Long-distance colonization, isolation by distance, and historical demography in a relictual Mexican pinyon pine (*Pinus nelsonii* Shaw) as revealed by paternally inherited genetic markers (cpSSRs)

A. CUENCA, A. E. ESCALANTE and D. PIÑERO

Departamento de Ecología Evolutiva, Instituto de Ecología, Universidad Nacional Autónoma de México, Apartado Postal 70–275, México D.F. 04510, México

Abstract

Pinus nelsonii **is a relictual pinyon pine distributed across a wide altitudinal range in semiarid zones in Mexico near the border between the States of Nuevo León and Tamaulipas. It also occurs in small patches in the State of San Luis Potosí.** *Pinus nelsonii* **is classified in the monotypic subsection** *Nelsoniae,* **separated from other pinyon pines (subsection** *Cembroides*), **because it possesses several distinctive characters including persistent fascicle sheaths, connate needles, and a distinctive wood anatomy. In the present study, chloroplast simple sequence repeats (cpSSRs) were used to estimate genetic variation in most known populations (nine) of** *P. nelsonii.* **The genetic variation ($H_T = 0.73$; 27 haplotypes in 256 individuals) is moderate when compared to other pine species. Population differentiation ranged between low and moderate ($F_{ST} = 0.13$ and $R_{ST} = 0.05$), as did the Nei and Goldstein genetic distances between populations. However, this pattern varied depending on whether the infinite alleles or stepwise mutation model was used. In the former case a significant isolation by distance was found, but not in the latter. A significant association between geographical and genetic structure in one clade, through a nested clade analysis, was found, which suggested long-distance colonization between 125 000 and 309 000 years ago. We found weak evidence for a population expansion. A mismatch distribution suggests that** *P. nelsonii* **populations underwent an expansion 4.25 times their size between 59 000 and 146 000 years ago. On the other hand, the populations' star-like phylogeny and a slight parabolic relationship between coalescence times and lineage number also suggest weak population expansion. Overall, this species appears to have been in demographic stasis for a large proportion of the time detected by the markers used.**

Keywords: cpSSRs, historical demography, isolation by distance, long-distance colonization, phylogeography, *Pinus nelsonii*

Received 13 November 2002; revision received 20 February 2003; accepted 25 April 2003

Introduction

Patterns of genetic variability in Mexican pines are important in terms of breeding, biological conservation and evolutionary considerations. Pines form an integral part of the Mexican forests, in which conifer and oak–pine communities cover around one-third (15 million ha) of the present forested areas in the country. Pine lineages in

Correspondence: D. Piñero. Fax: 52 (55) 56161976; E-mail: pinero@servidor.unam.mx

Mexico include species from both the *Pinus* and *Strobus* subgenera. In both there are not only relictual species, usually endangered, with just a few populations, but also species forming complexes that frequently hybridize in sympatry (Lanner & Van Devender 1998). Particularly, within subgenus *Strobus*, subsection *Cembroides*, studies of genetic structure and variability have been published regarding some relictual and endangered species including *P. maximartinezii*, known from only one population (Ledig *et al.* 1999), *P. rzedowskii*, known from nine populations (Delgado *et al.* 1999), and *P. pinceana*, known from

about 20 populations (Ledig *et al.* 2001; Molina *et al.* 2001). Results have consistently shown high levels of genetic variation in all three species and high genetic differentiation in *P. rzedowskii* and *P. pinceana*. Phylogenetically, these three species, together with *Pinus nelsonii*, represent basal lineages within the pinyon pines (Gernandt *et al.* 2001; Gernandt personal communication), while the more recent taxa include a group of species with a still unclear taxonomy (Perry 1991; Farjon & Styles 1997; Price *et al.* 1998; Gernandt *et al.* 2001; Gernandt *et al.* personal communication).

Pinus nelsonii is an endemic species of Mexico. It is distributed in few isolated populations in the Sierra Madre Oriental, in the states of Tamaulipas, Nuevo León and San Luis Potosí. This region is characterized by high mountain ranges surrounded by more or less arid zones and the presence of pine species adapted to aridity and with many local endemic species (Farjon 1996). This region of the Sierra Madre Oriental is ancient when compared with other Mexican mountainous areas both biogeographically (Marshall & Liebherr 2000; Luna-Vega *et al.* 2001) and phylogeographically (Harris *et al.* 2000).

Pinus nelsonii has been generally included within the pinyon pines because it has wingless seeds, though it is now believed to be relictual or an old lineage because morphological and molecular characters distinguish it from other pinyon pines (Farjon & Styles 1997; Liston *et al.* 1999; Gernandt *et al.* 2001). In fact, other authors (Price *et al.* 1998) have suggested that *P. nelsonii* is the only member of the subsection *Nelsoniae*. Despite its restricted distribution, *P. nelsonii* is the second most economically important pinyon pine (after *P. cembroides*) in Mexico (Suzán-Azpiri *et al.* 2002). In addition, this species is subject to special protection by the Mexican environmental law (NOM-ECOL-059-2001), implying that it is necessary to develop conservation strategies.

There are few methods that can distinguish if gene flow or population divergence are responsible for the genetic distance patterns found in natural populations. In fact, geographical patterns of genetic variation have been shown to be the result of present population phenomena, population history, and a combination of both (Templeton *et al.* 1995).

Microsatellites or simple sequence repeats (SSRs) are molecular markers that show high mutation rates and high levels of heterozygosity and polymorphism and their molecular structure and evolution are relatively well understood (e.g. Jarne & Lagoda 1996; Estoup & Cornuet 1999; Xu *et al.* 2000). The most widely used mutational models for microsatellites are the infinite alleles model (Kimura & Crow 1964) and the stepwise mutation model (Kimura & Ohta 1978). Recently, new models of microsatellite mutation have been developed (Di Rienzo *et al.* 1994; Estoup & Cornuet 1999), which try to explain their

mutational dynamics in a more accurate way. Estimates of population structure, number of migrants and effective population size are highly dependent on the mutation model assumed. For microsatellites, in particular, this dependence is strong because of their high mutation rate (Estoup & Angers 1998; Van Oppen *et al.* 2000). DNA markers for haploid, uniparental genomes are more sensitive indicators of population subdivision. This is because in monoecious species, the effective population size is twofold smaller than diploid nuclear genomes (Echt *et al.* 1998; Schaal *et al.* 1998). In addition, these paternal lineages within a species retain a clonal record of new mutations, whereas this record is obscured in genomes that have recombination (Echt *et al.* 1998). As a result of the high mutation rate in microsatellite loci, these markers could show homoplasy. This occurs when different copies of a locus are identical in state, although not identical by descent. This process is related to the way mutation produces new alleles and is expected in most mutation models, except under the infinite allele model. If homoplasy is prevalent in a population this would seriously undermine the conclusions drawn in a phylogeographic study. When considering haplotypes (several loci), as in chloroplast DNA (cpDNA), the extent of homoplasy is less but it still could be a problem.

The phylogeographic approach (Avise 2000) has proved to be important in uncovering comparative patterns of haplotype geographical distribution for different groups of species. In particular, *P. nelsonii* grows in areas with relictual flora and fauna in the Sierra Madre Oriental (Marshall & Liebherr 2000; Luna-Vega *et al.* 2001). Thus, *P. nelsonii* is also a source of data to explore phylogeographic relations within this area.

In this work, our first objective is to study the genetic variation in *P. nelsonii* populations using cpSSRs, as well as to investigate whether there is a geographical structure in this variation. Second, we analyse the historical population phenomena that might have affected the genetic variation in these populations. In particular we would like to test different scenarios that have probably shaped the population genetic structure in *P. nelsonii* populations and uncover the relative importance of long-distance colonization, isolation by distance, population expansion and fragmentation during its history. Finally, we also want to contribute information about the phylogeography and recent history of the Sierra Madre Oriental.

Materials and methods

Plant material

Nine populations of *Pinus nelsonii* were collected in the Mexican states of Tamaulipas, Nuevo León and San Luis Potosí. These cover all the known populations for this species, except for the one at Las Tablas, San Luis Potosí

Table 1 Haplotype frequency in nine Mexican populations of *Pinus nelsonii*

H*	Locus† 1	2	3	4	Population‡ LD ($n = 30$)	LT ($n = 23$)	PM ($n = 18$)	M ($n = 20$)	PN ($n = 28$)	SL ($n = 24$)	P ($n = 26$)	T ($n = 17$)	A ($n = 46$)	Total ($n = 232$)
H1	158	146	153	93	27	20	12	1	16	6	11	9	12	0.491
H2	158	146	153	95	1			1	1					0.013
H3	158	146	153	92	1			1						0.009
H4	159	146	153	93	1	1	4		3	3	4	1	6	0.099
H5	159	145	153	93		1		1					1	0.013
H6	158	147	153	93		1	1	2	2	6	1			0.056
H7	157	145	153	95			1	1						0.009
H8	157	146	153	93				7	2	4	3	2	12	0.129
H9	158	145	153	93				3	1	1		1	6	0.052
H10	159	145	153	95				1						0.004
H11	157	146	153	95				1						0.004
H12	158	147	153	95				1				1		0.009
H13	158	145	154	93					1				1	0.009
H14	159	146	153	95					1					0.004
H15	159	147	153	93					1				2	0.013
H16	158	146	152	93						1	1			0.009
H17	157	145	154	93						1				0.004
H18	157	147	153	93						1		1	1	0.013
H19	158	147	153	94						1				0.004
H20	160	146	153	93							3	1	1	0.022
H21	160	146	153	94							1			0.004
H22	157	146	152	93							1			0.004
H23	158	146	153	94							1	1		0.009
H24	158	145	153	94									1	0.004
H25	158	148	153	93									1	0.004
H26	157	146	154	93									1	0.004
H27	159	147	153	94									1	0.004

*Haplotype code.

†Loci 1, 2, 3 and 4 are the Pt30204, Pt71936, Pt3678 and Pt63718 (Vendramin *et al*. 1996) respectively.

‡The populations are: Los Duraznillos (LD), La Tapona (LT), Puerto Membrillo (PM), Miquihuana (M), Peña Nevada (PN), San Lázaro (SL), Palmillas (P), Tula (T) and Antonias (A).

(Perry 1991), which we could not find. In each population we sampled between 17 and 46 individuals, 50 m apart, along two or three transects depending on tree density (Table 1). DNA was extracted from needles using a modified CTAB protocol (Vázquez-Lobo 1996).

cpSSR markers

Eleven cpDNA pairs of primers derived from *Pinus thunbergii* chloroplast sequence were used: Pt1254, Pt9383, Pt30204, Pt71936, Pt87268, Pt15169, Pt36081, Pt63480, Pt41093, Pt6718, Pt110048 (Vendramin *et al*. 1996). Polymerase chain reactions (PCR) were performed in 25 μL containing 10–50 ng genomic DNA, 10 mM Tris–HCl, 50 mM KCl, 2 mM MgCl$_2$, 200 μM of each dNTP, 5 pmol ^{32}P end-labelled forward primer, 5 pmol forward primer, 10 pmol reverse primer and 0.5 units *Taq* DNA polymerase (Gibco BRL). Reactions were carried out on a PTC-100

Programmable Thermal Controller (MJ Research Inc.) using the following parameters: (1) initial denaturation at 95° for 5 min, (2) 5 min at 80°, (3) 33 cycles of denaturalization at 94° for 1 min, annealing at 60° for 1 min and extension at 72° for 1 min, (4) final extension at 72° for 8 min. Products were resolved on 6% denaturing polyacrylamide gels (7 M urea) at 60 W constant power for 3 h. The plasmid pUC18 sequence was used as a size standard (Life Technologies, Gibco BRL). Gels were transferred onto 3-mM blotting paper (Whatman), dried and exposed to X-ray film overnight without intensification screens.

Statistical analyses

Haplotypic variation within populations was estimated using the unbiased genetic diversity (H_E, Nei 1987), and the global expected heterozygosity or gene diversity for the total sample of individuals (H_T). For both mutational

models (stepwise and infinite allele), θ was estimated indirectly based on heterozygosity (in the case of an haploid marker, θ = 2Nu, were N is the population size and u is the mutation rate; Schneider *et al.* 2000). The mismatch distribution was obtained using the difference in nucleotide size among haplotypes. A χ^2 test was used to assess its statistical significance. These analyses were performed with ARLEQUIN version 2000 (Schneider *et al.* 2000).

A haplotype network was inferred with a statistical parsimony method (Templeton *et al.* 1992; Carbone & Kohn 2001), which interconnects all haplotypes by a series of mutational events derived from a matrix of the number of pairwise nucleotide differences (TCS: estimating gene genealogies, Clement *et al.* 2000). The result is a network of equally parsimonious paths in which haplotypes are placed both as nodes and branch tips. Population structure was also examined using an analysis of molecular variance (AMOVA; Excoffier *et al.* 1992) with ARLEQUIN version 2000 (Schneider *et al.* 2000). The variance components were used to estimate the fixation index and coalescent times as proposed by Slatkin (1995). The genetic distance between pairs of haplotypes was estimated using both the infinite alleles and the stepwise mutation models in which fixation indices were F_{ST} and R_{ST}, respectively. The null distribution of pairwise F_{ST} and R_{ST} values under the hypothesis of panmixia was obtained with 1000 random permutations of haplotypes between populations. Genetic distances among populations were determined using two estimates, D_a (Nei *et al.* 1983) and $(\delta\mu)^2$ (Goldstein *et al.* 1995). A neighbour-joining tree was used to reconstruct population relations (NJBAFD program, Naoko Takezaki, National Institute of Genetics, Japan). The statistical significance between D_a and geographical distance was tested with a Mantel test (Mantel 1967), as well as the correlation between the logarithm of migrants and the geographical distance among populations (Slatkin 1995).

To evaluate clades for significant association between haplotypes and their geographical distribution we used a random, two-way, contingency permutation analysis (Templeton & Sing 1993; Templeton 1998; Carbone & Kohn 2001). Clades with significant haplotype–geographical association from a nested contingency analysis were interpreted using the biological inference key developed by Templeton *et al.* (1992, 1995; GEODIS 2.0, Posada *et al.* 2000). This key is based on the assumptions that older haplotypes are ancestral and should be located in the interior portions of the network, while the more recently derived haplotypes should be on the tips of the network, and older haplotypes are more geographically dispersed than recently derived haplotypes (Carbone & Kohn 2001).

Population expansion is usually studied through different analyses. One is the distribution of genetic mismatches for a population sample; another is the analysis of the haplotypic phylogeny that should be star-like when expansion

has occurred. A third involves the estimation of parameters usually used to study the effect of selection. Finally, other methods predict low levels of geographical structure detected by a molecular variance analysis (AMOVA; Schneider & Moritz 1999). To study the population's demographic history we used a method proposed by Nee *et al.* (1996) and a mismatch distribution analysis (Bernatchez 2001). The Nee *et al.* (1996) approach is based on the distance between nodes in a phylogenetic tree. A neighbour-joining tree was reconstructed with the distance between haplotypes with a stepwise mutation model (using the absolute size differences among haplotypes). We plotted the logarithm of lineage number and the coalescence times (estimated with the genetic distance). In this graphical representation, a parabolic curve, given by a star-like phylogeny, suggests population growth. An exponential curve, given by a structuring phylogeny, is expected from a population of constant size through its history.

Results

Genetic variation

Four (Pt30204, Pt36480, Pt63718 and Pt71936) of the 11 cpSSR loci were polymorphic in 232 *Pinus nelsonii* individuals, having either three or four alleles (Table 1). A total of 27 haplotypes was found, from a theoretical maximum of 192 based on the numbers of alleles. One of these, haplotype H1, was particularly common (49.1% for the whole sample), predominant in every population except in Miquihuana where H8 was the most common (35%). The Antonias population had the largest number of haplotypes (13), while Los Duraznillos, La Tapona and Puerto Membrillo had only four haplotypes each. Haplotype H8, which was the second most frequent (12.9%), occurred in the six populations of Tamaulipas and Nuevo León and was absent from the Southern populations (San Luis Potosí). Finally, haplotype H4 (9.9%) was present in all populations except Miquihuana (Fig. 1, Table 1).

Genetic or haplotypic diversity was highest in Miquihuana (0.868; Table 2), San Lázaro (0.859), and Antonias (0.842). San Luis Potosí populations showed less genetic variation. The smallest expected heterozygosity was found in Los Duraznillos (0.193). Average genetic diversity for the species was 0.727 while mean H_E was 0.640 (Table 2).

Since θ is obtained from heterozygosity data in each population, both follow the same pattern. Also, θ can be interpreted as a relative estimate of population size. The highest θ estimates, using the infinite allele model, were found in Miquihuana and San Lázaro (5.48 and 5.01, respectively), while the smallest estimate corresponds to Los Duraznillos (0.18). The θ estimate for the entire sample was 2.12. In all cases θ estimates were larger using the stepwise mutation model (Table 2).

Population	nh	Haplotypic diversity	($\theta = 2Nu$) SMM	($\theta = 2Nu$) IAM
Los Duraznillos	4	0.193 ± 0.095	0.268 ± 0.108	0.178 ± 0.108
La Tapona	4	0.249 ± 0.116	0.387 ± 0.153	0.247 ± 0.153
Puerto Membrillo	4	0.529 ± 0.117	1.758 ± 0.423	0.841 ± 0.401
Miquihuana	11	0.868 ± 0.064	28.380 ± 3.644	5.480 ± 3.376
Peña Nevada	9	0.669 ± 0.096	4.072 ± 0.774	1.538 ± 0.697
San Lázaro	9	0.859 ± 0.040	24.541 ± 1.992	5.005 ± 1.822
Palmillas	10	0.825 ± 0.061	15.755 ± 1.950	3.780 ± 1.713
Tula	8	0.728 ± 0.114	6.255 ± 1.424	2.062 ± 1.260
Antonias	12	0.842 ± 0.030	19.414 ± 1.184	4.317 ± 1.063
Total	27	0.648 ± 0.083	6.24 ± 0.619	2.12 ± 0.545

nh, number of haplotypes.

θ estimates correspond to the stepwise mutation model (SMM) and the infinite allele model (IAM).

Numbers following haplotypic diversity, SMM and IAM estimates are standard deviations.
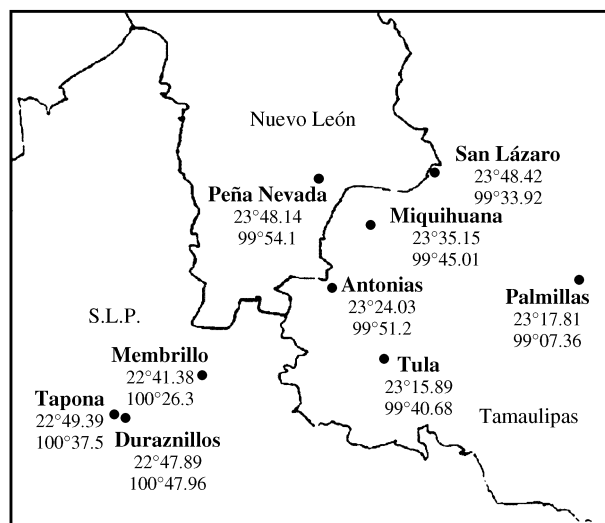
Fig. 1 Geographic distribution of the nine collected Mexican populations of *Pinus nelsonii*.

Genetic differentiation among populations

Analyses of molecular variance results were different depending on which mutation model was used. Both fixation indices were statistically significant ($F_{ST} = 0.131$, $R_{ST} = 0.047$). Miquihuana is the population with the highest genetic divergence when compared to the rest of the populations with both the infinite alleles and the stepwise mutation model. However, the largest differentiation was obtained between Los Duraznillos and Miquihuana (0.47), when D_a was estimated using the infinite alleles model whereas for the stepwise mutation model, we found the largest differentiation between Los Duraznillos and Palmillas (0.78). Also, when the stepwise mutation model was used, no divergence was detected between four population pairs (data not shown). A phenogram among populations using a neighbour-joining algorithm (Fig. 2) shows Miquihuana as the most divergent population and the Palmillas and San Luis Potosí populations (Los Duraznillos, La Tapona and Puerto Membrillo) form the group with the highest bootstrap (91) value. This result is the opposite to that expected based on geographical distances, because Palmillas is geographically closer to the rest of the populations (Fig. 1).

Geographic distribution of haplotypes

Haplotypes with the widest distribution are the central haplotypes in the minimum spanning network (Fig. 3). For the infinite alleles model, we found a significant association between genetic and geographical distances (Mantel 1967; $r = 0.382$; $P = 0.0360$), as well as between Log M (number of migrants per generation; Slatkin 1995) and the geographical distance ($r = 0.434$; $P = 0.0147$; d.f. = 30). Neither of these associations was significant if a stepwise mutation model was used to estimate the distances among haplotypes ($r = -0.037$; $P = 0.3820$ and $r = 0.089$; $P = 0.6859$; d.f. = 22). The same pattern was found when we clustered the populations into two groups (one with populations from the state of San Luis Potosí and the other formed by the rest of the populations). The analysis of molecular variance (AMOVA) shows a small but significant differentiation between these groups using the infinite allele model ($F_{CT} = 0.077$; $P = 0.016$) but not when using the stepwise mutation model ($F_{CT} = 0.018$; $P = 0.810$).
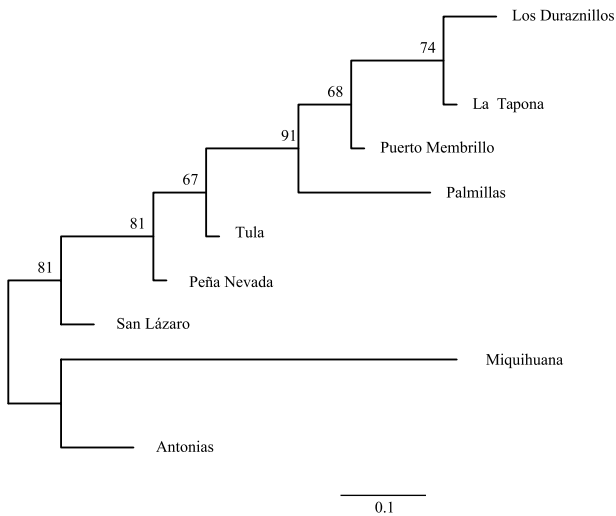
Fig. 2 Unrooted neighbour-joining tree for *Pinus nelsonii* populations based on Goldstein *et al.* (1995) distance among populations $(\delta\mu)^2$. Bootstrap estimates were obtained with 1000 permutations.

Nested clade analysis

The nested clade analysis uses a statistically supported haplotype network to define a nested series of clades. The null hypothesis of no geographical association of genetic variation is not rejected in almost every nested clade,

Table 3 Nested contingency analysis

Clade	χ^2	P
1.9	29.582	0.042
3.1	58.768	0.000
3.2	18.146	0.009

χ^2, Permutational chi-square statistic.

except for clades 1.9, 3.1 and 3.2 (Table 3). Following the biological inference key for each of these clades, only clade 3.1 showed statistically significant geographical dispersion of the haplotypes within a clade (*Dc*; the clade distance) and their dispersion relative to all haplotypes in the next higher clade (*Dn*; the nesting clade distance). Following Templeton's inference key, we found significant long-distance colonization.

Mismatch distribution and demographic history

Mismatch distribution was clearly unimodal corresponding to the number of differences among individuals within lineages (~2; Fig. 4). This distribution is not consistent with a historical population expansion because there is a significant deviation from a sudden expansion model ($\chi^2 = 14731.4$, d.f. = 14). Estimated θ parameters ($\theta_0 = 1.467$
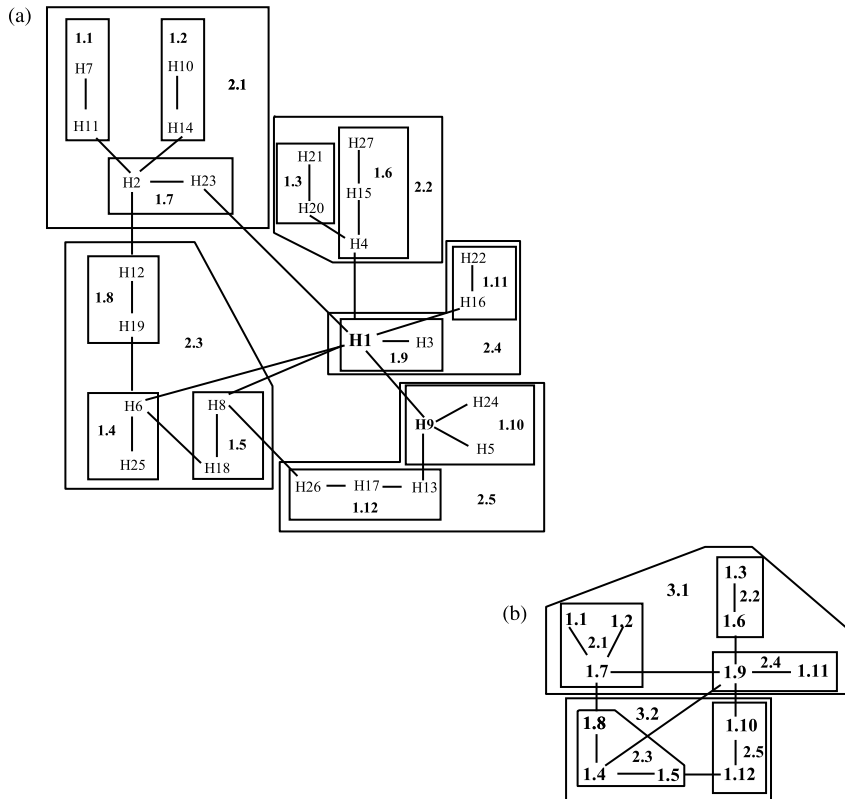

Fig. 3 *Pinus nelsonii* haplotypes network using TCS: estimating gene genealogies version 1.13 (Clement *et al.* 2000). The different clades were constructed following the nested rules (Templeton *et al.* 1992). Each line represents a mutational event. (a) Network of one-step and two-step clades. (b) Network of two-step and three-step clades. There are some alternative connections between haplotypes that are not shown: H13–H16, H8–H22, H27–H19, H8–H25, H23–H24, H23–H19, H6–H15.
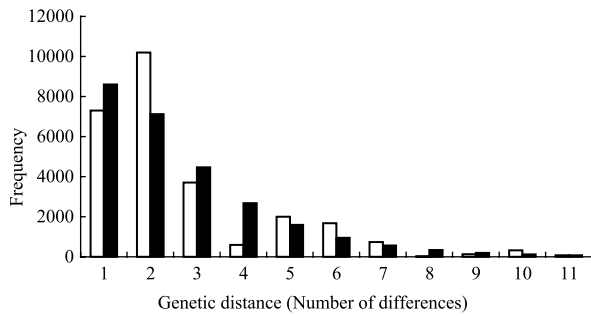
Fig. 4 Mismatch distribution of pairwise differences in size among 27 haplotypes found in *Pinus nelsonii* populations in México. Filled bars represent expected values under a sudden expansion model. Observed mean is 1.67 differences and variance is 3.67. Parameters estimated under the sudden expansion model of Rogers (1995) were: (1) $\theta_0 = 1.411$ (2) $\tau = 0.264$. The χ^2 value is 4731 ($P = 0.000$, d.f. = 14).

Fig. 5 Relation between the logarithm of lineage number and genetic distances among haplotypes (coalescence times; Nee *et al*. 1996). Coalescence times among haplotypes were inferred from the genetic distance among them. The lineage number was shown in an unrooted neighbour-joining tree for the 27 haplotypes found in *Pinus nelsonii* (data not shown).

and $\theta_1 = 6.235$) indicate a slight population growth of *P. nelsonii*. Based in the *P. torreyana* cpSSRs mutation rate (3.2×10^{-5}–7.9×10^{-5}; Provan *et al*. 1999) and the estimated units of mutation time ($\tau = 0.264$), it is possible to estimate that the population grew 4.25-fold between 7281 and 2949 generations ago. If the approximate generation time of *P. nelsonii* is 20 years (17–22; Suzán-Azpiri *et al*. 2002), the event is estimated to have occurred between 146 000 and 59 000 years ago. Accordingly, the analysis by Nee *et al*. (1996) resulted in a slight parabolic relationship (particularly for ancestral clades) between coalescence time and the logarithm of the number of lineages (Fig. 5). This analysis reinforces the conclusion of a population stasis or no significant sudden population expansion found with the mismatch distribution between haplotypes.

Discussion

Genetic diversity

In contrast to the patterns described using molecular markers for other endangered pinyon pines (Ledig *et al*. 1999, 2001; Molina *et al*. 2001), *Pinus nelsonii* is not genetically depauperate, but shows a moderate amount of chloroplast genetic diversity, similar to that found for other pine species with a wider geographical distribution (reviewed in Delgado *et al*. 2002; Marshall *et al*. 2002; Richardson *et al*. 2002).

According to the coalescent theory (Schaal *et al*. 1998), H1 (Table 2) should be the most ancestral haplotype because it shows the largest number of connections with other haplotypes and is central in the network. Also, the network shows a short internodal distance that is consistent with a star-like phylogeny. This is indicative of population expansion. The probability of parsimonious connections across the network is estimated to be > 95%
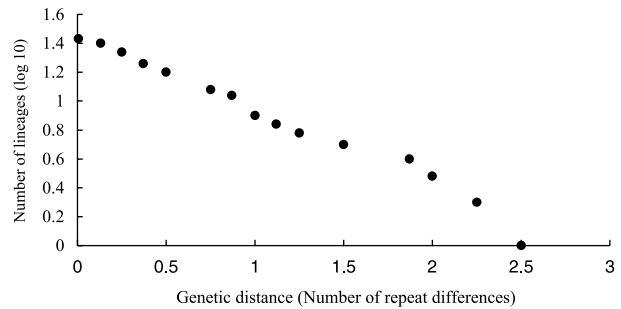
(Clement *et al*. 2000). A moderate level of homoplasy is apparent in the network, resulting in closed loops. The presence of one very common haplotype in all populations suggests that this species went through a bottleneck after which new allelic variants have originated in low frequencies. This scenario is reinforced because we have also shown evidence of historic population stasis evidenced from the mismatch distribution (Fig. 4) and the results from Nee *et al*. (1996) (Fig. 5). On the other hand, recent population stasis is not supported by the fact that effective population size estimates from heterozygosity are about 10 000–23 500, which are relatively low when compared to the densities estimated from population demography (Suzán-Azpiri *et al*. 2002) and distribution area for this species (Perry 1991; Farjon & Styles 1997).

Of particular interest is the low variation found in the San Luis Potosí populations, mainly Los Duraznillos and La Tapona where we found the most frequent haplotype (H1) almost exclusively. This could suggest either that these populations went through a recent bottleneck or were recently colonized.

The average expected coalescence time in generations is twice the effective number of females in a population assuming equilibrium between drift and mutation and a low variance in family size (Hartl & Clark 1989). Estimates of θ for the infinite allele model (Table 2) show a large variation among populations (from 0.178 in Los Duraznillos to 5.480 in Miquihuana). We do not have estimates of mutation rates for *P. nelsonii*, however, Provan *et al*. (1999) estimated mutation rates of 3.2×10^{-5}–7.9×10^{-5} for chloroplast markers for *P. torreyana*. If we assume similar rates for *P. nelsonii*, coalescence time estimates range from as low as 20 000–55 000 years for Los Duraznillos to 700 000–1 752 000 years for Miquihuana. The average for all populations is 330 000–820 000 years, which are approximate coalescence times for the whole species. Approximate coalescence times for nuclear genes should be about twice

those found for chloroplast markers (Schaal *et al.* 1998). This would move the coalescence times for the species to 660 000–1 640 000 years. These estimates are low when compared to the divergence times estimated for the divergence of *P. nelsonii* from its closest relatives obtained with internal transcribed spacer sequences (49 million years ago, Gernandt *et al.* 2001).

Genetic differentiation among populations

There is currently no simple model that describes with precision the population dynamics of microsatellites in species with contrasting genetic structures (Di Rienzo *et al.* 1994). Consequently neither of the two alternative models (infinite allele model and stepwise mutation model) could be used for contrasting genetic structures. The genetic structure data of *P. nelsonii* are no exception, and we found discrepancies in the results obtained for these two models. Nevertheless, it is possible to find certain common conclusions for the results with these two models in relation to the way *P. nelsonii* populations are genetically structured. Both models show a significant population differentiation (Table 2), but a larger differentiation is apparent when the infinite allele model is assumed (F_{ST}), which is an indication of recent population divergence (Hedrick 1999). Also, some estimates of pairwise distance among populations are negative which could be caused by a high allele size variance within populations. The same has been described for *P. albicaulis* in which F_{ST} was much more sensitive than R_{ST} in detecting population genetic structure in allopatric populations or populations of recent divergence (Richardson *et al.* 2002). The same pattern was observed in two sheep species *Ovis aries* and *O. canariensis* (Forbes *et al.* 1995) and in *Sorex araneus*, the common shrew (Lugon-Moulin *et al.* 1999), where R_{ST} was a better predictor of population divergence in cases of more ancient historical divergence, such as in the case of two close species or races, while F_{ST} appeared to be more sensitive for the detection of intraspecific differentiation, or much more recent divergence events. This could be the result of the time and mutation rate needed (500 generations and 10^{-4}; Goldstein *et al.* 1995) for R_{ST} to detect significant population structure (e.g. Richardson *et al.* 2002). On the other hand, the high variance found in the within-population variation produced by differences in allelic sizes originated a number of negative R_{ST} estimates (not shown) that do not have biological significance. Furthermore, despite the fact that both models assume a large number of alleles, restrictions to allelic size will have a more profound effect on R_{ST} (Nauta & Weissing 1996). These could be related to the significant genetic structure and correlation between geographical and genetic distance with the infinite allele model, when compared to the stepwise mutation model.

Long-distance colonization

Long-distance colonization can be deduced using at least two different approaches. The first approach uses the congruence between phylogenetic lineages and their geographical distribution. This approach has been used, for example, to detect multiple events of transoceanic colonization in Gekkonid lizards (Carranza *et al.* 2000). The second approach is to use nested clade analysis (e.g. in the fungus *Schizolobium commune*; James *et al.* 2001). Following this second approach we found significant long-distance colonization for one of the two most inclusive clades, 3.1. Tomback & Linhart (1990) have described the way in which colonization could take place at long distances in pinyon species in which dispersal is probably carried out by bird species. Unfortunately, there are no data on the dispersal of *P. nelsonii*, but estimates of coalescence times for this clade (125 000–309 000 years ago assuming a molecular clock) suggest that this dispersal mechanism did have an important historical impact on the geographical distribution of this species.

Isolation by distance

Isolation by distance can be studied through the correlation of genetic and geographical distances and also through the application of the nested clade analysis. Isolation by distance through restricted gene flow has been described for many plant species (e.g. Olsen 2002). Recent analyses have found contrasting results for pinyon pines. Ledig *et al.* (2001) found a statistically significant association for *P. pinceana* and, contrary to Delgado *et al.* (1999), found no significant association for *P. rzedowskii*. We found, for clade 2.3 (Fig. 3), evidence for restricted gene flow and isolation by distance without significant nested contingency analysis (Table 3). In more recent clades, we did not find a significant association between haplotype distribution and their geographical distribution. This could be the result of an inadequate sampling strategy to detect geographical structure, though this possibility can be ruled out in *P. nelsonii* because we sampled all but one of the populations that have been described. Another possibility would be that the population is panmictic and the differences in haplotypic frequencies could be the result of sampling. This last possibility contradicts the fact that we found for *P. nelsonii* a significant fit to the isolation by distance model with the infinite allele model (F_{ST}).

The main limitation of *F* statistics when microsatellites are used as markers is their sensitivity to the mutation rate when migration rate is low (Hedrick 1999; Balloux & Lugon-Moulin 2002). In cases where migration rate is sufficiently large to overcome the mutation effect, F_{ST} is usually the best estimator because it has a lower variance than R_{ST}. That the nested clade analysis did not detect

isolation by distance could be because of the recent divergence and the high among-population migration rate in this species. Conversely, high mutation rates might produce homoplasy among different alleles, part of which could be deduced from the alternative routes found in the haplotypic network (Fig. 3).

Population expansion

Pinus nelsonii has some specific characteristics typical of species that have gone through a population expansion. These include a wide distribution of the older haplotypes and a haplotypic genealogy with a star-like form, with very few ancestral haplotypes and a high proportion of derived ones. These derived haplotypes are expected, in a population expansion scenario, to be in low frequencies at only one or two mutational steps from ancestral haplotypes. Also, AMOVA shows relatively low levels of genetic structure (see Results section). Finally, the observed mismatch distribution is statistically different from the one expected under a sudden population expansion model (Fig. 4).

For *P. nelsonii*, a population expansion is supported by the analysis by Nee *et al.* (1996) (Fig. 5), where a large proportion of comparisons among haplotypes occur in relatively short coalescence times. This is despite the fact that the mismatch distribution did not significantly fit a sudden expansion model (Rogers & Harpending 1992; Harpending 1994). The estimated parameters in the mismatch analysis suggest an increase of 4.25 times the population size between 146 000 and 59 000 years ago. Unfortunately there are no palaeobotanical records from northeastern Mexico that would enable us to relate these dates to environmental phenomena from those times. We found a clade (2.3) with significant Dc and Dn but without significance in the nested contingency analysis. We followed the inference key in this case and found evidence for restricted gene flow and isolation by distance; the same evidence was found using the infinite allele model. This would mean that some events could not be detected efficiently with a nested contingency analysis (Table 3). Similarly, Paulo *et al.* (2002) found, for lizards from the Iberian peninsula, that area expansion was not detected for ancestral haplotypes using this same approach.

Our phylogeographic inferences could not be accurate if homoplasy is prevalent in cpDNA microsatellite loci in *P. nelsonii*. Estoup *et al.* (2002) concluded using simulations that the large amount of variability at microsatellite loci often largely compensates for their homoplasious evolution. The cases in which size homplasy may be a problem include high mutation rates and large population sizes. There are no estimates of mutation rates in *P. nelsonii* but Provan *et al.* (1999) found a low mutation rate for another pine species, *P. torreyana*. Also, *P. nelsonii* does not have large demographic population sizes (Suzán-Azpiri *et al.*

2002). Also, homoplasy does not occur under the infinite allele model. We tested different mutation models (Cornuet & Luikart 1996) and found that three out of the four loci that we used are more closely described by the infinite allele model while just one (Pt30204) was statistically better described by the stepwise mutation model (data not shown) which suggests that our inferences are reasonably supported.

In conclusion, we have shown that genetic structure in *P. nelsonii* can be explained by a combination of historical processes, that include a slight population expansion and long-distance colonization, and more recently restricted gene flow with isolation by distance that results in a moderate genetic variation and genetic differentiation for chloroplast markers.

References

Avise JC (2000) *Phylogeography*. Harvard University Press, Cambridge, MA.

Balloux F, Lugon-Moulin N (2002) The estimation of population differentiation with microsatellite markers. *Molecular Ecology*, **11**, 155–165.

Bernatchez L (2001) The evolutionary history of brown trout (*Salmo trutta* L.) inferred from phylogeographic, nested clade, and mismatch analyses of mitochondrial DNA variation. *Evolution*, **55**, 351–379.

Carbone I, Kohn LM (2001) A microbial population–species interface: nested cladistic and coalescent inference with multilocus data. *Molecular Ecology*, **10**, 947–964.

Carranza S, Arnold EN, Mateo JA, López-Jurado LF (2000) Long-distance colonization and radiation in gekkonid lizards, *Tarentola* (Reptilia: Gekkonidae), revealed by mitochondrial DNA sequence. *Proceedings of the Royal Society of London Series B*, **267**, 637–649.

Clement M, Derington J, Posada D (2000) *TCS: Estimating Gene Genealogies, Version 1.13*. Brigham Young University, Provo, UT.

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

Delgado P, Piñero D, Chaos A, Pérez-Nasser N, Alvarez-Buylla E (1999) High population differentiation and genetic variation in the endangered Mexican pine, *Pinus rzedowskii* (Pinaceae). *American Journal of Botany*, **86**, 669–676.

Delgado P, Cuenca A, Escalante AE, Piñero D (2002) Comparative genetic structure in pines: evolutionary and conservation consequences. *Revista Chilena de Historia Natural*, **75**, 27–37.

Di Rienzo A, Peterson AC, Garza JC, Valdes AM, Slatkin M (1994) Mutational processes of simple-sequence repeat loci in human populations. *Proceedings of the National Academy of Sciences of the USA*, **91**, 3166–3170.

Echt CS, DeVerno L, Anzidei M, Vendramin GG (1998) Chloroplast microsatellites reveal population genetic diversity in red pine, *Pinus resinosa* Ait. *Molecular Ecology*, **7**, 307–316.

Estoup A, Angers B (1998) Microsatellites and minisatellites for molecular ecology: theoretical and empirical considerations. In: *Advances in Molecular Ecology* (ed. Carvalho GR), pp. 56–89. IOS Press, Amsterdam, the Netherlands.

Estoup A, Cornuet JM (1999) Microsatellites evolution: inferences from population data. In: *Microsatellites. Evolution and Applications* (eds Golsdstein DB, Schlotterer C), pp. 49–79. Oxford University Press, Oxford.

Estoup A, Jarne P, Cornuet JM (2002) Homoplasy and mutation model at microsatellite loci and their consequences for population genetic analysis. *Molecular Ecology*, **11**, 1591–1604.

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

Farjon A (1996) Biodiversity of *Pinus* (Pinaceae) in Mexico: speciation and paleo-endemism. *Botanical Journal of the Linnean Society*, **121**, 365–384.

Farjon A, Styles BT (1997) *Pinus (Pinaceae). Flora Neotropica*, Monograph 75. New York Botanical Garden, New York.

Forbes SH, Hogg JB, Buchannan FC, Crawfor A, Allendorf FW (1995) Microsatellite evolution in congeneric mammals: domestic and bighorn sheep. *Molecular Biology and Evolution*, **12**, 1106–1113.

Gernandt DS, Liston A, Piñero D (2001) ITS region variation in *Pinus* subsection *Cembroides*: implications for molecular systematic studies in pine species complexes. *Molecular Phylogenetics and Evolution*, **21**, 449–467.

Goldstein D, Ruiz BA, Cavalli-Sforza L, Feldman W (1995) Genetic absolute dating based on microsatellites and their origin of modern humans. *Proceedings of the National Academy of Sciences of the USA*, **92**, 6723–6727.

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

Harris D, Rogers DS, Sullivan J (2000) Phylogeography of *Peromyscus furvus* (Rodentia; Muridae) based on cytochrome *b* sequence data. *Molecular Ecology*, **9**, 2129–2135.

Hartl DL, Clark AG (1989) *Principles of Population Genetics*, 2nd edn. Sinauer Associates, Sunderland, MA.

Hedrick PW (1999) Highly variable loci and their interpretation in evolution and conservation. *Evolution*, **53**, 313–318.

James TY, Moncalvo JM, Li S, Vilgalys R (2001) Polymorphism at the ribosomal DNA spacer and its relation to breeding structure of the widespread mushroom *Schizolobium commune*. *Genetics*, **157**, 149–161.

Jarne P, Lagoda JL (1996) Microsatellites, from molecules to populations and back. *Trends in Ecology and Evolution*, **11**, 424–429.

Kimura M, Crow JF (1964) The number of alleles that can be maintained in a finite population. *Genetics*, **49**, 725–738.

Kimura M, Ohta T (1978) Stepwise mutation model and distribution of allelic frequencies in a finite population. *Proceedings of the National Academy of Sciences of the USA*, **75**, 2868–2872.

Lanner RM, Van Devender R (1998) The recent history of pinyon pines in the American Southwest. In: *Ecology and Biogeography of Pinus* (ed. Richardson DM), pp. 171–180. Cambridge University Press, Cambridge, UK.

Ledig FT, Conkle MT, Bermejo B, *et al.* (1999) Evidence for an extreme bottleneck in a rare Mexican pinyon: Genetic diversity, disequilibrium and the mating system in *P. maximartinezii*. *Evolution*, **53**, 91–99.

Ledig FT, Capó-Arteaga M, Hodgskiss Sbay H, Flores-López C, Conkle MT, Bermejo-Velázquez B (2001) Genic diversity and the mating system of a rare mexican piñon, *Pinus pinceana*, and a comparison with *Pinus maximartinezii* (Pinaceae). *American Journal of Botany*, **88**, 1977–1987.

Liston A, Robinson WA, Piñero D, Alvarez-Buylla E (1999) Phylogenetics of *Pinus* (Pinaceae) based on nuclear ribosomal DNA internal transcribed spacer region sequences. *Molecular Phylogenetics and Evolution*, **11**, 95–109.

Lugon-Moulin NA, Wyttenbach H, Brunner H, Hausser J, Goudet J (1999) Hierarchical analysis of genetic differentiation in a hybrid zone of *Sorex araneus* (Insectivora, Soricidae). *Molecular Ecology*, **8**, 419–431.

Luna-Vega I, Morrone JJ, Alcántara-Ayala O, Espinosa-Organista D (2001) Biogeographical affinities among neotropical cloud forests. *Plant Systematics and Evolution*, **228**, 229–239.

Mantel N (1967) The detection of disease clustering and a generalized regression approach. *Cancer Research*, **27**, 209–220.

Marshall CJ, Liebherr JK (2000) Cladistic biogeography of the Mexican transition zone. *Journal of Biogeography*, **27**, 203–216.

Marshall HD, Newton C, Ritland K (2002) Chloroplast phylogeography and evolution of highly polymorphic microsatellite in Lodgepole pine (*Pinus contorta*). *Theoretical and Applied Genetics*, **104**, 367–378.

Molina-Freaner F, Delgado P, Piñero D, Pérez-Nasser N, Alvarez-Buylla E (2001) Do rare pines need different conservation strategies? Evidence from three mexican rare species. *Canadian Journal of Botany*, **79**, 131–138.

Nauta MJ, Weissing FJ (1996) Constraints on allelic size at microsatellite loci: implications for genetic differentiation. *Genetics*, **143**, 1021–1032.

Nee S, Holmes EC, Rambaut A, Harvey PH (1996) Inferring population history from molecular phylogenies. In: *New Uses for New Phylogenies* (eds Harvey P, Leigh Brown A, Maynard Smith J, Nee S), pp. 66–80. Oxford University Press, Oxford.

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

Nei MF, Tajima F, Tateno Y (1983) Accuracy of estimated phylogenetic trees from molecular data. II. Gene frequencies data. *Journal of Molecular Evolution*, **19**, 153–170.

Olsen KM (2002) Population history of *Manihot esculenta* (Euphorbiaceae) inferred from nuclear DNA sequences. *Molecular Ecology*, **11**, 901–911.

Paulo OS, Jordan WC, Bruford MW, Nichols RA (2002) Using nested clade analysis to asses the history of colonization and the persistence of population of an Iberian lizard. *Molecular Ecology*, **11**, 809–819.

Perry JP (1991) *The Pines of Mexico and Central America*. Timber Press, Portland, OR.

Posada D, Crandall KA, Templeton AR (2000) GEODIS: a program of the cladistic nested analysis of the geographical distribution of the genetic haplotypes. *Molecular Ecology*, **9**, 487–488.

Price RA, Liston A, Strauss SH (1998) Phylogeny and systematics

of *Pinus*. In: *Ecology and Biogeography of Pinus* (ed. Richardson DM), pp. 49–93. Cambridge University Press, Cambridge.

Provan J, Soranzo N, Wilson N, Goldstein DB, Powell W (1999) A low mutational rate for chloroplast microsatellites. *Genetics*, **153**, 943–947.

Richardson BA, Brunsfeld J, Klopfenstein NB (2002) DNA from bird-dispersed seed and wind-disseminated pollen provides insights into postglacial colonization and population genetic structure of white bark pine (*Pinus albicaulis*). *Molecular Ecology*, **11**, 215–227.

Rogers AR (1995) Genetic evidence for a Pleistocene population explosion. *Evolution*, **49**, 608–615.

Rogers AR, Harpending H (1992) Population growth makes waves in the distribution of pairwise genetic differences. *Molecular Biology and Evolution*, **9**, 552–569.

Schaal BA, Hayworth DA, Olsen KM, Rauscher JT, Smith WA (1998) Phylogeographic studies in plants: problems and prospects. *Molecular Ecology*, **7**, 465–474.

Schneider C, Moritz C (1999) Rainforest refugia and evolution in Australia's wet tropics. *Proceedings of the Royal Society of London Series B-Biological Sciences*, **266**, 191–196.

Schneider S, Roessli D, Excoffier L (2000) *Arlequin, Version 2000. A Software for Population Genetic Data Analysis*. Genetics and Biometry Laboratory. University of Geneva, Geneva Switzerland.

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

Suzán-Azpiri H, Sánchez-Ramos G, Martínez-Ávalos JG, Villa-Melgarejo S, Franco M (2002) Population structure of *Pinus nelsoni* Shaw, an endemic pinyon pine in Tamaulipas, Mexico. *Forest Ecology and Management*, **165**, 193–203.

Templeton AR (1998) Nested clade analyses of phylogeographic data: testing hypotheses about gene flow and population history. *Molecular Ecology*, **7**, 381–397.

Templeton AR, Sing CF (1993) A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping IV. Nested analyses with cladogram uncertainly and recombination. *Genetics*, **134**, 659–669.

Templeton AR, Crandall KA, Sing CF (1992) A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping and DNA sequence data. III. Cladogram estimation. *Genetics*, **132**, 619–633.

Templeton AR, Routman E, Phillips CA (1995) Separating population structure from population history: a cladistic analysis of the geographical distribution of mitochondrial DNA haplotypes in the tiger salamander, *Ambystoma tigrinum*. *Genetics*, **140**, 167–782.

Tomback DF, Linhart YB (1990) The evolution of bird-dispersed pines. *Evolutionary Ecology*, **4**, 185–219.

Van Oppen M, Rico C, Turner GR, Hewitt G (2000) Extensive homoplasy, non stepwise mutations and shared ancestral polymorphism at a complex microsatellite locus in lake Malawi Cichlids. *Molecular Biology and Evolution*, **17**, 489–498.

Vázquez-Lobo A (1996) Filogenia de hongos endófitos del género *Pinus* L. implementación de técnicas moleculares y resultados preliminares. Tesis de Licenciatura, Facultad de Ciencias, UNAM.

Vendramin G, Lelli L, Rossi P, Morgante M (1996) A set of primers for the amplification of 20 chloroplast microsatellites in *Pinaceae*. *Molecular Ecology*, **5**, 111–114.

Xu X, Peng M, Fang Z, Xu X (2000) The detection of microsatellites mutation is dependent upon length. *Nature*, **24**, 396–399.

Argelia Cuenca is a Master's student in Biological Sciences at the National University of México (UNAM) on the phylogeography of *Pinus cembroides* in Mexico. Ana Elena Escalante is now working on her PhD thesis in Biomedical Sciences at UNAM on population genetics of archeobacteria. Daniel Piñero is a researcher at the Ecology Institute at UNAM and has been studying the evolution of Mexican conifers for several years.