Ciência Rural, Santa Maria, v.53:10, e20220327, 2023
http://doi.org/10.1590/0103-8478cr20220327
ISSNe 1678-4596
BIOLOGY

Genome-enabled prediction through quantile random forest for complex traits

Cristiane Botelho Valadares[1] Moysés Nascimento[1] Maurício de Oliveira Celeri[1] Ana Carolina Campana Nascimento[1] Laís Mayara Azevedo Barroso[2] Isabela de Castro Sant'Anna[3] Camila Ferreira Azevedo[1]

[1]Departamento de Estatística, Universidade Federal de Viçosa (UFV), 36570-900, Viçosa, MG, Brasil. E-mail: moysesnascim@ufv.br. *Corresponding author.
[2]Departamento de Matemática, Universidade Federal de Rondônia (UNIR), Paraná, RO, Brasil.
[3]Centro de Seringueira e Sistemas Agroflorestais, Instituto Agronômico de Campinas (IAC), Votuporanga, SP, Brasil.

ABSTRACT: Quantile Random Forest (QRF) is a non-parametric methodology that combines the advantages of Random Forest (RF) and Quantile Regression (QR). Specifically, this approach can explore non-linear functions, determining the probability distribution of a response variable and extracting information from different quantiles instead of just predicting the mean. This evaluated the performance of the QRF in the genomic prediction for complex traits (epistasis and dominance). In addition, compare the accuracies obtained with those derived from the G-BLUP. The simulation created an F2 population with 1,000 individuals and genotyped for 4,010 SNP markers. Besides, twelve traits were simulated from a model considering additive and non-additive effects, QTL (Quantitative trait loci) numbers ranging from eight to 120, and heritability of 0.3, 0.5, or 0.8. For training and validation, the 5-fold cross-validation approach was used. For each fold, the accuracies of all the proposed models were calculated: QRF in five different quantiles and three G-BLUP models (additive effect, additive and epistatic effects, additive and dominant effects). Finally, the predictive performance of these methodologies was compared. In all scenarios, the QRF accuracies were equal to or greater than the methodologies evaluated and proved to be an alternative tool to predict genetic values in complex traits.
Key words: genomic selection, accuracy, epistasis, dominance, prediction.

Predição genômica por meio do random forest quantílitico para características complexas

RESUMO: *Quantile Random Forest* (QRF) é uma metodologia não paramétrica, que combina as vantagens do *Random Forest* (RF) e da Regressão Quantílica (QR). Especificamente, essa abordagem pode explorar funções não lineares, determinando a distribuição de probabilidade de uma variável resposta e extraindo informações de diferentes quantis em vez de apenas prever a média. O objetivo deste trabalho foi avaliar o desempenho do QRF em predizer o valor genético genômico para características com arquitetura genética não aditiva (epistasia e dominância). Adicionalmente, as acurácias obtidas foram comparadas com aquelas advindas do G-BLUP. A simulação criou uma população F2 com 1.000 indivíduos genotipados para 4.010 marcadores SNP. Além disso, doze características foram simuladas a partir de um modelo considerando efeitos aditivos e não aditivos, com número de QTL (Quantitative trait loci) variando de oito a 120 e herdabilidade de 0,3, 0,5 ou 0,8. Para treinamento e validação foi usada a abordagem da validação cruzada 5-fold. Para cada um dos folds foram calculadas as acurácias de todos os modelos propostos: QRF em cinco quantis diferentes e três modelos do G-BLUP (com efeito aditivo, aditivo e epistático, aditivo e dominante). Por fim, o desempenho preditivo dessas metodologias foi comparado. Em todos os cenários, as acurácias do QRF foram iguais ou superiores às metodologias avaliadas e mostrou ser uma ferramenta alternativa para predizer valores genéticos em características complexas.
Palavras- chave: Seleção genômica, precisão, epistasia, dominância, predição.

INTRODUCTION

Genomic selection (GS) presents high accuracy in predicting genomic breeding values, accelerating the process of genetic improvement (SINGH et al., 2019; LIU et al., 2019). Statistical methodologies generally used for genomic prediction, such as RR-BLUP, G-BLUP, Bayes A, and Bayes B (MEUWISSEN et al., 2001), are based on errors and; consequently, phenotypic values normality assumptions.

The employment of computational intelligence-based methods to predict genomic breeding values is increasing (SOUSA et al., 2021; KUJAWA & NIEDBAŁA, 2021). Compared to statistical methods for predicting genomic values, such methodologies do not require assumptions about the model, making them more flexible for a wide range of problems (ROSADO et al., 2022). Specifically, such flexibility allows one to deal naturally with different types of non-additive genetic effects, like dominance and epistasis.

Received 06.06.22 Approved 09.29.22 Returned by the author 01.19.23
CR-2022-0327.R1
Editors: Leandro Souza da Silva Alberto Cargnelutti Filho

Among computational intelligence methodologies, Random Forest (RF) has proven to be an interesting alternative for genomic prediction; in addition to possibly increasing the predictive performance of the method, it reduces, through the selection of variables, problems related to correlated variables (BREIMAN, 2001). Such methodology has been successfully employed to predict genomic breeding values, like the study by BARBOSA et al. (2021) that used RF in simulated populations with different levels of heritability and loci numbers of quantitative characteristics (QTL) in the presence of dominant and epistatic effects. These authors reported that machine-learning methodologies, such as RF, are powerful tools to predict genomic breeding values for traits that comprise non-additive genetic architecture. In addition, SOUSA et al. (2021) successfully applied the methodology in predicting rust resistance in *Coffea arabica*. This study has verified that the RF presented higher results, in terms of apparent error rate, compared to generalized Bayesian LASSO.

Another method used in GS that is robust to the break of assumptions and allows one to adjust models along the entire probability distribution of the characteristic of interest is quantile regression (QR) (KOENKER & BASSET, 1978). The QR allows the investigation of possible issues related to asymmetry and heteroscedasticity present in data sets (NASCIMENTO et al., 2019). Concerning GS, such a method was used by NASCIMENTO et al. (2017) that employed QR to estimate genomic breeding value from simulated data in different asymmetry scenarios and observed that the technique provided better results regarding accuracy in the presence of asymmetric phenotypic values. In addition, OLIVEIRA et al. (2021) evaluated the use of QR considering simulated data of autogamous plants with oligogenic characteristics and observed better or equal results to those obtained by RR-BLUP and BLASSO.

A methodology that seizes qualities from both QR and RF is the Quantile Random Forest (QRF) (MEINSHAUSEN, 2006). This approach combined the best explanation of a phenomenon obtained through QR and increases predictive power by using the RF.

The QRF was ranked among those with the best performance in the challenge of predicting drug sensitivity in cancer treatment (FANG et al., 2018; LIND & ANDERSON, 2019), efficiently predicted heat waves in Pakistan (KHAN et al., 2019) and marine flooding (ROHMER et al., 2020).

Despite its potential, the QRF has not yet been used in the context of GS, so this study aimed: i) to propose and evaluate the use of QRF in genomic prediction for complex characteristics (inclusion of dominance and epistasis effects); ii) to compare the accuracy of the QRF with those resulting from the G-BLUP methodology.

MATERIALS AND METHODS

Experimental data

An F2 population of a diploid species (2n = 20) was simulated, containing 1,000 individuals. A co-dominant 4,010 markers (locus) of bi-allelic single nucleotide polymorphisms (SNPs) were considered, distributed equally and equidistantly into 10 binding groups (chromosomes) with a size of 200 cm each (401 markers on each chromosome). With the simulated genotypic data of the F2 population, 12 scenarios (C1 to C12) were considered, with the number of controlling genes (QTLs) equal to 8, 40, 80 or 120, distributed equally among the first eight linkage groups and heritabilities of 0.3, 0.5 or 0.8 (Table 1).

The phenotypic characteristics of the 12 scenarios were simulated considering the mean (μ) equal to 100 and, coefficient of variation of 10%, average degree of dominance (d_i) equal to 0.5 and controlled by a model that also includes epistatic effect: $Y_i = \mu + \sum_j \alpha_j + \sum_j \sum_{j'} \alpha_j \alpha_{j'} + \varepsilon_i$ in which Y_i is the phenotypic value for the *i* observation; μ the overall mean; α_j is the effect of the favorable allele on the *j* locus and assume the values $u + a_i$, $u + d_i$ and $u - a_i$ the genotypic values associated with the classes AA, Aa, and aa, respectively, with *u* as the mean between the dominant homozygous (AA) and the recessive homozygous (aa); $\alpha_j \alpha_{j'}$ represents the interaction between favorable alleles in different loci (epistasis). For the variance of errors, we have the errors vector $\varepsilon \sim N(0, I V_\varepsilon)$, in which $V_\varepsilon = [(1-h^2)V_g]/h^2$ with V_ε as the residual variance, V_g the genotypic variance and h^2 the heritability.

Table 1 - Evaluated scenarios.

Heritability (h^2)	---number of controlling genes (QTLs)---			
	8	40	80	120
0.3	C1	C4	C7	C10
0.5	C2	C5	C8	C11
0.8	C3	C6	C9	C12

Quantile random forest (QRF)

For the construction of the QRF, it is necessary to obtain T regression trees generated from bootstrap samples considering subsets of markers under study, i.e., the construction of trees based on Random Forest (HASTIE et al., 2008). Later, for each generated tree (T_t), conditional distribution is obtained by weighting the observed values of the studied characteristic. Specifically, given an observation, X = X, it is defined for each terminal node (adjusted tree leaf), $F(X, T_{tf})$, the following weighting factor: $w_i(x, T_{tf}) = \frac{I_{\{x \in F(x,T_{tf})\}}}{\#\{m: X_m \in F(x,T_{tf})\}}$ with $\sum_{i=1}^{n} w_i(x, T_{tf}) = 1$, $I_{\{X_i \in F(x_i,T_{tf})\}}$ an indicator variable stating that the observed value (X = x) belongs to f-th leaf and $\#\{m: X_m \in F(x,T_{tf})\}$ represents the number of observations on the f-th leaf.

The prediction of a tree T_t, according to MEINSHAUSEN (2006), for a new point, $X = x_{new}$ is given by the weighted average of the observations, that is, $\hat{\mu}(x_{new}) = \sum_{i=1}^{n} w_i(x, T_{tf}) Y_i$. In this way, the prediction for a given observation, X = x, after the construction of T trees is given by: $\hat{\mu}_{RF}(x) = \sum_{i=1}^{n} w_i(x) Y_i$, in which: $w_i(x) = \frac{1}{T} \sum_{t=1}^{T} w_i(x, T_{tf})$. Taking into consideration that the estimated cumulative distribution function is given by: $\hat{F}(y|X = x) = \sum_{i=1}^{n} w_i(x) I_{\{Y_i \leq y\}}$ in which $I_{\{Y_i \leq y\}}$ is an indicator function, the predicted value for the -th quantile is given by $Q_\tau(x) = \inf \{y: \hat{F}(y|X = x) \geq \tau\}$,for any τ,$0<\tau<1$.

The main difference between QRF and RF is that for each node in each tree, the RF maintains only the average of the observations that fall into that node and neglects any other information. Conversely, the QRF maintains the value of all node observations (not just the average) and evaluates conditional distribution based on this information (MEINSHAUSEN, 2006).

Genomic best linear unbiased predictor (G-BLUP)

In order to compare the results obtained by Quantile Random Forest, the adjustment of three G-BLUP models was considered: i. G-BLUP-A (only the additive component), ii. G-BLUP-AD (additive component and due to the dominance effect); iii. G-BLUP-AE (additive component and due to epistasis additive x additive). The description of such models can be seen in detail in RESENDE et al. (2014).

Comparison of methodologies

In order to access the prediction quality of the evaluated models, accuracy was used. This measure is defined as Pearson's correlation coefficient between the individuals' simulated genetic values and those predicted by the adjusted model. The higher the accuracy value, the better the model is in terms of prediction capacity, and it can be used in the selection phase of new individuals. In this research, the accuracy of the QRF was calculated in the quantiles 0.1, 0.3, 0.5, 0.7, and 0.9 and compared to the accuracy of the adjusted G-BLUPs models.

For training and validation, 5-fold cross-validation was performed. The data set is divided into 5 populations. At the k-th fold (k = 1,..,5), the k-th population is used as a validation population. The remaining populations were used as a validation population. For each of the five folds, the accuracy of all the proposed genomic selection models was calculated: QRF at quantiles 0.1, 0.3, 0.5, 0.7 and 0.9, G-BLUP-A, G-BLUP-AD and G-BLUP-AE and at the end, the mean and the standard error between the folds was estimated.

Computational resources

Data simulation was performed on the Genes software (CRUZ, 2016). All G-BLUP models (with additive, additive and epistatic, additive and dominant effect) were adjusted using the Genomic Land software (AZEVEDO et al., 2019). To run the QRF, it was used the "quantreg Forest" package (MEINSHAUSEN, 2017) of the R software (R CORE TEAM et al., 2020).

RESULTS AND DISCUSSION

Overall, among the three evaluated G-BLUPs models (Additive - A, Additive and Dominant - AD, Additive and Epistatic - AE), those that have non-additive effects in their adjustment, i.e., GBLUP - AD and GBLUP - AE, were the ones that presented the best accuracy values for all evaluated scenarios (Figure 1). This result is reasonable since the data used in this study consider dominance and epistasis effects in its simulation, indicating that the adjustment of non-additive effects is an essential factor to be considered in the modeling. Similar results have been reported in the literature (CALLEJA-RODRIGUES et al., 2021; YADAV et al., 2021). Specifically, CALLEJA-RODRIGUES et al. (2021) showed that genomic prediction is more accurate

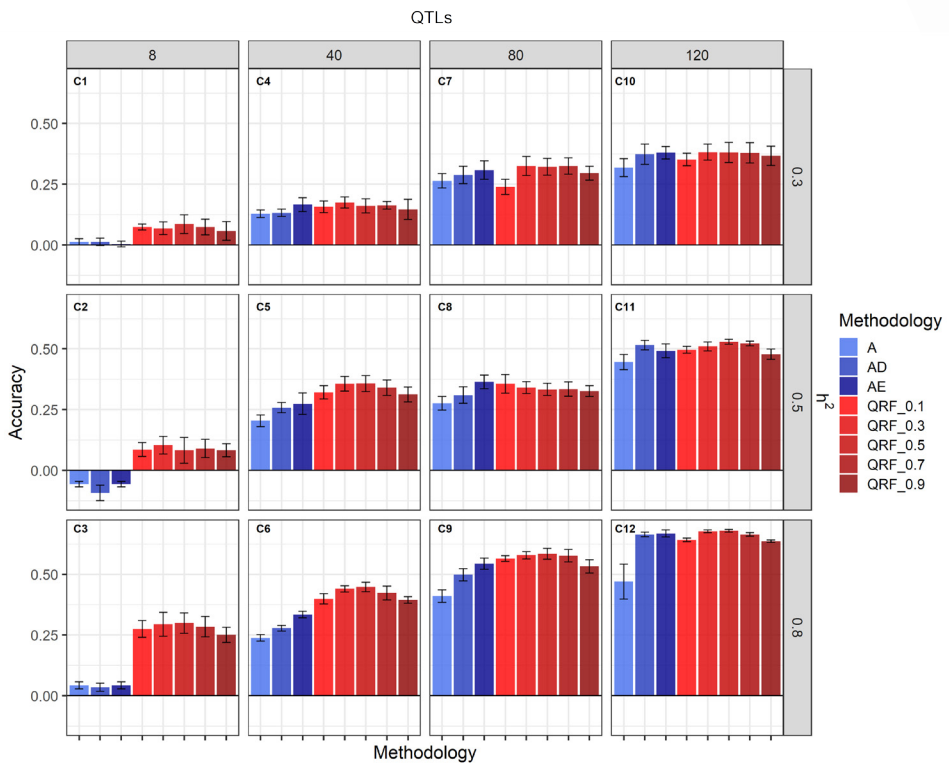

Figure 1 - Estimation of the accuracy of the QRF's models (QRF_0.1, QRF_0.3, QRF_0.5, QRF_0.7 e QRF_0.9), G-BLUP-A (only the additive component), G-BLUP-AD (additive component and due to the dominance effect); G-BLUP-AE (additive component and due to epistasis additive x additive) for combinations of the number of controlling genes (8, 40, 80 and 120) and heritability (0.3, 0.5 and 0.8), represented by scenarios C1 to C12.

when considering the inclusion of non-additive effects in *Pinus sylvestris*. YADAV et al. (2021) also obtained results that indicate the superiority of the adjustment of models with non-additive effects in sugarcane.

Another result that can be highlighted is that, with the increase in the number of QTLs, there was a significant improvement in the accuracy of G-BLUPs methods. This result is justified once G-BLUP considers the infinitesimal model (AZEVEDO et al., 2015), that is, many genes of little effect controlling the characteristic. Thus, genomic predictions are based on the kinship obtained from all markers and, thus, when more markers have a genetic effect, the accuracy of the prediction increases (WANG et al., 2018; ZHANG et al., 2019) exhibit an advantage on dense markers, and offer the flexibility of using different priors. In contrast, genomic best linear unbiased prediction (gBLUP).

Considering, as a basis for comparison the G-BLUP models with non-additive effects, the construction of models based on the QRF showed greater accuracy in the scenarios in which the characteristics were controlled by 8 and 40 QTLs (Figure 1, Scenarios C1 to C6). In the other scenarios, in which at least 80 QTLs control the characteristics, the adjustment through the QRF presented results similar to those obtained by adjusting the G-BLUP models considering non-additive effects (Figure 1, Scenarios C7 to C12). BARBOSA et al. (2021), considering simulated data, observed that for characteristics with the lowest number of QTLs, the multiplicative effects of the controlling genes (epistasis) may be more important since the individual effect of each gene is more significant than in the characteristics controlled by a larger number of QTLs. In SOUSA et al. (2021), machine learning methodologies were

applied for genomic prediction of rust resistance in *Coffea arabica*. Such a study verified that such methodologies, including the RF, presented higher results concerning apparent error rate compared to generalized Bayesian LASSO.

An interesting feature of the QRF is that, like QR, this method enables the adjustment of models for all parts of the probability distribution of the characteristic, allowing the conditional quantile that "better" describes the functional relationship between dependent and independent variables to be used for prediction (NASCIMENTO et al., 2019). In addition, QR does not require assumptions as to probability distribution and is robust to outliers (OLIVEIRA et al., 2021). In the present study, the quantiles 0.1, 0.3, 0.5, 0.7, and 0.9 were considered for the construction of the models. It was possible to observe, especially for a smaller number of QTLs (8 and 40 QTLs), that the highest accuracy values were observed when considering the models QRF_0.3 (quantile 0.3) and QRF_0.5 (quantile 0.5).

CONCLUSION

The QRF proved to be able to predict genetic values with epistatic gene control in characteristics with different degrees of heritability and different numbers of QTLs. It was equal to or superior to the G-BLUP methodology in all evaluated scenarios, presenting higher accuracy even when the characteristic is of low heritability.

ACKNOWLEDGMENTS

The authors are grateful to the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) for granting scholarships (code 001).

DECLARATION OF CONFLICT OF INTEREST

The authors declare no conflict of interest.

AUTHORS' CONTRIBUTIONS

All authors contributed equally for the conception and writing of the manuscript. All authors critically revised the manuscript and approved of the final version.

REFERENCES

AZEVEDO, C. F. et al. GenomicLand: Software for genome-wide association studies and genomic prediction. *Acta Scientiarum*. **Agronomy**, v.41, 2019. Available from: <https://doi.org/10.4025/actasciagron.v41i1.45361>. Accessed: Mar. 02, 2021. doi: 10.4025/actasciagron.v41i1.45361.

AZEVEDO, C. F. et al. Ridge, Lasso and Bayesian additive-dominance genomic models. **BMC Genetics** (Online), v.16, p.105, 2015. Available from: <https://doi.org/10.1186/s12863-015-0264-2>. Accessed: Jun. 14, 2021. doi: 10.1186/s12863-015-0264-2.

BARBOSA, I. P. et al. Genome-enabled prediction through machine learning methods considering different levels of trait complexity. **Crop Science**, v.61, p.1890–1902, 2021. Available from: <https://doi.org/10.1002/csc2.20488>. Accessed: Jan. 12, 2022. doi: 10.1002/csc2.20488.

BREIMAN, L. Random forests. **Machine Learning**, v.45, p.5–32, 2021. Available from: <https://doi.org/10.1023/A:1010933404324>. Acessed: Jan. 12, 2022. doi: 10.1023/A:1010933404324.

CALLEJA-RODRIGUEZ A. et al. Genomic Predictions With Nonadditive Effects Improved Estimates of Additive Effects and Predictions of Total Genetic Values in Pinus Sylvestris. **Frontiers in Plant Science**, v.12, 2021. Available from: <https://doi.org/10.3389/fpls.2021.66682>. Accessed: Jan. 11, 2022. doi: 10.3389/fpls.2021.666820.

CRUZ, C. D. Genes Software - extended and integrated with the R, Matlab and Selegen. *Acta Scientiarum*. **Agronomy**. v.38, n.4, 2016. Available from: <https://doi.org/10.4025/actasciagron.v38i4.32629>. Accessed: Mar. 02, 2021. doi: 10.4025/actasciagron.v38i4.32629.

FANG Y. et al. A quantile regression forest based method to predict drug response and assess prediction reliability. **PLoS ONE**, v.13, n.10, 2018. Available from: <https://doi.org/10.1371/journal.pone.0205155>. Accessed: Aug. 05, 2021. doi: 10.1371/journal.pone.0205155.

HASTIE, T. et al. **The Elements of Statiscal Learning**: Data Mining, Inference, and Prediction. Springer, 2. ed., 745p, New York, NY, USA, 2008.

KHAN, N. et al. Prediction of heat waves in Pakistan using quantile regression forests. **Atmospheric Research**, 2019. Available from: <https://doi.org/10.1016/j.atmosres.2019.01.024>. Accessed: Sept. 16, 2021. doi: 10.1016/j.atmosres.2019.01.024.

KOENKER, R.; BASSET, G. Regression Quantiles. **Econometrica**, v.46, p.33–50, 1978. Available from: <https://doi.org/10.2307/1913643>. Accessed: Jun. 14, 2021. doi: 10.2307/1913643.

KUJAWA, S.; NIEDBAŁA, G. Artificial Neural Networks in Agriculture. **Agriculture**, v.11, p.497, 2021. Available from: <https://doi.org/10.3390/agriculture11060497>. Accessed: Jan. 11, 2022. doi: 10.3390/agriculture11060497.

LIND A. P.; ANDERSON, P. C. Predicting drug activity against cancer cells by random forest models based on minimal genomic information and chemical properties. **PLoS ONE**, v.14, n.7, 2019. Available from: <https://doi.org/10.1371/journal.pone.0219774>. Accessed: Sept. 16, 2021. doi: 10.1371/journal.pone.0219774.

LIU, X. et al. Enhancing genomic selection with quantitative trait loci and nonadditive effects revealed by empirical evidence in maize. **Frontiers in plant science**, v.10, 2019. Available from: <https://doi.org/10.3389/fpls.2019.01129>. Accessed: Aug. 05, 2021. doi: 10.3389/fpls.2019.01129.

MEUWISSEN, T. H. E. et al. Prediction of Total Genetic Value Using Genome-Wide Dense Marker Maps. **Genetics**, v.157, 2001.

Available from: <https://doi.org/10.1093/genetics/157.4.1819>. Accessed: Mar. 02, 2021. doi: 10.1093/genetics/157.4.1819.

MEINSHAUSEN, N. Quantile regression forests. **Journal of Machine Learning Research**, v.7, p.983–999, 2006.

MEINSHAUSEN, N. quantregForest: Quantile Regression Forests. **R package version 1.3-7**. 2017. Available from: <https://cran.r-project.org/web/packages/quantregForest/quantregForest.pdf>. Accessed: Jul. 21, 2021.

NASCIMENTO, A. C. et al. Quantile Regression Applied to Genome-Enabled Prediction of Traits Related to Flowering Time in the Common Bean. **Agronomy**, v.9, n.12, 2019. Available from: <https://doi.org/10.3390/agronomy9120796>. Accessed: Aug. 02, 2021. doi: 10.3390/agronomy9120796.

NASCIMENTO, M. et al. Regularized quantile regression applied to genome-enabled prediction of quantitative traits. **Genetics and Molecular Research**, Ribeirão Preto, v.16, n.1, p.1-12, 2017. Available from: <https://doi.org/10.4238/gmr16019538>. Accessed: Sept. 16, 2021. doi: 10.4238/gmr16019538.

OLIVEIRA G. F. et al. Quantile regression in genomic selection for oligogenic traits in autogamous plants: A simulation study. **PLoS ONE**, v.16, n.1, 2021. Available from: <https://doi.org/10.1371/journal.pone.0243666>. Accessed: Jan. 12, 2022. doi: 10.1371/journal.pone.0243666.

R CORE TEAM, R. F. for S. Computing, and R. C. Team. 2020. **R**: A Language and Environment for Statistical Computing. Available from: <https://www.r-project.org/>. Accessed: Mar. 01, 2021. doi: 10.1371/journal.pone.0243666.

RESENDE, M. D. V. et al. **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. Viçosa: Suprema, 881p. 2014.

ROHMER, J. et al. A nuanced quantile random forest approach for fast prediction of a stochastic marine flooding simulator applied to a macrotidal coastal site. **Stoch Environ Res Risk Assess**, v.34, p.867–890, 2020. Available from: <https://doi.org/10.1007/s00477-020-01803-2>. Accessed: Sep. 17, 2021. doi: 10.1007/s00477-020-01803-2.

ROSADO, R. D. S. et al. Artificial neural network as an alternative for peach fruit mass prediction by non-destructive method. **Scientia Horticulturae**, v.299, 2022. Available from: <https://doi.org/10.1016/j.scienta.2022.111014>. Accessed: Jun. 02, 2022. doi: 10.1016/j.scienta.2022.111014.

SINGH, B. et al. Whole-Genome Selection in Livestock. In: **Advances in Animal Biotechnology**. Springer, Cham, p.349-364, 2019.

SOUSA, I. C. de. et al. Genomic prediction of leaf rust resistance to Arabica coffee using machine learning algorithms. **Scientia Agricola**, v.78, p.1, 2021. Available from: <https://doi.org/10.1590/1678-992X-2020-0021>. Accessed: Jan. 18, 2022. doi: 10.1590/1678-992X-2020-0021.

WANG J. et al. Expanding the BLUP alphabet for genomic prediction adaptable to the genetic architectures of complex traits. **Heredity**, v.121, p.648–662, 2018. Available from: <https://doi.org/10.1038/s41437-018-0075-0>. Accessed: Oct. 20, 2021. doi: 10.1038/s41437-018-0075-0 .

YADAV, S. et al. Improved genomic prediction of clonal performance in sugarcane by exploiting non-additive genetic effects. **Theoretical and Applied Genetics**, v.134, p.2235–2252, 2021. Available from: <https://doi.org/10.1007/s00122-021-03822-1>. Accessed: Jan. 24, 2022. doi: 10.1007/s00122-021-03822-1.

ZHANG, H. et al. Factors affecting the accuracy of genomic selection for agricultural economic traits in maize, cattle, and pig populations. **Frontiers in genetics**, v.10, 2019. Available from: <https://doi.org/10.3389/fgene.2019.00189>. Accessed: Oct. 20, 2021. doi: 10.3389/fgene.2019.00189.