

## Multiple-trait model through Bayesian inference applied to flood-irrigated rice (*Oryza sativa* L)

Antônio Carlos da Silva Júnior<sup>(1)</sup>, Isabela de Castro Sant'Anna<sup>(2)</sup>, Marco Antônio Peixoto<sup>(1)</sup>, Livia Gomes Torres<sup>(1)</sup>, Michele Jorge Silva<sup>(1)</sup>, Weverton Gomes da Costa<sup>(1)</sup>, Camila Ferreira Azevedo<sup>(3)</sup>, Plínio César Soares<sup>(4)</sup> and Cosme Damião Cruz<sup>(1)</sup>

<sup>(1)</sup> Departamento de Biologia Geral, Universidade Federal de Viçosa, CEP: 36.571-000, Viçosa, Minas Gerais, Brasil

<sup>(2)</sup> Centro de Seringueira e Sistemas Agroflorestais, Instituto Agrônomo (IAC), Rodovia Péricles Beline, 1.481, Botafogo, Campinas, CEP: 13020-902, São Paulo, Brasil

<sup>(3)</sup> Departamento de Estatística, Universidade Federal de Viçosa, CEP: 36.571-000, Viçosa, Minas Gerais, Brasil

<sup>(4)</sup> Empresa de Pesquisa Agropecuária de Minas Gerais – EPAMIG, Viçosa, Minas Gerais, Brazil.

\*Corresponding author: [antonio.silva.c.junior@gmail.com](mailto:antonio.silva.c.junior@gmail.com)

Tel.: +55 0313885-3256

### Abstract

The giant challenge breeding flood-irrigated rice is to identify superior genotypes that present high-yielding with specific grain qualities, resistance to abiotic and biotic stresses, excellent adaptation to the target environment. Thus, the objectives of this study were to propose a bayesian multi-trait model, estimate genetic parameters, and select flood-irrigated rice genotypes with better genetic potentials in different evaluation environments. For this, twenty-five rice genotypes belonging to the flood-irrigated rice improvement program were evaluated. The grain yields, grain length, width and thickness, grain length, and grain width and weight of 100 grains in the agricultural year 2016/2017. The experimental design used in all experiments was a randomized block design with three replications. The Monte Carlo Markov Chain algorithm estimated genetic parameters and genetic values. The grain thickness trait was considered highly heritable, with a credibility interval ranging from:  $h^2$ : 0.9480; 0.9440; 0.8610, in environments 1, 2, and 3, respectively. The grain yields showed a low correlation estimate

between grain thickness and 100-grain weight, in all environments, with a credibility interval ranging from ( $\rho = 0.5477$ ;  $0.5762$ ;  $0.5618$  and  $0.5973$ ;  $0.5247$ ;  $0.5632$ , grain thickness and 100-grain weight, in environments 1, 2, and 3, respectively). The Bayesian multi-trait model proved to be an adequate strategy for the genetic improvement of flood-irrigated. Genotypes 2 and 15 had similar potential in the three environments, they should be selected as high-performance multi-trait genotypes for the genetic breeding of flood-irrigated rice in the program.

**Keywords:** high-performance, heritable, multi-environments, credibility interval

## Introduction

Rice (*Oryza sativa* L.) is one of the most important crops in the world and is considered one of the main annual crops in Brazil (Silva Júnior et al., 2020). Plant breeding is an effective way to increase crop productivity. Its objective is to recommend the most suitable cultivars for planting, whether from final productivity, plant health, or operational (harvest facilitator). It is necessary to consider the interaction between genotypes and environments (GE) (Silva Júnior et al., 2021). The occurrence of this interaction in multi-trait datasets poses an even more significant challenge for the breeder (Torres et al., 2018; Volpato et al. 2019, Peixoto et al., 2021). Therefore, a statistical methodology must evaluate the information from a network of experiments with a multi-trait structure that correctly represents genetic and phenotypic variation in the data (Malosetti et al., 2008). For that, Bayesian inference has become a useful statistical tool for dealing with complex models (Torres et al., 2018).

Bayesian inference has surpassed traditional analyses by providing different results to the classical approach, such as longer credibility intervals, genetic parameter estimates, and genetic values with greater precision (Peixoto et al., 2021). The Bayesian inference is a flexible methodology that allows the estimation of precise genetic values and variance components, even from small samples (Resende et al., 2014; Jarquín et al., 2016; Peixoto et al., 2021; Schoot et al., 2021).

Bayesian multi-trait models (MTM) have become a proper statistical method for genetic evaluations of plants and animals (Junqueira et al., 2016 Volpato et al., 2019). In addition, this model allows the estimation of variance components and breeding values for each trait (Peixoto et al., 2021), jointly modeling multiple traits compared to the analysis of each trait separately. The inference process adequately explains the correlation

between the traits, which helps to increase prediction accuracy, statistical power, parameter estimation accuracy, and reduce the character selection bias (Henderson and Quaas, 1976; Pollak et al., 1984; Schaeffer, 1984). The common use of multiple traits benefits from the genetic correlation between traits and significantly improves prediction accuracy compared to single-trait methods, specifically for low heritability traits that are genetically correlated with a high heritability trait (Jia and Jannink, 2012; Guo et al., 2014; Jiang et al., 2015; Montesinos-Lopez et al., 2016).

Some studies have shown the potential of the Bayesian approach for genetic evaluation in plant breeding, considering models with multiple traits or multiple environments (Cané-Retamales et al., 2011; Arriagada et al., 2012; Mora et al., 2014; Junqueira et al., 2016; Torres et al., 2018; Volpato et al., 2019). However, few studies combine multi-trait models under a Bayesian view for flood-irrigated rice cultivation. Therefore, the objectives of this study were to propose a Bayesian multi-trait model, estimate genetic parameters, and select flood-irrigated rice genotypes with better genetic potentials (desirable agronomic traits) in different evaluation environments.

## **Material and Methods**

### **Field experiments**

The experiments were carried out in the State of Minas Gerais, Brazil, on the experimental fields of Agricultural Research Institute of Minas Gerais State (EPAMIG) in the cities of Leopoldina (21° 31' 48.01" S, 42° 38' 24" W), Lambari (21° 58' 11.24" S, 45° 20' 59.6" W) and Janaúba (15° 48' 77" S, 43° 17' 59.09" W).

Twenty-five rice (*Oryza sativa* L) genotypes belonging to the flood-irrigated rice breeding program of the state of Minas Gerais were evaluated, and five of these genotypes were used as experimental controls (Rubelita, Seleta, Ourominas, Predileta, and Rio Grande). These genotypes were evaluated in comparative trials after multiple generations of selection, and in addition, they are known for their high yield, uniform growth rate and plant growth, resistance to major diseases, and for their excellent grain quality. The traits evaluated were grain yields (GY, Kg ha<sup>-1</sup>), grain length (GL, mm), grain width (GW, mm) and grain thickness (GT, mm), grain length, and grain width and weight of 100 grains (GWH, g) in the agricultural year 2016/2017.

The design used in all experiments was a randomized complete block design with three replications. The experiments were conducted in floodplain soils with continuous

flood irrigation. Management practices were carried out according to recommendations for flood-irrigated rice in the relevant regions (Soares et al., 2005).

The useful plot area consisted of 4 m central of three internal rows (4 m x 0.9 m, 3.60 m<sup>2</sup> total). The soil preparation was carried out by plowing and harrowing around 30 days before sowing and harrowing on the eve of the installation of the tests. For planting fertilization, a mixture of 100 kg ha<sup>-1</sup> of ammonium sulfate, 300 kg ha<sup>-1</sup> of simple superphosphate, and 100 kg ha<sup>-1</sup> of potassium chloride was used, applied in the plot, and incorporated into the soil before planting. The fertilization in the top dressing was carried out approximately 60 days after the installation of the experiments, with 200 kg ha<sup>-1</sup> of ammonium sulfate. The weeds were controlled with the use of herbicides and manual weeding. Sowing was carried out in the planting line with a density of 300 seeds m<sup>-2</sup>. The irrigation started around 10-15 days after the implantation of the experiments, and the water was only removed close to the maturation of the material later. The harvest was carried out when the grains reached a humidity of 20-22%. Grain production data were obtained by weighing all grains harvested in the useful plot, after cleaning and uniform drying in the sun, until they reached a humidity of 13%.

### Biometric Analysis

The measured traits were analyzed using the univariate model and the multi-trait model through the Bayesian approach of Markov Chain Monte Carlo (MCMC).

The multi-trait model was given by:

$$y = X\beta + Zg + \varepsilon$$

where  $y$  is the vector of phenotypic data, and the conditional distribution was given by:  $y | \beta, g, i, G, R \sim N(X\beta + Zg, R \otimes I)$ ,  $G$  is the matrix of genotypic covariance,  $R$  is the matrix of residual covariance.  $I$  is an identity matrix,  $\beta$  is vector of systematic effects (genotypes mean and replication effects), assumed as  $\beta \sim N(\beta, \Sigma\beta \otimes I)$ .  $g$  is the vector of genotype effects, assumed as  $g | G, \sim N(0, G \otimes I)$ .  $e$  is the vector of residuals, assumed as  $e | R, \sim N(0, R \otimes I)$ . The uppercase bold letters  $\mathbf{X}$  and  $\mathbf{Z}$  refer to the incidence matrices for the effects  $\beta$  and  $g$ , respectively.

We assume that  $G$  and  $R$  follow an inverted Wishart distribution  $WI(v, V)$ , with hyperparameters  $v$  and  $V$  (Sorensen and Gianola, 2002). Hyperparameters for all prior distributions have been selected to provide non-informative or flat prior distributions. For the systematic effect ( $\beta$ ), a pre-uniform distribution was assigned. In addition, the

parameters  $\beta$ ,  $g$ ,  $G$ ,  $R$  were estimated following the set posterior distribution:  $P(\beta, g, G, R | y) \propto P(y | \beta, g, G, R) \times P(\beta, g, G, R)$ .

For the model, the package was used “MCMCglmm” (Hadfield et al., 2010, Hadfield et al., 2010) of the R software (R Development Core Team, 2020). A total of 10,000,000 samples were generated and assumed a burn-in period and thin range of 500,000 and 10 iterations, respectively, resulting in a final total of 50,000 samples. The convergence of the MCMC was verified by the criterion of Geweke et al. (1992), carried out in two R software packages: "boa" (Smith et al., 2007) and "CODA" (Convergence Diagnosis and Output Analysis) (Plummer et al., 2006).

The model was compared using the deviation information criterion (DIC) proposed by Spiegelhalter et al. (2002):

$$DIC = D(\bar{\theta}) + 2p_D$$

where  $D(\bar{\theta})$  is a point estimate of the deviance obtained by replacing the parameters with their posterior means estimates in the likelihood function and  $p_D$  is the effective number of model parameters. Models with a lower DIC should be preferred over models with a higher DIC.

The components of variance, broad-sense heritability, genotypic correlation coefficients between traits and breeding values were calculated from the posterior distribution. The package “boa” (Smith et al., 2007) R software was used to calculate the intervals of higher posterior density (HPD) for all parameters. A posteriori estimates for broad-sense heritability ( $h^2$ ) of the six traits for each iteration were calculated from the later samples of the variance components obtained by the univariate (1) and multivariate (2) model, using the expression:

$$h^2 = \frac{\sigma_g^2}{(\sigma_g^2 + \sigma_l^2 + \sigma_e^2)} \quad (1)$$

$$h^2 = \frac{\sigma_g^2}{(\sigma_g^2 + \sigma_e^2)} \quad (2)$$

For the multi-trait model, the genetic correlation coefficients between the pairs of traits in each environment were obtained, as suggested by Piepho et al. (2018), using the expression below for all models:

$$\rho_{l(1,2)} = \frac{\sigma_{gl(1,2)}}{\sqrt{\sigma_{gl(1)}^2 \sigma_{gl(2)}^2}}$$

where  $\hat{\sigma}_{gl}^2$  represents the genetic variance of the evaluated trait and  $\sigma_{gl(1,2)}$  represents the genetic covariance between pairs of traits.

### *Genetic selection based on selection index*

The multi-trait index based on factor analysis and genotype-ideotype distance (FIA-BLUP) (Rocha et al., 2018) was used to identify superior genotypes to be selected in the flood-irrigated rice breeding program.

$$P_{ij} = \frac{\frac{1}{d_{ij}}}{\sum_{i=1:n; j=1:m} \frac{1}{d_{ij}}},$$

where,  $P_{ij}$ : probability of the  $i^{\text{th}}$  genotype ( $i = 1, 2, \dots, n$ ) to be similar to the  $j^{\text{th}}$  ideotype ( $j = 1, 2, \dots, m$ );  $d_{ij}$ : genotype-ideotype distance from  $i^{\text{th}}$  genotype to  $j^{\text{th}}$  ideotype – based on standardized mean distance.

Selection gains (SG) were obtained directly from the FIA-BLUP result considering four different selection intensities: 12%, 20%, 40%, and 60%, which referred to the selection of 3, 5, 10, and 15 genotypes, respectively, as follows:

$$SG(\%) = \left( \frac{X_s - X_0}{X_0} \right),$$

where  $X_s$  is the overall mean of the estimated breeding values of the selected genotypes, and  $X_0$  is the general population average.

## **Results**

Geweke's convergence criterion indicates convergence for all dispersion parameters by generating 10,000,000 MCMC iterations, 500,000 samples for burn-in, and a sampling interval of 10, totaling 50,000 effective samples used for estimating variance components (Figure 1). However, all chains [(co)variance components] reached convergence by this criterion. Similar posterior mean, median and modal estimates were obtained for variance components, suggesting normal-appearing density. According to

the deviation information criteria (DIC), there was evidence that the complete model for multi-trait (DIC = 348.84, 466.01 and 671.70, environment 1, 2, and 3, respectively) is the one that best fits the data, which reveals the significance of genotypic effects (DIC = 675.40, 675.52 and 898.84, environments 1, 2 and 3, respectively) (Table 1). Therefore, the DIC values were lower when using the complete model (considering the effects of genotype x environment interaction), in which the difference about the genotype model was greater than 1.30 (Table 1). Hence according to Spiegelhalter et al. (2002), suffices to suggest that the use of the complete model can lead to greater precision in the estimation of parameters (Table 1). Since the DIC values are higher, it is possible to indicate the superiority of the complete model over the restricted models. On the other hand, as this component of the model is important, the “best” genotypes measured in different environments may not be the same.

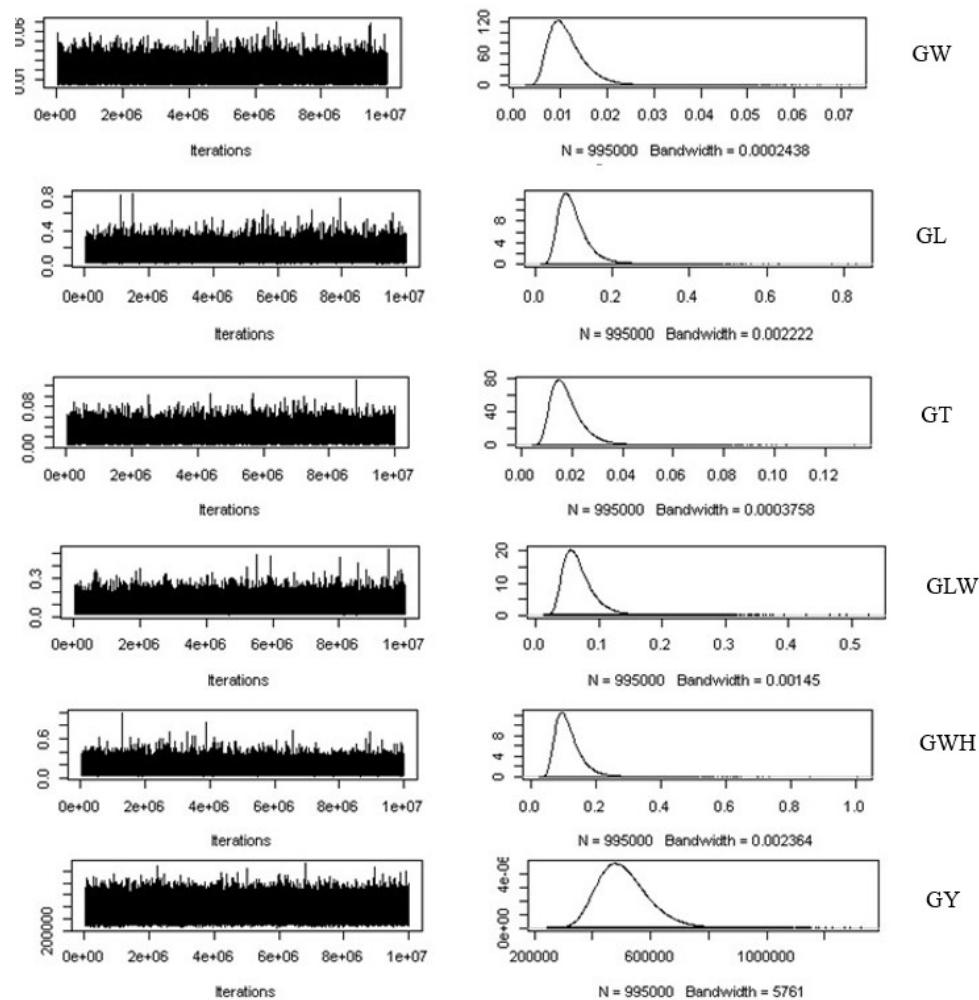
**Table 1.** Deviance information criteria for the full (considering the G x E interaction) and null (not considering the interaction) models.

Model	EM	Trait	Deviance information criteria (DIC)		
			Full	Genotype	GE
Single-Trait		GY	3609.89	3609.86	3665.60
Single-Trait		GL	18.05	18.04	206.61
Single-Trait		GW	-678.09	-676.38	-487.08
Single-Trait		GT	-798.52	-798.53	-610.88
Single-Trait		GLW	-260.72	260.07	-70.93
Single-Trait		GWH	-648.09	-647.86	-223.68
Multi-Trait	1		348.84	675.40	
Multi-Trait	2		466.01	675.52	
Multi-Trait	3		671.70	898.84	

GY: grain yields ( $\text{Kg ha}^{-1}$ ); GL: grain length (mm); GW: grain width (mm); GT: grain thickness (mm); GLW: grain length and width ratio; GWH: 100-grain weight (g); GE: genotype x environment interaction.

The posterior mean estimates for the variance components, which suggested density with chi-square and normal distributions (Figure 1). The GW to GWH traits showed a chi-square distribution (of which the Wishart distribution is a generalization) and only GY shows a normal distribution appearance.





**Figure 1.** Convergence for the genotypic variance of the six characteristics analyzed in the multi-trait model. The numbers on the right refer to the posterior density of the genetic variance estimates. The numbers on the left refer to Markov Chains for genetic variance estimates. GY: grain yields ( $\text{Kg ha}^{-1}$ ); GL: grain length (mm); GW: grain width (mm); GT: grain thickness (mm); GLW: grain length and width ratio; GWH: 100 grain weight (g).

Table 2 corresponds to the result of the estimate of heritability in the broad sense and the confidence interval with 95% probability for the multi-trait model. The estimates were different for mode, mean, median, and posterior density range (HPD). Highest estimates of heritability in the broad sense were for the GL, GW, and GWH traits, in all environments were observed. On the other hand, the lowest estimates consisted of the

grain yield and length-width ratio traits. The GT trait was considered highly heritable, with a credibility interval (95% probability) ranging from:  $h^2$ : 0.7890 – 0.9480; 0.7640 – 0.9440; 0.5640– 0.8610, in environments 1, 2 and 3, respectively (Table 2).

**Table 2.** Posterior inferences for mode, mean, median, and posterior density range (HPD) of heritability in the broad sense ( $h^2$ ), in three environments (E1, E2, and E3), considering the multi-trait model.

Trait	E1				
	Mean	Median	Mode	Lower Bound	Upper Bound
GY	3.03E-06	2.57E-07	-6.53E-08	7.59E-09	3.99E-06
GL	0.7074	0.7164	0.7302	0.5250	0.870
GW	0.8741	0.8793	0.8888	0.7890	0.9480
GT	0.40	0.3553	4.5914	6.630	7.0619
GLW	5.87E-02	0.00812	0.1519	-0.1730	0.1680
GWH	1.00	1.0001	1.0002	0.9660	1.04
Trait	E2				
	Mean	Median	Mode	Lower Bound	Upper Bound
GY	1.84E-05	1.93E-07	5.32E-08	4.66E-09	3.30E-06
GL	0.7040	0.7125	0.7307	0.5270	0.8670
GW	0.8590	0.8650	0.8769	0.7640	0.9440
GT	0.5440	0.2978	1.5010	3.71	4.07
GLW	1.34E-02	0.004371	5.5272	0.130	0.1380
GWH	0.9790	0.9811	0.9839	0.9110	1.04
Trait	E3				
	Mean	Median	Mode	Lower Bound	Upper Bound
GY	1.86E-06	1.82E-07	-5.11E-08	3.90E-09	3.08E-06
GL	0.7130	0.7202	0.7299	0.5450	0.8680
GW	0.7180	0.7202	0.7130	0.5640	0.8610
GT	-0.1680	0.7228	0.7357	-0.7470	1.80
GLW	-5.58E-03	-0.00523	0.2230	-7.59E-02	6.53E-02
GWH	0.9790	0.9808	0.9840	0.8920	1.06

GY: grain yields (Kg ha<sup>-1</sup>); GL: grain length (mm); GW: grain width (mm); GT: grain thickness (mm); GLW: grain length and width ratio; GWH: 100 grain weight (g); E: environments E1, E2, and E3, respectively.

The posterior inferences for mean and posterior density range (HPD) of the correlation between six characteristics of flood-irrigated rice, in three environments, considering the MTM model, is described in Table 3. GY showed a low correlation estimate between GT and GWH, in all environments, with a credibility interval (95% probability) ranging from ( $\rho$ = -0.5444 – 0.5477; -0.5806 – 0.5762; -0.5574 – 0.5618 e -

0.5980 – 0.5973; -0.5287 – 0.5247; -0.5561 – 0.5632, GT and GWH, in environments 1, 2 and 3, respectively). About GY, the traits that presented the highest correlation estimates were GL and GLW.

**Table 3.** Posterior inferences for mean and posterior density range (HPD) of the correlation between six traits, in three environments, considering the multi-trait model.

$\rho$	E1			E2			E3		
	Mean	Lower	Bound	Mean	Lower	Bound	Mean	Lower	Bound
GYxGL	-0.0035	-0.8758	0.8787	-7.82E-06	-0.9136	0.9180	2.61E-03	-0.9165	0.9249
GYxGW	0.0001	-0.7280	0.7271	-2.16E-03	-0.7512	0.7425	-9.90E-05	-0.6699	0.6760
GYxGT	0.00033	-0.5444	0.5477	-1.19E-03	-0.5806	0.5762	5.16E-04	-0.5574	0.5618
GYxGLW	-0.0031	-0.8282	0.8221	1.73E-03	-0.8309	0.8322	1.55E-03	-0.8318	0.8329
GYxGWH	0.00033	-0.5980	0.5973	-9.19E-04	-0.5287	0.5247	8.39E-04	-0.5561	0.5632
GLxGW	0.2383	0.1341	0.3529	0.2770	0.1566	0.4106	0.350	0.1966	0.5225
GLxGT	0.01885	-0.6007	0.6305	0.1940	-0.5355	0.9231	-2.58E-02	-0.5954	0.5310
GLx GLW	0.55150	0.2210	0.8371	0.510	0.1575	0.8225	0.570	0.2397	0.8540
GLxGWH	-0.0077	-0.4421	0.4357	0.120	-0.3182	0.5504	0.290	-0.1275	0.6759
GTxGLW	-0.2659	-0.6575	0.1517	-9.56E-02	-0.5262	0.3448	-8.87E-02	-0.5253	0.3513
GTxGWH	0.1794	-0.2442	0.5840	0.1640	-0.2594	0.5733	0.196	-0.2221	0.6032

GY: grain yields (Kg ha<sup>-1</sup>); GL: grain length (mm); GW: grain width (mm); GT: grain thickness (mm); GLW: grain length and width ratio; GWH: 100 grain weight (g); E: environments E1, E2, and E3, respectively.

*Variance Estimate*

Estimates of genotypic variances, residuals, and genotype x environment interaction in the multi-trait models were very different between environments (Table 4). The GY trait had a higher estimate of genotypic variance compared to the other traits. On the other hand, all traits showed similarities in the estimation of genotypic variance in each environment. Smaller interaction variance estimates were observed for GW and GT traits.

**Table 4.** Genetic parameters of six traits of flood-irrigated rice, in three environments, using multi-trait models.

Trait	EN	$\sigma_g^2$	$\sigma_i^2$	$\sigma_e^2$		
				$\sigma_{e1}^2$	$\sigma_{e2}^2$	$\sigma_{e3}^2$
GY	1	1.32	5.08E+05	4780	466.0	-1330
	2	72.28	7.99E+05	42020	-328.4	-36910
	3	1.71	8.307E+05	5480	79.86	-10930
GL	1	0.0983	0.0375	-4770	-0.123	-0.357
	2	0.1577	0.0611	41940	-0.2323	-0.415
	3	0.1641	0.0610	-54680	-0.3468	-0.6497
GW	1	0.018	2.39E-03	-4780	-0.0477	-0.0912
	2	0.022	3.24E-03	-4200	-0.050	-0.1028
	3	0.014	5.156E-03	-54760	-0.0628	-0.1272
GT	1	0.0117	3.14E-03	-4780	-0.0320	-0.0592
	2	0.010	3.26E-03	-4200	-0.0344	-0.0804
	3	0.014	4.810E-03	-54760	-0.0344	-0.1104
GLW	1	0.068	0.0135	-4770	0.014	-0.0340
	2	0.078	0.0212	-41990	-0.0291	-0.0324
	3	0.082	0.0177	-54730	-0.0164	-0.3350
GWH	1	0.11	0.018	-4780	-7.62E-03	-8.09E-04
	2	0.077	0.0186	-41990	-0.0172	-0.0252
	3	0.082	0.0193	-54750	5.962E-03	-6.01E-03

GY: grain yields (Kg ha<sup>-1</sup>); GL: grain length (mm); GW: grain width (mm); GT: grain thickness (mm); GLW: grain length and width ratio; GWH: 100 grain weight (g); E: environments E1, E2, and E3, respectively.

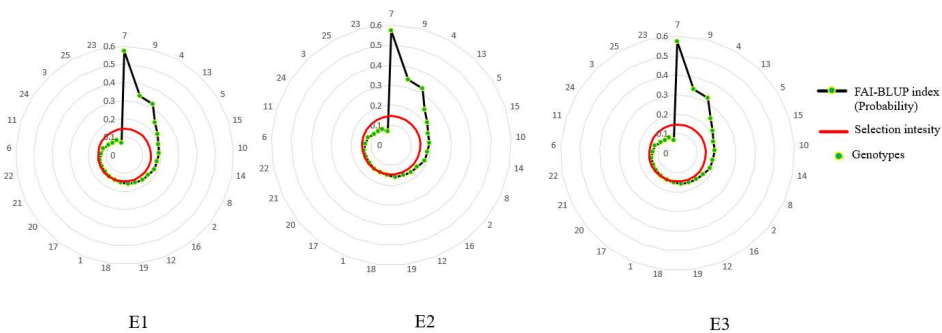
The selection gains obtained by the FIA-BLUP index considering four different selection intensities: 12%, 20%, 40%, and 60%, which referred to the selection of 3, 5, 10, and 15 genotypes, for six traits of flood-irrigated rice in three environments it represents in Table 5. The FIA-BLUP index indicated discrepant selection gains between environments for the same trait. Selection gains decreased with increasing selection intensity for the GL and GT traits in all environments (Table 5). The selection gain for the grain yield trait was approximately zero in environments 1 and 3. On the other hand, in environment 2, the greatest genetic gain was observed for the selection intensity of 10 genotypes. In this environment, the evaluated genotypes showed greater genetic variation compared to other environments. This environment showed greater genetic variance than other environments (Table 4). Regarding the GWH trait, it was the one with the greatest genetic gain in all environments (Table 5).

**Table 5.** Selection gains obtained by the FIA-BLUP index considering four different selection intensities: 12%, 20%, 40%, and 60%, which referred to the selection of 3, 5, 10, and 15 genotypes, for six traits of flood irrigated rice in three environments.

Trait	E1				E2				E3			
	3	5	10	15	3	5	10	15	3	5	10	15
GL	5.76	3.68	2.09	1.17	2.82	2.34	1.88	1.30	3.38	2.80	2.25	1.54
GT	3.57	2.37	2.37	1.49	-2.09	-1.46	-5.35	-1.93	4.27	2.98	1.09	0.39
GW	3.09	3.71	4.00	2.47	1.80	1.41	5.22	1.29	4.46	3.48	1.29	0.00
GWH	7.00	10.68	6.99	5.36	4.74	1.41	2.32	9.57	7.10	6.20	3.47	1.43
GY	0.00	0.00	0.00	0.00	1.57	4.14	9.00	5.67	0.00	0.00	0.00	0.00
GLW	2.60	-0.11	-1.98	-1.31	-5.10	1.40	1.17	1.66	-0.43	-0.70	1.05	1.46

GY: grain yields (Kg ha<sup>-1</sup>); GL: grain length (mm); GW: grain width (mm); GT: grain thickness (mm); GLW: grain length and width ratio; GWH: 100 grain weight (g); E: environments E1, E2, and E3, respectively.

Figure 2 shows the ranking of the 25 genotypes according to the FAI-BLUP index and their associated spatial probability, and the complete ranking was presented in (Table S). The results allowed for a unique and easy genotype selection process. We observed that genotypes 2 and 15 were similar in the three environments, they should be selected as high-performance multi-trait genotypes.



**Figure 2.** Genetic selection considering the selection intensity of 60% (15 genotypes). The green dotted line indicates the flood irrigated rice genotypes. Outside the red line are the selected genotypes. E: environments E1, E2, and E3, respectively.

**Discussion**

The successful evaluation of a breeding program is related to the accuracy of the prediction of genotypic values, which is closely linked to the adoption of adequate models. The implementation of bayesian multi-trait models is straightforward and, currently, it has been widely used due to the possibility of considering a priori knowledge in the modeling.

One of the main limitations of using multi-trait models is the correlations between traits that are in practice undesirable for breeders. Guo et al. (2020), demonstrated the ability of multi-trait models for four traits when compared to the single trait in wheat in response to selection and prediction accuracy. These authors showed a genetic gain of 22% compared to the single trait model across the environment reflected by the response to selection.

Chain convergence ensures that the most likely estimate for each co(variance) component is reached. The significance of genetic effects indicates genetic variability among the twenty-five flood-irrigated rice genotypes, which allows for genetic selection. The deviation information criterion is widely applied as a criterion to assess the best fit of models in Bayesian inference (Resende et al., 2014). HPD ranges also indicate the significance of genetic effects (genotypes). An advantage of Bayesian inference over frequentist inference is the possibility to obtain HPD intervals (Peixoto et al., 2021).

The HPD intervals are more accurate when compared to the confidence intervals of frequentist inference, which increases the reliability of the variance components and genetic parameters estimated through Bayesian inference. In rice, we did not find studies related to this interval, especially for these traits used in this study. On the other hand, in maize, the estimates of genetic parameters for efficiency of uptake and use of nitrogen under contrasting levels of nitrogen in the soil were demonstrated using multi-trait models (Torres et al., 2018). In soybean, estimates of genetic parameters for genetic selection of segregating progenies were reported using multi-trait models (Volpato et al., 2019).

The difference between mean, mode, and median of the broad-sense heritability estimates and the correlations between traits (Tables 2 & 3) reflect some lack of symmetry in the posterior distribution estimates. The lack of symmetry between mean, mode, and median heritability estimates in posterior distribution estimates was reported by Moura et al. (2014) and Torres et al. (2018). The low broad-sense heritability observed in the traits does not depend on the number of samples evaluated, since the Bayesian structure used is essentially recommended for situations involving small samples. On the other hand, quantitative characters are traits of agronomic interest, determined by several genes,

showing low expression, and significantly influenced by the environment (Falconer, 1981), reflected in the grain yield trait. Magalhães Junior et al. (2020), state that in rice breeding programs, productivity is identified as the main objective, however, the grain quality attributes, for example, long and fine grains, can directly reflect the market value and acceptance of the product by the consumer, making rice dependent on increased productivity and grain quality.

Low genetic gain for the grain yield trait using the FAI-BLUP index in environments 1 and 3 (Table 5). One possible justification is that genotypes belong to advanced comparative experiments after going through several generations of selection. The GWH trait showed the greatest genetic gain in all environments (Table 5). This high gain proves recent efforts by breeders in the search for advances in grain quality. Grain weight is crucial in determining rice yield; therefore, it is a trait that breeders have been directed towards these traits. The grain size, in addition to being important for yield, is also an important suggestion of intrinsic quality (Custodio et al., 2019).

The GWH trait is determined by grain size and fill rate, which is characterized by grain length, width, and height (Huang et al., 2013; Xie et al., 2015). Grain weight is related to genetic factors, and grain filling rate is affected by environmental conditions (Li et al., 2019). This trait negatively correlates with GLW (Song et al., 2015; Li et al., 2019). However, grain length and width are important factors influencing grain yield in rice (Zhang et al., 2012; Si et al., 2016). Compared to other traits, GWH was consistent in different environments and may help in the genetic gain of grain yield (Yang and Zhang, 2010). The GL, GW, and GT traits show moderate heritability (Table 2) and are difficult to estimate due to grain size. Therefore, the use of multi-trait models helps the breeder in genetic progress for the selection of various traits, environments, and genotypes. Guo et al. (2020), argues that joint prediction of multiple traits benefits from the genetic correlation between traits and indirect selection of a target trait with relatively low heritability that genetically correlated with other traits of high heritability. The joint multi-trait model obtained greater predictive accuracy than the single trait methods, especially for a trait with low heritability (Guo et al., 2020).

This result is in agreement with previous studies that reported that multi-trait models could be implemented to increase selection response for low heritability traits correlated with high heritability traits (Jia and Jannink 2012; Schulthess and Tal 2016; Rutkoski et al., 2016; Montesinos-López et al., 2018; Ward et al., 2019; Guo et al., 2020). Jia and Jannink (2012), also indicated that a multi-trait model is more effective when the

genetic correlation between these traits is moderate. Traits with lower heritability, such as grain yield, showed more benefits compared to high heritability traits such as GLW using the MTM model. Guo et al. (2020), reported that the traits with lower heritability performed better than traits with high heritability through the MTM model, as it contemplates the two-way interaction (Traits x Genotypes), and provides a better correlation estimate between dashes.

It has been reported in the literature that multi-trait analyzes improve parameter estimates (Schulthess et al., 2017; Montesinos-López et al., 2018). These authors also showed that the performance of multi-trait analysis depends considerably on whether only a few traits are missing in just some individuals or all individuals. Precise estimates of genetic parameters bring new perspectives on the application of bayesian methods to solve modeling problems in the genetic improvement of flood-irrigated rice.

One of the great contributions of biometrics is the evaluation of the indirect response by selecting a certain trait. However, the problem related to the indirect response is when the traits present unfavorable correlations causing undesirable changes in others. When the selection is for the grain yield trait, other components of these traits are indirectly selected, such as 100-grain weight, grain length, width and thickness, and the grain length and grain width ratio, which is associated with grain yield (Li et al., 2019).

The results of the present study can potentially be applied in plant breeding to achieve more selection cycles per unit of time for multiple traits, to accurately assess genotype performance due to the low number of test environments or due to lack of replication, and to predict the performance of genotypes for different environments that present low heritability.

## **Conclusion**

The Bayesian multi-trait model proved to be an adequate strategy for the genetic improvement of flooded rice. Furthermore, the bayesian multi-trait model has the potential for genetic evaluation of other crops.

The genotypes 2 and 15 were similar in the three environments, they should be selected as high-performance multi-trait genotypes.

## **Acknowledgment**



The authors would like to thank the Research Support Foundation of the State of Minas Gerais, the National Council for Scientific and Technological Development, and the Coordination for the Improvement of Higher Education Personnel for the financial support and researcher of Embrapa Rice and Beans Dr. Orlando Peixoto de Moraes (*in memory*) and Prof. Dr. Fabyano Fonseca e Silva (*in memory*). This study was financed in part by the Coordination for the Improvement of Higher Education Personnel - Brazil (CAPES) - Financial Code 001. The authors gratefully acknowledge the Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) for researcher fellowship to ICS 2018/26408-0.

### **Conflict of interest**

The authors declare that they have no conflict of interest.

### **Author contributions**

All authors contributed equally to the idea and preparation of the manuscript, and all authors read and approved the manuscript.

### **References**

- Arriagada, O., Mora, F., Dellarossa, J.C. et al. (2012). Bayesian mapping of quantitative trait loci (QTL) controlling soybean cyst nematode resistant. *Euphytica*. 186: 907. <https://doi.org/10.1007/s10681-012-0696-y>.
- Bocianowski, J; Nowosad, K; Szulc, P; Tratwal, A; Bakinowska, E & Piesik, D. (2019) Genetic parameters and selection of maize cultivars using Bayesian inference in a multi-trait linear model, *Acta Agriculturae Scandinavica, Section B — Soil & Plant Science*, 69:6, 465-478. <https://doi.org/10.1080/09064710.2019.1601764>.
- Burgueño, J., G. de los Campos, K. Weigel, and J. Crossa, 2012 Genomic prediction of breeding values when modeling genotype·environment interaction using pedigree and dense molecular markers. *Crop Sci.* 52(2): 707–719.

- Cané-Retamales, C., Mora, F., Vargas-Reeve, F. et al. Bayesian threshold analysis of breeding values, genetic correlation and heritability of flowering intensity in *Eucalyptus cladocalyx* under arid conditions. (2011). *Euphytica* 178: 177. <https://doi.org/10.1007/s10681-010-0292-y>.
- Custodio, M.C., Cuevas, R. P., Ynion, J., Laborte, A. G., Velasco, M. L., & Demont, M. (2019). Quality of rice: How is it defined by consumers, industry, food scientists and geneticists. *Trends in food science and technology*, 92, 122-137. <https://doi.org/10.1016/j.tifs.2019.07.039>.
- Duhnen A, Gras A, Teyssèdre S, Romestant M, Claustres B, Daydé J, et al. Genomic selection for yield and seed protein content in Soybean: A study of breeding program data and assessment of prediction accuracy. *Crop Sci.* 2017;57: 1325–1337.
- Falconer DS, Mackay TFC. 1996. Introduction to quantitative genetics. Edinburgh, SC: Addison Wesley Longman.
- Gianola D, Fernando RL. Bayesian Methods in Animal Breeding Theory. *J Anim Sci.* 1986;63: 217–244.
- Guo J., Khan J, Pradhan S, Shahi D, Khan N, Avci M, Mcbreen J, Harrison S, Brown-Guedira G et al., (2020). Multi-Trait Genomic Prediction of Yield-Related Traits in US Soft Wheat under Variable Water Regimes. *Genes* 2020, 11, 1270; doi:10.3390/genes11111270.
- Guo, G., F. Zhao, Y. Wang, Y. Zhang, L. Du et al., 2014 Comparison of single-trait and multiple-trait genomic prediction models. *BMC Genet.* 15(1): 30.
- Hayashi, T., Iwata, H. A Bayesian method and its variational approximation for prediction of genomic breeding values in multiple traits. *BMC Bioinformatics* 14, 34 (2013) doi:10.1186/1471-2105-14-34.
- Henderson, C. R., and R. L. Quaas, 1976 Multiple trait evaluation using relatives' records. *J. Anim. Sci.* 43(6): 1188–1197.
- Heslot, N., D. Akdemir, M. E. Sorrells, and J. L. Jannink, 2014 Integrating environmental covariates and crop modeling into the genomic selection framework to predict genotype by environment interactions. *Theor. Appl. Genet.* 127(2): 463–480.

- Huang, R., Jiang, L., Zheng, J., Wang, T., Wang, H., Huang, Y., et al. (2013). Genetic bases of rice grain shape: so many genes, so little known. *Trends Plant Sci.* 18, 218–226. doi: 10.1016/j.tplants.2012.11.001.
- Jarquín, D., J. Crossa, X. Lacaze, P. D. Cheyron, J. Daucourt et al., 2014 A reaction norm model for genomic selection using high-dimensional genomic and environmental data. *Theor. Appl. Genet.* 127(3): 595–607.
- Jarquín, D., J. Crossa, X. Lacaze, P. D. Cheyron, J. Daucourt et al., 2014 A reaction norm model for genomic selection using high-dimensional genomic and environmental data. *Theor. Appl. Genet.* 127(3): 595–607.
- Jia, Y.; Jannink, J.-L. Multiple-trait genomic selection methods increase genetic value prediction accuracy. *Genetics* 2012, 192, 1513–1522.
- Jiang, J., Q. Zhang, L. Ma, J. Li, Z. Wang et al., 2015 Joint prediction of multiple quantitative traits using a Bayesian multivariate antedependence model. *Heredity* 115(1): 29–36.
- Johnson HW, Robinson HF, Comstock RE. Genotypic and Phenotypic Correlations in Soybeans and Their Implications in Selection<sup>1</sup>. *Agronomy Journal*. 1955. p. 477.
- Junqueira VS, Azevedo Peixoto Ld, Galvêas Laviola B, Lopes Bhering L, Mendonça S, et al. (2016) Correction: Bayesian Multi-Trait Analysis Reveals a Useful Tool to Increase Oil Concentration and to Decrease Toxicity in *Jatropha curcas* L.. *PLOS ONE* 11(8): e0161046. <https://doi.org/10.1371/journal.pone.0161046>.
- Li R, Li M, Ashraf U, Liu S and Zhang J (2019) Exploring the Relationships Between Yield and Yield-Related Traits for Rice Varieties Released in China From 1978 to 2017. *Front. Plant Sci.* 10:543. doi: 10.3389/fpls.2019.00543.
- Liersch A, Bocianowski J, Woś H, Szała L, Sosnowska K, Cegielska-Taras T, Nowosad K, Bartkowiak-Broda I. 2016. Assessment of genetic relationships in breeding lines and cultivars of *Brassica napus* and their implications for breeding winter oilseed rape. *Crop Sci.* 56(4):1540– 1549.
- Magalhaes Junior AM, Streck EA, Aguiar GA, Facchinello PHK. (2020) Industrial Quality. In: Oliveira, A. C.; Pegoraro, C.; Viana, V. E. *The Future of Rice*

- Demand: Quality Beyond Productivity. 1 ed. Springer Nature Switzerland AG, Cap. 3, p. 47-67.
- Malosetti, M., Ribaut, J.M., Vargas, M. et al. A multi-trait multi-environment QTL mixed model with an application to drought and nitrogen stress trials in maize (*Zea mays* L.). *Euphytica* (2008) 161: 241. <https://doi.org/10.1007/s10681-007-9594-0>.
- Mathew, B., Holand, A.M., Koistinen, P. et al. Reparametrization-based estimation of genetic parameters in multi-trait animal model using Integrated Nested Laplace Approximation. *Theor Appl Genet* 129, 215–225 (2016). <https://doi.org/10.1007/s00122-015-2622-x>.
- Momen M, Campbell MT, Walia H, Morota G. Predicting Longitudinal Traits Derived from High-Throughput Phenomics in Contrasting Environments Using Genomic Legendre Polynomials and B-Splines. *G3* (Bethesda, Md.). 2019 Oct;9(10):3369-3380. <https://doi.org/10.1534/g3.119.400346>.
- Montesinos-López OA, Montesinos-López A, Crossa J, Kismiantini, Ramírez-Alcaraz JM, Singh R, et al. A singular value decomposition Bayesian multiple-trait and multiple-environment genomic model. *Heredity* (Edinb). Springer US; 2018; pmid:30120367.
- Montesinos-López, AO; Montesinos-López, A; Crossa, J; Toledo, FH, Pérez-Hernández, O; Eskridge, KM and Rutkoski, J. A Genomic Bayesian Multi-trait and Multi-environment Model. *G3* (Bethesda). 2016. 6(9):2725-44. doi: 10.1534/g3.116.032359.
- Montesinos-López, O. A., A. Montesinos-López, P. Pérez-Rodríguez, G. de los Campos, K. M. Eskridge et al., 2015 Threshold models for genome-enabled prediction of ordinal categorical traits in plant breeding. *G3* (Bethesda) 5(1): 291–300.
- Montesinos-López, O.A.; Montesinos-López, A.; Montesinos-López, J.C.; Crossa, J.; Luna-Vázquez, F.J.; Salinas-Ruiz, J. A Bayesian Multiple-Trait and Multiple-Environment Model Using the Matrix Normal Distribution. *Phys. Methods Stimul. Plant Mushroom Dev.* 2018, 19.
- Montesinos-López, O.A.; Montesinos-López, A.; Pérez-Rodríguez, P.; de los Campos, G.; Eskridge, K.; Crossa, J. Threshold models for genome-enabled prediction of

- ordinal categorical traits in plant breeding. *G3 Genes Genomes Genet.* 2015, 5, 291–300.
- Mora, F. & Serra, N. (2014). Bayesian estimation of genetic parameters for growth, stem straightness, and survival in *Eucalyptus globulus* on an Andean Foothill site. *Tree Genetics & Genomes*.10: 711. <https://doi.org/10.1007/s11295-014-0716-2>.
- Peixoto MA, Evangelista JSPC, Coelho IF, Alves RS, Laviola BG, Fonseca e Silva F, et al. (2021) Multiple-trait model through Bayesian inference applied to *Jatropha curcas* breeding for bioenergy. *PLoS ONE* 16(3): e0247775. <https://doi.org/10.1371/journal.pone.0247775>.
- Piepho HP. Allowing for the structure of a designed experiment when estimating and testing trait correlations. *J Agric Sci.* 2018;156: 59–70.
- Pollak, E. J., J. Van der Werf, and R. L. Quaas, 1984 Selection bias and multiple trait evaluation. *J. Dairy Sci.* 67(7): 1590–1595.
- Radkowski A, Sosin-Bzducha E, Radkowska I. 2017. Effects of silicon foliar fertilization of meadow plants on the nutritional value of silage fed to dairy cows. *J Elem.* 22(4):1311–1322.
- Resende MDV, Duda LL, Guimarães PRB, Fernandes JSC. Análise de modelos lineares mistos via inferência Bayesiana. *Rev Mat Estat.* 2001;19: 41–70. Available: [http://jaguar.fcav.unesp.br/RME/fasciculos/v19/A3\\_Artigo.pdf](http://jaguar.fcav.unesp.br/RME/fasciculos/v19/A3_Artigo.pdf).
- Rocha JRASC, Machado JC, Carneiro PCS. Multitrait index based on factor analysis and ideotype-design: proposal and application on elephant grass breeding for bioenergy. *GCB Bioenergy* (2018) 10, 52–60. <https://doi.org/10.1111/gcbb.12443>.
- Rolf MM, Garrick DJ, Fountain T, Ramey HR, Weaber RL, Decker JE, et al. Comparison of Bayesian models to estimate direct genomic values in multi-breed commercial beef cattle. *Genetics Selection Evolution.* 2015;47(1):23.
- Rutkoski, J.; Poland, J.; Mondal, S.; Autrique, E.; Pérez, L.G.; Crossa, J.; Reynolds, M.; Singh, R. Canopy temperature and vegetation indices from high-throughput phenotyping improve accuracy of pedigree and genomic selection for grain yield in wheat. *G3 Genes Genomes Genet.* 2016, 6, 2799–2808.

- Schaeffer, L. R., 1984 Sire and cow evaluation under multiple trait models. J. Dairy Sci. 67(7): 1567–1580.
- Schulthess AW, Zhao Y, Longin CFH, Reif JC (2017). Advantages and limitations of multiple-trait genomic prediction for Fusarium head blight severity in hybrid wheat (*Triticum aestivum* L.). Theor Appl Genet. <https://doi.org/10.1007/s00122-017-3029-7>.
- Schulthess, A.W.; Wang, Y.; Miedaner, T.; Wilde, P.; Reif, J.C.; Zhao, Y. Multiple-trait- and selection indices-genomic predictions for grain yield and protein content in rye for feeding purposes. Theor. Appl. Genet. 2016, 129, 273–287.
- Si, L., Chen, J., Huang, X., Gong, H., Luo, J., Hou, Q., et al. (2016). OsSPL13 controls grain size in cultivated rice. Nat. Genet. 48, 447–456. doi: 10.1038/ng.3518.
- Silva Júnior, AC, Silva, MJ, Cruz, CD, Nascimento, M, Azevedo, CF & Soares, PC. (2020). Patterns recognition methods to study genotypic similarity in flood-irrigated rice. *Bragantia*, 79(3), 356-363. <https://doi.org/10.1590/1678-4499.20200232>.
- Silva Júnior, A. C., da Silva, M. J., da Costa, W. G., de Sousa, I. C., Cruz, C. D., Nascimento, M., & Soares, P. C. (2021). Application of fuzzy logic for adaptability and stability studies in flood-irrigated rice (*Oryza sativa*). Plant Breeding, 1–9. <https://doi.org/10.1111/pbr.12973>.
- Song, X. J., Kuroha, T., Ayano, M., Furuta, T., Nagai, K., Komeda, N., et al. (2015). Rare allele of a previously unidentified histone H4 acetyltransferase enhances grain weight, yield, and plant biomass in rice. Proc. Natl. Acad. Sci. U.S.A. 112, 76–81. doi: 10.1073/pnas.1421127112.
- Sorensen DA, Gianola D. Likelihood, Bayesian and MCMC methods in quantitative genetics: statistics for biology and health. Springer-Verlag; 2002.
- Spiegelhalter DJ, Best NG, Carlin BP, van der Linde A. Bayesian Measures of Model Complexity and Fit. J R Stat Soc Ser B. Statistical Methodol. 2002;64: 583–639.
- Torres LG, Rodrigues MC, Lima NL, Trindade TFH, Silva FFe, Azevedo CF, et al. (2018) Multi-trait multi-environment Bayesian model reveals G x E interaction for

- nitrogen use efficiency components in tropical maize. PLoS ONE 13(6): e0199492. <https://doi.org/10.1371/journal.pone.0199492>.
- Valente BD, Rosa GJ, de Los Campos G, Gianola D, Silva MA. 2010. Searching for recursive causal structures in multivariate quantitative genetics mixed models. *Genetics*;185(2):633–644. <https://doi.org/10.1534/genetics.109.112979>.
- Van de Schoot, R., Depaoli, S., King, R. *et al.* Bayesian statistics and modelling. *Nat Rev Methods Primers* 1, 1 (2021). <https://doi.org/10.1038/s43586-020-00001-2>.
- Ventura HT, Silva FF, Varona L, de Figueiredo EAP, Costa EV, da Silva LP, et al. Comparing multi-trait Poisson and Gaussian Bayesian models for genetic evaluation of litter traits in pigs. *Livestock Science*. 2015;176:47–53.
- Volpato L, Alves RS, Teodoro PE, Vilela de Resende MD, Nascimento M, Nascimento ACC, et al. (2019) Multi-trait multi-environment models in the genetic selection of segregating soybean progeny. PLoS ONE 14(4): e0215315. <https://doi.org/10.1371/journal.pone.0215315>.
- Ward, B.P.; Brown-Guedira, G.; Tyagi, P.; Kolb, F.L.; Van Sanford, D.A.; Sneller, C.H.; Griffey, C.A. Multienvironment and multitrait genomic selection models in unbalanced early-generation wheat yield trials. *Crop Sci*. 2019, 59, 491–507.
- Xie, Q., Mayes, S., and Sparkes, D. L. (2015). Carpel size, grain filling, and morphology determine individual grain weight in wheat. *J. Exp. Bot.* 66, 6715–6730. doi: 10.1093/jxb/erv378.
- Yang, J., and Zhang, J. (2010). Grain-filling problem in ‘super’ rice. *J. Exp. Bot.* 61, 1–5. doi: 10.1093/jxb/erp348.
- Yu, H, Campbell, T, Zhang, Qi, Walia, H, Morota, G. Genomic Bayesian Confirmatory Factor Analysis and Bayesian Network To Characterize a Wide Spectrum of Rice Phenotypes. 2019. G3: GENES, GENOMES, GENETICS. vol. 9 n. 6. 1975-1986; <https://doi.org/10.1534/g3.119.400154>.
- Zhang, X., Wang, J., Huang, J., Lan, H., Wang, C., Yin, C., et al. (2012). Rare allele of OsPPKL1 associated with grain length causes extra-large grain and a significant yield increase in rice. *Proc. Natl. Acad. Sci. U.S.A.* 109, 21534–21539. doi: 10.1073/pnas.1219776110.