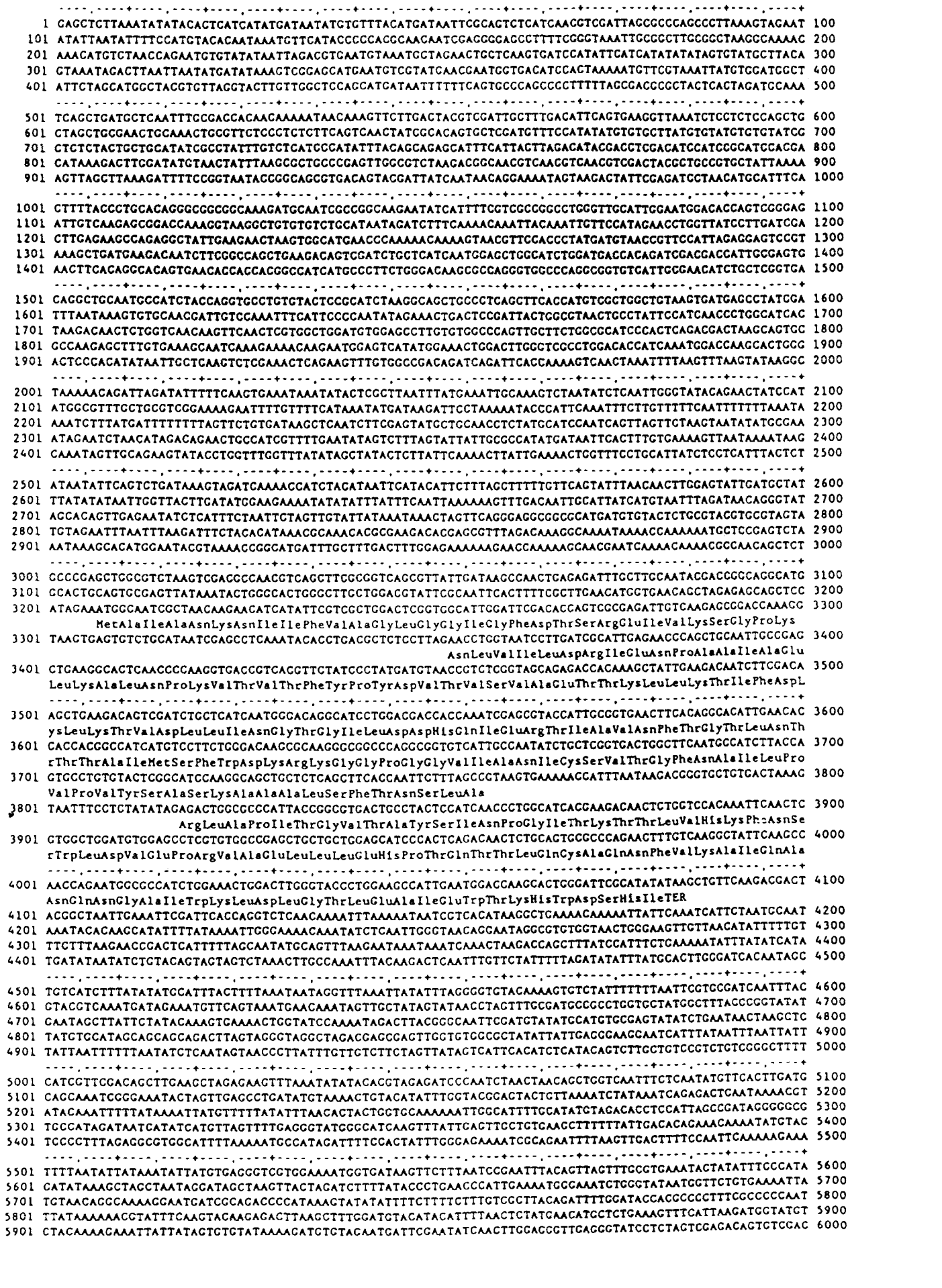
The basic structure of the *Adh* region in *D. mulleri* has been found by FISCHER and MANIATIS (1985) to consist of two functional genes, *Adh-1* and *Adh-2*, and one pseudogene arranged in tandem along approximately 10 kb of DNA. This is consistent with the structure of the *D. mojavensis* gene proposed by MILLS *et al.* (1986). MILLS *et al.* (1986) also reported that the functional genes, *i.e.*, those which encode active enzymes, are closely linked and located at a single chromosomal site.

We report here the nucleotide sequence of an 8.8-kb section of the *Adh* region of *D. mojavensis*. This includes *Adh-1* and *Adh-2* and an *Adh* pseudogene. The comparison of these sequences with both the nucleotide sequence of the *Adh* genes from other species and among one another has allowed us to estimate when the *Adh* duplication(s) originated and to propose a series of steps which might have occurred during the evolution of the *Adh* region as it is currently represented in *D. mojavensis*.

MATERIALS AND METHODS

The clone of *D. mojavensis Adh* DNA was obtained from an EMBL-4 genomic library as described previously (MILLS *et al.* 1986).

Nucleotide sequences were determined by the chain termination method, HONG (1982), using ^{35}S-labeled dATP (New England Nuclear). The buffer gradient gels of BIGGIN, GIBSON and HONG (1983) were used for separation of ter-



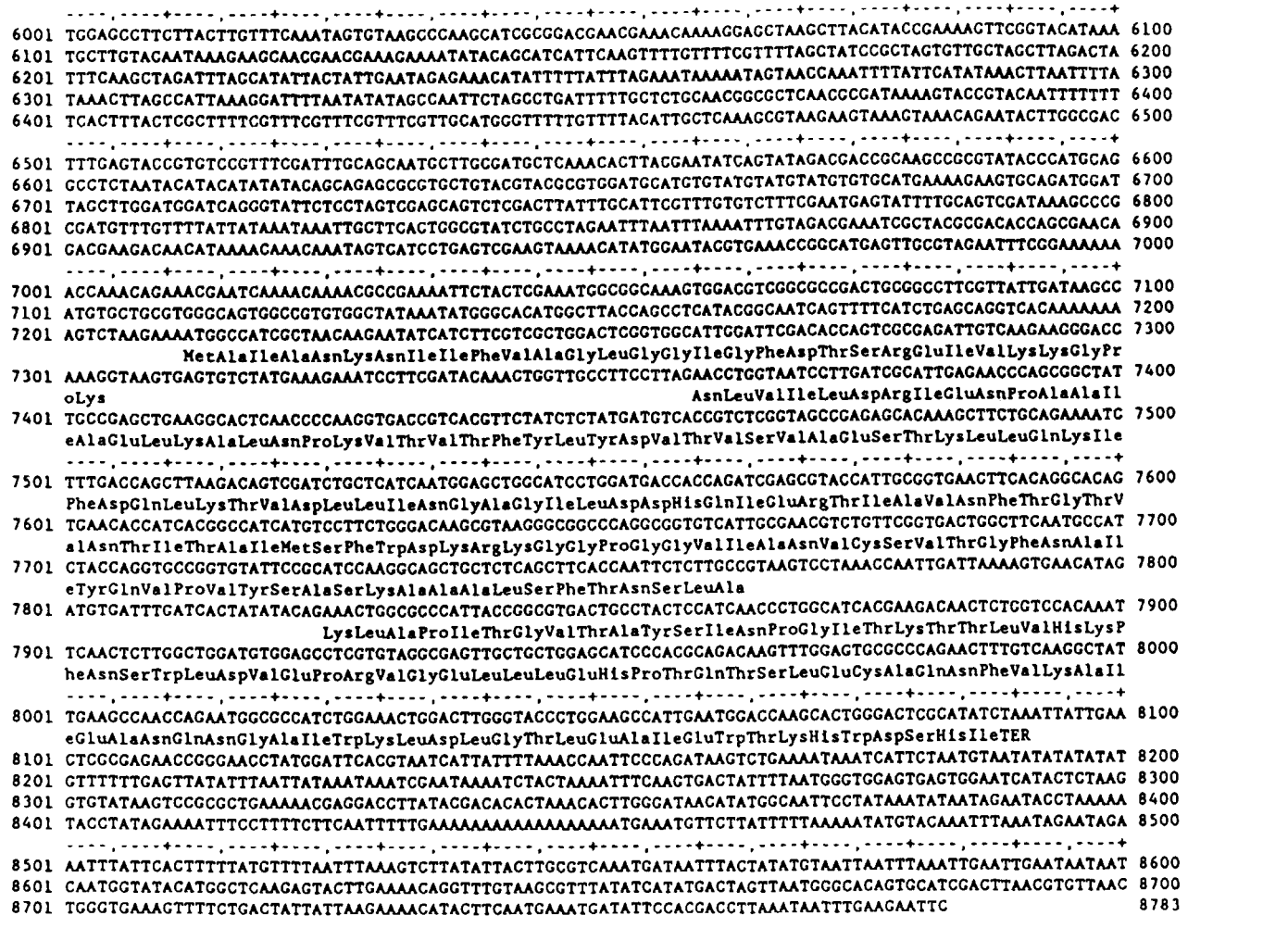

6001 TGGAGCGTTCTTACTTGTTTCAAATAGTGTAAGCCCAAGCATCGCGGACGAACGAAACAAAAGGAGGTAAGCTTACATACCGAAAAGTTCGGTACATAAA 6100
6101 TGCTTGTACAATAAAGAAGCAACGAACGAAACGAAAGAAAATATACAGCATCATTCAAGTTTTGTTTTCGTTTTAGCTATCCGCTAGTGTTGCTAGCTTAGACTA 6200
6201 TTTCAAGCTAGATTTAGCATATTACTATTGAATAGAGAAACATATTTTTATTTAGAAATAAAAATAGTAACGAAATTTTATTCATATAAACTTAATTTTA 6300
6301 TAAACTTAGCCATTAAAGGATTTTAATATATAGCCAATTCTAGCCTGATTTTTGCTCTGCAACGGCGCTCAACGCGATAAAAGTACCGTACAATTTTTTT 6400
6401 TCACTTTACTCGCTTTTCGTTTCGTTTCGTTTCGTTGCATGGGTTTTTGTTTTACATTGCTCAAAGCGTAAGAAGTAAAGTAAACAGAATACTTGGCGAC 6500
6501 TTTGAGTACCGTGTCCGTTTCGATTTGCAGCAATGCTTGCGATGCTCAAACACTTACGAATATCAGTATAGACGACCGCAAGCCGCGTATACCCATGCAG 6600
6601 GCCTCTAATACATACATATATACAGCAGAGCGCGTGCTGTACGTACGCGTGGATGCATGTGTATGTATGTATGTGTGCATGAAAAGAAGTGCAGATGGAT 6700
6701 TAGCTTGGATGGATCAGGGTATTCTCCTAGTCGAGCAGTCTCGACTTATTTGCATTCGTTTGTGTCTTTCGAATGAGTATTTTGCAGTCGATAAAGCCCG 6800
6801 CGATGTTTGTTTTATTATAAATAAATTGCTTCACTGGCGTATCTGCCTAGAATTTAATTTAAAATTTGTAGACGAAATCGCTACGCGACACCAGCGAACA 6900
6901 GACGAAGACAACATAAAACAAACAAATAGTCATCCTGAGTCGAAGTAAAACATATGGAATAGGTCAAACCGGCATGAGTTGCGTAGAATTTCGGAAAAAA 7000
7001 ACCAAACAGAAACGAATCAAAACAAAACGCCGAAAATTCTACTCGAAATGGCGGCAAAGTGGACGTCGGCGCCGACTGCGGCCTTCGTTATTGATAAGCC 7100
7101 ATGTGCTGCGTGGGCAGTGGCCGTGTGGCTATAAATATGGGCACATGGCTTACCAGCCTCATACGGCAATCAGTTTTCATCTGAGCAGGTCACAAAAAAA 7200
7201 AGTCTAAGAAAATGGCCATCGCTAACAAGAATATCATCTTCGTCGGTGGACTCGGTGGCATTGGATTCGACACCAGTCGGCGAGATTGTCAAGAAGGGACC 7300
MetAlaIleAlaAsnLysAsnIleIlePheValAlaGlyLeuGlyGlyIleGlyPheAspThrSerArgGluIleValLysLysGlyPr
7301 AAAGGTAAGTGAGTGTCTATGAAAGAAATCCTTCGATACAAACTGGTTGCCTTCCTTAGAACCTGGTAATCCTTGATCGCATTGAGAACCCAGCGGCTAT 7400
oLys AsnLeuValIleLeuAspArgIleGluAsnProAlaAlaIl
7401 TGCCGAGGTGAAGGCACTCAACCCCAAGGTGACCGTCACGTTCTATCTCTATGATGTCACCGTCTCGGTAGCCGAGAGCACAAAGCTTCTGCAGAAAATC 7500
eAlaGluLeuLysAlaLeuAsnProLysValThrValThrPheTyrLeuTyrAspValThrValSerValAlaGluSerThrLysLeuLeuGlnLysIle
7501 TTTGACCAGCTTAAGACAGTCGATCTGCTCATCAATGGAGCTGGCATCCTGGATGACCACCAGATCGAGCGTACCATTGCGGTGAACTTCACAGGCACAG 7600
PheAspGlnLeuLysThrValAspLeuLeuIleAsnGlyAlaGlyIleLeuAspAspHisGlnIleGluArgThrIleAlaValAsnPheThrGlyThrV
7601 TGAACACCATCACGGCCATCATGTCCTTCTGGGACAAGCGTAAGGGCGGCCCAGGCGGTGTCATTGCGAACGTCTGTTCGGTGACTGGCTTCAATGCCAT 7700
alAsnThrIleThrAlaIleMetSerPheTrpAspLysArgLysGlyGlyProGlyGlyValIleAlaAsnValCysSerValThrGlyPheAsnAlaIl
7701 CTACCAGGTGCCGGTGTATTCCGCATCCAAGGCAGCTGCTCTCAGCTTCACCAATTCTCTTGCCGTAAGTCCTAAAGCAATTGATTAAAAGTGAACATAG 7800
eTyrGlnValProValTyrSerAlaSerLysAlaAlaAlaLeuSerPheThrAsnSerLeuAla
7801 ATGTGATTTGATCACTATATACAGAAACTGGCGCCCATTACCGGCGTGACTGCCTACTCCATCAACCCTGGCATCACGAAGACAACTCTGGTCCACAAAT 7900
LysLeuAlaProIleThrGlyValThrAlaTyrSerIleAsnProGlyIleThrLysThrThrLeuValHisLysP
7901 TCAACTCTTGGCTGGATGTGGAGCCTCGTGTAGGCGAGTTGCTGCTGGAGCATCCCACGCAGACAAGTTTGGAGTCGCGCCCAGAACTTTGTCAAGGCTAT 8000
heAsnSerTrpLeuAspValGluProArgValGlyGluLeuLeuLeuGluHisProThrGlnThrSerLeuGluCysAlaGlnAsnPheValLysAlaIl
8001 TGAAGCCAACCAGAATGGCGCCATCTGGAAACTGGACTTGGGTACGCTGGAAGCCATTGAATGGACCAAGCACTGGGACTCCGCATATCTAAATTATTGAA 8100
eGluAlaAsnGlnAsnGlyAlaIleTrpLysLeuAspLeuGlyThrLeuGluAlaIleGluTrpThrLysHisTrpAspSerHisIleTER
8101 CTCGCGAGAACCGGCAACCTATGGATTCACGTAATCATTATTTTAAACCAATTCCCAGATAAGTCTGAAAATAAATCATTCTAATGTAATATATATATAT 8200
8201 GTTTTTTGAGTTATATTTAATTATAAATAAATCGAATAAAATCTACTAAAATTTCAAGTGACTATTTTAATGGGTGGAGTGAGTGGAATCATACTCTAAG 8300
8301 GTGTATAAGTCCGCGCTGAAAAACGAGGACCTTATACGACACACTAAACACTTGGGATAACATATGGCAATTCGTATAAATATAATAGAATACCTAAAAA 8400
8401 TACCTATAGAAAATTTCCTTTTCTTCAATTTTTGAAAAAAAAAAAAAAAAAAATGAAATGTTCTTATTTTTAAAAATATGTACAAATTTAAATAGAATAGA 8500
8501 AATTTATTCACTTTTTATGTTTTAATTTAAAGTCTTATATTACTTGCGTCAAATGATAATTTACTATATGTAATTAATTTAAATTGAATTGAATAATAAT 8600
8601 CAATGGTATACATGGCTCAAGAGTACTTGAAAACAGGTTTGTAAGCGTTTATATCATATGACTAGTTAATGGGCACAGTGCATCGACTTAACGTGTTAAC 8700
8701 TGGGTGAAAGTTTTCTGACTATTATTAAGAAAACATACTTCAATGAAATGATATTCCACGACCTTAAATAATTTGAAGAATTC 8783

FIGURE 1.—Nucleotide sequence of the *D. mojavensis Adh* gene region. The sequence runs from a *Sst*I site to an *Eco*RI site of the map shown in MILLS *et al.* (1986).

minated nucleotide fragments. Gels were read and sequences compared using a digitizer and computer programs from DNASTAR, Madison, Wisconsin. Comparisons of *Adh* genes both between and within species were performed by the algorithm devised by WILBUR and LIPMAN (1983).

Clones for sequencing were obtained using two strategies. In the first, an 11-kb fragment containing the entire *Adh* region was sonicated into 300–500-bp fragments. These were isolated following agarose gel electrophoresis and made "blunt" using phage T4 DNA polymerase and then ligated to *Sma*I digested M13mp8 DNA. The second strategy was to subclone three non-overlapping fragments which covered this region of DNA into M13mp18. A nested set of deletions was subsequently generated using the procedures described by HENIKOFF (1984) and these were sequenced by the above methods.

Sequences were compared between pairs of species to determine relative divergence of the species (orthologous comparisons), while sequence comparisons between the different genes of *D. mojavensis* or *D. mulleri* were made to determine relative divergence of the duplicate genes (paralogous comparisons). Intron, 5′ untranslated and 3′ untranslated regions were compared by using the Tajima and Nei method as outlined by LI, LUO and WU (1985). This method which estimates K (the mean number of substitutions per nucleotide site) is a modification of the one-parameter method of JUKES and CANTOR (LI, LUO and WU 1985). Exons were compared by the new method of LI, LUO and WU (1985); however, we did not weight possible paths between two codons according to the relative frequencies of codon changes in mammalian genes. Instead, where appropriate (paths through stop codons were not allowed) we weighted paths as equally probable. We choose to calculate K_A (substitutions per nonsynonomous site) and K_S (substitutions per synonomous site) for the pseudogenes even though they no longer have coding function since a comparison with K_A and K_S for functional genes will give us an estimate as to when the pseudogene became inactivated and some idea about the validity of using K_S as a means of estimating neutral substitution rates in *Drosophila Adh* genes. Alignment of introns was accomplished by inspecting the sequence and then by adding or deleting a minimum number of nucleotides at appropriate positions in order to maximize similarity.

In order to align exon-1 of *D. melanogaster*, six-nucleotides just following the start codon were deleted since the *D. mojavensis Adh* is two amino acids shorter at the N-terminal end. Time since divergence of two sequences from the ancestral sequence was calculated by using the substitution rate of $\alpha = 5.5 \times 10^{-9}$ substitutions per synonymous site per year as estimated by MIYATA, YASUNAGA and MISHIDA (1980) and HAYASHIDSA and MIYATA (1983). The rate was used with the estimate of K to calculate T in the formula $K = (\alpha)\ (2T)$. Percent similarity between sequences was

calculated as 100 × (number of nucleotides in common)/(total number of nucleotides compared).

RESULTS

The nucleotide sequence of an 8.8-kb region of the *D. mojavensis* genome which includes the *Adh* genes is presented in Figure 1. This region contains three *Adh* regions, the most 5′ of which is an *Adh* pseudogene. The *ATG* codon analogous to an ADH translation start point is at nucleotide position 1030. This pseudogene contains several frame shift mutations and stop codons which preclude the production of an active ADH molecule. The pseudogene contains sequences which are homologous to intron splice sites at nucleotide positions *1122, 1178* and *1581, 1646.* These are the expected position for introns in a *Drosophila Adh* gene.

Downstream of the pseudogene are two *Adh* genes whose conceptual translation is indicated in Figure 1. The more 5′ of the two encodes the more basic protein and is consequently judged to be *Adh-2* based on the previously described properties of the *D. mojavensis* ADH molecules (BATTERHAM *et al.* 1983a). The 3′ gene therefore encodes ADH-1. These two genes have previously been shown to encode electrophoretically separable proteins which are genetically closely linked. Each of the *Adh* genes has two introns located in the identical positions of other *Drosophila Adh* genes. The *Adh* region of *D. mojavensis* described here is fundamentally similar to the *Adh* region of *D. mulleri* described by FISCHER and MANIATIS (1985). A major difference seems to be an increase in the spacing between the *Adh-2* and *Adh-1* genes which is due to a 1.1 kb insertion (see below).

In order to study the origin of the *Adh* genes of *D. mojavensis* we have compared the extent of nucleotide substitution between the three *D. mojavensis* genes and between the *D. mojavensis* genes and the *Adh* genes of other species of the genus for which sequence information is available. These comparisons are presented in Table 1.

Comparisons of the sequence divergence between each of the genes within a species, in this case *D. mojavensis*, allow for the study of the sequence of events which occurred in the evolution of an *Adh* locus containing only one gene to the state now found in several *Drosophila* species of the repleta group. A graphical comparison of the extent of nucleotide substitution between the *Adh* genes of *D. mojavensis* in pairwise comparison is shown in Figure 2. In a qualitative sense three points are of note. The extent of substitution at synonymous sites, K_S, measured in comparing *Adh-1* and *Adh-2* is lower than comparing measurements of either gene to the pseudogene. Second, the extent of substitutions in the introns, K_I, is similar in each comparison and greater than K_S between the coding genes. Finally, there is a suggestion of an increase in the extent of nucleotide substitution at non-synonymous nucleotides in comparisons involving either coding gene and the pseudogene. While these data do not provide a statistically significant demonstration of this point, the small increase in the mean value of K_A is greater in each comparison involving the pseudogene within *D. mojavensis, D. mulleri* or between these species. In any case, it is clear that there is a large difference in the magnitude of increase of K observed at non-synonymous sites as compared to the increase in K at synonymous sites in coding-pseudogene as compared to coding-coding gene comparisons.

In the following argument the difference in the amount of substitution for synonymous vs. nonsynonymous sites for the pseudogene and *Adh-1* or *Adh-2* (*i.e.*, $K_{S(\psi-1 \text{ or } 2)} - K_{A(\psi-1 \text{ or } 2)}$) is compared with the difference in amount of substitution for the same categories of sites in *Adh-1* and *Adh-2* ($K_{S(1-2)} - K_{A(1-2)}$). This comparison can be used to evaluate the likely history of the evolution of the three genes. Using the values from Table 1 and estimating the standard error of the difference as the square root of the sum of the two variances, $K_{S(\psi-1 \text{ or } 2)} - K_{A(\psi-1 \text{ or } 2)} = 0.33 \pm 0.097$ or 0.34 ± 0.096 and $K_{S(1-2)} - K_{A(1-2)} = 0.17 \pm 0.05$. The ratio of these two differences is 1.94 or 2.00 (average of 1.97).

Figure 3 shows three possible evolutionary histories of the three *Adh* genes. Let T be the time since the first duplication, x be the time the gene destined to become the pseudogene remained active after its origin and y be the time between the two duplication events (model 2 does not involve a second event). Let α_S be the substitution rate for synonymous substitutions as estimated by the comparison of *Adh-1* and *Adh-2*. For model 1,

$\alpha_S = K_{S(1-2)}/(2(T - y))$, while
$\alpha_S = K_{S(1-2)}/(2T)$ for model 2 and 3.

Let $\alpha_S\psi$ be the substitution rate for synonymous sites for comparison of the nonfunctional pseudogene to the other genes. In a similar manner, let α_A be the substitution rate for nonsynonymous substitutions as estimated by the comparison of *Adh-1* and *Adh-2*. For model 1,

$\alpha_A = K_{A(1-2)}/[2(T - y)]$ while
$\alpha_A = K_{A(1-2)}/(2T)$ for model 2 and 3.

Let $\alpha_A\psi$ = the substitution rate for nonsynonymous sites for comparison involving *Adh-1* or *Adh-2* and the pseudogene after it became nonfunctional. We assume $\alpha_S\psi = \alpha_A\psi = \alpha\psi$, in all three models, since all codon sites should be equivalent in terms of substitution rate after the pseudogene became nonfunctional.

For model 1 the value for K_s of the pseudogene (ψ)

TABLE 1

Nucleotide substitution comparisons of *D. mojavensis Adh* genes

Species	*Adh-1*					*Adh-2*					*Adh-ψ*				
	$\%S_E$	K_S	K_A	$\%S_I$	K_I	$\%S_E$	K_S	K_A	$\%S_I$	K_I	$\%S_E$	K_S	K_A	$\%S_I$	K_I
D. mojavensis Adh-1						93.33	0.197	0.035	63.39	0.517	84.26	0.443	0.112	64.04	0.506
							(0.037)	(0.033)		(0.095)		(0.064)	(0.070)		(0.093)
D. mojavensis Adh-2	93.33	0.197	0.035	63.39	0.517						84.52	0.441	0.106	64.29	0.509
		(0.037)	(0.033)		(0.095)							(0.063)	(0.072)		(0.096)
D. mojavensis Adh-ψ	84.26	0.443	0.112	64.04	0.506	84.52	0.441	0.106	64.29	0.509					
		(0.064)	(0.070)		(0.093)		(0.063)	(0.072)		(0.096)					
D. mulleri Adh-1	94.51	0.184	0.022	67.83	0.441	93.59	0.210	0.028	82.30	0.206	84.92	0.427	0.104	63.48	0.528
		(0.036)	(0.029)		(0.085)		(0.038)	(0.029)		(0.049)		(0.063)	(0.070)		(0.099)
D. mulleri Adh-2	93.73	0.200	0.029	61.95	0.565	94.38	0.208	0.017	71.17	0.372	85.10	0.437	0.099	60.53	0.594
		(0.037)	(0.029)		(0.107)		(0.038)	(0.021)		(0.089)		(0.062)	(0.065)		(0.110)
D. mulleri Adh-ψ	81.22	0.629	0.123	60.71	0.585	81.76	0.639	0.111	62.73	0.532	91.72	0.297	0.035	84.17	0.180
		(0.086)	(0.079)		(0.109)		(0.086)	(0.075)		(0.098)		(0.050)	(0.029)		(0.043)
D. affinidisjuncta	80.78	1.01	0.076	ND	ND	81.70	0.945	0.067	ND	ND	76.06	1.25	0.123	ND	ND
		(0.140)	(0.056)				(0.124)	(0.060)				(0.178)	(0.075)		
D. melanogaster	77.91	1.001	0.126	ND	ND	77.52	1.035	0.123	ND	ND	74.34	1.080	0.182	ND	ND
		(0.143)	(0.077)				(0.144)	(0.071)				(0.182)	(0.100)		

$\%S_E$ = % similarity of exons; K_I = substitution per nucleotide in introns; K_S = substitution per nucleotide for synonomous sites; K_A = substitution per nucleotide for nonsynonomous sites; $\%S_I$ = % similarity of introns; ND = not determined, numbers in parentheses are SE. Variances were estimated according to equations 20 (K_S) and 22 (K_A) of LI, LUO and WU (1985).

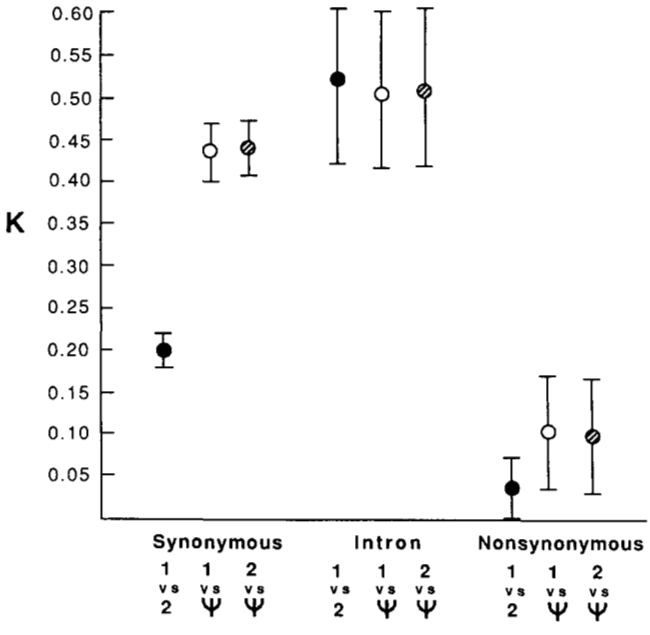


FIGURE 2.—Graphical comparison of the extent of sequence divergence in pairwise comparison between the *Adh* genes of *D. mojavensis* for synonymous codon substitutions, introns and nonsynonymous codon substitutions. Closed circles, *Adh-1* vs. *Adh-2* comparisons; open circles, *Adh-1* vs. *Adh-ψ* comparisons; hatched circles, *Adh-2* vs. *Adh-ψ*.

vs. either *Adh-1* or *Adh-2* can be written as

$$K_{S(\psi-1 \text{ or } 2)} = [K_{S(1-2)}/\{2(T-y)\}]\cdot(T+x) + \alpha\psi(T-x).$$

Likewise the value for K_A of the pseudogene ( ) *vs.* either *Adh-1* or *Adh-2* in model 1 can be written as

$$K_{A(\psi-1 \text{ or } 2)} = [K_{A(1-2)}/\{2(T-y)\}] \cdot(T+x) + \alpha\psi(T-x).$$

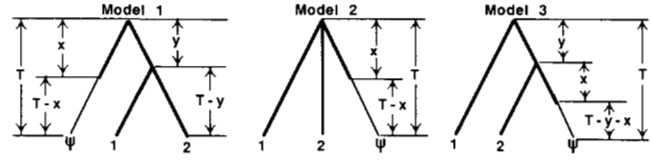


FIGURE 3.—Three models describing the evolutionary history of the *D. mojavensis Adh* locus. T is time since the first duplication event, y is the time between the first and the second duplication event, x is time that the ancestor to the pseudogene had coding function.

The difference between $K_S(\psi -1 \text{ or } 2)$ and $K_A(\psi -1 \text{ or } 2)$ is

$$[K_{S(\psi-1 \text{ or } 2)} - K_{A(\psi-1 \text{ or } 2)}] = [K_{S(1-2)} - K_{A(1-2)}]\cdot[(T+x)/\{2(T-y)\}].$$

Substituting the estimates for the values of K (from Table 1), $1.92 = (T + x)/\{2(T - y)\}$ and setting y as a function of x and T, $y = 0.75T - 0.25x$. This formulation can be used to interpret the possible time that the pseudogene became inactive relative to the origin of *Adh-1* and *Adh-2*. When $x < 0.60T$ then the pseudogene became inactive before the origin of *Adh-1* and *Adh-2*. When $x > 0.60T$ then the pseudogene became inactive after the other two genes were duplicated.

Model 2 assumes all three *Adh* genes originated from a common event (*i.e.*, model 1 with $y = 0$, $x = 3.0T$). This result shows x to be outside of the realm of possible values for x and thus the model is considered inappropriate.

Model 3 can be constructed with either *Adh-1* or *Adh-2* as the coproduct of the second duplication event. Since the differences between K_S and K_A are

similar, both constructions yield the same comparative information. A similar algebraic formulation yields two equations, one involving *Adh-1* and one involving *Adh-2*; $y = 3.0\ T - x$ and $y = x - 3.0T$. Since $x < T$ then $y > T$ or $y < 0$ and thus model 3 is judged to be inappropriate.

Model 1 is the only model which has a reasonable interpretation and thus provides a point of comparison for when the pseudogene became inactive relative to the origin of the two functional genes. The length of time since the divergence of *Adh-1* and *Adh-2* (y estimated from the $K_{S(1-2)}$ value for *D. mojavensis*) is approximately the same as the length of time since *D. mojavensis* and *D. mulleri* lineages separated (T estimated from the interspecific $K_{S(1-1)}$ or $K_{S(2-2)}$ values) and it is therefore difficult to determine the value of y or when the second duplication event occurred. Sequence information on a more distant member of the mulleri group might provide a better relative measure of the time since the first duplication event, since the widespread occurrence of the duplication in the group and complex suggests it is monophyletic.

Given that model 1 represents a likely history of events occurring during the evolution of the *Adh* duplication as represented in *D. mojavensis* we consider one issue related to comparisons between these three genes. The value of K_S for comparisons between pseudogenes is larger than between coding genes. A higher rate of nucleotide substitutions has also been noted in comparisons of globin genes by LI, GOJOBORI and NEI (1981). A likely interpretation of the increase in K_S is that the codon bias seen among synonymous codons is no longer relevant to the sequence of the pseudogene. This has been suggested previously by ASHBURNER, BODMER and LEMEUNIER (1984) and MIYATA and HAYASHIDA (1981). This interpretation is at least partially correct because the *Adh* pseudogene has experienced several nucleotide deletions that should render the gene unconstrained with regard to codon usage. Since we have argued above that the pseudogene was functional for a significant fraction of its history during which the constraints operating on coding genes would apply it is likely that the pseudogene has not yet attained codon randomization.

A focus of our studies has been to understand the evolutionary events which resulted in changes in gene expression since the origin of the *Adh* duplication. Towards this end we have compared the 5′ ends of each of the genes of *D. mojavensis*. As shown in Figure 4, an alignment performed according to the algorithm devised by WILBUR and LIPMAN (1983) of 400 nucleotides upstream and inclusive of the translation start site demonstrates that there is extensive similarity of the 5′ flanking regions of *Adh-1* and *Adh-2*. The overall sequence similarity of these two regions is 75%. However, it is evident there are several blocks of identical or almost identical sequence. First, the TATA box and adjacent nucleotides immediately downstream are almost identical in each gene. Another long stretch of about 250 nucleotides that is highly similar starts about 30 to 40 nucleotides 5′ to the TATA box. This region does contain pentanucleotide sequences, indicated in Figure 4, similar to the repeats found in regions involved in binding of the *Adf-1* transcription factor identified by HEBERLEIN, ENGLAND and TJIAN (1985). Further 5′ there are additional smaller blocks of identical sequences. One that is particularly striking is a sequence of 13 nucleotides that are identical in *Adh-1* and *Adh-2* and includes a second TATA like element. However, there is no reason to suspect that these are functional with respect to transcription initiation since all the transcripts from these *Adh* genes originate in the expected positions downstream of the TATA boxes underlined in Figure 4 (W. CARROLL and D. SULLIVAN, unpublished data). Whether these conserved regions represent regions that are conserved for functional reasons or represent areas that are similar simply due to common origins cannot be decided in the absence of experimental tests. The similarity of these regions does seem to point out the likely common origin of these genes and their associated 5′ flanking regions.

Additional information about the relationship of these *Adh* genes can be gained by comparing the sequence divergence between the *Adh* genes of *D. mojavensis* and its close relative, *D. mulleri* whose sequence has previously been determined by FISCHER and MANIATIS (1985). A comparison of the coding genes between the two species reveals that the K_S values are consistently lower than the K_S values between genes within a species. These values while not statistically significantly different, indicate that *Adh-1* and *Adh-2* began to diverge from each other near the time of or possibly slightly earlier than the species divergence time. A graphical presentation of the sequence similarities between these two species is shown in Figure 5. The regions immediately 5′ to each coding gene are highly similar in these two species. It has been demonstrated that sequences located within 350 bp upstream of the transcription start points of both *Adh-1* genes are sufficient for the regulation of each of these genes when transformed into *D. melanogaster* (FISCHER and MANIATIS 1986; C. BAYER and D. SULLIVAN, unpublished data). The *Adh-1* genes are 83% similar approximately 360 bp upstream. The *Adh-2* genes are even more similar, 92% to approximately 400 bp upstream from the transcription start positions. It may be relevant that the *Adh-2* genes in these two species have a similar time and tissue specific pattern of expression. However, the *Adh-1* genes dif-

```
Adh-2 TAGAATTTAATTTAAGATTTCTACACATAAA-CGCAAACACGCGAAGACACGAGCGTTTAGACAAAGGCAAAATAAAACCAAA-AAATGC---TCG-GAGTCTAAATAAA
Adh-1 TAGAATTTAATTTAAAATTTGTAGACG-AAATCGCT---ACGCGA---CACCAGCGAACAGACGAAGACAACATAAAAC-AAACAAATAGTCATCCTGAGTCGAAGTAAA

Adh-2 GCACATGGAATACGTAAAACCGGCATGATTTGCTTTGACTTTGGAGAAAAAAGAACCAAAAAGCAACGAATCAAAACAAAACGCCAACAGCTCTGCCCGAGCTGGCGTCT
Adh-1 ACATATGGAATACGTGAAACCGGCATGAGTTGCGTAGAATTTCG-GAAAAAA--ACCAAACAGAAACGAATCAAAACAAAACGCCGAAAATTCTACTCGAAATGGCGGCA

          <----
Adh-2 AAGTCGACGCCAACGTCAGCTTCG----------CGGTCAGCGTTATTGATAAGCCAACTGAGAGATTTGCTTGCAATACGACCGGCAGGCATGGCACTGCAGTGCGAGT
Adh-1 AAGTGGACG----------TCGGCGCCGACTGCGGCCTTCGTTATTGATAAGCCATG---------TGC-TGCGTG------GGCAG---TGGCCGT---GTGGC---
            <-          ---                                  -- -->

Adh-2 TATAAATACTGGGCACTGGGCTTGCTGG------ACGTATTCGCAATTCACTTTTCGCTTGAACATGGTGAACAG--------CTAGAGAGCAGCTCCATAGAAATG
Adh-1 TATAAATA-TGGGCACATGGCTTACCAGCCTCATACG-----GCAA-TCAGTTTTCATCTGAGCA-GGTC-ACAAAAAAAAGTCTA-AGA----------AAATG
```

FIGURE 4.—Alignment of the nucleotide sequences 5′ to the *Adh-1* and *Adh-2* genes of *D. mojavensis*. The *Adh-2* sequence begins at 2803 of Figure 1. The *Adh-1* sequence begins at 6848 of Figure 1. The TATA box and translational start signals are underlined. Pentanucleotide sequences identified as *Adf-1* binding sites by HEBERLEIN, ENGLAND and TJIAN (1985) are marked with arrows.

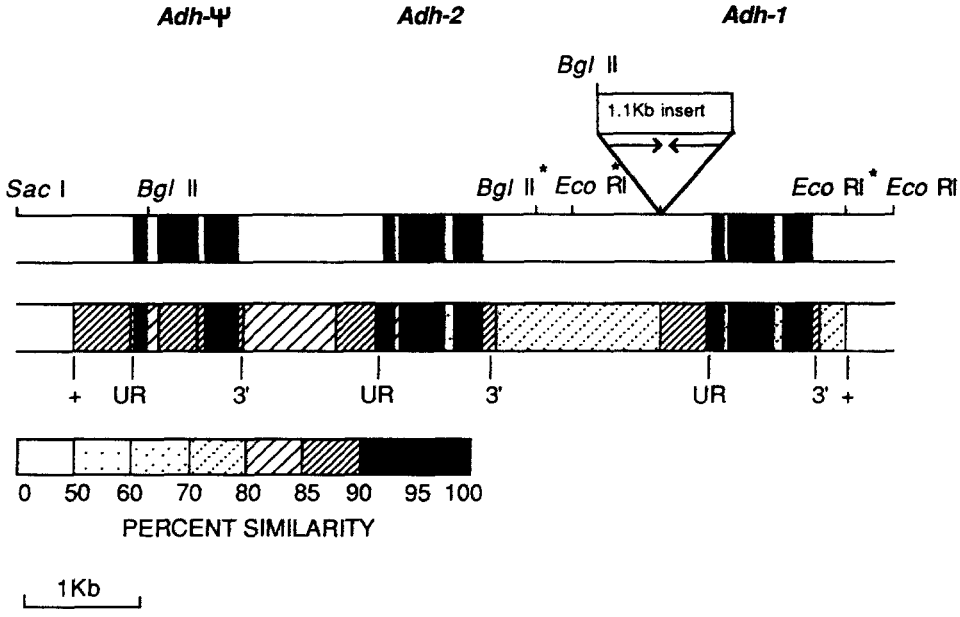


FIGURE 5.—Regional similarity between *Adh* genes of *D. mojavensis* and *D. mulleri*. Alignment of flanking sequences was performed according to the algorithm devised by WILBUR and LIPMAN (1983). + represents the limit of published *D. mulleri* sequence, UR represents the 5′ untranslated region, 3′ represents the untranslated region of the ADH transcript up to the putative poly A addition sites. * are restriction sites specific to *D. mulleri*.

fer between the two species in that *Adh-1* is abundantly expressed in the ovaries of *D. mojavensis* (BATTERHAM *et al.* 1983b, 1984) while there is no *Adh* expression in the ovaries of *D. mulleri* (FISCHER and MANIATIS 1986; C. BAYER and D. SULLIVAN, unpublished data).

Figure 5 also shows that there are regions of high sequence similarity between these species upstream from *Adh-2* as far as comparative sequence data is available. Two regions are of particular note. The introns of the pseudogenes are very similar in sequence $K_I = 0.18$ (Table 1). In addition there is a region immediately 5′ to the pseudogenes that appears highly conserved. The sequence of this region is shown in Figure 6. Of note is a region, shown underlined, that is almost identical to the TATA box region of a distal promoter of a *D. melanogaster Adh* gene (see also FISCHER and MANIATIS 1985). There is no fully adequate explanation for the similarity of the intron and 5′ flanking regions of the pseudogenes of *D. mulleri* and *D. mojavensis*. Several possibilities are considered in the discussion below.

The comparison of the *Adh* genes of *D. mojavensis* and *D. mulleri* reveals one major difference in structure. There is a 1.1-kb insertion located upstream from *Adh-1* of *D. mojavensis* (Figure 5). Close inspection of this insertion reveals it to entirely consist of two juxtaposed imperfect inverted repeats whose center is at nucleotide position 6254 (Figure 1). As such this element is similar to the foldback (FB) transposable element of *D. melanogaster*. However its internal structure contains no sequence similarity to FB and, in addition, this element does not contain small direct repeats within each large inverted repeat as is found in FB elements (POTTER 1982).

Comparison of the coding region of *D. mojavensis Adh-1* or *Adh-2* with the single *Adh* gene of *D. affini-*

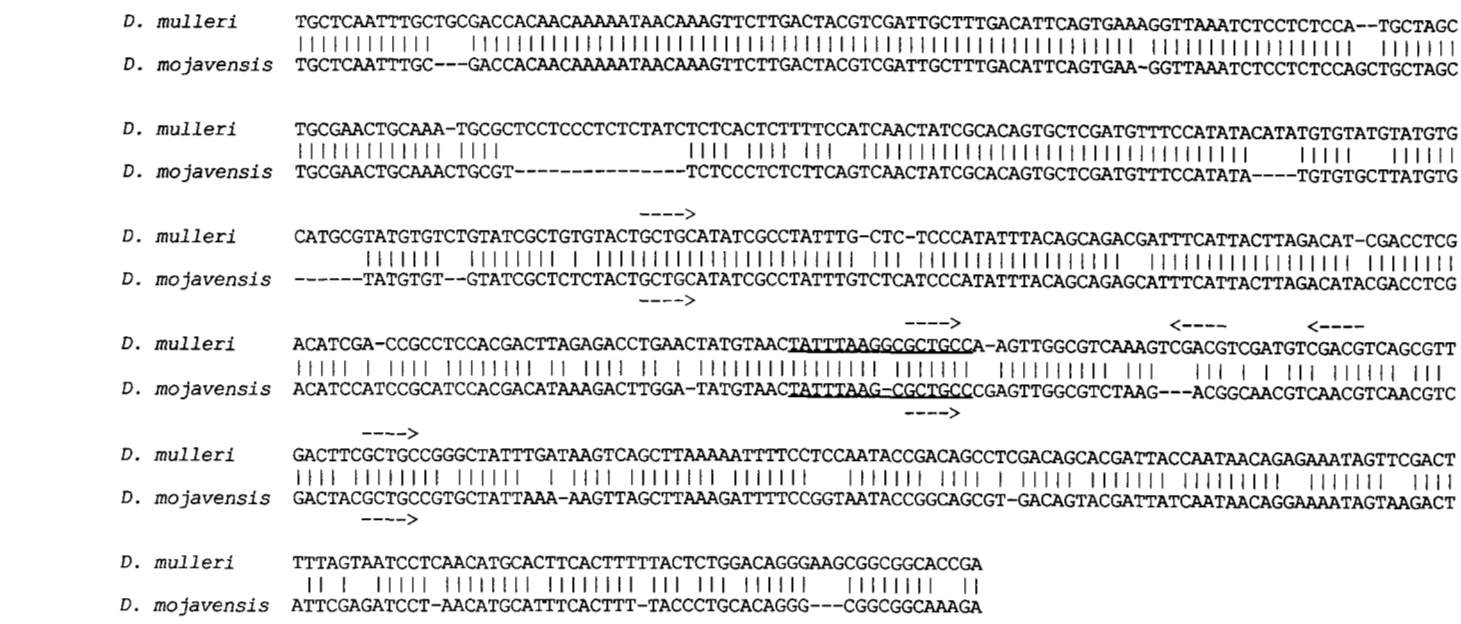


FIGURE 6.—Alignment of the nucleotide sequence 5′ to the pseudogenes of *D. mojavensis* and *D. mulleri*. The *D. mojavensis* sequence starts at position 495 of Figure 1. A sequence of 19 nucleotides is underlined which is highly similar to the TATA box of the distal promoter region of the *Adh* gene of *D. melanogaster* [see text and FISCHER and MANIATIS (1985)]. Pentanucleotide sequences identified as *Adf-1* binding sites by HEBERLEIN, ENGLAND and TJIAN (1985) are marked with arrows.

disjunta and with *D. melanogaster* is shown in Table 1 and indicates an appreciable similarity with the *Adh* gene of each species. The magnitude of divergence between *D. mojavensis* and *D. affinidisjunta* is similar to that between *D. mojavensis* and *D. melanogaster*. *D. affinidisjunta* and *D. mojavensis* are both members of the subgenus *Drosophila* while *D. melanogaster* is a member of the subgenus *Sophophora*. The similarity in the extent of nucleotide substitutions in these comparisons indicates that the lineage leading to *D. mojavensis* and *D. affinidisjunta* split shortly after the divergence of the two subgenera.

We have attempted to compare the regions 5′ to the *Adh* genes of *D. mojavensis* with comparable regions of the proximal and distal promoters of the *Adh* genes of *D. affinidisjunta* and *D. melanogaster* (data not shown). The comparison reveals significant stretches of similar sequence only at the region of the TATA boxes and the pentanucleotides sequences that are putative transcription factor binding sites (HEBERLEIN, ENGLAND and TJIAN 1985). The TATA boxes and immediately adjacent regions of both *D. mojavensis Adh-2* and *Adh-1* are similar to the sequence of the TATA box regions of only the proximal promoters of the *Adh* genes of species that have dual promoters. This has also been noted by FISCHER and MANIATIS (1985) for the TATA box regions of *D. mulleri Adh* genes. Small stretches of similarity can be observed in any pairwise comparison between these three species. However, in only a few cases of short sequence are the same regions identified in separate paired comparisons. This lack of recognizable similarity is intriguing for two reasons. First, the developmental time and tissue expression pattern of *Adh* in the three species is quite similar. Second, transformants having *D. mojavensis Adh* genes introduced into *D. melanogaster* are expressed according to the developmental program of *D. mojavensis* (C. BAYER and D. SULLIVAN, unpublished data). Presumably any relevant host *D. melanogaster trans*-acting factors used to express the transduced *Adh* genes can recognize these analogous yet dissimilar sequences. This situation is reminiscent of the properties of the yeast regulatory protein *HAP1* which is able to regulate different genes, *CYC1* and *CYC7* by binding to small 5′ regions whose sequences are not similar (PFEIFER, PREZANT and GUARENTE 1987).

DISCUSSION

From the sequence comparisons presented here we have developed a model for the evolution of the *Adh* duplication found in the mulleri subgroup of *Drosophila*. This model is consistent with the evolutionary history of the *D. mojavensis* genes developed above. In addition it includes some assumptions concerning the structure and functions of the *Adh* genes in the genus. First, we assume that the basic *Adh* gene structure for the genus *Drosophila* is essentially that which has been presented for the *D. melanogaster* locus by BENYAJATI *et al.* (1983). A similar structure is also found in *D. pseudoobscura* and *D. affinidisjunta* (SCHAEFFER and AQUADRO 1987; ROWAN and DICKINSON 1988). Second, we assume that no species of *Drosophila* would evolve that does not have ADH

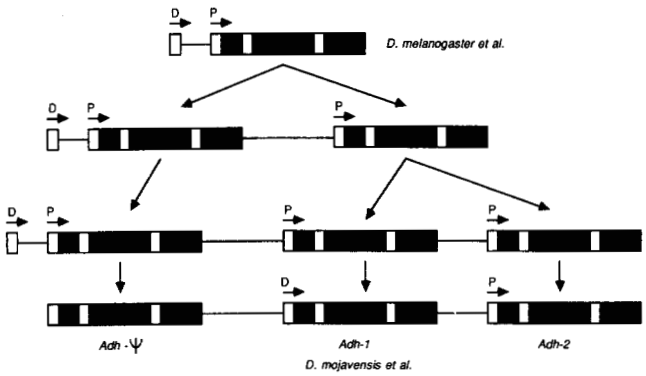


FIGURE 7.—Model of *Adh* gene evolution. D and P, functional distal and proximal promoters, respectively.

activity in both larval and adult stages. Therefore, we propose that an initial duplication event or events occurred starting from a gene similar in structure to that of *D. melanogaster*. This generated an *Adh* locus with one gene similar to *Adh* of *D. melanogaster*, having a proximal and distal promoter separated by a 5′ intron, and one gene that had only a proximal promoter. This second gene would be 3′ to the original gene and might have lost the distal promoter by reason of the extent of the duplication not including this region. Alternatively a deletion of the distal promoter region might have occurred following the duplication. In any case, we find no evidence of the sequences specific to a distal promoter region upstream of the 3′ gene. A species having this *Adh* locus structure would express two *Adh* genes in larvae and one in adults. At a significantly later point in evolution a second event occurred that generated three *Adh* genes arranged in tandem. This second event involved duplication of the most 3′ gene and therefore resulted in two genes each having only a proximal promoter. The lineage represented by this species would have three *Adh* genes, all of which could be expressed in larvae but only one of which would be expressed in adults. Following the second duplication, we propose that the promoter region of the middle gene evolved or more likely had superimposed on it (possibly by upstream enhancers) the capability of acting like a distal promoter. This lineage would then have two genes expressed in adults. In *D. mojavensis* and *D. mulleri* we propose that the most 5′ gene became mutationally inactivated to become a pseudogene. The model is summarized in Figure 7.

The evidence obtained to date which supports this model derives from the DNA sequence comparisons between genes within species and between analogous *Adh* genes of *D. mojavensis* and *D. mulleri*. First, we have argued above that the pseudogene found in these species appears to have been functional for a substantial period. Since *Adh-1* and *Adh-2* genes are more similar to each other than either is to the pseudogene, their origin from a common ancestor is suggested. Furthermore, inspection of the region upstream of the pseudogene reveals a sequence which is identical to the TATA region of the distal promoter of the *D. melanogaster* gene (Figure 6). This supports the hypothesis that the upstream gene is the ancestral gene and once had the dual promoter structure typical of a *Drosophila melanogaster* type *Adh* gene. Further evidence in support of the model will be obtained by analyzing species which have preserved an *Adh* locus structure which represents one of the intermediate structures proposed to link the *D. melanogaster* like gene structure and the *D. mojavensis* structure presented here. Several candidate species have been identified and their analysis is underway.

If this model of *Adh* evolution becomes further substantiated, an interesting issue arises concerning the evolution of the promoter region of *Adh-2*. The 5′ region of *Adh-2* shows significant similarity to the 5′ region of *Adh-1* and the sequence divergence comparisons of these genes and their flanking regions suggest they derive from a common ancestor (Figure 4). However, the regulation of expression of *Adh-1* and *Adh-2* during development is totally different. *Adh-1* of *D. mojavensis* is expressed in cell types in which *D. melanogaster* utilizes the proximal *Adh* promoter. *Adh-2* of *D. mojavensis* is expressed in cell types in which *D. melanogaster* utilizes the distal promoter (BATTERHAM *et al.* 1983b; SAVAKIS, ASHBURNER and WILLIS 1986; FISCHER and MANIATIS 1986). Therefore, it appears that the promoter region of *Adh-2* of *D. mojavensis* is homologous to a proximal promoter yet analogous to a distal promoter. Three mechanisms might have resulted in this pattern of expression. There may have been a deletion in the 5′ region of a gene early in the evolution of the *Adh* duplication that resulted in a distal promoter being brought closer to the gene. Alternatively sequence divergence of a proximal promoter region could have resulted in its gaining the ability to support transcription in adult tissues. Finally, it is possible that the developmental specificity of *Adh-2* expression is generated by sequences further upstream than the region of *Adh-1–Adh-2* similarity. In this regard, FISCHER and MANIATIS (1986) have demonstrated that a region important for *Adh-2* expression is located near or upstream of the *D. mulleri* pseudogene. Similar results have been obtained in our laboratory (C. BAYER and D. SULLIVAN, unpublished data). Consequently we currently favor this last mechanism. The region immediately 5′ to the *Adh-2* gene is probably involved in some aspect of transcriptional control thereby explaining the sequence conservation of this region, but the developmental specificity for *Adh-2* transcription appears to be generated through the function of sequences either within or upstream of the pseudogene. These are probably the same sequence elements which directed the develop-

mental expression from the distal promoter of the ancestral *Adh* gene.

The extensive sequence similarity of *D. mulleri* and *D. mojavensis* in the region upstream of *Adh-2* and extending through and beyond the pseudogene was unexpected. This similarity might be due to selective pressure preserving a function. Alternatively, the sequence similarity could be due to one or more gene conversion events. If the sequence similarity is due to selection for a function, then it is not clear what that function might be. It is clear that these regions contain regulatory sequences which affect *Adh-2* expression. However, it is unlikely that the entire pseudogenic region of several kilobases is involved in *Adh-2* regulation. There are no significant open reading frames on either DNA strand in this region. Any other function remains obscure and could even be related to a flanking gene located further upstream and different from *Adh*.

There is always a likelihood of gene conversion events along a stretch of tandemly repeated DNA. A conversion event in the *Adh* region of *D. mulleri* has been pointed out by FISCHER and MANIATIS (1985). We have inspected the *Adh* region of *D. mojavensis* for evidence of past conversion events. The results were ambiguous. In any case there are several reasons to argue that even though gene conversions may have occurred they are not the basis of the sequence similarity of the pseudogenes of *D. mulleri* and *D. mojavensis*. The two most compelling reasons are that the intron sequences of the pseudogenes are more similar to one another than to the intron sequences of either coding gene of the same species and that the region 5′ to the pseudogene is not at all similar to the region 5′ to *Adh-2*. For gene conversion to be the basis of pseudogene similarity, it would have to be by conversion from the *Adh-2* gene of each species and the resultant similarity of both the pseudogene introns and the pseudogene 5′ region to its adjacent *Adh-2* gene should be obvious. No such similarity is apparent in these regions. Consequently, no fully adequate explanation for the high sequence similarity of the two pseudogene regions is available.

An issue that arises in making comparisons of sequence divergence is deciding what class of sequences to choose in making the comparison. There has been much discussion of this, *e.g.*, see LI, LUO and WU (1985). Ideally, one would like sequences which are varying in response to the mutation rate without selective constraints. The sequence comparison between *Adh* genes within species and between species presented here offer several cautionary examples. It is extremely difficult to define, for the purposes of comparison, 5′ or 3′ flanking nucleotides that have specific function, *e.g.*, the proximal promoter region of the *D. melanogaster* gene and the promoter regions of *Adh-1* of *D. mojavensis* function in a similar manner. However, attempts at locating sequences relevant to the control of expression of these genes by identifying conserved nucleotides have not been fruitful, despite the fact that these control regions can be identified by functional tests. Intron sequences are often suggested as a basis for comparison since these nucleotides do not have an apparent function. Our results indicate that the rates of sequence divergence of the introns in the *Adh* genes of *D. mojavensis* are greater than the rates of synonomous codon substitution in the coding genes. Coding-pseudogene comparison indicates that K_S increases after the gene became a pseudogene, implying release from the selective constraints of codon utilization bias. However the value of K_I remains approximately constant despite the loss of gene function indicating that the selective constraints, if any, have not changed. However it is not evident what selective constraints are operating on the *Adh* pseudogene introns and, as has been discussed above, it is possible that the entire pseudogenic regions of both *D. mojavensis* and *D. mulleri* have an undetermined function. Conservation of sequences in other introns, possibly for different functional reasons, has also been observed (see discussion in KASSIS *et al.* 1986).

The use of synonymous codons substitution is probably the most commonly used parameter in making comparisons. Since codon bias is not constant in all lineages (*e.g.*, discussion in ASHBURNER, BODMER and LEMEUNIER 1984), caution must be exercised in using these sequences. The use of changes in synonymous codons therefore seems most justifiable in making comparisons in relatively recent diverged lineages.

We have refrained from calculating the divergence times of the genes within a species or of the species we have compared since our arguments do not depend on the absolute value of divergence times. Calculation of the divergence time requires an assumption as to the average rate of nucleotide substitutions (α). This is a controversial parameter. One approach that could be used to put our results in perspective with other analyses of the molecular evolution in the genus *Drosophila*, is to use a mammalian nucleotide substitution rate of 5.5×10^{-9} nucleotide per site per year as used by BODMER and ASHBURNER (1984). Using this value and the synonymous codon substitutions, K_S, we estimate that the time since divergence of *D. mojavensis* and *D. mulleri* has been approximately 16.7–18.9 million years and that *Adh-1* and *Adh-2* diverged from each other about 17.9 million years ago. The time of the first duplication event which generated the ancestors of the pseudogene and of *Adh-1* and *Adh-2* is not possible to estimate because the K_S for the pseudogene vs. either *Adh-1* or *Adh-2* probably results in two separate rates, one before and one after the gene

became a pseudogene. The timing for the first duplication event is thought to be coincident with the radiation of the mulleri subgroup into arid regions more or less during the Miocene epoch (BATTERHAM *et al.* 1984). This view is supported by the widespread existence of duplicate *Adh* genes in mulleri complex species implicating a single initial event.

Another approach to these calculations is to take the time of origin of the genus *Drosophila* as 60 million years ago (THROCKMORTON 1975) and assume that the divergence of the lineages leading to *D. mojavensis* and *D. melanogaster*, representatives of the two major subgenera, occurred at about that time. In this case the *Adh-1–Adh-2* divergence based on relative K_S would be about 20% of the *D. mojavensis–D. melanogaster* divergence or about 12 million years. The two approaches yield values which are reasonably similar and represent our present best guesses on the time of the *Adh-1–Adh-2* duplication.

This research was supported by U.S. Public Health Service grant GM 31857 to D.T.S. We thank JANICE FISCHER, TOM MANIATIS and W. J. DICKINSON for sharing their results with us prior to publication. ANNE SMARDON and BENJAMIN METCALF provided excellent technical assistance.

LITERATURE CITED

ASHBURNER, M., M. BODMER and J. LEMEUNIER, 1984 On the evolutionary relationships of *Drosophila melanogaster*. Dev. Genet. **4:** 295–312.

BATTERHAM, P., E. GRITZ, W. T. STARMER and D. T. SULLIVAN, 1983a Biochemical characterization of the products of the *Adh* loci of *Drosophila mojavensis*. Biochem. Genet. **21:** 871–883.

BATTERHAM, P., J. A. LOVETT, W. T. STARMER and D. T. SULLIVAN, 1983b Differential regulation of duplicate alcohol dehydrogenase genes in *Drosophila mojavensis*. Dev. Biol. **96:** 5553–5567.

BATTERHAM, P., G. K. CHAMBERS, W. T. STARMER and D. T. SULLIVAN, 1984 Origin and expression of an alcohol dehydrogenase gene duplication in the genus *Drosophila*. Evolution **38:** 644–657.

BENYAJATI, C., N. SPOEREL, H. HAYMERLE 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.

BIGGIN, M. D., T. S. GIBSON and G. G. HONG, 1983 Buffer gradient gels and ^{35}S label as an aid to rapid DNA sequences determination. Proc. Natl. Acad. Sci. USA **80:** 3963–3965.

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.

FISCHER, J. A., and T. MANIATIS, 1985 Structure and transcription of the *Drosophila mulleri* alcohol dehydrogenase genes. Nuclei Acids Res. **13:** 6899–6917.

FISCHER, J. A., and T. MANIATIS, 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,

HAYASHIDA, H., and T. MIYATA, 1983 Unusual evolutionary conservation and fragment DNA segment exchange in class 1 genes of the major histocompatibility complex. Proc. Natl. Acad. Sci. USA **80:** 2671–2675.

HEBERLEIN, U., B. ENGLAND and R. TJIAN, 1985 Characterization of *Drosophila* transcription factors that activate the tandem promoters of the alcohol dehydrogenase gene. Cell **41:** 965–977.

HENIKOFF, S., 1984 Unidirectional digestion with exonuclease III creates targeted breakpoints for DNA sequencing. Gene **28:** 351–359.

HONG, G. F., 1982 A systematic DNA sequencing method. J. Mol. Biol. **158:** 539–549.

KASSIS, J. A., S. J. POOLE, D. K. WRIGHT and P. O'FARRELL, 1986 Sequence conservation in the protein coding and intron regions of the *engrailed* transcription unit. EMBO J. **5:** 3583–3589.

LI, W.-H., T. GOJOBRI and M. NEI, 1981 Pseudogenes as a paradigm of neutral evolution. Nature **292:** 237–239.

LI, W.-H., C. C. LUO and C.-I. WU, 1985 Evolution of DNA sequences. In: *Molecular Evolutionary Genetics*, Edited by R. J. MACINTYRE. Plenum Press, New York.

MILLS, L. E., P. BATTERHAM, J. ALEGRE, W. T. STARMER and D. T. SULLIVAN, 1986 Molecular genetic characterization of a locus that contains duplicate *Adh* genes in *Drosophila mojavensis* and related species. Genetics **112:** 295–310.

MIYATA, T., and H. HAYASHIDA, 1981 Extraordinary high evolutionary rate of pseudogenes: evidence for the presence of selective pressure against changes between synonymous codons. Proc. Natl. Acad. Sci. USA **78:** 5739–5743.

MIYATA, T., T. YASUNAGA and T. MISHIDA, 1980 Nucleotide sequence divergence and functional constraint in mRNA evolution. Proc. Nat. Acad. Sci. USA **77:** 7328–7332.

OAKESHOTT, J. G., G. K. CHAMBERS, P. D. EAST, J. B. GIBSON and J. S. F. BARKER, 1982 Evidence for a genetic duplication involving alcohol dehydrogenase genes in *Drosophila buzzatii* and related species. Aust. J. Biol. Sci. **35:** 73–84.

PFEIFER, K., T. PREZANT and L. GUARENTE, 1987 Yeast *HAP1* activator binds to two upstream activation sites of different sequence. Cell **49:** 19–27.

POTTER, S. S., 1982 DNA sequence of a foldback transposable element in *Drosophila*. Nature **297:** 201–204.

ROWAN, R. G., and W. J. DICKINSON, 1988 Nucleotide sequence of the genomic region coding for alcohol dehydrogenase in *Drosophila affinidisjuncta* (in press).

SAVAKIS, C., M. ASHBURNER and J. H. WILLIS, 1986 The expression of the gene coding for alcohol dehydrogenase during the development of *Drosophila melanogaster*. Dev. Biol. **114:** 207.

SCHAEFFER, S. W., and C. F. AQUADRO, 1987 Nucleotide sequence of the *Adh* gene region of *Drosophila pseudoobscura*: evolutionary change and evidence of an ancient gene duplication. Genetics **117:** 61–73.

THROCKMORTON, L. H., 1975 The phylogeny, ecology and geography of *Drosophila*. In: *Handbook of Genetics*, Edited by R. C. KING. Plenum Press, New York.

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.

Communicating editor: C. C. LAURIE