doi:10.1111/j.1420-9101.2006.01111.x

Microsatellite variation among diverging populations of *Drosophila mojavensis*

C. L. ROSS & T. A. MARKOW

Department of Ecology and Evolutionary Biology, University of Arizona, Tucson, AZ, USA

Keywords:

Drosophila mojavensis;
microsatellites;
population structure;
speciation.

Abstract

Divergence and speciation may occur by various means, depending on the particular history, selective environments, and genetic composition of populations. In *Drosophila mojavensis*, a good model of incipient speciation, understanding the population genetic structure within this group facilitates our ability to understand the context in which reproductive isolation among populations is developing. Here we report the genetic structure and relationships of *D. mojavensis* populations at nuclear loci. We surveyed 29 populations throughout the distribution of *D. mojavensis* for four microsatellite loci to differentiation among populations of this species. These loci reveal four distinct geographical regions of *D. mojavensis* populations in the south-western United States and north-western Mexico – (i) Baja California peninsula (Baja), (ii) Sonora, Mexico–southern Arizona, United States (Sonora), (iii) Mojave Desert and Grand Canyon (Mojave), and (iv) Santa Catalina Island (Catalina). While all regions show strong isolation, Mojave and Catalina are highly diverged from other regions. Within any region, populations are largely homogenous over broad geographical distances. Based on the population structure, we find clear geographical barriers to gene flow appear to have a strong effect in isolating populations across regions for this species.

Introduction

Our ability to observe the process of reproductive isolation before it is complete, and hence 'speciation in action', is crucial to understanding the genetics of speciation. Identifying populations of the same species that exhibit various levels of reproductive isolation is an essential first step. Many studies of *Drosophila* and other taxa have primarily utilized closely related pairs of species (Coyne & Orr, 1989, 1997) rather than populations of the same species at an earlier stage along the 'life history' of the speciation process (Harrison, 1998). While interspecific studies are informative about the relationship between genetic differentiation and the strength of a given isolating mechanism after speciation has occurred, they cannot definitively answer questions regarding earlier stages in the process of speciation – when various isolating mechanisms initially arose or what their relationship is to degrees of genetic divergence among genomes as a whole. Ideally, then, evolutionarily diverging populations could be identified prior to speciation and measured both for genetic divergence and the presence of emerging reproductive isolating mechanisms.

Unfortunately, relatively few examples exist in which diverging populations of the same species have been characterized with respect to both reproductive isolation and degree of genetic differentiation among them (e.g. Hollocher *et al.*, 1997; Feder, 1998; Via, 1999; Tregenza *et al.*, 2000). An important exception is *Drosophila mojavensis*, a cactophilic species found in the Sonoran and Mojave Deserts of North America (Markow & Hocutt, 1998). Four geographically separated populations of *D. mojavensis* exist (here abbreviated as Baja, Sonora, Mojave and Catalina) each utilizing a different locally abundant species of cactus as its host plant (Ruiz & Heed, 1988; Fig. 1). In the Baja California peninsula (Baja), these flies utilize agria (*Stenocereus gummosus*) exclusively despite the presence of other columnar cacti, including

Correspondence: Therese A. Markow, Department of Ecology and Evolutionary Biology, University of Arizona, Tucson, AZ 85721, USA. Tel.: 520 621 3323; fax: 520 626 3522; e-mail: tmarkow@arl.arizona.edu
Present address: Charles L. Ross, Department of Biology, New Mexico State University, Las Cruces, NM 88003, USA.

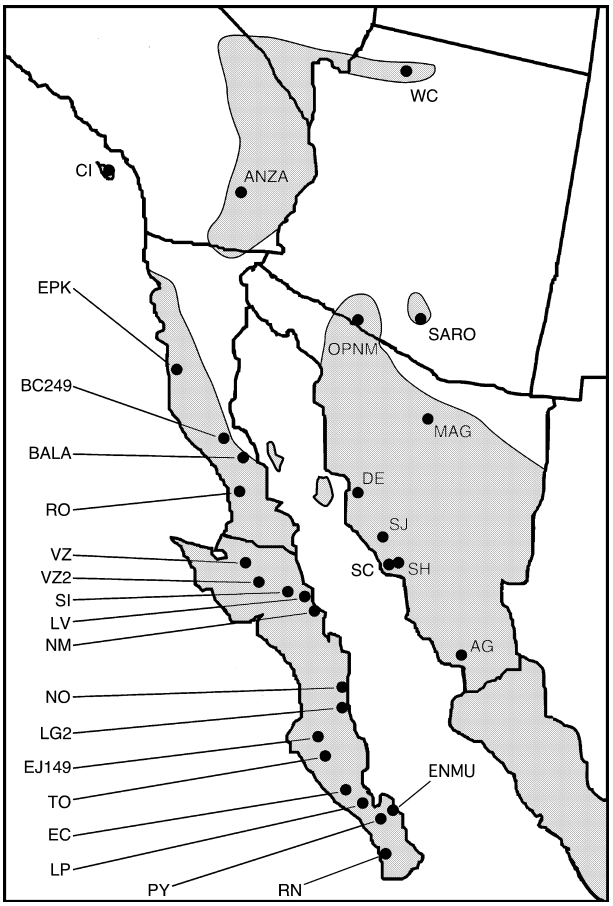

Fig. 1 Geographical distribution of populations sampled for this study. Shaded areas demarcate the approximate distribution of *Drosophila mojavensis*.

occasional stands of organ pipe (*Stenocereus thurberi*). In the Sonoran Desert of Mexico and Arizona (Sonora), *D. mojavensis* exclusively utilize organ pipe with the exception of the limited region near Desemboque (DE) where they use the localized stand of agria that occurs there. In the Mojave Desert and Grand Canyon (Mojave), flies utilize barrel cactus (*Ferocactus cylindraceus*), and on Santa Catalina Island (Catalina), flies use prickly pear (*Opuntia littoralis*). Among these four hosts, *D. mojavensis* utilizes the dominant or exclusive host species in each area of its range. In spite of this, all *D. mojavensis* populations show the highest oviposition preference and performance on agria (Heed & Magnan, 1986; S. Castrezana, personal communication). These four populations of *D. mojavensis* show various degrees of pre- and post-mating reproductive isolation from each other (Zouros & D'Entremont, 1980; Krebs & Markow, 1989; Markow & Hocutt, 1998; Knowles & Markow, 2001). In addition, reproductive isolation between *D. mojavensis* and its sister species *D. arizonae* is incomplete and dependent upon the geographical strain of *D. mojavensis* (Reed & Markow, 2004).

The differing levels of reproductive isolation provide an opportunity to ask questions about the relationship between degree of genetic differentiation and the order of appearance of specific types of isolating mechanisms during, rather than subsequent to the completion of, the process of speciation. Genetic relationships among *D. mojavensis* populations, however, still are unclear. Based upon analyses of chromosome polymorphism, morphology and behaviour, Mettler (1963) originally divided *D. mojavensis* into two subspecies, *D. mojavensis mojavensis* from the Mojave Desert of California and *D. mojavensis baja* from the Sonoran Desert in Baja California, southern Arizona and Sonora. These two subspecies were also referred to as races A and B respectively (Zouros, 1973). Subsequent allozyme studies (Zouros, 1973) lead to the further subdivision of the Sonoran Desert race into two races: B1 in Sonora, and B2 in Baja California, separated by the Sea of Cortez. One locus, alcohol dehydrogenase (ADH), however, contributed disproportionately to the estimated degree of differentiation between Baja and Sonora populations. This locus is likely under strong selection (Starmer *et al.*, 1977) because of its importance in the metabolism of various alcohols in the necrotic tissue among the cactus hosts, so this gene may not reveal a 'neutral' evolutionary history of *D. mojavensis* populations. Additionally, ADH is now known to be duplicated in this species, so questions of homology are introduced (Begun, 1997; Matzkin & Eanes, 2003; Matzkin, 2004). Consequently, these, patterns should be verified with additional molecular markers. Finally, the disjunct populations of *D. mojavensis* later discovered on Santa Catalina Island off Southern California and the Grand Canyon of Arizona were grouped with race A based upon similarities in chromosome polymorphism (Ruiz *et al.*, 1990)

Several lines of evidence question the validity of these groups. First, more recent allozyme studies (Hocutt, 2000) that were extended to include the Santa Catalina Island population suggest that this population may not, in fact, be the most similar to those in the Mojave Desert as suggested by Ruiz *et al.* (1990). The second line of evidence concerns the degrees to which populations of *D. mojavensis* are reproductively isolated from the sister species, *D. arizonae*. Traditionally, hybrid male sterility was thought to be unidirectional: *D. arizonae* females crossed to *D. mojavensis* males always produce sterile sons (Baker, 1947; Patterson, 1947; Wasserman & Koepfer, 1977), while the F_1 males from the reciprocal cross were observed, until recently to be fully fertile. With the discovery of *D. mojavensis* on Santa Catalina Island, however, the reciprocal crosses of this population with *D. arizonae* populations show that mothers from this island produced sterile hybrid sons (Ruiz *et al.*, 1990; Reed & Markow, 2004). These observations suggest that *D. mojavensis* from Santa Catalina may be much more genetically differentiated from other conspecific populations than originally thought.

No molecular studies have yet been reported that use locus-specific, DNA-based markers to determine the levels of genetic differentiation among all of the various geographical populations of *D. mojavensis*. Given the likelihood of natural selection acting on the ADH and other allozyme loci, allozyme data are not likely to provide a 'neutral' historical description of populations within *D. mojavensis*. Moreover, the conserved evolution of chromosome inversions and many allozymes suggest these markers are unlikely to provide a recent temporal resolution of population divergence. Without a reliable picture of the degree to which these populations differ genetically, the evolutionary interpretation of the patterns of emerging reproductive isolation is far less meaningful. We developed microsatellite markers for *D. mojavensis* (Ross *et al.*, 2003) in order to examine genetic differentiation among populations of *D. mojavensis*. Specifically we asked the following questions: (i) What are the genetic relationships among the different geographical host races of *D. mojavensis* based on microsatellite loci? and (ii) How do they differ from relationships based upon allozymes and chromosomal variants?

Materials and methods

Population sampling

Populations were sampled between May 2000 and October 2002 (Table 1, Fig. 1), throughout the four geographical areas where *D. mojavensis* is found. Because the ability to collect flies depends upon the availability of cactus necroses, which can vary spatially and temporally (Breitmeyer & Markow, 1998), the collection strategy was to collect as many flies from as many sites within an area as possible. For Catalina Island, which is small, there is only one population, but it was sampled twice. Some populations were sampled more than once, and these temporal samples were kept distinct during subsequent analyses. In total, 31 population samples were collected representing 29 distinct geographical populations. Adult individuals were caught using two methods: bait trapping using a mixture of fermenting banana mash spiked with the appropriate host plant in the area, and direct aspiration of adults off active rots. Additionally, for a few populations, larvae were also extracted from active rots and collected after eclosion to adults. Males were directly identified to species and frozen at −80 °C, while females were frozen after their male offspring were identified to species in the lab. Overall 1657 individuals were used for genotyping.

Table 1 Populations sampled.

Population	Region	State	Race	N
AG.0201	Sonora	Sonora	B1	50
SC.1100	Sonora	Sonora	B1	46
SH.1100	Sonora	Sonora	B1	74
SJ.1100	Sonora	Sonora	B1	28
DE.1100	Sonora	Sonora	B1	86
MAG.0202	Sonora	Sonora	B1	48
OPNM.1000	Sonora	Arizona	B1	232
SARO.0501	Sonora	Arizona	B1	308
SARO.1001	Sonora	Arizona	B1	290
EPK.0101	Baja	Baja Norte	B2	32
BC249.0101	Baja	Baja Norte	B2	20
BALA.0101	Baja	Baja Norte	B2	66
RO.0101	Baja	Baja Norte	B2	28
VZ.0101	Baja	Baja Sur	B2	236
VZ2.0101	Baja	Baja Sur	B2	70
SI.0101	Baja	Baja Sur	B2	108
LV.0101	Baja	Baja Sur	B2	38
NM.0101	Baja	Baja Sur	B2	66
NO.0101	Baja	Baja Sur	B2	48
LG2.0101	Baja	Baja Sur	B2	32
EJ149.0101	Baja	Baja Sur	B2	358
TO.0101	Baja	Baja Sur	B2	188
EC.0101	Baja	Baja Sur	B2	14
LP.0101	Baja	Baja Sur	B2	42
ENMU.0101	Baja	Baja Sur	B2	260
PY.0101	Baja	Baja Sur	B2	76
RN.0101	Baja	Baja Sur	B2	80
WC.0302	Mojave	Arizona	A	46
ANZA.0402	Mojave	California	A	48
CI.0401	Catalina Island	California	C	118
CI.1002	Catalina Island	California	C	188

The four numbers after each population name indicate the 2-digit month and year each population was sampled. For example, 0101 = January 2001. Two populations (SARO and CI) were sampled on two separate dates. Race designations are described in text. N = number of alleles sampled per population.

Microsatellite genotyping

DNA was extracted from all samples using a modified squish prep (Gloor *et al.*, 1993). For genotyping, we used four microsatellite loci that were developed and characterized previously from enriched clonal libraries of *D. mojavensis* or *D. arizonae* (Table 2; Ross *et al.*, 2003). Three of these loci were dinucleotide repeats, and one (M3147) was a trinucleotide repeat. Two loci, M2192 and A2131, each contained one imperfection within the microsatellite, so these loci likely do not follow simple mutational models. One primer for each locus was fluorescently tagged with either ABI dyes 6-FAM or HEX (Applied Biosystems – ABI – Foster City, CA, USA), and microsatellite loci were multiplexed during polymerase chain reaction (PCR). For each 25-μL reaction, 1/50th of the DNA sample was added to a reaction mix (1x PCR buffer; Invitrogen, Carlsbad, CA, USA; 1.5 mM $MgCl_2$, 0.4 μM of each primer, 0.5 U *Taq*, Invitrogen, and 0.2 mM dNTPs). After an initial 3 min soak at 95 °C, the multiplexed reaction was run 35 times through a temperature profile of 94 °C for 20 s, 53 °C for 45 s, and 72 °C for 90 s, followed by a 10 min extension at 72 °C.

The PCR products were diluted 2 : 3 with H_2O and genotyped using an ABI 3100 genetic analyser and the

 doi:10.1111/j.1420-9101.2006.01111.x

Table 2 Microsatellite loci surveyed. All loci and primers were characterized from *Drosophila mojavensis* except A2131, which was characterized from *D. arizonae* (see Ross *et al.*, 2003).

Locus	Motif	Forward primer (5′–3′)	Reverse primer (5′–3′)	Chromosome
A2131	ac	CAGAAATCGTTTCATTCATGC	CGCTTGGACAACTTTCAGC	4
M2192	ac	CCTTATCGCTGCTCGACTCC	AGGAAAACTTCAGCCAGACG	4
M3147	agc	CAAGATAGCCACAATCAAGTCG	TGTAACCCACTCGCTAAATGC	?
M496	ac	TCAACTGGAAGCTGTTAAATATCG	CATGCATCAGGCTTATCTCC	5

program GENESCAN ver. 3.1 (ABI) at the Genomics and Technology Core facility of the University of Arizona. Standard samples of known size for each locus were run with every plate (94 samples) to adjust for variation among gels and scoring of allele sizes. Alleles were scored using GENOTYPER ver. 1.1 (ABI) and alleles were binned into natural clusters based on size in base pairs (bp). These clusters closely tracked stepwise differences in intervals of the locus' motif size (i.e. 2 bp for dinucleotides, 3 bp for trinucleotides), after adjusting for plate-to-plate variation using the sequenced standards of known size. Allele size (number of repeats) was determined by comparing binned clusters to sequenced standards of known repeat numbers.

Analysis of variation

Genetic diversity for each locus and each population, as well as over all loci and populations, was quantified using MSA ver. 3.00 (Dieringer & Schlötterer, 2003) or ARLEQUIN ver. 2.000 (Schneider *et al.*, 2000) by calculating the number of alleles (N), variance in allele size (v), range in allele size (r), expected heterozygosity (H_e), observed heterozygosity (H_{obs}) and other relevant measures. Deviations from Hardy–Weinberg equilibrium were tested for each locus and over all loci in ARLEQUIN using a Markov chain approximation (Guo & Thompson, 1992), and genotypic phase disequilibrium was estimated to test for independence of loci. All estimates were assessed for significance using Fisher's exact test and permuting the data 10 000 times to create a null distribution. Critical levels of significance were determined after applying sequential Bonferroni adjustments.

We partitioned genetic variance hierarchically across populations using Analysis of Molecular Variance (AMOVA), implemented in ARLEQUIN. We calculated AMOVAS for our populations first with no regional hierarchy, and then with the four major regions. Pairwise F_{st} for all population pairs were estimated, as well as among all regions, and significance for all values was estimated after applying a sequential Bonferroni adjustment on critical levels. We do not report estimates of R_{st} as they yielded similar results to our F-statistics. Even though R_{st} may incorporate a more appropriate, stepwise mutational model (though even this may not be appropriate with 'imperfect' microsatellites), F_{st} estimates have been shown to more reliably estimate population structure with data sets such as the one reported here because the variance associated with the R_{st} estimates are generally high (Balloux & Goudet, 2002; Balloux & Lugon-Moulin, 2002).

Results

Intrapopulation variation

Each locus revealed pronounced variation for all populations typical of microsatellite loci (Appendix S1, Table 3). Allele number for four major geographical regions of *D. mojavensis* varied from four (M3147 locus, Mojave region) to 26 (M496 locus, Sonora region). For each locus, mean allele size was relatively consistent among major geographical regions. Mean numbers of repeats over all populations for the three dinucleotide repeat loci (A2131, M2192, M496) were 10.42, 15.03, 11.25, respectively, and the mean number for the trinucleotide locus (M3147) was 7.14. As expected, the trinucleotide locus showed smaller average allele size and lower variation than the dinucleotide repeat loci (Schug *et al.*, 1998; Ross *et al.*, 2003). These values approximate those estimated from an analysis of *D. mojavensis* microsatellites cloned from 258 dinucleotide repeat loci (mean = 13.34 repeats) and 59 trinucleotide loci (mean = 8.70 repeats) (Ross *et al.*, 2003).

For the four loci across all populations, we conducted 128 assessments for deviation from Hardy–Weinberg equilibrium (Appendix S1). Of these tests, 37 were significant – not unusual for microsatellite loci but a greater number than expected by chance – indicating that several populations are not in mutation/drift-selection/migration balance. Only one major geographical region of *D. mojavensis* does not violate Hardy–Weinberg equilibrium: when the two relatively disjunct populations sampled within the Mojave Desert are combined (Mojave), all loci fall within Hardy–Weinberg expectations. This effect undoubtedly is partially due to the relatively low sample size for this region. Tests of linkage disequilibrium indicated no evidence of linkage among these loci, including M2192 and A2131, which are both located on chromosome 4 (data not shown).

Expected heterozygosity (H_e) for each locus in every population was predictably very high (Appendix S1), with H_e ranging from 0.75 to 0.95 for most populations. Across loci, H_e varied somewhat for populations. Over all

 doi:10.1111/j.1420-9101.2006.01111.x

Table 3 Intra-regional variation for *Drosophila mojavensis* populations per locus.

Locus	Population	*N*	H_{obs}	H_{eq}	Variance in number repeats	Mean number repeats	No. of alleles	n_e (SMM)
A2131	Sonora	1162	0.7745*	0.8628	7.8290	11.0207	19	9
A2131	Baja	1784	0.7029*	0.8785	7.3498	10.0244	23	10
A2131	Mojave	94	0.2766	0.3235	1.6688	11.3617	6	3
A2131	Catalina	306	0.6013*	0.6481	3.5273	9.9624	13	5
M2192	Sonora	1146	0.8307*	0.8305	8.0974	15.1278	24	8
M2192	Baja	1778	0.8234*	0.8620	8.6126	14.8841	24	9
M2192	Mojave	94	0.4255	0.3811	10.6899	13.7021	6	3
M2192	Catalina	306	0.6732*	0.6991	13.4463	16.2516	13	5
M3147	Sonora	1158	0.7219*	0.7930	4.2859	7.2565	12	7
M3147	Baja	1784	0.8150*	0.8317	3.3964	7.0594	13	8
M3147	Mojave	94	0.4043	0.3590	1.0464	9.9149	4	3
M3147	Catalina	306	0.0980*	0.1323	1.0088	6.1438	10	2
M496	Sonora	1130	0.7646*	0.8310	16.3183	12.2637	26	8
M496	Baja	1762	0.8570*	0.8937	10.9609	10.9671	25	11
M496	Mojave	92	0.6087	0.5592	13.9111	13.6087	6	4
M496	Catalina	300	0.4800*	0.5343	6.9570	8.4467	9	4

All values calculated in MSA VER. 3.00 (Dieringer & Schlötterer, 2003). *N*, number of chromosomes surveyed; SMM, stepwise mutational model.

*Significant deviation from Hardy–Weinberg expectations.

populations median H_e for A2131, M2192, M3147, M496 was 0.87, 0.88, 0.81, 0.86, respectively. The trinucleotide locus (M3147) showed significantly less variation than the dinucleotide loci (Tukey–Kramer LSD for M3147 vs. A2131, M2192, M496 = 0.035, 0.017 and 0.031, respectively, where positive values indicate significant interactions. Catalina and Mojave populations were excluded for reasons below). This lower variation at trinucleotide loci is expected due to theoretical predictions and empirical observations of microsatellite motif size class evolution (Kruglyak *et al.*, 1998; Schug *et al.*, 1998; Ross *et al.*, 2003).

Two geographical regions showed reduced variation in different ways (Table 3). Flies from Catalina Island (two temporal samples) showed slightly (but significant) reduced heterozygosity at all loci, but the variation at the trinucleotide locus, M3147, was dramatic. Mojave Desert flies (Mojave), however, showed a pronounced reduction in variation at all loci, with no differences between tri- and dinucleotide loci. These differences in pattern of heterozygosity among loci for these two regions may reflect different histories, different historical population sizes, or different forces acting on populations for these two areas.

Population structure

Overall, most genetic variation at these microsatellite loci for *D. mojavensis* is concentrated among individuals within the same population (Table 4). From our AMOVA using the distribution of different alleles (i.e. an F_{st} measure), 88.95% of all variation in *D. mojavensis* is found within populations. Most of the remaining significant genetic variation resides among major geographical regions (10.80%), and only a minute fraction (0.25%) can be attributed to differences among populations within major geographical regions. These data indicate that each major region is relatively isolated from others, but within any one region there is extensive gene flow.

Pairwise F_{st}s for all populations reinforce the conclusion that the four major geographical regions represent relatively isolated groups of fly populations (Appendix S2, Table 5). We found a high degree of genetic differentiation among all populations (overall $F_{st} = 0.11$), and pairwise F_{st}s ranged from 0.00 to 0.52. Within each of the four regions (Sonora, Baja, Mojave, Catalina), pairwise F_{st} values are all consistently low and only occasionally significantly different from zero, indicating extensive gene flow among populations within each

Table 4 Hierarchical Analysis of Molecular Variance (AMOVA) for *Drosophila mojavensis* populations grouped by major geographical region, and resulting fixation indices.

Source of variation	d.f.	Sum of squares	Variance components	Percentage of variation	Fixation indices
Among regions	4	394.733	0.19226 Va	10.8	F_{it}: 0.11053*
Among populations within regions	27	54.783	0.00444 Vb	0.25	F_{is}: 0.00279*
Within populations	3336	5280.457	1.58287 Vc	88.95	F_{st}: 0.10804*
Total	3367	5729.973	1.77956		

*$P < 0.0001$.

 doi:10.1111/j.1420-9101.2006.01111.x

Table 5 Pairwise F_{st} estimates across populations grouped into major regions. Above diagonal = F_{st}, below diagonal = unadjusted *P*-value.

	Sonora	Baja	Mojave	Catalina
Sonora	0	0.044147	0.246244	0.204662
Baja	0.001	0	0.221827	0.166536
Mojave	0.001	0.001	0	0.479056
Catalina	0.001	0.001	0.001	0

region. For example, within Sonora, pairwise F_{st} values range from 0.000 to 0.022, and within Baja, pairwise F_{st}s range between 0.000 and 0.021, while between Sonora and Baja, F_{st} values range between 0.017 and 0.088. Only 10 of 36 pairwise comparisons within Sonora are significant, and only 18 of 153 comparisons are significant within Baja, but all pairwise population comparisons (162/162) are significantly different from zero between Sonoran and Baja populations. Likewise, for Mojave and Catalina, no within-region comparison is significant, but all among-region comparisons are significant.

Variation among the California *D. mojavensis* populations is especially notable. Catalina and Mojave populations show exceptionally high pairwise F_{st}s with populations in other regions (Appendix S1, Table 5). Whereas the overall F_{st} estimate between Sonora and Baja is approximately 0.04, F_{st}s between Catalina and Sonora or Baja are 0.20 and 0.17, respectively. Similarly, F_{st} estimates between the Mojave region and Sonora or Baja are 0.24 and 0.22, respectively. Consequently, both Catalina and Mojave appear to represent isolated and relatively independent regions of populations. More importantly, the F_{st} estimate between Catalina and Mojave is extremely high ($F_{st} = 0.48$). Therefore, these regions effectively are completely isolated from each other despite their relative close spatial proximity, and they exhibit greater differentiation than what is observed between currently recognized 'races' in Sonora and on the Baja Peninsula. The apparent extensive gene flow between the two sampled populations in the Mojave region (ANZA and WC, $F_{st} = 0.020$) is surprising considering their geographical separation (straight line distance = 425 km) and habitat differences between the Anza Borrego Desert and the Grand Canyon.

Populations within the Baja and Sonora region are also isolated from other populations in other regions. The F_{st} values between Baja and Sonora populations suggest restriction of gene flow between these regions, though not nearly at the level of isolation of Mojave and Catalina Island regions. The Sea of Cortez provides a formidable yet apparently incomplete barrier to gene exchange for these flies. Islands within the gulf may act as 'stepping stones' across the sea, as suggested by other genetic analyses (Zouros, 1973; Johnson, 1980; Hocutt, 2000). Essentially, each of these regions represents large, relatively homogenous distributions of flies.

Discussion

The evolution of host races in *D. mojavensis*

Drosophila mojavensis populations show striking genetic differentiation across the geographical distribution of the species, even when only four microsatellite loci are surveyed. Specifically, four major groups emerge from our genetic survey: one is consistent with what has been called race B1 [designated *D. mojavensis sonora* by Hocutt (2000)] in Sonora and southern Arizona, which breeds in organ pipe cactus, another with race B2 in Baja (*D. mojavensis baja*) breeding in agria, a third matches race A breeding in barrel cactus in the Mojave Desert and the Grand Canyon (*D. mojavensis mojavensis*), and a fourth distinct group, using prickly pear, is found exclusively on Catalina Island [designated *D. mojavensis wrigleyi* by Hocutt (2000) or race C by Pfeiler *et al.* (2005)]. Because of the large number of individuals sampled and the concordance of pattern among all microsatellite loci even though they are unlinked, it is unlikely that additional microsatellite loci would provide a different picture. Thus our microsatellite data support the three established races (A, B1 and B2) found in a previous investigation using allozymes (Zouros, 1973), but not the inclusion of Catalina into race A as presented by Ruiz *et al.* (1990). Lacking the benefit of allozyme or locus-specific DNA data for the Catalina population, Ruiz *et al.* (1990) based their conclusions on morphology (Mettler, 1963) and common chromosome inversion patterns between Catalina and Mojave populations, and they suggested a more recent common ancestor for these groups than other *D. mojavensis* races. Hocutt (2000) suggested, based on a subsequent study of allozyme variation in *D. mojavensis*, that the Catalina Island population should be considered a separate race, *D. mojavensis wrigleyi*. Considering the allozyme (Hocutt, 2000) and our microsatellite data, the Catalina Island population is indeed as divergent as other previously described races. The divergences among Catalina Island or Mojave and other regions are as high or higher than divergences seen in many sibling species (e.g. *D. simulans*, *D. mauritiana* and *D. sechellia*; Perez *et al.*, 1993).

Influence of vicariance, environment and current gene flow

Besides *D. mojavensis* and *D. arizonae*, the plate boundary expansion 3–5 Ma between the North American and Pacific plates resulting in the separation of Baja from the mainland (Gastil *et al.*, 1983; Lonsdale, 1989; Helenes & Carreño, 1999) has been used to explain the divergence and speciation of many other species groups on the Baja peninsula and mainland Sonora (Riddle *et al.*, 2000c; Nason *et al.*, 2002). Two additional vicariant events have been proposed within the peninsula in the form of transpeninsular seaways, one isolating the cape region

and the other in the mid-peninsular region (Upton & Murphy, 1997; Riddle *et al.*, 2000c). Genetic data from vertebrates and cacti support these scenarios (Riddle *et al.*, 2000a, b, c; Nason *et al.*, 2002); however, we find no support that *D. mojavensis* was affected by later vicariant events on the Baja peninsula. Populations within Sonora or Baja show no evidence of structure using an isolation-by-distance model ($P = 0.5$, Mantel tests using 'IBD' program; Bohonak, 2002). This conclusion also is readily apparent by examining the lack of divergence, isolation and partitioned variation in Appendix S2 and Table 4.

Unfortunately, the distribution of host plant, and thus host utilization, is almost completely confounded with geography. Agria, which is the preferred host for all *D. mojavensis*, is found almost exclusively on the Baja peninsula, with only a small stand near Desemboque (DE) in Sonora. Whenever agria is present, *D. mojavensis* individuals utilize this host exclusively, even when other suitable hosts are present, such as in the Baja California peninsula. In Sonora, Mojave and Catalina Island, *D. mojavensis* utilizes organ pipe, barrel cactus and prickly pear, respectively. Though shifts in host utilization may be in part or wholly responsible for the isolation and divergence among *D. mojavensis* races, we cannot exclude the equally parsimonious explanation of geographical vicariance.

Of the other three *Drosophila* endemic to the desert (*D. nigrospiracula*, *D. mettleri* and *D. pachea*), only *D. mettleri* exhibits no differentiation across the Sea of Cortez or within geographical areas. Of the other two species, *D. nigrospiracula* and *D. pachea*, only in the latter is there significant genetic differentiation between Baja and mainland populations (Hurtado *et al.*, 2004). This difference between *D. nigrospiracula* and *D. pachea* is more likely to reflect the relatively greater dispersal abilities of *D. nigrospiracula* (Markow & Castrezana, 2000) rather than adaptations to different hosts, as *D. nigrospiracula* shifts host cacti across the Gulf of California, while *D. pachea* does not. Thus, the Sea of Cortez provides a barrier to gene flow only for two species, *D. mojavensis* and *D. pachea*.

Overall, the regions of Baja and Sonora represent relatively large, homogenous populations throughout each region, with extensive gene flow among all populations. We base this conclusion on low F_{st} values among populations within each region, no evidence of historical patterns of expansion or bottlenecks, no correspondence with past vicariant events within regions or current physiographic regions, and direct estimates of dispersal in these species (Markow & Castrezana, 2000). Furthermore, a Bayesian likelihood analysis of population structure within each region using STRUCTURE (Pritchard *et al.*, 2000) reveals no differentiation below the 'region' level (data not shown). Additionally, patterns of genetic variation within regions show no congruence for an isolation-by-distance model. Similarly, populations of the other three endemic cactophilic *Drosophila* species within a given geographical region are quite panmictic (Markow *et al.*, 2002; Hurtado *et al.*, 2004), probably owing to their dispersal abilities (Markow & Castrezana, 2000). This is not to say, however, that subsequent analyses, using additional loci, might not have the power to detect the existence of substructure within one or more of the major geographical regions inhabited by any of these species.

Divergence of Catalina and Mojave races

One major question that emerges from this study is why Mojave and Catalina Island populations are each so highly diverged from other *D. mojavensis* populations. Both of these regions show low variation and high isolation from Baja and Sonora. For Mojave populations, all four loci show severely reduced heterozygosity compared with Baja and Sonora populations and F_{st} estimates are very high. Catalina Island, however, shows moderate reductions in heterozygosity compared with Baja and Sonora at the three dimer loci, but extremely low heterozygosity at the trimer locus. As with Mojave, F_{st} estimates are very high. Additionally, between Catalina and Mojave there is almost no gene flow. Of the other three *Drosophila* endemic to the desert (*D. nigrospiracula*, *D. mettleri* and *D. pachea*), *D. mettleri* is also found on Catalina Island. As observed for *D. mojavensis*, the Catalina population of *D. mettleri* shows significant genetic differentiation from all other conspecific populations (Markow *et al.*, 2002; Hurtado *et al.*, 2004). In contrast to these microsatellite markers, allozymes, which generally are more conserved and have slower rates of evolution than microsatellites, do not show any reduction in variation or gene flow in these *D. mojavensis* populations (Hocutt, 2000).

Microsatellite loci, with numerous alleles at only moderate-to-rare population frequencies, likely are more sensitive to reduction in effective population size due to bottlenecks, founder events, and selective sweeps than are allozyme loci (Estoup *et al.*, 2001), which frequently have only a small number of alleles all at appreciable frequencies. During a bottleneck or founder event, loci with one or two alleles at moderate frequencies and numerous other alleles at low frequencies should experience a reduction in polymorphism (number of alleles) but not suffer a great reduction in heterozygosity because those alleles at moderate frequencies are likely to be retained in the population through the bottleneck (Nei *et al.*, 1975; Chakraborty & Nei, 1977). This may lead to a transient 'excess' observed heterozygosity (H_{obs}) compared with a population with the same number of alleles at equilibrium (allelic diversity – H_e) because of the faster loss (due to genetic drift) of rare alleles, which contribute relatively little to heterozygosity, compared with the loss of heterozygosity due to the bottleneck (Maruyama & Fuerst, 1985; Cornuet & Luikart, 1996; Garza &

 doi:10.1111/j.1420-9101.2006.01111.x

Williamson, 2001). We find no evidence of a sustained reduction in population size using this criterion. Indeed, all four microsatellite loci show a significant *deficit* in H_{obs} for Catalina and Mojave populations ($P = 0.03$ for each population, Wilcoxon test of two-phased model using BOTTLENECK; Cornuet & Luikart, 1996). However, if a bottleneck was short and followed by a rapid population expansion, then we may not see an effect of heterozygosity excess but rather a heterozygosity deficit due to the rapid expansion, which will likely have a greater effect on genetic variance than the short bottleneck (Maruyama & Fuerst, 1984). This is in fact what we see for the Catalina Island and Mojave regions. Furthermore, relative to Baja and Sonora, both Catalina and Mojave show significantly reduced allelic diversity, measured as n_e, the effective number of alleles (Tukey–Kramer HSD values greater than zero for all appropriate comparisons, indicating significant differences; Table 3), where n_e is the number of equally frequent alleles in an ideal population that would be required to produce the same homozygosity as in an actual population. Reduced levels of allelic diversity (n_e) at microsatellite loci have been shown to be a sensitive indicator of bottlenecks (Spencer *et al.*, 2000). In addition to a loss of genetic variation, bottlenecks can cause rapid increases in genetic distance at microsatellite loci among populations (Chakraborty & Nei, 1977), a pattern consistent with the large divergences of both Catalina and Mojave from other *D. mojavensis* regions.

Populations will recover genetic variation after a bottleneck as a function primarily of migration from other populations and mutation rates at specific loci. Because dinucleotide microsatellite loci generally have higher mutation rates than trinucleotide loci (Schug *et al.*, 1998; Bachtrog *et al.*, 2000), one might expect dinucleotide loci to recover from reductions in genetic variation due to bottlenecks more quickly than trinucleotide loci (Kimmel *et al.*, 1998). If this is the case, then differences we see in heterozygosity and n_e between di- and trinucleotide loci on Catalina Island may be the signature of a severe bottleneck (followed by a very rapid population expansion) where higher mutation rates have allowed dimer loci to partially recover. Genetic variation in Mojave populations may reflect a very recent reduction in the effective population size as all loci show relatively high and equivalent reductions in heterozygosity and n_e. The origin of these populations and demographic events could be explained by the north and south migration of the frost-sensitive host cacti during the Pleistocene glaciation cycles (Van Devender, 1990; Van Devender *et al.*, 1994), leaving relictual populations using prickly pear and barrel cacti as hosts in these areas.

Our observation that the population from Santa Catalina Island is significantly differentiated not only from the Baja and Mainland populations but from the Mojave Desert populations as well, is consistent with the degree of reproductive isolation between *D. arizonae* and the Santa Catalina Island population of *D. mojavensis* (Ruiz *et al.*, 1990; Reed & Markow, 2004). Having an understanding of the degrees of differentiation among the populations of *D. mojavensis* will allow studies of reproductive isolation emerging among *D. mojavensis* populations to be placed in a more realistic evolutionary framework.

Acknowledgments

We are greatly indebted to T. Erez, S. Castrezana, T. Watts, L. Reed, M. Mateos and L. Hurtado, for collecting flies for this project. A. Holyoake and M. Kaplan provided valuable advice with molecular biology approaches and techniques. We also thank the Catalina Island Nature Conservancy for permitting us to collect flies on Santa Catalina Island and providing support while on the island. This work was supported by a National Science Foundation traineeship to C.L.R. under the Interdisciplinary Research Training Group on Plant–Insect Interactions (DBI 9602249), and National Science Foundation grants to T.A.M. (DEB 95-10645 and DEB 0075312).

References

Bachtrog, D., Agis, M., Imhof, M. & Schlötterer, C. 2000. Microsatellite variability differs between dinucleotide repeat motifs: evidence from *Drosophila melanogaster*. *Mol. Biol. Evol.* **17**: 1277–1285.

Baker, W.K. 1947. A study of the isolation mechanisms found in *Drosophila arizonensis* and *Drosophila mojavensis*. *Univ. Texas Publ.* **4752**: 126–136.

Balloux, F. & Goudet, J. 2002. Statistical properties of population differentiation estimators under stepwise mutation in a finite island model. *Mol. Ecol.* **11**: 771–783.

Balloux, F. & Lugon-Moulin, N. 2002. The estimation of population differentiation with microsatellite markers. *Mol. Ecol.* **11**: 155–165.

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

Bohonak, A.J. 2002. IBD (Isolation By Distance): a program for analyses of isolation by distance. *J. Hered.* **93**: 154–155.

Breitmeyer, C. & Markow, T.A. 1998. Resource availability and population size in Sonoran Desert *Drosophila*. *Funct. Ecol.* **12**: 14–21.

Chakraborty, R. & Nei, M. 1977. Bottleneck effects on average heterozygosity and genetic distance with the stepwise mutation model. *Evolution* **31**: 347–356.

Cornuet, J.M. & Luikart, G. 1996. Description and power analysis of two tests for detecting recent population bottlenecks from allele frequency data. *Genetics* **144**: 2001–2014.

Coyne, J.A. & Orr, H.A. 1989. Patterns of speciation in *Drosophila*. *Evolution* **43**: 362–381.

Coyne, J.A. & Orr, H.A. 1997. Patterns of speciation in *Drosophila* revisited. *Evolution* **51**: 295–303.

Dieringer, D. & Schlötterer, C. 2003. Microsatellite analyzer (MSA) – a platform independent analysis tool for large microsatellite data sets. *Mol. Ecol. Notes* **3**: 167–169.

Estoup, A., Wilson, I.J., Sullivan, C., Cornuet, J.M. & Moritz, C. 2001. Inferring population history from microsatellite and enzyme data in serially introduced cane toads, *Bufo marinus*. *Genetics* **159**: 1671–1687.

Feder, J. 1998. The Apple maggot fly, *Rhagoletis pomonella*: flies in the face of conventional wisdom about speciation?. In: *Endless Forms: Species and Speciation* (D. J. Howard & S. H. Berlocher, eds), pp. 130–144. Oxford Press, Oxford.

Garza, J.C. & Williamson, E.G. 2001. Detection of reduction in population size using data from microsatellite loci. *Mol. Ecol.* **10**: 305–318.

Gastil, G., Minch, J. & Phillips, R.P. 1983. The geology and ages of the islands. In: *Island Biogeography in the Sea of Cortez* (T. J. Case & M. L. Cody, eds), pp. 13–25. University of California Press, Berkeley, CA.

Gloor, G., Preston, C.R., Johnson-Schlitz, D.M., Nassif, N.A., Phillis, R.W., Robertson, H.M. & Engels, W.R. 1993. Type I repressors of P element mobility. *Genetics* **135**: 81–95.

Guo, S. & Thompson, E. 1992. Performing the exact test of Hardy-Weinberg proportion for multiple alleles. *Biometrics* **48**: 361–372.

Harrison, R.G. 1998. Linking evolutionary pattern and process: the relevance of species concepts for the study of speciation. In: *Endless Forms: Species and Speciation* (D. J. Howard & S. H. Berlocher, eds), pp. 19–31. Oxford Press, Oxford.

Heed, W.B. & Magnan, R.L. 1986. Community ecology of Sonoran Desert *Drosophila*. In: *The Genetics and Biology of Drosophila*, Vol. 3e (M. Ashburner, H. L. Carson & J. N. Thompson, eds), pp. 311–345. Academic Press, London.

Helenes, J. & Carreño, A.L. 1999. Neogene sedimentary evolution of Baja California in relation to regional tectonics. *J. S. Am. Earth Sci.* **12**: 589–605.

Hocutt, G. 2000. Genetic population structure across comparable geographic regions in the sister species *Drosophila arizonae* and *D. mojavensis*: potential for development of reproductive isolation mechanisms. PhD Thesis, Arizona State University, Tempe, AZ.

Hollocher, H., Ting, C.T., Wu, M.L. & Wu, C.I. 1997. Incipient speciation by sexual isolation in *Drosophila melanogaster*: extensive genetic divergence without reinforcement. *Genetics* **147**: 1191–1201.

Hurtado, L.A., Erez, T., Castrezana, S. & Markow, T.A. 2004. Contrasting population genetic patterns and evolutionary histories among sympatric Sonoran Desert cactophilic *Drosophila*. *Mol. Ecol.* **13**: 1365–1375.

Johnson, W.R. 1980. Chromosomal polymorphism in natural populations of the desert-adapted species *Drosophila mojavensis*. PhD Thesis, University of Arizona, Tucson, AZ.

Kimmel, M., Chakraborty, R., King, J.P., Bamshad, M., Watkins, W.S. & Jorde, L.B. 1998. Signatures of population expansion in microsatellite repeat data. *Genetics* **148**: 1921–1930.

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

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

Kruglyak, S., Durrett, R.T., Schug, M.D. & Aquadro, C.F. 1998. Equilibrium distributions of microsatellite repeat length resulting from a balance between slippage events and point mutations. *Proc. Nat. Acad. Sci. U.S.A.* **95**: 10774–10778.

Lonsdale, P. 1989. Geology and tectonic history of the Gulf of California. In: *The Geology of North America. Vol. N. The Eastern Pacific Ocean and Hawaii* (E. L. Winterer, D. M. Hussong & R. W. Decker, eds), pp. 499–521. The Geological Society of America, Boulder, CO.

Markow, T.A. & Castrezana, S. 2000. Dispersal in cactophilic *Drosophila*. *Oikos* **89**: 378–386.

Markow, T.A. & Hocutt, G.D. 1998. Reproductive isolation in Sonoran Desert *Drosophila*: testing the limits of the rules. In: *Endless Forms: Species and Speciation* (D. J. Howard & S. H. Berlocher, eds), pp. 234–244. Oxford Press, Oxford.

Markow, T.A., Castrezana, S. & Pfeiler, E. 2002. Flies across the water: genetic differentiation and reproductive isolation in allopatric desert *Drosophila*. *Evolution* **56**: 546–552.

Maruyama, T. & Fuerst, P.A. 1984. Population bottlenecks and nonequilibrium models in population genetics. I. Allele numbers when populations evolve from zero variability. *Genetics* **1108**: 745–763.

Maruyama, T. & Fuerst, P.A. 1985. Population bottlenecks and nonequilibrium models in population genetics. II. Number of alleles in a small population that was formed by a recent bottleneck. *Genetics* **111**: 675–689.

Matzkin, L.M. 2004. Population genetics and geographic variation of alcohol dehydrogenase (Adh) paralogs and glucose-6-phosphate dehydrogenase (G6pd) in *Drosophila mojavensis*. *Mol. Biol. Evol.* **21**: 276–285.

Matzkin, L.M. & Eanes, W.F. 2003. Sequence variation of alcohol dehydrogenase (adh) paralogs in cactophilic *Drosophila*. *Genetics* **163**: 181–194.

Mettler, L.E. 1963. *D. mojavensis baja*, a new form in the mulleri complex. *Drosoph. Inf. Serv.* **38**: 57–58.

Nason, J.D., Hamrick, J.L. & Fleming, T.H. 2002. Historical vicariance and postglacial colonization effects on the evolution of genetic structure in *Lophocereus*, a sonoran desert columnar cactus. *Evolution* **56**: 2214–2226.

Nei, M., Maruyama, T. & Chakraborty, R. 1975. The bottleneck effect and genetic variability in populations. *Evolution* **29**: 1–10.

Patterson, J.T. 1947. Sexual isolation in the mulleri subgroup. *Univ. Texas Publ.* **4752**: 32–40.

Perez, D.E., Wu, C.I., Johnson, N.A. & Wu, M.L. 1993. Genetics of reproductive isolation in the *Drosophila simulans* clade: DNA-marker assisted mapping and characterization of a hybrid male sterility gene, Odysseus (Ods). *Genetics* **134**: 261–275.

Pfeiler, E., Reed, L.K. & Markow, T.A. 2005. Inhibition of alcohol dehydrogenase after 2-propanol exposure in different geographic races of *Drosophila mojavensis*: lack of evidence for selection at the Adh-2 locus. *J. Exp. Zool. B.* **304**: 159–168.

Pritchard, J.K., Stephens, M. & Donnelly, P. 2000. Inference of population structure using multilocus genotype data. *Genetics* **155**: 945–959.

Reed, L.K. & Markow, T.A. 2004. Early events in speciation: polymorphism for hybrid male sterility in *Drosophila*. *Proc. Natl. Acad. Sci. U.S.A.* **101**: 9009–9012.

Riddle, B.R., Hafner, D.J. & Alexander, L.F. 2000a. Phylogeography and systematics of the *Peromyscus eremicus* species group and the historical biogeography of North American warm regional deserts. *Mol. Phylogenet. Evol.* **17**: 145–160.

Riddle, B.R., Hafner, D.J. & Alexander, L.F. 2000b. Comparative phylogeography of baileys' pocket mouse (*Chaetodipus baileyi*)

 doi:10.1111/j.1420-9101.2006.01111.x

and the *Peromyscus eremicus* species group: historical vicariance of the Baja California peninsular desert. *Mol. Phylogenet. Evol.* **17**: 161–172.

Riddle, B.R., Hafner, D.J. & Alexander, L.F. & Jaeger, J.R. 2000c. Cryptic vicariance in the historical assembly of a Baja California peninsular desert biota. *Proc. Natl. Acad. Sci. U.S.A.* **97**: 14438–14443.

Ross, C.L., Dyer, K.A., Erez, T., Miller, S., Jaenike, J. & Markow, T.A. 2003. Macroevolutionary patterns of microsatellites: a comparison of variation in five species of *Drosophila*. *Mol. Biol. Evol.* **20**: 1143–1157.

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

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

Schneider, S., Roessli, D. & Excoffier, L. 2000. ARLEQUIN ver. 2.000: a Software Program for Population Genetics Data Analysis. Genetics and Biometry Laboratory, University of Geneva, Geneva, Switzerland.

Schug, M.D., Hutter, C.M., Wetterstrand, K.A., Gaudette, M.S., Mackay, T.F.C. & Aquadro, C.F. 1998. The mutation rates of di-, tri- and tetranucleotide repeats in *Drosophila melanogaster*. *Mol. Biol. Evol.* **15**: 1751–1760.

Spencer, C.C., Neigel, J.E. & Leberg, P.L. 2000. Experimental evalution of the usefulness of microsatellite DNA for detecting demographic bottlenecks. *Mol. Ecol.* **9**: 1517–1528.

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

Tregenza, T., Pritchard, V.L. & Butlin, R.K. 2000. Patterns of trait divergence among populations of the meadow grasshopper, *Chorthippus parallelus*. *Evolution* **54**: 574–585.

Upton, D.E. & Murphy, R.W. 1997. Phylogeny of the side-blotched lizards (Phrynosomatidae: *Uta*) based on mtDNA sequences: support for midpeninsular seaway in Baja California. *Mol. Phylogenet. Evol.* **8**: 104–113.

Van Devender, T.R. 1990. Late Quaternary vegetation and climate of the Sonoran Desert, United States and Mexico. In: *Packrat Middens: the Last 40,000 years of Biotic Change* (J. L. Betancourt, T. R. Van Devender & P. S. Martin, eds), pp. 134–165. University of Arizona Press, Tucson, AZ.

Van Devender, T.R., Burgess, T.L., Piper, J.C. & Turner, R.M. 1994. Paleoclimatic implications of Holocene plant remains from the Sierra Bacha, Sonora, Mexico. *Quaternary Research* **41**: 99–108.

Via, S. 1999. Reproductive isolation between sympatric races of pea aphids. I. Gene flow restriction and habitat choice. *Evolution* **53**: 1446–1457.

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

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

Zouros, E. & D'Entremont, C.J. 1980. Sexual isolation among populations of *Drosophila mojavensis* response to pressure from a related species. *Evolution* **34**: 421–430.

Supplementary Material

The following supplementary material is available for this article online:

Appendix S1. Intra-populational variation for *Drosophila mojavensis* populations per locus.

Appendix S2. Pairwise F_{st} estimates across populations.

This material is available as part of the online article from http://www.blackwell-synergy.com.

Received 1 December 2005; revised 22 January 2006; accepted 31 January 2006