Sequence Variation of Alcohol Dehydrogenase (*Adh*) Paralogs in Cactophilic Drosophila

Luciano M. Matzkin[1] **and Walter F. Eanes**

Department of Ecology and Evolution, State University of New York, Stony Brook, New York 11794-5245

Manuscript received August 9, 2002
Accepted for publication October 4, 2002

ABSTRACT

This study focuses on the population genetics of alcohol dehydrogenase (*Adh*) in cactophilic Drosophila. *Drosophila mojavensis* and *D. arizonae* utilize cactus hosts, and each host contains a characteristic mixture of alcohol compounds. In these Drosophila species there are two functional *Adh* loci, an adult form (*Adh-2*) and a larval and ovarian form (*Adh-1*). Overall, the greater level of variation segregating in *D. arizonae* than in *D. mojavensis* suggests a larger population size for *D. arizonae*. There are markedly different patterns of variation between the paralogs across both species. A 16-bp intron haplotype segregates in both species at *Adh-2*, apparently the product of an ancient gene conversion event between the paralogs, which suggests that there is selection for the maintenance of the intron structure possibly for the maintenance of pre-mRNA structure. We observe a pattern of variation consistent with adaptive protein evolution in the *D. mojavensis* lineage at *Adh-1*, suggesting that the cactus host shift that occurred in the divergence of *D. mojavensis* from *D. arizonae* had an effect on the evolution of the larval expressed paralog. Contrary to previous work we estimate a recent time for both the divergence of *D. mojavensis* and *D. arizonae* (2.4 ± 0.7 MY) and the age of the gene duplication (3.95 ± 0.45 MY).

THE strength of selection at a particular locus can be affected by changes in the environment experienced by its product (Dykhuizen and Hartl 1980; Hartl *et al.* 1985; Dykhuizen *et al.* 1987). Interspecific comparisons can identify differences in the evolutionary history of loci between species, but may not identify the environmental factors that changed the shape of the adaptive landscape. It is often possible to determine changes in the external environment experienced by species. An approach to this problem is to compare the evolutionary trajectories of recently duplicated loci, which have complete or partial nonoverlapping modes of expression.

Gene duplications underlie the diversification of genes and the origination of novel gene functions (Ohno 1970). Most models of gene evolution via duplication offer two possible fates for duplicated loci, either the creation of a gene with novel function or the formation of a pseudogene (Ohno 1970; Ohta 1988a, 1989). Other models have offered a third fate for recently duplicated loci. If the original locus had several modes of expression controlled by different regulatory sequences, after the duplication, mutations may occur in one or more of these regulatory sequences resulting in two loci having at least partially nonoverlapping modes of expression (Hughes 1994; Force *et al.* 1999). At this point adaptive evolution can potentially occur since each locus will be released from the constraints of being expressed in all regions, effectively being placed in a new environment.

In Drosophila and other Diptera, there have been several independent duplications of the alcohol dehydrogenase (*Adh*) locus (Gasperi *et al.* 1992; Russo *et al.* 1995; Goulielmos *et al.* 2001). One such set of duplications in Drosophila occurs in the *mulleri* species complex. Within the *mulleri* species complex three *Adh* loci have been described (Atkinson *et al.* 1988). An early duplication created a locus of abnormal function positioned in the 5′ region of the *Adh* gene cluster. The exact function of this locus (*Adh* Finnegan) is unknown (Begun 1997). A later duplication resulted in *Adh-2* and *Adh-1*, positioned ~3 kb apart. ADH-1 activity is observed from egg until the 5-day-old larva stage (Batterham *et al.* 1983). Just prior to pupal eclosion, both ADH-1 and ADH-2 activities are observed. In adults, only ADH-2 is expressed, except in females, where ADH-1 activity is limited to the ovaries.

The cactophilic *mojavensis* species cluster is within a set of 24 species that compose the *mulleri* species complex (Wasserman 1982). The *mojavensis* cluster is composed of *D. navojoa*, *D. arizonae*, and *D. mojavensis*, from basal to most derived (Heed 1982; Ruiz *et al.* 1990). Cactus hosts tend not to be shared across species: agria (*Stenocereus gummosus*) and organpipe (*S. thurberi*) are utilized by *D. mojavensis*, cina (*S. alamosensis*) and *Opuntia* spp. by *D. arizonae*, and *Opuntia wilcoxii* by *D. navojoa*

Sequence data from this article have been deposited with the EMBL/GenBank Data Libraries under accession nos. AY154827–AY154876, AY156524, and AY156525.

[1]*Corresponding author:* Department of Ecology and Evolution, State University of New York, Stony Brook, NY 11794-5245.
E-mail: lmatzkin@life.bio.sunysb.edu

(Fellows and Heed 1972; Ruiz and Heed 1988). Cactus host use does, however, vary with geographic location. Most importantly, the alcohol composition of the cactus rots varies with species (Heed 1978; Vacek 1979; Fogleman 1982; Kircher 1982).

The *D. melanogaster* ADH allozyme polymorphism has become a textbook example of selection on a simple gene. The Fast/Slow (F/S) allozyme polymorphism shows parallel latitudinal clines in several continents, where the Fast allele is at its highest frequency in higher latitudes (Oakeshott *et al.* 1982). Population studies at the *Adh* sequence level have proposed that the pattern of variation observed at *Adh* is indicative of balancing selection (Kreitman 1983). Centered around the replacement substitution producing the F/S polymorphism is an excess of silent polymorphisms that is consistent with models of balancing selection (Hudson and Kaplan 1988; Hudson 1990). Berry and Kreitman (1993) observed that an indel (∇1) exhibited a stronger clinal pattern than that of the replacement polymorphism producing the F/S change. The ∇1 indel, located in the first intron of the larval transcript, has a significant effect on ADH activity (Laurie and Stam 1994). Also, there is evidence of selection for the maintenance of pre-mRNA structure in Drosophila *Adh* (Kirby *et al.* 1995), which could have potential effects on ADH activity. Hence the actual target of the proposed selection in *D. melanogaster* has not been unequivocally determined. In addition to *D. melanogaster*, population level sequence data have been collected in only a few species: *D. pseudoobscura* (Prakash 1977; Schaeffer and Miller 1992a,b), *D. simulans* (Kliman *et al.* 2000), *D. willistoni* (Griffith and Powell 1997), and *D. americana americana* (McAllister and Charlesworth 1999).

Alcohol dehydrogenase has been proposed to play an important role in the adaptation to alcohol environments (Mercot *et al.* 1994). In Drosophila, a physiological and behavioral response to ethanol can be observed at both the larval and the adult stage (Starmer *et al.* 1977; Gelfand and McDonald 1980; Geer *et al.* 1990). There is a large amount of variation in hosts across Drosophila species, from flowers to crab excrement, and each particular host has its specific chemical makeup (Powell 1997). Given what is known about the alcohol-dependent ecology of cactophilic Drosophila, the duplication and nonoverlapping modes of expression of *Adh* in *D. mojavensis* and *D. arizonae*, and the possible adaptive significance of *Adh* in Drosophila, it is of great value to examine the variation at *Adh* in these species. In this study we have focused on the allozyme and nucleotide variation of *Adh* paralogs of two cactophilic Drosophila sister species (*D. mojavensis* and *D. arizonae*) that utilize different cactus hosts.

MATERIALS AND METHODS

Isofemale line collections: The *D. mojavensis* (MJBC) lines were collected in February 2001 from a geographic location 30 km south of La Paz, Baja California Sur, Mexico. Flies were aspirated from rotting sections of agria cactus and briefly placed in collecting tubes. Isofemale lines were set up in the field by placing a gravid female in an 8-dram vial containing standard banana-molasses medium. A similar technique was used in the collection of *D. arizonae* isofemale lines. *D. arizonae* isofemale lines were collected by Therese Markow in Tucson, Arizona (ARTU) and Guaymas, Sonora, Mexico (AR00) in 2000. The *D. navojoa* line is from the University of Arizona Drosophila species stock center (no. 15081-1374.0). While in the lab all lines were maintained in 8-dram vials containing standard banana-molasses media sprinkled with a few granules of live yeast. Lines were stored in a 25° incubator with a 14:10 light:dark cycle. Lines were transferred into new food vials every 3–4 weeks.

Allozyme survey: Starch gel electrophoresis was used to estimate the level of ADH allozyme variation. A modified version of the Batterham *et al.* (1983) protocol was used. Three one-fly samples per isofemale line were homogenized in 15 μl of Tris-boric acid buffer (41 mm Tris; 6 mm boric acid, pH 8.8) and transferred to filter paper. Samples were run at 4° through a 12% starch gel at 30 V/cm for 5 hr. A 1% agar overlay stain (100 mm Tris-HCl, pH 8.8; 260 mm 2-propanol; 0.75 mm NAD+; 0.61 mm methylthiazoletetrazolium; 0.03 mm phenazine methosulfate) was used to visualize ADH.

Sampling: According to the allozyme variation observed in the *D. mojavensis* population, a constructed random sample (Hudson *et al.* 1994) of 13 isofemale lines was PCR amplified and sequenced for both *Adh-2* and *Adh-1*. The *D. arizonae* isofemale lines sampled varied slightly between *Adh-2* (7 ARTU and 4 AR00 lines) and *Adh-1* (6 ARTU and 7 AR00 lines). Overall there were 6 ARTU and 4 AR00 lines in common between the two samples that were sequenced for both paralogs. Additionally, one *D. navojoa* line was used to polarize the *D. mojavensis* and *D. arizonae Adh* sequence data for the purpose of performing a lineage-specific McDonald-Kreitman test.

PCR amplification and sequencing: PCR amplification of CTAB genomic DNA preps (Winnepenninckx *et al.* 1993) was done in a thermal cycler (MJ Research, Waltham, MA) and an Air-Thermo-Cycler (Idaho Technologies, Idaho Falls, ID) using GIBCO BRL (Carlsbad, CA) Taq DNA polymerase. Locus-specific primers were designed from the published *D. mojavensis Adh-2/Adh-1* sequence (Atkinson *et al.* 1988). Primers designed from the *D. mojavensis Adh-1* sequence successfully amplified sequence from *D. arizonae* and *D. navojoa*. However, only a partial sequence (~250 bp) was obtained for *Adh-2* in *D. navojoa* using *D. mojavensis* primers. We used an inverse PCR technique explained by Triglia *et al.* (1988) to obtain the complete *D. navojoa Adh-2* sequence.

To examine the association between polymorphic sites between and within paralogs, we retrieved full haplotypes. Due to the fact that the *Adh* paralogs are autosomal loci, it was necessary to clone the *Adh-2/Adh-1* gene cluster. This technique was used only for the *D. mojavensis* sequences. A 5.5-kb fragment containing *Adh-2* and *Adh-1* was amplified using the Expand Long Template PCR System (Roche, Mannheim, Germany). Fragments were then purified using QIAquick columns (QIAGEN, Valencia, CA) and cloned into a pCR 2.1 vector (TA cloning kit; Invitrogen, Carlsbad, CA). Colonies with inserts were picked and disrupted in 500 μl of ddH_2O. This solution was used as a template to PCR amplify *Adh-2* and *Adh-1* fragments individually. All PCR fragments were cleaned using the Prep-A-Gene kit (Bio-Rad, Hercules, CA) prior to sequencing.

Fragments were amplified and sequenced from three colonies per *D. mojavensis* individual to obtain the linkage phase and to correct for errors. The sequencing reactions were performed using the ABI Prism BigDye cycle sequencing kit v2.0, and reactions were run in an ABI 3100 genetic analyzer (Ap-

plied Biosystems, Foster City, CA). The sequencing of the *D. arizonae* isofemale lines and the stock center *D. navojoa* line was done manually using the Sequenase kit v2.0 (United States Biochemical, Cleveland) and [^{35}S]dATP (Amersham, Buckinghamshire, England). All sequences are stored under GenBank accession nos. AY154827–AY154876, AY156524, and AY156525.

Data analysis: Descriptive and statistical analysis of the sequence data was produced using SITES (Hey and Wakeley 1997) and DnaSP version 3.53 (Rozas and Rozas 1999). The neighbor-joining gene tree using third base positions and bootstrapping was created using MEGA version 2.1 (Kumar *et al.* 2001).

RESULTS

Allozyme variation: Three individuals from each of the 41 *D. mojavensis* isofemale lines were used in the allozyme survey. Consistent with previous work (Batterham *et al.* 1983) the ADH-2 enzyme had the strongest cathodal mobility. For ADH-2, 2 lines were homozygous for the Slow allele, 32 lines were homozygous for the Fast allele, and both alleles segregated in the 7 remaining lines. The weighted frequency of the ADH-2 Fast allele was 0.90; no allozyme variation was observed at ADH-1, which is in agreement with previous surveys (Heed 1978).

Sequence variation at alcohol dehydrogenase-2: The gene structure in *D. mojavensis* and *D. arizonae Adh-2* was similar to that observed in *D. melanogaster Adh* (Kreitman 1983). In both species, the locus consists of a total of 880 bp, with two small introns of 55 bp (position 94–148) and 60 bp (position 554–613). *D. mojavensis* and *D. arizonae Adh* possess 2 fewer amino acid residues compared to the *D. melanogaster* locus. The inferred protein consists of 254 amino acid residues, which is the same as that observed in *D. lebanonensis*, the species for which a refined crystal structure of ADH is known (Benach *et al.* 1998, 1999).

For *D. mojavensis* we sequenced a total of 12 *Adh-2-F* and 2 *Adh-2-S* alleles. The constructed random sample (CRS) consisted of 12 Fast alleles and 1 Slow allele; the variable sites are shown in Figure 1. All parameter estimates and tests were performed using this data set. The second Slow allele was omitted from the constructed random sample because there was evidence of a substantial gene conversion event between *Adh-1* and *Adh-2* (data not shown). There are a total of six replacement polymorphisms, where five separate the Fast and Slow alleles. Two of these replacement polymorphisms produce a charge change, which could be responsible for the Fast/Slow allozyme polymorphism: an arginine-to-serine change at position 84 and a histidine-to-tyrosine change at position 347. In an additional survey (L. M. Matzkin and W. F. Eanes, unpublished data) of sequence variation across *D. mojavensis*, we observed that only the replacement polymorphism at position 84 is unique to all Slow alleles. In the coding region there are 19 silent polymorphisms, plus a substantial level of variation segregating in the introns. Twenty-four segregating sites plus two small deletions are in the introns, and 20 segregating sites occur in the first intron alone.

The *Adh-2* gene structure in *D. arizonae* was identical to *D. mojavensis* with the exception of a single-base deletion in both introns (Figure 1). Overall, we observed a greater number of segregating sites in *D. arizonae*. There were a total of 33 silent and 6 replacement polymorphisms, none of which produced a charge change (Figure 1). In 13 out of the 33 silent sites we found one or more individuals segregating for both bases, *i.e.*, heterozygous sites. In such cases the alternative state was never a singleton. Similar to *D. mojavensis* there was a substantial amount of variation segregating in the introns, 38 sites, of which 30 occur in the first intron. Of the sites segregating in the first intron, 16 are shared with *D. mojavensis* (Figure 1). We found that the shared polymorphisms are due to gene conversion between the *Adh-2* and *Adh-1* intron 1, because the sequence of one of the *Adh-2* intron haplotypes is identical to the *Adh-1* intron.

The silent diversity (θ) in *D. arizonae* was greater than that observed in *D. mojavensis* (Table 1). Given the nature of the sequence data in *D. arizonae*, an estimate of pairwise differences (π) can be performed only if we assume a random distribution of the variation across the alleles. A way to check for the validity of this assumption is by estimating pairwise differences from the seven *D. arizonae* individuals that contained zero or one heterozygous site. The estimated π from these seven individuals ($\pi = 0.05928$) was similar to the overall estimate of π (Table 1).

Sequence variation at alcohol dehydrogenase-1: The gene structure observed at *Adh-1* is similar to that of *Adh-2* with the exception of a few small indels in the introns (Figure 2). As with ADH-2, 254 amino acid residues were in the inferred ADH-1 protein. For the *D. mojavensis Adh-1* data set, we used the same population sample as in *Adh-2*. In *D. mojavensis* there were a total of 12 segregating sites; 10 occurred in the exons and none of those were replacement changes. There was very little variation in the introns. In *D. arizonae*, a total of 35 segregating sites were observed in the coding region, 31 silent and 4 replacement. Line ARTU 10 was not present in the *Adh-1* analysis, unlike the *D. arizonae Adh-2* data set. Additionally, lines AR00 2, 4, and 5 were included in the *Adh-1* set but excluded for *Adh-2*. Due to the 16 changes that occur in intron 1 and the heterozygous nature of the data, no intron 1 data were collected for lines AR00 4, 6, and 14, because of difficulty in interpreting the autoradiographs. Once again, the overall level of variation in *D. arizonae* was greater than that observed in *D. mojavensis* (Table 1).

Interspecific and interparalog divergence: Two different estimates of *D. mojavensis/D. arizonae* silent divergence were calculated, one for each paralog. Since nucleotide divergence K_s is based on pairwise differences (Nei 1987), for the comparison between species within paralog, we used only one *D. arizonae* line (ARTU 34).

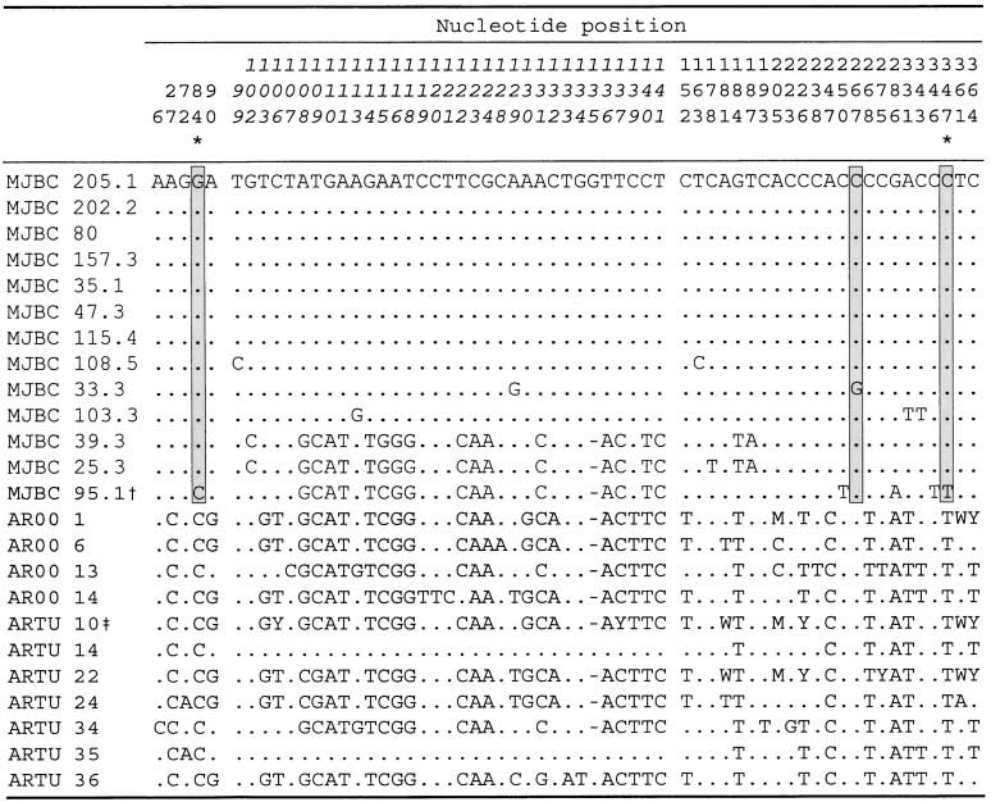

FIGURE 1.—A list of inter- and intraspecific variation at *Adh-2* in *D. mojavensis* and *D. arizonae*. The italicized nucleotide positions are sites located in the introns. Dagger indicates the *D. mojavensis* ADH-2 Slow allele. Shaded boxes indicate a replacement polymorphism within a species. Asterisk indicates replacement polymorphism producing a charge change. Dashes are sequence gaps produced by an insertion/deletion. Standard nomenclature was used to identify heterozygous sites: Y (C/T), W (A/T), M (A/C), and R (A/G). Double dagger indicates that *D. arizonae* line (ARTU 10) is not present in the *Adh-1* data set.

ARTU 34 was chosen for the analysis because it contained no heterozygous sites in *Adh-1* and only one in *Adh-2*. POWELL (1997) suggested that the rate of silent substitution in Drosophila is between 1.0 and 2.7×10^{-8} site changes per year with a mean of 1.9×10^{-8} site changes per year; we used the mean rate of silent substitution to calculate divergence times. Using the locus divergence between *Adh-1* ($K_s = 0.065$) and *Adh-2* ($K_s =$

TABLE 1

Estimates of synonymous and nonsynonymous variation in *Adh-2* and *Adh-1* in *D. mojavensis* and *D. arizonae*

	Adh-2				*Adh-1*			
	θ_s	π_s	θ_n	π_n	θ_s	π_s	θ_n	π_n
D. mojavensis (n = 13)	0.0324	0.0206	0.0034	0.0016	0.0174	0.0130	0	0
D. arizonae (n = 22/26[a])	0.0479	0.0571	0.0029	0.0046	0.0438	0.0318	0.0018	0.0014

Diversity (θ) and pairwise (π) differences were calculated from synonymous (s) and nonsynonymous (n) sites.

[a] Sample size of *Adh-2* and *Adh-1*, respectively. The value of n for *D. arizonae* is double the number of sequenced individuals because the individuals were heterozygous (see text). These values of n were used to standardize the estimates of variation.

0.117) we obtained an estimate of time of divergence between *D. mojavensis* and *D. arizonae* of 1.7 and 3.1 million years (MY; for *Adh-1* and *Adh-2*, respectively). Additionally, we calculated the silent divergence between *D. navojoa* and *D. mojavensis* (0.158 for *Adh-1* and 0.133 for *Adh-2*) and *D. arizonae* (0.177 for *Adh-1* and 0.155 for *Adh-2*). We obtained similar estimates for the divergence time between *D. navojoa* and *D. mojavensis* (3.5–4.1 MY) and *D. navojoa* and *D. arizonae* (4.1–4.6 MY). This method of estimating divergence ignores intraspecific variation; therefore it estimates only the divergence times of the genes, not the species (see DISCUSSION).

We also estimated the age of the duplication that gave rise to *Adh-1* and *Adh-2*. The neighbor-joining tree in Figure 3 should be analogous to the estimated K_s values since the tree was constructed using third position sites only. As shown in Figure 3, *D. mojavensis* and *D. arizonae* share the same *Adh* duplication. Common ancestry of the *Adh* duplication is not evident for *D. navojoa*, but this may be largely due to the sequence homogenizing effect of gene conversion. The silent divergence between the *D. mojavensis* paralogs (0.133) resembles that of the *D. arizonae* paralogs (0.165). Therefore we can date the duplication event to 3.5–4.4 MYA, close to the time of the *D. navojoa* and *D. mojavensis/D. arizonae* divergence.

Linkage disequilibrium and recombination: The level

```
                                  Nucleotide position
                    1 1122222333333333333444445555 55555666 6666667777788888888
                  388 0 8812223011333466679933469000 4 66789001 1456892349924467777
                 6334 1 1470697104478014561909806258 7 56804382 8624841450611780127
                                                                                *
MJBC 205.1 CCAG A AGGGCGTTCCCCCGTCACGTTCTCGGGTT ACAGGCAA TCCCACACTTCGATCCGCC
MJBC 202.2 .... . ............................ ........ ...................
MJBC 80    .... . ..A.................A........ ........ ........C..........
MJBC 157.3 .... . ........T.................T.C. ........ ........C..........
MJBC 35.1  .... . ............................ ........ ........C..........
MJBC 47.3  T... . ............................ ........ ........C..........
MJBC 115.4 .... . ....................A....... ........ ........C..........
MJBC 108.5 .G.. . ............................ ........ ........C..........
MJBC 33.3  .... . ..............T.........T.C. ........ ........C..........
MJBC 103.3 .... . ............................ .....A.. ........C..........
MJBC 39.3  .... G ........................T.C. ........ ........C..........
MJBC 25.3  .... . ............................ ........ ........C.........T
MJBC 95.1† .... . ........................T.C. ........ ........C..........
AR00 1     ..GC . .T....CK....Y.C.GT.[.]...T..A.. .......T [.]......GTC..C..T[.].[.].
AR00 2     ..GC . .T....C.....T.C.GT.[.]..AT..A.. ...A...T [.]......GTCC.C..T[.].[.].
AR00 4‡    ..GC - .T....C.......C.GTR..YW..RA.. R..R...T .Y....GTCC.CR.T[.].[.].
AR00 5‡    ..GC . .T....CK..Y.YRC.GT.[.]..AT..A.. ...R...T YY....GTCC.CR.T[.].[.].
AR00 6‡    ..GC - .T.R..C....T..C.GT.[.]..AT..A.. .......T [.]..Y..YGTCY.C..T[.].[.].
AR00 13    ..GC . .T...RC.....Y.C.GT.[.]..AY..A.. ...R..R. [.]......G.CC.C..T[.].[.].
AR00 14    ..GC - WT..Y.C.......C.GT.[.]..A...A.. ...R.... [.]......G.CC.C.WT[.].[.].
ARTU 14    ..GC . WT....C..Y....C.GTR..YW..RA.Y .S.R.... [.]T....G.CC.C..TGC..
ARTU 22    ..GC . .T....C.......C.GT.[.]..WY..A.. ..TA.... [.]Y.YR.G.CY.CR.TGC..
ARTU 24    ..GC . .T.R..C.......C.GT.YY..Y..A.. ..MRR... [.]......G.CY.C..TGCY.
ARTU 34    ..GC . .T....C..T....C.GT.[.]..A..AA.. ...A.... [.]T....G.CC.C..T[.].[.].
ARTU 35    ..GC . .T....C.......C.GT.[.]..W.R.RY. ...R.... [.]....R.G.CCSC..T[.].[.].
ARTU 36    ..GC . WT....C..T....C.GT.[.]..A..AA.Y ...A.... [.]T....G.CC.C..T[.].[.].
```

FIGURE 2.—A list of inter- and intraspecific variation at *Adh-1* in *D. mojanvensis* and *D. arizonae*. The italicized nucleotide positions are sites located in the introns. Dagger indicates the *D. mojaensis* ADH-2 Slow allele. Shaded boxes indicate a replacement polymorphism within a species. Asterisk indicates replacement polymorphism producing a charge change. Dashes are sequence gaps produced by an insertion/deletion. Standard nomenclature was used to identify heterozygous sites: Y (C/T), W (A/T), K (T/G), S (C/G), and R (A/G). Double dagger indicates *D. arizonae* lines (AR00 4, 5, and 6) not present in the *Adh-2* data set.

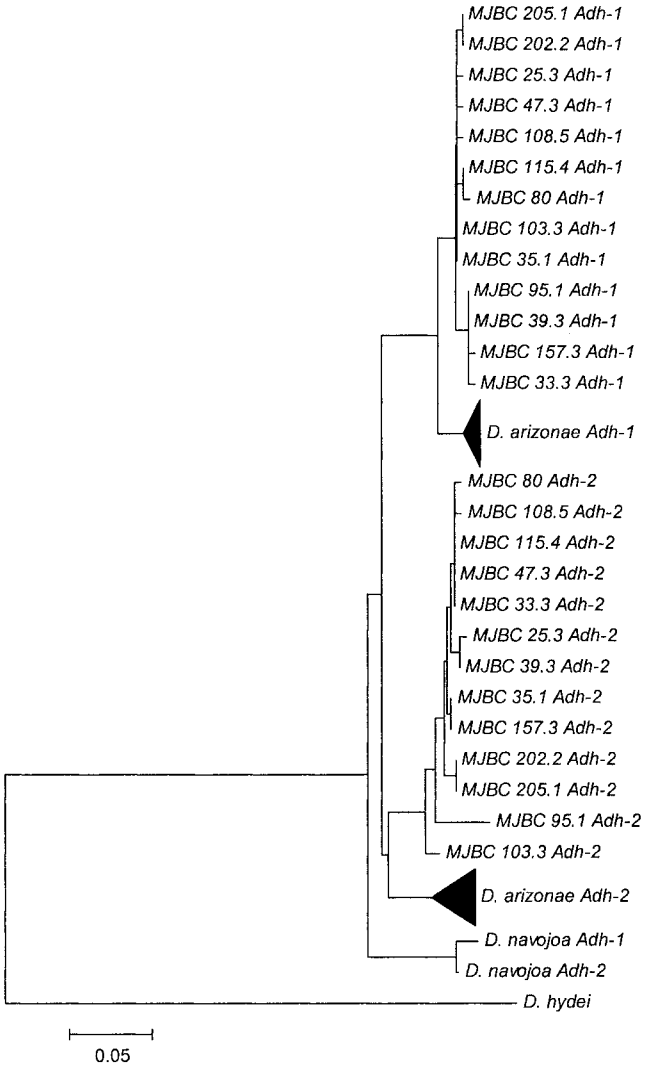

FIGURE 3.—Neighbor-joining tree of the *Adh* paralogs. Only third base positions were used to create the tree. The *D. hydei* (*Adh-2*) sequence was obtained from GenBank (accession no. X58694).

of linkage disequilibrium was determined only for *D. mojavensis,* since sequences in complete linkage phase were collected for *D. mojavensis.* Even though the paralogs were only 3 kb apart, there was little disequilibrium between the paralogs (Figure 4) as determined by the chi-square test using the Bonferroni correction for multiple comparisons. Markedly different levels of linkage disequilibrium were observed within the paralogs. There was a large amount of disequilibrium in *Adh-2,* mostly centered on the first intron. The cause of the disequilibrium in intron 1 is the segregation of two intron haplotypes that differ at 16 sites. The low-frequency (0.23) haplotype class in *D. mojavensis* was found to be segregating at a high frequency (0.82) in *D. arizonae.*

We determined the level of recombination using two methods: solving for the estimator C (HUDSON 1987) and solving for the less biased estimator γ (HEY and WAKELEY 1997). Both estimators of recombination are best presented as the ratio of the estimator over θ. This method provides an estimate of the rate of recombination (c) over that of the mutation rate (μ). Using C, the ratio is 0.173, while it was considerably greater (1.203) using the less biased estimator γ. The low c/μ ratio estimated from C was partly due to the intron 1 structure. Removing the effect of the intron 1 haplotype polymorphism greatly increased the c/μ ratios (1.072 and 2.700, estimated from C and γ, respectively).

Distribution of variation: The pockets of linkage disequilibrium produced a heterogeneous distribution of variation (Figure 5). We examined the distribution of variation across the region using MCDONALD's (1996) DNA Slider program. There is a significant nonrandom distribution of segregating sites at *Adh-2,* both in *D. mojavensis* ($G_{mean} = 4.416$, $P = 0.045$) and in *D. arizonae* ($G_{mean} = 5.080$, $P = 0.024$). This heterogeneous pattern of variation was not observed at *Adh-1,* either in *D. mojavensis* ($G_{mean} = 0.274$, $P = 0.962$) or in *D. arizonae* ($G_{mean} = 3.218$, $P = 0.096$).

Frequency and test of independence analysis: Both the TAJIMA (1989) and the FU and LI (1993) tests examine the frequency distribution of the variants. Significant negative values of the test statistic indicate a greater proportion of low-frequency variants, usually explained as being a result of an adaptive sweep or as a population expansion, while positive values indicate a greater proportion of high-frequency variants, possibly a result of balancing selection or population admixture. In *D. mojavensis Adh-2,* both Tajima's ($D_{Tajima} = -1.549$) and Fu and Li's ($D_{Fu\&Li} = -1.139$) statistic were negative but not significantly. A similar result was observed at *Adh-1* ($D_{Tajima} = -1.007$ and $D_{Fu\&Li} = -1.365$). In *D. arizonae* Tajima's statistic was not significant for *Adh-2* and *Adh-1* ($D_{Tajima} = 0.733$ and $D_{Tajima} = -1.023$, respectively). The FU and LI (1993) test could not be performed for *D. arizonae* given the nature of the data.

Since polymorphism can eventually lead to fixation at both silent and replacement sites, fixed and polymorphic variation should be correlated. This can be addressed by implementing a MCDONALD and KREITMAN (1991) test. We examined this correlation in our data by using a *G*-test or Fisher's exact test when appropriate. In addition to analyzing the combined locus correlation (both *D. mojavensis* and *D. arizonae*), we also analyzed the lineage-specific correlation (see Tables 2 and 3). An overall pooling of the fixed differences between two species in a McDonald/Kreitman (MK) test will tend to mask the pattern of evolution within each species lineage. Therefore, a way to implement an MK test is to partition the fixed differences into each species lineage. For the analysis of lineage-specific evolution, we used the *D. navojoa* sequence to polarize the data. As seen in Table 2 the correlation is present for *Adh-2* at both the total locus and the species level. Conversely, this correlation does not seem to hold up when examining *Adh-1* (Table 3). By doing the lineage-specific analysis

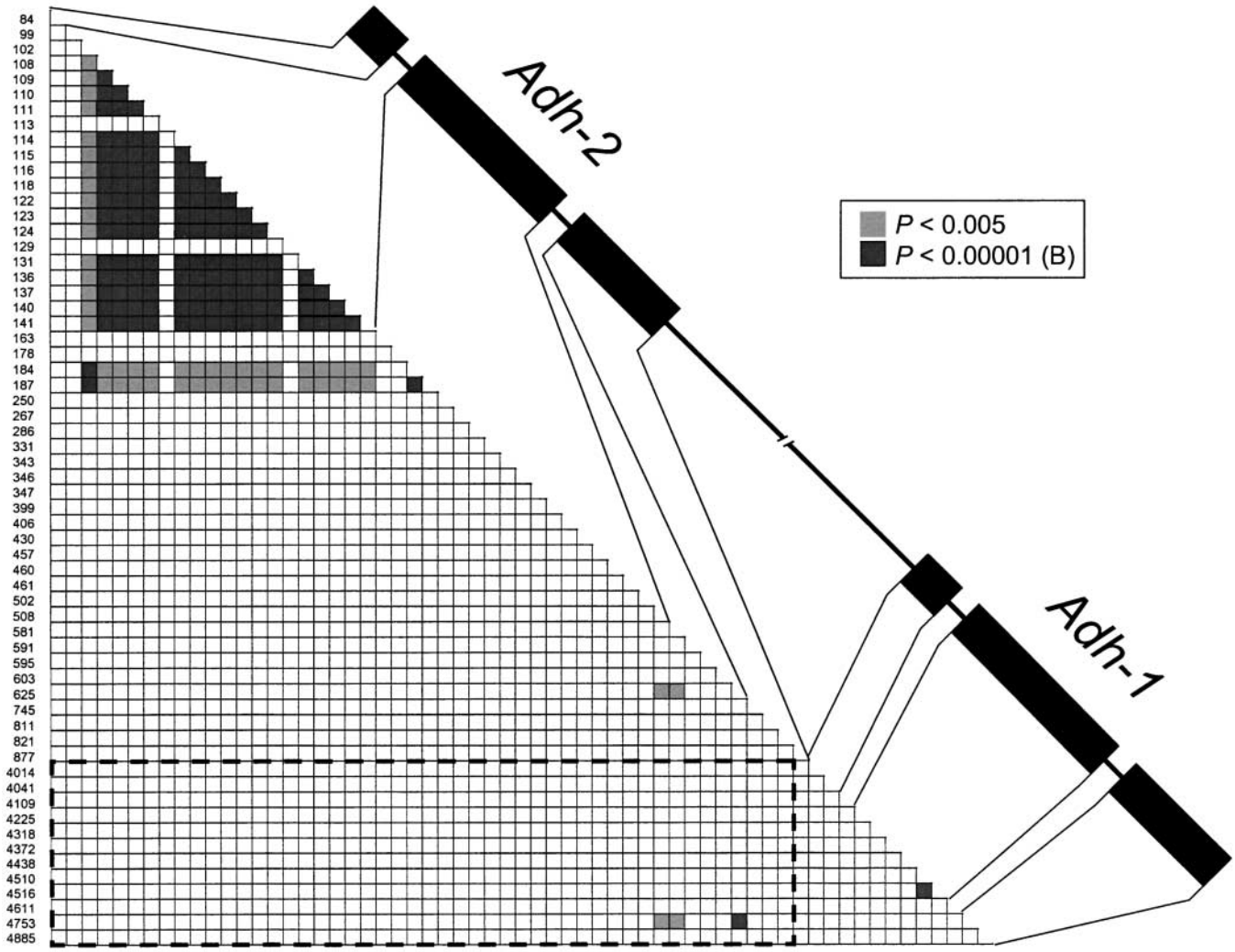

FIGURE 4.—Linkage disequilibrium in the *Adh* cluster of *D. mojavensis*. Linkage disequilibrium was determined using the chi-square test. The significance level was adjusted using the Bonferroni (B) correction for multiple comparisons. The darkened boxes represents exons, while the lines separating the boxes are introns. The dashed box reflects comparisons between the paralogs.

we can observe that the significant deviation found at *Adh-1* ($G = 7.21$, $P = 0.0072$) is due to *D. mojavensis* only. Two of the fixed replacement sites found in *D. mojavensis* are adjacent to each other, so it could be considered as either one or two mutational events. Overall this does not affect the result of the analysis if we assume it to be two mutational events; the deviation is significant ($P = 0.015$) as well if we assume it to be one event ($P = 0.035$).

DISCUSSION

Population size differences: The overall level of variation in both *D. mojavensis* and *D. arizonae* differs in similar proportion in the *Adh* paralogs. *D. arizonae* has a level of gene diversity ~1.5–2.5 times higher than that of *D. mojavensis* (Table 1). Our calculation of diversity from segregating sites (HUDSON 1990) estimates θ (WATTERSON 1975), which equals $4N\mu$ where N is the effective population size and μ is the neutral mutation rate. Assuming that the neutral mutation rates between *D. mojavensis* and *D. arizonae* are not different, this suggests a difference in population size between the species, *D. arizonae* being larger. Using the silent mutation rate (4.5×10^{-9} per base pair per generation) calculated from *D. melanogaster Adh* (HEY and WAKELEY 1997) and the overall θ for *Adh-1* and *Adh-2* (0.0249 and 0.0469, for *D. mojavensis* and *D. arizonae*, respectively) we estimated the effective population sizes in *D. mojavensis* and *D. arizonae* to be 1.4×10^6 and 2.6×10^6, respectively. We can also compare the variation at *Adh* in these two species with what is known in other Drosophila species. The level of variation in both paralogs of *D. arizonae* resembles that found for *D. pseudoobscura* (MORIYAMA and POWELL 1996). There is a difference in the level of variation between the *D. mojavensis* paralogs. The

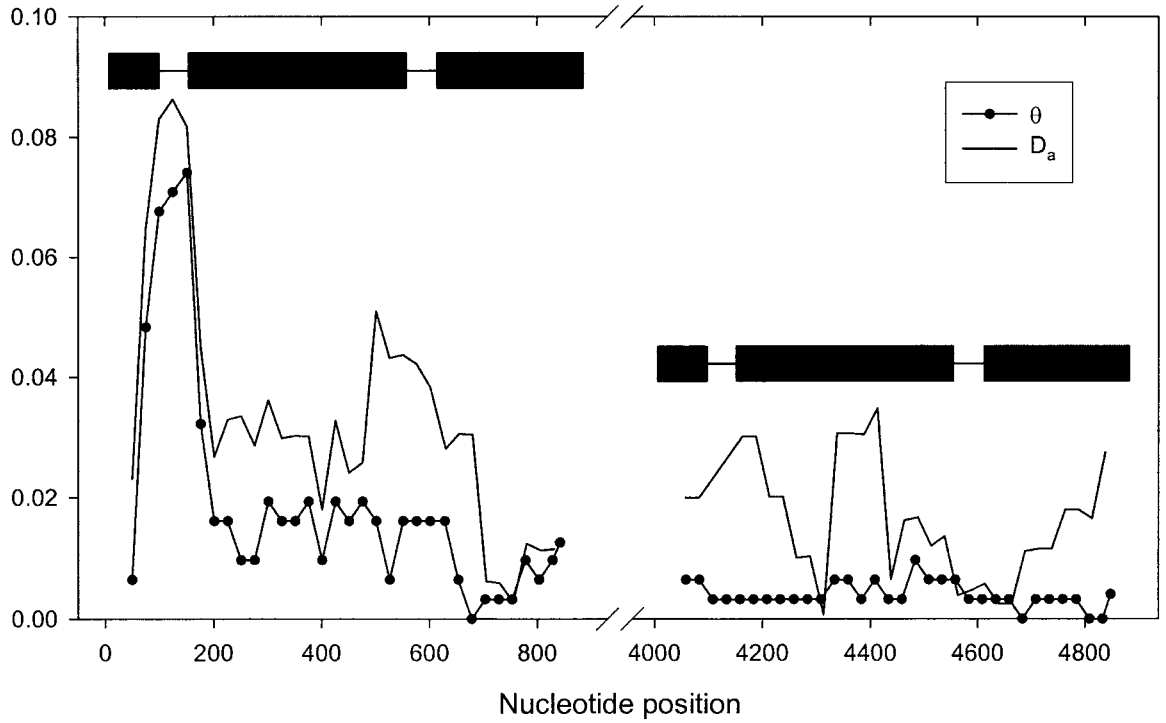

FIGURE 5.—Sliding-window analysis of inter- and intraspecific variation in *D. mojavensis Adh-2* and *Adh-1*. The total (synonymous, nonsynonymous, and intron) level of variation (θ) was determined for a sliding-window size of 100 bp and a step size of 25. The net nucleotide divergence (D_a) was determined using one *D. arizonae* sequence (ARTU 34).

variation found at *Adh-1* is similar to what is observed in *D. melanogaster*, while the variation observed in *Adh-2* resembles *D. simulans* (MORIYAMA and POWELL 1996). Given the significant MK test on *Adh-1* (Table 3), the *Adh-2* comparison is most adequate because hitchhiking can sweep out variation leading to a lower estimate of *N*.

Age of duplication and speciation: In the original sequence characterization of the *D. mojavensis Adh* cluster, ATKINSON *et al.* (1988) concluded that the duplication event that produced *Adh-1* and *Adh-2* occurred about 17.9 MYA. Our estimate of the age of the duplication is ~4.5-fold lower. There are two reasons for this large discrepancy between the estimates. First, sequencing errors in the initial inference of the sequence created an elevated level of divergence between the paralogs. The divergence between the paralogs (K_s = 0.197) in ATKINSON *et al.* (1988) was ~1.5 times greater than that observed in this study. Second and most important, the rate of silent substitution used to estimate the age of the duplication was much slower than the rate used in this study. That study used the mammalian rate of silent substitution that was used by BODMER and ASHBURNER (1984) in a study of Drosophila *Adh* in the *D. melanogaster* species group. In this study we used the mean rate of silent substitution calculated from Drosophila (MORIYAMA and GOJOBORI 1992). Even if we

TABLE 2

Total and lineage-specific intra- and interspecific variation at synonymous (syn.) and replacement (rep.) sites in *Adh-2*

		Fixed	Polymorphic	
D. mojavensis/D. arizonae	Syn.	8	49	Fisher's exact test, $P = 0.33$
	Rep.	0	12	
D. mojavensis	Syn.	5	19	Fisher's exact test, $P = 0.55$
	Rep.	0	6	
D. arizonae	Syn.	2	33	Fisher's exact test, $P = 0.99$
	Rep.	0	6	

The lineage-specific intra- and interspecific variation was calculated by polarizing the data using the *D. navojoa Adh-2* sequence. One synonymous fixed difference was removed from the lineage-specific analysis because of ambiguity. The total number of synonymous polymorphisms did not add up to the lineage-specific estimates due to shared polymorphisms between the species.

TABLE 3

Total and lineage-specific intra- and interspecific variation at synonymous and replacement sites in *Adh-1*

		Fixed	Polymorphic	
D. mojavensis/D. arizonae	Syn.	5	40	$G = 0.626$ (with William's correction), $P = 0.0072$
	Rep.	5	4	
D. mojavensis	Syn.	3	10	Fisher's exact test, $P = 0.035$
	Rep.	3[a]/4[b]	0	Fisher's exact test, $P = 0.015$
D. arizonae	Syn.	2	31	$G = 0.626$ (with William's correction), $P = 0.43$
	Rep.	1	4	

The lineage-specific intra- and interspecific variation was calculated by polarizing the data using the *D. navojoa Adh-1* sequence. Two fixed replacement differences (positions 83 and 84) in *D. mojavensis* occur adjacent to each other.

[a] Sites 83 and 84 considered as a one independent mutational event.

[b] Sites 83 and 84 considered as two independent mutational events.

estimate the age of the duplication using the slowest rate reported, 1.1×10^{-8} sites/year, the age of the duplication estimated for *D. mojavensis* and *D. arizonae* (6.1 and 7.5 MY, respectively) is still about three times younger than the Atkinson *et al.* (1988) estimate. This second, older calculation of the age of duplication using the slowest rate reported is in accordance with the estimate by Russo *et al.* (1995) of 6.5 ± 0.90 MY.

The phylogenetic relationships of the species in the *D. mojavensis* cluster are supported by previous work on the cytological evolution of these species (Heed 1982; Ruiz *et al.* 1990). In all studies, *D. navojoa* was found to be the most basal species of the species group, while *D. mojavensis* is the most derived (Heed 1982; Ruiz *et al.* 1990). Our estimate of the divergence between the *D. mojavensis* and *D. arizonae Adh* loci (1.7 and 3.1 MY for *Adh-1* and *Adh-2*, respectively) was lower than that previously reported (Russo *et al.* 1995). To obtain an estimate of the divergence of the species, intraspecific variation must be included in our analysis. Hudson *et al.* (1987) provide a set of equations in which the time of divergence T (in units of $2N$ generations) can be calculated taking into account intraspecific variation. We calculated the time of divergence (T) of *D. mojavensis* from *D. arizonae* to be $4.48N_e$ generations ago. Assuming that *D. mojavensis* goes through 6 generations in 1 year and using the above estimate of N_e we can estimate that the split between the species occurred $\sim 10^6$ years ago, more recent than our estimate using K_s. Using the HKA program (J. Hey) we performed 10,000 coalescent simulations based on observed levels of variation to estimate the 95% confidence intervals of T. We calculate the 95% confidence interval of the divergence time to be 0.19×10^6 to 3.4×10^6 years. This recent time of divergence supports many of the behavioral, physiological, and morphological studies on these species (Markow 1981; Krebs and Markow 1989; Markow and Toolson 1990; Markow 1991; Etges 1992; Stennett and Etges 1997; Knowles and Markow 2001). Evidence on the reproductive behavior and physiology of *D. mojavensis* and *D. arizonae* suggests a possible rapid divergence propelled, at least in part, by sexually antagonistic coevolution (Knowles and Markow 2001).

We find that a gene conversion event has occurred between intron 1 of *Adh-2* and *Adh-1*. Gene conversion will tend to homogenize the variation between the paralogs, causing an underestimation of the divergence between paralogs. In addition, gene conversion and hybridization will cause an underestimation of the sequence divergence. Species hybridization, followed by gene conversion between the paralogs will tend to have an effect on both paralogs, not just one intron; unless recombination and selection removed all the sequence belonging to one species, maintaining only the hybrid intron sequence, this scenario is highly unlikely. Therefore to avoid erroneous results we used only exons, more specifically third base positions, for constructing the gene tree (Figure 3).

Adaptive protein evolution at *Adh-1*: Overall we found a significant MK test at *Adh-1*. The significance of the test was due to an excess in replacement fixations (Table 3). This pattern of variation has been previously described to be indicative of selection for advantageous amino acid mutations (McDonald and Kreitman 1991; Eanes *et al.* 1993). We were also interested in determining if this pattern of evolution is reflected in both the *D. mojavensis* and the *D. arizonae* lineages. It appears that the excess in replacement fixations is largely associated with the *D. mojavensis* lineage (Table 3).

Two out of the three amino acid differences that occur in the *D. mojavensis* lineage are near a functionally important region of the molecule. Residue Val-236 fixed in *D. mojavensis* is next to Asp-237, a residue involved in the noncovalent interaction at the dimer surface (Benach *et al.* 1998, 1999). Residue Leu-61 is adjacent to residue Tyr-62, which comprises the coenzyme binding zone (Benach *et al.* 1998, 1999). Much is known about the kinetics and functional importance of the ADH Fast/Slow allozyme polymorphisms in *D. melanogaster* (Gelfand and McDonald 1980; Chambers 1988; Geer *et al.* 1990; Freriksen *et al.* 1994), even though it is

unclear if selection is specifically acting on the allozyme polymorphism itself or a linked site (Berry and Kreitman 1993). In *D. melanogaster*, the Fast/Slow polymorphism is due to a threonine-to-lysine substitution that occurs at residue 192. That residue is equivalent to Thr-191 in the *D. lebanonensis* ADH, for which the crystal structure is known (Benach *et al.* 1998, 1999). Threonine 191 acts as a gate on the active site once the NAD-alcohol ternary complex is formed. The adaptive significance, if any, of the Fast allele in *D. melanogaster* is still not fully understood. What is known is that the Fast allele possesses higher protein levels (Laurie and Stam 1988) and therefore exhibits a higher activity (Fletcher *et al.* 1978; Savakis *et al.* 1986), has different kinetic properties (Winberg *et al.* 1985), and segregates at greater frequencies at higher latitudes (Oakeshott *et al.* 1982).

Without a detailed examination of the differences in the kinetic characteristics between the *D. mojavensis* and *D. arizonae* ADH-1 enzyme it is difficult to examine the possible adaptive significance of the three amino acid residues fixed between the two lineages. Other sources of evidence may shed some light on the possible selective forces responsible for the rapid fixation of amino acids in the *D. mojavensis* lineage. A cactus host shift occurred during the divergence of *D. mojavensis* from *D. arizonae*; this host shift might be responsible for creating the selective force (Heed 1982; Ruiz and Heed 1988; Ruiz *et al.* 1990). ADH-1 is expressed in the larval stage of *D. mojavensis* (Batterham *et al.* 1983), and it is known that different cactus hosts exhibit different mixtures of alcohols (Vacek 1979; Fogleman 1982; Heed 1982; Kircher 1982). The larval stage of the fly is intimately associated with its host. Since *Adh-1* is expressed in the larvae, the host shift could have been responsible for the apparent adaptive evolution of *Adh-1*.

Maintenance of intron structure in *Adh-2*: Unlike the pattern of adaptive protein evolution observed at *Adh-1*, *Adh-2* does not seem to have undergone any recent episode of strong directional or balancing selection. One aspect of the variation at *Adh-2* that is striking is the heterogeneous distribution of the polymorphisms, more specifically, the fact that there is notable variation in intron 1, which is due to the segregation of two distinct intron haplotypes in both species.

One of the *Adh-2* intron 1 haplotypes is identical to the *Adh-1* intron 1 [we refer to this haplotype as LA1 (like *Adh-1*) and the other as LA2]. Intron 1 is 55 bp long, but the intron haplotype is only 34 bp long (from position 108 to 141, see Figure 1). We propose that the LA1 haplotype originated from a gene conversion event between *Adh-1* and *Adh-2*. The *Adh-2* intron sequence was replaced by that of *Adh-1*. In *D. mojavensis* the LA1 haplotype is found at high frequency, while in *D. arizonae* it segregates at a low frequency. The *Adh-1* intron 1 of *D. mojavensis* and the *Adh-1* intron 1 of *D. arizonae* are almost identical (Figure 2). Gene conversion is a common phenomenon in duplicated genes in Drosophila [*e.g.*, *Amylase* (Hickey *et al.* 1991; Popadic and Anderson 1995), *Attacin* (Lazzaro and Clark 2001), and *hsp70* (Bettencourt and Feder 2002)], as well as in other taxa [*e.g.*, opsins in primates (Boissinot *et al.* 1998), MHC in red-winged black birds (Edwards *et al.* 1998), and HLA in humans (Takahata and Satta 1998)]. Gene conversion has been generally viewed as a force slowing or even preventing the diversification of duplicated loci (Walsh 1987). Conversely, under certain circumstances gene conversions could aid in the adaptive evolution of paralogs by transferring mutation *en masse* (Hansen *et al.* 2000). Such models of the possible effect of gene conversion do not place much emphasis on the transition stage (*i.e.*, polymorphic). In this study we have observed gene conversion alleles segregating in *Adh-2* in both *D. mojavensis* and *D. arizonae*.

There are three explanations for the existence of the same intron polymorphism in both species. The simplest of all possible explanations is that the LA1 haplotype independently originated in both species. This would encompass two independent and precisely identical gene conversion events, one in *D. mojavensis* and one in *D. arizonae*. Although simple, this seems highly improbable. Another explanation is that there was a gene conversion event between *Adh-1* and *Adh-2* in one species and a recent introgression introduced the LA1 haplotype into the other species. In lab conditions the two species can be forced to hybridize, but at very low frequency (Wasserman and Koepfer 1977). Several character differences can affect the frequency of copulation in these species, such as behavior (Markow 1981) and cuticle hydrocarbon composition (Markow and Toolson 1990). As well as prezygotic mechanisms, postzygotic mechanisms such as a reduction in offspring viability and fertility of species hybrids have been observed (Ruiz *et al.* 1990). Introgression will produce a genome-wide effect; therefore evidence of this should be visible in several regions of the genome. If introgression were the source of the polymorphism, our data would suggest that only the intron 1 of *Adh-2* was introgressed and nothing else, and this seems highly unlikely. As can be observed by the neighbor-joining tree (Figure 3), all *D. mojavensis* sequences within each paralog coalesce to one another. Therefore, there is no strong evidence of hybridization from the sequence data, and from the known behavioral data this seem unlikely. Therefore, an introgression event between the species might not be the most likely explanation for the intron 1 polymorphism in *Adh-2*. The apparent shared polymorphism observed in exon 2 (sites 430 and 460, see Figure 1) of *Adh-2* is due to a second gene conversion between *Adh-1* that occurred only in the *D. mojavensis* lineage. This will further be described in future work.

Third, the polymorphism in both species may extend from an ancient gene conversion that predates the diver-

gence of *D. mojavensis* and *D. arizonae* and has been maintained in both species. Selection needs not be implied when explaining the sharing of a polymorphism between two species. Assuming the two alleles were equally common in the ancestral population the expected mean time loss of the shared polymorphism is $1.7N$ generations (Clark 1997). The distribution of time to loss has a long tail; 95% of the time the losses occur by $3.8N$ generations and 99% of the time by $5.3N$ generations (Clark 1997). Our data suggest that the divergence between *D. mojavensis* and *D. arizonae* occurred $4.48N$ generations ago; therefore it is not very likely that the shared polymorphism has been neutrally maintained.

Additionally, if the shared polymorphism is neutrally maintained, then we would expect that in each species a considerable amount of variation would be segregating in the two intron haplotypes. To examine this we have split our intron 1 data into three classes per species: LA2 haplotype, LA1 haplotype, and all *Adh-1* haplotypes. The level of diversity (θ) observed for the LA2 haplotype in *D. mojavensis* and *D. arizonae* was 0.025 and 0.097, respectively. The variation observed for the *Adh-1* intron in *D. mojavensis* and *D. arizonae* was lower, 0.006 and 0, respectively. Additionally, the level of variation within the LA1 haplotype was also lower (0.019 and 0, for *D. mojavensis* and *D. arizonae*, respectively). Although variation has accumulated in the LA2 haplotype, variation has not greatly accumulated in the LA1 haplotype; this same pattern of low variation is also present in intron 1 of *Adh-1*. Therefore it is possible that selection might be preventing the accumulation of variation at the *Adh-1* intron and the homologous LA1 haplotype. The difference in diversity between the LA1 and LA2 might be hard to reconcile with the proposed age of haplotypes. However, the possibility exists that the converted allele has simply had historically a smaller effective population size.

The pattern of linkage disequilibrium across intron 1 is striking given the observed levels of overall recombination apparent in the *D. mojavensis Adh* cluster. If selection is not involved, then we would expect that recombination would have broken down the linkage disequilibrium. Given the value of γ (0.0134/bp) estimated for *D. mojavensis* we can obtain an estimate c (9.6×10^{-9}), the recombination rate per base per generation. Using the *D. mojavensis* and *D. arizonae* divergence of $4.48N$ or 6.3×10^6 generations and the fact that the intron haplotype is 34 bp, we can estimate that about two recombination events should occur in a 34-bp span of sequence since the species split. The estimate of two recombination events should be taken to be a maximum only. The number of sampled alleles and the fact that a recombination event between homologous alleles cannot be observed could be an explanation for the lack of recombinant introns in our sample. Although this is a possibility, we believe that the expected estimate of recombination does not agree with the levels and distribution of disequilibrium observed. This suggests that there might be selection to maintain the intron haplotype structure in *D. mojavensis* and *D. arizonae*. Similar patterns of disequilibrium in introns have been observed in other Drosophila.

Independent of *D. mojavensis* and *D. arizonae*, *D. pseudoobscura* also possesses a duplication of the *Adh* locus (Schaeffer and Aquadro 1987). In *D. pseudoobscura*, two significant clusters of linkage disequilibrium were found, both occurring within introns (Schaeffer and Miller 1993). The adult intron, intron 1, and intron 2 all form pre-mRNA structures, but the two major haplotypes segregating in intron 2 form highly stable pre-mRNA stem-loop structures (Kirby *et al.* 1995). Pre-mRNA structures in *Adh* have also been suggested in several Drosophila species (Stephan and Kirby 1993). The model for the maintenance of pre-mRNA structure is based on the selection for compensatory mutations that stabilize RNA structure (Stephan 1996). The function of pre-mRNA structures has recently been linked to the regulation of expression (Antezana and Kreitman 1999; Carlini *et al.* 2001). Loci of low expression have low levels of codon bias and a more even distribution of G/A/T/C in the third position. Such low-expression loci tend to have more stable pre-mRNA structures (Carlini *et al.* 2001). Highly stable pre-mRNA structures produce pauses in the translation of proteins, which may aid in protein expression and folding (Netzer and Hartl 1997; Niepel *et al.* 1999).

In our survey of *Adh-2*, no intron 1 recombinants were observed in either species. Evidence suggests that the structure of intron 1, especially in *Adh-1* and the LA1 haplotype, has been independently maintained in both species. Both the LA2 and the LA1 haplotypes form pre-mRNA stem-loop structures that are more stable (calculated as the change in free energy) than all the potential recombinants between the two haplotypes (data not shown). Therefore, selection might be responsible for maintaining the intron haplotype structure of *Adh* in *D. mojavensis* and *D. arizonae* as it has been observed to occur in other Drosophila. The regulation of expression and translation seems the most likely role for the *Adh* pre-mRNA stem-loop structures. Further analysis of this latter point is needed.

Subfunctionalization of *Adh* paralogs: This study shows that the *Adh* paralogs of *D. mojavensis* have traveled different evolutionary paths. It is of fundamental interest to understand why their evolutionary paths have diverged. Duplication followed by amino acid fixations has been traditionally placed as the major force in the evolution of novel gene functions (Ohno 1970; Ohta 1974, 1987, 1988a,b). These models assume that the duplicated locus is bombarded by mutations, most of them being deleterious, until a beneficial combination producing a novel function is created and fixed. These models do not take into account the role of gene expression in the

evolution of paralogs. If the original locus had several modes of expression, following fixation, specialization of each paralog could occur if they limit their expression to that of a subset of the original locus (Hughes 1994). This model has been further developed by Force *et al.* (1999) in which they propose that following the fixation of the paralogs, mutations will occur in one of the regulatory regions of a paralog that will silence expression. Assuming the original locus had several modes of expression, the end product of subfunctionalization will be the existence of paralogs with no or nearly no overlapping expression. Loci being expressed in different tissues or at different times during development will experience different environmental pressures, which may lead to the specialization and divergent evolution between the paralogs. The *Adh* system provides an especially good fit to the model, given the fact that it is known that the duplication event that produced *Adh-1* and *Adh-2* was not complete: the distal promoter for *Adh-2* was lost during the duplication (Atkinson *et al.* 1988).

The efficacy of natural selection on an enzyme locus is highly dependent on the environment in which the enzymes are expressed (Dykhuizen and Hartl 1980; Hartl *et al.* 1985; Dykhuizen *et al.* 1987). In *D. mojavensis* and *D. arizonae* the external and internal/cellular environments experienced by each paralog differ. According to metabolic flux analysis, the relative role of enzyme activity to overall pathway flux is quantified as the control coefficient (Fell 1997). In *D. melanogaster*, the control coefficient for ADH recorded in adults is 0.022 (Middleton and Kacser 1983), while that measured in larvae is ~1.0 (Freriksen *et al.* 1991). This implies that changes that affect ADH enzyme activity will have a much greater effect on flux in the larval stage than in the adult stage. The differences in the modes of expression of *Adh-1* and *Adh-2* have provided for the possibility of natural selection to aid in their divergence.

The study of duplicated loci, most specifically recently duplicated loci, offers a very powerful system for the study of evolution at a locus. Paralogs with nonoverlapping modes of expression highlight how changes in the environment experienced by an enzyme have the potential to drastically alter its evolution. The combination of knowledge of the ecology of the study organism, population genetics studies of duplicated metabolic enzymes, and studies of enzyme function will aid in the study of the fundamental question of maintenance of natural variation.

The authors thank Therese Markow for assistance in the field and for providing the *D. arizonae* isofemale lines and Brian Verrelli and Thomas Merritt for comments and suggestions on earlier versions of the manuscript. This research was supported by U.S. Public Health Service grant GM-45247 to W.F.E. and by National Science Foundation Predoctoral Fellowship and W. Burghardt Turner Fellowship to L.M.M. This is contribution no. 1103 from the Graduate Program in Ecology and Evolution, State University of New York at Stony Brook.

LITERATURE CITED

Antezana, M. A., and M. Kreitman, 1999 The nonrandom location of synonymous codons suggests that reading frame-independent forces have patterned codon preferences. J Mol. Evol. **49:** 36–43.

Atkinson, P. W., L. E. Mills, W. T. Starmer and D. L. Sullivan, 1988 Structure and evolution of the *Adh* genes of *Drosophila mojavensis*. Genetics **120:** 713–723.

Batterham, P., J. A. Lovett, W. T. Starmer and D. T. Sullivan, 1983 Differential regulation of duplicate alcohol dehydrogenase genes in *Drosophila mojavensis*. Dev. Biol. **96:** 346–354.

Begun, D. J., 1997 Origin and evolution of a new gene descended from alcohol dehydrogenase in *Drosophila*. Genetics **145:** 375–382.

Benach, J., S. Atrian, R. Gonzalez-Duarte and R. Ladenstein, 1998 The refined crystal structure of *Drosophila lebanonensis* alcohol dehydrogenase at 1.9 Å resolution. J. Mol. Biol. **282:** 383–399.

Benach, J., S. Atrian, R. Gonzalez-Duarte and R. Ladenstein, 1999 The catalytic reaction and inhibition mechanism of *Drosophila* alcohol dehydrogenase: observation of an enzyme-bound NAD-ketone adduct at 1.4 Å resolution by X-ray crystallography. J. Mol. Biol. **289:** 335–355.

Berry, A., and M. Kreitman, 1993 Molecular analysis of an allozyme cline alcohol dehydrogenase in *Drosophila melanogaster* on the East Coast of North America. Genetics **134:** 869–893.

Bettencourt, B. R., and M. E. Feder, 2002 Rapid concerted evolution via gene conversion at the *Drosophila* hsp70 genes. J. Mol. Evol. **54:** 569–586.

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.

Boissinot, S., Y. Tan, S. K. Shyue, H. Schneider, I. Sampaio *et al.*, 1998 Origins and antiquity of X-linked triallelic color vision systems in New World monkeys. Proc. Natl. Acad. Sci. USA **95:** 13749–13754.

Carlini, D. B., Y. Chen and W. Stephan, 2001 The relationship between third-codon position nucleotide content, codon bias, mRNA secondary structure and gene expression in the drosophilid alcohol dehydrogenase genes *Adh* and *Adhr*. Genetics **159:** 623–633.

Chambers, G. K., 1988 The *Drosophila* alcohol dehydrogenase gene enzyme system. Adv. Genet. **25:** 39–107.

Clark, A. G., 1997 Neutral behavior of shared polymorphism. Proc. Natl. Acad. Sci. USA **94:** 7730–7734.

Dykhuizen, D., and D. L. Hartl, 1980 Selective neutrality of 6pgd allozymes in *Escherichia coli* and the effects of genetic background. Genetics **96:** 801–817.

Dykhuizen, D. E., A. M. Dean and D. L. Hartl, 1987 Metabolic flux and fitness. Genetics **115:** 25–31.

Eanes, W. F., M. Kirchner and J. Yoon, 1993 Evidence for adaptive evolution of the *G6pd* gene in the *Drosophila melanogaster* and *Drosophila simulans* lineages. Proc. Natl. Acad. Sci. USA **90:** 7475–7479.

Edwards, S. V., J. Gasper and M. March, 1998 Genomics and polymorphism of Agph-DAB1, an Mhc class II B gene in red-winged blackbirds (Agelaius phoeniceus). Mol. Biol. Evol. **15:** 236–250.

Etges, W. J., 1992 Premating isolation is determined by larval substrates in cactophilic *Drosophila mojavensis*. Evolution **46:** 1945–1950.

Fell, D., 1997 *Understanding the Control of Metabolism*. Portland Press, London.

Fellows, D. F., and W. B. Heed, 1972 Factors affecting host plant selection in desert-adapted cactiphilic *Drosophila*. Ecology **53:** 850–858.

Fletcher, T. S., F. J. Ayala, D. R. Thatcher and G. K. Chambers, 1978 Structural analysis of the ADH-S electromorph of *Drosophila melanogaster*. Proc. Natl. Acad. Sci. USA **75:** 5609–5612.

Fogleman, J. C., 1982 The role of volatiles in the ecology of cactophilic *Drosophila*, pp. 191–208 in *Ecological Genetics and Evolution: The Cactus-Yeast-Drosophila Model System*, edited by J. S. F. Barker and W. T. Starmer. Academic Press, New York.

Force, A., M. Lynch, F. B. Pickett, A. Amores, Y. L. Yan *et al.*, 1999 Preservation of duplicate genes by complementary, degenerative mutations. Genetics **151:** 1531–1545.

Freriksen, A., D. Seykens, W. Scharloo and P. W. H. Heinstra, 1991 Alcohol dehydrogenase controls the flux from ethanol

into lipids in *Drosophila* larvae: a C-13 NMR-study. J. Biol. Chem. **266:** 21399–21403.

Freriksen, A., B. L. A. Deruiter, H. J. Groenenberg, W. Scharloo and P. W. H. Heinstra, 1994 A multilevel approach to the significance of genetic variation in alcohol dehydrogenase of *Drosophila*. Evolution **48:** 781–790.

Fu, Y. X., and W. H. Li, 1993 Statistical tests of neutrality of mutations. Genetics **133:** 693–709.

Gasperi, G., L. Baruffi, A. R. Malacrida and A. S. Robinson, 1992 A biochemical genetic study of alcohol dehydrogenase isozymes of the medfly, Ceratitis capitata wied. Biochem. Genet. **30:** 289–304.

Geer, B. W., P. W. H. Heinstra, A. M. Kapoun and A. V. D. Zel, 1990 Alcohol dehydrogenase and alcohol tolerance in *Drosophila melanogaster*, pp. 231–252 in *Ecological and Evolutionary Genetics of Drosophila*, edited by J. S. F. Barker, W. T. Starmer and R. J. MacIntyre. Plenum Press, New York.

Gelfand, L. J., and J. F. McDonald, 1980 Relationship between ADH activity and behavioral response to environmental alcohol in *Drosophila*. Behav. Genet. **10:** 237–249.

Goulielmos, G. N., N. Cosmidis, M. Loukas, S. Tsakas and E. Zouros, 2001 Characterization of two alcohol dehydrogenase (*Adh*) loci from the olive fruit fly, *Bactrocera (Dacus) oleae* and implications for *Adh* duplication in dipteran insects. J. Mol. Evol. **52:** 29–39.

Griffith, E. C., and J. R. Powell, 1997 *Adh* nucleotide variation in *Drosophila willistoni*: high replacement polymorphism in an electrophoretically monomorphic protein. J. Mol. Evol. **45:** 232–237.

Hansen, T. F., A. J. R. Carter and C. H. Chiu, 2000 Gene conversion may aid adaptive peak shifts. J. Theor. Biol. **207:** 495–511.

Hartl, D. L., D. E. Dykhuizen and A. M. Dean, 1985 Limits of adaptation: the evolution of selective neutrality. Genetics **111:** 655–674.

Heed, W. B., 1978 Ecology and genetics of Sonoran desert *Drosophila*, pp. 109–126 in *Ecological Genetics: The Interface*, edited by P. F. Brussard. Springer-Verlag, New York.

Heed, W. B., 1982 The origin of *Drosophila* in the Sonoran desert, pp. 65–80 in *Ecological Genetics and Evolution: The Cactus-Yeast-Drosophila Model System*, edited by J. S. F. Barker and W. T. Starmer. Academic Press, New York.

Hey, J., and J. Wakeley, 1997 A coalescent estimator of the population recombination rate. Genetics **145:** 833–846.

Hickey, D. A., L. Ballycuif, S. Abukashawa, V. Payant and B. F. Benkel, 1991 Concerted evolution of duplicated protein coding genes in *Drosophila*. Proc. Natl. Acad. Sci. USA **88:** 1611–1615.

Hudson, R. R., 1987 Estimating the recombination parameter of a finite population model without selection. Genet. Res. **50:** 245–250.

Hudson, R. R., 1990 Gene genealogies and the coalescent process, pp. 1–44 in *Oxford Surveys in Evolutionary Biology*, edited by D. J. Futuyma and J. Antonovics. Oxford University Press, Oxford.

Hudson, R. R., and N. L. Kaplan, 1988 The coalescent process in models with selection and recombination. Genetics **120:** 831–840.

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

Hudson, R. R., K. Bailey, D. Skarecky, J. Kwiatowski and F. J. Ayala, 1994 Evidence for positive selection in the superoxide dismutase (*Sod*) region of *Drosophila melanogaster*. Genetics **136:** 1329–1340.

Hughes, A. L., 1994 The evolution of functionally novel proteins after gene duplication. Proc. R. Soc. Lond. Ser. B Biol. Sci. **256:** 119–124.

Kirby, D. A., S. V. Muse and W. Stephan, 1995 Maintenance of pre-mRNA secondary structure by epistatic selection. Proc. Natl. Acad. Sci. USA **92:** 9047–9051.

Kircher, H. W., 1982 Chemical composition of cacti and its relationship to Sonoran Desert *Drosophila*, pp. 143–158 in *Ecological Genetics and Evolution: The Cactus-Yeast-Drosophila Model System*, edited by J. S. F. Barker and W. T. Starmer. Academic Press, New York.

Kliman, R. M., P. Andolfatto, J. A. Coyne, F. Depaulis, M. Kreitman *et al.*, 2000 The population genetics of the origin and divergence of the *Drosophila simulans* complex species. Genetics **156:** 1913–1931.

Knowles, L. L., and T. A. Markow, 2001 Sexually antagonistic coevolution of a postmating-prezygotic reproductive character in desert *Drosophila*. Proc. Natl. Acad. Sci. USA **98:** 8692–8696.

Krebs, R. A., and T. A. Markow, 1989 Courtship behavior and control of reproductive isolation in *Drosophila mojavensis*. Evolution **43:** 908–913.

Kreitman, M., 1983 Nucleotide polymorphisms at the alcohol dehydrogenase locus of *Drosophila melanogaster*. Nature **304:** 412–416.

Kumar, S., K. Tamura, I. B. Jakobsen and M. Nei, 2001 MEGA2: molecular evolutionary genetics analysis software. Bioinformatics **17:** 1244–1245.

Laurie, C. C., and L. F. Stam, 1988 Quantitative analysis of RNA produced by Slow and Fast alleles of *Adh* in *Drosophila melanogaster*. Proc. Natl. Acad. Sci. USA **85:** 5161–5165.

Laurie, C. C., and L. F. Stam, 1994 The effect of an intronic polymorphism on alcohol dehydrogenase expression in *Drosophila melanogaster*. Genetics **138:** 379–385.

Lazzaro, B. P., and A. G. Clark, 2001 Evidence for recurrent paralogous gene conversion and exceptional allelic divergence in the *Attacin* genes of *Drosophila melanogaster*. Genetics **159:** 659–671.

Markow, T. A., 1981 Courtship behavior and control of reproductive isolation between *Drosophila mojavensis* and *Drosophila arizonensis*. Evolution **35:** 1022–1026.

Markow, T. A., 1991 Sexual isolation among populations of *Drosophila mojavensis*. Evolution **45:** 1525–1529.

Markow, T. A., and E. C. Toolson, 1990 Temperature effects on epicuticular hydrocarbons and sexual isolation in *Drosophila mojavensis*, pp. 315–331 in *Ecological and Evolutionary Genetics of Drosophila*, edited by J. S. F. Barker, W. T. Starmer and R. J. MacIntyre. Plenum Press, New York.

McAllister, B. F., and B. Charlesworth, 1999 Reduced sequence variability on the neo-Y chromosome of *Drosophila americana americana*. Genetics **153:** 221–233.

McDonald, J. H., 1996 Detecting non-neutral heterogeneity across a region of DNA sequence in the ratio of polymorphism to divergence. Mol. Biol. Evol. **13:** 253–260.

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

Mercot, H., D. Defaye, P. Capy, E. Pla and J. R. David, 1994 Alcohol tolerance, ADH activity, and ecological niche of *Drosophila* species. Evolution **48:** 746–757.

Middleton, R. J., and H. Kacser, 1983 Enzyme variation, metabolic flux and fitness: alcohol dehydrogenase in *Drosophila melanogaster*. Genetics **105:** 633–650.

Moriyama, E. N., and T. Gojobori, 1992 Rates of synonymous substitution and base composition of nuclear genes in *Drosophila*. Genetics **130:** 855–864.

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

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

Netzer, W. J., and F. U. Hartl, 1997 Recombination of protein domains facilitated by co-translational folding in eukaryotes. Nature **388:** 343–349.

Niepel, M., J. Ling and D. R. Gallie, 1999 Secondary structure in the 5′-leader or 3′-untranslated region reduces protein yield but does not affect the functional interaction between the 5′-cap and the poly(A) tail. FEBS Lett. **462:** 79–84.

Oakeshott, J. G., J. B. Gibson, P. R. Anderson, W. R. Knibb, D. G. Anderson *et al.*, 1982 Alcohol dehydrogenase and glycerol-3-phosphate dehydrogenase clines in *Drosophila melanogaster* on different continents. Evolution **36:** 86–96.

Ohno, S., 1970 *Evolution by Gene Duplication*. Springer-Verlag, Heidelberg, Germany.

Ohta, T., 1974 Mutational pressure as main cause of molecular evolution and polymorphism. Nature **252:** 351–354.

Ohta, T., 1987 Simulating evolution by gene duplication. Genetics **115:** 207–213.

Ohta, T., 1988a Evolution by gene duplication and compensatory advantageous mutations. Genetics **120:** 841–847.

Ohta, T., 1988b Time for acquiring a new gene by duplication. Proc. Natl. Acad. Sci. USA **85:** 3509–3512.

Ohta, T., 1989 Role of gene duplication in evolution. Genome **31:** 304–310.

Popadic, A., and W. W. Anderson, 1995 Evidence for gene conver-

sion in the amylase multigene family of *Drosophila pseudoobscura*. Mol. Biol. Evol. **12:** 564–572.

Powell, J. R., 1997 *Progress and Prospects in Evolutionary Biology: The Drosophila Model*. Oxford University Press, New York.

Prakash, S., 1977 Genetic divergence in closely related sibling species *Drosophila pseudoobscura*, *Drosophila persimilis* and *Drosophila miranda*. Evolution **31:** 14–23.

Rozas, J., and R. Rozas, 1999 DnaSP version 3: an integrated program for molecular population genetics and molecular evolution analysis. Bioinformatics **15:** 174–175.

Ruiz, A., and W. B. Heed, 1988 Host-plant specificity in the cactophilic *Drosophila mulleri* species complex. J. Anim. Ecol. **57:** 237–249.

Ruiz, A., W. B. Heed and M. Wasserman, 1990 Evolution of the Mojavensis cluster of cactophilic *Drosophila* with descriptions of two new species. J. Hered. **81:** 30–42.

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

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:** 194–207.

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

Schaeffer, S. W., and E. L. Miller, 1992a Estimates of gene flow in *Drosophila pseudoobscura* determined from nucleotide sequence analysis of the alcohol dehydrogenase region. Genetics **132:** 471–480.

Schaeffer, S. W., and E. L. Miller, 1992b Molecular population genetics of an electrophoretically monomorphic protein in the alcohol dehydrogenase region of *Drosophila pseudoobscura*. Genetics **132:** 163–178.

Schaeffer, S. W., and E. L. Miller, 1993 Estimates of linkage disequilibrium and the recombination parameter determined from segregating nucleotide sites in the alcohol dehydrogenase region of *Drosophila pseudoobscura*. Genetics **135:** 541–552.

Starmer, W. T., W. B. Heed and E. S. Rockwood-Sluss, 1977 Extension of longevity in *Drosophila mojavensis* by environmental ethanol: differences between subraces. Proc. Natl. Acad. Sci. USA **74:** 387–391.

Stennett, M. D., and W. J. Etges, 1997 Premating isolation is determined by larval rearing substrates in cactophilic *Drosophila mojavensis*. III. Epicuticular hydrocarbon variation is determined by use of different host plants in *Drosophila mojavensis* and *Drosophila arizonae*. J. Chem. Ecol. **23:** 2803–2824.

Stephan, W., 1996 The rate of compensatory evolution. Genetics **144:** 419–426.

Stephan, W., and D. A. Kirby, 1993 Rna folding in *Drosophila* shows a distance effect for compensatory fitness interactions. Genetics **135:** 97–103.

Tajima, F., 1989 Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics **123:** 585–595.

Takahata, N., and Y. Satta, 1998 Selection convergence, and intragenic recombination in HLA diversity. Genetica **103:** 157–169.

Triglia, T., M. G. Peterson and D. J. Kemp, 1988 A procedure for in vitro amplification of DNA segments that lie outside the boundaries of known sequences. Nucleic Acids Res. **16:** 8186.

Vacek, D. C., 1979 The microbial ecology of the host plants of *Drosophila mojavensis*. Thesis, University of Arizona, Tucson, AZ.

Walsh, J. B., 1987 Sequence dependent gene conversion: Can duplicated genes diverge fast enough to escape conversion? Genetics **117:** 543–557.

Wasserman, M., 1982 Cytological evolution in the *Drosophila repleta* species group, pp. 49–64 in *Ecological Genetics and Evolution: The Cactus-Yeast-Drosophila Model System*, edited by J. S. F. Barker and W. T. Starmer. Academic Press, New York.

Wasserman, M., and H. R. Koepfer, 1977 Character displacement for sexual isolation between *Drosophila mojavensis* and *Drosophila arizonensis*. Evolution **31:** 812–823.

Watterson, G. A., 1975 On the number of segregating sites in genetical models without recombination. Theor. Popul. Biol. **20:** 256–276.

Winberg, J. O., R. Hovik and J. S. McKinley-McKee, 1985 The alcohol dehydrogenase alleloenzymes ADH-S and ADH-F from the fruitfly *Drosophila melanogaster*: an enzymatic rate assay to determine the active site concentration. Biochem. Genet. **23:** 205–216.

Winnepenninckx, B., T. Backeljau and R. Dewachter, 1993 Extraction of high molecular weight DNA from mollusks. Trends Genet. **9:** 407.

Communicating editor: S. W. Schaeffer