Theor Appl Genet (1998) 97: 456–463

G. G. Vendramin · M. Anzidei
A. Madaghiele · G. Bucci

Distribution of genetic diversity in *Pinus pinaster* Ait. as revealed by chloroplast microsatellites

Received: 5 November 1997 / Accepted: 5 January 1998

Abstract Using nine chloroplast simple sequence repeats (cpSSRs) markers, we evaluated haplotypic variation within and among natural populations of Maritime pine (*Pinus pinaster* Ait.) in order to shed light on the history of this species. Seven out of the nine cpSSRs analysed were polymorphic, giving a total of 24 different variants. The 24 variants combined in 34 different haplotypes. The populations which generally showed the lowest level of haplotypic diversity are those located in Portugal. The Landes (France) and Pantelleria (Italy) populations represent the two main reservoirs of haplotypic diversity. The proportion of genetic differentiation among populations, estimated using R_{st}, which is a measure based upon a strict stepwise mutation model, was 0.235. The high level of differentiation was also confirmed by the AMOVA analysis ($\Phi_{ST} = 0.254$, $P < 0.001$). Four main groups of populations were identified on the basis of Principal Component Analysis, with the differences being statistically significant ($\Phi_{CT} = 0.299$, $P < 0.001$). Based on our results the presence of refugia located in the South of Portugal, previously proposed for this species, may be excluded, and a different possible recolonization process of Maritime pine in the post-glacial period has been proposed. Populations from North Africa and France might have represented a starting point of the recolonization process of Portugal and of the Italian part of the natural range, respectively. This hypothesis seems to be confirmed by the analysis of the distribution of the pairwise differences among individuals within populations: Landes and Pantelleria populations showed a bimodal distribution, as would be expected for ancient gene pools.

Key words cpSSR · Haplotypic diversity · Recolonization pathway · Evolutionary history

Communicated by P. M. A. Tigerstedt

G. G. Vendramin (✉) · M. Anzidei · A. Madaghiele · G. Bucci
Istituto Miglioramento Genetico Piante Forestali,
Consiglio Nazionale delle Ricerche, Via Atto Vannucci 13,
50134 Firenze, Italy
Fax: +39 55 486604
E-mail: vendramin@imgpf.fi.cnr.it

Introduction

A knowledge of the evolutionary history of a species is of particular importance for devising conservation strategies of genetic resources. The short time that has elapsed since the last glaciation, in evolutionary terms, and the low number of generations that have occurred (due to the long life-cycles of forest species) has surely played a fundamental role in determining the amount and actual distribution of the genetic resources of conifers.

Analysis of polymorphism of the uniparentally inherited genomes may be very useful for collecting information about the evolutionary history of populations. In particular the chloroplast (cp) genome, for which paternal inheritance has been demonstrated in *Pinus* (Cato and Richardson 1996; Watano et al. 1996) and *Abies* (Vendramin and Ziegenhagen 1997) species, has several features which may facilitate molecular evolutionary analyses; for example, its small size and the large amount of information on its structure. Recently, non-coding chloroplast intergenic regions were explored for intraspecific variation in conifers; low intraspecific variation was found in *Abies alba*, (Ziegenhagen et al. 1995), and a complete absence was found in *Pinus leucodermis*, (Boscherini et al. 1994). Most recently, the analysis of chloroplast simple sequence repeats (cpSSRs, microsatellites) revealed a larger polymorphism than that previously found in conifer cpDNA, due to the very high mutation rate of these regions (e.g. Powell et al. 1995; Vendramin and Ziegenhagen 1997; Echt et al. 1998).

Microsatellite variants are supposed to be generated by a mechanism of adding and subtracting (with equal probability) a single repeat to or from the current allele. Under this stepwise mutation model (Valdes et al. 1993), similar-sized alleles are less different in terms of mutational steps than alleles with larger differences in size and, consequently, it can be considered that the process of mutation has a memory (Jarne and Lagoda 1996). Therefore, genetic distances and population parameters based on microsatellite markers may be correctly estimated taking into account the specific mutational mechanism and including such theoretical considerations into equations. Simulation work has provided linear relationships between genetic distance based on the size differences of the SSR alleles and the time of divergence (Di Rienzo et al. 1994; Slatkin 1995; Goldstein et al. 1995; Feldman et al. 1997). The fact that the chloroplast genome does not undergo recombination and that chloroplast microsatellites show a high rate of variation allows the accumulation of a relevant number of time-related mutations. Therefore, using information about the SSR mutation process enables phylogenetic inferences based on chloroplast lineages to be made. In general, cytoplasmic markers are expected to be very useful for assesing the post-glacial recolonization pathways that are directly dependent on pollen and seed dispersal into new habitats.

Maritime pine (*Pinus pinaster* Ait.) is one of the most important species in the Mediterranean area. Its natural distribution is limited to the western Mediterranean basin, where it occurs as scattered nuclei of different sizes from Portugal to Italy, including also some islands such as Sardinia, Corsica and Pantelleria. Some small and scattered populations are also present in Morocco. It has been hypothesized that the actual distribution of this species is the result of events that occurred during the last glaciation (Baradat and Marpeaux 1988). The typical scattered distribution may have prevented, or limited, gene flow among the different groups of populations, thereby determining genetic drift and high genetic divergence among populations. It has been reported that uniparentally inherited markers, like chloroplast microsatellites, should be more sensitive indicators of the possible effects of bottlenecks and genetic drift because of their smaller effective population size with respect to diploid nuclear markers (Mitton 1993).

Our main objective of the study presented here was to make phylogenetic inferences using chloroplast microsatellites (cpSSRs, Powell et al. 1995; Vendramin et al. 1996) in *Pinus pinaster* Ait. Analyses of the genetic variation were carried out taking into account specific mutation mechanisms in order to shed light on the history of natural populations of this species.

Materials and methods

Plant material, DNA extraction and sizing of polymerase chain reaction (PCR) products

Seeds were collected from 30 randomly selected trees (at least 100 m apart) from ten natural populations sampled in different parts of the natural range (Table 1). For the Tamjout (Morocco) population no details on the sampling scheme were available. For each population, DNA was extracted from 24 germinated embryos following the protocol developed by Doyle and Doyle (1994). Primers designed from the chloroplast genome of *Pinus thunbergii* (Vendramin et al. 1996) were used to amplify nine microsatellites regions (Pt9379, Pt15169, Pt26081, Pt30204, Pt36480, Pt41093, Pt71936, Pt87268, Pt110048; codes as in Vendramin et al. 1996). PCR amplifications were performed as previously described (Vendramin et al. 1996) using a Perkin Elmer model 9600 thermal cycler: a final volume of 25 μl contained 0.2 m*M* dNTPs, 2.5 m*M* $MgCl_2$, 0.2 μ*M* of primers

Table 1 Geographical and genetic characteristics of the ten populations of *Pinus pinaster* analysed. [n_e = Effective number of haplotypes. H_e unbiased haplotypic diversity (after Nei, 1987). *Mean* D^2_{sh} average distance of individuals within populations (stepwise haplotype approach; Echt et al. 1998; see also Materials and methods). S_{within}, S_{all} unbiased within-population and total variance of allele size, respectively (after Slatkin, 1995; see Materials and methods). R_{ST} genetic divergence among populations based on a stepwise mutation model (after Slatkin, 1995)]

Population	Label	District (State)	Country	Latitude	Longitude	n_e	H_e	Mean D^2_{sh}
Tamjout	MOR	–	Morocco	33° 52′ N	4° 02′ W	4.235	0.797	0.330
Medoc	LAN	Landes	France	45° 20′ N	0° 58′ E	4.114	0.790	1.093
Monte Pino	SAR	Sardinia	Italy	40° 55′ N	9° 25′ E	4.966	0.833	0.478
Poggio Adorno	TUS	Tuscany	Italy	43° 45′ N	10° 44′ E	3.945	0.779	0.216
La Spezia	LIG	Liguria	Italy	44° 06′ N	9° 50′ E	3.236	0.721	0.264
Montagna Grande	PAN	Pantelleria	Italy	36° 48′ N	12° 00′ E	4.174	0.793	1.554
Valle Restonica	COR	Corsica	France	42° 18′ N	9° 20′ E	4.966	0.833	0.557
Alcacer do Sal	ALC	Setubal	Portugal	38° 23′ N	8° 31′ W	2.102	0.547	0.089
Moncao/Valinha	MON	Viana	Portugal	42° 04′ N	8° 23′ W	1.920	0.500	0.079
Marinha Grande	LEI	Leiria	Portugal	39° 46′ N	8° 56′ W	3.840	0.772	0.465
Average						3.750	0.737	0.512
Overall						5.216	0.812	0.621
S_{within}								2.306
S_{all}								3.015
R_{ST}								0.235

(either 5′-fluoresceine labelled), 10 × reaction buffer (Pharmacia), 25 ng of template DNA and 1 unit of *Taq* polymerase (Pharmacia). Amplifications were performed using the following profile: 5 min at 95°C, 1 min at 80°C during which the enzyme was added, followed by 25 cycles of 1 min at 94°C, 1 min at 55°C and 1 min at 72°C, with a final extension step of 8 min at 72°C. PCR products were sized on an ALF automatic sequencer (Pharmacia). Up to three microsatellites fragments (different in size) were multiplexed before being loaded on a 6%, 20-cm-long denaturating polyacrylamide gel (Pharmacia) and run at 35 W constant power for approximately 80 min. Gels were run at least twice. Sizing of the amplified fragments was carried out by Fragment Manager version 1.1 conversion software (Pharmacia) using both external and internal standards (50, 100, 150 and 200 bps) as reference molecular weights. All of the amplification reactions and the sample mixture were automatically prepared using a robotic workstation Biomek 2000 (Beckman Instruments).

Data analysis

Haplotypes were inferred from individual allele size profiles at the nine cpSSR loci analysed. The effective number of haplotypes for each population was calculated as $n_e = 1/(\Sigma p_i^2)$, as well as the unbiased haplotype diversity as $H_e = [n/(n-1)] \cdot (1 - \Sigma p_i^2)$, where p_i is the frequency of the i-th haplotype and n is the number of individuals analysed (Nei 1987; Avise 1994).

Genetic distances among individuals within populations were estimated using the $\bar{D}_{sh}^2$ measure (stepwise haplotype approach, Echt et al. 1998)

$$\bar{D}_{sh}^2 = \frac{1}{[n \cdot (n-1)]/2} \cdot \frac{1}{L} \cdot \sum_{i=1}^{n} \sum_{j=(i+1)}^{n} \left[\sum_{k=1}^{L} |a_{ik} - a_{jk}| \right]^2$$

where n is the total number of individual analysed within each population, L is the number of loci, a_{ik} and a_{jk} are the allele size of the i-th and j-th individual at the k-th locus, respectively and $|a_{ik} - a_{jk}|$ is the absolute difference between allele size of individuals considered.

Within-population haplotypic diversity was also estimated by calculating the parameter S_w (Slatkin 1995, eq. 9), based upon a strict stepwise mutation model.

Genetic differentiation among populations was assessed by R_{ST} following Slatkin (1995, eq. 13), which is the ratio of the between-population variance and the total variance of allele size.

A hierarchical analysis of molecular variance (AMOVA, Excoffier et al. 1992) was performed using the $\bar{D}_{sh}^2$ as a measure of distance between haplotype pairs. Genetic divergence between regions, among populations within regions and within populations were inferred from Φ_{CT}, Φ_{SC} and Φ_{ST} values, respectively. Significance of the obtained values was inferred by bootstrap analysis (Excoffier et al. 1992).

Pairwise haplotypic difference between individuals belonging to the same population was calculated as the number of mutational steps intervening between their haplotypes

$$X_{ij(i<j)} = \sum_{k=1}^{L} |a_{ik} - a_{jk}|$$

where a_{ik} and a_{jk} are the allele size at the k-th locus of haplotypes carried by the i-th and j-th individuals ($i < j$), and L is the total number of loci compared. The number of mutations that occur is a random variable drawn from a Poisson distribution (Slatkin and Hudson 1991; Slatkin 1995, eq. 6). Frequency distribution of the observed within-population x_{ij} values was calculated independently for each population. Distribution of within-population pairwise x_{ij} was tested for deviation from the expected distribution by the Kolmogorov-Smirnov test ($\alpha = 0.05$).

Ordination of the populations was carried out by multivariate analysis. Principal Component Analysis (PCA) was performed both on transformed haplotype frequency ($x_i = \arcsin(\sqrt{p_i})$, where p_i is the haplotypic frequency of the i-th population) and on a coancestry coefficient distance matrix [$D = -\ln(1 - \Phi_{ST})$; Excoffier et al. 1992].

Results

Seven (Pt9379, Pt15169, Pt30204, Pt36480, Pt41093, Pt71936, Pt87268) out of the nine chloroplast microsatellites analysed were polymorphic, giving a total of 24 different variants. The 24 variants were combined in 34 different haplotypes, the frequency distributions of which are reported in Table 1. Only 1 haplotype was common to all populations, while 25 of them were found in single populations ("population-specific haplotypes").

The populations which generally showed the lowest value of within-population diversity are those located in Portugal, as revealed by the estimated parameters that are based on haplotypic frequencies (effective number of haplotypes, n_e, and unbiased haplotypic diversity, H_e) as well as by the mean D_{sh}^2 and S_w (within-population variance of allele size) (Table 1 and Fig. 3a). In particular, the mean pairwise distance within a population (mean D_{sh}^2) for the Pantelleria population is about 15–20 times greater than those of the Moncao and Alcacier populations.

The distribution of the pairwise differences among individuals within each population is reported in Fig. 1. With the exceptions of the Landes and Pantelleria populations, which show a somewhat uneven, bimodal distribution, all the other populations have a skewed distribution (in particular the Portuguese ones). The observed distribution did not show significant deviation from a Poisson expectation, except for the Landes and Pantelleria populations (KS d = 0.1484, $P < 0.01$; KS d = 0.1749, $P < 0.01$, respectively).

The proportion of genetic differentiation among populations, estimated using R_{ST}, was 0.235, thus indicating that more than 23% of the haplotypic variation was due to differences among populations. The high level of haplotypic divergence among populations was confirmed by the analysis of the molecular variance (AMOVA) (Table 3). The Φ_{ST} value, which indicates the degree of correlation of the molecular diversity of random haplotypes within populations relative to that of random pairs of haplotypes from the whole species, was 0.254 and highly significant ($P < 0.001$) (Table 3a). When the populations were arbitrarily grouped into two regions, one including the Eastern populations and the second the Western ones, no statistically significant divergence was found among regions. Grouping of the populations on the basis of the results obtained performing the PCA analysis (Fig. 2a,b) on the haplotype frequencies and in particular on the pairwise D_{sh}^2 estimated between haplotypes revealed statistically significant differences among the following four groups ($\Phi_{CT} = 29.9$, $P < 0.001$, Table 3b): (1) Landes, Liguria, Tuscany; (2) Alcacier, Leiria, Moncao, Morocco, Sardinia; (3) Pantelleria; (4) Corsica (Fig. 3b). The

Table 2 Haplotype frequency (sorted by overall abundance) in ten populations of *Pinus pinaster*. In the first column the allele size at each of the nine cpSSR loci studied is displayed (Pt110048, Pt26081, Pt71936, Pt9379, Pt30204, Pt87268, Pt41093, Pt15169, Pt36480, respectively). In the last column the overall frequency of haplotypes is reported. The last row shows the number of different haplotypes observed for each population

Haplotype	Label	Mor	Lan	Sar	Tus	Lig	Pan	Cor	Alc	Mon	Lei	All
83/107/143/90/144/163/77/115/145	A	0.4167	0.2083	0.375	0.25	0.3333	0.4167	0.1667	0.6667	0.7083	0.4583	0.400
83/107/143/90/144/163/77/114/145	B	0.0833	0.0833	0.0417	0.375	0.4167	0	0	0	0.0833	0	0.108
83/107/143/90/145/163/77/115/145	C	0.0417	0.375	0	0.1667	0.125	0	0.2083	0	0.0417	0	0.096
83/107/143/90/143/163/77/115/145	D	0.1667	0.2083	0.125	0	0	0	0.0417	0.0833	0	0	0.063
83/107/143/90/144/162/77/115/145	E	0.125	0	0.125	0	0	0	0	0.125	0.0417	0.0417	0.046
83/107/143/90/142/163/77/115/145	F	0	0	0	0	0	0	0.3333	0	0	0	0.033
83/107/143/90/144/163/77/115/146	G	0	0	0	0.125	0	0	0	0.0833	0.0417	0.0833	0.033
83/107/143/90/144/163/77/116/145	H	0	0	0	0	0	0	0	0.0417	0.0833	0.125	0.025
83/107/145/90/142/162/77/115/145	I	0	0	0	0	0	0.2083	0	0	0	0	0.021
83/107/144/90/144/163/77/115/145	J	0	0	0.0833	0	0	0.0417	0	0	0	0.0417	0.017
83/107/143/90/144/163/77/112/145	K	0	0	0.125	0	0	0	0	0	0	0	0.013
83/107/144/90/143/162/77/115/145	L	0	0	0	0	0	0	0	0	0	0.125	0.013
83/107/143/90/143/162/77/115/145	M	0	0	0.0417	0	0	0	0	0	0	0.0417	0.008
83/107/143/90/142/163/77/114/145	N	0	0	0	0	0	0	0.0833	0	0	0	0.008
83/107/143/90/143/163/77/114/145	O	0	0	0	0	0	0	0.0833	0	0	0	0.008
83/107/143/90/144/162/77/116/145	P	0	0	0	0	0	0.0833	0	0	0	0	0.008
83/107/143/90/144/163/78/115/145	Q	0	0	0	0.0833	0	0	0	0	0	0	0.008
83/107/143/90/145/163/77/116/145	R	0	0	0	0	0	0	0	0	0	0.0833	0.008
83/107/143/90/146/163/77/115/145	S	0	0	0	0	0.0833	0	0	0	0	0	0.008
83/107/144/90/143/163/77/114/145	T	0	0	0	0	0	0.0833	0	0	0	0	0.008
83/107/144/90/143/163/77/115/145	U	0.0833	0	0	0	0	0	0	0	0	0	0.008
83/107/144/90/145/163/77/115/140	V	0	0.0833	0	0	0	0	0	0	0	0	0.008
83/107/143/89/144/163/77/114/145	W	0	0.0417	0	0	0	0	0	0	0	0	0.004
83/107/143/90/141/163/77/114/145	X	0	0	0	0	0	0	0.0417	0	0	0	0.004
83/107/143/90/141/163/77/115/145	Y	0	0	0	0	0	0	0.0417	0	0	0	0.004
83/107/143/90/144/163/77/115/144	Z	0	0	0.0417	0	0	0	0	0	0	0	0.004
83/107/143/90/144/164/77/115/145	AA	0	0	0.0417	0	0	0	0	0	0	0	0.004
83/107/143/90/144/164/77/116/145	AB	0	0	0	0	0.0417	0	0	0	0	0	0.004
83/107/143/90/145/162/77/115/145	AC	0.0417	0	0	0	0	0	0	0	0	0	0.004
83/107/143/90/145/162/77/115/146	AD	0.0417	0	0	0	0	0	0	0	0	0	0.004
83/107/144/90/142/162/77/115/145	AE	0	0	0	0	0	0.0417	0	0	0	0	0.004
83/107/144/90/142/162/77/116/146	AF	0	0	0	0	0	0.0417	0	0	0	0	0.004
83/107/145/90/142/162/77/114/145	AG	0	0	0	0	0	0.0417	0	0	0	0	0.004
83/107/145/90/142/162/77/116/145	AH	0	0	0	0	0	0.0417	0	0	0	0	0.004
Count		8	6	9	5	5	9	8	5	6	8	34

populations from Portugal and Morocco tend to group together while the Corsica and Pantelleria populations are clearly differentiated. Moreover Landes, Liguria and Tuscany form a distinct cluster. Significant differences ($\Phi_{CT} = 10.6$, $P < 0.05$) were also detected when the overall variance was partitioned on the basis of the population grouping proposed by Baradat and Marpeau (1988) and Bahrman et al. (1994) (see Fig. 3b and Table 3c).

The distribution of the populations is graphically displayed in the plane defined by the first two principal components (Fig. 2a, b), explaining more than 68% (haplotype frequencies) and 95% (pairwise D^2_{sh}) of the total variation, respectively.

Discussion

In this investigation the genetic differentiation of populations of *Pinus pinaster* was found to be higher than those previously reported by other authors using allozyme markers (Petit et al. 1995; Torres de Castro 1989; Nemoz 1997), though in the last two investigations only populations from a more restricted geographical areas were analysed. The level of within-population haplotypic diversity in maritime pine (mean $D^2_{sh} = 0.512$) was higher than in *Pinus resinosa* (mean $D^2_{sh} = 0.154$, Echt et al. 1998) and lower than in *Pinus halepensis* (mean $D^2_{sh} = 3.58$, Morgante et al. 1997). This lower level of haplotypic diversity may be hypothesized to be due to the fragmentation of the distribution range. Species with a more disjunct and scattered natural range, such as *P. resinosa* and *P. pinaster*, are expected to show lower within-population genetic diversity due to genetic drift and limited gene flow among populations. Our data seem to confirm this trend. Haploid, uniparentally inherited DNA markers are considered to be more efficient in detecting the effects of genetic drift because the number of effective markers is half that of diploid, biparentally inherited markers (Mitton 1993).

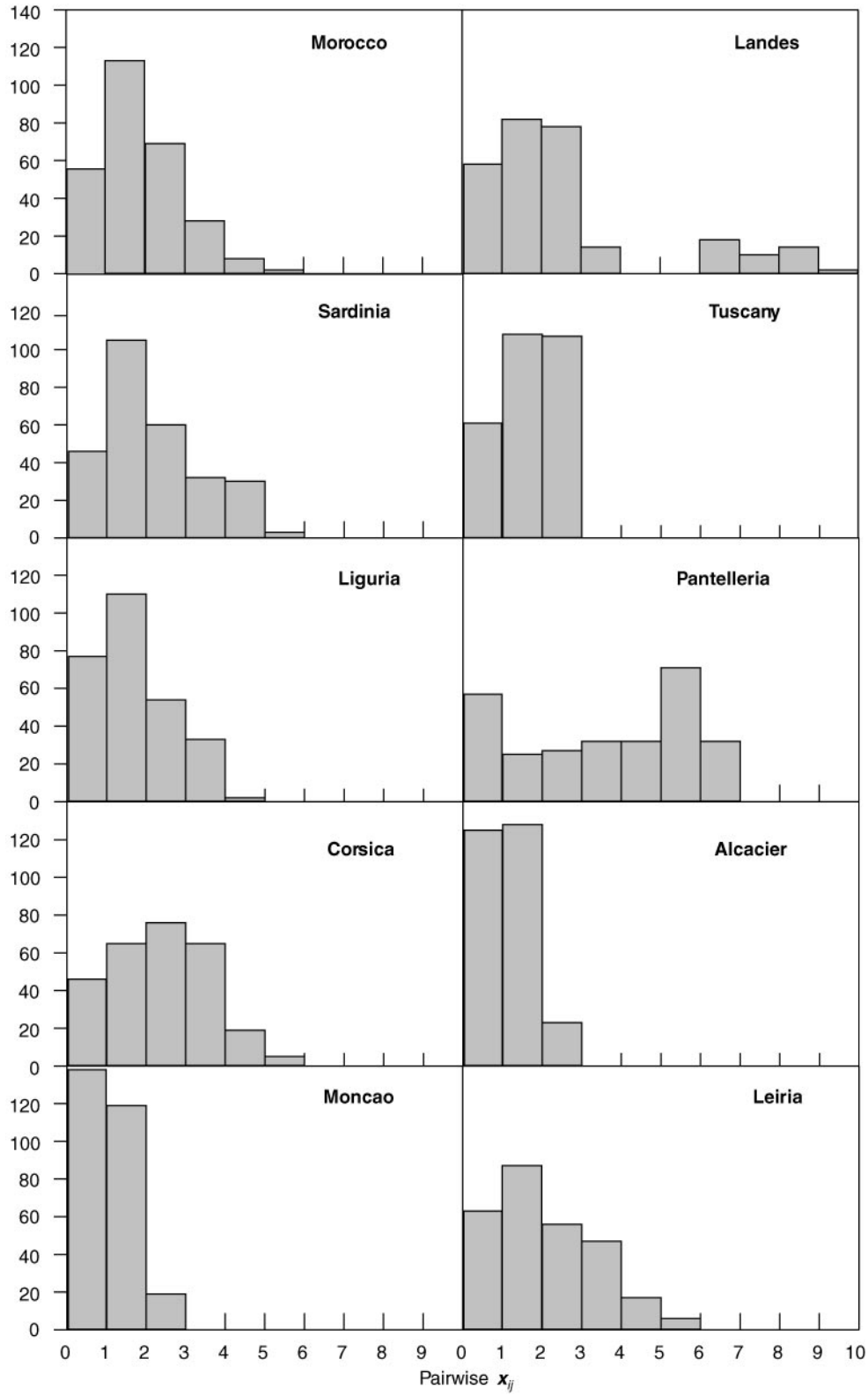

Fig. 1 Distribution of the observed pairwise haplotypic differences (x_{ij}) between individuals within the ten populations analysed. Observed distributions did not show significant differences from a Poisson expectation, except for the Landes and Pantelleria stands (Kolmogorov-Smirnov test; $P < 0.05$)

Parcelling of the natural range of *Pinus pinaster* after the Wurm III glaciation may have determined the strong differentiation observed in this species. This hypothesis was proposed by Baradat and Marpeau (1988) and partially confirmed by Bahrman et al. (1994), based on the analysis of total proteins. In particular, three distinct groups of maritime pine populations have been distinguished on the basis of palynological and palaeoclimatological (Baradat and Marpeau 1988, and references therein), terpene (Baradat and Marpeau 1988)

Table 3 Hierarchical analysis of molecular variance (AMOVA) based on D_{sh}^2 haplotype distance. (a) Partitioning of the variance assuming no hierarchical structure. (b) Partitioning of the molecular variance assuming a population structure as displayed in Fig. 3a (four groups). (c) Partitioning of the molecular variance assuming a population structure as displayed in Fig. 3b (three groups; Baradat and Marpeau 1988; Bahrman et al. 1994). (*SS* Sum of squares · *MS* mean squares)

Variance component	*df*	SS	MS	Variance	%Total	Φ-Stats	*P*
(a)							
Among populations	9	21.105	2.345	0.0870	25.41	0.254	< 0.001
Within populations	230	58.792	0.256	0.2556	74.59		
Total	239	79.898					
(b)							
Among groups	3	18.532	6.177	0.1122	29.9	0.299	< 0.001
Among populations within groups	6	2.573	0.429	0.0072	1.9	0.027	0.017
Within populations	230	58.792	0.256	0.2556	68.2	0.319	< 0.001
Total	239	79.898					
(c)							
Among groups	2	9.188	4.594	0.0376	10.65	0.106	0.027
Among populations within groups	7	11.917	1.702	0.0602	17.05	0.191	< 0.001
Within populations	230	58.792	0.256	0.2556	72.30	0.277	< 0.001
Total	239	79.898					

and total protein composition data (Bahrman et al. 1994) (Fig. 3b): (1) the Atlantic group, which includes populations from Portugal, Spain and France; (2) the Mediterranean group which consists of populations from Italy, Sardinia and Corsica and (3) the North African group which comprises also the Pantelleria population. The chloroplast microsatellite data partly confirm the proposed subdivision, considering that statistically significant differences among the three groups were observed performing the AMOVA analysis. However, a more complex picture seems to derive from our data. On the basis of the PCA analysis performed considering the D_{sh}^2 between haplotypes, two main groups can be distinguished, one including North African, the Portuguese and Sardinia populations, the second including the others, and two minor groups comprising the Corsica and Pantelleria populations, respectively. The AMOVA analysis revealed highly significant differences among the above groups ($P < 0.001$). Moreover, the French population is in an intermediate position between the Italian populations (Liguria and Tuscany) and Portuguese, Sardinian and Moroccan population groups. The movement of plant material by human activities in the past may partially explain the above picture as well as local gene flow among geographically close populations. In fact, the distribution of the variation at the level of the chloroplast genome, which is paternally inherited in conifers (Cato and Richardson 1996; Vendramin and Ziegenhagen 1997), is strictly related to the *via* pollen gene flow that might have reduced the effects of genetic drift to some extent.

Baradat and Marpeau (1988) hypothesized the presence of a centre of origin of *Pinus pinaster* during the last glaciation that was located in the South of Portugal, where the most ancient fossil findings were discovered (dated about 3 million years, Pliocene) (Baradat and Marpeau 1988), and of two different pathways of recolonization, before and after the Wurm glaciation, respectively. The pre-glacial recolonization hypothesis presupposes the presence of three distinct routes – towards the North of Portugal, Spain and France, towards the South of Spain, France and Italy and towards the North of Africa. The post-glacial hypothesis assumes that the recolonization occurred only along the first two routes. If the scenario proposed by Baradat and Marpeau (1988) holds, one should expect a decrease in genetic variability eastwards, with the populations located in the centre of origin of the species (South Portugal) characterized by a high level of genetic diversity. Contrarily low levels of within-population diversity were observed in the Portuguese populations: a partial exception is represented by the Leiria stand, which is however considered to be of non-natural origin (Baradat and Marpeau 1988). Similar results were obtained by Nemoz (1997) who observed a generally higher level of allozyme diversity in the Spanish populations than in the Leiria population and hypothesised a migration pathway from Spain to Portugal.

The Landes population showed a high value of within-population diversity, higher than those observed in all the populations located along the hypothesized migration route (Baradat and Marpeau 1988). This evidence seems to reject the hypothesis of an origin of the French population from the refugia located in Portugal. One evident and surprising result is represented by the high level of variability detected in the small and completely isolated population of Pantelleria. This result appears more relevant if we consider that in this area fire represents an important ecological factor which may determine population bottlenecks and genetic drift.

Based on our data a more complex picture of the possible recolonization pathways can be drawn. Two main reservoirs of haplotypic diversity (Landes and

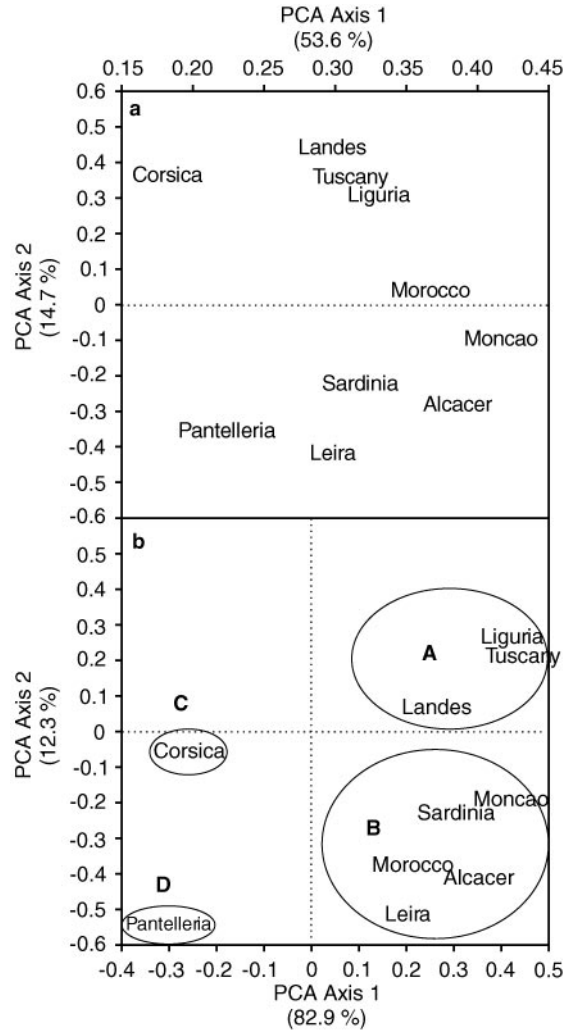

Fig. 2 **a** Ordination of the ten populations of *Pinus pinaster* based on haplotype frequency. Data was transformed as $x_i = \arcsin(\sqrt{p_i})$, where p_i is the observed haplotype frequency for the i-th population. The proportion of variance accounted for by the first two axes was 68.3%. **b** Ordination of the same populations based on the coancestry distance matrix $[D = -\ln(1 - \Phi_{ST})]$ obtained from AMOVA analysis (Excoffier et al. 1992), carried out on pairwise D^2_{sh} between haplotypes (stepwise haplotype approach – see Materials and methods). The proportion of variance explained was 95.2%. *Block capital letters* refer to the four putative genetic groups identified (see also Fig. 3a)

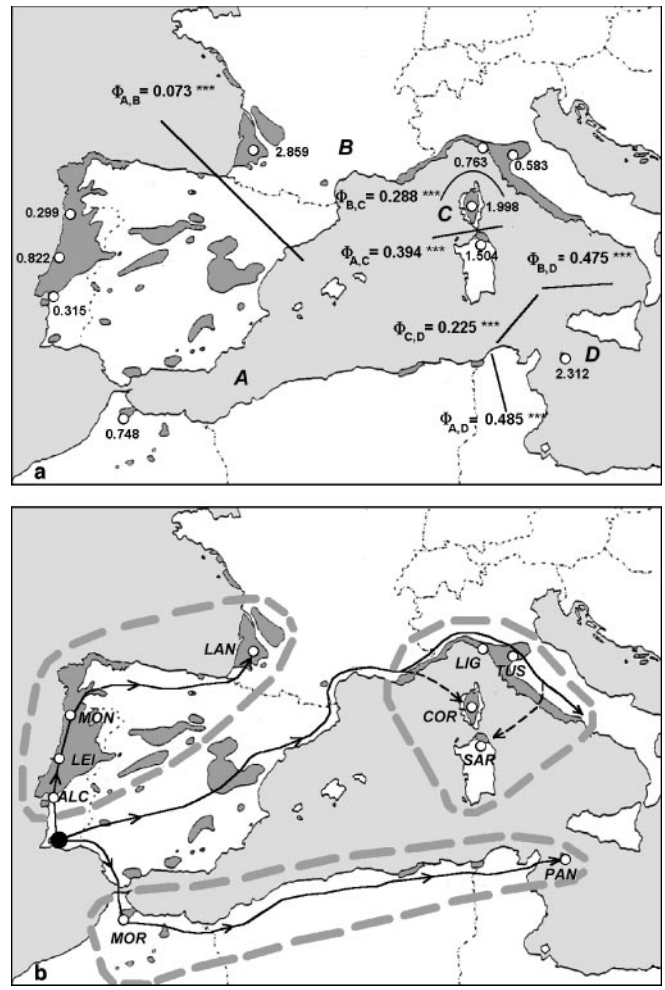

Fig. 3a, b Natural range of *Pinus pinaster*, population structure and hypothetical post-glacial recolonization pathways. **a** Unbiased, within-population variance in allele size (S_w – Slatkin 1995) is shown for each population; *black lines* represent significant discontinuity between putative nuclei. For each discontinuity, Φ_{CT} values and their significance are reported. *Capital letters* refer to the four groups of populations detected by multivariate analysis (see Fig. 2). **b** Grouping of populations based on terpene and allozymic polymorphisms and proposed post-glacial recolonization pathways of *Pinus pinaster* (redrawn from Baradat and Marpeau 1988 and Bahrman et al. 1994). Population labels are the same as in Table 1

Pantelleria) may be identified. The Pantelleria population (or populations which were located in the north of Africa) might have represented a starting point of the recolonization process. The presence of a putative refugia in the Maghreb region was also suggested by Baradat and Marpeau (1988). Pantelleria, therefore, might represent an ancient population that originated in the pre-glaciation period from the refugia located in the central part of North Africa, from which the recolonization might have started both west- and eastwards. A possible confirmation of this hypothesis may reside in the fact that populations from Portugal, and in particular Alcacier and Moncao, show a significant skewed distribution of the pairwise differences between individuals within populations, which is indicative of recently established gene pools (founder effects). On the contrary, the Pantelleria population exhibits a bimodal distribution, which suggests an ancient gene pool and constant population size through time (Slatkin and Hudson 1991). Analogous considerations can be made for the Landes population: the French nucleus might have represented a refugia from which the recolonization of the Italian part of the natural range of this species began. However, the hypothesis that these populations are 'melting spots' (mixtures) rather than 'hot spots' of haplotypic diversity cannot be excluded.

Our data does not support the results of Bahrman et al. (1994), who found the lowest level of variability in the Sardinian stand and hypothesised that it might have originated from an Italian population and subsequent differentiation by genetic drift. The high values of haplotypic diversity of this population seem to exclude this conclusion. The clustering of the Sardinia stand with the Portuguese and North African populations may be explained assuming a pre-glaciation colonization of the island from the south, from Africa: moreover, the possibility of a non-natural origin of this population cannot be completely ruled out, even though the clear genetic differentiation seems to exclude the Corsica as the possible region of origin. Puzzling evidence about the history of the Corsica population was obtained. Its clear differentiation from all the other populations, including Liguria from which it supposedly originated (Bahrman et al. 1994) and the relatively high value of genetic diversity, seems to indicate an independent origin even if no definitive conclusions can be drawn on the basis of our results. The anthropic influence might also have played an important role in determining the actual distribution of this species considering that for a long period the maritime pine has been extensively cultivated and its wood used for the construction of ships (Corsica and Sardinia belonged to the Carthaginian empire).

As emphasized by Baradat and Marpeau (1988) the actual palaeological information about Maritime pine is very limited and too incomplete to clearly reconstruct the possible recolonization routes: in this context genetic information obtained in this study seems to shed light on the possible history of this species.

Acknowledgements This work has been financed by the European Union [INCO project: "Global, physiological and molecular responses to climatic stresses of three mediterranean conifers (pines of the *halepensis – brutia* section, maritime pine and mediterranean cedars). Consequences for optimal use, conservation and sustainable improvement of their genetic resources" Contract n. ERBIC 18CT 970200.]. Experiments carried out for this work comply with the current laws of the country in which the experiments were performed.

References

Avise JC (1994) Molecular markers, natural history and evolution. Chapman and Hall, New York

Baradat Ph, Marpeau A (1988) Le pin maritime, *Pinus pinaster* Ait. Biologie and génétique des terpènes pour la connaissance et l'améloriation de l'espèce. PhD thesis, University of Bordeaux I, Bordeaux

Bahrman N, Zivy M, Damerval C, Baradat Ph (1994) Organisation of the variability of abundant proteins in seven geographical origins of maritime pine (*Pinus pinaster* Ait.). Theor Appl Genet 88 : 407–411

Boscherini G, Morgante M, Rossi P, Vendramin GG (1994) Allozyme and chloroplast DNA variation in Italian and Greek populations of *Pinus leucodermis*. Heredity 73 : 284–290

Cato SA, Richardson TE (1996) Inter- and intraspecific polymorphism at chloroplast SSR loci and the inheritance of plastids in *Pinus radiata* D. Don. Theor Appl Genet 93 : 587–592

Di Rienzo A, Peterson AC, Garza JC, Valdes AM, Slatkin M, Freimer NB (1994) Mutational processes of simple-sequence repeat loci in human populations. Proc Natl Acad Sci USA 91 : 3166–3170

Doyle JJ, Doyle JL (1990) Isolation of plant DNA from fresh tissue. Focus 12 : 13–15

Echt CS, De Verno LL, Anzidei M, Vendramin GG (1998) Chloroplast microsatellites reveal population genetic diversity in red pine, *Pinus resinosa* Ait. Mol Ecol 7 : 307–317

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

Feldman MW, Bergman A, Pollock DD, Goldstein DB (1997) Microsatellite genetic distances with range constraints: analytic description and problems of estimation. Genetics 145 : 207–216

Goldstein DB, Ruiz Linares A, Cavalli-Sforza LL, Feldman MW (1995) An evaluation of genetic distances for use with microsatellite loci. Genetics 139 : 463–471

Jarne P, Lagoda JL (1996) Microsatellites, from molecules to population and back. Tree 10 : 424–429

Mitton JB (1993) Molecular approaches to population biology. Ann Rev Ecol Syst 25 : 45–69

Morgante M, Felice N, Vendramin GG (1997) Analysis of hypervariable chloroplast microsatellites in *Pinus halepensis* reveals a dramatic genetic bottleneck. In: Karp A, Isaac PG, Ingram DS (eds) Molecular tools for screening biodiversity: plants and animals. Chapman and Hall, London, pp 407–412

Nei M (1987) Molecular evolutionary genetics. Columbia University Press, New York

Nemoz LS (1997) Variabilidad genetica de *Pinus pinaster* Aiton en Espana: analysis mediante marcadores proteicos e isoenzimaticos. PhD thesis, Universidad Politecnica de Madrid, Spain

Petit RJ, Bahrman N, Baradat Ph (1995) Comparison of genetic differentiation in maritime pine (*Pinus pinaster* Ait.) estimated using isozyme, total protein and terpenic loci. Heredity 75 : 382–389

Powell W, Morgante M, McDevitt R, Vendramin GG, Rafalski JA (1995) Polymorphic simple sequence repeat regions in chloroplast genomes: application to the population genetics of pines. Proc Natl Acad Sci USA 92 : 7759–7763

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

Slatkin M, Hudson RR (1991) Pairwise comparisons of mitochondrial DNA sequences in stable and exponentially growing populations. Genetics 129 : 555–562

Torres de Castro LF (1989) Isoenzimas do *Pinus pinaster* Ait. numa perspectiva de aplicacao ao melhoramento genetico da especie. PhD thesis Universidade de Tras-os-Montes e Alto Douro, Vila Real, Portugal

Valdes AM, Slatkin M, Freiner NB (1993) Allele frequencies at microsatellite loci: the stepwise mutation model revisited. Genetics 133 : 737–749

Vendramin GG, Ziegenhagen B (1997) Characterisation and inheritance of polymorphic plastid microsatellites in *Abies*. Genome 40 : 857–864

Vendramin GG, Lelli L, Rossi P, Morgante M (1996) A set of primers for the amplification of 20 chloroplast microsatellites in *Pinaceae*. Mol Ecol 5 : 585–598

Watano Y, Imazu M, Shimizu T (1996) Spatial distribution of cpDNA and mtDNA haplotypes in a hybrid zone between *Pinus pumila* and *P. parviflora* var '*pentaphylla*' (*Pinaceae*). J Plant Res 109 : 403–408

Ziegenhagen B, Kormutak A, Shauerte M, Scholz F (1995) Restriction site polymophism in chloroplast DNA of silver fir (*Abies alba* Mill.). For Genet 2 : 99–107