Ann. For. Sci. 65 (2008) 706
DOI: 10.1051/forest:2008049

Available online at:
www.afs-journal.org

Original article

Genetic variation, mating patterns and gene flow in a *Pinus pinaster* Aiton clonal seed orchard

Lúcia FERNANDES[1], Margarida ROCHETA[1,6], Jorge CORDEIRO[1], Sandra PEREIRA[1], Sophie GERBER[2,3], Maria M. OLIVEIRA[1,4], Maria M. RIBEIRO[1,5]*

[1] IBET/ITQB, Quinta do Marquês, 2784-505 Oeiras, Portugal
[2] INRA – UMR BIOGECO 69 route d'Arcachon 33612 Cestas Cedex, France
[3] Université de Bordeaux 1, UMR1202 BIOGECO, 69 route d'Arcachon, F-33612 Cestas, France
[4] Universidade de Lisboa, Faculdade de Ciências, Campo Grande, 1749-016, Lisbon, Portugal
[5] Departamento de Silvicultura e Recursos Naturais, Escola Superior Agrária de Castelo Branco, Quinta da Senhora de Mércules, 6001-909 Castelo Branco, Portugal
[6] Present address: Centro de Botânica Aplicada à Agricultura, Secção de Genética, Instituto Superior de Agronomia (ISA), Techn. Univ. Lisbon, Portugal

(Received 23 November 2007; revised version 15 April 2008; accepted 16 May 2008)

Keywords:
genetic variation /
mating system /
gene flow /
simple sequence repeats /
maritime pine

Abstract

• Relatedness among parents, variation in clonal fertility and background pollination deviate the realized genetic gain and the gene diversity of open pollinated seed orchard from expectation, in particular in wind pollinated species such as *Pinus pinaster* Aiton.

• This work investigates the genetic variation, the mating system and the pollen contamination in a *P. pinaster* clonal seed orchard (CSO), by screening the 60 clones from the CSO and the seeds collected from 21 mother-trees with three nuclear microsatellites.

• The expected diversity was similar, but the observed heterozygosity decreased 20% in the progenies compared with the parental trees. The outcrossing rate was 90.1%, the biparental inbreeding 21.7% computed through a multilocus approach, and the observed selfing 3.9%. The observed gene flow from outside the CSO was 52.4%.

• From the results we concluded that the observed gene flow and the biparental inbreeding were high, and care should be taken in the implementation and management of future CSO, in particular clones should be checked for relatedness and the ramet number could be directly proportional to their breeding value.

Mots-clés :
variation génétique /
système de reproduction /
flux de gènes /
microsatellites /
pin maritime

Résumé – Variation génétique, système de reproduction et flux de gènes dans un verger à graines de clones de *Pinus pinaster* Aiton.

• L'apparentement entre parents, la variation de fertilité entre clones et la pollution pollinique font dévier le gain génétique réalisé et la diversité génétique des valeurs attendues en verger à graines de clones à pollinisation libre, en particulier pour une espèce à pollinisation anémophile comme *P*inus pinaster Aiton.

• Ce travail étudie la variation génétique, le système de reproduction et la contamination du pollen d'un verger à graines de clones (VGC) de *P. pinaster*, en génotypant les 60 clones du VGC et les graines récoltées sur 21 arbres-mères pour trois marqueurs microsatellites nucléaires.

• La diversité attendue s'est révélée similaire mais l'hétérozygotie observée a diminué de 20 % chez les descendants par rapport aux parents. Le taux d'allofécondation était de 90,1 %, la consanguinité biparentale calculée grâce à une approche multilocus était de 21,7 % et le taux d'autofécondation observé était de 52,4 %.

• D'après ces résultats, nous avons conclu que les flux de gènes observés et la consanguinité biparentale étaient élevés et que des précautions devaient être prises dans la mise en place et la gestion de futurs VGC, l'apparentement entre clones devrait en particulier être vérifié et le nombre de copies d'un individu pourrait être directement proportionnel à sa valeur génétique.

* Corresponding author: mribeiro@esa.ipcb.pt

1. INTRODUCTION

Seed orchards represent the link between breeding programs and reforestation activities through the consistent delivery of genetically improved material, with their genetic potential being dependent on several factors. The breeding population should reflect the diversity of the original population and be large enough to maintain diversity for many ensuing generations. The trees should be unrelated, and there should be neither flowering asynchronization nor pollen contamination problems. Furthermore, domestication of forest trees may lead to genetic erosion in several steps of the breeding cycle (phenotypic selection, breeding, seed and seedling production), and systematic monitoring of those various steps should be kept in mind. Moreover, parental contribution to the resultant seed crops is consistently asymmetrical, and this is a major cause of genetic erosion, due to the reduction of the genetic base (El-Kassaby, 1995).

Several studies concluded that considerable pollen contamination occurs in *Pinus* spp. seed orchards, with values ranging from 2% in *P. sylvestris* to 84% in *P. ellioiii* (Kang, 2001). High levels of pollen contamination constitute a serious problem because seed resulting from fertilization by alien pollen is expected to have lower genetic gain than seed fertilized by orchard pollen. For instance, in a *P. pinaster* polycross seed orchard (constituted by the seedlings derived from the pollination of elite trees with a pollen cocktail – a mix of pollen coming from different elite trees), 36% of pollen contamination was observed, and the estimated decrease in genetic gain was greater than to 50% (Plomion et al., 2001). High levels of background pollination are observed even when the stands are isolated by several hundreds of meters from natural populations (Adams and Burczyk, 2000), thus contamination levels should be monitored in open pollinated orchards. Recently, molecular markers have been used to trace details of reproductive processes. Most studies published to date used allozyme markers for mating system, gene flow and pollen contamination estimation (reviewed in Wheeler and Jech, 1992). In particular, mating patterns were studied using allozymes in conifers natural stands (Adams and Birkes, 1991; Burczyk et al., 1996), in an angiosperm's seed orchard (e.g. Burczyk et al., 2002) and in a conifer clonal archives for ex situ conservation programs (Burczyk et al., 2004b). The efficiency and utility of estimating mating system, gene flow and pollen contamination are greatly increased by the use of microsatellite markers, ideal for parentage studies and fingerprinting because of their high discrimination power (e.g. Chaix et al., 2003; Gerber et al., 2000; Moriguchi et al., 2004; 2007; Plomion et al., 2001; Slavov et al., 2005a; 2005b).

Pinus pinaster Aiton is one of the most important conifer species in the southwestern countries of Europe, particularly in Portugal, where it occupies the largest area as compared with the area of the country (Ribeiro et al., 2001). In Portugal, a breeding program was started during the 50s and the Leiria provenance was chosen for plant material selection in establishing a clonal seed orchard (CSO) as field trials demonstrated that this provenance was much superior to others (Perry, 1940). This species is wind-pollinated and has known problems of flowering phenology asymmetry, observed in another *P. pinaster* CSO study (Varela, 1989).

In the present study, three SSR loci were used to screen samples from the *P. pinaster* CSO, with the following specific objectives:

(i) to determine the average selfing, single locus and multilocus outcrossing rate;

(ii) to estimate the existence of biparental inbreeding and the outside CSO gene flow (GFO);

(iii) to assign paternity by using embryos collected in three plots along the main wind direction so as to verify if the GFO is windward oriented;

(iv) to compute the effective number of clones;

(v) to compare parents' and progenies' genetic diversity.

2. MATERIALS AND METHODS

2.1. Study site and plant material

The selected clonal seed orchard is located in Escaroupim National Forest (latitude 39° 4' 33.76" N and longitude 8° 44' 38.77" O) and was established by grafting between 1970 and 1975, surrounded by *Eucalyptus globulus* Labill. and *Pinus pinea* L. stands. There are scattered *P. pinaster* trees within the orchard's isolation area surrounding the CSO, and a stand of the same species in less than 2 km.

The initial spacing between trees was 4 × 4 m in about a 4 ha area. The material for the CSO establishment was initially selected in the Leiria provenance in Portugal, following phenotypic evaluation (volume, height, bole straightness, branching habits and spiral grain). There are 728 ramets (a vegetatively reproduced copy of a plant, in principle with the same genotype as the original parent tree) in the CSO, and they were distributed using three different incomplete random bloc designs, repeated 13 times each, and avoiding close vicinity of ramets from the same clone (Fig. 1). The ramets were based on 60 clones and the number of ramets per clone is uneven, ranging from five to 23 for clones 4 and 41, respectively. Three plots were selected in three different positions in the CSO, along the main wind direction (NW), and, in the same plot, seven trees were selected as mother-trees, afterwards two cones were collected from each one of the mother-trees in opposite sides of the crown (see location of plots and mother-trees in Fig. 1). For each mother-tree the seeds from the two collected cones were pooled together and mixed, and then 10 seeds were sampled, except for that one mother-tree, for which only six seeds were sampled. Needles were collected from each one of the 60 clones in the ramets, randomly chosen inside the seed orchard, and its location is marked in Figure 1 (cells with grey background). We decreased the number of seeds sampled per tree to minimize seed relatedness, and to increase the number of clones as maternal trees, as well as the number of mother-trees sampled per plot, so as to boost the accuracy of the among plots gene flow comparison.

2.2. DNA extraction, PCR amplification and electrophoresis

The needles from the 60 genotypes and the embryos from the 206 progenies were used for DNA extraction following the protocol

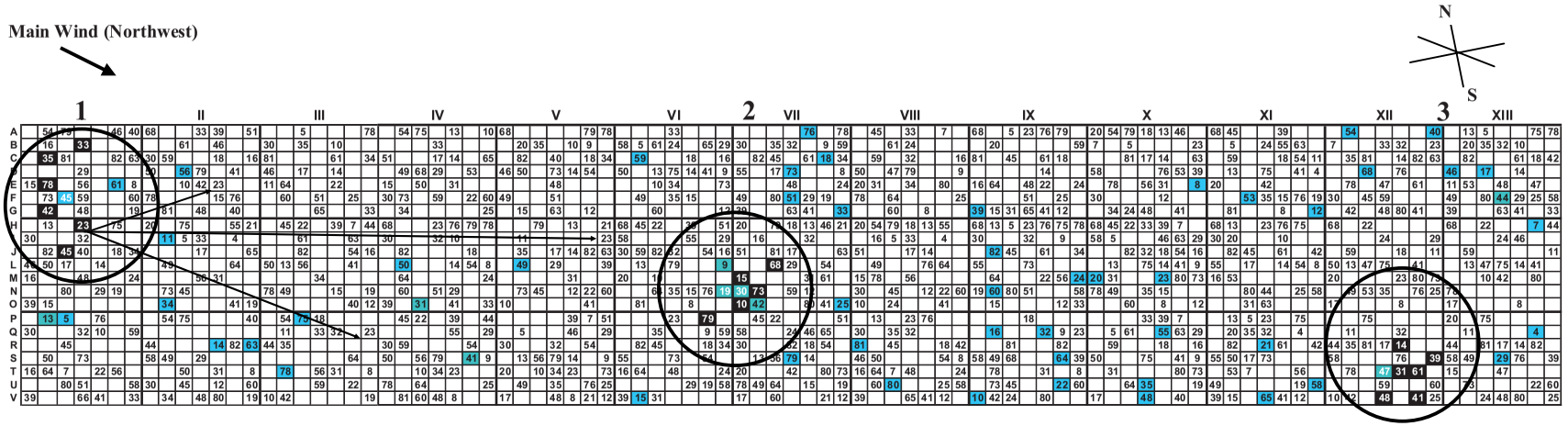

Figure 1. Representation of the four ha CSO with the plots and trees location. Trees location are indicated by Latin numerals (I to XIII) and capital letters (A to V). There are 13 blocs (I to XIII) and 3 repetitions (A-G, H-O and P-V) in the CSO. The circles represent the plots location according to the main wind direction, from 1 to 3. The trees were numbered and the same number represents a ramet of the same clone. The trees where the progenies were picked were numbered in white in a black background, seven per plot (inside the circles), and in all the other trees numbered in black in a grey background (60) the leaves were sampled for DNA extraction. In four cases, the mother-trees No. 45, 19, 13 and 47, besides the progenies the leaves were also sampled, and they are numbered in white with grey background. In the total the CSO includes 728 ramets from 60 different clones.

described Doyle and Doyle (1990) and Szmidt et al. (1996). Three pair-primers flanking pine nuclear microsatellites (Mariette et al., 2001) (coded FRPP91, FRPP94 and ITPH4516) were used to amplify the genomic DNA. Each forward *primer* was labeled with CEQ dye blue (D4, Beckman Coulter).

PCR reactions were performed in a final volume of 25 μL that contained 1.25 U of *Taq* DNA polymerase (Invitrogen Life Technologies), 2.5 μL of 10×reaction buffer (Invitrogen Life Technologies), 0.5 μL of each dNTP (5 mM) and, for each primer, an optimised concentration of DNA, *primer* and $MgCl_2$. Amplifications were carried out using an Uno-Thermoblock thermal cycler. The amplification conditions followed the protocol explained by González-Martínez et al. (2001). After the amplification, 1 μL of PCR product (for FRPP91 and ITPH4516) and 0.5 μL (FRPP94) were mixed with 24 μL of deionised formamide (SLS, Sample Loading Solution), with 0.5 μL of CEQ Well Red fluorescent dye and one drop of mineral oil for fragment analysis. Fragments were separated in a capillary sequencer (Beckman Coulter, CEQ™ 8000) with linear polyacrylamide (LPA) denaturing gel and CEQ separation buffer. The samples were genotyped twice to avoid genotyping errors, and no mismatch was found between genotype of the offspring and the genotype of the mother.

2.3. Diversity estimates, mating system, paternity assignment and effective number of clones

The diversity parameters and the polymorphic information content (PIC) (Botstein et al., 1980) for the 60 parental genotypes and for the 206 progenies were computed for each microsatellite by using the Cervus 2.0 software (Marshall et al., 1998). The diversity parameters comprised the observed number of alleles (A_o), the observed heterozygosity (H_o), the expected heterozygosity (H_e) (Nei, 1987), and the fixation index ($F = 1 - (H_o/H_e)$) (Weir and Cockerham, 1984). Hardy-Weinberg equilibrium (HWE) departures were also tested for each locus within the parental and progeny groups (see details in Marshall et al., 1998).

Mating system parameters were estimated using the Multilocus Mating System Program (MLTR) version 2.4, adapted to highly polymorphic (microsatellites), and based on a mixed mating model, where a fraction of the progeny of the mother-tree are derived from self-fertilization and the remainder derived from outcrossing at random (Ritland, 2002). The multilocus population outcrossing rate (*tm*) and the minimum variance single-locus population outcrossing rate (*ts*) were estimated from offspring and seed parent genotypes. If mating occurs between relatives (biparental inbreeding), some outcrossing events would be confounded with selfing events, therefore (*tm* – *ts*) is an estimate of the minimal fraction of apparent selfing events due to biparental inbreeding (Ritland, 2002). Sampling from the original dataset with replacement created replicated datasets and 1000 bootstraps estimated the parameters' standard errors. Twenty different families based on the 21 mother-trees – two mother-trees had the some genotype – were analysed, in a total of 206 progenies. Since the maternal genotypes were known, the parameters described above were estimated for each family.

The allele frequencies of the 60 genotypes (potential fathers) in the CSO and of the 206 offspring were available. The identity and the exclusion probabilities were computed using the FaMoz software (Gerber et al., 2003). The expected or observed identity probabilities represent the probability that two individuals drawn at random or observed from a population will have the same genotype at multiple loci. The exclusion probability can be defined as the average capability of any marker system to exclude any given relationship, and, in the case reported here, it is the exclusion probability of a mother-offspring pair compared with a potential father, this probability is dependent on the genotypes of the reported relatives, the frequency of alleles at the loci and the number of independent loci tested. The microsatellites used in this study are independent, being located on different linkage groups (González-Martínez et al., 2002).

For the paternity analysis we considered as potential fathers all genotyped trees (60 clones). Maximum-likelihood estimates and simulations were performed using the FaMoz software for parentage assignment and gene flow estimates. For each offspring, the most-likely fathers were identified by LOD score calculation: the likelihood of an individual being the father of a given offspring is divided by the likelihood of any other individual from the population being the father (Meagher and Thompson, 1986).

Since the statistical laws associated to the likelihood ratio are unknown, simulations were used to calculate the threshold value of the LOD score for the parentage assignment, and if a parent had an LOD score exceeding the single parent threshold, it was considered as a true potential parent (Gerber et al., 2003; Meagher and Thompson, 1986). The simulations were done by generating 10 000 random offspring with the father inside the CSO or outside the CSO. For each offspring the most-likely father among the genotyped parents was identified and the LOD score value recorded (see details in Gerber et al., 2003). We defined a LOD score threshold of 2.5 for resolving paternity, and it was chosen in the intersection of the two distributions to minimize type I and type II errors. The test was further applied to simulated data to check the number of correct assignments (Gerber et al., 2003). An error due to mistyping was introduced at a rate of 0.0001, both in the simulation procedure, in the assignation of the most-likely father, and gene flow estimates, because mistyping is very likely to occur when scoring microsatellites and any parentage analysis should allow for at least a low rate (Gerber et al., 2000). After the real data analysis three alternative conclusions were then possible after simple exclusion analysis: (i) no father was found inside the CSO, (ii) more than one father was found inside the CSO, and (iii) a single father was found within the CSO (Gerber et al., 2003).

The orchard was not isolated from other pollen sources, and the genotyped trees were only a subset of the potential parents. We estimated the percentage of pollination events originated from outside the CSO (GFO, gene flow from outside) for each mother sampled, and then averaged those values to obtain an estimate for the entire CSO. GFO is the percentage of each mother's progeny with no pollen donor identified inside the plot and the GFI is the gene flow from inside the seed orchard. The *G*-test (Sokal and Rohlf, 1981) was used to test the differences found among the three plots in the rate of gene flow from outside the CSO (GFO) using the statistical package SPSS version 14.

The effective number of clones, N_c, was calculated based on the variation of ramet numbers among clones:

$$N_c = \frac{n_{\text{total}}^2}{\sum_{i=1}^{N} n_i^2} \text{ (Kang et al., 2001a),}$$

where n_{total} is the total number of ramets in the seed orchard (728), n_i is the number of ramets of clone i and N is the census number of clones in the CSO (60).

Table I. Diversity parameters (A_o observed number of alleles, H_o, H_e, observed and expected heterozygosity, $F = 1 - (H_o/H_e)$ fixation index), polymorphic information content (PIC), paternity exclusion probability (EP), and identity probability (IP) for the 60 parental genotypes for each microsatellite.

Locus	A_o	H_o	H_e	F	EP	IP	PIC
FRPP91	11	0.817	0.866	0.057^{NS}	0.748	0.02240	0.844
FRPP94	9	0.783	0.678	-0.155^{NS}	0.405	0.16322	0.618
ITPH516	13	0.517	0.824	0.373^{**}	0.687	0.03539	0.795
Mean	11	0.706	0.789	0.091	0.953[a]	0.00013[a]	0.753

[a] Cumulative over loci. Significance level: NS = non-significant; ** $p < 0.01$.

Table II. Diversity parameters for three microsatellite loci, and the polymorphic information content (PIC) for the 206 offspring. Abbreviations as in Table I.

Locus	A_o	H_o	H_e	F	PIC
FRPP91	11	0.485	0.874	0.445^{**}	0.858
FRPP94	7	0.563	0.639	0.119^{NS}	0.589
ITPH516	14	0.709	0.837	0.153^{**}	0.819
Mean	10.7	0.586	0.784	0.239	0.755

Significance level: NS = non-significant; ** $p < 0.01$.

3. RESULTS

3.1. Genetic diversity, identity and exclusion probabilities

The 60 genotypes were unambiguously identified with the three-microsatellite markers, except two pairs, which had the same genotype (clones no. 14/42 and 11/47). In the case of the parental genotypes, locus FRPP91 had the highest paternity exclusion probability (0.75), followed by ITPH4516 (0.69) and, finally, locus FRPP94 with the lowest value, 0.41. The polymorphic information content (PIC) followed the same pattern and the identity probability a conversed one. A cumulative exclusion probability of 0.95 and cumulative identity probability of 0.00013 were found (Tab. I). The mean observed number of alleles and observed genetic diversity of the three SSRs were respectively 11 and 0.706 for the clones, and 10.7 and 0.586 in the case of the progenies (Tabs. I and II). The observed genetic diversity was much lower in the case of progenies as compared to the parental one (about 20% less), but the average number of alleles is similar. The expected diversity was higher for the most informative locus (FRPP91), both in the parental genotypes and in the progenies, as compared to the other loci. The heterozygote deficiency (F) was significantly higher in the case of locus ITPH4516 in the parental population (Tab. I). In the case of the progenies a sounding excess of heterozygotes was found for the three loci (average 0.24, see Tab. II), in particular the loci FRPP91, with 45%, and two other loci showed a significant departure from the Hardy-Weinberg equilibrium (HWE).

3.2. Mating system estimates

The single-locus outcrossing rate (t_s) was higher in plot 2 (79.4 ± 4.8) as compared to the other plots. No differences were found among the three plots for the multilocus outcrossing rate (t_m), and the probability that an embryo sampled from a mother-tree is derived from an outcross was 90.1 ± 2.3. When the multilocus is higher than the single-locus outcrossing rate, this means that there is an effect of the population substructure on the male similarity between outcrosses, and, in our case, the minimum estimate of biparental inbreeding ($t_m - t_s$) was quite high: 21.7 ± 2.9. Biparental inbreeding was found higher in plot 1 (22.3 ± 5.2), followed by plot 3 (18.8 ± 6.1) and 2 (12.9 ± 4.5) (Tab. III).

3.3. Paternity assignment, pollen contamination estimates and effective number of clones

No possible male parent was found within the 60 potential pollen donors in the CSO for 108 embryos out of the 206 analysed, yielding a minimum estimate of pollen contamination of 52.4%: the observed gene flow (GFO). A single matching father was found within the CSO for 51 offspring, and more than one father could be found among the 60 clones for 47 of the sampled seeds, meaning that the trees inside the CSO fathered 47.6% of the offspring (51 + 47 = 98 out of 206 sampled seeds). Nineteen single matching fathers were found in plot 3, a value slightly higher than the 16 ones found in each of the other two plots. Gene flow from outside the CSO (GFO) yield very small differences among plots, the highest value was found in plot 2 (60.0%), against 49% in the other plots (see Tab. III), but the *G*-test revealed that those differences were not significant ($G = 2.451$, $P = 0.294$). The minimum and maximum GFO per mother-tree was 16.7 and 90% in trees no. 33 I/B and 15 VII/M (see location in Fig. 1), respectively, and 16 mother-trees had progenies pollinated by at least 40% of fathers from inside the CSO (data not shown).

Putative selfing events happened in 5 out of the 20 genotypes (21 mother-trees), and the number of putative selves per tree varied from one to two seeds (Tab. IV); however, we should not exclude outcross events, in this case, since clones are represented by several ramets in the CSO. The average selfing event in the CSO was 3.4%, the lowest value in plot 2 (1.4) and the highest in plot 3 (5.7) (Tab. III).

The trees that pollinated at least one time (single-matching father), ranging from one to six seeds, and that 30% of those paternal trees were involved in only one mating event. Two parental trees fathered a slightly higher number of seeds than the others (clones 47 and 61 fathered six and four seeds, respectively). Finally, the effective number of clones (N_c) was 36.

4. DISCUSSION

4.1. Genetic variation

In the fingerprinting analysis, only two individuals showed the same multilocus genotype, for the three analysed loci, but

Table III. Paternity assignment and mating system estimate (in percentage). *GFO*, gene flow from outside, meaning the percentage of seeds for which no pollen donors were found inside the CSO; *Selfing*, seeds with father and mother with the same genotype (selfing or outcrossing events: see text for details); *tm*, the seed orchard the multilocus outcrossing rate; *ts*, the single-locus outcrossing rates; *tm* – *ts*, the biparental inbreeding. Standard errors of estimates are based on 1000 bootstraps using progeny arrays as the unit of resampling.

Parameters	*GFO*[a]	*Selfing* (100-*Selfing*)[a]	t_m (s.d.)[b]	t_s(s.d.)[b]	$t_m - t_s$ (s.d.)[b]
Plot 1	48.5	3.0 (97.0)	93.7 (3.7)	71.4 (7.7)	22.3 (5.2)
Plot 2	60.0	1.4 (98.6)	92.3 (3.8)	79.4 (4.8)	12.9 (4.5)
Plot 3	48.6	7.1 (92.9)	89.3 (5.2)	70.5 (8.7)	18.8 (6.1)
CSO	52.4	3.9 (96.1)	90.1 (2.3)	68.3 (4.2)	21.7 (2.9)

[a] Estimated with paternity analysis.
[b] Estimated by multilocus approach.

Table IV. The number of seeds from the mother-tree assigned to each clone/pollen donor (single matching father). Dark shaded cells correspond to apparent selfing events.

		Pollen donor																								
	mother tree	5	7	8	9	12	14	17	23	24	29	30	33	34	39	41	47	48	55	56	58	61	64	65	68	79	81	Total
PLOT 1	23I/H								1												1							2
	33I/B			1									1															2
	35I/C														1		2		1			1						5
	42I/G				1												1											2
	45I/F, I/J																1			3								4
	78I/E												1															1
PLOT 2	10VII/O																							1				1
	15VII/M																	1										1
	19VI/N					1								2												1		4
	30VII/N													1														1
	68VII/L							1															2		1			4
	73VII/N			1														1				2						4
	79VI/P																									1		1
PLOT 3	14XII/R		2				1									1												4
	31XII/T										1	1																2
	39XII/S																		1									1
	41XII/V	1				1				1															1		1	5
	47XII/T																2											2
	48XII/V						1		1									1										3
	61XII/T											1										1						2
	Total	1	2	2	1	2	2	1	2	1	1	2	2	3	1	1	6	3	2	3	1	4	2	1	2	2	1	51

the high exclusion probability and the low identity probability indicated that the three loci were polymorphic enough for a paternity analysis. In a similar study, three pairs of clones with identical multilocus genotype were also found, based on six SSR loci (Buiteveld et al., 2001). Microsatellites have a large number of rare alleles that are required to perform paternity analysis by exclusion. In the present case, the large majority of the alleles (72%) in the 60 parental trees occurred at frequencies bellow 10%. This was also reported in studies inferring paternity in oak (Dow and Ashley, 1998; Streiff et al., 1999) and in *Eucalyptus* (Chaix et al., 2003).

In the case of the 60 parental trees, only one locus was not in HWE, but relatedness among parental trees is not to be excluded, as was further confirmed by a sounding probability of biparental inbreeding. This is a thorough hypothesis since the 60 plus trees used in the CSO were all selected within the same provenance: Leiria. In a similar study, three out of six SSR loci also displayed significant departures from the HWE (Chaix et al., 2003) both in the parental trees and in the progeny. In the current study, the fixation index was found to be much higher in the progeny (0.239) than in the parental population (0.091) and in a population from the provenance where the trees were selected (Leiria: 0.003, data from Derory et al. 2002 using the same three loci). The three loci displayed very positive F-values in the case of the progenies, though one of them in HWE. Deviations from Hardy-Weinberg equilibrium at many or all loci are an indicator of population substructure and absence of panmixia, and those values are probably due to relatedness among groups of progenies and to the unevenness of pollination. Evidences of mating among relatives existed in this CSO (biparental inbreeding: follow discussion below). Moreover, in general, conifers show an excess of homozygotes over panmictic expectations at the embryo stage, which later disappears at the adult stage (e.g.: *Abies* spp. Parducci et al., 2001). In *P. pinaster*, according to González-Martínez et al. (2003), the carriers of most lethal or sublethal alleles are probably eliminated during seed formation and germination, as well as during the first growing season. The hypothesis of genotyping problems due to the presence of null alleles for those loci was discarded because very low F-values after

genotyping 76 adult trees and 132 progenies (seedlings and saplings) were found in a study using the same three loci and the same species (González-Martínez et al., 2002). Additionally, data obtained after screening 47 *P. pinaster* populations from the range-wide of the species, with the same set of loci used in the current paper, showed no deficiency of heterozygotes in the western populations, which included Portuguese populations (Derory et al., 2002). In particular, the Leiria population from the Derory et al. (2002) study had F = 0 for the most problematic locus (ITPH516) of the current study. In a *Pseudotsuga menziesii* CSO study, the fixation indices were also higher in progenies than in mother-trees, but an excess of homozygocity was found both in the seed orchard and in the natural stands (Prat and Burczyk, 1998). As a consequence of mating among relatives, an increase in homozygosity may lead to a reduction of fitness (Spielman et al., 2004), and thus have consequences in the CSO genetic gains.

The expected heterozygosity (H_e) was similar in the progenies (0.784) and in the CSO clones (0.789), both slightly higher compared to the Leiria's population value (0.709, data from Derory et al., 2002). Also the average number of alleles was similar both in the progeny and in the parental clones. If the seed orchard consisted of clones collected from a wide geographic area this could result in a diversity increase in the orchard compared to the populations of the species. However, this does not appear to be the case, since a single provenance was used to select the plus trees. Similar findings were reported in a Douglas fir SO, the H_e of natural and seed orchard parental population were also similar, though slightly lower in the seed orchard progeny where the clones were also collected in only one provenance (Prat and Burczyk, 1998). The observed diversity was reduced by about 20% in the progeny as compared with the parental trees, but this was probably due to the biparental inbreeding, the asymmetric contribution of parental clones and to the progenies substructure, i.e., the relatedness among the seeds collected in the same mother-tree. Conversely, Chaix et al. (2003) in a seedling seed orchard of *Eucalyptus*, observed similar values of observed heterozygosity in the parental trees and in the progenies, but biparental inbreeding was not present.

4.2. Mating system

Selfing also reduces the genetic value of the orchard seeds, but the observed selfing events were low and affected only 3.9% of the progenies of the current study. A similar result (3.8%) was obtained in an adult stand of *P. pinaster* by using the same set of SSR (González-Martínez et al., 2003). Nevertheless, in the CSO we studied, outcross events among trees with the same genotype were possible for several ramets of the same clone (in Fig. 1 some possibilities of apparent selfing for mother-tree 23 are shown with arrows).

The minimum estimate of biparental inbreeding was 21.7%, a very high value considering the caution in selection of the parental trees (plus trees) for establishing the CSO, but since they were collected in the same provenance, a relationship among different genotypes is a reasonable hypothesis. Indeed, a certain level of biparental inbreeding is common, as shown by differences between the multilocus estimate of outcrossing and the average of the single-locus estimate using progeny arrays, in different *Pinus* species, though low genetic relatedness was found in *P. pinaster* natural stands (González-Martínez et al., 2003; Lucas et al., 2008). Forest trees are known to suffer from high inbreeding depression (Williams and Savolainen, 1996), and both biparental inbreeding and selfing may have harsh consequences in this species. The consequences of inbreeding depression had consequences in several important traits, in an 11-year progeny trial of *P. pinaster* (Durel et al., 1996), the mean inbreeding depressions were 27% for height, 37% for circumference at breast height (63% for bole volume) and 89% for female fertility.

The observed outcrossing rate, computed through paternity analysis was 96.2%, and high outcrossing rates are common in conifer species, in particular pines (Burczyk et al., 1996 and references therein). A similar value (96%) was also estimated in two studies using *P. pinaster* natural stands and chloroplast and nuclear microsatellites (González-Martínez et al., 2003; Lucas et al., 2008). In the three plots, the mean multilocus estimates of outcrossing rate and paternity analysis gave similar results, considering that the paternity analysis values overlapped the mean multilocus estimates upper values (Tab. III). Nevertheless, the probability that an embryo sampled from a mother-tree is derived from an outcross (90.1 ± 2.3) was slightly lower than the observed outcrossing rate. This could be explained by the lower sensitiveness of multilocus estimators to the mixed mating model assumptions, mating among relatives also affects those estimates, albeit in a smaller degree as compared to the single-locus one (Prat and Burczyk, 1998).

4.3. Gene flow

The observed gene pollen flow from outside the CSO, 52.4%, was very high, but still within the range of wind-pollinated species, which are known to experience high pollen contamination rates, approaching or exceeding 40% (Burczyk et al., 2004a and references therein). Indeed, this species is known to have an extensive historical gene flow (*Nm* = 2.99: Salvador et al., 2000). The observed high gene flow value could partially be explained by the scattered *P. pinaster* trees within the orchard's isolation zone and a *P. pinaster* stand closer than 2 km. A similar situation was observed in a Scots pine seed orchard isolated from nearby native stands by at least 2 km, where the pollen contamination was 48%, but the isolation zone contained scattered Scots pine individuals (Harju and Nikkanen, 1996). Plomion et al. (2001) observed that the minimum pollen contamination rate was 36%, in a *P. pinaster* polycross seed orchard (PSO), using chloroplast microsatellites, but the studied orchard block was bordered by other PSO blocks and by mature *P. pinaster* natural stands. Minimum pollen flow within a *P. pinaster* natural stand was around 30% based only on exclusion analysis, but the true level of pollen gene flow was probably much higher (González-Martínez et al., 2003). Reports about pollen contamination in seed orchards demonstrated that gene flow can be extensive

and there is evidence that pollen of widely distributed forest tree species can disperse over large distances, up to tens or hundred kilometers (Burczyk et al., 2004a). Even with spatial isolation from stands of the same species an existence of pollen contamination of 48% was observed by Moriguchi et al. (2004). It seems likely that contaminating pollen will generally have a lower breeding value than that of the orchard (Kang et al., 2001b), but it is difficult to assess the genetic quality and quantity of incoming pollen from surrounding stands. In a *P. pinaster* SO the pollen contamination decreased the genetic gain between 50 and 82% of what was originally expected (Plomion et al., 2001).

The pollen contamination was found to be higher in the external plots as compared to the inner plot (N° 2 (60%) < N° 3 (48.6%) and N° 1 (48.5%)), but the differences among plots were non-significant, therefore no windward oriented gene flow was observed in the CSO we studied. Nevertheless, in their study (Plomion et al., 2001) found that the pollen flow from outside the orchard was windward oriented, and, as a consequence, the contamination was not evenly distributed in that direction. The influence of the wind in the seed orchard's pollen pollution seems to follow a complex pattern and contradictory data are found in the literature (Dow and Ashley, 1998).

4.4. Paternity assignment and effective number of clones

The number of mother-trees that had values superior or equal to 50% of the progeny with at least one father inside the seed orchard was 13 out of the 20 genotypes, and the values of progenies pollinated by trees from inside the seed orchard, per mother-tree, ranged from 10% to 83%. We can conclude that the number of progenies pollinated by trees from inside the CSO varied with the mother tree, which might be related to several factors, such as flowering phenology variability and putative variation in clonal fertility. Burczyk et al. (2002) referred that the reproductive system of a CSO they studied was far from panmixia and that the large phenology differences among individuals prevent pollination from early and late flowering trees. Several authors found that the earlier or later receptive mother-trees were more prone to be fertilized by pollen from outside the orchard (Harju and Nikkanen, 1996; Slavov et al., 2005a). In the CSO since about only half of the progenies were pollinated by trees from within the SO, therefore the genetic gain could be reduced; probably flowering asymmetry and pollen production were key factors for fertilization success. Also, in *Q. robur* the studied mother-trees had values of progenies fathered by trees inside the CSO consistently below 40% (Buiteveld et al., 2001). We could only assign 25% of the sampled seeds to single matching fathers. In their study, Buiteveld et al. (2001) could only assign 36% of the sampled acorns with the six SSR loci they used.

In the process of orchard establishment, variation in the ramets of the selected clones occurred. To study the impact of such phenomenon we estimated the effective number of clones N_c in the CSO, which was 36. We also considered the relative effective number of clones ($N_r = N_c/N = 0.6$), based on the census number ($N = 60$). Kang et al. (2001a) studied 255 conifers CSO and the effective number of clones varied from 10 to 421, averaging 66. The same authors, based on simulations observed that the reduction in gene diversity is profound when N_c becomes smaller than 10, a value much lower than the value we obtained in the current study.

4.5. Final considerations

The base material for the CSO establishment was initially selected in the Leiria provenance, which has a large plasticity and growth potential demonstrated in several provenance trials (e.g. Alía et al., 1995) and was, probably, the seed source for reforestation in the country (Ribeiro et al., 2001). Nevertheless, the base population of the seed orchard may not fully represent the country's species genetic variability. A study comparing the diversity of the CSO and a representative sample of the range of species in the country should to be performed to assess if the allelic richness and genetic diversity of the breeding population is large enough to support reduction in the number of genotypes and to maintain the frequency of the desirable genes in the seed-orchard crops for future breeding cycles. Mislabeling may also be a source of error and the clonal identities of CSO ramets, which should be checked with SSR markers, and an appropriate sampling design. Finally, a phenology study of the CSO should detect the variation among clones regarding flowering characteristics, which usually causes unbalanced male and female contribution, reducing the effective population and decreasing the genetic diversity of the orchard's progeny.

In practice, due to relatedness among parents, variation in clonal fertility, ramet number, and background pollination in clonal seed orchard, the realized genetic gain and gene diversity of open pollinated seed orchard deviate from expectation. Therefore, the results presented in this paper may have several practical implications, which could help in designing and managing other CSO to be installed: (i) establish seed orchards in areas well-isolated from putative contamination sources, by at least few kilometers, (ii) use of gibberellins to increase reproductive success, (iii) use of controlled pollination whenever possible, and (iv) clones should be checked for relatedness and the ramet number could be directly proportional to the breeding value, thus clones with the highest breeding value are permitted to contribute more to diversity loss, and to achieve the highest genetic gain with a given desirable gene diversity (Kang et al., 2001a).

Acknowledgements: L Fernandes was supported by a PRODEP III – Action 3.2 fellowship and the research was supported by ITQB research line ICL38. Special thanks are due to A. Aguiar for exchanging stimulating ideas and for his help in interpreting the data. This work could not have been done without the collaboration of M. Veloso from Estação Agronómica Nacional, who facilitated access to the sequencing unit. M.O. Krause is gratefully acknowledged for the final English revision.

REFERENCES

Adams W.T. and Birkes D.S., 1991. Estimating mating patterns in forest tree populations. In: Fineschi S., Malvolti M.E., Cannata F., and Hattemer H.H. (Eds.), Biochemical markers in the population genetics of forest trees, SPB Academic Publishing, The Hague, pp. 157–172.

Adams W.T. and Burczyk J., 2000. Magnitude and implications of gene flow in gene conservation reserves. In: Young A., Boshier D., and Boyle T. (Eds.), Forest conservation: principles and practice, CSIRO Publisher, Collingwood, pp. 215–224.

Alía R., Gil L. and Pardos J., 1995. Performance of 43 *Pinus pinaster* Ait. provenances on 5 locations in central Spain. Silvae Genet. 44: 75–81.

Botstein D., White R.L., Skolnick K., and Davis R.W., 1980. Construction of a genetic linkage map in man using restriction fragment length polymorphism. Am. J. Hum. Genet. 32: 314–331.

Buiteveld J., Bakker E.G., Bovenschen J., and de Vries S.M.G., 2001. Paternity analysis in a seed orchard of *Quercus robur* L. and estimation of the amount of background pollination using microsatellite markers. For. Genet. 8: 331–337.

Burczyk J., Adams W., and Shimizu J., 1996. Mating patterns and pollen dispersal in a natural knobcone pine (*Pinus attenuata* Lemmon.) stand. Heredity 77: 251–260.

Burczyk J., Adams W.T., Moran G.F., and Griffin A.R., 2002. Complex patterns of mating revealed in a *Eucalyptus regnans* seed orchard using allozyme markers and the neighbourhood model. Mol. Ecol. 11: 2379–2391.

Burczyk J., DiFazio S.P., and Adams W.T., 2004a. Gene flow in forest trees: How far do genes really travel? For. Genet. 11: 1–14.

Burczyk J., Lewandowski A., and Chalupka W., 2004b. Local pollen dispersal and distant gene flow in Norway spruce (*Picea abies* [L.] Karst.). For. Ecol. Manag. 197: 39–48.

Chaix G., Gerber S., Razafimaharo V., Vigneron P., Verhaegen D., and Hamon S., 2003. Gene flow estimation with microsatellites in a Malagasy seed orchard of *Eucalyptus grandis*. Theor. Appl. Genet. 107: 705–712.

Derory J., Mariette S., González-Martínez S.C., Chagné D., Madur D., Gerber S., Ribeiro M.M., and Plomion C., 2002. What can nuclear microsatellites tell us about maritime pine genetic resources conservation and provenances certification strategies? Ann. For. Sci. 59: 699–708.

Dow B.D. and Ashley M.V., 1998. High levels of gene flow in bur oak revealed by paternity analysis using microsatellites. J. Hered. 89: 62–70.

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

Durel C.E., Bertin P., and Kremer A., 1996. Relationship between inbreeding depression and inbreeding coefficient in maritime pine (*Pinus pinaster*). Theor. Appl. Genet. 92: 347–356.

El-Kassaby Y.A., 1995. Evaluation of the tree-improvement delivery system: factors affecting genetic potential. Tree Physiol. 15: 545–550.

Gerber S., Chabrier P., and Kremer A., 2003. FaMoz: a software for parentage analysis using dominant, codominant and uniparentally inherited markers. Mol. Ecol. Notes 3: 479–481.

Gerber S., Mariette S., Streiff R., Bodénès C., and Kremer A., 2000. Comparison of microsatellites and amplified fragment length polymorphism markers for parentage analysis. Mol. Ecol. 9: 1037–1048.

González-Martínez S.C., Agúndez D., Alía R., Salvador L., and Gil L., 2001. Geographical variation of gene diversity of *Pinus pinaster* Ait. in the Iberian Peninsula. In: Müller-Starck G. and Schubert R. (Eds.), Genetic Response of Forest Systems to Changing Environmental Conditions. Kluwer Academic Press, Dordrecht, pp. 161–171.

González-Martínez S.C., Gerber S., Cervera M.T., Martínez-Zapater J.M., Gil L., and Alía R., 2002. Seed gene flow and fine-scale structure in a Mediterranean pine (*Pinus pinaster* Ait.) using nuclear microsatellite markers. Theor. Appl. Genet. 104: 1290–1297.

González-Martínez S.C., Gerber S., Cervera M.T., Martínez-Zapater J.M., Gil L., and Alía R., 2003. Selfing and sibship structure in a two-cohort stand of maritime pine (*Pinus pinaster* Ait.) using nuclear SSR markers. Ann. For. Sci. 60: 115–121.

Harju A.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.

Kang K.-S., 2001. Genetic gain and gene diversity of seed orchard crops. The Swedish University of Agricultural Sciences, Umeå.

Kang K.S., Harju A.M., Lindgren D., Nikkanen T., Almqvist C., and Suh G.U., 2001a. Variation in effective number of clones in seed orchards. New For. 21, 17.

Kang K.S., Lindgren D., and Mullin T.J., 2001b. Prediction of genetic gain and gene diversity in seed orchard crops under alternative management strategies. Theor. Appl. Genet. 103: 1099–1107.

Lucas A.I., Robledo-Arnuncio J.J., Hidalgo E., and Gonzalez-Martinez S.C., 2008. Mating system and pollen gene flow in Mediterranean maritime pine. Heredity 100: 390–399.

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

Marshall T.C., Slate J., Kruuk L., and Pemberton J.M., 1998. Statistical confidence for likelihood-based paternity inference in natural populations. Mol. Ecol. 7: 639–655.

Meagher T.R. and Thompson E., 1986. The relationship between single parent and parent pair genetic likelihoods in genealogy reconstruction. Theor. Popul. Biol. 29: 87–106.

Moriguchi Y., Taira H., Tani N., and Tsumura Y., 2004. Variation of paternal contribution in a seed orchard of *Cryptomeria japonica* determined using microsatellite markers. Can. J. For. Res. 34: 1683–1690.

Moriguchi Y., Tsuchiya S., Iwata H., Itoo S., Tani N., Taira H., and Tsumura Y., 2007. Factors influencing male reproductive success in a *Cryptomeria japonica* seed orchard revealed by microsatellite marker analysis. Silvae Genet. 56: 207–214.

Nei M., 1987. Molecular evolutionary genetics. Columbia University Press, New York, 512 p.

Parducci L., Szmidt A.E., Ribeiro M.M., and Drouzas A.D., 2001. Taxonomic position and origin of the endemic sicilian fir *Abies nebrodensis* (Lojac.) Mattei based on allozyme analysis. For. Genet. 8: 119–127.

Perry D.H., 1940. *Pinus pinaster* in Western Australia. Aust. For. 5: 85–87.

Plomion C., LeProvost G., Pot D., Vendramin G., Gerber S., Decroocq S., Brach J., Raffin A., and Pastuszka P., 2001. Pollen contamination in a maritime pine polycross seed orchard and certification of improved seeds using chloroplast microsatellites. Can. J. For. Res. 31: 1816–1825.

Prat D. and Burczyk J., 1998. Genetic variation and mating system in a native provenance and the derived seed orchard of douglas-fir (*Pseudotsuga menziesii* (Mirb.) Franco). For. Genet. 5: 201–209.

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

Ritland K., 2002. Extensions of models for the estimation of mating systems using n independent loci. Heredity 88: 221–228.

Salvador L., Alía R., Agúndez D., and Gil L., 2000. Genetic variation and migration pathways of maritime pine (*Pinus pinaster* Ait.) in the Iberian Peninsula. Theor. Appl. Genet. 100: 89–95.

Slavov G.T., Howe G.T., and Adams W.T., 2005a. Pollen contamination and mating patterns in a Douglas-fir seed orchard as measured by simple sequence repeat markers. Can. J. For. Res. 35: 1592–1603.

Slavov G.T., Howe G.T., Gyaourova A.V., Birkes D.S. and Adams W.T., 2005b. Estimating pollen flow using SSR markers and paternity exclusion: accounting for mistyping. Mol. Ecol. 14: 3109–3121.

Sokal R.R. and Rohlf F.J., 1981. Biometry. W.H. Freeman and Co., San Francisco, 859 p.

Spielman D., Brook B.W., Briscoe D.A., and Frankham R., 2004. Does inbreeding and loss of genetic diversity decrease disease resistance? Conserv. Genet. 5: 439–448.

Streiff R., Ducousso A., Lexer C., Steinkellner H., Gloessl J., and Kremer A., 1999. Pollen dispersal inferred from paternity analysis in a mixed oak stand of *Quercus robur* L. and *Q. petraea* (Matt.) Liebl. Mol. Ecol. 8: 831–841.

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

Varela C., 1989. Comportamento reprodutivo de *Pinus pinaster* Ait. num pomar clonal de sementes. Universidade Técnica de Lisboa, Lisbon, 107 p.

Weir B.S. and Cockerham C.C., 1984. Estimating F-statistics for the analysis of population structure. Evolution 38: 1358–1370.

Wheeler N.C. and Jech K.S., 1992. The use of electrophorectic markers in seed orchard research. New For. 92: 311–328.

Williams C.G. and Savolainen O., 1996. Inbreeding depression in conifers: Implications for breeding strategy. For. Sci. 42: 102–117.