Molecular Ecology (2007) **16**, 3009–3024 doi: 10.1111/j.1365-294X.2007.03325.x

Multilocus nuclear sequences reveal intra- and interspecific relationships among chromosomally polymorphic species of cactophilic *Drosophila*

CARLOS A. MACHADO, LUCIANO M. MATZKIN, LAURA K. REED and THERESE A. MARKOW
Department of Ecology and Evolutionary Biology, University of Arizona, Tucson, AZ 85721, USA

Abstract

***Drosophila mojavensis* and *Drosophila arizonae*, a pair of sibling species endemic to North America, constitute an important model system to study ecological genetics and the evolution of reproductive isolation. This species pair can produce fertile hybrids in some crosses and are sympatric in a large part of their ranges. Despite the potential for hybridization in nature, however, evidence of introgression has not been rigorously sought. Further, the evolutionary relationships within and among the geographically distant populations of the two species have not been characterized in detail using high-resolution molecular studies. Both species have six chromosomes: five large acrocentrics and one 'dot' chromosome. Fixed inversion differences between the species exist in three chromosomes (X, 2 and 3) while three are colinear (4, 5 and 6), suggesting that were introgression to occur, it would be most likely in the colinear chromosomes. We utilized nucleotide sequence variation at multiple loci on five chromosomes to test for evidence of introgression, and to test various scenarios for the evolutionary relationships of these two species and their populations. While we do not find evidence of recent introgression, loci in the colinear chromosomes appear to have participated in exchange in the past. We also found considerable population structure within both species. The level of differentiation discovered among *D. arizonae* populations was unexpectedly high and suggests that its populations, as well as those of *D. mojavensis*, may be themselves undergoing incipient speciation and merit further attention.**

Keywords: *Drosophila*, introgression, multilocus, population structure, speciation

Received 5 January 2007; revision accepted 19 February 2007

Introduction

While species of the genus *Drosophila* have been popular model systems for studies of speciation, only a few of the species offer the advantage of a well-known ecology and geographical distribution. Prominent among these are the cactophilic *Drosophila* endemic to the deserts of North America, and in particular the sibling species *Drosophila mojavensis* and *D. arizonae*. Reproductive isolation between these two species is incomplete when measured in the laboratory (Wasserman & Koepfer 1977; Markow & Hocutt 1998), and it depends upon which populations of *D. mojavensis* are used to estimate reproductive isolation and the particular reproductive isolating mechanism tested (Wasserman & Koepfer 1977; Ruiz *et al*. 1990; Reed & Markow 2004; Massie 2006). In addition, some prezygotic isolation has been detected between different *D. mojavensis* populations (Zouros & D'Entremont 1980; Markow 1991; Hocutt 2000; Knowles & Markow 2001), suggesting that populations of this species may be in the very early stages of species divergence. However, despite being one of the best *Drosophila* systems to study ecological genetics and speciation, we still do not know the divergence history of this group in detail.

The two species are sympatric in Sonora, southern Arizona and northern Sinaloa, but allopatric in large parts of their ranges. Further, there is strong evidence that in areas of sympatry, the two species share some of the same primary host plants where they mate and complete their larval development (Fellows & Heed 1972; Ruiz & Heed 1988). This sharing of feeding and mating grounds in areas of sympatry, coupled with the observation of incomplete reproductive isolation in the laboratory suggests that

Correspondence: Carlos A. Machado, Fax: 520-621-9190; E-mail: cmachado@email.arizona.edu

hybridization and introgression between *D. mojavensis* and *D. arizonae* should be observed in nature. Further, as the species differ in the presence of fixed chromosomal inversions in three of their six chromosomes (see below), it is expected that any evidence of genetic exchange between them should be noticeable in colinear regions of their genomes, those that can recombine in hybrid females, as suggested by recent results from other species groups (Rieseberg *et al*. 1999; Machado *et al*. 2002; Besansky *et al*. 2003; Machado & Hey 2003; Panithanarak *et al*. 2004; Feder *et al*. 2005; Machado *et al*. 2007).

Large chromosome surveys in the two species have not found evidence of hybrids or introgression of chromosomal arrangements (Johnson 1980; Etges *et al*. 1999), but such cytological surveys would fail to detect introgression at a finer scale in colinear chromosomes. Comprehensive studies using multiple molecular markers located in all chromosomes are required for determining whether introgression between *D. mojavensis* and *D. arizonae* is occurring or has occurred during their history of divergence. This is an important question to be addressed because introgression can seriously impact interpretations of the evolutionary histories of these two species and their populations, as well as on the historical and geographical context in which different mechanisms of reproductive isolation have evolved. A recent molecular study using sequences from loci located in three chromosomes (X, 2, 4) (Counterman & Noor 2006), found no evidence of historical introgression between the two species at colinear or inverted regions of their genomes. Although those results are important, the question of differential introgression across different genomic regions, especially in the colinear chromosomes themselves where introgression is most likely, is still not fully resolved, because not all chromosomes were sampled.

In the present study, we use 10 newly developed variable DNA sequence markers, two on each Muller element A–E (chromosomes X, 2, 3, 4 and 5), to address a series of questions on the evidence of historical introgression between *D. mojavensis* and *D. arizonae* and their population structure. Specifically, we address the following questions: (i) What are the patterns of divergence and introgression between species based on the multilocus data set we developed? (ii) Do those patterns differ in the context of geographical location (allopatry vs. sympatry)? (iii) What are the patterns of population structure of both species when using multiple loci across the genome, and how do these patterns compare to those reported using mtDNA or microsatellite data? (iv) Do inverted chromosomes (X, 2 and 3) show a different pattern of divergence between *D. mojavensis* and *D. arizonae* than colinear ones (4 and 5)? If any patterns of introgression are detected, we expect that they will reflect the constraints of the fixed chromosomal differences between the species and/or constraints on the location of hybrid dysfunction genes or genes important in species-specific adaptations. In particular, we expect to see less divergence in colinear chromosomes than in chromosomes with fixed inversion differences. Furthermore, we expect to observe strong population structure in *D. mojavensis*, and to a lesser extent in *D. arizonae*, in agreement with a recent study using the mitochondrial locus COI (Reed *et al*. 2006). The results will provide a meaningful context for testing hypotheses about the evolution of reproductive isolation in this important species pair.

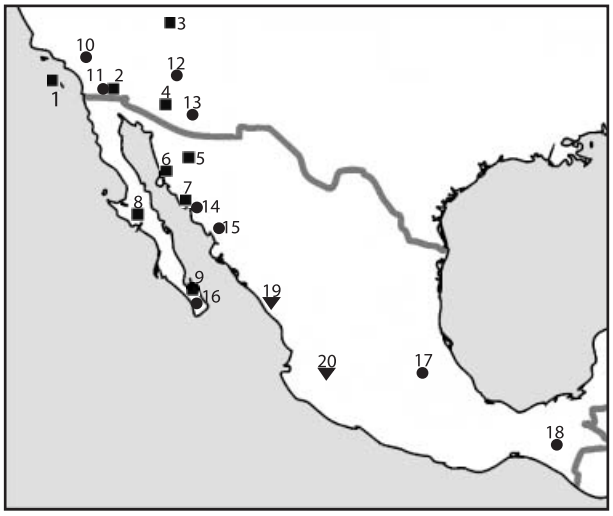

Fig. 1 Collection locations for the *Drosophila* lines used in this study. The squares, circles and triangles indicate the collection sites of *D. mojavensis*, *D. arizonae* and *D. navojoa*, respectively. (1) Catalina Island (CI401-9, CI401-12, CI401-21, CI1002-27); (2) Anza-Borrego Desert (ANZA-1, ANZA-16); (3) Whitmore Canyon (WC302-9, WC302-20); (4) Santa Rosalia (SARO-1); (5) Magdalena de Kino (NS-7, NS-10, NS-13); (6) Desemboque (DE101); (7) Guaymas/San Carlos (MJ122, MJ163, SC1100), (8) Vizcaino (VZ101-24, VZ101-90); (9) La Paz (MJBC178); (10) Riverside (RVSD-10, RVSD-11, RVSD-12); (11) Anza-Borrego Desert (ANZA-18); (12) Peralta Canyon (PERAib-10); (13) Tucson (ARTU1, ARTU2, ARTU6); (14) Guaymas (AROO1, AROO2, AROO3); (15) Navojoa (ARNA5, ARNA24, ARNA28); (16) Ensenada de los Muertos (ENMUib-12); (17) Hidalgo (Hid, MXT8-11, MXT8-16, MXT-9); (18) Chiapas (Chiapas-13); 19. El Dorado (Nav-10); 20. Jalisco (Nav-12).

The Drosophila mojavensis-D. arizonae *system*

The distribution of *D. mojavensis* and *D. arizonae* include the deserts of the US southwest and Mexico (see Fig. 1). Several features of their distribution are notable. First, the two species are sympatric in Sonora and southern Arizona, but allopatric in large parts of their ranges. Second, *D. mojavensis* actually exists as four regional populations (Santa Catalina Island, Mojave Desert, Baja California and Sonora), specializing on necroses of different host cacti in each area (Heed 1982; Reed *et al*. 2006; Ross & Markow 2006), while *D. arizonae* is more of a generalist, using various

cactus necroses in the desert, and domestic fruit in urban areas (Markow *et al*. 1999). Third, the chemical composition within the different cactus necrosis can drastically differ (Heed 1978; Vacek 1979; Kircher 1982). In areas of sympatry, however, adult flies have been collected in large numbers from each other's primary host plants, and when cacti have been brought into the laboratory, the emergence of adults of both species from a given cactus illustrates the extensive overlap of larval habitat of these flies (Fellows & Heed 1972; Ruiz & Heed 1988).

The evolutionary relationships between these two species and between populations of each species remain unclear. Wasserman (1992) used chromosome rearrangements to describe the events that gave rise to these species. A close relative, *D. navojoa*, possesses a more ancestral chromosome arrangement, while *D. mojavensis* and *D. arizonae* have derived chromosomal arrangements (Ruiz *et al*. 1990). All three species have six pairs of acrocentric chromosomes: five rods and a dot. Between *D. mojavensis* and *D. arizonae*, there are fixed inversion differences in chromosomes X, 2 and 3, while chromosomes 4, 5 and 6 are homosequential, or colinear, and thus could recombine in hybrid females (male *Drosophila* do not exhibit recombination). Within *D. mojavensis*, there is considerable chromosomal polymorphism for inversions in chromosomes 2 and 3 (Johnson 1980; Ruiz *et al*. 1990), while no chromosome polymorphism whatsoever has been reported for *D. arizonae*. Relationships among populations of *D. mojavensis* differ depending upon whether inferred from levels of chromosome polymorphism or from molecular sequence data. Heed (1982) places the ancestral populations of *D. mojavensis* in the Baja California Peninsula because this area contains the highest degree of inversion polymorphism for the species. On the other hand, Reed *et al*. (2006) concluded, based on mtDNA COI sequences, that the barrel cactus breeding population of *D. mojavensis* from southern California (Mojave Desert) is basal to all three others. In contrast, no biogeographical scenario has been proposed for *D. arizonae*, and, aside from the Reed *et al*. study on COI variation (2006), only limited population genetic studies addressing the issue of population structure have been conducted in this species. Population genetic analyses of the COI data suggest that there is little population structure within the species, and that it has a population size smaller than that of *D. mojavensis*. The latter result stands in stark contrast with results using sequence data from the nuclear gene *Adh* (Matzkin & Eanes 2003; Matzkin 2004).

Materials and methods

Fly Lines

A total of 19 lines of *D. mojavensis*, 20 lines of *D. arizonae* and 2 lines of the outgroup species *Drosophila navojoa* were utilized in this study (Table 1, Fig. 1). Our sampling scheme included lines representative of all four *D. mojavensis* host races and lines representing all the geographical range of *D. arizonae*. Sonoran Desert *D. mojavensis* have historically been in sympatry with *D. arizonae*, although in recent years *D. arizonae* has been found in the southern tip of Baja California (hereafter 'Baja'), and both in the Mojave Desert and Riverside, California (T.A. Markow and L.K. Reed, unpublished data). For the purpose of the analyses, only the *D. mojavensis* Sonoran population will be classified as sympatric. All *D. mojavensis* and *D. arizonae* lines utilized in this study are isofemale, with the exception of SC1100, DE101, Chiapas-13, and Hid which were started from mass collections (PERAib-10 was mass collected but has been inbred for over 10 generations). The two outgroup *D. navojoa* lines were obtained from the Tucson Drosophila Stock Center (Stock numbers 15081–1374.10 and 15081–1374.12).

Table 1 *Drosophila mojavensis* and *Drosophila arizonae* isofemale lines used in this study

D. mojavensis		*D. arizonae*	
Population	Line	Population	Line
Catalina Island	CI401-9	Sonora	ANZA-18
	CI1002-27		ARNA24
	CI401-12		ARNA28
	CI401-21		ARNA5
Mojave Desert	WC302-20		AROO1
	WC302-9		AROO2
	ANZA-1		AROO3
	ANZA-16		ARTU1
Baja California	VZ101-90		ARTU2
	VZ101-24		ARTU6
	MJBC178		PERAib-10
Sonora	SARO-1	Baja California	ENMUib-12
	MJ122	Southern Mexico	Chiapas-13
	MJ163		Hid
	SC1100		MXT-9
	DE101		MXT8-11
	NS-7		MXT8-16
	NS-10	Riverside, CA	RVSD-10
	NS-13		RVSD-11
			RVSD-12

See Fig. 1 for geographical locations of the collections.

Loci

Table 2 shows the 10 markers used in this study and the primers used for polymerase chain reaction (PCR) and sequencing. Two markers per Muller elements A–E were used; no markers were sampled in Muller element F (the dot chromosome). Nucleotide sequences from all markers were approximately 900-bp long on average. Four of the

Table 2 Chromosome location and structure of the sequenced markers

Locus	Chromosome location	Muller element	Coding regions*	*e* value§	Primers
X100	X	A	*fs*(1)*h* (1–47†)	2 H 10^{-91}	GTTTGCATGCTTATGGCTAAG/ ATTTCAGGCGATGCTGAC
3196	X	A	*fz4* (719–910)	0	ACGTTTGCAGCGTATACTCCAC/ TCGATTTGCCGCCACTTAG
5246	2	E	Unknown (159–248)	—	CCGGACTTTGGACACGTTG/ TTTGTGGGAGCTTTCACGG
5307	2	E	CG18519 (1–441, 532–922)	0	TGGCTGTCACAAAGGAGTTAGC/ TGGATGGACGTGACCCAAG
996	**3**	B	CG31826 (1–161†) *Ku80* (759–854)	1 H 10^{-116}	GCCAGATATTCACTCACGATGG/ TCGCTGACAATCTGGTAGAAGG
A4125	3	B	*Pkd2* (515–924)	1 H 10^{-30}	ACGCCCACTGAAATGAAAGC/ ACAAAGTGCAGGGTGTCTGC
1343	4	D	CG10274 (1–893)	0	CAGGCGAAATCCTTAACCTTC/ AGAGACTATCCCTGCACGGAG
A4115	4	D	CG9194 (129–335, 404–682 or 404–707‡)	3 H 10^{-70}	CATATGCATTTAATTTTCAATAAGTGG/ TCATGCTCAAGACTCCAACG
5239	5	C	CG30127 (1–785, 867–882)	1 H 10^{-120}	ATTGACAGGAGAGCCGTCATC/ TTGTGTGTATCACGGAATCGG
M491	5	C	CG15712 (6–527, 582–722)	6 H 10^{-43}	CCTGCAACTCAAACTTCACC/ GCTTCAGCTACCCAAAAGTCC

*The base positions of the exons are in parentheses; †the exon starts at the second codon position; ‡three *D. mojavensis* lines contain an indel which produces a elongated exon with seven additional amino acids; §the *e* value result from a protein BLAST to *Drosophila melanogaster*.

markers (X100, A4125, M491 and A4115) correspond to the flanking regions of previously developed and mapped microsatellite loci (L.K. Reed, B.A. La Flamme and T.A. Markow, in preparation). To develop the markers, we first used BLAST (Altschul *et al*. 1990) to localize the microsatellite locus plus its flanking region in the sequenced *D. mojavensis* genome (December 2004 Agencourt Assembly). We then chose regions in the vicinity of the microsatellite locus, 4.5–9.0 kb away from the locus, that could be amplified by PCR. The remaining markers were produced by randomly choosing regions of the sequenced *D. mojavensis* genome (December 2004 Agencourt Assembly) that were assigned to Muller elements via syntenic analysis (W. Gelbart, personal communication). The locations of all markers were further verified using a second *D. mojavensis* genome assembly (December 2005) and the *Drosophila melanogaster* Alignment Net algorithm (UCSC Genome Browser, http://genome.ucsc.edu/). The Alignment Net algorithm determines the largest orthologous nucleotide chain between two genomes.

Within all markers there were regions pertaining to exons of confirmed or predicted coding genes (Table 2). Marker 996 contains exons from two different loci, CG31826 and Ku80. Marker A4115 (CG9194) is particularly interesting as three *D. mojavensis* lines, two from Catalina Island and one from Baja, contain a 4-bp indel that produces an exon with a predicted product that is seven amino acid longer than the product predicted from all other lines from the three species. The 4-bp indel is located 15 bases before the putative stop codon in all other lines, and generates 11 additional nucleotide polymorphisms in *D. mojavensis* compared to an alignment including the-4 bp indel inside the exon (see Table 3). In order to be conservative, we used the alignment that included the 4-bp indel inside the exon in our analyses, and thus we did not take into account the additional nucleotide polymorphism generated by moving the indel to the following intron. The coding exons of all markers had strong BLAST hits to *D. melanogaster* proteins (Table 2), with the exception of marker 5246 which had no hit although a coding gene was predicted using GENSCAN. With very few exceptions we were able to PCR amplify and do bidirectional sequencing of all 10 markers for each of the 41 lines described above. Nucleotide sequences are available in GenBank under Accession nos EF436596–EF436986.

Data analyses

Sequences from each data set were edited and initial alignments obtained with the program SEQUENCHER (Genecodes Corporation). Alignments were improved with CLUSTAL_X (Thompson *et al*. 1997), and manual alignments were performed in some data sets to further improve the CLUSTAL alignments because of the presence

Table 3 Polymorphism statistics of the sequenced loci

Locus	Species	*n*	*L*	*S*	syn[a]	repl[†]	$\hat{\theta}$[‡]	π[§]	*D*[#]	$D_{Fu\&Li}$[**]	Divergence[††]
X100	*mojavensis*	19	882.9	40	0	0	0.0129	0.0104	–0.791	–1.843	0.1877
	arizonae	20	874.3	76	0	0	0.0245	0.0206	–0.648	–0.256	0. 1853
	navojoa	2	782	25	0	0	0.0319	0.0320	—	—	—
3196	*mojavensis*	19	874.9	29	5	1	0.0095	0.0072	–0.945	–1.317	0.0393
	arizonae	20	873.8	27	4	1	0.0087	0.0053	–1.518	–2.592*	0.0414
	navojoa	2	871	0	0	0	0	0	—	—	—
5246	*mojavensis*	18	859.8	27	0	1	0.0091	0.0066	–1.096	–1.725	0.0485
	arizonae	19	832.7	33	1	2	0.0113	0.0082	–1.103	–1.976*	0.0475
	navojoa	2	874	8	0	0	0.0091	0.0091	—	—	—
5307	*mojavensis*	19	918	30	8	8	0.0093	0.0084	–0.421	0.579	0.0729
	arizonae	20	915.6	17	9	7	0.0052	0.0041	–0.774	–0.804	0.0681
	navojoa	2	922	10	6	4	0.0108	0.0108	—	—	—
996	*mojavensis*	19	842.7	18	2	0	0.0061	0.0060	–0.076	0.051	0.1431
	arizonae	19	833.5	48	6	1	0.0165	0.0118	–1.142	–1.672	0.1336
	navojoa	2	830	5	1	0	0.0060	0.0060	—	—	—
A4125	*mojavensis*	19	907.6	25	4	8	0.0079	0.0086	0.339	0.0400	—
	arizonae	18	882.7	63	8	4	0.0207	0.0140	–1.352	–1.327	—
	navojoa	—	—	—	—	—	—	—	—	—	—
1343	*mojavensis*	19	869.5	30	26	4	0.0099	0.0086	–0.494	–0.767	0.0539
	arizonae	20	850.5	27	25	2	0.0089	0.0058	–1.354	–1.594	0.0517
	navojoa	2	855	1	1	0	0.0012	0.0012	—	—	—
A4115	*mojavensis*	18	827.2	23	9	2	0.00808	0.0063	–0.891	–1.191	0.0342
	mojavensis‡‡	18	827.2	34	9	13	0.0119	0.0101	–0.611	–0.221	0.0366
	arizonae	13	826.2	23	6	0	0.0090	0.0066	–1.144	–0.646	0.0382
	navojoa	2	828	9	4	0	0.0109	0.0109	—	—	—
5239	*mojavensis*	19	867.3	30	17	10	0.0099	0.0056	–1.702*	–2.786*	0.0379
	arizonae	20	865	40	26	14	0.0130	0.0077	–1.624*	–2.084*	0.0401
	navojoa	2	858	10	3	5	0.0117	0.0117	—	—	—
M491	*mojavensis*	19	773.7	39	26	10	0.0144	0.0124	–0.553	–0.865	0.0371
	arizonae	20	769.9	48	26	19	0.0176	0.0094	–1.867*	–3.396*	0.0340
	navojoa	2	754	5	2	2	0.0066	0.0066	—	—	—

*Significant at $P < 0.05$; n, number of lines sequenced; L, average length (bp) of the sequences from each species; S, number of polymorphic sites; (—) values could not be obtained for small samples or for groups of sequences with few informative sites. Marker A4125 could not be sequenced in *Drosophila navojoa*; †number of synonymous (syn) and replacement (rep) polymorphisms in the coding regions; ‡estimate of $4N$ per base pair using the number of polymorphic sites (Watterson 1975); §estimate of $4N$ using the average number of nucleotide differences per site (Nei 1987); ¶Tajima's statistic (1989). Significance was determined using 10 000 coalescent simulations; **Fu and Li's statistic (1993). Significance was determined using 10 000 coalescent simulations; ††average divergence per base pair between alleles from each species and the alleles of *D. navojoa*; ‡‡based on alternative alignment of *Drosophila mojavensis* that does not include a 4-bp deletion inside the last exon of three strains (see text). The alignment difference generates 11 polymorphic bases in *D. mojavensis*.

of multiple indels. Basic polymorphism analyses were performed with the programs SITES (Hey & Wakeley 1997) and DNASP version 4.10 (Rozas *et al*. 2003). Indels were not included in the polymorphism analyses. Population structure analyses using analysis of molecular variance (AMOVA, Excoffier *et al*. 1992) were performed with ARLEQUIN version 2 (Schneider *et al*. 2000). The statistical significance of F_{ST} values was assessed by 1000 permutations of haplotypes between populations. The probabilities of the pairwise F_{ST} values were combined into the quantity $-2\Sigma \ln P$ to generate an overall test of significance, or meta-analysis, as described in pp. 779–782 of Sokal & Rohlf (1995). The combined $-2\Sigma \ln P$ quantity is χ^2 distributed with $2k$ degrees of freedom, where k is the number of unlinked loci [see Schöfl *et al*. (2005) for an example of implementation of this method in analyses of population structure].

Phylogenetic analyses were conducted using version 4.0b1 of PAUP* (Swofford 1998) using a matrix of 8052 bp from 36 fly lines (19 *D. mojavensis* and 17 *D. arizonae*), that was generated by concatenating sequences from nine loci. Locus A4115 was not included in that data matrix because the sample size in *D. arizonae* was small (13 sequences). Gene trees were reconstructed using the neighbour joining (NJ) algorithm (Saitou & Nei 1987) with Tamura–Nei distances (Tamura & Nei 1993). Because the phylogeny was reconstructed using a multilocus data set from nuclear

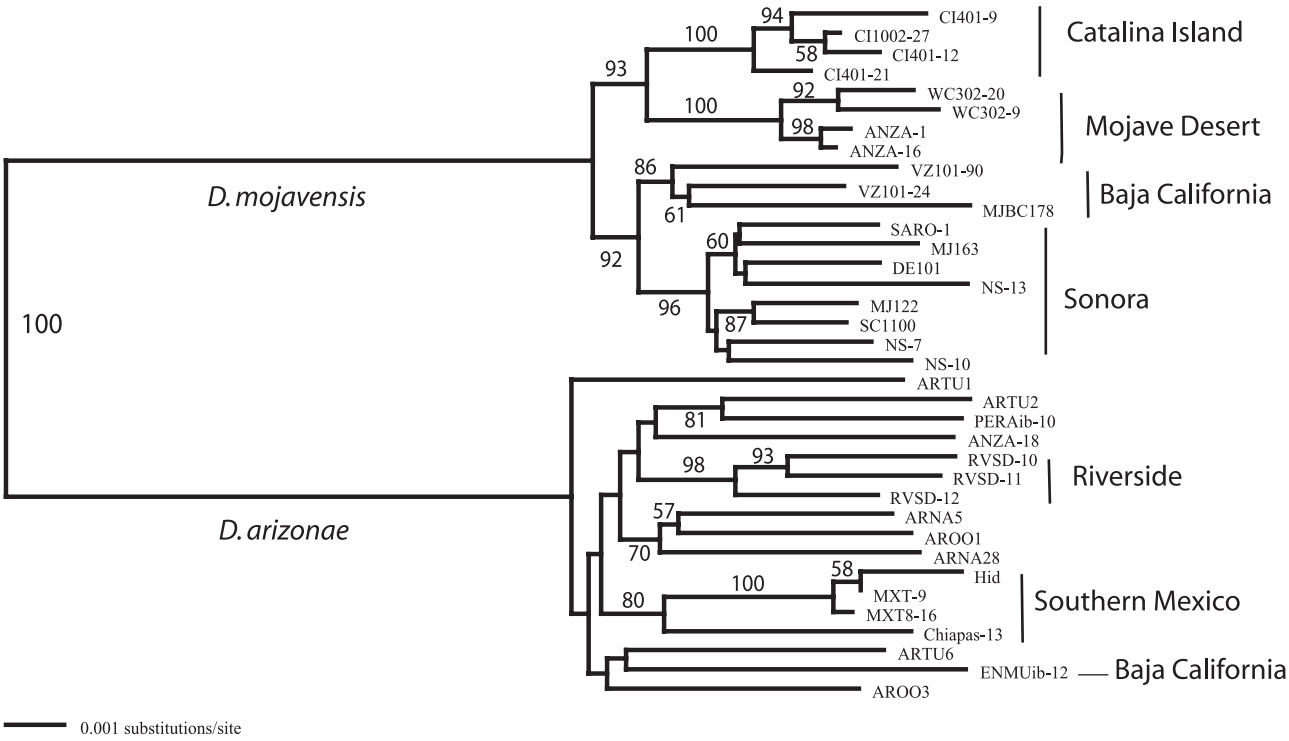

Fig. 2 Neighbour-joining (NJ) tree of nine concatenated nuclear loci (8052 bp). Numbers associated with branches represent bootstrap values greater than 50% (based on 500 pseudoreplications). All the *D. arizonae* strains not grouped by population are from Sonora.

recombining genes, the NJ gene tree shown in Fig. 2 is intended to provide a basic picture of the history of divergence of the *populations* sampled and not of the *fly strains* surveyed. This is justified by our finding of strong population structure in both species (see Results), which reduces the problems caused by intragenic recombination on gene tree reconstruction as it is expected that recent recombination should not have occurred among sequences from different populations.

Tests of neutrality were conducted using DNASP version 4.10 (Rozas *et al*. 2003) or the HKA program (Jody Hey, Rutgers University). We conducted single-locus neutrality tests with (Fu & Li 1993) or without (Tajima 1989) an outgroup. In these tests, what is evaluated is whether the frequency spectrum of mutations is significantly different from that expected under neutrality. For the Fu & Li test (Fu & Li 1993) we used a single randomly chosen sequence of either species to represent the outgroup. The Hudson-Kreitman-Aguade (HKA) test (Hudson *et al*. 1987) was applied either by including all sequences from each species or by using a single sequence of either one of *D. mojavensis* or *D. arizonae* as outgroups. We also present results of the HKA test using the outgroup *D. navojoa*, although we do not discuss those results in detail because the estimated divergence time between this species and the *D. mojavensis*/*D. arizonae* cluster is considerable [7.8 million years, inferred from Russo *et al*. (1995), or between 2.9 and 4.5 million years ago as reported by Matzkin & Eanes (2003), Matzkin (2004), and Reed *et al*. (2006)]. The assumption of constant historical population size in each species was tested by determining whether the observed average value of Tajima's *D* and Fu & Li's *D* across loci departed significantly from zero. Significance for all test statistics was determined by simulating 10 000 neutral genealogies with the program HKA (Jody Hey, Rutgers University), under the conservative assumption of no recombination, and using parameters estimated from the data. We also performed the McDonald–Kreitman test (McDonald & Kreitman 1991) on every coding region. Significance was determined using *G*-tests of independence using Williams' correction or Fisher exact tests (Sokal & Rohlf 1995).

The polymorphism data was fitted to a model of speciation with no gene flow (isolation model) (Wakeley & Hey 1997) using the method described by Wang *et al.* (1997) implemented in the program WH. We used two test statistics, WWH and χ^2, to evaluate the fit of the data to the model. The WWH test statistic is the difference between the highest and lowest number of fixed differences across loci plus the difference between the highest and lowest number of shared polymorphisms (Wang *et al*. 1997). We also computed a χ^2 statistic by comparing the observed and expected values of shared polymorphisms, exclusive

polymorphisms and fixed differences (Kliman *et al*. 2000). Coalescent simulations (10 000) were conducted to determine the level of significance. Poor fit of the data to the isolation model leads to a high value of the test statistic and suggest the occurrence of gene flow at some loci during species divergence. The *P* values presented in Table 5 correspond to the proportion of values of the test statistic estimated from the simulated data that are greater than or equal to the observed. The tests are one-tailed because the focus is on detecting a departure from the model in the direction expected if historical gene flow had occurred. We conducted comparisons between all strains or among sympatric strains of both species, as we predict that if historical introgression has occurred it would be more likely to be detected in the sympatric strains.

Results

Levels of DNA polymorphism are not significantly different between the two species

Table 3 shows the basic polymorphism statistics for the 10 loci sequenced. All loci are polymorphic in the two species, and only one (3196) has no variation in the outgroup *D. navojoa*. The weighted average values of $\hat{\theta}$ and π per base pair for the eight autosomal genes are, respectively, 0.0093 and 0.0078 for *D. mojavensis*, and 0.0128 and 0.0084 for *D. arizonae*. Although averages are larger in *D. arizonae*, the differences across all 10 genes are not significant (paired Wilcoxon signed-rank test; $\hat{\theta}$: $Z = 16.5$, $P = 0.10$; π: $Z = 6.5$, $P = 0.55$). Despite the lack of significance, the observed trend agrees with previous suggestions of a larger historical effective population size for *D. arizonae* (Matzkin & Eanes 2003; Matzkin 2004). If only silent sites are included, the weighted average values of $\hat{\theta}$ and π for the autosomal loci are, respectively, 0.0049 and 0.0041 for *D. mojavensis*, and 0.0066 and 0.0039 for *D. arizonae*. Similarly, differences using silent sites are not significantly different from zero across all 10 genes ($\hat{\theta}$: $Z = 10.5$, $P = 0.25$; π: $Z = -2.5$, $P = 0.79$).

Tests of the neutral model

The neutral model predicts a correlation between levels of polymorphism and divergence across loci, and the HKA test (Hudson *et al*. 1987) is used to test that prediction. No significant departure from the neutral model was detected in any comparison, neither when all sequences from the two species were included ($\chi^2 = 12.48$, $P = 0.5664$), nor when analyses were conducted with a single sequence from either *D. mojavensis* or *D. arizonae* as outgroup (*D. mojavensis:* $\chi^2 = 11.34$, $P = 0.1002$; *D. arizonae*: $\chi^2 = 3.7201$, $P = 0.7132$). Similarly, analyses conducted using the divergent *D. navojoa* outgroup do not reject the neutral model (*D. mojavensis*: $\chi^2 = 24.92$, $P = 0.0619$; *D. arizonae*: $\chi^2 = 16.10$, $P = 0.3901$). Thus, there is no evidence of recent selection at these 10 loci using the HKA test.

The McDonald–Kreitman test (McDonald & Kreitman 1991) was applied to every coding region. This test examines whether the ratio of silent to replacement variation is the same for polymorphisms as it is for fixed differences between species. Under the assumption that these two kinds of variation are selectively neutral, the ratios are expected to be the same. We did not observe evidence of violation of the neutral model in any of the eight loci in which this test could be applied (996: $G = 1.821$, $P = 0.177$; 5307: $G = 0.289$, $P = 0.591$; A4115: $G = 0.049$, $P = 0.825$; A4125: $G = 2.820$, $P = 0.093$; for 1343, 3196, 5239 and M491: Fisher exact test, $P = 1.0$).

We also tested the neutral model at each independent locus within each species using Tajima's *D* (Tajima 1989) and Fu and Li's *D* (Fu & Li 1993) statistics. Tajima's *D* statistic is proportional to the difference between two estimates of the population mutation parameter $4N$, the mean pairwise difference between the sampled sequences (π) and Watterson's estimator ($\hat{\theta}$) (Watterson 1975). Under a neutral model with constant population size, both estimators have the same expected value, and therefore the value of Tajima's *D* under neutrality is zero. Tajima's *D* was negative in most loci, but it was significantly different from zero only in locus 5239 in both species and in locus M491 just in *D. arizonae* (Table 3). Fu and Li's *D* statistic (Fu & Li 1993) is used to determine whether there is a significant excess of singletons (unique mutations) in the sample. Fu and Li's *D* was also negative in most cases and significantly different from zero in locus 5239 of *D. mojavensis*. In *D. arizonae*, however, Fu and Li's *D* was significantly negative in four loci (5239, 5246, M491 and 3196) (Table 3). These single locus tests then reject the neutral model in several loci. This could indicate either the recent action of natural selection at any of these loci, or could reflect a violation of the constant population size assumption of the tests (see below).

The constant population size assumption of the neutral model is rejected by the data. The average value of Tajima's *D* and Fu and Li's *D* across loci departs significantly from zero in both species. For *D. mojavensis* the observed mean values of both test statistics were more negative than all of the means found in 10 000 simulations (Tajima's $D = -0.663$, $P = 0.019$; Fu and Li's $D = -0.982$, $P = 0.038$). This was also true, but more extreme, for *D. arizonae* (Tajima's $D = -1.253$, $P < 0.0001$; Fu and Li's $D = -1.635$, $P < 0.0001$). The observation of a significant skewed mutation frequency spectrum across multiple loci of both species, suggests that *D. mojavensis* and *D. arizonae* are undergoing a population size expansion.

It is important to note that rejection of the constant population size assumption in our study was not caused by the

Table 4 The number of shared polymorphisms, fixed differences and estimated migration rate (Nm) between *D. mojavensis* and *D. arizonae*

		All lines			Sympatric *mojavensis*/sympatric *arizonae*		
Locus	Chromosome*	Shared	Fixed	Nm	Shared	Fixed	Nm
X100	X (inv.)	2	26	0.103	1	28	0.093
3196	X (inv.)	2	4	0.200	1	5	0.124
5246	2 (inv.)	2	7	0.130	0	12	0.066
5307	2 (inv.)	0	12	0.090	0	14	0.064
996	3 (inv.)	0	27	0.059	0	28	0.066
A4125	3 (inv.)	3	30	0.067	0	37	0.046
1343	4 (col.)	2	3	0.228	1	3	0.199
A4115	4 (col.)	1	4	0.138	1	7	0.117
5239	5 (col.)	3	3	0.188	3	6	0.242
M491	5 (col.)	4	1	0.385	0	1	0.400

Nm estimated using the method of Hudson *et al*. (1992). *It is indicated in parenthesis whether the chromosome has a fixed inverted region between species (inv.) or whether it is colinear (col.).

significant population subdivision detected in both species (see below). Although it has been shown that strong population subdivision can cause standard tests of neutrality to reject the null model with high probability (Nielsen 2001; Ingvarsson 2004), population subdivision tends to generate positive rather than negative values of the test statistics *D*, *F** and *D** (Simonsen *et al*. 1995). Thus, observation of significantly negative values is not likely to be caused by strong population structure. Thus, the deviations from neutrality we detected are likely the result of population expansion and not of population structure.

Tests of species divergence and historical introgression: lower differentiation in colinear regions of the genome

Table 4 shows the number of shared polymorphisms and fixed differences between all *D mojavensis* and *D. arizonae* strains and between their sympatric populations. This information was used to fit the multilocus data set to a model of species divergence with no gene flow (i.e. isolation model) (Wakeley & Hey 1997) (Table 5). The isolation model could not be rejected in any of the four comparisons we conducted, which involved either data from all strains of each species or just from sympatric strains. These results are robust and were not affected by the inclusion of data from loci that did not fit neutral expectations using single-locus tests (not shown), and suggest that there is little evidence of historical introgression between the species and that the observed shared variation is due either to ancestral polymorphism or homoplasy. We also conducted the linkage disequilibrium (LD) test of historical introgression proposed by Machado *et al*. (2002), that looks at patterns of LD among shared polymorphisms and its difference with LD among shared and exclusive polymorphisms. The test could only be applied to locus M491, which is the only one with the required minimum of shared polymorphisms (4; see Table 4). The null hypothesis of no gene flow at M491 could not be rejected using this test (*D. mojavensis*: $D' = -0.531$, $P = 0.990$; *D. arizonae*: $D' = 0.333$, $P = 0.095$).

Table 5 Results from fitting the data to the isolation model of species divergence

Species 1 Species 2	χ^2	$P\chi^2$	WWH	P_{WWH}
D. mojavensis *D. arizonae*	104.1	0.2539	33	0.5115
Sympatric D. mojavensis *D. arizonae*	115.5	0.1340	38	0.3330
D. mojavensis *Sympatric D. arizonae*	104.7	0.2139	33	0.5079
Sympatric D. mojavensis *Sympatric D. arizonae*	110.1	0.1501	39	0.3534

P values are not corrected for multiple tests.

However, as the test of the isolation model is very conservative, we explored other aspects of the data to determine whether there is remnant evidence of historical introgression between the two species at colinear regions of the genome. *D. mojavensis* and *D. arizonae*, have fixed inversion differences in chromosomes X, 2 and 3, while chromosomes 4, 5 and 6 are homosequential and thus could recombine in hybrid females. Therefore, we expect that loci in colinear chromosomes will show lower differentiation between species if there has been some gene flow during their divergence. It is important to note that for the comparisons shown below, there is a complication in that the current low resolution of the genome assembly and

Table 6 Population structure analyses using AMOVA

	D. mojavensis			*D. arizonae*		
Locus	Within†	Among‡	F_{ST}	Within†	Among‡	F_{ST}
X100	45.24	54.76	0.5476***	58.98	41.02	0.4102***
3196	56.53	43.47	0.4347***	80.03	19.97	0.1997**
5246	67.55	32.45	0.3244***	46.37	53.63	0.5363***
5307	55.88	44.12	0.4411***	99.57	0.43	0.0043
996	83.26	16.74	0.1674	72.04	27.96	0.2795***
A4125	56.42	43.58	0.4260***	71.92	28.08	0.2808**
1343	90.58	9.42	0.0941	28.18	71.82	0.7182***
A4115	81.74	18.26	0.1826*	76.47	23.53	0.2353
5239	84.64	15.36	0.1536*	57.84	42.16	0.4216***
M491	73.67	26.33	0.2632***	54.02	45.98	0.4598***
All loci§	70.05	29.95	0.2995***	59.11	40.89	0.4088***

†Percent of the total genetic variation due to within-population variation; ‡percent of the total genetic variation due to differences among populations; §locus A4115 was not included because of smaller sample size in *D. arizonae*. Total length was 8052 bp; *$P < 0.05$; **$P < 0.01$; ***$P < 0.001$.

physical map of *D. mojavensis* did not allows us to assign loci located in the chromosomes with fixed inversions to either the inverted or colinear regions of those chromosomes. Although on the surface this may seem problematic, it turns out that it leads to a more conservative test. If inverted regions are expected to show higher levels of differentiation than colinear regions, pooling loci from colinear and inverted regions, as carried out when pooling all loci from the inverted chromosomes, is likely to reduce any differences between colinear and inverted chromosomes, making our test conservative.

We standardized the number of fixed differences across loci based on the average divergence of both species to the outgroup *D. navojoa* and compared the corrected numbers across loci from chromosomes that have fixed inversions (X, 2, 3) and from colinear chromosomes (4, 5). The number of corrected fixed differences is significantly higher in chromosomes that have fixed inversions when all strains are compared (Mann–Whitney *U*-test; $Z = -2.08$, $P = 0.037$), although it is not significant for sympatric strains ($Z = -1.34722$, $P = 0.178$). This result provides some evidence of historical introgression, and is not explained by higher mutation rates in the inverted chromosomes than in the colinear chromosomes, because neither the average sequence divergence to the outgroup ($Z = -1.83$, $P = 0.07$) nor the levels of polymorphism in *D. mojavensis* ($\hat{\theta}$: $Z = 0.53$; $P = 0.59$; π: $Z = -0.42$, $P = 0.67$) or *D. arizonae* ($\hat{\theta}$: $Z = -0.53$; $P = 0.59$; π: $Z = -0.95$, $P = 0.33$) are significantly higher in chromosomes carrying the fixed inverted regions. In the latter comparison, nucleotide diversities for X-linked loci were corrected to reflect the expected differences in effective population size between X-linked and autosomal loci.

Table 4 also shows migration rates (Nm) estimated using F_{ST} values (Hudson *et al*. 1992). There are no significant differences between the Nm values estimated for all strains or for sympatric strains (paired Wilcoxon signed-rank test; $Z = 15.5$, $P = 0.125$). Interestingly, Nm values are significantly lower in loci from chromosomes that have fixed inversions (X, 2, 3) than in loci from the colinear chromosomes 4 and 5 (Mann–Whitney *U*-test; $Z = 2.02$, $P = 0.043$). This difference also holds when only sympatric strains are compared ($Z = 2.24$, $P = 0.025$). These results also suggest lower differentiation between *D mojavensis* and *D. arizonae* in colinear regions of the genome that could be the result of historical introgression. However, one may also consider the possibility that lower levels of shared variation in chromosomes with inverted regions, which may reduce Nm estimates, may reflect hitch-hiking if inversions were fixed by selection and some of the loci are located inside the inversions.

Evidence of strong population differentiation in D. mojavensis *and* D. arizonae

AMOVA analyses (Excoffier *et al*. 1992) reveal strong evidence of population structure in both *D. mojavensis* and *D. arizonae* throughout their geographical range (Tables 6 and 7). In *D. mojavensis*, all F_{ST} values are large, ranging from 0.0941 to 0.5476 (Table 6), and all are significant except at loci 996 ($F_{ST} = 0.1674$) and 1343 ($F_{ST} = 0.0941$). In *D. arizonae*, F_{ST} values range from 0.0043 to 0.7182, and are not significant only at A4115 ($F_{ST} = 0.2353$) and 5307 ($F_{ST} = 0.0043$) (Table 6). Lack of statistical significance at locus A4115 despite its large F_{ST} value could be the result of the smaller sample size of this locus (13 sequences) compared to the other loci surveyed in *D. arizonae*, as significance is determined by permutation of haplotypes, and thus this approach has low power to detect significance when sample sizes are small. To determine genetic differentiation

Table 7 Pairwise genetic differentiation among populations

D. mojavensis	Catalina (4)	Mojave (4)	Baja (3)	Sonora (8)
Catalina (4)	—	70.37	94.10	122.03
Mojave (4)	**< 0.00001**	—	86.77	186.51
Baja (3)	**< 0.00001**	**< 0.00001**	—	45.25
Sonora (8)	**< 0.00001**	**< 0.00001**	**0.001**	—
D. arizonae	Sonora (9)	Baja (1)	Mexico (4)	Riverside (3)
Sonora (9)	—	139.47	169.19	84.06
Baja (1)	**< 0.00001***	—	27.17	87.86
Mexico (4)	**< 0.00001**	0.0759	—	69.43
Riverside (3)	**< 0.00001**	**< 0.00001***	**< 0.00001**	—

Values above the diagonal in each half of the table are the combined probabilities ($-2\Sigma \ln P$) of the significance tests of pairwise differentiation at 10 unlinked loci for each of the species (comparisons including the Baja population of *D. arizonae* only include nine loci, because of lack of data for locus A4115). Values below the diagonal in each half of the Table are the *P* values from the significance tests (see Methods). Sample sizes per population shown in parentheses. Significant comparisons after Bonferroni correction ($P < 0.008$) are marked in bold. *We make no strong claims about the significance of these comparisons because of the minimal size of the Baja sample.

among population pairs within each one of the two species, pairwise probabilities for the 10 loci were combined (Sokal & Rohlf 1995). This meta-analysis indicated highly significant levels of population structure in both *D. mojavensis* and *D. arizonae* (Table 7). In *D. mojavensis*, all pairwise comparisons are significant, suggesting very strong population structure across all its geographical range. In *D. arizonae*, the Sonoran population shows strong differentiation with the other three populations of this species, and there is also significant differentiation between the Southern Mexican lines and the Riverside lines. Although we present results for comparisons including the Baja population of *D. arizonae*, we make no strong claims about their statistical significance because of the minimal size of this sample (a single line sampled).

We reconstructed a phylogeny using concatenated sequences from the nine loci (see Methods) to explore the level of differentiation among the surveyed populations of the two species (Fig. 2). The phylogeny shows a very clear separation between the two species, in agreement with results suggesting limited evidence of historical introgression (see below). Further, there is more clear phylogenetic differentiation among the four *D. mojavensis* populations than among the *D. arizonae* populations, although this is due to the lack of reciprocal monophyly of the Sonoran strains of *D. arizonae*. In *D. mojavensis*, the four populations form strongly supported monophyletic groups (bootstrap > 86%). Further, the Santa Catalina Island and Mojave Desert populations of *D. mojavensis* are highly supported as sister populations as are the Baja and Sonora populations, suggesting a strong Northwest/Southeast split in this species. In *D. arizonae*, the Southern Mexico and Riverside strains form strongly supported monophyletic clades that fall among the nine Sonora strains and the single Baja strain. The Sonora strains of *D. arizonae* do not form a reciprocally monophyletic group. Furthermore, among the non-Sonoran Mexican samples the sample collected in the southern state of Chiapas is quite different from the strains collected in Hidalgo in central Mexico.

Discussion

The history of divergence of D. mojavensis *and* D. arizonae

We conducted population genetic analyses of species divergence and population structure in the *D. mojavensis-D. arizonae* species pair using a multilocus sequence data set composed of 10 nuclear loci. No evidence of recent introgression was observed between the two species, and this conclusion did not change when sympatric and allopatric populations were analysed separately. The null model of speciation, a model in which divergence occurs without gene flow, could not be rejected (Table 5). These results are similar to recent findings in this system using sequences from loci located in three chromosomes (X, 2, 4) (Counterman & Noor 2006), where no evidence of introgression was detected and no differences in patterns of introgression were observed between colinear and inverted regions of the genome. Further, we did not observe instances of shared haplotypes among species, and we could infer strong differentiation between species from AMOVA analyses (not shown) or phylogenetic analyses of concatenated sequences that showed highly supported reciprocally monophyletic nodes for each species (Fig. 2).

Rejection of recent introgression is puzzling because these species show incomplete reproductive isolation in

the laboratory and are often found in the same cactus rots in nature in areas of sympatry, suggesting the presence of strong mechanisms of reproductive isolation in nature. In general, reproductive isolation is incomplete and asymmetrical in the laboratory. Sexual isolation between *D. arizonae* females and *D. mojavensis* males is almost complete. When matings occur, however, they produce five times as many offspring as the reciprocal cross, but hybrid males are sterile. Sexual isolation in the reciprocal cross, on the other hand, depends upon the source of the *D. mojavensis* females. Strong sexual isolation is only observed when female *D. mojavensis* are from Sonora, where they are sympatric with *D. arizonae.* Hybrid male sterility also depends upon the source of the *D. mojavensis* mothers (Reed & Markow 2004). While the greatest hybrid male sterility is observed when the *D. mojavensis* females are from Catalina Island, not all females from this population produce sterile sons, leading to the conclusion that the responsible genetic factors are not yet fixed.

Although we were able to reject historical introgression using a conservative test (Table 5), by examining five of the six chromosomes simultaneously we did detect an interesting pattern of divergence across chromosomes suggestive of older introgression between the species in colinear chromosomes. We observed significantly lower differentiation and significantly higher inferred levels of interspecific gene flow in colinear chromosomes (4 and 5) than in chromosomes harbouring fixed inverted regions (X, 2 and 3). This result is consistent with predictions from models of speciation in which inverted regions play an important role (Noor *et al*. 2001; Rieseberg 2001). Under those models, inverted regions, and regions close to them, are predicted to be less prone to introgress than colinear regions in part because crossover products typically fail to be recovered in inversion heterozygotes, but more importantly because inverted regions may disproportionately bear alleles under divergent selection for species-specific adaptations or conferring reproductive isolation, that, by definition, cannot introgress. Those models lead to expectations of higher divergence in inverted than colinear regions of the genome as we observed here. These predictions have been previously confirmed in numerous empirical studies (Rieseberg *et al*. 1999; Machado *et al*. 2002; Besansky *et al*. 2003; Machado & Hey 2003; Panithanarak *et al*. 2004; Feder *et al*. 2005; Machado *et al*. 2007). However, despite our observations, it is clear that introgression between these species has not occurred recently because of the strong phylogenetic differentiation between the species (Fig. 2) and the lack of shared or similar haplotypes in the sampled loci.

Interestingly, although speciation in the repleta species group of *Drosophila* has been associated with chromosome rearrangements (Wasserman 1963), the relationship of the inversions to the genes underlying reproductive isolation is unclear. Zouros (1982) observed that the genes controlling both prezygotic and postzygotic reproductive isolation between *D. mojavensis* and *D. arizonae* tend to reside in the colinear rather than the inverted chromosomes. More recently, however, Reed *et al*. (2006) found quantitative trait loci (QTL) affecting hybrid male sterility to be in *D. mojavensis* chromosomes 2, 3, and 5, two inverted and one colinear chromosome. Hence, these studies do not resolve the role of chromosomal inversions as the major force behind reproductive isolation in this species group. Alternatively, the pattern of high numbers of fixed differences between inverted chromosomes could be the result of an ancient inversion polymorphism not linked to the speciation event. In the history of 46 species of the repleta group to which *D. mojavensis* and *D. arizonae* belong, chromosomes 2 and 3 have participated in 103 and 18 inversion events, respectively (Wasserman 1963). Chromosomes 4 and 5, on the other hand, having only experienced four and 10, respectively, are comparatively inversion poor. Additionally, in *D. mojavensis*, inversion polymorphisms are only found in chromosomes 2 and 3 (Johnson 1980; Ruiz *et al*. 1990). Thus, the propensity to undergo and/or sustain inversions differs dramatically among the major chromosomes. Given the level of fixed inversions in chromosomes 2 and 3 and the known polymorphisms, it is likely that these chromosomes segregated for inversion polymorphism at the time of the split between *D. mojavensis* and *D. arizonae.* Therefore, it is plausible that the fixed inversion differences observed at chromosomes 2 and 3 are due to the alternative fixation of an ancient inversion polymorphism in the lineages leading to *D. mojavensis* and *D. arizonae* and that the inversions may not have played an important role in the process of species divergence. However, that scenario cannot explain our observation of lower differentiation in colinear chromosomes, which remains puzzling. Further studies with greater marker densities scattered across all chromosomes are thus needed to further resolve the role of inversions in the divergence of *D. mojavensis* and *D. arizonae*.

Population history, population structure and historical biogeography

Drosophila mojavensis and *D. arizonae* harbour significant genetic variation in the sampled loci, showing average levels of nucleotide variation similar or greater to those of *D. simulans* and *D. persimilis*, two species of *Drosophila* traditionally considered to be polymorphic (Kliman *et al*. 2000; Machado *et al*. 2002). More nucleotide variation is observed in *D. arizonae* than in *D. mojavensis* in this multilocus data set, but the difference is not statistically significant. Although this result agrees with previous observations from a pair of nuclear loci (*Adh-1* and *Adh-2*) (Matzkin & Eanes 2003; Matzkin 2004) and a recent survey of nuclear sequence variation (Counterman & Noor 2006), it does not

match recent results from an extensive population genetic study using the COI mitochondrial locus (Reed *et al*. 2006). In the latter study, *D. mojavensis* was found to harbour more than twice nucleotide variation than *D. arizonae*, reflecting a recent population history that has seemingly affected nuclear and mitochondrial loci differentially. Such observed discrepancy between levels of nuclear and mitochondrial variation illustrates the importance of using multilocus data sets for estimating population parameters and for inferring population history (Hey & Machado 2003).

Our multilocus nuclear data also show that *D. mojavensis* is a highly structured species that has at least four strongly differentiated populations (Baja, Sonora, Mojave, Catalina Island) (Table 7), in agreement with a recent study by Ross and Markow using four microsatellite loci (Ross & Markow 2006). The strong population structure inferred for *D. mojavensis* also agrees with recent results by Reed *et al*. (2006) using mitochondrial sequences (COI). However, while Reed *et al*.'s data showed lack of differentiation between the Baja and Sonora populations, the nuclear data presented here show that there is strong differentiation between those two populations. Previous studies also suggested a strong split between Baja and Sonora based on *Adh-1*/*Adh-2* or *Acp* sequences (Matzkin & Eanes 2003; Matzkin 2004; Wagstaff & Begun 2005). Further, in contrast to Reed *et al*. (2006), our phylogenetic analyses of *D. mojavensis* show that the closest relative to the Catalina Island population is the Mojave population and not the Baja and Sonora populations (Fig. 2). While we observe amounts of nuclear variation within Catalina that are similar to those observed in the other three populations, the COI data shows no variation whatsoever in that population suggesting a recent selective sweep or demographic event in this population that differentially affected the mitochondria. Sonora was suggested to be the place of origin of *D. mojavensis* based on the structure of the COI phylogeny (Reed *et al*. 2006). Further, Baja was also suggested as the place of origin or centre of diversity of *D. mojavensis* because of higher inversion polymorphism (Johnson 1980) and higher variation at nuclear loci (*Adh-1* and *Adh-2*) (Matzkin & Eanes 2003; Matzkin 2004) observed in flies from that region. Our data shows that there is indeed an increased level of nucleotide variation in Baja than in the other three populations (not shown). However, it is not possible to determine which one of the four geographical regions may have been the site of origin of this species or the specific order of split of the populations (see Fig. 2); further sampling and analyses are needed.

We also observed strong evidence of population structure in *D. arizonae* (Table 7), in contrast with previous molecular studies. For instance, the COI study of Reed *et al*. (2006) showed little evidence of population structure in this species although there was strong phylogenetic evidence of a deep division between populations from northwestern Mexico and southwestern United States and those from southern and eastern Mexico. Our results support those phylogenetic findings, but also reveal additional and significant genetic structure among regional *D. arizonae* populations. Two previous studies using allozymes (Hocutt 2000) and *Adh* sequences (Matzkin & Eanes 2003) did not detect population structure in *D. arizonae* since neither included the Southern Mexico population nor the recently discovered Riverside population. Based on the phylogeny presented in Fig. 2 we conclude that Sonora is the centre of diversity of this species, from which the Southern and Eastern Mexico and Riverside populations have split in two independent events.

A significant excess of rare mutations across the 10 loci was inferred in both species, leading to significantly negative values of Tajima's *D* and Fu and Li's *D* across all loci. This result suggests that *D. mojavensis* and *D. arizonae* have recently undergone or are undergoing a population size expansion. This result is not due to the strong population structure detected in both species (see Results). Rather, the inferred population expansion in both species likely reflects common historical events that have affected most biota in the Sonoran Desert and vicinities in the recent past. Pleistocene glaciation cycles have been proposed to be the driving force under inferred range expansions or contractions in Sonora and Baja (Nason *et al*. 2002; Smith & Farrell 2005). For instance, population genetic analyses of allozyme data from *Lophocereus schottii*, the senita cactus, suggest a clear northward expansion in Baja that contrasts with weaker evidence of expansion in Sonora (Nason *et al*. 2002; Dyer & Nason 2004). Further, population genetic and phylogeographical evidence also show that in *D. pachea*, a species of cactophilic *Drosophila* associated with *Lophocereus*, both peninsular and mainland populations have experienced recent population expansions (Hurtado *et al*. 2004; Pfeiler *et al*., in press). Finally, phylogeographical evidence of ancient northern expansions from multiple southern refugia in Sonora has recently been reported for the longhorn cactus beetle (*Moneilema gigas*) (Smith & Farrell 2005), a species not restricted to a single host. Our results therefore match observations in other plant and insect species, and further support the proposed scenario in which recent range expansions caused by Pleistocene glaciation cycles have occurred in phytophagous insects and their host plants. The fact that *D. mojavensis* and *D. arizonae* may have diverged just before (Matzkin & Eanes 2003), or during (Reed *et al*. 2006), the Pleistocene further suggest the plausibility of such biogeograpic history.

Implications for speciation studies in this system: intraspecific differentiation

A major unresolved issue in speciation is the earliest appearance of reproductive isolation. Of particular interest

Table 8 Host relationships and chemistry for *D. mojavensis* and *D. arizonae*

			Cactus chemistry	
Species	Population	Host plant	Triterpenes	Alkaloids
D. mojavensis	Baja	Agria (*Stenocereus gummosus*)	+	–
	Sonora	Organ pipe (*Stenocereus thurberi*), Agria (*Stenocereus gummosus*)	+	–
	Catalina	Prickly pear (*Opuntia* sp.)	–	+
	Mojave	Barrel (*Ferocactus cylindraceus*)	?	?
D. arizonae	Sonora	Cina (*Stenocereus alamosensis*), Citrus	+	–
	Baja	Opuntia	–	+
	Riverside	Citrus	–	–
	Mexico	Opuntia	–	+

Data from Kircher (1982).

is whether one category of isolating mechanism appears before others, for example, does sexual isolation appear before postzygotic isolation? Also unknown is at what level of genetic differentiation the first signs of reproductive isolation appear. Neither of these questions can be addressed, however, without knowing the levels of divergence among the populations of interest. For the *D. mojavensis-D. arizonae* species pair, we now have a more clear picture of the genetic relationships among populations of both species, and this framework may now be used for testing hypotheses about early events in speciation. In the case of these two species, reproductive isolation between them not only is incomplete, but also is dependent upon the population of each species being used in laboratory tests of reproductive isolation. These patterns are likely due to a combination of factors such as genetic divergence, ecology, and whether or not the populations come from regions of sympatry. Fortunately, in the case of this species pair, sufficient information exists to untangle at least some of the likely contributions of each of these factors. Below we discuss patterns of isolation at the intraspecific level for each species.

For *D. mojavensis* the four geographical populations exhibit significant genetic differentiation (Table 7), a result illustrated by the topology of the neighbour-joining tree (Fig. 2). The Catalina and Mojave populations, however, are in a separate lineage from the Baja-Sonora populations. If historical relationships are the primary predictors of reproductive isolation, the greatest isolation is expected between populations that belong to the two main lineages. On the other hand, if additional factors, such as host use or sympatry with *D. arizonae*, influence reproductive isolation, different patterns may be observed. In the case of the two major *D. mojavensis* clades, the Catalina Island population utilizes prickly pear (*Opuntia* sp.) and the California populations breed in barrel cactus (*Ferocactus cylindraceus*), while the Baja-Sonora clade populations both use columnar cacti of the genus *Stenocereus* (Table 8). With respect to host use, the Catalina Island and Mojave populations, whose hosts are the most distinctive, should exhibit the greatest isolation between each other as well as with the Baja and Sonora populations. At the same time, the Sonora populations are sympatric with *D. arizonae*. If sympatry is important in isolation, then sexual isolation and reinforcement, would be expected between the Baja and Sonora populations even though they are part of the same lineage.

Two types of reproductive isolation have been well studied among *D. mojavensis* populations: sexual isolation (Zouros & D'Entremont 1980; Markow *et al*. 1983; Markow 1991; Markow & Hocutt 1998) and postmating-prezygotic isolation (Baker 1947; Knowles & Markow 2001). Postzygotic isolation patterns are unknown. Significant sexual isolation among *D. mojavensis* populations is only observed between *D. mojavensis* from Sonora and Baja. Specifically, females from Sonora, where the two species are sympatric, discriminate against Baja males, consistent with reinforcement (Markow & Hocutt 1998). In conclusion, for *D. mojavensis* the presence of reinforcement in the Sonoran population, sympatric with *D. arizonae*, has had an effect on intraspecific sexual isolation, and neither host use nor genetic distance seem to have had an effect on sexual isolation among populations of this species. Studies of postzygotic isolation among populations of *D. mojavensis* are sorely needed. Hybrid male sterility is expected to be consistent with F_{ST}, while hybrid inviability may also depend upon the host material upon which the hybrids are reared.

Drosophila arizonae has not been studied as intensively as *D. mojavensis* with respect to either genetic relationships among its populations or potential intraspecific reproductive isolation. Insufficient information exists to permit speculation about the role of host use in the observed differentiation of *D. arizonae*. While it breeds in cactus of various species, including columnar and opuntia, it also has been reared from citrus (Vacek *et al*. 1979; Markow *et al*. 1999) suggesting that it is much more of a generalist than *D. mojavensis* (Table 8). At the same time, however, vicariant events in the Mexican mainland may have shaped the

diversification of *D. arizonae*. For example, the Sierra Madre Occidental likely provides a strong barrier to gene flow between eastern and western populations, while southern populations could be isolated from the north by the TransMexican Volcanic Belt. More intensive sampling of *D. arizonae* will be required to assess the impacts of these geological features on gene flow within species.

Given that the differentiation among *D. arizonae* populations is of a magnitude similar or greater to that observed among *D. mojavensis* where significant reproductive isolation is seen, *D. arizonae* could also be at the early stages of speciation. Preliminary data suggest that this may be the case. Knowles & Markow (2001) reported divergence in postmating–prezygotic interactions between a Baja and a Sonora strain of *D. arizonae*. Further, Massie (2006) detected significant sexual isolation between a strain of *D. arizonae* from Hidalgo, representing the southeastern clade and another from near Phoenix, Arizona, representing the northwestern clade. Additional, more comprehensive studies clearly are warranted to further explore the relationship between genetic divergence and isolation and the nature of the earliest isolating mechanisms to appear in this species.

Conclusions

From an extensive geographical sampling of *D. mojavensis* and *D. arizonae* we found no evidence for recent introgression between the two species. We cannot, however, rule out the existence of some historical introgression as we observed sharp differences in the pattern of variation between inverted and colinear chromosomes, a result that raises interesting questions about the role of inversions and inverted chromosomes in the differentiation of this species pair. This study further supported the existence of strong population structure in *D. mojavensis* and provided evidence of strong geographical structure in *D. arizonae*. While numerous studies of reproductive isolation have been performed with *D. mojavensis* populations, the finding of significant structure in *D. arizonae* suggests that this species will prove to be an equally valuable group for studies of incipient speciation.

Acknowledgements

We thank three anonymous reviewers and C. Schlötterer for their comments and suggestions to improve the manuscript. Kristin Sweetser helped with the data collection. L.M.M. was supported by an NIH Postdoctoral Excellence in Research and Teaching (PERT) Fellowship from the Center for Insect Science, University of Arizona. L.K.R. was supported by a University of Arizona NSF Integrative Graduate Education Research Training (IGERT) grant Genomics Initiative (DGE-0114420). Research was supported by funds from the University of Arizona and grants from the National Science Foundation to C.A.M. (DEB 0520535), and T.A.M. (DEB 9510645, DEB 0075312).

References

Altschul SF, Gish W, Miller W *et al*. (1990) Basic local alignment search tool. *Journal of Molecular Biology*, **215**, 403–410.

Baker WK (1947) A study of isolating mechanisms found in *Drosophila arizonensis* and *Drosophila mojavensis*. *University Of Texas Publications*, **4720**, 126–136.

Besansky NJ, Krzywinski J, Lehmann T *et al*. (2003) Semipermeable species boundaries between *Anopheles gambiae and Anopheles arabiensis*: evidence from multilocus DNA sequence variation. *Proceedings of the National Academy of Sciences, USA*, **100**, 10 818–10 823.

Counterman BA, Noor MA (2006) Multilocus test for introgression between cactophilic *Drosophila mojavensis* and *D. arizonae*. *American Naturalist*, **168**, 682–696.

Dyer RJ, Nason JD (2004) Population graphs: the graph theoretic shape of genetic structure. *Molecular Ecology*, **13**, 1713–1727.

Etges WJ, Johnson WR, Duncan GA *et al*. (1999) Ecological genetics of cactophilic *Drosophila*. In: *Ecology of Sonoran Desert Plants and Plant Communities* (ed. Robichaux R), pp. 164–214. University of Arizona Press, Tucson, Arizona.

Excoffier L, Smouse PE, Quattro JM (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes: applications to human mitochondiral DNA restriction data. *Genetics*, **131**, 479–491.

Feder JL, Xie X, Rull J *et al*. (2005) Mayr, Dobzhansky, and Bush and the complexities of sympatric speciation in Rhagoletis. *Proceedings of the National Academy of Sciences, USA*, **102**, 6573–6580.

Fellows DP, Heed WB (1972) Factors affecting host plant selection in desert-adapted cactophilic *Drosophila*. *Ecology*, **3**, 850–858.

Fu YX, Li WH (1993) Statistical tests of neutrality of mutations. *Genetics*, **133**, 693–709.

Heed WB (1978) Ecology and genetics of Sonoran Desert *Drosophila*. In: *Symposium on Genetics and Ecology: the Interface* (ed. Brussard PF). Springer-Verlag, Heidelberg, Germany.

Heed WB (1982) The origin of *Drosophila* in the Sonoran Desert. In: *Ecological Genetics and Evolution: the Cactus-Yeast-Drosophila Model System* (eds Barker JSF, Starmer WT), pp. 65–80. Academic Press Inc., New York.

Hey J, Machado CA (2003) The study of structured populations — new hope for a difficult and divided science. *Nature Reviews Genetics*, **4**, 535–543.

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

Hocutt GD (2000) *Reinforcement of premating barriers to reproduction between* Drosophila arizonae *and* Drosophila mojavensis. PhD, Arizona State University, Tempe, Arizona.

Hudson RR, Kreitman M, Aguade M (1987) A test of neutral molecular evolution based on nucleotide data. *Genetics*, **116**, 153–159.

Hudson RR, Slatkin M, Maddison WP (1992) Estimation of levels of gene flow from DNA sequence data. *Genetics*, **132**, 583–589.

Hurtado LA, Erez T, Castrezana S *et al*. (2004) Contrasting population genetic patterns and evolutionary histories among sympatric Sonoran Desert cactophilic *Drosophila*. *Molecular Ecology*, **13**, 1365–1375.

Ingvarsson PK (2004) Population subdivision and the Hudson-Kreitman-Aguade test: testing for deviations from the neutral model in organelle genomes. *Genetical Research*, **83**, 31–39.

Johnson WRJ (1980) *Chromosomal polymorphism in natural populations of the desert adapted species*, Drosophila mojavensis. PhD, University of Arizona, Tucson, Arizona.

Kircher H (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.

Kliman RM, Andolfatto P, Coyne JA *et al.* (2000) The population genetics of the origin and divergence of the *Drosophila simulans* complex species. *Genetics*, **156**, 1913–1931.

Knowles LL, Markow TA (2001) Sexually antagonistic coevolution of a postmating-prezygotic reproductive character in desert *Drosophila*. *Proceedings of the National Academy of Sciences, USA*, **98**, 8692–8696.

Machado CA, Hey J (2003) The causes of phylogenetic conflict in a classic *Drosophila* species group. *Proceedings of the Royal Society of London. Series B, Biological Sciences*, **270**, 1193–1202.

Machado CA, Kliman RM, Markert JA *et al.* (2002) Inferring the history of speciation using multilocus DNA sequence data: the case of *Drosophila pseudoobscura* and close relatives. *Molecular Biology and Evolution*, **19**, 472–488.

Machado CA, Haselkorn TS, Noor MA (2007) Evaluation of the genomic extent of effects of fixed inversion differences on intraspecific variation and interspecific introgression in *Drosophila pseudoobscura* and *D. persimilis*. *Genetics*, **175**, 1289–1306.

Markow TA (1991) Sexual isolation among populations of *Drosophila mojavensis*. *Evolution*, **45**, 1525–1529.

Markow TA, Hocutt GD (1998) Speciation in Sonoran Desert *Drosophila*: testing the limits of the rules. In: *Endless Forms: Species and Speciation* (eds Howard D, Berlocher SH), pp. 234–244. Oxford University Press, Oxford, UK.

Markow TA, Fogleman JC, Heed WB (1983) Reproductive isolation in Sonoran Desert *Drosophila*. *Evolution*, **37**, 649–652.

Markow T, Raphael B, Breitmeyer CM *et al.* (1999) Elemental stoichiometry of *Drosophila* and their hosts. *Functional Ecology*, **13**, 78–84.

Massie K (2006) *Sexual Isolation between* Drosophila Mojavensis *and* Drosophila Arizonae *MS*. University of Arizona, Tucson, Arizona.

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.

McDonald JH, Kreitman M (1991) Adaptive protein evolution at the Adh locus in *Drosophila*. *Nature*, **351**, 652–654.

Nason JD, Hamrick JL, Fleming TH (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 (1987) *Molecular Evolutionary Genetics*. Columbia University Press, New York.

Nielsen R (2001) Statistical tests of selective neutrality in the age of genomics. *Heredity*, **86**, 641–647.

Noor MA, Grams KL, Bertucci LA *et al.* (2001) Chromosomal inversions and the reproductive isolation of species. *Proceedings of the National Academy of Sciences, USA*, **98**, 12 084–12 088.

Panithanarak T, Hauffe HC, Dallas JF *et al.* (2004) Linkage-dependent gene flow in a house mouse chromosomal hybrid zone. *Evolution*, **58**, 184–192.

Pfeiler E, Erez T, Hurtado LA *et al.* (2007) Genetic differentiation in *Drosophila* pachea across multiple geographic scales as revealed by nuclear and mitochondrial markers. *Hereditas*, in press.

Reed LK, La Flamme BA, Markow TA (submitted) Genetic architecture of Male Sterility in *Drosophila*: Direct Analysis in F1 Hybrids.

Reed LK, Markow TA (2004) Early events in speciation: polymorphism for hybrid male sterility in *Drosophila*. *Proceedings of the National Academy of Sciences, USA*, **101**, 9009–9012.

Reed LK, Nyboer M, Markow T (2007) Evolutionary relationships of *Drosophila mojavensis* geographic host races and their sister species *Drosophila arizonae*. *Molecular Ecology*, **16**, 1007–1022.

Rieseberg LH (2001) Chromosomal rearrangements and speciation. *Trends in Ecology & Evolution*, **16**, 351–358.

Rieseberg LH, Whitton J, Gardner K (1999) Hybrid zones and the genetic architecture of a barrier to gene flow between two sunflower species. *Genetics*, **152**, 713–727.

Ross CL, Markow TA (2006) Microsatellite variation among diverging populations of *Drosophila mojavensis*. *Journal of Evolutionary Biology*, **19**, 1691–1700.

Rozas J, Sanchez-DeI, Barrio JC, Messeguer X *et al.* (2003) DNASP: DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics*, **19**, 2496–2497.

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.

Russo CAM, Takezaki N, Nei M (1995) Molecular phylogeny and divergence times of drosophilid species. *Molecular Biology and Evolution*, **12**, 391–404.

Saitou N, Nei M (1987) The neighbor-joining method: a new method for constructing phylogenetic trees. *Molecular Biology and Evolution*, **4**, 406–425.

Schneider S, Roessli D, Excoffier L (2000) *ARLEQUIN: A Software for Population Genetics Data Analysis*. Genetics and Biometry Lab, Department of Anthropology, University of Geneva, Switzerland.

Schöfl G, Catania F, Nolte V *et al.* (2005) African sequence variation accounts for most of the sequence polymorphism in non-African *Drosophila melanogaster*. *Genetics*, **170**, 1701–1709.

Simonsen KL, Churchill GA, Aquadro CF (1995) Properties of statistical tests of neutrality for DNA polymorphism data. *Genetics*, **141**, 413–429.

Smith CI, Farrell BD (2005) Range expansions in the flightless longhorn cactus beetles, *Moneilema gigas* and *Moneilema armatum*, in response to Pleistocene climate changes. *Molecular Ecology*, **14**, 1025–1044.

Sokal RR, Rohlf FJ (1995) *Biometry*, 3rd edn. W.H. Freeman, New York.

Swofford DL (1998) *PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods)*. Sinauer & Associates, Sunderland, Massachusetts.

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

Tamura K, Nei M (1993) Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Molecular Biology and Evolution*, **10**, 512–526.

Thompson JD, Gibson TJ, Plewniak F *et al*. (1997) The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Research*, **24**, 4876–4882.

Vacek DC (1979) *The Microbial Ecology of the Host Plants of* Drosophila Mojavensis. PhD, University of Arizona, Tucson, Arizona.

Vacek DC, Starmer WT, Heed WB (1979) Relevance of the ecology of citrus yeasts to the diet of *Drosophila*. *Microbial Ecology*, **5**, 43–49.

Wagstaff BJ, Begun DJ (2005) Comparative genomics of accessory gland protein genes in *Drosophila melanogaster* and *D. pseudoobscura*. *Molecular Biology and Evolution*, **22**, 818–832.

Wakeley J, Hey J (1997) Estimating ancestral population parameters. *Genetics*, **145**, 847–855.

Wang RL, Wakeley J, Hey J (1997) Gene flow and natural selection in the origin of *Drosophila pseudoobscura* and close relatives. *Genetics*, **147**, 1091–1106.

Wasserman M (1963) Cytology and phylogeny of *Drosophila*. *American Naturalist*, **97**, 333–352.

Wasserman M (1992) Cytological evolution of the *Drosophila repleta* species group. In: *Drosophila Inversion Polymorphism* (eds Krimbas CB, Powell JR), pp. 455–552. CRC Press, Boca Raton, Florida.

Wasserman M, Koepfer HR (1977) Character displacement for sexual isolation between *Drosophila mojavensis* and *Drosophila arizonensis*. *Evolution*, **31**, 812–823.

Watterson GA (1975) On the number of segregating sites in genetical models without recombination. *Theoretical Population Biology*, **7**, 256–276.

Zouros E (1982) On the role of chromosome inversions in speciation. *Evolution*, **36**, 414–416.

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

Carlos A. Machado is an Assistant Professor interested in population genetics, phylogenetics, plant-insect coevolution, and the genetic basis of speciation in insects. His main research systems are *Drosophila* and the fig/fig wasp mutualism. Luciano M. Matzkin is a PERT Postdoctoral Fellow interested in the genetical basis of adaptation focusing on the transcriptional and molecular variation of ecologically interesting organisms (cactophilic *Drosophila*). Laura K. Reed recently completed her Ph.D. on speciation in the cactophilic *Drosophila*. She continues her work on evolutionary genetics, focusing on *D. melanogaster* as a model for human Metabolic Syndrome. Therese A. Markow is a Regents Professor who has extensively worked on the ecology and evolution of cactophilic *Drosophila*.
