Journal of Forestry Research (2026) 37:172
https://doi.org/10.1007/s11676-026-02108-w

ORIGINAL PAPER

Cross-generational decline in genomic selection accuracy in Norway spruce (*Picea abies* (L.) H. Karst) and strategies to mitigate it

Edward A. Carlsson[1,2] · Henrik R. Hallingbäck[3] · Jon Ahlinder[4] · Mari Suontama[4] · Harry X. Wu[1]

Received: 27 March 2026 / Accepted: 9 July 2026
© The Author(s) 2026

Abstract Genomic selection (GS) could be used to reduce the long cycle time for tree breeding. This assumes that marker-based predictions are sufficiently accurate without phenotypes. We evaluated GS across two linked generations of Norway spruce (*Picea abies* (L.) H. Karst), the first consisting of 954 plus-trees (G0) and the second of 956 progeny trees representing 34 full-sib families (G1), using 16 clonal field trials across mid- and southern Sweden. Both generations were measured for height and genotyped using a 50 K SNP chip array. Cross-validations within and across generations were compared. GS efficiency was evaluated using global prediction accuracy and within-family predictive ability, using GBLUP with phenotypes for independent validation. We used computer simulations to emulate the experimental data and repeat the same analysis under different assumptions of effective population size. Additional simulations were performed to investigate the cross-generation GS accuracy in future generations. Simulations assuming a historical effective population size of 1000 or 5000, together with experimental results, indicate that prediction accuracy decreased by 49–76% for global prediction and by 15–65% for within-family prediction when G1 was predicted from G0, compared with cross-validation within the G1 generation. Increasing the relatedness at the expense of training set size increased global accuracy but decreased within-family accuracy. Simulations of advanced generations showed that training on multiple generations increases GS accuracy, both for global and within-family prediction. Access to multiple generations for training and/or a higher density of markers may be recommended to increase accuracy for cross-generation and within-family GS in conifers.

Keywords Genomic selection (GS) · Cross-generation · Norway spruce · AlphaSimR · Within-family prediction

Corresponding editor: Lei Yu

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s11676-026-02108-w.

✉ Harry X. Wu
Harry.wu@slu.se

[1] Department of Forest Genetics and Plant Physiology, Umeå Plant Science Centre (UPSC), Swedish University of Agricultural Sciences, 901 83 Umeå, Sweden

[2] Forestry Research Institute of Sweden (Skogforsk), Ekebo 2250, 268 90 Svalöv, Sweden

[3] Forestry Research Institute of Sweden (Skogforsk), Uppsala science park, 751 83 Uppsala, Sweden

[4] Forestry Research Institute of Sweden (Skogforsk), P.O. Box 3, 918 21 Sävar, Sweden

Introduction

The methodology of genomic selection (GS), introduced by Meuwissen et al. (2001), involves the use of genome wide markers, such as single-nucleotide polymorphisms (SNPs), to predict genomic estimated breeding values (GEBVs). GS overcomes the challenges of early marker-based selection by assuming that a trait is controlled by a large number of small genetic effects, which are all estimated simultaneously. Using tens of thousands of SNPs across the whole genome, GS relies on the linkage disequilibrium (LD) between markers and quantitative trait loci (QTL). Selection based on GEBVs reduces the need for phenotyping and allows the breeders to make selections earlier, potentially increasing the genetic gain per unit of time. GS has been successfully implemented in dairy cattle and other domestic animals (Hayes et al. 2009; Meuwissen et al. 2016) where the time

Published online: 01 August 2026

for selection for milk production has decreased from seven years with traditional breeding, to twelve months. Consequently, GS has doubled the rate of genetic gain over the last decade (García-Ruiz et al. 2016; Xu et al. 2020).

Conifer breeding programs generally have prolonged breeding cycles, which have motivated research into the implementation of GS in forest trees, including simulation and experimental studies (Grattapaglia et al. 2018; Lebedev et al. 2020; Beaulieu et al. 2022). These research articles suggest that GS can increase genetic gain by shortening the breeding cycle and by capturing the variation from Mendelian segregation. This last point is vital for selection of top breeding candidates within families (Papin et al. 2024). Typically, GS efficiency is evaluated in terms of the accuracy of prediction models on trees without phenotype data, using cross-validation. Studies on GS in conifers have indicated that early selection at the seedling stage and omitting field testing is possible without loss of genetic gain (Resende et al. 2012; Beaulieu et al. 2014; Ratcliffe et al. 2015; Bartholomé et al. 2016; Thistlethwaite et al. 2017; Chen et al. 2018, 2019; Calleja-Rodriguez et al. 2020). For Nordic tree species such as Norway spruce (*Picea abies* (L.) H. Karst)) and Scots pine (*Pinus sylvestris* L.) it has been shown that 86–107% of the gain from traditional selection can be achieved using GS, based on exome sequencing, 50 K SNP array genotyping and genotyping-by-sequencing (GBS) (Calleja-Rodriguez et al. 2019; Chen et al. 2018, 2023).

The Norway spruce breeding cycle in Sweden requires around 25 years (Rosvall 2019). Combined with clonal forestry, GS could accelerate the deployment of the elite clones at the seedling or embryonic stage by somatic mass-propagation (Wu 2019). Assuming that flowering can be initiated in young trees, implementing GS could potentially shorten the breeding cycle of Norway spruce to merely 8–11 years. This would enable tree breeders to complete two to three generations of selective breeding within the time frame of one conventional breeding cycle.

However, most studies on GS efficiency in trees have been conducted using cross-validation within the same generation. In operational tree breeding programs, a GS model based on the phenotypic and genotypic data of parents and their relatives would be used to predict breeding values or total genetic values of their progenies for selection purposes (Fig. 1). Therefore, both selection effects (e.i., only a subpopulation of elite trees will be used for the next generation) and mating effects (e.i., new trees will be produced from crosses among the selected elite trees) on GS efficiency need to be studied using trees across generations, before implementation in operational breeding.

Few publications on GS in conifers have performed cross-generation validations using two or more generations. In maritime pine (*Pinus pinaster* Aiton), estimated breeding values (EBVs) from grandparental (G0) and parental (G1) trees were used to train models to predict breeding values of

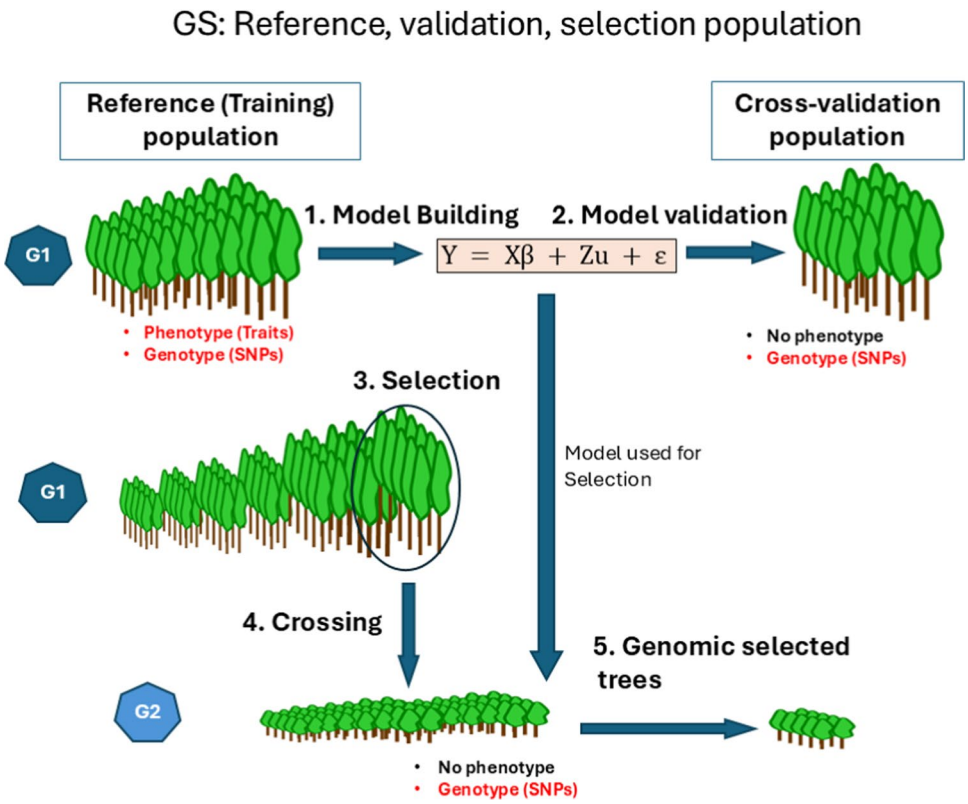


Fig. 1 Conceptual description of genomic selection in tree species. Step (1) genetic effects are estimated using a model with a reference population which is both phenotyped and genotyped; (2) the accuracy of the model is validated on a separate set of trees; (3) selection of the best genetic material is carried out; (4) selected trees are crossed to generate seed for the next generation; step (5) previously developed model is used to make early selection based on genomic estimated breeding values (GEBVs) before the trees can be phenotyped

a G2 generation for stem size, straightness and tree height (Bartholomé et al. 2016). It was observed that the combined average accuracy of all three traits using genomic best linear unbiased prediction (GBLUP) was higher across generations, using G0 and G1 to predict G2, than cross-validated GBLUP accuracy within the G2 generation. Two GS studies of Douglas-fir (*Pseudotsuga menziesii* (Mirbel) Franko) with shared sites and genetic material jointly suggest that accuracy of GBLUP across generations was higher than the within-generation cross-validation accuracy for juvenile height (Thistlethwaite et al. 2017; 2019). These results are surprising for at least two reasons. First, recombination between generations is expected to weaken LD between QTL and markers, reducing GS accuracy (Habier et al. 2007; Xu et al. 2020). Secondly, directional selection of elite trees within a generation will likely change the allele frequencies in the progeny generation (Wientjes et al. 2022). GBLUP implicitly assumes a common allele frequency distribution for the population, and deviations from that will affect the estimated genomic relationship (Isidro y Sánchez and Akdemir 2021). A common pitfall of validating accuracy of genomic predictions is the failure to ensure that the training and validation populations are independent. Estimated breeding values (EBVs) derived from pedigree (ABLUP) with a large number of measurements have commonly been used for validation (Bartholomé et al. 2016; Lenz et al. 2017; Chen et al. 2018; Klápště et al. 2018). If there is an overlap in the phenotype information used to estimate the GEBVs and EBVs, this may inflate the GS accuracy (Amer and Banos 2010). This method also treats the EBVs from ABLUP as true breeding values (TBVs), a simplification that is difficult to verify since the TBV are never known. Computer simulation is a useful tool in this respect, as the breeder gets access to TBV and can compare different strategies. Many studies have already made use of simulations to make projections about future trends in tree breeding (Iwata et al. 2011; Hall et al. 2016; McLean et al. 2023; Papin et al. 2024). Ideally, the results from simulation should be validated against experimental results.

We evaluated GS efficiency for cross- and within-generation prediction in Norway spruce. We first conducted an experimental analysis on a founder generation (G0) and its progeny generation (G1), with a total of 1910 clones. The trees were genotyped with a 50 k SNP array and measured for height growth across 16 clonal field trials. Different cross-validation schemes were conducted to investigate GS accuracy both within and across generations. GEBVs were predicted using GBLUP. The experimental dataset was replicated using stochastic simulations and analyzed with similar methodology. An additional set of simulations were carried out for up to six generations to explore future trends. The objectives of this study were to (i) compare the accuracy of within-generation GS to cross-generation GS using independent validation, (ii) investigate if increasing the relatedness between training and validation populations can increase GS accuracy across generations, (iii) to explore projected changes in cross-generation GS accuracy when multiple generations are used for training. Throughout the study, the accuracy of GS models was evaluated using a population-based genomic model for breeding selection (global selection) for within- and across-generation and for within-family clonal selection.

Materials and methods

Plant material

Tree data was collected from the Norway spruce clonal breeding trials owned by the Forestry Research Institute of Sweden (Skogforsk). A total of 16 clonal field trials were used, located in mid-southern Sweden (Fig. 2). Thirteen clonal trials were established between 1982 and 1993 with a total of 1193 plus-trees propagated from seed-lots collected in Slovakia and Belarus (G0), vegetatively propagated using rooted cuttings. Based on seed-lot information of geographical origin, three separate field trial series were established with either mainly Belarusian (1), a mix of Belarusian and Slovakian (2), or mainly Slovakian clones (3). The clones of each series were planted across three to five field trials with varying site characteristics, so that all trials within the same series had roughly the same genetic material (Table 1). The number of planted ramets per clone for each trial was ten, ten, and six for series 1, 2 and 3 respectively. A subset of 30 superior clones was selected as parents based on field trial performance and crossed to create 44 full-sib families (G1) with an average size of 28 (SD = 11). The G1 clones were planted in a new series of clonally propagated field trials in 2014 and 2015 at three new sites, with three to four ramets per clone in each trial.

At each field trial, tree height was measured over multiple years. Measurements up to the age of nine years were made with a height stick to the centimeter precision. For ages ten and above, measurements were made with a Vertex height measurement device at decimeter precision. Each G1-field trial was measured once, at ages of seven, eight or nine years. For the G0-trials, measurements from multiple years were available, and a single year was selected for each trial with the aim of choosing a similar assessment age among the G1 trial series.

Spatial analysis of individual trials

Forest tree trials are subject to microsite environmental effects, such as variation in local topology, soil conditions and the direction of planting (Fagroud and Van

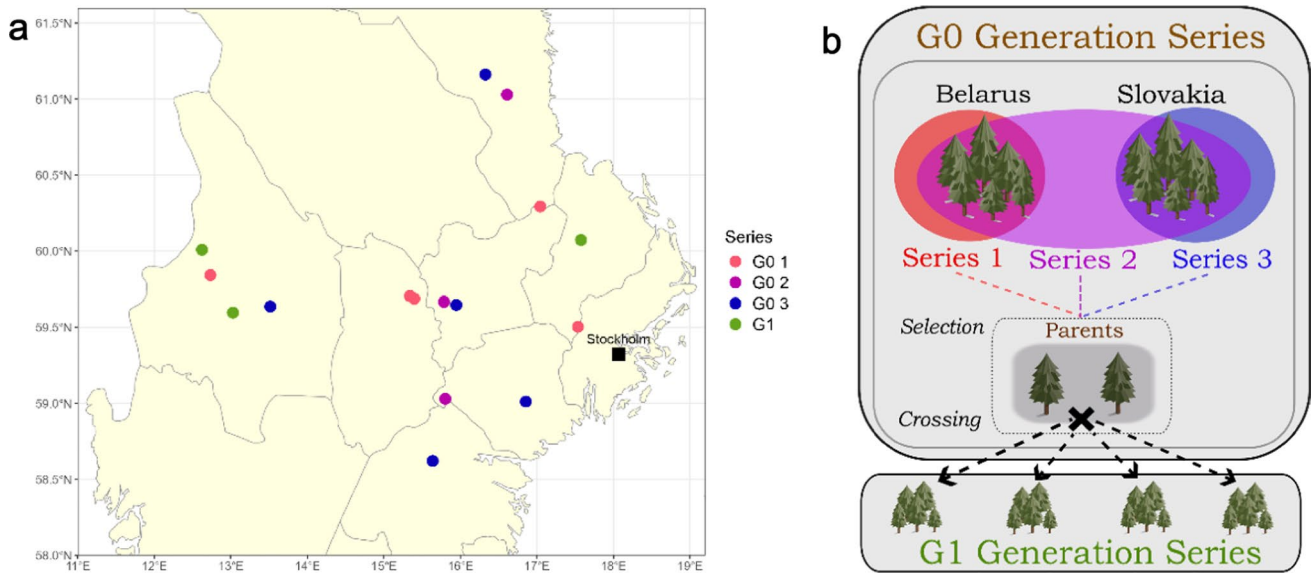


Fig. 2 Overview of field trial locations and clones from different generations. **a** Locations of the 16 field trials in mid-southern Sweden, colored by trial series. **b** Illustration of generations and origins for all clones included. The "Parents" group includes clones among all G0 series that were selected and crossed to produce the G1 generation

Table 1 Overview of field trial data. "All trees" refer to all available trees within each trial. "Genotyped" refer to the subset of trees that had SNP marker data. Standard deviations are in parentheses

Trial name	Gen	Series	Year	Age	Measurements		Clones		Growth rate (cm / year)	
					All trees	Genotyped	All trees	Genotyped	All trees	Genotyped
Salungen	G0	1	1990	8	3036	2187	481	415	33.0 (9.1)	32.9 (9.1)
Dunderbo	G0	1	1990	8	2760	2027	481	415	27.9 (9.1)	27.9 (9.1)
Ålbrunna	G0	1	1992	8	1234	646	373	209	16.7 (5.7)	16.5 (5.7)
Morskoga	G0	1	1992	8	1970	976	379	210	16.9 (6.6)	16.5 (6.5)
Hade gravfält	G0	1	1993	8	1825	880	372	206	13.0 (4.4)	12.8 (4.4)
Vingåker	G0	2	1983	15	1359	890	222	162	29.1 (11.6)	29.1 (11.5)
Jönsarbo20	G0	2	1983	15	1529	1002	222	162	42.2 (11.0)	42.3 (11.4)
Lingbo	G0	2	1983	15	1030	667	220	160	11.5 (4.4)	11.5 (4.3)
Finspång	G0	3	1982	10	2520	1708	490	377	23.3 (11.0)	23.3 (10.9)
Frändesta	G0	3	1982	10	2485	1640	487	375	21.0 (7.7)	21.0 (7.7)
Forshaga	G0	3	1982	10	2668	1795	489	376	22.6 (9.9)	22.8 (9.8)
Jönsarbo08	G0	3	1982	10	2598	1745	487	401	17.3 (5.8)	17.2 (5.7)
Kilafors	G0	3	1982	10	3092	2076	487	377	17.3 (7.2)	17.5 (7.3)
Savolax	G1		2014	8	3973	2875	1255	943	43.7 (10.5)	44.1 (10.4)
Malmen	G1		2014	7	3876	2840	1247	953	44.8 (10.0)	45.3 (9.9)
Bunäset	G1		2015	9	5313	3655	1333	913	39.6 (10.9)	39.9 (10.7)

Meirvenne 2002; Bian et al. 2017). Adjustments for heterogenous residual variation for tree height were performed for all measurements within each trial. The adjustments were performed on each trial separately (including non-genotyped trees), as fitting a single model across all trials was considered too computationally demanding. A linear mixed model was fitted with one spatially independent residual term and a spatially dependent residual term using two first-order autoregressive (AR1) terms in the row and column direction (Dutkowski et al. 2002; 2006; Gilmour et al. 2021). Spatially dependent residual variance was removed from each measurement, creating a new set of spatially adjusted phenotypes used for downstream

analysis (Calleja-Rodriguez et al. 2019). Details on the spatial analysis can be found in Supplementary Methods.

Genotype data

Buds and needles were sampled from 2073 individual trees, corresponding to 2034 clones from the trials Jönsarbo08, Jönsarbo20, Dunderbo and Malmen, between June and November of 2020. Immediately after sampling, the plant tissue was kept in a freeze box with ice packs, and later at -20℃ for long term storage. The DNA was extracted at the Umeå Plant Science Center (UPSC) using the E.Z.N.A SP Plant DNA Kit where some steps of the protocol were modified (Supplementary Methods). Genotyping was performed using the Norway spruce Piab50K SNP array chip (Bernhardsson et al. 2021).

Merging of individual genotyping calls and SNP filtering was performed with the Axiom Analysis Suite software (v.5.3) using the default "Genotyping" workflow. SNPs were filtered for a call rate of minimum 0.97 and conversion type of "PolyHighResolution", "NoMinorHom" or "MinorHighResolution". Pedigree correction and sample filtering were carried out on the original 2073 samples using custom R scripts, version 4.3.3 (R Core Development Team 2024) (Supplementary Methods). A second SNP filtering was performed in Axiom Analysis Suite using a final subset of 1910 unique clones. Next SNPs with a minor allele frequency below 0.01 were removed, retaining 39 710 SNPs for downstream analysis.

Genomic prediction

The marker matrix of SNPs was coded as −1, 0 and 1 and converted to a G-matrix with the A.mat function of the R package rrBLUP (Endelman 2011). Missing values were imputed using the "means" option. Since the number of missing marker calls was only 0.8%, more advanced imputation methods were not pursued. A single G-matrix was generated for both generations and used for all downstream experimental analyses. To ensure stable invertibility, the original G-matrix was blended by adding an identity matrix with identical dimensions multiplied by 0.01 and was converted to a sparse format using the G.inverse function of the ASRGenomics R-package (Gezan et al. 2022).

Population structure was investigated by comparing the G-matrix and the expected pedigree-based relationship, and a PCA analysis using the stats R-package (R Core Development Team 2024).

A linear mixed model was used on genotyped trees to estimate variance components and predict GEBVs with ASReml-R (Butler et al. 2023):

$$\boldsymbol{y_{adj}} = \boldsymbol{X\beta} + \boldsymbol{Zu} + \boldsymbol{\varepsilon} \tag{1}$$

where $\boldsymbol{y_{adj}}$ is a vector of spatially adjusted height growth rate, $\boldsymbol{\beta}$ is a vector of fixed effects, including overall mean and site effects, and $\boldsymbol{u}$ is a vector of random effects distributed as $\sim \boldsymbol{N}(0, \sigma_A^2\boldsymbol{G})$ where σ_A^2 is the additive genetic variance and G is the marker-based realized genomic relationship matrix. $\boldsymbol{X}$ and $\boldsymbol{Z}$ are incidence matrices for fixed and random effects respectively. The residual error term $\boldsymbol{\epsilon}$ was distributed as $\boldsymbol{\varepsilon} \sim \boldsymbol{N}(0, \boldsymbol{R})$, with:

$$\boldsymbol{R} = \begin{pmatrix} \sigma_{e_1}^2 \boldsymbol{I}_{n_1} & 0 & \dots & 0 \\ 0 & \sigma_{e_2}^2 \boldsymbol{I}_{n_2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma_{e_S}^2 \boldsymbol{I}_{n_S} \end{pmatrix} \tag{2}$$

where $\sigma_{e_S}^2$ is the residual variance for site S, and $\boldsymbol{I}_{n_S}$ is an identity matrix with dimensions n equal to the number of measurements in site S. The solutions for fixed and random effects were obtained by solving the mixed model equation according to Henderson (1984) (see Supplementary methods).

Site-specific narrow-sense heritability for the 16 experimental trials was estimated as:

$$\widehat{h}_s^2 = \frac{\widehat{\sigma}_A^2}{\left(\widehat{\sigma}_A^2 + \widehat{\sigma}_{e_S}^2\right)} \tag{3}$$

Standard errors of the heritability estimates were obtained using the methods of Ahlinder et al. (2024). The model for estimating variance components included 1910 genotyped clones from both generations. Predictive ability (PA) was estimated as the Pearson correlation between the GEBVs and phenotypes. Prediction accuracy (PACC), defined as the correlation between GEBVs and TBV, was estimated using the methods proposed by (Legarra et al. 2008):

$$PACC = \widehat{r}(TBV, GEBV) = \frac{PA}{\sqrt{\widehat{h}_s^2}} \tag{4}$$

As all predictions were made for G1 clones, only heritabilities from Savolax, Malmen and Bunäset were used to calculate PACC.

The same model for global prediction was used for within-family prediction, referring to picking the top-ranking siblings in a full-sib family. GS studies on forest trees rarely report accuracy within families, likely due to the difficulties in estimating family-specific variance components. In this case, model performance was only expressed in terms of PA. This was done in two steps. First, PA was calculated for each individual full-sib family. Next, we took the average PA across all families to express a general within-family PA.

Cross-validation schemes

Four cross-validation scheme were set up for prediction of genotyped G1 clones (Fig. 3). Each scheme consisted of sampling and prediction. We used the caret R-package (Kuhn 2008) to generate five folds of genotyped clones from the specified "sampling set". Four out of five folds would then be used as a training set. This was repeated five times, each time excluding a new fold from training. Each cross-validation was repeated 20 times to generate 100 iterations per scheme. By sampling based on clone id, we ensured that measurements of the same clone were never present in the training and validation set simultaneously. During prediction, the training set was used in a GBLUP model to predict GEBVs, which were used to predict the phenotypes in a validation set. We used two Cross-Generation schemes which both used the whole G1 generation as a fixated validation set. The first sampling set consisted of the whole G0 generation (CG1). The second sampling set consisted of the G0 clones which had been selected for crossing (parents) plus any G0 clone with a genomic relationship coefficient of minimum 0.05 with at least one parent according to the G-matrix (CG1R). We used two Within-Generation schemes. In the first scheme, we sampled from the G1 generation and validated against the excluded G1 clones (WG1). For the second scheme, we also added the whole G0 generation to the training set (WG01).

Simulating data

Additional datasets were generated using stochastic simulations with the R-package AlphaSimR (Gaynor et al. 2021), aiming to recreate the experimental study setup. Twelve chromosome pairs were generated using the Markovian coalescent simulator (MaCS) (Chen et al. 2009), each with a physical distance of 1.63 Gb and genetic distance of 2.96 M (Nystedt et al. 2013; Bernhardsson et al. 2019). A previous study on Norway spruce used a mutation rate of 2.2×10^{-9} (Nystedt et al. 2013), expressing the probability that a new mutation occurs per locus per meiosis per year. Here, this number was multiplied by 45 as an approximation of the age at which new seeds were created in wild stands. The initial trait "genetic mean" and variance were set to 200 and 3600 respectively to mimic height in Norway spruce at age ten in centimeters, as an approximation based on the G0 trials from the experimental data. For simplicity, the simulated trait was only controlled by additive genetic effects. The total number of segregating loci per chromosome was arbitrarily set to one million, as a compromise between realism and ease of computation. For each chromosome, 42 segregating sites were selected as quantitative trait nucleotides (QTNs) (Hall et al. 2016). QTN effects were randomly sampled from a gamma distribution with shape = 0.5 (Denis and Bouvet 2011; Li and Dungey 2018).

The historic genetic diversity of the breeding population was simulated with three separate scenarios where the initial effective population- (N_e) and census size were identical, either 100, 1000 or 5000. Each founder population went through 100 generations of burn-in to generate more realistic patterns of LD. Each generation cycle, individuals were randomly crossed with the number of crosses equal to the initial census size and generated 2 new progenies per cross, similarly to the methods of McLean et al. (2023). Next, 1000 individuals were randomly sampled from the last burn-in

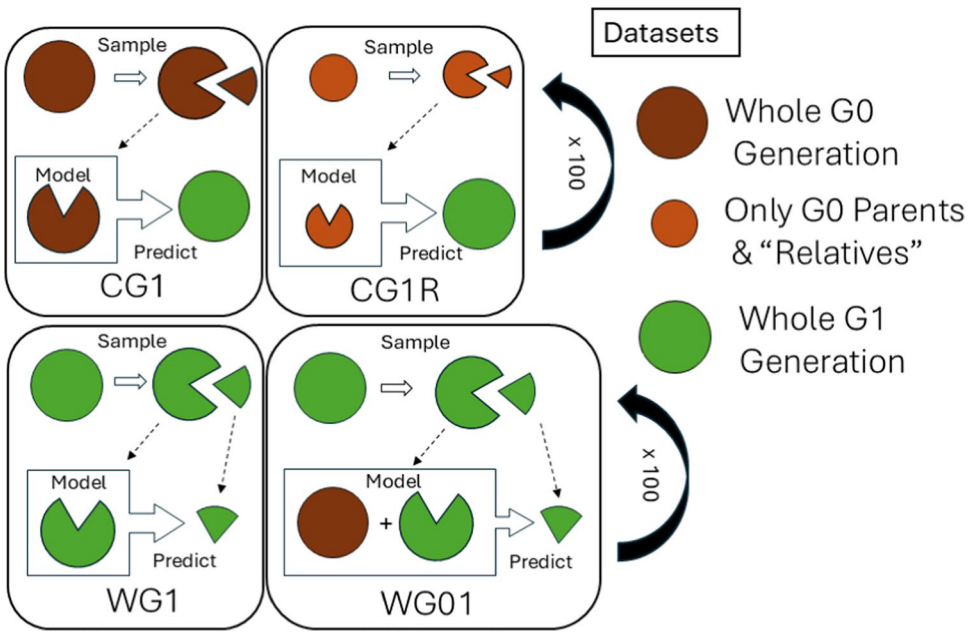


Fig. 3 Illustration of cross-validation schemes. The two Cross-Generation (CG) schemes both had the whole G1 generation as a fixated validation set (VS), while the training set (TS) was sampled from either the whole G0 generation (CG1) or the G0 clones used as parents and their close relatives identified by the G-matrix (CG1R). The Within-Generation (WG) schemes used a subset of G1 clones as VS, while the TS was either the remaining G1 clones (WG1) or the remaining G1 clones plus the whole G0 generation (WG01)

generation to represent the G0 population. In the N_e 100 scenario, this was achieved by performing an additional generation of burn-in with 10 progenies per cross.

To simulate clonal propagation, 20 phenotypic values were generated for each G0 clone using a heritability of 0.2. One tenth of G0 phenotype values were randomly removed afterwards to simulate varying number of ramets per clone. The mean phenotype per clone was used to select the top 30 parents, which were used to make 35 random crosses. The family sizes were set to emulate the experimental dataset. For each cross, the number of offspring were sampled from a normal distribution with a mean of 30 and standard deviation of 10, rounded to integers and the minimum possible number was 1, constituting the new G1 generation. Phenotypic values were generated for the G1 generation with a heritability of 0.2, but with ten values per clone. Again, one tenth of simulated G1 phenotypes were randomly removed.

All clones were genotyped using genomic marker information of a simulated SNP chip with a total of 40 464 biallelic markers, meaning 3372 per chromosome, which were sampled from the available segregating sites with a minimum MAF of 0.01, using the G0 generation as reference. The causal QTNs were excluded as candidate markers for sampling, under the assumption that perfect overlap between SNP markers and causal segregating sites is unlikely. Each N_e scenario was replicated 10 times, creating 30 simulated datasets of genotyped and clonally replicated phenotypic data from two generations.

The simulation setup was extended to study the efficiency of GS in more advanced generations. The same settings were used to generate the G0 and G1 generation as previously described, but only using the historic N_e of 1000. New selections were simulated among the G1 based on EBVs calculated using pedigree-based ABLUP with the kin.blup function the R-package rrBLUP (Endelman 2011). The pedigree information was reset at the G0 generation to replicate assumptions made in a typical breeding program, meaning that all founders were treated as completely unrelated. Two sets of simulations were generated where selection of new parents was based on either the 35 best clones overall (global selection) or the best clone from each of the 35 families (within-family selection). Crossing scheme, number of progenies per cross and phenotypic data were simulated as previously described for the G1 generation. Each new calculation of EBVs made use of all phenotypic and pedigree information available at that point in the simulation. This process was repeated to generate five successive full-sib generations, using either global-selection or within-family selection, plus the founder generation. Again, all clones were genotyped with a simulated SNP-array as before. The same simulated SNP chip was used for all subsequent generations, within the given iteration. This simulated advanced breeding cycle was again replicated 10 times for each of the two selection methods.

Genomic prediction on simulated data

Genomic prediction and cross-validation schemes were repeated for each simulated dataset similarly to the experimental data analysis, except no type of data preprocessing was done, as no sites or spatial patterns were part of the simulation. The model for estimating breeding values was a simplified version of Eq. 1, with only an overall mean as fixed effect and a single residual variance. Since only additive genetic effects were simulated, the total genetic value was treated as the TBV and "true prediction accuracy" was calculated as the Pearson correlation between the GEBVs and the TBV within the validation set. This was performed family-wise and then averaged across families for each iteration to calculate the within-family PACC. For comparison with the experimental results, within-family PA was estimated on the simulated data as well.

The first set of simulations with only two generations used the same cross-validation schemes as the experimental data. Essentially, this implies that a fivefold cross-validation was repeated 20 times per each of 10 simulated datasets for each cross-validation scheme, ultimately giving 1000 iterations per scheme. For the advanced generations, GS accuracy was only evaluated across generations, similarly to scheme CG1. In this case, the accuracy metric was the same as the selection method, i.e. within-family prediction accuracy for breeding programs based on within-family selection. Each new generation was used as a fixated validation set in separate analyses. The training set consisted of phenotypes sampled from either the one previous generation or all previous generations, relative to the given generation in the validation set. Similarly, the G-matrix was generated using markers from the validation generation and either the previous generation, or all previous generations.

Inspecting LD

LD was assessed on the simulated advanced generations using PLINK (Purcell et al. 2007), calculated as the squared correlation coefficient (r^2) between pairs of loci (Hedrick 2017). Separate analyses were made on each unique generation within all simulated iterations. All r^2 values were retained within a 1500 SNP window and maximum distance of 100 000 kb. The rate of LD decay was estimated by fitting a nonlinear regression model to the expected value of r^2 between adjacent sites (Hill and Weir 1988; Marroni et al. 2011). The model was used to find the distance at which the expected r^2 was 0.2, used as a benchmark to compare LD between different generations and selection methods.

Results

Evaluating genomic relationships and population structure with marker data

The marker-based G-matrix closely mirrored the expected pedigree-based relationship (Fig. S1). The G0 plus-trees displayed a modest population structure in the G-matrix on two levels. Firstly, trees that had previously been recorded as either from Belarus or Slovakia generally showed noticeably higher realized genomic relationship with clones of the same origin than between different origins. Secondly, some clones within the Belarusian G0 population formed more closely related clusters, with some pairwise relationships similar to half-sibs. Distributions of genetic relationship coefficients for the G1 clones displayed a left skewed peak in relation to the expected value from a pedigree-based matrix for half- and full-sib relations (Fig. 4). Among the pairwise relationships which were unrelated according to pedigree, genomic relatedness showed a second peak around 0.22. Values along the diagonal showed a bimodal distribution with a few outliers above 1.2. One full-sib family showed signs of unintentional inbreeding, as the parents had a genomic relationship close to that of two half-sibs.

The PCA plot showed weak population structure in our data (Fig. S2). The first PC (explaining 3.31% of the variation) roughly corresponded to the difference between the two historical populations and their hybrid offspring, while the second PC (explaining 2.26% of the variation) explained more of the within-population variation. The G1 clones formed several distinct clusters, likely corresponding to the different full-sib families.

Experimental GS accuracy

Site-specific narrow-sense heritability ranged from 0.056 to 0.375 with a mean of 0.171 (Table S1). For sites with G1 clones, used to estimate PACC, the heritability had a smaller range of 0.108 to 0.117. GS models predicting across generations had a substantially lower PACC compared to the within generations (Fig. 5a). For the cross-generation schemes, sampling from all available G0 data (CG1) gave an average PACC of 0.251. Retaining only G0 clones exhibiting a certain genomic relationship to any of the actual parents of the G1 (CG1R) increased the PACC to 0.313. When the G1 clones were cross-validated within the same generation (WG1), the mean PACC was 0.610. Adding all G0 clones to the training set (WG01) slightly increased mean PACC to 0.628. In effect, this means that including measurements from the G0 did not appreciably improve the global PACC of G1 clones compared to restricting both training and validation to the G1 generation. This was also the case when the training set comprised G1 clones together with the same subset of G0 clones used for CG1R (Fig. 5, Fig. S3). The relative performance of each scheme was similar in terms of global prediction regardless of whether PACC or PA was considered (Fig. 5a, Fig. S4), but the magnitude of PA was, as expected, substantially lower than the corresponding PACC.

Average PAs estimated within-family were close to zero in all cross-validation schemes, but always positive (Fig. 5b). Using the smaller but more related training set did not increase the within-family PA compared with the larger training set. Rather, the within-family PA of CG1R (0.033) was slightly lower than that of CG1 (0.041). Similarly to the global prediction estimates, the within-generation schemes had higher within-family PA, 0.049 and 0.061 for WG1 and

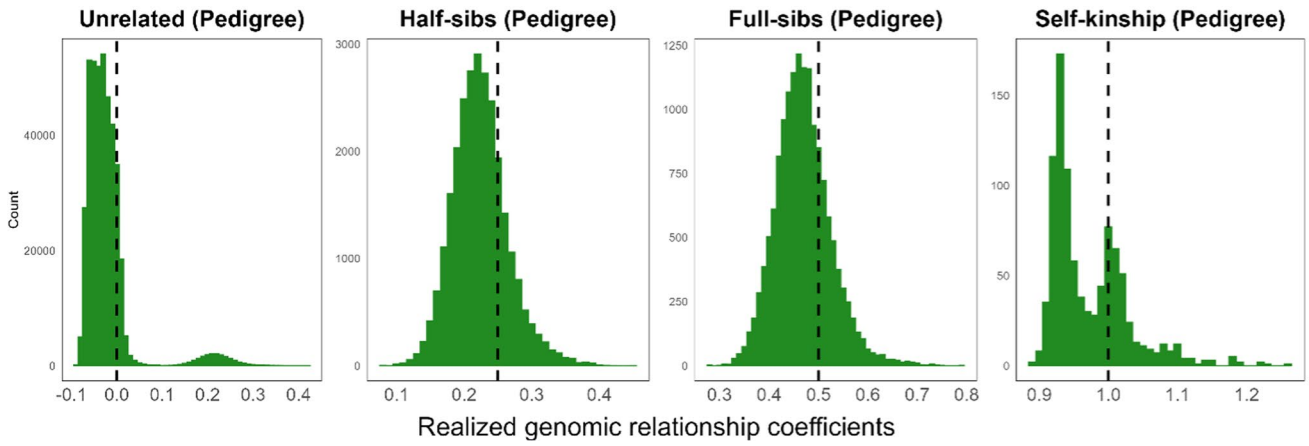


Fig. 4 Distribution of realized genomic relationship coefficients of clones in the G1 generation, indicated by green bar plots. Relationships were grouped by expected kinship from pedigree, indicated by the dashed horizontal lines. Self-kinship represents the diagonal elements of the G-matrix

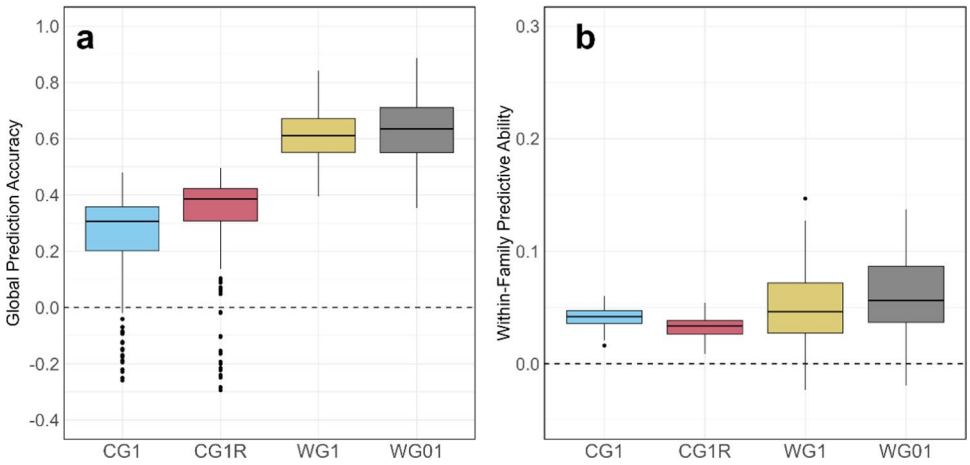


Fig. 5 Global prediction accuracy **a** and within-family predictive ability **b** of the experimental dataset. Each boxplot of global prediction accuracy has a sample size of 300. For within-family predictive ability the sample sizes are 100. Colors as well as labels below the boxplots correspond to cross-validation schemes described previously (see Fig. 3)

WG01 respectively, but values varied more between iterations. Individual PA estimates for specific families often fluctuated between large negative and positive estimates. For within-generation schemes, some iterations were unable to estimate PA for specific families, likely caused by too few siblings being represented in the validation set in that iteration.

Simulated GS accuracy

The experimental data was replicated by simulation, assuming three levels of historic effective population size (N_e). There was a similar trend between simulated and experimental data in terms of how the cross-validation schemes compared with each other (Fig. 6a–i). Cross-generation PACC was lower than within-generation PACC both globally and within-family, regardless of historic effective population size. Lower N_e tended to result in higher PACC overall. Global PACC across generations (CG1) decreased from 0.411 to 0.150 and 0.138 at N_e-levels of 100, 1000 and 5000 respectively (Fig. 6a–c). The change across N_e levels was smaller for the within-generation schemes, going from 0.629 to 0.595 and 0.569 for WG1 (Fig. 6a–c). At N_e 100, within-generation prediction of G1 clones were noticeably improved when training included clones from the previous generation, both in terms of global and within-family PACC (Fig. 6a). Meanwhile, at the higher effective population sizes there was almost no difference (Fig. 6b, c). Cross-generation within-family PACC decreased when using a smaller training set with higher relatedness to the validation set. This was true for all schemes but showed most clearly for N_e 1000 and N_e 5000 scenarios (Fig. 6e, f). The within-family PA followed the same pattern as the within-family PACC, except absolute values were lower (Fig. 6g–i).

Comparing the simulated datasets to the experimental data, the N_e 5000 scenario was most similar in terms of training set size for the CG1R scheme, which is determined by the internal relatedness of the founder population (Tables S2, S3). The average simulated training set size was 80 clones and 1447 individual tree phenotypes, compared to 82 and 1589 respectively for the experimental data. The N_e 1000 scenario was closest to experimental data for the CG1 scheme in terms of global PACC (0.251 vs 0.150) (Fig. 5a, Fig. 6b). For within-family PA using the same cross-validation scheme, the N_e 5000 scenario was closer to the experimental data (0.045 vs 0.041 respectively) (Fig. 5b, Fig. 6i). One surprising finding in the simulation was that within-family PACC was 25% higher than global PACC using CG1 for the N_e 1000 scenario (Fig. 6b, e). Experimental within-family PA on the other hand was 51% lower than global PA (Fig. 5b, Fig. S4).

The simulations were extended to more advanced generations (G2 – G5) with a historic N_e of 1000. Two sets of breeding programs were simulated, where the selection of parents was based on either global- or within-family selection. Again, genomic predictions were made across generations using either one previous generation or multiple generations for training (Fig. 7a-d). Cross-generation selection accuracy increased in later generations compared to G1. This was true both for global PACC and within-family PACC, regardless of whether only the last generation or all previous generations were used for model training. The increase was more rapid when all generations were included. Predicting G5 with multiple generations for training resulted in accuracies of 0.636 and 0.542 using global- and within-family selection respectively (Fig. 7b, d), while for single generation training it was 0.478 and 0.355 (Fig. 7a, c). Using only the previous generation, PACC showed a tendency to plateau at later generations, even slightly decreasing in terms of within-family PACC at G4 compared to G3 (Fig. 7c). Multiple generations for training continually increased PACC and did not show signs of leveling of even at G5 (Fig. 7b, d). All prediction and selection methods had the largest increase in PACC in the earlier stages of the simulated breeding

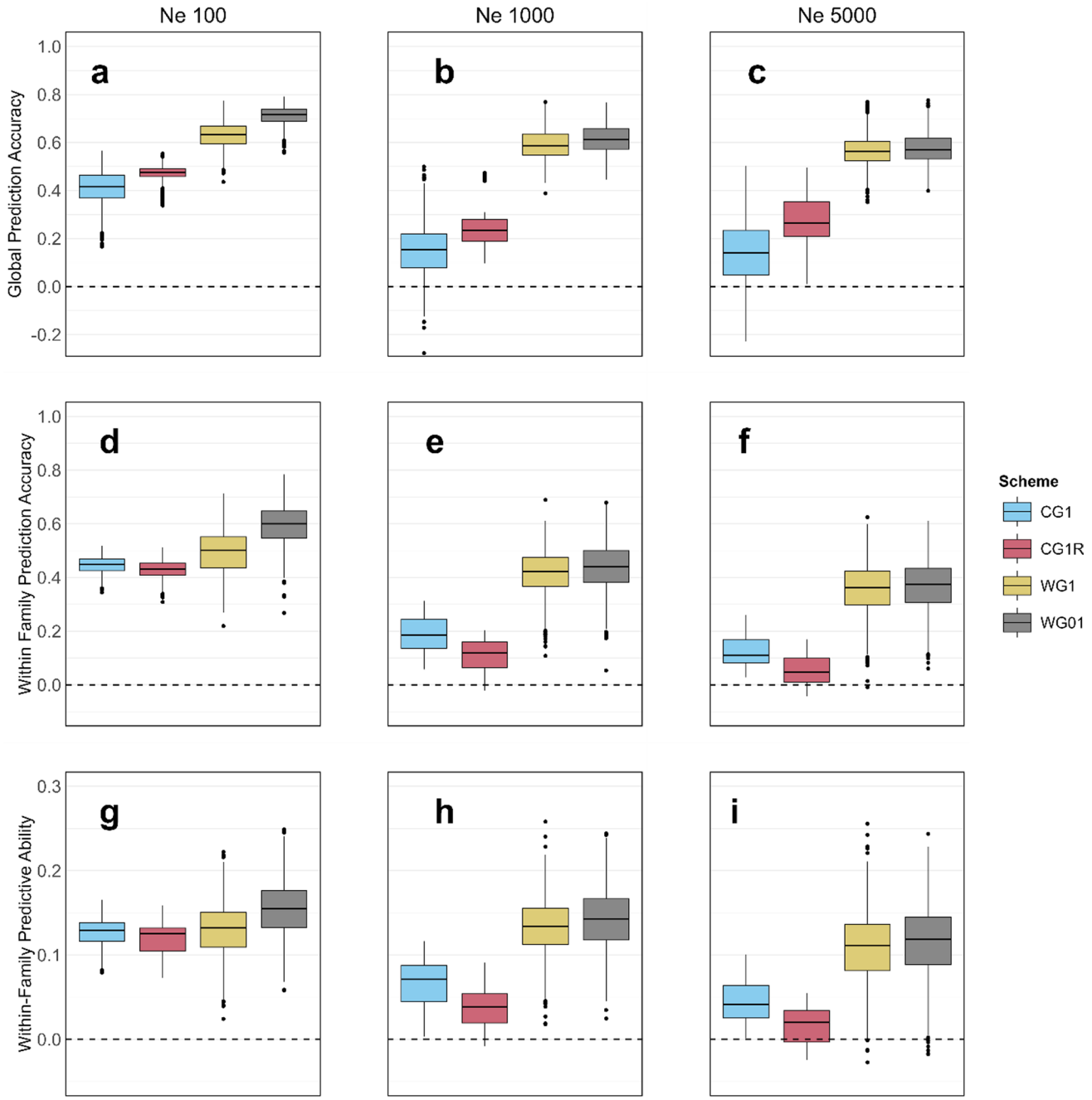


Fig. 6 Simulated replications of experimental analysis described in this study. True global prediction accuracy **a**–**c**, true within-family prediction accuracy **d**–**f** and within-family predictive ability **g**–**i** across three simulated levels of historical effective population size (N_e). Each boxplot has a sample size of 1000. Scheme correspond to cross-validation schemes described previously (see Fig. 3)

program. Going from predicting G1 to G2 almost doubled the global PACC and increased within-family PACC by 81%, if both G0 and G1 were used for training (Fig. 7b, d).

Generation-wide LD increased with each new round of selection and was generally higher when using global selection (Fig. 8a, b). At the fifth generation, the distance at which r^2 was expected to be 0.2 was between 5.0 to 20.6 kb and 1.3 to 1.7 kb using global selection and within-family selection respectively. There was also greater variation between iterations using global selection than within-family selection, which showed a more stable increase in LD for each new generation.

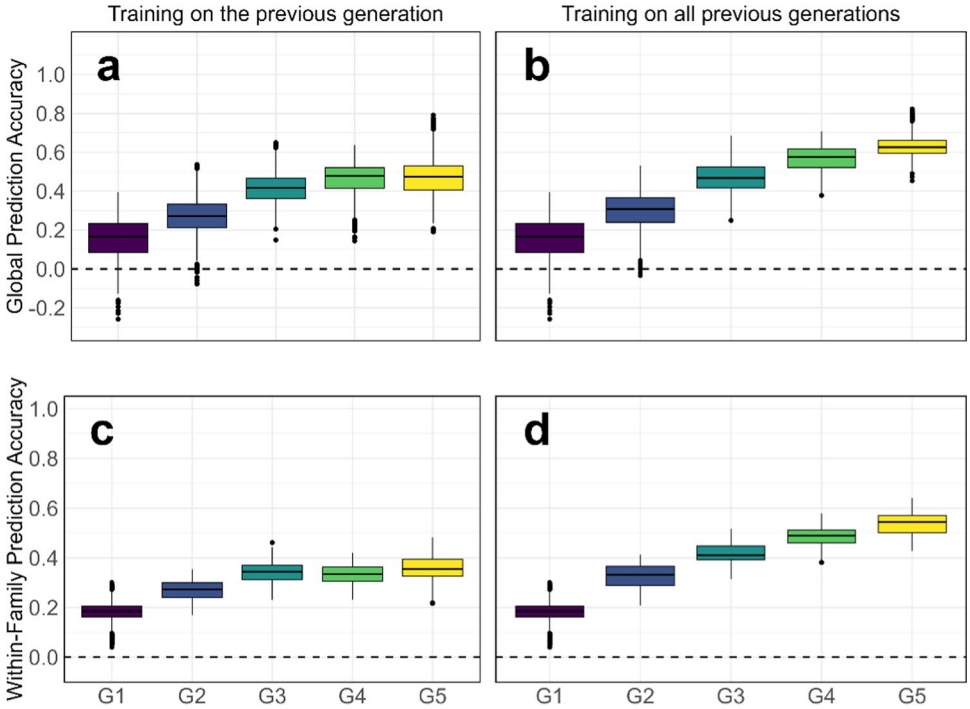


Fig. 7 Prediction accuracies of genomic selection across five simulated generations, generated using global- **a**, **b** or within-family selection **c**, **d**. The x-axis indicates the generation being used as validation set. Training sets were sampled from either the one previous generation **a**, **c** or all available previous generations **b**, **d**, relative to each validation set. Each boxplot has a sample size of 1000

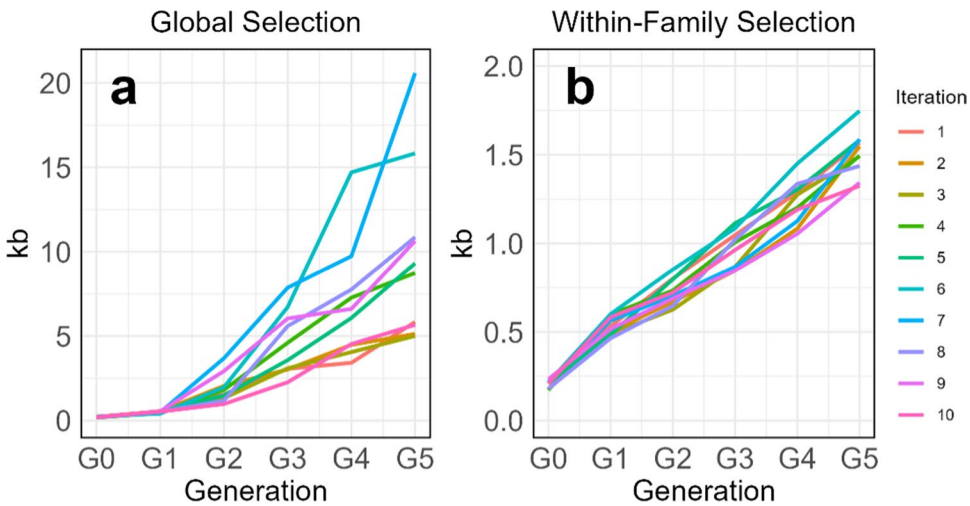


Fig. 8 Changes in Linkage disequilibrium (LD) across simulated generations, generated using global- **a** or within-family selection **b**. Values on the y-axis indicate the distance at which the expected LD (r^2) was 0.2

Discussion

Genomic selection could shorten the long generation interval by selecting new breeding candidates early at the seedling stage. However, GS models still need to be sufficiently accurate to not decrease genetic gain per unit of time compared to traditional breeding methods. We have investigated GS accuracy across two successive generations using experimental material from the Swedish Norway spruce breeding program. One key difference to most Norway spruce breeding populations in Sweden is that the plus-trees themselves were here tested as rooted cuttings in a series replicated field trials (Haapanen et al. 2015). Traditionally, plus trees are evaluated using progeny trials. Our estimates of narrow-sense heritability for height growth were generally comparable to previous GS studies on Norway spruce (Chen et al. 2018; 2019; Lenz et al. 2020a). Only the Lingbo trial had a surprisingly high heritability of 0.375, along with the highest standard error (Table S1). This was also the trial with the worst annual growth, and lowest phenotypic variance (Table 1).

Independent validation across generations

GS cross-validation studies have been conducted in many commercially important conifer species, usually by

assessing model accuracy using cross-validation within the same generation (Cappa et al. 2022; Klápště et al. 2020; Calleja-Rodriguez et al. 2020; Ratcliffe et al. 2015; Beaulieu et al. 2014; Zapata-Valenzuela et al. 2013). Our results show that this approach is an overoptimistic estimation of cross-generation accuracy. The associations between QTLs and markers are altered by each new recombination event (Habier et al. 2007; Xu et al. 2020). Predictions carried out within the same generation might benefit from long-range LD, which is likely to break up in the next breeding cycle. Directional selection (particularly with high intensity) would increase LD. If the loci contributing to the trait phenotype are physically close on a chromosome, recombination can't easily break up favored combinations, so LD builds up even more (Smith and Haigh 1974; Stephan 2019).

A recent GS study on Norway spruce also used independent validation across two generations (Hayatgheibi et al. 2025). However, the setup differed from ours in three important ways. Firstly, the studied traits were mainly solid wood properties. These typically have a higher heritability compared to growth traits (Hong et al. 2014; Zhang et al. 2022; Hayatgheibi et al. 2025). Secondly, all individuals in the first generation were included as parents, meaning no selection was applied. Directional selection of the best trees will raise the frequency of favorable alleles and lower that of unfavorable alleles. Our experimental dataset showed a clear difference in the distribution of GEBVs between G0 and G1, indicating selection (Fig. S5). Such selection and mating effects will influence GS accuracy, since they affect LD and the estimation of allelic effects (Hospital and Chevalet 1996; Hedrick 2017). Thirdly, the phenotypes from the parental generation were not collected from proper field trials, but clonal archives. Such conditions could provide a skewed comparison with expectations under traditional forest management.

Previous GS studies across generations in conifers have often used different estimates of prediction accuracy compared to the methodology described here (Bartholomé et al. 2016; Thistlethwaite et al. 2019; Isik et al. 2026). They all estimated prediction accuracy as the correlation between EBVs estimated from a large set of phenotypic data, and GEBVs. This typically results in higher PACC compared with other common methods (Calleja-Rodriguez et al. 2020). By using phenotypes excluded from training as validation, we have ensured that validation is independent. The drawback of this approach is that it relies on accurate estimation of variance components for scaling (Legarra et al. 2008). By using simulations, calculating prediction accuracy becomes straightforward because the breeder has access to TBVs. Our experimental and simulated analysis showed very similar trends for accuracy. Comparing scenarios CG1 and WG1, experimental cross-generation global PACC was around 59% lower than within-generation PACC, while for the simulation the same value was around 75% (Fig. 5a, Fig. 6b, c).

Within-family prediction

The superiority of GBLUP over ABLUP relies on the ability of markers to improve rankings within families by capturing the Mendelian segregation variation (Papin et al. 2024). This point becomes even more pronounced in the context of predictions across generations. In this scenario, ranking within families with ABLUP is not simply difficult, but impossible, since siblings of the same family will have identical breeding values in the absence of phenotypic data. Attempts have been made to estimate within-family accuracy in trees, either by averaging across families (Resende et al. 2017; Chen et al. 2023; Papin et al. 2024) or subtracting parental means (Thistlethwaite et al. 2017; 2019). The experimental studies done on conifers report negligible or zero accuracy, even using within-generation prediction. One recent study on Maritime pine has reported similarly low or zero efficacy of within-family selection using a GS model trained from a population (global model) (Papin et al. 2024). They attributed such low within-family accuracy to inadequate size of the genomic training sets and proposed that 40–65 individuals per full-sib family are needed in the genomic training set, from a total of 1600–2000 individuals for efficient within-family selection. Nevertheless, within-family prediction accuracy would likely still be lower than for global prediction. While global accuracy reflects the ability to predict both the parent average component and the mendelian sampling term, only the latter is measured by within-family accuracy (Werner et al. 2020). In the current study, average experimental PA within family was still very low, but positive across all cross-validation schemes. Similarly to global accuracy, both our experimental and simulation results show that within-family PACC and PA are lower across generations compared to within generations.

The low within-family GS accuracy may be explained by the fact that the all-additive genetic variance (among-family and within-family) are used in a global model building while only within-family variation (mainly Mendelian segregation) is used for prediction. In addition, the phenotypic within-family variation also includes non-additive genetic variation, which was not accounted for in our experimental GS model. In our simulation, genetic variance was completely additive. This could also explain why the simulation sometimes had higher within-family PACC than for global prediction, specifically for cross-generation prediction. Because our main training set (G0) had extremely limited family structure (Fig. S1), separating additive from non-additive effects would be challenging.

The relatedness between reference and a target population has been shown to have a major impact on the accuracy of

GS in forest trees (Lorenz and Smith 2015; Edwards et al. 2019; Grattapaglia 2022). As the average relationship coefficient between the training and validation sets increases, so does the probability that they share chromosome segments which are identical-by-decent. Following this logic, predicting offspring performance using only the selected parents for training can increase accuracy (Thistlethwaite et al. 2019). This strategy is sensitive to the small size of the training set, which could hinder the estimation of variance components. For this reason, we chose to combine parent trees with what we have called "relatives" based on the relatedness to parents according to the G-matrix. Even when the training set was ten times smaller, global prediction accuracy was still higher for CG1R than for CG1 (Fig. 5a), but the effect was either neutral or detrimental for within-family prediction. Our simulation shows especially clearly that both within-family PA and PACC decreased when relatedness was prioritized over sample size (Fig. 6h, i). Similarly, a study on loblolly pine (*Pinus taeda* L.) found no clear trend for within-family prediction accuracy and the relatedness between training and validation (Walker et al. 2022). Rather than relatedness, within-family prediction accuracy relies more on LD (Habier et al. 2013). Additionally, as the training sample size increases, the contribution of the pedigree relationship for accuracy also decreases relative to LD (Clark et al. 2012; Wientjes et al. 2013). This might explain why training only on "relatives" was less effective for ranking within families.

Tracking pedigree relationships can be accomplished with relatively few SNP markers even in large conifer genomes (Lenz et al. 2020b; Chen et al. 2023). Tracking LD between QTLs and SNPs on the other hand might require considerably more markers (De Roos et al. 2008; Isik et al. 2016; Resende et al. 2017; Thistlethwaite et al. 2020). In tree species with smaller genomes, like eucalyptus, current marker densities have been high enough to capture LD, and consequently within-family predictions were also high (Resende et al. 2017). The same study found that their marker density was close to the distance where LD decayed to $r^2 = 0.2$ in their population. Using a corresponding estimate of LD decay in Norway spruce (Chen et al. 2023), the number of required markers would be 19.6×10^9 divided by 42.9×10^3 equals around 450 000, assuming even spacing between SNPs. For cross-generation predictions, even more markers might be required. This point is somewhat contradicted by many studies, showing that accuracy reaches a plateau with marker densities available today (Chen et al. 2018; Thistlethwaite et al. 2019; Papin et al. 2024). However, they only reported changes in global PACC. As stated previously, our results show that trends in global and within-family prediction can differ. The effect of increased marker densities on GS performance should therefore be investigated further, specifically on the ability of GS to rank siblings within each family without phenotypes. Ideally, this should be compared within and across generations.

Training using multiple generations

Cattle breeders have been able to use the same GS model, that is, the same training data, for two generations of forward selection (Hayes et al. 2009). In trees, however, simulation studies show that the accuracy of GS will drop off in subsequent generations if the prediction model is not updated (Iwata et al. 2011; Vanavermaete et al. 2020). Again, additional recombination will cause LD between marker and QTL to decline (Habier et al. 2007; Xu et al 2020). Our simulation indicates that continually updating the GS models with additional generations of genotype and phenotype data for training could improve the PACC of cross-generation GS models. This is likely related to the buildup of LD and relatedness in each successive generation. These results are in agreement with simulation studies on animal breeding (Muir 2007; Sonesson and Meuwissen 2009; Dekkers et al. 2021) and adds that this trend applies for within-family PACC as well.

Even if the new GS model was only updated using the previous parental generation, PACC still increased in the initial generations. Since the training set size and number of markers was essentially the same, they cannot explain the improved GS accuracy. More likely, this was caused by the buildup of relatedness and LD. Training with more generations could improve predictions simply because of increased sample size. In this study, we did not attempt to discern if the composition of samples from many generations was better than an equal training size from one generation. However, (Muir 2007) observed that 4 generations of training were better than 2 for the reliability of GEBVs when the total number of training individuals were equal.

As stated previously, increased training sample size increases the relative importance of LD. While recombination breaks down LD between marker and QTLs, the selection carried out in breeding programs can generate stronger genome-wide LD in subsequent generations (Joukhadar et al. 2019). The extent of LD is influenced by the relatedness within a population (Mangin et al. 2012). The fact that LD increased faster with global selection likely reflects the fact that within-family selection minimizes the increase in group co-ancestry (Rosvall et al. 1998). The success of GS in animal breeding can be explained in part because of multiple completed breeding cycles (García-Ruiz et al. 2016). Meanwhile, many forest trees can be considered as largely undomesticated (Lebedev et al. 2020; Cappa et al. 2013). Rapid LD decay has been reported in many conifers (Neale and Savolainen 2004), with a notable exception of western redcedar (*Thuja*

plicata Donn ex D. Don) (Shalev et al. 2022; Gamal El-Dien et al. 2024). This adds to the challenge of implementing GS.

Simulated results should ideally be validated against experimental data when possible. To our knowledge, no experimental analysis has been done on forest trees on the effects of increasing the number of generations for training GS models. Experimental studies on several generations forward have been conducted on chicken (Picard Duret et al. 2020; Weng et al. 2016; Wolc et al. 2011), strawberry (Osorio et al. 2021) cattle and pigs (Lourenco et al. 2014). Increasing the number of generations for training generally increased GS accuracy. For some traits, however, continually adding generations of data eventually had a neutral or even a detrimental effect (Weng et al. 2016; Lourenco et al. 2014; Osorio et al. 2021).

Our results indicate that, currently, cross-generation GS of the G1 generation of Norway spruce could be too inefficient, especially when practicing within-family selection. That being said, genotyping of today's populations will likely still improve the GS accuracy of future generations, as shown by our simulation results. This highlights the need for tree breeders to consider keeping records of past generations to be available for future genotyping, as well as to continue field testing of new material.

Conclusion

We investigated the accuracy of GS of Norway spruce, comparing estimates within and across generations. Trees from two experimental generations of the Swedish Norway spruce breeding program were phenotyped for annual height growth rate in clonally replicated field trials and genotyped with a 50 K SNP array. GEBVs of the second generation were predicted using GBLUP, based on training sets comprising either the first generation, a subset of the second generation, or a combination of both. The experimental dataset was emulated using simulation and previous analysis was repeated. Additional simulations were performed to assess trends in cross-generation GS accuracy for future generations. Both experimental and simulated data showed a substantial drop in GS efficiency when estimated across generations relative to within-generation cross-validation. We also observed that cross-generation GS efficiency increased when training the prediction model on multiple generations of phenotypic and genotypic data. These results showcase a potential pitfall when evaluating GS accuracy. Our simulation highlights the value of retaining data from older material. Tree breeders looking to implement GS should consider the current number of generations in their breeding program, as this will impact the prediction accuracy of the next generation.

Acknowledgements Field sampling was carried out by ZhiqiangChen, Anders Fries, Tom van der Valk, Björn Nystedt and Per Unneberg. Laboratory DNA-extraction was carried out by Rajiv Chaudhary and Tomas Funda. Initial genotyping preparation was carried out by Zhiqiang Chen. Mats Berlin assisted with collation of historic pedigree and phenotypic data.

Author contributions E.A.C. designed the methods in later stages of the study, wrote the R scripts, analyzed the data, interpreted the results, generated majority of figures and wrote the manuscript. H.R.H. compiled relevant information on the plant material, designed the methodology, interpreted the results, and edited the manuscript. J.A. designed the methodology, interpreted the results, and edited the manuscript. M.S.wrote the initial study design and edited the manuscript. H.X.W. wrote the initial design of the study and contributed to later design of the manuscript, interpreted the results, generated figures, wrote and edited the manuscript.

Funding Open access funding provided by Swedish University of Agricultural Sciences. This research was financed by the Swedish Foundation for Strategic Research (SSF, grant number FFF20-0014). Access to genetic data was supported by the KAW project (2018.0272). The computer simulations were enabled by resources provided by the National Academic Infrastructure for Supercomputing in Sweden (NAISS), partially funded by the Swedish Research Council through grant agreement no. 2022–06725.

Data availability R codes for analysis and simulation will be available upon publication at [https://github.com/Edwcar/Cross-Generation-GS-in-Norway-Spruce.git]. Phenotypic and SNP data will be available upon publication at Zenodo repository: https://doi.org/10.5281/zenodo.18801992.

Declarations

Conflict of interest The authors declare no conflict of interest.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

References

Ahlinder J, Hall D, Suontama M, Sillanpää MJ (2024) Principal component analysis revisited: fast multitrait genetic evaluations with smooth convergence. G3: Genes|Genomes|Genetics 14(12):1–17. https://doi.org/10.1093/g3journal/jkae228

Amer PR, Banos G (2010) Implications of avoiding overlap between training and testing data sets when evaluating genomic predictions of genetic merit. Journal of Dairy Science 93(7):3320–3330. https://doi.org/10.3168/jds.2009-2845

Bartholomé J, Van Heerwaarden J, Isik F, Boury C, Vidal M, Plomion C, Bouffier L (2016) Performance of genomic prediction within and across generations in maritime pine. BMC Genomics 17(1):1–14. https://doi.org/10.1186/s12864-016-2879-8

Beaulieu J, Doerksen TK, MacKay J, Rainville A, Bousquet J (2014) Genomic selection accuracies within and between environments and small breeding groups in white spruce. BMC Genomics 15(1):1–16. https://doi.org/10.1186/1471-2164-15-1048

Beaulieu J, Lenz P, Bousquet J (2022) Metadata analysis indicates biased estimation of genetic parameters and gains using conventional pedigree information instead of genomic-based approaches in tree breeding. Sci Rep 12:1–10. https://doi.org/10.1038/s41598-022-06681-y

Bernhardsson C, Vidalis A, Wang X, Scofield DG, Schiffthaler B, Baison J, Street NR, García-Gil MR, Ingvarsson PK (2019) An ultra-dense haploid genetic map for evaluating the highly fragmented genome assembly of Norway spruce (*Picea abies*). G3: Genes|Genomes|Genetics 9(5):1623–1632. https://doi.org/10.1534/g3.118.200840

Bernhardsson C, Zan Y, Chen Z, Ingvarsson PK, Wu HX (2021) Development of a highly efficient 50K single nucleotide polymorphism genotyping array for the large and complex genome of Norway spruce (*Picea abies* L. Karst) by whole genome resequencing and its transferability to other spruce species. Mol Ecol Resour 21(3):880–896. https://doi.org/10.1111/1755-0998.13292

Bian L, Zheng R, Su S, Lin H, Xiao H, Wu HX, Shi J (2017) Spatial analysis increases efficiency of progenytesting of Chinese fir. J For Res 28:445–452. https://doi.org/10.1007/s11676-016-0341-z

Butler DG, Cullis BR, Gilmour AR, Gogel BJ, and Thompson R (2023). ASReml-R Reference Manual Version 4.2 ASReml estimates variance components under a general linear mixed model by residual maximum likelihood (REML). https://asreml.kb.vsni.co.uk/article-categories/asreml-r-resources/

Calleja-Rodriguez A, Li Z, Hallingbäck HR, Sillanpää MJ, Wu HX, Abrahamsson S, García-Gil MR (2019) Analysis of phenotypic- and estimated breeding values (EBV) to dissect the genetic architecture of complex traits in a Scots pine three-generation pedigree design. J Theor Biol 462:283–292. https://doi.org/10.1016/j.jtbi.2018.11.007

Calleja-Rodriguez A, Pan J, Funda T, Chen Z, Baison J, Isik F, Abrahamsson S, Wu HX (2020) Evaluation of the efficiency of genomic versus pedigree predictions for growth and wood quality traits in Scots pine. BMC Genomics 21(1):1–17. https://doi.org/10.1186/s12864-020-07188-4

Cappa EP, El-Kassaby YA, Garcia MN, Acuña C, Borralho NMG, Grattapaglia D, Marcucci Poltri SN (2013) Impacts of population structure and analytical models in genome-wide association studies of complex traits in forest trees: a case study in *Eucalyptus globulus*. PLoS ONE 8(11):1–16. https://doi.org/10.1371/journal.pone.0081267

Cappa EP, Ratcliffe B, Chen C et al (2022) Improving lodgepole pine genomic evaluation using spatial correlation structure and SNP selection with single-step GBLUP. Heredity 128:209–224. https://doi.org/10.1038/s41437-022-00508-2

Chen GK, Marjoram P, Wall JD (2009) Fast and flexible simulation of DNA sequence data. Genome Res 19(1):136–142. https://doi.org/10.1101/gr.083634.108

Chen ZQ, Baison J, Pan J, Karlsson B, Andersson B, Westin J, García-Gil MR, Wu HX (2018) Accuracy of genomic selection for growth and wood quality traits in two control-pollinated progeny trials using exome capture as the genotyping platform in Norway spruce. BMC Genomics 19(1):1–16. https://doi.org/10.1186/s12864-018-5256-y

Chen ZQ, Baison J, Pan J, Westin J, Gil MRG, Wu HX (2019) Increased prediction ability in Norway spruce trials using a marker X environment interaction and non-additive genomic selection model. J Hered 110(7):830–843. https://doi.org/10.1093/jhered/esz061

Chen ZQ, Klingberg A, Hallingbäck HR, Wu HX (2023) Preselection of QTL markers enhances accuracy of genomic selection in Norway spruce. BMC Genomics 24(1):1–16. https://doi.org/10.1186/s12864-023-09250-3

Clark SA, Hickey JM, Daetwyler HD, van der Werf JHJ (2012) The importance of information on relatives for the prediction of genomic breeding values and the implications for the makeup of reference data sets in livestock breeding schemes. Genet Sel Evol 44:1–9. https://doi.org/10.1186/1297-9686-44-4

De Roos APW, Hayes BJ, Spelman RJ, Goddard ME (2008) Linkage disequilibrium and persistence of phase in Holstein-Friesian, Jersey and Angus cattle. Genetics 179(3):1503–1512. https://doi.org/10.1534/genetics.107.084301

Dekkers JCM, Su H, Cheng J (2021) Predicting the accuracy of genomic predictions. Genet Sel Evol 53:1–23. https://doi.org/10.1186/s12711-021-00647-w

Denis M, Bouvet JM (2011) Genomic selection in tree breeding: testing accuracy of prediction models including dominance effect. BMC Proc 5(Suppl 7):O13. https://doi.org/10.1186/1753-6561-5-S7-O13

Dutkowski GW, Costa e Silva J, Gilmour AR, Lopez GA (2002) Spatial analysis methods for forest genetic trials. Can J For Res 32(12):2201–2214. https://doi.org/10.1139/x02-111

Dutkowski GW, Costa E Silva J, Gilmour AR, Wellendorf H, Aguiar A (2006) Spatial analysis enhances modelling of a wide variety of traits in forest genetic trials. Can J For Res 36(7):1851–1870. https://doi.org/10.1139/X06-059

Edwards SM, Buntjer JB, Jackson R, Bentley AR, Lage J, Byrne E, Burt C, Jack P, Berry S, Flatman E, Poupard B, Smith S, Hayes C, Gaynor RC, Gorjanc G, Howell P, Ober E, Mackay IJ, Hickey JM (2019) The effects of training population design on genomic prediction accuracy in wheat. Theor Appl Genet 132:1943–1952. https://doi.org/10.1007/s00122-019-03327-y

Endelman JB (2011) Ridge regression and other kernels for genomic selection with R package rrBLUP. The Plant Genome 4(3):250–255. https://doi.org/10.3835/plantgenome2011.08.0024

Fagroud M, Van Meirvenne M (2002) Accounting for soil spatial autocorrelation in the design of experimental trials. Soil Sci Soc Am J 66(4):1134–1142. https://doi.org/10.2136/sssaj2002.1134

Gamal El-Dien O, Shalev TJ, Yuen MMS, van der Merwe L, Kirst M, Yanchuk AD, Ritland C, Russell JH, Bohlmann J (2024) Genomic selection in western redcedar: from proof of concept to operational application. New Phytol 244(2):588–602. https://doi.org/10.1111/nph.20022

García-Ruiz A, Cole JB, Vanraden PM, Wiggans GR, Ruiz-López FJ, Van Tassell CP (2016) Changes in genetic selection differentials and generation intervals in US Holstein dairy cattle as a result of genomic selection. Proc Natl Acad Sci U S A 113(28):3995–4004. https://doi.org/10.1073/pnas.1519061113

Gaynor RC, Gorjanc G, Hickey JM (2021) AlphaSimR: An R package for breeding program simulations. G3: Genes|Genomes|Genetics 11(2):1–5. https://doi.org/10.1093/G3JOURNAL/JKAA017

Gezan S, de Oliveira AA, Galli G, and Murray D (2022). ASRgenomics: An R package with complementary genomic functions. Version 1.1.0 VSN International, Hemel Hempstead, United Kingdom. https://vsni.co.uk/free-software/asrgenomics

Gilmour AR, Gogel BJ, Cullis BR, Welham SJ, Thompson R (2021). ASReml User Guide Release 4.2 Functional Specification, VSN International Ltd, Hemel Hempstead, HP2 4TP, UK, www.vsni.co.uk

Grattapaglia D (2022) Twelve years into genomic selection in forest trees: climbing the slope of enlightenment of marker assisted

tree breeding. Forests 3(10):1–25. https://doi.org/10.3390/f13101554

Grattapaglia D, Silva-Junior OB, Resende RT, Cappa EP, Müller BSF, Tan B, Isik F, Ratcliffe B, El-Kassaby YA (2018) Quantitative genetics and genomics converge to accelerate forest tree breeding. Front Plant Sci 9:1-10. https://doi.org/10.3389/fpls.2018.01693

Haapanen M, Jansson G, Nielsen UB, Steffenrem A, Stener L-G (2015) The status of tree breeding and its potential for improving biomass production: A review of breeding activities and genetic gains in Scandinavia and Finland. SkogForsk.

Habier D, Fernando RL, Dekkers JCM (2007) The impact of genetic relationship information on genome-assisted breeding values. Genetics 177(4):2389–2397. https://doi.org/10.1534/genetics.107.081190

Habier D, Rohan LF, Dorian JG (2013) Genomic BLUP decoded: a look into the black box of genomic prediction. Genetics 194(3):597–607. https://doi.org/10.1534/genetics.113.152207

Hall D, Hallingbäck HR, Wu HX (2016) Estimation of number and size of QTL effects in forest tree traits. Tree Genet Genomes 12:1–17. https://doi.org/10.1007/s11295-016-1073-0

Hayatgheibi H, Hallingbäck HR, Gezan SA et al (2025) Cross-generational genomic prediction of Norway spruce (*Picea abies*) wood properties: an evaluation using independent validation. BMC Genomics 26(1):1–14. https://doi.org/10.1186/s12864-025-11861-x

Hayes BJ, Bowman PJ, Chamberlain AJ, Goddard ME (2009) Invited review: genomic selection in dairy cattle: progress and challenges. J Dairy Sci 92(2):433–443. https://doi.org/10.3168/jds.2008-1646

Hedrick PW (2017) Assortative mating and linkage disequilibrium. G3: Genes|Genomes|Genetics 7(1):55–62. https://doi.org/10.1534/g3.116.034967

Henderson CR (1984) Applications of linear models in animal breeding. University of Guelph, Guelph

Hill WG, Weir BS (1988) Variances and covariances of squared linkage disequilibria in finite populations. Theor Popul Biol 33:54–78. https://doi.org/10.1016/0040-5809(88)90004-4

Hong Z, Fries A, Wu HX (2014) High negative genetic correlations between growth traits and wood properties suggest incorporating multiple traits selection including economic weights for the future Scots pine breeding programs. Ann For Sci 71:463–472. https://doi.org/10.1007/s13595-014-0359-3

Hospital F, Chevalet C (1996) Interactions of selection, linkage and drift in the dynamics of polygenic characters. Genet Res 67(1):77–87. https://doi.org/10.1017/s0016672300033498

Isidro y Sánchez J, Akdemir D (2021) Training set optimization for sparse phenotyping in genomic selection: a conceptual overview. Front Plant Sci 12:1-14. https://doi.org/10.3389/fpls.2021.715910

Isik F, Bartholomé J, Farjat A, Chancerel E, Raffin A, Sanchez L, Plomion C, Bouffier L (2016) Genomic selection in maritime pine. Plant Sci 242:108–119. https://doi.org/10.1016/j.plantsci.2015.08.006

Isik F, Shalizi MN Walker TD (2026) Genomic selection validated across two generations of loblolly pine breeding. G3: Genes|Genomes|Genetics 16(7):1–11. https://doi.org/10.1093/g3journal/jkag122

Iwata H, Hayashi T, Tsumura Y (2011) Prospects for genomic selection in conifer breeding: a simulation study of *Cryptomeria japonica*. Tree Genet Genomes 7:747–758. https://doi.org/10.1007/s11295-011-0371-9

Joukhadar R, Daetwyler HD, Gendall AR, Hayden MJ (2019) Artificial selection causes significant linkage disequilibrium among multiple unlinked genes in Australian wheat. Evol Appl 12(8):1610–1625. https://doi.org/10.1111/eva.12807

Klápště J, Suontama M, Dungey HS, Telfer EJ, Graham NJ, Low CB, Stovold GT (2018) Effect of hidden relatedness on single-step genetic evaluation in an advanced open-pollinated breeding program. J Hered 109(7):802–810. https://doi.org/10.1093/jhered/esy051

Klápště J, Dungey HS, Telfer EJ, Suontama M, Graham NJ, Li Y, McKinley R (2020) Marker selection in multivariate genomic prediction improves accuracy of low heritability traits. Front Genet 11:1–15. https://doi.org/10.3389/fgene.2020.499094

Kuhn M (2008). Building Predictive Models in R Using the caret Package. http://www.jstatsoft.org/

Lebedev VG, Lebedeva TN, Chernodubov AI, Shestibratov KA (2020) Genomic selection for forest tree improvement: methods, achievements and perspectives. Forests 11:1–36. https://doi.org/10.3390/f11111190

Legarra A, Robert-Granié C, Manfredi E, Elsen JM (2008) Performance of genomic selection in mice. Genetics 180(1):611–618. https://doi.org/10.1534/genetics.108.088575

Lenz PRN, Beaulieu J, Mansfield SD, Clément S, Desponts M, Bousquet J (2017) Factors affecting the accuracy of genomic selection for growth and wood quality traits in an advanced-breeding population of black spruce (*Picea mariana*). BMC Genomics 18:1–17. https://doi.org/10.1186/s12864-017-3715-5

Lenz PRN, Nadeau S, Mottet MJ, Perron M, Isabel N, Beaulieu J, Bousquet J (2020a) Multi-trait genomic selection for weevil resistance, growth, and wood quality in Norway spruce. Evol Appl 13(1):76–94. https://doi.org/10.1111/eva.12823

Lenz PRN, Nadeau S, Azaiez A, Gérardi S, Deslauriers M, Perron M, Isabel N, Beaulieu J, Bousquet J (2020b) Genomic prediction for hastening and improving efficiency of forward selection in conifer polycross mating designs: an example from white spruce. Heredity 124(4):562–578. https://doi.org/10.1038/s41437-019-0290-3

Li Y, Dungey HS (2018) Expected benefit of genomic selection over forward selection in conifer breeding and deployment. PLoS ONE 13(12):1–21. https://doi.org/10.1371/journal.pone.0208232

Lorenz AJ, Smith KP (2015) Adding genetically distant individuals to training populations reduces genomic prediction accuracy in barley. Crop Sci 55(6):2657–2667. https://doi.org/10.2135/cropsci2014.12.0827

Lourenco DAL, Misztal I, Tsuruta S, Aguilar I, Lawlor TJ, Forni S, Weller JI (2014) Are evaluations on young genotyped animals benefiting from the past generations? J Dairy Sci 97(6):3930–3942. https://doi.org/10.3168/jds.2013-7769

Mangin B, Siberchicot A, Nicolas S, Doligez A, This P, Cierco-Ayrolles C (2012) Novel measures of linkage disequilibrium that correct the bias due to population structure and relatedness. Heredity 108(3):285–291. https://doi.org/10.1038/hdy.2011.73

Marroni F, Pinosio S, Zaina G, Fogolari F, Felice N, Cattonaro F, Morgante M (2011) Nucleotide diversity and linkage disequilibrium in *Populus nigra* cinnamyl alcohol dehydrogenase (CAD4) gene. Tree Genet Genomes 7:1011–1023. https://doi.org/10.1007/s11295-011-0391-5

McLean D, Apiolaza L, Paget M, Klápště J (2023) Simulating deployment of genetic gain in a radiata pine breeding program with genomic selection. Tree Genet Genomes 19:1–16. https://doi.org/10.1007/s11295-023-01607-9

Meuwissen THE, Hayes BJ, Goddard ME (2001) Prediction of total genetic value using genome-wide dense marker maps. Genetics 157(4):1819–1829. https://doi.org/10.1093/genetics/157.4.1819

Meuwissen T, Hayes B, Goddard M (2016) Genomic selection: a paradigm shift in animal breeding. Anim Front 6(1):6–14. https://doi.org/10.2527/af.2016-0002

Muir WM (2007) Comparison of genomic and traditional BLUP-estimated breeding value accuracy and selection response under alternative trait and genomic parameters. J Anim Breed Genet

124(6):342–355. https://doi.org/10.1111/j.1439-0388.2007.00700.x
Neale DB, Savolainen O (2004) Association genetics of complex traits in conifers. Trends Plant Sci 9(7):325–330. https://doi.org/10.1016/j.tplants.2004.05.006
Nystedt B, Street N, Wetterbom A et al (2013) The Norway spruce genome sequence and conifer genome evolution. Nature 497:579–584. https://doi.org/10.1038/nature12211
Osorio LF, Gezan SA, Verma S, Whitaker VM (2021) Independent validation of genomic prediction in strawberry over multiple cycles. Front Genet 11:1–13. https://doi.org/10.3389/fgene.2020.596258
Papin V, Gorjanc G, Pocrnic I, Bouffier L, Sanchez L (2024) Unlocking genome-based prediction and selection in conifers: the key role of within-family prediction accuracy illustrated in maritime pine (*Pinus pinaster* Ait.). Ann For Sci 81:1–22. https://doi.org/10.1186/s13595-024-01269-0
Picard Duret D, Varenne A, Herry F, Hérault F, Allais S, Burlot T, le Roy P (2020) Reliability of genomic evaluation for egg quality traits in layers. BMC Genet 21(1):1–11. https://doi.org/10.1186/s12863-020-0820-2
Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, Maller J, Sklar P, de Bakker PIW, Daly MJ, Sham PC (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet 81(3):559–575. https://doi.org/10.1086/519795
R Core Development Team. (2024). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/
Ratcliffe B, El-Dien OG, Klápště J, Porth I, Chen C, Jaquish B, El-Kassaby YA (2015) A comparison of genomic selection models across time in interior spruce (*Picea engelmannii* × *glauca*) using unordered SNP imputation methods. Heredity 115(6):547–555. https://doi.org/10.1038/hdy.2015.57
Resende JFR, Muñoz P, Resende MDV, Garrick DJ, Fernando RL, Davis JM, Jokela EJ, Martin TA, Peter GF, Kirst M (2012) Accuracy of genomic selection methods in a standard data set of loblolly pine (*Pinus taeda* L.). Genetics 190(4):1503–1510. https://doi.org/10.1534/genetics.111.137026
Resende RT, Resende MDV, Silva FF, Azevedo CF, Takahashi EK, Silva-Junior OB, Grattapaglia D (2017) Assessing the expected response to genomic selection of individuals and families in *Eucalyptus* breeding with an additive-dominant model. Heredity 119(4):245–255. https://doi.org/10.1038/hdy.2017.37
Rosvall O (2019) Using Norway spruce clones in Swedish forestry: Swedish forest conditions, tree breeding program and experiences with clones in field trials. Scand J For Res 34(5):342–351. https://doi.org/10.1080/02827581.2018.1562566
Rosvall O, Lindgren D, Mullin TJ (1998) Sustainability robustness and efficiency of a multi-generation breeding strategy based on within-family clonal selection. Silvae Genetica 47(5–6):307–321
Shalev TJ, Gamal El-Dien O, Yuen MMS, Shengqiang S, Jackman SD, Warren RL, Coombe L, van der Merwe L, Stewart A, Boston LB et al (2022) The western redcedar genome reveals low genetic diversity in a self-compatible conifer. Genome Res 32(10):1952–1964. https://doi.org/10.1101/gr.276358.121
Smith JM, Haigh J (1974) The hitch-hiking effect of a favourable gene. Genet Res 23:23–35. https://doi.org/10.1017/S0016672300014634
Sonesson AK, Meuwissen TH (2009) Testing strategies for genomic selection in aquaculture breeding programs. Genet Sel Evol 41(37):1–9. https://doi.org/10.1186/1297-9686-41-37
Stephan W (2019) Selective sweeps. Genetics 211(1):5–13. https://doi.org/10.1534/genetics.118.301319
Thistlethwaite FR, Ratcliffe B, Klápště J, Porth I, Chen C, Stoehr MU, El-Kassaby YA (2017) Genomic prediction accuracies in space and time for height and wood density of Douglas-fir using exome capture as the genotyping platform. BMC Genomics 18(1):1–17. https://doi.org/10.1186/s12864-017-4258-5
Thistlethwaite FR, Ratcliffe B, Klápště J, Porth I, Chen C, Stoehr MU, El-Kassaby YA (2019) Genomic selection of juvenile height across a single-generational gap in Douglas-fir. Heredity 122(6):848–863. https://doi.org/10.1038/s41437-018-0172-0
Thistlethwaite FR, Gamal El-Dien O, Ratcliffe B, Klápště J, Porth I, Chen C, Stoehr MU, Ingvarsson PK, El-Kassaby YA (2020) Linkage disequilibrium vs. pedigree: genomic selection prediction accuracy in conifer species. PLoS ONE 15(6):1–14. https://doi.org/10.1371/journal.pone.0232201
Vanavermaete D, Fostier J, Maenhout S, de Baets B (2020) Preservation of genetic variation in a breeding population for long-term genetic gain. G3: Genes|Genomes|Genetics 10(8):2753–2762. https://doi.org/10.1534/g3.120.401354
Walker TD, Cumbie WP, Isik F (2022) Single-step genomic analysis increases the accuracy of within-family selection in a clonally replicated population of *Pinus taeda L.* Forest Sci 68(1):37–52. https://doi.org/10.1093/forsci/fxab054
Weng Z, Wolc A, Shen X, Fernando RL, Dekkers JCM, Arango J, Settar P, Fulton JE, O'Sullivan NP, Garrick DJ (2016) Effects of number of training generations on genomic prediction for various traits in a layer chicken population. Genet Sel Evol 48(22):1–21. https://doi.org/10.1186/s12711-016-0198-9
Werner CR, Gaynor RC, Gorjanc G, Hickey JM, Kox T, Abbadi A, Leckband G, Snowdon RJ, Stahl A (2020) How population structure impacts genomic selection accuracy in cross-validation: implications for practical breeding. Front Plant Sci 11:1–14. https://doi.org/10.3389/fpls.2020.592977
Wientjes YCJ, Veerkamp RF, Calus MPL (2013) The effect of linkage disequilibrium and family relationships on the reliability of genomic prediction. Genetics 193(2):621–631. https://doi.org/10.1534/genetics.112.146290
Wientjes YCJ, Bijma P, Calus MPL, Zwaan BJ, Vitezica ZG, van den Heuvel J (2022) The long-term effects of genomic selection: 1. Response to selection, additive genetic variance, and genetic architecture. Genet Sel Evol 54(1):1–21. https://doi.org/10.1186/s12711-022-00709-7
Wolc A, Arango J, Settar P, Fulton JE, O'Sullivan NP, Preisinger R, Habier D, Fernando R, Garrick DJ, Dekkers JCM (2011) Persistence of accuracy of genomic estimated breeding values over generations in layer chickens. Genet Sel Evol 43(1):1–8. https://doi.org/10.1186/1297-9686-43-23
Wu XH (2019) Benefits and risks of using clones in forestry – a review. Scand J Forest Res 34(5):352–359. https://doi.org/10.1080/02827581.2018.1487579
Xu Y, Liu X, Fu J, Wang H, Wang J, Huang C, Prasanna BM, Olsen MS, Wang G, Zhang A (2020) Enhancing genetic gain through genomic selection: from livestock to plants. Plant Commun 1(1):1–21. https://doi.org/10.1016/j.xplc.2019.100005
Zapata-Valenzuela J, Whetten RW, Neale D, McKeand S, Isik F (2013) Genomic Estimated Breeding Values Using Genomic Relationship Matrices in a Cloned Population of Loblolly Pine. G3: Genes|Genomes|Genetics 3(5):909–916. https://doi.org/10.1534/g3.113.005975
Zhang Q, Pei X, Lu X, Zhao C, Dong G, Shi W, Wang L, Li Y, Zhao X, Tigabu M (2022) Variations in growth traits and wood physicochemical properties among *Pinus koraiensis* families in Northeast China. J For Res 33:1637–1648. https://doi.org/10.1007/s11676-022-01455-8

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.