

# Genetic differentiation in the *Agave deserti* (Agavaceae) complex of the Sonoran desert

A Navarro-Quezada<sup>1</sup>, R González-Chauvet<sup>2</sup>, F Molina-Freaner<sup>3</sup> and LE Eguiarte<sup>4</sup>

<sup>1</sup>Biology Department McGill University, 1205 Ave. Dr Penfield, Montreal, Quebec Canada, H3A 1B1; <sup>2</sup>Departamento de Ecología Evolutiva, Instituto de Ecología, Universidad Nacional Autónoma de México, Apartado Postal 70–275, México, D.F. 04510, México; <sup>3</sup>Instituto de Ecología, Universidad Nacional Autónoma de México, Estación Regional del Noroeste, Apartado Postal 1354, Hermosillo, Sonora, 83000, México; <sup>4</sup>Departamento de Ecología Evolutiva, Instituto de Ecología, Universidad Nacional Autónoma de México, Apartado Postal 70–275, México, D.F. 04510, México

The *Agave deserti* complex, comprising *A. deserti*, *A. cerulata* and *A. subsimplex*, represents a group of species and subspecies with a near allopatric distribution and clear differences in morphology. Genetic differentiation and taxonomic status with respect to spatial distribution of 14 populations of the complex were analyzed in an effort to understand the evolution and speciation process within the genus. Allelic frequencies, levels of genetic variation, expected heterozygosity ( $H_S$ ), proportion of polymorphic loci ( $P$ ), and genetic differentiation ( $\theta$  and Nei's genetic distance) were estimated using 41 putative RAPD loci. All three species show high levels of genetic variation ( $H_S = 0.12–0.29$ ,  $P = 63.4–95.1$ ), and low genetic differentiation between

populations and species ( $\theta$  populations =  $0.14 \pm 0.02$  (SE);  $G_{ST} = 0.11 \pm 0.02$ ). Accordingly, gene flow among populations was estimated as high by three different methods ( $N^m = 2.91–6.14$ ). Nei's genetic distances between the three species were low compared to the values obtained from other Agavaceae, and there was no clear correlation with taxonomic divisions. In a UPGMA analysis, *A. subsimplex* and *A. cerulata* formed exclusive monospecific clusters, whereas the *A. deserti* populations appear in more than one cluster together with other species. The results were consistent with a pattern of genetic isolation by distance.

*Heredity* (2003) **90**, 220–227. doi:10.1038/sj.hdy.6800216

**Keywords:** Agavaceae; *Agave deserti*; genetic structure; RAPDs; Sonoran desert; speciation

## Introduction

The Sonoran desert is relatively young (Axelrod, 1979; Van Devender, 1990). Its present day configuration results from the physical separation of the Baja California Peninsula from mainland Mexico and the consequent formation of the Gulf of California 5 million years ago (Lonsdale, 1989). The present distribution of many Sonoran desert taxa on both sides of the Gulf (Turner *et al*, 1995) offers a good system to study population differentiation across a geographic barrier.

The genus *Agave* comprises more than 160 species, most of them with similar ecologies and morphologies, although their distributions are usually geographically restricted (Gentry, 1982; García-Mendoza and Galván, 1995). Among the species in the genus, the *A. deserti* complex represents an interesting system, having species/subspecies with near allopatric distributions and differences in morphology (Gentry, 1982), which possibly reflect patterns of genetic differentiation because of independent evolution among the populations. The species in the complex are an important floristic component of the Sonoran desert vegetation, since they usually have large populations. For example, Gentry

(1982) indicates that *A. cerulata* probably has the second largest population of all the *Agave* species.

Morphological studies recognize three species and seven subspecies within this complex. It comprises *A. deserti* from the northern part of the Sonoran desert (with three subspecies), plus two closely related species, *A. cerulata* (with four subspecies) and *A. subsimplex* (Gentry, 1982; Turner *et al*, 1995, Figure 1). The last two species have been considered to be closely related (Gentry, 1982), but parallel adaptation to similar environmental conditions and/or conservation of shared ancestral characters might explain their morphological similarity despite being separated by the geographical barrier created by the Gulf of California (see Figure 1). Some of the subspecies have been considered as species in the past (Gentry, 1982) or as populations undergoing speciation, as in the case of *A. cerulata* ssp. *dentiens*, from the San Esteban Island.

Population genetic structure was analysed in order to understand genetic differentiation, taxonomy and geographical distribution in the heterogeneous Sonoran desert could explain the evolution and speciation process of this species complex.

In order to analyze the genetic differentiation in this group of species, we used RAPDs, as these markers allow sampling a large number of loci that, in principle, represent a random sample of the genome, and thus guarantee a genome-wide estimate of the population genetic structure parameters. RAPDs have been criticized as tools for estimating population genetic para-

Correspondence: A Navarro-Quezada, Centre de Recherche en Biologie Forestière, Université Laval, Quebec, Canada G1K 7P4.  
E-mail: [aura.navarro@mail.mcgill.ca](mailto:aura.navarro@mail.mcgill.ca)

Received 17 December 2001; accepted 27 August 2002

meters (Hedrick, 1999); nevertheless, they have proven successful in many studies where they have been used in combination with other markers to test their effectivity. Once good quality controls have been established, RAPDs are reliable genetic markers for population genetic studies (eg, Besnard *et al*, 2001).

## Materials and methods

### Field collection

In all, 14 populations of the three species were sampled between November 1997 and March 1998 (Table 1 and Figure 1). At least 30 individuals per population were collected, which has been suggested as a minimum number for relatively accurate estimates of the allelic frequencies (Nei, 1987). Plants collected were at least 2 m apart from one another to avoid collecting the same

genets. Each population was 50–100 km apart. The location of each population was determined using a GPS device (Ensign GPS, Tremble Navigation) and INEGI maps (scale 1:250 000). Approximately 25 cm<sup>3</sup> of central leaf tissue was collected. It was kept in liquid nitrogen and transferred to storage at –80°C.

### Genetic analysis

DNA was extracted from all individuals using a modified CTAB technique (Doyle and Doyle, 1987). The tissue was ground to a fine powder using liquid nitrogen. DNA concentrations were measured with a Perkin-Elmer Lambda Bio UV/Vis spectrophotometer and standardized for amplification.

The procedure of Williams *et al* (1990) for RAPDs was used with few modifications: 6 × PCR buffer with MgCl<sub>2</sub> (Boehringer-Mannheim), 4 × PCR buffer without MgCl<sub>2</sub>, 0.2 mM nucleotide mix (Boehringer-Mannheim), 1 U *Taq* polymerase (Boehringer-Mannheim), 2.5 ng of DNA template and 10 pM of primer. The PCR was carried out in an ERICOMP Deltacycler II™ with the following program: 94°C (1 min), 43 cycles (94°C, 1 min; 36°C, 1 min.; 72°C, 2 min.); 72°C, 15 min; 4°C. The ethidium bromide-stained agarose gels were photographed using a digital camera (Kodak) and analyzed with help of the program Bio Max 1D from Kodak Scientific Imaging Systems.

In total, 40 primers from Operon Technologies (Kits A and B) were assayed. Clear bands were obtained from five primers, but two yielded repeatable results: primers OPA-10 and OPB-20.

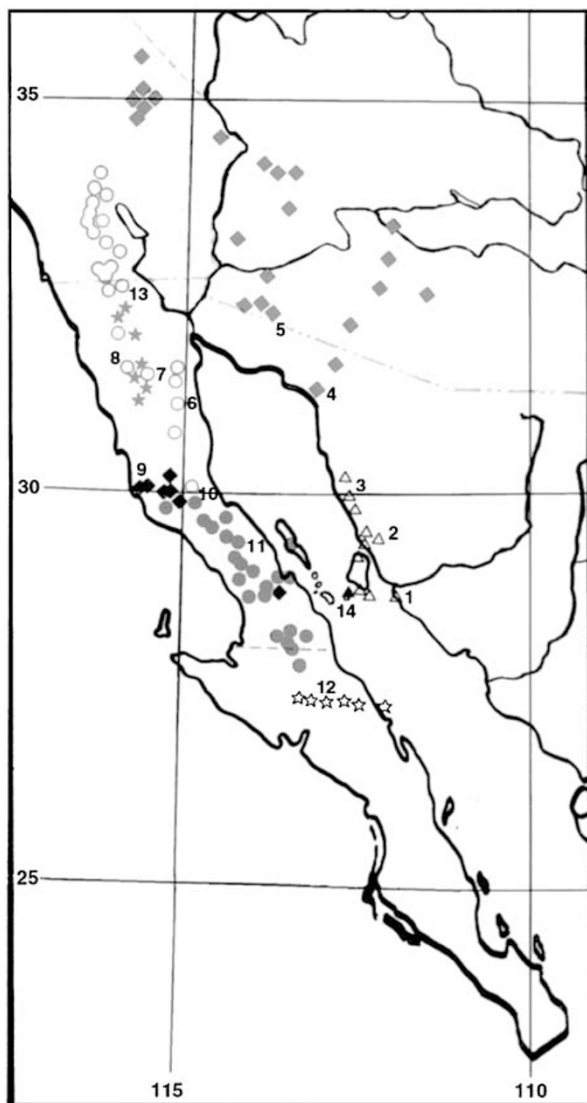
### Data analysis

From the gels, a presence/absence matrix of RAPD bands was constructed. This matrix was analyzed with the program Tools for Population Genetic Analyses (TFPGA, Mark P Miller, Northern Arizona University, AZ, USA). We obtained the allelic frequencies using Lynch and Milligan (1994) correction, since the markers are dominant. This assumes that the populations are in Hardy–Weinberg equilibrium, a reasonable assumption, given preliminary allozyme population genetics data for the species (M Salas and A Valera, UNAM, unpublished data).

From the RAPD data for each population, we estimated the expected (heterozygosity) ( $H_S$ ) per locus and per population, as well as the proportion of polymorphic loci per population ( $P$ , 99% criterion). To test for significant differences in (heterozygosity) values between populations, we performed an ANOVA using SYSTAT to test for the effect of species and populations when comparing  $H_S$ .

Population differentiation analyses were undertaken for each individual population together and for the combined populations of each species.

- (1) Differences in allelic frequencies per locus were compared using Fisher's combined probability test (in TFPGA, from Sokal and Rohlf, 1995).
- (2)  $\theta$  ( $= F_{ST}$ , coefficient of coancestry, Weir, 1996) was estimated for each locus in all populations. The  $\theta$  value was estimated for all populations and bootstrapped 1000 times, to obtain a 95% confidence interval.  $G_{ST}$  was also estimated after Hamrick and



**Figure 1** Geographic distribution of the *A. deserti* complex (modified from Gentry, 1982). Numbers 1–14 indicate populations collected (Table 1). Symbols represent: ● *Agave deserti* ssp. *deserti*, ◆ *Agave deserti* ssp. *simplex*, ★ *Agave deserti* ssp. *pringlei*, ◆ *Agave cerulata* ssp. *nelsonii*, ○ *Agave cerulata* ssp. *cerulata*, ☆ *Agave cerulata* ssp. *subcerulata*, ▲ *Agave cerulata* ssp. *dentiens*, △ *Agave subsimplex*.

Godt (1996). Gene flow was measured by  $N^m$ , which is equivalent to the number of migrants per generation (Hartl and Clark, 1996). Gene flow was indirectly calculated using the Crow and Aoki (1984) formula,  $\theta = 1/(4\alpha N^m + 1)$ , where  $\alpha = (n/n-1)^2$ , and  $n$  is the number of subpopulations. The subpopulations were the populations analyzed for each species.

- (3)  $N^m$  was calculated using the private alleles method of Slatkin (1985). In this method, the average frequency of alleles found in only one population  $p(1)$  is  $\log_{10}[p(1)] = a \log_{10}(N^m) + b$ . Based on simulation studies, Slatkin (1985) determined  $a = -0.61$  and  $b = -1.2$  for 50 individuals, and  $a = -0.58$  and  $b = -1.1$  for 25 individuals.
- (4) Gene flow was estimated by an analysis of average allele frequencies per population (Slatkin, 1981) to obtain, the proportion of migrants per generation ( $m$ ). A qualitative comparison was made by overlaying the graphs obtained with those obtained by Slatkin (1981) for both the island and the stepping-stone models to obtain an approximate value for  $m$ .
- (5) Nei's genetic distances (Nei, 1987) were estimated. With the matrix of Nei's minimum distances (Nei, 1987), an UPGMA dendrogram from the presence/absence matrix using 1000 bootstrap replicates was constructed.
- (6) Isolation-by-distance was tested by correlating genetic and geographic distances between populations using the Mantel analysis (Manly, 1991) with 10000 permutations to obtain an unbiased result for the probability.

## Results

### Genetic variation

In all, 41 polymorphic bands were amplified, 23 with primer B-20 and 18 with primer A-10 (the presence/absence matrix can be obtained from the corresponding author). DNA was extracted from all the individuals collected, although for some the concentration or quality

of DNA was unsuitable for RAPD amplifications, hence the number of analyzed individuals was less than the sampled number (Table 1, average  $n = 41.4 \pm 1.83$  SE).

The genetic variation was high compared to values obtained for plants using allozymes (Hamrick et al, 1992) ranging from 0.121 to 0.292 in  $H_S$  and from 63.41 to 95.12% for  $P$ , according to both estimators used (average  $H_S = 0.199 \pm 0.01$  SE, average  $P = 82.6 \pm 2.46$  SE, Table 1). The results of the ANOVA on the heterozygosities (Table 2) indicate that there are significant differences among all populations of the three species taken together, as well as among the *A. deserti* populations. No significant differences were detected among populations of *A. cerulata* and *A. subsimplex* (Table 2).

The northern *A. deserti* ssp. *deserti* (La Rumorosa) showed the highest variation, whereas one of the southern populations of *A. subsimplex* (Rancho San Gabriel) showed the least variation there was no overall correlation between latitude and  $H_S$  ( $r^2 = 0.536$ ,  $P = 0.59$ ). Both estimators of genetic variation ( $H$  and  $P$ ) are correlated ( $r^2 = 0.7809$ ,  $P = 0.01$ ), indicating that higher polymorphism corresponds to higher heterozygosity.

### Genetic differentiation

The Fisher combined probability test showed significant differences in allelic frequencies for all loci among species in the *A. deserti* complex and also among populations of *A. deserti* (Table 3). All but six loci among populations of *A. cerulata* were significantly different, whereas in *A. subsimplex* there were no significant differences in 30 loci, which indicate less differentiation among its populations than in the other two species.

Following the same pattern as the differences in allelic frequencies, population differentiation as estimated by  $\theta$  and  $G_{ST}$  shows small values particularly for *A. subsimplex* and *A. cerulata*, although significantly different from 0 (Table 3). The estimator  $\theta$  ranges from 0.084 to 0.137 for all species and  $G_{ST}$  varies from 0.07 to 0.149. The estimators  $\theta$  and  $G_{ST}$  were not significantly different ( $t$ -test  $P < 0.001$ ).

**Table 1** Populations of *Agave cerulata*, *A. deserti* and *A. subsimplex* used in this study (see also Figure 1) and levels of genetic variation at 41 RAPD loci for 14 populations.

Pop	Species	$H_S$	$P$	$N$
1	<i>A. subsimplex</i> (Bahía Kino)	0.161 ( $\pm 0.021$ )	82.9	39
2	<i>A. subsimplex</i> (San Gabriel)	0.121 ( $\pm 0.024$ )	63.4	29
3	<i>A. subsimplex</i> (La Placa)	0.151 ( $\pm 0.023$ )	80.5	34
	<i>A. subsimplex</i>	0.144 ( $\pm 0.021$ )	75.6 ( $\pm 6.14$ )	34 ( $\pm 2.88$ )
4	<i>A. deserti</i> ssp. <i>simplex</i> (Quitobac)	0.247 ( $\pm 0.022$ )	95.1	45
5	<i>A. deserti</i> ssp. <i>simplex</i> (El Pinacate)	0.218 ( $\pm 0.023$ )	87.8	51
6	<i>A. deserti</i> ssp. <i>deserti</i> (Laguna Amarga)	0.247 ( $\pm 0.024$ )	92.7	50
7	<i>A. deserti</i> ssp. <i>Pringlei</i> (Paso San Matías)	0.248 ( $\pm 0.027$ )	87.8	50
8	<i>A. deserti</i> ssp. <i>Deserti</i> (Leyes de Reforma)	0.168 ( $\pm 0.023$ )	80.5	36
13	<i>A. deserti</i> ssp. <i>Deserti</i> (La Rumorosa)	0.292 ( $\pm 0.026$ )	95.1	47
	<i>A. deserti</i>	0.186 ( $\pm 0.049$ )	78.1 ( $\pm 2.52$ )	39.8 ( $\pm 2.51$ )
9	<i>A. cerulata</i> ssp. <i>nelsonii</i> (Rosario)	0.177 ( $\pm 0.026$ )	73.2	42
10	<i>A. cerulata</i> ssp. <i>cerulata</i> (Cataviña)	0.167 ( $\pm 0.025$ )	73.2	40
11	<i>A. cerulata</i> ssp. <i>cerulata</i> (Punta Roca)	0.163 ( $\pm 0.023$ )	75.6	40
12	<i>A. cerulata</i> ssp. <i>subcerulata</i> (Abulón)	0.207 ( $\pm 0.025$ )	80.5	44
14	<i>A. cerulata</i> ssp. <i>dentiens</i> (Isla San Esteban)	0.217 ( $\pm 0.022$ )	87.8	33
	<i>A. cerulata</i>	0.237 ( $\pm 0.078$ )	89.8 ( $\pm 2.53$ )	44.17 ( $\pm 1.69$ )
	All three species	0.199 ( $\pm 0.013$ )	82.6 ( $\pm 2.46$ )	41.4 ( $\pm 1.83$ )

$H_S$ =expected heterozygosity,  $P$ =proportion of polymorphic loci are shown for all populations and for each separate species ( $\pm$ SE).

**Table 2** Analysis of variance of  $H_S$  (expected heterozygosity) within and between *Agave* species

Source of variation	SS	df	MS	F	P-value
ANOVA for $H_S$					
<i>A. subsimplex</i>	0.035	2	0.017	0.829	0.439
<i>A. deserti</i>	0.347	5	0.069	2.895	0.015
<i>A. cerulata</i>	0.099	4	0.025	1.032	0.392
All species together	1.231	13	0.095	4.051	<0.0001

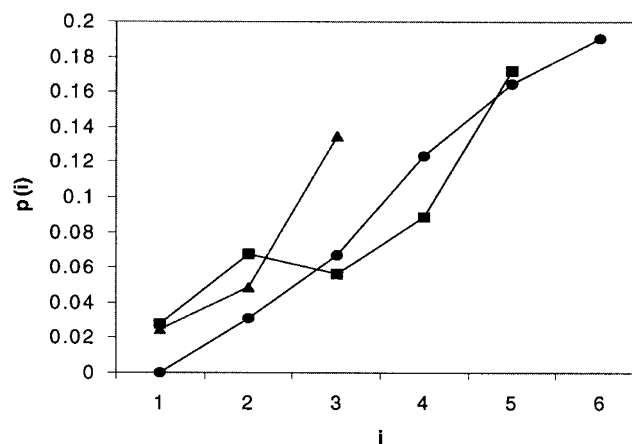
The  $N^m$  values obtained either from  $\theta$  or private allele estimates were similar, ranging between 2.91 and 6.14 for the first analysis, and between 3.36 and 3.83 with the rare alleles (Table 3). An estimate for the rare alleles was not possible for *A. deserti* as private alleles were not found. In all cases,  $N^m$  estimates were greater than 1.

According to the analysis of average allelic frequencies (Slatkin, 1981) *A. subsimplex* and *A. cerulata* have a migration rate of approximately  $m=0.1$ , whereas *A. deserti* adjusted to the simulation of a lower migration rate, near  $m=0.05$  (Figure 2).

The average genetic distance among all populations was  $D=0.040$  ( $SE \pm 0.002$ , 91 comparisons). When comparing the genetic distances observed at three taxonomic levels (Figure 3), we found that the average genetic distance within subspecies was  $D=0.043$  ( $SE \pm 0.015$ , five pairs of comparisons), within species  $D=0.032$  ( $SE \pm 0.004$ , 23 comparisons), while between species the average is  $D=0.041$  ( $SE \pm 0.003$ , 63 comparisons).

The low correlation between current taxonomy and genetic differentiation is clearly shown in the UPGMA dendrogram (Figure 4), although this distance tree reflects geographical patterns. *Agave deserti* seems to be a polyphyletic group, as one *A. deserti* spp. *deserti* population (8) appears inside the cluster with *A. cerulata*, which may be because of an introgression event, since this *A. deserti* population (Leyes de Reforma) is close to the north-western limit of the *A. cerulata* distribution. The subspecies *A. deserti* spp. *simplex* forms a cluster genetically closer to *A. subsimplex* than to other *A. deserti* subspecies. These populations of *A. deserti* spp. *simplex* (Quitobac and El Pinacate) are geographically near the distribution of *A. subsimplex*, in Sonora. Low bootstrap values in some of the branches indicate that these groups are not strong, something that might be a result of high gene flow among the populations.

There is only a weak correlation between geographic distance and genetic distance according to the Mantel tests, when considering all the species together ( $r=0.401$ ,



**Figure 2** Analysis of average allele frequencies per population as performed in Slatkin (1981), where  $i$  is the number of populations with a certain allele, and  $p(i)$  is the average frequency of all alleles displayed by  $i$  populations.  $m$  (number of migrants per generations) within a species was estimated quantitatively by comparing the graphs with those obtained from simulations by Slatkin (1981). ● *Agave deserti* approximates to  $m=0.05$ , ■ *A. cerulata* and ▲ *A. subsimplex* approximate to  $m=0.1$ .

$P=0.085$ ). Since this weak correlation could be because of the isolation effect produced by the Gulf of California, we also estimated the distances in the populations for each species. For *A. deserti* and *A. cerulata*, we found a correlation between genetic and geographic distance ( $r=0.920$ ,  $P=0.006$  and  $r=0.735$ ,  $P=0.009$ , respectively, Figure 5). In *A. subsimplex*, Figure 5 genetic distances seemed not to be related to geography, but bias may have been introduced by the small sample size (three populations). When only the continental populations of *A. cerulata* (ie, removing the San Esteban Island, population 14) were taken into account, we observed a higher correlation ( $r=0.892$ ,  $P=0.001$ ) in the Mantel test (data not shown). This suggests that the island population is not following the same pattern of geographic–genetic differentiation as other populations in the species.

## Discussion

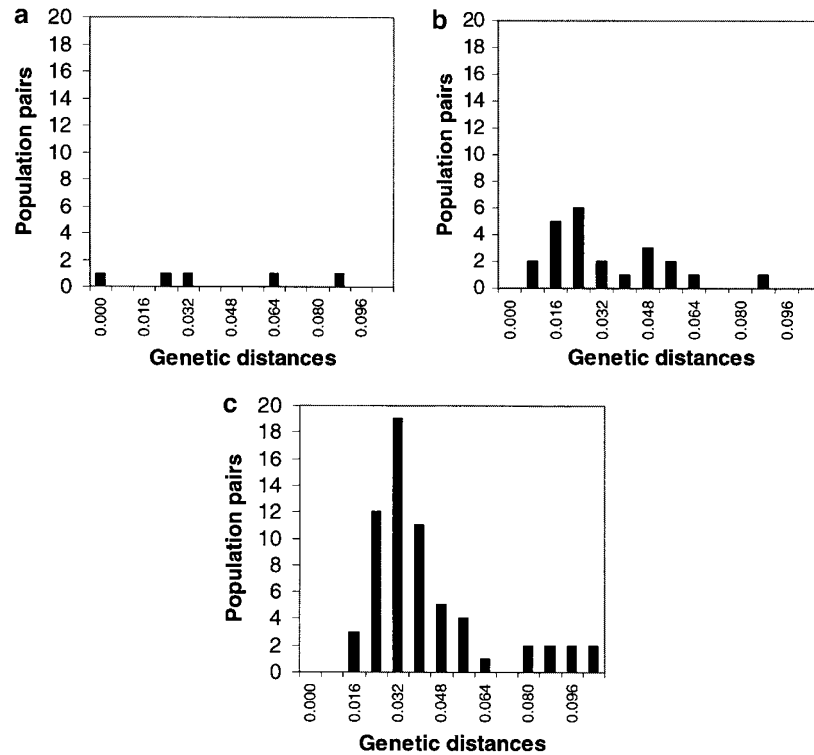
### Levels of genetic variation

Genetic variation is high in all the populations sampled, compared with the values obtained for plants in general using allozyme markers ( $H_S=0.15$  and  $P=50\%$ , Hamrick *et al* 1992). Our  $H_S$  estimates are lower than an average obtained from a survey of published plant RAPD data (A Navarro-Quezada, personal communica-

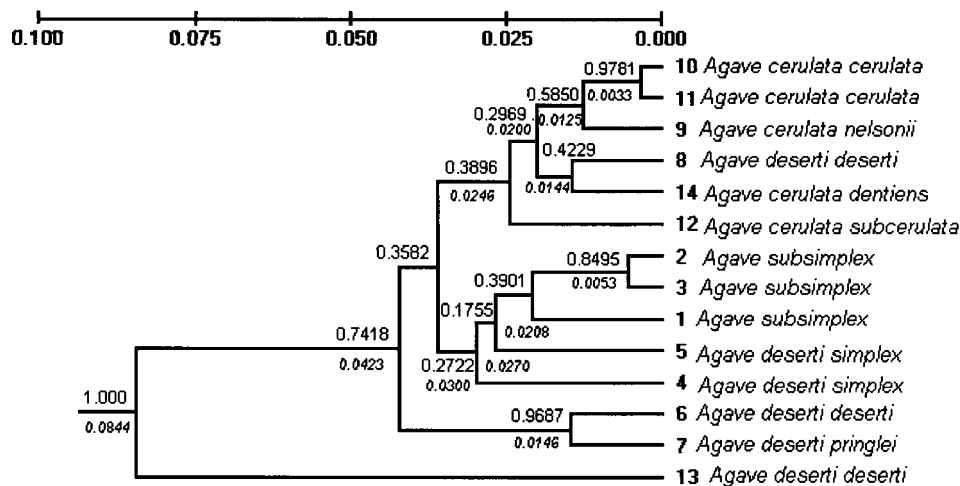
**Table 3** Population differentiation as estimated with Fisher's combined probability test, Weir and Cockerham's  $\theta$  and gene flow estimates derived from  $\theta$  and using the private alleles method of Slatkin (1981) ( $\pm SE$ ) for 14 populations of the *Agave deserti* complex in the Sonoran desert, Mexico, using 41 RAPD loci

	Fisher's combined probability	$\theta$ ( $F_{ST}$ ) ( $\pm SE$ )	$G_{ST}^b$ ( $\pm SE$ )	$N^m$ from $\theta$ ( $\pm SE$ )	$N^m$ from private alleles
All populations	All loci different $P=0.001$	0.137 ( $\pm 0.02$ )	0.113 ( $\pm 0.02$ )	2.91 ( $\pm 0.64$ )	3.51
<i>A. subsimplex</i>	11/41 different loci $P=0.001$	0.084 ( $\pm 0.05$ )	0.07 ( $\pm 0.03$ )	6.14 ( $\pm 0.07$ )	3.83
<i>A. deserti</i>	All loci different $P=0.001$	0.135 ( $\pm 0.03$ )	0.149 ( $\pm 0.03$ )	2.99 ( $\pm 0.41$ )	— <sup>a</sup>
<i>A. cerulata</i>	35/41 different loci $P=0.001$	0.098 ( $\pm 0.02$ )	0.095 ( $\pm 0.02$ )	4.38 ( $\pm 0.02$ )	3.36

<sup>a</sup>No private alleles found in this species. <sup>b</sup>Hamrick and Godt's (1989)  $G_{ST}$ .



**Figure 3** Nei's (1978 in Nei, 1987) genetic distances for pairs of populations at three taxonomic levels. (a) within subspecies, (b) between subspecies and (c) between the three species, *Agave deserti*, *A. cerulata* and *A. subsimplex*.

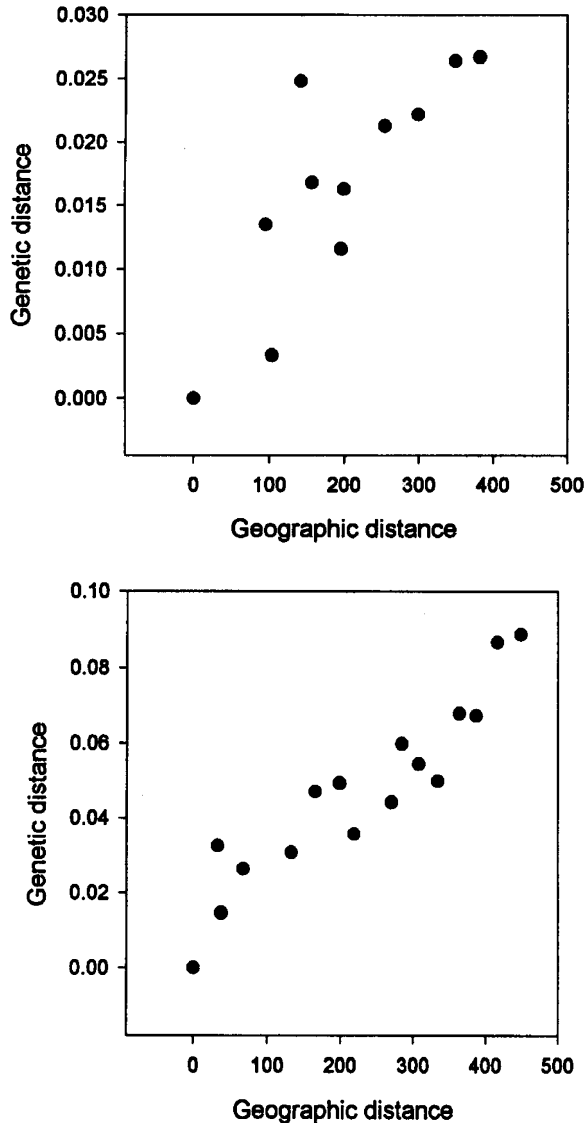


**Figure 4** UPGMA dendrogram constructed from RAPD data. 1–14 are the numbers given to the collected populations (see Table 2, Figure 1). The bootstrap consensus value (proportion) is given at each node, the Nei (1978) genetic distance between clusters or taxa separated at the nodes are given below in italics.

tion), but higher for *P* (average  $H_S=0.271$  and  $P=61.83\%$ ). The higher *P* values could be because of the use of only two RAPD primers, since other primers assayed yielded less polymorphism (eg, OPB-17), but they were not consistent enough to be used in our analyses. According to Hamrick and Godt's (1996) allozyme averages, the values obtained for *H* and *P* would be an indication of a predominantly outcrossing plant with gravity seed dispersal, a long-lived perennial habit and a regional distribution. All traits are in concordance with the ecology of the genus *Agave*. Similar

genetic variation values have been obtained for other Agavaceae (Massey and Hamrick, 1998; Martínez-Palacios *et al*, 1999) and other long-lived monocots (Eguiarte *et al*, 1993; Izquierdo and Piñero, 1998).

Genetic variation of the *A. cerulata* population from the San Esteban Island is not significantly different from that of other populations of the same species. Island populations generally have less variation than mainland populations (Frankham, 1997). One possible explanation for higher variation on islands is the presence of gene flow from continental populations. This could be the case



**Figure 5** Correlation between genetic (Nei, 1987) and geographic distance (in km). Upper graph: *Agave cerulata*, lower: *A. deserti*.

for *A. cerulata* ssp. *dentiens* if it is pollinated by bats. San Esteban Island is within flight distance from a known roost of *Leptonycteris curasoae* on Tiburon Island (Horner et al, 1998). Nevertheless, a large population size might be a factor for the maintenance of high genetic variation (A Navarro-Quezada, F Molina-Freaner, unpublished data).

**Genetic differentiation and indirect estimates of gene flow**  
The values obtained for population differentiation ( $\theta$ ,  $G_{ST}$ ) are close to the ones surveyed for long-lived perennial, regional and outcrossing monocots in allozyme studies (Hamrick and Godt, 1996), which again coincides with the *Agave* life history. The measure  $\theta$  is a good estimator of population differentiation and gene flow when populations are large and gene flow is moderate, introducing less bias for the estimation of  $F_{ST}$  (Slatkin, 1989). Differences within allelic frequencies turned out to be low and  $\theta$  and  $G_{ST}$  values were hardly,

although significantly, different from zero. This shows that there is a lower genetic differentiation within and among the populations of the species comprising the *A. deserti* complex, than was expected from the extant morphological differentiations that led Gentry (1982) to divide them into different species/subspecies. This suggests that there is relatively high gene flow among populations.

Differentiation values are relatively low for *A. sub-simplex* and *A. cerulata*, which agrees with the differentiation observed in another widespread species of *Agave* (Silva, 2001). Nevertheless, larger differentiation is observed when considering all populations of the three species together and the species *A. deserti* alone. Of the three species, the one with the most disjunct distribution is also the more differentiated (*A. deserti*) and suggests that an isolation-by-distance model is responsible for the larger differentiation values found for this species.  $N^m$  was highest for *A. subsimplex*, the species whose populations are geographically closer than the other two more widespread species. The analysis of average allelic frequencies as described in Slatkin (1981) supports that *A. deserti* has a lower gene flow within its populations than the other two species.

Using a Mantel test, a positive correlation between geographic and genetic distance was found not only for *A. deserti*, but also for *A. cerulata* (Figure 5). Data on the pollination/dispersal biology of the species are needed to confirm that distances between populations are a determinant factor in differentiating them.

$N^m$  estimates suggest that gene flow is a relevant evolutionary force in the complex (Kimura, 1983; Slatkin, 1985). Nevertheless, indirect estimates of  $N^m$  can be biased upwards by different factors violating the assumptions made when estimating  $F_{ST}$  and  $N^m$  from allelic frequencies (Whitlock and MacCauley, 1998). The upward bias for  $N^m$  (downward for  $F_{ST}$ ) can occur because of overdominance and spatially uniform selection, both of which can be a possibility in our system. Nevertheless, RAPDs have been considered to be neutral and less prone to mutation than other markers, such as microsatellites and mtDNA. A large number of markers (41 in this study) help assure that  $N^m$  estimates are less affected by non-neutral processes.

The small differentiation of the populations and species is also reflected in the genetic distance analysis. In contrast to what was expected from taxonomy, Nei's genetic distances among populations of the same species are low ( $D=0.016$  for *A. subsimplex*,  $D=0.018$  for *A. cerulata* and  $D=0.050$  for *A. deserti*, see also Figure 3). They are lower compared to the genetic distances estimated among populations of *A. victoriae-reginae* (average genetic distance among populations = 0.182, Martínez-Palacios et al, 1999). *Agave deserti* has genetic distances similar to the ones obtained for *A. lechuguilla* (average distance among populations = 0.07, Silva, 2001), another species with a large distribution, albeit these last studies were done using allozyme markers. The small genetic distances between populations and species could thus indicate that either the species complex has diverged recently and/or the presence of gene flow between species.

In addition, if genetic distances are an indication of reproductive isolation (Wu and Horlocher, 1998), the

results are not necessarily in accordance with current taxonomic divisions. The species category in the *A. deserti* complex is not clearly reflected by the genetic data (Figure 3). This complex may represent just a single genetic species with geographic differentiation. It is also clear from our analysis that *A. cerulata* ssp. *dentiens* is not more differentiated than the other populations sampled from the complex, despite growing in apparent isolation on the San Esteban island and showing morphological differences. Genetic distances were only large for *A. deserti* (La Rumorosa population), the most northern and western population, which appears at the base of the UPGMA dendrogram (Figure 4).

It is important to mention that highly polymorphic neutral markers such as RAPDs, have been criticized for their tendency to underestimate differentiation between populations, since they average differentiation over different loci in the genome, which may not be necessarily the ones responsible for adaptive change (Hedrick, 1999).

#### General patterns of evolution in the complex

*A. deserti* might be a polyphyletic species, since its populations (belonging to different subspecies) group with three different clades in the UPGMA analysis, although there is no statistical support for this assertion. Morphological similarities that exist between *A. cerulata* and *A. subsimplex* seem to be mainly because of parallel adaptation from a similar genetic pool or to the species having retained ancestral characters, since they appear as a relatively distant group in the tree.

In conclusion, the differentiation of the currently described subspecies is not reflected in the genetic distances estimated by RAPDs. Gentry (1982) had already reduced the number of species in the complex, but, at least from a genetic standpoint, there seem to be even fewer species or subspecies than he proposes. Another possibility is that the changes important for speciation might reside in a few loci not detected by the RAPD markers. The problem of whether the complex is a single species could be addressed by quantitative morphological studies in common garden experiments (eg, Colunga-García-Marín and May-Pat, 1997). Together with crosses among individuals from different taxonomic groups, these studies may give us an idea of reproductive isolation and/or the existence of ecotypes within a single species.

Geography seems to be important in differentiating populations of *Agave*. For instance as seen in the UPGMA clusters *A. subsimplex* and *A. deserti* ssp. *simplex* from Sonora fall together, and an *A. deserti* geographically close to *A. cerulata* appears in the cluster including the populations of this species. Nevertheless, only *A. deserti* and *A. cerulata* showed a correlation between geographic distance and genetic distance. The same phenomenon might occur in *A. subsimplex*, but our small sample size (three populations) does not allow a complete investigation.

The results suggest that hybridization among populations of different species might be frequent, as has been observed by Gentry (1982) for *A. deserti* and *A. cerulata*, and could be responsible for the observed pattern of nesting of *A. deserti* ssp. *deserti* in the *A. cerulata* group seen in the UPGMA analysis. Large population sizes may be preventing drift from differentiating them genetically,

although the heterogeneous environment of the Sonoran desert is responsible for relatively different morphologies. Future studies on the reproductive biology of these species together with the estimation of the effective population sizes could further reveal the importance of gene flow in shaping the genetic structure of this complex.

#### Acknowledgements

This work was supported by grants DGAPA-PAPIIT IN21197 and CONACyT -27983N to LEE and a Fundación UNAM and CONACyT scholarship to ANQ. The authors thank Juan Nuñez-Farfán and César Domínguez-Pérez-Tejada for help in project design, during the field work and for comments on previous versions of the study. We also thank Arturo Silva, Nestor Mariano, César Abarca, Martha Rocha and Aldo Valera for technical assistance in the laboratory and with the data analyses. We acknowledge Ricardo Clark and Antonio Hernandez for their help during the field collections. Daniel J Schoen made useful comments on a previous version of the paper.

#### References

- Axelrod DI (1979). Age and origin of Sonoran desert vegetation. *Occas Pap Calif Acad Sci* **132**: 74.
- Besnard G, Baradat PH, Chevalier D, Tagmount A, Berville A (2001). Genetic differentiation in the Olive complex (*Olea europaea*) revealed by RAPDs and RFLPs in the rRNA genes. *Gen Res Crop Evol* **48**: 165–182.
- Colunga-García-Marín P, May-Pat F (1997). Morphological variation of Henequen (*Agave fourcroydes* Agavaceae) germplasm and its wild ancestor (*A. angustifolia*) under uniform growth conditions, diversity and domestication. *Am J Bot* **84**: 1449–1465.
- Crow JF, Aoki K (1984). Group selection for a polygenic behavioral trait, estimating the degree of population subdivision. *Proc Natl Acad Sci USA* **81**: 6073–6077.
- Doyle JJ, and Doyle JL (1987) A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochem Bull* **19**: 11–15.
- Eguiarte LE, Burquez A, Rodríguez J, Martínez-Ramos M, Sarukhan J, Pinero D (1993). Direct and indirect estimates of neighborhood and effective population size in a tropical palm, *Astrocaryum mexicanum*. *Evolution* **47**: 75–87.
- Frankham R (1997). Do island populations have less genetic variation than mainland populations? *Heredity* **38**: 311–327.
- García-Mendoza A, Galván R (1995). Riqueza de las familias Agavaceae y Nolinaceae en México. *Bol Soc Bot Méx* **56**: 7–24.
- Gentry HS (1982). *Agaves of Continental North America*. The University of Arizona Press, Tucson, Arizona.
- Hamrick JL, Godt MJW, Herman-Broyles S (1992). Factors influencing levels of genetic diversity in woody plant species. *New Forests* **6**: 95–124.
- Hamrick JL, Godt MJW (1996). Effects of life history traits on genetic diversity in plant species. *Proc Roy Soc London B* **351**: 1291–1298.
- Hartl DL, Clark AG (1996). *Principles of Population Genetics*. Sinauer Associates: Sunderland, MA.
- Hedrick PW (1999). Perspective, highly variable loci and their interpretation in evolution and conservation. *Evolution* **53**: 313–318.
- Horner MA, Fleming TH, Sahley CT (1998). Foraging behaviour and energetics of a nectar-feeding bat, *Leptonycteris curasoae* (Chiroptera, Phyllostomidae). *J Zool* **244**: 575–586.

- Izquierdo LY, Pinero D (1998). Allozyme divergence among four species of *Podaechmea s.l.* and the status of *Ursulaea* (Bromeliaceae, Bromelioideae). *Plant Syst Evol* **213**: 207–215.
- Kimura M (1983). *The Neutral Theory of Molecular Evolution*. Cambridge University Press: Cambridge.
- Lonsdale P (1989) Geology and tectonic history of the Gulf of California. In: Winterer H, Hussong D, Decker WR (eds.) *The Geology of North America, the Eastern Pacific Ocean and Hawaii*. The Geological Society of America: Boulder, CO, pp 449–521.
- Lynch M, Milligan BG (1994). Analysis of population genetic structure with RAPD markers. *Mol Ecol* **3**: 91–99.
- Manly BFJ (1991). *Randomization and Monte Carlo Methods in Biology*. Chapman & Hall: London.
- Martinez-Palacios A, Eguiarte LE, Furnier GR (1999). Genetic diversity of the endangered endemic *Agave victoria-reginae* (Agavaceae) in the Chihuahuan desert. *Am J Bot* **86**: 1093–1099.
- Massey LK, Hamrick JL (1998). Genetic diversity and population structure of *Yucca filamentosa* (Agavaceae). *Am J Bot* **85**: 340–345
- Nei M (1987). *Molecular Evolutionary Genetics*. Columbia University Press: New York.
- Silva JA (2001) Ecología reproductiva y genética de poblaciones de *Agave lechuguilla* (Torr.) en un gradiente latitudinal. Ph.D. Thesis, Universidad Nacional Autónoma de México.
- Slatkin M (1981). Estimating levels of gene flow in natural populations. *Genetics* **99**: 323–335.
- Slatkin M (1985). Rare alleles as indicators of gene flow. *Evolution* **39**: 53–65.
- Slatkin M (1989). A comparison of three indirect methods for estimating average levels of gene flow. *Evolution* **43**: 1349–1367.
- Sokal RR, Rohlf FJ (1995). *Biometry*. 3rd ed. W.H. Freeman: New York.
- Turner RM, Bowers JE, Burgess TL (1995). *Sonoran Desert Plants*. The University of Arizona Press, Tucson, Arizona.
- Van Devender TR (1990). Late quaternary vegetation and climate of the Sonoran desert, United States and Mexico. In: Betancourt JL, Van Devender JL, Martin PS (eds) *Packrat Middens, the Last 40,000 Years of Biotic Change*. University of Arizona Press, Tucson, Arizona: pp 134–164.
- Weir S (1996). *Genetic Data Analysis II*. Sinauer Associates. Sunderland, MA.
- Williams JGK, Hanafey MK, Rafalski JA, Tingey SV (1990). Genetic analysis using random polymorphic DNA markers. *Methods Enzymol* **218**: 704–740.
- Whitlock MC, MacCauley DE (1998). Indirect measures of gene flow and migration:  $F_{ST} \neq 1/(4N_m+1)$ . *Heredity* **82**: 117–125
- Wu CL, Hollocher H (1998). Subtle is nature: the genetics of species differentiation and speciation. In: Howard DJ, Berlocher SH (eds.) *Endless Forms: Species and Speciation*. Oxford University Press: New York; pp 339–351.