

RESEARCH ARTICLE

Multi-trait and multi-environment Bayesian analysis to predict the G x E interaction in flood-irrigated rice

Antônio Carlos da Silva Júnior^{1*}, Isabela de Castro Sant'Anna², Michele Jorge Silva Siqueira¹, Cosme Damião Cruz¹, Camila Ferreira Azevedo³, Moyses Nascimento³, Plínio César Soares⁴

1 Departamento de Biologia Geral, Universidade Federal de Viçosa, Viçosa, Minas Gerais, Brasil, **2** Centro de Seringueira e Sistemas Agroflorestais, Instituto Agrônomo (IAC), São Paulo, Brasil, **3** Departamento de Estatística, Universidade Federal de Viçosa, Viçosa, Minas Gerais, Brasil, **4** Empresa de Pesquisa Agropecuária de Minas Gerais–EPAMIG, Viçosa, Minas Gerais, Brazil

* antonio.silva.c.junior@gmail.com



OPEN ACCESS

Citation: da Silva Júnior AC, Sant'Anna IdC, Silva Siqueira MJ, Cruz CD, Azevedo CF, Nascimento M, et al. (2022) Multi-trait and multi-environment Bayesian analysis to predict the G x E interaction in flood-irrigated rice. PLoS ONE 17(5): e0259607. <https://doi.org/10.1371/journal.pone.0259607>

Editor: Mehdi Rahimi, KGUT: Graduate University of Advanced Technology, ISLAMIC REPUBLIC OF IRAN

Received: October 16, 2021

Accepted: April 15, 2022

Published: May 3, 2022

Peer Review History: PLOS recognizes the benefits of transparency in the peer review process; therefore, we enable the publication of all of the content of peer review and author responses alongside final, published articles. The editorial history of this article is available here: <https://doi.org/10.1371/journal.pone.0259607>

Copyright: © 2022 da Silva Júnior et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: The data belong to the Agricultural Research Company of Minas Gerais in Brazil. There are ethical restrictions on

Abstract

The biggest challenge for the reproduction of flood-irrigated rice is to identify superior genotypes that present development of high-yielding varieties with specific grain qualities, resistance to abiotic and biotic stresses in addition to superior adaptation to the target environment. Thus, the objectives of this study were to propose a multi-trait and multi-environment Bayesian model to estimate genetic parameters for the flood-irrigated rice crop. To this end, twenty-five rice genotypes belonging to the flood-irrigated rice breeding program were evaluated. Grain yield and flowering were evaluated in the agricultural year 2017/2018. The experimental design used in all experiments was a randomized block design with three replications. The Markov Chain Monte Carlo algorithm was used to estimate genetic parameters and genetic values. The flowering is highly heritable by the Bayesian credibility interval: $h^2 = 0.039\text{--}0.80$, and $0.02\text{--}0.91$, environment 1 and 2, respectively. The genetic correlation between traits was significantly different from zero in the two environments (environment 1: -0.80 to 0.74 ; environment 2: -0.82 to 0.86). The relationship of CV_e and CV_g higher for flowering in the reduced model ($CV_g/CV_e = 5.83$ and 13.98 , environments 1 and 2, respectively). For the complete model, this trait presented an estimate of the relative variation index of: $CV_e = 4.28$ and 4.21 , environments 1 and 2, respectively. In summary, the multi-trait and multi-environment Bayesian model allowed a reliable estimate of the genetic parameter of flood-irrigated rice. Bayesian analyzes provide robust inference of genetic parameters. Therefore, we recommend this model for genetic evaluation of flood-irrigated rice genotypes, and their generalization, in other crops. 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.

sharing the used data set, because this one still contains important about the genotypes information; in addition, the data are owned by state-owned organization. However, data access requests may be directed to: epamigsudeste@epamig.br.

Funding: 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 research 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. The funders had a role in study design, data collection and analysis, the decision to publish, and the preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

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 [1]. Rice breeding is primarily aimed at the development of high-yielding varieties with specific grain qualities, resistance to abiotic and biotic stresses in addition to superior adaptation to the target environment [2, 3]. In this case, the breeder needs to realize mutual relationships and knowledge of the interdependence between agronomically important traits that can improve accuracy selection [3].

Specifically, in the rice crop, evaluating multiple traits rather than a single trait aims to maximize grain yield and quality. This is possible through the exploration of genetic correlations between traits. In multi-trait analysis, the prediction of secondary traits can be used to improve the prediction of primary characteristics, especially when they have low heritability. Although consideration of the genetic correlation between traits is essential, modeling interactions between phenotypes provides essential information for developing breeding strategies that cannot be carried out with conventional multivariate approaches alone [2, 4, 5].

Biometric methods available are useful for analyzing a single trait measured in a single environment or across multiple environments with the genotype x environment (G x E) interaction [6–9]. This interaction can be defined as the differential response of genotypes to environmental variation [10], and offers an even greater challenge for the breeder [11]. The information from a network of experiments obtains a multi-trait and multi-environment (MTME) structure but presents limited statistical methodology that not correctly represents genetic and phenotypic variation in the data [12]. Therefore, genetic correlations and G x E interaction require more complex models that are difficult to converge in the context of mixed linear models [11]. Thus, Bayesian inference has become a useful statistical tool for dealing with complex models [13].

Bayesian inference has been used successfully in studies with complex models. [11] evaluated the multi-trait and multi-environment Bayesian model considering the G x E interaction for nitrogen use efficiency components in tropical corn. [13], applied such models through Bayesian inference applied to the breeding of *Jatropha curcas* for bioenergy. [14], used these models in the genetic selection of soybean progenies. [15], demonstrate such models in phenotypic and genotypic data in corn and wheat. However, few studies combine models of multiple traits in a multi-environment under a Bayesian point of view, mainly for rice cultivation.

Therefore, the objectives of this study were to propose a multi-trait and multi-environment Bayesian model to estimate genetic parameters for the flood-irrigated rice crop. In addition to comparing: (i) the complete model (considering the interaction between genotypes and environment) with the restricted model (not considering the interaction); (ii) estimates of genetic parameters of models with single and multiple traits, for grain yield and flowering.

Material and methods

Description of the experiment

The experiments were carried out in the State of Minas Gerais—Brazil, in the experimental fields of Agricultural Research Corporation of Minas Gerais State (EPAMIG), in the cities of Lambari (21° 58' 11.24" S, 45° 20' 59.6" W) and Janaúba (15° 48' 77" S, 43° 17' 59.09" W). Twenty-five genotypes belonging to the flood-irrigated rice breeding program of the Southeast region 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. Grain yield (GY, Kg ha⁻¹) and flowering period in days (FL) were evaluated in the agricultural year 2017/2018. The experimental design was a randomized complete block design with three replications. The experiments were conducted in floodplain soils with continuous flood irrigation. Grain production data in grams per useful plot were used, later converted into kilograms per hectare. Management practices were carried out according to recommendations for flood-irrigated rice in the region.

The useful area consisted of 4 m central of three internal rows (4 m x 0.9 m, totaling 3.60 m²). The soil preparation was carried out by plowing and harrowing around 30 days before sowing. For planting fertilization, a mixture of 100 kg ha⁻¹ of ammonium sulfate, 300 kg ha⁻¹ of simple superphosphate, and 100 kg ha⁻¹ 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⁻¹ 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⁻². 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 materials. 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%. Days for flowering correspond to the number of days from sowing to flowering when the plot presented approximately 50% of plants with panicles.

Biometric analysis

Grain yield (GY) and flowering period in days (FL) were analyzed using single- and multi-trait models using the Bayesian Markov Chain Monte Carlo (MCMC) approach. The objective was to compare: (i) the complete model (considering the interaction between genotypes and environments) with the restricted model (not considering the interaction); (ii) estimates of genetic parameters of models with single and multiple traits, for grain yield and flowering period.

The multi-trait and multi-environment (MTME) model was given by:

$$y = X\beta + W_1r + W_2u + \varepsilon$$

which can be rewritten as:

$$\begin{pmatrix} Y_1 \\ \dots \\ Y_2 \end{pmatrix} = X \begin{pmatrix} \beta(E1, 1) \\ \beta(E2, 1) \\ \dots \\ \beta(E1, 2) \\ \beta(E2, 2) \end{pmatrix} + W_1 \begin{pmatrix} r_1(E1, 1) \\ r_2(E2, 1) \\ \dots \\ r_1(E1, 2) \\ r_2(E2, 2) \end{pmatrix} + W_2 \begin{pmatrix} u(E1, 1) \\ u(E2, 1) \\ \dots \\ u(E1, 2) \\ u(E2, 2) \end{pmatrix} + \begin{pmatrix} \varepsilon(E1, 1) \\ \varepsilon(E2, 1) \\ \dots \\ \varepsilon(E1, 2) \\ \varepsilon(E2, 2) \end{pmatrix},$$

where: y is the vector of the phenotypic values of the two evaluated traits (grain yield, Y_1 ; and flowering period, Y_2); X is the incidence matrix of the systematic effects represented by β , assuming $\beta \sim N(\mu_\beta, I_\otimes \Sigma_\beta)$ so that $E1$ and $E2$ represent the two environments studied; W_1 it is the incidence matrix of the random effect of the environment; r is the ambient effect vector, $r \sim N(0, I_\otimes \Sigma_r)$; W_2 is the incidence matrix of the effects of the genotype x environment interaction; u is the random effects vector of the genotype x environment interaction, $u \sim N(0, I_\otimes \Sigma_u)$; and ε is the residual effects vector, $\varepsilon \sim N(0, I_\otimes \Sigma_\varepsilon)$.

The (co)variance matrices are given by:

$$\sum_r = \begin{pmatrix} \sigma_{ry(1)}^2 & \sigma_{ry(1,2)} & \sigma_{ry,E(1)} & \sigma_{ry,E(1,2)} \\ \sigma_{ry(1,2)} & \sigma_{ry(2)}^2 & \sigma_{ry,E(2,1)} & \sigma_{ry,E(2)} \\ \sigma_{ry,E(1)} & \vdots & \sigma_{rE(1)}^2 & \dots \\ \sigma_{ry,E(1,2)} & \dots & \dots & \sigma_{rE(2)}^2 \end{pmatrix}$$

$$\sum_u = \begin{pmatrix} \sigma_{uy(1)}^2 & \sigma_{uy(1,2)} & \sigma_{uy,E(1)} & \sigma_{uy,E(1,2)} \\ \sigma_{uy(1,2)} & \sigma_{uy(2)}^2 & \sigma_{uy,E(2,1)} & \sigma_{uy,E(2)} \\ \sigma_{uy,E(1)} & \vdots & \sigma_{uE(1)}^2 & \dots \\ \sigma_{uy,E(1,2)} & \dots & \dots & \sigma_{uE(2)}^2 \end{pmatrix}$$

$$\sum_\varepsilon = \begin{pmatrix} \sigma_{\varepsilon y(1)}^2 & \sigma_{\varepsilon y(1,2)} & \sigma_{\varepsilon y,E(1)} & \sigma_{\varepsilon y,E(1,2)} \\ \sigma_{\varepsilon y(1,2)} & \sigma_{\varepsilon y(2)}^2 & \sigma_{\varepsilon y,E(2,1)} & \sigma_{\varepsilon y,E(2)} \\ \sigma_{\varepsilon y,E(1)} & \vdots & \sigma_{\varepsilon E(1)}^2 & \dots \\ \sigma_{\varepsilon y,E(1,2)} & \dots & \dots & \sigma_{\varepsilon E(2)}^2 \end{pmatrix},$$

where: y represents grain yield and h represents flowering period in days; 1 and 2 represent the two environments studied. The variance-covariance matrices follow an inverted Wishart distribution, which was used as a priori to model the variance-covariance matrix [16].

The package “MCMCglmm” [17] was used. A total of 10,000,000 samples were generated and assuming a flare period and sampling interval of 500,000 and 10 iterations, respectively, this resulted in a final total of 50,000 samples. The convergence of the MCMC was verified by the criterion of [18], performed in two packages R “boa” [19] and “CODA” (convergence diagnosis) [20]. Even though Bayesian and frequentist structures were not directly compared, especially in the field of genetics [21], the same models were also adjusted based on the REML estimation method (Restricted Maximum Likelihood).

The complete models (considering the interaction between genotypes and environments) were compared with the null models (not considering the interaction) by the deviation information criterion (DIC) proposed by [22]:

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

where $\bar{\theta}$ is a point estimate of the deviation obtained by replacing the parameters with their later means estimates in the probability 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, coefficient of variation residual and genetic, variation index, and genotypic correlation coefficients between genetic traits and values were calculated from the posterior distribution. The package “boa” [19] was used to calculate the intervals of greater posterior density (HPD) for all parameters. The later estimates for the broad-sense heritability of grain yield and flowering period in days for each interaction were calculated from the later samples of the variance components obtained by the model,

using the expression:

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

where: $\sigma_g^{2(i)}$, $\sigma_r^{2(i)}$, and $\sigma_e^{2(i)}$ are the genetic, replication, and residual variations for each iteration, respectively.

The genetic correlation coefficients between the pairs of traits in each environment were obtained, as suggested by [23], using the expression below for all models: $\rho_{l(1,2)} = \frac{\sigma_{gl(1,2)}}{\sqrt{\sigma_{g(1)}^2 \sigma_{g(2)}^2}}$:

genetic correlation between environment and grain yield and $\rho_{h(1,2)} = \frac{\sigma_{gh(1,2)}}{\sqrt{\sigma_{gh(1)}^2 \sigma_{gh(2)}^2}}$: genetic cor-

relation between environment and flowering period in days. All data analysis was conducted using the statistical software package R version 4.1.0.

Results and discussion

Geweke's convergence criterion indicated convergence for all dispersion parameters by generating 10,000,000 MCMC strings, 500,000 samples for burn-in, and a sampling interval of 10, totaling 50,000 effective samples used for estimating variance components. Similar posterior mean, median and modal estimates were obtained for variance components, suggesting normal-appearing density. However, all chains reached convergence by this criterion. According to the deviation information criteria (DIC), there was positive evidence of interactions between genotypes and environments for all analyzed models (Table 1). However, the DIC values were lower when using the complete model (considering the effects of genotype x environment interaction), in which the difference in relation to the complete model was greater than 2 (Table 1), which according to [22] the use of full model can lead to greater precision in estimating parameters (Table 1). Therefore, since the obtained DIC values were greater than two, 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 cannot be the same. However, convergence was not achieved by the AI (Average Information) and EM (Expectation-Maximization) algorithms.

The posterior inferences for mean and higher posterior density range (HPD) considering the multi-trait and multi-environment (MTME) model are described in Table 2. The average values for the grain yield trait varied from 4210.91–3901.56 kg ha⁻¹ and flowering in days of 99.40–76.43, in environments 1 and 2, respectively (Table 2). The Bayesian credibility interval (95% probability) for average grain yield corresponds to 4191.70–4227.86 kg ha⁻¹, and 3852.60–3946.07 kg ha⁻¹ (p<1e-05), in environments 1 and 2, respectively. In relation to the flowering period in days, this interval corresponds to 98.09–100.69 e 74.56–78.22 (p<1e-05), environments 1 and 2, respectively.

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

Model	Trait	Deviance information criteria (DIC)	
		Full Model	Null Model
Multi-trait	GY, FL	-308.83	1967.79
Single-trait	GY	1867.49	1868.28
Single-trait	FL	162.19	697.67

GY: Grain Yield; FL: Flowering Period.

<https://doi.org/10.1371/journal.pone.0259607.t001>

Table 2. Posterior inferences for mean and highest posterior density range (HPD) considering the proposed complete multi-trait multi-environment model.

Trait	EN	post.mean	HPD 95%	
			LOWER	UPPER
GY	1	4210.91***	4191.7	4227.86
FL	1	99.4***	98.09	100.69
GY	2	3901.56***	3852.6	3946.07
FL	2	76.43***	74.56	78.22

*** Significância estatística: $p \leq 0.001$. GY: Grain Yield; FL: Flowering Period; EN: environment.

<https://doi.org/10.1371/journal.pone.0259607.t002>

Table 3 presents the subsequent inferences for mean and genetic variance; mode, mean, median, and highest posterior density range (HPD) of heritability in the broad sense; and the mode, mean, median, and greater posterior density interval (HPD) of the genetic correlation, considering MTME. The grain yield trait in environment 2 was considered weakly heritable with Bayesian credibility interval (95% probability): $h^2 = 4.36E-07-9.21E-06$, and in environment 1, it was moderately heritable $h^2 = 0.39-0.757$. The low estimate of heritability observed does not depend on the number of samples evaluated, since the Bayesian structure used is essentially recommended for situations involving small sample sizes. In addition, the grain yield trait is largely influenced by the environment as it is a quantitative character [24], which reflects this low estimate of heritability. In relation to flowering period, it is highly heritable by the Bayesian credibility interval (95% probability): $h^2 = 0.039-0.80$, and $0.02-0.91$, environment 1 and 2, respectively. And the posterior mean of the genetic correlation between traits was significantly different from zero (95% credibility intervals) in the two environments (environment 1: -0.80 to 0.74; environment 2: -0.82 to 0.86) (Table 3).

Bayesian methods access the posterior density ranges of genetic parameters (Fig 1). The genetic parameters of the flood-irrigated rice genotypes for each trait and their larger posterior density ranges (HPD) were obtained to assist in the selection of genotypes.

In Fig 2, there is a unimodal distribution in which there is a mixture of two distinct populations for posterior density in relation to the genotypic correlation between traits for the complete MTME model. The red line represents the posterior density for environment 1, while the blue line represents the posterior density for environment 2.

Table 3. Resume of inferences for mean and genetic variance; mode, mean, median, and highest posterior density range (HPD) of heritability in the broad sense; and the mode, mean, median, and highest posterior density range (HPD) of the genetic correlation, considering the complete model multi-trait and multi-environment.

Trait	EN	h^2			HPD 95%	
		Mode	Mean	Median	LOWER	UPPER
GY	1	0.11	0.28	0.18	0.39	0.757
GY	2	1.30E-06	3.32E-06	2.24E-06	4.36E-07	9.21E-06
FL	1	0.12	0.31	0.24	0.039	0.80
FL	2	0.06	0.27	0.13	0.02	0.91
		Genotypic Correlation			HPD 95%	
		Mode	Mean	Median	LOWER	UPPER
GY, FL	1	-0.0076	-0.027	-0.028	-0.80	0.74
GY, FL	2	0.037	0.018	0.019	-0.82	0.86

EN: environment; GY: Grain Yield; FL: Flowering Period; h^2 : heritability.

<https://doi.org/10.1371/journal.pone.0259607.t003>

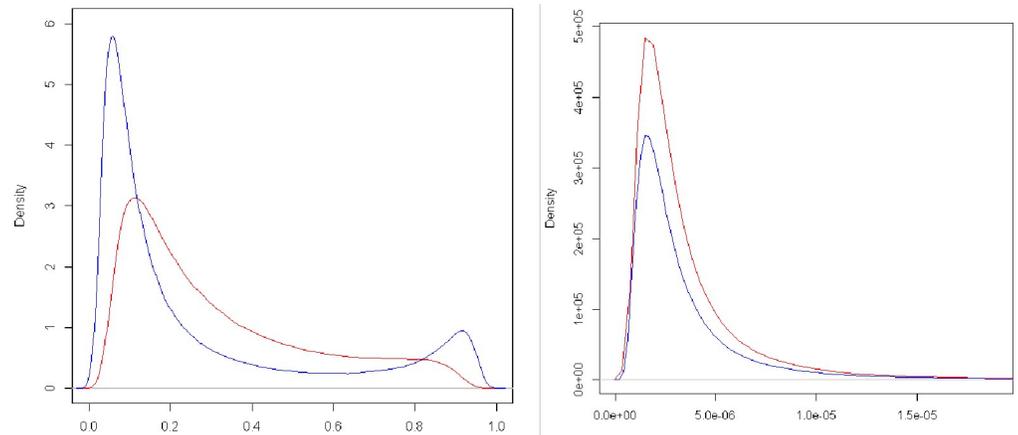


Fig 1. Posterior density for the complete model proposed by multi-traits multi-environment (left: flowering period and right: grain yield). The red line represents the posterior density for environment 1, while the blue line represents the posterior density for environment 2.

<https://doi.org/10.1371/journal.pone.0259607.g001>

Variance estimate

The a posteriori estimates of the genotypic and residual variances for the reduced model (MTME) were very discrepant among the environments (Table 4). The grain yield trait in the

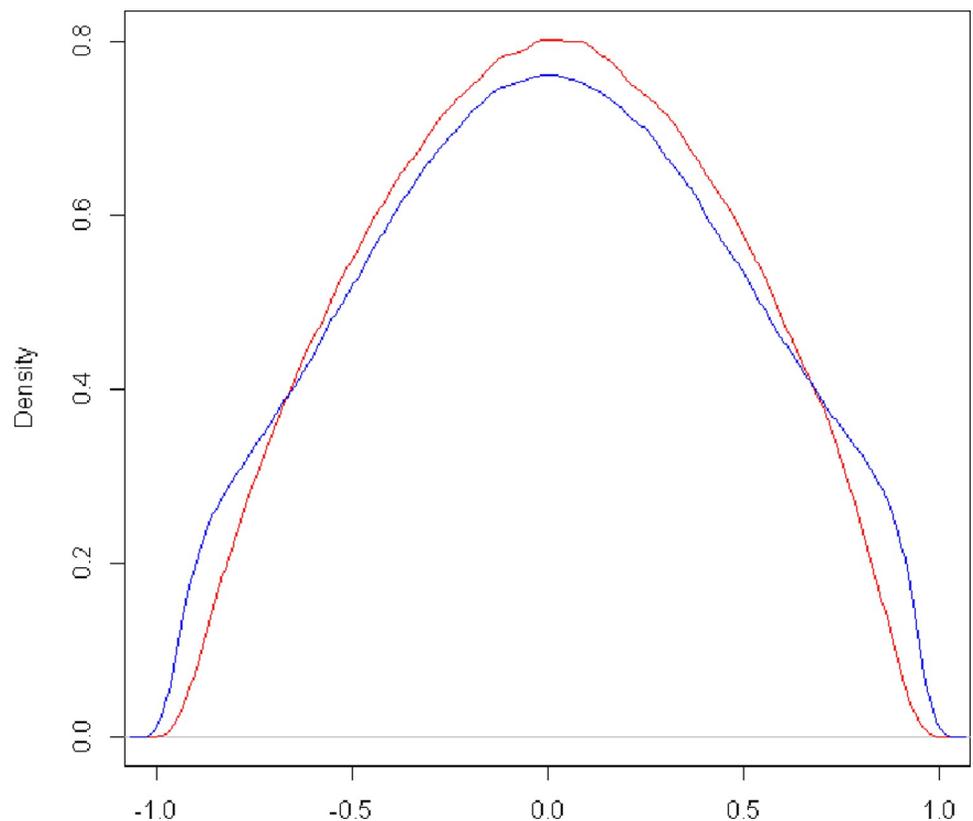


Fig 2. Posterior density for the genotypic correlation between the grain yield trait and flowering period in days for the model proposed by multi-traits and multi-environment. The red line represents the posterior density for environment 1, while the blue line represents the posterior density for environment 2.

<https://doi.org/10.1371/journal.pone.0259607.g002>

Table 4. Genetic parameters for traits grain yield and flowering period in days, in two environments, using multi-trait multi-environment (MTME) models.

	Model	Trait	EM	Component		
				σ_g^2	σ_r^2	σ_{int}^2
Multi trait	Null	GY	1	38.93	8.86	-
		GY	2	23.19	6.13	-
		FL	1	14.95	4.45E-5	-
		FL	2	32.63	0.19	-
	Full	GY	1	2.63	7.24	4.64
		GY	2	2.70	7.39	4.77
		FL	1	3.35	0.18	7.57
		FL	2	5.69	0.34	15.91

EN: environment; GY: Grain Yield; FL: Flowering Period; $\sigma_g^2, \sigma_r^2, \sigma_{int}^2$: are the genetic, replication, and interaction variations, respectively.

<https://doi.org/10.1371/journal.pone.0259607.t004>

reduced model obtained a greater estimate of genotypic variance in environment 1, corresponding to a difference of approximately 60% greater, in relation to environment 2. This indicates a greater influence of genetic components on environmental components in the expression of traits. On the other hand, this estimate for the complete model showed less variation between environments, especially for the flowering period in days trait. This result showed the best consistency for the complete model (MTME). Estimates of the variance of the greatest interaction were observed for the flowering period trait.

Relative variation index. The ratio of the coefficient of variation genotypic and the coefficient of variation residual (CV_g/CV_e) corresponded to the relative variation index. When this index is greater than one unit, it suggests that genetic variation is more influential than residual variation. This was observed in this study for both traits, in the reduced model proposed by MTME (Table 5). For the complete model in the grain yield trait, this relationship was less than one unit. The relationship of CV_e and CV_g higher for flowering period trait in the reduced model ($CV_g/CV_e = 5.83$ and 13.98 , environments 1 and 2, respectively). This trait has greater variability and is highly promising for selection. This is due to its complex genetic inheritance resulting from the involvement of several genes with little effect on the phenotype [24, 25]. For the complete model, this trace presented an estimate of the relative variation index of: $CV_g/CV_e = 4.28$ and 4.21 , environments 1 and 2, respectively (Table 5). Then, this trait presented a higher posterior density interval (HPD) for 80 and 91% heritability, environment 1 and 2,

Table 5. Coefficient of variation residual (CV_e , %), coefficient of variation genotypic (CV_g , %) and relative variation index (CV_g/CV_e) for the multi-trait and multi-environment model.

Model	Trait	EN	CV_g (%)	CV_e (%)	CV_g/CV_e
Null	GY	1	0.149	0.071	2.11
	GY	2	0.122	0.063	1.92
	FL	1	3.89	0.67	5.83
	FL	2	7.47	0.57	13.98
Full	GY	1	0.039	0.064	0.61
	GY	2	0.042	0.070	0.60
	FL	1	1.84	0.43	4.28
	FL	2	3.20	0.76	4.21

EN: environment; GY: Grain Yield; FL: Flowering Period; EN: environment.

<https://doi.org/10.1371/journal.pone.0259607.t005>

respectively, reinforcing the possibility of genetic gain (Table 3). This implies that the variability observed in these traits has genetic predominance, which is interesting in the process of genetic gain in a flood-irrigated rice breeding program. The coefficient of variation genotypic for the grain yield trait was low in both environments. This is justified by the fact that these genotypes belong to advanced comparative trials in which they have gone through several generations of selection.

One of the ways to succeed in breeding programs is related to the accurate prediction of genotypic values, which is closely related to the adoption of adequate models. Thus, in this study, we apply a new statistical approach for estimating variance components in floodplain rice breeding schemes. The implementation of multi-trait multi-environment models Bayesian is straight forward and currently has been widely used due to the possibility of considering a priori knowledge in modeling. In addition to its application wide application in animal breeding [26, 27], Bayesian multi-trait analysis has been reported in plant breeding [11, 13–15, 28].

Bayesian inference has been used since 1986 [29] and has been further explored in recent years [2, 11, 13–15, 30, 31] due to major computational advances and new applications and methodologies [32]. However, Bayesian analysis is based on knowledge of the posterior distribution of the parameters to be estimated. This allows the construction of exact credibility intervals for estimates of random variables and variance components [33]. Values for the 95% distribution credibility interval for the broad-sense heritability parameter found in this study (Table 3) were also presented in the study by [11] to estimate genetic parameters for efficiency of uptake and efficiency of use of N under contrasting soil N levels via MTME models. Another study, based on the estimation of genetic parameters for genetic selection of segregating soybean progenies using the MTME model [14]. The difference between mean, mode, and median of the broad-sense heritability estimates (Table 3) reflects 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 [11, 30].

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 traits of agronomic interest, determined by several genes, demonstrate low expression and significantly influenced by the environment [24], reflected in the traits grain yield and flowering period in days.

Based on the results of heritability estimation in the broad sense for the GY and FL traits varied in: $h^2 = 0.39–0.757$ and $h^2 = 4.36E-07–9.21E-06$; $h^2 = 0.039–0.80$, and $0.02–0.91$, environment 1 and environment 2, respectively (Table 3). [34] found heritability estimates of 0.48 and 0.94 for GY and FL, respectively, using 198 rice progenies by the ANOVA technique. [35] evaluated upland rice genotypes, by this same technique, and obtained an estimate of h^2 for GY and FL traits 0.35 and 0.77, respectively. [36] obtained results of h^2 ranging from 0.44–0.87 for GY, 0.46–0.94 for FL. [37] applied a mixed model in studies using FL and GY traits, found an estimate of h^2 de 0,88, and 0,71, respectively. [38], using the ANOVA method, found h^2 of 0.76 for FL and [39] estimate of 0.30 for GY. Regarding the estimate of CVs using the ANOVA method for the FL and GY traits, representing 2.98% and 15.28%, respectively [34].

The amounts of data that breeding programs around the world are generating continue to increase; consequently, there is a growing need to extract more knowledge from the data being produced. For this, MTME models are commonly used to take advantage of correlated traces to improve parameter estimation and prediction accuracy. However, when there are a large number of features, implementing these types of models is a challenge. Therefore, it is necessary to develop efficient models of multiple trait and multiple environments for selection, in order to take advantage of multiple correlated features. In this work, was proposed an alternative method to analyze MTME model data that could be useful for genotype selection, and

estimation of genetic parameters in flooded rice in the context of an abundance of traits and environments.

Full-model multi-trait analysis tends to be powerful and provide more accurate estimates than single-trait analysis because the previous method can take into account the underlying correlation structure found in a multi-trait dataset. However, Bayesian and non-Bayesian inferences from the MTME model analysis are complex and computationally demanding [40]. [28] argue that Bayesian multi-trait analysis is more appropriate than ANOVA to perform analyzes and select superior genotypes for genetic improvement since the Bayesian model can capture small genetic differences between families, while ANOVA cannot. In this study, was explained how to make Bayesian inference using a multi-trait multi-environment Bayesian model in plant breeding. These results showed that the approach was efficient in estimating genetic parameters in flooded rice.

The correlation study revealed favorable associations for the traits in studies in two settings. This result indicates that the selection of genotypes characterized by longer flowering period favors grain yield in flood-irrigated rice, which is desirable for rice cultivation since later plants tend to be more productive. However, late cycle cultivars tend to be more productive in relation to the early cycle, since they obtain an increase in the amount of photoassimilates that are translocated to grains [3]. This result justifies the significant correlation between the traits. According to [24], correlations between traits may be the result of pleiotropy or genetic linkage. Thus, if the undesirable correlations are caused by genetic linkage, these associations can be broken by recombination caused by crossing or self-fertilization; consequently, these factors do not necessarily become major impediments to breeding programs [41–43].

Another point that we would like to highlight is that our proposed model is of MTME, but with the restriction that an identity matrix is assumed for the variance-covariance matrix of the environments. However, even with this restrictive assumption in the variance-covariance matrix of the environments, the model has the advantage of taking into account the terms of interaction x trait, genotype x trait, and the environment x genotype x trait. Furthermore, it takes into account the correlated traits. The program of irrigated rice program aims to obtain the desired results in a short period and with precision, therefore, the choice of the model will be use as breeding strategies.

Conclusion

The multi-trait and multi-environment Bayesian model were efficient to estimate genetic parameters for the flood-irrigated rice crop.

The 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.

Author Contributions

Conceptualization: Antônio Carlos da Silva Júnior, Isabela de Castro Sant'Anna, Michele Jorge Silva Siqueira, Cosme Damião Cruz, Camila Ferreira Azevedo, Moyses Nascimento.

Data curation: Antônio Carlos da Silva Júnior, Isabela de Castro Sant'Anna, Michele Jorge Silva Siqueira, Cosme Damião Cruz, Plínio César Soares.

Formal analysis: Antônio Carlos da Silva Júnior, Isabela de Castro Sant'Anna, Michele Jorge Silva Siqueira, Cosme Damião Cruz, Camila Ferreira Azevedo, Moyses Nascimento, Plínio César Soares.

Funding acquisition: Antônio Carlos da Silva Júnior, Cosme Damião Cruz, Plínio César Soares.

Investigation: Antônio Carlos da Silva Júnior, Cosme Damião Cruz, Camila Ferreira Azevedo, Plínio César Soares.

Methodology: Antônio Carlos da Silva Júnior, Isabela de Castro Sant'Anna, Michele Jorge Silva Siqueira, Cosme Damião Cruz, Camila Ferreira Azevedo, Moyses Nascimento.

Project administration: Cosme Damião Cruz, Plínio César Soares.

Resources: Antônio Carlos da Silva Júnior, Cosme Damião Cruz, Plínio César Soares.

Software: Antônio Carlos da Silva Júnior, Isabela de Castro Sant'Anna, Michele Jorge Silva Siqueira, Cosme Damião Cruz, Camila Ferreira Azevedo, Moyses Nascimento.

Supervision: Antônio Carlos da Silva Júnior, Michele Jorge Silva Siqueira, Cosme Damião Cruz, Camila Ferreira Azevedo, Moyses Nascimento.

Validation: Antônio Carlos da Silva Júnior, Isabela de Castro Sant'Anna, Michele Jorge Silva Siqueira, Cosme Damião Cruz, Moyses Nascimento.

Visualization: Antônio Carlos da Silva Júnior, Isabela de Castro Sant'Anna, Michele Jorge Silva Siqueira, Cosme Damião Cruz, Camila Ferreira Azevedo, Moyses Nascimento.

Writing – original draft: Antônio Carlos da Silva Júnior, Isabela de Castro Sant'Anna, Michele Jorge Silva Siqueira, Cosme Damião Cruz, Camila Ferreira Azevedo, Moyses Nascimento, Plínio César Soares.

Writing – review & editing: Antônio Carlos da Silva Júnior, Isabela de Castro Sant'Anna, Michele Jorge Silva Siqueira, Cosme Damião Cruz, Camila Ferreira Azevedo, Moyses Nascimento, Plínio César Soares.

References

1. 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>.
2. 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>.
3. Silva Júnior AC, Santanna IC, Silva GN, Cruz CD, Nascimento M, Bhereng LL, et al. (2022). Computational intelligence to study the importance of characteristics in flood-irrigated rice. *Acta Scientiarum-Agronomy*.
4. 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> PMID: 20351220
5. 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> PMID: 31427454
6. Burgueño J., de los Campos G., Weigel K., and Crossa J. (2012). Genomic prediction of breeding values when modeling genotype-environment interaction using pedigree and dense molecular markers. *Crop Sci.* 52(2): 707–719.
7. Heslot N., Akdemir D., Sorrells M. E., and Jannink J. L. (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. <https://doi.org/10.1007/s00122-013-2231-5> PMID: 24264761
8. Jarquín D., Crossa J., Lacaze X., Cheyron P. D., Daucourt J. et al. (2014). A reaction norm model for genomic selection using high-dimensional genomic and environmental data. *Theor. Appl. Genet.* 127(3): 595–607. <https://doi.org/10.1007/s00122-013-2243-1> PMID: 24337101

9. Montesinos-López O. A., Montesinos-López A., Pérez-Rodríguez P., de los Campos G., Eskridge K. M. et al. (2015). Threshold models for genome-enabled prediction of ordinal categorical traits in plant breeding. *G3 (Bethesda)* 5(1): 291–300.
10. Silva Júnior AC, Silva MJ, Sousa I, Costa WG, Cruz CD, Nascimento M, et al. (2021). Fuzzy logic for adaptability and stability studies in irrigated rice (*Oryza Sativa* L.) genotypes. *Plant Breeding*. <https://doi.org/10.1111/pbr.12896> PMID: 34239217
11. 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> PMID: 29949626
12. Malosetti M., Ribaut J.M., Vargas M. et al. (2008). A multi-trait multi-environment QTL mixed model with an application to drought and nitrogen stress trials in maize (*Zea mays* L.). *Euphytica*. 161: 241. <https://doi.org/10.1007/s10681-007-9594-0>.
13. Peixoto MA, Evangelista JSPC, Coelho IF, Alves RS, Laviola BG, Fonseca 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> PMID: 33661980
14. 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> PMID: 30998705
15. Montesinos-López OA, Montesinos-López A, Montesinos-López JC, Crossa J, Luna-Vázquez FJ, Salinas-Ruiz JA. (2018). Bayesian Multiple-Trait and Multiple-Environment Model Using the Matrix Normal Distribution. *Phys. Methods Stimul. Plant Mushroom Dev.* 19.
16. Van Tassel CP, Van Vleck LD. (1996). Multiple-trait Gibbs sampler for animal models: flexible programs for Bayesian and likelihood-based (co)variance component inference. *J Anim Sci*. 74:2586–2597. PMID: 8923173. <https://doi.org/10.2527/1996.74112586x>.
17. Hadfield J. MCMC Methods for Multi-Response Generalized Linear Mixed Models: The MCMCglmm R Package. *J Stat Softw.* 2010; 33:1–22. <https://doi.org/10.18637/jss.v033.i02>. PMID: 20808728
18. Geweke J. Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments. In: Bernardo JM, Berger JO, Dawid AP, Smith AFM, editors. *Bayesian Statistics 4*. Oxford University Press; 1992. pp 625–631.
19. Smith BJ. boa: an R package for MCMC output convergence assessment and posterior inference. *J Stat Softw.* 2007; 21:1–37. <https://doi.org/10.18637/jss.v021.i11>.
20. Plummer M, Best N, Cowles K, Vines K. CODA: Convergence diagnosis and output analysis for MCMC. *R news.* 2006; 6:7–11.
21. Sorensen DA, Gianola D. *Likelihood, Bayesian and MCMC methods in quantitative genetics: statistics for biology and health*. Springer-Verlag; 2002.
22. 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.
23. Piepho HP. Allowing for the structure of a designed experiment when estimating and testing trait correlations. *J Agric Sci.* 2018; 156: 59–70.
24. Falconer DS, Mackay TFC. 1996. *Introduction to quantitative genetics*. Edinburgh, SC: Addison Wesley Longman.
25. Cruz CD; Regazzi AJ; Carneiro PCS. (2012). *Modelos biométricos aplicados ao melhoramento genético*. UFV. Viçosa.
26. 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.
27. Rolf MM, Garrick DJ, Fountain T, Ramey HR, Weaver 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. <https://doi.org/10.1186/s12711-015-0106-8> PMID: 25884158
28. 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> PMID: 27500402
29. Gianola D, Fernando RL. *Bayesian Methods in Animal Breeding Theory*. *J Anim Sci.* 1986; 63: 217–244.
30. 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>.

31. 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.
32. 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>.
33. 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.
34. Guimarães P. H. R., Melo P. G. S., Cordeiro A. C. C., Torga P. P., Rangel P. H. N., Castro A. P. Index selection can improve the selection efficiency in a rice recurrent selection population. *Euphytica* (2021) 217:95. <https://doi.org/10.1007/s10681-021-02819-7>.
35. Sari W. K., Nualsri C., Junsawang N., Soonsuwon W. Combining ability and heritability for yield and its related traits in Thai upland rice (*Oryza sativa* L.). *Agr. Nat. Resour.* Vol. 54 No. 3 (2020)
36. Margaret Catolos, Nitika Sandhu, Shalabh Dixit, Shamsudin Noraziya A. A., Naredo Ma E. B., McNally Kenneth L., et al. Genetic Loci Governing Grain Yield and Root Development under Variable Rice Cultivation Conditions. *Front. Plant Sci.*, 16 October 2017 | <https://doi.org/10.3389/fpls.2017.01763>.
37. Bhandari A., Bartholomé J., Cao T.-V., Kumari N., Frouin J., Kumar A., et al. (2019). Selection of trait-specific markers and multi-environment models improve genomic predictive ability in rice. *PLoS ONE*, 14(5). <https://doi.org/10.1371/journal.pone.0208871> PMID: 31059529
38. Sangaré J. R., Konaté A. K., Cissé F., Sanni A. Assessment of genetic parameters for yield and yield related-traits in an intraspecific rice (*Oryza sativa* L.) population. *J. Plant Breed. Genet.* 05 (02) 2017. 45–56.
39. Xu Y., Wang X., Ding X. et al. Genomic selection of agronomic traits in hybrid rice using an NCII population. *Rice* 11, 32 (2018). <https://doi.org/10.1186/s12284-018-0223-4> PMID: 29748895
40. 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> PMID: 26582509
41. Liersch A, Bocianowski J, Woś H, Szala L, Sosnowska K, Cegielska-Taras T, et al. (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.
42. 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.
43. 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>.