Molecular Ecology (2005) **14**, 2223–2231 doi: 10.1111/j.1365-294X.2005.02532.x

Activity variation in alcohol dehydrogenase paralogs is associated with adaptation to cactus host use in cactophilic *Drosophila*

LUCIANO M. MATZKIN*
Department of Ecology and Evolution, State University of New York at Stony Brook, Stony Brook, New York 11794–5245

Abstract

***Drosophila mojavensis* and *Drosophila arizonae* are species of cactophilic flies that share a recent duplication of the alcohol dehydrogenase (*Adh*) locus. One paralog (*Adh-2*) is expressed in adult tissues and the other (*Adh-1*) in larvae and ovaries. Enzyme activity measurements of the ADH-2 amino acid polymorphism in *D. mojavensis* suggest that the Fast allozyme allele has a higher activity on 2-propanol than 1-propanol. The Fast allele was found at highest frequency in populations that utilize hosts with high proportions of 2-propanol, while the Slow allele is most frequent in populations that utilize hosts with high proportions of 1-propanol. This suggests that selection for ADH-2 allozyme alleles with higher activity on the most abundant alcohols is occurring in each *D. mojavensis* population. In the other paralog, ADH-1, significant differences between *D. mojavensis* and *D. arizonae* are associated with a previously shown pattern of adaptive protein evolution in *D. mojavensis*. Examination of protein sequences showed that a large number of amino acid fixations between the paralogs have occurred in catalytic residues. These changes are potentially responsible for the significant difference in substrate specificity between the paralogs. Both functional and sequence variation within and between paralogs suggests that *Adh* has played an important role in the adaptation of *D. mojavensis* and *D. arizonae* to their cactophilic life.**

Keywords: activity variation, alcohol dehydrogenase, cactophilic *Drosophila*, duplication, host adaptation, subfunctionalization

Received 14 November 2004; revision accepted 7 February 2005

Introduction

Population-level studies have provided several examples suggesting natural selection has a role in shaping the pattern of variation at a locus (reviewed in Eanes 1999). Although such studies are informative, one criticism is that there is often a lack of connection between the ecology of the organism and the pattern of selection observed at the locus. In fact, the complexity of some systems may prevent such connections from being discovered. In contrast, the association between variation at alcohol dehydrogenase (*Adh*) in *Drosophila* and the alcohol environment of the flies' host is a tractable system in which to study such connections.

Correspondence: Luciano M. Matzkin, Present address: Department of Ecology and Evolutionary Biology, University of Arizona, PO Box 210088, Tucson, AZ 85721–0088; Fax: (520) 626 3522; E-mail: lmatzkin@email.arizona.edu

Drosophila species utilize a myriad of hosts with varying composition of alcohols (Powell 1997), and it has been suggested that *Adh* has played a role in the adaptation of certain *Drosophila* species to their specific host (Mercot *et al*. 1994; Atrian *et al*. 1998).

Adh duplications have independently occurred in several species of *Drosophila*. One such duplication is shared by *Drosophila mojavensis* and *Drosophila arizonae* (see Fig. 1; Atkinson *et al*. 1988). These species contain two functional *Adh* loci, *Adh-2* and *Adh-1*, physically separated by about 3 kb. The age of the duplication that created *Adh-2* and *Adh-1* has been estimated to have occurred about 4 million years ago (Ma), while the species diverged about 2 Ma (Matzkin & Eanes 2003; Matzkin 2004). In *D. mojavensis* and *D. arizonae*, *Adh-1* is expressed in the larvae and ovaries while *Adh-2* is expressed in the remaining adult tissues (Batterham *et al*. 1983). *Drosophila mojavensis* and *D. arizonae* are cactophilic species, inhabiting the deserts of northwestern

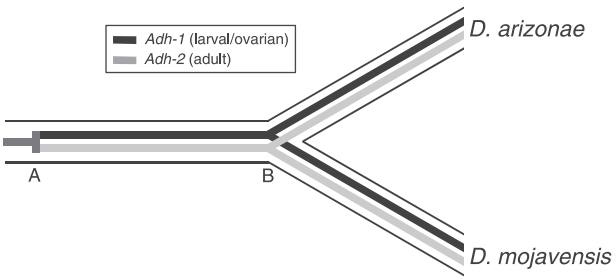

Fig. 1 The relationship between the *Adh* paralogs and *Drosophila mojavensis* and *Drosophila arizonae*. The spatio-temporal pattern of expression of each paralog is noted in parenthesis. Point A refers to the duplication of the paralogs *c.* 4 Ma. Point B refers to the divergence of the two species *c.* 2 Ma.

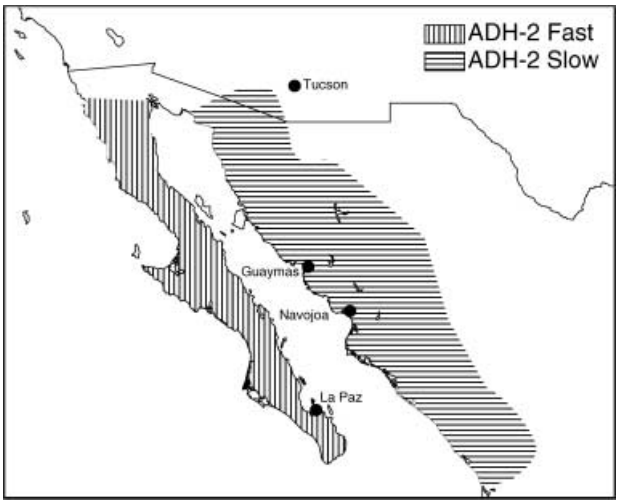

Fig. 2 Distribution of *Drosophila mojavensis* in Baja California and mainland Mexico (Mojave and Catalina Island populations omitted). The frequency distribution of the ADH-2 allozyme allele are shown; Fast allele found at highest frequency in Baja (> 0.9) while the Slow allele is found at highest frequency in the mainland (> 0.8). Collecting sites of *D. mojavensis* and *Drosophila arizonae* isofemale lines are noted.

Mexico and southwestern United States. Each species has its own set of cactus hosts, utilizing the necrotic tissues, called rots, of the cactus as well as its fruit (mostly by *D. arizonae*). *Drosophila arizonae* tends to be a generalist, primarily utilizing the cina (*Stenocereus alamosensis*) and *Opuntia* cacti (pads and fruit) as its hosts, but has been known to utilize citrus and melon. The cactus hosts of *D. mojavensis* vary with geography, populations in the Baja California peninsula utilize the agria cactus (*Stenocereus gummosus*), mainland desert populations utilize organpipe cactus (*Stenocereus thurberi*), Mojave desert population utilize barrel cactus (*Ferrocactus cylindraceus*), and Catalina Island (southern California) utilize prickly pear (*Opuntia* spp.) (Fellows & Heed 1972; Ruiz & Heed 1988). The alcohol composition of the agria and organpipe rots differs; the agria rots utilized by Baja populations contain a greater proportion of 2-propanol relative to the organpipe rots, which contain a relatively higher concentration of 1-propanol. Interestingly, an allozyme polymorphism in *D. mojavensis* ADH-2 exhibits a pattern of geographical variation that mostly parallels the pattern of host alcohol composition (Heed 1978; Vacek 1979; Kircher 1982).

The ADH-2 Fast allele is found at high frequency (*c.* 0.9) in the Baja California population while the ADH-2 Slow allele is found at highest frequency (*c.* 0.9) in mainland populations (see Fig. 2; Zouros 1973; Heed 1978; Matzkin & Eanes 2003; Matzkin 2004). The maintenance of a high frequency of ADH-2 Fast alleles in Baja California has been associated with a higher 2-propanol vapor tolerance experienced by homozygous Fast individuals (Batterham *et al.* 1982). The pattern of geographical distribution of ADH-2 alleles may thus be related to the ability of each allele to metabolize the different alcohols found in each cactus host species. In prior studies of the sequence variation of the *Adh* paralogs in *D. mojavensis*, it was observed that the change responsible for the Fast/Slow allozyme polymorphism is an arginine to a serine change at residue 28 (Matzkin & Eanes 2003; Matzkin 2004). Unlike ADH-2, ADH-1 lacks amino acid variation in *D. mojavensis* (Matzkin & Eanes 2003; Matzkin 2004). At ADH-1, the pattern of variation suggests that it has gone through an episode of adaptive protein evolution; possibly associated with the cactus host and alcohol environment shift that occurred during the divergence of *D. mojavensis* and *D. arizonae* (Matzkin & Eanes 2003; Matzkin 2004). In these same studies it was evident that the duplication at *Adh* created two loci that have gone through different evolutionary trajectories.

The goal of this study is to determine if the functional variation in the ADH paralogs are in congruence with the known ecology of these species and the pattern of intra- or interspecific sequence variation found at *Adh*. There are three main questions addressed by this study: (i) Are there functional differences between the *D. mojavensis* allozyme alleles? (ii) Are there functional differences associated with the pattern of adaptive protein evolution found in *D. mojavensis Adh-1*? (iii) Does the pattern of variation support the idea that the *Adh* paralogs in *D. mojavensis* and *D. arizonae* have functionally diverged? The results of this study will be placed in the context of the known ecology of *D. mojavensis* and *D. arizonae*.

Materials and methods

Isofemale lines

Drosophila mojavensis and *Drosophila arizonae* isofemale lines were all collected from rotten cacti in the field (Matzkin & Eanes 2003; Matzkin 2004). In this study, two populations

of *D. mojavensis* isofemale lines were compared: one collected in La Paz, Baja California (MJBC) and a mainland population collected from Guaymas, Mexico (MJS) (see Fig. 2). Three populations of *D. arizonae* isofemale lines were used, collected from Tucson, Arizona (ARTU), Guaymas, Mexico (AR00), and Navojoa, Mexico (ARNA) (see Fig. 2). Isofemale lines were maintained at 25 °C in a 14:10 light:dark cycle. All stocks were reared in standard banana-molasses food sprinkled with a few granules of live bakers yeast.

D. mojavensis *ADH-2 Fast and Slow activity*

Differences between the *D. mojavensis* ADH-2 Fast and Slow allozyme alleles were determined using the Baja California (25 isofemale lines) and mainland (22 isofemale lines) populations. Assay flies were reared at constant egg densities (50–60 eggs per vial) and aged in mixed sex vials (*c*. 20 flies per vial) for 14 days prior to assay. After the 14-day ageing period, males were frozen and stored at –80 °C. Only males were used because adult males only express ADH-2 (Batterham *et al*. 1983). Two replicates of one male per line were used to determine ADH activity. To determine genotype at *Adh-2*, males were decapitated and their heads were used for quick genomic DNA preps using the squishing buffer method of Gloor & Engels (1992). The remaining body of the fly was stored at –80 °C. Genotypes were determined by PCR amplifying a region of *Adh-2* containing the site producing the Fast/Slow change and using a restriction enzyme (*Bsr*BI) that cuts the Slow allele. Only homozygous individuals were used to measure activity because it is found in this state in the majority of the time (80%) in each population. The concentrations of NAD + and the alcohols were optimized to measure ADH activity at its highest rate (V_{max}). As described in (Fell 1997), the two-substrate concentration method (at 0.2 and 0.8 M) of Cornish-Bowden (1995) was utilized to determined V_{max}. Each sample was homogenized in 125 µL of phosphate buffer (25 mM KH_2PO_4, pH 7.4) and 40 µL of each sample was assayed in 500 µL of reaction buffer. ADH activity on two substrates (1-propanol and 2-propanol) was measured. The reaction buffer was composed of 25 mM KH_2PO_4, 4 mM NAD+ and 0.8 M alcohol. All measurements were performed at 25 °C in a Beckman DU 640 Spectrophotomer (Beckman Coulter). Absorbance was measured at 340 nm every 12 s for 4 min (activity reduction of NAD+/s). Soluble protein content was measured from 25 µL of every homogenate in 1.25 mL of Bio-Rad Protein Assay (Bio-Rad) and calculated using four BSA (bovine serum albumin) standards.

Activity differences between ADH paralogs

Both *D. mojavensis* and *D. arizonae* were used to determine activity differences between ADH-2 and ADH-1. Eight MJBC and eight MJS *D. mojavensis* isofemale lines and eight ARTU, eight AR00, and 10 ARNA *D. arizonae* isofemale lines were utilized. Flies for assays were reared at a constant egg density (40–45 eggs/vial). After eclosion, flies were kept in vials at low densities (*c.* 20 flies per vial) in equal sex ratios. Flies were aged for 14 days and then frozen and stored at –80 °C.

The tissue-specific expression of the ADH paralogs was exploited to calculate the activity of each enzyme. For each line, four types of sample preparations were used to examine the activity of each paralog independently and in combination. The four sample types were males (ADH-2 only expressed), females (ADH-1 and ADH-2 expressed), ovaries (ADH-1 only), and carcasses (no ovaries) (ADH-2 only). All dissections were carried out on ice. Each sample contained four flies or four ovaries. Samples were randomized prior to homogenization and assayed in that order. Sample homogenizations were performed the same day as the assay in groups of 40 (10 lines × 4 sample types per line). The samples containing the four ovaries were homogenized in 300 µL of phosphate buffer (25 mM KH_2PO_4, pH 7.4), and 50 µL of homogenate used per 500 µL assay. The bodies and carcasses were homogenized in 1 mL phosphate buffer, and 25 µL of homogenate was used per 500 µL assay. Activity and soluble protein content were measured as described above.

Activity data analysis

Analysis of activity was performed on each substrate separately [$\log_2$(activity/soluble protein)] and as a $\log_2$-ratio [$\log_2$(activity on 2-propanol/activity on 1-propanol); this will be referred to as the log-ratio]. The lines used were not controlled for genetic background, which could have an effect on the amount of enzyme produced. Because of this, the focus of this study will be on log-ratios. By analysing the data as a log-ratio, it removes the effects of possible expression level differences between alleles. Statistical analyses of activity ratios were performed using JMP version 4.0 (SAS Institute Inc.) and BIOMSTAT version 3.30j (Applied Biostatistic Inc.).

Results

Activity of ADH-2 Fast and Slow allozyme alleles

The mean log-ratio for ADH-2 Fast (FF) males was 1.63 (SE = 0.033) while that of Slow (SS) males was 1.25 (SE = 0.022). Given the fact that in each population one allozyme allele was prevalent, a within-population analysis of the differences between Fast and Slow alleles could not be performed. Therefore, a nested ANOVA was performed pooling all the Fast alleles together and all the Slow alleles together (genotype as fixed effect). The ANOVA indicated a large significant difference in the log-ratio between the genotypes (Table 1). The different genotypes accounted

Table 1 Activity log-ratio differences between ADH-2 Fast and Slow alleles

Level	d.f.	MS	*F*	Variance component (%)
Genotypes	1	3.906	81.7***	70.5
Line	53	0.0478	4.2***	18.2
Within	55	0.0114		11.3

Nested ANOVA for 2-propanol/1-propanol activity log-ratio.
***$P < 0.001$.

Table 2 Activity log-ratios for the different sample types assayed

	D. mojavensis	*D. arizonae*
Females (ADH-2/ADH-1)	1.72 (0.13)	1.53 (0.04)
Ovaries (ADH-1)	2.10 (0.11)	1.86 (0.06)
Carcass (ADH-2)	1.56 (0.15)	1.48 (0.05)
Males (ADH-2)	1.62 (0.07)	1.38 (0.02)

Mean and standard error (in parentheses) of 2-propanol/1-propanol activity log-ratios of *Drosophila mojavensis* and *Drosophila arizonae* for all the sample types assayed.

for about 70.5% of the total variation in activity ratio. Although there was a significant line effect, it only accounted for 18.2% of the variation. The Fast allele had a significantly higher activity (1.38 times greater) than the Slow allele for 2-propanol (*t*-test, $P < 0.001$). No activity difference between the alleles was observed when utilizing 1-propanol as a substrate (*t*-test, $P = 0.67$). Hence, the increase in log-ratio of ADH-2 Fast males is largely due to an increase in the 2-propanol activity of the Fast allele.

Activity differences between paralogs

The log-ratios for all sample types assayed are shown in Table 2. For *Drosophila mojavensis* and *Drosophila arizonae*, the log-ratio was lowest for the male and carcass samples, and highest for ovaries, while females had an intermediate ratio (Table 2). Given the known expression pattern, the carcass and male samples should reflect activity of the same paralog (ADH-2). The log-ratio of males and carcasses were not significantly different for either *D. mojavensis* (*t*-test, $P = 0.69$) or *D. arizonae* (*t*-test, $P = 0.10$). A two-way ANOVA for each species was used to examine activity differences between the paralogs and populations (Tables 3 and 4). The analysis was performed utilizing only the ovaries and carcass data to reduce the variance because such tissues were extracted from the same set of flies. In *D. mojavensis* (Table 3), there was a highly significant difference between paralogs and also between the two populations, as might have been expected given the significant differences between the Fast and Slow allozyme alleles, while the interaction

Table 3 Activity log-ratio differences between ADH-2 and ADH-1 in *D. mojavensis*

Source	d.f.	MS	*F*
Populations	1	1.75	11.0**
Paralogs	1	2.91	18.4***
Populations * paralogs	1	0.50	3.18n.s.
Within	27	0.19	

Two-way ANOVA for 2-propanol/1-propanol activity log-ratio.
$P < 0.01$,*$P < 0.001$, n.s. (not significant).

Table 4 Activity log-ratio differences between ADH-2 and ADH-1 in *Drosophila arizonae*

Source	d.f.	MS	*F*
Populations	2	0.16	1.95n.s.
Paralogs	1	1.84	21.9***
Populations * paralogs	2	0.04	0.49n.s.
Within	47	0.08	

Two-way ANOVA for 2-propanol/1-propanol activity log-ratio.
***$P < 0.001$, n.s. (not significant).

term was not significant. In *D. arizonae*, there was only a significant difference between the paralogs (Table 4). As shown in Table 2 the direction of the difference between the paralogs (ADH-1 > ADH-2) is the same in both species.

Activity differences between the species

Activity differences between the species within a given paralogs were analysed. The log-ratio of *D. mojavensis* ADH-1 is significantly greater (Table 2) than that of *D. arizonae* (*t*-test, $P = 0.002$). The activity of ADH-1 on 2-propanol was significantly greater than that of ADH-2 in both *D. mojavensis* (2.35 times; *t*-test, $P < 0.001$) and *D. arizonae* (1.72 times; *t*-test, $P < 0.001$). Similarly, although not to as great an extent, utilizing 1-propanol as a substrate produced a significantly higher activity of ADH-1 in *D. mojavensis* (1.61 times; *t*-test, $P < 0.01$) and *D. arizonae* (1.30 times; *t*-test, $P = 0.025$). Analysis of differences between the species at ADH-2 can be carried out utilizing the data from both comparisons (males and carcasses), although the first comparison (males) will have more power than the second (carcasses) given the greater sample size of the former. Differences between the *D. mojavensis* alleles and *D. arizonae* are shown in Fig. 3. The *D. mojavensis* ADH-2 Fast allele has a significantly greater (*t*-test, $P < 0.001$) log-ratio than *D. arizonae*, while the Slow allele is significantly lower (*t*-test, $P < 0.001$). An identical result is observed with the carcass data set (data not shown).

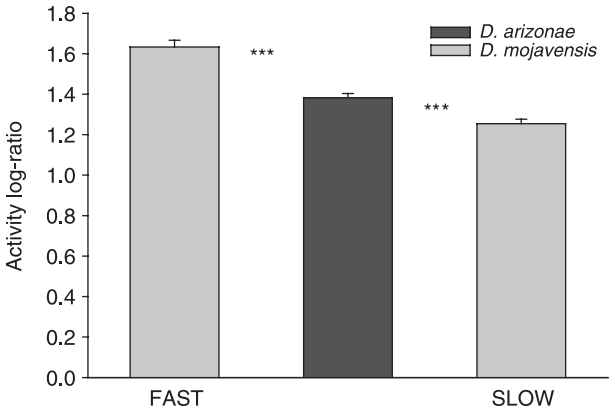

Fig. 3 The mean 2-propanol/1-propanol activity log-ratio of ADH-2 (with standard error bars) from male samples in *Drosophila mojavensis* and *Drosophila arizonae*. The activity of both ADH-2 allozyme alleles (Fast and Slow) are shown. Significance (*t*-test; ***$P < 0.001$) between adjacent bars is noted.

Discussion

In this study, ADH activity differences within and between paralogs were measured. Activity and substrate specificity differences were observed between *Drosophila mojavensis* ADH-2 Fast (highest frequency in Baja California) and Slow alleles (highest frequency in the mainland). These differences match the environmental differences experienced by Baja and mainland populations. Changes in ADH-1 activity in *D. mojavensis*, further supports previous observed pattern of adaptive protein evolution occurring at that locus. There are significant differences in activity between the paralogs associated with amino acid changes, suggesting that following the duplication each paralog has diverged and adapted to function in its environment of expression, i.e. subfunctionalization.

Activity differences between ADH-2 alleles

Cactophilic *Drosophila* have adapted to the alcohol environment offered by its cactus hosts (Brazner *et al*. 1984; Etges 1989; Ganter *et al*. 1989). Ruiz *et al*. (1990) proposed that the mainland populations of *D. mojavensis* are a product of a colonization event from Baja California. This is supported by the observed pattern of derived *Adh-1* alleles in mainland populations (Matzkin 2004). Associated with the colonization of mainland deserts has been a shift in cactus host use. In Baja California, the agria cactus (*Stenocereus gummosus*) is the primary cactus host of *D. mojavensis* (Fellows & Heed 1972; Ruiz & Heed 1988), while the mainland populations tend to utilize the organpipe cactus (*Stenocereus thurberi*). Agria cactus rots have a relatively greater quantity of 2-propanol than 1-propanol and conversely, organpipe contains a higher proportion of 1-propanol (Heed 1978; Vacek 1979; Kircher 1982). Therefore, Baja California and mainland populations of *D. mojavensis* experience a different alcohol environment and the two ADH-2 alleles are associated with the different alcohol environments. The work reported here shows that each allozyme allele tends to have a greater activity (Table 1 and Fig. 3) in the environment in which it is found at highest frequency: Fast allele in Baja and Slow allele in the mainland (Zouros 1973; Heed 1978; Matzkin & Eanes 2003; Matzkin 2004). In each population, selection appears to maintain the allele that can more rapidly remove the most abundant alcohol compounds found in the cactus host. It is expected that the larval expressed ADH-1, like ADH-2, would experience the same differences in alcohol environment between Baja California and mainland. However, no replacement polymorphisms were found segregating in either population at *Adh-1*. In fact, given the pattern of adaptive protein evolution found in both populations of *D. mojavensis*, *Adh-1* may be under a distinct type of selection (Matzkin & Eanes 2003; Matzkin 2004).

The site responsible for the Fast/Slow allozyme polymorphism is an arginine to serine change at residue 28 (Matzkin & Eanes 2003; Matzkin 2004). This residue does not appear to be located in a functionally important portion of the molecule (Benach *et al*. 1998, 2000), although this does not necessarily mean that it can not have an effect on activity (see Chen *et al*. 1994). Four additional amino acid polymorphisms (residues 98, 115, 136 and 236) are associated, but not in complete linkage disequilibrium, with the Fast allozyme allele in the Baja population. Similarly to residue 28, these amino acid polymorphisms do not appear to occur in a functionally important region of the enzyme. The assumption that only active site residues can cause activity changes is too conservative, therefore in the absence of a mutagenesis study, it is not possible at this time to determine which of the polymorphic residue or residues are responsible for the substrate specificity differences between the Fast and Slow allozyme alleles.

The previous *Adh-2* sequence survey revealed that there are no fixed amino acid differences between *D. mojavensis* and *Drosophila arizonae* (Matzkin & Eanes 2003; Matzkin 2004). Although there are no fixed differences, there is not an instance where both species share the same identical protein haplotype. A common feature among all the *D. arizonae* haplotypes is that they are fixed for a serine at residue 28 (the Slow amino acid polymorphism in *D. mojavensis*). Hence, if residue 28 is the factor responsible for the *D. mojavensis* Fast/Slow differences, then it might be expected that the activity ratio of *D. arizonae* would be similar to that of the Slow allele. As illustrated in Fig. 3 this is not necessarily true. The activity ratio of *D. arizonae* is between the activity of both *D. mojavensis* alleles. A complicating factor is that within each species, there are different amino acid polymorphisms segregating. A possible solution would be to examine a common ancestor (*Drosophila navojoa*), but

because amino acid changes that increase or decrease activity ratio might totally mask prior levels (i.e. non-additivity), determining ancestral activity may not be possible and/or informative.

Earlier work further supports the notion that different alcohol environments may ultimately select for different ADH-2 alleles. Batterham *et al*. (1982) showed that the *D. mojavensis* ADH-2 Fast allele is associated with a higher resistance to 2-propanol vapor. It appears that the higher the ADH activity on 2-propanol (ADH-2 Fast), the faster alcohols are metabolized, increasing the resistance to 2-propanol vapors. This study supports the idea that ADH has played an important role in the adaptation of *D. mojavensis* to its cactus hosts. Although this study cannot rule out the role of differential expression associated with cactus host shifts. The two cactus rots differ by more than just 2-propanol and 1-propanol; it is probable that other pathways and other members of the alcohol metabolism pathways could be involved in the adaptation to cactus host.

Functional and molecular divergence of ADH-1

The log-ratio for ADH-1 significantly differs between the species (*D. mojavensis* > *D. arizonae*). Previous studies suggested that the pattern of variation observed at *Adh-1* is consistent with that of adaptive protein evolution (Matzkin & Eanes 2003; Matzkin 2004). More specifically, we observed a significant excess of amino acid substitutions in the *D. mojavensis* lineage, and that two out of the three amino acid fixations in *D. mojavensis* occur in a functionally important region of the enzyme. The natural history of *D. mojavensis* further supports the idea of both molecular and functional adaptive evolutionary change in the *Adh-1* locus. *Drosophila mojavensis* and *D. arizonae* utilize different cactus hosts and likely the alcohol composition of the cactus rots are distinct. In several groups of insects, host shifts have frequently been associated with speciation, such as in phytophagous insects (Funk *et al*. 1995; Futuyma *et al*. 1995). It has been suggested that the host shift and species divergence between *D. mojavensis* and *D. arizonae* occurred in allopatry (Ruiz *et al*. 1990). The interspecific difference in substrate specificity of ADH-1 could be a response to the cactus host shift that occurred between the species.

It would be expected that loci expressed in larvae would be greatly affected by a cactus host shift, as both species spend all of pre-adult stages imbedded in the necrotic tissues of the cactus host. The pattern of adaptive protein evolution at *Adh-1* in *D. mojavensis* following the divergence from *D. arizonae* is clear. But why don't we see amino acid variation between Baja California and mainland populations of *D. mojavensis* given their different host use? One obvious answer is that selection might be on expression and not necessarily at the protein level (King & Wilson 1975). For example, an intron polymorphism affecting expression in *Drosophila melanogaster Adh* explains more of the clinal variation than the Fast/Slow amino acid polymorphism (Berry & Kreitman 1993; Laurie & Stam 1994). Alternatively, expression of *Adh-1* might be sufficiently high in the larvae that the differences in host use in *D. mojavensis* might not be a factor. Finally, activity differences in the *D. mojavensis* ovaries might not be reflective of expression differences in larvae, therefore an analysis of expression level differences in larvae between the *D. mojavensis* populations is needed.

Divergence of the Adh *paralogs*

One classical model of paralog evolution assumes that following a duplication event there is a period of relaxed selection in which a relatively large number of amino acid polymorphisms are tolerated (Ohno 1970; Ohta 1987, 1989). In contrast, more recent models have focused on change in the regulation and expression of paralogs as a major force in their maintenance (Piatigorsky & Wistow 1991; Hughes 1994; Force *et al*. 1999). On such model is the duplication–degeneration–complementation (DDC) model proposed by Force *et al*. (1999). The DDC model is based on the assumption that the ancestral locus had a developmentally complex spatio-temporal pattern of expression. Following the duplication, degenerative mutations in the regulatory regions will tend to create paralogs with no overlap of expression or complementarity, providing the possibility for each paralog to respond independently to natural selection (i.e. subfunctionalization). Because each paralog will be released from the constraints of being expressed in all tissues or developmental times, each paralog can specialize in function with respect to its environment of expression (Lynch & Force 2000). Another possible fate for duplicated genes is neofunctionalization, which refers to when one or both copies acquire a new function (expression) due to mutations in the regulatory region (Force *et al*. 1999). Currently, there is no evidence suggesting the existence of a new expression pattern for either of the *Adh* paralogs, but yet the occurrence of neofunctionalization can not completely be ruled out. Among previous studies (De Martino *et al*. 2000; Bruce *et al*. 2001; Chong *et al*. 2001; Sauka-Spengler *et al*. 2001; Altschmied *et al*. 2002; Gibert 2002; Yan *et al*. 2002), there are no examples of subfunctionalization of relatively recent duplicated metabolic genes. The model proposed by Hughes (1994) is similar in the sense that it assumes that the ancestral locus was 'gene sharing' (performing two distinct functions). According to Hughes (1994), following duplication positive selection at the coding region in one or both paralogs produces two functionally distinct loci. In this study, it was observed that ADH-1 and ADH-2 in *D. mojavensis* and *D. arizonae* have diverged at both the functional and molecular level. In both species, the log-ratio is significantly greater for ADH-1 than ADH-2.

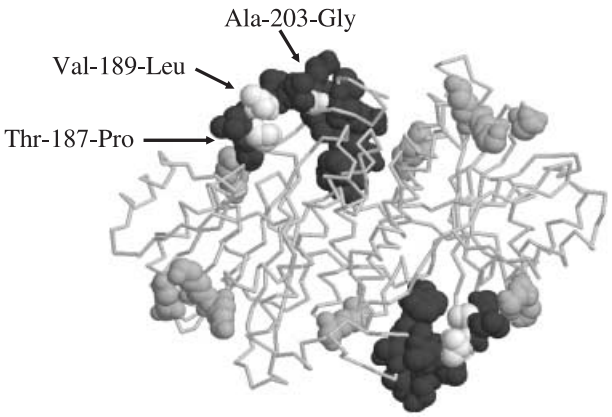

Fig. 4 Three-dimensional structure of ADH with residues forming the active site pocket labelled in black and white. Gray residues are fixed differences between the paralogs, while the labelled white residues are fixed differences that occur in the active site of the enzyme.

The amino acid changes that have occurred since the duplication of *Adh* are largely the cause for the observed difference in the activity ratio, which suggests that positive selection might have played a role in the divergence of the paralogs.

The duplication that produced *Adh-2* and *Adh-1* occurred about 4 Ma (Matzkin & Eanes 2003; Matzkin 2004). From that time until the divergence between *D. mojavensis* and *D. arizonae* (about 2 Ma), there have been nine amino acid fixations between ADH-2 and ADH-1 (Fig. 4). The enzyme is 254 amino acids in length, 16 of which comprise the active site pocket (Benach *et al*. 1998, 1999). Three of the nine differences occur in the residues forming the active site pocket. Overall, 18.8% of the active site pocket residues changed, while only 2.5% of the residues outside the active pocket of ADH changed. This difference in the ratio of fixations is significant (Table 5). Residues 186–191 act as the gate of the ADH active site (Benach *et al*. 1999). Two of the three active site differences observed between ADH-1 and ADH-2 occur in residues 187 and 189. The Thr187Pro change occurring between ADH-2 and ADH-1 is also present when comparing the *D. lebanonensis* and *D. melanogaster* sequences. This change in *D. lebanonensis* has been hypothesized to decrease enzyme activity by affecting the flexibility of the gate (Benach *et al*. 2000). The third change (residue 203) is in the substrate binding portion of ADH (Benach *et al*. 1999). Although the three active site pocket changes could be the cause of the log-ratio differences between the paralogs, it is not the only potential explanation, as changes at non-active site residues have been shown to affect ADH activity (Chen *et al*. 1994).

Table 5 Differences in the number of fixed amino acids between paralogs in non-active and active site residues of ADH

	Non-active site	Active site
Fixed residues	6	3
Total residues	238	16

Significance was tested using *G*-test, $G = 5.547$, $P < 0.05$. Residues were assigned to the active site pocket based on the *Drosophila lebanonensis* ADH structure (Benach *et al*. 1998; Benach *et al*. 1999).

A prerequisite for either Hughes (1994) or the DDC model is the need for the ancestral gene to have multiple functions or a complex spatio-temporal pattern of expression. There is no information on the preduplication pattern of expression of the ancestral *Adh* locus in the *mojavensis*-species cluster, although the notion of multiple patterns of expression in the preduplication *Adh* is supported by work in *D. melanogaster*. In *D. melanogaster*, the *Adh* locus produces two types of transcript, a larval (proximal) and adult (distal) transcript (Benyajati *et al*. 1983), which by themselves resemble the expression of *Adh-1* and *Adh-2*, respectively (Savakis *et al*. 1986). The pattern of expression of distal and proximal *Adh* transcript may be a common feature in *Drosophila* (Powell 1997). It can be speculated that the expression pattern of the preduplication *Adh* in the *mojavensis* species cluster may have been the sum of the expression patterns of *Adh-2* and *Adh-1* seen in *D. mojavensis* and *D. arizonae*.

Differences in the spatio-temporal environment of each paralog may lead to molecular and functional divergence. Experimental evidence suggests that the cellular environment can potentially alter the strength of natural selection on enzymes and the control of metabolic pathways (Dykhuizen & Hartl 1980; Hartl *et al*. 1985; Dykhuizen *et al*. 1987). Metabolic control theory states that members of a pathway will have a certain level of control of flux, measured as a control coefficient (Fell 1997). Dramatically different estimates of the control coefficients have been calculated for ADH in larvae and adults of *D. melanogaster* (Middleton & Kacser 1983; Freriksen *et al*. 1991), suggesting that the selective environment experienced by ADH in larval and adult tissues could be very distinct (Freriksen *et al*. 1994).

Building upon previous work, this study further illustrates that given a locus' function, position in a pathway and its relation to the ecology of an organism, a locus can provide foci for evolutionary change. The *Drosophila* alcohol dehydrogenase locus is one such example. Its specialization can be observed at many levels: between population, between species and between paralogs. We can observe changes both at the molecular and functional level, which can be associated to environment change. The fact that independent duplications of *Adh* are maintained throughout *Drosophila* hints at the evolutionarily multifaceted and variable role of this locus.

Acknowledgements

I thank Walter F. Eanes for his assistance and support in this project. I also thank Therese Markow for assistance in the field and for providing the *Drosophila arizonae* isofemale lines. Additionally, I thank Robert Sokal, John True, Thomas Merritt, Allen Gibbs and anonymous reviewers for providing productive comments on the manuscript. This research was supported by a NSF Predoctoral Fellowship and a W. Burghardt Turner Fellowship to L.M.M. and in part by U.S. Public Health Service Grant GM-45247 to W.F.E. This is contribution number 1113 from the Graduate Program in Ecology and Evolution, State University of New York at Stony Brook.

References

Altschmied J, Delfgaauw J, Wilde B *et al.* (2002) Subfunctionalization of duplicate mitf genes associated with differential degeneration of alternative exons in fish. *Genetics*, **161**, 259–267.

Atkinson PW, Mills LE, Starmer WT, Sullivan DL (1988) Structure and evolution of the *Adh* genes of *Drosophila mojavensis*. *Genetics*, **120**, 713–723.

Atrian S, Sanchez-Pulido L, Gonzàlez-Duarte R, Valencia A (1998) Shaping of *Drosophila* alcohol dehydrogenase through evolution: relationship with enzyme functionality. *Journal of Molecular Evolution*, **47**, 211–221.

Batterham P, Starmer WT, Sullivan DL (1982) Biochemical genetics of alcohol longevity response of *Drosophila mojavensis*. In: *Ecological Genetics and Evolution: The Cactus-Yeast-Drosophila Model System* (eds Barker JSF, Starmer WT), pp. 307–322. Academic Press, New York.

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

Benach J, Atrian S, Gonzàlez-Duarte R, Ladenstein R (1998) The refined crystal structure of *Drosophila lebanonensis* alcohol dehydrogenase at 1.9 å resolution. *Journal of Molecular Biology*, **282**, 383–399.

Benach J, Atrian S, Gonzàlez-Duarte R, Ladenstein R (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. *Journal of Molecular Biology*, **289**, 335–355.

Benach J, Atrian S, Fibla J, Gonzàlez-Duarte R, Ladenstein R (2000) Structure-function relationships in *Drosophila melanogaster* alcohol dehydrogenase allozymes ADH (S), ADH (F) and ADH (UF), and distantly related forms. *European Journal of Biochemistry*, **267**, 3613–3622.

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

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

Brazner J, Aberdeen V, Starmer WT (1984) Host-plant shifts and adult survival in the cactus breeding *Drosophila mojavensis*. *Ecological Entomology*, **9**, 375–381.

Bruce AEE, Oates AC, Prince VE, Ho RK (2001) Additional hox clusters in the zebrafish: divergent expression patterns belie equivalent activities of duplicate *hoxB5* genes. *Evolution & Development*, **3**, 127–144.

Chen Z, Tsigelny I, Lee WR, Baker ME, Chang SH (1994) Adding a positive charge at residue 46 of *Drosophila* alcohol dehydrogenase increases cofactor specificity for NADP+. *FEBS Letters*, **356**, 81–85.

Chong SW, Emelyanov A, Gong ZY, Korzh V (2001) Expression pattern of two zebrafish genes, *cxcr4a* and *cxcr4b*. *Mechanisms of Development*, **109**, 347–354.

Cornish-Bowden A (1995) *Fundamentals of Enzyme Kinetics*, 2nd edn. Portland Press, London.

De Martino S, Yan YL, Jowett T *et al.* (2000) Expression of *sox11* gene duplicates in zebrafish suggests the reciprocal loss of ancestral gene expression patterns in development. *Developmental Dynamics*, **217**, 279–292.

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

Dykhuizen DE, Dean AM, Hartl DL (1987) Metabolic flux and fitness. *Genetics*, **115**, 25–31.

Eanes WF (1999) Analysis of selection on enzyme polymorphisms. *Annual Review of Ecology and Systematic*, **30**, 301–326.

Etges WJ (1989) Divergence in cactophilic *Drosophila*: the evolutionary significance of adult ethanol metabolism. *Evolution*, **43**, 1316–1319.

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

Fellows DF, Heed WB (1972) Factors affecting host plant selection in desert-adapted cactiphilic *Drosophila*. *Ecology*, **53**, 850–858.

Force A, Lynch M, Pickett FB *et al.* (1999) Preservation of duplicate genes by complementary, degenerative mutations. *Genetics*, **151**, 1531–1545.

Freriksen A, Seykens D, Heinstra PWH (1994) Differences between larval and adult *Drosophila* in metabolic degradation of ethanol. *Evolution*, **48**, 504–508.

Freriksen A, Seykens D, Scharloo W, Heinstra PWH (1991) Alcohol dehydrogenase controls the flux from ethanol into lipids in *Drosophila* larvae: a C-13 NMR-study. *Journal of Biological Chemistry*, **266**, 21399–21403.

Funk DJ, Futuyma DJ, Orti G, Meyer A (1995) A history of host associations and evolutionary diversification for *Ophraella* (Coleoptera, Chrysomelidae): New evidence from mitochondrial DNA. *Evolution*, **49**, 1008–1017.

Futuyma DJ, Keese MC, Funk DJ (1995) Genetic constraints on macroevolution: The evolution of host affiliation in the leaf beetle genus Ophraella. *Evolution*, **49**, 797–809.

Ganter PF, Peris F, Starmer WT (1989) Adult lifespan of cactophilic *Drosophila*: interactions among volatiles and yeasts. *American Midland Naturalist*, **121**, 331–340.

Gibert JM (2002) The evolution of engrailed genes after duplication and speciation events. *Development Genes and Evolution*, **212**, 307–318.

Gloor G, Engels W (1992) Single-fly DNA preps for PCR. *Drosophila Information Service*, **71**, 148–149.

Hartl DL, Dykhuizen DE, Dean AM (1985) Limits of adaptation: The evolution of selective neutrality. *Genetics*, **111**, 655–674.

Heed WB (1978) Ecology and genetics of Sonoran desert *Drosophila*. In: *Ecological Genetics: The Interface* (ed. Brussard PF), pp. 109–126. Springer-Verlag, New York.

Hughes AL (1994) The evolution of functionally novel proteins after gene duplication. *Proceedings of the Royal Society of London. Series B,Biological Sciences*, **256**, 119–124.

King M-C, Wilson AC (1975) Evolution at two levels in humans and chimpanzees. *Science*, **188**, 107–116.

Kircher HW (1982) Chemical composition of cacti and its relationship to Sonoran Desert *Drosophila*. In: *Ecological Genetics and Evolution: The Cactus-Yeast-Drosophila Model System* (eds Barker JSF, Starmer WT), pp. 143–158. Academic Press, New York.

Laurie CC, Stam LF (1994) The effect of an intronic polymorphism on alcohol dehydrogenase expression in *Drosophila melanogaster*. *Genetics*, **138**, 379–385.

Lynch M, Force A (2000) The probability of duplicate gene preservation by subfunctionalization. *Genetics*, **154**, 459–473.

Matzkin LM (2004) Population genetics and geographic variation of alcohol dehydrogenase (*Adh*) paralogs and glucose-6-phosphate dehydrogenase (*G6pd*) in *Drosophila mojavensis*. *Molecular Biology and Evolution*, **21**, 276–285.

Matzkin LM, Eanes WF (2003) Sequence variation of alcohol dehydrogenase (*Adh*) paralogs in cactophilic *Drosophila*. *Genetics*, **163**, 181–194.

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

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

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

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

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

Piatigorsky J, Wistow G (1991) The recruitment of crystallins: new functions precede gene duplication. *Science*, **252**, 1078–1079.

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

Ruiz A, Heed WB (1988) Host-plant specificity in the cactophilic *Drosophila mulleri* species complex. *Journal of Animal Ecology*, **57**, 237–249.

Ruiz A, Heed WB, Wasserman M (1990) Evolution of the Mojavensis cluster of cactophilic *Drosophila* with descriptions of two new species. *Journal of Heredity*, **81**, 30–42.

Sauka-Spengler T, Baratte B, Shi DL, Mazan S (2001) Structure and expression of an *Otx5*-related gene in the dogfish *Scyliorhinus canicula*: evidence for a conserved role of *Otx5 and Crx* genes in the specification of photoreceptors. *Development Genes and Evolution*, **211**, 533–544.

Savakis C, Ashburner M, Willis JH (1986) The expression of the gene coding for alcohol dehydrogenase during the development of *Drosophila melanogaster*. *Developmental Biology*, **114**, 194–207.

Vacek DC (1979) *The microbial ecology of the host plants of* Drosophila mojavensis. PhD thesis, University of Arizona, Tucson.

Yan YL, Miller CT, Nissen R *et al.* (2002) A zebrafish *sox9* gene required for cartilage morphogenesis. *Development*, **129**, 5065–5079.

Zouros E (1973) Genic differentiation associated with the early stages of speciation in the mulleri subgroup of *Drosophila*. *Evolution*, **27**, 601–621.

This study is the last section of Luciano Matzkin's doctoral work at Stony Brook University. Luciano Matzkin is currently a PERT (Postdoctoral Excellence in Research and Teaching) Fellow at the University of Arizona. He is a population geneticist interested in the genetical basis of adaptation. Most of his work focuses on examining patterns of sequence and functional variation in ecologically interesting organisms, that is cactophilic *Drosophila*. He is currently examining the changes in gene expression associated with host shifts in cactophilic *Drosophila*.
