Mol Breeding (2017) 37: 125
DOI 10.1007/s11032-017-0725-6

Improving accuracy of breeding values by incorporating genomic information in spatial-competition mixed models

Eduardo P. Cappa · Yousry A. El-Kassaby · Facundo Muñoz · Martín N. Garcia · Pamela V. Villalba · Jaroslav Klápště · Susana N. Marcucci Poltri

Received: 2 February 2017 / Accepted: 4 September 2017 / Published online: 13 September 2017

Abstract Climate change and the increasing demand for sustainable energy resources require urgent strategies to increase the accuracy of selection in tree breeding (associated with higher gain). We investigated the combined pedigree and genomic-based relationship approach and its impact on the accuracy of predicted breeding values using data from 5-year-old *Eucalyptus grandis* progeny trial. The number of trees that can be genotyped in a tree breeding population is limited; therefore, the combined approach can be a feasible and efficient strategy to increase the genetic gain and provide more accurate predicted breeding values. We calculated the accuracy of predicted breeding values for two growth traits, diameter at breast height and total height, using two evaluation approaches: the combined approach and the classical pedigree-based approach. We also investigated the influence of two different trait heritabilities as well as the inclusion of competition genetic effects or environmental heterogeneity in an individual-tree mixed model on the estimated variance components and accuracy of breeding

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s11032-017-0725-6) contains supplementary material, which is available to authorized users.

E. P. Cappa (✉)
Instituto de Recursos Biológicos (IRB), Centro de Investigación en Recursos Naturales (CIRN), Instituto Nacional de Tecnología Agropecuaria (INTA), De Los Reseros y Dr. Nicolás Repetto s/n, 1686 Hurlingham, Buenos Aires, Argentina
e-mail: cappa.eduardo@inta.gob.ar

E. P. Cappa
Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Buenos Aires, Argentina

Y. A. El-Kassaby · J. Klápště
Department of Forest and Conservation Sciences, Faculty of Forestry, The University of British Columbia, 2424 Main Mall, Vancouver, BC V6T 1Z4, Canada

F. Muñoz
Unité Amélioration, Génétique et Physiologie Forestières (UR UGPF), Institut National de la Recherche Agronomique (INRA), 45160 Ardon, France

M. N. Garcia · P. V. Villalba · S. N. Marcucci Poltri
Instituto de Biotecnología (IB), Centro de Investigación en Ciencias Veterinarias y Agronómicas (CICVyA), Instituto Nacional de Tecnología Agropecuaria (INTA), De Los Reseros y Dr. Nicolás Repetto s/n, 1686 Hurlingham, Buenos Aires, Argentina

J. Klápště
Department of Genetics and Physiology of Forest Trees, Faculty of Forestry and Wood Sciences, Czech University of Life Sciences Prague, Kamycka 129, 165 21 Prague 6, Czech Republic

Present Address:
J. Klápště
Scion (New Zealand Forest Research Institute Ltd.), 49 Sala Street, Whakarewarewa, Rotorua 3046, New Zealand

values. The genomic information of genotyped trees is automatically propagated to all trees with the combined approach, including the non-genotyped mothers. This increased the accuracy of overall breeding values, except for the non-genotyped trees from the competition model. The increase in the accuracy was higher for the total height, the trait with low heritability. The combined approach is a simple, fast, and accurate genomic selection method for genetic evaluation of growth traits in *E. grandis* and tree species in general. It is simple to implement in a traditional individual-tree mixed model and provides an easy extension to individual-tree mixed models with competition effects and/or environmental heterogeneity.

Keywords Genomic selection · *Eucalyptus grandis* · Individual-tree mixed model · Genetic and environmental competition effects · Environmental heterogeneity · breedR

Introduction

The recent availability of affordable high-density genomic marker panels (e.g., diversity array technology (DArT) and/or single nucleotide polymorphisms (SNP)) created an opportunity, whereby the expected pedigree-based relationship (Wright 1922) used for estimating genetic parameters of complex traits following Fisher's (1918) infinitesimal model is replaced by molecular marker-based realized pairwise genetic similarity between individuals (VanRaden 2008). Studies in domesticated crop plants and animals have demonstrated that the marker-based realized kinship matrix (***G*** matrix) is superior to the common pedigree-based average numerator relationship matrix (***A*** matrix; Daetwyler et al. 2012; El-Kassaby et al. 2012; Klápště et al. 2014; Cappa et al. 2016a; Gamal El-Dien et al. 2016). This superiority is attributable to the marker-based realized kinship ability to account for (1) Mendelian segregation among family members (also known as Mendelian sampling term) as their genetic similarity is estimated precisely instead of using a "universal" expected relationship such as 0.25 and 0.5 given to half- and full-sib family members, respectively; and (2) historical pedigree that cannot be ascertained by the known contemporary pedigree.

The ***G*** matrix can be used instead of the ***A*** matrix in the individual-tree mixed model to predict breeding values by best linear unbiased prediction (BLUP). Fitting ***G*** matrix in the mixed model equations is the genomic selection method known as genomic BLUP or GBLUP (VanRaden 2008). This method is equivalent to estimating individual marker effects using ridge regression and assuming such effects to be mutually independent (Meuwissen et al. 2001). GBLUP is a very promising approach for tree breeding (El-Kassaby et al. 2012; Klápště et al. 2014; Zapata-Valenzuela et al. 2013; Muñoz et al. 2014). However, forest tree breeding populations are usually large, with thousands of progenies from many tested parents. Thus, genotyping costs and logistical issues hinder the successful large-scale implementation of GBLUP approach in tree species. Moreover, in an early program of open-pollinated tests, genotyping of mother trees from the native stand is often not possible. Therefore, a more realistic scenario would include only a sub-set of genotyped trees to obtain the ***G*** matrix (Isik 2014).

The combined pedigree-genomic relationship matrix approach (***H*** matrix) was recently proposed by Misztal et al. (2009), Legarra et al. (2009), and Christensen and Lund (2010), where the traditional ***A*** matrix is concurrently utilized with the ***G*** matrix. Therefore, this blended relationship ***H*** matrix combines two types of genetic information: (1) pedigree information (***A*** matrix) of the many non-genotyped individuals and (2) marker-based relationship (***G*** matrix) of a sub-set of genotyped individuals. The ***H*** matrix (or its inverse $\boldsymbol{H}^{-1}$) can be seen as a projection of genetic merit (or marker genotype) from genotyped to non-genotyped individuals using pedigree relationships (Vitezica et al. 2011). Therefore, additional information generated by including the genomic information in the combined approach acts as genetic relationship bridges connecting information across individuals and parents, thus ultimately facilitating better information utilization during the BLUP analysis. As a result, more reliable and accurate breeding values (in terms of the correlation between the true and predicted breeding value) should be obtained, thus increasing the probability of correct ranking of selection candidates. The combined approach has been widely applied in animal breeding with many successful applications including pigs (Christensen et al. 2012), chickens (Legarra et al. 2011; Chen et al. 2011), dairy cattle (Aguilar et al. 2010), dairy sheep (Baloche et al. 2014), dairy goat (Carillier et al. 2014), and beef cattle (Onogi et al. 2014). A priori, the combined approach should produce more accurate breeding value predictions than the pedigree approach alone and also in forest genetic evaluations. However, only two references on the combined approach have been found in

forest genetic trials (Ogut 2012; Ratcliffe et al. 2017). Using a real cloned *Pinus taeda* population and simulated data, Ogut (2012) obtained higher accuracies of predictions from the combined relationship ***H*** matrix than from the traditional BLUP based on the ***A*** matrix. Ratcliffe et al. (2017) in an open-pollinated population of white spruce (*Picea glauca*) showed also an improvement in the accuracy of predicted breeding value for tree height and wood density traits when compared to the combined approach vs. the utilization of the pedigree-based approach.

The accuracy of breeding values predictions also depends on the covariance structure of the random effects. Therefore, the specification of the dispersion parameters should consider the negative correlation caused by competition among individuals and/or the positive spatial correlation due to the environmental heterogeneity. Negative competition and positive spatial correlations are observable in forest genetic trials, either for each component alone (e.g., Thomson and El-Kassaby 1988; Dutkowski et al. 2006; Cappa and Cantet 2007 and 2008; Ye and Jayawickrama 2008; Costa e Silva et al. 2013) or in combination with each other (Costa e Silva et al. 2013; Cappa et al. 2015; Cappa et al. 2016b). Forest genetic trial analysis including competition effects and/or environmental heterogeneity often result in greater accuracy of breeding values and greater genetic gain than different a priori experimental designs (Dutkowski et al. 2006; Costa e Silva and Kerr 2013). Computationally, using an individual-tree mixed model to account for competition (Cappa and Cantet 2008) and environmental heterogeneity (Cappa and Cantet 2007) or both effects simultaneously (Cappa et al. 2015; Cappa et al. 2016b) with pedigree-based relationship matrix is similar to applying the competition or spatial analyses with the combined ***H*** matrix. Since the combined approach uses traditional BLUP mixed model equations (Legarra et al. 2014), the extension to an individual-tree mixed model with competition genetic or environmental heterogeneity effects is straightforward. However, the use of the combined approach in these models has not been reported in forest tree breeding.

We assessed a *Eucalyptus grandis* (Hill ex Maiden) population consisting of 164 open-pollinated families including 2026 trees measured for two growth attributes (diameter at breast height—DBH—and total height—TH), which are known to possess different heritabilities. A random sample of 187 trees from 132 families with up to three individuals per family was also genotyped with DArT (7680 markers), and subsequently, the genotypes were used to compute the ***G*** matrix. The objectives of the study were to (1) compare the accuracies of the predicted breeding values of genotyped offspring and non-genotyped mothers and offspring from the combined approach vs. the traditional pedigree-based approach; and (2) study the effects of DBH and TH contrasting heritabilities after the inclusion of competition effects or environmental heterogeneity in an individual-tree mixed model on the estimated variance components and on the accuracy of predicted breeding values.

Material and methods

Progeny trial data

An open-pollinated (OP) progeny trial of *Eucalyptus grandis* (Hill ex Maiden) (hereafter *E. grandis*) at Gobernador Virasoro (lat. 28° 02′ S, long. 56° 03′ W alt., 105 m), northern Corrientes province, Argentina, was used. The trial comprised 164 OP families from native forest: 92 from New South Wales and 56 from southeastern Queensland, Australia and 16 from two local land-race sources from Concordia, Entre Rios province, Argentina. Nineteen genetic groups were formed according to provenance. A detailed description of this genetic material can be found in Marcó and White (2002). Briefly, this progeny trial was established as a randomized complete block design with 20 replications and one tree per plot at each replication (i.e., single-tree plot). Five-year diameter at breast height (1.3 m above the ground level) over bark (DBH, cm) and total height (TH, m) were measured on all surviving trees (N = 2026). A random sample of 187 trees originating from 131 families of the OP progeny trial was genotyped with a range of one to three trees per family. The total number of phenotyped trees with at least one genotyped half-sib was 1650 (see Table 1 for summary).

Molecular markers

Total genomic DNA was extracted from young leaves in November 2009 using the CTAB method (Hoisington et al. 1994) with minor modifications to avoid oil precipitation. Instead of using isopropanol in the precipitation step, samples were diluted in twice the volume of 10 mM Tris-HCl and 1 mM EDTA buffer and

Table 1 Open-pollinated families and genotyped and non-genotyped individuals' statistics for diameter (DBH) and total height (TH) in the *Eucalyptus grandis* data set

	No. of records	Mean (SD[a])	
		DBH (cm)	TH (m)
All trees with phenotype	2026	18.85 (4.27)	18.87 (2.68)
Trees from mothers with genotyped offspring	1650	18.87 (4.24)	18.87 (2.65)
Genotyped offspring	187	20.81 (3.07)	20.57 (1.67)
All mothers	164	–	–
Total number of mothers with genotyped offspring	131	–	–

[a]*SD* standard deviation

precipitated with 2.5 volumes of ethanol and 300 mM sodium acetate.

All 187 randomly selected individuals were genotyped by Diversity Arrays Technology Pty Ltd. (DArT P/L, Canberra, Australia) for 2816 DArT molecular markers selected from an operational array with 7680 (Sansaloni et al. 2010). The selected markers showed call rate values > 0.8, reproducibility values > 0.97 (reproducibility of scoring between replicated target assays), and minor allele frequency (MAF) > 0.05.

Statistical models of analysis

Preliminary analyses indicated that the DBH and TH showed strong competition and considerable spatial heterogeneity, respectively (see Online Resource 1). Moreover, according to the Akaike information criterion (AIC), competition model (see below model Eq. (2)) for DBH and spatial model (see below model Eq. (3)) for TH were significantly better than standard models (see below model Eq. (1)) for the two genetic evaluation methods (pedigree-based and combined), except to the combination DBH trait and competition model for the combined approach (results no shown). Therefore, the DBH trait was analyzed as follows:

1. Standard mixed model (TM)

$$\boldsymbol{y} = \boldsymbol{X}\,\beta \;+\; \boldsymbol{Z}_r\boldsymbol{r} + \boldsymbol{Z}\boldsymbol{a} \;+\; \boldsymbol{e} \qquad (1)$$

where $\boldsymbol{y}$ is the vector of phenotypic data, $\boldsymbol{\beta}$ is the vector of genetic groups as fixed effects, $\boldsymbol{r}$ is the vector of random replicate effects, $\boldsymbol{a}$ is the vector of random additive genetic effects of individual trees (i.e., breeding values), and $\boldsymbol{e}$ is the vector of random residuals; $\boldsymbol{X}$, $\boldsymbol{Z}_r$, and $\boldsymbol{Z}$ are incidence matrices relating the observations ($\boldsymbol{y}$) to the model effects in $\boldsymbol{\beta}$, $\boldsymbol{r}$, and $\boldsymbol{a}$, respectively. The vector $\boldsymbol{e}$ is distributed as $\boldsymbol{e} \sim N\left(\boldsymbol{0},\ \boldsymbol{I}\,\sigma_e^2\right)$ where $\boldsymbol{I}$ is an identity matrix and σ_e^2 is the residual variance. For the pedigree-based approach, the $\boldsymbol{a}$ vector is assumed to be distributed as $\boldsymbol{a} \sim N\,(\boldsymbol{0}, \boldsymbol{A}\ \sigma^2{}_a)$ where $\sigma^2{}_a$ is the additive genetic variance, and $\boldsymbol{A}$ is the average numerator relationship matrix from the pedigree information.

2. Competition mixed model (CM)

$$\boldsymbol{y} = \boldsymbol{X}\,\boldsymbol{\beta} + \boldsymbol{Z}_r\boldsymbol{r} + \boldsymbol{Z}_d\boldsymbol{a}_d + \boldsymbol{Z}_c\boldsymbol{a}_c + \boldsymbol{Z}_p\boldsymbol{p}_c + \boldsymbol{e} \qquad (2)$$

where the effects $\boldsymbol{\beta}$, $\boldsymbol{r}$, and $\boldsymbol{e}$ and matrices $\boldsymbol{X}$ and $\boldsymbol{Z}_r$ were specified as described above. The direct and competition breeding values for mothers and offspring are included in the random vectors $\boldsymbol{a}_d$ and $\boldsymbol{a}_c$, linked to the phenotypic data $\boldsymbol{y}$ through the incidence matrices $\boldsymbol{Z}_d$ and $\boldsymbol{Z}_c$, respectively. Matrix $\boldsymbol{Z}_d$ has all elements equal to 0 except for a 1 in the corresponding column. Similarly, the ith row of matrix $\boldsymbol{Z}_c$ has all elements equal to 0 except in the position corresponding to the m_i neighbors of the tree i, with values f_{ij}, $j = 1, \ldots, m_i$. These positive coefficients can be interpreted as the *intensity of competition* (IC) that each neighbor exerts over the phenotype of the ith tree. Following Cappa and Cantet (2008) and assuming that the intensity of competition is related to the inverse of the distance between i and j, the IC for the competitors that lie either in the same row or column (R-C) of the tree i ($f_{i\text{R-C}}$) and for competitors laying in the diagonal positions (D) with respect to tree i ($f_{i\text{D}}$) are

$$f_{i\text{R-C}} = \sqrt{\frac{2}{2\,n_{i\,\text{R-C}} + n_{i\,\text{D}}}}\quad f_{i\text{D}} = \frac{1}{\sqrt{2\,n_{i\,\text{R-C}} + n_{i\,\text{D}}}}$$

where $n_{i\text{R-C}}$ and $n_{i\text{D}}$ are the respective numbers of competitors laying in R-C and D. For the pedigree-based approach, the stacked vector of breeding values is distributed as

$$\begin{bmatrix}\boldsymbol{a}_d\\ \boldsymbol{a}_c\end{bmatrix} \sim N\left(\begin{bmatrix}\boldsymbol{0}\\ \boldsymbol{0}\end{bmatrix}, \begin{bmatrix}\sigma^2_{\text{A}d} & \sigma_{\text{A}d\text{A}c}\\ \sigma_{\text{A}d\text{A}c} & \sigma^2_{\text{A}c}\end{bmatrix} \otimes \boldsymbol{A}\right)$$

where σ^2_{Ad} is the variance of the direct additive genetic effects, σ^2_{Ac} is the variance of the competition breeding values, and σ_{AdAc} is their covariance. The vector $\boldsymbol{p}_c$ includes the environmental competition effects (or permanent environmental competition effects, Cappa and Cantet 2008; Cappa et al. 2015; Cappa et al. 2016b) such that $\boldsymbol{p}_c \sim N(\boldsymbol{0}, \boldsymbol{I}\sigma^2_p)$ where σ^2_p is the variance of environmental competition effects. Finally, the matrix $\boldsymbol{Z}_p$ is composed of the non-zero columns of $\boldsymbol{Z}_c$.

The TH trait was analyzed using the standard model (Eq. (1)) and with the following spatial mixed model:

3. Spatial mixed model (SM)

$$\boldsymbol{y} = \boldsymbol{X}\beta + \boldsymbol{Z}_r\boldsymbol{r} + \boldsymbol{Z}\boldsymbol{a} + \xi + \eta \quad (3)$$

where the effects β, $\boldsymbol{r}$, and $\boldsymbol{a}$, and matrices $\boldsymbol{X}$, $\boldsymbol{Z}_r$, and $\boldsymbol{Z}$ are specified as described above. In Eq. (3), an autoregressive spatial component was added, where the residual vector $\boldsymbol{e}$ was partitioned into a spatially dependent (ξ) and independent (η) residuals (e.g., Costa e Silva et al. 2001). Gilmour et al. (1997) suggested to model the ξ vector as a separable Kronecker product of first-order autoregressive covariance structures (AR1) on the rows (*row*) and the columns (*col*), i.e., $\sigma^2_\xi\ [\mathrm{AR1}(\rho_{col}) \otimes \mathrm{AR1}(\rho_{row})]$, where σ^2_ξ is the spatially dependent residual variance, and ρ_i the autocorrelation parameters. The vector η was assumed pairwise independent, i.e., $\boldsymbol{I}\,\sigma^2_\eta$, where σ^2_η is the independent residual variance.

In the combined approach, the matrix $\boldsymbol{A}$ of the previous mixed models, Eqs. (1)–(3), was substituted by the combined pedigree- and marker-based pairwise relationship $\boldsymbol{H}$ matrix**.** Therefore, the vector $\boldsymbol{a}_d$ of TM and SM is distributed $\boldsymbol{a}_d \sim N(\boldsymbol{0}, \boldsymbol{H}\,\sigma^2{}_a)$, while the stacked vector of breeding values for the CM is distributed as

$$\begin{bmatrix}\boldsymbol{a}_d\\ \boldsymbol{a}_c\end{bmatrix} \sim N\left(\begin{bmatrix}\boldsymbol{0}\\ \boldsymbol{0}\end{bmatrix}, \begin{bmatrix}\sigma^2_{Ad} & \sigma_{AdAc}\\ \sigma_{AdAc} & \sigma^2_{Ac}\end{bmatrix} \otimes \boldsymbol{H}\right)$$

The inverse of the relationship matrix that combines pedigree and genomic information ($\boldsymbol{H}^{-1}$) was derived by Misztal et al. (2009), Legarra et al. (2009), Aguilar et al. (2010), and Christensen and Lund (2010) as

$$\boldsymbol{H}^{-1} = \boldsymbol{A}^{-1} + \begin{bmatrix}\boldsymbol{0} & \boldsymbol{0}\\ \boldsymbol{0} & \lambda\left(\boldsymbol{G}^{-1} - \boldsymbol{A}_{22}^{-1}\right)\end{bmatrix}$$

where λ scales differences between genomic and pedigree-based information, $\boldsymbol{G}^{-1}$ is the inverse of the genomic-based relationship matrix, and $\boldsymbol{A}^{-1}{}_{22}$ is the inverse of the pedigree-based relationship matrix for the genotyped individuals ($\boldsymbol{A}_{22}$). We studied the influence of weighting factor λ in the estimation of variance components for the TM model and for DBH and TH traits. However, only slight differences were found in both studied traits when λ decreased from 1.0 (i.e., only contributions from genomic relationships) to 0.0 (i.e., only contributions from pedigree relationships) by 0.2 (i.e., λ = 1.0, 0.8, 0.6, 0.4, 0.2, 0.0). For instance, the additive (and residual) variance varied from 5.76 (and 12.11) for λ = 0.0 to 6.21 (and 11.76) for λ = 1.0 for DBH (see Online Resource 2). Therefore, the weighting factor λ was set to 1.0 for both traits.

The genomic relationship $\boldsymbol{G}$ matrix from the dominant DArT markers was calculated following the formula suggested by Resende et al. (2010):

$$\boldsymbol{G} = \frac{(\boldsymbol{Z}-\boldsymbol{P})(\boldsymbol{Z}-\boldsymbol{P})'}{\sum_k p_k(1-p_k)}$$

where p_k is the frequency of the code 1 at locus k, $\boldsymbol{Z}$ is a $n \times m$ matrix (n = number of genotyped trees, m = number of DArT markers) that specifies the genotypes expressed as 0/1 denoting the absence/presence of the DArT marker, and $\boldsymbol{P}$ is a matrix containing p_k in the kth column.

A potential problem of the combined approach is that $\boldsymbol{G}$ and $\boldsymbol{A}_{22}$ have to be expressed on the same scale (Meuwissen et al. 2011). However, $\boldsymbol{A}_{22}$ involves relationships of genotyped individuals with reference to the base population (mothers), and $\boldsymbol{G}$ corresponds to relationships within the current population. Several methods have been developed to overcome this scale problem (Aguilar et al. 2010; Chen et al. 2011; Forni et al. 2011; Vitezica et al. 2011; Christensen et al. 2012). In this study, we scaled the $\boldsymbol{G}$ matrix following closely the work of Chistensen et al. (2012; Eq. 4): $\boldsymbol{G}_c = \beta\boldsymbol{G} + \alpha$, where $\boldsymbol{G}_c$ is the scaled $\boldsymbol{G}$ matrix and β and α are parameter values calculated by solving the following system of equations:

$$\mathrm{Avg}(\mathrm{diag}\,(\boldsymbol{G}))\,\beta = \mathrm{Avg}(\mathrm{diag}\ (\boldsymbol{A}_{22}))$$
$$\mathrm{Avg}(\boldsymbol{G})\,\beta + \alpha = \mathrm{Avg}(\boldsymbol{A}_{22})$$

The actual parameter values of α and β were 0.0062 and 0.9938, respectively. Following Vitezica et al. (2011), other two possible corrections of $\boldsymbol{G}$ were also studied: (1)

$\boldsymbol{G}_{\delta} = \boldsymbol{G} + \mathbf{1}\,\mathbf{1}'\,\delta$, where $\delta = \frac{1}{n^2}\left(\sum_i \sum_j \boldsymbol{A}_{22(i,j)} - \sum_i \sum_j \boldsymbol{G}_{i,j}\right)$ (Eq. 5 in Vitezica et al. 2011); and (2) $\boldsymbol{G}_{\mathrm{Fst}} = (1 - ½\,\delta)\,\boldsymbol{G} + \mathbf{1}\,\mathbf{1}'\,\delta$, correction suggested by Powell et al. (2010) in Vitezica et al. (2011). However, given that β was very close to 1 (0.9938), this parameter δ in Vitezica et al. (2011) and Powell et al. (2010) was equal to α, i.e., 0.0062. Consequently, the $\boldsymbol{G}$ matrix corrected by the equation proposed by Christensen et al. (2012) was equivalent to those proposed by Vitezica et al. (2011) and Powell et al. (2010).

All the analyses were performed in R (R Core Team 2015) with the package breedR (Muñoz and Sanchez, 2014) using restricted maximum likelihood (REML, Patterson and Thompson 1971) inference.

The theoretical accuracy $\boldsymbol{r}$ of the predicted breeding values was calculated using the following expression:

$$\boldsymbol{r} = \sqrt{1 - \frac{\mathrm{PEV}}{\hat{\sigma}_a^2(1 + F_i)}}$$

The acronym PEV stands for "prediction error variance" of predicted breeding values and are calculated as the diagonal elements of the inverse of the coefficient matrix from the mixed model equations of each model studied (Gilmour et al. 1995), and F_i is the inbreeding coefficients of tree i. For the CM, the estimated additive genetic variance $\hat{\sigma}_a^2$ was replaced with the estimated direct additive genetic variance $\hat{\sigma}_{Ad}^2$.

The gain in accuracy of a tree's breeding value when using its markers' information depends on the trait, the model, and the group to which the tree belongs (i.e., genotyped offspring, non-genotyped offspring or parents). Furthermore, even within each of these combinations, the gains vary due to differences in relatedness among individuals as measured by the markers and ultimately reflecting the underlying Mendelian segregation process. We estimated the expected gain in accuracy for an individual by maximum likelihood using a linear regression on the trait, the model, and the group.

The narrow-sense individual heritability $\left(\hat{h}^2\right)$ was estimated as

$$\hat{h}^2 = \frac{\hat{\sigma}_a^2}{\hat{\sigma}_a^2 + \hat{\sigma}_e^2.}$$

where $\hat{\sigma}_a^2$ is the estimated additive genetic variance and $\hat{\sigma}_e^2$ is the estimated error variance. For CM, the additive variance $\hat{\sigma}_a^2$ was replaced with the estimated total additive genetic variance (Cappa and Cantet 2008). Assuming that the focus tree is unrelated to its competitors and the competitors are unrelated among themselves, the total additive variance is $\hat{\sigma}_{Ad}^2 + \hat{\sigma}_{Ac}^2$. Additionally, for CM, the denominator of the above expression also included the estimated variance of the environmental competition effects, $\hat{\sigma}_p^2$.

Results

Pedigree and combined relationship matrices

When the pedigree-based $\boldsymbol{A}$ matrix analysis was conducted, the studied 164 OP families were treated as unrelated; conversely, when the genomic-based $\boldsymbol{G}$ matrix of the genotyped offspring was implemented, the ancestral relationship among their mother trees was projected and implied that 132 mothers, with genotyped offspring, are related. The mother trees in the combined $\boldsymbol{H}$ matrix had 4084 (30.6%) out of 13,366 (i.e., (164) × (164–1)/2) pairwise relatedness coefficients higher than zero while that of the genotyped and the non-genotyped offspring, from families with genotyped offspring, had 56.3% of the pairwise relationships higher than zero as compared to that of the $\boldsymbol{A}$ matrix which produced 0.51% higher than zero. The average relationship coefficient for the non-genotyped offspring from the same half-sib family varied from 0.10 to 0.38 for the combined $\boldsymbol{H}$ matrix (consistent with the expected value of 0.25 from the $\boldsymbol{A}$ matrix). Moreover, the relationships between mother and offspring varied from 0.32 to 0.75 (consistent with the expected value of 0.5 from the $\boldsymbol{A}$ matrix). Briefly, genotyped offspring generated additional information for the non-genotyped half-sib mothers and their respective offspring, while, as expected, offspring from mothers with non-genotyped offspring did not produce any additional information.

Variance components

Generally, the TM model produced lower direct additive genetic variance for both traits, while CM resulted in higher additive variance for DBH, and SM showed only slightly higher additive variance for TH, highlighting the effect of competition and environmental heterogeneity (Table 2). These differences among the studied

models showed opposite effects on the residual variance terms as they were higher for the TM model that only focused on direct genetic effects or experimental design consideration (Table 2).

Using the ***H*** matrix generally produced slightly higher additive genetic (direct and indirect) and smaller residual variance estimates than the relationship ***A*** matrix (Table 2), resulting in similar heritability estimates. However, for TH, the spatial variance from the combined approach (2.71) was notably smaller than that from the pedigree approach (23.24). Finally, as expected, the replicate variance decreased after spatial effects (i.e., SM) were included (Table 2). However, this reduction was more pronounced for the SM with the ***H*** matrix than with the ***A*** matrix.

Accuracy of predicted breeding values

Mean accuracies of breeding values from the extended models (i.e., CM or SM) of both investigated traits (i.e., DBH or TH) were higher than those from TM for both pedigree and combined analyses (from 3.0 to 12.5%) (Table 3). However, the increments in the average accuracy across non-genotyped mothers and non-genotyped and genotyped offspring were slightly higher from the pedigree than for the combined approaches: 10.9 vs. 6.6% (CM vs. TM) and 4.5 vs. 4.2% (SM vs.TM) for ***A*** and ***H*** analyses, respectively.

Accuracy comparisons favored the combined analyses over the pedigree for all individual-tree mixed models in both traits, except for the non-genotyped parent and offspring from the CM (Table 3; Fig. 1; Online Resource 3). For the TM, the expected accuracy averaged across groups increased from 0.62 to 0.64 (+ 3.6%) for DBH and from 0.49 to 0.58 (+ 20.7%) for TH. For the extended mixed models (i.e., CM and SM), the increments were slightly negative for DBH (from 0.70 to 0.69, − 1.3%) or positive for TH (from 0.51 to 0.61, + 21.8%). However, the expected gain in the accuracy with respect to the classical pedigree approach varied among the non-genotyped mothers and offspring and genotyped offspring for the different models and traits. Thus, associated with lower heritability for TH, the expected gain in the accuracies of breeding values for parents and all offspring using the combined approach was higher for TH (varying from 9.0 to 43.5%) than for DBH (varying from − 2.7 to 6.2%) (Table 3; Fig. 1). In short, the accuracy of overall predicted breeding values from the combined approach increased with respect to the classical pedigree-based approach.

As expected, when the number of offspring per mother increases, the accuracy of mother breeding values from the *A* and *H* matrices increases for both traits (Fig. 2). For both traits, more complex models produced higher accuracies than the reduced model (i.e., TM), and the combined approach (***H*** matrix) was superior to that of pedigree approach (***A*** matrix), except for the trait-model combination DBH-CM. Therefore, the observed increase in accuracy resulting from using genomic data could potentially reduce the number of tested offspring.

Table 2 Estimation of genetic parameters with pedigree and combined approaches for diameter at breast height (DBH) and total height (TH) in *Eucalyptus grandis* using different individual-tree mixed models with spatial and competition effects

Method of genetic evaluation	DBH		TH	
	TM[a]	CM[a]	TM[a]	SM[a]
Pedigree				
Additive (direct) variance	5.76	7.56	1.08	1.16
Additive indirect variance	–	1.12	–	–
Direct and competition correlation	–	−0.94	–	–
Replicate variance	0.19	0.25	0.75	0.06
Permanent environmental variance	–	0.16	–	–
Spatial variance	–	–	–	23.24
Residual Variance	12.11	9.30	5.38	4.96
Heritability	0.32	0.48	0.17	0.19
Combined				
Additive (direct) variance	6.21	6.26	1.14	1.26
Additive indirect variance	–	0.43	–	–
Direct and competition correlation	–	−0.99	–	–
Replicate variance	0.19	0.25	0.75	0.02
Permanent environmental variance	–	0.10	–	–
Spatial variance	–	–	–	2.71
Residual Variance	11.76	8.52	5.34	4.85
Heritability	0.35	0.44	0.18	0.21

[a] See text for models' abbreviations

Discussion

In forest tree breeding, there is growing interest in applying GBLUP using a genomic individual-tree mixed model to increase genetic gains per unit of time. However, the application of GBLUP requires genotyping all the trees involved in the prediction. Although genotyping costs have dramatically decreased in recent years, genotyping a

Table 3 Mean and standard deviations of estimated theoretical accuracies for the predicted breeding value based on the pedigree and combined approaches for diameter at breast height (DBH) and total height (TH) in *Eucalyptus grandis* using different individual-tree mixed models

Method of genetic evaluation	DBH		TH	
	TM[a]	CM[a]	TM[a]	SM[a]
Pedigree				
Mothers	0.67 ± 0.07	0.72 ± 0.07	0.54 ± 0.07	0.57 ± 0.08
Genotyped offspring	0.60 ± 0.02	0.68 ± 0.02	0.46 ± 0.02	0.48 ± 0.02
Non-genotyped offspring	0.60 ± 0.03	0.68 ± 0.03	0.46 ± 0.02	0.48 ± 0.02
Average	0.62 ± 0.04	0.70 ± 0.04	0.49 ± 0.04	0.51 ± 0.04
Combined				
Mothers	0.68 ± 0.07	0.71 ± 0.07	0.60 ± 0.08	0.62 ± 0.07
Genotyped offspring	0.64 ± 0.02	0.69 ± 0.02	0.66 ± 0.03	0.68 ± 0.04
Non-genotyped offspring	0.61 ± 0.03	0.67 ± 0.03	0.50 ± 0.03	0.53 ± 0.03
Average	0.64 ± 0.04	0.69 ± 0.04	0.58 ± 0.05	0.61 ± 0.04

[a] See text for models' abbreviations

large number of trees may not be cost-effective (Beaulieu et al. 2014). An alternative genomic selection method involves combining information from a few genotyped and many non-genotyped trees with pedigree-recorded data, into an overall relationship ***H*** matrix. The combined ***H*** matrix can be seen as a modification of the regular ***A*** matrix to accommodate genomic relationship (Legarra et al. 2014). We investigated the accuracy of breeding values of non-genotyped mothers and offspring and genotyped offspring in a population of *E. grandis* as well as the influence of the heritability of the trait and the inclusion of competition genetic effects or environmental heterogeneity in an individual-tree mixed model on the estimated variance components and accuracy of breeding values. This preliminary study intended to assess the feasibility and usefulness of including marker data in the genomic evaluation of trees, using the combined approach methodology.

Cappa et al. (2016) highlighted the advantages of the marker-based relatedness estimates over the expected categorical measure of relationships using a sample of 166 genotyped offspring from the same data-set used in this study. The combined relationship ***H*** matrix diffused the information from genomic markers to non-genotyped offspring and parent trees and leveraged this information by more accurately reconstructing family relationships of all trees: genotyped and non-genotyped and with or without phenotype. For example, based on

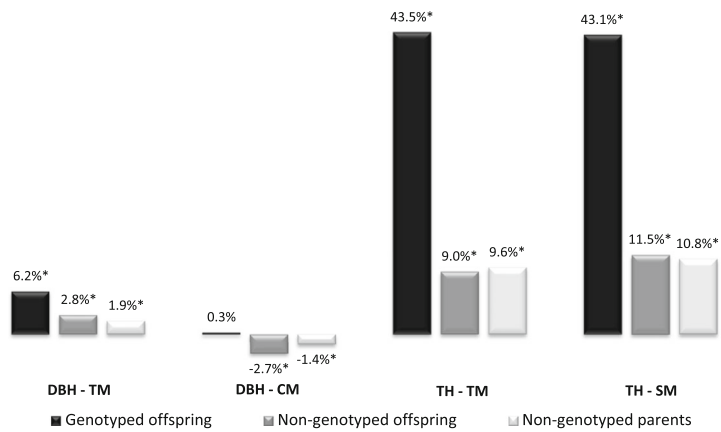

Fig. 1 Expected average percent increases of the accuracy of breeding values for mothers and offspring from the combined approach (***H*** matrix) compared with the classical pedigree (***A*** matrix) using the standard model (TM), competition model (CM), and spatial model (SM) for diameter at breast height (DBH) and total height (TH)

Note: * Significant expected percent increases with t-test $p < 0.05$

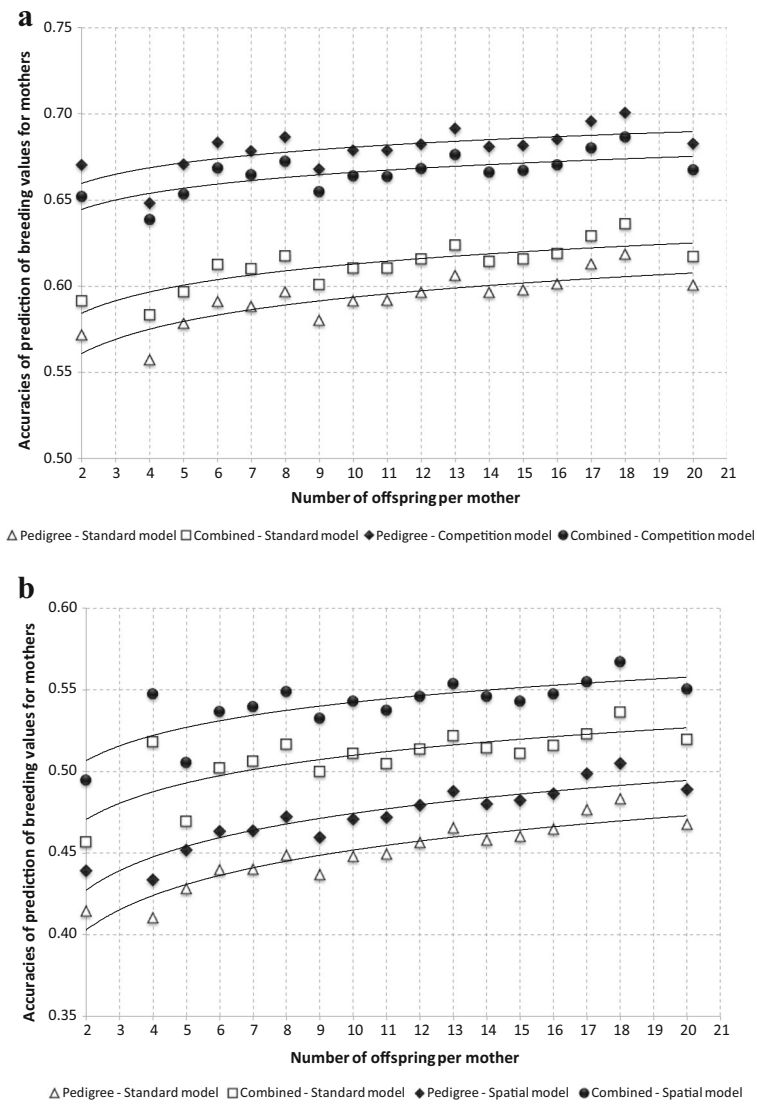

Fig. 2 Estimated theoretical accuracy of breeding values for mothers for the pedigree (***A*** matrix) and the combined (***H*** matrix) approaches using the standard, spatial, and competition individual-tree mixed models vs. the number of phenotyped offspring per mother for (**a**) diameter at breast height (DBH) and (**b**) total height (TH). The trend lines are logarithmic curves and show the increment in the estimated accuracy with the increment in the number of phenotyped offspring per mother

their contemporary pedigree, several pairs of seemingly unrelated mothers in ***A*** appeared as related in ***H***, highlighting the value of capturing historical pedigree through their DNA fingerprints resulting from relatedness among their offspring in ***G*** matrix. Therefore, this methodology is advantageous even when molecular information from the parents is lacking as it offers a backward projection of the relationships between genotyped offspring towards the relationship between mothers.

From a breeding perspective, the accuracy of the selection criteria is important for the realized genetic progress (Loberg et al. 2015). It is thus essential to know how relationship matrices (pedigree or combined) using different individual-tree mixed models affect the accuracy of the breeding values. Although conventional selection using pedigree information has been shown to effectively improve growth traits in *E. grandis* (Marcó and White 2002; Harrand et al. 2009), including more information (i.e., genomic) is expected to increase the accuracy and therefore the probability of correct ranking of selection candidates. However, the increase in the accuracy of breeding values in forest tree breeding populations based on joint phenotypic, pedigree, and genomic information remains uncertain. For both traits (DBH and TH), our results showed that the combined approach led to an increase in the average of theoretical accuracies for genotyped offspring, but also for non-

genotyped offspring and mother trees using the TM and SM models. These results confirmed that the incorporation of genomic information (i.e., DArT markers) increased the accuracies with respect to a pedigree-based approach. Several studies in animal breeding showed that the accuracy of the combined approach is greater than that of the pedigree approach (from 16.7 to 94.4%) for both phenotyped and non-phenotyped individuals (see Table 1 in Legarra et al. 2014). It should be pointed out that the number of genetic markers used in genotyping is essential in resolving historical relationships as indicated by the lack of accuracy improvement when a modest number of markers was used when compared to conventional pedigree approach (831 SNPs; Legarra et al. 2011). The authors (Legarra et al. 2011) argued that denser SNP marker coverage of the genome and a greater number of animals were needed to improve the accuracy of genomic prediction.

Genomic selection is promising for low-heritability traits (e.g., Calus et al. 2008; Grattapaglia and Resende 2011). Thus, for these traits, genomic selection should be more efficient than standard pedigree selection. Our results showed that the advantage of the combined over pedigree approach was larger under low-heritability traits (TH) than in high-heritability traits (DBH). However, the variation (i.e., standard deviation) in the accuracy of breeding values of genotyped and non-genotyped offspring from the combined approach increased for the low-heritability trait (TH) with respect to the pedigree-based approach for the two mixed models studied (TM and SM; Table 3). Even under this condition, however, the combined approach was more accurate than the pedigree approach in all cases. These findings confirm that genomic selection is particularly beneficial for low-heritability traits, even when using the more complex mixed model with competition genetic effects or environmental heterogeneity.

Genetic and environmental competition effects and environmental heterogeneity can affect the precision of the predicted breeding value and need to be considered for obtaining accurate breeding value predictions. Accounting for competition and spatial effects has been shown to increase the accuracy of breeding values for both parents and offspring when pedigree was used (Dutkowski et al. 2006; Costa e Silva and Kerr 2013). Although with lower increments, our results suggest that it is also true for the *E. grandis* evaluation using the proposed combined approach. To our knowledge, this is the first study applying competition or spatial analysis with the combined genomic selection method in forest genetic trials. Therefore, making comparisons with other studies is not possible. In a crop breeding context, Lado et al. (2013) used different models to adjust a spatial trend in phenotypic data of 384 wheat (*Triticum aestivum*) genotypes and concluded that a correction of spatial variation is an essential to increasing the prediction accuracy in genomic selection models.

The scaled $\boldsymbol{G}$ matrix reflects a translation from the relationships relative to the genotyped trees to relationships relative to the base population defined by pedigree (i.e., mothers). An incorrect scale of $\boldsymbol{G}$ matrix can bias the predicted breeding values of genotyped individuals relative to those of the non-genotyped ones (Chen et al. 2011; Forni et al. 2011; Simeone et al. 2011; Vitezica et al. 2011; Christensen et al. 2012). We scaled the genomic relationship matrix $\boldsymbol{G}$ following closely the work of Christensen et al. (2012). In a simulation study, Vitezica et al. (2011) observed less accurate genomic breeding values with incorrectly scaled $\boldsymbol{G}$ but only under strong selection (i.e., non-random) of genotyped individuals and when genotyping was across 10 generations. However, in the present study, the genotyped trees were randomly selected (i.e., unselected). Moreover, when different scale parameters (λ) were used to test the effects of the differences between genomic and pedigree-based information on the dispersion parameters (Aguilar et al. 2010), only slight differences were found in both studied traits (see an example for DBH in Online Resource 2).

Although the small number of genotyped trees (188 out of 2026) and relatively low number of markers assayed (2816 DArT markers) is still less than optimal for a powerful genomic selection study, the combined approach used for the current generation should be applied to the subsequent generation to test the predictive ability of these DArT markers. Cross-validation is difficult to implement due to the scarcity of available trees with genomic information. However, preliminary results on genotyped offspring without phenotype information showed, for the two traits analyzed, that the theoretical accuracy of breeding values from the combined approach decreased (with respect to the offspring with phenotype) but remained always higher than that obtained by the pedigree approach in all the mixed models studied (results not shown). For example, under TM, the accuracies for the non-phenotyped and

genotyped offspring were 0.41 for DBH and 0.33 for TH using the combined approach, while for the pedigree, the respective values were 0.33 and 0.27. Results from this preliminary study indicate that the ***H*** matrix also has superiority over the ***A*** matrix in the prediction for non-phenotyped and genotyped offspring. These predictive abilities are lower (0.54 and 0.51 for DBH and TH, respectively) than those reported by Resende et al. (2012) for two *Eucalyptus* populations genotyped with more than 3000 DArT markers. However, these values were obtained for small effective population size (11 and 51) and high heritabilities (0.53 and 0.42 for DBH and TH, respectively). The predictive ability of the proposed genomic selection method needs to be investigated under higher number of genotyped trees and denser DArT marker coverage of the genome.

Conclusions

The combined approach provides a simple and efficient genomic selection methodology to jointly evaluate genotyped and non-genotyped trees, even when phenotype has not been assessed for parental trees. The method can be easily extended to competition or spatial individual-tree mixed model analysis. This empirical study shows the importance of using all available information (i.e., phenotype, pedigree, and genomic) to improve the accuracy of breeding values for growth traits with low to moderate heritability in an *E. grandis* population. Genomic prediction using an individual-tree mixed model that takes into account competition effects or environmental heterogeneity was more accurate than a prediction based on the standard individual-tree mixed model. These results indicated that genomic selection for *E. grandis* is promising. However, using a higher number of markers and more trees would allow a more efficient evaluation of the genomic selection potential for growth traits in the *E. grandis* population.

Acknowledgements The authors are grateful to Forestal Las Marias for providing the land and logistical support for the test. Thanks to Martín Marcó, Leonel Harrand, Javier Oberschelp, and Juan López, who assisted with field work and data collection. We also wish to thank Cintia Acuña and María Carolina Martínez for their help with the DNA extraction and molecular data analysis and Dario Grattapaglia, Carolina Sansaloni, Cesar Petroli, and Danielle Paiva for generating DArTs data.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

References

Aguilar I, Misztal I, Johnson DL, Legarra A, Tsuruta S, Lawlor TJ (2010) Hot topic: a unified approach to utilize phenotypic, full pedigree, and genomic information for genetic evaluation of Holstein final score. J of Dairy Sci 93:743–752

Baloche G, Legarra A, Sallé G, Larroque H, Astruc JM, Robert-Granié C, Barillet F (2014) Assessment of accuracy of genomic prediction for French Lacaune dairy sheep. J of Dairy Sci 97:1107–1116

Beaulieu J, Doerksen T, Clément S, MacKay J, Bousquet J (2014) Accuracy of genomic selection models in a large population of open-pollinated families in white spruce. Heredity 113: 343–352

Calus MPL, De Roos APW, Veerkamp RF (2008) Accuracy of genomic selection using different methods to define haplotypes. Genetics 178:553–561

Cappa EP, Cantet RJC (2007) Bayesian estimation of a surface to account for a spatial trend using penalized splines in an individual-tree mixed model. Can J For Res 37:2677–2688

Cappa EP, Cantet RJC (2008) Direct and competition additive effects in tree breeding: Bayesian estimation from an individual tree mixed model. Silvae Genet 57:45–56

Cappa EP, Muñoz F, Sanchez L, Cantet RJC (2015) A novel individual-tree mixed model with competition effects and environmental heterogeneity: a Bayesian approach. Tree Genet Genomes 11:120–135

Cappa EP, Klápště J, Garcia MN, Villalba PV, Marcucci Poltri S (2016a) SSRs, SNPs and DArTs comparison on estimation of relatedness and genetic parameters' precision from a small half-sib sample population of *Eucalyptus grandis*. Mol Breeding 36:97

Cappa EP, Stoehr MU, Xie C, Yanchuk AD (2016b) Identification and joint modeling of competition effects and environmental heterogeneity in three Douglas-fir trials. Tree Genet Genomes 12:102

Carillier C, Larroque H, Robert-Granié C (2014) Comparison of joint versus purebred genomic evaluation in the French multi-breed dairy goat population. Genet Select Evol 46:67

Chen CY, Misztal I, Aguilar I, Legarra A, Muir WM (2011) Effect of different genomic relationship matrices on accuracy and scale. J Anim Sci 89:2673–2679

Christensen OF, Lund MS (2010) Genomic prediction when some animals are not genotyped. Genet Select Evol 42:2

Christensen OF, Madsen P, Nielsen B, Ostersen T, Su G (2012) Single-step methods for genomic evaluation in pigs. Animal 6:1565–1571

Costa e Silva J, Dutkowski GW, Gilmour AR (2001) Analysis of early tree height in forest genetic trials is enhanced by including a spatially correlated residual. Can J For Res 31: 1887–1893

Costa e Silva J, Kerr RJ (2013) Accounting for competition in genetic analysis, with particular emphasis on forest genetic trials. Tree Genet Genomes 9:1–17

Costa e Silva J, Potts BM, Bijma P, Kerr RJ, Pilbeam DJ (2013) Genetic control of interactions among individuals: contrasting outcomes of indirect genetic effects arising from neighbour disease infection and competition in a forest tree. New Phytol 197:631–641

Daetwyler HD, Swan AA, van der Werf JHJ, Hayes BJ (2012) Accuracy of pedigree and genomic predictions of carcass and novel meat quality traits in multi-breed sheep data assessed by cross-validation. Genet Select Evol 44:33

Dutkowski GW, Costa e Silva J, Gilmour AR, Wellendorf H, Aguiar A (2006) Spatial analysis enhances modeling of a wide variety of traits in forest genetic trials. Can J For Res 36: 1851–1870

El-Kassaby YA, Klápště J, Guy RD (2012) Breeding without breeding: selection using the genomic best linear unbiased predictor method (GBLUP). New For 43:631–637

Fisher RA (1918) The correlation between relatives on the supposition of Mendelian inheritance. Trans R Soc Edinburgh 52: 399–433

Forni S, Aguilar I, Misztal I (2011) Different genomic relationship matrices for single-step analysis using phenotypic, pedigree and genomic information. Genet Sel Evol 43:1

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

Gilmour AR, Thompson R, Cullis BR (1995) Average information REML, an efficient algorithm for variance parameter estimation in linear mixed models. Biometrics 51:1440–1450

Gilmour AR, Cullis BR, Verbyla AP (1997) Accounting for natural and extraneous variation in the analysis of field experiments. J Agric Biol Environ Stat 2:269–293

Grattapaglia D, Resende MD (2011) Genomic selection in forest tree breeding. Tree Genet Genomes 7:241–255

Harrand L, Hernandez JV, Upton JL, Valverde GR (2009) Genetic parameters of growth traits and wood density in *Eucalyptus grandis* progenies planted in Argentina. Silvae Genet 58:11–19

Hoisington D, Khairallah M, Gonzalez-De-Leon D (1994) Laboratory protocols: CIMMYT. Applied molecular genetics laboratory, 3rd edn. CIMMYT D.F., Mexico

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

Klápště J, Lstibůrek M, El-Kassaby YA (2014) Estimates of genetic parameters and breeding values from western larch open-pollinated families using marker-based relationship. Tree Genet Genomes 10:241–249

Lado B, Matus I, Rodríguez A, Inostroza L, Poland J, Belzile F, Quincke M, Castro M, von Zitzewitz J (2013) Increased genomic prediction accuracy in wheat breeding through spatial adjustment of field trial data. G3: Genes|Genomes|Genetics 3: 2105–2114

Legarra A, Aguilar I, Misztal I (2009) A relationship matrix including full pedigree and genomic information. J Dairy Sci 92:4656–4663

Legarra A, Calenge F, Mariani P, Velge P, Beaumont C (2011) Use of a reduced set of single nucleotide polymorphisms for genetic evaluation of resistance to salmonella carrier state in laying hens. Poultry Sci 90:731–736

Legarra A, Christensen OF, Aguilar I, Misztal I (2014) Single step, a general approach for genomic selection. Livestock Sci 166: 54–65

Loberg A, Dürr JW, Fikse WF, Jorjani H, Crooks L (2015) Estimates of genetic variance and variance of predicted genetic merits using pedigree or genomic relationship matrices in six Brown Swiss cattle populations for different traits. J Anim Breed Genet 132:376–385

Marcó M, White T (2002) Genetic parameter estimates and genetic gains for *Eucalyptus grandis* and *E. dunnii* in Argentina. For Gent 9:205–215

Meuwissen THE, Hayes BJ, Goddard ME (2001) Prediction of total genetic value using genome wide dense marker maps. Genet 157:1819–1829

Meuwissen THE, Luan T, Woolliams JA (2011) The unified approach to the use of genomic and pedigree information in genomic evaluations revisited. J Anim Breed Genet 128: 429–439

Misztal I, Legarra A, Aguilar I (2009) Computing procedures for genetic evaluation including phenotypic, full pedigree, and genomic information. J Dairy Sci 92:4648–4655

Muñoz F, Sanchez L (2014) breedR: statistical methods for forest genetic resources analysts. R package version 0.7–16 Available from: https://github.com/famuvie/breedR

Muñoz PR, Resende MFR Jr, Gezan SA, Deon VRM, de los Campos G, Kirst M, Huber D, Peter GF (2014) Unraveling additive from non-additive effects using genomic relationship matrices. Genet 198:1759–1768

Ogut F (2012) Predictions of genetic merit in tree breeding using factor analytic linear mixed models and blended genomic relationship matrices. PhD Thesis, North Carolina State University, Raleigh, NC, USA. Available from: http://www.lib.ncsu.edu/resolver/1840.16/8252

Onogi A, Ogino A, Komatsu T, Shoji N, Simizu K, Kurogi K, Yasumori T, Togashi K, Iwata H (2014) Genomic prediction in Japanese black cattle: application of a single-step approach to beef cattle. J Animal Sci 92:1931–1938

Patterson HD, Thompson R (1971) Recovery of inter-block information when block sizes are unequal. Biometrika 58:545–554

Powell JE, Visscher PM, Goddard ME (2010) Reconciling the analysis of IBD and IBS in complex trait studies. Nat Rev Genet 11:800–805

R Core Team (2015) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/

Ratcliffe B, Gamal El-Dien O, Cappa EP, Porth I, Klapste J, Chen C, El-Kassaby Y (2017) Single-step BLUP with varying genotyping effort in open-pollinated *Picea glauca*, G3 g3.116.037895; Early online January 25, 2017 doi: https://doi.org/10.1534/g3.116.037895

Resende MDV, Resende Jr. MFRR, Aguiar AM, Abad JIM, Missiaggia AA, Sansaloni C, Petroli C, Grattapaglia D (2010) Computação da seleção genômica ampla (GWS). Colombo, Embrapa Florestas. 79p. (Série Documentos, 210)

Resende MDV, Resende MFR, Sansaloni CP, Petroli CD, Missiaggia AA, Aguiar AM, Abad JM, Takahashi EK, Rosado AM, Faria DA (2012) Genomic selection for growth and wood quality in *Eucalyptus*: capturing the missing

heritability and accelerating breeding for complex traits in forest trees. New Phytol 194:116–128

Sansaloni CP, Petroli CD, Carling J, Hudson CJ, Steane DA, Myburg AA, Grattapaglia D, Vaillancourt RE, Kilian A (2010) A high-density diversity arrays technology (DArT) microarray for genome-wide genotyping in *Eucalyptus*. Plant Meth 6:16–26

Simeone R, Misztal I, Aguilar I, Legarra A (2011) Evaluation of the utility of diagonal elements of the genomic relationship matrix as a diagnostic tool to detect mislabelled genotyped animals in a broiler chicken population. J Anim Breed Genet 128:386–393

Thomson AJ, El-Kassaby YA (1988) Trend surface analysis of a Douglas-fir provenance-progeny transfer test. Can J For Res 18:515–520

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

Vitezica ZG, Aguilar I, Misztal I, Legarra A (2011) Bias in genomic predictions for populations under selection. Genet Res 93:357–366

Wright S (1922) Coefficients of inbreeding and relationship. Am Nat 56:330–338

Ye TZ, Jayawickrama KJS (2008) Efficiency of using spatial analysis in first-generation coastal Douglas-fir progeny tests in the US Pacific northwest. Tree Genet Genomes 4:677–692

Zapata-Valenzuela J, Whetten RW, Neale DB, McKeand SE, Isik F (2013) Genomic estimated breeding values using genomic relationship matrices in a cloned population of loblolly pine. G3: Genes|Genomes|Genetics 3:900–916

Data archiving statement

Genomic data used in this study is available in the Zenodo repository, https://zenodo.org/record/887732. Supplementary information of the *Eucalyptus grandis* trial and pedigree data including identity information of trees, family, and provenance is also available in the Zenodo repository, https://zenodo.org/record/887732. In addition, phenotypic data of the trial *E. grandis* trial will be available upon request.