Pollen contamination in a maritime pine polycross seed orchard and certification of improved seeds using chloroplast microsatellites

Christophe Plomion, Grégoire Le Provost, David Pot, Giovanni Giuseppe Vendramin, Sophie Gerber, Stéphane Decroocq, J. Brach, Annie Raffin, Patrick Pastuszka

Pollen contamination in a maritime pine polycross seed orchard and certification of improved seeds using chloroplast microsatellites

C. Plomion, G. LeProvost, D. Pot, G. Vendramin, S. Gerber, S. Decroocq, J. Brach, A. Raffin, and P. Pastuszka

Abstract: A new concept of seed orchard was developed by Institut National de la Recherche Agronomique for the maritime pine (*Pinus pinaster* Ait.) breeding program: the polycross seed orchard (PSO). The expected genetic gain of the PSO can only be reached if the fathers used in the pollen mix contribute equally to the next generation (i.e., to the base material of the PSO) and if pollen contamination from the surrounding stands is limited. Using chloroplast simple-sequence repeats (cpSSR), we showed that the chloroplast DNA was unipaternally inherited in maritime pine and verified that the chloroplast haplotype composition of the megagametophyte tissue corresponded to the chloroplast haplotype of the female parent. As a practical application, a statistical test based on cpSSR markers and simulation was established to verify the PSO origin of maritime pine seed lots. As a result of the cpSSR test, it was observed that (*i*) departure from even distribution of the fathers in the PSO was barely significant, (*ii*) the minimum pollen contamination rate in the PSO was 36%, and (*iii*) the contamination was not evenly distributed in the PSO. As a consequence, the expected genetic gain will range between 50 and 82% of what was initially foreseen.

Résumé : Dans le cadre du programme d'amélioration génétique du pin maritime (*Pinus pinaster* Ait.), un nouveau concept de verger à graines, le verger à graines de familles polycross (VGFP), a été développé par l'Institute National de la Recherche Agronomique. Pour installer un VGFP, des arbres élites sont croisés selon un schéma polycross, puis les descendants sont plantés à densité définitive. La variété améliorée est finalement produite par pollinisation libre entre les arbres du verger. Le gain génétique espéré d'un VGFP ne peut être réalisé que si les parents mâles contribuent également à la génération suivante (le matériel de base du verger), et si la pollution pollinique due aux arbres extérieurs au VGFP est limitée. En utilisant des marqueurs microsatellites chloroplastiques (cpSSR), nous avons tout d'abord montré que le génome chloroplastique présentait une hérédité paternelle. Dans un second temps, nous avons pu vérifier que les mégagamétophytes d'un même arbre possédaient le même haplotype chloroplastique que l'arbre mère. Un test statistique empirique construit par simulation a alors été établi pour décider de l'appartenance ou non d'un lot de graines quelconque au VGFP. Il a permi de montrer que l'écart à l'égale participation des pères dans le VGFP était tout juste significatif, et que le taux de pollution était au minimum de 36% et réparti de façon hétérogène dans le VGFP. En conséquence, le gain génétique espéré sera compris entre 50 et 82% de celui annoncé.

Introduction

Maritime pine (*Pinus pinaster* Ait.) is the first conifer species used for reforestation in France, where it covers about 1.4×10^6 ha. In the early 1960s, the Institut National de la Recherche Agronomique (INRA) initiated a breeding program to improve genetically the Aquitaine provenance (southwestern France) of this species. A recurrent selection scheme was adopted followed by seed orchard establishment (Baradat and Pastuszka 1992). To provide high genetic gains with low financial input for establishment and management, Baradat (1987) developed a new concept of seed orchard: the polycross seed orchard (PSO). This type of wind-pollinated seed orchard constitutes an alternative to the classical clonal seed orchard (CSO). As for a CSO based on elite clones, the best trees (about 30 clones) from the breeding population are selected for their general combining ability estimated in progeny tests. Instead of being grafted at final stocking to constitute an elite CSO, the selected trees (generation G_n) are then mated according to a polycross mating design, and the seedlings (generation $G_n + 1$) are planted at large spacing using a complete random design to establish the PSO. Open pollination among orchard trees produces the improved variety (generation $G_n + 2$). For discussion about the advantages and limitations of the PSO over a classical CSO, the different possible types of polycross mating designs to produce the PSO base material, and the probability of inbred seeds in the improved variety, see Baradat (1987) and Baradat et al. (1992). From 1988 to 1991, 180 ha of PSO were established in southwestern France to produce an improved maritime pine variety (Baradat and Pastuszka 1992), which will provide most of

Received January 17, 2001. Accepted June 11, 2001. Published on the NRC Research Press Web site at http://cjfr.nrc.ca on October 3, 2001.

C. Plomion,[1] G. LeProvost, D. Pot, S. Gerber, S. Decroocq, J. Brach, A. Raffin, and P. Pastuszka. Institut National de la Recherche Agronomique, Equipe de Génétique et Amélioration des Arbres Forestiers, BP45, F-33610 Cestas, France.
G. Vendramin. Istituto Miglioramento Genetico Piante Forestali, Consiglio Nazionale delle Ricerche, Via Atto Vannucci 13, 50134, Firenze, Italy.

[1]Corresponding author (e-mail: plomion@pierroton.inra.fr).

 DOI: 10.1139/cjfr-31-10-1816

the maritime pine improved seeds commercialized in France for the next 15 years. However, the PSO can only reach the expected genetic gains if the different clones are equally represented among the seed orchard trees (i.e., equal contribution of the father clones used in the pollen mix) and, as for any open pollinated seed orchard, if pollen contamination from surrounding stands is limited. In this context, our objectives were (*i*) to evaluate the contribution of each father in the polycross used to produce the PSO base material, (*ii*) to estimate the pollen contamination and study its distribution in the PSO, and (*iii*) to design a DNA-based genetic test to assess the putative PSO origin of any seed lot collected in France. To achieve these goals, we used chloroplast simple-sequence repeats (cpSSR). These markers were developed by Powell et al. (1995) and Vendramin et al. (1996) from the *Pinus thunbergii* chloroplast DNA (cpDNA) sequence (Wakasugi et al. 1994). They were recently used in maritime pine for population genetic studies (Vendramin et al. 1998). As prerequesites for the correct use of these markers for the above-mentioned objectives, the unipaternal inheritance of cpDNA in maritime pine and the presence of plastid DNA in the megagametophyte were also verified.

Materials and methods

Plant material

G_0 tree collection

Thirty-four elite trees were selected (the selection criteria were height and bole straightness) amongst the first maritime pine breeding population (G_0 generation) to establish a PSO. Each elite tree was pollinated by the pollen mix of all, or a subgroup, of the same elite trees in four successive years (1988–1991), and seeds (G_1 generation) were sown to establish the PSO. In this study, only the first set of crosses, which included 28 elite trees was used. Needles from the 28 elite trees were collected in a grafted clonal archive.

Seed orchard description and G_1 trees collection

The PSO was established in southwestern France (Mimizan; 44°08′N, 01°18′W, 35 m altitude). The seedlings were randomly planted at 5 × 5 m spacing in the orchard. Needles or megagametophytes were collected from 320 G_1 trees distributed throughout one single block of the PSO (Fig. 1A). This block was bordered to the north and south by other orchard blocks and to the west and east by mature natural stands of the local maritime pine provenance. Comparison of cpSSR allele frequencies between the G_0 and G_1 populations made it possible to test if the pollen contribution of the G_0 parents to the orchard trees was even. Moreover, knowledge about cpSSR haplotype frequencies in the G_1 population provided us with a null hypothesis to design the varietal test (see Statistical analysis section).

Improved seeds (G_2) collection

In 1999 (corresponding to a pollination in 1997 on 9-year-old trees), 309 viable G_2 seeds were sampled from G_1 trees, before their 11th growing season, to estimate the pollen contamination rate. A first set of 93 G_2 seeds was taken from a bulked seed lot. A second set of 216 G_2 seeds was sampled on 36 G_1 trees (four groups of nine closely located trees, six seeds per tree) along a west–east gradient, parallel to the main wind direction (Fig. 1B). By acting as "pollen traps" the sampling within these four groups was performed with the objective of evaluating whether or not the pollen contamination was evenly distributed in the PSO. Seeds were germinated in vermiculite for 2 weeks (16 h light : 8 h dark; 25:21°C, day:night). After emergence, just before the seed coat was cast off, both the megagametophytes and the embryos were collected and stored at –80°C.

Tested seed lots

As a control population for the varietal test, 50 viable seeds were sampled in a commercial seed lot (called CEMAGREF) originated from selected stands (non improved material) of the northern Aquitaine region (Médoc).

cpSSR analysis

DNA was extracted using the hexadecyltrimethylammonium bromide method of Doyle and Doyle (1990) with some modifications as described by Plomion et al. (1995). In this study, six primers pairs flanking pine chloroplast microsatellites (Pt1254, Pt15169, Pt30204, Pt36480, Pt71936, and Pt87268) were used. They were designed according to sequences in the *Pinus thunbergii* chloroplast genome (Vendramin et al. 1996). These primers were chosen because they detected a relatively high level of polymorphism in an analysis of a subsample of individuals. Polymerase chain reaction (PCR) and electrophoresis were performed according to the protocol described for maritime pine by Ribeiro et al. (2001*a*). The presence of the PCR fragments was visually scored by two persons, with the help of fragment size standards. Whenever the two scores disagreed, the PCR and the electrophoresis were repeated until 100% of scoring agreement was reach.

For ease of presentation, the term "locus" will refer to a cpSSR site and "allele" will refer to a size variant at a given cpSSR site. Assuming that the chloroplast genome does not recombine in maritime pine, it may be viewed as a single locus, and the combination of the detected alleles at each locus, as the different haplotypes. In this study, the scores for the studied loci were combined to derive the chloroplast haplotype of each individual.

Statistical analysis

Testing the equal contribution of the fathers in the PSO

A statistical test was developed to compare the G_0 and G_1 populations based on the six cpSSR haplotype frequencies. Firstly, a statistic, *S*, was computed as follows:

$$[1] \qquad S = \sum_{i=1}^{n} (x_{i\text{ref}} - x_i)^2$$

where *n* is the total number of haplotypes found in the G_0 and G_1 populations, $x_{i\text{ref}}$ is the frequency of the *i*th haplotype in the G_0 reference population, and x_i is the frequency of the *i*th haplotype in the G_1 population. Since nothing was known a priori about the expected distributions of this statistic and because most of the haplotypes were represented by a rather small number of individuals (so the haplotype frequencies were subject to a rather large sampling error), simulations were performed to obtain the distribution of the null hypothesis (H_0: the G_1 population does not depart from the G_0 population, i.e., the contribution of the fathers to the G_1 population is proportional to their frequencies in the G_0 population). Resampling with replacement was performed 10 000 times, and the density function was obtained from the 10 000 S_j values:

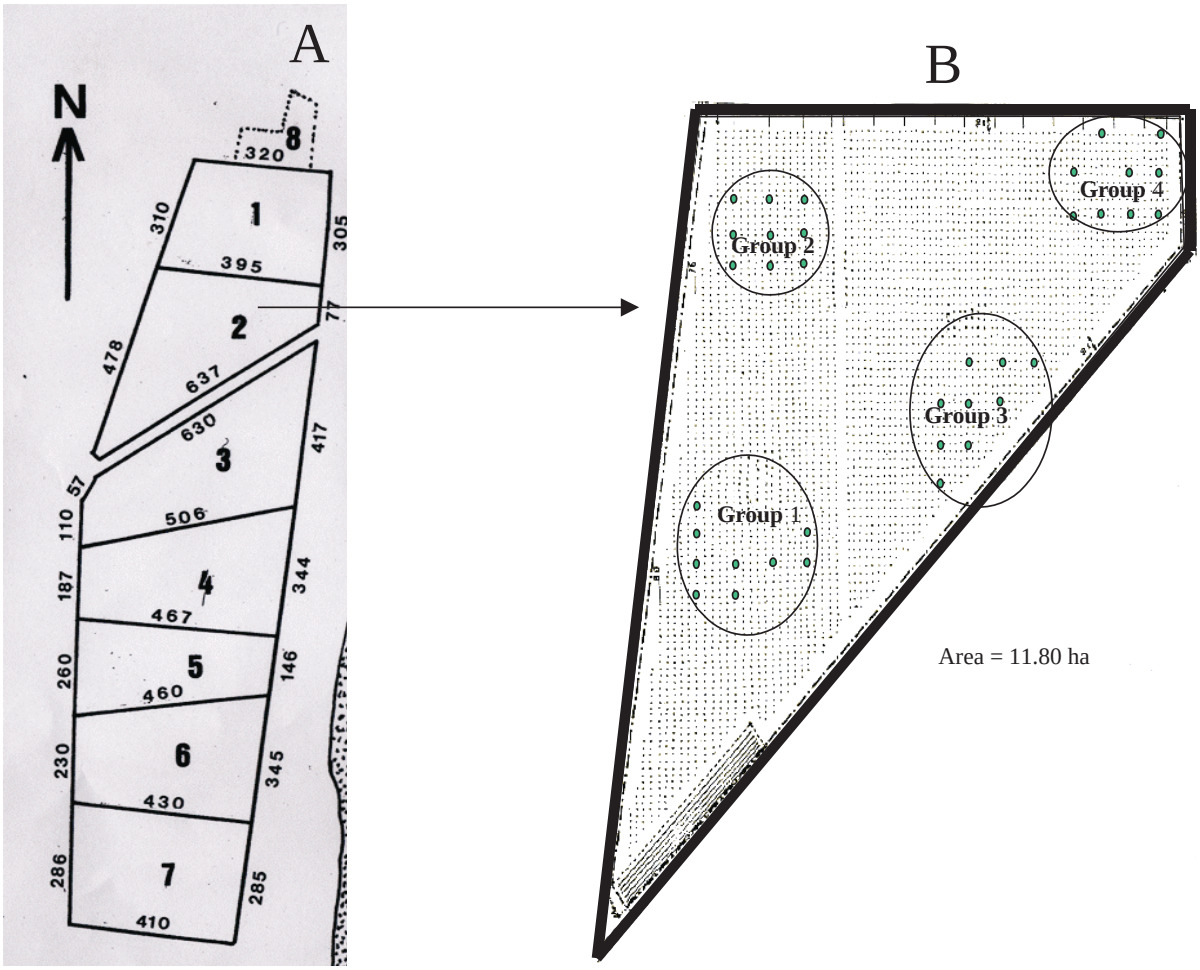
$$[2] \qquad S_j = \sum_{i=1}^{n} (x_{i\text{ref}} - x_{ij})^2$$

where x_{ij} is the frequency of the *i*th haplotype in a sample of 320 trees of the G_0 population under the *j*th outcome ($j = 1 - 10\,000$).

Testing seed lots for PSO origin

The null hypothesis H_0, the seed lot originates from the PSO, was tested against an alternative hypothesis H_1, the seed lot be-

Fig. 1. Polycross seed orchard (PSO) of Mimizan, indicating the studied block (A) and showing the four groups of trees used to study the distribution of pollen contamination (B). Maritime pine is the only forest tree species growing in the surrounding of the PSO.



longs to the Aquitaine provenance. The Aquitaine reference material included 371 trees collected from 13 natural populations along the Atlantic coast (Mariette et al. 2001*a*).

In a first step, the G_1 and Aquitaine populations were analysed at the six cpSSR loci separately. The aim was to identify the loci that could best differentiate between the PSO and the Aquitaine populations. Equation 2 was used to obtain the distribution of the null and the alternative hypotheses, where n is the total number of haplotypes found in both the G_1 and Aquitaine populations, $x_{i\text{ref}}$ is the frequency of the ith haplotype in the G_1 population, and x_{ij} is the frequency of the ith haplotype in a sample from the G_1 population (to obtain H_0) or the Aquitaine population (to obtain H_1) under the jth outcome. Resampling with replacement was performed 10 000 times (j = 1 – 10 000), and the sample sizes were 50, 54, and 93. Afterwards, the density functions were generated, locus by locus, based on the 10 000 S_j values. In a second step, individuals from the G_1 and Aquitaine populations were analysed at the two most informative loci revealed in the first step. The data obtained from the haplotypes derived from the two selected loci and eq. 2 were used to obtain the density functions for the G_1 and Aquitaine population, based on the 10 000 S_j values. The tested sample sizes were the same as before.

The haplotypes of 50 (CEMAGREF seed lot), 93 (bulk of commercialized G_2 seeds), and 54 (four G_2 seed lots geographically localized in the PSO) trees were recorded at the two selected loci. The S statistic was computed for each seed lot to determine their degree of similarity with the G_1 population. Equation 1 was used, where n is the total number of haplotypes found both in the G_1 reference population and in the tested seed lot, $x_{i\text{ref}}$ is the frequency of the ith haplotype in the G_1 reference population and x_i is the frequency of the ith haplotype in the tested seed lot.

Results and discussion

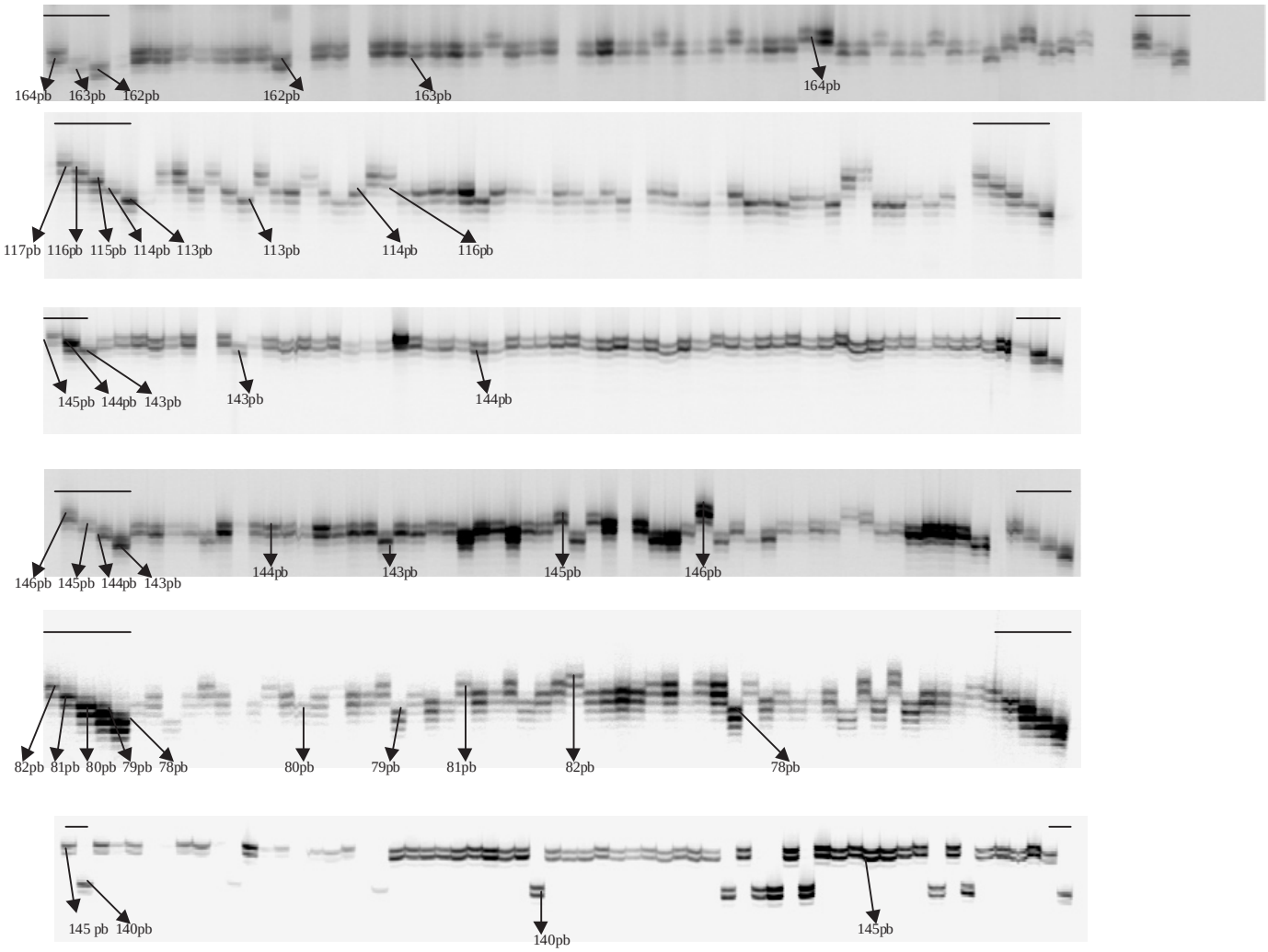
Polymorphism at simple sequence repeat markers in the chloroplast genome of maritime pine

Considering the 28 elite trees analysed in this study, the six cpSSR loci (Pt87268, Pt15169, Pt71936, Pt30204, Pt1254, and Pt36480) exhibited three, four, two, four, four, and two alleles, respectively. The cpSSR profiles can be obtained upon request from the corresponding author. Example of cpSSR profiles are shown in Fig. 2. When all alleles from the 28 trees were combined at the 6 cpSSR loci, 17 different haplotypes were found.

Inheritance of cpDNA in maritime pine

One assumption to meet the goals of this study was the paternal inheritance of the chloroplast genome in maritime pine. To verify the mode of inheritance of cpDNA, we used an intraspecific cross (80 full-sib progeny) obtained by the controlled pollination of two elite trees (accession Nos. 9-106-3 and 101-59-3 used as female and male parents, re-

Fig. 2. The cpSSR profiles at the six studied loci showing from top to bottom: Pt87268, Pt15169, Pt71936, Pt30204, Pt1254, and Pt36480. Fragment size standards are indicated on the left and right sides of each panel.



spectively). Polymorphism between the parents was revealed for two of the six cpSSRs: Pt15169 and Pt36480. All offspring displayed the haplotype of the male parent, indicating for the first time in maritime pine that the cpDNA was paternally inherited. Another three generations inbred pedigree was screened, and the progeny also displayed the haplotype of the male parent (Fig. 3A).

Paternal inheritance of cpDNA in conifers had been studied previously. Although a largely uniparental–paternal mode of inheritance has been confirmed for cpDNA, occasional offspring with maternal or biparental cpDNA genotypes have been observed (reviewed by Cato and Richardson 1996). We used the binomial model of organelle inheritance (Milligan 1992) to evaluate the maximum rate of transmission (T) from the female parent by the following formula: $T = 1 - (1 - P)^{1/N}$, where N is the total number of progeny and P the power of the test. With a β threshold of 5% of falsely accepting the strict uniparental inheritance hypothesis and considering the absence of maternal cpDNA haplotypes among the $N = 80$ offsprings analysed, the maximum rate of maternal transmission of cpDNA that may have overlooked with this sample size is 3.6%. Based on this result, we concluded that the cpDNA genome is mainly paternally inherited in maritime pine, but considering the sample size used, rare events of "maternal leakage" cannot be discarded.

Presence of plastid DNA in the megagametophyte of maritime pine

Another assumption of our study was the occurrence of plastid DNA in the megagametophyte tissue. To verify this assumption, the plastid DNA of megagametophytes was studied. Nine seeds obtained from a self-fertilized individual (accession No. H12) were used. All megagametophytes displayed the haplotype of the mother tree (Fig. 3B). The presence of maternal plastid DNA in the megagametophytes of maritime pine agrees with the results obtained by Wang et al. (1996) and Bucci et al. (1998) in *Pinus* species and by Ziegenhagen et al. (1996) in *Abies* species. Therefore, as a practical application, the analysis of megagametophyte cpDNA facilitates the identification of the genetic origin (seed orchard origin) of any seed lots by using cpDNA markers, avoiding the pollen contamination problem with the embryo analysis. Similar analysis using embryos would require the analysis of mitochondrial DNA (mt DNA; maternally inherited genome in the conifer species tested so far; reviewed by Wang et al. (1996)). However, compared with

Fig. 3. (A) The cpSSR profile at locus Pt15169 showing paternal inheritance of cpDNA. The female parent (lane 2, accession No. L146) differs from the male parent (lane 1, accession No. C10) and the progeny (lane 3, accession No. H12) by one base pair difference. (B) Haplotype identity between maternal and megagametophyte progeny. Megagamethophytes (lanes 4–12) of seeds carried by a mother tree (lane 3, accession No. H12) have the same haplotype. Other lanes are fragment size standards.

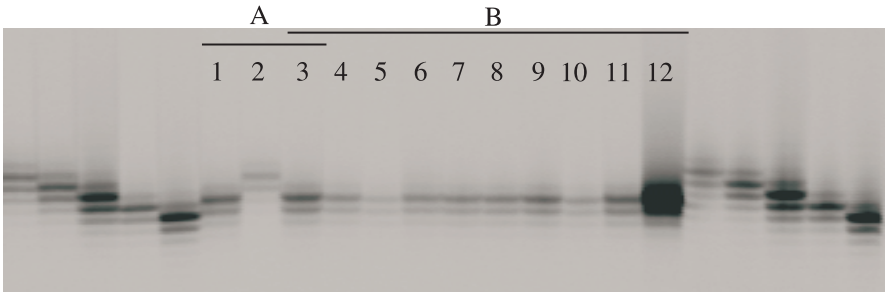
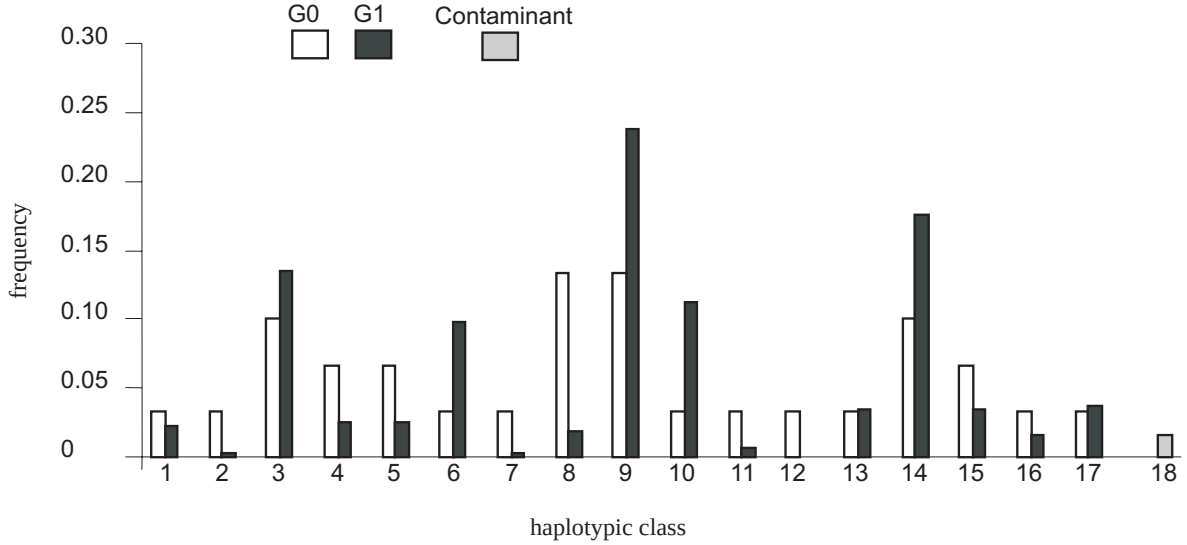


Fig. 4. Comparison of the six cpSSR haplotype distributions in the G_0 and G_1 populations.



cpDNA markers, mtDNA markers are less developed in conifers (Soranzo et al. 2000) and mtDNA in maritime pine reveals low levels of polymorphism (C. Burban, personal communication).

Contribution of the fathers in the seed orchard trees

A total of 17 haplotypes were found among the 28 G_0 trees. Of the 320 G_1 trees analysed, only 5 (1.5%) showed a new haplotype. Three main explanations can account for the presence of contaminants in the orchard trees. Because most alleles differ by only a single base pair, scoring errors cannot be excluded. The use of allele size standards reduces but does not eliminate this type of error resulting from the nature of cpSSR. Natural regenerations in the site where the PSO was established is a second possibility. A third possibility is the contamination by external pollen during the controlled polycrosses. A fourth explanation, less likely to occur, is derived from cpDNA events such as heteroplasmy, hairpin loops creating inversions, and mutational hotspots.

The value of the *S* statistic for the G_1 population allowed to reject the null hypothesis of even contribution of the fathers to the seed orchard trees (Table 1). Genetic and (or) environmental hypothesis can both explain the lack of even contribution of the male parents in the PSO. Firstly, we cannot exclude a problem of pollen viability and (or) competition between pollen coming from different trees during the fertilization, resulting in an over- or under-representation of some fathers in the following generation. Secondly, in the case of the PSO base material tested in this paper, different contributions of father trees could be due to different seed germination rates in the nursery or to seedling survival rates in the PSO.

However, when the distributions of the 17 haplotypic classes were compared between the G_0 and G_1 populations (Fig. 4) the following observations could be made: (*i*) all but one class (No.12) were present in the G_1 population; (*ii*) the most frequent classes in the G_0 population (Nos. 3, 8, 9, and 14) were also the most frequent in the G_1 population except for one (No. 8); and (*iii*) the absolute difference of haplotype frequencies between both populations was at most 12%. In addition, when simulations were performed for each locus separately, only locus Pt30204 showed departure from even paternal contribution (Table 1) without any obvious explanation to this. Therefore, the conclusion based on the statistical test should be taken with caution. In fact, the test is based on a rather limited sample size (320 trees), and even a small difference in the haplotype frequencies can be significant.

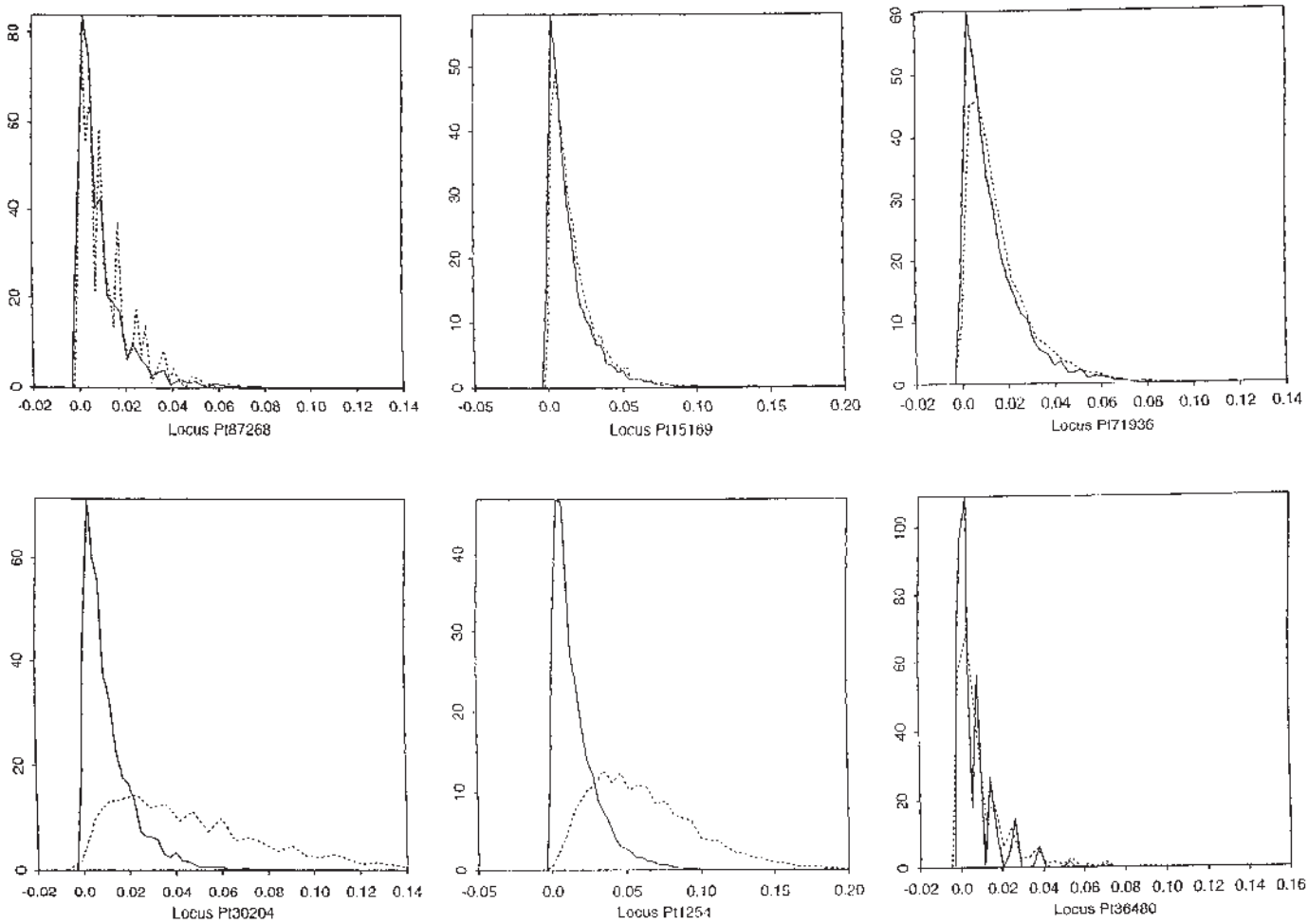
The paternal contribution in a controlled polycross was investigated in different forest tree species using genetic markers. Differential male reproductive success was demonstrated

Table 1. Summary of the statistical tests.

H_0[a]	H_1[b]	Locus[c]	Value of the *S* statistic[d]	Critical value[e]	Sample size[f]	α[g] (%)	β[h] (%)	Decision
Equiproportional contribution of the fathers	—	Pt87268, Pt15169, Pt71936, Pt30204, Pt1254, and Pt36480	$S_{G_1} = 0.05$	$S = 0.0012$ and 0.0061 (two-sided test)	320	1	—	H_0 rejected
	—	Pt30204	$S_{G_1} = 0.012$	$S = 0.000085$ and 0.0063 (two-sided test)	320	1	—	H_0 rejected
CEMAGREF seed lot is of PSO origin	The CEMAGREF seed lot is of Aquitaine origin	Pt30204, Pt1254	$S_{CEMAGREF} = 0.093$	$S_c = 0.037$	50	8	8.4	H_0 rejected
Bulked seed lot of G_2 seeds is of PSO origin	The bulked seed lot of G_2 seeds is of Aquitaine origin	Pt30204, Pt1254	$S_{G_2} = 0.033$	$S_c = 0.026$	93	1.1	1.6	H_0 rejected
Group No. 1 is of PSO origin	Group No. 1 is of Aquitaine origin	Pt30204, Pt1254	$S_{group1} = 0.028$	$S_c = 0.032$	54	4.6	5.7	H_0 accepted
Group No. 2 is of PSO origin	Group No. 2 is of Aquitaine origin	Pt30204, Pt1254	$S_{group2} = 0.048$	$S_c = 0.032$	54	4.6	5.7	H_0 rejected
Group No. 3 is of PSO origin	Group No. 3 is of Aquitaine origin	Pt30204, Pt1254	$S_{group3} = 0.15$	$S_c = 0.032$	54	4.6	5.7	H_0 rejected
Group No. 4 is of PSO origin	Group No. 4 is of Aquitaine origin	Pt30204, Pt1254	$S_{group4} = 0.044$	$S_c = 0.032$	54	4.6	5.7	H_0 rejected

[a]Null hypothesis.
[b]Alternative hypothesis.
[c]Locus used in the test.
[d]Value of the *S* statistic obtained using eq. 1 from the Materials and methods.
[e]For the two-sided test the critical value corresponds to a type I error of 1%, for a one-sided test, the critical value is the *S* value that minimizes both type I and II error rates.
[f]Number of trees genotypes with the loci indicated in the second column.
[g]Type I error risk associated with the test.
[h]Type II error risk associated with the test.

Fig. 5. Density functions for the G_1 and the Aquitaine populations (solid and broken lines, respectively) for each locus. The data points were obtained after 10 000 simulations. Sample size was $N = 93$.



for *Picea mariana* (Rogers and Boyle 1991), *Pseudotsuga menziesii* (Apsit et al. 1989), *Picea abies* (Schoen and Cheliak 1987), *Pinus radiata* (Moran and Griffin 1985), and *Picea glauca* (Schoen and Stewart 1986), but equal mating hypothesis was verified for *Pinus taeda* (Wiselogel and van Buijtenen 1988). In the present study, the consequence of nonequal paternal contribution in the controlled polycross is a bias in the estimation of the genetic gain that should be realized by improved seeds produced by the PSO and a reduction in the genetic diversity relative to the diversity assumed by the number of selected parents of the PSO. However, the bias in genetic gain is likely to be negligible, considering the equivalence of breeding values of the different parents in the polycross composition, compared with nonimproved material. Likewise, the reduction of diversity should be low; Baradat (1987) showed that even a reduction of 50% of the number of fathers in the polycross would have negligible effects on the proportion of noninbred seeds produced in the PSO, provided that the number of mothers is around 10, which is verified in our case (28 mother trees).

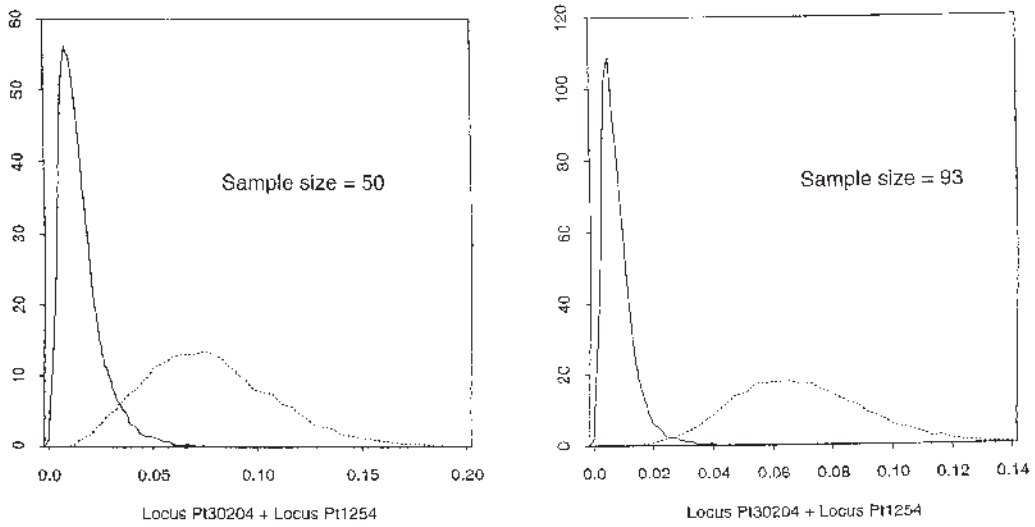
Genetic identification of commercialized seed lots

In a first screening step, the density functions for the S_j statistics were obtained for the six cpSSR loci separately (Fig. 5). Four loci (Pt87268, Pt15169, Pt71936, and Pt36480) showed overlapping density functions for the G_1 and Aquitaine populations, whereas two loci (Pt30204 and Pt1254) were discriminant. In a second step, the haplotypes were combined at these two most informative loci. Figure 6 shows the density functions for the null and alternative hypotheses based on the two selected loci, for tested sample sizes (50 and 93) in the resampling procedure. The critical value S_c that minimizes both type I (α) and type II (β) error rates is located in the intersection of the distributions defining the null and the alternative hypothesis. The threshold values are given in Table 1 for the different tested seed lots. Whenever the S value of the tested seed lot was found to be smaller than S_c, it was supposed of PSO origin. Using this criterion, the following results were obtained:

(1) The CEMAGREF seed lot was unambiguously rejected from being of PSO origin. The S value was perfectly located in the center of the Aquitaine distribution (Fig. 6, left panel).

(2) The bulked G_2 seed lot (93 seedlings) had to be rejected, although the S value was close to the critical value (Fig. 6, right panel). Based on the six-locus haplotype frequencies, the 93 G_2 seedlings were distributed in 34 haplotypic classes, 10 of which were common to the G_0 haplotypes (representing 69% of the seedlings); however, 24 corresponded to new haplotypes

Fig. 6. Density functions for the G_1 and Aquitaine reference populations (solid and broken lines, respectively) for the two most discriminant loci combined (Pt30204 and Pt1254). The data points were obtained after 10 000 simulations. Sample sizes were $N = 50$ (left panel) and $N = 93$ (right panel).



represented by one or two seedlings. The new haplotypes found in the G_2 seed lot represented 31% of the individuals screened and were most probably the results of external pollen contamination. New alleles were not identified in this seed lot.

(3) Of the four geographically localized groups of G_2 seeds, only one (group No. 1) was accepted as being of PSO origin; the three others had to be rejected, although the S values were close to S_c for group Nos. 2 and 4. Interestingly, the S value of group No. 3 did not even allow to accept this group as being of Aquitaine origin (alternative hypothesis). The percentage of six-locus G_0 haplotypes was 68, 59, 24, and 81% for group Nos. 1–4, respectively, showing that the pollen contamination was unequally distributed in the PSO. This result allowed to draw some hypotheses regarding the distribution of pollen contamination in the PSO (see next section). In contrast to the bulked G_2 seed lot, three new alleles (not present in the G_0 population) were found: two at locus Pt1254 and one at locus Pt15169.

Overall, these results demonstrate that pollen contamination has to be taken into account in any statistical analysis aiming at testing the orchard origin of any seed lot.

Estimation of pollen contamination and loss of genetic gain

Taking the 309 G_2 seeds into account, the average "visible" contamination rate based on the six-loci haplotype was 36.5%. This value should be viewed as a minimum estimate of pollen contamination, because some contaminants are likely to have haplotypes indistinguishable from those produced by the surrounding natural stands. Multiallelic highly polymorphic markers such as nuclear SSR would be required to estimate the cryptic gene flow (Streiff et al. 1998) and to develop a paternity exclusion test. Such markers are under development in maritime pine (Mariette et al. 2001*b*). This result shows that, without spatial isolation from nonorchard pollen sources, levels of pollen contamination are high in juvenile maritime pine seed orchards.

Several studies using genetic markers have demonstrated that a large proportion of a young orchard's seed crop is fertilized by unselected contaminant sources (Stoehr et al. 1998; Harju and Nikkanen 1996; Paule et al. 1993; Wheeler and Jech 1992; Yazdani and Lindgren 1991). It is also the case in mature seed orchards (reviewed by Adams et al. 1997). Recommendations to limit pollen contamination have been proposed (reviewed by Caron and Leblanc 1992), including alteration of phenological patterns in orchard by water spray, isolation barriers, windbreaks, and supplemental mass pollination.

A high level of pollen contamination is a major problem in conifer seed orchards. If seed-orchard trees are pollinated by surrounding unselected trees, the consequence is a decrease in the genetic gain of the commercialized variety, with a maximum of 50% if all flowers are fertilized by pollen from nonimproved trees. In addition, the higher the difference between the genetic values of the seed orchard and its surrounding stands, the higher the loss in genetic gain in absolute value. In the case of the PSO, expected genetic gains for selected traits were estimated at 35% compared with nonselected trees in the absence of pollen contamination. A minimum contamination rate of 36.5% corresponds to a minimum decrease of genetic gain of 18.25%. The actual maximum genetic gain will then be $\Delta g = \Delta G_0^*(1 - C/2) \approx 28.3\%$, where C is the minimum pollen contamination rate and ΔG_0 is the expected genetic gain in the absence of pollen contamination.

Distribution of pollen contamination

According to their level of pollen contamination, the four groups of geographically localized G_1 trees could be ranked from the less to the most polluted ones as follows: group No. 4 (19%) < No. 1 (32%) < No. 2 (41%) < No. 3 (76%). To understand such heterogeneous results in contamination rate, both the proximity of the groups from the adult stands surrounding the PSO and the main wind direction (westward) had to be considered. Group No. 3 was located in the center of the PSO and was the most polluted group, group

Nos. 1 and 2 (probably too close to surrounding adult stands) and No. 4 (probably too far from the adult stands) tend to be less polluted. The results reported here are in agreement with those of Castaing and Vergeron (1973) in another maritime pine seed orchard. They estimated that the pollen was dispersed according to a parabolic curve from the receptor tree, with an optimum dispersal distance of 500 m. Group No. 3 was indeed located at a distance of about 400 m from the border of the most probable source of contamination considering the main wind direction (westward). Conversely, in a similar size seed orchard, Yazdani and Lindgren (1991) showed that the contamination was higher in the corner of the seed orchard compared with that in the center, although the reduction in the center was low. Pollen movements in a wind-pollinated seed orchard are complex and unpredictable (Di-Giovanni and Kevan 1991) depending largely upon meteorological conditions (mainly wind speed and wind direction). It is, therefore, difficult to compare the results. What is known for sure is that a large fraction of the seeds harvested in conifer seed orchards are fertilized by trees more than 100 m away (Friedman and Adams 1985; Wang et al. 1991; Di-Giovanni and Kevan 1991). This means that reduction in contamination cannot be achieved by collecting preferentially seeds from trees in the middle or borders of orchards. Instead, isolation from any sources of contamination may well be justified for the next maritime pine seed orchards generation.

Conclusion

Chloroplast DNA microsatellites have the potential to be the first cost-effective genetic marker system for supporting management objectives in the maritime pine breeding program. In previous reports we demonstrated its use for (*i*) the description of the genetic diversity and the study of its structuration (Ribeiro et al. 2001*a*) and (*ii*) origin identification of maritime pine stands (Ribeiro et al. 2001*b*). In this study we showed that these markers can be used to routinely estimate pollen contamination in seed orchards and to monitor the genetic integrity of plants propagated by seeds. It is also worth noting that, in a conifer seed, the cpSSR marker haplotype for the female parent can be determined from a DNA sample extracted from the megagametophyte, and the haplotype from the male parent can be determined from the embryo or seedling. Therefore, cpSSR haplotypes can be used to determine both the paternal and maternal parentage and ultimately to verify the parentage of controled crosses.

Acknowledgments

This work was supported by grants from Ministère de l'Agriculture et de la Pêche – Direction de l'Espace Rural et de la Forêt (No. 01.40.07/99 and No. 01.40.26/99) and the European Union (IC18-97-02-00). We thank Maria Ribeiro for valuable comments and critical reading of the manuscript.

References

Adams, W.T., Hipkins, V.D., Burczyk, J., and Randall, W.K. 1997. Pollen contamination trends in a maturing Douglas-fir seed orchard. Can. J. For. Res. **27**: 131–134.

Apsit, V.J., Nakamura, R.R., and Wheeler, N.C. 1989. Differential male reproductive success in Douglar-fir. Theor. Appl. Genet. **77**: 681–684.

Baradat, P. 1987. Méthode d'evalution de la consanguinité chez les plants issus des vergers à graines de semis de première génération. Silvae Genet. **36**: 134–144.

Baradat, P., and Pastuszka, P. 1992. Le pin maritime. *In* Amélioration des espèces végétales cultivées. *Edited by* A. Gallais and H. Bannerot. Institut National de la Recherche Agronomique, Paris. pp. 695–709.

Baradat, P., Durel, C.-E., and Pastuszka, P. 1992. The polycross seed orchard: an original concept. *In* Proceeding of the IUFRO–AFOCEL Symposium, Mass Production Technology for Genetically Improved Fast Growing Forest Tree Species. Tome II. Editions AFOCEL, Bordeaux, France. pp. 53–62.

Bucci, G., Anzidei, M., Madaghiele, A., Vendramin, G.G. 1998. Detection of haplotypic variation and natural hybridization in *halepensis*-complex pine species using chloroplast simple sequence repeat (SSR) markers. Mol. Ecol. **12**: 1633–1645.

Caron, G.E., and Leblanc, R. 1992. Pollen contamination in a small black spruce seedling seed orchard for three consecutive years. For. Ecol. Manage. **53**: 245–261.

Castaing, J.-P., and Vergeron, P. 1973. Principes et méthodes d'études expérimentale de la dispersion du pollen de pin maritime dans le Massif Landais. Pollen et spores. Muséum national d'histoire naturelle, Paris. Publ. 15. pp. 255–280.

Cato, S.A., and Richardson, T.E. 1996. Inter- and intraspecefic polymorphism at chloroplast SSR loci and the inheritance of plastids in *Pinus radiata* D. Don. Theor. Appl. Genet. **93**: 587–592.

Di-Giovani, F., and Kevan, P.G. 1991. Factors affecting pollen dynamics and its importance to pollen contamination: a review. Can. J. For. Res. **21**: 1155–1170.

Doyle, J.J., and Doyle, J.L. 1990. Isolation of plant DNA from fresh tissue. Focus, **12**: 13–15.

Friedman, S.T., and Adams, W.T. 1985. Estimation of gene flow into two seed orchards of loblolly pine (*Pinus taeda* L.) Theor. Appl. Genet. **69**: 609–615.

Harju, M., and Nikkanen, T. 1996. Reproductive success of orchard and nonorchard pollens during different stages of pollen shedding in a Scots pine seed orchard. Can. J. For. Res. **26**: 1096–1102.

Mariette, S., Chagne, D., Lézier, C., Pastuszka, P., Raffin, A., Plomion, C., and Kremer, A. 2001*a*. Genetic diversity within and among *Pinus pinaster* populations: comparison between AFLP and microsatellite markers. Heredity, **86**: 469–479.

Mariette, S., Chagne, D., Decroocq, S., Vendramin, G.G., Lalanne, C., Madur, D., and Plomion, C. 2001*b*. Microsatellite markers for *Pinus pinaster* Ait. Ann. Sci. For. **58**: 203–206.

Milligan, M.G. 1992. Is organelle DNA strictly maternally inherited? Power analysis of a binomial distribution. Am. J. Bot. **79**: 1325–1328.

Moran, G.F., and Griffin, A.R. 1985. Non-random contribution of pollen in polycrosses of *Pinus radiata*. D. Don. Silvae Genet. **34**: 117–121.

Paule, L., Lindgren, D., and Yazdani, R. 1993. Allozyme frequencies, outcrossing rate and pollen contamination in *Picea abies* seed orchards. Scand. J. For. Res. **8**: 8–17.

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

Powell, W., Morgante, M., Mcdevitt, R., Vendramin, G.G., and Rafalski, J.A. 1995. Polymorphic simple sequence repeat regions in chloroplast genome: applications to the population genetics of pine. Proc. Natl. Acad. Sci. U.S.A. **92**: 7759–7763.

Ribeiro, M.M., Plomion, C., Petit, R., Vendramin, G.G., and Szmidt, A.E. 2001*a*. Variation of chloroplast single-sequence repeats in Portuguese maritime pine (*Pinus pinaster* Ait.) Theor. Appl. Genet. **102**: 97–103.

Ribeiro, M.M., LeProvost, G., Gerber, S., Vendramin, G.G., Anzidei, M., Decroocq, S., Marpeau, A., Mariette, S., and Plomion, C. 2001*b*. Origin identification of maritime pine stands in France using chloroplast single-sequence repeats. Ann. For. Sci. In press.

Rogers, D.L., and Boyle, T.J.B. 1991. Unequal paternal contribution in black spruce polycross seedlots. Heredity, **67**: 373–379.

Schoen, D.J., and Cheliak, W.M. 1987. Genetics of the polycross. 2. Male fertility variation in Norway Spruce, *Picea abies* (L.) Karst. Theor. Appl. Genet. **74**: 554–559.

Schoen, D.J., and Stewart, S.C. 1986. Variation in male reproductive investment and male reproductive success in white spruce. Evolution, **40**: 1109–1120.

Soranzo, N., Alia, R., and Powell, W. 2000. Patterns of variation at a mitochondrial sequence-tagged-site locus provides new insights into the postglacial history of European *Pinus sylvestris* populations. Mol. Ecol. **9**: 1205–2111.

Stoehr, M.U., Orvar, B.L., Vo, T.M., Gawley, J.R., Webber, J.E., and Newton, C.H. 1998. Application of a chloroplast DNA marker in seed orchard management evaluations of Douglas-fir. Can. J. For. Res. **28**: 187–195.

Streiff, R., Labbe, T., Bacilieri, R., Steinkellner, H., Gloessl, J., and Kremer, A. 1998. Within population genetic structure in *Quercus robur* L., and *Quercus petraea* (Matt.) Liebl. assessed with isozymes and microsatellites. Mol. Ecol. **7**: 317–328.

Vendramin, G.G., Lelli, L., Rossi, P., and Morgante, M. 1996. A set of primers for the amplification of 20 chloroplast microsatellites in *Pinaceae*. Mol. Ecol. **5**: 595–598.

Vendramin, G.G., Anzidei, M., Madaghiele, A., and Bucci, G. 1998. Distribution of genetic diversity in *Pinus pinaster* Ait. as revealed by chloroplast microsatellites. Theor. Appl. Genet. **97**: 456–463.

Wakasugi, T., Tsudzuki, J., Ito, S., Nakashima, K., Tsudzuki, T., and Sugiura, M. 1994. Loss of all *ndh* genes as determined by sequencing the entire chloroplast genome of the black pine *Pinus thunbergii*. Proc. Natl. Acad. Sci. U.S.A. **91**: 9794–9798.

Wang, X., Lindgren, D., Szmidt, A., and Yazdani, R. 1991. Pollen migration into a seed orchard of *Pinus sylvestris* L. and the efficiency of its estimation using allozyme markers. Scand. J. For. Res. **6**: 379–385.

Wang, X.R., Szmidt, A.E., and Lu, M.Z. 1996. Genetic evidence for the presence of cytoplasmic DNA in pollen and megagametophytes and maternal inheritance of mitochondrial DNA in *Pinus*. For. Genet. **3**: 37–44.

Wheeler, N., and Jech, K. 1992. The use of electrophoretic markers in seed orchard research. New. For. **6**: 311–328.

Wiselogel, A.E., and van Buijtenen, J.P. 1988. Probability of equal mating in polymix pollinations of loblolly pine (*Pinus taeda* L.) Silvae Genet. **37**: 184–187.

Yazdani, R., and Lindgren, D. 1991. Variation of pollen contamination in a scot pine seed orchard. Silvae Genet. **40**: 243–358.

Ziegenhagen, B., Schauerte, M., Kormutak, A., and Scholz., F. 1996. Plastid DNA polymorphism of megagametophytes and pollen in two *Abies* species. Silvae Genet. **45**: 355–358.