

Multi-trait genomic selection indexes applied to identification of superior genotypes

Lidiane Aparecida Silva¹ , Marco Antônio Peixoto¹ , Leonardo de Azevedo Peixoto¹ , Juan Vicente Romero¹ , Leonardo Lopes Bhering^{1,*} 

1. Universidade Federal de Viçosa – Departamento de Biologia Geral – Laboratório de Biometria – Viçosa (MG), Brazil.

Received: Aug. 24, 2020 | **Accepted:** May 3, 2021

Section Editor: Freddy Mora

***Corresponding author:** leonardo.bhering@ufv.br

How to cite: Silva, L. A., Peixoto, M. A., Peixoto, L. A., Romero, J. V. and Bhering, L. L. (2021) Multi-trait genomic selection indexes applied to identification of superior genotypes. *Bragantia*, 80, e3621. <https://doi.org/10.1590/1678-4499.20200381>

ABSTRACT: Most studies on genomic selection in plant breeding compare different statistical methods of univariate approach. However, multi-trait methodologies should be considered since they allow the simultaneous selection of superior genotypes in several economic traits. Here, the aims were to compare the selection accuracy and efficiency of the multivariate partial least square (MPLS) method compared with random regression best linear unbiased predictor (rrBLUP), Bayesian Lasso (Blasso) and univariate partial least square (UPLS) and to develop genomic selection indexes efficient for superior genotypes identification in plant breeding. Ten F_2 populations with 800 individuals were simulated, considering four traits with different heritabilities. Genomic selection analyses using rrBLUP, Blasso, UPLS, and MPLS were conducted. Four genomic selection indexes were elaborated by the sum of the marker effects obtained for each trait, weighted by the respective residual variance. Multi-trait indexes were developed based on the assumptions of each methodology mentioned (rrBLUP, Blasso, UPLS, and MPLS), and were denominated I-rrBLUP, I-Blasso, I-UPLS, and I-MPLS. Processing time, selective accuracy, selection gains, and selection coincidence were used to compare the methods and the selection indexes proposed. The MPLS method had similar results compared to UPLS method for the low heritability traits and was less efficient than the rrBLUP and Blasso. The genome selection indexes provided the highest total genetic gains. The I-rrBLUP and I-MPLS indexes stood out for high efficiency in selecting superior genotypes in the shortest processing time. Results suggest that the genomic selection indexes proposed in this study may be promising for plant breeding programs.

Key words: quantitative genetics, biometry, multi-trait analysis, genomic selection, lower heritability, phenotypic selection.

INTRODUCTION

Genomic selection (Meuwissen et al. 2001) is based on the prediction of the genomic estimated breeding value (GEBV) from the relationship between the individuals' phenotype and thousands of molecular markers widely distributed in the genome. This high density of markers increases the probability of each quantitative trait loci (QTL) of the target trait to be in linkage disequilibrium with at least one marker. It allows for better selection accuracy and early direct selection without limiting the number of QTL that control the expression of a quantitative trait (Bernardo and Yu 2007; Meuwissen et al. 2001; Xu et al. 2017).

Genomic selection has been currently described by several authors and has provided promising results in several areas of genetics and plant breeding (Peixoto et al. 2017; Persa et al. 2020; Silva et al. 2017). Many statistical methods can be applied to genomic selection (Crossa et al. 2017), and the choice and efficiency of application of these methods depend mainly on the genetic architecture of the quantitative trait under study (Daetwyler et al. 2013; de los Campos et al. 2013).

The multi-trait selection is relevant, hence superior varieties combine optimal attributes for several traits simultaneously. In addition, it can impact positively decreasing the cycle time in plant breeding programs, especially using genomic selection



(Covarrubias-Pazarán et al. 2018; Watson et al. 2019). However, combining multi-traits in an efficient structure of selection indexes is often a complex and difficult task, once there is reduction in selection gains and correlation between traits as the number of traits assessed increases (Cruz et al. 2012). Moreover, univariate approaches are most used in genomic selection (Sant'Anna et al. 2019). Thus, studies proposing and testing the efficiency of multi-trait genomic selection in plant breeding programs should be developed.

The regression method via partial least square (PLS) can be an efficient alternative for allowing the multi-trait selection (Azevedo et al. 2013). The general purpose of PLS is the formation of components that capture the highest amount of information possible from the explanatory variables to predict the principal dependent variable (Boulesteix and Strimmer 2007). Moreover, this method is efficient to solve the problems of multicollinearity and high dimensionality caused by the high number of markers in relation to the number of individuals genotyped (Azevedo et al. 2013).

Breeders and biometricians have advocated the use of selection indexes. This strategy consists in establishing a new trait from a linear combination of all the target traits, considering that the weight coefficients are estimated to maximize the correlation between the index and the true genetic values of the genotypes to be selected. Several studies have addressed the application of selection indexes to plant breeding, leading to gains in a set of traits simultaneously (França et al. 2016; Junqueira et al. 2016; Kumar et al. 2016; Silva et al. 2016). Thus, the interest for studying the methodology efficiency in genomic selection has increased (Ceron-Rojas et al. 2015; Covarrubias-Pazarán et al. 2018; Fernandes et al. 2018).

Therefore, this study aimed to: (i) compare the selection accuracy and efficiency of the multivariate partial least square (MPLS) method with univariate genomic selection methods as random regression best linear unbiased predictor (rrBLUP), Bayesian Lasso (Blasco), and univariate partial least square (UPLS); and, (ii) develop genomic selection indexes efficient in the identification of superior genotypes in different traits, simultaneously, to be applied in plant breeding programs.

MATERIAL AND METHODS

Simulation of phenotypic and genotypic data

The phenotypic and genotypic data used in the application of genomic selection statistical methods were simulated using the Genes software (Cruz 2013). First, the genome was simulated considering ten linkage groups, similar to a diploid species $2n = 2x = 20$. Each linkage group was simulated with 100 cM, comprising 100 codominant molecular markers, spaced equidistantly (1 cM), totaling 1000 markers. The number of markers used in the simulation is considered adequate for the study purpose and is similar with other simulations studies (Oliveira et al. 2021; Peixoto et al. 2016, Sant'Anna et al. 2019, 2021).

Afterward, contrasting homozygous parents were simulated. For each locus, one parent was designated as A_1A_1 homozygote, and the other was designated as A_2A_2 contrasting homozygote. Thus, the cross between parent 1 and parent 2 generated the F_1 population, with all the markers being heterozygous (A_1A_2).

Using the genome and the parents obtained from the selfing of individuals of the F_1 population, ten F_2 mapping populations were simulated, each one containing 800 individuals. Each individual from the F_1 population produced 5,000 gametes, and the random combination of two of these gametes generated the F_2 population. This procedure was repeated until forming all the individuals in each population. The simulated F_2 populations were coded with 0, 1, and 2, where 0 and 2 corresponded to homozygous individuals (A_1A_1 or A_2A_2) and 1 referred to heterozygote individuals (A_1A_2), for a given locus.

Then, the phenotypes, i.e., the quantitative traits, were simulated considering the binomial distribution, the additive gene action, and the absence of dominance among the alleles (additive model). Four quantitative traits, controlled by 200 loci each, were simulated. Traits were simulated with heritability values of 20, 40, 60, and 80 (hereafter C1, C2, C3, and C4, respectively). Two alleles per locus without the presence of QTL with major effects were considered. Thus, the effect of each QTL was defined by $A_1A_1 = \mu + a$; $A_1A_2 = \mu$; $A_2A_2 = \mu - a$, in which a expresses the additive effect of each gene in the F_2 population. Therefore, the phenotypes of the individuals (Y_i) were generated according to the model in Eq. 1:

$$Y_i = \mu + \sum_{j=1}^{200} \alpha_j + \varepsilon_i \quad (1)$$

where μ is the overall mean of the trait; α_j is the genetic effect in each locus; and ε_i is the environmental effect.

Cross-validation

For all genomic selection methodologies evaluated, cross-validation was performed considering five folds with 50 replications. Thus, each F_2 population with 800 individuals was divided into five equal groups containing 160 individuals each and used as a validation population. The training populations were composed of 640 individuals and used to estimate the marker effects. The agreement between the genetic values predicted through estimates from the training population was validated in each group considered as a validation population.

Genomic selection methods used

The rrBLUP (Meuwissen et al. 2001) uses best linear unbiased prediction (BLUP), considering that all markers have the same variance (absence of major effect genes). The rrBLUP was analyzed using the mixed.solve function of the rrBLUP package (Endelman 2011).

The Blasso method (Park and Casella 2008) considers a variance for each marker. This method was analyzed using the Bayesian generalized linear regression (BGLR) function of the BGLR package (Pérez-Rodríguez and de los Campos 2011). A total of 20,000 burn-ins, 10 thin, and 100,000 saved iterations, as obtained from the Markov chain Monte Carlo (MCMC) method, was used.

The UPLS and MPLS methods were analyzed in the R software (R Development Core Team 2020), using the PLS package and the partial least square regression (pls) function. For these methods, the optimal number of components that explained 80% of the total variation of the markers was used (variable X). Thus, approximately 30 components were considered in the analysis for the different simulated populations.

Elaboration of genomic selection indexes

The principal of the genomic selection index proposed in this study is the estimation of the final effect for each marker, using the sum of the effects of this marker for each trait, which are weighted by their respective residual variance, as shown in Eq. 2:

$$Ef_m = \sum_{i=1}^4 \frac{M_{mi}}{\sigma_i^2} = \frac{M_{m1}}{\sigma_1^2} + \frac{M_{m2}}{\sigma_2^2} + \frac{M_{m3}}{\sigma_3^2} + \frac{M_{m4}}{\sigma_4^2} \quad (2)$$

where Ef_m is the final effect of marker m weighted by the four traits evaluated, considering $m = 1, 2, \dots, 1000$ markers; M_{mi} is the estimated effect for marker m for trait i ($i = 1, 2, 3, 4$); and σ_i^2 is the residual variance obtained for trait i .

Following this principle, four genomic selection indexes were developed, one for each evaluated method: rrBLUP selection index (I-rrBLUP); Blasso selection index (I-Blasso); UPLS selection index (I-UPLS); and MPLS selection index (I-MPLS). Thus, the final effects of the markers ($Ef_1, Ef_2, \dots, Ef_{1000}$) estimated by the genomic selection index proposed for each method were considered in the estimation of the GEBVs.

Comparison between methods and indexes of genomic selection

The phenotypic (ac_p) and genotypic (ac_g) selective accuracy were calculated, as in Eqs. 3 and 4:

$$ac_f = \frac{cor(PV, GEBV)}{\sqrt{h^2}} \quad (3)$$

$$ac_g = cor(GV, GEBV) \quad (4)$$

where h^2 is the heritability value of the trait; PV is the phenotypic value; GV is the true genetic value; and $GEBV$ is the genomic estimated breeding value.

The processing time of each method was computed in seconds. An AxB factor analysis was performed in a completely randomized design (CRD). The selection accuracy (A factor) and time of each genomic selection method (B factor) were compared by Tukey's test at the 5% probability, considering each evaluated trait.

Afterward, 80 individuals (10%) with the highest phenotypic values and true genetic values were selected, and the selection gains (SG) were estimated by the direct (selection of individuals with superior performance in a trait and estimation of the gain with selection is calculated in the same trait) and indirect selection (selection of individuals with superior performance in a trait and gain with selection is calculated in the other traits) methods, using Eq. 5:

$$SG (\%) = \frac{SD \times h^2 \times 100}{\bar{X}_o} \quad (5)$$

where SD is the selection differential, which was estimated by: $SD = \bar{X}_s - \bar{X}_o$, where \bar{X}_s is the mean of the selected individuals, and \bar{X}_o is the mean of the initial population.

Similarly, 80 individuals with the highest GEBVs, estimated from each genomic selection method and genomic selection index, were selected. By the individuals ranked from the highest GEBVs, the respective SG was calculated, using the phenotypic values and true genetic values.

The calculation of the SG using the true genetic values simulated for each trait considered a heritability value of 100%, i.e., $h^2 = 1$. Conversely, the calculation of the SG using the phenotypic values considered simulated heritability values of 20, 40, 60 and 80% for traits C1, C2, C3 and C4, respectively. The selection efficiency of the methodologies evaluated was estimated through the selection coincidence (SC), for 10% of superior individuals, identified for the true genetic value, phenotypic value and each genomic selection method or index. Thus, the SC was estimated as in Eq. 6:

$$SC \% = \frac{NS}{NT} \times 100 \quad (6)$$

where NS is the number of individuals selected on both methods in contrast and NT is the total number of individuals selected.

All measures used to compare the different methodologies evaluated in this study were estimated considering the mean of the ten populations simulated for the four traits with different heritability values.

RESULTS

Table 1 shows that the analysis of processing time for the Blasso method was significantly greater than that of the other genomic selection methods, evaluated by Tukey's test at 5% probability level, regardless of the heritability value of the trait. Thus, considering the lower computational demand (shorter processing time), these results suggest that the rrBLUP, UPLS, and MPLS methods were more advantageous.

Table 1. Comparison between the genomic selection methods random regression best linear unbiased predictor (rrBLUP), Bayesian Lasso (Blasso), univariate partial least square (UPLS) and multivariate partial least square (MPLS) regarding the processing time (in seconds) analysis for the four traits with different heritability values.

Methods	C1	C2	C3	C4
Blasso	304.27 ^a	295.89 ^a	273.24 ^a	253.57 ^a
rrBLUP	2.31 ^b	2.62 ^b	2.63 ^b	2.30 ^b
UPLS	0.47 ^b	0.46 ^b	0.53 ^c	0.53 ^b
MPLS	0.72 ^b	0.98 ^{bc}	0.86 ^{bc}	0.94 ^b

Note. Means followed by the same letter in the column do not differ statistically at the 5% probability level by Tukey's test. Simulated traits C1, C2, C3, and C4, with heritability values of 20, 40, 60 and 80%, respectively.

Tukey's test revealed significant differences for selection accuracy for the genomic selection methods, considering the four traits with different heritability values from the phenotypic value and true genetic value (Table 2). No significant differences were observed between rrBLUP and Blasso methods, in any of the selection accuracy evaluations. The selection accuracy of the UPLS method was inferior compared with other genomic selection methods, considering the low heritability traits. However, for the trait of higher heritability value, results obtained by the UPLS method were statistically equal to those of the rrBLUP and Blasso methods (Tables 2).

Table 2. Comparison between the genomic selection methods random regression best linear unbiased predictor (rrBLUP), Bayesian Lasso (Blasso), univariate partial least square (UPLS) and multivariate partial least square (MPLS) regarding the genotypic and phenotypic selection accuracy for the four traits evaluated.

Methods	C1	C2	C3	C4
Blasso	0.91 ^a	0.94 ^a	0.96 ^a	0.98 ^a
rrBLUP	0.91 ^a	0.94 ^a	0.96 ^a	0.98 ^a
UPLS	0.46 ^b	0.65 ^b	0.79 ^b	0.90 ^{ab}
MPLS	0.79 ^a	0.84 ^a	0.78 ^b	0.84 ^b
Phenotypic accuracy				
Blasso	0.89 ^a	0.94 ^a	0.96 ^a	0.98 ^a
rrBLUP	0.89 ^a	0.94 ^a	0.96 ^a	0.97 ^a
UPLS	0.46 ^b	0.64 ^b	0.79 ^b	0.90 ^{ab}
MPLS	0.80 ^a	0.84 ^a	0.79 ^b	0.82 ^b

Note. Means followed by the same letter in the column do not differ statistically at the 5% probability level by Tukey's test. Simulated traits C1, C2, C3 and C4, with heritability values of 20, 40, 60 and 80%, respectively.

Unlike the other methods, the selection accuracy results for the MPLS method did not increase with the increase in the heritability value of the trait (Table 2). The selection accuracy results obtained for the MPLS method were statistically equal to those detected for the rrBLUP and Blasso methods for the low heritability traits. However, for the high heritability traits, the results of the MPLS were lower than those from rrBLUP and Blasso methods.

Table 3 shows the estimates of direct and indirect selection gains, considering the 80 individuals selected from the highest phenotypic values and true genetic values for the four traits evaluated. The greatest gains were obtained by direct selection. As expected, the direct gains in each trait obtained from the true genetic values were greater than those based on the phenotypic values.

For phenotypic values, high heritability traits (C3 and C4) had the greatest total selection gains, 5.52 and 5.92% (Table 3). These values correspond to only 50.41 and 54.31%, respectively, of the maximum total gains that can be achieved when considering the true genetic values for these traits (10.95 and 10.59%). Total gains obtained for the low heritability traits, considering the phenotypic values (3.96 and 4.55%), were far from the maximum total gain that can be achieved with the true genetic values (10.75 and 10.34%) (Table 3). This fact confirms that phenotypic selection is less efficient for low heritability traits.

Table 3. Direct and indirect selection gains, based on the individuals selected from the phenotypic values and true genetic values, for the four traits evaluated with different heritability values.

Indirect response Direct selection	Traits				Total
	C1	C2	C3	C4	
	Phenotypic value				
C1	1.68	0.47	0.74	1.07	3.96
C2	0.30	2.36	0.85	1.04	4.55
C3	0.35	0.68	2.89	1.60	5.52
C4	0.49	0.69	1.36	3.38	5.92
	True genetic value				
C1	3.79	2.25	2.33	2.38	10.75
C2	2.26	3.70	2.34	2.04	10.34
C3	2.34	2.36	3.73	2.53	10.95
C4	2.35	2.00	2.47	3.78	10.59

Note. Simulated traits C1, C2, C3 and C4 with heritability values of 20, 40, 60 and 80%, respectively.

Table 4 shows the direct and indirect selection gains estimated in the true genetic values, based on the 80 individuals ranked by the GEBVs, obtained in each genomic selection method. Direct and indirect gains, as well as total gains provided by the rrBLUP and Blasso methods, were remarkably like each other (Table 4). These results were as expected when considering the maximum gains that could be achieved via true genetic values (Table 3). These methods stood out due to higher total gains obtained in the selection for the low heritability traits (C1 and C2), showing superiority when compared with the selection based only on the phenotypes of the individuals (Table 3).

Table 4. Direct and indirect selection gains estimated in the true genetic values based on the individuals ranked in descending order of GEBVs, obtained by the genome-wide selection methods, for the simulated traits C1, C2, C3 and C4, with heritability values of 20, 40, 60 and 80%, respectively.

Indirect response Direct selection	Traits				Total
	C1	C2	C3	C4	
	rrBLUP				
C1	3.46	2.27	2.39	2.53	10.65
C2	2.30	3.51	2.26	2.07	10.15
C3	2.47	2.44	3.59	2.59	11.09
C4	2.43	2.06	2.45	3.70	10.63
	Blasso				
C1	3.46	2.29	2.40	2.54	10.69
C2	2.30	3.51	2.26	2.08	10.15
C3	2.45	2.43	3.60	2.58	11.05
C4	2.45	2.06	2.46	3.70	10.67
	UPLS				
C1	1.94	1.17	1.26	1.44	5.82
C2	1.51	2.51	1.52	1.44	6.99
C3	1.94	1.91	3.04	2.10	8.99
C4	2.17	1.85	2.26	3.45	9.73
	MPLS				
C1	1.98	1.24	1.31	1.35	5.88
C2	1.97	3.18	1.94	1.77	8.87
C3	2.38	2.37	3.56	2.52	10.82
C4	2.37	2.05	2.47	3.67	10.57

Note. rrBLUP: random regression best linear unbiased predictor; Blasso: Bayesian Lasso; UPLS: univariate partial least square; MPLS: multivariate partial least square.

The UPLS and MPLS methods exhibited the lowest direct gains for the four traits evaluated when compared with the rrBLUP and Blasso methodologies (Table 4). Furthermore, these methods were efficient only for the high heritability traits since the total selection gains tend to be close to the gains obtained with the phenotypic selection for the low heritability traits (Table 3). The total gains obtained from the true genetic values when applying any genomic selection index proposed in this study (Table 5) are greater than those obtained by direct and indirect selection in any methodology presented previously (Tables 3 and 4). These results suggested the superiority of the selection indexes in relation to both the multi-trait method MPLS and the univariate methods rrBLUP, Blasso and UPLS.

Table 5. Gains with selection estimated in phenotypic values and true genetic values, based on individuals ranked in descending order of GEBVs, obtained by genomic selection indexes, for the simulated traits C1, C2, C3 and C4, with heritability values of 20, 40, 60 and 80%, respectively.

Index	C1	C2	C3	C4	Total
I-Blasso	0.60	1.05	1.90	2.88	6.43
I-rrBLUP	0.60	1.05	1.90	2.88	6.43
I-UPLS	0.65	1.09	1.94	2.95	6.63
I-MPLS	0.66	1.09	1.94	2.82	6.51
True genetic value					
I-Blasso	2.73	2.55	3.06	3.57	11.90
I-rrBLUP	2.73	2.55	3.05	3.57	11.90
I-UPLS	2.52	2.38	2.88	3.36	11.15
I-MPLS	2.72	2.60	3.10	3.50	11.92

Note. I-rrBLUP: rrBLUP selection index; I-Blasso: Blasso selection index; I-UPLS: UPLS selection index; and I-MPLS: MPLS selection index.

The total gain for the phenotypic selection did not exceed 6% in the best case (Table 3), while with the use of the indexes the total gain was above 6.5% (Table 5), with a better distribution of gains among the four traits. As expected, the selection indexes can achieve higher total gains, but the gains obtained for each trait are lesser than the maximum gains that could be achieved in the direct selection of each trait based on the true genetic values (Table 3). Among the selection indexes, the I-UPLS had the least total gain when selecting the individuals based on the true genetic values. Conversely, it presented the highest total selection gain when considering the phenotypic values (Table 5).

The selection coincidence of 80 individuals, between the methods genomic selection, genome selection indexes, phenotypic values, and true genetic values for the low (C1 and C2) and high (C3 and C4) heritability traits are shown in Tables 6 and 7, respectively.

The selection coincidence estimated among the individuals selected from the phenotypic values and true genetic values ranged from 30.75 to 65.13%, from the lowest to the highest heritability value (Tables 6 and 7). The rrBLUP and Blasso methods were more efficient in the selection of superior genotypes based on the phenotypic values, mainly for the low heritability traits. This fact is evidenced in Table 6, which shows that the selection coincidences of the individuals selected by the GEBVs obtained by the rrBLUP and Blasso methods, considering that the true genetic value for C1 was twice (70% for both methods) when compared with the selection coincidence between the phenotypic value and the true genetic value (30.75%).

Also, the selection coincidence obtained between the rrBLUP and Blasso methods in the selection of 80 individuals were higher than 98%, regardless of the heritability of the trait (Tables 6 and 7). These results confirm once again that the rrBLUP and Blasso methods, despite the different approaches, showed similar behavior and results are equally efficient in the identification and selection of genetically superior individuals.

Conversely, for C1 trait (Table 6), the selection coincidences obtained by the individuals selected based on the GEBVs, estimated by the UPLS and MPLS methods, and the true genetic values were 32.25 and 33.75%, respectively. These values are remarkably similar to that obtained by the individuals selected via phenotypic value and true genetic value (30.75%). Therefore, the MPLS and UPLS methods had the same efficiency as the phenotypic selection for the low heritability traits, not justifying the application of these methodologies in these cases.

Table 6. Comparison between genomic selection methods and indexes, by the selection coincidence analysis, for the simulated traits C1 ($h^2 = 20\%$), above the main diagonal, and C2 ($h^2 = 40\%$), below the main diagonal.

	GV	PV	Blasso	rrBLUP	UPLS	MPLS	I-Blasso	I-rrBLUP	I-UPLS	I-MPLS
GV		30.75	70.00	70.00	32.25	33.75	48.00	47.88	43.50	47.63
PV	42.25		34.00	34.38	56.50	54.50	26.00	26.25	27.38	29.00
Blasso	76.38	44.25		99.13	36.88	38.13	50.88	50.75	46.25	50.00
rrBLUP	76.50	44.38	99.50		36.63	37.63	50.75	50.75	46.38	49.88
UPLS	45.50	62.13	50.13	50.00		58.25	26.13	26.25	28.88	28.75
MPLS	63.00	46.88	69.50	69.63	50.38		26.88	26.88	27.88	29.63
I-Blasso	45.75	29.25	48.13	47.88	32.50	40.25		99.25	76.25	88.00
I-rrBLUP	45.88	29.38	48.25	48.00	32.75	40.25	99.25		76.38	88.25
I-UPLS	42.25	31.13	43.50	43.25	35.38	37.13	76.25	76.38		72.38
I-MPLS	47.00	30.50	48.88	48.88	34.00	42.75	88.00	88.25	72.38	

Note. GV: true genetic value; PV: phenotypic value; rrBLUP: random regression best linear unbiased predictor; Blasso: Bayesian Lasso; UPLS: univariate partial least square; MPLS: multivariate partial least square; I-rrBLUP: rrBLUP selection index; I-Blasso: Blasso selection index; I-UPLS: UPLS selection index; and I-MPLS: MPLS selection index.

It was seen that the indexes produce a ranking that generates a greater gain in all the traits. Comparing the selection through phenotypic value and selection through I-Blasso index, both for C3 trait (Table 7), the I-Blasso index has a 57.37% higher efficiency than the ranking from the phenotypic value for the trait C3. In most scenarios, the selection coincidences between the individuals selected by the genomic selection indexes and the true genetic values, regardless of the heritability of the trait, were higher when compared with those obtained between the individuals selected by phenotypic values and true genetic values.

Table 7. Comparison between genomic selection methods and indexes by the selection coincidence analysis for the simulated traits: C3 ($h^2 = 60\%$), above the main diagonal, and C4 ($h^2 = 80\%$), below the main diagonal.

	GV	PV	Blasso	rrBLUP	UPLS	MPLS	I-Blasso	I-rrBLUP	I-UPLS	I-MPLS
GV		54.38	81.25	80.63	58.63	78.88	58.00	57.88	53.75	58.38
PV	65.13		56.00	56.13	69.38	53.13	42.63	42.75	44.88	43.88
Blasso	85.13	68.50		98.75	61.63	85.00	60.75	60.50	56.13	60.63
rrBLUP	85.00	68.50	98.63		61.75	84.75	61.00	60.75	56.38	60.88
UPLS	70.00	78.75	74.75	75.63		57.13	45.75	45.88	51.63	45.75
MPLS	82.38	63.88	87.75	88.13	70.13		57.75	57.88	53.13	59.50
I-Blasso	75.50	61.50	80.38	80.13	67.63	77.63		99.25	76.25	88.00
I-rrBLUP	75.50	61.38	80.63	80.38	67.75	77.75	99.25		76.38	88.25
I-UPLS	66.38	66.50	70.25	70.25	73.75	66.38	76.25	76.38		72.38
I-MPLS	73.00	58.50	74.63	74.38	62.00	76.25	88.00	88.25	72.38	

Note. GV: true genetic value; PV: phenotypic value; rrBLUP: random regression best linear unbiased predictor; Blasso: Bayesian Lasso; UPLS: univariate partial least square; MPLS: multivariate partial least square; I-rrBLUP: rrBLUP selection index; I-Blasso: Blasso selection index; I-UPLS: UPLS selection index; and I-MPLS: MPLS selection index.

Among the proposed indexes, the highest selection coincidence of the superior individuals was observed between the indexes I-Blasso and I-rrBLUP (99.25%) for C4 trait, which selected practically the same individuals. Despite the low to moderate efficiency revealed by the MPLS method in the selection of superior genotypes for the low heritability traits (C1 and C2), the I-MPLS was indicated as more efficient in some scenarios (Table 6). Conversely, the individuals selected by the I-UPLS resulted in the lowest selection coincidence with the individuals selected via true genetic values when compared with the other indexes. Moreover, for some traits, this selection coincidence was lower than or equal to the selection coincidence between the phenotypic value and the true genetic value (Tables 6 and 7).

DISCUSSION

Phenotypic selection in plant breeding is difficult since most of the traits of interest are of quantitative nature, i.e., they are controlled by several genes, have low heritability value, and are consequently strongly influenced by the environment (Falconer and Mackay 1996). This fact is evidenced by the present results since the phenotypic selection showed less efficiency in the selection coincidence with the superior genotypes, especially for the low heritability traits (Tables 6 and 7). In addition, the direct and indirect selection gains through phenotypic values were lower compared with the maximum genetic gains that can be achieved with the selection (Table 3), resulting in lower gains in relation to what could be achieved with the use of more accurate selection methods, such as genomic selection (Spindel et al. 2015; Zhang et al. 2016).

The UPLS and MPLS methods allowed a high dimensionality reduction (97%), which resulted in a shorter analysis processing time, being more advantageous than Blasso (Table 1). Methodologies based on partial least squares regression allow large amounts of data to be quickly analyzed; they have high statistical efficiency and computational speed to obtain the GEBVs estimates of the individuals (Boulesteix and Strimmer 2007; Solberg et al. 2009). However, these methods did not show greater computational advantages over the rrBLUP method, considering the evaluation of 800 individuals and 1000 markers (Table 1).

The number of markers used in this study was low, however was considered adequate for the purpose and agree in other similar studies (Oliveira et al. 2021; Sant'Anna et al. 2019, 2021). Bhering et al. (2015) and Spindel et al. (2015) showed that a small number of markers can be used for high precision genomic selection for many traits. Peixoto et al. (2016), in a study with simulated data, concluded that a genomic selection model that uses 300–800 markers is sufficient to capture all the genetic variance, and to decrease the residual variance, to obtain the maximum prediction precision of an F_2 population.

In general, the UPLS method provided low selection accuracy for the low heritability traits (Table 2). Solberg et al. (2009), based on simulated data, compared different genomic selection methodologies to be applied in animal breeding and verified the superiority of the selection accuracy of a Bayesian method (Bayes B) in relation to the UPLS method. These authors demonstrated that the selection accuracy of the UPLS method decreased for the low heritability traits. Azevedo et al. (2014) reported low predictive capacity for the UPLS method, indicating the inferior performance of this method in relation to rrBLUP.

The UPLS method was very similar to the phenotypic selection, mainly for the low heritability traits, resulting in low total gains (Tables 3 and 4). This fact reveals the high selection coincidence of the superior individuals by the UPLS method and the phenotypic selection. This selection coincidence was higher than that obtained when considering the true genetic values of the individuals (Tables 6 and 7). Thus, these results suggest that the UPLS method was inefficient in the identification and accurate selection of genetically superior individuals. The prediction inefficiency of this method was also reported by Azevedo et al. (2014) in a comparative analysis of six genomic selection methodologies applied to animal breeding.

The selection accuracy obtained by the MPLS method was greater than or equal to UPLS method (Table 2). Azevedo et al. (2013) reported higher accuracy for the MPLS method and inferred that this methodology is more appropriate for genomic selection than the UPLS method, since the MPLS realistically captures the nature of the traits, considering the correlations between them. However, the present study revealed that, for traits of lower heritability values, considering the selection gains (Table 4) and the selection coincidences in the selection of the genetically superior individuals (Tables 6 and 7), the MPLS had similar behavior compared to UPLS, resulting in low total gains and low efficiency in the identification and selection of genetically superior individuals. These results indicate that, for the low heritability traits, the MPLS approach is similar to the phenotypic selection and is recommended for genomic selection only for traits with 60% heritability.

In this study, 200 QTL were simulated for each quantitative trait, distributed randomly in the genome, without the presence of QTLs of major effects, which resulted in similarity between the Blasso and rrBLUP methods. Considering the normal distribution of QTL throughout the genome, methods based on mixed models are equally efficient to Bayesian methods. Bhering et al. (2015), comparing rrBLUP and Blasso, verified no significant difference between the methods of the selection accuracy. Therefore, results of the present study indicate that the rrBLUP method was the most efficient, since it presented high values of selection accuracy, high gains, and selection coincidence of superior genotypes in the shortest processing time.

The selection indexes proposed in this study were more efficient and provided the highest total selection gains when compared with the other univariates (Blasso, rrBLUP, and UPLS) and multi-trait (MPLS) methods (Tables 4 and 5). According to Dekkers (2007), the application of the index theory in genomic selection is promising and can significantly increase selection gains, especially for the low heritability traits, which was evidenced in this study by the results.

However, the direct gains estimated through selection indexes for each trait were lesser than those obtained by the other genomic selection methods (Tables 4 and 5). These results are in agreement with the theory of selection indexes, which provide relatively higher total gains, causing reduced gains in each individual trait, compensating for this reduction by the better distribution of favorable gains in the other traits (Resende et al. 2016). In this sense, in relation to the GEBVs, the selection coincidences obtained for the selection indexes were always lesser than those obtained via the Blasso and rrBLUP methods (Tables 6 and 7). However, the individuals selected by the indexes have higher performance in all the traits, simultaneously, unlike those selected by the genomic selection methods, which consider the improvement in one trait at a time.

The I-MPLS led to high total genetic gains and high efficiency in the selection and identification of the superior individuals, showing results similar to those observed for I-Blasso and I-rrBLUP (Table 5). Covarrubias-Pazaran et al. (2018) demonstrated that the use of multi-trait models could increase selection accuracy and consequently result in more powerful and reliable selection indexes. To obtain the proposed genomic selection indexes, first, data should be analyzed by genomic selection methods, aiming to estimate markers effects. Thus, considering the low efficiency demonstrated by the UPLS, the lowest selection coincidences between the indexes in the selection of superior individuals are in relation to the I-UPLS (Tables 6 and 7). This index had the lowest efficiency in the identification and selection of superior genotypes, resulting in the lowest selection gains and coincidences when considering the true genetic value, being the least indicated for application in breeding programs.

On the other hand, the I-rrBLUP, I-Blasso and I-MPLS indexes properly identified the best genotypes, allowing gains in all traits simultaneously. Genomic selection indexes are a promising tool to be applied in plant breeding programs, since they allow the direct early selection by the information of molecular markers, as well as the selection of superior individuals in a set of traits of economic interest, in a shorter processing time (Ceron-Rojas et al. 2015). The results showed that the I-rrBLUP and I-MPLS indexes are the most indicated since they were highly efficient in the identification and selection of superior genotypes, in a shorter time, and with less computational demand. The genomic selection indexes developed in this study were efficient in the identification of superior genotypes in different traits, simultaneously, and are suitable for plant breeding programs.

CONCLUSION

The rrBLUP is the most recommended method for selecting superior genotypes based on selection accuracy, selection coincidence, selection gain and processing time. However, from the selection indexes proposed, I-rrBLUP and I-MPLS indexes were the most advantageous due to the higher efficiency in the selection of superior genotypes in shorter processing time. The indexes proposed here overcome the results from the univariate models, regardless the heritability, being its use largely indicated genomic selection in any breeding program aiming the improvement of several traits, simultaneously.

AUTHORS' CONTRIBUTION

Conceptualization: Silva, L. A., Peixoto, L. A. and Romero, J. V.; **Methodology:** Silva, L. A., Peixoto, L. A. and Romero, J. V.; **Investigation:** Silva, L. A., Peixoto, M. A., Peixoto, L. A. and Romero, J. V.; **Writing – Original Draft:** Silva, L. A.; **Writing – Review and Editing:** Silva, L. A. and Peixoto, M. A.; **Funding Acquisition:** Bhering, L. L.; **Resources:** Bhering, L. L.; **Supervision:** Bhering, L. L.

DATA AVAILABILITY STATEMENT

All datasets were generated or analyzed in the current study.

FUNDING

Coordenação de Aperfeiçoamento de Pessoal de Nível Superior
[<https://doi.org/10.13039/501100002322>]
Finance Code 001

Conselho Nacional de Desenvolvimento Científico e Tecnológico
[<https://doi.org/10.13039/501100003593>]
Grants No. 301630/2017-8 and 405679/2018-1.

ACKNOWLEDGMENTS

Not applicable.

REFERENCES

- Azevedo, C. F., Fonseca e Silva, F., Rezende, M. D. V., Peternelli, L. A., Guimaraes, S. E. F. and Lopes, P. S. (2013). Quadrados mínimos parciais uni e multivariado aplicados na seleção genômica para características de carcaça em suínos. *Ciência Rural*, 43, 1642-1649. <https://doi.org/10.1590/S0103-84782013000900017>
- Azevedo, C F, Silva, F. F., Resende, M. D. V, Lopes, M. S., Duijvesteijn, N., Guimarães, S. E. F., Lopes, P. S., Kelly, M. J., Viana, J. M. S. and Knol, E. F. (2014). Supervised independent component analysis as an alternative method for genomic selection in pigs. *Journal of Animal Breeding and Genetics*, 131, 452-461. <https://doi.org/10.1111/jbg.12104>
- Bernardo, R. and Yu, J. (2007). Prospects for Genomewide Selection for Quantitative Traits in Maize. *Crop Science*, 47, 1082-1090. <https://doi.org/10.2135/cropsci2006.11.0690>
- Bhering, L. L., Junqueira, V. S., Peixoto, L. A., Cruz, C. D. and Laviola, B. G. (2015). Comparison of methods used to identify superior individuals in genomic selection in plant breeding. *Genetics and Molecular Research*, 14, 10888-10896. <https://doi.org/10.4238/2015.September.9.26>
- Boulesteix, A.-L. and Strimmer, K. (2007). Partial least squares: a versatile tool for the analysis of high-dimensional genomic data. *Briefings in Bioinformatics*, 8, 32-44. <https://doi.org/10.1093/bib/bbl016>
- Ceron-Rojas, J. J., Crossa, J., Arief, V. N., Basford, K., Rutkoski, J., Jarquín, D., Alvarado, G., Beyene, Y., Semagn, K. and DeLacy, I. (2015). A Genomic Selection Index Applied to Simulated and Real Data. *G3: Genes, Genomes, Genetics*, 5, 2155-2164. <https://doi.org/10.1534/g3.115.019869>
- Covarrubias-Pazaran, G., Schlautman, B., Diaz-Garcia, L., Grygleski, E., Polashock, J., Johnson-Cicalese, J., Vorsa, N., Iorizzo, M. and Zalapa, J. (2018). Multivariate GBLUP Improves Accuracy of Genomic Selection for Yield and Fruit Weight in Biparental Populations of *Vaccinium macrocarpon* Ait. *Frontiers in Plant Science*, 9, 1310. <https://doi.org/10.3389/fpls.2018.01310>
- Crossa, J., Pérez-Rodríguez, P., Cuevas, J., Montesinos-López, O., Jarquín, D., de los Campos, G., Burgueño, J., González-Camacho, J. M., Pérez-Elizalde, S., Beyene, Y., Dreisigacker, S., Singh, R., Zhang, X., Gowda, M., Roorkiwal, M., Rutkoski, J. and Varshney, R. K.

- (2017). Genomic Selection in Plant Breeding: Methods, Models, and Perspectives. *Trends in Plant Science*, 22, 961-975. <https://doi.org/10.1016/j.tplants.2017.08.011>
- Cruz, C. D., Regazzi, A. J., Carneiro, P. C. S. (2012). Modelos biométricos aplicados ao melhoramento genético. Viçosa: UFV.
- Cruz, C. D. (2013). GENES - a software package for analysis in experimental statistics and quantitative genetics. *Acta Scientiarum Agronomy*, 35, 271-276. <https://doi.org/10.4025/actasciagron.v35i3.21251>
- Daetwyler, H. D., Calus, M. P. L., Pong-Wong, R., de los Campos, G. and Hickey, J. M. (2013). Genomic Prediction in Animals and Plants: Simulation of Data, Validation, Reporting, and Benchmarking. *Genetics*, 193, 347-365. <https://doi.org/10.1534/genetics.112.147983>
- de los Campos, G., Vazquez, A. I., Fernando, R., Klimentidis, Y. C. and Sorensen, D. (2013). Prediction of Complex Human Traits Using the Genomic Best Linear Unbiased Predictor. *PLoS Genetics*, 9, e1003608. <https://doi.org/10.1371/journal.pgen.1003608>
- Dekkers, J. C. M. (2007). Prediction of response to marker-assisted and genomic selection using selection index theory. *Journal of Animal Breeding and Genetics*, 124, 331-341. <https://doi.org/10.1111/j.1439-0388.2007.00701.x>
- Endelman, J. B. (2011). Ridge Regression and Other Kernels for Genomic Selection with R Package rrBLUP. *The Plant Genome*, 4, 250-255. <https://doi.org/10.3835/plantgenome2011.08.0024>
- Falconer, D. S. and Mackay, T. F. C. (1996). Introduction to Quantitative Genetics. Harlow: Longman.
- Fernandes, S. B., Dias, K. O. G., Ferreira, D. F. and Brown, P. J. (2018). Efficiency of multi-trait, indirect, and trait-assisted genomic selection for improvement of biomass sorghum. *Theoretical and Applied Genetics*, 131, 747-755. <https://doi.org/10.1007/s00122-017-3033-y>
- França, Â. E. D., Parrella, R. A. C., Souza, V. F., Bastos, G. Q., Nunes, J. A. R. and Schaffert, R. E. (2016). Seleção simultânea em progênies de sorgo-sacarina por meio de índices de seleção. *Pesquisa Agropecuária Brasileira*, 51, 1737-1743. <https://doi.org/10.1590/s0100-204x2016001000005>
- Junqueira, V. S., Peixoto, L. A., Laviola, B. G., Bhering, L. L., Mendonça, S., Costa, T. S. A. and Antoniassi, R. (2016). Bayesian Multi-Trait Analysis Reveals a Useful Tool to Increase Oil Concentration and to Decrease Toxicity in *Jatropha curcas* L. *PLoS ONE*, 11, e0157038. <https://doi.org/10.1371/journal.pone.0157038>
- Kumar, B., Guleria, S. K., Khanorkar, S. M., Dubey, R. B., Patel, J., Kumar, V., Parihar, C. M., Jat, S. L., Singh, V. and Yatish, K. R. (2016). Selection indices to identify maize (*Zea mays* L.) hybrids adapted under drought-stress and drought-free conditions in a tropical climate. *Crop and Pasture Science*, 67, 1087-1095. <https://doi.org/10.1071/CP16141>
- Meuwissen, T. H. E., Hayes, B. J. and Goddard, M. E. (2001). Prediction of Total Genetic Value Using Genome-Wide Dense Marker Maps. *Genetics*, 157, 1819-1829. <https://doi.org/10.1093/genetics/157.4.1819>
- Oliveira, G. F., Nascimento, A. C. C., Nascimento, M., Sant'Anna, I. C., Romero, J. V., Azevedo, C. F., Bhering, L. L. and Moura, E. T. C. (2021). Quantile regression in genomic selection for oligogenic traits in autogamous plants: A simulation study. *Plos one*, 16, e0243666. <https://doi.org/10.1371/journal.pone.0243666>
- Park, T. and Casella, G. (2008). The Bayesian Lasso. *Journal of the American Statistical Association*, 103, 681-686. <https://doi.org/10.1198/016214508000000337>
- Peixoto, L. A., Bhering, L. L., and Cruz, C. D. (2016). Determination of the optimal number of markers and individuals in a training population necessary for maximum prediction accuracy in F_2 populations by using genomic selection models. *Genetics and Molecular Research*, 15, gmr15048874. <https://doi.org/10.4238/gmr15048874>
- Peixoto, L. A., Laviola, B. G., Alves, A. A., Rosado, T. B. and Bhering, L. L. (2017). Breeding *Jatropha curcas* by genomic selection: A pilot assessment of the accuracy of predictive models. *PLoS ONE*, 12, e0173368. <https://doi.org/10.1371/journal.pone.0173368>

- Pérez-Rodríguez, P. and de los Campos, G. (2011). BGLR: A Statistical Package for Whole Genome Regression and Prediction (p. 21). [Accessed Jun. 09, 2018]. Available at: <https://r-forge.r-project.org/projects/bglr/>
- Persa, R., Bernardeli, A. and Jarquin, D. (2020). Prediction Strategies for Leveraging Information of Associated Traits under Single- and Multi-Trait Approaches in Soybeans. *Agriculture*, 10, 308. <https://doi.org/10.3390/agriculture10080308>
- R Development Core Team. (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing. [Accessed Aug. 14, 2018]. Available at: <https://www.r-project.org/>
- Resende, M. D. V., Ramalho, M. A. P., Carneiro, P. C. S., Carneiro, J. E. S., Batista, L. G. and Gois, I. B. (2016). Selection Index with Parents, Populations, Progenies, and Generations Effects in Autogamous Plant Breeding. *Crop Science*, 56, 530-546. <https://doi.org/10.2135/cropsci2015.05.0303>
- Sant'Anna, I. C., Ferreira, R. A. D. C., Nascimento, M., Silva, G. N., Carneiro, V. Q., Cruz, C. D., Oliveira, M. S. and Chagas, F. E. (2019). Multigenerational prediction of genetic values using genome-enabled prediction. *PLoS ONE*, 14, e0210531. <https://doi.org/10.1371/journal.pone.0210531>
- Sant'Anna, I. C., Silva, G. N., Nascimento, M. and Cruz, C. D. (2021). Subset selection of markers for the genome-enabled prediction of genetic values using radial basis function neural networks. *Acta Scientiarum. Agronomy*, 43. <https://doi.org/10.4025/actasciagron.v43i1.46307>
- Silva, L. A., Resende, R. T., Ferreira, R. A. D. C., Silva, G. N., Kist, V., Barbosa, M. H. P., Nascimento, M. and Bhering, L. L. (2016). Selection index using the graphical area applied to sugarcane breeding. *Genetics and Molecular Research*, 15, gmr15038711. <https://doi.org/10.4238/gmr.15038711>
- Silva, L. A., Peixoto, L. A., Teodoro, P. E., Rodrigues, E. V., Laviola, B. G. and Bhering, L. L. (2017). Path analysis and canonical correlations for indirect selection of *Jatropha* genotypes with higher oil yield. *Genetics and Molecular Research*, 16, gmr16019562. <https://doi.org/10.4238/gmr16019562>
- Solberg, T. R., Sonesson, A. K., Woolliams, J. A. and Meuwissen, T. H. E. (2009). Reducing dimensionality for prediction of genome-wide breeding values. *Genetics Selection Evolution*, 41, 29. <https://doi.org/10.1186/1297-9686-41-29>
- Spindel, J., Begum, H., Akdemir, D., Virk, P., Collard, B., Redoña, E., Atlin, G., Jannink, J.-L. and McCouch, S. R. (2015). Correction: 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*, 11, e1004982. <https://doi.org/10.1371/journal.pgen.1005350>
- Watson, A., Hickey, L. T., Christopher, J., Rutkoski, J., Poland, J. and Hayes, B. J. (2019). Multivariate Genomic Selection and Potential of Rapid Indirect Selection with Speed Breeding in Spring Wheat. *Crop Science*, 59, 1945-1959. <https://doi.org/10.2135/cropsci2018.12.0757>
- Xu, Y., Li, P., Zou, C., Lu, Y., Xie, C., Zhang, X., Prasanna, B. M. and Olsen, M. S. (2017). Enhancing genetic gain in the era of molecular breeding. In *Journal of Experimental Botany*, 68, 2641-2666. <https://doi.org/10.1093/jxb/erx135>
- Zhang, J., Song, Q., Cregan, P. B. and Jiang, G. L. (2016). Genome-wide association study, genomic prediction, and marker-assisted selection for seed weight in soybean (*Glycine max*). *Theoretical and Applied Genetics*, 129, 117-130. <https://doi.org/10.1007/s00122-015-2614-x>