Pollen contamination and mating structure in maritime pine (*Pinus pinaster* Ait.) clonal seed orchards revealed by SNP markers

Laurent Bouffier [a,*], Sandrine Debille [b], Pierre Alazard [c], Annie Raffin [d], Patrick Pastuszka [d], Jean-François Trontin [b]

[a] INRAE, Univ. Bordeaux, BIOGECO, F-33610 Cestas, France

[b] FCBA, BioForBois, Pôle Industrie Bois & Construction, F-33610 Cestas, France

[c] FCBA, Pôle Biotechnologie et Sylviculture avancée, F-33610 Cestas, France

[d] INRAE, UEFP, F-33610 Cestas, France

* Corresponding author

Running title

Pollen contamination in maritime pine seed orchards

Keywords

Pollen contamination, paternity reconstruction, SNP markers, *Pinus pinaster* Ait., clonal seed orchard

Corresponding author

Laurent Bouffier

laurent.bouffier@inrae.fr

ORCID : 0000-0001-7493-5077

Abstract

Maritime pine (*Pinus pinaster* Ait.) is a major forest tree species in south-western Europe. In France, an advanced breeding program for this conifer species has been underway since the early 1960s. Open-pollinated seed orchards currently supply more than 90% of maritime pine seedlings for plantation forestry. However, little is known about pollen contamination and mating structure in such seed orchards for maritime pine. We analyzed these features here, focusing on: i) the location of the seed orchard, ii) the year of pollination, and iii) the genotype of the tree from which seeds were collected. Parental analyses based on an optimized set of 60 SNP markers were performed on 2,552 seedlings with Cervus software (likelihood inference methodology).

Pollen contamination rates were highly variable between seed lots (from 20 to 96%), with a mean value of 50%. Several interpretative factors were highlighted, including the distance between the seed orchard and external pollen sources, rain during the pollination period, seed orchard age, soil characteristics and seed parent identity. All parental genotypes contributed to the offspring as pollen parents, but differences in paternal reproductive success were detected. These differences were only partly explained by differences in the number of ramets of each parental genotype deployed in each seed orchard. Finally, the overall self-fertilization rate was estimated at 5.4%, with considerable variability between genotypes. These findings are useful to formulate recommendations for seed orchard management and for identifying new research perspectives.

Introduction

The first breeding programs for forest trees were initiated in the mid-20th century, to improve forest genetic resources in terms of productivity, wood quality and adaptation to environmental conditions (Burdon et al. 2008; McKeand et al. 2003; Pâques 2013). These programs have been shaped by environmental, biological, economic, institutional and sociopolitical factors (Mullin and Lee 2013). However, they have always been constructed around two main populations (Namkoong et al. 1988): a breeding population initiated by selecting superior trees, generally managed over successive cycles of crossing-testing-selection (recurrent selection scheme), and a deployment population (improved Forest Reproductive Material – FRM -) released for commercial plantations. The breeding population carries a high level of genetic diversity, to limit inbreeding (targeted effective size of 30 to 70) and ensure future genetic gains (Danusevicius and Lindgren 2005). By contrast, the deployment population is selected to maximize genetic gains for selection criteria. The genetic diversity of deployment populations varies between breeding programs, from single genotypes (clonal forestry) or mixtures of selected genotypes (multiclonal forestry, also known as multivarietal forestry, Weng et al. 2011) to synthetic populations produced through sexual reproduction in seed orchards.

Maritime pine (*Pinus pinaster* Ait.), which covers 4.2 million hectares, is a major forest tree species in south-western Europe (Abad Viñas et al. 2016). It covers 7% of the forested area in France (1.03 million ha), mostly in the large Landes de Gascogne forest (0.81 million ha), but it provides 18% of lumber and 27% of pulpwood production nationally (FCBA 2020). A breeding program was initiated in the 1960s from a base population selected in South-West France for two main criteria: growth rate and stem straightness. The breeding population was subjected to three cycles of recurrent selection (Mullin and Lee 2013). Over the past ten years (2011-2021), 375 million improved seedlings were sold by forest nurseries (French Ministry of Agriculture 2022), making maritime pine the leading species for plantation forestry in France. These seedlings were obtained from seeds collected in open-pollinated clonal and polycross seed orchards (Baradat 1987) displaying an expected genetic gain of 30% for breeding objective traits.

However, the expected gains in such synthetic populations are achieved only if two major assumptions are fulfilled: i) an equal contribution of parental genotypes and ii) an absence of contaminating pollen from surrounding or more distant stands. Many studies on conifer species have suggested that there may be significant levels of pollen contamination, based on pedigree reconstruction, initially with allozymes (Harju and Nikkanen 1996; Yazdani and Lindgren 1991) or RAPD (Goto et al. 2002), and subsequently with microsatellites (Slavov et al. 2005; Torimaru et al. 2009). Maritime pine is no exception: Plomion et al. (2001) reported a pollen contamination rate of 36% in French seed orchards, based on analyses of chloroplast microsatellite DNA, and an even higher level of contamination (32% to 81%) in analyses based on nuclear microsatellites (Plomion et al. 2005). Cryptic gene flow was suspected in these studies due to the low discrimination power of the microsatellites markers used. The recent development of single-nucleotide polymorphisms (SNP) has opened up new opportunities for analyzing mating patterns in seed orchards. Individually, SNP are considered less informative than microsatellites markers, but they are potentially numerous and easy to multiplex. Moreover, null alleles can be highly detrimental for parentage analyses based on microsatellites (Moriguchi et al. 2004; Plomion et al. 2005; Telfer et al. 2015). Finally, the simpler genotyping process and the higher repeatability of SNP over microsatellites (Jones et al. 2007) render the use of SNP an efficient and cost-effective tool for high-throughput analyses.

We used an SNP array previously optimized by Vidal et al. (2015) for pedigree reconstruction in maritime pine. We genotyped 2,552 seedlings at 80 loci to investigate pollen contamination and parental contributions in maritime pine open-pollinated clonal seed orchards (CSO), considering three main factors: i) the location of the seed orchard, ii) the year of pollination, and iii) the genotype of the tree from which the seeds were collected. We discuss these results from the perspective of optimizing the deployment of new seed orchards of this key tree species for plantation forestry in France.

Materials and methods

Plant material

Sampling was carried out in third-generation CSO (VF3) currently producing maritime pine seed lots (improved FRM) by open pollination. For each seed collected, the seed parent (maternal genotype) is known, as the parental genotypes are identified in the field, but the pollen parent (paternal genotype) is unknown, due to the system of open pollination. We studied seeds from three CSO established between 2002 and 2006. Each CSO was composed of the same 50 selected genotypes deployed through grafting. An additional genotype has been introduced following a restocking operation after establishment of the seed orchards. The actual contribution of each genotype varies within and between CSO due to differences in the number of ramets per genotype. The three CSO (Table 1) differ principally in terms of their location and soil characteristics:

- CSO-1, established in the northern part of the Landes de Gascogne forest (i.e. surrounded by maritime pine plantations) on sandy soils;
- CSO-2, established on the eastern outskirts of the Landes de Gascogne forest (i.e. nearest maritime pine plantations a few kilometers away) on clay loam;
- CSO-3, established at the southernmost eastern location, outside the Landes de Gascogne forest (i.e. nearest maritime pine plantations more than 20 kilometers away) on clay loam

In total, 2,552 seedlings were considered, through three sampling strategies (Table 2):

- Sampling strategy 1 (SS1): in the fall of 2012 and 2014 (pollination years 2011 and 2013), one central sampling zone was defined per CSO, except for CSO-1, in which two zones were considered (center vs. border). Four genotypes (denoted by A, B, C and D) were selected according to seed parent flowering phenology (Trontin et al. 2019): two early (A, B) and two late flowering genotypes (C and D). Each year, four to six cones were collected from two ramets (the same ramets were sampled in 2012 and 2014) per genotype and per sampling zone. After germination, 17 to 60 seedlings per genotype (1,524 in total) were sampled at random.

- Sampling strategy 2 (SS2): in the fall of 2014 (pollination year 2013), three cones were collected for 20 additional genotypes, randomly selected from CSO-2, and 27 to 30 seedlings per genotype were sampled at random (590 in total).
- Sampling strategy 3 (SS3): commercial seed lots, i.e. seeds extracted from bulked cones collected from 40 randomly selected trees from each CSO, were sampled in the fall of 2015 (pollination year 2014) and 142 to 149 seedlings per CSO were sampled (438 in total).

For each lot harvested, the seeds were germinated and grown in greenhouse conditions for 6 months. Seed parent identity was recorded for each seedling in SS1 and SS2, whereas the identities of both seed and pollen parents were unknown in SS3. There is a one-year time lag between pollination and fertilization in maritime pine. To avoid confusion, the years specified for each seed lots hereafter are the pollination years and not the sampling (fertilization) years.

Table 1. Characteristics of the three maritime pine clonal seed orchards (CSO) sampled.

Code	**Name (reference)**	**Establishment year**	**Location***	**Soil type**	**Area (ha)**	**No. genotypes (no. trees)**
CSO-1	Saint-Laurent2-VF3 (PPA-VG-014)	2006	Within (north)	Sandy	15.0	46 (3171)
CSO-2	Beychac-VF3 (PPA-VG-011)	2002-2003	Outskirts (northeast)	Clay loam	15.5	47 (3676)
CSO-3	Saint-Sardos VF3 (PPA-VG-015)	2003	Outside (southeast)	Clay loam	6.5	48 (1565)

*With reference to the Landes de Gascogne Forest

124 **Table 2.** Sampling strategies (SS1, SS2 and SS3) of 2,552 seedlings in three maritime pine clonal seed
125 orchards (CSO) over three pollination years.

Seed orchard	Pollination year	Sampling strategy	Sampling zone	No. seed parent genotypes*	No. seedlings genotyped (no. seedlings per genotype)
CSO-1	2011	SS1	Center	4 (A, B, C, D)	240 (60)
			Border	4 (A, B, C, D)	240 (60)
	2013	SS1	Center	4 (A, B, C, D)	116 (27-30)
			Border	4 (A, B, C, D)	120 (30)
	2014	SS3	Unknown	Unknown	147 (unknown)
CSO-2	2011	SS1	Center	4 (A, B, C, D)	240 (60)
	2013	SS1	Center	4 (A, B, C, D)	238 (58-60)
	2013	SS2	Center	20	590 (27-30)
	2014	SS3	Unknown	Unknown	142 (unknown)
CSO-3	2011	SS1	Center	4 (A, B, C, D)	240 (60)
	2013	SS1	Center	4 (A, B, C, D)	90 (17-29)
	2014	SS3	Unknown	Unknown	149 (unknown)

126 * A, B, C, D: reference genotypes with early (A, B) or late (C, D) female flowering (4-6 cones collected
127 each year on the same 2 ramets per genotype and sampling zone).
128

129 DNA extraction and genotyping

130 Needle tissues from the 2,552 six-month-old seedlings described above and from the 51 seed orchard
131 parental genotypes (two ramets sampled per genotype) were ground to a fine powder in liquid nitrogen
132 and subjected to DNA extraction with the Qiagen DNeasy® 96 Plant Kit, in accordance with the
133 manufacturer's protocol. The DNA was quantified with a NanoDrop microvolume spectrophotometer
134 (Thermo Fisher Scientific Inc., Waltham, CA, USA) and diluted to 10 ng/µl. DNA samples were then
135 genotyped with the 80 SNP markers developed by Vidal et al. (2015). Genotyping was performed with
136 the Sequenom MassARRAY iPLEX Gold assay (Sequenom, San Diego, CA, USA). SNP markers were

discarded if genotype calling was unsuccessful for more than 10% of the samples or if they deviated from Hardy-Weinberg equilibrium.

Pedigree reconstruction

Pedigree reconstruction was performed by likelihood inference with Cervus 3.0.7 (Kalinowski et al. 2007). Paternity reconstruction analysis was performed when the seed parent was known (SS1 and SS2). By contrast, parental reconstruction analysis was performed when both seed and pollen parents were unknown (SS3). We assumed a 0.1% genotyping error rate (estimation based on repeated genotyping of the 51 parental genotypes). The delta score (i.e. the difference in LOD scores of the two most likely candidate parents) was used as a criterion for the assignment of paternity with 99% confidence. The critical values of delta scores were determined from simulations of 100,000 offspring. We allowed only one mismatched allele between a given offspring and its parents. A seedling was considered to result from pollen contamination (pollination by a pollen grain originating from outside the seed orchard) if no pollen parent from the 51 parental genotypes was found in the paternity reconstruction analysis or if only one parent was identified in the parental reconstruction analysis.

Parental contribution

The paternal contribution for a given genotype was estimated, for SS1 and SS2, as the number of seedlings in which the pollen parent was identified divided by the total number of pollen parents recovered. This estimate was compared to a theoretical paternal contribution to assess the deviation from equal paternal contributions. The theoretical paternal contribution for a given genotype *i* was calculated with weighting according to the number of ramets per CSO, as follows:

$$Theoretical\ contribution_i = \sum_{j=1}^{3} \frac{Ramet_{i,j} * \sum_{i=1}^{51} Pollen_{i,j}}{\sum_{i=1}^{51} Ramet_{i,j}}$$

where: $Ramet_{i,j}$ is the number of ramets *i* in CSO-*j*

$Pollen_{i,j}$ is the number of pollen parents *i* recovered in CSO-*j*

The self-fertilization rate was estimated by dividing the number of seedlings with two identical parental genotypes by the total number of seedlings for which both parental genotypes were recovered.

The significance of frequency differences for contamination rate and parental contribution was estimated with a chi-squared test of homogeneity ($\alpha = 0.05$).

Genetic diversity parameters

As described above, the seed parent genotypes were clonally represented, within the three CSO, by different numbers of ramets. The census number of seed parents (N) per CSO was, therefore, different from the effective number of seed parents (N_{eff}) defined by $N_{eff} = \frac{1}{\sum_{i=1}^{51} p_i^2}$ where pi is the contribution of genotype i (Kang et al. 2001) The effective number was also calculated for the contribution of the pollen parent, initially without considering pollen contamination ($N_{eff,\ pollen\ parent}$) and then considering each immigrant pollen grain as a unique pollen parent ($N_{eff,\ pollen\ parent*}$).

Results

Genotyping and pedigree reconstruction

We retained 60 of the 80 SNP tested for pedigree reconstruction based on the genotyping call restriction and Hardy-Weinberg equilibrium criteria. The mean polymorphic information content (Hearne et al. 1992) per marker was 0.372. The mean non-exclusion probability (Marshall et al. 1998), defined as the probability of not excluding a candidate parent of a given offspring, was 4×10^{-4}, and the probability of genotypes not differing between two randomly chosen individuals was 4×10^{-26}. The 2,552 seedlings were successfully genotyped for 35 to 60 SNP (mean of 56.8 SNP per seedling). After concatenation of the genotyping data obtained from two different ramets, data for 59 to 60 SNP were available for the 51 seed orchard parental genotypes.

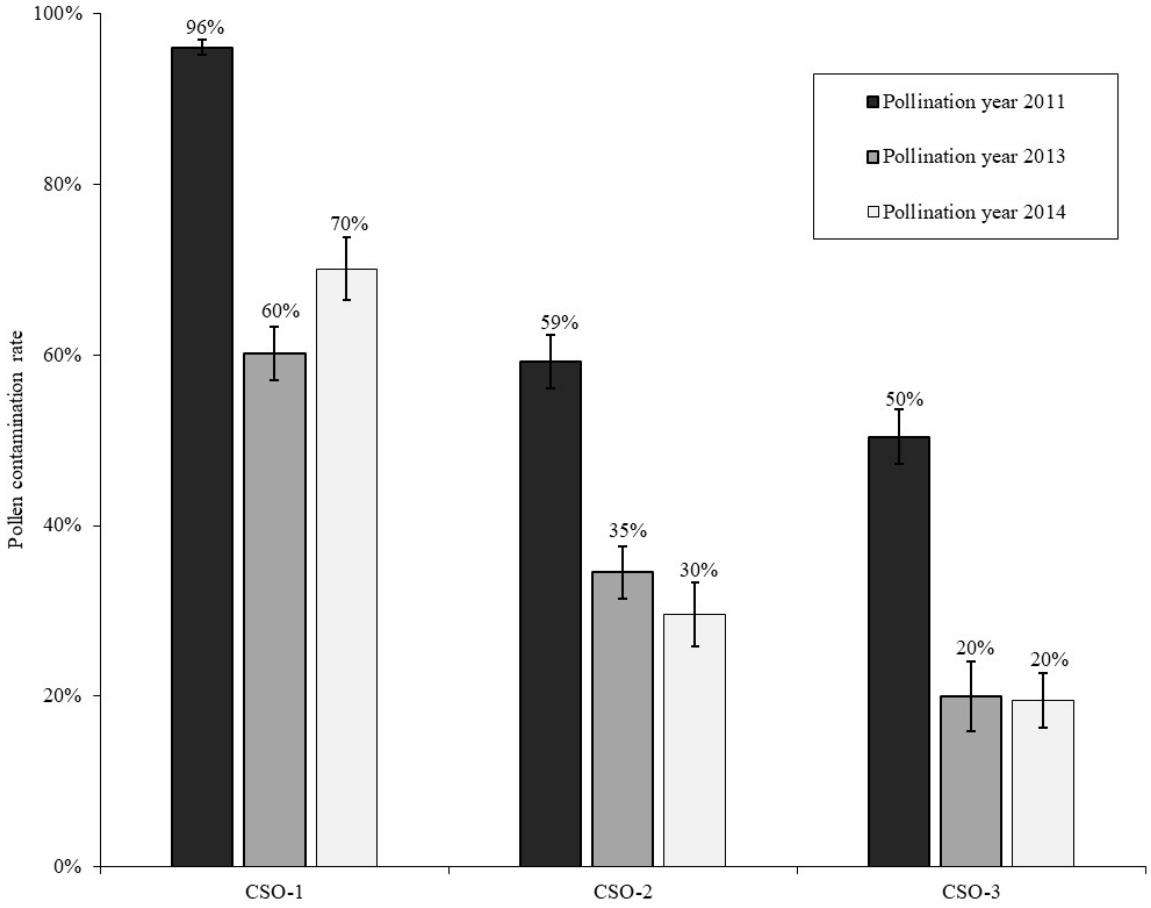
All seed parent identities were confirmed for seeds collected in SS1 and SS2. In addition, based on paternity analyses, a pollen parent was identified from among the 51 parental genotypes for 1,023 of the 2,114 samples collected from a known seed parent (48.4%). These samples included 57 samples arising from the self-pollination of parental genotypes (5.6%). A parental analysis was performed on the commercial seed lots (SS3: 438 samples): only one parent from the 51 parental genotypes was recovered for 174 samples (39.7%) and both parents were identified for the remaining 264 samples (60.3%), 12 of which were generated by self-fertilization (4.5%).

188 Pollen contamination

189 Overall, pollen contamination rates was estimated at 50% for the 2,552 samples analyzed: 558 pollen
190 parents were recovered for 1524 samples in SS1, 465 pollen parents were recovered for 590 samples in
191 SS2 and both seed and pollen parents were recovered for 264 of 438 samples in SS3.
192 Pollen contamination rates are expressed by CSO and by pollination year in Figure 1 (SS1 and SS3).
193 Contamination rates were significantly higher in 2011 than in 2013 and 2014 for a given CSO (no
194 significant difference was found between 2013 and 2014). The pollen parent originated from outside the
195 CSO-1 orchard for 96% of the samples collected in 2011 vs. 60% in 2013 and 70% in 2014. A similar
196 inter-annual trend was observed in CSO-2 (59% vs. 35% and 30%, respectively) and CSO-3 (50% vs.
197 20% and 20%, respectively). Whatever the pollination year considered, contamination rates were
198 significantly higher in CSO-1 than in CSO-2 and CSO-3 (CSO-2 and CSO-3 differed significantly in
199 2013, but not in 2011 and 2014). Within CSO-1, no significant difference was found in contamination
200 rates between the two sampling zones (central vs. border), for either 2011 or 2013 (data not shown).
201 Seed parent identity had no significant effect on contamination rates estimated over the three CSO based
202 on SS1 (62.1%, 61.0%, 61.9% and 68.7% for seed parents A, B, C and D, respectively, see Table 3).

203 CSO-2 pollen contamination rates in 2013 are reported in Figure 2, for each of the 20 seed parents
204 investigated in SS2 and for the seed parents A, B, C, D (SS1). The variability of contamination rates
205 was high and depended on seed parent identity, ranging from 10% to 45%. The mean value over the 24
206 seed parents identities was 25%.

207 **Figure 1.** Pollen contamination rates (%) observed in three maritime pine clonal seed orchards
208 (CSO-1, CSO-2, CSO-3) over three pollination years (2011, 2013, 2014).
209 Pollination years 2011 and 2013 corresponds to the sampling strategy SS1 (seeds collected on the four
210 reference seed parents A, B, C, D). Pollination year 2014 corresponds to the sampling strategy SS3
211 (commercial seedlots from unknown seed parents) (see Table 2). Bars: standard errors.



212

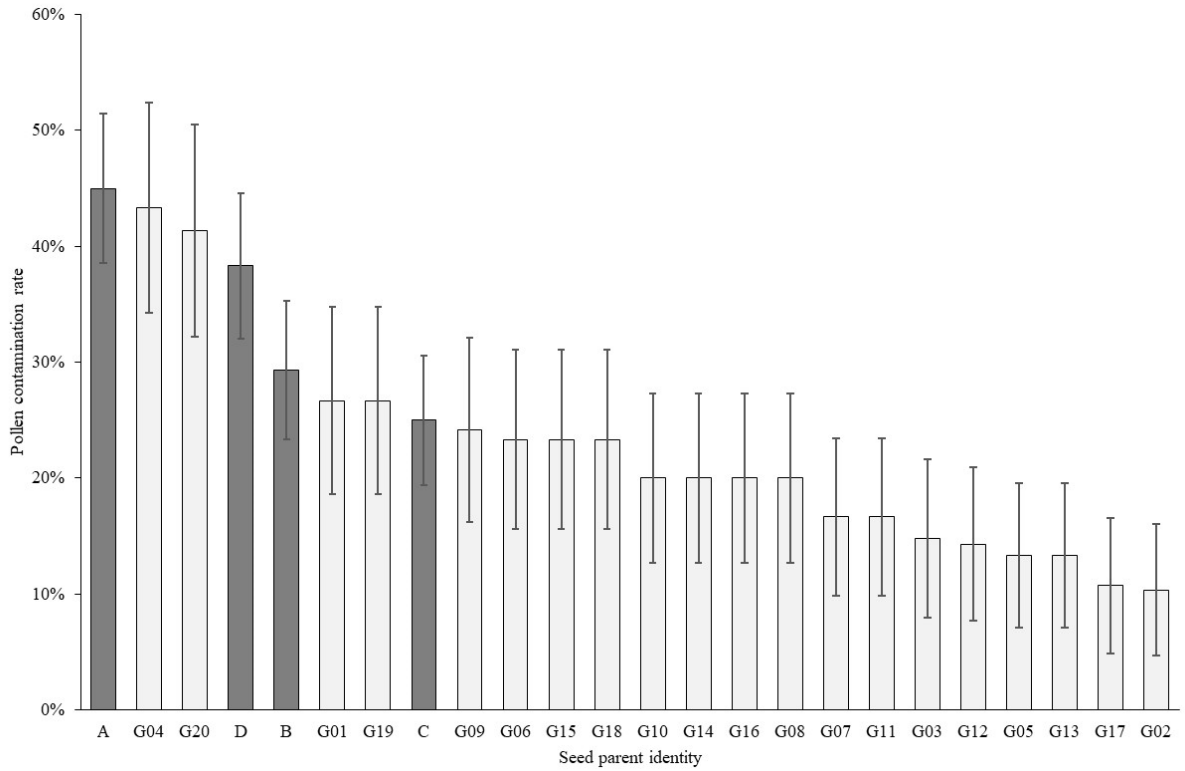
213 **Table 3**. Pollen contamination and self-fertilization rates per reference seed parent genotype estimated
214 through SNP analyses (60 markers) in three maritime pine clonal seeds orchards (CSO-1, CSO-2,
215 CSO-3) over 2011 and 2013 (sampling strategy SS1, see Table 2).

Seed parent identity	**A**	**B**	**C**	**D**
Seed parent flowering phenology	Early	Early	Late	Late
No. seeds analyzed	385	387	378	374
Pollen contamination rate (SE) (%)	62.1 (2.5)	61.0 (2.5)	61.9 (2.5)	68.7 (2.4)
No. self-fertilization	6	6	12	7
Self-fertilization rate (SE) (%)	4.1 (1.6)	4.0 (1.6)	8.3 (2.3)	6.0 (2.0)

216 SE: standard error.

217

218 **Figure 2.** Pollen contamination rates (%) on 24 seed parent genotypes of maritime pine clonal seed
219 orchard CSO-2 pollinated in 2013.
220 Dark grey: four reference seed parents (sampling strategy SS1); Light grey: 20 seed parents (sampling
221 strategy SS2); Bars: standard errors.



222

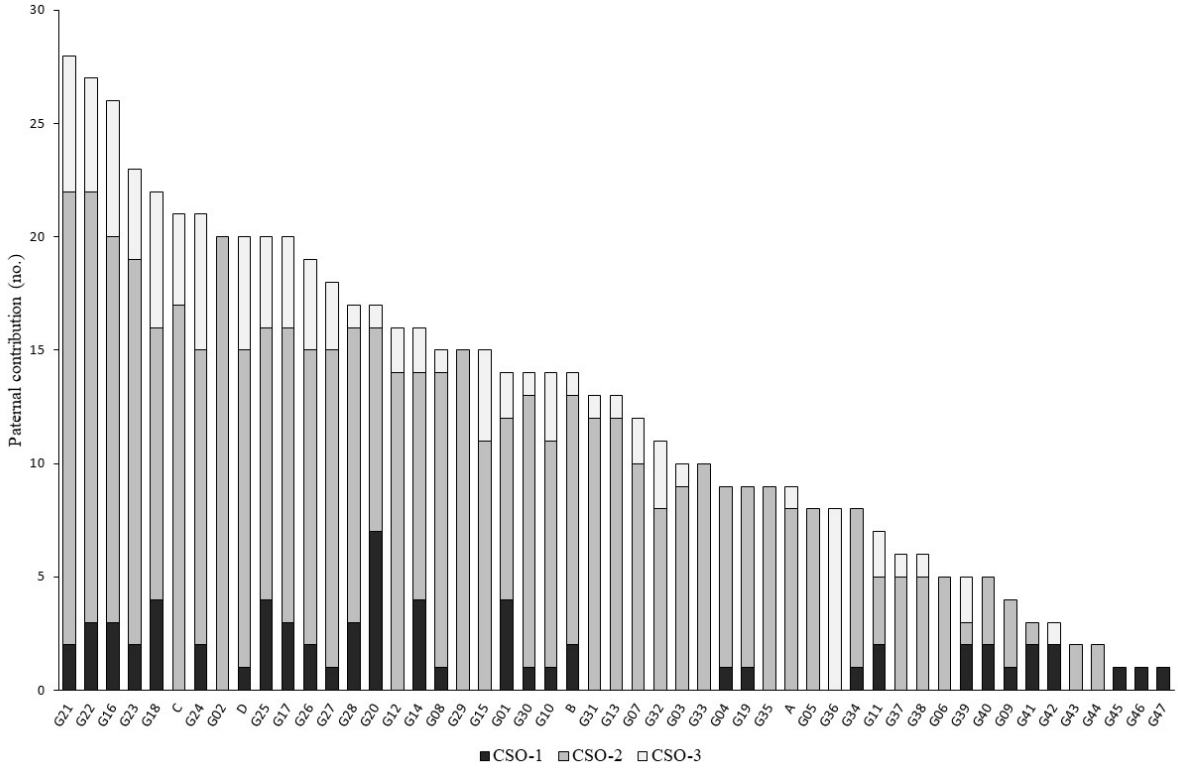
223 Paternal contribution

224 It was possible to estimate paternal contributions only with SS1 and SS2, for which the seed parent was
225 known. These contributions are reported, by CSO, for each genotype in Figure 3. All parental genotypes
226 contributed as pollen parents, but a high degree of heterogeneity was observed (parental genotypes were
227 identified as pollen parents from 1 to 28 times). As SS2 focused exclusively on CSO-2, a larger number
228 of pollen parents were recovered for CSO-2 (719) than for CSO-1 (113) and CSO-3 (191). The number
229 of ramets per genotype and per CSO partly accounted for the heterogeneity of paternal contributions
230 (Figure 4). Paternal contribution was, indeed, significantly correlated with genotype representativeness
231 (expressed as the percentage of ramets per genotype in orchard); Pearson's correlation coefficient was
232 significant and estimated at 0.45 in CSO-1, 0.52 in CSO-2 and 0.48 in CSO-3.

233 Self-fertilization was estimated for all three sampling sets and amounted to 5.4% over the 2,552 samples
234 analyzed. Results for SS1 are reported in Table 3, with no significant differences detected between the
235 four seed parents (selfing rates were 4.1%, 4.0%, 8.3%, 6.0% for seed parent A, B, C, D, respectively).
236 By contrast, in SS2, selfing rates were variable and ranged from 0 to 26% (Figure 5). The rate of self-
237 fertilization was not correlated with the number of ramets per genotype (data not shown).
238

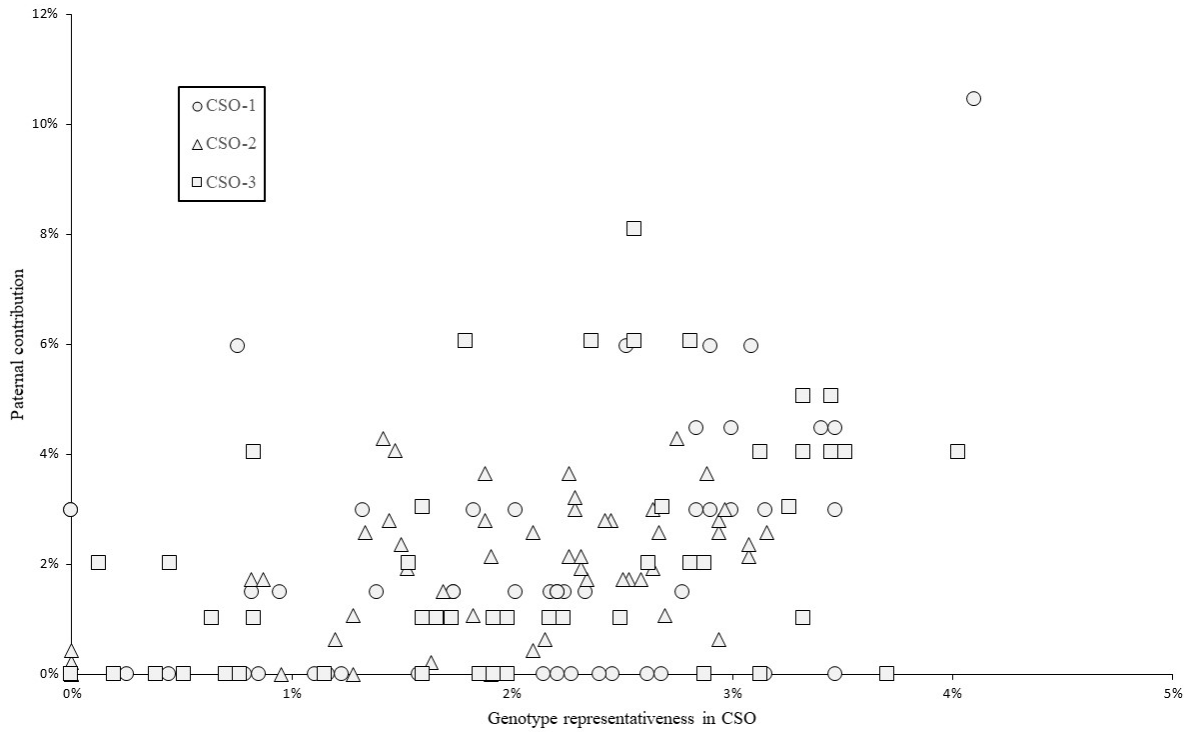
239 **Figure 3.** Paternal contribution (number of pollen parents) of each genotype observed in three maritime

240 pine clonal seed orchards (CSO) over 2 pollination years (2011, 2013, sampling strategies SS1 and SS2).



241

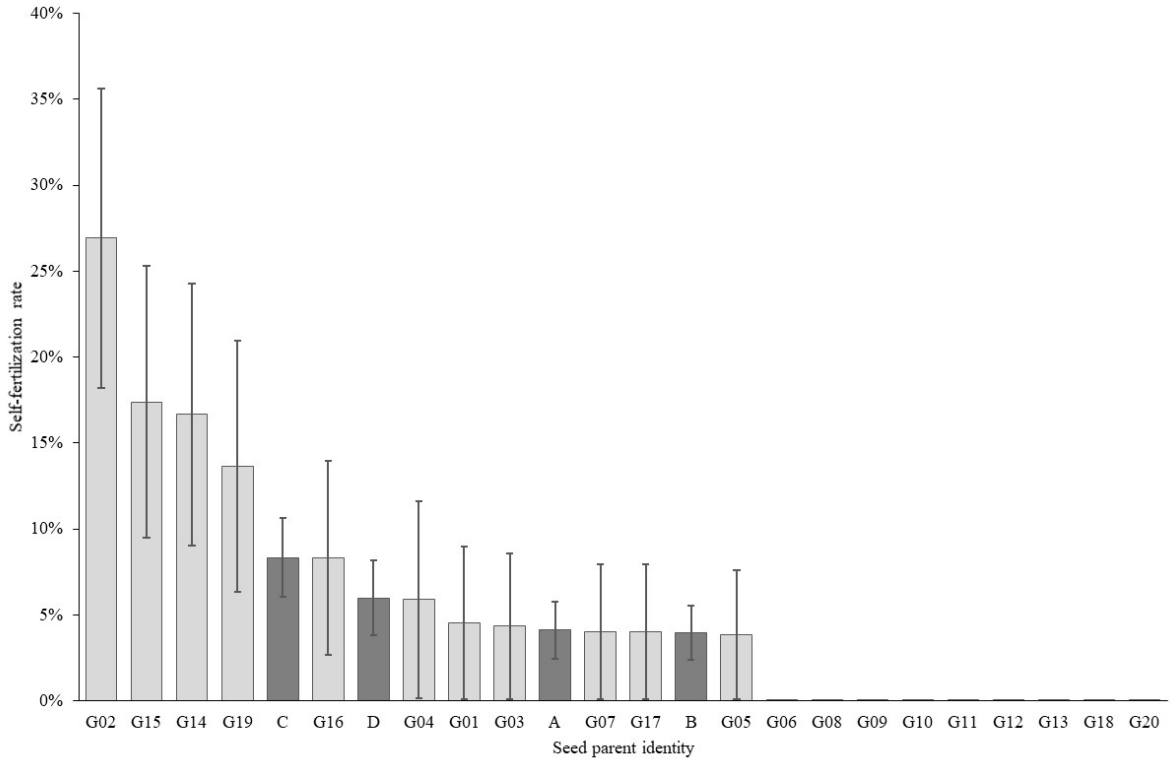
242 **Figure 4**. Paternal contribution (% of pollen parents) according to genotype representativeness (% of
243 total number of ramets) in three maritime pine clonal seed orchards (CSO) over 2 pollination years
244 (2011, 2013, sampling strategies SS1 and SS2).



245

Figure 5. Self-fertilization rates (%) on 24 seed parent genotypes of maritime pine clonal seed orchard CSO-2 pollinated in 2013.

Dark grey: four reference seed parents (sampling strategy SS1); Light grey: 20 seed parents (sampling strategy SS2); Bars: standard errors.



Genetic diversity

Differences between the census number (N) and effective number (N_{eff}) of parental genotypes per CSO resulted from the deployment of variable numbers of ramets per genotype (Table 4). Considering only pollen parents from within the CSO, the low $N_{eff,\ pollen\ parent}$ (14.6 in CSO-1, 31.2 in CSO-2 and 13.1 in CSO-3) reflected a highly heterogeneous paternal contribution, as shown in Figure 3.The consideration of pollen parents from outside the CSO greatly inflated genetic diversity, particularly in CSO-1 ($N_{eff,\ pollen\ parent}$ = 346.9).

259 **Table 4.** Genetic diversity parameters computed from SNP data (60 markers) for three maritime pine
260 clonal seed orchards (CSO-1, CSO-2, CSO-3) over 2011 and 2013 (sampling strategies SS1 and SS2,
261 see Table 2).

Clonal seed orchard	**CSO-1**	**CSO-2**	**CSO-3**
Census no. seed parent (N)	46	47	48
Effective no. seed parents (N_{eff})	39.1	43.1	38.6
Effective no. pollen parents, excluding pollen contamination ($N_{eff,\ pollen\ parent}$)	14.6	31.2	13.1
Effective no. pollen parents, including pollen contamination ($N_{eff,\ pollen\ parent*}$)	346.9	67.4	41.8

262

263 **Discussion**

264 Previous studies (Plomion et al. 2001; Plomion et al. 2005) aiming to estimate pollen contamination in
265 maritime pine seed orchards suffered from a limited power of discrimination due to the small number
266 of microsatellite markers available. Consequently, the reliability of the high rates of pollen
267 contamination estimated in these studies remained questionable.

268 In this study, we used the SNP set developed by Vidal et al. (2015), and demonstrated the power of a
269 set of 60 SNP markers (parental exclusion probability exceeding 99.99%) to estimate pollen
270 contamination rates accurately. In most species, SNP markers have become a tool of choice for
271 parentage analyses (Flanagan and Jones 2019). The main advantages of SNP markers over traditional
272 microsatellites include their availability (SNP resources have been published for most conifer species,
273 including maritime pine, Plomion et al. 2016), low genotyping error rate, and the possibilities for
274 multiplexing and automation of genotyping for high-throughput analyses. However, only a few studies
275 to date implemented SNP markers to study pollen contamination in forest trees (Galeano et al. 2021;
276 Hall et al. 2020; Suharyanto et al. 2012).

We found a mean pollen contamination rate of 50% over the 2,552 seedlings genotyped, consistent with the minimum rate of 36% estimated in maritime pine polycross seed orchards by Plomion et al. (2001), using chloroplast microsatellites. Plomion et al. (2005) subsequently used nuclear microsatellites and revealed similar high pollen contamination rates in second-generation seed orchards (32% to 81%). Medium-to-high levels of pollen contamination have been reported in conifer species: 28% in *Pinus thumbergii* (Suharyanto et al. 2012), 49% in *Cryptomeria japonica* (Moriguchi et al. 2010), 12-35% in *Pseudotsuga menziesi* (Korecký and El-Kassaby 2016; Slavov et al. 2005), 27% in *Picea glauca* (Galeano et al. 2021), 58% in *Picea abies* (Dering et al. 2014), 5-52% in *Pinus sylvestris* (Funda et al. 2015; Torimaru et al. 2009), and 86% in *Pinus brutia* (Kaya et al. 2006).

However, the various seed lots and sampling strategies as well as the large number of seedlings analyzed in this study provided new insights into the mating structure in maritime pine seed orchards.

First, the location of the CSO emerged as a key factor explaining pollen contamination. Indeed, CSO-1 (located within the Landes de Gascogne forest) had higher pollen contamination rates than CSO-2 (outskirts of the forest) and CSO-3 (outside the forest), whatever the year considered. This result can be related to the vicinity of CSO-1 with the potential massive source of external pollen. *In situ* experimental studies in a maritime pine seed orchard located in the Landes de Gascogne forest showed that about 20% of pollen contamination could be explained by distant pollen flows, the remaining 80% being explained by local input within a range of ten to several hundred meters (Baradat et al. 1984; Castaing and Vergeron 1976). In other conifer species, such as Scots pine and loblolly pine, viable pollen can travel very long distances — several hundred kilometers, reviewed by Kremer et al. (2012) — with as much as 4.3% of viable pollen covering distances of more than 100 km in Scots pine. Assuming a similar pollen dispersal profile in maritime pine, CSO-1 would experience massive local pollen flow from the Landes de Gascogne forest, whereas CSO-2 and CSO-3 would potentially receive more limited and distant pollen flow from this source. Another plausible explanation concerns the environmental (pedoclimatic) conditions of the seed orchards, which were located on either sandy soils (CSO-1) or a clay loam soil (CSO-2 and CSO-3). Clay loam soils are known in maritime pine to be associated with the earlier formation of strobili, about 7-10 days ahead of most of the Landes de Gascogne forest located

304 on sandy soils (Table 1). The receptivity of the female strobili in CSO-2 and CSO-3 may therefore be
305 optimal well before the emission of massive amounts of pollen from the Landes de Gascogne forest, in
306 which CSO-1 is located.

307 Second, the age of the seed orchard may also partially account for higher levels of contamination in
308 CSO-1 (planted in 2006) than in CSO-2 and CSO-3 (planted in 2002-2003). The intensity of flowering
309 increases with tree age and becomes optimal for commercial harvesting after about 8-10 years. At the
310 time of first sampling in our experiments (2011), CSO-1 trees were only 5 years old while the CSO-2
311 and CSO-3 trees were 3-4 years older (age 8-9). A lower rate of fertile male and female strobilus is
312 therefore likely in CSO-1 compared to CSO-2 and CSO-3. Internal pollen flows at CSO-1 could be
313 insufficient to compete with external sources and result in very high rates of pollen contamination (96%
314 in 2011). Accordingly, pollen contamination observed 2 and 3 years later in CSO-1 was reduced (60%
315 and 70%). However, contamination levels can remain high in old seed orchards, as highlighted by
316 Torimaru et al. (2009). This is consistent with the rates observed in CSO-2 and CSO-3 which remain at
317 quite high level in 2013 and 2014 (20-35%, age 10-12).

318 Third, our specific sampling in CSO-1 (SS1) revealed that pollen contamination did not vary within the
319 seed orchard. Identical contamination rates in the center and at the edge of CSO-1 were detected in 2011
320 and 2013 confirming the results published by Funda et al. (2015) in Scots pine and suggesting that the
321 whole orchard is subject to homogeneous contamination with outside pollen due to long-distance pollen
322 flows.

323 Fourth, we found a clear effect of pollination year, with 2011 associated with a much higher rate of
324 pollen contamination than 2013 and 2014, whatever the CSO considered. One major explanation may
325 be the changes in the maturity of the trees in these orchards, as discussed above. However, these
326 differences between years may also reflect meteorological factors (such as temperature, rainfall, wind
327 strength and direction), which can affect the formation, persistence and outcome of pollen clouds, as
328 well as the viability of pollen released in spring. Mean daily rainfall during the period of female strobilus
329 receptivity (estimated from control crosses in the framework of the maritime pine breeding program)
330 was 0.5 mm in 2011 vs. 1.9 mm in 2013 and 2.4 mm in 2014 (Météo France data). Dry periods, such as

331 that observed in 2011, favor pollen flow over long distances, as rain is known to affect the extent of
332 pollen dispersal (Di-Giovanni and Kevan 1991).

333 Fifth, the identity of the seed parent also had a significant impact on pollen contamination rate
334 (Figure 2). However, unlike Slavov et al. (2005), who reported higher contamination rates for genotypes
335 with early female receptivity, we found no relationship between the timing of pollen receptivity
336 (estimated during pollination years 2015 and 2016) and pollen contamination (Trontin et al. 2019). Our
337 findings suggest that the female flowering phenology within the orchard had little impact on pollen
338 contamination, probably due to the extended period of pollen release. The four seed parents sampled for
339 SS1 were ranked among the most contaminated seed parents sampled in SS2 (Figure 2). We currently
340 have no explanation for this observation, but it may have biased the pollen contamination rates estimated
341 in 2011 and 2013 upwards in the three CSO.

342 With the sampling strategy used here, based on the collection of cones from specific seed parents, we
343 were unable to study the contribution of the seed parent, but it was possible to analyze paternal
344 contribution based on SS1 and SS2. All pollen parents were recovered at least once in the seed lots
345 genotyped, but a high level of variability was observed for paternal contribution, as also reported by
346 Suharyanto et al. (2012) in *Pinus thunbergii.* The weak correlation between genotype representativeness
347 (based on the number of ramets per genotype) and paternal contribution (Figure 3) suggests that different
348 genotypes released different amounts of pollen. Trontin et al. (2019) suggested that the intensity of
349 pollen production, estimated during pollination years 2015 and 2016 as the density score for male
350 strobili, could partly explain the heterogeneity of paternal contributions.

351 Seed orchard design is optimized to minimize self-fertilization, which leads to inbreeding depression in
352 conifer species. The overall rate of self-fertilization was estimated at 5.4% at the seedling stage, a value
353 below the 13% reported by Baradat et al. (1984) for maritime pine, but within the range of estimates for
354 pine seed orchards (Funda et al. 2015; Suharyanto et al. 2012; Torimaru et al. 2009). As previously
355 reported by Funda et al. (2015), self-fertilization rates depended strongly on seed parent identity and
356 was as high as 26.9% for one genotype in our study.

Conclusion

This study of 2,552 maritime pine seedlings is the largest study to date estimating pollen contamination and mating structure in forest tree seed orchards. Deployment based on open-pollinated CSO is a cost-effective strategy for delivering genetic gains. However, two major assumptions must be satisfied for the objectives for genetic gain and diversity to be fully achieved: i) no contamination with foreign pollen and, ii) random mating between the parental genotypes. Our study reveals a mean pollen contamination rate of 50% and an unbalanced paternal contribution. Based on breeding values estimations, VF3 seed lots from the analyzed CSO should deliver 30% genetic gains for volume and stem straightness relative to unimproved material. Assuming that the foreign pollen comes from unimproved stands, we calculated that the expected genetic gain would fall from 30% to 24%, with an accompanying increase in genetic diversity. Unbalanced parental contributions do not induce a systematic bias in genetic gain, but they do decrease genetic diversity. The level of genetic diversity, initially fixed by the breeder when selecting parental genotypes, is thus subjected to two adverse forces: pollen contamination and unbalanced parental contribution.

Various management practices have been proposed for reducing pollen contamination in forest tree seed orchards. These practices include supplemental mass pollination (Korecký and El-Kassaby 2016; Stoehr et al. 2006), water cooling to delay strobilus production (El-Kassaby and Davidson 1991; Song et al. 2018) and greenhouse-like structures (Funda et al. 2016; Moriguchi et al. 2010; Torimaru et al. 2013). Our study reveals that pollen contamination in French maritime pine CSO can be strongly reduced by: 1) choosing the location of the orchard carefully, in terms of its distance from external pollen sources and the nature of the soil, so as to advance (or delay) strobilus production in the seed orchard, 2) not collecting seeds from young trees. However, some yearly variations in pollen contamination were observed and may be associated with lower rainfall levels during the pollination period.

The methodology used here, based on a set of 60 SNP markers, proved cost-effective and highly powerful for parentage reconstruction. Our results suggest that sampling 100 seeds annually should be sufficient to estimate pollen contamination (this sample size provides estimates with a standard error of

5%) for both applied uses (seed lot quality certification) and for research purposes (e.g. exploring potential links between pollen contamination and climatic data).

Finally, flowering phenology, as well as pollen and cone productivity are known to be under strong genetic control in conifers (Wu et al. 2021) and to define the mating structure within the seed orchard. Our preliminary studies of both flowering phenology and intensity in CSO during two pollination years already support strong genetic control in maritime pine (Trontin et al. 2019). A better knowledge of these flowering traits in the whole breeding population is required, to optimize seed orchard composition and to hone estimates of the expected genetic gain.

Author contributions

LB and JFT conceived and designed the study with expert support of PA, AR and PP. LB coordinated the study. PA collected samples. SD contributed to the molecular laboratory work. LB carried out the genetic analyses and drafted the first version of the manuscript with contributions from SD and JFT. All authors contributed to the discussion and approved the final version of the manuscript.

Data availability

Information about seedlings (CSO, pollination year, sampling strategy, sampling zone, seed parent genotype), SNP description, as well as genotyping data for the 2,552 seedlings and the 51 CSO parents are available from https://doi.org/10.57745/SR2HAJ

Acknowledgments

This study was supported by a national grant (QUASEGRAINE project, French Ministry of Agriculture/DGAL, no. 2014-352, coordinated by ONF/B. Musch) and regional funds from the Conseil Régional d'Aquitaine (IMAF project, no. 12009468-052, coordinated by FCBA/L. Harvengt) and the Conseil Régional Centre Val de Loire (IMTEMPERIES project, no. 2014-00094511, coordinated by INRAE/M.-A. Lelu-Walter). We thank Vilmorin and Forelite for providing access to the orchards and the Maritime Pine Breeding Cooperative (GIS Groupe Pin Maritime du Futur) for its support through the FORTIUS project (grants from the Conseil Régional d'Aquitaine and the French Ministry of

Agriculture, coordinated by INRAE/P. Pastuszka). Samples were prepared at Xylobiotech (https://www.xylobiotech.org/) — the shared FCBA/INRAE platform dedicated to forest biotechnology supported by the ANR (ANR-10-EQPX-16 Xyloforest) — with contributions by Francis Canlet (FCBA Research Technician), Marie Chambard and Mathilde Staat (FCBA trainees). The SNP genotyping was performed at the Bordeaux Genome Transcriptome Facility (doi:10.15454/1.5572396583599417E12). The authors would like to dedicate a special tribute to Marjorie Vidal who passed away in 2021. Marjorie's PhD work (2016, supervision INRAE/FCBA, University of Bordeaux) has notably contributed to the development of SNP arrays for pedigree reconstruction in maritime pine (Vidal et al. 2015).

Declaration of competing interest

There is no conflict of interests to declare.

References

Abad Viñas R, Caudullo G, Oliveira S, de Rigo D (2016) *Pinus pinaster* in Europe: distribution, habitat, usage and threats. In: San-Miguel-Ayanz J, de Rigo D, Caudullo G, Houston Durrant T, Mauri A (Eds), European Atlas of Forest Tree Species. Publi. Off. EU, Luxembourg. https://ies-ows.jrc.ec.europa.eu/efdac/download/Atlas/pdf/Pinus_pinaster.pdf

Baradat P (1987) Méthode d'évaluation de la consanguinité chez les plants issus des vergers à graines de semis de première génération. III- Application à l'optimisation d'une alternative économique aux vergers à graines de clones d'élite: les vergers d'équivalents-clones. Silvae Genet 36(3-4):134-144. https://hal.inrae.fr/hal-03745006

Baradat P, Marpeau A, Bernard-Dagan C (1984) Les terpènes du pin maritime, aspects biologiques et génétiques VI. - Estimation du taux moyen d'autofécondation et mise en évidence d'écarts à la panmixie dans un verger à graines de semis. Ann Sci For 41(2):107-134. https://doi.org/10.1051/forest:19840201

Burdon RD, Carson MJ, Shelbourne CJA (2008) Achievements in forest tree genetic improvement in Australia and New Zealand. Aust For 71(4):263-279. https://doi.org/10.1080/00049158.2008.10675045

Castaing JP, Vergeron P (1976) Étude Expérimentale de la contamination pollinique du verger à graine de pin maritime de Sore (Landes). Ann Sci For 33(3):161-175. https://doi.org/10.1051/forest/19760304

Danusevicius D, Lindgren D (2005) Optimization of breeding population size for long-term breeding. Scand J For Res 20(1):18-25. https://doi.org/10.1080/02827580410019517

Dering M, Misiorny A, Chałupka W (2014) Inter-year variation in selfing, background pollination, and paternal contribution in a Norway spruce clonal seed orchard. Can J For Res 44(7):760-767. https://doi.org/10.1139/cjfr-2014-0061

Di-Giovanni F, Kevan PG (1991) Factors affecting pollen dynamics and its importance to pollen contamination: a review. Can J For Res 21(8):1155-1170. https://doi.org/10.1139/x91-163

El-Kassaby YA, Davidson R (1991) Impact of pollination environment manipulation on the apparent outcrossing rate in a Douglas-fir seed orchard. Heredity 66:55-59. https://doi.org/10.1038/hdy.1991.7

FCBA (2020) Memento FCBA 2020. https://www.fcba.fr/ressources/memento-2020/. Accessed 11 August 2022

Flanagan SP, Jones AG (2019) The future of parentage analysis: from microsatellites to SNPs and beyond. Mol Ecol 28(3):544-567. https://doi.org/10.1111/mec.14988

French Ministry of Agriculture (2022) Statistiques annuelles sur les ventes de graines et plants forestiers. (11 August 2022). https://agriculture.gouv.fr/statistiques-annuelles-sur-les-ventes-de-graines-et-plants-forestiers

Funda T, Wennström U, Almqvist C, Gull BA, Wang X-R (2016) Mating dynamics of Scots pine in isolation tents. Tree Genet Genomes 12:112. https://doi.org/10.1007/s11295-016-1074-z

Funda T, Wennström U, Almqvist C, Torimaru T, Gull BA, Wang X-R (2015) Low rates of pollen contamination in a Scots pine seed orchard in Sweden: the exception or the norm? Scand J For Res 30(7):573-586. https://doi.org/10.1080/02827581.2015.1036306

Galeano E, Bousquet J, Thomas BR (2021) SNP-based analysis reveals unexpected features of genetic diversity, parental contributions and pollen contamination in a white spruce breeding program. Sci Rep 11:4990. https://doi.org/10.1038/s41598-021-84566-2

Goto S, Miyahara F, Ide Y (2002) Identification of the male parents of half-sib progeny from Japanese black pine (*Pinus thunbergii* Parl.) clonal seed orchard using RAPD markers. Breed Sci 52:71-77. http://dx.doi.org/10.1270/jsbbs.52.71

Hall D, Zhao W, Wennstrom U, Andersson Gull B, Wang XR (2020) Parentage and relatedness reconstruction in Pinus sylvestris using genotyping-by-sequencing. Heredity 124(5):633-646. https://doi.org/10.1038/s41437-020-0302-3

Harju AM, Nikkanen T (1996) Reproductive success of orchard and non-orchard pollens during differents stages of pollen shedding in a Scots pine seed orchard. Can J For Res 26:1096-1102. https://doi.org/10.1139/x26-121

Hearne C, Ghosh S, Tood JA (1992) Microsatellites for linkage analysis of genetic traits. Trends Ecol Evol 8(8):288-294. https://doi.org/10.1016/0168-9525(92)90256-4

Jones ES, Sullivan H, Bhattramakki D, Smith JS (2007) A comparison of simple sequence repeat and single nucleotide polymorphism marker technologies for the genotypic analysis of maize (*Zea mays* L.). Theor Appl Genet 115(3):361-71. https://doi.org/10.1007/s00122-007-0570-9

Kalinowski ST, Taper ML, Marshall TC (2007) Revising how the computer program CERVUS accommodates genotyping error increases success in paternity assignment. Mol Ecol 16(5):1099-106. https://doi.org/10.1111/j.1365-294x.2007.03089.x

Kang KS, Harju AM, Lindgren D, Nikkanen T, Almqvist C, Suh GU (2001) Variation in effective number of clones in seed orchards. New For 21(1):17-33. https://doi.org/10.1023/A:1010785222169

Kaya N, Isik K, Adams WT (2006) Mating system and pollen contamination in a *Pinus brutia* seed orchard. New For 31(3):409-416. https://doi.org/10.1007/s11056-005-0876-x

Korecký J, El-Kassaby Y (2016) Pollination dynamics variation in a Douglas-fir seed orchard as revealed by microsatellite analysis. Silva Fenn 50(4):1682. https://doi.org/10.14214/sf.1682

Kremer A, Ronce O, Robledo-Arnuncio JJ, Guillaume F, Bohrer G, Nathan R, Bridle JR, Gomulkiewicz R, Klein EK, Ritland K, Kuparinen A, Gerber S, Schueler S (2012) Long-distance gene flow and adaptation of forest trees to rapid climate change. Ecol Lett 15(4):378-92. https://doi.org/10.1111/j.1461-0248.2012.01746.x

Marshall DF, Slate J, Kruuk LEB, Pemberton JM (1998) Statistical confidence for likelihood-based paternity inference in natural populations. Mol Ecol 7:639-655. https://doi.org/10.1046/j.1365-294x.1998.00374.x

McKeand S, Mullin TJ, Bryam T, White TL (2003) Deployment of genetically improved loblolly and slash pines in the South. J For 101(3):32-37. https://doi.org/10.1093/jof/101.3.32

Moriguchi Y, Taira H, Tani N, Tsumura Y (2004) Variation of paternal contribution in a seed orchard of *Cryptomeria japonica* determined using microsatellite markers. Can J For Res 34(8):1683-1690. https://doi.org/10.1139/x04-029

Moriguchi Y, Yamazaki Y, Taira H, Tsumura Y (2010) Mating patterns in an indoor miniature *Cryptomeria japonica* seed orchard as revealed by microsatellite markers. New For 39(3):261-273. https://doi.org/10.1007/s11056-009-9169-0

Mullin TJ, Lee SJ (2013) Best practice for tree breeding in Europe. Skogforsk, Uppsala, Sweden. ISBN: 978-91-977649-6-4. https://www.skogforsk.se/english/news/2014/best-practice-for-tree-breeding/

Namkoong G, Kang G, Brouard JS (1988) Tree breeding: principal and strategies. Springer, New York. ISBN: 978-1-4612-3892-8. https://doi.org/10.1007/978-1-4612-3892-8

Pâques L (2013) Forest tree breeding in Europe: current state-of-the-art and perspectives. Springer. ISBN: 978-94-007-6146-9. https://doi.org/10.1007/978-94-007-6146-9

Plomion C, Bartholome J, Lesur I, Boury C, Rodríguez-Quilón I, Lagraulet H, Ehrenmann F, Bouffier L, Gion J-M, Grivet D, de Miguel M, de María N, Cervera MT, Bagnoli B, Isik F, Vendramin GG, González-Martínez SC (2016) High-density SNP assay development for genetic analysis in maritime pine (*Pinus pinaster*). Molecular Ecology Ressources 16(2):574-587. https://doi.org/10.1111/1755-0998.12464

Plomion C, Le Provost G, Pot D, Vendramin G, Gerber S, Decroocq S, Brach J, Raffin A, 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(10):1816-1825. https://doi.org/10.1139/x01-115

Plomion C, Léger V, Gerber S, Harvengt L, Trontin J-F, Quoniou S, Canlet F, Alazard P (2005) Développement de marqueurs microsatellites nucléaires chez le pin maritime et utilisation dans le cadre de la traçabilité des lots de graines issus des vergers à graines. Rapport final DERF/DGFAR n°6145801502. https://hal.inrae.fr/hal-03410193/document

Slavov GT, Howe GT, Adams WT (2005) Pollen contamination and mating patterns in a Douglas-fir seed orchard as measured by simple sequence repeat markers. Can J For Res 35(7):1592-1603. https://doi.org/10.1139/x05-082

Song J, Ratcliffe B, Kess T, Lai BS, Korecký J, El-Kassaby YA (2018) Temporal quantification of mating system parameters in a coastal Douglas-fir seed orchard under manipulated pollination environment. Sci Rep 8:11593. https://doi.org/10.1038/s41598-018-30041-4

Stoehr M, Mehl H, Nicholson G, Pieper G, Newton C (2006) Evaluating supplemental mass pollination efficacy in a lodgepole pine orchard in British Columbia using chloroplast DNA markers. New For 31(1):83-90. https://doi.org/10.1007/s11056-004-5398-4

Suharyanto, Nose M, Shiraishi S (2012) Development and application of a multiplex SNP system to evaluate the mating dynamics of *Pinus thunbergii* clonal seed orchards. Mol Breed 30(3):1465-1477. https://doi.org/10.1007/s11032-012-9733-8

Telfer EJ, Stovold GT, Li Y, Silva-Junior OB, Grattapaglia DG, Dungey HS (2015) Parentage reconstruction in *Eucalyptus nitens* using SNPs and microsatellite markers: a comparative analysis of marker data power and robustness. PLoS One 10(7). https://doi.org/10.1371/journal.pone.0130601

Torimaru T, Wang X-R, Fries A, Andersson B, Lindgren D (2009) Evaluation of pollen contamination in an advanced Scots pine seed orchard. Silvae Genet 58(5-6):262-269. https://doi.org/10.1515/sg-2009-0033

Torimaru T, Wennström U, Andersson B, Almqvist C, Wang X-R (2013) Reduction of pollen contamination in Scots pine seed orchard crop by tent isolation. Scand J For Res 28(8):715-723. https://doi.org/10.1080/02827581.2013.838298

Trontin J-F, Alazard P, Debille S, Bouffier L (2019) Flowering traits as a component of reproductive success in maritime pine clonal seed orchards. In: Bonga JM, Park YS, Trontin JF (Eds) Proceedings of the 5th International Conference of the IUFRO Unit 2.09.02 on "Clonal Trees in the Bioeconomy Age: Opportunities and Challenges." Sept.10-15, 2018, Coimbra, Portugal. pp.173-179. https://www.iufro.org/fileadmin/material/publications/proceedings-archive/20902-coimbra18.pdf

Vidal M, Plomion C, Harvengt L, Raffin A, Boury C, Bouffier L (2015) Paternity recovery in two maritime pine polycross mating designs and consequences for breeding. Tree Genet Genomes 11:105. https://doi.org/10.1007/s11295-015-0932-4

Weng Y, Park Y-S, Krasowski MJ, Mullin TJ (2011) Allocation of varietal testing efforts for implementing conifer multi-varietal forestry using white spruce as a model species. Annals of Forest Science 68(1):129-138. https://doi.org/10.1007/s13595-011-0014-1

Wu HX, Ker R, Chen Z, Ivkovic M (2021) Balancing breeding for growth and fecundity in radiata pine (*Pinus radiata* D. Don) breeding programme. Evol Appl 14(3):834-846. https://doi.org/10.1111/eva.13164

Yazdani R, Lindgren D (1991) Variation of pollen contamination in a Scots pine seed orchard. Silvae Genet 40(5-6):243-246. https://www.thuenen.de/media/institute/fg/PDF/Silvae_Genetica/1991/Vol._40_Heft_5-6/40_5-6_243.pdf