



**HAL**  
open science

## 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

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

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



Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License

1 **Title**

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

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

5 \* corresponding author

6

7 Laurent Bouffier:

8 BIOGECO, INRA, Univ. Bordeaux, 33610 Cestas, France

9 laurent.bouffier@inra.fr / +33 5 35 38 53 60

10

11 Jaroslav Klápště :

12 Scion, Te Papa Tipu Innovation Park, 49 Sala Street, Private Bag 3020 Rotorua,

13 3046, New Zealand

14 jaroslav.klapste@scionresearch.com

15

16 Mari Suontama :

17 Scion, Te Papa Tipu Innovation Park, 49 Sala Street, Private Bag 3020 Rotorua,

18 3046, New Zealand

19 Present address: Skogforsk (The Swedish Forestry Research Institute), Box 3, Sävar, Sweden S-

20 918 21

21 Mari.Suontama@skogforsk.se

22

23 Heidi S. Dungey,

24 Scion, Te Papa Tipu Innovation Park, 49 Sala Street, Private Bag 3020 Rotorua,

25 3046, New Zealand

26 Heidi.Dungey@scionresearch.com

27  
28 Tim J. Mullin,  
29 Skogforsk (The Swedish Forestry Research Institute), 224 rue du Grand Royal Est, Shefford,  
30 QC, Canada J2M 1R5; and Box 3, Sävar, Sweden S-918 21  
31 Tim.Mullin@skogforsk.se

32 **Abstract**

33

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

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

50

51 **Keywords**

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

53

54

55 **Résumé**

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

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

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

74

75 **Introduction**

76

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

90

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

100 forest tree breeding has been reported to our knowledge even if many proof-of-concept studies  
101 have been published (Grattapaglia 2017).

102  
103 Furthermore, molecular markers have also been used with success in various forest tree studies  
104 for pedigree reconstruction to monitor genetic diversity and level of pollen contamination in seed  
105 orchards (Dering et al. 2014; Lai et al. 2010) or to improve accuracy of genetic parameters and  
106 breeding values (Doerksen and Herbinger 2010; El-Kassaby et al. 2011; Klápště et al. 2017;  
107 Vidal et al. 2015). Only a limited number of markers are required for such studies, which makes  
108 the approach applicable to most forest tree species. Besides, pedigree reconstruction could  
109 enhance the development of new breeding schemes as proposed by Lambeth et al. (2001) with  
110 the “polymix breeding with paternity analysis” strategy, then by El-Kassaby and Lstiburek  
111 (2009) with the “breeding without breeding” concept and similarly by Hansen and McKinney  
112 (2010) with the “quasi-field trial” approach. In all cases, the traditional bi-parental mating design  
113 is substituted by a two-step process: first, a designed mating scheme using polycross or open-  
114 pollination, then parental reconstruction allowing the completion of incomplete pedigree,  
115 subsequent breeding values calculation and selection. Breeding schemes based on pedigree  
116 reconstruction have several advantages such as a simplification of the crossing process,  
117 generation of a large number of families with a timely mixing of the breeding population and  
118 verification of identities for selected genotypes. However, the attractiveness of such breeding  
119 strategies based on pedigree reconstruction depends on the specifics of the breeding programme  
120 under consideration (current breeding strategy, biological constraints of the species, annual  
121 investments, technical skills available, etc.).

122  
123 The objective of this study is to evaluate the efficiency of breeding strategies based on partial  
124 pedigree reconstruction in two contrasting breeding programmes: *Pinus pinaster* (Ait.) in France  
125 and *Eucalyptus nitens* (H. Deane & Maiden) in New-Zealand. The first is characterised by

126 forward selection in bi-parental control-cross trials associated with a large investment (Bouffier  
127 et al. 2008), whereas the second is managed by forward selection in open-pollinated trials with a  
128 low level of investment (Klápště et al. 2017). For each case-study, alternative strategies based on  
129 pedigree reconstruction were proposed, and genetic gains achieved in the deployment population  
130 were compared with those from the current strategy through stochastic simulations. Quantifying  
131 and tracking the genetic variability in forest tree breeding populations is vital in order to keep  
132 enough variability for further breeding cycles (sustainable generation of long-term genetic gain)  
133 and to avoid inbreeding in the seed orchards (Stoehr et al. 2008). That is why, for a fair  
134 comparison, the same level of genetic variability was considered, regardless of the strategy.  
135 Finally, the advantages and limits of breeding strategies based on pedigree reconstruction are  
136 discussed in the context of forest tree breeding.



## 137 **Material and methods**

138

### 139 ***1. Stochastic simulations***

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

159 
$$N_s = \frac{1}{2 \times \theta}$$
 where  $\theta$  is the group coancestry (Cockerham 1967).

160

### 161 ***2. Simulation process***

162 Simulations were performed on a single trait, which can also be considered as a selection index.  
163 Typical genetic parameters for traits related to forest tree growth were considered (Cornelius  
164 1994). The additive coefficient of variation was fixed at 15% and heritability at 0.2 (i.e., trait  
165 mean = 100, additive variance = 225, environmental variance = 900). The BP size was fixed at  
166 150 trees in accordance with the two case-studies considered (*Pinus pinaster* and *Eucalyptus*  
167 *nitens*), and various mating designs specific for each case-study (see below) were applied to  
168 generate a RP of 15 000 trees. For each breeding cycle, a first selection step was carried out  
169 through BLUP evaluation (based on phenotypes and pedigree information) to form the next-  
170 generation BP of 150 trees. A second selection step was performed in this new BP to select  
171 genotypes according to their BLUP EBVs for the establishment of a SO. The selected proportion  
172 is thus 1% for the BP, which is the order of magnitude currently considered in the *Pinus pinaster*  
173 and *Eucalyptus nitens* breeding programmes. Some diversity constraints, detailed below for each  
174 case-study, were applied for these two selection steps on  $N_s$  if the full pedigree is known or, if  
175 not, on the maximum number of trees per family. For each simulation, 5 successive breeding  
176 cycles were generated with the same BP size, mating design and diversity constraints. When the  
177 full pedigree of the RP was not known, paternity testing was applied on a subset (RP') created  
178 through balanced within-family selection. In that case, the selection carried out to generate the  
179 new BP and the SO was realised only within the subset where paternal identity was recovered.  
180 This selection was based on a BLUP evaluation integrating the recovered pedigree information.  
181 A paternity-testing factor was defined as the ratio between the RP' size and the BP size. For  
182 example, with a BP size of 150, a paternity testing factor of 20 means that 3000 trees (out of the  
183 15 000 trees of the RP) were genotyped for paternity testing. This step carried out with POPSIM  
184 simulates a partial paternity recovery based on molecular markers (no error or unknown parents  
185 from pollen contamination were considered in the paternity recovery process).

186 In this study, a breeding scenario, illustrated in Figure 2, will be defined as a combination of a  
187 breeding strategy (a specific mating design eventually associated with partial paternity recovery)

188 and a set of diversity constraints applied on the BP and SO. Each scenario was simulated over 5  
189 breeding cycles with 100 iterations, producing a mean (and a standard deviation) for the genetic  
190 gain and diversity in the 5 successive BP and SO. The breeding scenarios are identified and  
191 summarized in Table 1 for *Pinus pinaster* and *Eucalyptus nitens* case-studies.

192

### 193 **3. *Pinus pinaster* case-study**

#### 194 **Breeding scenarios Pinus\_A and Pinus\_B**

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

196 i. Double-pair mating (DPM) in Pinus\_A1

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

203 ii. Polycross mating (PCM) in Pinus\_A2 and Pinus\_B

204 The best 50 trees of the BP were selected based on their BLUP EBVs and each was  
205 crossed with three different polymixes of 50 pollen parents to generate 100 progenies per  
206 polycross mating. Each tree of the BP was randomly assigned to one, and only one, of the  
207 three polymixes, which means that all 150 trees of the BP contribute as a pollen parent in  
208 the mating design. This design produced 50 half-sib families (same seed parent) with 300  
209 progenies each. Paternity testing factors varying from 0 to 50 (0, 5, 10, 25, or 50) were  
210 applied to genotype the trees with the highest BLUP EBVs within each half-sib family.  
211 The strategy is designated by PCM followed by the paternity testing factor. For example,  
212 the breeding scenario Pinus\_A2 involves the PCM\_5 strategy which corresponds to the  
213 genotyping of 750 trees (i.e., 5 x 150) selected as the best 15 trees from each of the 50

214 half-sib families. The best 150 trees in the RP were then selected under a diversity  
215 constraint using the MIQCO option in OPSEL.

216 In addition, an alternative PCM breeding strategy was considered in the breeding scenario  
217 Pinus\_B where the 50 trees in the BP used as seed parents were assigned randomly.  
218 Pinus\_B differs from Pinus\_A2 only by the selection process (random selection for  
219 Pinus\_B vs. the 50 best trees for Pinus\_A2).

### 220 iii. Optimal contribution (OC) in Pinus\_A3

221 The contribution from each parental selection to a controlled cross mating design was  
222 optimized with OPSEL, based on second-order cone programming (SOCP) methodology  
223 (Yamashita et al. 2018) to maximize the genetic gain at a given diversity level. The  
224 mating design for the optimized contributions was then generated with XDesign (Mullin  
225 2017b), avoiding coancestry between mates greater than 0.01. A total of 150 crosses were  
226 generated in the OC strategy with 100 progenies per cross.

227  
228 These three breeding strategies were evaluated under identical constraints on genetic diversity.  
229 When full pedigree was known (DPM, OC or PCM associated with paternity testing), an  
230 increase of group coancestry of maximum 0.00267 per generation was allowed in the BP (i.e.,  $N_s$   
231  $\geq 30$  after 5 cycles) and the genetic diversity of the SO was fixed at  $N_s \geq 10$  (i.e., group-  
232 coancestry  $\leq 0.05$ ), this was accomplished using the SOCP option in OPSEL to optimize unequal  
233 numbers of ramets to a grafted orchard. For the PCM\_0 strategy (in the breeding scenario  
234 Pinus\_A2'), as the paternal identity was not known, the three best trees per half-sib family were  
235 selected for the BP and a maximum of one progeny per seed parent was selected for the SO.

236

### 237 **Breeding scenarios Pinus\_C and Pinus\_D**

238 In order to estimate the sensitivity of the ranking among the three breeding strategies evaluated  
239 in Pinus\_A, RP size and diversity constraints were modified. The breeding scenarios Pinus\_C

240 and Pinus\_D involved the same breeding strategies evaluated in Pinus\_A, but the RP was  
241 increased by three times in Pinus\_C (i.e., RP = 45 000 instead of 15 000,  $N_s \geq 30$  after 5 cycles in  
242 the BP, and  $N_s \geq 10$  in the SO) and the diversity constraints were relaxed in Pinus\_D (i.e., RP =  
243 15 000,  $N_s \geq 15$  instead of 30 after 5 cycles in the BP, and  $N_s \geq 5$  instead of 10 in the SO).

244

#### 245 4. *Eucalyptus nitens* case-study

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

248 i. Open-pollination with seeds collected from all the trees (OP\_all)

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

251 ii. Open-pollination with seeds collected from 50 of the trees with the highest BLUP EBVs  
252 (OP\_best)

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

255

256 For each main breeding strategy, two paternity testing factors (0 and 10) were considered. When  
257 genotyping was performed, an identical number of trees per family was genotyped (10 trees /  
258 family for OP\_all, 30 trees / family for OP\_best), i.e., the paternity testing factor in each case  
259 equals 10. The strategies are designed by OP\_all or OP\_best followed by the paternity testing  
260 factor (0 or 10). Thus, four breeding strategies were considered for *Eucalyptus nitens*: OP\_all\_0,  
261 OP\_all\_10, OP\_best\_0 and OP\_best\_10.

262

263 These breeding strategies were evaluated under the same diversity constraints as for the breeding  
264 scenarios Pinus\_A, i.e.,  $N_s \geq 30$  after 5 cycles in the BP and  $N_s \geq 10$  in the SO when the full  
265 pedigree of the selection candidates was known (OP\_all\_10 and OP\_best\_10). When the pollen

266 parent identity was unknown, 1 tree (OP\_all\_0) or 3 trees (OP\_best\_0) per half-sib family were  
267 selected based on BLUP EBVs for the BP, and a maximum of one progeny per parent was  
268 selected for the SO in both cases. The *Eucalyptus* breeding scenarios, designated Eucalyptus\_A  
269 (no paternity testing) and Eucalyptus\_B (paternity testing), are summarized in Table 1.

270

## 271 **5. Complementary simulations**

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

278

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

284

285 In the second series, seven levels of bi-parental cross size were randomly generated (50, 75, 150,  
286 500, 1000, 2500 and 5000 crosses) with an equal contribution from each parent. The number of  
287 progenies per family was adjusted in order to keep a RP of 15,000 trees.

288

289 As these simulations were time-demanding, 50 iterations were run within each series. One  
290 scenario (150 crosses with 100 progenies per family) was performed with both 50 and 100

291 iterations and no significant difference in genetic gain was found (gain in SO is  $80.3\% \pm 2.9$  with  
292 50 iterations and  $79.5\% \pm 3.5$  with 100 iterations).

293 **Results**

294

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

297

298 ***Pinus pinaster* case-study**

299 Results from scenarios Pinus\_A are illustrated in Figure 3 (scenarios Pinus\_A1, Pinus\_A2 and  
300 Pinus\_A3). The OC strategy achieved greater genetic gain in the SO at cycle 5 (86.6%) than the  
301 DPM strategy (79.5%) and PCM\_5 strategy (70.1%), which means that the DPM and OC  
302 strategies outperformed the PCM\_5 strategy by a margin of 13.4% and 23.5%, respectively.  
303 Genetic gains for the PCM\_5 strategy were associated with a larger standard deviation (4.3) than  
304 for the OC (3.6) and DPM (3.5) strategies. As genetic diversity was constrained, the three  
305 strategies had  $N_s \geq 10$  in the SO regardless of the cycle and showed similar diversity decrease in  
306 the BP to reach  $N_s \geq 30$  after 5 breeding cycles. Various paternity testing factors (from 5 to 50)  
307 were considered for the PCM strategy, but these had no significant impact on the results. Thus,  
308 only the PCM\_5 (scenario Pinus\_A2) was reported in Figure 3. When a paternity testing factor  
309 below 5 was considered, it was not possible to meet the constraints for genetic diversity. A  
310 paternity testing factor greater than 50 was not considered because the simulations were too time  
311 consuming to be realized with 100 iterations. PCM\_100 was run over one iteration (simulation  
312 time = 11 hours) and achieved a genetic gain in the SO of 74.0%. However, as it was considered  
313 unrealistic to genotype more than 50% of the BP (7500 trees) as well as excessive simulation  
314 time, PCM strategies with a paternity testing factor greater than 50 were not evaluated. The  
315 genetic diversity constraints could not be applied for the PCM\_0 strategy due the unknown  
316 pollen parent identity. Even considering the stronger diversity constraint that can be applied on  
317 to this breeding strategy (i.e., selection of the 3 best trees for each of the 50 half-sib families),  $N_s$   
318 fell below 30 after 2 generations and reached 12.4 at cycle 5.



319  
320 Scenario Pinus\_B (PCM\_5 with a random selection of the 50 seed parents) showed a non-  
321 significant reduced genetic gain in the SO when compared to scenario Pinus\_A2 (67.2% with  
322 sd=4.1 vs 70.1% with sd=4.3).

323  
324 In the Pinus\_C scenarios, where the recruitment population was inflated from 15 000 to 45 000  
325 trees for the three breeding strategies, genetic gain in the SO increased whatever the strategy, but  
326 the differences between strategies slightly decreased compared with a RP of 15 000 trees: gain  
327 for PCM\_5 was 77.6%, 87.8% for DPM (+13.1%) and 94.7% for OC (+19.5%).

328  
329 In the Pinus\_D scenarios, where the diversity constraints were relaxed, greater genetic gain in  
330 the SO and lower differences between strategies were found: gain was 88.4% for PCM\_5, 95.8%  
331 for DPM (+8.4%) and 98.9% for OC (+11.0%).

332

### 333 ***Eucalyptus nitens* case-study**

334 When no paternity testing was performed, OP\_best\_0 (scenario Eucalyptus\_A2) exhibited a  
335 greater genetic gain in the SO than did OP\_all\_0 (scenario Eucalyptus\_A1) (85.0% vs. 71.0%),  
336 but genetic diversity in both the SO and BP decreased more rapidly for OP\_best\_0 than  
337 OP\_all\_0 (Figure 4). As paternity identity was unknown, the genetic diversity was controlled  
338 only based on the seed parent identity and both strategies were below  $N_s = 30$  in the BP at cycle  
339 5.  $N_s$  in the SO was greater than 10 for OP\_all\_0 up to cycle 5, but dropped below 10 at cycle 5  
340 for OP\_best\_0 (data not shown). When a paternity testing factor of 10 was considered, diversity  
341 constraints on  $N_s$  can be fulfilled, and thus both strategies were compared at a given diversity  
342 level ( $N_s \geq 30$  in the BP at cycle 5 and  $N_s \geq 10$  in the SO). In that case, OP\_all\_10 (scenario  
343 Eucalyptus\_B1) outperformed OP\_best\_10 (scenario Eucalyptus\_B2) for genetic gain in SO  
344 (77.2% vs. 70.4% at cycle 5) as illustrated in Figure 5.

345

### 346 **Complementary simulations**

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

350

351 For a given number of crosses (150 crosses considered here), the number of progenies had a  
352 strong impact on the genetic gain in the SO (Figure 6). As expected, the more progenies that  
353 were evaluated (from 10 to 300 per cross, i.e., an increasing RP size from 1500 to 45 000 trees),  
354 the more genetic gain was achieved. The gain increase was strong from 10 to 100 progenies per  
355 family: genetic gain was 56.9% with 10 progenies per cross and 80.3% with 100 progenies per  
356 cross. However, with over 100 progenies per cross, the gain increase was moderate (87.8% for  
357 300 progenies per cross).

358

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

365 **Discussion**

366  
367 Genetic diversity management is of fundamental importance in tree breeding (Rosvall 1999),  
368 first to limit inbreeding which induces inbreeding depression (Durel et al. 1996), and, second, to  
369 keep enough genetic variation for long-term genetic gains for the current selection criteria or for  
370 traits of interest in the future (Bouffier et al. 2008). Genetic diversity constraints are a key  
371 component when designing a breeding strategy. In the present study, when no diversity  
372 constraints were applied in the simulated breeding strategies,  $N_s$  in the BP dropped rapidly below  
373 20 (data not shown) regardless of the breeding strategy considered (at the latest in the second  
374 generation for the *Pinus pinaster* case-study and in the third cycle for the *Eucalyptus nitens* case-  
375 study). This erosion of genetic diversity is clearly not compatible with the management of a  
376 breeding programme for the long-term, where an effective population size in the order of 20 to  
377 80 is recommended (Danusevicius and Lindgren 2006; Ruotsalainen 2002). Diversity constraints  
378 were here considered in accordance with the current requirements of the two case-studies under  
379 study ( $N_s \geq 30$  after 5 cycles in the BP and  $N_s \geq 10$  in the SO).

380  
381 It is difficult to compare breeding strategies when the genetic diversity levels in the breeding and  
382 deployment populations are not fully controlled. This is why the simulation tool POPSIM, with  
383 its options for selection constrained on diversity with OPSEL, has been used in this study,  
384 allowing a given level of genetic diversity in both the BP and SO to be fixed and thus making  
385 breeding strategies readily comparable. The *Pinus pinaster* case-study deals with the polymix  
386 breeding with paternity recovery as an alternative to bi-parental breeding schemes. The  
387 *Eucalyptus nitens* case-study highlights the potential of adding a paternity recovery step in an  
388 open-pollination breeding scheme to better manage genetic diversity over time. Identical genetic  
389 parameters and population size were considered for both case-studies, as it was relevant to the  
390 characteristics of these two breeding programmes. It is also helpful for easy comparison,

391 enabling the formulation of some general conclusions on the use of paternity recovery in tree  
392 breeding strategies.

393

### 394 **Paternity testing**

395 The simulations performed in this study are based on the hypothesis that each genotype of the  
396 polymix (PCM or OP design) contributes equally and at random to the RP. Paternity recovery in  
397 two polycross trials of the French *Pinus pinaster* breeding programme (Vidal et al. 2015) broadly  
398 confirms this hypothesis, as only a few genotypes contribute more or less than expected.  
399 Paternity recovery carried out in open-pollinated *Eucalyptus nitens* seed orchard (Grosser et al.  
400 2010) exhibits a larger variation in pollen contribution, partially due to the variation in the umbel  
401 number per tree. If the pollen contribution is unbalanced, the optimal level of genotyping could  
402 be underestimated in these simulations. However, in both case-studies, paternity testing rate does  
403 not significantly impact the genetic gain in the SO. This can be explained by clear identification  
404 of the best performers with a low paternity testing rate which did not change with a higher  
405 paternity testing rate. The advantage of scenarios using higher paternity rates is not to generate  
406 an additional genetic gain but rather to identify genotypes to meet genetic diversity constraints  
407 even if they do not perform quite as well. Therefore, paternity testing has to be sufficient to fulfil  
408 diversity constraints, but once this threshold is achieved, there is no gain from increasing the  
409 genotyping rate. Based on our diversity constraints, the paternity testing factor can be set to 5  
410 (750 trees). When the paternity testing factor is below 5, the diversity constraints cannot be  
411 fulfilled. A paternity testing factor of 5 was chosen for *Pinus pinaster*, as no major deviation  
412 from equal contribution is expected for the pollen parents; but a higher level (10%) was  
413 considered for *Eucalyptus nitens* as more variable paternal contributions were expected.

414

415 Another underlying assumption is the absence of error in paternity recovery. Whereas, the  
416 hypothesis seems reasonable in the PCM strategy, where mixes of 50 pollen parents are

417 considered, it can be more challenging in the OP strategy when dealing with 150 potential pollen  
418 parents that will be more and more related in subsequent generations. This increased relatedness  
419 between candidate parents will decrease discrimination power of parentage assignment and will  
420 result in increased occurrence of false positives (Olsen et al. 2001). The best option to bypass  
421 this difficulty would be to estimate the relatedness with the markers instead of recovering the  
422 pedigree. This implies using a G-matrix (markers-based) instead of the A-matrix (pedigree  
423 based), a methodology used in genomic selection studies (El-Dien et al. 2016). Whereas less than  
424 100 well-chosen markers can be sufficient for a paternity analysis (Vidal et al. 2015), several  
425 thousands of markers will be necessary to recover relatedness with a G-matrix (Wang 2016).

426

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

428 The DPM strategy, currently used for the French *Pinus pinaster* breeding programme, mixes  
429 efficiently the BP as all trees contribute equally to the RP. Nevertheless, the mating design is  
430 challenging to fulfil as it involves bi-parental crosses. The PCM\_5 strategy can be an alternative  
431 to simplify the mating as a large number of families is generated with a limited number of  
432 control-crosses. In the simulated PCM design, three mixes of 50 pollen parents, which represent  
433 the entire BP, were applied on a restricted set of 50 seed parents. The limited number of parents  
434 involved in the pollen mixes is chosen to facilitate the pedigree recovery, particularly in the  
435 advanced cycles where the relatedness in the BP tends to be high. In addition, the restriction on  
436 the number of seed parents allows the focus of important resources on these genotypes (grafting,  
437 flowering induction, pruning), whereas it is generally not necessary to graft in a clonal archive  
438 the genotypes used only for pollen collection.

439

440 The DPM strategy outperforms the PCM\_5 strategy in the delivery of genetic gain by 13.4%  
441 after 5 cycles, when RP=15 000 (scenarios Pinus\_A1 vs. Pinus\_A2), and this difference is  
442 similar (+13.1%) for a larger RP (RP=45 000 in scenarios Pinus\_C). The DPM and PCM mating

443 designs involve exactly the same number of crosses (150 bi-parental crosses for DPM vs. 50  
444 seed parents x 3 polymixes =150 control-crosses for PCM), but the PCM design generates more  
445 full-sib families (potentially  $50 \times 50 \times 3 = 7500$  families) than the DPM design (150 families).  
446 However, for a given RP size, the number of crosses is not a key driver for the gain as shown in  
447 Figure 7, at least over a threshold of about 150 crosses. This can be explained by a trade-off  
448 between within-family and between-family selection intensity. In contrast, diversity constraints  
449 play a major role in the BP and SO selection. The DPM design involves each parent equally in  
450 the RP, whereas the PCM design simulated over-represents the 50 best parents selected as seed  
451 parents. When the selection is carried out under high diversity, the PCM design is penalized for  
452 gains due to this over-representation of the best parents. The decrease of gain difference between  
453 DPM and PCM\_5 strategies when the diversity constraints are relaxed (5.3% when  $N_s \geq 15$  in  
454 the BP at cycle 5 and  $N_s \geq 5$  in the SO in scenarios Pinus\_D) confirms that the diversity  
455 constraint is a key driver for the selection step. If the 150 parents have been represented equally  
456 in the PCM design (scenario equivalent to scenario Eucalyptus\_B1 in Table 1), PCM would have  
457 performed equally well ( $77.2\% \pm 4.1$ ) as DPM ( $79.5\% \pm 3.5$  in scenario Pinus\_A1). We conclude  
458 that there is no disadvantage or advantage to generate more full-sib families for a given RP size;  
459 but when the diversity constraints are strong, the DPM strategy outperforms the PCM\_5 strategy  
460 if parents contribute unequally in the PCM mating design. However, both DPM and PCM\_5  
461 strategies are outperformed by the OC strategy where the 150 crosses use unequal contributions  
462 from selected parents to optimize gain under a specific level of diversity constraints. Scenario  
463 Pinus\_A3 (OC strategy) displays 23.5% more gain in the SO than scenario Pinus\_A2 (strategy  
464 PCM\_5). This result confirms the superiority of the OC approach to maximize gain under a  
465 given level of genetic diversity (Hallander and Waldmann 2009; Meuwissen 1997).

466  
467 The superiority of the OC strategy over DPM and PCM strategies has to be mitigated based on  
468 the time required to complete the design, which has an impact on the length of the breeding

469 cycle, and thus the genetic gain per unit time. Generally, bi-parental crosses (OC and DPM  
470 designs) are more complex to implement than polymix crosses (PCM design), as mentioned  
471 earlier. In addition, as the random selection of the seed parent subset for the PCM design in  
472 scenario Pinus\_B has only a small impact on genetic gain, in comparison to scenario Pinus\_A2,  
473 one can focus on the 50 genotypes more convenient to breed as seed parents (earlier flowering,  
474 greater number of flowers, etc.) making the PCM design even easier to complete. Furthermore,  
475 the OC design is also probably more sophisticated than a DPM design, as there is a unique  
476 mating design that reaches the optimum gain, whereas several mating alternatives can respect the  
477 DPM strategy, giving greater flexibility (the only requirement for the DPM design is to cross  
478 each parent twice). For the French *Pinus pinaster* breeding programme, the current generation  
479 time is roughly 20 years. As DPM and OC designs outperform PCM design from 2.7 to 4.7% per  
480 breeding cycle on average, a cycle shortening of one year due to faster completion of the mating  
481 design with PCM strategy would give greater gain per unit time. However, the precise time  
482 required for each mating operation is currently difficult to estimate for the French breeding  
483 programme, as it faces low success for control crosses.

484

485 The genotyping cost is an extra investment for polycross breeding with paternity recovery. It has  
486 to be taken into account when selecting the most appropriate scenario, as breeders generally  
487 focus on the genetic gain per unit time for a given budget. With the decreasing cost of  
488 genotyping, the genotyping of 750 to 1500 trees should not be a major limitation. In the French  
489 *Pinus pinaster* breeding programme, the cost of paternity recovery is currently roughly 10€ per  
490 tree, which means an additional cost of 7500€ per cycle for the Pinus\_A2 scenario. However,  
491 one has to keep in mind that this genotyping step also allows the verification of relatedness at  
492 each cycle. If no genotyping is performed in the DPM and OC strategies, it is likely that there  
493 will be an accumulation of errors in the pedigree (Munoz et al. 2014) that will affect the gain.

494

## 495 ***Eucalyptus nitens* and open-pollination breeding**

496 The current OP strategy implemented for the *Eucalyptus nitens* breeding programme in New  
497 Zealand, adopted due to biological constraints, does not allow the full management of genetic  
498 diversity as only the identity of the seed parent is known. While this is not necessarily an issue in  
499 the first few breeding cycles, it becomes a major concern after 4 cycles. Indeed, even with the  
500 strongest constraints applied on the seed parent identity in the BP (selection of 1 tree per half-sib  
501 family in scenario Eucalyptus\_A1 and 3 trees per half-sib family in scenario Eucalyptus\_A2), *N<sub>s</sub>*  
502 drops below 30 in the BP at cycle 5 (Figure 4) if the full pedigree is unknown. As bi-parental  
503 crosses are not practicable due to biological and technical constraints, a genotyping step has to  
504 be carried out to recover the full-pedigree before selecting trees for the BP and SO. When  
505 paternity testing is considered, OP\_all design (in scenario Eucalyptus\_B1) outperforms OP\_best  
506 design (in scenario Eucalyptus\_B2). As discussed for polymix breeding scenario, the equal  
507 contribution of each parent in OP\_all design makes it easier to meet diversity constraints, and  
508 thus achieve greater gains in SO. We conclude that it is preferable to collect seeds from the  
509 whole population, rather than focusing on genotypes with the highest BLUP EBVs. Besides the  
510 diversity management issue, paternity recovery also increases the EBVs accuracy for the selected  
511 trees. Additionally, in species with viable selfing such as Eucalyptus, identification of  
512 individuals coming from selfing through pedigree reconstruction allows their elimination from  
513 the BP and improvement of accuracy in genetic parameter estimates, especially in traits suffering  
514 from inbreeding depression (Klápště et al. 2017).

515

## 516 **Simulation limitations**

517 The POPSIM software is specially designed to simulate various forest tree breeding strategies at  
518 a given level of diversity in the successive BP and SO. However, in the PCM design, all the  
519 polymixes must be mated with the same set of seed parents. Without this limitation, and since an  
520 equal-parental contribution in the BP seems preferable to better manage the diversity, two



521 alternative PCM mating designs could have been considered: the first one where each parent  
522 would be used as either pollen or seed parent, and the second where the whole BP would be used  
523 as both pollen and seed parents. This implies, for the first alternative, to divide the BP into two  
524 sets of 75 trees: 75 pollen parents (eventually two polymixes of 37 and 38 pollen parents can be  
525 designed to facilitate the paternity recovery) applied on the 75 remaining trees. The second  
526 alternative would be similar to that considered in this paper, but with each of the three polymixes  
527 being crossed with three different sets of 50 seed parents. It is expected that, for the same  
528 number of crosses (150 crosses), these strategies might achieve slightly greater gains than the  
529 PCM design evaluated in the paper. However, both alternatives imply grafting more than 50  
530 trees (75 for the first and 150 for the second), which would complicate the field operations.

531  
532 The mating design initially planned by the breeder is sometimes difficult to complete in the field  
533 (Kerr et al. 2015). A key issue is thus to estimate the consequences on the gain when there is a  
534 deviation for the initial mating design. This has not been studied by Kerr et al, but should be  
535 taken into account when selecting the more appropriate breeding scenario. Similarly, the impact  
536 of deviation from equal pollen contribution for PCM and OP designs was not evaluated in this  
537 paper. It is probably not a major issue for PCM design, as discussed previously with the *Pinus*  
538 *pinaster* case-study; but for OP design, phenology and pollen production variation could result in  
539 unequal pollen contributions in the RP. On the contrary, the pollination systems, wind  
540 pollination for conifers (Owens et al. 1998) and insect pollination for *Eucalyptus nitens* (Barbour  
541 et al. 2005), should not impact pollen contribution in the OP design, as pollen is dispersed far  
542 away from its source in both cases.

543  
544 Finally, two additional limitations can be highlighted. First, simulations carried out in this paper  
545 considered only discrete generations. The “rolling front” breeding scheme has been proposed as  
546 an alternative: parents are crossed as soon as possible after selection and progeny trials are

547 established as soon as the seeds are available, rather than waiting for all crosses in that  
548 generation to be completed. It has been demonstrated that the “rolling front” breeding scheme  
549 might achieve greater genetic gain per unit time in comparison with schemes based on discrete  
550 generations (Borralho and Dutkowski 1998). However, all the breeding strategies evaluated here  
551 can also be implemented as “rolling front”, although not specifically addressed here. Second, the  
552 level of genotyping chosen was applied by selecting, within each family, the trees with the  
553 highest BLUP EBVs (evaluated with their seed-parent identity and their own phenotype). This  
554 could bias the genetic parameter estimations carried out after paternity recovery; adding  
555 “random” trees in the genotyping process could counteract this. However, as a limited number of  
556 trees were genotyped (5 to 10% of the BP), only slightly biased estimates are expected.

557 **Conclusion**

558

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

565 1. The knowledge of the full pedigree is mandatory to manage genetic diversity in the long  
566 term. This can be achieved through bi-parental crosses or through paternity recovery in  
567 polymix or open-pollination breeding.

568 2. For a given RP size, increasing the number of full-sib families above a threshold (which  
569 corresponds approximately to the number of parents in the BP) is not of interest whatever  
570 the mating design (bi-parental or polymix breeding). In contrast, if additional investments  
571 are available, an increase of the RP size is useful to enhance gain in the SO (Figure 6).

572 3. The mating design is a key issue to maximize the gain for a given level of diversity,  
573 especially if diversity constraints are strong. The OC strategy is clearly the most effective  
574 way to optimize the gain per cycle (+23.5% over 5 cycles in the *Pinus pinaster* case-  
575 study in comparison to the PCM strategy), as the mating design is carried out taking into  
576 account relatedness constraints. However, this optimal mating design can be time-  
577 consuming to complete and, even sometimes, not feasible. Depending on the breeding  
578 context, alternative mating design can be considered, some followed by a paternity  
579 recovery step. If they are faster to complete, they can produce more genetic gain per time  
580 unit than the OC strategy.

581 Moreover, the genetic gain is expressed, in this paper, per cycle with no time scale.

582 Breeders have to keep in mind that it is generally more efficient to rapidly turn over a

583 new cycle, than trying to optimize the gain per cycle by extending the period required to  
584 complete the mating design.

585 4. Alternative strategies that include paternity recovery are based on polymix and open-  
586 pollination breeding. In contrast with the OC strategy, where each parental contribution is  
587 optimized for gain given the diversity constraints, these alternative strategies are based on  
588 random mating between sets of parents. In that context, optimal selection must achieve  
589 gain through an equal contribution from each parent selected from the RP. This means,  
590 for OP strategy, the collection of an equal number of seeds on each parent. The trial  
591 design must also prevent a large deviation from an equal paternal contribution. For the  
592 PCM strategy, an equal contribution of each parent could be time-consuming to achieve.  
593 Crosses can be focused on a limited number of seed parents and keeping wide genetic  
594 diversity in the polymix composition as considered in this paper. The gain is not optimal,  
595 but it allows the completion of the crosses in limited time, as the number of trees grafted  
596 is reduced.

597 5. Interestingly, a limited level of paternity recovery is sufficient to achieve an optimal gain  
598 in the SO (a paternity testing factor of 5 for PCM strategy and 10 for OP strategy in this  
599 paper). The paternity testing rate must be adjusted depending on the expected deviation  
600 from equal paternal contribution (higher rate when contributions are more unequal), the  
601 diversity constraints (higher rate with stronger constraints), the pedigree errors suspected  
602 in BP (higher rate if pedigree records are not accurate) and genotyping costs.

603 6. Two major benefits are associated with the strategies based on paternity testing. The first  
604 is to cleanse the pedigree of errors at each generation, at least for the genotypes selected.  
605 This is a key point to achieve more accurate EBVs. The second is to facilitate the  
606 predictable transition from current breeding strategies, based only on pedigree, to more  
607 advanced breeding strategies based on molecular markers and genomic selection, more  
608 particularly single-step methodology (Klápště et al. 2018).

609 **Acknowledgements**

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

618 **References**

- 619 Barbour, R.C., Potts, B.M., and Vaillancourt, R.E. 2005. Pollen dispersal from exotic eucalypt plantations.  
620 Conserv. Genet. **6**(2): 253-257. doi: 10.1007/s10592-004-7849-z.  
621
- 622 Borralho, N.M.G., and Dutkowski, G.W. 1998. Comparison of rolling front and discrete generation  
623 breeding strategies for trees. Can. J. For. Res. **28**: 987-993.  
624
- 625 Bouffier, L., Klápště, J., Bartholome, J., Plomion, C., Telfer, E., Graham, N., and Dungey, H. 2018.  
626 Genomic selection proof-of-concept in two major conifer species. In Proceedings of the World Congress  
627 on Genetics Applied to Livestock Production, Auckland, New Zealand.  
628
- 629 Bouffier, L., Raffin, A., and Kremer, A. 2008. Evolution of genetic variation for selected traits in successive  
630 breeding populations of maritime pine. Heredity (Edinb) **101**(2): 156-165. doi:  
631 <https://doi.org/10.1038/hdy.2008.41>.  
632
- 633 Cockerham, C.C. 1967. Group inbreeding and coancestry. Genetics **56**(89-104).  
634
- 635 Cornelius, J. 1994. Heritabilities and additive genetic coefficients of variation in forest trees. Can. J. For.  
636 Res. **24**(2): 372-379.  
637
- 638 Danusevicius, D., and Lindgren, D. 2006. Optimization of breeding population size for long-term breeding.  
639 Scand. J. For. Res. **20**(1): 18-25. doi: 10.1080/02827580410019517.  
640
- 641 Dering, M., Misiorny, A., and Chałupka, W. 2014. Inter-year variation in selfing, background pollination,  
642 and paternal contribution in a Norway spruce clonal seed orchard. Can. J. For. Res. **44**(7): 760-767. doi:  
643 10.1139/cjfr-2014-0061.  
644
- 645 Doerksen, T.K., and Herbinger, C.M. 2010. Impact of reconstructed pedigrees on progeny-test breeding  
646 values in red spruce. Tree Genetics & Genomes **6**(4): 591-600. doi: 10.1007/s11295-010-0274-1.  
647
- 648 Dungey, H., Brawner, J., Burger, F., Carson, M., Henson, M., Jefferson, P., and Matheson, A. 2009. A  
649 new breeding strategy for *Pinus radiata* in New Zealand and New South Wales. Silvae Genet. **58**(1-2):  
650 28-38.  
651
- 652 Durel, C.E., Bertin, P., and Kremer, A. 1996. Relationship between inbreeding depression and inbreeding  
653 coefficient in maritime pine (*Pinus pinaster*). Theor Appl Genet **92**: 347-356.  
654
- 655 El-Dien, O.G., Ratcliffe, B., Klápště, J., Porth, I., Chen, C., and El-Kassaby, Y.A. 2016. Implementation of  
656 the realized genomic relationship matrix to open-pollinated white spruce family testing for disentangling  
657 additive from nonadditive genetic effects. G3: Genes Genom. Genet. **6**: 743-753. doi:  
658 10.1534/g3.115.025957/-/DC1.  
659
- 660 El-Kassaby, Y.A., Cappa, E.P., Liewlaksaneeyanawin, C., Klápště, J., and Lstibůrek, M. 2011. Breeding  
661 without breeding: is a complete pedigree necessary for efficient breeding? PLoS One **6**(10). doi:  
662 10.1371/journal.pone.0025737.g001.  
663
- 664 El-Kassaby, Y.A., and Lstiburek, M. 2009. Breeding without breeding. Genetics research **91**(2): 111-120.  
665 doi: 10.1017/S001667230900007X.  
666
- 667 Gilmour, A.R., Gogel, B.J., Cullis, B.R., and Thompson, R. 2009. ASReml user guide release 3.0 VSN  
668 International Ltd, Hemel Hempstead, HP1 1ES, UK.  
669
- 670 Grattapaglia, D. 2017. Status and perspectives of genomic selection in forest tree breeding. In Genomic  
671 Selection for Crop Improvement Edited by Springer, New York. pp. 199-249.  
672
- 673 Grattapaglia, D., and Resende, M.D.V. 2011. Genomic selection in forest tree breeding. Tree Genet.  
674 Genomes **7**(2): 241-255. doi: 10.1007/s11295-010-0328-4.  
675
- 676 Grosser, C., Potts, B.M., and Vaillancourt, R.E. 2010. Microsatellite based paternity analysis in a clonal  
677 *Eucalyptus nitens* seed orchard. Silvae Genet. **59**(1-6): 57-62. doi: 10.1515/sg-2010-0007.

678  
679 Hallander, J., and Waldmann, P. 2009. Optimum contribution selection in large general tree breeding  
680 populations with an application to Scots pine. *Theor Appl Genet* **118**(6): 1133-1142. doi: 10.1007/s00122-  
681 009-0968-7.  
682  
683 Hansen, O.K., and McKinney, L.V. 2010. Establishment of a quasi-field trial in *Abies nordmanniana*—test  
684 of a new approach to forest tree breeding. *Tree Genet. Genomes* **6**(2): 345-355.  
685  
686 Isik, F. 2014. Genomic selection in forest tree breeding: the concept and an outlook to the future. *New*  
687 *Forest*. **45**(3): 379-401. doi: 10.1007/s11056-014-9422-z.  
688  
689 Kerr, R.J., MacRae, T.A., Dutkowski, G.W., and Tier, B. 2015. Managing the rate of increase in average  
690 co-ancestry in a rolling front tree breeding strategy. *J. Anim. Breed. Genet.* **132**: 109-120.  
691  
692 Klápště, J., Suontama, M., Dungey, H.S., Telfer, E.J., Graham, N.J., Low, C.B., and Stovold, G.T. 2018.  
693 Effect of hidden relatedness on single-step genetic evaluation in an advanced open-pollinated breeding  
694 program. *J. Hered.* **109**(7): 802-810. doi: 10.1093/jhered/esy051.  
695  
696 Klápště, J., Suontama, M., Telfer, E., Graham, N., Low, C.B., Stovold, T., McKinley, R.B., and Dungey, H.  
697 2017. Exploration of genetic architecture through sib-ship reconstruction in advanced breeding population  
698 of *Eucalyptus nitens*. *PLoS One* **12**(9). doi: 10.1371/journal.pone.0185137.  
699  
700 Lai, S.B., Funda, T., Liewlaksaneeyanawin, C., Klápště, J., Niejenhuis, A., Cook, C., Stoehr, M.U.,  
701 Woods, J., and El-Kassaby, Y.A. 2010. Pollination dynamics in a Douglas-fir seed orchard as revealed by  
702 pedigree reconstruction. *Ann. For. Sci.* **67**(8): 808-808. doi: 10.1051/forest/2010044.  
703  
704 Lambeth, C., Lee, B.-C., O'Malley, D., and Wheeler, N. 2001. Polymix breeding with parental analysis of  
705 progeny: an alternative to full-sib breeding and testing. *Theor Appl Genet* **103**: 930-943.  
706  
707 Lee, S. 2001. Selection of parents for the Sitka spruce breeding population in Britain and the strategy for  
708 the next breeding cycle. *Forestry* **74**(2): 129-143.  
709  
710 Lindgren, D., Gea, L., and Jefferson, P. 1996. Loss of genetic diversity monitored by status number.  
711 *Silvae Genet.* **45**(1): 52-58.  
712  
713 McKeand, S., and Bridgwater, F. 1998. A strategy for the third breeding cycle of loblolly pine in the  
714 Southeastern U.S. *Silvae Genet.* **47**(4): 223-234.  
715  
716 Meuwissen, T.H. 1997. Maximizing the response of selection with a predefined rate of inbreeding. *J.*  
717 *Anim. Breed. Sci.* **75**: 934-940.  
718  
719 Mullin, T.J. 2017a. OPSEL 2.0: A computer program for optimal selection in tree breeding.  
720  
721 Mullin, T.J. 2017b. XDesign 1.0: a tool for generation of complex mating designs with optimal  
722 contributions. *Edited by* S. Arbetsrapport från Skogforsk Nr. 956-2017, Uppsala, SE, 15 pp.  
723  
724 Mullin, T.J. 2018. POPSIM: a computer program for simulation of tree breeding programs over multiple  
725 generations. *Arbetsrapport från Skogforsk Nr. 984-2018*, Skogforsk, Uppsala, SE, 34 pp.  
726  
727 Mullin, T.J., Andersson, B., Bastien, J.-C., Beaulieu, J., Burdon, R.D., Dvorak, W.S., King, J.N., Kondo,  
728 T., Krakowski, J., Lee, S.J., McKeand, S.E., Pâques, L., Raffin, A., Russell, J., Skrøppa, T., Stoehr, M.,  
729 and Yanchuk, A.D. 2011. Chapter 2: Economic importance, breeding objectives and achievements. *In*  
730 *Genetics, genomics and breeding of conifers. Edited by* J.B.a.C.K. C. Plomion. Science Publishers, Inc.,  
731 New Hampshire; Edenbridge Ltd., UK.  
732  
733 Mullin, T.J., and Belotti, P. 2016. Using branch-and-bound algorithms to optimize selection of a fixed-size  
734 breeding population under a relatedness constraint. *Tree Genetics & Genomes* **12**(4). doi:  
735 10.1007/s11295-015-0961-z.  
736  
737 Mullin, T.J., and Lee, S.J. 2013. Best practice for tree breeding in Europe. Skogforsk, Uppsala, SE.  
738

739 Mullin, T.J., and Park, Y.S. 1995. Stochastic simulation of population management strategies for tree  
740 breeding: a new decision-support tool for personal computers. *Silvae Genet.* **44**(2-3): 132-141.  
741

742 Munoz, P.R., Resende, M.F.R., Huber, D.A., Quesada, T., Resende, M.D.V., Neale, D.B., Wegrzyn, J.L.,  
743 Kirst, M., and Peter, G.F. 2014. Genomic relationship matrix for correcting pedigree errors in breeding  
744 populations: impact on genetic parameters and genomic selection accuracy. *Crop Sci.* **54**(3): 1115-1123.  
745 doi: 10.2135/cropsci2012.12.0673.  
746

747 Muranty, H., Jorge, V., Bastien, C., Lepoittevin, C., Bouffier, L., and Sanchez, L. 2014. Potential for  
748 marker-assisted selection for forest tree breeding: lessons from 20 years of MAS in crops. *Tree Genet.*  
749 *Genomes* **10**: 1491-1510.  
750

751 Namkoong, G., Kang, G., and Brouard, J.S. 1988. *Tree breeding: principal and strategies.* . Springer-  
752 Verlag New York.  
753

754 Olsen, J.B., Busack, C., Britt, J., and Bentzen, P. 2001. The aunt and uncle effect: an empirical evaluation  
755 of the confounding influence of full sibs of parents on pedigree reconstruction. *J. Heredity* **92**: 243-247.  
756 Owens, J.N., Takaso, T., and Runions, C.J. 1998. Pollination in conifers. *Trends Plant Sci.* **3**(12): 479-  
757 484.  
758

759 Rosvall, O. 1999. Enhancing gain from long-term forest tree breeding while conserving genetic diversity.  
760 *In* Swedish University of Agricultural Sciences, Umeå, Sweden.  
761

762 Ruotsalainen, S. 2002. Managing breeding stock in the initiation of a long-term tree breeding program. *In*  
763 Finnish Forest Research Institute, Punkaharju Research Station.  
764

765 Stoehr, M., Yanchuk, A., Xie, C.-Y., and Sanchez, L. 2008. Gain and diversity in advanced generation  
766 coastal Douglas fir selections for seed production populations. *Tree Genet. Genomes* **4**(2): 193-200.  
767

768 Thistlethwaite, F.R., Ratcliffe, B., Klápště, J., Porth, I., Chen, C., Stoehr, M.U., and El-Kassaby, Y.A.  
769 2017. Genomic prediction accuracies in space and time for height and wood density of Douglas-fir using  
770 exome capture as the genotyping platform. *BMC Genomics* **18**(1): 930. doi: 10.1186/s12864-017-4258-5.  
771

772 Vidal, M., Plomion, C., Harvengt, L., Raffin, A., Boury, C., and Bouffier, L. 2015. Paternity recovery in two  
773 maritime pine polycross mating designs and consequences for breeding. *Tree Genet. Genomes* **11**(5):  
774 105.  
775

776 Wang, J. 2016. Pedigrees or markers: Which are better in estimating relatedness and inbreeding  
777 coefficient? *Theor. Pop. Biol.* **107**: 4-13. doi: 10.1016/j.tpb.2015.08.006.  
778

779 Wiggans, G.R., Cole, J.B., Hubbard, S.M., and Sonstegard, T.S. 2017. Genomic selection in dairy cattle:  
780 the USDA experience. *Annu. R. A. Biosci.* **5**: 309-327. doi: 10.1146/annurev-animal-021815-111422.  
781

782 Wu, H.X., Eldridge, K.G., Matheson, A., Powell, M.P., and McRae, T.A. 2007. Successful introduction and  
783 breeding of radiata pine to Australia. *Austral. For.* **70**(4): 215-225.  
784

785 Yamashita, M., Mullin, T.J., and Safarina, S. 2018. An efficient second-order cone programming approach  
786 for optimal selection in tree breeding. *Optim. Lett.* **12**: 1-15. doi: 10.1007/s11590-018-1229-y.  
787



788  
789

**Table 1: Description of the simulated breeding scenarios**

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

790  
791  
792  
793  
794  
795  
796

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

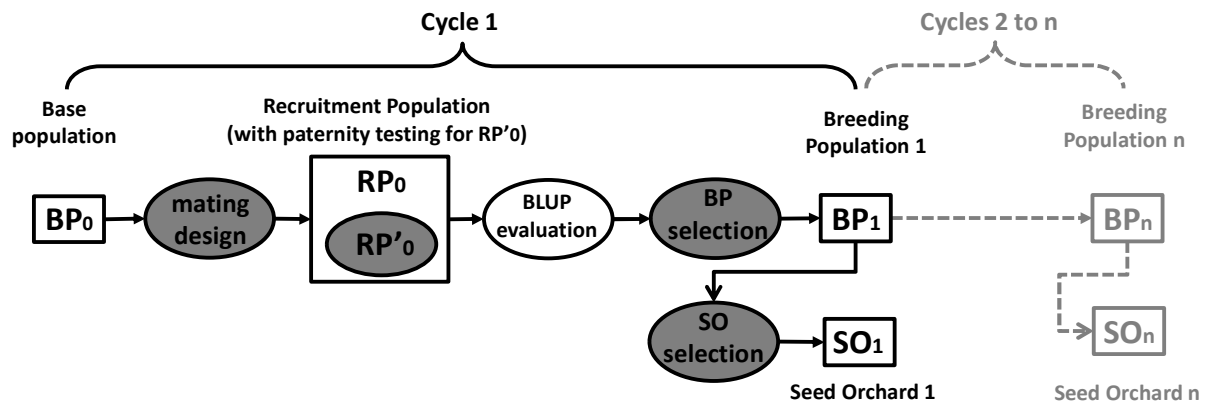
**Table 2: Genetic gain and genetic diversity of the simulated breeding scenarios**

Breeding scenario		Results at cycle 5		Results over 5 cycles	
		Genetic gain in SO	Ns in BP		
<i>Pinus pinaster</i>	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)	-	
<i>Eucalyptus nitens</i>	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	

797  
798  
799  
800

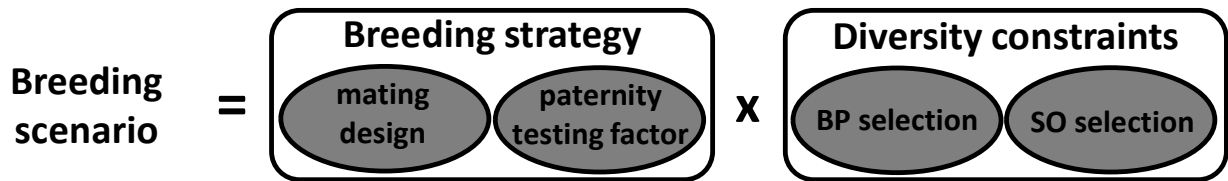
Note: The main simulated breeding scenarios are highlighted in grey.

801 **Figure 1: Simulation process with POPSIM**  
 802



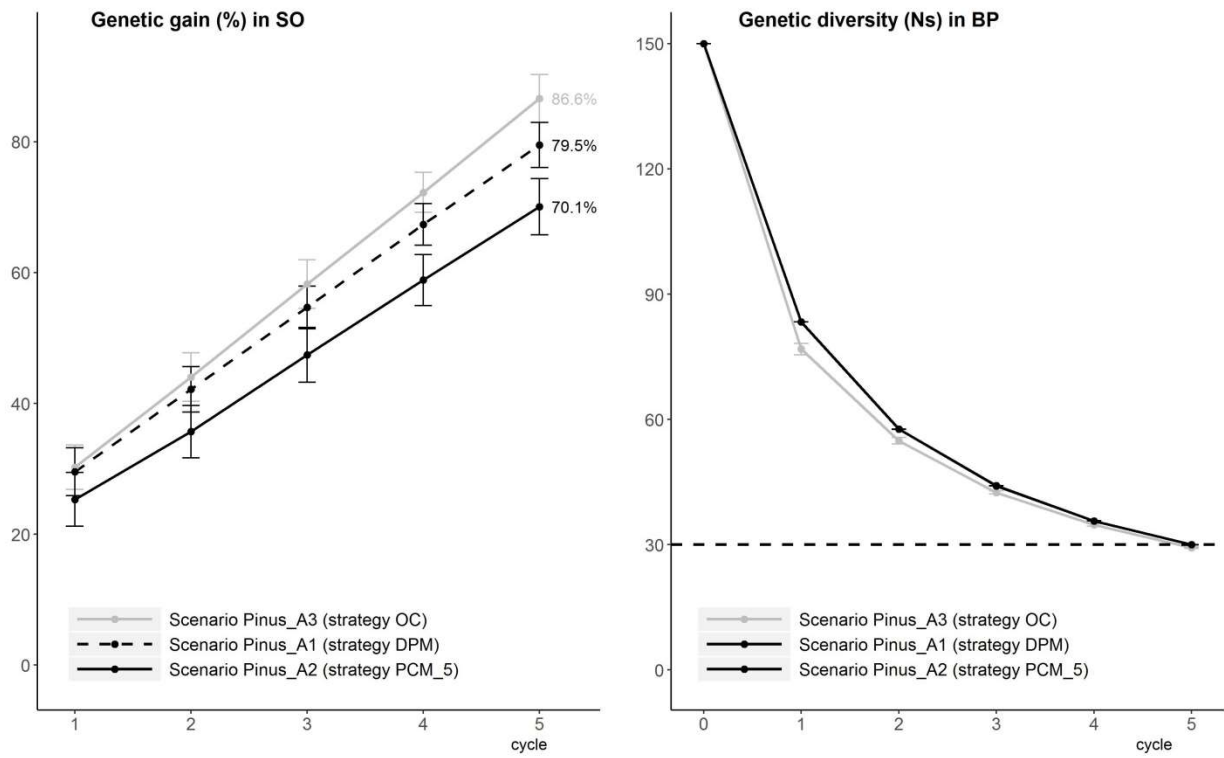
803  
 804  
 805 Note: BP: Breeding Population; RP: Recruitment Population; RP': subset of the Recruitment  
 806 Population considered for paternity testing; SO: Seed Orchard  
 807  
 808

809 **Figure 2: A breeding scenario is defined as a breeding strategy applied with specific**  
810 **diversity constraints**  
811



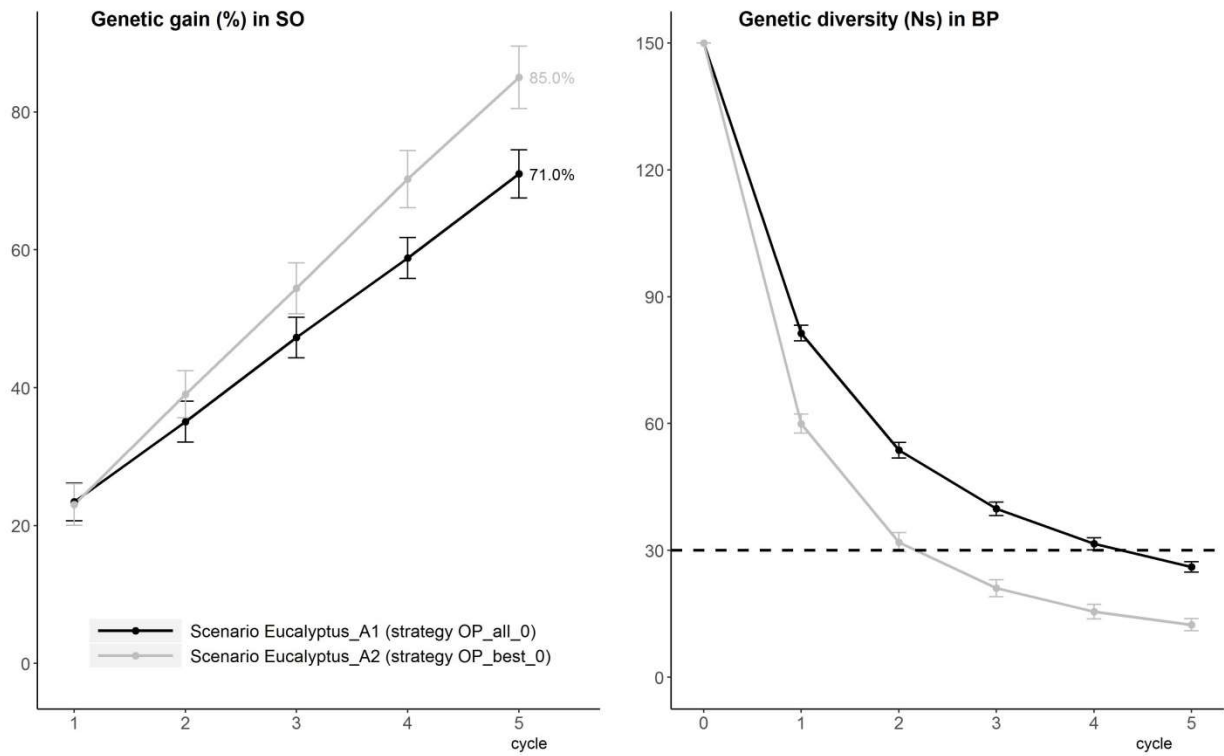
812  
813  
814 Note:  
815 Three main breeding strategies were considered for the *Pinus pinaster* case-study:  
816 DPM: Double-Pair Mating  
817 PCM\_x: PolyCross Mating with a paternity testing factor of x (PCM\_5 for example)  
818 OC: Optimal Contribution  
819 Two main breeding strategies were considered for the *Eucalyptus nitens* case-study:  
820 OP\_all\_x: Open-Pollination with seeds collected on the whole BP and a paternity testing  
821 factor of x (OP\_all\_10 for example)  
822 OP\_best\_x: Open-Pollination with seeds collected on the 50 best trees of the BP and a  
823 paternity testing factor of x (OP\_best\_10 for example)  
824 For most breeding scenarios, diversity constraints were fixed at  $N_s \geq 30$  in BP and  $N_s \geq 10$  in  
825 SO.  
826  
827

828 **Figure 3: *Pinus pinaster* case-study**  
 829 Genetic gain in SO and genetic diversity in BP over 5 breeding cycles (each point is a mean of  
 830 100 iterations and the error bar represents  $\pm 1$  standard deviation).  
 831



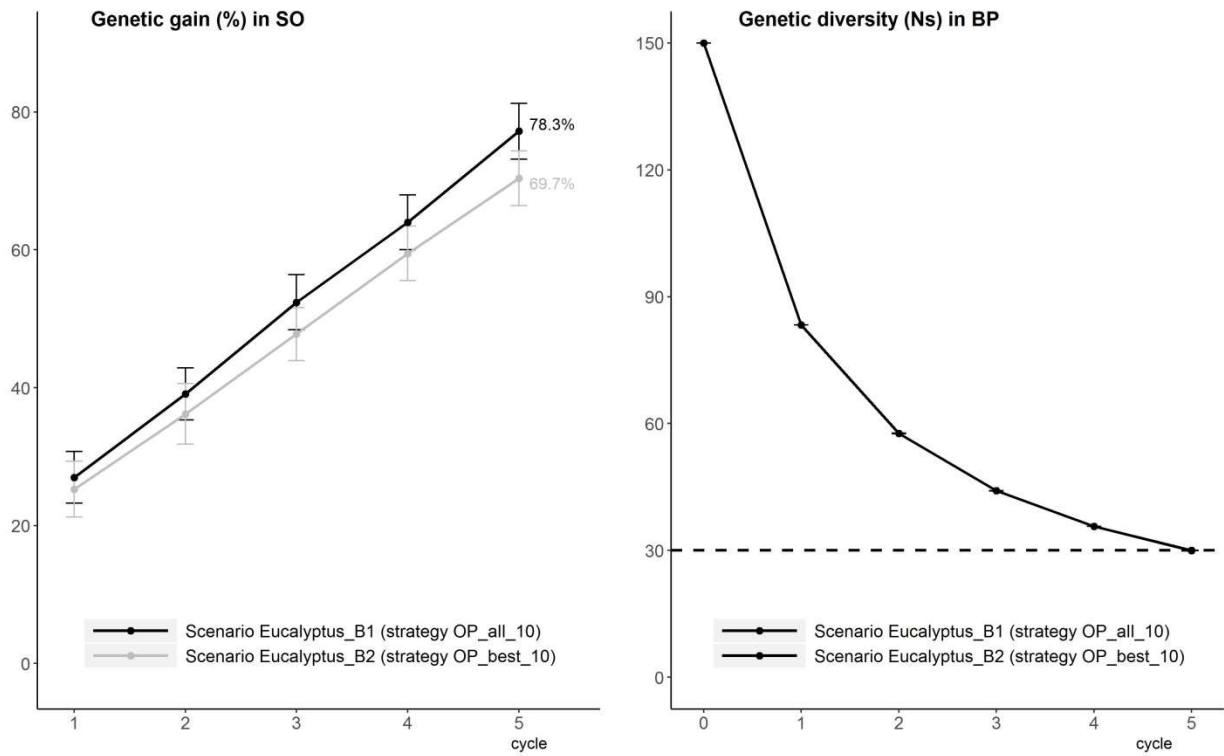
832  
833

834 **Figure 4: *Eucalyptus nitens* case-study – Breeding strategies without paternity testing**  
 835 Genetic gain in SO and genetic diversity in BP over 5 breeding cycles (each point is a mean of  
 836 100 iterations and the error bar represents  $\pm 1$  standard deviation).  
 837



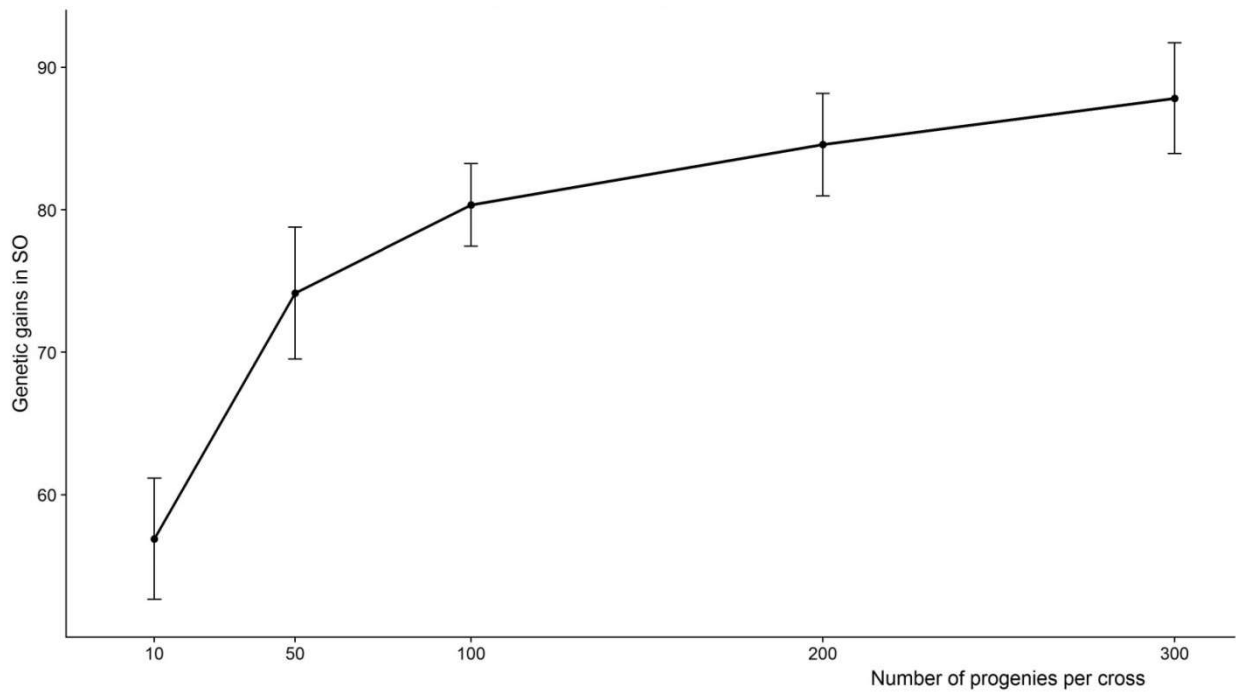
838  
 839  
 840

841 **Figure 5: *Eucalyptus nitens* case-study – Breeding strategies with paternity testing**  
 842 Genetic gain in SO and genetic diversity in BP over 5 breeding cycles (each point is a mean of  
 843 100 iterations and the error bar represents  $\pm 1$  standard deviation).  
 844



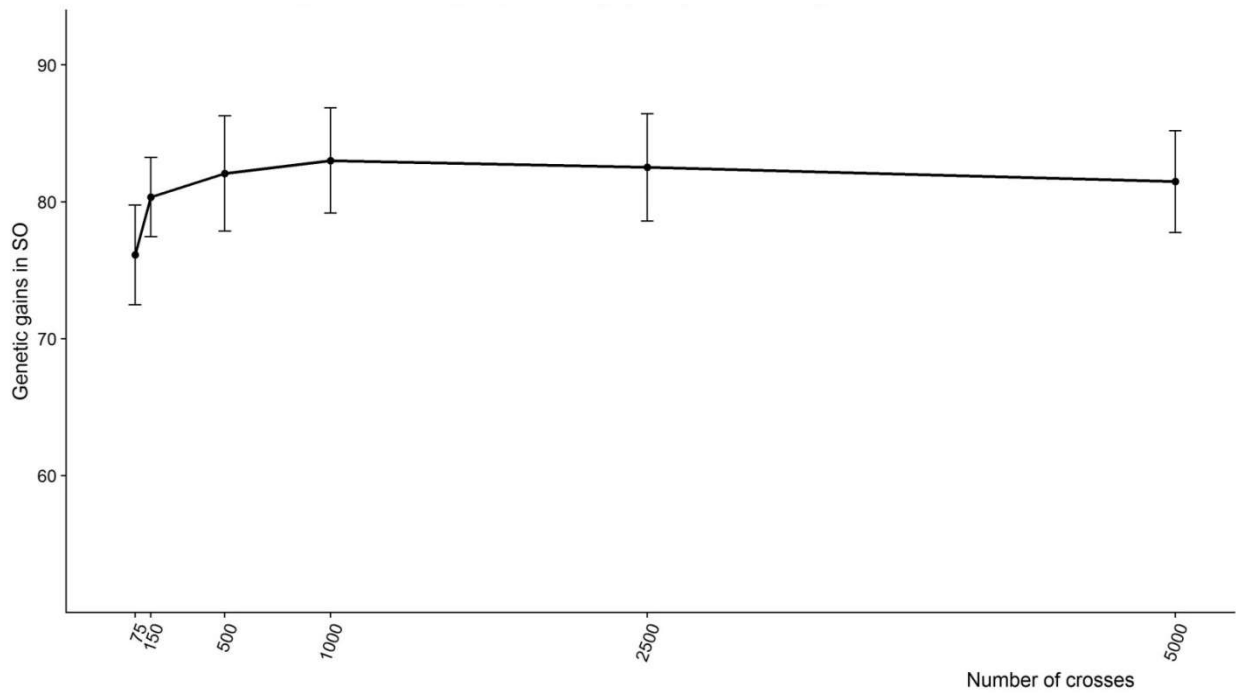
845  
 846  
 847

848 **Figure 6: Genetic gain at breeding cycle 5 for increasing family sizes**  
849 Genetic gain in SO over 5 breeding cycles was simulated for DPM strategy (150 crosses or  
850 families) with 5 levels of family size (each point is a mean of 50 iterations and the error bar  
851 represents  $\pm 1$  standard deviation).  
852



853  
854

855 **Figure 7: Genetic gain at breeding cycle 5 for increasing numbers of crosses**  
856 Genetic gain in SO over 5 cycles was simulated for increasing numbers of bi-parental crosses  
857 under a constant RP size of 15 000 (each point is a mean of 50 iterations and the error bar  
858 represents  $\pm 1$  standard deviation).  
859



860  
861