

Accounting for population structure in genomic predictions of *Eucalyptus globulus*

Andrew N. Callister,^{1*} Matias Bermann,² Stephen Elms,³ Ben P. Bradshaw,⁴ Daniela Lourenco,² Jeremy T. Brawner⁵

¹Treehouse Forest Research LLC, Check, VA 24072, USA,

²Department of Animal and Dairy Science, University of Georgia, Athens, GA 30602, USA,

³HVP Plantations, Churchill, VIC 3842, Australia,

⁴Australian Bluegum Plantations, Albany, WA 6330, Australia,

⁵Department of Plant Pathology, University of Florida, Gainesville, FL 32611, USA

*Corresponding author: Treehouse Forest Research LLC, 249 Level Bottom Road, Check, VA 24072, USA. Email: andrew@treehouse-research.com

Abstract

Genetic groups have been widely adopted in tree breeding to account for provenance effects within pedigree-derived relationship matrices. However, provenances or genetic groups have not yet been incorporated into single-step genomic BLUP (“HBLUP”) analyses of tree populations. To quantify the impact of accounting for population structure in *Eucalyptus globulus*, we used HBLUP to compare breeding value predictions from models excluding base population effects and models including either fixed genetic groups or the marker-derived proxies, also known as metafounders. Full-sib families from 2 separate breeding populations were evaluated across 13 sites in the “Green Triangle” region of Australia. Gamma matrices (Γ) describing similarities among metafounders reflected the geographic distribution of populations and the origins of 2 land races were identified. Diagonal elements of Γ provided population diversity or allelic covariation estimates between 0.24 and 0.56. Genetic group solutions were strongly correlated with metafounder solutions across models and metafounder effects influenced the genetic solutions of base population parents. The accuracy, stability, dispersion, and bias of model solutions were compared using the linear regression method. Addition of genomic information increased accuracy from 0.41 to 0.47 and stability from 0.68 to 0.71, while increasing bias slightly. Dispersion was within 0.10 of the ideal value (1.0) for all models. Although inclusion of metafounders did not strongly affect accuracy or stability and had mixed effects on bias, we nevertheless recommend the incorporation of metafounders in prediction models to represent the hierarchical genetic population structure of recently domesticated populations.

Keywords: breeding value accuracy, cross-validation, forest tree breeding, genetic groups, genomic selection, LR method, metafounders, Myrtaceae, single-step GBLUP; MPP; Multiparental Populations; Multiparent Advanced Generation Inter-Cross (MAGIC)

Introduction

Provenance variation in growth and commercial value is usually characterized at an early stage in the domestication of forest trees and can play an important role in subsequent generations of breeding (White *et al.* 2007). Forest geneticists have widely adopted genetic groups to account for population differences in linear mixed models (LMM) (e.g. McRae *et al.* 2004; Brawner *et al.* 2010; Callister *et al.* 2011; Klápště *et al.* 2019), typically as unknown parent groups (UPGs) that were developed to account for missing parents in livestock populations (Quaas 1988; Westell *et al.* 1988). This approach treats provenances as fixed effects, and it results in more conservative estimates of heritability relative to treating founders as unrelated individuals (Callister *et al.* 2021).

Widespread adoption of genomics in breeding programs has prompted the development of the single-step genomic BLUP method, which merges empirical relationship estimates from a genotyped subset (\mathbf{G} matrix) with the pedigree-derived relationship coefficients (\mathbf{A}) for the entire population (Legarra *et al.* 2009; Aguilar *et al.* 2010; Christensen and Lund 2010). The resulting

inverse relationship matrix (\mathbf{H}^{-1}) is then used in place of \mathbf{A}^{-1} in LMM to produce (genomic) estimated breeding values ((G)EBV). Although there have been a number of studies applying this “HBLUP” approach to forest tree populations (e.g. Ratcliffe *et al.* 2017; Klápště *et al.* 2018; Thavamanikumar *et al.* 2020; Ukrainetz and Mansfield 2020; Callister *et al.* 2021; Jurcic *et al.* 2021), none have yet considered multiple base populations such as provenances.

UPGs have been incorporated into livestock HBLUP analyses through various modifications of \mathbf{H}^{-1} that extend the method of Quaas (1988) and Westell *et al.* (1988) (Misztal *et al.* 2013; Tsuruta *et al.* 2019; Masuda *et al.* 2021). A similar approach is to constitute genetic groups as “metafounders” (MFs), which are pseudo-individuals acting as proxies for UPGs in base populations (Legarra *et al.* 2015; Masuda *et al.* 2022). Genetic diversity within, and relationships among MFs, are computed using the genotypes of their descendants. MF models have performed favorably in both simulated and actual livestock populations (Bradford *et al.* 2019; Kudinov *et al.* 2020; Macedo, Christensen, *et al.* 2020; Masuda *et al.* 2021). The MF approach is of particular interest for

Received: March 28, 2022. Accepted: June 29, 2022

© The Author(s) 2022. Published by Oxford University Press on behalf of Genetics Society of America.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

representing provenances of forest populations because it may estimate biologically meaningful relationships among geographically or genetically defined races that arise from evolutionary processes. MFs may also reveal cryptic relationships between these races and infusions from external populations.

Predictions from quantitative genetic models can be assessed against 3 important parameters: accuracy, bias, and dispersion. Accuracy is a measure of the reliability of EBVs and is used to predict the response to selection (Bijma 2012). Bias is the difference between means of EBVs and true breeding values (TBVs), whereas dispersion is calculated as the slope of the regression of TBVs on EBVs and compares the scale or variation of EBVs against TBVs. Dispersion values less than 1 indicate over-dispersion and inflation of EBV relative to TBV, whereas values greater than 1 indicate under-dispersion of EBV. Bias and dispersion influence expected genetic gain, with potentially significant impacts on selection decisions (Macedo, Reverter, et al. 2020). Legarra and Reverter (2018) introduced the linear regression (LR) method to evaluate accuracy, bias, and dispersion using a cross-validation approach.

Eucalyptus globulus Labill. is a commercially important plantation species in Mediterranean climates around the world, and it is particularly favored for pulpwood production. Its native range in Southeast Australia has been classified into 13 races and 20 subraces using morphological and molecular markers (Dutkowski and Potts 1999; Potts et al. 2004). Landraces have also developed where *E. globulus* has been naturalized in areas such as California, Portugal, Spain, and Chile. Genetic associations among races have been shown to follow geographical patterns (Costa et al. 2017), with Portuguese and Californian landraces shown to be most closely related to the east Tasmanian races (Freeman et al. 2007; Costa et al. 2017; Yost et al. 2021).

Callister et al. (2021) used HBLUP to conduct a joint analysis of 2 disconnected *E. globulus* breeding populations that were linked using marker-derived pedigree. Relationships within the resulting **H** matrix tended to be significantly positive among population founders from the same population, while across-race relationships among founders ranged from significantly positive to significantly negative. The authors suggested that true average relationships within races were probably substantially larger than those estimated in **H** and that accommodating ancestral race effects in HBLUP should be prioritized. The goals for the present study extend from this recent work (Callister et al. 2021) using a subset of their data. Our goals were to (1) fit models including MF and examine relationships corresponding to race, landrace, and plus tree (infusion) populations and (2) use the LR approach (Legarra and Reverter 2018) to evaluate goodness of fit for genetic models with and without genomic relationship information, with genetic groups excluded or incorporated as UPGs, or as MFs.

Materials and methods

Experimental populations

This study used a subset of the *E. globulus* phenotype data analyzed by Callister et al. (2021). Full-sib family data for 13 progeny tests in the “Green Triangle” region of Australia were provided by 2 separate multigenerational commercial tree improvement programs, EG1 (Australian Bluegum Plantations) and EG2 (HVP Plantations) (Fig. 1).

Pedigree founders or base population parents were allocated to 14 genetic groups, which consisted of 9 races (following Dutkowski and Potts 1999), 2 landraces (Portugal and California),

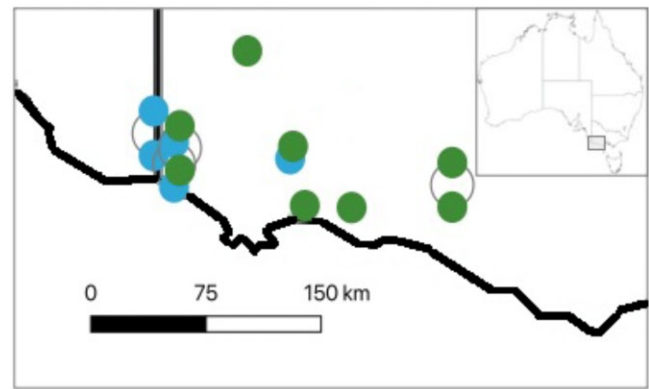


Fig. 1. Location of 5 progeny trials from the EG1 program (blue symbols) and 8 progeny trials from the EG2 program (green symbols) used for this study. Inset provides context.

and 3 plus tree populations (see Supplementary Table 1). Plus tree populations 1 and 2 corresponded to external sources of improved populations from which phenotypic selections were made. Miscellaneous unpedigreed individuals from unknown origins or external sources that lacked sufficient representation to constitute their own groups were allocated to plus tree population 3. The pedigree contained 35,533 progeny (35,247 of which were phenotyped), 368 parents of progeny, 175 progenitors of those parents, and 172 other seed orchard individuals that were both related and unrelated to the tested progeny (see Supplementary Table 2). Progeny represented 650 full-sib families from EG1 and 592 full-sib families from EG2.

HBLUP was necessary to analyze joint-program data due to the inadequacy of inter-program pedigree relationships. Callister et al. (2021) reported 2,071 inter-program relationship coefficients greater than 0.1, 428 greater than 0.2, and 70 greater than 0.3 in a genomic relationship matrix (**G**) using a subset of the genotyped individuals from the present analysis, which provides ample connectivity for joint HBLUP models.

Field trials were primarily established in randomized incomplete-block designs. Four to eight replications of each family were established in contiguous row-plots of 4–5 trees. One EG1 trial was established as a single-tree plot design. Incomplete blocks were arranged into replicates that were generally contiguous, enabling the resolution of spatial trends within each trial.

Stem volume calculation

Diameter at breast height (DBH) for each tree was measured with a tape at 3 years (4 sites), 4 years (3 sites), 5 years (5 sites), or 8 years (1 site) after planting. Tree height (HT) was measured with a hypsometer for each tree in the 5 EG1 trials, and for 19% of trees in EG2 trials. In these cases, unmeasured HT data were predicted from DBH-HT relationships established among measured trees for each trial. Stem volume (VOL) was calculated for each tree using DBH, HT (measured or predicted), and the volume function provided by each breeding program. DBH of all stems forking below breast height was measured in ten trials and in these cases, VOL was aggregated to the tree level before analysis.

Genotyping

Leaves were sampled from 148 parents and 89 others in the EG1 program and from 80 parents, 958 progeny, and 32 others in the EG2 program, where “others” refers to individuals that are not parents or tested progeny. DNA extraction and genotyping were conducted by Gondwana Genomics Pty Ltd, Canberra

(Thavamanikumar et al. 2020). The *E. globulus* marker panel consisted of 2,568 single-nucleotide polymorphism (SNP) and small biallelic insertion/deletion (INDEL) markers within candidate genes identified in previous studies (Southerton et al. 2011). Missing values were imputed using LinkImpute (Money et al. 2015). Pedigree errors were identified and corrected within the genotyped population using the ASRgenomics package (Gezan et al. 2021) in R version 4.1.0 (R Core Team 2021).

Statistical analyses

We conducted 2 sets of analyses, each including various models to compare methods of incorporating genetic groups into analyses. A set of *single-population* analyses was completed using the EG2 population to contrast pedigree-based BLUP (ABLUP) with HBLUP models. A set of *joint-population* HBLUP analyses were completed using information from both programs, with a joint \mathbf{H} for unification.

Calculation of $\mathbf{\Gamma}$

The relationship matrix across MF, $\mathbf{\Gamma}$, is a function of genetic similarity across populations following $\Gamma_{ij} = 8\text{cov}(p_i, p_j)$, where p_i represents allele frequencies in population i (Garcia-Baccino et al. 2017). We calculated $\mathbf{\Gamma}$ using the gammaf90 program of the BLUPF90 software suite (Misztal et al. 2014), which estimates allele frequencies, and thus $\hat{\Gamma}_{ij} = 8\text{cov}(\hat{p}_i, \hat{p}_j)$ using the GLS method of Garcia-Baccino et al. (2017). The $\mathbf{\Gamma}$ used for *single population* analyses was calculated across 9 genetic groups represented in the EG2 program (8 races and 1 landrace; see Supplementary Table 1). $\mathbf{\Gamma}$ for the *joint population* analyses was calculated across all 14 genetic groups.

Relationship matrix calculation

The $\mathbf{\Gamma}$ -augmented inverse of the pedigree relationship matrix was calculated as:

$$\mathbf{A}_\Gamma^{-1} = \begin{bmatrix} \mathbf{A}_\Gamma^{11} & \mathbf{A}_\Gamma^{12} & \mathbf{A}_\Gamma^{1m} \\ \mathbf{A}_\Gamma^{21} & \mathbf{A}_\Gamma^{22} & \mathbf{A}_\Gamma^{2m} \\ \mathbf{A}_\Gamma^{m1} & \mathbf{A}_\Gamma^{m2} & \mathbf{A}_\Gamma^{mm} + \mathbf{\Gamma}^{-1} \end{bmatrix}, \quad (1)$$

where the superscript “1” represents ungenotyped individuals, “2” represents genotyped individuals, and “m” represents MF.

Without consideration of genetic groups, the inverse of \mathbf{H} was calculated as:

$$\mathbf{H}^{-1} = \mathbf{A}^{-1} + \begin{bmatrix} 0 & 0 \\ 0 & \mathbf{G}^{-1} - \mathbf{A}_{22}^{-1} \end{bmatrix}. \quad (2)$$

Following Christensen et al. (2012) and Aguilar et al. (2010), \mathbf{G} was blended as $0.95\mathbf{G}_{\text{scaled}} + 0.05\mathbf{A}_{22}$, where $\mathbf{G}_{\text{scaled}}$ was a rescaled version of \mathbf{G}_{obs} , originally computed from observed allele frequencies by the first method of VanRaden (2008). For HBLUP with MFs, the $\mathbf{\Gamma}$ -augmented \mathbf{H}^{-1} (\mathbf{H}_Γ^{-1}) was calculated as:

$$\mathbf{H}_\Gamma^{-1} = \mathbf{A}_\Gamma^{-1} + \begin{bmatrix} 0 & 0 & 0 \\ 0 & \mathbf{G}_{05}^{-1} - \mathbf{A}_{\Gamma 22}^{-1} & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad (3)$$

where $\mathbf{A}_{\Gamma 22}^{-1}$ is the inverse of $\mathbf{\Gamma}$ -augmented pedigree relationship matrix (\mathbf{A}_Γ) for genotyped individuals and \mathbf{G}_{05} was calculated in the same manner as \mathbf{G} , except that the initial computation of the relationship matrix was based on 0.5 allele frequencies (Legarra et al. 2015).

To examine relationships among genotyped individuals of the *single program*, \mathbf{A}_{22} and $\mathbf{A}_{\Gamma 22}$ were calculated from versions of \mathbf{A}^{-1}

and \mathbf{A}_Γ^{-1} produced with a subset of the pedigree. \mathbf{H}_Γ^{-1} was also produced with a *single program* pedigree subset small enough for convenient inversion and inspection.

Model fitting

Stem volume data were first analyzed at the individual site level using ASReml-R version 4 (Butler et al. 2018) in R version 4.1.0 to adjust the data for any trial design effects (incomplete block, row, column, or plot) and spatial trend (AR1xAR1), represented as random effects. The adjusted data were then standardized by trial to a mean of zero and phenotypic standard deviation of 1.

All cross-site models were fitted using BLUPF90 programs and the general LMM framework:

$$\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{Z}_1\mathbf{a} + \mathbf{Z}_2\mathbf{f} + \mathbf{e}, \quad (4)$$

where \mathbf{y} is the vector of phenotypic values, \mathbf{b} is a vector of fixed effects for sites, \mathbf{a} is the vector of random additive genetic effects with $E(\mathbf{a}) = 0$, \mathbf{f} is a vector of random family-specific effects with $E(\mathbf{f}) = 0$ and $\text{var}(\mathbf{f}) = \mathbf{I}\sigma_f^2$, and \mathbf{e} is the vector of residual effects with $E(\mathbf{e}) = 0$ and $\text{var}(\mathbf{e}) = \mathbf{I}\sigma_e^2$. \mathbf{X} , \mathbf{Z}_1 , and \mathbf{Z}_2 are incidence matrices relating phenotypic records in \mathbf{y} to effects in vectors \mathbf{b} , \mathbf{a} , and \mathbf{f} .

Five models were fitted to the *single population* which differed only in the specification of $\text{var}(\mathbf{a})$. Three pedigree models were fitted: without groups (ABLUP), with fixed genetic groups (ABLUP_UPG), and with MFs (ABLUP_MF). For ABLUP, $\text{var}(\mathbf{a}) = \mathbf{A}\sigma_a^2$. For ABLUP_UPG, $\text{var}(\mathbf{a}) = \mathbf{A}^*\sigma_a^2$, where \mathbf{A}^* is adjusted to include fixed group effects (Quaas 1988; Westell et al. 1988). For ABLUP_MF, $\text{var}(\mathbf{a}) = \mathbf{A}_\Gamma\sigma_a^2$, where \mathbf{A}_Γ was calculated using Equation 1. Two HBLUP models were fitted to the *single population* and *joint population*. Model “HBLUP” without groups assumes $\text{var}(\mathbf{a}) = \mathbf{H}\sigma_a^2$, whereas HBLUP models with MFs (HBLUP_MF) were fitted assuming $\text{var}(\mathbf{a}) = \mathbf{H}_\Gamma\sigma_a^2$.

Variances were specified for all random effects at values estimated from ABLUP for the *single population* analyses and from HBLUP for the *joint population* analyses. Additive variance for MF models was divided by a scalar $k = 1 + -\text{diag}(\mathbf{\Gamma})/2 - \mathbf{\Gamma}$ (Legarra et al. 2015), which was 1.010 for the *single population* and 0.987 for the *joint population*.

Evaluation of model performance

Model performance was evaluated using the LR method (Legarra and Reverter 2018). This approach involves removing phenotypes from a focal group to create a partial dataset, which is analyzed with the same model specification as the complete phenotypic dataset. Solutions from analysis of the whole dataset are then regressed against solutions obtained from the partial dataset to provide model validation statistics. We formed focal groups consisting of 5,400 progeny from the *single population* (all progeny of 20 parents representing a diversity of origins) and 8,367 progeny from the *joint population* (all progeny of 60 parents representing a diversity of origins and both programs). Legarra and Reverter (2018) showed that regressions may be formed among solutions for either the retained individuals or the focal individuals. We used the focal group as our goal was to determine the best model for predicting breeding values of unphenotyped individuals in the pedigree. For each model specification, prediction accuracy (ACC)

was calculated as $\sqrt{\frac{\text{cov}(\hat{a}_w, \hat{a}_p)}{(1-F)\sigma_a^2}}$ and stability (STAB) as the correlation $\frac{\text{cov}(\hat{a}_w, \hat{a}_p)}{\sqrt{\text{var}(\hat{a}_w)\text{var}(\hat{a}_p)}}$, where $\text{cov}(\hat{a}_w, \hat{a}_p)$ is the covariance between focal group solutions from the whole and partial datasets, $-F$ is average inbreeding coefficient of the focal group calculated from

the **H** matrix without groups using the INBUPGF90 program (Misztal et al. 2014), and $\hat{\sigma}_a^2$ is the REML estimate of the additive genetic variance. Dispersion (DISP) was calculated as the regression slope: $\text{cov}(\hat{a}_w, \hat{a}_p) / \text{var}(\hat{a}_p)$. BIAS was calculated as $-\hat{a}_p - -\hat{a}_w$, where $-\hat{a}_p$ and $-\hat{a}_w$ are mean focal group solutions from analysis of the partial and whole datasets, respectively.

Results

Relationship matrices

The **Γ** estimate for the joint population was:

	Port	Calif	W.Otw	E.Otw	Strz	Furnx	NE.Tas	SE.Tas	S.Tas	W.Tas	KI	P.T.P.1	P.T.P.2	P.T.P.3	
Γ =	0.25	0.19	0.14	0.15	0.15	0.14	0.19	0.23	0.22	0.17	0.14	0.15	0.16	0.17	Port
		0.38	0.13	0.15	0.13	0.12	0.18	0.21	0.21	0.14	0.15	0.14	0.15	0.17	Calif
			0.30	0.21	0.21	0.14	0.13	0.15	0.18	0.17	0.22	0.19	0.16	0.20	W.Otw
				0.37	0.21	0.16	0.15	0.16	0.19	0.15	0.19	0.19	0.18	0.18	E.Otw
					0.29	0.15	0.14	0.17	0.17	0.16	0.17	0.21	0.16	0.18	Strz
						0.24	0.16	0.14	0.16	0.13	0.14	0.17	0.19	0.15	Furnx
							0.40	0.23	0.22	0.17	0.11	0.15	0.17	0.15	NE.Tas
								0.34	0.26	0.18	0.15	0.17	0.18	0.19	SE.Tas
									0.46	0.20	0.17	0.18	0.19	0.19	S.Tas
										0.56	0.21	0.17	0.16	0.20	W.Tas
											0.41	0.17	0.17	0.23	KI
												0.26	0.19	0.18	PlusTreePop1
													0.26	0.18	PlusTreePop2
														0.25	PlusTreePop3

Diagonal elements of **Γ** ranged from 0.24 to 0.56 with a mean of 0.34, which corresponds to a mean inbreeding coefficient of -0.64 within MF. Plus-Tree Populations were more genetically diverse than the average MF, with Γ_{ij} estimates lower than the average of the diagonal elements. Off-diagonal elements of **Γ** provide estimates of relatedness among populations that are derived from correlations of allele frequencies. **Γ** showed there are closer affinities amongst East Otways, West Otways, and Strzelecki races from the mainland of Australia, and King Island population was most similar to the West Otways population. Closer affinities between the 3 east-coast Tasmanian races were evident in **Γ**, with the greatest between-population relationship coefficient found between South-eastern Tasmania and Southern Tasmania ($\Gamma_{ij} = 0.26$). A UPGMA phylogenetic tree representing the among-race relationships in **Γ** demonstrates these 2 clusters, while associating Western Tasmania loosely with east Tasmanian races and showing that Furneaux is the most distantly related race (Fig. 2). The Portuguese and Californian landraces were most closely related to South-eastern Tasmania and Southern Tasmania in **Γ**. Plus-Tree Population 1 was most closely related to the Strzelecki race ($\Gamma_{ij} = 0.21$) and Plus-Tree Population 2 was not strongly associated with any particular race in **Γ**, while Plus-Tree Population 3 was most closely aligned with King Island ($\Gamma_{ij} = 0.23$).

In contrast with the HBLUP method, which modifies the genomic relationship matrix to be compatible with **A**, the HBLUP_MF method modifies **A** to be compatible with **G**₀₅. Table 1 shows correlations among elements of the **A**₂₂, **A**_{Γ22}, **G**_{obs}, and **G**₀₅ matrices in the EG2 population. Correlations between diagonal and off-diagonal elements of **A**₂₂ and **G**_{obs} were 0.08 and 0.85, respectively (Table 1). Inclusion of MF relationships in **A** increased the similarity of pedigree and genomic relationship matrices. The correlations between diagonal and off-diagonal elements of **A**_{Γ22} and **G**₀₅ were 0.81 and 0.87, respectively (Table 1).

As expected, **G**_{obs}, calculated with observed allele frequencies, displayed a similar mean to **A**₂₂, but with greater variation (Table 2). **A**_{Γ22} had slightly higher diagonal values than **A**₂₂, as inbreeding values were translated from **Γ** throughout the pedigree and each founder's inbreeding value in **A**_Γ incorporates half of Γ_{ii} from its respective MF. The mean off-diagonal element of **A**_{Γ22} was substantially greater than that of **A**₂₂ (Table 2), as relationships within and among MFs in **Γ** were distributed throughout **A**_Γ. For example, founders with a common MF that were unrelated in **A** were related in **A**_Γ by Γ_{ii} , the diagonal value corresponding to their common MF. Similarly, founders from different

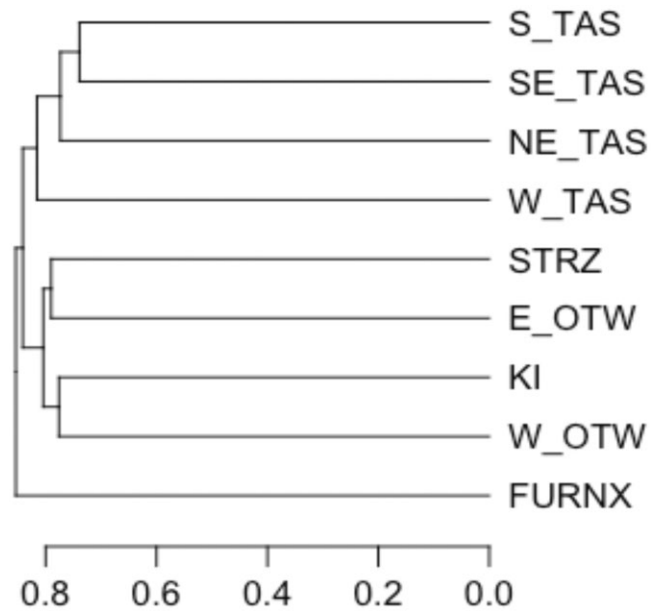


Fig. 2. UPGMA plot of similarity among *E. globulus* races based on **Γ**. See Supplementary Table 1 for key to race names.

Table 1. Correlations among diagonal (upper triangle) and off-diagonal (lower triangle) elements of **A**₂₂, **A**_{Γ22}, **G**_{obs}, and **G**₀₅.

	A ₂₂	A _{Γ22}	G _{obs}	G ₀₅
A ₂₂	1	0.62	0.08	0.14
A _{Γ22}	0.96	1	0.89	0.81
G _{obs}	0.85	0.30	1	0.30
G ₀₅	0.77	0.87	0.91	1

Table 2. Mean, minimum, and maximum element values of \mathbf{A}_{22} , $\mathbf{A}_{\Gamma 22}$, \mathbf{G}_{obs} , and \mathbf{G}_{05} on the diagonal and off-diagonal.

Element	Matrix	Mean	Min	Max
Diagonal	\mathbf{A}_{22}	1.00	1.00	1.25
	$\mathbf{A}_{\Gamma 22}$	1.07	1.04	1.35
	\mathbf{G}_{obs}	1.04	0.78	1.38
	\mathbf{G}_{05}	1.34	1.13	1.71
Off-diagonal	\mathbf{A}_{22}	0.03	0.00	0.75
	$\mathbf{A}_{\Gamma 22}$	0.19	0.08	0.92
	\mathbf{G}_{obs}	0.00	-0.24	0.85
	\mathbf{G}_{05}	0.61	0.40	1.22

MF that were unrelated in \mathbf{A} were related in \mathbf{A}_{Γ} by Γ_{ij} , the off-diagonal value corresponding to the relationship between their respective MFs. Relationships at the founder level therefore influence relationships among all parents and progeny in \mathbf{A}_{Γ} . Values in \mathbf{G}_{05} were considerably greater than all other relationship matrices among genotyped individuals (Table 2). Diagonal elements of the \mathbf{H}_{Γ} matrix representing *single program* parents and progenitors were consistently greater than the corresponding elements of \mathbf{A} with UPG, ranging from 1.10 to 1.67, with a mean of 1.41.

Model results

Additive variance estimates were similar between the *single population* and *joint population* (around 0.17; Table 3), corresponding to narrow-sense heritability estimates of 0.168 for the *single population* and 0.161 for the *joint population* (Table 3). Family-specific variances were also similar between the *single population* and *joint population* (around 0.04; Table 3). Genetic group solutions from ABLUP_UPG were strongly correlated with MF solutions from ABLUP_MF and HBLUP_MF in the *single population* (r 0.97 and 0.96, respectively).

The inclusion of MF influenced (G)EBVs of individuals from HBLUP_MF analyses. This is demonstrated in Fig. 3, where founders' HBLUP_MF solutions are both higher and lower than their HBLUP solutions, depending on genetic group assignment.

Evaluation of models using the LR method

Accuracy of the *single population* focal group was increased by addition of genomic information, from 0.41 for ABLUP to 0.47 for HBLUP (Table 4). Inclusion of genetic groups also increased ACC of pedigree-based solutions, to 0.50 with UPG or 0.46 with MF (Table 4). The ACC of HBLUP solutions was slightly lower with the inclusion of MF for both the *single population* and *joint population*, reduced by 0.02 and 0.01, respectively.

Stability was increased in the *single population* from 0.68 for ABLUP to 0.71 for HBLUP, 0.76 for ABLUP_UPG, and 0.77 for ABLUP_MF (Table 4). Stability was increased by addition of MF to HBLUP, by 0.01 for *single population* and 0.03 for *joint population*.

Dispersion was generally acceptable as estimates were within ± 0.1 of 1.0 (Table 4). Nevertheless, ABLUP and ABLUP_MF models were over-dispersed, and ABLUP_UPG produced the DISP value closest to 1.0. DISP was similar between HBLUP models with and without MF (Table 4).

Bias was similar among ABLUP models regardless of how genetic groups were treated (Table 4). Inclusion of MF in HBLUP models produced contrasting results. For the *single population*, MF reduced bias from +0.066 to -0.006, whereas for the *joint population*, MF increased bias from +0.034 to +0.067 (Table 4).

Table 3. Variance components from *single population* ABLUP and *joint population* HBLUP.

Var. comp. ^a	Single pop.	Joint pop.
$\hat{\sigma}_a^2$	0.171	0.164
$\hat{\sigma}_f^2$	0.043	0.042
$\hat{\sigma}_e^2 \pm \text{SD}(\hat{\sigma}_e^2)^b$	0.805 ± 0.040	0.812 ± 0.038
\hat{h}_{Xsite}^2	0.168	0.161

^a $\hat{\sigma}_a^2$ is the additive variance, $\hat{\sigma}_f^2$ is the family-specific variance, $\hat{\sigma}_e^2$ is the site error variance, and \hat{h}_{Xsite}^2 is the cross-site narrow-sense heritability.

^b Mean and standard deviation of 8 *single population* site error estimates and 13 *joint population* site error estimates.

Discussion

Forest tree breeding programs are often founded on diverse sets of parents from wild populations with distinct characteristics that have developed as adaptations to local environments. Inclusion of provenances as genetic groups uses these similarities to improve the estimation of genetic parameters and effects in a range of tree breeding contexts (e.g. Callister et al. 2011; Lee et al. 2015; Ukrainetz et al. 2018). Thoughtful genetic group definitions for unpedigreed individuals can elucidate patterns of gene flow as well as improve predictions (Klápště et al. 2019). There is growing interest in the HBLUP approach to integrate genomic, pedigree, and phenotypic information and provide unbiased predictions of genetic merit for tree breeding programs (Callister et al. 2021). Accounting for population structure with HBLUP is therefore an important development for tree breeding programs with genotyped subpopulations.

The MF approach is an elegant solution to the challenge of accounting for distinct base populations in HBLUP and it is under extensive investigation in livestock breeding (Masuda et al. 2022), where applications include accounting for trends in genetic means of unknown parents across time (e.g. Granado-Tajada et al. 2020) and accounting for breeds in crossbred populations (Xiang et al. 2017; Kluska et al. 2021; Poulsen et al. 2022). Our use of MF in a forest tree population now expands the scope of this approach to the plant kingdom, where it may be used to address numerous applications in applied breeding, population genetics, and conservation genetics. The distinct advantage of the MF approach over other genomic methods for population genetic studies is that it provides estimates of relationships among base populations and diversity within groups using the genotype and pedigree of individuals that may be multiple generations removed (Legarra et al. 2015).

We have demonstrated that a Γ matrix used to describe MF relationships can be constructed with relatively small founder populations that represent races, landraces, and infused plus tree populations. Diagonal elements of Γ representing our 14 *E. globulus* MFs were substantially smaller than 2/3, which is the expected value if the allele frequencies in the base generation were uniformly distributed (van Grevenhof et al. 2019). Estimates of relationships among races in Γ validated previous *E. globulus* population studies that used microsatellite markers (Jones et al. 2002; Costa et al. 2017) and elucidated relationships between landraces and native races (Freeman et al. 2007; Yost et al. 2021). Although we expect these patterns across races and landraces to have an empirical basis beyond this particular population, the diagonal elements of Γ appeared to be influenced by sample size of the population. Further estimates of Γ based on different samples would provide welcome validation, particularly from programs with larger *E. globulus* base populations.

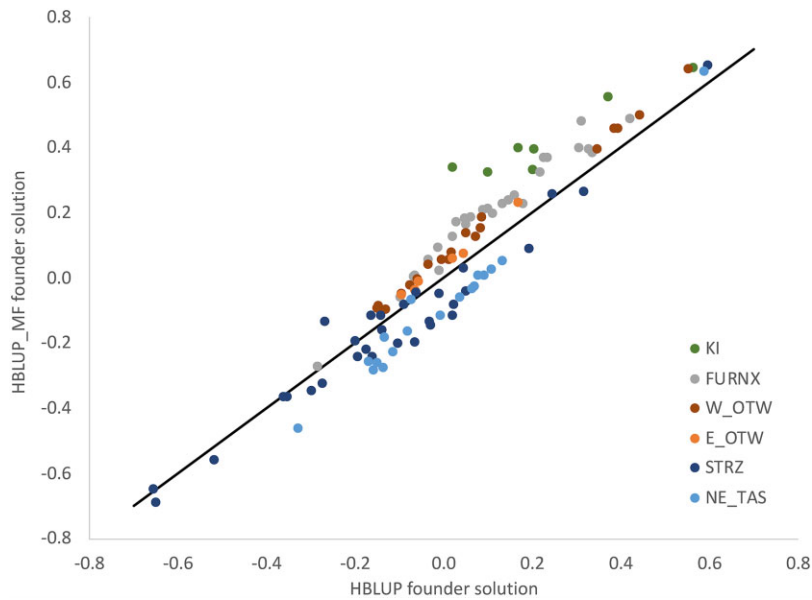


Fig. 3. Relationships between joint population HBLUP_MF and HBLUP solutions for founders from 6 races selected for demonstration purposes. See [Supplementary Table 1](#) for key to race names.

Table 4. Results of linear regression validation: accuracy, stability, dispersion, and bias, expressed in dm^3 units with genetic standard deviation units in parentheses.

Population	Model	Acc	Stab	Disp	Bias (SD_a)
Single program	ABLUP	0.41	0.68	1.09	0.038 (0.14)
	ABLUP_UPG	0.50	0.76	0.97	0.046 (0.17)
	ABLUP_MF	0.46	0.77	1.10	0.048 (0.18)
	HBLUP	0.47	0.71	0.95	0.066 (0.25)
	HBLUP_MF	0.45	0.72	0.96	-0.006 (-0.02)
Joint program	HBLUP	0.45	0.66	0.90	0.034 (0.13)
	HBLUP_MF	0.44	0.69	0.91	0.067 (0.27)

As this is the first forestry study with MF, Γ comparisons are restricted to results with livestock populations. Working with cattle, [Kluska et al. \(2021\)](#) reported diagonal Γ similar to ours for 4 biological types including subspecies *Bos taurus indicus* and *B. taurus taurus* and 6 combinations among them. Their diagonal Γ values ranged from 0.15 (for the group combining all 4 primary biological types) to 0.65, with a mean of 0.41. Off-diagonal elements of Γ in [Kluska et al. \(2021\)](#) tended to be smaller than ours with *E. globulus*, including values of -0.08 and -0.11 between *Bos taurus indicus* and taurine animals from Europe, and from Britain, respectively. [Macedo, Christensen, et al. \(2020\)](#) reported diagonal Γ values between 0.52 and 0.96 for diary sheep, with off-diagonal Γ between 0.22 and 0.44. [Granado-Tajada et al. \(2020\)](#) also found high diagonal Γ values with diary sheep, ranging up to 1.45, indicating that unknown parents within the 3-year period represented by this certain MF were considerably inbred. Working with red dairy cattle, [Kudinov et al. \(2020\)](#) reported Γ with diagonal values from 0.57 to 0.74 and off-diagonals from 0.45 to 0.59, and with chickens, [Bermann et al. \(2021\)](#) constructed a Γ with diagonal elements of 0.50 and 0.57 for 2 MF and 0.38 in the off-diagonal.

The values of Γ impact the values of \mathbf{A}_F , including the sub-component \mathbf{A}_{F22} . [Kudinov et al. \(2020\)](#) created an \mathbf{A}_{F22} with diagonal and off-diagonal elements on average 0.29 and 0.55 units greater than those in their \mathbf{A}_{22} , respectively. On the other hand,

with considerably smaller values in Γ than those reported by [Kudinov et al. \(2020\)](#), our \mathbf{A}_{F22} was only 0.07 units greater than \mathbf{A}_{22} on the diagonal and 0.16 units greater on the off-diagonal, on average (see [Table 2](#)). Mean and range of our \mathbf{G}_{05} were generally similar to those of [Kudinov et al. \(2020\)](#) and [Forni et al. \(2011\)](#), who compared various formulations of \mathbf{G} in a swine population. Our correlations between \mathbf{G}_{05} and \mathbf{A}_{F22} were also similar to those presented by [Kudinov et al. \(2020\)](#).

MF effects are treated as random effects, which varies from the typical practice of treating founding populations as fixed effects in forest genetics ([White et al. 2007](#)). Nevertheless, there are some advantages to treating genetic groups as random effects, and this approach has been considered more seriously by animal geneticists (reviewed by [Masuda et al. 2022](#)). Theoretical concerns with fixed-effect genetic groups include the lack of inbreeding within groups and the inconsistency of allowing for selection to have altered the means of groups but not their genetic variance ([Kennedy 1991](#)). A more pragmatic concern in our experience with tree breeding is that poorly represented genetic groups can be assigned extreme fixed-effect solutions which cause biased EBVs for their relatives throughout the pedigree. Including MF in HBLUP produced changes to (G)EBVs that were similar to including UPG in ABLUP, variously elevating or depressing solutions for founders depending on their race (see [Fig. 3](#)). This is an important demonstration of the impact MF are likely to have in applied tree breeding with HBLUP models.

We found that accuracy increased with the addition of genomic information in the single population. Breeding value accuracy calculated as a function of prediction error variance has typically been shown to increase from ABLUP to HBLUP in studies of forest trees (e.g. [Cappa et al. 2017, 2018](#); [Thavamanikumar et al. 2020](#); [Callister et al. 2021](#)). ACC has been shown to be higher in HBLUP than ABLUP models using the LR method in livestock populations ([Cesarani, Biffani, et al. 2021](#); [Kluska et al. 2021](#); [Sungkhapreecha et al. 2021](#)) and by comparing true and estimated (G)EBVs in simulated experiments ([Garcia-Baccino et al. 2017](#); [Bradford et al. 2019](#); [van Grevenhof et al. 2019](#)). Addition of MFs to HBLUP models did not improve ACC in our study, which differs from results of simulated populations ([Garcia-Baccino et al. 2017](#); [Bradford et al. 2019](#); [van](#)

Grevenhof et al. 2019) that showed substantial increases in ACC with the inclusion of MF. Like our result, Kluska et al. (2021) found ACC from HBLUP to be lower with MF for most traits, while Bermann et al. (2021) showed mixed results across 5 traits. Model specification may affect ACC more for genotyped than ungenotyped individuals (e.g. Ukrainetz and Mansfield 2020) and further comparison of HBLUP models with and without MF is recommended using a larger cohort of genotyped trees and separate LR validations for genotyped and ungenotyped focal groups.

Stability (the correlation between \hat{a}_p and \hat{a}_w) is also called the “ratio of accuracies” and the expected value is $\text{acc}_p/\text{acc}_w$ (Legarra and Reverter 2018; Macedo, Christensen, et al. 2020). Our results showed only a small improvement in stability between ABLUP and HBLUP in the *single population*, whereas other studies have demonstrated more marked increases caused by the addition of genomic information (e.g. Cesarani, Biffani, et al. 2021; Kluska et al. 2021; Sungkhapreecha et al. 2021). Including MF improved our HBLUP model stability by 0.01 in the *single population* and 0.03 in the *joint population*, while it had mixed impacts on stability across 4 traits in Kluska et al. (2021).

When dispersion is equal to 1.0, the focal group’s (G)EBVs are expressed on the same scale regardless of whether they are phenotyped. Over-dispersion (slope less than 1.0) will cause unphenotyped progeny to be selected in error, whereas under-dispersion (slope greater than 1.0) will cause the best unphenotyped progeny to be overlooked for selection despite their merit. Addition of genomic information has corrected strong over-dispersion found for ABLUP models in actual datasets (e.g. milk yield in Sungkhapreecha et al. 2021; direct weaning weight in Kluska et al. 2021; all traits in Cesarani, Masuda, et al. 2021) and simulated populations (Bradford et al. 2019). Other studies (Macedo, Christensen, et al. 2020; Cesarani, Biffani, et al. 2021) have reported dispersion of various models to be mostly within 10% of the expected value (1.0), which was also our result.

Genetic models ideally have zero bias, so that the mean (G)EBV of unphenotyped trees is correctly predicted. Our bias results were mixed. In the *single population* we found a small positive bias for ABLUP and a larger positive bias for HBLUP, which was effectively eliminated by the inclusion of MF. In contrast, a small positive bias for HBLUP in the *joint population* was doubled by the inclusion of MF, which could be a result of the relatively small number of genotyped individuals contributing to the estimation of Γ . The livestock breeding literature contains a wide range of reports about bias in ABLUP and HBLUP models. Garcia-Baccino et al.’s (2017) simulation study found bias to increase markedly with the addition of genomic information and was effectively eliminated with the subsequent inclusion of MF. In Bradford et al.’s (2019) simulation study, an incomplete population displayed large positive bias by ABLUP, a much smaller negative bias by HBLUP, and effectively no bias by HBLUP with MF. Cesarani, Biffani, et al. (2021) and Sungkhapreecha et al. (2021) both found that biases present in ABLUP models were greatly diminished using HBLUP on the same populations. In contrast, Granado-Tajada et al. (2020) found that bias in dairy sheep evaluations was greater for HBLUP than ABLUP, although the difference was not statistically significant. Macedo, Christensen, et al. (2020) reported the least bias for HBLUP with MF, as did Macedo et al. (2022) using a method of Γ construction based on inbreeding trend over time.

Conclusions

Our results demonstrate that the MF method is well suited to representing multiple founder populations in tree breeding analyses

that utilize HBLUP. The Γ matrix can be interpreted to reveal cryptic relationships among UPGs and biologically meaningful associations among provenances or races. Greater confidence in such results will result from validation in future studies with a larger sample of genotyped individuals.

Including MF in HBLUP did not substantially alter model accuracy or dispersion, while it slightly improved stability and presented mixed results for bias in this study. Nevertheless, it is a preferable model on account of its improved approximation of the hierarchical genetic population structure, which we know is present in recently domesticated populations of forest trees. Application of the HBLUP_MF model will therefore provide greater certainty that provenance-level trait variations are being well represented in (G)EBVs of subsequent generations.

Data availability

The genomic, pedigree, and phenotypic data are available at <https://doi.org/10.25387/g3.19487696>.

Supplemental material is available at G3 online.

Acknowledgments

The authors are grateful to Australian Bluegum Plantations and HVP Plantations for permitting the publication of this work. We also acknowledge 2 anonymous reviewers for thoughtful suggestions and improvements.

Author contributions

The project was conceptualized by ANC. Statistical analyses were conducted by ANC with assistance from MB and DL. Manuscript writing was conducted by ANC with revision by MB, JTB, and DL. Trial establishment, maintenance, measurement, sampling, and data preparation were conducted by BPB and SE.

Funding

Genotyping was funded in part through the Forest and Wood Products Australia Project No PNC378-1516. All other experimental costs were funded by Australian Bluegum Plantations and HVP Plantations.

Conflicts of interest

None declared.

Literature cited

- Aguilar I, Misztal I, Johnson DL, Legarra A, Tsuruta S, Lawlor TJ. A unified approach to utilize phenotypic, full pedigree, and genomic information for genetic evaluation of Holstein final score. *J Dairy Sci.* 2010;93(2):743–752.
- Bermann M, Lourenco D, Breen V, Hawken R, Lopes FB, Misztal I. Modeling genetic differences of combined broiler chicken populations in single-step GBLUP. *J Anim Sci.* 2021;99(Supplement_3):254–257.
- Bijma P. Accuracies of estimated breeding values from ordinary genetic evaluations do not reflect the correlation between true and estimated breeding values in selected populations. *J Anim Breed Genet.* 2012;129(5):345–358.

- Bradford HL, Masuda Y, VanRaden PM, Legarra A, Misztal I. Modeling missing pedigree in single-step genomic BLUP. *J Dairy Sci.* 2019;102(3):2336–2346.
- Brawner JT, Bush DJ, Macdonell PF, Warburton PM, Clegg PA. Genetic parameters of red mahogany breeding populations grown in the tropics. *Aust For.* 2010;73(3):177–183.
- Butler DG, Cullis BR, Gilmour AR, Gogel BG, Thompson R. ASReml-R Reference Manual Version 4. Hemel Hempstead (UK): VSN International Ltd; 2018.
- Callister AN, Bradshaw BP, Elms S, Gillies RAW, Sasse JM, Brawner JT. Single-step genomic BLUP enables joint analysis of disconnected breeding programs: an example with *Eucalyptus globulus* Labill. B3 (Bethesda). 2021;11(10):jkab253
- Callister AN, England N, Collins S. Genetic analysis of *Eucalyptus globulus* diameter, straightness, branch size, and forking in Western Australia. *Can J For Res.* 2011;41(6):1333–1343.
- Cappa EP, El-Kassaby YA, Muñoz F, Garcia MN, Villalba PV, Klápště J, Marcucci Poltri SN. Improving accuracy of breeding values by incorporating genomic information in spatial-competition mixed models. *Mol Breed.* 2017;37(125). <https://doi.org/10.1007/s11032-017-0725-6>
- Cappa EP, El-Kassaby YA, Muñoz F, Garcia MN, Villalba PV, Klápště J, Marcucci Poltri SN. Genomic-based multiple-trait evaluation in *Eucalyptus grandis* using dominant DArT markers. *Plant Sci.* 2018;271:27–33.
- Cesarani A, Biffani S, Garcia A, Lourenco D, Bertolini G, Neglia G, Misztal I, Macciotta NPP. Genomic investigation of milk production in Italian buffalo. *Ital J Anim Sci.* 2021;20(1):539–547.
- Cesarani A, Masuda Y, Tsuruta S, Nicolazzi EL, VanRaden PM, Lourenco D, Misztal I. Genomic predictions for yield traits in US Holsteins with unknown parent groups. *J Dairy Sci.* 2021;104(5):5843–5853.
- Christensen OF, Lund MS. Genomic prediction when some animals are not genotyped. *Genet Sel Evol.* 2010;42:2.
- Christensen OF, Madsen P, Nielsen B, Ostersen T, Su G. Single-step methods for genomic evaluation in pigs. *Animal.* 2012;6(10):1565–1571.
- Costa J, Vaillancourt RE, Steane DA, Jones RC, Marques C. Microsatellite analysis of population structure in *Eucalyptus globulus*. *Genome.* 2017;60(9):770–777.
- Dutkowski GW, Potts BM. Geographic patterns of genetic variation in *Eucalyptus globulus* ssp. *globulus* and a revised racial classification. *Aust J Bot.* 1999;47(2):237–263.
- Forni S, Aguilar I, Misztal I. Different genomic relationship matrices for single-step analysis using phenotypic, pedigree and genomic information. *Genet Sel Evol.* 2011;43(1):1.
- Freeman JS, Marques CMP, Carocha V, Borralho N, Potts BM, Vaillancourt RE. Origins and diversity of the Portuguese Landrace of *Eucalyptus globulus*. *Ann For Sci.* 2007;64(6):639–647.
- Garcia-Baccino CA, Legarra A, Christensen OF, Misztal I, Pocrnic I, Vitezica ZG, Cantet RJC. Metafounders are related to Fst fixation indices and reduce bias in single-step genomic evaluations. *Genet Sel Evol.* 2017;49(1):34.
- Gezan SA, de Oliveira AA, Murray D. ASRgenomics: An R package with Complementary Genomic Functions. Version 1.0.0. Hemel Hempstead (UK): VSN International; 2021.
- 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(7):6346–6353.
- Jones RC, Steane DA, Potts BM, Vaillancourt RE. Microsatellite and morphological analysis of *Eucalyptus globulus* populations. *Can J For Res.* 2002;32(1):59–66.
- Jurcic EJ, Villalba PV, Pathauer PS, Palazzini DA, Oberschelp GPJ, Harrand L, Garcia MN, Aguirre NC, Acuña CV, Martínez MC, et al. Single-step genomic prediction of *Eucalyptus dunnii* using different identity-by-descent and identity-by-state relationship matrices. *Heredity.* 2021;127(2):176–189.
- Kennedy BW. C. R. Henderson: the unfinished legacy. *J Dairy Sci.* 1991;74(11):4067–4081.
- Klápště J, Suontama M, Dungey HS, Telfer EJ, Graham NJ, et al. Effect of hidden relatedness on single-step genetic evaluation in an advanced open-pollinated breeding program. *J Hered.* 2018;109(7):802–810.
- Klápště J, Suontama M, Dungey HS, Telfer EJ, Stovold GT. Modelling of population structure through contemporary groups in genetic evaluation. *BMC Genomics.* 2019;20(81). <https://doi.org/10.1186/s12863-019-0778-0>
- Kluska S, Masuda Y, Ferraz JBS, Tsuruta S, Eler JP, Baldi F, Lourenco D. Metafounders may reduce bias in composite cattle genomic predictions. *Front Genet.* 2021;12:678587.
- Kudinov AA, Mäntysaari EA, Aamand GP, Uimari P, Strandén I. Metafounder approach for single-step genomic evaluations of Red Dairy cattle. *J Dairy Sci.* 2020;103(7):6299–6310.
- Lee DJ, Brawner JT, Pegg GS. Screening *Eucalyptus cloeziana* and *E. argophloia* populations for resistance to *Puccinia psidii*. *Plant Dis.* 2015;99(1):71–79.
- Legarra A, Aguilar I, Misztal I. A relationship matrix including full pedigree and genomic information. *J Dairy Sci.* 2009;92(9):4656–4663.
- Legarra A, Christensen OF, Vitezica ZG, Aguilar I, Misztal I. Ancestral relationships using metafounders: finite ancestral populations and across population relationships. *Genetics.* 2015;200(2):455–468.
- 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(1):53.
- Macedo F, Reverter A, Legarra A. Behavior of the linear regression method to estimate bias and accuracies with correct and incorrect genetic evaluation models. *J Dairy Sci.* 2020;103(1):529–544.
- Macedo FL, Astruc JM, Meuwissen THE, Legarra A. Removing data and using metafounders alleviates biases for all traits in Lacauene dairy sheep predictions. *J Dairy Sci.* 2022;105(3):2439–2452.
- Macedo FL, Christensen OF, Astruc J-M, Aguilar I, Masuda Y, Legarra A. Bias and accuracy of dairy sheep evaluations using BLUP and SSGBLUP with metafounders and unknown parent groups. *Genet Sel Evol.* 2020;52(1):47.
- Masuda Y, Tsuruta S, Bermann M, Bradford HL, Misztal I. Comparison of models for missing pedigree in single-step genomic prediction. *J Anim Sci.* 2021;99(2):1–10.
- Masuda Y, VanRaden PM, Tsuruta S, Lourenco DAL, Misztal I. Invited review: unknown-parent groups and metafounders in single-step genomic BLUP. *J Dairy Sci.* 2022;105(2):923–939.
- McRae TA, Dutkowski GW, Pilbeam DJ, Powell MB, Tier B. 2004. Genetic evaluation using the TREEPLAN system. 2004 IUFRO Joint Conference of Division 2 “Forest Genetics and Tree Breeding in the Age of Genomics: Progress and Future”. Charleston (SC). Raleigh NC, USA: North Carolina State University.
- Misztal I, Tsuruta S, Lourenco DAL, Masuda Y, Aguilar I, et al. Manual for BLUPF90 Family of Programs. Athens (USA): University of Georgia; 2014.
- Misztal I, Vitezica ZG, Legarra A, Aguilar I, Swan AA. Unknown-parent groups in single-step genomic evaluation. *J Anim Breed Genet.* 2013;130(4):252–258.
- Money D, Gardner K, Migicovsky Z, Schwaninger H, Zhong G-Y, et al. LinkImpute: fast and accurate genotype imputation for non-model organisms. *G3 (Bethesda).* 2015;5(11):2383–2390.

- Potts BM, Vaillancourt RE, Jordan G, Dutkowski G, Costa e Silva J, et al. Exploration of the *Eucalyptus globulus* gene pool. In: Borralho NMG, Pereira JS, Marques C, Coutinho J, Madeira M, et al., editors. *Eucalyptus in a Changing World*. Aveiro, Portugal: RAIZ, Instituto Investigação de Floresta e Papel; 2004. p. 46–61.
- Poulsen BG, Ostersen T, Nielsen B, Christensen OF. Predictive performances of animal models using different multibreed relationship matrices in systems with rotational crossbreeding. *Genet Sel Evol*. 2022;54(1):1–17.
- Quaas RL. Additive genetic model with groups and relationships. *J Dairy Sci*. 1988;71(5):1338–1345.
- R Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing; 2021. <https://www.R-project.org/>.
- Ratcliffe B, El-Dien OG, Cappa EP, Porth I, Klápště J, Chen C, El-Kassaby YA. Single-step BLUP with varying genotyping effort in open-pollinated *Picea glauca*. *G3 (Bethesda)*. 2017;7(3):935–942.
- Southerton S, Dillon S, Thumma B. Identification of genes and alleles influencing wood development in *Eucalyptus*. *BMC Proc*. 2011; 5(S7):I5.
- Sungkhapreecha P, Misztal I, Hidalgo J, Lourenco D, Buaban S, Chankitisakul V, Boonkum W. Validation of single-step genomic predictions using the linear regression method for milk yield and heat tolerance in a Thai-Holstein population. *Vet World*. 2021; 14(12):3119–3125.
- Thavamanikumar S, Arnold RJ, Luo J, Thumma BR. Genomic studies reveal substantial dominant effects and improved genomic predictions in an open-pollinated breeding population of *Eucalyptus pellita*. *G3 (Bethesda)*. 2020;10(10):3751–3763.
- Tsuruta S, Lourenco DAL, Masuda Y, Misztal I, Lawlor JT. Controlling bias in genomic breeding values for young genotyped bulls. *J Dairy Sci*. 2019;102(11):9956–9970.
- Ukrainetz NK, Mansfield SD. Prediction accuracy of single-step BLUP for growth and wood quality traits in the lodgepole pine breeding program in British Columbia. *Tree Genet Genom*. 2020;16(64).
- Ukrainetz NK, Yanchuk AD, Mansfield SD. Climatic drivers of genotype–environment interactions in lodgepole pine based on multi-environment trial data and a factor analytic model of additive covariance. *Can J For Res*. 2018;48(7):835–854.
- van Grevenhof EM, Vandenplas J, Calus MPL. Genomic prediction for crossbred performance using metafounders. *J Anim Sci*. 2019; 97(2):548–558.
- VanRaden PM. Efficient methods to compute genomic predictions. *J Dairy Sci*. 2008;91(11):4414–4423.
- Westell RA, Quaas RL, van Vleck LD. Genetic groups in animal models. *J Dairy Sci*. 1988;71(5):1310–1318.
- White TL, Adams WT, Neale DB. *Forest Genetics*. Oxfordshire (UK): CABI Publishing; 2007.
- Xiang T, Christensen O, Legarra A. Genomic evaluation for crossbred performance in a single-step approach with metafounders. *J Anim Sci*. 2017;95(4):1472–1480.
- Yost JM, Wise SL, Love NLR, Steane DA, Jones RC, Ritter MK, Potts BM. Origins, diversity and naturalization of *Eucalyptus globulus* (Myrtaceae) in California. *Forests*. 2021;12(8):1129.

Communicating editor: P. Ingvarsson