HAL
open science

Forward selection in a maritime pine polycross progeny trial using pedigree reconstruction

Marjorie Vidal, Christophe Plomion, Annie Raffin, Luc Harvengt, Laurent Bouffier

To cite this version:

Marjorie Vidal, Christophe Plomion, Annie Raffin, Luc Harvengt, Laurent Bouffier. Forward selection in a maritime pine polycross progeny trial using pedigree reconstruction. Annals of Forest Science, 2017, 74 (1), pp.1-12. 10.1007/s13595-016-0596-8 . hal-01602767

HAL Id: hal-01602767
https://hal.science/hal-01602767v1

Submitted on 21 Feb 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Copyright

Annals of Forest Science (2017) 74: 21
DOI 10.1007/s13595-016-0596-8

ORIGINAL PAPER

Forward selection in a maritime pine polycross progeny trial using pedigree reconstruction

Marjorie Vidal[1,2,3] · **Christophe Plomion**[1,2] · **Annie Raffin**[1,2] · **Luc Harvengt**[3] · **Laurent Bouffier**[1,2]

Received: 2 February 2016 / Accepted: 28 November 2016 / Published online: 21 February 2017
© INRA and Springer-Verlag France 2017

Abstract

• ***Key message*** **Molecular markers were used for paternity recovery in a maritime pine (*Pinus pinaster* Ait.) polycross trial, facilitating forward selection. Different breeding strategies for seed orchard establishment were evaluated by comparing genetic gains and diversity. This work opens up new perspectives in maritime pine breeding.**

• ***Context*** Polycross mating designs are widely used in forest tree breeding to evaluate parental breeding values for backward selection. Alternatively, polycross progeny trials may be used to select the best trees on the basis of individual breeding values and molecular pedigree analysis.

• ***Aims*** This study aimed to test such a forward selection strategy for the maritime pine breeding program.

• ***Methods*** In a maritime pine polycross trial, progeny with higher breeding values for growth and stem straightness was first preselected with or without relatedness constraints. After paternity recovery, the preselected trees were ranked on the basis of their breeding values, estimated from the recovered full pedigree. Finally, the best candidates were selected with three different strategies (forward, backward, mixed) and three levels of coancestry constraints to establish a virtual clonal seed orchard.

• ***Results*** Complete pedigrees were successfully recovered for most of the preselected trees. There was no major difference in expected genetic gains between the two preselection strategies which differed for relatedness constraints. Genetic gains were slightly higher for forward selection than for classical backward selection.

• ***Conclusion*** This seminal study opens up new perspectives for using forward selection within the French maritime pine breeding program.

Keywords Breeding strategy · Polymix breeding · Paternity recovery · SNP markers · *Pinus pinaster* Ait

Handling Editor: Ricardo Alia

Contribution of the co-authors MV sampled plant material and extracted DNA. MV and LB analyzed the data. MV wrote the manuscript, helped by AR, CP, and LB. AR, LB, CP, and LH read and revised the manuscript. LB designed and coordinated the study. All authors read and approved the final manuscript.

Electronic supplementary material The online version of this article (doi:10.1007/s13595-016-0596-8) contains supplementary material, which is available to authorized users.

✉ Laurent Bouffier
laurent.bouffier@inra.fr

Marjorie Vidal
marjorie.vidal@fcba.fr

Christophe Plomion
christophe.plomion@inra.fr

Annie Raffin
annie.raffin@inra.fr

Luc Harvengt
luc.harvengt@fcba.fr

[1] INRA, BIOGECO, UMR 1202, 33610 Cestas, France

[2] University Bordeaux, BIOGECO, UMR 1202, 33400 Talence, France

[3] Biotechnology and Advanced Silviculture Department, Genetics & Biotechnology Team, FCBA, 33610, Cestas, France

1 Introduction

Progeny testing for parental ranking is widely used in forest tree breeding (Zobel and Talbert 1984). In the French maritime pine (*Pinus pinaster* Ait.) breeding program, multisite

polycross progeny trials have been established in the last 20 years for assessing breeding values and ranking female parents for backward selection. The ranking of parents according to the performance of their progeny is particularly important for traits with a low heritability (Falconer and Mackay 1996), such as many traits of interest in forest trees (Burdon and Kumar 2004; Cornelius 1994; Pâques 2013). In such polycross trials, the progeny phenotyped to evaluate parental breeding values is not used to generate the next generation in the breeding population (which is actually selected from the progeny of biparental crosses) or included in production populations (commercial seed orchards). The lack of information about the male parent greatly limits selection of advanced generation. However, progress in molecular genetics, such as the development of highly informative and cost-effective DNA markers, has made new approaches possible in tree breeding. One such new approach, pedigree reconstruction, makes it possible to reconstruct genealogical relationships between individuals, providing opportunities for the development of new breeding strategies. For example, controlled crosses can be replaced by pedigree recovery in open pollinated populations for the estimation of genetic parameters and the prediction of breeding values in a strategy known as "breeding without breeding" (El-Kassaby et al. 2011; El-Kassaby and Lstiburek 2009; Lstiburek et al. 2011). This approach can be used to initiate a tree improvement program without the need for the initial cycle of breeding and testing (Lstiburek et al. 2015). Another approach, "polymix breeding with parental analysis (PMX/WPA)" was developed by Lambeth et al. (2001) and combines controlled crosses and pedigree recovery. These authors proposed the use of molecular markers to identify the parents of potential selection candidates in polycross mating designs for the evaluation of breeding values and the selection of progeny for the next generation in the breeding program. Three different scenarios are presented, depending on the progeny set genotyped: (i) partial population paternity analysis (pedigree analysis only for the best progeny using female general combining ability and individual performance as selection criteria); (ii) full population paternity analysis (pedigree analysis for all progeny); and (iii) full population parental analysis (identities of both female and male parents recovered by molecular marker analysis as the identities of the mothers are not recorded in this scenario to decrease logistical costs). Lambeth et al. (2001) claimed that PMX/WPA was a "viable alternative to full-sib breeding and testing system." Their approach presents several advantages. Polymix crosses are more cost-efficient than full-sib crosses for a given number of parents. They lead to a larger number of recombination events for fewer crosses, and breeding values are more reliably estimated than for other methods, because each individual is crossed with a larger number of parents. If pollens used for the polymix crosses are from trees with high breeding values, the genetic gain from forward selection with PMX/WPA should be greater. Thus, it should be possible to deploy this gain more rapidly than that obtained with classical backward selection. Moreover, mislabeled clones can be eliminated in the genotyping phase, potentially increasing selection efficiency.

We investigated the feasibility of using a forward selection strategy in a maritime pine polycross trial associated with an analysis of the parentage of the progeny. It should be noted that this polycross trial was not designed initially to perform forward selection but backward selection. Forward selection may have two goals: recruitment of the best genotypes for the next generation of the breeding population and the formation of a production population, such as a clonal seed orchard (CSO). This study focuses on selection for the constitution of a CSO.

The successive stages of forward selection strategies studied here are presented in Fig. 1. Candidate trees were preselected in a polycross trial, using two different preselection options without (PS1) and with (PS2) constraints on relatedness. The preselected trees were then genotyped for single-nucleotide polymorphism (SNP) markers. Once their complete pedigrees were recovered, the best individuals were selected for the CSO, using three levels of coancestry constraints (none, status number Ns = 10, Ns = 20). Forward selection strategies were then compared, in terms of possible genetic gains, with backward and mixed (i.e., a combination of both) selection strategies, using the same three levels of coancestry constraints. Finally, ways of optimizing the preselection and final selection options for the establishment of a CSO were considered in the framework of the French maritime pine breeding program.

2 Materials and methods

2.1 Plant material and mating design

In this article, the successive maritime pine breeding populations were named as follows:

- G0 trees, the "plus" trees mass selected from the Landes provenance; they constitute the base population of the French maritime pine breeding program (Illy 1966)
- G1 trees, the selected progeny from G0 trees; they constitute the second generation of the breeding program
- G2 trees, the progeny from G1 progeny trials.

Six polycross progeny trials for the maritime pine breeding program were established from 1994 to 2002 in southwestern France for the prediction of second-generation (G1) parental breeding values. In total, 960 G1 trees were evaluated (as seed donors) within these six polycross progeny trials, each of which took place on three sites. This study focuses on one of these trial sites, established in 1996 (at 44° 42′ 32″ N/0° 46′ 8″ W) for the evaluation of 166 G1 trees as seed parents.

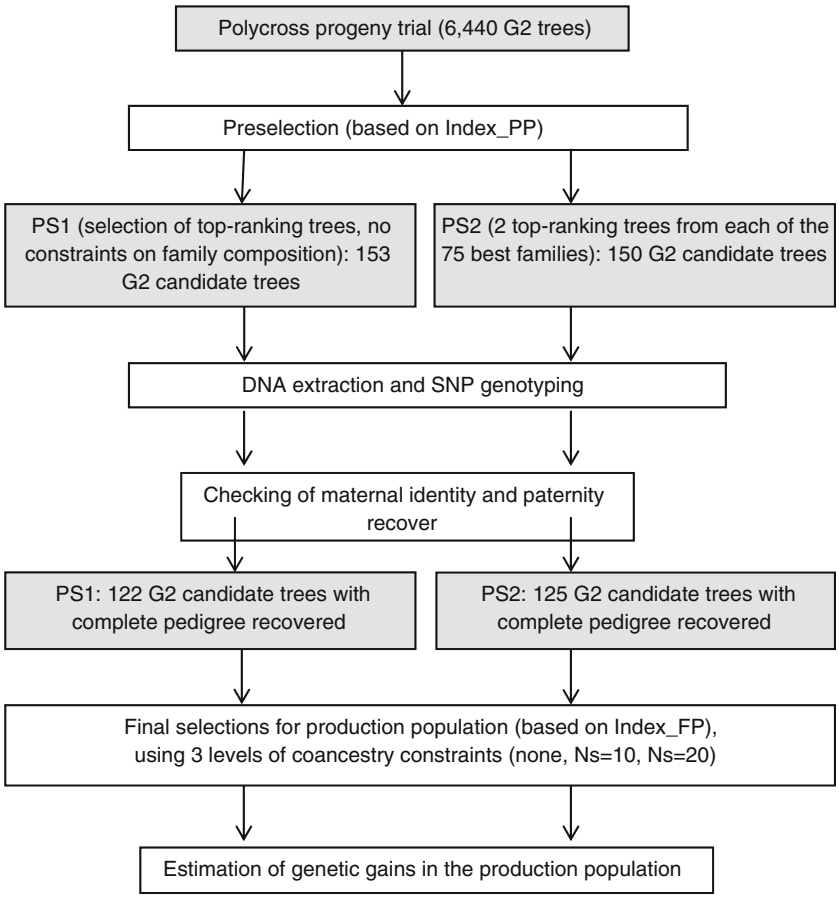

Fig. 1 Main steps in forward selection (with two preselection options) in the maritime pine polycross trial. *Index_PP* is the index calculated with the partial-pedigree information (only the mother identity is known) whereas *Index_FP* is calculated with the full-pedigree information (i.e., after paternity recovery)

Two different pollen mixes were used: 98 seed parents pollinated with one polymix (47 G1 pollen donors, Ns = 19, four pollen donors were also used as females) and 76 seed parents pollinated with the other polymix (43 G1 pollen donors, Ns = 43, ten pollen donors were also used as females). There were no pollen donors common to both polymixes. Eight of the G1 seed parents were pollinated with both polymixes, but the families resulting from identical seed parents and different PMX were considered to be different. The progeny trial thus consisted of 174 half-sib families plus five checklots, planted in a randomized block design, corresponding to a total of 6440 trees (35 complete blocks with one tree plot per family and two trees per checklot in each block).

Progeny (G2 trees) of both polycrosses was phenotyped for selection criteria. Tree girth at breast height (GBH) and tree height (HT) were measured (in cm) at the age of 12 years, and stem sweep (SWE; stem deviation from verticality at 1.5 m from the ground, expressed in cm) was measured at the age of 8 years.

2.2 Breeding value prediction and genetic index

Breeding values for growth (height and girth) and stem sweep were estimated with the TREEPLAN genetic evaluation system (McRae et al. 2004), which includes a database of all available data from the genetic trials of the French maritime pine breeding program. The phenotypic data were first spatially adjusted within each trial. A joint multivariate analysis of all trials based on the best linear unbiased prediction (BLUP) method was then carried out, taking into account both the pedigree relationships between the trees and the correlations between traits. Estimated breeding values (EBVs) and their accuracy were calculated for GBH, HT, and SWE at both measurement ages (i.e., 8 and 12 years). EBVs were also estimated at harvest age (i.e., 35 years), on the basis of age-age correlations for SWE and volume (VOL). Age-age correlations were estimated with the Lambeth correlation model (Lambeth 1980) from multiage data available on other maritime pine trials (unpublished data). The Lambeth coefficient was set at 0.10 for SWE and 0.15 for VOL.

EBVs were obtained with two different pedigree models (before and after pedigree recovery): EBV_PP were calculated with the partial-pedigree model, in which only the theoretical seed donors were known, and EBV_FP were calculated with the full-pedigree model, which included the complete pedigree of the genotyped G2 trees. EBVs are expressed in units of additive standard deviation with the G0 population as the reference population.

Selection decisions were based on multiple-trait selection indexes, combining EBVs calculated at the harvest age.

Index_PP and Index_FP were successively considered depending on pedigree information used to calculate the EBVs.

$$\text{Index_}PP = \text{EBV_PP_VOL} - \text{EBV_PP_SWE}$$

where EBV_PP_VOL and EBV_PP_SWE are the EBVs estimated with the partial-pedigree model for volume and stem sweep at 35 years of age

$$\text{Index_}FP = \text{EBV_FP_VOL} - \text{EBV_FP_SWE}$$

where EBV_FP_VOL and EBV_FP_SWE are the EBVs estimated with the full-pedigree model for volume and stem sweep at 35 years of age.

2.3 Sampling in the polycross trial, with two different preselection strategies

The G2 trees of the progeny polycross trial were ranked according to Index_PP. As this index includes no evaluation of major defects, G2 trees were also scored visually (binary score: 0 for trees with major defects, such as bad branching, forks, disease, or pest damage; 1 for trees without major defects). Trees with a score of 0 were excluded from the preselection process described below.

Two different options were used in the polycross trial to preselect candidates with high growth and low sweep for pedigree recovery. The two options differed in terms of the contribution of the maternal family:

- In preselection 1 (PS1), no restriction was placed on relatedness. PS1 involved the preselection of trees with no major defects ranked among the 200 best individuals (based on Index_PP). In total, 153 G2 trees were sampled.
- Preselection 2 (PS2) included a restriction on relatedness. PS2 involved preselection of the two top-ranking trees with no major defects from each of the 75 best families in the progeny trial. The families and the trees within each family were ranked according to Index_PP. Thus, 150 G2 trees (2 individuals × 75 families) were sampled.

Overall, 57 preselected individuals were common in PS1 and PS2 which means that 246 G2 trees were sampled in total. Young needles were collected from the preselected trees and their potential parents (seed donors and pollen donors of both polymixes) and stored at −80 °C for DNA extraction.

2.4 DNA extraction and fingerprinting

Frozen needle tissues were ground to a fine powder and used for DNA extraction with an Invisorb® DNA Plant HTS 96 Kit (Stratec Molecular, Berlin, Germany), according to the manufacturer's instructions. The DNA was quantified with a NanoDrop microvolume spectrophotometer (Thermo Fisher Scientific Inc., Waltham, CA, USA). The sampled individuals were genotyped with SNP markers, in the *Sequenom MassARRAY* iPLEX Gold assay (Sequenom, San Diego, CA, USA), performed at the genotyping and sequencing facility of Bordeaux, France (http://www.pgtb.u-bordeaux2.fr/). The 80 SNPs used here were originally developed for paternity recovery in a maritime pine breeding population (Vidal et al. 2015). These SNPs were selected from a 12-k Infinium SNP-array (Illumina, San Diego, USA) developed by Chancerel et al. (2013), and each had a minor allele frequency greater than 0.45 and low levels of linkage disequilibrium ($r_v^2 < 0.3$).

2.5 Assignment of parentage for the preselected trees

Likelihood inference was carried out with Cervus 3.0 (Kalinowski et al. 2007; Marshall et al. 1998), both to check the identity of the maternal parent and to recover the identity of the paternal parent for each of the preselected G2 trees. Cervus was run assuming a 0.1% genotyping error rate. The female parent was confirmed if the LOD score (likelihood ratio estimated over all loci, Marshall et al. 1998) was positive, and only one mismatch allele was allowed for each progeny and its supposed female parent. For paternity recovery, 90% of the pollen donors were considered to have been sampled (Vidal et al. 2015). The delta score (i.e., the difference in LOD scores of the two most likely candidate parents) was used as a criterion for paternity assignment at the 99% confidence level. The critical values of delta scores were based on simulations of 100,000 progeny. One mismatch allele was allowed between a given progeny and its male parent.

2.6 Final selection for clonal seed orchard establishment

OPSEL 1.0 software (Mullin 2014) was used for the optimal selection of a production population (virtual CSO), maximizing genetic gains while imposing various constraints on coancestry within the selected population. Constraints on coancestry were based on the minimum status number Ns. The status number of a population describes the effective number of individuals, i.e., the corresponding number of unrelated and non-inbred individuals (Lindgren et al. 1997). Three levels of coancestry constraints were tested: either no restriction on Ns, Ns = 10, or Ns = 20.

The "optimum selection of seed orchard method" was used, allowing unequal numbers of ramets per genotype in the CSO.

The final selection strategies studied were as follows:

- Forward (FOR) selection based on preselection PS1 or PS2: The candidate genotypes were G2 trees for which a

complete pedigree had been recovered. Genetic evaluation was carried out with Index_FP (i.e., with EBVs estimated from the full pedigree model)

– Backward (BACK) selection: all the 166 G1 seed donors evaluated in the polycross trial were candidates. Genetic evaluation was carried out with Index_PP (i.e., with EBVs estimated from the partial-pedigree model)
– -Mixed (MIX) forward-backward selection: G2 trees for which a complete pedigree had been recovered, and all 166 seed donors were candidates. The genetic evaluation was carried out with Index_FP for G1 and G2 individuals.

The target number of selected ramets constituting the CSO was set at 600 (named "census size" in OPSEL). For logistical reasons, the number of ramets per genotype was set at a maximum of 50 for G2 trees and a maximum of 200 for G1 trees (several ramets of G1 trees were available from clonal archives, but this was not the case for G2 trees, with only one tree per genotype, limiting the number of available scions for grafting).

2.7 Estimation of genetic gain for seed orchards

The expected genetic gain (ΔG) was calculated as $\Delta G = \mathrm{CV_a}\sum_{i=1}^{n}\mathrm{EBV}i\ \mathrm{p}i$, where $\mathrm{CV_a}$ is the additive coefficient of variation of the base population (G0 trees), EBVi and pi are the estimated breeding value and the proportion of ramets in the CSO of genotype i, respectively, and n is the number of different genotypes in the CSO.

$\mathrm{CV_a}$ values for height, girth, and stem sweep were extracted from the article by Bouffier et al. (2008) and were calculated as $\mathrm{CV_a} = \sigma_a/\mu$, where σ_a is the square root of the additive genetic variance and μ is the mean value for the trait. Expected genetic gains are expressed as a percentage relative to G0 trees (plus trees) performances.

3 Results

In this study, a breeding strategy was defined as a combination of two selection steps (preselection and final selection) at a given diversity level. Two preselection options (PS1, with no restriction of within-family selection, and PS2, with restriction), three final selection strategies (FOR, BACK, and MIX selection), and three diversity levels (no constraint on Ns, Ns = 10, and Ns = 20) were investigated. The resulting breeding strategies were named according to the combination of these three features. For example, in strategy "FOR_PS1_Ns10," forward selection was performed on the candidate trees from preselection option PS1, with a minimum status number of 10 in the CSO.

3.1 Sampling and genotyping

Two different types of preselection were applied to candidate genotypes in the polycross trial studied: (i) PS1 provided 153 G2 trees from 35 half-sib families (with a family size of 1 to 21 individuals/family), and (ii) PS2 provided 150 G2 trees from 75 half-sib families (2 individuals/family). Genotyping was successfully achieved for 146 (PS1) and 147 (PS2) G2 individuals (minimum = 45 SNPs, maximum = 63 SNPs, mean = 60.5 SNPs), which were analyzed for paternity recovery (Table 1). The dataset (Vidal et al. 2016) is available in the Zenodo repository (http://dx.doi.org/10.5281/zenodo.165158).

3.2 Pedigree recovery on preselected trees

The identity of the maternal parent was not confirmed for one of the 146 individuals in PS1 and five of the 147 individuals in PS2 analyzed for paternity recovery with Cervus software. These individuals were thus excluded from the paternity analysis. The identity of the maternal parent was confirmed for 122 individuals in PS1 and 125 individuals in PS2, and paternity was recovered with 99% confidence for these individuals (see Table 1). In total, 23 individuals in PS1 and 17 individuals in PS2, respectively, were clearly fathered by outside pollen (i.e., not from the two polymixes).

Only individuals for which a complete pedigree was recovered (i.e., the mother confirmed and the father identified) were considered as candidate trees for final selection. All subsequent analyses therefore focus on G2 trees with a complete pedigree. Their pedigree information is summarized in Fig. 2. Overall, PS1 and PS2 G2 candidate trees came from 73 seed donors and 53 pollen donors; 30 seed donors and 34 pollen donors were common to both PS1 and PS2.

The 122 G2 candidate trees from PS1 came from 30 different seed donor clones (maternal contribution of 1 to 13) and from 42 pollen donor clones (paternal contribution of 1 to 8) (Fig. 2). Mean coancestry within these candidate trees was 0.029 (equivalent to Ns = 17). The best seed donor clones contributed more than the others (Fig. 2), because the best preselected trees were from the best maternal families.

The 125 G2 candidate trees from PS2 came from 73 different seed donor clones (maternal contribution of 1 to 2) and from 45 pollen donor clones (paternal contribution of 1 to 10) (Fig. 2). Mean coancestry within these candidate trees was 0.017 (equivalent to Ns = 29).

3.3 Correlation between the EBV_PP and the EBV_FP of candidate trees

There was a strong correlation between the breeding values estimated with the partial (EBV_PP) and full (EBV_FP) pedigree models for the three traits in candidate trees from PS1 and PS2. Pearson's correlation coefficients ranged

Table 1 Number of G2 trees sampled with two preselection options (PS1 and PS2) and paternity recovery statistics

	Preselection PS1	Preselection PS2
Number of preselected G2 trees	153[a]	150[a]
Number of successfully genotyped G2 trees	146	147
Number of G2 trees with seed donor clone confirmed[b]	145	142
Number of G2 trees with complete pedigree recovered	122	125

[a] Fifty-seven preselected G2 were common in PS1 and PS2

[b] and thus analyzed for paternity recovery

from 0.72 to 0.79 for candidate trees from PS1 and from 0.81 to 0.86 for candidate trees from PS2, depending on the trait considered (Fig. 3). These correlation coefficients were slightly higher in the sample from PS2, probably because the range of EBVs was larger in PS2 than in PS1, due to the limitation of relatedness in PS1.

Moreover, the accuracy of EBV_FP (0.70 for girth, 0.74 for height, and 0.75 for stem sweep) was much higher than that of EBV_PP (0.56 for girth, 0.63 for height, and 0.65 for stem sweep). As expected, paternity recovery resulted in a better EBV estimation for G2 trees.

3.4 Final selection to establish clonal seed orchards

The last step in the selection process was the selection of the best genotypes for which a complete pedigree had been recovered, from the PS1 and PS2 candidate trees, to obtain a virtual CSO. OPSEL software was used to obtain an optimal selection, maximizing genetic gain while maintaining genetic diversity by imposing a constraint on mean relatedness (a minimum status number). The optimization of CSO composition by OPSEL resulted in different numbers of ramets for different genotypes. Genotypes with higher breeding values tended to be represented by larger numbers of ramets, but this trend was counterbalanced by relatedness between these genotypes.

Forward selection strategies were evaluated with OPSEL and compared with backward and mixed selection strategies, as explained in the materials and methods section. Detailed results are presented in Supplemental Data I and summarized in Fig. 4.

Forward selection with the two preselection options was first studied, with the imposition of different levels of coancestry in the CSO: either no restriction on Ns, Ns = 10, or Ns = 20. In

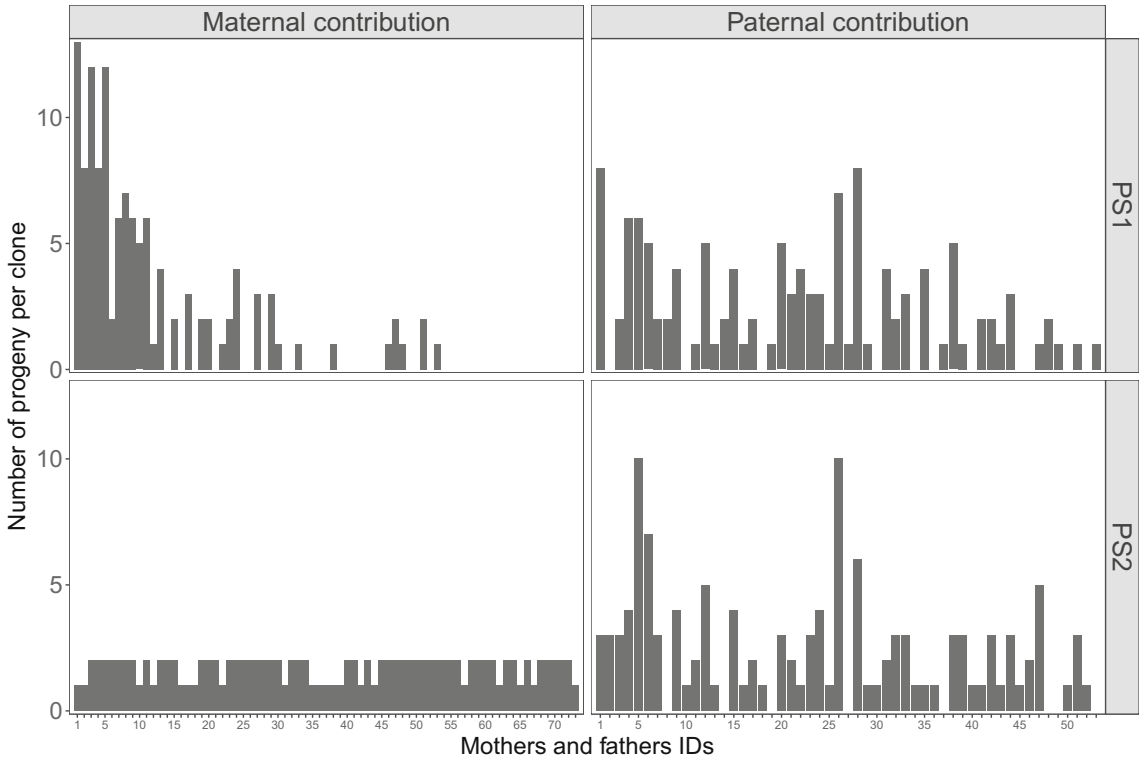

Fig. 2 Maternal and paternal contributions (number of progeny per parent) for candidate trees for which a complete pedigree was recovered, for preselection options *PS1* (122 G2 candidates) and *PS2* (125 G2 candidates). The identities of the maternal (73 different maternal genotypes in total) and paternal parents (53 different paternal genotypes in total) are listed in decreasing order of Index_PP. A maternal family size of 1 in PS2 means that one of the two preselected G2 individuals has an unknown father (i.e., not included in the polymixes)

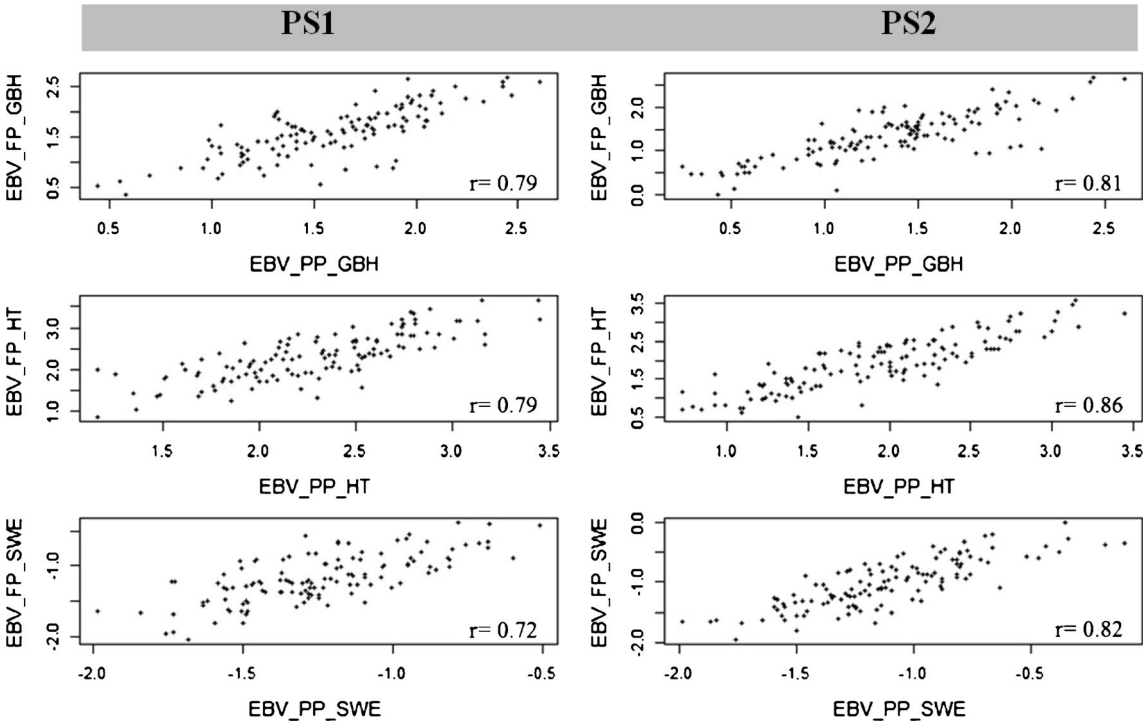

Fig. 3 Correlation between *EBV_PP* (estimated with the partial-pedigree model) and *EBV_FP* (estimated with full-pedigree model) of the G2 candidate trees preselected with *PS1* or *PS2* for girth (*GBH*) and height (*HT*) at 12 years and for stem sweep (*SWE*) at 8 years, where *r* is the Pearson product moment correlation coefficient

FOR selection without restriction on Ns, the 12 best genotypes (Top_12) were selected (ranking on Index_FP) with either the PS1 or the PS2 option. Each genotype contributed equally to the CSO, with 50 ramets per genotype (600 in total). PS1 and PS2 gave similar expected gains for height and girth (additional gains of about 16% for HT and GBH; Fig. 4) but PS1 resulted in lower diversity (Ns = 5 in strategy FOR_PS1_Top12 whereas Ns = 7 in strategy FOR_PS2_Top12). In FOR selection with constraints on coancestry (Ns = 10 or 20), different genotypes contributed different numbers of ramets. As expected, increasing the minimum target status number increased the number of genotypes selected: 37 genotypes (with strategy FOR_PS1_Ns10) or 20 genotypes (with strategy FOR_PS2_Ns10) contributed to the CSO with 2 to 50 ramets

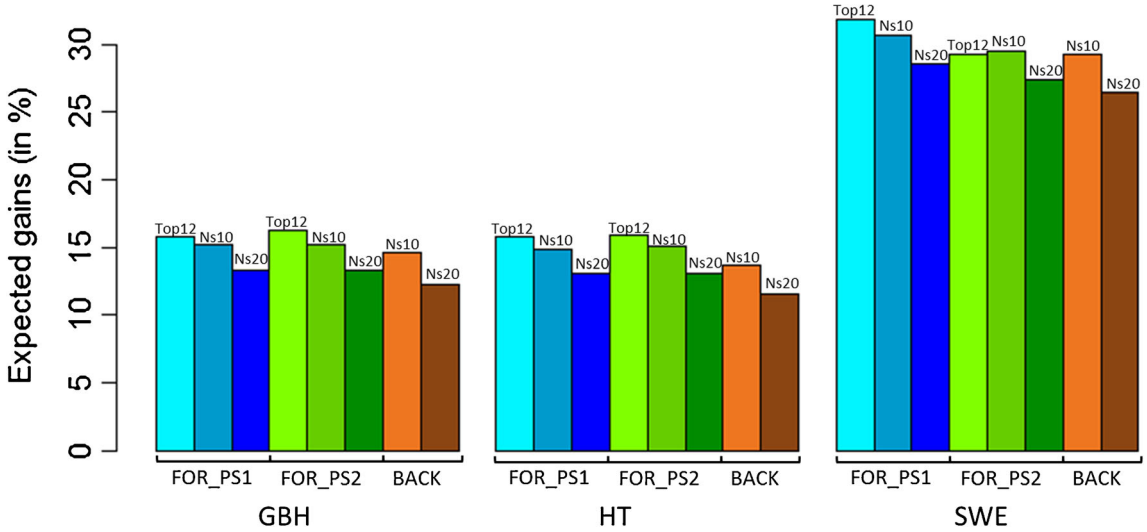

Fig. 4 Expected genetic gain (in %, relative to the base population G0) for girth (*GBH*), height (*HT*), and stem sweep (*SWE*; here, positive gains for SWE indicate greater stem straightness) with different breeding strategies: forward selection with PS1 (*FOR_PS1*, *in blue*) or PS2 (*FOR_PS2*, *in green*) and either no restriction on Ns (*top12*), Ns = 10, or Ns = 20; backward selection with Ns constraint (*BACK_Ns10* and *BACK_Ns20*, *in orange*). Selection was optimized with OPSEL software

per genotype, whereas 77 (with strategy FOR_PS1_Ns20) or 57 (with strategy FOR_PS2_Ns20) genotypes contributed to the CSO with 1 to 31 ramets per genotype. The expected genetic gain was decreased slightly with increasing strength of constraint on diversity (Ns), regardless of the preselection option (Fig. 4).

Forward selection was then compared with backward selection. All 166 seed donors (G1) evaluated in the polycross trial were candidates for selection on the basis of their Index_PP (no pedigree recovery for classical backward selection). Forward selection provided a slightly greater genetic gain than backward selection at equivalent Ns values (Fig. 4). For example, FOR_PS1_Ns10 gave an additional gain of 1.4% for SWE, 0.5% for GBH, and 1.2% for HT (equivalent to an additional gain of 2.2% for volume) over BACK_Ns10. However, the expected gain was more reliable in backward selection than in forward selection. The mean EBV accuracy for G1 trees (in backward selection) was about 0.95, whereas that for G2 trees (in forward selection) was about 0.73.

Finally, mixed selection strategies were evaluated. In this case, 37 genotypes (31 G2 and 6 G1) were involved in the CSO for the MIX_PS1_Ns10 strategy, and 26 genotypes (19 G2 and 7 G1) were involved in the CSO for the MIX_PS2_Ns10 strategy. The two preselection options provided equivalent genetic gains at equivalent Ns values. Moreover, mixed selection provided gains similar to those achieved with forward selection at equivalent Ns values. For example, the MIX_PS1_Ns10 strategy yielded an additional gain of 1.4% for SWE, 0.3% for GBH, and 0.2% for HT (equivalent to an additional gain of 0.8% for volume) over the FOR_PS1_Ns10 strategy.

4 Discussion

The main objective of this study was to assess the feasibility of forward selection associated with parental analysis of the progeny in an existing maritime pine polycross progeny trial in order to accelerate the breeding cycles. A few theoretical studies have been carried out but, to our knowledge, this is the first example of a practical study of forward selection in a polycross trial. Different options for the forward and classical backward selection of a production population (establishment of a virtual CSO) were studied and compared on the basis of genetic gains for growth traits and stem straightness.

The various stages of forward selection, and some considerations about the PMX/WPA strategy, are discussed below.

4.1 Preselection options in a polycross trial

The genotyping of all individuals in a progeny trial is currently too costly, so the preselection of trees is a necessary step before paternity recovery. This step must provide candidate trees for the final selection with two goals: maximizing genetic gain while limiting the relatedness between candidates to ensure that the CSO population contains sufficient diversity.

The identities of the pollen donors were unknown. Consequently, one limitation of this approach was that the set of preselected candidates may not have included some of the best individuals from the polycross trial due to inaccurate EBV estimations (obtained with the partial-pedigree model). Nevertheless, we showed that between EBVs estimated with the partial and full-pedigree models were highly correlated. The ranking of G2 trees on the basis of their Index_FP would therefore have been relatively similar to that obtained with Index_PP if pedigrees had been determined for all the trees.

The preselection of candidate trees may affect final genetic gain and diversity in the CSO. We therefore considered two contrasting preselection options, one with (PS2) and the other without (PS1) restrictions on relatedness between the preselected candidates. In this study, the choice of PS1 or PS2 had little effect on the final selection, as these options yielded similar genetic gains at equivalent Ns values, mostly because the number of preselected individuals was high, and the bias in the EBVs estimated with the partial-pedigree model was small. The same number of individuals was sampled in PS1 and PS2, but the mean coancestry (calculated with complete pedigree information) was, as expected, higher for PS1 than for PS2. Thus, for equivalent genetic gain and diversity in the CSO, PS1 resulted in the selection of a larger number of different genotypes, with fewer ramets per clone required than PS2. PS1 was therefore more logistically efficient, as fewer scions per tree were required. Thus, PS1 seems to be the most appropriate preselection approach for our breeding program, and it does not seem to be necessary to apply constraints on relatedness between preselected individuals, provided that enough trees are preselected.

4.2 Genetic gain and diversity in commercial seed orchards

A large proportion of the planting material for cultivated forests today originates from seed orchards. For maritime pine, more than 90% of the seedlings used for the reforestation of the Landes in Gascony are improved seedlings originating from seed orchards (GIS PMF 2014). Seed orchards consist of selected superior individuals, and the main objective of their establishment is to generate genetically improved forest tree seeds by maximizing genetic gain (Funda and El-Kassaby 2012). The challenge for tree breeders is thus to create seed orchards in which breeding progress is maximal (maximum performance), but with a sufficient degree of genetic diversity to ensure a reasonable degree of genetic heterogeneity in the final forest (Hosius et al. 2000; Lindgren et al. 2009; Stoehr et al. 2004). Genetic diversity plays an important role in the sustainability of forest ecosystems and is essential for a

population to adapt to new environmental factors, such as climate change and diseases (Hansen 2008; Johnson and Lipow 2002; Muller-Starck 1995). How much genetic diversity should be present in a CSO depends on the length of the rotation and the environmental variation to which the planting material originating from the CSO will need to adapt during its lifetime (Johnson and Lipow 2002). Johnson and Lipow showed that a seed orchard with "25 unrelated selections contains about 92 percent of the genetic variation of the natural population" and that a minimum of "20 unrelated selections should provide the same level of risks as seed collected from the natural population." Moreover, restrictions on relatedness between the individuals selected for the CSO can limit inbreeding depression, with potential effects on the performance of the planting material (Durel et al. 1996; Olsson et al. 2001; Stoehr et al. 2008). However, the management of diversity and relatedness (expressed as group coancestry here) between selections becomes relatively complicated at the third generation of breeding. OPSEL software proposes an optimal selection, "not to completely avoid kinship, but rather to find the set of selections that maximizes gain under a relatedness constraint" (Mullin 2014). In this study, we used the status number Ns to quantify coancestry in the CSO. Ns is a useful parameter for evaluating trade-offs between gain and diversity (Lindgren et al. 1997; Lindgren and Kang 1997; Lindgren and Mullin 1998). The minimum Ns was set at 10 or 20 for a population census size of 600 (total number of ramets in the CSO).

In the French maritime pine breeding program, the establishment of a CSO based on forward selection could involve the best G2 trees from several polycross progeny trials. Our standard CSO area is at least 10 ha, so about 2400 grafted trees would be required. Four polycross trials are currently available for forward selection, so the selection of the best genotypes providing 600 ramets within a trial was considered here. Due to the relatedness between the parents used in the different polycross trials, we set the diversity in the studied trial at Ns = 10 to ensure that the minimum diversity required was attained (as described above and in accordance with Johnson and Lipow 2002) in the final complete CSO.

The number of ramets per genotype is limited in forward selection approaches, because each selected clone is represented by a single tree (giving few scions). By contrast, in backward selection, the parent trees selected are often grafted with several replicates in clonal archives. The development of efficient vegetative propagation methods (such as micropropagation, somatic embryogenesis, or micrografting *through tissue culture*) would increase the number of ramets available for the best genotypes and provide powerful tools for scaling up the production of genetically improved planting material (Bonga 2015; Lelu-Walter et al. 2013). However, such methods are not yet available for use in this species (and were therefore not considered in our options).

Whatever the preselection option used, forward selection resulted in a slightly higher genetic gain than backward selection. It should be borne in mind that the polycross trial studied was not designed with forward selection in mind. In particular, the pollen mixes were mostly of random composition rather than based on high EBVs. The expected genetic gain obtained with forward selection in this trial would therefore be far from optimal.

Finally, genetic gain and diversity in the production population were estimated under an assumption of random mating, equal reproductive success, and no pollen contamination within the CSO. However, many factors can affect the genetic quality of orchard seedlots. Both genetic gain and diversity depend on the variation of reproductive success in the CSO, synchrony in reproductive phenology, pollen quality and contamination, self-fertilization rates, seed germination, and other factors. Many studies have shown that there can be a considerable gulf between expectations and reality (Askew 1988; Burczyk et al. 1997; Edwards and ElKassaby 1996; Funda et al. 2009; Gomory et al. 2003; Hansen 2008; Kang and Lindgren 1998; Machanska et al. 2013; Matziris 1994; Na et al. 2015), making it difficult to predict genetic quality. Moreover, the selfing rate in the CSO and inbreeding depression were not taken into account in the estimation of genetic gain and diversity. However, absolute values were not of prime importance here as the aim was to compare different breeding strategies.

4.3 Towards the implementation of a PMX/WPA strategy?

This study shows that forward selection associated with molecular pedigree analysis of progeny is feasible in real-life conditions. In the polycross trial analyzed here, the expected genetic gain in the production population did not much exceed that obtained with classical backward selection, largely because this polycross trial was not designed for this kind of selection.

The implementation of a PMX/WPA strategy, as proposed by Lambeth et al. (2001) is a broader issue, because the best progeny is selected for the next generation of breeding, rather than just for a CSO. This strategy has clear advantages, including the need for only one round of crossing (polycross), for simultaneous testing and recruitment for forward selection. The many costly full-sib crosses required for classical approaches are replaced by a small number of polycrosses (with one or several different polymixes), followed by genotyping and paternity recovery to identify the best progeny for use in the next generation of breeding. This approach is thus easier to implement than classical approaches, and, as breeding and testing are performed at the same time, the interval between the generations of consecutive seed orchards is shortened

(Fig. 5). In our current breeding cycle, because of technical and economic constraints (number of crosses possible in each year) and because the results of polycross trials are used to choose parents for full-sib crosses, there is a time lag between polymix crosses (for parental EBV estimation) and full-sib crosses (used to select the next generation for breeding). This time lag can be estimated to around 12 years for the French maritime pine breeding program. The use of a PMX/WPA strategy would eliminate this time lag (Fig. 5), accelerating the breeding cycle. Improved planting material can be renewed more quickly and adapt better to changing economic and environmental contexts.

Polycrosses also maximize the number of full-sib families obtained with a smaller number of crosses than for full-sib designs. However, the combinations of parents are not precisely chosen. The parents transferring their genes to the next breeding generation are therefore determined, to some extent, at the time of selection, but not at the time of the cross. However, by contrast to the "breeding without breeding" approach (pedigree identification from open pollination rather than controlled crosses), the potential fathers are at least partially selected through the choice of polymix composition.

The successful use of polymix breeding and testing systems requires accurate pedigree reconstruction, small differences in male reproductive success (to prevent the difficulties involved in managing coancestry within the preselected subset in situations in which the best fathers contribute more to the progeny than others), and a low rate of pollen contamination (because only trees with a full-pedigree can be selected). Vidal et al. (2015) recently showed that these requirements are fulfilled in the maritime pine polycross progeny trial studied here. Additional studies are required to optimize the polymix trial design (e.g., composition and number of polymixes) in particular. A relatively large number of pollen donors with low levels of relatedness and high EBVs is required within the polymix to ensure that there will be sufficient genetic diversity in the next generation of the breeding population.

5 Conclusion

This study shows that forward selection with pedigree reconstruction is feasible for maritime pine. Complete pedigrees were recovered for most of the preselected (and thus genotyped)

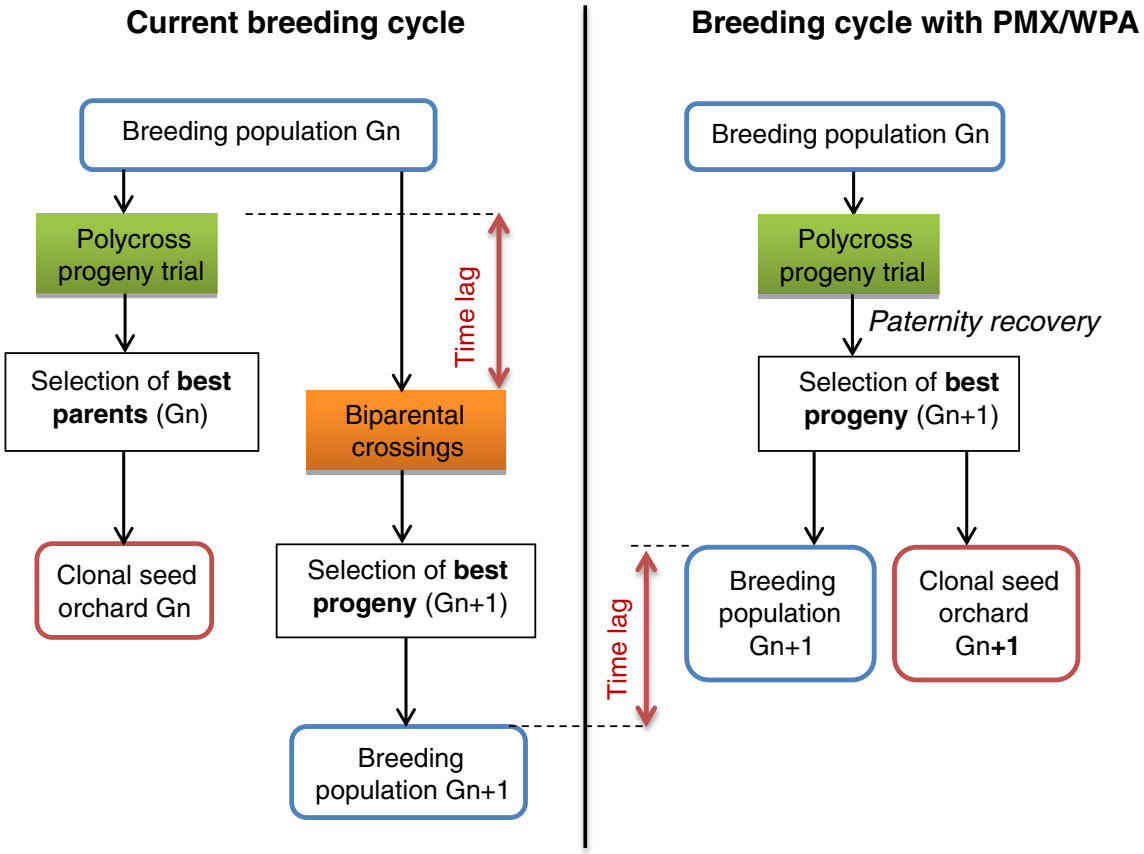

Fig. 5 Comparison of the current breeding cycle based on backward selection and a breeding cycle based on forward selection as in a "polymix breeding with parental analysis" (PMX/WPA) strategy. A breeding cycle based on a PMX/WPA strategy runs faster only because it eliminates the time lag between polymix crossing and biparental crossing. This time lag reaches 12 years old in the French maritime pine breeding program. G*n* and G*n* + 1 are generations *n* and *n* + 1 of the breeding population, respectively

progeny, a prerequisite for selection for the production population. In the polycross progeny trial analyzed, forward selection gave a slightly greater genetic gain (despite the absence of optimization of polymix composition) than classical backward selection. No major differences in expected genetic gain in the production population were observed between two contrasting preselection options (with and without constraints on relatedness).

The implementation of a PMX/WPA strategy, speeding up the production of the next breeding population and decreasing the workload, would be possible. However, simulation studies are required to optimize the general design of such breeding strategies, and a cost/benefit analysis should be performed to assess their economic efficiency, given the specific cost and time components of our maritime pine breeding program.

Forward selection also provides a favorable context for genomic selection. Indeed, the additive genetic relationship matrix (derived from the pedigree) could be replaced with a genomic relationship matrix to improve the estimation of EBVs and to ensure the maintenance of higher levels of genetic diversity within the breeding program.

Acknowledgments This study would not have been possible without the support of the maritime Pine Breeding Cooperative (GIS "Pin Maritime du Futur"). We gratefully acknowledge all its members. The authors also thank the INRA Experimental Unit (UE0570) for field measurements, Jean-Mathieu De Boisseson (FCBA) for needles sampling, Tim Mullin for providing access to OPSEL Software, and Jérôme Bartholomé (INRA) for useful advices in R.

The genotyping was performed at the Genomic Facility of Bordeaux (grants from the *Conseil Regional d'Aquitaine*, nos. 20030304002FA and 20040305003FA; the European Union, FEDER no. 2003227; and ANR, no. ANR-10-EQPX-16 Xyloforest), with help from Christophe Boury and Adline Delcamp (INRA).

Compliance with ethical standards

Data availability The dataset analyzed during the current study is available in the Zenodo repository [http://dx.doi.org/10.5281/zenodo.165158].

Funding This study was funded by INRA (EFPA division "projet innovant"), the European-Union (ProCoGen project: no. 289841), and *Conseil Regional d'Aquitaine* (IMAF project cofunded by FCBA: no. 120009468-052). Marjorie Vidal received a CIFRE Ph.D. fellowship (Public/Private Research Partnerships between FCBA and the French Ministry of Higher Education and Research).

References

Askew GR (1988) Estimation of gamete pool compositions in clonal seed orchards. Silvae Genet 37:227–232

Bonga JM (2015) A comparative evaluation of the application of somatic embryogenesis, rooting of cuttings, and organogenesis of conifers. Can J For Res 45:379–383. doi:10.1139/cjfr-2014-0360

Bouffier L, Raffin A, Kremer A (2008) Evolution of genetic variation for selected traits in successive breeding populations of maritime pine. Heredity 101:156–165. doi:10.1038/hdy.2008.41

Burczyk J, Nikkanen T, Lewandowski A (1997) Evidence of an unbalanced mating pattern in a seed orchard composed of two larch species. Silvae Genet 46:176–181

Burdon RD, Kumar S (2004) Forwards versus backwards selection: trade-offs between expected genetic gain and risk avoidance. N Z J Forest Sci 34:3–21

Chancerel E, Lamy JB, Lesur I, Noirot C, Klopp C, Ehrenmann F, Boury C, Le Provost G, Label P, Lalanne C, Leger V, Salin F, Gion JM, Plomion C (2013) High-density linkage mapping in a pine tree reveals a genomic region associated with inbreeding depression and provides clues to the extent and distribution of meiotic recombination. BMC Biol 11:19. doi:10.1186/1741-7007-11-50

Cornelius J (1994) Heritabilities and additive genetic coefficients of variation in forest trees. Can J For Res 24:372–379. doi:10.1139/x94-050

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

Edwards DGW, ElKassaby YA (1996) The biology and management of coniferous forest seeds: genetic perspectives. For Chron 72:481–484

El-Kassaby YA, Cappa EP, Liewlaksaneeyanawin C, Klapste J, Lstiburek M (2011) Breeding without breeding: is a complete pedigree necessary for efficient breeding? PLoS One 6:11. doi:10.1371/journal.pone.0025737

El-Kassaby YA, Lstiburek M (2009) Breeding without breeding. Genet Res 91:111–120. doi:10.1017/s001667230900007x

Falconer DS, Mackay TFC (1996) Introduction to quantitative genetics, vol Ed. 4. Longman Group Limited, Harlow

Funda T, El-Kassaby YA (2012) Seed orchard genetics. CAB Reviews 7: 1–23. doi:10.1079/pavsnnr20127013

Funda T, Lstiburek M, Lachout P, Klapste J, El-Kassaby YA (2009) Optimization of combined genetic gain and diversity for collection and deployment of seed orchard crops. Tree Genet Genomes 5:583–593. doi:10.1007/s11295-009-0211-3

GIS PMF (2014) GIS Groupe Pin maritime du futur. Les cahiers de la reconstitution n°4: matériel végétal de reboisement. http://www.onf.fr/outils/medias/20130708-143100-661300/++files++/4

Gomory D, Bruchanik R, Longauer R (2003) Fertility variation and flowering asynchrony in *Pinus sylvestris*: consequences for the genetic structure of progeny in seed orchards. For Ecol Manag 174: 117–126. doi:10.1016/s0378-1127(02)00031-2

Hansen OK (2008) Mating patterns, genetic composition and diversity levels in two seed orchards with few clones—impact on planting crop. For Ecol Manag 256:1167–1177. doi:10.1016/j.foreco.2008.06.032

Hosius B, Bergmann F, Konnert M, Henkel W (2000) A concept for seed orchards based on isoenzyme gene markers. For Ecol Manag 131: 143–152. doi:10.1016/s0378-1127(99)00209-1

Illy G (1966) Recherches sur l'amélioration génétique du pin maritime. Ann Sci forest 23:765–948. doi:10.1051/forest/19660401

Johnson R, Lipow S (2002) Compatibility of breeding for increased wood production and long-term sustainability: the genetic variation of seed orchard seed and associated risks. In: Johnson AC, Haynes RW, Monserud RA (eds) Congruent management of multiple resources: proceedings from the wood compatibility initiative workshop, vol 563. USDA Forest Service General Technical Report Pacific Northwest. Us Dept Agr, Forest Serv Pacific Nw Research Stn, Portland, pp 169–179

Kalinowski ST, Taper ML, Marshall TC (2007) Revising how the computer program Cervus accommodates genotyping error increases success in paternity assignment. Mol Ecol 16:1099–1106. doi:10.1111/j.1365-294X.2007.03089.x

Kang KS, Lindgren D (1998) Fertility variation and its effect on the relatedness of seeds in *Pinus densiflora*, *Pinus thunbergii* and *Pinus koraiensis* clonal seed orchards. Silvae Genet 47:196–201

Lambeth C, Lee BC, O'Malley D, Wheeler N (2001) Polymix breeding with parental analysis of progeny: an alternative to full-sib breeding and testing. Theor Appl Genet 103:930–943. doi:10.1007/s001220100627

Lambeth C (1980) Juvenile-mature correlation in Pinaceae and its implications for early selection. For Sci 26:571–580

Lelu-Walter MA, Thompson D, Harvengt L, Sanchez L, Toribio M, Paques LE (2013) Somatic embryogenesis in forestry with a focus on Europe: state-of-the-art, benefits, challenges and future direction. Tree Genet Genomes 9:883–899. doi:10.1007/s11295-013-0620-1

Lindgren D, Danusevicius D, Rosvall O (2009) Unequal deployment of clones to seed orchards by considering genetic gain, relatedness and gene diversity. Forestry 82:17–28. doi:10.1093/forestry/cpn033

Lindgren D, Gea LD, Jefferson PA (1997) Status number for measuring genetic diversity. For Genet 4:69–76

Lindgren D, Kang K (1997) Status number—a useful tool for tree breeding. Research Report of the Forest Genetics Research Institute (Suwon)154–165

Lindgren D, Mullin TJ (1998) Relatedness and status number in seed orchard crops. Can J For Res-Rev Can Rech For 28:276–283. doi:10.1139/cjfr-28-2-276

Lstiburek M, Hodge GR, Lachout P (2015) Uncovering genetic information from commercial forest plantations-making up for lost time using "breeding without breeding". Tree Genet Genomes 11:12. doi:10.1007/s11295-015-0881-y

Lstiburek M, Ivankova K, Kadlec J, Kobliha J, Klapste J, El-Kassaby YA (2011) Breeding without breeding: minimum fingerprinting effort with respect to the effective population size. Tree Genet Genomes 7:1069–1078. doi:10.1007/s11295-011-0395-1

Machanska E, Bajcar V, Longauer R, Gomory D (2013) Effective population size estimation in seed orchards: a case study of *Pinus nigra* ARNOLD and *Fraxinus excelsior* L./*F. angustifolia* VAHL. Genetika-Belgrade 45:575–588. doi:10.2298/gensr1302575m

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

Matziris DI (1994) Genetic variation in the phenology of flowering in black pine. Silvae Genet 43:321–328

McRae TA, Dutkowski GW, Pilbeam DJ, Powell MB, Tier B (2004) Genetic evaluation using the TREEPLAN® system. Paper presented at the IUFRO Joint Conference of Division 2 "Forest Genetics and Tree Breeding in the Age of Genomics: Progress and Future" Charleston, SC, USA, 1–5 November 2004

Muller-Starck G (1995) Protection of genetic variability in forest trees. For Genet 2:121–124

Mullin TJ (2014) OPSEL 1.0: a computer program for optimal selection in forest tree breeding. Technical Report Nr 841-2014, Arbetsrapport Från Skogforsk

Na SJ, Lee HS, Han SU, Park JM, Kang KS (2015) Estimation of genetic gain and diversity under various genetic thinning scenarios in a breeding seed orchard of *Quercus acutissima*. Scand J Forest Res 30:377–381. doi:10.1080/02827581.2015.1018936

Olsson T, Lindgren D, Li B (2001) Balancing genetic gain and relatedness in seed orchards. Silvae Genet 50:222–227

Pâques L (2013) Forest tree breeding in Europe: current state-of-the-art and perspectives. In: Pâques L E (ed) Managing Forest Ecosystems, Vol. 25. Springer, doi:10.1007/978-94-007-6146-9

Stoehr M, Webber J, Woods J (2004) Protocol for rating seed orchard seedlots in British Columbia: quantifying genetic gain and diversity. Forestry 77:297–303. doi:10.1093/forestry/77.4.297

Stoehr M, Yanchuk A, Xie CY, Sanchez L (2008) Gain and diversity in advanced generation coastal Douglas-fir selections for seed production populations. Tree Genet Genomes 4:193–200. doi:10.1007/s11295-007-0100-6

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 11:1–13. doi:10.1007/s11295-015-0932-4

Vidal M, Plomion C, Raffin A, Harvengt L, Bouffier L (2016) Forward selection in a maritime pine polycross progeny trial using pedigree reconstruction. V1. INRA [Data set] doi: 10.5281/zenodo.165158

Zobel BJ, Talbert JT (1984) Applied forest tree improvement. Wiley, New York, p 528