Genetic evaluation and selection in *Jatropha curcas* through Frequentist and Bayesian inferences

Jeniffer Santana Pinto Coelho Evangelista[1] (iD), Marcos Antonio Peixoto[1] (iD), Igor Coelho[1] (iD), Rodrigo Alves[1] (iD), Marcos Deon Vilela Resende[2] (iD), Fabyano Fonseca e Silva[3] (iD), Bruno Laviola[4] (iD), Leonardo Lopes Bhering[1,*] (iD)

1. Universidade Federal de Viçosa ᴿᴼᴿ – Departamento de Biologia Geral – Viçosa (MG), Brazil.
2. Embrapa Florestas ᴿᴼᴿ – Pesquisa e Desenvolvimento – Colombo (PR), Brazil.
3. Universidade Federal de Viçosa ᴿᴼᴿ – Departamento de Zootecnia – Viçosa (MG), Brazil.
4. Embrapa Agroenergia ᴿᴼᴿ – Laboratório de Genética e Biotecnologia – Brasília (DF), Brazil.

ABSTRACT: An accurate and efficient statistical method for genetic evaluation is a key requirement for progress in any breeding program. Thus, the present study aimed to evaluate the performance of Frequentist and Bayesian inferences for repeated measures analysis in *Jatropha curcas* breeding. To this end, 730 individuals from 73 half-sib families were evaluated for grain yield trait, over six crop years. Frequentist and Bayesian analyses were made considering repeatability models with different residual variance structures. Variance components were estimated through restricted maximum likelihood (REML) and Markov Chain Monte Carlo (MCMC). Genetic values were predicted through best linear unbiased prediction (BLUP) and estimated through MCMC. Variance components and genetic and non-genetic parameters estimated by the Frequentist inference presented values similar to those estimated by the Bayesian inference. The selective accuracy presented high magnitude (0.84) by the Frequentist and Bayesian inferences, indicating high reliability. Confidence and highest posterior density (HPD) intervals were similar for the genetic parameters, however the HPD intervals range was slightly short. This study highlighted the importance of testing the residual variance structure and pointed out that the Frequentist and Bayesian inferences presented similar results when using non-informative prior. Then, both inferences can be efficiently applied in *Jatropha curcas* breeding.

Key words: restricted maximum likelihood, best linear unbiased prediction, Markov Chain Monte Carlo, repeated measures, genetic selection.

INTRODUCTION

Renewable energy sources have been the focus of numerous research in recent years (Bhering et al. 2015; Junqueira et al. 2016). Considering the renewable energy sources, the agricultural sector presents some crops, such as *Jatropha curcas* L., with great potential for bioenergy production (Nithiyanantham et al. 2012). The *J. curcas* oil presents relevant characteristics, in comparison with other vegetable oils, such as better stability to oxidation and lower viscosity (Bailis and McCarthy 2011). In view of this, *J. curcas* breeding programs are focused on creation and selection of genotypes with high productivity.

Repeated measures consist in phenotyping the same individual repeatedly over time and are common in perennial breeding (Ferreira et al. 2020). The repeated measures analysis presents own features due to the various crop years being correlated with each other and the possibility of occurrence of heterogeneous residual variance among crop years. The repeated measures analysis allows to identify the genetic and residual covariance structures over crop years

(Alves et al. 2018). Therefore, it is possible to comprehend the trait behavior over time and, consequently, do more accurate genetic selection.

Several statistical methods can be used for repeated measures analysis and support decision-making in perennials breeding programs (Resende et al. 2014). In this context, the Frequentist inference (restricted maximum likelihood – REML/best linear unbiased prediction – BLUP) (Patterson and Thompson 1971; Henderson 1974) has been commonly used in plant breeding (Resende 2016), providing good results for several crops, including *J. curcas* (Alves et al. 2018; Rodrigues et al. 2020; Peixoto et al. 2021). However, the Bayesian inference (Markov Chain Monte Carlo – MCMC) can be more informative and flexible for the genetic evaluation and selection, when compared with the REML/BLUP (Resende et al. 2014).

The Bayesian inference allows the analysis of the small samples, whereas the REML/BLUP procedure presents limitation to deal with this sort of data (Gianola and Fernando 1986). Further, the Bayesian inference not only takes the observed data into account, but also allows the input prior information, which can increase the selective accuracy (Hayashi and Iwata 2013). It has excellent theoretical and practical properties, which indicates that this method will become routine in genetics and breeding (Silva et al. 2013).

Thus, the present study aimed to evaluate the performance (in terms of variance components, genetic parameters, genetic values, selective accuracy, and selection gains) of Frequentist and Bayesian inferences for repeated measures analysis in *J. curcas* breeding.

MATERIAL AND METHODS

Experimental data

The experiment was implemented in November 2008, in a randomized complete block design, with two replications and five plants per plot, arranged in rows, spaced 4 m between rows and 2 m between plants. It was settled in the experimental field of Empresa Brasileira de Pesquisa Agropecuária (Embrapa) Agroenergia in Distrito Federal, Brazil (15°35'30"S and 47°42'30"W; 1,007 m a.s.l.). The experiment consisted of the performance evaluation of 730 individuals from 73 half-sib families regarding the grain yield trait, over six crop years (2010 to 2015 – M1 to M6). The management practices were based on Bahadur et al. (2012).

Statistical analyses

The repeatability model was determined by Equation 1:

$$y = Xm + Zg + Wp + Qi + Ts + e \tag{1}$$

in which: y = the vector of phenotypic data; m = the vector of measurement-replication combinations (fixed factor in the Frequentist context and systematic factor in the Bayesian context), which comprises the effects of measurement, replication and measurement × replication interaction, added to the overall mean; g = the vector of genetic effects (random factor) [$g \sim N(0, I\sigma_g^2)$ in which σ_g^2 is the genetic variance between progenies]; p = the vector of plot effects (random factor) [$p \sim N(0, I\sigma_p^2)$, in which σ_p^2 is the plot variance]; i = the vector of genotype × measurement (GM) interaction effects (random factor) $i \sim N(0, I\sigma_i^2$, in which σ_i^2 is the GM interaction variance]; s = the vector of permanent environmental effects (random factor) [$s \sim N(0, I\sigma_s^2$, in which σ_s^2 is the permanent environmental variance]; e = the vector of residuals (random factor) [($e \sim N(0, R)$, in which R is the residual variance matrix]. Capital letters (X, Z, W, Q, and T) = the incidence matrices for *m, g, p, i,* and *s*, respectively.

Different residual variance structures, identity (IDV) (homogeneous) and diagonal (DIAG) (heterogeneous) were tested. The IDV and DIAG residual variance structures were given as Eqs. 2 and 3:

$$IDV = \begin{bmatrix} \sigma_e^2 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sigma_e^2 \end{bmatrix} \tag{2}$$

$$DIAG = \begin{bmatrix} \sigma_{e\_M1}^2 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sigma_{e\_M6}^2 \end{bmatrix} \tag{3}$$

in which: $\sigma_{e\_}^2$ = the residual variance.

The phenotypic variance ($\hat{\sigma}_{phen}^2$), broad sense individual heritability (h_g^2), coefficient of determination of permanent environmental effects (c_s^2), coefficient of determination of GM interaction effects (c_i^2), coefficient of determination of plot effects (c_p^2), coefficient of determination of residual effects (c_{res}^2), individual repeatability (ρ), and genetic correlation across measurements (r_{gm}) were estimated, respectively, by Eqs. from 4 to 11 (Resende et al. 2014):

$$\hat{\sigma}_{phen}^2 = \hat{\sigma}_g^2 + \hat{\sigma}_p^2 + \hat{\sigma}_i^2 + \hat{\sigma}_s^2 + \hat{\sigma}_e^2 \tag{4}$$

$$h_g^2 = \frac{\hat{\sigma}_g^2}{\hat{\sigma}_{phen}^2} \tag{5}$$

$$c_s^2 = \frac{\hat{\sigma}_s^2}{\hat{\sigma}_{phen}^2} \tag{6}$$

$$c_i^2 = \frac{\hat{\sigma}_i^2}{\hat{\sigma}_{phen}^2} \tag{7}$$

$$c_p^2 = \frac{\hat{\sigma}_p^2}{\hat{\sigma}_{phen}^2} \tag{8}$$

$$c_e^2 = \frac{\hat{\sigma}_e^2}{\hat{\sigma}_{phen}^2} \tag{9}$$

$$\rho = \frac{\hat{\sigma}_g^2 + \hat{\sigma}_p^2 + \hat{\sigma}_s^2}{\hat{\sigma}_{phen}^2} \tag{10}$$

$$r_{gm} = \frac{\hat{\sigma}_g^2}{\hat{\sigma}_g^2 + \hat{\sigma}_i^2} \tag{11}$$

In which: n = the number of individuals per plot (five individuals); r = the number of replications (two replications).

Frequentist analysis

Variance components were estimated through REML (Patterson and Thompson 1971), and genetic values were predicted through BLUP (Henderson 1974), using the ASReml-R package (Butler et al. 2009). Repeatability models with homogeneous and heterogeneous residual variance structures were tested by the Akaike information criterion (AIC) (Akaike 1974) and by

the Bayesian information criterion (BIC) (Schwarz 1978). For the random effects of the statistical model (genotype, plot, GM interaction, and permanent environmental effects), significance was tested by the likelihood ratio test (LRT) (Wilks 1938).

Standard deviations of the estimates of the genetic and non-genetic parameters were obtained through the pin function of the ASReml-R software (Butler et al. 2009).

Bayesian analysis

Variance components and genetic values were estimated via Gibbs sampling through MCMC algorithm, using the *MCMCglmm* package (Hadfield 2010) in R (R Core Team 2019. A total of 1,100,000 interactions were generated, burn-in period of 100,000 and thin (sampling interval) of 10 iterations, totalizing 100,000 effective MCMC samples.

The convergence was checked by Geweke (1991) and Heidelberger and Welch (1983) criteria using the *boa* package (Smith 2007) in R (R Core Team 2019). Repeatability models with homogeneous and heterogeneous residual variance structures were tested by the deviance information criterion (DIC) (Spiegelhalter et al. 2002). DIC was also used to test the significance of the random effects of the statistical model (genotype, plot, GM interaction, and permanent environmental effects). Highest posterior density (HPD) intervals of the estimates of the genetic and non-genetic parameters were obtained through MCMC samples. As proposed by Resende and Alves (2020), different degrees of freedom (nu) (0.002, 0.02, 2, 4 and 7) were considered to test informative priors.

Selective accuracy

The selective accuracies ($r_{\hat{g}g_i}$) were estimated by the Eqs. 12 and 13 (Resende et al. 2014; Resende and Alves 2020):

$$r_{\hat{g}g_i} = \sqrt{1 - \frac{PEVi}{\hat{\sigma}_g^2}} \tag{12}$$

$$r_{\hat{g}g_i} = 1 - s(g_i)/g_i \tag{13}$$

in which: PEV = the prediction error variance; $s(g_i)$ = the standard deviation of the estimated genetic value (g_i).

Genetic selection

Genetic gains through Frequentist and Bayesian inferences were calculated from the predicted/estimated genetic values of the progenies. The selection gains (SG), in percentage, were obtained by Eq. 14:

$$SG\ (\%) = \left(\frac{X_s - X_o}{X_o}\right) x\ 100 \tag{14}$$

in which: X_s = the mean of the genetic values of the selected progenies; X_o = the overall mean.

A selection intensity of 27% (20 progenies) was considered.

RESULTS

Frequentist analysis

According to the AIC and BIC, the best repeatability model was that with heterogeneous residual variance (lowest AIC and BIC) (Table 1). Thus, this model was adopted to estimate the variance components and to predict the genetic values. In addition, all random effects of the statistical model were statistically significant by the LRT ($p < 0.05$) (Table 2).

Table 1. Akaike information criterion (AIC), Bayesian information criterion (BIC) and deviance information criterion (DIC), for repeatability models with homogeneous and heterogeneous residual variance structures.

Inference	Test	Residual variance	
		Homogeneous	**Heterogeneous**
Frequentist	AIC	-1,985.67	-3,992.85†
	BIC	-1,961.82	-3,945.17†
Bayesian	DIC	5,065.16	2,827.05†

†Selected model.

The broad sense individual heritability between progenies ranged from 0.10 (M6) to 0.34 (M1), whereas the repeatability coefficients ranged from 0.11 (M6) to 0.38 (M1) (Table 3). Furthermore, the genetic correlation through measures was equal to 0.36.

Table 2. Log likelihood (Log), likelihood ratio test (LRT), and deviance information criterion (DIC) for the repeatability model with heterogeneous residual variance structure.

Effect	Log	DIC	LRT
Genetic	1,972.08	2,828.16	68.69
GM interaction	1,528.91	3,065.35	955.03
Plot	1,959.27	4,629.35	94.32
Permanent environment	1,987.05	2,838.72	38.76
Full model	2,006.43	2,828.16	-

Table 3. Estimates of variance components and genetic and non-genetic parameters through Frequentist inference (REML) for the grain yield trait, evaluated in 730 individuals from 73 half-sib families of *Jatropha curcas*, in six harvests (M1 to M6). Values between parentheses refers to the standard deviations.

Component / Parameter	M1	M2	M3	M4	M5	M6
$\hat{\sigma}_g^2$			0.0568 (0.0132)			
$\hat{\sigma}_a^2$			0.0028 (0.0005)			
$\hat{\sigma}_s^2$			0.1001 (0.0091)			
$\hat{\sigma}_i^2$			0.0037 (0.0008)			
$\hat{\sigma}_p^2$	0.0032 (0.0005)	0.0421 (0.0024)	0.1840 (0.0103)	0.2635 (0.0147)	0.2814 (0.0158)	0.4238 (0.0243)
$\hat{\sigma}_e^2$	0.1666 (0.0154)	0.2054 (0.0154)	0.3473 (0.0154)	0.4269 (0.0154)	0.4448 (0.0154)	0.5871 (0.0154)
$\hat{\sigma}_{phen}^2$	0.3409 (0.0568)	0.2765 (0.0490)	0.1635 (0.0324)	0.1331 (0.0272)	0.1277 (0.0262)	0.0961 (0.0205)
h_g^2	0.0168 (0.0034)	0.0136 (0.0027)	0.0081 (0.0015)	0.0066 (0.0012)	0.0063 (0.0012)	0.0048 (0.0009)
c_s^2	0.6007 (0.0538)	0.4871 (0.0399)	0.2881 (0.0222)	0.2344 (0.0183)	0.2249 (0.0176)	0.1704 (0.0136)
c_i^2	0.0223 (0.0053)	0.0181 (0.0042)	0.0107 (0.0024)	0.0087 (0.0019)	0.0083 (0.0019)	0.0063 (0.0014)
c_p^2	0.0194 (0.0018)	0.2048 (0.0154)	0.5297 (0.0235)	0.6173 (0.0233)	0.6327 (0.0220)	0.7218 (0.0321)

continue...

Table 3. Continuation...

Component / Parameter	M1	M2	M3	M4	M5	M6
ρ	0.3754	0.3049	0.1809	0.1473	0.1414	0.1071
μ	0.201	0.5286	1.483	1.3095	2.0392	2.6521
CV	28.14	12.27	28.92	40.51	25.17	24.55

$\hat{\sigma}_g^2$: genetic variance between progenies; $\hat{\sigma}_s^2$: permanent environmental variance; $\hat{\sigma}_i^2$: GM interaction variance; $\hat{\sigma}_p^2$: plot variance; $\hat{\sigma}_r^2$: residual variance; $\hat{\sigma}_{phen}^2$: phenotypic variance; h_g^2: broad sense individual heritability; c_s^2: coefficient of determination of permanent environmental effects; c_i^2 coefficient of determination of GM interaction effects; c_p^2: coefficient of determination of plot effects; c_{res}^2: coefficient of determination of residual effects; ρ: individual repeatability; μ: phenotypic mean; CV: variance coefficient in percentage.

Bayesian analysis

According to the DIC, the best repeatability model was that with heterogeneous residual variance (lowest DIC) (Table 1). Thus, this model was adopted to estimate the variance components and genetic values. Moreover, the DIC values indicated significance for all random effects, since the DIC value of the model under test was higher than the DIC value of the full model (Table 2). Geweke and Heidelberger and Welch criteria indicated convergence for all variance components whit 100,000 effective MCMC samples.

The HPD intervals confirmed the significance of all variance components, i.e., none interval involved the value 0 (Table 4). The broad sense individual heritability between progenies ranged from 0.10 (M6) to 0.35 (M5), whereas the repeatability coefficients ranged from 0.11 (M6) to 0.38 (M5) (Table 4). Further, the genetic correlation through measurements was equal to 0.36.

Table 4. Estimates of variance components and genetic and non-genetic parameters through Bayesian inference (MCMC) for the grain yield trait, evaluated in 730 individuals from 73 half-sib families of *Jatropha curcas*, in six harvests (M1 to M6). Values between parentheses refers to the HPD intervals.

Component / Parameter	M1	M2	M3	M4	M5	M6
$\hat{\sigma}_g^2$			0.0568 (0.0132)			
$\hat{\sigma}_a^2$			0.2271 (0.0528)			
$\hat{\sigma}_s^2$			0.0028 (0.0005)			
$\hat{\sigma}_i^2$			0.1001 (0.0091)			
$\hat{\sigma}_p^2$	0.0037 (0.0008)	0.0421 (0.0024)	0.1840 (0.0103)	0.2635 (0.0147)	0.2814 (0.0158)	0.4238 (0.0243)
$\hat{\sigma}_e^2$	0.0032 (0.0005)	0.0421 (0.0024)	0.1840 (0.0103)	0.2635 (0.0147)	0.2814 (0.0158)	0.4238 (0.0243)
$\hat{\sigma}_{phen}^2$	0.1666 (0.0154)	0.2054 (0.0154)	0.3473 (0.0154)	0.4269 (0.0154)	0.4448 (0.0154)	0.5871 (0.0154)
h_g^2	0.3409 (0.0568)	0.2765 (0.0490)	0.1635 (0.0324)	0.1331 (0.0272)	0.1277 (0.0262)	0.0961 (0.0205)
c_s^2	0.0168 (0.0034)	0.0136 (0.0027)	0.0081 (0.0015)	0.0066 (0.0012)	0.0063 (0.0012)	0.0048 (0.0009)
c_i^2	0.6007 (0.0538)	0.4871 (0.0399)	0.2881 (0.0222)	0.2344 (0.0183)	0.2249 (0.0176)	0.1704 (0.0136)
c_p^2	0.0223 (0.0053)	0.0181 (0.0042)	0.0107 (0.0024)	0.0087 (0.0019)	0.0083 (0.0019)	0.0063 (0.0014)
c_p^2	0.0194 (0.0018)	0.2048 (0.0154)	0.5297 (0.0235)	0.6173 (0.0233)	0.6327 (0.0220)	0.7218 (0.0321)
ρ	0.3754	0.3049	0.1809	0.1473	0.1414	0.1071
μ	0.201	0.5286	1.483	1.3095	2.0392	2.6521
CV	28.14	12.27	28.92	40.51	25.17	24.55

$\hat{\sigma}_g^2$: genetic variance between progenies; $\hat{\sigma}_s^2$: permanent environmental variance; $\hat{\sigma}_i^2$: GM interaction variance; $\hat{\sigma}_p^2$: plot variance; $\hat{\sigma}_r^2$: residual variance; $\hat{\sigma}_{phen}^2$: phenotypic variance; h_g^2: broad sense individual heritability between progenies; c_s^2: coefficient of determination of permanent environmental effects; c_i^2 coefficient of determination of GM interaction effects; c_p^2: coefficient of determination of plot effects; c_{res}^2: coefficient of determination of residual effects; ρ: individual repeatability; μ: phenotypic mean; CV: variance coefficient in percentage.

Selective accuracy

The mean selective accuracy by the Frequentist inference was equal to 0.84, ranging from 0.82 to 0.84 (Table 5). In the Bayesian inference, with degrees of freedom equals to 0.002, 0.02, 2, 4, and 7, similar results were obtained. By the Bayesian inference, the modes of the selective accuracy estimates were 0.84, for degrees of freedom 0.002, 0.02, 2, 4 and 7, identical to the value obtained by the Frequentist inference (Table 5).

Table 5. Selective accuracies by Frequentist (REML/BLUP) and Bayesian (MCMC) inferences.

|	REML/BLUP	MCMC$^{0.002}$	MCMC$^{0.02}$	MCMC^2	MCMC^4	MCMC^7
Minimum	0.82	0.80	0.80	0.80	0.84	0.81
Mean	0.84	0.84	0.84	0.84	0.84	0.84
Mode	0.84	0.84	0.84	0.84	0.84	0.84
Maximum	0.84	0.84	0.85	0.84	0.85	0.85
Range	0.02	0.04	0.04	0.04	0.01	0.04

[0.002, 0.02, 2, 4 and 7]: degrees of freedom; REML: restricted maximum likelihood; BLUP: best linear unbiased prediction; MCMC: Markov Chain Monte Carlo.

Genetic selection

The same progenies were selected by Frequentist and Bayesian inferences. The 20 selected progenies were: 6, 15, 16, 10, 41, 39, 70, 1, 34, 54, 48, 56, 67, 29, 73, 37, 35, 9, 42, and 61 (Fig. 1). Consequently, the predicted selection gains were quite similar by both classes of inference, namely, 13.31 and 13.20%. Also, for the 20 selected progenies, similarities were observed between the values of confidence (Frequentist inference) and HPD (Bayesian inference) intervals (Fig. 2).
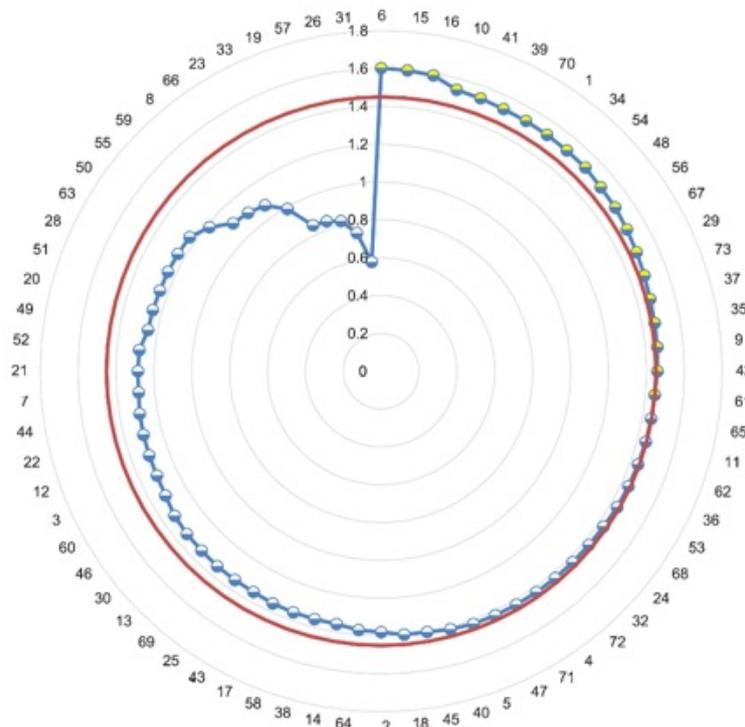
Figure 1. Ranking of the 73 *Jatropha curcas* progenies for the grain yield trait by Frequentist and Bayesian inferences (yellow dot highlights the 20 selected progenies).
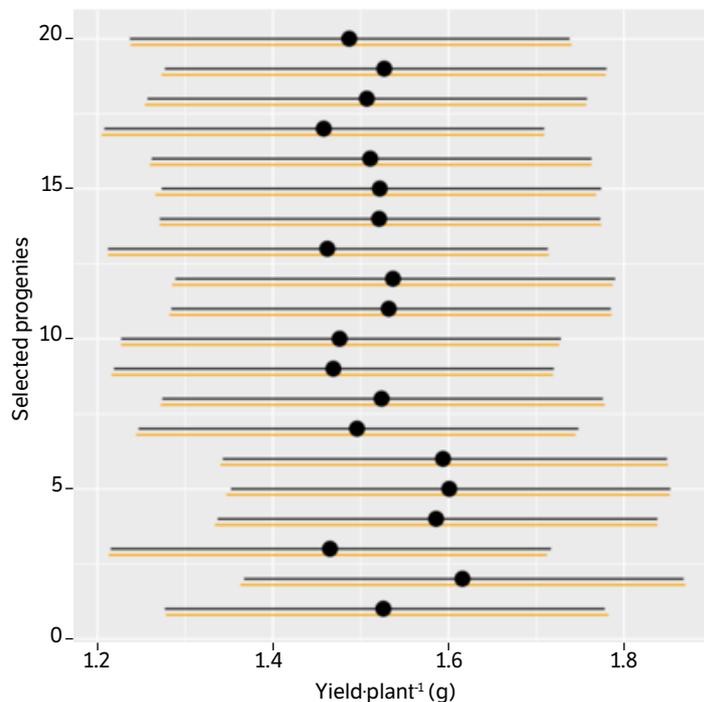
Figure 2. Predicted/estimated genetic values (black dot) for the 20 selected *Jatropha curcas* progenies for the grain yield trait and the confidence (black line) and highest posterior density (orange line) intervals obtained through Frequentist and Bayesian inferences.

DISCUSSION

In the analysis of repeated measures, it is very important to evaluate different residual variance structures (homogeneous, heterogeneous), since the prediction of the genetic values directly depends on the precision of the estimates of variance components (Peixoto et al. 2020; Peixoto et al. 2021).

In the Frequentist context, AIC and BIC are widely used for model selection (Casella and Berger 2002), while LRT is widely used to test the significance of the random effects of the statistical model. On the other hand, in the Bayesian context, the DIC is the most used for model selection and also to test the significance of the random effects of the statistical model (Lenarcic et al. 2012; Volpato et al. 2019; Evangelista et al. 2021). In this study, AIC and BIC, together with the LRT (for Frequentist inference), and DIC (for Bayesian inference), indicated the same statistical model, as best fitted: the full repeatability model with heterogeneous residual variance, showing the effects of the biotic and abiotic factors that are in constant change over time.

The significance of the genetic and GM interaction effects demonstrates the occurrence of genetic variability and GM interaction for the grain yield trait. Similar results were reported by Alves et al. (2018). The variance components estimated through REML and MCMC were remarkably similar. Consequently, similarities were found for the genetic parameters, predicted/estimated genetic values and selection gains. These results were expected, since no informative prior (nu > 2) was found in this study. Under such conditions, the results obtained through Frequentist inference are expected to be equal to those obtained through Bayesian inference (Gianola and Fernando 1986), since the Frequentist inference is equal to the Bayesian inference, except for the prior distribution.

For both classes of inference, the broad sense individual heritabilities ranged from low ($0.01 \leq h^2 \leq 0.15$) to moderate ($0.15 < h^2 < 0.50$) magnitudes (Resende 2015). Besides that, the individual repeatability coefficient presented low magnitude ($\rho \leq 0.30$) (Resende 2015). This last result indicates low genetic correlation across measurements, which implies the need to evaluate several crop years for an accurate genetic selection (Alves et al. 2018).

Confidence and HPD intervals presented similar magnitudes. However, the Frequentist inference provide only approximate confidence intervals using approximate information and assumptions of asymptotic normality (Silva et al. 2020). In this subject, the use of Bayesian inference was advantageous, compared with Frequentist inference, as it is based on the knowledge of the posterior distribution of the parameters. This fact allows the construction of exact creditability intervals for the estimates of variance components and genetic and non-genetic parameters (Resende et al. 2001).

In the Frequentist inference, the desirable properties of the point and interval estimates are based on hypothetical repetitions of the experiment (Resende et al. 2001). The credibility and HPD intervals are constructed to qualify in terms of final probability, which is, being the valid probability for that observed sample rather than possible repetitions or hypothetical results (Ehlers and Brooks 2004). This inference also allows the analysis of the small samples, without consuming degrees of freedom (Wakefield 2013). The latter statement (no consumption of degrees of freedom) is also valid for the random effects in the REML.

The selective accuracy shows how close the predicted/estimated genetic values are to the true genetic values. The closer those values, the higher the selective accuracy. According to Resende and Duarte (2007), the mean selective accuracy obtained by Frequentist and Bayesian inferences (0.84) indicates high selective accuracy ($0.70 \leq r_{\hat{g}g} < 0.90$). The approach for computing the Bayesian selective accuracy was proposed by Resende et al. (2014) and showed coherent and consistent results (Volpato et al. 2019). Like expected, the use of variance components estimated by REML like prior information in the Bayesian analysis did not maximize the selective accuracy. But it is important to report that this procedure made it easier to achieve convergence (Resende and Alves 2020).

Other advantages of the Bayesian inference in relation to the Frequentist inference are:
• The Bayesian inference allows flexibility in choosing the distributions for the set of unknown data and parameters;
• Possibility of prior incorporation of knowledge about the model's parameters, facilitating the model's convergence (Sorensen 2009).

Therefore, when compared to Frequentist inference, it seems reasonable that Bayesian inference presents better results due to its statistical properties (Silva et al. 2013; Silva et al. 2020; Peixoto et al. 2021).

Jatropha curcas is an undomesticated species in Brazil (Laviola and Rodrigues 2019). Therefore, the predicted selection gain ($\approx 14\%$) highlighted the great potential of the evaluated progenies for recombination and for development of a commercial variety. Similar predicted selection gains were reported by Alves et al. (2019), Peixoto et al. (2020) and Peixoto et al. (2021). Few studies exist in the literature evaluating repeated measures in *J. curcas* breeding. Thus, the results obtained in the present study using the Frequentist and Bayesian inferences contribute to filling a gap in the *J. curcas* breeding. In addition, these methods can be used for genetic evaluation and selection in other perennial species.

CONCLUSION

The Frequentist and Bayesian inferences conducted to similar results regarding variance components, genetic and non-genetic parameters, genetic values, and selection gains. The selected *J. curcas* progenies present great potential for recombination and for development of a commercial variety.

AUTHORS' CONTRIBUTION

DATA AVAILABILITY STATEMENT

The dataset is available in the supplement file.

FUNDING

ACKNOWLEDGMENTS

REFERENCES

Akaike, H. (1974). New look at the statistical model identification. IEEE Transactions on Automatic Control, 19, 716-723. https://doi.org/10.1109/TAC.1974.1100705

Alves, R. S., Peixoto, L. A, Teodoro, P. E., Silva, L. A., Rodrigues, E. V., Resende, M. D. V., Laviola, B. G. and Bhering, L. L. (2018). Selection of *Jatropha curcas* families based on temporal stability and adaptability of genetic values. Industrial Crops and Products, 119, 290-293. https://doi.org/10.1016/j.indcrop.2018.04.029

Alves, R. S., Teodoro, P. E., Peixoto, L. A., Rocha, J. R. A. S. C., Silva, L. A., Laviola, B. G., Resende, M. D. V. and Bhering, L. L. (2019). Multiple-trait BLUP in longitudinal data analysis on *Jatropha curcas* breeding for bioenergy. Industrial Crops and Products, 130, 558-561. https://doi.org/10.1016/j.indcrop.2018.12.019

Bahadur, B., Sujatha, M. and Carels, N. (2012). Jatropha, Challenges for a new energy crop. Cham: Springer.

Bailis, R. O. B. and McCarthy, H. (2011). Carbon impacts of direct land use change in semiarid woodlands converted to biofuel plantations in India and Brazil. GCB Bioenergy, 3, 449-460. https://doi.org/10.1111/j.1757-1707.2011.01100.x

Bhering, L. L., Almeida Filho, J. E., Peixoto, L. A., Laviola, B. G. and Gomes, B. E. L. (2015). Plateau regression reveals that eight plants per accession are representative for Jatropha germplasm bank. Industrial Crops and Products, 65, 210-215. https://doi.org/10.1016/j.indcrop.2014.11.056

Butler, D. G., Cullis, B. R., Gilmour, A. R. and Gogel, B. J. (2009). ASReml-R reference manual (version 3). Brisbane: State of Queensland, Department of Primary Industries and Fisheries.

Casella, G. and Berger, R. L. (2002). Statistical inference (v. 2). Pacific Grove: Duxbury.

Ehlers, R. S. and Brooks, S. P. (2004). Bayesian analysis of order uncertainty in ARIMA models. Brazil: Universidade Federal do Paraná.

Evangelista, J. S. P. C., Peixoto, M. A., Coelho, I. F., Alves, R. S., Silva, F. F., Resende, M. D. V., Silva, F. L. and Bhering, L. L. (2021). Environmental stratification and genotype recommendation toward the soybean ideotype: a Bayesian approach. Crop Breeding and Applied Biotechnology, 21, e359721111. https://doi.org/10.1590/1984-70332021v21n1a11

Ferreira, F. M., Carvalho Rocha, J. R. D. A. S., Alves, R. S., Elizeu, A. M., Benites, F. R. G., Resende, M. D. V. and Bhering, L. L. (2020). Estimates of repeatability coefficients and optimum number of measures for genetic selection of Cynodon spp. Euphytica, 216, 70. https://doi.org/10.1007/s10681-020-02605-x

Geweke, J. (1991). Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments (v. 196, p. 1-30). St. Louis: Federal Reserve Bank of Minneapolis, Research Department Minneapolis.

Gianola, D., and Fernando, R. L. (1986). Bayesian methods in animal breeding theory. Journal of Animal Science, 63, 217-244. https://doi.org/10.2527/jas1986.631217x

Hadfield, J. D. (2010). MCMC methods for multi-response generalized linear mixed models: the MCMCglmm R package. Journal of Statistical Software, 33, 1-22. https://doi.org/10.18637/jss.v033.i02

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

Heidelberger, P., and Welch, P. D. (1983). Simulation run length control in the presence of an initial transient. Operations Research, 31, 1109-1144. https://doi.org/10.1287/opre.31.6.1109

Henderson, C. R. (1974). General flexibility of linear model techniques for sire evaluation. Journal of Dairy Science, 57, 963-972. https://doi.org/10.3168/jds.S0022-0302(74)84993-3

Junqueira, V. S., Azevedo Peixoto, L., Laviola, B. G., Bhering, L. L., Mendonca, S., Costa, T. S. A. and Antoniassi, R. (2016). Bayesian multi-trait analysis reveals a useful tool to increase oil concentration and to decrease toxicity in *Jatropha curcas* L. PloS One, 11, e0157038. https://doi.org/10.1371/journal.pone.0157038

Laviola, B. G. and Rodrigues, E. V. (2019). Pinhão-manso: pesquisas, conhecimentos e práticas. Brasília: Embrapa Agroenergia.

Lenarcic, A. B., Svenson, K. L., Churchill, G. A. and Valdar, W. (2012). A general Bayesian approach to analyzing diallel crosses of inbred strains. Genetics, 190, 413-435. https://doi.org/10.1534/genetics.111.132563

Nithiyanantham, S., Siddhuraju, P., and Francis, G. (2012). Potential of *Jatropha curcas* as a biofuel, animal feed and health products. Journal of the American Oil Chemists' Society, 89, 961-972. https://doi.org/10.1007/s11746-012-2012-3

Patterson, H. D. and Thompson, R. (1971). Recovery of inter-block information when block sizes are unequal. Biometrika, 58, 545-554. https://doi.org/10.1093/biomet/58.3.545

Peixoto, M. A., Alves, R. S., Coelho, I. F., Evangelista, J. S. P. C., Resende, M. D. V., Rocha, J. R. A. S. C., Silva, F. F., Laviola, B. G. and Bhering, L. L. (2020). Random regression for modeling yield genetic trajectories in *Jatropha curcas* breeding. PLoS One, 15, e0244021. https://doi.org/10.1371/journal.pone.0244021

Peixoto, M. A., Evangelista, J. S. P. C., Coelho, I. F., Alves, R. S., Laviola, B. G., Fonseca e Silva, F., Resende, M. D. V. and Bhering, L. L. (2021). Multiple-trait model through Bayesian inference applied to *Jatropha curcas* breeding for bioenergy. PloS One, 16, e0247775. https://doi.org/10.1371/journal.pone.0247775

R Core Team (2019). R: A Language and Environment for Statistical Computing. R Development Core Team.

Resende, M. D. V. (2015). Genética quantitativa e de populações. Visconde do Rio Branco: Suprema.

Resende, M. D. V. (2016). Software Selegen-REML/BLUP: a useful tool for plant breeding. Crop Breeding and Applied Biotechnology, 16, 330-339. https://doi.org/10.1590/1984-70332016v16n4a49

Resende, M. D. V. and Alves, R. S. (2020). Linear, generalized, hierarchical, Bayesian and random regression mixed models in genetics/genomics in plant breeding. Functional Plant Breeding Journal, 3, 121-152. https://doi.org/10.35418/2526-4117/v2n2a1

Resende, M. D. V. and Duarte, J. B. (2007). Precisão e controle de qualidade em experimentos de avaliação de cultivares. Pesquisa Agropecuária Tropical, 37, 182-194. https://doi.org/10.5216/pat.v37i3.1867

Resende, M. D. V., Duda, L. L., Guimarães, P. R. B. and Fernandes, J. S. C. (2001). Análise de modelos lineares mistos via inferência Bayesiana. Revista de Matemática e Estatística, 19, 41-70.

Resende, M. D. V, Silva, F. F. and Azevedo, C. F. (2014). Estatística matemática, biométrica e computacional: modelos mistos, multivariados, categóricos e generalizados (REML/BLUP), inferência bayesiana, regressão aleatória, seleção genômica, QTL-QWAS, estatística espacial e temporal, competição, sobrevivência. Visconde do Rio Branco: Suprema.

Rodrigues, E. V., Rocha, J. R. D. A. S. D. C., Alves, R. S., Teodoro, P. E., Laviola, B. G., Resende, M. D. V. D. and Bhering, L. L. (2020). Selection of Jatropha genotypes for bioenergy purpose: an approach with multitrait, multiharvest and effective population size. Bragantia, 79, 346-355. https://doi.org/10.1590/1678-4499.20200046

Schwarz, G. (1978). Estimating the dimension of a model. The Annals of Statistic, 6, 461-464.

Silva, F. A., Viana, A. P., Corrêa, C. C. G., Carvalho, B. M., Sousa, C. M. B., Amaral, B. D., Ambrósio, M. and Glória, L. S. (2020). Impact of Bayesian Inference on the Selection of Psidium guajava. Scientific Reports, 10, 1999. https://doi.org/10.1038/s41598-020-58850-6

Silva, F. F., Viana, J. M. S., Faria, V. R. and Resende, M. D. V. (2013). Bayesian inference of mixed models in quantitative genetics of crop species. Theoretical and Applied Genetics, 126, 1749-1761. https://doi.org/10.1007/s00122-013-2089-6

Smith, B. J. (2007). boa: an R package for MCMC output convergence assessment and posterior inference. Journal of Statistical Software, 21, 1-37. https://doi.org/10.18637/jss.v021.i11

Sorensen, D. (2009). Developments in statistical analysis in quantitative genetics. Genetica, 136, 319-332. https://doi.org/10.1007/s10709-008-9303-5

Spiegelhalter, D. J., Best, N. G., Carlin, B. P. and van der Linde, A. (2002). Bayesian measures of model complexity and fit. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 64, 583-639. https://doi.org/10.1111/1467-9868.00353

Volpato, L., Alves, R. S., Teodoro, P. E., Resende, M. D. V., Nascimento, M., Nascimento, A. C. C., Ludke, W. H., Silva, F. L. and Borém, A. (2019). Multi-trait multi-environment models in the genetic selection of segregating soybean progeny. PLoS One, 14, e0215315. https://doi.org/10.1371/journal.pone.0215315

Wakefield, J. (2013). Bayesian and frequentist regression methods. New York: Springer.

Wilks, S. S. (1938). The large-sample distribution of the likelihood ratio for testing composite hypotheses. The Annals of Mathematical Statistics, 9, 60-62. https://doi.org/10.1214/aoms/1177732360