

RESEARCH

Open Access



Genomic dissection of additive and non-additive genetic effects and genomic prediction in an open-pollinated family test of Japanese larch

Leiming Dong^{1,2}, Yunhui Xie¹, Yalin Zhang¹, Ruizhen Wang² and Xiaomei Sun^{1*}

Abstract

Genomic dissection of genetic effects on desirable traits and the subsequent use of genomic selection hold great promise for accelerating the rate of genetic improvement of forest tree species. In this study, a total of 661 offspring trees from 66 open-pollinated families of Japanese larch (*Larix kaempferi* (Lam.) Carrière) were sampled at a test site. The contributions of additive and non-additive effects (dominance, imprinting and epistasis) were evaluated for nine valuable traits related to growth, wood physical and chemical properties, and competitive ability using three pedigree-based and four Genomics-based Best Linear Unbiased Predictions (GBLUP) models and used to determine the genetic model. The predictive ability (PA) of two genomic prediction methods, GBLUP and Reproducing Kernel Hilbert Spaces (RKHS), was compared. The traits could be classified into two types based on different quantitative genetic architectures: for type I, including wood chemical properties and Pilodyn penetration, additive effect is the main source of variation (38.20–67.46%); for type II, including growth, competitive ability and acoustic velocity, epistasis plays a significant role (50.76–91.26%). Dominance and imprinting showed low to moderate contributions (< 36.26%). GBLUP was more suitable for traits of type I (PAs = 0.37–0.39 vs. 0.14–0.25), and RKHS was more suitable for traits of type II (PAs = 0.23–0.37 vs. 0.07–0.23). Non-additive effects make no meaningful contribution to the enhancement of PA of GBLUP method for all traits. These findings enhance our current understanding of the architecture of quantitative traits and lay the foundation for the development of genomic selection strategies in Japanese larch.

Keywords Genomic prediction, Dominance, Epistasis, GBLUP, RKHS, Japanese larch

Introduction

Genetic improvement of economically and ecologically important species depends on accurate decomposition of phenotypic variation and prediction of breeding values for traits of interest. Phenotypic variation can be decomposed into genetic, environmental, and interaction variance according to quantitative genetic theory [1]. Genetic variance can be further decomposed into additive and non-additive variance (dominance, imprinting and epistasis). Traditionally, the genetic effects are estimated upon pedigree information. For instance, the additive

*Correspondence:

Xiaomei Sun
xmsun@caf.ac.cn

¹ State Key Laboratory of Tree Genetics and Breeding, Key Laboratory of Tree Breeding and Cultivation of State Forestry and Grassland Administration, Research Institute of Forestry, Chinese Academy of Forestry, Beijing 100091, China

² Key Laboratory of National Forestry and Grassland Administration on Plant Ex situ Conservation, Beijing Floriculture Engineering Technology Research Centre, Beijing Botanical Garden, Beijing 100093, China



© The Author(s) 2023. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

effect can easily be obtained from a half-sib progeny test. In contrast, the analysis of non-additive variance requires the establishment of a more complex family structure (e.g., full-sib families) [2]. Typically, full-sib families are derived from crossbreeding, which is time-consuming and logistically challenging, especially in the early stages of breeding cycles, or impractical for some species in the early stages of breeding, such as many forest trees. Pedigree reconstruction is an alternative method to complete the unknown male-parent information by molecular markers and paternity analysis on half-sib families [3]. However, the expected theoretical relationships between members of half- or full-sib families may bias the estimates of variance component, breeding value, and expected genetic gain for the Mendelian sampling effect [4]. Fortunately, the use of Genomic-based Best Linear Unbiased Prediction (GBLUP) and the reduced cost of genotyping provide the opportunity not only to estimate the non-additive effect [5–7] but also to obtain more accurate genetic parameters [8–10] using the realized relationship matrix constructed by molecular markers, facilitating a better understanding of the genetic basis of key traits. Moreover, the non-additive effects could be selectively considered into genomic selection models based on their contributions.

Genomic selection (GS), introduced by Meuwissen et al. in 2001 [11], uses the developed models and genomic markers to predict genomic breeding values of selection candidates, offering the potential to significantly shorten the breeding cycle and increase gains per unit time. These advantages make GS attractive for improving the traits of animals (e.g. [12, 13]), crops (e.g. [14]), and especially long-lived forest trees (e.g. [15, 16]). Typically, non-additive effects and prediction methods are important aspects that need to be carefully considered when developing GS models. Both simulation and empirical studies have been carried out to evaluate the impact of non-additive effects on the predictive ability of the models [17–20]. Dominance improved the prediction of complex traits in plants, such as grain production and drought tolerance in maize hybrids [21], yield in sorghum [22], or growth in hybrid *Eucalyptus* [23]. Epistasis can influence the accuracy of GS for complex traits in plants. Jiang et al. [24] found that epistatic effects played a more prominent role in grain-yield heterosis in wheat using GS. Raffo et al. [25] showed that epistatic models can be useful to enhance predictions of total genetic merit for wheat grain yield. However, the inclusion of epistasis in GS models may reduce the accuracy of the models [26]. In addition, few studies have been focused on imprinting in GS [7, 27], although this non-additive effect has been shown to play a role in the expression of several phenotypes in plants [28] and animals [29]. It also should be

noted that the influence of non-additive effects on GS may depend on factors such as the type of trait under consideration, the genetic architecture of the trait, and the size and structure of the population being analyzed. A number of prediction methods have been proposed for GS, such as Genomic Best Linear Unbiased Prediction (GBLUP), Ridge Regression, Bayesian-based models, and Reproducing Kernel Hilbert Spaces (RKHS) [30, 31], and the performance of these methods needed to be evaluated in a GS programme. Previous empirical studies have shown that for traits genetically regulated by a large number of genes with small effects, there is no statistically significant difference in performance between GBLUP and Bayesian-based models [32–37]. The Reproducing Kernel Hilbert Space (RKHS) model is a powerful mathematical framework used in machine learning and statistical learning theory that provides a way to represent and analyze data in a high-dimensional feature space using kernel functions [30, 31]. When non-additive effects, particularly epistasis, play a role in a trait, Reproducing Kernel Hilbert Spaces (RKHS), which combine features of non-parametric kernel regression with mixed-effect linear models [38], usually have a better predictive ability [39–41]. Therefore, it is both necessary and promising to determine the contribution of non-additive effects and then to assess the predictive ability of RKHS models using GBLUP as a benchmark for those traits that are significantly regulated by non-additive effects.

Japanese larch (*Larix kaempferi* (Lam.) Carrière) is an economically and ecologically important tree characterized by rapid juvenile growth rate and wide environmental adaptability. It is one of the most successful exotic species and is rapidly becoming the preferred conifer for lumber and pulpwood production in northern China and the southern sub-alpine region. In 1965, a 14.7 ha clonal seed orchard was established at the Dagujia Tree Farm in Liaoning Province, China (42.3–42.4°N latitude, 124.8–125.2°E longitude, with an altitude of 200–600 m). Plantations established in the northeastern China in the early twentieth century provided the clones with outstanding growing performance. For many years, these clones have been the subject of progeny testing through open-pollinated and control-pollinated crosses. The genetic material for this study came from one of the progeny tests, which was established in 1988, which consists of 94 open-pollinated (OP) and 55 control-pollinated (CP) families. Based on this population, we have investigated the additive and dominant effects, spatial patterns and competition for growth traits [42, 43], implemented pedigree reconstruction using microsatellite markers [44], and evaluated the influence of pedigree reconstruction on genetic parameter estimation [45]. In this study, the genetic research and the GS model development were

carried out on a sample of 661 trees from 66 OP families. Individual trees were phenotyped for nine valuable traits related to growth traits, wood physical and chemical traits, and competition index and were genotyped using 11,333 high-quality SNPs captured by the Genotyping by Sequencing (GBS). The main objectives of this study were to (1) compare the model fitting and parameter estimates between the OP-pedigree and full-sib (FS) pedigree reconstructed in our previous study; (2) determine the contribution of additive, dominance, imprinting, and first-order epistasis using the GBLUP method and construct a GS model for each trait based on the importance of non-additive components; and (3) assess the predictive ability of the RKHS models using GBLUP as a benchmark for each trait, and select the better one for further application.

Materials and methods

Genetic material

The data used in this study are from a progeny trial of the Japanese larch tree improvement programme in the Temperate Breeding Zone, China. The field trial, transplanted to the field in 1988 with 2-year-old seedlings, comprises 94 open-pollinated families. The seeds of the open-pollinated families were collected from the first-generation seed orchard of Japanese larch in China, established with 152 maternal clones [44]. The field design was a randomized complete block design with 5 replications for each family. The number of trees per family varied from 5 to 15 per replications and the number of 5-tree row plot per family varied from 1 to 3. 5-tree row plots with 1–3 columns due to seedling availability and 2×2 m spacing. This study involved 661 randomly selected trees from 66 open-pollinated families with 9–12 trees per family.

Phenotypic data

Growth

In autumn of 2005 (age 19), all trees in the trial were assessed and measured for growth traits. Tree height (H, m) was measured using a Vertex III sonic clinometer (Haglöf Sweden AB, Västernorrland, Sweden). Diameter at breast height (DBH, cm) was measured using a diameter tape. Tree stem volume (V, dm³) was calculated by the following equation developed for larch trees in eastern Liaoning Province (Liao Q1667-83):

$$V = 0.0592372 \times DBH^2 \times H^{0.98098926} \quad (1)$$

Competition index

DBH data were used to calculate competition indices to characterize the competitive ability of individuals to their

neighbors. The Hegyi index, proposed by Hegyi in 1974 [46], is the most classical spatial competition index used in forest research to quantify the competitive pressure of trees in forests [47]. Hegyi's competition index depends on the size and distance of neighboring trees, and was calculated as follows:

$$CI = \sum_{j=1}^n (DBH_j/DBH_i)/D_{ij} \quad (2)$$

where i is the i th tree, j is the j th neighbor of this tree, $1 \leq n \leq 8$, D_{ij} is the distance between the i th tree with its j th neighbor. The first-order neighbors were considered in this study, so $D_{ij}=2$ m (in row or column direction) or $2\sqrt{2}$ m (in diagonal direction).

Wood physical traits

Pilodyn penetration The Pilodyn 6 J Forest penetrometer (PROCEQ, Switzerland) with a 2.0 mm diameter pin was inserted into each stem twice, without removing the bark, in the southern and northern directions at approximately 1.3 m above ground. The average Pilodyn penetration (PILODYN) was used as a surrogate for the wood density characteristic.

Acoustic assessment Acoustic velocity (AV) is directly related to dynamic wood stiffness or modulus of elasticity (MOE) in conifers and has been widely used as a proxy [48–50]. The AV of each selected and sampled tree was measured using the Fakkop Microsecond Timer tool (Fakopp Enterprise, Ágfalva, Hungary) to substitute the dynamic modulus of elasticity estimate at age 32. Two probes were inserted lengthwise into the tree trunks at heights of 0.3 and 1.3 m. Two measurements taken from both the north and south sides of the trunk and averaged before analysis.

Wood chemical traits

Sample preparation In winter 2017 (age 27), one to three 5 mm increment cores were extracted from each sampled tree at breast height ± 10 cm from bark to bark with an increment borer to meet the needs of subsequent NIR scanning. Cores without bark were air dried to constant weight (wood samples were dried to an equilibrium moisture content of 12%), ground to wood meal, filtered through a 60-mesh sieve and stored in sealed plastic bags for subsequent NIR spectra scanning.

Spectra collection Near infrared (NIR) spectra were recorded using a Field Spec[®] spectrophotometer (ASD Inc., Boulder, CO) with wavelengths between 350 and 2500 nm at 1 nm intervals. The room temperature and humidity were kept constant. The spectrophotometer was pre-run for 30 min before each regular measurement. Three measurements of NIR spectra were taken per sample and averaged before statistical analysis. The contents of three wood chemical properties, including holo-cellulose, hemi-cellulose and lignin, were predicted using NIR-PLS models developed for *L. kaempferi* [51] and denote by HOLOCEL, HEMICEL and LIGNIN.

Pedigree reconstruction

Because we have previously reported the results of the pedigree reconstruction [44], the material, methods and results are briefly outlined here. A total of 17 simple sequence repeat (SSR) markers were used for genotyping 647 progenies from 63 open-pollinated families and 140 existing maternal clones. 62 families were shared between the previous and the present studies. The CER-VUS software (version 4.0) [52] was used to perform paternity analysis with a strict confidence level (CL) of 95% and a relaxed CL of 80% in 10,000 simulation cycles. 572 progenies were implicitly assigned to 97 paternal clones, of which 223 were at the strict CL and the rest at the relaxed CL, generating 433 full-sib families with a mean number of trees of 1.21 (SD=0.63) per full-sib family. The limited number of trees per full-sib family may bias genetic parameter estimates [53, 54]. Therefore, the reconstructed pedigree provides preliminary estimates of the genetic parameters of various traits.

SNP genotyping

DNA extraction was performed on fresh needles from the sample trees using a CTAB procedure modified from Doyle and Doyle [55]. To generate a high-density SNP profile for the 661 DNA extracts, we conducted a multiplexed, high-throughput Genotyping-by-Sequencing (GBS) following Elshire et al. [56]. A 48-plex GBS library containing 47 DNA samples and a negative control (no DNA) was prepared and each of the 47 DNA extracts was barcoded. Briefly, each DNA extract (500 ng) was digested with the restriction enzyme ApeKI for 2 h. Ligation products from each DNA extract were pooled and purified using the QIAquick PCR purification kit (Qiagen). The amplified 48-plex libraries were diluted and sequenced twice (single-end reads only) on the Illumina HiSeq 2000 to achieve the sequencing coverage equivalent to 24-plex. Raw DNA short-read sequences were analyzed with a pipeline, the Universal Network Enabled Analysis Kit (UNEAK), tailored to species lacking

reference genome information. To reduce sequencing errors in genotyping, we set the error tolerance rate to 0.03 (to exceed the expected Illumina sequencing error rate of 0.4%). The resulting SNP table was further filtered using the minimum value of inbreeding coefficient (mnF=0.05) and minimum minor allele frequency (mnMAF=0.05), and SNPs that are present in less than 40% of the samples were eliminated from further analysis. Finally, a total of 11,333 SNPs was used for genotyping the individual trees (see Additional file 1: Table S1). Accuracy validation of SNP calling was performed using conventional PCR and Sanger-based sequencing (see Additional file 1: Table S2) in 30 progenies. A total of eight fragments containing 69 SNPs were randomly selected and 54 SNPs (78.3%) were validated. The mean missingness of genotypic data was 17.7%. Missing data were imputed using random imputation method implemented in the R package synbreed [57] in the R environment (version 4.0.2) [58].

Statistical models

For each trait, three pedigree-based individual tree models (ABLUP-OP, ABLUP-FS-A and ABLUP-FS-AD) and four genomic-based individual tree models (GBLUP-A, GBLUP-AD, GBLUP-ADI, GBLUP-ADIE) were fitted.

For the pedigree-based models, ABLUP-AD was the full model, and the rest were reduced models. The form of ABLUP-AD was as follows:

$$y = X\beta + Z_p p + Z_a a + Z_d d + \epsilon \quad (3)$$

where y is the vector of phenotypic observations of a single trait; β is the vector of fixed effects, including a grand mean and block effects; p , a , d , and ϵ are the vectors of random plot, additive, dominance, and residual effects, respectively; X , Z_p , Z_a , Z_d and Z_e are the incidence matrices for p , a , d , and ϵ , respectively.

Assuming that the random effects in formula (3) follow a joint distribution as:

$$\begin{bmatrix} p \\ a \\ d \\ \epsilon \end{bmatrix} \sim N \left(\begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} I\sigma_p^2 & 0 & 0 & 0 \\ 0 & A_A\sigma_a^2 & 0 & 0 \\ 0 & 0 & A_D\sigma_d^2 & 0 \\ 0 & 0 & 0 & I\sigma_\epsilon^2 \end{bmatrix} \right) \quad (4)$$

where A_A and A_D are pedigree-based relationship matrices for additive and dominance effects, respectively; I is the identical matrix; σ_p^2 , σ_a^2 , σ_d^2 , and σ_ϵ^2 are variances of plot, additive, dominance, and residual effects, respectively.

The matrices A_A and A_D were constructed by the kin function in the R package synbreed [57].

For the genomic-based models, GBLUP-ADIE was the full model, and the rest were reduced models. The form of GBLUP-ADIE was as follows:

$$y = X\beta + Z_p p + Z_a a + Z_d d + Z_i i + Z_e e + \epsilon \quad (5)$$

where i and e are the vectors of random imprinting and additive-by-additive epistatic effects, respectively; Z_i and Z_e are the incidence matrices for i and e , respectively. The other terms are as defined in formula (3).

Assuming that the random effects in formula (5) follow a joint distribution as:

$$\begin{bmatrix} p \\ a \\ d \\ i \\ e \\ \epsilon \end{bmatrix} \sim N \left(\begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} I\sigma_p^2 & 0 & 0 & 0 & 0 & 0 \\ 0 & G_A\sigma_a^2 & 0 & 0 & 0 & 0 \\ 0 & 0 & G_D\sigma_d^2 & 0 & 0 & 0 \\ 0 & 0 & 0 & G_I\sigma_i^2 & 0 & 0 \\ 0 & 0 & 0 & 0 & G_E\sigma_e^2 & 0 \\ 0 & 0 & 0 & 0 & 0 & I\sigma_\epsilon^2 \end{bmatrix} \right) \quad (6)$$

where $G_A, G_D, G_I,$ and G_E are genomic-based relationship matrices for additive, dominance, imprinting, and additive-by-additive epistatic effects, respectively; σ_i^2 and σ_e^2 are variances of imprinting and additive-by-additive epistatic effects, respectively. The other terms are as defined in formula (4).

The genomic-based relationship matrices $G_A, G_D, G_I,$ and G_E were constructed from SNP data as follows [59, 60]:

$$G_A = \frac{M_A M_A'}{2 \sum p_j (1 - p_j)} \quad (7)$$

$$G_D = \frac{M_D M_D'}{\sum (2p_j (1 - p_j))^2} \quad (8)$$

$$G_I = \frac{M_I M_I'}{\sum (2p_j (1 - p_j))^2} \quad (9)$$

where M_A, M_D and M_I are $n \times p$ matrices, n and p are the number of individuals and SNPs, respectively. p_j is the observed minor allele frequency of the j th SNP. The elements of M_A, M_D and M_I for the i th individual at the j th SNP are calculated as follows:

$$M_{Aij} = \begin{cases} 0 - 2p_j & (A_1A_1) \\ 1 - 2p_j & (A_1A_2 \text{ and } A_2A_1) \\ 2 - 2p_j & (A_2A_2) \end{cases} \quad (10)$$

$$M_{Dij} = \begin{cases} -2p_j^2 (A_1A_1) & (A_1A_1) \\ 2p_j(1 - p_j) & (A_1A_2 \text{ and } A_2A_1) \\ -2(1 - p_j)^2 & (A_2A_2) \end{cases} \quad (11)$$

$$M_{Iij} = \begin{cases} 0 & (A_1A_1) \\ 1 & (A_1A_2) \\ -1 & (A_2A_1) \\ 0 & (A_2A_2) \end{cases} \quad (12)$$

The first-order epistatic relationship matrix G_{AA} was computed using the Hadamard product (cell by cell multiplication, denoted by #). The additive-by-additive term is $G_{AA} = G_A \# G_A$. These matrices were constructed using the sommer-R package (version 4.0) [58, 61].

For ABLUP and GBLUP model analyses, restricted maximum likelihood (REML) estimates of (co)variance components were obtained by using the average information (AI) algorithm, implemented in the sommer-R package (version 4.0) [61]. All (co)variance component estimates were positively constrained.

The coefficient of variation CV was calculated as follows:

$$CV = \sigma_p / \bar{X} \times 100\% \quad (13)$$

where σ_p and \bar{X} are the square root of the phenotypic variance and the grand mean, respectively.

The narrow and broad sense heritability were estimated as follows:

$$h^2 = \frac{\sigma_a^2}{\sigma_p^2} \quad (14)$$

$$H^2 = \frac{\sigma_g^2}{\sigma_p^2} \quad (15)$$

where σ_a^2 is the estimated additive genetic variance; σ_g^2 is the sum of all the genetic effect variances; σ_p^2 , the phenotypic variance, is the sum of all the variances of the random effects. Both σ_g^2 and σ_p^2 were different in different models. In model (5), for example, $\sigma_g^2 = \sigma_a^2 + \sigma_d^2 + \sigma_i^2 + \sigma_e^2$, $\sigma_p^2 = \sigma_p^2 + \sigma_g^2 + \sigma_\epsilon^2$. In the remaining models, the term for genetic effect variance was specifically omitted from the formular according to the form of the model.

The Akaike Information Criterion (AIC, [62]) and SchwarzBayesian Information Criterion (BIC, [63])

were used to compare the relative quality of the goodness-of-fit of the different models with a threshold of 2. $AIC = -2\log L + 2\rho$, $BIC = -2\log L + 2\log(n)\rho$, where $\log L$ is the REML log-likelihood, ρ is the number of estimated parameters, n is the number of observations. A smaller AIC or BIC value indicates a better quality of fit.

Based on the estimates of the (co)variance components and the information criteria values, we omitted the genetic terms that provide a negligible contribution to the phenotypic variance and re-fitted the GBLUP models and RKHS models, i.e. the optimal model for each trait.

RKHS models assume that the random effects in formula (5) follow a joint distribution as:

$$\begin{bmatrix} p \\ a \\ d \\ i \\ e \\ \epsilon \end{bmatrix} \sim N \left(\begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} I\sigma_p^2 & 0 & 0 & 0 & 0 & 0 \\ 0 & K_A\sigma_a^2 & 0 & 0 & 0 & 0 \\ 0 & 0 & K_D\sigma_d^2 & 0 & 0 & 0 \\ 0 & 0 & 0 & K_I\sigma_i^2 & 0 & 0 \\ 0 & 0 & 0 & 0 & K_E\sigma_e^2 & 0 \\ 0 & 0 & 0 & 0 & 0 & I\sigma_\epsilon^2 \end{bmatrix} \right) \tag{16}$$

each of K_A , K_D , K_I , and K_E depends on a reproducing kernel function with a smoothing semi-parameter h , which controls how quickly the prior covariance function declines with increasing genomic distance between genotypes and can be interpreted as a correlation matrix [64]. The other terms are as defined in formula (6). The RKHS models were implemented using the BGLR function from the BGLR package in R [65]. The Gibbs chain length was 20,000 iterations with the first 2000 iterations discarded as burn-in and a thinning interval set to 100.

A 10-fold cross-validation scheme with 10 replicates was used to assess accuracy and predictive ability (PA). The PA of the models was evaluated as the Pearson correlation coefficient between the predicted breeding values of the validation trees and the phenotypes adjusted for block effects and was compared using GBLUP-A as a benchmark.

Results

Phenotypic variation

Descriptive statistics for each trait are presented in Table 1. The CV of wood chemical traits were generally lower than those of growth and wood physical traits. Estimates of CV for HOLOCEL and LIGNIN were 4.14% and 7.44%. HEMICEL had the highest CV among the wood chemical attributes (11.12%). In wood physical properties, CV for PILODYN was 11.42%, higher than that for AV (CV=9.24%). Estimates of CV for H, DBH and V were 11.37%, 14.55% and 34.97%, respectively. The CV of CI was the highest of all traits, at 44%.

Table 1 Descriptive statistics for the analyzed traits

| Traits | Mean | SD | Min | Max | CV/% |
|-------------------|--------|-------|--------|--------|-------|
| H/m | 17.72 | 2.02 | 10.50 | 24.00 | 11.37 |
| DBH/cm | 16.66 | 2.42 | 9.10 | 31.10 | 14.55 |
| V/dm ³ | 195.68 | 68.43 | 43.16 | 621.68 | 34.97 |
| CI | 1.08 | 0.47 | 0.22 | 3.23 | 43.49 |
| PILODYN | 20.20 | 2.31 | 13.00 | 26.00 | 11.42 |
| AV | 232.67 | 21.50 | 186.00 | 399.00 | 9.24 |
| HOLOCEL/% | 70.81 | 2.93 | 63.53 | 82.04 | 4.14 |
| HEMICEL/% | 29.94 | 3.33 | 17.79 | 39.42 | 11.12 |
| LIGNIN/% | 27.35 | 2.04 | 21.16 | 32.98 | 7.44 |

Note: H Tree height, DBH Diameter at breast height, V Stem volume, CI Competition index, PILODYN Pilodyn penetration, AV Acoustic velocity, HOLOCEL Holo-cellulose, HEMICEL Hemi-cellulose, LIGNIN Lignin, Max Maximum value, Min Minimum value, SD Standard deviation, CV Coefficient of variation

Model comparison

AIC and BIC values obtained from the 7 models are presented in Table 2. The AIC values were generally in agreement with the BIC values of the same model. The comparison of ABLUP-OP and ABLUP-FS showed that ABLUP-OP outperformed ABLUP-FS for growth and wood chemical properties except for DBH. ABLUP-OP and ABLUP-FS showed no differences for CI and wood physical properties. The comparison of ABLUP and GBLUP showed that ABLUP outperformed GBLUP for wood chemical properties, while GBLUP performed better than ABLUP for CI and wood physical properties; the patterns for growth were irregular. The comparison of GBLUP showed that GBLUP-A and other models showed no differences for wood chemical properties; GBLUP-ADIE was generally the best for growth, CI and AV. These results showed that (1) the performance of ABLUP was generally better than GBLUP-A for wood chemical traits, (2) epistasis played a more important role in model fitting for growth, CI and AV, and (3) dominance and imprinting contributed slightly to model fitting for almost all traits.

Genetic variance components

The proportions of variance components (additive σ_a^2 , dominance σ_d^2 , imprinting σ_i^2 , epistatic σ_e^2 , plot σ_p^2 and residual effects σ_ϵ^2) obtained from the 7 models are presented in Fig. 1.

The comparison of ABLUP-OP and ABLUP-FS showed that the proportions of σ_a^2 estimated from ABLUP-OP (7.88-100%) were larger than those from ABLUP-FS (0-81.41%) for all traits, indicating that they were estimated upwards from ABLUP-OP; σ_d^2 was observed in ABLUP-FS-AD (8.48-35.61%) for growth and AV, as well as in GBLUP-AD (6.48-34.30%). The comparison

Table 2 AIC and BIC values of the 7 models (3 ABLUP and 4 GBLUP models) for the analyzed traits

| Traits | Information criterion | ABLUP-OP | ABLUP-FS-A | ABLUP-FS-AD | GBLUP-A | GBLUP-AD | GBLUP-ADI | GBLUP-ADIE |
|---------|-----------------------|----------|------------|-------------|---------|----------|-----------|------------|
| H | AIC | 380.95 | 383.44 | 383.38 | 380.22 | 380.15 | 380.15 | 377.81 |
| H | BIC | 434.85 | 437.35 | 437.29 | 434.13 | 434.06 | 434.06 | 431.72 |
| DBH | AIC | 540.30 | 542.43 | 541.89 | 545.82 | 544.99 | 544.99 | 543.83 |
| DBH | BIC | 594.20 | 596.34 | 595.80 | 599.72 | 598.90 | 598.90 | 597.73 |
| V | AIC | 507.72 | 510.19 | 510.17 | 510.42 | 509.18 | 509.18 | 507.23 |
| V | BIC | 501.79 | 564.09 | 564.07 | 501.09 | 500.12 | 500.12 | 492.84 |
| CI | AIC | 587.40 | 587.98 | 587.93 | 583.46 | 583.46 | 580.73 | 578.08 |
| CI | BIC | 640.94 | 641.52 | 641.47 | 637.00 | 637.00 | 634.27 | 631.62 |
| PILODYN | AIC | 529.82 | 528.30 | 528.30 | 514.51 | 514.42 | 510.97 | NA |
| PILODYN | BIC | 559.86 | 558.34 | 558.34 | 544.55 | 544.46 | 541.01 | NA |
| AV | AIC | 522.91 | 524.02 | 523.30 | 518.65 | 516.46 | 516.46 | 502.68 |
| AV | BIC | 576.78 | 577.89 | 577.17 | 572.52 | 570.33 | 570.33 | 556.55 |
| HOLOCEL | AIC | 366.71 | 440.06 | 438.85 | 463.51 | 463.51 | 462.90 | 462.90 |
| HOLOCEL | BIC | 414.26 | 487.60 | 486.40 | 511.06 | 511.06 | 510.44 | 510.44 |
| HEMICEL | AIC | 390.52 | 459.08 | NA | 469.59 | 469.59 | 469.59 | 469.59 |
| HEMICEL | BIC | 438.07 | 506.62 | NA | 517.14 | 517.14 | 517.14 | 517.14 |
| LIGNIN | AIC | 363.73 | 423.96 | 423.96 | 462.94 | 462.56 | 462.56 | NA |
| LIGNIN | BIC | 411.27 | 471.51 | 471.51 | 510.49 | 510.10 | 510.10 | NA |

Note: See Table 1 for full description of traits. AIC: Akaike Information Criterion; BIC: SchwarzBayesian information criterion; ABLUP-OP: the OP pedigree-based individual tree model (using the additive relationship matrix A_A estimated from the OP pedigree with known maternity only); ABLUP-FS-A: the FS pedigree-based individual tree model (using the additive relationship matrix A_A matrix estimated from the FS pedigree reconstructed using paternity assignment analysis); ABLUP-FS-AD: the FS pedigree-based individual tree model (using the additive and dominance matrices A_A and A_D estimated from the FS pedigree reconstructed using paternity assignment analysis); GBLUP-A: the genomic selection model (using the realized additive genomic relationship matrix G_A estimated from SNPs); GBLUP-AD: the genomic selection model (using the realized additive and dominance genomic relationship matrices G_A and G_D estimated from SNPs); GBLUP-ADI: the genomic selection model (using the realized additive, dominance and imprinting genomic relationship matrices G_A , G_D and G_I estimated from SNPs); GBLUP-ADIE: the genomic selection model (using the realized additive, dominance, imprinting and additive-by-additive epistatic genomic relationship matrices G_A , G_D , G_I and G_E estimated from SNPs); NA: not available due to failure of the model to reach convergence

of ABLUP-OP and GBLUP-A showed that the proportions of σ_a^2 estimated from ABLUP-OP (95.48-100%) were larger than those from GBLUP-A (38.20-67.46%) for wood chemical properties; in contrast, for CI and wood physical properties, σ_a^2 contributed more of the variance from GBLUP-A (38.17-54.11%) than ABLUP-OP (7.88-30.67%).

The comparison of GBLUP showed that additive effect is the main source of variation for wood chemical properties (38.20-67.46%). Except for growth and AV, σ_a^2 had contribution to CI and HOLOCEL in ABLUP-FS-AD and to PILODYN and LIGNIN in GBLUP-AD. Dominance variances were generally separated from the residual variances in ABLUP-FS-AD. However, in GBLUP-AD, the dominance variances were more from additive genetic variances and a small part of residual variance: up to half of σ_a^2 was reduced for AV, and almost all of σ_a^2 was reduced to near zero for DBH, V and PILODYN. These results indicated that σ_a^2 estimated from GBLUP-A may confounded with the σ_d^2 , and that including dominance effect could improve the explanation of residual variance for those traits that are considerable genetically

controlled by dominance effect. The imprinting effects had contributions to CI, PILODYN and HOLOCEL, with proportions estimated by GBLUP-ADI of 17.33%, 19.71% and 7.08%, respectively. The imprinting effect variances were dissected from the residual variances. For growth, CI and AV, when epistasis was considered, a surprisingly significant proportion of the variance was explained by this effect (50.76-91.26%), which was decomposed from other sources of variation, particularly the residual variance. These results were generally consistent with the AIC values. Therefore, the traits could be classified into two types: for type I, including wood chemical properties and PILODYN, additive effect is the main source of variation; for type II, including growth, CI and AV, epistasis plays a significant role.

Heritability estimates

Narrow and broad sense heritability (h^2 and H^2) exhibited similar trends as additive and total genetic variance respectively from different models and are shown in Fig. 2. h^2 estimates from ABLUP-FS were lower than those from ABLUP-OP for all traits. For growth traits,

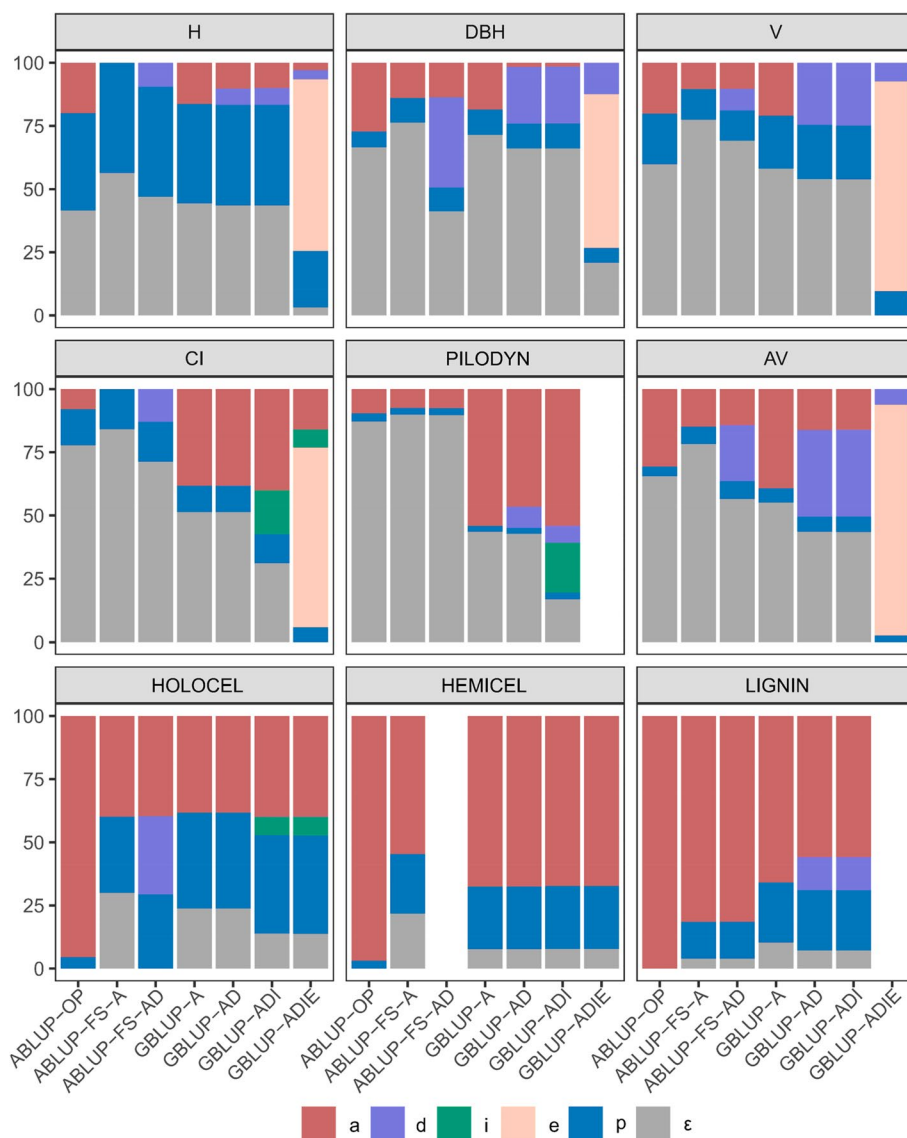


Fig. 1 The proportion of variance for analyzed traits from the 7 genetic models fitted. See Table 1 for full description of traits. See Table 2 for full description of models. *a, d, i, e, p, and ε* are the variance proportion of additive, dominance, imprinting, additive-by-additive epistatic effects, plot, and residual effects, respectively

h^2 estimates from ABLUP-OP indicated that these traits were moderately controlled by additive effects, ranging from 0.20 to 0.27, which were slightly higher than those from GBLUP-A (0.16–0.20). For wood physical properties and CI, h^2 estimates from GBLUP-A, 0.38–0.54, were much higher than those from ABLUP-OP (0.08–0.31).

As the number of genetic terms increased, the σ_a^2 were partitioned into other components, thus h^2 were gradually reduced to zero for traits such as DBH. On the other hand, the H^2 were increased from 0.17 to around 0.80, due to the inflating non-additive components, which are mainly epistatic effects.

The wood chemical properties showed the highest h^2 estimates in this study. h^2 estimates from ABLUP-OP were almost one seemingly unrealistic. The h^2 and H^2 estimates did not show much fluctuation when more genetic effects were included in the GBLUP models, ranging from 0.38 (HOLOCEL) to 0.67 (HEMICEL). The h^2 and H^2 estimates from ABLUP-OP and ABLUP-FS had higher standard errors than those from GBLUP-A in most cases, indicating that more accurate h^2 and H^2 estimates could be obtained using the GBLUP method.

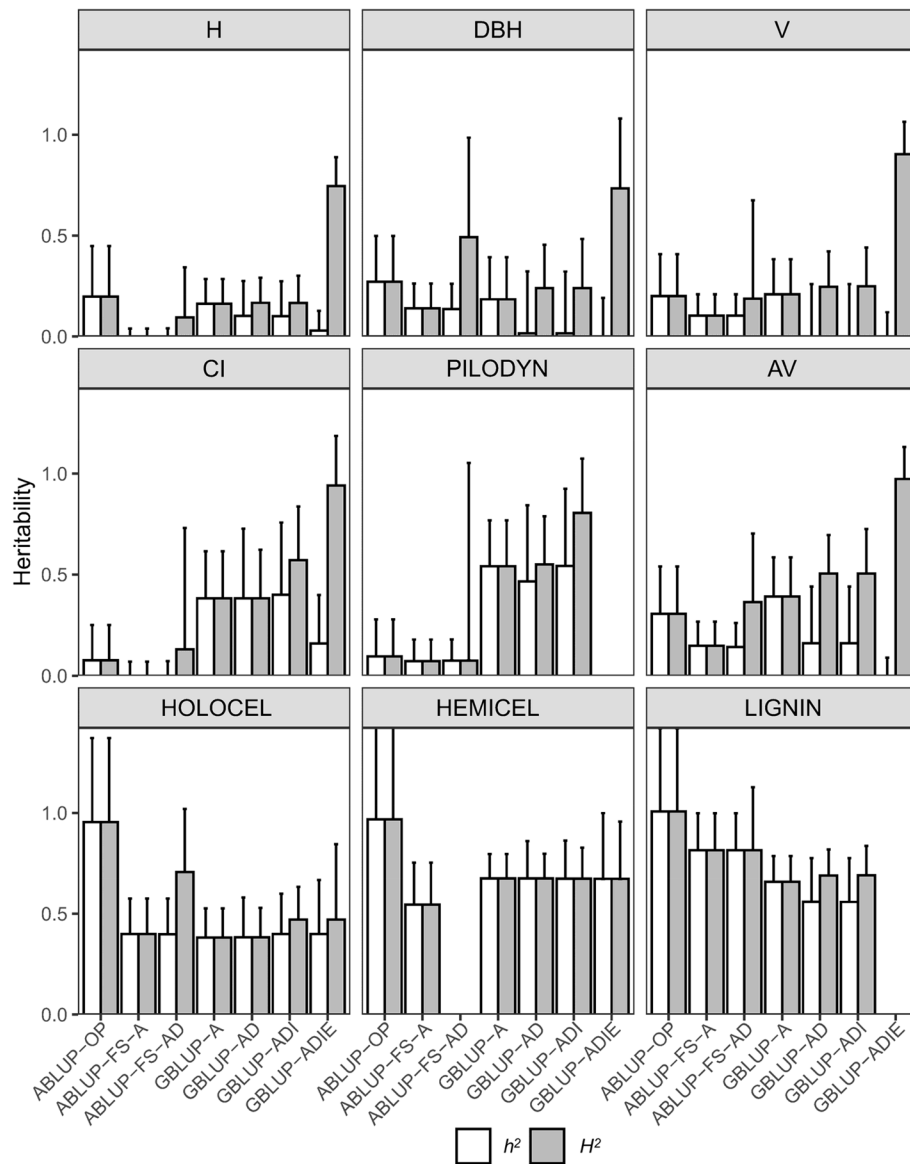


Fig. 2 The narrow and broad sense heritability for each trait from the 7 genetic models fitted. h^2 and H^2 are the narrow and broad sense heritability, respectively. Error bars above each bar indicate standard errors (SE). See Table 1 for full description of traits. See Table 2 for full description of models

Predictive ability

The PAs from the three models for each trait are presented in Fig. 3 using box plots combining the results of the hypothesis tests.

No significant difference was found between the PAs of GBLUP-A and the corresponding GBLUP models considering non-additive effects (GS-GBLUP) for all the traits except for DBH and HOLOCEL. For growth-related traits, the PAs of GS-RKHS were significantly higher than those of GBLUP-A and GS-GBLUP ($p < 0.01$). The PAs of GS-RKHS for H, DBH, and V were ~0.29, 89.30-148.26% higher than those of GS-GBLUP, which were 0.15, 0.10,

and 0.15, respectively. For CI, the PA of GS-RKHS was 0.35, 149.08 times higher than that of GS-GBLUP. For PILODYN, a wood physical trait, the PA of GS-RKHS was 0.05, significantly lower than 0.18 of GS-GBLUP and 0.22 of GBLUP-A. For AV, the pattern of PAs of the GS models was similar to that of the growth-related traits. The PA of GS-RKHS was 0.27, significantly higher than the 0.13 of GS-GBLUP. For wood chemical traits, GS-GBLUP and GS-RKHS were not successfully fitted for HEMICEL. The PAs of GS-RKHS were 0.20 and 0.26 for HOLOCEL and LIGNIN respectively, significantly lower than those of GS-GBLUP and GBLUP-A (0.30 and 0.38).

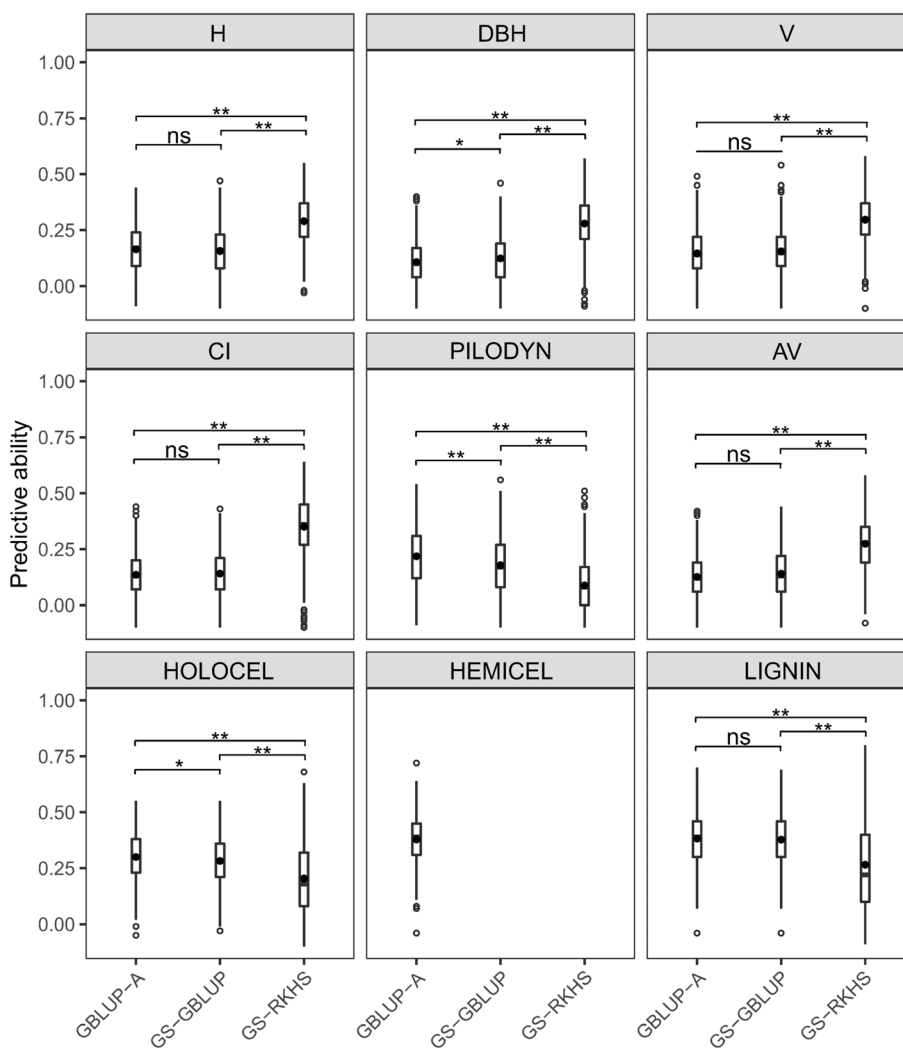


Fig. 3 Distribution of predictive abilities for each trait. Each panel contains box plots showing the distribution of predictive abilities from the base models (GBLUP-A), and the optimal models (without genetic terms that provide a negligible contribution to the phenotypic variance) considering non-additive effects fitted by the GBLUP (GS-GBLUP) and RKHS (GS-RKHS) approaches. ns represents non-significance; * and ** indicate significant at $p=0.05$ and $p=0.01$. Hollow circles represent outliers. The black solid points and bars represent means and medians. See Table 1 for full description of traits

There was no significant difference between the last two models. The PA of GBLUP-A for HEMICEL was 0.38.

Discussion

Unraveling the genetic control of economically and adaptively important traits in forest tree species will help meet the growing global demand for high quality wood and the increasing challenges posed by changing climates and environments. OP progeny testing is the most efficient and widely used approach to genetic analysis and screening of large numbers of individuals in terms of low cost and less time. However, under this family structure, only the additive genetic variance

component can be obtained, the estimate of which is somewhat biased due to the common violation of the true half-sib assumption [45, 66]. The considerable available DNA-marker information and GS models provided us with an unprecedented opportunity to understand the genetic control of key traits of Japanese larch, which is extremely important for decision making of breeding strategy and facilitating a better exploitation of the available inherent variation for these traits in breeding programmes. This study is the first comprehensive assessment of contribution of the additive and non-additive effects on growth, wood and competitive traits of Japanese larch using a realized-genetic based

model that has been proven effective in animal and plant breeding (e.g. [67, 68]).

Genetic control of growth, wood and competitive traits

For forest tree species used for pulp production, much attention has been paid by breeders to wood chemical properties such as cellulose and lignin content, which can be readily predicted by NIR-based calibrations. In this study, cellulose and lignin content and Pilodyn penetration are mainly controlled by additive effects in both ABLUP-OP ($h^2=0.95-1.00$) and GBLUP-A ($h^2=0.38-0.67$), which have been reported in various forest tree species in both conventional and GS studies [36, 69–73]. Therefore, it is promising to achieve significant genetic gain through GS-based recurrent selection. Meanwhile, the minimal contribution of non-additive effects on these traits also demonstrated the efficacy of the previous studies that did not consider non-additive effects. Additionally, the moderate to high narrow-sense heritability indicates that the discovery of the genes underlying the biosynthesis of these traits is prospective using genome-wide association mapping studies (GWAS). In Japanese larch, 77 SNPs were significantly associated with cellulose and lignin content and were located in 54 genes (unpublished data). In *Populus tomentosa* Carrière, Li et al. [74] reported that selection of specific SNPs in functional genes could regulate the cellulose and lignin content. In *Picea abies* L. Karst, Chen et al. [75] detected four SNPs associated with Pilodyn penetration. Alternatively, the genetic gain could be increased by manipulating the genetic elements involved using molecular biology approaches.

For growth, AV and CI, additive-by-additive epistatic effects play more important role (50.76–91.26%, Fig. 1). For AV, the measurement of acoustic velocity, which represents the mechanical stiffness of wood, was moderately controlled by the additive effect from the model ABLUP-OP ($h^2=0.31$) and GBLUP-A ($h^2=0.39$) considering only the additive effect, which is consistent with other tree species reports [37, 76, 77]. However, when the epistatic effect was accounted for, almost all the phenotypic variation was explained by the epistatic and dominance effects, 91.26% and 6.08% respectively. No similar results have not been found. Epistasis showed no effect on wood stiffness in Scots pine (*Pinus sylvestris* L.) with 695 progeny trees from 184 full-sib families [9] and in control-pollinated Norway spruce (*Picea abies* (L.) Karst.) with 1370 progeny trees from 128 full-sib families [78]. Further research is required to identify the specific reasons for this result in Japanese larch. The significant contribution of epistatic effects was also found for growth and competitive traits and has also been reported in other tree species. Tree height was controlled by a significant

additive-by-additive epistatic effect in an open-pollinated white spruce (*Picea glauca* (Moench) Voss) trial of 214 families [79]. In contrast, the epistatic effect could be ignored for tree height Norway spruce trials [78] and in an open-pollinated interior spruce (*Picea glauca* x *engelmannii*) test of 25 families [19]. Furthermore, an opposite contribution by epistatic effect was shown for wood density in spruce, Scots pine and hybrid *Eucalyptus* [9, 23, 79]. These inconsistent results may hinder their implementation in breeding programs.

The dominance variance was mainly absorbed by the additive variance and a small amount by the residual variance. The imprinting variance was fully extracted from the residual variance. Furthermore, the proportions of these two effects were reduced when epistasis was taken into account due to its strong capacity to absorb variance from other components [19, 80]. The dominance effect has a contribution for AV and LIGNIN (6.08% and 13.02% respectively). Dominance contributed 9.8–18.1% of phenotypic variation for tree height in Norway spruce [78] and 16.4% and 5.7% of phenotypic variation for LIGNIN and cellulose content in loblolly pine [81]. In contrast, Lenz et al. [37] reported that the dominance effect was not significant for all traits including weevil resistance, growth, and wood quality traits in Norway spruce and did not improve the model fit.

As in traditional pedigree-based genetic analysis, the population-specific genetic analysis still shown inconsistent contributions of additive and non-additive effects for the same traits in most cases in the GS period [23, 78, 82]. For those traits that are primarily controlled by non-additive effects, short-term genetic gains could be captured through vegetative propagation of individuals with high genetic values by rooted cuttings or grafts. Although somatic embryogenesis is a more efficient method for conifers, its practical application in Japanese larch is not yet well developed. For long-term genetic gain, individuals with additional genetic variation and superior characteristics should be introduced into this breeding population. Alternatively, the systematic introduction of germplasm from other larch species into the current breeding population through hybridization is a promising approach.

RKHS with epistatic genetic effects improves predictive ability

Maximizing the predictive ability (PA) of predictive models is one of the main goals in GS, and PA has been influenced by several factors. Recent studies have shown that the PA can be improved when additive and non-additive effects are considered simultaneously in a predictive model [17–20]. In the present study, the influence of statistical model and non-additive effects on the PAs

of predictive models was evaluated. The results showed that using the GBLUP approach, non-additive effects have no significant contribution to the improvement of PA for all the traits (Fig. 3), suggesting that GBLUP as a linear model may not capture complex patterns like additive-by-additive epistatic effects. In contrast, the PAs from RKHS models were significantly higher than those from GBLUP models considering additive effects only (GBLUP-A), and accounting for additive and non-additive effects simultaneously (GS-GBLUP) for those traits controlled by epistatic effects. These results showed that the main advantage of the RKHS is on its superiority in capturing epistatic effects [64, 83]. However, for traits not significantly controlled by epistatic effects but by other non-additive components, holo-cellulose and lignin content, the PAs from RKHS models were remarkably lower than those from GBLUP models, which could be attributed to the inappropriate definition of a kernel [64]. Similarly, Tan et al. [84] reported that RKHS was the worst performing method for pulp yield in two *Eucalyptus* species and their F_1 hybrids. Therefore, for these types of traits, GBLUP models are more suitable for predicting GEBVs, in addition to their speed of fast computation. Most of the previous empirical studies showed that there was a slight difference in prediction performance for quantitative traits between GBLUP and RKHS models [32, 84–86], which could be caused by the negligible contribution of epistasis on the traits analyzed or by not accounting for epistasis in the models.

Some issues affecting the efficiency of genomic prediction

Sample size is primarily considered in genetic research for its critical role in statistical power [87, 88]. The sample size in this study, consisting of 661 trees from 66 OP families (~10 individuals/family), can be considered relatively small for generating highly reliable estimates of genetic effects and genomic predictions, particularly for traits with low narrow-sense heritability (Figs. 2 and 3). In addition, the sample size is not sufficient to obtain an optimal number of individuals per full-sib family after pedigree reconstruction to precisely estimate genetic parameters for low heritability traits such as height (Fig. 2). Therefore, the reconstructed pedigree provides preliminary estimates of the genetic parameters of various traits. It is expected that the statistical power would further improve with a larger sample size [37]. However, we would like to emphasize that practical constraints, such as limited financial resources, often impose limitations on sample sizes in genetic research involving molecular genotyping. Achieving a balance between sample size and available budget is a challenge faced by many breeders [87].

Accuracy of phenotypes has significant impact on the efficiency of genomic prediction. Various non-destructive evaluation methods have been used to assess wood properties in standing trees [89]. We used Pilodyn penetration, acoustic velocity and spectroscopy techniques to proxy the wood physical and chemical properties. The Pilodyn penetration may not be a good surrogate for wood density in tree species with very high variation between earlywood and latewood densities within a growth ring, such as radiata pine (*Pinus radiata* D. Don) and *Larix kaempferi*, as the pin penetration may be affected by the number of latewood bands encountered [90]. There was also a relatively large measurement error in Pilodyn penetration, which could be caused by differences in bark thickness between trees of *Larix kaempferi*. In contrast, the Resistograph can overcome some of the limitations of Pilodyn and can serve as a more efficient alternative tool for evaluating wood density in standing trees [91] and will be used in our future investigations in *Larix kaempferi* to collect more accurate data.

Conclusions

Genomic-based models (GBLUP) provide a more comprehensive and accurate estimate of the contribution of the additive and non-additive effects than pedigree-based models (ABLUP). In the GBLUP model with all non-additive effects, the traits considered in this study could be divided into two types based on different quantitative genetic architectures: type I, additive controlling traits, including wood chemical traits and Pilodyn penetration; type II, epistatic controlling traits, including growth traits, competitive ability and acoustic velocity. Dominance and imprinting showed low contributions to the phenotypic variance of the traits. The GBLUP and RKHS methods were preferred in terms of predictive ability for type I and type II traits, respectively.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12864-023-09891-4>.

Additional file 1: Table S1. Genotyping data of 11,333 SNPs. **Table S2.** Validation of SNP calling accuracy by Sanger sequencing

Acknowledgements

We would like to thank Dr. Weibo Xiang, Dr. Xingbin Chen and Dr. Chao Sun for their excellent technical assistance in preparing the genotype data. We sincerely acknowledge the editor and two anonymous referees for their constructive criticism and helpful comments, which have greatly improved the manuscript.

Authors' contributions

L.D. and X.S. designed the study and wrote the manuscript. L.D. and Y.X. collected the phenotypic measurements. L.D. and R.W. performed the statistical

analysis. Y.Z. prepared all figures. All authors contributed to the article and approved the final manuscript.

Funding

This study was supported by the National Key R&D Program of China (2022YFD2200302), the Beijing Natural Science Foundation (6224061) and the Special Funds of Research Institute of Forestry (LYSZX202002).

Availability of data and materials

The raw DNA short-read sequences generated in this paper are not publicly available due to the large file size (~8 TB) but are available from the corresponding author on reasonable request. The original genotype and phenotype data and results were included in the article and its additional files.

Declarations

Ethics approval and consent to participate

All field studies and experimental research conducted on plants, including the procurement of plant materials, were carried out in strict accordance with institutional, national, and international guidelines and regulations. The corresponding author, Xiaomei Sun, is the head of the Conifer Genetic Improvement Team at the Institute of Forestry, Chinese Academy of Forestry, and possesses the necessary authorization to gather plant materials.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Received: 14 June 2023 Accepted: 11 December 2023

Published online: 02 January 2024

References

- Falconer DS, Mackay TFC. Introduction to Quantitative Genetics. 4th ed. Harlow: Longman Group; 1996.
- Foster GS, Shaw DV. Using clonal replicates to explore genetic variation in a perennial plant species. *Theor Appl Genet*. 1988;76:788–94.
- Lambeth C, Lee B-C, O'Malley D, Wheeler N. Polymix breeding with parental analysis of progeny: an alternative to full-sib breeding and testing. *Theor Appl Genet*. 2001;103:930–43.
- Hayes BJ, Visscher PM, Goddard ME. Increased accuracy of artificial selection by using the realized relationship matrix. *Genet Res*. 2009;91:47–60.
- Nishio M, Satoh M. Including dominance effects in the genomic BLUP method for genomic evaluation. *PLoS ONE*. 2014;9: e85792.
- Lopes MS, Bastiaansen JWM, Janss L, Knol EF, Bovenhuis H. Estimation of additive, dominance, and imprinting genetic variance using genomic data. *G3 (Bethesda, Md)*. 2015;5:2629–37.
- Guo X, Christensen OF, Ostensen T, Wang Y, Lund Mogens Sandø and Su G. Genomic prediction using models with dominance and imprinting effects for backfat thickness and average daily gain in Danish duroc pigs. *Genet Selection Evol*. 2016;48:67.
- Morley KI, Martin NGA, Visscher PM, Medland SE, Ferreira MAR, Morley KI, et al. Assumption-free estimation of heritability from genome-wide identity-by-descent sharing between full siblings. *PLoS Genet*. 2006;2:e41.
- Calleja-Rodriguez A, Chen Z, Suontama M, Pan J, Wu HX. Genomic predictions with Nonadditive effects Improved estimates of Additive effects and Predictions of Total Genetic Values in *Pinus sylvestris*. *Front. Plant Sci*. 2021;12:666820.
- Beaulieu J, Lenz P, Bousquet J. Metadata analysis indicates biased estimation of genetic parameters and gains using conventional pedigree information instead of genomic-based approaches in tree breeding. *Sci Rep*. 2022;12:3933.
- Meuwissen THE, Hayes BJ, Goddard ME. Prediction of total genetic value using genome-wide dense marker maps. *Genetics*. 2001;157:1819–29.
- Hayes BJ, Bowman PJ, Chamberlain AJ, Goddard ME. Invited review: genomic selection in dairy cattle: progress and challenges. *J Dairy Sci*. 2009;92:433–43.
- Goddard ME, Hayes BJ. Genomic selection. *J Anim Breed Genet*. 2007;124:323–30.
- Jannink J-L, Lorenz AJ, Iwata H. Genomic selection in plant breeding: from theory to practice. *Brief Funct Genomics*. 2010;9:166–77.
- Grattapaglia D. Breeding forest trees by genomic selection: current progress and the Way Forward. In: *Genomics of plant genetic resources*. Dordrecht: Springer Netherlands; 2014. p. 651–82.
- Isik F. Genomic selection in forest tree breeding: the concept and an outlook to the future. *New For*. 2014;45:379–401.
- Muñoz PR, Resende MFR, Gezan SA, Resende MDV, Campos G, de los, Kirst M, et al. Unraveling additive from nonadditive effects using genomic relationship matrices. *Genetics*. 2014;198:1759–68.
- de Almeida Filho JE, Guimarães JFR, Fonsceca e Silva F, Vilela de Resende MD, Muñoz P, Kirst M, et al. Genomic prediction of additive and non-additive effects using genetic markers and pedigrees. *G3 (Bethesda Md)*. 2019;9:2739–48.
- Gamal El-Dien O, Ratcliffe B, Klápště J, Porth I, Chen C, El-Kassaby YA. Multi-environment genomic variance decomposition analysis of open-pollinated Interior spruce (*Picea glauca* × *Engelmannii*). *Mol Breeding*. 2018;38:26.
- Varona L, Legarra A, Toro MA, Vitezica ZG. Non-additive effects in genomic selection. *Front Genet*. 2018;9:1–12.
- Dias KODG, Gezan SA, Guimarães CT, Nazarian A, Da Costa E, Silva L, Parentoni SN, et al. Improving accuracies of genomic predictions for drought tolerance in maize by joint modeling of additive and dominance effects in multi-environment trials. *Heredity*. 2018;121:24–37.
- Hunt CH, Hayes BJ, van Eeuwijk FA, Mace ES, Jordan DR. Multi-environment analysis of sorghum breeding trials using additive and dominance genomic relationships. *Theor Appl Genet*. 2020;133:1009–18.
- Tan B, Grattapaglia D, Wu HX, Ingvarsson PK. Genomic relationships reveal significant dominance effects for growth in hybrid *Eucalyptus*. *Plant Sci*. 2018;267:84–93.
- Jiang Y, Schmidt RH, Zhao Y, Reif JC. A quantitative genetic framework highlights the role of epistatic effects for grain-yield heterosis in bread wheat. *Nat Genet*. 2017;49:1741–6.
- Raffo MA, Sarup P, Guo X, Liu H, Andersen JR, Orabi J, et al. Improvement of genomic prediction in advanced wheat breeding lines by including additive-by-additive epistasis. *Theor Appl Genet*. 2022;135:965–78.
- Lorenzana RE, Bernardo R. Accuracy of genotypic value predictions for marker-based selection in biparental plant populations. *Theor Appl Genet*. 2009;120:151–61.
- Jiang J, Shen B, O'Connell JR, VanRaden PM, Cole JB, Ma L. Dissection of additive, dominance, and imprinting effects for production and reproduction traits in Holstein cattle. *BMC Genomics*. 2017;18:425.
- Bai F, Settles AM. Imprinting in plants as a mechanism to generate seed phenotypic diversity. *Front Plant Sci*. 2015;5:1–10.
- Reik W, Walter J. Genomic imprinting: parental influence on the genome. *Nat Rev Genet*. 2001;2:21–32.
- Gianola D. Priors in whole-genome regression: the bayesian alphabet returns. *Genetics*. 2013;194:573–96.
- Campos G, de los, Hickey JM, Pong-Wong R, Daetwyler HD, Calus MPL. Whole-genome regression and prediction methods applied to plant and animal breeding. *Genetics*. 2013;193:327–45.
- Chen Z-Q, Baisson J, Pan J, Karlsson B, Andersson B, Westin J, et al. Accuracy of genomic selection for growth and wood quality traits in two control-pollinated progeny trials using exome capture as the genotyping platform in Norway spruce. *BMC Genomics*. 2018;19:946.
- Ratcliffe B, El-Dien OG, Klápště J, Porth I, Chen C, Jaquish B, et al. A comparison of genomic selection models across time in Interior spruce (*Picea engelmannii* × *glauca*) using unordered SNP imputation methods. *Heredity*. 2015;115:547–55.
- Thistlethwaite FR, Ratcliffe B, Klápště J, Porth I, Chen C, Stoehr MU, et al. Genomic prediction accuracies in space and time for height and wood density of Douglas-fir using exome capture as the genotyping platform. *BMC Genomics*. 2017;18:930.
- Beaulieu J, Doerksen T, Clément S, MacKay J, Bousquet J. Accuracy of genomic selection models in a large population of open-pollinated families in white spruce. *Heredity*. 2014;113:343–52.

36. Isik F, Bartholomé J, Farjat A, Chancerel E, Raffin A, Sanchez L, et al. Genomic selection in maritime pine. *Plant Sci.* 2016;242:108–19.
37. Lenz PRN, Nadeau S, Mottet M, Perron M, Isabel N, Beaulieu J, et al. Multi-trait genomic selection for weevil resistance, growth, and wood quality in Norway spruce. *Evol Appl.* 2020;13:76–94.
38. Gianola D, Fernando RL, Stella A. Genomic-assisted prediction of genetic value with semiparametric procedures. *Genetics.* 2006;173:1761–76.
39. Jiang Y, Reif JC. Modeling epistasis in genomic selection. *Genetics.* 2015;201:759–68.
40. Bandeira e Sousa M, Cuevas J, de Oliveira Couto EG, Pérez-Rodríguez P, Jarquín D, Fritsche-Neto R, et al. Genomic-enabled prediction in Maize Using Kernel Models with genotype \times environment Interaction. *G3 (Bethesda Md).* 2017;7:1995–2014.
41. Olatoye MO, Hu Z, Aikpokpodion PO. Epistasis detection and modeling for genomic selection in Cowpea (*Vigna unguiculata* L. Walp). *Front Genet.* 2019;10:1–14.
42. Dong L, Xie Y, Sun X. Full-diallel-based analysis of genetic parameters for growth traits in Japanese larch (*Larix kaempferi*). *New For.* 2020;51:261–71.
43. Dong L, Xie Y, Wu HX, Sun X. Spatial and competition models increase the progeny testing efficiency of Japanese larch. *Can J for Res.* 2020;50:1373–82.
44. Chen X, Sun X, Dong L, Zhang S. Mating patterns and pollen dispersal in a Japanese larch (*Larix kaempferi*) clonal seed orchard: a case study. *Sci China Life Sci.* 2018;61:1011–23.
45. Dong LM, Zhang SG, Sun XM. Impact of pedigree reconstruction on estimates of genetic parameters for growth traits and wood properties of open-pollinated progeny of Japanese Larch. *For Res.* 2018;31:27–35.
46. Hegyi F. A simulation model for managing jack-pine stands. In: Fries J, editor. *Growth models for tree and stand simulation.* Royal College of Forestry: Royal College of Forestry; 1974. p. 74–90.
47. Sun Z, Wang Y, Pan L, Sun Y. Hegyi competition index decomposition to improve estimation accuracy of *Larix olgensis* crown radius. *Ecol Ind.* 2022;143: 109322.
48. Lenz P, Auty D, Achim A, Beaulieu J, Mackay J. Genetic improvement of White Spruce Mechanical Wood traits—early screening by means of Acoustic Velocity. *Forests.* 2013;4:575–94.
49. Chen Z-Q, Karlsson B, Lundqvist S-O, García Gil MR, Olsson L, Wu HX. Estimating solid wood properties using Pilodyn and acoustic velocity on standing trees of Norway spruce. *Ann for Sci.* 2015;72:499–508.
50. Walker TD, Isik F, McKeand SE. Genetic variation in acoustic time of flight and drill resistance of juvenile wood in a large loblolly pine breeding population. *For Sci.* 2019;65:469–82.
51. Zhang S, Sun X, Chen D, Xie Y, Liu C. Method for detecting wood chemical properties of Japanese Larch. *Applied spectroscopy.* 2022;61:882–8.
52. Kalinowski ST, Taper ML, Marshall TC. Revising how the computer program CERVUS accommodates genotyping error increases success in paternity assignment. *Mol Ecol.* 2007;16:1099–106.
53. Robertson A. Experimental design in the evaluation of genetic parameters. *Biometrics.* 1959;15:219.
54. Isik F, Boos DD, Li B. The distribution of genetic parameter estimates and confidence intervals from small disconnected diallels. *Theor Appl Genet.* 2005;110:1236–43.
55. Doyle JJ. Isolation of plant DNA from fresh tissue. *Focus.* 1990;12:13–5.
56. Elshire RJ, Glaubitz JC, Sun Q, Poland JA, Kawamoto K, Buckler ES, et al. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS ONE.* 2011;6: e19379.
57. Wimmer V, Albrecht T, Auinger HJ, Schön CC. synbreed: a framework for the analysis of genomic prediction data using R. *Bioinformatics.* 2012;28:2086–7.
58. R Core Team. R: a language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2014.
59. VanRaden PM. Efficient methods to compute genomic predictions. *J Dairy Sci.* 2008;91:4414–23.
60. Nishio M, Satoh M. Genomic best linear unbiased prediction method including imprinting effects for genomic evaluation. *Genet Selection Evol.* 2015;47:32.
61. Covarrubias-Pazaran G. Genome-assisted prediction of quantitative traits using the R Package Sommer. *PLoS ONE.* 2016;11: e0156744.
62. Akaike H. Information theory and an extension of the maximum likelihood principle. In: Petrov BN, Csáki F, editors. *2nd international symposium on information theory.* Budapest: Akademiai Kiado; 1973.
63. Schwarz G. Estimating the dimension of a model. *Ann Stat.* 1978;6:461–4.
64. de los Campos G, Gianola D, Rosa GJM, Weigel KA, Crossa J. Semi-parametric genomic-enabled prediction of genetic values using reproducing kernel Hilbert spaces methods. *Genetics research.* 2010;92:295–308.
65. Pérez P, Campos G. De Los. genome-wide regression and prediction with the BGLR Statistical Package. *Genetics.* 2014;198:483–95.
66. Namkoong G. Inbreeding effects on estimation of genetic additive variance. *For Sci.* 1966;12:8–13.
67. Heslot N, Jannink JL, Sorrells ME. Perspectives for genomic selection applications and research in plants. *Crop Sci.* 2015;55:1.
68. Aliloo H, Pryce JE, González-Recio O, Cocks BG, Hayes BJ. Accounting for dominance to improve genomic evaluations of dairy cows for fertility and milk production traits. *Genet Selection Evol.* 2016;48:8.
69. Apiolaza LLA, Raymond CA, Yeo BJJ. Genetic variation of physical and chemical wood properties of *Eucalyptus globulus*. *Silvae Genetica.* 2005;54:160–6.
70. Ukrainetz NK, Kang K-Y, Aitken SN, Stoehr M, Mansfield SD. Heritability and phenotypic and genetic correlations of coastal Douglas-fir (*Pseudotsuga menziesii*) wood quality traits. *Can J for Res.* 2008;38:1536–46.
71. Esteban LG, Gril J, De Palacios PDP, Casasús AG. Reduction of wood hygroscopicity and associated dimensional response by repeated humidity cycles. *Ann for Sci.* 2005;59:563–75.
72. Marco de Lima B, Cappa EP, Silva-Junior OB, Garcia C, Mansfield SD, Grattapaglia D. Quantitative genetic parameters for growth and wood properties in *Eucalyptus "urograndis"* hybrid using near-infrared phenotyping and genome-wide SNP-based relationships. *PLoS ONE.* 2019;14:e0218747.
73. Bouvet J-M, Makouanzi Ekomo CG, Brendel O, Laclau J-P, Bouillet J-P, Epron D. Selecting for water use efficiency, wood chemical traits and biomass with genomic selection in a *Eucalyptus* breeding program. *For Ecol Manag.* 2020;465: 118092.
74. Li P, Xiao L, Du Q, Quan M, Song Y, He Y, et al. Genomic insights into selection for heterozygous alleles and woody traits in *Populus tomentosa*. *Plant Biotechnol J.* 2023;21:2002–18.
75. Chen Z-Q, Zan Y, Milesi P, Zhou L, Chen J, Li L, et al. Leveraging breeding programs and genomic data in Norway spruce (*Picea abies* L. Karst) for GWAS analysis. *Genome Biol.* 2021;22:179.
76. Hong Z, Fries A, Wu HX. High negative genetic correlations between growth traits and wood properties suggest incorporating multiple traits selection including economic weights for the future scots pine breeding programs. *Ann for Sci.* 2014;71:463–72.
77. Beaulieu J, Nadeau S, Ding C, Celedon JM, Azaiez A, Ritland C, et al. Genomic selection for resistance to spruce budworm in white spruce and relationships with growth and wood quality traits. *Evol Appl.* 2020;2:eva.13076.
78. Chen ZQ, Baison J, Pan J, Westin J, García Gil MR, Wu HX. Increased prediction ability in Norway spruce trials using a marker \times environment interaction and non-additive genomic selection model. *J Hered.* 2019;40 C:1–30.
79. Gamal El-Dien O, Ratcliffe B, Klápště J, Porth I, Chen C, El-Kassaby Y. Implementation of the realized genomic relationship matrix to open-pollinated white spruce family testing for disentangling additive from nonadditive genetic effects. *G3 (Bethesda, Md).* 2016;6:743–53.
80. Jannink J-L. Identifying quantitative trait locus by genetic background interactions in association studies. *Genetics.* 2007;176:553–61.
81. Zapata-Valenzuela J, Isik F, Maltecca C, Wegrzyn J, Neale D, McKeand S, et al. SNP markers trace familial linkages in a cloned population of *Pinus taeda*—prospects for genomic selection. *Tree Genet Genomes.* 2012;8:1307–18.
82. Crossa J, Pérez P, Hickey J, Burgueño J, Ornella L, Cerón-Rojas J, et al. Genomic prediction in CIMMYT maize and wheat breeding programs. *Heredity.* 2014;112:48–60.
83. Gianola D, van Kaam JBCHM. Reproducing Kernel Hilbert spaces regression methods for genomic assisted prediction of quantitative traits. *Genetics.* 2008;178:2289–303.
84. Tan B, Grattapaglia D, Martins GS, Ferreira KZ, Sundberg B, Ingvarsson PK. Evaluating the accuracy of genomic prediction of growth and wood traits in two *Eucalyptus* species and their F1 hybrids. *BMC Plant Biol.* 2017;17:110.

85. Kadam DC, Lorenz AJ. Evaluation of nonparametric models for genomic prediction of early-stage single crosses in Maize. *Crop Sci.* 2019;59:1411–23.
86. Juliana P, Singh RP, Singh PK, Crossa J, Huerta-Espino J, Lan C, et al. Genomic and pedigree-based prediction for leaf, stem, and stripe rust resistance in wheat. *Theor Appl Genet.* 2017;130:1415–30.
87. Isidro y Sánchez J, Akdemir D. Training set optimization for sparse phenotyping in genomic selection: a conceptual overview. *Front Plant Sci.* 2021;12:715910.
88. Wang M, Xu S. Statistical power in genome-wide association studies and quantitative trait locus mapping. *Heredity.* 2019;123:287–306.
89. Schimleck L, Dahlen J, Apiolaza LA, Downes G, Emms G, Evans R, et al. Non-destructive evaluation techniques and what they tell us about Wood Property Variation. *Forests.* 2019;10: 728.
90. Cown DJ. Comparison of the Pilodyn and torsionmeter methods for the rapid assessment of wood density in living trees. *N Z J Forest Sci.* 1978;8:384–91.
91. Isik F, Li B. Rapid assessment of wood density of live trees using the Resistograph for selection in tree improvement programs. *Can J for Res.* 2003;33:2426–35.

Publisher's Note

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

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

