G3
Genes | Genomes | Genetics

Genomic Prediction of Additive and Non-additive Effects Using Genetic Markers and Pedigrees

Janeo Eustáquio de Almeida Filho,* João Filipi Rodrigues Guimarães,[†] Fabyano Fonsceca e Silva,[‡]
Marcos Deon Vilela de Resende,[§] Patricio Muñoz,** Matias Kirst,[††] and Marcio Fernando Ribeiro de
Resende Júnior**,[1]

*Genetic Plant Improvement Lab, Norte Fluminense "Darcy Ribeiro" State University, Campos dos Goytacazes RJ,
28013-602, Brazil, [†]Futuragene Ltda, Itapetininga SP, 18207-780, Brazil, [‡]Animal Science Department, Federal University
of Viçosa, Viçosa, MG 36570-000, Brazil, [§]EMBRAPA Forestry, Colombo, PR 83411-000, Brazil / Statistics Department,
Federal University of Viçosa, Viçosa, MG 36570-000, Brazil, **Horticultural Sciences Department, University of Florida,
Gainesville, FL 32611, and [††]University of Florida Genetics Institute, School of Forest Resources and Conservation,
Gainesville, FL 32611

ORCID IDs: 0000-0002-3906-8884 (J.E.d.A.F.); 0000-0002-3849-9834 (J.F.R.G.); 0000-0001-9536-1113 (F.F.e.S.); 0000-0001-8973-9351 (P.M.);
0000-0002-2367-0766 (M.F.R.d.R.J.)

ABSTRACT The genetic merit of individuals can be estimated using models with dense markers and pedigree information. Early genomic models accounted only for additive effects. However, the prediction of non-additive effects is important for different forest breeding systems where the whole genotypic value can be captured through clonal propagation. In this study, we evaluated the integration of marker data with pedigree information, in models that included or ignored non-additive effects. We tested the models Reproducing Kernel Hilbert Spaces (RKHS) and BayesA, with additive and additive-dominance frameworks. Model performance was assessed for the traits tree height, diameter at breast height and rust resistance, measured in 923 pine individuals from a structured population of 71 full-sib families. We have also simulated a population with similar genetic properties and evaluated the performance of models for six simulated traits with distinct genetic architectures. Different cross validation strategies were evaluated, and highest accuracies were achieved using within family cross validation. The inclusion of pedigree information in genomic prediction models did not yield higher accuracies. The different RKHS models resulted in similar predictions accuracies, and RKHS and BayesA generated substantially better predictions than pedigree-only models. The additive-BayesA resulted in higher accuracies than RKHS for rust incidence and in simulated additive-oligogenic traits. For DBH, HT and additive-dominance polygenic traits, the RKHS- based models showed slightly higher accuracies than BayesA. Our results indicate that BayesA performs the best for traits with few genes with major effects, while RKHS based models can best predict genotypic effects for clonal selection of complex traits.

Pedigree and high-density DNA markers have been used to predict genetic merit of individuals in animal (Wiggans *et al.* 2011) and plant

breeding (Resende *et al.* 2012a; Resende *et al.* 2012; Crossa *et al.* 2014). Initially, genomic prediction models were developed to select sires in animal breeding, accounting only for additive effects (Meuwissen *et al.* 2001). However, the prediction of dominance effects represents an important component of models designed for breeding program that focus on cross-bred populations and/or hybrid production (Zeng *et al.* 2013; Nishio and Satoh 2014; Resende *et al.* 2017). In vegetatively-propagated species, non-additive effects are especially relevant because selection takes advantages of dominance effects. Hence, breeders can transfer whole-genotypic values of individuals to the next generation through clonal selection strategies.

Many statistical methods for genomic prediction have been proposed, which differ with regards to the assumed trait genetic architecture

(Gianola 2013; de los Campos *et al.* 2013). For instance, BayesA consists of a whole-genome multiple regression (WGR) model where each marker regression coefficient assumes one specific variance (Meuwissen *et al.* 2001). Consequently, BayesA should provide a good fit for oligogenic traits where few genes explain a large proportion of the observed genetic variation (Meuwissen *et al.* 2001). However, prediction of non-additive effects introduces new SNP-covariates associated with these effects (Toro and Varona 2010), which may prohibitively increase the number of estimated parameters. The semi-parametric Reproducing Kernel Hilbert Space (RKHS) method has been proposed to account for non-additive effects, and is less computationally demanding than BayesA, especially when the number of individuals is substantially lower than the number of markers. RKHS combines features of non-parametric kernel regression with mixed-effect linear models (Gianola *et al.* 2006; Gianola and van Kaam 2008). RKHS is equivalent to the animal model in BLUP (de los Campos *et al.* 2009a), but uses kernel matrices that are different than the numerator relationship matrix (Morota *et al.* 2013; Tusell *et al.* 2014; Morota and Gianola 2014). In fact, the numerator relationship matrix is a special case of RKHS matrices.

Empirical results showed that the inclusion of dominance in BayesA improved the prediction accuracy in complex real and simulated traits in a pine breeding context. The additive-dominance BayesA was also superior to other Bayesian models for oligogenic traits (de Almeida Filho *et al.* 2016). In addition, reports that analyzed complex traits in other forest species showed improvement of accuracy in genomic selection models with inclusion of pedigree information in additive-only models (Beaulieu *et al.* 2014a, b). Thus, it is hypothesized that inclusion of pedigree information in additive-dominance BayesA can also increase prediction accuracy in structured breeding populations. The same hypothesis can be considered for alternative methods to BayesA, such as RKHS, which may also provide higher prediction accuracies when combined with pedigree data (Crossa *et al.* 2010, 2013).

To date, few studies have compared the prediction of non-additive effects from models using only genetic marker data, with models that integrate genetic markers and pedigree information. Moreover, the ability of RKHS models to predict whole genotypic values has not been assessed in the context of forest tree breeding. This study had as objectives: 1) evaluate the impact of including pedigree information in genomic prediction models applied to forest tree breeding, 2) evaluate RKHS and compare it with BayesA when it includes non-additive effects or not and 3) contrast different cross-validation strategies in the context of a structured breeding population. We applied and evaluated these methods to traits with distinct genetic architecture measured in a loblolly pine breeding population, and to simulated oligogenic and polygenic traits with different levels of dominance.

MATERIAL AND METHODS

Loblolly pine population and phenotypes
In this study, we used the traits tree height (HT), diameter at breast height (DBH) and two measures of fusiform rust infection: presence or absence of rust (RFbin) and gall volume (RFgall). These traits were measured in the loblolly pine (*Pinus taeda*) CCLONES breeding population, as previously described (de Almeida Filho *et al.* 2016). The traits HT and DBH are polygenic (Resende *et al.* 2012b), and HT also shows a significant non-additive genetic contribution (Muñoz *et al.* 2014). In contrast, rust resistance is thought to be controlled by fewer genes, including loci of large effect (Resende *et al.* 2012b; Quesada *et al.* 2014). The CCLONES population where these traits were measured was created from 42 founders, which were crossed to create 71 full-sib families, with an average of 13 individuals per family (SD = 5). In total,

923 individuals were genotyped for 7,216 single nucleotide polymorphic markers (SNP) as previously described (de Almeida Filho *et al.* 2016). From the total set, all 4,722 polymorphic loci were used in this study. Traits HT and DBH were measured in a field trial established in Nassau under alpha lattice design (Florida, USA), when the plants were six years old, in eight clonal replicates (Baltunis *et al.* 2007). Rust resistance traits (RFbin and RFgall) were measured in greenhouse, under randomized complete block design, as previously described (Kayihan *et al.* 2005, 2010). The phenotypes for these traits were adjusted with the following linear model:

$$y_{ijk} = \mu + b_k(r_j) + r_j + g_i + e_{ijk}$$

Where y_{ijk} is phenotype of the ith clone evaluated in the jth repetition and the kth incomplete block, μ is the intercept, $b_k(r_j)$ is the random effect of kth incomplete block nested in the jth repetition $b_k(r_j) \sim N(0, \sigma_b^2)$, r_j is the fixed effect of the jth repetition and g_i is the effect of the ith clone (considered as fixed to estimate the least-square means or adjusted means) and e_{ijk} is the error of observation ijk $e_{ijk} \sim N(0, \sigma^2)$. This model was used for analysis of DBH and HT. For rust resistance, the incomplete block term was excluded. The analysis of variance of these traits is described in Table S1.

Simulated population and phenotypes
We also analyzed six traits with different genetic architecture, in a simulated population described previously (de Almeida Filho *et al.* 2016). This analysis included traits with oligogenic and polygenic inheritance, and three levels of dominance (none, median and high). The simulated population was created following a standard forest breeding program model, in two steps. First, a base population with 1,000 individuals was created by randomly sampling 2,000 haplotypes from a population with effective size of 10,000, generated by 1,000 generations of a neutral coalescence model, with mutation rate 2.5×10^{-8} per generation (Willyard *et al.* 2007). Next, the breeding population was created by phenotypic selection of 100 individuals from the base population. These selected individuals were randomly mated to create 1,000 individuals to be used in the first breeding cycle. From the 1,000 individuals of the first breeding cycle, 42 were phenotypically selected and crossed following the same mating design used in the loblolly pine CCLONES population (Baltunis *et al.* 2007). Ten independent replicates were carried out for each simulated trait.

The simulated genome had 12 chromosomes with 100 cM. A set of 10,000 bi-allelic loci (*e.g.*, SNP) were used in developing prediction models, and 30 or 1,000 QTL were simulated for oligogenic and polygenic traits, respectively. All traits had narrow sense heritability of 0.25. Three levels of dominance were evaluated: d^2: 0, 0.1 and 0.2, where d^2 is the proportion of phenotypic variance explained by dominance deviation - $d^2 = V_d/V_p$; V_d and V_p are dominance deviation and phenotypic variances respectively (Falconer and Mackay 1996). The additive effect of a gene (*a*) was defined as half of the difference between the alternative homozygotes, and the dominance effect (*d*) was estimated by the difference between the heterozygote and the mean of the homozygotes. The distribution used for *a* in oligogenic traits were gamma (rate = 1.66, shape = 0.4) with sign (positive or negative) sampled with equal probability (Meuwissen *et al.* 2001). For polygenic traits, *a* was simulated with standard normal distribution (mean = 0, sd = 1) (Hickey and Gorjanc 2012). When dominance was present, it was simulated by: $d_i = a_i \times \tau_i$, where τ_i was sampled from a normal distribution with mean zero and standard deviation 1 and 2 for traits with medium- and high-dominance-levels, respectively. To achieve the targeted values of d^2, only samples that provided d^2 between 0.09 and 0.11 for medium-

dominance traits and between 0.19 and 0.21 for high-dominance traits were kept. The simulated populations showed very similar allele frequencies, when compared across scenarios. Hence, d^2 was mostly defined by a larger or smaller sampled value of τ_i.

Statistical methods

We evaluated models that consider just SNP or pedigree information, and models that combined both. The DNA marker (SNP) component was fitted with the following methods: a) semi parametric Reproducing Kernel Hilbert Space models (RKHS), using different kernels (Ka and Ka-Kd), and b) BayesA considering additive and additive-dominant effects. BayesA was the choice of WGR because it was previously shown to generate better predictions than the Bayesian Lasso and Bayesian Ridge Regression, and to have similar accuracies as BayesB, for oligogenic traits (de Almeida Filho *et al.* 2016). The full base model can be represented by:

$$y_j^* = \mu + g_j + u_j + \delta_j + e_j$$

Where y_j^* is the phenotype (adjusted clonal mean in real traits) of individual j; μ is the intercept; g_j is the genotypic value for each jth individual, estimated from SNP data — this term changed among methods adopted, as can be seen bellow; u_j is the additive polygenic effect (when included) of individual j; δ_j is the dominant polygenic effect (when included) of individual j; e_j is the error term. The joint data distribution, the prior distribution for the constant μ, and the prior distributions for the vectors g, u and δ (containing information of all evaluated individuals) are given by:

$$y_j^* \big| \mu, g_j, u_j, \delta_j, \sigma_e^2 \sim IID\ N(\mu + g_j + u_j + \delta_j, \sigma_e^2);$$

$$\mu \sim N(0, 10^6);$$

$$u \big| A\sigma_u^2 \sim N(0, A\sigma_u^2);\ \text{and}$$

$$\delta \big| D\sigma_\delta^2 \sim N(0, D\sigma_\delta^2).$$

Where A and D are the additive relationship matrix and dominance relationship matrix, respectively, calculated using standard methods (Henderson 1984; Lynch and Walsh 1998; Mrode 2014). In the pedigree model, the g_j term was absent and u and δ are the breeding values and dominance deviation vectors, respectively.

BayesA: The BayesA model adopted in this study used the changes in the formulation proposed by Gianola *et al.* (2009), modified from the original version (Meuwissen *et al.* 2001) to estimate the shape parameter of the inverted chi-square (χ^{-2}). This modification is expected to reduce the influence of the hyperparameter and consequently improve the bayesian learn. The full BayesA model included additive and dominant effects from SNP and pedigree. This model can be represented by:

$$y_j^* = \mu + \sum_{i=1}^{k} (x_{ij} a_i + w_{ij} d_i) + u_j + \delta_j + e_j$$

Where x_{ij} and w_{ij} are the functions of SNP i in individual j, for genotypes AA, Aa and aa. x_{ij} is composed of values 1 (AA), 0 (Aa) and -1 (aa) and w_{ij} of 0 (AA), 1 (Aa) and 0 (aa). a_i and d_i are the additive and dominance effect of marker i, respectively. The dominance effect was fitted only in the additive-dominance model. The priors used in linear regression coefficients for additive-dominance and additive models are described below:

$$a_i \big| \sigma_{a_i}^2 \sim N(0, \sigma_{a_i}^2);\ \sigma_{a_i}^2 \big| \nu_a, S_\alpha \sim \chi^{-2}(\nu_a, S_\alpha); S_a \big| s_a, r_a \sim G(s_a, r_a);$$
$$d_i \big| \sigma_{d_i}^2 \sim N(0, \sigma_{d_i}^2);\ \ \sigma_{d_i}^2 \big| \nu_d, S_d \sim \chi^{-2}(\nu_d, S_d);\ \ S_d \big| s_d, r_d \sim G(s_d, r_d).$$
Where χ^{-2} and G represent the scaled inverse chi square and Gamma distributions, respectively.

RKHS Kernel averaging model: From the full base model represented above, the term g in RKHS represents the whole genotypic values explained by the markers, including additive and non-additive effects such as dominance and gene interactions (Gianola *et al.* 2006; Gianola and van Kaam 2008). Here, g was modeled in two forms referred to as RKHS Ka and RKHS Ka-Kd.

RKHS-Ka: In the RKHS-Ka, the whole genotypic effect (g) was explained by 3 SNP functions:

$$g = \sum_{r=1}^{3} g_r$$

where

$$g_r \big| K_{ar} \sigma_{g_r}^2 \sim N(0, K_{ar} \sigma_{g_r}^2)$$

$$\sigma_{g_r}^2 \big| \nu_g, S_g \sim \chi^{-2}(\nu_g, S_g)$$

$$K_{a_r} = \exp(-\varphi_{a_r} D_a^2)$$

D_a^2 is the square of the Euclidean distance matrix among the individuals using the SNP additive incidence matrix **X** that takes values -1, 0 and 1. The φ_a is a bandwidth parameter that controls for the relationship measure between individuals j and j', for a given distance (squared Euclidean in this case). Large positive values of bandwidth result in the relationship of j and j' being close (or equal) to 0. Small positive values result in the relationship of j and j' being close (or equal) to 1. The kernel averaging method (De los Campos *et al.* 2010) was used to determinate the bandwidth components in this study. In this approach, each SNP function g is replaced for two or more SNP functions with the same distance (squared Euclidean in this case), but with different bandwidth values — in kernel averaging these bandwidths are not regular parameters, they are constants. The bandwidth values (φ_{a_r}) used in g_1, g_2, and g_3 are 5/h, 1/h and 0.2/h respectively, where h is 5th percentile of D_a^2 leading to local, intermediate and global kernels, respectively (González-Camacho *et al.* 2012; Tusell *et al.* 2014).

RKHS-Ka-Kd: In addition to the information contained in the **X** matrix described above, RKHS-Ka-Kd also includes **W**, that is the SNP incidence matrix for dominance effects in additive-dominance Bayes A. The g in this case for each SNP matrix (X and W) used three SNP functions:

$$g = \sum_{r=1}^{3} (g_{a_r} + g_{d_r})$$

The g_{ar} consists of the same SNP functions used in RKHS-Ka, and for g_{d_r} we assumed:

$$g_{dr} \big| K_{d_r} \sigma_{g_{d_r}}^2 \sim N(0, K_{d_r} \sigma_{g_{d_r}}^2)$$

$$\sigma_{g_{d_r}}^2 \big| \nu_{g_d}, S_{g_d} \sim \chi^{-2}(\nu_{g_d}, S_{g_d})$$

$$K_{d_r} = \exp(-\varphi_{d_r} D_d^2)$$

D_d^2 is the square of the Euclidean distance matrix using the dominance SNP matrix **W** that takes values 1 for heterozygote and 0 for both homozygotes, as described in the additive-dominance-BayesA model. The same bandwidth values used in RKHS-KA were used for g_{a_r} in RKHS-Ka-Kd, and for g_{d_1}, g_{d_2} and g_{d_3}, the bandwidths values (φ_{d_r}) were 5/h$_d$, 1/h$_d$ and 0.2/h$_d$ respectively, where h$_d$ is 5th percentile of D_d^2, similar to Morota *et al.* (2014). In both RKHS models (Ka and Ka-Kd) the whole genotypic value was predicted but cannot be separated into the components: breeding value, dominance deviation and epistasis. The summary of models compared are available in Table S2.

Models validation

The prediction accuracies were calculated using 10-fold cross-validation (Resende *et al.* 2012b; de Almeida Filho *et al.* 2016). In order to infer the impact of the training population (TP) on the prediction accuracy, the TP was defined following three different schemes: a) prediction across families –individuals from a group of families were used to fit the model and the genetic potential was predicted in individuals from different families; b) prediction within families – the genetic merit of was predicted in individuals belonging to the same families included in the TP; c) random sample of individuals for TP. These validation approaches were also applied to each of the ten replicates of the simulated data. For data collected from the CCLONES population, the 10-fold based on random sample process was applied 10 times with independent groups of individuals in each fold. Prediction of accuracies and regression coefficients of parametric values on validation data were estimated for each of the 10 folds. A graphical representation of the 10-folds cross-validation schemes used in this study is available in Figure S1.

Breeding values and dominance deviation

The expected breeding value (EBV) and the expected dominance deviation (EDD) were estimated as described below:

$$E\hat{B}V_j = \sum_i \left[I(x_{ij}=1)2q_i + I(x_{ij}=0)(q_i - p_i) - I(x_{ij} = -1)2p_i \right]\hat{\alpha}_i + \hat{u}_j$$

and

$$E\hat{D}D_j = \sum_i \left[-I(x_{ij}=1)2q_i^2 + I(x_{ij}=0)2p_iq_i - I(x_{ij}=-1)2p_i^2 \right]\hat{d}_i + \hat{\delta}_j$$

Where p_i is allele frequency of allele A of SNP i, $q_i = 1-p_i$, $\hat{\alpha}_i$ is the average effect of substitution, $\hat{\alpha}_i = \hat{a}_i + \hat{d}_i(q_i - p_i)$, and I is an indicator function of SNPs; $\hat{u}_j$ and $\hat{\delta}_j$ are terms from additive and dominant polygenic effects, respectively. The whole genotypic value is the sum of $E\hat{B}V_j$ and $E\hat{D}D_j$. In RKHS-based models, the whole genotypic value is predicted, and therefore the partitions between $E\hat{B}V_j$ and $E\hat{D}D_j$ cannot be estimated.

Variance components

The variance components from WGR are extensions of estimators previously reported (Zeng *et al.* 2013; Ertl *et al.* 2014) that assume absence of epistasis, linkage equilibrium among markers and Hardy-Weinberg equilibrium (Gianola *et al.* 2009). The general estimator of additive variance (V_A) and the variance due to dominance deviation (V_D) are:

$$\hat{V}_A = 2\sum_i p_iq_i \left[\hat{\sigma}_{a_i}^2 + (q_i - p_i)^2 \hat{\sigma}_{d_i}^2 \right] + \hat{\sigma}_u^2$$

and

$$\hat{V}_D = 4\sum_i (p_iq_i)^2 \hat{\sigma}_{d_i}^2 + \hat{\sigma}_\delta^2$$

The components $\hat{\sigma}_u^2$ and $\hat{\sigma}_\delta^2$ in each model above are associated with the polygenic effect. The remaining equation of $\hat{V}_A$ and $\hat{V}_D$ are due to marker effects. The whole genotypic variance is the sum of additive and dominance variance. The h^2, d^2 and H^2 are the proportion of additive, dominance and genotypic variance in the phenotypic variance. In RKHS models the genetic variance estimated by markers is the whole genotypic variance. All models were fitted with the R package BLGR (de los Campos and Perez 2014), using 100,000 iterations, burn-in of 20,000, thinning of 3 and default hyperparameters previously described (Pérez and de los Campos 2014).

Data availability

All phenotypic and genotypic data utilized in this study have been previously published as a standard data set for development of genomic prediction methods (Resende *et al.* 2012b). Simulated data available from the Dryad Digital Repository: http://dx.doi.org/10.5061/dryad.3126v. Supplemental material available at FigShare: https://doi.org/10.25387/g3.8379059.

RESULTS

The prediction within families was more accurate than prediction across families

To infer the impact of population structure and the choice of training population (TP) in the estimates of prediction accuracy, we carried the cross-validation across families, within families and at random. For all models and traits, the prediction accuracies were lowest when the predicted individuals came from families not included in the TP (across family cross-validation). The reductions in accuracy ranged from ~15–59% in both real and simulated data (Figure 1-2). Conversely, the within-family cross validation resulted in highest accuracy for the majority of cases. In CCLONES, the superiority of the within-family TP ranged from 1.5–2.4%, 2.7–6.8% and 3.6–5.6% for HT, RFbin and RFgall, respectively (Figure 1). In simulated traits, the within-family cross validation was slightly superior (0.15–3.3%) than the random sample TP in ~85% and ~80% of cases for prediction of breeding values and whole genotypic values respectively (Figure 2). There model comparison, however, was very similar across TP schemes, and no interaction between models and TP was observed (data not shown). Hence, the following model comparisons and general conclusions, were performed using random-sample cross-validation.

Pedigree information in model prediction

The use of pedigree information has been shown previously to improve the accuracy of genotypic predictions (Crossa *et al.* 2014). Here, we evaluated if this gain in accuracy is also observed in our populations, based on correlation using parametric genetic values from simulated data, and phenotypic values from real data. Initially we compared the prediction accuracy of models based only on pedigree information against models with markers-only data (Table 1). Overall, models based only on pedigree data had modestly lower accuracy than those using marker information for traits measured in the pine CCLONES population. The improvement in accuracy for breeding and genotypic values
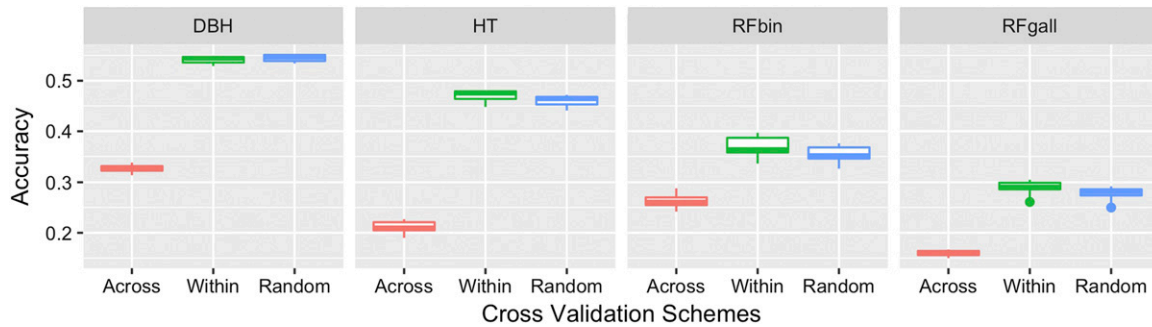
Figure 1 Accuracy distribution of all genomic prediction models fitted for tree height (HT), diameter at breast height (DBH) and two measures of fusiform rust infection: presence or absence of rust (RFbin) and gall volume (RFgall). These results were achieved from three different 10-fold cross validation orientations: a) Across families: Each fold is a group of distinct families; b) Within families: The folds were grouped inside families; and c) Random sample: Each fold is a group of distinct individuals random sampled ignoring family information.

prediction was much more apparent when markers were used in the simulated population, compared to when only pedigree data were utilized. This improvement was particularly pronounced for simulated oligogenic traits (Table 2). Next, we extended the comparison of accuracies to include models that combined both pedigree and marker information. We observed that in only few instances the expanded models had higher accuracies, and that in general the accuracies were comparable to those with marker only in the real and simulated data (Table 1-2). One of such cases was the accuracy in predicting dominance deviation in the simulated traits (Table 2).

Genotypic predictive model accuracy depends on trait architecture

The two genomic prediction methods tested in this study differ with respect to the assumptions regarding the genetic architecture of the trait being predicted. BayesA is a linear regression model that assumes that each marker has different variance. Consequently, some markers can explain the effect of major loci, such as in oligogenic traits (*e.g.*, rust resistance). On the other hand, RKHS is a semi-parametric model that assumes that all markers with the same MAF contribute equally to the relationship among individuals. Thus, this assumption has closer resemblance with polygenic trait regulation. For phenotypic predictions, RKHS had higher accuracy for DBH and HT (Figure 3) when

compared to BayesA. Similar results were observed in the polygenic traits with presence of dominance effects ($d^2 = 0.1$ and $d^2 = 0.2$). In both cases, the RKHS method resulted in higher accuracies than BayesA models for genotypic and phenotypic prediction (Figure 4, Table S3). In contrast, BayesA resulted in higher correlation for RFbin (Figure 3) and all oligogenic simulated traits (Figure 4), independent of the inclusion of pedigree in the genomic prediction and independent of the simulated dominance effect. The superiority of the additive-dominance models was trait dependent and in some cases opposite accuracy patterns were observed, as it is the case for HT and RFbin (Figure 3).

Genotypic predictive model strength depends on non-additive effects

The prediction of genotypic values is valuable for genetic improvement programs where breeders can capture additive and non-additive effects by cloning selected individuals. Prediction of dominance effects for each locus is also critical to optimize crossing designs, such as in mate-pair allocation. The inclusion of the Kd kernels matrix in RKHS was equal or worse than the Ka-only kernels for both real traits (Figure 3) and simulated traits (Figure 4). On the other hand, the inclusion of dominance effects in the BayesA model resulted in a better prediction for HT (Figure 3). For simulated traits, the additive-dominance-BayesA showed considerably higher genotypic (Figure 4) and phenotypic
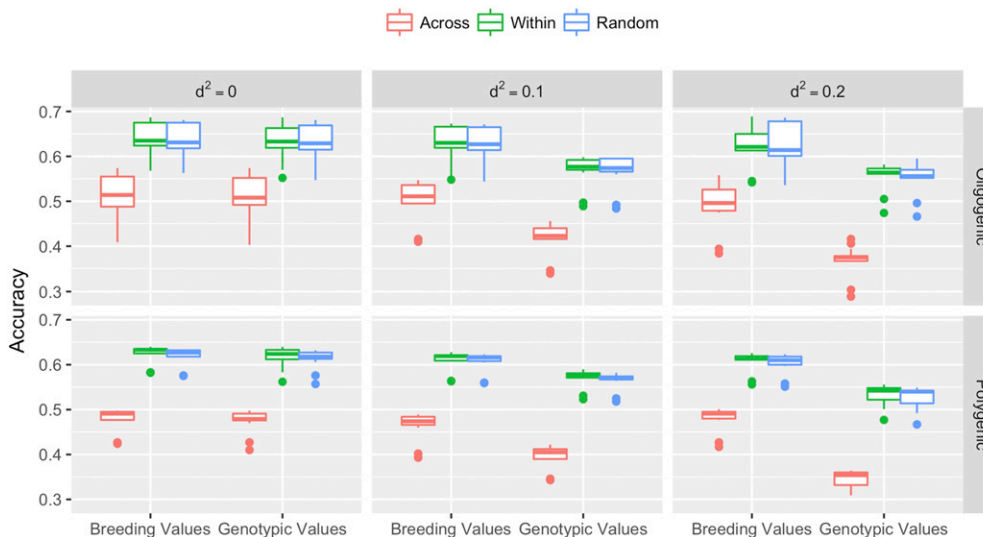

Figure 2 Accuracy average of prediction of breeding values and genotypic values for all genomic prediction models fitted in six simulated traits: Oligogenic and Polygenic with three degrees of dominance ($d^2 = 0$; $d^2 = 0.1$ and $d^2 = 0.2$). These results were achieved from three different 10-fold cross validation orientations: a) Across families: Each fold is a group of distinct families; b) Within families: The folds were grouped inside families; and c) Random sample: Each fold is a group of distinct individuals random sampled ignoring family information.

Models	DBH	HT	RFbin	RFgall
Pedigree	0.536	0.450	0.331	0.255
Markers	0.545	0.459	0.361	0.288
Markers + Pedigree	0.548	0.465	0.356	0.279

Only-Markers models: Additive-, Additive-dominance-BayesA, RKHS Ka and RKHS Ka-Kd; Only-Pedigree models: Additive-, Additive-dominance-Pedigree; Markers + Pedigree are the models id: 2,4,5,7,8,10,11 (Table S2).

(Table S3) prediction accuracy compared to the additive-BayesA in traits with high dominance.

BayesA models generated higher accuracies in breeding value prediction

The breeding value of one individual represents the component of its genotypic value that is directly transmitted to the progeny. Thus, the breeding value is a critical parameter in the selection of individuals to be used extensively in mating with other individuals in the population. With BayesA and pedigree-based models, it was possible to predict directly the breeding values. The RKHS models, however, generate a prediction of the whole genotypic values, which cannot be split in breeding values, dominance deviation and epistasis. In the simulated the traits, we still reported the correlation between predicted genotypic values and parametric breeding values to assess the accuracy in breeding value prediction. The traditional additive-BayesA and additive-dominance-BayesA based only in marker information estimated the higher accuracy for breeding value prediction for all simulated traits (Figure 4), while pedigree-based models showed the worst accuracies (Table 2).

Variance components and heritability

One of the most important tasks for a breeder is to make decisions regarding the breeding strategy. Several genetic parameters are evaluated to determine breeding strategies, including the estimation of variance components and the proportion of the genetic variance relative to the phenotypic variance. Here these parameters were estimated using genetic markers, pedigree information or both, for the real and simulated traits. For simulated traits, the parametric values are known, which allowed us to evaluate the capacity of the methods and models to accurately estimate them.

Considering the parametric values of h^2 and d^2 in the simulated population, the BayesA models that used only genetic markers resulted in the least biased estimate of the genetic parameters (Table S4). The inclusion of pedigree information in BayesA models increased the estimates of heritabilities, and in most cases these parameters were overestimated. Models based only on pedigree also overestimated these genetics parameters: the additive-only pedigree model over estimated h^2, and the additive-dominance pedigree models estimated high d^2 for traits with no dominance.

The RKHS based models predict the whole genotypic values. Consequently, estimation of both h^2 and d^2 is not possible. Nonetheless, H^2 can be obtained. The H^2 was calculated for the simulated traits and showed that all RKHS based models, regardless of the pedigree inclusion, overestimated this parameter substantially (on average by ~106%, Table S4). Overestimation is higher in models that included the Kd kernel (RKHS Ka-Kd). Table S5 reports the results of heritabilities estimated with all models for real traits. Similarly, to the simulated data, the estimation of h^2 and d^2 increased with the pedigree information, and the RKHS based models resulted in much larger estimates of H^2.

Prediction bias

The regression coefficient (slope) of observed values *vs.* predicted values was used to measure model bias – a slope of one indicates the absence of any bias. The linear regression of simulated data included parametric genotypic values and prediction. For the real data, where the parametric values are unknown, we calculated the slope using phenotype values. The predictions using real data (Table S6) resulted in regression coefficients near one. Most predictions using simulated data resulted in a slope similar to one, with the exception of dominance deviation predictions. In this case, the slopes showed a biased dominance prediction. More specifically, the slope for dominance deviation prediction was less than one in the marker-only models with additive-dominance effects, indicating an underestimation of dominance effects. In contrast, when the pedigree information was considered, the slopes were higher than one, suggesting that the pedigree information contributed to an overestimation of dominance effects (Table S7).

DISCUSSION

We tested the performance of genomic- and pedigree-based models, with and without non-additive effects, for the prediction of genetic values in complex traits. We used real data from a standard pine tree
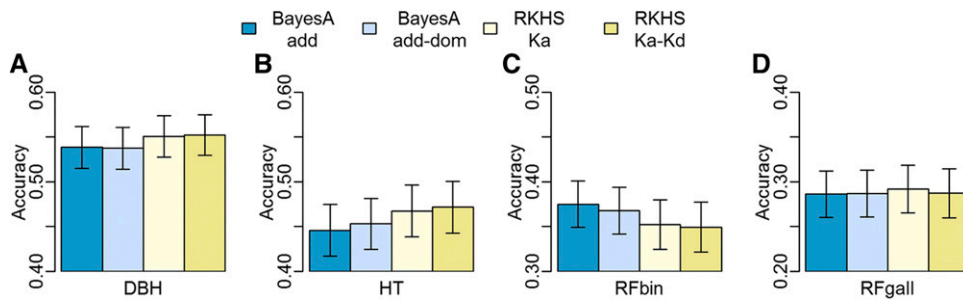
Accuracy	Models	$d^2 = 0$		$d^2 = 0.1$		$d^2 = 0.2$	
		Olig	Poly	Olig	Poly	Olig	Poly
Breeding Value	Pedigree	0.564	0.576	0.545	0.560	0.538	0.554
	Markers	0.653	0.627	0.645	0.618	0.645	0.613
	Markers + Pedigree	0.646	0.626	0.639	0.615	0.638	0.610
Dominance Deviation	Pedigree	—	—	0.179	0.202	0.271	0.259
	Markers	—	—	0.175	0.169	0.273	0.243
	Markers + Pedigree	—	—	0.186	0.185	0.284	0.258
Genotypic Value	Pedigree	0.556	0.567	0.488	0.521	0.481	0.479
	Markers	0.652	0.626	0.586	0.575	0.569	0.537
	Markers + Pedigree	0.638	0.619	0.578	0.571	0.566	0.536
Phenotypic Value	Pedigree	0.251	0.259	0.284	0.306	0.313	0.335
	Markers	0.297	0.286	0.338	0.331	0.378	0.373
	Markers + Pedigree	0.290	0.282	0.335	0.331	0.373	0.373

Only-Markers models: Additive-, Additive-dominance-BayesA, RKHS Ka and RKHS Ka-Kd; Only-Pedigree models: Additive-, Additive-dominance-Pedigree; Markers + Pedigree are the models id: 2,4,5,7,8,10,11 (Table S2).

Figure 3 Average of phenotypic prediction accuracies and standard error for four markers-only models: additive- and additive-dominance-BayesA, RKHS-Ka and RKHS Ka-Kd for diameter at breast height (DBH), height (HT) and Rust resistance evaluated as gall volume (RFgall) and presence or absence (RFbin) in loblolly pine. The standard errors (s.e.=s.d.(x)/sqrt(10)) were calculated for each ten-fold procedure. The error bars are the averages of s.e. across the ten independent cross validations.

breeding program in its third generation. Pine traits used for model testing included plant height (HT), diameter at breast height (DBH), and the rust resistance measures RFbin and RFgall, whose narrow sense heritability, prior to least square mean adjustment, were previous reported as 0.31, 0.31, 0.21 and 0.12, respectively (Resende *et al.* 2012b). These traits have different genetic architecture; DBH and HT are polygenic whereas rust resistance is thought to be an oligogenic trait (Resende *et al.* 2012b; Quesada *et al.* 2014). Moreover, HT has significant non-additive effects (Muñoz *et al.* 2014). To expand and validate our conclusions, we simulated six traits with distinct genetic architectures (polygenic and oligogenic) and three dominance levels.

Pedigree information in genomic predictions

Pedigree and marker information were used separately and in combination, to predict genetic values. In the prediction of breeding and whole genotypic values, the genomic selection model and the combined model (genetic markers + pedigree) were substantially better than the models accounting only for pedigree information. However, the combined model did not improve the prediction accuracies. Previous studies suggest that the use of pedigree data only results in an improvement in prediction when a low marker density is used, as seen in simulated studies (Calus and Veerkamp 2007), as well as in wheat (Crossa *et al.* 2010) and dairy cattle (Vazquez *et al.* 2010). In contrast, pedigree
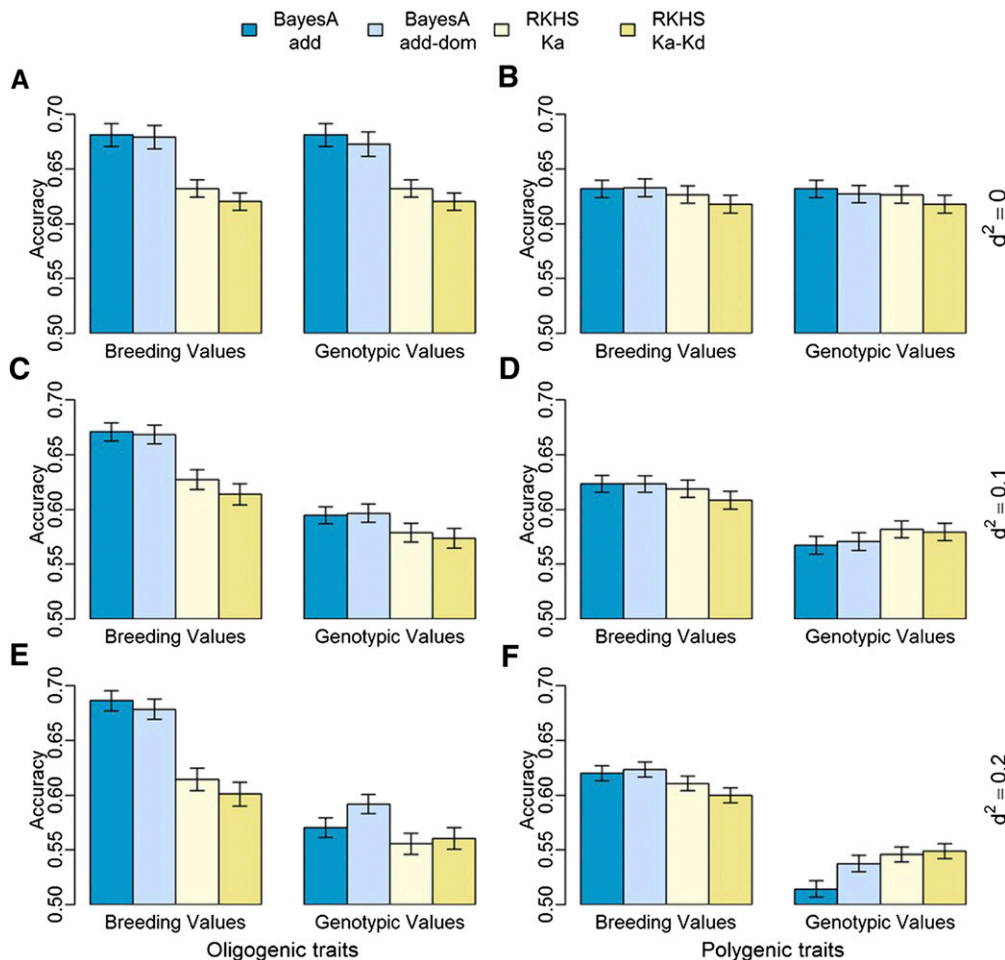

Figure 4 Accuracy average of prediction of breeding values and genotypic values for four marker-only models: additive-, additive-dominance-BayesA, RKHS-Ka and RKHS Ka-Kd for six simulated traits: Oligogenic and Poligenic with three degrees of dominance ($d^2 = 0$; $d^2 = 0.1$ and $d^2 = 0.2$). Error bars are standard error (s.e.) considering the 100 independent samples used to calculate the mean (s.e.=s.d.(x)/sqrt(100)).

information did not improve model predictions when a higher density of genetic markers were used (Calus and Veerkamp 2007; de los Campos *et al.* 2009b; Vazquez *et al.* 2010). An exception was reported in maize, when prediction models generated with a high-density SNP panel from genotyping-by-sequencing (GBS) was improved with the addition of pedigree data (Crossa *et al.* 2013). A possible explanation for this outcome rests in the fact that the GBS data contains a high frequency of missing data (Crossa *et al.* 2013), resulting in incomplete genomic information.

Semi-parametric kernel choice

Different kernels can be used to improve the predictions of complex traits in semi-parametric RKHS models. In the current study, the genomic predictions of RKHS Ka-Kd models and the simpler RKHS Ka models yielded similar results. These findings are in agreement with those previously reported by Morota *et al.* (2014) for dairy cattle. These authors did not find increased accuracies with the inclusion of extra kernels in the RKHS Ka. Other kernel comparisons in RKHS also showed that RKHS Ka is a robust choice for the prediction of additive and non-additive effects (Morota *et al.* 2013; Tusell *et al.* 2014).

BayesA resulted in the highest breeding values accuracy

The BayesA models based only on genetic marker information resulted in the highest accuracies of breeding values prediction in all the simulated scenarios, regardless of the dominance effects. The additive-BayesA and additive-dominance-BayesA models also resulted in similar accuracies for breeding values in additive-dominance traits, despite the fact that the breeding value is also function of dominance effects (Falconer and Mackay 1996). While this result was unexpected, one possible reason for this is the overall low accuracy in the prediction of dominance. The results were also similar to another simulation study that showed higher genetic gain with the additive model compared to an additive-dominance model, even in presence of higher dominance effects (Denis and Bouvet 2012). Furthermore, Nishio and Satoh (2014) also showed that dominance inclusion did not provide higher accuracies in breeding value prediction, even in traits with dominance effects.

Prediction of whole-genotypic values of traits with distinct genetic architecture

Models with built-in assumptions that some loci have major effects, such as BayesA, usually provide better genomic predictions for simulated oligogenic traits (de los Campos *et al.* 2013), and also for real traits controlled by few genes, such as fat percentage in milk (Habier *et al.* 2011). For genotypic prediction in this study, BayesA generated better predictions than RKHS for RFbin and simulated oligogenic traits. This finding is in agreement with other studies that suggest that rust resistance is an oligogenic trait (Resende *et al.* 2012b; Quesada *et al.* 2014). In the case of simulated additive-dominance oligogenic traits, the additive-dominant-BayesA model resulted in the best predictions of genotypic values.

Our analyses of polygenic simulated traits showed that, for additive-dominant polygenic traits, the RKHS models were better than the additive-BayesA for whole genotypic predictions. These findings agree with those of other authors who argue that RKHS addresses non-additive variation in a non-explicit manner (Gianola *et al.* 2006; Gianola and van Kaam 2008; Morota and Gianola 2014). In addition, RKHS was modestly more accurate when compared to the additive-dominant-BayesA, confirming that RKHS can be explored for predictions in polygenic traits.

In CCLONES population, the RKHS models was slightly more accurate than additive- and additive-dominant-BayesA for HT and DBH. In addition, the additive-dominant-BayesA model was slightly more accurate than the additive-BayesA for HT. This result suggests the presence of non-additive effects in *P. taeda* tree height, which was also previously suggested by Muñoz *et al.* (2014). Similar results were also observed in a genomic selection study by Bouvet *et al.* (2016) which reported higher prediction accuracy of HT in Eucalyptus using models with non-additive effects. In contrast, El-dien *et al.* (2016), did not achieved prediction accuracy increase with inclusion of non-additive effects in HT in an open-pollinated white Spruce population. Altogether, and as expected, we conclude that the increase of prediction accuracy with inclusion of dominance depend of the population and trait.

Similar to Xavier *et al.* (2016), we also fit a model that combined BayesA and RKHS in a single model. This combined model resulted in lower accuracies than additive BayesA for RFbin and simulated oligogenic traits with $d^2 = 0$ and $d^2 = 0.1$. In addition, the BayesA+RKHS combined model provides lower accuracy than RKHS for HT, DBH and polygenic additive-dominance traits (Table S8). However, the BayesA+RKHS model was slightly more accurate than BayesA for genotypic prediction in oligogenic traits with high dominance ($d^2 = 0.2$) and more accurate than RKHS in additive polygenic traits for breeding and genotypic values prediction (Tables S9). In soybean, Xavier *et al.* (2016) reported that BayesA+RKHS and BayesB+RKHS were the most accurate genomic prediction models. These authors suggested that this model could be capturing the breeding values through BayesA and the remaining genetic components with RKHS. In our dataset, this advantage of the combined model was not clear, but further evaluation of this model could be made.

Heritabilities estimation

When using RKHS based models and pedigree information, the parameters h^2, d^2 and H^2 were often overestimated. Similarly, others genomic prediction studies in forest species also suggested that pedigree based models generate large and unrealistic estimates of h^2, when compared with GBLUP (Bouvet *et al.* 2016; El-dien *et al.* 2016).

In quantitative genetics theory, the additive and non-additive effects are typically assumed to be orthogonal and can be divided as independent components from the whole genotypic variance (Falconer and Mackay 1996; Hallauer *et al.* 2010; de los Campos *et al.* 2015). However, in practical situations (absence of Hardy-Weinberg equilibrium, non-random matting), these presuppositions do not hold, and additive and dominance effects are related to each other and not orthogonal. The inability to partition the components can lead to bias estimates of the variance components, affecting the estimates of selection gain as well as the interpretation of the trait architecture. Muñoz *et al.* (2014) indicated that this bias was stronger in pedigree models and recommended the use of genetic markers to partition the additive and non-additive components.

The marker specific variances and therefore the variance component estimates can be influenced by the effect of the priors in Bayesian models (Gianola *et al.* 2009). Nonetheless, in our study, the BayesA models using only markers, generated more reasonable variance components estimates, compared to the parametric values. The inclusion of dominance effects in BayesA resulted in less biased heritabilities estimation for additive-dominance traits, and the traditional additive-BayesA were less biased for h^2 estimation in additive traits. This suggests that additive-dominant models should be used for estimation of heritability in cases where the inclusion of dominance effects increase the prediction

accuracy. Based on this observation, the estimates for tree height of h^2 and d^2 using the additive-dominant model equal 0.37 and 0.17 respectively. For DBH, RFbin and RFgall, the additive model is suggested and theirs respective h^2 are estimated in 0.52, 0.39 and 0.29 (Table S13). The h^2 estimated for HT in this report were similar to the value found in Resende *et al.* (2012b), but higher than reported in Resende *et al.* (2012a). The estimates of d^2 for HT were similar to previous report (Muñoz *et al.* 2014), which used the same population but different methods. For DBH, and rust resistance, the h^2 obtained in this study, were higher than other authors (Resende *et al.* 2012a, 2012b). Calus and Veerkamp (2007) reported that the inclusion of polygenic effects resulted in better estimates of variance components when compared with models that included only markers - this result that was not replicated in our study.

Prediction accuracy is affected by the choice of training population

This study contrasted three schemes of training population (TP). The results showed that across families cross validation resulted the lowest accuracy, whereas the setup to within families cross validation produced the best results in most part of cases, being slightly superior than random sample individuals for TP. These results were expected, and further emphasize the importance of genetic relationship among individuals used to fit the model and target individuals for genomic prediction. Genomic Selection models have been shown to capture not only linkage disequilibrium between the markers and causal alleles, but also genetic relationship (Habier *et al.* 2007). The scheme based on random sampling has been commonly used, because it represents a balance between within and across families. These results agree with other studies in forest breeding that showed higher accuracies in the prediction within families than across families (Beaulieu *et al.* 2014b, 2014a). In addition, Albrecht *et al.* (2014) reported higher accuracies in the prediction inside genetic groups than across genetics groups. Hence, one of the most important factors when outlining a genomic selection breeding program is the definition of the target population. The TP needs to resemble, as close as possible, the level of relationship of the target population where GS will be applied.

CONCLUSION

In summary, we conclude that, in the CCLONES pine population and in our simulation studies, genomic selection is effective compared to phenotypic pedigree selection. The inclusion of pedigree information did not improve the prediction accuracies, suggesting that pedigree computation is not required in genomic prediction programs with reasonable SNP panels. This study also supports the findings that the individuals used to compose the training population should be genetically related to the individuals in the target breeding population. Finally, the BayesA models overcame the RKHS-based models for breeding and genotypic prediction for oligogenic traits, while in polygenic traits, BayesA was suitable for breeding value prediction and RKHS for whole-genotypic prediction.

LITERATURE CITED

Albrecht, T., H.-J. Auinger, V. Wimmer, O. J. Ogutu, C. Knaak *et al.*, 2014 Genome-based prediction of maize hybrid performance across genetic groups, testers, locations, and years. Theor. Appl. Genet. 127: 1375–1386. https://doi.org/10.1007/s00122-014-2305-z

de Almeida Filho, J. E., J. F. R. Guimarães, F. F. Silva, M. D. V. De Resende, P. Muñoz *et al.*, 2016 The contribution of dominance to phenotype prediction in a pine breeding and simulated population. Heredity 117: 33–41. https://doi.org/10.1038/hdy.2016.23

Baltunis, B. S., D. A. Huber, T. L. White, B. Goldfarb, and H. E. Stelzer, 2007 Genetic analysis of early field growth of loblolly pine clones and seedlings from the same full-sib families. Can. J. For. Res. 37: 195–205. https://doi.org/10.1139/x06-203

Beaulieu, J., T. Doerksen, S. Clément, J. Mackay, and J. Bousquet, 2014a Accuracy of genomic selection models in a large population of open-pollinated families in white spruce. Heredity 113: 343–352. https://doi.org/10.1038/hdy.2014.36

Beaulieu, J., T. K. Doerksen, J. MacKay, A. Rainville, and J. Bousquet, 2014b Genomic selection accuracies within and between environments and small breeding groups in white spruce. BMC Genomics 15: 1048. https://doi.org/10.1186/1471-2164-15-1048

Bouvet, J., G. Makouanzi, D. Cros, and P. Vigneron, 2016 Modeling additive and non-additive effects in a hybrid population using genome-wide genotyping : prediction accuracy implications. Heredity 116: 146–157. https://doi.org/10.1038/hdy.2015.78

Calus, M. P. L., and R. F. Veerkamp, 2007 Accuracy of breeding values when using and ignoring the polygenic effect in genomic breeding value estimation with a marker density of one SNP per cM. J. Anim. Breed. Genet. 124: 362–368. https://doi.org/10.1111/j.1439-0388.2007.00691.x

Crossa, J., Y. Beyene, S. Kassa, P. Pérez, J. M. Hickey *et al.*, 2013 Genomic prediction in maize breeding populations with genotyping-by-sequencing. G3 Genes, Genomes. Genet. 3: 1903–1926.

Crossa, J., L. Campos Gde, P. Pérez, D. Gianola, and J. Burgueño *et al.*, 2010 Prediction of genetic values of quantitative traits in plant breeding using pedigree and molecular markers. Genetics 186: 713–724. https://doi.org/10.1534/genetics.110.118521

Crossa, J., P. Pérez, J. Hickey, J. Burgueño, L. Ornella *et al.*, 2014 Genomic prediction in CIMMYT maize and wheat breeding programs. Heredity 112: 48–60. https://doi.org/10.1038/hdy.2013.16

Denis, M., and J. Bouvet, 2012 Efficiency of genomic selection with models including dominance effect in the context of Eucalyptus breeding. Tree Genet. Genomes 9: 37–51. https://doi.org/10.1007/s11295-012-0528-1

El-dien, O. G., B. Ratcliffe, J. Klápstě, I. Porth, C. Chen *et al.*, 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. Genet. 6: 743–753.

Ertl, J., A. Legarra, Z. G. Vitezica, L. Varona, C. Edel *et al.*, 2014 Genomic analysis of dominance effects on milk production and conformation traits in Fleckvieh cattle. Genet. Sel. Evol. 46: 40. https://doi.org/10.1186/1297-9686-46-40

Falconer, D. S., and T. F. C. Mackay, 1996 *Introduction to Quantitative Genetics*, Longman. Harlow, UK.

Gianola, D., 2013 Priors in whole-genome regression: The Bayesian alphabet returns. Genetics 194: 573–596. https://doi.org/10.1534/genetics.113.151753

Gianola, D., R. L. Fernando, and A. Stella, 2006 Genomic-assisted prediction of genetic value with semiparametric procedures. Genetics 173: 1761–1776. https://doi.org/10.1534/genetics.105.049510

Gianola, D., and J. B. C. H. M. van Kaam, 2008 Reproducing Kernel Hilbert Spaces Regression Methods for Genomic Assisted Prediction of Quantitative Traits. Genetics 178: 2289–2303. https://doi.org/10.1534/genetics.107.084285

Gianola, D., G. de los Campos, W. G. Hill, E. Manfredi, and R. Fernando, 2009 Additive genetic variability and the Bayesian alphabet. Genetics 183: 347–363. https://doi.org/10.1534/genetics.109.103952

González-Camacho, J. M., G. de los Campos, P. Pérez, D. Gianola, J. E. Cairns *et al.*, 2012 Genome-enabled prediction of genetic values using radial basis function neural networks. Theor. Appl. Genet. 125: 759–771. https://doi.org/10.1007/s00122-012-1868-9

Habier, D., R. L. Fernando, and J. M. Dekkers, 2007 The impact of genetic relationship information on genome-assisted breeding values. Genetics 177: 2389–2397.

Habier, D., R. L. Fernando, K. Kizilkaya, and D. J. Garrick, 2011 Extension of the bayesian alphabet for genomic selection. BMC Bioinformatics 12: 186. https://doi.org/10.1186/1471-2105-12-186

Hallauer, A. R., M. J. Carena, and J. B. Miranda Filho, 2010 *Quantitative Genetics in Maize Breeding*, Springer, New York.

Henderson, C. R., 1984 *Applications of Linear Models in Animal Breeding*, University of Guelph. Guelph, Ontario.

Hickey, J. M., and G. Gorjanc, 2012 Simulated Data for Genomic Selection and Genome-Wide Association Studies Using a Combination of Coalescent and Gene Drop Methods. G3 Genes, Genomes. Genet. 2: 425–427.

Kayihan, G. C., D. A. Huber, A. M. Morse, T. L. White, and J. M. Davis, 2005 Genetic dissection of fusiform rust and pitch canker disease traits in loblolly pine. Theor. Appl. Genet. 110: 948–958. https://doi.org/10.1007/s00122-004-1915-2

Kayihan, G. C., C. D. Nelson, D. A. Huber, H. V. Amerson, T. L. White *et al.*, 2010 Clonal evaluation for fusiform rust disease resistance: effects of pathogen virulence and disease escape. Can. J. For. Res. 40: 1042–1050. https://doi.org/10.1139/X10-045

de los Campos, G., D. Gianola, and G. J. M. Rosa, 2009a Reproducing kernel Hilbert spaces regression: a general framework for genetic evaluation. J. Anim. Sci. 87: 1883–1887. https://doi.org/10.2527/jas.2008-1259

De los Campos, G., D. Gianola, G. J. M. Rosa, K. A. Weigel, and J. Crossa, 2010 Semi-parametric genomic-enabled prediction of genetic values using reproducing kernel Hilbert spaces methods. Genet. Res. 92: 295–308. https://doi.org/10.1017/S0016672310000285

de los Campos, G., J. M. Hickey, R. Pong-Wong, H. D. Daetwyler, and M. P. L. Calus, 2013 Whole-Genome Regression and Prediction Methods Applied to Plant and Animal Breeding. Genetics 193: 327–345. https://doi.org/10.1534/genetics.112.143313

de los Campos, G., H. Naya, D. Gianola, J. Crossa, A. Legarra *et al.*, 2009b Predicting quantitative traits with regression models for dense molecular markers and pedigree. Genetics 182: 375–385. https://doi.org/10.1534/genetics.109.101501

de los Campos, G., and P. R. Perez, 2014 BGLR: Bayesian Generalized Linear Regression.

de los Campos, G., D. Sorensen, and D. Gianola, 2015 Genomic Heritability: What Is It? PLoS Genet. 11: e1005048. https://doi.org/10.1371/journal.pgen.1005048

Lynch, M., and B. Walsh, 1998 *Genetics and Analysis of Quantitative Traits*, Sinauer, Sunderland, MA.

Meuwissen, T. H. E., B. J. Hayes, and M. E. Goddard, 2001 Prediction of Total Genetic Value Using Genome-Wide Dense Marker Maps. Genetics 157: 1819–1829.

Morota, G., P. Boddhireddy, N. Vukasinovic, D. Gianola, and S. DeNise, 2014 Kernel-based variance component estimation and whole-genome prediction of pre-corrected phenotypes and progeny tests for dairy cow health traits. Front. Genet. 5: 56. https://doi.org/10.3389/fgene.2014.00056

Morota, G., and D. Gianola, 2014 Kernel-based whole-genome prediction of complex traits: a review. Front. Genet. 5: 363. https://doi.org/10.3389/fgene.2014.00363

Morota, G., M. Koyama, G. J. M. Rosa, K. A. Weigel, and D. Gianola, 2013 Predicting complex traits using a diffusion kernel on genetic markers with an application to dairy cattle and wheat data. Genet. Sel. Evol. 45: 17. https://doi.org/10.1186/1297-9686-45-17

Mrode, R. A., 2014 *Linear Models for the Prediction of Animal Breeding Values*, CABI Pub. https://doi.org/10.1079/9781780643915.0000

Muñoz, P. R., M. F. R. Resende, S. A. Gezan, M. D. V. Resende, G. de los Campos *et al.*, 2014 Unraveling Additive from Non-Additive Effects Using Genomic Relationship Matrices. Genetics 198: 1759–1768. https://doi.org/10.1534/genetics.114.171322

Nishio, M., and M. Satoh, 2014 Including Dominance Effects in the Genomic BLUP Method for Genomic Evaluation. PLoS One 9: e85792. https://doi.org/10.1371/journal.pone.0085792

Pérez, P., and G. de los Campos, 2014 Genome-Wide Regression & Prediction with the BGLR Statistical Package. Genetics 198: 483–495. https://doi.org/10.1534/genetics.114.164442

Quesada, T., M. F. R. Resende, Jr, P. Muñoz, J. Wegrzyn, D. Neale *et al.*, 2014 Mapping Fusiform Rust Resistance Genes within a Complex Mating Design of Loblolly Pine. Forests 5: 347–362. https://doi.org/10.3390/f5020347

Resende, M. F. R., Jr, P. Muñoz, J. J. Acosta, G. F. Peter, J. M. Davis *et al.*, 2012a Accelerating the domestication of trees using genomic selection: accuracy of prediction models across ages and environments. New Phytol. 193: 617–624. https://doi.org/10.1111/j.1469-8137.2011.03895.x

Resende, M. F. R., Jr, P. Muñoz, M. D. V. Resende, D. J. Garrick, R. L. Fernando *et al.*, 2012b Accuracy of genomic selection methods in a standard data set of loblolly pine (Pinus taeda L.). Genetics 190: 1503–1510. https://doi.org/10.1534/genetics.111.137026

Resende, M. D. V., M. F. R. Resende, Jr, C. P. Sansaloni, C. D. Petroli, A. A. Missiaggia *et al.*, 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. https://doi.org/10.1111/j.1469-8137.2011.04038.x

Resende, R. T., M. D. V. Resende, F. F. Silva, C. F. Azevedo, E. K. Takahashi *et al.*, 2017 Assessing the expected response to genomic selection of individuals and families in Eucalyptus breeding with an additive-dominant model. Heredity 119: 245–255. https://doi.org/10.1038/hdy.2017.37

Toro, M. A., and L. Varona, 2010 A note on mate allocation for dominance handling in genomic selection. Genet. Sel. Evol. 42: 33. https://doi.org/10.1186/1297-9686-42-33

Tusell, L., P. Pérez-Rodríguez, S. Forni, and D. Gianola, 2014 Model averaging for genome-enabled prediction with reproducing kernel Hilbert spaces: a case study with pig litter size and wheat yield. J. Anim. Breed. Genet. 131: 105–115. https://doi.org/10.1111/jbg.12070

Vazquez, A. I., G. J. M. Rosa, K. A. Weigel, G. de los Campos, D. Gianola *et al.*, 2010 Predictive ability of subsets of single nucleotide polymorphisms with and without parent average in US Holsteins. J. Dairy Sci. 93: 5942–5949. https://doi.org/10.3168/jds.2010-3335

Wiggans, G. R., P. M. Vanraden, and T. A. Cooper, 2011 The genomic evaluation system in the United States: past, present, future. J. Dairy Sci. 94: 3202–3211. https://doi.org/10.3168/jds.2010-3866

Willyard, A., J. Syring, D. S. Gernandt, A. Liston, and R. Cronn, 2007 Fossil calibration of molecular divergence infers a moderate mutation rate and recent radiations for pinus. Mol. Biol. Evol. 24: 90–101. https://doi.org/10.1093/molbev/msl131

Xavier, A., W. M. Muir, and K. M. Rainey, 2016 Assessing Predictive Properties of Genome-Wide Selection in Soybeans. G3 Genes, Genomes. Genet. 6: 2611–2616.

Zeng, J., A. Toosi, R. L. Fernando, J. C. Dekkers, and D. J. Garrick, 2013 Genomic selection of purebred animals for crossbred performance in the presence of dominant gene action. Genet. Sel. Evol. 45: 11. https://doi.org/10.1186/1297-9686-45-11

Communicating editor: J.-L. Jannink