PAB Pesquisa Agropecuária Brasileira 60 years

Genetics/ Original Article

ISSN 1678-3921

Journal homepage: www.embrapa.br/pab

For manuscript submission and journal contents, access: www.scielo.br/pab

Marker selection using posterior inclusion probability in genomic prediction models for rice data

Abstract – The objective of this work was to evaluate the performance of the combination of the BayesDπ model for marker selection based on posterior inclusion probability (PIP) and of BayesA in assessing the predictive ability, heritability, and predictive bias for a set of rice phenotypic traits. The Markov chain Monte Carlo algorithm was used for the data analysis. For the calculation of PIP, marker effects were estimated using the BayesDπ method. Subsequently, the ratio between the number of iterations in which each marker had a non-zero effect and the total number of iterations was calculated. The markers were allocated into groups of 2,000, 4,000, 6,000, ..., and 36,901 (entire data set), in descending order of importance. The BayesA method was used to re-estimate the effect of the markers in each group. For comparison purposes, marker effects were also calculated using the BayesA and BayesDπ methods separately. In the proposed model, the PIP proved to be effective in understanding genetic architecture, resulting in a higher predictive ability, as well as in a higher heritability and a lower bias in the selection of the most important markers for genomic prediction compared with the other methods without prior marker selection.

Index terms: *Oryza sativa*, Bayesian inference, genetic breeding, genomic selection.

Seleção de marcadores com uso de probabilidade a posteriori de inclusão em modelos para predição genômica em dados de arroz

Resumo – O objetivo deste trabalho foi avaliar o desempenho da combinação do modelo BayesDπ para seleção de marcadores com base na probabilidade de inclusão a posteriori (PIP) e do BayesA na avaliação da capacidade preditiva, da herdabilidade e do viés preditivo de um conjunto de características fenotípicas de arroz. O algoritmo *Markov chain Monte Carlo* foi utilizado para a análise de dados. Para o cálculo da PIP, os efeitos dos marcadores foram estimados com uso do método BayesDπ. Posteriormente, foi calculada a razão entre o número de iterações em que cada marcador apresentou efeito diferente de zero e o número total de iterações. Os marcadores foram distribuídos em grupos de 2.000, 4.000, 6.000, ..., e 36.901 (conjunto de dados inteiro), em ordem decrescente de importância. Utilizou-se o método BayesA para reestimar o efeito dos marcadores em cada grupo. Para fins de comparação, os efeitos dos marcadores também foram calculados com uso dos métodos BayesA e BayesDπ separadamente. No modelo proposto, a PIP se mostrou eficaz na compreensão da arquitetura genética, tendo resultado em maior capacidade preditiva, bem como em maior herdabilidade e menor viés na seleção dos marcadores mais importantes para predição genômica, em comparação a outros métodos sem seleção prévia de marcadores.

Termos para indexação: *Oryza sativa*, inferência bayesiana, melhoramento genético, seleção genômica.

Vinicius Silva Begnami
Universidade Federal de Viçosa, Departamento de Estatística, Viçosa, MG, Brazil.
E-mail: vinicius.begnami@ufv.br

Camila Ferreira Azevedo✉
Universidade Federal de Viçosa, Departamento de Estatística, Viçosa, MG, Brazil.
E-mail: camila.azevedo@ufv.br

Moysés Nascimento
Universidade Federal de Viçosa, Departamento de Estatística, Viçosa, MG, Brazil.
E-mail: moysesnascim@ufv.br

Ana Carolina Campana Nascimento
Universidade Federal de Viçosa, Departamento de Estatística, Viçosa, MG, Brazil.
E-mail: ana.campana@ufv.br

Leísa Pires Lima
Instituto Federal de Educação, Ciência e Tecnologia do Sudeste de Minas Gerais, Campus Muriaé, Muriaé, MG, Brazil.
E-mail: leisapires@hotmail.com

Laís Mayara Azevedo Barroso
Instituto Federal de Educação, Ciência e Tecnologia de Mato Grosso, Campus Sorriso, Sorriso, MT, Brazil.
E-mail: lais.barroso@ifmt.edu.br

✉ Corresponding author

Received
May 8, 2024

Accepted
February 6, 2025

How to cite
BEGNAMI, V.S.; AZEVEDO, C.F..; NASCIMENTO, M.; NASCIMENTO, A.C.C.; LIMA, L.P.; BARROSO, L.M.A. Marker selection using posterior inclusion probability in genomic prediction models for rice data. **Pesquisa Agropecuária Brasileira**, v.60, e03762, 2025. DOI: https://doi.org/10.1590/S1678-3921.pab2025.v60.03762.

This is an open-access article distributed under the Creative Commons Attribution 4.0 International License

Introduction

Rice (*Oryza sativa* L.) is one of the most produced and consumed cereals worldwide, serving as the primary food source for over half of the world's population (Mohidem et al., 2022). Due to its importance, studies aiming at its genetic breeding are necessary, considering the possible enhancement of both the productivity and quality of the crop (Silva Junior et al., 2022).

The breeding process may be accelerated by the application of genome wide selection, which was developed by Meuwissen et al. (2001). This technique uses information from molecular markers to predict the genomic estimated breeding values of individuals, but requires linkage disequilibrium between the marker and quantitative trait loci (QTL) according to Xu et al. (2021).

Various methods for genomic prediction have been proposed over the years, including Bayesian methods such as BayesA, proposed by Meuwissen et al. (2001), and BayesDπ, by Habier et al. (2011). BayesA computes the effects of all markers, considering that each one can explain a specific proportion of genetic variance, whereas BayesDπ selects a random quantity of markers based on a mixture distribution. By using the mixture distribution, the selected markers will have non-zero effects for model adjustment at each interaction. This allows of calculating the posterior inclusion probability (PIP) of each marker in the model, which enables the selection of the most relevant markers, contributing to genetic variance and the predictive ability of the prediction model.

Other authors, such as Dai et al. (2022) and Wang et al. (2020), have also addressed the term PIP, but in the contexts of genomic association or using other Bayesian prediction methods, such as Bayesian variable selection regression and stepwise Bayesian iterative selection. In other studies, Bishara & Hittner (2017) employed PIP to compare the accuracy of various confidence intervals for non-normal data, while Mollandin et al. (2021) used PIP to examine the effectiveness of QTL mapping in genomic data simulated from real data. In the prediction context, breeding programs targeting specific traits could consider PIP to customize marker chips for genotyping individuals, as performed by Liu et al. (2021).

In genomic selection analyses across various crops, prediction ability may be improved by the use of marker subsets (Sousa et al., 2019; Liu et al., 2021; Marquez et al., 2024). In this line, low-density marker panels have been investigated over the years to reduce genotyping costs, dimensional complexity, and computational demands, especially for public institutions and small breeding programs aiming to implement genomic selection. According to the obtained results, in the short term, low-density panels achieved a higher prediction accuracy, whereas, in the long term, high-density panels provided greater genetic gains throughout the breeding cycles. For a sustainable use of marker subsets, prediction models should be retrained more frequently using small datasets (with fewer individuals genotyped at a high density) in order to enhance the ability of the marker set to capture linkage disequilibrium over time.

The objective of this work was to evaluate the performance of the combination of the BayesDπ model for marker selection based on PIP and of BayesA in assessing the predictive ability, heritability, and predictive bias for a set of rice phenotypic traits.

Materials and Methods

The public dataset used in this study consists of 11 traits related to 413 rice genotypes, which were genotyped for 44,100 single nucleotide polymorphism (SNP) markers. The data, part of the OryzaSNP Project and the OMAP Project (Ammiraju et al., 2006; Zhao et al., 2011), are publicly available and can be accessed at the Rice Diversity Project website (2025). Markers with a call rate below 70% and a minor allele frequency below 1% were excluded. After this filtering process, 36,901 markers remained. The 11 traits used were: blast resistance, flag leaf length, flag leaf width, florets per panicle, number of panicles per plant, number of primary panicle branches, number of seeds per panicle, panicle fertility, panicle length, plant height, and protein content.

The Bayesian methods proposed for genomic prediction are based on the statistical model of multiple linear regression with k independent variables, expressed as follows:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_k X_{ik} + \epsilon_i,$$
$$i = 1, \cdots, n \text{ and } j = 1, ..., k$$

where Y_i represents the phenotypic value for the ith individual, β_0 is the overall mean, β_j is the effect of the jth

marker on the phenotype, X_{ij} represents the incidence of the jth marker for the ith individual, ϵ_i is the random value assigned to the ith individual following a normal distribution N(0, σ^2), n is the number of phenotypic observations, and k is the number of markers. In this model, the data distribution and the prior distribution of the residual variance component are defined as follows:

$$y_i \sim N(\beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_k X_{ik}, \sigma^2) \text{ and } \sigma^2 \sim IG(a, b)$$

where N is the normal distribution, σ^2 is the residual variance, and IG is the inverse gamma distribution for which and are hyperparameters, defined based on Azevedo et al. (2022a).

The Bayesian methods applied to genome wide selection use different prior distributions for regression coefficients. In the case of BayesA, proposed by Meuwissen et al. (2001), the prior distributions can be defined as:

$$\beta_j \sim N(0, \sigma^2_{\beta j}) \text{ and } \sigma^2_{\beta j} \sim IG(a_j, b_j)$$

where β_j is the effect of the jth marker on the phenotype, $\sigma^2_{\beta j}$ is the genetic variance associated with the effect of the jth marker, and a_j and b_j are the hyperparameters of the inverse gamma distribution, also defined based on Azevedo et al. (2022a). That method induces heterogeneous shrinkage on marker effects, causing their values only to approach but never to become exactly zero (Meuwissen et al., 2001). It also includes the effects (whether small or large) of all markers in the model and even markers not in linkage disequilibrium with any QTL, which are the majority for a specific trait (Simeão et al., 2021).

The BayesDπ method, at each iteration, assumes that a fraction of markers have effects following a normal distribution with mean zero and specific variance, similar to the BayesA method. However, the remaining markers 1-π have their effects nullified (Habier et al., 2011), resulting in a reduction in the number of marker effects to be estimated in the iterations, which may increase the precision in model fitting (Marquez et al., 2024). For the fraction π of markers, a uniform distribution with hyperparameters 0 and 1 is assumed. The prior distributions are defined as follows:

$$\pi \sim (0,1), \beta_j \sim \pi N(0, \sigma^2_{\beta j}) + (1-\pi) N(0,0) \text{ and}$$

$$\sigma^2_{\beta j} \sim IG(a_j, b_j)$$

where β_j is the effect of the jth marker on the phenotype, $\sigma^2_{\beta j}$ is the genetic variance associated with the effect of the jth marker, is a degenerate distribution centered at zero, and a_j and b_j are the hyperparameters of the inverse gamma distribution.

A total of 300,000 iterations were used in the Markov chain Monte Carlo algorithm, but 50,000 were discarded (burn-in) for chain warm-up. To decrease the correlations between iterations, thinning was adopted by selecting one observation for every ten iterations. Geweke (1992) and Raftery & Lewis (1992) were used for the convergence analysis.

At the conclusion of the BayesDπ method, using the marker-effect chains, the PIP of each marker was estimated as the ratio between the number of iterations in which the marker had a non-zero effect and the total number of iterations. After calculating all PIPs, groups of SNP markers with the highest probabilities were selected for evaluation using the BayesA method. These groups were formed by N markers, ranging from 2,000 to 36,901, with an increase of 2,000 markers each time.

The BayesA, BayesDπ, and the combination of the BayesA and BayesDπ methods with PIP selection criteria (proposed method) were performed under a ten-fold cross-validation procedure to evaluate the following metrics: predictive ability, which is Pearson's correlation between the phenotypic and predicted values of the three methods; prediction bias, evaluated by the regression coefficient between the phenotypic and predicted values, considered not biased or underestimated/overestimated when the coefficient is 1 or higher/lower than 1, respectively; and the 95% confidence interval of the regression coefficients, to determine if their values are statistically different from 1.

The heritability of the trait estimated for each method was also calculated as the posterior mean of the following values: $\frac{\sigma_a^{2(t)}}{\sigma_a^{2(t)} + \sigma_e^{2(t)}}$, where $h^{2(t)}$ is the heritability in the tth iteration of the Markov chain Monte Carlo; $\sigma_a^{2(t)}$ is the additive genetic variance in the tth iteration, $\sigma_a^{2(t)} = \sum_{j=1}^{k} 2p_j(1-p_j)\beta_j^{2(t)}$, where p_j is the allelic frequency of the jth marker and $\beta_j^{(t)}$ is the effect of the jth marker; and $\sigma_e^{\{2(t)\}}$ is the residual variance in the tth iteration.

Results and Discussion

Markers with a lower inclusion were more frequent than those with a higher inclusion, indicating a right-skewed distribution for PIP. Additionally, more markers had effects close to zero, which may suggest a more oligogenic nature of the PIP distribution (Figure 1).

Considering markers with PIP values higher than 0.5 as relevant for the trait, resistance to blast disease exhibited a greater variation in PIP values and fewer markers with a high PIP, also displaying a right-skewed distribution. Moreover, there was a large number of markers with an effect close to zero and fewer with negative effects, further supporting a right-skewed distribution and suggesting that this trait may be oligogenic in nature. Although this finding is consistent with those of Li et al. (2019), further investigations are needed.

The BayesDπ and BayesA methods presented similar metrics, while the proposed method in which both were combined showed improvements in terms of heritability, predictive ability, and predictive bias (Figure 2). The traits flag leaf length, number of seeds

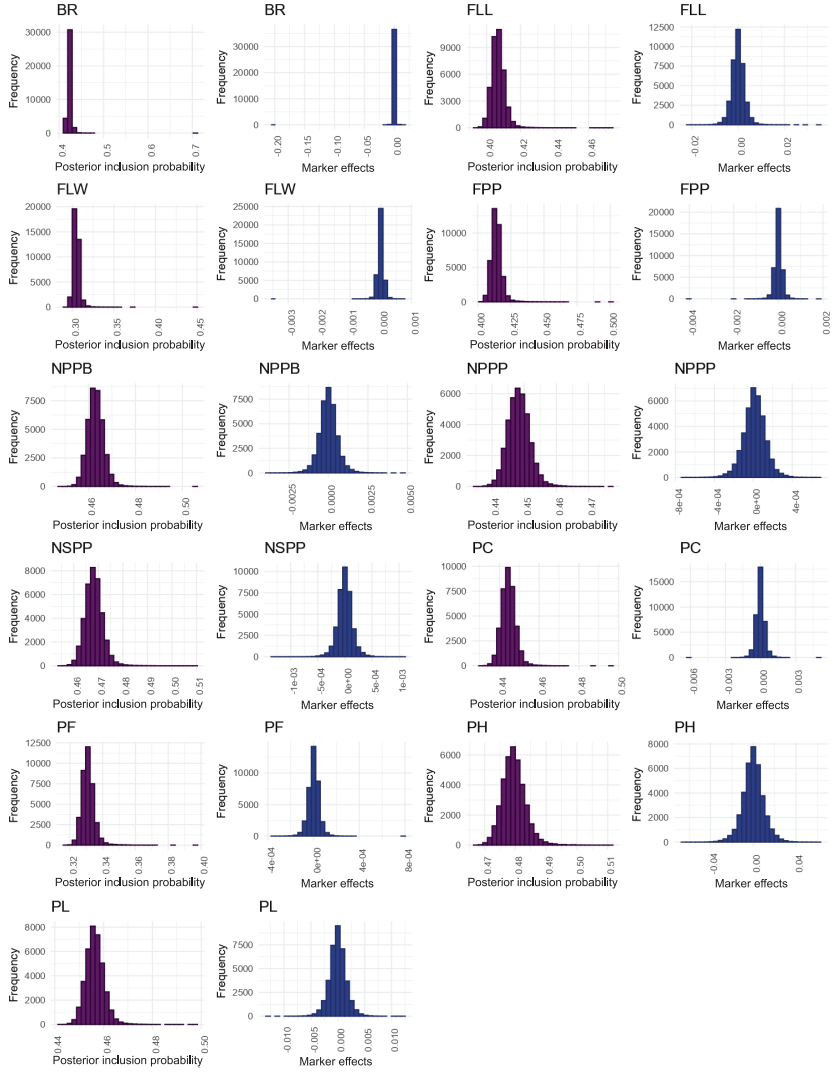

Figure 1. Histograms showing posterior inclusion probability (in purple) and effects (in blue) of molecular markers for each of the following rice (*Oryza sativa*) traits: blast resistance (BR), flag leaf length (FLL), flag leaf width (FLW), florets per panicle (FPP), number of primary panicle branches (NPPB), number of panicles per plant (NPPP), number of seeds per panicle (NSPP), protein content (PC), panicle fertility (PF), plant height (PH), and panicle length (PL).

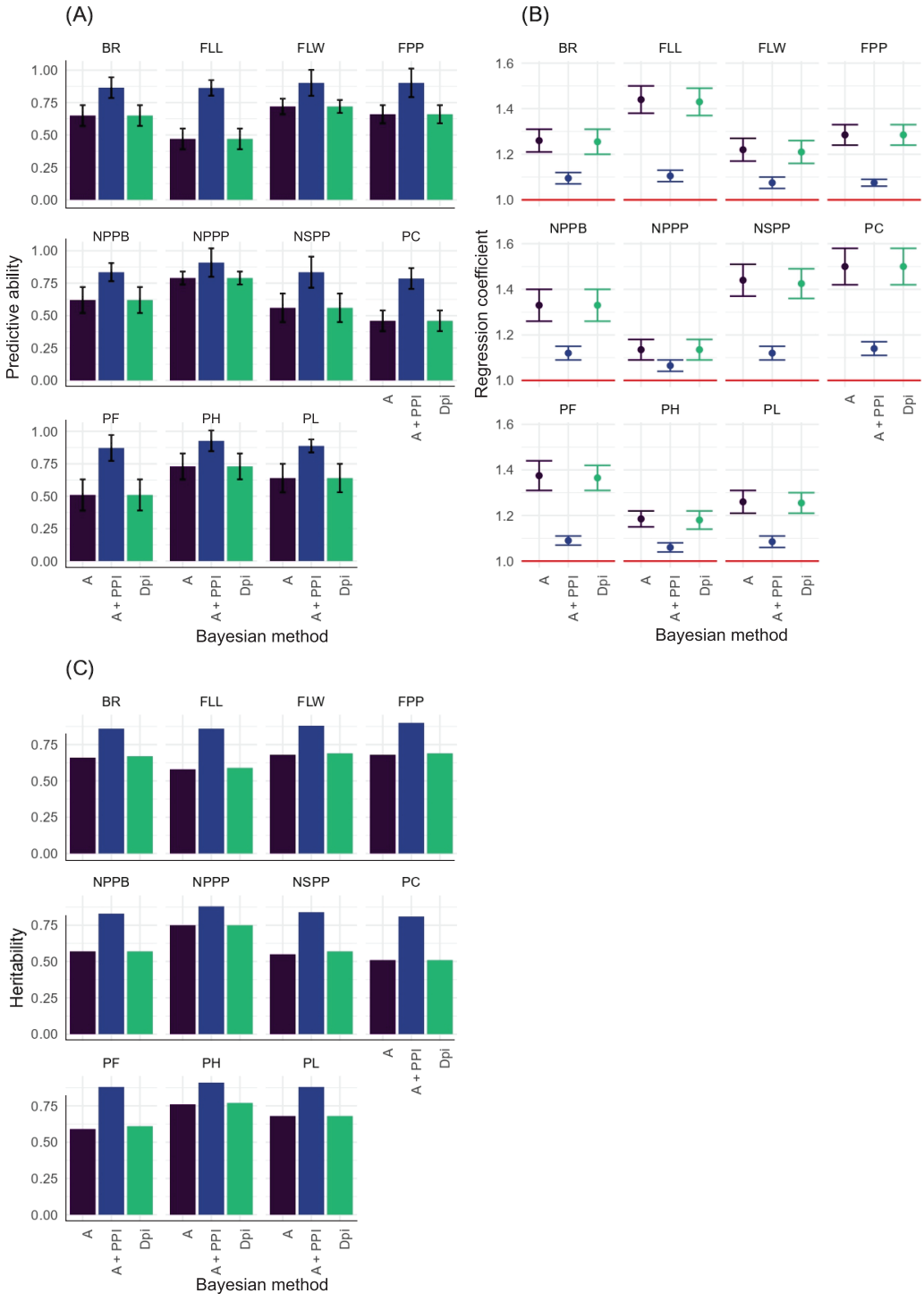

Figure 2. Graphs showing: mean predictive ability and standard deviation bars (A), confidence interval of the regression coefficient between predicted and observed phenotypic values (B), and posterior mean of heritability of the studied rice (*Oryza sativa*) traits (C). BR, blast resistance; FLL, flag leaf length; FLW, flag leaf width; FPP, florets per panicle; NPPB, number of primary panicle branches; NPPP, number of panicles per plant; NSP, number of seeds per panicle; PC, protein content; PF, panicle fertility; PH, plant height; and PL, panicle length. The horizontal red line represents the regression coefficient equal to 1. A, BayesA method; A+ PIP, the combination of the BayesA and BayesDπ methods with PIP selection criteria; and Dpi, BayesDπ method.

per panicle, number of primary panicle branches, panicle fertility, and protein content exhibited a higher predictive ability (Figure 2 A), a lower prediction bias (Figure 2 B), and a higher heritability (Figure 2 C) than those tested in each isolated method. These results are in agreement with those reported by Huang et al. (2019), who evaluated the accuracy of seven models for marker resistance to blast disease. He et al. (2022) concluded that, since there are currently no comprehensive control measures, cultivating resistant varieties remains the most effective strategy to combat the disease, a major threat to crop production, with potential to compromise up to 50% of total yield.

The traits number of panicles per plant, number of primary panicle branches, number of seeds per panicle, plant height, and panicle length showed less variation in PIP and a Gaussian distribution for marker effects (Figure 1), which suggests they may be more polygenic in nature. This observation is in alignment with association studies such as that of Suela et al. (2022). Among the analyzed traits, protein content, number of seeds per panicle, and number of primary panicle branches exhibited the highest gains in predictive ability (0.23, 0.21, and 0.18, respectively) and heritability (0.15, 0.16, and 0.19, respectively), as indicated in Table 1. The use of PIP appears particularly relevant for number of seeds per panicle and protein content, which stood out among the top three traits with the most significant improvements.

In the BayesA prediction model, marker groups ranged from 2,000 to 36,000, increasing in increments of 2,000 (Figure 3). The markers were selected based on the highest PIP in the model, along with the posterior mean heritability of the rice traits and prediction bias. In this case, the selection process is expected to be more effective since adjusting a prediction model with a smaller number of markers increases its predictive ability (Sousa et al., 2019; Liu et al., 2021). In fact, there was an increase in predictive ability and heritability

Table 1. Descriptive summary of predictive ability (r) and heritability (h^2) across groups of markers selected for rice (*Oryza sativa*) traits, ranging from 2,000 to 36,000, increasing in increments of 2,000.

Trait[(1)]	Group	Predictive ability (r)				Heritability (h^2)			
		Min	Mean	Max	SD	Min	Mean	Max	SD
BR	All	0.67	0.76	0.76	0.06	0.67	0.77	0.77	0.06
FLL	All	0.50	0.69	0.69	0.12	0.61	0.75	0.75	0.08
FLW	All	0.73	0.81	0.81	0.05	0.69	0.79	0.79	0.06
FPP	All	0.66	0.79	0.79	0.08	0.69	0.82	0.82	0.07
NPPB	All	0.62	0.72	0.72	0.07	0.58	0.71	0.71	0.08
NPPP	All	0.78	0.84	0.84	0.04	0.75	0.81	0.81	0.04
NSPP	All	0.57	0.70	0.70	0.09	0.57	0.72	0.72	0.09
PC	All	0.46	0.61	0.61	0.09	0.52	0.68	0.68	0.09
PF	All	0.52	0.70	0.70	0.11	0.61	0.77	0.77	0.08
PH	All	0.73	0.83	0.83	0.06	0.78	0.85	0.85	0.04
PL	All	0.63	0.77	0.77	0.07	0.69	0.79	0.79	0.06
BR	2,000		0.93				0.91		
FLL	2,000		0.86				0.86		
FLW	2,000		0.89				0.88		
FPP	2,000		0.87				0.88		
NPPB	2,000		0.90				0.90		
NPPP	2,000		0.90				0.88		
NSPP	2,000		0.91				0.88		
PC	2,000		0.84				0.83		
PF	2,000		0.84				0.84		
PH	2,000		0.87				0.86		
PL	2,000		0.79				0.81		

[(1)]BR, blast resistance; FLL, flag leaf length; FLW, flag leaf width; FPP, florets per panicle; NPPB, number of primary panicle branches; NPPP, number of panicles per plant; NSPP, number of seeds per panicle; PC, protein content; PF, panicle fertility; PH, plant height; and PL, panicle length.

when using the 2,000 most important markers for each trait (Table 1).

Across all groups of markers and traits, predictive values ranged from 0.46 to 0.84 and heritabilities from 0.52 to 0.85 (Table 1). For the group of the 2,000 most important markers, predictive ability and heritability were high, ranging from 0.84 to 0.93 and from 0.83 to 0.91, respectively; the exception was panicle length, which showed a predictive ability of 0.79 and a heritability of 0.81.

As the number of less important markers increased, predictive ability and heritability gradually decreased. The decrease in the predictive ability of the model and in average heritability was the greatest for the traits flag leaf length, panicle fertility, and protein content, but the smallest for flag leaf width, number of panicles per plant, and plant height. A decrease in predictive ability was also observed by Sousa et al. (2019) for a few rice traits when using the mixed models methods and by Marquez et al. (2024) for sweet corn (*Zea*

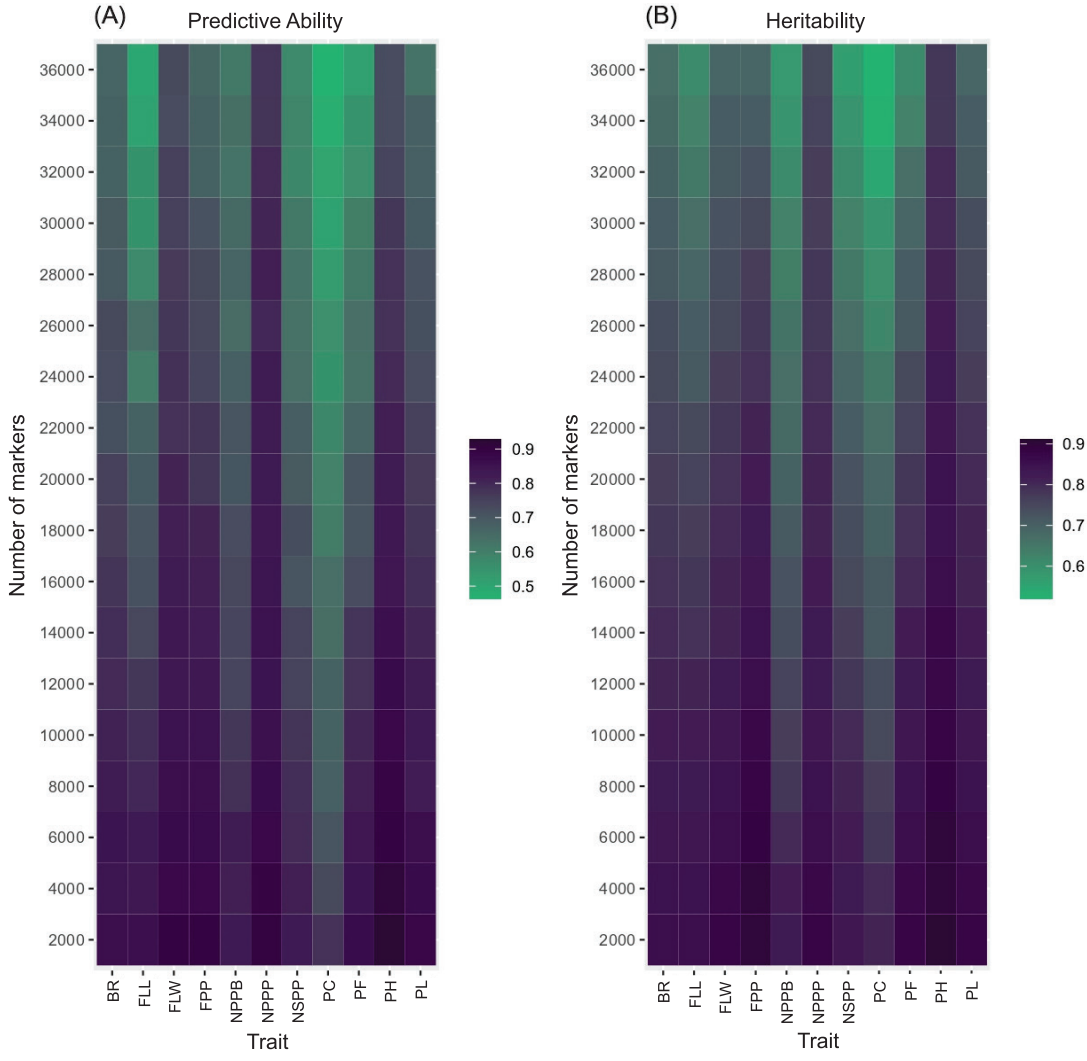

Figure 3. Graphs showing: the predictive ability of the BayesA prediction model using groups of markers ranging from 2,000 to 36,000, increasing by increments of 2,000, selected based on highest posterior inclusion probabilities in the model (A); and posterior mean of heritability of the studied rice (*Oryza sativa*) traits (B). BR, blast resistance; FLL, flag leaf length; FLW, flag leaf width; FPP, florets per panicle; NPPB, number of primary panicle branches; NPPP, number of panicles per plant; NSPP, number of seeds per panicle; PC, protein content; PF, panicle fertility; PH, plant height; and PL, panicle length.

mays L.). Wray et al. (2013) noted that, as irrelevant or redundant markers with a small effect are included to predict the trait of interest, the theoretical predictive ability decreases.

According to Azevedo et al. (2022b), assuming some marker effects equal to zero leads to a more favorable statistical condition, meaning that the lower the number of marker effects to be estimated, the more accurate the estimation process considering the same number of observations. In the present study, as the number of markers was reduced, the metrics increased gradually (Figure 3). In addition, the use of a smaller number of more important markers (2,000 for each trait under study) resulted in a better predictive ability and in higher heritability estimates (Table 1). Therefore, focusing on more relevant markers possibly allows for better predictions, which aligns with the concept discussed by Azevedo et al. (2022b).

In relation to heritabilities, high estimates are more stable and accurate when using a reduced set of more relevant markers that better capture genetic variation. De los Campos et al. (2015) concluded that, if all causal variants are in the marker subset, there should be no missing heritability. In this line, Porto-Neto et al. (2015) observed that the missing heritability can be largely recovered using an accurately selected subset of markers, with different marker allele frequencies between individuals. Sousa et al. (2019) added that the increase in the number of SNPs can cause changes in genetic and residual variances and, consequently, in heritability estimates.

The BayesDπ and BayesA methods showed similar heritabilities, whereas the proposed method consistently showed a higher estimate (Figure 2 C). According to Jilo et al. (2018), heritability values ranging from 0.30 to 0.60 are considered of moderate magnitude, while those above 0.60 are of high magnitude. Therefore, in the present study, the heritabilities ranged from moderate to high magnitude, varying from 0.51 to 0.76 for the BayesA method, from 0.51 to 0.77 for the BayesDπ method, and from 0.77 to 0.80 for the combination of both methods. This shows the superior performance of the proposed method in capturing the genetic variance of the studied traits.

The estimated heritabilities were also comparable to those found by Guo et al. (2014), who used the G-BLUP method to adjust the prediction model for the same dataset. Compared with the results of Wang et al. (2017), who applied G-BLUP to rice data from Wuhan University, a higher heritability was obtained for plant height in the present work, but a slightly higher predictive ability and slightly lower heritability for panicle length. This balance between predictive ability and heritability highlights the robustness of the proposed method in adapting to different traits while maintaining a high precision.

Regarding prediction bias, none of the traits presented a confidence interval for the regression coefficient between predicted and observed phenotypic values that included 1 (Figure 2 B). It should be noted that the forecasts were systematically underestimated since all intervals had a lower bound greater than 1. Similarly, Sousa et al. (2019) also reported bias in the predictions across different marker subsets. However, the method proposed here led to less biased predictions for most of the evaluated traits, with smaller confidence intervals for the regression coefficients, providing more precise estimates of the genetic breeding values, which is of extreme importance for selection efficiency (Bishara & Hittner, 2017).

For predicted genetic values, Costa et al. (2022) observed both over- and underestimation when using the same dataset to compare the efficiency of some prediction methods, including Bayesian Lasso, methods based on mixed models, and dimensionality reduction techniques. Compared with these methods, the approach proposed in the present work not only minimized bias but also provided more consistent and reliable estimates across traits.

Conclusions

1. The combination of the BayesDπ and BayesA methods proposed for genomic prediction shows a higher predictive ability than the isolated methods without prior marker selection, as well as a lower prediction bias in the selection of the most important markers based on posterior inclusion probability (PIP).

2. The use of PIP to select more important markers contributes to gains in heritability, enabling the capture of a greater proportion of genetic variance.

3. Dimensionality reduction results in a lower prediction bias, enhancing the reliability of the obtained estimates.

Acknowledgements

To Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), for financing, in part, this study (Finance Code 001); and to Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), for research productivity fellowships.

References

AMMIRAJU, J.S.S.; LUO, M.; GOICOECHEA, J.L.; WANG, W.; KUDRNA, D.; MUELLER, C.; TALAG, J.; KIM, H.; SISNEROS, N.B.; BLACKMON, B.; FANG, E.; TOMKINS, J.B.; BRAR, D.; MACKILL, D.; MCCOUCH, S.; KURATA, N.; LAMBERT, G.; GALBRAITH, D.W.; ARUMUGANATHAN, K.; RAO, K.; WALLING, J.G.; GILL, N.; YU, Y.; SANMIGUEL, P.; SODERLUND, C.; JACKSON, S.; WING, R.A. 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, p.140-147, 2006. DOI: https://doi.org/10.1101/gr.3766306.

AZEVEDO, C.F.; LIMA, L.P.; NASCIMENTO, M.; NASCIMENTO, A.C.C. Bayesian methods for genomic association of chromosomic regions considering the additive-dominance model. **Crop Breeding and Applied Biotechnology**, v.22, e418422310, 2022a. DOI: https://doi.org/10.1590/1984-70332022v22n3a33.

AZEVEDO, C.F.; RESENDE, M.D.V.; SILVA, F.F.; NASCIMENTO, M. Métodos estatísticos aplicados à seleção genômica. In: PEIXOTO, L. de A.; BHERING, L.L.; CRUZ, C.D. (Ed.). **Seleção genômica aplicada ao melhoramento genético**. Viçosa: Ed. da UFV, 2022b.

BISHARA, A.J.; HITTNER, J.B. Confidence intervals for correlations when data are not normal. **Behavior Research Methods**, v.49, p.294-309, 2017. DOI: https://doi.org/10.3758/s13428-016-0702-8.

COSTA, J.A. da; AZEVEDO, C.F.; NASCIMENTO, M.; SILVA, F.F. e; RESENDE, M.D.V. de; NASCIMENTO, A.C.C. Determination of optimal number of independent components in yield traits in rice. **Scientia Agricola**, v.79, e20200397, 2022. DOI: https://doi.org/10.1590/1678-992X-2020-0397.

DAI, X.; FU, G.; REESE, R.; ZHAO, S.; SHANG, Z. An approach of Bayesian variable selection for ultrahigh-dimensional multivariate regression. **Stat**, v.11, e476, 2022. DOI: https://doi.org/10.1002/sta4.476.

DE LOS CAMPOS, G.; SORENSEN, D.; GIANOLA, D. Genomic heritability: what is it? **PLoS Genetics**, v.11, e1005048, 2015. DOI: https://doi.org/10.1371/journal.pgen.1005048.

GEWEKE, J. Evaluating the accuracy of sampling-based approaches to the calculation of posterori moments. In: BERNARDO, J.M.; BERGER, J.O.; DAWID, A.P.; SMITH, A.F.M. (Ed.). **Bayesian statistics 4**: Proceeding of the Fourth Valencia International Meeting, Dedicated to the memory of Morris H. Degroot, 1931-1989. Oxford: Oxford University Press, 1992. p.169-194. DOI: https://doi.org/10.1093/oso/9780198522669.003.0010.

GUO, Z.; TUCKER, D.M.; BASTEN, C.J.; GANDHI, H.; ERSOZ, E.; GUO, B.; XU, Z.; WANG, D.; GAY, G. The impact of population structure on genomic prediction in stratified populations. **Theoretical and Applied Genetics**, v.127, p.749-762, 2014. DOI: https://doi.org/10.1007/s00122-013-2255-x.

HABIER, D.; FERNANDO, R.L.; KIZILKAYA, K.; GARRICK, D.J. Extension of the Bayesian alphabet for genomic selection. **BMC Bioinformatics**, v.12, art.186, 2011. DOI: https://doi.org/10.1186/1471-2105-12-186.

HE, Y.; ZHAO, X.; ZHANG, W.; HE, X.; TONG, L. Study on the identification of resistance of rice blast based on near infrared spectroscopy. **Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy**, v.266, art.120439, 2022. DOI: https://doi.org/10.1016/j.saa.2021.120439.

HUANG, M.; BALIMPONYA, E.G.; MGONJA, E.M.; MCHALE, L.K.; LUZI-KIHUPI, A.; WANG, G.-L.; SNELLER, C.H. Use of genomic selection in breeding rice (*Oryza sativa* L.) for resistance to rice blast (*Magnaporthe oryzae*). **Molecular Breeding**, v.39, art.114, 2019. DOI: https://doi.org/10.1007/s11032-019-1023-2.

JILO, T.; TULU, L.; BIRHAN, T.; BEKSISA, L. Genetic variability, heritability and genetic advance of maize (*Zea mays* L.) inbred lines for yield and yield related traits in southwestern Ethiopia. **Journal of Plant Breeding and Crop Science**, v.10, p.281-289, 2018. DOI: https://doi.org/10.5897/JPBCS2018.0742.

LI, C.; WANG, D.; PENG, S.; CHEN, Y.; SU, P.; CHEN, J.; ZHENG, L.; TAN, X.; LIU, J.; XIAO, Y.; KANG, H.; ZHANG, D.; WANG, G.-L.; LIU, Y. Genome-wide association mapping of resistance against rice blast strains in South China and identification of a new *Pik* allele. **Rice**, v.12, art.47, 2019. DOI: https://doi.org/10.1186/s12284-019-0309-7.

LIU, Z.; SUN, C.; YAN, Y.; LI, G.; LI, X.C.; WU, G.; YANG, N. Design and evaluation of a custom 50K Infinium SNP array for egg-type chickens. **Poultry Science**, v.100, art.101044, 2021. DOI: https://doi.org/10.1016/j.psj.2021.101044.

MARQUEZ, G.R.; ZHANG-BIEHN, S.; GUO, Z.; MORO, G.V. Effects of marker density on genomic prediction for yield traits in sweet corn. **Euphytica**, v.220, art.52, 2024. DOI: https://doi.org/10.1007/s10681-024-03313-6.

MEUWISSEN, T.H.E.; HAYES, B.J.; GODDARD, M.E. Prediction of total genetic value using genome-wide dense marker maps. **Genetics**, v.157, p.1819-1829, 2001. DOI: https://doi.org/10.1093/genetics/157.4.1819.

MOHIDEM, N.A.; HASHIM, N.; SHAMSUDIN, R.; CHE MAN, H. Rice for food security: revisiting its production, diversity, rice milling process and nutrient content. **Agriculture**, v.12, art.741, 2022. DOI: https://doi.org/10.3390/agriculture12060741.

MOLLANDIN, F.; RAU, A.; CROISEAU, P. An evaluation of the predictive performance and mapping power of the BayesR model for genomic prediction. **G3 Genes, Genomes, Genetics**, v.11, jkab225, 2021. DOI: https://doi.org/10.1093/g3journal/jkab225.

PORTO-NETO, L.R.; BARENDSE, W.; HENSHALL, J.M.; MCWILLIAM, S.M.; LEHNERT, S.A.; REVERTER, A.

Genomic correlation: harnessing the benefit of combining two unrelated populations for genomic selection. **Genetics Selection Evolution**, v.47, art.84, 2015. DOI: https://doi.org/10.1186/s12711-015-0162-0.

RAFTERY, A.L.; LEWIS, S.M. Comment: one long run with diagnostics: implementation strategies for Markov chain Monte Carlo. **Statistical Science**, v.7, p.493-497, 1992. DOI: https://doi.org/10.1214/ss/1177011143.

RICE DIVERSITY PROJECT. Available at: <https://ricediversity.org/data/>. Accessed on: Jan. 29 2025.

SILVA JUNIOR, A.C. da; SANT'ANNA, I. de C.; PEIXOTO, M.A.; TORRES, L.G.; SIQUEIRA, M.J.S.; COSTA, W.G. da; AZEVEDO, C.F.; SOARES, P.C.; CRUZ, C.D. Multiple-trait model through Bayesian inference applied to flood-irrigated rice (*Oryza sativa* L). **Euphytica**, v.218, art.124, 2022. DOI: https://doi.org/10.1007/s10681-022-03077-x.

SIMEÃO, R.M.; RESENDE, M.D.V.; ALVES, R.S.; PESSOA-FILHO, M.; AZEVEDO, A.L.S.; JONES, C.S.; PEREIRA, J.F.; MACHADO, J.C. Genomic selection in tropical forage grasses: current status and future applications. **Frontiers in Plant Science**, v.12, art.665195, 2021. DOI: https://doi.org/10.3389/fpls.2021.665195.

SOUSA, M.B. e; GALLI, G.; LYRA, D.H.; GRANATO, I.S.C.; MATIAS, F.I.; ALVES, F.C.; FRITSCHE-NETO, R. Increasing accuracy and reducing costs of genomic prediction by marker selection. **Euphytica**, v.215, art.18, 2019. DOI: https://doi.org/10.1007/s10681-019-2339-z.

SUELA, M.M.; AZEVEDO, C.F.; NASCIMENTO, M.; NASCIMENTO, A.C.C.; RESENDE, M.D.V. de. Regional heritability mapping and genome-wide association identify loci for rice traits. **Crop Science**, v.62, p.839-858, 2022. DOI: https://doi.org/10.1002/csc2.20706.

WANG, G.; SARKAR, A.; CARBONETTO, P.; STEPHENS, M. A simple new approach to variable selection in regression, with application to genetic fine mapping. **Journal of the Royal Statistical Society Series B: Statistical Methodology**, v.82, p.1273-1300, 2020. DOI: https://doi.org/10.1111/rssb.12388.

WANG, X.; LI, L.; YANG, Z.; ZHENG, X.; YU, S.; CHU, C.; HU, Z. Predicting rice hybrid performance using univariate and multivariate GBLUP models based on North Carolina mating design II. **Heredity**, v.118, p.302-310, 2017. DOI: https://doi.org/10.1038/hdy.2016.87.

WRAY, N.R.; YANG, J.; HAYES, B.J.; PRICE, A.L.; GODDARD, M.E.; VISSCHER, P.M. Pitfalls of predicting complex traits from SNPs. **Nature Reviews Genetics**, v.14, p.507-515, 2013. DOI: https://doi.org/10.1038/nrg3457.

XU, Y.; MA, K.; ZHAO, Y.; WANG, X.; ZHOU, K.; YU, G.; LI, C.; LI, P.; YANG, Z.; XU, C.; XU, S. Genomic selection: a breakthrough technology in rice breeding. **The Crop Journal**, v.9, p.669-677, 2021. DOI: https://doi.org/10.1016/j.cj.2021.03.008.

ZHAO, K.; TUNG, C.-W.; EIZENGA, G.C.; WRIGHT, M.H.; ALI, M.L.; PRICE, A.H.; NORTON, G.J.; ISLAM, M.R.; REYNOLDS, A.; MEZEY, J.; MCCLUNG, A.M.; BUSTAMANTE, C.D.; MCCOUCH, S.R. Genome-wide association mapping reveals a rich genetic architecture of complex traits in *Oryza sativa*. **Nature Communications**, v.2, art.467, 2011. DOI: https://doi.org/10.1038/ncomms1467.