

Article

Validating Single-Step Genomic Predictions for Growth Rate and Disease Resistance in *Eucalyptus globulus* with Metafounders

Milena Gonzalez ¹, Ignacio Aguilar ², Matias Bermann ³, Marianella Quezada ⁴, Jorge Hidalgo ³,
Ignacy Misztal ³, Daniela Lourenco ³ and Gustavo Balmelli ^{1,*}

¹ Instituto Nacional de Investigación Agropecuaria (INIA), Tacuarembó 45000, Uruguay; mileg2014@gmail.com

² Instituto Nacional de Investigación Agropecuaria (INIA), Montevideo 11500, Uruguay; iaguilar@inia.org.uy

³ Department of Animal and Dairy Science, University of Georgia, Athens, GA 30602, USA; mbermann@uga.edu (M.B.); jh37900@uga.edu (J.H.); ignacy@uga.edu (I.M.); danilino@uga.edu (D.L.)

⁴ Departamento de Biología Vegetal, Facultad de Agronomía, Universidad de la República, Montevideo 12900, Uruguay; mquezada@fagro.edu.uy

* Correspondence: gbalmelli@inia.org.uy

Abstract: Background: Single-step genomic BLUP (ssGBLUP) has gained increasing interest from forest tree breeders. ssGBLUP combines phenotypic and pedigree data with marker data to enhance the prediction accuracy of estimated breeding values. However, potential errors in determining progeny relationships among open-pollinated species may result in lower accuracy of estimated breeding values. Unknown parent groups (UPG) and metafounders (MF) were developed to address missing pedigrees in a population. This study aimed to incorporate MF into ssGBLUP models to select the best parents for controlled mating and the best progenies for cloning in a tree breeding population of *Eucalyptus globulus*. Methods: Genetic groups were defined to include base individuals of similar genetic origin. Tree growth was measured as total height (TH) and diameter at breast height (DBH), while disease resistance was assessed through heteroblasty (the transition from juvenile to adult foliage: ADFO). All traits were evaluated at 14 and 21 months. Two genomic multi-trait threshold linear models were fitted, with and without MF. Also, two multi-trait threshold-linear models based on phenotypic and pedigree information (ABLUP) were used to evaluate the increase in accuracy when adding genomic information to the model. To test the quality of models by cross-validation, the linear regression method (LR) was used. Results: The LR statistics indicated that the ssGBLUP models without MF performed better, as the inclusion of MF increased the bias of predictions. The ssGBLUP accuracy for both validations ranged from 0.42 to 0.68. Conclusions: The best model to select parents for controlled matings and individuals for cloning is ssGBLUP without MF.

Keywords: metafounders; linear regression method; ssGBLUP; genomic selection; Eucalyptus; tree breeding



Academic Editor: Sheng Zhang

Received: 4 May 2025

Revised: 2 June 2025

Accepted: 4 June 2025

Published: 10 June 2025

Citation: Gonzalez, M.; Aguilar, I.; Bermann, M.; Quezada, M.; Hidalgo, J.; Misztal, I.; Lourenco, D.; Balmelli, G. Validating Single-Step Genomic Predictions for Growth Rate and Disease Resistance in *Eucalyptus globulus* with Metafounders. *Genes* **2025**, *16*, 700. <https://doi.org/10.3390/genes16060700>

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Forest tree breeding has successfully delivered genetically improved material for multiple traits through recurrent selection, mating, and testing [1]. However, long breeding cycles, late flowering, variable juvenile-mature correlations, emerging pests, diseases, and climate and market changes pose significant challenges [2]. Open-pollinated populations are commonly used in tree breeding due to their simplicity and cost-effectiveness [3]. This mating design is one of the most frequently employed, as its simplicity and low cost enable the evaluation of the genetic merit of numerous individuals without requiring controlled

mating [4]. Nevertheless, missing pedigree information in such populations complicates accurate genetic evaluation and selection [5]. This challenge arises because open pollination involves random mating with unknown paternal contributors [6].

Genomic selection (GS) is expected to bring a paradigm shift in tree breeding by enhancing its speed and efficiency [7]. By simultaneously fitting all genome-wide markers, GS can capture much of the “missing heritability” associated with complex traits [8]. Single-step genomic best linear unbiased prediction (ssGBLUP) is a widely adopted method in animal breeding that integrates pedigree, phenotypic, and genomic information into a unified evaluation framework [9–12]. In GS, an estimation or “training” population of several hundred or thousands of individuals is genotyped using a genome-wide marker panel and phenotyped for target traits of interest [13]. From these datasets, prediction models are developed and validated in a “validation” set, employing appropriate methods to prevent overfitting [14].

In *Eucalyptus* species, several studies have demonstrated that using genomic tools to estimate relationships improves genetic evaluation accuracy for traits like growth, wood density, and disease resistance [8,15,16]. However, missing pedigree information can introduce biases and inflate genomic estimated breeding values (GEBVs) in ssGBLUP [17]. This issue arises due to pedigree-based and genomic relationship matrices incompatibility, leading to inaccurate genetic evaluations [18]. Unknown parent groups (UPG) were developed to account for missing pedigrees in a population [19,20]; assigning UPG to missing parents helps mitigate biases in genetic evaluations [21,22]. Nevertheless, the effectiveness of UPG in ssGBLUP depends on the availability of sufficient data; inadequate data can lead to biased genetic trends and underestimated UPG effects [22,23]. The UPG approach still assumes that the base populations are unrelated, which is often not true [22,24]. In practice, the pedigrees are incomplete, and individuals with information descend from different base populations [25].

Metafounders (MF) also model missing pedigrees by assuming that an ancestral population is represented as a single pseudo-individual (a metafounder) with a particular self-relationship (a measure of homozygosity) that represents a pool of gametes [26]. The MF approach generalizes UPG by assuming them as inbred and related by a covariance matrix called Γ [26–28]. This method has shown promise in reducing biases, especially in multibreed or crossbred evaluations, provided the dataset allows for estimating Γ reasonably well [26,29]. When MF are used, individuals from different breeds can be related through pedigree, improving the compatibility between genomic and pedigree relationships [24,30]. In the MF approach, the goal is to adjust the pedigree-based relationship matrix (A) to achieve compatibility with a genomic relationship matrix (G) constructed with an assumed allelic frequency of 0.5 [26]. The vector of additive genetic effects (u) in this model encompasses genotyped and non-genotyped individuals and metafounders.

Eucalyptus globulus is a globally significant species in forestry, particularly valued for its rapid growth and high-quality wood for pulp and paper production [31]. In recent years, *E. globulus* plantations have been severely affected by the disease caused by *Teratosphaeria nubilosa* [32], significantly damaging young plantations [33]. *T. nubilosa* is a serious pathogen affecting many *Eucalyptus* spp. leaves [32,34]. This pathogen primarily infects juvenile and intermediate foliage, causing severe leaf spotting, necrosis, defoliation, and shoot blight [35]. In *E. globulus*, an early heteroblastic transition has been observed as a natural defense mechanism [36]. Adult leaves demonstrate significantly lower susceptibility to *T. nubilosa* than juvenile foliage [32]. In regions where *T. nubilosa* is prevalent, alternative *Eucalyptus* spp. or hybrids with inherent resistance are being explored [37]. GS models offer an effective tool to predict genetic potential for disease resistance by analyzing the entire genome, enabling breeders to select individuals with superior resistance traits [38,39].

In 2017, the National Agricultural Research Institute (INIA) in Uruguay initiated a program to develop GS models to predict growth and disease resistance in *E. globulus*. The first approach, carried out by Quezada et al. [40] using an open-pollinated population, achieved high accuracies in estimating genetic parameters and improving the prediction of breeding values for ranking and selecting individuals. To further advance the effective incorporation of GS into the breeding program for *E. globulus*, the main objective of this study was to compare the GS using in a multi-trait threshold linear models with and without MF, to select the best parents for controlled mating and the best progenies for cloning.

2. Materials and Methods

2.1. Study Population

The study population is derived from the *E. globulus* tree improvement program coordinated by INIA Uruguay. It includes both parent trees and their progeny. Parent material originated from two seed orchards established in 1996 (first generation) and 2002 (second generation). Progeny were evaluated in six trials located in southeastern Uruguay (Lavalleja and Rocha), regions commonly used for *E. globulus* plantations. The first trial, established in 2011, consists of 3853 individuals from 194 open-pollinated (OP) half-sib families, primarily from the first-generation orchard (see Quezada et al. [40]). The other five trials, established between 2015 and 2019, comprise 6051 individuals from 137 full-sib families generated through controlled matings (CM) from the second-generation orchard (Table 1).

Table 1. Summary of phenotypic and genotypic data across all trials. The table includes the installation year, tree age at measurement, mating type (open pollination, OP; or controlled mating, CM), number of trees phenotyped, and number of trees genotyped for each trial. Progeny tests from OP represent half-sib families, while those from CM represent full-sib families.

Trials	Installation Year	Age (Year)	Matting Type	Number of Trees Phenotyped	Number of Trees Genotyped
Parents (seed orchard first generation)	1996	28	OP	604	80
Parents (seed orchard second generation)	2002	22	OP	455	446
Progeny test 1	2011	13	OP	3756	975
Progeny test 2	2014	10	CM	551	87
Progeny test 3	2015	9	CM	644	108
Progeny test 4	2016	8	CM	989	194
Progeny test 5	2017	7	CM	1591	199
Progeny test 6	2019	5	CM	2276	320
Total				10,373	2409

Tree growth was assessed by measuring total height (TH) with a hypsometer (Vertex IV HS 102, Haglöf Sweden, Långsele, Sweden) and diameter at breast height (DBH) using a diameter tape at 14 and 21 months, respectively. In addition, the precocity of the transition from juvenile to adult foliage (ADFO) was evaluated at the same time points. ADFO was visually estimated as the percentage of the crown exhibiting adult foliage, using a scale in 10% increments. Due to an uneven distribution of observations across these intervals, values were consolidated into three categories for analysis: 1 = no adult foliage, 2 = up to 50%, and 3 = more than 50%. ADFO was analyzed as a categorical trait, while TH and DBH were treated as quantitative traits. At 14 months, most individuals exhibited only juvenile

foliage; by 21 months, the majority showed up to 50% adult foliage, with relatively few exceeding this threshold (Table 2).

Table 2. Distribution of trees across adult foliage (ADFO) categories at two time points: 14 months (ADFO14) and 21 months (ADFO21). Category 1 = no adult foliage, Category 2 = up to 50% adult foliage, and Category 3 = more than 50% adult foliage. The number of trees per category is shown for each trait.

Trait	Number of Trees	Category
ADFO14	6145	1
ADFO14	3565	2
ADFO14	192	3
ADFO21	2139	1
ADFO21	5658	2
ADFO21	1828	3

The genotypic dataset comprised 2,409 individuals, including both parents and progeny from OP and CM families (Table 1). Genotyping was performed using the *Eucalyptus* EUchip60K SNP chip [41] by GeneSeek Inc. (Lincoln, NE, USA). Quality control filtering was conducted with PREGSF90 version 1.26 [42], retaining markers with a call rate > 0.95 and a minor allele frequency (MAF) > 0.05, as well as individuals with a call rate > 0.80. Following filtering, 15,821 SNP markers and 2359 individuals remained for downstream analyses. Pedigree correction was carried out using SEEKPARENTF90 version 1.56 from the BLUPF90 software suite version 2.6 [43].

2.2. Multi-Trait Threshold Linear ssGBLUP Model

Four multi-trait threshold linear models were compared to check the effect of including MF with and without genomic information. The matrix notation is specified by Equation (1):

$$\begin{matrix} y_1 \\ y_2 \\ l_3 \\ l_4 \end{matrix} = \begin{bmatrix} X_1 & O & O & O \\ O & X_2 & O & O \\ O & O & X_3 & O \\ O & O & O & X_4 \end{bmatrix} \begin{matrix} b_1 \\ b_2 \\ b_3 \\ b_4 \end{matrix} + \begin{bmatrix} W_1 & O & O & O \\ O & W_2 & O & O \\ O & O & W_3 & O \\ O & O & O & W_4 \end{bmatrix} \begin{matrix} p_1 \\ p_2 \\ p_3 \\ p_4 \end{matrix} + \begin{bmatrix} Z_1 & O & O & O \\ O & Z_2 & O & O \\ O & O & Z_3 & O \\ O & O & O & Z_4 \end{bmatrix} \begin{matrix} u_1 \\ u_2 \\ u_3 \\ u_4 \end{matrix} + \begin{matrix} e_1 \\ e_2 \\ e_3 \\ e_4 \end{matrix} \quad (1)$$

where y_1 and y_2 is the vector of observed phenotypes for TH and DBH, l_3 and l_4 is a vector of unobserved liabilities for ADFO at 14 months and ADFO at 21 months, b is the vector of fixed effects, including the overall mean of sites and blocks (within site) with incidence matrix X ; p is the vector of random plot effects with incidence matrix W ; u is the vector of random genetic additive effects of the individual trees, with incidence matrix Z ; and e is the vector of random residuals. Under the assumed threshold model, the observed phenotypes for individual i is dictated by its liability (l_3 and l_4), with the following relationship: if l_3 and l_4 exceeds an unknown fixed threshold (t), then y_3 and $y_4 = 1$ (no adult foliage), y_3 and $y_4 = 2$ (up to 50% adult foliage) and y_3 and $y_4 = 3$ (more than 50% adult foliage). We assumed that liability was normally distributed with mean vector θ and variance σ_e^2 as specified by Equation (2):

$$l_3 \text{ and } l_4 | \theta, \sigma_e^2 \sim N(S\theta, I\sigma_e^2); \sigma_e^2 \quad (2)$$

where $\theta' = (b', p', u')$ is a vector of systematic and random effects, and S is an incidence matrix (containing the X , W and Z) for linking θ to the phenotypic records. The FOAD response at 14 months, given the liability and the threshold, can be expressed as Equation (3):

$$p(y_3|l_3, t) = \prod_{i=1}^n [I(l_3 \leq t_1)I(y_3 = 1) + I(t_1 < l_3 \leq t_2)I(y_3 = 2) + I(t_2 < l_3)I(y_3 = 3)] \quad (3)$$

where t_1 and t_2 are thresholds that define the three categories of response and I is an indicator function that takes value 1 if the condition specified is true, otherwise the value is 0. The FOAD response at 21 months, given the liability and the threshold, is expressed in the same way.

The models differed in the relationship matrices used for the variance of additive genetic effects. The first genomic model assumed $\text{Var}(\mathbf{u}) = \mathbf{H} \otimes \Sigma$ (Aguilar et al., 2010 [11]), whereas the second included metafounders and assumed that $\text{Var}(\mathbf{u}) = \mathbf{H}^F \otimes \Sigma$. Similarly, the first model assumed $\text{Var}(\mathbf{u}) = \mathbf{A} \otimes \Sigma$, and the other assumed $\text{Var}(\mathbf{u}) = \mathbf{A}^F \otimes \Sigma$.

The (co)variance structure was defined to include additive genetic, plot, and residual variances for each trait, incorporating both pedigree and genomic information, as shown in (4):

$$\text{Var} \begin{bmatrix} \mathbf{p} \\ \mathbf{u} \\ \mathbf{e} \end{bmatrix} = \begin{bmatrix} \mathbf{I} \otimes \mathbf{P}_0 & 0 & 0 \\ 0 & \mathbf{H} \otimes \Sigma / \mathbf{H}^F \otimes \Sigma / \mathbf{A} \otimes \Sigma / \mathbf{A}^F \otimes \Sigma & 0 \\ 0 & 0 & \mathbf{I} \otimes \mathbf{R}_0 \end{bmatrix} \quad (4)$$

where Σ , $\mathbf{P}_0$ and $\mathbf{R}_0$ are 4×4 matrices of (co)variances for additive genetic, plot effects, and residual; $\mathbf{I}$ is the identity matrix, a square matrix with 1 on the diagonal and 0 on the off-diagonal, $\mathbf{H}$ is the realized relationship matrix that combines pedigree and genomic relationships. $\mathbf{H}^F$ and $\mathbf{A}^F$ are the relationship matrices augmented with metafounders.

The inverse of $\mathbf{H}$ ($\mathbf{H}^{-1}$) was derived by Aguilar et al. [11] as defined by Equation (5):

$$\mathbf{H}^{-1} = \mathbf{A}^{-1} + \begin{bmatrix} 0 & 0 \\ 0 & \mathbf{G}^{-1} - \mathbf{A}_{22}^{-1} \end{bmatrix} \quad (5)$$

where $\mathbf{A}^{-1}$ and $\mathbf{A}_{22}^{-1}$ are the inverse of the pedigree relationship matrices for all individuals and only for the genotyped individuals, respectively, and $\mathbf{G}^{-1}$ is the inverse of the genomic relationship matrix, calculated according to VanRaden [44]. The MF were defined based on the origin and breeding generations of the seeds: Australia, Uruguay, and two seed orchards (established in 1996 and 2002).

2.3. Variance Components

Variance components and (G)EBVs for each model were estimated using GIBBSF90+ [45] version 3.2. For variance components, the Gibbs sampling process comprised 300,000 samples, and one every 50th sample was stored. After discarding the first 10,000 samples as burn-in, posterior means were calculated. For (G)EBV, variance components were fixed to the posterior means, and 30,000 samples were drawn. Variance components were calculated with genomic information for four models. Posterior means of (G)EBV were used in the validation process.

The narrow sense heritability for all traits was estimated for the models without MF (ssGBLUP and ABLUP), as follows in (6):

$$h^2 = \frac{\sigma_a^2}{\sigma_a^2 + \sigma_p^2 + \sigma_e^2} \quad (6)$$

where σ_a^2 is the additive genetic variance, σ_p^2 is the plot variance, and σ_e^2 is the residual variance.

The genetic correlations for all traits were estimated for the models without MF (ssGBLUP and ABLUP) as follows in (7):

$$rg_{ij} = \frac{\sigma_{ij}}{\sqrt{\sigma_i^2 + \sigma_j^2}} \quad (7)$$

where σ_i^2 and σ_j^2 are the additive genetic variances for the i th and j th traits, respectively, and σ_{ij} is their covariance.

Non-additive genetic effects and genotype-by-environment ($G \times E$) interactions were excluded from narrow-sense heritability estimates, as only additive genetic variance contributes directly to predictable evolutionary change. In crosses within the same species, particularly among individuals from genetically similar populations, genetic divergence is limited, and most of the genetic variation is expected to be additive. $G \times E$ interactions were also omitted from the analysis because selection was aimed at a single breeding zone (southeastern Uruguay), where environmental conditions are relatively uniform. The focus was on identifying genotypes with superior performance in this specific context.

All analyses were carried out using BLUPF90 family software version 2.6 [46] and R package ggplot2 [47] version 3.4.2.

2.4. Model Validation Using LR

To complement cross-validation approaches, we followed a semiparametric procedure based on the classical theory of genetic evaluation, the linear regression method (LR) [48]. They proposed to test the quality of evaluation methods using cross-validation statistics based on successively estimated breeding values (EBV) from a set of “focal” individuals. These “focal” individuals can be the whole population or a group of individuals of interest, such as candidates for selection. The focal individuals have no phenotype in the partial dataset (p) but have a phenotype in the whole dataset (w). The whole dataset contains all phenotypes, and the partial dataset contains phenotypes of a smaller group of individuals. Then, validation statistics compare EBV obtained from the “whole” dataset versus EBV obtained in the “partial” dataset. According to Legarra and Reverter [48], the correlation between the whole and partial datasets is a measure of the expected magnitude of the change in accuracy with increasing information, as specified by Equation (8):

$$\rho_{wp} = \frac{cov(\hat{\mathbf{u}}_w, \hat{\mathbf{u}}_p)}{\sqrt{var(\hat{\mathbf{u}}_w) var(\hat{\mathbf{u}}_p)}} \quad (8)$$

where $\hat{\mathbf{u}}_w$ and $\hat{\mathbf{u}}_p$ are the GEBV estimated with the whole and partial data, respectively. The bias is defined as the difference in their means as follows in (9):

$$\mu_{wp} = \bar{\hat{\mathbf{u}}_p} - \bar{\hat{\mathbf{u}}_w} \quad (9)$$

and the dispersion as the slope of the regression of $\hat{\mathbf{u}}_w$ on $\hat{\mathbf{u}}_p$ as follows in (10):

$$b_{wp} = \frac{cov(\hat{\mathbf{u}}_w, \hat{\mathbf{u}}_p)}{var(\hat{\mathbf{u}}_p)} \quad (10)$$

The bias has an expected value of 0 if the evaluation is unbiased. The dispersion has an expected value of 1 if there is no over/under dispersion. The accuracy of GEBV in the partial dataset is calculated as follows in (11):

$$\rho_{cov(w,p)} = \sqrt{\frac{cov(\hat{\mathbf{u}}_w, \hat{\mathbf{u}}_p)}{(1 - \bar{F}) \sigma_u^2}} \quad (11)$$

where $\bar{F}$ is the average inbreeding coefficient of the focal set. Additionally, Bermann et al. [49] calculated the relative increase in accuracy by adding phenotypic information as $\rho_{wp}^{-1} - 1$, and the relative increase in accuracy by adding genomic information to the partial dataset as $\rho_{A,G}^{-1} - 1$. Both formulas were multiplied by 100 to express them in percentage.

We conducted two validation analyses to evaluate the different models' predictive performance. The first assessed the ability to predict the breeding values of parents (for controlled mating), while the second evaluated the ability to predict the breeding values of individuals (for cloning).

The parent's validation group consisted of parents without phenotyped progeny in the reduced dataset (validation set) but with phenotyped progeny in the full dataset (training set). Parent validation was performed under two scenarios: one involving selecting 100 parents and another selecting only 50 parents (Table 3).

Table 3. Validation design for parent selection under two scenarios (V1 and V2). Each scenario includes the number of phenotypic and genotypic records in the training and validation sets, the progeny tests included in training, and the number of parents evaluated in validation. Progeny tests are labeled from 1 to 6.

	Phenotypes	Set Training		Phenotypes	Set Validation	
		Genotypes	Number of Progeny Test		Genotypes	Parents
V1	9832	2359	1, 2, 3, 4, 5, 6	2121	478	100
V2	9832	2359	1, 2, 3, 4, 5, 6	1329	272	50

Cross-validation for individuals involved progenies without phenotypic data in the reduced dataset (validation set) but with phenotypic data in the full dataset (training set). For the reduced dataset, the progeny test 5 was used (Table 4). Ideally, evaluations at 14 or 21 months, the target selection age, would have been preferable. For this reason, only one validation was performed.

Table 4. Validation design for the selection of individuals for cloning. The table shows the number of phenotypic and genotypic records in the training and validation sets, along with the progeny tests included in each. Progeny tests 1 to 4 were used for training, while progeny test 5 was used for validation.

Phenotypes	Set Training		Phenotypes	Set Validation	
	Genotypes	Number of Progeny Test		Genotypes	Number of Progeny Test
5965	2359	1, 2, 3, 4	1591	199	5

The LR statistics were applied only to the genotyped individuals in all validations.

3. Results

3.1. Genetic Parameters

The narrow-sense heritability estimates obtained using the genomic model (ssGBLUP) were not significantly different from those using the pedigree-based model (ABLUP), except for DBH21 (Table 5). Heritability values ranged from 0.37 to 0.84 in the ssGBLUP model and from 0.33 to 0.86 in the ABLUP approach.

The genetic correlations among all traits were positive and ranged from moderate to high, with values between 0.51 and 0.97 across both models (Table 6). DBH showed the strongest correlations with all traits.

Table 5. Narrow-sense heritability estimates (h^2) and 95% high posterior density (HPD) intervals for four traits in *Eucalyptus globulus* under the ssGBLUP and ABLUP models. Traits include total tree height at 14 months (TH14), proportion of adult foliage at 14 and 21 months (ADFO14 and ADFO21), and diameter at breast height at 21 months (DBH21).

Trait	ssGBLUP		h^2	ABLUP HPD Interval (95%)
	h^2	HPD Interval (95%)		
TH14	0.37	0.31–0.43	0.33	0.26–0.40
ADFO14	0.84	0.79–0.89	0.84	0.77–0.88
DBH21	0.53	0.47–0.58	0.65	0.62–0.69
ADFO21	0.83	0.78–0.88	0.86	0.80–0.90

Table 6. Genetic correlations (rg) among traits in *Eucalyptus globulus* estimated using ssGBLUP (above the diagonal) and ABLUP (below the diagonal) models. Values in parentheses indicate the 95% high posterior density (HPD) intervals. Traits include total tree height at 14 months (TH14), proportion of adult foliage at 14 and 21 months (ADFO14, ADFO21), and diameter at breast height at 21 months (DBH21).

rg	TH14	ADFO14	DBH21	ADFO21
TH14	1	0.52 (0.44–0.60)	0.83 (0.79–0.86)	0.51 (0.44–0.59)
ADFO14	0.64 (0.56–0.71)	1	0.66 (0.60–0.72)	0.97 (0.95–0.98)
DBH21	0.77 (0.73–0.81)	0.69 (0.64–0.75)	1	0.70 (0.64–0.75)
ADFO21	0.61 (0.53–0.70)	0.97 (0.95–0.99)	0.73 (0.67–0.78)	1

3.2. Model Comparison for Parent Selection

The LR statistics estimated for both models showed that the accuracy of the genomic-based models (ssGBLUP) was always higher than that of the pedigree-based models (ABLUP) in both validation scenarios, ranging from 0.37 to 0.66 with genomic information and 0.23 to 0.51 using only pedigree information (Table 7). The estimated bias was smaller with genomic information than without it in both validation scenarios, ranging from 0.08 to 0.27 and 0.21 to 0.36, respectively. When comparing the dispersion in both validation scenarios, using pedigree information in the first scenario showed slightly better dispersion for growth traits (TH and DBH) and nearly equal dispersion for heteroblasty (ADFO), ranging from 0.77 to 0.95. However, in the second validation scenario, the dispersion was better for models based on genomic information, with estimated values close to one, ranging from 0.82 to 0.95. The correlation between whole and partial dataset for ssGBLUP was always higher than for ABLUP in both validation scenarios, ranging from 0.52 to 0.77 and 0.28 to 0.58, respectively. The second validation showed the best values of accuracy, bias, dispersion, and correlation for ssGBLUP, making this the optimal scenario for validation.

The increase in accuracy by adding genomic information was higher than by adding phenotypes to the partial dataset (Table 8) for all traits, especially for growth traits (TH and DBH).

The LR statistics estimated for both models with MF showed that the accuracy with genomic information (ssGBLUP_MF) was always higher than using only pedigree information (ABLUP_MF) models in both validation scenarios, ranging from 0.36 to 0.67 and 0.21 to 0.48, respectively (Table 9). Both models exhibited bias in both validation scenarios, with ssGBLUP_MF displaying a greater range, from 0.15 to 1.6, compared to ABLUP_MF, which ranged from 0.04 to 0.43. When comparing the dispersion in both validation scenarios, ssGBLUP_MF showed values closer to one, ranging from 0.72 to 0.93, whereas for ABLUP_MF the range was from 0.66 to 0.91. The first validation showed practically equal dispersion values for all traits. However, in the second validation scenario, when only pedigree information with MF was used, the estimated dispersion was more variable. In

contrast, with genetic information and MF, the estimates were closer to one. The correlation between whole and partial dataset was higher with ssGBLUP_MF than using ABLUP_MF, ranging from 0.57 to 0.80 and 0.29 to 0.58, respectively, in both validation scenarios. The second validation showed the best values of accuracy, bias, dispersion and correlation for ssGBLUP_MF, making this a better scenario for validation.

Table 7. Linear regression statistics for the ssGBLUP and ABLUP models under two validation scenarios (V1 and V2) for selected parents. Reported metrics include accuracy, standardized bias, dispersion, and the correlation between whole and partial dataset (p(wp)). Traits include total tree height at 14 months (TH14), proportion of adult foliage at 14 and 21 months (ADFO14, ADFO21), and diameter at breast height at 21 months (DBH21).

Traits	Validation	ssGBLUP				ABLUP			
		Accuracy	Standardized Bias	Dispersion	p(wp)	Accuracy	Standardized Bias	Dispersion	p(wp)
TH14	V1	0.37	−0.24	0.72	0.59	0.27	−0.26	0.77	0.43
ADFO14	V1	0.61	−0.16	0.94	0.72	0.43	−0.29	0.95	0.50
DBH21	V1	0.37	−0.27	0.73	0.52	0.23	−0.36	0.77	0.28
ADFO21	V1	0.59	−0.17	0.90	0.70	0.43	−0.29	0.90	0.49
TH14	V2	0.42	−0.15	0.88	0.69	0.33	−0.24	0.73	0.51
ADFO14	V2	0.66	−0.08	0.95	0.77	0.51	−0.21	0.90	0.58
DBH21	V2	0.41	−0.14	0.82	0.60	0.30	−0.31	0.80	0.37
ADFO21	V2	0.64	−0.08	0.92	0.75	0.50	−0.23	0.86	0.57

Table 8. Relative increase in prediction accuracy (%) for the second validation scenario when adding phenotypic (Inc Phen%) and genotypic (Inc Geno%) information to the partial dataset. Traits include total tree height at 14 months (TH14), proportion of adult foliage at 14 and 21 months (ADFO14, ADFO21), and diameter at breast height at 21 months (DBH21).

Traits	Inc Phen%	Inc Geno%
TH14	45	96
ADFO14	30	61
DBH21	67	194
ADFO21	33	59

Table 9. Linear regression statistics for the ssGBLUP_MF and ABLUP_MF models under two validation scenarios (V1 and V2) for selected parents. Metrics include prediction accuracy, standardized bias, dispersion, and the correlation between whole and partial dataset (p(wp)). Traits evaluated are total tree height at 14 months (TH14), proportion of adult foliage at 14 and 21 months (ADFO14, ADFO21), and diameter at breast height at 21 months (DBH21).

Traits	Validation	ssGBLUP_MF				ABLUP_MF			
		Accuracy	Standardized Bias	Dispersion	p(wp)	Accuracy	Standardized Bias	Dispersion	p(wp)
TH14	V1	0.36	−1.60	0.72	0.63	0.24	−0.35	0.72	0.40
ADFO14	V1	0.61	−1.14	0.93	0.76	0.40	0.18	0.91	0.49
DBH21	V1	0.39	−0.87	0.72	0.57	0.21	−0.12	0.73	0.29
ADFO21	V1	0.60	−0.72	0.89	0.74	0.40	0.14	0.88	0.48
TH14	V2	0.42	−0.64	0.81	0.71	0.30	−0.43	0.66	0.47
ADFO14	V2	0.67	−1.02	0.93	0.80	0.48	−0.43	0.85	0.58
DBH21	V2	0.42	−0.15	0.76	0.63	0.27	−0.04	0.73	0.38
ADFO21	V2	0.66	−0.31	0.88	0.78	0.48	−0.25	0.82	0.57

When comparing the four models, it is observed that the second validation scenario with the ssGBLUP model had the best predictive performance, with dispersion near one, relatively high accuracy (Figure 1), high correlation between whole and partial dataset and bias near zero (Figure 2).

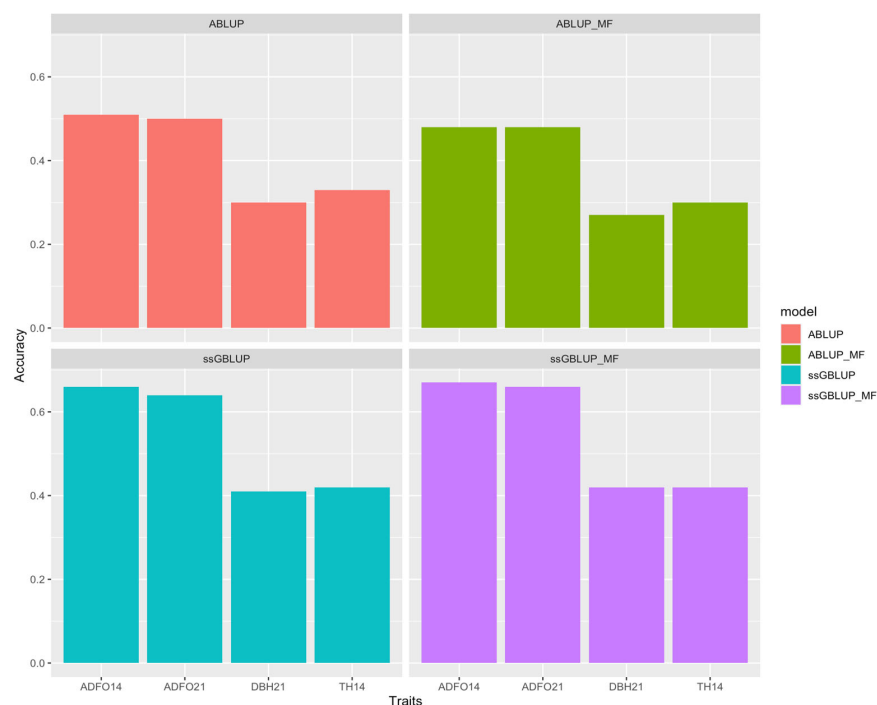


Figure 1. Accuracy of models for all traits in the second scenario of parent validation for ABLUP, ABLUP_MF, ssGBLUP, and ssGBLUP_MF models. ADFO14 and ADFO21: fraction of adult foliage in the canopy at 14 and 21 months; DBH21: diameter at breast height at 21 months; and TH14: total tree height at 14 months.

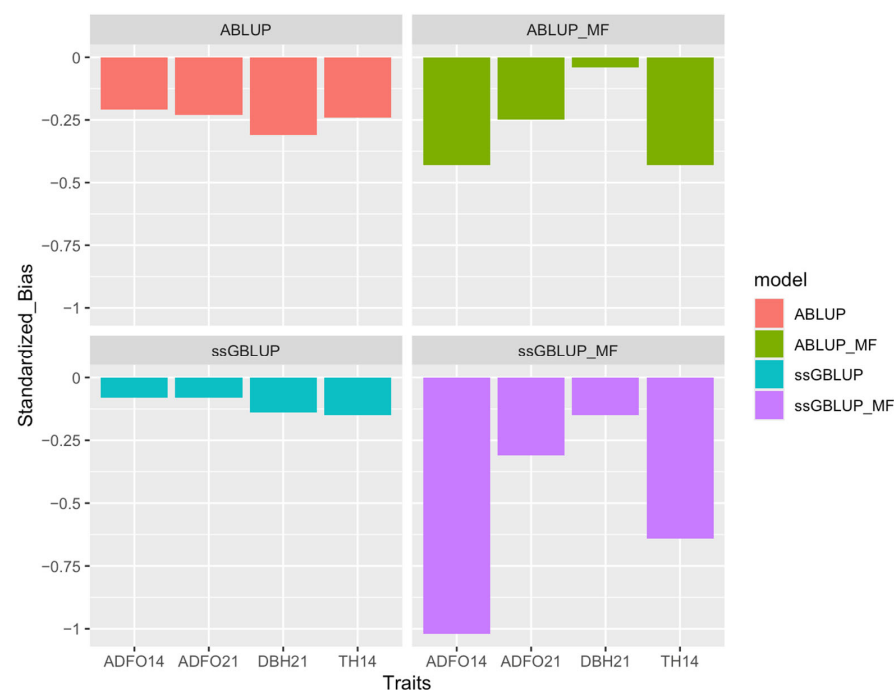


Figure 2. Bias of models for all traits in the second scenario of parent validation for ABLUP, ABLUP_MF, ssGBLUP, and ssGBLUP_MF models. ADFO14 and ADFO21: fraction of adult foliage in the canopy at 14 and 21 months; DBH21: diameter at breast height at 21 months; and TH14: total tree height at 14 months.

3.3. Model Comparison for Individuals

The LR statistics estimated for both models showed that the accuracy of the genomic-based models (ssGBLUP) was always higher than that of the pedigree-based (ABLUP) models, ranging from 0.48 to 0.68 with genomic information and 0.37 to 0.53 using only

pedigree information (Table 10). The estimated bias was near zero with both models, except for ADFO14 and ADFO21 using genomic information (ssGBLUP) and HT14 using ABLUP, ranging from 0.01 to 0.17 and 0.05 to 0.15, respectively. The dispersion using genomic information was slightly better for all traits (close to one) than using ABLUP, ranging from 0.90 to 1.09 and 0.83 to 1.20, respectively. The correlation between whole and partial dataset was higher with ssGBLUP than using ABLUP, ranging from 0.74 to 0.80 and 0.54 to 0.67, respectively.

Table 10. Linear regression statistics for the ssGBLUP and ABLUP models used in the selection of individual trees for cloning. Reported metrics include prediction accuracy, standardized bias, dispersion, and the correlation between whole and partial dataset (p(wp)). Traits include total tree height at 14 months (TH14), proportion of adult foliage at 14 and 21 months (ADFO14, ADFO21), and diameter at breast height at 21 months (DBH21).

Traits	ssGBLUP				ABLUP			
	Accuracy	Standardized Bias	Dispersion	p(wp)	Accuracy	Standardized Bias	Dispersion	p(wp)
HT14	0.48	−0.01	0.90	0.74	0.45	0.15	0.83	0.66
ADFO14	0.66	−0.15	1.08	0.80	0.53	−0.05	1.16	0.67
DBH21	0.48	−0.07	0.94	0.74	0.37	0.09	0.99	0.54
ADFO21	0.68	−0.17	1.09	0.80	0.53	−0.07	1.20	0.67

The LR statistics estimated for both models with MF showed that the accuracy with genomic information (ssGBLUP_MF) was always higher than when using only pedigree information with MF (ABLUP_MF), ranging from 0.57 to 0.85 and 0.37 to 0.53, respectively (Table 11). The bias estimation was higher with genomic information than without, except for DBH, ranging from 0.04 to 1.28 and 0.36 to 0.93, respectively. When comparing the dispersion in both models, the model with genomic information and MF had estimates close to one. In contrast, the model with only pedigree information and MF exhibited more variable dispersions, ranging from 0.82 to 1.17. The correlation between whole and partial dataset was higher with ssGBLUP_MF (p(wp)) = 0.99 than using ABLUP_MF (p(wp)) = 0.60 to 0.69.

Table 11. Linear regression statistics for the ssGBLUP_MF and ABLUP_MF models using a validation method for the selection of individual trees. Metrics reported include prediction accuracy, standardized bias, dispersion, and the correlation between whole and partial dataset (p(wp)). Traits evaluated are total tree height at 14 months (TH14), proportion of adult foliage at 14 and 21 months (ADFO14, ADFO21), and diameter at breast height at 21 months (DBH21).

Traits	ssGBLUP_MF				ABLUP_MF			
	Accuracy	Standardized Bias	Dispersion	p(wp)	Accuracy	Standardized Bias	Dispersion	p(wp)
HT14	0.57	−1.14	1.005	0.99	0.44	−0.93	0.82	0.69
ADFO14	0.82	−1.28	1.007	0.99	0.52	−0.36	1.14	0.68
DBH21	0.59	−0.04	1.004	0.99	0.37	−0.59	0.97	0.60
ADFO21	0.85	−0.41	1	0.99	0.53	−0.43	1.17	0.69

The statistics of predictive performance for the four models were better with the ssGBLUP model, with relatively high accuracy, except for ssGBLUP_MF (Figure 3), bias near zero (Figure 4), dispersion near one and high correlation between whole and partial dataset.

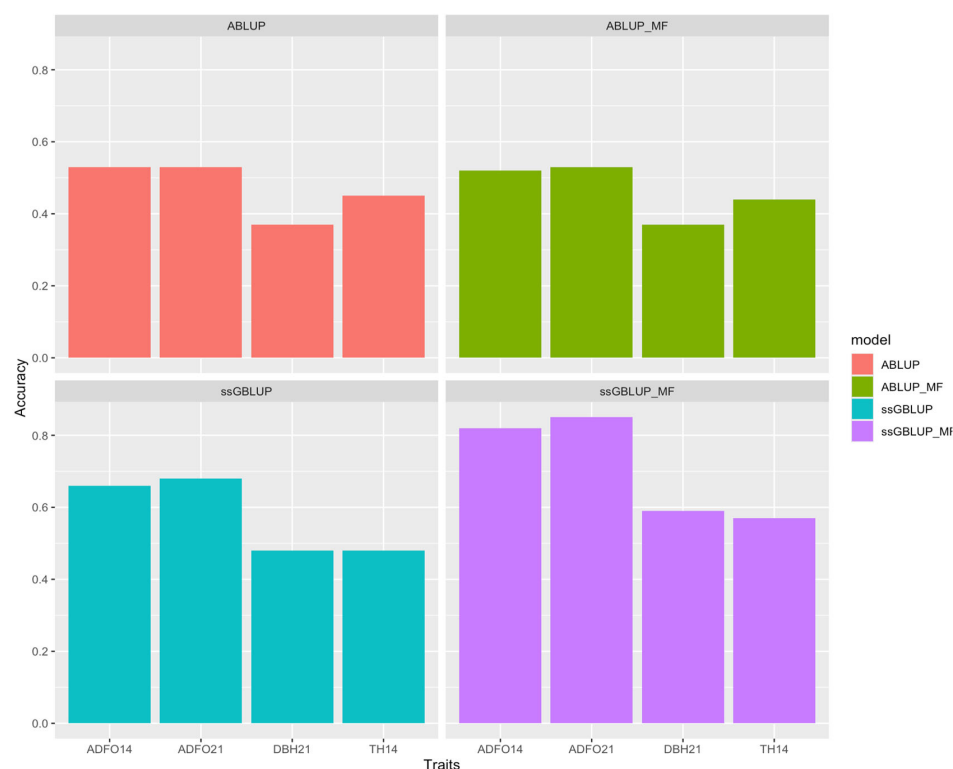


Figure 3. Accuracy of models for all traits in individual validation for ABLUP, ABLUP_MF, ssGBLUP, and ssGBLUP_MF models. ADFO14 and ADFO21: fraction of adult foliage in the canopy at 14 and 21 months; DBH21: diameter at breast height at 21 months; and TH14: total tree height at 14 months.

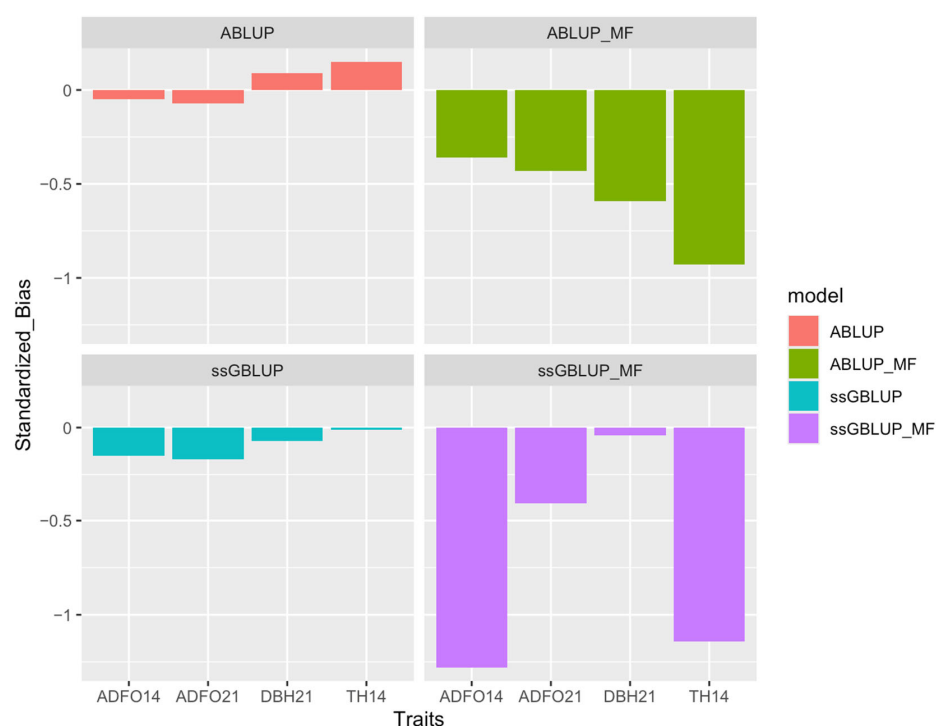


Figure 4. Bias of models for all traits in individual validation for ABLUP, ABLUP_MF, ssGBLUP, and ssGBLUP_MF models. ADFO14 and ADFO21: fraction of adult foliage in the canopy at 14 and 21 months; DBH21: diameter at breast height at 21 months; and TH14: total tree height at 14 months.

4. Discussion

Tree improvement programs globally are not very advanced due to the long generation cycles of even the fastest-growing species [50]. Tree selection based on genomic information

has become an important tool in forest tree breeding [51]. Estimating genetic parameters plays a crucial role in the management of seed orchards and provides valuable guidance for developing the evaluation and selection strategy for the next generation of improvement [52,53]. In this study, the heritabilities estimated using ssGBLUP were consistently lower than those obtained from ABLUP across all evaluated traits. This discrepancy is consistent with findings in previous research and is often attributed to the increased resolution provided by genomic information, which enables more accurate partitioning of genetic and residual variances [10,54]. Marker-based relationship matrices capture Mendelian sampling and cryptic relationships that are not reflected in pedigree-based models, thereby refining the estimation of additive genetic variance and potentially reducing the estimated heritability [44,55]. In contrast, pedigree-based models may overestimate heritability by attributing unaccounted environmental or non-additive effects to additive genetic variance [56,57]. Genomic models, particularly those incorporating dense marker information, can disentangle these effects more accurately by capturing the realized proportion of shared alleles among individuals, thus providing a more precise estimate of additive genetic variance [44,55,58].

4.1. Predictive Model Performance

The effectiveness of genomic selection in breeding programs depends on the phenotypic quality and depth, the prediction model, the number and type of molecular markers, the size and composition of the training population, the accuracy of predicted genomic breeding values, and the relatedness between the training and validation populations [54,59]. Therefore, the optimal training population design is one of the most challenging aspects of GS. Beaulieu et al. [60] and Klápště et al. [52] underline the importance of the relationship between training and test populations in genomic selection models. The cross-validation approaches proposed by Legarra and Reverter [48], including “partial” and “whole” data based on differences in means, covariance, and correlation, are not usually implemented in tree breeding evaluations. In this study, the increase in accuracy obtained with ssGBLUP suggests that incorporating genomic information significantly influences the selection of superior individuals. Several studies have demonstrated the effectiveness of ssGBLUP in enhancing prediction accuracy for growth and wood quality traits in tree species [61–64]. By using realized genomic relationships rather than expected pedigree-based ones, ssGBLUP enhances the accuracy of breeding value predictions [10,44]. This improvement is especially relevant in populations with unbalanced pedigrees, where traditional models may not fully capture the true genetic variability [54]. The inclusion of genomic data allows for better discrimination among individuals with similar pedigree backgrounds but different genetic merits [65]. This enhanced resolution improves the estimation of breeding values, allowing for more precise selection of superior individuals and reducing the risk of selecting suboptimal candidates based solely on pedigree [44,54]. As a result, genomic selection contributes to increased selection accuracy and, consequently, a higher rate of genetic gain per unit of time or generation [66–68]. In our study, integrating genomic information led to increased accuracy and reduced bias in estimated breeding values (EBVs) compared to traditional pedigree-based models. Additionally, the dispersion of EBVs improved, with values approaching the ideal of one, indicating more accurate and consistent predictions of genetic merit. These findings align with previous studies that have reported similar improvements in accuracy through the inclusion of genomic data [40,52,64] and observed less biased predictions compared to ABLUP [63,69].

Metafounders are useful for characterizing relationships within and across populations by modeling the means and variances of unknown base population individuals [26,70]. For instance, Macedo et al. [71] demonstrated that metafounders enhanced the accuracy of genomic predictions by accounting for additional genetic variation that is not captured

by traditional pedigree-based relationships. Similarly, Kudinov et al. [72] reported that incorporating metafounders in ssGBLUP models improved selection accuracy in both plant and animal breeding programs, particularly for traits with low heritability. Other reports demonstrated the advantages of metafounders in populations with complex or unknown pedigree structures [73–75]. However, in our study, except in the ssGBLUP model for individual selection, the inclusion of metafounders in both ssGBLUP and ABLUP models did not improve the prediction accuracy. These results agree with those of other studies, particularly in cases with well-connected populations or limited pedigree uncertainty [76–78].

The correlation between the full and partial datasets showed an improvement in stability for ssGBLUP compared to ABLUP, with considerable increases caused by the inclusion of genomic information. This trend is consistent with Berman et al. [49], who reported similar findings. In contrast, Callister et al. [78] found only a small improvement in stability between ABLUP and ssGBLUP. Including genomic information corrected the dispersion found for ABLUP models, particularly with the inclusion of MF. Similar results were obtained in other works [71,72,79], suggesting that MF could be an option for managing missing pedigree in ssGBLUP models. However, the inclusion of MF resulted in a significant increase in bias for all models, which could result from the relatively small number of genotyped individuals contributing to the estimation of gamma (γ), leading to a lack of connection between the genotyped individuals and the base population. Callister et al. [78] reported similar results when MF was included in the population. In contrast, in other studies, the bias was effectively eliminated with the inclusion of MF [71,79]. The study by Legarra et al. [26] underscores the importance of having sufficient representation within each metafounder group to ensure accurate estimation of the gamma matrix. Small or unbalanced groups can affect the stability and accuracy of relationship estimates, including those in the gamma matrix. In their 2024 study, Legarra et al. [80] further discuss the complexities of estimating gamma in highly unbalanced data, such as genotyped individuals far from base populations or many unknown parent groups within breeds. They propose maximum likelihood and pseudo-expectation-maximization methods to estimate gamma in these settings, emphasizing the need for careful consideration of group representation to avoid biased estimates. These results suggest that the metafounder approach needs to be tested with data that are more closely related to the base population and groups that have sufficient representation within each metafounder to improve the results.

After comparing different models with and without genomic information, including metafounders (MF) in both approaches, this study found that the best model to select parents for controlled matings and individuals for cloning is ssGBLUP without MF. The ssGBLUP accuracy for both validations (parents and individuals) ranged from 0.42 to 0.68, being lower for growth traits (TH and DBH) than for disease resistance (ADFO). These results confirm those obtained by Quezada et al. [40], encouraging the application of genomic prediction schemes for these traits in tree breeding programs. These results indicate that ssGBLUP models allow to predict BV with high accuracy, and thus to select parents for controlled matings and individuals for cloning with confidence, reducing the time and costs of implementing progeny trials.

4.2. Strategy of Selection

The adoption of GS in forest tree breeding allows for the shortening of breeding cycles, increased selection intensity, and improved accuracy of breeding values, all of which lead to increased genetic gains [81,82]. The findings from this study regarding growth and disease traits in *E. globulus* indicate moderate to high heritability for both traits. Therefore, the expected responses to selection for growth and disease resistance are favorable. Furthermore, the high age-age genetic correlations between ADFO at 14 months

and ADFO at 21 months demonstrate the effectiveness of early selection, allowing selection at 14 months to predict performance at 21 months without a significant decrease in genetic gain. This reduces evaluation costs, and the time required to obtain improved individuals. Early selection is a cornerstone of modern tree breeding programs, significantly improving their efficiency and effectiveness [1,8]. By identifying superior genotypes at juvenile stages, breeders can reduce generation intervals, increase selection intensity, and improve overall genetic gains [83,84].

Additionally, advancements in genomic selection have enhanced early selection methods by enabling accurate prediction of mature phenotypes from young plants, thus facilitating more precise and efficient breeding decisions [85]. The findings reported by this study suggest that two stages of selection are recommended in tree breeding programs aimed at selecting commercial clones. The first stage involves selecting the best parents for controlled matings. The second stage focuses on selecting the best progenies for cloning and their evaluation in clonal tests (CT), which evaluates many genotypes with few replicates, and amplified clonal testing (ACT), which evaluates fewer genotypes with more replicates (Figure 5). Including GS in both stages would reduce the breeding cycle by eliminating progeny testing and the first type of CT, thus reducing implementation costs and increasing genetic gains. The ssGBLUP model evaluated in this study would function at two levels: initially, by selecting the best parents in the first year of the forest genetic improvement cycle, and subsequently, by identifying the best individuals for cloning in the third year.

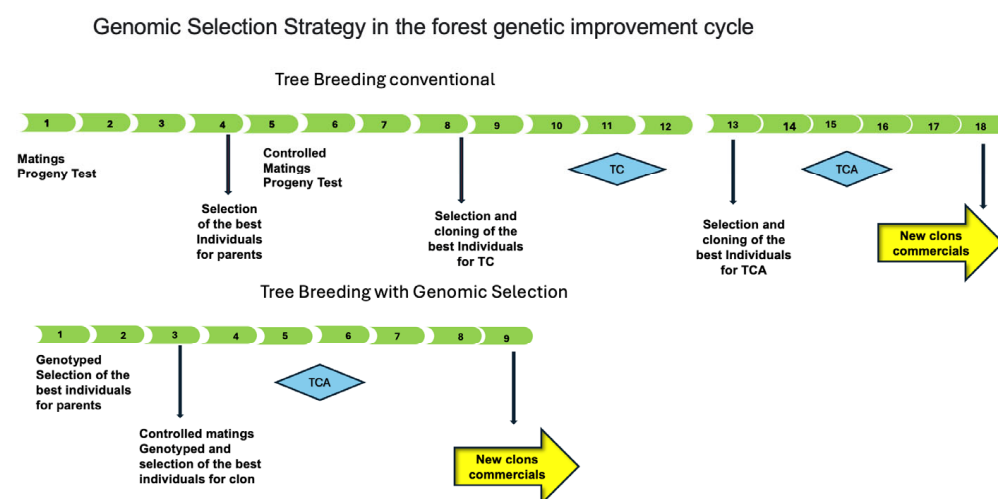


Figure 5. Conventional breeding cycle vs. breeding cycle with the inclusion of genomic selection for obtaining commercial clones of *Eucalyptus globulus*, the numbers correspond to the cycle years.

5. Conclusions

The implementation of GS demonstrated that it is a useful and straightforward tool for selection of candidates. While ABLUP relies exclusively on pedigree information and may overestimate additive genetic variance, ssGBLUP incorporates genomic relationships, leading to more accurate and less biased predictions of breeding values. However, the inclusion of metafounders (MF) increased bias, suggesting that MF should be tested using data more representative of the base population to achieve more reliable outcomes. Therefore, genomic selection (GS) can be effectively integrated at two key stages of the breeding cycle: the selection of superior parents for controlled matings, and the early selection of individuals in the nursery stage, potentially reducing the need for extensive field evaluation of selected clones.

Author Contributions: Conceptualization, M.G., I.A., M.Q. and G.B.; methodology, M.G., I.A., M.B., J.H., I.M. and D.L.; formal analysis, M.G.; writing—original draft preparation, M.G.; writing—review and editing, I.A., M.Q., M.B., D.L. and G.B.; supervision, I.A. and D.L. All authors have read and agreed to the published version of the manuscript.

Funding: The funding of this work came from INIA Uruguay (National Institute of Agricultural Research), project AMG-03 and ANII (National Association of Investigation and Innovation Research), project FMV_1_2021_166656.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available on request from the corresponding author due to privacy.

Conflicts of Interest: The authors declare no conflicts of interest.

References

- Grattapaglia, D.; Silva-Junior, O.B.; Resende, R.T.; Cappa, E.P.; Muller, B.S.F.; Tan, B.; Isik, F.; Ratcliffe, D.; El-Kassaby, Y.A. Quantitative genetics and genomics converge to accelerate forest tree breeding. *Front. Plant Sci.* **2018**, *9*, 1693. [\[CrossRef\]](#) [\[PubMed\]](#)
- El-Kassaby, Y.A.; Cappa, E.P.; Chen, C.; Ratcliffe, B.; Porth, I.M. Efficient genomics-based ‘end-to-end’ selective tree breeding framework. *Heredity* **2024**, *132*, 98–105. [\[CrossRef\]](#) [\[PubMed\]](#)
- Kang, K.S.; Bilir, N.; Jeon, K.; Kim, Y.J. Fertility Variation and Gene Diversity in Forest Populations. *Forests* **2023**, *14*, 2172. [\[CrossRef\]](#)
- El-Kassaby, Y.A.; Lstibůrek, M. Breeding without breeding. *Genet. Res.* **2009**, *91*, 111–120. [\[CrossRef\]](#)
- Klápšte, J.; Suontama, M.; Dungey, H.S.; Telfer, E.J.; Graham, N.J.; Low, C.B.; Stovold, G.T. Effect of Hidden Relatedness on Single-Step Genetic Evaluation in an Advanced Open-Pollinated Breeding Program. *J. Hered.* **2018**, *109*, 802–810. [\[CrossRef\]](#) [\[PubMed\]](#) [\[PubMed Central\]](#)
- Bouffier, L.; Debille, S.; Alazard, P.; Raffin, A.; Pastuszka, P.; Trontin, J.F. Pollen contamination and mating structure in maritime pine (*Pinus pinaster* Ait.) clonal seed orchards revealed by SNP markers. *Peer Community J.* **2023**, *3*, 68. [\[CrossRef\]](#)
- Grattapaglia, D. Twelve Years into Genomic Selection in Forest Trees: Climbing the Slope of Enlightenment of Marker Assisted Tree Breeding. *Forests* **2022**, *13*, 1554. [\[CrossRef\]](#)
- Resende, M.D.V.; Resende, M.F.R.; Sansaloni, C.P., Jr.; Petroli, C.D.; Missiaggia, A.A.; Aguiar, A.M.; Abad, J.M.; Takahashi, E.K.; Rosado, A.M.; Faria, D.A.; et al. Genomic selection for growth and wood quality in Eucalyptus: Capturing the missing heritability and accelerating breeding for complex traits in forest trees. *New Phytol.* **2012**, *194*, 116–128. [\[CrossRef\]](#)
- Park, M.N.; Alam, M.; Kim, S.; Park, B.; Lee, S.H.; Lee, S.S. Genomic selection through single-step genomic best linear unbiased prediction improves the accuracy of evaluation in Hanwoo cattle. *Asian-Australas. J. Anim. Sci.* **2020**, *33*, 1544–1557. [\[CrossRef\]](#) [\[PubMed\]](#) [\[PubMed Central\]](#)
- Curzon, A.Y.; Ezra, E.; Weller, J.I.; Seroussi, E.; Börner, V.; Gershoni, M. Single-step genomic BLUP (ssGBLUP) effectively models small cattle populations: Lessons from the Israeli-Holstein Herd-book. *BMC Genom.* **2024**, *25*, 1147. [\[CrossRef\]](#)
- Aguilar, I.; Misztal, I.; Johnson, D.; Legarra, A.; Tsuruta, S.; Lawlor, T. Hot topic: A unified approach to utilize phenotypic, full pedigree, and genomic information for genetic evaluation of Holstein final score. *J. Dairy Sci.* **2010**, *93*, 743–752. [\[CrossRef\]](#) [\[PubMed\]](#)
- Christensen, O.F.; Lund, M.S. Genomic prediction when some animals are not genotyped. *Genet. Sel. Evol.* **2010**, *42*, 2. [\[CrossRef\]](#) [\[PubMed\]](#)
- Sørensen, E.S.; Jansen, C.; Windju, S.; Crossa, J.; Sonesson, A.K.; Lillemo, M.; Alsheikh, M. Evaluation of strategies to optimize training populations for genomic prediction in oat (*Avena sativa*). *Plant Breed.* **2023**, *142*, 41–53. [\[CrossRef\]](#)
- Goddard, M.E.; Hayes, B.J. Mapping genes for complex traits in domestic animals and their use in breeding programmes. *Nat. Rev. Genet.* **2009**, *10*, 381–391. [\[CrossRef\]](#)
- Kullan, A.R.; van Dyk, M.M.; Hefer, C.A.; Jones, N.; Kanzler, A.; Myburg, A.A. Genetic dissection of growth, wood basic density and gene expression in interspecific backcrosses of *Eucalyptus grandis* and *E. urophylla*. *BMC Genet.* **2012**, *20*, 60. [\[CrossRef\]](#) [\[PubMed\]](#) [\[PubMed Central\]](#)
- Tan, B.; Grattapaglia, D.; Martins, G.S.; Ferreira, K.Z.; Sundberg, B.; Ingvarsson, P.K. 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. [\[CrossRef\]](#)
- Koivula, M.; Strandén, I.; Aamand, G.P.; Mäntysaari, E.A. Accounting for Missing Pedigree Information with Single-Step Random Regression Test-Day Models. *Agriculture* **2022**, *12*, 388. [\[CrossRef\]](#)

18. Christensen, O.F. Compatibility of pedigree-based and marker-based relationship matrices for single-step genetic evaluation. *Genet. Sel. Evol.* **2012**, *44*, 37. [\[CrossRef\]](#)
19. Thompson, R. Sire evaluation. *Biometrics* **1979**, *35*, 339–353. [\[CrossRef\]](#)
20. Quaas, R.L. Additive Genetic Model with Groups and Relationships. *J. Dairy Sci.* **1988**, *71* (Suppl. S2), 91–98. [\[CrossRef\]](#)
21. Karin, M. Impact of missing pedigrees in single-step genomic evaluation. *Anim. Prod. Sci.* **2021**, *61*, 1760–1773. [\[CrossRef\]](#)
22. Belay, T.K.; Eikje, L.S.; Gjuvsland, A.B.; Nordbø, Ø.; Tribout, T.; Meuwissen, T. Correcting for base-population differences and unknown parent groups in single-step genomic predictions of Norwegian Red cattle. *J. Anim. Sci.* **2022**, *100*, skac227. [\[CrossRef\]](#) [\[PubMed\]](#)
23. Patry, C.; Ducrocq, V. Evidence of biases in genetic evaluations due to genomic preselection in dairy cattle. *J. Dairy Sci.* **2011**, *94*, 1011–1020. [\[CrossRef\]](#) [\[PubMed\]](#)
24. VanRaden, P.M.; Tooker, M.E.; Cole, J.B.; Wiggans, G.R.; Megonigal, J.H., Jr. Genetic evaluations for mixed-breed populations. *J. Dairy Sci.* **2007**, *90*, 2434–2441. [\[CrossRef\]](#)
25. Strandén, I.; Aamand, G.P.; Mäntysaari, E.A. Single-step genomic BLUP with genetic groups and automatic adjustment for allele coding. *Genet. Sel. Evol.* **2022**, *54*, 38. [\[CrossRef\]](#)
26. Legarra, A.; Christensen, O.F.; Vitezica, Z.G.; Aguilar, I.; Misztal, I. Ancestral Relationships Using Metafounders: Finite Ancestral Populations and Across Population Relationships. *Genetics* **2015**, *200*, 455–468. [\[CrossRef\]](#)
27. Anglhuber, C.; Edel, C.; Pimentel, E.C.G.; Emmerling, R.; Götz, K.-U.; Thaller, G. Definition of metafounders based on population structure analysis. *Genet. Sel. Evol.* **2024**, *56*, 43. [\[CrossRef\]](#)
28. Himmelbauer, J.; Schwarzenbacher, H.; Fuerst, C.; Fuerst-Waltl, B. Exploring unknown parent groups and metafounders in single-step genomic best linear unbiased prediction: Insights from a simulated cattle population. *J. Dairy Sci.* **2024**, *107*, 8170–8192. [\[CrossRef\]](#)
29. Kudinov, E.A.; Mäntysaari, G.P.; Aamand, P.; Uimari, P.; Strandén, I. Metafounder approach for single-step genomic evaluations of Red Dairy cattle. *J. Dairy Sci.* **2020**, *103*, 6299–6310. [\[CrossRef\]](#)
30. Christensen, O.F.; Legarra, A.; Lund, M.S.; Su, G. Genetic evaluation for three-way crossbreeding. *Genet. Sel. Evol.* **2015**, *47*, 98. [\[CrossRef\]](#)
31. Arroja, L.; Dias Capela, I. The Role of *Eucalyptus Globulus* Forest and Products in Carbon Sequestration. *Clim. Change* **2006**, *74*, 123–140. [\[CrossRef\]](#)
32. Hunter, G.C.; Crous, P.W.; Carnegie, A.J.; Wingfield, M.J. *Teratosphaeria nubilosa*, a serious leaf disease pathogen of *Eucalyptus* spp. in native and introduced areas. *Mol. Plant Pathol.* **2009**, *10*, 1–14. [\[CrossRef\]](#) [\[PubMed\]](#)
33. Perez, G.; Slippers, B.; Wingfield, B.; Finkenauer, E.; Wingfield, M. Mycosphaerella leaf disease (MLD) outbreak on *Eucalyptus globulus* in Brazil caused by *Teratosphaeria* (*Mycosphaerella*) *nubilosa*. *Phytopathol. Mediterr.* **2009**, *48*, 302–306. Available online: <https://oajournals.fupress.net/index.php/pm/article/view/5293> (accessed on 27 February 2025).
34. Aylward, J.; Havenga, M.; Dreyer, L.L.; Roets, F.; Wingfield, B.D.; Pérez, C.A.; Ramírez-Berrutti, N.; Carnegie, A.J.; Wingfield, M.J. Genetic diversity of *Teratosphaeria pseudoeucalypti* in *Eucalyptus* plantations in Australia and Uruguay. *Australas. Plant Pathol.* **2021**, *50*, 639–649. [\[CrossRef\]](#)
35. Carnegie, A.J.; Ades, P.K. Mycosphaerella leaf disease reduces growth of plantation-grown *Eucalyptus globulus*. *Aust. For.* **2003**, *66*, 113–119. [\[CrossRef\]](#)
36. Hudson, C.J.; Freeman, J.S.; Jones, R.C.; Potts, B.M.; Wong, M.M.; Weller, J.L. Genetic control of heterochrony in *Eucalyptus globulus*. *G3 Genes Genomes Genet.* **2014**, *4*, 1235–1245. [\[CrossRef\]](#)
37. Pérez, S.; Renedo, C.; Ortiz, A.; Ortiz, F.; Tejedor, C. Strategies to Combat *Mycosphaerella* Leaf Disease in *Eucalyptus globulus* Plantations in Northern Spain. *Forests* **2016**, *7*, 190. [\[CrossRef\]](#)
38. Poland, J.; Rutkoski, J. Advances and Challenges in Genomic Selection for Disease Resistance. *Annu. Rev. Phytopathol.* **2016**, *54*, 79–98. [\[CrossRef\]](#)
39. Merrick, L.F.; Burke, A.B.; Chen, X.; Carter, A.H. Breeding With Major and Minor Genes: Genomic Selection for Quantitative Disease Resistance. *Front. Plant Sci.* **2021**, *12*, 713667. [\[CrossRef\]](#)
40. Quezada, M.; Aguilar, I.; Balmelli, G. Genomic breeding values prediction including populational selfing rate in an open-pollinated *Eucalyptus globulus* breeding population. *Tree Genet. Genomes* **2022**, *18*, 10. [\[CrossRef\]](#)
41. Silva-Junior, O.B.; Faria, D.A.; Grattapaglia, D. A flexible multi-species genome-wide 60K SNP chip developed from pooled resequencing of 240 *Eucalyptus* tree genomes across 12 species. *New Phytol.* **2015**, *206*, 1527–1540. [\[CrossRef\]](#) [\[PubMed\]](#)
42. Aguilar, I.; Tsuruta, S.; Masuda, Y.; Lourenco, D.A.L.; Legarra, A.; Misztal, I. BLUPF90 suite of programs for animal breeding with focus on genomics. In Proceedings of the World Congress on Genetics Applied to Livestock Production, Auckland, Australia, 13 February 2018.
43. Aguilar, I.; Misztal, I.; Tsuruta, S.; Legarra, A.; Wang, H. PREGSF90—POSTGSF90: Computational tools for the implementation of single-step genomic selection and genome-wide association with ungenotyped individuals in BLUPF90 programs. In Proceedings of the 10th World Congress on Genetics Applied to Livestock Production, Vancouver, BC, Canada, 17–22 August 2014.

44. VanRaden, P. Efficient methods to compute genomic predictions. *J. Dairy Sci.* **2008**, *91*, 4414–4423. [[CrossRef](#)] [[PubMed](#)]
45. Misztal, I.; Tsuruta, S.; Lourenco, D.; Aguilar, I.; Legarra, A.; Vitezica, Z. *Manual for BLUPF90 Family of Programs*; University of Georgia: Athens, GA, USA, 2014.
46. Misztal, I.; Tsuruta, S.; Strabel, T.; Auvray, B.; Druet, T.; Lee, D. BLUPF90 and related programs (BGF90). In Proceedings of the 7th World Congress on Genetics Applied to Livestock Production, CD-ROM Communication. Montpellier, France, 19–23 August 2002.
47. Wickham, H. *ggplot2: Elegant Graphics for Data Analysis*; Springer: Berlin/Heidelberg, Germany, 2016.
48. Legarra, A.; Reverter, A. Semi-parametric estimates of population accuracy and bias of predictions of breeding values and future phenotypes using the LR method. *Genet. Sel. Evol.* **2018**, *50*, 53. [[CrossRef](#)] [[PubMed](#)]
49. Bermann, M.; Legarra, A.; Hollifield, M.K.; Masuda, Y.; Lourenco, D.; Misztal, I. Validation of single-step GBLUP genomic predictions from threshold models using the linear regression method: An application in chicken mortality. *J. Anim. Breed. Genet.* **2021**, *138*, 4–13. [[CrossRef](#)] [[PubMed](#)]
50. Resende, M.F.R., Jr.; Muñoz Del Valle, P.R.; Acosta, J.J.; Resende, M.D.V.; Grattapaglia, D.; Kirst, M. Stability of Genomic Selection prediction models across ages and environments. *BMC Proc.* **2011**, *5* (Suppl. S7), O14. [[CrossRef](#)]
51. Lebedev, V.G.; Lebedeva, T.N.; Chernodubov, A.I.; Shestibratov, K.A. Genomic Selection for Forest Tree Improvement: Methods, Achievements and Perspectives. *Forests* **2020**, *11*, 1190. [[CrossRef](#)]
52. Klápště, J.; Ismael, A.; Paget, M.; Graham, N.J.; Stovold, G.T.; Dungey, H.S.; Slavov, G.T. Genomics-Enabled Management of Genetic Resources in Radiata Pine. *Forests* **2022**, *13*, 282. [[CrossRef](#)]
53. Alan, M.; Payn, K.; Mckeand, S.; Isik, F. Genetic Parameter Estimates from a Polymix Breeding Population of *Pinus taeda* L. *For. Sci.* **2023**, *69*, 415–423. [[CrossRef](#)]
54. Hayes, B.J.; Visscher, P.M.; Goddard, M.E. Increased accuracy of artificial selection by using the realized relationship matrix. *Genet. Res.* **2009**, *91*, 47–60. [[CrossRef](#)]
55. De los Campos, G.; Hickey, J.M.; Pong-Wong, R.; Daetwyler, H.D.; Calus, M.P. Whole-genome regression and prediction methods applied to plant and animal breeding. *Genetics* **2013**, *193*, 327–345. [[CrossRef](#)]
56. Hill, W.G.; Goddard, M.E.; Visscher, P.M. Data and theory point to mainly additive genetic variance for complex traits. *PLoS Genet.* **2008**, *4*, e1000008. [[CrossRef](#)] [[PubMed](#)]
57. Visscher, P.M.; Hill, W.G.; Wray, N.R. Heritability in the genomics era—concepts and misconceptions. *Nat. Rev. Genet.* **2008**, *9*, 255–266. [[CrossRef](#)] [[PubMed](#)]
58. Yang, J.; Benyamin, B.; McEvoy, B.P.; Gordon, S.; Henders, A.K.; Nyholt, D.R.; Madden, P.A.; Heath, A.C.; Martin, N.G.; Montgomery, G.W.; et al. Common SNPs explain a large proportion of the heritability for human height. *Nat. Genet.* **2010**, *42*, 565–569. [[CrossRef](#)] [[PubMed](#)]
59. Berro, I.; Lado, B.; Nalin, R.S.; Quincke, M.; Gutiérrez, L. Training Population Optimization for Genomic Selection. *Plant Genome* **2019**, *12*, 190028. [[CrossRef](#)]
60. 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*, 343–352. [[CrossRef](#)]
61. Ukrainetz, N.K.; Mansfield, S.D. Prediction accuracy of single-step BLUP for growth and wood quality traits in the lodgepole pine breeding program in British Columbia. *Tree Genet. Genomes* **2020**, *16*, 64. [[CrossRef](#)]
62. Chen, Z.-Q.; Baisou, J.; Pan, J.; Karlsson, B.; Andersson, B.; Westin, J.; García-Gil, M.R.; Wu, H.X. 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.* **2018**, *19*, 946. [[CrossRef](#)]
63. Ousmael, K.M.; Cappa, E.P.; Hansen, J.K.; Hendre, P.; Hansen, O.K. Genomic evaluation for breeding and genetic management in *Cordia africana*, a multipurpose tropical tree species. *BMC Genom.* **2024**, *25*, 9. [[CrossRef](#)]
64. Cappa, E.P.; de Lima, B.M.; da Silva-Junior, O.B.; Garcia, C.C.; Mansfield, S.D.; Grattapaglia, D. Improving genomic prediction of growth and wood traits in *Eucalyptus* using phenotypes from non-genotyped trees by single-step GBLUP. *Plant Sci.* **2019**, *284*, 9–15. [[CrossRef](#)]
65. Goddard, M.E.; Hayes, B.J. Genomic selection. *J. Anim. Breed. Genet.* **2007**, *124*, 323–330. [[CrossRef](#)]
66. Goddard, M.E.; Wray, N.R.; Verbyla, K.; Visscher, P.M. Estimating effects and making predictions from genome-wide marker data. *Stat. Sci.* **2010**, *24*, 517–529. [[CrossRef](#)]
67. Meuwissen, T.H.; Hayes, B.J.; Goddard, M. Prediction of Total Genetic Value Using Genome-Wide Dense Marker Maps. *Genetics* **2001**, *157*, 1819–1829. [[CrossRef](#)]
68. Daetwyler, H.D.; Pong-Wong, R.; Villanueva, B.; Woolliams, J.A. The impact of genetic architecture on genome-wide evaluation methods. *Genetics* **2010**, *185*, 1021–1031. [[CrossRef](#)] [[PubMed](#)]
69. Galeano, E.; Cappa, E.P.; Bousquet, J.; Thomas, B.R. Optimizing a Regional White Spruce Tree Improvement Program: SNP Genotyping for Enhanced Breeding Values, Genetic Diversity Assessment, and Estimation of Pollen Contamination. *Forests* **2023**, *14*, 2212. [[CrossRef](#)]

70. Strandén, I.; Matilainen, K.; Aamand, G.P.; Mäntysaari, E.A. Solving efficiently large single-step genomic best linear unbiased prediction models. *J. Anim. Breed. Genet.* **2017**, *134*, 264–274. [[CrossRef](#)] [[PubMed](#)]
71. Macedo, F.L.; Astruc, J.M.; Meuwissen, T.H.E.; Legarra, A. Removing data and using metafounders alleviates biases for all traits in Lacaune dairy sheep predictions. *J. Dairy Sci.* **2022**, *105*, 2439–2452. [[CrossRef](#)]
72. Kudinov, A.A.; Koivula, M.; Aamand, G.P.; Strandén, I.; Mäntysaari, E.A. Single-step genomic BLUP with many metafounders. *Front. Genet.* **2022**, *13*, 1012205. [[CrossRef](#)]
73. Van Grevenhof, E.M.; Vandenplas, J.; Calus, M.P.L. Genomic prediction for crossbred performance using metafounders. *J. Anim. Sci.* **2019**, *97*, 548–558. [[CrossRef](#)]
74. Xiang, T.; Christensen, O.F.; Legarra, A. Technical note: Genomic evaluation for crossbred performance in a single-step approach with metafounders. *J. Anim. Sci.* **2017**, *95*, 1472–1480. [[CrossRef](#)]
75. Meyer, K.; Tier, B.; Swan, A. Estimates of genetic trend for single-step genomic evaluations. *Genet. Sel. Evol.* **2018**, *50*, 39. [[CrossRef](#)]
76. Granado-Tajada, I.; Legarra, A.; Ugarte, E. Exploring the inclusion of genomic information and metafounders in Latxa dairy sheep genetic evaluations. *J. Dairy Sci.* **2020**, *103*, 6346–6353. [[CrossRef](#)]
77. Fu, C.; Ostensen, T.; Christensen, O.F.; Xiang, T. Single-step genomic evaluation with metafounders for feed conversion ratio and average daily gain in Danish Landrace and Yorkshire pigs. *Genet. Sel. Evol.* **2021**, *53*, 79. [[CrossRef](#)] [[PubMed](#)]
78. Callister, A.; Bermann, M.; Elms, S.; Bradshaw, B.; Lourenco, D.; Brawner, J. Accounting for population structure in genomic predictions of *Eucalyptus globulus*. *G3 Genes | Genomes | Genet.* **2022**, *12*, jkac180. [[CrossRef](#)] [[PubMed](#)]
79. Himmelbauer, J.; Schwarzenbacher, H.; Fuerst, C.; Fuerst-Waltl, B. Investigation On the Metafounder Concept in ssGBLUP Based On a Simulated Cattle Population. *Interbull Bull.* **2023**, *59*, 124–131.
80. Legarra, A.; Bermann, M.; Mei, Q.; Christensen, O.F. Estimating genomic relationships of metafounders across and within breeds using maximum likelihood, pseudo-expectation-maximization maximum likelihood and increase of relationships. *Genet. Sel. Evol.* **2024**, *56*, 35. [[CrossRef](#)] [[PubMed](#)] [[PubMed Central](#)]
81. Alves, F.C.; Balmant, K.M.; Resende, M.F.R.; Kirst, M.; de los Campos, G. Accelerating forest tree breeding by integrating genomic selection and greenhouse phenotyping. *Plant Genome* **2020**, *13*, e20048. [[CrossRef](#)]
82. Duarte, D.; Jurcic, E.J.; Dutour, J.; Villalba, P.V.; Centurión, C.; Grattapaglia, D.; Cappa, E.P. Genomic selection in forest trees comes to life: Unraveling its potential in an advanced four-generation *Eucalyptus grandis* population. *Front. Plant Sci.* **2024**, *15*, 1462285. [[CrossRef](#)]
83. Fries, A.; Ericsson, T. Genetic parameters for early wood and latewood densities and development with increasing age in Scots pine. *Ann. For. Sci.* **2011**, *66*, 404. [[CrossRef](#)]
84. Lee, K.; Oh, C.; Kim, I.S. Genetic parameter changes and age–age correlations in *Pinus koraiensis* growth over 40-year progeny testing. *BMC Plant Biol.* **2024**, *24*, 86. [[CrossRef](#)]
85. Souza, L.M.; Francisco, F.R.; Gonçalves, P.S.; Scaloppi Junior, E.J.; Le Guen, V.; Fritsche-Neto, R.; Souza, A.P. Genomic Selection in Rubber Tree Breeding: A Comparison of Models and Methods for Managing G×E Interactions. *Front. Plant Sci.* **2019**, *10*, 1353. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.