



# Heirloom tomato gene bank: assessing genetic divergence based on morphological, agronomic and molecular data using a Ward-modified location model

L.S.A. Gonçalves<sup>1</sup>, R. Rodrigues<sup>1</sup>, A.T. do Amaral Júnior<sup>1</sup>, M. Karasawa<sup>2</sup> and C.P. Sudré<sup>1</sup>

<sup>1</sup>Universidade Estadual do Norte Fluminense Darcy Ribeiro, Campos dos Goytacazes, RJ, Brasil

<sup>2</sup>Instituto Agronômico de Pernambuco, Belém de São Francisco, PE, Brasil

Corresponding author: L.S.A. Gonçalves  
E-mail: lsagrural@yahoo.com.br

Genet. Mol. Res. 8 (1): 364-374 (2009)

Received October 22, 2008

Accepted February 12, 2009

Published March 31, 2009

**ABSTRACT.** Accessions in gene banks need to be characterized and evaluated to determine their genetic diversity. We made a joint diversity analysis of the tomato gene bank of the Universidade Estadual do Norte Fluminense Darcy Ribeiro in Rio de Janeiro State, using the Ward-modified location model. Forty *Solanum lycopersicum* accessions were characterized and evaluated for 22 morphoagronomic descriptors and 131 random amplified polymorphic DNA markers. Based on the pseudo-F and pseudo-t<sup>2</sup> criteria, the optimal number of groups was established as five. Variability within groups was high for

both continuous and discrete nominal data. The first two canonical variables explained about 90% of the inter-group variability. Care should be taken in using the Ward-modified location model technique to avoid incorporating excessive and unnecessary markers, which could favor molecular markers when compared with morphoagronomic variables. However, the minimum number of markers is germplasm-dependent and must be recalculated for each new divergence analysis.

**Key words:** *Solanum lycopersicum*; Joint analysis; Ward-MLM; Random amplified polymorphic DNA markers

## INTRODUCTION

The tomato, *Solanum lycopersicum* (L.), is an autogamous species with a narrow genetic base (Rick, 1976; Saavedra et al., 2001; Barrero and Tanksley, 2004; Bai and Lindhout, 2007; Foolad, 2007; Gonçalves et al., 2008a). The introduction of the species in Europe, from Mexico, was crucial in the reduction of genetic variability, since in the European habitat tomatoes were generally cultivated in protected environments. This protected the wild forms, then allogamous, from the action of wind and insect pollinators, culminating in the maintenance of a germplasm adapted to autogamy only (Foolad, 2007).

The high degree of genetic uniformity in tomato cultivars is not only strongly influenced by domestication far from the center of origin, but above all by genetic improvement which, *per se*, culminated in the achievement of uniform forms, apart from the fact that only a limited number of genotypes were used for breeding (Saavedra et al., 2001). Moreover, due to the replacement or disappearance of wild and local species, countless genomic forms with genes that could be of high interest for future actions of breeders were lost due to pests and diseases (Bai and Lindhout, 2007).

The need for the preservation of wild species, local varieties and traditional genotypes in gene banks is evident, which have become an important form of gene maintenance (Gepts, 2006). However, the accessions in gene banks should be characterized and evaluated in order to determine genetic diversity, which would allow the identification of redundant accessions and genotypes of interest in breeding programs (Balestre et al., 2008; Gonçalves et al., 2008b; Terzopoulos and Bebel, 2008).

The classification of individuals and quantification of genetic diversity in gene banks is usually aimed at the identification of similarity groups based on separate analyses of continuous (e.g., plant height, fruit weight, days to flowering) and discrete variables (such as fruit color and shape, the presence or absence of a trait, or a molecular marker) (Mohammadi and Prasanna, 2003; Crossa and Franco, 2004; Sudré et al., 2007). The joint analysis of these variables is, however, becoming an interesting alternative for both breeders and gene bank curators for a better quantification of genetic variability. The reason is that the joint analysis of variables of different types (continuous and discrete, the latter nominal or binary-type) can deepen and complete the knowledge about a set of accessions.

In order to analyze all variables jointly, Olkin and Tate (1961) proposed the location model (LM) to classify  $n$  individuals when  $p$  continuous variables and  $q$  discrete variables are measured in an environment. LM combines the levels of all  $q$  discrete variables in a single multinomial variable,  $W$ , with  $m$  levels ( $w = 1, 2, \dots, m$ ). Franco et al. (1998) modified the LM and proposed the modified location model (MLM), assuming that  $m$  levels of variable  $W$  and  $p$ -multinomial variables for each sub-population are independent. According to Franco et al. (1998), the MLM strategy consists of two stages. In the first, the groups are defined by Ward's cluster method using the dissimilarity matrix of Gower (1971). In the second step, the mean of the vector of continuous variables for each sub-population independent of the  $W$  values is estimated by the MLM procedure.

The statistical method Ward-MLM has advantages such as the optimization of two objective functions related in two stages of the process, which are the sum of squares within groups in the first stage and the likelihood function of observations in the second. This method also allows the definition of the optimal number of groups and the calculation of a highly accurate measure of the groups, since the likelihood of each accession to be allocated to a specific group can be determined by this technique. This statistical approach exploits all available information on the gene bank accessions, whether the variables are continuous or discrete (Crossa and Franco, 2004; Ortiz et al., 2008).

Among the various tomato gene banks in Brazil, the Universidade Estadual do Norte Fluminense Darcy Ribeiro (UENF) has a collection of accessions considered heirloom seeds, maintained by public institutions in Rio de Janeiro for over 40 years (Jordan, 2007). These represent an important source of genetic resources for breeding and are being characterized based on morphological descriptors and agronomic and random amplified polymorphic DNA (RAPD) markers (Karasawa et al., 2005; Gonçalves et al., 2008a). The purpose of this study was to conduct a joint analysis of available data on the UENF tomato gene bank using the Ward-MLM strategy to determine the diversity in the accessions.

## MATERIAL AND METHODS

Forty *Solanum lycopersicum* accessions of the UENF gene bank were characterized and evaluated by means of continuous (morphological and agronomic) and discrete variables (morphological and RAPD markers) (Karasawa et al., 2005; Gonçalves et al., 2008a).

The morphological and agronomic traits were evaluated in a field test conducted in Campos dos Goytacazes, Rio de Janeiro, Brazil, in a randomized complete block design with three replications and 16 plants per plot, based on 22 descriptors proposed by Biodiversity International (Table 1). For the molecular characterization, 300 mg leaves from 35-day-old tomato plants grown in a greenhouse was used. DNA extraction followed the protocol of Doyle and Doyle (1987) and the DNA was amplified as described by Williams et al. (1990). In the electrophoretic procedure, the agarose gels were stained with ethidium bromide and photographed under UV light using an Eagle Eye II (Stratagene). The following primers (Operon Technologies) were selected: OPPA 03, OPAA 04, OPAA 18, OPAB 05, OPAB 07, OPAB 09, OPAB 14, OPAC 06, OPAH 01, OPC 08, OPC 09, OPC 11, OPC 15, OPE 06, OPE 07, OPE 18, OPG 16, OPI 12, OPK 16, OPN 06, OPN 08, OPO 10, OPT 16, OPW 06, OPW 13, and OPV 12 (Gonçalves et al., 2008a).

**Table 1.** Descriptors used for characterization and evaluation of 40 tomato accessions from gene bank of the UENF, Brazil.

Traits	Description
Plant growth type	1-4 scoring (1 = Dwarf, 2 = Determinate, 3 = Semi-determinate, 4 = Indeterminate)
Foliage density	1-3 scoring (1 = Sparse, 2 = Intermediate, 3 = Dense)
Leaf type	1-3 scoring (1 = Dwarf, 2 = Potato leaf type, 3 = Standard)
Number of days to flowering	From sowing until 50% of plants have at least one open flower in a uniform growing environment
Corolla color	1-4 scoring (1 = White, 2 = Yellow, 3 = Orange)
Exterior color of immature fruit	1-5 scoring (1 = Greenish-white, 2 = Light green, 3 = Green, 4 = Dark green, 5 = Very dark green)
Exterior color of mature fruit	1-5 scoring (1 = Green, 2 = Yellow, 3 = Orange, 4 = Pink, 5 = Red)
Predominant fruit shape	1-8 scoring (1 = Oblate, 2 = Slightly flattened, 3 = Rounded, 4 = High rounded, 5 = Heart-shaped, 6 = Long oblong, 7 = Pear-shaped, 8 = Plum-shaped)
Presence of green stripes on the fruit	0 (Absent), 1 (Present)
Intensity of green shoulder	1-3 scoring (1 = Absent; 2 = Slight, 3 = Intermediate, 4 = Strong)
Radial cracking	1-4 scoring (1 = Corky lines, 2 = Slight, 3 = Intermediate, 4 = Severe)
Concentric cracking	1-4 scoring (1 = Corky lines, 2 = Slight, 3 = Intermediate, 4 = Severe)
Number of locules per plant	Counted on at least 10 fruits
Presence of open locules	0 (Absent), 1 (Present)
Total number of fruits per plant	Counting every fruit harvested in each plant
Total fruit weight (g)	Assessed in 10 fruits considering all plants
Mean fruit weight (g)	Total fruit weight and plant number ratio
Fruit length (mm)	Recorded from stem end to blossom end, to one decimal place, at maturity
Fruit width (mm)	Recorded at the largest diameter of cross-sectioned fruits to one decimal place at maturity
Number of days to maturity	From sowing until 50% of plants have at least one mature fruit
Number of flowers per inflorescence	Mean of 10 plants
Soluble solids	Measures in Brix units from two composite raw juice samples of at least five fruits per juice sample

In the statistical analysis, the significance of continuous variables was first evaluated by the F-test. The continuous and discrete variables, together, were used according to the Ward-MLM procedure (Franco et al., 1998) to compose the groups of accessions using the procedures CLUSTER and interactive matrix programming (IML) of the SAS program (SAS Institute, 1999). To use the grouping method of Ward, the distance matrix was calculated using the algorithm of Gower (1971), which analyzes continuous and discrete variables jointly. For the definition of the number of groups, the procedure indicated for the MLM model was considered, which is based on pseudo-F and pseudo- $t^2$  criteria. Based on the definition of the optimal number of groups, the hierarchical classification was obtained by the Ward method, which provides the initial values for the parameters needed to implement the final step of the MLM model (Franco et al., 1998; Crossa and Franco, 2004).

The diagram of differences between the groups and the correlation of traits with canonical variables were established using the procedure CANDISC of SAS (SAS Institute, 1999). The coefficient proposed by Matusita (1956), adapted by Krzanowski (1983) and later by Franco et al. (1998) for the distribution of mixed variables (continuous and discrete) was used to determine similarity and dissimilarity values among the groups formed. Considering the homogeneity of variance and co-variance matrices of all groups and independence between discrete and continuous variables, the similarity coefficient between two groups is defined by

$$p_{ij} = I_{ij} \sum_{s=1}^m (p_{is} p_{js})^{1/2}, \quad (\text{Equation 1})$$

where  $I_{ij} = \exp\{-D_{ij}^2/8\}$  is the affinity due to continuous variables;  $D_{ij}^2$  the distance of Mahalanobis;  $\sum_{s=1}^m (p_{is} p_{js})^{1/2}$  is the similarity coefficient for discrete variables, and  $p_{is}$  and  $p_{js}$  are, respectively, the proportion of plants in groups  $i$  and  $j$ , corresponding to the  $s^{\text{th}}$  value for the multinomial variable in relation to all multicategorical variables.

## RESULTS AND DISCUSSION

All continuous variables but content of soluble solids were significant at 5% probability by the F-test, demonstrating variability in the accessions studied for these descriptors. For the discrete nominal variables, the traits plant growth habit and corolla color did not differ among accessions, and all accessions have an indeterminate growth habit and yellow corolla. For discrete binary variables, which were the RAPD markers, 131 bands amplified by 26 primers were studied. Of these, 89 were polymorphic, and consequently, each primer generated a mean of 3.42 polymorphic bands.

By the criteria pseudo-F and pseudo- $t^2$ , the optimal number of groups was established at five (A, B, C, D, and E). Five was therefore determined as the optimal number of groups to express the diversity in the genotypes studied, as suggested by the MLM analysis. Padilha et al. (2005) investigated 120 populations of *Brassica rapa* subsp. *rapa* L. and found that the increase in the likelihood function was maximal when five groups were considered. Ortiz et al. (2008), however, investigated maize races of high altitudes in Peru and observed greatest increases in the probability function at a level of four and eight groups (increases of 56.22 and 50.60, respectively). The number of groups may therefore vary depending on the species, the number of accessions and the number and type of descriptors. The method can, however, define more precise criteria of group formation, resulting in less subjective groups of accessions.

The groups A, B, C, D, and E comprised 10, 14, 9, 4, and 3 accessions, respectively (Table 2). The relative frequencies for discrete variables in the five groups are shown in Table 3, and means and ranges of continuous variables in Table 4. Preliminary analyses indicated that the variability within groups for continuous as well as for discrete and nominal data was high. This can be detected when the range of variation for the morphoagronomic descriptors is examined (Tables 3 and 4). In terms of fruit shape (Table 3), wide variation was observed within the groups A, B, C, and D, while in group E only a slightly flattened fruit shape was recorded. Even for discrete nominal variables, wide variability within groups was also observed for the number of locules per fruit, with bi-, tri-, and multi-locule fruit observed within the groups A, B, C, and E. In terms of continuous variables, for mean fruit weight (Table 4), for example, the range of variation was greater in group D, since the difference between the highest (88.30) and lowest (44.67) values for the trait was close to 1.0, indicating high variability for the clustering of only four accessions. Similarly, high variability was observed for total soluble solids, since the range of variation weighted by the lower and upper limits was estimated at 1.7, 2.0, 1.7, 2.0, and 1.0 for groups A, B, C, D, and E, respectively.

**Table 2.** Identification of the accessions, popular name and group into which each of the 40 tomato accessions of the Universidade Estadual do Norte Fluminense (UENF) germplasm bank clustered.

No.	Accession	Popular name	Group
1	UENF 176	PI-280060	B
2	UENF 177	Bould Mountain	E
3	UENF 178	Early Pak	C
4	UENF 179	Red Jacket	A
5	UENF 180	VR superior	A
6	UENF 182	ES-58	A
7	UENF 185	WV-139-1-2-1-1-1	E
8	UENF 186	VR Wisconsin	B
9	UENF 187	PI-95588 CGS	B
10	UENF 188	Heinz 14451 VF	B
11	UENF 189	VF-14	A
12	UENF 190	PI-262910	A
13	UENF 193	Heinz 2439	B
14	UENF 194	Acano	C
15	UENF 195	Valians	B
16	UENF 196	Bongned	D
17	UENF 197	Sundwarf	D
18	UENF 199	UTAH 4	C
19	UENF 200	Improved Garden State	A
20	UENF 201	-	B
21	UENF 202	KC 46J2J2	B
22	UENF 204	Cornell 61-56	B
23	UENF 205	Porte	B
24	UENF 206	PI-255829 CGS	C
25	UENF 208	V641	A
26	UENF 209	Short Stem Boone	B
27	UENF 210	C-49-59	A
28	UENF 210B	C-49-59	A
29	UENF 211	Cornell 54-17	D
30	UENF 212	Roma	B
31	UENF 213	Persimmon Type	B
32	UENF 215	PI-255839	C
33	UENF 217	WV-289-1-4-1	A
34	UENF 218	Manelee	B
35	UENF 219	Red Top	C
36	UENF 221	WV rutgers	C
37	UENF 222	ACC high pigment line	C
38	UENF 223	PI-105342	C
39	UENF 224	Manzaha	D
40	UENF 225	VR pnitic hard	E

For the discrete binary variables, which corresponded to the 131 RAPD markers, 43, 40, 32, 29, and 37 polymorphic bands were observed in the groups A, B, C, D, and E, respectively.

High variability for continuous and discrete variables within groups was found, which was not expected, since the tomato germplasm has a narrow genetic base. The most likely explanation for this is that the continuous and discrete variables may not have been the main determinants in the clustering of accessions. The molecular markers, which were represented in higher number than the phenotypic attributes, may have had a greater influence on the differentiation of the groups. Additionally, one should consider that the molecular markers may

**Table 3.** Relative frequency of nominal categorical variables (%) in each of the five groups (A, B, C, D, and E) formed by the Ward-modified location model strategy of 40 tomato accessions of the UENF gene bank.

Nominal variables	Groups				
	A	B	C	D	E
<b>Foliage density</b>					
Sparse	0	14.3	11.2	25.0	0
Intermediate	90.0	71.4	44.4	75.0	100.0
Dense	10.0	14.3	44.4	0.0	0
<b>Leaf type</b>					
Potato leaf type	100.0	100.0	11.2	100.0	100.0
Standard	0	0	88.8	0	0
<b>Exterior color of immature fruit</b>					
Green	50.0	0	22.2	75.0	33.3
Very dark green	50.0	100.0	77.8	25.0	66.7
<b>Predominant fruit shape</b>					
Slightly flattened	40.0	57.2	33.3	75.0	100.0
Rounded	20.0	28.6	55.6	25.0	0
High rounded	30.0	0	11.1	0	0
Heart-shaped	0.0	7.1	0	0	0
Long oblong	10.0	7.1	0	0	0
<b>Presence of green stripes on the fruit</b>					
Absent	10.0	0	0	0	0
Present	90.0	100.0	100.0	100.0	100.0
<b>Intensity of green shoulder</b>					
Absent	10.0	0	0	0	0
Strong	90.0	100.0	100.0	100.0	100.0
<b>Radial cracking</b>					
Slight	40.0	57.2	55.6	100.0	100.0
Intermediate	40.0	35.7	44.4	0	0
Severe	20.0	7.1	0	0	0
<b>Concentric cracking</b>					
Slight	20.0	42.9	77.8	100.0	33.3
Intermediate	70.0	42.9	22.2	0	66.7
Severe	10.0	14.2	0	0	0
<b>Number of locules per fruit</b>					
2	20.0	21.4	22.2	0	0
3	40.0	50.0	66.7	0	66.7
4	20.0	28.6	11.1	0	33.3
5	20.0	0	0	100.0	0
<b>Exterior color of mature fruit</b>					
Yellow	0	0	11.1	0	0
Pink	20.0	14.3	33.3	50.0	0
Red	80.0	85.7	55.6	50.0	100.0
<b>Presence of open locules</b>					
Absent	0	7.1	11.1	50.0	0
Present	100.0	92.9	88.9	50.0	100.0

represent coding or non-coding regions and are therefore weakly or not at all correlated with the phenotypes observed. Previous studies carried out with the same group of tomato accessions indicated a weak correlation between RAPD markers and discrete nominal variables



**Table 4.** Means and range of variation of continuous variables for each of the five groups (A, B, C, D, and E) by Ward-modified location model strategy of 40 accessions of the UENF tomato gene bank.

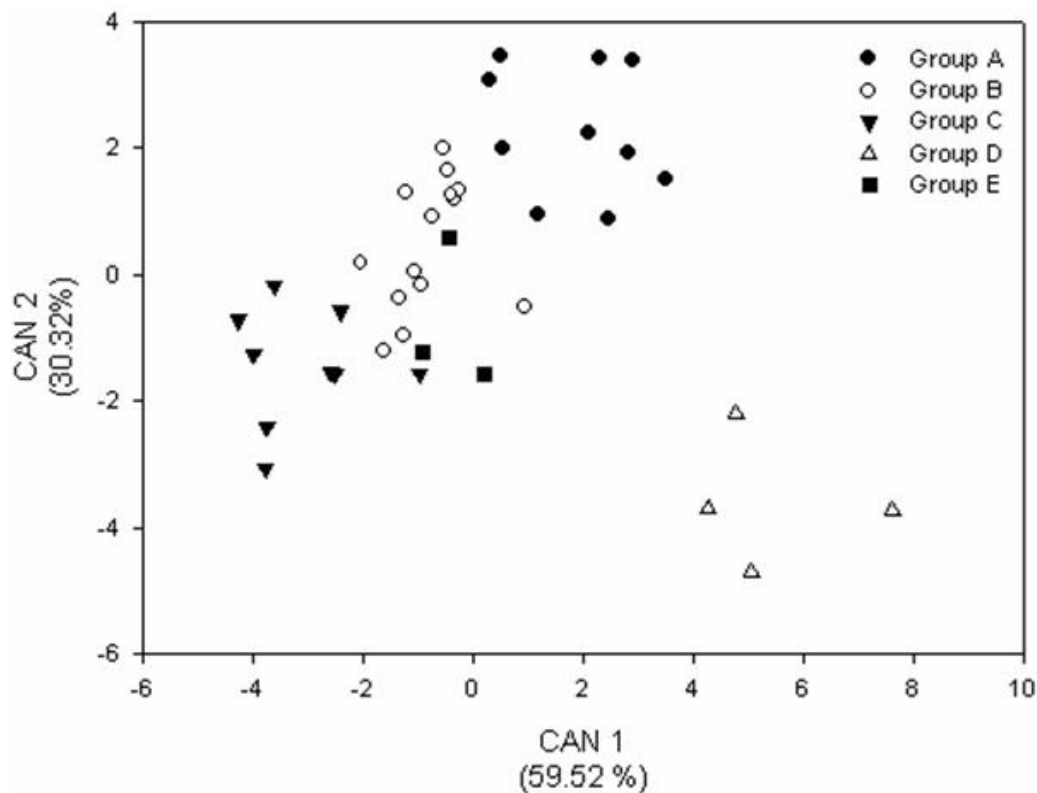
Agronomic descriptors	Groups				
	A	B	C	D	E
TFN	145.1 (64.0-325.7)	352.2 (106.7-717.3)	370.1 (196.3-815.3)	77.3 (26.7-126.3)	149.0 (90.3-241.7)
TFW (g)	4563.2 (2597.8-7306.5)	7148.6 (2313.7-12618.1)	7843.8 (5526.0-11123.0)	4484.8 (2316.9-8040.3)	5660.0 (3514.6-7570.3)
MFW (g)	33.7 (18.0-45.7)	22.02 (9.0-31.3)	23.5 (13.0-39.7)	68.0 (44.67-88.3)	42.2 (33.0-52.3)
FL (mm)	40.1 (34.3-46.3)	34.86 (27.0-40.0)	36.4 (34.3-39.7)	43.0 (39.0-46.7)	38.0 (36.3-40.0)
FW (mm)	44.4 (35.7-50.3)	38.8 (25.3-48.7)	40.4 (35.3-49.3)	55.8 (50.7-60.0)	49.4 (46.3-54.7)
NDF (days)	60.7 (58.9-64.3)	59.9 (56.4-61.1)	59.8 (58.6-61.1)	59.0 (57.8-60.9)	60.8 (59.6-61.7)
NDM (days)	104.9 (98.7-113.7)	98.5 (92.3-105.3)	94.5 (92.3-99.7)	99.2 (95.0-104.7)	98.9 (96.0-101.0)
NFI	5.3 (4.0-7.7)	6.2 (5.0-7.3)	6.1 (4.3-7.3)	6.5 (4.7-7.7)	6.4 (6.0-7.0)
TSS (°Brix)	3.8 (3.0-4.7)	4.4 (3.3-5.3)	3.9 (3.0-4.7)	3.9 (3.0-5.0)	3.4 (3.0-4.0)

TFN = total fruit number per plant; TFW = total fruit weight; MFW = mean fruit weight; FL = fruit length; FW = fruit width; NDF = number of days to flowering; NDM = number of days to maturity; NFI = number of flowers per inflorescence; TSS = total soluble solids.



(Gonçalves et al., 2008a). This suggests that the contribution of molecular markers to clustering is greater, mainly when the value of polymorphism is high, which can be bypassed by using the minimally sufficient number of markers to characterize the population variability (Franco et al., 2001). The Ward-MLM procedure should therefore be applied carefully to avoid the use of an excessive and unnecessary number of markers, which could favor the molecular markers in determining the divergence among accessions. However, since the minimal number of markers is germplasm-dependent and must be recalculated for each new divergence analysis, this strategy is too complex for direct use with techniques currently available.

The first two canonical variables explained about 90% of the variability among groups (Figure 1). This high value indicates that the two-dimensional graphical representation is appropriate for viewing the relationship among groups and among accessions within the groups. Regarding the distances between groups (Table 5), the highest value was observed between groups C and D, with a magnitude of 77.75, while the smallest distance was between the groups B and C, with an estimated value of 10.87.



**Figure 1.** Two-dimensional distribution of canonical variables (CAN) of the five groups (A, B, C, D, and E) established by the Ward-modified location model method for 40 accessions of the UENF tomato gene bank.

**Table 5.** Distance between groups observed for 40 accessions of tomato gene bank of the UENF using the generalized distance of Mahalanobis.

Groups	Groups			
	B	C	D	E
A	12.43	38.38	47.65	21.64
B		10.87	55.81	13.31
C			77.75	16.75
D				51.96

The results of this study show that continuous and discrete data (nominal and binary) can provide a more comprehensive analysis of the genetic diversity available in a gene bank. However, care should be taken in choosing descriptor types and descriptor numbers for use, in order to avoid a greater contribution of molecular markers in the quantification of diversity and discrimination in gene bank accessions. Summarizing, it is necessary for the scientific community to be involved in related research to evaluate the practicality and need of analyzing both phenotypic and molecular data.

## ACKNOWLEDGMENTS

L.S.A. Gonçalves was the recipient of a scholarship for Master's degree studies from the Fundação Carlos Chagas Filho de Amparo à Pesquisa do Estado do Rio de Janeiro (FAPERJ).

## REFERENCES

- Bai Y and Lindhout P (2007). Domestication and breeding of tomatoes: what have we gained and what can we gain in the future? *Ann. Bot.* 100: 1085-1094.
- 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.
- Barrero LS and Tanksley SD (2004). Evaluating the genetic basis of multiple-locule fruit in a broad cross section of tomato cultivars. *Theor. Appl. Genet.* 109: 669-679.
- Crossa J and Franco J (2004). Statistical methods for classifying genotypes. *Euphytica* 137: 19-37.
- Doyle JJ and Doyle JL (1987). A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytoch. Bull.* 19: 11-15.
- Foolad MR (2007). Genome mapping and molecular breeding of tomato. *Int. J. Plant Genomics* 2007: 64358.
- Franco J, Crossa J, Villasenõr J, Taba S, et al. (1998). Classifying genetic resources by categorical and continuous variables. *Crop Sci.* 38: 1688-1696.
- Franco J, Crossa J, Ribaut JM, Betran J, et al. (2001). A method for combining molecular markers and phenotypic attributes for classifying plant genotypes. *Theor. Appl. Genet.* 103: 944-952.
- Gepts P (2006). Plant genetic resources conservation and utilization. *Crop Sci.* 46: 2278-2292.
- Gonçalves LSA, Rodrigues R, Sudré CP, Bento CS, et al. (2008a). Divergência genética em tomate estimada por marcadores RAPD em comparação com descritores multicategóricos (Genetic divergence among tomato accessions using RAPD markers and its comparison with multicategorical descriptors). *Hortic. Bras.* 26: 364-370.
- Gonçalves LS, Rodrigues R, Amaral AT Jr, Karasawa M, et al. (2008b). Comparison of multivariate statistical algorithms to cluster tomato heirloom accessions. *Genet. Mol. Res.* 7: 1289-1297.
- Gower JC (1971). A general coefficient of similarity and some of its properties. *Biometrics* 27: 623-637.
- Jordan JA (2007). The Heirloom tomato as cultural object: investigating taste and space. *Sociol. Ruralis* 47: 20-41.
- Karasawa M, Rodrigues R, Sudré CP, Silva MP, et al. (2005). Aplicação de métodos de agrupamento na quantificação

- da divergência genética entre acessos de tomateiro (Cluster analysis in quantifying genetic divergence in tomato accessions). *Hortic. Bras.* 23: 1000-1005.
- Krzanowski WJ (1983). Distance between populations using mixed continuous and categorical variables. *Biometrika* 70: 235-243.
- Matusita K (1956). Decision rules based on the distance for problems of fit, two samples, and estimation. *Ann. Math. Stat.* 8: 67-77.
- Mohammadi SA and Prasanna BM (2003). Analysis of genetic diversity in crop plants - salient statistical tools and considerations. *Crop Sci.* 43: 1235-1248.
- Olkin I and Tate RF (1961). Multivariate correlation models with mixed discrete and continuous variables. *Ann. Math. Statist.* 32: 448-465.
- Ortiz R, Crossa J, Franco J, Sevilla R, et al. (2008). Classification of Peruvian highland maize races using plant traits. *Genet. Res. Crop Evol.* 55: 151-162.
- Padilla G, Carrea M, Rodríguez V and Ordás A (2005). Genetic diversity in a germplasm collection of *Brassica rapa* subsp *rapa* L. from northwestern Spain. *Euphytica* 145: 171-180.
- Rick CM (1976). Tomato, *Lycopersicon esculentum* (Solanaceae). In: Evolution of Crop Plants (Simmonds NW, ed.). Longman Group, London, 268-273.
- Saavedra G, Spoor W and Harrier L (2001). Molecular markers and genetic base broadening in *Lycopersicon* spp. *Acta Horticulturae* 546: 503-507.
- SAS Institute Inc. (1999). SAS OnlineDoc1. Version 8. SAS Institute Inc., Cary.
- Sudré CP, Leonardecz E, Rodrigues R, Amaral Júnior AT, et al. (2007). Genetic resources of vegetable crops: a survey in the Brazilian germplasm collections pictured through papers published in the journals of the Brazilian Society for Horticultural Science. *Hortic. Bras.* 25: 493-503.
- Terzopoulos PJ and Bebel PJ (2008). DNA and morphological diversity of selected Greek tomato (*Solanum lycopersicum* L.) landraces. *Sci. Horticulturae* 116: 354-361.
- 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.