Multivariate analysis in a genetic divergence study of *Psidium guajava*

A.M. Nogueira[1], M.F.S. Ferreira[1], J.H.S. Guilhen[2] and A. Ferreira[2]

[1]Laboratório de Genética e Melhoramento, Departamento de Biologia,
Centro de Ciências Agrárias, Universidade Federal do Espírito Santo, Alegre,
ES, Brasil
[2]Laboratório de Genética e Melhoramento, Departamento de Produção Vegetal,
Centro de Ciências Agrárias, Universidade Federal do Espírito Santo, Alegre,
ES, Brasil

Corresponding author: M.F.S. Ferreira
E-mail: mfloressf@gmail.com

ABSTRACT. The family Myrtaceae is widespread in the Atlantic Forest and is well-represented in the Espírito Santo State in Brazil. In the genus *Psidium* of this family, guava (*Psidium guajava* L.) is the most economically important species. Guava is widely cultivated in tropical and subtropical countries; however, the widespread cultivation of only a small number of guava tree cultivars may cause the genetic vulnerability of this crop, making the search for promising genotypes in natural populations important for breeding programs and conservation. In this study, the genetic diversity of 66 guava trees sampled in the southern region of Espírito Santo and in Caparaó, MG, Brazil were evaluated. A total of 28 morphological descriptors (11 quantitative and 17 multicategorical) and 18 microsatellite markers were used. Principal component, discriminant and cluster analyses, descriptive analyses, and genetic diversity analyses using simple sequence repeats were performed. Discrimination of accessions using molecular markers resulted in clustering of genotypes of the same origin, which

was not observed using morphological data. Genetic diversity was detected between and within the localities evaluated, regardless of the methodology used. Genetic differentiation among the populations using morphological and molecular data indicated the importance of the study area for species conservation, genetic erosion estimation, and exploitation in breeding programs.

Key words: Breeding; Conservation; Guava; Microsatellite markers; Morphological descriptors

INTRODUCTION

The family Myrtaceae includes 130-150 genera and 5650 species with centers of diversity in Australia, Southeast Asia, and between the tropical and subtropical Americas (Govaerts et al., 2008). The genus *Psidium* includes approximately 150 species (Pereira and Nachtigal, 2003), of which 59 are present in Brazil and 10 are present in Espírito Santo (Sobral et al., 2012). However, of these species, only *P. guajava* L. "guava" is cultivated. Brazil is one of the largest producers of guava worldwide (FAO, 2010).

Guava is native to the American continent and has a very broad center of origin from Mexico to Peru and Brazil (Pereira and Kavati, 2011). Guava cultivation is economically important in many tropical and subtropical countries (Rodríguez et al., 2010), and guava fruits are valued for their taste and high nutritional value because they contain vitamins A, B, and C, as well as calcium, zinc, phosphorus, and iron (Singh, 2005). Furthermore, the fruits, leaves, flowers, roots, bark, and stems are important for their medicinal properties (Gutiérrez et al., 2008).

Cross-pollination is the reproductive form most frequently exhibited by *P. guajava* (Alves and Freitas, 2007); however, self-pollination has also been reported (Singh and Seghal, 1968). This pollination mechanism is associated with seed dispersal and adaptability to different edaphoclimatic conditions and contributes to increased genetic variability within the species.

The cultivation of few guava cultivars is concerning as it may lead to genetic vulnerability of the crop. This situation can be observed in Brazil, where approximately 70% of the guava trees currently cultivated for industrial processing belong to the cultivar Paluma (Pereira and Kavati, 2011).

Genetic diversity studies are useful for pre-breeding and breeding programs, are the foundation of germplasm banks, and can be used to guide conservation strategies. These studies are performed using phenotypic and molecular data, which provide complementary information regarding each genotype. Because various characteristics are evaluated in these studies and these characteristics consist of different data types, multivariate techniques are useful for evaluating genetic diversity.

The genetic diversity of guava (wild and cultivated) and the identification of divergent genotypes worldwide has been reported in studies based on morphological (Hernández-Delgado et al., 2007; Urdaneta et al., 2007; Santos et al., 2008; Lozano et al., 2009; Fernandes Santos et al., 2010; Padilla-Ramírez and González-Gaona, 2010), molecular, and biochemical data. Molecular and biochemical studies used dominant markers, such as random amplified polymorphic DNAs (RAPDs) (Chen et al., 2007; Pessanha et al., 2011) and amplified fragment length polymorphisms (AFLPs) (Rodríguez et al., 2004; Hernández-Delgado et al., 2007; Corrêa et al., 2011). A number of studies have used simple sequence repeats (SSRs),

which are co-dominant markers (Valdés-Infante et al., 2007; Briceño et al., 2010; Herrero Juliette et al., 2010; Teyer et al., 2010). These markers are preferred because they are informative and provide a large amount of data for analyses, which helps to increase the understanding of genetic diversity within a species. However, few studies have used different types of markers simultaneously (Gomes-Filho et al., 2010).

Comparative studies using morphological and molecular data from SSRs to analyze genotypes from different localities can contribute to a better understanding of the genetic diversity of *P. guajava*, an estimation of genetic erosion, and the identification of promising diverging genotypes for breeding. In the present study, we evaluated the genetic diversity between and within guava trees sampled from 7 localities in the southern regions of Espírito Santo and Caparaó (Minas Gerais, Brazil) at different altitudes, which contain remnants of the Atlantic Forest, using morphological descriptors as well as SSR markers. Our goal was to identify diverging promising genotypes for breeding and to verify complementarity of the data for genetic diversity analysis using multivariate analysis.

MATERIAL AND METHODS

Samples from 66 guava trees were collected between November 2009 and January 2011 at different locations in southern Espírito Santo and Caparaó (Minas Gerais, Brazil) at different altitudes. Between 5 and 17 plants were collected from each location (Figure 1).
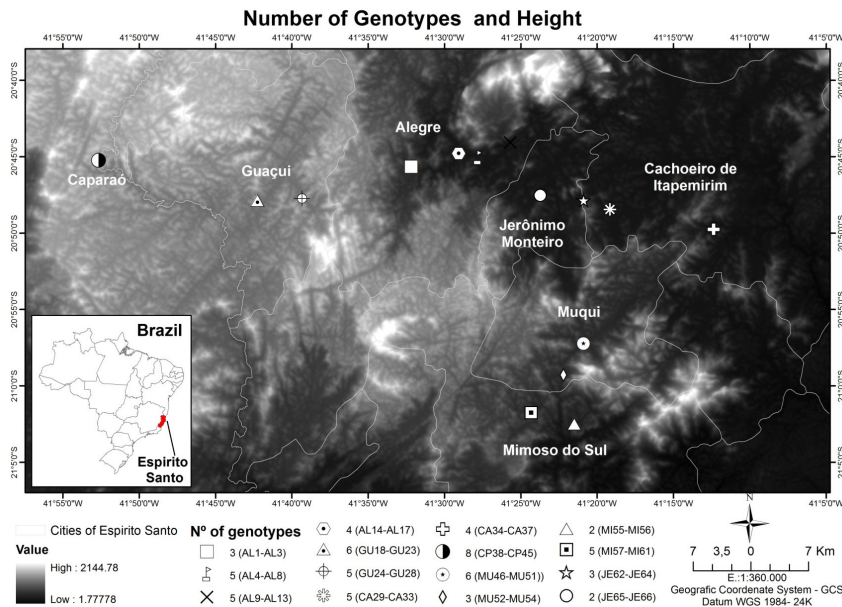

Figure 1. Geographic distribution of the genotypes sampled and the altitudes.

Twenty-eight descriptors, including 11 quantitative and 17 multicategorical, were evaluated in productive adult plants in accordance with the International Union for the Protection of New Varieties (UPOV, 1987). From each plant, 5 young leaves, 10 developed leaves,

and 5 ripe fruits were analyzed. Quantitative descriptors included the length and width of the young leaf (cm), length and width of the developed leaf (cm), fruit mass (g), fruit length and width (cm), pulp mass (g) and diameter (cm), and seed mass (g) and number. Multicategorical descriptors used included stem color and branch distribution of the superior appendages, shape of the leaf basis and apexes, shape of the stalk end, presence and degree of marginal undulation, presence and degree of nervure and transversal curvatures, fruit shape, external coloration, color uniformity, surface texture, base shape, and mesocarp coloration.

The genomic DNA of each plant was extracted in accordance with the protocol of Doyle and Doyle (1990). DNA was extracted from the bulk of young leaves collected from saplings of germinated fruit seeds or from young leaves of the matrix plant. The samples were quantified using spectrophotometry, after which the purity was estimated.

For the polymerase chain reaction (PCR) procedure, the following 18 primers were used: mPgCIR02, mPgCIR07, mPgCIR08, mPgCIR09, mPgCIR10, mPgCIR11, mPgCIR13, mPgCIR14, mPgCIR15, mPgCIR16, mPgCIR17, mPgCIR18, mPgCIR19, mPgCIR20, mPgCIR21, mPgCIR22, mPgCIR25, and mPgCIR26 (Risterucci et al., 2005). PCR amplification conditions were as follows: 60 ng DNA and 0.4 µM of each primer in a final volume of 17 µL, containing 1X PCR master mix (2x) (Fermentas, Vilnius, Lithuania). The PCR amplification program used was as follows: 4 min at 94°C, followed by 30 cycles of 45 s at 94°C, 1 min at 55°C, and 2 min at 72°C, followed by final extension of 8 min at 72°C (Risterucci et al., 2005). The PCR products were separated using 3% agarose gel electrophoresis and stained with ethidium bromide.

Genetic diversity of the genotypes based on quantitative and multicategorical data were analyzed by calculating the mean standardized Euclidean distance of the genotypes. Molecular data were used to calculate the dissimilarity of the genotypes from a weighted index (Cruz, 2013). Subsequently, cluster analysis of the genotypes was performed using the unweighted pair group method with arithmetic averages (UPGMA).

Principal component analysis was performed to verify the most important quantitative characteristics. Genotypes were evaluated in relation to localities (groups) using a discriminant analysis by the *k*-nearest neighbor method to estimate the apparent error rate (AER) relative to the formation of groups for the quantitative, multicategorical, and molecular data, considering k of 4 individuals. Cluster analysis using UPGMA was performed using the genetic distances obtained from the quantitative, multicategorical, and molecular data and was based on the Gower algorithm (Gower, 1971). Pearson correlation was calculated between distances obtained from the quantitative, multicategorical, and molecular data and using the Gower algorithm. Analyses were performed using the Genes program (Cruz, 2013) and the R Development Core Team program (2011).

RESULTS

A large degree of variation was verified among the 66 genotypes for the 11 quantitative characteristics, and the ratio between the highest and lowest values ranged from 5-6 for the weight of the pulp, fruit, and seed and number of seeds. For other characteristics, variation ranged from 2-3. The 3 principal components explained 87.86% of the variation in the genotypes in relation to these characteristics as follows: the first, second, and third components explained 49.15, 26.45, and 12.26% of the total variation, respectively. The variables of greatest importance were length, fruit diameter, pulp mass, developed leaf length, and seed number.

In the discriminant analysis using the *k*-nearest neighbor method, the lowest AER was obtained using molecular data (22.7%), followed by multicategorical (53.0%) and the quantitative (71.2%) data, and a higher concordance in genotype classification by origin was demonstrated using molecular data (Table 1).

Table 1. Discriminant analysis by *k*-nearest neighbor of guava trees from different localities (pre-defined groups) estimated from standardized mean Euclidean distance matrix of quantitative and multicategorical data and from the distance matrix by pondered index of molecular data.

Data type	No. of genotypes	Locality*	Percentage of genotype allocation in the groups						
			AL	GU	CA	CP	MU	MI	JE
	17	AL	41.2	-	58.8	-	-	-	-
	11	GU	18.2	45.6	36.4	-	-	-	-
Quantitative	9	CA	33.3	22.2	44.4	-	-	-	-
	8	CP	50.0	-	37.5	-	-	12.5	-
	9	MU	33.3	33.3	22.2	11.1	-	-	-
	7	MI	42.9	28.6	28.6	-	-	-	-
	5	JE	-	-	40.0	-	-	-	60.0
	Allocated *a posteriori*	47	31.9	14.9	48.9	2.1	-	2.1	-
	Total	66	22	12	27	1	-	1	3
	17	AL	-	47.1	5.9	17.6	5.9	-	23.5
	11	GU	-	90.9	-	-	-	-	9.1
Multicategorical	9	CA	-	22.2	33.3	-	-	-	44.4
	8	CP	-	25.0	-	50.0	12.5	-	12.5
	9	MU	-	-	33.3	-	33.3	-	33.3
	7	MI	-	-	-	-	-	100.0	-
	5	JE	-	-	20.0	-	-	-	80.0
	Allocated *a posteriori*	35	-	34.29	14.29	8.57	5.71	-	37.14
	Total	66	-	22	8	7	5	7	17
	17	AL	76.5	-	5.9	-	17.7	-	-
	11	GU	-	90.9	-	-	-	-	9.1
Molecular	9	CA	22.2	-	77.8	-	-	-	-
	8	CP	-	-	25.0	62.5	-	-	12.5
	9	MU	-	-	-	-	88.9	11.1	-
	7	MI	-	-	-	-	42.9	42.9	14.3
	5	JE	-	-	-	-	-	-	100.0
	Allocated *a posteriori*	15	13.33	-	20	-	40	6.67	20
	Total	66	15	10	10	5	14	4	8

Locality* = Alegre (AL), Guaçuí (GU), Cachoeiro (CA), Caparaó (CP), Muqui (MU), Mimoso (MI), and Jerônimo Monteiro (JE).

Based on the quantitative data in this analysis, the genotypes from Caparaó (CP), Muqui (MU), and Mimoso do Sul (MI) were allocated into groups (localities) that differed from those established previously. Groups from Jerônimo Monteiro (JE) and Muqui did not include genotypes from other localities. Genotypes from Alegre (AL), Guaçuí (GU), and Cachoeiro (CA) were allocated with the genotypes of nearly all groups (Table 1).

Using multicategorical data, the Mimoso do Sul genotypes were allocated within the original group without the inclusion of other genotypes. Groups from Guaçuí and Jerônimo Monteiro showed a high percentage of correct allocations but included genotypes from other groups. The group from Alegre did not cluster with any other genotype (Table 1).

Based on molecular data, a high percentage of genotype allocation in the pre-established groups was observed, with the exception of Mimoso do Sul. Genotypes from Jerônimo Monteiro were allocated within their own group along with other genotypes. Most genotypes from Guaçuí were allocated within their own group without other genotypes. Other groups exhibited a high percentage of correct clustering, and most groups showed grouping of geno-

types from 2 localities (Table 1).

Cluster analysis using UPGMA exhibited divergences between genotypes within and between localities depending on the characteristic evaluated. The greatest divergence of genotypes within localities was verified using quantitative traits and was demonstrated based on the sequential cluster of only 2 genotypes from the same locality (AL2-AL6, AL16-AL17, GU4-GU5, GU8-GU10, and CA2-CA6) (Figure 2A). This type of clustering was also observed in multicategorical data analysis (GU1-GU5, GU7-GU10, CP2-CP6, CP3-CP4, and MU6-MU8), which also verified more genotypes by locality (MI1, MI3, MI4, and MI7) (Figure 2B).
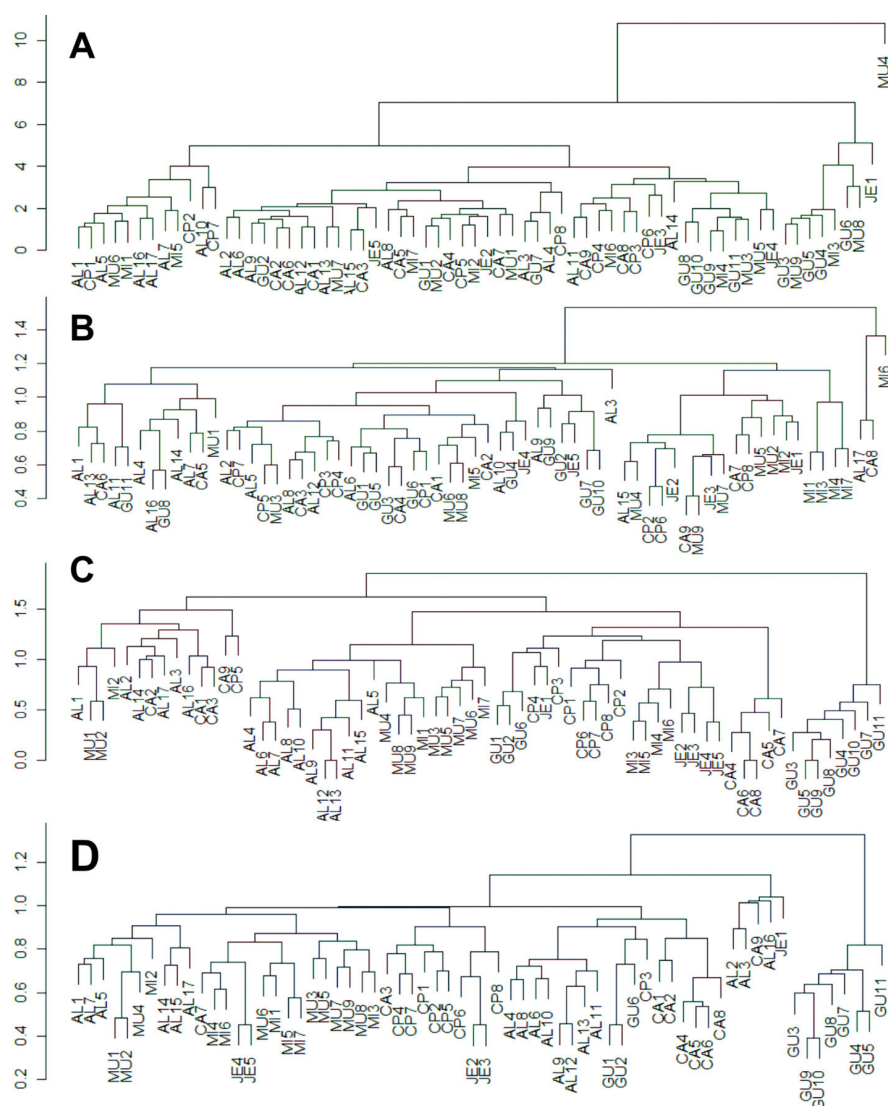

Figure 2. Dendrogram of *Psidium guajava* genotypes obtained from evaluating the quantitative (**A**) and multicategorical (**B**) characteristics and microsatellite markers (**C**) and using the Gower algorithm (**D**).

Molecular data allowed for the clustering of most genotypes according to origin (except for AL1, AL2, AL3, AL14, AL16, AL17, CA2, CA9, CP4, CP5, MI1, MI2, MI7, and JE1) (Figure 2C). These results corroborated the discriminant analysis data, reinforcing the higher concordance in genotype classification according to the origin provided by the molecular data and exhibiting greater similarities between the genotypes composing each group.

Analysis using the Gower algorithm of quantitative, multicategorical, and molecular data simultaneously allowed for identification of similarities and variabilities among genotypes of the same locality. Clustering by origin was verified with the exception of the genotypes AL16, CA3, CA7, CA9, CP3, CP8, MU6, MI1, MI2, MI3, and JE1 (Figure 2D).

The largest distances obtained with the quantitative data were observed between genotype MU8 (in relation to CP2, MI1, AL16, MU6, and AL17) and genotype GU6 (in relation to CP2 and MU4). Based on multicategorical data, the greatest dissimilarities were observed between genotypes CA8 and MI1 (0.74) and between MU7 and MI6 (0.72). Based on molecular data, the greatest dissimilarities occurred between genotypes CP5 and MU7 (0.82), CP5 and AL4 (0.80), and CP5 and MI1 (0.79). According to Gower analysis, the greatest dissimilarities occurred between genotypes GU4 and AL7 (0.72) and genotypes GU4 and AL16 (0.72).

Based on the analysis of dissimilarity by locality according to quantitative data, Guaçuí was clearly the most divergent locality in relation to the other areas (Figure 3A). In multicategorical analyses, the greatest similarity occurred between Cachoeiro de Itapemirim and Muqui, Alegre and Guaçuí, and Caparaó and Jerônimo Monteiro (Figure 3B). Molecular data showed that the most similar populations were Muqui and Mimoso do Sul, followed by Alegre, Guaçuí, and Caparaó (Figure 3C).
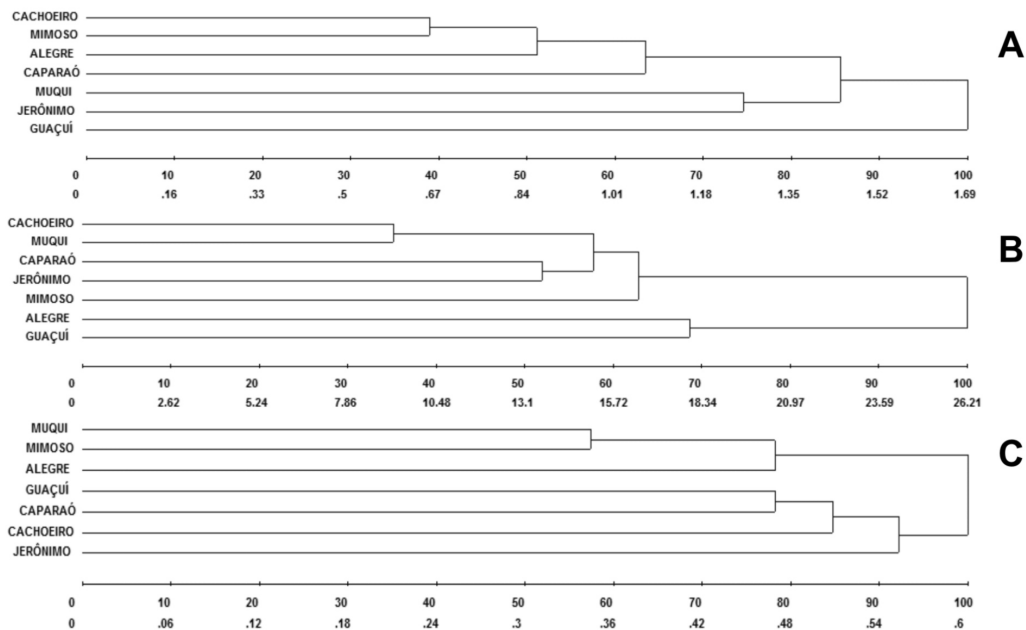

Figure 3. Dendrogram of genotype groups by locality established using the UPGMA method based on the quantitative (**A**), multicategorical (**B**), and molecular (**C**) data.

　　　　The greatest and most significant correlation (r = 0.73) was detected between the data matrix obtained using the Gower algorithm and the matrix of molecular data. Correlations between the Gower matrix and quantitative and multicategorical data were also significant (r = 0.26 and 0.21, respectively) (Figure 4).
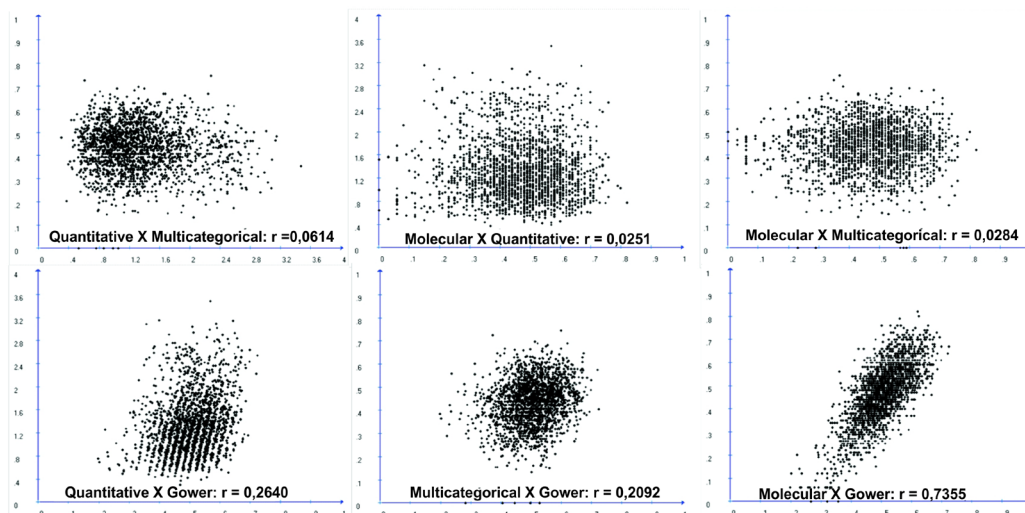

Figure 4. Graphic dispersion of the distance matrices for the quantitative, multicategorical, molecular, and Gower data and Spearman correlations (r) between the matrices.

DISCUSSION

　　　　The large degree of variation in quantitative characteristics revealed differences among genotypes. Based on these characteristics, principal component analysis demonstrated that 87.86% of the variance detected in the 66 genotypes could be explained by 3 principal components and that the fruit-related characteristics were important for discriminating genotypes.

　　　　In studies by Sanabria et al. (2005) and Lozano et al. (2009), fruit-related characteristics were also found to be the most important characteristic, accounting for 72% and 76.86% of accumulated variation, respectively, in the 3 principal components. Lozano et al. (2009) evaluated 12 quantitative characteristics in 22 accessions of *P. guajava*, while Sanabria et al. (2005) analyzed 25 characteristics in 53 accessions. In contrast, Urdaneta et al. (2007) observed 100% variation in 3 principal components when evaluating 18 quantitative and qualitative descriptors of plants, stems, and leaves; however, only 4 cultivars were analyzed, which may have influenced the variation observed in the 3 components. Furthermore, Hernández-Delgado et al. (2007) reported that less than 30% of the total variation could be explained by the 3 first principal components in the evaluation of 50 descriptors in 48 accessions of *P. guajava*, 2 of *P. cattleianum*, and 2 of *P. friedrichsthalianum*. In this study, the large number and classification types of descriptors, in addition to the inclusion of different species, may have accounted for the low total variance explained by the 3 principal components.

　　　　Discriminant analysis revealed dissimilarities of genotypes within and between localities, and was verified by the inclusion or removal of genotypes from pre-established groups (source group or pre-defined). Molecular data exhibited the greatest efficiency in allocating

genotypes according to the locality of origin, followed by quantitative and multicategorical data.

In wild (Nielsen et al., 2003) as well as in cultivated (Roldán-Ruiz et al., 2001) plants, low concordance has been reported between molecular and phenotypic genetic divergence. This scenario is commonly reported, except when the molecular markers used are closely linked to the genes controlling the morphological characteristics analyzed. Nielsen et al. (2003) applied discriminant analysis to 6 natural populations from the genus *Scalesia* and found an increase in the discrimination of individuals in 5 of these populations using AFLP markers compared with using morphometric data.

The faulty allocation of genotypes by locality using quantitative and qualitative data may be linked to the mixed reproductive system of guava, which may have contributed to increased genetic variability. Furthermore, quantitative characteristics are controlled by several genes and are influenced by environmental factors. In contrast, molecular markers allow for access to the entire genome without environmental influence, explaining their efficiency in allocating genotypes into their pre-established groups.

The results of cluster analysis corroborated the data from discriminant analysis, reinforcing the high concordance of genotype classification by origin obtained using molecular data. The number of genotypes from different locations in the same group was reduced by simultaneously using quantitative, multicategorical, and molecular data. Moreover, genotypes in the same locality exhibited morphological divergence, despite high molecular similarity, whereas other genotypes were not grouped, independent of data type, reflecting the high variability of these plants.

In a study of the genetic diversity of 25 guava trees, Gomes-Filho et al. (2010) also reported high discrimination of genotypes using molecular markers, as demonstrated by the 2-fold increase in the number of groups formed using data from the 28 RAPD descriptors compared with the 6 morphological characteristics.

Based on grouping analysis by locality, the results again demonstrated higher clustering adequacy using molecular data because grouping by locality was more coherent in relation to localization of sampling in the districts.

With respect to the distances between genotypes obtained from different data types, the most divergent genotypes did not coincide among characteristics. Therefore, diverging genotypes with adequate breeding characteristics must be defined for use in hybridization programs to develop new cultivars.

To investigate the diversity of genotypes while considering all data types simultaneously, the Gower algorithm was used, allowing for the grouping of all genotypes in a single dendrogram. Using this algorithm, genotype variability was represented by all data types at once, clarifying the variability of the genotypes between and within localities. This analysis also enabled identification of similarities between genotypes according to origin, as was achieved using molecular data. The genotypes AL16, CA9, MI1, MI2, and JE1 were not grouped in any locality, independently of grouping criterion, suggesting the high variabilities of these plants.

No correlation was observed between dissimilarity matrices of the quantitative, multicategorical, and molecular data. Nevertheless, the matrices were significantly correlated with the Gower matrix, although the correlation was much higher for the molecular data ($r = 0.73$) in relation to the other types of data ($r = 0.21$ and 0.26 for multicategorical and quantitative data, respectively), indicating that the Gower matrix better represented the data for discriminating between genotypes at the highest level, i.e., molecular data.

The combination of data from the 15 morpho-agronomical characteristics, 19 RAPDs, and 20 inter-simple sequence repeats using the Gower algorithm was considered more efficient for characterizing genetic diversity in 32 progenies of backcrossed papaya (*Carica papaya*) compared with individual analysis of each data source (Ramos et al., 2012).

Simultaneous evaluation of quantitative, qualitative (Rodríguez et al., 2005; Rocha et al., 2010), and molecular data (Krichen et al., 2008), or a combination of these data (Gonçalves et al., 2008; Ramos et al., 2012) using the Gower algorithm, is useful for estimating genetic divergence in different crops. However, few studies have adopted this strategy (Rocha et al., 2010). Because difficulties have been reported previously concerning the application of this algorithm (Gonçalves et al., 2008), the comparison of analyses with and without this algorithm can be used as a guide.

This method for the simultaneous analysis of quantitative and qualitative characteristics was proposed originally by Gower (1971), making it possible to estimate the similarity between 2 individuals using an algorithm. However, we found that the algorithm is equally efficient for molecular data, such as the then-employed microsatellites. The lack of correlation between matrices of dissimilarity for the different data types justified the different groupings and was confirmed by the presence of groups containing a single individual.

According to Souza Júnior et al. (2002), the guava tree is susceptible to interactions between the genotype and environment, which is an essential factor for phenotypic expression. In the analyses of morphological data, most genotypes exhibited variable distributions in the groupings, independently of origin. This variation may be related to the mixed pollination system of the plant, which amplifies genetic variability.

Based on these results, genetic divergence was verified between and within genotypes of different localities. This information is relevant for executing crossings between divergent genotypes as well as for directing the collection of germplasm samplings to represent the genetic diversity of a region and to minimize duplicates in germplasm banks.

In conclusion, genetic diversity was observed within and between genotypes from different study locations. Combining morphological and SSR molecular marker data was sufficient for determining the genetic divergence between and within the collection sites, independently of the analysis methodology adopted and revealing genetic variability for breeding and conservation.

ACKNOWLEDGMENTS

REFERENCES

Alves JE and Freitas BM (2007). Requerimentos de polinização da goiabeira. *Cienc. Rur.* 37: 1281-1286.
Briceño A, Aranguren Y and Fermin G (2010). Assessment of guava-derived SSR markers for the molecular characterization of Myrtaceae from different ecosystems in Venezuela. *Acta Hort.* 849: 139-146.
Chen TW, Chaing C, Wang CY and Shyu YT (2007). Molecular identification and analysis of *Psidium guajava* L. from indigenous tribes of Taiwan. *J. Food Drug. Anal.* 15: 82-88.
Corrêa LC, Santos CAF, Lima GPP and Rodrigues MA (2011). Genetic similarity among accessions of guava and Brazilian guava araçazeiros based on AFLP markers. *Rev. Bras. Frutic.* 33: 859-867.

Cruz CD (2013). GENES - a software package for analysis in experimental statistics and quantitative genetics. *Acta Sci.* 35: 271-276.

Doyle JJ and Doyle JL (1990). Isolation of plant DNA from fresh tissue. *Focus*13-15.

FAO (2010) - CCP. Food and Agriculture Organization of the United Nations - Committee on Commodity Problems. Available at [http://www.fao.org/unfao/bodies/ccp/ccp66/Index_en.htm]. Accessed June 2, 2012.

Fernandes Santos CA, Castro JM da C e, Souza FF and Vilarinho AA (2010). Prospecting and morphological characterization of Brazilian *Psidium* germplasm. *Acta Hortic.* 43: 69-76.

Gomes-Filho A, Oliveira JG, Viana AP and Siqueira APO (2010). RAPD molecular markers and morphological descriptors in the evaluation of genetic diversity of guava (*Psidium guajava* L.). *Acta Sci.* 32: 627-633.

Gonçalves LS, Rodrigues R, Amaral AT Jr, Karasawa M, et al. (2008). Comparison of multivariate statistical algorithms to cluster tomato heirloom accessions. *Genet. Mol. Res.* 7: 1289-1297.

Govaerts R, Sobral M, Ashton P and Barrie F (2008). World Checklist of Myrtaceae. Kew Publishing, Cumbria.

Gower JC (1971). A general coefficient of similarity and some of its properties. *Biometrics* 27: 857-874.

Gutiérrez RM, Mitchell S and Solis RV (2008). *Psidium guajava*: a review of its traditional uses, phytochemistry and pharmacology. *J. Ethnopharmacol.* 117: 1-27.

Hernández-Delgado S, Padilla-Ramírez JS, Nava-Cedillo A and Mayek-Pérez N (2007). Morphological and genetic diversity of Mexican guava germplasm. *Plant Genet. Res.* 5: 131-141.

Herrero Juliette VI, Narciso Nerdo RM, Alor Martin B and Matilde Margarita OG (2010). Microsatélites desarrollados en guayabo (*Psidium guajava* L.) y su utilidad para evaluar diversidad en la familia Myrtaceae. *Rev. Colomb. Biotecnol.* 12: 64-76.

Krichen L, Martins JMS, Lambert P and Daaloul A (2008). Using AFLP markers for the analysis of the genetic diversity of apricot cultivars in Tunisia. *Am. Soc. Hort. Sci.* 133: 204-212.

Lozano LJ, Pinzón MIA and Flórez JEM (2009). Morphological characterization of wild guava access. *Acta Agron.* 58: 69-73.

Nielsen LR, Cowan RS, Siegismund HR and Adsersen H (2003). Morphometric, AFLP and plastid microsatellite variation in populations of *Scalesia divisa* and *S. incisa* (Asteraceae) from the Galápagos Islands. *Bot. J. Linn. Soc.* 143: 243-254.

Padilla-Ramírez JS and González-Gaona EG (2010). Collection and characterization of Mexican Guava (*Psidium guajava* L.) Germplasm. *Acta Hort.*49-54.

Pereira FM and Nachtigal JC (2003). Melhoramento da Goiabeira. In: Cultura da Goiabeira: Tecnologia e Mercado (Rozane DE and Couto Fad'a, eds.). Universidade Federal de Viçosa, Viçosa, 53-78.

Pereira FM and Kavati R (2011). Contribuição da pesquisa científica brasileira no desenvolvimento de algumas frutíferas de clima subtropical. *Rev. Bras. Frutic.*108.

Pessanha PGO, Viana AP, Amaral Júnior AT and Souza RM (2011). Assessment of genetic diversity in access to *Psidum* spp. via RAPD markers. *Rev. Bras. Frutic.* 33: 129-136.

R Development Core Team. (2014) R: A Language and Environment for Statistical Computing. Vienna: R Foundation for Statistical Computing. Available at [http://www.R-project.org http://www.R-project.org].

Ramos HC, Pereira MG, Gonçalves LS, Berilli AP, et al. (2012). Multivariate analysis to determine the genetic distance among backcross papaya (*Carica papaya*) progenies. *Genet. Mol. Res.* 11: 1280-1295.

Risterucci AM, Duval MF, Rohde W and Billotte N (2005). Isolation and characterization of microsatellite loci from *Psidium guajava* L. *Mol. Ecol. Notes* 5: 745-748.

Rocha MC, Gonçalves LSA, Rodrigues R and Silva PRA (2010). Uso do algoritmo de Gower na determinação da divergência genética entre acessos de tomateiro do grupo cereja. *Acta Sci. Agron.* 32: 423-431.

Rodríguez NN, Valdéz-Infante J, Becker D and Velásques B (2004). Caratterizzazione morfologica, agronomica e molecolare delle accessioni cubane di guaiava (*Psidium guaiava* L.). *J. Genet. Breed.* 58: 79-90.

Rodríguez VM, Cartea ME, Padilla G and Velasco P (2005). The nabicol: A horticultural crop in northwestern Spain. *Euphytica* 142: 237-246.

Rodríguez NN, Valdés-Infante JJ, Rodríguez JA and Velásquez JB (2010). Genetic resources and breeding of guava (*Psidium guajava* L.) in Cuba. *Biotecnol. Apl.* 27: 238-241.

Roldán-Ruiz I, van Eeuwuk FA, Gilland TJ and Dubreul P (2001). A comparative study of molecular and morphological methods of descriting relationships between perennial ryegrass (*Lolium perenne* L.) varieties. *Theor. Appl. Genet.* 103: 1138-1150.

Sanabria HL, García MA, Diaz HA and Muñoz JE (2005). Morphological characterization of guava trees native in the Valle of Cauca. *Acta Agron.* 54: 1-6.

Santos CAF, Castro JMC, Souza FF and Vilarinho AA (2008). Preliminary characterization of *Psidium* germplasm in different Brazilian ecogeographic regions. *Pesq. Agropecu. Bras.* 43: 437-440.

Singh G (2005). High density planting in guava application of canopy architecture. *ICAR News* 11: 9-10.

Singh R and Sehgal OP (1968). Studies on the blossom biology of *Psidium guajava* L. (guava); 2, Pollen studies stigmatal receptivity pollination and fruit set. *Indian J. Hort.* 25: 52-59.

Sobral M, Proença C, Souza M and Mazine F (2012). Myrtaceae in Lista de Espécies da Flora do Brasil. Jardim Botânico do Rio de Janeiro. Available at [http://floradobrasil.jbrj.gov.br/2011/FB000171]. Accessed August 27, 2011.

Souza Júnior EE, Duarte JB and Chaves LJ (2002). Estabilidade fenotípica em goiabeira (*Psidium guajava* L.) com ênfase em peso de fruto, precocidade e período de colheita. *Pesq. Agropec. Trop.* 32: 97-103.

Teyer SLF, Moralez BA, Keb L and Barredo F (2010). Assessment of genetic diversity of Mexican guava germplasm using DNA molecular markers. *Acta Horticult.* 849: 133-138.

UPOV. (International Union for the Protection of new Varieties of Plants) (1987). Guidelines For the Conduct of Tests for Distinctness, Homogeneity and Stability in Guava (*Psidium guajava* L.). 1-27.

Urdaneta SAB, Colmenares C, Bracho B and Ortega J (2007). Morphological characterization of guava fruit (*Psidium guajava* L.) in a farm at Mara municipality, Zulia state. *Rev. Fac. Agron.* 24: 282-302.

Valdés-Infante J, Rodríguez NN, Becker D and Velázquez B (2007). Microsatellite characterization of guava (*Psidium guajava* L.) germoplasma collection in Cuba. *Cult. Trop.* 28: 61-67.