


RESEARCH ARTICLE

Open Access



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

Zhi-Qiang Chen¹, John Baisou¹, Jin Pan¹, Bo Karlsson², Bengt Andersson³, Johan Westin³,
María Rosario García-Gil¹ and Harry X. Wu^{1,4*} 

Abstract

Background: Genomic selection (GS) can increase genetic gain by reducing the length of breeding cycle in forest trees. Here we genotyped 1370 control-pollinated progeny trees from 128 full-sib families in Norway spruce (*Picea abies* (L.) Karst.), using exome capture as genotyping platform. We used 116,765 high-quality SNPs to develop genomic prediction models for tree height and wood quality traits. We assessed the impact of different genomic prediction methods, genotype-by-environment interaction ($G \times E$), genetic composition, size of the training and validation set, relatedness, and number of SNPs on accuracy and predictive ability (PA) of GS.

Results: Using G matrix slightly altered heritability estimates relative to pedigree-based method. GS accuracies were about 11–14% lower than those based on pedigree-based selection. The efficiency of GS per year varied from 1.71 to 1.78, compared to that of the pedigree-based model if breeding cycle length was halved using GS. Height GS accuracy decreased to more than 30% while using one site as training for GS prediction and using this model to predict the second site, indicating that $G \times E$ for tree height should be accommodated in model fitting. Using a half-sib family structure instead of full-sib structure led to a significant reduction in GS accuracy and PA. The full-sib family structure needed only 750 markers to reach similar accuracy and PA, as compared to 100,000 markers required for the half-sib family, indicating that maintaining the high relatedness in the model improves accuracy and PA. Using 4000–8000 markers in full-sib family structure was sufficient to obtain GS model accuracy and PA for tree height and wood quality traits, almost equivalent to that obtained with all markers.

Conclusions: The study indicates that GS would be efficient in reducing generation time of breeding cycle in conifer tree breeding program that requires long-term progeny testing. The sufficient number of trees within-family (16 for growth and 12 for wood quality traits) and number of SNPs (8000) are required for GS with full-sib family relationship. GS methods had little impact on GS efficiency for growth and wood quality traits. GS model should incorporate $G \times E$ effect when a strong $G \times E$ is detected.

Keywords: Norway spruce, Exome capture, Genotype-by-environment interaction, GBLUP, Bayesian LASSO, Bayesian ridge regression, RKHS

* Correspondence: Harry.wu@slu.se; Harry.wu@CSIRO.au

¹Umeå Plant Science Centre, Department of Forest Genetics and Plant Physiology, Swedish University of Agricultural Sciences, SE-90183 Umeå, Sweden

⁴CSIRO NRCA, Black Mountain Laboratory, Canberra, ACT 2601, Australia

Full list of author information is available at the end of the article



Background

Norway spruce (*Picea abies* (L.) Karst.) is one of the most important conifer species for commercial wood production and ecological integrity in Europe [1]. A conventional breeding program for Norway spruce based on pedigree-based phenotypic selection usually takes between 20 and 30 years in Scandinavian countries [2]. To shorten the breeding cycle, genomic selection (GS) has recently been proposed as an alternative in many tree species such as eucalypts (*Eucalyptus*) [3–5], maritime pine (*Pinus pinaster* Aiton) [6, 7], loblolly pine (*Pinus taeda* L.) [8, 9], white spruce and its hybrids (*Picea glauca* [Moench] Voss) [10–12], and black spruce (*Picea Mariana* [Mill] B.S.P.) [13].

Hayes et al. [14] considered four major factors affecting the accuracy of GS: 1) heritability of the target trait; 2) the extent of linkage disequilibrium (LD) between the marker and the quantitative trait locus (QTL); 3) the size of training population and the degree of relationship between training set (TS) and validation set (VS); and 4) the genetic architecture of the target trait. The genetic architecture and heritability of the target/breeding traits are intrinsic of the nature of the traits and environment in the test trials. Thus it is difficult to change in a practical breeding process. However, some factors such as LD between the marker and the QTL and the size of training population are relatively easy to be managed by increasing the number of markers and relationship between training and prediction (or validation) populations, reducing effective population size (N_e), and increasing training population size [15].

Model selection in GS is quite important for the prediction of genomic estimated breeding values (GEBVs) [14]. In GS, model adequacy is more related to the genetic architecture of the target trait. Based on genome-wide association studies (GWAS), most growth and wood quality traits for conifer species have polygenic inheritance with a gamma or exponential distribution of allelic effects [16]. To account for these skewed distributions of a few genes with large effects and most of genes with small effects, Bayes A, B, and C π and Bayesian Least Absolute Shrinkage and Selection Operator (BLASSO) were developed to fit the models more accurately, in contrast to Genomic best linear unbiased prediction (GBLUP) model that assumes a normal distribution of allelic effect. In most studies for growth and wood quality traits, the results were similar regardless of models used [4, 9]. Resende et al. [9] reported that fusiform rust in loblolly pine may be controlled by a few genes with large effects and Bayesian-based models had a higher predictive ability (PA), which is defined as the correlation between adjusted phenotype values and GEBVs. Thus, it is worthwhile to test different models on traits that may have different genetic architectures for specific tree species.

So far, several genotyping technologies have been employed in GS, such as diversity array technology (DArT) array [5, 17], SNP chip/array [4–6, 9, 10, 13], genotyping by sequencing (GBS) [10, 11], and exome capture [18]. Those technologies were developed to genotype a subset of a whole genome, especially for conifer species with a large genome size. From the published papers, the number of used markers in tree species varies from 2500 to 69,511 single nucleotide polymorphisms (SNPs), most with a few thousands of SNPs. With the large genome size in most commercial conifer species (~ 20Gb) [19], for example, 20 Gb in Norway spruce [20], such small number of markers may not be able to capture most of QTL effects with short-range marker-QTLs LD in undomesticated populations or large breeding populations. Thus, such studies mostly capture those QTL effects with long-range marker-QTLs LD and relationships in highly related populations, such as full-sib families in tree breeding programs with a small N_e .

Evaluations of accuracy in GS have been performed with phenotypic and dense marker data from a single site and multiple sites in several tree species [8, 11, 17]. Substantial $G \times E$ effects for growth traits have been found in conifer species [21–23]. However, wood quality traits usually have a low or non-significant $G \times E$ [24–26]. Thus, a GS model for growth traits using data from a single site and used to predict genomic breeding values in another site, may produce a low accuracy.

The aims of this study were to 1) evaluate the accuracy of GS on tree height and wood quality traits; 2) assess the GS accuracy for single site, cross-site, and joint-sites (e.g. $G \times E$ effect on GS selection); 3) examine effect of different statistical models and ratios between TS and VS for GS; 4) explore the roles of relatedness (full-sib, half-sib and unrelated) on accuracy of GS; 5) test the accuracy using subsets of random markers and markers with the largest positive effects; 6) estimate number of trees within-family and number of families required for effective GS for tree height and wood quality traits.

Materials and methods

Sampling of plant material

In this study, 1370 individuals were selected from two 28-year-old control-pollinated progeny trials with the same 128 families from a partial diallel mating design that consisted of 55 parents originating from Northern Sweden. 5–20 trees per family per site were selected in Vindeln (64.30°N, 19.67°E, altitude: 325 m) and Hädanberg (63.58°N, 18.19°E, altitude, 240 m). Buds and the first year fresh needles from 46 parents were sampled in a grafted archive at Skogforsk, Sävar (63.89°N, 20.54°E) and in a grafted seed orchard at Hjässjö (63.93°N, 20.15°E). Progenies were raised in the nursery at

Sävar, and the trials were established in 1988 by Skogforsk in Vindeln and Hädanberg.

A completely randomized design without designed pre-block was used in the Vindeln trial (site 1), which was divided into 44 post-blocks. Single-tree plot with a spacing of 1.5 m × 2 m was used in each rectangular block with 60 trees (6 × 10). The same design was also used in the Hädanberg trial (site 2) with 44 post-blocks, but for the purpose of demonstration, there was an extra design with 47 extra plots, each plot with 16 trees (4 × 4). Based on the spatial analysis, in the final model, 47 plots were combined into two big post-blocks.

Phenotyping

Tree height was measured in 2003 at the age of 17 years. Solid-wood quality traits including Pilodyn penetration (Pilodyn) and acoustic velocity (velocity) were measured in October 2016. Surrogate wood density trait was measured using Pilodyn 6 J Forest (PROCEQ, Zurich, Switzerland) with a 2.0 mm diameter pin, without removing the bark. Velocity is highly related to microfibril angle (MFA) in Norway spruce [27] and was determined using Hitman ST300 (Fiber-gen, Christchurch, New Zealand). By combining the Pilodyn penetration and acoustic velocity, indirect modulus of elasticity (MOE) was estimated using the equation developed by Chen et al. [27].

Genotyping

Total genomic DNA was extracted from 1, 370 control-pollinated progeny and their 46 unrelated parents using the Qiagen Plant DNA extraction protocol with DNA quantification performed using the Qubit® ds DNA Broad Range Assay Kit, Oregon, USA. Probe design and evaluation were described in Vidalis et al. [28]. Sequence capture was performed using the 40,018 probes previously designed and evaluated for the materials [28] and samples were sequenced to an average depth of 15x at an Illumina HiSeq 2500 platform. Raw reads were mapped against the *P. abies* reference genome v1.0 using BWA-mem [29, 30]. SAMTools [31] and Picard [32] were used for sorting and removal of PCR duplicates and the resulting BAM files were subsequently reduced to containing the probe only bearing scaffolds (24,919) before variant calling. Variant calling was performed using the Genome Analysis Toolkit (GATK) HaplotypeCaller [32] in Genome Variant Call Format (gVCF) output format. Samples were then merged into batches of ~200 before all samples were jointly called.

As per the recommendations from GATK's best practices, Variant Quality Score Recalibration (VQSR) method was performed in order to avoid the use of hard filtering for exome/sequence capture data. The VQSR method utilizes machine-learning algorithms to learn

from a clean dataset to distinguish what a good versus bad annotation profile of variants for a particular species should be like. For the VQSR analysis two datasets were created, a training subset and the final input file. The training dataset was derived from a Norway spruce genetic mapping population with loci showing expected segregation patterns [33]. The training dataset was designated as true SNPs and assigned a prior value of 12.0. The final input file was derived from the raw sequence data using GATK best practices with the following parameters: extended probe coordinates by + 100 excluding INDELS, excluding LowQual sites, and keeping only bi-allelic sites. The following annotation parameters QualByDepth, MappingQuality and BaseQRankSum, with tranches 100, 99.9, 99.0 and 90.0 were then applied to the two files for the determination of the good versus bad variant annotation profiles.

The recalibrated Variant Call Format was filtered based on the following steps: (1) removing indels; (2) keeping only biallelic loci; (3) treating genotype with a genotype quality (GQ) < 6 as missing; (4) filtering read depth (DP) < 2; (5) removing individual call rate < 50%; (6) removing variant call rate ("missingness") < 90%; (7) minor allele frequency (MAF) < 0.01. After steps 1–4, we calculated discordance between 148 pairs technique replicates, the average discordance was less 1%. Thus, conditions of GQ < 6 and DP < 2 as missing are sufficient to do downstream analysis. After all filtering, 116,765 SNPs were kept for downstream analysis and 77,116 SNPs were independent based on LD ($r^2 < 0.2$) calculated in PLINK [34]. The resultant SNPs were annotated using the default parameters for snpEff 4. The Ensembl general feature format (GTF, gene sets) information for the *P. abies* genome was utilized to build an annotation database. This analysis revealed that 90% of the variants were located within gene coding regions, with only 10% variants in intronic regions.

LD K-nearest neighbour genotype imputation approach [35] was used to impute missing genotypes in TASSEL 5 [36]. After several rounds of imputation, a few of missing genotypes were imputed using random imputation of the codeGeno function in the synbreed package in R [37]. In total, 5.9% of missing genotypes were imputed.

Estimating breeding values

Breeding values for the genotypes in the trials were predicted using the following model:

$$y = X\beta + Wb(s) + Z_1a + e \quad (1)$$

Where y is a vector of phenotypic observations of a single trait; β is a vector of fixed effects, including a grand mean and site effects, $b(s)$ is a vector of post-block within site effects, a is a vector of site by

additive effects of individuals. X , W , and Z_1 are incidence matrices for β , $b(s)$, and a , respectively. For join-site cross-validation, the average of breeding values was assumed as estimated (true) breeding values (EBVs) when unstructured variance and covariance were used for additive effects. Otherwise, EBVs in the single site were assumed as true or reference breeding values. The random additive effects (a) in equation (1) were assumed

to follow $a \sim N\left(0, A \begin{bmatrix} \sigma_{a1}^2 & \sigma_{a12} \\ \sigma_{a12} & \sigma_{a2}^2 \end{bmatrix}\right)$, where A is the additive genetic relationship matrix, σ_{a1}^2 and σ_{a2}^2 are the additive genetic variances for site 1 and site 2, respectively, σ_{a12} is additive genetic covariance between site 1 and

site 2. The residual e was assumed to follow $e \sim N(0, \begin{bmatrix} I_{n1}\sigma_{e1}^2 & 0 \\ 0 & I_{n2}\sigma_{e2}^2 \end{bmatrix})$, where σ_{e1}^2 and σ_{e2}^2 are the residual variances for site 1 and site 2, I_{n1} and I_{n2} are identity matrices, $n1$ and $n2$ are the number of individuals in each site, 0 is the zero matrix.

To obtain accurate heritability estimates for joint-site models, we used the equation as following:

$$y = X\beta + Wb(s) + Z_1a + Z_2sa + e \tag{2}$$

where sa is a vector of site-by-additive interaction effects, a , sa , and e were assumed to be homogenous between the two sites.

Statistical analyses for genomic predictions

GBLUP, Bayesian ridge regression (BRR), BLASSO, and reproducing kernel Hilbert space (RKHS) were used to estimate GEBVs. We implemented GBLUP calculations using ASReml R [38]. And we implemented the BRR, BLASSO, and RKHS methods using BGLR function from the BGLR package in R [39]. The details of these statistical methods will be defined later. The GEBVs were estimated using the following mixed linear model:

$$y' = X\beta + Za + e \tag{3}$$

Where y' is a vector of adjusted phenotypic observations by post-block effects and standardized site effect (transforming it to have zero mean and unit variance for each site), β is a vector of fixed effect, including a grand mean), a and e are vectors of random additive and random error effects, respectively, and X and Z are the incidence matrices.

The four genomic-based best linear unbiased prediction methods were compared with the traditional pedigree-based best linear unbiased prediction (ABLUP)

1) ABLUP

The ABLUP is the traditional method that utilizes a pedigree relationship matrix (A) to predict the EBVs. For ABLUP the vector of random additive effect (a) in equation (3) is assumed to follow a normal distribution $a \sim N(0, A\sigma_a^2)$, where σ_a^2 is the additive genetic variance. The residual vector e is assumed as $e \sim N(0, I\sigma_e^2)$, where I is the identity matrix. The mixed model equation (3) was solved to obtain EBVs as:

$$\begin{bmatrix} X'X & X'Z \\ Z'X & Z'Z + A^{-1}\alpha \end{bmatrix} \begin{bmatrix} b \\ u \end{bmatrix} = \begin{bmatrix} X'y \\ Z'y \end{bmatrix} \tag{4}$$

The scalar α is defined as $\alpha = \sigma_e^2/\sigma_a^2$, where σ_e^2 is the residual variance, σ_a^2 is the additive genetic variance.

2) GBLUP

The GBLUP model is the same as ABLUP, with the only difference being that the genomic relationship matrix (G) replaces the A matrix. The G matrix is calculated as $G = \frac{(M-P)(M-P)^T}{2\sum_{i=1}^M p_i(1-p_i)}$, where M is the matrix of samples with SNPs encoded as 0, 1, 2 (i.e. the number of minor alleles), P is the matrix of allele frequencies with the i^{th} column given by $2(p_i - 0.5)$, where p_i is the observed allele frequency of all genotyped samples. In GBLUP, the random additive effect (a) in equation (3) is assumed to follow $a \sim N(0, G\sigma_g^2)$, where σ_g^2 is the genomic-based genetic variance and GEBVs (\hat{a}) are predicted from equation (4), but with A^{-1} replaced by G^{-1} and σ_a^2 replaced by σ_g^2 . The inverse of G matrix was estimated using write.relationshipMatrix function in the synbred package in R [40].

3) Bayesian ridge regression (BRR)

BRR is a Bayesian version of ridge regression with the shrinkage merit that was originally intended to deal with the problem of high correlation among predictors in linear regression models [41]. The random additive vector a is assigned a multivariate normal prior distribution with a common variance to all marker effects, that is $a \sim N(0, I_p\sigma_m^2)$, where p is the number of markers. Parameter σ_m^2 denotes the unknown genetic variance contributed by each individual marker and is assigned as $\sigma_m^2 \sim \chi^{-2}(df_a, S_m)$, where df_a is degrees of freedom, S_m is the scale parameter. Finally, the residual variance is assigned as $\sigma_e^2 \sim \chi^{-2}(df_e, S_e)$, where df_e is degrees of freedom for residual variance, S_e is the scale parameter for residual variance.

4) Bayesian LASSO (BLASSO)

BLASSO is a Bayesian version of LASSO regression with two properties of LASSO: 1) shrinkage and 2) variable selection. BLASSO assumes that the random additive effects in equation (3) is given a multivariate normal distribution with marker specific prior variance, which is assigned as $a \sim N(0, T\sigma_m^2)$, where $T = \text{diag}(\tau_1^2, \dots, \tau_p^2)$, Parameter τ_j^2 is assigned as $\tau_j^2 \sim \text{Exp}(\lambda^2)$ for $j = 1, \dots, p$, where λ^2 is assigned as $\lambda^2 \sim \text{Gamma}(r, \delta)$. The residual variance is assigned as $\sigma_e^2 \sim \chi^{-2}(df_e, S_e)$, where df_e is degrees of freedom, S_e is the scale parameter.

5) RKHS

RKHS assumes that the random additive marker effects in equation (3) are distributed as $a \sim N(0, \bar{K}\sigma_a^2)$, where \bar{K} is computed by means of a Gaussian Kernel that is given by $K_{ij} = \exp(-hd_{ij})$, where h is a semi-parameter that controls how fast the prior covariance function declines as genetic distance increase and d_{ij} is the genetic distance between two samples computed as $d_{ij} = \sum_{k=1}^p (x_{ik} - x_{jk})^2$, where x_{ik} and x_{jk} are the k^{th} SNPs for the i^{th} and j^{th} samples, respectively [42]. RKHS method uses a Gibbs sampler for the Bayesian framework and assigned the prior distribution of σ_a^2 and σ_e^2 as $\sigma_a^2 \sim \chi^{-2}(df_a, S_a)$ and $\sigma_e^2 \sim \chi^{-2}(df_e, S_e)$, respectively. Here we chose a Single-Kernel model as suggested by Perez and de Los Campos [39], where h value was defined as $h = 0.25$.

Model convergence and prior sensitivity analysis

The algorithm is extended by Gibbs sampling for estimation of variance components. The Gibbs sampler was run for 150,000 iterations with a burn-in of 50,000 iterations. A thinning interval was set to 1000. The convergence of the posterior distribution was verified using trace plots. Flat priors were given to all the models.

Heritability and type-B genetic correlation estimates

Pedigree-based individual narrow-sense heritabilities (h_a^2) and marker-based individual narrow-sense heritabilities (h_g^2) were calculated as

$$h_a^2 = \frac{\sigma_a^2}{\sigma_{pa}^2}, h_g^2 = \frac{\sigma_g^2}{\sigma_{pg}^2}$$

respectively, where σ_a^2 is the pedigree-based additive variance estimated from ABLUP, while σ_g^2 is the marker-based additive variance estimated from GBLUP. σ_{pa}^2 and σ_{pg}^2 are phenotypic variances for pedigree-based and marker-based models, respectively. Type-B genetic correlation was calculated as $r_{12} = \sigma_{a12} / \sqrt{\sigma_{a1}^2 \sigma_{a2}^2}$, where σ_{a12} is covariance between additive effects of the same

traits in different sites and σ_{a1}^2 and σ_{a2}^2 are estimated additive variances for the same traits in different sites, respectively [43]. One-tailed likelihood ratio test (LRT) was used to check the significance of the Type-B genetic correlation against one.

Model validation and estimation of GS accuracy

Tenfold cross-validation (90% of training and 10% validation) was performed for all the models, except in testing the various sizes of training data sets and the number of trees per family. GEBVs in VS for Bayesian and RKHS methods were estimated as

$$\hat{g}_i = \sum_{j=1}^n Z'_{ij} \hat{a}_j$$

Where Z'_{ij} is the indicator covariate (-1, 0, or 1) for the i^{th} tree at the j^{th} locus and \hat{a}_j is the estimated effect at the j^{th} locus. In this study, prediction accuracy (accuracy) was defined as the Pearson correlation between cross-validated GEBVs and EBVs as “true” or reference breeding values estimated from ABLUP using all the phenotypic data (y). Predictive ability (PA) was defined as the Pearson correlation between GEBVs and adjusted phenotypic values y' in equation (3).

Testing different statistical models and the size of training set on GS accuracy

To test the effect of different statistical models, we used the five models (ABLUP, GBLUP, BRR, BLASSO, and RKHS) and five TS/VS sizes (ratio of 1:1, 3:1, 5:1, 7:1, and 9:1 of training/validation population size, respectively) to evaluate the accuracies of different models. For each of 25 models, 10 replicate runs were carried out for each scenario of four traits.

Testing the number of families on GS accuracy

To test the effect of the number of families on GS, we randomly selected family numbers from 10 to 120 to test the efficiency of GS. The purpose is to examine the efficiency of using a small subset of families in clonal selection.

Testing the number of trees per family on GS accuracy

To test the effect of the number of trees per family on GS, we randomly selected 1 to 20 trees per family as TS, remaining trees as VS.

Testing site effect on GS accuracy

In order to consider genotype and environment interaction ($G \times E$) effect on GS, different GS scenarios were tested: 1) within-site GS, both training and validation sets are from one single site, where EBVs from single site model including $G \times E$ term was assumed as true breeding values of the site; 2) cross-site GS, using one site data as TS to predict GEBVs in another site; 3) joint-site

GS, average EBVs were assumed as true breeding values when unstructured additive variance-covariance were used in equation (1).

Testing the effect of relatedness on GS accuracy

In order to test whether the different relatedness (family structures) affect the accuracy and PA, three different scenarios were used in this study. There were: 1) TS and VS selected based on random sampling, but more likely from the same full-sib families; 2) for comparison purpose, TS and VS selected based on half-sib family structure in which the TS and VS shared female parents, but from different families; 3) for comparison purpose, TS and VS selected based on unrelated family structure in which the TS and VS shared different female and male parents.

Testing subset of markers (SNPs) on GS accuracy

To test the impact of the number of SNPs on the accuracy of genomic prediction models, 15 subsets of SNPs (10, 25, 50, 100, 250, 500, 750, 1 K, 2 K, 4 K, 6 K, 8 K, 10 K, 50 K, and 100 K) and two different types of sampling strategies: 1) randomly selected SNP subsets and 2) SNP subsets selected with the largest positive effects were implemented. The single marker effects were estimated using single-marker regression for association testing in training data set. To examine the impact of relatedness on the number of markers, we used full-sib and half-sib family structures between training and validation populations.

Selection response with genomic selection

Selection response could be calculated as the ratio between selection accuracy and breeding cycle length in years.

The relative efficiency of GS to traditional BLUP-based selection (TBS) is

$$RE = \frac{r(\text{EBVs}_{\text{GS}}, \text{EBVs})}{r(\text{EBVs}_{\text{TBS}}, \text{EBVs})}$$

Thus, the relative efficiency of GS to TBS per year is

$$RE/\text{year} = \frac{r(\text{EBVs}_{\text{GS}}, \text{EBVs})}{r(\text{EBVs}_{\text{TBS}}, \text{EBVs})} \times \frac{T_{\text{TS}}}{T_{\text{GS}}}$$

Where EBVs are the estimated breeding values using the full data from equation (1), and T_{TBS} and T_{GS} are the breeding cycle lengths under TBS and GS, respectively [15]. We assumed that the T_{GS} is reduced to 12.5 years from the current approximately 25 years of a breeding cycle by omitting or reducing progeny testing time about 10–15 years.

Results

Heritability and type-B genetic correlation

Heritabilities of the tree height based on GBLUP from two single sites and the joint sites were higher than those from ABLUP models (Table 1). This is in contrast with the wood quality traits (Pilodyn, velocity, and MOE) where all heritabilities from the GBLUP models were lower than those obtained from the ABLUP models. For example, heritability of Pilodyn (0.34) from the GBLUP model was smaller than that from the ABLUP model (0.41).

The type-B genetic correlations for tree height estimated from ABLUP and GBLUP models (as 0.48 and 0.41, respectively), were significantly different from 1, however, the type-B genetic correlations of wood quality traits (0.80–0.94) were higher and statistically were not significantly different from 1.

Accuracy of different statistical methods and the size of training set

Estimates of accuracy were obtained using different statistical methods and for different ratios of TS/VS for each of four traits (Fig. 1). It was observed that ABLUP had higher accuracy than that using the four genomic selection methods (GBLUP, BRR, BLASSO, and RKHS) for tree height, Pilodyn, velocity, and MOE. Tree height had higher accuracy than the three wood quality traits using ABLUP and GS (Fig. 1 and Table 2). Among the four GS methods, GBLUP, BLASSO, and RKHS had approximately similar accuracies, but a slightly higher accuracy was observed when BRR was implemented for tree height. Nevertheless, these four GS methods had little differences on accuracy for the three wood quality traits.

GS accuracy increased as the ratio of training to validation population size increased. However, for GS methods, the maximum accuracy was reached at the ratio of 5:1 between training to validation populations for tree height while maximum accuracy seemed to change minimally after the 3:1 ratio for wood quality traits.

Impact of the number of families on GS accuracy

We estimated the effect of the number of families on GS. The accuracies of all four traits increased when the number of families increased from 10 to 120 families in both the ABLUP and GBLUP model building (Fig. 2). PA had a similar trend, except for tree height. PA of tree height increased from 10 to 30 families and then had similar values up to 120 families in the model building.

Impact of the number of trees per family on GS accuracy

In this study, all 128 full-sib families planted in both progeny trials were selected. Five trees per family in site 1 (Vindeln) and 20 trees per family were selected if there were enough trees for some families in site 2 (Hädanberg).

Table 1 Estimates of variance components and narrow-sense heritabilities from the conventional pedigree-based relationship matrix model (ABLUP) and genomic-based relationship matrix model (GBLUP) in two single sites and a joint-site analysis

Trait		Model	Additive ^a	Residual ^b	Heritability (SE ^c)	Type-B ^d
Height	Site1	ABLUP	690.8	5260.5	0.12 (0.06)	
		GBLUP	902.2	5064.1	0.15 (0.06)	
	Site2	ABLUP	2007.1	8604.5	0.19 (0.06)	
		GBLUP	2140.6	8461.8	0.20 (0.06)	
	Joint-site	ABLUP	1104.2	7576.4	0.13 (0.04)	0.48
		GBLUP	1351.8	7393.5	0.15 (0.05)	0.41
Pilodyn	Site1	ABLUP	2.3	3.2	0.42 (0.11)	
		GBLUP	1.7	3.6	0.31 (0.08)	
	Site2	ABLUP	2.3	2.7	0.46 (0.10)	
		GBLUP	1.9	2.9	0.39 (0.07)	
	Joint-site	ABLUP	2.2	3.0	0.41 (0.09)	0.88
		GBLUP	1.7	3.3	0.34 (0.06)	0.90
Velocity	Site1	ABLUP	0.039	0.036	0.52 (0.11)	
		GBLUP	0.036	0.037	0.49 (0.08)	
	Site2	ABLUP	0.034	0.039	0.47 (0.10)	
		GBLUP	0.026	0.044	0.37 (0.07)	
	Joint-site	ABLUP	0.033	0.041	0.45 (0.09)	0.88
		GBLUP	0.027	0.045	0.37 (0.06)	0.80
MOE	Site1	ABLUP	8.0	8.5	0.48 (0.11)	
		GBLUP	6.2	9.7	0.39 (0.08)	
	Site2	ABLUP	5.8	6.4	0.47 (0.10)	
		GBLUP	4.7	6.9	0.40 (0.07)	
	Joint-site	ABLUP	6.2	7.7	0.44 (0.09)	0.88
		GBLUP	4.7	8.5	0.36 (0.06)	0.94

^aAdditive is the additive genetic variance^bResidual is the residual variance^cSE is the standard error^dType-B represents type-B genetic correlation

Thus, in joint-site cross-validation model, maximum 20 trees per family were tested. Accuracies and PA from ABLUP for all the traits were higher than those from GBLUP when we randomly selected a subset trees per family as TS (Fig. 3). It was also observed that accuracy and PA had similar increased trends as the number of trees within-family increased, but it flattened (stabilized) as the tree numbers reached between 6 and 14, depending on method (e.g. ABLUP and GBLUP) and traits. The accuracies of tree height, Pilodyn, velocity, and MOE increased initially from 0.48, 0.39, 0.52, and 0.41 to 0.84, 0.64, 0.78, and 0.72, respectively, and then stabilized after tree number reached 14, 6, 12, and 12 per family for ABLUP, respectively. The GS accuracies of tree height, Pilodyn, velocity, and MOE also stabilized after the tree number reached 18, 6, 10 and 10 per families for GBLUP, respectively. This may indicate that more trees within a family (16–19) are required for a reliable training set in GS for growth trait than for wood quality traits (6–12).

Testing site effect on GS accuracy

Accuracy and PA of ABLUP and GBLUP were listed in Table 2 for three selection scenarios: Within-site training and selection, cross-site training and selection (e.g. training based on one site while selection for the second site) and joint-site training and selection using full-sib family structure. For all four traits, accuracy of within-site training and selection were always higher than the second scenario of cross-site training and selection. However, the accuracy differences between within-site and cross-site were larger for tree height than for three wood quality traits. For example, the average accuracy of two within-site selections is 0.76 relative to 0.44 for average across-site for tree height from GBLUP. However, for MOE, the average accuracy of two within-site selection is 0.64 relative to 0.59 for average cross-site selection. PA had a similar pattern. The joint-site model accuracies from both ABLUP and GBLUP were the higher than those of within-site and cross-site training and selection.

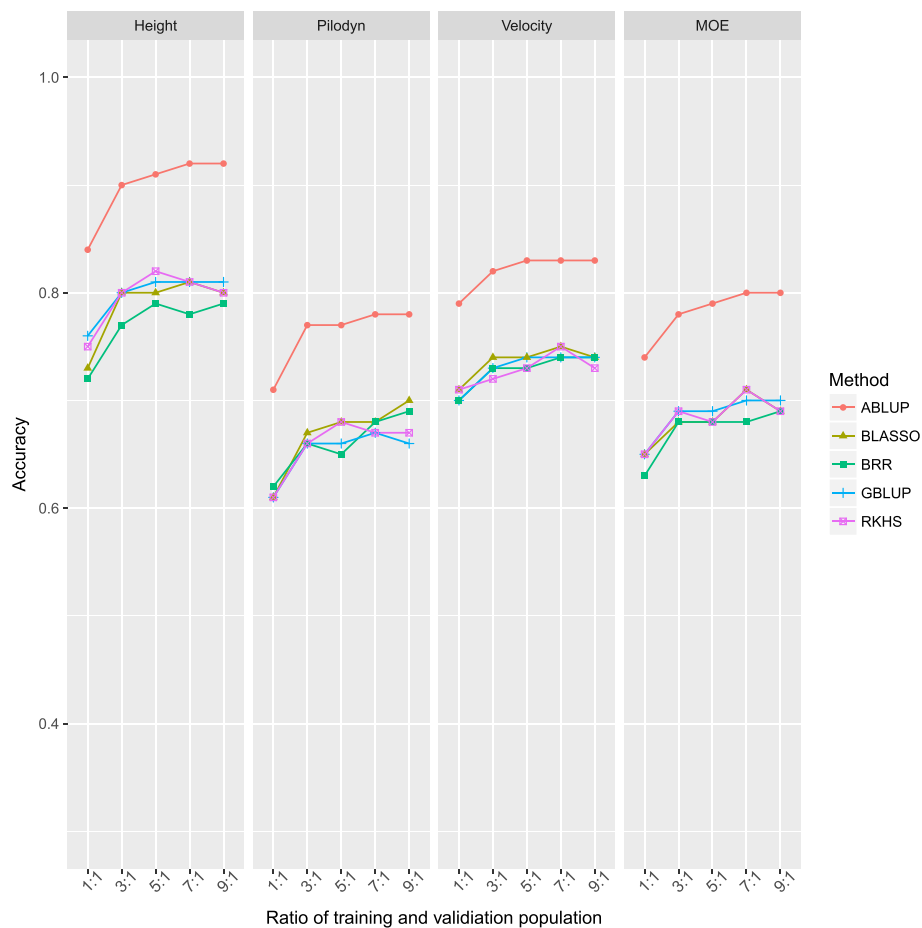


Fig. 1 Accuracy of different methods and increasing ratios of training set (TS) and validation set (VS)

Especially, for instance, tree height accuracy in the joint-site (0.81) was higher than average of within-site (0.76) and average of cross-site (0.44). Similarly, PAs from the joint-site model for all four traits were higher than cross-site training and selection model, but were not always higher than within-site training and selection model. For instance, PAs from ABLUP (0.26) and GBLUP (0.23) models in site 2 were higher than those from the joint-site (0.20 in both models) for tree height.

Relative efficiency (RE) of GBLUP to ABLUP was lower than 1 for all the selection scenarios and traits, ranging from 0.80 to 0.95 with an average of 0.88. However, RE per year (assuming halving a breeding cycle time) reached from 1.60 to 1.89 with an average of 1.76 and were not related to any traits and selection scenarios.

Relatedness

Compared with the full-sib family structure, GS models built with a half-sib family structure led to a considerable decrease in accuracy and PA (Table 3). For instance, in the half-sib family structure, GS accuracy and PA from GBLUP model decreased from 0.81 and 0.20 to 0.55 and

0.11, respectively for tree height and from 0.69 and 0.36 to 0.50 and 0.26, respectively for MOE. However, both RE and RE per year changed little between half-sib and full-sib structure in GS selection. For example, RE and RE per year increased slightly from 0.89 and 1.78 to 0.98 and 1.97, respectively for velocity from full-sib to half-sib population while RE and REs per year decreased slightly from 0.89 and 1.78 to 0.83 and 1.67, respectively from full-sib to half-sib population for tree height.

Compared with half-sib family structure, GS models built with an unrelated family between TS and VS had a considerable decrease in accuracy and PA from GBLUP (Table 3). For example, in the unrelated family structure, GS accuracy and PA from GBLUP model decreased from 0.55 and 0.11 to 0.24 and 0.06, respectively for tree height and from 0.50 and 0.26 to 0.19 and 0.10, respectively for MOE. Especially for velocity, GS accuracy and PA from GBLUP model with a marked decrease from 0.62 and 0.32 to 0.09 and 0.02, respectively, was observed. However, it is worth to note that ABLUP models built with unrelated family had zero accuracies between training and validation populations.

Table 2 Accuracy, predictive ability (PA), relative efficiency (RE), and relative efficiency per year (RE per year) based on all the markers and five genomic selection scenarios for height, Pilodyn, velocity, and MOE

Trait	GS scenario	Accuracy (Error)		Predictive ability (Error)		RE	RE per year
		ABLUP	GBLUP	ABLUP	GBLUP	GBLUP/ABLUP	GBLUP/ABLUP
Height	Site1	0.82 (0.01)	0.74 (0.02)	0.15 (0.03)	0.16 (0.04)	0.91	1.81
	Site2	0.89 (0.01)	0.77 (0.02)	0.26 (0.03)	0.23 (0.03)	0.87	1.73
	Site1 → 2	0.48 (0.04)	0.39 (0.04)	0.10 (0.04)	0.11 (0.03)	0.81	1.63
	Site2 → 1	0.54 (0.02)	0.48 (0.02)	0.02 (0.03)	0.01 (0.03)	0.89	1.77
	Joint-site	0.91 (0.00)	0.81 (0.02)	0.20 (0.02)	0.20 (0.02)	0.89	1.78
Pilodyn	Site1	0.69 (0.02)	0.60 (0.03)	0.29 (0.04)	0.24 (0.05)	0.87	1.74
	Site2	0.72 (0.02)	0.58 (0.03)	0.33 (0.03)	0.29 (0.04)	0.80	1.60
	Site1 → 2	0.57 (0.03)	0.52 (0.03)	0.28 (0.03)	0.28 (0.03)	0.92	1.84
	Site2 → 1	0.58 (0.02)	0.52 (0.02)	0.24 (0.04)	0.23 (0.03)	0.90	1.80
	Joint-site	0.77 (0.01)	0.66 (0.01)	0.32 (0.01)	0.30 (0.02)	0.86	1.71
Velocity	Site1	0.75 (0.01)	0.70 (0.02)	0.42 (0.03)	0.44 (0.04)	0.93	1.87
	Site2	0.80 (0.01)	0.69 (0.02)	0.41 (0.02)	0.37 (0.03)	0.86	1.72
	Site1 → 2	0.72 (0.02)	0.63 (0.02)	0.40 (0.03)	0.36 (0.03)	0.88	1.76
	Site2 → 1	0.69 (0.02)	0.60 (0.03)	0.37 (0.02)	0.34 (0.03)	0.87	1.75
	Joint-site	0.83 (0.01)	0.74 (0.01)	0.43 (0.02)	0.41 (0.02)	0.89	1.78
MOE	Site1	0.70 (0.01)	0.64 (0.03)	0.34 (0.03)	0.35 (0.05)	0.91	1.83
	Site2	0.76 (0.02)	0.64 (0.03)	0.38 (0.04)	0.35 (0.05)	0.84	1.69
	Site1 → 2	0.66 (0.02)	0.59 (0.02)	0.34 (0.02)	0.32 (0.03)	0.89	1.78
	Site2 → 1	0.61 (0.03)	0.58 (0.02)	0.29 (0.04)	0.31 (0.03)	0.95	1.89
	Joint-site	0.79 (0.01)	0.69 (0.03)	0.38 (0.02)	0.36 (0.03)	0.87	1.75

Impact of the number of SNPs on GS accuracy

Accuracy and PA using subsets of markers with the largest positive effects were slightly higher than those using subsets of random markers until the subset of random markers reached 100 K SNPs (Fig. 4). Accuracy and PA using a subset of random markers increased with the increase in the number of markers, until all the markers were included, except that tree height accuracy stabilized with 4 K SNPs. However, accuracy and PA using the subset of markers with the largest positive effects showed different trends. It increased initially, then stabilized and finally decreased until it got to the same level as the random markers selection at the highest number of markers of 100 K SNPs. The use of subsets of markers with the largest positive effects had higher accuracy and PA than random markers until most of the markers were used. For example, PA using a subset of markers with the largest positive effects increased initially from 0.10 to 0.23 with 1 K SNPs and then decreased to 0.19 with 10 K SNPs for tree height. Trait had a similar influence on the number of markers to reach the plateau of the accuracy and PA using a subset of markers with the highest positive effects. The accuracy and PA reached a plateau when the number of markers with the highest positive effects increased to 4 K–6 K for all the traits.

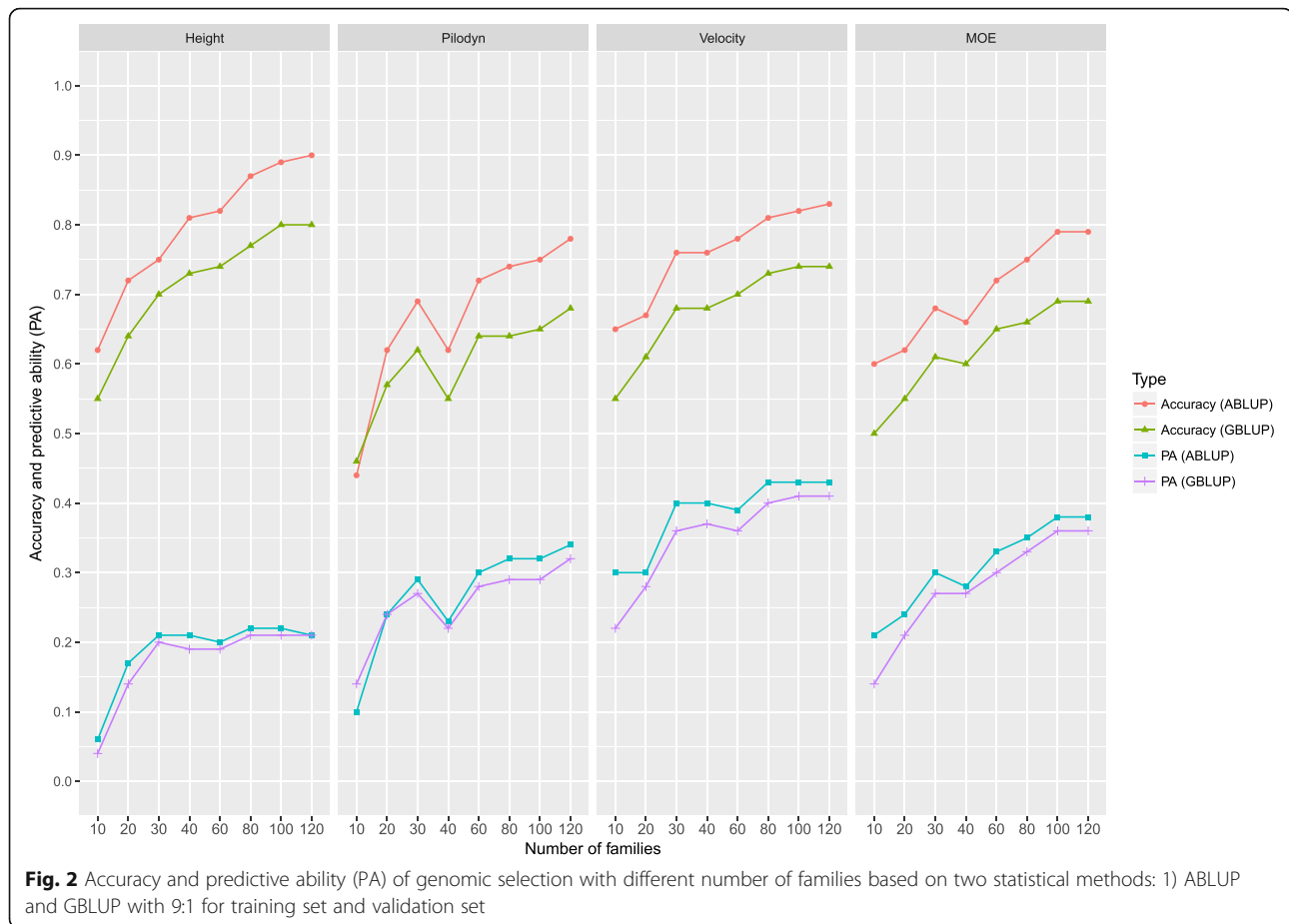
Impact of the number of SNPs and relatedness on GS accuracy

It required fewer number of markers in the full-sib family structure than the half-sib family structure to reach the same accuracy and PA (Fig. 5). For example, the accuracies observed with 250 markers in the full-sib family structure were similar as using all 100 K markers in the half-sib family structure for tree height, Pilodyn, velocity, and MOE. The same PA with random 750 markers used in the full-sib family structure for all four traits required at least 100 K markers in half-sib family structure.

Discussion

Heritability estimates

Heritability is an essential genetic parameter in selective breeding and its values are dependent on the relative contributions of genetic and environmental variations, and vary among traits [43]. In our study, heritabilities for wood quality traits were higher than those for tree height, which is expected and agrees with previous reports for Norway spruce [1, 27]. Tan et al. [4] reported that heritability estimates obtained from GBLUP were higher than those from ABLUP for growth and wood quality traits in *Eucalyptus*. In contrast, Lenz et al. [13] and Gamal EL-Dien et al. [11] reported that heritability

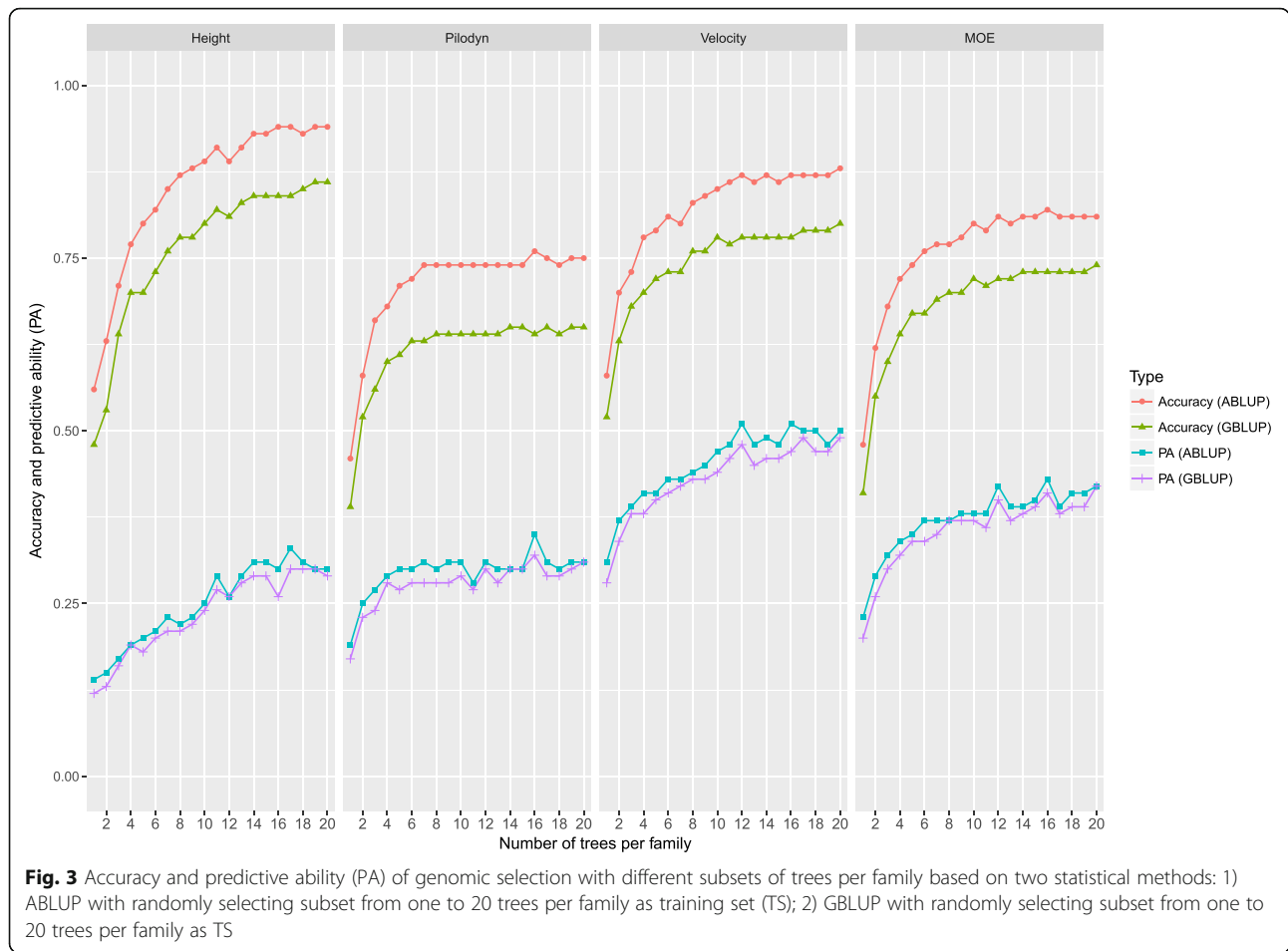


estimates obtained from GBLUP were lower than those from ABLUP for similar growth and wood quality traits in black spruce and interior spruce (*Picea glauca* [Moench] Voss × *Picea engelmannii* Parry ex Engelm.). In this study, the heritability estimates for tree height obtained from GBLUP were slightly higher than those from ABLUP, but there is no significant difference from ABLUP considering the estimated standard errors. The heritability estimates for wood quality traits obtained from GBLUP were slightly lower than those from ABLUP. The above two opposite situations indicate that pedigree-based ABLUP model may inflate heritability estimates for tree height and deflate heritability estimates for wood quality traits if estimates from GBLUP reflect true genetic relationships among families and account for Mendelian segregation within families.

The impact of heritability on the accuracy seems to be low in this study, in line with the report in Douglas-fir [18] and interior spruce [11]. In our joint-site analyses, the heritabilities of tree height, Pilodyn, velocity, and MOE were low to moderate (0.15, 0.34, 0.37, and 0.36, respectively), but the accuracies were high (0.81, 0.66, 0.74, and 0.83,

respectively). Several factors may explain this: (1) the large sample size of 1370 and the relatively small effective population size ($N_e = 55$) likely negate the effect of low trait heritability on prediction accuracy [18]. Märtens et al. [44] demonstrated that increasing relatedness between training and validation populations leads to high prediction accuracy on yeast; (2) the accuracy is the correlation between EBVs assuming as true breeding values from ABLUP and GEBVs from GBLUP in VS in that heritability may affect PAs of ABLUP and GBLUP, but may have little influence on the correlation; (3) the accuracy estimate only represents the additive genetic effect. We found that PA is more similar to the narrow-sense heritability because PA involved both phenotypic and genetic values. For example, heritability of MOE from GBLUP model (0.36) is the same as from ABLUP model (0.36).

In the present study, tree height accuracy (0.81 from GBLUP in Table 2) in full-sib family structure with 1233 individuals in TS was similar as in the deterministic simulation with 50 QTLs and heritability of 0.2 [15]. Wood quality traits were similar to that in the simulation with 100 QTLs and a heritability of 0.4.



Different GS methods show similar results

As expected, the accuracies of four different genomic statistical methods did not clearly outperform each other. This contrasts with previous evaluation of the RKHS method that was reported to outperform other GS methods for low heritability growth traits [4]. It was usually observed that different genomic statistical methods produce similar results for growth and wood quality traits in other forest trees species [10, 18, 45]. With an exception, Resende et al. [9] compared Ridge Regression-BLUP (RR-BLUP), Bayes A, C π , and

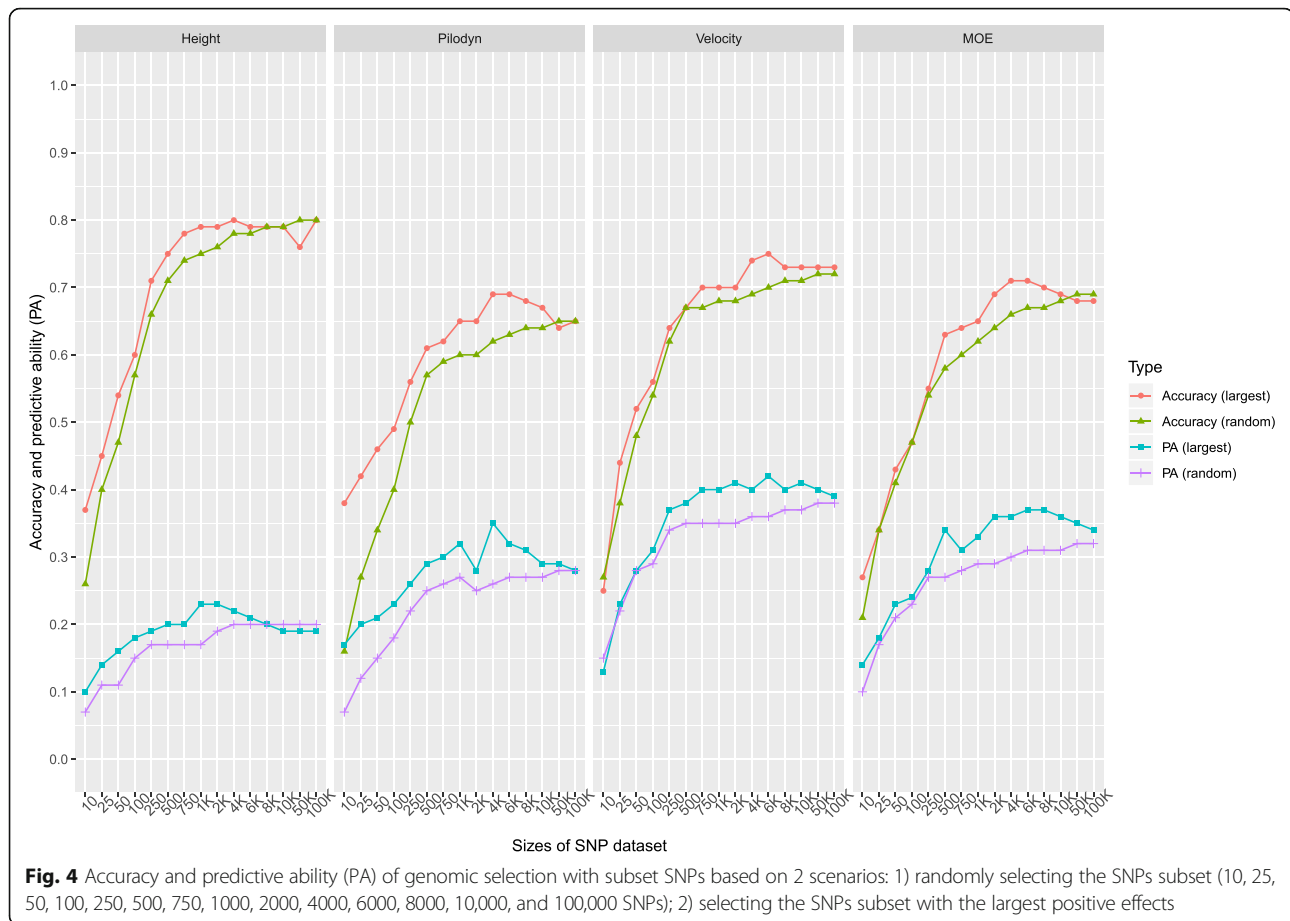
BLASSO for 17 traits in loblolly pine, and found that Bayes A and Bayes C π have higher PA than RR-BLUP and BLASSO for fusiform rust disease-resistance trait. They attribute this to a few genes with large effects that control disease resistance. When the number of markers increases the computational time for Bayesian methods take longer time to converge. Therefore, we also support the proposal that GBLUP is an effective method in providing the best compromise between computational time and prediction efficiency if there are no major gene effects [4, 46].

Table 3 Accuracy, predictive ability (PA), relative efficiency (RE), and relative efficiency per year (RE per year) of genomic selection model based on half-sib families and unrelated families using all markers in a joint-site analysis

Trait	Accuracy (error)			Predictive ability (error)			RE	RE per year
	ABLUP	GBLUP ^a	GBLUP ^b	ABLUP	GBLUP ^a	GBLUP ^b		
Height	0.66 (0.03)	0.55 (0.04)	0.24 (0.10)	0.16 (0.04)	0.11 (0.04)	0.06 (0.06)	0.83	1.67
Pilodyn	0.59 (0.03)	0.44 (0.04)	0.21 (0.05)	0.23 (0.04)	0.19 (0.04)	0.12 (0.04)	0.75	1.49
Velocity	0.63 (0.02)	0.62 (0.03)	0.09 (0.09)	0.29 (0.03)	0.32 (0.03)	0.02 (0.06)	0.98	1.97
MOE	0.59 (0.02)	0.50 (0.03)	0.19 (0.09)	0.28 (0.03)	0.26 (0.03)	0.10 (0.06)	0.85	1.69

^arepresents GBLUP model based on half-sib family and

^brepresents GBLUP model based on unrelated family in training and validation population. The results from ABLUP models based on unrelated families did not present in the table because they were zero



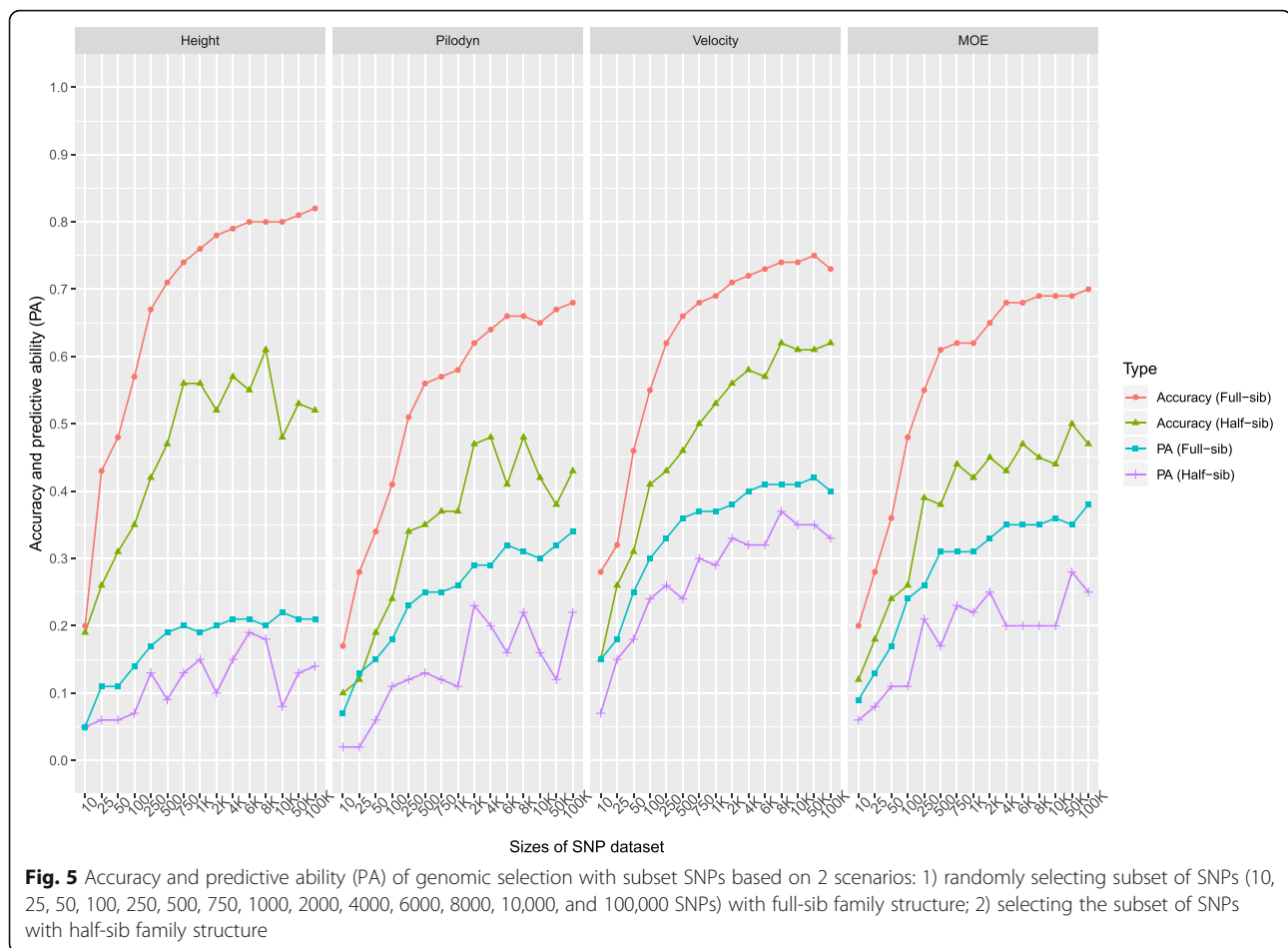
The effect of training dataset and the number of trees per family on GS accuracy

In this study, ratio of TS/VS varied from 1:1 to 9:1. We found that the accuracy improved from the TS/VS ratios of 1:1 to 3:1, but accuracy only improved a little after the TS/VS ratio of 3:1. This is different from other studies [4, 13] showing that an increased ratio of the TS/VS beyond 3:1 still increases the accuracy. However, we found that our result concurs with other reports when the ratio of TS/VS is related to the number of trees per family [13]. In our case, each family under 1:1 ratio of the TS/VS, has 5 trees. There is an average of 10.7 trees per each of the 128 families. The ratios of TS/VS with 1:1, 3:1, 5:1, 7:1, and 9:1 equate to average numbers of TS trees per family of 5.3, 8.0, 8.9, 9.4, and 9.6, respectively. This may indicate that after 8 trees per family on average as TS, there is little increase of GS efficiency for the full-sib family. Based on resampling technique, Perron et al. [47] reported that the number of trees per family has an important effect on the magnitude and precision of genetic parameter estimate. For tree height in that study, at least four trees per family at each site should be included in a half-sib family in order to estimate a more accurate heritability and 4–8 trees/per

family per site still could improve the accuracy of heritability. However, further increase in the number of tree trees per family had little contribution toward increase of accuracy. Wood quality traits usually have higher heritabilities than those of growth traits and also have less $G \times E$. Therefore, wood quality traits may need a lower number of trees than growth traits for obtaining a similar accurate estimate for genetic parameters. Such calculations could guide us to make a more accurate estimate on the number of trees per family required for phenotyping and genotyping.

The effect of the number of families on GS accuracy

The effect of the number of families used in cross-validation test was found important in this study with full-sib family structure. We found that the PA and accuracy increased greatly for all the traits from 10 to 120 families, except the PA for tree height, which stabilized after 30 families for cross-validation test. Based on resampling technique, Perron et al. [47] reported that the number of families has a less important effect on the magnitude and precision of genetic parameter estimate in a half-sib family structure. However, in this study, we found that the number of families is also important for estimates of



GS accuracy and PA. It may be due to the small effective population size (55 unrelated parents) as compared to the study in Perron et al. [47]. One application of GS is in clonal forestry to select the best clones after selection and mating of several best parents (5–10). One question is how to build the training equation for such clonal selection. Should we use progenies of the selected parents or all parents from the trial? From this study, it seems that it is more efficient involving progenies from a larger number of parents (families). Model building and selection using 10 families seem to have lower accuracy than using a larger number of families. It must be tested whether it is due to the small size of families (10–20 trees per family) used in this study, and whether increasing the family size (for example 40 trees per family) as used in clonal progeny testing in Norway spruce in Swedish tree breeding program [2] would increase GS accuracy in a small group of elite families.

Genotype-by-environment interaction

$G \times E$ is usually important for growth traits when the seedlings are planted in different growth environments. We found that tree height in these two northern trials had a

significantly strong $G \times E$, indicated by the type-B genetic correlations (0.48 and 0.41) from ABLUP and GBLUP, respectively (Table 1). Such strong $G \times E$ resulted in a low accuracy and PA when one site is used as TS to predict BVs in another site as VS. A moderate to strong $G \times E$ for growth traits has been reported in several studies in southern and central Sweden [21, 48, 49], but not documented in northern Sweden. Chen et al. [21] reported that within a test series in southern and central Sweden, the averages of type-B genetic correlations varied from 0.60 to 0.89 in 6 test series and their type-B correlations were higher than those in this study. Such strong $G \times E$ in the present study should be considered in model fitting in order to improve PA and accuracy. Several advanced models have been built and tested in crops [50, 51] and one study in tree species [52]. For instance, Oakey et al. [50] used marker and marker by environment interaction in RR-BLUP method to extend the genomic selection to multiple environments.

As expected, we found that wood quality traits have no significant type-B genetic correlations and negligible change of accuracy and PA. A similar result has been reported in two southern Norway spruce open-pollinated

progeny trials [27]. All these indicate that we could use a genomic model in one site to predict GEBV in another site in the same test series for wood quality traits and use of joint-site models could slightly improve the accuracy as in Table 2.

Effect of different family structures

We found that our genomic accuracy for tree height and Pilodyn using half-sib family structure was lower than that reported by Lenz et al. [13] in black spruce, even though more SNPs were used in this study (116 K in 20,695 contigs vs 0.49 K from SNP chip). The difference is likely due to lower heritabilities (i.e. 0.15 vs 0.42 for tree height in GBLUP) and the larger N_e in this study (55 vs 27 unrelated parents). Accuracy and PA decreased from full-sib to half-sib family structure and from half-sib to unrelated family structure, indicating that GS model is more efficient in strongly structured populations where relatedness and LD are higher, and full-sib families also needed less number of markers than half-sib families for obtaining similar accuracy (Fig. 5). Similar results were obtained by other studies [12, 13, 45, 53]. However, the relative efficiency between ABLUP and GBLUP (the accuracy ratio) is more or less similar in both full-sib and half-sib populations, which indicates that GS could be used in both half-sib and full-sib populations. The lower estimates of accuracy and PA (Table 3) obtained from GBLUP for unrelated family structure may be due to the low LD between marker and QTL in the unrelated population, which indicates that GS may not be used in unrelated family structure based on the current exome capture data.

The effect of the number of SNPs and LD in genomic prediction

To our knowledge, this study has used the largest number of SNPs (116,765 SNPs from 20,695 contigs) for GS in tree species [3, 7, 18]. When we used GS in the half-sib family structure, the accuracy and PA reduced about 20%. Moreover, the values of square of PA in half-sib and full-sib family structure accounted for less than 50% of heritability for all traits, which may indicate that even with such a large number of SNPs, we may still have not captured and explained most of the QTL effects. This may be attributed to the low LD in spruce in Fig. 6 (approximate 84 base pairs based on 517 unrelated individuals in Baisson et al. [54]) and only exome regions being used. In humans, the exome constitutes a mere 1% of the whole genome (3Gb) [55]. For Norway spruce, however, the genome size is ca. 20 Gb and LD is lower than Humans [19]. Norway spruce has a mapped genome size of 3326.3 centiMorgan (cM) [33], which is larger compared to ~2100 cM in white spruce (*Picea glauca* (Moench) Voss) [56]. There were about 5.6 SNPs per contig/gene on average based on 20,695 contigs/

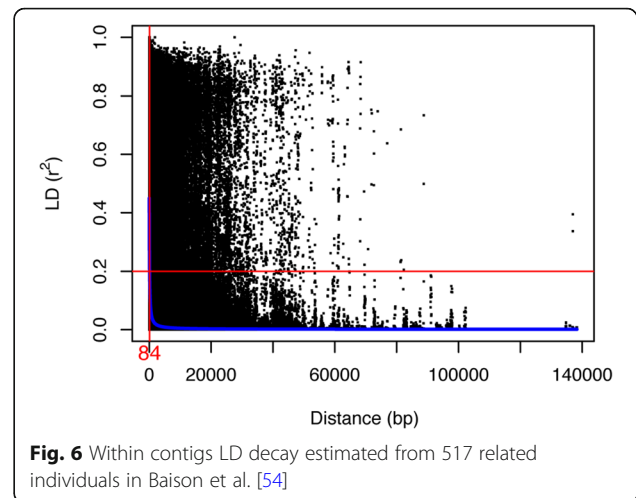


Fig. 6 Within contigs LD decay estimated from 517 related individuals in Baisson et al. [54]

genes that our SNPs come from. This translates to an average genome coverage of ~6.2 markers/cM. If we use all independent 77,116 SNPs, then, it amounts to 3.7 SNPs per contig used.

Accuracy and PA obtained from the subset of markers with the largest positive effects were slightly higher than those from a subset of markers based on random selection, implying that using the subset of markers with the largest effects in genomic regions with small LD decay could track relationship more effectively than random markers. It also implies that using the subset of markers with the largest positive effect could also obtain some effects based on part of the short-range LD [13, 45]. Thus, this factor could also be potentially useful to reduce genotyping the number of SNPs and GS cost, in highly structured population when the genome locations of markers are known.

Efficiency of genomic selection

In the present study, we observed that the efficiency of GS per year is greater than that based on traditional progeny test selection if GS is used to half generation time. In traditional pedigree-based progeny selection, the generation time for Norway spruce in Northern Sweden is at least 25 years, which is based on clonal replicated progeny trial of selected breeding trees. The clonal based progeny test procedure includes: seed sowing and growing to a sufficient size (2 yrs), vegetative propagation of seed plants by rooted cuttings (2 yrs), testing in field trials (15 yrs), and assessment of trials (1 yr). The final stage, the completion of a crossing scheme to create the next generation, is about 5 yrs. If we could omit the progeny testing of the first three stages (a total of 19 yrs) and complete flowering induction and mating within 15 years of GS selection, the time for a breeding cycle could be halved for Norway spruce. In this study, we only considered the efficiency of GS based on the timing of the

breeding cycle. Considering the cost for field testing, the benefit of GS might be higher if the cost for establishing and maintaining 3–4 progeny trials in each breeding population is more than genotyping.

Conclusions

As expected, the main advantage of genomic selection is the potential to shorten the breeding cycle. We observed that:

1. Using G matrix slightly altered heritability relative to A matrix with a slight increase for tree height and a decrease for wood quality traits.
2. ABLUP is about 11–14% more efficient than GBLUP for tree height and wood quality traits, while the four GS methods (GBLUP, BRR, BLASSO, RKHS) had a similar accuracy.
3. Efficiency of GS increased from 49 to 97% among four growth and wood quality traits if GS could reduce the generation time for a breeding cycle to half.
4. The GS accuracy improved from the TS/VS ratios of 1:1 to 3:1, but accuracy only improved marginally after a TS/VS ratio of 3:1.
5. The number of families and the number of trees per family also had an effect on GS efficiency. Wood quality traits need fewer number of trees within-family than tree height for a similar GS efficiency.
6. GS accuracy decreased from full-sib to half-sib family structure and from half-sib to unrelated family structure. The number of markers need to be increased greatly for half-sib family structure to have a similar efficiency to the full-sib family structure.
7. GS accuracy increased as the number of markers increased. The accuracy almost reached a plateau when the number of markers increased to 4 K–8 K for all the traits.

Abbreviations

ABLUP: Pedigree-based best linear unbiased prediction; BLASSO: Bayesian LASSO regression; BRR: Bayesian ridge regression; DP: Read depth; EBV: Estimated breeding value; G × E: Genotype-by-environment interaction; GATK: Genome Analysis Toolkit; GBLUP: Genomic best linear unbiased prediction; GEBV: Genomic breeding values; GQ: Genotype quality; GS: Genomic selection; LD: Linkage disequilibrium; MAF: Minor allele frequency; MFA: Microfibril angle; MOE: Modulus of elasticity; PA: Predictive ability; Pilodyn: Pilodyn penetration; QTL: Quantitative trait locus; RE: Relative efficiency; RKHS: Reproducing Kernel Hilbert Space; SNP: Single nucleotide polymorphism; TS: Training set; Velocity: Acoustic velocity; VQSR: Variant quality score recalibration; VS: Validation set

Acknowledgements

The computations were performed on resources by the Swedish National Infrastructure for Computing (SNIC) at UPPMAX and HPC2N. We thank Dr. Junjie Zhang, Tianyi Liu, and Ms. Xinyu Chen, Linghua Zhou for help of the DNA extraction and field assistance and Anders Fries for field work.

Funding

Financial support was received from Formas (grant number 230–2014-427) and the Swedish Foundation for Strategic Research (SSF, grant number RBP14–0040). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Availability of data and materials

The datasets supporting the conclusions of this article are available upon request.

Authors' contributions

ZQC designed sampling strategy, coordinated field sampling, analyzed data, and drafted the manuscript. BA, BK, and JW participated in the selection of the breeding populations, providing access to field experiments, tree height data and edited the manuscript. JB, JP, and MRGG participated in collection of phenotypic data, extraction of the DNA, SNP calling, and editing of the manuscript. HXW conceived and designed the study, and assisted writing of the manuscript. All authors read and approved the final manuscript.

Ethics approval and consent to participate

The plant materials analyzed for this study comes from common garden experiments (Plantation and clonal archives) that were established and maintained by the Forestry Research Institute of Sweden (Skogforsk) for breeding selections and research purposes. Three tree breeders in Sweden were coauthors in this paper. They agreed to access the materials.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Publisher's Note

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

Author details

¹Umeå Plant Science Centre, Department of Forest Genetics and Plant Physiology, Swedish University of Agricultural Sciences, SE-90183 Umeå, Sweden. ²Skogforsk, Ekebo 2250, SE-268 90 Svalöv, Sweden. ³Skogforsk, Box 3, SE-918 21 Sävar, Sweden. ⁴CSIRO NRCA, Black Mountain Laboratory, Canberra, ACT 2601, Australia.

Received: 19 March 2018 Accepted: 16 November 2018

Published online: 18 December 2018

References

1. Hannrup B, Cahalan C, Chantre G, Grabner M, Karlsson B, Le Bayon I, et al. Genetic parameters of growth and wood quality traits in *Picea abies*. *Scand J For Res.* 2004;19(1):14–29.
2. Karlsson B, Rosvall O. Progeny testing and breeding strategies. *Proceedings of the Nordic group for tree breeding.* Edinburgh; 1993.
3. Tan B, Grattapaglia D, Wu HX, Ingvarsson PK. Genomic relationships reveal significant dominance effects for growth in hybrid *Eucalyptus*. *Plant Sci.* 2018;267:84–93.
4. Tan B, Grattapaglia D, Martins GS, Ferreira KZ, Sundberg B, Ingvarsson PK. Evaluating the accuracy of genomic prediction of growth and wood traits in two *Eucalyptus* species and their F1 hybrids. *BMC Plant Biol.* 2017;17(1):110.
5. Resende MDV, Resende MFR, Sansaloni CP, Petrolini CD, Missiaggia AA, Aguiar AM, et al. Genomic selection for growth and wood quality in *Eucalyptus*: capturing the missing heritability and accelerating breeding for complex traits in forest trees. *New Phytol.* 2012;194(1):116–28.
6. Isik F, Bartholomé J, Farjat A, Chancerel E, Raffin A, Sanchez L, et al. Genomic selection in maritime pine. *Plant Sci.* 2016;242:108–19.
7. Bartholomé J, Van Heerwaarden J, Isik F, Boury C, Vidal M, Plomion C, et al. Performance of genomic prediction within and across generations in maritime pine. *BMC Genomics.* 2016;17(1):604.
8. Zapata-Valenzuela J, Whetten RW, Neale D, McKeand S, Isik F. Genomic estimated breeding values using genomic relationship matrices in a cloned population of loblolly pine. *G3 Genes Genom Genet.* 2013;3(5):909–916.
9. Resende MFR, Muñoz P, Resende MDV, Garrick DJ, Fernando RL, Davis JM, et al. Accuracy of genomic selection methods in a standard data set of loblolly pine (*Pinus taeda* L.). *Genetics.* 2012;190(4):1503–10.
10. Ratcliffe B, El-Dien OG, Klápště J, Porth I, Chen C, Jaquish B, et al. A comparison of genomic selection models across time in interior spruce (*Picea engelmannii* × *glauca*) using unordered SNP imputation methods. *Heredity.* 2015;115(6):547–55.

11. Gamal El-Dien O, Ratcliffe B, Klápště J, Chen C, Porth I, El-Kassaby YA. Prediction accuracies for growth and wood attributes of interior spruce in space using genotyping-by-sequencing. *BMC Genomics*. 2015;16(1):1–16.
12. Beaulieu J, Doerksen TK, Clement S, MacKay J, Bousquet J. Accuracy of genomic selection models in a large population of open-pollinated families in white spruce. *Heredity*. 2014;113(4):343–52.
13. Lenz PRN, Beaulieu J, Mansfield SD, Clément S, Despots M, Bousquet J. 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*. 2017;18(1):335.
14. Hayes BJ, Bowman PJ, Chamberlain AJ, Goddard ME. Invited review: genomic selection in dairy cattle: Progress and challenges. *J Dairy Sci*. 2009;92(2):433–43.
15. Grattapaglia D, Resende MDV. Genomic selection in forest tree breeding. *Tree Genet Genomes*. 2011;7(2):241–55.
16. Hall D, Hallingbäck HR, Wu HX. Estimation of number and size of QTL effects in forest tree traits. *Tree Genet Genomes*. 2016;12(6):110.
17. Resende MFR, Muñoz P, Acosta JJ, Peter GF, Davis JM, Grattapaglia D, et al. Accelerating the domestication of trees using genomic selection: accuracy of prediction models across ages and environments. *New Phytol*. 2012; 193(3):617–24.
18. Thistlethwaite FR, Ratcliffe B, Klápště J, Porth I, Chen C, Stoehr MU, et al. Genomic prediction accuracies in space and time for height and wood density of Douglas-fir using exome capture as the genotyping platform. *BMC Genomics*. 2017;18(1):930.
19. Mackay J, Dean JFD, Plomion C, Peterson DG, Canovas FM, Pavy N, et al. Towards decoding the conifer giga-genome. *Plant Mol Biol*. 2012;80(6):555–69.
20. Nystedt B, Street NR, Wetterbom A, Zuccolo A, Lin Y-C, Scofield DG, et al. The Norway spruce genome sequence and conifer genome evolution. *Nature*. 2013;497:579–84.
21. Chen Z-Q, Karlsson B, Wu HX. Patterns of additive genotype-by-environment interaction in tree height of Norway spruce in southern and Central Sweden. *Tree Genet Genomes*. 2017;13(1):25.
22. Cullis BR, Jefferson P, Thompson R, Smith AB. Factor analytic and reduced animal models for the investigation of additive genotype-by-environment interaction in outcrossing plant species with application to a *Pinus radiata* breeding programme. *Theor Appl Genet*. 2014;127(10):2193–210.
23. Wu HX, Matheson AC. Genotype by environment interactions in an Australia-wide radiata pine diallel mating experiment: implications for regionalized breeding. *For Sci*. 2005;51(1):29–40.
24. Baltunis BS, Gapare WJ, Wu HX. Genetic parameters and genotype by environment interaction in radiata pine for growth and wood quality traits in Australia. *Silvae Genet*. 2010;59:113–24.
25. Gapare WJ, Ivković M, Baltunis BS, Matheson CA, Wu HX. Genetic stability of wood density and diameter in *Pinus radiata* D. Don plantation estate across Australia. *Tree Genet Genomes*. 2010;6(1):113–25.
26. Chen Z-Q, Karlsson B, Mörling T, Olsson L, Mellerowicz EJ, Wu HX, et al. Genetic analysis of fiber dimensions and their correlation with stem diameter and solid-wood properties in Norway spruce. *Tree Genet Genomes*. 2016;12(6):123.
27. Chen Z-Q, Karlsson B, Lundqvist S-O, García Gil MR, Olsson L, Wu HX. Estimating solid wood properties using Pilodyn and acoustic velocity on standing trees of Norway spruce. *Ann For Sci*. 2015;72(4):499–508.
28. Vidalis A, Scofield DG, Neves LG, Bernhardsson C, García-Gil MR, Ingvarsson P. Design and evaluation of a large sequence-capture probe set and associated SNPs for diploid and haploid samples of Norway spruce (*Picea abies*). *bioRxiv*. 2018.
29. Langmead B, Salzberg SL. Fast gapped-read alignment with bowtie 2. *Nat Methods*. 2012;9:357.
30. Li H, Durbin R. Fast and accurate short read alignment with burrows-wheeler transform. *Bioinformatics*. 2009;25(14):1754–60.
31. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The sequence alignment/map format and SAMtools. *Bioinformatics*. 2009;25(16):2078–9.
32. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytzky A, et al. The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res*. 2010;20(9):1297–303.
33. Bernhardsson C, Vidalis A, Wang X, Scofield DG, Shiffthaler B, Baison J, et al. An ultra-dense haploid genetic map for evaluating the highly fragmented genome assembly of Norway spruce (*Picea abies*). *bioRxiv*. 2018.
34. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet*. 2007;81(3):559–75.
35. Money D, Gardner K, Migicovsky Z, Schwaninger H, Zhong G-Y, Myles S. LinkImpute: fast and accurate genotype imputation for nonmodel organisms. *G3 Genes Genom Genet*. 2015;5(11):2383–90.
36. Bradbury PJ, Zhang Z, Kroon DE, Casstevens TM, Ramdoss Y, Buckler ES. TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics*. 2007;23(19):2633–5.
37. R Core Team. R: a language and environment for statistical Computing R Foundation for Statistical Computing. Vienna; 2014.
38. Butler DG, Cullis BR, Gilmour AR, Gogel BJ. ASReml-R reference manual. The State of Queensland, Department of Primary Industries and Fisheries. Brisbane; 2009.
39. Pérez P, de Los Campos G. Genome-wide regression & prediction with the BGLR statistical package. *Genetics*. 2014;198(2):483–95.
40. Wimmer V, Albrecht T, Auinger H-J, Schoen C-C. Synbreed: a framework for the analysis of genomic prediction data using R. *Bioinformatics*. 2012;28(15): 2086–7.
41. Hoerl AE, Kennard RW. Ridge regression: biased estimation for nonorthogonal problems. *Technometrics*. 1970;12(1):55–67.
42. De Los Campos G, Gianola D, Rosa GJ, Weigel KA, Crossa J. Semi-parametric genomic-enabled prediction of genetic values using reproducing kernel Hilbert spaces methods. *Genet Res*. 2010;92(4):295–308.
43. Falconer D, Mackay T. Introduction to quantitative genetics. 4th ed. New York: Longman; 1996.
44. Märtens K, Hallin J, Warringer J, Liti G, Parts L. Predicting quantitative traits from genome and phenome with near perfect accuracy. *Nat Commun*. 2016;7:11512.
45. Beaulieu J, Doerksen TK, MacKay J, Rainville A, Bousquet J. Genomic selection accuracies within and between environments and small breeding groups in white spruce. *BMC Genomics*. 2014;15(1):1048.
46. Lorenz AJ, Chao S, Asoro FG, Heffner EL, Hayashi T, Iwata H, et al. Genomic selection in plant breeding: knowledge and prospects. In: *Advances in agronomy*, vol. 110; 2011. p. 77–123.
47. Perron M, DeBlois J, Despots M. Use of resampling to assess optimal subgroup composition for estimating genetic parameters from progeny trials. *Tree Genet Genomes*. 2013;9(1):129–43.
48. Kroon J, Ericsson T, Jansson G, Andersson B. Patterns of genetic parameters for height in field genetic tests of *Picea abies* and *Pinus sylvestris* in Sweden. *Tree Genet Genomes*. 2011;7(6):1099–111.
49. Berlin M, Jansson G, Högborg K-A. Genotype by environment interaction in the southern Swedish breeding population of *Picea abies* using new climatic indices. *Scand J For Res*. 2014;30(2):112–21.
50. Oakey H, Cullis B, Thompson R, Comadran J, Halpin C, Waugh R. Genomic selection in multi-environment crop trials. *G3 Genes Genom Genet*. 2016; 6(5):1313–26.
51. Pérez-Rodríguez P, Crossa J, Rutkoski J, Poland J, Singh R, Legarra A, et al. Single-step genomic and pedigree genotypexenvironment interaction models for predicting wheat lines in international environments. *Plant Genome*. 2017;10(2).
52. Ventorim Ferrão LF, Gava Ferrão R, Ferrão MAG, Francisco A. Garcia AAF. A mixed model to multiple harvest-location trials applied to genomic prediction in *Coffea canephora*. *Tree Genet Genomes*. 2017;13(5):95.
53. Zapata-Valenzuela J, Isik F, Maltecca C, Wegrzyn J, Neale D, McKeand S, et al. SNP markers trace familial linkages in a cloned population of *Pinus taeda*—prospects for genomic selection. *Tree Genet Genomes*. 2012;8(6):1307–18.
54. Baison J, Vidalis A, Zhou L, Chen Z-Q, Li Z, Sillanpää MJ, et al. Association mapping identified novel candidate loci affecting wood formation in Norway spruce. *bioRxiv*. 2018.
55. Ng SB, Turner EH, Robertson PD, Flygare SD, Bigham AW, Lee C, et al. Targeted capture and massively parallel sequencing of 12 human exomes. *Nature*. 2009;461:272.
56. Pavy N, Namroud MC, Gagnon F, Isabel N, Bousquet J. The heterogeneous levels of linkage disequilibrium in white spruce genes and comparative analysis with other conifers. *Heredity*. 2012;108(3):273–84.