# Multi-trait multi-environment models in the genetic selection of segregating soybean progeny

**Leonardo Volpato**[1]*, **Rodrigo Silva Alves**[2], **Paulo Eduardo Teodoro**[3], **Marcos Deon Vilela de Resende**[4], **Moysés Nascimento**[4], **Ana Carolina Campana Nascimento**[4], **Willian Hytalo Ludke**[1], **Felipe Lopes da Silva**[1], **Aluízio Borém**[1]

1 Federal University of Viçosa—Department of Plant Science, University Campus, Viçosa, Minas Gerais, Brazil, 2 Federal University of Viçosa—Department of General Biology, University Campus, Viçosa, Minas Gerais, Brazil, 3 Federal University of Mato Grosso do Sul—Department of Plant Science, University Campus, Chapadão do Sul, Mato Grosso do Sul, Brazil, 4 Federal University of Viçosa—Department of Statistics, University Campus, Viçosa, Minas Gerais, Brazil

* leo.agroufv@gmail.com

## Abstract

At present, single-trait best linear unbiased prediction (BLUP) is the standard method for genetic selection in soybean. However, when genetic selection is performed based on two or more genetically correlated traits and these are analyzed individually, selection bias may arise. Under these conditions, considering the correlation structure between the evaluated traits may provide more-accurate genetic estimates for the evaluated parameters, even under environmental influences. The present study was thus developed to examine the efficiency and applicability of multi-trait multi-environment (MTME) models by the residual maximum likelihood (REML/BLUP) and Bayesian approaches in the genetic selection of segregating soybean progeny. The study involved data pertaining to 203 soybean $F_{2:4}$ progeny assessed in two environments for the following traits: number of days to maturity (DM), 100-seed weight (SW), and average seed yield per plot (SY). Variance components and genetic and non-genetic parameters were estimated via the REML/BLUP and Bayesian methods. The variance components estimated and the breeding values and genetic gains predicted with selection through the Bayesian procedure were similar to those obtained by REML/BLUP. The frequentist and Bayesian MTME models provided higher estimates of broad-sense heritability per plot (or heritability of total effects of progeny; $h^2_{prog}$) and mean accuracy of progeny than their respective single-trait versions. Bayesian analysis provided the credibility intervals for the estimates of $h^2_{prog}$. Therefore, MTME led to greater predicted gains from selection. On this basis, this procedure can be efficiently applied in the genetic selection of segregating soybean progeny.

## Introduction

Soybean [*Glycine max* (L.) Merrill] is the fourth most widely grown crop in the world. This species is originally from China and is the major crop in the USA, Brazil, Argentina, and many

other countries [1]. Soybean is currently grown from low to high latitudes, where it is used as a source of oil, protein, biodiesel, etc. [2]. In this scenario, because the genotype × environment (G×E) interaction plays an essential role in genotypic expression, it must be considered in the evaluation and selection of superior genotypes [3–5].

The selection of segregating soybean progeny is a rather complex process because the traits of agronomic importance (e.g., maturity, yield, etc.) are of quantitative nature. Furthermore, because some traits are correlated with each other, selection based on one of them leads to alterations in others [6,7]. In soybean breeding, most of the available methods for the selection of progeny or lines are useful in the analysis of a single trait measured either in a single environment [8,9] or in various environments with the incorporation of the G×E interaction [10–12]. However, researchers often face situations in which multiple traits are measured across multiple environments [13]. Moreover, selection bias may arise when genetic selection is performed based on two or more genetically correlated traits (due to pleiotropism, imbalance in the gametic phase, and/or the common influence of the environment) and these are analyzed individually. This is especially true in sequential selection [13–15].

To reduce selection bias, Henderson *et al.* [16] proposed the multiple-trait BLUP method. An additional advantage of jointly modeling multiple traits compared to analyzing each trait separately is that the inference process appropriately accounts for the correlation between the traits, which helps to increase prediction accuracy, statistical power, and parameter estimation accuracy [14,16,17]. Despite the improvements provided by the multiple-trait BLUP (Best Linear Unbiased Prediction) method, to the best of our knowledge, there are few studies combining multi-trait models under a multi-environment approach.

The multiple-trait BLUP procedure is ideal, as it makes it possible to simultaneously analyze traits that are correlated with each other and that exhibit covariance heterogeneity across experimented environments [18,19]. In this way, a covariance structure is applied to each random factor in the model; e.g., the progeny effects, the G×E interaction effects, and the residual effects [20,21]. Even though data collected in plant breeding studies often present a multi-trait multi-environment structure, as they take into consideration the genetic correlations and the G × E interaction, more complex models are required, rendering the computation process more laborious. Some studies have demonstrated the potential of the Bayesian approach for genetic evaluation in plant breeding considering multi-trait or multi-environment models [22–25]. In this approach, the parameters are interpreted as random variables, following the law of probability, which assumes a priori knowledge [25,26].

In view of the above-described situation, the present study proposes to examine the efficiency and applicability of multi-trait multi-environment (MTME) models in the selection of segregating soybean progeny, using phenotypic data, by the frequentist (FMTME) and Bayesian (BMTME) methodologies. Effective biometric tools were exploited and compared with the usual tools employed in genetic breeding aiming at increasing the predicted genetic gain. In this way, study demonstrated it was possible to attain these goals by selecting soybean genotypes with better genetic potentials (desirable phenotypic traits) in different evaluation environments.

## Material and methods

### Experimental data

Three populations (Pop) belonging to the Soybean Breeding Program at the Federal University of Viçosa (UFV) were obtained from crosses between divergent inbred lines (Pop1: TMG123RR/M7211RR; Pop2: UFVSCitrinoRR/UFVSTurquezaRR; and Pop3: M7908RR/M7211RR). These lines were classified into different relative maturity groups according to the

soybean crop management classification [27], aiming to exploit genetic variability for the selection of productive progeny. Seventy-two, 71, and 60 $F_2$ plants of populations 1, 2, and 3, respectively, were separately bulk-harvested and threshed. One sample was collected from each $F_{2:3}$ progeny to compose the 203 $F_{2:4}$ progeny that were used in this study by the within-progeny bulk method [28,29].

To evaluate these progeny, two trials were conducted in the 2016/2017 crop year at the Teaching, Research, and Academic-Extension Units at UFV. One of them took place in Capinópolis—MG, Brazil (18˚40'48" S latitude, 49˚33'58" W longitude; 530 m altitude) and the other in Viçosa—MG, Brazil (20º45'45" S latitude, 42º49'27" W longitude; 647 m altitude).

The experiments were set up as a randomized complete block design with three replicates per site. Plots consisted of two 3.0-m rows spaced 0.5 m apart, with a plant stand density of 13 seeds per meter, totaling a density of 256,000 plants ha$^{-1}$. All plant management operations were undertaken in accordance with the requirements of the crop in the region [30].

Three target agronomic traits were evaluated, namely: number of days to maturity (DM, days), 100-seed weight (SW, g), and average seed yield per plot (SY, g). The first variable, DM, corresponds to the number of days before 95% of the pods were mature, as indicated by their color [31]. To evaluate the SW and SY traits, the grains were dried to 13% moisture. All data and code to run the main models used in this study are available in S1 Table and S2 Table, respectively.

## Frequentist statistical analyses

The Restricted Maximum Likelihood/Best Linear Unbiased Prediction (REML/BLUP) procedure was adopted for statistical analyses under a frequentist approach (Patterson and Thompson [32] and Henderson [33]). The frequentist single-trait multi-environment (FSTME) statistical model associated with the evaluation of segregating progeny in a randomized block design in two environments, with one observation per plot, is given by:

$$y = Xb + Zg + Wi + e,$$

where $y$ is the vector of phenotypes; $b$ is the vector of block effects added to the overall mean (assumed fixed); $g$ is the vector of progeny effects (assumed random), in which $g \sim N(0, \sigma_g^2)$; $i$ is the vector of the G×E interaction effects (random), in which $i \sim N(0, \sigma_{int}^2)$; and $e$ is the vector of residuals (random), in which $e \sim N(0, \sigma_e^2)$. The capital letters ($X$, $Z$ and $W$) represent the incidence matrices for the effects of $b$, $g$, and $i$, respectively. The $b$ vector encompasses all replicates of all locations. In this case, this vector comprises the effects of locations and of replicates within locations. The goodness of fit was obtained using Akaike information criteria (AIC [34]), defined by AIC = $-2LogL + 2p$, in which $LogL$ is the log-likelihood function and $p$ is the number of estimated parameters, according to Little $et$ $al.$ [35]; and the Likelihood Ratio Test (LRT), following Wilks [36], using Chi-square Statistics with one degree of freedom, which is calculated by the following equation: $\lambda = 2[Log_eL_{p+1} - Log_eL_p]$, in which $L_{p+1}$ and $L_p$ are the maximum likelihood associated with the full and the reduced models, respectively.

In the frequentist multi-trait multi-environment (FMTME) approach, this model was transformed into $g \sim N(0, \Sigma_g \otimes I)$, $i \sim N(0, \Sigma_{int} \otimes I)$ and $e \sim N(0, \Sigma_e \otimes I)$, where $\Sigma_g$ is the progeny covariance matrix; $\Sigma_{int}$ is the G×E interaction covariance matrix; $\Sigma_e$ is the residual covariance matrix; $I$ is an identity matrix with order appropriate to the respective random effect; and $\otimes$ denotes the Kronecker product. $\Sigma_g$, $\Sigma_{int}$, and $\Sigma_e$ are also unstructured covariance structures (US) [18,19]. The following $Y_1$, $Y_2$, and $Y_3$ are the vectors of observed responses for each trait

(DM, SW, and SY) and (co)variance matrix estimates:

$$Y = \begin{bmatrix} Y_1 \\ \cdots \\ Y_2 \\ \cdots \\ \vdots \\ Y_3 \end{bmatrix}; \sum_g = \begin{bmatrix} \sigma_{g1}^2 & \sigma_{g1g2}^2 & \sigma_{g1g3}^2 \\ \sigma_{g2g1}^2 & \sigma_{g2}^2 & \sigma_{g2g3}^2 \\ \sigma_{g3g1}^2 & \sigma_{g3g2}^2 & \sigma_{g3}^2 \end{bmatrix}; \sum_{int} = \begin{bmatrix} \sigma_{i1}^2 & \sigma_{i1i2}^2 & \sigma_{i1i3}^2 \\ \sigma_{i2i1}^2 & \sigma_{i2}^2 & \sigma_{i2i3}^2 \\ \sigma_{i3i1}^2 & \sigma_{i3i2}^2 & \sigma_{i3}^2 \end{bmatrix} \text{ and } \sum_e = \begin{bmatrix} \sigma_{e1}^2 & \sigma_{e1e2}^2 & \sigma_{e1e3}^2 \\ \sigma_{e2e1}^2 & \sigma_{e2}^2 & \sigma_{e2e3}^2 \\ \sigma_{e3e1}^2 & \sigma_{e3e2}^2 & \sigma_{e3}^2 \end{bmatrix}$$

To perform the statistical analyses of the FSTME and FMTME models and obtain the variance components and breeding values, we applied the ASReml 4.1 package [21] of integrated R software (Development Core Team—[37]).

## Bayesian statistical analyses

The single-trait multi-environment and multi-trait multi-environment models as well as the covariance matrix structures (US) of the frequentist analyses were used according to the Bayesian approach (BSTME and BMTME, respectively) via Markov Chain Monte Carlo (MCMC) to estimate the variance components and genetic parameter. To gather further information about representation of model, matrix, and vectors structures as well as priori probability distributions, we recommend reading Mrode [18] and Gilmour *et al.* [21] and recent publications Junqueira *et al.* [23], Torres *et al.* [25] and Mora and Serra [38].

We assumed that the variance-covariance matrices follow an Inverse-Wishart (IW) distribution and independent Inverse-Gamma (IG) distributions, which were used as a priori to model the variance-covariance matrix [39–41]. The ensuing covariance matrix distribution is such that all standard-deviation parameters have Half-t distributions and the correlation parameters have uniform distributions on (-1,1) for a particular choice of the IW shape parameter [13]. The advantage of this approach is that it allows us to choose the shape and scale parameters that achieve high non-arbitrary information of all standard deviations and correlation parameters [40].

The full (considering the genotype and G × E interaction effects) models were compared with the reduced (disregarding the genotype or G × E interaction effects) models by the deviance information criterion (DIC) proposed by Spiegelhalter *et al.* [42]: $DIC = D(\bar{\theta}) + 2p_D$, where $D(\bar{\theta})$ is a point estimate of the deviance obtained by replacing the parameters with their posterior mean estimates in the likelihood function and $p_D$ is the effective number of parameters in the model. Models with lower DIC should be preferred over models with higher DIC.

The Bayesian models (BSTME and BMTME) were implemented in the "MCMCglmm" package [43,44] of R software (R Development Core Team—[37]). A total of 1,000,000 samples were generated, and assuming a burn-in period and sampling interval of 500,000 and 5 iterations, respectively, this resulted in a final total of 100,000 samples. The convergence of MCMC was checked by the criterion of Geweke [45], which was performed using the "boa" [46] and "CODA" (Convergence Diagnosis and Output Analysis) [47] R packages.

## Genetic and non-genetic components

Broad-sense heritability per plot (or heritability of total effects from progeny) for the frequentist and Bayesian models were computed based on the approximated estimators, as discussed

in Piepho *et al.* [48], using the following expression:

$$h^2_{prog} = \frac{\sigma^2_g}{\sigma^2_g + \frac{\sigma^2_{int}}{n} + \frac{\sigma^2_e}{nr}}$$

where $\sigma^2_g$: variance of progeny; $\sigma^2_{int}$: variance of the progeny $\times$ environment interaction; $\sigma^2_e$: error variance; $n$: number of locations; and $r$: number of replicates. For the Bayesian models, the posterior estimates were calculated from the posterior samples of the variance components obtained by the model.

The accuracy of progeny selection for the frequentist models (FSTME and FMTME) was estimated based on the following expression [49]:

$$r_{\hat{g}g} = \sqrt{(\sigma^2_g - PEV)/\sigma^2_g},$$

where $\sigma^2_g$ is the genotypic variance and *PEV* is the prediction error variance extracted from the diagonal of the generalized inverse of the coefficient matrix of the mixed-model equations.

For the Bayesian approach (BSTME and BMTME models), the accuracy of progeny selection was estimated according to Resende *et al.* [19] from the posterior distribution, given by:

$$\tilde{r}_{\hat{g}g} = [1 - s(\tilde{g})/\tilde{g}],$$

where $s(\tilde{g})$ is the standard deviation of the predicted breeding value. According to Resende *et al.* [19], in Bayesian inference, the variance of the very parameter that is assumed as a random variable is computed.

The "boa" [46] and "bayesplot" [50,51] packages of R software were used to calculate and plot the highest posterior density (HPD) intervals for all parameters, respectively. Estimates of the coefficient of experimental variation (*CVe*) and selective accuracy ($r_{\hat{g}g}$) were used to evaluate the experimental quality of the models [52].

## Genetic correlation

To determine the genetic covariance by the frequentists and Bayesian single-trait models (FSTME and BSTME, respectively), a pairwise analysis of the sum of phenotypic values of the traits was performed. Thus, the covariances were obtained, as proposed by Resende *et al.* [53], using the following expression:

$$\sigma_{g(trait_i, trait_j)} = \frac{\sigma^2_{g(trait_i + trait_j)} - \sigma^2_{g(trait_i)} - \sigma^2_{g(trait_j)}}{2},$$

where $\sigma^2_{g(trait_i + trait_j)}$ is the variance of the sum of phenotypic values of traits $i$ and $j$; $\sigma^2_{g(trait_i)}$ is the genotypic variance of trait $i$; and $\sigma^2_{g(trait_j)}$ is the genotypic variance of trait $j$. Genetic covariances by the multi-trait multi-environment models (FMTME and BMTME models) were obtained directly by the mixed-model output from each applied methodology.

The genetic correlation coefficients between the DM, SW, and SY traits were obtained, as suggested by Piepho *et al.* [54], using the expression below for all models:

$$\rho_{(trait_i, trait_j)} = \frac{\sigma_{g(trait_i, trait_j)}}{\sqrt{\sigma^2_{g(trait_i)} \sigma^2_{g(trait_j)}}}$$

## Progeny selection

The Spearman rank correlation coefficient was calculated between the BLUP (breeding value) of the progeny from the analyses of FSTME, FMTME, BSTME, and BMTME, and its significance was verified using nonparametric bootstrap in the R package 'boot' [55,56]. The agreement between the selected progeny was also checked by the coincidence index (CI) proposed by Hamblin and Zimmermann [57], as shown below:

$$CI = \frac{(A - C)}{(M - C)},$$

where $A$ is the number of coincident progeny in the two methods, $M$ is the number of selected progeny, and $C$ is the number of progeny coincident due to chance ($C = bM$, where $b$ is the selection intensity = 0.15; i.e., 15%).

Selection gain was predicted for each trait (DM, SW, and SY) based on the expression below:

$$SG = \frac{\sum_{i=1}^{n} GV_i}{n},$$

where $GV_i$ is the predicted genotypic value of progeny $i$ and $n$ is the number of selected progeny (30).

In order to perform a simultaneous selection and infer about the efficiency of selection gain for each evaluated trait between the frequentist (FSTME and FMTME) and Bayesian (BSTME and BMTME) models, we applied the additive genetic index using Selegen REML/BLUP software (AGI–[58]). The predicted breeding values of the selected progeny were thus compared by the frequentist and Bayesian approaches. The weights for each trait were defined based on the coefficients of genetic variation [59]. For all traits, the progeny were selected to increase the phenotypic expression, or provide the highest BLUP. After the direction of selection was defined, the genotypic values for each progeny [weighted by the pre-established weights ($CV_g$) for each trait] were summed, generating the AGI value. Subsequently, they were organized in descending order.

## Results

### Analysis of deviance and model fitting

The significance of progeny (G) and G×E interaction effects of the FSTME model were evaluated. Significant G and G×E interaction effects (P ≤ 0.01) were detected in the LRT for the DM, SW, and SY traits (Table 1). According to the AIC from the results obtained with the FSTME model, the model including the G and G×E interaction effects (full model) showed the best fit (lowest AIC value) for all traits (Table 2). Thus, according to the two methodologies,

**Table 1. Deviance and likelihood ratio test (LRT) for number of days to maturity (DM), 100-seed weight (SW) (grams), and average seed yield per plot (SY) (grams) evaluated in 203 soybean $F_{2:4}$ progeny for single-trait multi-environment (FSTME) frequentist analysis.**

| Effect | DM | | SW | | SY | |
|---|---|---|---|---|---|---|
| | Deviance | LRT | Deviance | LRT | Deviance | LRT |
| Progeny (G) | 7048.74 | 141.17** | 2398.04 | 94.15** | 5837.97 | 15.68** |
| G×E interaction | 7010.14 | 102.57** | 2362.57 | 58.68** | 5853.38 | 31.09** |
| Full model | 6907.57 | | 2303.89 | | 5822.29 | |

** Significant at the 0.01 probability level according to the chi-square test.

https://doi.org/10.1371/journal.pone.0215315.t001

**Table 2. Akaike information criteria for the full model and deviance information criteria for the full (considering genotype and G × E interaction effects) and reduced [disregarding the genotype (−*prog*) and interaction (−int) effects] models for number of days to maturity (DM), 100-seed weight (SW) (grams), and average seed yield per plot (SY) (grams) via frequentist and Bayesian single-trait multi-environment (FSTME and BSTME) and multi-trait multi-environment (FMTME and BMTME) models.**

| Model | Trait | Akaike information criteria (AIC) | | |
| --- | --- | --- | --- | --- |
| | | Full model | Reduced model1 (−*prog*) | Reduced model2 (−int) |
| FMTME | DM, SW, SY | 14861.22 | 15184.53 | 15051.87 |
| FSTME | DM | 6913.569 | 7052.743 | 7014.143 |
| FSTME | SW | 2309.892 | 2402.042 | 2366.568 |
| FSTME | SY | 5828.291 | 5841.969 | 5857.377 |
| | | Deviance information criteria (DIC) | | |
| BMTME | DM, SW, SY | 20577.3 | 20694.71 | 21019.89 |
| BSTME | DM | 8700.467 | 8741.243 | 8938.965 |
| BSTME | SW | 4157.547 | 4198.951 | 4303.262 |
| BSTME | SY | 7890.409 | 7900.89 | 7968.122 |

the full FSTME model was the most suitable to estimate the genetic parameters and predict the genotypic values. Likewise, the FMTME model was also appropriate, as it showed the lowest AIC value. In the Bayesian models, all chains achieved convergence by the criterion of Geweke [45]. Overall, the DIC values were smaller when using the full Model (considering genotype and G × E interaction effects), being the difference in relation to full Model higher than 2 (Table 2), which according to Spiegelhalter *et al.* [42] it's enough to suggest that the use of full Model can lead to higher accuracy in estimating the parameters (Table 2). Therefore, since this model component is important the "best" genotypes measured in different environments couldn't the same.

### Genetic and non-genetic components

The estimates of genotypic variance ($\sigma_g^2$) and G×E interaction variance ($\sigma_{int}^2$) were higher and lower, respectively, when obtained via FMTME and BMTME, for all traits. By contrast, the residual variance ($\sigma_{res}^2$) and individual phenotypic variance ($\sigma_{phen}^2$) estimates were similar regardless of the procedure used for the frequentist models, whereas in the Bayesian models these estimates were higher for all traits, even for the genotypic component ($\sigma_g^2$). However, the $\sigma_{res}^2$ and $\sigma_{phen}^2$ estimates were similar between the BSTME and BMTME models. As a consequence, both the frequentist and Bayesian MTME models led to higher estimates of $h_{prog}^2$, $CV_g$, and $Ac_{prog}$ compared with the frequentists and Bayesian single-trait models. The exception was the SW trait, for which no significant differences were observed (Table 3) between single- and multi-trait multi-environment models. For the other components, both procedures led to results proportional to the estimates of $\sigma_g^2$ and $\sigma_{int}^2$.

Compared with FSTME, the FMTME model provided increases of the orders of 1.81, 0.22, and 21.72% in the estimate of $\sigma_g^2$ and reductions of 6.63, 0.26, and 17.61% in the estimate of $\sigma_{int}^2$ for the respective traits DM, SW, and SY. As a result, the $h_{prog}^2$ estimates increased by 1.03, 0.1, and 15.94%; $CV_g$, by 0.88, 0.14 and 10.28%; and $Ac_{prog}$, by 4.38, 0.25, and 42.29% for the respective traits.

Compared with BSTME, the BMTME model led to 1.56, 0.02, and 31.81% higher estimates of $\sigma_g^2$ and 8.56, 0.4, and 32.57% lower estimates of $\sigma_{int}^2$ for the respective traits DM, SW, and SY. The reduced ranging of $h_{prog}^2$ (lower and upper difference) with Bayesian MTME credible intervals (probability of 95%) were 9.66, 0.22, and 42.44% for the DM, SW, and SY traits,

**Table 3. Estimates of variance components and genetic and non-genetic parameters for number of days to maturity (DM), 100-seed weight (SW) (grams), and average seed yield per plot (SY) (grams) evaluated in 203 soybean $F_{2:4}$ progeny via frequentist single-trait multi-environment (FSTME) and multi-trait multi-environment (FMTME) and Bayesian single-trait multi-environment (BSTME) and multi-trait multi-environment (BMTME) models.**

| Component | FSTME | | | FMTME | | | BSTME | | | BMTME | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | DM | SW | SY | DM | SW | SY | DM | SW | SY | DM | SW | SY |
| $\sigma^2_g$ | 128.625 | 1.8045 | 8.6171 | 130.947 | 1.8084 | 10.4882 | 129.598 | 1.8158 | 8.3801 | 131.616 | 1.8161 | 11.0459 |
| $\sigma^2_{int}$ | 33.9228 | 0.6244 | 10.0727 | 31.6741 | 0.6228 | 8.2995 | 34.2204 | 0.6278 | 10.2834 | 31.2917 | 0.6253 | 6.9344 |
| $\sigma^2_{res}$ | 56.5253 | 1.5008 | 36.4566 | 56.5234 | 1.5006 | 36.4419 | 56.7499 | 1.5081 | 36.6665 | 57.1089 | 1.5152 | 37.5989 |
| $\sigma^2_{phen}$ | 219.073 | 3.9297 | 55.1464 | 219.144 | 3.9318 | 55.2297 | 220.569 | 3.9518 | 55.3299 | 220.017 | 3.9566 | 55.5792 |
| $h^2_{prog}$ | 0.8298 | 0.7624 | 0.4368 | 0.8383 | 0.7631 | 0.5064 | Table 4[a] | Table 4[a] | Table 4[a] | Table 4[a] | Table 4[a] | Table 4[a] |
| $C^2_{int}$ | 0.1548 | 0.1589 | 0.1827 | 0.1445 | 0.1584 | 0.1503 | 0.1552 | 0.1588 | 0.1859 | 0.1422 | 0.158 | 0.1248 |
| $CV_g$ | 0.0791 | 0.0729 | 0.141 | 0.0798 | 0.073 | 0.1555 | 0.0794 | 0.0732 | 0.1390 | 0.0801 | 0.0732 | 0.1596 |
| $CV_e$ | 0.0524 | 0.0665 | 0.2901 | 0.0524 | 0.0665 | 0.2899 | 0.0525 | 0.0666 | 0.2908 | 0.0527 | 0.0668 | 0.2945 |
| $Ac_{prog}$ | 0.9087 | 0.8691 | 0.6557 | 0.9485 | 0.8713 | 0.933 | 0.9633 | 0.6337 | 0.7367 | 0.9733 | 0.6355 | 0.8941 |

$\sigma^2_g$: genotypic variance; $\sigma^2_{int}$: genotype × environment interaction variance; $\sigma^2_{res}$: residual variance; $\sigma^2_{phen}$: individual phenotypic variance; $h^2_{prog}$: mean broad-sense heritability per plot; $C^2_{int}$: coefficient of determination of the genotype × environment interaction effects; $CV_g$: genotypic coefficient of variation; $CV_e$: residual coefficient variation; and $Ac_{prog}$: mean accuracy of progeny.

[a]Results showed in Table 4.

respectively (Table 4). Thus, the mean increased by 1.08, 0.00, and 24.01%; $CV_g$ increased by 0.88, 0.14, and 10.82%; and $Ac_{prog}$ increased by 1.04, 0.28, and 21.37% for the respective traits.

The posterior density intervals of heritability genetic parameters (Table 4, and Figs 1–3) were accessed to assist in the selection of genotypes in the Bayesian models. Thus, the breeding values and their HPD intervals obtained from the Bayesian STME and MTME models for each trait can be useful tools in progeny selection. Posterior density intervals of estimates of genotypic variance for BSTME and BMTME were showed in S1–S4 Figs.

## Genetic correlation

Genetic correlations between the DM, SW, and SY traits obtained by the frequentist and Bayesian models are given in Table 5. For the DM-SW and SW-SY pairs, low correlations were detected in every comparison, indicating absence of linear associations. However, a high and positive correlation was found between DM and SY in both methodologies. Additionally, the correlation between this pair of traits estimated via FSTME and BSTME exceeded the parameter space. The same was not true when it was estimated via FMTME and BMTME, which showed similar values within the parameter space and were thus more realistic (not biased).

**Table 4. Posterior inferences for mode, mean, median, and higher posterior density (HPD) interval of the broad-sense heritability per plot, considering the proposed Bayesian single-trait multi-environment (BSTME) and multi-trait multi-environment (BMTME) models for number of days to maturity (DM), 100-seed weight (SW) (grams), and average seed yield per plot (SY) (grams).**

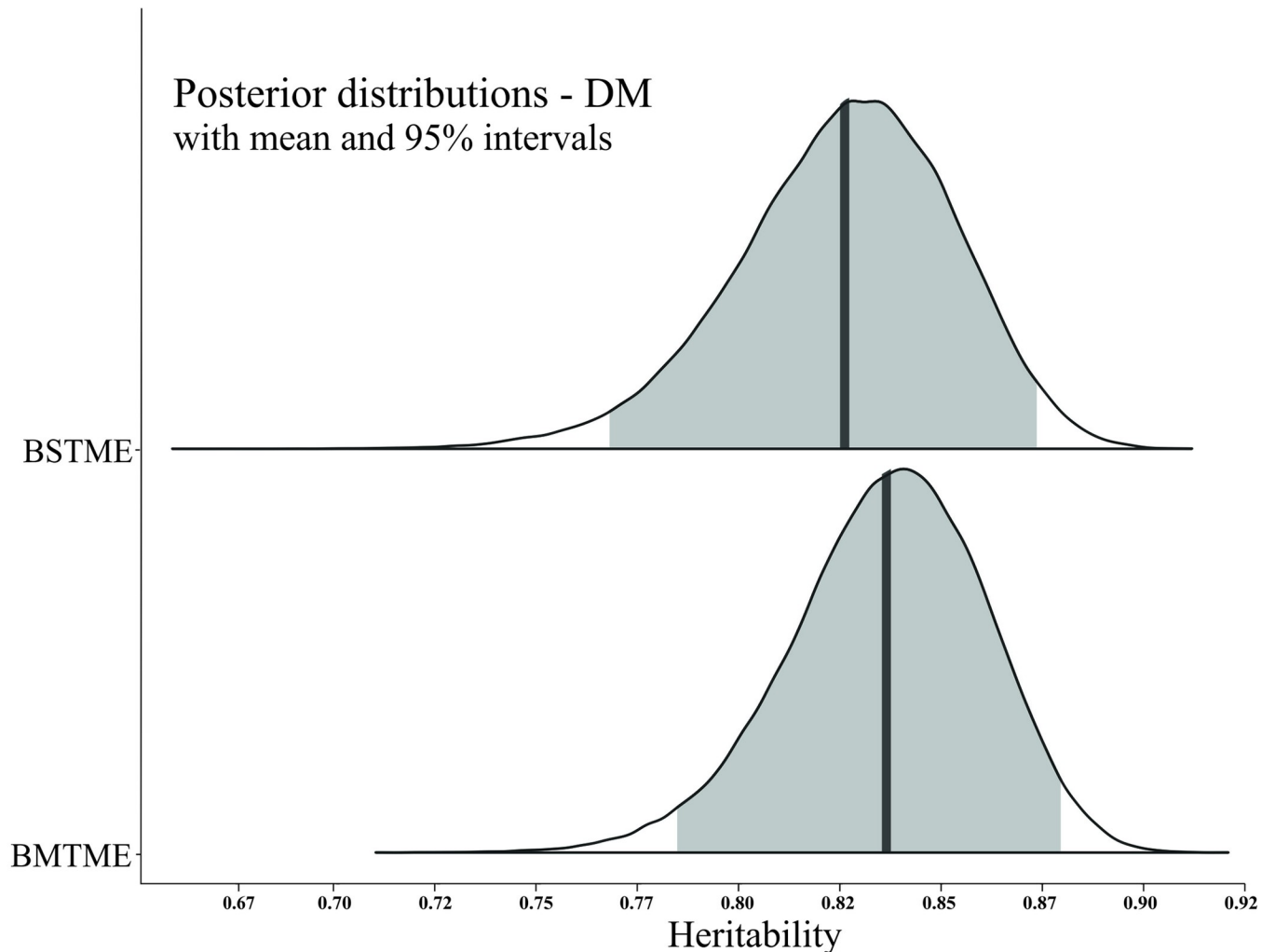| Trait | BSTME | | | | | BMTME | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | $h^2_{prog}$ | | | HPD (95%) | | $h^2_{prog}$ | | | HPD (95%) | |
| | Mode | Mean | Median | Lower | Upper | Mode | Mean | Median | Lower | Upper |
| DM | 0.8354 | 0.8282 | 0.8298 | 0.7798 | 0.874 | 0.842 | 0.8375 | 0.8388 | 0.7936 | 0.8787 |
| SW | 0.765 | 0.76 | 0.7628 | 0.6905 | 0.8247 | 0.775 | 0.765 | 0.7628 | 0.6908 | 0.8247 |
| SY | 0.44 | 0.4204 | 0.4282 | 0.2406 | 0.586 | 0.5361 | 0.5292 | 0.531 | 0.4304 | 0.6292 |

**Fig 1.** Posterior density for the Bayesian single-trait multi-environment (BSTME) (top) and multi-trait multi-environment (BMTME) (bottom) models of the broad-sense heritability per plot for number of days to maturity (DM). The solid color represents the posterior density of 95% intervals and the solid vertical line indicates the mean.

### Progeny selection

The Spearman rank correlations between breeding values via single- and multi-trait multi-environment models (FSTME and FMTME; BSTME and BMTME) were significant in all comparisons. This correlation was medium for SY (79.16 and 70.92) in the frequentist and Bayesian models, respectively, and high for the other comparisons. These results were confirmed by the agreement (Table 6).

In the frequentist and Bayesian analyses, the predicted selection gains were equivalent for all traits. However, the MTME models showed greater gains. For the SW trait, both the frequentist and the Bayesian procedures generated very similar results (Table 6). For DM and SY, however, there was less agreement between the selected progeny, especially for the SY trait, which culminated in greater discrepancy between the gains predicted from selection.

For the DM trait, the MTME models led to increased gains predicted from selection: 4.01 and 5.13% for the frequentist and Bayesian methodologies, respectively. Considering the SY trait, for which the agreement between the selected progeny was lower than 50%, the MTME
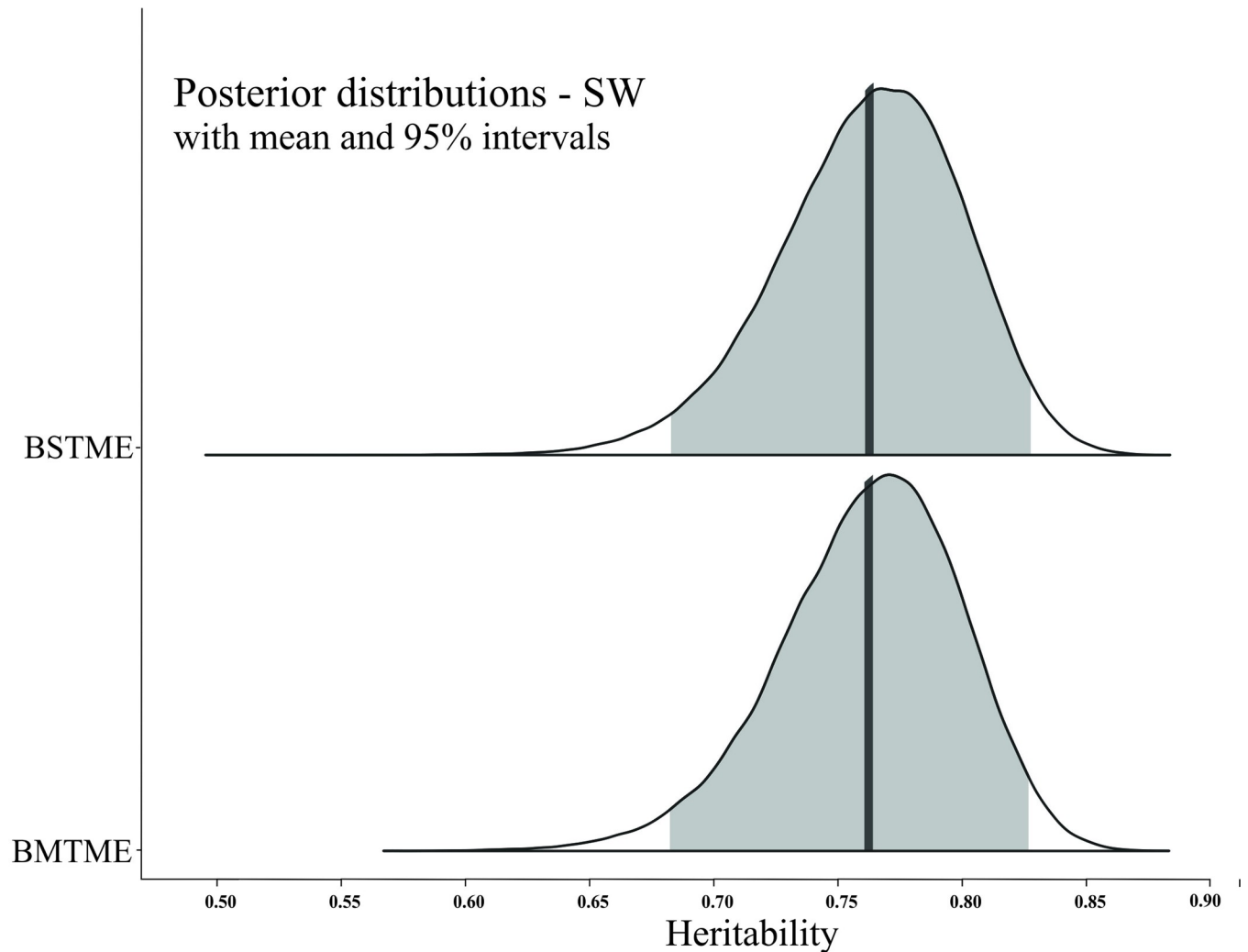
**Fig 2.** Posterior density for the Bayesian single-trait multi-environment (BSTME) (top) and multi-trait multi-environment (BMTME) (bottom) models of the broad-sense heritability per plot for 100-seed weight (SW; grams). The solid color represents the posterior density of 95% intervals and the solid vertical line indicates the mean.

https://doi.org/10.1371/journal.pone.0215315.g002

models showed increases of 20.26% and 28.43% (frequentist and Bayesian models, respectively) in the gain predicted from selection, as compared with their respective single-trait models (Table 6).

When the AGI was used for the simultaneous selection of the 30 best soybean progeny, higher index gains were found for the FMTME and BMTME models (Tables 7 and 8) compared with the gain of the overall mean of AGI. The gains predicted from selection were similar for both models (FMTME and BMTME), for all traits. Moreover, greater gains were observed for the SY variable (15.03 and 15.71% for FMTME and BMTME, respectively).

## Discussion

### Analysis of deviance

The LRT for the FSTME model revealed that the progeny and G×E interaction effects are significant (P < 0.01) for the DM, SW, and SY traits. Consequently, the respective variance
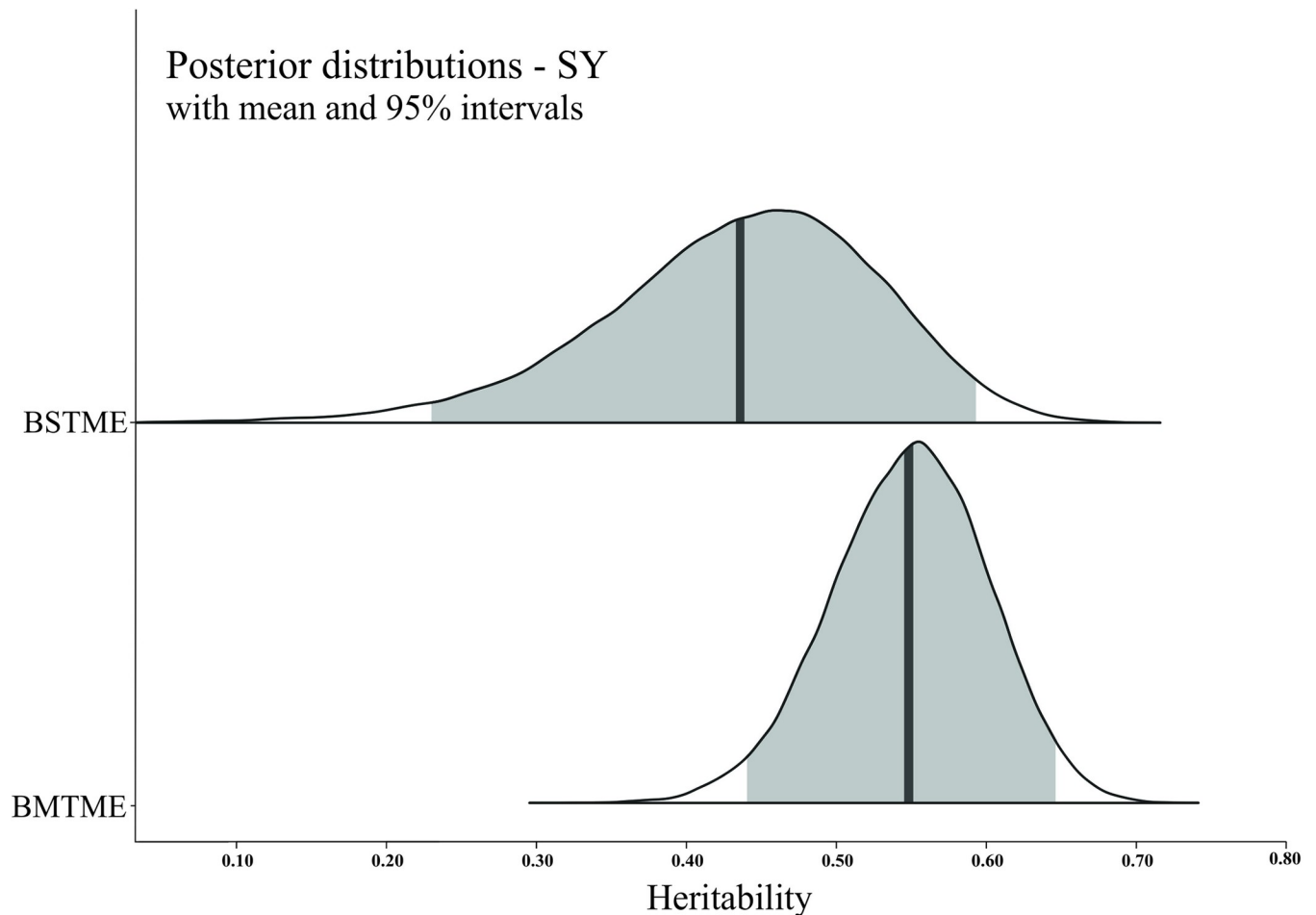
**Fig 3.** Posterior density for the Bayesian single-trait multi-environment (BSTME) (top) and multi-trait multi-environment (BMTME) (bottom) models of the broad-sense heritability per plot for average seed yield per plot (SY; grams). The solid color represents the posterior density of 95% intervals and the solid vertical line indicates the mean.

components are significantly different from zero and so are the respective coefficients of determination (Table 1). The fit of the frequentist models was checked by AIC. This criterion indicated the full model as the most suitable to estimate the variance components and predict the genotypic values (Table 2). For the Bayesian approach, the full (considering genotype and G × E interaction effects) and reduced (only genotype or G x E interactions effects) models were compared through DIC (Deviance Information Criterion), which suggests that models

**Table 5.** Genetic correlations between number of days to maturity (DM), 100-seed weight (SW) (grams), and average seed yield per plot (SY) (grams) evaluated in 203 soybean $F_{2:4}$ progeny via frequentist single-trait multi-environment (FSTME) and multi-trait multi-environment (FMTME) and Bayesian single-trait multi-environment (BSTME) and multi-trait multi-environment (BMTME) models.

| Correlation | FSTME | FMTME | BSTME | BMTME |
|---|---|---|---|---|
| $\rho$ (DM,SW) | −0.1090 | −0.1107 | −0.1048 | −0.1078 |
| $\rho$ (DM,SY) | ¶ | 0.9718 | ¶ | 0.9724 |
| $\rho$ (SW,SY) | −0.0153 | −0.0217 | −0.0164 | −0.0296 |

¶: value higher than unity

**Table 6. Predicted selection gain, agreement, and Spearman rank correlation in the selection of the 30 best soybean progeny via frequentist (FSTME and FMTME) and Bayesian (BSTME and BMTME) models for number of days to maturity (DM), 100-seed weight (SW; grams), and average seed yield per plot (SY; grams) evaluated in 203 soybean $F_{2:4}$ progeny.**

| Trait | Predicted selection gain | | | | Agreement (%) | |
|---|---|---|---|---|---|---|
| | FSTME | FMTME | BSTME | BMTME | FSTME × FMTME | BSTME × BMTME |
| DM | 12.70 | 13.22 | 12.67 | 13.32 | 93.30 (0.9349[b]) | 93.31 (0.9289) |
| SW | 1.92 | 1.93 | 1.92 | 1.92 | 100.00 (0.9963) | 100.00 (0.9965) |
| SY | 3.11 | 3.74 | 2.99 | 3.84 | 46.40 (0.7916) | 46.60 (0.7992) |

[b] Spearman rank correlation

https://doi.org/10.1371/journal.pone.0215315.t006

with smaller DIC are better supported by the data. According to Spiegelhalter *et al*. [42], models with differences in DIC values lower than 2 need to be considered as equally well. Therefore, since DIC values obtained were higher than 2, it is possible to indicate the superiority of full model over the restrict models.

The generalization of AIC is the most common method of assessing the fit of a statistical model estimated via Bayesian inference (DIC). The effects of Bayesian models can be used as an inference to the test of hypothesis [19]. In Bayesian statistics, the lowest expected deviance has the HPD [60], and this was observed in the present study for the MTME model (Tables 2 and 3).

In both criteria (AIC and DIC) for the choice of statistical models, the obtained results revealed that the MTME models showed the best fit, explaining the genetic variability of the experiment and selection considering the genetic (progeny) and environmental interaction effects (Table 2).

## Variance components

Variance components are the variances associated with the random effects of a model. Knowing them is of great importance in genetics and breeding, since the population and the breeding method to be used depend on information that can be obtained from these components. The solution of mixed-model equations depends on knowledge of the variance and covariance matrix, whose structure is known, but its components often are not. At present, the standard method for the estimation of variance components is REML, developed by Patterson and Thompson [32].

The BLUP method [33] maximizes the correlation between the predicted and the true genotypic value; i.e., it minimizes the prediction error variance (PEV). Additionally, it is not biased, as we expect the predicted genotypic value to be equal to the true genotypic value [61]. Further, BLUP allows for the simultaneous use of several sources of information as well as information originating from experiments carried out in one or various locations and evaluated in one or various harvests [62].

Although the mixed-model methodology by the frequentist approach has several desirable characteristics [49], the adoption of Bayesian statistical inference for genetic evaluation in the breeding of crop species has shown to be advantageous. Bayesian models have been used since 1986 [63] and further exploited in recent years [23–25,64,65] due to the great computational advancements and new methodological applications and elucidations.

Bayesian analysis is based on the knowledge of the posterior distribution of the parameters to be estimated. This allows for the construction of exact credibility intervals for the estimates of random variables, variance components, and fixed effects [66]. Higher values for the interval with 95% credibility of distribution for the *broad-sense heritability* parameter found in this study (Table 4) were also presented in the study of Torres *et al*. [25] to estimate genetic

**Table 7. Order, progeny (Prog), breeding value (u+g), and Additive Genetic Index (AGI) of the 30 progeny selected simultaneously via frequentist single-trait multi-environment (FSTME) and multi-trait multi-environment (FMTME) models for number of days to maturity (DM), 100-seed weight (SW; grams), and average seed yield per plot (SY; grams) evaluated in 203 soybean $F_{2:4}$ progeny.**

| Order | Prog | FSTME Trait | | | AGI | Prog | FMTME Trait | | | AGI |
|---|---|---|---|---|---|---|---|---|---|---|
| | | DM | SW | SY | | | DM | SW | SY | |
| 1 | 521 | 156.92 | 20.66 | 23.70 | 31.39 | 521 | 158.31 | 20.54 | 25.49 | 29.79 |
| 2 | 267 | 148.06 | 20.36 | 26.51 | 31.13 | 550 | 157.50 | 20.22 | 25.15 | 29.47 |
| 3 | 550 | 157.05 | 20.34 | 22.93 | 31.02 | 267 | 155.95 | 20.22 | 25.06 | 29.33 |
| 4 | 556 | 152.91 | 19.98 | 22.83 | 30.44 | 556 | 153.87 | 19.90 | 24.09 | 28.80 |
| 5 | 235 | 152.08 | 19.13 | 25.31 | 30.38 | 235 | 157.97 | 19.00 | 25.25 | 28.72 |
| 6 | 545 | 149.86 | 20.16 | 23.07 | 30.36 | 537 | 151.56 | 20.14 | 23.46 | 28.67 |
| 7 | 537 | 151.66 | 20.20 | 22.17 | 30.33 | 545 | 151.65 | 20.08 | 23.55 | 28.66 |
| 8 | 554 | 151.38 | 20.32 | 21.83 | 30.31 | 562 | 157.65 | 19.03 | 24.86 | 28.66 |
| 9 | 520 | 153.18 | 19.77 | 22.49 | 30.24 | 554 | 150.68 | 20.27 | 23.21 | 28.65 |
| 10 | 262 | 154.56 | 18.17 | 26.44 | 30.20 | 262 | 162.68 | 17.99 | 26.43 | 28.63 |
| 11 | 56 | 154.15 | 19.80 | 21.79 | 30.18 | 520 | 153.64 | 19.69 | 23.96 | 28.63 |
| 12 | 562 | 158.58 | 19.12 | 21.98 | 30.14 | 56 | 153.21 | 19.74 | 23.79 | 28.60 |
| 13 | 568 | 157.47 | 19.60 | 20.81 | 30.10 | 568 | 154.41 | 19.55 | 23.98 | 28.59 |
| 14 | 272 | 154.70 | 19.24 | 22.65 | 30.06 | 567 | 153.70 | 19.61 | 23.72 | 28.54 |
| 15 | 515 | 149.72 | 20.63 | 20.35 | 30.04 | 272 | 155.49 | 19.16 | 24.37 | 28.49 |
| 16 | 266 | 147.93 | 19.03 | 25.45 | 30.01 | 515 | 146.55 | 20.62 | 22.04 | 28.36 |
| 17 | 567 | 158.44 | 19.63 | 19.86 | 29.98 | 215 | 155.10 | 18.97 | 24.35 | 28.33 |
| 18 | 215 | 151.38 | 19.07 | 23.96 | 29.97 | 569 | 155.41 | 18.91 | 24.13 | 28.28 |
| 19 | 546 | 145.02 | 20.03 | 22.67 | 29.80 | 266 | 154.80 | 18.91 | 24.39 | 28.28 |
| 20 | 571 | 154.15 | 19.50 | 20.77 | 29.76 | 269 | 157.11 | 18.49 | 24.68 | 28.23 |
| 21 | 269 | 156.09 | 18.58 | 22.69 | 29.75 | 571 | 151.64 | 19.46 | 23.22 | 28.20 |
| 22 | 569 | 158.02 | 18.97 | 20.90 | 29.75 | 51 | 155.26 | 18.80 | 24.05 | 28.19 |
| 23 | 514 | 151.80 | 19.49 | 21.60 | 29.75 | 239 | 155.65 | 18.60 | 24.34 | 28.13 |
| 24 | 531 | 151.52 | 19.69 | 20.96 | 29.71 | 514 | 151.12 | 19.44 | 23.15 | 28.13 |
| 25 | 239 | 153.60 | 18.69 | 23.08 | 29.70 | 531 | 149.66 | 19.65 | 22.74 | 28.09 |
| 26 | 51 | 158.16 | 18.83 | 20.61 | 29.61 | 546 | 146.99 | 19.98 | 22.25 | 28.02 |
| 27 | 212 | 150.14 | 18.09 | 25.59 | 29.60 | 524 | 152.57 | 19.01 | 23.47 | 28.02 |
| 28 | 516 | 150.42 | 19.58 | 21.20 | 29.60 | 544 | 149.87 | 19.51 | 22.71 | 28.01 |
| 29 | 524 | 153.04 | 19.07 | 21.70 | 29.59 | 512 | 154.70 | 18.58 | 23.97 | 27.99 |
| 30 | 544 | 152.91 | 19.54 | 20.36 | 29.59 | 516 | 149.23 | 19.54 | 22.62 | 27.96 |
| Mean of sel. | | 153.16 | 19.51 | 22.54 | 30.08 | | 153.80 | 19.45 | 23.95 | 28.48 |
| Overall mean | | 143.45 | 18.43 | 20.82 | 28.21 | | 143.45 | 18.43 | 20.82 | 26.49 |
| Gain (%) | | 6.77 | 5.85 | 8.27 | 6.64 | | 7.21 | 5.55 | 15.03 | 7.50 |

parameters for N-uptake efficiency and N-utilization efficiency under contrasting N levels in the soil via BMTME models. The difference between mean, mode, and median of broad-sense heritability estimates (Table 4) reflects some lack of symmetry in the posterior distribution estimates [38]. However, for the SY trait, differences between the BSTME and BMTME are clear when we analyze the posterior densities, mainly because the posterior MTME resulted in a more narrow and symmetric distribution, confirming the increase in precision (Fig 3). When the prior distribution is informative, the credibility interval tends to be narrower than the confidence intervals. When the mixed-model parameters are assigned non-informative distributions, Bayesian and frequentist inferences should be equivalent [67].

**Table 8. Order, progeny (Prog), breeding value (u+g), and Additive Genetic Index (AGI) of the 30 progeny selected simultaneously via Bayesian single-trait multi-environment (BSTME) and multi-trait multi-environment (BMTME) models for number of days to maturity (DM), 100-seed weight (SW; grams), and average seed yield per plot (SY; grams) evaluated in 203 soybean $F_{2:4}$ progeny.**

| Order | Prog | BSTME Trait | | | AGI | Prog | BMTME Trait | | | AGI |
|---|---|---|---|---|---|---|---|---|---|---|
| | | DM | SW | SY | | | DM | SW | SY | |
| 1 | 521 | 156.86 | 20.65 | 23.60 | 31.23 | 521 | 158.49 | 20.53 | 25.58 | 29.96 |
| 2 | 267 | 148.06 | 20.37 | 26.28 | 30.95 | 550 | 157.61 | 20.22 | 25.23 | 29.64 |
| 3 | 550 | 157.02 | 20.34 | 22.85 | 30.86 | 267 | 156.48 | 20.22 | 25.23 | 29.50 |
| 4 | 556 | 152.90 | 19.98 | 22.75 | 30.28 | 556 | 154.00 | 19.90 | 24.16 | 28.96 |
| 5 | 235 | 152.08 | 19.13 | 25.14 | 30.21 | 235 | 158.30 | 19.01 | 25.41 | 28.90 |
| 6 | 545 | 149.87 | 20.16 | 22.99 | 30.20 | 537 | 151.64 | 20.13 | 23.49 | 28.83 |
| 7 | 537 | 151.62 | 20.19 | 22.12 | 30.17 | 562 | 157.59 | 19.03 | 24.92 | 28.82 |
| 8 | 554 | 151.38 | 20.32 | 21.80 | 30.16 | 545 | 151.83 | 20.08 | 23.62 | 28.82 |
| 9 | 520 | 153.16 | 19.77 | 22.42 | 30.09 | 262 | 163.15 | 18.01 | 26.68 | 28.81 |
| 10 | 56 | 154.13 | 19.81 | 21.74 | 30.03 | 554 | 150.73 | 20.26 | 23.24 | 28.80 |
| 11 | 262 | 154.55 | 18.18 | 26.24 | 30.03 | 520 | 153.73 | 19.69 | 24.01 | 28.79 |
| 12 | 562 | 158.54 | 19.12 | 21.94 | 29.98 | 56 | 153.22 | 19.74 | 23.82 | 28.75 |
| 13 | 568 | 157.48 | 19.60 | 20.81 | 29.95 | 568 | 154.28 | 19.54 | 23.99 | 28.75 |
| 14 | 272 | 154.68 | 19.25 | 22.58 | 29.90 | 567 | 153.49 | 19.60 | 23.70 | 28.70 |
| 15 | 515 | 149.69 | 20.63 | 20.37 | 29.90 | 272 | 155.53 | 19.16 | 24.43 | 28.65 |
| 16 | 567 | 158.42 | 19.63 | 19.89 | 29.84 | 515 | 146.49 | 20.60 | 22.00 | 28.50 |
| 17 | 266 | 147.92 | 19.03 | 25.27 | 29.84 | 215 | 155.30 | 18.97 | 24.45 | 28.50 |
| 18 | 215 | 151.37 | 19.07 | 23.84 | 29.81 | 266 | 155.18 | 18.92 | 24.55 | 28.44 |
| 19 | 546 | 145.02 | 20.04 | 22.60 | 29.64 | 569 | 155.26 | 18.91 | 24.15 | 28.44 |
| 20 | 571 | 154.10 | 19.50 | 20.77 | 29.61 | 269 | 157.13 | 18.50 | 24.76 | 28.40 |
| 21 | 569 | 157.98 | 18.97 | 20.90 | 29.60 | 571 | 151.54 | 19.45 | 23.22 | 28.35 |
| 22 | 514 | 151.77 | 19.49 | 21.55 | 29.59 | 51 | 155.06 | 18.80 | 24.06 | 28.35 |
| 23 | 269 | 156.07 | 18.59 | 22.61 | 29.59 | 239 | 155.75 | 18.60 | 24.43 | 28.30 |
| 24 | 531 | 151.51 | 19.69 | 20.96 | 29.56 | 514 | 151.13 | 19.44 | 23.18 | 28.28 |
| 25 | 239 | 153.57 | 18.69 | 22.99 | 29.54 | 531 | 149.61 | 19.65 | 22.74 | 28.24 |
| 26 | 51 | 158.15 | 18.83 | 20.62 | 29.46 | 524 | 152.56 | 19.01 | 23.51 | 28.17 |
| 27 | 516 | 150.40 | 19.57 | 21.19 | 29.45 | 546 | 147.20 | 19.99 | 22.29 | 28.16 |
| 28 | 524 | 153.04 | 19.07 | 21.66 | 29.44 | 544 | 149.77 | 19.50 | 22.70 | 28.16 |
| 29 | 544 | 152.89 | 19.54 | 20.36 | 29.44 | 512 | 154.69 | 18.58 | 24.03 | 28.15 |
| 30 | 212 | 150.11 | 18.09 | 25.41 | 29.43 | 212 | 157.80 | 17.97 | 25.11 | 28.13 |
| Mean of sel. | | 153.14 | 19.51 | 22.47 | 29.93 | | 154.15 | 19.40 | 24.09 | 28.64 |
| Overall mean | | 143.45 | 18.43 | 20.82 | 28.06 | | 143.45 | 18.43 | 20.82 | 26.63 |
| Gain (%) | | 6.76 | 5.86 | 7.95 | 6.64 | | 7.46 | 5.26 | 15.71 | 7.56 |

Mathew *et al.* [68] showed that Bayesian inference is superior to frequentist inference when the posterior distribution of a variance component is bimodal. Mathew *et al.* [68], Waldmann *et al.* [69], Schenkel *et al.* [70], and Harville *et al.* [71] did not find relevant differences between the breeding values predicted by frequentist or Bayesian approach. Schenkel *et al.* [70] also observed that the breeding values presented the same bias and accuracy. Silva *et al.* [72] found results from noninformative analyses and results from REML/BLUP analyses (frequentist) for some components of variance and heritability and for breeding values. The specific results obtained by the frequentist and Bayesian approaches were similar (Table 3). This was expected,

since non-informative prior distributions were used in Bayesian analysis. The modes of the marginal posterior distributions of the genetic parameters were similar to the corresponding REML estimates. From the Bayesian point-of-view, the estimates obtained via REML correspond to the modes of the combined posterior distributions of the variance components, obtained by Bayesian approach, given the use of uniform priors for the fixed effects and variance components [66].

The frequentist and Bayesian MTME models provided higher $\sigma_g^2$ estimates and lower $\sigma_{int}^2$ estimates, which resulted in higher $h_{prog}^2$ for all evaluated traits. The highest $h_{prog}^2$ estimates were found for DM and SW and the lowest for SY, which confirms that DM and SW are less complex traits and are thus less influenced by the environment than SY [73–75].

The genotypic coefficient of variation ($CV_g$) quantifies the magnitude of genetic variation available for selection, and thus high values are desirable [76]. In this way, the increase seen in this parameter with the use of the MTME models is important for breeding programs. The residual coefficient of variation ($CV_e$) is a measurement of experimental precision of statistical and non-genetic nature. According to Resende and Duarte [52], $CV_e$ is of moderate magnitude for the SY trait and low magnitude for the DM and SW traits, indicating good experimental precision. Moreover, as expected, there were no alterations in the $CV_e$ estimates when the FSTME and FMTME models were used (Table 3).

### Genetic correlations

Studies on genotypic, phenotypic, and environmental correlations in soybean involve traits that are evaluated from flowering to maturity; notably, yield and its components [77–80]. Our results corroborate those reported by Cober *et al.* [81], who also obtained a high correlation between DM and SY and no linear associations between the DM-SW and SW-SY pairs. The authors argued that the genes controlling maturity in soybean have pleiotropic effects with grain yield. Ablett *et al.* [82] and Lee *et al.* [83] reported that late maturity was associated with high yield. Liu *et al.* [84] investigated the genetic architecture of three growth period traits and confirmed that the soybean growth stages are highly correlated with grain yield. Li *et al.* [85] and Zhang *et al.* [75] observed that the association between the BARC-016957-02165 marker and seed yield was located in the same region as a QTL controlling pod maturity on chromosome 6, which explains the high correlation observed between these traits, as the QTL were closer to each other.

According to Pollak *et al.* [14], selection biases may occur when traits are analyzed individually. This bias was observed in the present study, where the genetic correlation value between the DM and SY traits exceeded the parameter space (value higher than 1) (Table 5). Viana *et al.* [15] evaluated two traits in popcorn and also found that the genetic correlation obtained with the single-trait model exceeded the parameter space.

According to Thompson and Meyer [86], the increase in accuracy obtained with the use of multi-trait BLUP analysis compared with single-trait analysis is proportional to the difference between the genetic and environmental correlations of the analyzed traits. In the context of whole-genome prediction, Jia *et al.* [87], Guo *et al.* [88], and Jiang *et al.* [22] found that joint prediction of multiple traits benefits from genetic correlations 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. This fact was observed in our study, in which the DM variable showed high heritability and high correlation with SY, consequently generating significant increases in selection accuracy for SY. However, for both methodologies—frequentist and Bayesian—there was no significant increase in selection accuracy for the SW trait, as verified by its low correlation with the other evaluated traits.

## Progeny selection

The observed differences in the genotypic values predicted by the Bayesian and BLUP/REML procedures were small, leading to a slight alteration in the ranking of the progeny selected by both procedures. This finding was confirmed by Rank Spearman correlation (Table 6). However, compared with the FSTME and BSTME models, the respective MTME models showed higher genetic gains for DM and, especially, for SY. According to Resende *et al.* [66], these are the conditions for there to be correspondence between the frequentist and Bayesian methodologies for the fixed and random effect parameters: attribution of non-informative priori for the fixed effects, normal priori for the random effects, and normal likelihood for the observations vector. These promises were used in the present study, which explains the obtained results.

Despite the high agreement between the progeny selected for the DM and SW traits by both procedures, there was little agreement for the SY trait, which resulted in greater gains predicted from selection via MTME. Piepho *et al.* [3] and Piepho *et al.* [62] recommended the use of multiple-trait models to predict breeding values in annual crops, because this procedure has the best statistical properties and provides more-accurate results.

Resende *et al.* [19] and Okeke *et al.* [89] stated that one of the main advantages of using multivariate models is higher selection accuracy. Okeke *et al.* [89] also reported that multi-environment models were useful for understanding G×E interactions. Higher $Ac_{prog}$ were observed when obtained via MTME for all traits, with SY standing out with 42.29 and 21.37% increases in the frequentist and Bayesian models, respectively. Greater accuracy and efficiency of multiple-trait models were also reported by Viana *et al.* [15] in selection among and within half-sib families.

However, it must be stressed that the BSTME model obtained superior accuracy in comparison with the FSTME model for DM and SY, despite the similar broad-sense heritability values. This is explained by the use of the estimator of selection accuracy. In this regard, Resende *et al.* [19] described that when Bayesian accuracy is higher than frequentist accuracy, the distribution of the parameters attributed to Bayesian approach were probably more adequate than those associated with the traditional model. The opposite can be considered true for the SW variable, for which the frequentist models obtained better results due to a better adjustment of the normal distribution of the parameters attributed to the data. These conclusions are also valid for the MTME models, which exhibited different obtained accuracies; however, for the SY variable, the FMTME model showed the best fit according to the mean accuracy of the progeny (Table 3).

As can be seen in Tables 7 and 8, desirable gains are obtained in selecting the best progeny for the DM, SW, and SY variables for all models based on the AGI. However, the high positive correlation between the DM and SY traits can favor the selection of high-yielding and late-cycle progeny. In this case, selection indices can help breeders select progeny that exhibit gains for both traits simultaneously [90]. Although similar gains were found for both approaches employed, it should be stressed that there was a slight increase in gain (%) when the 30 best progeny were selected for the DM and SY traits using Bayesian approach. This is a desirable factor that should be taken into account by breeders.

According to Silva *et al.* [1], individual plants or progeny in the $F_{2:3}$ or $F_{2:4}$ generations can be selected aiming at the adoption of the recurrent selection method described by Hallauer *et al.* [91], by exploiting genetic variability among and within progeny. Silva *et al.* [1] stated that the $F_{2:4}$ generation is suitable for selection, since 87.5% (1.75) of the total additive genetic variance $(2\sigma_a^2)$ that will be available in $F_\infty$ is already available in $F_{2:4}$. Thus, progeny selection in $F_{2:4}$ through a more precise method is relevant.

Early in the generation of the base populations of the soybean breeding programs, many populations are commonly obtained at the expense of the number of progeny to be evaluated;

i.e., the evaluation of future lines, be them for high or low heritability, is based on samples with a finite (small) number of progeny. Thus, Bayes' theorem is recommended for those situations, as it gives precise solutions to the problem of finite-size samples, because for each data set—large or small—there is an exact posterior distribution to draw inferences.

The MTME models provided better results than the single-trait models using frequentist and Bayesian approach. Therefore, the former procedures can be efficiently applied in the genetic selection of segregating soybean progeny. However, it is necessary to use an adequate statistical tool that provides algorithms and routines to efficiently perform the analyses. Though not necessarily easy, the use of Bayesian inference in quantitative genetics in the breeding of crop species [69,72] is a tendency in breeding programs [5].

## Conclusion remarks

For our data set, the average BMTME processing time using an Intel(R) i7-5500U (2.4 GHz) processor with 8 GB of RAM was 1 h 40 min and 35 s, corresponding to approximately 0.006 s for each MCMC iteration. Silva *et al.* [72] considered this performance plausible, but pointed out that improvements can be obtained using the conditional decompositions proposed by Hallander *et al.* [92]. For the same purpose, in addition to improving the prior information, Montesinos-López *et al.* [13] proposed a Bayesian model for analyzing multiple traits and multiple environments for the whole-genome prediction model. The authors also developed an R-software package that offers specialized and optimized routines to efficiently perform the analyses under the proposed model. By contrast, the FMTME model took approximately 14 s to converge. Despite the considerable difference in processing time of the analysis and output size of the results (around 1.03 GB) due to the high number of interactions adopted, the Bayesian model showed to be efficient for the proposed objective. Furthermore, it provided additional results to those obtained by the frequentist approach, with noteworthy credibility intervals.

The Bayesian models have desirable potentials when using informative prior distributions, providing parameters with lower standard deviations and/or possible genetic gains. However, the quality of the informative prior may have questionable origins and may not generate considerable advantages. Silva *et al.* [72] showed it can be advantageous to implement a Bayesian framework for mixed-model analysis in the breeding of crop species using informative priors. However, for potential future studies in plant breeding, the implementation of informative prior fitted to MTME models can be the next step to be assessed.

## Supporting information

**S1 Table. Data set necessary to replicate the findings of our research.**
(TXT)

**S2 Table. Scripts to run the multi-trait multi-environment models.** Data set necessary to replicate the findings of our research.
(TXT)

**S1 Fig. On the diagonal, posterior density of 95% intervals of the estimate of variance components via the Bayesian multi-trait multi-environment (BMTME) model; off the diagonal the traits' correlations by score for number of days to maturity (DM), 100-seed weight (SW; grams), and average seed yield per plot (SY; grams).**
(TIFF)

**S2 Fig. Posterior density for the Bayesian single-trait multi-environment model (BSTME) of the estimate of variance components for maturity (DM).** The solid color represents the

posterior distributions of 95% intervals and the solid vertical line indicates the mean for number of days.
(TIFF)

**S3 Fig. Posterior density for the Bayesian single-trait multi-environment model (BSTME) of the estimate of variance components for 100-seed weight (SW).** The solid color represents the posterior distributions of 95% intervals and the solid vertical line indicates the mean.
(TIFF)

**S4 Fig. Posterior density for the Bayesian single-trait multi-environment model (BSTME) of the estimate of variance components for average seed yield per plot (SY).** The solid color represents the posterior distributions of 95% intervals and the solid vertical line indicates the mean.
(TIFF)

## Author Contributions

**Conceptualization:** Leonardo Volpato, Marcos Deon Vilela de Resende, Moysés Nascimento.

**Data curation:** Leonardo Volpato, Felipe Lopes da Silva.

**Formal analysis:** Leonardo Volpato, Rodrigo Silva Alves.

**Funding acquisition:** Felipe Lopes da Silva, Aluízio Borém.

**Investigation:** Leonardo Volpato, Rodrigo Silva Alves, Paulo Eduardo Teodoro, Marcos Deon Vilela de Resende, Moysés Nascimento, Ana Carolina Campana Nascimento, Willian Hytalo Ludke.

**Methodology:** Rodrigo Silva Alves, Paulo Eduardo Teodoro, Marcos Deon Vilela de Resende, Moysés Nascimento, Ana Carolina Campana Nascimento.

**Project administration:** Felipe Lopes da Silva, Aluízio Borém.

**Software:** Leonardo Volpato, Rodrigo Silva Alves.

**Supervision:** Paulo Eduardo Teodoro, Moysés Nascimento, Ana Carolina Campana Nascimento, Willian Hytalo Ludke, Felipe Lopes da Silva, Aluízio Borém.

**Visualization:** Paulo Eduardo Teodoro, Marcos Deon Vilela de Resende, Moysés Nascimento, Ana Carolina Campana Nascimento, Willian Hytalo Ludke, Felipe Lopes da Silva, Aluízio Borém.

**Writing – original draft:** Leonardo Volpato, Rodrigo Silva Alves.

**Writing – review & editing:** Leonardo Volpato.

## References

1. Silva FL, Borém A, Sediyama T, Ludke WH. Soybean breeding. Soybean Breeding. 2017. https://doi.org/10.1007/978-3-319-57433-2

2. Liu X, Wu J, Ren H, Qi Y, Li C, Cao J, et al. Genetic variation of world soybean maturity date and geographic distribution of maturity groups. Breed Sci. 2017; 67: 221–232. https://doi.org/10.1270/jsbbs.16167 PMID: 28744175

3. Piepho HP, Williams ER. A comparison of experimental designs for selection in breeding trials with nested treatment structure. Theor Appl Genet. 2006; 113: 1505–1513. https://doi.org/10.1007/s00122-006-0398-8 PMID: 17028902

4. Kang MS, Gauch HG. Genotype -by- environment interaction. CRC Press; 1996.

5. Van Eeuwijk FA, Bustos-Korts D V., Malosetti M. What should students in plant breeding know about the statistical aspects of genotype x Environment interactions? Crop Sci. 2016; 56: 2119–2140. https://doi.org/10.2135/cropsci2015.06.0375

6. Pathan SM, Vuong T, Clark K, Lee JD, Grover Shannon J, Roberts CA, et al. Genetic mapping and confirmation of quantitative trait loci for seed protein and oil contents and seed weight in soybean. Crop Sci. 2013; 53: 765–774. https://doi.org/10.2135/cropsci2012.03.0153

7. Teng W, Feng L, Li W, Wu D, Zhao X, Han Y, et al. Dissection of the genetic architecture for soybean seed weight across multiple environments. Crop Pasture Sci. 2017; 68: 358–365. https://doi.org/10.1071/CP16462

8. Teixeira FG, Hamawaki OT, Hamawaki RL, Jorge GL, Hamawaki CL. Genetic parameters and selection of soybean lines based on selection indexes. Genet Mol Res. 2017; 16: 1–17.

9. Bizari EH, Val BHP, Pereira E de M, Mauro AO Di, Unêda-Trevisoli SH. Selection indices for agronomic traits in segregating populations of soybean. Rev Cienc Agron. 2017; 48: 110–117. https://doi.org/10.5935/1806-6690.20170012

10. Zhe Y, Lauer JG, Borges R, de Leon N. Effects of genotype × environment interaction on agronomic traits in soybean. Crop Sci. 2010; 50: 696–702. https://doi.org/10.2135/cropsci2008.12.0742

11. Pereira F, Bruzi AT, de Matos JW, Rezende BA, Prado LC, Nunes JAR. Implications of the population effect in the selection of soybean progeny. Plant Breed. 2017; 136: 679–687. https://doi.org/10.1111/pbr.12512

12. Volpato L, Simiqueli GF, Alves RS, Rocha JR do AS de C, Del Conte MV, Resende MDV de, et al. Selection of inbred soybean progeny (*Glycine max*): an approach with population effect. Plant Breed. 2018; 1–8. https://doi.org/10.1111/pbr.12648

13. Montesinos-López OA, Montesinos-López A, Crossa J, Toledo FH, Pérez-Hernández O, Eskridge KM, et al. A Genomic Bayesian Multi-trait and Multi-environment Model. G3&#58; Genes|Genomes|Genetics. 2016; 6: 2725–2744. https://doi.org/doi:10.1534/g3.116.032359

14. Pollak EJ, van der Werf J, Quaas RL. Selection Bias and Multiple Trait Evaluation. J Dairy Sci. 1984; 67: 1590–1595. https://doi.org/10.3168/jds.S0022-0302(84)81481-2

15. Viana JMS, Sobreira FM, De Resende MDV, Faria VR. Multi-trait BLUP in half-sib selection of annual crops. Plant Breed. 2010; 129: 599–604. https://doi.org/10.1111/j.1439-0523.2009.01745.x

16. Henderson CR, Quaas RL. Multiple trait evaluation using relatives records. J Anim Sci. 1976; 43: 1188–1197. https://doi.org/10.2527/jas1976.4361188x

17. Schaeffer LR. Sire and Cow Evaluation Under Multiple Trait Models. J Dairy Sci. Elsevier; 1984; 67: 1567–1580. https://doi.org/10.3168/jds.S0022-0302(84)81479-4

18. Mrode RA. Linear Models for the Prediction of Animal Breeding Values. 2014. https://doi.org/10.1017/CBO9781107415324.004

19. Resende MDV, Silva FF, Azevedo CF. Estatística Matemática, Biométrica e Computacional. 1st ed. Viçosa: Editora Rio Branco; 2014.

20. Malosetti M, Ribaut JM, van Eeuwijk FA. The statistical analysis of multi-environment data: Modeling genotype-by-environment interaction and its genetic basis. Front Physiol. 2013;4 MAR: 1–17. https://doi.org/10.3389/fphys.2013.00001

21. Gilmour R, Gogel BJ, Cullis BR, Welham SJ, Thompson R. ASReml User Guide Release 4.1 Structural Specification. VSN Int Ltd Hemel Hempstead. 2015; 1–30. https://doi.org/10.1017/CBO9781107415324.004

22. Jiang J, Zhang Q, Ma L, Li J, Wang Z, Liu JF. Joint prediction of multiple quantitative traits using a Bayesian multivariate antedependence model. Heredity (Edinb). 2015; 115: 29–36. https://doi.org/10.1038/hdy.2015.9 PMID: 25873147

23. Junqueira VS, Peixoto LA, Laviola BG, Bhering LL, Mendonça S, Costa TSA, et al. Bayesian multi-trait analysis reveals a useful tool to increase oil concentration and to decrease toxicity in Jatropha curcas L. PLoS One. 2016; 11: 1–14. https://doi.org/10.1371/journal.pone.0161046

24. Arriagada O, Mora F, Dellarossa JC, Ferreira MFS, Cervigni GDL, Schuster I. Bayesian mapping of quantitative trait loci (QTL) controlling soybean cyst nematode resistant. Euphytica. 2012; 186: 907–917. https://doi.org/10.1007/s10681-012-0696-y

25. Torres LG, Rodrigues MC, Lima NL, Horta Trindade TF, Fonseca e Silva F, Azevedo CF, et al. Multi-trait multi-environment Bayesian model reveals g x e interaction for nitrogen use efficiency components in tropical maize. PLoS One. 2018; 13: 1–15. https://doi.org/10.1371/journal.pone.0199492 PMID: 29949626

26. Hayashi T, Iwata H. A Bayesian method and its variational approximation for prediction of genomic breeding values in multiple traits. BMC Bioinformatics. 2013; 14. https://doi.org/10.1186/1471-2105-14-34 PMID: 23363272

**27.** Carpentieri-Pípolo V, De Almeida LA, De Souza Kiihl RA, Rosolem CA. Inheritance of long juvenile period under short day conditions for the BR80-6778 soybean (Glycine max (L.) Merrill) line. Euphytica. 2000; 112: 203–209. https://doi.org/10.1023/A:1003927817278

**28.** Ramalho MA, Patto Carvalho BL, Nunes JA, Rodrigues. Perspectives for the use of quantitative genetics in breeding of autogamous plants. ISRN Genet. 2013; 2013: 1–6. https://doi.org/10.5402/2013/718127

**29.** Ramalho MAP, Abreu AFB, Santos JB, Nunes JAR. Aplicações da Genética Quantitativa no Melhoramento de Plantas Autógamas. 1st ed. Lavras: UFLA; 2012.

**30.** Sediyama T, Silva F, Borém A. Soja do Plantio à Colheita. 1st ed. Viçosa: Editora UFV; 2015.

**31.** Fehr WR, Caviness CE. Stages of Soybean Development. Spec Rep. 1977; 80: 11.

**32.** Patterson HD, Thompson R. Recovery of inter-block information when block sizes are unequal. Biometrika. 1971; 58: 545–554. https://doi.org/10.1093/biomet/58.3.545

**33.** Henderson CR. Best linear unbiased estimation and prediction under a selection model published by: international biometric society stable. Biometrics. 1975; 31: 423–447. PMID: 1174616

**34.** Akaike H. A new look at the statistical model identification. IEEE Trans Autom Control. 1974; 19: 716–723. https://doi.org/10.1109/TAC.1974.1100705

**35.** Littell RC, Littell RC, Henry PR, Henry PR, Ammerman CB, Ammerman CB. Statistical analysis of repeated measures data using SAS procedures. J Anim Sci. 1998; 76: 1216–31. doi:/1998.7641216x PMID: 9581947

**36.** Wilks SS. The Large-Sample Distribution of the Likelihood Ratio for Testing Composite Hypotheses. Ann Math Stat. 1938; 9: 60–62. https://doi.org/10.1214/aoms/1177732360

**37.** R Core Team. R: The R Project for Statistical Computing. Vienna; 2018.

**38.** Mora F, Serra N. Bayesian estimation of genetic parameters for growth, stem straightness, and survival in Eucalyptus globulus on an Andean Foothill site. Tree Genet Genomes. 2014; 10: 711–719. https://doi.org/10.1007/s11295-014-0716-2

**39.** Tassell CP Van, Vleck LD Van. Multiple-trait Gibbs sampler for animal models: flexible programs for Bayesian and likelihood-based (co)variance component inference The online version of this article, along with updated information and services, is located on the World Wide Web at. 1996; 2586–2597.

**40.** Huang A, Wand MP. Simple marginally noninformative prior distributions for covariance matrices. Bayesian Anal. 2013; 8: 439–452. https://doi.org/10.1214/13-BA815

**41.** Villemereuil P. Estimation of a biological trait heritability using the animal model: how to use the MCMCglmm R package. Guide. 2012; 36. Available: http://devillemereuil.legtux.org/wp-content/uploads/2012/12/tuto_en.pdf

**42.** Spiegelhalter DJ, Best NG, Carlin BP, van der Linde A. Bayesian Measures of Model Complexity anf Fit. J R Stat Soc Ser B. Statistical Methodol. 2002; 64: 583–639. https://doi.org/10.1111/1467-9868.00353

**43.** Hadfield JD. MCMC methods for multi-response generalized linear mixed models: the MCMCglmm R package. J Stat Softw. 2010; 33: 1–22. https://doi.org/10.1002/ana.22635

**44.** Hadfield JD, Nakagawa S. General quantitative genetic methods for comparative biology: Phylogenies, taxonomies and multi-trait models for continuous and categorical characters. J Evol Biol. 2010; 23: 494–508. https://doi.org/10.1111/j.1420-9101.2009.01915.x PMID: 20070460

**45.** Geweke J. Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments. Bayesian Stat 4. 1992; 169–193. 1176289

**46.** Smith BJ. boa: An R Package for MCMC Output Convergence. J Stat Softw. 2007; 21: 1–37. http://dx.doi.org/10.18637/jss.v021.i11

**47.** Plummer M, Best N, Cowles K, Vines K. CODA: convergence diagnosis and output analysis for MCMC. R News. 2006; 6: 7–11. https://doi.org/10.1159/000323281

**48.** Piepho HP, Möhring J. Computing heritability and selection response from unbalanced plant breeding trials. Genetics. 2007; 177: 1881–1888. https://doi.org/10.1534/genetics.107.074229 PMID: 18039886

**49.** Henderson CR. Applications of Linear Models in Animal Breeding Models. Univ Guelph. 1984; 384. https://doi.org/10.1002/9780470316856.ch7

**50.** Gabry J, Simpson D, Vehtari A, Betancourt M, Gelman A. Visualization in Bayesian workflow. 2017; 2. Available: http://arxiv.org/abs/1709.01449

**51.** Gabry J, Mahr T, Burkner P-C, Modrák M, Barrett M. Plotting for Bayesian Models [Internet]. 2018. Available: http://mc-stan.org/bayesplot

**52.** Resende MDV de Duarte JB. Precisão E Controle De Qualidade Em Experimentos De Avaliação De Cultivares. Pesqui Agropecuária Trop (Agricultural Res Trop. 2007; 37: 182–194. https://doi.org/10.5216/pat.v37i3.1867

**53.** Resende MDV. Genética Quantitativa e de Populações. 1st ed. Viçosa: Visconde de Rio Branco; 2015.

**54.** Piepho HP. Allowing for the structure of a designed experiment when estimating and testing trait correlations. J Agric Sci. 2018; 156: 59–70. https://doi.org/10.1017/S0021859618000059

**55.** Canty A, Ripley B. Package 'boot'. Bootstrap Functions. CRAN R Proj. 2017;

**56.** Davison AC, Hinkley D V. Bootstrap Methods and their Application. Technometrics. 1997; 42: 216. https://doi.org/10.2307/1271471

**57.** Hamblin J, Zimmermann MJ de O. Breeding Common Bean for Yield in Mixtures [Internet]. Plant Breeding Reviews. Plant Breeding Reviews; 1986. pp. 04, 245–272. https://doi.org/10.1002/9781118061015.ch8

**58.** Resende MDV de. Software Selegen-REML / BLUP: a useful tool for plant breeding. Crop Breed Appl Biotechnol. 2016; 16: 330–339. http://dx.doi.org/10.1590/1984- 70332016v16n4a49

**59.** Bhering LL, Laviola BG, Salgado CC, Sanchez CFB, Rosado TB, Alves AA. Genetic gains in physic nut using selection indexes. Pesqui Agropecu Bras. 2012; 47: 402–408. https://doi.org/10.1590/S0100-204X2012000300012

**60.** Gelman A, Carlin JB, Stern HS, Rubin DB. Bayesian Data Analysis, Third Edition (Texts in Statistical Science). Book. 2014. https://doi.org/10.1007/s13398-014-0173-7.2

**61.** White TL, Hodge GR. Predicting breeding values with applications in forest tree improvement. Kluwer Academic Publishers; 1989.

**62.** Piepho HP, Möhring J, Melchinger AE, Büchse A. BLUP for phenotypic selection in plant breeding and variety testing. Euphytica. 2008; 161: 209–228. https://doi.org/10.1007/s10681-007-9449-8

**63.** Gianola D, Fernando RL. Bayesian Methods in Animal Breeding Theory. J Anim Sci. 1986; 63: 217–244. https://doi.org/10.2527/jas1986.631217x

**64.** 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; https://doi.org/10.1038/s41437-018-0109-7 PMID: 30120367

**65.** 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. https://doi.org/10.2135/cropsci2016.06.0496

**66.** 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

**67.** Sorensen D, Gianola D. Likelihood of Bayesian, and MCMC Methods in Quantitative Genetics. Dietz K, M G, K K I, Samet J, Tsiatis A, editors. Springer. New York: Springer-Verlag; 2002. https://doi.org/10.2135/cropsci2003.1574

**68.** Mathew B, Bauer AM, Koistinen P, Reetz TC, Léon J, Sillanpää MJ. Bayesian adaptive Markov chain Monte Carlo estimation of genetic parameters. Heredity (Edinb). 2012; 109: 235–245. https://doi.org/10.1038/hdy.2012.35 PMID: 22805656

**69.** Waldmann P, Hallander J, Hoti F, Sillanpää MJ. Efficient Markov chain Monte Carlo implementation of Bayesian analysis of additive and dominance genetic variances in noninbred pedigrees. Genetics. 2008; 179: 1101–1112. https://doi.org/10.1534/genetics.107.084160 PMID: 18558655

**70.** SCHENKEL FS SCHAEFFER LR, BOETTCHER PJ. Comparison between estimation of breeding values and fixed effects using Bayesian and empirical BLUP estimation under selection on parents and missing pedigree information. Genet Sel Evol. 2002; 34: 41–59. https://doi.org/10.1186/1297-9686-34-1-41 PMID: 11929624

**71.** Harville DA, Carriquiry AL. Classical and Bayesian Prediction as Applied to an Unbalanced Mixed Linear Model. Biometrics. 1992; 48: 987–1003. doi:2532693

**72.** Silva FF, Viana JMS, Faria VR, de Resende MDV. Bayesian inference of mixed models in quantitative genetics of crop species. Theor Appl Genet. 2013; 126: 1749–1761. https://doi.org/10.1007/s00122-013-2089-6 PMID: 23604469

**73.** Panthee DR, Pantalone VR, Saxton AM, West DR, Sams CE. Quantitative trait loci for agronomic traits in soybean. Plant Breed. 2007; 126: 51–57. https://doi.org/10.1111/j.1439-0523.2006.01305.x

**74.** Lee SH, Bailey MA, Mian MAR, Shipe ER, Ashley DA, Parrott WA, et al. Identification of quantitative trait loci for plant height, lodging, and maturity in a soybean population segregating for growth habit. Theor Appl Genet. 1996; 92: 516–523. https://doi.org/10.1007/BF00224553 PMID: 24166318

**75.** Zhang J, Song Q, Cregan PB, Nelson RL, Wang X, Wu J, et al. Genome-wide association study for flowering time, maturity dates and plant height in early maturing soybean (Glycine max) germplasm. BMC Genomics. ???; 2015; 16: 1–11. https://doi.org/10.1186/1471-2164-16-1

**76.** Johnson HW, Robinson HF, Comstock RE. Genotypic and Phenotypic Correlations in Soybeans and Their Implications in Selection1. Agronomy Journal. 1955. p. 477. https://doi.org/10.2134/agronj1955.00021962004700100008x

**77.** Akhter M, Sneller CH. Yield and yield components of early maturing soybean genotypes in Mid-South. Crop Sci. 1996; 36: 877–882.

**78.** Mansur L, Orf J, Chase K, Jarvik T, Cregan P, Lark K. Genetic mapping of agronomic traits using recombinant inbred lines of soybean. Crop Science. 1996. pp. 1327–1336.

**79.** Orf JH, Chase K, Jarvik T, Mansur LM, Cregan PB, Adler FR, et al. Genetics of soybean agronomic traits: I. Comparison of three related recombinant inbred populations. Crop Sci. 1999; 39: 1642–1651. https://doi.org/10.2135/cropsci1999.3961642x

**80.** Guzman PS, Diers BW, Neece DJ, St. Martin SK, LeRoy AR, Grau CR, et al. QTL associated with yield in three backcross-derived populations of soybean. Crop Science. 2007. pp. 111–122. https://doi.org/10.2135/cropsci2006.01.0003

**81.** Cober ER, Morrison MJ. Regulation of seed yield and agronomic characters by photoperiod sensitivity and growth habit genes in soybean. Theor Appl Genet. 2010; 120: 1005–1012. https://doi.org/10.1007/s00122-009-1228-6 PMID: 20012856

**82.** Ablett GR, Beversdorf WD, Dirks VA. Performance and Stability of Indeterminate and Determinate Soybean in Short-Season Environments. Crop Science. 1989. p. 1428. https://doi.org/10.2135/cropsci1989.0011183X002900060019x

**83.** Lee S, Jun TH, Michel AP, Rouf Mian MA. SNP markers linked to QTL conditioning plant height, lodging, and maturity in soybean. Euphytica. 2015; 203: 521–532. https://doi.org/10.1007/s10681-014-1252-8

**84.** Liu Z, Li H, Fan X, Huang W, Yang J, Li C, et al. Phenotypic characterization and genetic dissection of growth period traits in soybean (Glycine max) using association mapping. PLoS One. 2016; 11: 1–18. https://doi.org/10.1371/journal.pone.0158602 PMID: 27367048

**85.** Li W, Zheng D, Van K, Lee S. QTL mapping for major agronomic traits across two years in soybean (*Glycine max* L. Merr.). J Crop Sci Biotech. 2008; 11: 171–190.

**86.** Thompson R, Meyer K. A review of theoretical aspects in the estimation of breeding values for multi-trait selection. Livest Prod Sci. 1986; 15: 299–313. https://doi.org/10.1016/0301-6226(86)90071-0

**87.** Jia Y, Jannink JL. Multiple-trait genomic selection methods increase genetic value prediction accuracy. Genetics. 2012; 192: 1513–1522. https://doi.org/10.1534/genetics.112.144246 PMID: 23086217

**88.** Guo G, Zhao F, Wang Y, Zhang Y, Du L, Su G. Comparison of single-trait and multiple-trait genomic prediction models. BMC Genet. 2014; 15: 1–7. https://doi.org/10.1186/1471-2156-15-1

**89.** Okeke UG, Akdemir D, Rabbi I, Kulakow P, Jannink JL. Accuracies of univariate and multivariate genomic prediction models in African cassava. Genet Sel Evol. BioMed Central; 2017; 49: 1–10. https://doi.org/10.1186/s12711-016-0283-0

**90.** Bernardo R. Breeding for Quantitative Traits in Plants. 2nd Editio. Stemma Press. University of Minnesota–Twin Cities : Stemma Press; 2010.

**91.** Hallauer AR, Darrah LL. Compendium of recurrent selection methods and their application. CRC Crit Rev Plant Sci. 1985; 3: 1–33. https://doi.org/10.1080/07352688509382202

**92.** Hallander J, Waldmann P, Wang C, Sillanpää MJ. Bayesian inference of genetic parameters based on conditional decompositions of multivariate normal distributions. Genetics. 2010; 185: 645–654. https://doi.org/10.1534/genetics.110.114249 PMID: 20351218