Molecular Ecology (2002) **11**, 869–877

Comparison of genetic diversity estimates within and among populations of maritime pine using chloroplast simple-sequence repeat and amplified fragment length polymorphism data

M. M. RIBEIRO,* S. MARIETTE,†‡ G. G. VENDRAMIN,§ A. E. SZMIDT,¶ C. PLOMION‡ and A. KREMER‡

*Unidade Departamental de Silvicultura e Recursos Naturais, Escola Superior Agrária, 6001-909 Castelo Branco, Portugal, †CEMAGREF, Nogent-sur-Vernisson, France, ‡INRA, Laboratoire de Génétique et Amélioration des Arbres Forestiers, Cestas, France, §Istituto Miglioramento Genetico Piante Forestali, Firenze, Italy, ¶Laboratory of Molecular Population Genetics, Kyushu University, Fukuoka, Japan

Abstract

We compared the genetic variation of *Pinus pinaster* populations using amplified fragment length polymorphism (AFLP) and chloroplast simple-sequence repeat (cpSSR) loci. Populations' levels of diversity within groups were found to be similar with AFLPs, but not with cpSSRs. The high interlocus variance associated with the AFLP loci could account for the lack of differences in the former. Although AFLPs revealed much lower genetic diversity than cpSSRs, the levels of among-population differentiation found with the two types of marker were similar, provided that loci showing fewer than four null-homozygotes, in any population, were pruned from the AFLP data. Moreover, the French and Portuguese populations were clearly differentiated from each other, with both markers. The Mantel test showed that the genetic distance matrix calculated using the AFLP data was correlated with the matrix derived from the cpSSRs. Because of the concordance found between markers we conclude that gene flow was indeed the predominant force shaping nuclear and chloroplastic genetic variation of the populations within regions, at the geographical scale studied.

Keywords: AFLP, cpSSR, chloroplast microsatellites, gene diversity, population differentiation, *Pinus pinaster*

Received 23 July 2001; revision received 21 January 2002; accepted 24 January 2002

Introduction

The combined effects of natural selection, genetic drift, mutation and gene flow determine the patterns of genetic structure of subdivided populations. Genetic drift tends to increase genetic structure, whereas gene flow among populations slows down the differentiation process, until a steady state is eventually reached between the opposing effects of gene flow and genetic drift (Wright 1943; Slatkin 1987). The frequencies of neutral alleles are not influenced by natural selection, thus gene flow and genetic drift are expected to affect all loci uniformly when the mutation rate is much lower than the migration rate (Kimura 1968). Mutation rates are generally ignored because they are considered to be much lower than the migration rates, but this might not always be valid. Recent data show that mutation rates are higher at the chloroplast microsatellite (cpSSR) loci than substitution rates elsewhere in the chloroplast genome, and generally higher than in the nuclear genome sequences, except for the nuclear microsatellites (SSRs) (Provan *et al*. 1999 and references therein), and higher mutation rates may underestimate population differentiation (Hedrick 1999). A high mutation rate enhances homoplasy, which may also lead to underestimates of differentiation when cpSSRs are used, by erasing some of the differences in haplotypes that have arisen in the past (Doyle *et al*. 1998).

Correspondence: M. M. Ribeiro. Tel. 351 272 339970; Fax: 351 272 339901; E-mail: mribeiro@esa.ipcb.pt
M. M. Ribeiro and S. Mariette contributed equally to this work.

However, the chloroplast genome is haploid and does not undergo recombination. Therefore, population subdivision is expected to be more prevalent for chloroplastic genes than for nuclear genes, as a consequence of the uniparental inheritance and lower effective number of chloroplastic genes in diploid species, and because of differences in seed and pollen migration (Petit *et al*. 1993; Ennos 1994).

The amplified fragment length polymorphism (AFLP) technique, first described by Vos *et al*. (1995) generates a large number of markers that are representative of the nuclear genome as a whole with a high degree of reproducibility. Nevertheless, they are generally dominant markers when diploid material is used and a bias is introduced into the estimation of population-genetic parameters (Isabel *et al*. 1995; Szmidt *et al*. 1996). According to Lynch & Milligan (1994), dominant markers can be used to estimate unbiased population-genetic parameters, provided that the loci with low-frequency null alleles are pruned from the analysis.

Higher levels of differentiation would be expected for cpSSR markers than for AFLP markers, because of differences in inheritance and ploidy, but higher mutation rate and homoplasy would tend to decrease the level of differentiation for cpSSR markers, comparative to AFLP markers. If the properties of each type of marker only are considered, the situation is complex and there is no simple theoretical expectation against which to test observed levels of differentiation obtained with AFLP and cpSSR markers. Nevertheless, diversity and differentiation are dependent not only on the marker-specific properties, such as mutation rate, homoplasy and inheritance, but also on intensity of gene flow among populations.

Pinus pinaster is distributed throughout the western part of the Mediterranean region, but discontinuously, because of the geographical isolation of pine populations and the impact of ancient human activities in the Mediterranean. The generally scattered distribution of this species may have prevented or limited gene flow among the different groups of populations, promoting high levels of genetic divergence among regions due to genetic drift (Baradat & Marpeau-Bezard 1988; Bahrman *et al*. 1994; Petit *et al*. 1995; Vendramin *et al*. 1998). At a fine geographical scale, within the French and Portuguese groups of populations, gene flow and human activity were probably responsible for the weak among-population differentiation that has been found in *P. pinaster* using both nuclear (isozymes, Castro 1989; nuclear SSR and AFLP markers, Mariette *et al*. 2001b) and chloroplast SSR markers (Ribeiro *et al*. 2001).

In this study we examine the genetic variation at six cpSSR and 100 AFLP loci in 24 *P. pinaster* populations. The 12 Portuguese populations of *P. pinaster* were analysed in a previous study with the six cpSSRs (Ribeiro *et al*. 2001), and the 12 French populations analysed using AFLPs in the study published by Mariette *et al*. (2001b) were also incorporated. Our primary purpose was to test for concordance in patterns of genetic differentiation and diversity with two different genetic markers (AFLP and cpSSR). We should expect congruence in the estimates of diversity and differentiation with AFLPs and cpSSRs if a high amount of gene flow exits, because in conifers migration is mainly via the pollen. In the case of differentiation, the absolute values are directly comparable, as it is an independent measure (two diversities ratio). In the case of the diversity, higher absolute values are expected for microsatellites (higher mutation rates), as assessment of diversity is dependent on the measure used. Consequently, we compared the relative values of the within-population diversity to test the congruence between marker systems.

Materials and Methods

Plant material and DNA extraction

The French group included 12 *Pinus pinaster* populations located in the Aquitaine region of France in which the needles were collected from trees randomly selected within stands as described by Mariette *et al*. (2001b). The Portuguese group included 12 populations in which needles were collected from trees randomly selected within stands as described by Ribeiro *et al*. (2001). The DNA was extracted according to the Doyle & Doyle (1990) method with modifications described by Lerceteau & Szmidt (1999). The location of populations and the number of individuals per population are listed in Table 1.

AFLP and cpSSR analysis

AFLP analysis was performed according to the protocol developed for this species by Costa *et al*. (2000). Selective amplifications were performed on the pre-amplified fragments using two primer–enzyme combinations (PECs): *Eco*RI-ACC/*Mse*I-CCTG (PEC1) and *Eco*RI-ACG/*Mse*I-CCCA (PEC2). The PECs used were selected according to the number of amplified loci (110 and 81, respectively) and the number of polymorphic loci (29 and 28, respectively) detected in a mapped population (Costa *et al*. 2000). RFLPSCAN Version 3.0 software (Scanalytics) was used to score the AFLP fragments of the French populations. For the Portuguese populations the protocol was modified slightly. The *Eco*RI primer was radiolabelled with γ-^{33}P ATP. Following amplification, reaction products were mixed with an equal volume of loading buffer (95% formamide, 200 μM EDTA, 0.1% bromophenol blue and 0.1% xylene cyanol). After denaturation for 8 min at 90 °C, 4 μL of each sample was loaded on a 4% acrylamide/bisacrylamide (19:1) denaturating gel. The gel was pre-equilibrated by passing an electric current through it (at a constant 80 W) for 60 min. The reaction products were then

Table 1 Geographical parameters from the 24 *Pinus pinaster* populations, and numbers of individuals sampled per population, *n*

Country	*n*	Population	Alt. (m)	Latitude	Longitude
France	30	Lit-et-Mixe	30–40	44°03′ N	1°19′ W
France	28	St-Julien-en-Born	20	44°06′ N	1°19′ W
France	27	Boul. Allemands	20	44°05′ N	1°19′ W
France	29	Ste-Eulalie-en-Born	40–50	44°20′ N	1°14′ W
France	25	Mimizan	35–40	44°08′ N	1°18′ W
France	30	Vielle St-Girons	35	43°56′ N	1°28′ W
France	29	Biscarrosse	25–60	44°20′ N	1°13′ W
France	32	Lège	15	44°43′ N	1°12′ W
France	26	Lacanau	10–15	45°02′ N	1°09′ W
France	26	Pointe de Grave	10–15	45°34′ N	1°04′ W
France	26	Carcans	10–15	45°06′ N	1°09′ W
France	29	Hourtin	25–45	45°10′ N	1°08′ W
Portugal	20	Aveiro	30	40°39′ N	8°36′ W
Portugal	20	Oleiros	750	39°55′ N	7°50′ W
Portugal	19	Alcácer do Sal	20	37°52′ N	8°30′ W
Portugal	20	Bragança	800	41°52′ N	6°32′ W
Portugal	20	Figueira da Foz	30	40°18′ N	8°44′ W
Portugal	19	Lousã	250	40°09′ N	8°11′ W
Portugal	19	Monção	310	42°04′ N	8°23′ W
Portugal	20	Mondim de Basto	480	41°25′ N	7°55′ W
Portugal	20	Leiria	50	39°46′ N	8°57′ W
Portugal	20	Manteigas	625	40°24′ N	7°26′ W
Portugal	20	Montalegre	690	41°49′ N	7°56′ W
Portugal	18	Sintra	250	38°46′ N	9°22′ W

separated, under these conditions, for ≈ 150 min After the fragments had been separated, the gel was fixed in 10% acetic acid for ≈ 20 min, rinsed with water, dried and exposed to Konica X-ray film for 5–7 days. The presence and absence of the bands were then scored visually. In each gel, the same two individuals from the French populations were included as controls, to calibrate each gel to a known band level and thus avoid any shifts in scoring from one gel to another for a given PEC. To check the reproducibility of the visual vs. RFLPSCAN readings, the same sample set, consisting of 48 individuals from different Portuguese populations, was used for each PEC.

Six of the 20 primers flanking pine chloroplast microsatellites, based on sequences of the *Pinus thunbergii* chloroplast genome (Vendramin *et al*. 1996), were used in this study: Pt1254, Pt15169, Pt30204, Pt36480, Pt71936 and Pt87268. The cpSSR PCR conditions and analysis were performed according to the protocol described for *P. pinaster* by Ribeiro *et al*. (2001). The amplified fragments were scored and sized as described by Ribeiro *et al*. (2002).

Data analysis

For each AFLP locus two alleles were considered, one for the presence (*A*) and the other for the absence (*a*) of the corresponding band. Individuals that showed the presence of a band were considered to be dominant homozygotes (*AA*) or heterozygotes (*Aa*), whereas those in which it was absent were scored as null homozygotes (*aa*). Loci that showed the presence of both phenotypes were considered polymorphic, and the polymorphic loci found in the two groups were compared using a χ^2 goodness-of-fit test (Steel & Torrie 1981).

Two different genotypic analyses (G_1 and G_2) were performed on the AFLP data following different criteria. In the first genotypic analysis (G_1), the frequencies of the genotypes (*AA*, *Aa* and *aa*) were deduced according to F_{IS} values, i.e. the deficiency in heterozygotes, which indicates the departure from Hardy–Weinberg equilibrium (HWE). An unbiased estimate of *q*, the frequency of the allele that generates no band (the null allele), was obtained according to Mariette *et al*. (2001b), assuming that the true value of F_{IS} is known. Chong *et al*. (1994) showed that available data from co-dominant markers could be used to correct the potential deviations from HWE. Therefore, we used the fixation index, $F_{IS} = 0.05$, obtained for *P. pinaster* in an analysis of nuclear microsatellites by Mariette *et al*. (2001a). In the second genotypic analysis (G_2), the correction introduced for dominant markers by Lynch & Milligan (1994) in order to obtain unbiased estimates of the population-genetic parameters (designated LM restriction) was incorporated into the G_1 analysis. Any locus that showed fewer than four individuals with the null phenotype in any population was pruned from the analysis.

The AFLP allele frequencies obtained with the two types of analysis (G_1 and G_2) were used to compute the following

parameters: the unbiased gene diversity within each population, H_E; the average for all populations, H_S; and the unbiased total gene diversity, H_T (Nei 1987). The estimate of the variance associated with H_{Ei} was computed by intralocus and interlocus bootstrapping (1000 replicates), SD^2_{1i} and SD^2_{2i}, for the i-th population. The total variance of diversity was derived using the additive model $SD^2_{Ti} = SD^2_{1i} + SD^2_{2i}$. The differences between populations were computed based on a Z-test: $z = \frac{H_{Ei} - H_{Ej}}{\sqrt{SD^2_{Ti} + SD^2_{Tj}}}$, in which H_{Ei} and H_{Ej} are the unbiased diversity within the i-th and j-th populations, respectively, and SD^2_{Ti} and SD^2_{Tj} are the variances associated with the within population diversity for the i-th and j-th population, respectively.

Because the chloroplast genome is haploid and does not undergo recombination, it can be viewed as a single locus, and the size scores for the six fragments analysed were combined in order to derive the chloroplast haplotype of each individual. Nevertheless, we use the term locus to refer to a cpSSR site, and allele to refer to a size variant at a given cpSSR site. The variation of the chloroplast haplotype within populations was computed by estimating the total number of haplotypes, n_h; the effective number of haplotypes, n_E; the unbiased haplotypic diversity, H_E; the average over populations, H_S; and the total gene diversity, H_T (Nei 1987). The variance associated with the within population haplotypic diversity, SD^2_T, was estimated by bootstrapping over individuals (1000 replicates). The differences between populations for the parameter H_E were computed using a Z-test as described above.

For both types of markers, the degree of genetic differentiation among populations, for all populations used in this study and for each group taken separately, were estimated using the parameter G_{ST} (Nei 1987). Nei's unbiased distances were obtained for pair-wise populations, using the cpSSR-haplotype and AFLP-genotypic allelic frequencies (Nei 1978). The degree of relatedness between the genetic distance matrices generated by the two types of markers was measured using the Mantel matrix-correspondence test (Mantel 1967). In order to combine the information provided by the different loci, a multivariate approach based on principal component analysis (PCA) was performed with the informative ALFP loci allele frequencies (after the LM restriction) and with the transformed haplotype frequencies, $x_i = \arcsin(\sqrt{p_i})$, where p_i is the haplotypic frequency of the *i-th* population from the cpSSR data set.

For the two types of markers, parameters were compared. Resampling procedures were used to estimate the differences between the two groups of populations for each computed genetic parameter K ($K = H_S$, H_T and G_{ST}) for the two types of markers used. For a given K parameter, the associated variance was obtained by bootstrapping over populations (1000 replicates), as recommended by Petit & Pons (1998). The nuclear and the chloroplastic K estimates were compared between the two groups. The nuclear G_{ST} estimates were also compared with the chloroplastic G_{ST} estimates.

The programs used to compute the required parameters were HAPLOID and HAPDOM (Antoine Kremer, Equipe de Génétique et Amélioration des Arbres Forestiers, Cestas, France), TFPGA Version 1.3 (Mark P. Miller, Northern Arizona University, USA) and SAS SYSTEM Version 8 for Unix.

Results

AFLP and cpSSR analysis

A total of 100 reproducible AFLP bands was detected, 62% of which were polymorphic with the two PECs, and similar percentages of polymorphism were found with both combinations. Based on the AFLP data set, when all the populations were considered, the average total diversity based on the 100 loci was $H_T = 0.179$ and the average within-population diversity was $H_S = 0.152$. These parameters increased ≈ 1.5-fold when the LM restriction was applied, and the number of loci was reduced to 19 informative loci (Table 2).

In total 25 alleles (2–7 per locus) were detected in the 572 individuals at the 6-cpSSR loci. When all alleles were combined, 108 different haplotypes were found. The level of polymorphism obtained with the cpSSR for all the populations was very high, the total diversity and the average diversity were 0.944 and 0.908, respectively (Table 2).

A standardized Mantel statistics of $r = 0.344$ was obtained for the association of the genetic distances computed with both markers, with a probability of $P \leq 0.001$. Both matrices are correlated and they detected similar trends with respect to the genetic distances.

Genetic differentiation among populations and between groups

The levels of among-population differentiation found with the AFLP markers depended on the method of analysis employed (Table 3). When all the populations were considered, the G_1 method gave higher levels of among-population diversity (15.2%), compared with the G_2 approach (4.7%). The G_2 method detected similar levels of among-population differentiation in the Portuguese and French groups, whereas the AFLP G_1 method of analysis showed significantly ($P \leq 0.01$) higher among-population differentiation in the Portuguese group (6.9%) compared with the French (2.3%). The among-population differentiation values estimated using the cpSSR data were similar (2%) for both groups of populations (Table 3).

The G_1 method of analysis gave significantly higher levels of among-population differentiation compared with the

Table 2 Diversity statistics of the amplified fragment length polymorphism (AFLP) and chloroplast simple-sequence repeat (cpSSR) for the 24 *Pinus pinaster* populations

	H_{EG1} (SD_T)	H_{EG2} (SD_T)	H_{ECP} (SD_T)	n_h	n_E
French population					
Lit-et-Mixe	0.162 (0.022)	0.345 (0.026)	0.913 (0.033)	15	8.5
St-Julien Born	0.154 (0.021)	0.325 (0.028)	0.950 (0.024)	18	11.9
Boul. Allemands	0.154 (0.021)	0.334 (0.027)	0.974 (0.016)	19	16.2
Ste-Eulalie Born	0.146 (0.021)	0.311 (0.029)	0.948 (0.025)	18	11.8
Mimizan	0.155 (0.021)	0.326 (0.029)	0.957 (0.028)	18	12.3
Vielle St-Girons	0.158 (0.021)	0.341 (0.027)	0.903 (0.039)	15	7.9
Biscarrosse	0.147 (0.019)	0.313 (0.027)	0.983 (0.014)	23	19.6
Lège	0.163 (0.022)	0.344 (0.027)	0.919 (0.031)	17	9.1
Lacanau	0.131 (0.020)	0.273 (0.027)	0.988 (0.014)	22	19.9
Pointe de Grave	0.151 (0.021)	0.320 (0.027)	0.985 (0.014)	21	18.8
Carcans	0.150 (0.021)	0.323 (0.028)	0.935 (0.029)	15	9.9
Hourtin	0.147 (0.020)	0.312 (0.027)	0.926 (0.035)	17	9.4
H_S	0.151 (0.002) NS[a]	0.322 (0.005)***[a]	0.948(0.008)***[a]	18.2[b]	
H_T	0.155 (0.002) NS[a]	0.327 (0.005)***[a]	0.967(0.008)***[a]		
Portuguese population					
Aveiro	0.141 (0.023)	0.236 (0.043)	0.763 (0.065)	6	3.6
Oleiros	0.126 (0.022)	0.177 (0.042)	0.900 (0.041)	10	6.9
Alcácer do Sal	0.136 (0.022)	0.201 (0.049)	0.842 (0.054)	8	4.9
Bragança	0.162 (0.022)	0.240 (0.036)	0.847 (0.061)	9	5.1
Figueira da Foz	0.152 (0.022)	0.243 (0.042)	0.821 (0.072)	9	4.5
Lousã	0.143 (0.022)	0.199 (0.046)	0.912 (0.045)	11	7.4
Monção	0.157 (0.024)	0.198 (0.043)	0.848 (0.048)	7	5.1
Mondim Basto	0.158 (0.020)	0.222 (0.043)	0.884 (0.046)	10	6.2
Leiria	0.170 (0.022)	0.286 (0.034)	0.926 (0.034)	11	8.3
Manteigas	0.158 (0.022)	0.210 (0.039)	0.874 (0.046)	9	5.9
Montalegre	0.173 (0.023)	0.188 (0.043)	0.868 (0.039)	8	5.7
Sintra	0.144 (0.022)	0.241 (0.035)	0.909 (0.053)	11	7.0
H_S	0.152 (0.004)	0.220 (0.009)	0.866 (0.013)	9.1[b]	
H_T	0.163 (0.004)	0.223 (0.008)	0.884 (0.012)		
All populations					
H_S	0.152 (0.002)	0.226 (0.008)	0.908 (0.011)		
H_T	0.179 (0.002)	0.237 (0.008)	0.944 (0.011)		

H_{EG1} = the genotypic genetic diversity; H_{EG2} = the genotypic genetic diversity after the LM restriction; H_{ECP} = the unbiased haplotypic diversity; H_S = the average diversity; H_T = the total diversity; n_h = the number of haplotypes; n_E = effective number of haplotypes; SD_T = the standard deviation.

a. Significance of differences between the French and Portuguese group's H_S *and* H_T parameters: NS is nonsignificant and ***$P \leq 0.001$.

b. Average number of haplotypes.

value obtained with the cpSSR markers when all the populations were considered ($P \leq 0.001$) and when the Portuguese populations were considered separately ($P \leq 0.01$), but not when the French populations were considered separately (Table 4). When the LM restriction was applied, no differences in differentiation were found between the values computed for the cpSSRs and AFLPs, whichever group of populations was considered (Table 4).

The French and Portuguese groups of populations could be clearly distinguished by both principal component analyses performed with the AFLP and cpSSR data sets (Fig. 1A and B), and the percentage of the total variation explained was ≈ 51 and 25%, respectively. The centroids are more evenly dispersed for AFLP data than for cpSSR markers, thus indicating that the among-population differentiation, even if is about the same, on average, between the two types of markers, does not have the same type of distribution among populations from one type of marker to the other. For instance, in the case of cpSSR, all Portuguese populations are poorly differentiated, unlike with the AFLP. The French populations are also quite poorly differentiated, with the exception of four outlier populations, which are probably responsible for most of the among-population differentiation in cpSSR among French populations. This is quite a contrast with the evenly dispersed pattern of among-population differentiation observed with AFLP.

Table 3 Genetic differentiation statistics among French and Portuguese populations, and among all populations

	AFLP markers		cpSSR
	G_{STG1}	G_{STG2}	G_{STCP}
French pop.	0.023 (0.005)** [a]	0.016 (0.004) NS[a]	0.019 (0.003) NS[a]
Number of loci	100	40	1[b]
Portuguese pop.	0.069 (0.010)	0.015 (0.005)	0.020 (0.012)
Number of loci	100	17	1
All populations	0.152 (0.011)	0.047 (0.006)	0.039 (0.006)
Number of loci	100	19	1

AFLP = Amplified fragment length polymorphism. G_{STG1} = the genotypic genetic differentiation; G_{STG2} = the genotypic genetic differentiation after the LM restriction; G_{STCP} = the haplotypic genetic differentiation; the standard deviations in parentheses.
a. Significance of differences between the G_{ST} values of the French and Portuguese groups: NS nonsignificant or $^{**}P \leq 0.01$.
b. The chloroplast genome can be viewed as a single locus because it does not undergo recombination.

Table 4 Comparison of among-population differentiation derived from nuclear and chloroplastic marker analyses. Abbreviations as in Tables 2 and 3

	cpSSR G_{STCP}
French population	
G_{STG1}	NS
G_{STG2}	NS
Portuguese population	
G_{STG1}	**
G_{STG2}	NS
All populations	
G_{STG1}	***
G_{STG2}	NS

Diversity comparisons between the French and Portuguese groups

For the AFLPs left after LM restriction was applied, the average ($H_S = 0.322$) and the total diversity ($H_T = 0.327$) found in the French group were significantly higher ($P \leq 0.001$) than the estimates computed for the Portuguese group ($H_S = 0.220$ and $H_T = 0.223$, respectively). However, both parameters displayed similar values in the two groups of populations without LM restriction, i.e. when the G_1 approach was adopted (Table 2).

When populations were compared within group for the parameter H_E based on the AFLP data set, no differences were found among them with the G_1 approach. However, in the G_2 analysis, the Leiria population showed a significantly higher level of diversity than the other Portuguese populations ($P \leq 0.05$). No significant differences in the within population diversity estimates were found among

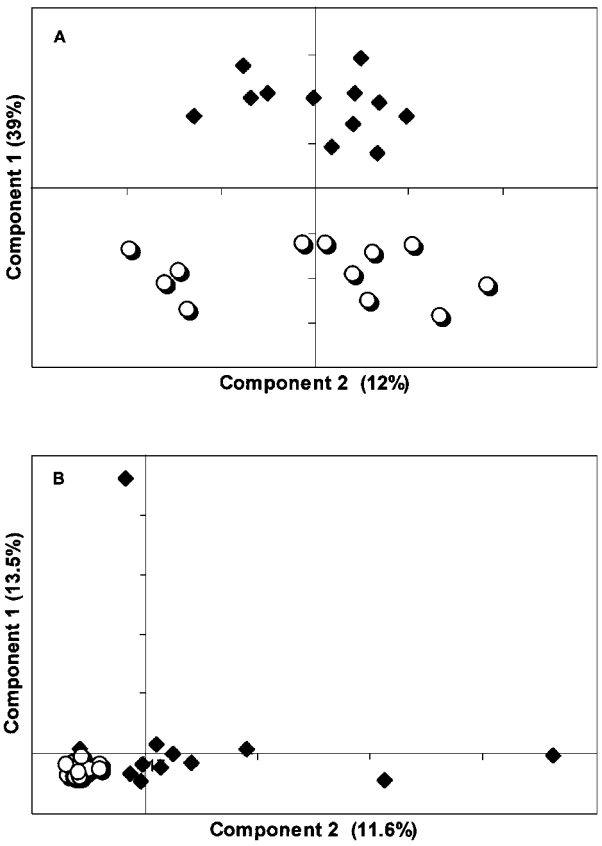

Fig. 1 Plots of the first two components of the standardized principal components analysis based on AFLP allelic frequencies from the 19 informative loci of the 24 studied populations (A), and the transformed haplotypic frequencies based on the cpSSR data set (B). The cpSSR data set was transformed as $x_i = \arcsin(\sqrt{p_i})$, where p_i is the observed haplotype frequency for the ith population. With both markers, a fairly good separation between the French populations (◆) and the Portuguese populations (○) could be obtained.

the populations of the French group. The interlocus variance had a larger contribution to the total variance associated with the parameter H_E, whichever type of AFLP analysis was considered, more than twice the intralocus variance (data not shown).

In estimates based on the cpSSR data, both H_S and H_T were found to be significantly higher in the French than in the Portuguese group ($P \leq 0.001$). In the French group, both the mean number of haplotypes (n_h) and the effective number of haplotypes (n_E) were found to be about twice as high as in the Portuguese group (Table 2).

Discussion

AFLP vs. cpSSR genetic variation

To our knowledge there are no published figures comparing chloroplast microsatellite population genetic structure estimates with any other markers' estimates. In this study, cpSSR markers detected a higher level of diversity than AFLP markers, which is a result predicted by theory because of the nature of these markers. Microsatellites are known to be highly variable within-population (Lefort *et al.* 1999, and references therein). Hedrick (1999) suggested that we need to evaluate the data obtained with highly variable loci, such as microsatellites, because the information they provide can be quite different to that provided by less variable markers. Nevertheless, in this study, although both types of marker have different properties and reveal different absolute values of diversity, the trend in genetic variation was the same (after LM restriction in the case of the AFLP loci), i.e. lower diversity in the Portuguese group of populations than the French group of populations. Furthermore, the Mantel test showed that the genetic distance matrix calculated from the AFLPs was significantly correlated with the cpSSR-based matrix, and we can infer that the gene flow through pollen within groups was of more importance than the marker-specific factors.

At a more regional geographical scale, significant differences among populations within groups for the H_E parameters were found with the cpSSRs, but not with the AFLPs. The high value of the interloci variance found with the AFLPs could explain the lack of difference among populations for the within-population genetic diversities, because all the estimates for this parameter across the populations lay within the *z*-test-based computed confidence interval. No differences at the within-population diversity level were also found in an allozyme analysis of Portuguese populations (Castro 1989).

Concerning differentiation, the AFLPs exhibited similar levels of differentiation compared with the cpSSRs, for both groups and for all the populations studied, when loci for which the null allele was present at a low frequency were ignored, as recommended by Lynch & Milligan (1994). In general, population subdivision is expected to be lower for nuclear markers than for cytoplasmic markers, but several possible explanations can be proposed for the results obtained in this study. First, one possible explanation is linked with the property of cpSSR markers: higher mutation rates. According to Hedrick (1999) the size of G_{ST} for isolated populations is strongly influenced by the amount of variation determined by the mutation rate. Second, size homoplasy has been observed at two chloroplast microsatellites in *Glicinia* accessions and in higher taxonomic groups. CpSSRs are generated by mutations at a limited number of hotspots, so they are prone to undergo identical mutations independently in different populations (Doyle *et al*. 1998 and references therein), which leads to underestimates of differentiation. Third, extensive gene flow via pollen could explain the similar among-population differentiation values found using nuclear and chloroplastic markers, by smoothing differences due to variations in effective population sizes and genetic drift. Moreover, the data gathered in this study support this hypothesis, as the genetic differentiation among the Portuguese populations estimated from both the AFLP data (after LM restriction) and the cpSSR data was found to be similar to that reported by Castro (1989) using allozyme markers ($G_{ST} = 0.020$) in six populations of *Pinus pinaster* widely spaced across Portugal. The differentiation estimates obtained with the cpSSR and AFLP data (G_2) in this study for the French group were also similar to the estimate obtained by Mariette *et al*. (2001b) in an analysis of nuclear microsatellite markers in the same group of populations. The French populations included in this study also show low differentiation values with both types of markers, providing further strong evidence that extensive gene flow has had an homogenizing effect on the populations. In pines, population subdivision may be found to be weaker when chloroplast DNA (cpDNA) markers are used compared with mitochondrial DNA (mtDNA) markers, because wind-dispersed pollen is the main agent of gene flow (e.g. Dong & Wagner 1994; Mitton *et al*. 2000). Latta & Mitton (1997) observed that the among-population differentiation in *P. flexilis* was much higher for mtDNA than for cpDNA and allozymes, which, in turn, revealed low and similar levels of population differentiation. Moreover, in a study with *P. sylvestris* from Finland (Karhu *et al*. 1996), the allozymes, random amplified polymorphic DNA (RAPDs) and nuclear microsatellites, all showed very little differentiation between southern and northern populations, reflecting the high level of gene flow in the studied area.

The differentiation obtained with all populations considered together (Portuguese and French) was higher than within each provenance separately, and both provenances

could clearly be distinguished with both types of markers (AFLP and cpSSR). The spatial isolation of the two regions, with the Pyrenees mountain range constituting a natural barrier, probably prevented the possibility of the genetic homogenizing effect of gene flow, and allowed the genetic divergence between the two groups of populations due to isolation-by-distance.

Population differentiation and diversity estimates with AFLPs

According to several authors (Lynch & Milligan 1994; Isabel *et al*. 1995) when dominant markers are used to estimate parameters of population genetics, loci with low null allele frequencies should be discarded, and larger numbers of individuals should be sampled per population unless haploid tissues are available to circumvent the dominance problem. For the Portuguese, but not for the French group, a significant difference was found between the G_{ST} estimates computed by the G_1 and G_2 methods. The number of loci pruned by applying the LM restriction was twice as high in the Portuguese group, than the French. The polymorphism found in both groups with both PECs was similar. Therefore, the numbers of loci for which the null allele was present at a low frequency (i.e. in fewer than four null homozygotes, according to the LM restriction) in the Portuguese group were higher. Thus, the differences in the G_{ST} estimates could have been due to the pruning of the loci. The bias introduced in estimating null allele frequencies can be substantial if the null allele is rare, as predicted by Lynch & Milligan (1994). Wu *et al*. (1999) found that applying the LM restriction to RAPD data from three pine species approximately halved the among-population differentiation estimates. In three other studies with conifers, the RAPD phenotypes produced inflated among-population diversity estimates compared with direct RAPD estimates derived from haploid tissue (Isabel *et al*. 1995, 1999; Szmidt *et al*. 1996).

Our estimates for the total diversity (H_T) and the average within population diversity (H_S) of the AFLP data set increased ≈ 1.5-fold using the G_2 compared with the G_1 approach, when all the populations were considered, i.e. the calculated level of diversity was inflated by the exclusion of monomorphic and nearly monomorphic loci. But as only polymorphic loci for both types of markers were used, the comparison of diversity and differentiation measures between AFLP and cpSSR should not be affected by the removal of monomorphic or nearly monomorphic loci. Moreover, Petit *et al*. (1995) after comparing different nuclear makers in *P. pinaster*, showed that differentiation estimates are diversity independent by removing those loci characterized by low values of diversity.

In this study we aim to compare levels of diversity with different types of markers exhibiting different marker-specific properties, such as are AFLPs and cpSSRs, in a set of French and Portuguese populations. Because there is usually a strong asymmetry between pollen and seed flow in conifers, one would expect that pollen flow would be the major factor of population subdivision. Therefore, congruent results with the two types of markers are expected. Indeed, we did find in the French and Portuguese populations similar trends of diversity and differentiation with the AFLP and cpSSR data.

Acknowledgements

We are very grateful for the technical help provided by C Lalanne. Thanks are extended to S González-Martínez for his help in the analysis performed with the SAS system, and to J Bousquet and X-R Wang for critically reading the manuscript. MMR received a PRODEP II — Acção 5.2 fellowship, a SLU grant and a KSLA/Carl Fredrik von Horn's fund (H-389). SM was funded by an INRA/SJFR 2000 Cooperation in Agricultural Research scholarship. This work was supported by the European Union (IC 18970200) and by a grant from Bureau des Ressources Génétiques (BRG), the Swedish Council for Forestry and Agricultural Research (SJFR) and the Swedish International Development Cooperation Agency (SIDA) funded AES. We would like to thank two anonymous reviewers for their helpful comments on an earlier version of this paper.

References

Bahrman N, Zivy M, Baradat P, Damerval C (1994) Organization of the variability of abundant proteins in seven geographical origins of maritime pine (*Pinus pinaster* Ait.). *Theoretical and Applied Genetics*, **88**, 407–411.

Baradat PH, Marpeau-Bezard A (1988) Le pin maritime, *Pinus pinaster* Ait. Biologie et génétique des terpènes pour la connaissance et l'amélioration de l'espèce. PhD Thesis, University of Bordeaux, France.

Castro LFT (1989) Isoenzimas do *Pinus pinaster* Ait. numa prespectiva de aplicação ao melhoramento genético da espécie. PhD Thesis, Universidade de Trás-os-Montes e Alto Douro, Portugal.

Chong DKX, Yang R-C, Yeh FC (1994) Nucleotide divergence between populations of trembling aspen (*Populus tremuloides*) estimated with RAPDs. *Current Genetics*, **26**, 374–376.

Costa P, Pot D, Dubos C *et al.* (2000) A genetic map of maritime pine based on AFLP, RAPD and protein markers. *Theoretical and Applied Genetics*, **100**, 39–48.

Dong JS, Wagner DB (1994) Paternally inherited chloroplast polymorphism in *Pinus*: estimation of diversity and population subdivision, and tests of disequilibrium with a maternally inherited mitochondrial polymorphism. *Genetics*, **136**, 1187–1194.

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

Doyle JJ, Morgante M, Tingey SV, Powell W (1998) Size homoplasy in chloroplast microsatellites of wild perennial relatives of soybean (*Glycine* subgenus *Glycine*). *Molecular Biology and Evolution*, **15**, 215–218.

Ennos R (1994) Estimating the relative rates of pollen and seed migration among plant populations. *Heredity*, **72**, 250–259.

Hedrick RP (1999) Prespective: highly variable loci and their interpretation in evolution and conservation. *Evolution*, **53**, 313–318.

Isabel N, Beaulieu 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. *Proceedings of the National Academy of Sciences of the USA*, **92**, 6369–6373.

Isabel N, Beaulieu J, Theriault P, Bousquet J (1999) Direct evidence for biased gene diversity estimates from dominant random amplified polymorphic DNA (RAPD) fingerprints. *Molecular Ecology*, **8**, 477–483.

Karhu A, Hurme P, Karjalainen M *et al.* (1996) Do molecular markers reflect patterns of differentiation in adaptive traits of conifers? *Theoretical and Applied Genetics*, **93**, 215–221.

Kimura M (1968) Evolutionary rate at the molecular level. *Nature*, **217**, 624–626.

Latta RG, Mitton JB (1997) A comparison of population differentiation across four classes of gene marker in limber pine (*Pinus flexilis* James). *Genetics*, **146**, 1153–1163.

Lefort F, Echt C, Streiff R, Vendramin GG (1999) Microsatellite sequences: a new generation of molecular markers for forest genetics. *Forest Genetics*, **6**, 15–20.

Lerceteau E, Szmidt AE (1999) Properties of AFLP markers in inheritance and genetic diversity studies of *Pinus sylvestris* L. *Heredity*, **82**, 252–260.

Lynch M, Milligan BG (1994) Analysis of population genetic structure with RAPD markers. *Molecular Ecology*, **3**, 91–99.

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

Mariette S, Chagne D, Decroocq S *et al.* (2001a) Microsatellite markers for *Pinus pinaster* Ait. *Annals of Forest Science*, **58**, 203–206.

Mariette S, Chagne D, Lezier C *et al.* (2001b) Genetic diversity within and among *Pinus pinaster* populations: comparison between AFLP and microsatellite markers. *Heredity*, **86**, 469–479.

Mitton JB, Kreiser BR, Rehfeldt GE (2000) Primers designed to amplify a mitochondrial *nad1* intron in ponderosa pine, *Pinus ponderosa*, limber pine, *P. flexilis*, and Scots pine, *P. sylvestris*. *Theoretical and Applied Genetics*, **101**, 1269–1272.

Nei M (1978) Estimation of average heterozygosity and genetic distance from a small number of individuals. *Genetics*, **89**, 583–590.

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

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

Petit RJ, Kremer A, Wagner DB (1993) Finite island model for organelle and nuclear genes in plants. *Heredity*, **71**, 630–641.

Petit RJ, Pons O (1998) Bootstrap variance of diversity and differentiation estimators in a subdivided population. *Heredity*, **80**, 56–61.

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

Ribeiro MM, LeProvost G, Gerber S *et al.* (2002) Origin identification of maritime pine stands in France using chloroplast simple-sequence repeats. *Annals of Forest Science*, **59**, 53–62.

Ribeiro MM, Plomion C, Petit R, Vendramin GG, Szmidt AE (2001) Variation of chloroplast simple-sequence repeats in Portuguese maritime pine (*Pinus pinaster* Ait.). *Theoretical and Applied Genetics*, **102**, 97–103.

Slatkin M (1987) Gene flow and the geographic structure of natural populations. *Science*, **236**, 787–792.

Steel RG, Torrie JH (1981) *Principles and Procedures of Statistics a Biometrical Approach*. McGraw-Hill, Singapore.

Szmidt AE, Wang X-R, Lu M-Z (1996) Empirical assessment of allozyme and RAPD variation in *Pinus sylvestris* L. using haploid tissue analysis. *Heredity*, **76**, 412–420.

Vendramin GG, Anzidei M, Madaghiele A, Bucci G (1998) Distribution of genetic diversity in *Pinus pinaster* Ait. as revealed by chloroplast microsatellites. *Theoretical and Applied Genetics*, **97**, 456–463.

Vendramin GG, Lelli L, Rossi P, Morgante M (1996) A set of primers for the amplification of 20 chloroplast microsatellites in *Pinaceae*. *Molecular Ecology*, **5**, 595–598.

Vos P, Hogers R, Bleeker M *et al.* (1995) AFLP: a new technique for DNA fingerprinting. *Nucleic Acids Research*, **23**, 4407–4414.

Wright S (1943) Isolation by distance. *Genetics*, **28**, 114–138.

Wu J, Krutovskii K, Strauss SH (1999) Nuclear DNA diversity, population differentiation and phylogenetic relationships in the California closed-cone pines based on RAPD and allozyme markers. *Genome*, **42**, 893–908.

This study is part of MM Ribeiro's PhD thesis about the genetic structure of *Pinus pinaster* by using nuclear and cytoplasmic markers. S Mariette compared different molecular markers to assess genetic diversity of different tree species. AE Szmidt is working with the application of molecular approaches to the study of plant evolutionary biology. GG Vendramin is involved in microsatellite marker analysis of many forest tree species. C Plomion is interested in the application of QTL analysis and mapping to operational tree breeding programme. A Kremer's research is focused on the organization and dynamics of gene diversity in forest trees.
