Online journal ISSN - 1676-5680
GMR Genetics and Molecular Research
EVOLUTION AND TECHNOLOGY

Comparison of coefficients and distance measurements in passion fruit plants based on molecular markers and physicochemical descriptors

C.B.M. Cerqueira-Silva[1], C.B. Cardoso-Silva[2], L.D.H.C.S. Conceição[3], J.V.A. Nonato[2], A.C. Oliveira[2] and R.X. Corrêa[3]

[1]Departamento de Biologia Vegetal, Universidade Estadual de Campinas, Campinas, SP, Brasil
[2]Departamento de Ciências Naturais, Universidade Estadual do Sudoeste da Bahia, Vitória da Conquista, BA, Brasil
[3]Departamento de Ciências Biológicas, Universidade Estadual de Santa Cruz, Ilhéus, BA, Brasil

Corresponding author: C.B.M. Cerqueira-Silva
E-mail: cerqueirasilva1@yahoo.com.br

Genet. Mol. Res. 8 (3): 870-879 (2009)
Received March 12, 2009
Accepted June 2, 2009
Published July 28, 2009

ABSTRACT. We investigated seven distance measures and 14 similarity coefficients in a set of observations of variables of the ‘yellow’ passion fruit plant (*Passiflora edulis* Sims), submitted to multivariate analyses (distance, projection and grouping). Fourteen genotypes were characterized, based on DNA amplification with 16 random amplified polymorphic DNA primers and the assessment of nine fruit physical-chemical descriptors. The distance measurements and the similarity coefficients were compared by the Spearman correlation test, projection in two-dimensional space and grouping efficiency, using five grouping methods; the genotype ranking varied with the different techniques. Coler-Rodger distance measures, Euclidean distance square measures and Yule similarity coefficients proved to be inadequate for projection in two-dimensional space or for grouping matrices. Regardless of the origin of the distance matrix, the unweighted pair group method with arithmetic mean grouping method proved to be the most adequate. The various distance measurements, similarity coefficients and grouping methods gave different values of distortion, cophenetic correlation and stress; they influence the characterization of genetic variability and this should be taken into account for studies of yellow passion fruit plants.

Key words: Grouping analysis; Multivariate statistics;
Genetic divergence; *Passiflora edulis* Sims

INTRODUCTION

In genetic improvement programs, the detection and quantification of available intra- and interspecific variability is of fundamental importance. It allows more efficient use of genetic resources by the breeder and optimizes prioritization in the conservation of specimens and access to the species (Dias, 1998; Benin et al., 2003; Emygdio et al., 2003). To estimate genetic variability between genotypes, the investigation of DNA polymorphism using molecular markers (Duarte et al., 1999) and agronomic descriptors (Gonçalves et al., 2008) are commonly employed. Also, various uni- and multivariate statistical methods are available to estimate dissimilarity between genotypes, while similarity coefficients are used for the assessment of qualitative data, as are distance measurements, based on quantitative data (Dias, 1998; Mohammadi and Prasanna, 2003).

As there are numerous methodologies available for estimating genetic dissimilarity, attention must be paid to choosing the method to be used, since this choice may influence considerably the estimates made (Jackson et al., 1989; Duarte et al., 1999). The determination of the method to estimate similarity (and its complement; dissimilarity), depends on the objectives proposed in the research, the data characteristics and the properties inherent to distances and coefficients (Cruz, 1990; Dias, 1998; Benin et al., 2003).

Comparisons made between distance measurements, with a preliminary assessment of agronomic descriptors as their basis, between oat (Benin et al., 2003) and tomato (Gonçalves et al., 2008) genotypes, as well as comparisons between similarity coefficients based on molecular markers, involving the genotypes of beans (Emygdio et al., 2003), sesame (Arriel et al., 2006) and corn (Meyer et al., 2004; Balestre et al., 2008), are all recent examples of how the efficiency of distances and coefficients may vary according to the nature of the variables measured. However, in genetic variability studies, the reason or criterion used to choose the similarity coefficient and the distance measure employed is not commonly reported; different coefficients and measures may be employed for the same purpose or a particular coefficient is adopted for different approaches (Emygdio et al., 2003).

In studies of passion fruit plants (Passifloraceae: *Passiflora*), various distance measures and similarity coefficients have been used to characterize intra- and interspecific genetic variability. The following are examples of these studies i) using agronomic descriptors and multivariate statistics, such as estimates of the Euclidean distance (Crochemore et al., 2003a) and the Mahalanobis generalized distance (Cardoso-Silva et al., 2007; Araújo et al., 2008); ii) using molecular markers and distance measurements (Ganga et al., 2004), Jaccard (Aukar et al., 2002;

Crochemore et al., 2003b; Viana et al., 2003) and Nei and Li (Bellon et al., 2007) coefficients.

Similar to what occurs with estimates of dissimilarity, various grouping strategies are available and adopted in research on *Passiflora*: Tocher (Araújo et al., 2008), Ward (Viana et al., 2003), neighbor-joining (Ganga et al., 2004), and unweighted pair group method with arithmetic mean (Aukar et al., 2002; Crochemore et al., 2003a,b; Bellon et al., 2007). However, these publications on *Passiflora* genus do not give a description of the criteria that were used to choose methods for the estimation of distance and genotype grouping.

Considering i) the use of different approaches regarding method and statistical analysis in the characterization of diversity in passion fruit plants, ii) the influence that these methods may exert on the results obtained, and iii) the absence of studies related to the efficiency of these methods for the passion fruit culture, we evaluated the influence of similar (and dissimilar) coefficients and of distance measurements in the characterization of the diversity of ‘yellow’ passion fruit genotypes, with the use of random amplified polymorphic DNA (RAPD) markers and of fruit physicochemical descriptors.

MATERIAL AND METHODS

The ‘yellow’ passion fruit genotypes (*Passiflora edulis* Sims) used in this study belong to the Active Collection of *Passiflora* Work Germplasm ‘Planalto de Conquista’ from Universidade Estadual do Sudoeste da Bahia (CAGT-*Passiflora* ‘Planalto de Conquista’/UESB), Vitória da Conquista campus, Bahia (14°53’20”S and 40°47’54”O, altitude 900 m; with an average annual precipitation of 700-800 mm, concentrated between November and March, an average annual temperature of 20-22°C) (Instituto Nacional de Meteorologia/Ministério da Agricultura e Abastecimento). This collection has been characterized for the reaction to Cowpea aphid-borne mosaic virus (Cerqueira-Silva et al., 2008), as well as how much is the prolificity.

To estimate the distance measurements, a total of 14 ‘yellow’ passion fruit genotypes were analyzed between the months of May and July 2008, using nine fruit physicochemical descriptors (fruit weight, equatorial diameter, longitudinal diameter, weight of the pulp with seeds, weight of the skin, thickness of the skin, pH, total soluble solids in °Brix, and total titratable acidity), with the use of a digital scale (precision of 0.01 g), Starrett® 727 digital pachymeter, bench Quimis® pH meter and manual refractometer with direct reading and corrected to 20°C (Instrutherm RT-30ATC).

To estimate the similarities and genetic distance coefficients, a total of 20 ‘yellow’ passion fruit plants were genotyped through the access and identification of DNA polymorphism with the use of 16 RAPD primers (OPD-03, -05, -07, -11, -13, -18, and -20; OPE-01, -02, -03, -04, -05, -07, -14, -16, and -17), preliminarily selected among the 40 primers that comprise the Operon© OPD and OPE series, since they result in the best standards of amplification of molecular markers.

Genomic DNA used in the polymerase chain reactions was extracted according to the method described by Doyle and Doyle (1990) and the amplification reactions performed according to Williams et al. (1990). The amplification products were then mixed with bromophenol blue and loaded on an agarose gel (1.6%) with ethidium bromide, submerged in Tris-borate and EDTA buffer, and electrophoretically resolved for 4 h at 90 V. Finally, the gels were exposed to ultraviolet light, photodocumented (Kodak EDAS

290) and assessed for the construction of a binary data matrix (0 for absence and 1 for the presence of bands).

To estimate the distances, based on results derived from physicochemical descriptors, the following strategies for the calculation of measurements were employed: Coler-Rodgers distance (C-DR); Euclidean distance (ED); average Euclidean distance (AED); Gower distance (GD); Mahalanobis' generalized distance (MGD); weighted distance by the square of the residue; Euclidean distance squared (EDS). To estimate the genetic distances (gd_{ij}), with RAPD binary data, the Baroni (B), Hamman (H), Phi, Kulczynski, Jaccard (J), Ochiai, Ochiai2, Roger and Tamino (RT), Russel and Rao (RR), simple matching (SM), Sorense-Dice (SD), Sokal and Sneath (SS), Yule (Y) similarity (gs_{ij}), and the Solkal dissimilarity (gd_{ij}) coefficients were obtained. The similarities obtained (sg_{ij}) were transformed into genetic distance by the following equation: $gd_{ij} = 1 - gs_{ij}$, meeting the presuppositions for transformation, that is, the genetic distance matrices were defined as non-negative, according to Johnson and Wichern (1998).

Alterations in genotype ranking, obtained by the distance measurements and similarity (and dissimilarity) coefficients, were analyzed by the Spearman (r_s) correlation. The efficacy of the measurements and coefficients, with regard to the different grouping methods (closest neighbor, farthest neighbor, Ward, Gower, and unweighted pair group method with arithmetic mean - UPGMA), was estimated on the basis of the original and simplified dissimilarity matrices; the last resulting from the use of one of the grouping methods with the following parameters: distortion values (D), cophenetic correlation coefficient (r_c) and stress (S).

Alterations in the efficiency of data projection in two-dimensional space, due to the choice of different measures and coefficients, were also assessed. For this, the D values, correlation coefficients (r) and S were calculated based on the original and graphic distances (adjusted for two-dimensional space).

The similarity and genetic distance analyses, as well as the estimates of D, r_c, S, and projection efficiency in two-dimensional space, were calculated using the Genes program (Cruz, 2001). Additionally, the BioEstat 5.0 program was adopted for Spearman correlation analyses (Ayres et al., 2005).

RESULTS AND DISCUSSION

The Spearman correlation coefficients between the seven distance measurements were, except for the MGD ($r_s \leq 0.65$), elevated and significant ($r_s \geq 0.80$; $P < 0.001$), indicating that the calculated distances are highly correlated and show few alterations in genotype ranking (Table 1). The $r_s = 1$ ($P < 0.001$) observed between ED, AED, and EDS stands out, allowing us to infer that these three measurements show the same distance ranking between the genotypes and have differences with regard to the ranking obtained by the GD and MGD. A low correlation value between distance measures ($r_s = 0.53$; between ED and MGD) was also found by Benin et al. (2003) when assessing oat genotypes using agronomic descriptors. This low correlation may be associated with the fact that the Euclidean distances assess the phenotype variation, based on the average of the characteristics, while the Mahalanobis distance assesses the genetic variation, taking repetitions and their deviations as a basis (Cruz, 1990; Benin et al., 2003).

Table 1. Spearman correlation coefficients[1] between seven distance measurements of nine fruit physicochemical descriptor variables measured in 14 ‘yellow’ passion fruit genotypes.

Distance measurement	C-DR	ED	AED	GD	MGD	WDSR	EDS
Coler-Rodgers distance (C-DR)	-						
Euclidean distance (ED)	0.99	-					
Average Euclidean distance (AED)	0.99	1.00	-				
Gower distance (GD)	0.97	0.98	0.98	-			
Mahalanobis generalized distance (MGD)	0.65	0.58	0.58	0.59	-		
Weighted distance by the square of the residue (WDSR)	0.86	0.81	0.81	0.81	0.58	-	
Euclidean distance squared (EDS)	0.99	1.00	1.00	0.98	0.58	0.81	-

[1]All correlation coefficients in the table are highly significant ($P < 0.001$).

In regard to similarity (and dissimilarity) coefficients, elevated and significant levels of the Spearman correlation ($r_s \geq 0.66$; $P < 0.001$) (Table 2) were also observed, except for the distance obtained by the RR coefficient’s complement ($0.31 \leq r_s \leq 0.80$). Among others, the maximum correlations ($r_s = 1$; $P < 0.001$) observed between coefficients B, H, RT, S, SM, and SS stand out, allowing us to infer that distance rankings between the genotypes obtained by these six coefficients are identical. Other coefficients, such as J and SD, also showed $r_s = 1$; however, they were different from the others because they had r_s values < 1. High Spearman correlation values between coefficients were also observed in other crops, such as with maize, genotyped using dominant (RAPD) (Meyer et al., 2004) and co-dominant (microsatellites - simple sequence repeats) (Balestre et al., 2008) markers, and beans, genotyped by RAPD markers (Corrêa et al., 1999). However, the correlation coefficient values were, at least in part, different between these works and in regard to the results obtained from this study. We found it impossible to generalize results obtained in other plant species with regard to the passion fruit plant culture.

Table 2. Spearman correlation coefficient[1] between 14 similarity coefficients related to DNA polymorphism identified based on DNA amplifications with 16 RAPD primers, detected in 20 ‘yellow’ passion fruit plants.

Coefficient	B	H	P	K	J	O	OII	RT	RR	S	SM	SD	SS	Y
Baroni (B)	-													
Hamman (H)	1.00	-												
Phi (P)	0.91	0.91	-											
Kulczynski (K)	0.98	0.98	0.83	-										
Jaccard (J)	0.90	0.98	0.82	1.00	-									
Ochiai (O)	0.98	0.98	0.82	1.00	1.00	-								
Ochiai2 (OII)	0.89	0.89	1.00	0.79	0.79	0.79	-							
Roger and Tamino (RT)	1.00	1.00	0.91	0.98	0.98	0.98	0.89	-						
Russel and Rao (RR)	0.68	0.68	0.36	0.79	0.80	0.79	0.31	0.68	-					
Sokal (S)	1.00	1.00	0.91	0.98	0.98	0.98	0.89	1.00	0.68	-				
Simple matching (SM)	1.00	1.00	0.91	0.98	0.98	0.98	0.89	1.00	0.68	1.00	-			
Sorensen-Dice (SD)	0.98	0.98	0.82	1.00	1.00	1.00	0.79	0.98	0.80	0.98	0.98	-		
Sokal and Sneath (SS)	1.00	1.00	0.91	0.98	0.98	0.98	0.89	1.00	0.68	1.00	1.00	0.98	-	
Yule (Y)	0.93	0.93	1.00	0.86	0.85	0.86	0.99	0.93	0.41	0.93	0.93	0.85	0.93	-

[1]All correlation coefficients in the table are highly significant ($P < 0.001$).

With regard to the projection efficiency in two-dimensional space, the distance measurements showed great variation, with stress values oscillating between 22.6 and 49.9% (Table 3). The highest stress values were observed based on the C-DR (41.3) and the EDS (49.9) distances, while the lowest stress values were obtained based on the GD (22.7) and the MGD (22.6). Because stress values higher than 40% were obtained, it is possible to classify the C-DR and EDS measures as “inadequate”, according to Kruskal’s classification (1964).

Table 3. Efficacy of the projection of similarity (and dissimilarity) measurements and coefficients in two-dimensional space, in ‘yellow’ passion fruit plants, based on distortion percentage, the correlation between the original and the projected (r) distance and stress values.

Measurements and coefficients[1]	Distortion (%)	Correlation (r)	Stress (%)
Coler-Rodgers distance (C-DR)	26.44	0.88	41.3
Euclidean distance (ED)	20.7	0.91	26.3
Average Euclidean distance (AED)	20.7	0.91	26.3
Gower distance (GD)	15.5	0.92	22.7
Mahalanobis generalized distance (MGD)	14.4	0.94	22.6
Weighted distance by the square of the residue (WDSR)	18.5	0.87	32.9
Euclidean distance squared (EDS)	28.2	0.83	49.9
Baroni (B)	51.3	0.32	57.8
Hamman (H)	52.9	0.28	58.9
Phi (P)	50.3	0.29	57.6
Kulczynski (K)	53.8	0.29	60.3
Jaccard (J)	58.5	0.25	63.7
Ochiai (O)	53.7	0.29	59.9
Ochiai2 (OII)	55.3	0.27	61.2
Roger and Tamino (RT)	59.7	0.23	64.7
Russel and Rao (RR)	69.9	0.44	73.8
Sokal (S)	59.2	0.13	71.9
Simple matching (SM)	52.9	0.28	58.9
Sorensen-Dice (SD)	54.1	0.28	60.1
Sokal and Sneath (SS)	47.5	0.30	55.0
Yule (Y)	4.7	0.21	74.3

[1] The seven first lines of this column concern the distance measurements, and the other 14 lines are in regard to similarity (and dissimilarity) coefficients.

For distances estimated based on similarity (and dissimilarity) coefficients, the projection efficiencies in two-dimensional space displayed elevated stress values between 55 and 74.3% (Table 3). The highest stress values were observed for RT (64.7%), S (71.9%), RR (73.8%), and Y (74.3%). These results concur, at least in part, with those obtained by Meyer et al. (2004) and Balestre et al. (2008), who found the highest stress values for RT and RR coefficients with regard to the other coefficients considered in these studies. The results obtained allowed the classification, according to Kruskal (1964), of the coefficients determined as being inadequate for assessments in two-dimensional space, based on the matrix of binary data from molecular markers of the dominant type, as with RAPD, in passion fruit plants.

The different combinations between the distance measurements and the five grouping methods gave distinct results concerning the efficacy of the grouping matrix in presenting the original distance matrix ($-2311.0 \leq D \leq 93.2$; $0.33 \leq r_c \leq 0.75$; $22.3 \leq S \leq 395.7$) (Table 4). UPGMA was the most efficient among the groupings assessed, showing all distance measurements, distortion and stress values as being closer to zero ($5 \leq D \leq 18.9$; $22.3 \leq S \leq 43.5$) and the highest cophenetic correlation values ($0.65 \leq r_c \leq 0.75$). Thus, the Ward method showed, based on all the distance measurements, the lowest efficiency in the genotype grouping ($-2311.0 \leq D \leq -1477.0$; $0.57 \leq r_c \leq 0.70$; $395.7 \leq S \leq 312.7$). These results indicate that UPGMA has the highest/best efficiency as a grouping method, in regard to an assessment of ‘yellow’ passion fruit plant quantitative variables, as well as the inefficiency of the Ward grouping method for this purpose. Results similar to these were also found by Gonçalves et al. (2008), while assessing the genetic distance of tomato genotypes with agronomic descriptors and by Arriel et al. (2006), while assessing sesame genotypes, with DNA amplifications using RAPD markers.

Table 4. Efficacy of five grouping methods (closest neighbor - CN; farthest neighbor - FN; Ward - W; Gower - WPGMC, and unweighted pair group method with arithmetic mean - UPGMA) from different similarity (and dissimilarity) measures and coefficients, based on criteria of distortion percentage (D), cophenetic correlation (r_c) and stress percentage values (S).

Measurements and coefficients [1]	CN			FN			W			WPGMC			UPGMA		
	D	r_c	S	D	r_c	S	D	r_c	S	D	r_c	S	D	r_c	S
Coler-Rodgers distance (C-DR)	83.2	0.38	64.7	-155.9	0.57	76.1	-2059.1	0.57	373.9	68.6	0.61	51.7	11.0	0.65	33.2
Euclidean distance (ED)	59.9	0.67	42.2	-156.7	0.63	71.2	-2311.0	0.70	395.7	71.0	0.67	49.8	5.0	0.75	22.3
Average Euclidean distance (AED)	59.9	0.67	42.2	-156.7	0.63	71.2	-2311.0	0.70	395.7	71.0	0.67	49.8	5.0	0.75	22.3
Gower distance (GD)	66.1	0.64	47.7	-163.4	0.61	75.2	-2277.6	0.62	384.2	64.1	0.66	45.6	6.8	0.71	26.1
Mahalanobis' generalized distance (MGD)	85.7	0.64	67.8	-221.6	0.65	98.4	-1529.4	0.60	318.1	51.3	0.64	51.8	13.6	0.75	36.1
Weighted distance by the square of the residue (WDSR)	93.2	0.33	79.4	-180.5	0.71	85.2	-1556.4	0.71	317.8	52.5	0.72	45.2	14.7	0.73	38.4
Euclidean distance squared (EDS)	88.4	0.56	72.3	-476.6	0.51	164.4	-1477.0	0.60	312.9	65.4	0.67	53.6	18.9	0.67	43.5
Baroni (B)	39.3	0.19	26.2	-71.4	0.44	35.9	-6711.7	0.46	727.9	64.3	0.13	42.5	1.9	0.48	14.0
Hamman (H)	37.9	0.19	25.0	-64.7	0.45	33.1	-6754.3	0.46	730.4	64.7	0.11	43.1	1.7	0.49	13.2
Phi (P)	40.2	0.39	26.1	-65.5	0.45	34.2	-6536.3	0.49	717.3	62.5	0.25	40.9	1.76	0.55	13.3
Kulczynski (K)	39.2	0.21	25.8	-64.7	0.34	34.3	-6746.7	0.44	730.1	61.4	0.12	41.1	1.9	0.45	13.8
Jaccard (J)	0.34	0.23	22.2	-52.3	0.34	28.3	-6874.4	0.45	737.5	61.8	0.09	40.2	1.35	0.47	11.6
Ochiai (O)	39.3	0.21	25.9	-65.2	0.34	34.5	-6747.0	0.44	730.1	61.5	0.12	41.2	1.9	0.45	13.9
Ochiai2 (OII)	35.1	0.44	22.4	-51.8	0.48	27.4	-6865.1	0.46	737.2	57.6	0.35	36.7	1.2	0.57	11.2
Roger and Tamino (RT)	30.9	0.22	19.9	-46.2	0.46	24.5	-6920.5	0.47	740.1	61.8	0.12	39.6	1.0	0.51	10.2
Russel and Rao (RR)	19.4	0.54	19.4	-27.2	0.51	15.4	-6578.2	0.41	719.7	53.6	0.54	32.5	0.39	0.67	6.2
Sokal (S)	25.1	0.24	15.5	-29.4	0.44	16.6	-6976.4	0.44	743.5	58.5	0.18	36.2	0.51	0.54	7.1
Simple matching (SM)	37.8	0.19	25.0	-64.7	0.45	33.1	-6754.2	0.46	730.4	64.9	0.11	43.1	1.7	0.49	13.2
Sorensen-Dice (SD)	39.4	0.23	25.9	-66.3	0.33	34.8	-6722.7	0.44	728.7	65.4	0.04	44.0	1.9	0.46	13.8
Sokal and Sneath (SS)	42.7	0.18	28.7	-80.9	0.44	40.1	-6623.5	0.45	722.8	66.8	0.05	45.2	2.4	0.47	15.5
Yule (Y)	79.8	0.26	61.3	-278.3	0.39	110.3	-4942.4	0.44	618.8	70.7	0.32	54.3	12.5	0.48	35.4

[1] The seven first lines of this column concern the distance measurements, and the other 14 lines are in regard to similarity (and dissimilarity) coefficients.

The grouping methods, similar to circumstances surrounding the distance measurements, showed different results regarding the efficacy of the grouping matrices in representing the distance matrices obtained by the adopted coefficients ($-6976.4 \leq D \leq 0.39$; $0.04 \leq r_c \leq 0.67$; $6.2 \leq S \leq 743.5$) (Table 4). UPGMA was again identified as being the most efficient among the groupings assessed, yielding for all similarity coefficients, distortion and stress values closer to zero ($0.39 \leq D \leq 12.5$; $6.2 \leq S \leq 35.4$) and showing the highest cophenetic correlation values ($0.45 \leq r_c \leq 0.67$). The Ward method also demonstrated, for the similarity coefficients studied, the lowest efficiency in the genotype grouping ($6976.4 \leq D \leq -4942.4$; $0.41 \leq r_c \leq 0.49$; $618.8 \leq S \leq 743.5$).

In preliminary genetic diversity studies involving passion fruit plants, whose distance was estimated using agronomic variables or molecular markers of the dominant type, the grouping method UPGMA is becoming the most adopted one, such as in the study of Aukar et al. (2002), Crochemore et al. (2003a,b), and Bellon et al. (2007). Other methods such as the neighbor-joining method used by Ganga et al. (2004) and the Ward method used by Viana et al. (2003) are also selected. The utilization of different grouping methods for the same purpose, without indicating the criterion of choice, may at least make it difficult to compare results obtained by different studies, since the results may be influenced by the method selected for the construction of the grouping matrix.

Grouping efficiency results obtained by UPGMA, based on the distance measurements assessed, showed little variation in regard to the cophenetic correlation coefficient ($0.65 \leq r_c \leq 0.75$) and great variation in the stress values ($11.9 \leq S \leq 43.9$) (Table 3). With the exception of stress values observed for the EDS and for C-DR (43.5 and 41.5, respectively), stress values for the other measurements were less than 38.5. The ED and AED that showed the lowest stress values (11.9) stand out. These results make it possible to classify the EDS and C-DR distance measurements as “inadequate” and ED and AED as “good”, according to the classification suggested by Kruskal (1964).

With regard to the similarity (and dissimilarity) coefficients, the efficiency results of groupings obtained with UPGMA were similar to those obtained for distance measurements, that is, a lower variation of the cophenetic correlation coefficient ($0.45 \leq r_c \leq 0.67$) and great variation in the stress values ($6.2 \leq S \leq 35.4$) (Table 4). For the stress values, the estimated distances showed, except for the Yule coefficient (stress = 35.4), values less than 15.4. The RR and S coefficients stand out as the lowest stress values (6.3 and 7.1, respectively). The results obtained allow us to classify, according to Kruskal (1964), the RR and S coefficients as “excellent” and the others as, at least, “regular”.

The different distance measurements and the similarity (and dissimilarity) coefficients influence the results of characterizations of the passion fruit genotype groupings, since they show a wide variation in their results. Therefore, in diversity characterizations of passion fruit plants, based on quantitative data and dominant molecular markers, the use of Coler-Rodger distance measurements, the Euclidean distance squared and the Yule coefficient must be avoided, since they show high stress values in the projection of data in two-dimensional space and in the construction of the grouping matrix. In the same way, the Roger and Tamino, and Russel and Rao similarity coefficients must also be avoided in characterizations of passion fruit genetic variability, since they have a contrasting efficiency with regard to the projection of data in two-dimensional space and in the generation of the cophenetic matrix (grouping matrix). One should be aware of the type of population studied, because as discussed by Corrêa et al. (1999) the efficiency of methods may be influenced by the level of heterozygosity associated with the population studied, since the number of heterozygous loci is usually different between natural and improved populations as well as autogamous and allogamous species.

With regard to the grouping of passion fruit genotypes, UPGMA may be adopted, based on different strategies for the estimation of distance in passion fruit plants, considering that it shows better values of distortion, cophenetic correlation and stress, compared to the other grouping methods tested.

ACKNOWLEDGMENTS

Elisa Susilene Lisboa dos Santos contributed to this study during the writing of this article, and Alan Silva Pereira, graduate student, also contributed during laboratory phases. Research supported by CNPq and Programa de Pós-Graduação em Genética e Biologia Molecular, UESC, for having allowed the development of this research, which is part of the Master's dissertation of C.B.M. Cerqueira-Silva.

REFERENCES

Araújo FP, Silva N and Queiroz MA (2008). Genetic divergence among *Passiflora cincinnata* Mast accessions based on morphoagronomic descriptors. [Divergência genética entre acessos de *Passiflora cincinnata* Mast com base em descritores morfoagronômicos]. *Rev. Bras. Frutic.* 30: 723-730.

Arriel NHC, Di Mauro AOD, Di Mauro SMZ, Bakke OA, et al. (2006). Multivariate techniques for the determination of genetic diversity in sesame using RAPD markers. [Técnicas multivariadas na determinação da diversidade genética em gergelim usando marcadores RAPD]. *Pesq. Agropec. Bras.* 41: 801-809.

Aukar APA, Lemos EGM and Oliveira JC (2002). Genetic variations among passion fruit species using RAPD markers. [Variação genética entre espécies de maracujá utilizando marcadores RAPD]. *Rev. Bras. Frutic.* 24: 738-740.

Ayres M, Ayres Junior M, Ayres DL and Santos AS (2005). Programa BioEstat 5.0. Aplicações Estatísticas nas Áreas das Ciências Biológicas e Biomédicas. Sociedade Civil Mamirauá, Belém.

Balestre M, Von Pinho RG, Souza JC and Lima JL (2008). Comparison of maize similarity and dissimilarity genetic coefficients based on microsatellite markers. *Genet. Mol. Res.* 7: 695-705.

Bellon G, Faleiro FG, Junqueira KP, Junqueira NTV, et al. (2007). Genetic variability of wild and commercial passion fruit (*Passiflora edulis* Sims.) accessions using RAPD markers. [Variabilidade genética de acessos silvestres e comerciais de *Passiflora edulis* Sims. com base em marcadores RAPD]. *Rev. Bras. Frutic.* 29: 124-127.

Benin G, Carvalho FIF, Oliveira AC, Marchioro VS, et al. (2003). Comparisons among dissimilarity measures and multivariate statistics as criterions for directing hibridizations in oat. [Comparações entre medidas de dissimilaridade e estatísticas multivariadas como critérios no direcionamento de hibridações em aveia]. *Cienc. Rural* 33: 657-662.

Cardoso-Silva CB, Melo JRF, Pereira AS, Cerqueira-Silva CBM, et al. (2007). Study of the genetic diversity of sleep-passion fruit native of Bahia State, Brazil. [Estudo da diversidade genética mediante caracterização físico-química de frutos de maracujazeiros-do-sono nativos]. *Magistra* 19: 352-358.

Cerqueira-Silva CB, Moreira CN, Figueira AR, Correa RX, et al. (2008). Detection of a resistance gradient to Passion fruit woodiness virus and selection of 'yellow' passion fruit plants under field conditions. *Genet. Mol. Res.* 7: 1209-1216.

Corrêa RX, Abdelnoor RV, Faleiro FG, Cruz CD, et al. (1999). Genetic distances in soybean based on RAPD markers. *Bragantia* 58: 15-22.

Crochemore ML, Molinari HB and Stenzel NMC (2003a). Agro-morphological characterization of passion fruit (*Passiflora* spp.) germplasm. [Caracterização agromorfológica do maracujazeiro (*Passiflora* spp.)]. *Rev. Bras. Frutic.* 25: 5-10.

Crochemore ML, Molinari HBC and Vieira LGE (2003b). Genetic diversity in passion fruit (*Passiflora* spp.) evaluated by RAPD markers. *Braz. Arch. Biol. Technol.* 46: 521-527.

Cruz CD (1990). Aplicação de Algumas Técnicas Multivariadas no Melhoramento de Plantas. Doctoral thesis, Genética e Melhoramento, Escola Superior de Agronomia Luiz de Queiroz, Universidade de São Paulo, Piracicaba.

Cruz CD (2001). Programa Genes: Aplicativo Computacional em Genética e Estatística. Editora Universidade Federal de Viçosa, Viçosa.

Dias LAS (1998). Análises Multidimensionais. In: Eletroforese de Isoenzimas e Proteínas Afins: Fundamentos e Aplicações em Plantas e Microrganismos (Alfenas AC, ed.). Editora da Universidade Federal de Viçosa, Viçosa, 405-476.

Doyle JJ and Doyle JL (1990). Isolation of plant DNA from fresh tissue. *Focus* 12: 13-15.

Duarte JM, dos Santos JB and Melo LC (1999). Comparison of similarity coefficients based on RAPD markers in the common bean. *Genet. Mol. Biol.* 22: 427-432.

Emygdio BM, Antunes IF, Choer E and Nedel JL (2003). Efficiency of similarity coefficients based on RAPD markers in common bean genotypes. [Eficiência de coeficientes de similaridade em genótipos de feijão mediante marcadores RAPD]. *Pesq. Agropec. Bras.* 38: 243-250.

Ganga RMD, Ruggiero C, Lemos EGM, Grili GVG, et al. (2004). Genetic diversity in yellow passion fruit utilizing fAFLP molecular markers. [Diversidade genética em maracujazeiro-amarelo utilizando marcadores moleculares fAFLP]. *Rev. Bras. Frutic.* 26: 494-498.

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.

Jackson AA, Somers KM and Harvey HH (1989). Similarity coefficients: measures of co-occurrence and association or simply measures of occurrence? *Am. Nat.* 133: 436-453.

Johnson RA and Wichern DW (1998). Applied Multivariate Statistical Analysis. Prentice-Hall, New Jersey.

Kruskal J (1964). Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika* 29: 1-27.

Meyer AS, Garcia AAF, Souza AP and Souza CL Jr (2004). Comparison of similarity coefficients used for cluster analysis with dominant markers in maize (*Zea mays* L). *Genet. Mol. Biol.* 27: 83-91.

Mohammadi SA and Prasanna BM (2003). Analysis of genetic diversity in crop plants - salient statistical tools and considerations. *Crop Sci.* 43: 1235-1248.

Viana AP, Pereira TNS, Pereira MG, Souza MM, et al. (2003). Genetic diversity among yellow passion fruit commercial genotypes and among passiflora species using RAPD. *Rev. Bras. Frutic.* 25: 489-493.

Williams JG, Kubelik AR, Livak KJ, Rafalski JA, et al. (1990). DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. *Nucleic Acids Res.* 18: 6531-6535.