JUNIOR e SILVA WC; CARVALHO SIC; DUARTE JB. 2013. Identification of minimum descriptors for characterization of *Capsicum* spp. germplasm. *Horticultura Brasileira* 31: 190-202.

Identification of minimum descriptors for characterization of *Capsicum* spp. germplasm

Waldir C Junior e Silva[1]; Sabrina Isabel C de Carvalho[2]; João B Duarte[1]

[1]UFG, Escola de Agronomia, C. Postal 131, 74001-970 Goiânia-GO; waldircamargo@gmail.com; jbduarte@agro.ufg.br; [2]Embrapa Hortaliças, C. Postal 218, 70351-970 Brasilia-DF; sabrina.carvalho@embrapa.br

ABSTRACT

This study was proposed to select minimum descriptors to the characterization of accessions of *Capsicum* spp. from Embrapa Hortaliças germplasm bank, which are characterized mainly by qualitative data. From this collection, 893 (420 *C. annuum*, 106 *C. baccatum*, 307 *C. chinense* and 60 *C. frutescens*) were considered because they were completely characterized by 56 morphological descriptors. The statistical treatment was performed by factorial multiple correspondence analysis, combined with estimation of matrices of genetic dissimilarity among accessions for multicategoric data. Lists of minimum descriptors were proposed for the whole collection and for the subcollections related to the four cultivated species. Around thirty minimum descriptors were selected for each subcollection, maintaining a minimum of 0.90 for the correlation between the dissimilarity matrices that consider all the descriptors or only the selected minimum descriptors. These proposed lists allow a reduction of around 50% in the number of initial descriptors. The filament colour, mature fruit colour, number of locules, fruit position, origin, pungency, fruit surface, stem length, plant height and fruit length were selected as minimum descriptors, both to the whole collection and their subcollections. Most of these descriptors are related to the fruit traits of sweet and chilli peppers, which are important for genetic breeding of *Capsicum* because are related to storage, processing, marketing and consumption of commercial derived products.

Keywords: qualitative data, multicategoric variables, multiple correspondence analysis, discarding variables.

RESUMO

Identificação de descritores mínimos para caracterização de germoplasma de *Capsicum* spp.

Neste estudo buscou-se selecionar descritores mínimos para caracterização de acessos do banco de germoplasma de *Capsicum* spp. da Embrapa Hortaliças, caracterizado principalmente com base em dados de natureza qualitativa. Desta coleção foram considerados 893 acessos (420 *C. annuum*, 106 *C. baccatum*, 307 *C. chinense* e 60 *C. frutescens*), que estavam completamente caracterizados segundo 56 descritores morfológicos. O tratamento estatístico foi feito por análise fatorial de correspondência múltipla, combinada com estimação de matrizes de distâncias genéticas entre acessos, para dados multicategóricos. Foram propostas listas de descritores mínimos para a coleção completa e para as subcoleções relacionadas às quatro espécies. Cerca de trinta descritores foram selecionados para cada subcoleção, garantindo-se correlação mínima de 0,90 entre as matrizes de distâncias que consideram todos os descritores ou apenas os descritores mínimos selecionados. As listas propostas possibilitam reduções de quase 50% no número de descritores. Cor do filamento, cor do fruto maduro, número de lóculos, posição do fruto, procedência, pungência, superfície do fruto, comprimento da haste, comprimento da planta e comprimento do fruto foram eleitos como descritores mínimos, tanto para a coleção completa quanto para suas subcoleções. Grande parte destes descritores são características do fruto em pimentas e pimentões, importantes para o melhoramento genético de *Capsicum*, haja vista suas relações com armazenamento, processamento, comercialização e consumo dos produtos comerciais derivados.

Palavras-chave: dados qualitativos, variáveis multicategóricas, análise de correspondência múltipla, descarte de variáveis.

(Recebido para publicação em 16 de abril de 2012; aceito em 18 de fevereiro de 2013)
(*Received on April 16, 2012; accepted on February 18, 2013*)

Plant genetic diversity is a valuable guarantee against possible adversaries that threaten the survival of biological species. Research on genetic resources and plant breeding is one of the innovation activities most relevant for agriculture sustainability. In this sense, the access, organization and maintenance of important collections of exotic variability and the significant efforts with collection, characterization, conservation and promotion for the use of native species of agricultural food interest have helped to create the bases of sustainability for various genetic breeding programs.

Germplasm characterization consists of programmed and systematic obtaining of data based on the assessment of characteristics that describe and differentiate the existing accessions (Almeida *et al*., 2005). This characterization therefore focuses basically on morphological and phenological aspects observed in the accessions and their comparison with lists of descriptive characteristics (descriptors). The state of the descriptors can vary among the accessions, beyond that foreseen on the lists, and some may be shown to be invariable for whole collections. It can sometimes be observed that certain descriptors have no value for identifying duplicates or reliable discrimination of the accessions.

There are several examples

in the literature that show certain discriminatory incapacity of the traits used to classify accessions in germplasm banks. Onyilgha (1986) worked with the *Dioscorea* species and concluded that only 15 of 40 morphological traits were sufficient to discriminate the accessions. Bekele *et al*. (1994) selected only 23 descriptors from a list of 68 variables used to characterize a collection of cacao. Strapasson *et al*. (2000) discarded 86% of the descriptors used in characterizing of *Paspalum* sp. accessions.

Some descriptor lists, mostly prepared by research groups commissioned by Biodiversity International, the former IPGRI (International Plant Genetic Resources Institute), require taking a great number of observations, making it a very time-consuming and expensive task. Frankel & Brown (1984) criticized the extension and level of detailing on these lists. They point out that this effort did not bring the equivalent gains to the heavy workload imposed on curators and breeders. For these authors, the morphological and phenological data taking should be restricted to identifying duplicates and establishing nuclear collections.

Reliable methods and processes for germplasm characterization are fundamental to increase the use of the variability available. Its efficiency would anticipate the potential of each accession in the face of the desired genetic characteristics. For this, the use of multivariate techniques has been efficient not only in identifying the descriptors of greater interest, but also in discarding those that are not so relevant (Pereira, 1989). These procedures have been used in field crops such as *Paspalum* sp. (Strapasson *et al*., 2000), *Capsicum chinense* (Luz, 2007), tomato (Gonçalves *et al*., 2009) and *Capsicum* spp. (Ortiz *et al*., 2010; Sudré *et al*., 2010).

The objective of the present study was to select, from the descriptors available in the characterization of the *Capsicum* spp. collection of Embrapa Hortaliças, in Brazil, those essential for a reliable discrimination of the accessions, both for the whole collection and for the subcollections related to their four cultivated species (*C. annuum*, *C. baccatum*, *C. chinense* and *C. frutescens*). This will reduce efforts in characterizing the collections of this kind, based on smaller lists of descriptors.

MATERIAL AND METHODS

The data for this research were obtained at Embrapa Hortaliças that has a germplasm collection of chilli and sweet peppers (*Capsicum* spp.) with about 4,000 accessions. From this total, 893 accessions were considered in the study. They were brought from different regions of Brazil and abroad, and are completely characterized by 56 morphological descriptors (Table 1). As most of these descriptors are qualitative data, the other traits were also categorized, to gather them together in the same type of analysis. The 893 accessions are distributed in 420 *C. annuum* accessions, 106 *C. baccatum* accessions, 307 *C. chinense* accessions, and 60 *C. frutescens* accessions.

Multiple Correspondence Analysis (MCA) (Escofier & Pagès, 1992) was used to select the minimum descriptors. Analogically to the known Principle Components Analysis (PCA), applied to quantitative data, this analysis also allows to study a population of *n* individuals assessed according to *p* qualitative variables. The selection of descriptors was based mainly on the approach proposed by Curry (1993) and Dias (1994), based on method of Jolliffe (1973), substituting PCA for MCA. According to this proposal, after carrying out the analysis with all the descriptors (complete analysis), the one that most contributed to the last factorial axis is discarded. Given that importance of the principal components or factorial axes decreases from the first to the last, this one explains usually a very small portion of the total variance. For this, the descriptor with the highest coefficient on the principal axis with the smallest eigenvalue (the last axis) can be discarded (Pereira, 1989). For each discarding, another analysis is carried out with the remaining descriptors, and thus successively until the ranking of the descriptors is established by their importance. Fifty-five analyses (*p*-1) were carried out using the *corresp* procedure from the SAS computation system (SAS Institute, 2009).

Based on the results of these analyses, three selection criteria for minimum descriptors were assessed: C1 – selection according the proposal by Jollife (1973) method, based on the inverse order of the descriptor for the p^{th} (last) factorial axis (O'_p); C2 - selection based on the mean of the contribution orders of the descriptor (O_s) on the first three factorial axes in the complete analysis, and O'_p, so that $O_s = (O_1 + O_2 + O_3 + O'_p)/4$; and C3 - selection based on the weighted mean of the contribution orders of the descriptor (O_z) for the three first axes, with weights (w_1, w_2 and w_3) defined by the respective eigenvalues in the complete analysis: $O_z = (w_1O_1 + w_2O_2 + w_3O_3)/(w_1+w_2+w_3)$.

Before applying and assessing the three criteria, uniform size was established for the list of minimum descriptors to be constituted within each group (complete collection or subcollections). This size (around thirty descriptors) was defined having as target a correlation equal or greater to 0.90 between the dissimilarity matrices of accessions estimated with all the descriptors (M) or only with the minimum descriptors (E). The dissimilarity index proposed by Cruz & Carneiro (2006) was used for multi-categorical variables: [$d_{ii'} = D/(C+D)$], where *i* and *i'* correspond to a pair of accessions (*i* e *i'*= 1, 2, ..., *n*), C is the number of category agreements, and D is the number of disagreements. This was performed in the Genes computer program (Cruz, 2006).

The correlation between M and E dissimilarity matrices, for each germplasm group, taking account the different criteria, was used as the measure of the quality of the minimum descriptor lists. The best criterion, chosen to establish the definitive list of minimum descriptors, was that which resulted in the highest correlation between these matrices. The significance of these correlations was assessed by the Mantell test, carried out on the NTSys software, with 10,000 randomizations (Rohlf, 2005).

RESULTS AND DISCUSSION

***Capsicum* spp. -** Table 2 shows the ranking of the descriptors for the whole germplasm collection, according to their contributions to the last factorial axis in the successive analyses (MCA). It was observed in the analysis with all the descriptors that the variable "species" (Esp) was that most contributed to explain this factor axis and, in principle, it would be of less importance for the characterization of these accessions. In theory, it means that the information associated to this variable is already contemplated by other variables that determine the previous axes. And in this case, this certainly happened because the identification of the species is related mainly to differences in flower characteristics (Moscone *et al*., 2007; Ince *et al*., 2009).

On the other hand, this variable (Esp) showed the second highest contribution to the first factorial axis, and the greatest contribution to the second and third axes. Partial results such as these are not presented here, but are available in Silva (2008), and also for the four subcollections. This same behavior was observed in other variables: fruit weight (PFt), with the greatest contribution to the first factorial axis; corolla colour (CrC) - fourth greatest contribution to the first axis; calyx annular constriction (CCl) - second greatest contribution to the second axis; and corolla spot colour (CCr) - second greatest contribution to the third axis. Because of this it was decided to also consider the participation of the descriptors in the first three factorial axes (criteria C2 and C3), instead of discarding variables only with base on the C1 criterion (Jollife, 1973).

Assessment of the contributions of the descriptors to the first three factorial axes in the complete MCA showed that about 50% of the descriptors with the greatest contributions explained 90% or more of the total variability associated to each axis (results available in Silva, 2008). This is only possible because of the considerable amount of redundancy or association among the variables; that

Table 1. List of descriptors adopted by Embrapa Hortaliças for germplasm of *Capsicum* spp. with their number of categories (*K*) and the variable code (lista de descritores adotada pela Embrapa Hortaliças para germoplasma de *Capsicum* spp. com os respectivos número de categorias (*K*) e código da variável). Brasília, Embrapa Hortaliças, 2005.

Descriptors	*K*	Code	Descriptors	*K*	Code
Nodal anthocyanin	4	AtN	Stem shape	3	FHt
Fruit blossom end appendage	2	APF	Fruit shape at blossom end	4	FPF
Aroma	3	Arm	Fruit shape	6	FFt
Tillering	4	BtF	Plant growth habit	4	Hbt
Stem length	4	CHt	Plant canopy width	4	LPt
Placenta length	3	CPl	Fruit width	5	LFt
Plant height	5	CPt	Male-sterility	2	McE
Fruit length	6	CFt	Calyx margin	3	MgC
Fruit pedicel length	4	CpP	Varietal mixture condition	2	MVr
Calyx annular constriction	2	CCl	Number of flowers per axil	8	NFA
Anther colour	6	CAt	Number of locules	5	NLl
Corolla colour	11	CrC	Number of seeds per fruit	3	NSF
Leaf colour	8	CrF	Fruit shape at pedicel attachment	5	OFt
Stem colour	3	CrH	Ripe fruit persistence	3	PtF
Corolla spot colour	6	CCr	Neck at base of fruit	2	PcF
Seed colour	4	CSt	Fruit weight	6	PFt
Filament colour	7	CFn	Calyx pigmentation	2	Pgl
Immature fruit colour	10	CFI	Flower position	7	PFr
Mature fruit colour	14	CFM	Stigma exsertion	6	PEt
Leaf density	3	DsF	Fruit position	7	PsF
Branching habit	3	DsR	Origin	12	Pro
Stem diameter	4	DHt	Leaf pubescence	3	PbF
Days to flowering	4	DFl	Stem pubescence	3	PHt
Days to fruiting	4	DFt	Pungency	4	Pug
Species	4	Esp	Fruit cross-sectional corrugation	3	STv
Fruit wall thickness	6	EFt	Segregation	7	Seg
Corolla shape	3	FmC	Seed surface	3	SSt
Leaf shape	3	FmF	Fruit surface	5	SFt

is, some descriptors capture information of the same nature about the variability of the accessions. However, for the purpose of the study, this redundancy was the guarantee that the number of descriptors needed to characterize the germplasm could be reduced without great harm to the representation of their genetic variability. In this case, it was observed that the accumulated contribution of the first variable (PFt) up to the 28th (PFr), on the first factorial axis, was 92.5% (Silva, 2008). Therefore to guarantee minimum explanation of 90% of the variability captured by the first and main spectre of the factorial analysis, about thirty variables should be selected as minimum descriptors.

The ranking of the descriptors according to the three selection criteria (C1, C2 and C3) for this collection is shown in Table 3. The best descriptive ability is associated to the descriptor with the lowest of these order. Taking account the thirty best classified descriptors in each criterion, significant correlations ($p<0.01$) among the matrices M (estimated with all the descriptors) and E (estimated only with the minimum descriptors) were observed, with magnitude of 0.80 (C1), 0.89 (C2) and 0.87 (C3). Therefore, the criterion C2 was chosen to select the minimum descriptors.

The definitive list of the minimum descriptors for *Capsicum* spp. was defined as: species (Esp), fruit wall thickness (EFt), fruit width (LFt), number of flowers per axil (NFA), fruit weight (PFt), fruit position (PsF), origin (Pro), ripe fruit persistence (PtF), calyx margin (MgC), fruit shape (FFt), fruit shape at pedicel attachment (OFt), number of seeds per fruit (NSF), anther colour (CAt), number of locules (NLl), filament colour (CFn), placenta length (CPl), corolla colour (CrC), mature fruit colour (CFM), fruit shape at blossom end (FPF), leaf shape (FmF), pungency (Pug), fruit pedicel length (CpP), plant canopy width (LPt), stigma exsertion (PEt), fruit cross-sectional corrugation (STv), aroma (Arm), days to fruiting (DFt), fruit surface (SFt), nodal anthocyanin (AtN), plant growth habit (Hbt) and flower position (PFr).

Regarding the presence of the variable "species" (Esp) among the minimal descriptors of this collection, it is needed to point out, firstly, that if the accession is a cultivated material or if the species has already been identified, its characterization can be realized directly with base on the list of minimum descriptors of its species. Otherwise, the inclusion of this variable should be interpreted as the need to maintain, among the minimum descriptiors of a generic collection, all the morphological attributes involved in the identification key of species in *Capsicum* spp. This naturally is irrelevant within the specific subcollections.

Even though the variable "origin" (Pro) is not an intrinsic attribute to the morphology of the accessions, although it is important passport data, its inclusion in this list shows that, regardless of the species, the geographic origin of the germplasm is directly associated to their genetic divergence. As will be shown further, this fact was also observed in the study of the subcollections. According to Galwey (1995), good passport data (e.g. geographic origin) are probably unrivaled in supplying concise and inexpensive information about the structure of the genetic variability of a germplasm collection.

The analysis of this collection further highlights the high proportion of minimum descriptors related to the fruit characteristics. This is interesting

Table 2. Descending order of morphological descriptors (Var) to accessions of *Capsicum* spp. after sucessive multiple correspondence analyses (MCA)[1], and their maximum contribution percentage (%) for the last factorial axis in each analysis (ordenamento decrescente dos descritores morfológicos (Var) para acessos de *Capsicum* spp. após análises de correspondência múltipla (ACM) sucessivas[1], e respectivas contribuições percentuais máximas (%) para o último eixo fatorial em cada análise). Brasília, Embrapa Hortaliças, 2005.

ACM	Var	%	ACM	Var	%	ACM	Var	%	ACM	Var	%
1	Esp	61.84	15	CAt	15.13	29	AtN	20.61	43	FmF	39.88
2	DHt	58.19	16	FFt	23.24	30	DFl	31.51	44	SFt	24.19
3	CrC	57.74	17	PFr	25.62	31	MgC	31.86	45	PHt	33.50
4	CCl	30.52	18	CFM	33.03	32	CFI	30.70	46	FHt	25.48
5	PFt	56.68	19	CrF	28.83	33	Hbt	27.06	47	CpP	46.79
6	CCr	53.21	20	LPt	36.43	34	FPF	46.16	48	Seg	49.56
7	FmC	53.24	21	STv	25.60	35	BtF	30.89	49	DsF	43.24
8	CFt	35.43	22	DsR	29.61	36	Arm	31.11	50	PcF	27.57
9	LFt	69.85	23	Pro	31.23	37	CPl	28.69	51	APF	26.90
10	Oft	30.84	24	Pug	44.60	38	CFn	19.03	52	MVr	33.09
11	NFA	60.93	25	NLl	29.71	39	CHt	23.82	53	Pgl	40.45
12	PbF	19.00	26	NSF	27.61	40	CrH	19.10	54	SSt	50.00
13	EFt	29.52	27	CPt	19.03	41	PEt	33.14	–	CSt	50.00
14	DFt	19.43	28	PsF	14.89	42	PtF	23.28	–	McE	0.00

[1]Analyses where was eliminated, in each cycle, the variable with the largest contribution to the last factorial axis of the previous analysis (análises em que se eliminava, em cada ciclo, a variável com a maior contribuição para o último eixo fatorial da análise anterior).

because they are traits commonly evaluated in the chilli and sweet pepper consumer markets (Onoyama *et al.*, 2010). Thus, the characterization of accessions by these descriptors is directly linked to the interests of the *Capsicum* breeders (e.g. identification of potential parents for their crossing programs). Furthermore, this corroborates the results obtained by Teixeira (1996) and Sudré (2003), who emphasized length, diameter and mean weight of the fruit as the most important descriptors for genetic divergence in *Capsicum*.

***Capsicum annuum* -** Considering only the *C. annuum* accessions, the variable with greatest contribution to the last factorial axis in the analysis with all the descriptors was the corolla colour (CrC). Thus, according criterium C1 (Jollife, 1973), this variable could

Table 3. Ascending order of the descriptors (Var) used to the characterization of accessions of *Capsicum* spp. from the Embrapa Horticultura germplasm bank, obtained from the application of three criteria (C1, C2 and C3)[1] for the choice of the list of minimum descriptors (ordem crescente dos descritores (Var) usados para a caracterização de acessos de *Capsicum* spp. do banco de germoplasma da Embrapa Hortaliças, obtida a partir da aplicação de três critérios (C1, C2 e C3)[1] para a escolha da lista de descritores mínimos). Brasília, Embrapa Hortaliças, 2005.

C1		C2		C3		C1		C2		C3	
Var	O'_p	Var	O_s	Var	Oz	Var	O'_p	Var	O_s	Var	Oz
CSt	1.50	Esp	14.75	Esp	1.48	CPt	29.00	AtN	27.38	AtN	28.73
SSt	1.50	EFt	16.00	PFt	6.10	NSF	30.00	Hbt	28.00	Hbt	29.51
Pgl	3.00	LFt	17.88	EFt	7.22	NLl	31.00	PFr	28.50	BtF	29.94
MVr	4.00	NFA	18.13	LFt	7.24	Pug	32.00	CCr	29.13	FmF	30.69
APF	5.00	PFt	18.25	NFA	7.76	Pro	33.00	CFt	29.50	Pet	31.90
PCF	6.00	PsF	18.25	CrC	9.85	DsR	34.00	CHt	29.88	Dfl	32.01
DsF	7.00	Pro	18.50	Oft	10.81	STv	35.00	CCl	30.00	SFt	32.84
Seg	8.00	PtF	19.38	Pro	13.75	LPt	36.00	APF	30.50	FmC	33.75
CpP	9.00	MgC	19.50	NSF	14.07	CrF	37.00	BtF	31.25	DsR	33.91
FHt	10.00	FFt	19.75	FFt	15.10	CFM	38.00	DFl	31.75	CHt	34.28
PHt	11.00	Oft	19.75	PsF	15.75	PFr	39.00	CrH	32.75	CPF	34.66
SFt	12.00	NSF	20.13	CAt	16.74	FFt	40.00	DsF	33.13	CPt	35.41
FmF	13.00	CAt	21.25	MgC	18.39	CAt	41.00	Seg	34.50	CrF	38.21
PtF	14.00	NLl	21.25	CFM	18.46	DFt	42.00	CFI	35.00	CrH	38.97
Pet	15.00	CFn	22.00	CCL	19.08	EFt	43.00	FHt	35.00	CFI	39.35
CrH	16.00	CPl	22.38	Dft	19.12	PbF	44.00	PcF	35.00	DsF	42.19
CHt	17.00	CrC	22.38	Pug	19.23	NFA	45.00	CPt	35.75	PbF	43.01
CFn	18.00	CFM	22.50	CFt	19.24	OFt	46.00	FmC	36.25	Seg	43.52
CPl	19.00	FPF	22.75	LPt	20.20	LFt	47.00	DsR	36.38	APF	43.93
Arm	20.00	FmF	24.13	NLL	20.75	CFt	48.00	CrF	36.50	FHt	44.24
BtF	21.00	Pug	24.50	PtF	20.92	FmC	49.00	CSt	38.75	DHt	44.30
FPF	22.00	CpP	25.50	CPl	21.68	CCr	50.00	Pgl	38.88	PCF	45.74
Hbt	23.00	LPt	25.75	CFn	22.58	PFt	51.00	SSt	40.00	CSt	50.64
CFI	24.00	Pet	25.75	FPF	23.11	CCl	52.00	MVr	40.25	PHt	51.22
MgC	25.00	STv	26.13	PFr	24.50	CrC	53.00	PHt	40.88	Pgl	51.64
DFl	26.00	Arm	26.50	STv	25.08	DHt	54.00	PbF	42.25	MVr	51.71
AtN	27.00	DFt	26.50	Arm	25.13	Esp	55.00	DHt	46.88	SSt	53.04
PsF	28.00	SFt	27.13	CCR	28.44	McE	56.00	McE	55.38	McE	54.88

[1]C1: reverse order of the contribution of the descriptor for the last factorial axis (O'_p), obtained from successive multiple correspondence analyses (MCA), where is eliminated in each analysis the variable with the largest contribution to this axis; C2: average order (O_s) of the contributions of the descriptor in the first three factorial axes (O_1, O_2 and O_3, respectively) of the MCA with all the descriptors, and of the order O'_p already defined, such that $O_s = (O_1 + O_2 + O_3 + O'_p)/4$; C3: average order ($O_z$) of the first three factorial axes of the MCA complete, weighted by contributions (w_1, w_2 and w_3) of these axes, such that $O_z = (w_1O_1 + w_2O_2 + w_3O_3)/(w_1+w_2+w_3)$ [C1: ordem inversa da contribuição do descritor para o último eixo fatorial (O'_p), obtida a partir de análises de correspondência múltipla (ACM) sequenciais, em que se elimina, em cada uma, a variável com a maior contribuição para este eixo; C2: ordem média (O_s) das contribuições do descritor nos três primeiros eixos fatoriais (O_1, O_2 e O_3, respectivamente) da ACM com todos os descritores, e da ordem O'_p já definida, tal que $O_s = (O_1 + O_2 + O_3 + O'_p)/4$; C3: ordem média ($O_z$) dos três primeiros eixos fatoriais da análise ACM completa, ponderada pelas contribuições (w_1, w_2 e w_3) desses eixos, tal que $O_z = (w_1O_1 + w_2O_2 + w_3O_3)/(w_1+w_2+w_3)$].

be discarded. However, as already discussed, this descriptor also showed the fourth greatest contribution to the second factorial axis, and the greatest to the third axis. This fact reiterates the need to aggregate information on the contribution of the descriptors in the first axes of the MCA.

Some descriptors did not contribute to the factorial axes, that is, they were invariant in this germplasm group: calyx annular constriction (CCl), male-sterility (McE), neck at base of fruit (PcF) and, obviously, species (Esp). Thus they were discarded. In a similar case, Simpson *et al*. (2002) highlighted the trait leaf colour when separating cultivated peanut accessions, because the descriptor was invariable in 681 accessions. However, it should be emphasized that in this type of study commonly the relative contribution of each factorial axis is fairly small, bearing in mind the great number of variables and categories per variable.

The variables that most contributed to the first factorial axis of the analysis with all the descriptors in this group of accessions were: fruit weight (PFt), fruit width (LFt), fruit wall thickness (EFt), fruit shape at pedicel attachment (OFt), pungency (Pug), placenta length (CPl), fruit shape (FFt), number of locules (NLl), branching habit (DsR), and fruit shape at blossom end (FPF). It was again observed that most of these were related to fruit characteristics, which was an interesting fact because these traits are used in breeding of the species, especially in the sweet pepper group. This may have resulted further from the collect and domestication processes that overvalue this type of characteristic. The same tendency was observed in the second factorial axis, with greatest contribution from fruit weight (PFt), fruit length (CFt), fruit width (LFt), corolla colour (CrC), flower position (PFr), fruit position (PsF), and fruit wall thickness (EFt). In the *Capsicum* genus, fruit shape, and immature and mature fruit colour are important traits for the consumer market (Onoyama *et al.*, 2010; Sudré *et al.*, 2006).

It was also observed in this group (*C. annuum*) that 50% of the descriptors explained almost 90% of the variability captured by the first factorial axis (Table 4); that is, the accumulated contribution of the first variable up to the 28th variable corresponded to approximately 90%. Thus selecting at least the first 29 descriptors (about thirty variables), guarantees again this minimal explanation of the variability associated to the first principal axis of the analysis.

There was significant correlation ($p<0.01$) among M and E dissimilarity matrices for all criteria of descriptor selection (0.84 for C1 criterion, and 0.91 for C2 and C3). The C2 criterion was chosen because of its better performance also in the whole collection study. The list of minimum descriptors for this group was defined as: fruit length (CFt), calyx pigmentation (Pgl), fruit wall thickness (EFt), fruit width (LFt), origin (Pro), fruit weight (PFt), stem colour (CrH), corolla colour (CrC), immature fruit colour (CFI), plant growth habit (Hbt), filament colour (CFn), stem length (CHt), anther colour (CAt), fruit shape at pedicel attachment (OFt), fruit shape at blossom end (FPF), tillering (BtF), fruit position (PsF), leaf density (DsF), leaf colour (CrF), fruit shape (FFt), number of seeds per fruit (NSF), pungency (Pug), fruit surface (SFt), calyx margin (MgC), mature fruit colour (CFM), plant height (CPt), flower position (PFr), number of locules (NLl), stem pubescence (PHt), and plant canopy width (LPt). In this case, eleven of these are also fruit characteristics. Immature fruit colour, fruit shape, and length and width of the fruit are related mainly to fresh fruit commercialization; wall thickness and number of seeds per fruit are important traits when selling the dried product; and mature fruit colour is interesting in both forms of commercialization.

***Capsicum baccatum* -** Observation of the *C. baccatum* accessions showed that some descriptors had no descriptive importance (invariant) for the characterization of the species: male sterility (McE), calyx annular constriction (CCl), seed colour (CSt), seed surface (SSt), corolla spot colour (CCr) and, obviously, species (Esp). Thus these variables should not be part of the respective descriptor list. In this group, the variables that most contributed to the first factorial axis in the complete analysis were: fruit weight (PFt), fruit length (CFt), ripe fruit persistence (PtF), fruit wall thickness (EFt), number of seeds per fruit (NSF), fruit width (LFt), days to fruiting (DFt), leaf pubescence (PbF), nodal anthocyanin (AtN), and number of flowers per axil (NFA). Also in this case, fruit-related descriptors presented greater variability that is a marked characteristic in the *Capsicum* genus.

From the ranking of the descriptors according to the three criteria (Table 5) and the respective lists of selected descriptors, the criterion which generated the highest correlation (0.94) between M and E matrices was again C2. The C1 and C3 criteria produced the correlations 0.77 and 0.64, respectively. All these correlations were significant ($p<0.01$). Thus, the minimum descriptor list was defined by C2 criterion as: fruit shape at blossom end (FPF), fruit position (PsF), fruit length (CFt), fruit weight (PFt), fruit shape (FFt), fruit wall thickness (EFt), fruit surface (SFt), corolla colour (CrC), stem length (CHt), plant height (CPt), number of locules (NLl), immature fruit colour (CFI), origin (Pro), fruit pedicel length (CpP), pungency (Pug), days to flowering (DFl), number of flowers per axil (NFA), tillering (BtF), filament colour (CFn), number of seeds per fruit (NSF), fruit width (LFt), leaf pubescence (PbF), stigma exsertion (PEt), stem pubescence (PHt), fruit shape at pedicel attachment (OFt), placenta length (CPl), mature fruit colour (CFM), days to fruiting (DFt), branching habit (DsR), and plant growth habit (Hbt).

As observed in the previous analyses, twelve of the thirty descriptors selected are fruit characteristics. This corroborates the importance of these traits to discriminate accessions in germplasm bank of *Capsicum*. Thus maintaining these descriptors not only guarantees parsimony in the germplasm characterization but also supplies information of immediate interest to the breeders, because it is linked to the market demands regarding the cultivars and commercial products derived from these peppers (Sudré *et al.*, 2006).

Capsicum chinense **-** Table 6 shows the ranking of the variables according to the criteria for the composition of the minimal descriptor list for *C. chinense*. Choosing thirty variables according to each criterion, the correlations between M and E matrices, all significant ($p<0.01$) were 0.82 for C1, 0.88 for C2, and 0.88 for C3. Thus, once again, the C2 and C3 criteria were outstanding. Due to the same reasons previously reported, the following descriptors were chosen to compose the list according C2 criterion: flower position (PFr), number of locules (NLl), fruit pedicel length

Table 4. Ascending order of the descriptors (Var) used to the characterization of accessions of *Capsicum annuum* from the Embrapa Hortaliças germplasm bank, obtained from the application of three criteria (C1, C2 and C3)[1] to choose the list of minimum descriptors (ordem crescente dos descritores (Var) usados para a caracterização de acessos de *Capsicum annuum* do banco de germoplasma da Embrapa Hortaliças, obtida a partir da aplicação de três critérios (C1, C2 e C3)[1] para escolha da lista de descritores mínimos). Brasília, Embrapa Hortaliças, 2005.

C1		C2		C3		C1		C2		C3	
Var	O'_p	Var	O_s	Var	Oz	Var	O'_p	Var	O_s	Var	Oz
CCr	1.50	CFt	13.50	PFt	3.94	CFt	29.00	PHt	26.00	NSF	27.17
Pgl	1.50	Pgl	14.63	LFt	4.55	Pro	30.00	LPt	26.13	STv	27.38
PHt	3.00	EFt	15.00	EFt	5.96	CPl	31.00	Pet	26.38	CHt	27.50
CpP	4.00	LFt	15.00	CrC	9.77	LPt	32.00	DHt	26.50	Pet	28.97
NSF	5.00	Pro	15.00	CFt	9.92	FPF	33.00	Arm	26.75	CFM	29.04
DHt	6.00	PFt	15.50	Pro	10.39	CFI	34.00	DFt	26.75	PFr	29.48
PEt	7.00	CrH	17.00	Oft	11.89	Hbt	35.00	DsR	27.50	PHt	32.54
DsF	8.00	CrC	17.38	FFt	12.52	DFt	36.00	STv	28.13	PbF	33.93
FmF	9.00	CFI	18.75	Hbt	14.91	Pug	37.00	DFl	28.50	Arm	34.06
DFl	10.00	Hbt	18.75	Pug	15.08	DsR	38.00	FmF	29.13	Seg	34.97
Arm	11.00	CFn	19.88	PsF	15.31	EFt	39.00	CPl	29.75	DHt	35.63
CHt	12.00	CHt	20.63	FPF	15.60	PbF	40.00	Seg	29.75	FmF	36.65
Seg	13.00	Cat	21.00	CFI	16.89	OFt	41.00	CCr	32.38	Dfl	37.80
AtN	14.00	Oft	21.00	CrH	17.30	CrF	42.00	CpP	34.13	PtF	39.32
APF	15.00	FPF	21.25	CFn	18.52	LFt	43.00	AtN	35.38	MVr	40.75
PtF	16.00	BtF	21.50	CrF	19.28	PsF	44.00	PtF	35.63	FHt	41.87
PFr	17.00	PsF	21.50	BtF	19.97	FFt	45.00	PbF	36.00	CPF	41.98
STv	18.00	DsF	21.63	DsR	20.12	PFt	46.00	APF	38.25	AtN	44.27
CFM	19.00	CrF	21.75	MgC	20.36	CrC	47.00	NFA	39.25	CCR	44.77
BtF	20.00	FFt	22.25	CAt	20.67	FHt	50.00	MVr	41.75	NFA	44.82
MgC	21.00	NSF	22.25	NLL	20.69	FmC	50.00	FHt	44.50	FmC	46.48
SFt	22.00	Pug	22.75	SFt	22.73	MVr	50.00	FmC	47.25	APF	47.23
NFA	23.00	SFt	22.88	Pgl	23.01	CSt	50.00	CSt	50.88	CSt	50.79
CrH	24.00	MgC	23.00	DsF	23.55	SSt	50.00	SSt	50.88	SSt	51.83
CPt	25.00	CFM	23.88	CPl	23.78	Esp	54.50	CCl	54.00	CCL	53.89
CAt	26.00	CPt	24.00	LPt	23.99	McE	54.50	Esp	54.00	Esp	53.89
CFn	27.00	PFr	25.38	CPt	25.15	CCl	54.50	McE	54.00	McE	53.89
NLl	28.00	NLl	25.50	Dft	25.38	PcF	54.50	PcF	54.00	PCF	53.89

[1]C1: reverse order of the contribution of the descriptor for the last factorial axis (O'_p), obtained from successive multiple correspondence analyses (MCA) where is eliminated in each analysis the variable with the largest contribution to this axis; C2: average order (O_s) of the contributions of the descriptor in the first three factorial axes (O_1, O_2 and O_3, respectively) of the MCA with all the descriptors, and of the order O'_p already defined, such that $O_s = (O_1 + O_2 + O_3 + O'_p)/4$; C3: average order ($O_z$) of the first three factorial axes of the MCA complete, weighted by contributions (w_1, w_2 and w_3) of these axes, such that $O_z = (w_1O_1 + w_2O_2 + w_3O_3)/(w_1+w_2+w_3)$ [C1: ordem inversa da contribuição do descritor para o último eixo fatorial (O'_p), obtida a partir de análises de correspondência múltipla (ACM) sequenciais, em que se elimina, em cada uma, a variável com a maior contribuição para este eixo; C2: ordem média (O_s) das contribuições do descritor nos três primeiros eixos fatoriais (O_1, O_2 e O_3, respectivamente) da ACM com todos os descritores, e da ordem O'_p já definida, tal que $O_s = (O_1 + O_2 + O_3 + O'_p)/4$; C3: ordem média ($O_z$) dos três primeiros eixos fatoriais da análise ACM completa, ponderada pelas contribuições (w_1, w_2 e w_3) desses eixos, tal que $O_z = (w_1O_1 + w_2O_2 + w_3O_3)/(w_1+w_2+w_3)$].

(CpP), mature fruit colour (CFM), fruit surface (SFt), anther colour (CAt), fruit position (PsF), fruit length (CFt), fruit wall thickness (EFt), fruit shape at pedicel attachment (OFt), fruit shape (FFt), origin (Pro), fruit weight (PFt), number of seeds per fruit (NSF), stem shape (FHt), days to fruiting (DFt), fruit width (LFt), ripe fruit persistence (PtF), days to flowering (DFl), calyx margin (MgC), leaf pubescence (PbF), plant canopy width (LPt), fruit shape at blossom end (FPF), stem diameter (DHt), plant height (CPt), leaf shape (FmF), immature fruit colour (CFI), stem length (CHt), filament colour (CFn), placenta length (CPl), corolla shape (FmC), and pungency (Pug).

Table 5. Ascending order of the descriptors (Var) used to the characterization of accessions of *Capsicum baccatum* from the Embrapa Hortaliças germplasm bank, obtained from the application of three criteria (C1, C2 and C3)[1] to choose the list of minimum descriptors (ordem crescente dos descritores (Var) usados para a caracterização de acessos de *Capsicum baccatum* do Banco de Germoplasma da Embrapa Hortaliças, obtida a partir da aplicação de três critérios (C1, C2 e C3)[1] para escolha da lista de descritores mínimos). Brasília, Embrapa Hortaliças, 2005.

C1		C2		C3		C1		C2		C3	
Var	O'_p	Var	O_s	Var	Oz	Var	O'_p	Var	O_s	Var	Oz
CPl	1.50	FPF	12.75	CFt	3.64	CrH	29.00	DsR	25.38	CPt	27.61
Pgl	1.50	PsF	13.38	PFt	4.16	PEt	30.00	Hbt	25.75	PHt	27.80
APF	3.00	CFt	13.50	EFt	5.79	Pug	31.00	MgC	26.25	CAt	28.37
FmF	4.00	PFt	14.75	FFt	9.45	DsR	32.00	STv	26.88	CFI	28.84
PHt	5.00	FFt	15.50	PsF	9.81	NFA	33.00	AtN	27.25	BtF	30.58
MVr	6.00	EFt	16.38	LFt	11.62	PtF	34.00	PtF	28.13	DsF	31.38
CrF	7.00	SFt	17.00	FPF	12.18	NSF	35.00	APF	28.75	CFM	31.59
Seg	8.00	CrC	20.38	NSF	14.45	FFt	36.00	FmF	29.75	CrH	34.74
Arm	9.00	CHt	20.75	PbF	15.81	Pro	37.00	Pgl	30.38	PFr	35.15
CPt	10.00	CPt	20.75	CFn	16.17	PbF	38.00	PFr	30.75	CPl	35.36
FmC	11.00	NLl	20.75	Dft	16.44	DFt	39.00	CrF	31.13	DHt	35.81
BtF	12.00	CFI	21.63	NFA	16.59	DsF	40.00	Arm	32.00	MgC	36.59
MgC	13.00	Pro	22.25	Hbt	17.61	CFt	41.00	CAt	32.63	Arm	37.07
PcF	14.00	CpP	22.63	SFt	18.34	PFt	42.00	FmC	33.00	FmF	37.27
CrC	15.00	Pug	22.75	NLL	18.89	STv	43.00	LPt	33.00	APF	37.44
FHt	16.00	DFl	23.00	PtF	19.03	EFt	44.00	DsF	33.13	MVr	39.50
DFl	17.00	NFA	23.00	Pug	19.28	CFn	45.00	MVr	33.75	CrF	40.59
CFM	18.00	BtF	23.13	CrC	19.32	Hbt	46.00	DHt	33.88	FmC	42.86
NLl	19.00	CFn	23.25	Pro	19.75	OFt	47.00	Seg	36.38	Pgl	43.52
SFt	20.00	NSF	23.50	Pet	21.44	CAt	48.00	PcF	37.25	PCF	45.23
CHt	21.00	LFt	23.63	AtN	21.73	LFt	49.00	CrH	37.38	Seg	45.38
CFI	22.00	PbF	24.00	CPF	22.52	LPt	50.00	FHt	38.00	FHt	45.84
FPF	23.00	PEt	24.25	Dfl	22.63	Esp	53.50	CCl	53.00	CCL	53.05
CpP	24.00	PHt	24.25	CHt	22.68	CCr	53.50	CCr	53.00	CCR	53.05
AtN	25.00	OFt	24.38	STv	23.74	McE	53.50	CSt	53.00	CSt	53.05
PFr	26.00	CPl	25.13	Oft	23.95	CCl	53.50	Esp	53.00	Esp	53.05
PsF	27.00	CFM	25.25	DsR	24.99	CSt	53.50	McE	53.00	McE	53.05
DHt	28.00	DFt	25.38	LPt	27.18	SSt	53.50	SSt	53.00	SSt	53.05

[1]C1: reverse order of the contribution of the descriptor for the last factorial axis (O'_p), obtained from successive multiple correspondence analyses (MCA), where is eliminated in each analysis the variable with the largest contribution to this axis; C2: average order (O_s) of the contributions of the descriptor in the first three factorial axes (O_1, O_2 and O_3, respectively) of the MCA with all the descriptors, and of the order O'_p already defined, such that $O_s = (O_1 + O_2 + O_3 + O'_p)/4$; C3: average order ($O_z$) of the first three factorial axes of the MCA complete, weighted by contributions (w_1, w_2 and w_3) of these axes, such that $O_z = (w_1O_1 + w_2O_2 + w_3O_3)/(w_1+w_2+w_3)$ [C1: ordem inversa da contribuição do descritor para o último eixo fatorial (O'_p), obtida a partir de análises de correspondência múltipla (ACM) sequenciais, em que se elimina, em cada uma, a variável com a maior contribuição para este eixo; C2: Ordem média (O_s) das contribuições do descritor nos três primeiros eixos fatoriais (O_1, O_2 e O_3, respectivamente) da ACM com todos os descritores, e da ordem O'_p já definida, tal que $O_s = (O_1 + O_2 + O_3 + O'_p)/4$; C3: ordem média ($O_z$) dos três primeiros eixos fatoriais da análise ACM completa, ponderada pelas contribuições (w_1, w_2 e w_3) desses eixos, tal que $O_z = (w_1O_1 + w_2O_2 + w_3O_3)/(w_1+w_2+w_3)$].

Again, of the 32 descriptors selected twelve are fruit traits, corroborating the previous results from *C. annuum* and *C. baccatum*, and those reported by Teixeira (1996) and Sudré (2003).

Capsicum frutescens - Table 7 shows the rankings of the variables in this germplasm group, according to the criteria for establishment of the minimum descriptors. The correlations between M and E matrices, all significant ($p<0.01$), were: 0.83 for C1, 0.89 for C2, and 0.92 for C3. Thus, in this case the advantage was in favor of C3 criterion, that resulted in the following list of minimal descriptors: mature fruit colour (CFM), nodal anthocyanin (AtN), fruit position (PsF), origin (Pro), days to

Table 6. Ascending order of the descriptors (Var) used to the characterization of accessions of *Capsicum chinense* from the Embrapa Hortaliças germplasm bank, obtained from the application of three criteria (C1, C2 and C3)[1] to choose the list of minimum descriptors (ordem crescente dos descritores (Var) usados para a caracterização de acessos de *Capsicum chinense* do Banco de Germoplasma da Embrapa Hortaliças, obtida a partir da aplicação de três critérios (C1, C2 e C3)[1] para escolha da lista de descritores mínimos). Brasília, Embrapa Hortaliças, 2005.

C1		C2		C3		C1		C2		C3	
Var	O'_p	Var	O_s	Var	Oz	Var	O'_p	Var	O_s	Var	Oz
CpP	1.50	PFr	10.75	CFM	5.15	DsR	29.00	CFn	28.38	CHt	28.54
APF	1.50	NLl	12.50	PFr	5.30	NLl	30.00	CPl	28.50	Hbt	28.62
PcF	3.00	CpP	12.88	CFt	5.83	CrH	31.00	FmC	28.50	DHt	30.79
CCl	4.00	CFM	15.75	NLL	7.06	CrC	32.00	Pug	28.50	PbF	31.35
MVr	5.00	SFt	15.75	Pro	9.35	PsF	33.00	STv	28.75	PHt	31.74
CFn	6.00	CAt	16.00	PFt	9.70	OFt	34.00	Arm	29.13	CrF	33.53
CPt	7.00	PsF	16.25	PsF	10.21	CrF	35.00	Hbt	29.25	CrH	33.82
BtF	8.00	CFt	16.50	EFt	10.41	EFt	36.00	PcF	31.25	Pet	34.51
Arm	9.00	EFt	17.00	FFt	12.25	FFt	37.00	PEt	31.75	Arm	34.66
FmC	10.00	OFt	17.88	Oft	12.62	LPt	38.00	Seg	31.75	AtN	34.73
PbF	11.00	FFt	18.00	LFt	13.03	Pug	39.00	DsF	32.00	CPt	34.95
PtF	12.00	Pro	18.75	CAt	13.24	Hbt	40.00	AtN	32.25	CFn	35.04
DHt	13.00	PFt	19.00	Dfl	15.98	CFI	41.00	BtF	32.50	FmC	36.02
Seg	14.00	NSF	19.13	CPF	16.42	STv	42.00	CCl	32.63	Seg	37.40
DsF	15.00	FHt	20.25	SFt	16.44	DFl	43.00	NFA	33.38	DsF	38.11
SFt	16.00	DFt	21.00	NSF	19.11	PFt	44.00	CrF	33.50	CrC	38.46
NSF	17.00	LFt	22.00	Dft	19.92	LFt	45.00	CrH	34.13	DsR	39.19
FHt	18.00	PtF	22.13	FHt	21.66	CCr	46.00	PHt	35.00	PCF	40.07
AtN	19.00	DFl	23.00	STv	21.90	CFt	47.00	DsR	35.38	CCL	40.96
MgC	20.00	MgC	23.25	LPt	22.67	PHt	48.00	CrC	36.50	BtF	41.23
FmF	21.00	PbF	25.00	CFI	22.77	NFA	49.00	MVr	36.50	MVr	46.79
FPF	22.00	LPt	25.63	Pug	22.97	CFM	50.00	APF	38.75	CCR	48.13
DFt	23.00	FPF	26.00	MgC	26.14	Pro	51.00	CCr	47.50	APF	51.19
CAt	24.00	DHt	26.50	CPl	26.57	CSt	54.00	CSt	53.88	CSt	53.85
PEt	25.00	CPt	26.88	PtF	26.73	Esp	54.00	Esp	53.88	Esp	53.85
CPl	26.00	FmF	27.25	NFA	27.56	McE	54.00	McE	53.88	McE	53.85
PFr	27.00	CFI	27.75	FPF	27.60	Pgl	54.00	Pgl	53.88	Pgl	53.85
CHt	28.00	CHt	28.13	FmF	28.33	SSt	54.00	SSt	53.88	SSt	53.85

[1]C1: reverse order of the contribution of the descriptor for the last factorial axis (O'_p), obtained from successive multiple correspondence analyses (MCA) where is eliminated in each analysis the variable with the largest contribution to this axis; C2: average order (O_s) of the contributions of the descriptor in the first three factorial axes (O_1, O_2 and O_3, respectively) of the MCA with all the descriptors, and of the order O'_p already defined, such that $O_s = (O_1 + O_2 + O_3 + O'_p)/4$; C3: average order ($O_z$) of the first three factorial axes of the MCA complete, weighted by contributions (w_1, w_2 and w_3) of these axes, such that $O_z = (w_1O_1 + w_2O_2 + w_3O_3)/(w_1+w_2+w_3)$ [C1: ordem inversa da contribuição do descritor para o último eixo fatorial (O'_p), obtida a partir de análises de correspondência múltipla (ACM) sequenciais, em que se elimina, em cada uma, a variável com a maior contribuição para este eixo; C2: ordem média (O_s) das contribuições do descritor nos três primeiros eixos fatoriais (O_1, O_2 e O_3, respectivamente) da ACM com todos os descritores, e da ordem O'_p já definida, tal que $O_s = (O_1 + O_2 + O_3 + O'_p)/4$; C3: ordem média ($O_z$) dos três primeiros eixos fatoriais da análise ACM completa, ponderada pelas contribuições (w_1, w_2 e w_3) desses eixos, tal que $O_z = (w_1O_1 + w_2O_2 + w_3O_3)/(w_1+w_2+w_3)$].

flowering (DFl), ripe fruit persistence (PtF), aroma (Arm), number of flowers per axil (NFA), stem shape (FHt), fruit cross-sectional corrugation (STv), plant height (CPt), branching habit (DsR), plant canopy width (LPt), days to fruiting (DFt), anther colour (CAt), stigma exsertion (PEt), fruit surface (SFt), stem length (CHt), stem colour (CrH), filament colour (CFn), number of locules (NLl), leaf colour (CrF), tillering (BtF), calyx margin (MgC), leaf density (DsF), corolla shape (FmC), pungency (Pug), fruit length (CFt), flower position (PFr), and stem diameter (DHt). Again, about a third of these descriptors are fruit traits, reinforcing the results reported previously.

Table 7. Ascending order of the descriptors (Var) used to the characterization of accessions of *Capsicum frutescens* from the Embrapa Hortaliças germplasm bank, obtained from the application of three criteria (C1, C2 and C3)[1] to choose of the list of minimum descriptors (ordem crescente dos descritores (Var) usados para a caracterização de acessos de *Capsicum frutescens* do Banco de Germoplasma da Embrapa Hortaliças, obtida a partir da aplicação de três critérios (C1, C2 e C3)[1] para escolha da lista de descritores mínimos). Brasília, Embrapa Hortaliças, 2005.

C1		C2		C3		C1		C2		C3	
Var	O'_p	Var	O_s	Var	Oz	Var	O'_p	Var	O_s	Var	Oz
CFI	1.50	PsF	10.25	CFM	1.38	DsR	29.00	Pug	24.75	PFr	26.63
CpP	1.50	CFM	11.50	AtN	7.96	CrH	30.00	FmF	25.25	DHt	27.04
SSt	4.50	AtN	11.75	PsF	8.60	STv	31.00	NLl	25.75	Seg	27.67
CPt	4.50	CPt	12.38	Pro	9.71	PFt	32.00	FPF	26.38	NSF	28.99
Seg	4.50	Arm	14.00	Dfl	10.87	NFA	33.00	DsF	27.50	PFt	29.05
LFt	4.50	CHt	16.00	PtF	12.49	FFt	34.00	NSF	27.50	CPl	29.27
PFr	7.00	FHt	16.25	Arm	12.94	CFn	35.00	CPl	28.25	FmF	29.28
FmC	8.00	Pro	16.63	NFA	13.55	LPt	36.00	PFt	29.25	FPF	29.47
BtF	9.00	SFt	16.75	FHt	13.67	Pro	37.00	SSt	29.38	LFt	29.54
CHt	10.00	DFt	18.00	STv	14.58	DsF	38.00	CFI	30.00	EFt	33.73
Hbt	11.00	DsR	18.38	CPt	14.89	NLl	39.00	CpP	30.00	Oft	37.02
FmF	12.00	NFA	18.75	DsR	15.04	DHt	40.00	DHt	30.00	SSt	37.95
OFt	13.00	STv	19.00	LPt	15.55	CAt	41.00	OFt	30.75	CFI	39.05
CFt	14.00	BtF	19.25	Dft	15.76	CFM	42.00	EFt	30.88	CPF	39.05
FPF	15.00	PEt	19.38	CAt	16.60	PtF	43.00	Hbt	33.13	Hbt	40.09
PsF	16.00	DFl	19.88	Pet	17.02	DFl	44.00	FFt	45.63	APF	49.55
SFt	17.00	PtF	20.00	SFt	17.69	Esp	50.50	APF	49.75	CCL	49.55
CrF	18.00	FmC	20.50	CHt	18.19	PHt	50.50	CCl	49.75	CCR	49.55
Arm	19.00	Seg	20.88	CrH	19.83	PbF	50.50	CCr	49.75	CrC	49.55
EFt	20.00	LPt	21.00	CFn	20.08	CrC	50.50	CrC	49.75	CSt	49.55
FHt	21.00	PFr	21.00	NLL	22.13	CCr	50.50	CSt	49.75	Esp	49.55
DFt	22.00	CrF	22.25	CrF	22.23	McE	50.50	Esp	49.75	FFt	49.55
Pug	23.00	CrH	22.50	BtF	23.04	Pgl	50.50	McE	49.75	McE	49.55
AtN	24.00	CFt	23.25	MgC	23.22	CCl	50.50	MVr	49.75	MVr	49.55
NSF	25.00	CAt	23.38	DsF	24.13	APF	50.50	PbF	49.75	PbF	49.55
MgC	26.00	LFt	23.38	FmC	24.90	PcF	50.50	PcF	49.75	PCF	49.55
PEt	27.00	MgC	24.13	Pug	25.78	MVr	50.50	Pgl	49.75	Pgl	49.55
CPl	28.00	CFn	24.25	CFt	26.25	CSt	50.50	PHt	49.75	PHt	49.55

[1]C1: reverse order of the contribution of the descriptor for the last factorial axis (O'_p), obtained from successive multiple correspondence analysis (MCA), where is eliminated in each analysis the variable with the largest contribution to this axis; C2: average order (O_s) of the contributions of the descriptor in the first three factorial axes (O_1, O_2 and O_3, respectively) of the MCA with all the descriptors, and of the order O'_p already defined, such that $O_s = (O_1 + O_2 + O_3 + O'_p)/4$; C3: average order ($O_z$) of the first three factorial axes of the MCA complete, weighted by contributions (w_1, w_2 and w_3) of these axes, such that $O_z = (w_1O_1 + w_2O_2 + w_3O_3)/(w_1+w_2+w_3)$ [C1: ordem inversa da contribuição do descritor para o último eixo fatorial (O'_p), obtida a partir de análises de correspondência múltipla (ACM) sequenciais, em que se elimina, em cada uma, a variável com a maior contribuição para este eixo; C2: ordem média (O_s) das contribuições do descritor nos três primeiros eixos fatoriais (O_1, O_2 e O_3, respectivamente) da ACM com todos os descritores, e da ordem O'_p já definida, tal que $O_s = (O_1 + O_2 + O_3 + O'_p)/4$; C3: ordem média ($O_z$) dos três primeiros eixos fatoriais da análise ACM completa, ponderada pelas contribuições (w_1, w_2 e w_3) desses eixos, tal que $O_z = (w_1O_1 + w_2O_2 + w_3O_3)/(w_1+w_2+w_3)$].

Table 8. Lists of minimum descriptors proposed to the characterization of *Capsicum* spp. and its cultivated species *C. annuum*, *C. baccatum*, *C. chinense* and *C. frutescens*[1], obtained from the analyses of 893 accessions from the Embrapa Hortaliças germplasm bank, assessed by 56 morphological descriptors[2] (listas de descritores mínimos propostas para a caracterização de *Capsicum* spp. e de suas espécies cultivadas *C. annuum*, *C. baccatum*, *C. chinense* e *C. frutescens*[1], obtidas a partir da análise de 893 acessos do Banco de Germoplasma da Embrapa Hortaliças, caracterizados segundo 56 descritores morfológicos[2]). Brasília, Embrapa Hortaliças, 2005.

***Capsicum* spp. (34)**	***C. annuum* (30)**	***C. baccatum* (30)**	***C. chinense* (32)**	***C. frutescens* (30)**
Species (Esp)	Fruit length (CFt)	Fruit shape at blossom end (FPF)	Flower position (PFr)	Mature fruit colour (CFM)
Fruit wall thickness (EFt)	Calyx pigmentation (Pgl)	Fruit position (PsF)	Number of locules (NLl)	Nodal anthocyanin (AtN)
Fruit width (LFt)	Fruit wall thickness (EFt)	Fruit length (CFt)	Fruit Pedicel length (CpP)	Fruit position (PsF)
Number of flowers per axil (NFA)	Fruit width (LFt)	Fruit weight (PFt)	Mature fruit colour (CFM)	Origin (Pro)
Fruit weight (PFt)	Origin (Pro)	Fruit shape (FFt)	Fruit surface (SFt)	Days to flowering (DFl)
Fruit position (PsF)	Fruit weight (PFt)	Fruit wall thickness (EFt)	Anther colour (CAt)	Ripe fruit persistence (PtF)
Origin (Pro)	Stem colour (CrH)	Fruit surface (SFt)	Fruit position (PsF)	Aroma (Arm)
Ripe fruit persistence (PtF)	Corolla colour (CrC)	Corolla colour (CrC)	Fruit length (CFt)	Number of flowers per axil (NFA)
Calyx margin (MgC)	Immature fruit colour (CFI)	Stem length (CHt)	Fruit wall thickness (EFt)	Stem shape (FHt)
Fruit shape (FFt)	Plant growth habit (Hbt)	Plant height (CPt)	Fruit shape at pedicel attachment (OFt)	Fruit cross-sectional corrugation (STv)
Fruit shape at pedicel attachment (OFt)	Filament colour (CFn)	Number of locules (NLl)	Fruit shape (FFt)	Plant height (CPt)
Number of seeds per fruit (NSF)	Stem length (CHt)	Immature fruit colour (CFI)	Origin (Pro)	Branching habit (DsR)
Anther colour (CAt)	Anther colour (CAt)	Origin (Pro)	Fruit weight (PFt)	Plant canopy width (LPt)
Number of locules (NLl)	Fruit shape at pedicel attachment (OFt)	Fruit Pedicel length (CpP)	Number of seeds per fruit (NSF)	Days to fruiting (DFt)
Filament colour (CFn)	Fruit shape at blossom end (FPF)	Pungency (Pug)	Stem shape (FHt)	Anther colour (CAt)
Placenta length (CPl)	Tillering(BtF)	Days to flowering (DFl)	Days to fruiting (DFt)	Stigma exsertion (PEt)
Corolla colour (CrC)	Fruit position (PsF)	Number of flowers per axil (NFA)	Fruit width (LFt)	Fruit surface (SFt)
Mature fruit colour (CFM)	Leaf density (DsF)	Tillering(BtF)	Ripe fruit persistence (PtF)	Stem length (CHt)
Fruit shape at blossom end (FPF)	Leaf colour (CrF)	Filament colour (CFn)	Days to flowering (DFl)	Stem colour (CrH)
Leaf shape (FmF)	Fruit shape (FFt)	Number of seeds per fruit (NSF)	Calyx margin (MgC)	Filament colour (CFn)
Pungency (Pug)	Number of seeds per fruit (NSF)	Fruit width (LFt)	Leaf pubescence (PbF)	Number of locules (NLl)
Fruit Pedicel length (CpP)	Pungency (Pug)	Leaf pubescence (PbF)	Plant canopy width (LPt)	Leaf colour (CrF)
Plant canopy width (LPt)	Fruit surface (SFt)	Stigma exsertion (PEt)	Fruit shape at blossom end (FPF)	Tillering(BtF)
Stigma exsertion (PEt)	Calyx margin (MgC)	Stem pubescence (PHt)	Stem diameter (DHt)	Calyx margin (MgC)
Fruit cross-sectional corrugation (STv)	Mature fruit colour (CFM)	Fruit shape at pedicel attachment (OFt)	Plant height (CPt)	Leaf density (DsF)
Aroma (Arm)	Plant height (CPt)	Placenta length (CPl)	Leaf shape (FmF)	Corolla shape (FmC)
Days to fruiting (DFt)	Flower position (PFr)	Mature fruit colour (CFM)	Immature fruit colour (CFI)	Pungency (Pug)
Fruit surface (SFt)	Number of locules (NLl)	Days to fruiting (DFt)	Stem length (CHt)	Fruit length (CFt)
Nodal anthocyianin (AtN)	Stem pubescence (PHt)	Branching habit(DsR)	Filament colour (CFn)	Flower position (PFr)
Plant growth habit (Hbt)	Plant canopy width (LPt)	Plant growth habit (Hbt)	Placenta length (CPl)	Stem diameter (DHt)
Flower position (PFr)			Corolla shape (FmC)	
Fruit length (CFt)			Pungency (Pug)	
Stem length (CHt)				
Plant height (CPt)				

[1]The value between parentheses in the header represents the number of descriptors in each list (o valor entre parênteses no cabeçalho representa o número de descritores em cada lista); [2]the order in that variables presented in each list is related to the importance of the descriptor to the characterization of the corresponding germplasm (a ordem de apresentação das variáveis em cada lista tem relação com a importância do descritor para a caracterização do respectivo germoplasma).

Synthesis of the minimum descriptors - Table 8 shows the minimal descriptor lists proposed generically for *Capsicum* spp., and, in particular, for the cultivated species (*C. annuum*, *C. baccatum*, *C. chinense* and *C. frutescens*). Stem length (CHt), plant height (CPt) and fruit length (CFt) were added to the listing for *Capsicum* spp. because these descriptors were always included in the listings established for the specific subcollections.

Eight of the 56 descriptors studied (fruit blossom end appendage, calyx annular constriction, corolla spot colour, seed colour, male-sterility, varietal mixture condition, neck at base of fruit, and seed surface) were not essential for characterizing the accessions of this germplasm. Some of these, however, can be occasionally considered in this type of characterization; for example, cayx annular constriction that is one of the main traits for morphological distinction between *C. chinense* and *C. frutescens* (Sudré *et al*., 2006). Five other descriptors were selected exclusively for one of the collections ("species" for *Capsicum* spp.; calyx pigmentation for *C. annuum*; and nodal anthocyanin, aroma and fruit cross sectional corrugation for *C. frutescens*), although some with relatively little importance (e.g. Pgl and STv).

These facts showed the possibility of reducing the list of descriptors currently adopted at the *Capsicum* germplasm bank of Embrapa Hortaliças, and without harming the representation of their genetic variability. In general, the minimal descriptor lists proposed in the present study (Table 8) allow approximately 50% of the original descriptors to be discarded.

It is pointed out that ten of the minimal descriptors indicated (Table 8) are present in the five established lists: filament colour (CFn), mature fruit colour (CFM), number of locules (NLl), fruit position (PsF), origin (Pro), pungency (Pug), fruit surface (SFt), stem length (CHt), plant height (CPt), and fruit length (CFt). These descriptors were shown therefore to be relevant for germplasm characterization in the four *Capsicum* species studied.

A further fourteen descriptors also deserve mention because were included in the lists of three of the four species: tillering (BtF), anther colour (CAt), immature fruit colour (CFI), days to flowering (DFl), days to fruiting (DFt), fruit wall thickness (EFt), fruit shape (FFt), fruit shape at blossom end (FPF), plant canopy width (LPt), calyx margin (MgC), number of seeds per fruit (NSF), fruit shape at pedicel attachment (OFt), flower position (PFr), and fruit weight (PFt). Only three of these were not repeated in the generic *Capsicum* spp. list (BtF, CFI and DFl). Another fifteen were part of two of these lists, and eight (CPl, CpP, CrC, FmF, Hbt, NFA, PtF and PEt) are also on the generic list. Thus, the minimal descriptors selected for the whole collection are in general represented among those of the subcollections. In spite of this, according to Valls (2007), the most suitable descriptors to discriminate among species of a genus tend to be the least suitable for differentiating cultivars within of species.

Another fact that deserves further emphasis is that most of the descriptors selected are fruit characteristics. Given the importance of these traits for *Capsicum* breeding, it was found that it is possible to characterize these germplasm collections to meet simultaneously the criteria of representing genetic variability and the most immediate interests of the breeder. In this sense, if the germplasm characterization takes into account traits of interest for breeding, the possibilities of its use are increased, as stipulated in public policies on conservation and use of plant genetic sources (FAO, 1991).

From the methodological point of view, the results of this study show, firstly, that the multiple correspondence factorial analysis (MCA) is an useful alternative in the multivariate analysis of data from germplasm collections, especially when most of the descriptors have a qualitative nature. Secondly, it enabled to state that the criteria of discarding of variables based on the original proposal by Jolliffe (1973) is less efficient than the alternatives that value more the contribution of the variables on the first factorial axes.

ACKNOWLEDGEMENTS

The authors thank Embrapa Hortaliças by providing the data, and National Council for Scientific and Technological Development (CNPq) for offering the first author the conditions to conclude this study.

REFERENCES

ALMEIDA CMCV; DIAS LAS; OKABE ET; MEDEIROS JRP. 2005. Variability in genetic resources of cacao in Rondônia, Brazil. *Crop Breeding and Applied Biotechnology* 5: 318-324.

BEKELE FL; KENNEDY AJ; MCDAVID C; LAUCKNER FB; BEKELE I. 1994. Numerical taxonomic studies on cacao (*Theobroma cacao* L.) in Trinidad. *Euphytica* 75: 231-240.

CRUZ CD. 2006. *Programa Genes*. Software. Versão Windows: aplicativo computacional em genética e estatística. Viçosa: UFV. (CD-ROM).

CRUZ CD; CARNEIRO PCS. 2006. *Modelos biométricos aplicados ao melhoramento genético*. 2 ed. Viçosa: UFV. v. 2, 585p.

CURY R. 1993. *Dinâmica evolutiva e caracterização de germoplasma de mandioca (*Manihot esculenta *Crantz) na agricultura autóctone do sul do Estado de São Paulo*. Piracicaba: USP-ESALQ. 103p (Dissertação mestrado).

DIAS LAS. 1994. *Divergência genética e fenética multivariada na predição de híbrido e preservação de germoplasma de cacau (*Theobroma cacao *L.)*. Piracicaba: USP-ESALQ. 94p. (Tese doutorado).

ESCOFIER B; PAGÈS J. 1992. *Análisis factoriales simples y múltiples: objetivos, métodos e interpretación.* Bilbao: Servicio Editorial de la Universidad del País Vasco. 285p.

FAO. 1991. *Relatório da quarta reunião da Comissão de Recursos Fitogenéticos*. Roma: FAO. 96p.

FRANKEL OH; BROWN AHD. 1984. Plant genetic resources today: a critical reappraisal. In: HOLDEN JHW; WILLIAMS JT (eds). *Crop genetic resources: conservation and evaluation*. London: George Allen & Unwin, p. 249-257.

GALWEY NW. 1995. Verifying and validating the representativeness of a core collection. In: HODGKIN T; BROWN AHD; VAN HINTUM TJL; MORALES EAV (eds). *Core collections of plant genetic resources*. New York: J. Wiley. p. 187-198.

GONÇALVES LS; RODRIGUES R; AMARAL JUNIOR AT; KARASAWA M; SUDRÉ CP. 2009. Heirloom tomato gene bank: assessing genetic divergence based on morphological, agronomic and molecular data using a Ward-modified location model. *Genetics and Molecular Research*, 8: 364-374.

INCE AG; KARACA M; ONUS AN. 2009.

Development and utilization of diagnostic DAMD-PCR markers for *Capsicum* accessions. *Genetic Resources and Crop Evolution* 56: 211-221.

JOLLIFFE IT. 1973. Discarding variables in a principal component analysis, II Real data. *Applied Statistics* 22: 21-31.

LUZ FJF. 2007. *Caracterizações morfológica e molecular de pimentas (*Capsicum chinense *Jack.)*. Jaboticabal: UNESP-FCAV. 150p (Tese doutorado).

MOSCONE EA; SCALDAFERRO MA; GRABIELE M; CECCHINI NM; SÁNCHEZ GARCÍA Y; JARRET R; DAVIÑA JR; DUCASSE DA; BARBOZA GE; EHRENDORFER F. 2007. The evolution of Chili Peppers (*Capsicum* - Solanaceae): a cytogenetic perspective. VI International Solanaceae Conference: Genomics Meets Biodiversity. *Acta Horticulturae* 745: 137-170.

ONOYAMA SS; REIFSCHNEIDER FJB; MOITA AW; SOUZA GS. 2010. Atributos de hortaliças sob a ótica de consumidores: estudo de caso do pimentão no Distrito Federal. *Horticultura Brasileira* 28: 124-132.

ONYILAGHA JC. 1986. Numerical analysis of variation among Nigerian *Dioscorea* rotunda accessions. *Euphytica* 35: 413-419.

ORTIZ R; DE LA FLOR FD; ALVARADO G; CROSSA J. 2010. Classifying vegetable genetic resources. A case study with domesticated *Capsicum* spp. *Scientia Horticulturae* 126: 186-191.

PEREIRA AV. 1989. *Utilização de análise multivariada na caracterização de germoplasma de mandioca* (Manihot esculenta *Crantz*). Piracicaba: USP-ESALQ. 180p (Tese doutorado).

ROHLF FJ. 2005. *NTSYS-PC: Numerical taxonomy and multivariate analysis system*. Version 2.2: software. New York: Exeter/ Setauket.

SAS INSTITUTE. 2009. *SAS® System for Windows*. Version 9.0: software. SAS Institute: Cary, NC.

SILVA WCJ. 2008. *Escolha de descritores mínimos e estabelecimento de coleções nucleares em* Capsicum *spp*. Goiânia: UFG. 139p (Tese doutorado).

SIMPSON CE; HOWARD ER; HIGGINS DL; THOMAS GD. 2002. *Catalog of minimum descriptors of* Arachis hypogaea *L. germplasm collections from South America*. College Station: Texas A&M University, p. 1977-1982.

STRAPASSON E; VENCOVSKY R; BATISTA LAR. 2000. Seleção de descritores na caracterização de germoplasma de *Paspalum* sp. por meio de componentes principais. *Revista Brasileira de Zootecnia* 29: 373-381.

SUDRÉ CP. 2003. *Divergência genética e avaliação da resistência à mancha bacteriana em* Capsicum *spp*. Campos dos Goytacazes: UENF. 112p (Tese doutorado).

SUDRÉ CP; CRUZ CD; RODRIGUES R; RIVAL EM; AMARAL JÚNIOR AT; SILVA DJH; PEREIRA TNS. 2006. Variáveis multicategóricas na determinação da divergência genética entre acessos de pimenta e pimentão. *Horticultura Brasileira* 24: 88-93.

SUDRÉ CP; GONÇALVES LSA; RODRIGUES R; AMARAL JÚNIOR AT; RIVA-SOUZA EM; BENTO CS. 2010. Genetic variability in domesticated *Capsicum* spp. as assessed by morphological and agronomic data in mixed staatistical analysis. *Genetics and Molecular Research 9: 283-294*.

TEIXEIRA R. 1996. *Diversidade em Capsicum: análise molecular, morfoagronômica e química*. Viçosa: UFV. 81p (Dissertação mestrado).

VALLS JFM. 2007. Caracterização de recursos genéticos vegetais. In: Nass LL (ed). *Recursos genéticos vegetais*. Brasília: Embrapa Recursos Genéticos e Biotecnologia, v. 1, cap. 8, p. 281-305.