FUNCTIONAL PLANT BREEDING JOURNAL
Scientific Journal. ISSN 2595-9433
Volume 1, Number 2, Article n. 6, November/December
D.O.I. http:// dx.doi.org/10.35418/2526-4117/v1n2a6
Accepted: 15/07/2019 Published: 02/12/19

COMPARATIVE ANALYSIS OF BAYESIAN AND FREQUENCY-BASED METHODS IN GENOMIC SELECTION FOR POPCORN POPULATION BREEDING AND OPTIMIZATION OF SNP MARKER DENSITY

Ismael Albino Schwantes[1], Antonio Teixeira do Amaral Júnior[1], Janeo Eustáquio de Almeida Filho[2], Pedro Henrique Araújo Diniz Santos[1], Fernando Rafael Alves Ferreira[1], Gabrielle Sousa Mafra[1], Marcelo Vivas[3], Yure Pequeno de Souza[1], Fabio Tomaz de Oliveira[1], Ismael Fernando Schegoscheski Gerhardt[1], Juliana Saltires Santos[1]

[1]*Laboratório de Melhoramento Genético Vegetal, Universidade Estadual do Norte Fluminense Darcy Ribeiro (UENF), Campos dos Goytacazes, RJ.*
[2] *Bayer Crop Science, Bayer, Passo Fundo, RS.*
[3] *Laboratório de Engenharia Agrícola, Universidade Estadual do Norte Fluminense Darcy Ribeiro (UENF), Campos dos Goytacazes, RJ.*

Corresponding author: Ismael Albino Schwantes (ismael.schwantes31@gmail.com)

Abstract - Bayesian methods and frequency-based approaches such GBLUP are used to estimate genomic genetic values in recurrent genomic selection. An important factor in genetic gain is prediction accuracy; therefore, the objective of the present study was to estimate the prediction accuracy of the following methods: GBLUP, Bayes A, Bayes B, Bayes Cπ, Bayes Lasso, and RKHS. After establishing the best method, different densities of SNP markers were tested. The experiment was implemented using an incomplete block design with three repetitions in two locations. Ninety-eight individuals were evaluated using 10,507 SNPs; the assessed traits were grain yield, popping expansion, and popcorn volume. The analyses were performed using R software and a ten-fold cross-validation system. The methods were compared using the *t*-test, via correlation networks and according to the time required to perform the analysis. The obtained results showed that the methods did not differ statistically with regard to selection accuracy, with high correlation estimates (<0.98); however, the GBLUP test stood out for its short analysis time and simplicity of execution. With regard to the distinct scenarios of SNP density tested with the GBLUP method, it was concluded that the use of ~ 4,800 SNPs led to results similar to those obtained with 10,507 SNPs. The GBLUP method is recommended for its greater speed in generating analytical results and for exhibiting robustness similar to that of other tested methods with the use of a small panel of SNPs.

Keywords: Genomic selection, popcorn breeding, SNP markers, GBLUP method

Introduction

Recurrent selection (RS) is one of the most important breeding methods to obtain popcorn varieties; however, it is a time consuming and effort intensive procedure as each selection cycle includes three steps: progeny development, evaluation, and recombination of superior families. Recurrent genomic selection (RGS) a method that uses the principles of genomic selection (GS) in RS can reduce the time required foreach RS cycle because it allows performing evaluation and recombination steps simultaneously. This method provides a form of direct early selection, i.e., it acts prematurely on genes expressed in adult hood (Resende et al., 2010; Fritsche-Neto et al., 2012). Thus, the time required for each selection cycle is reduced to just one crop season.

Meuwissen et al. (2001) marker haplotypes simultaneously from a limited number of phenotypic records. A genome of 1000cM was simulated with a marker spacing of 1cM. The markers surrounding every 1cM region were combined into marker haplotypes. Due to finite population size (Ne = 100 proposed GS, which is an approach that consists of predicting the genetic potential of individuals using data of a large number of markers widely distributed across the genome. This prediction is performed using models where genomic characterization that results from genotyping is used as a source of an explanatory variable, and the phenotype of a given trait of interest is used as the response variable. This genomic prediction model should be added to a training population (TP) where individuals are genotyped and phenotyped and if this prediction is accurate, the model can be applied to breeding populations related to the TP (Resende Jr et al., 2012; Almeida Filho et al., 2016) it is unknown how accurate genomic selection prediction models remain when used across environments and ages. This knowledge is critical for breeders to apply this strategy in genetic improvement. Here, we evaluated the utility of genomic selection in a *Pinus taeda* population of c. 800 individuals clonally replicated and grown on four sites, and genotyped for 4825 single-nucleotide polymorphism (SNP).

An important factor in genetic gain is the accuracy of the prediction; in GS, this accuracy may be maximized using the model that best fits the genetic complexity of the traits of interest (Rabier et al., 2016). A substantial range of statistical methods with distinct assumptions on the genetic architecture of the traits is available in the literature (Meuwissen et al., 2001; Gianola and van Kaam, 2008; de los Campos et al., 2009) marker haplotypes simultaneously from a limited number of phenotypic records. A genome of 1000 cM was simulated with a marker spacing of 1cM. The markers surrounding every 1cM region were combined into marker haplotypes. Due to finite population size (Ne=100). In addition, several studies have been published that compare the main GS methods (Pszczola et al., 2011; Heslot et al., 2012; Thavamanikumar et al., 2015), however, there are no studies that compare these methods in real popcorn breeding populations.

According to Crossa et al. (2017), the use of GS in the breeding of plants may be limited owing to two major factors: (i) the cost of genotyping and (ii) unclear guidelines on when (i.e. at which step) GS can be applied efficiently in the breeding program. The cost of genotyping decreases with the decreasing density of the SNP panel; thus, the use of a small number of SNPs allows genotyping a higher number of individuals using the same resource, which thereby increases selection intensity, and consequently, the genetic gain. Despite the importance of determining the correct density of SNPs for the routine application of GS, it is yet to be fully addressed in the literature on popcorn.

The objective of the present study is to estimate selection accuracy using the following methods: GBLUP, Bayes A, Bayes B, Bayes Cπ, Bayes Lasso (BL), and RKHS. Once the best method is established, different densities of SNP markers are tested to determine whether it is possible to obtain satisfactory accuracy values using a lower density of SNP markers; the aim is to optimize of the use of GS in the genetic popcorn breeding.

Materials and methods

Study population

The study population is taken from the UNB 2U population, which is an open-pollinated variety. Before being delivered to the Northern Fluminense State University Darcy Ribeiro (UENF), it consisted of an indigenous sample donated to the University of Brasilia (UNB) by the ESALQ/USP, thus getting the nameUNB-1. UNB-1was brought to the UENF by Professor Joachim Friedrich Wilhelm Von Bülowin 1993, and it was crossed with the South American Mushroom (SAM) popcorn variety. This first filial generation was then crossed with a variety of popcorn resistant to *Exserohilum turcicum* (*Helminthosporium*). Three backcrossings with the SAM variety were performed after two cycles of mass selection to obtain the UNB-2 population. The latter produced the UNB-2U population after two cycles of mass selection (Pereira and Amaral Júnior, 2001). Eight cycles of RS were then performed with the aim of gaining the main traits of economic importance popping expansion and grain yield. The first cycle was composed of full-sibs families (Daros et al., 2002); the second, S_1 families (Daros et al., 2004); the third, half-sibs families (Santos et al., 2008); and the

remaining, full-sibs families (Freitas Júnior et al., 2009; Rangel et al., 2011; Ribeiro et al., 2012; Freitas et al., 2014; Guimarães et al., 2018).

The UENF-14 population used in the present study was obtained after the eighth cycle of RS. The seeds of the resulting population were sown after the recombination of the selected progenies, thus initiating the ninth RS cycle that was used in the present study.

Genotyping

The characterization of the polymorphism in the genome among individuals of this population was performed using 200 DNA samples from the collected seedlings, and by using the Capture Seq method (Neves et al., 2014) with 5,000 probes evenly distributed in the corn reference genome. This genome sequencing approach was implemented in the collaboration with Rapid Genomics LLC (Florida, USA), and it resulted in 21,442 SNPs.

With this genotypic information and using the Plink software (Purcell et al., 2007), three filters were applied in the following sequence: a) exclusion of individuals with > 10% missing data; b) exclusion of SNPs with > 5% missing data; and c) exclusion of SNPs with minor allele frequency (MAF)< 5%. The application of these filters resulted in 196 individuals and 10,507 SNPs, of which 98 individuals were used in the S1 progeny trial.

Phenotyping

Two experiments were designed using the S1 families obtained from the above method. The first experiment was conducted in the State Agricultural College Antônio Sarlo, in Campos dos Goytacazes, in the northern region of the Rio de Janeiro State, located at 21º 45' S and 41º 20' W; it has an altitude of 11 m, with a climate classified as tropical rainforest with a mean annual rainfall of 1.023 mm, a potential annual evapotranspiration of 1.601 mm, and a mean annual temperature of 23 °C. The second experiment was conducted in the PESAGRO-RIO Experimental Station in Itaocara, located in the northwest region Fluminense, at 21º 39′ 12″ S and 42º 04′ 36″ W; it has an altitude of 60 m, with a mean annual temperature of 22.5 °C and a mean annual rainfall of 1.041 mm. Both experiments were set up in August 2016. An incomplete block design was used with three repetitions; the population density was 60.000 plants per hectare. Management practices were implemented according to the crop's requirements.

The phenotyped traits were the following: grain yield, expressed in kg ha^{-1}; ii) popping expansion, determined in a plastic bowl without oil, with three repetitions per plot, using samples of 30 g of grain. Popping was performed in the microwave oven for 2 min. Subsequently, the volume of expanded popcorn was measured in a 2.000 mL cylinder and the result was divided by the initial grain weight of 30 g and expressed in mL g^{-1}; and iii) volume of expanded popcorn per hectare (PV), obtained by multiplying the mean productivity of the plot by popping expansion, which yielded the mean volume of expanded popcorn per hectare of crop, expressed in m^3 ha^{-1} (PV=GY x PE/1000). The adjusted means (LS-Means) were estimated based on the phenotypical observations and accounting for crop moisture, initial stand, and experimental design.

Statistical methods and cross-validation

In this study, the prediction accuracies (for the traits under the study) of the following methods were compared: GBLUP, Bayes A, Bayes B, Bayes Cπ, BL, and RKHS. The t-test was used to identify the differences among the methods, with a confidence interval of 95%. A ten-fold cross-validation was used to assess the efficacy of the selection (Resende Jr et al., 2012; Almeida Filho et al., 2016;).In this process, the estimates were adjusted with 90% of the population's individuals, whereas the genetic merit of the remaining 10% was predicted considering only marker data. This model fitting process was repeated 10 times, and in each cycle, the genetic merit of a different group of individuals was predicted while ignoring their phenotype.

GBLUP

The genomic best linear unbiased predictor (GBLUP) method is a model of individuals where the pedigree-based kinship matrix is replaced by the kinship matrix, which is estimated using markers. The generation of genetic relationship matrix (GRM) markers with the same MAF contribute equally to genetic variance. Therefore, this model is in accordance with the infinitesimal model of inheritance, which is a widely used model in classical quantitative genetics and assumes that the traits are polygenic, with the number of genes tending to infinity and all contributing equally to phenotypic variation.

The adjusted GBLUP model is formulated as

$$y = \mu + g + \varepsilon \quad \text{Eq. 1}$$

Where **y** is the vector of phenotypic data (adjusted mean) of a given trait, **μ** is the intercept of the model, **g** is the random effect of the genomic estimated breeding value (GEBV), and **ε** is the effect of random error.

For this model, it was assumed that

$$g \sim N(0, G\alpha^2_A)$$

$$\varepsilon \sim N(0, I\,\alpha^2_\varepsilon), \quad \text{Eq. 2}$$

Where **I** and **G** are identity and GRM matrices, respectively. The components of variance were estimated with the AI-REML algorithm using the ASReml package (Butler, 2009) of the R software (R Core Team, 2013).

Bayesian methods

Alternatively, regression models that fit the SNPs as co-variables were compared (de los Campos et al., 2013; Gianola, 2013). The versatility of the regression models stems from different presumptions assumed by the effects of the markers, which are integrated into these models via Bayesian inference. The following general model is assumed.

$$y_j = \mu + \sum_{i=1}^{n^o\ SNPs} x_{ij}\, a_i + \varepsilon \quad \text{Eq. 3}$$

Where $\mathbf{y_j}$ is the phenotypic value previously corrected by experimental design (adjusted mean); **μ** is the intercept; ai is the effect of the allelic replacement of the marker **i**; x_{ij} is the co-variable related to the genotype of the individual **j** and to SNP **i**; x_{ij} is the number of the copies of the less frequent allele of individual **j** for marker **i** (2, 1, or 0), and ε_j is the effect of random error.

In general, the following may be assumed under the Bayesian approach (Pérez and de los Campos, 2014):

$$\mu \sim N(0,10^6);$$
$$\varepsilon_j | \sigma_e^2 \sim N(0, \sigma_\varepsilon^2$$
$$\sigma_\varepsilon^2 | gl, S_\varepsilon \sim \chi^{-2}(gl, S_\varepsilon). \quad \text{Eq. 4}$$

For the effect of allelic replacement (a_i), the assumptions were formulated according to the assumed priors, among the several priors available in the literature (Habier et al., 2011; De los Campos et al., 2013; Gianola, 2013; Pérez and de los Campos, 2014). A brief summary of each tested method is provided below.

Bayes A

The assumptions of the Bayes A method allow markers with the same MAF to have different contributions to genetic variance because the variances of the effect of the markers are heterogeneous (Pérez and De los Campos 2014) e.g., the number of marker effects. The Bayes A method was proposed by Meuwissen et al. (2001) and assumes that

$$a_i | \sigma_{a_i}^2 \sim N(0, \sigma_{a_i}^2);$$
$$\sigma_{a_i}^2 | gl, S_a \sim \chi^{-2}(gl, S_a);$$
$$S_a | r, s \sim G(r, s). \quad \text{Eq. 5}$$

Bayes B

The Bayes B method can be understood as a complement to the Bayes A method, because in addition to fitting markers with heterogeneous variances, it assumes that some markers are not in linkage disequilibrium (LD) with any gene. Thus, their effect is null. The Bayes B method is formulated via a mixture of distributions (Pérez and De los Campos, 2014) considering the following assumptions.

$$a_i | \sigma_{a_i}^2 \begin{cases} \sim N(0, \sigma_{a_i}^2) & \text{with probability } 1-\pi \\ = 0 & \text{with probability } \pi \end{cases}$$

$$\pi | \pi_0, p \sim beta(\pi_0, p);$$
$$\sigma_{a_i}^2 | gl, S_a \sim \chi^{-2}(gl, S_a);$$
$$S_a | r, s \sim G(r, s). \quad \text{Eq. 6}$$

Bayes Lasso (BL)

Similar to the previous Bayesian methods, the BL method assumes heterogeneous variances for the marker effect, and it also assumes that several markers may not be in LD with any gene. However, the selection of markers in BL is performed indirectly, via the marginal distribution of marker effects, which is double exponential (DE) (Park and Casella, 2008; de los Campos et al., 2009). This distribution is more leptokurtic than the prior marginal t-Student distribution used in Bayes A and B (Gianola et al., 2009). The BL that was adjusted in this study assumes the following.

$$a_i | \sigma_\varepsilon^2, \tau_i^2 \sim N(0, \sigma_\varepsilon^2 \tau_i^2);$$
$$\tau_i^2 | \lambda \sim Exponential(0.5\lambda^2);$$
$$\lambda | r, s \sim Gamma(r, s). \quad \text{Eq. 7}$$

According to Park and Casella (2008) and los Campos et al. (2009),

$$a_i | \lambda \sim DE(\lambda). \quad \text{Eq. 8}$$

Bayes Cπ

The Bayes Cπ method was proposed by (Habier et al., 2011). In this method, similar to the Bayes B method, it is assumed that some markers are not in LD with any gene and, therefore, has null effect. However, all SNPs of non-null effect exhibit homogeneous variance. Thus, in this model, SNPs with the same MAF that are in LD with the genes that govern the trait have the same proportion of genetic variance.

$$a_i|\sigma_a^2 \begin{cases} \sim N(0,\sigma_a^2) & \text{with probability } 1\text{-}\pi \\ = 0 & \text{with probability } \pi \end{cases}$$

$$\pi|\pi_0, p \sim beta(\pi_0, p);$$

$$\sigma_a^2|gl, S_a \sim \chi^{-2}(gl, S_a). \quad \text{Eq. 9}$$

RKHS

The semi-parametric regression method reproducing kernel Hilbert spaces (RKHS) is a procedure that predicts the genetic merit of the individuals directly, i.e., it forgoes the estimation of the marker effect. In fact, GBLUP is a special case of RKHS (De los Campos, 2009); however, in the present study, the RKHS model was adjusted according to the kernel averaging approach (De los Campos et al., 2010) using three functions of genotype data.

$$y = \mu + \sum_{i=1}^{3} g_i + e \quad \text{Eq. 10}$$

where

$$[g_1 \quad g_2 \quad g_3]' \sim N(0, \bigoplus_{i=1}^{3} K_i \sigma_{g_i}^2);$$

$$\sigma_{g_i}^2|\nu, S \sim \chi^{-2}(\nu, S);$$

$$K_i = \exp(-\varphi_i D^2); \quad \text{Eq. 11}$$

Where **D** is the Euclidean distance matrix considering the data set of SNPs (dosage given allele). φ_i is the i-th bandwidth value that controls the magnitude of the covariance among individuals; the used bandwidth values were 5/h, 1/h, and 0.2/h, where h corresponds to the 0.05 percentile of D, which leads to local, intermediate, and global Gaussian kernels (González-Camacho et al., 2012; Tusell et al., 2014). $\bigoplus$ is the direct sum operator; the other components of the model have been previously explained. In this model, the estimation of the total genotypic value (additive and non- additive effects) was confounded (Morota and Gianola, 2014) by the component, $\sum_{i=1}^{3} g_i$, and it was not possible to perform orthogonal decomposition.

Correlation network and computational processing times

After the GS models were fitted, correlation network analysis was performed among the different methods to identify the most optimal method. In addition, the computational processing time of each method was calculated. The frequency-based model was fitted using the rrBLUP package and the Bayesian models were fitted using the BGLR package of the R software. Each method was analyzed ten times in random order

Different densities of SNP markers

After the identification of the best method for GRS, different densities of SNP markers were evaluated using LD, which is defined as the non-random association of alleles at different loci; it is calculated as the difference between the observed frequency and the expected frequency of the haplotypes considering the independent segregation of the alleles (Weiss and Clark, 2002).

To assess the possibility of using less dense panels of SNPs, filters were applied using the LD-pruning approach, which consists of keeping only one SNP when two SNPs are in LD above the threshold; there- fore, thresholdsfrom0.05(resulting in 418 SNPs) to1(complete panel with 10,507 SNPs) were used for LD r2 statistics, considering the entire length of the chromosome in the comparisons and filters. Thus, the selective accuracies for the traits under study were obtained to detect the lowest density of SNPs with satisfactory accuracy. LD-pruning was performed using Plink (Purcell et al., 2007).

Results and discussion

Comparison of the methods

The selective accuracies of the tested methods were statistically similar according to the t-test, with a confidence interval of 95%. For the GY trait, the accuracies were between 0.1922 for the BL method and 0.2685 for the Bayes A method, with an overall mean of 0.2428. The Bayes A, Bayes B, Bayes Cπ, and GBLUP methods provided similar results, with a prediction accuracy of ~0.26. The BL and RKHS methods provided inferior results, with prediction accuracies of 0.1922 and 0.2030, respectively (Figure 1). Wang et al. (2015) also obtained similar results with different models for GY in wheat. Riedelsheimer et al. (2012) did not find major differences between the Bayes B and other models in the prediction of several traits, including traits with QTLs with major effects in maize.

The accuracy for PE was between 0.3811 (RKHS) and 0.4164 (GBLUP), with an estimated overall mean of 0.4007. The results of the Bayes Cπ, BL, and RKHS methods were lower, estimated around 0.39. The GBLUP method stood out with an accuracy of 0.4164 (Figure 1). Xu et al. (2018) reported that GBLUP was better than the Bayesian methods in the prediction of six maize traits. Some authors have reported the predominance of additive effects in the gene expression of PE (Burnham Larish and Brewbaker, 1999; Pereira and Amaral Júnior,

2001; Freitas Júnior et al., 2006; Santos et al., 2008). According to Wang et al. (2015), the existing GS methods are mostly based on an additive model and hence it is difficult to precisely estimate non-additive variation. It is thus inferred that the best values of accuracy obtained for PE are explained by the higher heritability, i.e., lower environmental influence on the manifestation of the trait.

The accuracy for PV was between 0.2517 (RKHS) and 0.3037 (Bayes Cπ), with an overall mean of 0.2829. The Bayes B, Bayes Cπ, and GBLUP methods provided similar results, with estimated values around 0.30. The BL and RKHS methods resulted in lower estimates, with values around 0.25 (Figure 1) and were therefore less accurate. According to Amaral Junior et al. (2016), the use of PV as a super-trait circumvents the challenge faced by specialists in popcorn breeding, with regards to the occurrence of a negative correlation between PE and GY, because PV allows obtaining gains simultaneously in the two main traits of economic importance. The accuracy estimates for the PV trait were intermediate compared to those for the PE and GY traits; the GBLUP, Bayes A, and Bayes B methods provided the most satisfactory results for selection purposes, as was the case for GY and PE.

In correlation network analysis, variables are represented by points connected by arrows of varying thickness, depending directly on the intensity of the correlation. The stronger the correlation between two variables, the thicker is the line that connects the points of the mesh of the network (Silva et al., 2016). Correlation networks have been used to characterize complex systems in several areas, including biology (Di Leo et al., 2011; Pearce et al., 2015), public health (Saba et al., 2014), food science (Monforte et al., 2015), and breeding (Silva et al., 2016). In the present study, the correlation between the results obtained for the traits in the different methods was higher than 0.98 (Figure 2). This result is a further indication that there are no significant differences between the tested methods.

Heslot et al. (2012) compared different methods to analyze the computation time required to conduct GS studies. In the present study, the analysis of the computation time needed to obtain the results for each method showed that it was considerably less for the GBLUP method as compared to those of the other methods, with an almost instant output. The results of Bayesian models were obtained in approximately 4 min, whereas the results of GBLUP were obtained in less than 1 s (Table 1). According to Wang et al. (2018), the estimation process using the Bayesian approach is usually lengthy, which restricts its application.

In summary, considering the results obtained using the different methods for the traits under study, it is concluded that there were no statistically significant differences between the methods. However, the GBLUP, Bayes A, and Bayes B methods stood out for providing-numerically more accurate estimates for all traits in the study. However, because the GBLUP method had simplicity in execution, a shorter analysis time, and because it was numerically more accurate for the main trait of popcorn quality (PE), this method is indicated for GRS in the popcorn crop.

Different densities of SNPs

The estimation of LD should be one of the first analyses in GS, with the aim of investigating the utilized marker informativity. According to Resende (2008), the genetic variation of a quantitative trait may be explained by the presence of markers in LD with minor or major effect QTLs, because only the markers in disequilibrium are used to determine phenotypes from genotypes.

Table 1. Estimate of the times required to obtain results using the Bayes A, Bayes B, and GBLUP methods.

Model	Time (s)	
	Mean	Standard deviation
Bayes A	229.3784	1.2024
Bayes B	282.6650	1.7503
GBLUP	0.0143	0.0038

Considering that the assumptions of the GBLUP provided the most satisfactory results, different densities of SNPs were subsequently evaluated to identify the most beneficial density, considering the significant cost of generating a high amount of markers. Thus, 20 different densities of SNPs were tested – varying between 10,507 (r^2=1.00) and 418 (r^2=0.05) – using the LD (r^2) as a filter (Figure 3). For example, the threshold for 6,766 SNPs was r^2=0.7, i.e., if two markers had r^2>0.7, only one was retained. For the trait GY, the highest accuracy value (0.303) was obtained with the use of 6,766 SNPs (r^2=0.70); however, a similar accuracy (0.296) was obtained with 4,848 SNPs (r^2=0.45). For the trait PV, a panel of 4,848 SNPs (r^2=0.45) also provided a good accuracy result, with an estimate of 0.311 (Figure 3).

For the GY and PV traits, densities lower than 4.406 SNPs (r^2=0.40) led to lower accuracies, with the exception of the accuracy value obtained with a density of 418 SNPs (r^2=0.05). However, the latter result was probably a false-positive result; this is in accordance with Dickson et al. (2011) study which states are variants may lead to unacceptable rates of false-positives. According to Litonjua and Celedo (2006), there are three potential conditions that make an association between a polymorphism and a phenotype significant: (1) the occurrence of aspurious relationship; (2) the proximity of a functional variant at an adjacent locus (LD); and (3) the locus directly affecting the expression of the phenotype. Utsunomiya et al. (2013) reported that small errors in the determination of genotypes can also result in unacceptable levels of type I (false-positive) and II (false-negative) errors.

For PE, the highest accuracy (0.416) was obtained with 10,507 SNPs (r^2=1.00); however, the use of 4,406 SNPs (r^2=0.40) also resulted in a similar accuracy (0.402). Accuracy varied between 0.416 (10,507 SNPs) and 0.332 (418 SNPs), which was considerably lower than the variation observed for additive trait GY (Figure 3). Wang et al. (2015) analyzed the data of the "Wheat Global Pro- gram" of the International Center for Maize and Wheat Breeding (CIMMYT) and observed that the precision and predictive capacity of GBLUP remained relatively constant for additive traits, and it was independent of the number of QTLs used.

According to Hiremath et al. (2012), SNPs markers are receptive to automation and to high yield approaches, which can be implemented to reduce the costs of genotyping. Considering the results obtained in the different scenarios of SNP density assessed in the present study, it is in ferred that the use of ~ 4,800 SNPs (r^2=0.45) leads to similar or better results than those obtained with 10,507 SNPs (r^2=1.00). Therefore, it is possible to reduce the density of SNPs by approximately 50%, which results in a significant reduction of the costs of GRS in popcorn**.**

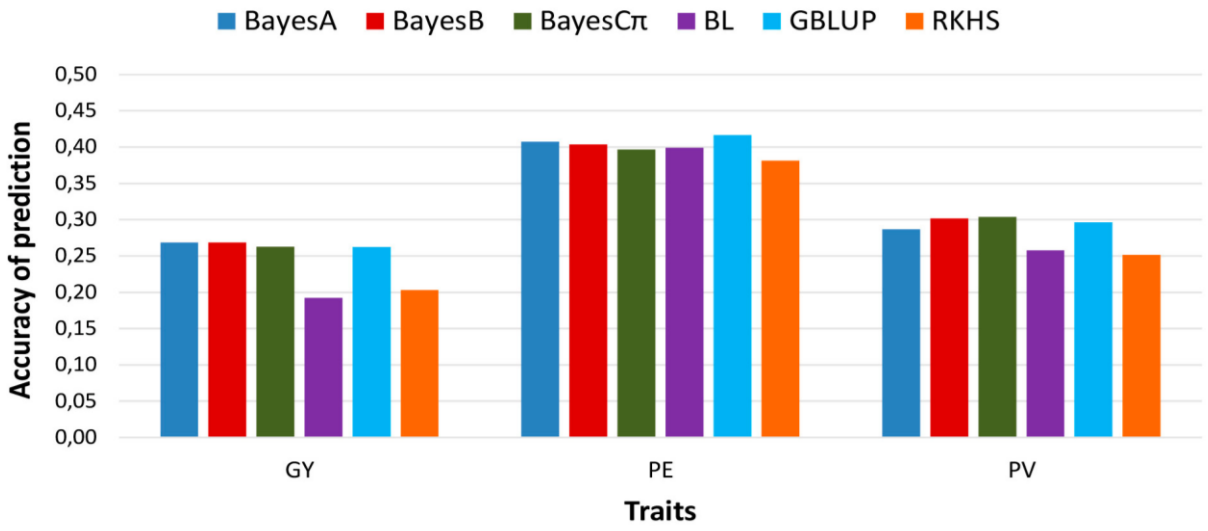

Figure 1. Estimates of the predictive accuracy of the means obtained for the Bayes A, Bayes B, Bayes C, Bayes Lasso (BL), GBLUP, and RKHS methods in a popcorn population for the traits grain yield (GY), popping expansion (PE), and popcorn volume (PV).

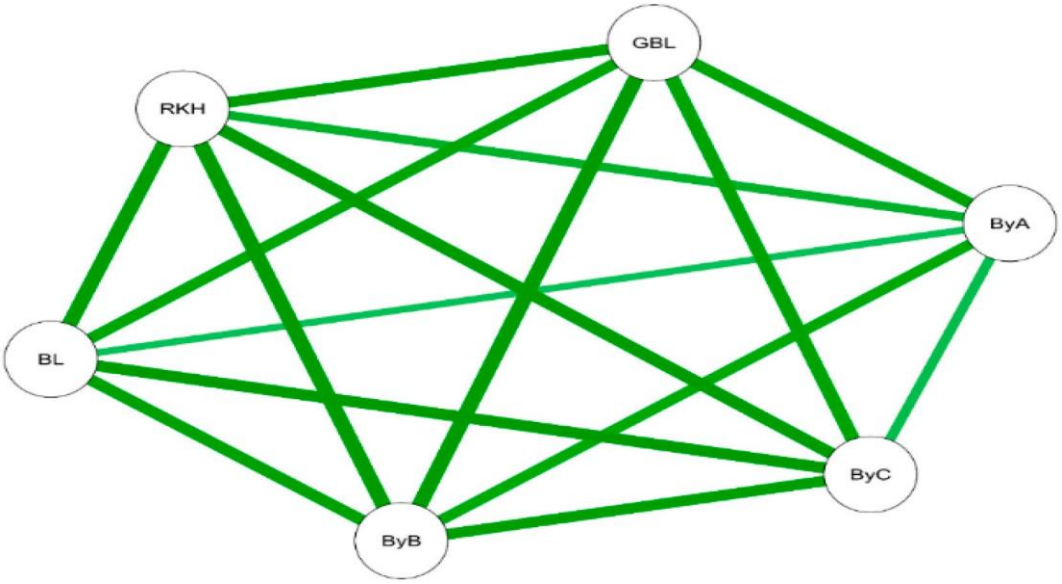

Figure 2. Correlation networks obtained in the GBLUP, Bayes A, Bayes B, Bayes C, BL, and RKHS methods for the traits GY, PE, and PV (p<005).

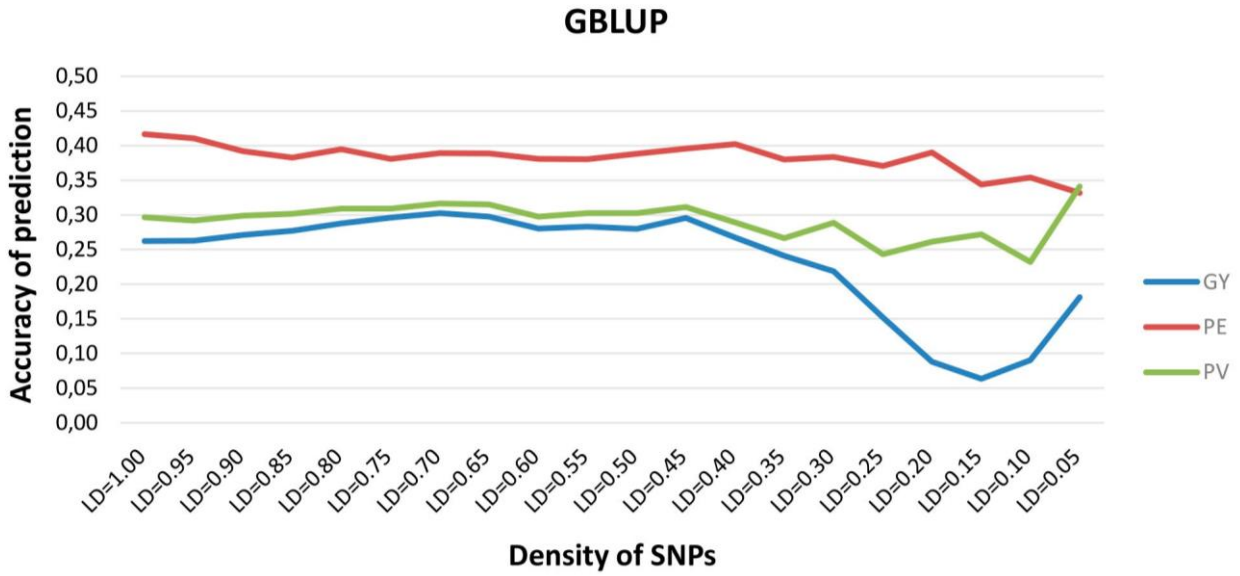

Figure 3. Prediction accuracies obtained with different densities of SNPs using the GBLUP method for the traits GY, PE, and PV.

Conclusions

The GBLUP method can be applied successfully. A considerable reduction in costs was obtained with the use of a small panel of SNPs.

References

ALMEIDA FILHO, J.E. DE; GUIMARÃES, J.F.R.; SILVA, F.F. E; RESENDE, M. DE V. DE; MUÑOZ, P.; KIRST, M.; JR, M.R. 2016. The contribution of dominance to phenotype prediction in a pine breeding and simulated population. **Heredity**, 117:33–41.

AMARAL JUNIOR, A.T. DO; DE JESUS FREITAS, I.L.; GUIMARÃES, A.G.; MALDONADO, C.; ARRIAGADA, O.; MORA, F. 2016. Bayesian analysis of quantitative traits in popcorn (*Zea mays* L.) through four cycles of recurrent selection. **Plant Production. Science,** 19:574–578.

BURNHAM LARISH, L.L.; BREWBAKER, J.L.1999. Diallel analyses of temperate and tropical popcorns. **Maydica,** 44:279–284.

BUTLER, D.; CULLIS, B.R.; GILMOUR, A. 2007. Asreml-R: An R package for mixed models using residual maximum likelihood. Accessed: 07 December 2018.

CROSSA, J.; PÉREZ-RODRÍGUEZ, P.; CUEVAS, J.; MONTESINOS-LÓPEZ, O.; JARQUÍN, D.; DE LOS CAMPOS, G.; BURGUEÑO, J.; CAMACHO-GONZÁLEZ, J.M.; PÉREZ-ELIZALDE, S.; BEYENE, Y.; DREISIGACKER, S.; SINGH, R.; ZHANG, X.; GOWDA, M.; ROORKIWAL, M.; RUTKOSKI, J.; VARSHNE, R.K.2017. Genomic selection in plant breeding: methods, models and perspectives. **Trends in Plant Science**, 22:1–15.

DAROS, M.; AMARAL JÚNIOR, A.T.; PEREIRA, M.G. 2002. Genetic gain for grain yield and popping expansion in full-sib recurrent selection in popcorn. **Crop Breeding and Applied Biotechnology,** 2:339–344.

DAROS, M.; DO AMARAL JR., A.T.; PEREIRA, M.G.; SANTOS, F.S.; GABRIEL, A.P.C.; SCAPIM, C.A.; FREITAS JR., S. DE P.; SILVÉRIO, L.2004. Recurrent selection in inbred popcorn families. **Scientia Agricola**, 61:609–614.

DE LOS CAMPOS, G., GIANOLA, D., ROSA, G.J.M.2009. Reproducing kernel Hilbert spaces regression: a general framework for genetic evaluation. **J. Anim. Sci.**, 87:1883–1887.

DE LOS CAMPOS, G.; GIANOLA, D.; ROSA, G.J.M.; WEIGEL, K. A.; CROSSA, J. 2010. Semi-parametric genomic-enabled prediction of genetic values using reproducing kernel Hilbert spaces methods. **Genetics Research**, 92:295–308.

DE LOS CAMPOS, G.; HICKEY, J.M.; PONG-WONG, R.; DAETWYLER, H.D.; CALUS, M.P.L. 2013. Whole genome regression and prediction methods applied to plant and animal breeding. **Genetics**, 193: 327–345.

DE LOS CAMPOS, G.; NAYA, H.; GIANOLA, D.; CROSSA, J.; LEGARRA, A.; MANFREDI, E.; WEIGEL, K.; COTES, J.M. 2009. Predicting quantitative traits with regression models for dense molecular markers and pedigree. **Genetics**,182:375–385.

DILEO, M. V.; STRAHAN, G.D.; DEN BAKKER, M.; HOEKENGA, O.A. 2011. Weighted Correlation Network Analysis (WGCNA) Applied to the Tomato Fruit Metabolome. **PLoS One**, 6:e26683.

FREITAS, I.L.J.; DO AMARAL JÚNIOR, A.T.; FREITAS JR., S.P.; CABRAL, P.D.S.; RIBEIRO, R.M.; GONÇALVES, L.S.A. 2014. Genetic gains in the UENF-14 popcorn population with recurrent selection. **Genetics and Molecular Research**, 13:518–527.

FREITAS JÚNIOR, S. DE P.; AMARAL JÚNIOR, A.T. DO; PEREIRA, M.G.; CRUZ, C.D.; SCAPIM, C.A. 2006. Capacidade combinatória em milho-pipoca por meio de dialelo circulante. **Pesquisa Agropecuária Brasileira**, 41:1599–1607.

FREITAS JÚNIOR, S.P. DE P.; AMARAL JÚNIOR, A.T. DO; RANGEL, R.M.; VIANA, A.P. 2009. Genetic gains in popcorn by full-sib recurrent selection. **Crop Breeding and Applied Biotechnology**, 9:1–7.

FRITSCHE-NETO, R.; RESENDE, M.D.V.; MIRANDA, G.V.; DOVALE, J.C. 2012. Seleção genômica ampla e novos métodos de melhoramento do milho. **Revista Ceres**, 59:794–802.

GIANOLA, D. 2013. Priors in whole-genome regression: The Bayesian alphabet returns. **Genetics**, 194:573–596.

GIANOLA, D.; DE LOS CAMPOS, G.; HILL, W.G.; MANFREDI, E.; FERNANDO, R. 2009. Additive genetic variability and the Bayesian alphabet. **Genetics**, 183:347–363.

GIANOLA, D.; VAN KAAM, J.B.C.H.M. 2008. Reproducing Kernel Hilbert Spaces Regression Methods for Genomic Assisted Prediction of Quantitative Traits. **Genetics**, 178:2289–2303.

GONZÁLEZ-CAMACHO, J.M.; DE LOS CAMPOS, G.; PÉREZ, P.; GIANOLA, D.; CAIRNS, J.E.; MAHUKU, G.; BABU, R.; CROSSA, J. 2012. Genome-enabled prediction of genetic values using radial basis function neural networks. **Theoretical and Applied Genetics**, 125:759–771.

GUIMARÃES, A.G.; AMARAL JÚNIOR, A.T. DO; LIMA, V.J. DE; LEITE, J.T.; SCAPIM, C.A.; VIVAS, M. 2018. Genetic gains and selection advances of the UENF-14 popcorn population. **Revista Caatinga**, 31:271–278.

HABIER, D.; FERNANDO, R.L.; KIZILKAYA, K.; GARRICK, D.J. 2011. Extension of the bayesian alphabet for genomic selection. **BMC Bioinformatics**,12:186.

HESLOT, N.; YANG, H.P.; SORRELLS, M.E.; JANNINK, J.L. 2012. Genomic Selection in Plant Breeding: A Comparison of Models. **Crop Science**, 52:146.

HIREMATH, P.J.; KUMAR, A.; PENMETSA, R.V.; FARMER, A.; SCHLUETER, J.A.; CHAMARTHI, S.K.; WHALEY, A.M.; CARRASQUILLA-GARCIA, N.; GAUR, P.M.; UPADHYAYA, H.D.; KAVI KISHOR, P.B.; SHAH, T.M.; COOK, D.R.; VARSHNEY, R.K. 2012. Large-scale development of cost-effective SNP marker assays for diversity assessment and genetic mapping in chickpea and comparative mapping in legumes. **Plant Biotechnology Journal**, 10:716–732.

MEUWISSEN, T.H.E.; HAYES, B.J.; GODDARD, M.E. 2001. Prediction of Total Genetic Value Using Genome-Wide Dense Marker Maps. **Genetics**, 157:1819–1829.

MONFORTE, A.R.; JACOBSON, D.; SILVA FERREIRA, A.C. 2015. Chemiomics: Network Reconstruction and Kinetics of Port Wine Aging. **Journal of Agricultural and Food Chemistry**, 63:2576–2581.

MOROTA, G.; GIANOLA, D. 2014. Kernel-based whole-genome prediction of complex traits: a review. **Frontier Genetics**, 5:363.

NEVES, L.G.; DAVIS, J.M.; BARBAZUK, W.B.; KIRST, M. 2014. A High-Density Gene Map of Loblolly Pine (*Pinus taeda* L.) Based on Exome Sequence Capture Genotyping. **G3 Genes, Genomes, Genetics**, 4:29–37.

PARK, T.; CASELLA, G. 2008. The Bayesian Lasso. **Journal of the American Statistical Association**, 103:681–686.

PEARCE, S.; FERGUSON, A.; KING, J.; WILSON, Z.A. 2015. Flower Net: A Gene Expression Correlation Network for Anther and Pollen Development. **Plant Physiology**, 167:1717–1730.

PEREIRA, M.G.; AMARAL JÚNIOR, A.T. 2001. Estimation of Genetic Components in Popcorn Based on the Nested Design. **Crop Breeding and Applied Biotechnology**, 1:3–10.

PÉREZ, P.; DE LOS CAMPOS, G. 2014. Genome-Wide Regression & Prediction with the BGLR Statistical Package. **Genetics**, 198:483–495.

PSZCZOLA, M.; MULDER, H.A.; CALUS, M.P.L. 2011. Effect of enlarging the reference population with (un) genotyped animals on the accuracy of genomic selection in dairy cattle. **Journal of Dairy Science**, 94:431–441.

PURCELL, S.; NEALE, B.; TODD-BROWN, K.; THOMAS, L.; FERREIRA, M.A.R.; BENDER, D.; MALLER, J.; SKLAR, P.; DE BAKKER, P.I.W.; DALY, M.J.; SHAM, P.C. 2007. PLINK: A Tool Set for Whole Genome Association and Population-Based Linkage Analyses. **American Journal of Human Genetics**, 81:559–575.

R CORE TEAM.2013. R: A language and environment for statistical computing. R Foundation for Statistical Computing.

RABIER, C.E.; BARRE, P.; ASP, T.; CHARMET, G;, MANGIN, B. 2016. On the Accuracy of Genomic Selection. **PLoS One**, 11:e0156086.

RANGEL, R.M.; TEIXEIRA, A.; SIMÕES, L.; GONÇALVES, A.DE,S.; 2011. Análise biométrica de ganhos por seleção em população de milho pipoca de quinto ciclo de seleção recorrente. **Revista Ciência Agronômica**, 43:473–481.

RESENDE, M.D.V. DE; RESENDE JÚNIOR, M.F.R.; AGUIAR, A.M.; ABAD, J.I.M.; SANSALONI, A.A.M.C.; PETROLI, C.; GRATTAPAGLIA, D. 2010. Computação da Seleção Genômica Ampla (GWS). **EMBRAPA Florestas.** Colombo.

RESENDE JR, M.F.R.; MUÑOZ, P.; ACOSTA, J.J.; PETER, G.F.; DAVIS, J.M.; GRATTAPAGLIA, D.; RESENDE, M.D. V.; KIRST, M. 2012. Accelerating the domestication of trees using genomic selection: Accuracy of prediction models across ages and environments. **New Phytologist**, 193:617–624.

RESENDE, M.D.V. 2008. Genomica Quantitative e Seleção no Melhoramento de Plantas e Animais. **Embrapa Florestas.**

RIBEIRO, R.M.; AMARAL JÚNIOR, A.T.; GONÇALVES, L.S.A.; CANDIDO, L.S.; SILVA, T.R.C.; PENA, G.F. 2012. Genetic progress in the UNB-2U population of popcorn under recurrent selection in Rio de Janeiro, Brazil. **Genetics and Molecular Research**, 11:1417–1423.

RIEDELSHEIMER, C.; TECHNOW, F.; MELCHINGER, A.E. 2012. Comparison of whole-genome prediction models for traits with contrasting genetic architecture in a diversity panel of maize inbred lines. **BMC Genomics**, 13:452.

SABA, H.; VALE, V.C.; MORET, M.A.; MIRANDA, J.G.V. 2014. Spatio-temporal correlation networks of dengue in the state of Bahia. **BMC Public Health,** 14:1085.

SANTOS, F.S.; DO AMARAL JÚNIOR, A.T.; FREITAS JÚNIOR, S.D.P.; RANGEL, R.M.; SCAPIM, C.A.; MORA, F. 2008. Genetic gain prediction of the third recurrent selection cycle in a popcorn population. **Acta Scientiarum Agrononomy**,30:435-441.

SILVA, A.R. DA; RÊGO, E.R. DO; PESSOA, A.M. DOS S.; RÊGO, M.M. DO. 2016. Correlation network analysis between phenotypic and genotypic traits of chili pepper. **Pesquisa Agropecuária Brasileira**, 51:372–377.

THAVAMANIKUMAR, S.; DOLFERUS, R.; THUMMA, B.R. 2015. Comparison of Genomic Selection Models to Predict Flowering Time and Spike Grain Number in Two Hexaploid Wheat Doubled Haploid Populations. **G3: Genes, Genomes, Genetics**, 5:1991–1998.

TUSELL, L.; PÉREZ-RODRÍGUEZ, P.; FORNI, S.; GIANOLA, D. 2014. Model averaging for genome-enabled prediction with reproducing kernel Hilbert spaces: a case study with pig litter size and wheat yield **Journal of Animal Breeding and Genetics**, 131:105–115.

UTSUNOMIYA, Y.T.; DO CARMO, A.S.; CARVALHEIRO, R.; NEVES, H.H.; MATOS, M.C.; ZAVAREZ, L.B.; PÉREZ O'BRIEN, A.M.; SÖLKNER, J.; MCEWAN, J.C.; COLE, J.B.; VAN TASSELL, C.P.; SCHENKEL, F.S.; DA SILVA, M.V.; PORTO NETO, L.R.; SONSTEGARD, T.S.; GARCIA, J.F. 2013. Genome-wide association study for birth weight in Nellore cattle points to previously described orthologous genes affecting human and bovine height. **BMC Genetics**, 14:52.

WANG, X.; XU, Y.; HU, Z.; XU, C. 2018. Genomic selection methods for crop improvement: Current status and prospects. **The Crop Journal**, 6:330–340.

WANG, X.; YANG, Z.; XU, C. 2015. A comparison of genomic selection methods for breeding value prediction. **Science Bulletin**, 60:925–935.

WEISS, K.M.; CLARK, A.G. 2002. Linkage disequilibrium and the mapping of complex human traits. **Trends Genetics**, 18:19–24.

XU, Y.; WANG, X.; DING, X.; ZHENG, X.; YANG, Z.; XU, C.; HU, Z. 2018. Genomic selection of agronomic traits in hybrid rice using an NCII population. **Rice**, 11:32.