Molecular Ecology (2006) doi: 10.1111/j.1365-294X.2006.02941.x

Evolutionary relationships of *Drosophila mojavensis* geographic host races and their sister species *Drosophila arizonae*

L. K. REED, M. NYBOER and T. A. MARKOW
Department of Ecology and Evolutionary Biology, PO Box 210088, University of Arizona, Tucson, Arizona 85721, USA

Abstract

The cactophilic *Drosophila mojavensis* species group living in the deserts and dry tropical forests of the southwestern United States and Mexico provides a valuable system for studies in diversification and speciation. Rigorous studies of the relationships between host races of *D. mojavensis* and the relationships among the members of the species group (*D. mojavensis*, *Drosophila arizona*, and *Drosophila navojoa*) are lacking. We used mitochondrial CO1 sequence data to address the phylogenetics and population genetics of this species group. In this study we have found that the sister species *D. mojavensis* and *D. arizonae* share no mitochondrial haplotypes and thus show no evidence for recent introgression. We estimate the divergence time between *D. mojavensis* and *D. arizonae* to be between 1.91 and 2.97 million years ago. *D. arizonae* shows little structure in our population genetic analyses but there is phylogenetic differentiation between southeastern and northern populations of *D. arizonae*. *Drosophila mojavensis* shows significant population and phylogenetic structure across the four geographic regions of its distribution. The mitochondrial data support an origin of *D. mojavensis* on the mainland with early differentiation into the populations now found in the Mojave Desert and the Mainland Sonoran Desert and later colonization of the Baja Peninsula, in contrast to previous models. Also, the sister clade to *D. mojavensis*/*D. arizonae* includes *D. navojoa* and *Drosophila huaylasi*. By defining the genetic relationships among these populations, we provide a foundation for more sophisticated hypothesis testing regarding the timing of early speciation events and host switches in this species group.

Keywords: *Drosophila mojavensis*, mitochondrial CO1, Mojave Desert, phylogeography, population structure, Sonoran Desert

Received 27 November 2005; revision accepted 13 February 2006

Introduction

Among the unresolved issues in speciation genetics is the relative timing of the appearance of various reproductive isolating mechanisms relative to degree of genetic differentiation. A preponderance of speciation genetics studies have employed *Drosophila* and have focused primarily upon closely related pairs of species; estimating genetic differentiation from allozyme data and relating pre- and postzygotic isolation in species pairs with varying levels of divergence (Coyne & Orr 1989, 1997). This approach has been immensely informative regarding the relationship between genetic differentiation and the strength of a given isolating mechanism after speciation has occurred, but is unable to address questions about the earliest appearance of reproductive isolation relative to genetic differentiation. Detecting early events in speciation and relating them to genetic divergence requires that we examine populations of the same species that are just beginning to exhibit reproductive isolation.

An exceptional model system for such studies is the group of cactophilic *Drosophila* species endemic to North America, in particular *Drosophila arizonae* and its sister species *Drosophila mojavensis*. The two species have been the subject of numerous studies of interspecific and intraspecific reproductive isolation (reviewed in Markow & Hocutt 1998) and between them and their various geographic populations, provide a continuum of reproductive incompatibilities. For example, the degree and nature of

Correspondence: L. K. Reed, Fax: 520-626-3522; E-mail: laurak@email.arizona.edu

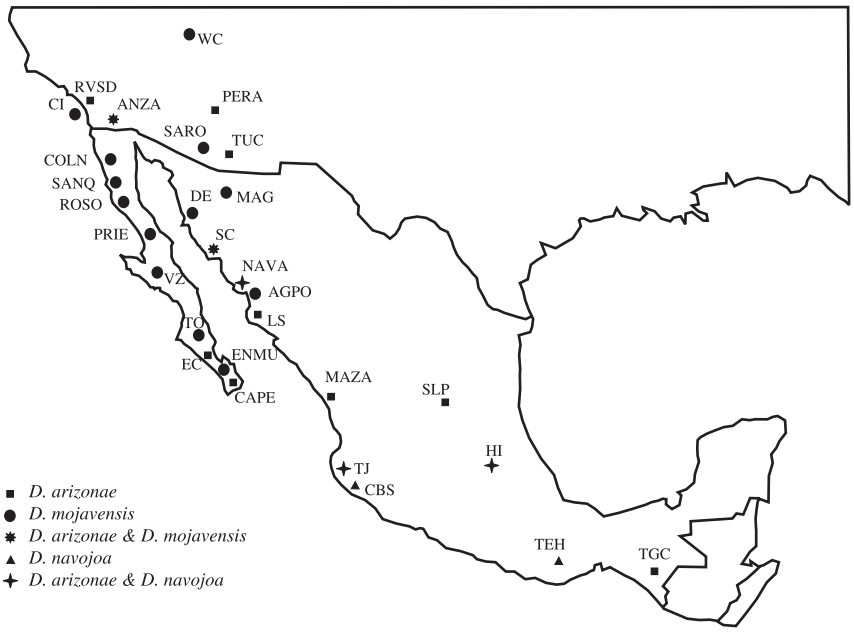

Fig. 1 Collection sites for populations and mass cultures used in this study. MAZA, Mazatlan; AGPO, Agiabampo; NAVA, Navajoa City; SC, San Carlos; DE, Desemboque; MAG, Magdalena; SARO, Santa Rosa Mountains; TUC, Tucson; ANZA, Anza Borrego; WC, Whitmore Canyon; COLN, Punte Colnett; SANQ, San Quintin; ROSO, El Rosario; PRIE, Punta Prieta; VZ, Vizcaino Desert; TO, Torete; EC, El Cien; CAPE, Cape; ENMU, Ensenada de los Muertos; LS, Los Mochis; HI, Hildalgo; PERA, Peralta Canyon; SLP, San Luis Potosi; TJ, Tomatlan; TGC, Tuxtla Gulierrez; CBS, Chamela Biological Station; TEH, Tehuantepec.

reproductive isolation between *D. arizonae* and *D. mojavensis* depend largely on the population of origin of *D. mojavensis* (Wasserman & Koepfer 1977; Ruiz *et al*. 1990; Reed & Markow 2004; Massie & Markow 2005). Second, geographic populations of *D. mojavensis* differ significantly in the strength of reproductive isolation that exists among them (Zouros & D'Entremont 1980; Markow 1991; Hocutt 2000).

The two species have different distributions within North America. *Drosophila arizonae* has a widespread distribution: from southwestern New Mexico, southern Arizona, and mainland Mexico all the way south into Guatemala (Heed 1982). It also has been found in increasing numbers on the tip of the Baja Peninsula (Markow, unpublished) and recently has been found breeding in citrus in the Anza Borrego desert and in Riverside, CA (Markow & Reed, unpublished). *Drosophila arizonae* uses the columnar cina cactus (*Stenocereus alamosensis*) as its primary host in the Sonoran Desert but has also been found using saguaro (*Carnegiea gigantea*), organ pipe (*Stenocereus thurberi*), pitaya agria (*Stenocereus gummosus*), and various opuntias (Ruiz & Heed 1988), and, as mentioned above, it has been found using citrus fruits as well. There is little information on its host use in the southern and eastern portions of its range. *Drosophila arizonae* can be considered a relative host generalist that is capable of using columnar and opuntia cacti hosts. It is sympatric with *D. mojavensis* in the Mexican states of Sonora and Sinaloa where they have some niche overlap (Ruiz & Heed 1988).

Drosophila mojavensis is found in four different geographic areas and utilizes a different host cactus in each (Fig. 1). Based upon population differences in colour and morphology (Mettler 1963), in alcohol dehydrogenase (ADH) allele frequencies (Zouros 1973; Heed 1982) and chromosomal inversion frequencies among its populations, *D. mojavensis* is considered to occur as two races: Race A (*Drosophila mojavensis mojavensis*) in the Mojave Desert of California, Race B (*Drosophila mojavensis baja*), from the Sonoran Desert in Baja California, southern Arizona, and Sonora. Based upon allozyme studies, Zouros (1973) further subdivided the Sonoran Desert or B race into B1 in Sonora, and B2 in the Baja California peninsula. Finally, an additional population was discovered on Santa Catalina Island off the coast of California. Based upon chromosome inversions, Ruiz *et al*. (1990) grouped the Santa Catalina Island population with the flies from the Mojave Desert, or Race A. There are reasons, however, to question the lumping of the Santa Catalina Island population with the Race A flies found in the Mojave Desert. First, an allozyme survey suggested that the Santa Catalina Island population was different from those in southern California (Hocutt 2000). Second, the nature of reproductive isolation between female *D. mojavensis* from Santa Catalina and male *D. arizonae*, is quite different from what is observed when the same cross is made with *D. mojavensis* females from the other localities (Ruiz *et al*. 1990; Reed & Markow 2004).

Populations of *D. mojavensis* are separated by areas in which there are no host plants and thus potentially reduced gene flow (Turner *et al*. 1995). For example, the Sea of Cortez, approximately 70 miles wide, separates the populations in Baja California from those in the mainland Mexican states of Sonora and Sinaloa. Similarly, the 26-mile Catalina Channel separates Santa Catalina Island from the mainland of California. At the same time that we have geographic barriers to gene flow between these regions, we see the utilization of different host cacti in each. These four geographically separated regional groups of *D. mojavensis*

each specialize on different local cactus hosts. Sonoran Desert populations use different host plants on the two sides of the Sea of Cortez. On the Mexican mainland, including southern Arizona (hereafter referred to as Mainland Sonoran Desert regional group), *D. mojavensis* is primarily associated with organ pipe cactus (*Stenocereus thurberi*) although it occasionally uses cina (*Stenocereus alamosensis*) (Ruiz & Heed 1988). In a small area near Desemboque, Sonora (DE), they utilize the few pitaya agria (*Stenocereus gummosus*) found there. In Baja California, these flies utilize agria (*Stenocereus gummosus*) almost exclusively, although organ pipe is abundant as are other occasional columnar hosts. Barrel cactus (*Ferocactus cylindraceus*), is the host for those populations found in the Mojave Desert and Grand Canyon, while on Catalina Island, prickly pear (*Opuntia* spp.) serves as the host because columnar cacti are absent.

Despite the importance of the *D. mojavensis* model system for speciation studies, and the need to place observed levels of reproductive isolation in the context of genetic divergence, existing levels of genetic differentiation among the four *D. mojavensis* populations has never been examined using high resolution techniques. All published relationships are inferred from chromosome polymorphisms and allozyme studies on limited samples, and these are subject to conflicting interpretations. In addition, for *D. mojavensis* a major contributor to the levels of allozyme-based differentiation is the ADH locus. Because this locus is likely to be under selection (Matzkin & Eanes 2003), the evolutionary relationships among the populations based upon allozyme studies may be biased. Thus, the degrees to which the different populations are genetically differentiated remain unknown.

Here we utilize sequence variation in the mitochondrial CO1 gene to ask the following questions: (i) What is the degree of genetic differentiation among the regional host areas of *D. mojavensis*? (ii) What is the degree of differentiation among regional populations of *D. arizonae*? and (iii) What are the evolutionary relationships among *D. mojavensis*, *D. arizonae*, and *D. navojoa*, the third member of the *mojavensis* group?

Materials and methods

Population samples

Population genetic variation was examined in two species of the *Drosophila mojavensis* species group: *D. mojavensis* and *Drosophila arizonae*. We collected 174 genetic individuals from 15 populations of *D. mojavensis* and 102 genetic individuals from eight populations of *D. arizonae* (Table 1, Fig. 1, Appendix S1). Genotyped individuals were either single flies from isofemale lines established from wild inseminated females or individual wild-caught males. All collections were made between October 2000 and May 2003.

Table 1 Populations and number of individuals collected

Populations	*Drosophila arizonae*	*Drosophila mojavensis*
Mainland Sononan Desert		
Mazatlan (MAZA)	1	
Agiabampo (AGPO)		6
Navojoa City (NAVA)	20	
San Carlos (SC)	10	9
Desemboque (DE)		8
Magdalena (MAG)		17
Santa Rosa Mountains (SARO)		24
Tucson (TUC)	12	
Mojave Desert		
Anza Borrego (ANZA)	1	9
Whitmore Canyon (WC)		12
Baja Peninsula		
Punta Colnett (COLN)		7
San Quintin (SANQ)		18
El Rosario (ROSO)		9
Punta Prieta (PRIE)		10
Vizcaino Desert (VZ)		13
Torete (TO)		13
El Cien (EC)	5	
Cape (CAPE)	35	
Ensenada de los Muertos (ENMU)		10
Other		
Riverside (RVSD)	18	
Santa Catalina Island (CI)		9
Total	102	174

For the phylogenetic analyses, we augmented the population samples with stocks of *D. arizonae, Drosophila navojoa*, and *Drosophila huaylasi* (Fontedvila *et al*. 1990) from the Tucson Drosophila Species Stock centre and mass cultures of *D. arizonae* from our own laboratory for locations where we were not able to take recent population samples (Table 2, Fig. 1, Appendix S2). *Drosophila huaylasi* was included in this study because it was found to group with the *D. mojavensis* species group in another phylogenetic study (Durando *et al*. 2000). At least two flies from each mass culture were sampled. These samples were not included in any population genetic analyses.

DNA data collection

We used the manufacture's protocol for the DNeasy kit (QIAGEN) to extract total DNA from single whole flies. A 710-bp fragment of the mitochondrial encoded CO1 gene was amplified by polymerase chain reaction (PCR) for each fly from raw genomic DNA. PCR primers (LCO and HCO) and PCR conditions are described in Folmer *et al*. (Folmer *et al*. 1994). PCR products were sequenced in both directions using the PCR primers in an ABI3700 analyser. Sequences were proofread and aligned using SEQUENCHER 4.1 (GeneCodes

Table 2 Species and localities of additional stocks from the Tucson *Drosophila* Species Stock Center

Species	Locality	Stock number
D. arizonae	Los Mochis, Sinaloa, Mexico (LS)	NA
D. arizonae	Navojoa Airport, Hildalgo, Mexico (HI)	15081-1271.07
D. arizonae	Peralta Canyon, Arizona, USA (PERA)	NA
D. arizonae	San Luis Potosi, Mexico (SLP)	15081-1271.06
D. arizonae	Tomatlan, Jalisco, Mexico (TJ)	NA
D. arizonae	Tuxtla Gulierrez, Chiapas, Mexico (TGC)	15081-1271.14
D. arizonae	Venados, Hildalgo, Mexico (HI)	15081-1271.05
D. navojoa	Chamela Biological Station, Jalisco, Mexico (CBS)	NA
D. navojoa	Navojoa, Sonora, Mexico (NAVA)	15081-1374.00
D. navojoa	Tehuantepec, Oaxaca, Mexico (TEH)	15081-1374.01
D. navojoa	Tomatlan, Jalisco, Mexico (TJ)	15081-1374.11
D. huaylasi	Quives, Peru	15081-1303.00

Corp.). Sequences were truncated to remove the primer binding sequence and ambiguous sites leaving a high quality 658-bp sequence for all individuals used in this study. Sequences were translated into amino acids and no stop codons or indels were found. They are also similar to CO1 sequences reported for other *Drosophila* species thus we are confident that our sequences represent the functional mitochondrial CO1 and not a pseudogene. We also re-amplified and sequenced any samples that had singleton haplotypes to rule out PCR or sequencing error producing excess variation.

Population genetic analyses

We calculated the following CO1 genetic diversity indices, as implemented by ARLEQUIN version 2.000 (Schneider *et al*. 2000), for each species and for the regions within species as defined by Table 1: heterozygosity (*h*) (equation 8.6 in Nei 1987), the mean number of pairwise differences (π) (Tajima 1983), and θ_S (Watterson 1975) (Table 3). ARLEQUIN 2.000 was also used to calculate all pairwise genetic differentiation (F_{ST}) values between populations of *D. arizonae* and between populations of *D. mojavensis* (Reynolds *et al*. 1983; Slatkin 1995) and significance was determined by permutation at the 0.05 and the Bonferroni corrected levels. The corresponding average number of historical migrants between populations (*Nm*) was calculated using the equation for haploid data $[(1/F_{ST}) - 1]/2$ (Wright 1951). This estimate is the number of migrants between two populations each generation required to generate the observed F_{ST} values assuming mutation–drift equilibrium within each population. Analysis of molecular variance (AMOVA) (Excoffier *et al*. 1992) was also calculated using ARLELQUIN 2.000 for *D. arizonae* and *D. mojavensis*. AMOVA identifies the amount of total genetic differentiation that can be attributed to differentiation between groups (regions) (F_{CT}),

Table 3 Diversity indices

Species	Region partitioned by F_{ST} groups	*N*	Haplotypes observed	Polymorphic sites	Heterozygosity (*h*)	Mean number of pairwise differences (π)	θ_S
D. arizonae	Overall	100	14	16	0.72 ± 0.03	2.15 ± 1.20	3.09 ± 1.05
	Baja Peninsula	40	7	10	0.59 ± 0.06	1.78 ± 1.05	2.35 ± 0.98
	Mainland Sonoran Desert (overall)	42	10	10	0.74 ± 0.06	2.32 ± 1.31	2.32 ± 0.97
	Mainland Sonoran Desert (minus Tucson)	30	8	9	0.59 ± 0.10	1.82 ± 1.08	2.27 ± 1.00
	Tucson	12	4	5	0.70 ± 0.09	1.20 ± 0.82	1.66 ± 0.93
	Riverside	18	2	1	0.47 ± 0.08	0.47 ± 0.43	0.29 ± 0.29
D. mojavensis	Overall	174	46	49	0.90 ± 0.01	6.31 ± 3.01	8.55 ± 2.21
	Mainland Sonoran Desert (overall)	64	18	24	0.80 ± 0.04	5.58 ± 2.72	5.08 ± 1.66
	Mainland Sonoran Desert (minus Magdelena)	47	16	21	0.87 ± 0.03	5.51 ± 2.70	4.76 ± 1.65
	Magdelena	17	4	13	0.33 ± 0.14	1.82 ± 1.10	3.85 ± 1.67
	Mojave Desert (overall)	21	7	6	0.70 ± 0.11	1.34 ± 0.87	1.67 ± 0.85
	Anza Borrego	9	1	0	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00
	Whitmore Canyon	12	7	6	0.89 ± 0.06	1.96 ± 1.19	1.99 ± 1.07
	Baja Peninsula (overall)	80	22	25	0.75 ± 0.05	1.81 ± 1.05	5.05 ± 1.60
	Baja Peninsula (minus San Quintin)	62	21	24	0.71 ± 0.06	1.64 ± 1.00	5.11 ± 1.68
	San Quintin	18	5	6	0.77 ± 0.05	1.61 ± 1.00	1.74 ± 1.00
	Catalina Island	9	1	0	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00

differentiation between populations within regions (F_{SC}), and within populations (F_{ST}).

Statistical parsimony haplotype networks of the unique haplotypes represented in the population genetic sample were constructed by TCS program (Clement *et al*. 2000).

Phylogenetic analyses

Phylogenetic analyses were conducted on all of the unique haplotypes from population genetic samples of *D. mojavensis* and *D. arizonae* as well as additional samples from laboratory-maintained samples for *D. arizonae*, *D. navojoa*, *D. hauylasi* (Tables 1 and 2). The outgroup sequence was *Drosophila nigrospiracula* (GenBank Accession no. AY533813) from Hurtado *et al*. (2004). In the phylogenetic analyses, each haplotype was represented by a single sequence (Appendices S1 and S2). Phylogenetic reconstructions were conducted using PAUP* 4.0 (Swofford 2000) and MRBAYES (Huelsenbeck & Ronquist 2001). PAUP was used to perform heuristic parsimony and maximum-likelihood searches. Heuristic searches were conducted using 100 independent random addition trees and tree-bisection–reconnection branch swapping. The parsimony trees were also confirmed with bootstrapping. Bayesian analysis using MRBAYES was conducted using six chains (1 cold, and 5 hot) and was run for 10 000 000 generations with a burn-in of 10 000 trees, sampling every 100 generations. Results of the original Bayesian analysis were confirmed with four additional analyses with four chains each, run for 6 000 000 generations with a burn-in of 6000 trees.

Since maximum-likelihood and Bayesian analyses suggested that *D. arizonae* was paraphyletic, parametric bootstrapping (Huelsenbeck *et al*. 1995, 1996; Swofford *et al*. 1996; Goldman *et al*. 2000) was conducted to test for the rejection of monophyly of *D. arizonae*. The likelihood models used for the parametric bootstrapping were estimated using both MODELTEST (Posada & Crandall 1998) in conjunction with PAUP* and DT-MODSEL (Minin *et al*. 2003) in conjunction with PAUP*. The reason for using both model selection methods is that it is argued (Minin *et al*. 2003) that the Akaike information criterion (AIC; Akaike 1974) and hierarchical likelihood ratio tests (LRT; Huelsenbeck & Crandall 1997) selection criterion used by MODELTEST tend to select overparameterized models while DT-MODSEL using the Bayesian information criterion (BIC) modified by inclusion of a decision-theory framework (Bernardo & Smith 1994) selection criterion, tends to select likelihood models with a more reasonable number of parameters. We wanted to be certain that the model selection criteria did not bias the results of our parametric tests. Once the likelihood models were selected, heuristic maximum-likelihood analyses under each model for the unconstrained and constrained trees were conducted using 20 random addition sequences in PAUP*, and the parsimony scores for each tree was calculated. MESQUITE version 1.05 (Maddison & Maddison 2004) was used to efficiently simulate data matrices on the constrained maximum-likelihood trees estimated by PAUP*, generate batchfiles for PAUP* to estimate the topology of those matrices with and without constraint using parsimony, and then to determine the distribution of differences in tree length between the constrained and unconstrained trees (Maddison 2004). When the difference in tree length of the constrained and unconstrained trees for the real data is compared to the distribution of differences from the simulated data, one can determine if there is significant support for rejection of monophyly in the real data. We choose to use monophyly as the null hypothesis due to the controversial phylogenetic species concept (Cracraft 1989) which states that species are by definition monophyletic. We were interested if the apparent paraphyly of *D. arizonae* was indeed inconsistent with the phylogenetic species concept.

Divergence times from important nodes on the phylogeny of these species was estimated using the average of all pairwise synonymous site changes (*Ks*) between haplotypes on each side of the node calculated in DNASP (Rozas *et al*. 2003). We choose to use average pairwise Ks values as opposed to more complex methods of divergence estimation such as net nucleotide divergence (Nei 1987; equation 10.21) or methods that account for population size (e.g. Arbogast *et al*. 2002), for two reasons. First, the phylogenetic dataset used for these analyses contained stock centre samples that could not be used in population-based analyses. And second, an accurate estimate of divergence estimates relative to absolute time is only available for this gene in *Drosophila* using *Ks*. A divergence rate of between 12.3% and 18.5% per million years was assumed based on the percent synonymous divergence at CO1 between *Drosophila melanogaster* and *Drosophila simulans* of 36.9% (Moriyama & Powell 1997) and the estimated divergence time between *D. melanogaster* and *D. simulans* of 2–3 million years (Lachaise *et al*. 1988).

Results

Population genetics

Diversity indices. We found a lower level of diversity in *Drosophila arizonae* than in *Drosophila mojavensis* (Table 3). We found 14 distinct haplotypes in our overall sample of 100 individuals of *D. arizonae* (GenBank Accession nos DQ383668–DQ383684) and 46 haplotypes out of 174 individuals of *D. mojavensis* (GenBank Accession nos DQ383685–DQ383730). *D. mojavensis* and *D. arizonae* had no haplotypes in common. The overall number of polymorphic sites, heterozygosity, pairwise differences, and theta were greater for *D. mojavensis* (49, 0.90, 6.31 and 8.55, respectively) than for *D. arizonae* (16, 0.72, 2.15 and 3.09, respectively).

Table 4 Pairwise F_{ST} values (below diagonal) and N_m (above diagonal) for *Drosophila arizonae*. Significant pairwise differences after bonferroni correction for multiple comparisons ($P < 0.0033$) shown in bold italics. * is N_m undefined and approaches panmixia

		Baja Peninsula		Mainland Sonoran Desert			Riverside
		EC	CAPE	NAVA	SC	TUC	RVSD
Baja Peninsula	EC	—	*	*	7.50	0.86	0.33
	CAPE	0.00	—	14.60	6.65	0.84	0.54
Mainland Sonoran Desert	NAVA	0.00	0.03	—	46.40	0.60	0.36
	SC	0.06	0.07	0.01	—	0.29	0.14
	TUC	***0.37***	***0.37***	***0.45***	***0.64***	—	1.45
Riverside	RVSD	***0.60***	***0.48***	***0.58***	***0.78***	***0.26***	—

Table 5 F_{ST} values (below diagonal) and N_m (above diagonal) for *Drosophila mojavensis*. Significant pairwise differences ($P < 0.05$) shown in bold and significance after Bonferroni correction for multiple comparisons ($P < 0.00048$) is shown in bold italics. * is N_m undefined and approaches panmixia

		Mainland Sonoran Desert					Mojave Desert		Baja Peninsula							Catalina Island
		AGPO	SC	DE	MAG	SARO	ANZA	WC	COLN	SANQ	ROSO	PRIE	VZ	TO	ENMU	CI
Mainland	AGPO	—	*	126.73	0.45	98.12	0.15	0.20	0.80	0.39	0.57	0.56	0.42	0.51	0.63	0.18
Sonoran	SC	0.00	—	9.49	0.34	*	0.13	0.18	0.48	0.29	0.37	0.36	0.29	0.34	0.41	0.16
Desert	DE	0.00	0.05	—	1.39	*	0.16	0.23	1.73	0.69	1.25	1.25	0.92	1.06	1.69	0.25
	MAG	**0.53**	***0.59***	**0.26**	—	0.78	0.05	0.09	1.20	0.57	1.32	1.85	1.16	1.20	4.62	0.10
	SARO	0.01	0.00	0.00	***0.39***	—	0.24	0.26	1.01	0.62	0.85	0.81	0.72	0.77	0.92	0.31
Mojave	ANZA	***0.77***	***0.79***	***0.75***	***0.91***	***0.68***	—	1.38	0.04	0.05	0.03	0.04	0.03	0.05	0.05	0.00
Desert	WC	***0.71***	***0.73***	***0.69***	***0.85***	***0.65***	**0.27**	—	0.10	0.08	0.08	0.08	0.07	0.09	0.09	0.05
Baja	COLN	**0.38**	**0.51**	0.22	**0.29**	**0.33**	***0.92***	***0.84***	—	10.61	*	15.64	10.40	*	11.21	0.07
Peninsula	SANQ	***0.56***	***0.63***	***0.42***	***0.47***	***0.45***	***0.91***	***0.86***	0.05	—	1.58	1.22	1.11	2.61	1.18	0.09
	ROSO	**0.47**	**0.58**	**0.29**	***0.27***	***0.37***	***0.94***	***0.86***	0.00	**0.24**	—	*	*	*	113.40	0.05
	PRIE	**0.47**	**0.58**	**0.28**	**0.21**	**0.38**	***0.93***	***0.86***	0.03	***0.29***	0.00	—	–18.79	159.24	*	0.06
	VZ	**0.54**	***0.63***	**0.35**	**0.30**	***0.41***	***0.95***	***0.88***	0.05	**0.31**	0.00	0.00	—	*	21.19	0.04
	TO	**0.49**	**0.59**	**0.32**	**0.29**	**0.39**	***0.92***	***0.85***	0.00	**0.16**	0.00	0.00	0.00	—	12.09	0.07
	ENMU	**0.44**	**0.55**	**0.23**	**0.10**	**0.35**	***0.91***	***0.84***	0.04	**0.30**	0.00	0.00	0.02	0.04	—	0.08
Catalina Island	CI	***0.74***	***0.76***	***0.67***	***0.84***	***0.62***	***1.00***	***0.91***	***0.87***	***0.85***	***0.91***	***0.89***	***0.92***	***0.87***	***0.86***	—

Genetic differentiation. *Drosophila arizonae* showed relatively little genetic differentiation between regions and populations (Table 4). There was no evidence for significant pairwise F_{ST} values within the Baja Peninsula, and the only significant pairwise F_{ST} values within the Mainland Sonoran Desert region was between the most northern population included, Tucson (TUC), and the other two populations. TUC also showed differentiation from the two Baja Peninsula populations. The recently discovered Riverside (RVSD) population showed significant differentiation from all other populations, presumably due to the high frequency of a Riverside unique haplotype (az47, Appendix S3). Analysis of molecular variation (AMOVA) for *D. arizonae* (Table 6) showed that there is significant differentiation between populations within regions ($F_{SC} = 0.344$) but not between regions ($F_{CT} = 0.020$). Most of the total variation (64.3%) is harboured within populations of *D. arizonae* (i.e. shared across all populations) while only 2.0% of the total variation is partitioned by region.

Drosophila mojavensis showed substantially greater population structure than *D. arizonae*. The pairwise population F_{ST} comparisons revealed significant differentiation within each of the regions having more than one population (Table 5). Within the Mainland Sonoran Desert region, one population showed significant differentiation from the others (MAG) as did one population within the Baja Peninsula (SANQ). The two populations from the Mojave Desert were significantly different from each other ($F_{ST} = 0.27$) at the 0.05 level. There were substantial differences between regions; all but one of the pairwise comparisons between the Baja Peninsula and the Mainland Sonoran Desert were significant at the either the 0.05 or the Bonferroni-corrected levels. The Mojave Desert populations

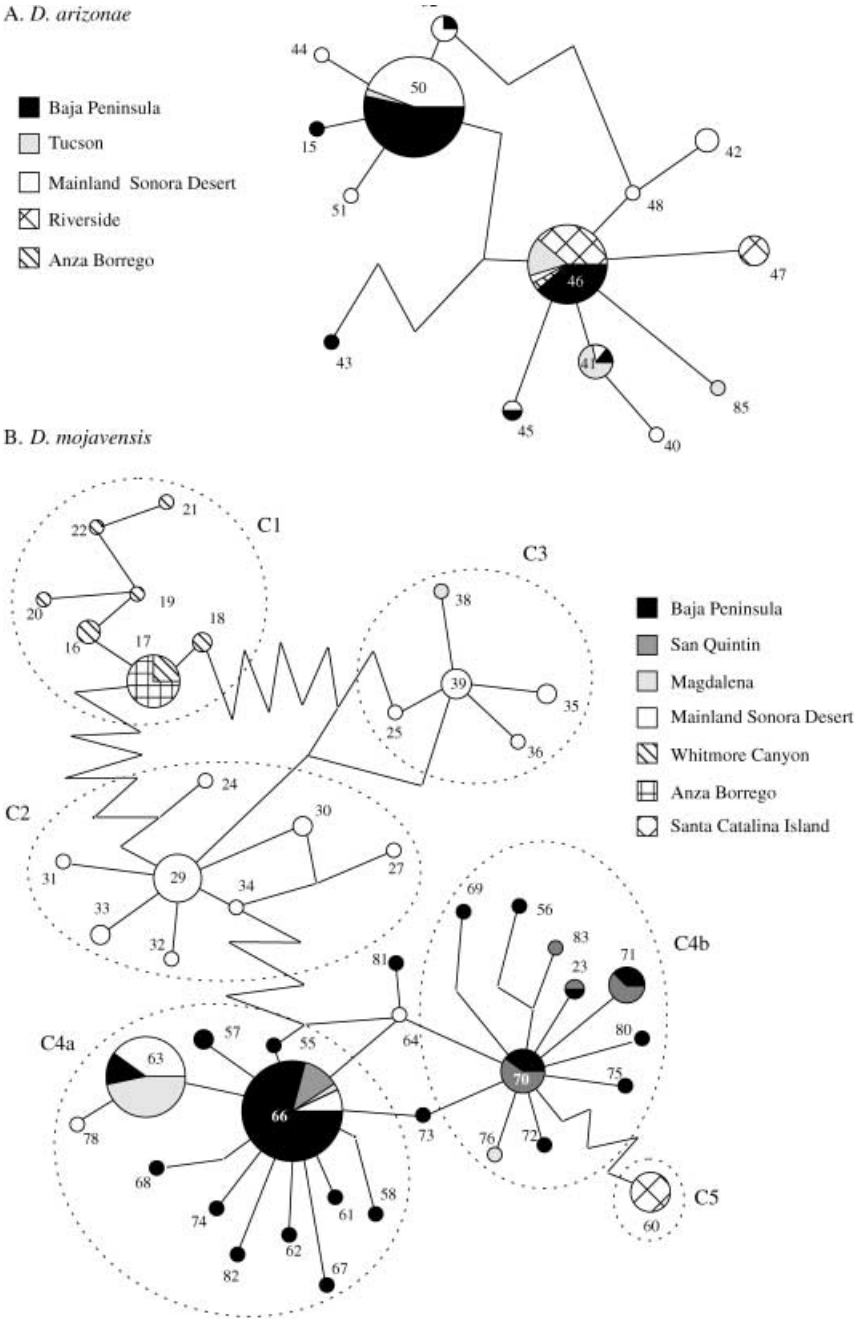

Fig. 2 Haplotype networks for *Drosophila mojavensis* (A) and *Drosophila arizonae* (B) for haplotypes found in wild-caught individuals. Circles at nodes are proportional in size to the number of individuals with that haplotype. Where one haplotype was found in more than one location the circle is divided into a pie diagram where the slices correspond to the proportion of individuals from a given location. Circles are shaded based on the location(s) where that haplotype is found. The origin of the haplotypes is designated by region or by population within a region if that population showed significant differentiation (F_{ST}, Tables 4 and 5) from other populations in the same region. Numbers at nodes indicate the haplotype designation. Each line segment represents one mutational step.

were significantly different from all other populations at the Bonferroni-corrected level and the one Santa Catalina Island population was significantly different from all others at the Bonferroni-corrected level. The especially high levels of differentiation between Santa Catalina Island and other populations (average $F_{ST} = 0.84$), and between the Mojave Desert populations (average $F_{ST} = 0.84$) and other regions is observed because each of these regions has only unique haplotypes not shared with any other region. There is lower average genetic differentiation between the Baja Peninsula and the Mainland Sonoran Desert (0.41) which makes sense due to some shared haplotypes between the regions and a close phylogenetic relationship among those haplotypes that are not shared. The AMOVA is consistent with the pairwise F_{ST}'s for *D. mojavensis* (Table 5). There is significant variation within populations ($F_{ST} = 0.683$, 31.7%), between populations within regions ($F_{SC} = 0.208$, 8.3%), and between regions ($F_{CT} = 0.60$, 60%).

Haplotype networks. For *D. arizonae* (Fig. 2a), the Mainland Sonoran Desert region is represented by two groups, labelled Tucson (just the TUC population) and Mainland

Sonoran Desert (NAVA and SC populations combined) because Tucson showed significant differentiation from the rest of the Mainland Sonoran Desert. One individual *D. arizonae* from Anza Borrego is also included in the haplotype network. For *D. arizonae*, the two most common haplotypes were az46 and az50. There is no obvious clustering of haplotypes by region with the exception of Riverside, where its two haplotypes only differ by one mutational step. Both Mainland Sonoran Desert and Baja Peninsula derived individuals were found with each of the common haplotypes and as well as ones within a few mutational steps of the common haplotypes.

The picture is quite different for *D. mojavensis* (Fig. 2B). The two populations from the Mojave Region (Whitmore Canyon and Anza Borrego) are represented separately since they showed significant differentiation from each other. Also, San Quintin is represented separately from the rest of the Baja Peninsula populations and Magdalena is represented separately from the rest of the Mainland Sonoran Desert populations for the same reason. The haplotype network for *D. mojavensis* shows, very clearly, five clusters of haplotypes. One cluster (C1) is made up exclusively of haplotypes derived from the Mojave Desert (Whitmore Canyon and Anza Borrego). C1 is separated from the next closest cluster by nine mutational steps. There are two clusters that contain only individuals from Mainland Sonoran Desert [the common haplotype of one is moj29 (C2) and the common haplotype of the other moj39 (C3)]. C2 is separated from the next cluster by six mutational steps. The fourth and largest cluster is composed of two subclusters; one (C4a) that contains the two most common haplotypes (moj63 and moj66) and is composed of individuals derived from both the Baja Peninsula (including San Quintin) and Mainland Sonoran Desert (including Magdalena), and the second (C4b) that contains the haplotypes moj70 and moj71 and is composed of individuals from primarily Baja Peninsula populations (including San Quintin) along with one individual from Magdalena. These two subclusters are completely linked by two extant, singleton haplotypes (moj73 and moj64). The final cluster (C5) is composed of only one haplotype (moj60), is derived entirely from Santa Catalina Island, and is six mutational steps away from the next closest (sub)cluster (C4b).

We estimate, based on average divergence between *D. mojavensis* and *D. arizonae* sequences derived from DNASP (Rozas *et al*. 2003), that about 20 mutational steps would separate the haplotype networks of the two species. The TCS program (Clement *et al*. 2000) is not capable of connecting the two networks over such a distance.

Tests for population expansion were conducted using Tajima's D (Tajima 1989), Fu's F_S (Fu 1996), and mismatch distribution (Harpending 1994; Schneider & Excoffier 1999). Neutrality could not be rejected generally and the details of the analyses are given in Appendix S4.

Phylogenetic analyses

Some, but not all phylogenetic patterns of the relationships in the *D. mojavensis* species cluster mitochondrial CO1 were robust to the method of analysis (parsimony, maximum likelihood, and Bayesian). The Bayesian results are shown in Fig. 3 and results of the parsimony analysis are shown in Fig. 4. Maximum-likelihood results are not shown due to their similarity to the Bayesian results. First, *D. huaylasi*, an enigmatic species described from Peru (Fontedvila *et al*. 1990), came out as the consistent sister taxon to *D. navojoa*. *D. navojoa* forms a monophyletic group, and along with *D. huaylasi*, is the sister group to *D. mojavensis* and *D. arizonae*. The relationship among *D. arizonae* lineages is less well resolved (which we will explore below) but one clear pattern in *D. arizonae* is that three haplotypes derived from Stock Center samples of collections for southeastern Mexico (az53, az54, and az92) form a monophyletic group on a long branch. *D. mojavensis* forms a monophyletic group and is composed of three major monophyletic clades. The first major *D. mojavensis* clade is composed of haplotypes from the Mojave Desert and is equivalent to C1 in the *D. mojavensis* haplotype network. The second major *D. mojavensis* clade is composed of only haplotypes found in the Mainland Sonoran Desert and is equivalent to the haplotype network clusters C2 and C3. C3 forms its own monophyletic group within this Mainland Sonoran Desert only clade. The last major monophyletic clade of *D. mojavensis* is composed of haplotypes found in the Mainland Sonoran Desert, all of the haplotypes found on the Baja Peninsula (C4), and the one haplotype found on Santa Catalina Island (C5).

Several relationships were less clear from the analyses. In the parsimony analyses, the consensus of 864 most parsimonious trees (MPT) found a tritomy, with two *D. arizonae* clades and one *D. mojavensis* clade, each with 100% support (Fig. 4). One of the *D. arizonae* clades is composed of the three southeastern Mexico haplotypes and the other is composed of the remaining *D. arizonae* haplotypes found in both southeastern and northern populations. Bootstrapping the parsimony analysis produces strong support (87.9%) for the monophyly of the southeastern Mexico clade and moderate support (80.3%) for the monophyly of the other *D. arizonae* clade. The maximum likelihood and Bayesian analyses tell a different story. They find that the three southeastern haplotypes form a monophyletic group that is the sister group to *D. mojavensis* with a Bayesian support value of 99%. The rest of the *D. arizonae* haplotypes fall out as many paraphyletic lineages basal to the Southeastern clade (Fig. 3). We were surprised by the suggestions in these analyses that *D. arizonae* may be paraphyletic and were also surprised that the Southeastern *D. arizonae* lineage might be the sister group to *D. mojavensis*. Thus, we conducted parametric bootstrapping analyses to test for

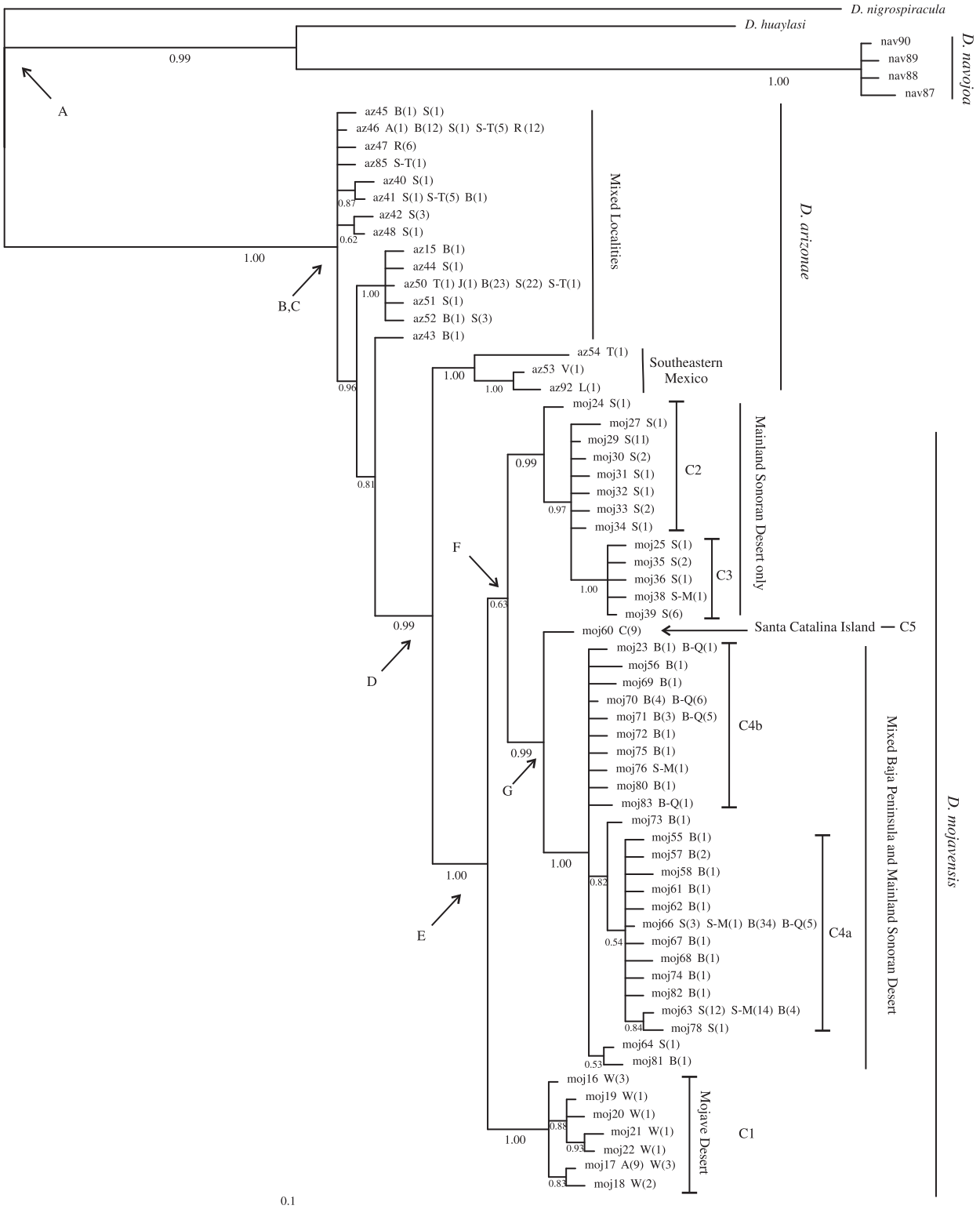

Fig. 3 Bayesian consensus tree of all haplotypes. Posterior probability values are shown below the branch for each clade. Letters indicate nodes at which date of divergence is estimated (Table 7). Haplotype designation given at each twig (e.g. moj57) followed by the localities and numbers of individuals from each locality having that haplotype (e.g. B(2)). S, Mainland Sonoran Desert (minus Magdalena and Tucson); S-M, Magdalena; S-T, Tucson; C, Catalina Island; B, Baja Peninsula (minus San Quintin); B-Q, San Quintin; T, Tuxtla Gulierrez; V, Venados; L, San Luis Potosi; J, Tomatlan; R, Riverside; W, Whitmore Canyon; A, Anza Borrego. Clades or haplotype groups that correspond to the clusters in Fig. 2 are given as C# and are discussed in the text. The model of evolution used in the Bayesian analysis GTR + γ + I, where codon positions were partitioned.

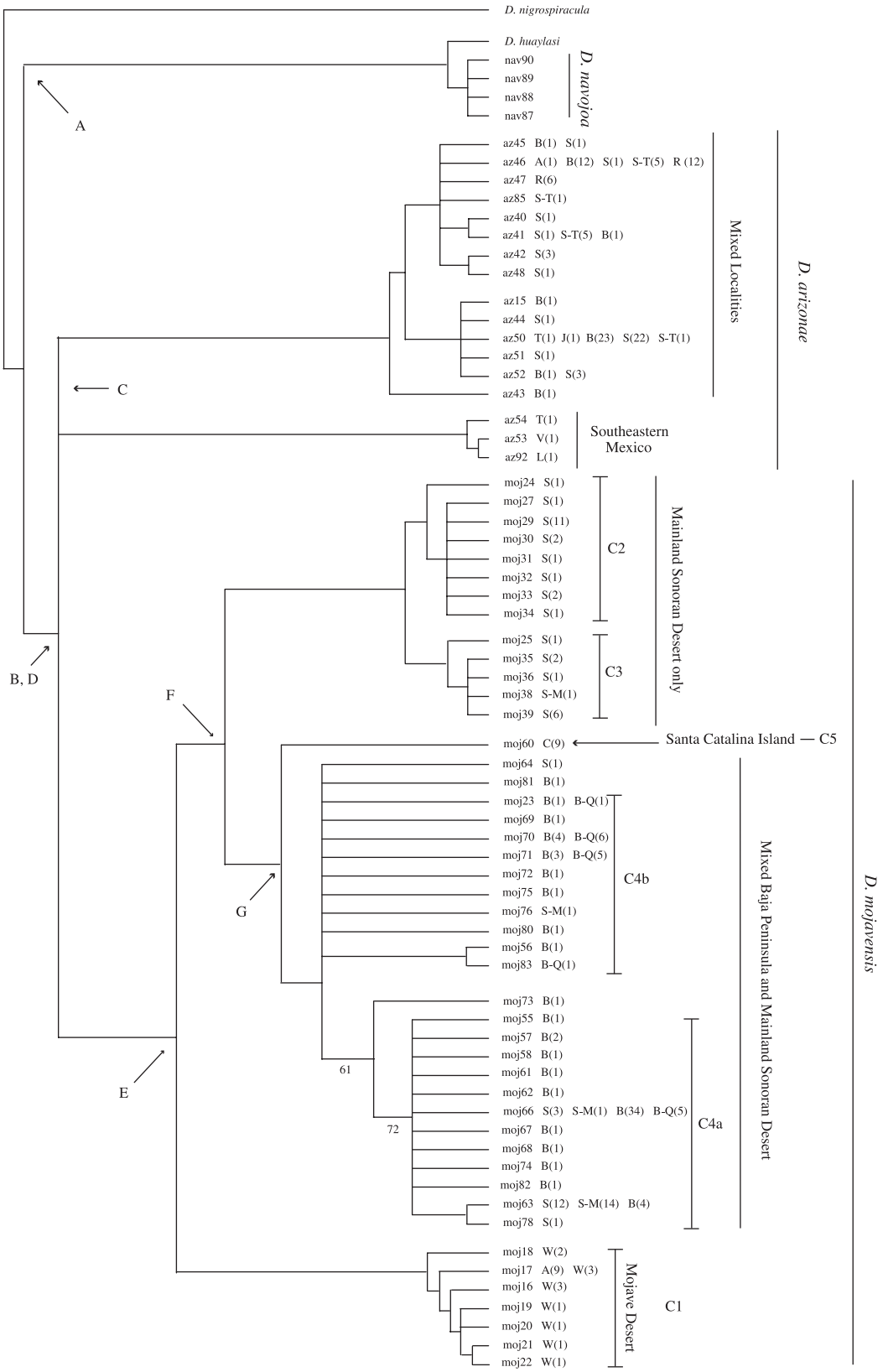

Fig. 4 Majority rule consensus tree of 864 most parsimonious trees. All support values are 100 unless otherwise noted. Letters indicate nodes at which date of divergence is estimated (Table 7). Haplotype designation given at each twig (e.g. moj57) followed by the localities and numbers of individuals from each locality having that haplotype [e.g. B(2)]. S, Mainland Sonoran Desert (minus Magdalena and Tucson); S-M, Magdalena; S-T, Tucson; C, Catalina Island; B, Baja Peninsula (minus San Quintin); B-Q, San Quintin; T, Tuxtla Gulierrez; V, Venados; L, San Luis Potosi; J, Tomatlan; R, Riverside; W, Whitmore Canyon; A, Anza Borrego. Clades or haplotype groups that correspond to the clusters in Fig. 2 are given as C# and are discussed in the text.

monophyly of *D. arizonae*. We estimated the likelihood models for DNA substitution on the neighbour-joining tree for the data, MODELTEST, using both LRT and AIC choose the Transversional model with invariant sites and a gamma shape parameter of among site rate variation (TVM + I + G) as the best fit of the data with base frequencies: A = 0.3162, C = 0.1517, G = 0.1463, T = 0.3858, substitution rates: A-C = 0.00, A-G = 13.58, A-T = 2.38, C-G = 0.00, C-T = 13.58, G-T = 1.00, proportion invariant sites (I) = 0.524, and gamma distribution shape parameter = 0.382. DT-MODSEL, using BIC, chose the less parameter rich model, HKY + I + G with base frequencies: A = 0.3289, C = 0.1398, G = 0.1356, T = 0.3956, transition–transversion ratio = 6.221, proportion invariant sites (I) = 0.3269, and gamma distribution shape parameter (G) = 0.1550. Each model was used to test two hypotheses of monophyly, one in which all *D. arizonae* haplotypes are forced to be monophyletic and a second where the haplotypes not included in the southeastern clade are forced to be monophyletic (as was suggested by the parsimony analysis). We found that regardless of the substitution model, the constrained trees were not significantly less likely than the unconstrained. The threshold for rejection of monophyly of *D. arizonae* was a difference in tree length of 10 and 7 for each model, respectively, when the actual difference was 4 and the threshold for rejection of monophyly of the nonsoutheastern haplotypes was a difference in tree length of 10 and 8 for each model, respectively, when the actual difference was 6. So, we cannot reject monophyly of *D. arizonae* despite the maximum likelihood and Bayesian suggestions of paraphyly.

The precise relationship among the three major clades of *D. mojavensis* differs between the methods of phylogenetic analysis. All three analysis methods (parsimony, maximum likelihood, and Bayesian) find some support for the initial split of the Mojave clade from the rest of the species, followed by a subsequent split between the Mainland Sonoran Desert-only clade and the Sonoran-Baja-Catalina Island clade. The Bayesian support value for the first split is fairly weak (63%) and, although the consensus of the 864 MPT finds 100% support for the first split, the bootstrap parsimony analyses finds a tritomy between the three major *D. mojavensis* clades. Thus, while there clearly are three major *D. mojavensis* clades, their exact relationships remain somewhat unresolved.

Of particular interest is the relationship of the *D. mojavensis* Santa Catalina Island haplotype to the rest of the Sonoran-Baja-Catalina clade. The Bayesian analysis finds that the first split within the clade is between Santa Catalina Island and the rest of the clade with a support value of 100%. The consensus of the 864 MPT also finds 100% support for that initial split. The bootstrap parsimony analysis, in contrast, finds that all 26 haplotypes in that clade collapse into a polytomy, so there is no information on the relative position of the Santa Catalina Island haplotype.

Divergence times. Significant nodes for which divergence times are estimated are indicated in Figs 3 and 4. Divergence times were based on a molecular clock and average *Ks* values of all pairwise comparisons across the node (Table 7). The first comparison (A) between *D. navojoa* and the *D. arizonae–D. mojavensis* clade estimates the nodes age as between 2.91 and 4.38 million years ago (Ma). The second comparison (B) is treating all of *D. arizonae* as a monophyletic group (as supported by parametric bootstrapping) and comparing it to *D. mojavensis* and the node's age is estimated at between 0.66 and 0.99 Ma, this value is inconsistent with the divergence time estimate of 2.4 million years from ADH sequence data (Matzkin & Eanes 2003). The third comparison is within *D. arizonae*, treating the three Southeastern haplotypes as one clade and the other *D. arizonae* haplotypes as the other monophyletic clade (as supported by parsimony and parametric bootstrapping) estimates the node's age at between 0.61 and 0.91 Ma. The fourth contrast (D) is comparable to B but is only comparing the *D. arizonae* haplotypes from southeastern Mexico to *D. mojavensis* (this node is supported by the Bayesian analysis) and is estimated to be 0.68–0.69 Ma. Nodes E, F and G are within the *D. mojavensis* clade. The Mojave clade verses all other *D. mojavensis* (E) is 0.45–0.68 Ma. The Mainland Sonoran Desert only clade verses the Sonoran-Baja-Catalina clade (F) node is 0.34 to 0.51 Ma. And the final comparison of the Santa Catalina Island haplotype to the rest of the Sonoran-Baja-Catalina clade (G) is 0.27 to 0.41 Ma.

Discussion

Within each species, *Drosophila arizonae* and *Drosophila mojavensis*, there is evidence of significant genetic differentiation. *Drosophila arizonae* has received less attention than has *D. mojavensis* with respect to studies of differentiation and reproductive isolation. Contrary to initial impressions based upon allozyme (Hocutt 2000) and ADH sequence data (Matzkin & Eanes 2003), however, *D. arizonae* exhibits significant genetic structure among a portion of the geographic populations sampled.

What is the degree of genetic differentiation among the regional host areas of D. mojavensis*?*

Significant genetic structure among *D. mojavensis* geographic host populations was expected, based upon earlier studies. Our data suggest, however, that the degree to which particular geographic host populations are differentiated, are more complex than what has been assumed (Wasserman & Koepfer 1977, 1980; Johnson 1980; Heed 1982; Ruiz *et al*. 1990; Etges *et al*. 1999). Most striking is that contrary to the suggestion by Ruiz *et al*. (1990) that populations from Santa Catalina Island and from the Mojave

Desert belong to the same race, the Santa Catalina Island population actually exhibits significant differentiation from all three of the other geographic regions and appears to be most closely related to Baja Peninsula and Mainland Sonoran populations. Despite being derived from the Baja/Mainland Sonoran clade the Santa Catalina Island population(s) appears to have been small and isolated for a long period of time since it exhibits no variation and a high degree of divergence from its closest relatives. Another surprise is that rather than being a recent arrival to the Mojave Desert, the populations from this area form a lineage that is basal to the other *D. mojavensis* populations, suggesting that it could have existed in the region for some time. Finally, while elevated chromosomal polymorphism in the Baja Peninsula led to the conclusion that this area served as the site of origin of *D. mojavensis*, phylogenetic analysis of mtDNA sequence data suggest that the Mexican mainland is more likely to be the area where *D. mojavensis* first arose.

It is difficult to disentangle the impact of geographic isolation vs. host use in the explaining patterns of differentiation among regional populations of *D. mojavensis*. Three of the host taxa for *D. mojavensis*, organ pipe (*Stenocereus thurberi*), agria (*Stenocereus gummosus*), and prickly pear (*Opuntia* spp.) have been characterized with respect to their chemistry (Kircher 1982) and major differences found to exist. *Ferocactus cylindraceus* has not been analysed for its chemical composition. Both species of *Stenocereus* contain triterpene glycosides but are low in alkaloids, while the opposite is true for *Opuntia*. If adaptation to these differences in host chemistry plays a role in differentiation of *D. mojavensis*, we would expect to see the greatest levels of genetic divergence between populations utilizing the least similar host plants. In fact, while all geographic host populations of *D. mojavensis* are significantly different from each other, the greatest differences are between the Santa Catalina population (which breeds in *Opuntia*) and the Mojave population (which breeds on *Ferocactus cylindraceus*) and the other regional populations. In contrast, the lowest F_{ST}'s are between the Baja and Sonora populations, which breed in the closely related species of *Stenocereus*. While such an observation does not prove that host chemistry drives of genetic divergence in *D. mojavensis*, it certainly is consistent with the predictions. The alternative explanation is that the physical isolation between regions allows for the genetic differentiation and there is simply higher gene flow between the Baja and Mainland Sonoran Desert populations, than among any other regional pairs.

What is the degree of differentiation among regional populations of D. arizonae*?*

Drosophila arizonae shows significant differentiation between its Sonoran Desert Tucson population and all other populations and between the Riverside population and all other populations. There is not, however, consistent evidence of isolation between the Baja Peninsula populations and the Mainland Sonoran Desert populations (Table 4). In the AMOVA, only 2% of the total variation can be attributed to regional groups and the majority (64.3%) of the genetic variation is contained within populations (Table 6). Unfortunately, we were unable to obtain population samples from the full distribution of *D. arizonae* (e.g. Southeastern Mexico and New Mexico), and thus we cannot determine if there is significant population genetic differentiation between those regions and the populations that were sampled. But, we can surmise that there is likely to be some significant population genetic differentiation between Southeastern Mexico and the Sonoran populations of *D. arizonae* due to the deep divergence between haplotypes from Southeastern Mexico found in Tucson Stock Center stocks and the haplotypes in the Sonoran population samples (Figs 3 and 4). These findings suggest that additional sampling be undertaken in those regions, as it is likely to reveal evolutionarily important patterns of differentiation in this species.

Table 6 Analyses of molecular variance for *Drosophila arizonae* and *Drosophila mojavensis* grouped by region. F_{CT} represents genetic differentiation between groups, F_{SC} represents genetic variation among populations within groups, and F_{ST} represents overall genetic variation among populations. Bold values indicate significant ($P < 0.05$) genetic differentiation

Source of variation	*D. arizonae*	*D. mojavensis*
Among groups	2.0%	60.0%
Among populations within groups	33.7%	8.3%
Within populations	64.3%	31.7%
F_{CT}	0.020	**0.600**
F_{SC}	**0.344**	**0.208**
F_{ST}	**0.357**	0.683

What are the evolutionary relationships among D. mojavensis, D. arizonae, *and* D. navojoa, *the third member of the mojavensis group?*

In an earlier study using a different population sample, Oliveira *et al*. (2003) reported evidence for introgression at the mitochondrial level, but we did not find this result in our sampling and analyses. It is possible that there has been ancient introgression between the species that might explain the grouping of *D. arizonae* from Southern Mexico with *D. mojavensis* in the Bayesian analyses with 99% support values, but that can also be explained as a more recent divergence of *D. mojavensis* from Southern Mexico populations than from more Northern populations. If introgression occurs or has occurred between the species more recently it is not reflected in our mitochondrial data. The discrepancy between these studies requires further analyses of

this system. The phylogenetic analyses show clearly that there is no sharing of haplotypes between the species. *Drosophila mojavensis* is monophyletic by all analyses and the parametric bootstrapping cannot reject the monophyly of *D. arizonae*; though, it is important to emphasize, that the actual data support paraphyly of *D. arizonae*. If there is no active introgression occurring between the species, then the reproductive barriers at work in areas of sympatry must be highly effective. Also, the evidence for reinforcement in sympatric populations (Wasserman & Koepfer 1977; Markow 1981; Massie & Markow 2005) is probably due to selection against hybridization in the past since barriers appear to be sufficient to prevent recent introgression. It is possible that there is active introgression in other portions of the genome and the permeability of the mitochondria is somehow reduced, perhaps due to mitochondria–nuclear interactions in hybrids (Reed *et al*., unpublished).

It has been argued that *D. mojavensis* and *D. arizonae* first began to speciate as allopatric populations separated by the Sea of Cortez (Wasserman & Koepfer 1977, 1980; Johnson 1980; Heed 1982; Ruiz *et al*. 1990; Etges *et al*. 1999). The population on the Baja Peninsula turned into *D. mojavensis* and the population on the Mainland turned into *D. arizonae*. There are two forms of genetic evidence from the nuclear genome that generally support this notion of a Baja Peninsula origin of *D. mojavensis*, the centres of diversity for both chromosomal inversions (Johnson 1980) and ADH sequence variation (Matzkin & Eanes 2003) for *D. mojavensis* is on the Baja Peninsula. Phylogenetic studies of nuclear sequence for these regions of *D. mojavensis* have not been performed. The phylogenetic evidence from the mitochondria though, does not support the origin of *D. mojavensis* on the Baja Peninsula. There are three major clades of *D. mojavensis*, only one of which has haplotypes found on the Baja Peninsula and that one clade also has representatives on the mainland. We believe that a more plausible explanation for the origin of *D. mojavensis* is that it occurred on the mainland with a very early lineage colonizing the Mojave Desert and another occupying in the Mainland Sonoran Desert, and followed by a later colonization of the Baja Peninsula.

Drosophila mojavensis could have originated on the Baja Peninsula as nuclear data suggests, or the nuclear data could be reflecting a complicated demographic and selective history that the mitochondria have resisted. A phylogenetic analysis of nuclear markers is needed to determine the actual evolutionary history of the nuclear genome. The second argument cited by others (Johnson 1980; Heed 1982) for *D. mojavensis*' origin on the Baja Peninsula is the preference of all populations of *D. mojavensis* for the Baja host plant, agria, and the presence of all 'secondary' hosts on the Baja peninsula (Johnson 1980). Using extant host plant distributions as evidence for a historical event is dangerous. The Sonoran Desert's distribution has been dynamic over the timescale of the evolution of *D. mojavensis* due repeated glaciation events during the Pleistocene (Van Devender 1990, 2002) and there is no reason to expect that the current Baja host, agria, could not have been at one time distributed throughout the Mainland Sonoran Desert, especially since agria is found now on the mainland at Desemboque, Sonora. *Drosophila mojavensis*' preference for agria could have just as easily evolved from historical use of agria on the mainland. If *D. mojavensis* originated on the mainland, then it either had to undergo sympatric speciation with *D. arizonae*, or *D. arizonae* had to be living somewhere else. From the data we have here, we cannot determine whether these two species speciated in sympatry but more extensive population sampling of *D. arizonae* throughout the rest of its range might provide some insight on this issue.

The third member of the *D. mojavensis* species group, *D. navojoa*, is diverged from the *D. mojavensis*/*D. arizonae* clade by 53.9% at silent sites (Table 7) estimating a 2.91- to 4.38-million-year divergence from these two species. Interestingly, *D. huaylasi*, the obscure species from Peru that has only been collected on one occasion, is grouping as a sister species to *D. navojoa*. In the analyses done by Durando *et al*. (2000) which included nuclear genes, *D. huaylasi* was found

Table 7 Estimates of divergence times. Node on phylogeny corresponds to labelled node in Figs 3 and 4. Mean *Ks* for all pairwise comparisons of haplotypes in the contrasting clades. Estimated time since last common ancestor assuming 0.123 (low) and 0.185 (high) synonymous changes per million years

Node on phylogeny	Contrast	Mean *Ks*	Range of time since last common ancestor (million years ago)	
			Low	High
A	nav vs. moj/az	0.5390	4.38	2.91
B	az vs. moj	0.1222	0.99	0.66
C	SE az vs. other az	0.1120	0.91	0.61
D	SE az vs. moj	0.1270	1.03	0.69
E	Mojave moj vs. other moj	0.0840	0.68	0.45
F	pure Sonoran moj vs. mixed moj	0.0630	0.51	0.34
G	CI moj vs. mixed Baja/Sonoran moj	0.0506	0.41	0.27

to be the sister group to *D. mojavensis*/*D. arizonae* species pair. *D. huaylasi* may be a key understanding the origin of the *D. mojavensis* species group and requires, if additional strains can be collected, further study of its ecology, genetics, and reproductive isolation from the other species.

Finally, *D. mojavensis* is one of four cactophilic species considered to be endemic to the Sonoran Desert. Evolutionary relationships among populations of the other three, *Drosophila nigrospiracula*, *Drosophila mettleri*, and *Drosophila pachea*, across similar geographic regions, have been examined previously using sequence variation in CO1 (Hurtado *et al*. 2004). With respect to genetic differentiation in the Sonoran Desert proper, *D. mojavensis* is most similar to *D. pachea*, in that they both show differentiation across the Sea of Cortez, while *D. nigrospiracula* and *D. mettleri* do not. The similarity between *D. mojavensis* and *D. pachea*, ends there, however. Unlike *D. mojavensis*, *D. pachea* does not switch host cacti between the two regions and is a poor disperser (Markow & Castrezana 2000). *D. mettleri* is the only other species of the four found on Santa Catalina Island, off the southern California coast. Like *D. mojavensis*, on the island it is associated with *Opuntia* owing to the absence of columnar cacti. Interestingly, it is the only population of *D. mettleri* that exhibits significant genetic differentiation from the other regions.

Summary

In this study we have found that the sister species *Drosophila mojavensis* and *Drosophila arizonae* do not share mitochondrial haplotypes and thus show no evidence for recent introgression. We estimate the divergence time between *D. mojavensis* and *D. arizonae* to be between 0.66 and 0.99 Ma. *D. arizonae* shows little population structure in our population genetic analyses but there is phylogenetic differentiation between southeastern and northern populations of *D. arizonae*. *D. mojavensis* shows significant population and phylogenetic structure across the four geographic regions of its distribution. We believe the mitochondrial data support an origin of *D. mojavensis* on the mainland with early colonization of the Mojave Desert and later colonization of the Baja Peninsula, in contrast to previous models. Also, the sister clade to *D. mojavensis*/*D. arizonae* includes *Drosophila navojoa* and *Drosophila huaylasi*. By defining the genetic relationships among these populations, we provide a foundation for more sophisticated hypothesis testing regarding the timing of early speciation events and host switches in this species group.

Acknowledgements

We thank Jeffery Good, Luis Hurtado, Carlos Machado, Mariana Mateos, Luciano Matzkin, Wendy Moore, Jeffery Oliver, Edward Pfeiler, Michael Worobey and Tom Van Devender and four anonymous reviewers for advice on this study and manuscript. We thank Sergio Castrezana, Tamar Erez, Luis Hurtado, Mariana Mateos, and Charles Ross for assistance in collecting specimens. We are also greatly indebted to the Santa Catalina Island Conservancy for collecting permits and logistical support in collecting specimens. National Science Foundation support for this work was provided by Interdisciplinary Research Training Group on Plant–Insect Interactions (DBI-9602249) to the University of Arizona and the Integrative Graduate Education and Research Traineeship in Evolutionary, Computational, and Molecular Approaches to Genome Structure and Function (DGE-0114420) to the University of Arizona, Doctoral Dissertation Improvement Grant (DEB 0414184) to L.K.R., and grants (DEB 95–10645 and DEB 0075312) to T.A.M.

Supplementary material

The supplementary material is available from http://www.blackwellpublishing.com/products/journals/suppmat/MEC/MEC2941/MEC2941sm.htm

Appendix S1 Haplotype distribution in populations of *Drosophila arizonae* and *Drosophila mojavensis*.

Appendix S2 Haplotype distribution in additional stocks.

Appendix S3 Species and population distribution for haplotypes.

Appendix S4 Demographic tests to population expansion.

References

Akaike H (1974) A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, **19**, 716–723.

Arbogast BS, Edwards SV, Wakeley J, Beerli P, Slowinski JB (2002) Estimating divergence times from molecular data on phylogenetic and population genetic timescales. *Annual Review of Ecology and Systematics*, **33**, 707–740.

Bernardo JM, Smith AFM (1994) *Bayesian Theory*. Wiley & Sons, New York.

Clement M, Posada D, Crandall KA (2000) TCS: a computer program to estimate gene geneologies. *Molecular Ecology*, **9**, 1657–1659.

Coyne JA, Orr HA (1989) Patterns of speciation in *Drosophila*. *Evolution*, **43**, 362–381.

Coyne JA, Orr HA (1997) 'Patterns of speciation in *Drosophila* revisited. *Evolution*, **51**, 295–303.

Cracraft J (1989) Speciation and its ontology: the empirical consequences of alternative species concepts for understanding patterns and processes of differentiation. In: *Speciation and its Consequences* (eds Otte D, Endler JA), pp. 28–59. Sinauer Associates, Sunderland, Massachusetts.

Durando CM, Baker RH, Etges WJ *et al*. (2000) Phylogenetic analysis of the repleta species group of the genus *Drosophila* using multiple sources of characters. *Molecular Phylogenetics and Evolution*, **16**, 296–307.

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

Excoffier L, Smouse P, Quattro J (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes:

application to human mitochondrial DNA restriction data. *Genetics*, **131**, 479–491.

Folmer O, Black M, Hoeh W, Lutz R, Vrijenhoek R (1994) DNA primers for amplification of mitochondrial cytochrome *c* oxidase subunit 1 from metazoan invertebrates. *Molecular Marine Biology and Biotechnology*, **3**, 294–299.

Fontedvila A, Wasserman M, Pla C *et al*. (1990) Description and evolutionary relationships of two species of the *Drosophila mulleri* cluster (Diptera: Drosophilidae). *Annals of the EntoMolecular Society of America*, **83**, 444.

Fu YX (1996) New statistical tests of neutrality in DNA samples from a population. *Genetics*, **143**, 557–570.

Goldman N, Anderson JP, Rodrigo AG (2000) Likelihood-based tests of topologies in phylogenetics. *Systematic Biology*, **49**, 652–670.

Harpending HC (1994) Signature of ancient population-growth in a low-resolution mitochondrial DNA mismatch distribution. *Human Biology*, **66**, 591–600.

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, Sydney.

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

Huelsenbeck JP, Crandall KA (1997) Phylogeny estimation and hypothesis testing using maximum likelihood. *Annual Review of Ecology and Systematics*, **28**, 437–466.

Huelsenbeck JP, Ronquist F (2001) MRBAYES: Bayesian inferences of phylogeny. *Bioinformatics*, **17**, 754–755.

Huelsenbeck JP, Hillis DM, Jones R (1995) Parametric bootstrapping in molecular phylogenetics: Applications and performance. In: *Molecular Zoology: Advances, Strategies, and Protocols* (eds Ferraris JD, Palumbi SR), pp. 19–45. Wiley-Liss Inc., New York.

Huelsenbeck JP, Hillis DM, Nielsen R (1996) A likelihood-ratio test of monophyly. *Systematic Biology*, **45**, 546–558.

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

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

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

Lachaise D, Cariou ML, David JR, Lemeunier F, Ashburner M (1988) Historical biogeography of the *D. melanogaster* species subgroup. *Evolutionary Biology*, **22**, 159–226.

Maddison DR (2004) *Testing Monophyly of a group of beetles. Study 1 in* MESQUITE*: A modular system for evolutionary analysis, Version 1.04.* http://mesquiteproject.org.

Maddison WP, Maddison DR (2004) MESQUITE*: A modular system for evolutionary analysis*, Version 1.05. http://mesquiteproject.org.

Markow TA (1981) Courtship behavior and control of reproductive isolation between *Drosophila mojavensis* and *Drosophila arizonensis*. *Evolution*, **35**, 1022–1026.

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

Markow TA, Castrezana S (2000) Dispersal in cactophilic *Drosophila*. *Oikos*, **89**, 378–386.

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

Massie KR, Markow TA (2005) Sympatry, allopatry and sexual isolation between *Drosophila mojavensis* and *D. Arizonae*. *Hereditas*, **142**, 51–55.

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

Mettler LE (1963) *Drosophila mojavensis* baja, a new form in the mulleri complex. *Drosophila Information Service*, **38**, 57–58.

Minin V, Abdo Z, Joyce P, Sullivan J (2003) Performance-based selection of likelihood models for phylogenetic estimation. *Systematic Biology*, **52**, 674–683.

Moriyama EN, Powell JR (1997) Synonymous substitution rates in *Drosophila*: mitochondrial versus nuclear genes. *Journal of Molecular Evolution*, **45**, 378–391.

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

Oliveira DCSG, Wintermute EH, Etges WJ *et al*. (2003) Hierarchical structure in the *Drosophila mojavensis* cluster (Diptera: Drosophilidae). *Hereditas*, **139**, 223–227.

Posada D, Crandall KA (1998) MODELTEST: testing the model of DNA substitution. *Bioinformatics*, **14**, 817–818.

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

Reynolds J, Weir BS, Cockerham CC (1983) Estimating coancestry coefficient: basis for a short-term genetic distance. *Genetics*, **105**, 767–779.

Rozas J, Sanchez-De-lBarrio JC, Messeguer X, Rozas R (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.

Schneider S, Excoffier L (1999) Estimation of past demographic parameters from the distribution of pairwise differences when the mutation rates vary among sites: application to human mitochondrial DNA. *Genetics*, **152**, 1079–1089.

Schneider S, Roessli D, Excoffier L (2000) ARLEQUIN*: A software package for population genetics data analysis, Version 2000*. Genetics and Biometry Laboratory, Department of Anthropology, University of Geneva, Geneva, Switzerland.

Slatkin M (1995) A measure of population subdivision based on microsatellite allele frequencies. *Genetics*, **139**, 457–462.

Swofford DL (2000) PAUP*: *Phylogenetic Analysis Using Parsimony (*and Other Methods)*. Sinauer Associates, Sunderland, Massachusetts.

Swofford DL, Olson GJ, Waddell PJ, Hillis DM (1996) Phylogenetic inference. In: *Molecular Systematics* (eds Hillis DM, Moritz C, Mable BK), pp. 407–514. Sinauer Associates, Sunderland, Massachusetts.

Tajima F (1983) Evolutionary relationship of DNA sequences in finite populations. *Genetics*, **105**, 437–460.

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

Turner RM, Bowers JE, Burgess TL, Hastings JR (1995) *Sonoran Desert Plants: an Ecological Atlas*. University of Arizona Press, Tucson, Arizona.

Van Devender TR (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* (eds Betancourt JL, Van Devender TR, Martin PS), pp. 134–165. University of Arizona Press, Tucson, Arizona.

Van Devender TR (2002) Environmental history of the Sonoran desert. In: *Columnar Cacti and Their Mutualists* (eds Fleming TH, Valiente-Banuet A), pp. 3–24. University of Arizona Press, Tucson, Arizona.

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

Wasserman M, Koepfer HR (1980) Does asymmetrical mating preference show the direction of evolution? *Evolution*, **34**, 1116–1124.

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

Wright S (1951) The genetical structure of populations. *Annals of Eugenics*, **15**, 323–354.

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

Laura K. Reed has recently completed her Ph.D. in the laboratory of T.A. Markow on the evolutionary genetics of hybrid male sterility in the species pair *Drosophila arizonae* and *D. mojavensis*. She is generally interested in the evolutionary genetics of complex phenotypes. Therese A. Markow and members of her lab study the ecology and population biology of desert-adapted *Drosophila*, with special interests in speciation and the evolution of mating systems.
