



## Combined index of genomic prediction methods applied to productivity traits in rice

Matheus Massariol Suela<sup>1</sup>  Leísa Pires Lima<sup>1</sup>  Camila Ferreira Azevedo<sup>1\*</sup>   
 Marcos Deon Vilela de Resende<sup>1,2</sup>  Moysés Nascimento<sup>1</sup>  Fabyano Fonseca e Silva<sup>3</sup> 

<sup>1</sup>Laboratório de Inteligência Computacional, Departamento de Estatística, Universidade Federal de Viçosa (UFV), 36570-900, Campus Universitário, Viçosa, MG, Brasil. E-mail: [camila.azevedo@ufv.br](mailto:camila.azevedo@ufv.br). \*Corresponding author.

<sup>2</sup>Departamento de Engenharia Florestal, Embrapa Florestas, Universidade Federal de Viçosa (UFV), Viçosa, MG, Brasil.

<sup>3</sup>Departamento de Zootecnia, Universidade Federal de Viçosa (UFV), Viçosa, MG, Brasil.

**ABSTRACT:** Rice cultivation has great national and global importance, being one of the most produced and consumed cereals in the world and the primary food for more than half of the world's population. Because of its importance as food, developing efficient methods to select and predict genetically superior individuals in reference to plant traits is of extreme importance for breeding programs. The objective of this research was to evaluate and compare the efficiency of the Delta-p, G-BLUP (Genomic Best Linear Unbiased Predictor), BayesCpi, BLASSO (Bayesian Least Absolute Shrinkage and Selection Operator), Delta-p/G-BLUP index, Delta-p/BayesCpi index, and Delta-p/BLASSO index in the estimation of genomic values and the effects of single nucleotide polymorphisms on phenotypic data associated with rice traits. Use of molecular markers allowed high selective efficiency and increased genetic gain per unit time. The Delta-p method uses the concept of change in allelic frequency caused by selection and the theoretical concept of genetic gain. The Index is based on the principle of combined selection, using the information regarding the additive genomic values predicted via G-BLUP, BayesCpi, BLASSO, or Delta-p. These methods were applied and compared for genomic prediction using nine rice traits: flag leaf length, flag leaf width, panicles number per plant, primary panicle branch number, seed length, seed width, amylose content, protein content, and blast resistance. Delta-p/G-BLUP index had higher predictive abilities for the traits studied, except for amylose content trait in which the method with the highest predictive ability was BayesCpi, being approximately 3% greater than that of the Delta-p/G-BLUP index.

**Key words:** genomic prediction, selection index, genetic gain.

## Índice combinado de métodos de predição genômica aplicado em características de produtividade em arroz

**RESUMO:** A cultura do arroz tem grande importância nacional e mundial por ser um dos cereais mais produzidos e consumidos no mundo, caracterizando-se como o principal alimento de mais da metade da população mundial. Em função de sua importância alimentar, desenvolver métodos eficientes que visam a predição e a seleção de indivíduos geneticamente superiores, quanto a características da planta, é de extrema importância para os programas de melhoramento. Diante disso, o objetivo deste trabalho foi avaliar e comparar a eficiência do método Delta-p, G-BLUP, BayesCpi, BLASSO e o índice Delta-p/G-BLUP, índice Delta-p/BayesCpi e índice Delta-p/BLASSO, na estimação de valores genômicos e dos efeitos de marcadores SNPs (Single Nucleotide Polymorphisms) em dados fenotípicos associados a características de arroz. A utilização de marcadores moleculares permite alta eficiência seletiva e o aumento do ganho genético por unidade de tempo. O método Delta-p utiliza o conceito de mudança na frequência alélica devido à seleção e o conceito teórico de ganho genético. O Índice é baseado no princípio da seleção combinada, utiliza conjuntamente as informações dos valores genômicos aditivos preditos via G-BLUP, BayesCpi ou BLASSO e via Delta-p. Estes métodos foram aplicados e comparados quanto à predição genômica utilizando nove características de arroz (*Oryza sativa*), sendo elas: comprimento da folha bandeira, largura da folha bandeira; número de panículas por planta; número de ramos da panícula primária; comprimento de semente; largura de semente; teor de amilose; teor de proteína; resistência a bruzone. O índice Delta-p/G-BLUP obteve maiores capacidades preditivas para as características estudadas, exceto para a característica Conteúdo de amilose, em que o método que obteve maior capacidade preditiva foi o BayesCpi, sendo aproximadamente 3% superior ao índice Delta-p/G-BLUP.

**Palavras-chave:** predição genômica, índice de seleção e ganho genético.

## INTRODUCTION

Rice (*Oryza sativa*) is one of the most important crops in the world. Increased rice production has played key roles in food security, especially in developing countries in Asia and Africa (CHEN, 2017). Currently, the production

of this crop is approximately 12,327.8 thousand tons (CONAB, 2018). Despite supplying world's current population, it is estimated that by 2050, rice production in the world must increase from 60 to 110% to meet population demand (GODFRAY et al., 2010; TILMAN et al., 2011 RAY et al., 2013). Thus, there is a need for development of new lines

considering improvements in yield over existing varieties. According to SPINDEL et al. (2015), because the process is extremely time-consuming, using conventional breeding and selection methods, it takes ten years on average for elite varieties to be developed and identified.

The continued evolution of sequencing and genotyping technologies has led to a breakthrough in molecular genetics. Such advances have promoted the direct use of the information from the DNA in the identification of genetically superior individuals, thereby shortening the selection cycle to the benefit of plant breeding programs. For this purpose, MEUWISSEN et al. (2001) devised a Genome Wide Selection (GWS), which consists of the analysis of a large number of single nucleotide polymorphisms (SNPs) distributed in the genome, capturing the genes affecting the quantitative trait of interest. According to MEUWISSEN et al. (2001), it can be assumed that some of these molecular markers are in linkage disequilibrium (LD) with quantitative trait loci (QTL), allowing its direct use in the prediction of genomic breeding values (GEBVs) of the individuals subject to selection.

Several methods, such as Bayesian methods, Bayesian Least Absolute Shrinkage and Selection Operator (BLASSO), BayesCpi, and the mixed-model method, Genomic Best Linear Unbiased Predictor (G-BLUP), have been extensively applied to GWS and are recommended for genomic prediction (DE LOS CAMPOS et al., 2012, AZEVEDO et al., 2015). However, new methodologies have been proposed, such as the method called Delta-p (RESENDE, 2015; LIMA et al., 2019), which does not demand an iterative computational method and consequently, does not require evaluation regarding the convergence of results. Such methodology divides the estimation population into two subpopulations, one associated with higher phenotypic values and the other associated with lower phenotypic values. Effects of the markers were estimated non-parametrically using the difference between the allelic frequencies and the genetic gain associated with these two subpopulations. With the goal of combining good properties of different methodologies, LIMA et al. (2019) proposed the use of a genomic index, called the Delta-p/G-BLUP index, which combines estimated genomic values obtained by Delta-p and G-BLUP. The Delta-p/G-BLUP index was more accurate than G-BLUP in genomic prediction. However, the genomic index can combine predictions from several statistical methodologies, as indicated by the literature, such as BLASSO and BayesCpi. Use of other methodologies to compose

the index can be interesting because it allows the use of specific properties of each method of genomic selection in terms of architecture of the evaluated traits. In addition, the Bayesian approach has been used successfully in other areas (MACEDO et al., 2014, GARNERO et al., 2014).

Consequently, the goals of the present study were to evaluate the Delta-p/BLASSO and Delta-p/BayesCpi genomic indexes and compare them to the Delta-p/G-BLUP index in terms of prediction efficiency of additive genomic values of the individuals in reference to photosynthetic yield traits, grain quality, yield, and blast resistance in rice.

## MATERIALS AND METHODS

### *Description of the database*

Database used in this study was composed of nine traits referring to 352 rice accessions (*Oryza sativa*), which were genotyped for 44,100 SNPs markers. The dataset is publicly available, part of two projects, the OryzaSNP Project and the OMAP Project (AMMIRAJU et al., 2006), and available at <https://ricediversity.org/data/>.

Plantations were supervised throughout the access phase, from May to October of 2006 and 2007. A complete block design with two replications was used, in which the planting lines had a length equal to 5 m. Plants were spaced 25 cm apart and there was 0.50 m between rows. Further details can be reported in ZHAO et al. (2011). Quality control procedures were made considering a call rate of 70% and minor allele frequency (MAF) less than 1%. After the quality control of the genomic database, the total was 36,901 SNPs markers.

The nine traits analyzed were considered to have relevant effects on the improvement of rice. They were flag leaf length (FLL), flag leaf width (FLW), amylose content (AC), protein content (PC), panicles number per plant (PNPP), seed length (SL), seed width (SW), primary panicle branch number (PPBN), and blast resistance (BR). The first two traits (FLL and FLW) are associated with the photosynthetic yield of the plant and traits AC and PC are associated with grain quality. Traits PNPP, SL, SW, and PPBN are associated with the production of the plant, whereas BR is associated with the main rice disease.

### *Delta-p method*

The Delta-p method proposed by RESENDE (2015) and LIMA et al. (2019) is based on the concept of changes in allele frequency because of selection and the genetic gain theory (the contrast

between the averages of two subpopulations). The method consists of the following steps:

i) The training population is divided into two subpopulations (subpopulation one and subpopulation two), according to the phenotype of the character corrected for systematic effects;

ii) Calculation of difference between allelic frequencies of the subpopulations,  $\Delta p_i = p_{i1} - p_{i2}$ , where  $p_{i1}$  is the allele frequency of allele A of the  $i$ th marker ( $i = 1, 2, \dots, n$  being the total number of markers) in subpopulation one and  $p_{i2}$  is allele frequency of allele A of the  $i$ th marker in subpopulation two;

iii) Calculation of the average difference between the allelic frequencies of the subpopulations,  $\Delta p_m = \frac{\sum_{i=1}^n \Delta p_{il}}{n}$ ;

iv) Calculation of the average allelic substitution effect  $\alpha_m = \frac{0.5h_a^2(u_1 - u_2)}{2n\Delta p_m}$ , where  $h_a^2$  is the heritability in the restricted sense of the trait,  $u_1$  and  $u_2$  are the averages of phenotypic values in subpopulations one and two, respectively;

v) Calculation of the allelic substitution effect of the  $i$ th marker,  $\hat{\alpha}_i = \frac{\Delta p_i}{\Delta p_m} \alpha_m$ .

vi) Calculation of the additive genomic value of the  $j$ th individual ( $j = 1, 2, \dots, N$  with  $N$  being the total number of individuals in the validation population),  $\hat{a}_j = \sum_{i=1}^n w_{ji} \hat{\alpha}_i$ , where  $w_{ji}$  are the elements of the  $j$ th line of the centered incidence matrix of marker  $W_v$  of the validation population. Incidence matrix for the vectors of additive effects of markers ( $\alpha$ ) is parameterized according to VAN RADEN (2008), VITEZICA et al. (2013), and RESENDE et al. (2014) and is presented below:

$$W = \begin{cases} Se MM, \text{então } 2 - 2p \\ Se Mm, \text{então } 1 - 2p \\ Se mm, \text{então } 0 - 2p \end{cases}$$

#### G-BLUP method

Genomic Best Linear Unbiased Predictor (G-BLUP) method is based on the following linear mixed model:

$y = 1\mu + Za + e$ , where  $y$  is a vector of phenotypes ( $N \times 1$ ,  $N$  being the number of genotyped and phenotyped individuals);  $\mu$  is the general mean and  $1$  the vector with dimension ( $N \times 1$ );  $a$  is the vector of additive genomic values ( $N \times 1$ ) with incidence matrix  $Z$  ( $N \times N$ ), whose assumed distribution is  $a \sim N(0, G_a \sigma_a^2)$ , where  $\sigma_a^2$  is the genetic additive variance and  $G_a$  ( $N \times N$ ) is the additive genomic relationship matrix; and  $e$  is the random residual vector, assumed to be  $e \sim N(0, I \sigma_e^2)$ , where  $\sigma_e^2$  is the residual variance and an identity matrix.

Equations of mixed models for prediction via the G-BLUP method are equivalent to:

$$\begin{bmatrix} X'X & X'Z \\ Z'X & Z'Z + G_a^{-1} \frac{\sigma_e^2}{\sigma_a^2} \end{bmatrix} \begin{bmatrix} \hat{\beta} \\ \hat{a} \end{bmatrix} = \begin{bmatrix} X'y \\ Z'y \end{bmatrix},$$

with the components of variance,  $\sigma_a^2$  and  $\sigma_e^2$ , estimated by the restricted maximum likelihood (REML) method. According to VITEZICA et al. (2013), the genomic relationship matrix for additive effect,  $G_a$ , is given by:

$$G_a = \frac{WW'}{\sum_{i=1}^n (2p_i q_i)},$$

where  $p_i$  and  $q_i$  are the allelic frequencies of the  $i$ th marker,  $W$  is the incidence matrix for the markers in the training population.

#### BLASSO Method

The Bayesian version of the LASSO regression (BLASSO) for genomic selection was proposed by DE LOS CAMPOS et al. (2009). BLASSO (*Bayesian Least Absolute shrinkage and Selection Operator*) includes a common variance term for the genetic and residual effects of markers. Basic linear model for predicting the effects of markers is presented below:

$$y = 1\mu + W\alpha + e, \tag{1}$$

where  $y$  is the vector of phenotypes of training population,  $\mu$  is the general mean,  $1$  is the vector with the same dimension of  $y$  whose elements are equal to 1,  $\alpha$  is the vector of allelic substitution effects of the markers with incidence matrix  $W$ , and  $e$  is the residual vector. The *a priori* distributions of the parameters in terms of an increased hierarchical model are presented below:

$$e | \sigma^2 \sim MVN(0, I\sigma^2)$$

$$\alpha_i | \lambda_a, \sigma^2 \sim N(0, D_a \sigma^2)$$

$$p(\tau_a^2 | \lambda_a^2) = \prod_i \left( \frac{\lambda_a^2}{2} \right) e^{\left[ \frac{-\lambda_a^2 \tau_{ai}^2}{2} \right]}$$

in which MNV represents the multivariate normal distribution,  $\lambda_a$  is the “sharpness” parameter that can be estimated from the data by the MCMC (*Markov chain Monte Carlo*) method using a non-informative *priori*,  $\sigma^2$  has an *a priori* distribution consisting of a scaled inverse chi-square and  $D_a = \text{diag}(\tau_{1a}^2, \tau_{2a}^2, \dots, \tau_{ma}^2)$ . This leads to a double exponential distribution for the effects of allelic substitution (PARK & CASELLA, 2008), as follows:

$$\alpha_i | \lambda_a^2 \sim \text{ExpDupla} \left( 0, \frac{\sigma}{\lambda_a} \right).$$

The additive genetic variance of each marker is given by  $\sigma_{\alpha_i}^2 = \tau_{ai}^2 \sigma^2$  with  $i = 1, 2, \dots, n$ . In this way, the additive genetic variance can be

estimated using the relationship  $\sigma_a^2 = \sum_{i=1}^m 2p_i q_i \sigma_{a_i}^2$  and thus,  $\sigma_a^2 = \sum_{i=1}^m 2p_i q_i \tau_{a_i}^2 \sigma^2$ . The additive genomic values are estimated through the  $\hat{a} = W\hat{\alpha}$ . The full conditional distributions for the parameters of the BLASSO are presented in detail by DE LOS CAMPOS et al. (2009).

#### BayesCpi method

The Bayes Cpi method was proposed by HABIER et al. (2011) to allow election of variables and Bayesian learning with data. The *a priori* distributions assumed for the parameters in the model (1) considering this method are:

$$\alpha_i | \pi, \sigma_{ma}^2 \sim (1 - I_{ai}) 0 + I_{ai} N(0, \sigma_{\alpha}^2)$$

$$\sigma_{\alpha}^2 \sim \chi^{-2}(v_a, s_a^2)$$

in which the indicator variable  $I_a = (I_{a1} \dots I_{am})$  follows a binomial distribution with probability  $\pi$ . Thus, the probability of mixing  $\pi$  will be assigned an *a priori* distribution beta. The additive genetic variance is given by  $\sigma_a^2 = \sum_{i=1}^m 2p_i q_i \sigma_{a_i}^2 I_{ai}$ .

In this study, for Bayesian methods, 300,000 iterations were used for the MCMC algorithms, of which 20,000 were discarded (*burn-in*) to guarantee the heating of the chain and there was a selection of one in 10 iterations (*thin*). The convergence analysis was performed using the criterion proposed by GEWEKE (1992).

#### Genomic index

The genomic index is defined as (RESENDE, 2015; LIMA et al., 2019):

$$I = b_1 \hat{a}_1 + b_2 \hat{a}_2,$$

in which it combines the genomic values predicted through G-BLUP, BLASSO, or BayesCpi ( $\hat{a}_1$ ) and through the Delta-p ( $\hat{a}_2$ ) method, weighted by coefficients  $b_1$  and  $b_2$ , respectively. The weights,  $b_1$  and  $b_2$ , are given respectively by:

$$b_1 = \frac{1 - r_{\hat{a}_2 y}^2 \Delta^2}{1 - r_{\hat{a}_1 y}^2 r_{\hat{a}_2 y}^2 \Delta^2}; b_2 = \frac{1 - r_{\hat{a}_1 y}^2}{1 - r_{\hat{a}_1 y}^2 r_{\hat{a}_2 y}^2 \Delta^2};$$

in which  $r_{\hat{a}_1 y}^2$  is the square of the correlation between the phenotype and the additive genomic values predicted via G-BLUP, BLASSO, or BayesCpi ( $\hat{a}_1$ ),  $r_{\hat{a}_2 y}^2$  is the square of the correlation between the phenotype and the additive genomic values predicted through Delta-p ( $\hat{a}_2$ ) and  $\Delta^2 = \sigma_{\hat{a}_2}^2 / \sigma_{\hat{a}_1}^2$ , that is, the ratio between the additive genetic variances.

#### Cross-validation and comparison between methods

The validation procedure chosen was the *k-fold* process with  $k = 4$ . Thus, the phenotypic

dataset composed of 352 individuals was divided into 4 groups with 88 individuals each. Thus, for each replicate of the analysis, three groups were considered as training populations and used to obtain effects of SNP markers. The other group was considered a validation population and was used to predict the additive genomic values through estimations of effects of markers obtained in the estimation population. Later, the calculation of the efficiency measures was possible, as described below. The process was repeated such that at each step, one of the four groups constituted the validation population. After the end of the validation process, the arithmetic averages and standard deviations of the efficiency measures were used, such that it was possible to report the general results. The efficiency measures used are described below: (i) the molecular heritability was given by  $h^2 = \frac{\sigma_a^2}{\sigma_y^2}$ , where  $\sigma_y^2$  is the phenotypic variance and is the additive genetic variance estimated by REML in G-BLUP and estimated through  $\sigma_a^2 = \sum_{i=1}^m 2p_i q_i \sigma_{a_i}^2$ . Bayesian methods, where  $p_i$  and  $q_i$  were the allelic frequencies, and  $\sigma_{a_i}^2$  was the additive genetic variance of the *i*th marker; (ii) the predictive ability ( $r_{\hat{a}_y}$ ) of the method consisted of the correlation between the estimated genomic value and the phenotypic value; (iii) the regression coefficient between the estimated genomic value and the phenotypic value was given by  $\hat{\beta}_{\hat{a}_y}$ ; (iv) the predictive ability,  $r_{I,y}$ , consisted of correlation of the genomic value estimated by index and phenotypic value; and (v) the regression coefficient was between estimated genomic value via index and phenotypic value given by  $\hat{\beta}_{I,y}$ .

#### Computational resource

All the computational routines of the proposed methods were implemented in R software (R Development Core Team, 2018). For the G-BLUP, the *sommer* package and the *mmer* function were used; for the BLASSO method and BayesCpi, the *BGLR* package and the *BGLR* function were used. Algorithms used for the development of the Delta-p and Delta-p/G-BLUP index methods were implemented by LIMA et al. (2019) and are available at <https://licauefv.wordpress.com/pesquisas-research/>.

## RESULTS AND DISCUSSION

Average results and the respective estimated standard deviations relative to molecular heritability, predictive ability, and regression coefficient between genomic value and phenotypic value associated with the Delta-p and G-BLUP methods, as well as predictive ability and regression

coefficient between the estimated genomic value through the Delta-p/G-BLUP index and phenotypic value are shown in table 1.

The average results and respective standard deviations relative to molecular heritability, predictive ability, and regression coefficient between genomic value and phenotypic value associated with Bayesian methods (BLASSO and BayesCpi), as well as predictive ability and regression coefficient between the estimated genomic value through the index (Delta-p/BLASSO index and Delta-p/BayesCpi index) and phenotypic value are shown in table 2.

#### Predictive ability

Results showed that the Delta-p/G-BLUP, Delta-p/BayesCpi, and Delta-p/BLASSO indices for all traits presented higher predictive abilities than the G-BLUP, BayesCpi, and BLASSO methods, respectively. This can be easily seen when evaluating the relationship between the predictive abilities of the methods, which showed that the Delta-p/G-BLUP index, Delta-p/BLASSO index, and Delta-p/BayesCpi index were, on average, 9.7%, 3.6% and

3.3%, respectively, more efficient in the genomic prediction than the traditionally applied methods, G-BLUP, BLASSO and BayesCpi. It is important to point out that one substantial advantage is that these percentage points in predictive ability have no additional computational cost. Moreover, according to RESENDE et al. (2015), gains of 5% in predictive ability and accuracy are already significant in plant breeding, often equivalent to the gain that is obtained in a complete cycle of improvement genetics. Thus, under genomic selection performed in a short time, these gains are cumulative and grow rapidly. Therefore, it has been shown that the indices caused an improvement in the prediction of the GEBVs because they provided superior predictive abilities over other methods.

For all traits, the Delta-p method presented lower predictive values compared to that of the G-BLUP, BLASSO and BayesCpi methods because of the different genetic information used in each of the methods. According to LIMA et al. (2019), the Delta-p method uses only linkage unbalance information, whereas the other methods, such as

Table 1 - Average and standard deviation of molecular heritability ( $h^2$ ), predictive ability ( $r_{\hat{a}_y}$ ) and regression coefficient between the genomic value and the phenotypic value ( $\hat{\beta}_{\hat{a}_y}$ ) from each method (Delta-p, G-BLUP), predictive ability ( $r_{I_y}$ ) and regression coefficient between the estimated genomic value via index and the phenotypic value ( $\hat{\beta}_{I_y}$ ) from the method Delta-p/G-BLUP index.

Method	Trait	$h^2$	$r_{\hat{a}_y}$	$\hat{\beta}_{\hat{a}_y}$	$r_{I_y}$	$\hat{\beta}_{I_y}$
Delta-p	FLL	0.34±0.08	0.21±0.07	0.39±0.18		
	FLW	0.25±0.01	0.55±0.08	1.14±0.25		
	PNPP	0.13±0.02	0.69±0.07	1.93±0.31		
	PPBN	0.20±0.03	0.36±0.06	0.79±0.14		
	SL	0.43±0.02	0.41±0.18	0.67±0.32		
	SW	0.16±0.01	0.50±0.06	1.26±0.26		
	AC	0.10±0.02	0.61±0.04	1.98±0.16		
	PC	0.13±0.02	0.31±0.06	0.86±0.15		
	BR	0.25±0.04	0.42±0.19	0.84±0.37		
G-BLUP	FLL	0.22±0.28	0.47±0.03	1.00±0.14	0.63±0.14	0.28±0.10
	FLW	0.50±0.05	0.77±0.04	1.09±0.14	0.80±0.04	0.85±0.14
	PNPP	0.66±0.03	0.83±0.03	1.02±0.09	0.83±0.03	0.92±0.10
	PPBN	0.36±0.08	0.63±0.05	1.06±0.17	0.70±0.08	0.60±0.08
	SL	0.56±0.08	0.75±0.09	1.01±0.10	0.77±0.09	0.76±0.19
	SW	0.65±0.04	0.84±0.03	1.04±0.08	0.84±0.03	0.93±0.09
	AC	0.59±0.03	0.80±0.09	1.04±0.09	0.80±0.09	0.93±0.13
	PC	0.25±0.05	0.46±0.05	0.92±0.08	0.59±0.06	0.46±0.06
	BR	0.44±0.08	0.67±0.08	1.02±0.11	0.72±0.06	0.69±0.05

Flag leaf length (FLL); Flag leaf width (FLW); Amylose content (AC); Panicles number per plant (PNPP); Primary panicle branch number (PPBN); Seed length (SL); Seed width (SW); Protein content (PC); Blast resistance (BR).

Table 2 - Average and standard deviation of molecular heritability ( $h^2$ ), Predictive ability ( $r_{\hat{a}_y}$ ) and regression coefficient between the genomic value and the phenotypic value ( $\hat{\beta}_{\hat{a}_y}$ ) from each method (BLASSO and BayesCpi), predictive ability ( $r_{I_y}$ ) and regression coefficient between the estimated genomic value via the index and the phenotypic value ( $\hat{\beta}_{I_y}$ ) from each method (Delta-p/BLASSO index and Delta-p/BayesCpi index).

Method	Trait	$h^2$	$r_{\hat{a}_y}$	$\hat{\beta}_{\hat{a}_y}$	$r_{I_y}$	$\hat{\beta}_{I_y}$
BLASSO	FLL	0.48±0.15	0.48±0.09	1.11±0.29	0.53±0.09	0.57±0.19
	FLW	0.70±0.10	0.76±0.03	1.11±0.09	0.78±0.03	0.91±0.09
	PNPP	0.77±0.02	0.81±0.04	1.01±0.11	0.82±0.04	0.91±0.11
	PPBN	0.49±0.03	0.63±0.02	1.11±0.09	0.65±0.02	0.80±0.07
	SL	0.83±0.18	0.74±0.08	1.07±0.17	0.75±0.08	0.82±0.23
	SW	0.84±0.03	0.83±0.04	1.08±0.12	0.83±0.03	0.97±0.15
	AC	0.81±0.02	0.81±0.01	1.12±0.06	0.82±0.01	1.00±0.04
	PC	0.41±0.01	0.48±0.06	1.00±0.04	0.52±0.03	0.70±0.05
	BR	0.65±0.05	0.68±0.07	1.06±0.12	0.71±0.09	0.77±0.14
BayesCpi	FLL	0.47±0.17	0.50±0.09	1.17±0.32	0.54±0.09	0.60±0.20
	FLW	0.71±0.09	0.76±0.03	1.10±0.09	0.78±0.03	0.90±0.08
	PNPP	0.78±0.02	0.81±0.04	1.01±0.11	0.82±0.04	0.90±0.11
	PPBN	0.51±0.01	0.64±0.02	1.11±0.10	0.66±0.02	0.79±0.07
	SL	0.83±0.18	0.74±0.08	1.07±0.16	0.76±0.08	0.82±0.23
	SW	0.84±0.03	0.83±0.04	1.08±0.12	0.83±0.04	0.97±0.16
	AC	0.82±0.01	0.82±0.02	1.14±0.06	0.83±0.02	1.02±0.05
	PC	0.42±0.03	0.48±0.06	1.00±0.05	0.52±0.04	0.69±0.06
	BR	0.62±0.04	0.69±0.08	1.09±0.13	0.71±0.09	0.80±0.17

Flag leaf length (FLL); Flag leaf width (FLW); Amylose content (AC); Panicles number per plant (PNPP); Primary panicle branch number (PPBN); Seed length (SL); Seed width (SW); Protein content (PC); Blast resistance (BR).

G-BLUP and Bayesian methods, also used the relationship information between individuals. In addition, AZEVEDO et al. (2016) reported that when genomic prediction considered only linkage imbalance, the predictive ability reported should be less than or equal to that derived from the genomic prediction that also considers the relationship between individuals, which corroborates the results reported in our study. However, it was perceived that the index together with the G-BLUP is able to capture more genetic information that benefits genomic prediction.

In addition, the G-BLUP, BLASSO, and BayesCpi methods presented similar predictive abilities, being in agreement with the results reported in the literature (AZEVEDO et al., 2015, GIANOLA, 2013, DE LOS CAMPOS et al., 2012) that point out the similarity of several methods in terms of predictive ability regarding the prediction of genomic values. GUO et al. (2014), using G-BLUP, also reported similar values for the predictive ability for the same traits analyzed.

#### Regression Coefficient

Interest in GWS is that the regression coefficient between the phenotype and the estimated

genomic value is close to one, indicating that these values are non-biased. For regression coefficients below one (1), it is understood that the genomic values are overestimated and for coefficients above one (1), genomic values are underestimated. Thus, according to the results, genomic values estimated by the three indices considered were overestimated, except for the trait of amylose content that obtained a regression coefficient equal to one in the Delta-p/BLASSO index method and underestimated the values in the Delta-p/BayesCpi method. In addition, G-BLUP obtained regression coefficients closer to one than did the Delta-p and Delta-p/G-BLUP index methods. Additionally, the BayesCpi method exhibited values closer to one than did the Delta-p/BayesCpi index. Lower values of regression coefficients reported for these indices may have occurred because of the Delta-p method because as previously reported, this method generates regression coefficients more than one. In turn, for the traits flag leaf width, seed width, and amylose content, it was observed that Delta-p/BLASSO index method obtained regression coefficient values closer to one in relation to the BLASSO method.

### Heritabilities

When analyzing heritability, it was observed that the BLASSO and BayesCpi methods presented similar heritability estimates, which corroborates the results obtained by AZEVEDO et al. (2015) who also verified similarities between Bayesian methods to estimate genomic heritability. Heritability estimated by G-BLUP were similar to those reported by GUO et al. (2014) considering the same dataset. It was verified that the methods Delta-p and G-BLUP, resulted in smaller values for heritability in relation to the Bayesian methods. According to XING & ZHANG (2010) and VALLURU et al. (2014), quantitative traits, as were the traits used in this study, are generally known because they have low heritability and are difficult to investigate.

Reports of estimates of heritability obtained through *pedigree* for some of the traits, such as, panicles number per plant, flag leaf width, flag leaf length, and seed length, are reported in the literature (XU et al., 2018; SUMANTH et al., 2017; AKINWALE et al., 2011; SEYOUM et al., 2012; SINGH et al., 2011; OLADOSU et al., 2014). However, according to DE LOS CAMPOS & SORENSEN (2013) and DE LOS CAMPOS et al. (2015), these heritability values are always superior to genomic or molecular heritability. This superiority is caused by molecular heritability being a fraction of the heritability obtained via the *pedigree* that is captured by the markers.

### CONCLUSION

In general, the Delta-p/G-BLUP index has more predictive ability for genomic values than traditional methods (G-BLUP, BLASSO, and BayesCpi) and Bayesian indexes, besides being easy to implement and requiring cost computation. Conversely, the genomic indexes presented greater bias in the predictions of the individual genomic values. Results indicated a greater potentiality of use of rank indexes for the selection of genetically superior individuals and not the exact inference about how much they will produce when commercially planted.

### ACKNOWLEDGMENTS

The authors thank the following Brazilian funding organizations: the Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG), the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) and the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) for their financial support.

### DECLARATION OF CONFLICT OF INTERESTS

The authors declare no conflict of interest. The founding sponsors had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, and in the decision to publish the results.

### AUTHOR'S CONTRIBUTIONS

All authors contributed equally to the design and writing of the manuscript. All authors critically reviewed the manuscript and approved the final version.

### REFERENCES

- AKINWALE, M. G. et al. Heritability and correlation coefficient analysis for yield and its components in rice (*Oryza sativa L.*). **African Journal of Plant Science**, v.5, n.3, p.207-212, 2011. Available from: <[http://www.academicjournals.org/app/webroot/article/article1379945851\\_Akinwale%20et%20al.pdf](http://www.academicjournals.org/app/webroot/article/article1379945851_Akinwale%20et%20al.pdf)>. Accessed: Sep. 13, 2018. doi: 10.5897/AJPS.
- AMMIRAJU, J.S.S. et al. The *Oryza* bacterial artificial chromosome library resource: construction and analysis of 12 deep-coverage large-insert BAC libraries that represent the 10 genome types of the genus *Oryza*. **Genome Research**, v.16, n.1, p.140-147, 2006. Available from: <<http://www.genome.org/cgi/doi/10.1101/gr.3766306>>. Accessed: Nov. 9, 2018. doi: 10.1101/gr.3766306.
- AZEVEDO, C.F. et al. New accuracy estimators for genomic selection with application in a cassava (*Manihot esculenta*) breeding program. **Genet Mol Res**, v.15, n.4, 2016. Available from: <<http://dx.doi.org/10.4238/gmr.15048838>>. Accessed: Nov. 7, 2018. doi: 10.4238/gmr.15048838.
- AZEVEDO, C.F. et al. Ridge, LASSO and Bayesian Additive-dominance genomic models. **BMC Genetics**, v.16, p.105, 2015. Available from: <<https://bmcgenet.biomedcentral.com/articles/10.1186/s12863-015-0264-2>>. Accessed: Sep. 11, 2018. doi: 10.1186/s12863-015-0264-2.
- CHEN, S. et al. Genome-wide study of an elite rice pedigree reveals a complex history of genetic architecture for breeding improvement. **Scientific reports**, v.7, p.45685, 2017. Available from: <<https://www.nature.com/articles/srep45685>>. Accessed: Aug. 23, 2018. doi: 10.1038/srep45685.
- CONAB. COMPANHIA NACIONAL DE ABASTECIMENTO. **Acompanhamento de safra brasileiro – grãos: Nono levantamento, junho 2018 – safra 2017/2018**. Brasília: Companhia Nacional de Abastecimento. Available from: <<http://www.conab.gov.br>>. Accessed: Nov. 12, 2018.
- DE LOS CAMPOS, G. et al. Genomic heritability: what is it?. **PLoS Genetics**, v.11, n.5, p.e1005048, 2015. Available from: <<https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1005048>>. Accessed: Sep. 20, 2018. doi: 10.1371/journal.pgen.1005048.
- DE LOS CAMPOS, G. et al. Predicting quantitative traits with regression Models for Dense Molecular Markers and Pedigree. **Genetics**, v.182, n.1, p.375-385, 2009. Available from: <<https://doi.org/10.1534/genetics.109.101501>>. Accessed: Nov. 7, 2018. doi: 10.1534/genetics.109.101501.

- DE LOS CAMPOS, G. et al. Whole genome regression and prediction methods applied to plant and animal breeding. **Genetics**, v.193, p.327–45, 2012. Available from: <<https://doi.org/10.1534/genetics.112.143313>>. Accessed: Nov. 10, 2018.
- DE LOS CAMPOS, G.; SORENSEN D. A. A commentary on Pitfalls of predicting complex traits from SNPs. **Nature Reviews Genetics**, 14: 894-894, 2013. Available from: <<http://dx.doi.org/10.1038/nrg3457-c1>>. Accessed: Sep. 20, 2018.
- GARNERO, A. V. et al. Inferência bayesiana aplicada à estimação de herdabilidades dos parâmetros da curva de crescimento de fêmeas da raça Nelore. **Ciência Rural**, v.43, n.4, p.702-708, 2014. Accessed: Sep. 20, 2018. doi: 10.1590/S0103-84782013005000029.
- GEWEKE, J. **Evaluating the Accuracy of Sampling-Based Approaches to the Calculation of Posterior Moments**. In Bayesian Statistics 4 (eds. J.M. Bernardo, J. Berger, A.P. Dawid and A.F.M. Smith), Oxford: Oxford University Press, 169-193, 1992.
- GIANOLA, D. Priors in whole-genome regression: the bayesian alphabet returns. **Genetics**, v.194, n.3, p.573–96, 2013. Available from: <<https://doi.org/10.1534/genetics.113.151753>>. Accessed: Nov. 7, 2018. doi: 10.1534/genetics.113.151753.
- GODFRAY, H. C. J. et al. Food Security: The Challenge of Feeding 9 Billion People. **Science**, 327(5967), p.812–818, 2010. Available from: <<http://science.sciencemag.org/content/327/5967/812>>. Accessed: Nov. 9, 2018. doi: 10.1126/science.1185383.
- GUO, Z. et al. The impact of population structure on genomic prediction in stratified populations. **Theoretical and applied genetics**, v.127, p.749-762, 2014. Available from: <<https://www.ncbi.nlm.nih.gov/pubmed/24452438>>. Accessed: Nov. 08, 2018. doi: 10.1007/s00122-013-2255-x.
- HABIER, D. et al. Extension of the Bayesian alphabet for genomic selection. **BMC bioinformatics**, v.12, n.1, p.186, 2011. Available from: <<https://doi.org/10.1186/1471-2105-12-186>>. Accessed: Nov. 7, 2018. doi: 10.1186/1471-2105-12-186.
- LIMA, L.P. et al. New insights into genomic selection through population-based non-parametric prediction methods. **Sci. Agric**, v.76, n.4, p.290-298, July/August, 2019. doi: <http://dx.doi.org/10.1590/1678-992X-2017-0351>.
- MACEDO, L.R. et al. Modelagem hierárquica Bayesiana na avaliação de curvas de crescimento de suínos genotipados para o gene halotano. **Ciência Rural**, v.44, n.10, p.1853-1859, 2014. Accessed: Nov. 08, 2018. doi: 10.1590/0103-8478cr20131278.
- MEUWISSEN, T.H.E. et al. Prediction of total genetic value using genome-wide dense marker maps. **Genetics**, v.157, p.1819–29, 2001. Available from: <<https://doi.org/10.1534/genetics.116.189795>>. Accessed: Nov. 10, 2018.
- OLADOSU, Y. et al. Genetic Variability and Selection Criteria in Rice Mutant Lines as Revealed by Quantitative Traits. **The Scientific World Journal**, v. 2014, p.190531, 2014. Available from: <<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4241329/>>. Accessed: Sep. 23, 2018. doi: 10.1155/2014/190531.
- PARK, T.; CASELLA, G. The Bayesian LASSO. **Journal of the American Statistical Association**, v.103, n.482, p.681-686, 2008. Available from: <<https://doi.org/10.1198/01621450800000337>>. Accessed: Nov. 7, 2018. doi: 10.1198/01621450800000337.
- R Development Core Team. **R: A language and environment for statistical computing**. R Foundation for Statistical Computing, Vienna, Austria, 2018. Available from: <<http://www.R-project.org/>>.
- RAY, D. K. et al. Yield Trends Are Insufficient to Double Global Crop Production by 2050. **PLoS One**, 8(6): e66428, 2013. Available from: <<https://doi.org/10.1371/journal.pone.0066428>>. Accessed: Nov. 9, 2018. doi: 10.1371/journal.pone.0066428.
- RESENDE, M.D.V. et al. **Estatística Matemática, Biométrica e Computacional: Modelos Mistos, Multivariados, Categóricos e Generalizados (REML/BLUP), Inferência Bayesiana, Regressão Aleatória, Seleção Genômica, QTL-GWAS, Estatística Espacial e Temporal, Competição, Sobrevida.** 1. ed. Visconde do Rio Branco: Suprema, 1v, 881p, 2014.
- RESENDE, M.D.V. et al. Multi generation index in the within progenies bulk method for breeding of self-pollinated plants. **Crop Science**, v.55, p.1202-1211, 2015. Available from: <<https://dl.sciencesocieties.org/publications/cs/abstracts/55/3/1202>>. Accessed: Nov. 7, 2018. doi: 10.2135/cropsci2014.08.0580.
- RESENDE, M.D.V. **Genética Quantitativa e de Populações**. Visconde do Rio Branco: Suprema, 2015. 463p.
- SEYOUM, M. et al. Genetic variability, heritability, correlation coefficient and path analysis for yield and yield related traits in upland rice (*Oryza sativa* L.). **Journal of Plant Sciences**, v.7, n.1, p.13-22, 2012. Available from: <<http://docsdrive.com/pdfs/academicjournals/jps/2012/13-22.pdf>>. Accessed: Sep. 25, 2018. ISSN 1816-4951 / doi: 10.3923/jps.2012.13.22.
- SINGH, S. K. et al. Assessment of genetic variability for yield and its component characters in rice (*Oryza sativa* L.). **Research in Plant Biology**, v.1, n.4, p.73-76, 2011. Available from: <<http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.863.3827&rep=rep1&type=pdf>>. Accessed: Sep. 13, 2018. doi: 10.23910/ijbsm/2018.9.1.3c0818.
- SPINDEL, J. et al. Genomic selection and association mapping in rice (*Oryza sativa*): effect of trait genetic architecture, training population composition, marker number and statistical model on accuracy of rice genomic selection in elite, tropical rice breeding lines. **PLoS genetics**, v.11, n.2, p.e1004982, 2015. Available from: <<http://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1004982>>. Accessed: Sep. 10, 2018. doi: 10.1371/journal.pgen.1004982.
- SUMANTH, V. et al. Estimation of genetic variability, heritability and genetic advance for grain yield components in rice (*Oryza sativa* L.). **Journal of Pharmacognosy and Phytochemistry**, v.6, n.4, p.1437-1439, 2017. Available from: <<http://www.phytojournal.com/archives/2017/vol6issue4/PartU/6-4-59-319.pdf>>. Accessed: Sep. 20, 2018. E-ISSN: 2278-4136.
- TILMAN, D. et al. Global Food Demand and the Sustainable Intensification of Agriculture. **Proceedings of the National Academy of Sciences**, 108(50):20260–20264, 2011. Available from: <<https://doi.org/10.1073/pnas.1116437108>>. Accessed: Nov. 9, 2018. doi: 10.1073/pnas.1116437108.
- VALLURU, R. et al. Genetic and molecular bases of yield associated traits: A translational biology approach between rice and wheat. **Theoretical and Applied Genetics**, v.127, n.7,

- p.1463–1489, 2014. Available from: <[https://www.researchgate.net/profile/Matthew\\_Reynolds5/publication/262978594\\_Genetic\\_and\\_molecular\\_bases\\_of\\_yield-associated\\_traits\\_A\\_translational\\_biology\\_approach\\_between\\_rice\\_and\\_wheat/links/55b7c1c308ae092e965742e5/Genetic-and-molecular-bases-of-yield-associated-traits-A-translational-biology-approach-between-rice-and-wheat.pdf?origin=publication\\_detail](https://www.researchgate.net/profile/Matthew_Reynolds5/publication/262978594_Genetic_and_molecular_bases_of_yield-associated_traits_A_translational_biology_approach_between_rice_and_wheat/links/55b7c1c308ae092e965742e5/Genetic-and-molecular-bases-of-yield-associated-traits-A-translational-biology-approach-between-rice-and-wheat.pdf?origin=publication_detail)>. Accessed: Sep. 11, 2018. doi: 10.1007/s00122-014-2332-9.
- VAN RADEN, P. M. Efficient Methods to compute genomic predictions. **Journal of Dairy Science**, Champaign, v.91, n.11, p.4414-4423, 2008. Available from: <<https://doi.org/10.3168/jds.2007-0980>>. Accessed: Nov. 7, 2018. doi: 10.3168/jds.2007-0980.
- VITEZICA, Z. G. et al. On the additive and dominance variance and covariance of individuals within the genomic selection scope. **Genetics**, Austin, v.195, n.4, p.1223-1230, 2013. Available from: <<https://doi.org/10.1534/genetics.113.155176>>. Accessed: Nov. 7, 2018. doi: 10.1534/genetics.113.155176.
- XING, Y.; ZHANG, Q. Genetic and Molecular Bases of Rice Yield. **Annual Review of Plant Biology**, v. 61, p. 421–442, 2010. Available from: <[https://www.researchgate.net/profile/Yongzhong\\_Xing/publication/41654874\\_Genetic\\_and\\_Molecular\\_Bases\\_of\\_Rice\\_Yield/links/02e7e51de056f2702c000000/Genetic-and-Molecular-Bases-of-Rice-Yield.pdf?origin=publication\\_detail](https://www.researchgate.net/profile/Yongzhong_Xing/publication/41654874_Genetic_and_Molecular_Bases_of_Rice_Yield/links/02e7e51de056f2702c000000/Genetic-and-Molecular-Bases-of-Rice-Yield.pdf?origin=publication_detail)>. Accessed: Sep. 16, 2018. doi: 10.1146/annurev-arplant-042809-112209.
- XU, Y. et al. Genomic selection of agronomic traits in hybrid rice using an NCII population. **Rice**, v.11, p.32, 2018. Available from: <<https://www.ncbi.nlm.nih.gov/pubmed/29748895>>. Accessed: Sep. 03, 2018. doi: 10.1186/s12284-018-0223-4.
- ZHAO, K. et al. Genome-wide association mapping reveals a rich genetic architecture of complex traits in *Oryza sativa*. **Nature Communications**, v.2, p.467, 2011. Available from: <[www.nature.com/naturecommunications](http://www.nature.com/naturecommunications)>. Accessed: Nov. 9, 2018. doi: 10.1038/ncomms1467.