Incorporating genome-wide association study results into genomic prediction enhances the predictive ability of growth and phenology traits in Norway spruce

Zhi-Qiang Chen[1*], Adam Klingberg[2], Henrik R. Hallingbäck[3], Harry X. Wu[1,4]

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

[2]Skogforsk, SE-91821 Sävar, Sweden

[3]Skogforsk, Uppsala Science Park, SE-75183 Uppsala

[4]CSIRO National Collection Research Australia, Black Mountain Laboratory, Canberra, ACT 2601, Australia

[*]Corresponding author

Zhi-Qiang Chen

Email: zhiqiang.chen@slu.se

Abstract:

Genome-wide association (GWA) has been used to detect quantitative trait loci (QTLs) in plant and animal species. Genomic prediction is to predict an accumulative effect of all QTL effects by capturing the linkage disequilibrium between markers and QTLs. Thus, marker preselection is considered a promising method to capture Mendelian segregation effects, especially for an oligogenic trait. Using QTLs detected in the GWA study could improve genomic prediction (GP), including informative marker selection and adding a QTL with the largest effect size as a fixed effect. In this study, we performed GWA and GP studies in a population with 904 clones from 32 full-sib families planted in four trials using a newly developed 50k SNP Norway spruce array. In total, GWAS identified 41 SNPs associated with the budburst stage (BB) and the SNP with the largest effect size explained 5.1% of the phenotypic variation (PVE). For other traits like height at tree ages six and 12, diameter at breast height, frost damage and Pilodyn penetration, only 2 – 13 SNP associations were detected and the PVE of the strongest effects ranged from 1.2% to 2.0%. GP with approximately 100 preselected SNPs based on the smallest *p*-values from GWAS showed the largest predictive ability (PA) for the oligogenic trait BB. But for the other polygenic traits, approximate 2000-4000 preselected SNPs, indicated by the smallest Akaike information criterion (AIC) to offer the best model fit, still resulted in PA being similar to that of GP models using all markers. Analyses on both real-life and simulated data also showed that the inclusion of a large QTL SNP in the model as a fixed effect could improve the PA and accuracy of GP provided that the PVE of the QTL was ≥2.5%.

Introduction

Genomic prediction (GP) using genome-wide dense markers has been widely adopted in animal breeding and extensively studied in crops (Hickey *et al*, 2017) and tree plant species (Grattapaglia, 2022) in the last decade. GP assumes that individual quantitative trait loci (QTL) are linked with at least one DNA marker. Therefore, linkage disequilibrium (LD) between QTLs and markers plays an important role in genomic prediction efficiency (Meuwissen et al. 2001). Most GP results showed that stronger LD between markers and causative mutation resulted in higher accuracy of genomic predictions. It is suggested that training to use multiple generations (tracing LD) or including causative mutations will increase GP efficiency. This is because there is no need to trace the causative mutations with LD markers when the causative mutations are among the genotypes (Meuwissen et al. 2013). Therefore, the inclusion of markers tightly associated with a few large-effect QTLs, detected by genome-wide association (GWA) or validated by gene transformation, could be incorporated into the GP model development. Including such large-effect SNPs as fixed effects in GP modelling is also considered as an ideal approach to increasing GP efficiency, which has been verified by several studies in crops, including simulations (Bernardo, 2014; Bian and Holland, 2017; Sehgal *et al*, 2020). GWA is considered a powerful approach to dissecting the genetic architecture of different traits in animals and plants (Du *et al*, 2019; Tam *et al*, 2019). Usually, a locus accounting for 1-15% of phenotypic variation is often detected by GWA studies if population sizes are large enough, from a few hundred to a few thousand in plants (Hall *et al*, 2016), and a few thousand to a few million in humans (Visscher *et al*, 2017).

Several studies in trees have shown that a few thousand randomly selected SNPs may capture most of the variation, and similar GP efficiency as using all available markers (Chen *et al*, 2018; Estopa *et al*, 2022; Tan *et al*, 2017; Thumma *et al*, 2022). It has also been reported that marker preselection could slightly to moderately improve genomic prediction accuracy in tree species (Cappa *et al*, 2022; Resende *et al*, 2012; Tan and Ingvarsson, 2022). In theory, GP would like to capture all QTL effects by LD between markers and causative loci. Provided that one or two of several strongly linked markers for each QTL is selected, the model should almost capture the QTL effects. Meanwhile, if one can select the tightly linked markers or mutants themselves through GWAS for GP, it is possible to reduce the cost of genotyping while maintain or improve the accuracy of GP. Recently, following the development of high throughput genotyping techniques and tools, more than several hundred thousand markers have been commonly and easily produced by several genotyping platforms, such as SNP array, exome capture, and genotyping-by-sequencing (GBS). Thus, marker preselection could become a very useful and common pre-step for GP.

Norway spruce (*Picea abies*) is one of the most important economic species in Europe, especially in the Nordic countries and Northwestern Russia (Schmidt-Vogt, 1977). The current breeding program of Norway spruce, similar to other conifer species, mainly focuses on the use of additive genetic effects (i.e. breeding values). However, the non-additive genetic effects are considered important if a clonal deployment is considered as a deployment strategy in the future (Chen *et al*, 2020; Nguyen *et al*, 2022; Wu, 2018). Recent several studies on the cost, benefits and genetic diversity with the deployment of clonal forestry for Norway spruce indicated that a considerable productivity increase of planted forests could be achieved with an acceptable genetic diversity (Berlin *et al*, 2019; Bradshaw *et al*, 2019;

Chen *et al*, 2020; Egertsdotter, 2019; Ingvarsson and Dahlberg, 2018; Rosvall, 2019a; Rosvall, 2019b; Rosvall *et al*, 2019; Wu, 2018).

In Norway spruce breeding, traditional breeding values, dominance, and epistatic genetic effects were predicted for seedling or clonally propagated progenies based on the theoretical expectation using pedigree data (Wu, 2018). When genome-wide dense markers are available, the estimates of genetic parameters for additive, dominance, and epistatic effects may be more accurate since the real genomic relationships were captured by alleles (Muñoz *et al*, 2014; Tan *et al*, 2018). Several studies in tree species have demonstrated an increase in the accuracy of genetic parameters estimates using genomic models (Chen *et al*, 2019; Gamal El-Dien *et al*, 2016; Walker *et al*, 2021). The accurate estimation of non-additive effects, as well as additive effects, should therefore be an important objective to increase genetic gains and improve the efficiency of the Norway spruce tree breeding program.

In this study, we explored the use of detected GWA QTLs by including the most closely associated SNPs in GP for tree breeding value prediction through empirical experiments and simulations. The detailed aims of this study were to 1) dissect the total genetic component into additive, dominant, and epistatic variances using genomic-based relationship matrices and compare with the results estimated from models using traditional pedigree-based relationship matrices; 2) test the efficiency using different number of SNPs, different number of clones per family with informative maker preselection; 3) optimize the training population size and population structure for the current Swedish breeding population; and 4) test whether the GP could be improved by including the most significant GWA marker as a fixed effect in the genomic prediction model.

Materials and methods

Plant materials

A Norway spruce breeding population using 32 control-pollinated families from 49 parents was established in 2007 at four different field sites. A total of 1430 unique clones were derived from the 32 families with an average of about 45 clones per family and about three ramets per genotype were planted in each site. The detailed descriptions of the four sites in this breeding population were presented in the published paper (Chen *et al*, 2020). Generally, the field site series were established using a randomized incomplete block design with single-tree plots. Meanwhile, 98% of clones were replicated among the four sites. In this study, we selected all available clones from ten families (ca. 50 clones per family) and 20 clones from each of the remaining 22 families (a total of 904 clones) from a single site S1389 (Rössjöholm) for genotyping (Table 1).

Phenotyping

Tree height was measured at field age six (HT6) at all four sites and at twelve years (HT12) at one of the sites. Diameter at breast height (DBH) was measured at field age twelve at all four sites. Budburst stage (BB) was scored at field age six at three sites based on eight categories (Krutzsch, 1975). Two further traits were measured only in a single site. Pilodyn penetration (PILO) as a proxy of wood density was measured by Pilodyn 6J Forest (PROCEQ, Zurich, Switzerland) at field age ten. Frost damage (FD) was quantified after the site was exposed to a severe frost event at field age six. FD was scored as a categorical variable from zero (without frost damage) to three (the most severe damage). The detailed descriptions of traits measured in each of the four sites are shown in Table 1.

Genotyping

Newly fresh needles were sampled from 904 clones in the spring of 2018. Total genomic DNA was extracted using the Qiagen plant DNA extraction protocol with DNA quantification performed using the Qubit® ds DNA Broad Range Assay Kit (Qiagen, Oregon, USA). Genotypic data were generated using the Norway spruce *Piab50K* SNP array chip (Bernhardsson *et al*, 2021). Genotype calling of the 50K Axiom array was performed using the Axiom analysis suite (V4.0), following best practice with default parameters (a sample call using a Dish-QC threshold [axiom_dishqc_DQC] ≥ 0.82 and an average SNP call rate cutoff per sample [cr-cutoff] ≥ 0.97). In this study, missing SNPs were imputed by Beagle v4.0 (Browning and Browning, 2007). Due to several parents only involving a single controlled cross/family, the unique rare allele of the parent only contributes one allele to each progeny in the family. Thus, SNPs with a minor allele frequency (MAF) less than $M/(2n)$ will be filtered out for all models, where M is the harmonic number of clones per family and n is the population size in each GP model.

Pedigree correction

Since a couple of clear discrepancies between the additive relationship and the genomic relationship matrices ($\boldsymbol{A}$ and $\boldsymbol{G_a}$, respectively) were detected for some individuals, we performed pedigree correction for this population based on $\boldsymbol{A}$ and a heatmap of $\boldsymbol{G_a}$. The number of parents increased from 49 based on the documented pedigree to 55 based on

$\boldsymbol{G_a}$ and the number of families also increased from the original 32 to 56 (Table S1). Finally, the number of clones per family after correction varied from 1 to 56.

Spatial analysis

Spatial analysis based on a two-dimensional separable autoregressive (AR1) model was used to fit the row and column directions for phenotypic data from each site using ASReml v4.1 (Gilmour *et al*, 2015). Block effects were estimated simultaneously and all significant block and spatial effects were removed from the raw data. The spatially adjusted phenotypic data were used for downstream analysis.

Variance component and heritability estimates

Four univariate models were used to estimate variance components for each of three traits (HT6, DBH, and BB) based on pedigree-based best linear unbiased prediction (PBLUP) and genomic-based best linear unbiased prediction (GBLUP) as following:

$$y = \boldsymbol{X}\beta + \boldsymbol{Z_1}a + \boldsymbol{Z_2}as + \varepsilon \quad [1]$$

$$y = \boldsymbol{X}\beta + \boldsymbol{Z_1}a + \boldsymbol{Z_2}as + \boldsymbol{Z_6}r + \boldsymbol{Z_7}rs + \varepsilon \quad [2]$$

$$y = \boldsymbol{X}\beta + \boldsymbol{Z_1}a + \boldsymbol{Z_2}as + \boldsymbol{Z_3}d + \boldsymbol{Z_4}ds + \boldsymbol{Z_6}r + \boldsymbol{Z_7}rs + \varepsilon \quad [3]$$

$$y = \boldsymbol{X}\beta + \boldsymbol{Z_1}a + \boldsymbol{Z_2}as + \boldsymbol{Z_3}d + \boldsymbol{Z_4}ds + \boldsymbol{Z_5}e_{xx} + \boldsymbol{Z_6}r + \boldsymbol{Z_7}rs + \varepsilon \quad [4]$$

where $\boldsymbol{y}$ is the vector of adjusted phenotypic observations of a single trait; β is the vector of fixed effects, including a grand mean and site effects; a and d are the vectors of random additive and dominance effects, respectively; as and ds are the vectors of random additive-by-site and dominance-by-site effects, respectively; e_{xx} is one of $e_{aa}, e_{ad}, and\ e_{dd}$, which are the vectors of random additive-by-additive, additive-by-dominance, and dominance-by-dominance epistatic effects, respectively; r is the vector of residual genotypic effects, referring to an un-dissectable combination of dominance and epistatic effects in equation [2], epistatic effects in equation [3], epistatic effects excluding e_{xx} effects in equation [4]; rs is the vector of residual genotypic-by-site effects; ε is the vector of random residual effects. $\boldsymbol{X}$, $\boldsymbol{Z_1}$, $\boldsymbol{Z_2}$, $\boldsymbol{Z_3}$, $\boldsymbol{Z_4}$, $\boldsymbol{Z_5}$, $\boldsymbol{Z_6}$, and $\boldsymbol{Z_7}$ are the incidence matrices for β, a, as, d, ds, e_{xx}, r, and rs, respectively. The random additive effects (a) in equation [1–4] were assumed to follow $a \sim N(0, \boldsymbol{A}\sigma_a^2)$ with σ_a^2 the additive variance, where $\boldsymbol{A}$ is the pedigree-based additive relationship matrix in PBLUP (or replaced by a genomic-based additive relationship matrix $\boldsymbol{G_a}$ in GBLUP). The random dominance effects (d) in equations [1–4] were assumed to follow $d \sim N(0, \boldsymbol{D}\sigma_d^2)$ with σ_d^2 the dominance variance, where $\boldsymbol{D}$ is the pedigree-based dominance relationship matrix in PBLUP (or replaced by a genomic-based dominance relationship matrix $\boldsymbol{G_d}$ in GBLUP). The as and ds are the random additive-by-site and dominance-by-site interaction effects following $as \sim N(0,\ \sigma_{as}^2 \boldsymbol{I_s} \otimes \boldsymbol{A})$, and $ds \sim N(0, \sigma_{ds}^2 \boldsymbol{I_s} \otimes \boldsymbol{D})$ in the PBLUP model (or replaced by $\boldsymbol{G_a}$ and $\boldsymbol{G_d}$ in the GBLUP model). $\sigma_{as}^2\ and\ \sigma_{ds}^2$ are the additive-by-site and dominance-by-site variances, respectively. $e_{aa}, e_{ad}, and\ e_{dd}$ are the vectors of the random additive-by-additive, additive-by-dominance and dominance-by-dominance epistatic effects following $e_{aa} \sim N(0, \boldsymbol{P_{aa}}\sigma_{aa}^2)$, $e_{ad} \sim N(0, \boldsymbol{P_{ad}}\sigma_{ad}^2)$, and $e_{dd} \sim N(0, \boldsymbol{P_{dd}}\sigma_{dd}^2)$, respectively. $\boldsymbol{P_{aa}}$, $\boldsymbol{P_{ad}}$, and $\boldsymbol{P_{dd}}$ are the pedigree-

based additive-by-additive, additive-by-dominance, and dominance-by-dominance epistatic relationship matrices in the PBLUP model ($\boldsymbol{G_{aa}}, \boldsymbol{G_{ad}},$ and $\boldsymbol{G_{dd}}$ in GBLUP model), respectively and σ^2_{aa}, σ^2_{ad}, and σ^2_{dd} are their variance components. The *r* and *rs* are the vectors of residual genotypic effects and residual genotypic-by-site effects following $r \sim N(0,\ \boldsymbol{I_{nc}}\sigma^2_r)$ and $rs \sim N(0,\ \boldsymbol{I_{ncs}}\sigma^2_{rs})$, respectively, where *nc* is the number of clones, *ncs* is the number of clones multiplied by the number of sites. The vector of residual *e* was assumed to follow

$$\varepsilon \sim \mathrm{N}(0, \begin{bmatrix} \boldsymbol{I}_{\mathrm{n1}}\,\sigma^2_{\mathrm{e1}} & 0 & 0 & 0 \\ 0 & \boldsymbol{I}_{\mathrm{n2}}\,\sigma^2_{\mathrm{e2}} & 0 & 0 \\ 0 & 0 & \boldsymbol{I}_{\mathrm{n3}}\,\sigma^2_{\mathrm{e3}} & 0 \\ 0 & 0 & 0 & \boldsymbol{I}_{\mathrm{n4}}\,\sigma^2_{\mathrm{e4}} \end{bmatrix}),$$

where σ^2_{e1}, σ^2_{e2}, σ^2_{e3}, and σ^2_{e4} are the residual variances for sites 1, 2, 3, and 4, respectively; $\boldsymbol{I_{\mathrm{n1}}}$, $\boldsymbol{I_{\mathrm{n2}}}$, $\boldsymbol{I_{\mathrm{n3}}}$, and $\boldsymbol{I_{\mathrm{n4}}}$ are their identity matrices, and n1, n2, n3, and n4 are the number of individuals at each of the four sites, respectively.

For the other three traits (HT12, PILO, and FD), four similar models as equations [1–4] were used, except that the interaction terms with site were excluded because those traits were only measured at a single site (Table 1).

Pedigree-based BLUP (PBLUP) models based on equations [1] to [4] are called as PBLUP-A, PBLUP-AR, PBLUP-ADR, and PBLUP-ADR-xx, respectively. PBLUP-ADR-xx included three models of different epistatic effects called as PBLUP-ADR-aa, PBLUP-ADR-ad, and PBLUP-ADR-dd. Genomic-based BLUP models based on questions [1-4] could be called GBLUP-A, GBLUP-AR, GBLUP-ADR, and GBLUP-ADR-xx, respectively.

Pedigree-based and genomic-based relationship matrix estimates

The pedigree-based additive ($\boldsymbol{A}$) and dominance ($\boldsymbol{D}$) relationship matrices were constructed based on information from pedigrees. The diagonal elements (*i*) of the $\boldsymbol{A}$ were calculated as $A_{ii} = 1 + f_i = 1 + A_{gh}/2$, where g and h are the parents of the ith individual, while the off-diagonal element is the relationship between individuals ith and jth calculated as $A_{ij} = A_{ji} = (A_{jg} + A_{jh})/2$ (Mrode and Thompson 2005). In the $\boldsymbol{D}$ matrix, the diagonal elements were all one ($D_{ii} = 1$), while the off-diagonal elements between the individual *i*th and *j*th can be calculated as $D_{ij} = (A_{gk}A_{hl} + A_{gl}A_{hk})/4$, where g and h are the parents of the *i*th individual and k and l are the parents of the *j*th individual. $\boldsymbol{A}$ and $\boldsymbol{D}$ relationship matrices were produced using the AGHmatix package (Amadeu, 2021).

The genomic-based additive ($\boldsymbol{G_{\mathrm{a}}}$) and dominance ($\boldsymbol{G_{\mathrm{d}}}$) relationship matrices were constructed based on imputed SNP data as described by VanRaden (2008) (VanRaden, 2008) for $\boldsymbol{G_{\mathrm{a}}}$ and by Vitezica et al. (2013) (Vitezica *et al*, 2013) for $\boldsymbol{G_{\mathrm{d}}}$ using AGHmatrix package in R (Amadeu, 2021).

The relationship matrices due to the first-order epistatic interactions were computed using the Hadamard product (cell by cell multiplication, denoted #) and trace (*tr*) (Vitezica *et al*, 2018). In the pedigree-based model, the additive-by-additive terms are calculated as $\boldsymbol{P_{\mathrm{aa}}} = \frac{\boldsymbol{A\#A}}{tr(\boldsymbol{A\#A})/n}$, additive-by-dominance terms as $\boldsymbol{P_{\mathrm{ad}}} = \frac{\boldsymbol{A\#D}}{tr(\boldsymbol{A\#D})/n}$, and dominance-by-dominance terms as $\boldsymbol{P_{\mathrm{dd}}} = \frac{\boldsymbol{D\#D}}{tr(\boldsymbol{D\#D})/n}$. The *n* is the number of genotyped individuals. In genomic-based relationship

matrix models, additive-by-additive terms are calculated as $\boldsymbol{G_{aa}} = \frac{\boldsymbol{G_a \# G_a}}{tr(\boldsymbol{G_a \# G_a})/n}$, additive-by-dominance terms as $\boldsymbol{G_{ad}} = \frac{\boldsymbol{G_a \# G_d}}{tr(\boldsymbol{G_a \# G_d})/n}$, and dominance-by-dominance terms as $\boldsymbol{G_{dd}} = \frac{\boldsymbol{G_d \# G_d}}{tr(\boldsymbol{G_d \# G_d})/n}$.

Proportion of variance component to phenotypic variance

The narrow-sense heritability can be estimated as $h_i^2 = \sigma_a^2/\sigma_p^2$, using the models constructed above, σ_p^2 is the total phenotypic variance (e.g. in equation 3 and within a multi-site model, $\sigma_p^2 = \sigma_a^2 + \sigma_{as}^2 + \sigma_d^2 + \sigma_{ds}^2 + \sigma_r^2 + \sigma_{rs}^2 + \sigma_e^2$). Similarly, the dominance variance to the total phenotypic variance ratio was calculated as $d^2 = \sigma_d^2/\sigma_p^2$, the first-order epistatic variance to the total phenotypic variance ratio was calculated as $i^2 = \sigma_{xx}^2/\sigma_p^2$, the residual genotypic variance to the total phenotypic variance ratio was calculated as $r^2 = \sigma_r^2/\sigma_p^2$, and the broad-sense heritability was estimated as $H_i^2 = \sigma_g^2/\sigma_p^2$, where $\sigma_g^2 = \sigma_a^2 + \sigma_r^2$ is based on equation [2], $\sigma_g^2 = \sigma_a^2 + \sigma_d^2 + \sigma_r^2$ is based on equation [3], and $\sigma_g^2 = \sigma_a^2 + \sigma_d^2 + \sigma_{xx}^2 + \sigma_r^2$ is based on equation [4]. Clone mean narrow-sense heritability was estimated as $h_c^2 = \sigma_a^2/\sigma_{\bar{p}}^2$, where $\sigma_{\bar{p}}^2 = \sigma_a^2 + \frac{\sigma_{as}^2}{s} + \sigma_d^2 + \frac{\sigma_{ds}^2}{s} + \sigma_r^2 + \frac{\sigma_{rs}^2}{s} + \frac{\sigma_{\bar{e}}^2}{n}$ based on equation [3], *n* is the harmonic mean of the total number of ramets per clone and *s* is the harmonic mean number of sites in which each clone was represented by one or more ramets.

Association mapping

To check the additive genetic architecture of the six traits in the breeding population, we also performed GWAS based on clone mean values across site for all traits using the multi-locus BLINK model (Huang *et al*, 2019) conducted in GAPIT V3.0 R Software package (Wang and Zhang, 2021). Principal components were used to control population structure if the genomic inflation factor (i.e. lambda) is less than 0.95 or more than 1.05 (Yang *et al*, 2011). For the GWAS model, clonal means calculated across sites were used as adjusted phenotypic values. The genome-wide significance of associations was determined at an experiment-wise false discovery rate of <0.05 according to Benjamini & Hochberg (1995). The percentage of phenotypic variance explained (PVE) for each significant association was obtained from results using a mixed linear model (MLM) conducted in GAPIT V3.0 R Software package (Wang and Zhang, 2021).

Linkage disequilibrium

Genome-wide analysis of linkage disequilibrium (LD) was conducted in the F1 full-sib progeny population. All SNPs were mapped into Norway spruce genome v2.0 (In preparation). Out of the 47,445 SNPs available from the *Piab50K* SNP array, 43,267 SNPs were successfully mapped and evenly distributed across the 12 chromosomes (Figure 1). The remainder 4178 could not be correctly mapped to any particular chromosome, therefore they were instead grouped together in an assumed "chromosome 13" (Table S2). LD values for Pair-wise SNPs within each chromosome were calculated using VCFtools (Danecek *et al*, 2011).

Cross-validation test

Due to the negligible effects of dominance, dominance-by-site, and also first-order epistatic effects for all traits (Table S3), clone means calculated across sites were used as phenotype values to perform cross-validations. Ten sets of 10-fold cross-validations were performed. When a major-effect SNP was included as a fixed effect, the GWAS-analyses underlying the SNP-selection was performed on the training-portion of the cross-validation performance as well. In summary, we employed a model as below:

$$y' = \boldsymbol{X}\beta + \boldsymbol{Z_1}a + \varepsilon \qquad [5]$$

Where y' is the vector (904, 1) of the clonal means across sites, β is the vector of fixed effects, including a grand mean and a single-locus effect for the SNP showing the most significant association (smallest *p*-value) to the trait (when included). Furthermore, a is the vector of additive effects, ε is the vector of random residual effects. $\boldsymbol{X}$ and $\boldsymbol{Z_1}$ are the matrices related to the β and a. The random additive effects (a) were assumed to follow $a \sim N(0, \boldsymbol{A}\sigma_a^2)$ with σ_a^2 the additive variance in PBLUP-A whereas the matrix A will be replaced by G_a in the GBLUP-A model.

The prediction ability (PA) was defined as the Pearson correlation between predicted breeding values (EBVs) and clonal means. Furthermore, we evaluated the PA of within-family predictions. For this purpose, Pearson correlations were calculated between the EBV deviations from the EBV family means and the clonal mean deviation from the family means used as benchmark validation values. An estimate of the selection accuracy was calculated by dividing PA with the square root of the narrow-sense clonal heritability (PA/h_c).

Testing the efficiency of genomic prediction

Marker density and maker preselection

To test the impact of the number of SNPs on the PA of GBLUP, we performed GPs using 14 subsets of SNPs (25, 50, 100, 200, 500, 1K, 2K, 4K, 8K, 10K, 20K, 30K, 40K, and all SNPs) and using two different types of sampling strategies: 1) randomly selected SNP subsets and 2) SNP subsets selected based on the smallest *p*-values shown in the GWAS for additive effects using BLINK method in the training population. We performed these steps in both the whole population with 10-fold cross-validation replicated 10 times (n=100) and in 482 clones from the 10 largest families. In the subset population (n=482), we also focused on PA for within-family variation using marker preselection. The subset population was evaluated by inspecting their Aikake Information Criterion values (AIC) and the subset sample of markers showing the smallest AIC-values were selected for further investigations (henceforth called GBLUP-S).

Different statistical models

In most conifer tree species, including Norway spruce (Chen *et al*, 2021), growth traits, such as tree height and DBH, are commonly assumed as polygenic traits, but phenology traits, such as budburst stage may be considered as an oligogenic trait. The recorded pedigree usually has certain pedigree errors due to errors made in the field or due to pollen contamination at pollination. Thus, we tested the efficacy for four different statistical models:

1) PBLUP-C: the traditional pedigree-based BLUP but where marker-based pedigree correction was performed.

2) GBLUP: a BLUP model with a genomic-based relationship matrix (G) estimated from all markers. In this model, the prediction could simultaneously capture family and population structures, and also Mendelian segregation effects.
3) GBLUP-S: a genomic-based BLUP with a G matrix estimated from a subset of preselected markers where the number of markers depended on the genetic architecture of traits (i.e. 100 SNPs for BB, and 2000 SNPs for all other traits) based on GWAS results using the BLINK approach (Huang *et al*, 2019).
4) GBLUP-F: a genomic-based BLUP with a G matrix estimated from all markers, except the marker with the greatest significance included as a fixed regression effect and the marker was selected from GWAS result.

In addition to the four models mentioned above, we also performed several extra strategies.

Relationship between training and validation sets

Compared with the previously random sampling strategy with 10-fold cross-validation, another strategy entailed the prediction of genomic breeding values of all progenies within a specific full-sib family based on model training using all the other families using GBLUP. In all, there were 32 cross-validation repeats and the relationship between training and validation sets were thus consistently weak (unrelated or half-sib). This scenario represents a common breeding situation in which a new full-sib family is produced. Thus, predictive ability in the within-family selection depends on whether the model could capture the Mendelian segregation effects.

Family size (i.e. number of clones per family)

To test if an increase in family size could improve the PA of PBLUP-C and GBLUP, we randomly selected five to 30 clones per family as a training set for the largest ten families with 48-56 clones per family (Fig. S1), using the remaining clones in the ten families as a validation set. After pedigree correction, the number of clones for each of the ten families varied from 32 to 56 (Table S1). Thus, the cross-validation was performed based on the corrected pedigree.

We also performed GP of within-family phenotypic values by using the clonal means as y' in equation [5] and subsequently cross-validating this in the ten large families (n=482) by using within-family clonal means (clone mean-family mean) as benchmark. As previously, GBLUP-S was applied with 100 preselected SNPs for BB and with 2000 preselected SNPs for the rest of the traits.

Simulations of large-effect SNPs and their inclusion in the genomic prediction model

To verify whether the inclusion of a major-effect locus as a fixed effect in the model would improve prediction ability in genomic selection, we conducted finite-locus model simulations of a simplified breeding population undergoing one generation of directional selection. We would thereby demonstrate the value of identifying major-effect loci e.g. by GWAS in a context of genomic selection and breeding. We used the simulation software Metagene (Sanchez *et al*, 2008) to simulate one generation of genomic selection and breeding.

Simulated genomic architecture

A simplified genomic architecture setup comprising 15,000 biallelic loci, uniformly distributed along 12 chromosomes, each with a genetic length of 250 cM, and thus similar to the architecture of Norway spruce (Bernhardsson *et al*, 2019), was used. Metagene regulated the genetic length of the genome by allowing completely random recombination between chromosomes at meiosis and thereafter randomly assigning 30 additional crossovers uniformly throughout the genome for each simulated meiosis event. Thus, a total genome length of 3,000 cM was achieved (0.2 cM between each neighboring pair of intrachromosomal loci). In order to emulate the highly skewed allele frequency distributions observed in many populations, allele frequencies for all loci (p) were drawn from a U-shaped probability distribution (Hill *et al*, 2008) while setting the effective population size parameter (N_e) equal to the number of simulated founders (100). For simplification, the founder haplotypes were sampled according to these specifications without the presence of any "historical" linkage disequilibrium (LD) thus implying that alleles would roughly adhere to Hardy-Weinberg equilibrium (HW-eq) and that the founder population would be devoid of any systematic substructures. Out of the 15,000 available loci, 500 loci were randomly designed to have minor additive effects (α_i for effective loci $I = 1\ldots500$) on a virtual quantitative trait of interest in accordance to the results of previous meta-analyses (Hall *et al*, 2016). To emulate the usually observed distribution of a few loci with slightly larger effects and many loci with very small effects, the sizes of the 500 additive allelic effects were randomly drawn from a negative exponential distribution (Otto and Jones, 2000) with the rate parameter (λ) set at 1. To regulate the heritability of the virtual trait, the total additive genetic variance of the founder population was first estimated (Lynch and Walsh, 1998) as:

$$\sigma_a^2 = 2\sum_{i=1}^{500} \alpha_i^2 p_i(1-p_i) \qquad [6]$$

Since the founder population adheres to HW-eq and the founder population exhibited no population structure (no systematic LD), this simplified equation should be adequate for estimating σ_a^2. Thereafter the virtual phenotypic trait was constructed by adding random environmental noise effects from a normal distribution with a mean of 0 and a standard deviation of $\sqrt{3}\sigma_a$ to the additive genetic effects ($2\sum_{i=1}^{500} \alpha_i p_{ij}$) of each individual j. Thus, a virtual quantitative trait showing a narrow-sense heritability (h^2) of ~0.25 at the founder level was achieved. It should be noted that the additive genetic variance in subsequent simulated generations was allowed to be influenced by any simulated events that could occur (selection effects, deviations from HW-eq and accumulation of LD) whereas the environmental noise variance (σ_e^2) was fixated at the level determined at the founder level (G_0).

Simulation of a single-population breeding program

The 100 founders were crossed 50 times according to a randomly allocated single-pair mating design (SPM) yielding 40 offspring per cross and in total 2000 individuals distributed within 50 full-sib families in a segregation population. For the study of conventional genomic selection, 1500 loci, randomly selected among the total 15,000 loci, were "genotyped" and were subsequently subjected to the same type of GBLUP-analysis and cross-validation procedure as described previously for the real-life data (see equation [5]). However, for the simulated data, phenotypic data was used directly for training and as a validation benchmark because the simulations did not include clonal replicates.

The evaluation of genomic prediction PA was done as previously described for the real-life data whereas prediction accuracy was calculated as the Pearson-correlation between predicted breeding values and true breeding values as the

latter is reported by Metagene. Within-family prediction accuracy was also estimated as the Pearson-correlation between predicted breeding value deviations from the family means and the true EBV – true family EBV mean as the validation benchmark. For conventional genomic prediction and selection, 10 repeated simulations were performed completely resampling genomic architecture and founder population makeup according to previously reported specifications and for each such simulation, one set of 10-fold cross-validations was performed.

Simulations of major-effect loci

Apart from conventional genomic prediction and selection, simulation scenarios were designed in which the presence of a "major-effect" locus was added. To produce such a locus, the previously designed set of genomic architectures was modified so that a major locus was chosen within the existing genomic architectures. The major effect locus for each architecture was arbitrarily selected within chromosome 1 but the choice of the locus was nonetheless done according to a number of requirements. The major effect locus was required to: 1) not exhibit any causal effect in the original genomic architecture; 2) exhibit a close-to-intermediate allele frequency ($0.45 < p_{maj} < 0.55$); 3) be situated at least 10 positions (2 cM) away from any other effective locus in order to avoid the development of overly tight linkage with these; and 4) not be previously assigned as a "genotyped locus" in the original setup. When added, the number of effective loci for the modified scenario thus became 501 rather than 500 for the conventional scenario. The effect size of the major locus (α_{maj}) was regulated in terms of the percentage of the genetic variance that it could explain (PGVE) thus being calculated as:

$$\alpha_{maj} = \sqrt{\frac{PGVE}{100} \cdot \frac{\sigma_a^2}{2 \cdot p_{maj}(1-p_{maj})}} \qquad [7]$$

Then, the effect size of all other effective loci was adjusted (rescaled) as:

$$\alpha_{i,adj} = \alpha_i \sqrt{1 - \frac{PGVE}{100}}$$

so that overall σ_a^2 remained constant. Environmental variance and heritability were regulated as previously stated. Different PGVE-values for the major locus were tested in different scenarios (0%, 1%, 5%, 10% and 20%). Because the previously used heritability of 0.25 was retained this translates to 0%, 0.25%, 1.25%, 2.5%, 5% in terms of PVE. It should be noted that PVE = 0% (no major locus effect) would be equivalent to the conventional genomic architecture and a PVE at 0.25% would only imply that the locus would mainly fall within the conventional range of the negative exponential sampling distribution (i.e. a minor effect locus). However, a major locus with a PVE at 1.25% would equal the very strongest QTL that sampling of 500 effects from a negative exponential distribution could produce and major loci with PVE-values at 2.5% and 5% would be considerably stronger than what could be thus sampled with any reasonable likelihood (i.e. a true major QTL).

Using all these genomic architecture setups, with the PVE of the major locus varying from 0% (fake) to 5% (true major QTL), conventional genomic prediction procedure was first simulated. As the major locus was not assigned as being genotyped in the conventional cross-validation procedure, these analyses represent the situation where a major locus (of various strength) is present in the genome but is missed in genotyping and thus not directly utilized by the

GP models. But in addition to this, a second set of cross-validation analyses were made where the major QTL was indeed genotyped and directly used as a fixed regression term in model training and application. This would represent the situation where the QTL was indeed observed and subsequently utilized in GP model training. It should be noted that the context of this strategy varies according to the relative strength (PVE) of the major locus, ranging from the attempted use of a locus that is not truly causal or major (PVE 0-0.25%) to the utilization of a true major effect locus (PGVE 2.5-5%). Consequently, no GWAS analysis was performed to determine which locus would be used for model training and cross-validation in contrast to the real-data analysis.

Results

Genetic parameter comparisons between models

To compare the goodness of fit of four models described in equations [1-4], the results of the different fitted models were shown in Table S3. For all traits, the GBLUP-AR model had the smallest Akaike information criterion (AIC) value, except for frost damage (FD) with a zero non-additive variance (i.e. GBLUP-A). This indicates that the fitting of GBLUP-AR was generally better than all other models, both pedigree-based and genomic-based, and implies that genetic parameters of the GBPLUP-AR models should be better estimated than for the other models. When comparisons were made only among PBLUP models, patterns were less consistent where PBLUP-AR or even PBLUP-ADR were indicated (by low AIC-values) to be the best models whereas, for PILO and FD, PBLUP-A appeared to be the model with the best fit. Based on the AIC, we did not see that models with first-order epistatic effect terms (PBLUP-ADR-xx and GBLUP-ADR-xx) showed any better fit than the best pedigree- or genome-based models in that respect (usually PBLUP-AR and GBLUP-AR, respectively).

Estimates of variance component and heritability

We found that the additive genetic variance under PBLUP-AR and GBLUP-AR decreased compared to that under PBLUP-A and GBLUP-A for HT6, HT12, DBH and BB, respectively (Table 2). Given that the PBLUP-AR and GBLUP-AR exhibited the best fit for all these traits, the previously observed patterns indicate that the additive genetic variance in the PBLUP-A and GBLUP-A may be inflated due to the inadvertent inclusion of non-additive effects within σ_a^2. For the growth traits (HT6, HT12 and DBH) the estimates of the non-specific non-additive genetic variance (σ_r^2) were substantial in size relative to the additive genetic variance estimates ($\sigma_r^2 > 0.2\sigma_a^2$) under the best-fit model according to AIC (GBLUP-AR). For BB and PILO, the additive genetic variance under GBLUP-AR decreased only marginally compared to σ_a^2-estimates under GBLUP-A and the non-additive variance estimates were relatively small in comparison to the additive ($\sigma_r^2 < 0.2\sigma_a^2$). Finally, FD did not exhibit any non-additive effects regardless of whether pedigree- or genome-based matrix models were used. Taken together, we may conclude that non-additive effects are important for all growth traits, whereas non-additive effects had a limited impact on non-growth trait BB and PILO and to be absent in FD. Due to the non-significant and small amount of dominance variance generally estimated, we found that the additive variance under GBLUP-ADR was comparable to that estimated under GBLUP-AR (Table S3). Also, most of the first-order epistatic interactions under PBLUP-ADR-xx and GBLUP-ADR-xx were in boundary close to zero based on Equation [4] (Table S3). Thus, we do not discuss the result of the PBLUP-ADR, GBLUP-ADR, PBLUP-ADR-xx and GBLUP-ADR-xx models any further. Broad-sense heritabilities under GBLUP-AR, GBLUP-ADR, PBLUP-AR, and PBLUP-ADR were very similar for all traits.

Summary of Norway spruce 50K SNPs array, LD decay, and association mapping

For the *Piab50K* SNP array, 41,236 of the total 47,445 SNPs were mapped onto each of the 12 chromosomes in the Norway spruce genome v2 (In preparation). The number of SNPs in each chromosome varied from 3158 to 3991 SNPs (Table S2. The physical extent of LD ($r^2 > 0.2$) within each chromosome varied from 33.7 kb in chromosome 2 to 54.6 kb in chromosome 9, with an average of 42.9 kb for the whole genome based on the SNP array in the studied

full-sib family population (Fig. S2). The family clusters was clearly separated by the first two principal components (Fig. S3). In total, GWAS identified 41, 11, 2, 4, 4, 11, and 13 SNPs as having a significant effect on BB, DBH, FD, HT12, HT6, and PILO, respectively, under a false discovery rate (FDR) of 0.05 (Benjamini and Hochberg 1995). The largest effect sizes for trait-associated SNPs explained 5.1, 1.6, 1.7, 2.0, 1.2, and 1.4% of the phenotypic variation of clonal means in BB, DBH, FD, HT12, HT6, and PILO, respectively (Table S4). There were six highly significant SNP-trait associations with considerable PVE (>2.5%) for budburst, indicating oligogenic nature whereas such observations were not made for the other traits.

Predictive ability comparisons between PBLUP and GBLUP

To confirm the previous observations that GBLUP models in general offered a better fit to the data than PBLUP, and to evaluate the efficacy of genomic prediction in general, we initially compared predictive abilities (PAs) estimated from cross-validation under PBLUP and GBLUP models with the square root of the clone mean narrow-sense heritability (h_c, Table 3 and Table S5). Estimates of h_c are interpreted as the theoretical PA and accuracy when applying mass selection based on phenotypic clonal means. In agreement with previous comparisons between PBLUP and GBLUP in terms of AIC, it was observed that PA-estimates of GBLUP (0.23 – 0.67) were consistently higher than corresponding PA-estimates for PBLUP (0.20 – 0.61). The increases in PA under GBLUP were the greatest for HT6 and BB. For example, PA of GBLUP for HT6 increased by 15.8% when compared to PBLUP. However, the PA-estimates for GBLUP in a 10-fold cross-validation scheme (phenotypic data absent) were in turn consistently and considerably lower in comparison to the corresponding theoretical PAs in the presence of phenotypic data (h_c in the range 0.37 – 0.90).

The impact of the number of SNPs included in the G-matrix

For all traits, the PA of the GBLUP model increased from 25 randomly selected SNPs to ca. 4000 SNPs (Fig. 3a, Table 3) and PA plateaued when even more SNPs were selected all to the full set of SNPs in the array (46,482). Similarly, the standardized AIC values of the genomic prediction model decreased from the model with 25 SNPs to ca. 4000 SNPs and then the AIC stabilized after that (Fig. 3b). For example, for BB, PA under the GBLUP-model increased from 0.36 using 25 SNPs to 0.63 using 4000 SNPs in the estimation of a genomic-based relationship matrix (G).

When marker preselection from the GWAS-analyses were performed, the resulting prediction models showed two types of trends for six traits. First, for BB, the genomic prediction model (GBLUP-S) with 100-200 preselected SNPs with the smallest *p*-values obtained a higher PA than that using all other numbers of SNPs (Fig. 3a). To check if GP with a marker preselection by GWAS captured more Mendelian segregation effects, we also calculated the correlation between estimated breeding values (EBVs) and within-family phenotypic variation (clone mean-family mean). We found that GBLUP-S using 100-200 informative markers captured more of the phenotypic variation and exhibited the lowest AIC value (Fig. 3c). This suggests that using all available SNPs for a trait indicated to be oligogenic may introduce misleading noise in the prediction models.

Second, for the other five traits (DBH, FD, HT12, HT6 and PILO), we found that GP using a preselected set of SNPs of the smallest *p*-value was better than GP with random marker selection if the number of pre-selected SNPs ranged from 25 to a few hundred. However, when more SNPs were preselected, GP did not outperform random marker selection. Furthermore, preselecting more markers beyond a few hundred still increased the PA of GP for these five polygenic traits with PA being the highest when all 46,482 markers were used (Fig. 3a). When evaluating GP for within-family variation, we found that the model with 2000-4000 preselected markers for traits HT6, HT12, and PILO, actually captured more, albeit still very low amount of the trait within-family variation (0.05-0.06, 0.02-0.002, and 0.03-0.04 of pAs for HT6, HT12, and PILO, respectively), than including all markers. For all traits, within-family PA values were consistently higher when marker selection was based on low *p*-values in comparison to random selection. AIC values for all traits except BB showed a similar trend (Fig. 3c) where the lowest AIC value occurred at 2000-4000 *p*-value preselected markers.

Predictive ability comparisons between advanced GBLUP models

As already indicated in Fig. 3, only BB showed appreciable increases in overall PA from 0.61 for GBLUP to 0.72 for GBLUP-S when the 100 SNPs exhibiting the smallest *p*-values in GWAS were included in the G_a-matrix (Table 3). For all other traits the GBLUP-S model, which included the 2000 SNPs with the smallest *p*-values in GWAS, showed similar or even lower PA than the conventional GBLUP including all SNPs. In addition, we attempted the fitting of a single SNP showing the smallest *p*-value (the highest significance) in GWAS as an additional fixed regression effect in the model (GBLUP-F). For HT12, BB and FD, the GBLUP-F model produced a slightly higher PA than GBLUP (by 0.02 – 0.03) but for DBH, HT6 and PILO, GBLUP-F and GBLUP showed similar values. Furthermore, we calculated an approximation of cross-validation accuracy in the absence of phenotypic data as the PA of PBLUP and GBLUP divided by h_c. We found that the overall accuracy of GP for all traits ranged from 0.58 for HT6 to 0.78 for BB which was the highest observed accuracy.

Simulations of including a large effect SNP as fixed in the GBLUP

To verify whether including a major gene locus as a fixed effect in the GBLUP model (GBLUP-F) would increase PA of GP, we also performed 10 repeats for 10-fold cross-validation on simulated data where single SNPs exerting a range of effect sizes (in terms of PVE) were included (Table 4). Simulation data is easier to interpret than real-life data since it offers the possibility to robustly estimate GP accuracies merely by calculating the correlations between predicted and true breeding values. Results showed that the overall accuracies of the GBLUP-F model were higher than that of GBLUP when provided a major-effect SNP with PVE>=2.5% (accuracy ranges of 0.66-0.71 and 0.60-0.63 for GBLUP-F and GBLUP respectively). Following a similar trend, the overall PA of the GBLUP-F model increased from 0.31 to 0.35 for a fixed SNP explaining 0% to 5% of the phenotypic variation whereas no increasing trend could be observed for the GBLUP model which did not include the major QTN.

For the within-family accuracy (Table 4), the GBLUP-F model increased from a very low level (0.02) when the fitted fixed-effect SNP explained 0% of the variation (i.e. fitting a false-positive effect SNP) to 0.21 for a fixed effect SNP with a PVE at 5% (true major gene). In contrast, for the model where the assumed major SNP was not fitted as a fixed

effect, within-family accuracies remained at very low levels (0.02-0.03) regardless of the PVE of the major SNP. The trends of within-family PA were similar as those for accuracy but were all lower due to the environmental noise that always influences PA-estimates. The above results indicate that including a large-effect SNP (PVE>=2.5%) as a fixed effect in the GBLUP model may improve the accuracy and PA, especially with respect to the within-family accuracy.

Number of clones per family

To test the effect of the number of clones per family available to the training dataset on the PA using PBLUP and GBLUP (Fig. 4), we sampled 5, 10, 15, 20, 25, or 30 clones from each of the ten largest families as a training data set and the rest of clones in those families as a validation set. We found that the PA of both PBLUP and GBLUP consistently increased from 5 clones per family to 30 clones per family for all traits except for PBLUP for BB where an optimum was reached at 20 offspring clones per family. Based on the trends and with genomic prediction in mind (GBLUP), more than 30 clones would be better regardless of the trait under study. Following the increase of the number of clones per family, GBLUP showed a higher PA than PBLUP, except for DBH where PA-estimates were similar.

Number of clones per family and marker preselection by GWAS also affect the within-family variation

To test further whether combining GWAS *p*-value based marker preselection and the number of clones per family would affect the predictive ability (PA) of within-family variation, we performed 10-fold cross-validation using GBLUP based on clonal means as training phenotypic value (y') within the largest ten families (Fig. 5). We found that the PA within-families of the ten largest families using GBLUP with GWAS *p*-value based preselection of a number of markers produced similar trends (Fig. 5a) as the corresponding PA values estimated for the population as a whole (Fig. 3a). In similarity to the results of the whole population, GBLUP of BB with 100-200 preselected SNPs produced higher predictive ability and lower AIC than using a lesser or greater number of SNP for G_a-matrix calculation (Fig. 5a and 5b). For the other traits, 2000-4000 preselected SNPs produced the lowest AIC in agreement with the whole-population analysis. But the trends for within-family PA were less obvious although it was nonetheless clear that some sort of SNP preselection resulted in higher PA than uncritically using all SNPs in the model. One notable difference between this analysis of the ten largest families and that of the whole population was that the PA for BB peaked at a higher value (0.42) when only the ten largest families were included in comparison with the corresponding PA estimate for the whole population (0.36). These results indicate that more within-family variation linked to Mendelian segregation could be captured by preselecting influential markers for predicting within-family variation and that the capture of such segregation variation is likely easier in a population where families are fewer and larger and where the average relationship thus is higher.

Based on the marker preselection results for within-family prediction (Fig. 5c), we also found that PA increased quickly from 5 clones per family to 30 clones per family for all traits. For example, PA within-family variation for BB increased from 0.02 at family size 5 to 0.42 when 30 clones per family were available for model training.

In practical breeding, genomic prediction will usually be performed for a new full-sib family with half-sib or unrelated relationships with some of the families in a training set. Thus, we also evaluated one special type of cross-validation

scheme in which the clonal mean data for each of 32 families were in turn held out from the GBLUP training dataset, while the family for which clonal mean data was missing was used for the prediction validation (across-family validation). As previously, we performed the GBLUP with 1) all markers for G_a -matrix calculation (GBLUP), 2) 100 preselected SNPs for BB and 2000 preselected SNPs for the rest of the six traits based on the smallest GWAS *p*-values (GBLUP-S), and 3) all markers for the G_a -matrix plus the marker with the smallest GWAS *p*-value fitted as a fixed regression effect in the model (GBLUP-F). We found that marker preselection increased pAs for some traits (HT12, PILO, BB, and FD), but not for others (HT6 and DBH). The PA-estimates across families were very low for most traits (0.03 to 0.11) but for BB the GBLUP PA was 0.19 and increased to 0.25 when using GBLUP-S (Table 5). To fit the most powerful SNP as a fixed effect (GBLUP-F) was generally not helpful except for BB where the PA was increased to 0.24 for GBLUP-F.

Discussion

Genomic-based BLUP model could enhance the accuracy of the estimated additive and non-additive genetic variances

In tree or conifer breeding programs, a control-pollinated clonal test provides an opportunity to simultaneously estimate the additive, non-additive, dominance, and epistatic variances using pedigree. However, most heritability estimates in many traditional tree breeding programs were based on open-pollinated or control-pollinated progeny trials without vegetative propagation. In this context, such estimates of additive genetic variation may be biased due to the difficulty in separating the additive genetic variance from parts of the dominance and additive-by-additive epistatic variances (Mullin and Park, 1992). Using a model that includes genomic data, the additive and dominance variances may still show bias if models do not include a residual genetic effect if vegetatively propagated material used (Tan *et al*, 2018; Walker *et al*, 2021). Potentially, the genomic data could capture and discriminate between the different sources of the non-additive genetic variance such as dominant effects within a locus and epistatic interaction effects among loci if an appropriate model is used. Based on different genomic prediction models, GBLUP-AR showed the smallest AIC values for most of the traits, indicating that non-additive genetic effects are significant for all traits, with frost damage being the only exception. Also, the fact that GBLUP-AR model AIC values for these traits were systematically lower than the corresponding PBLUP-AR values, implies that genomic prediction models perform better with respect to separating additive and non-additive variances from each other than pedigree-based models. For example, PILO showed non-significant and negligible non-additive effects under PBLUP-AR but showed nonetheless significant non-additive effects under the GBLUP-AR model, indicating that the GBLUP-AR model with a realized relationship matrix could better capture and separate the additive genetic variation and thus improve estimates of genetic parameters.

Significant marker-trait associations and genomic prediction

Recently, several studies on tree species have reported that selecting markers that have a particular influence over the studied trait could improve the predictive ability (Thumma *et al*, 2022). Tan and Ingvarsson (2022) (Tan and Ingvarsson, 2022) reported that a careful 1% preselection of markers could improve the estimate of heritability and genomic prediction in a *Eucalyptus* population. In *Pinus contorta* Douglas ex Loudon var. *latifolia*, Cappa et al. (2022) reported that selecting informative markers, in particular markers capturing ancestry/population structure, can improve the predictive ability. In *Eucalyptus nitens*, Thumma et al. (2022) developed a marker panel based on selected candidates and genotyped with a "targeted genotyping-by-sequencing (TGD)" method and revealed dominance effects influencing growth.

The recently developed 50K SNP array for Norway spruce included several QTLs per trait detected in our previous GWAS (Baison *et al*, 2019; Baison *et al*, 2020; Chen *et al*, 2021; Elfstrand *et al*, 2020). In this study, GWAS identified 44 associated SNPs for the budburst stage and a particular SNP (MA_12842_2274, Table S4) with the second largest effect size explaining more than 4% phenotypic variation. Within 400 base pairs of this SNP, one of the QTLs related to the budburst stage was previously detected by GWAS in a different population of more than 4000 individuals (Chen

et al, 2021). We also observed two to 13 SNPs significantly associated to the other five traits (aside from BB), which makes the possibility to test if including the most significant SNP in GP models could improve the PA.

Marker preselection and inclusion of large effect QTL as fixed effects could enhance GP predictive ability

For budburst stage, we found that a G_a-matrix built from 100 preselected markers resulted in GBLUP-models having lower AIC values and being better at predicting the genetic value in the absence of phenotypic data than did a G_a-matrix model using all available markers. Such preselected SNPs also were observed to capture a considerable amount of within-family variation/mendelian segregation effects. However, for other more polygenic traits, genomic models using ca. 4000 preselected markers showed a comparable PA compared to the model using all markers, even though such models exhibited lower AIC values. This indicates that the preselection of influential markers is more likely to be successful when applied to a limited set of markers showing highly significant associations to an oligogenic trait where a relatively limited number of QTL are likely to exhibit major individual effects (PVE > 1%).

In this study, we also investigated a realized GP model for the budburst stage where a large-size QTLwith a PVE of ca. 5% was explicitly included as a fixed regression effect and we observed a 4.4% improvement in overall PA (Table 3) in comparison to a GP model without such a modification. Such enhancement was also observed in several empirical crop studies (Merrick *et al*, 2021; Sarinelli *et al*, 2019) and a simulation study conducted by Bernardo (2014) (Bernardo, 2014). Our finite-locus simulations also showed a similar result with 13% and 18% of improvement, in terms of PA and accuracy respectively, for the model including a locus of large effect size (PVE at 5%) as a fixed effect compared with the model without the fixed effect. The model improvement was particularly notable when the objective was the capture of within-family Mendelian segregation effects (Table 4). However, if the QTL effect size was less than ca. 1.25% of the phenotypic variation, the simulated data analyses did not indicate any advantage for the model including the locus as a fixed effect in terms of PA or accuracy. Based on more than dozens GWAS results in tree species (Hall, et al. 2016), SNPs detected with a PVE >1.25% is not uncommon, indicating that genomic prediction adding a QTL of large size should be a very useful approach to improve PA and accuracy. This may indicate that genomic prediction using GWAS results would be more efficient for oligogenic traits and maybe less effective for polygenic traits in which GWAS only detected a few SNP-trait associations with a small effect size.

Family size matters for the efficiency of genomic prediction

A larger family size (number of clones per family) is usually expected to capture more Mendelian segregation effects (Arenas *et al*, 2021) and improves the PA and accuracy of genetic parameter estimates (Perron *et al*, 2013). In this study, we observed that increasing the family size is important to improve the PA, both with respect to overall phenotypic prediction (Fig. 3) but also to the prediction of within-family variation (Fig. 5). This is especially important when the family size is small (less than 10 or 15 clones per family). For example, in Fig 5, within-family PA appeared to be even less than 0 for DBH when the number of clones per family was less than 15. We again observed that GP using marker preselection was generally better than that using all markers for capturing within-family variation based on Mendelian segregation.

Relationship between training and validation datasets highly important for genomic prediction

Forward-selection tree breeding usually entails the selection of a few unrelated progenies within a segregation test population (F1). A few new controlled crosses among the selected elite trees produce a new batch of progenies (F2). Thus, the GP using existing F1 as a training dataset to predict F2, the relationship between F1 to F2 will decrease compared with relationships within the same F1 generation as was used for 10-fold entirely random cross-validation in this study. We therefore also investigated the situation where training and validation datasets contained individuals from separate families and we could produce a 32-fold cross-validation scheme by removing phenotypic data for one particular family at the time. The relationships between the validation family and the training dataset were thus restricted to the level of half-sibs or even weaker. Thus, the PA for predicting a particular validation family was considerably lower than for conventional random cross-validation (Table 5), which was a result similar to that in *Pinus taeda* L. (Walker *et al*, 2021). However, it should nonetheless be noted that the seemingly oligogenic BB still offered a PA-estimate that was appreciable (0.19), and this estimate was further improved if marker preselection was performed for calculating the G_a-matrix of the model (GBLUP-S, 0.25) or by fitting the most significant marker as a fixed effect in the model (GBLUP-F, 0.24).

Marker density and LD affects predictive ability of phenotypic variation and within-family variation

In forest tree breeding, the capture of the expected pedigree relationships (e.g. 0.25 among half-sib siblings), only requires a few thousand SNPs. Literatures indicate that such a number of markers would also be enough to achieve an overall predictive ability similar to models utilizing all markers (Calleja-Rodriguez *et al*, 2020; Cappa *et al*, 2022; Chen *et al*, 2018; Tan *et al*, 2017). More markers within a limited whole-genome coverage may not be able to improve GP (Chen *et al*, 2018). In such situations, a few thousand SNPs preselected based on GWAS or other priorly known information could instead improve the PA, especially for traits where several large-effect QTNs have been observed (Cappa *et al*, 2022; Chen *et al*, 2021; Tan and Ingvarsson, 2022). In theory, coefficients of the pedigree relationship matrix describe additive genetic relationships between individuals at quantitative traits loci (QTL) (Habier *et al*, 2013), but in reality, it is not obvious to what extent the genomic relationship matrix explains a genetic covariance matrix between individuals for QTLs, especially for a targeted trait.

Marker density (i.e. the number of markers) is usually considered as one of the most important factors affecting GP performance (Grattapaglia and Resende, 2011). To capture LD between markers and QTL, genomic datasets consisting of more densely located markers usually capture more variation and improve GP (Isik, 2014). In many animals and crop GP, using the whole-genome resequencing markers or SNP array with whole genome-level imputation were believed to improve the PA of GP, especially when low minor allele frequency loci in the low linkage disequilibrium with neighboring variants were included in the prediction model (Wainschtein *et al*, 2022). In tree species, GP is usually performed with 20k-50k SNP arrays (Cappa *et al*, 2022; Shalizi *et al*, 2021) or 100k to 200k exome capture markers or genotyping-by-sequencing (GBS) markers (Calleja-Rodriguez *et al*, 2020; Chen *et al*, 2018).

On the other hand, LD matters when genomic prediction is conducted using a constant number of markers and for the estimation of marker-based narrow-sense heritability. A few hundred thousand random sampled markers usually only capture a limited proportion of the heritability when estimated using a population with unrelated individuals (the *missing heritability* phenomenon), especially for polygenic traits in conifer species (Chen *et al*, 2021). The extent of

LD (based on the threshold $r^2 \geq 0.2$) was observed to be 42.9kb when based on analyses using the 50k SNP array in this studied population (Fig. 1). However, based on such an extent of LDs between markers and QTLs, the 50k SNP array only covered ca. 25% of the total genomic size of Norway spruce (20G) (Nystedt *et al*, 2013). This could be a reason why the PA of GP did not reach the standard value of the square root of additive clone mean heritability (Table 3) and why within-family and across-family PA was very low for most of the studied traits (Table 5 and Fig. 5). In order to capture more Mendelian segregation between QTL and makers, we would then need more markers for marker preselection for oligogenic traits and to increase the number of informative markers for polygenic traits. For example, for the budburst stage, the model with 100 pre-selected markers based on GWAS may have captured LD between markers and QTLs and indeed it captured a considerable amount of within-family variation (Mendelian segregation, Fig. 5 and Table 5). This is agreed with a few studies on other tree species (Cappa *et al*, 2022; Tan and Ingvarsson, 2022; Thistlethwaite *et al*, 2017). Cappa et al. (2022) showed that trait-specific SNP selection including only those SNPs that maximize the ancestry informativeness coefficient (AIM) could improve the PA, in which those markers efficiently capture population structure. On the other hand, the results of this study also indicated that further increases in PA and accuracy could be possible by further increasing the number of clones for GP model training, in particular the number of clones per family. There was a clear positive relationship between the number of clones per family and overall and within-family PA and this relationship did not appear to plateau or be broken, with population PBLUP PA for BB being the only possible exception, within the ranges of family size available for this study.

Acknowledgements

We greatly appreciated that Dr. Wu Yanfang help us with field samplings. The research was mainly funded by a KAW project (2018.0272). Moreover, Z.-Q. Chen was partly supported by the European Union Horizon 2020 research and innovation programme under grant No 773383 (B4EST project) while H.R. Hallingbäck was partly supported by FORMAS general grant No 2019-00874. The computations were performed on resources provided by the Swedish National Infrastructure for Computing (SNIC) at UPPMAX.

References:

Amadeu R (2021). AGHmatrix: An R package to compute relationship matrices for diploid and autopolyploid species.

Arenas S, Cortés AJ, Mastretta-Yanes A, Jaramillo-Correa JP (2021). Evaluating the accuracy of genomic prediction for the management and conservation of relictual natural tree populations. *Tree Genet Genomes* **17**(1)**:** 12.

Baison J, Vidalis A, Zhou L, Chen Z-Q, Li Z, Sillanpaeae MJ *et al* (2019). Genome-wide association study identified novel candidate loci affecting wood formation in Norway spruce. *Plant J* **100**(1)**:** 83–100.

Baison J, Zhou L, Forsberg N, Mörling T, Grahn T, Olsson L *et al* (2020). Genetic control of tracheid properties in Norway spruce wood. *Sci Rep* **10**(1)**:** 18089.

Berlin M, Jansson G, Högberg K-A, Helmersson A (2019). Analysis of non-additive genetic effects in Norway spruce. *Tree Genet Genomes* **15**(3)**:** 42.

Bernardo R (2014). Genomewide selection when major genes are known. *Crop Sci* **54**(1)**:** 68-75.

Bernhardsson C, Vidalis A, Wang X, Scofield DG, Schiffthaler B, Baison J *et al* (2019). An ultra-dense haploid genetic map for evaluating the highly fragmented genome assembly of Norway spruce (Picea abies). *G3: Genes, Genom, Genet* **9**(5)**:** 1623-1632.

Bernhardsson C, Zan Y, Chen Z, Ingvarsson PK, Wu HX (2021). Development of a highly efficient 50K SNP genotyping array for the large and complex genome of Norway spruce (*Picea abies* L. Karst) by whole genome re-sequencing and its transferability to other spruce species. *Mol Ecol Resour* **21**(3)**:** 880–896.

Bian Y, Holland J (2017). Enhancing genomic prediction with genome-wide association studies in multiparental maize populations. *Heredity* **118**(6)**:** 585-593.

Bradshaw RH, Ingvarsson PK, Rosvall O (2019). The ecological consequences of using clones in forestry. *Scand J For Res***:** 1-10.

Browning SR, Browning BL (2007). Rapid and Accurate Haplotype Phasing and Missing-Data Inference for Whole-Genome Association Studies By Use of Localized Haplotype Clustering. *Am J Hum Genet* **81**(5)**:** 1084-1097.

Calleja-Rodriguez A, Pan J, Funda T, Chen Z, Baison J, Isik F *et al* (2020). Evaluation of the efficiency of genomic versus pedigree predictions for growth and wood quality traits in Scots pine. *BMC Genom* **21**(1)**:** 796.

Cappa EP, Ratcliffe B, Chen C, Thomas BR, Liu Y, Klutsch J *et al* (2022). Improving lodgepole pine genomic evaluation using spatial correlation structure and SNP selection with single-step GBLUP. *Heredity*.

Chen Z-Q, Baison J, Pan J, Karlsson B, Andersson B, Westin J *et al* (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 Genom* **19**(1)**:** 946.

Chen Z-Q, Baison J, Pan J, Westin J, Garcia-Gil MR, 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.

Chen Z-Q, Hai HNT, Helmersson A, Liziniewicz M, Hallingbäck HR, Fries A *et al* (2020). Advantage of clonal deployment in Norway spruce (*Picea abies* (L.) H. Karst). *Ann For Sci* **77**(1)**:** 14.

Chen Z-Q, Zan Y, Milesi P, Zhou L, Chen J, Li L *et al* (2021). Leveraging breeding programs and genomic data in Norway spruce (*Picea abies* L. Karst) for GWAS analysis. *Genome Biol* **22**(1)**:** 179.

Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA *et al* (2011). The variant call format and VCFtools. *Bioinformatics* **27**(15)**:** 2156−2158.

Du Q, Yang X, Xie J, Quan M, Xiao L, Lu W *et al* (2019). Time-specific and pleiotropic quantitative trait loci coordinately modulate stem growth in Populus. *Plant biotechnology journal* **17**(3)**:** 608-624.

Egertsdotter U (2019). Plant physiological and genetical aspects of the somatic embryogenesis process in conifers. *Scand J For Res* **34**(5)**:** 360-369.

Elfstrand M, Baison J, Lundén K, Zhou L, Vos I, Capador-Barreto HD *et al* (2020). Association genetics identifies a specifically regulated Norway spruce laccase gene, *PaLAC5*, linked to *Heterobasidion parviporum*-resistance. *Plant, Cell & Environment* **43**(7)**:** 1779–1791.

Estopa RA, Paludeto JGZ, Muller BSF, de Oliveira RA, Azevedo CF, de Resende MDV *et al* (2022). Genomic prediction of growth and wood quality traits in Eucalyptus benthamii using different genomic models and variable SNP genotyping density. *New Forests*.

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

Gilmour AR, Gogel BJ, Cullis BR, Welham SJ, Thompson R (2015). ASReml user guide release 4.1 structural specification. *Hemel hempstead: VSN international ltd*.

Grattapaglia D (2022). Twelve Years into Genomic Selection in Forest Trees: Climbing the Slope of Enlightenment of Marker Assisted Tree Breeding. *Forests* **13**(10)**:** 1554.

Grattapaglia D, Resende MDV (2011). Genomic selection in forest tree breeding. *Tree Genet Genomes* **7**(2)**:** 241−255.

Habier D, Fernando RL, Garrick DJ (2013). Genomic BLUP Decoded: A Look into the Black Box of Genomic Prediction. *Genetics* **194**(3)**:** 597-607.

Hall D, Hallingbäck HR, Wu HX (2016). Estimation of number and size of QTL effects in forest tree traits. *Tree Genet Genomes* **12**(6)**:** 110.

Hickey JM, Chiurugwi T, Mackay I, Powell W, Hickey JM, Chiurugwi T *et al* (2017). Genomic prediction unifies animal and plant breeding programs to form platforms for biological discovery. *Nat Genet* **49**(9)**:** 1297-1303.

Hill WG, Goddard ME, Visscher PM (2008). Data and theory point to mainly additive genetic variance for complex traits. *PLOS Genet* **4**(2)**:** e1000008.

Huang M, Liu X, Zhou Y, Summers RM, Zhang Z (2019). BLINK: a package for the next level of genome-wide association studies with both individuals and markers in the millions. *GigaScience* **8**(2)**:** 1–12.

Ingvarsson PK, Dahlberg H (2018). The effects of clonal forestry on genetic diversity in wild and domesticated stands of forest trees. *Scand J For Res***:** 1-10.

Isik F (2014). Genomic selection in forest tree breeding: the concept and an outlook to the future. *New Forests* **45**(3)**:** 379-401.

Krutzsch P (1975). *Die Pflanzschulenergebnisse eines inventierenden Fichtenherkunftsversuches*. Department of Forest Genetics, Royal College of Forestry: Stockholm.

Lynch M, Walsh B (1998). Genetics and analysis of quantitative traits.

Merrick LF, Burke AB, Chen XM, Carter AH (2021). Breeding With Major and Minor Genes: Genomic Selection for Quantitative Disease Resistance. *Front Plant Sci* **12**.

Mullin TJ, Park YS (1992). Estimating genetic gains from alternative breeding strategies for clonal forestry. *Can J For Res* **22**(1)**:** 14–23.

Muñoz PR, Resende MFR, Gezan SA, Resende MDV, de los Campos G, Kirst M *et al* (2014). Unraveling additive from nonadditive effects using genomic relationship matrices. *Genetics* **198**(4)**:** 1759−1768.

Nguyen HTH, Chen Z-Q, Fries A, Berlin M, Hallingbäck HR, Wu HX (2022). Effect of additive, dominant and epistatic variances on breeding and deployment strategy in Norway spruce. *Forestry: An International Journal of Forest Research* **95**(3)**:** 416 - 427.

Nystedt B, Street NR, Wetterbom A, Zuccolo A, Lin Y-C, Scofield DG *et al* (2013). The Norway spruce genome sequence and conifer genome evolution. *Nature* **497:** 579–584.

Otto SP, Jones CD (2000). Detecting the Undetected: Estimating the Total Number of Loci Underlying a Quantitative Trait. *Genetics* **156**(4)**:** 2093-2107.

Perron M, DeBlois J, Desponts M (2013). Use of resampling to assess optimal subgroup composition for estimating genetic parameters from progeny trials. *Tree Genet Genomes* **9**(1)**:** 129−143.

Resende MFR, Muñoz P, Resende MDV, Garrick DJ, Fernando RL, Davis JM *et al* (2012). Accuracy of genomic selection methods in a standard data set of loblolly pine (*Pinus taeda* L.). *Genetics* **190**(4)**:** 1503−1510.

Rosvall O (2019a). Using Norway spruce clones in Swedish forestry: general overview and concepts. *Scand J For Res* **34**(5)**:** 336-341.

Rosvall O (2019b). 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.

Rosvall O, Bradshaw RHW, Egertsdotter U, Ingvarsson PK, Mullin TJ, Wu H (2019). Using Norway spruce clones in Swedish forestry: Implications of clones for management. *Scand J For Res***:** 1−48.

Sanchez L, Yanchuk AA, King JN (2008). Gametic models for multitrait selection schemes to study variance of response and drift under adverse genetic correlations. *Tree Genet Genomes* **4**(2)**:** 201-212.

Sarinelli JM, Murphy JP, Tyagi P, Holland JB, Johnson JW, Mergoum M *et al* (2019). Training population selection and use of fixed effects to optimize genomic predictions in a historical USA winter wheat panel. *Theor Appl Genet* **132**(4)**:** 1247-1261.

Schmidt-Vogt H (1977). Die Fichte. Vol. 1. *Verlag P Parey, Hamburg*.

Sehgal D, Rosyara U, Mondal S, Singh R, Poland J, Dreisigacker S (2020). Incorporating Genome-Wide Association Mapping Results Into Genomic Prediction Models for Grain Yield and Yield Stability in CIMMYT Spring Bread Wheat. *Front Plant Sci* **11**.

Shalizi MN, Cumbie WP, Isik F (2021). Genomic prediction for fusiform rust disease incidence in a large cloned population of Pinus taeda. *G3* **11**(9)**:** jkab235.

Tam V, Patel N, Turcotte M, Bosse Y, Pare G, Meyre D (2019). Benefits and limitations of genome-wide association studies. *Nat Rev Genet* **20**(8)**:** 467–484.

Tan B, Grattapaglia D, Martins GS, Ferreira KZ, Sundberg B, Ingvarsson PK (2017). Evaluating the accuracy of genomic prediction of growth and wood traits in two *Eucalyptus* species and their F1 hybrids. *BMC Plant Biol* **17**(1)**:** 110.

Tan B, Grattapaglia D, Wu HX, Ingvarsson PK (2018). Genomic relationships reveal significant dominance effects for growth in hybrid *Eucalyptus*. *Plant Sci* **267:** 84−93.

Tan B, Ingvarsson PK (2022). Integrating genome-wide association mapping of additive and dominance genetic effects to improve genomic prediction accuracy in Eucalyptus. *Plant Genome* **15**(2)**:** e20208.

Thistlethwaite FR, Ratcliffe B, Klápště J, Porth I, Chen C, Stoehr MU *et al* (2017). Genomic prediction accuracies in space and time for height and wood density of Douglas-fir using exome capture as the genotyping platform. *BMC Genom* **18**(1)**:** 930.

Thumma BR, Joyce KR, Jacobs A (2022). Genomic studies with preselected markers reveal dominance effects influencing growth traits in Eucalyptus nitens. *G3-Genes Genomes Genetics* **12**(1).

VanRaden PM (2008). Efficient methods to compute genomic predictions. *J Dairy Sci* **91**(11)**:** 4414−4423.

Visscher PM, Wray NR, Zhang Q, Sklar P, McCarthy MI, Brown MA *et al* (2017). 10 years of GWAS discovery: biology, function, and translation. *Am J Hum Genet* **101**(1)**:** 5–22.

Vitezica ZG, Reverter A, Herring W, Legarra A (2018). Dominance and epistatic genetic variances for litter size in pigs using genomic models. *Genet Sel Evol* **50**(1)**:** 71.

Vitezica ZG, Varona L, Legarra A (2013). On the additive and dominant variance and covariance of individuals within the genomic selection scope. *Genetics* **195**(4)**:** 1223−1230.

Wainschtein P, Jain D, Zheng Z, Aslibekyan S, Becker D, Bi W *et al* (2022). Assessing the contribution of rare variants to complex trait heritability from whole-genome sequence data. *Nat Genet*

Walker TD, Cumbie WP, Isik F (2021). Single-Step Genomic Analysis Increases the Accuracy of Within-Family Selection in a Clonally Replicated Population of Pinus taeda L. *For Sci*.

Wang J, Zhang Z (2021). GAPIT Version 3: boosting power and accuracy for genomic association and prediction. *Genomics, proteomics & bioinformatics* **19**(4)**:** 629-640.

Wu HX (2018). Benefits and risks of using clones in forestry – a review. *Scand J For Res***:** 1−8.

Yang J, Weedon MN, Purcell S, Lettre G, Estrada K, Willer CJ *et al* (2011). Genomic inflation factors under polygenic inheritance. *Eur J Hum Genet* **19**(7)**:** 807–812.

Table 1 Descriptive statistics of traits observed in four sites measured for the Norway spruce sampled and analyzed

Trait	S1387	S1388	S1389	S1390	N	Mean	SD	Min	Max	CV	Unit
HT6	√	√	√	√	914	26.7	3.5	16.3	36.7	13.1%	dm
HT12	√				909	67.0	10.8	14.0	111.0	16.1%	dm
DBH	√	√	√	√	914	81.9	10.4	41.6	122.3	12.7%	mm
PILO	√				900	19.1	1.9	13.5	26.3	9.9%	mm
BB		√	√	√	914	4.4	1.1	1.6	6.0	25.0%	Category
FD		√			908	0.38	0.40	0	2.00	105.2%	Category

Abbreviation: HT6, total tree height at age 6; HT12, total tree height at age 12; DBH, diameter at breast height at age 12; PILO, Pilodyn penetration; BB, budburst stage; FD, frost damage of fresh new buds.

Table 2 Ratios of additive, dominance, and residual of genetic variance component to the total phenotypic variance for different PBLUP and GBLUP models.

Model	σ_a^2	σ_{as}^2	σ_r^2	σ_{rs}^2	$\bar{\sigma}_e^2$	h_i^2 (SE)	H_i^2(SE)	AIC
HT6								
PBLUP-A	6.48 (0.58)	1.13 (0.30)			28.99 (0.48)	0.18 (0.01)		43680.9
PBLUP-AR*	4.44 (1.33)	0.60 (0.23)	0.92 (0.78)	1.36 (0.39)	28.24 (0.50)	0.13 (0.04)	0.15 (0.02)	43670.3
GBLUP-A	6.37 (0.56)	1.08 (0.28)			29.05 (0.47)	0.17 (0.01)		43672.7
GBLUP-AR*	4.02 (0.84)	0.62 (0.23)	1.13 (0.49)	1.31 (0.39)	28.24 (0.50)	0.11 (0.02)	0.15 (0.02)	43656.4
HT12								
PBLUP-A	30.7 (5.0)				125.2 (4.0)	0.20 (0.03)		16649.4
PBLUP-AR*	13.0 (5.2)		12.5 (4.3)		123.6 (4.0)	0.09 (0.03)	0.17 (0.02)	16646.2
GBLUP-A	26.6 (4.5)				127.0 (4.0)	0.17 (0.03)		16651.1
GBLUP-AR*	12.9 (4.5)		12.28(4.0)		123.6 (3.7)	0.09 (0.03)	0.17 (0.02)	16644.6
DBH								
PBLUP-A	60.0 (5.5)	8.8 (2.8)			313.4 (1.9)	0.16 (0.01)		64017.5
PBLUP-AR	57.4 (5.7)	5.9 (2.6)	0 (0)	8.1 (4.3)	313.8 (5.8)	0.15 (0.01)	0.15 (0.01)	64017.9
GBLUP-A	56.8 (5.6)	8.7 (2.7)			319.6 (5.3)	0.15 (0.01)		64023.2
GBLUP-AR*	31.2 (7.3)	6.4 (2.6)	14.8 (4.7)	7.5 (4.3)	313.7 (5.8)	0.09 (0.02)	0.13 (0.01)	64011.3
PILO								
PBLUP-A*	1.78 (0.14)				1.51 (0.05)	0.54 (0.02)		4583.2
PBLUP-AR	1.78 (0.14)		0 (0)		1.51 (0.05)	0.54 (0.02)		4585.2
GBLUP-A	1.73 (0.14)				1.53 (0.05)	0.53 (0.02)		4580.9
GBLUP-AR*	1.33 (0.20)		0.25 (0.11)		1.51 (0.05)	0.43 (0.05)	0.51 (0.02)	4576.9
BB								
PBLUP-A	1.30 (0.07)	0.14 (0.01)			0.33 (0.007)	0.73 (0.01)		2671.3
PBLUP-AR*	0.98 (0.23)	0.09 (0.02)	0.15 (0.12)	0.04 (0.01)	0.33 (0.007)	0.62 (0.11)	0.71 (0.02)	2667.5
GBLUP-A	1.25 (0.07)	0.13 (0.01)			0.33 (0.007)	0.73 (0.01)		2591.5
GBLUP-AR*	1.05 (0.12)	0.08 (0.02)	0.09 (0.05)	0.04 (0.01)	0.32 (0.007)	0.66 (0.04)	0.72 (0.01)	2579.1
FD								
PBLUP-A*	0.014 (0.004)				0.201 (0.006)	0.06 (0.02)		-1354.3
PBLUP-AR	0.014 (0.004)		0 (0)		0.208 (0.006)	0.06 (0.02)	0.06 (0.02)	-1352.3
GBLUP-A*	0.015 (0.004)				0.208 (0.006)	0.07 (0.02)		-1357.0
GBLUP-AR	0.015 (0.005)		0 (0)		0.208 (0.007)	0.07 (0.02)	0.07 (0.02)	-1355.0

$\sigma_a^2, \sigma_{as}^2, \sigma_r^2, \sigma_{rs}^2, \bar{\sigma}_e^2$ represents variances of additive, additive-by-site, residual of genetic, residual of genetic-by-site, and the average of residual effects, respectively. h_i^2 and H_i^2 represent the narrow-sense and broad-sense heritabilities, respectively. AIC represents Akaike information criterion. * represents that the model showed the smallest AIC value compared to the same other four GBLUP or PBLUP models.

Table 3 Predictive ability (PA) of different types of pedigree- and genome-based prediction models following a 10-fold cross-validation procedure.

Model	HT6	HT12	DBH	PILO	BB	FD
h_c	0.72	0.42	0.69	0.87	0.89	0.37
PBLUP-C	0.38 (0.01)	0.22 (0.01)	0.44 (0.01)	0.62 (0.01)	0.61 (0.01)	0.21 (0.01)
GBLUP	***0.42 (0.01)***	0.24 (0.01)	***0.44 (0.01)***	***0.63 (0.01)***	0.67 (0.01)	0.23 (0.01)
GBLUP-S	0.35 (0.01)	0.22 (0.01)	0.37 (0.01)	0.58 (0.01)	***0.72 (0.01)***	0.20 (0.01)
GBLUP-F	***0.42 (0.01)***	***0.26 (0.01)***	***0.44 (0.01)***	***0.63 (0.01)***	0.70 (0.00)	***0.25 (0.01)***
Accuracy	0.58	0.62	0.64	0.72	0.81	0.68

The prediction models showing the highest PA for a trait have their PA-estimated highlighted in ***italic bold***.

The h_c is the square root of the clonal mean narrow sense heritability (h_c^2) based on PBLUP-AR model. PBLUP-C is the traditional pedigree-based best linear unbiased prediction (BLUP) including marker-based pedigree correction; GBLUP is the genomic-based BLUP; GBLUP-S is genomic-based BLUP and with marker-preselection based on the smallest *p*-values from genome-wide association analyses. In this table, the 100 smallest *p*-value SNPs were preselected for the budburst stage (BB) and 2000 SNPs were preselected for the other traits. GBLUP-F is a genomic-based BLUP model plus the SNP with the smallest *p*-value fitted as an additional fixed regression effect. The accuracy for the model showing the highest PA is here roughly estimated as PA/h_c.

Table 4 Predictive abilities and accuracy estimates with genomic prediction model with and without adding an SNP with the largest effect size as a fixed effect (GBLUP and GBLUP-F) based on simulated data using a trait with a heritability of 0.25 and evaluated using 10-fold cross-validations.

PVE (%)	Overall predictive ability		Within-family predictive ability		Overall accuracy		Within-family accuracy	
	- QTN	+ QTN	- QTN	+ QTN	- QTN	+ QTN	- QTN	+ QTN
5	0.31 (0.06)	0.35 (0.06)	-0.04 (0.07)	0.05 (0.06)	0.60 (0.05)	0.71 (0.04)	0.02 (0.08)	0.21 (0.06)
2.5	0.33 (0.07)	0.34 (0.06)	-0.02 (0.08)	0.00 (0.07)	0.63 (0.06)	0.66 (0.05)	0.03 (0.07)	0.13 (0.07)
1.25	0.31 (0.06)	0.31 (0.07)	-0.04 (0.06)	-0.02 (0.08)	0.63 (0.05)	0.64 (0.05)	0.02 (0.07)	0.08 (0.07)
0.25	0.30 (0.07)	0.30 (0.06)	-0.04 (0.07)	-0.03 (0.06)	0.61 (0.05)	0.60 (0.06)	0.03 (0.06)	0.04 (0.08)
0	0.32 (0.06)	0.31 (0.07)	-0.03 (0.07)	-0.03 (0.07)	0.62 (0.06)	0.61 (0.07)	0.02 (0.08)	0.02 (0.07)

PVE is the percentage of phenotypic variance explained by a large effect size SNP. QTN represents the quantitative trait nucleotide with a large effect size where - and + signifies the absence (GBLUP) and presence (GBLUP-F) respectively, of this QTN in the genomic prediction model as a fixed effect. The value in parenthesis indicates the standard deviation across replicate simulations.

Table 5 Predictive ability for clone mean phenotypic variation and their standard errors in parenthesis using two different models.

Model	HT6	HT12	DBH	PILO	BB	FD
GBLUP-All	0.11 (0.05)	0.03 (0.04)	0.05 (0.04)	0.05 (0.04)	0.19 (0.05)	0.03 (0.04)
GBLUP-S	0.03 (0.04)	0.07 (0.04)	-0.02 (0.04)	0.08 (0.03)	0.25 (0.06)	0.04 (0.03)
GBLUP-F	0.09 (0.04)	0.02 (0.04)	0.04 (0.04)	0.04 (0.04)	0.24 (0.06)	-0.01 (0.04))

In this scenario (GBLUP), each of the 32 full-sib families was a validated set and a training set using the rest of the 31 families. In total, 32 cross-validations were repeated to calculate the predictive ability for phenotypic variation.

GBLUP-All, GBLUP model with all SNPs; GBLUP-S, GBLUP model with preselected SNPs (100 SNPs for budburst stage (BB), 2000 SNPs for the rest of the traits); GBLUP-F, GBLUP model with the most significant p value SNP as a fixed effect.

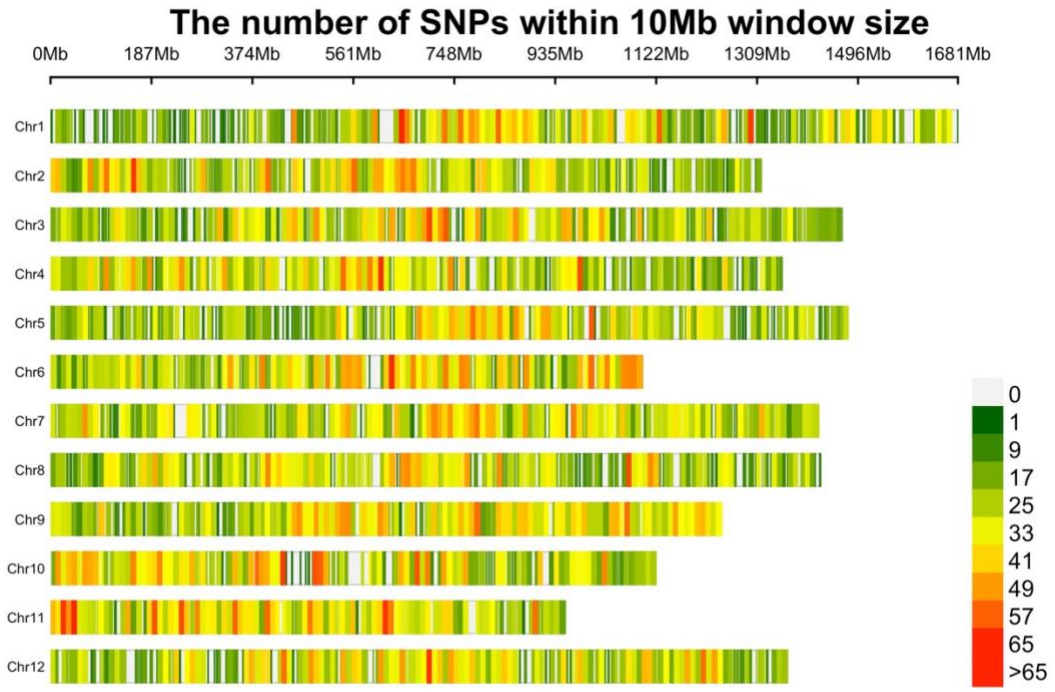

Fig.1 Marker density of the 50k Norway spruce array based on a 10Mb window size for each of 12 chromosomes.

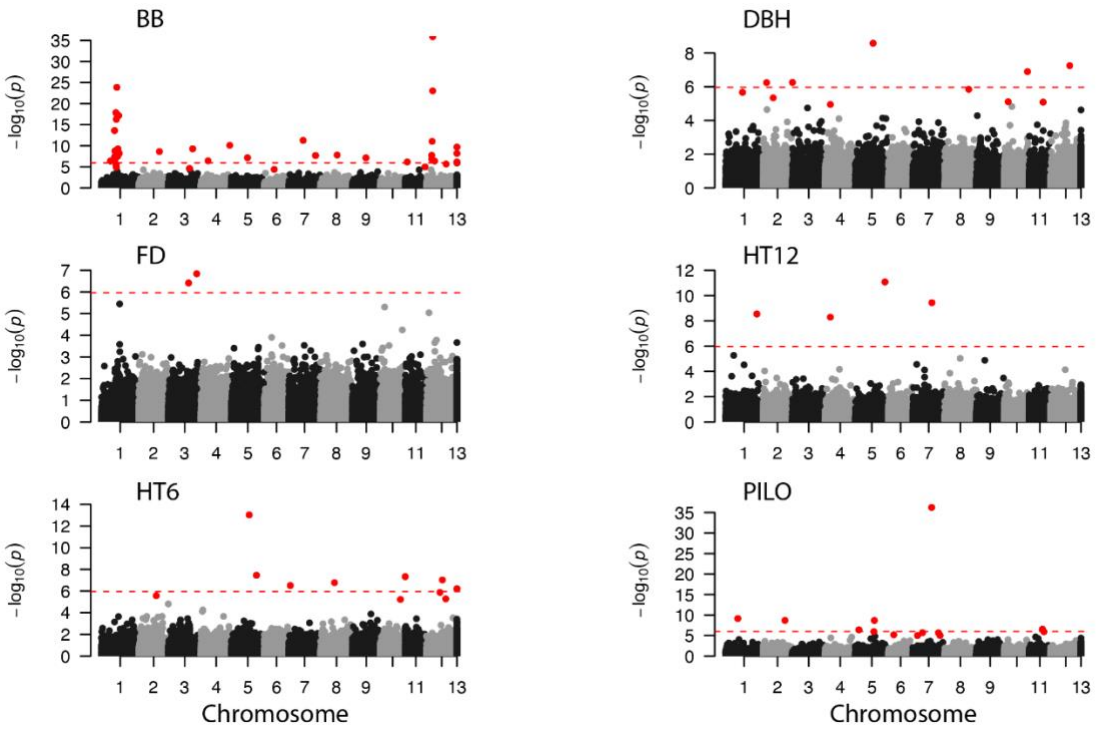

Fig. 2. Manhattan plots for six traits. The red dashed line represents the significant threshold of $p = 1.7 \times 10^{-6}$ after the Bonferroni correction. The red dots represent that the SNPs passed the false discovery rate test threshold of 0.05 based on Benjamini and Hochberg (1995). SNPs which were not mapped into the Norway spruce genome v2 (In preparation) were grouped into a region assumed as chromosome 13 in this study. BB, bud burst stage; DBH, diameter at breast height; FD, frost damage; HT12, tree height at field tree age 12; HT6, tree height at field age six; PILO, Pilodyn penetration.

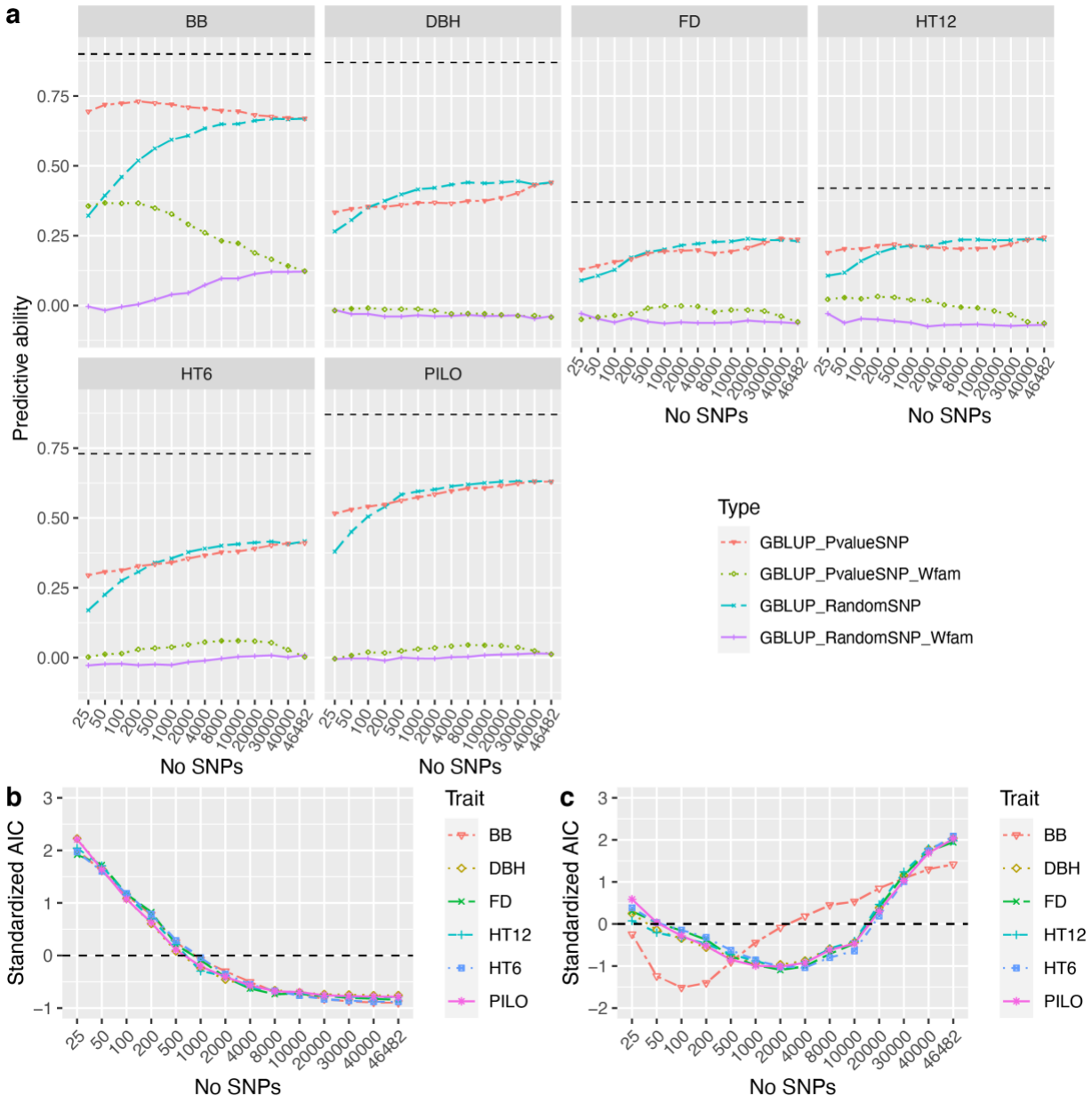

Fig. 3. Predictive ability estimates (PA) for six traits using GBLUP-A models based on 10 repeats of 10-fold cross-validations (n=100) and 904 clone means. a) PAs for trait's phenotypic values based on GBLUP-A using different numbers of SNPs randomly selected (blue longdashed line), SNPs selected based on lowest *p*-values in GWAS performed on the training dataset using BLINK method (red twodashed line), and PAs of within family variation (clone mean - family mean) using different number of SNPs based on both randomly selected (purple solid line) and the lowest *p*-values based on GWAS (green dotted line). The black horizontal dashed line in a) is the square root of narrow-sense clone mean heritability estimated based on the PBLUP-AR model. Matching to GBLUP-A models in a), the standardized Akaike information criterion (AIC) values for each model is shown in b) based on randomly selected SNPs and in c) based on SNPs selected by smallest *p*-values.

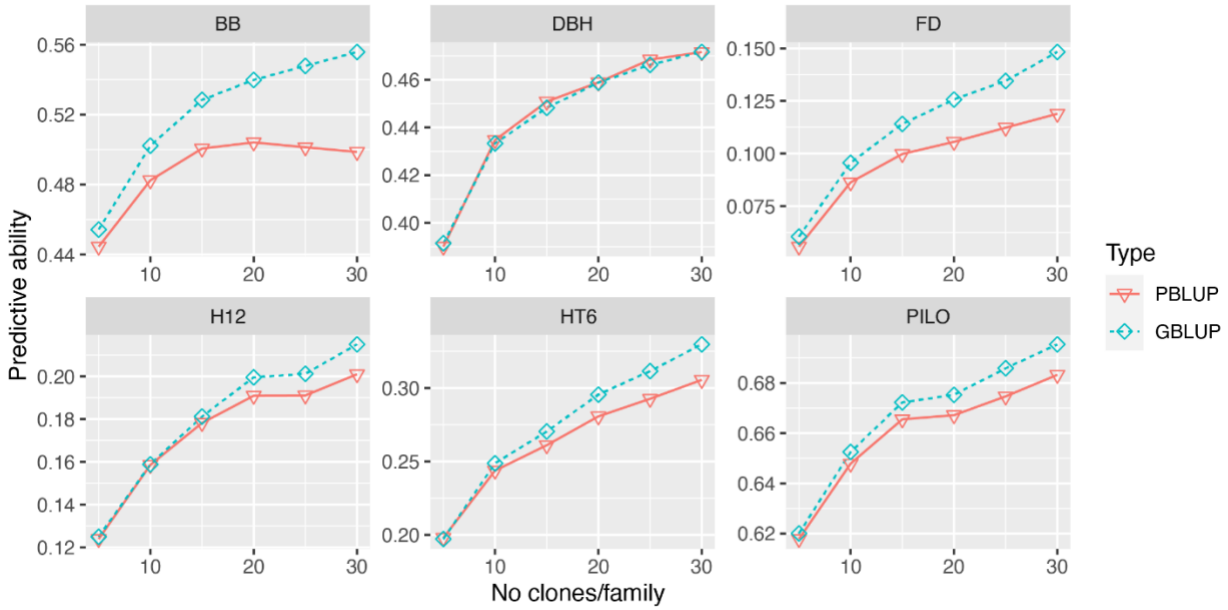

Fig. 4. Predictive abilities (PAs) for six traits using PBLUP-A and GBLUP-A models using all available SNPs but employing different number (5, 10, 15, 20, 25, and 30) of clones per family for model training leaving the remainder number of clones for validation purposes. Only ten families comprising the largest numbers of offspring were included in this analysis.

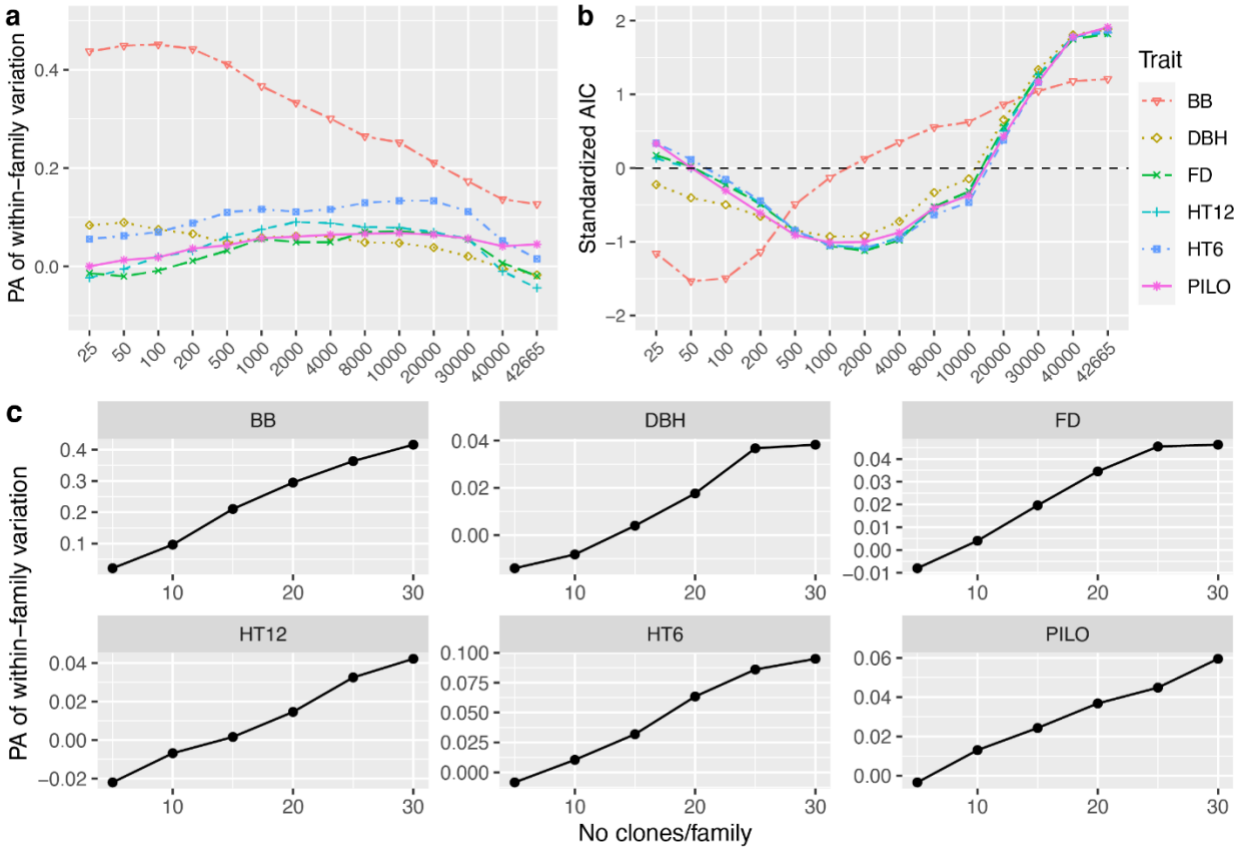

Fig. 5. a) Predictive ability (PA) for within-family variation using 10-fold cross-validation for the ten largest families with 482 clones. b) the standardized Akaike information criterion (AIC) values for different simulation strategies based on different numbers of SNPs with the smallest p values from GWAS. c) PA for within-family variation based on the GBLUP model with G matrix estimated from 100 SNPs for budburst and 2000 SNPs for the rest of the traits based on cross-validation GWAS in the ten families.

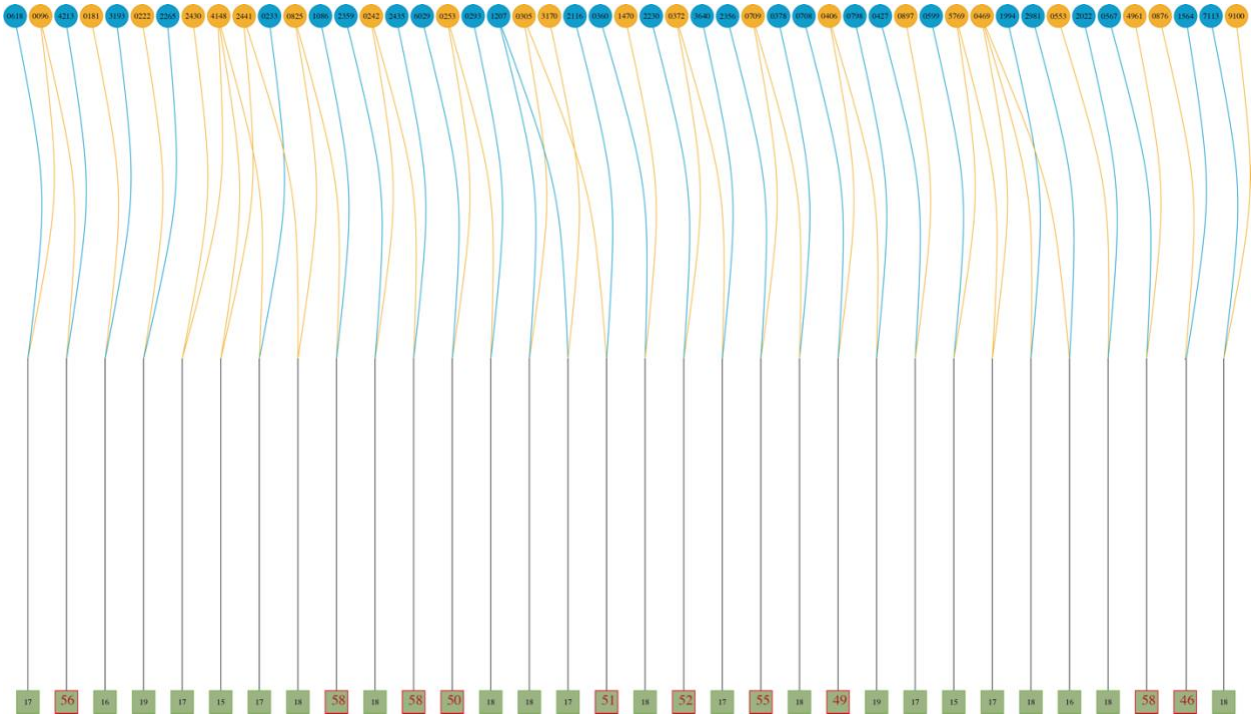

Fig. S1 A mating design of 49 parents constructed into 32 full-sib families. Each circle in the top level represents a parent and the number is the identity of the parent. Each square at the bottom represents a full-sib family. The number within each square represents the number of clones per family genotyped. The families whose squares are framed in red color, were selected to test the effect of family size to predictive ability.

If the parents only cross once, then dark goldenrod means male and dark skyblue means female. Otherwise, the color of parents could be any of both colors.

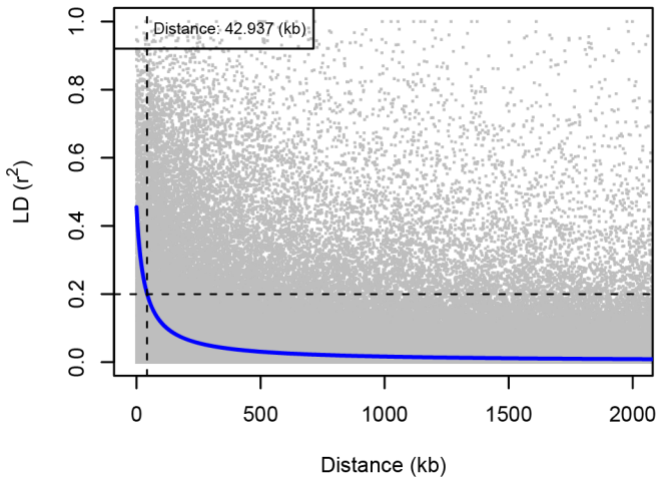

Fig. S2 LD decay (r2) in a full-sib progeny population (F1-generation, n=904 from 49 parents based on the original pedigree) based on Norway spruce 50k SNP array. The Vertical dashed line represents LD decay distance when r^2=0.2.

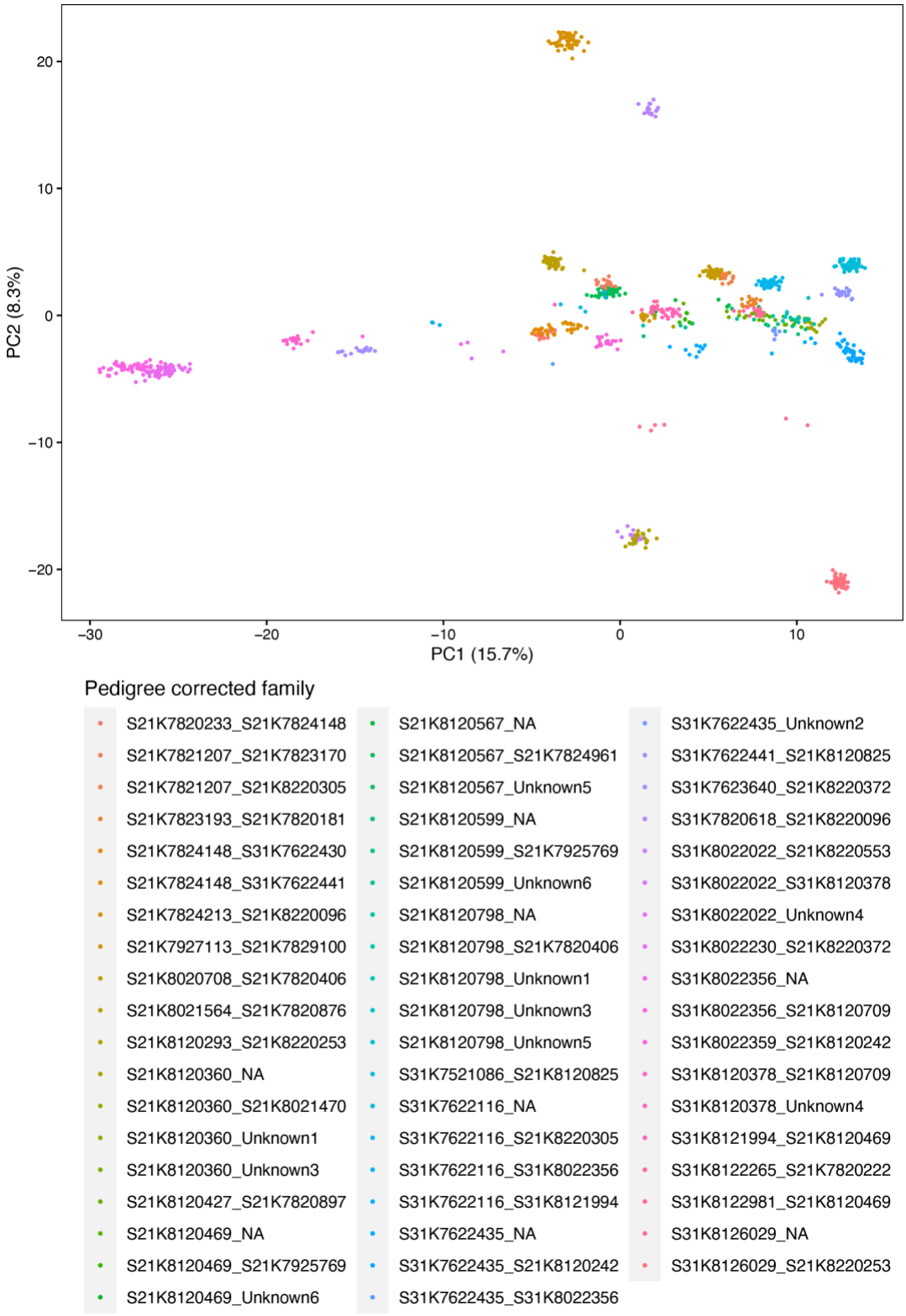

Fig. S3 Population structure for the 904 clones coloured based on the family identities.