Australian Journal of Crop Science AJCS

AJCS 12(10):1596-1602 (2018) ISSN:1835-2707
doi: 10.21475/ajcs.18.12.10.p1152

The integration of quantitative and multicategorical data for the analysis of genetic divergence in germplasm of cassava (*Manihot esculenta* Crantz.): A new approach

Ronaldo Silva Gomes*, Francisco Charles dos Santos Silva, Ronaldo Machado Junior, Cleverson Freitas de Almeida, Fabio Teixeira Delazari, Renata Dias Freitas Laurindo, Rafael Henrique Fernandes, Derly José Henriques da Silva

Universidade Federal de Viçosa, Viçosa-MG, Brazil

***Corresponding author: ronaldo.s.gomes@ufv.br**

Abstract

The analysis of genetic divergence in plant germplasm based on a set of quantitative data has revealed a low association to the divergence estimated from multicategorical data and vice versa. On the other hand, strategies involving the integration of quantitative and multicategorical information may provide more accurate estimations. Thus, this study aimed to estimate the genetic divergence in the germplasm of cassava based on the integration of quantitative and multicategorical data, targeting a greater comprehensiveness and accuracy in the estimation. Data from 10 genotypes of M. *esculenta*, characterized by 20 quantitative and 24 multicategorical characteristics was used in this work. The genotypes were collected from autochthonous fields from five different microregions of Brazil and evaluated in a completely randomized block design experiment with four replications. First, the analysis of genetic divergence was performed based on the distances individually obtained from the quantitative and multicategorical information. For the data integration, three strategies were adopted, the Gower's algorithm, the transformation of quantitative data by the equitable division of its amplitude into several classes, and the sum of the distance matrices, obtained from quantitative and multicategorical data. The estimates of genetic divergence from the quantitative data had a low association with that expected from the multicategorical data and vice versa. The transformation of data and the algorithm of Gower were not efficient, which resulted in low correlations with the matrices of distances obtained from the original data. The sum of matrices consisted on the strategy of higher efficiency and provided a higher comprehensiveness and accuracy in the analysis of genetic divergence in the germplasm of cassava.

Keywords: Cassava, data integration, genetic resources**,** genetic diversity, morphoagronomic characterization.
Abbreviation: EDDA_strategy of equitable division of the data amplitude; SUM_strategy of the sum of the distance matrices.

Introduction

The genetic resources of cassava (*Manhiot esculenta* Crantz.) are fundamental components of the breeding programs of this crop (Mezette et al., 2013). Brazil, the center of domestication and dispersal of M. *esculenta*, holds the largest genetic reserve of this species. The magnitude of this variability stems mainly from the variation in soil and climatic conditions found in this country, which decisively contributes to the increase in the variability of this species' germplasm (Valois, 2005).

The Brazilian germplasm of cassava encompasses a series of favorable genes, which allow the selection of genotypes possessing resistance to the main diseases that affect this crop (Do Carmo et al., 2015; Nery-Silva et al., 2007; Vilas Boas et al., 2016), besides those with better feeding properties (Coronado et al., 2011; Da Silva et al., 2014; Mezette et al., 2009). In addition, these characteristics confer to the Brazilian germplasm the capacity of adaptation to different edaphoclimatic conditions, constituting, therefore, to fundamental resource building for the breeding and sustainability of this crop.

The morphoagronomic characterization of plant germplasm is an important step in the genetic breeding programs, especially in the determination of genetic diversity. The diversity of cassava germplasm has been accessed through a series of strategies. They include characterization based on pedigree information, passport, and the usual morpho-agronomic characterization through multicategorical and quantitative descriptors (Gomes et al., 2007; Mezette et al., 2013; Okpara et al., 2014). Recently, this characterization has also been performed based on the use of molecular markers.

The estimates of genetic diversity based on morpho-agronomic characterization are advantageously more economical as compared to the use of molecular markers. With a wide application, this strategy is also suitable for regions with low technological levels, where access to tools such as molecular markers is still limited, which allows practical and economical evaluations of the germplasm polymorphism.

The estimation of diversity obtained from morphoagronomic characterization is performed mainly on isolated groups of descriptors, particularly on multicategorical or quantitative descriptors. This approach has the disadvantage of inconsistency, since the correlations between the estimates of genetic divergence established from an isolated group of descriptors are generally of low magnitude.

The low correlations between the estimates of genetic divergence obtained from the isolated groups of descriptors compromises the extrapolation of this estimate from one specific group to another. In order to circumvent this limitation, multivariate statistics have been proposed as a strategy that allows the integration of information of different natures for estimating genetic divergence, thereby optimizing the differentiation and selection of parentals (Fonseca et al., 2006).

A series of multivariate methods can be applied for the prediction of genetic divergence, depending on the desired accuracy, the ease of analysis, and the method by which the data was obtained. (Cruz et al., 2011). The conversion of all variables into a single pattern, such as the transformation of quantitative into multicategorical variables, incorporates a promising strategy for the study of genetic diversity, which allows the integration of variables of different natures. A series of studies involving the estimation of genetic divergence in the germplasm of a series of crops based on the integration of data from different natures has already been conducted, revealing the existence of different strategies for this integration (Alves et al., 2013; Da Silva et al., 2013; Martins et al., 2011).

On the other hand, studies that have accessed the genetic divergence in cassava germplasm based on the integration of data remain incipient due to greater difficulties in the acquisition of information of different natures and mainly as a result of the lack of using efficient methodologies for such practices. This observation demonstrates the need for developing methodologies that enable optimization of estimation of genetic divergence in the germplasm under study for economic purpose and, most importantly, for developing more accurate methodologies.

The objective of this work was to estimate the genetic diversity in cassava germplasm by integrating quantitative and multicategorical data, targeting a greater comprehensiveness and accuracy of estimation.

Results and Discussion

The clustering and genetic divergence analysis based on quantitative and multicategorical data showed different patterns.

The clustering of the germplasm by the method proposed by Tocher (Table 1) based on the distances obtained only from the quantitative data resulted in the formation of 4 groups, while the clustering from the distances based on the multicategorical data formed 5 groups .

The pattern of clustering based on the distances of each set of data was also differentiated. In clustering based on the distances obtained from the quantitative data, most (60%) of the germplasm were concentrated in Group 1. For the multicategorical data, most of the germplasm were grouped in either Group 1 (40%) or 2 (30%), which resulted in a more equitable distribution of other genotypes in the following groups, including the formation of an additional group in relation to grouping based on quantitative information (Table 1).

The estimates of genetic divergence obtained from quantitative data (data not shown), identified the genotypes Talo Vermelho and Rosa as the most similar and the genotypes R 01 and BRS Dourada as the most divergent ones.

When considering all pairs of distances between the genotypes, as obtained from the quantitative information, it was noticed that the genotype R 01 was the most divergent as it manifested greater average distance between the pairs of distances to which it belonged and also that the genotype Rampa (with the lowest average distance) was the most similar. Based on the multicategorical data, the R 01 genotype continued as the most divergent and the genotype Gameleira was found to be the most similar, based on the inferences previously mentioned.

The pattern of clustering based on quantitative or multicategorical data may be associated with the nature of each of these datasets. Unlike multicategorical data, the quantitative characteristics possess polygenic control attributable to a series of genes of smaller effects as well as are intrinsically influenced by the environment. As a consequence, the phenotypic variation for these characteristics tends to be continuous, disallowing the establishment of discrete classes. These aspects reduce the distances between genotypes, while increasing the difficulty in their differentiation.

Thus, the formation of different numbers of groups and differentiated distribution of germplasm in the groups formed from each dataset revealed a discrepancy between the analyses of genetic divergence performed individually based on quantitative or multicategorical information. This approach limited the extrapolation of the inferences of genetic divergence obtained from the set of quantitative data to the expected divergence from multicategorical data and vice versa. This represents a limitation, considering that the estimates of divergence should consider both quantitative and multicategorical characteristics, which, in some cases, are equally important for the characterization of plant germplasm.

In their study with accessions of tomato, Martins et al. (2011) compared the clustering obtained from quantitative information and that formed from multicategorical data and found a differentiated result for the number of groups and distribution accessions in the groups. Silva et al. (2013) reported a similar result for the assessment of genetic divergence between accessions of *Coffea arabica.* In addition, evaluation of sunflower germplasm based on quantitative and multicategorical characteristics by Purwaity and Herwaity (2016) revealed a greater divergence from the multicategorical characteristics.

The analysis of correlations between distance matrices showed that the sum of matrices was efficient for the analysis of genetic divergence

Only the strategy of the sum of the distance matrices (SUM) was efficient for the analysis of genetic divergence based on the simultaneous integration of quantitative and multicategorical data.

The estimations of correlations between the matrices of distances obtained from the quantitative and multi-categorical data were of low magnitude (Table 2).

The low magnitude of correlations between the matrices of distances obtained from the quantitative and multicategorical data revealed that the estimates of distances obtained from a dataset do not encompass the estimates of distances obtained from another dataset. In this way, the genetic divergence obtained from quantitative characteristics does not predict the divergence expected from multicategorical information and vice versa.

The estimations of genetic divergence from a dataset of different natures have provided inconsistent results, as demonstrated by Alves et al. (2013), who studied the germplasm of Brazilian physic nut (*Jatropha curcas*), and by Scarano et al. (2014), who assessed beans cultivars (*Phaseolus vulgaris* L.).

According to Rohlf (2000), the values of correlation <0.7 are inconsistent, indicating that the estimations of divergence obtained from a set of characters does not explain those obtained from another set of data. Thus, the genetic divergence estimated from a set of quantitative data cannot be extrapolated to the divergence expected from multicategorical data and vice versa due to the low magnitude of correlation between the distances matrices obtained from these dataset.

Considering the inconsistency in estimating genetic divergence from each dataset, the decision about which groups of variables should be used for determining the genetic divergence depends on the study objective. According to Gomes et al. (2016), the use of multicategorical information such as morphological descriptors allows evaluation of the genetic divergence in cassava germplasm more efficiently in relation to the practical and economical aspects. On the other hand, the association between multicategorical and quantitative information, even considering a greater difficult in the assessment of quantitative, is indispensable considering the economic importance of most of the quantitative descriptors.

The strategy of equitable division of the data amplitude (EDDA), was not efficient, providing matrices of distances in low correlations with those obtained from the quantitative original data. On comparison, the pairs of distances between the germplasm in the matrix obtained by quantitative evaluation and those observed in the new matrices of distances obtained by EDDA, a considerable disagreement was noted with respect to an increasing distances obtained from the EDDA. The increase in the distances in the new matrices was probably the main cause of the low correlation between these matrices.

The transformation of the quantitative data into multicategorical data and, consequently, the establishment of discrete classes in the converted data increase the dissimilarity between the germplasm, contributing to lower correlations of new matrices and distance matrix from the original data.

The results of the transformation of quantitative data into multicategorical data by EDDA in this study differ from those reported by Martins et al. (2011) and Silva et al. (2013). These authors considered the EDDA in tree classes as the most appropriated strategy for the integration of quantitative and multicategorical data in the study of genetic divergence in the germplasm of tomato and coffee plants, respectively.

Moreover, the Gower algorithm was not an efficient strategy. Although the correlation between the distance matrix obtained from this strategy and the matrix obtained from the multicategorical data were high (0.82), it was much lower when compared to the quantitative data (0.62), revealing an inconsistency in the correlations (Table 2).

Only the strategy of the sum of the distance matrices (SUM) was efficient for the analysis of genetic divergence based on the simultaneous integration of quantitative and multicategorical data. As preconized by Rohlf (2000), the correlations between the matrix SUM and the original matrices of distance, obtained from the quantitative and multicategorical data, was adequate (>0.70) and nearly similar, besides showing significance at 1% of probability by Z test of Mantel.

Thus, the proposed analysis of genetic divergence was performed by integrating quantitative and multicategorical data by means of SUM, based on the arithmetic complement of a simple coincidence index.

Clustering and genetic divergence based on the integration of quantitative and multicategorical data

As shown in Figure 1, a cut made at the distance of 0.94, allowed the establishment of 5 groups: first, with the genotype R 01; second, with BRS Dourada; third, including the genotypes BRS Gema de Ovo and Orelha de Leão; forth, Gameleira and Turiaçu; and fifth, with the highest number, including the genotypes Pão, Rampa, Rosa and Talo Vermelho.

The principal of the clustering method of complete linkage includes the initial identification of the most similar pair of genotypes (of lower distance), followed by the determination of the distances between this pair and each one of the remaining genotypes (Cruz et al., 2011). Based on this fact, a new dissimilarity matrix with the distances between the pair of higher similarity and each one of the remaining genotypes is obtained, which establishes the remaining pairs of genotypes with lower distances as the net group (Cruz et al., 2011). This operation is conducted repeatedly until all genotypes were clustered. This method allowed the inclusion of real values of minimum and maximum distances between the genotypes in the dendogram, allowing inference to be drawn from the dendogram with respect to the real values of these distances. The estimates of distances based on the integrated data identified the genotypes Talo Vermelho and Rosa as the most similar pair of genotypes and the genotypes R 01 and BRS Dourada as the most divergent ones. On considering all the pairs of distances to which each genotype belonged, it was observed that the genotype R 01 was the most divergent, as it manifested the highest average distance in the pairs of distances to which it belonged and that the genotype Rampa, with the lowest average distance, was the most similar. The objective for which the cassava germplasm was selected by farmers over several years appears to be the main cause of similarity in the clustering in each group. The genotypes clustered in the fifth group corresponded to those with the highest average productivities of tuberous roots, in addition to sharing a common upright habit of growing (Supplementary Data).

Table 1. Clustering of the germplasm of cassava from the Middle North Region of Brazil by the Tocher method, based on the distances obtained from quantitative and multicategorical characteristics.

Groups	Quantitative Data	Multicategorical Data
1	Talo Vermelho, Rosa, Rampa, Orelha de Leão, Pão, BRS Gema de Ovo	Rosa, Rampa, Talo vermelho, Orelha de Leão
2	BRS Dourada, Gameleira	Turiaçu, Gameleira, Pão
3	Turiaçu	BRS Dourada
4	R 01	BRS Gema de Ovo
5		R 01

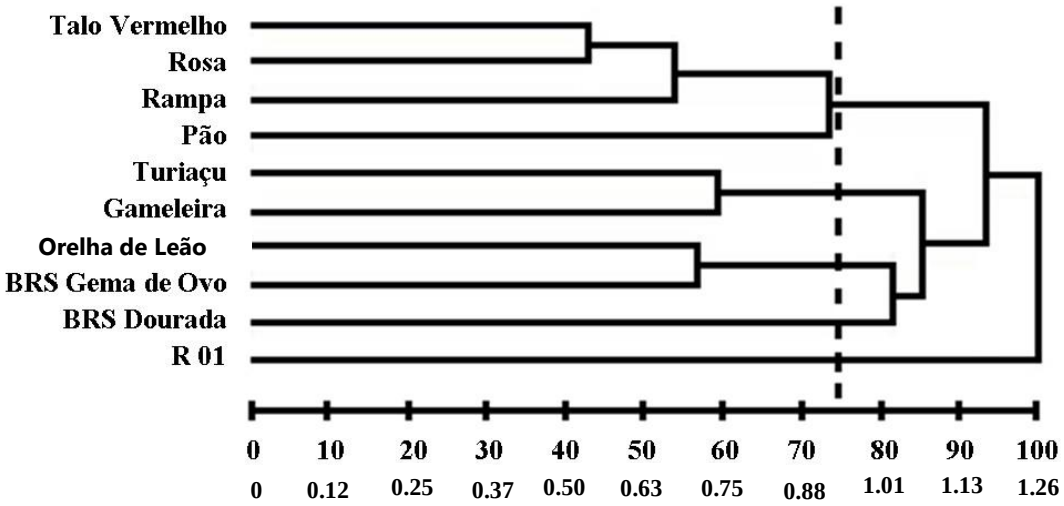

Fig 1. Dendogram showing the clustering of the genotypes of cassava based on the sum of the distance matrices, obtained by the method of complete linkage by means of the arithmetic complement of simple coincidence index.

Table 2. Correlation coefficient (*r*) between the matrices of distances obtained from the quantitative and multicategorical original data, between the matrices obtained from the different strategies and the matrices obtained from the original data.

----------Estrategies----------	---------- *r*----------	
Quantitative x Multicategorical	0.37	+
Gower x Quantitative	0.62	++
Gower x Multicategorical	0.82	++
Sum x Quantitative	0.84	++
Sum x Multicategorical	0.81	++
EDDA		
2 Classes x Quantitative	0.57	++
3 Classes x Quantitative	0.47	++
4 Classes x Quantitative	0.44	++
5 Classes x Quantitative	0.39	+
8 Classes x Quantitative	0.40	+
10 Classes x Quantitative	0.30	ns

EDDA: equitable division of the data amplitude. (**) (*) Significance at 1% and 5% by the Z test Z of Mantel, respectively, based on 10, 000.00 simulations.

Table 3. Cassava germplasm from the Middle North Region of Brazil, Chapadinha, MA, 2013.

Code	**Genotypes**	**Origin**
56- MA	BRS Dourada	Embrapa-PI
57-MA	BRS Gema de Ovo	Embrapa-PI
58-MA	Orelha Leão	Middle Mearim-MA
59-MA	Gameleira	Middle Mearim-MA
60-MA	Pão	São Luís- MA
61-MA	Turiaçu	Gurupi- MA
62-MA	R 01	Chapadinha-MA
63-MA	Rampa	Itapecuru Mirim- MA
64-MA	Rosa	Chapadinha-MA
65-MA	Talo Vermelho	Bico do Papagaio-TO

Table 4. Categories of descriptors and their classes used for the morphological characterization of the cassava germplasm (Fukuda and Guevara, 2010), Chapadinha, MA, 2013.

Number	Plant Descriptors	Given Categories
1	Branching habit	1-Erect; 2-Dichotomous; 3-Trichotomous and 4-Tetrachotomous.
2	Type of plant	1-Open; 2-Umbrella type and 3-Compact
	Leaf descriptors	
3	Apical leaf color	1- Light green; 2- Dark green; 3- Purplish-green and 4- Purple.
4	Pubescence of apical bud	1-Present and 2- Absente.
5	Petiole color	1-Yellowish-green; 2- Green; 3-Redish-green; 4- greenish-red; 5-red and 6-Purple
6	Developed leaf color	1-Light-green; 2-Dark-green; 3-Purplesh-green and 4-purple
7	Terminal branches color	1-Light-green; 2-Dark-green; 3-Purplesh-green and 4-purple
8	Leave´s rib color	1- Green; 2-Redish-green; and 3- Greenish-red
9	petiole position	1-Tilted up; 2-Horizontal; 3-Angled down and 4-Irregular
10	Prominence of leaf scars	1-Without prominence and 2-Proeminent.
	Stem Descriptors	
11	Color of stem cortex	1-Light yellow; 2-Light green; 3-Green and 4-Dark green.
12	Length of phyllotaxis	1-Short; 2-Middle and 3-Large.
13	External Color of steam	1-Orange; 2-Yllowish-green; 3-Golden; 4-Light brown; 5-Gray; 6- Silvery; 7- Gray; 8- Silvery; 9- Dark brown.
14	Color of stem epidermis	1- Cream; 2- Light brown; 3- Dark brown; 4- Yellow.
15	Growth habit of the stem	1-Straight and 2-Forked.
	Root descriptors	
16	Presence of peduncle in roots	1-Present and 2-Absent.
17	External color of roots	1-White; 2-Yellow; 3-Light brown; 4-Brown and 5-Dark brown.
18	Color of root Cortex	1-White; 2-Yellow and 3-Pinkish.
19	Texture of root epidermis	1-Smooth and 2-Rough.
20	Constriction of roots	1-Absent; 2-Little or none and 3-Average.
21	Root shape	1-Conical; 2-Cylinder and 3-Spindle.
22	Highlight pellicle from roots	1-Easy release and 2- Difficult release.
23	Highlight of roots cortex	1-Easy release and 2-Difficult release.
24	Position of roots	1-Horizontal and 2- Vertical tendency.

These characteristics of the main parameters considered by farmers in the selection of the genotypes of cassava in the region of germplasm collection corroborate the greater similarity observed between these genotypes. The dendogram cophenetic coefficient of correlation (r = 0.76) demonstrated a satisfactory adjustment of the graphic representation of the distances obtained from SUM, confirming the inferences based on the visual interpretation of Figure 1.

Materials and Methods

Germplasm

The germplasm was collected from autochthonous fields maintained by local families, which represented the most common model of cultivation of this crop in these states. A total of 10 genotypes were collected, of which, eight were ethno-varieties (Table 3). The two checks (BRS Dourada and BRS Gema de Ovo), developed for the Brazilian Northeast region by the Brazilian Enterprise of Agriculture and Research (Embrapa), were donated by one of the Embrapa Units (Embrapa Meio Norte), based in Teresina, PI.

The germplasm was collected from five different micro-regions (Table 3). In the fields, the cultivars considered different by farmers were collected, which characterized a fix sampling with a fix and directed model (Hershey, 1992; Martins, 1994).

Establishment and layout of the experiment

The experiment was established in January 2013 at the beginning of the regional raining season and conducted in the community of Vila União (3°44'34''S 43°21'07''W), in the rural zone of the municipality of Chapadinha, MA, Brazil.

The experiment was conducted under rain-fed conditions and eventual irrigation was used to compensate for the unusual shortage of rainfall that occurred during the experiment. The experiment was arranged in a completely randomized block design, with 4 replications and 10 treatments.

The experimental plots constituted of simple rows, 5-m long, with a spacing of 1.20 x 0.6 m between and within rows, respectively. Each row was composed by one only genotype and 10 plants. Of the 10 plants/row, the 6 central plants were evaluated, totalizing 24 plants for each genotype.

The selection of planting material, soil preparation, plantation, and fertilization were performed in accordance with the recommendation of the cassava system of production for the region of Cerrado (Sousa and Fialho, 2008).

Characterization of germplasm

The morphological characterization was conducted at 8 months after plantation, based on the descriptors and methodology proposed by Fukuda and Guevara (2010). The genotypes were characterized by 20 morphological (multicategorical) and 15 quantitative descriptors. The morphological descriptors were divided into the following 5 categories: i) plant descriptor, ii) leaf descriptor, iii) stem descriptor, and iv) tuberous root descriptor (Table 4).

The characterization by vegetative-quantitative descriptors was also performed at 8 months of plantation. For this purpose, the genotypes were evaluated with respect to the number of leaf lobes, length and width of the central leaf lobe, the relationship between the length and width of the

central lobe, the length of petiole, plant height, height of the first branching, the level of branching, and the number of stems. The agronomic-quantitative characterization was performed at the harvest, in the twelfth month after plantation. For this purpose, the genotypes were evaluated with respect to the mean productivity of tuberous roots, the harvest index, the number of roots per plant, the root length and diameter, and the length of phyllotaxis.

Statistical analysis

Initially, the analyses of genetic divergence were individually performed based on quantitative and multicategorical data. For the quantitative data, the matrix of distances was obtained from the mean values of each characteristic using the standard Euclidean distance, according to Equation 1. For the multicategorical data, the distance matrix was obtained by means of an arithmetic complement of a simple coincidence index, according to Equation 2.

$$d^2_{ii'} = \frac{1}{v} \sum j \, (\mathrm{Y}ij - Yi'j \,)^2 \qquad (1)$$

Where,
v = Number of characteristics studied;
$\mathrm{Y}ij$ = Value of the variable j for the individual i;
$Yi'j$ = Value of the variable j for the individual i´.

$$dii' = \frac{D}{C+D} \qquad (2)$$

Where,
C = Category agreement;
D = Category disagreement.

The analysis of correlation between the matrices of distances was verified by the Z test of Mantel at 5% probability, based on 10,000.00 simulations. After obtaining the matrices of distances, the analysis of genetic divergence was performed by the clustering method of Tocher, according to Cruz et al. (2011).
For the analysis of divergence based on the integration of quantitative and multicategorical data, three strategies were adopted: the Gower dissimilarity index (Equation 3), the transformation of quantitative data into multicategorical data, and the SUM of the distance matrices.

$$Sij = \frac{\sum_{k=1}^{p} Wijk \, . \, Sijk}{\sum_{k=1}^{p} Wijk} \qquad (3)$$

Where,
K corresponds to the number of variables (1, 2, 3, ..., p), and i and j correspond to the two individuals that represent the genotypes; Wijk corresponds to the weight of the comparison ijk, where the value 1 is assigned to the valid comparisons and 0 to the invalid ones; and Sijk corresponds to the contribution of variable K between the individuals i and j, with values varying from 0 to 1.
In the transformation of the data, the information of each characteristic was codified in the multicategorical data using the equitable division of the data amplitude (EDDA). With this strategy, the means of the characteristics were grouped into 2, 3, 4, 5, 8, or 10 classes, with the limits of each class established based on the EDDA.
For the SUM, the matrices of distances obtained from the quantitative and multicategoric data were summed, resulting in a third matrix denominated here by a "sum matrix", containing information from both the datasets.
After the integration of the data, 8 matrices of distances were obtained, one based on the Gower dissimilarity index and the others on the arithmetic complement of simple coincidence index. These matrices were also compared to the matrix of distance obtained from the quantitative characteristics by means of the Z test of Mantel in order to establish the most adequate strategy for data integration.
Once the most appropriate strategy for data integration was determined, the analysis of divergence was performed using the complete linkage method, from a single matrix of distance simultaneously comprising of quantitative and multicategorical information.
The analysis of genetic divergence and correlations were conducted with the aim of the Genes Statistical Program (Cruz, 2016).

Conclusion

The pattern of clustering of germplasm based on quantitative and multicategorical information was differentiated and seems to be associated to the nature of each dataset. The discrepancy in the estimation of genetic divergence obtained from each dataset represents the limitation in the assessment of genetic divergence in the germplasm of cassava, which limits the extrapolation of the estimates obtained from a dataset to another. The estimates of genetic divergence from a set of quantitative data do not predict the divergence expected from the multicategorical data and vice versa. The integration of data by the transformation of quantitative information into multicategorical one by means of the EDDA and the algorithm of Gower were not efficient and therefore resulted in low correlations with the distance matrix obtained from the original data. The SUM is consisted of the strategy of higher efficiency for the estimation of genetic divergence in the germplasm of cassava based on the integration of quantitative and multicategorical information. The SUM may provide a higher comprehensiveness and accuracy in the analysis of genetic in the germplasm of cassava.

Acknowledgments

This research was supported by the Foundation of Support to Research and Scientific Development of the State of Maranhão (Fundação de Amparo à Pesquisa e Desenvolvimento Científico e Tecnológico do Estado do Maranhão- FAPEMA). The support was granted to the first author as an extension/research scholarship under the grant AEXT-04127/10.

References

Alves AA, Bhering LL, Rosado TB, Laviola BG, Formighieri EF, Cruz CD (2013) Joint analysis of phenotypic and molecular diversity provides new insights on the genetic variability of

the Brazilian physic nut germplasm bank. Genet Mol Biol. 36: 371-381.

Coronado ACM, Coronado YM, Fregene M, Ramíres H, Chávez AL, Sánchez T, Morante N, Lascano HC (2011) Diversidad genética y contenido de carotenos totales en accesiones de yuca (*Manihot esculenta* Crantz). Acta Agron. 60: 97-107.

Cruz CD, Ferreira FM, Pessoni LA (2011) Biometria aplicada ao estudo da diversidade genética. Suprema, Visconde do Rio Branco, Brazil, 620p.

Cruz CD (2016) Genes software – extended and integrated with the r, matlab and selegen. Acta Sci. 38: 547-552.

Da Silva FL, Baffa DCF, Oliveira ACB de, Pereira AA, Bonono VS (2013) Integration of quantitative and multicategorical data in determining the genetic divergence among accessions of coffee trees. Bragantia. 44: 224-229.

Da Silva KN, Vieira EA, Fialho JF, Carvalho LJCB, Silva MS (2014) Agronomic potential and carotenoid contents within cassava storage roots. Cienc Rural. 44: 1348-1354.

Do Carmo CD, Silva MS, Oliveira GAF, Oliveira EJ (2015) Molecular-assisted selection for resistance to cassava mosaic disease in *Manihot esculenta* Crantz. Sci Agric. 72: 520-527.

Fonseca AFA da, Sediyama T, Cruz CD, Sakaiyama NS, Ferrão MAG, Ferrão RG, Bragança SM (2006) Genetic divergence in conilon coffee. Pesqui Agropecu Bras. 41: 599-605.

Fukuda WMG, Guevara CL, Kawuki R, Ferguson ME (2010) Selected morphological and agronomic descriptors for the characterization of cassava. International Institute of Tropical Agriculture (IITA), Ibadan, Nigeria, 19p.

Gomes CN, Carvalho SP de, Jesus MAS, Custódio TN (2007) Morphoagronomic characterization and path analysis of production components in cassava clones. Pesqui Agropecu Bras. 42: 1121-1130.

Gomes RS, De Almeida CF, COSTA JRS , Junior RM, Delazari FT, Santos FCS, Da Silva DJH (2016) Genetic diversity in sweet cassava from the Brazilian Middle North Region and selection of genotypes based on morpho-agronomical descriptors. Afr J Agr Res. 11: 3710-3719.

Hershey CH (1992) Manihot esculenta diversity. In: CIAT, IITA and IBPGR (ed) International network for cassava genetic resources. International Plant Genetic Resources Institute Press, Cali. 4.

Martins FA, Carneiro PCS, Silva DJH da, Cruz CD, Carneiro JE de S (2011) Integration of data in studies of genetic diversity of tomato. Pesqui Agropecu Bras. 46: 1496-1502.

Martins PS (1994) Biodiversity and agriculture: patterns of domestication of Brazilian native plants species. An Acad Bras Cienc. 66: 219-226.

Mezette TF, Blumer CG, Veasey EA (2013) Morphological and molecular diversity among cassava genotypes. Pesqui Agropecu Bras. 48: 510-518.

Mezette TF, Carvalho CRL, Morgano MA, Silva MG da, Parra ESB, Galera JMSV, Valle TL (2009) Selection of sweet cassava elite-clones for agronomical, technological and chemical characteristics. Bragantia. 68: 601-609.

Nery-Silva FA, Fernandes JJ, Juliatti FC, Melo B de (2007) Reaction of cassava's germoplasm to X*anthomonas axonopodis* pv. *manihots*. Semin Cienc Agrar. 28: 3-10

Okpara DA, Mbah EU, Ojikpong TO (2014) Association and path coefficients analysis of fresh root yield of high and low cyanide cassava (*Manihot esculenta* Crantz.) genotypes in the humid tropics. J Crop Sci Biotechnol. 17: 103-109.

Purwaity RD, Herwati (2016) A Short Communication: Evaluation of quantitative and multicategoricalmorphological characters of sunflower (*Helianthus* annuus.) germplasm. Biodivers. 17: 461-465.

Rohlf FJ (2000) NTSYS-pc: numerical taxonomy and multivariate analysis system. New York, Exeter Software.

Scarano D, Rubio F, Ruiz JJ , Rao R, Corrado G (2014) Morphological and genetic diversity among and within common bean (*Phaseolus vulgaris* L.) landraces from the campania region (Southern Italy). Sci Hortic. 80: 72–78.

Sousa LS, Fialho JF (2008) Cassava system of production for the region of Cerrado. Embrapa Press, Cruz das Almas, Brazil, 208p.

Valois ACC (2005) Acesso aos recursos genéticos e repartição de benefícios: uma visão atual e de futuro. In: Lima M da C (ed) Recursos genéticos de hortaliças: riquezas naturais, 1 edn, Inter-American Institute for Cooperation on Agriculture, São Luis. 3.

Vilas Boas SA, Hohenfeld CS, Oliveira SAS de, Vanderlei SS da, Oliveira EJ de (2016) Sources of resistance to cassava root rot caused by fusarium spp.: a genotypic approach. Euphytica. 209:237–251.