

Multidimensional scaling for divergence analysis in pepper

Júlio Carlos Polimeni de Mesquita¹, Elizanilda Ramalho do Rêgo^{2*}, Angela Maria dos Santos Pessoa³, Anderson Rodrigo da Silva⁴, João José da Silva Neto⁵, Maílson Monteiro do Rêgo²

¹Agronomical Institute of Pernambuco, Brazil

²Federal University of Paraíba, Brazil

³Federal University of Ceará, Brazil

⁴Federal Institute of Education, Science, and Technology of Goiano, Brazil

⁵Institute of Education, Science, and Technology of Mato Grosso, Brazil

*Corresponding author: elizanildaramalho@hotmail.com

Abstract

Pepper plants of the genus *Capsicum* present wide genetic variability for phenotypical characters. The joint analysis of quantitative and qualitative characters is a more complete indicator to identify the existing variability in populations of plants. This work aimed to evaluate the genetic divergence within F₃ populations of *Capsicum annuum* L. through the multidimensional scaling based on the joint analysis of quantitative and qualitative data. The work was developed in at the Center of Agrarian Sciences of the Federal University of Paraíba, state of Paraíba, Brazil. Eight populations of an F₃ generation were employed, performing the characterization of 36 qualitative and quantitative characteristics. The experimental design was in randomized blocks, with eight treatments and 50 replications. The non-metric multidimensional scaling (nMDS) was performed based on Gower's dissimilarity matrixes. Dispersion diagrams were built to present the nMDS solutions in the bidimensional space. The nMDS mapping was efficient on estimating divergence within populations with Kruskal's Stress below 20%. The internal uniformity presented by populations P-4, P-5, P-6, P-7, and P-8 is an indicator that the genotypes of these populations are suitable to be used in future genetic breeding programs of ornamental peppers. Conversely, populations P-1, P-2, and P-3 presented the greatest dispersion, characterizing phenotypical variability. The multidimensional scaling was efficient to estimate the genetic divergence among and within F₃ populations of ornamental pepper.

Keywords: *Capsicum*; genetic variability; genetic breeding; ornamental pepper; multivariate analysis.

Abbreviations: nMDS_non-metric multidimensional scaling, PCA_principal component analysis.

Introduction

Pepper plants possess a wide genetic diversity associated with morpho-agronomical aspects. The exploitation of this variability allows their use in sectors of the food, pharmaceutical, and cosmetical industry, besides the most recent use for ornamental purposes (Albrecht et al., 2012; Hill et al., 2013; Rêgo and Rêgo, 2018).

The study of the divergence among individuals, populations, or groups of individuals is of fundamental importance for the knowledge of the genetic variability and has been performed through morphological descriptors, using several types of quantitative and qualitative variables (Bento et al., 2007; Gonçalves et al., 2009; Barbé et al., 2010; Rêgo et al., 2016). However, although the joint analysis of quantitative and qualitative analysis is potentially a more complete indicator of the variability in germplasm banks, few works have used this strategy, which is probably related to the still insufficient knowledge on the statistical techniques that allow this approach, the lack of free softwares to jointly analyze these data, as well as the tendency of researchers in giving more importance to those variables which are directly related to characters that are worked in improvement programs (Gonçalves et al., 2008), mostly of a quantitative nature.

A technique that allows the joint analysis of quantitative and qualitative data was proposed by Gower (1971), through an algorithm that estimates the similarity between two individuals using data with continuous and discrete distributions.

Once calculated the similarities between individuals, multivariate analysis techniques are in general employed for the quantification of the genotypical and phenotypical divergence in several horticultural species, such as peppers and tomato (Pessoa et al., 2018; Del Medico et al., 2019).

Several techniques can be used to study the phenotypical variability, with grouping analysis and the principal component analysis (PCA) ranking among the most used. Some general application algorithms such as the UPGMA, Ward, and nearest neighbor are conditioned to the property of the hierarchy of groups, which might not exist (Silva et al., 2014). The PCA, by its turn, does not present such a problem, although it is restricted to quantitative variables.

The MDS is an ordination technique for dimensionality reduction that allows placing individuals as points in space, generally bi or tridimensional (Manly, 2004; Borg and Groenen, 2005). This is a general technique since it operates

directly on dissimilarity or similarity matrixes, and is especially useful in cases in which the relation between individuals is unknown, common in germplasm banks, although it is possible to estimate a dissimilarity matrix between genotypes.

Using grouping analysis and the non-metric multidimensional scaling for joint analysis of phenotypical and genotypical data, Pessoa et al (2019) identified variability among the individuals of an F_3 population of pepper plants. Other authors reported that this method was efficient in representing the genetic distance between individuals of ornamental pepper plants. The MDS technique was also used in a genetic divergence study in cocoa (Leal et al., 2008), *Tibouchina papyrus* (Telles et al., 2010), tomato leafminer (*Tuta absoluta*) in tomato leaves (Bettaïbi et al., 2012), and in garlic (Silva et al., 2014).

This work aimed to estimate the genetic divergence within F_3 populations of *Capsicum annum* L. using a multidimensional scaling based on the joint analysis of quantitative and qualitative data.

Results

The dispersion quantified among and within the eight populations of ornamental peppers in the bidimensional space, obtained through the non-metric multidimensional scaling (nMDS) in two years of evaluation is presented in Figures 1 and 2.

First year

It was observed that populations P-4, P-6, and P-7 presented the lowest intragroup divergence (Fig. 1a). The lower distance between the individuals of each of these populations can be verified by a phenotypical analysis, through which the existing uniformity among these plants can be observed, for example, the individuals that compose the P-4 population present low stature, uniform canopy diameter, and fruits with little variability of shape and size, but varying mainly as for the coloration (data not presented), thus confirming the low variability of the population.

The individuals of the P-6 population formed a distinct group compared to the behavior observed in the remaining populations (Fig. 1a). All plants from this population (P-6) presented greater uniformity for the morphological characteristics of plants, flowers, and fruits (data not presented). Likewise, the P-7 population presented little variation as for the morphological characteristics.

In the P-3 population, some individuals presented similarities while others remained dispersed (Fig. 1a). Conversely, a wide intragroup genetic divergence was verified for populations P-1 and P-2. The genotypes in these populations were highly dispersed (Fig. 1a).

It was also verified that populations P-1, P-4, and P-7 presented overlap of genotypes, which also occurred among some individuals of populations P-2 and P-3 (Fig. 1a).

The populations P-5 and P-8 presented relevant phenotypical distance from one to another (Fig. 1b). In spite of the distance observed between these populations, within each of them the genotypes were uniform for the evaluated characteristics, and those that formed the P-5 population presented smaller size and alike leaf and fruit characteristics, whereas in the genotypes of the P-8 population the plants were highlighted for being taller and more uniform (data not presented), which justified that graphic distance observed.

Using Kruskal's Stress calculation for the level of adjustment of the multidimensional scaling for the characteristics evaluated in the first evaluation year, maximum values of 16.8% were observed for the stress values in the mixed data (Fig. 1a).

Second year

In spite of the arrangement of the populations in distinct groups in the second year (Fig.2), it can be observed that the behavior of the genotypes that formed populations P-4, P-6, P-7, and P-8 was little modified throughout time, that is, from one year to another, except for what was observed in the genotypes of the P-5 population, which was more dispersed in the second year.

The Kruskal's Stress calculation for the level of adjustment of the multidimensional scaling for the characteristics evaluated in the second evaluation year, maximum values of 19.18% (Fig. 2b).

For the two evaluation years of the characteristics and simultaneous systematization of the data, the stress values are within the acceptable limits.

The multidimensional scaling is efficient to estimate the genetic divergence within F_3 populations of ornamental pepper plants through the joint analysis of quantitative and qualitative data.

Discussion

The dispersion observed in the populations for the two evaluation years is probably associated with the segregation of the genes for the evaluated characters, an understandable factor for an F_3 generation originated from common parents. Mesquita et al (2016) reported that F_3 populations of ornamental peppers are divergent, allowing the formation of groups according to the similarity of the evaluated individuals through multivariate analyses. Pessoa et al (2019) also reported the existence of genetic variability among individuals of an F_3 population of *C. annum* through mixed data analysis, in which the non-metric multidimensional method was efficient in representing the genetic distance between the individuals of the population.

In the populations (P-4, P-6, and P-7) in which the genotypes were grouped in the bidimensional space of disparity, they showed to be more similar for the phenotypical characteristics evaluated. The homogeneity for the individuals in these populations is favorable in the improvement of pepper plants, providing a faster uniformity in these individuals and, consequently, the performing of comparative assays for the indication of cultivars, as long as it presents the target characters for selection.

In the population (P-3) that presented variation in the behavior of some genotypes, demonstrating uniformity among some individuals and dispersion in others, this is an aspect of interest in pepper plants with ornamental potential, since genotypes with distinct purposes can be selected, such as to open lineages and advance generation for further selection. Since selection is the base for all improvers to obtain superior varieties, genetic diversity is an important component for selection (Anira et al., 2017).

In the dispersion verified in the genotypes of some specific populations, such as P-1 and P-2. These are characteristic aspects of genetic diversity among the individuals within populations. These genotypes probably present specific characteristics that differentiate them from each other and

Table 1. Quantitative and qualitative for plant and fruit in ornamental pepper (*Capsicum annuum* L.).

Plant Descriptors	Fruit Descriptors
Plant Height	Fruit Length
Canopy Diameter	Largest Fruit Diameter
First Bifurcation Height	Smallest Fruit Diameter
Stem Diameter	Peduncle Length
Leaf Length	Pericarp Thickness
Leaf Width	Placenta Length
Stem Color	Fruit Length/Diameter Ratio
Presence of Anthocyanin in the Node	Fruit Weight
Stem Shape	Content of Dry Matter
Stem Pubescence	Number of Seeds per Fruit
Growth Habit	Color of the Immature Fruit
Branching Density	Color of the Intermediary Fruit
Tillering	Color of the Mature Fruit
Leaf Density	Presence of Anthocyanin Stain
Leaf Color	Fruit Shape
Leaf Shape	Fruit Apex Shape
Blade Margin of the Leaf	Shape of the Union of the Pedicel with the Fruit
Leaf Pubescence	Shape of the Persistence of the Pedicel with the Stem

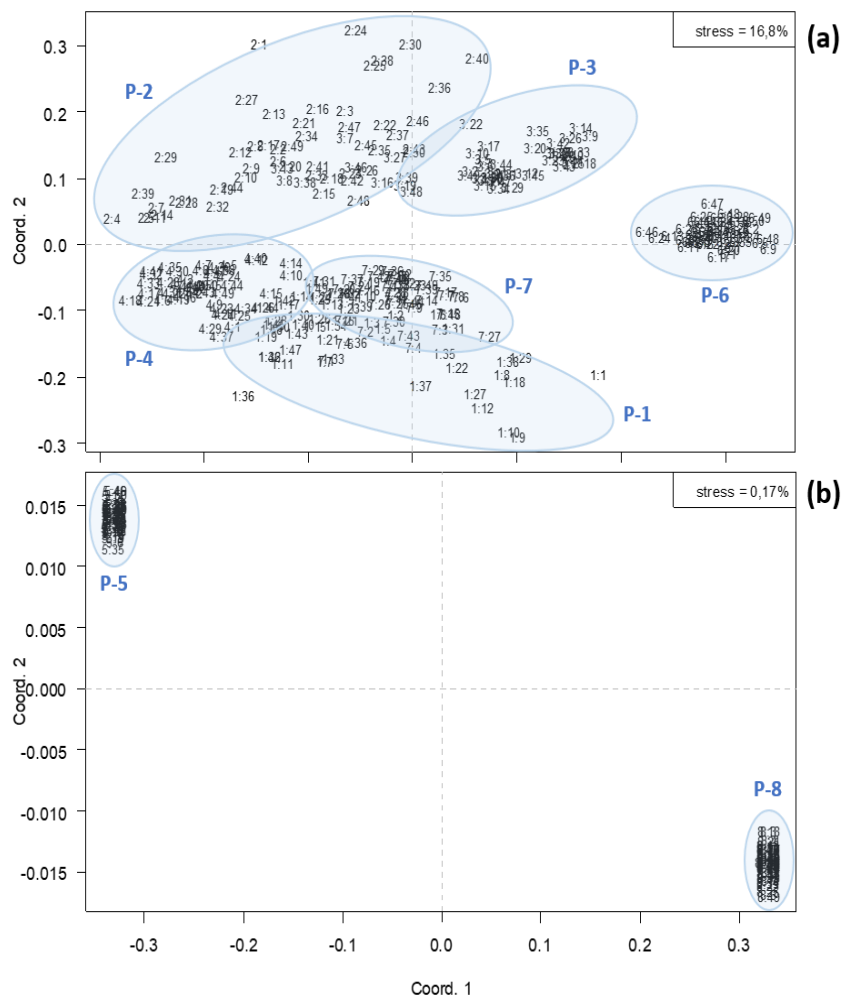


Fig 1. Dispersion in the bidimensional space of phenotypic disparities among F3 population of ornamental pepper clusters formed by genotypes of the populations (a) P-1, P-2, P-3, P-4, P-6, and P-7; and (b) P-5 and P-8. First year (2015), Areia-PB, Brazil. The elongation of the ellipses indicate within population divergence.

Table 2. Grouping among ornamental pepper populations (*C. annuum* L.) in two years of evaluation, 2013 and 2014.

Years	Group I (Populations)	Group II (Populations)
First Year	1, 2, 3, 4, 6 and 7	5 and 8
Second Year	1, 5, 6 and 8	2, 3, 4 and 7

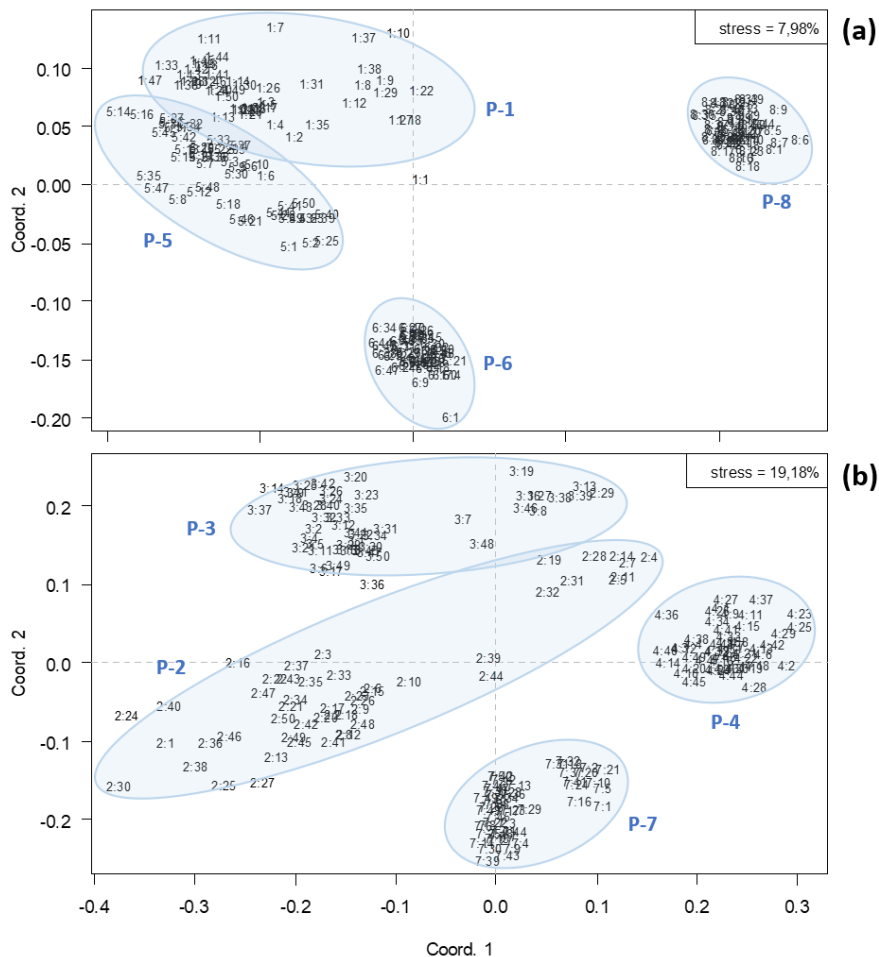


Fig 2. Dispersion in the bidimensional space of phenotypic disparities in F₃ populations of ornamental pepper clusters formed by genotypes of the populations (a) P-1, P-5, P-6, and P-8 and (b) P-2, P-3, P-4, and P-7. Second year (2016), Areia-PB, Brazil. The elongation of the ellipses indicate within population divergence.

from the remaining individuals in the population, becoming unique and able to be chosen for selection, since they present specific characters that can be selected in later generations, as long as they present characters of interest. This diversity occurs due to multiple genes that control these characteristics and continue to segregate, undergoing influence from the environment, which may harden the progress in the selection of progenies containing favorable genes (Htwe et al., 2017). Lima et al (2019) reported that morpho-agronomical characters are efficient for the determination of genetic diversity in populations of ornamental pepper plants.

The overlay of some individuals in distinct populations (P-1, P-4, and P-7) might be related to the similarity for some characteristics and, probably because they are originated from identical parents, the genes segregate similarly for the evaluated characters.

The change in the dispersion of some individuals in the population, in the second evaluation year, may be

associated with the genotype x environment interaction for the phenotypical characteristics evaluated. The component of the genotype x environment interaction in pepper plants may be related to the genotypes, since they do not present the same production performance in the different evaluation years. Moreira et al. (2008) also reported the influence of the genotype x environment interaction in several characteristics of pepper plants in the F₇ generation. Besides allowing to expressively facilitate the examination of the dissimilarity matrix in a high number of genotypes, the non-metric multidimensional scaling in two dimensions allowed to efficiently identify the variability between and within F₃ populations based on quantitative and qualitative morphological descriptors. Bento et al (2007) highlight the advantages that the multi-categorical variables present compared to quantitative variables, being easier to observe and sparing both time and work. Moura et al (2010) also detected phenotypical variability in pepper genotypes based

on morpho-agronomical descriptors and using the estimation of genetic distance using Gower's algorithm. For the two evaluation years of the characteristics and simultaneous systematization of the data, the stress values are within the acceptable limits. According to Kruskal et al (1964), stress values up to 20% are acceptable; above this value, the representation of the distances through this technique is not efficient (Sturrock et al., 2000).

Materials and methods

Plant materials

Fifty plants from each of the eight F_3 populations or ornamental pepper were evaluated: F_{21} (P-1), F_{24} (P-2), F_{25} (P-3), F_{27} (P-4), F_{29} (P-5), F_{210} (P-6), F_{211} (P-7), and F_{231} (P-8), originated from the crossing between accessions 134 (P-9) and 77.2 (P-10), belonging to the Active Germplasm Bank of the Laboratory of Biotechnology and Vegetal Improvement of UFPB.

Treatments and experimental design

The experimental design employed was completely randomized, in a factorial scheme (8 x 2), eight treatments (populations) and two years of evaluation with 50 replications per population.

Conduction study

The work was performed in a plant nursery of the Laboratory of Biotechnology and Vegetal Improvement at the Center of Agrarian Sciences of the Federal University of Paraíba (CCA-UFPB), Areia – PB, Brazil.

The sowing was performed in expanded polystyrene trays with two hundred cells filled with the commercial substrate Plantmax HT[®]. Thirty-five days after sowing the plantlets were transplanted to plastic containers with volumetric capacity for 900 ml, containing the same substrate. The plants were daily irrigated in alternate days, receiving a nutritive solution that was prepared based on Furlani et al (1999), with the following composition in g/1000 L: 1000 g of calcium nitrate; 1250 g of potassium nitrate; 250 g of MKP; 500 g of magnesium sulfate; 1.5 g of boric acid; 25 g of quelatec AZ; 25 g of ultra-iron; 110 g of potassium chloride, and 150 g potassium sulfate. During the experimental period, the monitoring of pests and diseases was performed, along with phytosanitary measures with the preventive objective of minimizing possible damages to the plants caused by pests and diseases.

Evaluated traits

The morpho-agronomical characterization was performed according to the recommendations of the descriptors of the genus *Capsicum*, proposed by the IPGRI (1995). The first evaluation was performed when the plants presented more than half of the fruits in the intermediate stage for ripe, and the second evaluation was performed one year after the first. 36 morphological descriptors were used (16 quantitative and 20 qualitative) (Table 1).

Statistical analysis

The data were previously subjected to a two-way multivariate analysis of variance (two-way MANOVA). The effect of the interaction between populations and years was evaluated. Once detected an interaction effect, the analysis of genetic divergence between populations was performed separately for each year through grouping analysis and

canonical variables, observing the formation of two population groups in the first and second year (Table 2) (Mesquita et al., 2016). Afterward, based on the previously described population groups, the dissimilarities between genotypes within populations were estimated for each evaluation year. Gower's dissimilarity coefficient (1971) was used for that purpose, based on the arithmetic mean of the quantitative characters and the modal value of the qualitative characters. A non-metric multidimensional scaling (nMDS) was performed based on the Gower dissimilarity matrixes. Dispersion diagrams were built to present the nMDS solutions in the bidimensional space. The adjustment level of the nMDS mapping was calculated by the $Stress_1$ of Kruskal.

All analyses were performed with the R software, version 3.0.3 (R Core Team, 2014).

Conclusion

The multidimensional scaling is efficient to estimate the genetic divergence within F_3 populations of ornamental pepper plants through the joint analysis of quantitative and qualitative data. There is a genotype x environment interaction in the F_3 population of ornamental pepper plants, consequently, there is a need to performed studies in populations in other evaluation years.

Acknowledgements

The authors are thankful to Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) and Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) by grant of their research fellowship and financial support.

References

- Albrecht E, Zhang D, Saftner RA, Stommel RJ (2012) Genetic diversity and population structure of *Capsicum baccatum* genetic resources. *Genet Resour Crop Evol.* 59 (4): 517-538.
- Anira R, Herison C, Suryati D (2017) Genetic Variability and Heritability Estimates of Growth and Yield Component of Curly Chili Pepper. *Akta Agrosia.* 20 (2): 48-53.
- Barbé TC, Amaral Júnior AT, Gonçalves LSA, Rodrigues R, Scapim CA (2010) Association between advanced generations and genealogy in inbred lines of snap bean by the ward-modified location model. *Euphytica.* 173(3): 337-343.
- Bento CS, Sudré CP, Rodrigues R, Riva EM, Pereira MG (2007) Descritores qualitativos e multicategóricos na estimativa da variabilidade fenotípica entre acessos de pimenta. *Scientia Agrária.* 8(2):146-153.
- Bettaïbi A, Mezghani-Khemakhem M, Bouktila D, Makni H, Makni M (2012) Genetic variability of the tomato leaf miner (*Tuta absolutameyrick*; LEPIDOPTERA: GELECHIIDAE), In Tunisia, inferred from RAPD-PCR. *Chilean Journal of Agricultural Research.* 72(2): 212-216.
- Borg I, Groenen PJF (2005) *Modern Multidimensional Scaling: Theory and Applications.* Springer, 2 ed. Springer Science + Business Media.
- Del Medico AP, Cabodevia VG, Vitelleschi MS, Pratta GR (2019) Multivariate estimate of heritability for quality traits in tomatoes by the multiple factor analysis. *Pesq Agropec Bras.* 54: 1-8.

- Gonçalves LSA, Rodrigues R, Amaral Junior AT, Karasawa M, Sudré CP (2009) Heirloom tomato gene bank: assessing genetic divergence based on morphological, agronomic and molecular data using a Ward-modified location model. *Genet Mol Res.* 8(1): 364-374.
- Gonçalves LS, Rodrigues R, Amaral Junior AT, Karasawa M (2008) Comparison of multivariate statistical algorithms to cluster tomato heirloom accessions. *Genet Mol Res.* 7(4): 1289-1297.
- Gower JC (1971) A general coefficient of similarity and some of its properties. *Biometrics.* 27(4): 857-874.
- Hill TA, Ashrafi H, Wo RCS, Yao J, Stoffel K, Truco JM, Kozik A, Michelmore RW, Deynze AV (2013) Characterization of *Capsicum annuum* genetic diversity and population structure based on parallel polymorphism discovery with a 30K unigene pepper gene chip. *Plos One.* 8(2): 1-16.
- Htwe CSS, Thwe AA, Htwe NM, Aung MZN, Myint KT (2017) Variation and Segregation in F2 Population of Hot Pepper (*Capsicum annuum*). *Journal of Agricultural Research.* 4: 86-94.
- IPGRI (1995) International plant genetic resources institute. Descriptors for *Capsicum*. Rome, IBPGR. p.49.
- Kruskal JB (1964) Multidimensional scaling by optimizing goodness of fit to a non-metric hypothesis. *Psychometrika.* 29(1): 01-27.
- Leal JB, Santos LM, Santos CAP, Pires JL, Ahnert D, Corrêa RX (2008) Diversidade genética entre acessos de cacau de fazendas e de banco de germoplasma na Bahia. *Pesq Agropec Bras.* 43(7): 851-858.
- Lima JAM, Rêgo ER, Porcino MM, Silva GHN, Carvalho MG, Pessoa AMS, Rêgo MM (2019) Selection in Base Population of Ornamental Peppers (*Capsicum annuum* L.). *Journal of Experimental Agriculture International.* 31(3): 1-7.
- Manly B (2004) *Multivariate and Statistical Methods: A Primer*, Third Edition. Taylor & Francis. p.253.
- Mesquita JCP, Rêgo ER, Silva A, Silva Neto JJ, Cavalcante LC, Rêgo MM (2016) Multivariate analysis of the genetic divergence among populations of ornamental pepper (*Capsicum annuum* L.). *Afr J Agric Res.* 11(42): 4289-4194.
- Moreira SO, Rodriguês R, Araújo ML, Sudré CP, Riva-Souza EM (2008) Desempenho agrônômico de linhas endogâmicas recombinadas de pimenta em dois sistemas de cultivo. *Ciência Rural.* 39(5):1387-1393.
- Moura MCCL, Gonçalves LSA, Sudré CP, Rodrigues R, Amaral Júnior AT, Pereira TNS (2010) Algoritmo de Gower na estimativa da divergência genética em germoplasma de pimenta. *Hortic Bras.* 28(2): 155-161.
- Pessoa AMS, Rêgo ER, Carvalho MG, Santos CAP, Rêgo MM (2018) Genetic diversity among accessions of *Capsicum annuum* L. through morphoagronomic characters. *Genet Mol Res.* 17(1): 01-15.
- Pessoa AMS, Rêgo ER, Silva APG, Mesquita JCP, Silva AR, Rêgo MM (2019) Genetic diversity in F3 population of ornamental peppers (*Capsicum annuum* L.). *Rev Ceres.* 66(6): 442-450.
- R CORE TEAM. R: A language and environment for statistical computing. (2013) R foundation for Statistical Computing, Vienna, Austria.
- Rêgo ER, Rêgo MM (2018) Ornamental Pepper. In: Van Huylbroeck J (Ed.) *Ornamental Crops*. Cham, Springer International Publishing. 529-565.
- Rêgo ER, Rêgo MM, Finger FL (2016) Production and Breeding of Chilli Peppers (*Capsicum* spp.). Cham, Springer International Publishing Switzerland. 129.
- Silva AR, Cecon PR, Dias CTS, Puiatti M, Finger FL, Carneiro APS (2014) Morphological phenotypic dispersion of garlic cultivars by cluster analysis and multidimensional scaling. *Scientia Agricola.* 71(1): 38-43.
- Sturrock K, Ocha JA (2000) Multidimensional scaling stress evaluation table. *Field Methods.* 12:49-60.
- Telles MPC, Silva SP, Ramos JR, Soares TN, Melo DB, Resende LV, Batista EC, Vasconcelos BF (2010) Estrutura genética em populações naturais de *Tibouchina papyrus* (pau-papel) em áreas de campo rupestre no cerrado. *Revista Brasil Bot.* 33(2): 291-300.