

# Parental selection based on molecular information under a population genetics approach

Renato Domiciano Silva Rosado<sup>1,\*</sup> , Ana Maria Cruz e Oliveira<sup>2</sup> , Iara Gonçalves dos Santos<sup>2</sup> , Pedro Crescêncio Souza Carneiro<sup>2</sup> , Cosme Damião Cruz<sup>1</sup>  and Paulo Roberto Cecon<sup>1</sup> 

<sup>1</sup>Universidade Federal de Viçosa (UFV), Departamento de Estatística, CEP 36570-900 Viçosa, MG, Brazil. <sup>2</sup>Universidade Federal de Viçosa (UFV), Departamento de Biologia Geral, CEP 36570-900 Viçosa, MG, Brazil. \*Corresponding author, E-mail: renato.rosado@ufv.br

## ABSTRACT

The correct choice of parents that will compose optimal segregating populations is the key to success for breeding programs. It was postulated the hypothesis that this choice of these parents could be made based on information of molecular markers analyzed in the context of population structure. Ten parental populations were simulated and 45 hybrid combinations were obtained from the diallel crosses. Each population consisted of 200 individuals with 50 independent loci. The populations were evaluated for the Hardy-Weinberg Equilibrium (HWE), Coefficient of Inbreeding (F), Heterozygosity (H), and the Polymorphic Information Content (PIC). Genetic diversity between pairs of parental populations was evaluated using five dissimilarity measures. Values of Mantel correlation were obtained for the pairs of the dissimilarity matrices, and the PIC, H, and F values were obtained in the hybrid combinations. All parental populations were under HWE, and the combination that emerged from this condition was the hybrid 3x5, with only 26% of the loci manifesting HWE. This same hybrid was among those with lower F estimates and higher values of H, which indicated the existence of greater divergence between their parentals. There was agreement on the indication of the more and less divergent hybrid combinations for the dissimilarity measures. This fact is important because the variability, associated with the good average potential, are important criteria for the formation of an initial population in breeding programs of any kind, involving sexual processes.

**Keywords:** Genetic diversity, quantitative genetic, biometric genetic, biometrical techniques, germplasm, molecular data.

## INTRODUCTION

One of the most important steps in a breeding program is the selection of parents to compose promising segregating populations since this determines the success of the subsequent stages and the effectiveness of the program (Bertan, Carvalho, & Oliveira, 2007; Pereira, Santos, Abreu, & Couto, 2007). In this context, the selection of the most potential segregating populations optimizes the use of resources spent on a breeding program (Pimentel et al., 2013).

Emphasis has been given to the study of genetic diversity in several crops such as in cotton (Santos et al., 2017), bean (Carović-Stanko *et al.*, 2017) and soybean (Santos et al., 2014), in order to identify promising genotypes for genetic breeding purposes, to quantify the genetic variability (Rigon et al., 2012; Hiremath & Nagaraja, 2016), and to conduct the breeders to the most appropriate choices for the formation of superior hybrids (Ferreira et al., 2012) among others. In general, the parental selection for the formation of a base population has been made from phenotypic information of traits of agronomic importance, in test crosses which are modeled and analyzed following the principles of quantitative or biometric genetics. However, different data sets may be used, including pedigree data (Teixeira-Neto, Cruz, Carneiro, Malhado, & Faria, 2013), biochemical data (Signorini, Renesto, Machado, Bernal, & Monteiro, 2013), molecular markers data (Silva et al., 2017), and others.

Molecular markers have become important and efficient tools, and, and its combined evaluation with agronomic traits can increase the selection process accuracy, can optimize field work, and can ensure greater success in breeding programs (Annicchiarico, Nazzicari, Carelli, Wei, & Brummer, 2016). Using molecular markers in the studies of genetic diversity guarantees the possibility of using different biometrical techniques, based on means and variances. However, it is worth emphasizing that the choice of a method will always depend on the study objective, the level of response required, the necessary

technological infrastructure and the available time.

When molecular information is available, there is a possibility of studying genetic diversity at different levels, from the point of view of population genetics that includes individual genotypes, germplasm accessions, and populations. The diversity analysis at the population level is considered the most complex, since it is influenced by the number of individuals sampled, number of loci, genotypic constitution, and effective size (Cruz, Ferreira, & Pessoni, 2011).

The population structure is defined by the frequency of the alleles that compose the different genotypes that constitute it, and their understanding can direct decision-making in breeding programs (Cruz et al., 2011). In the analysis of the population structure, identifying the occurrence of Hardy-Weinberg Equilibrium (HWE), linkage disequilibrium (LD), and estimating parameters such as polymorphic information content (PIC), coefficient of inbreeding (F), and heterozygosity (H) are primordial (Santos et al., 2012).

This is a work that reflects the studies done in the area of population genetics. Thus, it is sufficient to consider only the genotypic information of individuals and populations generated with the observance of a meiotic process and the gametic encounter, subject to observance of the type of mating involved that, in the study, were random mating to generate targeted parents and crossbreeding to generate the hybrids.

The phenotypic evaluation of the potential of populations and their genetic diversity has been indispensable in the choice and orientation of crossing between potential parents in any breeding program based on sexual reproduction processes. The objective of this study was to evaluate the potential of the population and the degree of differentiation between pairs of populations, in the context of conventional breeding while using a population genetic approach based on molecular data, inbreeding, heterozygosity, Hardy-Weinberg equilibrium, and differentiation aspects.

## **MATERIALS AND METHODS**

For simulation purposes the only parameters required are the number of loci, the number of alleles per locus and the dominance relationship between these alleles. It is considered, at random, a certain allele frequency, for each locus, from which a gametic pool is established for possible ancestors. The genotype of each individual formed from the parent population is a consequence of the union between two gametes taken at random from a set of 10,000 gametes of the ancestors. The validation parameters of the simulation are inherent to the result of this work, which is based on the manifestation of Hardy-Weinberg equilibrium, heterozygosity and PIC. Additional information on simulation in the genes program can be found in Cruz (2006).

The Genes program (Cruz, 2016) is capable of simulating genotypic genome data (with parameters related to the size of the link group, distance, link phase, etc.), genotypic and phenotypic data of individuals and populations (with information on heritability, dominance, epistasis, averages, etc.). It is also capable of generating data on individuals and populations derived by random mating, self-fertilization and hybridizations.

Ten parental populations in Hardy-Weinberg equilibrium were simulated. Each population had 50 independent loci and two co-dominant alleles per locus, and consisted of 200 individuals. Moreover, 45 hybrid combinations were obtained from the crossing of these ten populations in a diallel scheme.

### **Step 1 - Assessment of population potential**

In order to validate the simulation process, the ten parental populations were evaluated based on their Hardy-Weinberg equilibrium condition by the chi-square test, which is presented in detail in Cruz et al. (2011). After confirming the HWE condition, data from the hybrid populations were used to estimate the descriptors of the population structure, including the coefficient of inbreeding, based on the heterozygote frequencies in the population, compared to the expected heterozygote frequencies in the all population, and relative to the polymorphic information content (PIC), which were analyzed according to Botstein, Skolnick and Davis (1980).

### **Step 2 - Degree of population differentiation**

In order to estimate the degree of differentiation among pairs of parental populations, three distance measures were estimated: Euclidian, Angular, and Genotypic of Hedrick. Furthermore, two fixation indices

were estimated: Nei's fixation index (GST) (Nei, 1973) and Wright's fixation index (FST) (Wright, 1965). The description of these methods are also presented in Cruz et al. (2011).

### Step 3 - Correlation between matrices

In order to predict the hybrid populations performance, the Mantel test (Manly, 1997) was used to correlate the distance matrices generated for the parental populations and the matrices generated for the hybrid populations considering the population descriptors, including the polymorphic information content, heterozygosity, and the coefficient of inbreeding.

### Step 4 - Computational resources for data analysis

The simulation and data analysis were performed at the Biometrics Laboratory of the Department of General Biology of the Federal University of Viçosa, using the computational resources of the Genes software (Cruz, 2013).

## RESULTS AND DISCUSSION

The parental populations were all in Hardy-Weinberg equilibrium (HWE), with at least 86% of the loci manifesting HWE (Table 1). All estimates of the coefficient of inbreeding in both parental and hybrid population were negative or close to zero. A null estimate for F is expected in sufficiently large populations and under random mating. Note that the species can be considered allogamous, but not exclusively. Any species that has mating at random occurring naturally (which is the case with allogamous) or guided by human action. This concept applies to plants and animals. Negative coefficients for F translate to an excess of heterozygous forms in the population resulting from crosses between divergent parents. Thus, the absence of disturbing factors, including natural and artificial selection, mutation, migration, and inbreeding, can be confirmed. The value of  $F = 0$  means that the genotypic frequencies were maintained as expected in HWE, which are  $p^2$ ,  $2pq$  and  $q^2$ . If there was an inbreeding, the genotypic frequencies would pass to  $p^2 + pqF$ ,  $2pq(1-F)$  and  $q^2 + pqF$ . Inbreeding does not affect the allele frequency, but it does affect the genotype frequency. Mutation and migration, if it existed, would alter the allele frequency and, consequently, the genotype frequency.

Mayo (2008) claims that we can compare the HWE rule with Newton's first Law of Motion, which states that a physical body will either remain at rest, or continue to move at a constant speed, unless forces act upon it. If stability is the rule, it will also be the basis for identifying effects on the population. These results, obtained from simulation, were evidence of practical applications of these measures in the conventional breeding, whereby the endogamic phenomenon and the genetic complementarities could not be understood by the simple examination of means and variances of phenotypic values.

The hypothesis was built on a fact. This fact is that information about the potential of possible parents and diversity are important for the formation of a base population for improvement. It is known that conventional breeding has some means to infer about genetic diversity using distance measurements based on phenotypic or genotypic information, but population structure measures are neglected. Thus, the hypothesis of this work was to work with molecular information, within the perspective of population genetics, but with the purpose of helping conventional breeding. There would be several possibilities for studies, using genetic designs such as diallel, or analysis of generations (P1, P2, F1 and F2), or by segregating generations advanced by self-fertilization or random mating or backcrossing generations, among others. The choice of this work, for the use of the diallel, proved to be adequate and points out that statistics, common in the area of population genetics, but not used in genetic improvement, such as F, PIC and EHW measures, were important and useful. Currently, conventional breeding has successfully included information on molecular markers for the purpose of prediction, classification and pattern recognition in broad genomic selection approaches and little credit has been given to the dynamics of the population under study.

In biometric studies based on phenotypic information, inference on gene complementation is done in a predictive way by means of distance measurements, or directly, by quantifying heterosis or specific capacity. As expected, the hybrid populations demonstrated that the equilibrium condition was lost. The hybrid combination originated from the crossing between the parental populations 3 and 5, which showed

98% and 88% of the loci in HWE, respectively, had only 26% of the loci manifesting HWE (Table 1). This finding was expected, since this hybrid population also stood out with a high heterozygosity value -0.56 (Table 2) and was created from the parental populations most divergent (Table 5). Thus, it is a satisfactory combination for use in cross-breeding systems that maximize genetic variability (Santos et al., 2017).

**Table 1.** Percentage of loci expressing Hardy-Weinberg equilibrium (HWE) and coefficient of inbreeding (F) observed in 10 parental populations (Pop.) and their hybrid combinations.

Pop.	1	2	3	4	5	6	7	8	9	10	HWE
1		-0.22	-0.24	-0.20	-0.24	-0.25	-0.11	-0.22	-0.26	-0.19	100%
2	48%		-0.18	-0.23	-0.18	-0.22	-0.23	-0.15	-0.18	-0.22	86%
3	44%	50%		-0.19	-0.29	-0.19	-0.24	-0.15	-0.08	-0.24	98%
4	54%	32%	56%		-0.17	-0.21	-0.14	-0.21	-0.13	-0.20	90%
5	56%	54%	26%	60%		-0.19	-0.20	-0.20	-0.20	-0.22	88%
6	48%	54%	52%	52%	54%		-0.21	-0.21	-0.21	-0.26	96%
7	46%	50%	34%	58%	56%	46%		-0.16	-0.21	-0.18	94%
8	46%	42%	56%	44%	58%	52%	62%		-0.21	-0.29	98%
9	32%	60%	50%	62%	48%	56%	36%	56%		-0.19	96%
10	52%	40%	46%	60%	54%	52%	62%	32%	56%		88%
F	0.01	0.05	0.02	0.06	0.02	-0.01	-0.04	-0.03	-0.03	0.06	

**Table 2.** Comparative values of heterozygosity (H) and polymorphic information content (PIC) of 10 parental populations (Pop.) and their hybrid combinations.

Pop.	1	2	3	4	5	6	7	8	9	10	PIC
1		0.5074	0.5275	0.5094	0.5004	0.5144	0.4442	0.5200	0.5280	0.4434	0.2467
2	0.3234		0.5043	0.5303	0.4707	0.4956	0.5433	0.4958	0.5009	0.5158	0.2760
3	0.3309	0.3309		0.5325	0.5620	0.4883	0.5507	0.4912	0.4643	0.5239	0.2817
4	0.3272	0.3332	0.3440		0.4850	0.5153	0.4922	0.5409	0.4765	0.4883	0.2890
5	0.3157	0.3117	0.3370	0.3228		0.4553	0.4939	0.4931	0.4874	0.4727	0.2433
6	0.3202	0.3184	0.3186	0.3310	0.3006		0.5097	0.5089	0.5168	0.5076	0.2457
7	0.3124	0.3415	0.3420	0.3345	0.3195	0.3275		0.4864	0.5107	0.4703	0.2756
8	0.3295	0.3324	0.3302	0.3438	0.3204	0.3268	0.3272		0.5484	0.5645	0.2844
9	0.3261	0.3312	0.3340	0.3274	0.3187	0.3297	0.3282	0.3477		0.4847	0.2870
10	0.2933	0.3285	0.3273	0.3160	0.3047	0.3142	0.3109	0.3367	0.3182		0.2353
H	0.3072	0.3279	0.3498	0.3435	0.2942	0.3084	0.3602	0.3723	0.3730	0.2725	

Other information that may collect knowledge about the potential variability of a population, available through population approaches, were H and PIC. Table 2 shows the heterozygosity values and the polymorphic information content observed in the parental populations and their respective hybrid combinations. The highest values of H were observed in the hybrid populations, with emphasis on the combinations 8 x 10 (H = 0.5645), 3 x 10 (H = 0.5620) and 3 x 7 (H = 0.5507). The PIC values ranged between 0.2353 and 0.2890 in the parental populations, and between 0.2933 and 0.3477 in the hybrid populations. All these values were lower than the heterozygosity values, which ranged from 0.2942 to 0.5645, in hybrid and parental populations. The loci studied, considering the content of the polymorphic information, were not very informative, as they should be between 0.25 and 0.50 for the marker to be considered moderately polymorphic and greater than 0.50 for highly polymorphic information (Bolstein et al., 1980). According to Ott (1992), PIC values should always be lower than those estimated for heterozygosity, such that the PIC values observed in Table 2 were in agreement with their expectations. Several authors confirmed this expectation, where these values reinforce the importance of using markers since they present high quality of information that can be extracted for different studies, such as characterization and genetic diversity studies, as well as paternity analyzes Reis et al. (2011) and Crispim, Silva, Banari, Seno and Grisolia (2012).

The estimates of association between the predicted diversity in the parental populations and the performance observed in the hybrids were shown in Table 3. Four of the five diversity measures were strongly correlated with estimated performance of hybrids with high magnitudes and statistically significant levels at the level of 1 percent probability. Hedrick's Genotypic Distance was the only one with significant values at 1% for heterozygosity and coefficient of inbreeding, and at 5% for polymorphic information content. In general, correlation estimates of low magnitude were observed between the dissimilarity measures and the PIC values of the hybrids. These low values may be due to the low variation observed in PIC values. The PIC values depend on two factors associated with the studied location. The first refers to the quantity of the locus allele, which in this study was fixed in two. The second refers to the frequency of the allele in the population considered and is, in this study, the cause of the largest and smallest variation observed.

**Table 3.** Relationship between the parents prediction and the hybrids performance.

Parentals diversity	Hybrid performance		
	PIC <sup>1</sup>	H	F
Euclidian Distance	0.1566	0.6137**	-0.7910**
Angular Distance	0.2000	0.6320**	-0.7786**
Hedrick's Genotypic Distance	0.2950*	0.6717**	-0.7239**
Nei's fixation index (G <sub>ST</sub> )	-0.1138	0.3941**	-0.7641**
Wright's fixation index (F <sub>ST</sub> )	-0.1138	0.3941**	-0.7641**

\*, \*\*Significant at 1 and 5% probability level according Mantel test, respectively. <sup>1</sup>Polymorphic Information Content (PIC), Heterozygosity (H) and Coefficient of Inbreeding (F).

The populations that had deviated the least from the HWE condition were those generated from the crosses between populations 4 & 9, 7 & 8, and 7 & 10, with 62% of the loci manifesting HWE. These populations were considered the least divergent. There was a great agreement among the more and less divergent populations, even under different measures of similarity (Tables 4 and 5). This result was surprising, since each dissimilarity measure addresses different philosophies. The Euclidean and Angular distances were based on geometric properties and considered the allelic frequencies information for discriminating populations that were more and less divergent. According to Dias (1988), for two populations were considered similar only if they occur in the same region of the geometric space, with a small distance

between them. Hedrick's genotypic distance proposes an alternative way to quantify the dissimilarity among populations, considering statistics based on the genotype frequencies and not solely on the allelic frequencies.

**Table 4.** Pairs of divergent populations according to Euclidian, Angular and Hedrick's Genotypic distance, and Wright and Nei's fixation indices ( $G_{ST}$  and  $F_{ST}$ ).

Parental diversity	More divergent	Less divergent
Euclidian Distance	(8x10) e (2x10)	(4x9) e (3x8)
Angular Distance	(8x10) e (2x10)	(4x9) e (3x8)
Hedrick's Genotypic Distance	(2x7) e (2x10)	(4x9) e (1x10)
Nei's fixation index ( $G_{ST}$ )	(6x10) e (8x10)	(4x9) e (3x8)
Wright's fixation index ( $F_{ST}$ )	(6x10) e (8x10)	(4x9) e (3x8)

**Table 5.** More and less divergent populations.

Parental Population	More Divergent					Parental Population	Less Divergent				
	Euclidian	Angular	Hedrick	$G_{ST}$	$F_{ST}$		Euclidian	Angular	Hedrick	$G_{ST}$	$F_{ST}$
1	6	6	6	6	6	1	7	10	10	7	7
2	10	10	10	10	10	2	8	5	4	4	4
3	5	5	5	5	5	3	8	8	8	8	8
4	6	6	6	6	6	4	9	9	9	9	9
5	3	3	3	3	3	5	2	2	6	2	2
6	10	10	10	10	10	6	3	3	5	3	3
7	2	2	2	2	2	7	8	8	9	8	8
8	10	10	10	10	10	8	3	3	3	3	3
9	6	1	6	6	6	9	4	4	4	4	4
10	8	8	8	6	8	10	1	1	1	4	4

Hedrick (1971) states that his methodology was advantageous in relation to the others, since populations that were completely distinct genotypically will not be wrongly labeled as identical, based on their allelic constitution. Phenotypic information has allowed the use of a range of biometric procedures to infer information about populations for the purpose of forming a base population that manifests high vigor and wide variability to be explored by selection (Mohammadi & Prasanna, 2003; Cruz, Carneiro, & Regazzi, 2014).

Biometric techniques based on measures of dissimilarity (distances) and clustering have been extensively explored (Santos, Carneiro, Silva Junior, Cruz, & Soares, 2019). However, unique information from molecular markers can provide equally important information when processed and analyzed using population parameters (Milligan et al., 2018). Recent technological advances have allowed the routine evaluation of genetic diversity at the genome level (Narum, Buerkle, Davey, Miller, & Hohenlohe, 2013; Meirmans, 2015; Garner et al., 2016).



## CONCLUSIONS

The relative diversity of the parental populations, based on the five FOR dissimilarity measures, agreed with the population descriptors of their respective hybrid combinations. Thus, estimates of H, PIC, HWE, and F, which were measured in the parental populations, may help in the prediction of the genetic diversity of hybrid combinations and assist in the assertiveness of the parents' choice for the formation of base populations.

## ACKNOWLEDGEMENTS

The authors are indebted to the Coordination for the Improvement of Higher Education Personnel (Capes), for a scholarship.

## FULL DISCLOSURE

Authors declare there is no conflict of interests in carrying the research and publishing this manuscript.

## REFERENCES

- Annicchiarico, P., Nazzicari, N., Carelli, M., Wei, Y., & Brummer, E.C. (2016) Assessment of cultivar distinctness in alfalfa: A comparison of genotyping-by-sequencing, simple-sequence repeat marker, and morphophysiological observations. *The Plant Genome*, 9, 1-12. <https://doi.org/10.3835/plantgenome2015.10.0105>
- Bertan, I., Carvalho, F. I. F., & Oliveira, A. C. (2007) Parental selection strategies in plant breeding programs. *Journal of Crop Science and Biotechnology*, 10, 211–222.
- Botstein, D., R. L., Skolnick, M., & Davis, R. W. (1980) Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *American Journal of Human Genetics*, 32, 314-331.
- Carović-Stanko, K., Liber, Z., Vidak, M., Baresic, A., Grdisa, M., Lazarević, B., & Satović, Z. (2017) Genetic Diversity of Croatian Common Bean Landraces. *Frontiers on Plant Science*, 8,1-8. <https://doi.org/10.3389/fpls.2017.00604>
- Crispim, B. A., Silva, D. B. S., Banari, A. C. B., Seno, L. O., & Grisolia, A. B. (2012) Discriminação alélica em ovinos naturalizados do Pantanal Sul-Matogrossense por meio de marcadores microssatélites. *Journal of the Selva Andina Research Society*, 3, 3-13.
- Cruz, C. D. (2006) *Programa Genes - Análise Multivariada e simulação*. Viçosa, MG: Editora UFV.
- Cruz, C. D., Ferreira, F. M., & Pessoni, L. A. (2011) *Biometria aplicada ao estudo da diversidade genética*. Visconde do Rio Branco, MG: Editora Suprema.
- Cruz, C. D. (2013) GENES: a software package for analysis in experimental statistics and quantitative genetics. *Acta Scientiarum Agronomy*, 35, 271-276. <https://doi.org/10.4025/actasciagron.v35i3.21251>
- Cruz, C.D. (2016) Genes Software – extended and integrated with the R, Matlab and Selegen. *Acta Scientiarum Agronomy*, 38:547-552. <https://doi.org/10.4025/actasciagron.v38i4.32629>
- Cruz, C. D., Carneiro, P. C. S., & Regazzi, A. J. (2014) *Modelos Biométricos Aplicados ao Melhoramento Genético*. (3<sup>rd</sup> ed.). Viçosa, MG: Editora UFV.
- Ferreira, C. B. B., Lopes, M. T. G., Lopes, R., Cunha, R. N. V., Moreira, D. A. M., Barros, W. S., & Matiello, R. R. (2012) Diversidade genética molecular de progênies de dendezeiro. *Pesquisa Agropecuária Brasileira*, 47, 378-384. <https://doi.org/10.1590/S0100-204X2012000300009>

- Garner, B. A., Hand, B. K., Amish, S. J., Bernatchez, L., Foster, J. T., Miller, K. M., Morin, P. A., Narum, S. R., O'Brien, S. J., Roffler, G., Templin, W. D., Sunnucks, P., Strait, J., Warheit, K. I., Seamons, T.R., Wenburg, J., Olsen, J., & Luikart, G. (2016) Genomics in conservation: Case studies and bridging the gap between data and application. *Trends in Ecology and Evolution*, 31, 81–83. <http://dx.doi.org/10.1016/j.tree.2015.10.009>
- Hedrick, P. W. (1971) A new approach to measuring genetic similarity. *Evolution*, 25:276-280. <https://doi.org/10.2307/2406918>
- Hiremath, G., & Nagaraja, T. E. (2016) Genetic variability and heritability analysis in selected clones of sugarcane. *International Journal of Science Technology & Engineering*, 2, 341-343.
- Mayo, O. (2008) A century of Hardy–Weinberg equilibrium. *Twin Research and Human Genetics*, 11, 249-256. [10.1375/twin.11.3.249](https://doi.org/10.1375/twin.11.3.249). <https://doi.org/10.1375/twin.11.3.249>
- Meirmans, P. G. (2015) Seven common mistakes in population genetics and how to avoid them. *Molecular Ecology*, 24, 3223-3231. <https://doi.org/10.1111/mec.13243>
- Milligan, B. G., Archer, F. I., Ferchaud, A. L., Hand, B. K., Kierepka, E. M., & Waples, R. S. (2018) Disentangling genetic structure for genetic monitoring of complex populations. *Evolutionary Applications*, 11, 1149-1161. <https://doi.org/10.1111/eva.12622>
- Mohammadi, S. A., & Prasanna, B. M. (2003). Analysis of genetic diversity in crop plants—salient statistical tools and considerations. *Crop science*, 43(4), 1235-1248. <https://doi.org/10.2135/cropsci2003.1235>
- Narum, S.R., Buerkle, C. A., Davey, J. W., Miller, M. R., & Hohenlohe, P. A. (2013) Genotyping-by-sequencing in ecological and conservation genomics. *Molecular Ecology*, 22, 2841-2847. <https://doi.org/10.1111/mec.12350>
- Nei, M. (1973) Analysis of gene diversity in subdivided populations. *Proceedings of the National Academy of Sciences of USA*, 70, 3321-3323. <https://doi.org/10.1073/pnas.70.12.3321>
- Pereira, H.S., Santos, J. B., Abreu, A. F. B., & Couto, K. R. (2007) Informações fenotípicas e marcadores microssatélites de QTL na escolha de populações segregantes de feijoeiro. *Pesquisa Agropecuária Brasileira*, 42, 707-713. <https://doi.org/10.1590/S0100-204X2007000500014>
- Pimentel, A. J. B., Ribeiro, G., Souza, M. A., Moura, L. M., Assis, J. C., & Machado, J. C. (2013) Comparação de métodos de seleção de genitores e populações segregantes aplicados ao melhoramento de trigo. *Bragantia*, 72, 113-121. <https://doi.org/10.1590/S0006-87052013005000026>
- Reis, R. V., Oliveira, E. J., Viana, A. P., Pereira, T. N. S., Pereira, M. G., & Silva, M. G. M. (2011) Diversidade genética em seleção recorrente de maracujazeiro-amarelo detectada por marcadores microssatélites. *Pesquisa Agropecuária Brasileira*, 46:51-57. <https://doi.org/10.1590/S0100-204X2011000100007>
- Rigon, J. P. G., Capuani, S., Brito-Neto, J. F., Rosa, G. M., Wastowski, A. D., & Rigon, C. A. G. (2012) Dissimilaridade genética e análise de trilha de cultivares de soja avaliada por meio de descritores quantitativos. *Revista Ceres*, 59:233-240. <https://doi.org/10.1590/S0034-737X2012000200012>
- Santos, I. G., Teodoro, P. E., Farias, F. J. C., Carvalho, L. P., Rodrigues, J. I., & Cruz, C. D. (2017) Genetic diversity among cotton cultivars in two environments in the State of Mato Grosso. *Genetics and Molecular Research*, 16:gmr16029628. <http://dx.doi.org/10.4238/gmr16029628>



- Santos, I. G., Carneiro, V. Q., Silva Junior, A. C., Cruz, C. D., & Soares, P. C. (2019) Self-organizing maps in the study of genetic diversity among irrigated rice genotypes. *Acta Scientiarum Agronomy*, 41:e39803. <https://doi.org/10.4025/actasciagron.v41i1.39803>
- Santos, J. A. S., Teodoro, P. E., Correa, A. M., Soares, C. M. G., Ribeiro, L. P., & Abreu, H. K. A. (2014) Desempenho agrônômico e divergência genética entre genótipos de feijão-caupi cultivados no ecótono Cerrado/Pantanal. *Bragantia*, 73:377-382. <https://doi.org/10.1590/1678-4499.0250>
- Santos, L. H., Oliveira, S. M. P., Malhado, C. H. M., Carneiro, P. L., Martins Filho, R., Lôbo, R. N. B., & Rodrigues, D. S. (2012) Estrutura populacional e tendências genéticas e fenotípicas da raça Guzerá no Nordeste do Brasil. *Revista Brasileira Saúde Produção Animal*, 13:1032-1043. <https://doi.org/10.1590/S1519-99402012000400007>
- Signorini, T., Renesto, E., Machado, M. F. P. S., Bepalhok, D. N., & Monteiro, E. R. (2013) Diversidade genética de espécies de *Capsicum* com base em dados de isozimas. *Horticultura Brasileira*, 31:534-539. <https://doi.org/10.1590/S0102-05362013000400005>
- Silva, M. J., Pastina, M. M., Souza, V. F., Schaffert, R. E., Carneiro, P. C. S., Noda, R. W., Carneiro, J. E. S., Damasceno, C. M. B., & Parrella, R. A. C. (2017) Phenotypic and molecular characterization of sweet sorghum accessions for bioenergy production. *PLoS ONE*, 12:e0183504. <https://doi.org/10.1371/journal.pone.0183504>
- Teixeira-Neto, M. R., Cruz, J. F., Carneiro, P. L. S., Malhado, C. H. M., & Faria, H. H. N. (2013) Parâmetros populacionais da raça ovina Santa Inês no Brasil. *Pesquisa Agropecuária Brasileira*, 48:1589-1595. <https://doi.org/10.1590/S0100-204X2013001200008>
- Wright, S. (1965) The interpretation of population structure by F-Statistics with special regard to systems of mating. *Evolution*, 19: 395-420 <https://doi.org/10.2307/2406450>

**Received: March 29, 2021.**

**Accepted: April 21, 2021.**

**Published: May 22, 2021.**

**English by: Renato Domiciano Silva Rosado and Ana Maria Cruz e Oliveira.**