Genetic differentiation and demographic history in *Drosophila pachea* from the Sonoran Desert

EDWARD PFEILER[1], TAMAR EREZ [2], LUIS A. HURTADO[2,3] and THERESE A. MARKOW[2]

[1]*Centro de Investigación en Alimentación y Desarrollo, A.C., Unidad Guaymas, Guaymas, Sonora, Mexico*
[2]*Department of Ecology and Evolutionary Biology, University of Arizona, Tucson, Arizona USA*
[3]*Present address: Department of Wildlife and Fisheries Sciences, Texas A&M University, College Station, Texas, USA*

Pfeiler, E., Erez, T., Hurtado, L. A. and Markow, T. A. 2007. Genetic differentiation and demographic history in *Drosophila pachea* from the Sonoran Desert. — *Hereditas 00*: 000–000. Lund, Sweden. eISSN 1601-5223. Received February 9, 2007. Accepted March 2, 2007

Genetic variation at six microsatellite DNA loci and a segment of the mitochondrial cytochrome oxidase subunit I (*COI*) locus was used to estimate gene flow, population structure, and demographic history in the cactophilic *Drosophila pachea* from the Sonoran Desert of North America, a species that shows a strict association with its senita host cactus (genus *Lophocereus*). For microsatellite analyses, thirteen populations of *D. pachea* were sampled, five in mainland Mexico and the southwestern USA, and eight on the Baja California (Baja) peninsula, covering essentially the entire range of the species. Analysis of molecular variance (AMOVA) of microsatellite data revealed that populations from both the mainland and the Baja peninsula generally showed little structure, although there were a few exceptions, suggesting some local differentiation and restriction of gene flow within both regions. Pairwise comparisons of F_{ST} among each of the mainland and Baja populations showed evidence of both panmixia and population subdivision. AMOVA performed on grouped populations from both the mainland and Baja, however, revealed significant partitioning of genetic variation among the two regions, but no partitioning among localities within each region. Bayesian skyline analyses of the *COI* data set, consisting of four mainland and seven peninsular populations, revealed population expansions dating to the Pleistocene or late Pliocene in *D. pachea* from both regions, although regional differences were seen in the estimated timing of the expansions and in changes in effective population size over time.

Edward Pfeiler, Centro de Investigación en Alimentación y Desarrollo (CIAD), A.C., Unidad Guaymas, Carretera al Varadero Nacional km 6.6, Apartado Postal 284, Guaymas, Sonora MX-85480, Mexico. E-mail: epfeiler@asu.edu

Many species of *Drosophila* have distributions and ecologies which should promote substantial differentiation at the intraspecific level. This is especially true of the cactophilic *Drosophila* species endemic to the deserts of North America (HEED 1978, 1982). There are four of these species, each having specialized on the necrotic tissues of a different columnar cactus host. One of the four, *D. pachea*, is particularly attractive for studies of genetic differentiation at multiple geographic scales. Throughout essentially its entire range in the Sonoran Desert, *D. pachea* has an obligate relationship with senita (*Lophocereus schottii*) because of its nutritional dependence upon schottenol, a sterol unique to this cactus species (HEED and KIRCHER 1965). Furthermore, few other species of *Drosophila* have the ability to detoxify the particular alkaloids found in senita (FOGLEMAN and DANIELSON 2001). Only in a highly-localized region near Punta Conejo, in the southwestern part of the Baja California (Baja) peninsula (Fig. 1), does a host shift occur. In this region, *D. pachea* utilize the local senita, *L. gatesii*, a species endemic to Baja California Sur (LINDSAY 1963). Chloroplast and mitochondrial DNA studies, however, have called into question whether the morphologically distinct *L. schottii* and *L. gatesii* are in fact distinct species (HARTMANN et al. 2002; but see NASON et al. 2002).

In *D. pachea*, the opportunity for genetic differentiation exists at multiple spatial scales. First, the range of *D. pachea* is bisected by the ~120 km wide Gulf of California (“Gulf” and “Gulf of California” are used interchangeably here; also known as the Sea of Cortez; Fig. 1), which provides a potentially major barrier to gene flow. Second, within the range of *D. pachea* on each side of the Gulf, the senita host cactus exists in morphologically distinct varieties (LINDSAY 1963). *Lophocereus schottii* var. *schottii* is the most prevalent, found in most of the Baja peninsula and from southern Arizona, USA to northwestern Sonora, Mexico on the mainland. In southern Sonora to northern Sinaloa, *L. schottii* var. *tenuis* replaces *L. s.* var. *schottii*, and in portions of the peninsular cape region, *L. s.* var. *australis* replaces *L. s.* var. *schottii*. While varieties *tenuis* and *australis* of *L. schottii*, as well as *L. gatesii*, have not been characterized chemically, the possibility remains that the *D. pachea* breeding in them may have undergone some degree of local adaptation. Recent allozyme

DOI: 10.1111/j.2007.0018-0661.01997.x

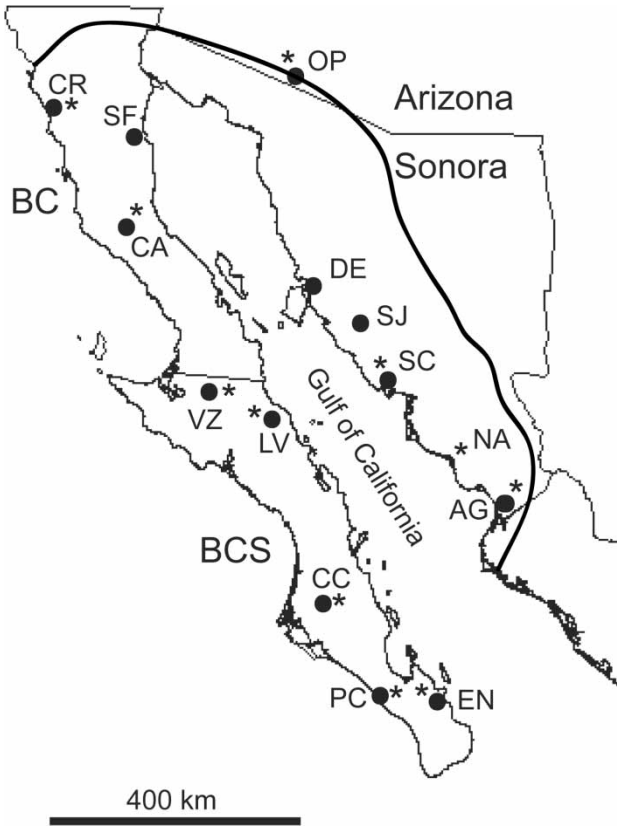

Fig. 1. Map showing collecting localities for *Drosophila pachea* used for microsatellite analyses (solid circles). Abbreviations: (mainland) AG, Agiabampo; SC, San Carlos; SJ, San Juanico; DE, Desemboque; OP, Organ Pipe Cactus National Monument; (Baja Peninsula) EN, Ensenada de los Muertos; PC, Punta Conejo; CC, Ciudad Constitución; LV, Las Vírgenes; VZ, Vizcaino Desert; CA, Cataviña; CR, Rancho Costa Rica; SF, San Felipe. Mitochondrial DNA (*COI*) sequence data (Hurtado et al. 2004) were also obtained for *D. pachea* from all the above localities (marked with asterisks), except DE and SJ on the mainland, and SF on the Baja Peninsula. In addition, *COI* data were obtained for flies from Navojoa, Sonora (NA) on the mainland, a locality not sampled for the microsatellite study. BC, Baja California; BCS, Baja California Sur. The curved thick line represents the approximate distribution limits of the senita host cactus.

studies, however, have called into question the distinctiveness of the three traditional varieties of *L. schottii*, suggesting the presence of only two phylogroups, one inhabiting the Baja peninsula and the other the mainland (Nason et al. 2002). The studies of Nason et al. (2002) also provided evidence of a postglacial northward population expansion of *L. schottii* from southern refugia in Baja, but not on the mainland. Finally, compared to other cactophilic species in the Sonoran Desert, *D. pachea* exhibits the most limited dispersal behavior, such that gene flow among regions is expected to be reduced (Markow and Castrezana 2000).

The possibility that these factors could promote genetic differentiation among populations of *D. pachea* has been borne out by some genetic studies, but the observations appear to be dependent upon the genetic markers used and the spatial scale examined. Across approximately 500 km of the Mexican mainland, no population structure was detected with allozymes (Rockwood-sluss et al. 1973; Pfeiler and Markow 2001), although the existence of a cline in chromosome inversion polymorphism (Ward et al. 1974) suggests some local adaptation within this area. A subsequent study utilized mitochondrial cytochrome oxidase subunit I (*COI*) sequence variation to test for genetic differentiation across the entire range of *D. pachea*, including seven localities on the Baja peninsula (Hurtado et al. 2004), and the only significant structure detected was between populations on opposite sides of the Gulf of California.

Microsatellite loci are both highly polymorphic and unlikely to be under selection (but see Nielsen et al. 2006), making them more useful nuclear markers than allozymes for examining population genetic structure. Microsatellites were developed for *D. pachea* (Ross et al. 2003), six of which were utilized here to search further for evidence of population genetic structure across the entire range of this species. We were also interested in determining if regional differences in demographic history of *D. pachea* could be detected, both within and among peninsular and mainland populations, which might be associated with the differences in demographics of the obligate senita host reported from the two regions (Nason et al. 2002). Neutrality tests and mismatch distribution analyses of *COI* sequence data have previously provided evidence for population expansions in *D. pachea* from both regions (Hurtado et al. 2004), but analysis of changes in population size through time has not been investigated. Therefore, we also report here results of Bayesian skyline analyses of *COI* sequences in *D. pachea* from both Baja and the mainland, which estimate a posterior distribution of effective population size through time utilizing a Markov chain Monte Carlo (MCMC) sampling of sequence data (Drummond et al. 2005).

MATERIAL AND METHODS

Collection of flies

For the microsatellite analyses we sampled 542 individuals of *D. pachea* from 13 localities (Fig. 1, Table 1) during March–April 2000. On the mainland, samples were taken from five localities that covered the region from southern Arizona, USA (Organ Pipe Cactus National Monument), which represents the

Table 1. *Localities from which Drosophila pachea were sampled and the number of flies collected. Localities are listed from south to north. BC, Baja California; BCS, Baja California Sur.*

Collecting localities	Abbreviation	Total no. of flies	Female	Male
Mainland				
Agiabampo, Sonora	AG	53	34	19
San Carlos, Sonora	SC	39	15	24
San Juanico, Sonora	SJ	31	23	8
Desemboque, Sonora	DE	42	35	7
Organ Pipe Cactus National Monument, Arizona	OP	36	22	14
Baja peninsula				
Ensenada de los Muertos, BCS	EN	27	13	14
Punta Conejo, BCS	PC	35	22	13
Ciudad Constitución, BCS	CC	92	64	28
Las Vírgenes, BCS	LV	29	15	14
Vizcaino Desert, BCS	VZ	62	15	47
Cataviña, BC	CA	40	40	0
Rancho Costa Rica, BC	CR	10	4	6
San Felipe, BC	SF	46	12	34
Total		542	314	228

northernmost part of the species range, to Agiabampo, Sonora, Mexico, near the Sinaloa border, ~700 km to the south and at the opposite extreme of the mainland distribution of *D. pachea.* On the Baja peninsula, samples were taken from eight localities from points along entire length of the species range, from San Felipe, in the northeastern corner of the state of Baja California, to Ensenada de los Muertos, ~900 km south in the southeastern section of Baja California Sur. Flies were keyed out, sorted by sex and frozen shortly after they were collected from their necrotic hosts. The *COI* sequence data of Hurtado et al. (2004) used for the demographic analyses were obtained from 203 individuals of *D. pachea* from the same collection as described above (Fig. 1).

Microsatellite genotyping

We followed the manufacturer's protocol for the DNeasy kit (Qiagen, Inc., Valencia, CA, USA) to extract total DNA from each individual. Whole flies were ground for DNA extraction. Six of the microsatellite loci developed by Ross et al. (2003) were chosen for use in this study (Table 2), two of which were dinucleotide repeats (AC13, AC19) and four of which were trinucleotide repeats (CAG9, CAG11, CAG12, CTG7). One primer for each locus was fluorescently tagged with either ABI dyes 6-FAM or HEX, and microsatellite loci were multiplexed during the polymerase chain reaction (PCR). For each 25 μl reaction, 1/50th of the DNA sample was added to a reaction mix (1 × PCR buffer [Invitrogen, Carlsbad, CA, USA], 1.5 mM $MgCl_2$, 0.4 μM of each primer, 0.5 units TAQ [Invitrogen], and 0.2 mM dNTPs). After an initial 3 min soak at 95°C, the multiplexed reaction was run 35 times through a temperature profile of 94°C for 20 s, 53°C for 45 s and 72°C for 90 s, followed by a 10 min extension at 72°C.

PCR products were diluted 2:3 with H_2O and genotyped using an ABI 3100 genetic analyzer (Applied Biosystems, Foster City, CA, USA) and the

Table 2. *Primer sequences and chromosome location for the six microsatellite loci used in this study.*

Locus	Primer sequence (5′–3′) (forward/reverse)	Chromosome
AC13	AAAGCCTAAATCAATATGCATCG CCCCAGCCACTTGTCTAGC	X
AC19	GCCAACAAGAGGCTTTAGACC GTTGCTGTCGAGTTCAATGC	?
CAG9	CTTCCGAAAATCCCACTGC ACTGTGGCTCGATGTTTGC	X
CAG11	CAGAAGCCCCTTATCTTATCG CAATGAACCATACAAAGGTTGG	?
CAG12	GGGCATGCAAATGAAAGG TGAAGTGGATTTGCGGTAGC	?
CTG7	GGAAATCGTCAAGCTGTGC TGTTGCTGTCTCGCCTTCC	X

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

Population genetic analyses

Genetic diversity for each locus and each population, as well as over all loci and populations, was quantified using Microsatellite Analyser (MSA) ver. 4.00 (Dieringer and Schlötterer 2003) or ARLEQUIN version 3.1 (Excoffier et al. 2005) by calculating a variety of parameters, including the number of alleles, expected number of alleles (SMM; Kimura and Ohta 1975), variance in allele size, expected heterozygosity (H_{exp}), and observed heterozygosity (H_{obs}). Deviations from Hardy-Weinberg equilibrium (HWE) were tested for each locus and over all loci in ARLEQUIN using a Markov chain approximation (Guo and Thompson 1992). All estimates were assessed for significance using a test analogous to Fisher's exact test with 100 000 steps in the Markov chain and 5000 dememorization steps. Gametic (linkage) disequilibrium tests to assess independence of loci were performed in ARLEQUIN using 10 000 permutations of the data matrix. Significance for all estimates was placed at the 0.05 level.

Inspection of the raw data, and preliminary analyses of the entire microsatellite data set using the MSA program (not shown), revealed that three of the loci we had selected, AC13, CAG9 and CTG7, were sex-linked. Males of *D. pachea* were hemizygous at these three loci, and for all populations in which males comprised at least 30% of the sample (Table 1), the three loci showed significant deviations from HWE, with the largest deviations present in populations where males comprised $>75\%$ of the sample (i.e. Vizcaino Desert and San Felipe). For all population genetic analyses, therefore, data for male flies were excluded. The trimmed data set was comprised of 314 female *D. pachea* (Table 1).

Analysis of molecular variance (AMOVA; Excoffier et al. 1992), performed in ARLEQUIN, was used to test for population structure in *D. pachea*. AMOVAs were conducted considering each locality separately, and then by grouping the localities into mainland and Baja peninsula. The calculation of significance (0.05 level) of pairwise comparisons of F_{ST} (Weir and Cockerham 1984; Michalakis and Excoffier 1996) was based on 10 000 permutations of the data matrix. Genetic variation was partitioned among localities relative to the total sample (F_{ST}), among localities within groups (F_{SC}), and among the mainland and Baja peninsula groups (F_{CT}). We report F_{ST} values rather than R_{ST} because estimates of the former have been shown to more reliably estimate population structure with microsatellite data sets because the variances associated with the R_{ST} values are generally high (Balloux and Goudet 2002; Balloux and Lugon-moulin 2002).

Demographic history

Bayesian skyline analyses of the *COI* data set (661 bp) were performed in BEAST ver. 1.2 (Drummond et al. 2005) using the GTR+I+G substitution model with four gamma categories. The number of grouped intervals (m) was set to ten. Five million iterations of the MCMC chains were run, sampling every 1000 iterations; the first 500 000 chains were discarded as burnin. The Bayesian skyline plots were generated with the program Tracer version 1.2.1. (Drummond et al. 2005).

The values generated for changes in effective population size over time in BEAST depend on a specified value for the mean mutation rate per site per generation for a particular gene. Because this rate has not been determined for *COI* in *D. pachea*, it was estimated from rates determined for other insects. The rate we used, 2.0×10^{-9}, was obtained assuming an average pairwise sequence divergence rate of 2.3% per million years (Brower 1994), with five generations per year. The generation time was based on laboratory observations of captive *D. pachea* held under fluctuating temperatures designed to reflect those seen in nature. Pairwise sequence divergence rates for the *COI* gene in some insects have been shown to range from ~ 0.6–1.5% per million years (Farrell 2001; Pfeiler et al. 2006), thus there is the possibility that the mutation rate we used for *D. pachea* was overestimated. If an average pairwise sequence divergence rate of 1.0% per million years is assumed, the time scale shown on the skyline plots (Fig. 2) is approximately doubled. Given the various assumptions, specific values shown in Fig. 2 should be considered only rough estimates, but nonetheless they are informative for interpreting differences in demographic trends among populations.

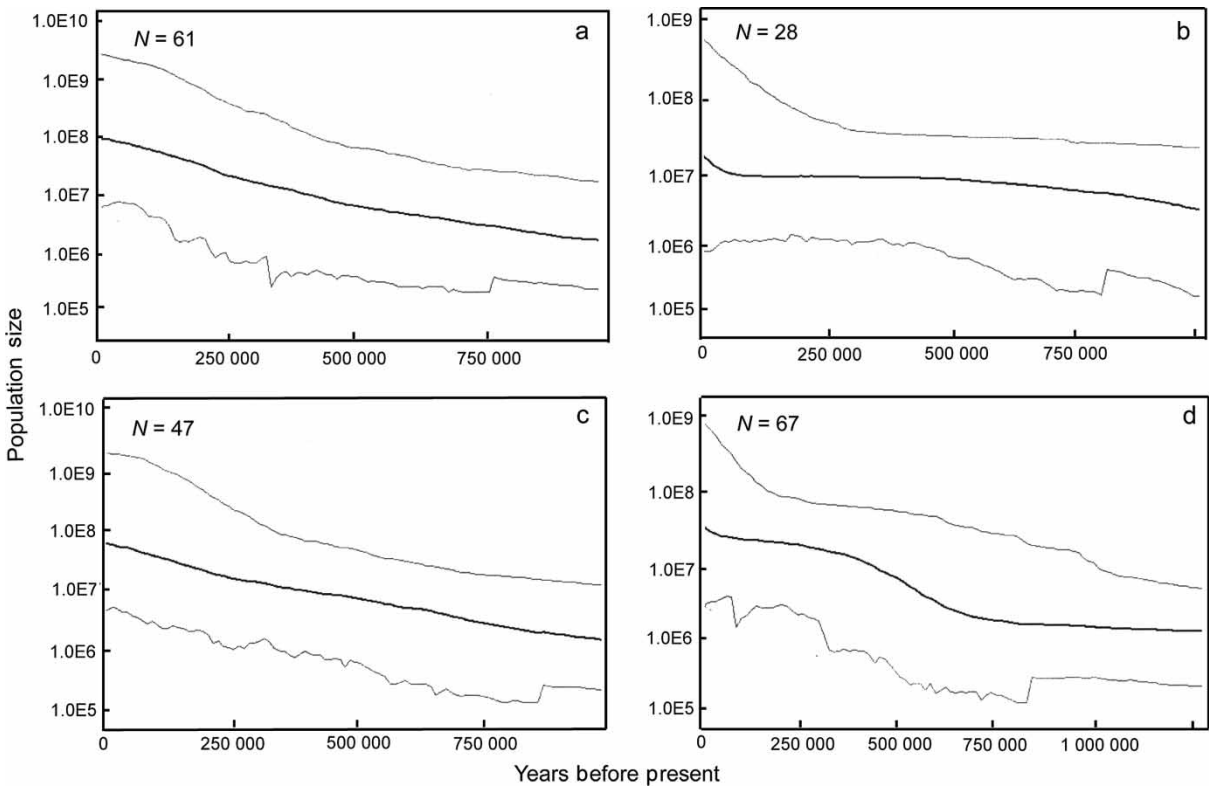

Fig. 2. Bayesian skyline plots showing changes in effective population sizes over time in populations of *Drosophila pachea* from the mainland (**a**) and the northern (**b**), central (**c**), and southern regions (**d**) of the Baja Peninsula. Plots were generated from aligned *COI* sequences (Hurtado et al. 2004) using the BEAST and Tracer computer programs (Drummond et al. 2005). *N* = total number of sequences from each region. Time estimates were obtained by assuming an average pairwise sequence divergence rate of 2.3% per million years for the *COI* gene in *D. pachea* and a generation time of five per year. Population size is given on a logarithmic scale. The thick solid line represents the median estimate of population size; the thin solid lines show the 95% HPD (highest posterior density) intervals.

RESULTS

Microsatellite variation

Variation at the six microsatellite loci over all populations in females of *D. pachea* is shown in Table 3. Each locus showed pronounced variation, typical of microsatellite loci. Average heterozygosity over all populations was high, ranging from 0.743–0.892. Three of the six loci, AC13, AC19 and CAG9, showed significant deviations from HWE. Mean allele size and variation in allele size was generally greater for the two dinucleotide repeat loci than the trinucleotide repeat loci, as expected (Schug et al. 1998; Ross et al. 2003). Mean number of repeats in the CTG7 locus, however, was only slightly lower than that of the AC19 locus.

When each of the 13 populations was compared separately (Appendix 1), variation at the six micro-

Table 3. *Summary of variation at each microsatellite locus averaged over all populations of Drosophila pachea females.*

Locus	n	H_{obs}	H_{exp}	Variance in repeat no.	Mean no. repeats	No. alleles	Expected no. alleles (SMM)
AC13	565	0.851	0.881*	10.96	22.83	21	11.19
AC19	548	0.892	0.926*	28.57	18.81	27	17.06
CAG9	582	0.743	0.770*	2.62	12.77	13	6.39
CAG11	558	0.806	0.838	6.60	12.87	16	8.47
CAG12	580	0.787	0.813	4.68	8.24	13	8.19
CTG7	600	0.789	0.816	3.08	17.64	12	8.17

n, number of chromosomes surveyed; H, observed and expected heterozygosity.*, significant deviation from Hardy-Weinberg expectations.

satellite loci was again high for each population, but deviations from HWE were seen in only twelve of 78 instances. No deviations from HWE were seen in the six loci in populations of females from San Juanico and Desemboque on the mainland, or those from Las Vírgenes and Rancho Costa Rica on the Baja peninsula, although in the latter, sample size was small. No significant deviations from HWE were found at the CTG7 locus across all populations. For each locus, mean allele size was relatively consistent among all populations (Appendix 1). Tests of linkage disequilibrium revealed no consistent evidence of linkage among the six loci (not shown).

Population differentiation

AMOVA indicated that essentially all genetic variation at the six microsatellite loci for female *D. pachea* (99.72%) was found among individuals within populations (Table 4). Less than 1% of the variation was attributed to differences among the two geographic regions. With one exception, pairwise comparisons of F_{ST} among mainland populations indicated the absence of structure and high gene flow characteristic of a panmictic population (Table 5). The population from San Juanico in west-central Sonora, however, showed a significant F_{ST} when compared with the population from Agiabampo in extreme southern Sonora. On the Baja peninsula, pairwise comparisons of F_{ST} again showed evidence of high gene flow, but only when the relatively isolated population at San Felipe in the northeastern region of the peninsula was omitted. The population at San Felipe showed significant F_{ST} values compared with populations from Cataviña, the Vizcaino Desert, Las Vírgenes, and Ciudad Constitución (Table 5).

Pairwise comparisons of F_{ST} among mainland and peninsular populations revealed a contrasting pattern of both population structure and panmixia (Table 5). The population from San Juanico on the mainland was significantly different from all populations on the Baja peninsula, except for those at Rancho Costa Rica in the north, and Punta Conejo in the south. In addition, the population from Desemboque on the mainland was significantly different from the peninsular Cataviña population, and Agiabampo on the mainland was significantly different from San Felipe. Although the remaining pairwise comparisons of F_{ST} among mainland and peninsular populations were not significant, hierarchical AMOVA indicated significant partitioning of genetic variation ($F_{CT} = 0.005$, $P = 0.008$) when comparisons were made between the grouped peninsular and mainland populations, but no significant partitioning was evident among localities within each of the two regions ($F_{SC} = -0.002$, $P = 0.83$; Table 4). None of the pairwise comparisions of F_{ST} described above were significant when a Bonferroni correction for multiple comparisons was applied (critical P value = 0.00064), but it should be noted that this correction also increases the probability of rejecting significant relationships (type II error).

Finally, we found no evidence that changes in the host species of *Lophocereus* affects genetic differentiation in *D. pachea*. No evidence for structure was found between the peninsular population of *D. pachea* from Punta Conejo that utilizes *L. gatesii* as host, and any of the other populations, both peninsular and mainland, that utilize *L. schottii* (Table 5).

Demographic history

The demographic histories of mainland and peninsular populations of *D. pachea* assessed using Bayesian skyline analyses of *COI* sequence data are shown in Fig. 2. For these analyses we subdivided the peninsular populations into three groups, northern (Cataviña and Rancho Costa Rica [*COI* data were not available for San Felipe]), mid-peninsular (Las Vírgenes and Vizcaino Desert), and southern (Ensenada de los Muertos, Punta Conejo, and Ciudad Constitución).

Table 4. *Hierarchical analysis of molecular variance (AMOVA) for Drosophila pachea populations, and resulting fixation indices, grouped by mainland (n = 5) vs Baja peninsula (n = 8) geographic regions.*

Source of variation	DF	Sum of squares	Variance components	% variation
Among groups	1	5.157	0.01100 Va	0.51
Among populations within groups	11	21.195	−0.00492 Vb	−0.23
Within populations	611	1312.180	2.14759 Vc	99.72
Total	623	1338.532	2.15368	

Fixation indices
F_{ST} 0.00282 P = 0.29 (ns)
F_{SC} −0.00229 P = 0.83 (ns)
F_{CT} 0.00511 P = 0.008*

*, significant at the 0.05 level; ns, not significant.

Table 5. *Pairwise F_{ST} values (above diagonal) and estimated number of migrants per generation (Nm, below diagonal) for populations of Drosophila pachea. Population abbreviations are given in Table 1. Mainland populations in bold. Results for pairwise comparisons between mainland and peninsular populations are italicized.* *, *significant at the 0.05 level.*

	AG	**SC**	**SJ**	**DE**	**OP**	EN	PC	CC	LV	VZ	CA	CR	SF
AG	–	−0.021	0.012*	−0.020	−0.026	*0.001*	*−0.026*	*−0.008*	*−0.009*	*0.005*	*−0.009*	*0.011*	*0.032**
SC	inf	–	−0.007	0.007	−0.007	*0.013*	*−0.004*	*0.008*	*0.000*	*−0.007*	*0.007*	*0.022*	*0.008*
SJ	20.2	inf	–	0.004	0.000	*0.020**	*0.002*	*0.011**	*0.017**	*0.016**	*0.011**	*0.015*	*0.032**
DE	inf	36.0	68.3	–	−0.010	*0.010*	*0.007*	*0.004*	*0.007*	*−0.005*	*0.007**	*0.010*	*0.009*
OP	inf	inf	546.9	inf	–	*0.004*	*−0.011*	*−0.002*	*−0.001*	*−0.001*	*−0.007*	*0.008*	*0.003*
EN	*363.0*	*18.4*	*12.0*	*25.3*	*63.6*	–	−0.005	0.004	0.011	0.000	0.000	0.014	−0.009
PC	*inf*	*inf*	*106.3*	*33.6*	*inf*	inf	–	−0.005	−0.009	−0.018	−0.009	0.025	−0.001
CC	*inf*	*29.5*	*21.9*	*70.1*	*inf*	57.0	inf	–	0.005	−0.014	−0.001	0.005	0.028*
LV	*inf*	*536.3*	*14.1*	*36.9*	*inf*	21.5	inf	46.7	–	−0.005	0.000	0.030	0.037*
VZ	*50.6*	*inf*	*15.3*	*inf*	*inf*	747.1	inf	inf	inf	–	−0.015	0.022	0.023*
CA	*inf*	*35.9*	*22.6*	*34.6*	*inf*	923.6	inf	inf	inf	inf	–	0.025	0.013*
CR	*22.5*	*11.3*	*16.4*	*25.5*	*32.8*	17.9	9.8	53.2	8.1	11.2	9.6	–	0.038
SF	*7.6*	*31.9*	*7.5*	*27.7*	*74.8*	inf	inf	8.7	6.5	10.6	18.3	6.3	–

Although a low, but significant, F_{ST} value was seen between the two northern populations (Hurtado et al. 2004), the number of migrants per generation between the two was high (Nm = 12.9), thus we combined the northern populations from Cataviña and Rancho Costa Rica for this analysis. Our rationale for subdividing the Baja populations into three groups was based on the possibility that several physical and biological factors may have had, either separately or in combination, a differential effect on population demographics in different regions of the peninsula. These potential factors include the geographical distances separating the northern and southern populations (~600 km), differences in climatological conditions, availability of the alternate host, *L. gatesii*, in the southern populations, and evidence for northward range expansion of *L. schottii* from Pleistocene glacial refugia (Nason et al. 2002). In addition, transpeninsular seaways are hypothesized to have bisected both the mid-peninsular and southern peninsular regions during the mid Pleistocene and late Pliocene, respectively, potentially affecting gene flow of terrestrial organisms along the north-south peninsular axis (Upton and Murphy 1997; Riddle et al. 2000).

The Bayesian skyline analyses revealed similar demographic histories in populations of *D. pachea* from the mainland (Fig. 2a) and the central peninsular region (Fig. 2c). Flies from both regions showed the signature of a progressive population expansion over approximately the last 1–2 million years. Evidence for a population expansion was also seen in flies from the southern peninsular region (Fig. 2d), but the expansion was relatively more abrupt, beginning approximately 750 000 year before present (BP) based on the 2.3% calibration (Material and methods). Because *D. pachea* from Punta Conejo utilize a different host (*L. gatesii*), we conducted separate skyline analyses on this population (n = 24), as well as on each of the other two southern Baja populations which utilize *L. schottii* (Ciudad Constitución, n = 19, and Ensenada de los Muertos, n = 24), and found demographic trends in each that were identical to that seen in the combined populations in Fig. 2d (not shown). In the northern peninsula, a different trend was observed. The northern population revealed no signature of an expansion, remaining relatively stable in size over approximately the last million years (Fig. 2b).

DISCUSSION

Population structure

Previous population studies on *D. pachea* using allozymes (Rockwood-sluss et al. 1973; Pfeiler and Markow 2001) and mitochondrial *COI* se-

quences (HURTADO et al. 2004) have shown high gene flow and a lack of structure throughout the mainland portion of its range, results largely congruent with those seen with the microsatellite markers. The only evidence for structure we found on the mainland was between populations at San Juanico in west-central Sonora and Agiabampo in extreme southern Sonora. All other pairwise comparisons of F_{ST} were not significant.

The only previous population studies of peninsular populations of *D. pachea* were those of HURTADO et al. (2004). It is worth noting here that all peninsular collecting sites sampled in the present study, except for San Felipe, were sampled by HURTADO et al. (2004). The microsatellite data showed that for all localities, except San Felipe, no population structure was evident, in agreement with results from the mitochondrial DNA (mtDNA) studies. The relatively isolated population at San Felipe, however, showed significant structure compared with most other peninsular populations, including Cataviña in the north, Las Vírgenes and Vizcaino Desert in the mid-peninsular region, and Ciudad Constitución in the south. Interestingly, pairwise comparison of F_{ST} indicated the absence of population structure between San Felipe in the north and the geographically distant populations at both Ensenada de los Muertos and Punta Conejo in the southern peninsular region.

The main difference seen between mtDNA and microsatellite markers in *D. pachea* is that in the former all pairwise comparisons of F_{ST} among peninsular and mainland populations were significant. The hierarchical AMOVA in the mtDNA study also showed a significant F_{CT} value when the grouped peninsular populations were compared with the mainland group, suggesting that the Gulf of California represents a major barrier to gene flow in *D. pachea* (HURTADO et al. 2004). The microsatellite data, however, showed that only eight of the 40 possible pairwise comparisons of F_{ST} between peninsular and mainland populations were significant, and most of those included the mainland San Juanico population (Table 5). But in agreement with the mtDNA data, the hierarchical AMOVA showed a significant, albeit low, F_{CT} value when the grouped peninsular populations were compared with the mainland group (Table 4). A significant F_{CT} value was also obtained after removing the San Juanico population from the analysis ($F_{CT} = 0.003$, $P = 0.011$). Thus, in contrast to results obtained with mtDNA, the microsatellite data suggest that the Gulf, although having an effect on dispersal, provides an incomplete barrier to gene flow. Even the limited number of microsatellite loci utilized here was able to confirm the general pattern detected with mtDNA while revealing additional evidence of structure at a local scale.

Population studies have also been conducted on mainland and peninsular populations of the other three species of cactophilic *Drosophila* endemic to the Sonoran Desert region, and a comparison of these results with those from *D. pachea* provides a basis for understanding the overall potential importance of the Gulf of California on restricting gene flow in flies with differing dispersal capabilities. Results from both mtDNA and allozyme studies suggest that the Gulf does not restrict gene flow in *D. nigrospiracula* and *D. mettleri* (MARKOW et al. 2002; HURTADO et al. 2004), two species with high dispersal capabilities (MARKOW and CASTREZANA 2000). For *D. mojavensis*, however, an analysis of four microsatellite loci in 29 populations from throughout the range of the species, including 8 mainland and 18 peninsular populations, revealed a significant value for F_{ST} among grouped peninsular and mainland populations, suggesting a restriction of gene flow between the two regions, but the value was not as high as that seen among populations of this species in other regions (Mojave Desert and Santa Catalina Island) (ROSS and MARKOW 2006). These workers concluded that the Gulf of California provides an apparently incomplete barrier to gene flow in *D. mojavensis*, possibly as a result of the mid-riff islands acting as "stepping stones" and thus facilitating the migration of flies across the water. Results from mtDNA analyses (658 bp segment of *COI*) also revealed significant structure among peninsular and mainland populations of *D. mojavensis* (all but one of the pairwise comparisons of F_{ST} among the two regions were significant; REED et al. 2007), but the average differentiation among the two groups was less than that seen among the other northern populations, a result similar to that seen with microsatellite markers. The presence of shared *COI* haplotypes among peninsular and mainland populations also suggests that the Gulf barrier to gene flow may be incomplete, supporting the "stepping stone" hypothesis. This hypothesis also could explain the apparent lack of complete isolation of populations of *D. pachea* detected with microsatellite markers, as the senita host occurs on several of the mid-riff islands (TURNER et al. 1995). In agreement with results obtained with *D. pachea*, the microsatellite and mtDNA data from *D. mojavensis* indicate, with a few exceptions, high gene flow and panmixia within the mainland and peninsular regions (ROSS and MARKOW 2006; REED et al. 2007).

Demographic history

Because *D. pachea* shows an obligate relationship with its senita host, we would predict that any host range expansions or contractions driven by Pleistocene glaciation cycles might be reflected in the demographic history of the fly. Based on analyses of allozyme data in *Lophocereus* from throughout its range, Nason et al. (2002) hypothesized that contrasting patterns of postglacial range expansions occurred on the mainland and Baja peninsula, possibly related to a greater number of southern refugia on the mainland. In contrast to the mainland, a clear signature of a postglacial northern range expansion was seen for senita on the Baja peninsula (Nason et al. 2002). However, these authors found no signatures of vicariant events related to the hypothesized transpeninsular seaways in the southern and mid-peninsular regions (Upton and Murphy 1997; Riddle et al. 2000). Overall, these results predict that *D. pachea* might also show contrasting patterns of historical demography between mainland and peninsular populations.

Bayesian skyline analyses revealed differences in the demographic histories of the southern and northern peninsular populations of *D. pachea*, as well as differences between each of these populations and the mainland population (Fig. 2). The estimated timing of the population expansions in the southern peninsular region (~750 000 years BP), and the mid-peninsular and mainland regions (~1–2 million years BP), however, are much older than the end of the last glacial cycle (~15 000 years BP), and appear to predate the beginning of glacial/interglacial cycling (~680 000 years BP), suggesting that factors other than Pleistocene climate changes were responsible for the population expansions in *D. pachea* and the regional differences observed. Similar results and conclusions on the timing of northern population expansions from multiple refugia in southern Sonora were reported for the longhorn cactus beetle (*Moneilema gigas*) (Smith and Farrell 2005), a species that is not restricted to a single host. The unique signature of population expansion seen in the three southern peninsular populations of *D. pachea* in both combined and separate skyline analyses is consistent with the dynamic geological history of this region over the last several million years (Upton and Murphy 1997; Riddle et al. 2000), but whether it is a direct result of hypothesized transpeninsular seaways is unknown.

Finally, the maintenance of a stable effective population size in northern populations of *D. pachea* during the Pleistocene (Fig. 2b) is not consistent with a cycle of local extinctions and recolonization events driven by range contractions and expansions of the senita host. Also, if northern range expansions from southern refugia occurred in *D. pachea*, one would predict reduced genetic diversity in the northern populations, a result which we did not see. Genetic diversity [haplotype diversity (h) and nucleotide diversity (π), with standard deviations] of the *COI* sequences, determined in DnaSP ver. 3.51 (Rozas and Rozas 1999), was essentially identical in the northern ($n = 28$) and southern ($n = 67$) peninsular populations ($h = 0.934 \pm 0.026$ and 0.933 ± 0.022; $\pi = 0.0060 \pm 0.0008$ and 0.0063 ± 0.0005). These observations suggest that northern populations of *D. pachea* may have also been able to survive in isolated refugia during periods of glaciation.

Acknowledgements – We thank Sergio Castrezana for assistance in the field and laboratory. This work was supported by NSF grants (DEB 95-10645, DEB 00–75312) to TAM and funds from the Univ. of Arizona.

REFERENCES

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

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

Brower, A. V. Z. 1994. Rapid morphological radiation and convergence among races of the butterfly *Heliconius erato* inferred from patterns of mitochondrial DNA evolution. – Proc. Natl Acad. Sci. USA 91: 6491–6495.

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

Drummond, A. J., Rambaut, A., Shapiro., B. et al. 2005. Bayesian coalescent inference of past population dynamics from molecular sequences. – Mol. Biol. Evol. 22: 1185–1192.

Excoffier, L., Smouse, P. E. and Quattro, J. M. 1992. Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. – Genetics 131: 479–491.

Excoffier, L., Laval, G. and Schneider, S. 2005. Arlequin ver. 3.0: an integrated software package for population genetics data analysis. – Evol. Bioinformatics Online 1: 47–50.

Farrell, B. D. 2001. Evolutionary assembly of the milkweed fauna: cytochrome oxidase I and the age of *Tetraopes* beetles. – Mol. Phylogenet. Evol. 18: 467–478.

Fogleman, J. C. and Danielson, P. B. 2001. Chemical interactions in the cactus-microorganism-*Drosophila* model system of the Sonoran Desert. – Am. Zool. 41: 877–889.

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

Hartmann, S., Nason, J. D. and Bhattacharya, D. 2002. Phylogenetic origins of *Lophocereus* (Cactaceae) and the

senita cactus-senita moth pollination mutualism. – Am. J. Bot. 89: 1085–1092.

Heed, W. B. 1978. Ecology and genetics of Sonoran desert *Drosophila*. – In: Brussard, P. F. (ed.), Ecological genetics: the interface. Springer-Verlag, p. 109–126.

Heed, W. B. 1982. The origin of *Drosophila* in the Sonoran Desert. – In: Barker, J. S. F. and Starmer, W. T. (eds), Ecological genetics and evolution: the cactus-yeast-*Drosophila* model system. Academic Press, p. 65–80.

Heed, W. B. and Kircher, H. W. 1965. A unique sterol in the ecology and nutrition of *Drosophila pachea*. – Science 149: 758–761.

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

Kimura, M. and Ohta, T. 1975. Distribution of allelic frequencies in a finite population under stepwise production of neutral alleles. – Proc. Natl Acad. Sci. USA 72: 2761–2764.

Lindsay, G. 1963. The genus *Lophocereus*. – Cactus Succulent J. 35: 176–192.

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

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

Michalakis, Y. and Excoffier, L. 1996. A generic estimation of population subdivision using distances between alleles with special reference to microsatellite loci. – Genetics 142: 1061–1064.

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

Nielsen, E. E., Hansen, M. M. and Meldrup, D. 2006. Evidence of microsatellite hitch-hiking selection in Atlantic cod (*Gadus morhua* L.): implications for inferring population structure in nonmodel organisms. – Mol. Ecol. 15: 3219–3229.

Pfeiler, E. and Markow, T. A. 2001. Ecology and population genetics of Sonoran Desert *Drosophila*. – Mol. Ecol. 10: 1787–1791.

Pfeiler, E., Bitler, B. G., Ramsey, J. M. et al. 2006. Genetic variation, population structure and phylogenetic relationships *of Triatoma rubida* and *T. recurva* (Hemiptera: Reduviidae: Triatominae) from the Sonoran Desert, insect vectors of the Chagas' disease parasite *Trypanosoma cruzi*. – Mol. Phylogenet. Evol. 41: 209–221.

Reed, L. K., Nyboer, M. and Markow, T. A. 2007. Evolutionary relationships of *Drosophila mojavensis* geographic host races and their sister species *Drosophila arizonae*. – Mol. Ecol. 16: 1007–1022.

Riddle, B. R., Hafner, D. J., Alexander, L. F. et al. 2000. Cryptic vicariance in the historical assembly of a Baja California Peninsular Desert biota. – Proc. Natl Acad. Sci. USA 97: 14438–14443.

Rockwood-Sluss, E. S, Johnston, J. S. and Heed, W. B. 1973. Allozyme genotype-environment relationships. I. Variation in natural populations of *Drosophila pachea*. – Genetics 73: 135–146.

Ross, C. L. and Markow, T. A. 2006. Microsatellite variation among diverging populations of *Drosophila mojavensis*. – J. Evol. Biol. 19: 1691–1700.

Ross, C. L., Dyer, K. A., Erez, T. et al. 2003. Rapid divergence of microsatellite abundance among species of *Drosophila*. – Mol. Biol. Evol. 20: 1143–1157.

Rozas, J. and Rozas, R. 1999. DnaSP version 3: an integrated program for molecular population genetics and molecular evolution analysis. – Bioinformatics 15: 174–175.

Schug, M. D., Hutter, C. M., Wetterstrand, K. A. et al. 1998. The mutation rates of di-, tri- and tetranucleotide repeats in *Drosophila melanogaster*. – Mol. Biol. Evol. 15: 1751–1760.

Smith, C. I. and Farrell, B. D. 2005. Range expansions in the flightless longhorn cactus beetles, *Moneilema gigas* and *Moneilema armatum*, in response to Pleistocene climate changes. – Mol. Ecol. 14: 1025–1044.

Turner, R., Bowers, J. E. and Burgess, T. 1995. Sonoran Desert plants: an ecological atlas. – Univ. of Arizona Press.

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

Ward, B. L, Starmer, W. T., Russell, J. S. et al. 1974. The correlation of climate and host plant morphology with a geographic gradient of an inversion polymorphism in *Drosophila pachea*. – Evolution 28: 565–575.

Weir, B. S. and Cockerham, C. C. 1984. Estimating F-statistics for the analysis of population structure. – Evolution 38: 1358–1370.

Appendix 1. Variation at the six microsatellite loci for each population of *Drosophila pachea* females. Population abbreviations are given in Table 1. n = number of chromosomes surveyed; H = observed and expected heterozygosity; SMM = expected number of alleles. *, significant deviation from Hardy – Weinberg expectations.

Locus/region	Pop	n	H_{obs}	H_{exp}	Variance in repeat no.	Mean no. repeats	No. alleles	SMM
AC13								
Mainland	AG	56	1.000	0.868	12.00	22.96	11	9.30
	SC	30	0.800	0.887	9.69	22.97	10	9.66
	SJ	36	0.833	0.889	23.35	24.28	14	10.10
	DE	67	0.882	0.902	12.78	23.28	15	11.90
	OP	30	0.867	0.892	9.98	23.23	10	9.95
Peninsular	EN	26	0.923	0.892	8.40	22.35	10	9.63
	PC	44	0.727	0.890*	7.27	22.11	11	10.48
	CC	124	0.871	0.896	10.75	23.05	17	11.53
	LV	28	0.857	0.894	15.84	22.71	11	9.93
	VZ	20	0.700	0.911	5.84	21.45	9	9.88
	CA	78	0.769	0.863*	5.70	22.41	12	9.19
	CR	8	0.500	0.750	1.36	24.25	4	3.93
	SF	18	0.667	0.915*	19.48	21.78	9	9.70
AC19								
Mainland	AG	40	0.800	0.942*	27.65	20.20	16	15.73
	SC	30	0.867	0.903	23.65	16.73	13	10.75
	SJ	36	0.833	0.927	30.26	18.97	15	13.62
	DE	56	0.821	0.921	22.99	17.66	17	14.01
	OP	44	0.773	0.938*	24.86	17.50	17	15.98
Peninsular	EN	24	1.000	0.938	53.97	18.67	14	13.01
	PC	44	0.909	0.937	38.16	18.07	16	15.79
	CC	116	0.862	0.924*	22.36	19.51	19	15.14
	LV	28	0.857	0.899	24.30	18.82	10	10.28
	VZ	24	1.000	0.924	24.72	20.88	11	11.62
	CA	78	0.795	0.930*	29.68	19.08	18	15.96
	CR	8	1.000	0.929	27.14	19.00	6	6.29
	SF	20	0.900	0.926	21.63	19.45	11	10.93
CAG9								
Mainland	AG	46	0.609	0.720	2.53	13.22	6	5.17
	SC	30	0.667	0.747	1.39	12.70	6	5.40
	SJ	40	0.700	0.782	1.95	12.48	7	6.20
	DE	70	0.857	0.804	4.15	12.77	10	6.95
	OP	44	0.682	0.833	5.91	13.36	10	7.62
Peninsular	EN	26	0.615	0.692	1.87	12.88	7	4.60
	PC	44	0.682	0.745	1.50	12.89	6	5.58
	CC	122	0.689	0.793*	3.15	12.80	10	6.76
	LV	28	0.643	0.728	1.66	12.21	7	5.07
	VZ	26	0.615	0.843*	4.07	12.92	7	7.44
	CA	76	0.789	0.756	2.19	12.70	9	5.94
	CR	8	0.750	0.857	1.84	12.13	5	5.09
	SF	22	0.727	0.710	1.90	13.00	6	4.68

Appendix 1 (*Continued*)

Locus/region	Pop	n	H_{obs}	H_{exp}	Variance in repeat no.	Mean no. repeats	No. alleles	SMM
CAG11								
Mainland	AG	56	0.893	0.845	7.69	12.63	11	8.22
	SC	30	1.000	0.871	9.45	13.00	10	8.79
	SJ	40	0.750	0.779	4.27	12.80	8	6.15
	DE	70	0.857	0.870	6.08	13.24	12	9.52
	OP	40	0.750	0.900*	11.99	12.90	12	11.11
Peninsular	EN	26	0.692	0.815	3.04	12.65	7	6.63
	PC	44	1.000	0.844	5.44	13.05	10	8.00
	CC	104	0.808	0.827	4.27	13.29	12	7.71
	LV	24	0.750	0.837	5.87	12.96	8	7.13
	VZ	28	0.857	0.849	6.53	13.36	9	7.75
	CA	74	0.865	0.836	5.40	13.12	12	7.94
	CR	8	0.750	0.857	9.84	10.88	5	5.09
	SF	14	0.571	0.758	5.96	13.50	7	4.83
CAG12								
Mainland	AG	58	0.828	0.780	4.54	9.09	10	6.33
	SC	28	0.571	0.833*	5.60	7.75	8	7.22
	SJ	38	0.632	0.727	6.08	8.16	9	5.23
	DE	58	0.724	0.855	7.93	8.29	12	8.65
	OP	42	0.714	0.851	4.71	8.14	10	8.28
Peninsular	EN	24	0.833	0.819*	3.28	8.67	8	6.62
	PC	44	0.909	0.845	2.81	8.57	8	8.04
	CC	120	0.817	0.840	4.85	8.21	12	8.21
	LV	30	0.867	0.841	4.92	8.20	9	7.56
	VZ	30	0.933	0.844	5.51	7.93	9	7.64
	CA	78	0.872	0.856	5.11	8.44	10	8.82
	CR	8	0.750	0.750	2.98	7.88	4	3.93
	SF	22	0.545	0.732	2.56	7.77	7	4.95
CTG7								
Mainland	AG	64	0.625	0.751	2.09	17.81	6	5.71
	SC	30	0.600	0.802	2.66	17.60	6	6.45
	SJ	38	0.789	0.844	3.34	17.47	7	7.88
	DE	68	0.794	0.802	3.25	17.63	8	6.87
	OP	44	0.955	0.837	3.00	17.48	6	7.76
Peninsular	EN	26	0.846	0.837	3.28	17.65	6	7.24
	PC	44	0.818	0.850	3.14	17.52	8	8.26
	CC	120	0.783	0.826	3.70	17.39	11	7.70
	LV	28	0.643	0.812	3.39	17.86	9	6.63
	VZ	28	0.857	0.847	3.74	17.43	8	7.66
	CA	78	0.846	0.857	3.11	17.53	10	8.86
	CR	8	0.750	0.750	1.71	17.50	3	3.93
	SF	24	0.667	0.797	3.65	18.50	6	6.12