



# GENOME-ENABLED PREDICTION OF GENETIC VALUES FOR USING RADIAL BASIS FUNCTION NEURAL NETWORKS

Isabela de Castro Sant'Anna<sup>1,3</sup>, Moyses Nascimento<sup>1</sup>, Gabi Nunes Silva<sup>2</sup>,  
Cosme Damião Cruz<sup>3</sup>, Camila Ferreira Azevedo<sup>1</sup>, Leonardo Siqueira Gloria<sup>4</sup>, Fabyano Fonseca e Silva<sup>5</sup>

<sup>1</sup>Laboratório de Inteligência Computacional e Aprendizado Estatístico, Universidade Federal de Viçosa (UFV), Viçosa, MG, 36570-900.

<sup>2</sup>Departamento de Matemática e Estatística, Universidade Federal de Rondônia (UNIR), Ji-Paraná, RO, 76900-726.

<sup>3</sup>Laboratório de Bioinformática, Universidade Federal de Viçosa (UFV), Viçosa, MG, 36570-900.

<sup>4</sup>Centro de Ciências e Tecnologias Agropecuárias, Universidade Estadual do Norte Fluminense Darcy Ribeiro (UENF), RJ, 28013-602.

<sup>5</sup>Departamento de Zootecnia, Universidade Federal de Viçosa (UFV), Viçosa, MG, 36570-900.

Corresponding author: Isabela de Castro Sant'Anna ([isabelacsantanna@gmail.com](mailto:isabelacsantanna@gmail.com))

**Abstract** - The objective of this work is to evaluate the efficiency of genomic selection (GS) of genome-enabled prediction by Radial Basis Function Neural Networks (RBFNN) in the prediction of genetic values considering dominance effects and different degrees of heritability. In addition, the results were compared with those obtained by G-BLUP. An F<sub>1</sub> population with 500 individuals genotyped with 1000 SNP-type markers was simulated. The phenotypic traits were determined by adopting two different gene action models: additive and dominance admitting heritability levels ( $h^2$ ) 30 and 60%, each is controlled by 50 loci, considering two alleles per loco. The accuracy and the mean squared error root (MSER) were estimated using a five-fold cross-validation scheme. For the low heritability scenario,  $h^2 = 0.3$  in the additive scenario, the accuracy of validation was 31% for RBFNN, 58% for RR-BLUP, and in the complete dominance scenario the values were 28% e 25%, respectively. Additionally, when analyzing the MSER the difference in performance of the techniques is even greater. For additive scenario, the estimates were 97.33 RR-BLUP and 5.80 for RBFNN, in the most critical scenario, 91.31 GBLUP and 14.55 for RBFNN. Overall, a RBFNN shows accuracy lower than those obtained through G-BLUP. On the other hand, the RBFNN has low prediction bias. Finally, the adjustment of the GBLUP to dominance models despite increasing the complexity of the model also increased the predictive accuracy compared to the model without considering the dominance effect.

**Keywords:** Dominance, neural network, genomic selection

## Introduction

One of the main contributions of molecular genetics for the benefit of plant breeding is the possibility of using directly DNA information in the selection of individuals. Genomic Selection (GS) (Meuwissen et al., 2001) enabling to estimate the genomic value of individuals (GEBV) without the need of phenotyping reducing time and money (Cossa et al., 2017).

For some species, as example maize, eucalyptus, cotton, rice and pinnus (Denis and Bouvet, 2011; Liu et al., 2012; Technow et al., 2012; Liang 2015; Almeida Filho et al., 2016), where there is commercial interest in hybrids and heterosis, the contribution of dominance presents high importance (Almeida Filho et al., 2016). The inclusion of these effects in GS models has been carried out by some authors (Azevedo et al., 2015; Almeida Filho et al., 2016; Denis and Bouvet, 2011; Technow et al., 2012; Santos et al., 2016; Sant'Anna et al., 2019; Viana

et al., 2017; Vitezica et al., 2017). Despite huge efforts in the development of statistical models for the implementation of GS with non-additive effects (as dominance), there are still some issues that have to be dealt with (Crossa et al., 2017). For example, either in the Bayesian or frequentist approaches it is necessary to define a priori the statistical model to be fit. On the other hand, methodologies based on Computational Intelligence (CI) allows to infer the trait architecture directly from the used data set. Therefore, differently from the Bayesian and frequentist methods there is no need to address assumptions about the trait architecture, distributions of data and residues (Haykin, 2009; Long et al., 2010; Gianola et al., 2011; Howards et al., 2014). Among several methodologies based on CI, Radial Basis Function Neural Networks (RBFNN) are a particular class of Neural Networks (NN) that presents attractive properties to GS. RBFNNs have the ability to learn from the data used in their training (Gianola et al., 2011), presents universal approximation properties (González-Camacho et al., 2012), give a unique solution (Cruz and Nascimento, 2018) and are faster than standard NNs (González-Camacho et al., 2018). RBFNN has been successful used in GS (Long et al., 2010, 2011; Pérez-Rodríguez et al., 2012; González-Camacho et al., 2012, 2016). In general, these studies concludes the application of RBFNN in GS demonst great potential for capturing complex interactions by recognizing the importance of the non-additive effects in training data and incorporating them into the predictions. However, they also recommend highly that the use of RBFNN needs further investigation to better understand the complexity of quantitative traits in order to obtain more accurate results.

In light of expose, this paper aimed to evaluate the efficiency of genome-enabled prediction by RBFNN in the prediction of genetic values considering dominance effects and different degrees of heritability. The results were compared with those obtained by one of the standard GS model: G-BLUP.

## Materials and methods

### Origin of populations

In order to assess the accuracy of GS predictions, data were simulated by considering a diploid species with  $2n = 2x = 20$  chromosomes as the reference, and the total length of the genome was stipulated in 1,000 cM. Genomes were generated with a saturation level of 101 molecular markers spaced by 1 cM per linkage group, totaling 1010 markers. Divergent parental line genomes were simulated. The effective size of the base population is the size of  $F_1$  itself, since the base population ( $F_1$ ) was derived from two contrasting homozygous parents.

### Simulation of quantitative traits

The genotypic value for the monogenic model is defined by  $u + a$ ,  $u + d$ ,  $u - a$  for the genotypes AA, Aa e aa, respectively. In a polygenic model, the total genotypic value expressed by a given individual belonging to the population was the sum of each additive effects of individual locus estimated by the following expression

$$G_i = \mu + a_i + d_i \text{ (Eq. 1)}$$

where the additive effect (a) of each locus is one half the difference in mean phenotype between the two homozygous genotypes (for each individual i). The dominance effect (d) is the difference between the mean phenotype of the heterozygous genotype and the average phenotype of the two homozygous genotypes. In our simulation we defined 50 loci to control the trait. Therefore, the additive effect is given by:

$$a_i = \sum_{j=1}^{50} p_j \alpha_j \text{ (Eq. 2)}$$

With  $\alpha_j$  being the effect of the favorable allele in locus  $j$ , considered equal to 1, 0 or -1 for the genotypic classes AA, Aa and aa, respectively, and  $p_j$  being the contribution of locus  $j$  to the manifestation of the trait under consideration. In this study, it was established as being equivalent to the probability of the set generated by the binomial distribution  $(a+b)^s$ , where  $a = b = 0.5$  and  $s = 49$ . The value of  $d_i$  was defined according to the average degree of dominance expressed in each trait. The quantitative traits were simulated in three scenarios considering three degrees of dominance ( $d/a = 0, 0.5$  and  $1$ ) and two broad sense heritability ( $h^2 = 0.30$  and  $0.60$ ), totaling six genetic architectures.

The phenotypic values of the  $i^{\text{th}}$  individuals were obtained according to the model:  $P_i = G_i + E_i$ , where  $G_i$  is the genetic effect given by the sum of the genetic effects in each locus, and  $E_i$  is the environmental effect, generated according to a normal distribution with means equal to zero and variance given by the equation bellow:

$$\sigma_e^2 = \frac{\sigma_g^2(1-h^2)}{h^2} \text{ (Eq. 3)}$$

where  $\alpha_e^2$  is the variance given by the environmental values,  $\alpha_g^2$  is the variance of the genetic values, and  $h^2$  is the heritability defined for the trait. The genetic variance is defined for each population from the information of the genetic control and the importance of each locus in the polygenic model.

$$\bar{a}^2 + \frac{1}{4}\bar{d}^2 \text{ (Eq. 4)}$$

where  $\bar{a}^2$ ,  $\bar{d}^2$  were defined by the mean values of the effects associated with the homozygote and heterozygous genotypes for each one of the 50 loci, respectively.

In the additive model the phenotypic value of the  $i^{\text{th}}$  individuals were obtained according to equation below:

$$Y_i = \mu + \sum_{j=1}^{50} p_j \alpha_j + E_i \text{ (Eq. 5)}$$

Where  $Y_i$  is the phenotypic value for individual (changed from 1 to 500 individuals),  $\mu$  is the general average, with  $\alpha_j$  being the effect of the favorable allele in locus  $j$ , considered equal to 1, 0 or -1 for the

genotypic classes AA, Aa and aa, respectively, and  $p_j$  being the contribution of locus  $j$  to the manifestation of the trait under consideration, in this work as having binomial distribution. The value of  $d_i$  was established from the average degree of dominance manifested in each trait as part of the equation  $\alpha_j = a_i + d_i e d_i/a_i = (\text{average degree of dominance})$ , being  $\mu + a_j$ ,  $\mu + d_j$  e  $\mu - a_j$  for the genotypic classes AA, Aa and aa, respectively, where  $a_i$  and  $d_i$  representing the deviation.

**Table 1.** Simulated scenarios composed by combination of traits, action genit model, heritability and dominance degree.

Scenarios	Heritability (%)	Model	Dominance
V1 - D0H30_Ad	30	additive	0
V2 - D0.5H30_Ado	30	additive-dominant	0.5
V3 - D1H30_Ad0	30	additive-dominant	1
V4 - D0H60_Ad	60	additive	0
V5 - D0.5H60_Ado	60	additive-dominant	0.5
V6 - D1H60_Ado	60	additive-dominant	1

**Genomic selection procedures**

The additive dominance model for the REML/G-BLUP method is given by Azevedo et al. (2019):

where  $y$  is the vector of phenotypic observations,

$$y = Xb + Zu_a + Zu_d + e \text{ (Eq.6)}$$

$b$  is the vector of fixed effects,  $u_a$  is the vector of random of additive marker effects,  $u_d$  is the vector of random of dominance marker effects and  $e$  refers to the vector of random errors; The variance structure is given by  $u_a \sim N(0, G_a \sigma_{ua}^2)$ ;  $u_d \sim N(0, G_d \sigma_{ud}^2)$ ; by  $e \sim N(0, I \sigma_e^2)$ .

An equivalent model at the marker level is given by

$$y = Xb + ZWm_a + ZSm_d + e \text{ (Eq. 7)}$$

where:  $u_a = Wm_a$ ;  $Var(Wm_a) = WI\sigma_{ma}^2W' = WW' \sigma_{ma}^2$ ;  $u_d = Sm_d$ ;  $Var(Sm_d) = SI\sigma_{md}^2S' = \sigma_{md}^2$ ;  $X$  and  $Z$  are matrices of incidence for the vectors additive ( $m_a$ ) and dominance ( $m_d$ ) marker genetic effects. The variance components associated to these effects are  $\sigma_{ma}^2$ ; and  $\sigma_{md}^2$ , respectively.  $G_a$  and  $G_d$  are the genomic relationship matrices for the additive and dominance effects. The quantity  $m_a$  in one locus is the allele substitution effect and is given by  $m_a = \alpha_i = a_i + (q_i - p_i)$ , where  $p_i$  and  $q_i$  are allelic frequencies and  $a_i$  and  $d_i$  are the genotypic values for one homozygote and heterozygote, respectively, at locus  $i$ . In turn, the quantity  $m_d$  can be directly defined as  $m_{di} = d_i$ . The matrices  $W$  and  $S$  are defined based on the values 0, 1 and 2 for the number of one of the alleles at the  $i$  marker locus in a diploid individual.

The correct parameterization of  $W$  and  $S$  is as follows, according to the marker genotypes at a locus  $m$ .

$$W = \begin{cases} MM: 2 - 2p \rightarrow 2q \\ Mm: 1 - 2p \rightarrow q - p \\ mm: 2 - 2p \rightarrow 2q \end{cases}$$

$$S = \begin{cases} MM: 0 \rightarrow -2q^2 \\ Mm: 1 \rightarrow 2pq \\ mm: 0 \rightarrow -2p^2 \end{cases}$$

The covariance matrix for the additive effects is given by  $G_a \sigma_{ma}^2 = Var(Wm_a) = WW' \sigma_{ma}^2$ , which leads to:

$$G_a = WW' / (\sigma_{ma}^2 / \sigma_a^2) = WW' / \sum_{i=1}^n [2p_i(1 - p_i)]$$

$$\sigma_d^2 = \sum_{i=1}^n [2p_i(1 - p_i)]^2 \sigma_{md}^2$$

The covariance matrix for the dominance effects is given by  $G_d \sigma_d^2 = Var(Sm_d)SS' \sigma_{md}^2$ . Thus

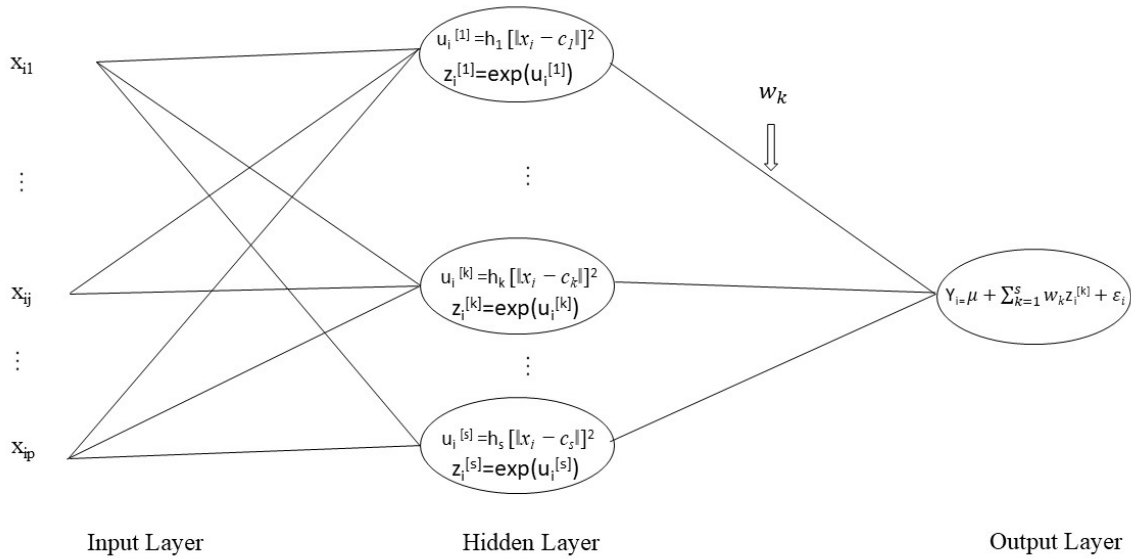
$$G_d = SS' / (\sigma_{md}^2 / \sigma_d^2) = SS' / \sum_{i=1}^n [2p_i(1 - p_i)]^2 \text{ as}$$

$$\sigma_d^2 = \sum_{i=1}^n [2p_i(1 - p_i)]^2 \sigma_{md}^2$$

## Radial Basis Function Neural Network (RBFNN)

A RBFNN is an artificial neural network that uses radial basis functions as activation functions. The RBFNN in the present study is a three layered feed-

forward neural network, where the first layer is linear and only distributes the input signal, while the next layer is nonlinear and uses Gaussian functions (Figure 1).



**Figure 1.** Structure of a radial basis function neural network (RBFNN).

In the hidden layer, each input vector ( $x_{i1}, \dots, x_{ip}$ ) is summarized by the Euclidean distance between the input vectors  $x_i$  and the centers  $c_m$ , ( $m = 1, \dots, M$ ) neurons, i.e.,  $h_m ||x_i - c_m||$ , where  $h_m$  is a band width parameter. Then distances are transformed by the Gaussian kernel  $\exp(-(h_m ||x_i - c_m||)^2)$  for obtaining the response,  $y_i = w_0 + \sum_{m=1}^M w_m z_{mi} + \epsilon_i$  (extracted from Gonzalez-Camacho et al., 2012).

The training of RBFNN optimization includes: the weights between the hidden layer and the output layer, the activation function, the center of activation functions, the distribution of center of activation functions, and the number of hidden neurons (Cruz and Nascimento, 2018). During the training process, only the weights between the hidden layer and the output layer are modified. The vector of weights  $\omega = \{w_1, \dots, w_s\}$  of the linear output layer is obtained using the ordinary least-squares fit that minimizes the mean squared differences between  $\hat{y}_i$  (from RBFNN) and the observed  $\hat{y}_i$  observed in the training set, provided that the Gaussian RBFs for centers  $c_k$  and  $h_k$  of the hidden layer are defined.

The radial basis function selected is usually a Gaussian kernel selected using K-means clustering algorithm. The centers are selected using the orthogonalization least-squares learning algorithm (Chen et al., 1991) and implemented in (Matlab, 2011). The centers are added iteratively such that each new selected center is orthogonal to the others. The selected centers maximize the decrease in the mean squared error of the RBFNN, and the algorithm stops when the number of centers (neurons) added to

the RBFNN attains a desired precision (goal error) or when the number of centers is equal to the number of input vectors, that is, when  $S = n$ .

To select the best RBFNN, a grid for training the net was generated, containing different spread values and different precision values (goal error). The spread value ranging from 5 to 100 and an initial value of 0.01 for the goal error was considered. The spread parameter allows adjusting the form of the Gaussian RBFNN such that it is sufficiently large to respond to overlapping regions of the input space but not so big that it could induce the Gaussian RBFNN to have a similar response (Asvadi et al., 2011).

## Comparison of methodologies under a GS approach

The models were compared using the accuracy, defined as the correlation between the GEBVs of the the true breeding values (TBV), and the mean squared error root (MSER). A five-fold cross-validation scheme was used to assess the performance of both genomic prediction methods. Specifically, the individuals (500) were randomly split into five equal-size groups and each group with about 100 individuals (20% of the population) was in turn assigned with phenotypic values and used as the validation set.

The simulations and the RBFNN was implemented using Genes (Cruz, 2016) software in integration with Matlab (Matlab, 2011). The additive and additive-dominance G-BLUP models were fit using Genomic Land software (Azevedo et al, 2019)

via REML through mixed model equations.

## Results and discussion

Table 2 displays the accuracy and meansquare error root (RMSE) values for the scenario without dominance. G-BLUP presented accuracy values higher than those obtained through RBFNN. Specifically, the accuracy values were equal to 0.58 and 0.77 for the low and high heritability scenarios, respectively (Table 2). Such an assessment, given by the correlation between predicted and actual values, is only possible due to the availability of simulated data and, according to Long et al. (2010), is the best measure to describe accuracy since the environment is not a disturbing factor. On the other hand, the prediction of genomic genetic values based on RBFNN shows lower MSER values when

compared to those obtained through G-BLUP for the two scenarios evaluated. According to Howards et al. (2014), for characteristics determined by additive genetic effects, SG models, such as G-BLUP, are proven to be efficient. Literature (Gianola et al., 2008; Gianola et al., 2011; Gonzalez-Camacho et al., 2012; Howards et al., 2014) has presented various studies aimed at comparing linear models (Bayesian LASSO - BLASSO, Bayesian ridge regression - BRR, Bayes A and Bayes B) and non-parametric models (reproducing kernel Hilbert space regression - RKHS, Bayesian regularized neural networks - BRNN, and radial basis function neural networks - RBFNN) and, in general, these studies have indicated the efficiency of linear models and/or a small improvement using non-parametric methods (Howard et al., 2014; Gonzalez-Recio et al., 2014).

**Table 2.** Accuracy and MSER in two scenarios of broad sense heritability 0.30 (d0h30) and 0.60 (d0h60) obtained through GBLUP and RBFNN in scenarios without dominance.

Models	Scenarios	Accuracy	MSER
GBLUP	d0h30	0.58 ± 0.04	8.71 ± 0.65
	d0h60	0.77 ± 0.05	9.91 ± 0.85
RBFNN	d0h30	0.31 ± 0.05	5.24 ± 0.29
	d0h60	0.65 ± 0.06	3.82 ± 0.35

In the scenario with dominance, the accuracy and mean squared error root (MSER) are presented in Table 3. G-BLUP, parametrized for dominance or not, showed accuracy values higher than those obtained through RBFNN in the dominance scenario. Specifically, in the low heritability scenario, the G-BLUP accuracy values were, respectively, for models parameterized for dominance or not, equal to 0.75 and 0.65 considering degree of dominance equal to  $d = 0.5$  (Table 3). In turn,

for dominance equal to 1 ( $d = 1$ ), the accuracy values, for models parameterized for dominance or not, were equal to 0.50 and 0.48. For RBFNN, the values were 0.31 and 0.25 for the low heritability scenarios. The same behavior was observed for the scenario with high heritability. On the other hand, for all the evaluated scenarios, the prediction of genomic genetic values based on RBFNN showed lower RMSE values when compared to those obtained through G-BLUP (Table 3).

**Table 3.** Accuracy and MSER in four scenarios of broad sense heritability 0.30 (d0.5h30; d1h30) and 0.6 (d0.5h60; d1h60) obtained through GBLUP and RBFNN in scenarios with dominance equal 0.5 or 1.

Models	Scenarios	Accuracy	MSER
G-BLUP	d0.5h30	0.65 ± 0.05	12.49 ± 1.18
	d1h30	0.48 ± 0.03	32.62 ± 2.58
	d0.5h60	0.65 ± 0.05	12.49 ± 1.14
	d1h60	0.47 ± 0.05	18.06 ± 1.48
G-BLUP dom	d0.5h30	0.75 ± 0.04	28.55 ± 1.41
	d1h30	0.50 ± 0.05	91.31 ± 5.64
	d0.5h60	0.75 ± 0.03	33.00 ± 2.31
	d1h60	0.66 ± 0.08	21.35 ± 1.98
RBFNN	d0.5h30	0.31 ± 0.05	5.47 ± 2.03
	d1h30	0.25 ± 0.06	14.55 ± 3.04
	d0.5h60	0.60 ± 0.05	4.03 ± 2.06
	d1h60	0.61 ± 0.03	4.27 ± 1.07

In general, the results showed a decrease in accuracy when dominance increased in both methods. This result differs from those obtained by some authors (Denis et al., 2011; Almeida-Filho et al., 2016; Santos et al., 2016; Toro and Varona, 2018), who reported that the incorporation of the dominance component in the linear model used lead to an improvement in the prediction process of complex traits. Concerning the RMSE values, for G-BLUP, the higher the influence of dominance, the higher the RMSE value obtained, and the adjustment of the model to the effects of dominance greatly diminishes the estimates of the error.

Overall, in terms of accuracy, G-BLUP presents better results compared to those obtained by RBFNN. However, such an approach is more biased. In practical terms, bias may not present itself as a problem since selection, in general, is based on ranking individuals. Finally, it is worth stressing that RBFNN comprises only one hidden layer, and networks that are more complex may be more suitable for the modeling of non-additive effects such as epistasis and dominance.

## Conclusion

Considering the values of accuracy, RBFNN produced lower results than those obtained by G-BLUP. On the other hand, RBFNN has low prediction bias. Furthermore, the adjustment of GBLUP to dominance models, despite increasing the complexity of the model, increases predictive accuracy.

## Acknowledgments

CAPES – Coordenação de Aperfeiçoamento de Pessoal de Nível Superior –, FAPEMIG – Fundação de Amparo à Pesquisa do Estado de Minas Gerais – and CNPq – Conselho Nacional de Desenvolvimento Científico e Tecnológico, for the financial support and the grant conceded.

## References

- ALMEIDA FILHO, J.E.; GUIMARÃES, J.F.R.; SILVA, F.F.; DE RESENDE, M.D.V.; MUÑOZ, P.; KIRST, M.; RESENDE JUNIOR, M.F.R. 2016. The contribution of dominance to phenotype prediction in a pine breeding and simulated population. **Heredity**,117(1).
- ASVADI A.; KARAMI M.; BALEGHI, Y. 2011. Efficient object tracking using optimized K-means segmentation and radial basis function neural networks. **International Journal of Information and Communication Technology Research**, 4(1):29-39.
- AZEVEDO, C.F.; NASCIMENTO, M.; FONTES, V.C.; SILVA, F.F.; RESENDE, M.D.V.R.; CRUZ, C.D. 2019. GenomicLand: Software for genome-wide association studies and genomic prediction. **Acta Scientiarum**, 41:e45361.
- AZEVEDO, C.F.; RESENDE, M.D.V.DE; E SILVA, F.F.; VIANA, J.M.S.; VALENTE, M.S.F.; RESENDE, M.F.R.; MUÑOZ, P. 2015. Ridge, Lasso and Bayesian additive-dominance genomic models. **BMC Genetics**,16(1).
- CHEN, S.; COWAN, C.F.; GRANT, P.M. 1991. Orthogonal least squares learning algorithm for radial basis function networks. **IEEE Transactions on neural networks**, 2(2):302-309.
- CROSSA, J.; PÉREZ-RODRÍGUEZ, P.; CUEVAS, J.; MONTESINOS-LÓPEZ, O.; JARQUÍN, D.; DE LOS CAMPOS, G.; BURGUEÑO, J.; CAMACHO-GONZÁLEZ, J.M.; PÉREZ-ELIZALDE, S.; BEYENE, Y.; DREISIGACKER, S. 2017. Genomic selection in plant breeding: methods, models, and perspectives. **Trends Plant Science**, 22(11):961-975.
- CRUZ, C.D.; NASCIMENTO, M. 2018. Inteligência Computacional aplicada ao melhoramento genético. 1st ed UFV,Viçosa, 414p.
- CRUZ, C.D. 2016. Genes Software ampliado e integrado com o R, Matlab e Selegen. **Acta Scientiarum Agronomy**, 38(4):547-552.

- DENIS, M.; BOUVET, J.M. 2012. Efficiency of genomic selection with models including dominance effect in the context of *Eucalyptus* breeding. **Tree Genetics Genomes**, 9:37–51.
- GIANOLA, D.; OKUT, H.; WEIGEL, K.A.; ROSA G.J.M. 2011. Predicting complex quantitative traits with neural networks: a case study with Jersey cows and wheat. **BMC Genetics**, 12(1).
- GIANOLA, D.; DE LOS CAMPOS, G. 2008. Inferring genetic values for quantitative traits non-parametrically. **Genetics Research**, 90(6):525-540.
- GONZÁLEZ-CAMACHO, J.M.; ORNELLA, L.; PÉREZ-RODRÍGUEZ, P.; GIANOLA, D.; DREISIGACKER, S.; CROSSA, J., 2018. Applications of machine learning methods to genomic selection in breeding wheat for rust resistance. **The plant genome**, 11:1–15.
- GONZÁLEZ-CAMACHO, J.M.; CROSSA, J.; PÉREZ-RODRÍGUEZ, P.; ORNELLA, L.; GIANOLA, D. 2016. Genome-enabled prediction using probabilistic neural network classifiers. **BMC genomics**, 17(1).
- GONZÁLEZ-CAMACHO, J.M.; DELOS CAMPOS, G.; PÉREZ, P.; GIANOLA, D.; CAIRNS, J.E.; MAHUKU, G.; BABU, R.; CROSSA, J. 2012. Genome-enabled prediction of genetic values using radial basis function neural networks. **Theoretical and Applied Genetics**, 125(4):759-771.
- GONZÁLEZ-RECIO, O.; ROSA, G.J.; GIANOLA, D. 2014. Machine learning methods and predictive ability metrics for genome-wide prediction of complex traits. **Livestock Science**, 166:217-231.
- HAYKIN, S. 2009. Neural Networks and Learning Machines. 3rd ed Prentice Hall, New York, 437p.
- HOWARD, R.; CARRIQUIRY, A.L.; BEAVIS, W.D. 2014. Parametric and nonparametric statistical methods for genomic selection of traits with additive and epistatic genetic architectures. **Genetic**, 4(6):1027-1046.
- LIANG, Q.; SHANG, L.; WANG, Y.; HUA, J. 2015. Partial dominance, overdominance and epistasis as the genetic basis of heterosis in upland cotton (*Gossypium hirsutum* L.). **PLoS One**, 10(11): p.e0143548.
- LIU, R.; WANG, B.; GUO, W.; QIN, Y.; WANG, L.; ZHANG, Y.; ZHANG, T. 2012. Quantitative trait loci mapping for yield and its components by using two immortalized populations of a heterotic hybrid in *Gossypium hirsutum* L. **Molecular breeding**, 29(2):297-311.
- LONG, N.; GIANOLA, D.; ROSA, G.J.; WEIGEL, K.A. 2011. Marker-assisted prediction of non-additive genetic values. **Genetica**, 139(7):843-854.
- LONG, N.; GIANOLA, D.; ROSA, G.J.; WEIGEL, K.A.; KRANIS, A.; GONZALEZ-RECIO, O. 2010. Radial basis function regression methods for predicting quantitative traits using SNP markers. **Genetics research**, 92(3):209-225.
- MATLAB. 2011. Matlab Version 7.10.0. Natick, Massachusetts: The Math Works Inc.
- MEUWISSEN, T.H.E.; HAYES, B.J.; GODDARD, M.E. 2001. Prediction of total genetic values using genome-wide dense marker maps. **Genetics**, 157(4):1819-1829.
- PÉREZ-RODRÍGUEZ, P.; GIANOLA, D.; GONZÁLEZ-CAMACHO, J.M.; CROSSA, J.; MANÈS, Y.; DREISIGACKER, S. 2012. Comparison between linear and non-parametric regression models for genome-enabled prediction in wheat. **G3:Genes, Genomes, Genetics**, 2(12):1595-1605.
- R CORE TEAM. 2014. R: A language and environment for statistical computing. Vienna: R Foundation for Statistical Computing, 2017. Available at: <<https://www.R-project.org/>>

- 
- SANT'ANNA, I. C.; SILVA, G.N.; NASCIMENTO, M; CRUZ, C.D. 2018. Subset selection of markers for genome-enabled prediction of genetic values using radial basis function neural networks. **Acta Scientiarum Agronomy**, Forthcoming.
- SANTOS, V.S.; MARTINS FILHO, S.; DE RESENDE, M.D.V.; AZEVEDO, C.F.; LOPES, P.S.; GUIMARÃES, S.E.F.; SILVA, F.F. 2016. Genomic prediction for additive and dominance effects of censored traits in pigs. **Genetics and molecular research**, 15(4).
- TECHNOW, F.; RIEDELSHEIMER, C.; SCHRAG, T.A.; MELCHINGER, A.E. 2012. Genomic prediction of hybrid performance in maize with models incorporating dominance and population specific marker effects. **Theoretical and Applied Genetics**, 125(6):1181-1194.
- TORO, M.A.; VARONA, L. 2010. A note on mate allocation for dominance handling in genomic selection. **Genetics Selection Evolution**, 42(1).
- VIANA, J.M.S.; PIEPHO, H.P. 2017. Quantitative genetics theory for genomic selection and efficiency of genotypic value prediction in open-pollinated populations. **Scientia Agricola**, 74(1):41-50.
- VITEZICA, Z.G.; LEGARRA, A.; TORO, M.A.; VARONA, L. 2017. Orthogonal estimates of variances for additive, dominance, and epistatic effects in populations. **Genetics**, 206(3):1297-1307.