



# Genome-wide prediction of maize single-cross performance, considering non-additive genetic effects

J.P.R. Santos<sup>1</sup>, H.D. Pereira<sup>2</sup>, R.G. Von Pinho<sup>2</sup>, L.P.M. Pires<sup>2</sup>, R.B. Camargos<sup>2</sup>  
and M. Balestre<sup>3</sup>

<sup>1</sup>Programa de Pós-Graduação em Genética e Melhoramento de Plantas,  
Universidade Federal de Lavras, Lavras, MG, Brasil

<sup>2</sup>Programa de Pós-Graduação em Fitotecnia, Universidade Federal de Lavras,  
Lavras, MG, Brasil

<sup>3</sup>Programa de Pós-Graduação em Estatística e Experimentação Agropecuária,  
Universidade Federal de Lavras, Lavras, MG, Brasil

Corresponding author: M. Balestre

E-mail: [marciobalestre@dex.ufla.br](mailto:marciobalestre@dex.ufla.br)

Genet. Mol. Res. 14 (4): 18471-18484 (2015)

Received September 4, 2015

Accepted December 11, 2015

Published December 23, 2015

DOI <http://dx.doi.org/10.4238/2015.December.23.35>

**ABSTRACT.** The prediction of single-cross hybrids in maize is a promising technique for optimizing the use of financial resources in a breeding program. This study aimed to evaluate Genomic Best Linear Unbiased Predictors models for hybrid prediction and compare them with the Bayesian Ridge Regression, Bayes A, Bayesian LASSO, Bayes C, Bayes B, and Reproducing Kernel Hilbert Spaces Regression models, with inclusion or absence of non-additive effects under three heritability scenarios. Data from a maize germplasm bank belonging to USDA were used to determine the effects of molecular markers, which were considered to be parametric, to build 400 single-cross hybrids between two line groups via simulation. The following parameters were used to compare the models: predictive ability, estimation of variance components, heritability of genetic effects present in all situations, and the sum of squares of the predicted errors.

The models responded positively when dominance effects were included in non-additive models, with all models tending to show an increase in the values of heritability parameters under all scenarios. Differences occur between models depending on the heritability range considered. Estimates of additive and dominant effects were better than estimates of epistatic effects. Estimates increased in accuracy for all models when non-additive effects for maize cob weight were considered.

**Key words:** Plant breeding; Epistasis; Heterosis; Bayesian models

## INTRODUCTION

One of the main problems in maize breeding programs is the need for a large number of crossings to obtain single-cross hybrids, which generates considerable demand for labor and financial resources (Bernardo, 1994). This is mainly due to the need for these hybrids to be evaluated in experiments with large numbers of replicates and often in multiple locations. Thus, effective ways of predicting the performance of traits of economic importance have great potential to reduce costs due to fewer hybrids being tested in the field. A pioneering study investigating performance prediction of maize hybrids was carried out by Jenkins (1934). In this study, four double-cross hybrid performance prediction methods were proposed using data on the performance of single-cross hybrids. The most effective method was method B, in which predictions were based on the average performance of four non-parental single-cross hybrids. At that time, the methodology was widely accepted and significantly reduced the number of double hybrids tested, which led to a considerable increase in the efficiency of breeding programs (Hallauer et al., 2010). Later, Eberhart (1964) carefully studied the methods proposed by Jenkins (1934) and suggested a new methodology for the prediction of double hybrids using average yield values of single and triple-cross hybrids. Cockerham (1967) also proposed a new prediction methodology for double-cross hybrids, in which both genetic and environmental effects were considered and values of single-cross hybrid performance were also used in predictions of double-cross hybrids.

Although there is now high interest in the use of phenotypic information in predictive processes, the discovery of DNA in the 1950s led researchers to explore information relating to molecular markers to forecast plant performance. Many studies have exploited the information of molecular markers in predictive performance procedures of single-cross hybrids of maize (Lee et al., 1989). The main justification for using molecular markers in predictions of single-cross hybrids came from the theory of quantitative genetics, in which hybrids containing loci with a high level of positive directional dominance obtained from lines with greater genetic diversity show higher heterosis (Falconer and Mackay, 1996). Therefore, it was believed that the genetic divergence estimated by molecular markers in the form of genetic distance could predict the best combinations among the lines used to generate the best single-cross hybrids (Lee et al., 1989). However, over the years, this hypothesis was not proven correct; in fact, genetic divergence only considered the genome in its state, regardless of whether the genomic regions are responsible for the predicted traits or if they are unimportant for the inheritance of these traits, which are indispensable factors for the success of the prediction using genetic information for molecular markers.

Subsequently, based on the mixed models theory proposed by Henderson (1985a), Bernardo (1994) proposed its application to predict quantitative traits with the relationship among

hybrids being estimated by molecular markers. This relationship can also be determined on the basis of pedigree information as originally proposed; however, such information is not always available in breeding programs, and, when available, does not always have high reliability. Thus, markers could be used to estimate kinship between individuals; as a result, it was possible to determine hereditary information for individuals and use this information in mixed models. However, despite the greater efficiency observed with this methodology, it also presented a constraint because it encourages the use of low density molecular markers during kinship estimation, which limits the linkage disequilibrium among markers and genes, resulting in rapid reduction in the efficiency of predictions across generations.

More recently, a new method for predicting quantitative traits was presented by Meuwissen et al. (2001), known as genome wide selection (GWS). The concept of genomic selection is based on the hypothesis that, with a high density of markers, all polymorphisms that contribute to trait variation will be in linkage disequilibrium with markers, which will segregate with the same pattern in the population under study (Goddard and Hayes, 2007). At that time, one of the main limitations of this method was the high cost of genotyping individuals, which was not yet economically feasible for breeding programs. However, with a significant decrease in the price of genotyping, and increased phenotyping, the adoption of a GWS method in breeding programs has become financially feasible. Currently, this is a successful technique used in animal breeding and is still in the early stages of testing and adoption for the improvement of annual crops (Heffner et al., 2009).

In terms of predictive models, it is common to use mixed and Bayesian models in GWS, which differ with regard to hypotheses about marker variance. The main feature of mixed models is their hypothesis of the infinitesimal gene with respect to the genetic control of traits. These models assume that the genetic effects of markers follow a common and normal probability distribution. The most popular mixed models in GWS are ridge regression BLUP (rr-BLUP) and G-BLUP (Meuwissen et al., 2001; Vanraden, 2008). Bayesian models allow the incorporation of *a priori* information, modeling of the variance heterogeneity of loci, and the selection and shrinkage of markers of small effect, thereby relaxing strong assumptions and permitting more modeling versatility. Many Bayesian models have been proposed in recent years with hypotheses of changes in *a priori* distributions of the effect estimates of molecular markers (Gianola, 2013). Despite the large number of models used for trait predictions with information from molecular markers, currently, little is known about the genetic architecture of quantitative traits.

In particular, some information is already known about some traits of economic importance and the nature of their genetic architecture. Currently, these traits are known to be controlled by a large number of genes; some genes are common for small effects and other rare genes have a high effect and are strongly affected by the environment when co-expressed (Mackay et al., 2009). In addition, these genes may present highly complex patterns of inheritance, including their allelic interactions given rise to so-called epistasis effects and its pervasive influence across several traits (pleiotropic effects) (Mackay et al., 2009).

Heterosis is a specific genetic event of great importance in maize crops and its genetic nature remains unknown. This event has been widely studied and exploited in maize crops, but the physiological and biochemical mechanisms remain unclear (Reif et al., 2005). However, non-additive genetic effects such as dominance and epistasis are necessary for heterosis to occur. Some traits can show a strong non-additive inheritance. If plant breeders were aware of the relative magnitude of these components, more accurate selection could be practiced (Henderson, 1985b).

Currently, there is great interest in improving GWS predictive models with the incorporation

of non-additive effects to increase the efficiency of complex trait predictions. However, until date, few comparative studies have been performed to evaluate the impact of non-additive effects on the predictive performance of both mixed and Bayesian models. Thus, the main objective of this study was to evaluate G-BLUP models and to compare them with Bayesian Ridge Regression (BRR), Bayes (BA), Bayes LASSO, Bayes C (BC), Bayes B (BB), and Bayesian Reproducing Kernel Hilbert Spaces Regression (RKHS) models with the inclusion or absence of non-additive genetic effects in the prediction of performance of maize cob weight of single-cross hybrids of maize.

## MATERIAL AND METHODS

### Description of database used

Genotypic and phenotypic data of 3215 lines of the Nested Association Mapping population (NAM population), belonging to the maize germplasm bank of the United States Department of Agriculture (USDA-ARS), available through the Panzea project (<http://www.panzea.org/>) were used. Genotyping data were provided by Romay et al. (2013); however, systematic sampling of these markers was carried out in this study at intervals 25 positions apart to obtain 27,000 molecular markers distributed across the genome. Among the available phenotypic data, the weight of maize cobs was chosen.

### Genetic analysis of the lines

The Bayes B method was used to estimate the genetic effects of 27,000 molecular markers, and the phenotypic data of maize cob weight were considered as the response variable. The genetic model used for analysis with BB is given below:

$$\mathbf{y} = \mathbf{j}\boldsymbol{\mu} + \mathbf{W}_1\mathbf{a} + \mathbf{W}_2\mathbf{d} + \boldsymbol{\varepsilon} \quad (\text{Equation 1})$$

where  $\mathbf{y}$  is a vector of  $n \times 1$  observations, in which  $n$  is the number of observations;  $\mathbf{j}$  is an incidence vector of fixed effects  $n \times 1$ ;  $\boldsymbol{\mu}$  is a scalar of fixed effects (mean);  $\mathbf{W}_1$  and  $\mathbf{W}_2$  are deviation matrices of allelic substitution and dominance effects  $n \times q$ , respectively, in which  $q$  is the number of random effects;  $\mathbf{a}$ ,  $\mathbf{d}$  are vectors of the marker's allelic substitution and dominance effects  $q \times 1$ , respectively; and  $\boldsymbol{\varepsilon}$  is the vector of residual effects  $n \times 1$ . Matrices of  $\mathbf{W}_1$  and  $\mathbf{W}_2$  incidence were built following Cockerham's metric (Cockerham, 1954; Zeng et al., 2005):

$$\mathbf{W}_1 \begin{cases} 2 - 2p_k \\ 1 - 2p_k \\ -2p_k \end{cases} \text{ for genotypes} \begin{cases} \mathbf{A}^1\mathbf{A}^1 \\ \mathbf{A}^1\mathbf{A}^2 \\ \mathbf{A}^2\mathbf{A}^2 \end{cases} \quad (\text{Equation 2})$$

$$\mathbf{W}_2 \begin{cases} -2(1 - p_k)^2 \\ 2p(1 - p_k) \\ -2p_k^2 \end{cases} \text{ for genotypes} \begin{cases} \mathbf{A}^1\mathbf{A}^1 \\ \mathbf{A}^1\mathbf{A}^2 \\ \mathbf{A}^2\mathbf{A}^2 \end{cases} \quad (\text{Equation 3})$$

where  $p$  is the frequency of the favorable allele at  $k$  locus.

A full description of specifications of the probability distributions of the random effects and the Bayes B model parameters can be found in de los Campos et al. (2013). All statistical analyzes were performed on Bayesian Generalized Linear Regression (BGLR) package (Pérez and de los Campos, 2014) of the R software (R Core Team, 2014), with the BGLR function, adjusted to 10,000 iterations with the first 1000 cycles discarded as burn-in, according to Meuwissen et al. (2001). The  $\pi$  mixing parameter (100/27,000) of the Bayes B model was introduced to adjust 100 allele substitution effects and 100 dominance effects of molecular markers. All other settings of BGLR function were maintained in the standard package mode.

### Simulation of hybrid yield

Four-hundred single-cross hybrids were simulated using artificial crossing in a partial diallel scheme without reciprocals (20 x 20). The two line groups were defined graphically by principal component analysis using spectral decomposition of the additive relationship matrix among lines. Group 1 was set by the lines next to the B73 line; and Group 2 was set by the lines next to Mo17 line (Table 1). Many commercial hybrids have these two lines in their genetic background.

**Table 1.** Groups of lines defined by principal component analysis.

Group 1		Group 2	
B73	GEMS-0086	Mo17	EZ18
NC328	PI01004	Mo44	NC44
Ames22753	PI559382	R177	Va17
PI539927	PI539923	B97	GEMN-0081
Ames27151	NSL438033	GA224	IDS91
Ames27218	GEMS-0223	PI601685	PI601416
PI538009	N192	A682	Ames19008
NSL437913	PI46485	Ames10261	Ames19287
A679	Ames30797	Ames27178	Ames26764
B109	PI550473	NSL437903	PI542778

Source: Panzea (2014).

The genotypic status of each hybrid locus during artificial crossing was defined by expectation of the allelic contribution of each parental line in the concerned hybrid. The mathematical expectation was calculated by:

$$E(Z_{ij}) = p_{\lambda j} p_{\Omega j} 2 + p_{\lambda j} (1 - p_{\Omega j}) 1 + p_{\Omega j} (1 - p_{\lambda j}) 1 \quad (\text{Equation 4})$$

where  $E(Z_{ij})$  is the expectation of genotypic value in  $j$  locus of the  $i$  hybrid coming from the crossing between  $\lambda$  and  $\Omega$  lines;  $p_{\lambda j}$ : favorable allele frequency of the  $\lambda$  line to the  $j$  locus; and  $p_{\Omega j}$ : favorable allele frequency of  $\Omega$  line in  $j$  locus.

The phenotypic value of the hybrid was obtained by:

$$y_{HS} = \mu + W_1 a^* + W_2 d^* + W_3 aa^* + W_4 ad^* + W_5 da^* + W_6 dd^* + \varepsilon \quad (\text{Equation 5})$$

where  $y_{HS}$  is a vector of observations of single-cross hybrids;  $\mu$  is the sample mean obtained by the Bayes B model using the actual phenotypic data of the lines;  $a^*$  is the vector of allele substitution

effects of the artificially constructed genome;  $d^*$  is the vector of dominance effects;  $aa^*$  (epistatic effect of the additive-additive type),  $ad^*$  (epistatic effect of the additive-dominant type),  $da^*$  (epistatic effect of the dominant-additive type), and  $dd^*$  (epistatic effect of the dominant-dominant type) are vectors of the simulated epistatic effects;  $\varepsilon$  is the sampled residual vector of a univariate normal distribution adjusting observations for heritability scenarios of 0.3, 0.5, and 0.7;  $W_1$  and  $W_2$  are the matrices of additive and dominance deviations built by the  $M$  matrix of hybrids;  $W_3$ ,  $W_4$ ,  $W_5$ , and  $W_6$  are epistatic deviation matrices built as established by Cockerham (1954). Columns related to the 233 largest additive effects estimated in the analysis of lines and columns involving the 233 largest dominance effects were separated in order to construct these matrices. These effects come from 225 quantitative trait loci (QTLs) that showed only additive effects, 225 QTLs that showed only dominance effects, and a further eight QTLs that showed both effects, additive and dominance.

Epistatic effects were simulated by the product of all two-by-two combinations among the values of the genetic markers (233 higher values) for each class of epistatic effects without the reciprocal to obtain vectors of classes  $aa^*$ ,  $ad^*$ ,  $da^*$ , and  $dd^*$ .

The heritability ranges were adjusted in by varying the magnitude of residual values ( $\varepsilon$ ), which were sampled from normal distributions centered on zero. The residual variance parameters of these distributions were defined by:

$$\sigma^2_{\varepsilon} = [(1 - h^2) / h^2] * \sigma^2_g \quad (\text{Equation 6})$$

where  $h^2$  is the broad sense heritability desired for adjustment;  $\sigma^2_g$  is the total genetic variance of the trait.

### Estimates of the weight of maize cobs with and without the inclusion of non-additive effects

After the simulation procedure, the data from single-cross hybrids were analyzed. The overall G-BLUP model used to predict additive effects and non-additive genetic values of hybrids was as follows:

$$\mathbf{y} = \mathbf{j}\boldsymbol{\mu} + \mathbf{Z}\mathbf{g} + \boldsymbol{\varepsilon} \quad (\text{Equation 7})$$

where  $\mathbf{y}$  ( $n \times 1$ ) is the vector of phenotypic observations of single-cross hybrids;  $\mathbf{j}$  ( $n \times 1$ ) is the incidence vector of fixed effects;  $\boldsymbol{\mu}$  ( $1 \times p$ ) is the vector of fixed effects (sample average);  $\mathbf{Z}$  ( $n \times q$ ) is the incidence matrix of random effects and in the case of G-BLUP, it is an identity matrix of dimension ( $n \times n$ );  $\mathbf{g}$  is the genetic value (random) of hybrids, which, depending on the model, were decomposed into  $\alpha$  (allele substitution effects),  $\delta$  (dominance deviations), and epistatic effects,  $\alpha\alpha$ ,  $\alpha\delta$ , and  $\delta\delta$ ; and finally  $\boldsymbol{\varepsilon}$  ( $n \times 1$ ) is the vector of residual effects.

For the  $\mathbf{g}$  decomposition, three configurations of G-BLUP models were used: i) G-BLUP A considering only the  $\alpha$  effects in the model; ii) G-BLUP AD with the inclusion of  $\alpha$  and  $\delta$  effects; iii) G-BLUP EP including  $\alpha$ ,  $\delta$ ,  $\alpha\alpha$ ,  $\alpha\delta$ , and  $\delta\delta$  effects.

The variance components for the analysis were obtained by maximizing the residual maximum likelihood estimation (REML) function of Patterson and Thompson (1971), using the

expectation maximization algorithm (EM) of Dempster (1977).

The predictive ability of the G-BLUP model was compared with some predictive models available in the BGLR package. Thus, analysis of the simulated hybrid was performed using the Bayesian Ridge Regression (BRR), Bayes, Bayes LASSO, Bayes C, Bayes B, and RKHS models. Analyses in the BGLR package were performed with the default settings of the program. Probability distribution specifications and detailed descriptions of Bayesian and RKHS models can be found in the studies by Gianola (2009) and de los Campos et al. (2013).

A preliminary study was carried out on epistatic effects in the G-BLUP model, which is less computationally intensive, and aimed to determine the relevance of their inclusion in other models.

Effects of allele substitution and dominance deviation from these models were estimated, except for the RKHS method. This predictive methodology lacks a genetic basis and does not consider additive or dominant relationship matrices, using only a Euclidean distance function to map the covariance between the genetic effects (our study case).

### Comparison of models

The predictive ability of the models was evaluated using Pearson's correlation among the estimated and parametric genetic values. In addition, the heritability of genetic effects was calculated by:

$$h_{gy_i}^2 = \text{COV}(\hat{g}, y) / \text{var}(y) \quad (\text{Equation 8})$$

where  $\hat{g}$  is only  $\alpha$  (additive model),  $\alpha + \delta$  (additive-dominant model), or  $\alpha + \delta + \alpha\alpha + \alpha\delta + \alpha\delta + \delta\delta$  (epistatic model).

Furthermore, we evaluated the ability of the G-BLUP model to partition the additive and non-additive components of genetic variance, given that it is the model most consistent with the theory of quantitative genetics. All of the other models tested assume prior distributions (Bayesian models) or modeling structure (RKHS model) that do not permit inference of genetic variance components.

Another statistic used was the prediction sum of squares (*PRESS*):

$$\text{PRESS} = \sum_{i=1}^n (\hat{O}_i - \hat{O})^2 \quad (\text{Equation 9})$$

where  $\hat{O}_i$  is the estimated value calculated individually for each variable referred to  $\alpha_i$ ;  $\delta_i$ ;  $\alpha_i\alpha_i$ ; and  $\alpha_i\delta_i$ , and  $\delta_i\delta_i$  variables, and  $\hat{O}$ , the parameter values.

## RESULTS

When comparing the quality of G-BLUP models in the estimation of variance components (Table 2), it can be observed that, for the variance of additive effects ( $V_a$ ), a closer approximation to the parametric value occurred when a complete model was used, i.e., by including dominance deviation effects and epistasis in the model. Considering all the effects, the approximation of

estimates means that the quality of variance estimates gradually improved when using the complete model. The closer approximation to the parameter value indicates the decomposition ability of variance components, namely, the confounding reduction among the additive, dominant, and epistatic effects.

**Table 2.** Parametric variance components (S) and the estimated components by models at the three heritability ranges ( $h^2$ ).

Parameters*	S	$h^2 = 0.3$	$h^2 = 0.5$	$h^2 = 0.7$
		G-BLUP A		
Va	-	25.8227	26.6841	41.0145
Vg	-	25.8227	26.6841	41.0145
		G-BLUP AD		
Va	-	24.7678	21.2032	26.6961
Vd	-	12.5113	15.8379	14.8276
Vg	-	37.2791	37.0410	38.5237
		G-BLUP EP		
Va	20.7	23.8922	20.0839	21.1368
Vd	14.3	11.1155	13.9725	11.9878
Vaa	2.17	2.0353	2.2003	5.1167
Vad	0.0792	0.8026	1.0696	0.3495
Vdd	0.00249	0.6784	0.8224	0.2824
Vg	37.3	38.5240	38.1488	38.8731

\*Va = variance of additive effects; Vd = variance of dominance effects; Vaa = variance of additive-additive epistatic effects; Vad = variance of additive-dominant epistatic effects; Vdd = variance of dominant-dominant epistatic effects; Vg = total genetic variance for maize cob weight character.

The inclusion of epistatic effects had little effect on estimates of Va, Vd, and especially Vg in relation to the estimates obtained by the reduced model only for additive and dominance effects (Table 2). In this context, it is expected that other types of genomic selection models with only additive and dominance effects show the same performance as their corresponding models, also considering the epistatic effects. Because of this, no further analyses were carried out in this study that included epistatic effects in other models of genomic selection.

In general, the inclusion of non-additive effects improved estimates of the total genetic variance (Vg) when comparing the estimates of the additive only model with others (Table 2). Therefore, it is noteworthy that the most accurate estimates of genetic variance were made using the G-BLUP EP model in the heritability setting of 0.5, for example, Vd value of 13.97, whereas the simulated value was 14.3.

The results suggest that the models responded positively in the three heritability scenarios with the inclusion of non-additive effects in predictive ability for the maize cob weight character (Table 3).

For the additive only models, RKHS presents good predictive ability for the heritability of 0.7, but high PRESS estimates were observed, ranging from 63,223.56 (simulated in  $h^2 = 0.3$ ) to 302,923.44 (simulated in  $h^2 = 0.5$ ). These high estimates of PRESS suggest that the RKHS model has only a low ability for isolating the additive genetic effects without confounding other components; thus the model interpretation might be prejudiced.

Despite the significant increase in analysis efficiency with the inclusion of dominance effects in a strictly additive model, all genomic selection models showed similar responses with the inclusion of this effect. The predictive ability of the G-BLUP EP and G-BLUP AD models did not differ from each other significantly.

**Table 3.** Predictive ability and PRESS (values in parentheses) for the models including only additive effects and with inclusion of non-additive effects for the three considered heritability ranges.

Models	Heritabilities		
	0.3	0.5	0.7
Including only additive effects			
G-BLUP A	0.6966 (7066.08)	0.7599 (5839.64)	0.7725 (5483.76)
Bayes A	0.7073 (7048.04)	0.7764 (5416.12)	0.7729 (5478.56)
Bayes B	0.7164 (7001.80)	0.7771 (5400.80)	0.7909 (5095.12)
Bayes C	0.7132 (7001.08)	0.7735 (5482.28)	0.7853 (5213.28)
Bayes LASSO	0.7359 (6774.84)	0.7898 (5128.76)	0.7769 (5392.44)
BRR	0.7056 (7057.36)	0.7616 (5773.44)	0.7702 (5537.40)
RKHS	0.6967 (63,223.56)	0.7956 (302,923.44)	0.9155 (15,276.16)
Including non-additive effects			
G-BLUP AD	0.8331 (4255.36)	0.9039 (2527.92)	0.9439 (1711.92)
G-BLUP EP	0.8333 (4228.40)	0.9019 (2550.40)	0.9436 (1666.32)
Bayes A	0.8406 (4676.64)	0.9025 (559.79)	0.9424 (1801.12)
Bayes B	0.8400 (4727.32)	0.8997 (2701.83)	0.9421 (1814.40)
Bayes C	0.8399 (4533.60)	0.9043 (2513.27)	0.9436 (1728.16)
Bayes LASSO	0.8389 (4758.52)	0.9029 (2539.81)	0.9438 (1703.52)
BRR	0.8410 (4184.16)	0.9041 (2521.10)	0.9444 (1697.20)

Although simulated epistasis was low, its inclusion in the model improved estimates of the variance components and, in addition, there was a slight improvement in correlations and in the reduction of PRESS with the three heritability scenarios. The major PRESS estimates were always associated with the RKHS model without compromising the estimated correlations.

Taking into account all models, heritability ranges, and the inclusion of non-additive effects, the magnitude of predictive abilities can be considered high compared to other simulation studies and also with real data reported in the literature.

Despite varying the conditions of genetic heritability based on three scenarios, the behavior of the evaluated models remained stable and efficient. For larger heritability, an increase in the predictive ability occurred.

On the basis of the results found here, it is possible to obtain good predictions of single-cross hybrids when non-additive effects are included in the model, validating the potential of genomic prediction as a tool in breeding programs. As the cost of genotyping decreases, breeding programs could make use of genomic prediction models to predict the genotypic value of new crossings before reaching the stage of field tests (Pérez-Rodríguez et al., 2012).

In addition, the present results are consistent with those reported by Jia and Jannink (2012), who found no difference in the predictive ability of the G-BLUP and Bayes A models under conditions of non-additive effects, presenting as low (0.1) or medium (0.5) heritability and small (20) or large (200) numbers of QTLs controlling a trait.

Table 4 shows the estimates of heritability in the narrow sense for models considering only additive effects when predicting the weight of maize cobs hybrids.

For heritability in the broad sense simulated at 0.3, the G-BLUP model provided the best estimate of heritability in the narrow sense. This was not observed for heritability in the broad sense at 0.7. The closer the estimate in relation to the parametric value, the more the model was able to discriminate the genetic effects in the data. This is another measure that can determine the quality of the predictive ability of the models.

Table 5 shows the estimates of heritability when including non-additive effects in predicting the heritability simulated at 0.3.

**Table 4.** Estimates of the heritability of additive effects for models considering only additive effects for the three simulated heritability ranges.

Broad (Narrow) sense heritability	0.3 (0.166)	0.5 (0.2775)	0.7 (0.388)
Bayes A	0.2028	0.2849	0.4381
Bayes B	0.2178	0.2855	0.4663
Bayes C	0.2112	0.2784	0.4602
Bayes LASSO	0.2457	0.3042	0.4466
BRR	0.2002	0.2546	0.4320
RKHS	0.1819	0.2467	0.4389
G-BLUP A	0.1695	0.2705	0.6420

**Table 5.** Estimates of heritability for models considering non-additive effects for heritability simulated at 0.3.

	Nature of heritabilities					Broad sense
	Additive	Dominant	Epistatic*			
			A-A	A-D	D-D	
Parametric heritabilities	0.1665	0.1150	0.0175	0.00064	0.00002	0.3
Bayes A	0.2047	0.1402	-	-	-	0.3449
Bayes B	0.2083	0.1374	-	-	-	0.3457
Bayes C	0.2025	0.1255	-	-	-	0.3280
Bayes LASSO	0.2055	0.1417	-	-	-	0.3472
BRR	0.1876	0.1050	-	-	-	0.2926
G-BLUP AD	0.1818	0.0809	-	-	-	0.2627
G-BLUP EP	0.1755	0.0717	0.0104	0.0051	0.0047	0.2674

\*Additive-additive epistasis (AA), additive-dominant epistasis (AD), and dominant-dominant epistasis (DD).

Table 5 shows that the models differed in the quality of the estimate among the dominant and additive genetic effects. The G-BLUP models estimated the additive effects accurately but not when regarding the effects of dominance. For the G-BLUP EP model, which was the only one to consider epistasis, the epistatic variance was of a small magnitude as expected, since the epistatic effects explain little about the weight of maize cobs in the simulation scenario used in this study.

For heritability of 0.5 (Table 6), the extent of estimates of dominance effects decreased in relation to heritability of 0.3. Even so, this was overestimated by all models. Again, the epistatic effects were estimated with lower accuracies.

**Table 6.** Estimates of heritability for models considering non-additive effects for heritability simulated at 0.5.

	Nature of heritabilities					Broad sense
	Additive	Dominant	Epistatic*			
			A-A	A-D	D-D	
Parametric heritabilities	0.2775	0.1917	0.00291	0.0011	0.00003	0.5
Bayes A	0.2353	0.2439	-	-	-	0.4722
Bayes B	0.2530	0.2667	-	-	-	0.5196
Bayes C	0.2467	0.2303	-	-	-	0.4770
Bayes LASSO	0.2444	0.2240	-	-	-	0.4684
BRR	0.2447	0.2275	-	-	-	0.4722
G-BLUP AD	0.2372	0.2257	-	-	-	0.4629
G-BLUP EP	0.2229	0.2015	0.0242	0.0143	0.0111	0.4739

\*Additive-additive epistasis (AA), additive-dominant epistasis (AD), and dominant-dominant epistasis (DD).

**Table 7.** Estimates of heritability for models considering non-additive effects for heritability simulated at 0.7.

	Nature of heritabilities					Broad sense
	Additive	Dominant	Epistatic*			
			A-A	A-D	D-D	
Parametric heritabilities	0.3884	0.2683	0.0407	0.0015	0.00005	0.7
Bayes A	0.4133	0.3183	-	-	-	0.7316
Bayes B	0.4126	0.3204	-	-	-	0.7330
Bayes C	0.4145	0.2937	-	-	-	0.7082
Bayes LASSO	0.4148	0.2827	-	-	-	0.6975
BRR	0.4085	0.2912	-	-	-	0.6997
G-BLUP AD	0.4116	0.2913	-	-	-	0.7029
G-BLUP EP	0.3796	0.2473	0.0731	0.0062	0.0058	0.7121

\*Additive-additive epistasis (AA), additive-dominant epistasis (AD), and dominant-dominant epistasis (DD).

When heritability was simulated at 0.7 (Table 7), differences in the estimates of heritability in a broad sense among the models were reduced compared with the heritabilities simulated at 0.3 and 0.5. This indicates that all models benefitted in this more favorable scenario, which decreased the discrepancies among the estimates and, therefore, hindered their discrimination.

The good performance of the G-BLUP models in estimating the heritability of genetic effects reinforces the attribute of such models for discriminating genetic effects, except for G-BLUP A in the heritability of 0.7.

Another observation is that the analysis of models with the heritability criterion of the estimates of genetic effects is not in accordance entirely with the analysis performed with the predictive ability criterion, showing that heritability could provide further information on the choice of models.

## DISCUSSION

Many GWS models were used in this study. In some situations, it was possible to use non-additive effects and, in others, it was possible to use only the additive effects due to the analysis (specific case of RKHS). This comparison between models was motivated by the fact that no studies in the literature have compared the G-BLUP model, including non-additive effects, with other described models. Most studies consider only additive effects in the exploited models.

Since Meuwissen et al. (2001), many studies have compared the predictive performance of additive models in genomic selection and concluded that there is no great discrepancy among models and that when there is, it varies depending on the species and the genetic trait architecture (de Los Campos et al., 2013; Howard et al., 2014). However, the models tested only considered additive effects. One of the goals of the present study was to verify whether this also occurs when non-additive effects are included in the models commonly used in genomic selection.

As a result of a non-linear regression process that exploits information on the development of kernels without a genetic basis, the RKHS model does not favor the orthogonal decomposition of genetic effects in additive and non-additive effects. This predictive model appears to restrictively capture additive and non-additive genetic signals confounded in a single component, an event that therefore precludes genetic interpretation.

One reason for the widespread lack of response to the inclusion of epistatic effects is the low ability of parametric models based on linear regression for capturing multiplicative effects such as epistasis (Gianola, 2006). Furthermore, the magnitude of epistatic variance components in this

study was low, which may have contributed to the small difference inferred by the inclusion of this effect in the models.

The smallest epistasis values were observed, because in the hybrids, this was generated by a multiplicative effect of the genes obtained from the lines. The justification for using multiplicative epistasis of genes was discussed by Schnell and Cockerham (1992).

In accordance with the findings of Crossa et al. (2014) and Technow et al. (2014), the size of the studied population affects such estimates. Even using a genomic construction of the hybrids based on the effects of genes obtained from the genomic analysis of the lines, this simulation process was found to be efficient for the construction of phenotypic values preserving the trait's architecture under analysis.

In one of the few studies comparing prediction of maize hybrids among G-BLUP and Bayesian models with dominance effects, Technow et al. (2012) found differences in the predictive ability of G-BLUP and Bayes B, when considering the effects of dominance on models, which did not occur in this study.

Bernardo (2014) simulated a maize breeding program and found no difference in predictive ability for the first selection cycle, both for equality or differences of variance among markers, using a mixed model methodology for prediction. This scenario showed that the incorrect, but convenient, hypothesis of equal variances among markers does not seem to significantly affect the responses to selection, as was also found by Bernardo and Yu (2007). Those authors found an average loss in response to selection, taking into account different numbers of QTLs and heritability of only 2% when comparing markers with equal or specific variances per locus. They highlight that, even for genomic selection, schemes using Bayesian procedures for modeling loci with specific variances would have little advantage. This was consistent with the results of the present study, since no Bayesian model was found to have better predictive ability in any of the situations.

Technow et al. (2014) stated that Bayesian models require large populations to estimate marker effects so that they can express any potential advantage; this condition was met in this study, since 3.215 lines were used to find the parameter values. With these estimates, 400 hybrids were simulated to be predicted by all models.

When referring to results already found in maize and wheat breeding programs by the International Maize and Wheat Improvement Center (CIMMYT) research center, Crossa et al. (2014) stated that none of the prediction models used fit all situations. However, empirical evidence has shown the inclusion of non-linear functions of markers tend to increase the predictive ability of the model when used in complex traits and when compared with linear models, a fact that was also noted by Pérez-Rodríguez et al. (2012).

Confirming the results reported by Crossa et al. (2014), Howard et al. (2014) found that the parametric methods studied (e.g., Bayes A, Bayes B, Bayes C, Bayes LASSO, BRR, G-BLUP, Least Squares) performed slightly better in terms of predictive ability when only additive effects were present. In the presence of epistasis, nonparametric methods (e.g., RKHS, neural networks) exceeded the parametric methods. The authors noted that the genetic architecture of a given trait was the main factor responsible for the differences in predictive ability among the 14 methods of genomic selection evaluated; this opinion was shared by Crossa et al. (2010).

Estimates obtained using RKHS methodology should be interpreted with caution, because this model does not estimate an isolated genetic effect, rather it estimates a set of effects with bias tending to be high, as already emphasized. This model is exalted by the "capacity" to better estimate nonlinear effects as epistatic effects, but it does not allow conclusions at this level, i.e.,

conclusions of the exact genetic nature of what is being estimated. This does not mean that it has no value in the predictive context since the ranking of the genotypes were equivalent, but careful should be taken about genetic interpretation related to effects of this model.

Riedelsheimer et al. (2012) concluded that a small gain in predictive ability can be obtained if a genomic prediction model is chosen in accordance with the genetic architecture of the trait. However, Heffner et al. (2009) noted that the basic genetic architecture of complex traits is unknown and, therefore, there is no knowledge *a priori* for this.

In contrast to the present results, Riedelsheimer et al. (2012) used real data from maize lines (289) to evaluate the predictive ability of genomic selection models for two classes of traits classified as agronomic or metabolic, and each class was controlled by large number of QTLs with little effect or few QTLs with great effect, respectively. They found that the RKHS model was equivalent to the G-BLUP model when considering additive effects for the agronomic traits. For the metabolic traits, Bayes LASSO was highlighted.

Simulation studies allow a greater variety of situations to be explored regarding the genetic architecture of traits and the population under study, whereas the actual data reflect the complexity involved. However, the simulation of genomes, and their genetic mechanisms, is complex. There are different forms of genomic variability and a variety of populations with different backgrounds, as well as considerable uncertainty about the distribution of gene action (Daetwyler et al., 2013).

The inclusion of non-additive effects in the models was found to improve the predictive ability of untested maize hybrids. In general, there was an increase in the predictive ability of the models with increased heritability.

For the population and traits considered in this study, epistasis did not affect the estimates of predictive ability, but they were a determining factor in improving the estimates of variance components and heritability.

## Conflicts of interest

The authors declare no conflict of interest.

## ACKNOWLEDGMENTS

Research supported by the Coordination for the Improvement of Higher Education Personnel (CAPES) and the National Council for Scientific and Technological Development (CNPq).

## REFERENCES

- Bernardo R (1994). Prediction of maize single-cross performance using RFLPs and information from related hybrids. *Crop Sci.* 34: 20-25.
- Bernardo R (2014). Genome wide selection when major genes are known. *Crop Sci.* 54: 68-75.
- Bernardo R and Yu J (2007). Prospects for genomewide selection for quantitative traits in maize. *Crop Sci.* 47: 1082-1090.
- Cockerham CC (1954). An extension of the concept of partitioning hereditary variance for analysis of covariances among relatives when epistasis is present. *Genetics* 39: 859-882.
- Cockerham CC (1967). Prediction of double crosses from single crosses. *Theor. Appl. Genet.* 37: 160-169.
- Crossa J, Campos G de L, Pérez P, Gianola D, et al. (2010). Prediction of genetic values of quantitative traits in plant breeding using pedigree and molecular markers. *Genetics* 186: 713-724.
- Crossa J, Pérez P, Hickey J, Burguén J, et al. (2014). Genomic prediction in CIMMYT maize and wheat breeding programs. *Heredity* 112: 48-60.

- Daetwyler HD, Calus MPL, Pong-Wong R, de los Campos G, et al. (2013). Genomic prediction in animals and plants: simulation of data, validation, reporting, and benchmarking. *Genetics* 193: 347-365.
- de los Campos G, Hickey JM, Pong-Wong R, Daetwyler HD, et al. (2013). whole-genome regression and prediction methods applied to plant and animal breeding. *Genetics* 193: 327-345.
- Dempster AP (1977). Maximum likelihood from incomplete data via the EM algorithm. *J. R. Stat. Soc. Ser. B* 39: 1-38.
- Eberhart SA (1964). Theoretical relations among single, three-way, and doubled cross hybrids. *Biometrics* 20: 522-539.
- Falconer DS and Mackay TF (1996). Introduction to quantitative genetics. 4th edn. Longman Group Ltd., London.
- Gianola D (2013). Priors in whole-genome regression: the Bayesian alphabet returns. *Genetics* 194: 573-596.
- Gianola D, Fernando RL and Stella A (2006). Genomic-assisted prediction of genetic value with semiparametric procedures. *Genetics* 173: 1761-1776.
- Gianola D, de Los Campos G, Hill WG, Manfredi E, et al. (2009). Additive genetic variability and the Bayesian alphabet. *Genetics* 183: 347-363.
- Goddard ME and Hayes BJ (2007). Genomic selection. *J. Anim. Breed. Genet.* 124: 323-330.
- Hallauer AR, Carena MJ and Miranda Filho JB (2010). Quantitative genetics in maize breeding. 2nd edn. Iowa State University, Ames.
- Heffner EL, Sorrells ME and Jannink JL (2009). Genomic selection for crop improvement. *Crop Sci.* 49: 1-12.
- Henderson CR (1985a). Best linear unbiased prediction of nonadditive genetic merits in non inbreds populations. *J. Anim. Sci.* 60: 111-117.
- Henderson CR (1985b). MIVQUE and REML estimation of additive and nonadditive genetic variances. *J. Anim. Sci.* 61: 113-121.
- Howard R, Carriquiry AL and Beavis WD (2014). Parametric and nonparametric statistical methods for genomic selection of traits with additive and epistatic genetic architectures. *G3* 4: 1027-1246.
- Jenkins MT (1934). Methods of estimating the performance of double crosses in corn. *J. Am. Soc. Agron.* 26: 199-204.
- Jia Y and Jannink J-L (2012). Multiple-trait genomic selection methods increase genetic value prediction accuracy. *Genetics* 192: 513-1522.
- Lee M, Godshalk EB, Lamkey KR and Woodman WW (1989). Association of restriction length polymorphism among maize inbreds with agronomic performance of their crosses. *Crop Sci.* 29: 1067-1071.
- Mackay TF, Stone EA and Ayroles JF (2009). The genetics of quantitative traits: challenges and prospects. *Nat. Rev. Genet.* 10: 565-577.
- Meuwissen THE, Hayes BJ and Goddard ME (2001). Prediction of total genetic value using genome-wide dense marker maps. *Genetics* 157: 1819-1829.
- PANZEA (2014) Data sets for download. [[http://www.panzea.org/lit/data\\_sets.html#NAM\\_map](http://www.panzea.org/lit/data_sets.html#NAM_map)]. Accessed August 22, 2014.
- Patterson HD and Thompson R (1971). Recovery of inter-block information when block sizes are unequal. *Biometrika* 58: 545-554.
- Pérez P and de los Campos G (2014). Genome-wide regression and prediction with the BGLR statistical package. *Genetics* 198: 483-495.
- Pérez-Rodríguez P, Gianola D, Gonzalez-Camacho JM, Crossa J, et al. (2012). Comparison between linear and non-parametric regression models for genome-enabled prediction in Wheat. *G3* 2: 1595-1605.
- R Core Team (2014). R: A language and environment for statistical computing. Available at [<http://www.Rproject>]. Accessed June 2, 2014.
- Reif JC, Hallauer AR and Melchinger AE (2005). Heterosis and heterotic patterns in maize. *Maydica* 50: 215-223.
- Riedelsheimer C, Technow F and Melchinger AE (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.
- Romay MC, Millard MJ, Glaubitz JC, Peiffer JA, et al. (2013). Comprehensive genotyping of the USA national maize inbred seed bank. *Genome Biol.* 14: R55.
- Schnell FW and Cockerham CC (1992). Multiplicative vs. arbitrary gene action in heterosis. *Genetics* 131: 461-469.
- Technow F, Riedelsheimer C, Schrag TA and Melchinger AE (2012). Genomic prediction of hybrid performance in maize with models incorporating dominance and population specific marker effects. *Theor. Appl. Genet.* 125: 1181-1194.
- Technow F, Schrag TA, Schipprack W, Bauer E, et al. (2014). Genome properties and prospects of genomic prediction of hybrid performance in a breeding program of maize. *Genetics* 197: 1343-1355.
- VanRaden PM (2008). Efficient methods to compute genomic predictions. *J. Dairy Sci.* 91: 4414-4423.
- Zeng ZB, Wang T and Zou W (2005). Modeling quantitative trait loci and interpretation of models. *Genetics* 169: 1711-1725.