FUNCTIONAL PLANT BREEDING JOURNAL
Scientific Journal. ISSN 2595-9433
Volume 4, Number 1, Article n. 6, January/June
D. O. I. http://dx.doi.org/10.35418/2526-4117/v4n1a6
Received: 08/10/2021 - Accepted: 22/10/2021

DETECTING EPISTATIC EFFECTS UNDER A QUALITATIVE CONTEXT: IMPORTANCE AND QUANTIFICATION

Iara Gonçalves dos Santos[1], Pedro César de Oliveira Ribeiro[1], Odimar Ferreira de Almeida[1], Vinícius Quintão Carneiro[2], Cosme Damião Cruz[1]

1 Federal University of Viçosa, Viçosa, MG, Brazil;
2 Federal University of Lavras, Lavras, MG, Brazil.
** Corresponding author: Iara Gonçalves dos Santos (iara.santos@ufv.br).*

Abstract: The objectives of this work were to simulate and quantify epistatic effects on oligogenic traits and to verify the efficiency of Artificial Neural Networks (ANN) and Ridge Regression Best Linear Unbiased Predictor (RRBLUP) in the prediction of genetic values of oligogenic traits controlled by epistatic genes. We simulated 10 F_2 populations in Hardy-Weinberg equilibrium with 800 individuals, each. The individuals were genotyped with 105 codominant markers equidistantly distributed along 10 chromosomes. Genotypic values were simulated for each individual considering five oligogenic traits controlled by two biallelic loci, according to five different epistatic models: duplicate recessive genes, dominant and recessive interaction, duplicate dominant genes, recessive epistasis, and dominant epistasis. RRBLUP, as well as an ANN, were used to perform genomic selection. The coefficient of determination of the regression model revealed a mean epistatic effect ranging from 13.3% in the duplicate recessive genes to 62.5% in the duplicate dominant genes. The identification of epistatic genes was superior in the ANN model compared with the RRBLUP approach. Our result reinforces the potential of ANN in predicting genetic values in situations where other than linear or quadratic relationships are present. The results presented in this work offer important insights about the exploration of epistasis in the qualitative context. In situations where epistatic effects are completely ignored, they can play a role on genetic values, as seen for the duplicative dominant genes.

Keywords: oligogenic traits, epistasis, RRBLUP, artificial neural networks.

Introduction

One of the biggest challenges plant breeding faces is the understanding on how genes interact to cause phenotypic changes (Mackay, 2014). Phenotypic expression relies on linear and non-linear gene interactions (Carlborg and Haley, 2004). Bateson (1909) was the first to use the term epistasis to describe a situation where the action of a locus masks the effect at another locus, leading to deviations from the classical segregation ratio. From the qualitative standpoint, epistasis regards gene interaction involving non-allelic loci in which the final phenotypic expression depends on this interaction (Carlborg and Haley, 2004).

From the quantitative standpoint, however, epistasis is a multiplicative effect among loci in which the final phenotypic effect cannot be predicted by the sum of the individual allelic effects (Cheverud and Routman, 1995; Mackay, 2014). Epistasis between two loci is the phenotypic change caused by additive and dominance effects (Figure 1). The epistatic model for a diploid organism with alleles A1 and A2 in locus A and B1 and B2 in locus B is

$$Y_{A_1A_2B_1B_2} = \alpha_{A_1} + \alpha_{A_2} + \alpha_{B_1} + \alpha_{B_2} + \delta_{A_1} + \delta_{A_2} + \delta_{B_1} + \delta_{B_2} + \alpha\alpha + \alpha\delta + \delta\alpha + \delta\delta$$

in which $Y_{A_1A_2B_1B_2}$ represents the phenotypic value of individual A1A2B1B2; α_{A_1}, α_{A_2}, α_{B_1}, α_{B_2} are additive values of A1, A2, B1, and B2 alleles, respectively; δ_{A_1}, δ_{A_2}, δ_{B_1}, δ_{B_2} are dominance values of the alleles A1, A2, B1 e B2, respectively; $\alpha\alpha$, $\alpha\delta$, $\delta\alpha$, $\delta\delta$ are the additive x additive, additive x dominant, dominant x additive, and dominant x dominant epistatic interactions.

i)

	B1B1	B1B2	B2B2
A1A1	$2\alpha_{A_1} + 2\alpha_{B_1} + 2\delta_{A_1} + 2\delta_{B_1} + \alpha\alpha + \alpha\delta + \delta\alpha + \delta\delta$	$2\alpha_{A_1} + \alpha_{B_1} + \alpha_{B_2} + 2\delta_{A_1} + \delta_{B_1} + \delta_{B_2} + \alpha\alpha + \alpha\delta + \delta\alpha + \delta\delta$	$2\alpha_{A_1} + 2\alpha_{B_2} + 2\delta_{A_1} + 2\delta_{B_2} + \alpha\alpha + \alpha\delta + \delta\alpha + \delta\delta$
A1A2	$\alpha_{A_1} + \alpha_{A_2} + 2\alpha_{B_1} + \delta_{A_1} + \delta_{A_2} + 2\delta_{B_1} + \alpha\alpha + \alpha\delta + \delta\alpha + \delta\delta$	$\alpha_{A_1} + \alpha_{A_2} + \alpha_{B_1} + \alpha_{B_2} + \delta_{A_1} + \delta_{A_2} + \delta_{B_1} + \delta_{B_2} + \alpha\alpha + \alpha\delta + \delta\alpha + \delta\delta$	$\alpha_{A_1} + \alpha_{A_2} + 2\alpha_{B_2} + \delta_{A_1} + \delta_{A_2} + 2\delta_{B_2} + \alpha\alpha + \alpha\delta + \delta\alpha + \delta\delta$
A2A2	$2\alpha_{A_2} + 2\alpha_{B_1} + 2\delta_{A_2} + 2\delta_{B_1} + \alpha\alpha + \alpha\delta + \delta\alpha + \delta\delta$	$2\alpha_{A_2} + \alpha_{B_1} + \alpha_{B_2} + 2\delta_{A_2} + \delta_{B_1} + \delta_{B_2} + \alpha\alpha + \alpha\delta + \delta\alpha + \delta\delta$	$2\alpha_{A_2} + 2\alpha_{B_2} + 2\delta_{A_2} + 2\delta_{B_2} + \alpha\alpha + \alpha\delta + \delta\alpha + \delta\delta$

ii)

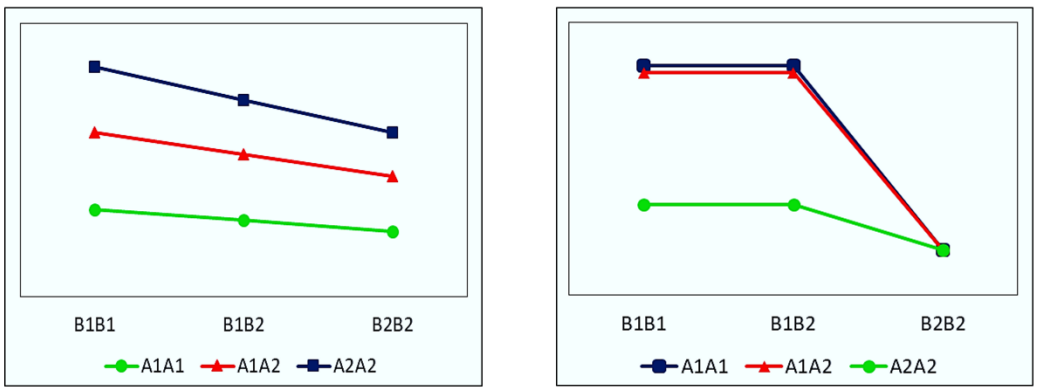

Figure 1. Two-gene model to describe the genetic value of an individual (i). Description of the allelic interactions of gene A and B (ii). Left-hand figure represents the additive gene action for loci A and B in the absence of dominance deviation and epistasis. The right-hand figure represents recessive epistasis (9:4:3).

The contribution of non-additive effects has often been neglected in the estimation of genetic variance (Crow, 2010). However, epistasis seems to have major importance in the study of qualitative traits. Examples of epistatic effects are fur color in animals, Bombain phenotype in ABO blood group in humans, crest phenotype in chicken (Carlborg and Haley, 2004). Examples in plant breeding include the finding of Krishna et al. (2018). The authors

found out statistical significance for epistatic effects in the study of traits related to quality of rice. They highlighted the importance of exploring epistatic variance when studying the genetic behind the quality of rice. Epistasis can contribute to a better understanding of genetic events, the functional relationship among genes, on the sort of biosynthetic routes as well as in identifying quantitative differences caused by effects of specific alleles (Phillips, 2008).

In theory, additive variance accounts for the major fraction of genetic variance (Hill et al., 2008). Ávila et al. (2014) found out small contributions of epistatic variance on multiloci models. There are situations where epistasis cannot be ignored though. Singh et al. (2011) stated that epistatic interaction ate responsible for wheat leaf rust resistance. The *Sr25* gene only has phenotypic expression when *Sr2* is present. In general, we consider important to quantify the epistatic proportion in the genetic variance. The highest probability of epistatic variance being identified happens when the allelic frequencies of the interacting loci are intermediates (Mackay, 2014).

Genomic selection (GS) has been an important strategy for predicting genetic values of individuals (Varona et al., 2018). Including epistatic effects in GS models is not straightforward due to the high dimensionality of data (McKinney et al., 2006; Desta and Ortiz, 2014) and also by the difficulty of modeling so many additional effects. Among the most common GS models are the Ridge Regression Best Linear Unbiased Predictor (RRBLUP), Genomic Best Linear Unbiased Predictor (GBLUP), Bayesian approaches (Bayes A, Bayes B, Bayes Cπ, etc), and lately, computational intelligence (CI) methods (Crossa et al., 2017). Each method has peculiarities regarding the marker's variance and the gene action (Salla et al., 2015). RRBLUP assumes marker's effects as random covariates that account for the same genetic variance (Meuwissen et al., 2001). This method has been broadly applied in GS due to its high stability and accuracy (Heslot et al., 2012; Zhao et al., 2015). The output of RRBLUP is on a marker basis, the importance of each marker is released with satisfactory robustness and low computational effort (Zhao et al., 2015).

Even though linear approaches present satisfactory solutions, they have limitations when non-additive effects are included. Using CI methods can be more efficient since they capture all kinds of effects (González-Camacho et al., 2018). The upside of CI methods is either the non-need for any statistical modeling and no assumption regarding data distribution (Zhou et al., 2015).

Among the vast range of CI methods, Artificial Neural Networks have been used to predict quantitative traits (Motsinger-Reif et al., 2008; Beam et al., 2014; Zhou et al., 2015; Crossa et al., 2017).The phenotypic prediction of ANN for GS is usually made by simple networks, using one hidden layer (González-Camacho et al, 2018). The input layer holds the marker vector, the hidden layer has a variable number of neurons (each neuron has an activation function), and the output layer has the predicted values.

The objectives of the work were (i) to simulate and quantify epistatic effects on oligogenic traits and (ii) to verify the efficiency of ANNs and RR-BLUP in the prediction of genetic values of oligogenic traits controlled by epistatic genes.

Material and methods

We simulate 10 F_2 populations in Hardy-Weinberg equilibrium with 800 individuals, each. The individuals were genotyped with 105 codominant markers equidistantly distributed along 10 chromosomes. Codes 1, 0, or -1 represented dominant homozygotes, heterozygotes, and recessive homozygotes individuals, respectively. Genotypic values were simulated for each individual considering five oligogenic traits controlled by two biallelic loci, according to five different epistatic models (Table 1). The environmental effect was not considered in the model as the work aimed to verify the capacity of different genomic prediction models to detect epistasis. Heritability was 1 for all traits *i.g.* genetic value is equal to phenotypic value.

Table 1. Segregation and genetic values of the five simulated epistatic traits. Trait 1 represents the duplicate recessive genes, 2 is the dominant and recessive interaction, 3 is the duplicate dominant genes, 4 is the recessive epistasis, and 5 is the dominant epistasis.

Trait	Segregation	Genetic value
1	9:7	A_B_ (4)*; A_bb (2); aaB_ (2); aabb (2)
2	13:3	A_B_ (4); A_bb (4); aaB_ (2); aabb (4)
3	15:1	A_B_ (4); A_bb (4); aaB_ (4); aabb (2)
4	9:4:3	A_B_ (4); A_bb (3); aaB_ (2); aabb (3)
5	12:3:1	A_B_ (4); A_bb (3); aaB_ (4); aabb (2)

* Genetic values for genotypes are shown in parentheses.

The proportion of epistasis presented in each trait was quantified by the linear regression analysis through the coefficient of determination (R^2). Each marker controlling the trait was considered as independent variable while the genetic value, as dependent variable. Thus, it was possible to verify which trait had the genomic prediction most influenced by epistasis.

We use a stochastic approach (RRBLUP) as well as a CI method (ANN) for performing genomic selection. The RRBLUP model is described as follows

$$Y = \mu + Wu + \varepsilon$$

where Y is vector of phenotypic values,μis the overall mean, W is the design matrix for random effects, u is the random effect with u ~ N $(0, I\sigma_u^2)$, ε is the vector of random residuals following distribution N $(0, I\sigma_\varepsilon^2)$. The BLUP solution for marker effects is

$$\hat{u} = Z'(ZZ' + \lambda I)^{-1}$$

Where Z = Wu and λ is the residual variances and marker effect ratio $(\lambda = \sigma_\varepsilon^2/\sigma_u^2)$.

The ANN consisted of a multilayer perceptron architecture with one input, one hidden, and one output layer (Figure 2). Each marker consisted of one input in the input layer. Markers were standardized to the interval from 0 to 1 to allow greater computational efficiency. The hidden layer had two neurons, similar to Gloria et al. (2016). Neurons in the hidden layer carried a logistic function that maps the real internal for the internal from 0 to 1, as follows

$$f(u_{ij}) = \frac{1}{(1 + e^{-u_{ij}})}$$

Where u_{ij} is the activation potential of the ith neuron given by $u_i = (\sum_{j=1}^{n} x_j w_{ij}) - w_{i0}$, in which x_{ij} is the value of the ith input (marker), w_{ij} is the weight of the ith neuron of the jth input, and w_{i0} is the activation threshold of the hidden layer. The output layer had a single neuron, with an identity function $f(u_j) = u_j$, representing a linear combination of all outputs in the hidden layer.

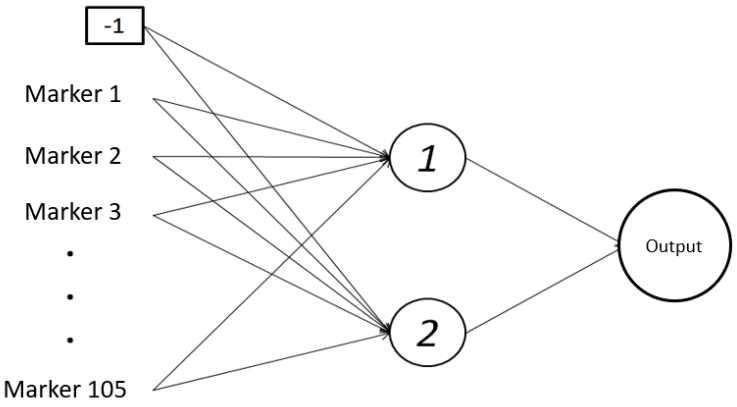

Figure 2. ANN architecture for genomic prediction of epistatic traits in F_2 populations.

The training algorithm back propagation with bayesian regularization was considered for training the ANN with the maximum number of 5,000 iterations. In order to evaluate the training efficiency, we used the mean squared error, given by

$$MSE = \frac{1}{n}\sum_{k=1}^{n}(\hat{y}_j - y_j)^2$$

Where k represents the individuals of the F_2 populations varying from 1 to 800; $\hat{y}_j$ is the vector of predicted genetic values of jth individual, and y_j is the real genotypic value of the jth individual. We considered an MSE lower than 0.01 as a threshold for stopping the training in a given iteration. If the iteration would not achieve the threshold, the training would stop after all iterations.

K-fold (Bengio and Grandvalet, 2004) sampling technique was used to partitioning the data for the RRBLUP approach and to estimate weights in the ANN method. The 800 individuals of each F_2 population where partitioned in 10 sets (k = 10), i.g. training and validation data were composed by 720 and 80 individuals, respectively. Genetic values were obtained for each method in each partitioning. In order to evaluate the accuracy for RRBLUP and ANN in each of the k folds, we use the squared correlation between real and the predicted genetic value ($r^2_{\hat{y}y}$) for both training and validation steps.

The data was simulated by the software GENES (Cruz, 2016). RRBLUP genomic prediction was performed by the software GENES integrated with R (R Core Team, 2018) using the package rrBLUP (Endelman, 2011). ANN prediction was performed by the software GENES integrated with Matlab (Matlab, 2011).

Results and Discussion

The coefficient of determination of the regression model revealed a mean epistatic effect of 13.3% in the ten F_2 populations for the trait governed by the duplicate recessive genes (Table 2). This interaction happens when there is a recessive allele masking the expression of dominant alleles at two loci.

This type of epistatic interaction has been reported in plant disease resistance trials, as observed by Devi et al. (2019) in a study on the genetic control of disease caused by the TYLCV virus in okra. Elmassry et al. (2020) identified duplicate recessive genes in crosses between wheat cultivars to assess resistance to yellow rust in adult stage. The authors crossed Gemmeiza 11 x Giza 168 and obtained a 100% susceptible F_1 population. The F_2 population presented 45.33% of susceptible plants and 54.67% of resistant plants, equivalent to the proportion of 9 resistant: 7 susceptible.

Table 2. Proportion of duplicate recessive genes (9:7), dominant and recessive interaction (13:3), duplicate dominant genes (15:1), recessive epistasis (9:4:3), and dominant epistasis (12:3:1) in the genetic values of 10 F_2 populations obtained by the coefficient of determination of the linear regression.

Population	Epistasis				
	9:7	13:3	15:1	9:4:3	12:3:1
$F_2$1	0.134296	0.200825	0.641018	0.207465	0.101699
$F_2$2	0.130823	0.219018	0.614684	0.211796	0.101143
$F_2$3	0.133791	0.220776	0.637102	0.220899	0.104306
$F_2$4	0.133246	0.192122	0.613383	0.196586	0.097732
$F_2$5	0.126035	0.201696	0.646886	0.200446	0.100796
$F_2$6	0.124141	0.188429	0.653169	0.190147	0.098663
$F_2$7	0.126587	0.194825	0.656251	0.197682	0.100554
$F_2$8	0.147899	0.242029	0.576721	0.240409	0.103084
$F_2$9	0.146666	0.239142	0.599452	0.241572	0.105083
$F_2$10	0.139084	0.233363	0.600726	0.228448	0.103142
Mean	**0.133519**	**0.210357**	**0.625893**	**0.209631**	**0.101619**

Dominant and recessive interaction and recessive epistasis explained about 21% of the genetic value of the populations, each. In this type of epistasis, a dominant allele at one locus can mask the expression of both (dominant and recessive) alleles at second locus. This is also known as inhibitory gene interaction and it has been identified in the control of stem rust of wheat (Ellis et al., 2014). The recessive epistasis happens when one dominant gene has its own phenotypic effect and other dominant gene has no effect of its own but its presence with the first gene modified the phenotypic expression. This interaction was identified by Vandemark et al. (2008) in a study of common bacterial blight resistance in dry bean.

Duplicate dominant genes explained 62.5% of the genetic value in the populations. This interaction happens when a dominant allele at either of two loci can mask the expression of recessive alleles at the two loci. Finally, the dominant epistasis explained about 10% of the genetic value in the F_2 populations. Saroj et al. (2015) identified both duplicate dominant genes and dominant epistasis in cowpea F_2 populations.

Estimated values for the different types of epistasis ranged from 0.09 in the dominant epistasis to 0.65 in the duplicate dominant genes. Values of this magnitude indicate a great relevance of studying epistatic gene interactions, especially in the qualitative context, such as in disease control. Approaches that detect the magnitude of epistatic effects are not common. Viana and Franco-Garcia (2021) estimated the proportion between epistatic and genotypic variance and stated that the relationship is proportional to the percentage of epistatic genes. The authors also found estimates of epistatic variance due to duplicative genes of high magnitude (10 to 64%) in different generations of random mating and self-fertilization. Their results agree with those presented in this work.

The prediction of genetic values of epistatic traits was obtained by RRBLUP and RNA techniques in order to identify the simulated epistatic genes. RRBLUP was able to capture the effect of epistatic genes in 61% and 55% of individuals from the F_2 populations in the training and validation stages, respectively, for the trait with duplicate recessive genes epistasis (Table 3).

For dominant and recessive epistasis (13: 3), 56% of the individuals in the populations had the epistatic genes identified in the training stage and 50% in the validation. For the duplicate dominant genes (15: 1), 28% and 21% of the individuals had the genes detected in the training and validation, respectively. The percentage of detection for recessive epistasis (9: 4: 3) was 57% in the training and 50% in the validation. Whereas for dominant epistasis (12: 3: 1), 63% of the individuals had the epistatic genes detected in the training and 56% in the validation.

The identification of epistatic genes was superior in the ANN model than it was in the RRBLUP approach. The detection ranged from 99 to 100% for all traits in both training and validation for the ANN (Table 3). Our result reinforces the potential of ANN in predicting genetic values in situations where other than linear or quadratic relationships are present. Regardless of the contribution of epistasis to genetic values, ANN correctly detected the epistatic genes, the same did not happen for RRBLUP. For traits with a low epistatic contribution (12:3:1 and 9:7), the values for training and validation of the RRBLUP model reached values ranging from 50 to 60%. However, when epistasis had the highest contribution for the genetic value (15:1), the technique reached very low values both in training (22 - 35%) and validation (12 - 33%).

The results presented in this work offer important insights about the exploration of epistasis in the qualitative context. In situations where epistatic effects are completely ignored, they can play a role on genetic values, as seen for the duplicative dominant genes (15:1). The result obtained in this simulation study reinforces that there are situations in which genomic prediction using stochastic models is limited. Future investigations involving real data will confirm the indications that the use of CI models is a powerful alternative for further investigation and exploration of different types of epistasis.

Table 3. Proportion of markers expressing 9:7, 13:3, 15:1, 9:4:3, and 12:3:1 epistasis detected by RRBLUP and ANN.

Population	9:7				13:3			
	RRBLUP		ANN		RRBLUP		ANN	
	Training	Validation	Training	Validation	Training	Validation	Training	Validation
$F_2$1	0.59988	0.52591	0.99997	0.99997	0.58555	0.56310	0.99997	0.99998
$F_2$2	0.62277	0.54828	0.99999	0.99999	0.55045	0.39765	0.99998	0.99998
$F_2$3	0.60575	0.55058	0.99997	0.99997	0.55785	0.48830	0.99999	0.99998
$F_2$4	0.59555	0.57018	0.99998	0.99998	0.59311	0.52330	0.99998	0.99997
$F_2$5	0.61780	0.52609	0.99996	0.99996	0.57005	0.51201	0.99999	0.99998
$F_2$6	0.61874	0.52279	0.99998	0.99997	0.60873	0.46550	0.99998	0.99999
$F_2$7	0.62276	0.49536	0.99999	0.99998	0.61523	0.46559	0.99999	0.99997
$F_2$8	0.58605	0.59464	0.99998	0.99998	0.53132	0.51066	0.99998	0.99998
$F_2$9	0.64960	0.57134	0.99998	0.99998	0.54997	0.51880	0.99997	0.99995
$F_2$10	0.57804	0.58828	0.99999	0.99999	0.54741	0.41768	0.99997	0.99999

Population	15:1				9:4:3			
	RRBLUP		ANN		RRBLUP		ANN	
	Training	Validation	Training	Validation	Training	Validation	Training	Validation
$F_2$1	0.31316	0.15397	0.99995	0.99993	0.56684	0.51062	0.99999	0.99999
$F_2$2	0.24171	0.33329	0.99996	0.99998	0.57698	0.44623	0.99999	0.99999
$F_2$3	0.26554	0.22268	0.99996	0.99994	0.54966	0.49556	0.99998	0.99998
$F_2$4	0.29863	0.21618	0.99985	0.99988	0.57823	0.54579	0.99999	0.99998
$F_2$5	0.26530	0.20296	0.99992	0.99992	0.57389	0.50042	0.99997	0.99997
$F_2$6	0.22207	0.18697	0.99999	0.99998	0.60527	0.46079	0.99995	0.99995
$F_2$7	0.27924	0.20337	0.99991	0.99991	0.59813	0.44633	0.99997	0.99996
$F_2$8	0.35834	0.27016	0.99992	0.99992	0.52439	0.54178	0.99996	0.99996
$F_2$9	0.32101	0.12247	0.99997	0.99996	0.57216	0.53751	1.00000	0.99997
$F_2$10	0.27456	0.28827	0.99996	0.99996	0.52615	0.50225	0.99998	0.99998

Population	12:3:1			
	RRBLUP		ANN	
	Training	Validation	Training	Validation
$F_2$1	0.64038	0.57353	0.99989	0.99991
$F_2$2	0.61515	0.57536	0.99991	0.99992
$F_2$3	0.62027	0.58848	0.99991	0.99985
$F_2$4	0.63746	0.60239	0.99992	0.99993
$F_2$5	0.61594	0.53694	0.99996	0.99997
$F_2$6	0.63340	0.52768	0.99996	0.99996
$F_2$7	0.64049	0.54868	0.99987	0.99986
$F_2$8	0.64072	0.58556	0.99998	0.99998
$F_2$9	0.62997	0.57395	0.99997	0.99997
$F_2$10	0.62178	0.56742	0.99995	0.99994

Conclusion

Epistasis can play a role on the genetic control of traits governed by a few genes. ANN models are superior than RRBLUP to identify epistatic genes in a qualitative stand point.

Acknowledgments

The authors thank the CNPq. This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001.

References

ÁVILA, V.; PÉREZ-FIGUEROA, A.; CABALLERO, A.; HILL, W.G.; GARCÍA-DORADO, A.; LÓPEZ-FANJUL, C. 2014. The action of stabilizing selection, mutation and drift on epistatic quantitative traits. **Evolution**, 68(7):1974–1987

BATESON, W. 1909. Mendel's Principles of Heredity. Cambridge University Press, Cambridge.

BEAM, A.L.; MOTSINGER-REIF, A.; DOYLE, J. 2014. Bayesian neural networks for detecting epistasis in genetic association studies. **BMC Bioinformatics**, 15(1):368-380.

CARLBORG, Ö.; HALEY, C.S. 2004. Epistasis: too often neglected in complex trait studies. **Nature Review Genetics**, 5:618-625.

BENGIO, Y.; GRANDVALET, Y. 2004. No unbiased estimator of the variance of k-fold cross-validation. **Journal of Machine Learning Research**, 5:1089-1105.

CHEVERUD, J.M.; ROUTMAN, E.J. 1995. Epistasis and its contribution to genetic variance components. **Genetics**, 139(3):1455–1461.

CROSSA, J.; PÉREZ-RODRIGUEZ, P.; CUEVAS, J.; MONTESINOS-LÓPEZ, O.; JARQUÍN, D.; DE LOS CAMPOS, G.; BURGUEÑO, J.; GONZÁLEZ-CAMACHO, J.M.; PÉREZ-ELIZALDE, S.; BEYENE, Y.; DREISIGACKER, S.; SINGH, R.; ZHANG, X.; GOWDA, M.; ROORKIWAL, M.; RUTKOSKI, J.; VARSHNEY, R.K. 2017. Genomic Selection in Plant Breeding: Methods, Models and Perspectives. **Trends in Plant Science**, 22(11):961-975.

CROW, J.F. 2010. On epistasis: why it is unimportant in polygenic directional selection. Philosophical Transactions of the Royal Society B: **Biological Sciences**, 365(1544):1241-1244.

CRUZ, C.D. 2016. Genes Software – extended and integrated with the R, Matlab and Selegen. **Acta Scientiarum Agronomy**, 38(4):547-552.

CRUZ, C.D. 2006. Programa genes: análise multivariada e simulação. Editora UFV, Viçosa.

DESTA, Z.A.; ORTIZ, R. 2014.Genomic selection: genome-wide prediction in plant improvment. **Trends in Plant Science**, 19(9):592-601.

DEVI, A.P.; BANERJEE, S.; BHATTACHARJEE, T.; MAURYA, P.K.; MANDAL, A.K.; DUTTA, S.; CHATTOPADHYAY, A. 2019. Genetic control of enation leaf curl virus disease tolerance in cultivated species of okra. **International Journal of Vegetable Science**, 26(2):150-162.

ELMASSRY, E.L.; DARWISH, M.A.H.; ELKOT, A.F. 2020; Inheritance of Stripe Rust Resistance in Adult Plants and Some Economic Traits of Some Bread Wheat Crosses. **Journal of Plant Production**, 11(12):1587-1596.

ELLIS, J.G.; LAGUDAH, E.S.; SPIELMEYER, W.; DODDS, P.N. 2014. The past, present and future of breeding rust resistant wheat. **Frontiers in Plant Sciences**, 5:1-13.

ENDELMAN, J.B. 2011. Ridge Regression and Other Kernels for Genomic Selection with R Package rrBLUP. **Plant Genome**, 4(3):250-255.

GLÓRIA, L.S.; CRUZ, C.D.; VIEIRA, R.A.M.; RESENDE, M.D.V.; LOPES, P.S.; SIQUEIRA, O.H.G.B.D.; SILVA, F.F. 2016.Accessing marker effects and heritability estimates from genome prediction by Bayesian regularized neural networks. **Livestock Science**, 191:91-96.

GONZÁLEZ-CAMACHO, J.M.; ORNELLA, L.; PÉREZ-RODRIGUEZ, P.; GIANOLA, D.; GREISIGACKER, S.; CROSSA, J. 2018. Applications of Machine Learning Methods to Genomic Selection in Breeding Wheat for Rust Resistance. **Plant Genome**, 11(2):170104.

HESLOT, N.; YANG, H.P.; SORRELLS, M.E.; JANNINK, J.L. 2012. Genomic selection in plant breeding: a comparison of models. **Crop Science**, 52(1):146-160.

HILL, W.G.; GODDARD, M.E.; VISSCHER, P.M. 2008. Data and theory point to mainly additive genetic variance for complex traits. **PLoS Genetics**, 4(2):e1000008.

KRISHNA, L.; RAJU, C.S.; KUMAR, S.S.; BHADRU, J.; MOHAN, Y.C. 2018. Genetics of Quality and Yield Traits using Aromatic and Non Aromatic Genotypes through Generation Mean Analysis in Rice (*Oryza sativa* L.). **International Journal of Current Microbiology and Applied Science**, 7(1):2907-2914.

MACKAY, T.F.C. 2014. Epistasis and quantitative traits: using model organisms to study gene–gene interactions. **Nature Reviews**, 15(1):22-33.

MATHWORKS. 2011. MATLAB – The Language of Technical Computing (R2011a), Natick. http://www.mathworks.com/products/matlab

MCKINNEY, B.A.; REIF, D.M.; RITCHIE, M.D.; MOORE, J.H. 2006. Machine learning for detecting gene-gene interactions: A review. **Applied Bioinformatics**, 5(2):77–88.

MEUWISSEN, T.H.E.; HAYES, B.J.; GODDARD, M.E. 2001. Prediction of total genetic value using genome-wide dense marker maps. **Genetics**, 157(4):1819–1829.

MOTSINGER-REIF, A.A.; DUDEK, S.M.; HAHN, L.W.; RITCHIE, M.D. 2008. Comparison of approaches for machine-learning optimization of neural networks for detecting gene-gene interactions in genetic epidemiology. **Genetic Epidemiology**, 32(4):325–340.

PHILLIPS, P.C. 2008. Epistasis - the essential role of gene interactions in the structure and evolution of genetic systems. **Nature Reviews Genetics**, 9(11):855–867.

R DEVELOPMENT CORE TEAM. 2018. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna. https://www.R-project.org/.

SALLA, A.H.; ENDELMAN, J.B.; JANNINK, J.L.; SMITH, K.P. 2015. Assessing Genomic Selection Prediction Accuracy in a Dynamic Barley Breeding Population. **The Plant Genome**, 8(1):1-15.

SAROJ, S.K.; SINGH, M.N.; VISHWAKARMA, M.K.; SINGH, T.; MISHRA, V.K. 2015. Identification of stable restorers and genetics of fertility restoration in late-maturing pigeonpea [*Cajanus cajan* (L.) Millspaugh]. **Plant Breeding**, 134(6):696-702.

SINGH, R.P.; HODSON, D.P.; HUERTA-ESPINO, J.; JIN, Y.; BHAVANI, S.; NJAU, P.; HERRERA-FOESSEL, S.; SINGH, P.K.; SINGH, S.; GOVINDAN, V. 2011. The emergence of *Ug99* races of the stem rust fungus is a threat to world wheat production. **Annual Review of Phytopathology**, 49:465–481.

VANDEMARK, G.J.; FOURIE, D.; MIKLAS, N.P. 2008.Genotyping with real-time PCR reveals recessive epistasis between independent QTL conferring resistance to common bacterial blight in dry bean. **Theoretical and Applied Genetics**, 117:513-522.

VARONA, L.; LEGARRA, A.; TORO, M.A.; VITEZICA, Z.G. 2018. Non-additive Effects in Genomic Selection. **Frontiersin Genetics**, 9(78):1-12.

VIANA, J.M.S.; FRANCO-GARCIA, A.A. 2021. Significance of linkage disequilibrium and epistasis on the genetic variances and covariance between relatives in non-inbred and inbred populations. bioRxiv.

ZHAO, Y.; METTE, M. F.; REIF, J.C. 2015. Genomic selection in hibrid breeding. **Plant Breeding**, 134:1-10.

ZHOU, Z.; LIU, G.; SU, L.; TIAN, Y. 2015. Detect epistasis based on optimization of artificial neural network architecture and parameters. **ICIC Express Letters**, 9(5):1499-1506.