



## Classification of cassava genotypes based on qualitative and quantitative data

E.J. Oliveira<sup>1</sup>, O.S. Oliveira Filho<sup>2</sup> and V.S. Santos<sup>1</sup>

<sup>1</sup>Embrapa Mandioca e Fruticultura, Cruz das Almas, BA, Brasil

<sup>2</sup>Universidade Federal do Recôncavo da Bahia, Cruz das Almas, BA, Brasil

Corresponding author: E.J. Oliveira

E-mail: [eder.oliveira@embrapa.br](mailto:eder.oliveira@embrapa.br)

Genet. Mol. Res. 14 (1): 906-924 (2015)

Received May 30, 2014

Accepted October 23, 2014

Published February 2, 2015

DOI <http://dx.doi.org/10.4238/2015.February.2.14>

**ABSTRACT.** We evaluated the genetic variation of cassava accessions based on qualitative (binomial and multicategorical) and quantitative traits (continuous). We characterized 95 accessions obtained from the Cassava Germplasm Bank of Embrapa Mandioca e Fruticultura; we evaluated these accessions for 13 continuous, 10 binary, and 25 multicategorical traits. First, we analyzed the accessions based only on quantitative traits; next, we conducted joint analysis (qualitative and quantitative traits) based on the Ward-MLM method, which performs clustering in two stages. According to the pseudo- $F$ , pseudo- $t^2$ , and maximum likelihood criteria, we identified five and four groups based on quantitative trait and joint analysis, respectively. The smaller number of groups identified based on joint analysis may be related to the nature of the data. On the other hand, quantitative data are more subject to environmental effects in the phenotype expression; this results in the absence of genetic differences, thereby contributing to greater differentiation among accessions. For most of the accessions, the maximum probability of classification was  $>0.90$ , independent of the trait analyzed, indicating a good fit of the clustering method. Differences in clustering according to the type of data implied that analysis of quantitative and qualitative traits in cassava germplasm

might explore different genomic regions. On the other hand, when joint analysis was used, the means and ranges of genetic distances were high, indicating that the Ward-MLM method is very useful for clustering genotypes when there are several phenotypic traits, such as in the case of genetic resources and breeding programs.

**Key words:** *Manihot esculenta* Crantz; Genetic resources; Ward-MLM; Genetic diversity; Breeding

## INTRODUCTION

Cassava (*Manihot esculenta* Crantz) is a native species from the Brazilian Amazon region. However, the crop is widely cultivated in many countries and is the staple diet of more than 500 million people. The species belongs to the Euphorbiaceae family, and hundreds of landraces have been described by taxonomists, botanists, and breeders (Rival and McKey, 2008). The wide genetic variability of cassava is, in part, the result of sexual reproduction by outcrossing, accompanied by clonal multiplication; consequently, high heterozygosity exists among the varieties cultivated by farmers.

Despite the high genetic diversity reported in cassava, the global variability tendency decreases in many production systems, because of the replacement of landraces with improved varieties (Willems et al., 2007); hence, there are concerns regarding the use and conservation of genetic resources. Moreover, the maintenance of variability is critical for achieving genetic progress in breeding programs throughout the recombination and selection cycles.

Knowledge of the genetic relations among germplasm accessions is an important component of breeding programs, because it provides information on the genetic diversity available for generating new allelic combinations in segregated populations. Accurate assessment of genetic diversity levels and patterns is crucially important for i) organization and prioritization of genetic variability in specific phenotypic traits, ii) identification of the optimal parental combinations for maximizing genetic gains in segregating populations for further selection, and iii) introgression of desirable genes from exotic germplasms in the genetic background used for the cassava production system. Therefore, any strategy related to germplasm evaluation and characterization is a powerful tool for promoting the use of cassava genetic resources.

In general, the study of genetic diversity is the process by which the variation between individuals, groups of individuals, or populations is analyzed by using specific methods or a combination of these methods. Typically, the available phenotypic data are related to the characterization and evaluation of qualitative and quantitative descriptors. However, the choice of strategy for genetic diversity detection and clustering is dependent on the germplasm bank status, the objective of the work, the required resolution level, and the available resources and infrastructure.

In previous studies of cassava, the main types of information used in genetic variability description were the pedigree and microsatellite markers (Pariyo et al., 2013), morphological and microsatellite data (Asare et al., 2011; Mezette et al., 2013), biochemical markers (Lefèvre and Charrier, 1993), and morpho-agronomic data (Kawuki et al., 2011; Muluaem et al., 2013). However, in these investigations, integration between the different information types to provide a more comprehensive review of the existing genetic variation was not performed. This integration is of critical importance, because each data set provides different information types. Given the requirement for a complete understanding of the cassava germ-

plasm, it is important simultaneously to analyze genetic diversity based on morphological and agronomic data (variables with continuous, discrete, ordinal, binomial, and multicategorical distributions), in order to obtain a clear view of the actual diversity available.

One of the first studies to propose data integration with different categories was performed by Lawrence and Krzanowski (1966), who used the location model (LM). This method classifies  $n$  individuals for  $p$  quantitative and  $q$  qualitative variables. The LM combines all levels of qualitative variables in single multinomial variable,  $W$ , with  $m$  levels. However, the use of the LM strategy in non-simulated data presents some problems, because the clustering strategy proposed by Ward (1963) groups the observations in  $g$  groups, some of which have no value for the  $W$  polynomial variable; i.e., the combination of  $m$  polynomial levels with  $g$  sub-populations produces  $m \times g$  cells, many of which are empty (Crossa and Franco, 2004).

On the basis of the problems inherent in the LM model, Franco et al. (1998) modified this model and proposed the modified location model (MLM), which considers  $m$  levels of the variable  $W$ , and  $p$  multinomial variables, independently for each sub-population. The MLM strategy includes two steps. In the first step, the groups are defined by using the clustering method of minimum variance between groups, as proposed by Ward (1963) based on the Gower distance matrix (Gower, 1971); this step is called the Ward-MLM strategy. In the second step, the MLM procedure is used to estimate the vector average of the quantitative variable for each independent sub-population, regardless of their  $W$  values. The MLM determines the probability of each accession being allocated to a specific group, thereby enabling the identification of the optimum number of clusters with high accuracy (Franco et al., 1998).

To date, several studies have used the Ward-MLM procedure to quantify variability based on quantitative and qualitative variables, in corn (Franco et al., 2005; Ortiz et al., 2008), tomato (Gonçalves et al., 2009), beans (Barbé et al., 2010; Cabral et al., 2010), and *Capsicum* spp (Sudré et al., 2010).

There are no previous reports of the use of Ward-MLM strategy in cassava, and therefore, the main objectives of the present study were i) to provide a better understanding of joint diversity analysis based on qualitative (morphological) and quantitative (agronomic) data, ii) to verify whether joint diversity analysis can facilitate the discrimination of cassava accessions, and iii) to determine whether joint analysis of phenotypic diversity can reveal new clustering patterns. Our results provide valuable information for curators and breeders focusing on research related to the conservation and use of cassava genetic resources.

## MATERIAL AND METHODS

### Plant material

From 2011 to 2013, we characterized 95 cassava accessions originating from several ecosystems in Brazil, Colombia, Venezuela, and Nigeria. The accessions were obtained from the Cassava Germplasm Bank (CGB) of Embrapa Mandioca e Fruticultura (Cruz das Almas, BA, Brazil). This germplasm bank contains landraces and improved varieties derived from conventional breeding procedures such as crossing and selection, and also the selection of landraces with high yield potential based on identification by farmers or research institutions.

We conducted our planting at the start of the rainy season (May to July) in Cruz das Almas (BA, Brazil) by sowing 15-20-cm stem cuttings in single rows. We used a spacing of 0.90 m between rows, and 0.80 m between plants, and maintained the crop according to cassava recommendations. The plants were harvested between 11 months and 12 months after planting.

## Evaluation of traits

Initially we evaluated 51 descriptors, which were classified into 13 minimum, 13 principal, 10 secondary, and 15 primary agronomic descriptors according to Fukuda et al. (2010); however 48 presented different categories (Table 1). We further grouped the descriptors into qualitative binary (10), qualitative multicategorical (25), and quantitative (13). With the exception of resistance to diseases, mites, and leaf retention, all of the data were evaluated following cassava manual descriptors (Fukuda et al., 2010).

**Table 1.** Descriptors used for the characterization and evaluation of cassava germplasm, together with the respective number of descriptor categories (K1) according to Fukuda et al. (2010), the code for each descriptor, and the type of variable analyzed.

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

<sup>1</sup>Bin = binary variable; Mult = multicategorical variable; Quant = quantitative variable.

The severities of anthracnose (*Colletotrichum gloeosporioides*) and bacterial blight (*Xanthomonas axonopodis* pv. *manihot*) were evaluated under field conditions, and symptoms were observed at 10 months after planting. For bacteriosis, we used the following scale: 0, no symptoms; 1, symptoms (angular spots) on leaves only; 2, presence of necrotic lesions on stems or petioles; 3, severe symptoms on leaves and/or presence of necrotic lesions with gum exudation; and 4, whole leaves lost, with apical death or plant death. For anthracnose, we used the following scale: 0, no symptoms; 1, presence of small cancers in the lower half of the plant; 2, presence of deep cancers in the upper half of the plant; 3, presence of deep cancers with sporulation, leaf distortion, or wilting and apex drying; 4, apical or plant death.

Regarding tolerance to mites, we used a scale ranging from 1 to 4 as follows: 1, no damage or presence of a few whitish yellow spots from the leaf base to the apical bud; 2, presence of a moderate number of yellow spots on a few leaves; 3, presence of abundant yellow spots on the leaves of the middle third of the plant, accompanied by apical bud deformation; 4, presence of severe leaf deformation of the apical bud, accompanied by whitish leaves, defoliation, and twisted or dead apical bud.

For leaf retention, we used the following scale: 1, low leaf retention, located only at the plant apex; 2, moderate leaf retention, with leaves being distributed in  $\leq 15\%$  of the volume of the aerial part of the plant; 3, high leaf retention, with leaves being distributed in  $> 15\%$  of the volume of the aerial part of the plant.

## Data analysis

We used two strategies for data analysis. In the first strategy, we considered only quantitative variables; in the second strategy, namely, joint analysis, we simultaneously analyzed qualitative (binary and multicategorical) and quantitative data. For both strategies, we used the Ward-MLM method.

Initially, we identified the similarity groups by using the Ward clustering strategy (Ward, 1963), based on the Gower distance matrix (Gower, 1971). Next, we used the MLM procedure to estimate a quantitative variable mean vector for each subgroup, regardless of the  $W$  value. Finally, we simultaneously analyzed the quantitative and qualitative data, by using the Ward-MLM procedure to group the genotypes according to the CLUSTER and IML procedures of SAS (SAS Institute, 2011).

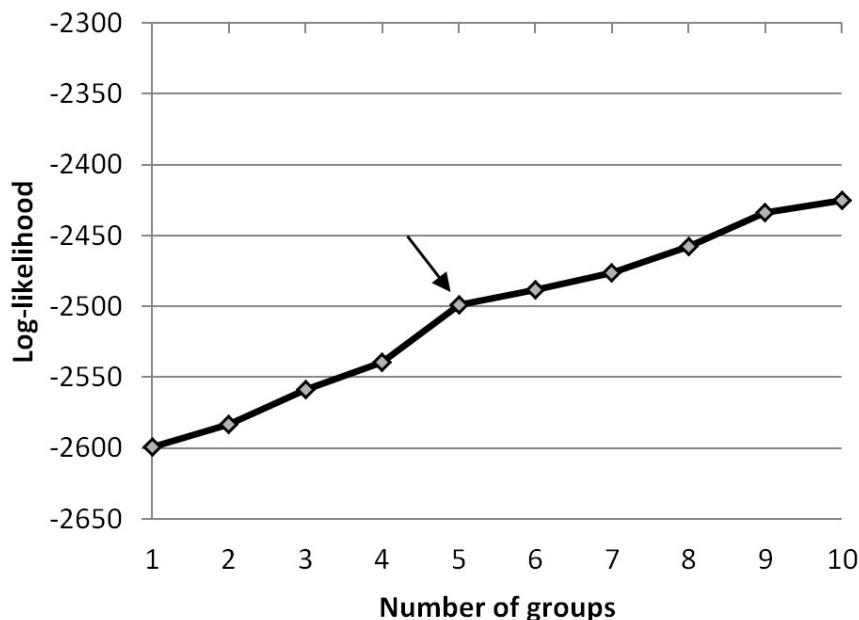
We defined the optimal number of groups based on the logarithmic likelihood function, according to the pseudo- $F$  and pseudo- $t^2$  criteria, combined with the likelihood profile associated with the likelihood ratio test. The difference between groups and data correlation with the canonical variable was plotted by using the CANDISC procedure of SAS (SAS Institute, 2011).

## RESULTS AND DISCUSSION

### Quantitative trait analysis

According to the pseudo- $F$  and pseudo- $t^2$  criteria, and the increased likelihood caused by the larger cluster number, the optimal number of groups required to represent the genetic variability derived by using quantitative data was five, with a maximum absolute value of 40.48 (Figure 1 and Table 2). The indication of the number of groups is an innovative aspect of

the Ward-MLM procedure compared with other hierarchical methods, and it results in a more accurate and less subjective cluster formation.



**Figure 1.** Distributions of the log-likelihood function according to the number of groups, based on quantitative data.

**Table 2.** Number of groups formed according to the Ward-MLM method based on the log-likelihood function (log-likelihood) and its increment to each group, derived by using quantitative data.

Number of clusters	Log-likelihood	Increment
1	-2599.27	0.00
2	-2583.54	15.74
3	-2558.89	24.64
4	-2539.22	19.67
5	-2498.74	40.48*
6	-2488.31	10.43
7	-2476.18	12.14
8	-2457.95	18.23
9	-2434.01	23.93
10	-2424.96	9.05

\*Greater increment.

The number of accessions per cluster ranged from four in Group 5 to 47 in Group 2 (Table 3). In general, the groups were reasonably consistent in terms of maximum likelihood accession assignment, the values of which ranged from 0.58 to 1.00 (Table 3). This range seems to be relatively high; however, only six accessions had a maximum likelihood assignment of <0.90. These accessions were BGM1153, BGM0116, and BGM1681 in Group 1 (maximum likelihood assignment values of 0.58, 0.65, and 0.69, respectively) and BGM1198, BGM1490, and BGM1549 in Group 2 (maximum likelihood assignment values of 0.61, 0.82, and 0.84, respectively). Franco et al. (1988) previously used several practical examples to

demonstrate that the use of the MLM method enabled an average maximum likelihood classification of >0.90 for most individuals, and that the proportion of individuals classified with a low probability (i.e., <0.50) was <5%.

**Table 3.** Clustering formed according to the Ward-MLM method, by using quantitative data in cassava germplasm accessions.

Group	No. of accessions	Range of assignment probability	Accessions
1	28	0.58-1.00	BGM0116, BGM0145, BGM0155, BGM0179, BGM0212, BGM0298, BGM0399, BGM0472, BGM1130, BGM1153, BGM1291, BGM1447, BGM1515, BGM1518, BGM1579, BGM1608, BGM1610, BGM1622, BGM1645, BGM1671, BGM1672, BGM1681, BGM1685, BGM1692, BGM1716, BGM1726, BGM1814, BGM2018
2	47	0.61-1.00	BRS Verdinha, BRS Tapioqueira, 9624-09, BRS Caipira, BGM0162, BGM0164, BGM0215, BGM0236, BGM0324, BGM0374, BGM0388, BGM0426, BGM0510, BGM0517, BGM0557, BGM0579, BGM0846, BGM0873, BGM1024, BGM1067, BGM1078, BGM1118, BGM1198, BGM1203, BGM1254, BGM1280, BGM1362, BGM1363, BGM1367, BGM1413, BGM1432, BGM1457, BGM1481, BGM1490, BGM1521, BGM1549, BGM1586, BGM1619, BGM1620, BGM1643, BGM1666, BGM1677, BGM1701, BGM1721, BGM1728, BGM1732, BGM1942, 98150-06, BGM0440, BGM0471, BGM0660, BGM0943, BGM1226, BGM1535, BGM1611, BGM1667, BGM1822
3	10	0.95-1.00	BGM0226, BGM0489, BGM0576, BGM1138, BGM1456, BGM1884
4	6	0.94-1.00	BGM1366, BGM1669, BGM1956, BGM1957
5	4	1.00	

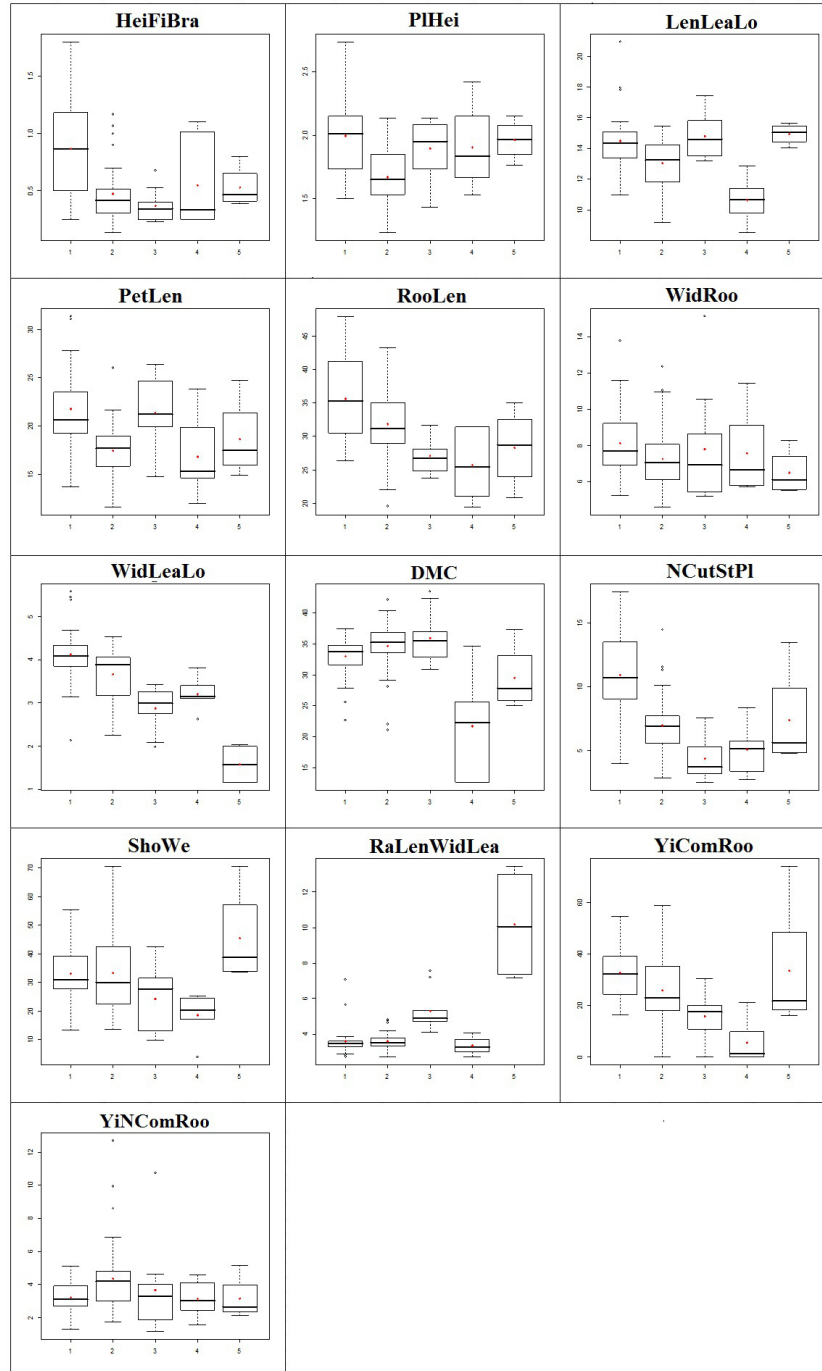
Group 1 was composed of 28 germplasm accessions and was characterized by having high amplitude for *HeiFiBra*, *PIHei*, *PetLen*, *RooLen*, *NCutStPl*, and *ShoWe* descriptors (Figure 2). However, in general, the main characteristics of this group were the largest values of height to first branching, plant height, root length and width, leaf lobe width, number of cutting stems per plant, and yield of commercial root.

Group 2 was composed of 43 germplasm accessions and four clones from the breeding program (namely, BRS Verdinha, BRS Tapioqueira, 9624-09, and BRS Caipira). The main characteristics of this group were i) smaller values of plant height and height to first branching, petiole length, number of cutting stems per plant, leaf lobe length/width ratio compared with all the other groups; ii) high values of leaf lobe width and dry matter content; and iii) moderate yield of commercial root and high yield of non-commercial root.

Group 3 was composed of nine germplasm accessions and clone 98150-06. The main characteristics of this group were lower height to first branching, but high plant height. In addition, we observed high average values of leaf lobe length and petiole length, small root length, high dry matter content, and a low number of cutting stems per plant (lower than the yield of commercial root).

Group 4 was composed of six germplasm accessions. The main characteristics of this group were i) low values of petiole length, leaf lobe length, and root length; and ii) low values of dry matter content, number of cutting stems per plant, shoot weight, and yield of commercial root.

Group 5 was composed of four germplasm accessions and was characterized by i) high values of plant height and leaf lobe length; and ii) low values of root length and width, leaf lobe width, dry matter content, and number of cutting stems per plant. Moreover, this group was distinctive because of its high shoot weight, ratio of leaf lobe length/width, and moderate yield of commercial root.



**Figure 2.** Boxplots of the mean values for quantitative traits of cassava germplasm accessions in each group formed according to the Ward-MLM method.



The results of canonical analysis of the variables obtained by using the Ward-MLM method for the 13 quantitative traits revealed that the characteristics showing the largest contributions to the genetic divergence were *WidLeaLo*, *RaLenWidLea*, and *NCutStPl*, with absolute values of 0.748, -0.659, and 0.591, respectively, in the first canonical variable. Regarding the second and third canonical variables, *RaLenWidLea* and *DMC* showed the largest contributions, with absolute values of 0.652 and 0.641, respectively (Table 4). By contrast, the results of analysis of the distribution of quantitative traits in the boxplot (Figure 2) revealed that only the yield of non-commercial root was not useful for discriminating accessions in the different groups formed based on quantitative data. This observation was confirmed by the results of analysis based on the canonical variable (Table 4).

**Table 4.** Estimates of the canonical variable analysis obtained according to the Ward-MLM method, by using quantitative data in cassava accessions.

Variable	Can1	Can2	Can3
<i>HeiFiBra</i>	0.467	0.340	-0.101
<i>PIHei</i>	0.144	0.418	-0.320
<i>LenLeaLo</i>	0.044	0.332	0.277
<i>PetLen</i>	0.206	0.282	-0.107
<i>RooLen</i>	0.477	0.115	0.274
<i>WidRoo</i>	0.160	0.030	-0.124
<i>WidLeaLo</i>	0.748	-0.251	0.056
<i>DMC</i>	0.025	-0.252	0.641
<i>NCutStPl</i>	0.591	0.376	0.154
<i>ShoWe</i>	0.046	0.214	0.449
<i>RaLenWidLea</i>	-0.659	0.652	0.193
<i>YiComRoo</i>	0.308	0.285	0.484
<i>YiNComRoo</i>	-0.070	-0.251	0.229

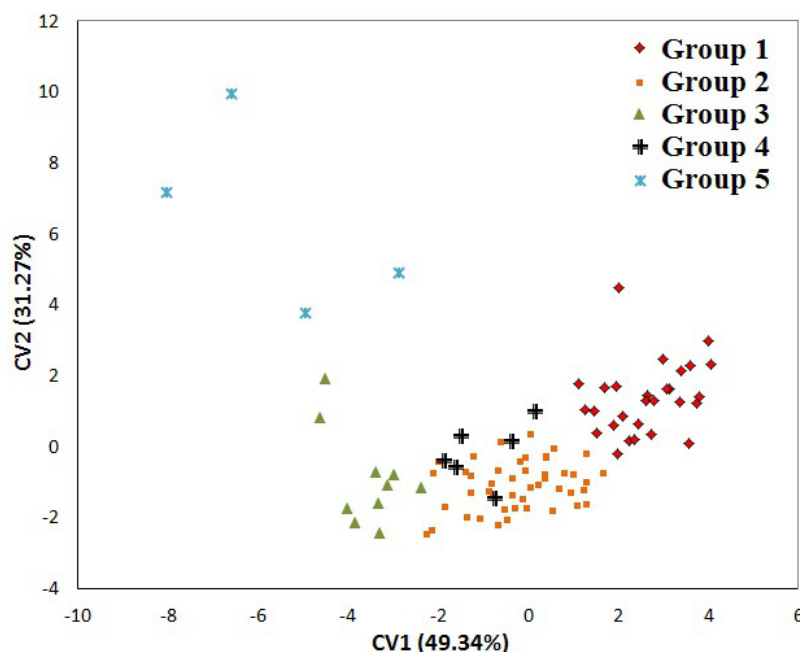
The first two canonical variables accounted for 80.61% of the quantitative data variation, indicating a good graphical representation of the cassava accessions (Figure 3). On the basis of graphic dispersion and the distances between pairs of groups (Table 5), Group 5 was the generally more divergent than Groups 1-4. Additionally, Group 1 was similar to Group 2 (distance of 15.70); however, these two groups differed in some characteristics, such as *HeiFiBra*, *PIHei*, *PetLen*, *RooLen*, *NCutStPl*, and *YiComRoo*. On the basis of canonical variable analysis, Group 4 was similar to Group 2; however, these two groups differed in some characteristics, such as *LenLeaLo*, *RooLen*, *DMC*, and *YiComRoo* (Figure 2).

In common bean, Cabral et al. (2010) used the Ward-MLM procedure to demonstrate a greater clustering separation of accessions belonging to Andean groups than of accessions belonging to Mesoamerican groups; the largest seed size of Andean groups was a key descriptor for the best discrimination.

### Joint analysis of quantitative and qualitative traits

The results of joint analysis based on the Ward-MLM method indicated the formation of four divergent groups according to the pseudo-*F* and pseudo-*t*<sup>2</sup> criteria, taking into account the fact that the maximum value (107.30) of increased likelihood occurs at this point (Figure 4 and Table 6). The smaller number of groups obtained by using joint analysis than by using quantitative trait analysis can be explained by the fact that quantitative traits are

markedly influenced by the environment. Therefore, there is a high variability associated with uncontrolled variations. When qualitative traits were grouped with quantitative data, some accessions previously present in different groups were clustered, possibly because of small phenotypic differences that were detected only by using quantitative trait analysis.



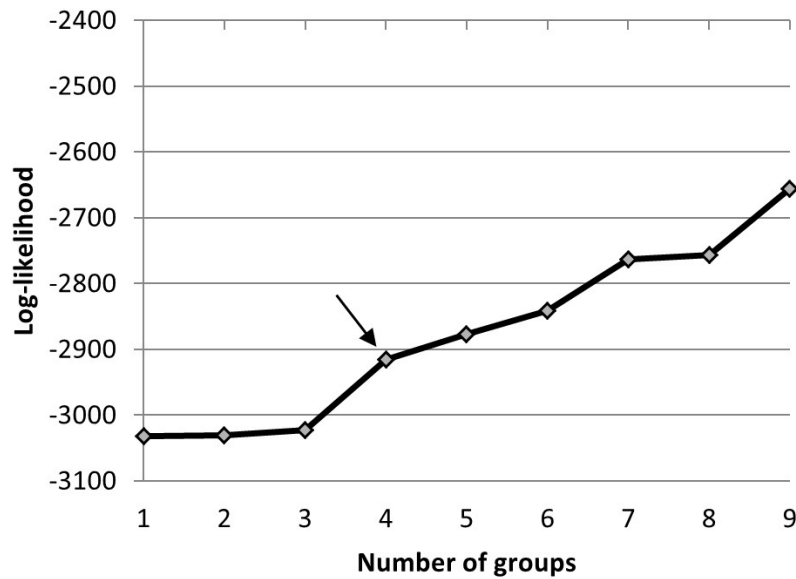
**Figure 3.** Dispersion of cassava accessions based on the first two canonical variables obtained by using quantitative data analysis, according to the five groups identified by using the Ward-MLM method.

**Table 5.** Distances between groups formed according to the Ward-MLM method based on quantitative data.

Group	2	3	4	5
1	15.70	45.43	31.29	97.07
2	28.18	21.46	86.84	
3		26.93	69.23	
4		89.50		

In principle, qualitative data are expected to provide additional information on groups, and therefore a larger number of groups. In the present study, the smaller number of groups may indicate a lower reliability in the clustering. Nevertheless, the increased quantitative (Table 2) and joint analysis (Table 6) likelihood data are relevant in indicating five and four groups, respectively. Previous studies have demonstrated that pseudo- $F$  and pseudo- $t^2$  analysis resembles a hypothesis test that compares the probability of the average vectors of two clusters to form a single group in each clustering stage. In this case, the highest pseudo- $F$  or pseudo- $t^2$  values are related to a lower probability of the significance test, and consequently a low probability of rejecting the hypotheses of equal means. If the mean vector equality is rejected,

the two groups are not clustered and the test will be performed with other groups. Therefore, according to Mingoti (2007), this methodology indicates the number of groups with relatively high accuracy.



**Figure 4.** Distributions of the log-likelihood function according to the number of groups, based on joint analysis.

**Table 6.** Number of groups formed according to the Ward-MLM method based on the log-likelihood function (log-likelihood) and its increment to each group, derived by using joint analysis.

Number of clusters	Log-likelihood	Increment
1	-3031.89	0.00
2	-3030.33	1.56
3	-3023.18	7.15
4	-2915.88	107.30*
5	-2877.72	38.16
6	-2842.01	35.71
7	-2763.72	78.29
8	-2757.30	6.42
9	-2656.59	100.71

\*Greater increment.

When the accessions were grouped based on quantitative and qualitative traits, the number of accessions per group ranged from 17 in Group 4 to 28 in Group 2 (Table 7). The maximum likelihood accession assignment in different groups ranged from 0.73 to 1.00 (Table 7); however, only accessions BGM1667 ( $P = 0.73$ ) in Group 1 and BGM0471 ( $P = 0.88$ ) in Group 3 had assignment probabilities of  $<0.90$ .

Group 1 was composed of 27 germplasm accessions and was characterized by having a wide range for most of the quantitative descriptors (Figure 5). However, in general, the main agronomic characteristics of this group were the large values of height to first branching, plant

height, root length and width, leaf lobe width, number of cutting stems per plant, and yield of commercial root. The results of distribution analysis for the qualitative descriptor classes among the four groups formed by Ward-MLM method revealed no particular descriptors that better characterized certain groupings (Tables 8-11). In general, most of the descriptor classes were present in all of the clusters. Therefore, specific morphological characters were not attributed to certain diversity groups. Previous studies have indicated no strong association between physiological characteristics and groupings formed by using simultaneous analysis of various data types. Cabral et al. (2010) analyzed 57 common bean accessions based on five agronomic traits, five morphological descriptors, and 16 microsatellite markers and observed a relatively homogeneous distribution among the different descriptor classes in all of the groups analyzed.

**Table 7.** Clustering formed according to the Ward-MLM method, by using joint analysis in cassava germplasm accessions.

Group	No. of accessions	Range of assignment probability	Accessions
1	27	0.73-1.00	BGM0145, BGM0215, BGM0226, BGM0399, BGM0426, BGM0510, BGM0517, BGM0576, BGM0660, BGM0943, BGM1138, BGM1198, BGM1203, BGM1226, BGM1363, BGM1366, BGM1456, BGM1490, BGM1515, BGM1535, BGM1622, BGM1667, BGM1685, BGM1701, BGM1728, BGM1822, BGM1884
2	28	0.97-1.00	BRS Tapioqueira, BGM0116, BGM0162, BGM0164, BGM0179, BGM0236, BGM0324, BGM0388, BGM0440, BGM0557, BGM0579, BGM0846, BGM0873, BGM1024, BGM1078, BGM1118, BGM1254, BGM1362, BGM1413, BGM1481, BGM1579, BGM1620, BGM1669, BGM1671, BGM1681, BGM1942, BGM1956, BGM1957
3	23	0.88-1.00	BRS Verdinha, 98150-06, BGM0155, BGM0298, BGM0374, BGM0471, BGM0489, BGM1067, BGM1153, BGM1280, BGM1457, BGM1518, BGM1521, BGM1549, BGM1611, BGM1619, BGM1643, BGM1645, BGM1677, BGM1721, BGM1726, BGM1732, BGM1814
4	17	0.98-1.00	9624-09, BRS Caipira, BGM0212, BGM0472, BGM1130, BGM1291, BGM1367, BGM1432, BGM1447, BGM1586, BGM1608, BGM1610, BGM1666, BGM1672, BGM1692, BGM1716, BGM2018

Group 2 was composed of 27 accessions and clone BRS Tapioqueira (Table 7). This group was characterized by i) low values of plant height and number of cutting stems per plant, ii) high values of leaf lobe width and dry matter content, and iii) a moderate yield of commercial root.

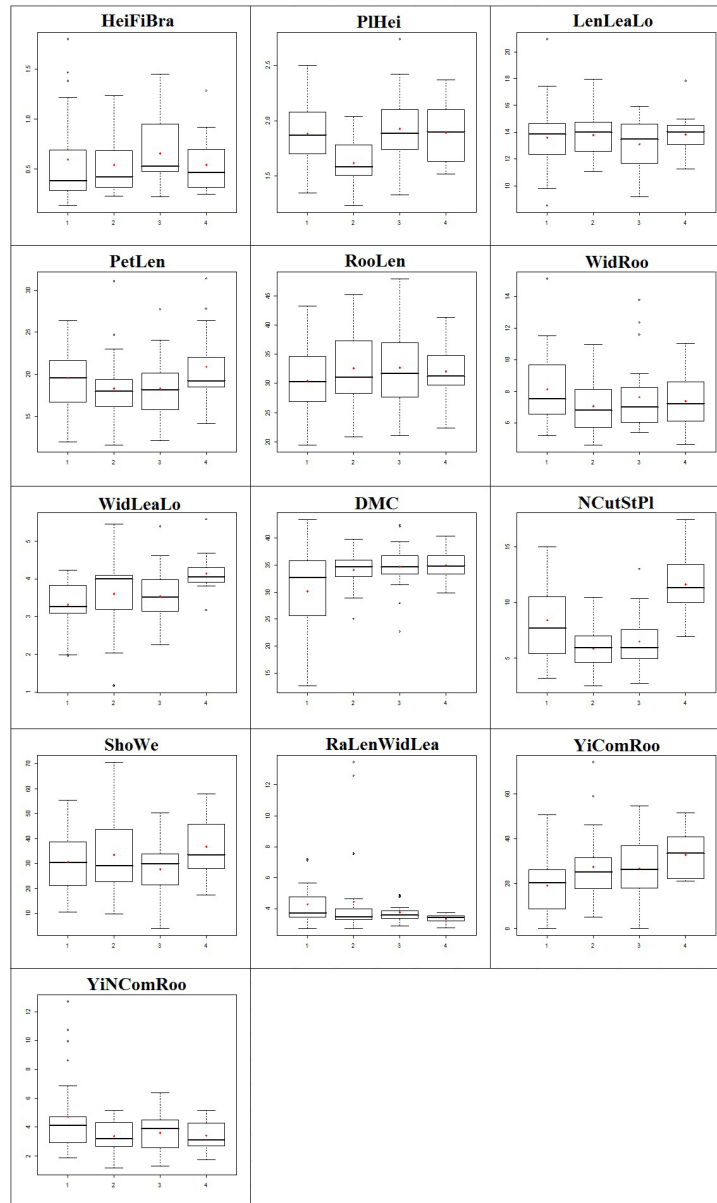
Group 3 was composed of 21 germplasm accessions and two clones (BRS Verdinha and 98150-06) and was characterized by high values of height to first branching and dry matter content, a low number of cutting stems per plant, and a moderate yield of commercial root.

Group 4 was composed of 15 germplasm accessions and two clones (9624-09 and BRS Caipira). The main characteristics of this group were high values of petiole length, leaf lobe width, dry matter content, number of cutting stems per plant, and yield of commercial root (Figure 5).

Previous studies have indicated more reliable clustering when agronomic, morphological, and molecular data were simultaneously analyzed by using the Ward-MLM method (Ortiz et al., 2008; Cabral et al., 2010; Sudré et al., 2010). For example, the use of this strategy in *Capsicum* enabled the correct grouping of 56 accessions according to their respective species, i.e., *Capsicum annuum*, *Capsicum frutescens*, *Capsicum chinense*, and *Capsicum baccatum* (Sudré et al., 2010).

In the present study, comparison between the groups formed by using quantitative data and the groups formed based on joint analysis indicated a weak association among the acces-

sions belonging to each group (0.38). This may indicate that the quantitative and qualitative data analysis considered different genomic regions. Our finding is in accordance with the results of Alves et al. (2013), who reported a relatively low correlation between dissimilarity matrices based on phenotypic and molecular data (RAPD and SSR) for the physic nut, implying that the molecular data sampled low phenotypic differentiation regions among accessions in this species.



**Figure 5.** Boxplots of the mean values for quantitative traits of cassava germplasm accessions in each group formed according to the Ward-MLM method based on joint analysis.

**Table 8.** Distribution of cassava accessions based on qualitative minimum descriptors, according to the groups formed by using the Ward-MLM method based on joint analysis.

Minimum descriptor	Class	G1	G2	G3	G4
Color of apical leaves	Dark green	0.15	0.29	0.00	0.18
	Purplish green	0.78	0.64	0.78	0.59
	Light green	0.07	0.07	0.22	0.24
Pubescence on apical leaves	Present	0.30	0.32	0.35	0.18
	Absent	0.70	0.68	0.65	0.82
Shape of central leaflet	Lanceolate	0.00	0.07	0.13	0.12
	Straight or linear	0.96	0.82	0.87	0.88
	Elliptic-lanceolate	0.04	0.07	0.00	0.00
	Pandurate	0.00	0.04	0.00	0.00
Petiole color	Greenish-red	0.07	0.11	0.00	0.00
	Green	0.11	0.21	0.13	0.12
	Red	0.22	0.18	0.39	0.24
	Reddish-green	0.33	0.32	0.17	0.47
Color of stem cortex	Yellowish-green	0.26	0.18	0.30	0.18
	Light green	0.00	0.04	0.00	0.00
	Dark green	0.19	0.25	0.09	0.41
Color of stem exterior	Orange	0.81	0.71	0.91	0.59
	Silver	0.07	0.11	0.04	0.24
	Light brown	0.11	0.07	0.00	0.12
	Dark brown	0.26	0.21	0.09	0.18
	Orange	0.11	0.04	0.35	0.12
	Greenish-yellowish	0.00	0.07	0.00	0.06
	Gray	0.44	0.50	0.52	0.29
Length of phyllotaxy	Medium	0.00	0.00	0.09	0.12
	Long	0.78	0.96	0.78	0.88
	Short	0.22	0.04	0.13	0.00
Extent of root peduncle	Mixed	0.00	0.00	0.04	0.00
	Pedunculate	0.11	0.04	0.00	0.00
	Sessile	0.89	0.96	0.96	1.00
External color of storage root	Dark brown	0.07	0.04	0.43	0.18
	White or cream	0.04	0.00	0.00	0.00
	Light brown	0.07	0.11	0.09	0.00
	Yellow	0.81	0.86	0.48	0.82
Color of root cortex	White or cream	0.63	0.71	0.65	0.53
	Purple	0.22	0.07	0.17	0.06
	Yellow	0.00	0.07	0.13	0.06
	Pink	0.15	0.14	0.04	0.35
Color of root pulp	Cream	0.74	0.96	0.83	0.65
	White	0.15	0.04	0.17	0.29
	Yellow	0.11	0.00	0.00	0.06
Texture of root epidermis	Rough	0.07	0.07	0.43	0.18
	Smooth	0.93	0.93	0.57	0.82
Flowering	Absent	0.63	0.43	0.70	0.41
	Present	0.37	0.57	0.30	0.59

**Table 9.** Distribution of cassava accessions based on qualitative principal descriptors, according to the groups formed by using the Ward-MLM method based on joint analysis.

Principal descriptor	Class	G1	G2	G3	G4
Leaf color	Dark green	0.96	0.89	0.96	1.00
	Light green	0.04	0.11	0.04	0.00
Number of leaf lobes	7	0.04	0.00	0.00	0.00
	5	0.11	0.21	0.17	0.18
	3	0.81	0.75	0.83	0.82
	9	0.04	0.04	0.00	0.00
Color of stem epidermis	Light brown	0.11	0.04	0.35	0.06
	Cream	0.26	0.39	0.17	0.41
	Dark brown	0.63	0.57	0.48	0.53
Color of end branches of adult plant	Green-purple	0.52	0.25	0.52	0.76
	Green	0.48	0.75	0.48	0.24
Levels of branching	1x	0.37	0.46	0.48	0.59
	2x	0.26	0.21	0.30	0.18
	4x	0.26	0.29	0.22	0.24
	3x	0.11	0.04	0.00	0.00
Root constrictions	Some	0.30	0.25	0.22	0.18
	Few to none	0.59	0.71	0.57	0.71
	Many	0.11	0.04	0.22	0.12

**Table 10.** Distribution of cassava accessions based on qualitative secondary descriptors, according to the groups formed by using the Ward-MLM method based on joint analysis.

Secondary descriptor	Class	G1	G2	G3	G4
Color of leaf vein	Reddish-green (>50% of the lobe)	0.85	0.86	1.00	0.82
	Green	0.04	0.14	0.00	0.18
	Reddish-green (<50% of the lobe)	0.07	0.00	0.00	0.00
	Red	0.04	0.00	0.00	0.00
Peduncle position	Horizontal	0.00	0.00	0.04	0.06
	Irregular	0.96	0.89	0.83	0.94
	Tilted up	0.04	0.11	0.13	0.00
Stipule margin	Entire	0.96	1.00	1.00	1.00
	Split or forked	0.04	0.00	0.00	0.00
Branching habit	Dichotomous	0.00	0.00	0.22	0.00
	Trichotomous	0.89	0.82	0.70	0.94
Angle of branching	Erect	0.11	0.18	0.09	0.06
	90°	0.00	0.04	0.00	0.00
	100°	0.11	0.14	0.22	0.00
	110°	0.70	0.46	0.48	0.53
	80°	0.11	0.32	0.22	0.35
Sinuosity of the leaf lobe	70°	0.07	0.04	0.09	0.12
	Smooth	0.81	0.89	0.78	0.71
Root shape	Sinuuous	0.19	0.11	0.22	0.29
	Conical-cylindrical	0.00	0.00	0.04	0.00
	Cylindrical	0.93	1.00	0.83	0.94
	Irregular	0.04	0.00	0.09	0.06
	Conical	0.04	0.00	0.04	0.00
Shape of plant	Open	0.44	0.39	0.48	0.41
	Compact	0.48	0.54	0.35	0.53
	Umbrella	0.04	0.07	0.04	0.06
	Cylindrical	0.04	0.00	0.13	0.00

**Table 11.** Distribution of cassava accessions based on qualitative preliminary agronomic traits, according to the groups formed by using the Ward-MLM method based on joint analysis.

Preliminary agronomic descriptor	Class	G1	G2	G3	G4
Vigor	High	0.22	0.04	0.17	0.18
	Medium	0.56	0.79	0.57	0.65
	Low	0.22	0.18	0.26	0.18
Periderm: ease of peeling	Difficult	0.22	0.39	0.48	0.53
	Easy	0.78	0.61	0.52	0.47
Cortex: ease of peeling	Easy	0.15	0.00	0.04	0.06
	Difficult	0.85	1.00	0.96	0.94
Root position	Irregular	0.04	0.04	0.00	0.00
	Horizontal	0.96	0.96	1.00	1.00
Resistance to bacterial blight	Grade 0	0.59	0.64	0.39	0.29
	Grade 1	0.30	0.15	0.26	0.24
	Grade 2	0.07	0.14	0.26	0.29
	Grade 3	0.00	0.07	0.09	0.06
	Grade 4	0.04	0.00	0.00	0.12
Resistance to anthracnose	Grade 0	0.26	0.52	0.48	0.65
	Grade 1	0.26	0.21	0.22	0.29
	Grade 2	0.00	0.00	0.04	0.00
	Grade 3	0.44	0.25	0.26	0.06
	Grade 4	0.04	0.04	0.00	0.00
Tolerance to mites	Grade 1	0.52	0.54	0.48	0.65
	Grade 2	0.22	0.14	0.26	0.24
	Grade 3	0.07	0.14	0.13	0.00
	Grade 4	0.19	0.18	0.13	0.12
Leaf retention	Grade 2	0.33	0.32	0.22	0.12
	Grade 1	0.63	0.64	0.61	0.82
	Grade 3	0.04	0.04	0.17	0.06

It seems that the selective forces in cassava acted in different directions when quantitative and qualitative descriptors were analyzed simultaneously. Additionally, quantitative traits are more strongly affected by the environment, and this may contribute to the inclusion of non-genetic variation in quantitative data. By contrast, Kawuki et al. (2011) conducted a comprehensive survey of genetic diversity in cassava germplasm from southern, eastern, and central regions of Africa, by evaluating 29 qualitative and four quantitative traits (dry matter content, harvest index, leaf retention, and root cortex thickness). The authors reported that qualitative traits showed a low discriminating power for cassava accessions, whereas quantitative traits showed high genetic variation, and were therefore suitable for obtaining high genetic gain in breeding programs. Thus, regardless of the advantages and disadvantages of quantitative and qualitative traits, both types of data are necessary to conduct a realistic analysis of genetic variability in *Manihot esculenta* Crantz.

The results of canonical variable analysis for the 13 quantitative, 10 binary, and 25 multicategorical traits revealed that the traits showing the largest contribution to the genetic divergence were *NCutStPl* for the first and second canonical variables (0.53 and 0.49, respectively), *WidLeaLo* for the second canonical variable (0.46), and *PIHei* for the third canonical variable (0.59). Interestingly, the *WidLeaLo* and *NCutStPl* traits were also important during the analysis of quantitative data. By contrast, the *PIHei* trait was of considerable importance when the qualitative and quantitative data were simultaneously analyzed, whereas the *DMC* trait was important only during the analysis of quantitative data (Table 12).

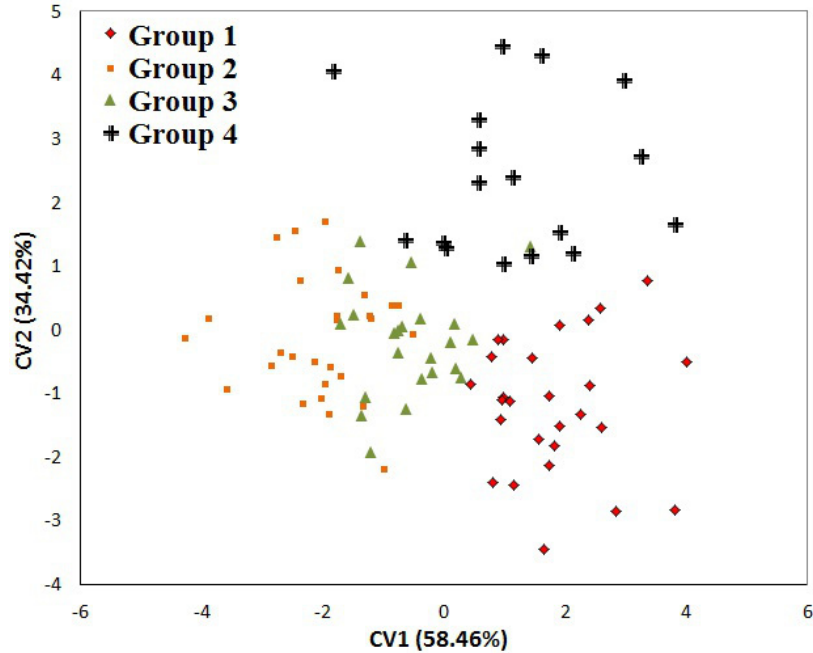
**Table 12.** Estimates of the canonical variables obtained according to the Ward-MLM method, by using joint analysis in cassava accessions based on the Ward-MLM method.

Variable	Can1	Can2	Can3
<i>HeiFiBra</i>	0.03	-0.07	0.24
<i>PIHei</i>	0.40	0.05	0.59
<i>LenLeaLo</i>	0.00	0.06	-0.28
<i>PetLen</i>	0.23	0.19	-0.13
<i>RooLen</i>	-0.15	0.08	0.12
<i>WidRoo</i>	0.21	-0.13	0.10
<i>WidLeaLo</i>	-0.03	0.46	-0.04
<i>DMC</i>	-0.28	0.34	0.27
<i>NCutStPl</i>	0.53	0.49	-0.18
<i>ShoWe</i>	0.00	0.22	-0.35
<i>RaLenWidLea</i>	-0.10	-0.25	-0.27
<i>YiComRoo</i>	-0.17	0.40	0.08
<i>YiNComRoo</i>	0.27	-0.26	-0.08

Regarding the results of joint analysis, the first two canonical variables explained 92.88% of the quantitative data variation (Figure 6), indicating a marked improvement in the graphical cassava accession representation; in comparison, the first two canonical variables accounted for only 80.61% of the quantitative data variation (Figure 3).

On the basis of graphic dispersion and distances between pairs of groups (Table 13), Groups 1 and 3 were similar, whereas Group 4 was the most divergent (distances ranging from 10.74 to 16.56). Groups 1 and 3 showed differences only regarding *DMC*, *NcutStPl*, and *YiComRoo* (Figure 5).





**Figure 6.** Dispersion of cassava accessions based on the first two canonical variables obtained by using joint analysis, according to the four groups identified by using the Ward-MLM method.

**Table 13.** Distances between groups formed according to the Ward-MLM method based on joint analysis.

Group	2	3	4
1	15.70	7.94	13.18
2		4.10	16.56
3			10.74

The average distances (Gower, 1971) between groups based on quantitative data and joint analysis were 0.41 and 0.63, respectively. The variation of genetic distances was higher for joint analysis (0.34-0.85) than for quantitative analysis (0.21-0.67), indicating that the simultaneous analysis of different variables provided a better representation of the actual genetic variability in the data set analyzed.

Phenotypic variation in cassava has previously been evaluated based on quality (Benesi, 2005; Kizito, 2006) and morpho-agronomic traits (Kawuki et al., 2011; Muluaem et al., 2013). However, few studies to evaluate these traits jointly, especially by using the Ward-MLM method, have been conducted. In the present study, we have shown that i) the results of cassava diversity analysis based on qualitative traits differ from those obtained by using quantitative data; ii) joint analysis of qualitative and quantitative traits enables different genomic regions to be sampled, and this can potentially increase the genetic diversity of segregated populations when the information obtained is used to select contrasting parental lines; iii) the results of joint analysis of phenotypic diversity reveal a relatively high variability of cassava,

and this information may be valuable for increasing the use of germplasm accessions in breeding programs, by indicating groups with superior agronomic traits.

## ACKNOWLEDGMENTS

Research supported by Fundação de Amparo à Pesquisa do Estado da Bahia (Fapesb) and Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq).

## REFERENCES

- Alves AA, Bhering LL, Rosado TB, Laviola BG, et al. (2013). Joint analysis of phenotypic and molecular diversity provides new insights on the genetic variability of the Brazilian physic nut germplasm bank. *Genet. Mol. Biol.* 36: 371-381.
- Asare PA, Galyuon IKA, Sarfo JK and Tetteh JP (2011). Morphological and molecular based diversity studies of some cassava (*Manihot esculenta* crantz) germplasm in Ghana. *Afr. J. Biotechnol.* 10: 13900-13908.
- Barbé TC, Amaral Júnior AT, Gonçalves LSA and Rodrigues R (2010). Association between advanced generations and genealogy in inbred lines of snap bean by the Ward-Modified Location Model. *Euphytica* 173: 337-343.
- Benesi IRM (2005). Characterisation of Malawian Cassava Germplasm for Diversity, Starch Extraction and its Native and Modified Properties. Doctoral thesis. Department of Plant Sciences: Plant Breeding, Faculty of Natural and Agricultural Sciences, University of the Free State, Bloemfontein, South Africa.
- Cabral PDS, Soares TCB, Gonçalves LSA and Amaral Júnior AT (2010). Quantification of the diversity among common bean accessions using Ward-MLM strategy. *Pesqui. Agropecu. Bras.* 45: 1124-1132.
- Crossa J and Franco J (2004). Statistical methods for classifying genotypes. *Euphytica* 137: 19-37.
- Franco J, Crossa J, Villaseñor J and Taba S (1988). Classifying genetic resources by categorical and continuous variables. *Crop Sci.* 38: 1688-1696.
- Franco J, Crossa J, Taba S and Shands H (2005). A sampling strategy for conserving genetic diversity when forming core subsets. *Crop Sci.* 45: 1035-1044.
- Fukuda WMG, Guevara CL, Kawuki R and Ferguson ME (2010). Selected Morphological and Agronomic Descriptors for the Characterization of Cassava. International Institute of Tropical Agriculture (IITA), Ibadan.
- Gonçalves LS, Rodrigues R, do Amaral Júnior AT, Karasawa M, et al. (2009). Heirloom tomato gene bank: assessing genetic divergence based on morphological, agronomic and molecular data using a Ward-modified location model. *Genet. Mol. Res.* 8: 364-374.
- Gower JC (1971). A general coefficient of similarity and some of its properties. *Biometrics* 27: 857-874.
- Kawuki RS, Ferguson M, Labuschagne MT and Herselman L (2011). Variation in qualitative and quantitative traits of cassava germplasm from selected national breeding programmes in sub-Saharan Africa. *Field Crops Res.* 122: 151-156.
- Kizito EB (2006). Genetic and Root Growth Studies in Cassava (*Manihot esculenta* Crantz): Implications for breeding. Doctoral thesis. Faculty of Natural Resources and Agricultural Sciences, Department of Plant Biology and Forest Genetics, Swedish University of Agricultural Sciences, Uppsala.
- Lawrence CJ and Krzanowski WJ (1996). Mixture separation for mixed-mode data. *Stat. Comput.* 6: 85-92.
- Lefèvre F and Charrier A (1993). Isozyme diversity within African *Manihot* germplasm. *Euphytica* 66: 73-80.
- Mezette TF, Blumer CG and Veasey EA (2013). Morphological and molecular diversity among cassava genotypes. *Pesqui. Agropec. Bras.* 48: 510-518.
- Mingoti AS (2007). Análise de dados através de métodos de estatística multivariada: uma abordagem aplicada. Editora UFMG, Belo Horizonte.
- Mulalem T, WeldeMichael G, Benti T and Walle T (2013). Genetic diversity of cassava (*Manihot esculenta* Crantz) genotypes in Ethiopia. *Greener J. Agric. Sci.* 3: 636-642.
- Ortiz R, Crossa J, Franco J and Sevilla R (2008). Classification of Peruvian highland maize races using plant traits. *Genet. Resour. Crop Evol.* 55: 151-162.
- Pariyo A, Tukamuhabwa P, Baguma Y and Kawuki RS (2013). Simple sequence repeats (SSR) diversity of cassava in South, East and Central Africa in relation to resistance to cassava brown streak disease. *Afr. J. Biotechnol.* 12: 4453-4464.
- Rival L and McKey D (2008). Domestication and diversity in Manioc (*Manihot esculenta* Crantz ssp. *esculenta*, Euphorbiaceae). *Curr. Anthropol.* 49: 1119-1128.
- SAS Institute Inc (2011). SAS® 9.3 System Options: Reference. 2nd edn. SAS Institute Inc., Cary.

- Sudré CP, Goncalves LS, Rodrigues R, do Amaral Junior AT, et al. (2010). Genetic variability in domesticated *Capsicum* spp as assessed by morphological and agronomic data in mixed statistical analysis. *Genet. Mol. Res.* 9: 283-294.
- Ward JH (1963). Hierarchical grouping to optimize an objective function. *J. Am. Stat. Assoc.* 58: 236-244.
- Willemen L, Scheldeman X, Soto Cabellos V and Salazar SR (2007). Spatial patterns of diversity and genetic erosion of traditional cassava (*Manihot esculenta* Crantz) in the Peruvian Amazon: An evaluation of socio-economic and environmental indicators. *Genet. Resour. Crop Evol.* 54: 1599-1612.