

Uncovering genetic information from commercial forest plantations—making up for lost time using “Breeding without Breeding”

Milan Lstibůrek¹ · Gary R. Hodge² · Petr Lachout³

Received: 6 December 2014 / Revised: 21 April 2015 / Accepted: 27 April 2015 / Published online: 14 May 2015
© Springer-Verlag Berlin Heidelberg 2015

Abstract An application of “Breeding without Breeding” (BwB) is proposed to uncover or extract genetic information from existing plantations, using pedigree reconstruction and BLUP to predict breeding values and identify genetically superior individuals. The focus is on the use of the methodology at the initiation of an operational breeding program to circumvent the first cycle of breeding and testing, but it could also have application in more advanced tree improvement programs. A simulation study was done to examine different sizes of three conceptual populations used in the BwB approach, and to compare the genetic gains achieved using that approach with those that would have been achieved with a full-sib breeding and testing strategy if it had been started years before. The BwB approach is based on pedigree reconstruction with a relatively small number of trees (from 1,200 to 3,600), comprised of a randomly selected sub-population of size $N_R = 600$ to 3,000, and a top-phenotype sub-population of size $N_T = 600$ (pre-selected out of 5,940 to 23,760 trees on the basis of phenotype alone). With the reconstructed pedigree, a combined REML/BLUP analysis of phenotypic data is done to

predict breeding values, and a linear optimization is done to make the final selections to maximize gain while constraining relatedness to a given effective population size $N_e = 5, 10$, or 20. Results indicate that the BwB strategy can achieve substantial levels of genetic gain, equivalent to 80 to 98 % of the gain that could have been achieved using a full-sib strategy.

Keywords Pedigree reconstruction · BLUP · Tree breeding · Breeding strategy

Introduction

Trees are long-lived organisms, and consequently often take a number of years to reach reproductive maturity. Commercial forestry also involves time spans much longer than are typical for most crop plants or animals: an 8-year rotation length for eucalypts is considered quite fast, 15 to 25 years for temperate and sub-tropical pines is very common, and tree species from northern climates can have rotations of 30 to 80 years, or even longer. Consequently, tree breeding is a long-term endeavor.

Traditionally, tree breeding programs are initiated by making mass selections in natural stands, or for exotic plantation species, in commercial plantations. Sometimes breeders will collect seed directly from the selected trees in the natural stands or plantations, but often selected trees are grafted in seed orchards or breeding arboreta with a subsequent wait of a number of years until seed collection can be done. With that seed, breeders typically establish a number of replicated progeny tests with well-structured mating designs and again wait for some number of years for the trees to reach an age where meaningful data can be obtained (e.g., 4 to 20 years, depending on the species) (White et al. 2007).

For over two decades, geneticists have known of the potential use of molecular marker data to estimate relatedness

Communicated by J. Beaulieu

This article is part of the Topical Collection on *Breeding*

✉ Milan Lstibůrek
lstiburek@fld.czu.cz

¹ Faculty of Forestry and Wood Sciences, Czech University of Life Sciences, Kamýcká 129, 165 21 Praha 6, Czech Republic

² Camcore, Department of Forestry and Environmental Resources, North Carolina State University, 2720 Faucette Drive, Raleigh, NC 27695, USA

³ Faculty of Mathematics and Physics, Charles University in Prague, Sokolovská 83, Praha 8, Czech Republic

(e.g., Queller and Goodnight 1989). Advances in molecular genetics, and in particular the development of highly informative DNA markers such as microsatellites (SSRs) and single-nucleotide polymorphisms (SNPs), have introduced the possibility of new approaches to tree breeding. One of these new approaches is pedigree reconstruction, which is the use of molecular data to determine the genealogical relationships among a population of individuals (e.g., Meagher and Thompson 1986).

Tree breeders have proposed a number of ways that pedigree reconstruction could be used in operational breeding programs. For example, Lambeth et al. (2001) introduced a testing strategy called Polymix Breeding with Parental Analysis. The motivation behind the strategy was that breeding with a pollen mix would be more cost efficient than full-sib breeding, and that markers could subsequently be used to identify male parents of potential selection candidates in order to control relatedness, or to identify the male parents of all the progeny in order to produce a full-sib mating pedigree. El-Kassaby et al. (2007) and El-Kassaby and Lstibůrek (2009) extended the idea to open-pollinated populations, noting that even with natural matings and widely varying family sizes, pedigree reconstruction could be used to estimate genetic parameters and predict breeding values. Their emphasis was on the idea that breeders could make genetic gain without making any crosses, and they called this approach “Breeding without Breeding” (BwB). Hansen and McKinney (2010) have applied this approach to a commercial planting of *Abies nordmanniana* arising from a seed orchard with 99 clones. Using a set of 12 SSR markers, and with genotype information on the parents, they were able to assign parentage (both parents) for 98 % of offspring with an 80 % confidence level. Hansen and McKinney called their approach “quasi-field testing,” to emphasize that breeders can create something like normal progeny test data using the pedigree reconstruction approach.

In the current study, we will discuss an approach to uncover or extract pedigree information from existing plantations and to predict breeding values and identify genetically superior individuals. Specifically, we seek to describe how a BwB approach on standing plantations of a particular species could be done to initiate an operational breeding program. In effect, this approach could be used to circumvent the first cycle of breeding and testing in a forest tree improvement program. The goal is to outline an approach that would achieve most of the genetic gain that would have been made if a full-scale traditional selection and breeding program had been started years before. First, we briefly outline the conceptual idea, and then conduct a thorough quantitative genetic analysis.

Outline of BwB methodology

Assume that we have a species of commercial interest, some plantations of this species, and no genetic tests. Furthermore, assume that the parents of the plantations are unknown, i.e., unavailable for use in breeding or orchards, and unavailable for genotyping. The goal is extract “progeny-test data” from the commercial plantations, and conduct genetic evaluation (prediction of breeding values) for the purposes of selection. The process involves three conceptual populations: the BwB population, and two subsets of that population, a random sub-population and a top-phenotype sub-population. The process is outlined below.

1. Measure and/or inspect the entire BwB population of size N_{BwB} , which will be a subset of the plantation base of the species.
2. Identify a top-phenotype sub-population of size N_{T} , consisting of the best phenotypes in the BwB population. These trees will be the final selection candidates, and provide a high selection differential.
3. Select a random sub-population of size N_{R} from the BwB population. The purpose of the random subpopulation is to provide data for estimates of family means (or equivalently, parental breeding values).
4. Use the molecular marker genotype information for the random and top-phenotype sub-populations for combined full-pedigree reconstruction ($N_{\text{R}} + N_{\text{T}}$).
5. Conduct a REML-BLUP evaluation (utilizing pedigree and phenotypic measurements) to predict breeding values (BVs) for the combined random and top-phenotype sub-populations.

The genetic gains made from the BwB approach will be compared with those that could be obtained from a large-scale full-sib breeding and testing strategy. We do not intend to suggest that the particular full-sib testing strategy described below is optimum in terms of efficiency; rather, we simply want to compare the BwB strategy with a large and comprehensive full-sib testing effort that should result in good genetic gain.

The specific objectives of this study are to:

1. Examine the genetic parameter estimates and accuracy of BV predictions from the BwB approach.
2. Determine the expected genetic gains from the BwB approach, and compare them to gains from a traditional testing strategy (full-sib mating, replicated progeny tests, etc.).
3. Examine different sizes of the three populations used in the BwB strategy (N_{BwB} , N_{R} , and N_{T}), in order to suggest near-optimum sizes for operational tree improvement programs.

Materials and methods

Simulation model

A stochastic simulation model was developed in the R system (R Core Team 2014), featuring 1 main routine and 10 subroutines. Two external software libraries were incorporated and invoked within the R code: (1) ASReml for R (Butler et al. 2007), and (2) Gurobi Optimizer (Gurobi Development Team 2014). The simulation experiment was repeated using 500 independent iterations for every scenario and statistical inference was drawn across these.

Data generation

Additive polygenetic effects of $N_P=99$ unrelated and non-inbred parental trees were sampled from the normal distribution $N(0, \sigma_A^2)$, assuming the infinitesimal genetic model. Correspondingly, additive by environmental effects were sampled from $N(0, \sigma_{AE}^2)$ for all experimental test sites considered throughout the study. Actual input parameters are provided in Table 1.

Genetic variance components are based on typical values for volume growth for four pine species (Hodge and Dvorak 2012). Two hypothetical breeding strategies were modeled: a full-sib mating design (FS) tested on six sites, and a Breeding without breeding (BwB) strategy applied across two sites. The choice of six sites for the FS strategy was somewhat arbitrary, but the idea was to sample a sufficient number of sites to give a good estimate of genotype $\times$ environment interaction and to make accurate predictions of parental breeding values on an across-site basis. The choice of two sites for the BwB strategy was because two is the minimum number of sites that allows an estimate of genotype $\times$ environment interaction. For both strategies, offspring genotypic and phenotypic values were generated as follows.

First, the polygenic additive genetic value was drawn from $N(\bar{a}, 0.5\sigma_A^2)$, where $\bar{a}$ is the respective mid-parental additive genetic value. Second, the polygenic additive genetic $\times$

environment interaction value was drawn from $N(\bar{ae}, 0.5\sigma_{AE}^2)$, where $\bar{ae}$ is the respective mid-parental additive genetic $\times$ environment value for the specific environment. Third, the expected dominance genetic value for a given parental combination was drawn from $N(0, 0.25\sigma_D^2)$, and correspondingly, the remaining within-family dominance deviate was sampled from $N(0, 0.75\sigma_D^2)$. Finally, the environmental deviate was drawn from $N(0, \sigma_E^2)$.

Full-sib strategy (FS)

Controlled crosses among all parental trees were created following a circular mating design (Huber et al. 1992), with each tree crossed with four other parents; correspondingly, the total number of crosses was equal to 198. Equal family size (120) was assumed across all combinations. In total, $N_{FS}=23,760$ offspring genotypes were generated and evenly distributed into six independent test locations (3,960 progenies per location).

Breeding without breeding strategy (BwB)

Open-pollinated mating was assumed, which would potentially generate a complete (but unbalanced) half-diallel mating among all parental trees. All parents were randomly assigned a gametic contribution probability from a theoretical distribution. The distribution was based on data in the literature regarding family size variation and gametic contribution variation in open-pollinated seed orchards, which typically find that there are a few genotypes that are very fecund, and many that contribute relatively little to the gametic population (El-Kassaby and Cook 1994; Funda et al. 2011; Lai et al. 2010). Briefly, with the distribution assumed in this study, one-half of the population produces only 33 % of the gametes, and the more fecund half of the population produces 67 % of the gametes; furthermore, the most fecund 10 % of the parents produce 25 % of the gametes (Fig. 1). For each possible combination of parents, family size was determined by multiplying the two parental gametic contribution probabilities, then multiplying by the total experimental test size N_{BwB} , and then rounding to the nearest integer value (the actual family size). Three values of N_{BwB} were assumed in different scenarios: 5,940 = $0.25 \times N_{FS}$, 11,880 = $0.5 \times N_{FS}$, and 23,760 = $1 \times N_{FS}$; in other words, the number of progeny equivalent to $\frac{1}{4}$, $\frac{1}{2}$, and the full number of progeny used in the full-sib strategy. Offspring genotypes within each parental combination were randomly allocated into two independent test locations.

For each run of each scenario, a random sub-population (size = N_R) was hypothetically fingerprinted. Next, a phenotypic truncation was performed to identify the top-phenotype sub-population, with size (N_T). Pedigree information of the two sub-populations was revealed, assuming hypothetical

Table 1 Simulation input parameters

Parameter	Value
σ_A^2	375
σ_D^2	75
σ_{AE}^2	200
σ_E^2	1,850
N_{FS}	23,760
N_{BwB}	5,940; 11,880; 23,760
N_R	0; 600; 1,200; 1,800; 2,400; 3,000
N_T	600
N_e	5; 10; 20

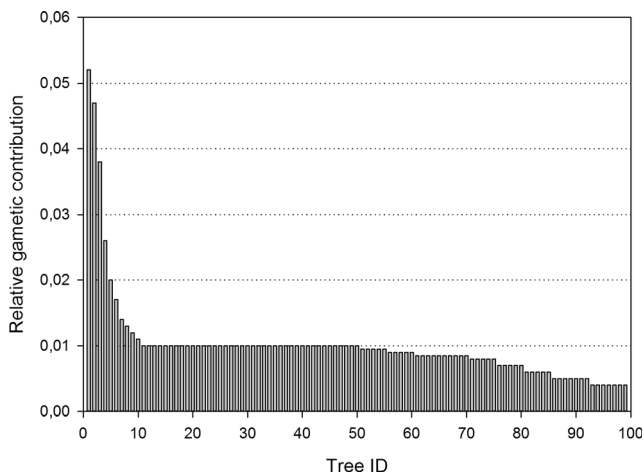


Fig. 1 Distribution of parental gametic contributions [in percent] for the respective 99 parents

pedigree reconstruction with 100 % accuracy. The size of the random sub-population N_R was varied from 600 to 3,000 in individual simulation scenarios. Preliminary investigation revealed that $N_T=600$ was sufficient to meet the prescribed relatedness restrictions (described below) for even the largest size of the final selected population ($N_e=20$), so N_T was restricted to 600 individuals for all scenarios. Under real circumstances, it might be possible to meet the declared N_e with even lower N_T , that is, with less genotyping effort, as suggested by Lstibůrek et al. (2011).

Phenotypic strategy (PH)

In addition to the FS and BwB strategies, scenarios with $N_R=0$ were examined where selection of the $N_T=600$ individuals was based on phenotypic evaluation alone. This was done to examine the genetic gain that could be achieved based on phenotypic selection with pedigree control, following the approach suggested by El-Kassaby and Lstibůrek (2009).

Genetic evaluation

Standard full-sib genetic evaluation was conducted under the FS strategy, assuming parents and offspring individuals were included into the combined REML-BLUP analyses within ASReml featuring the animal genetic evaluation model:

$$y_{ijkl} = a_{ijkl} + sca_{jk} + fe_{jl} + fe_{kl} + e_{ijkl} \quad (1)$$

where a_{ijkl} is the random additive genetic effect of the i th tree in family jk in the l th environment, sca_{jk} is the random dominance effect of family jk (i.e., the cross of the j th parent and the k th parent), fe_{jl} and fe_{kl} are the random additive $\times$ environment effects for the j th parent in the l th environment and the k th parent in the l th environment, respectively, and e_{ijkl} is the random error term.

Following the hypothetical pedigree reconstruction, genetic evaluation of the BwB data was performed using the same animal genetic model as used for the FS evaluation (Eq. 1), but with the sca effect removed. Preliminary analyses showed that the BwB approach had a tendency to dramatically overestimate dominance variance (from $2\times$ to $10\times$), due to very small number of full-sibs, so this term was dropped from the model. The BwB evaluation was done using ASReml, including parents and all offspring individuals from the random and top-phenotype sub-populations in a combined REML-BLUP analysis.

For both FS and BwB genetic evaluation, and for each run of each scenario, the estimated genetic parameters from the REML analysis were used in the BLUP of genetic values, exactly as would be the case under real circumstances.

Selection

In this final step, the goal was to select the best set of offspring trees, maximizing the genetic response to selection, while meeting the declared effective population size ($N_e=5$, 10, or 20). It was assumed that the census number of the selected set was equivalent to the effective population size, meaning that no relatedness among the selected trees was permitted. The objective function was

$$\sum_{i=1}^N \hat{a}_i x_i \rightarrow \max \quad (2)$$

where $N=N_{FS}$ (FS scenario), $N=N_R+N_T$ (BwB scenario), or $N=N_T$ (PH scenario), the $\hat{a}_i$ values are the predicted additive genetic (breeding) values (from the ASReml genetic evaluation) and the x_i values are selection pointers (binary variables: 1=tree selected, 0=tree not selected). The optimization utilized the following linear constraint to limit the number of selected individuals to the declared N_e (integer value):

$$\sum_{i=1}^N x_i = N_e \quad (3)$$

To ensure that the selection outcome consists of unrelated individuals, the following constraints were added to the optimization (one for a given parental tree, if applicable):

$$\begin{array}{ccccccc} y_{11}x_1 & + & y_{12}x_2 & + & \cdots & + & y_{1N}x_N & \leq & 1 \\ y_{21}x_1 & + & y_{22}x_2 & + & \cdots & + & y_{2N}x_N & \leq & 1 \\ \vdots & & \vdots & & & & \vdots & & \vdots \\ y_{Np1}x_1 & + & y_{Np2}x_2 & + & \cdots & + & y_{NpN}x_N & \leq & 1 \end{array} \quad (4)$$

where y_{ij} values are binary pointers linking a given parental tree to its respective offspring trees, thus a maximum parental contribution is one offspring per parent.

For each run of a given scenario (N_R, N_T), a set of parents and offspring for FS and BwB strategies were simulated, and all variance components estimated. For the offspring selected

for a given N_e , the true genetic gain for both the FS and BwB strategies (R_{FS} and R_{BwB} , respectively) were calculated as the mean of the true breeding values. Finally, an efficiency parameter $Q_{BwB/FS}$ was calculated as the ratio of the mean gains from the BwB and FS strategies, $Q_{BwB/FS} = \frac{R_{BwB}}{R_{FS}} \times 100\%$.

Results

Genetic parameter estimates and accuracy of breeding value prediction

For the FS strategy, mean genetic parameter estimates across 500 iterations were very accurate. For example, the mean additive variance estimate was $\hat{\sigma}_A^2 = 377.4$, with a target value of $\sigma_A^2 = 375$. The standard deviation for the 500 independent estimates of additive variance was 66.0. Similarly, the mean estimate of the additive $\times$ environment variance was $\hat{\sigma}_{AE}^2 = 200.0$, with a target value of $\sigma_{AE}^2 = 200$, and the standard deviation of the 500 additive $\times$ environment variance estimates was 5.1. Since the estimates from each iteration were based on a population of 23,760 trees in a balanced mating design, one would expect the estimates to be both accurate and precise, so these results are not surprising.

For the BwB scenarios, the number of trees used to estimate variance components was the sum of the number of trees in the random and top phenotype sub-populations ($N_R + N_T$), with $N_T = 600$ for all scenarios, and N_R ranging from 600 to 3,000. Thus, the number of genotyped trees used in the BwB strategies for variance component estimation and BLUP prediction ranged from 1,200 to 3,600. For all scenarios, the mean genetic parameter estimates were very close to the target expected values (Table 2). As might be expected, however, the variation of the parameter estimates was quite high. For example, for the minimum effort BwB strategy, Scenario 1 with $N_{BwB} = 5,940$, $N_R = 600$, and $N_T = 600$, the mean estimate for additive variance was $\hat{\sigma}_A^2 = 380.5$ (with a true value of $\sigma_A^2 = 375$), but the standard deviation of the 500 independent estimates was 208.6, indicating a very large range in additive variance estimates (Table 2). Increasing the size of the random sub-population improves the precision of the variance component estimation. For example, for $N_{BwB} = 5,940$ and $N_R = 1,200$, the standard deviation of the 500 independent estimates of $\hat{\sigma}_A^2$ was much lower at 138.5, and for $N_{BwB} = 5,940$ and $N_R = 3,000$, the standard deviation of the 500 independent estimates of $\hat{\sigma}_A^2$ was 95.5 (Table 2). For larger BwB population sizes, increasing N_R gave similar improvements in the precision of the variance component estimates (Table 2).

Despite the fact that the variance component estimates from the FS strategy were much more precise than those from the BwB strategy, the accuracies of the individual tree

Table 2 Mean variance component estimates across 500 simulation iterations, the standard deviations of the 500 estimates, and the accuracy of breeding value predictions for different Breeding without Breeding (BwB), phenotypic (PH), and the full-sib testing scenarios (FS)

Scenario	N	N_R	$\hat{\sigma}_A^2 \pm \text{sd}$	$\hat{\sigma}_{AE}^2 \pm \text{sd}$	$r_{aa} \pm \text{sd}$
BwB 1	5,940	600	380.5 $\pm$ 208.6	203.4 $\pm$ 48.7	0.59 $\pm$ 0.05
BwB 2	5,940	1,200	381.6 $\pm$ 138.5	194.1 $\pm$ 30.1	0.61 $\pm$ 0.04
BwB 3	5,940	1,800	375.1 $\pm$ 115.3	196.7 $\pm$ 25.1	0.62 $\pm$ 0.04
BwB 4	5,940	2,400	374.6 $\pm$ 101.3	200.6 $\pm$ 20.7	0.62 $\pm$ 0.04
BwB 5	5,940	3,000	373.9 $\pm$ 95.5	200.8 $\pm$ 17.8	0.62 $\pm$ 0.04
BwB 6	11,880	600	363.3 $\pm$ 214.0	214.7 $\pm$ 51.7	0.61 $\pm$ 0.05
BwB 7	11,880	1,200	383.1 $\pm$ 147.7	201.4 $\pm$ 34.6	0.63 $\pm$ 0.05
BwB 8	11,880	1,800	378.1 $\pm$ 122.1	201.3 $\pm$ 25.6	0.64 $\pm$ 0.04
BwB 9	11,880	2,400	377.4 $\pm$ 107.5	199.4 $\pm$ 21.1	0.64 $\pm$ 0.04
BwB 10	11,880	3,000	376.6 $\pm$ 99.6	199.7 $\pm$ 18.4	0.64 $\pm$ 0.04
BwB 11	23,760	600	368.0 $\pm$ 208.2	222.8 $\pm$ 53.0	0.64 $\pm$ 0.05
BwB 12	23,760	1,200	384.6 $\pm$ 143.5	192.5 $\pm$ 32.4	0.65 $\pm$ 0.04
BwB 13	23,760	1,800	381.6 $\pm$ 116.0	193.0 $\pm$ 25.0	0.65 $\pm$ 0.04
BwB 14	23,760	2,400	381.8 $\pm$ 106.5	196.3 $\pm$ 20.6	0.65 $\pm$ 0.04
BwB 15	23,760	3,000	380.9 $\pm$ 97.3	197.0 $\pm$ 18.4	0.65 $\pm$ 0.04
PH 1	5,940	0	NA	NA	0.39 $\pm$ 0.04
PH 2	11,880	0	NA	NA	0.39 $\pm$ 0.04
PH 3	23,760	0	NA	NA	0.39 $\pm$ 0.04
FS	23,760	0	377.4 $\pm$ 66.0	200.0 $\pm$ 5.1	0.68 $\pm$ 0.03
Target			375.0	200.0	

N = total number of trees in the respective population [i.e., $N = N_{BwB}$ in BwB scenarios, $N = N_{PH}$ in PH scenarios, $N = N_{FS}$ in FS scenario, N_R = number of trees in the random sub-population, $\hat{\sigma}_A^2$ = estimate of additive variance, $\hat{\sigma}_{AE}^2$ = estimate of additive $\times$ environment variance, r_{aa} = accuracy of prediction (correlation of the true and predicted breeding values)]

breeding value predictions were quite similar for the FS and BwB strategies. Accuracy of the predicted breeding values is measured as the correlation between the true and the predicted breeding values (r_{aa}) for the 23,760 progeny in the FS strategy, and the total number of progeny in the random and top-phenotype sub-populations ($N_R + N_T$). For phenotypic selection (PH strategy), the accuracy of selection was $r_{aa} = 0.39$, while for the FS strategy, the accuracy of selection was $r_{aa} = 0.68$ (Table 2). For all BwB scenarios, r_{aa} is lower than for the FS strategy than for the BwB strategy, but only slightly (Table 2). For example, for scenario 8 with $N_{BwB} = 11,880$ and $N_R = 1,800$, $r_{aa} = 0.64$, compared with 0.68 for the FS strategy. The fact that r_{aa} for the BwB and FS strategies are generally very similar suggests that both strategies would produce similar genetic rankings.

Although the accuracies of the BwB and the FS strategies were similar, there was a slight tendency for the predicted genetic gains from the BwB approach to overestimate the true genetic gains. For example, for scenario 7 with $N_{BwB} = 11,880$ and $N_R = 1,200$, and for selection of $N_e = 5$, the average true

genetic gain was 34.0 % (see Fig. 2), while the average predicted genetic gain was 39.1 %. For the same scenario, for selection of $N_e=20$, the average true genetic gain was 23.9 % (see Fig. 3), while the average predicted genetic gain was 26.0 %. In contrast, the FS average predicted genetic gains were almost exactly equal to the FS average true genetic gains for all sizes of N_e . The variance component estimates from the FS strategy are much more precise than the estimates from the BwB strategy (Table 2), and it may be that large variance in the variance component estimates for the BwB strategy contribute to this tendency for the BwB predicted gains to be too high. It is important to note, however, that this has no impact on the true genetic gains achieved using the BwB approach; selection based on the predicted genetic values $\hat{a}$ will result in gain in the true genetic values a even if the predicted gain is overestimated. However, breeders should be aware of this tendency, and use the predicted genetic values and gains from the BwB BLUP analyses with some caution.

Genetic gains

Figures 2 and 3 present the relative efficiency of the BwB and the FS strategies, and the true genetic gain for different sizes of the BwB population (N_{BwB}) and the random sub-population (N_R). The two figures are interpreted identically, but present results for two different sizes of selected populations, $N_e=5$ or 20 for Figs. 2 and 3, respectively. The different N_e represent different selection intensities that might apply in operational breeding programs. Gains for selected population size $N_e=10$ are not shown, but follow a similar pattern and are very nearly the average of gains for $N_e=5$ and $N_e=20$. Average genetic responses to selection for BwB strategies are plotted both as genetic gains ($\bar{R}_{BwB}$, right axis), and as the relative efficiency compared to the gain from the FS strategy ($Q_{BwB/FS}$, left axis).

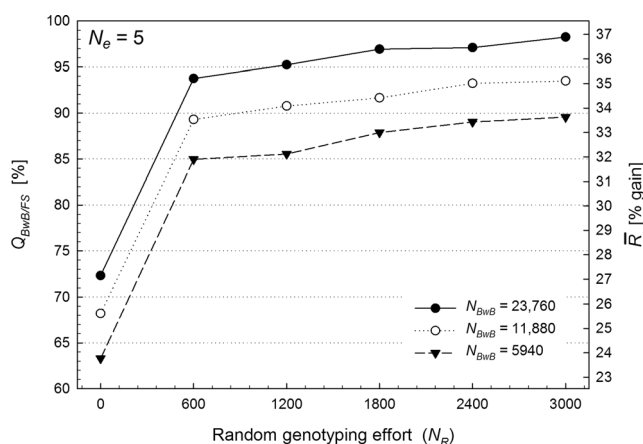


Fig. 2 Genetic gains from the proposed BwB strategy for a selected population size of $N_e=5$ for different sizes of the BwB population (N_{BwB}) and the random sub-population (N_R). The left axis indicates the relative efficiency compared to the gain from the FS strategy ($Q_{BwB/FS}$), and the right axis indicates absolute genetic gains ($\bar{R}_{BwB}$)

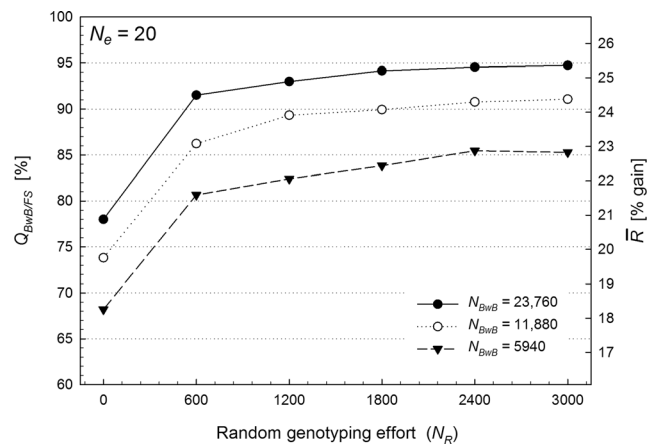


Fig. 3 Genetic gains from the proposed BwB strategy for a selected population size of $N_e=20$ for different sizes of the BwB population (N_{BwB}) and the random sub-population (N_R). The left axis indicates the relative efficiency compared to the gain from the FS strategy ($Q_{BwB/FS}$), and the right axis indicates absolute genetic gains ($\bar{R}_{BwB}$)

A t test was done to compare the difference of the FS gain and the BwB gain, ($\bar{R}_{FS} - \bar{R}_{BwB}$), and confidence intervals calculated for the ratio of the gains. For all scenarios, BwB gain was lower than FS gain, and all differences were statistically significant. For all scenarios with the smaller BwB populations ($N_{BwB}=5,940$, $N_{BwB}=11,880$), all differences between BwB gain and FS gain were significant at $P \leq 0.0001$. For the largest BwB population ($N_{BwB}=23,760 = N_{FS}$), all differences were significant at $P \leq 0.05$, and most were significant at $P \leq 0.0001$. It is clear that the genetic gain from the BwB strategy is lower than what could be achieved from the FS strategy, but this result is not surprising.

More interesting is the large amount of genetic gain that can be captured with the BwB strategy, which ranges between 80 and 98 % of the gain achieved by the FS strategy, depending on the exact scenario (depending on the size of the BwB population, the random sub-population, and the final selected population, N_{BwB} , N_R , N_e , respectively). The 95 % confidence intervals around $Q_{BwB/FS}$ (the ratio of BwB gain and FS gain) were very near to ± 2.0 % for all scenarios.

For the minimum effort BwB strategy examined in this study (scenario 1, with $N_{BwB}=5,940$, $N_R=600$), the ratio of BwB gain to FS gain was 85.1 and 81.4 % for selection of $N_e=5$ and 20 unrelated offspring, respectively (Figs. 2 and 3). For the maximum effort BwB strategy examined in this study (scenario 15, with $N_{BwB}=23,760$, $N_R=3,000$), the ratio of BwB gain to FS gain was 97.9 and 95.0 % for selection of $N_e=5$ and 20 unrelated offspring, respectively (Figs. 2 and 3).

The inclusion of a random sub-population is important to the BwB strategy. It is possible to make genetic gain using only phenotypic selection with pedigree control (PH strategy), but it is clear that there is a significant incremental gain from having a random sub-population (N_R) of at least 600 trees. For $N_e=5$, selection based only on phenotype produced from 23.8

to 27.1 % genetic gain ($Q_{\text{BwB/FS}}=63.3$ to 72.3 %), depending on N_{BwB} (see the points with $N_{\text{R}}=0$ in Fig. 2). The use of a genotyped random sub-population of size $N_{\text{R}}=600$ increases genetic gain by 8.0 to 9.5 % (equivalent to increasing $Q_{\text{BwB/FS}}$ by 21.1 to 21.8 %).

Increasing the size of the random sub-population (N_{R}) from 600 to 3,000 produces a small amount of incremental genetic gain. For example, for $N_{\text{R}}=600$, $N_{\text{BwB}}=11,880$, and $N_{\text{e}}=5$, the BwB genetic gain was 33.5 %, equivalent to $Q_{\text{BwB/FS}}=89.3$ % of the FS gain (Fig. 2). Increasing $N_{\text{R}}=3,000$ gives a BwB genetic gain of 35.0 %, equivalent to $Q_{\text{BwB/FS}}=93.5$ % of the FS gain (Fig. 2). Across all scenarios, increasing N_{R} from 600 to 3,000 produced about 1 to 2 % additional genetic gain from the BwB strategy, equivalent to an increase of 3 to 5 % in $Q_{\text{BwB/FS}}$.

Increasing the size of the BwB population also produces incremental genetic gain. This can be seen in Figs. 2 and 3 by comparing the heights of the three lines for the different sizes of N_{BwB} . For any given size of N_{R} and N_{e} , increasing N_{BwB} from 5,940 to 11,880, or increasing N_{BwB} from 11,880 to 23,760 produces about 1 to 2 % additional genetic gain.

Discussion

Genetic gains

The application of the BwB strategy in commercial plantations offers the opportunity to make substantial genetic gains and, perhaps more importantly, save years of time in testing. The reconstruction of the pedigree of 1,800 trees (1,200 random + 600 top phenotypes selected from a larger population) can return approximately 85 to 95 % of the genetic gain that would have been achieved by a large-scale full-sib testing program. This 85 to 95 % efficiency applies over the range of selection intensity, from $N_{\text{e}}=5$ to 20. A target of $N_{\text{e}}=20$ might be what the breeder would use if there was only one genetic source of plantation material, and the breeder wanted to construct a seed orchard. Or if there were multiple known genetic sources of the plantations, the breeder might want to construct a seed orchard using material from all of them, and a target of $N_{\text{e}}=5$ for each source might be appropriate. Alternately, the breeder might want to use a target of $N_{\text{e}}=5$ to select parents for a small full-sib production population (or a clonal production population).

The fact that the BwB approach is very efficient and comparable to the FS strategy in terms of genetic gain might initially be somewhat surprising. In this study, the focus was on progeny selection at the initiation of a tree improvement program, with parental genotypes unavailable for selection. Part of the genetic gain from progeny selection arises from identifying the best families, or in other words, accurately predicting the breeding values of the best parents (even though the

parents themselves are unknown or unavailable). The parental breeding values determine the additive genetic value of the full-sib family (i.e., the mid-parent BV). The other part of the genetic gain from progeny selection comes from identifying outstanding progeny from top families. Thus, for any individual progeny, predicted genetic gain can be thought of as the sum of the mid-parent predicted breeding values plus the within-family predicted genetic gain. The underlying genetic variance of both the mid-parent values and the within-family genetic values is equivalent to $\frac{1}{2}$ of the additive genetic variance.

The mid-parent genetic value prediction is based on observations of many progeny of the two parents. But the within-family genetic deviations are determined (conceptually) by comparing the phenotype of an individual progeny to the rest of its siblings. In other words, a substantial portion of the predicted genetic value for a specific progeny is determined by a single phenotypic observation (adjusted for environmental error effects). A single phenotypic observation for a low heritability trait is inherently an imprecise measurement of the underlying genetic value (in this case, the within-family genetic value), and thus the overall accuracy of selection $r_{\hat{a},a}$ will be limited. For phenotypic selection or mass selection based on a single trait, the accuracy of prediction is theoretically equal to the square root of the heritability (Falconer 1981; Hodge and White 1992). In this study, with $h^2=0.15$, the accuracy of the PH strategies should therefore be $r_{\hat{a},a}=0.39$, which exactly matches the average accuracies calculated for the PH strategies in the 500 simulation runs in this study (Table 2). With both the FS strategy and the BwB strategies, the BLUP analyses will simultaneously predict mid-parent genetic values and within-family genetic deviations to predict the genetic values for each offspring, and accuracies will be above this baseline of $r_{\hat{a},a}=0.39$. One can derive the theoretical maximum accuracy for offspring selection for the FS and BwB strategies, assuming infinite progeny testing, for a given heritability (see Appendix). In this study, with $h^2=0.15$, the maximum accuracy possible for offspring selection is $r_{\hat{a},a}=0.74$.

For the FS strategy, the average accuracy was $r_{\hat{a},a}=0.68$ (Table 2), near the maximum possible with the genetic parameters assumed in this study; this is a moderately high correlation of the true and predicted genetic values for progeny. The BwB strategies approached this accuracy (Table 2), which suggests that with the FS strategy the breeder would have invested more resources than necessary in predicting the parental genetic values. Furthermore, the fact that the BwB and FS accuracies are similar implies that the genetic gains from selection using the two strategies would also be similar, as they were (Figs. 2 and 3). It is important to note that one might find a very different result if the comparison between the BwB strategy and the FS strategy was made on the basis of parental genetic value predictions. This comparison would be relevant

if the parental genotypes were available to the breeder for selection; in that case, the FS strategy would likely have a substantial advantage.

If substantial genetic gains can be made with the suggested BwB strategy, this would be of particular value to circumvent a testing cycle at the initiation of a tree improvement program. Regardless of whether the BwB strategy is used or not, a breeding program would begin with inspection and/or mass selection in natural stands or plantations, either to identify the top-phenotype population for the BwB, or to identify a population of plus-trees for subsequent progeny testing. The BwB approach would accurately identify outstanding genotypes without grafting, seed collection, or progeny testing, potentially saving many of years of work. The best genotypes would then be immediately available to the breeder for orchard establishment or deployment to plantations.

Depending on species biology, there might be a substantial time lag to seed production following selection and grafting. In this case, a breeder could choose to initiate a tree improvement program by establishing a seed orchard with a population of plus-trees, and progeny testing those selections using OP seed collected from the plus-trees in the plantations. By the time the orchard begins to produce seed, the progeny tests might be old enough to provide data to rank parents in the orchard. The expected genetic gain from this approach would depend on a number of factors, including the number of initial selections grafted into the orchard and progeny tested, the number and design of the progeny tests, and whether or not the breeder chose to use pedigree reconstruction to control relatedness in the orchard. The gain from the initial selection would probably be approximately what we have estimated for phenotypic selection alone (Figs. 2 and 3), with a slight marginal gain after roguing the seed orchard (e.g., Lindgren and El-Kassaby 1989; Prescher et al. 2008). In other words, in this scenario, the BwB approach might not save time, but would make additional genetic gain. In addition, the breeder could focus accelerated breeding efforts, such as top grafting (Hartman and Kestler 1968) or flower stimulation treatments (e.g., Pharis and Kuo 1977) on only the best genotypes at the very beginning of the program.

Pedigree reconstruction accuracy

The potential genetic gains discussed above assume 100 % accuracy in pedigree reconstruction. Pedigree reconstruction can be done with SNPs or SSRs, but with both technologies, there will invariably be some errors in the genotyping process. For example, some individuals may have missing data for some or all loci, and some individuals with data cannot be assigned to parents with a high degree of confidence. However, with improving technology and sufficient numbers of markers, rather high accuracies can be achieved.

As mentioned earlier, Hansen and McKinney (2010) were able to use 12 SSR markers to assign both parents to 98 % of the offspring in an *A. nordmanniana* plantations. They note, however, that pollen contamination is extremely low (around 3 %) in the orchard where this seed was collected. Higher levels of pollen contamination are common among forest tree seed orchards (El-Kassaby et al. 1989; Adams et al. 1997; Slavov et al. 2005), and this can make pedigree reconstruction more challenging. For example, in a simulation study based on *P. sylvestris* data, Wang et al. (2010) were able to assign both parents to 97 % of offspring from mating among orchard clones, but for matings involving foreign (non-orchard) pollen, their accuracy to assign even the maternal parent was significantly lower, with around 78 % of offspring correctly assigned. This study was done with 9 SSR loci with an average of 9.3 alleles. For a population of *Pseudotsuga menziesii* arising from a seed orchard of 59 clones with a contamination rate of 40 %, El-Kassaby et al. (2007) used nine SSR markers to assign both parents with 43 to 58 % success. Given that there may be less than 100 % success in assigning parents, the sizes for the random and top-phenotype sub-populations discussed in this study could be considered as minimums. For example, a breeder might choose to genotype a larger number of trees, but then only use data for which parentage or sibship could be assigned with a high level of confidence.

In terms of the BwB strategy proposed in this study, the primary effect of pollen contamination would be on the random sub-population. A theoretical study by Lstibůrek et al. (2012) demonstrated that phenotypic pre-selection among bulk progeny from a seed orchard would significantly reduce the frequency of offspring resulting from pollen contamination, dependent on a number of factors, including genetic superiority of the orchard, the actual level of pollen contamination, the heritability of the trait, and the intensity of phenotypic pre-selection. This was followed by an applied study on a *P. sylvestris* first generation seed orchard with reported pollen contamination of 21 to 70 %. The top 10 % of the phenotypes of a bulk progeny population from the orchard were selected, and subsequent genotyping revealed that the pre-selected population had less than 5 % pollen contamination (or in other words, less than 5 % of those top phenotypes had a parent from outside the orchard) (Korecký et al. 2014). Relevant to the BwB strategy, these results suggest that the Top Phenotype sub-population would probably be relatively free of pollen contamination; however, the unselected random sub-population could have high levels of pollen contamination. For a random sub-population of size $N_R = 1,800$, a number of those progeny would be offspring from a parent outside the seed orchard. Data on these offspring would thus provide less information on the parents in the orchard, and therefore on the progeny of interest, that is, those of higher genetic quality that would be candidates for the final selected population. How much this would impact the genetic gains made by the BwB

strategy can be approximated by moving leftward (toward lower N_R) along the lines in Figs. 2–3, for example, by reducing N_R from 1,800 to 1,200. In this example, even the loss of 1/3 of the data from the random sub-population would reduce genetic gain by less than 1 %, so it seems clear that the effect of pollen contamination in the random sub-population would reduce overall genetic gain only by a small amount.

Most of the pedigree reconstruction and BwB literature (whether simulations or actual applications) has been based on known parental genotypes, which makes the job of pedigree reconstruction easier. If parental genotypes are unknown, pedigree reconstruction becomes more difficult, but still theoretically possible. For forest trees, Massah et al. (2010) have studied pedigree reconstruction in populations of yellow cedar (*Callitropsis nootkatensis*) with unknown parents, and were able to identify population and family relationships. Grattapaglia et al. (2014) used 10 SSR markers with an average of 13.2 alleles per locus to confirm parentage of mass control pollination families in a clonal seed orchard of *Pinus taeda*. They reported high probabilities of identity and parentage exclusion, and stated that the allelic variability should provide good power of identification even for more complex mating patterns with some unknown parents.

Fisheries geneticists have been very active in this area of research. Rodríguez-Barreto et al. (2013) have looked at pedigree construction in the absence of parental information on a wild population of amberjack (*Seriola dumerili*). Using only four highly informative SSRs (7 alleles per marker, $H_0 > 0.80$), they were able to estimate the number of parents and construct families. Rodríguez-Ramilo et al. (2007) studied a population of *Scophthalmus maximus* (turbot) consisting of 139 full-sib families with known pedigree structure, but assumed no knowledge of parental genotypes. Using 10 SSRs with an average of 15.1 alleles, they converted molecular relatedness to coancestry (Queller and Goodnight 1989). For a population of 560, all possible pairwise relationships were examined, and classification of unrelated pairs, half-sibs, and full-sibs was fairly accurate, around 80 %. It seems likely that larger numbers of markers would increase the accuracy of pedigree reconstruction. In a simulation study based on data for oil palm (*Elaeis guineensis*), Cros et al. (2014) examined the correlation between true coancestry and ancestry estimated from molecular relatedness. They found that 10 SSR would give a correlation of around 0.80, while 36 SSR would increase the correlation to above 0.90. This result supports the idea that a breeder could simply increase the number of markers in order to reach a desired level of pedigree reconstruction accuracy.

A slightly different approach might be to use the molecular relatedness matrices directly to estimate variances and predict breeding values. This would avoid the difficulty of getting all individuals correctly classified as unrelated, half-sib or full-sib, and have the additional advantage of being able to use all

of genotyped individuals. As an example, Blonk et al. (2010) compared breeding value predictions made using a reconstructed pedigree and molecular relatedness for a large, natural mating population of *Solea solea* (common sole) with unknown parents. Population size was 1,953 individuals, and pedigree reconstruction was able to assign parentage for 1,338 of these. Using cross validation, they determined accuracy of estimated breeding values was 0.54 with pedigree reconstruction and 0.55 with molecular relatedness for the 1,338 individuals. However, they noted that if all 1,953 individuals were analyzed with molecular relatedness, accuracy increased to 0.60.

Phenotyping accuracy

The potential genetic gains discussed above also assumed accurate phenotyping. In other words, the assumed genetic parameters were derived from well-designed and maintained progeny tests with good survival; therefore, for the gains predicted here under the BwB scenarios to be accurate, the selection conditions must be similar. If the plantations are significantly more heterogeneous than progeny tests, then heritability would be lower and the BwB gains and efficiencies estimated in this study would be biased upwards. However, with a large plantation land base, it seems likely that the breeder could locate some small blocks with good growth and good uniformity, similar to a maintained progeny test. In this study, the necessary total size of such blocks would be roughly between 6 and 24 ha.

In the BwB strategy proposed in this study, there are three populations which must be phenotyped (i.e., measured or inspected): the total BwB population as well as the random and top-phenotype sub-populations. We envision the random sub-population will be phenotyped with one approach, and that the other two populations would be handled together with a different approach.

For random sub-population, the phenotyping should be done in at least two sections of the plantation (i.e., two different sites). The sites should have high survival, uniform spacing, and preferably with little within-site variation (i.e., low levels of environmental variation with the section of the plantation where the random population will be measured). To reduce environmental error, one option would be to map out the site with a row-column grid, and use spatial analysis techniques to model the environmental variation (Dutkowski et al. 2002). Another option would be use post hoc blocking as suggested by Gezan et al. (2006), to effectively create small randomized incomplete blocks of 25 to 50 trees for inclusion in the BLUP models.

For the BwB and top-phenotype populations, the situation is somewhat different. In the scenarios discussed here, the BwB population could range in size from 5,940 trees (approximately 6 ha of plantation assuming a spacing of 3 × 3 m) to

23,760 trees (approximately 24 ha of plantation). We envision a process very similar to that used for mass selection (Zobel and Talbert 1984; White et al. 2007). In summary, field crews would walk the plantations scanning from four to six rows at a time for outstanding phenotypes. When a potential top-phenotype is located, the crew would evaluate characteristics such as stem form, absence of disease, survival of near-neighbors, etc. If the tree appeared to be suitable, the crew would then take measurements of that tree and 15–20 of its nearest neighbors, in order to express the measurement of the top-phenotype tree as a deviation from the neighbor trees (or “comparison trees”). In this study, the size of the final top-phenotype population (which will be genotyped to determine parentage) was $N_T=600$. With $N_T=600$, the use of 15–20 comparison trees per candidate would result in between 9,000 and 12,000 comparison trees to be measured. This is a significant effort, roughly equivalent to half of the trees measured in the maximum FS strategy (23,760). The number of neighbors to measure could be reduced, either by reducing the number of comparison trees, or by decreasing the number in the top-phenotype population.

Regarding a reduction in the number of comparison trees, for mass selection the suggested number is often around 4–6 neighbors (e.g., Brown and Goddard 1961; White et al. 2007, p. 333). The purpose of the comparison trees is to estimate the site potential (or in other words, the population mean on that micro-site) in order to precisely calculate the deviation of the top-phenotype tree from the population. Use of 15–20 neighbors should give a fairly precise measurement of the deviation. It might be possible to reduce the number of neighbors to be measured (perhaps to 6–10), but this might also reduce genetic gains slightly.

Regarding a reduction in the number of Top Phenotypes, it seems there is relatively little benefit to genotyping more than $N_T=600$, as the only goal of this is to ensure sufficient genotypes to select enough unrelated trees to meet the required effective population size (N_e). Depending on the size of N_e , it may be possible to use $N_T<600$ with little impact on genetic gain. The optimum balance between workload and precision (i.e., the size of N_T and number of comparison trees per top-phenotype tree) is an area for future research.

Larger BwB populations

It is important to note here that it should be possible to make additional genetic gain by increasing N_{BwB} beyond the maximum of 23,760 examined in this study. Recall that the FS strategy outlined here was chosen only to provide some basis of comparison of the BwB gains with gains that could be made with a traditional breeding and testing program. Thus, the “maximum” gains in Figs. 2–3, i.e., the 100 % FS gain, are arbitrary. From Figs. 2 and 3, one can see that doubling the size of BwB population from $N_{BwB}=5,940$ to $N_{BwB}=11,880$ increases gain by about 2 % (and increases $Q_{BwB/FS}$ by about 3 to 5 %). Doubling the size of BwB again from $N_{BwB}=11,880$ to

$N_{BwB}=23,760$ increases gain by similar amounts. This implies that further increases in N_{BwB} would produce further gains. Operationally, this means that more hectares of plantation would be inspected to identify additional top phenotypes and thus increase selection intensity and within-family gain. Clearly, at some point, the marginal return from increasing N_{BwB} will be too small to justify the additional effort, but since walking plantations is likely to be a low-cost operation, breeders should consider BwB populations larger than 23,760 trees.

BwB beyond the first cycle

Although the focus of this study was on the initiation of tree improvement programs, the efficiency of the BwB approach and the potential genetic gains estimated in this study suggest that the BwB approach might have advantages beyond the first cycle of improvement. There is no reason that the approach could not be utilized with advanced generations. If it is possible to “uncover” genetic information from unstructured plantations, a breeder might choose to simply establish plantations, and later recover “progeny test information” when it is necessary and expedient. In other words, a breeder could pay upfront costs to establish progeny tests and then wait for the information, or he could pay costs later to do the BwB genotyping to gather the necessary genetic information.

Alternately, a breeder could consider a complementary approach, combining a small number of replicated progeny tests, with a BwB approach where all the trees in a large plantation landbase are considered as selection candidates. Some parental/family information would come from the replicated progeny tests, and a small number of top-phenotypes could be selected from the plantations to increase selection intensity and genetic gain, and these would be genotyped to control pedigree and optimize combinations of the best parental genotypes.

The utility of BwB beyond the first cycle would depend to a great extent on the accuracy of parental breeding value predictions that would arise from this approach, and this should be the subject of additional research. Another important topic for research would be a comprehensive study of the impact of errors in pedigree reconstruction on the efficiency of BwB.

Conclusions

The objective of this study was to examine the effectiveness of a BwB strategy to uncover genetic information from plantations as a way to begin a tree improvement program and avoid the initial cycle of breeding and testing. Assuming that pedigree reconstruction can be done accurately, it is clear that the BwB approach will be effective, and the strategy would return 80 to 98 % of the genetic gain from progeny selection that would have been achieved if a breeding and testing program had been started years earlier.

This study did not undertake a cost-benefit analysis, as genotyping costs continue to decrease, and other costs associated with field operations (breeding, test establishment, inspection of plantations for mass selection, etc.) would vary widely among countries and programs. A detailed cost/benefit analysis should be performed to assess the actual economic efficiency given specific cost and time components representing realistic tree improvement programs. However, we believe that the BwB approach will be cost-effective, requiring low levels of genotyping and a relatively low investment in evaluating the BwB and random populations (at least compared to the costs of progeny test establishment, maintenance, and measurement). The BwB strategy requires that 1,200 to 3,600 trees be genotyped, and based on the results of this study, we suggest that genotyping a random sub-population of 1,200 to 1,800 trees and a top-phenotype sub-population of no more than 600 trees should give near-optimum results. If the final selected population is small (e.g., $N_e=5$), the size of the top-phenotype sub-population could possibly be reduced in order to save on genotyping costs. The top-phenotype sub-population should be pre-selected from a very large population (the BwB population), on the order of 24,000 trees or more, probably established in typical plantations. The large BwB population does not need to be measured in its entirety, but it should be inspected thoroughly and accurate measurements taken on a small top-phenotype population.

Acknowledgments We would like to acknowledge the financial support from the Ministry of Education, Youth and Sports, Czech Republic (KONTAKT II; grant LH13021), and members of Camcore, North Carolina State University, USA.

Data archiving statement In our study, we described a stochastic simulation model and conducted a comparison of hypothetical breeding strategies. No real-world data of any species have been used throughout the study. Output data have been prepared and published in tables and figures and they are included in the manuscript.

Conflict of interest The authors declare that they have no conflict of interest.

Appendix: accuracy of prediction ($r_{\hat{a},a}$) for phenotypic, within-family, and mid-parent + within-family selection

1. Define $a=f+w$
where a =overall genetic value, f =midparent genetic value, and w =within-family genetic value, and $\text{Var}(a)=\sigma_A^2$, $\text{Var}(f)=1/2\sigma_A^2$, and $\text{Var}(w)=1/2\sigma_A^2$
2. Define $\hat{a}=\hat{f}+\hat{w}$
where $\hat{a}$ =predicted genetic value, $\hat{f}$ =predicted mid-parent genetic value, and $\hat{w}$ =predicted within-family genetic value.

3. The accuracy of prediction is defined as

$$r_{\hat{a},a} = [\text{Var}(\hat{a})/\text{Var}(a)]^{1/2}$$

4. For mass selection based on phenotypic value y ,

$$\hat{a} = h^2 S,$$

where h^2 =heritability= $\text{Var}(a)/\text{Var}(y)$, and S =phenotypic deviation of y , and accuracy of prediction= $[\text{Var}(\hat{a})/\text{Var}(a)]^{1/2}=h$ (Falconer 1981; Hodge and White 1992). So if $h^2=0.15$, then $r_{\hat{a},a}=0.39$.

5. Assuming infinite testing (number of tests, number of crosses, number of progeny), then mid-parent genetic values f will be predicted without error, that is, accuracy of prediction=1.00.

$$r_{f,f} = 1.00 = [\text{Var}(\hat{f})/\text{Var}(f)]^{1/2},$$

therefore $\text{Var}(\hat{f})=\text{Var}(f)=1/2\sigma_A^2$

6. For within full-sib family selection based on within full-sib family phenotypic deviation S_w ,

$$\hat{w} = h_w^2 S_w,$$

where h_w^2 =within-family heritability= $1/2\sigma_A^2 / [\text{Var}(y) - 1/2\sigma_A^2]$.

If $h^2=0.15$, $h_w^2=0.075/[1.00-0.075]=0.081$.

As above, accuracy of prediction equals the square root of the heritability, so accuracy of prediction= $h_w=[\text{Var}(\hat{w})/\text{Var}(w)]^{1/2}$

$$\begin{aligned}\text{Var}(\hat{w}) &= h_w^2 \text{Var}(w) \\ &= 0.081 \cdot 1/2\sigma_A^2 \\ &= 0.0405\sigma_A^2\end{aligned}$$

7. Now assume offspring selection (equivalent to mid-parent selection + within-family selection), with $h^2=0.15$, and with infinite progeny testing.

$$\hat{a} = \hat{f} + \hat{w}$$

$$\begin{aligned}\text{Var}(\hat{a}) &= \text{Var}(\hat{f}) + \text{Var}(\hat{w}) \\ &= 1/2\sigma_A^2 + 0.0405\sigma_A^2 \\ &= 0.5405\sigma_A^2\end{aligned}$$

The accuracy of selection would then be

$$\begin{aligned}r_{\hat{a},a} &= [\text{Var}(\hat{a})/\text{Var}(a)]^{1/2} \\ &= [0.5405\sigma_A^2/\sigma_A^2]^{1/2} \\ &= 0.7352\end{aligned}$$

- 8, Thus, with $h^2=0.15$, the theoretical maximum accuracy of prediction for offspring selection from full-sib families (mid-parent selection + within-family selection) is $r_{\hat{a},a}=0.7352$.

References

- Adams WT, Hipkins VD, Burczyk J, Randall WK (1997) Pollen contamination trends in a maturing Douglas-fir seed orchard. *Can J Forest Res* 27:131–134
- Blonk RJ, Komen H, Kamstra A, van Arendonk JA (2010) Estimating breeding values with molecular relatedness and reconstructed pedigrees in natural mating populations of common sole, *Solea solea*. *Genetics* 184:213–219
- Brown CL, Goddard RE (1961) Silvicultural considerations in the selection of plus phenotypes. *J Forest* 59:420–426
- Butler D, Cullis BR, Gilmour AR, Gogé BJ (2007) ASReml-R reference manual. Guide Release 2.00. VSN International Ltd, Hemel Hempstead, UK. URL <http://www.vsnl.co.uk/>
- R Core Team (2014) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <http://www.R-project.org/>
- Cros D, Sánchez L, Cochard B, Samper P, Denis M, Bouvet JM, Fernández J (2014) Estimation of genealogical coancestry in plant species using a pedigree reconstruction algorithm and application to an oil palm breeding population. *Theor Appl Genet* 127:981–994
- Dutkowski GW, Costa e Silva J, Gilmour AR, Lopez GA (2002) Spatial analysis methods for forest genetic trials. *Can J Forest Res* 32:2201–2214
- El-Kassaby YA, Cook C (1994) Female reproductive energy and reproductive success in a Douglas-fir seed orchard and its impact on genetic diversity. *Silvae Genet* 43:243–246
- El-Kassaby YA, Lstibůrek M (2009) Breeding without breeding. *Genet Res* 91:111–120
- El-Kassaby YA, Rudin D, Yazdani R (1989) Levels of outcrossing and contamination in two *Pinus sylvestris* L. seed orchards in northern Sweden. *Scand J Forest Res* 4:41–49
- El-Kassaby YA, Lstibůrek M, Liewlaksaneeyanawin C, Slavov GT, and Howe GT (2007) Breeding without breeding: approach, example, and proof of concept. In: Proc. IUFRO, Low input breeding and genetic conservation of forest tree species. Antalya, Turkey, pp. 43–54
- Falconer DS (1981) Introduction to Quantitative Genetics, Second Edition. Longman Group Limited, 340 p
- Funda T, Liewlaksaneeyanawin C, Fundová I, Lai BSK, Walsh C, Niejenhuis AV, Cook C, Graham H, Woods J, El-Kassaby YA (2011) Congruence between parental reproductive investment and success determined by DNA-based pedigree reconstruction in conifer seed orchards. *Can J Forest Res* 41:380–389
- Gezan SA, Huber DA, White TL (2006) Post hoc blocking to improve heritability and precision of best linear unbiased genetic predictions. *Can J Forest Res* 36:2141–2147
- Grattapaglia D, do Amaral Diener PS, dos Santos GA (2014) Performance of microsatellites for parentage assignment following mass controlled pollination in a clonal seed orchard of loblolly pine (*Pinus taeda* L.). *Tree Genet Genomes* 10:1631–1643
- Gurobi Development Team (2014) Gurobi Optimizer Reference Manual. Gurobi Optimization, Inc. Houston TX, USA. URL <http://www.gurobi.com/>
- Hansen OK, McKinney LV (2010) Establishment of a quasi-field trial in *Abies nordmanniana*—test of a new approach to forest tree breeding. *Tree Genet Genomes* 6:345–355
- Hartman HT, Kestler DE (1968) Plant propagation: Principles and Practices. Prentice Hall, Inc., 559 p
- Hodge GR, Dvorak WS (2012) Growth potential and genetic parameters of four Mesoamerican pines planted in the Southern Hemisphere. *South Forests* 74:27–49
- Hodge GR, White TL (1992) Selection and gain prediction. In: Fins L, Friedman ST, Brotschol JV (eds) Handbook of Quantitative Forest Genetics. Kluwer Academic Publishers, 403 pp
- Huber DA, White TL, Hodge GR (1992) Efficiency of half-sib, half-diallel, and circular mating designs of different sizes in the estimation of genetic parameters. *Forest Sci* 38:757–776
- Korecký J, Lstibůrek M, El-Kassaby YA (2014) Congruence between theory and practice: reduced contamination rate following phenotypic pre-selection within the Breeding without Breeding framework. *Scand J Forest Res* 29:552–554
- Lai BS, Funda T, Liewlaksaneeyanawin C, Klápště J, Niejenhuis AV, Cook C, Stoehr MU, Woods J, El-Kassaby YA (2010) Pollination dynamics in a Douglas-fir seed orchard as revealed by pedigree reconstruction. *Ann For Sci* 67:808
- Lambeth C, Lee B-C, O'Malley D, Wheeler N (2001) Polymix breeding with parental analysis of progeny: an alternative to full-sib breeding and testing. *Theor Appl Genet* 103:930–943
- Lindgren D, El-Kassaby YA (1989) Genetic consequences of combining selective cone harvesting and genetic thinning in clonal seed orchards. *Silvae Genet* 38:65–70
- Lstibůrek M, Ivanková K, Kadlec J, Klápště J, Kobliha J, El-Kassaby YA (2011) Minimum fingerprinting effort with respect to the effective population size. *Tree Genet Genomes* 7:1069–1078
- Lstibůrek M, Klápště J, Kobliha J, El-Kassaby YA (2012) Breeding without breeding: effect of gene flow on fingerprinting effort. *Tree Genet Genomes* 8:873–877
- Massah N, Wang J, Russell J, Van Niejenhuis A, El-Kassaby YA (2010) Genealogical relationship among members of selection and production populations of yellow cedar (*Callitropsis nootkatensis* [D. Don] Oerst.) in the absence of parental information. *J Hered* 101:154–163
- Meagher TR, Thompson E (1986) The relationship between single parent and parent pair genetic likelihoods in genealogy reconstruction. *Theor Popul Biol* 29:87–106
- Pharis RP, Kuo CG (1977) Physiology of gibberellins in conifers. *Can J For Res* 7:299–325
- Prescher F, Lindgren D, Karlsson B (2008) Genetic thinning of clonal seed orchards using linear deployment may improve both gain and diversity. *Forest Ecol Manag* 254:188–192
- Queller DC, Goodnight KF (1989) Estimating relatedness using genetic markers. *Evolution* 43:258–275
- Rodriguez-Barreto D, Consuegra S, Jerez S, Cejas JR, Martín V, Lorenzo A (2013) Using molecular markers for pedigree reconstruction of the greater amberjack (*Seriola dumerili*) in the absence of parental information. *Anim Genet* 44:596–600
- Rodríguez-Ramilo ST, Toro MA, Martínez P, Bouza C, Fernández J (2007) Accuracy of pairwise methods in the reconstruction of family relationships, using molecular information from turbot (*Scophthalmus maximus*). *Aquaculture* 273:434–442
- Slavov GT, Howe GT, Adams WT (2005) Pollen contamination and mating patterns in a Douglas-fir seed orchard as measured by simple sequence repeat markers. *Can J Forest Res* 35:1592–1603
- Wang X-R, Torimaru T, Lindgren D, Fries A (2010) Marker-based parentage analysis facilitates low input 'breeding without breeding' strategies for forest trees. *Tree Genet Genomes* 6:227–235
- White TL, Adams WT, Neale DB (2007) Forest Genetics. CAB International, 682 p
- Zobel B, Talbert J (1984) Applied forest tree improvement. Wiley, 505 p