Genetic variability in domesticated *Capsicum* spp as assessed by morphological and agronomic data in mixed statistical analysis

C.P. Sudré[1], L.S.A. Gonçalves[1], R. Rodrigues[1], A.T. do Amaral Júnior[1], E.M. Riva-Souza[2] and C. dos S. Bento[1]

[1]Universidade Estadual do Norte Fluminense Darcy Ribeiro,
Campos dos Goytacazes, RJ, Brasil
[2]Instituto Capixaba de Pesquisa, Assistência Técnica e Extensão Rural,
Vitória, ES, Brasil

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

ABSTRACT: *Capsicum* species are very important in Brazil because of economic, cultural and biological factors, and the country is considered to be a diversity center for this genus. Collection and maintenance of the genetic diversity in *Capsicum* are important to avoid genetic erosion. Besides the identification of species, the characterization and evaluation of accessions maintained in gene banks are of fundamental importance. For this purpose, multivariate methods have become an important tool in the classification of conserved genotypes. The objectives of this study were: i) to identify and characterize accessions of the *Capsicum* spp collection and draw conclusions about the potential use of certain accessions in different production sectors; ii) to estimate the genetic divergence among accessions using the Ward-MLM procedure, and iii) to evaluate the

efficiency of the analysis of continuous and categorical data using the Ward-MLM procedure. Fifty-six *Capsicum* spp accessions were evaluated based on 25 descriptors, 14 of which were morphological and 11 agronomic. Based on the qualitative descriptors, it was possible to identify all species and, together with the agronomic descriptors, genotypes could be indicated with potential for use in various production sectors. Five was determined as the ideal number of groups by the criteria pseudo-F and pseudo-t^2. The Ward-MLM procedure allowed the differentiation of the species *C. annuum*, *C. frutescens*, *C. baccatum*, and *C. chinense* in separate groups. The Ward-MLM procedure showed some level of efficiency in clustering *Capsicum* species analyzing morphological and agronomic data simultaneously.

Key words: Germplasm characterization; Plant descriptors; Categorical variables; Quantitative data; Ward-MLM procedure

INTRODUCTION

A correct botanical species classification is essential for the proper management of a germplasm. An erroneous identification of species maintained in gene banks can lead to losses ranging from the propagation and inadequate conservation of accessions to the delivery of misidentified genetic material to other institutions, resulting in a waste of time and financial resources (Gotor et al., 2008).

Capsicum species are traditionally identified by morphological descriptors or related traits. Flower morphology, including flower color, calyx constriction and the number of flowers per axil, is most used in taxonomic descriptions (Pickersgill, 1971; Moscone et al., 2007; Ince et al., 2009). However, other descriptors are considered essential for a more accurate germplasm characterization, such as the ones indicated by the IPGRI (1995).

In this genus, native to the Americas, more than 30 species have been described, but only five of them, *Capsicum annuum* var. *annuum*, *C. chinense*, *C. frutescens*, *C. baccatum* var. *pendulum*, and *C. pubescens* are considered to be domesticated (Pozzobon et al., 2006; Moscone et al., 2007). It is, however, expected that new *Capsicum* species will be identified, especially in Atlantic Forest areas in Brazil (Bianchetti, 1996; Pozzobon et al., 2006). Moreover, in Brazil *Capsicum* plants are of economic, cultural and biological importance due to the use of the fruits in preparing dishes of typical culinary traditions, which is diverse among regions and sometimes unique in each region of the country. The multiple and diverse ways to use these fruits are related to multicultural and ethnic aspects, and it is probably the result of the genetic diversity observed in each region of the country, since Brazil is considered to be one of the diversity centers of *Capsicum*.

The characterization and the evaluation of the *Capsicum* domesticated species are particularly interesting for gene bank curators, since a wide variability, not yet fully known and exploited, is available in these species (Guzmán et al., 2005; Sudré et al., 2006; Ince et al., 2009). Despite the accuracy in estimating genetic divergence among

accessions by molecular markers, knowledge of the phenotype given by morphological and agronomical descriptors is still important. Besides the aspect of correct species identification, the characterization and evaluation of conserved genotypes are of fundamental importance, by deepening our knowledge and making it possible to detect genotypes for use in breeding programs and plant duplicates in gene banks (Franco et al., 2005; Gonçalves et al., 2008; Laurentin, 2009).

With the advancement of computer technology, multivariate methods have become an important tool for the classification of conserved genotypes (Ortiz et al., 2008; Gonçalves et al., 2009). The statistical classification procedures require a data set based on a number of variables. In general, hierarchical ranking only considers continuous or discrete data separately. However, strategies for a hierarchical joint data analysis were proposed by Gower (1971) and Wishart (1986), which are strategies associated with the use of the unweighted pair group method with arithmetic mean (UPGMA) and the Ward, centroid and median methods (Franco et al., 1998).

New advances in joint and non-hierarchical analysis of multivariate data have been achieved by Wolfe (1970) who proposed the Gaussian model, which examines only continuous variables. This model was later modified by Lawrence and Krzanowski (1996) with the proposal of the location model (LM), which classifies n individuals when p quantitative variables and q qualitative variables are evaluated in an environment. LM combines all levels of qualitative variables in a multinomial variable, W, with m levels. However, LM is a conditional model since the mean estimates of the continuous variables depend on the values of variable W (Franco et al., 1998; Crossa and Franco, 2004).

The modified location model (MLM) was proposed by Franco et al. (1998) as a modification of LM, assuming that m levels of variable W and the p-multinormal variables for each sub-population are independent. MLM consists of two stages, where the first groups are defined by the grouping method of Ward (Ward Jr., 1963) using the dissimilarity matrix of Gower (Gower, 1971), and in the second stage, the mean of the vector of the quantitative variable for each sub-population, independent of the W values, is estimated by MLM.

The objectives of this study were: i) to identify and to characterize accessions of the *Capsicum* spp collection from different geographical regions of Brazil to detect potential uses of certain accessions in different production segments, and ii) to estimate the genetic divergence among accessions and to evaluate the efficiency of the analysis of continuous and categorical data using the Ward-MLM procedure for classifying and measuring the diversity among accessions.

MATERIAL AND METHODS

A collection of 56 *Capsicum* accessions of the gene bank of the Universidade Estadual do Norte Fluminense Darcy Ribeiro (UENF) was used in this study. The gene bank accessions were collected at local markets and fairs in different Brazilian states and were donated by farmers from four of the five different geographic regions of the country (North, Northeast, Central-West, and Southeast), and three commercial varieties were included from donations by seed companies. Four accessions came from Mexico (Table 1).

Table 1. Code number, species identification, fruit type (chili or sweet pepper) and source (county, state and country) of 56 *Capsicum* spp accessions.

Code number	Species	Fruit type	Source	
			County (State)	Country
UENF 1381	*C. annuum* var. *annuum*	Chili pepper	-	Mexico
UENF 1417	*C. baccatum* var. *pendulum*	Chili pepper	Lavras (Minas Gerais)	Brazil
UENF 1418	*C. chinense*	Chili pepper	Lavras (Minas Gerais)	Brazil
UENF 1419	*C. chinense*	Chili pepper	Lavras (Minas Gerais)	Brazil
UENF 1420	*C. annuum* var. *annuum*	Sweet pepper 'Apolo'	Commercial variety from seed company	Brazil
UENF 1421	*C. annuum* var. *annuum*	Sweet pepper 'Hércules'	Commercial variety from seed company	Brazil
UENF 1422	*C. annuum* var. *annuum*	Sweet pepper 'Ikeda'	Commercial variety from seed company	Brazil
UENF 1423	*C. annuum* var. *annuum*	Chili pepper	Aracaju (Sergipe)	Brazil
UENF 1424	*C. chinense*	Chili pepper	Campos dos Goytacazes (Rio de Janeiro)	Brazil
UENF 1425	*C. frutescens*	Chili pepper	Campos dos Goytacazes (Rio de Janeiro)	Brazil
UENF 1426	*C. baccatum* var. *pendulum*	Chili pepper	Campos dos Goytacazes (Rio de Janeiro)	Brazil
UENF 1489	*C. baccatum* var. *pendulum*	Chili pepper	Campos dos Goytacazes (Rio de Janeiro)	Brazil
UENF 1490	*C. baccatum* var. *pendulum*	Chili pepper	Campos dos Goytacazes (Rio de Janeiro)	Brazil
UENF 1491	*C. frutescens*	Chili pepper	Campos dos Goytacazes (Rio de Janeiro)	Brazil
UENF 1492	*C. baccatum* var. *pendulum*	Chili pepper	Campos dos Goytacazes (Rio de Janeiro)	Brazil
UENF 1493	*C. chinense*	Chili pepper	Campos dos Goytacazes (Rio de Janeiro)	Brazil
UENF 1494	*C. baccatum* var. *pendulum*	Chili pepper	Campos dos Goytacazes (Rio de Janeiro)	Brazil
UENF 1495	*C. baccatum* var. *baccatum*	Chili pepper	Campos dos Goytacazes (Rio de Janeiro)	Brazil
UENF 1496	*C. baccatum* var. *pendulum*	Chili pepper	Campos dos Goytacazes (Rio de Janeiro)	Brazil
UENF 1497	*C. chinense*	Chili pepper	Campos dos Goytacazes (Rio de Janeiro)	Brazil
UENF 1498	*C. chinense*	Chili pepper	Campos dos Goytacazes (Rio de Janeiro)	Brazil
UENF 1499	*C. baccatum* var. *pendulum*	Chili pepper	Campos dos Goytacazes (Rio de Janeiro)	Brazil
UENF 1500	*C. baccatum* var. *pendulum*	Chili pepper	Campos dos Goytacazes (Rio de Janeiro)	Brazil
UENF 1501	*C. baccatum* var. *pendulum*	Chili pepper	Campos dos Goytacazes (Rio de Janeiro)	Brazil
UENF 1502	*C. annuum* var. *annuum*	Chili pepper	-	Mexico
UENF 1503	*C. annuum* var. *annuum*	Chili pepper	-	Mexico
UENF 1504	*C. baccatum* var. *pendulum*	Chili pepper	Viçosa (Minas Gerais)	Brazil
UENF 1551	*C. chinense*	Chili pepper (cheiro)	Goiânia (Goiás)	Brazil
UENF 1552	*C. chinense*	Chili pepper (bode)	Goiânia (Goiás)	Brazil
UENF 1553	*C. chinense*	Chili pepper (bode)	Goiânia (Goiás)	Brazil
UENF 1554	*C. chinense*	Chili pepper (bode)	Goiânia (Goiás)	Brazil
UENF 1555	*C. chinense*	Chili pepper (cumari)	Goiânia (Goiás)	Brazil
UENF 1556	*C. baccatum* var. *pendulum*	Chili pepper (dedo-de-moça)	Goiânia (Goiás)	Brazil
UENF 1557	*C. frutescens*	Chili pepper (malagueta)	Goiânia (Goiás)	Brazil
UENF 1558	*C. chinense*	Chili pepper	Campos dos Goytacazes (Rio de Janeiro)	Brazil
UENF 1559	*C. annuum* var. *glabriusculum*	Ornamental chili pepper	Cachoeiras de Macacu (Rio de Janeiro)	Brazil
UENF 1560	*C. frutescens*	Ornamental chili pepper	Cachoeiras de Macacu (Rio de Janeiro)	Brazil
UENF 1562	*C. annuum* var. *annuum*	Sweet pepper	Viçosa (Minas Gerais)	Brazil
UENF 1565	*C. annuum* var. *annuum*	Sweet pepper	Viçosa (Minas Gerais)	Brazil
UENF 1566	*C. annuum* var. *annuum*	Sweet pepper	Viçosa (Minas Gerais)	Brazil
UENF 1567	*C. annuum* var. *annuum*	Sweet pepper	Viçosa (Minas Gerais)	Brazil
UENF 1568	*C. annuum* var. *annuum*	Sweet pepper	Viçosa (Minas Gerais)	Brazil
UENF 1569	*C. annuum* var. *annuum*	Sweet pepper	Viçosa (Minas Gerais)	Brazil
UENF 1570	*C. chinense*	Chili pepper	Belém (Pará)	Brazil
UENF 1571	*C. chinense*	Chili pepper	Aracaju (Sergipe)	Brazil
UENF 1572	*C. chinense*	Chili pepper	Aracaju (Sergipe)	Brazil
UENF 1573	*C. baccatum* var. *pendulum*	Chili pepper	Duas Barras (Rio de Janeiro)	Brazil
UENF 1575	*C. annuum* var. *annuum*	Ornamental chili pepper	Campos dos Goytacazes (Rio de Janeiro)	Brazil
UENF 1576	*C. annuum* var. *glabriusculum*	Ornamental chili pepper	Rio das Ostras (Rio de Janeiro)	Brazil
UENF 1577	*C. chinense*	Chili pepper (cheiro)	Goiânia (Goiás)	Brazil
UENF 1578	*C. annuum* var. *annuum*	Chili pepper	-	Mexico
UENF 1584	*C. baccatum* var. *baccatum*	Chili pepper	Rio das Ostras (Rio de Janeiro)	Brazil
UENF 1585	*C. chinense*	Chili pepper	Parintins (Amazonas)	Brazil
UENF 1586	*C. chinense*	Chili pepper	Parintins (Amazonas)	Brazil
UENF 1587	*C. frutescens*	Chili pepper	Parintins (Amazonas)	Brazil
UENF 1588	*C. frutescens*	Chili pepper	Parintins (Amazonas)	Brazil

The accessions were characterized and evaluated based on morphological and agronomic descriptors. For the morpho-agronomic characterization and evaluation, an experiment was conducted under field conditions in UENF in Campos dos Goytacazes, Rio de Janeiro, Brazil (21º45'S, 41º18'W) using a randomized block design, with 56 accessions, three replications and 10 plants per plot.

Sowing was performed in polystyrene trays (polystyrene) of 128 cells with organic plant substrate. The seedlings were transplanted when they had grown two pairs of definitive leaves. Spacing was 1.00 m between rows and 0.50 m between plants. The calculation of fertilization was based on soil analysis and the needs of the crop.

For the morpho-agronomic characterization and evaluation, 25 descriptors were considered, with 14 morphological and 11 agronomic traits proposed for *Capsicum* by the IPGRI (1995) (Table 2).

Table 2. Descriptors studied and discrimination of the classes observed for 56 *Capsicum* accessions according to the IPGRI (1995).

Descriptor	Classes observed according to descriptors for *Capsicum* (IPGRI, 1995)
Morphological traits	
Stem color	1: green; 2: green with purple stripes; 3: purple
Anther color	2: yellow; 3: pale blue; 4: blue; 5: purple
Corolla color	1: white; 8: purple; 9: white with yellow-green spots; 10: white-green; 11: yellow with purple base; 12: purple with yellow base
Number of flowers per axil	1: one; 2: two; 3: three or more
Flower position	3: pendant; 5: intermediate; 7: erect
Plant growth habit	5: intermediate (compact); 7: erect
Immature fruit color	2: yellow; 3: green; 4: orange; 5: purple; 7: other
Fruit color at mature stage	1: white; 3: pale orange-yellow; 4: orange-yellow; 5: pale orange; 6: orange; 7: light red; 8: red; 9: dark red; 10: purple
Fruit shape	1: elongate; 2: almost round; 3: triangular; 4: campanulate; 5: blocky; 6: tomato-pepper; 7: ellipse; 8: Scotch bonnet
Fruit surface	1: smooth; 2: semi-wrinkled; 3: wrinkled
Number of locules	Observed cross-section of 10 cut fruits
Cotyledon color	1: light green; 5: purple; 7: variegated
Calyx annular constriction	0: absent; 1: present
Neck at base of fruit	0: absent; 1: present
Agronomic traits	
Plant height (cm)	Recorded when in 50% of plants the first fruit has begun to ripen
Plant canopy width (cm)	Measured immediately after first harvest, at the widest point
Days to flowering	Number of days from transplanting until 50% of plants have at least one open flower
Days to fruiting	Number of days from transplanting until 50% of plants bear mature fruits
Fruit length (cm)	Measured at the largest point in 10 ripe fruits
Fruit width (cm)	Measured at the widest point, as an average of 10 ripe fruits
Mean fruit weight (g)	Ratio of total fruit weight/total fruit number
1000-seed weight (g)	Estimated counting 250 seeds
Number of seeds per fruit	Counting the total number of seeds per fruit. Average of at least 10 fruits from 10 random plants
Fruit number per plant	Total number of fruits harvested in each plant
Fruit yield per plant (g)	Total weight of all harvested fruits in each plant

The significances of the continuous variables were initially analyzed by the F-test. Subsequently, the continuous and discrete variables were assessed simultaneously using the Ward-MLM procedure for the formation of accession groups by the CLUSTER and IML procedures of the SAS program (SAS Institute, 2000). To use the Ward clustering method, the distance matrix was obtained by Gower's algorithm (Gower, 1971). The ideal number of groups was defined based on the criteria pseudo-F and pseudo-t^2 (SAS Institute, 2000).

The difference between groups and correlation of variables with the canonical variable was graphically represented using the CANDISC procedure of the SAS program (SAS Institute, 2000). The distance proposed by Matusita (1956), adapted by Krzanowski (1983) and later by Franco et al. (1998) for the distribution of the joint variables (categorical and continuous) was used to determine the dissimilarity between the groups formed.

RESULTS AND DISCUSSION

Based on characterization by flower morphology, anther color was predominantly blue (36%), followed by yellow, purple and light blue, with 27, 25 and 12%, respectively. This descriptor is important for the identification of *Capsicum* species. The IPGRI descriptor list for *Capsicum* includes this trait as an essential descriptor because it is highly discriminatory among accessions (IPGRI, 1995). Of all accessions studied, 27% had yellow anthers, and therefore, they belong to *C. baccatum*. According to Andrews (1995), although *C. baccatum* belongs to the white-flowered group, its flowers have yellow spots and anthers. Of the 14 accessions with purple anthers, 12 belong to *C. chinense* and two accessions were recognized as belonging to the species *C. annuum* var. *glabriusculum*.

Regarding the corolla color, greenish white was predominate (34%), followed by white spots and white with yellow/green (with 25% each). By this descriptor alone, 57% of the accessions of the species could be identified. It can be used to characterize the cultivated species and wild relatives. However, corolla color is less suitable when the variation of the flower color of more distant wild species is described (Moscone et al., 2007).

Most accessions (62.50%) had only one flower per axil, a typical trait of *C. annuum*, but also observed in *C. baccatum* var. *pendulum*. It should be noted that the descriptor number of flowers per axil is extremely important in differentiating the botanical varieties *C. baccatum* var. *baccatum* from *C. baccatum* var. *pendulum*, since the latter is characterized by bearing only one flower per node, while nodes with two flowers predominate in the variety *C. baccatum* var. *baccatum*.

Most accessions (63%) were classified as intermediate with regard to the trait growth habit, followed by erect habit (37%). This descriptor is important in terms of crop management because it can help in terms of defining the area for each plant, harvesting, weed control and feasibility to calculate the volume of chemical spray.

The immature fruit color was highly variable, especially when the fruit was observed in the different developmental stages. For some accessions, more than six different colors were reported until reaching the final color (mature stage), e.g., in UENF 1423 and UENF 1575. Other accessions, in turn, only varied in color intensity (UENF 1503, UENF 1565, UENF 1576, UENF 1585, and UENF 1586). The descriptor immature fruit color, as proposed by the IPGRI (1995), does not consider all maturation stages, but rather the phase prior to maturation only. Nevertheless, based on the description of the IPGRI, divergence was observed among accessions in the immature fruits; most were orange (52%), followed by green (29%), brown (10.3%), yellow (7%), and purple immature fruits (1.7%).

Nine categories were found for fruit color in the mature stage, confirming the divergence among the genotypes studied regarding the diversity of pigments observed in the fruits. The color red was predominant (34%), followed by light red (23%) and dark red (18%). The wine color was observed in only two accessions (UENF 1585 and UENF 1586) from the Amazon region, with plants that also have a purplish color of some plant parts (leaf and stem). There was variation in the color hues, ranging from white to orange, with colors close to yellow totaling 21.4%.

Fruit shape was determined based on comparison with the shapes proposed in the list of descriptors of the IPGRI (1995), and three fruit shapes observed did not match with any of the shapes described in this list. However, the fruit shapes had already been described: one of them is a *C. baccatum* fruit similar to *C. annuum* tomato-pepper or tomato (Andrews, 1995), with flattened fruits and several locules. The second fruit type not found in IPGRI descriptors had an

ellipse shape; the third one resembles the "Scotch bonnet" type, but flattened at its apex. The three fruit shapes received the categorical numbers, 6, 7 and 8, respectively. The predominant shape was elongate (35.7%) followed by triangular (23.2%) and square/block (12.5%).

The descriptor calyx constriction was not detected in 66% of the accessions and observed in 34%. This descriptor is of great importance in differentiating *C. chinense* from *C. frutescens*, since it is present in the first and absent in the second species (Pickersgill, 1971), and it is used to separate these two species from each other.

Based on the qualitative descriptors evaluated, it was possible to identify all species, because 50% of the accessions of the collection had not been properly classified from a botanical point of view. Besides species identification, it was also possible to distinguish two botanical varieties of *C. baccatum* (*C. baccatum* var. *pendulum* from *C. baccatum* var. *baccatum*) and two of *C. annuum* (*C. annuum* var. *annuum* from *C. annuum* var. *glabriusculum*). The collection of morphological data is practical and economical, compared to the collection of quantitative and molecular data (Sudré et al., 2006). However, each strategy has a specific value, and it is preferable that a germplasm collection be widely based for a solid support of research work and database of the collection.

In univariate analysis of variance of the 11 continuous variables, a significant effect (P < 0.01) for all traits was observed, demonstrating the existence of genetic variability among the accessions studied. On the basis of the means of agronomic and morphological traits, genotypes with potential for use in various production sectors can be indicated, e.g., for the use in sauces, canned food and raw, as well as in *Capsicum*-breeding programs.

The accessions of chili pepper UENF 1557 and UENF 1570 (both *C. frutescens*) and UENF 1490, UENF 1492, and UENF 1496 (*C. baccatum* var. *pendulum*), which have a peculiar aroma apart from high pungency, are suitable for use as condiments and dressings. Similarly, the *C. chinense* accessions UENF 1551 and UENF 1577 (pimenta-de-cheiro) may also be indicated for use as a spice; however, they have a strong aroma associated with little pungency. The accessions UENF 1559 and UENF 1576 (both *C. annuum* var. *glabriusculum*) have a greater decorative value, with purple flowers, fruits with varied colors and short plant height. Lastly, the accessions UENF 1423 and UENF 1575 (both *C. annuum* var. *annuum*) produce more than six fruit colors in the ripening stage, white flowers and short plant height, and also have a potential ornamental use.

Mainly due to the versatility of application, *Capsicum* plants have become important from an economic point of view in several countries since their fruits can be used for different purposes, such as in cooking (raw or processed), in industry (e.g., production of "pepper spray") and for medicinal and ornamental purposes (Pickersgill, 1971; Moscone et al., 2007). In Brazil, this vegetable is of great importance, ranking second in the generation of foreign currency among exported vegetables (Embrapa, 2009). Another highly relevant aspect is the social and economic importance, since the crop contributes to a higher income for farmers of small properties and helps maintain workers in the rural areas, mainly due to the high-labor demand during harvest.

In terms of diversity, both continuous and categorical variables were used for grouping the 56 accessions into homogeneous groups by the Ward-MLM procedure. First, the value of five was determined as the optimal number of groups based on the criteria of pseudo-F and pseudo-t^2, since the maximum value was reached at this point.

The sequence of the strategy of the Ward-MLM classification allowed a separation of the species *C. annuum*, *C. frutescens*, *C. baccatum*, and *C. chinense* into separate clusters

(Tables 3 and 4). However, only in two groups (G3 and G5), the accessions clustered belonged to the same species, with G3 grouping only *C. baccatum* accessions and G5 clustering just *C. annuum* accessions. Group G1 was formed by 12 accessions of *C. annuum* var. *annuum* and only one accession of *C. baccatum* var. *pendulum*. Of the 12 accessions of *C. annuum* var. *annuum*, aleven are sweet peppers because UENF 1381 is pungent. Group G2 was represented by six accessions, five *C. frutescens* and one *C. chinense*. Group G3 consisted of 13 accessions of *C. baccatum*, of which 11 belonged to *C. baccatum* var. *pendulum* and two to *C. baccatum* var. *baccatum*. Group G4 joined 19 accessions, with 17 *C. chinense*, one *C. frutescens* and one *C. baccatum* var. *pendulum*. Group G5 was composed of only five accessions, three *C. annuum* var. *annuum* and two *C. annuum var*. *glabriusculum*. Also, there was no association between clusters formed with geographic location where the accessions were collected.

Table 3. Variables and accession number per group for categorical traits in each of the five groups (G1, G2, G3, G4, and G5) formed by the Ward-MLM strategy from 56 *Capsicum* spp accessions.

Variable	Group				
	G1 (13)	G2 (6)	G3 (13)	G4 (19)	G5 (5)
Stem color					
Green	12	6	12	8	1
Green with purple stripes	1	-	1	8	3
Purple	-	-	-	3	1
Anther color					
Yellow	1	-	13	1	-
Pale blue	-	4	-	1	2
Blue	12	1	-	6	1
Purple	-	1	-	11	2
Corolla color					
White	12	-	-	-	3
White with yellow-green spots	1	-	13	1	-
White-green	-	6	-	14	-
Purple	-	-	-	-	2
Yellow with purple base	-	-	-	2	-
Purple with yellow base	-	-	-	2	-
Number of flowers per axil					
One	13	4	11	2	5
Two	-	-	2	7	-
Three	-	2	-	10	-
Flower position					
Pendant	1	-	-	-	-
Intermediate	9	-	-	10	1
Erect	3	6	13	9	4
Plant growth habit					
Intermediate	1	4	13	13	4
Erect	12	2	-	6	1
Immature fruit color					
Yellow	-	1	-	3	-
Green	8	1	-	6	1
Orange	2	4	13	7	3
Purple	-	-	-	1	-
Other	3	-	-	2	1
Fruit color at mature stage					
White	-	1	-	1	-
Pale orange-yellow	-	1	-	-	-
Orange-yellow	-	-	-	2	-
Pale orange	-	-	-	4	-
Orange	-	-	-	3	-
Light red	-	3	8	2	-
Red	4	1	5	5	4
Dark red	9	-	-	-	1
Purple	-	-	-	2	-

Table 3. Continued.

Variable	Group				
	G1 (13)	G2 (6)	G3 (13)	G4 (19)	G5 (5)
Fruit shape					
Elongate	6	5	6	2	1
Almost round	-	-	-	5	-
Triangular	6	-	2	2	3
Campanulate	-	-	-	5	-
Blocky	1	-	3	3	-
Other – Tomato-pepper	-	-	1	1	-
Other – Ellipse	-	1	-	1	1
Other – Scotch bonnet	-	-	1	-	-
Fruit surface					
Smooth	9	1	3	12	3
Semi-wrinkled	3	1	6	5	2
Wrinkled	1	4	4	2	-
Number of locules					
Two	7	4	3	2	4
Three	5	2	9	14	1
Four	1	-	1	3	-
Cotyledon color					
Light green	13	5	13	15	3
Purple	-	1	-	3	2
Variegated	-	-	-	1	-
Calix annular constriction					
Present	-	1	1	17	-
Absent	13	5	12	2	5
Neck at base of fruit					
Present	3	1	3	2	-
Absent	10	5	10	17	5

Table 4. Means for continuous traits for each of the five groups (A-E) formed by the Ward-MLM strategy and the first two canonical variables determined from 56 *Capsicum* spp accessions.

Variable	Group					CAN1	CAN2
	G1	G2	G3	G4	G5		
Fruit length (cm)	10.3	2.7	4.8	3.3	3.3	-0.484	0.682
Fruit diameter (cm)	3.5	1.0	2.5	2.4	1.4	-0.195	0.504
Seed number per fruit	155.5	26.2	61.2	48.5	55.4	-0.612	0.685
Plant height (cm)	80.4	84.7	92.7	90.6	38.3	0.538	0.353
Canopy diameter (cm)	59.7	79.6	114.4	93.5	45.5	0.869	0.112
Weight of 1000 seeds (g)	5.7	3.3	4.8	4.0	3.5	-0.296	0.798
Days to flowering	41.7	71.3	58.2	55.2	36.6	0.779	-0.088
Days to fruiting	129.8	160.7	140.9	136.2	118.8	0.466	0.047
Fruit number per plant	39.8	381.5	265.6	289.7	126.6	0.425	-0.231
Fruit weight per plant (g)	599.4	238.2	1063.0	688.7	238.9	0.413	0.361
Fruit mean weight (g)	29.2	0.6	4.9	4.6	2.7	-0.529	0.605

The Ward-MLM methodology in data analysis for morpho-agronomic characterization of accessions allowed, with some level of efficiency, the separation of *Capsicum* species with the simultaneous use of morphological and agronomic variables. Most of the *C. annuum* (12 accessions) were clustered in the same group (G1), although one *C. baccatum* accession was also clustered in that group. Also, for *C. baccatum*, 13 of 15 accessions were grouped together (G3). *C. frutescens* and *C. chinense* accessions were grouped together into two different clusters (G2 and G4). Not surprisingly, these two species have been considered part of the *C.*

annuum complex and the most important trait to separate the two of them is the calyx constriction, as already mentioned before. Some authors even considered that *C. chinense* and *C. frutescens* should probably be combined into one species (Heiser Jr., 1976; Andrews, 1995).

Apart from G2 and G4, all others were separated by the graphical representation of the first two canonical variables (Figure 1). These two groups (G2 and G4) united accessions of the species *C. frutescens* and *C. chinense*, which have great similarity in morphological and agronomic characteristics and are distinguished only by the morphological descriptor calyx constriction (Pickersgill, 1971). Nevertheless, the first two canonical variables explained 90.5% of the variability among groups (Figure 1). This high value indicates that the graphical representation of the first two canonical variables is appropriate to show the relationship among groups and among accessions within a group. The best correlations of variables with the first canonical variable were plant canopy width, days to flowering, number of seeds, plant height, and fruit weight, and with the second canonical variable, weight of 1000 seeds, number of seeds per fruit, fruit length, fruit weight, and fruit diameter (Table 4).
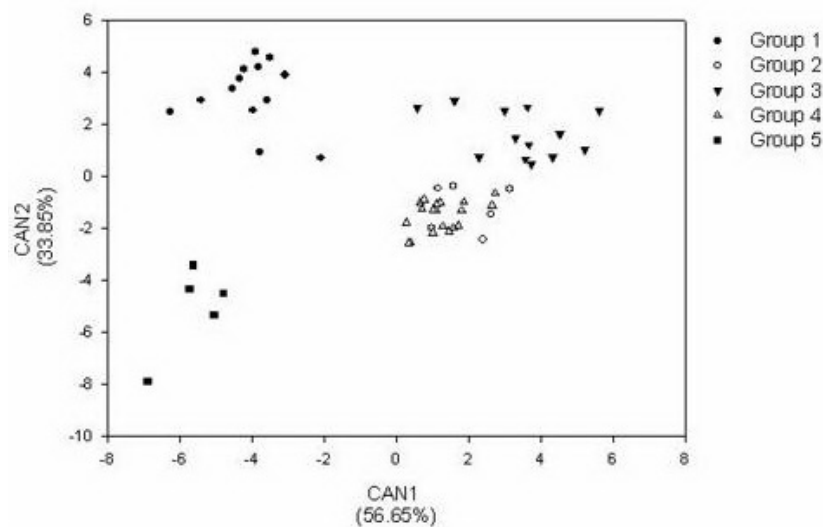

Figure 1. Plot of the first two canonical variables of the five groups (G1-G5) formed with the Ward-MLM strategy, considering 56 *Capsicum* spp accessions.

The shortest distance among groups was observed between G2 *(C. frutescens)* and G4 *(C. chinense)*, with a distance of 13.82 (Table 5), confirming the high similarity between the two species in morphological and agronomic traits. The greatest dissimilarity between groups, with a distance of 128.11, was observed between G3 (*C. baccatum)* and G5 *(C. annuum*), corroborating the proposed classification of these species into different gene complexes by Moscone et al. (2007) in investigations based on cytogenetic aspects and botanical characters.

Group	G1	G2	G3	G4	G5
G1	-	64.04	60.86	50.62	73.14
G2	64.04	-	26.35	13.82	88.98
G3	60.86	26.35	-	17.25	128.11
G4	50.62	13.82	17.25	-	63.22
G5	73.14	88.98	128.11	63.22	-

Table 5. Distance among clusters formed by the Ward-MLM based on Mahalanobis distance, for five clusters of 56 *Capsicum* spp accessions.

Similar results were obtained for other crops with the use of morphological and agronomic data only. Ortiz et al. (2008), working with 50 accessions of eight maize races from high-altitude regions in Peru, based on six agronomic descriptors, concluded that the Ward-MLM procedure produced an adequate genotype classification and that this analysis can be an additional refinement and a complement to the racial classification based on visual evaluation. Padilla et al. (2005), in studying genetic diversity and morphological similarities of 120 local *Brassica rapa* subsp. *rapa* L. varieties, concluded that the use of the Ward-MLM strategy results in the formation of homogeneous groups.

Considering the complexity of the taxonomic classification of the genus *Capsicum* (Hunziker, 2001; Barboza et al., 2005), morphological and agronomic data were sufficient for a good discrimination of the accessions. The strategy of analysis based on the Ward-MLM procedure was helpful in classifying and grouping the *Capsicum* species.

REFERENCES

Andrews J (1995). Peppers. The Domesticated Capsicums. University of Texas Press, Austin.

Barboza GE, Bianchetti LB and Lammers TG (2005). Three new species of *Capsicum* (Solanaceae) and a key to the wild species from Brazil. *Syst. Bot.* 30: 863-871.

Bianchetti LB (1996). Aspectos Morfológicos, Ecológicos e Biogeográficos de Dez Táxons de *Capsicum* (Solanaceae) Ocorrentes no Brasil. Msc. thesis, Universidade de Brasília/UNB, Brasília.

Crossa J and Franco J (2004). Statistical methods for classifying genotypes. *Euphytica* 137: 19-37.

Embrapa (2009). Exportações Brasileiras de Hortaliças 2000-2007. Available at [http:// www.cnph.embrapa.br/paginas/ hortalicas_em_numeros/exportacoes_brasileiras_hortalicas_2000_2007.pdf]. Accessed January 8, 2009.

Franco J, Crossa J, Villaseñor J, Taba S, et al. (1998). Classifying genetic resources by categorical and continuous variables. *Crop Sci.* 38: 1688-1696.

Franco J, Crossa J, Taba S and Shands H (2005). A sampling strategy for conserving genetic diversity when forming core subsets. *Crop Sci.* 45: 1035-1044.

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.

Goncalves LS, Rodrigues R, do Amaral Junior AT, Karasawa M, et al. (2009). Heirloom tomato gene bank: assessing genetic divergence based on morphological, agronomic and molecular data using a Ward-modified location model. *Genet. Mol. Res.* 8: 364-374.

Gotor E, Alercia A, Rao RV, Watts J, et al. (2008). The scientific information activity of Biodiversity International: the descriptor lists. *Genet. Res. Crop Evol.* 55: 757-772.

Gower JC (1971). A general coefficient of similarity and some of its properties. *Biometrics* 27: 857-871.

Guzmán FA, Azurdia HAC, Duque MC and Carmen de Vicente M (2005). AFLP assessment of genetic diversity of *Capsicum* genetic resources in Guatemala: home gardens as an option for conservation. *Crop Sci.* 45: 363-370.

Heiser CB Jr (1976). Peppers: *Capsicum* (Solanaceae). In: Evolution of Crop Plants. Longmans, London, 265-268.

Hunziker AT (2001). Genera Solanacearum: the Genera of Solanaceae, Illustrated, Arranged According to a New System. Koeltz Scientific Books, Königstein.

Ince AG, Karaca M and Onus AN (2009). Development and utilization of diagnostic DAMD-PCR markers for *Capsicum* accessions. *Genet. Res. Crop Evol.* 56: 211-221.

International Plant Genetic Resources Institute (IPGRI) (1995). Descriptores para *Capsicum* (*Capsicum* spp.). International Plant Genetic Resources Institute (IPGRI), Rome.

Krzanowski WJ (1983). Distance between populations using mixed continuous and categorical variables. *Biometrika* 70: 235-243.

Laurentin H (2009). Data analysis for molecular characterization of plant genetic resources. *Genet. Res. Crop Evol.* 56: 277-292.

Lawrence CJ and Krzanowski WJ (1996). Mixture separation for mixed-mode data. *Stat. Comput.* 6: 85-92.

Matusita K (1956). Decision rule, based on the distance, for the classification problem. *Ann. Inst. Statist. Math.* 8: 67-77.

Moscone EA, Scaldaferro MA, Grabiele M, Cecchini NM, et al. (2007). The evolution of Chili Peppers (*Capsicum* - Solanaceae): a cytogenetic perspective. VI International Solanaceae Conference: Genomics Meets Biodiversity. *Acta Hort.* 745: 137-170.

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, Cartea ME, Rodríguez VM and Ordás A (2005). Genetic diversity in a germplasm collection of *Brassica rapa* subsp *rapa* L. from northwestern Spain. *Euphytica* 145: 171-180.

Pickersgill B (1971). Relationships between weedy and cultivated forms in some species of chili peppers (Genus *Capsicum*). *Evolution* 25: 683-691.

Pozzobon MT, Schifino-Wittmann MT and Bianchetti LB (2006). Chromosome numbers in wild and semidomesticated Brazilian *Capsicum* L. (Solanaceae) species: do x = 12 and x = 13 represent two evolutionary lines? *Biol. J. Linn. Soc.* 151: 259-269.

SAS Institute (2000). SAS OnlineDOC. Version 8. SAS Institute Inc., Cary.

Sudré CP, Cruz CD, Rodrigues R, Riva EM, et al. (2006). Variáveis multicategóricas na determinação da divergência genética entre acessos de pimenta e pimentão. *Hortic. Bras.* 24: 88-93.

Ward JH Jr (1963). Hierarchical grouping to optimize an objective function. *J. Am. Stat. Assoc.* 58: 236-244.

Wishart D (1986). Hierarchical Cluster Analysis with Messy Data. In: Classification as a Tool of Research (Gaul W and Schader M, eds.). Elsevier, Amsterdam, 453-460.

Wolfe JH (1970). Pattern clustering by multivariate mixture analysis. *Multivariate Behav. Res.* 5: 329-350.