Scandinavian Journal of Forest Research

ISSN: 0282-7581 (Print) 1651-1891 (Online) Journal homepage: www.tandfonline.com/journals/sfor20

# The heritability accuracy in forest tree breeding programs relying on random mating

Christi Sagariya, Arne Steffenrem, Salvador Gezan, Torsten Pook, Rosario García-Gil, Chedly Kastally, Tanja Pyhäjärvi & Milan Lstibůrek

**To cite this article:** Christi Sagariya, Arne Steffenrem, Salvador Gezan, Torsten Pook, Rosario García-Gil, Chedly Kastally, Tanja Pyhäjärvi & Milan Lstibůrek (2025) The heritability accuracy in forest tree breeding programs relying on random mating, Scandinavian Journal of Forest Research, 40:5-6, 292-302, DOI: 10.1080/02827581.2025.2530427

**To link to this article:** https://doi.org/10.1080/02827581.2025.2530427

SCANDINAVIAN JOURNAL OF FOREST RESEARCH
2025, VOL. 40, NOS. 5–6, 292–302
https://doi.org/10.1080/02827581.2025.2530427

Taylor & Francis
Taylor & Francis Group



# The heritability accuracy in forest tree breeding programs relying on random mating

Christi Sagariya[a], Arne Steffenrem[b], Salvador Gezan[c], Torsten Pook[d], Rosario García-Gil[e], Chedly Kastally[f], Tanja Pyhäjärvi[f] and Milan Lstibůrek[a]

[a]Faculty of Forestry and Wood Sciences, Czech University of Life Sciences Prague, Praha 6, Czech Republic; [b]Norwegian Institute of Bioeconomy Research (NIBIO), Steinkjer, Norway; [c]VSN International, Hemel Hempstead, United Kingdom; [d]Animal Breeding and Genomics, Wageningen University & Research, Wageningen, The Netherlands; [e]Umeå Plant Science Centre (UPSC), Swedish University of Agricultural Sciences, Umeå, Sweden; [f]Department of Forest Sciences, University of Helsinki, Helsinki, Finland

**ABSTRACT**

In forest tree breeding, there is a growing trend towards using pedigree reconstruction after offspring are produced through natural random mating (open-pollination), as an alternative to traditional controlled crosses and full-sib genetic trials. Given that most forest tree species are predominantly outcrossing organisms, the accuracy of narrow-sense heritability ($h^2$) under natural mating conditions has not been thoroughly examined, particularly with regard to sample size. Our simulation study focuses on the genetic parameters specific to Norway spruce (*Picea abies* L.) in Norway. We used the stochastic model MoBPS to simulate a founder population with 100,000 SNP markers distributed across 12 haploid chromosomes, representative of many conifer species. Parental trees were selected from this population, followed by random mating and offspring evaluation. We focused on estimating the accuracy of $h^2$, with particular attention to its precision and bias. Our results suggest that the population sample sizes currently used in forest tree breeding are generally adequate for achieving precision. However, we identified two primary sources of bias: one due to dominance effects and the other from phenotypic parental selection. We discuss potential strategies to mitigate these biases in breeding programs.

**ARTICLE HISTORY**
Received 1 November 2024
Accepted 2 July 2025

**KEYWORDS**
Breeding without breeding; tree improvement; stochastic simulation; genetic gain and diversity

## Introduction

Forest tree breeding programs are long-term endeavors involving a complex process of mating, genetic evaluation, and selection. Traditional tree breeding relies on controlled pollination, requiring substantial investments and long-term commitment (White et al. 2007). Narrow-sense heritability ($h^2$) determines the rate of response to both artificial and natural selection, making its estimation a critical initial step in plant and animal breeding programs (Falconer and Mackay 1996). With advances in genomic tools, it has become feasible to determine the pedigree of offspring populations originating from seed orchards (Marshall et al. 1998; Kalinowski et al. 2007).

In view of this, El-Kassaby et al. (2007) and El-Kassaby and Lstibůrek (2009) proposed a cost-effective and streamlined method known as 'Breeding without Breeding' (BwB). The methodology is based on natural random mating (panmixia) in large seed orchards, followed by extensive phenotyping in forest stands that originate from a shared parental source. El-Kassaby and Lstibůrek (2009) proposed a genetic evaluation approach using random and top-phenotypic subsets within forest stands, which are jointly treated as a candidate population. The random subset serves to estimate genetic variances required for calculating breeding values of top-ranking individuals in the top-phenotypic subset. Details on sample sizes are extensively discussed by Lstibůrek et al. (2011, 2012, 2015). Subsequent pedigree reconstruction of the offspring population allows for the selection of superior individuals, establishing genetically improved seed orchards for the next breeding cycle. BwB has been successfully implemented on a large scale in the European larch (*Larix decidua* L.) breeding program in Austria, assuming multiple forest stands to capture environmental gradients and genotype-by-environment interactions (Lstibůrek et al. 2020). Alternative successful

**CONTACT** Milan Lstibůrek lstiburek@fld.czu.cz Faculty of Forestry and Wood Sciences, Czech University of Life Sciences Prague, Kamýcká 129, Praha 6 165 00, Czech Republic

Supplemental data for this article can be accessed online at http://dx.doi.org/10.1080/02827581.2025.2530427.

implementations of pedigree reconstructions in forest tree breeding have been reported by Vidal et al. (2015), Hansen and McKinney (2010) and Lambeth et al. (2001), eliminating the need for laborious control crosses.

Computer simulations have been employed in animal and plant breeding to optimize breeding programs. In particular, stochastic simulations have proven valuable in addressing the complex genetic architecture of multiple quantitative traits. Several computer simulation applications have been developed for this purpose. QuLine (Wang and Wolfgang 2007) is applied in crop improvement for inbred lines, QMSim (Sargolzaei and Schenkel 2009) is utilized for complex livestock pedigrees in mutation drift equilibrium. DeltaGen (Jahufer and Luo 2018) is a decision-support analytical tool used in plant breeding although it doesn't permit the use of real genomic maps. AlphaSim (Faux et al. 2016) and AlphaSimR (Gaynor et al. 2021) are more flexible and frequently used in plant and animal breeding. MoBPS (Pook et al. 2020) is highly flexible and computationally faster with bit-wise storage capabilities. It can handle very large-scale complex breeding programs in both plant and animal breeding and also allows users to customize functions for breeding value estimation. MoBPS is additionally linked to commercial software such as MiXBLUP (ten Napel et al. 2020) and blupf90 (Misztal et al. 2014) for breeding value estimation.

According to El-Kassaby and Lstibůrek (2009) the BwB provides multiple opportunities as a feasible alternative to resource-dependent controlled pollination breeding programs. It benefits in terms of time and resource savings by circumventing the crossing and field experimental trials. Moreover, this methodology is appropriate and suitable for minor commercial tree species facilitating the development of gene conservation and management programs. In an effort to delve deeper into these findings, Lstibůrek et al. (2015) conducted a more comprehensive quantitative-genetic evaluation of the BwB strategies. They estimated genetic parameters using additive relationship matrices for specific sub-populations, including random and top phenotypic subsets, while varying effective population sizes ($N_e$). The results of their study indicated that only a relatively small subset of the offspring population (1200 random + 600 top-ranking phenotypes) is necessary for pedigree reconstruction and genetic evaluation, potentially enabling the development of economically and logistically effective breeding strategies. Given the advancements in BwB, further evaluation of the aforementioned breeding approaches utilizing natural mating followed by pedigree reconstruction is necessary to enhance the accuracy of genetic parameter estimates.

Recent research initiatives, including the Swedish project "Landscape Breeding: A new paradigm in forest tree management" (initiated in 2022) and the project "Management of forest genetic resources under climate change" (focused on adapting BwB in Norway and the Czech republic, completed in 2024), have aimed to integrate the BwB methodology into conventional breeding programs, with an emphasis on enhancing resilience and managing genetic diversity in the face of climate change.

Our objective was to examine how key factors, such as plus tree selection, offspring population sampling (sizes of random and top-phenotypic subsets), QTL additive and dominance effects, and the presence of negative genetic correlations, influence the accuracy of $h^2$. Building on the simulation work by Lstibůrek et al. (2015), we employed stochastic simulations followed by a comprehensive sensitivity analysis. Through this exploratory framework, we identified conditions under which BwB could be feasible within the practical scenarios and genetic parameters typical of Norway spruce breeding programs in the Nordic region.

## Materials and methods

We employed an allelic-based stochastic simulation model tailored for forest tree breeding. To model the BwB strategy, we initiated a breeding program involving the selection of plus trees, random mating, and the subsequent genetic evaluation of the offspring population. This modeling process encompassed three key stages: (a) Population simulation; (b) Plus tree selection; and (c) Genetic evaluation. Our primary objective was to estimate ($h^2$) of two quantitative traits (height and stem straightness). The MoBPS (1.10.59) software (Pook et al. 2020) facilitated the population simulation and the selection of random offspring subset within the R system (R Core Team 2023). For subsequent genetic evaluations, we employed the ASReml-R (4.1.0.160) software (Butler et al. 2017). The entire simulation pipeline was initially repeated across 30 independent replications to balance computational time and obtain confidence intervals suitable for illustrating the precision of $h^2$. However, after detecting bias and aiming to refine the mean estimates, we incorporated 100 replications, which are presented in the Supplement.

### *Population simulation*

A founder population of $n_f = 5000$ haplotypes was generated, incorporating a set of 100,000 SNP markers equidistantly distributed across 12 haploid chromosomes (as in many conifers), each 100 cM in size. Allele frequencies were sampled from a beta distribution (Gupta and

**Table 1.** Input parameters for population simulation. The genetic parameters were adopted from a diallel experiment on Norway spruce, evaluated at 7–10 years. (Skrøppa et al. 2023).

| Parameter | Value |
|---|---|
| Founder population size, $n_f$ | 5000 |
| Number of parents, $n_p$ | 40, 60, 80, 160, 320 |
| Number of offspring individuals, $n_o$ | 6000 |
| Random subset of offspring, $n_r$ | 500, 1000, 3000 |
| Narrow-sense heritability, $h^2$ | height = 0.2, stem straightness = 0.3 |
| Dominance variance ratio, $\sigma_d^2/\sigma_a^2$ | 0.2 |
| Additive genetic correlation, $r_g$ | 0, −0.25 |

**Table 2.** Simulation scenarios.

| Scenario | Description |
|---|---|
| S1: Reference | No dominance, Random set of parents, and No additive genetic correlation |
| S2: Dominance variance | Dominance, Random set of parents, and No additive genetic correlation |
| S3: Phenotypic selection | No dominance, Phenotypically selected parents, and No additive genetic correlation |
| S4: Genetic correlation | No dominance, Random set of parents, and Negative additive genetic correlation |
| S5: Combined | Dominance, Phenotypically selected parents, and Negative additive genetic correlation |

Nadarajah 2004) with a shape parameter alpha = 0.2 and beta = 1.1. Individual alleles were randomly sampled, and 5000 individuals were randomly mated for 30 generations to generate a linkage-disequilibrium (LD) structure typical of Norway spruce (Heuertz et al. 2006; Larsson et al. 2013).

Tree height and stem straightness were simulated as quantitative traits with mixed genetic architecture and both were assumed to be controlled by 300 loci with additive effects sampled from $N(0, \sigma_a^2)$. For dominance effects, full dominance QTLs ($a = d$) was considered, with effect sizes sampled from $N(0, \sigma_d^2)$ (Falconer and Mackay 1996). Simulation input parameters for the respective traits are provided in Table 1.

### Selection

A subset consisting of $n_p = 40$; 60; 80; 160; and 320 parental trees (also referred to as plus trees) was selected from the simulated founder population. Subsequently, random (open-pollinated) mating among these selected plus trees was conducted to generate an offspring population ($n_o = 6000$). Random subsets of offspring ($n_r$) were then chosen, with sample sizes set at 500, 1000, and 3000 for the purpose of genetic evaluation.

### Genetic evaluation

A pedigree-based (additive genetic relationship) genetic analyses used the animal genetic model (Henderson 1984), as presented in Equation (1). The analysis involved a randomly selected subset of the offspring ($n_r$) population, assuming their pedigree is fully known (in reality, pedigree reconstruction is performed based on DNA marker data). The effects of $n_r$ and $n_p$ on the accuracy of $h^2$ were examined in consecutive simulation scenarios (Table 2).

### Simulation scenarios

In Scenario 1 (S1), we assumed a random sampling of parental trees from the founder population (reflecting the low selection accuracy during the initial phases of low-input breeding programs Lstibůrek et al. 2015), no dominance, and the additive genetic correlation was zero. This scenario is consequently used as a reference for comparison. In S2, we are introducing the additional effect of dominance on top of the reference S1. Similarly, in S3, we expand the reference scenario by phenotypic selection of parental trees (plus-tree selection), considering an index with equal weight given to both traits. The negative additive genetic correlation is investigated in S4. Next, in S5, we incorporate all described factors (dominance, phenotypic parental selection, negative genetic correlation). and study their joint effect on the accuracy of $h^2$. Refer to Table 2 for more detailed description.

### Statistical analyses

Based on phenotypic observations and the pedigree information of the parental population and offspring in the selected subset $n_r$, we performed a linear mixed-model (LMM) genetic evaluation (Henderson 1953). We begin by presenting the univariate LMM, which we later extend to the bivariate case used in our study:

$$\boldsymbol{y} = \mathbf{1}\mu + \boldsymbol{Z_1 a} + \boldsymbol{Z_2 f} + \boldsymbol{e} \quad (1)$$

where $\boldsymbol{y}$ is the response variable; $\mu$ is the overall mean; $\boldsymbol{a}$ denotes additive genetic effects (i.e. breeding values) with $\boldsymbol{a} \sim N(0, \sigma_a^2 \boldsymbol{A})$, where $\boldsymbol{A}$ is the average numerator (pedigree-based) relationship matrix and $\sigma_a^2$ is the additive genetic variance; $\boldsymbol{f}$ are family effects, with $\boldsymbol{f} \sim N(0, \sigma_f^2 \boldsymbol{I_f})$, where $\boldsymbol{I_f}$ is an identity matrix of order $f$ and $\sigma_f^2$ is the family genetic variance; $\boldsymbol{e}$ are random residual effects, with $\boldsymbol{e} \sim N(0, \sigma_e^2 \boldsymbol{I_n})$, where $\boldsymbol{I_n}$ is an identity matrix of an order $n$, and $\sigma_e^2$ is the residual variance. In addition, $\boldsymbol{Z_1}$ and $\boldsymbol{Z_2}$ are the incidence matrices for their respective terms.

The bivariate version of Equation (1) was fitted to model both traits simultaneously, as follows:

$$\boldsymbol{y} = \boldsymbol{Xt} + \boldsymbol{Z_1 a.t} + \boldsymbol{Z_2 f.t} + \boldsymbol{e} \quad (2)$$

where $\boldsymbol{y} = [\boldsymbol{y_1}' \boldsymbol{y_2}']'$ are the stacked phenotypic records of the two traits; $\boldsymbol{t}$ is the fixed effect of the overall mean of each trait; $\boldsymbol{a.t}$ represents additive genetic variance effects nested within each trait, with $\boldsymbol{a.t} \sim N(0, \boldsymbol{G_a} \otimes \boldsymbol{A})$,

where $\boldsymbol{G_a}$ is the "corh" variance structure of dimension $2 \times 2$ defined by the additive variance for each trait and the additive correlation between traits; $\boldsymbol{f.t}$ are family effects nested within each trait, with $\boldsymbol{f.t} \sim N(0, \boldsymbol{G_e} \otimes \boldsymbol{I_f})$; $\boldsymbol{e}$ are random residual effects nested within each trait, with $\boldsymbol{f.t} \sim N(0, \boldsymbol{G_e} \otimes \boldsymbol{I_e})$. Both the $\boldsymbol{G_f}$ and $\boldsymbol{G_e}$ are variance structures of dimension $2 \times 2$ defined in the same way as $\boldsymbol{G_a}$. All of the other terms were previously described. This full model was applied to S2 and S5 where the dominance was incorporated. In other scenarios, the model was reduced by dropping the dominance term from the equations. Since the random offspring segments ($n_r$) were considered to be located on a single site, the only fixed effect was the site's overall mean.

The variance component estimates were then used to calculate $h^2$ as:

$$\widehat{h^2} = \frac{\widehat{\sigma_a^2}}{\widehat{\sigma_a^2} + \widehat{\sigma_f^2} + \widehat{\sigma_e^2}} \quad (3)$$

## Results

After 30 generations of random mating within the founder population, we utilized the complete simulated SNP haplotypic datasets to generate the LD decay structure and allele frequency spectrum. We estimated the correlation coefficient ($r^2$) between pairs of loci, reflecting the level of LD. The $r^2$ values ranged from $4.2 \times 10^{-4}$ to $5 \times 10^{-3}$, with an average of $1.5 \times 10^{-3}$, indicating that most loci are in LD between each other.

In the following sections, we present the results of the respective simulation scenarios. It is important to note that each parameter has three distinct values. For instance, $h^2$ is initially set as a simulation input parameter. Subsequently, upon generating the population, the true $h^2$ is computed as the variance of the true (computer-generated) additive genetic (breeding) values divided by the variance of the true (computer-generated) phenotypic values. Lastly, we report $\widehat{h^2}$ following genetic evaluation using the ASReml software. This third value closely resembles the one calculated by tree breeders in real breeding programs. When examining the graphs, it is crucial to recognize that a single replication of simulation theoretically represents a single instance of the actual breeding program. Therefore, it is not only the averages of the presented parameters that are significant but also the corresponding confidence intervals.

### *Scenario 1: reference*

In Figure 1(a,b), we present the difference between $h^2$ (averaged across 30 stochastic iterations) and $\widehat{h^2}$, along with their respective confidence intervals. Assuming the 30 iterations and a significance level of $\alpha = 0.05$, we can conclude that there is no statistically significant difference between $h^2$ and $\widehat{h^2}$ in S1. As anticipated, higher values of $n_r$ led to improved precision (resulting in narrower confidence intervals, CIs), irrespective of $n_p$. This suggests that an optimal $n_p$ falls within the 60–80 range, while higher values of $n_r$, exceeding 1000, were necessary to achieve high accuracy in estimating $h^2$. With the increase in sample size $n_r$ from 500 to 3000, the CIs of $\widehat{h^2}$ decreased by 47% for height and 38% for stem straightness, assuming $n_p = 40$. Similarly, with $n_p = 320$, they decreased by 81% and 72%, respectively. $\widehat{h^2}$ showed relatively less variability for stem straightness compared to height, which can be attributed to differences in the initial $h^2$ and $\sigma_e^2$ between these traits in the founder population. As a result, better precision was achieved for both $h^2$ and $\widehat{h^2}$ for stem straightness compared to height. In summary, the results suggest that while the number of selected plus trees appears to have no discernible effect on the accuracy of $h^2$, the sample size $n_r$ plays a more influential role. Increasing the number of replications for the same strategy revealed no evidence of bias, as indicated in suppl. Figure 1(a,b). This suggests that the precision, and consequently the accuracy, of $\widehat{h^2}$ can be enhanced by adjusting $n_r$ primarily.

### *Scenario 2: dominance variance*

Both the true $h^2$ and observed $\widehat{h^2}$ values closely aligned with the initial input values, as expected (Figure 2(a,b)). The precision of $\widehat{h^2}$ improved significantly, with CIs decreasing by 26% for $n_p = 40$ and 70% for $n_p = 320$ in both traits, as the sample sizes $n_r$ increased from 500 to 3000. This reduction followed a similar pattern to S1. Interestingly, the CIs were wider in stem straightness than for height, which is contrary to the trend observed in S1, particularly for smaller $n_r$ sizes (500, 1000). These results suggest that $n_p$ sizes lower than 80, combined with larger $n_r$ sizes, are relatively optimal for achieving accuracy in estimating $h^2$ under both additive and dominant effects. We emphasize the importance of attaining precision in $h^2$ estimates by optimizing $n_r$ sizes, preferably aiming for 3000 alongside appropriate $n_p$ sizes. However, we must highlight a bias identified with 100 replications (Suppl. Figure 2(a,b)). Pairing smaller sample sizes ($n_r < 3000$) with a higher number of plus trees ($n_p > 60$), which effectively creates a larger number of smaller families, systematically underestimates $h^2$.

In Suppl. Figure 3(a), we present true and estimated $\sigma_d^2$ values. As expected, we found $\widehat{\sigma_d^2}$ ($4\widehat{\sigma_f^2}$) was biased upwards, with the bias proportional to family size. To obtain unbiased estimates of $\widehat{\sigma_d^2}$, the average family

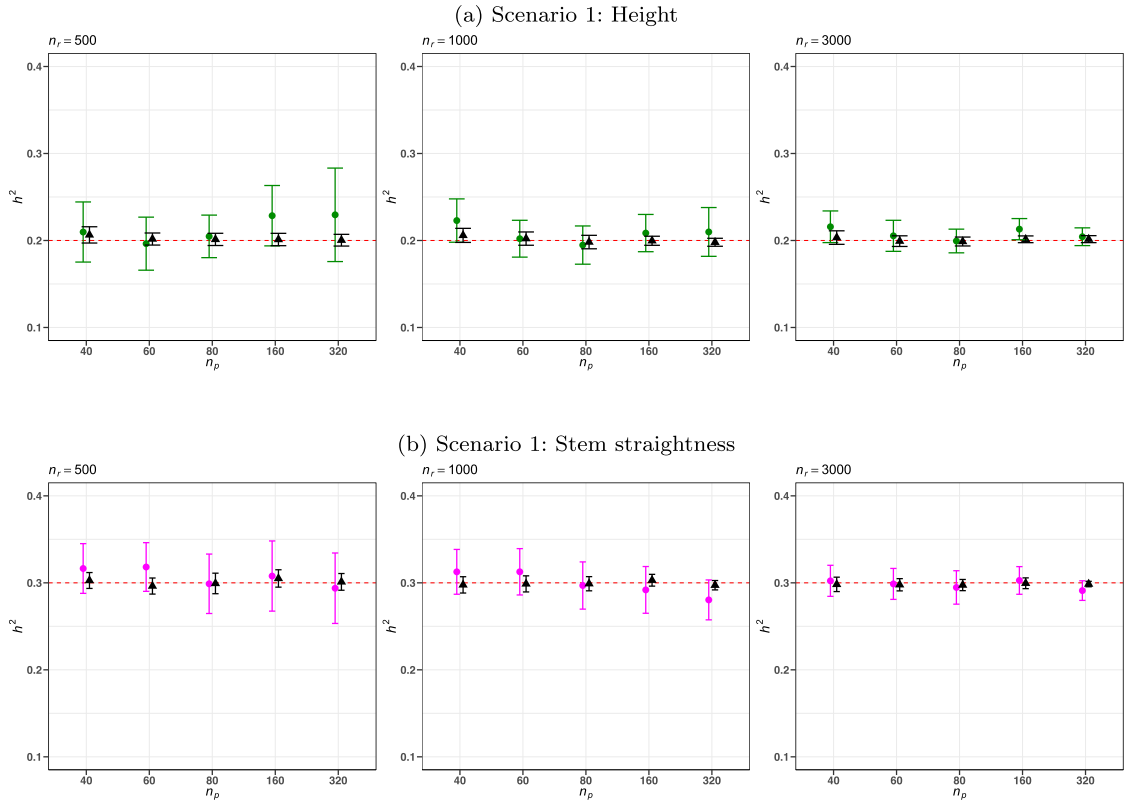


**Figure 1.** Scenario 1 "Reference". The *X*-axis indicates the number of plus trees ($n_p$), and the Y-axis displays the true narrow-sense heritability $h^2$ (black triangles with 95% CIs calculated across 30 replications) alongside its respective estimate (green and pink dots for height and stem straightness traits). The red dashed line indicates the initial $h^2$ values employed to generate the founder population (simulation input). The upper set of three graphs (a) illustrates height, while the lower set (b) depicts the stem straightness trait. The leftmost graphs correspond to a sample size of 500 ($n_r$), the center graphs to $n_r = 1000$, and the rightmost graphs to $n_r = 3000$. (a) Scenario 1: Height and (b) Scenario 1: Stem straightness.

size must reach approximately one offspring per family (assuming random mating), ideally close to four, as observed with $n_r = 3000$ and $n_p = 40$. Conversely, the maximum observed bias occurred with the smallest family sizes, specifically when $n_r = 500$ and $n_p = 320$, resulting in an average family size close to 0.01. In this extreme scenario, the ASReml model failed to converge.

To further investigate this phenomenon, we amplified the dominance variance by setting variance ratios $\sigma_d^2/\sigma_a^2$ to 0.5 and 1.0. As shown in Suppl. Figures 7 and 8, we observed a more pronounced downward bias under conditions of higher dominance variance, consistent with previously reported trends. Specifically, a combination of higher $n_p$ and lower $n_r$ values amplifies this bias.

### *Scenario 3: phenotypic selection of plus trees*

The results of S3, involving phenotypic selection of plus-trees, are illustrated in Figure 3(a,b). In comparison to the S1, here we revealed a noticeable reduction in the true $h^2$ due to the phenotypic selection of plus trees.

Suppl. Figure 4(a,b) reveal a significant bias resulting from the omission of phenotypic selection of plus trees in the genetic evaluation model. This leads to an underestimation of $h^2$, regardless of the values for $n_r$ and $n_p$. This issue is particularly concerning because it appears there is no solution to mitigate this bias when phenotypic selection is present, a common practice in the early stages of tree breeding programs.

The precision of the $\widehat{h^2}$ estimate could be improved by increasing the values of $n_r$ regardless of the $n_p$ sizes. Conversely, estimates are more precise with higher $h^2$ values and smaller $n_r$ sizes (500, 1000), particularly when the size of $n_p$ falls within the range of 60 to 80, similar to S1.

### *Scenario 4: negative additive genetic correlation*

The results of this scenario (Figure 4(a,b)) revealed significant variation in both $h^2$ and $\widehat{h^2}$, with simulated height displaying wider CIs than stem straightness, consistent with the reference scenario. Notably, CIs were substantially narrower with higher $n_r$ (3000) compared

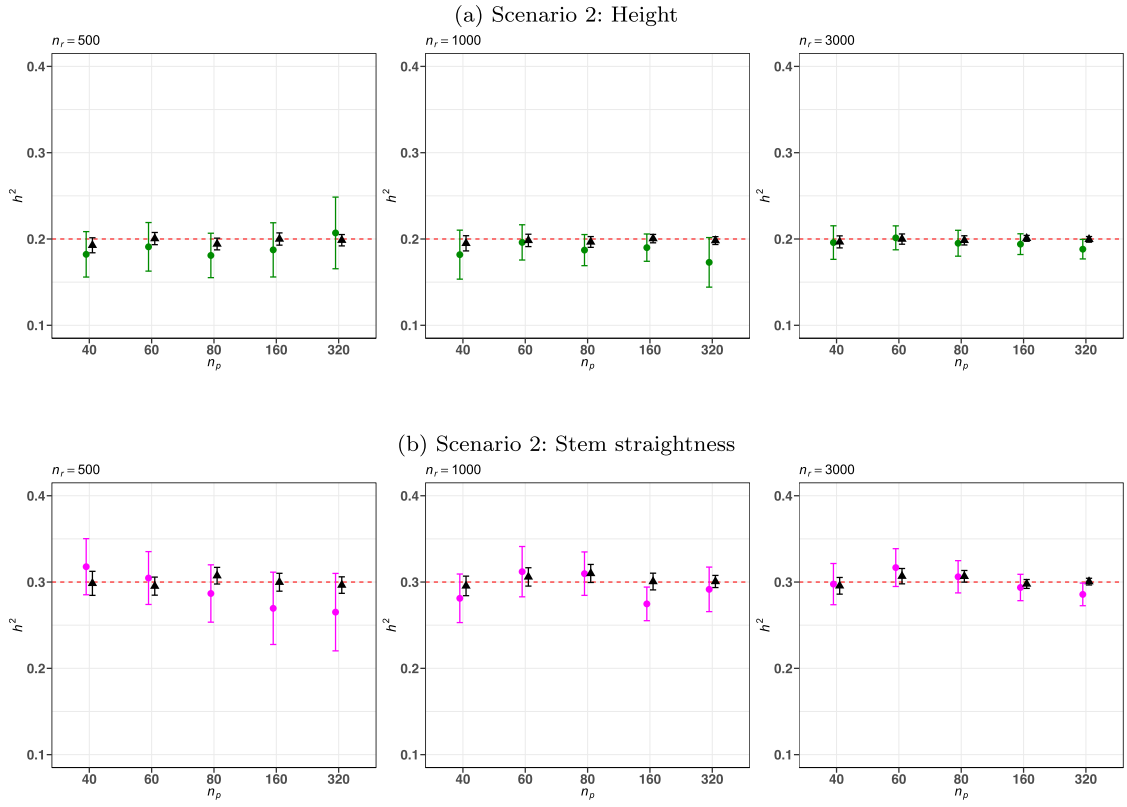


**Figure 2.** Scenario 2 "Dominance variance". The *X*-axis indicates the number of plus trees ($n_p$), and the *Y*-axis displays the true narrow-sense heritability $h^2$ (black triangles with 95% CIs across 30 replications) alongside its respective estimate (green and pink dots for height and stem straightness traits). The red dashed line indicates the initial $h^2$ values employed to generate the founder population (simulation input). The upper set of three graphs (a) illustrates height, while the lower set (b) depicts the stem straightness trait. The leftmost graphs correspond to a sample size of 500 ($n_r$), the center graphs to $n_r = 1000$, and the rightmost graphs to $n_r = 3000$. (a) Scenario 2: Height and (b) Scenario 2: Stem straightness.

to smaller values (500 and 1000). Moreover, larger sizes of $n_p > 80$ resulted in closer CIs than smaller $n_p$ sizes. Specifically, the width of the CIs for $n_p = 40$ with $n_r = 3000$ reduced by 42% in height and 38% in stem straightness compared to smaller $n_r$ sizes of 500. Similarly, for $n_p = 320$ with $n_r = 3000$, reductions of 76% and 65% were observed in height and stem straightness, respectively. Compared to S1, the current scenario reveals an average reduction of 5% for height and 4% for stem straightness in CIs, suggesting that differences between S4 and S1 are not significant but possibly interesting. The findings of S4 suggest that additive genetic correlation has no adverse effect on the accuracy of $h^2$ estimation. Plus tree selection, such as $n_p$ ranging from 60 to 80 combined with higher $n_r$, could lead to greater precision when considering negatively correlated traits in genetic evaluation, thereby enhancing the accuracy of $\widehat{h^2}$. Increasing the number of stochastic iterations within the same strategy revealed no evidence of bias, as shown in Suppl. Figure 5(a,b). This indicates that the precision, and thus the accuracy, of $\widehat{h^2}$ can primarily be improved by adjusting $n_r$.

### *Scenario 5: combined model*

The approach in S5 involves all the studied factors, resembling operational tree breeding programs in estimating $h^2$ (Figure 5(a,b)). Both $h^2$ and $\widehat{h^2}$ show high variation across different $n_p$ and $n_r$ sizes. In line with S1, the CIs for height were reduced by 49% and for stem straightness by 32% when $n_p = 40$, while for $n_p = 320$, reductions of 71% for height and 75% for stem straightness were observed. As $n_p$ values increased alongside $n_r$, the CIs narrowed, and the mean values approached the simulation input values, akin to S1.

The combined model exhibits two main sources of bias, as previously identified: dominance and phenotypic selection. Both factors consistently lower the estimates of $h^2$ beneath the actual $h^2$ value. This effect becomes more apparent when the number of stochastic iterations increases from 30 to 100, as demonstrated in Suppl. Figure 6(a,b). On average, across the studied parameters, the estimates are 7% lower than the true value for both traits. The most significant observed bias, a 20% reduction, was noted in stem straightness at $n_r = 500$

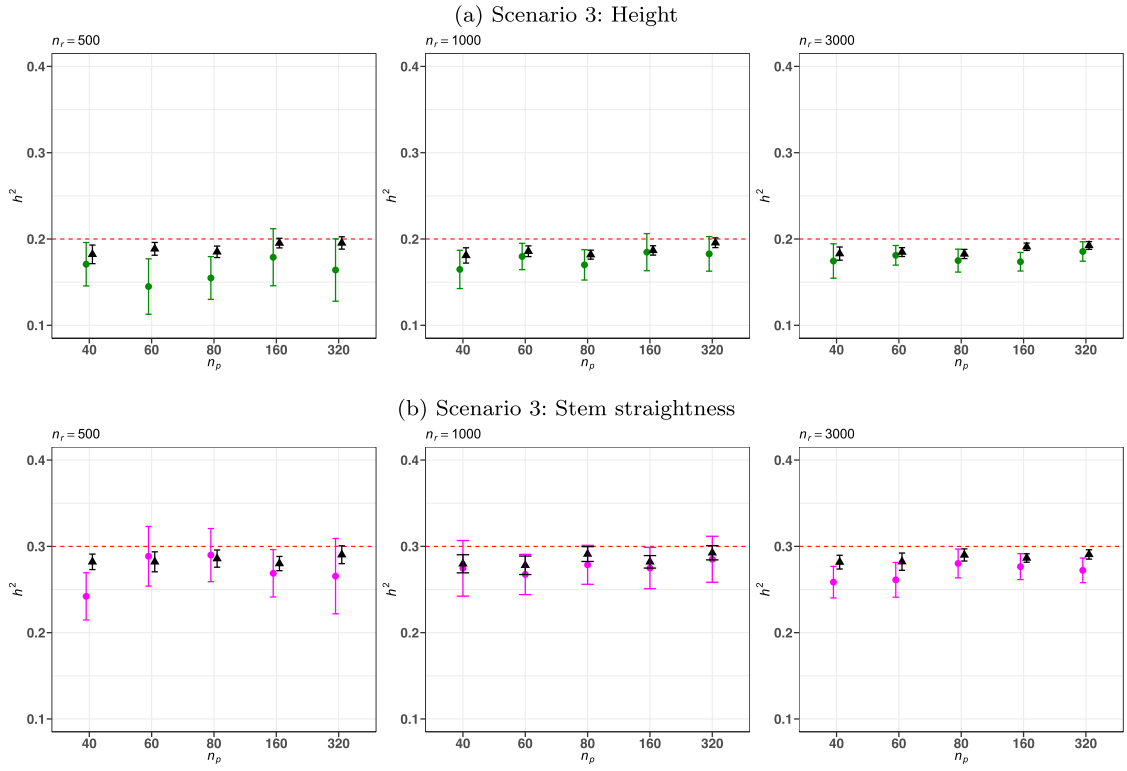


**Figure 3.** Scenario 3 "Phenotypic selection". The *X*-axis indicates the number of plus trees ($n_p$), and the *Y*-axis displays the true narrow-sense heritability $h^2$ (black triangles with 95% CIs across 30 replications) alongside its respective estimate (green and pink dots for height and stem straightness traits). The red dashed line indicates the initial $h^2$ values employed to generate the founder population (simulation input). The upper set of three graphs (a) illustrates height, while the lower set (b) depicts the stem straightness trait. The leftmost graphs correspond to a sample size of 500 ($n_r$), the center graphs to $n_r = 1000$, and the rightmost graphs to $n_r = 3000$. (a) Scenario 3: Height and (b) Scenario 3: Stem straightness.

and $n_p = 320$. The trends in $\sigma_d^2$ estimation bias observed previously (Suppl. Figure 3(b)), specifically the dependence on family size and the challenges with small families, remain fully applicable to the analyses presented here. This suggests that although breeders might address dominance by manipulating the sample size $n_r$, the issue of phenotypic selection will continue to pose challenges. Means and confidence intervals for all scenarios assuming the 30 simulation iterations are provided in Supplementary Tables 1 and 2.

## Discussion

Novel tree breeding strategies that combine random mating with subsequent pedigree reconstruction are proving to be effective alternatives in various species. By concentrating phenotyping and genotyping efforts on a relatively smaller subset of the candidate population, and fully abandoning full-sib artificial mating, significant time savings are achieved in operational breeding programs. Lstibůrek et al. (2015) estimated that evaluating approx. 1800 trees comprising 1200 random and 600 top phenotypes could yield 85 to 95% of the genetic gains comparable to those from large-scale full-sib testing programs. In a subsequent study, Lstibůrek et al. (2017) suggested that for Norway spruce in Norway, these 1800 trees should be selected from commercial forest plantations. With six such plantations, each containing a random mix of 6000 trees from open-pollinated seed orchard parents, there seems to be a sufficient candidate base to achieve genetic gains comparable to those from traditional breeding programs utilizing full-sib progeny trials.

In this study, we did not focus on directly estimating genetic gains (which would require evaluating the top-phenotypic segment of the candidate population). Instead, we concentrated on analyzing the accuracy of $h^2$, a reliable indicator of potential genetic gain. Theoretically, the standard error of $h^2$ from parent-offspring regression scales with the square root of twice the reciprocal of the sample size under ideal conditions (Falconer and Mackay 1996). Our simulations, incorporating complex genetic architecture and additional population characteristics, indicate that a genetic variance decomposition based on approx. 1000 offspring is adequate, regardless of the number of selected parental "plus" trees (40 to 320).

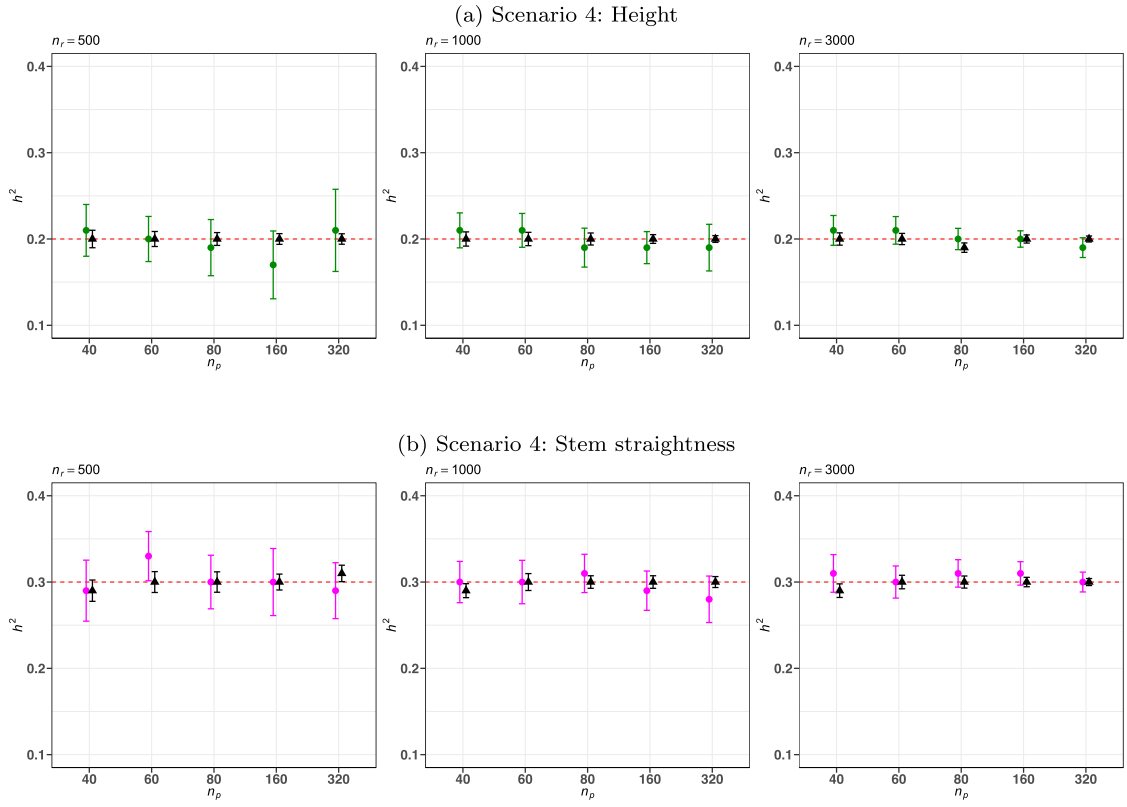


**Figure 4.** Scenario 4 "Genetic correlation". The *X*-axis indicates the number of plus trees ($n_p$), and the *Y*-axis displays the true narrow-sense heritability $h^2$ (black triangles with 95% CIs calculated across 30 replications) alongside its respective estimate (green and pink dots for height and stem straightness traits). The red dashed line indicates the initial $h^2$ values employed to generate the founder population (simulation input). The upper set of three graphs (a) illustrates height, while the lower set (b) depicts the stem straightness trait. The leftmost graphs correspond to a sample size of 500 ($n_r$), the center graphs to $n_r = 1000$, and the rightmost graphs to $n_r = 3000$. (a) Scenario 4: Height and (b) Scenario 4: Stem straightness.

However, using only 500 random offspring may compromise the precision of $h^2$ estimates, particularly when the ratio $\sigma_d^2/\sigma_a^2$ exceeds 0.2. If the dominance variance of certain traits is significant, larger sample sizes (around 3000 offspring) combined with fewer parents (less than 80) are necessary. Adequate family sizes are crucial for accurately capturing this component; otherwise, the $h^2$ estimate is likely biased downwards. Understanding the inherent variation in family sizes during natural random mating is crucial. A large number of parental combinations combined with variable family sizes affect genetic evaluation in forest tree breeding under open-pollination schemes, particularly when dominance variance is involved. For example, with 160 parents, there are $\binom{160}{2} = 12{,}720$ unique mating pairs. Assuming 1000 offspring, the mean family size is $1000 \times \frac{1}{12{,}720} \approx 0.079$, and the variance of family size is $1000 \times \frac{1}{12{,}720} \times (1 - \frac{1}{12{,}720}) \approx 0.079$, meaning the variance is roughly equal to the mean.

We did not observe any complications arising from the negative additive genetic correlation between the two traits. However, the bias introduced by phenotypic selection of plus trees, as identified in our study, is significant and warrants discussion within the tree breeding community. Typically, recurrent tree breeding programs begin by selecting plus trees, i.e. superior phenotypes, from even-aged unimproved stands. Breeders initially rely solely on individual phenotypic measurements, followed by systematic progeny testing to predict breeding values. They assume the selection is absent due to the lack of phenotypic and genotypic data from the broader founder population, i.e. natural stands (White et al. 2007). Hence during the genetic evaluation, the plus-trees are erroneously considered a random sample from the founder population. Our findings are relevant to all forest tree breeding programs, not limited to those utilizing random mating. It is important to highlight that bias introduced by the initial sampling of parents (plus trees) persists in the subsequent breeding cycles, which involves genetic evaluations across multiple generations. The extend to which this bias is amplified or mitigated through additional rounds of selection warrants further investigation.

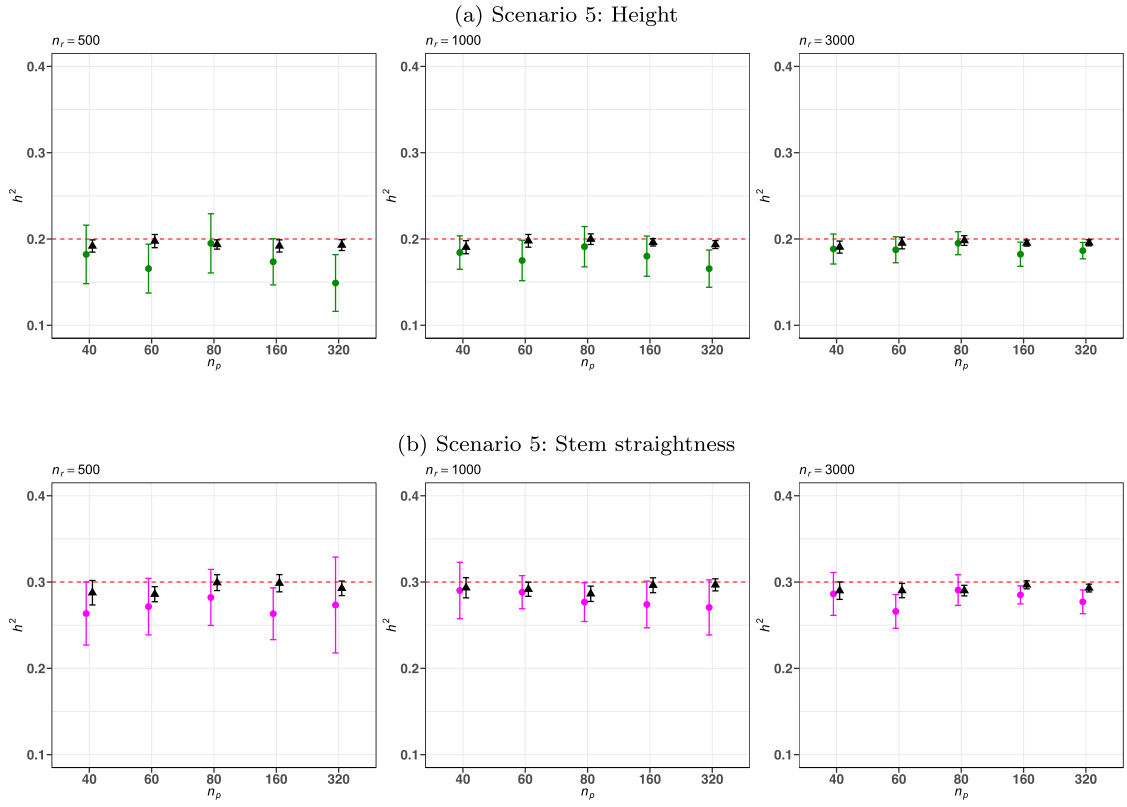


**Figure 5.** Scenario 5 "Combined". The *X*-axis indicates the number of plus trees ($n_p$), and the Y-axis displays the true narrow-sense heritability $h^2$ (black triangles with 95% CIs calculated across 30 replications) alongside its respective estimate (green and pink dots for height and stem straightness traits). The red dashed line indicates the initial $h^2$ values employed to generate the founder population (simulation input). The upper set of three graphs (a) illustrates height, while the lower set (b) depicts the stem straightness trait. The leftmost graphs correspond to a sample size of 500 ($n_r$), the center graphs to $n_r = 1000$, and the rightmost graphs to $n_r = 3000$. (a) Scenario 5: Height and (b) Scenario 5: Stem straightness.

The rationale for the reduction of $h^2$ following the phenotypic selection of parents is outlined by Falconer and Mackay (1996, p. 201–202); the effect of selection on additive genetic variance is proportional to the $h^2$ and the selection intensity. While this relationship is generally known and not in itself a concern, the problem arises in real tree breeding scenarios where this initial selection is overlooked due to lack of data on the wider founder population.

A potential remedy could involve genotyping and phenotyping a larger segment of the founder population, using the **G** matrix in genetic evaluations, which can effectively capture non-additive genetic variance (Muñoz et al. 2014; Gamal El-Dien et al. 2016). Alternatively, one could address the issue by altering the distributional assumptions of the genetic evaluation model (Tempelman 1998). A third strategy, as proposed by Lstibůrek et al. (2018), involves comparing the parentage of a truncated subset of the offspring with that of the corresponding truncated subset of the parental population. Unless such measures were implemented, the inherent bias will persist, and breeders will only be able to ascertain its negative direction, without a clear understanding of its magnitude.

The bias identified in this study has significant practical implications for forest tree breeding, with parallels to issues reported in other forest tree breeding contexts (Lu et al. 1999). When plus trees are selected solely on phenotype without incorporating broader information from the founder population, additive genetic variance tends to be underestimated, impairing the accuracy of breeding value predictions. This negatively impacts selection decisions and slows the rate of genetic gain. As the bias accumulates over successive generations, the overall effectiveness of breeding programs is compromised. Furthermore, overlooking the effects of dominance variance, exacerbated by the variability in family sizes under open-pollination systems, further impairs genetic evaluations, potentially resulting in suboptimal selection.

The findings of this study are particularly significant for tree breeding in Nordic countries, where the regeneration of commercial forests depends on open-pollinated seed pools from existing seed orchards comparable in

size to those examined here. However, since many breeding programs in other countries focus on coniferous tree species, where BwB could be effectively incorporated, the results of our study may have broader relevance within conifer breeding. The consistency of results across a wide range of parameters indicates that the implementation of BwB methods is likely to be robust, effective, and feasible within the region. The BwB approach enables the selection of superior trees directly from existing commercial plantations, while also managing the genetic diversity, enhancing the quality of the seed orchard as a source of genetic material, and assessing the extent of pollen and seed contamination. This comprehensive method holds promise for improving tree breeding outcomes.

## Conclusion

In conclusion, considering forest tree breeding programs dependent on random mating schemes, we confirmed that the current operational sizes of parental and offspring populations align with previous recommendations and should yield reasonably precise estimates of $h^2$. However, we observed a significant bias in $h^2$ due to dominance, which can be addressed by increasing the offspring sample size and reducing the number of parents. More concerning is the observation that phenotypic selection introduces a significant downward bias in $h^2$, necessitating further investigation due to its broad implications for the majority of tree breeding schemes.

In our future research, we plan to focus on the effects of epistasis and genotype-by-environment interactions on the accuracy of $h^2$. Additionally, we will explore alternative methods to utilizing genetic relationships. Further research is necessary to address the bias associated with phenotypic plus tree selection. The MoBPS software can handle real genomic data as input, enabling more accurate characterization of the genetic architecture of the studied traits (Pook et al. 2020). In forest trees, such approach is currently limited by huge genome sizes and still relatively sparse DNA marker coverage.

## Acknowledgments

The authors extend their gratitude to J. Stejskal, and V. Poupon for their valuable advice on the genetic evaluation aspects of this study. We also thank the anonymous reviewers for their valuable comments and suggestions, which helped improve the quality of this manuscript.

## Disclosure statement

No potential conflict of interest was reported by the author(s).

## Funding

The research leading to these results has received funding from the EEA/Norway Grants 2014–2021 and the Technology Agency of the Czech Republic, the Nordic Forest Research funds 2022–2024 (SNS-129), and The Swedish Foundation for Strategic Research (SSF). The authors extend their sincere gratitude to the Faculty of Forestry and Wood Sciences at the Czech University of Life Sciences in Prague for providing funding through an internal project grant (IGA No 1312).

## References

Butler DG, Cullis BR, Gilmour AR, Gogel BJ, Thompson R. 2017. ASReml-R reference manual version 4. Hemel Hempstead, HP1 1ES, UK: VSN International Ltd. https://vsni.co.uk/.

El-Kassaby YA, Lstibůrek M. 2009. Breeding without breeding. Genet Res. 91(2):111–120. doi: 10.1017/S001667230900007X

El-Kassaby YA, Lstibůrek M, Liewlaksaneeyanawin C, Slavov GT, Howe GT. 2007. Breeding without breeding: approach, example, and proof of concept. In: Isik F, editor. Proceedings of the IUFRO Division 2 Joint Conference: Low Input Breeding and Conservation of Forest Genetic Resources; 9-October 13, 2006; Antalya, Turkey; p. 86–92.

Falconer DS, Mackay TFC. 1996. Introduction to quantitative genetics. 4th ed. Essex, England: Longman.

Faux A-M, Gorjanc G, Chris Gaynor R, Battagin M, Edwards SM, Wilson DL, Hearne SJ, Gonen S, Hickey JM. 2016. AlphaSim: software for breeding program simulation. Plant Genome. 9(3):2. doi: 10.3835/plantgenome2016.02.0013

Gamal El-Dien O, Ratcliffe B, Klápště J, Porth I, Chen C, El-Kassaby YA. 2016. Implementation of the realized genomic relationship matrix to open-pollinated white spruce family testing for disentangling additive from nonadditive genetic effects. G3-Genes Genom Genet. 6(3):743–753. doi: 10.1534/g3.115.025957

Gaynor RC, Gorjanc G, Hickey JM. 2021. AlphaSimR: an R package for breeding program simulations. G3-Genes Genom Genet. 11(2):1–5. doi: 10.1093/g3journal/jkaa017

Gupta AK, Nadarajah S. 2004. Handbook of beta distribution and its applications. CRC Press.

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. doi: 10.1007/s11295-009-0253-6

Henderson CR. 1953. Estimation of variance and covariance components. Biometrics. 9(2):226–252. doi: 10.2307/3001853

Henderson CR. 1984. Applications of linear models in animal breeding. Guelph, Canada: University of Guelph.

Heuertz M, De Paoli E, Källman T, Larsson H, Jurman Ir, Morgante M, Lascoux M, Gyllenstrand N. 2006. Multilocus patterns of nucleotide diversity, linkage disequilibrium and demographic history of Norway spruce [Picea abies (L.) Karst]. Genetics. 174(4):2095–2105. doi: 10.1534/genetics.106.065102

Jahufer MZZ, Luo D. 2018. DeltaGen: A comprehensive decision support tool for plant breeders. Crop Sci. 58(3):1118–1131. doi: 10.2135/cropsci2017.07.0456

Kalinowski ST, Taper ML, Marshall TC. 2007. Revising how the computer program CERVUS accommodates genotyping

error increases success in paternity assignment. Mol Ecol. 16(5):1099–1106. doi: 10.1111/mec.2007.16.issue-5

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. doi: 10.1007/s001220100627

Larsson H, Källman T, Gyllenstrand N, Lascoux M. 2013. Distribution of long-range linkage disequilibrium and Tajima's D values in Scandinavian populations of Norway spruce (Picea abies). G3-Genes Genom Genet. 3(5):795–806. doi: 10.1534/g3.112.005462

Lstibůrek M, Bittner V, Hodge GR, Picek J, Mackay TFC. 2018. Estimating realized heritability in panmictic populations. Genetics. 208(1):89–95. doi: 10.1534/genetics.117.300508

Lstibůrek M, El-Kassaby YA, Skrøppa T, Hodge GR, Sønstebø JH, Steffenrem A. 2017. Dynamic gene-resource landscape management of Norway spruce: combining utilization and conservation. Front Plant Sci. 8(1810):1–6. doi: 10.3389/fpls.2017.01810

Lstibůrek M, Hodge GR, Lachout P. 2015. Uncovering genetic information from commercial forest plantations – making up for lost time using "breeding without breeding". Tree Genet Genomes. 11(55):1–12. doi: 10.1007/s11295-015-0881-y

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

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. doi: 10.1007/s11295-012-0472-0

Lstibůrek M, Schueler S, El-Kassaby YA, Hodge GR, Stejskal J, Korecký J, Škorpík P, Konrad H, Geburek T. 2020. In Situ genetic evaluation of European larch across climatic regions using marker-based pedigree reconstruction. Front Genet. 11:28. doi: 10.3389/fgene.2020.00028

Lu PX, Huber DA, White TL. 1999. Potential biases of incomplete linear models in heritability estimation and breeding value prediction. Can J Forest Research. 29(6):724–736. doi: 10.1139/x99-047

Marshall TC, Slate JBKE, Kruuk LEB, Pemberton JM. 1998. Statistical confidence for likelihood-based paternity inference in natural populations. Mol Ecol. 7(5):639–655. doi: 10.1046/j.1365-294x.1998.00374.x

Misztal I, Tsuruta S, Lourenco DAL, Aguilar I, Legarra A, Vitezica Z. 2014. Manual for BLUPF90 family of programs. University of Georgia; Athens, USA. https://nce.ads.uga.edu/html/projects/programs/docs.

Muñoz PR, Resende Jr MFR, Gezan SA, Deon Vilela Resende M, de Los Campos G, Kirst M, Huber D, Peter GF. 2014. Unraveling additive from nonadditive effects using genomic relationship matrices. Genetics. 198(4):1759–1768. doi: 10.1534/genetics.114.171322

Pook T, Schlather M, Simianer H. 2020. MoBPS-modular breeding program simulator. G3-Genes Genom Genet. 10(6):1915–1918. doi: 10.1534/g3.120.401193

R Core Team. 2023. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Sargolzaei M, Schenkel FS. 2009. QMSim: a large-scale genome simulator for livestock. Bioinformatics. 25(5):680–681. doi: 10.1093/bioinformatics/btp045

Skrøppa T, Mørtvedt Solvin T, Steffenrem A. 2023. Diallel crosses in Picea abies IV. Genetic variation and inheritance patterns in short-term trials. Silvae Genet. 72:58–71. doi: 10.2478/sg-2023-0006

Tempelman RJ. 1998. Generalized linear mixed models in dairy cattle breeding. J Dairy Sci. 81(5):1428–1444. doi: 10.3168/jds.S0022-0302(98)75707-8

Ten Napel J, Vandenplas J, Lidauer M, Stranden I, Taskinen M, Mäntysaari E, Calus MP, Veerkamp RF. 2020. MiXBLUP user's guide. Animal Breeding & Genomics, Wageningen University & Research; Wageningen, The Netherlands. https://www.mixblup.eu.

Vidal M, Plomion C, Harvengt L, Raffin A, Boury C, Bouffier L. 2015. Paternity recovery in two maritime pine polycross mating designs and consequences for breeding. Tree Genet Genomes. 11:1–13. doi: 10.1007/s11295-015-0932-4

Wang J-K, Wolfgang HP. 2007. Simulation modeling in plant breeding: principles and applications. Agric Sci China. 6(8):908–921. doi: 10.1016/S1671-2927(07)60129-1

White TJ, Adams WT, Neale DB. 2007. Forest genetics. Wallingford, UK: CAB International.