Evaluation of forest tree breeding strategies based on partial pedigree reconstruction through simulations: Pinus pinaster and Eucalyptus nitens as case studies

Laurent Bouffier, Jaroslav Klápště, Mari Suontama, Heidi S. Dungey, Tim J. Mullin

To cite this version:

Laurent Bouffier, Jaroslav Klápště, Mari Suontama, Heidi S. Dungey, Tim J. Mullin. Evaluation of forest tree breeding strategies based on partial pedigree reconstruction through simulations: Pinus pinaster and Eucalyptus nitens as case studies. Canadian Journal of Forest Research, 2019, 49 (12), pp.1504-1515. 10.1139/cjfr-2019-0145 . hal-02622287

HAL Id: hal-02622287
https://hal.inrae.fr/hal-02622287v1

Submitted on 30 Sep 2024

Title

Evaluation of forest tree breeding strategies based on partial pedigree reconstruction through simulations: *Pinus pinaster* and *Eucalyptus nitens* as case-studies.

Laurent Bouffier*, Jaroslav Klápště, Mari Suontama, Heidi S. Dungey, Tim J. Mullin
* corresponding author

Laurent Bouffier:
BIOGECO, INRA, Univ. Bordeaux, 33610 Cestas, France
laurent.bouffier@inra.fr / +33 5 35 38 53 60

Jaroslav Klápště :
Scion, Te Papa Tipu Innovation Park, 49 Sala Street, Private Bag 3020 Rotorua, 3046, New Zealand
jaroslav.klapste@scionresearch.com

Mari Suontama :
Scion, Te Papa Tipu Innovation Park, 49 Sala Street, Private Bag 3020 Rotorua, 3046, New Zealand
Present address: Skogforsk (The Swedish Forestry Research Institute), Box 3, Sävar, Sweden S-918 21
Mari.Suontama@skogforsk.se

Heidi S. Dungey,
Scion, Te Papa Tipu Innovation Park, 49 Sala Street, Private Bag 3020 Rotorua, 3046, New Zealand
Heidi.Dungey@scionresearch.com

Tim J. Mullin,

Skogforsk (The Swedish Forestry Research Institute), 224 rue du Grand Royal Est, Shefford,
QC, Canada J2M 1R5; and Box 3, Sävar, Sweden S-918 21

Tim.Mullin@skogforsk.se

Abstract

Despite recent developments in molecular markers, most forest tree breeding programmes do not use them routinely. One way to integrate markers would be to use them for pedigree reconstruction after a simplified mating design through polymix or open-pollinated breeding. Thanks to the latest developments in the POPSIM simulator, various breeding strategies, including some based on paternity recovery, were evaluated with specified constraints on the level of diversity over breeding cycles. These simulations were carried out in two case-studies: the French *Pinus pinaster* (Ait.) and the New-Zealand *Eucalyptus nitens* (H. Dean & Maiden) breeding programmes.

The *Pinus pinaster* case-study produced lower genetic gain for the polymix breeding strategy with paternity recovery, compared to double-pair mating or optimal-contribution strategies. However, the polymix breeding strategy could be of interest if the mating design is faster to complete. In the *Eucalyptus nitens* case-study, pedigree recovery was shown to be a mandatory step to control the erosion of diversity over breeding cycles. In both cases, the strategies based on pedigree reconstruction were applicable with a limited level of genotyping. Finally, these simulations allow some general recommendations to be drawn to help breeders when designing a strategy for forest tree breeding.

Keywords

forest tree, breeding strategy, simulation, molecular marker, pedigree reconstruction

Résumé

Dans la plupart des programmes d'amélioration forestiers, les marqueurs moléculaires ne sont pas utilisés en routine malgré les développements récents en génomique. Une des possibilités

d'intégration des marqueurs serait de les utiliser pour reconstituer les pedigrees après des croisements simplifiés de type polycross ou pollinisation libre. Grâce aux derniers développements du simulateur POPSIM, différentes stratégies d'amélioration, dont certaines incluant une recherche de paternité, ont été évaluées avec des contraintes spécifiques relatives au niveau de diversité au cours des cycles de sélection. Ces simulations ont été menées pour deux cas d'étude : les programmes d'amélioration de *Pinus pinaster* (Ait.) en France et d'*Eucalyptus nitens* (H. Deane & Maiden) en Nouvelle-Zélande.

Le cas d'étude *Pinus pinaster* révèle un gain génétique moindre pour la stratégie de croisements polycross avec recherche de paternité en comparaison aux stratégies de croisements double-paire ou de contributions optimales. Néanmoins, la stratégie de croisements polycross pourrait être intéressante si elle permet une réduction de la longueur des cycles de sélection. Pour le cas d'étude *Eucalyptus nitens*, la reconstitution de pedigree est une étape obligatoire pour contrôler l'érosion de la diversité au cours des cycles de sélection. Dans les deux cas, les stratégies basées sur la reconstitution de pedigree peuvent être mises en œuvre en génotypant un nombre limité d'individus. Finalement, ces simulations ont permis de formuler des recommandations générales afin d'aider les sélectionneurs à élaborer des stratégies d'amélioration chez les arbres forestiers.

Introduction

Forest tree breeding is a slow process due to the late onset of sexual maturity and late expression of selected traits, generally related to productivity and wood quality. Most tree breeding programmes were initiated from a base population mass-selected in natural forests or unimproved stands, and then bred following a recurrent selection scheme with successive cycles of "crossing-testing-selection" (Namkoong et al. 1988). Currently, the most advanced of them have completed only a few breeding cycles from the wild state, based on various breeding strategies, i.e., plan to achieve crossing, testing, selection and deployment activities (Dungey et al. 2009; Lee 2001; McKeand and Bridgwater 1998; Mullin et al. 2011; Mullin and Lee 2013; Wu et al. 2007). Development of molecular markers in the 80's has brought hopes to accelerate breeding cycles and to facilitate the introduction of new selection criteria through marker-assisted selection. Despite numerous studies related to QTL detection and association studies, no application of new breeding strategies based on molecular markers has been reported for forest trees (Muranty et al. 2014).

Dense marker coverage of the genome with high-throughput genotyping technologies could open a new area in forest tree breeding with the application of genomic selection to predict the breeding values without phenotyping (Grattapaglia and Resende 2011; Isik 2014). However, even if the genomic selection is currently applied successfully for dairy cattle breeding (Wiggans et al. 2017), several hurdles must be overcome for its application in forest tree breeding. The two major ones are probably: (1) the difficulty to predict the Mendelian sampling for efficient intra-family selection (Bouffier et al. 2018; Thistlethwaite et al. 2017) and, (2) the large investments required for development of genomic resources (if currently not available), which are only possible for major breeding programmes. No operational implementation of genomic selection in forest tree breeding has been reported to our knowledge even if many proof-of-concept studies have been published (Grattapaglia 2017).

Furthermore, molecular markers have also been used with success in various forest tree studies for pedigree reconstruction to monitor genetic diversity and level of pollen contamination in seed orchards (Dering et al. 2014; Lai et al. 2010) or to improve accuracy of genetic parameters and breeding values (Doerksen and Herbinger 2010; El-Kassaby et al. 2011; Klápště et al. 2017; Vidal et al. 2015). Only a limited number of markers are required for such studies, which makes the approach applicable to most forest tree species. Besides, pedigree reconstruction could enhance the development of new breeding schemes as proposed by Lambeth et al. (2001) with the “polymix breeding with paternity analysis” strategy, then by El-Kassaby and Lstiburek (2009) with the “breeding without breeding” concept and similarly by Hansen and McKinney (2010) with the “quasi-field trial” approach. In all cases, the traditional bi-parental mating design is substituted by a two-step process: first, a designed mating scheme using polycross or open-pollination, then parental reconstruction allowing the completion of incomplete pedigree, subsequent breeding values calculation and selection. Breeding schemes based on pedigree reconstruction have several advantages such as a simplification of the crossing process, generation of a large number of families with a timely mixing of the breeding population and verification of identities for selected genotypes. However, the attractiveness of such breeding strategies based on pedigree reconstruction depends on the specifics of the breeding programme under consideration (current breeding strategy, biological constraints of the species, annual investments, technical skills available, etc.).

The objective of this study is to evaluate the efficiency of breeding strategies based on partial pedigree reconstruction in two contrasting breeding programmes: *Pinus pinaster* (Ait.) in France and *Eucalyptus nitens* (H. Deane & Maiden) in New-Zealand. The first is characterised by forward selection in bi-parental control-cross trials associated with a large investment (Bouffier et al. 2008), whereas the second is managed by forward selection in open-pollinated trials with a low level of investment (Klápště et al. 2017). For each case-study, alternative strategies based on pedigree reconstruction were proposed, and genetic gains achieved in the deployment population were compared with those from the current strategy through stochastic simulations. Quantifying and tracking the genetic variability in forest tree breeding populations is vital in order to keep enough variability for further breeding cycles (sustainable generation of long-term genetic gain) and to avoid inbreeding in the seed orchards (Stoehr et al. 2008). That is why, for a fair comparison, the same level of genetic variability was considered, regardless of the strategy. Finally, the advantages and limits of breeding strategies based on pedigree reconstruction are discussed in the context of forest tree breeding.

Material and methods

1. *Stochastic simulations*

Breeding strategy simulations have been carried out with the POPSIM software based on stochastic samplings through a parametric genetic model (Mullin and Park 1995). This simulator, freely available (Mullin 2018), has been developed to evaluate forest tree breeding strategies and is well suited to handle a wide range of mating designs including polymix breeding with paternity recovery. POPSIM's main simulation steps, illustrated in Figure 1, are: breed unrelated founders (base population) according to a specific mating design, generate a recruitment population (RP) of a given size, evaluate BLUP estimated breeding values (EBVs), make selections for the new breeding population (BP) and generate a production population with a seed orchard (SO). In addition, paternity testing can be applied on a subset (RP') of the RP if the full pedigree is unknown. BLUP evaluation is performed using an internal call of ASReml v.3 software (Gilmour et al. 2009) based on pedigree and simulated phenotypes. Specific diversity constraints can be applied to both the BP and SO selection. More specifically, the recent development of a new selection tool, OPSEL (Mullin 2017a) allows to maximize genetic gain at a predetermined level of genetic diversity, either for equal (Mullin and Belotti 2016) or unequal (Yamashita et al. 2018) genetic contributions of each genotype to the selected population. The main outputs of these simulations are the genetic gain and genetic diversity in the BP and SO over a defined number of breeding cycles. Genetic gain is expressed as a percentage in comparison to the mean of the base population. Genetic diversity is expressed with the status number, *Ns* (Lindgren et al. 1996):

$$Ns = \frac{1}{2\times\theta} \text{ where } \theta \text{ is the group coancestry (Cockerham 1967).}$$

2. *Simulation process*

Simulations were performed on a single trait, which can also be considered as a selection index. Typical genetic parameters for traits related to forest tree growth were considered (Cornelius 1994). The additive coefficient of variation was fixed at 15% and heritability at 0.2 (i.e., trait mean = 100, additive variance = 225, environmental variance = 900). The BP size was fixed at 150 trees in accordance with the two case-studies considered (*Pinus pinaster* and *Eucalyptus nitens*), and various mating designs specific for each case-study (see below) were applied to generate a RP of 15 000 trees. For each breeding cycle, a first selection step was carried out through BLUP evaluation (based on phenotypes and pedigree information) to form the next-generation BP of 150 trees. A second selection step was performed in this new BP to select genotypes according to their BLUP EBVs for the establishment of a SO. The selected proportion is thus 1% for the BP, which is the order of magnitude currently considered in the *Pinus pinaster* and *Eucalyptus nitens* breeding programmes. Some diversity constraints, detailed below for each case-study, were applied for these two selection steps on *Ns* if the full pedigree is known or, if not, on the maximum number of trees per family. For each simulation, 5 successive breeding cycles were generated with the same BP size, mating design and diversity constraints. When the full pedigree of the RP was not known, paternity testing was applied on a subset (RP') created through balanced within-family selection. In that case, the selection carried out to generate the new BP and the SO was realised only within the subset where paternal identity was recovered. This selection was based on a BLUP evaluation integrating the recovered pedigree information. A paternity-testing factor was defined as the ratio between the RP' size and the BP size. For example, with a BP size of 150, a paternity testing factor of 20 means that 3000 trees (out of the 15 000 trees of the RP) were genotyped for paternity testing. This step carried out with POPSIM simulates a partial paternity recovery based on molecular markers (no error or unknown parents from pollen contamination were considered in the paternity recovery process).

In this study, a breeding scenario, illustrated in Figure 2, will be defined as a combination of a breeding strategy (a specific mating design eventually associated with partial paternity recovery) and a set of diversity constraints applied on the BP and SO. Each scenario was simulated over 5 breeding cycles with 100 iterations, producing a mean (and a standard deviation) for the genetic gain and diversity in the 5 successive BP and SO. The breeding scenarios are identified and summarized in Table 1 for *Pinus pinaster* and *Eucalyptus nitens* case-studies.

3. *Pinus pinaster case-study*

Breeding scenarios Pinus_A and Pinus_B

Three main breeding strategies were considered to generate the RP of 15 000 trees:

i. Double-pair mating (DPM) in Pinus_A1

   Each of the 150 parents of the BP was involved in two crosses (random mating avoiding crosses between half-sibs) with 100 progenies per full-sib family (i.e. 150 full-sib families in total). Selection of the 150 trees in the next BP was performed under the diversity constraint by Mixed Integer Quadratically Constrained Optimization (MIQCO) methodology in OPSEL (Mullin and Belotti 2016), to optimize the selection of 150 individuals that would contribute equally as breeding parents.

ii. Polycross mating (PCM) in Pinus_A2 and Pinus_B

   The best 50 trees of the BP were selected based on their BLUP EBVs and each was crossed with three different polymixes of 50 pollen parents to generate 100 progenies per polycross mating. Each tree of the BP was randomly assigned to one, and only one, of the three polymixes, which means that all 150 trees of the BP contribute as a pollen parent in the mating design. This design produced 50 half-sib families (same seed parent) with 300 progenies each. Paternity testing factors varying from 0 to 50 (0, 5, 10, 25, or 50) were applied to genotype the trees with the highest BLUP EBVs within each half-sib family. The strategy is designated by PCM followed by the paternity testing factor. For example, the breeding scenario Pinus_A2 involves the PCM_5 strategy which corresponds to the genotyping of 750 trees (i.e., 5 x 150) selected as the best 15 trees from each of the 50 half-sib families. The best 150 trees in the RP were then selected under a diversity constraint using the MIQCO option in OPSEL.

In addition, an alternative PCM breeding strategy was considered in the breeding scenario Pinus_B where the 50 trees in the BP used as seed parents were assigned randomly. Pinus_B differs from Pinus_A2 only by the selection process (random selection for Pinus_B vs. the 50 best trees for Pinus_A2).

iii. Optimal contribution (OC) in Pinus_A3

The contribution from each parental selection to a controlled cross mating design was optimized with OPSEL, based on second-order cone programming (SOCP) methodology (Yamashita et al. 2018) to maximize the genetic gain at a given diversity level. The mating design for the optimized contributions was then generated with XDesign (Mullin 2017b), avoiding coancestry between mates greater than 0.01. A total of 150 crosses were generated in the OC strategy with 100 progenies per cross.

These three breeding strategies were evaluated under identical constraints on genetic diversity. When full pedigree was known (DPM, OC or PCM associated with paternity testing), an increase of group coancestry of maximum 0.00267 per generation was allowed in the BP (i.e., *Ns* ≥ 30 after 5 cycles) and the genetic diversity of the SO was fixed at *Ns* ≥ 10 (i.e., group-coancestry ≤ 0.05), this was accomplished using the SOCP option in OPSEL to optimize unequal numbers of ramets to a grafted orchard. For the PCM_0 strategy (in the breeding scenario Pinus_A2'), as the paternal identity was not known, the three best trees per half-sib family were selected for the BP and a maximum of one progeny per seed parent was selected for the SO.

Breeding scenarios Pinus_C and Pinus_D

In order to estimate the sensitivity of the ranking among the three breeding strategies evaluated in Pinus_A, RP size and diversity constraints were modified. The breeding scenarios Pinus_C and Pinus_D involved the same breeding strategies evaluated in Pinus_A, but the RP was increased by three times in Pinus_C (i.e., RP = 45 000 instead of 15 000, *Ns* ≥30 after 5 cycles in the BP, and *Ns* ≥10 in the SO) and the diversity constraints were relaxed in Pinus_D (i.e., RP = 15 000, *Ns* ≥ 15 instead of 30 after 5 cycles in the BP, and *Ns* ≥ 5 instead of 10 in the SO).

4. *Eucalyptus nitens* case-study

Two main breeding strategies were considered for the *Eucalyptus nitens* case-study to generate the RP of 15 000 trees:

i. Open-pollination with seeds collected from all the trees (OP_all)

   OP_all was simulated considering a polymix of all 150 pollen parents from the BP applied on the same 150 trees, to generate 100 progeny per half-sib family.

ii. Open-pollination with seeds collected from 50 of the trees with the highest BLUP EBVs (OP_best)

   OP_best was simulated considering a polymix of all 150 pollen parents from the BP applied on the 50 best trees, to generate 300 progeny per half-sib family.

For each main breeding strategy, two paternity testing factors (0 and 10) were considered. When genotyping was performed, an identical number of trees per family was genotyped (10 trees / family for OP_all, 30 trees / family for OP_best), i.e., the paternity testing factor in each case equals 10. The strategies are designed by OP_all or OP_best followed by the paternity testing factor (0 or 10). Thus, four breeding strategies were considered for *Eucalyptus nitens*: OP_all_0, OP_all_10, OP_best_0 and OP_best_10.

These breeding strategies were evaluated under the same diversity constraints as for the breeding scenarios Pinus_A, i.e., *Ns* ≥ 30 after 5 cycles in the BP and *Ns* ≥ 10 in the SO when the full pedigree of the selection candidates was known (OP_all_10 and OP_best_10). When the pollen parent identity was unknown, 1 tree (OP_all_0) or 3 trees (OP_best_0) per half-sib family were selected based on BLUP EBVs for the BP, and a maximum of one progeny per parent was selected for the SO in both cases. The *Eucalyptus* breeding scenarios, designated Eucalyptus_A (no paternity testing) and Eucalyptus_B (paternity testing), are summarized in Table 1.

5. Complementary simulations

The evaluation of open-pollination and polymix breeding strategies raises the question of the optimal number of families and the optimal number of trees per family in the RP as, contrary of bi-parental mating designs, these strategies can easily generate a very large number of families. To tackle this issue, two simulation series were carried out, both under the genetic diversity constraints considered in the two case-studies (i.e., $Ns \geq 30$ at cycle 5 in the BP and $Ns \geq 10$ in the SO).

In the first series, a DPM strategy (random mating with two crosses per parent, i.e., 150 crosses or families in total) was considered with 5 levels of full-sib family size (10, 50, 100, 200 and 300 trees per family). This means a decreasing selected proportion (from 10% to 0.3% for the BP selection) as the BP size was constant (BP = 150) and the RP size varied from 1500 to 45 000 trees.

In the second series, seven levels of bi-parental cross size were randomly generated (50, 75, 150, 500, 1000, 2500 and 5000 crosses) with an equal contribution from each parent. The number of progenies per family was adjusted in order to keep a RP of 15,000 trees.

As these simulations were time-demanding, 50 iterations were run within each series. One scenario (150 crosses with 100 progenies per family) was performed with both 50 and 100 iterations and no significant difference in genetic gain was found (gain in SO is 80.3% ± 2.9 with 50 iterations and 79.5% ± 3.5 with 100 iterations).

Results

Genetic gain (percentage of the additive genetic effect relative to the base population mean) in the SO and genetic diversity (*Ns*) in the BP at cycle 5 are reported in Table 2.

Pinus pinaster case-study

Results from scenarios Pinus_A are illustrated in Figure 3 (scenarios Pinus_A1, Pinus_A2 and Pinus_A3). The OC strategy achieved greater genetic gain in the SO at cycle 5 (86.6%) than the DPM strategy (79.5%) and PCM_5 strategy (70.1%), which means that the DPM and OC strategies outperformed the PCM_5 strategy by a margin of 13.4% and 23.5%, respectively. Genetic gains for the PCM_5 strategy were associated with a larger standard deviation (4.3) than for the OC (3.6) and DPM (3.5) strategies. As genetic diversity was constrained, the three strategies had $Ns \geq 10$ in the SO regardless of the cycle and showed similar diversity decrease in the BP to reach $Ns \geq 30$ after 5 breeding cycles. Various paternity testing factors (from 5 to 50) were considered for the PCM strategy, but these had no significant impact on the results. Thus, only the PCM_5 (scenario Pinus_A2) was reported in Figure 3. When a paternity testing factor below 5 was considered, it was not possible to meet the constraints for genetic diversity. A paternity testing factor greater than 50 was not considered because the simulations were too time consuming to be realized with 100 iterations. PCM_100 was run over one iteration (simulation time = 11 hours) and achieved a genetic gain in the SO of 74.0%. However, as it was considered unrealistic to genotype more than 50% of the BP (7500 trees) as well as excessive simulation time, PCM strategies with a paternity testing factor greater than 50 were not evaluated. The genetic diversity constraints could not be applied for the PCM_0 strategy due the unknown pollen parent identity. Even considering the stronger diversity constraint that can be applied on to this breeding strategy (i.e., selection of the 3 best trees for each of the 50 half-sib families), *Ns* fell below 30 after 2 generations and reached 12.4 at cycle 5.

Scenario Pinus_B (PCM_5 with a random selection of the 50 seed parents) showed a non-significant reduced genetic gain in the SO when compared to scenario Pinus_A2 (67.2% with sd=4.1 vs 70.1% with sd=4.3).

In the Pinus_C scenarios, where the recruitment population was inflated from 15 000 to 45 000 trees for the three breeding strategies, genetic gain in the SO increased whatever the strategy, but the differences between strategies slightly decreased compared with a RP of 15 000 trees: gain for PCM_5 was 77.6%, 87.8% for DPM (+13.1%) and 94.7% for OC (+19.5%).

In the Pinus_D scenarios, where the diversity constraints were relaxed, greater genetic gain in the SO and lower differences between strategies were found: gain was 88.4% for PCM_5, 95.8% for DPM (+8.4%) and 98.9% for OC (+11.0%).

Eucalyptus nitens **case-study**

When no paternity testing was performed, OP_best_0 (scenario Eucalyptus_A2) exhibited a greater genetic gain in the SO than did OP_all_0 (scenario Eucalyptus_A1) (85.0% vs. 71.0%), but genetic diversity in both the SO and BP decreased more rapidly for OP_best_0 than OP_all_0 (Figure 4). As paternity identity was unknown, the genetic diversity was controlled only based on the seed parent identity and both strategies were below $Ns = 30$ in the BP at cycle 5. Ns in the SO was greater than 10 for OP_all_0 up to cycle 5, but dropped below 10 at cycle 5 for OP_best_0 (data not shown). When a paternity testing factor of 10 was considered, diversity constraints on Ns can be fulfilled, and thus both strategies were compared at a given diversity level ($Ns \geq 30$ in the BP at cycle 5 and $Ns \geq 10$ in the SO). In that case, OP_all_10 (scenario Eucalyptus_B1) outperformed OP_best_10 (scenario Eucalyptus_B2) for genetic gain in SO (77.2% vs. 70.4% at cycle 5) as illustrated in Figure 5.

Complementary simulations

In these complementary simulations, the genetic gain achieved in the SO after 5 breeding cycles was evaluated for various numbers of progenies per cross (under a constant number of crosses) and for various numbers of crosses (under a constant RP size).

For a given number of crosses (150 crosses considered here), the number of progenies had a strong impact on the genetic gain in the SO (Figure 6). As expected, the more progenies that were evaluated (from 10 to 300 per cross, i.e., an increasing RP size from 1500 to 45 000 trees), the more genetic gain was achieved. The gain increase was strong from 10 to 100 progenies per family: genetic gain was 56.9% with 10 progenies per cross and 80.3% with 100 progenies per cross. However, with over 100 progenies per cross, the gain increase was moderate (87.8% for 300 progenies per cross).

On the contrary, for a given RP size (15 000 trees considered here), the number of crosses had only a small impact on the genetic gain in the SO (Figure 7). A non-significant gain increase was found from 75 to 150 crosses (genetic gain was 76.1% with sd=3.6 for 75 crosses and 80.3% with sd=2.9 for 150 crosses). Over this number of crosses, the gain in the SO reached a plateau. When only 50 crosses were considered, the diversity constraints ($Ns \geq 30$ at cycle 5 in the BP and $Ns \geq 10$ in the SO) could not be fulfilled.

Discussion

Genetic diversity management is of fundamental importance in tree breeding (Rosvall 1999), first to limit inbreeding which induces inbreeding depression (Durel et al. 1996), and, second, to keep enough genetic variation for long-term genetic gains for the current selection criteria or for traits of interest in the future (Bouffier et al. 2008). Genetic diversity constraints are a key component when designing a breeding strategy. In the present study, when no diversity constraints were applied in the simulated breeding strategies, *Ns* in the BP dropped rapidly below 20 (data not shown) regardless of the breeding strategy considered (at the latest in the second generation for the *Pinus pinaster* case-study and in the third cycle for the *Eucalyptus nitens* case-study). This erosion of genetic diversity is clearly not compatible with the management of a breeding programme for the long-term, where an effective population size in the order of 20 to 80 is recommended (Danusevicius and Lindgren 2006; Ruotsalainen 2002). Diversity constraints were here considered in accordance with the current requirements of the two case-studies under study ($Ns \geq 30$ after 5 cycles in the BP and $Ns \geq 10$ in the SO).

It is difficult to compare breeding strategies when the genetic diversity levels in the breeding and deployment populations are not fully controlled. This is why the simulation tool POPSIM, with its options for selection constrained on diversity with OPSEL, has been used in this study, allowing a given level of genetic diversity in both the BP and SO to be fixed and thus making breeding strategies readily comparable. The *Pinus pinaster* case-study deals with the polymix breeding with paternity recovery as an alternative to bi-parental breeding schemes. The *Eucalyptus nitens* case-study highlights the potential of adding a paternity recovery step in an open-pollination breeding scheme to better manage genetic diversity over time. Identical genetic parameters and population size were considered for both case-studies, as it was relevant to the characteristics of these two breeding programmes. It is also helpful for easy comparison, enabling the formulation of some general conclusions on the use of paternity recovery in tree breeding strategies.

Paternity testing

The simulations performed in this study are based on the hypothesis that each genotype of the polymix (PCM or OP design) contributes equally and at random to the RP. Paternity recovery in two polycross trials of the French *Pinus pinaster* breeding programme (Vidal et al. 2015) broadly confirms this hypothesis, as only a few genotypes contribute more or less than expected. Paternity recovery carried out in open-pollinated *Eucalyptus nitens* seed orchard (Grosser et al. 2010) exhibits a larger variation in pollen contribution, partially due to the variation in the umbel number per tree. If the pollen contribution is unbalanced, the optimal level of genotyping could be underestimated in these simulations. However, in both case-studies, paternity testing rate does not significantly impact the genetic gain in the SO. This can be explained by clear identification of the best performers with a low paternity testing rate which did not change with a higher paternity testing rate. The advantage of scenarios using higher paternity rates is not to generate an additional genetic gain but rather to identify genotypes to meet genetic diversity constraints even if they do not perform quite as well. Therefore, paternity testing has to be sufficient to fulfil diversity constraints, but once this threshold is achieved, there is no gain from increasing the genotyping rate. Based on our diversity constraints, the paternity testing factor can be set to 5 (750 trees). When the paternity testing factor is below 5, the diversity constraints cannot be fulfilled. A paternity testing factor of 5 was chosen for *Pinus pinaster*, as no major deviation from equal contribution is expected for the pollen parents; but a higher level (10%) was considered for *Eucalyptus nitens* as more variable paternal contributions were expected.

Another underlying assumption is the absence of error in paternity recovery. Whereas, the hypothesis seems reasonable in the PCM strategy, where mixes of 50 pollen parents are considered, it can be more challenging in the OP strategy when dealing with 150 potential pollen parents that will be more and more related in subsequent generations. This increased relatedness between candidate parents will decrease discrimination power of parentage assignment and will result in increased occurrence of false positives (Olsen et al. 2001). The best option to bypass this difficulty would be to estimate the relatedness with the markers instead of recovering the pedigree. This implies using a G-matrix (markers-based) instead of the A-matrix (pedigree based), a methodology used in genomic selection studies (El-Dien et al. 2016). Whereas less than 100 well-chosen markers can be sufficient for a paternity analysis (Vidal et al. 2015), several thousands of markers will be necessary to recover relatedness with a G-matrix (Wang 2016).

Pinus pinaster **and polymix breeding with paternity recovery**

The DPM strategy, currently used for the French *Pinus pinaster* breeding programme, mixes efficiently the BP as all trees contribute equally to the RP. Nevertheless, the mating design is challenging to fulfil as it involves bi-parental crosses. The PCM_5 strategy can be an alternative to simplify the mating as a large number of families is generated with a limited number of control-crosses. In the simulated PCM design, three mixes of 50 pollen parents, which represent the entire BP, were applied on a restricted set of 50 seed parents. The limited number of parents involved in the pollen mixes is chosen to facilitate the pedigree recovery, particularly in the advanced cycles where the relatedness in the BP tends to be high. In addition, the restriction on the number of seed parents allows the focus of important resources on these genotypes (grafting, flowering induction, pruning), whereas it is generally not necessary to graft in a clonal archive the genotypes used only for pollen collection.

The DPM strategy outperforms the PCM_5 strategy in the delivery of genetic gain by 13.4% after 5 cycles, when RP=15 000 (scenarios Pinus_A1 vs. Pinus_A2), and this difference is similar (+13.1%) for a larger RP (RP=45 000 in scenarios Pinus_C). The DPM and PCM mating designs involve exactly the same number of crosses (150 bi-parental crosses for DPM vs. 50 seed parents x 3 polymixes =150 control-crosses for PCM), but the PCM design generates more full-sib families (potentially 50 x 50 x 3 = 7500 families) than the DPM design (150 families). However, for a given RP size, the number of crosses is not a key driver for the gain as shown in Figure 7, at least over a threshold of about 150 crosses. This can be explained by a trade-off between within-family and between-familiy selection intensity. In contrast, diversity constraints play a major role in the BP and SO selection. The DPM design involves each parent equally in the RP, whereas the PCM design simulated over-represents the 50 best parents selected as seed parents. When the selection is carried out under high diversity, the PCM design is penalized for gains due to this over-representation of the best parents. The decrease of gain difference between DPM and PCM_5 strategies when the diversity constraints are relaxed (5.3% when $Ns \geq 15$ in the BP at cycle 5 and $Ns \geq 5$ in the SO in scenarios Pinus_D) confirms that the diversity constraint is a key driver for the selection step. If the 150 parents have been represented equally in the PCM design (scenario equivalent to scenario Eucalyptus_B1 in Table 1), PCM would have performed equally well (77.2%±4.1) as DPM (79.5%±3.5 in scenario Pinus_A1). We conclude that there is no disadvantage or advantage to generate more full-sib families for a given RP size; but when the diversity constraints are strong, the DPM strategy outperforms the PCM_5 strategy if parents contribute unequally in the PCM mating design. However, both DPM and PCM_5 strategies are outperformed by the OC strategy where the 150 crosses use unequal contributions from selected parents to optimize gain under a specific level of diversity constraints. Scenario Pinus_A3 (OC strategy) displays 23.5% more gain in the SO than scenario Pinus_A2 (strategy PCM_5). This result confirms the superiority of the OC approach to maximize gain under a given level of genetic diversity (Hallander and Waldmann 2009; Meuwissen 1997).

The superiority of the OC strategy over DPM and PCM strategies has to be mitigated based on the time required to complete the design, which has an impact on the length of the breeding cycle, and thus the genetic gain per unit time. Generally, bi-parental crosses (OC and DPM designs) are more complex to implement than polymix crosses (PCM design), as mentioned earlier. In addition, as the random selection of the seed parent subset for the PCM design in scenario Pinus_B has only a small impact on genetic gain, in comparison to scenario Pinus_A2, one can focus on the 50 genotypes more convenient to breed as seed parents (earlier flowering, greater number of flowers, etc.) making the PCM design even easier to complete. Furthermore, the OC design is also probably more sophisticated than a DPM design, as there is a unique mating design that reaches the optimum gain, whereas several mating alternatives can respect the DPM strategy, giving greater flexibility (the only requirement for the DPM design is to cross each parent twice). For the French *Pinus pinaster* breeding programme, the current generation time is roughly 20 years. As DPM and OC designs outperform PCM design from 2.7 to 4.7% per breeding cycle on average, a cycle shortening of one year due to faster completion of the mating design with PCM strategy would give greater gain per unit time. However, the precise time required for each mating operation is currently difficult to estimate for the French breeding programme, as it faces low success for control crosses.

The genotyping cost is an extra investment for polycross breeding with paternity recovery. It has to be taken into account when selecting the most appropriate scenario, as breeders generally focus on the genetic gain per unit time for a given budget. With the decreasing cost of genotyping, the genotyping of 750 to 1500 trees should not be a major limitation. In the French *Pinus pinaster* breeding programme, the cost of paternity recovery is currently roughly 10€ per tree, which means an additional cost of 7500€ per cycle for the Pinus_A2 scenario. However, one has to keep in mind that this genotyping step also allows the verification of relatedness at each cycle. If no genotyping is performed in the DPM and OC strategies, it is likely that there will be an accumulation of errors in the pedigree (Munoz et al. 2014) that will affect the gain.

Eucalyptus nitens and open-pollination breeding

The current OP strategy implemented for the *Eucalyptus nitens* breeding programme in New Zealand, adopted due to biological constraints, does not allow the full management of genetic diversity as only the identity of the seed parent is known. While this is not necessarily an issue in the first few breeding cycles, it becomes a major concern after 4 cycles. Indeed, even with the strongest constraints applied on the seed parent identity in the BP (selection of 1 tree per half-sib family in scenario Eucalyptus_A1 and 3 trees per half-sib family in scenario Eucalyptus_A2), *Ns* drops below 30 in the BP at cycle 5 (Figure 4) if the full pedigree is unknown. As bi-parental crosses are not practicable due to biological and technical constraints, a genotyping step has to be carried out to recover the full-pedigree before selecting trees for the BP and SO. When paternity testing is considered, OP_all design (in scenario Eucalyptus_B1) outperforms OP_best design (in scenario Eucalyptus_B2). As discussed for polymix breeding scenario, the equal contribution of each parent in OP_all design makes it easier to meet diversity constraints, and thus achieve greater gains in SO. We conclude that it is preferable to collect seeds from the whole population, rather than focusing on genotypes with the highest BLUP EBVs. Besides the diversity management issue, paternity recovery also increases the EBVs accuracy for the selected trees. Additionally, in species with viable selfing such as Eucalyptus, identification of individuals coming from selfing through pedigree reconstruction allows their elimination from the BP and improvement of accuracy in genetic parameter estimates, especially in traits suffering from inbreeding depression (Klápště et al. 2017).

Simulation limitations

The POPSIM software is specially designed to simulate various forest tree breeding strategies at a given level of diversity in the successive BP and SO. However, in the PCM design, all the polymixes must be mated with the same set of seed parents. Without this limitation, and since an equal-parental contribution in the BP seems preferable to better manage the diversity, two alternative PCM mating designs could have been considered: the first one where each parent would be used as either pollen or seed parent, and the second where the whole BP would be used as both pollen and seed parents. This implies, for the first alternative, to divide the BP into two sets of 75 trees: 75 pollen parents (eventually two polymixes of 37 and 38 pollen parents can be designed to facilitate the paternity recovery) applied on the 75 remaining trees. The second alternative would be similar to that considered in this paper, but with each of the three polymixes being crossed with three different sets of 50 seed parents. It is expected that, for the same number of crosses (150 crosses), these strategies might achieve slightly greater gains than the PCM design evaluated in the paper. However, both alternatives imply grafting more than 50 trees (75 for the first and 150 for the second), which would complicate the field operations.

The mating design initially planned by the breeder is sometimes difficult to complete in the field (Kerr et al. 2015). A key issue is thus to estimate the consequences on the gain when there is a deviation for the initial mating design. This has not been studied by Kerr et al, but should be taken into account when selecting the more appropriate breeding scenario. Similarly, the impact of deviation from equal pollen contribution for PCM and OP designs was not evaluated in this paper. It is probably not a major issue for PCM design, as discussed previously with the *Pinus pinaster* case-study; but for OP design, phenology and pollen production variation could result in unequal pollen contributions in the RP. On the contrary, the pollination systems, wind pollination for conifers (Owens et al. 1998) and insect pollination for *Eucalyptus nitens* (Barbour et al. 2005), should not impact pollen contribution in the OP design, as pollen is dispersed far away from its source in both cases.

Finally, two additional limitations can be highlighted. First, simulations carried out in this paper considered only discrete generations. The “rolling front” breeding scheme has been proposed as an alternative: parents are crossed as soon as possible after selection and progeny trials are established as soon as the seeds are available, rather than waiting for all crosses in that generation to be completed. It has been demonstrated that the “rolling front” breeding scheme might achieve greater genetic gain per unit time in comparison with schemes based on discrete generations (Borralho and Dutkowski 1998). However, all the breeding strategies evaluated here can also be implemented as “rolling front”, although not specifically addressed here. Second, the level of genotyping chosen was applied by selecting, within each family, the trees with the highest BLUP EBVs (evaluated with their seed-parent identity and their own phenotype). This could bias the genetic parameter estimations carried out after paternity recovery; adding “random” trees in the genotyping process could counteract this. However, as a limited number of trees were genotyped (5 to 10% of the BP), only slightly biased estimates are expected.

Conclusion

The two case-studies considered in this paper can be extrapolated to other forest tree species. There is no “one” optimal breeding strategy; it has to be optimized depending on various parameters of the species biology, investments and genotyping facilities availability. However, some general conclusions can be drawn from these simulations. These are summarized below to help breeders when designing a breeding strategy that is considering the implementation of paternity testing.

1. The knowledge of the full pedigree is mandatory to manage genetic diversity in the long term. This can be achieved through bi-parental crosses or through paternity recovery in polymix or open-pollination breeding.
2. For a given RP size, increasing the number of full-sib families above a threshold (which corresponds approximately to the number of parents in the BP) is not of interest whatever the mating design (bi-parental or polymix breeding). In contrast, if additional investments are available, an increase of the RP size is useful to enhance gain in the SO (Figure 6).
3. The mating design is a key issue to maximize the gain for a given level of diversity, especially if diversity constraints are strong. The OC strategy is clearly the most effective way to optimize the gain per cycle (+23.5% over 5 cycles in the *Pinus pinaster* case-study in comparison to the PCM strategy), as the mating design is carried out taking into account relatedness constraints. However, this optimal mating design can be time-consuming to complete and, even sometimes, not feasible. Depending on the breeding context, alternative mating design can be considered, some followed by a paternity recovery step. If they are faster to complete, they can produce more genetic gain per time unit than the OC strategy.

   Moreover, the genetic gain is expressed, in this paper, per cycle with no time scale. Breeders have to keep in mind that it is generally more efficient to rapidly turn over a new cycle, than trying to optimize the gain per cycle by extending the period required to complete the mating design.

4. Alternative strategies that include paternity recovery are based on polymix and open-pollination breeding. In contrast with the OC strategy, where each parental contribution is optimized for gain given the diversity constraints, these alternative strategies are based on random mating between sets of parents. In that context, optimal selection must achieve gain through an equal contribution from each parent selected from the RP. This means, for OP strategy, the collection of an equal number of seeds on each parent. The trial design must also prevent a large deviation from an equal paternal contribution. For the PCM strategy, an equal contribution of each parent could be time-consuming to achieve. Crosses can be focused on a limited number of seed parents and keeping wide genetic diversity in the polymix composition as considered in this paper. The gain is not optimal, but it allows the completion of the crosses in limited time, as the number of trees grafted is reduced.
5. Interestingly, a limited level of paternity recovery is sufficient to achieve an optimal gain in the SO (a paternity testing factor of 5 for PCM strategy and 10 for OP strategy in this paper). The paternity testing rate must be adjusted depending on the expected deviation from equal paternal contribution (higher rate when contributions are more unequal), the diversity constraints (higher rate with stronger constraints), the pedigree errors suspected in BP (higher rate if pedigree records are not accurate) and genotyping costs.
6. Two major benefits are associated with the strategies based on paternity testing. The first is to cleanse the pedigree of errors at each generation, at least for the genotypes selected. This is a key point to achieve more accurate EBVs. The second is to facilitate the predictable transition from current breeding strategies, based only on pedigree, to more advanced breeding strategies based on molecular markers and genomic selection, more particularly single-step methodology (Klápště et al. 2018).

Acknowledgements

This publication was supported by the European Union's Horizon 2020 Research and Innovation Programme Project GenTree (grant agreement No 676876) and by the French Agricultural Ministry (project "Ressources génétiques et innovation variétale chez le pin maritime"). L. Bouffier's mobility at SCION was supported by the AgreenSkills+ fellowship programme which has received funding from the EU's Seventh Framework Programme under grant agreement N°FP7-609398, by the Fondation Bordeaux Université and by Scion Strategic Science Investment. The authors also acknowledge valuable discussions with Emily Telfer, Nathalie Graham and Ahmed Ismael from SCION.

References

Barbour, R.C., Potts, B.M., and Vaillancourt, R.E. 2005. Pollen dispersal from exotic eucalypt plantations. Conserv. Genet. **6**(2): 253-257. doi: 10.1007/s10592-004-7849-z.

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.

Bouffier, L., Klápště, J., Bartholome, J., Plomion, C., Telfer, E., Graham, N., and Dungey, H. 2018. Genomic selection proof-of-concept in two major conifer species. *In* Proceedings of the World Congress on Genetics Applied to Livestock Production, Auckland, New Zealand.

Bouffier, L., Raffin, A., and Kremer, A. 2008. Evolution of genetic variation for selected traits in successive breeding populations of maritime pine. Heredity (Edinb) **101**(2): 156-165. doi: https://doi.org/10.1038/hdy.2008.41.

Cockerham, C.C. 1967. Group inbreeding and coancestry. Genetics **56**(89-104).

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

Danusevicius, D., and Lindgren, D. 2006. Optimization of breeding population size for long-term breeding. Scand. J. For. Res. **20**(1): 18-25. doi: 10.1080/02827580410019517.

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

Doerksen, T.K., and Herbinger, C.M. 2010. Impact of reconstructed pedigrees on progeny-test breeding values in red spruce. Tree Genetics & Genomes **6**(4): 591-600. doi: 10.1007/s11295-010-0274-1.

Dungey, H., Brawner, J., Burger, F., Carson, M., Henson, M., Jefferson, P., and Matheson, A. 2009. A new breeding strategy for *Pinus radiata* in New Zealand and New South Wales. Silvae Genet. **58**(1-2): 28-38.

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

El-Dien, O.G., Ratcliffe, B., Klápště, J., Porth, I., Chen, C., and El-Kassaby, Y.A. 2016. Implementation of the realized genomic relationship matrix to open-pollinated white spruce family testing for disentangling additive from nonadditive genetic effects. G3: Genes Genom. Genet. **6**: 743-753. doi: 10.1534/g3.115.025957/-/DC1.

El-Kassaby, Y.A., Cappa, E.P., Liewlaksaneeyanawin, C., Klápště, J., and Lstibůrek, M. 2011. Breeding without breeding: is a complete pedigree necessary for efficient breeding? PLoS One **6**(10). doi: 10.1371/journal.pone.0025737.g001.

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

Gilmour, A.R., Gogel, B.J., Cullis, B.R., and Thompson, R. 2009. ASReml user guide release 3.0 VSN International Ltd, Hemel Hempstead, HP1 1ES, UK.

Grattapaglia, D. 2017. Status and perspectives of genomic selection in forest tree breeding. *In* Genomic Selection for Crop Improvement *Edited by* Springer, New York. pp. 199-249.

Grattapaglia, D., and Resende, M.D.V. 2011. Genomic selection in forest tree breeding. Tree Genet. Genomes **7**(2): 241-255. doi: 10.1007/s11295-010-0328-4.

Grosser, C., Potts, B.M., and Vaillancourt, R.E. 2010. Microsatellite based paternity analysis in a clonal *Eucalyptus nitens* seed orchard. Silvae Genet. **59**(1-6): 57-62. doi: 10.1515/sg-2010-0007.

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.

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.

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

Kerr, R.J., MacRae, T.A., Dutkowski, G.W., and Tier, B. 2015. Managing the rate of increase in average co-ancestry in a rolling front tree breeding strategy. J. Anim. Breed. Genet. **132**: 109-120.

Klápště, J., Suontama, M., Dungey, H.S., Telfer, E.J., Graham, N.J., Low, C.B., and Stovold, G.T. 2018. Effect of hidden relatedness on single-step genetic evaluation in an advanced open-pollinated breeding program. J. Hered. **109**(7): 802-810. doi: 10.1093/jhered/esy051.

Klápště, J., Suontama, M., Telfer, E., Graham, N., Low, C.B., Stovold, T., McKinley, R.B., and Dungey, H. 2017. Exploration of genetic architecture through sib-ship reconstruction in advanced breeding population of *Eucalyptus nitens*. PLoS One **12**(9). doi: 10.1371/journal.pone.0185137.

Lai, S.B., Funda, T., Liewlaksaneeyanawin, C., Klápště, J., Niejenhuis, A., Cook, C., Stoehr, M.U., Woods, J., and El-Kassaby, Y.A. 2010. Pollination dynamics in a Douglas-fir seed orchard as revealed by pedigree reconstruction. Ann. For. Sci. **67**(8): 808-808. doi: 10.1051/forest/2010044.

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.

Lee, S. 2001. Selection of parents for the Sitka spruce breeding population in Britain and the strategy for the next breeding cycle. Forestry **74**(2): 129-143.

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

McKeand, S., and Bridgwater, F. 1998. A strategy for the third breeding cycle of loblolly pine in the Southeastern U.S. Silvae Genet. **47**(4): 223-234.

Meuwissen, T.H. 1997. Maximizing the response of selection with a predefined rate of inbreeding. J. Anim. Breed. Sci. **75**: 934-940.

Mullin, T.J. 2017a. OPSEL 2.0: A computer program for optimal selection in tree breeding.

Mullin, T.J. 2017b. XDesign 1.0: a tool for generation of complex mating designs with optimal contributions. *Edited by* S. Arbetsrapport från Skogforsk Nr. 956-2017, Uppsala, SE, 15 pp.

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, 34 pp.

Mullin, T.J., Andersson, B., Bastien, J.-C., Beaulieu, J., Burdon, R.D., Dvorak, W.S., King, J.N., Kondo, T., Krakowski, J., Lee, S.J., McKeand, S.E., Pâques, L., Raffin, A., Russell, J., Skrøppa, T., Stoehr, M., and Yanchuk, A.D. 2011. Chapter 2: Economic importance, breeding objectives and achievements. *In* Genetics, genomics and breeding of conifers. *Edited by* J.B.a.C.K. C. Plomion. Science Publishers, Inc., New Hampshire; Edenbridge Ltd., UK.

Mullin, T.J., and Belotti, P. 2016. Using branch-and-bound algorithms to optimize selection of a fixed-size breeding population under a relatedness constraint. Tree Genetics & Genomes **12**(4). doi: 10.1007/s11295-015-0961-z.

Mullin, T.J., and Lee, S.J. 2013. Best practice for tree breeding in Europe. Skogforsk, Uppsala, SE.

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.

Munoz, P.R., Resende, M.F.R., Huber, D.A., Quesada, T., Resende, M.D.V., Neale, D.B., Wegrzyn, J.L., Kirst, M., and Peter, G.F. 2014. Genomic relationship matrix for correcting pedigree errors in breeding populations: impact on genetic parameters and genomic selection accuracy. Crop Sci. **54**(3): 1115-1123. doi: 10.2135/cropsci2012.12.0673.

Muranty, H., Jorge, V., Bastien, C., Lepoittevin, C., Bouffier, L., and Sanchez, L. 2014. Potential for marker-assisted selection for forest tree breeding: lessons from 20 years of MAS in crops. Tree Genet. Genomes **10**: 1491-1510.

Namkoong, G., Kang, G., and Brouard, J.S. 1988. Tree breeding: principal and strategies. . Springer-Verlag New York.

Olsen, J.B., Busack, C., Britt, J., and Bentzen, P. 2001. The aunt and uncle effect: an empirical evaluation of the confounding influence of full sibs of parents on pedigree reconstruction. J. Heredity **92**: 243-247.
Owens, J.N., Takaso, T., and Runions, C.J. 1998. Pollination in conifers. Trends Plant Sci. **3**(12): 479-484.

Rosvall, O. 1999. Enhancing gain from long-term forest tree breeding while conserving genetic diversity. *In* Swedish University of Agricultural Sciences, Umeå, Sweden.

Ruotsalainen, S. 2002. Managing breeding stock in the initiation of a long-term tree breeding program. *In* Finnish Forest Research Institute, Punkaharju Research Station.

Stoehr, M., Yanchuk, A., Xie, C.-Y., and Sanchez, L. 2008. Gain and diversity in advanced generation coastal Douglas fir selections for seed production populations. Tree Genet. Genomes **4**(2): 193-200.

Thistlethwaite, F.R., Ratcliffe, B., Klápště, J., Porth, I., Chen, C., Stoehr, M.U., and El-Kassaby, Y.A. 2017. Genomic prediction accuracies in space and time for height and wood density of Douglas-fir using exome capture as the genotyping platform. BMC Genomics **18**(1): 930. doi: 10.1186/s12864-017-4258-5.

Vidal, M., Plomion, C., Harvengt, L., Raffin, A., Boury, C., and Bouffier, L. 2015. Paternity recovery in two maritime pine polycross mating designs and consequences for breeding. Tree Genet. Genomes **11**(5): 105.

Wang, J. 2016. Pedigrees or markers: Which are better in estimating relatedness and inbreeding coefficient? Theor. Pop. Biol. **107**: 4-13. doi: 10.1016/j.tpb.2015.08.006.

Wiggans, G.R., Cole, J.B., Hubbard, S.M., and Sonstegard, T.S. 2017. Genomic selection in dairy cattle: the USDA experience. Annu. R. A. Biosci. **5**: 309-327. doi: 10.1146/annurev-animal-021815-111422.

Wu, H.X., Eldridge, K.G., Matheson, A., Powell, M.P., and McRae, T.A. 2007. Successful introduction and breeding of radiata pine to Australia. Austral. For. **70**(4): 215-225.

Yamashita, M., Mullin, T.J., and Safarina, S. 2018. An efficient second-order cone programming approach for optimal selection in tree breeding. Optim. Lett. **12**: 1-15. doi: 10.1007/s11590-018-1229-y.

Table 1: Description of the simulated breeding scenarios

Breeding scenario			Breeding strategy		Diversity constraints	
			Mating design	Paternity testing factor	BP	SO
Pinus pinaster	Pinus_A (RP size=15 000)	Pinus_A1	DPM 150 full-sib famillies with 100 progenies/familly	-	Ns≥30 after 5 cycles	Ns≥10
		Pinus_A2'	PCM: best 50 trees crossed with 3 polymixes 50 half-sib families with 300 progenies/familly	0	3 best trees/half-sib familly	1 progeny max/parent
		Pinus_A2		5	Ns≥30 after 5 cycles	Ns≥10
		Pinus_A2''		10		
		Pinus_A2'''		25		
		Pinus_A2''''		50		
		Pinus_A3	OC 150 full-sib famillies with 100 progenies/familly	-	Ns≥30 after 5 cycles	Ns≥10
	Pinus_B (RP size=15 000)	Pinus_B	PCM: 50 random trees crossed with 3 polymixes 50 half-sib families with 300 progenies/familly	5	Ns≥30 after 5 cycles	Ns≥10
	Pinus_C (RP size=45 000)	Pinus_C1	DPM 150 full-sib famillies with 300 progenies/familly	-	Ns≥30 after 5 cycles	Ns≥10
		Pinus_C2	PCM: best 50 trees crossed with 3 polymixes 50 half-sib families with 900 progenies/familly	15		
		Pinus_C3	OC 150 full-sib famillies with 300 progenies/familly	-		
	Pinus_D (RP size=15 000)	Pinus_D1	DPM 150 full-sib famillies with 100 progenies/familly	-	Ns≥15 after 5 cycles	Ns≥5
		Pinus_D2	PCM: best 50 trees crossed with 3 polymixes 50 half-sib families with 300 progenies/familly	5		
		Pinus_D3	OC 150 full-sib famillies with 100 progenies/familly	-		
Eucalyptus nitens (RP size=15 000)		Eucalyptus_A1	OP_all 150 half-sib famillies with 100 progenies/familly	0	1 best tree/half-sib familly	1 progeny max/parent
		Eucalyptus_B1		10	Ns≥30 after 5 cycles	Ns≥10
		Eucalyptus_A2	OP_best 50 half-sib families with 300 progenies/familly	0	3 best trees/half-sib familly	1 progeny max/parent
		Eucalyptus_B2		10	Ns≥30 after 5 cycles	Ns≥10

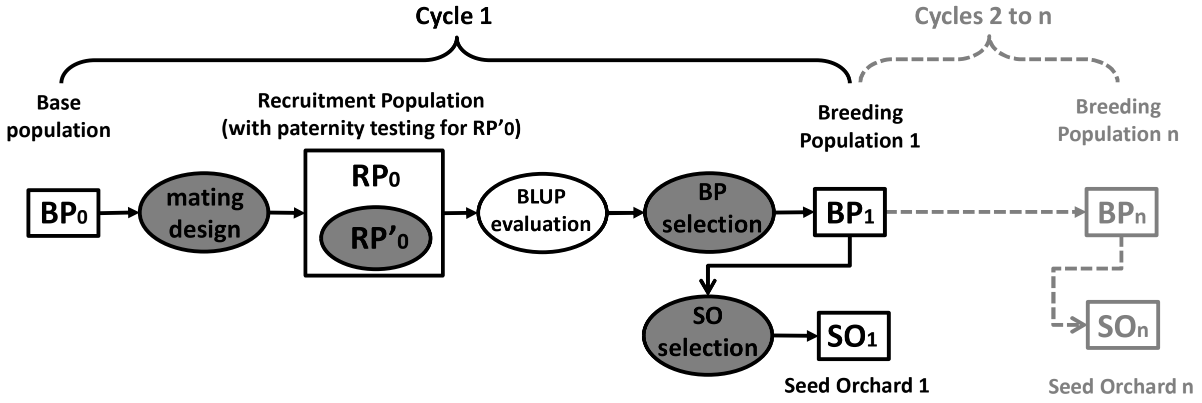
Note: The main simulated breeding scenarios are highlighted in grey.

Table 2: Genetic gain and genetic diversity of the simulated breeding scenarios

Breeding scenario			Results at cycle 5		Results over 5 cycles
			Genetic gain in SO	*Ns* in BP	
Pinus pinaster	Pinus_A	Pinus_A1	79.5 (3.5)	30.0 (0.0)	Figure 3
		Pinus_A2'	85.0 (4.3)	12.4 (1.5)	-
		Pinus_A2	70.1 (4.3)	30.0 (0.0)	Figure 3
		Pinus_A2''	69.5 (3.6)	30.0 (0.0)	-
		Pinus_A2'''	68.8 (4.6)	30.0 (0.0)	-
		Pinus_A2''''	70.7 (4.1)	30.0 (0.0)	-
		Pinus_A3	86.6 (3.6)	29.3 (0.2)	Figure 3
	Pinus_B	Pinus_B	67.2 (4.1)	30.0 (0.0)	-
	Pinus_C	Pinus_C1	87.8 (3.9)	30.0 (0.0)	-
		Pinus_C2	77.6 (3.2)	30.0 (0.0)	-
		Pinus_C3	94.7 (3.7)	29.2 (0.2)	-
	Pinus_D	Pinus_D1	95.8 (4.7)	15.0 (0.0)	-
		Pinus_D2	88.4 (6.4)	15.0 (0.0)	-
		Pinus_D3	98.9 (4.9)	14.9 (0.1)	-
Eucalyptus nitens	Eucalyptus_A1		71.0 (3.5)	26.1 (1.2)	Figure 4
	Eucalyptus_B1		77.2 (4.1)	30.0 (0.0)	Figure 5
	Eucalyptus_A2		85.0 (4.5)	12.4 (1.4)	Figure 4
	Eucalyptus_B2		70.4 (4.0)	30.0 (0.0)	Figure 5

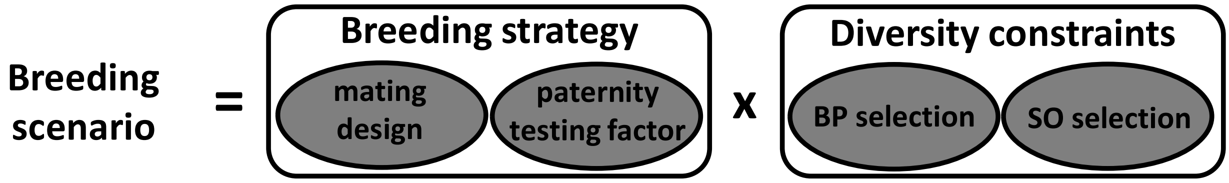
Note: The main simulated breeding scenarios are highlighted in grey.

Figure 1: Simulation process with POPSIM



Note: BP: Breeding Population; RP: Recruitment Population; RP': subset of the Recruitment Population considered for paternity testing; SO: Seed Orchard

Figure 2: A breeding scenario is defined as a breeding strategy applied with specific diversity constraints



Note:

Three main breeding strategies were considered for the *Pinus pinaster* case-study:

DPM: Double-Pair Mating

PCM_x: PolyCross Mating with a paternity testing factor of x (PCM_5 for example)

OC: Optimal Contribution

Two main breeding strategies were considered for the *Eucalyptus nitens* case-study:

OP_all_x: Open-Pollination with seeds collected on the whole BP and a paternity testing factor of x (OP_all_10 for example)

OP_best_x: Open-Pollination with seeds collected on the 50 best trees of the BP and a paternity testing factor of x (OP_best_10 for example)

For most breeding scenarios, diversity constraints were fixed at *Ns* ≥ 30 in BP and *Ns* ≥ 10 in SO.

Figure 3: *Pinus pinaster* case-study

Genetic gain in SO and genetic diversity in BP over 5 breeding cycles (each point is a mean of 100 iterations and the error bar represents ±1 standard deviation).

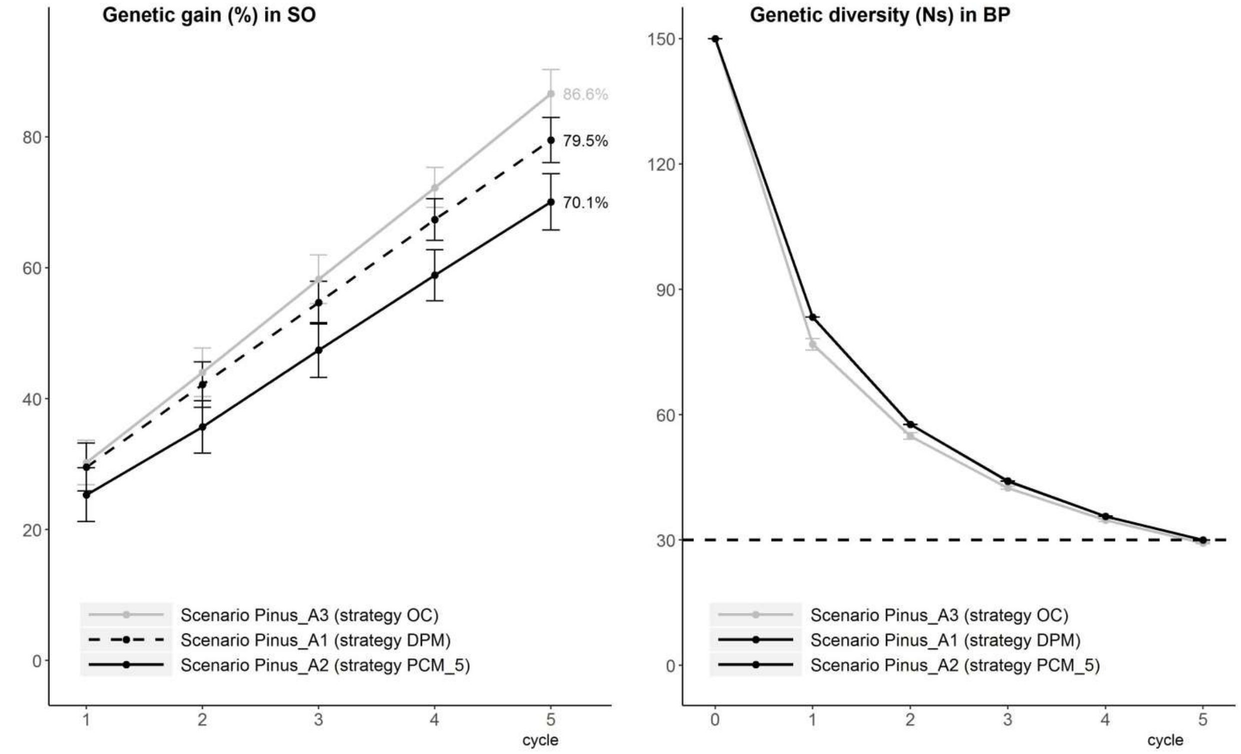


Figure 4: *Eucalyptus nitens* case-study – Breeding strategies without paternity testing
Genetic gain in SO and genetic diversity in BP over 5 breeding cycles (each point is a mean of 100 iterations and the error bar represents ±1 standard deviation).

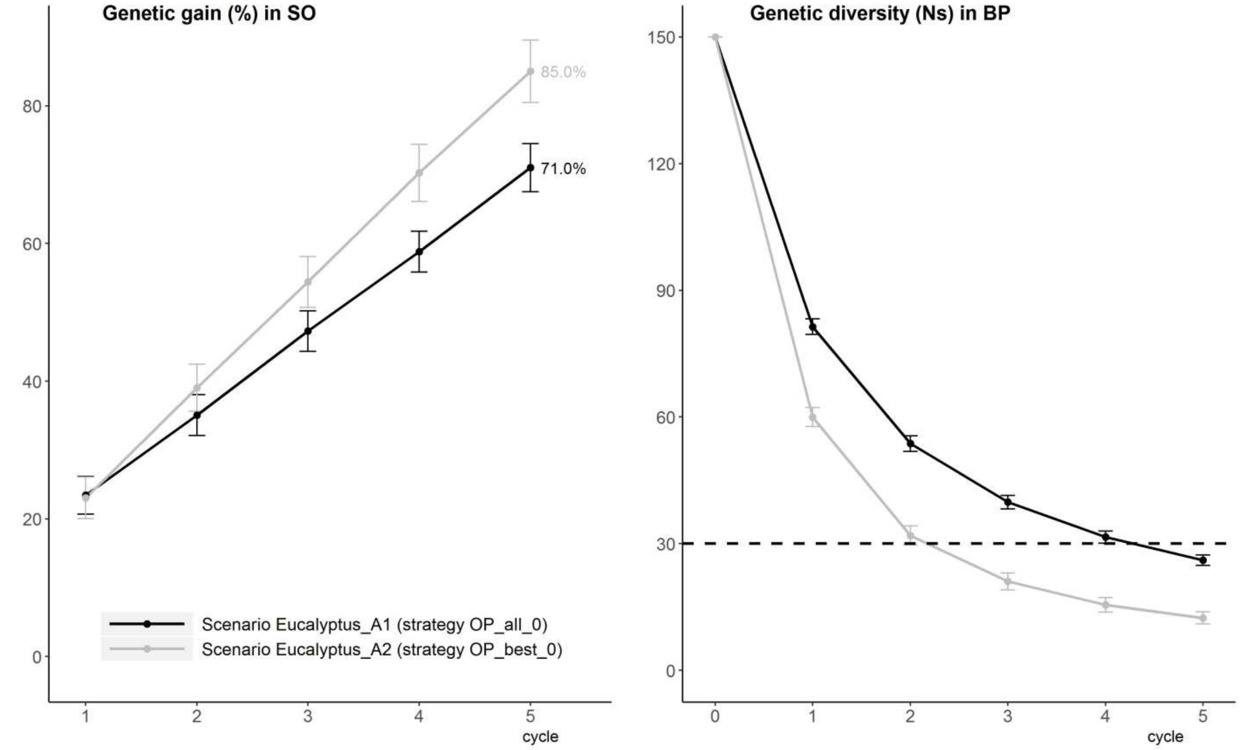


Figure 5: *Eucalyptus nitens* case-study – Breeding strategies with paternity testing

Genetic gain in SO and genetic diversity in BP over 5 breeding cycles (each point is a mean of 100 iterations and the error bar represents ±1 standard deviation).

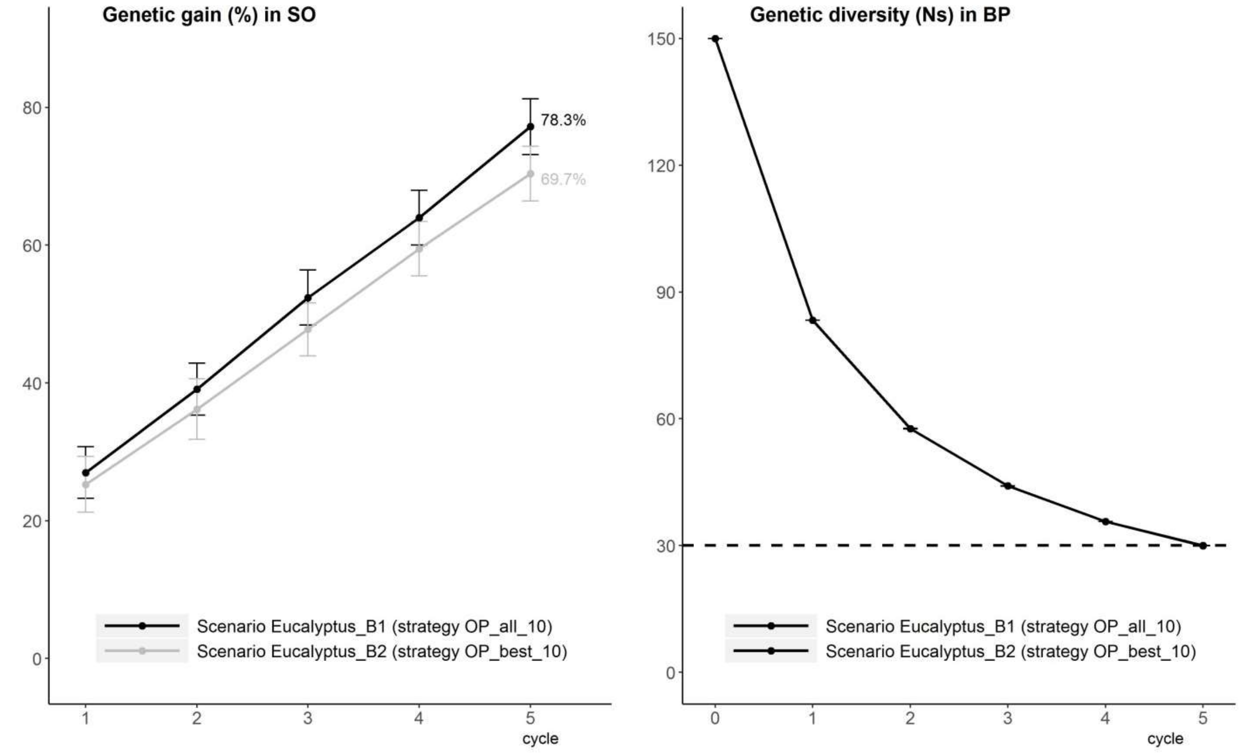


Figure 6: Genetic gain at breeding cycle 5 for increasing family sizes

Genetic gain in SO over 5 breeding cycles was simulated for DPM strategy (150 crosses or families) with 5 levels of family size (each point is a mean of 50 iterations and the error bar represents ±1 standard deviation).

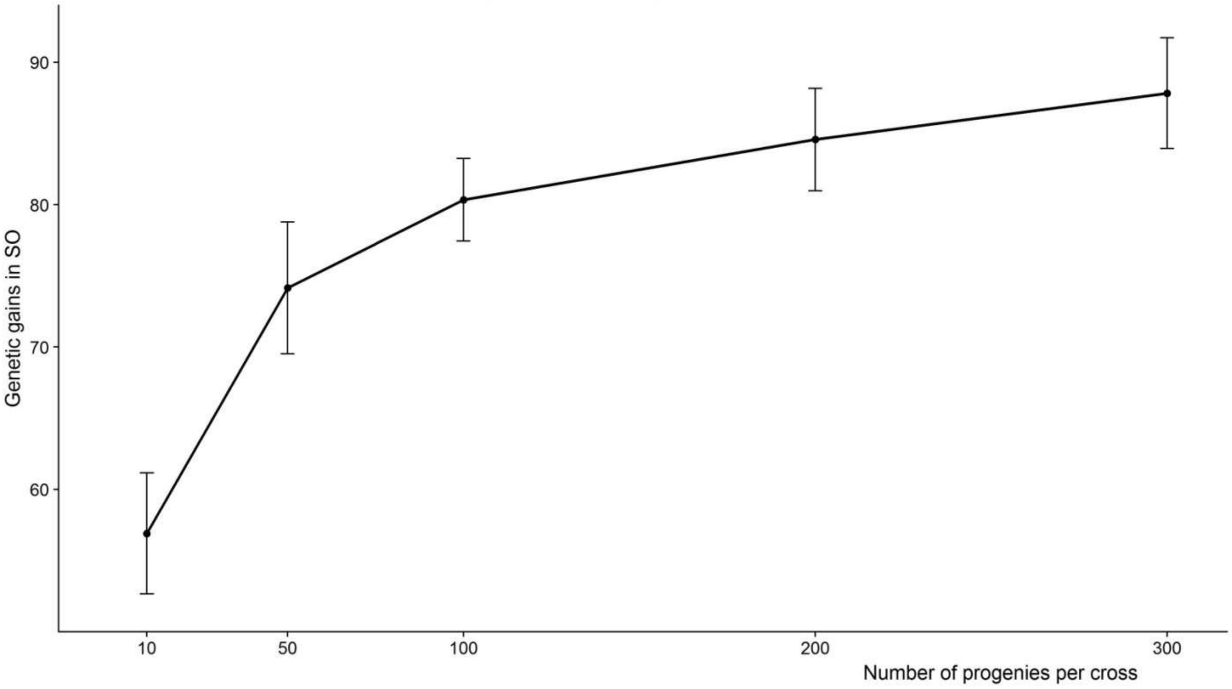


Figure 7: Genetic gain at breeding cycle 5 for increasing numbers of crosses

Genetic gain in SO over 5 cycles was simulated for increasing numbers of bi-parental crosses under a constant RP size of 15 000 (each point is a mean of 50 iterations and the error bar represents ±1 standard deviation).