Evaluation of breeding cost in the French maritime pine breeding program and perspectives for alternative strategies using molecular markers

Aline Fugeray-Scarbel, Stéphane Lemarié, Frédéric Bernier, Annie Raffin, Laurent Bouffier

Canadian Journal of Forest Research

Article

Evaluation of breeding cost in the French maritime pine breeding program and perspectives for alternative strategies using molecular markers

Aline Fugeray-Scarbel[a], **Stéphane Lemarié**[a], **Frédéric Bernier**[b], **Annie Raffin**[b], **and Laurent Bouffier**[c]

[a]Univ. Grenoble Alpes, INRAE, CNRS, Grenoble INP, GAEL, F-38000 Grenoble, France; [b]INRAE, UEFP, F–33610 Cestas, France; [c]INRAE, Univ. Bordeaux, BIOGECO, F–33610 Cestas, France

Corresponding author: **Laurent Bouffier** (email: laurent.bouffier@inrae.fr)

Abstract

The economic efficiency of conventional breeding strategies for forest trees based on biparental crosses is compared with that of alternative strategies based on pedigree reconstruction using molecular markers. Analyses of economic efficiency is based on comparisons of breeding scenarios corresponding to the same total investment. The first step is the description and cost evaluation of each basic operation, from crossing to genetic selection and clonal archive establishment. Breeding scenarios are then compared by stochastic sampling with a parametric genetic model (POPSIM), the comparison criteria in this case being genetic gain in the seed orchard for a given level of genetic diversity. Additionally, the economic gain resulting from the use of improved material is estimated for different levels of breeding investment. Our analysis shows that genotyping costs account for a much smaller proportion of total investment than phenotyping costs. We also show that, in comparisons of breeding scenarios corresponding to the same total investment, the three main breeding strategies (biparental crosses, polymix crosses, and open pollination) achieve similar genetic gains provided that sufficiently large numbers of parents are considered. These results open up promising perspectives for the wider integration of molecular markers into forest tree breeding strategies.

Key words: breeding strategy, simulation, *Pinus pinaster* Ait., pedigree reconstruction, economic efficiency

Introduction

New research methods and technologies, such as genomics and high-throughput phenotyping, are often proposed as a way to increase genetic gain in forest tree breeding (Grattapaglia 2022). Indeed, genetic gain in forest tree species is heavily constrained because of the duration of the breeding cycle and the high cost of phenotyping trees. However, even when these new technological opportunities are mature enough to be applied, it is important to assess their impact on genetic gain, taking into account their associated cost and the breeding context in which they are to be introduced. This assessment requires a multidisciplinary approach, combining skills in biology and economics. Very few such analyses have been reported to date. We present here an analysis of the benefits of using molecular markers for maritime pine breeding in the specific context of France. We explore the utility of molecular markers for simplifying mating designs through pedigree reconstruction. This strategy decreases the cost of the initial crosses, but this benefit is counterbalanced by an increase in cost due to the use of molecular markers. In this article, we develop a model calibrated on the case of the French maritime pine breeding program, for comparative assessments of the genetic gain achieved in seed orchards for alternative breeding strategies with the same total cost.

Maritime pine is grown on about one million ha in France (FCBA 2022), mostly in the Landes forest, the largest cultivated forest in Western Europe. Most of this area (90%) is owned by private forest owners, the majority of whom rely on contractors for plantation, stand management, and forest exploitation. The maritime pine stand rotation age is about 40 years and the wood production in France is, on average, 8 $m^3 \cdot ha^{-1} \cdot year^{-1}$ (FCBA 2022). The wood produced has various outlets: lumber for construction purposes, parquet flooring, furniture, packaging and pallets, industrial wood (production of particle board and paper pulp), green chemistry, and wood energy. Improved forest reproductive material (FRM) are currently grown on more than 60% of the Landes forest (Raffin personal communication) and 29 000 ha of maritime pine (mean value for the last 10 years) are planted with improved FRM from seed orchards every year (French Ministry of Agriculture 2022). Improved FRM are selected mostly on the basis of their growth and stem straightness, but also, to a lesser extent, for resistance to diseases such as twisting rust caused by the pathogen *Melampsora pinitorqua*. Breeding research began in the late 1950s with the selection of "Plus Trees", followed by a recurrent selection scheme based on progeny testing for genetic evaluation and on clonal archives in which the selected trees are maintained

and crossed. Three cycles of genetic selection have been carried out, with the establishment of four series of seed orchards: the first went into production in the early 1980s, the second in 1993, and the third in 2011. Seed orchards for the fourth series are currently being installed.

Maritime pine breeding is performed jointly by two public institutes: *Institut national de recherche pour l'agriculture, l'alimentation et l'environnement* (INRAE, a research institute) and *Forêt Cellulose Bois-construction Ameublement* (FCBA, a technological institute), within the *Groupement d'Intérêt Scientifique* "Groupe Pin Maritime du Futur" (GIS GPMF). The GIS GPMF provides improved genetic material to consortia of seed companies and nurseries, which are responsible for seed orchard establishment, management, and exploitation. Seed orchards are harvested by the seed companies, which process and package the seeds. The seeds are then delivered to nurseries, which use them to produce seedlings for sale to forest owners.

The mean annual investment in maritime pine breeding in France has been estimated at €693 000 for the 2015–2020 period. This estimate is based on an analysis of the GIS GPMF budget and additional side projects closely related to breeding activities. Public sources (regional, national, and Europe) provide 95% of the funding for this breeding program to cover operating and personnel costs (INRAE and FCBA). The 5% of private funding comes from the royalties paid on improved material by the companies managing the seed orchards (3%) and on endpoint royalties paid by forest owners (2%).

The analysis of breeding costs is of considerable interest, particularly when performed to evaluate the benefits of introducing a new technique. This is currently the case in forest tree breeding, in which the development of genomics (Neale and Kremer 2011; Sterck 2022) has opened up new breeding opportunities (Isik 2014). One such opportunity is the use of molecular markers in simplified mating designs involving pedigree reconstruction. The breeding program can be simplified by the use of polymix crosses rather than biparental crosses, as proposed in the "polymix breeding with paternity analysis" strategy of Lambeth et al. (2001). The need for controlled crosses can even be avoided altogether if seeds are collected after open pollination, as suggested by El-Kassaby and Lstiburek (2009) in their "breeding without breeding" concept and by Hansen and McKinney (2010) in their "quasi-field trial" approach.

These new strategies have been evaluated for maritime pine through comparisons of genetic gains to those obtained with a conventional breeding scheme (Bouffier et al. 2019). We present here a deeper evaluation in which breeding costs are also integrated into the comparison. As suggested by Fugeray-Scarbel et al. (2022*a*), we considered strategy A to be more efficient than strategy B if, for the same total breeding cost, A leads to higher genetic gains than B. Following Bouffier et al. (2019), we compared three forward breeding strategies differing in terms of the types of crosses initially performed: (i) CC strategies based on biparental crosses, (ii) PMX strategies based on polymix crosses (i.e., mother trees crossed with a mixture of pollen from different trees), and, (iii) OP strategies based on the collection of seeds from clonal archives. The genetic value of the candidates was evaluated based on pedigree information. The full pedigree of each candidate was either known precisely from the mating design (CC strategies) or was determined by pedigree recovery based on genotyping data (PMX and OP strategies). Strategies were compared for the same level of diversity in the deployment population (i.e., in the seed orchards). Comparisons were made for three levels of total breeding cost, and for each level, we compiled the expected economic gain generated by the genetic gain, taking into account the research investment.

Materials and methods

1) Maritime pine breeding cycle

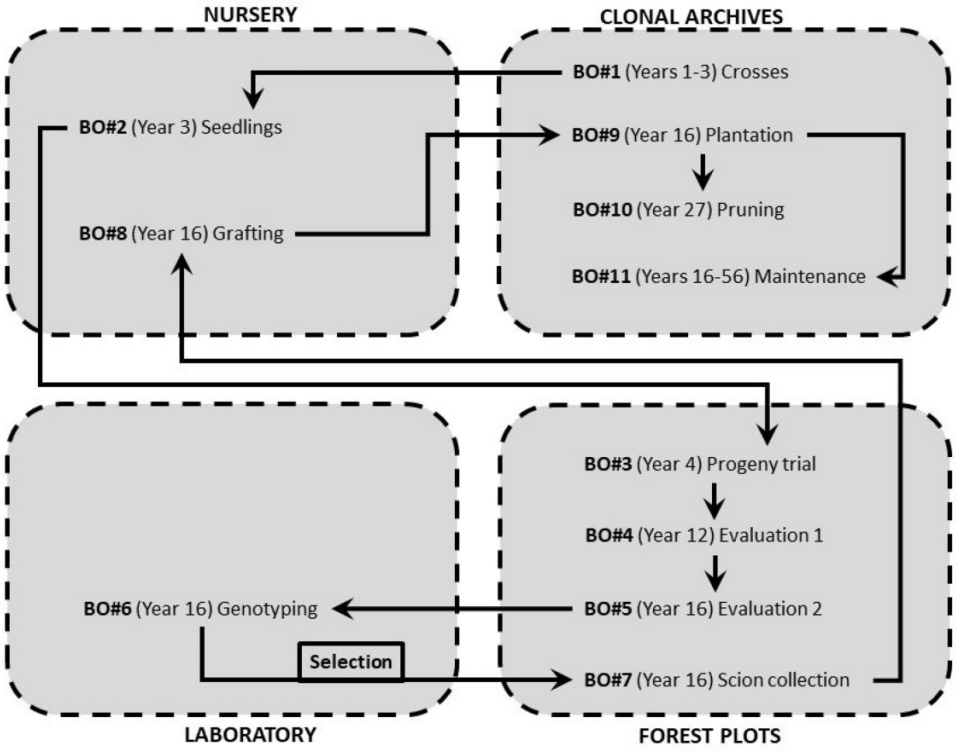
A maritime pine breeding cycle begins (year 1) with the crossing of the initial parents and culminates in the selection of the parents of the next generation 16 years later. From years 16 to 56, genotypes grafted in clonal archives are maintained. This breeding cycle can be described schematically (Fig. 1) by 11 basic operations (BO), the size of which depends on the strategy chosen: six BO (BO#1 to BO#6) from the start of the cycle until the selection step, followed by five BO (BO#7 to BO#11) for the installation and maintenance of the selected genotypes in clonal archives.

A detailed description of each BO is provided in Supplementary material S1. BO#1 can be broken down into several steps, including the crossing of the initial parents in clonal archives, cone collection, and seed extraction. In BO#2, harvested seeds are sown in the nursery to obtain 60 seedlings per cross. The seedlings are then installed in forest plots according to a randomized block design (BO#3). The resulting progeny trial is first evaluated at 8 years (BO#4) for growth, wood quality (stem straightness and branching), and sanitary status. A second evaluation is performed at 12 years (BO#5) for growth and complementary wood quality traits (wood density and wood grain angle). Genotyping (BO#6) is performed with 62 SNPs on a subset of selected trees to check tree identity or to determine parentage. The phenotyping and genotyping results are then used for a BLUP evaluation (Bouffier et al. 2016) leading to the final selection among the trees of the progeny trial. Scions are collected in the progeny trial (BO#7), with the objective of obtaining eight successful grafts per selected genotype (BO#8). These grafts are planted in clonal archives (BO#9) and are used as parents for the initial crosses for the next breeding cycle. Pruning (BO#10) is performed on 11-year-old trees in the clonal archives. Finally, the clonal archives are maintained to facilitate crossing operations, by mulching 15 times between plantation and the time point at which the trees reach the age of 40 years (BO#11).

2) Estimation of costs

Costs were estimated based on the maritime pine breeding activities carried out annually at INRAE. Four types of cost were considered. The first was personnel costs and expenditure on field visits by staff. For personnel costs, we estimated for each BO the number of days worked for each em-

Fig. 1. A maritime pine breeding cycle broken down into 11 basic operations (BO).



ployment category (researchers, technicians, and engineers) and then used the mean salary per category to calculate total salary costs. We also added 30% of this total salary cost to take into account the time for which staff were not assigned to a specific BO. The second type of cost considered was consumable costs, covering inputs (e.g., fertilizers, fuel, phytosanitary products, and water) and material supplies, such as pollination bags, labels, and stakes. The third type of cost considered related to equipment depreciation over a period of several years (measuring devices, shade houses, tractors, etc.). This cost was based on the purchase price for each item of equipment and the corresponding depreciation period. The final type of cost considered was rental costs and service costs in cases of outsourcing. Economies of scale were not taken into account. The raw unit cost of a given BO ($C^{\text{raw}}_{\text{BO\#}i}$) was calculated as the sum of costs for the four types of cost described above divided by the number of units (crosses, trees, and genotypes or samples) for the BO concerned, taking failure rate into account.

Cost estimates were used to simulate alternative breeding programs leading to the same total investment but with a different distribution of costs over time. For example, the PMX and OP strategies may require lower levels of expenditure for BO#1 than the CC strategy, but they entail additional genotyping costs for parental recovery at the end of the cycle in BO#6. Given the long duration of the breeding cycle, it is important to discount these costs when defining the parameters of breeding programs. Indeed, one Euro spent at the start of the breeding cycle weights more than one Euro spent latter in the breeding cycle. In accordance with standard practice in economics, we considered the discounted unit cost of each BO ($C_{\text{BO\#}i}$), taking the first year of the breeding cycle as a reference:

$$C_{\text{BO\#}i} = \frac{C^{\text{raw}}_{\text{BO\#}i}}{(1+r)^{t-1}}$$

where t is the year in which BO#i begins and r is the discount factor. We took into account a discount rate r of 4% in this study. This value is widely used in economic analyses and corresponds to the mean value used by both Chang et al. (2019*a*) and Chamberland et al. (2020) in their economic analyses of the application of genomic selection to tree breeding.

3) Simulation of breeding strategies

The three breeding strategies considered (CC, PMX, and OP) were characterized by four main variables:

N_C the number of crosses (CC and PMX strategies) or cones collected (OP strategy)
N_R the size of the population recruited
N_P the number of parents
N_G the number of genotyped trees

The CC strategy was initiated with N_C biparental crosses, generating N_R candidates. The genetic values of the candidates were determined in a BLUP analysis based on phenotypic data and the precisely known full pedigree (derived directly from the design of the crossing scheme). The final selection of N_P trees was preceded by the genotyping of $N_G = 1.1 \times N_P$ candidates to ensure that the pedigrees of the trees selected for the next breeding cycle are error-free.

The PMX strategy was initiated by N_C polymix crosses, generating a total of N_R candidates. At this stage, only the female parent of the candidates was known. A first BLUP analysis was performed to select a set of $N_G = 5 \times N_P$ candidates for genotyping to confirm the identity of the female parent and to determine the identity of the male parent based on the pollens present in the polymix. A second BLUP analysis was then performed, with full pedigree information, for the final selection of N_P trees.

The OP strategy was characterized by the absence of dedicated resources for crosses. Instead, N_C cones were collected in the clonal archives after open pollination. The female parent of each seed was known in this strategy, as the identity of the tree from which the seed was collected was recorded, but the male parent remained unknown at this stage. As in the PMX strategy, a first BLUP estimation of genetic value was performed with incomplete pedigree information for the prescreening of $N_G = 5 \times N_P$ candidates, which were then genotyped to determine their paternity (i.e., to identify the male parent from the trees maintained in the clonal archive). A second BLUP estimation with complete pedigree information was then performed to guide the final selection of N_P trees. No pollen contamination was considered for the OP strategy, whereas it has been shown that outside pollen can contribute to pollination in maritime pine seed orchards (Bouffier et al. 2023). However, an individual resulting from pollen contamination is supposed to have a lower breeding value than an individual resulting from a cross between two parents from the breeding population, making thus unlikely its selection during the prescreening process. However, if it happens, this individual would be excluded after the genotyping stage.

These three strategies were simulated with the POPSIM software (Mullin 2018; Mullin and Park 1995), which runs stochastic sampling according to a parametric genetic model. Simulations were performed on a single trait considering the typical genetic parameters for forest tree volume (Cornelius 1994). The heritability was fixed at 0.2 and the additive coefficient of variation was fixed at 15% (i.e., trait mean = 100, additive variance = 225, and environmental variance = 900). The strategies were compared on the basis of the genetic gains achieved for tree volume in seed orchards after five breeding cycles (expressed as a percentage of the base population mean). Twenty-five iterations were performed per strategy allowing to estimate the mean genetic gain and its associated standard deviation. As genetic gains are highly dependent on the level of genetic diversity, we maximized genetic gains for a predetermined level of genetic diversity (Mullin 2017). As described by Bouffier et al. (2019), genetic diversity expressed as the status number (Lindgren et al. 1996) was fixed at 30 in the breeding population after five breeding cycles and at 10 in the seed orchard, regardless of generation.

4) Sizing of the breeding strategies

Within each strategy, various sizes can be chosen for N_C, N_R, N_P, and N_G corresponding to the number of times each BO is realized. Below, we use the term "scenario" for a strategy associated with a specific set of these four variables. The relative efficiency of the various scenarios is determined by comparing them for the same total cost (CT), as follows:

$$(1) \quad CT = N_C \cdot \alpha \cdot C_C + N_R \cdot C_R + N_P \cdot C_P + N_G \cdot C_G + C_F$$

where $\alpha = 1$ for the CC and PMX strategies and $\alpha = 0.39$ for the OP strategy, and $N_G = \beta \cdot N_P$ with $\beta = 1.1$ for the CC strategy and $\beta = 5$ for the PMX and OP strategies.

In formulation (1), all BO applied to the same number of units are grouped together, and four cost categories are defined, with the cost for each category being identical in all breeding strategies:

- C_C is the discounted cost for one cross ($C_c = C_{BO\#1}$). We used $\alpha = 0.39$ for OP strategies, as cones were collected from clonal archives after open pollination. We estimated that 61.2% of the C_C is saved with the OP strategy because there are no crossing operations in this strategy, the remaining C_c being associated with cone collection and seed extraction.
- C_R is the discounted cost for growing and evaluating one candidate ($C_R = \sum_{i=2}^{5} C_{BO\#i}$).
- C_G is the discounted cost for genotyping one tree ($C_G = C_{BO\#6}$).
- C_P is the discounted cost for growing and maintaining one selected genotype ($C_P = \sum_{i=7}^{11} C_{BO\#i}$).
- C_F are the fixed costs considered to be identical for all breeding strategies, as some costs cannot be associated with a specific BO and are independent of the size of the BO. C_F relate to personnel costs associated with mating design, progeny trial design, BLUP evaluation, and the costs of the software used to perform these analyses.

Based on current practice in the French maritime pine breeding program, we set, as a reference, a CC strategy scenario called scenario CC150, with $N_C = 150$, $N_R = 15\,000$ (100 progenies per cross), $N_P = 150$, and $N_G = 165$. The total cost of this reference scenario is equal to CT^{REF}. Based on estimates of C_C, C_R, C_G, C_P, and C_F, $CT^{REF} = €355\,151$. As scenario size depends on four variables (N_C, N_R, N_P, and N_G), the number of alternatives resulting in the same total cost is very large, but it would not be realistic to consider all these alternatives.

In a first step, we fixed the total cost to CT^{REF} with N_P equals to 150 parents, which is the current level of this variable in the French maritime pine breeding program. For each of the three breeding strategies, several scenarios with different N_C values were considered. N_C ranged from 50 to 300 for CC strategies, but was restricted to the 50 to 150 range for PMX strategies, for which only 150 female parents were available. There are no controlled crosses in OP strategies, but we considered two scenarios differing in terms of the number of trees from which cones were collected (150 or 50). N_R was therefore adjusted according to the following equation to achieve the same total investment CT^{REF} for each of the scenarios simulated:

$$(2) \quad CT = CT^{REF} \Leftrightarrow N_R = \frac{CT^{REF} - C_F}{C_R} - N_C \cdot \frac{\alpha \cdot C_C}{C_R} - N_P \cdot \frac{C_P + \beta \cdot C_G}{C_R}$$

Table 1. Raw and discounted unit costs per BO and per BO category ($r = 4\%$).

BO#	BO description	BO year	BO raw unit cost (€)	BO discounted unit cost (€)	BO category	Discounted unit cost (€)
1	Crosses	1–3	369.5	369.5	Crosses (C)	$C_C = 369.5$
2	Seedlings	3	2.3	2.2	Candidates (R)	
3	Progeny trial	4	6.6	5.9	Candidates (R)	
4	Evaluation 1	12	4.3	2.8	Candidates (R)	
5	Evaluation 2	16	3.0	1.7	Candidates (R)	$C_R = 12.5$
6	Genotyping	16	15.8	8.8	Genotyping (G)	$C_G = 8.8$
7	Selection	16	255.7	142.0	Parents (P)	
8	Grafting	16	304.8	169.3	Parents (P)	
9	Plantation	16	242.9	134.9	Parents (P)	
10	Pruning	27	34.4	12.4	Parents (P)	
11	Maintenance	16–56	14.0	73.5	Parents (P)	$C_P = 532.0$
–	Overall management	1–16	46 017.0	30 980.0	–	$C_F = 30\,980.0$

We can see from (2) that, for a given scenario, N_R decreases as N_C increases; the magnitude of this decrease is increasing with C_C/C_R ratio.

In a second step, we considered two other levels of investment: $CT^{REF/2}$ (€177 575) and $CT^{REF\times2}$ (€710 570). For these two additional CT levels, various scenarios based on CC and PMX strategies were simulated (see size of breeding scenarios in Supplementary Table S2) following the same methodology as above. Scenarios were initially defined with $N_P = 150$ to allow a comparison with scenarios evaluated for the investment CT^{REF}, but other levels of N_P were then explored: $N_P = 100$ for $CT^{REF/2}$ and $N_P = 300$ for $CT^{REF\times2}$.

5) Estimation of economic gains

The economic gains correspond to a net present value (NPV), which is the difference in value of the forest tree production between a scenario with a research investment leading to genetic gains for tree volume and a scenario without research investment. The method used for estimating the NPV is detailed in Supplementary material S3. The general framework is similar to Chamberland et al. (2020) and Chang et al. (2019*a*). However, we compile the NPV from five successive generations of improved FRM issued from the five successive breeding cycles, while the literature generally analyzes the impact of only one generation over an infinite time horizon. To cope with the complexity related to this successive generations, we use a simple model to estimate the impact of genetic and wood production, without using forest growth model.

Results

1) Estimation of costs

The estimated BO costs are summarized in Table 1. All costs are presented both as if they occurred in the current year (i.e. raw costs) and discounted (i.e., taking into account the year in which the BO began). In the particular case of clonal archive maintenance (BO#11), this operation was repeated several times over the different breeding cycles. The BO raw unit cost corresponds to one occurrence of this operation, whereas the discounted cost corresponds to 15 occurrences from year 16 to year 56. Taking the discount factor into account, these 15 occurrences correspond to 5.3 times the (non-discounted) cost of one occurrence. For each BO, Supplementary Table S4 shows the distribution of costs between the four types. Personnel costs were the major cost type (between 42% and 97% depending on the BO). Table 1 also provides the costs per BO category, obtained by summing the corresponding BO costs. Finally, we also indicate the cost related to overall management (C_F) over the course of the 16-year cycle.

2) Breeding scenarios simulated and associated cost categories

The sizes of the breeding scenarios simulated for CT^{REF} are indicated in Table 2. The reference scenario, CC150, is a CC scenario with 150 crosses (i.e., each parent being involved in two crosses) and 100 offspring per cross. Alternative numbers of controlled crosses were also considered (50, 100, and 300 crosses, corresponding to CC50, CC100, and CC300, respectively) with the number of offspring adjusted so as to obtain the same CT (359, 165, and 35 offspring per cross, respectively). Three alternative PMX scenarios with different numbers of polymix crosses were considered: 50 (PMX50), 100 (PMX100), and 150 (PMX150). Each polymix contained pollen from the 150 parents. Two OP scenarios were evaluated: one based on the collection of seeds from all 150 trees in the parental population (OP150), and the other based on collection from 50 mother trees (OP50).

The right-hand side of Table 2 shows the distribution of CT between the various operations. Planting and maintaining the 150 parents accounted for 22.5% of CT, and this cost was identical for all strategies. Crossing accounted for between 2.0% (OP50) and 31.2% (CC300) of CT, depending on the number of crosses. Planting and phenotyping the recruitment population (candidates) accounted for the largest proportion of CT (from 37.2% to 64.9%), even if large numbers of crosses were performed. Finally, genotyping accounted for a small proportion of CT (from 0.4% to 1.9%), even in the PMX and OP scenarios.

Table 2. Size of breeding scenarios considered for the total cost CT^{REF}.

Breeding scenario	Size variables				Distribution of total cost			
	N_C	N_R	N_P	N_G	Crosses	Candidates	Parents	Genotyping
CC150	150	15 000	150	165	15.6%	52.8%	22.5%	0.4%
CC50	50	17 956	150	165	5.2%	63.2%	22.5%	0.4%
CC100	100	16 478	150	165	10.4%	58.0%	22.5%	0.4%
CC300	300	10 566	150	165	31.2%	37.2%	22.5%	0.4%
PMX50	50	17 546	150	750	5.2%	61.8%	22.5%	1.9%
PMX100	100	16 068	150	750	10.4%	56.6%	22.5%	1.9%
PMX150	150	14 590	150	750	15.6%	51.3%	22.5%	1.9%
OP50	50	18 450	150	750	2.0%	64.9%	22.5%	1.9%
OP150	150	17 303	150	750	6.1%	60.9%	22.5%	1.9%

Note: Fixed costs (C_F) account for 8.7% of the total costs in each row. N_C, N_R, N_P, and N_G are the number of crosses (CC and PMX strategies) or cones collected (OP strategy), the size of the population recruited, the number of parents, and the number of genotyped trees, respectively.

Table 3. Genetic gains (defined as the percentage of additive genetic effect relative to the mean of the base population) achieved in seed orchards at cycle 5 for scenarios corresponding to the total cost CT^{REF}, number of families, and family sizes in the recruitment population.

Breeding scenario	Number of families*	Number of offspring per family	Genetic gains at cycle 5 (SD)
CC50	50 FS	359	-**
CC100	100 FS	165	78.7% (3.7%)
CC150	150 FS	100	79.4% (3.2%)
CC300	300 FS	35	76.7% (3.8%)
PMX50	50 HS	349	70.8% (4.4%)
PMX100	100 HS	160	76.3% (3.6%)
PMX150	150 HS	97	77.4% (2.9%)
OP50	50 HS	369	72.7% (4.1%)
OP150	150 HS	115	78.2% (3.0%)

Full-sib (FS) or half-sib (HS) families
**Diversity constraints cannot be fulfilled*

The sizes of the breeding scenarios simulated for $CT^{REF/2}$ and $CT^{REF\times2}$ are presented in Supplementary Table S2. Considering these two additional contrasted CT allows to explore highly different scenarios. Generally speaking, scenarios associated with a low CT ($CT^{REF/2}$) are characterized by a large amount of resources dedicated to parents, as it is a mandatory condition to fulfill diversity constraints. In contrast, in scenarios associated with a higher level of investment ($CT^{REF\times2}$), more resources can be dedicated to planting and phenotyping the candidates.

3) Population recruitment, composition, and genetic gains

Population recruitment for CC scenarios involved the generation of large full-sib (FS) families (Table 3 for CT^{REF}): from 50 FS families with 359 trees per family (CC50) to 300 FS families with 35 trees per family (CC300). By contrast, the PMX and OP scenarios generated half-sib (HS) families (Table 3 for CT^{REF}) obtained with a mixture of 150 pollens (from a polymix or pollen cloud): 50 HS families in PMX50 (349 trees per family) and OP50 (369 trees per family), and 150 HS families in PMX150 (97 trees per family) and OP150 (115 trees per family). The population recruited in the PMX and OP scenarios therefore included a large number of FS families, with very few trees per family. For example, assuming that each pollen parent contributed equally to the polymix, PMX50 generated 7500 FS families with a mean of 2.3 trees per family.

The genetic gains in seed orchards after five breeding cycles are presented in Table 3 for scenarios associated with CT^{REF}. Given the number of parents (150) and the genetic diversity constraints considered, the optimal scenarios for each breeding strategy (CC, PMX, and OP) resulted in similar genetic gains (79.4% for CC150, 77.4% for PMX150, and 78.2% for OP150). Only the CC50 strategy did not reach a status number equals to 10 in seed orchards after five cycles. The evolution of genetic gains over the successive cycles is represented in Supplementary Figure S5 for the scenarios CC150, PMX150, and OP150. The genetic gains increased linearly over cycles as a parametric model was considered for the simulations. Indeed, in such a model, genetic effects are based on the infinitesimal quantitative genetic model and the genetic variance is assumed to be constant over generations.

The genetic gains for scenarios associated with $CT^{REF/2}$ and $CT^{REF\times2}$ are presented in Fig. 2 and in Supplementary Table S2. The genetic gains clearly increased with the level of breeding investment (the optimal scenario reached genetic gains of 60.8%, 79.4%, and 95.0% for $CT^{REF/2}$, CT^{REF}, and $CT^{REF\times2}$, respectively). However, as observed for CT^{REF}, the optimal scenario for CC and PMX strategies resulted in similar genetic gains when considering a given level of investment ($CT^{REF/2}$ or $CT^{REF\times2}$). Genetic diversity appeared as a major constraint with $CT^{REF/2}$ as several scenarios based on strategy CC did not allow to reach diversity constraints in seed orchard. When $CT^{REF\times2}$ is considered, diversity constraint is not anymore a major issue and increasing the number of parents to 300 appeared to be a valuable strategy to increase genetic gains.

4) Economic gains

The economic gain generated by the five successive breeding cycles are reported in Table 4 as the NPV based on genetic gains simulated for the best CC strategies for each level of in-

Fig. 2. Genetic gains (defined as the percentage of additive genetic effect relative to the mean of the base population) achieved in seed orchards at cycle 5 for scenarios corresponding to three levels of investment: CT^{REF}, $CT^{REF/2}$, and $CT^{REF\times2}$.

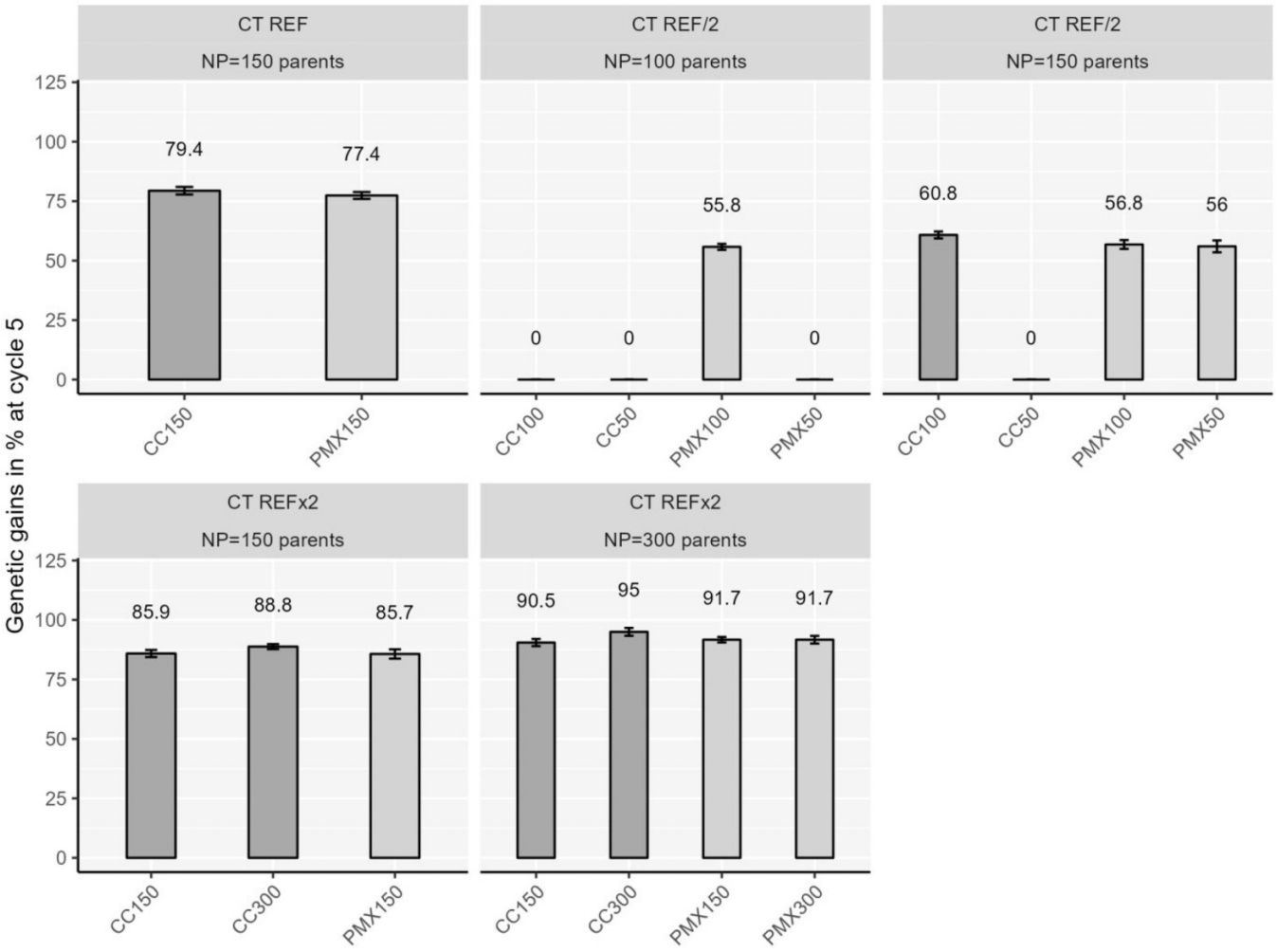


Table 4. Economic gains expressed with the net present value (NPV) from improved FRM issued from the five successive breeding cycles on the whole French maritime pine surface (1 million ha), per ha, as well as per ha and per year ($r = 4\%$).

CT	P	NPV (M€)	NPV·ha^{-1} (€)	NPV·ha^{-1}·an^{-1} (€)
CT^{REF}	30	173.37	6935	270
CT^{REF}	50	289.94	11 598	452
$CT^{REF/2}$	30	125.14	5005	195
$CT^{REF/2}$	50	209.06	8362	326
$CT^{REF\times2}$	30	193.95	7758	303
$CT^{REF\times2}$	50	325.23	13 009	507

Note: *P*, the wood unit price in €·m^{-3}, was considered to be constant. Genetic gains used for NPV calculations were based on scenarios CC150 ($N_P = 150$), CC100 ($N_P = 100$), and CC300 ($N_P = 300$) for CT^{REF}, $CT^{REF/2}$, and $CT^{REF\times2}$, respectively. The methodology used to estimate NPV is detailed in Supplementary material S3.

vestment. A 66% wood price increase from €30·t^{-1} to €50·t^{-1} leads to an increase of the economic gain of the same magnitude. The total economic gain over the 108 years ranges from €125 million to €325 million, a figure which is much higher compared to the total breeding investment corresponding to the five successive cycles that amounts from €0.65 million to €2.62 million. Considering various levels of CT, we find that the higher the breeding investment is, the higher the economic gain will be.

Discussion

Breeding strategies are generally compared on the basis of genetic gain, without consideration of costs: e.g., Lindgren and Mullin (1997), Rosvall and Mullin (2003), and Hallander and Waldmann (2009). However, several more recent studies have analyzed the economic costs and benefits of alternative breeding programs and deployment strategies. Chang et al. (2019*b*) reviewed 15 different articles, two of which, Chamberland et al. (2020) and Chang et al. (2019*a*), focused on the comparison of a breeding scheme based on genomic selection and a conventional breeding scheme. These breeding strategies differ in terms of the time required to obtain improved material and the productivity of this material. Chamberland et al. (2020) estimated an economic gain of between CAN$200·ha^{-1} and CAN$2000·ha^{-1}. They concluded that these estimates could guide policymakers as to

the level of investment required for genomic breeding programs. Chang et al. (2019*a*) assumed that genomic selection increased research costs by 50% per seedling. Based on this estimate, they concluded that the additional genetic gain and shorter breeding cycle provided by genomic selection were not necessarily associated with additional economic gain.

The analysis developed here differs from these previous studies by taking the breeding cost explicitly into account and compare alternative breeding scenarios corresponding to the same level of research investment. Our estimates of economic gain due to the deployment of improved FRM vary from €5000·ha^{-1} to €13 000·ha^{-1}. These results are consistent with those from Serrano-Leon et al. (2021) who analyze the economic impact of improved FRM for several species including maritime pine.

In cost–benefit-type analyses, breeding scenarios are often considered as a black box characterized by a total cost, a length of time, and a level of genetic gain, these parameters being defined in collaboration with experts. Our analysis is based on precise modeling of the breeding scenario by quantitative geneticists and economists working in collaboration. Finally, not only do we determine the total cost of the breeding scenarios, but we also analyze in detail the costs of the successive BO. This makes it possible to analyze how the alternative breeding scenarios result in reallocations of resources between BO categories, particularly as concerns the costs of crosses, phenotyping, and genotyping.

The main cost in all strategies was that associated with phenotyping, whereas genotyping made a smaller contribution to total costs (between 0.4% and 1.9% of CT^{REF}). These figures are lower than those reported in the few published articles taking costs into account in comparisons of breeding programs. To our knowledge, no other comparison of forest tree breeding programs has ever taken the costs of BO into account as we do here. In animal breeding, König et al. (2009) and Shumbusho et al. (2016) took costs into account, but König et al. (2009) did not provide any information about the relative contribution of genotyping to total breeding costs. According to Shumbusho et al. (2016), genotyping accounts for between 17% and 45% of total breeding investment. There are three reasons for this proportion being much higher than that in our analysis. First, in the study by Shumbusho et al. (2016), genotyping was performed for the entire recruitment population, for selection or pre-selection of the best progeny. By contrast, in this study, genotyping was performed only for *ex post* paternity determination. Our analysis therefore included the genotype of far fewer individuals. Second, the unit genotyping cost was lower for our study (€15.8) than in the study by Shumbusho et al. (2016) (€70 or €123). Indeed, genotyping costs have decreased significantly over time. As a result, costs are much lower for more recent publications. Moreover, paternity determination does not require the high-density genotyping required for genomic prediction, lowering costs still further. Finally, as costs are discounted, genotyping makes a smaller contribution to total costs in our analysis because it occurs later in the breeding cycle. Indeed, in the study by Shumbusho et al. (2016), genotyping is performed at the start of the breeding cycle to predict progeny value, whereas we performed genotyping to determine paternity at the end of the breeding cycle. Two other studies on wheat breeding (Longin et al. 2015; Ben-Sadoun et al. 2021) have confirmed this difference and advanced similar arguments. In the study by Longin et al. (2015), genotyping was used for genomic predictions for the selection or preselection of candidates and accounted for between 19% and 36% of the total breeding cost. Six years later, Ben-Sadoun et al. (2021) estimated genotyping costs at between 2.2% and 13.1% of the total breeding investment for similar breeding schemes. These figures are lower than those reported by Longin et al. (2015) due to a much lower per unit genotyping cost (more recent publication and data). However, they remain higher than that obtained here, because genotyping was performed on larger numbers of individuals. The limited additional cost estimated for the French breeding program opens up promising opportunities for new breeding strategies based on high-throughput genotyping, such as genomic selection.

In this article, breeding scenarios were simulated for the same total breeding cost and for the same level of diversity in the deployment populations. The control of diversity during the simulation process is of fundamental importance, as genetic gains are generally positively correlated with diversity loss. All the scenarios considered involved the production of seed orchards of identical diversity levels, making it possible to perform a fair comparison based on the genetic gains achieved after five breeding cycles. The main finding of these analyses was that, despite the different distributions of costs between BO categories, the best scenario for each breeding strategy (CC, PMX, and OP) reached a similar level of genetic gain. This result was consistent whatever the level of investment considered (CT^{REF}, $CT^{REF/2}$, or CT^{REFx2}). The PMX and OP scenarios generated more FS families than CC scenarios (one polymix cross can potentially generate 150 FS families in PMX scenarios versus one FS family in CC scenarios), but this did not result in a significant increase in genetic gain. The comparison of breeding scenarios at the same given total cost results in a trade-off, in the recruitment population, between the number of families generated and the number of offspring per family (Table 3). The number of families must be sufficient to allow interfamily selection, and the number of offspring per family must be high enough to allow intrafamily selection. We can hypothesize that intrafamily selection is limited in the PMX and OP scenarios due to the limited number of trees per FS family. Conversely, CC scenarios allow high levels of intrafamily selection, but only a limited number of parental combinations can be explored. For a given mating design, increasing the number of trees per family should increase genetic gains. The PMX150 and OP150 scenarios lead to the same mating design, but OP150 had a larger recruitment population as far fewer resources were dedicated to BO#1. Surprisingly, OP150 did not yield significantly higher genetic gains than PMX150 (78.2% versus 77.4%) for CT^{REF}, probably due to the small difference in family size (115 versus 97 trees per HS family). It is therefore tempting to design PMX or OP scenarios with smaller numbers of FS families to increase intrafamily selection. This could be achieved by limiting the number of crosses (PMX50) or by limiting the number of trees from which seeds are collected (OP50). However, the genetic

gains of these two scenarios are significantly lower than those for the optimal scenarios due to lower levels of genetic diversity in the recruitment population (as fewer maternal parents are involved in the crosses), making it necessary to select trees of low genetic value to satisfy the genetic constraints in seed orchards.

In CC scenarios, at least 100 crosses must be performed to satisfy the diversity constraint in seed orchards, whereas this constraint does not apply in PMX and OP scenarios. Thus, diversity constraints have a much greater impact on CC scenarios, which generate far fewer families than PMX and OP scenarios. This may be particularly important if certain crosses are unsuccessful due to biotic and abiotic stresses. Additional breeding alternatives (data not shown) were simulated in which 20% of the crosses were considered to be unsuccessful. In such conditions, genetic gains were only slightly lower in the PMX150 scenario (75.7% versus 77.4%), whereas it was impossible to create a seed orchard with the desired level of diversity in CC strategies.

As genetic gains in the French maritime pine breeding context are similar regardless of the breeding scenario when costs are taken into account, the final choice of breeding strategy will depend on seed yield per cross (low yields would argue for a switch from CC to PMX or OP strategies) and the capacity to integrate genotyping into routine practice in terms of organization (PMX and OP strategies require an efficient genotyping platform).

The French maritime pine breeding program was used as a reference for the calibration of our model. The total cost of the simulated breeding program was about €350 000, whereas annual investment was estimated at about €700 000 for maritime pine in France. These figures are different because overall annual investment includes research in support of the breeding program simulated here. These research activities account for about half the annual investment and include the development of genotyping tools, the management of genetic resources including provenances not used for breeding purposes, the production of grafted plants for the establishment of seed orchards, and studies on seed orchard management. It should also be borne in mind that the total cost of the breeding program is the sum of discounting costs over 16 years, while there is no discounting in the estimation of annual investment. For the reference strategy (CC150), with all BO performed in the same year (the current year, and, thus, not discounted), the total cost would be about €500 000. The simulated breeding program is therefore larger than the current breeding program. The reason for this choice is that the breeding program implemented in France for maritime pine resembles a "rolling front" strategy (Borralho and Dutkowski 1998) with overlapping cycles. However, such strategies cannot be simulated with POPSIM because this software can only simulate successive breeding cycles. With a "rolling front" strategy, several breeding cycles are launched in parallel, increasing the total range of parents used for crosses over that in programs based on successive cycles. We thus inflated the successive cycles simulated with POPSIM to achieve a realistic level of genetic diversity in the set of parents selected for the next breeding cycle.

Finally, the cost of the maritime pine program analyzed here is probably quite representative of other major tree breeding programs. Based on exchanges with breeders responsible for national programs for other forest tree species (poplar in France, Norway spruce in Finland and Norway, and *Pinus tadea* in the United States of America), annual investment in breeding appears to lie between €0.5 and 1 million, a range that contains our estimate for maritime pine in France (€0.7 million). A more precise comparison between these tree breeding programs would also require a comparison of the size of successive BO (e.g., number of crosses, size of recruitment population, and size of clonal archives).

Conclusion

This article is the first to compare forest tree breeding strategies based on a model combining genetic and economic approaches. We apply our approach to the French maritime pine breeding program and compare breeding strategies based on biparental crosses, polymix crosses, or seed collection after open pollination. After estimating the cost of each BO in the maritime pine breeding cycle, we were able to define different breeding scenarios for each strategy, representing the same total cost but with different numbers of crosses and different recruitment population sizes. The main finding of this study was that all three strategies (CC, PMX, and OP) designed for the level of investment CT^{REF} gave similar genetic gains when 150 FS (CC150) or HS families (PMX150 and OP150) were generated. This result contrasts with the findings of Bouffier et al. (2019), who showed that, for the same breeding program, strategies based on biparental crosses yielded higher genetic gains than strategies based on polymix or open pollination. This difference is due to the consideration of breeding costs in our study, but not in that of Bouffier et al. (2019), who compared scenarios with the same recruitment population size. Thus, taking breeding costs into account can substantially modify the results of comparisons of breeding strategies. Interestingly, our conclusions are consistent whatever the level of investment considered i.e., from breeding programs with limited resources ($CT^{REF/2}$) to ones with high resources ($CT^{REF\times2}$).

At last, we compile the economic gain from the five successive breeding cycles that are simulated. This gain is high and increases with the research investment, showing that research investment is worth making when considering the society as a whole. Moreover, high level of investments allows to manage more parents, resulting in a higher level of diversity in the breeding population and therefore facilitating the integration of new selection criteria. This could be a guarantee of adaptation for the future improved FRM in a context of climate change.

Finally, our analysis shows that the cost of genotyping accounts for a smaller proportion of the total cost than phenotyping, paving the way for the more widespread use of molecular markers in forest tree breeding programs, in genomic selection for example. These new breeding strategies can be analyzed with the framework combining genetic and economic approaches developed here and with collected breeding cost data.

Acknowledgements

This work was supported by the European Union's Horizon 2020 Research and Innovation Programme Project under grant agreement No. 773383 (B4EST).

Article information

History dates

Received: 25 May 2023
Accepted: 14 December 2023
Version of record online: 27 February 2024

Data availability

Data related to breeding costs are provided in full within the published article and its supplementary materials. Simulation parameters used to run POPSIM software and economic gain calculations are available in the *Recherche Data Gouv* repository, https://doi.org/10.57745/7AAZNU.

Author information

Author ORCIDs

Aline Fugeray-Scarbel https://orcid.org/0000-0002-2956-6943
Stéphane Lemarié https://orcid.org/0000-0003-1530-3787
Annie Raffin https://orcid.org/0000-0003-2459-4264
Laurent Bouffier https://orcid.org/0000-0001-7493-5077

Author contributions

Conceptualization: AF, SL, LB
Methodology: AF, SL, LB
Resources: AF, SL, FB, AR, LB
Software: LB
Supervision: AF, SL
Writing – original draft: AF, SL, LB
Writing – review & editing: AF, SL, FB, AR, LB

Competing interests

The authors declare there are no competing interests.

Supplementary material

Supplementary data are available with the article at https://doi.org/10.1139/cjfr-2023-0125.

References

Ben-Sadoun, S., Fugeray-Scarbel, A., Auzanneau, J., Charmet, G., Lemarié, S., and Bouchet, S. 2021. Integration of genomic selection into winter-type bread wheat breeding schemes: a simulation pipeline including economic constraints. Crop Breed. Genet. Genom. **3**(3). doi:cbgg.hapres.com/UpLoad/PdfFile/CBGG_1428.pdf.

Borralho, N.M.G., and Dutkowski, G.W. 1998. Comparison of rolling front and discrete generation breeding strategies for trees. Can. J. For. Res. **28**: 987–993. doi:10.1139/x98-07.

Bouffier, L., Raffin, A., and Dutkowski, G.W. 2016. Using pedigree and trait relationships to increase gain in the French maritime pine breeding program. *In* IUFRO Conference "Forest Genetics for Productivity". Rotorua, New Zealand, 14–18 March 2016. Available from https://hal.science/hal-02801580v1 [accessed January 2024].

Bouffier, L., Klápště, J., Suontama, M., Dungey, H.S., and Mullin, T.J. 2019. Evaluation of forest tree breeding strategies based on partial pedigree reconstruction through simulations: *Pinus pinaster* and *Eucalyptus nitens* as case studies. Can. J. For. Res. **49**(12): 1504–1515. doi:10.1139/cjfr-2019-0145.

Bouffier, L., Debille, S., Alazard, P., Raffin, A., Pastuszka, P., and Trontin, J.-F. 2023. Pollen contamination and mating structure in maritime pine (*Pinus pinaster* Ait.) clonal seed orchards revealed by SNP markers. Peer Community J. **3**(e68). doi:10.24072/pcjournal.302.

Chamberland, V., Robichaud, F., Perron, M., Gélinas, N., Bousquet, J., and Beaulieu, J. 2020. Conventional versus genomic selection for white spruce improvement: a comparison of costs and benefits of plantations on Quebec public lands. Tree Genet. Genomes, **16**(1). doi:10.1007/s11295-019-1409-7. PMID: 32256274.

Chang, W.-Y., Gaston, C., Cool, J., and Thomas, B.R. 2019a. A financial analysis of using improved planting stock of white spruce and lodgepole pine in Alberta, Canada: genomic selection versus traditional breeding. Forestry: Int. J. For. Res. **92**(3): 297–310. doi:10.1093/forestry/cpz011.

Chang, W.-Y., Wang, S., Gaston, C., Cool, J., An, H., and Thomas, B.R. 2019b. Economic evaluations of tree improvement for planted forests: a systematic review. BioProducts Business **4**(1). doi:10.22382/bpb-2019-001.

Cornelius, J. 1994. Heritabilities and additive genetic coefficients of variation in forest trees. Can. J. For. Res. **24**(2): 372–379. doi:10.1139/x94-050.

El-Kassaby, Y.A., and Lstiburek, M. 2009. Breeding without breeding. Genet. Res. **91**(2): 111–120. doi:10.1017/S001667230900007X.

FCBA. 2022. Memento FCBA. Available from https://www.fcba.fr/le-memento-2022-est-paru/ [accessed January 2024].

French Ministry of Agriculture. 2022. Statistiques annuelles sur les ventes de graines et plants forestiers. Available from https://agriculture.gouv.fr/statistiques-annuelles-sur-les-ventes-de-graines-et-plants-forestiers [accessed in January 2024].

Fugeray-Scarbel, A., Ben-Sadoun, S., Bouchet, S., and Lemarié, S. 2022a. Analyzing the economic effectiveness of genomic selection relative to conventional breeding approaches. *In* Genomic prediction of complex traits: methods and protocols. *Edited by* N. Ahmadi and J. Bartholomé. Springer, Berlin. pp. 619–644.

Grattapaglia, D. 2022. Twelve years into genomic selection in forest trees: climbing the slope of enlightenment of marker assisted tree breeding. Forests, **13**(10):. doi:10.3390/f13101554.

Hallander, J., and Waldmann, P. 2009. Optimum contribution selection in large general tree breeding populations with an application to Scots pine. Theor. Appl. Genet. **118**(6): 1133–1142. doi:10.1007/s00122-009-0968-7. PMID: 19183858.

Hansen, O.K., and McKinney, L.V. 2010. Establishment of a quasi-field trial in *Abies nordmanniana*—test of a new approach to forest tree breeding. Tree Genet. Genomes **6**(2): 345–355. doi:10.1007/s11295-009-0253-6.

Isik, F. 2014. Genomic selection in forest tree breeding: the concept and an outlook to the future. New. For. **45**(3): 379–401. doi:10.1007/s11056-014-9422-z.

König, S., Simianer, H., and William, A. 2009. Economic evaluation of genomic breeding programs. J. Dairy Sci. **92**(1): 382–391. doi:10.3168/jds.2008-1310. PMID: 19109296.

Lambeth, C., Lee, B.-C., O'Malley, D., and 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.

Lindgren, D., and Mullin, T.J. 1997. Balancing gain and relatedness in selection. Silvae Genet. **46**(2-3): 124–129.

Lindgren, D., Gea, L., and Jefferson, P. 1996. Loss of genetic diversity monitored by status number. Silvae Genet. **45**(1): 52–58.

Longin, C.F.H., Mi, X., and Würschum, T. 2015. Genomic selection in wheat: optimum allocation of test resources and comparison of breeding strategies for line and hybrid breeding. Theor. Appl. Genet. **128**(7): 1297–1306. doi:10.1007/s00122-015-2505-1. PMID: 25877519.

Mullin, T.J. 2017. OPSEL 2.0: A computer program for optimal selection in tree breeding. Arbetsrapport från Skogforsk Nr. 954-2017, Skogforsk, Uppsala, SE. 24p. doi:10.13140/RG.2.2.25809.43366.

Mullin, T.J. 2018. POPSIM: a computer program for simulation of tree breeding programs over multiple generations. Arbetsrapport från Skogforsk Nr. 984-2018, Skogforsk, Uppsala, SE. 34p. doi:10.13140/RG.2.2.35316.60804.

Mullin, T.J., and Park, Y.S. 1995. Stochastic simulation of population management strategies for tree breeding: a new decision-support tool for personal computers. Silvae Genet. **44**(2-3): 132–141.

Neale, D.B., and Kremer, A. 2011. Forest tree genomics: growing resources and applications. Nat. Rev. Genet. **12**(2): 111–122. doi:10.1038/nrg2931. PMID: 21245829.

Rosvall, O., and Mullin, T.J. 2003. Positive assortative mating with selection restrictions on group coancestry enhances gain while conserving genetic diversity in long-term forest tree breeding. Theor. Appl. Genet. **107**(4): 629–642. doi:10.1007/s00122-003-1318-9. PMID: 12851766.

Serrano-León, H., Ahtikoski, A., Sonesson, J., Fady, B., Lindner, M., Meredieu, C., et al. 2021. From genetic gain to economic gain: simulated growth and financial performance of genetically improved *Pinus sylvestris* and *Pinus pinaster* planted stands in France, Finland and Sweden. Forestry: Int. J. For. Res. **94**(4): 512–525, doi:10.1093/forestry/cpab004.

Shumbusho, F., Raoul, J., Astruc, J.M., Palhiere, I., Lemarié, S., Fugeray-Scarbel, A., and Elsen, J.M. 2016. Economic evaluation of genomic selection in small ruminants: a sheep meat breeding program. Animal, **10**(6): 1033–1041. doi:10.1017/S1751731115002049. PMID: 26446712.

Sterck, L. 2022. Maritime pine genomics in focus. *In* The pine genomes. Compendium of plant genomes. *Edited by* A.R. De La Torre. Springer, Cham, Switzerland.