Selection of the most informative morphoagronomic descriptors for cassava germplasm

Eder Jorge de Oliveira [1], Osvaldo Sebastião de Oliveira Filho[2] and Vanderlei da Silva Santos[1]

[1]Embrapa Mandioca e Fruticultura, Rua Embrapa, s/nº, Caixa Postal 007, CEP 44380-000 Cruz das Almas, BA, Brazil. E-mail: eder.oliveira@embrapa.br, vanderlei.silva-santos@embrapa.br [2]Universidade Federal do Recôncavo da Bahia, Campus Cruz das Almas, CEP 44380-000 Cruz das Almas, BA, Brazil. E-mail: osvaldoosof@yahoo.com.br

Abstract – The objective of this work was to select the most informative morphoagronomic descriptors for cassava (*Manihot esculenta*) germplasm and to evaluate the ability of different methods to select the descriptors. Ninety-five accessions were characterized using 51 morphoagronomic descriptors. Data were subjected to a multiple correspondence analysis (MCA), whose information was used in the following four methods of descriptor selection: reverse order of the descriptor for the pth factorial axis of the MCA (Jolliffe); sequential, multiple correspondence analysis (SMCA); mean of the contribution orders of the descriptor in the first three factorial axes (C3PA); and C3PA method weighted by the respective eigenvalues of the full analysis (C3PAWeig). The correlations between the dissimilarity matrix with all descriptors and the most informative descriptors were high and significant (0.75, 0.77, 0.83, and 0.84 for C3PAWeig, C3PA, SMCA, and Jolliffe, respectively). The less informative descriptors were discarded, considering those common among the selection methods and relevant for the breeding interests. Therefore, 32 morphoagronomic descriptors with correlation between the dissimilarity matrices (r=0.81) were selected, due to their high capacity to discriminate cassava germplasm and to their ability to maintain some preliminary agronomic traits, useful for the initial characterization of the germplasm.

Index terms: *Manihot esculenta*, descriptor selection, genetic resources, multiple correspondence analysis.

Seleção dos descritores morfoagronômicos mais informativos para germoplasma de mandioca

Resumo – O objetivo deste trabalho foi selecionar os descritores morfoagronômicos mais informativos para o germoplasma de mandioca (*Manihot esculenta*), e avaliar a capacidade de diferentes métodos em selecionar os descritores. Foram caracterizados 95 acessos com uso de 51 descritores morfoagronômicos. Os dados foram submetidos à análise de correspondência múltipla (ACM), cujas informações foram utilizadas nos seguintes quatro métodos de seleção de descritores: ordem inversa do descritor para o eixo fatorial pth da ACM (Jolliffe); análise de correspondência múltipla sequencial (ACMS); média das ordens de contribuição do descritor nos três primeiros eixos fatoriais (C3PE); e método C3PE ponderado pelos pesos dos respectivos autovalores da análise completa (C3PEPond). As correlações entre a matriz de dissimilaridade com todos os descritores e os descritores mais informativos foram altas e significativas (0,75, 0,77, 0,83 e 0,84 para C3PEPond, C3PE, ACMS e Jolliffe, respectivamente). Os descritores menos informativos foram descartados, tendo-se considerado apenas aqueles comuns entre os métodos de seleção e relevantes para os interesses do melhoramento. Assim, foram selecionados 32 descritores morfoagronômicos com correlação entre as matrizes de dissimilaridade (r=0,81), em razão de sua alta capacidade para discriminar o germoplasma de mandioca e de sua habilidade de manter alguns caracteres agronômicos preliminares, úteis para a caracterização inicial do germoplasma.

Termos para indexação: *Manihot esculenta*, seleção de descritor, recursos genéticos, análise de correspondência múltipla.

Introduction

Most of the produced cassava (*Manihot esculenta* Crantz) is destined for food consumption (human and animal) and for industrial purposes. Cassava starch is a worldwide, multibillion-dollar business, especially due to its many industrial applications (Tonukari, 2004).

The wide range of cassava uses should be linked to the genetic diversity of the species, which guides the development of varieties with traits that meet the diverse specificities of the consumer market as well as the crop production system. Therefore, the maintenance of cassava genetic diversity ensures the conservation of useful alleles that are related to resistance to pests and

diseases, conferring better root quality and differential starch characteristics (Raji et al., 2007; Oliveira et al., 2014). The availability of a wide genetic variability is essential for cassava breeding programs that are aimed at developing new varieties with levels of adaptation to cultivation in different environments and industrial adjustments.

The organization and maintenance of major germplasm banks and efforts for the collecting, characterization, evaluation, and use of wild species and landraces are of immeasurable importance in ensuring the sustainability of the cassava production chain, particularly in countries with continental dimensions, such as Brazil. The Brazilian cassava germplasm collections have been characterized primarily based on the analysis of morphological descriptors and cyanogenic compounds in the roots for the classification of genotypes as "sweet" or "bitter" cassava (Fukuda & Alves, 1987).

Some studies have been conducted for the Brazilian cassava germplasm using morphoagronomic descriptors and molecular markers, in order to evaluate the genetic diversity and associations between different traits (Vieira et al., 2011; Mezette et al., 2014). However, most of these studies used only part of the descriptors currently available for cassava and a small number of germplasm accessions, mainly due to the difficulty of using all descriptors in the entire collection. To reduce the work of characterization activities, several studies have been conducted to define a list of the most informative descriptors to distinguish accessions in gene banks (Strapasson et al., 2000; Castro et al., 2012; Oliveira et al., 2012; Silva et al., 2013). Reliable methods and procedures for germplasm characterization are essential to increase the use of available genetic variability in breeding programs (Oliveira et al., 2012).

In cassava, 75 descriptors have been used for germplasm characterization, of which 54 are morphological and 21 are agronomic traits (Fukuda & Guevara, 1998). Recently, Fukuda et al. (2010) published a revised version of their work, focusing on the documentation and characterization of cassava germplasm to analyze data and to draw comparisons among different countries. However, the selection of these descriptors was not carried out based on their discrimination power, which was identified using appropriate statistical tools. Furthermore, the number of descriptors still remains high for use in cassava germplasm characterization, requiring a large number of observations, which is extremely time-consuming and costly.

Although the definition of descriptors of greater importance has been made based on the experience of researchers (Coffelt & Johnson, 2011), the use of multivariate analysis techniques have been more effective in identifying descriptors of major interest, indicating the disposal of those less relevant descriptors (Strapasson et al., 2000; Giraldo et al., 2010; Castro et al., 2012; Oliveira et al., 2012; Silva et al., 2013). Besides, multivariate analyses, specifically the multiple correspondence analysis (MCA), have the advantage of assessing the importance of each studied descriptor in the total available variation among accessions, enabling the discard of the less discriminating descriptors which are invariant or correlated with other descriptors, similarly to the principal component analysis.

The objective of this work was to select the most informative descriptors for cassava (*Manihot esculenta* Crantz) germplasm evaluation and to evaluate the ability of different methods for descriptors selection.

Materials and Methods

Ninety-five germplasm accessions belonging to the Cassava Germplasm Bank (CGB) from Embrapa Mandioca e Fruticultura (Cruz das Almas, Brazil) were characterized from 2011–2013. The choice of the germplasm accessions was based on their high phenotypic contrasts observed under field conditions. This database consists of landraces and improved varieties which resulted from conventional breeding procedures, such as crossing and selection, as well as the selection of landraces with high yield potential, as identified by farmers or research institutions.

The accessions were planted at the beginning of the rainy season (May–July 2011–2013) using stem cuttings of 15–20 cm in single rows. Spacing was 0.90 m between rows and 0.80 m within rows, and the cultivation system was performed according to recommendations for cassava cultivation (Souza et al., 2006). Harvesting was done between 11 and 12 months after planting in the three years of evaluation.

Previously characterized with 51 descriptors according to Fukuda et al. (2010), the 95 accessions were used to define the list of most informative cassava descriptors (Table 1). The descriptors were divided

into four categories, 13 minimum, 13 principal, 10 secondary and 15 preliminary agronomic descriptors (Fukuda et al., 2010). As most traits are qualitative, the quantitative traits were also categorized into the same analysis. In this case, seven classes for each quantitative trait were distributed, and this categorization was based on the range of data variation.

The multiple correspondence analysis (MCA) (Escofier & Pagès, 1992) was used to select the minimum descriptors similarly to the principal component analysis (PCA). The selection of the most informative descriptors was based on four criteria – the Jolliffe method, the sequential multiple correspondence analysis, the selection by the mean contribution orders, and the selection by the weighted average contribution orders – which are described as follows.

The Jolliffe method uses the reverse order of the descriptor for the pth factorial axis (O'p) (Jolliffe, 1973). In this case, the descriptor with the highest contribution to the last factorial axis can be discarded, considering that the importance of the main principal components or factorial axes decreases from the first to the last, and also because this descriptor explains little of the total variance. Therefore, the descriptor with the highest coefficient in the principal axis and the lowest eigenvalue in the last axis can be discarded.

The sequential MCA (SMCA) has the descriptor with the highest contribution to the last factorial axis discarded, and further analyses are carried out with the other descriptors, until their ordering is established according to their importance (Silva et al., 2013)

The selection based on the mean of the contribution orders of the descriptor on the first three factorial axes of the full analysis (C3PA) along with O'p, i.e., OS = (O1 + O2 + O3 + O'p)/4 (Silva et al., 2013).

The selection based on the weighted mean of the contribution orders of the descriptor for the first three axes (C3PAWeig), with weights (w1, w2, and w3) defined by the respective eigenvalues of the full analysis, i.e.: Oz = (w1O1 + w2O2 + w3O3)/(w1 + w2 + w3) (Silva et al., 2013).

In the present work, all these analyses were performed using the FactoMineR package (Lê et al., 2008) for R version 3.0.1.

To compose the list of the most informative descriptors, a maximum of 30 descriptors was initially adopted. The efficiency of the four selection criteria in establishing the minimum list of descriptors was

Table 1. List of descriptors found in the 95 accessions (K2), used for the germplasm characterization of *Manihot esculenta*, with the respective number of defined categories in the descriptor manual (K1) (Fukuda et al., 2010), and the code for each descriptor.

Descriptor	K1	K2	Code
Minimum descriptors			
Color of apical leaves	4	3	ColApLea
Pubescence on apical leaves	2	2	PubApLea
Shape of central lobe	10	4	ShaCeLob
Petiole color	6	5	PetCol
Color of stem cortex	3	3	ColStCor
Color of stem exterior	7	6	ColStEx
Length of phyllotaxy	3	3	LenFil
Extent of root peduncle	3	3	ExtRooPe
External color of storage root	4	4	ExColRoo
Color of root cortex	4	4	ColRooCo
Color of root pulp	4	3	ColRooPu
Texture of root epidermis	2	2	TexRooEp
Flowering	2	2	Flo
Principal descriptors			
Leaf color	4	2	LeaCol
Number of leaf lobes	5	4	NLeaLo
Length of leaf lobe	7	6	LenLeaLo
Width of leaf lobe	7	7	WidLeaLo
Ratio of length/width of leaf lobe	7	5	RaLenWidLea
Petiole length	7	7	PetLen
Color of stem epidermis	4	3	ColStEp
Growth habit of stem	2	1	GroHabSt
Color of end branches of adult plant	3	2	ColBraAPl
Plant height	7	4	PlHei
Height to first branching	7	7	HeiFiBra
Levels of branching	7	4	LevBra
Root constrictions	3	3	RooConst
Secondary descriptors			
Color of leaf vein	4	4	ColLeaVe
Peduncle position	4	3	PedPos
Prominence of foliar scars	2	1	ProFoSca
Length of stipules	2	1	LenSti
Stipule margin	2	2	StiMarg
Branching habit	4	3	BraHab
Angle of branching	7	5	AngBra
Sinuosity of the leaf lobe	2	2	SinLea
Root shape	4	4	RooSh
Shape of plant	4	4	ShPl
Preliminary agronomic descriptors			
Vigor	3	3	Vig
Shoot weight	7	7	ShoWe
Number of cutting stems per plant	7	7	NCutStPl
Root length	7	7	RooLen
Root diameter	7	7	RooDiam
Periderm: ease of peeling	2	2	PerEasPel
Cortex: ease of peeling	2	2	CorEasPel
Root position	3	2	RooPos
Yield of commercial roots	7	7	YiComRoo
Yield of noncommercial roots	7	7	YiNComRoo
Dry matter content	7	7	DMC
Resistance to bacterial blight	7	5	Bac
Resistance to anthracnose	7	5	Ant
Tolerance to mites	7	4	Mite
Leaf retention	3	3	LeaRet

evaluated by the correlation between the dissimilarity matrices, which was estimated for all descriptors and was calculated using only the most informative ones. The dissimilarity matrices (simple matching) for the multicategorical variables were calculated in accordance with, dij = D/(C + D), in which: i and j correspond to a pair of accessions (and j = 1, 2, ..., n); C is the number of concordant classes; and D is the number of disagreeing classes.

Correlation between the dissimilarity matrices of all descriptors versus the most informative ones, according to the different selection criteria, was used as a quality pattern for selecting descriptors. The correlation significance was assessed by the t-test and Mantel test with ten thousand simulations using the Genes software (Cruz, 2006). The list of the most informative cassava descriptors was not adopted by a single criterion for discarding descriptors, but by a comparative analysis of those recommended for disposal by most of the analyzed selection criteria.

Results and Discussion

Initially, the descriptors growth habit of stem (GroHabSt), prominence of foliar scars (ProFoSca) and length of stipules (LenSti) did not show any variation in the set of evaluated cassava accessions; therefore, they were eliminated from further analysis.

Descriptor contributions for the first three factorial axes in the complete MCA showed that approximately 62% of the descriptors with the largest contributions accounted for more than 90% of the total variability that was associated with each axis (Table 2). This result can be explained by the existence of redundancy or association between the descriptors, allowing some of them to capture the same information in the germplasm variability.

Petiole length (PetLen) is the descriptor which showed the most contribution to explaining the last axis factor (9.87%) and, in principle, has less importance for the characterization of these accessions (Table 2). In theory, this result means that the information that is associated with this variable is already covered by other descriptors which contribute to other axes. In contrast, PetLen showed a high contribution to the first three factorial axes (12th, 10th and 1st largest contribution, respectively), and is therefore not suitable for disposal under this criterion. In contrast, other descriptors which showed great contribution to

the last factorial axes showed also low contribution in the first axes. For instance, the shape of plant (ShPl) had the second highest contribution to the last factor axis in MCA (8.54%), and the contributions for the first three factorial axes were lower compared to those of PetLen (38th, 15th and 25th, respectively); therefore, the criterion for disposal based on the higher contribution to the last axes would be correct. Additionally, the descriptor tolerance to mites (Mite) had the third largest contribution to the last factorial axis (7.52%), and had also a low contribution for the first three axes (34th, 17th and 35th higher contribution, respectively). Even with these inconsistencies, the informative descriptors were selected for comparison with other disposal methods. This same behavior was observed in *Capsicum* spp., when selecting minimum descriptors for germplasm characterization, in which the authors decided to consider different strategies, instead of discarding descriptors based only in one criterion (Silva et al., 2013).

The 30 descriptors that were selected by SMCA criteria were used to obtain the dissimilarity matrix, which was compared with that obtained for all descriptors. Overall, there was a significant correlation (p<0.01) between the dissimilarity matrices with all of the descriptors versus those selected ones by the SMCA criterion (0.83) (Figure 1).

Considering descriptor contributions for the first three factorial axes of MCA (C3PA), some descriptors were less informative in capturing most of the genetic variation data, as follows: the color of stem cortex (ColStCor), color of apical leaves (ColApLea), peduncle position (PedPos), length of phyllotaxy (LenFil), petiole color (PetCol), flowering (Flo), resistance to anthracnose (Ant), number of leaf lobes (NLeaLo), vigor (Vig), leaf retention (LeaRet), sinuosity of the leaf lobe (SinLea), periderm: ease of peeling (PerEasPel), extent of root peduncle (ExtRooPe), leaf color (LeaCol), cortex: ease of peeling (CorEasPel), stipule margin (StiMarg), color of end branches of adult plant (ColBraAPl) and root position (RooPos) (Table 2). By withdrawing these descriptors from the analyses, the correlations between the dissimilarity matrices were significant, but the absolute value of the correlation (0.77) was lower than that obtained by the SMCA criterion (Figure 1), indicating less adjustment between the discarded descriptors and the full list. The use of different methodologies leads to some

Table 2. The 30 most informative morphoagronomic descriptors of cassava ranked by: successive multiple correspondence analysis (SMCA), contribution to the first three factorial axes (C3PA), weighted criterion of contribution to the first three factorial axes (C3PAWeig), and the Jolliffe criterion (Jolliffe).

Descriptor[1]	Criterion							
	SMCA		C3PA		C3PAWeig		Jolliffe	
	Rank	C1 (%)[2]	Rank	Os (%)[2]	Rank	Oz (%)[2]	Rank	O'p (%)[2]
AngBra	16	26.87	25	1.94	35	1.65	33	2.99
Ant	18	21.15	37	0.90	30	1.91	17	0.80
Bac	28	15.50	15	2.64	23	2.45	39	3.69
BraHab	25	11.61	30	1.41	24	2.32	18	0.89
ColApLea	13	27.75	32	1.25	38	1.00	29	2.02
ColBraAPl	7	35.82	47	0.32	46	0.33	11	0.48
ColLeaVe	37	12.30	26	1.88	19	2.83	10	0.44
ColRooCo	24	21.52	14	2.92	14	4.05	22	1.21
ColRooPu	23	12.89	23	2.02	37	1.41	45	4.49
ColStCor	4	36.62	31	1.27	26	2.18	2	0.11
ColStEp	38	9.70	8	3.45	1	7.71	19	0.97
ColStEx	20	12.86	2	5.11	2	6.91	36	3.08
CorEasPel	12	27.34	45	0.40	44	0.60	1	0.05
DMC	45	7.91	28	1.60	33	1.74	25	1.35
ExColRoo	17	24.34	10	3.28	5	5.97	21	1.12
ExtRooPe	2	50.00	43	0.54	41	0.71	3	0.12
Flo	9	36.49	36	0.94	16	3.28	5	0.17
HeiFiBra	30	17.99	5	4.23	7	5.65	34	3.02
LeaCol	15	26.49	44	0.42	43	0.61	12	0.51
LeaRet	6	30.12	40	0.74	47	0.17	31	2.57
LenFil	27	10.56	33	1.15	18	3.11	15	0.68
LenLeaLo	32	17.87	9	3.30	10	4.97	27	1.94
LevBra	34	12.47	11	3.13	11	4.81	26	1.82
Mite	46	7.52	24	1.94	36	1.42	44	4.31
NCutStPl	39	11.94	4	4.56	3	6.44	46	5.88
NLeaLo	31	10.85	38	0.81	42	0.66	24	1.33
PedPos	5	62.73	34	1.15	40	0.81	30	2.30
PerEasPel	11	37.30	42	0.57	31	1.85	4	0.15
PetCol	22	17.72	35	1.11	34	1.71	20	0.98
PetLen	48	9.87	1	5.72	8	5.02	48	9.87
PlHei	21	14.82	13	3.00	9	4.99	35	3.07
PubApLea	41	8.09	20	2.14	22	2.50	42	4.05
RaLenWidLea	3	42.18	12	3.06	13	4.77	16	0.74
RooConst	33	11.51	27	1.69	21	2.58	14	0.57
RooLen	44	6.63	7	3.60	15	3.60	47	6.06
RooPos	1	0.00	48	0.09	48	0.01	8	0.36
RooSh	43	8.72	29	1.58	20	2.60	13	0.51
ShaCeLob	42	7.67	17	2.47	25	2.19	40	3.89
ShoWe	26	12.64	16	2.54	17	3.16	28	1.97
ShPl	47	8.54	21	2.11	32	1.78	43	4.14
SinLea	8	19.79	41	0.68	29	2.04	9	0.37
StiMarg	14	39.14	46	0.33	45	0.36	7	0.33
TexRooEp	36	13.40	19	2.20	12	4.78	6	0.29
Vig	10	25.04	39	0.76	39	0.95	23	1.32
WidLeaLo	29	37.06	3	4.66	6	5.68	38	3.14
RooDiam	35	10.79	22	2.06	28	2.06	32	2.88
YiComRoo	40	9.08	6	4.03	4	6.26	41	3.89
YiNComRoo	19	17.65	18	2.31	27	2.14	37	3.09

[1]Code for each descriptor is presented on Table 1. [2]C1, contributions (%) to the last factorial axes in each analysis, from which the variable with the largest contribution to the last factorial axis of the previous analysis (SMCA) was eliminated in each cycle; Os, contributions (%) of the descriptor to the first three factorial axes of the full analysis (C3PA); Oz, contributions (%) of the descriptor to the first three axes, with weights defined by the respective eigenvalues of the full analysis (C3PAWeig); O'p: contribution (%) of the descriptor to the last factorial axis (Jolliffe). Underlined numbers refer to the selected descriptors for each criterion.

inconsistencies which should be carefully analyzed. In the study of Oliveira et al. (2012), quantitative descriptors were subjected to principal components analyses using the Singh and direct selection methods, by which eighteen and fifteen quantitative descriptors were respectively discarded by the Singh's and direct selection methods. However, considering the simultaneous analyses of these methodologies, only 60% descriptors were selected to maximize the total variation of the genotypes. Therefore, simultaneous analyses using several methods seem to be an efficient strategy to minimize errors in the elimination of descriptors.

The criterion for the descriptor selection based on the weighted mean of the contribution orders of the descriptor for the first three axes, with weights defined by the respective eigenvalues of the full analysis (C3PAWeig), indicated that the PerEasPel, shape of plant (ShPl), dry matter content (DMC), PetCol, angle of branching (AngBra), Mite, color of root pulp (ColRooPu), ColApLea, Vig, PedPos, ExtRooPe, NLeaLo, LeaCol, CorEasPel, StiMarg, ColBraAPl, LeaRet and RooPos descriptors were less informative for cassava discrimination (Table 2). Although there is a significant correlation between the dissimilarity matrix of the complete data and descriptors based on the criterion C3PAWeig (0.75), the absolute value of the correlation was still less than that of the C3PA criterion (Figure 1). Similar results were observed in *Capsicum* spp., in which the thirty best descriptors
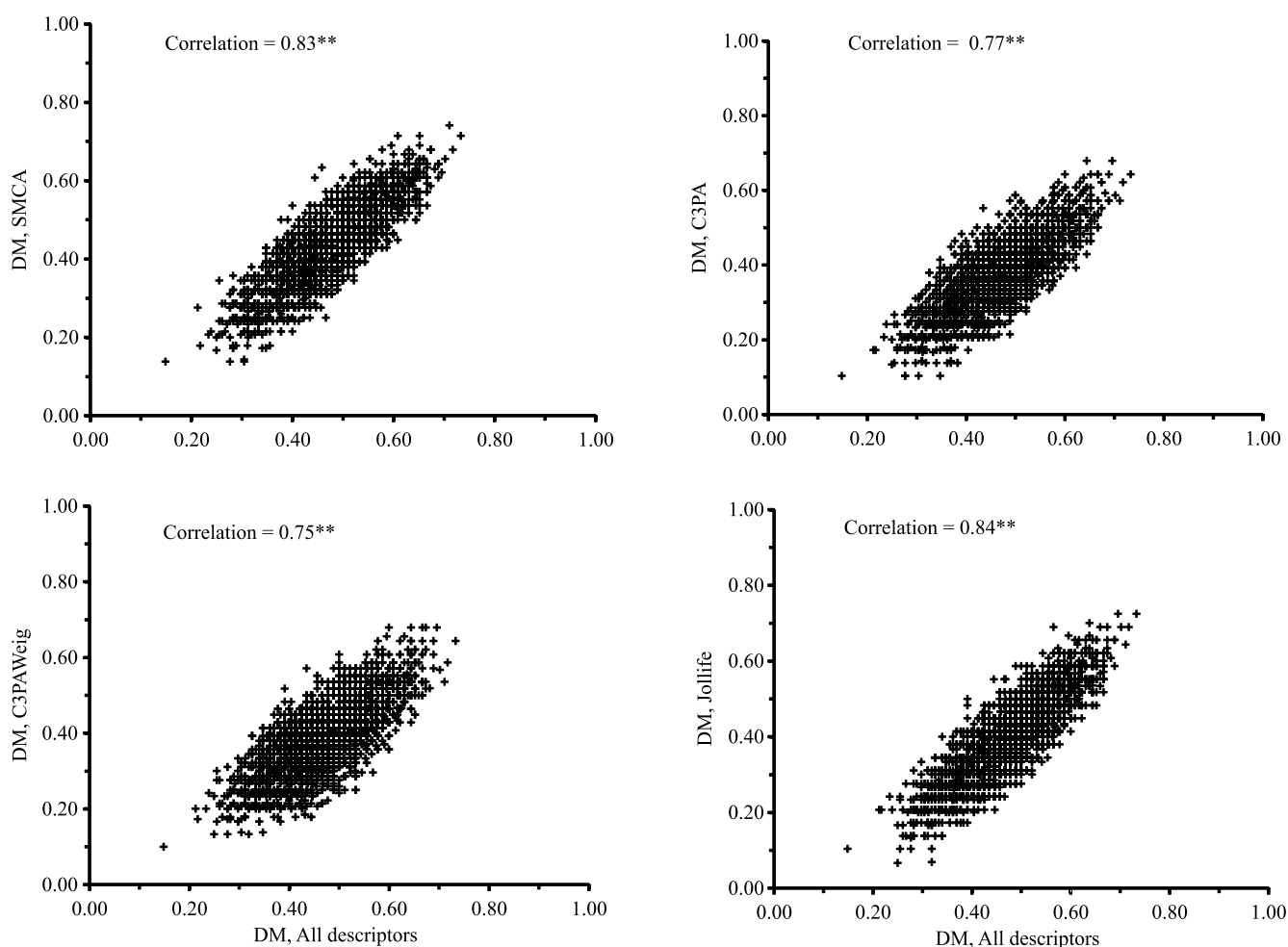

Figure 1. Correlation between the dissimilarity matrices (DM), comparing all the descriptors with those selected by successive, multiple correspondence analysis (SMCA), contribution to the first three factorial axes (C3PA), weighted contribution to the first three factorial axes (C3PAWeig), and Jolliffe criteria (Jolliffe). **Significant at 1% of probability both by t and by Mantel tests.

showed significant correlations (p<0.01) among the dissimilarity matrices, in which the correlation magnitude of C3PAWeig was lower (0.87) than C3PA (0.89) (Silva et al., 2013).

The criterion based on the higher eigenvectors elements in the last components (Jolliffe) indicated that LeaRet, root diameter (RooDiam), AngBra, height to first branching (HeiFiBra), plant height (PlHei), color of stem exterior (ColStEx), yield of noncommercial roots (YiNComRoo), width of leaf lobe (WidLeaLo), resistance to bacterial blight (Bac), yield of commercial roots (YiComRoo), shape of central lobe (ShaCeLob), pubescence on apical leaves (PubApLea), ShPl, Mite, ColRooPu, number of cutting stems per plant (NCutStPl), root length (RooLen) and PetLen (Table 2) should be discarded. Based on the Jolliffe criterion, there was a higher correlation between the dissimilarity matrix of the full data and the informative descriptors (0.84), indicating a better fit in the descriptor disposal (Figure 1). In contrast, Silva et al. (2013) decided to consider the participation of the descriptors in the first three factorial axes (C3PA and C3PAWeig), instead of discarding variables only based on Jolliffe (1973), since this criterion showed that the descriptor "species" (Esp) was the largest contributor to explain the last factorial axis, although showing the second highest contribution to the first factorial axis, and the greatest contribution to the second and third axes.

In general, there is a mismatch between descriptors listed for disposal using the four selection methods. Only the PetLen, RooLen, NCutStPl and YiComRoo descriptors were common to all the disposal methods (Table 2). Therefore, the decision about which descriptors should effectively be discarded was based on the coincidence of at least two methods. In this case, 23 descriptors could be dropped from the analysis (PetLen, RooLen, NCutStPl, YiComRoo, HeiFiBra, PlHei, length of leaf lobe (LenLeaLo), color of stem epidermis (ColStEp), ColStEx, ShaCeLob, WidLeaLo, levels of branching (LevBra), Mite, Bac, color of root cortex (ColRooCo), external color of storage root (ExColRoo), RooDiam, shoot weight (ShoWe), PubApLea, ratio of length/width of leaf lobe (RaLenWidLea), YiNComRoo, texture of root epidermis (TexRooEp) and ShPl. However, considering that the YiComRoo, PlHei, Mite, Bac, ShoWe, YiNComRoo and ShPl descriptors are directly related to the production capacity and resistance to

biotic stresses, being routinely assessed in cassava breeding programs, and that the number of accessions in the present work is only a sample of the germplasm that has been stored from the CGB (approximately 8%), the exclusion of these descriptors is not indicated. In passion fruit, Castro et al. (2012) selected the quantitative descriptors based on direct selection and Singh methods, whereas the qualitative descriptors were analyzed by correlation. These authors found that direct selection using principal component analysis pointed out eight characters to be discarded versus seven using Singh method. Therefore, to reduce inconsistencies in the elimination of descriptors, they adopted both procedures to indicate the most relevant descriptors.

Criteria for discarding descriptors showed very different directions; therefore, it is necessary to make comparisons between the methods, to critically analyze the common descriptors between the methodologies and, possibly, to place more value on the contribution of the variables on the first factorial axes.

Descriptor disposal indication according to selection methods of variables were also observed by Silva et al. (2013) for *Capsicum* spp. However, these authors made no comparison common descriptors to the different methods. Therefore, descriptor choice was based on the selection method of variables showing the highest correlation between the dissimilarity matrices of all descriptors versus the most informative list of descriptors.

The list of morphoagronomic descriptors associated with a higher discrimination capacity of cassava germplasm and important agronomic information to cassava breeders consisted of 32 descriptors (Table 3). The correlation between the dissimilarity matrix comparing the full and the minimum list of descriptors was significant and of high magnitude (0.81) (Figure 2). Although the correlation absolute value between the dissimilarity matrices is smaller than in SMCA methods and in that proposed by Jolliffe (1973), this list of the preliminary agronomic descriptors meets curator interests in the germplasm classification and breeder concerns in the preliminary characterization of germplasm for use in breeding programs. Considering that, in many cases, germplasm characterization for several species is linked to their use in breeding programs, it is important that some descriptors could allow a preliminary analysis of accession agronomic

potential, in order to use them to generate superior segregating populations. Thus, as reported for passion fruit (Castro el al., 2012) and papaya (Oliveira et al., 2012), descriptors related to commercial traits of interest should be kept in the analysis.

The analysis of cassava collection further highlights the high proportion of informative descriptors related to root traits (preliminary agronomic descriptors). This result is interesting because these traits are used to define the potential use of cassava germplasm, either per se or as a parental in cross blocks in breeding programs.

Table 3. List of the most informative descriptors for the characterization of *Manihot esculenta* germplasm, according to Fukuda et al. (2010).

Descriptor	Code
Minimum descriptors	
Color of apical leaves	ColApLea
Petiole color	PetCol
Color of stem cortex	ColStCor
Length of phyllotaxy	LenFil
Extent of root peduncle	ExtRooPe
Color of root pulp	ColRooPu
Flowering	Flo
Principal descriptors	
Leaf color	LeaCol
Number of leaf lobes	NLeaLo
Color of end branches of adult plant	ColBraAPl
Plant height	PlHei
Root constrictions	RooConst
Secondary descriptors	
Color of leaf vein	ColLeaVe
Peduncle position	PedPos
Stipule margin	StiMarg
Branching habit	BraHab
Angle of branching	AngBra
Sinuosity of the leaf lobe	SinLea
Root shape	RooSh
Shape of plant	ShPl
Preliminary agronomic descriptors	
Vigor	Vig
Shoot weight	ShoWe
Periderm: ease of peeling	PerEasPel
Cortex: ease of peeling	CorEasPel
Root position	RooPos
Yield of commercial roots	YiComRoo
Yield of non-commercial roots	YiNComRoo
Dry matter content	DMC
Resistance to bacterial blight	Bac
Resistance to anthracnose	Ant
Tolerance to mites	Mite
Leaf retention	LeaRet

Therefore, it is possible simultaneously to characterize these germplasm collections for the representation of genetic variability and to satisfy the immediate interests of breeders regarding the availability of readily accessions for practical use. Similarly, many of the standard *Capsicum* descriptors are also related to fruit traits in sweet and hot peppers, which has been the focus of the breeding of these species as for their relation with storage, processing, marketing, and the consumption of commercial derivatives (Silva et al., 2013). In this sense, if cassava germplasm characterization considers the traits of interest for breeding, the possibilities for their use are higher, as stipulated in the public policies of conservation and the use of plant genetic resources (Food and Agriculture Organization of the United Nations, 1991).

The list of the most informative morphoagronomic descriptors for cassava could reduce the number of descriptors which are currently in use in germplasm banks, without harming the representation of cassava genetic variability and the interests of breeding programs. In general, the list of the standard descriptors which were proposed in the present study (Table 3) indicates that 37% of the initial descriptors were discarded. This descriptor reduction is lower than those reported (%) in the literature, such as for: *Paspalum* sp., 86% (Strapasson et al., 2000); fig (*Ficus carica* L.), 74% (Giraldo et al., 2010); and *Capsicum* sp., 50% (Silva et al., 2013).
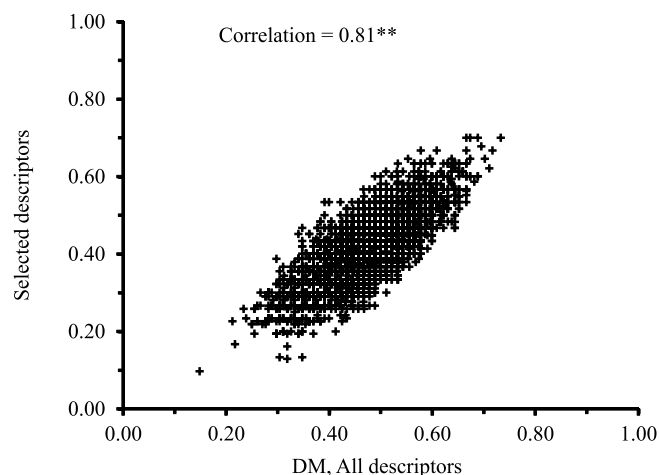

Figure 2. Correlation between the dissimilarity matrices (DM), comparing the descriptors and the final list of the most informative selected ones. **Significant at 1% of probability both by t and by Mantel tests.

Conclusions

1. The weak relationship between the different methods of descriptors selection makes it difficult the identification of the most important descriptors for the high discrimination ability using only one method.

2. The discard of 37% of the studied cassava descriptors is possible for cassava germplasm characterization.

3. For cassava characterization, the following descriptors can be discarded: pubescence on apical leaves, shape of central lobe, color of stem exterior, external color of storage root, color of root cortex, texture of root epidermis, length of leaf central lobe, width of leaf central lobe, ratio of length/width of leaf lobe, petiole length, color of stem epidermis, growth habit of stem, height to first branching, levels of branching, prominence of foliar scars, length of stipules, number of cutting stems per plant, root length, and root diameter.

4. The 32 selected cassava descriptors account for more than 90% of the total variability without significant losses for the preliminary characterization of the cassava germplasm.

Acknowledgments

References

CASTRO, J.A.; NEVES, C.G.; JESUS, O.N. de; OLIVEIRA, E.J. de. Definition of morpho-agronomic descriptors for the characterization of yellow passion fruit. **Scientia Horticulturae**, v.145, p.17-22, 2012. DOI: 10.1016/j.scienta.2012.07.022.

COFFELT, T.; JOHNSON, B.L. A set of descriptors for evaluating guayule germplasm. **Industrial Crops Products**, v.34, p.1252-1255, 2011. DOI: 10.1016/j.indcrop.2010.04.004.

CRUZ, C.D. **Programa Genes**: multivariate analysis and simulation. Viçosa: Ed. da UFV, 2006. 175p.

ESCOFIER, B.; PAGÈS, J. **Análisis factoriales simples y múltiples**: objetivos, métodos e interpretación. Bilbao: Universidad del País Vasco, 1992. 286p.

FOOD AND AGRICULTURE ORGANIZATION OF THE UNITED NATIONS. **Fourth Session of the FAO Commission on Plant Genetic Resources**. Rome: FAO, 1991. Available at: <ftp://ftp.fao.org/docrep/fao/meeting/015/aj420e.pdf>. Accessed in: 5 May 2013.

FUKUDA, W.M.G.; ALVES, A.A.C. Banco ativo de germoplasma de mandioca do Centro Nacional de Mandioca e Fruticultura. **Revista Brasileira de Mandioca**, v.6, p.65-97, 1987.

FUKUDA, W.M.G.; GUEVARA, C.L. **Descritores morfológicos e agronômicos para a caracterizaçâo de mandioca (***Manihot esculenta* **Crantz)**. Cruz das Almas: Embrapa Mandioca e Fruticultura, 1998. 38p. (Embrapa Mandioca e Fruticultura. Documentos, 78).

FUKUDA, W.M.G.; GUEVARA, C.L.; KAWUKI, R.; FERGUSON, M.E. **Selected morphological and agronomic descriptors for the characterization of cassava**. Ibadan: International Institute of Tropical Agriculture, 2010. 19p.

FUKUDA, W.M.G.; SILVA, S. de O. e; PORTO, M.C.M. **Caracterizaçâo e avaliaçâo de germoplasma de mandioca (***Manihot esculenta* **Crantz)**. Cruz das Almas: Embrapa-CNPMF, 1997. 161p.

GIRALDO, E.; LÓPEZ-CORRALES, M.; HORMAZA, J.I. Selection of the most discriminating morphological qualitative variables for characterization of fig germplasm. **Journal of the American Society for Horticultural Science**, v.135, p.240-249, 2010.

JOLLIFFE, I.T. Discarding variables in a principal component analysis. II: real data. **Journal of the Royal Statistical Society. Series C (Applied Statistics)**, v.22, p.21-31, 1973. DOI: 10.2307/2346300.

LÊ, S.; JOSSE, J.; HUSSON, F. FactorMineR: an R package for multivariate analysis. **Journal of Statistical Software**, v.25, p.1-18, 2008.

MEZETTE, T.F.; BLUMER, C.G.; VEASEY, E.A. Morphological and molecular diversity among cassava genotypes. **Pesquisa Agropecuária Brasileira**, v.48, p.510-518, 2013. DOI: 10.1590/S0100-204X2013000500007.

OLIVEIRA, E.J. de; DIAS, N.L.P.; DANTAS, J.L.L. Selection of morpho-agronomic descriptors for characterization of papaya cultivars. **Euphytica**, v.185, p.253-265, 2012. DOI: 10.1007/s10681-011-0565-0.

OLIVEIRA, E.J. de; FERREIRA, C.F.; SANTOS, V. da S.; JESUS, O.N.; OLIVEIRA, G.A.F.; SILVA, M.S. Potential of SNP markers for the characterization of Brazilian cassava germplasm. **Theoretical and Applied Genetics**, v.127, p.1423-1440, 2014. DOI: 10.1007/s00122-014-2309-8.

RAJI, A.A.; LADEINDE, T.A.O.; DIXON, A.G.O. Agronomic traits and tuber quality attributes of farmer grown cassava landraces in Nigeria. **Journal of Tropical Agriculture**, v.45, p. 9-13, 2007.

SILVA, W.C.J. e; CARVALHO, S.I.C. de; DUARTE, J.B. Identification of minimum descriptors for characterization of *Capsicum* spp. germplasm. **Horticultura Brasileira**, v.31, p.190-202, 2013. DOI: 10.1590/S0102-05362013000200004.

SOUZA, L. da S.; FARIAS, A.R.N.; MATTOS, P.L.P. de; FUKUDA, W.M.G. (Ed.). **Aspectos socioeconômicos e agronômicos da mandioca**. Cruz das Almas: Embrapa Mandioca e Fruticultura Tropical, 2006. 817p.

STRAPASSON, E.; VENCOVSKY, R.; BATISTA, L.A.R. Seleção de descritores na caracterização de germoplasma de

Paspalum sp. por meio de componentes principais. **Revista Brasileira de Zootecnia**, v.29, p.373-381, 2000. DOI: 10.1590/S1516-35982000000200008.

TONUKARI, N.J. Cassava and the future of starch. **Electronic Journal of Biotechnology**, v.7, 2004. DOI: 10.2225/vol7-issue1-fulltext-9.

VIEIRA, E.A.; FIALHO, J. de F.; FALEIRO, F.G.; BELLON, G.; FONSECA, K.G.; CARVALHO, L.J.C.B.; SILVA, M.S.; PAULA-MORAES, S.V.; OLIVEIRA, C.M.; DENKE, M.L. Characterization of sweet cassava accessions based on molecular, quantitative and qualitative data. **Crop Breeding and Applied Biotechnology**, v.11, p.232-240, 2011. DOI: 10.1590/S1984-70332011000300005.