Theor Appl Genet (1997) 95:1138–1146

G. Bucci · G. G. Vendramin · L. Lelli · F. Vicario

Assessing the genetic divergence of *Pinus leucodermis* Ant. endangered populations: use of molecular markers for conservation purposes

Received: 27 April 1997 / Accepted: 14 June 1997

Abstract In this study, 23 previously identified Mendelian RAPD markers and 16 polymorphic allozymic markers were used to assess divergence among two Greek populations and five Italian populations of *Pinus leucodermis*. Confidence intervals of observed genetic divergence were obtained using bootstrap analysis. Divergence among Italian populations was found to be about as large as that between Italian and Greek populations. Since it is likely that the split of two nuclei took place more than 10,000 years ago, a larger differentiation between, rather than within, the above nuclei was expected. If genetic drift was responsible for the larger divergence of Italian populations, large randomly generated disequilibrium between alleles at neutral, unlinked loci was expected. Indeed, the proportion of pairs of loci showing a non-random association of alleles within each of the Italian populations was larger than what was expected by pure chance (7.95–10.88%). Effective population size based on randomly generated disequilibrium was quite small for three out of the five populations considered ($N_e = 17.31 \pm 1.88$, 16.57 ± 1.73, and 31.41 ± 7.26, respectively). The implications of the results with respect to the conservation of endangered species of trees are discussed.

Key words Allozymes · RAPDs · Genetic differentiation · Linkage disequilibrium · Population genetics

Communicated by P. M. A. Tigerstedt

G. Bucci · G. G. Vendramin (✉) · L. Lelli · F. Vicario
Istituto Miglioramento Genetico Piante Forestali,
CNR Via Atto Vannucci 13, 50134 Florence, Italy
Fax: + 39 55 468804
E-mail: vendramin@imgpf.fi.cnr.it

Introduction

Many forest tree species are characterized by prevalent allogamous mating and continuous natural distribution over large areas (Müller-Stark et al. 1992). In such a situation, genetic divergence among populations by random genetic drift is expected to be balanced by high levels of gene flow (Wright 1965). Allozyme studies on natural populations of forest trees have often reported low levels of genetic divergence among populations, accounting on average for only 5% of the total genetic variation (Godt and Hamrick 1990). On the other hand, species with extremely small and disjunct geographic ranges are expected to diverge by a larger extent due to a lowered interpopulation gene flow and a larger influence of drift (Lesica and Allendorf 1994). Small, peripheral populations geographically or ecologically isolated from the main range are expected to show increased probability of losing genetic variation (Gilpin 1991; Young et al. 1996). Genetically depleted populations are less likely to face environmental stress by adaptation to changing environments and may be more prone to local extinction (Ellstrand 1992).

Loss of genetic variation has been reported to accompany a reduction in population size (Van Treuren et al. 1991; Raijmann et al. 1994; Prober and Brown 1994). Based on theoretical considerations, the rate of loss is inversely proportional to effective population size N_e (Nunney and Elam 1994). Therefore, the conservation of endangered species depends on maintaining a substantial population size (Carson 1990). In wild populations of forest trees, long life-cycles and overlapping generations make it unfeasible to directly estimate the loss of genetic variation through time, and an estimation of population size is often impractical.

Among forest tree species, puzzling evidence about the relationship between isolation, divergence and genetic variation among populations has been reported for *Pinus resinosa* (Mosseler et al. 1991; Simon et al.

1986) and *Pinus leucodermis* (Müller-Stark et al. 1992; Boscherini et al. 1994a), given their small discontinuous natural range and the low (or null) levels of genetic differentiation among populations detected by isozymes.

Pinus leucodermis is an endangered species with its main range in the Balkan peninsula and a smaller range in Southern Italy (Avolio 1984). The Italian range is restricted to four small and scattered nuclei established in harsh, mountainous areas partly situated within the Pollino National park (CS and PZ provinces) that cover 5,678 ha overall. Several genetic studies have been already carried out on *P. leucodermis* using allozymes, with the aim of estimating mating system parameters (Morgante et al. 1991, 1994) and genetic relationships among Greek and Italian populations (Boscherini et al. 1994a). Powell and colleagues (1996), using paternally inherited chloroplast simple-sequence-repeat (SSR) variants, reported that Italian populations are less variable than the Greek populations, with divergence among populations accounting for 22% of the total variation.

In the study reported here, 23 Mendelian random amplified polymorphic DNA (RAPD) markers (Boscherini et al. 1994b) and 16 allozyme markers (Morgante et al. 1994) were used to assess genetic divergence among seven populations of *Pinus leucodermis*. Linkage disequilibrium between marker pairs was estimated within populations, and effective population size for Italian stands tentatively estimated. Implications of the results with respect to the conservation of endangered species of trees are discussed.

Material and methods

Plant material and analysis procedure

Seeds were collected from five Italian populations (*La Spina*: Monte La Spina – 900 m a.s.l.; *Pollino*: Cima Pollino – 2,000 m; *Pollino A*: Valletorta – 900 m; *Pollino B*: Val Gaudolino – 1,400 m; *Pollino C*: Pollinello – 1,800 m) and two Greek populations (Olimpo – 1,500 m a.s.l.; Pindo – elevation not available). As for the Italian populations, seedlots were collected from each of 17–29 trees sampled and kept apart for further analysis, while for the Greek populations only bulked seedlots from more than 30 trees were available.

DNA extraction from megagametophytes of individual seeds was carried out as described in Doyle and Doyle (1987). DNA concentration of the samples was estimated using a fluorometric assay (Cesarone et al. 1979). Polymerase chain reaction (PCR) amplifications were carried out as previously described (Boscherini et al. 1994b) using random decamers OPG-02, OPG-06 and OPG-08 from Operon (Operon Technologies, Alameda, Calif.). Primers were selected on the basis of the largest number of Mendelian markers obtainable: using the above decamers, 7, 7 and 9 Mendelian markers, respectively, were identified (Boscherini et al. 1994b; G. G. Vendramin, unpublished data).

Allozyme analysis was carried out as previously described (Morgante et al. 1991). A total of 16 polymorphic markers encoded by 11 enzyme systems were used. In summary, 39 loci (23 RAPD and 16 allozymes) were analyzed in this study.

Data collection and analysis

Due to the limited amount to tissue available, two different sets of megagametophytes were analyzed using allozyme and RAPD markers. Therefore, cosegregation analysis between both kinds of markers could not be carried out.

For the Italian populations, genotypes at each locus (RAPD or allozyme markers) were inferred from the analysis of six megagametophytes per tree (*misclassifying probability* $\pi = 0.5^{k-1}$, where k is the number of megagametophytes per tree; $k = 6$, $\pi = 0.03125$). The least common alleles at the allozyme loci were pooled, thereby obtaining two allelic classes for every multiallelic locus. Allelic frequencies were calculated from genotypic data. For the Greek populations, allelic frequencies were computed from haplotypic data, assuming random sampling of gametes within the gene pool. Estimations of the population parameters by allozyme markers have been reported elsewhere (Boscherini et al. 1994a).

Principal Component Analysis (PCA) was carried out on trans formed genetic data $x_i = arcsin\sqrt{p_i}$, where p_i is the allelic frequency of the i-th population. Estimation of F_{ST} values was obtained by calculating the Wahlund variance (Wright 1951):

$$F_{ST_i} = \frac{V_{pi}}{\bar{p}_i \cdot (1 - \bar{p}_i)}$$

where V_{pi} is the variance among allele frequencies of the populations analyzed at the i-th locus, and $\bar{p}_i$ is their average allele frequency weighted on sample size. The significance of the observed population divergence over all loci was verified by the chi-square test (Weir 1996). Pairwise F_{ST} between populations were computed as the arithmetic average of F_{ST} at single loci. Pairwise genetic distances between populations were estimated following Nei (1973).

Numerical resampling of marker frequencies

In order to obtain confidence intervals for F_{ST} estimates, we carried out bootstrap analyses (Efron 1982). Null distribution of F_{ST} by bootstrapping loci were obtained, preserving the original structure of the dataset. Allelic frequencies were resampled with replacement, generating 1,000 new bootstrapped matrices, and F_{ST} values estimated from each resampled matrix as described above. The goodness-of-fit of the bootstrapped F_{ST} values to normal distribution was verified by the Kolmogorov-Smirnov test ($\alpha = 0.05$), and a 95% confidence interval was calculated from the middle 95% of the bootstrapped estimates (Weir 1996). In order to test differences in the degree of divergence among populations, we compared bootstrapped F_{ST} distributions by the t-test.

Numerical resampling procedures were carried out using an ad-hoc written computer programme in AWK language (GNU Corp) running a Sun SparcStation 10/20 (UNIX System). Source codes are available upon request to the authors. Results of the above analyses were verified using the GDA Statistical Package (courtesy of P. Lewis) and GENEPOP Statistical Package (Rousset and Raymond 1995).

Neutrality test of markers

Test of neutrality of the markers was carried out following Lewontin and Krakauer (1979). The parameter $(n-1)\hat{F}/\bar{F}$ was estimated for each marker, where $\hat{F}$ is the F_{ST} value estimated by each marker, $\bar{F}$ is the mean F_{ST} value over all loci, and n is the number of populations considered. The parameter above has a chi-square distribution with $n-1$ *df*. Goodness-of-fit of the observed distribution to expectation was verified by the Kolmogorov–Smirnov test. A fit to the expected distribution was also verified independently for RAPD and allozyme data.

Linkage disequilibrium analysis

Linkage disequilibrium analysis was carried out separately on Italian (diploid data) and Greek populations (haploid data). As for the latter, gametic disequilibrium was estimated using the following formula (Nei 1987):

$$D_{ij} = \frac{(n_1 \cdot n_4) - (n_2 \cdot n_3)}{n^2}$$

where n_1, n_2, n_3 and n_4 are the number of observed gametes '1/1', '1/0', '0/1' and '0/0' at i-th and j-th loci, respectively, and n is the total number of gametes analyzed. Variances and chi-square test for D_{ij} were computed following Weir (1996).

Gametic disequilibria were then tested for homogeneity (χ^2_H, with $1df$) and zero disequilibrium among all the populations [χ^2_D, with $(m-1)df$, where m is the number of populations analyzed] using the formula reported by Weir (1996).

For the Italian populations, pairwise composite genotypic disequilibria and the relative variance were estimated on genotypic data using the method proposed by Burrows (Cockerham and Weir 1977).

$$\Delta_{ij} = \frac{4n_1 + 2n_2 + 2n_3 + n_4}{2n} - 2p_i \cdot p_j$$

where n_1, n_2, n_3 and n_4 are the number of diploid genotypes showing the di-locus genotye '11/11', '11/10', '10/11', '10/10', respectively, n is the total number of individuals analyzed and p_i, p_j are the allelic frequencies at the two loci considered. This measure of non-random association of alleles at different loci is inclusive of within- and between-gametic disequilibria and therefore does not require distinguishing between the two types of double heterozygotes.

Composite genotypic disequilibria were also tested for homogeneity (χ^2_H, $4df$, see above) and zero disequilibrium among all the populations (χ^2_Δ, $1df$) as described above. Marker pairs showing significant, homogeneous genotypic disequilibrium among the Italian populations were then discarded from further analysis.

Correlation of alleles between locus pairs were calculated using the following formula:

$$r_{ij} = \frac{\Delta_{ij}}{\sqrt{[p_i(1-p_i) + D_i] \cdot [p_j(1-p_j) + D_j]}}$$

where Δ_{ij} is the composite genotypic disequilibrium between i-th and j-th alleles, D_i and D_j are the single-locus disequilibria at i and j, respectively, p_i and p_j are the marker frequencies at the loci considered (Weir 1996).

Estimation of effective population size

Estimates of effective population size (N_e) were obtained considering only unlinked, neutral loci using the least square estimator reported by Laurie-Ahlberg and Weir (1979):

$$N_e = \frac{m}{3 \cdot \sum_i \sum_j r^2_{ij} - (1/n_{ij})}$$

where m is number of pairs of loci considered, r^2_{ij} is the correlation of alleles between i-th and j-th loci and n_{ij} is the sample size. Confidence intervals for r^2_{ij} were computed using the following formula (Bartley et al. 1992):

$$(1-\alpha)CI = \frac{m\bar{r}^2}{\chi^2_{(1-\alpha/2)m}}, \frac{m\bar{r}^2}{\chi^2_{(\alpha/2)m}}$$

where m is the number of pairwise locus comparisons ($= df$), $\bar{r}^2$ is the average squared correlation between alleles and χ^2 is the chi-square value with $(1-\alpha/2)m$ and $(\alpha/2)m$ degrees of freedom. Confidence intervals for r^2_{ij} were then substituted into the above equation to obtain confidence intervals for N_e (Waples 1991).

Results

Population divergence by allozymes and RAPDs

Multivariate analysis (PCA) on allele frequencies underlined an obvious ordination of the populations analyzed. Italian and Greek populations were easily discriminated by the first two PCA axes (36.4% and 22.1% of the total variance explained, respectively). Overall, Italian populations turned out to be widely scattered along PCA axes 1 and 2, while Greek populations were fairly clumped (Fig. 1). No correlations were observed between elevation and the second PCA axis (data not shown).

Due to the heterogeneous dataset (including genotypes and gametes) computations were carried out on allelic frequencies (Weir 1996). The whole genetic dataset (both RAPD and allozyme markers) was considered at first. As expected for marginal populations, significant differences were detected among the Italian stands ($F_{ST} = 0.0755$; $\chi^2_{[1]} = 6.587$; $P < 0.05$), while no differences where found between the Greek populations ($F_{ST} = 0.0406$; $\chi^2_{[1]} = 1.738$; $P > 0.05$). Nineteen out of 23 RAPD markers and 6 out of 16 allozyme markers revealed overall significant differences among populations by Fisher's exact test (data not shown). Significant divergence between the Italian and Greek populations was also observed ($F_{ST} = 0.0740$; $\chi^2_{[1]} = 9.620$; $P < 0.01$) (Table 1).

In order to verify the ability of the two kinds of markers to distinguish among populations, interpopulation

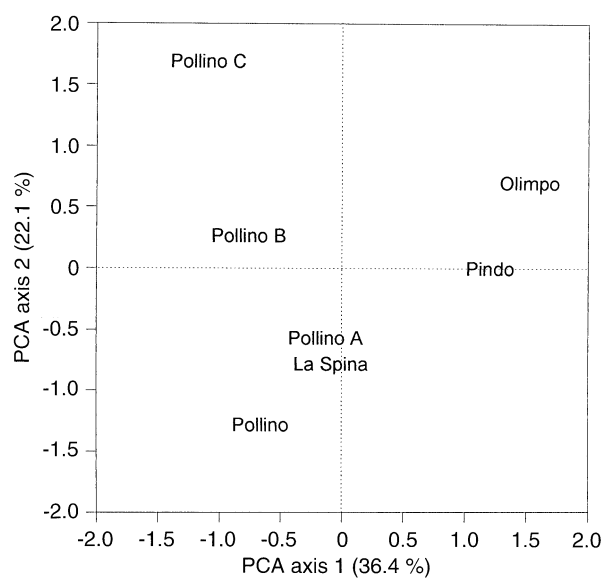

Fig. 1 Ordination of the Italian and Greek populations by principal component analysis of allelic frequencies

Table 1 Numerical resampling of allozyme and RAPD allele frequencies from Italian and Greek populations (*n.s.* non-significant)

	Observed data			Bootstrapped data					
	F_{ST}[a]	χ^{2}[b]	Probability	Valid N[c]	Mean null	SD[d]	SE[e]	C.I.[f]	
								Lower	Higher
RAPD markers									
Overall	0.1792	21.856	$P < 0.001$	1,000	0.1783	0.0246	0.0008	0.1295	0.2237
Between I & G	0.1055	12.872	$P < 0.001$	1,000	0.1047	0.0246	0.0008	0.0528	0.1504
Among Greeks	0.0413	1.768	n.s.	1,000	0.0404	0.0058	0.0002	0.0287	0.0509
Among Italians	0.0987	7.813	$P < 0.01$	1,000	0.0985	0.0097	0.0003	0.0800	0.1192
Allozyme markers									
Overall	0.0671	8.728	$P < 0.01$	1,000	0.0676	0.0114	0.0004	0.0433	0.0874
Between I & G	0.0289	3.752	n.s.	1,000	0.0290	0.0098	0.0003	0.0121	0.0425
Among Greeks	0.0164	0.702	n.s.	1,000	0.0239	0.0106	0.0003	0.0081	0.0487
Among Italians	0.0410	3.577	n.s.	1,000	0.0414	0.0054	0.0002	0.0304	0.0512
All markers									
Overall	0.1341	17.429	$P < 0.001$	1,000	0.1341	0.0161	0.0005	0.1002	0.1655
Between I & G	0.0740	9.620	$P < 0.01$	1,000	0.0740	0.0146	0.0005	0.0440	0.1019
Among Greeks	0.0406	1.738	n.s.	1,000	0.0401	0.0066	0.0002	0.0281	0.0521
Among Italians	0.0755	6.587	$P < 0.05$	1,000	0.0755	0.0071	0.0002	0.0618	0.0894

[a] Observed genetic divergence
[b] Chi-square value calculated following Weir (1996)
[c] Resampling size
[d] Standard deviation of the null distribution
[e] Standard errors of the null distribution
[f] Confidence intervals of the F_{ST} estimates obtained by the 2.5-percentil (lower) and 97.5-percentil (higher) of the null distribution

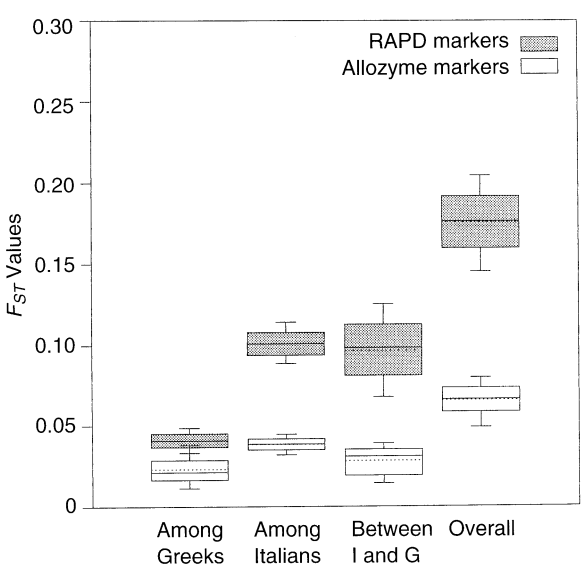

Fig. 2 Bootstrapping analysis of Italian and Greek populations using RAPD and allozyme markers. F_{ST} values were estimated from 1,000 new datasets obtained by resampling with replacement over loci. Mean values (*dotted lines*), median values (*plain lines*), standard deviations (*boxes*) and 95% confidence intervals (*error bars*) of the bootstrapped distributions are reported for both kinds of markers

divergence was estimated independently by allozymes and RAPDs, and the results compared (Fig. 2). The testing of the difference between the two types of markers can be better carried out comparing the distribution of null F_{ST} values obtained by bootstrapping loci instead of single observations (Table 1). In general, RAPD markers highlighted a larger amount of overall divergence between populations than allozymes (0.17483 vs. 0.0652, respectively; $t = 138.3$; $P < 0.001$). The same trend was observed for divergence between the Italian and Greek populations (0.0972 vs. 0.0281, $t = 88.94$; $P < 0.001$), among Italian populations (0.0961 vs. 0.0411, $t = 175.1$; $P < 0.01$) and among Greek populations (0.0426 vs. 0.0173; $t = 48.49$; $P < 0.001$). Divergence between Italian and Greek populations was not significantly different from the divergence observed among Italian populations either using RAPD (0.1055 vs. 0.0987) or allozymes (0.0289 vs. 0.0410, or both (0.074 vs. 0.075), given their overlapping confidence intervals obtained by bootstrapping the dataset (Fig. 2).

A comparison of RAPD and allozyme markers showed a good agreement between pairwise F_{ST} estimates (Fig. 3). The correlation found between F_{ST} calculated on RAPD and allozyme markers was fairly high (Pearson $r = 0.596$; $P < 0.01$). A higher correlation was found when Nei's genetic distances were considered (Pearson $r = 0.792$; $P < 0.001$). The relative power of RAPD markers with respect to allozymes can be assessed by considering the slope of the regressed function between RAPDs and allozymes distances (or divergences). The sensitivity of RAPD markers to population differentiation turned out to be 1/0.124–1/0.129 (7.75–8.03) times greater than that of the allozymes.

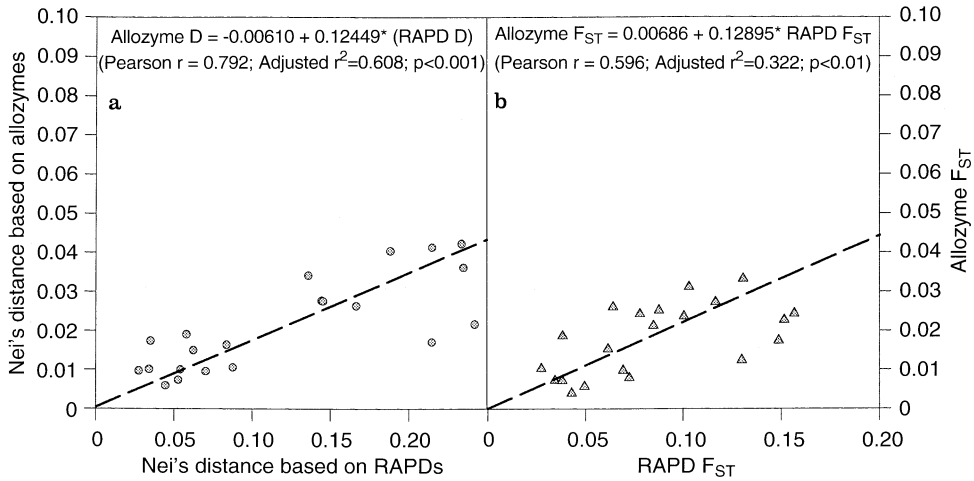

Fig. 3a, b Comparison of **a** pairwise Nei's genetic distances and **b** pairwise genetic divergence among seven populations of *Pinus leucodermis*, based on allozyme (Y axis) and RAPD (X axis) data. The RAPD markers are 1/0.124–1/0.129 (7.75–8.03) times more sensitive to population differentiation than the allozymes

Test of neutrality of genetic markers

Distribution of the parameter $(n-1)\hat{F}/\bar{F}$ did not show significant difference from the expected chi-square distribution ($df = 6$) for RAPD markers (KS d = 0.2113, $P > 0.05$), while a significant difference from expectation was found for allozymes (KS d = 0.5214, $P < 0.01$). Therefore, allozyme markers were discarded from further analysis.

Linkage disequilibrium analysis

Linkage disequilibrium analysis was carried out independently on Italian and Greek populations. Pairwise gametic (within-gametes) disequilibrium (D_{ij}) was estimated on megagametophytes from Greek seedlots. A non-random association of alleles was detected for 3 out 253 pairs of loci only (1.2%). Significant correlation of alleles was observed for marker pairs OPG08-E/OPG08-I ($r_{ij} = 0.163$; $\chi^2_H = 4.337$, $P < 0.05$; $\chi^2_D = 5.655$, $P < 0.05$), OPG08-A/OPG08-H ($r_{ij} = 0.153$; $\chi^2_H = 0.022$, *n.s.*; $\chi^2_D = 4.412$, $P < 0.05$) and OPG02-D/OPG02-F ($r_{ij} = 0.177$; $\chi^2_H = 1.237$, *n.s.*; $\chi^2_D = 6.724$, $P < 0.01$). Whatever the cause of the significant gametic disequilibria detected might be (linkage, selection, etc.), the above marker pairs showing significant, homogeneous gametic disequilibrium in both populations were discarded from further analysis.

Pairwise composite genotypic disequilibrium (Δ_{ij}) was computed for each of the five Italian populations. This measure is inclusive of the within- and between-gametes non-random association of alleles. When all of the within-population disequilibria were considered, significant Δ_{ij} values were detected for 261 out of 2,719 marker pairs (9.59%), a proportion larger than that expected by pure chance (Table 2). The proportion of significant cases was larger than 5% in all the populations considered (La Spina: 10.51%; Pollino: 10.88%; Pollino A: 7.95%; Pollino B: 8.47%; Pollino C: 10.00%). Pooling data over all the populations analyzed, we detected a significant correlation of alleles for 59 out 253 marker pairs, but only 19 values (7.51%) were homogeneous over the five Italian populations. Not surprisingly, homogeneous, significant Δ_{ij} was detected over the Italian populations for marker pairs showing significant D in the Greek populations (see above). Moreover, marker pairs that turned out to be tightly associated in a previous study (*6PGD-BIPGI-B* $rf = 0.080$; Morgante et al. 1993) showed significant disequilibrium over all the populations considered (data not shown). Whatever the reasons causing significant, homogeneous disequilibrium across different populations (linkage, selection etc.) the above 19 markers pairs were discarded from further analyses.

Estimation of effective population size

Estimation of effective population size (N_e) was carried out based upon the genetic relationship between randomly generated disequilibrium and N_e reported by Hill (1981), Laurie-Ahlberg and Weir (1979) and Waples (1991). Marker pairs showing some evidence of linkage based on a previously reported study (Morgante et al. 1993b) were discarded a priori. Moreover, we discarded marker pairs showing significant within-gamete (D_{ij}) or composite genotypic disequilibrium (Δ_{ij}) values homogeneous over all the populations analyzed.

When r_{ij} values obtained by RAPD markers only were used, the effective population size was quite small for three out of the five populations considered (for La Spina, Pollino and Pollino B; $N_e = 17.31$, 16.57 and 31.41, respectively, Table 2). Even when the upper confidence intervals were considered, a surprisingly small

Table 2 Randomly generated composite genotypic disequilibria, estimation of effective population size and genetic parameters for five Italian populations of *Pinus leucodermis* using RAPD markers

Population	Disequilibria			Effective population size			Population parameters	
	Sign. Δ_{ij}[a]	Valid	Prop.	N_e[b]	C.I. (lower)[c]	C.I. (higher)[c]	H_{exp}[d]	F_{IS}[e]
La Spina	60	571	0.1051	17.31	10.91	31.22	0.273	0.053
Pollino	57	524	0.1088	16.57	10.55	29.24	0.174	−0.028
Pollino A	47	591	0.0795	$+\infty$	31.39	$+\infty$	0.213	0.010
Pollino B	35	413	0.0848	31.41	13.94	647.38	0.251	0.000
Pollino C	62	620	0.1000	$+\infty$	46.79	$+\infty$	0.251	0.015
Overall	261	2719	0.0960			Average	0.231	0.012

[a] Composite genotypic disequilibria for marker pairs calculated following Weir (1996)
[b] Effective population size estimates obtained from randomly generated linkage disequilibrium data as reported by Lauhrie–Ahlberg and Weir (1979)
[c] Confidence intervals for N_e calculated following formulas proposed by Waples (1991)
[d] Corrected expected heterozygosity (Nei 1987)
[e] Average fixation index weighted on sample size (Nei 1987)

effective population size was estimated (lower than or around 30 for the first two populations). For the two remaining populations (Pollino A and Pollino C), negative values of N_e were obtained: as reported by Bartley et al. (1992), this could be due either to the sampling of an infinite population ($r_2 = 0.000$, no disequilibrium) or to sampling error ($r_2 < 1/S$, Hill 1981). In spite of the small sample size analyzed in this study, we obtained quite consistent confidence intervals for the other populations (N_e larger than zero). Therefore, we interpret the negative values obtained as being due to a sampling of infinite populations.

No correlations between N_e and either genetic diversity or the fixation index estimated by RAPD data was found (Table 2).

Discussion

The comparison of pairwise genetic divergence between populations using allozyme and RAPD markers showed a good agreement. Correlation coefficients between estimates obtained with the two types of markers were fairly high for both genetic distance and genetic divergence (0.792 and 0.596, respectively). Thormann and colleagues (1994) compared cophenetic values obtained for several accessions of cruciferous species using restriction fragment length polymorphism (RFLP) and RAPD data, obtaining a correlation of 0.932 for intraspecific comparisons and 0.363 for interspecific comparisons.

The greater sensitivity of RAPDs to population divergence (about 8 times greater than using allozymes) may derive from several causes. RAPDs are expected to originate primarily from non-coding, repetitive DNA (Williams et al. 1990; Paran et al. 1991; Plomion et al. 1995). It is well-known that the repetitive component of genomes can evolve rapidly both among and within species (Dover 1982; Cullis 1987; Charlesworth et al. 1984). Thus, the strong differentiation may result from the nature of the sequences sampled by random amplification techniques, whereas allozymes reflect the more conservative portion of the genome containing expressed genes. It has also been suggested that allozymes can be under the influence of balancing selection (Altukhov 1990, 1991; Karl and Avise 1992), though only indirect evidence has been reported in forest trees (Mitton 1994). As a consequence a shorter period of isolation is probably required for stronger RAPD differentiation to develop with respect to allozyme markers.

In this study, the population divergence obtained by RAPDs was remarkably larger than that generally reported using allozyme markers in forest trees (Godt and Hamrick 1990; Müller-Starck et al. 1992). RAPDs and isozymes gave comparable estimates of differentiation when populations of black spruce from a contiguous region were studied (Isabel et al. 1995). On the other hand, a strong differentiation among populations by RAPDs has been observed among races of Douglas-fir, with genetic differentiation revealed by RAPD exceeding by about 3 times that for allozymes (Aagaard et al. 1995).

In this investigation, multivariate analysis (PCA) pointed out an obvious ordination of populations analyzed. Based on allele frequencies, Italian populations were much more scattered than Greek populations. Interesting to note is that the divergence among Italian populations was about as large as that between Italian and Greek populations. Since it is likely that the split of two nuclei took place more than 10,000 years ago, one would expect a larger differentiation between rather than within the above nuclei.

According to theory, small populations located at the margin of a species' natural range are expected to

diverge by a larger extent based on the distance from the main gene pool (situated in Greece), thereby determining higher F_{ST} values. On the other hand, it could be envisaged that variance in allele frequencies may be larger for the Italian populations due to a larger sample size ($n = 5$ vs. $n = 2$ for Italian and Greek populations, respectively.) Based on our sampling, the above effects cannot be distinguished: further analyses are needed to confirm the above evidence.

Nonetheless, if genetic drift is responsible for the larger divergence of Italian populations, large randomly generated linkage disequilibrium can be expected between neutral, unlinked markers (Ohta and Kimura 1969; Hill 1981). The extent of the statistically significant disequilibria detected in this study was larger than that expected by chance alone and may be related to the demographical history of the populations considered (Slatkin 1994). In fact, Italian stands are partly characterized by low density, and fire represents one of the main ecological factors. Yang and Yeh (1993) found an extensive multilocus association of alleles in *Pinus contorta* thought to be due to a severe population bottleneck (reduction of the effective population size) formerly experienced by the populations as a consequence of fire.

On the other hand, linkage disequilibrium between pair of loci may stem from other reasons, like tight linkage (Epperson and Allard 1989; Hastings 1990), epistasis (Allard et al. 1972; Clegg et al. 1972; Brown et al. 1977), population subdivision (Prout et al. 1973; Smouse and Neel 1977) or hitch-hiking (Laurie-Ahlberg and Weir 1979; Hedrick 1980; Luckett and Edwards 1986). In our study, the neutrality of the two class of markers was tested, and marker pairs showing significant, homogeneous disequilibrium over all the Italian populations were discarded from the analysis. Indeed, marker pairs turned out to be tightly associated in a previous study (*6PGD-B/PGI-B rf* = 0.080 – Morgante et al. 1993) showed large, significant, homogeneous disequilibrium over all the populations considered (data not shown). However, a significant amount of random disequilibrium between loci was retained within populations. Moreover, previously reported evidence from two out of the five Italian populations considered in this study seems to reject the existence of substructuring within the stands (Morgante et al. 1991).

If genetic drift is responsible for the divergence observed, lowered genetic diversity is expected for small marginal populations, as reported for many forest tree species (Guries and Ledig 1982; Yeah and Layton 1979; Furnier and Adams 1986; Cwynar and McDonald 1987; Hamrick et al. 1989; Millar and Marshall 1992). On the other hand, many studies have failed to detect reduced genetic variation for peripheral compared to central populations of forest trees (Tigerstedt 1973; Yeh and O'Malley 1980; Betancourt et al. 1991). As for *Pinus leucodermis*, previously reported studies carried out using chloroplast SSR markers on the same populations considered in this investigation revealed a larger amount of within-population diversity for Greek stands ($H = 0.593 \pm 0.033$) than for the Italian ones ($H = 0.205 \pm 0.065$; Powell et al. 1996). In contrast, former allozymic studies on the same populations revealed an expected genetic diversity of 0.126 ± 0.014 and 0.102 ± 0.025 for Italian and Greek populations, respectively (Boscherini et al. 1994a). The RAPD markers considered in this study showed a slightly larger amount of genetic diversity for Italian ($H_e = 0.231 \pm 0.018$) than for Greek stands ($H_e = 0.216 \pm 0.013$). Further analyses based on larger samples are needed to throw light on the connection between genetic divergence, geographical isolation and amount of genetic variability detectable analyzing nuclear and organelle genomes.

Many genetic methods have been applied to estimate N_e in natural populations, most of which include the estimation of standardized variance of allele frequency through time (Nei and Tajima 1981; Waples 1990; Nunney and Elam 1994, and citation therein). A short-term genetic method for the estimation of N_e on randomly generated disequiibrium (Laurie-Ahlberg and Nei 1979) is thought to be less reliable due to a possible violation to the assumptions in natural populations (Hill 1981). In spite of that, confidence intervals associated with estimates obtained in this study were quite consistent for two of the five population analyzed. However, given the above considerations, estimates of N_e obtained in this study should be considered with caution.

Estimation of the effective population size (the "panmictic unit"; Wright 1961) is an important genetic parameter in conservation genetics and policies (Nunney and Elam 1994). A reduction in genetic variability along with a reduction in population size has been reported in *Eucalyptus albens* (Prober and Brown 1994). Moreover, the smaller the effective population size, the larger the expected increase of the level of inbreeding through generations (Falconer 1989). If the assumptions underlying the estimation of N_e in *Pinus leucodermis* hold, there is the possibility of losing genetic variation by drift in the next generations for at least two out of the five populations analyzed in this study.

Analysis carried out in this study can also be considered a helpful tool in conservation genetics of forest trees in establishing priority to be taken for conservation policies and management of endangered species. According to Lesica and Allendorf (1994), the strong divergence found among Italian populations of *Pinus leucodermis* suggests the high conservation value of these marginal populations.

Acknowledgements We thank Dr. A. Schettino for collecting seeds from the Italian populations. This work was partly supported by EC Project Bio2 CT93 0373.

References

Aagaard JE, Vollmer SS, Sorensen FC, Strauss FC (1995) Mitochondrial DNA products among RAPD profiles are frequent and strongly differentiated between races of Douglas-fir. Mol Ecol 4:441–446

Allard RW, Babbel GR, Clegg MT, Kahler AL (1972) Evidence for coadaptation in *Avena barbata*. Proc Natl Acad Sci USA 69:3043–3048

Altukhov YuP (1990) Population genetics: diversity and stability. Harwood Academic Publ, London

Altukhov YuP (1991) The role of balancing selection and overdominance in maintaining allozyme polymorphisms. Genetica 85:79–90

Avolio S (1984) II pino loricato (*Pinus leucodermis* Ant.). Ann 1st Sper Selv 17:79–153

Bartley D, Bagley M, Gall G, Bentley B (1992) Use of linkage disequilibrium data to estimate effective population size of hatchery and natural fish populations. Conserv Biol 6:365–375

Betancourt JL, Schuster WS, Mitton JB, Anderson RS (1991) Fossil and genetic history of a pinyon pine (*Pinus edulis*) isolate. Ecology 71:1685–1697

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

Boscherini G, Morgante M, Rossi P, Vendramin GG, Vicario F (1994b) Detection of DNA polymorphisms in *Pinus leucodermis* Ant. using random amplification. For Genet 1:131–137

Brown AHD, Nevo E, Zohary D (1977) Association of alleles at esterase loci in wild barley, *Hordeum spontaneum* L. Nature 268:430–431

Carson HL (1990) Increasing genetic variance after a population bottleneck. Trends Ecol Evol 5:228–231

Cesarone CF, Bolognesi C, Santi L (1979) Improved microfluorometric DNA determination in biological material using 33258 Hoechst. Anal Biochem 100:188–197

Charlesworth B, Sniegowski P, Stephan W (1994) The evolutionary dynamics of repetitive DNA in eukaryotes. Nature 371:215–220

Clegg MT, Allard RW, Kahler AL (1972) Is the gene the unit of selection? Evidence from two experimental populations. Proc Natl Acad Sci USA 69:2474–2478

Cockerham CC, Weir BS (1977) Digenic descent measures in a finite population. Genet Res 30:121–147

Cullis CA (1987) The generation of somatic and heritable variation in response to stress. Am Nat 130:62–73

Cwynar LC, MacDonald GM (1987) Geographical variation of lodgepole pine in relations to population history. Am Natur 129:463–469

Dover G (1982) Molecular drive: a cohesive mode of species evolution. Nature 299:111–117

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

Efron B (1982) The jackknife, the bootstrap and other resampling plans. Society for Industrial and Applied Mathematics, Philadelphia

Ellstrand NC (1992) Gene flow by pollen: implications for plant conservation genetics. Oikos 63:77–86

Epperson BK, Allard RW (1989) Spatial autocorrelation analysis of the distribution of genotypes within populations of lodgepole pine. Genetics 121:369–377

Falconer DS (1989) Introduction to quantitative genetics, 3rd edn. Longman, London

Furnier GR, Adams WT (1986) Geographic patterns of allozyme variation in Jeffrey pine. Am J Bot 73:1009–1015

Gibbs HL, Prior KA, Weatherhaed PJ (1994) Genetic analysis of populations of threatened snake species using RAPD markers. Mol Ecol 3:329–337

Gilpin M (1991) The genetic effective size of a metapopulation. Biol J Linn Soc 42:165–175

Godt MJ, Hamrick JL (1990) Allozyme diversity in plant species. In: Brown AHD, Clegg MT, Kahler AL, Weir BS (eds.) Plant population genetics, breeding, and genetic resources. Sinauer, Sunderland, Mass., pp 43–63

Guries RP, Ledig FT (1982) Genetic diversity and population structure in pitch pine (*Pinus rigida* Mill.). Evolution 36:387–402

Harmrick JL, Blanton HM, Hamrick KJ (1989) Genetic structure of geographically marginal populations of ponderosa pine. Am J Bot 76:1559–1568

Hastings A (1990) The interaction between selection and linkage in plant populations. In: Brown AHD, Clegg MT, Kahler AL, Weir BS (eds) Plant population genetics, breeding and genetic resources. Sinauer, Sunderland, Mass., pp 163–180

Hedrick PW (1980) Hitchhicking: a comparison of linkage and partial selfing. Genetics 94:791–808

Hill WG (1981) Estimation of effective population size from data on linkage disequilibrium. Genet Res 38:208–216

Isabel N, Beaulieau J, Bousquet J (1995) Complete congruence between gene diversity estimates derived from genotypic data at enzyme and random amplified polymorphic DNA loci in black spruce. Proc Natl Acad Sci USA 92:6369–6373

Karl SA, Avise JC (1992) Balancing selection at allozyme loci in oysters: implications from nuclear RFLPs. Science 256:100–102

Laurie-Ahlberg CC, Weir BS (1979) Allozymic variation and linkage disequilibrium in some laboratory populations of *Drosophila melanogaster*. Genetics 92:1295–1314

Lesica P, Allendorf FW (1994) When are peripheral populations valuable for conservation? Conserv Biol 9:753–760

Lewontin RC, Krakauer J (1979) Distribution of gene frequency as a test of the theory of the selective neutrality of polymorphism. Genetics 175–195

Luckett DJ, Edwards KJR (1986) Esterase genes in parallel composite cross barley populations. Genetics 114:289–302

Millar CI, Marshall KA (1992) Allozyme variation in Port-Oxford cedar (*Chamaecyparis lawsonii*): implications for genetic conservation. For Sci 37:1060–1077

Mitton JB (1994) Molecular approaches to population biology. Rev Ecol Sys 25:45–69

Morgante M, Vendramin GG, Olivier AM (1991) Mating system analysis in *Pinus leucodermis* Ant.: detection of self-fertilization in natural population Heredity 67:197–203

Morgante M, Vendramin GG, Giannini R (1993) Inheritance and linkage relationships of isozyme variants of *Pinus leucodermis* Ant. Silvae Genet 42:231–237

Morgante M, Rossi P, Vendramin GG, Giannini R (1994) Low levels of outcrossing in *Pinus leucodermis* Ant.: further evidences in artificial stands. Can J Bot 72:1289–1293

Mosseler A, Innes DJ, Roberts BA (1991) Lack of allozymic variation in disjunct Newfoundland populations of red pine (*Pinus resinosa*). Can J For Res 21:525–528

Müller-Starck G, Baradat Ph, Bergmann F (1992) Genetic variation within European tree species. New For 6:23–47

Nei (1973) Analysis of gene diversity in subdivided populations. Proc Natl Acad Sci USA 70:3321–3323

Nei M, Tajima F (1981) Genetic drift and estimation of effective population size. Genetics 98:625–640

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

Nunney L, Elam DR (1994) Estimating effective population size of conserved populations. Conserv Biol 8:175–184

Ohta T, Kimura M (1969) Linkage disequilibrium at a steady state determined by random genetic drift and recurrent mutation. Genetics 63:229–238

Paran I, Kesseli RV, Michelmore RW (1991) Identification of restriction fragment length polymorphisms and random amplified polymorphic DNA markers linked to downy mildew resistance genes in lettuce, using near-isogenic lines. Genome 34:1021–1027

Plomion C, Bahrman N, Durel C-E, O'Malley DM (1995) Genomic mapping in *Pinus pinaster* (Maritime pine) using RAPD and protein markers. Heredity 74:661–668

Powell W, Morgante M, McDewitt R, Vendramin GG, Rafalski AJ (1996) Polymorphic simple sequence repeat regions in chloroplast genome: application to the population genetics of pines. Proc Natl Acad Sci USA 92:7759–7763

Prober SM, Brown AHD (1994) Conservation of the grassy white box woodlands: population genetics and fragmentation of *Eucalyptus albens*. Conserv Biol 8:1003–1013

Prout TJ, Bundgaard J, Bryant S (1973) Population genetics of modifiers and meiotic drive. I. The solution of a special case and some general implications. Theor Pop Biol 4: 446–465

Raijmann LEL, Van Ginkel WE, Heckel DG, Menken SBJ (1997) Genetic variation and outcrossing rate in relation to population size in *Gentiana pneumonanthe* L. Heredity 8:1014–1026

Rousset F, Raymond M (1995) GENEPOP (version 1.2): a population genetics software for exact tests and ecumenicism. J Hered 86:248–249

Simon J, Bergeron Y, Gagnon D (1986) Isozyme uniformity in populations of red pine (*Pinus resinosa*) in the Abitibi region, Quebec. Can J For Res 16:1133–1135

Slatkin M (1994) Linkage disequilibrium in growing and stable populations. Genetics 137:331–336

Smouse PE, Neel JV (1977) Multivariate analysis of gametic disequilibrium in the Yanonama. Genetics 85:733–752

Thormann CE, Ferreira ME, Camargo LEA, Tivang JG, Osborn TC (1994) Comparison of RFLP and RAPD markers to estimating genetic relationships within and among cruciferous species. Theor Appl Genet 88:973–980

Tigerstedt PMA (1973) Studies of isozyme variation in marginal and central population of *Picea abies*. Hereditas 75: 47–60

Van Treuren, Bijlsma R, Van Delden W (1991) The significance of genetic erosion in the process of extinction. I. Genetic differentiation in *Salvia pratensis* and *Scabiosa columbaria* in relation to population size. Heredity 66:181–189

Waples RS (1990) Conservation genetics of pacific salmon. III. Estimating effective population size. J Hered 81:277–289

Waples RS (1991) Genetic methods for estimating the effective population size of cetacean populations. In: Hoelzel AR (ed) Genetic ecology of whales and dolphins. International Whaling Commission (Special Issue no. 13)

Weir BS (1996) Genetic data analysis II. Sinauer, Sunderland, Mass.

Williams JGK, Hanafey MK, Rafalski JA, Tingey SV (1993) Genetic analysis using random amplified polymorphic DNA markers. Methods Enzymol 218:701–740

Wright S (1951) The genetical structure of populations. Ann Eugen 15:323–354

Wright S (1965) The interpretation of population structure by F Statistics with special regard to system of mating. Evolution 19:395–420

Yang R-C, Yeh FC (1993) Multilocus structure in *Pinus contorta* Dougl. Theor Appl Genet 87:568–576

Yeh FC, Layton C (1979) The organization of genetic variability in central and marginal populations of lodgepole pine (*Pinus contorta* ssp *latifolia*). Can J Genet Cytol 21:487–503

Yeh FC, O'Malley D (1980) Enzyme variation in natural populations of Douglas-fir, *Pseudotsuges menziesii* (Mirb) Franco, from British Columbia. 1. Genetic variation patterns in coastal populations. Silvae Genet 29:83–92

Young A, Boyle T, Brown T (1996) The population genetic consequences of habitat fragmentation in plants. Trends Ecol Evol 11:413–419