Population Genetics and Geographic Variation of Alcohol Dehydrogenase (Adh) Paralogs and Glucose-6-Phosphate Dehydrogenase (G6pd) in Drosophila mojavensis

Luciano M. Matzkin

Department of Ecology and Evolution, State University of New York at Stony Brook

Populations of Drosophila mojavensis from the deserts of the Baja California peninsula and mainland Mexico utilize different cactus hosts with different alcohol contents. The enzyme alcohol dehydrogenase (ADH) has been proposed to play an important role in the adaptation of Drosophila species to their environment. This study investigates the role of ADH in the adaptation of the cactophilic D. mojavensis to its cactus host. In D. mojavensis and its sibling species, D. arizonae, the Adh gene has duplicated, giving rise to a larval/ovarian form (Adh-1) and an adult form (Adh-2). Studies of sequence variation presented here indicate that the Adh paralogs have followed different evolutionary trajectories. Adh-1 exhibits an excess of fixed amino acid replacements, suggesting adaptive evolution, which could have been a result of several host shifts that occurred during the divergence of D. mojavensis. A 17-bp intron haplotype polymorphism segregates in Adh-2 and has markedly different frequencies in the Baja and mainland populations. The presence of the intron polymorphism suggests possible selection for the maintenance of pre-mRNA structure. Finally, this study supports the proposed Baja California origination of D. mojavensis and subsequent colonization of the mainland accompanied by a host shift.

Introduction

The neutral theory proposes that the majority of the molecular variation segregating at a locus at any one time has no significant adaptive or deleterious effect; hence, it is neutral (Kimura 1981). Central to the prediction of the neutral theory is the assumption that population size and environment are constant. Changing these may drastically affect the fixation rate of segregating variation in a population (Ohta 1992). Duplicated genes with different spatiotemporal patterns of expression offer an interesting opportunity to investigate the role of environmental changes on the molecular evolution of a gene (Force et al. 1999). One such system is the alcohol dehydrogenase (Adh) paralogs in Drosophila mojavensis and D. arizonae.

Drosophila mojavensis and D. arizonae are cactophilic sibling species inhabiting the deserts of Sonora and Baja California. Both species utilize the necrotic tissues of only a few cactus species, and this is highly dependent on geography. In Baja California, D. mojavensis tends to utilize the agria cactus (Stenocereus gummosus), whereas in the mainland deserts of the Sonora, it utilizes the organ-pipe cactus (S. thurberi) as its main host (Fellows and Reed 1972; Ruiz and Reed 1988). Throughout its range, D. arizonae utilizes a different array of hosts, such as the cina (S. alamosensis) and Opuntia cactus (Fellows and Reed 1972; Ruiz and Reed 1988). Necrotic cactus tissues, or cactus rots, from different species contain distinct compositions of alcohol compounds. For example, agría rots contain a fourfold higher 2-propanol concentration than organ-pipe cactus rots (Heed 1978; Vacek 1979; Fogelman 1982; Kircher 1982). Given the close association of the life cycle of cactophilic Drosophila to their cactus hosts, a host shift will have a drastic effect on the environment experienced by the fly, especially the alcohol environment.

Alcohol dehydrogenase (ADH) plays a major role in the detoxification and metabolism of alcohols. The sequence, function, and geographical variation of this enzyme has been extensively studied in Drosophila (Chambers 1988; Heinstra 1993). Drosophila mojavensis and D. arizonae have two functional Adh loci separated by about 3 kb (Atkinson et al. 1988). The two loci are a product of a duplication event that occurred before the divergence of the species, about 3.5 to 4.4 MYA (Matzkin and Eanes 2003). One of the paralogs (Adh-1) is expressed from the egg until the 5-day larva stage, at which time ADH-2 activity can be observed (Batterham et al. 1983). In adults, ADH-1 activity is solely localized to the ovaries, whereas ADH-2 activity is observed in the remaining adult tissues (Batterham et al. 1983). A previous survey (Matzkin and Eanes 2003) of the sequence variation of Adh paralogs suggests that the host shift might have influenced the evolution of the Adh locus. This is based on the observation that the larval expressed paralog, Adh-1, exhibits a pattern of variation consistent with adaptive protein evolution in the D. mojavensis lineage (Matzkin and Eanes 2003). In the D. mojavensis lineage, two out of the three fixed amino acid mutations have the potential to affect the kinetic and substrate specificity properties of ADH (Matzkin and Eanes 2003). The present day distribution of the cactus species (Ruíz, Reed, and Wasserman 1990) suggests that during the divergence of D. mojavensis from D. arizonae, a host shift to the agría cactus occurred, possibly from the cina cactus (Ruíz, Reed, and Wasserman 1990).

The difference in host use between Baja California and mainland population of D. mojavensis suggests that at least one additional host shift has occurred. Drosophila mojavensis is believed to have originated in the Baja California peninsula, with the populations found in the mainland being products of a yet later colonization event (Ruíz, Reed, and Wasserman 1990). The colonization of the mainland by D. mojavensis must have occurred less
than 2.4 MYA, because that is the estimated mean divergence time between *D. mojavensis* and *D. arizonae* (Matzkin and Eanes 2003). Earlier studies of allozyme variation at ADH-2 showed a very distinct pattern of variation between the populations. In Baja California, the ADH-2 Fast allele is found at the highest frequency (about 90%), whereas in the mainland, the Slow allele is found at about 90% (Heed 1978). In addition to the host shift, the colonization of the mainland created *D. mojavensis* populations that were sympatric with *D. arizonae*, although utilizing different cactus hosts. The sympathy is believed to have promoted divergence of the mating system and preferences of *D. mojavensis*, resulting in significant reductions in the viability of F1 hybrids of crosses of *D. mojavensis* from Baja California and mainland strains compared with within-population crosses (Markow 1981; Ruiz, Heed, and Wasserman 1990; Markow 1991). Overall the Adh and life history data suggests that the mainland and Baja California populations of *D. mojavensis* have reduced contact. If true, over time, the subdivision would produce different patterns of neutral variation in each population, which should be observed at many loci.

Presented here is further evidence suggesting the role of ADH in the adaptation to cactus host use in cactophilic *Drosophila*. The changes at *Adh* are consistent with the host shift that occurred during the evolution of *D. mojavensis* from *D. arizonae*. To further investigate the role of ADH in the adaptation to cactus hosts, this study compares patterns of sequence variation at *Adh*-2 and *Adh*-1 between a mainland population of *D. mojavensis* and a previously collected Baja California population data set (Matzkin and Eanes 2003). In addition, to obtain an independent estimate of the level of genetic isolation between the populations, variation at the glucose-6-phosphate dehydrogenase (*G6pd*) gene for the same Baja California and mainland populations of *D. mojavensis* was examined.

**Materials and Methods**

**Isofemale Lines Collections**

Fifty-six isofemale lines from the mainland population of *D. mojavensis* (MJS) were established from a collection outside the town of Guaymas, in Sonora, Mexico. Flies were aspirated from organ-pipe cactus rotts and placed in 8-dram vials containing banana-molasses media. Soon after collection, flies were anesthetized and individual gravid females were placed into a fresh 8-dram vial with banana-molasses media. In the lab, isofemale lines were maintained in a 25°C incubator on a 14:10 h light:dark cycle. Isofemale lines were transferred every 3 to 4 weeks into new banana-molasses food vials sprinkled with a few granules of live yeast.

**Allozyme Survey**

A modified version of the Batterham et al. (1983) protocol for starch gel electrophoresis was used to determine the level of variation at ADH-2 and ADH-1. Six one-fly samples per isofemale line were homogenized individually 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°C through a 12% starch gel at 30 V/cm for 5 h. A 1% agar overlay stain (100 mM TrisHCl pH 8.8; 260 mM 2-propanol; 0.75 mM NAD+; 0.61 mM methylthiazolotetrazolium; 0.03 mM phenazine methosulfate) was used to visualize ADH.

**Sampling**

To be able to obtain the linkage phase of each allele sequenced for *Adh*-2 and *Adh*-1, I created a set of lines that were identical by descent at both loci. After the initial allozyme survey, *D. mojavensis* isofemale lines were produced that were fixed or almost fixed (>80%) for either the Fast or Slow allozyme allele. These lines were then inbred for at least three generations, resulting in final stocks that were fixed for either the Fast (F-stock) or Slow (S-stock) allele. Several individuals (>20) were assayed regularly using starch electrophoresis to make sure no contamination of stock occurred. Two or three males from one isofemale line were placed in a vial with two virgin F-stock females and allowed to mate. A single male from this cross was backcrossed to a virgin F-stock female. All progeny of the second cross were then transferred to a new vial and allowed to mate. This produced a vial in which 1/16 of all flies are expected to be *Adh*-2 S/S. Twenty to 30 adults from each cross were cut in half. Their abdomens were run in a starch gel to determine ADH genotype. The thorax/head of individuals that were determined to be homozygous Slow were saved for DNA extraction. Given the proximity of *Adh*-1 to *Adh*-2 (~3 kb), the *Adh*-1 locus was also expected to be identical-by-descent in these lines. A reciprocal design was used to recover independent Fast alleles.

To avoid biases in sampling, the samples were chosen to reflect the allozyme variation at ADH-2 in the population. The same set of individuals sequenced for *Adh*-2 in this constructed random sample (Hudson et al. 1994) were sequenced for *Adh*-1. For interspecific comparisons, sequences of *D. arizonae* and *D. navojoa* *Adh*-2 and *Adh*-1 were used from a previous study (Matzkin and Eanes 2003). Twelve *D. mojavensis* isofemale lines from each of the mainland and Baja California (MJBC) populations were randomly chosen to survey the variation of exon-4 of *G6pd*. *G6pd* was chosen to remove the need of performing crosses to obtain linkage phase. Because *G6pd* is X-linked, sequencing only males will provide the linkage phase data. In addition, one *D. arizonae* isofemale line was analyzed. The *D. arizonae* *G6pd* sequence was used to conduct a test of independence or a McDonald-Kreitman (MK) test (McDonald and Kreitman 1991) on the *D. mojavensis* *G6pd* data set.

**PCR Amplification and Sequencing**

Genomic DNA preparations of thorax were done using the CTAB method (Winnewinnerckx, Backeljau, and Dewachter 1993). PCR amplification was done in an Air-Thermo-Cycler (Idaho Technologies, Idaho Falls, Idaho) using Gibco BRL (Carlsbad, Calif.) Taq DNA polymerase. The PCR fragments were cleaned using the Prep-A-Gene Kit (Bio-Rad, Hercules, Calif.) before

**Sequence Variation in *D. mojavensis***

6 mM boric acid pH 8.8) and transferred to filter paper.
sequencing. Locus-specific primers were designed from the published *D. mojavensis* Adh-2/Adh-1 sequence (Atkinson et al. 1988). Manual sequencing was performed using the Sequenase Kit version 2.0 (United States Biochemical Co., Cleveland, Ohio) and [\[^{35}S\]] dATP (Amersham, UK). The *Adh* sequences are stored under GenBank accession numbers AY364493 to AY364522.

With the aid of prior *G6pd* studies (Jeffery et al. 1993; Eanes et al. 1996), primers were designed in conserved regions of exon-4 of *G6pd*. A small fragment (~300 bp) of *G6pd* exon-4 from *D. mojavensis* was initially amplified. An inverse PCR technique was used (Triglia, Peterson, and Kemp 1988) to obtain the complete sequence of *G6pd* exon-4. The primers designed for *D. mojavensis* were effective in amplifying *D. arizonae*. The sequencing reactions were performed using the ABI Prism BigDye Cycle Sequencing Kit version 2.0, and reactions were run in an ABI 3100 Genetic Analyzer (Applied Biosystems, Foster City, Calif.). The *G6pd* sequences are stored under GenBank accession numbers AY364481 to AY364492 (MJBC), AY364523 to AY364534 (MJS), and AY364480 (*D. arizonae*).

**Data Analysis**

Descriptive and statistical analysis of the sequence data was produced using SITES (Hey and Wakeley 1997), DnaSP version 3.53 (Rozas and Rozas 1999), and ProSeq version 2.91 (Filatov 2002). The neighbor-joining gene tree using third base positions was created using MEGA version 2.91 (Filatov 2002). The neighbor-joining gene tree was used to estimate evolutionary relationships among populations and to calculate divergence times.

**Results**

**Allozyme Variation**

Six individuals per isofemale line were screened for ADH-2 and ADH-1 allozyme variation. As previously described (Batterham et al. 1983; Matzkin and Eanes 2003), no allozyme polymorphism was observed for ADH-1. The variation observed at ADH-2 was similar to what has been previously described for mainland populations of *D. mojavensis* (Heed 1978). Out of a total of 56 isofemale lines screened, 35 were fixed for the ADH-2 Slow allele, and none were fixed for the Fast allele. Overall, the Slow allele segregated at a frequency of 0.89.

**Sequence Variation at the Alcohol Dehydrogenase Paralogs**

For *Adh*-2, 12 of the 15 randomly sampled alleles were Slow and three were Fast (the same individuals were sequenced for *Adh*-1). There were a total of seven unique polymorphisms in the exons of *Adh*-2, of which five were silent and two were replacement polymorphisms (fig. 1). The two replacement polymorphisms (positions 84 and 347) produced a charge change. Only the change at position 84 (serine to arginine at residue 24) is shared by all Fast alleles sampled in this population and in the Baja California (MJBC) population previously examined (Matzkin and Eanes 2003). The polymorphism at position 347 (tyrosine to histidine at residue 98) was found only once in the mainland population sample. There is a relatively large amount of variation segregating in *Adh*-1 (fig. 1). This variation separates into two haplotypes with 17 differences between them (Matzkin and Eanes 2003). At *Adh*-2, both silent and replacement variation in the mainland population were lower than observed in Baja (table 1). The observed level of variation at *Adh*-1 was lower than *Adh*-2. No replacement and only five silent polymorphisms were observed at *Adh*-1. Unlike *Adh*-2, no variation was found in *Adh*-1 of *D. arizonae*. However, similar to *Adh*-2, the level of silent and replacement variation observed at *Adh*-1 in the mainland was lower than that of the Baja California population (table 1).

**Interspecific and Interparalog Divergence of Adh**

The silent divergence at *Adh* between *D. mojavensis* and *D. arizonae* was calculated as in a previous study (Matzkin and Eanes 2003). For each paralog, the *D. mojavensis* population data set was compared with one *D. arizonae* sequence (ARTU 34; accession numbers AY154844 and AY154861, for *Adh*-2 and *Adh*-1, respectively). The per-site pairwise divergence ($K_{st}$) was 0.104 for *Adh*-2 and 0.079 for *Adh*-1. Moriyama and Gojobori’s (1992) mean rate of silent evolution in *Drosophila* of 1.9 × 10^{-8} site changes per year was utilized to estimate *Adh* divergence times between species. The time of divergence obtained using *Adh*-2 is 2.7 MYr, whereas that for *Adh*-1 is 2.1 MYr. The estimate of $K_{st}$ between *D. mojavensis* and *D. navajo* is 0.125 for *Adh*-2 and 0.157 for *Adh*-1 (divergence time of 3.3 MYr for *Adh*-2 and 4.1 MYr for *Adh*-1). Using the same technique, the time of the *Adh* duplication can be calculated, which is also shared by *D. arizonae* (fig. 3). Given an observed $K_{st}$ of 0.131 between the paralogs, the time since the duplication can be estimated to have occurred 3.4 MYA.

**Recombination and Linkage Disequilibrium in Adh-2 and Adh-1**

The level of recombination was determined by solving for the estimators $C$ (Hudson 1987) and $\gamma$ (Hey and Wakeley 1997). Overall, the value of both estimators for the entire *Adh-2/Adh-1* region was 0.038 and 3.3 for $C$ and $\gamma$, respectively. The estimator $\gamma$ is believed to be less biased than $C$ since $\gamma$ is independent of values of $\theta$, and its estimate is not greatly affected by either low or high numbers of polymorphisms (Hey and Wakeley 1997). The ratio of the estimator over $\theta$ provides the relative magnitude of the rate of recombination ($c$) to the mutation rate ($\mu$). For both estimators the ratio was less than 1. Using $\gamma$, the $c/\mu$ ratio was 0.347, and it was even lower using $C$ (0.004).

Linkage disequilibrium was not evenly distributed across the entire *Adh*-2 and *Adh*-1 region. Disequilibrium was determined by a chi-square test using a Bonferroni correction for multiple comparisons. Out of a total of 435 comparisons (excluding singletons), 113 were significant ($P < 0.0001$) using the Bonferroni correction. Intron-1 sites comprised 105 of those significant comparisons. The remaining eight significant comparisons were between
Distribution of Variation Across the Adh Paralogs

In addition to linkage disequilibrium, the distribution of fixed and polymorphic sites across Adh-2 and Adh-1 was examined. The neutral expectation is that fixed and polymorphic sites will be randomly distributed across a gene region (McDonald 1996). Homogeneity in the distribution of variation across each paralog was tested using McDonald’s (1996) DNA Slider program. There is a significant nonrandom distribution of polymorphisms centered around intron-1 of Adh-2 ($G_{mean} = 7.44$, $P = 0.0007$). Because there were only six polymorphisms throughout the entire Adh-1 region, the $G_{mean}$ statistic could not be calculated. Alternatively, the Kolmogorov-Smirnov statistic was calculated (0.052), which was not significant ($P = 0.83$).

Frequency and Test of Independence Analysis of Adh-2 and Adh-1

The frequency distribution of variants was examined by performing both the Tajima’s $D$ test (Tajima 1989) and the Fu and Li’s $D$ test (Fu and Li 1993). Significant positive values of either test have been associated with the presence of a few highly variable allele classes, whereas negative values tend to imply a large number of low-variation allele classes. For both paralogs, Tajima’s ($D_{Tajima} = 0.553$ and $D_{Tajima} = 1.522$, Adh-2 and Adh-1, respectively) and Fu and Li’s ($D_{Fu&Li} = 1.384$ and $D_{Fu&Li} = 1.354$, Adh-2 and Adh-1, respectively) statistics were not significant.

The MK test examined the expected correlation between polymorphism and fixation (McDonald and Kreitman 1991). The $G$-test or Fisher’s exact test were used to evaluate the significance of the correlation. Drosophila arizonae Adh-2 and Adh-1 sequences from a prior study (Matzkin and Eanes 2003) were used in the test. The use of an outgroup, D. navojoa, can partition the fixed differences into the D. mojavensis and D. arizonae lineages. The MK test was not significant for Adh-2 either with D. mojavensis and D. arizonae combined or by examining the D. mojavensis lineage by itself (table 2).

Additionally, no significant deviation was observed when we added the sequences from the Baja California population of D. mojavensis. The results for Adh-1 were very different. Overall (for D. mojavensis and D. arizonae), there was a significant MK test ($D = 5.818$, $P = 0.016$), although when the data were partitioned to the adjacent sites, and none occurred between the paralogs. The polymorphisms segregating en masse in intron-1 of Adh-2 are largely the cause of the amount of linkage disequilibrium observed.
D. mojavensis lineage, a not statistically significant trend towards an excess of replacement fixations was observed (table 3). Adding the Baja population data produces a significant \((P < 0.02)\) deviation of D. mojavensis Adh-1 from the expected correlation between polymorphic and fixed sites. In D. mojavensis, two of the fixed replacement sites are adjacent to each other (positions 83 and 84), but considering this pair as either one or two mutational events did not have an effect on the analysis (table 3).

Variation at G6pd
The sizes of D. mojavensis and D. arizonae G6pd exon-4 were identical \((1.077 \text{ bp})\) to D. melanogaster \((1.077 \text{ bp})\) (Eanes, Kirchner, and Yoon 1993). There were four replacement polymorphisms at positions 476, 508, 880, and 944 (residues Cys159Ser, Thr170Ala, Ser294Thr, and Arg315His); two \((476 \text{ and } 508)\) were singletons (fig. 4). Given the crystal structure of human G6PD, none of the replacement polymorphisms in D. mojavensis occur at or in the vicinity of a functional important region of the enzyme (Au et al. 2000). Two of the polymorphisms occur in a relatively variable region \(<50\% \text{ amino acid similarity}\), whereas two occur in a moderately conserved region \(51\% \text{ to } 71\% \text{ amino acid similarity}\) of the enzyme, as determined by a phylogenetic study of G6PD from 52 taxa (Notaro, Afolayan, and Luzzatto 2000). No fixed differences between the populations were observed, although several polymorphisms were unique to each population. Because G6pd is located in the X chromosome, to allow for comparison with the autosomal Adh paralogs, the estimates of diversity \((\theta \text{ and } \pi)\) in table 4 were adjusted by \(4/3\) to account for the expected lower effective population size of X-linked loci. The levels of silent variation in each D. mojavensis population were similar (table 4). The only difference observed was that all replacement polymorphisms were present in the mainland population.

Recombination and Linkage Disequilibrium in G6pd
The level of recombination was computed for each individual population. Both estimates of the \(c/\mu\) ratio using C (Hudson 1987) or \(\gamma\) (Hey and Wakeley 1997) were higher for Baja \((38.6 \text{ and } 6.92, \text{ respectively})\) than for the mainland \((16.9 \text{ and } 2.23, \text{ respectively})\). A similar difference between populations was observed in the level of linkage disequilibrium. Excluding singletons, none of the 30 pairwise comparisons were significant \((\text{chi-square test with a Bonferroni correction})\) in the Baja population, whereas 4 out of 45 comparisons were significant \((\text{chi-square test with a Bonferroni correction})\) in the mainland population. Overall the mean correlation coefficient \(r^2\) between variable sites for Baja was lower \((0.055)\) than for the mainland \((0.246)\).

Frequency and Test of Independence Analysis of G6pd
In Baja, there was a significant negative value of Fu and Li’s \(D\) test \((-2.384, P < 0.05)\). Tajima’s \(D\) test was also negative \((-1.430)\) but not significantly so. Similarly, Fu and Li’s \(D\) test \((-0.895)\) and Tajima’s \(D\) test \((-0.601)\) were negative but not significantly in the mainland. For Baja, the ratios of silent/replacement for fixed \((4/1)\) and polymorphic \((21/0)\) sites were not significantly different (Fisher’s exact test, \(P = 0.192)\). The MK test for the mainland was also not significant (Fisher’s exact test, \(P > 0.99)\) given the observed silent/replacement ratios of fixed \((3/0)\) and polymorphic \((16/4)\) sites.

Population Subdivision and Migration Between Baja and the Mainland
In conjunction with the prior data set of Adh-2 and Adh-1 from Baja (Matzkin and Eanes 2003), gene flow was estimated between D. mojavensis populations by calculating \(F_{ST}\) and \(Nm\) (Hudson, Slatkin, and Maddison 1992) and \(K_{ST}\) (Hudson, Boos, and Kaplan 1992). Significance of \(F_{ST}\) and \(K_{ST}\) was determined by performing 1,000 permutations in the ProSeq program (Filatov 2002). To remove the possible effects of selection at Adh and the geographic distribution of the Adh-2 intron-1 polymorphism, estimates of population subdivision were only performed utilizing silent sites. The estimates from Adh-2 of \(F_{ST}\) \((0.722, P < 0.001)\), \(K_{ST}\) \((0.381, P < 0.001)\), and \(N_{m}\ \text{of} \text{0.096)\ were similar to the} \text{F}_{ST} \((0.474, P < 0.001)\), \(K_{ST}\ \text{of} \text{0.216,} \text{P < 0.001)\, and \text{Nm of} \text{0.277)\ values from Adh-1. Similarly to Adh, only silent sites were utilized for analyzing G6pd. A lower, yet significant, level of population subdivision was observed at G6pd \((F_{ST} = 0.067, P = 0.009; K_{ST} = 0.018, P = 0.02\); and \(N_{m} = 4.6\), adjusted by \(4/3\)).

Discussion
Species Divergence and Age of Duplication
The estimates of divergence between D. mojavensis and D. arizonae of 2.7 MYr \((\text{Adh-2})\) and 2.1 MYr \((\text{Adh-1})\) calculated from \(K_{s}\) are similar to prior estimates of 3.1
MYr and 1.7 MYr for Adh-2 and Adh-1, respectively (Matzkin and Eanes 2003). The level of intraspecific variation can be used to estimate divergence time. This can be done by solving for $T$ (divergence time in units of $2N$ generations) in the equations (equation 5) of Hudson, Kreitman, and Aguade (1987). Using the HKA program (J. Hey), 10,000 coalescent simulations were performed and used to calculate the 95% confidence limits of $T$. Using this method, the divergence ($T$) between the mainland $D. mojavensis$ population and $D. arizonae$ is 7.9 ($2N$ generations), with a 95% confidence interval of 2.8 to 28.6 ($2N$ generations). Given a neutral mutation rate ($\lambda$) of $4.5 \times 10^{-9}$ per base pair per generation (Hey and Wakeley 1997) and the observed $\theta$, of Adh-2 and Adh-1 (0.0082), the $N_e$ of the mainland population of $D. mojavensis$ can be estimated to be $4.6 \times 10^5$. Assuming that $D. mojavensis$ goes through six generations a year, the divergence time between the mainland $D. mojavensis$ population and $D. arizonae$ occurred 1.21 MYA (95% confidence interval, 0.4 to 4.4 MYr). The species divergence ($T$) from this study was more than double the prior estimate (Matzkin and Eanes 2003) using $D. mojavensis$ from Baja, but the estimate of $N_e$ from the mainland $D. mojavensis$ was less than half that of the Baja population. Although the estimates of $T$ and $N_e$ were very different in the two $D. mojavensis$ populations, the final estimates of species divergence time (in MYr) calculated from each population were almost identical.

The original characterization of Adh-2 and Adh-1 placed the time of the gene duplication at 17.9 MYA (Atkinson et al. 1988). In our study, a more recent age of duplication, about 3.5 MYr, was determined (Matzkin and Eanes 2003). In this study, the estimate of the age of duplication, 3.4 MYr, supports the notion of a more recent age for the $Adh$ duplication than originally proposed. The disparity in the age of duplication between the original estimate (Atkinson et al. 1988) and the two recent ones (Matzkin and Eanes 2003; this study) is explained by the fact that the original study used a mammalian rate of substitution. The recent estimate of the age of duplication implies that the divergence between $D. navojoa$ and the $D. mojavensis$/$D. arizonae$ lineage occurred around the time of the Adh duplication. Although it may appear from
the neighbor-joining tree (fig. 3) that _D. navojoa_ does not share the same duplication as _D. mojavensis_ and _D. arizonae_, this may actually be because of the apparent gene conversion that has occurred between the paralogs in _D. navojoa_. Additional _D. navojoa_ sequences must be examined to better determine the origins of the _Adh_ paralogs in _D. navojoa_.

### Differential Modes of Evolution of _Adh_ Paralogs

Gene duplications have been proposed to play an important role in the evolution of novel gene functions (Ohno 1970; Ohta 1974, 1987, 1988a, 1988b). The classical model of gene evolution via gene duplication involves changes in the coding region of one paralog, creating a novel gene function. Other models suggest that adaptive changes in the coding region are rare, and functional divergence between paralogs is a function of changes in the regulatory region (Hughes 1994; Force et al. 1999). These models, such as Force et al.'s (1999) duplication-degeneration-complementation (DDC) model, require that the ancestral preduplication gene possesses a complex spatiotemporal pattern of expression. After the duplication, deleterious changes in the regulatory region of each paralog will limit the expression of each paralog. The outcome of regulatory mutations will be the creation of two genes with nonoverlapping spatiotemporal patterns of expression, thereby releasing each paralog from the constraints of having to function in the entire ancestral pattern. This subfunctionalization allows for the possible functional divergence of the paralogs.

In _Adh-1_, the apparent excess of fixed replacement changes is indicative of adaptive protein evolution (McDonald and Kreitman 1991; Eanes, Kirchner, and Yoon 1993; Matzkin and Eanes 2003). The inclusion of an outgroup sequence makes it possible to examine the lineage-specific pattern of variation in the mainland population of _D. mojavensis_. Although there was a non-significant trend towards excess replacement fixations in the mainland population, this pattern is highly significant when including the Baja population (table 3). Two (Val-236 and Leu-61) of the three amino acid fixations that have occurred in _D. mojavensis Adh-1_ after the divergence from _D. arizonae_ have the potential to be of functional importance. The fixed residue Val-236 is adjacent to a residue involved in the noncovalent interaction at the dimer surface, Asp-237 (Benach et al. 1998, 1999). These possible fixed functional differences in _D. mojavensis Adh-1_ could have played a role in the host shift that occurred after the divergence from _D. arizonae_ (Heed 1978; Ruiz and Heed 1988; Ruiz, Heed, and Wasserman 1990).

The pattern of evolution in _Adh-2_ is distinct from that of _Adh-1_. In _Adh-2_, there is a 17-bp haplotype segregating in intron-1. With the exception of a deletion at position 135, an identical polymorphism was found in _Adh-2_ of the Baja population (Matzkin and Eanes 2003). Surprisingly, the _Adh-2_ haplotype (LA1, \textquotedblright Like _Adh-1\textquotedblright) found at high frequency (0.77) in Baja was found at low frequency (0.13) in the mainland. Although associated with the LA1 haplotype, the _Adh-2_ Fast allozyme allele is not in complete linkage disequilibrium with the LA1 haplotype. The only charge change shared by all Fast alleles is a serine to arginine change at position 84. The residue at position 84 is not noticeably located at or near a point in which it could affect the kinetic properties of ADH (Benach et al. 1998, 1999). Further work is necessary to investigate the effects of the change at position 84 (residue 28) on enzyme activity and substrate specificity.

The LA1 haplotype is identical to the intron-1 sequence of _Adh-1_, and it is most likely the result of a gene conversion event between the paralogs that occurred before the _D. mojavensis_/_D. arizonae_ species split (Matzkin and Eanes 2003). The level of recombination estimated in the _Adh_ region in the mainland population is relatively high and similar to what has been previously observed in Baja (Matzkin and Eanes 2003). Hence lack of recombination alone cannot explain the persistence of the intron haplotype. A possible explanation for the persistence of the intron haplotype polymorphism in both populations, albeit at different frequencies, is selection for...
Origins of \textit{D. mojavensis} Populations

Cytological evidence suggests a Baja California origin of \textit{D. mojavensis} from an ancestral mainland population of \textit{D. arizonae} (Ruiz, Heed, and Wasserman 1990). The present-day mainland populations of \textit{D. mojavensis} are a result of a later colonization event (Ruiz, Heed, and Wasserman 1990). This model of evolution implies that \textit{D. mojavensis} has gone through several host shifts, given the geographic distribution of the cactus hosts (Ruiz and Heed 1988; Etges 1990). A probable shift occurred during the Baja California divergence of \textit{D. mojavensis} from \textit{D. arizonae}, which utilizes the cina cactus. Subsequent host shifts have occurred in the \textit{D. mojavensis} colonization to organ-pipe cactus in mainland desert of Mexico, to barrel cactus (\textit{Ferocactus acanthodes}) in southern California (U.S.), and to prickly pear cactus (\textit{Opuntia} spp.) in Santa Catalina Island (Ruiz and Heed 1988; Etges 1990).

Populations of \textit{D. mojavensis} from Baja and the mainland differ in their level of sequence variation. At all three loci examined (\textit{Adh-2}, \textit{Adh-1}, and \textit{G6pd}), the level of variation was higher in the Baja population than in the mainland (tables 1 and 4). Additionally, the levels of linkage disequilibrium for all three loci in the mainland population are greater than for Baja. Purifying selection occurring in one population and not in the other could produce such a pattern. It is unlikely, however, that all three loci are undergoing the same type of selection in only one population, especially since \textit{Adh-2}/\textit{Adh-1} and \textit{G6pd} are located in different chromosomes. Furthermore, the data suggest that a similar pattern of selection has occurred in both populations (e.g., \textit{Adh-1}). The lower levels of variation and higher levels of linkage disequilibrium suggest a lower overall effective population size of \textit{D. mojavensis} in the mainland. The effective population size of the Baja California population estimated from both \textit{Adh-2} and \textit{Adh-1} is \(1.4 \times 10^6\) (Matzkin and Eanes 2003), whereas the effective population size of the mainland population is \(4.6 \times 10^5\). Although not as great, a similar difference in effective population sizes of Baja (\(2.1 \times 10^6\)) and mainland (\(1.6 \times 10^6\)) populations is proposed for \textit{G6pd}. Furthermore, the replacement polymorphisms found segregating at \textit{G6pd} were only found in the mainland population. Because of a lower efficacy of selection, smaller population size might favor the persistence of slightly deleterious alleles (Ohta 1993).

An examination of the \textit{Adh-1} data indicates that a larger proportion of derived alleles are present in the mainland population, although this pattern is not evident at \textit{Adh-2} (fig. 3). The pattern at \textit{Adh-2} resembles that of an ancient population split, but it says nothing about the direction of colonization. As for \textit{G6pd}, the gene tree could not be resolved (data not shown). A higher effective population size does not necessarily imply ancestry. However, taken together, the higher effective population size in Baja and the presence of derived \textit{Adh-1} alleles in the mainland supports the previous notion of a Baja California origin of \textit{D. mojavensis}, followed by a colonization of...
the mainland and a subsequent cactus host shift from the agria to organ-pipe cactus. This host shift and the associated change in the alcohol environment, is potentially re sponsible for the geographic pattern of allozyme variation at Adh-2 and the overall level of genetic isolation.

Genetic Isolation Between D. mojavensis Populations

Genetically based differences between morphological, life history, and developmental characters have been observed between Baja California and mainland populations (Etges 1989, 1990, 1993). These differences have been attributed as adaptation to the mainland organ-pipe cactus habitat. Organ-pipe cactus rots are about 40 times less abundant per hectare than agria rots but are larger and longer lived than agria rots (Mangar 1982; Etges 1989). These characteristics of the mainland habitat have been associated with the differences between populations. For example, mainland individuals tend to have a larger thorax diameter (increased dispersal efficiency) and a longer developmental time (Etges 1990, 1993). Furthermore, there are significant mating preference differences between the mainland and Baja California populations (Markow, Fogleman, and Heed 1983; Markow 1991). Although there are significant morphological and life history differences between mainland and Baja, allozymic differentiation has been limited to the Adh locus (Zouros 1973; Heed 1978).

Similar to earlier work, this study has shown that the genetic differentiation between populations depends on the locus being examined. Both Adh-2 and Adh-1 show a high level of genetic divergence, whereas not as great a level was observed for G6pd. The signature of selection that has been observed in the Adh-2/Adh-1 cluster in both populations, rather than isolation, could be responsible for the high level of divergence observed in those loci. This suggests that Adh has played a role in the adaptation to a new cactus host. Therefore, it would not be appropriate to use Adh to examine the neutral divergence of two populations utilizing different cactus hosts. The pattern of variation observed at G6pd in this study suggests that the two populations may not be completely isolated. Yet, it is also possible that the persistence of shared lineages between populations at G6pd could have occurred even under complete isolation. Because the colonization of the mainland by D. mojavensis had to have occurred after the species divergence (~2 MYA), it may be that not a sufficient amount of time has passed to be able to detect the population subdivision. Further studies on loci not involved in the adaptation to cactus host are needed to fully determine the level of isolation between the D. mojavensis populations.

Acknowledgments

I would like to thank Walter Eanes for his assistance and support in this project. I also like to thank Therese Markow for assistance in the field and for providing the D. arizonae isofemale lines. Additionally, I would like to thank John True and Thomas Merritt for providing comments and suggestions on the manuscript. This research was in part supported by a U.S. Public Health Service Grant GM-45247 to W.F.E. and NSF Predoctoral Fellowship and W. Burghardt Turner Fellowship to L.M.M. This is contribution number 1105 from the Graduate Program in Ecology and Evolution, State University of New York at Stony Brook.

Literature Cited


