J Mol Evol (2001) 52:29–39
DOI: 10.1007/s002390010131

JOURNAL OF MOLECULAR EVOLUTION

Characterization of Two Alcohol Dehydrogenase (*Adh*) Loci from the Olive Fruit Fly, *Bactrocera* (*Dacus*) *oleae* and Implications for *Adh* Duplication in Dipteran Insects

George N. Goulielmos,[1] Nickolaos Cosmidis,[1] Michael Loukas,[1] Spyros Tsakas,[1] Eleftherios Zouros[2,3]

[1] Department of Agricultural Biotechnology, Agricultural University of Athens, Iera Odos 75, Votanikos, 118 55 Athens, Greece
[2] Department of Biology, University of Crete, Vasilika Vouton, 714 09 Iraklion, Crete, Greece
[3] Institute of Marine Biology of Crete, Iraklion, Crete, Greece

Received: 30 May 2000 / Accepted: 17 August 2000

Abstract. We report the cloning and structural characterization of two *Adh* loci of the olive fruit fly, *Bactrocera oleae*. Each of the two genes, named *Adh1* and *Adh2*, consists of three exons and two introns for a total length of 1981 and 988 nucleotides, respectively. Their deduced amino acid sequences of 257 and 258 residues exhibit a 77% identity and display the characteristics of the insect ADH enzymes, which belong to the short-chain dehydrogenases/reductases family. The *Adh* genes of *B. oleae* are compared to the two genes of the Mediterranean fly, *Ceratitis capitata*, the only other species of the *Tephritidae* family in which the *Adh* genes have been studied. On the basis of amino acid divergence the four genes form two clusters each containing one gene from each species, as expected if there was one duplication event before speciation. On the basis of nucleotide sequence the four sequences form two clusters each containing the two sequences from the same species, as expected if there was a separate duplication event in each species. To help decide between the two alternatives, we compared at both the amino acid and DNA level the *Adh* genes of five *Drosophila* species that are known to carry two such genes and observed that, with only one exception at the amino acid level, conspecific loci cluster together. We conclude that the information we have at present does not allow a firm choice between the hypothesis of a single duplication event that occurred before the split of *Bactrocera* and *Ceratitis* from their common ancestor and the hypothesis of two independent duplication events, one in each of the two genera.

Key words: Alcohol dehydrogenase gene — Gene duplication — *Tephritidae*

Correspondence to: G.N. Goulielmos

Introduction

The principal function of the enzyme alcohol dehydrogenase (ADH; EC 1.1.1.1) in insect metabolism is the catalysis of the reversible conversion of various alcohols generated by microbial fermentation in larval and adult feeding sites to their corresponding aldehydes and ketones.

The *Adh* genes from several *Drosophila* species have been cloned and sequenced. In *D. melanogaster* the ADH enzyme is encoded by a single gene with two promoters, activated at different developmental stages (Benyajati et al. 1983; Savakis and Ashburner 1985). This pattern appears to be true in all species of the subgenera *Sophophora* and *Scaptodrosophila*. In contrast, species belonging to subgenus *Drosophila*, such as *D. repleta* and *D. virilis*, contain a duplication of the *Adh* gene producing two distinct isozymes (Batterham et al. 1984; Nuzminsky et al. 1996; Begun 1997). The ADH enzyme system of *Drosophila* has also been studied with the aim of understanding the mechanisms of maintenance of genetic polymorphism in natural populations

(Gibson and Oakeshott 1982; Van Delden 1982; Kreitman 1983; Bodmer and Ashburnee 1984; David 1988; Kreitman and Hudson 1991). To detect biochemical differences of the enzyme among different species or among isozymes from the same species, the enzyme has been purified from several *Drosophila* species (Sofer and Ursprung 1968; Thatcher 1980; Juan and Gonzalez-Duarte 1980, 1981; Batterham et al. 1984; Moxon et al. 1985).

Beside *Drosophilidae*, our knowledge of ADH in insects is limited and is mostly concentrated on two *Tephritidae* species, the medfly *Ceratitis capitata* and the olive fruit fly, *Bactrocera oleae*, both of which are serious agricultural pests. Apart from evolutionary and functional aspects, the ADH enzyme system of these species is of interest because of its potential use in the biological control of the insects (e.g., Robinson et al. 1986; Robinson and MacLeod 1993 for medfly and Zouros et al. 1982 for *B. oleae*, respectively). The ADH from *C. capitata* (medfly, Diptera; *Tephritidae*) is well characterized. Biochemical and genetic studies have shown that two ADH proteins exist in this insect (Gasperi et al. 1992), encoded by two tightly linked genes on the second chromosome, probably generated by gene duplication (Malacrida et al. 1992; Gasperi et al. 1994).

In *B. oleae* biochemical studies have shown the existence of an allozyme polymorphism with three alleles, the fast (F), the intermediate (I), and the slow (S), mapping at a single locus (Zouros et al. 1982). Artificial rearing of the insect in the laboratory results in drastic changes in its physiology, behavior, and reproductive biology (Mazi et al. 1998a). These changes are accompanied with a rapid increase of the frequency of allele I from the low levels of 1% to 3% that occur in natural populations to about 35%. Even though a tight causal relationship between allozyme frequency change and changes in the physiology and behavior of the insect under laboratory conditions remains to be established, there is strong evidence that these changes are controlled by frequency-dependent selection (Cosmidis et al. 1999). Further insight into the evolutionary and adaptational significance of ADH enzymes in this insect requires that the *Adh* system be studied at the molecular level. This would allow the comparison of *Adh* system of this insect with that of the medfly and, at a broader level, with that in drosophilids. Recently, a cDNA encoding for an ADH enzyme has been cloned and sequenced in *B. oleae* (Benos et al. 2000). Even though this study referred to one *Adh* gene in *B. oleae*, there exist biochemical information that suggests the presence of two isozymes in this species (Gasperi et al. 1994).

We report here the successful cloning and sequence analysis of genomic regions that contain two *Adh* loci of *B. oleae*. We then take advantage of this finding to address the question of whether the presence of two *Adh* genes in tephritid species is the result of an early duplication event that predates the emergence of the two genera or of independent duplication events that occurred after the genera had emerged.

Materials and Methods

B. oleae *Flies.* The olive fruit fly stock used in this study was the ADH-FF homozygous line maintained at the Laboratory of Genetics, Agricultural University of Athens. The origin, extraction and maintenance of the stock is described in detail by Cosmides et al. (1997).

Cloning of the Olive Fruit Fly Adh1 *and* Adh2 *Genes.* Preparation of genomic DNA was done according to the protocol described by Holmes and Bonner (1973). The cDNA sequences of *C. capitata* (GenBank Accession nos. Z30194 and Z30195) and the single *Adh* cDNA of *B. oleae* (EMBL Accession no. AJ2500007) were used to design primers for PCR amplification of the corresponding genomic fragments. The upstream primers 5′-ACGCGTCGACGAATTCATGAG(C/T)TTGGCIGGIAAAAA(C/T)G-3′ and 5′-ACGCGTCGACGAATTCATGGGTTTGAGCGGCAAAAAT-3′ and the downstream 5′-ACGGAGCTC(G/A)TAIGTGGG(T/C)TCCCA(G/A)TAIAC-3′ and 5′-CCGAGCTCGGATCCCTAGTTGAATGTGGGTTGCCA-3′ were used for the search of an *Adh1* product and an *Adh2* product, respectively. The resulting PCR products contained *Eco*RI and *Sal*I overhangs, which allowed their directional cloning into the plasmid vector pBluescript II KS (Stratagene). Because of an *Eco*RI internal site in the fragment that corresponded to the *Adh1* gene, the pGEM (Promega) vector was used for the cloning of this fragment. In both cases standard PCR amplification procedures were followed (Sambrook et al. 1989). Additional degenerate primers were used for nested PCR, designed according to the sequence of the ADH1 cDNA of *C. capitata*, as a first step to verify that the PCR fragment tentatively referred to as *Adh1* was a genuine *Adh* locus. To this end, the upstream oligonucleotide 5′-ACGCGTCGACGAATTCTTCGTIGGIGGITTGGGCTTCATIG-3′ and the downstream oligonucleotides 5′-ACGGAGCTCCTC(C/T)ACATTGGGATCGGTGAG(G/T)ATGCC-3′ and 5′-ACGGAGCTCCAA(T/A)GTGCCTTGGTTGCTIAT(A/G)TA(A/G)ATG-3′ were used. Restriction and DNA modification enzymes were provided from MINOTECH and New England Biolabs. *Pwo* polymerase (a proofreading enzyme) (Boehringer-Mannheim) was used to get amplification products of high fidelity. Sequencing of the double-stranded plasmids was carried out according to the di-deoxy-chain termination method following the manufacturer's protocol (Sequenase, USB), using either vector-specific (T3, T7, SP6) or custom gene-specific primers. For each genomic region, both strands were completely sequenced and the consensus nucleotide sequence was obtained for two different clones from at least three sequencing reactions. Agarose gel electrophoresis and other recombinant DNA methods were performed essentially as described in Sambrook et al. (1989).

DNA Sequence Analysis. The DNA sequences were analyzed with the GCG Sequence Analysis Software Computer Package. The alignment of the sequences was done using the Clustal X program (Thompson et al. 1997). The nucleotide sequences reported in this study have been deposited in the EMBL Database (Accession nos.: *B. oleae Adh1,* AJ277835; *B. oleae Adh2,* AJ277834).

Phylogenetic Tree Construction. Phylogenetic trees were constructed using the neighbor-joining, UPGMA, and maximum parsimony methods, through the MEGA computer package (Kumar et al. 1993). Kimura's (1980) two-parameter distance was used. To assess the confidence of individual nodes a bootstrap analysis (Felsenstein 1985) with 1000 replications was performed using the same computer package. The same package was also used to estimate the transition/transversion ratio. The rates of synonymous (K_s) and nonsynonymous

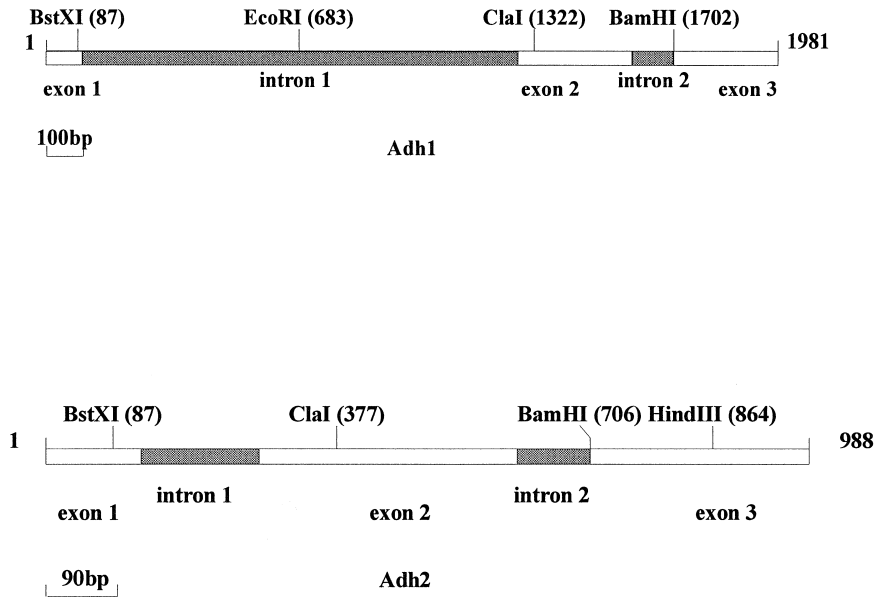

Fig. 1. Physical map of the *Adh1* and *Adh2* genomic DNA using the restriction enzymes *Eco*RI, *Hind*III, *Bst*XI, *Bam*HI, and *Cla*I. White boxes represent the exons and gray-shaded boxes the introns. Bars indicate the scale (in base pairs) that these physical maps were drawn.

substitutions (K_a) were estimated using the DnaSP computer program (Rozas and Rozas 1999).

The following nucleotide sequences, available from the EMBL and GenBank databases, were used: *C. capitata Adh1:* Z30194; *C. capitata Adh2:* Z30195; *B. oleae Adh1:* AJ277835; *B. oleae Adh2:* AJ277834; *D. melanogaster Adh:* U75652; *D. buzzatii Adh1* and *Adh2:* U65746; *D. montana Adh1:* U26842; *D. montana Adh2:* U26845; *D. hydei Adh1* and *Adh2:* X58694; *D. virilis Adh1* and *Adh2:* U26846; *D. mojavensis Adh1* and *Adh2:* X12536; and *Sarcophaga penegrina Adh:* D63669. The latter was used to provide a common outgroup for the comparison of the 4 tephritid and of the 10 drosophilid sequences.

Results

The Adh2 *Locus*

The deduced amino acid sequences of one PCR product from the *B. oleae* genomic DNA exhibited high homology (85%) with the *Adh2* sequence of *C. capitata.* We will therefore consider this to be homologous to the *Adh2* locus of *B. oleae.* The corresponding cDNA is identical to that reported by Benos et al. (2000) for the same species, with the exception of two amino acid residues that may correspond to the electrophoretic differences at the ADH enzyme of the two strains used; we have used a line that was fixed for the fast (F) allozyme, while Benos et al. (2000) used a line fixed for the slow (S) allozyme. This confirms that the cDNA isolated by Benos et al. (2000) corresponds to the *Adh2* locus of *B. oleae.* The position and size of the introns were obtained by sequencing the genomic DNA from the initial ATG codon to the stop codon (TAG). There are two introns of 125 and 82 nucleotides each, located immediately after amino acid residues 32 and 165 of the cDNA. The intron/exon organization and the restriction map of the genomic *Adh2* DNA are shown in Fig. 1.

The Adh1 *Locus*

The use of degenerate primers based on the cDNA sequence of the *Adh1* of *C. capitata* produced a ~2-kb fragment. To determine whether this product contained an *Adh*-like sequence, nested PCR was performed using a series of degenerate primers based on internal sequences of the *Adh1* of *C. capitata.* This produced positive evidence for the existence of a coding sequence, so we proceeded in the cloning and sequencing of the fragment. The *Adh1* genomic region that resulted consisted of three exons interrupted by two introns and its coding sequence, with 257 codons, exhibits a considerably high nucleotide sequence identity with the *Adh2* coding region of *B. oleae.* The positions of the introns are identical to the ones found for the *Adh2* gene, located again immediately after nucleotides 32 and 165 of the cDNA. However, the sizes of the introns are quite different from those of the *Adh2.* The first intron is 1071 bp long and the second 136 bp. The organization and the restriction map of the *Adh1* genomic region are shown in Fig. 1.

These data provide firm evidence for the presence of a second *Adh* gene in *B. oleae.* Benos et al. (2000) reported only one *Adh* gene in this species but left open the possibility of a second. The deduced amino acid sequence for *Adh1* of *B. oleae* was found to have a nucleotide divergence of 22% compared to *Adh1* of *C. capitata.* Moreover, the two ADH isozymes of *B. oleae* share approximately 86% amino acid sequence identity.

There is indirect evidence that the *Adh1* gene of *B. oleae* also produces an active protein product, even though this must not yet be considered as proven. Gasperi et al. (1994) reported that when ADH allozymes of *B. oleae* were separated by isoelectric focusing the zymograms were suggestive of two ADH allozyme loci. The difficulty in detecting the second ADH allozyme of

```
BoADH1   -MSLAGKNVVFVGGLGFIGYEACKTLITRDLASSLVFDIWDKPEAVKALEEINPKTKVFYT
BoADH2   -MGLSGKNVVFVGGLGFIGYEACKQIMTKNVASFFVFDVLENAENIKALQAINPKTKVYYT
CcADH1   -MSLAGKNVVFVGGLGFIAYEACKYLMNNDLASLFVFDVLDKPEAIKALQEINPKTKVYYT
CcADH2   -MGLSGKNVIFVGGLGFIGYEACKQLMAKNMASFFVFDVLDKPENIKALQALNPKTKVYYT
DmADH    SFTLTNKNVIFVAGLGGIGLDTSKELLKRDLKNLVILDRIENPAAIAELKAINPKVTVTFY
            *  *** ** *** *     *             *          *   ***  *

BoADH1   KFDITS-KDNIKQSLADVIAKVQYIDVLVNGAGILTDPNVEWTLNINLIGLINTTLEAIPL
BoADH2   KFDITN-KASIKSALADVIAKVQYIDVLVNGAGILTDPNVELTMNINLIGLINTTLEAIPL
CcADH1   KFDITN-KESIKQSLADVISKVQHIDALINGAGILTDPNVELTMNINLIGLINTTLEALPL
CcADH2   KFDITS-KQSIKSALADVVAKVKYIDALINGAGILTDPNVELTMNINLIGLINTTLEGLPL
DmADH    PYDVTVPIAETTKLLKTIFAQLKTVDVLINGAGILDDHQIERTIAVNYTGLVNTTTAILDF
           * *                    * * ****** *   * *   *  ** ***

BoADH1   MDKNKKGRGGVIVNIASVLGLEPGPPAAIYCASKFGVMGFSRSLGDPHYYEHTGIAVVTF
BoADH2   MDKNKKGRGGLIVNIASVLGLEPAPPGAIYCASKFGVMGFSRSISDPYYYNLTGIAVATF
CcADH1   MDKNKHGRGGVIVNIASVLGLEPCPPAAVYCASKFGVVGFSRSLGDPFYYEHTGVAVVTF
CcADH2   MDKNKQGRGGVIVNIASVLGLEPCPPAAVYCASKFGVMGFSRSIGDPYYYNITGVAVVTF
DmADH    WDKRKGGPGGIICNIGSVTGFNAIYQVPVYSGTKAAVVNFTSSLAK--LAPITGVTAYTV
          ** * * ** * ** ** *         *   *     *  *          **    *

BoADH1   CPGLTDTPLKNNFATKYTFDYSKEIGEKLNHCKTQKPEVCGAHLAQVVELRDNGAIYISN
BoADH2   CPGLTETPLKNNIATKYTFEYSKVIGDKLNNTKTQKPEACGAHLAQVLDTAENGGIYISN
CcADH1   CPGLTDTPLKNNIGSKYTFDYSKEIGEKLNSSKTQKPEVCGAHLAQAIELMDNGAIYISN
CcADH2   CPGLTETPLKNNIGSKYTFEYSKKISEELNSTKTQKPEVCGAHLAQVVESHENGGIYISN
DmADH    NPGITRTTLVHKFNS--WLDVEPQVAEKLLAHPTQPSLACAENFVKAIELNQNGAIWKLD
          ** * * *                   *    **    *            ** *

BoADH1   EGTLTKVKPSVYWEPTY-
BoADH2   QGTLSKVTPTVYWQPTFN
CcADH1   QGTLTKVKPSVYWEPTY-
CcADH2   QGTLAKVTPTVYWQPTYH
DmADH    LGTLEAIQWTKHWDSGI-
          ***        *
```

Fig. 2. Multiple alignment of the deduced amino acid sequence of the *Adh* genes of *B. oleae* (BoADH1 and BoADH2) and *C. capitata* (CcADH1 and CcADH2). The *Drosophila melanogaster* ADH protein (DmADH) is also given to highlight the conservation of residues characteristic of short-chain dehydrogenases. Gaps have been introduced to obtain maximum matching. Identical residues are indicated with asterisks. Bold residues are considered essential for enzymatic activity.

B. oleae is compatible with the fact that the second ADH allozyme of the medfly is also difficult to detect, as it is expressed only in the insect's muscle (Benos et al. 2000).

Amino Acid Alignment of ADH Sequences

Amino acid sequences of the two *B. oleae* and *C. capitata Adh* genes were aligned together with the corresponding sequence of *D. melanogaster* (Fig. 2). Ignoring terminal residues, all four tephritid sequences have an internal deletion of one amino acid and two additions of two amino acids each, for a net increase of three amino acid residues. The ADH1 of both *B. oleae* and *C. capitata* are shorter compared to ADH2 of the same species by one amino acid at the very end of the sequence. Nine amino acid residues that were determined to be particularly important for enzymatic activity in *D. melanogaster* (Chen et al. 1993; Cols et al. 1993; Jornvall et al. 1995) were also found to occur in the same positions in all four tephritid ADH products (Fig. 2). These amino acids are apparently conserved in many insect *Adh* loci as pointed out by Benos et al. (2000). Further similarities in amino acid sequence among *Adh* genes of the two tephritids or among tephritids and drosophilids can be attributed to either selective constraint or phylogenetic relationships (see below).

Phylogenetic Trees of the Tephritid Adh *Genes and Differences in Amino Acid and Nucleotide Substitution Rates*

The availability of cDNA sequences for the two *Adh* genes of *B. oleae* and *C. capitata* can be used to answer the question of whether there has been one duplication

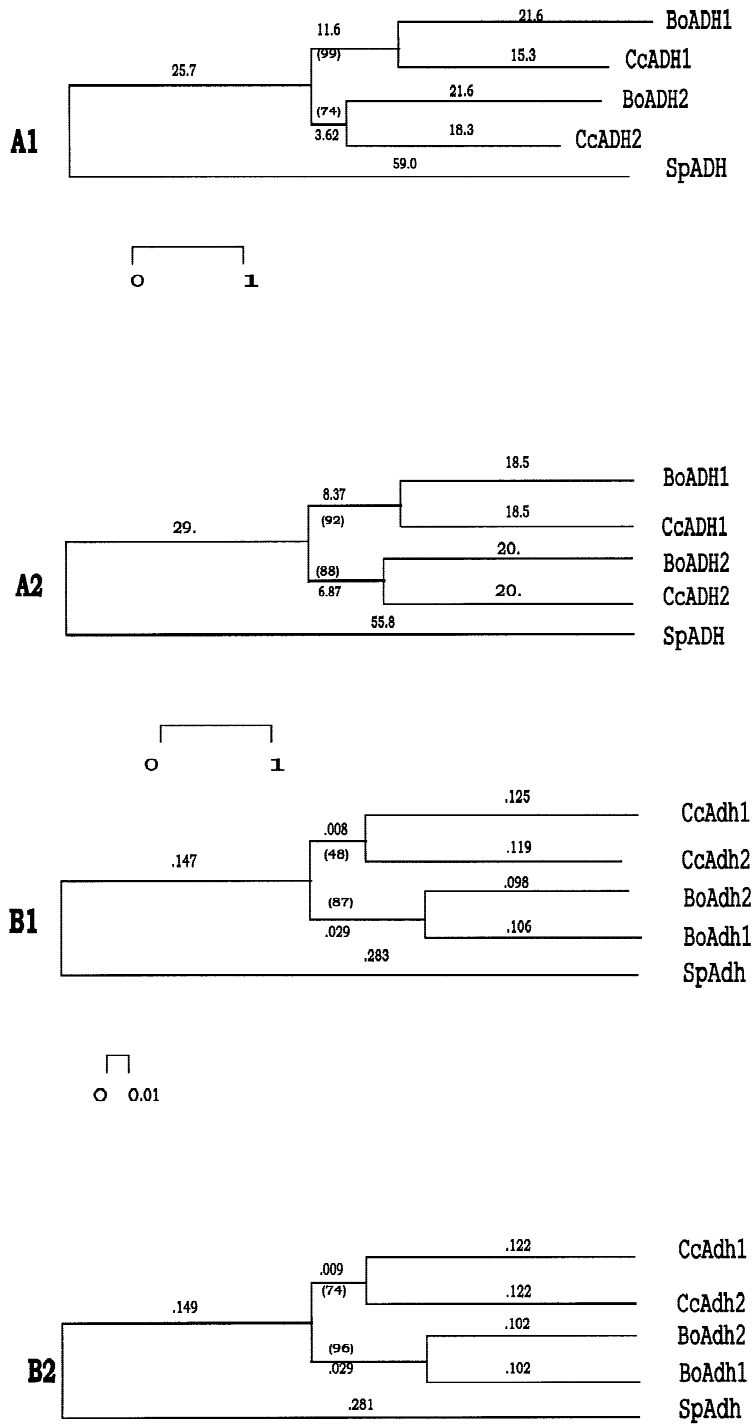

Fig. 3. Phylogenetic trees of four tephritid *Adh* genes. The *Adh* gene of *Sarcophaga penegrina* was used as outgroup. Numbers give branch lengths in number of amino acid **(A)** or nucleotide substitutions **(B)** and numbers in parentheses are bootstrap confidence values. **A:** Amino acid trees, **B:** Nucleotide trees. A1 and B1 are neighbor-joining and A2 and B2 are UPGMA trees. The bar below each tree indicates the distance measure.

event that preceded the separation of the two species and generated the *Adh1* and *Adh2* genes (which, then, would be considered as paralogous to each other, with *Adh1* or *Adh2* of one species being orthologous to the corresponding gene of the other species) or there have been two separate and independent duplication events, one in each of the two species (in which case *Adh1* and *Adh2* will be pairs of paralogous genes within each species). We have used the ADH amino acid sequence of *Sarcophaga penegrina* (Horio et al. 1996) to construct neighbor-joining and UPGMA trees (Fig. 3A). In both trees, the four ADH isozymes cluster according to "type" (i.e., *Adh1* with *Adh1* and *Adh2* with *Adh2*) than "species."

A careful inspection of the data suggests a more complex pattern in the rate of amino acid substitutions in the four proteins. The 257 amino acid positions can be classified in five classes according to the number of different amino acids in the four sequences (Table 1). The first class (4:0) consists of sites in which the same amino acid occurs in all four sequences. These are assumed to be the most conservative sites. The second most conservative

Table 1. Classification of amino acid sites of the four ADH proteins according to shared amino acid residues

	Number of sites
1. Residue same in all four sequences (pattern 4:0)	180
2. Residue same in three sequences (pattern 3:1)	
a. Residue different in BoADH1	12
b. Residue different in BoADH2	12
c. Residue different in CcADH1	5
d. Residue different in CcADH2	7
3. Two different residues, each in two sequences (pattern 2:2)	
a. BoADH1 = BoADH2 and CcADH1=CcADH2	7
b. BoADH1 = CcADH1 and BoADH2=CcADH2	16
c. BoADH1 = CcADH2 and BoADH2=CcADH1	1
4. Three different residues (pattern 2:1:1)	
a. BoADH1 = BoADH2 and CcADH1≠CcADH2	2
b. BoADH1 ≠ BoADH2 and CcADH1=CcADH2	2
c. BoADH1 = CcADH1 and BoADH2≠CcADH2	5
d. BoADH1 ≠ CcADH1 and BoADH2 = CcADH2	5
e. other	1
5. Four different residues (pattern 1:1:1:1)	1
Total	257

ADH products are symbolized as in Fig. 3. The symbols = and ≠ mean the amino acid residue is the same or different, respectively.

class is of the 3:1 form, i.e., sites in which only one sequence has a different amino acid. There are 36 such sites, 24 of which occur in one or the other sequence of *B. oleae* and 12 in one or the other sequence of *C. capitata*. This is a statistically significant difference. On the contrary, there is no difference among conspecific sequences either within *B. oleae* or *C. capitata* (Table 2). This suggests that the two ADH proteins of *B. oleae* evolve faster at the amino acid level than the corresponding ADH proteins of *C. capitata* (Table 2). The third class is of the 2:2 form, i.e., sites in which there occur two amino acids, each in two sequences. The important observation here is that the sites at which the proteins of same "type" (i.e., type "1" or "2") have the same amino acid are twice as common as the sites of same species. A similar trend occurs in the fourth class of the 2:1:1 form, i.e. three amino acids at a site of which one is shared by two sequences. Here the sites in which the common amino acid occurs in sequences of same type outnumber those in which in common amino acid occurs in sequences of same species. The latter two classes can be used jointly to ask what is the probability that a site will have the same amino acid residue in two sequences of same type as opposed to two sequences of same species. These two probabilities are significantly different in favor of sequences of the same type (Table 2). The overall conclusion is that the *B. oleae* genes evolve faster at the amino acid level than the *C. capitata* genes, yet there is a higher similarity between loci of same type than between loci of same species.

The trees produced from the nucleotide sequence of the four cDNAs are quite different from the ones obtained from amino acid sequences. Both the neighbor-joining and the UPGMA trees cluster the sequences according to species (Fig. 3B). The maximum parsimony method produced the same result (data not shown). Finally, from the calculation of the rates of synonymous (K_s) and non-synonymous (K_a) substitutions (Table 2), one may see that the two lower K_a values correspond to the "same type" comparisons (*Adh1* to *Adh1* and *Adh2* to *Adh2*), whereas the two lower K_s values correspond to the two "same species" comparisons. Consequently, trees constructed on K_a and K_s values differ in exactly the same way as the trees produced from amino acids and nucleotides: Synonymous substitutions clustered the sequences according to species, whereas non-synonymous substitutions clustered them according to type (data not shown). The possibility that the nucleotide divergence has reached the saturation point, so that no meaningful phylogenetic information can be extracted from their comparison, is not very strong given that the transversions/transitions ratio varies from 0.58 (*B. oleae Adh1* versus *C. capitata Adh1*) to 1.08 (*B. oleae Adh1* to *B. oleae Adh2*), but it exceeds the value of one only in the latter case. But even if we assumed that the nucleotide sequences have diverged to the point that no phylogenetic signal can be extracted from their comparison, the question remains what is the probability that the branch topography of Fig. 3B that joins the sequences according to species could be obtained by chance alone. There are 15 bifurcating rooted trees one may obtain for four endpoints (Li 1997), which means that each of the two probable evolutionary histories, that of Fig. 3A or Fig. 3B, have each a probability of 0.067 to arise by chance. This probability is slightly higher than the conventional 5% for the rejection of the null hypothesis that the clustering of nucleotide sequences by species is accidental. However, the observation that a similar clustering occurs in *Drosophila* (see below) reinforces the view that nucleotide saturation has not erased the phylogenetic signal in the four tephritid sequences and that duplication after speciation remains a viable hypothesis.

Discussion

This is the first report of a second *Adh* gene in *B. oleae*. To comply with the nomenclature followed for the *Adh* genes of *C. capitata*, we have named these sequences *Adh1* and *Adh2*. The cDNA of the *Adh2* locus was previously identified following a cloning procedure based on the functional complementation of appropriate yeast mutants (Benos et al. 2000). The present study provides data that fully demonstrate the presence of a second *Adh* gene in the olive fruit fly. Together with *C. capitata* these are the two species of the *Tephritidae* family whose *Adh* genes have been studied in any extent. The fact that both species have two *Adh* genes suggests that this may be a common feature of the family *Tephritidae*. This

Table 2. Tests for homogeneity in differences among the four *Adh* amino acid sequences (from Table 1) and nonsynonymous K_a and synonymous K_s values of the corresponding DNA sequences

1. Presence of a single amino acid difference (pattern 3:1)

Species	Type		
	Adh1		*Adh2*
B. oleae	12		12
C. capitata	5		7
	d.f	x^2	*p*
Between species	1	4.000	0.046
Between types	1	0.111	0.739
Interaction	1	0.223	0.637

2. Common amino acid (patterns 2:2 and 2:1:1)
   a. In at least one conspecific pair 11
   b. In at least one pair of same type 26
   $x^2 = 6.081$, d.f. = 1, P = 0.014
3. K_a (above the diagonal) and K_s (below the diagonal).

	Bo*Adh1*	Bo*Adh2*	Cc*Adh1*	Cc*Adh2*
Bo*Adh1*	—	0.1158	0.0799	0.1329
Bo*Adh2*	0.5292	—	0.1222	0.0944
Cc*Adh1*	1.4190	1.1052	—	0.1020
Cc*Adh2*	1.0636	1.0874	0.9198	—

conclusion may, however, be premature in view of the fact that in *Drosophilidae* there are species groups with one and others with two functional *Adh* genes.

The deduced amino-acid sequences confirmed that the two isozymes are highly related to each other and to known ADHs from other insects. The two ADH proteins of *B. oleae* are therefore classified as members of the short-chain dehydrogenases/reductases (SDR) family. All these proteins have a polypeptide consisting of about 250 amino acid residues. It is worth mentioning that most members of this family are characterized by distant duplications and divergence, are functionally and structurally related, and lack the zinc-liganding cysteine residues in their coenzyme binding regions (Jornvall et al. 1981, 1995; Persson et al. 1991). In contrast, most ADHs from other organisms, including those of mammals, plants, and yeasts, belong to the medium-chain family of dehydrogenases/reductases (MDRs) (formerly "long-chain") that have a longer polypeptide of about 370 residues and usually contain zinc ligands in their active site (Jornvall et al. 1981).

The *Adh1* and *Adh2* genes are known to have different expression patterns and tissue-specific distributions in *C. capitata* (Benos et al. 2000). *Adh1* seems to be expressed in muscle and was separated from third-instar larvae by ion exchange chromatography (Gasperi et al. 1994). *Adh2* is expressed in gut, ovaries, and fat body and can be detected in all stages. The *Adh* genes of *B. oleae* that we describe here have not been examined for differences in stage- or tissue-specific expression. It remains, therefore, a matter of interest whether the *Adh2* of *B. oleae* will have the expression characteristics of *Adh2* of *C. capitata,* which has been studied more intensively than *Adh1*. It also remains to identify which of the two *Adh* genes of *B. oleae* corresponds to the locus that segregates for the three ADH electrophoretic allozymes of *B. oleae.* Again the fact that these allozymes are abundantly expressed in all life stages of *B. oleae* suggests that they are coded by *Adh2.* The ADH allozymes of *B. oleae* represent a polymorphism under selection, possibly in response to the concentration of various types of alcohols in the larval food (Mazi et al. 1998b). Given that any attempt to control this insect would require its rearing on artificial food, the response of the ADH enzyme polymorphism to this food will determine the suitability of the artificial larval substrate. This study opens the way to the understanding of the allozyme polymorphism at the molecular level and of the role of selection in the evolution and maintenance of this polymorphism under natural and laboratory conditions.

We have used the available information from the *Adh* genes of *C. capitata* in an attempt to reconstruct the phylogenetic history of the *Adh* genes in the two *tephritid* species. The comparison of proteins suggests that the two *Adh1* and the two *Adh2* genes form two phylogenetic clusters. This type of grouping was produced by both methods of tree construction attempted and has good bootstrap support for the cluster of the two *Adh1* genes. The fact that the ADH2 polypeptides are longer by one amino acid also supports this grouping. But comparison of the full cDNA lengths produced a different result. The fact that the cDNA analysis failed to produce the same tree as the amino acid sequences is not surprising in itself, yet the fact that the latter analysis has joined the two conspecific *Adh* genes in separate pairs is remarkable. Again, the cDNA tree topography is the same

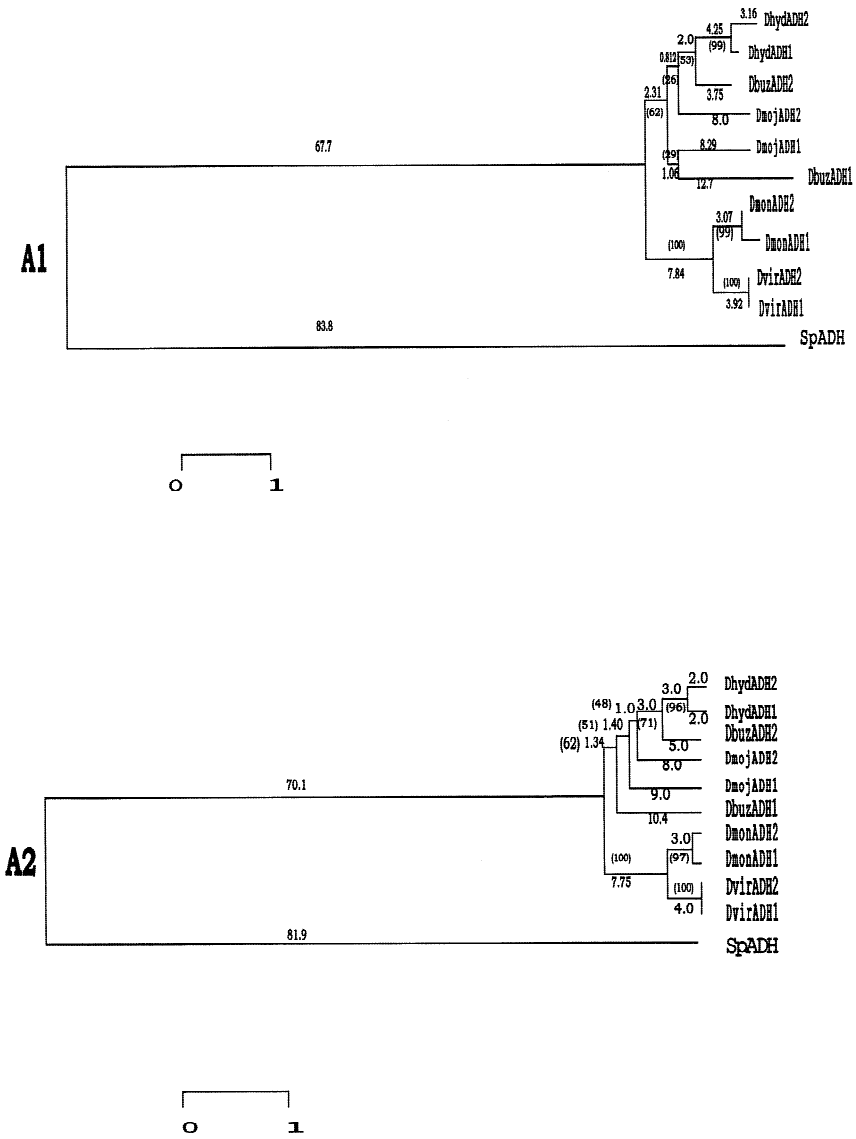

Fig. 4. Phylogenetic trees of 10 *Adh* genes from five species of *Drosophila: D. buzzatii* (DbuzAdh1 and DbuzAdh2), *D. hydei* (DhydAdh1 and DhydAdh2), *D. mojavensis* (DmojAdh1 and DmojAdh2), *D. montana* (DmonAdh1 and DmonAdh2), and *D. virilis* (DvirAdh1 and DvirAdh2). The *Adh* gene of *Sarcophaga penegrina* was used as outgroup. Numbers give branch lengths in number of amino acid **(A)** or nucleotide substitutions **(B)** and numbers in parentheses are bootstrap confidence values. **A:** Amino acid trees, **B:** Nucleotide trees. A1 and B1 are neighbor-joining and A2 and B2 are UPGMA trees. The bar below each tree indicates the distance measure.

whether the neighbor-joining or the UPGMA method is used. Both these trees are based on molecular distance. The maximum parsimony method which is based on character-state differences produced the same result (data not shown). As mentioned earlier, a regular pattern of this type has a low probability to arise by change alone. The difference between the trees of Fig. 4A and Fig. 4B is important, because each of these figures supports a different phylogenetic history of the genes, one of a unique duplication within the family *Tephritidae* (Fig. 4A), or one of separate duplications within each species (Fig. 4B). The question is not merely about the history of an important enzyme but also about the selective forces that have guided its evolution.

At present there is no further information from the *Adh* genes of *Tephritidae* that we may use to resolve this disagreement between amino acid and nucleotide sequences. On the other hand the *Adh* genes of *Drosophila* have been extensively studied either from the standpoint of their evolution (Sullivan et al. 1989) or as tools for reconstructing the phylogeny and estimating divergence times of drosophilid species (Russo et al. 1995). We have used here information from only those species of *Drosophila* for which two functional *Adh* genes, also named *Adh1* and *Adh2*, were known to exist. We have used this information in a way similar to that we have applied to the four tephritid *Adh* genes. The five species that we have examined (Fig. 4) belong to the *Drosophila* subgenus of the family *Drosophilidae.* Two of these (*D. virilis* and *D. montana*) belong to the *virilis* group and the other three to the *repleta* group. Of the latter species, *D. hydei* belongs to the *hydei* subgroup and *D. mojavensis* and *D. buzzatii* to the *mulleri* subgroup. The amino acid sequence trees produced a picture in which the *Adh* genes clustered according to species for three of the five species but in a mixed way for the two species of the *mulleri* subgroup. In contrast, cDNA nucleotide sequences produced very clear highly supported trees (Fig. 4A) that

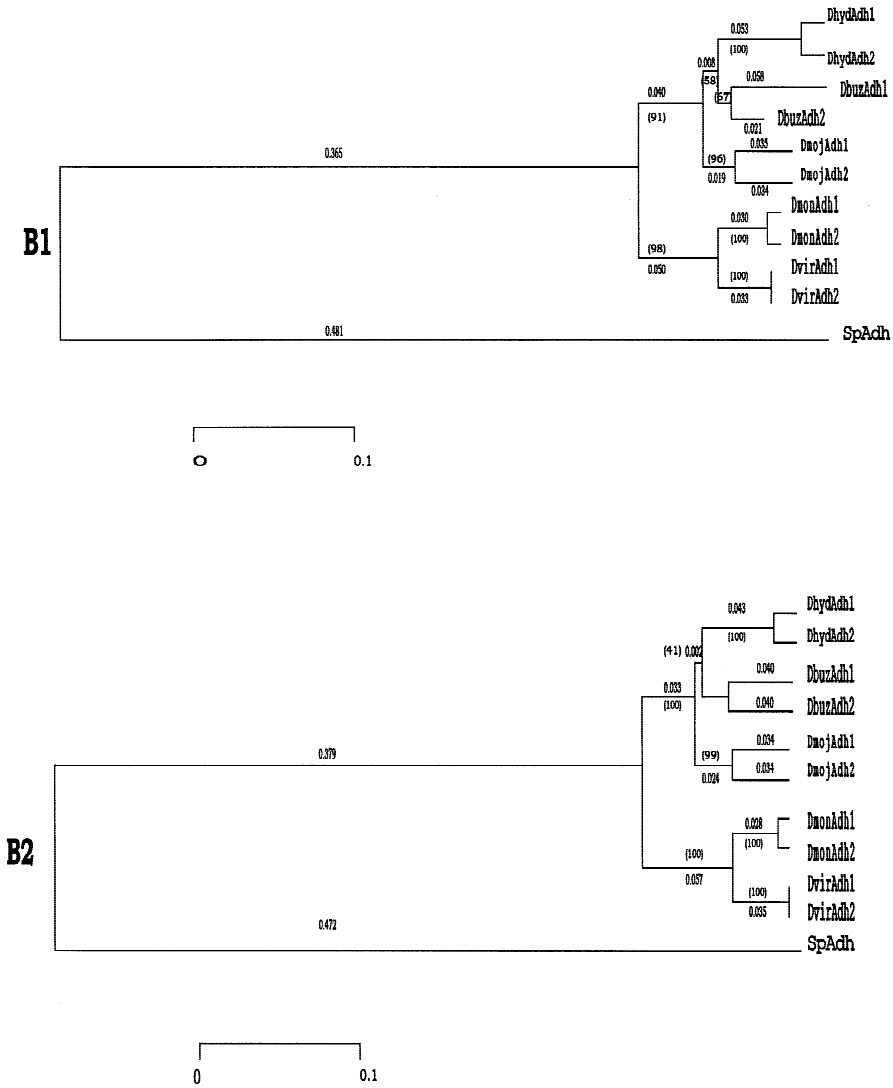

Fig. 4. Continued.

joined together the conspecific sequences for all five species (Fig. 4B). In addition, these trees successfully separated the species according to species groups and subgroups. No argument about substitution saturation may explain a pattern of such consistency. Examination of introns of the 10 *Adh* genes of Fig. 4 also supports a grouping according to species than type. In each pair of conspecific genes, the first intron is of the same size and the second intron is also of the same size, but sizes of homologous introns vary among *Adh* genes of same type from different species. When intron 1 and intron 2 of each locus are amalgamated to produce an intronic sequence, the 10 resulting sequences can be used to produce a new set of phylogenetic trees. These trees, based either on neighbor-joining or UPGMA methods, are similar in topology to those shown in Fig. 4B based on exonic sequences, i.e., they cluster the *Adh* genes according to species rather than type, with the exception of the *D. mojavensis* and *D. buzzatii* sequences, which cluster in a mixed way (data not shown). The conclusion must be either that in the species of the subgenus *Drosophila* the two *Adh* genes arose by separate duplication events that occurred after the emergence of each species or that the two genes have resulted from a prespeciation duplication event and have subsequently been subjected to concerted evolution, possibly through gene conversion.

The issue of *Adh* duplication in the genus *Drosophila* remains unsolved (Ashburner 1998). In the subgenera *Sophophora* and *Scaptodrosophila* there exist a functional gene and an apparently nonfunctional *Adh*-related (*Adhr*) gene, which is tightly linked to the functional gene. The situation is more complex in the subgenus *Drosophila* where species of the *repleta* group have as many as three genes. Of these one is a nonfunctional pseudogene (Sullivan et al. 1994), even though in some species, in particular *D. mettleri,* this would be functional (Begun 1997). Russo et al. (1995) have attempted to time the events of *Adh* duplication in species of the *repleta* group and suggested that a single duplication that happened 6–11 Mya could explain the *Adh* genes in the *D. mulleri* subgroup, except in *D. hydei* where this event cannot be older than 4 Mya. As a result these authors

have concluded that either there have been multiple duplication events within the *D. repleta* group or that gene conversion may have increased the similarity of conspecific genes in some species. The latter hypothesis was also entertained by Menotti-Reymond et al. (1991), who could not find compelling evidence for it. This is also supported by our observation that the divergence of intraspecific sequences vary considerably among species from nil in *D. virilis* to 0.08 in *D. buzzatii* (Fig. 4B). On face value this could be incompatible with a single prespeciation duplication event, since in this case homogenization through concerted evolution would have been the same for every two *Adh* genes drawn from the same species. Unfortunately, we know nothing about the probability distribution of the rate of this homogenization, so no statistical support can be provided for this observation. But one can observe that whatever the range of its distribution, stochastic homogenization could be hardly expected to cluster the species according to taxonomic subgroups and groups (Fig. 4B). On the other hand, if duplications occurred within species, the degree of divergence would be correlated to the time of duplication, which need not be the same in all species. Thus, present evidence is more supportive of the hypothesis of postspeciation *Adh* duplication than concerted evolution in *Drosophila*. By extension this provides additional support for our suggesting that the *Adh* genes of *tephritids* may have a similar evolutionary history. The issue cannot be settled without the accumulation of more data.

Acknowledgments. We thank Professor N.K. Moschonas and E. Ladoukakis for providing many helpful comments on the manuscript and M. Zervou for help with the design of the figures.

References

Ashburner M (1998) Speculations on the subject of alcohol dehydrogenase and its properties in *Drosophila* and other flies. BioEssays 20:949–954

Batterham P, Chambers GK, Starmer WT, Sullivan DT (1984) Origin and expression of an alcohol dehydrogenase gene duplication in the genus *Drosophila.* Evolution 38:644–657

Begun DJ (1997) Origin and evolution of a new gene descended from alcohol dehydrogenase in *Drosophila.* Genetics 145:375–382

Benos P, Tavernarakis N, Brogna S, Thireos G, Savakis C (2000) Acquisition of a potential marker for insect transformation: isolation of a novel alcohol dehydrogenase gene from *Bactrocera oleae* by functional complementation in yeast. Mol Gen Genet 263:90–95

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

Bodmer M, Ashburner M (1984) Conservation and change in the DNA sequences coding for alcohol dehydrogenase in sibling species of *Drosophila.* Nature 300:425–430

Chen Z, Jiang JC, Lin ZG, Lee WR, Chang SH (1993) Site-specific mutagenesis of *Drosophila* alcohol dehydrogenase: evidence for involvement of tyrosine-152 and lysine-156 in catalysis. Biochemistry 32:3342–3346

Cols N, Marfany G, Atrian S, Gonzalez-Duarte R (1993) Effect of site-directed mutagenesis of conserved positions of *Drosophila* alcohol dehydrogenase. FEBS Lett 319:90–94

Cosmides N, Loukas M, Zouros E (1997) Differences in fitness components among alcohol dehydrogenase genotypes of the olive fruit fly (Diptera: Tephritidae) under artificial rearing. Ann Entomol Soc Am 90(3):363–371

Cosmidis N, Loukas M, Zouros E (1999) Rarer need not be better if commoner is worse: frequency-dependent selection for developmental time at the alcohol dehydrogenase locus of the olive fruit fly, *Bactrocera oleae.* Evolution 53(2):518–526

David JR (1988) Ethanol adaptation and alcohol dehydrogenase polymorphism in *Drosophila:* from phenotypic functions to genetic structures. In: de Jong G (ed) Population genetics and evolution. Berlin: Springer, pp 163–172

Felsenstein J (1985) Confidence limits on phylogenies: an approach using the bootstrap. Evolution 39:783–791

Gasperi G, Baruffi L, Malacrida A, Robinson AS (1992) A biochemical genetic study of alcohol dehydrogenase isozymes of the medfly *Ceratitis capitata.* Wied Biochem Genet 30:289–304

Gasperi G, Kafetzopoulos D, Christodoulidou A, Bouriotis V, Savakis C (1994) Isolation and partial characterization of two alcohol dehydrogenase isozymes from the medfly *Ceratitis capitata.* Insect Biochem Molec Biol 24(1):87–94

Gibson J, Oakeshott J (1982) Tests of the adaptive significance of the alcohol dehydrogenase polymorphism in *Drosophila melanogaster:* path, pitfalls and prospects. In: Barker JSF, Starmer WT (eds) Ecological genetics and evolution. New York: Academic Press, pp 291–306

Holmes DS, Bonner J (1973) Preparation, molecular weight, base composition and secondary structure of giant nuclear ribonucleic acid. Biochemistry 12:2330–2338

Horio T, Kubo T, Natori S (1996) Purification and cDNA cloning of the alcohol dehydrogenase of the flesh fly *Sarcophaga penegrina.* A structural relationship between alcohol dehydrogenase and a 25-KDa protein. Eur J Biochem 237(3):698–703

Jornvall H, Persson M, Jeffery J (1981) Alcohol and polyol dehydrogenase are both divided into two protein types, and structural properties cross-relate the different enzyme activities within each type. Proc Natl Acad Sci USA 78:4226–4230

Jornvall H, Persson B, Krook M, Atrian S, Gonzalez-Duarte R (1995) Short-chain dehydrogenases/reductases (SDR). Biochemistry 34: 6003–6013

Juan E, Gonzalez-Duarte R (1980) Purification and enzyme stability of alcohol dehydrogenase from *Drosophila simulans, Drosophila virilis* and *Drosophila melanogaster.* Biochem J 189:105–110

Juan E, Gonzalez-Duarte R (1981) Determination of some biochemical and structural features of alcohol dehydrogenase from *Drosophila simulans* and *Drosophila virilis.* Biochem J 195:61–69

Kimura M (1980) A simple method for estimating evolutionary rate of base substitutions through comparative studies of nucleotide sequences. J Mol Evol 16:111–120

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

Kreitman M, Hudson R (1991) Inferring the evolutionary histories of the *Adh* and *Adh-dup* loci in *Drosophila melanogaster* from patterns of polymorphism and divergence. Genetics 127:565–582

Kumar S, Tamura N, Nei M (1993) MEGA: molecular evolutionary genetics analysis. University Park, PA: The Pennsylvania State University

Li W-H (1997) Molecular evolution. Sinauer Associates, Sunderland, MA

Malacrida AR, Gasperi G, Zacharopoulou A, Torti C, Riva-Francos E, Milani R (1992) Evidence for a genetic duplication involving alcohol dehydrogenase genes in *Ceratitis capitata.* Biochem Genet 30:35–48

Mazi VE, Cosmidis N, Clonis YD, Loukas M (1998a) Purification of alcohol dehydrogenase from four genotypes of the olive fruit fly *Bactrocera* (*Dacus*) *oleae.* Biotech Prog 36:259–269

Mazi V, Cosmidis N, Loukas M, Clonis Y, Zouros E (1998b) Biochemical differences between products of the ADH locus in olive fruit fly (*Bactrocera oleae*). Biochem Genet 36:259–269

Menotti-Raymond M, Starmer WT, Sullivan DT (1991) Characterization of the structure and evolution of the *Adh* region of *Drosophila hydei*. Genetics 127:355–366

Moxon LN, Holmes RS, Parsons PA, Irving MG, Doddrell DM (1985) Purification and molecular properties of alcohol dehydrogenase from *Drosophila melanogaster:* evidence from NMR and kinetic studies for function as an aldehyde dehydrogenase. Comp Biochem Physiol 80B:525–535

Nurminsky DI, Moriyama EN, Lozovskaya ER, Hartl DL (1996) Molecular phylogeny and genome evolution in the *Drosophila virilis* species group: duplications of the alcohol dehydrogenase gene. Mol Biol Evol 13:132–149

Persson B, Krook M, Jornvall JH (1991) Characteristics of short-chain alcohol dehydrogenases and related enzymes. Eur J Biochem 200: 537–543

Robinson AS, MacLeod EG (1993) Genetic sexing in *Drosophila melanogaster* using the alcohol dehydrogenase locus and a Y-linked translocation. Theor Appl Genet 59:23–24

Robinson AS, Riva ME, Zapater M (1986) Genetic sexing in the Mediterranean fruit fly, *Ceratitis capitata,* using the *Adh* locus. Theor Appl Genet 52:455–457

Rozas J, Rozas R (1999) DnaSp version 3: an integrated program for molecular population genetics and molecular evolution analysis. Bioinf Appl Note 15:174–175

Russo CAM, Takezaki N, Nei M (1995) Molecular phylogeny and divergence times of *Drosophila* species. Mol Biol Evol 12:391–404

Sambrook J, Fritsch EF, Maniatis T (1989) Molecular cloning: a laboratory manual, 2nd ed. Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press

Savakis C, Ashburner M (1985) A simple gene with a complex pattern of transcription: the alcohol dehydrogenase gene of *Drosophila melanogaster*. Cold Spring Harb Symp Quant Biol 50:505–514

Sofer W, Ursprung H (1968) *Drosophila* alcohol dehydrogenase. J Biol Chem 243:3110–3115

Sullivan DT, Atkinson PW, Starmer WT (1989) Molecular evolution of the alcool dehydrogenase genes in the genus *Drosophila.* Evol Biol 24:107–147

Sullivan DT, Starmer WT, Curtiss SW, Menotti-Raymond M, Yum Y (1994) Unusual molecular evolution of an *Adh* pseudogene in *Drosophila.* Mol Biol Evol 10:605–618

Thatcher DR (1980) The complete amino acid sequence of three alcohol dehydrogenase alloenzymes (Adh^{N-11}, Adh^{S} and Adh^{UF}) from the fruitfly *Drosophila melanogaster*. Biochem J 187:875–883

Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG (1997) The Clustal X Windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res 24:4876–4882

Van Delden W (1982) The alcohol dehydrogenase polymorphism in *Drosophila melanogaster*. Selection at an enzyme locus. Evol Biol 15:187–222

Zouros E, Loukas M, Economopoulos A, Mazomenos B (1982) Selection at the alcohol dehydrogenase locus of the olive fruit fly *Dacus oleae* under artificial rearing. Heredity 48:169–185