

Bayesian models applied to genomic selection for categorical traits

L.S. Silveira¹, S. Martins Filho¹, C.F. Azevedo¹; E.C. Barbosa¹,
M.D.V. Resende^{1,2} and E.K. Takahashi³

¹ Departamento de Estatística, Universidade Federal de Viçosa, Viçosa, MG, Brasil

² EMBRAPA Pesquisa Florestal, Colombo, PR, Brasil

³ CENIBRA Celulose Nipo Brasileira S.A., Belo Oriente, MG, Brasil

Corresponding author: L.S. Silveira
E-mail: lucasvrb@hotmail.com

Genet. Mol. Res. 18 (4): gmr18490

Received September 24, 2019

Accepted November 23, 2019

Published November 30, 2019

DOI <http://dx.doi.org/10.4238/gmr18490>

ABSTRACT. We compared two statistical methodologies applied to genetic and genomic analyses of categorical traits. The first one consists of a Bayesian approach to the Bayesian Linear Mixed Model (BLMM), which addresses the statistical problems of genomic prediction. The second methodology, called Bayesian Generalized Linear Mixed Model (BGLMM) is similar, but it is used when the distribution of the response variable is not Gaussian, as in the case of disease resistance phenotype categories. These models were compared according to predictive ability, bias, computational time and cross validation error rate (CVER). Additionally, an alternative classification method for the BLMM was proposed, which allowed us to obtain the CVER for this model. Estimates of the genetic parameters were obtained using BLASSO (Bayesian Least Absolute Shrinkage and Selection Operator) and Bayesian G-BLUP (Genomic Best Linear Unbiased Prediction) estimation methods applied to BLMM and BGLMM. The models were applied in two scenarios, with two and four classes for the phenotype of resistance to rust disease caused by the pathogen *Puccinia psidii* and classified as reaction types (two classes) and infection levels (four classes) recorded for 559 trees of *Eucalyptus urophylla* with 24,806 SNP markers. Modeling this trait through SNPs allow the next generation

of plants to be selected early, reducing time and costs. We found the same predictive ability for both models and a bias value closer to the ideal for BLMM (GBLUP). The BGLMM had the best CVER (0.29 against 0.32 and 0.47 against 0.51 for 2 and 4 categories, respectively), BLMM had a three times shorter computational time, and though BLMM is not the most appropriate model for handling categorical data, this model presented similar responses to BGLMM. Thus, we consider it as an appropriate alternative for categorical data modeling.

Key words: Plant breeding; Genetic improvement; Statistical methods; Bayesian inference

INTRODUCTION

Selection of plants that are disease resistant is of fundamental importance for agribusiness, because it minimizes risks and production losses and ensures the supply of a raw material of better quality to the final consumer. With the advent of genome-wide selection (GWS), proposed by Meuwissen et al. (2001), it has become possible to select plants based on DNA information; the identification of genetically superior plants can be made before its total development, only collecting the plant DNA data at the beginning of the life cycle. However, these studies deal with a large number of predictor variables, which leads to estimation problems of high dimensionality and multicollinearity. Also, disease resistance is recorded as categorical variables in many databases and variables of this nature are abundant; however, according to Biscarini et al. (2014), little studied in GWS.

Thus, for analyses of disease resistance and others categorical traits, the approaches Generalized Linear Mixed Model (GLMM) can be used. In addition, according to Biscarini et al (2014), statistical methods used for continuous traits may not be adequately applied to categorical traits due to the lack of linear relationship between predictors and categorical phenotype. It would be useful to know if modeling a categorical trait using a Bayesian approximation in a mixed linear model (BLMM), ignoring the assumption of normality of the data and residuals will produce results that are different from those obtained with the Bayesian approximation for mixed generalized linear model (BGLMM). The Bayesian approach addresses the classical estimation problems of GWS (high dimensionality and multicollinearity) using a priori distributions, without requiring the maximization of a specific function. Consequently, we compared these two models. Additionally, we developed an alternative way to allow categorization in BLMM, because this model does not provide a direct way to categorize as does BGLMM. The comparison of these two approaches was made by predictive ability, bias, computational time and cross validation error rate. The estimates of the genetic parameters were obtained using BLASSO (Bayesian Least Absolute Shrinkage and Selection Operator) and Bayesian G-BLUP (Genomic Best Linear Unbiased Prediction), according to Perez and De Los Campos (2014). These two estimation methods were applied in the BLMM and BGLMM frameworks. Then, four approaches were compared.

The objectives of this paper were to compare the strategies BLMM-BLASSO, BLMM-GBLUP, BGLMM-BLASSO, BGLMM-GBLUP using a database (genotypes

and phenotypes) of 559 trees of *E. urograndis*, originated from the crossing *E. urophylla* X *E. grandis*. The categorical phenotypes of rust disease resistance caused by *Puccinia psiidi* pathogen were evaluated with four and two categories.

MATERIAL AND METHODS

Data set

The data used in this analysis refers to 559 trees genotyped by SNP (single nucleotide polymorphism) markers and phenotyped for resistance to rust disease. The experiment was conducted in a randomized incomplete block design with single-tree plots and 24-36 repetitions per family in two types of soils (cambisol and latosol). The individuals came from 37 F2 full-sib families of *E. urograndis* from the cross of 10 progenies of *E. grandis* × *E. urophylla* hybrids breeding population belonging to Celulose Nipo-Brasileira (CENIBRA S.A.). For the prediction of genomic genetic values, a panel containing 24,806 SNPs markers was used. This database was previously treated according to Resende et al. (2017).

Genotyping

Genotyping was done using Illumina Infinium (Gunderson et al., 2005) and EuCHIP60K (Silva-Junior et al., 2015) generating information from 47,069 SNPs located at a distance of less than 10kb from 30,444 SNPs previously annotated in gene models. After quality control with a call rate >95% and MAF >1% (MAF- Minor Allele Frequency), 24,806 SNP markers were included. Further details of the genotyping and quality control process are available in Resende et al. (2017).

Phenotyping

The resistance to rust was measured according to the grading scale proposed by Junghans et al. (2003), which has a ordered policotomic classification in four classes according to the degree of infection (size of the pustule) or in two classes depending on the type of reaction (Table 1).

Table1. Evaluation levels of infections and types of reaction in *Eucalyptus urophylla* to inocula with the pathogen *Puccinia psiidi*.

Levels	Size of pustules	types of reaction
0	No pustules	Resistant
1	Small pustules (<0.8mm in diameter)	Resistant
2	Medium Pustules (between 0.8mm and 1.6mm in diameter)	Susceptible
3	Large Pustules (> 1.6mm in diameter)	Susceptible

Bayesian Linear Mixed Model (BLMM)

The Bayesian linear mixed model has components of random effects specified by a normal multivariate distribution with vector of means equal to zero and specific matrix of variances and covariances. As explained in Perez and De Los Campos (2014), the BLMM model with the parameters estimated by BLASSO method is described in Equation 1.

$$y = X\beta + Z_1b + Z_2f + Mm_a + e, \quad (\text{Eq. 1})$$

where

y is the vector (559×1) of the rust resistance variable;

β is the vector (3×1) containing the general mean and the effects of the evaluated soils under an a priori flat distribution, that is, constant, with incidence matrix X (559×3);

$b \sim N(0, K_b \sigma_b^2)$ is the vector (36×1) of random block effects, with incidence matrix Z_1 (559×36), covariance matrix of the block effect K_b and variance of the block effect σ_b^2 ;

$f \sim N(0, K_f \sigma_f^2)$ is the family random effects vector (37×1), with incidence matrix Z_2 (559×37), covariance matrix K_f and variance σ_f^2 of the family effect;

$m_a | \tau \sim N(0, D \sigma_m^2)$ is the vector (24.806×1) of random marker effects and $p(\tau^2 | \lambda^2) = \prod_i \left(\frac{\tau_i^2}{2} \right) e^{\left[\frac{-\tau_i^2 \lambda^2}{2} \right]}$ with variance matrix $D = \text{diag}(\tau_1^2, \tau_2^2, \dots, \tau_j^2)$ and λ is the penalty parameter that can be estimated using the Markov Chain Monte Carlo (MCMC) method; σ_m^2 is the variance component of markers and M (559×24806) is the incidence matrix, where M is parameterized according to Vitezica et al. (2013) so that:

$$M = \begin{cases} \text{if } AA, \text{ then } m_{ij} = 2 - 2p \\ \text{if } Aa, \text{ then } m_{ij} = 1 - 2p, \\ \text{if } aa, \text{ then } m_{ij} = 0 - 2p \end{cases}$$

which AA represents one of the homozygous genotypes, Aa the heterozygous genotypes, aa the other homozygous genotypes and p is the frequency of dominant alleles of the marker; $e \sim N(0, I\sigma^2)$ is the vector of random errors associated with model (1) with variance component σ^2 and I is the identity matrix.

When the BLMM has the parameters estimated by the individual GBLUP method, the model is defined in Equation 2

$$y = X\beta + Z_1b + Z_2f + Ig + e, \quad (\text{Eq. 2})$$

which the effects of markers are replaced by individual additive genetic effects $g \sim N(0, G_a \sigma_g^2)$ with incidence matrix given by the identity matrix I (559×559), genomic relationship matrix for additive effect G_a (obtained as $MM' / \Sigma 2p(1-p)$), with M marker matrix and p is the allele frequency of the markers) and σ_g^2 is the

additive genomic variance component. In BLASSO, the genomic genetic value (g) was obtained from the estimation of marker effects (m_a), so that $g = Mm_a$.

Bayesian Generalized Linear Mixed Model (BGLMM)

The generalized linear model (GLM) has the assumption that the response variable or random component of the model belongs to the exponential family of distributions, that is, the probability density function or the probability function of the variable of interest can be written in the form $f(y|\theta, \phi) = \exp\{\frac{y\theta - b(\theta)}{a(\phi)} + c(y, \phi)\}$ where θ is the canonical form of the location parameter (that is, θ is the own population mean parameter), ϕ is the dispersion parameter, the functions $a(\cdot)$, $b(\cdot)$ and $c(\cdot)$ are real and known, in addition $b(\cdot)$ is a differentiable function of the location parameter. Thus, GLM can be defined in three components, which are: i) Random component: a response variable belonging to the exponential family of distribution; ii) systematic component: a predictive variables of the model; iii) link function: a monotonous and differentiable function $g(\cdot)$, which establishes a linear relationship between the predictor variables of the model and the mean of the response variable (McCullagh and Nelder, 1989). In this work, the BGLMM allows random effects to be added and the model, estimated by the BLASSO method, can be described by Equation 3.

$$\ell = X\beta + Z_1b + Z_2f + Mm_a + e, \quad (\text{Eq. 3})$$

which ℓ is a vector (559×1) of latent variables (or that can not be directly observed) with standard normal distribution conditional on the threshold parameters ($t_i \in \{1, \dots, K - 1\}$, with K representing the number of classes), which allows predicting in which rust resistance class a given genotype belongs to. Thus, if the evaluation is performed using four levels of infection, the classification can be given by:

$$y_i = \begin{cases} 0, & \text{if } \ell_i < t_1 \\ 1, & \text{if } t_1 \leq \ell_i < t_2 \\ 2, & \text{if } t_2 \leq \ell_i < t_3 \\ 3, & \text{if } \ell_i \geq t_3 \end{cases}$$

If the resistance to rust has been evaluated as reaction types (susceptible or resistant), the classification will be given by:

$$y_i = \begin{cases} 0, & \text{if } \ell_i < t_1 \\ 1, & \text{if } \ell_i \geq t_1 \end{cases}$$

which 0 represents the class of resistant individuals and 1 susceptible individuals. The effects of this model are the same as previously specified in model (Eq. 1) with the difference in error variance $e \sim N(0,1)$. The model described in Equation 4 was adjusted by the Bayesian G-BLUP method,

$$\ell = X\beta + Z_1b + Z_2f + Ig + e, \quad (\text{Eq. 4})$$

which $e \sim N(0,1)$ and the others effects of this model are specified as in (2).

Categorization in the BLMM

Unlike BGLMM, which allows working with categorical dependent variables, BLMM does not perform classifications and requires alternative ways of doing so. In this study the classification was made according to the proportions of classes previously observed. The threshold values were pre-established by the quantile of normal distribution based on the proportions of each classes contained in the population. The sum of all proportions is 1(one).

Estimation of parameters

The parameters of the respective models BLMM-BLASSO (1), BLMM-G-BLUP (2), BGLMM-BLASSO (3) and BGLMM-G-BLUP (4) were estimated under the Bayesian approach using the BGLR package (Perez and De Los Campos, 2014) of R software (R Development Core Team, 2016). More specifically, 120,000 iterations were made with 50,000 burn-in used to remove the heat-up period from the chain and the thin of 10 to minimize the effect of autocorrelation between the samples of the marginal posterior distribution. Simulation was made using the Metropolis-Hasting algorithm of the Markov Chain Monte Carlo method. Subsequently, the parameters were estimated by taking the posterior mean of the parameters. The convergence of the Markov chains was evaluated by the Geweke criterion.

Model Validation

The validation of the estimates was performed through cross-validation under the 13-folds procedure. The initial population was equally and randomly divided into 13 groups, one of this group was considered as validation population and the other 12 groups used as training population. In the training population, the model parameters were estimated and then used to predict the genomic values of individuals in the validation population. Each of the 13 groups was once used as a validation population while the remaining 12 groups were used as training population.

Predictive Ability

The predictive ability in BGLMM was obtained through the Pearson's correlation between the observed phenotype (y) and the latent variable (ℓ) underlying the predicted phenotype. From the latent variable it was possible to categorize the eucalyptus tree as resistant or susceptible or at infection levels according to its position in relation to the estimated threshold parameters. The estimation of this correlation was given by $\rho_1 = \frac{cov(y,\ell)}{\sqrt{var(y)} \cdot \sqrt{var(\ell)}}$ or $\rho_2 = \frac{cov(y,\hat{y})}{\sqrt{var(y)} \cdot \sqrt{var(\hat{y})}}$, where ρ_1 is the correlation coefficient obtained in BGLMM and ρ_2 is the correlation coefficient obtained in BLMM.

Bias of prediction

The bias estimates were obtained from the regression coefficient of the response variable observed as a function of the predicted phenotype, or similarly by the Equation 5:

$$|1 - b_{y\hat{y}}| = \left| 1 - \frac{\text{cov}(y, \hat{y})}{\sigma_y^2} \right|, \quad (\text{Eq. 5})$$

which $b_{y\hat{y}}$ represents the bias and σ_y^2 is the variance of the phenotype. According to Resende et al. (2014), the unbiased estimate in genomic prediction is important when selection involves individuals of many generations using the effects of the markers obtained in just one generation. When BGLMM was used, the bias ($1 - \hat{b}_{y\ell}$) was obtained from the regression coefficient of the observed variable y as a function of the latent variable ℓ and not as a function of y .

Cross Validation Error Rate

The cross validation error rate (CVER) was calculated according to the categorizations made in each validation group, so that, if the predicted phenotypic value of the trait is equal to the observed value, the zero value is recorded for the error rate of this population validation, and if these values are different, 1 is counted for the error rate of this validation population. In this way,

$$\text{CVER} = \frac{1}{K} \sum_{k=1}^K \sum_{i=1}^{n_k} \frac{I_i}{n_k} \quad (\text{Eq. 6})$$

where K is the total number of validation populations, n_k is the number of individuals in each validation population and $I_i \forall (i = 1, 2, \dots, n_k)$ is an indicator variable, such that $I_i = 1$ if $\hat{Y} \neq Y$ and 0 otherwise, that is, $I_i = 0$ if $\hat{Y} = Y$.

RESULTS AND DISCUSSION

The Table 2 contains the CVER values for scenarios with two and four classes in the rust resistance assessment.

Table 2. Cross validation error rate (CVER) in two scenarios, the Bayesian Generalized Linear Mixed Model (BGLMM) and the Bayesian Linear Mixed Model (BLMM) estimated by BLASSO and GBLUP methods.

Number of classes	Model	BLASSO	GBLUP
2	BGLMM	0.29 ± 0.09	0.29 ± 0.08
	BLMM	0.32 ± 0.09	0.31 ± 0.09
4	BGLMM	0.47 ± 0.11	0.47 ± 0.11
	BLMM	0.51 ± 0.09	0.51 ± 0.09

Results showed that the BGLMM had lower CVER in both scenarios (two or four categories). This means that this model made better categorizations than BLMM. Curiously, if the categorization method used in BLMM is used in BGLMM, the error rates in both

models become equal. Therefore, BGLMM made better categorizations due to the way that the model categorize (according to the probit link function that relates the predictor variables to the mean of the response variable). A similar study was made by Montesinos-López et al. (2015), which compared the classification of the BGLMM (obtained by probit link function) with an implementation proposed by the authors that use a logit link function. In their paper the use of different link functions did not promote differences in results.

The predictive ability and bias measured for the BGLMM and BLMM models are presented in the Table 3. In both scenarios, there was no difference in predictive ability between BGLMM and BLMM. For bias values, BGLMM-BLASSO and BLMM-BLASSO were identical and the BLMM-GBLUP presented a bias value closer to the ideal than BGLMM-GBLUP.

Table 3. Predictive ability (Cor) and bias of Bayesian Linear Mixed Model (BLMM) and Bayesian Generalized Linear Mixed Model (BGLMM) with parameters estimated by BLASSO (models (1) and (3)) and GBLUP (models (2) and (4)) methods.

Number of Classes	Model	Cor. BLASSO	Cor. GBLUP	Bias-BLASSO	Bias-GBLUP
2	BGLMM	0.30 ± 0.10	0.30 ± 0.10	0.45 ± 0.17	0.39 ± 0.14
	BLMM	0.30 ± 0.10	0.30 ± 0.10	0.45 ± 0.17	1.04 ± 0.40
4	BGLMM	0.37 ± 0.11	0.37 ± 0.11	0.81 ± 0.24	0.87 ± 0.26
	BLMM	0.37 ± 0.11	0.38 ± 0.10	0.81 ± 0.24	1.01 ± 0.25

The bias in prediction is of great importance in genomic selection, especially when it comes to quantitative traits, because when individuals with overestimated /underestimated genetic values are selected, it can induce to economic losses. For the trait of resistance to rust in eucalyptus, biased models could indicate resistant individuals, when, in fact, they are susceptible. Therefore, it is usual to obtain the bias in studies of genetic improvement, for example, Azevedo et al. (2014) and (2015) in animal breeding and Sousa et al. (2018) in plant breeding.

Comparisons between BLMM and BGLMM have been previously made by Tiezzi et al. (2015). They compared similar models to BLMM and BGLMM (threshold model) presented here and evaluated them for detecting variance components and concluded that the BGLMM model explained a higher proportion of additive variance compared to BLMM.

Recent studies using the models compared in this work with other methodologies showed good results of BGLMM as Montesinos-López et al. (2019) that compared two machine learning methods with BGLMM (this model was called TGBLUP which resembles the BGLMM-GBLUP) in seven databases. BGLMM obtained the best percentage of cases correctly classified in four of the seven evaluated databases. Also, Ornella et al. (2012) compared BLASSO with Bayesian Ridge Regression (BRR) and two Machine Learning methods (support linear vector regression and support vector gaussian regression) to predict stem rust and yellow rust in wheat database. The results showed similarity between the BLASSO and BRR methods besides a slight superiority of BLASSO in relation to the Machine Learning methods.

The computational time of analyses with the BLASSO method is approximately fifteen times greater than with the GBLUP method. This is explained by the difference in

the number of parameters that each model estimates. While the BLASSO method estimates the effect of each marker, that is, 24,806 parameters of markers plus the respective variance components, the GBLUP method estimates the additive effects of the model at the individual level, that is, the number of genetic parameters corresponds to 559 plus its respective variance components. Therefore, the GBLUP estimates a lower number of parameters. Also, the BLMM has better computational time than BGLMM (time in BGLMM-BLASSO is approximately three times greater than BLMM-BLASSO), because it does not need to estimate threshold parameters, which has 1 (in the scenario with 2 classes) and 3 (in the scenario with 4 classes) addit parameter in each iteration of Markov Chain.

Studies in the same database was performed by Resende et al. (2017) with traits of complex growth, wood and disease resistance aiming to detect genomic regions in association studies. Associations for the rust resistance were detected on chromosome 3. According to Resende et al. (2017) genomic heritability for the resistance to *Puccinia psidii* captured 89% of the heritability trait which reaches the value of 0.36. Thus, it is possible to select more resistant trees only by using molecular marker data and, concomitantly, good predictive ability values and bias are required. We got the same heritability value using a BLMM model with the pedigree effect only.

CONCLUSIONS

Considering the results, we saw that BLMM had a similar error rate cross validation. However, it presented a predictive ability similar to BGLMM in both scenarios. Moreover, when GBLUP was used, the bias values in both scenarios were closer to optimal for BLMM. Thus, analyzing a categorical trait in genomic selection by a linear mixed model with Bayesian approximation becomes an alternative, especially when a lower bias value in prediction is desired.

ACKNOWLEDGMENTS

The authors would like to thank the Brazilian funding organizations: Coordination for the Improvement of Higher Education Personnel (CAPES) and CENIBRA- Celulose Nipo Brasileira.

CONFLICTS OF INTEREST

The authors declare no conflict of interest.

REFERENCES

- Azevedo CF, Silva FF, Resende MDV, Lopes MS, et al. (2014). Supervised independent component analysis as an alternative method for genomic selection in pigs. *J. Ani. Breed Genet.* 131: 452-461. doi: 10.1111/jbg.12104.
- Azevedo CF, Resende MDV, Silva FF, Viana JMS, et al. (2015). Ridge, Lasso and Bayesian additive-dominance genomic models. *BMC Genetics.* 16: 105. doi: 10.1186/s12863-015-0264-2.
- Biscarini F, Stevanato P, Broccanello C, Stella A, et al. (2014). Genome-enabled predictions for binomial traits in sugar beet populations. *BMC Genetics.* 15: 87. Doi: 10.1186/1471-2156-15-87.
- Gunderson KL, Steemers FJ, Lee G, Mendoza LG, et al. (2005). A genome-wide scalable SNP genotyping assay using microarray technology. *Na Genet.* 37: 549. doi: 10.1038/ng1547.
- Junghans DT, Alfenas AC and Maffia LA (2003). Escala de notas para quantificação da ferrugem em *Eucalyptus*. *Fitopatol. Bras.* 28: 184-188. doi: 10.1007/s00122-003-1415-9.

- McCullagh P and Nelder JA (1989). *Generalized linear models*. CRC press.
- Meuwissen TH, Hayes BJ and Goddard ME (2001). Prediction of total genetic value using genome-wide dense marker maps. *Genetics*. 157: 1819.
- Montesinos-López OA, Montesinos-López A, Crossa J, Burgueño J, et al. (2015). Genomic-enabled prediction of ordinal data with Bayesian logistic ordinal regression. *G3: Genes Genom. Genet.* 5: 2113-2126. doi: 10.1534/g3.115.021154.
- Montesinos-López OA, Martín-Vallejo J, Crossa J, Gianola D, et al. (2019). A benchmarking between deep learning, support vector machine and Bayesian threshold best linear unbiased prediction for predicting ordinal traits in plant breeding. *G3: Genes Genom. Genet.* 9: 601-618. doi: <https://doi.org/10.1534/g3.118.200998>.
- Ornella L, Singh S, Perez P, Burgueño J, et al. (2012). Genomic prediction of genetic values for resistance to wheat rusts. *Plant Genome*. 5: 136-148. doi: 10.3835/plantgenome2012.07.0017.
- Pérez P and de los Campos G (2014). Genome-wide regression and prediction with the BGLR statistical package. *Genetics*. 198: 483-495. doi: 10.1534/genetics.114.164442.
- R Core Team (2016). R: A language and environment for statistical computing, version 3.5.1. R Foundation for Statistical Computing, Vienna, Austria, <http://www.R-project.org> (accessed November 29, 2018).
- Resende MDV, Silva FF and Azevedo CF (2014). Estatística matemática, biométrica e computacional: modelos mistos, multivariados, categóricos e generalizados (reml/blup), inferência bayesiana, regressão aleatória, seleção genômica, qtl-gwas, estatística espacial e temporal, competição, sobrevivência. Suprema, Visconde do Rio Branco.
- Resende RT, Resende MDV, Silva FF, Azevedo CF, et al. (2017). Regional heritability mapping and genome-wide association identify loci for complex growth, wood and disease resistance traits in *Eucalyptus*. *New Phytol.* 213: 1287-1300. doi: 10.1111/nph.14266.
- Silva-Junior OB, Faria DA and Grattapaglia D (2015). A flexible multi-species genome -wide 60K SNP chip developed from pooled resequencing of 240 *Eucalyptus* tree genomes across 12 species. *New Phytol.* 206: 1527-1540. doi: 10.1111/nph.13322.
- Sousa TV, Caixeta ET, Alkimim ER, Oliveira ACB, et al. (2018). Early Selection Enabled by the Implementation of Genomic Selection in *Coffea arabica* Breeding. *Front Plant Sci.* 9: 1934. doi: 10.1111/nph.13322.
- Tiezzi F, Maltecca C, Cecchinato A and Bittante G (2015). Comparison between different statistical models for the prediction of direct genetic component on embryo establishment and survival in Italian Brown Swiss dairy cattle. *Livest Sci.* 180: 6-13. doi: <https://doi.org/10.1016/j.livsci.2015.06.029>.
- Vitezica ZG, Varona L and Legarra A (2013). On the additive and dominant variance and covariance of individuals within the genomic selection scope. *Genetics*. 195: 1223-1230. doi: <https://doi.org/10.1534/genetics.113.155176>.