Title: Accuracy of Genomic Selection Methods in a Standard Dataset of Loblolly Pine

(Pinus taeda L.)

Authors: Resende, M.F.R., Jr.*1,3, Muñoz, P.*2,3, Resende, M.D.V.4,5, Garrick, D.J.6, Fernando, R.L.6, Davis, J.M.3,7, Jokela, E.J.3, Martin, T.A.3, Peter, G.F.3,7, Kirst, M.3,7

* Both authors contributed equally

1Genetics and Genomics Graduate Program, University of Florida, P.O. Box 103610, Gainesville, FL 32611, USA

2Plant Molecular and Cellular Biology Graduate Program, University of Florida, P.O. Box 110690, Gainesville, FL 32611, USA

3School of Forest Resources and Conservation, University of Florida, P.O. Box 110410, Gainesville, FL 32611, USA

4EMBRAPA Forestry, Estrada da Ribeira, km 111 Caixa Postal 319, Colombo, PR 83411-000 Brazil;

5Department of Forest Engineering, Universidade Federal de Viçosa, Viçosa, MG 36571-000 Brazil

6Department of Animal Science, Iowa State University, Ames, IA 50011, USA

7University of Florida Genetics Institute, University of Florida, P.O. Box 103610, Gainesville, FL 32611, USA

Author for correspondence:

Matias Kirst

Phone: +1 352 846 0900

Fax: +1 352 846 1277

E-mail: mkirst@ufl.edu

Copyright 2012.
Keywords: Genomic selection, shared data resources, *Pinus taeda*, complex traits, fusiform rust, loblolly pine
ABSTRACT

Genomic selection can increase genetic gain per generation through early selection. Genomic selection is expected to be particularly valuable for traits that are costly to phenotype, and expressed late in the life-cycle of long-lived species. Alternative approaches to genomic selection prediction models may perform differently for traits with distinct genetic properties. Here the performance of four different original methods of genomic selection that differ with respect to assumptions regarding distribution of marker effects, including (i) Ridge Regression – Best Linear Unbiased Prediction (RR-BLUP), (ii) Bayes A, (iii) Bayes Cπ, and (iv) Bayesian LASSO are presented. In addition, a modified RR-BLUP (RR-BLUP B) that utilizes a selected subset of markers was evaluated. The accuracy of these methods was compared across 17 traits with distinct heritabilities and genetic architectures, including growth, development and disease-resistance properties, measured in a Pinus taeda (loblolly pine) training population of 951 individuals genotyped with 4,853 SNPs. The predictive ability of the methods was evaluated using a 10-fold, cross-validation approach, and differed only marginally for most method/trait combinations. Interestingly, for fusiform rust disease-resistance traits Bayes Cπ, Bayesian LASSO and RR-BLUP B had higher predictive ability than RR-BLUP and Bayes A. Fusiform rust is controlled by few genes of large effect. A limitation of RR-BLUP is the assumption of equal contribution of all markers to the observed variation. The genotypic and phenotypic data used in this study is publically available for comparative analysis of genomic selection prediction models.
INTRODUCTION

Plant and animal breeders have effectively used phenotypic selection to increase the mean performance in selected populations. For many traits, phenotypic selection is costly and time-consuming, especially so for traits expressed late in the life cycle of long-lived species. Genome-Wide Selection (GWS) (Meuwissen et al. 2001) was proposed as an approach to accelerate the breeding cycle. In GWS, trait-specific models predict phenotypes using dense molecular markers from a base population. These predictions are applied to genotypic information in subsequent generations to estimate Direct Genetic Values (DGV).

Several analytical approaches have been proposed for genome-based prediction of genetic values, and these differ with respect to assumptions about the marker effects (de los Campos et al. 2009a; Habier et al. 2011; Meuwissen et al. 2001). For example, Ridge Regression – Best Linear Unbiased Prediction (RR-BLUP) assumes that all marker effects are normally distributed, and that these marker effects have identical variance (Meuwissen et al. 2001). In Bayes A, markers are assumed to have different variances, and are modeled as following a scaled inverse chi-square distribution (Meuwissen et al. 2001). The prior in Bayes B (Meuwissen et al. 2001) assumes the variance of markers to equal zero with probability $\pi$, and the complement with probability $(1 - \pi)$ follows an inverse chi-square distribution, with $\nu$ degree of freedom and scale parameter $S$. The definition of the probability $\pi$ depends on the genetic architecture of the trait, suggesting an improvement to the Bayes B model, known as Bayes C$\pi$. In Bayes C$\pi$, the mixture probability $\pi$ has a prior uniform distribution (Habier et al. 2011). A drawback of Bayesian methods is the need for the definition of priors. The requirement of a prior for
the parameter $\pi$ is circumvented in the Bayesian LASSO method, which needs less information (de los Campos et al. 2009b; Legarra et al. 2011b). Methods for genomic prediction of genetic values may perform differently for different phenotypes (Habier et al. 2011; Meuwissen et al. 2001; Usai et al. 2009) and results may diverge because of differences in genetic architecture among traits (Grattapaglia and Resende 2011; Hayes et al. 2009). Therefore, it is valuable to compare performance among methods with real data and identify those which provide more accurate predictions.

Recently, GWS was applied to agricultural crops (Crossa et al. 2010) and trees (Resende et al. 2011). Here we report, for an experimental breeding population of the tree species loblolly pine (Pinus taeda L.), a comparison of GWS predictive models for 17 traits with different heritabilities and predicted genetic architectures. Genome-wide selection models included RR-BLUP, Bayes A, Bayes $C\pi$ and the Bayesian LASSO. In addition, we evaluated a modified RR-BLUP method that utilizes a subset of selected markers, RR-BLUP B. We show that, for most traits, there is limited difference among these four original methods in their ability to predict GBV. Bayes $C\pi$ performed better for fusiform rust resistance – a disease-resistance trait shown previously to be controlled in part by major genes and the proposed method RR-BLUP B was similar to or better than Bayes $C\pi$ when a subsample of markers was fitted to the model.

**MATERIALS AND METHODS**

**Training population and genotypic data:** The loblolly pine population used in this analysis is derived from 32 parents representing a wide range of accessions from the Atlantic Coastal Plain, Florida and Lower Gulf of the United States. Parents were crossed
in a circular mating design with additional off-diagonal crosses, resulting in 70 full-sib families with an average of 13.5 individuals per family (Baltunis et al. 2007a). This population is referred to hereafter as CCLONES (Comparing Clonal Lines ON Experimental Sites). A subset of the CCLONES population, composed of 951 individuals from 61 families (mean = 15, standard deviation = 2.2) was genotyped using an Illumina Infinium™ assay (Illumina, San Diego, CA; Eckert et al. 2010) with 7,216 SNP, each representing a unique pine EST contig. A subset of 4,853 SNPs were polymorphic in this population, and were used in this study. None of the markers were excluded based on minimum allele frequency. Genotypic data and pedigree information are available in the Supporting Information (File S1 and S2).

**Phenotypic data:** The CCLONES population was phenotyped for growth, developmental and disease resistance traits in three replicated studies. The first was a field study established using single-tree plots in eight replicates (one ramet of each individual is represented in each replicate) that utilized a resolvable alpha-incomplete block design (Williams et al. 2002). In that field trial, four replicates were grown under a high intensity and four were grown under a standard silvicultural intensity regime. The traits stem diameter (DBH, cm), total stem height (HT, cm) and total height to the base of the live crown (HTLC, cm) were measured in the eight replicates at years 6, 6 and 4, respectively. At year 6, crown width across the planting beds (CWAC, cm), crown width along the planting beds (CWAL, cm), basal height of the live crown (BLC, cm), branch angle average (BA, degrees) and average branch diameter (BD, cm) were measured only in the high intensity silvicultural treatment. Phenotypic traits tree stiffness (Stiffness, km²/s²), lignin content (Lignin), latewood percentage at year 4 (LateWood), wood
specific gravity (Density), and 5- and 6- carbon sugar content (C5C6) were measured only in two repetitions, in the high intensity culture (Baltunis et al. 2007a; Emhart et al. 2007; Li et al. 2007; Sykes et al. 2009).

The second study was a greenhouse disease resistance screen. The experimental design was a randomized complete block with single-tree plots arranged in an alpha lattice with an incomplete block (tray container). Fusiform rust (*Cronartium quercuum* Berk. Miyable ex Shirai f. sp. *fusiforme*) susceptibility was assessed as gall volume (Rust_gall_vol) and presence or absence of rust (Rust_bin) (Kayihan et al. 2005; Kayihan et al. 2010).

Finally, in the third study the rooting ability of cuttings was investigated in an incomplete block design (tray container) with four complete repetitions, in a controlled greenhouse environment. Root number (Rootnum) and presence or absence of roots (Rootnum_bin) were quantified (Baltunis et al. 2005; Baltunis et al. 2007b).

**Breeding value prediction:** Analyses were carried out using ASReml v.2 (Gilmour 2006) with the following mixed linear model:

\[ y = Xb + Z_1 i + Z_2 a + Z_3 c + Z_4 f + Z_5 d_1 + Z_6 d_2 + e \]

Where \( y \) is the phenotypic measure of the trait being analyzed, \( b \) is a vector of the fixed effects, \( i \) is a vector of the random incomplete block effects within replication \( \sim N(0, I_0 \sigma^2_{ibk}) \), \( a \) is a vector of random additive effects of clones, \( \sim N(0, A \sigma^2_a) \), \( c \) is a vector of random non-additive effects of clones \( \sim N(0, I_0 \sigma^2_c) \), \( f \) is a vector of random family effects \( \sim N(0, I_0 \sigma^2_f) \), \( d_1 \) and \( d_2 \) are described below, \( e \) is the vector of random residual effects \( \sim N(0, \text{DIAG}_e) \), \( X \) and \( Z_1-Z_6 \) are incidence matrices and \( I, A \) and \( \text{DIAG} \) are the identity, numerator relationship and block diagonal matrices, respectively. For traits measured in
the field study under both high and standard culture intensities, the model also included $d_1$, a vector of the random additive x culture type interaction $\sim N(0, \text{DIAG}\sigma^2_{d1})$, and $d_2$, a vector of the random family x culture type interaction $\sim N(0, \text{DIAG}\sigma^2_{d2})$. Narrow sense heritability was calculated as the ratio of the additive variance $\sigma^2_a$ to the total or phenotypic variance (e.g. for the field experiment total variance was $\sigma^2_a+\sigma^2_n+\sigma^2_t+\sigma^2_{d1}+\sigma^2_{d2}+\sigma^2_e$). Prior to use in GWS modeling, the estimated breeding values were deregressed into phenotypes (DP) following the approach described in Garrick et al. (2009), to remove parental average effects from each individual. Breeding values and deregressed phenotypes are available in Supporting Information (Files S3 and S4).

**Statistical methods:** The SNP effects were estimated based on five different statistical methods: RR-BLUP, Bayes A (Meuwissen et al. 2001), Bayes Cπ (Habier et al. 2011), the Improved Bayesian LASSO (BLASSO) approach proposed by Legarra et al. (2011b) and RR-BLUP B (a modified RR-BLUP). In all cases the genotypic information was fitted using the model:

$$DP = 1\beta + Zm + \varepsilon$$

Where DP is the vector of phenotypes deregressed from the additive genetic values (Garrick et al. 2009), $\beta$ is the overall mean fitted as a fixed effect, $m$ is the vector of random marker effects and $\varepsilon$ is the vector of random error effects, 1 is a vector of ones and $Z$ is the incidence matrix $m$, constructed from covariates based on the genotypes. No additional information, such as marker location, polygenic effects, or pedigree was used in those models.
Once the marker effects were estimated using one of the methods, the predicted direct genetic value (DGV) of individual $j$ for that method was given by:

$$\hat{g}_j = \sum^n_i Z_{ij} \hat{m}_i$$

Where $i$ is the specific allele of the $i$th marker on individual $j$ and $n$ is the total number of markers.

**Random Regression Best Linear Unbiased Predictor (RR-BLUP):** The RR-BLUP assumed the SNP effects, $m$, were random (Meuwissen et al. 2001). The variance parameters were assumed to be unknown and were estimated by restricted maximum likelihood (REML), which is equivalent to Bayesian inference using an uninformative, flat prior. The first and second moments for this model are described below:

- $m \sim (0, G = I\sigma^2_m)$  
  $E(y) = 1\beta$
- $\epsilon \sim (0, R = I\sigma^2_\epsilon)$  
  $Var(y) = V = ZGZ' + R$

Where $\sigma^2_m$ is the variance common to each marker effect and $\sigma^2_\epsilon$ is the residual variance.

The mixed model equation for the prediction of $m$ is equivalent to:

$$\begin{bmatrix} \sum X'X & X'Z \\ Z'X & Z'Z + I\frac{\sigma^2_\epsilon}{\sigma^2_a/\eta} \end{bmatrix} \begin{bmatrix} \hat{\beta} \\ \hat{m} \end{bmatrix} = \begin{bmatrix} X'y \\ Z'y \end{bmatrix}$$

Where $\sigma^2_a$ refers to the total additive variance of the trait and $\eta$, due to standardization of the $Z$ matrix, refers to the total number of markers (Meuwissen et al. 2009). The matrix $Z$ was parameterized and standardized to have a mean of zero and variance of one as previously described (Resende et al. 2010; Resende et al. 2011). The analyses were performed in the software R (available at CRAN, http://cran.r-project.org/) and the script is available in the Supporting Information (File S5).
**Bayes A:** The Bayes A method proposed by Meuwissen et al. (2001) assumes the conditional distribution of each effect (given its variance) to follow a normal distribution. The variances are assumed to follow a scaled inverse chi-square distribution with degrees of freedom $\nu_a$ and scale parameter $S^2a$. The unconditional distribution of the marker effects can be shown to follow a t-distribution with mean zero (Sorensen and Gianola 2002). Bayes A differs from RR-BLUP in that each SNP has its own variance. In this study, $\nu_a$ was assigned the value 4, and $S^2a$ was calculated from the additive variance according to Habier et al. (2011) as follows:

$$S^2_a = \sigma_a^2 (\nu_a - 2)$$

where:

$$\bar{\sigma}_a^2 = \frac{\sigma_s^2}{(1 - \pi) \sum_{k=1}^{k} 2p_k (1 - p_k)}$$

and $p_k$ is the allele frequency of the $k$-th SNP.

**Bayes Cπ:** Bayes Cπ was proposed by Habier et al. (2011). In this method, the SNP effects have a common variance, which follows a scaled inverse chi-square prior with parameters $\nu_a$, $S^2a$. As a result, the effect of a SNP fitted with probability $(1 - \pi)$ follows a mixture of multivariate Student’s t-distributions, $t(0, \nu_a, I S^2a)$, where $\pi$ is the probability of a marker having zero effect. Parameters $\nu_a$, and $S^2a$ were chosen as described for Bayes A. The $\pi$ parameter is treated as unknown with a uniform (0,1) prior distribution.

Bayes A and Bayes Cπ were performed using the software GenSel (Fernando and Garrick 2008); available at http://bigs.ansci.iastate.edu/bigsgui/) for which an R package is available in the Supporting Information (File S6). The marker input file was coded as -
10, 0 and 10 for marker genotypes 0,1 and 2, respectively. A total of 50,000 iterations were used, with the first 2,000 excluded as the burn-in.

**Bayesian LASSO:** The Bayesian LASSO method was performed as proposed by Legarra *et al.* (2011b), using the same model equation used previously for the estimation of the markers effects. However, in this case:

\[ m|\lambda \sim \prod_i \frac{\lambda}{2} \exp(-\lambda |a_i|); \quad e|\sigma_e^2 \sim MVN(0, I \sigma_e^2) \]

\[ \text{var}(m) = \frac{2}{\lambda^2} \]

Using a formulation in terms of an augmented hierarchical model including an extra variance component \( \tau_i^2 \) associated to each marker locus, we have:

\[ p(m|\tau) \sim N(0, D); \quad \text{diag}(D) = (\tau_1^2 \ldots \tau_n^2) \]

\[ p(\tau|\lambda) = \prod_i \left( \frac{\lambda^2}{2} \right) \exp\left( \frac{-\lambda^2 \tau_i^2}{2} \right) \]

Therefore, \( \text{Var}(m_i) = \sigma_{mi}^2 = \tau_i^2 \)

The prior distribution for \( \sigma_e^2 \) was an inverted chi-square distribution with 4 degrees of freedom and expectations equal to the value used in regular genetic evaluation for \( \sigma_e^2 \).

Analyses were performed using the software GS3 (Legarra *et al.* 2011a); available in http://snp.toulouse.inra.fr/~alegarra/. The chain length was 100,000 iterations, with the first 2,000 excluded as the burn-in and a thinning interval of 100.

**RR-BLUP B:** We also evaluated a modified, two-step RR-BLUP method that reduces the number of marker effects estimated. In this case, the DGV for each trait was generated based on a reduced subset of markers. In order to define the number of markers in the subset, the marker effects from the RR-BLUP were ranked in decreasing order by
their absolute values, and grouped in multiples of 10 (10, 20, 30, ..., 4800). Each group was used, with their original effects, to estimate DGV. The size, $q$, of the subset that maximized the predictive ability was selected as the optimum number of marker effects to be used in subsequent analyses. Next, markers effects were re-estimated in a second RR-BLUP, using only the selected $q$ markers within each training partition (see below). The estimated effects derived from this analysis were used to predict the merit of the individuals in the validation partition that were not present in the training partition. This process was repeated for different allocations of the data into training and validation partitions. In each validation, a different subset of markers was selected, based on the highest absolute effects within that training partition. Therefore, the only restriction applied to the second analysis was related to $q$, the number of markers to be included in each dataset. The same approach was performed with two additional subsets of markers of the same size as a control: the first subset contained randomly selected markers and the second subset contained markers with the smallest absolute effect values.

**Validation of the models:** Two cross validation schemes were tested in the RR-BLUP method: 10-fold and leave-one-out. For the 10-fold cross validation approach a random sub-sampling partitioning, fixed for all methods, was used (Kohavi 1995). Briefly, the data for each trait were partitioned into two subsets. The first one was composed by the majority of the individuals (90%) and was used to estimate the marker effects. The second one, the validation partition (10%), had their phenotypes predicted based on the marker effects estimated in the training set. Randomly taken samples of $N = (9/10) \times N_T$ individuals were used as training sets, while the remaining individuals were used for validation ($N_T= \text{total number of individuals in the population}$). The process was
repeated ten times, each time with a different set of individuals as the validation partition, until all individuals had their phenotypes predicted (Legarra et al. 2008; Usai et al. 2009; Verbyla et al. 2010). In the leave-one-out approach, a model was constructed using $N_T-1$ individuals in the training population and validated in a single individual that was not used in the training set. This was repeated $N_T$ times, such that each individual in the sample was used once as the validation individual. This method maximized the training population size.

**Accuracy of the models:** The correlation between the Direct Genetic Values (DGV) and the Deregressed Phenotype (DP) was estimated using the software ASReml v.2 (Gilmour 2006) from a bivariate analysis, including the validation groups as fixed effects since each validation group had DGV estimated from a different prediction equation and might have had a different mean. This correlation represented the predictive ability ($r_{y\hat{g}}$) of GS to predict phenotypes, and was theoretically represented (Resende et al. 2010) by:

$$r_{y\hat{g}} = r_{g\hat{g}} h$$

where $r_{g\hat{g}}$ was the accuracy of GS and $h$ was the square root of the heritability of adjusted phenotypes, which is associated to Mendelian sampling effects and is given by

$$h_m^2 = \frac{n 0.5 \sigma_a^2}{n 0.5 \sigma_a^2 + \sigma_e^2}$$

where $n$ was the number of ramets used in each study. To remove the influence of the heritability upon the predictive ability and thus estimate the accuracy, the following formula was applied

$$r_{g\hat{g}} = \frac{r_{y\hat{g}}}{h}$$
In addition, for each method and trait, the slope coefficient for the regression of DP on DGV was calculated as a measurement of the bias of the DGV. Unbiased models are expected to have a slope coefficient of 1, whereas values greater than 1 indicate a biased underestimation in the DGV prediction and values smaller than indicate a biased overestimation of the DGV.

**RESULTS**

**Cross validation method:** Testing the effect of cross validation using two methods, 10-fold and leave-one-out (N-fold), showed that their predictive ability was not significantly different (Supporting Information – Table S1). The largest difference was detected for the trait CWAC, where the leave-one-out method outperformed the 10-fold cross validation by 0.02 (standard error = 0.03). Likewise, no significant differences were found for bias of the regressions (slope) in both methods (Supporting Information – Table S2). Thus, the 10-fold approach was selected and used for comparing all methods.

**Predictive ability of the methods:** Four well-established genome-wide selection methods were compared in 17 traits with heritabilities ranging from 0.07 to 0.45. Overall, the ability to predict phenotype ($r_{y,y}$) ranged from 0.17 for Lignin to 0.51 for BA (Table 1). Although the methods differ in *a priori* assumptions about marker effects, their predictive ability was similar – no significant differences were detected for any of the 17 traits. The standard errors for each method and trait are described in Supporting Information – Table S3.

Bayesian approaches performed better for traits in the disease resistance category. For Rust_bin, the methods Bayes A and Bayes Cπ were 0.05 superior than RR-BLUP and
0.06 superior to BLASSO. For Rust_gall_vol, Bayes Cπ was 0.05 superior to RR-BLUP and BLASSO. The accuracy \( r_{g\hat{g}} \) for each genome-wide prediction method was also estimated, and varied from 0.37–0.77 (Supporting Information – Table S4).

For all methods, the ability to predict phenotypes \( r_{y\hat{g}} \) was linearly correlated with trait heritability. The strongest correlation (0.79) was observed for RR-BLUP (Figure 1). The correlation is expected, as traits with lower heritability have phenotypes less reflective of their genetic content, and are expected to be less predictable through genomic selection.

**Bias of the methods:** The coefficient of regression (slope) of DP on DGV was calculated as a measurement of the bias of each method. Ideally, a value of beta equal to one indicates no bias in the prediction. For all traits, the slopes of all the models were not significantly different than one, indicating no significant bias in the prediction. In addition, no significant differences among the methods were detected (Supporting Information – Table S5). Although no evidence of significant bias was detected, the value of beta derived from RR-BLUP was slightly higher for all traits (average across traits equal to 1.18).

**Markers Subset and RR-BLUP B:** Prediction of phenotype was also performed with RR-BLUP, but adding increasingly large marker subsets, until all markers were used jointly in the prediction. The predictive ability was plotted against the size of the subset of markers (Figure 2). The pattern of the prediction accuracy was similar for 13 out 17 traits (Figure 2A), where differences were mainly found in the rate with which the correlation reached the asymptote. In these cases, the size of the subset ranged from 820 to 4,790 markers. However, a distinct pattern was detected for disease resistance related
traits, density and CWAL (Figure 2B). In these cases, maximum predictive ability was reached with smaller marker subsets (110-590 markers), and decreased with the addition of more markers. An additional RR-BLUP was performed using as covariates only the marker subset for which maximum predictive ability was obtained. For traits where a large number of markers (> 600) explain the phenotypic variability, RR-BLUP B was similar to RR-BLUP or Bayesian methods (Supporting Information – Table S6). However, for traits where the maximum predictive ability (Density, Rust_bin, Rust_gall_vol) was reached with a smaller number of markers (<600), RR-BLUP B performed significantly better than RR-BLUP. For example, the predictive ability of the trait Rust_gall_vol was 61% higher using RR-BLUP B (0.37) compared to the traditional RR-BLUP (0.23), and also improved relative to BLASSO (0.24), Bayes A (0.28) and Bayes Cπ (0.29).

We also contrasted these results with the predictive ability using a subset of markers of similar size, but selected either randomly or to include those markers with lower effects. As expected, for the three traits the predictive ability was larger for the subset selected by RR-BLUP B over the subsets with lower effects and random effects (Figure 3). A significant difference over the lower and random subsets was found for rust resistance related traits (Rust_bin, Rust_gall_vol), while for Density the markers selected by RR-BLUP B were only significantly different than the lower marker subset but not different to the random marker subset.

DISCUSSION
We characterized the performance of RR-BLUP/RR-BLUP B, Bayes A, Bayesian LASSO regression and Bayes C_\pi for GWS of growth, developmental, disease resistance and biomass quality traits in common dataset generated from an experimental population of the conifer loblolly pine. In general, the methods evaluated differed only modestly in their predictive ability (defined by the correlation between the DGV and DP).

The suitability of different methods of developing GWS predictive models is expected to be trait dependent, conditional on the genetic architecture of the characteristic. RR-BLUP differs from the other approaches used in this study in that the unconditional variance of marker effects is normally distributed, with the same variance for all markers (Meuwissen et al. 2001). This assumption may be suitable when considering an infinitesimal model (Fisher 1918), where the characters are determined by an infinite number of unlinked and non epistatic loci, with small effect. Not surprisingly, BLUP-based methods underperformed relative to Bayesian approaches for oligogenic traits. For instance, Verbyla et al. (2009) showed that BLUP-based GWS had lower accuracy, compared to Bayesian methods, in prediction of fat percentage in a population where a single gene explains ~50% of the genetic variation. Similarly, our observation that Bayes A and Bayes C_\pi were more accurate in predicting fusiform rust resistance traits, compared to RR-BLUP, may reflect a simpler genetic architecture, with a few loci of large effect. While the causative genes that regulate fusiform rust resistance have not yet been uncovered, several genetic studies support the role of few major genes in the trait variation. For example, the Fr1 locus confers resistance to specific fungus aeciospore isolates (Wilcox et al. 1996), and at least five families within the CCLONES population segregate for this locus (Kayihan et al. 2010).
The under-performance of RR-BLUP for predicting oligogenic traits is a consequence of fitting a large number of markers to model variation at a trait controlled by few major loci, leading to model over-parameterization. In Bayes A and Bayes Cπ, the shrinkage of effects is marker-specific, while in BLUP all markers are penalized equally. To address this limitation, we proposed an alternative, RR-BLUP B, to Bayesian and the traditional RR-BLUP approaches, aimed at reducing the number of parameters. In RR-BLUP B, marker effects are initially estimated and ranked using RR-BLUP. Next, increasing markers subsets that include initially those with larger effect are used to estimate DGV. The number of markers that maximizes the predictive ability is then defined, and used in a second RR-BLUP model. For rust disease resistance and wood density traits, the modified RR-BLUP B approach performed better than traditional RR-BLUP, and as well as the Bayesian methods. Previous studies using simulated data have shown that improvements in predictive ability could be obtained by using a similar approach to the one proposed here (Zhang et al. 2010, Zhang et al. 2011), although with a different strategy of marker selection. While RR-BLUP B may add an additional step to the development of predictive models (i.e. initial marker selection), it is overall simpler and computationally less expensive than Bayesian approaches. Therefore, it may provide a suitable alternative to the use of BLUP-based methods for traits that do not fit an infinitesimal model, and are rather regulated by few major loci.

ACKNOWLEDGEMENTS

The authors wish to thank members of the Forest Biology Research Cooperative (FBRC) for their support in establishing, maintaining and measuring field trials used in this study.
The work was supported by the National Science Foundation Plant Genome Research Program (award no. 0501763 to GFP, JMD and MK) and the US Department of Agriculture National Institute of Food and Agriculture Plant Breeding and Education Program (award no. 2010-85117-20569 to MK, GFP and JMD). We also acknowledge the valuable input from two anonymous reviewers.

REFERENCES


FIGURE LEGENDS

Figure 1 – Regression of RR-BLUP predictive ability on narrow sense heritability for 17 traits (trend line is shown, $R^2=0.79$).

Figure 2 – Example of the two patterns of predictive ability observed among traits, as an increasing number of markers is added to the model. Each marker group is represented by a set of 10 markers. A) For DBH, the maximum predictive ability was detected when 380 groups of markers (3,800 markers) were included in the model. B) For the trait Rust_gall_vol, predictive ability pattern reached a maximum when only 10 groups (100 markers) were added. Lines indicate the predictive ability of RR-BLUP (filled line), Bayes C$\pi$ (dashed line), and RR-BLUP B (dotted line) as reported on Table 1 and in Supporting Information - Table S6.

Figure 3 – Predictive ability for subsets of 310 markers for Rust_bin, 110 markers for Rust_gall_vol and 240 markers for Density. Subsets were generated selecting markers with the lowest absolute effects (light grey), with random values (grey), including all markers (dark grey), and including only those markers with largest absolute effects (black).
TABLE

Table 1 – Predictive ability of Genomic Selection models using four different methods; $h^2$ is the narrow-sense heritability of the trait.

<table>
<thead>
<tr>
<th>Trait category</th>
<th>Trait</th>
<th>$h^2$</th>
<th>RR-BLUP</th>
<th>BLASSO</th>
<th>Bayes A</th>
<th>Bayes C$\pi$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Growth</td>
<td>HT</td>
<td>0.31</td>
<td>0.39</td>
<td>0.38</td>
<td>0.38</td>
<td>0.38</td>
</tr>
<tr>
<td></td>
<td>HTLC</td>
<td>0.22</td>
<td>0.45</td>
<td>0.44</td>
<td>0.44</td>
<td>0.44</td>
</tr>
<tr>
<td></td>
<td>BHLR</td>
<td>0.35</td>
<td>0.49</td>
<td>0.49</td>
<td>0.49</td>
<td>0.49</td>
</tr>
<tr>
<td></td>
<td>DBH</td>
<td>0.31</td>
<td>0.46</td>
<td>0.46</td>
<td>0.46</td>
<td>0.46</td>
</tr>
<tr>
<td>Development</td>
<td>CWAL</td>
<td>0.27</td>
<td>0.38</td>
<td>0.36</td>
<td>0.36</td>
<td>0.36</td>
</tr>
<tr>
<td></td>
<td>CWAC</td>
<td>0.45</td>
<td>0.48</td>
<td>0.46</td>
<td>0.47</td>
<td>0.47</td>
</tr>
<tr>
<td></td>
<td>BD</td>
<td>0.15</td>
<td>0.27</td>
<td>0.25</td>
<td>0.27</td>
<td>0.27</td>
</tr>
<tr>
<td></td>
<td>BA</td>
<td>0.33</td>
<td>0.51</td>
<td>0.51</td>
<td>0.51</td>
<td>0.51</td>
</tr>
<tr>
<td></td>
<td>Rootnum_bin</td>
<td>0.10</td>
<td>0.28</td>
<td>0.28</td>
<td>0.27</td>
<td>0.28</td>
</tr>
<tr>
<td></td>
<td>Rootnum</td>
<td>0.07</td>
<td>0.24</td>
<td>0.26</td>
<td>0.25</td>
<td>0.24</td>
</tr>
<tr>
<td>Disease</td>
<td>Rust_bin</td>
<td>0.21</td>
<td>0.29</td>
<td>0.28</td>
<td>0.34</td>
<td>0.34</td>
</tr>
<tr>
<td>resistance</td>
<td>Rust_gall_vol</td>
<td>0.12</td>
<td>0.23</td>
<td>0.24</td>
<td>0.28</td>
<td>0.29</td>
</tr>
<tr>
<td>Wood quality</td>
<td>Stiffness</td>
<td>0.37</td>
<td>0.43</td>
<td>0.39</td>
<td>0.42</td>
<td>0.42</td>
</tr>
<tr>
<td></td>
<td>Lignin</td>
<td>0.11</td>
<td>0.17</td>
<td>0.17</td>
<td>0.17</td>
<td>0.17</td>
</tr>
<tr>
<td></td>
<td>LateWood</td>
<td>0.17</td>
<td>0.24</td>
<td>0.24</td>
<td>0.23</td>
<td>0.24</td>
</tr>
<tr>
<td></td>
<td>Density</td>
<td>0.09</td>
<td>0.20</td>
<td>0.22</td>
<td>0.23</td>
<td>0.22</td>
</tr>
<tr>
<td></td>
<td>C5C6</td>
<td>0.14</td>
<td>0.26</td>
<td>0.25</td>
<td>0.25</td>
<td>0.25</td>
</tr>
</tbody>
</table>
FIGURE 2