Genetic diversity in sugarcane varieties in Brazil based on the Ward-Modified Location Model clustering strategy

B.P. Brasileiro[1], C.D. Marinho[1], P.M.A. Costa[1], E.F.A. Moreira[1], L.A. Peternelli[1] and M.H.P. Barbosa[2]

[1]Departamento de Estatística, Universidade Federal de Viçosa, Viçosa, MG, Brazil
[2]Departamento de Fitotecnia, Universidade Federal de Viçosa,Viçosa, MG, Brazil

Corresponding author: C.D. Marinho
E-mail: caillet.marinho@yahoo.com.br

ABSTRACT. We evaluated the genetic diversity of 77 clones of sugarcane used in crosses made by the Brazilian interuniversity network for the development of the sugar/energy sector (RIDESA) breeding program. Characterization of the genotypes was carried out at the ratoon stage, based on eight morphological traits and seven agronomic traits. Diversity analysis was carried out beginning with the Ward-Modified Location Model. The ideal number of groups was three. Groups 1, 2, and 3 were composed of 37, 21, and 19 accessions, respectively. Group 1 was formed entirely of commercial varieties (hybrids of advanced generations) and elite clones, with the exception of the old varieties 'Caiana Fita' and 'Cana Blanca' (hybrids of *Saccharum officinarum*). In general, group 2 had more divergent accessions regarding origin, including L60-14, NG57-6, TUC77-42, IN84-105 (hybrid of *S. officinarum*), and 28NG289 (species of *S. robustum*). Group 3 was formed entirely of commercial varieties and elite clones from the RIDESA program, with the exception of genotypes Co285 (India), Q124 (Australia) and VAT90-212 (unknown origin). The analysis based on the Ward-Modified Location Model

procedure resulted in an adequate and clearly discriminating grouping of sugarcane accessions, allowing the use of all the available information about the genotypes, in a mix of continuous and categorical variables.

Key words: *Saccharum* spp; Germplasm characterization; Genetic diversity; Multivariate analysis

INTRODUCTION

Sugarcane is cultivated in more than 100 tropical and subtropical nations and is of great economic importance due to the commercial value of its products, mainly sugar and ethanol. Furthermore, sugarcane has recently highlighted as a source of sustainable energy for the cogeneration of electricity and cellulosic ethanol from bagasse (Hofsetz and Silva, 2012).

Due to the increasing world demand for the production of biofuels, sugarcane production in Brazil has grown considerably in recent years, with the use of new agricultural areas, including regions with adverse edaphoclimatic conditions (Endres et al., 2010; Silva et al., 2012).

According to Waclawovsky et al. (2010), the average world yield is 80 t/ha and the theoretical maximum sugarcane yield is 380 t/ha. The average yield of Brazilian varieties (68 t/ha in 2011) falls quite short of what can actually be achieved with the genetic potential of the crop, which emphasises the necessity of breeding programs in achieving this objective. To attain the desired level of production without incorporating large new areas, breeding programs should indicate increasingly productive clones that are better adapted to these new regions in the shortest time.

Since the success of the development of new varieties of sugarcane is associated with the ability to correctly choose the best individuals to be the progenitors of the new varieties, a better understanding of the genetic diversity between the clones used as progenitors is essential in defining the new crosses (Alwala et al., 2006; Santos et al., 2012). To achieve this objective, it is necessary to characterize the germplasm collection of the breeding program.

The process of germplasm collection characterisation, in addition to providing the basic data necessary for the breeding programs, has a strategic application in the evaluation of genetic resources (Mohammadi and Prasanna, 2003). Also, advances in the areas of computational technology, numerical taxonomy, and methods of multivariate statistics have contributed to the conservation of genetic resources, the formation of nuclear collections and the identification of possible duplicates (Ortiz et al., 2008), furthermore providing parameters for progenitor choice that enable a greater heterotic effect on the progeny and increasing the chance of obtaining transgressive genotypes in segregant generations.

The best method of numerical classification is the one that defines the most compact, well-separated groups (Crossa and Franco, 2004). Franco et al. (1998) proposed the Ward Modified Location Model (Ward-MLM) as a substitute for the Location Model procedure proposed by Lawrence and Krzanowski (1996). This method allows the definition of the optimal number of groups and the calculation of an average of the groups with a high level of precision, making use of all of the available information about the genotypes, whether they are continuous, binary, or multi-categorical variables (Crossa and Franco, 2004).

The objective of this work was to evaluate the genetic diversity between 77 clones of sugarcane used in crosses by the RIDESA breeding program, using the Ward-MLM method, to obtain new varieties of sugarcane.

MATERIAL AND METHODS

Plant material

Seventy-seven clones of sugarcane belonging to the Active Germplasm Bank (AGB) of the Genetic Improvement Program of Sugarcane at Universidade Federal de Viçosa were used in this study, which integrates RIDESA (Table 1). AGB is composed of cultivated varieties, progenitors of commercial crosses, and old varieties. It is located in the municipality of Viçosa (20°45'S, 42°52'W; altitude, 648.74 m), in the state of Minas Gerais, Brazil.

AGB was implemented in 2009 in augmented blocks with two progenitors as controls (RB867515 and SP80-1816), with nine blocks of 15 plots each. The plots were constructed in furrows 3 m in length, which were spaced 1.4 m apart.

Phenotypic characterization

The characterization of the genotypes was carried out in August 2012, in the ratoon stage, 12 months after the last harvest, and was based on eight morphological traits and seven agronomic traits (Table 2).

Statistical analyses

The quantitative data was analyzed by restricted maximum likelihood (REML) and best linear unbiased prediction (BLUP), using the following statistical model: $y = Xf + Zg + Wb + e$, in which y is the vector of observations, where $y \sim N(XF, V)$, f is the vector of the effects assumed as fixed (averages of controls and average of the principal treatment population), is the vector of the genotypic effects (assumed as random), where $g \sim N(0, G)$, G is the matrix of genetic covariance of the genotypes $G = A\hat{\sigma}_g^2$, is the vector of the environmental effects of the blocks (assumed as random), where $b \sim N(0, I\hat{\sigma}_b^2)$, is the vector of errors or residuals (random), where $e \sim N(0, R)$, R is the matrix of residual covariance $(R = I\hat{\sigma}_e^2)$, X, Z and W are the incidence matrices for said effects, and $V = ZA\hat{\sigma}_g^2 Z' + WI\hat{\sigma}_b^2 W' + I\hat{\sigma}_e^2$. In this study, the matrix G was assumed to be equal to $I\hat{\sigma}_g^2$, that is $G = I$, where I = identity matrix. The variance components $\hat{\sigma}_g^2, \hat{\sigma}_b^2$, and $\hat{\sigma}_e^2$ correspond to the genotypic variance, block variance, and residual variance, respectively.

The diversity analysis was carried out beginning with the Ward-Modified Location Model with the use of the SAS software version 9.1.3 (SAS, 2002), starting with the qualitative variable mode and the average genotypic average of the quantitative variables estimated with the use of the SELEGEN-REML/BLUP software (Resende, 2007).

Relationship coefficient

After grouping, the coefficient of relationship (*f*) was obtained to investigate the relationship of the individuals within each group. Accordingly, the SAS software version 9.1.3 (SAS, 2002) was employed using the "proc inbreed" procedure. For the remote ancestors, *f* was considered to be equal to zero, as suggested by Cox et al. (1985), and the coefficient of endogamy equal to zero was assigned to each genotype, considering the heterozygous trait of this crop (Deren, 1995).

Table 1. Identification of the accessions, pedigree, and group into which each of the 77 sugarcane genotypes of the RIDESA/UFV germplasm bank were clustered.

Group	Accession	Female parent	Male parent	Origin[a]
1	Caiana Fita	*S. officinarum*	*S. officinarum*	Salinas, Brasil
1	Cana Blanca	*S. officinarum*	*S. officinarum*	Cuba
1	CB 47-355	POJ2878	Co413	C. Goytacazes, Brasil
1	IAC52-326	Co419	?	IAC, Brasil
1	RB001905	?	?	UFAL, Brasil
1	RB001913	RB931602	?	UFAL, Brasil
1	RB001914	RB72454	?	UFAL, Brasil
1	RB001922	?	?	UFAL, Brasil
1	RB00412	?	?	UFAL, Brasil
1	RB008004	SP80-3480	SP80-1836	UFV, Brasil
1	RB01406	?	?	UFAL, Brasil
1	RB72454	CP53-76	?	UFAL, Brasil
1	RB806043	NA56-79	?	UFPR, Brasil
1	RB825336	H53-3989	?	UFSCar, Brasil
1	RB825548	F150	?	UFSCar, Brasil
1	RB83102	NA56-79	SP70-1143	UFAL, Brasil
1	RB8317	POJ2878	H56-6724	UFAL, Brasil
1	RB835089	RB72454	NA56-79	UFSCar, Brasil
1	RB835486	L60-14	?	UFSCar, Brasil
1	RB845197	RB72454	SP70-1143	UFSCar, Brasil
1	RB845257	RB72454	SP70-1143	UFSCar, Brasil
1	RB8495	SP70-1143	?	UFAL, Brasil
1	RB855035	L60-14	SP70-1284	UFSCar, Brasil
1	RB855036	RB72454	SP70-1143	UFSCar, Brasil
1	RB855113	SP70-1143	RB72454	UFSCar, Brasil
1	RB876030	RB72454	?	UFPR, Brasil
1	RB928064	SP70-1143	?	UFV, Brasil
1	RB93509	RB72454	?	UFAL, Brasil
1	RB941537	RB83160	RB855035	UFSCar, Brasil
1	RB947501	SP71-1406	RB72454	UFV, Brasil
1	RB975950	RB825548	RB835486	UFSCar, Brasil
1	SP70-1143	IAC48/65	?	Copersucar, Brasil
1	SP71-3530	?	?	Copersucar, Brasil
1	SP77-5181	HJ57-41	?	Copersucar, Brasil
1	SP79-1011	NA56-79	Co775	Copersucar, Brasil
1	SP80-1816	SP71-1088	H57-5028	Copersucar, Brasil
1	SP83-2847	HJ5741	SP70-1143	Copersucar, Brasil
2	28NG289	*S. robustum*	?	Coimbatore, Índia
2	IAC87-3396	Co740	SP70-1143	IAC, Brasil
2	IN84-105	*S. officinarum*	?	Indonésia
2	L60-14	CP52-1	CP48-103	Louisiana, USA
2	NG57-6	*S. officinarum*	*S. officinarum*	Coimbatore, India
2	RB00507	RB931530	RB83594	UFAL, Brasil
2	RB00404	SP79-2233	?	UFAL, Brasil
2	RB721012	Co331	?	UFAL, Brasil
2	RB75126	C278	?	UFAL, Brasil
2	RB813804	CP48-124	?	UFPB, Brasil
2	RB835054	RB72454	NA56-79	UFSCar, Brasil
2	RB845210	RB72454	SP70-1143	UFSCar, Brasil
2	RB855156	RB72454	TUC71-7	UFSCar, Brasil
2	RB855206	RB72454	TUC71-7	UFSCar, Brasil
2	RB867515	RB72454	?	UFV, Brasil
2	RB925211	RB855206	?	UFSCar, Brasil
2	RB946903	RB765418	RB72454	UFPR, Brasil
2	SP80-1842	SP71-1088	H57-5028	Copersucar, Brasil
2	SP80-3280	SP71-1088	H57-5028	Copersucar, Brasil
2	TUC77-42	CP71-321	US72-19	Tucuman, Argentina
2	UVA	?	?	Salinas, Brasil
3	Co285	*S. spontaneum*	?	Coimbatore, India
3	Q124	Co310	54N7096	BSES, Austrália
3	RB001906	?	?	UFAL, Brasil

Table 1. Continued.

Group	Accession	Female parent	Male parent	Origin[a]
3	RB001916	?	?	UFAL, Brasil
3	RB001921	?	?	UFAL, Brasil
3	RB00410	SP80-1770	RB75126	UFAL, Brasil
3	RB00416	CB47-355	?	UFAL, Brasil
3	RB00504	?	?	UFAL, Brasil
3	RB00509	RB931530	RB83594	UFAL, Brasil
3	RB01413	?	?	UFAL, Brasil
3	RB855046	SP70-1143	TUC71-7	UFSCar, Brasil
3	RB855536	SP70-1143	RB72454	UFSCar, Brasil
3	RB925345	H59-1966	?	UFSCar, Brasil
3	RB92606	Q107	RB72454	UFAL, Brasil
3	RB931555	SP71-6113	?	UFAL, Brasil
3	RB945961	RB855206	?	UFSCar, Brasil
3	RB956911	RB855206	RB855035	UFPR, Brasil
3	RB977619	RB72454	RB806043	UFV, Brasil
3	VAT90-212	?	?	?

[a]IAC = Instituto Agronômico de Campinas; UFAL = Universidade Federal de Alagoas; UFSCar = Universidade Federal de São Carlos; UFV = Universidade Federal de Viçosa; UFPR = Universidade Federal do Paraná; UFPB = Universidade Federal de Pernambuco; BSES = Bureau of Sugar Experiment Stations. ? = parent/origin unidentified.

Table 2. Descriptors used for the characterisation and evaluation of 77 sugarcane accessions from the gene bank of RIDESA, Brazil.

Variables	Scale	Description
Morphological[a]		
SR	1-4	(1 = natural, 2 = easy, 3 = medium, 4 = difficult)
LH	1-4	(1 = absent, 2 = few, 3 = medium, 4 = very many)
BP	1-3	(1 = bud prominence < 1mm, 2 = 2mm, 3 = 3mm)
BG	1-3	(1 = absent, 2 = inclination < 30°, 3 = among 30° and 60°)
PS	1-3	(1 = absent, 2 = diameter <1mm , 3 = among 1 and 2mm)
EP	1-4	(1 = absent, 2 = 1/4 of stalk, 3 = 1/2, 4 ≥ 1/2)
LA	1-4	(1 = erect, 2 = tip curve, 3 = arched, 4 = base curve)
DC	1-3	(1 = absent, 2 = shallower, 3 = profound)
Agronomic[b]		
SH		Stalk height (cm)
SD		Stalk diameter
SN		Stalk number[c]
BRIX		Brix %
IL		Internode length (cm)
L+3		Leaf + 3 length (cm)
SW		Stalk weight (g)

[a]SR = Straw removal; LH = Leaf sheath hair; BP = Bud prominence; BG = Bending by excessive growth; PS = Piping of stalk; EP = Extension pith; LA = Leaf architecture and DC = depth of channel. [b]All of the agronomic variables were measured in three stalks in each plot. [c]Total Number of stalks in a three-meter plot.

RESULTS AND DISCUSSION

In defining the number of groups, the ideal quantity defined by the Ward-MLM method was equal to three, since the risk profile associated with the likelihood ratio test showed the greatest increase in the logarithmic function of probability in group 3 (71.77) (Figure 1).

According to Gonçalves et al. (2009), the logarithmic function of probability precisely defines the number of groups, resulting in less subjective groupings. Different studies have demonstrated that the number of groups can vary greatly depending on the species, the number of accessions, and the type/quantity of variables evaluated (Ortiz et al., 2008; Gonçalves et al., 2009; Barbé et al., 2010; Cabral et al., 2010).

In sugarcane, studies evaluating the genetic diversity have been conducted with the use of quantitative and qualitative traits, and/or with molecular markers (Da Silva et al., 2008; Oliveira et al., 2009; Duarte Filho et al., 2010; Dutra Filho et al., 2011; Sindhu et al., 2011; Santos et al., 2012; Santchurn et al., 2012; Perera et al., 2012). In these works, a high degree of genetic similarity between the genotypes has been found, resulting in the formation of few groups.
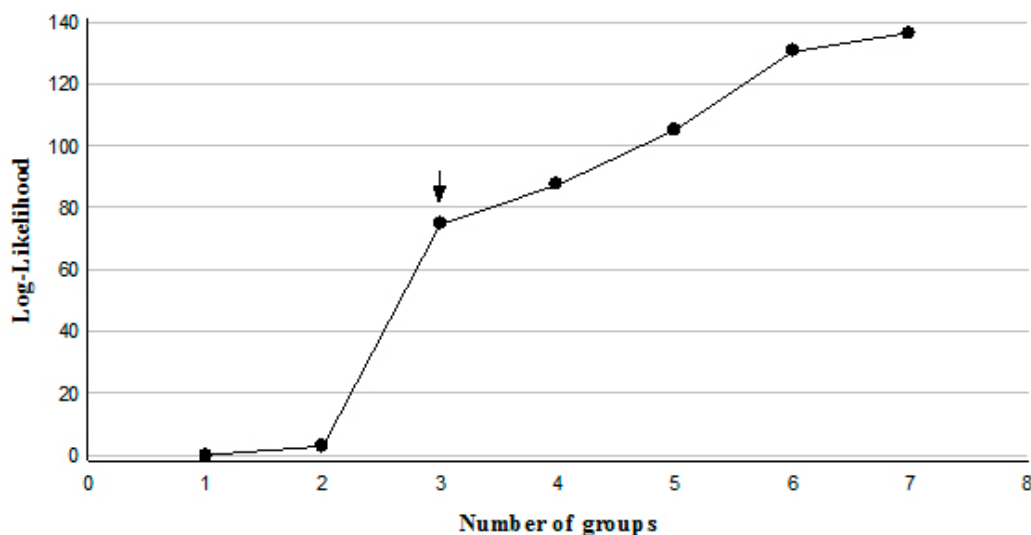

Figure 1. Graph expressing the logarithmic probability function (log-likelihood) according to the number of groups.

In the present study, the first two canonical variables obtained with the Ward-MLM explained 100% of the variation observed, permitting a satisfactory understanding of the genetic variability between the groups and between the clones of the same group. The formation of the three groups can be observed in the graphic representation of the first two canonical variables (Figure 2).

On the basis of the Mahalanobis distance matrix, groups 1 and 2 were the closest, with a distance of 14.75, while the greatest distance, 19.37, was found between groups 1 and 3 (Table 3).

The greatest dispersion of accessions in group 3 correlates with the genetic distance values between the groups. Greater variability within group 3 was noted via the graphic dispersion. Indeed, this was calculated by the averages of the coefficients of relationship between the individuals of the same group, in which, for groups 1, 2, and 3, the average values of the coefficients of relationship (*f*) were 0.0137, 0.0094, and 0.0037, respectively.

Groups 1, 2 and 3 were composed of 37, 21, and 19 accessions, respectively. Group 1 was formed entirely of commercial varieties (hybrids of advanced generations) and elite clones, with the exception of the old varieties Caiana Fita and Cana Blanca (hybrids of *S. officinarum*). In general, group 2 showed more divergent accessions in regard to the origin, such as L60-14, NG57-6, TUC77-42, IN84-105 (hybrid of *S. officinarum*) and 28NG289 (species *S. robustum*). Group 3 was formed totally of commercial varieties and elite clones from the RIDESA program, with the exception of the genotypes Co285 (India), Q124 (Australia), and VAT90-212 (unknown origin) (Table 1).
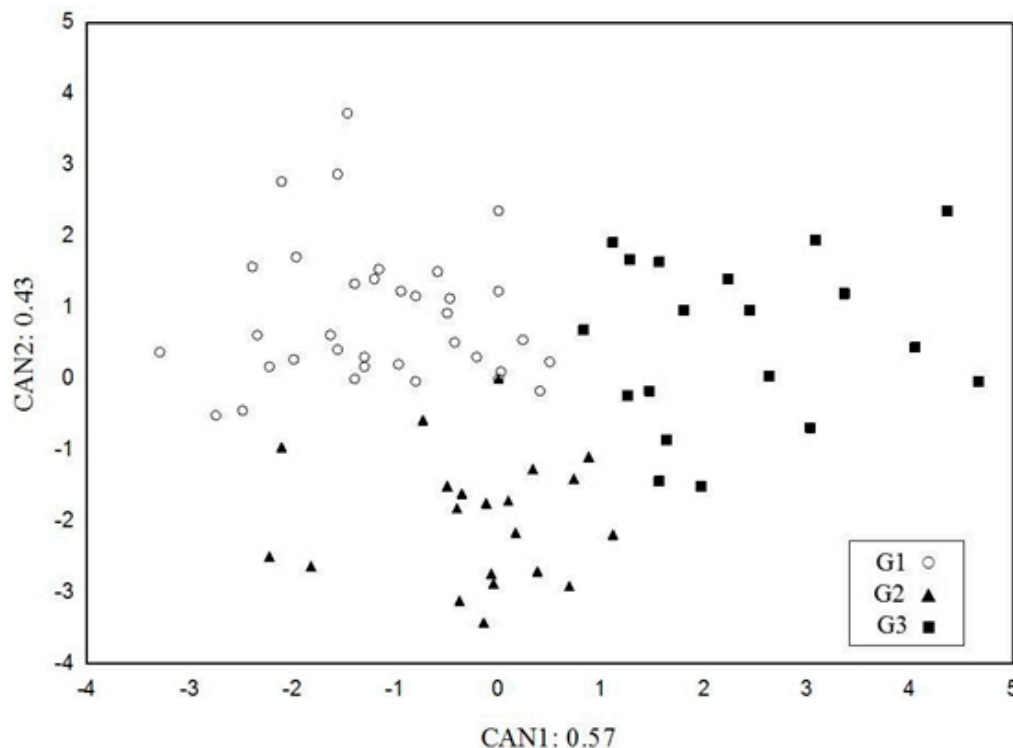
Figure 2. Representative graph of the first two canonical variables for the three groups formed by the Ward-MLM method.

Table 3. Mahalanobis distance between the groups formed by the Ward-MLM method.

Groups	G1	G2	G3
G1	-	14.75	19.37
G2	14.75	-	16.6
G3	19.37	16.6	-

   With regard to the morphological variables, these displayed broad dispersion in the different classification categories, demonstrating great variability from the genotypes in relation to such traits (Table 4). It is important to highlight that the evaluation of morphological traits is performed in a subjective manner through the assignment of classes by the evaluator, which explains the greater variability encountered. However, due to the ease of measurement and agronomic importance, one can conclude that they are useful in the study of genetic diversity.

   On the basis of the correlations of the agronomic traits with the two canonical variables, all of the traits evaluated showed considerable contributions to the genetic divergence, with the exception of SN the trait. The greatest magnitudes of correlation were observed between IL (CAN1), L+3 (CAN2), SW (CAN1), and SH (CAN1), with values of 0.8196, 0.7026, 0.6348, and 0.6105, respectively, demonstrating the importance of these variables for the study of genetic diversity in sugarcane (Table 5).

Table 4. Number of accessions according to some morphological characteristics in each of the three groups (G1, G2, G3) formed by the WARD-MLM strategy from 77 sugarcane accessions of the RIDESA gene bank.

Variables	Groups		
	G1	G2	G3
Straw removal			
Natural	1	1	1
Easy	9	7	4
Medium	21	7	9
Difficult	6	6	5
Hair number			
Absent	10	16	10
Few	16	3	8
Medium	8	2	-
Very many	3	-	1
Bud prominence			
Prominence of 1 mm	11	4	5
Prominence of 2 mm	23	11	7
Prominence of 3 mm	3	6	7
Bending by excessive growth			
Absent	31	14	14
Inclination <30°	4	2	5
Between 30° and 60°	2	5	-
Hollowing			
Absent	27	12	11
Diameter < 1mm	8	8	6
Between 1 and 2 mm	2	1	2
Extension pith			
Absent	35	15	10
1/4 of stalk	2	5	9
Between 1/4 and ½	-	-	-
≥ ½	-	1	-
Leaf architecture			
Erect	26	9	8
Tip curve	9	11	9
Arched	2	1	2
Base curve	-	-	-
Depth of channel			
Absent	21	14	9
Shallower	14	6	9
Profound	2	1	1

Table 5. Means of agronomic traits (variables) for each of the three groups formed by the Ward-MLM method and correlation coefficient of the quantitative with the first two canonical variables (CAN).

Variables[a]	Groups			CAN[b]	
	G1	G2	G3	CAN1	CAN2
IL	6.37	6.43	6.68	0.8196	0.0632
L+3	6.26	6.13	6.29	0.2128	0.7026
SN	10.64	10.59	9.48	-0.0685	-0.0162
BRIX	22.84	22.00	22.28	-0.2014	0.3326
SH	120.09	135.77	142.19	0.6105	-0.3358
SW	934.92	984.88	1094.1	0.6348	-0.0434
SD	37.89	37.81	38.02	0.3067	0.2793

[a]IL = internode length, cm; L+3 = leaf + 3 length, cm; SN = stalk number; BRIX = percentage of soluble solids in juice; SH = stalk height, cm; SW = stalk weight, g; SD = stalk diameter, mm. [b]Correlation coefficients of canonical variables 1 (CAN1) and 2 (CAN2).

Considering the general genotypic average of each trait evaluated for the three groups formed by the Ward-MLM method (Table 5), it can be seen that the greatest averages for the traits

SN and BRIX occurred in group 1. In contrast, the smallest averages for IL, SH, and SW were shown by the same group. For L+3, BRIX, and SD, the smallest averages were from group 2, and the smallest averages for SN were from group 3. Thus, for the quantitative traits evaluated, it can be seen that group 3 had the largest number of genotypes with superior averages in relation to the other groups, with the exception of SN and BRIX, whose averages from group 1 were superior.

The description of the data allocated within each group, for all of the quantitative variables, is presented in Figure 3. For IL, the amplitudes of the data were very similar between the three groups. However, the data from group 3 exhibited greater variation and more elevated averages. For L+3, the amplitudes remained similar, and again, group 3 presented greater values with greater variation of the data. However, in this case, it was group 2 that had the smallest values (Figure 3).
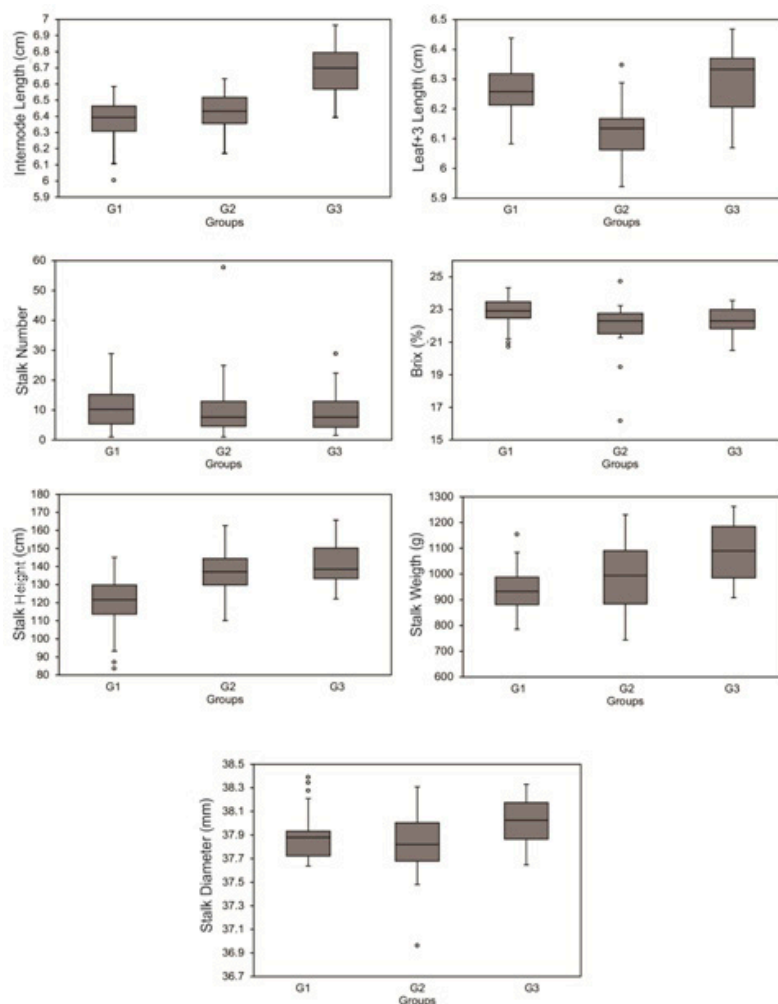

Figure 3. Box-plots showing variation of the data from the seven quantitative variables evaluated in the 77 accessions of sugarcane grouped into three groups by the Ward-MLM method.

In relation to SN and BRIX, the genotypic averages were very similar between and within the groups, with a slight superiority for group 1 (Figure 3). It is important to highlight that the outliers for SN (57.7) and BRIX (16.17) in group 2 were obtained from genotype 28NG289. This genotype in particular is the only representative of the species *S. robustum*, and is characterized by a high number of small diameter stalks and low sucrose content in the juice.

The genotypic averages for SH and SW increased from group 1 to group 3. Group 2 had intermediate averages. The trend of similarity between the groups for these two traits was expected, since the Pearson correlation between these variables was 0.70 (P < 0.01). Therefore, on average, it could be expected that a greater height would give rise to greater stalk weight. Ultimately, the trait SD showed similar averages between the groups, with a pronounced outlier (36.96) for group 2, once more arising from genotype 28NG289 (Figure 3).

Considering that the traits of greater agronomic interest in the current study were SN and BRIX, the genotypes that stood out in group 1 for SN were RB8495, RB001914, and RB845257, with genotypic averages of 28.88, 25.83, and 19.78 for stalks per plot, respectively. For BRIX, the prominent genotypes were RB001922 (24.34%), RB825336 (24.26%), and RB876030 (24.02%). In group 2, as previously stated, the genotype with the greatest genotypic average for SN was clone 28NG289 (57.71), followed by clones RB00507 (24.97) and UVA (18.50). With regard to BRIX, the clones that stood apart were SP80-1842 (24.73%), RB813804 (23.24%), and RB00404 (23%). With regard to SN in group 3, the genotypes with the greatest genotypic averages were the clones Co285 (28.85), RB00509 (22.38), and RB855536 (13.71). For BRIX, the clones RB925345 (23.55%), RB945961 (23.55%), RB855046 (23.40%), and Q124 (23%) had the highest genotypic averages.

The formation of groups is important in progenitor choice in breeding programs, since the new hybrid populations should be established on the basis of the magnitude of their dissimilarities and on the "*per se*" potential of the progenitors. The analysis strategy based on the Ward-MLM procedure resulted in an adequate and clearly discriminating grouping of accessions of sugarcane, allowing the use of all of the available information about the genotypes, in a mix of continuous and categorical variables. The Ward-MLM method has permitted the formation of compact, heterogeneous groupings in studies of snap beans, common bean, corn, *Brassica rapa* subsp. *rapa* L. and *Capsicum* spp. (Padilla et al., 2005; Ortiz et al., 2008; Barbé et al., 2010; Sudré et al., 2010; Cabral et al., 2010;). This technique is therefore an important statistical tool to be incorporated into sugarcane breeding programs to evaluate genetic diversity in germplasm banks, assisting in the definition of crossing strategies to increase efficiency in the generation of transgressive genotypes.

ACKNOWLEDGMENTS

REFERENCES

Alwala S, Suman A, Arro JA and Veremis JC (2006). Target region amplification polymorphism (TRAP) for assessing genetic diversity in sugarcane germplasm collections. *Crop Sci.* 46: 448-455.

Barbé TD, Amaral 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.

Cabral PDS, Soares TCB, Goncalves LSA and Amaral AT (2010). Quantification of the diversity among common bean accessions using Ward-MLM strategy. *Pesq. Agropec. Bras.* 45: 1124-1132.

Cox TS, Kiang YT, Gorman MB and Rodgers DM (1985). Relationship between Coefficient of Parentage and Genetic Similarity Indexes in the soybean. *Crop Sci.* 25: 529-532.

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

Da Silva CM, Mangolin CA, Mott AS and Machado MFPS (2008). Genetic diversity associated with *in vitro* and conventional bud propagation of *Saccharum* varieties using RAPD analysis. *Plant Breed.* 127: 160-165.

Deren CW (1995). Genetic base of US mainland sugarcane. *Crop Sci.* 35: 1195-1199.

Duarte Filho L, Silva P, Santos J and Barbosa G (2010). Genetic similarity among genotypes of sugarcane estimated by SSR and coefficient of parentage. *Sugar Tech.* 12: 145-149.

Dutra Filho JA, Melo LJOT, Resende LV and Anunciação Filho CJ (2011). Aplicação de técnicas multivariadas no estudo da divergência genética em cana-de-açúcar. *Rev. Ciênc. Agron.* 42: 185-192.

Endres L, Silva JV, Ferreira VM and Barbosa GVS (2010). Photosynthesis and water relations in Brazilian sugarcane. *The Open Agric. J.* 4: 31-37.

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

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

Hofsetz K and Silva MA (2012). Brazilian sugarcane bagasse: Energy and non-energy consumption. *Biomass Bioenerg.* 46: 564-573.

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

Mohammadi SA and Prasanna BM (2003). Analysis of genetic diversity in crop plants - Salient statistical tools and considerations. *Crop Sci.* 43: 1235-1248.

Oliveira KM, Pinto LR, Marconi TG, Mollinari M, et al. (2009). Characterization of new polymorphic functional markers for sugarcane. *Genome* 52: 191-209.

Ortiz R, Sevilla R, Alvarado G and Crossa J (2008). Numerical classification of related Peruvian highland maize races using internal ear traits. *Genet. Resour. Crop. Ev.* 55: 1055-1064.

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

Perera MF, Arias ME, Costilla D and Luque AC (2012). Genetic diversity assessment and genotype identification in sugarcane based on DNA markers and morphological traits. *Euphytica* 185: 491-510.

Resende MDV (2007). Software SELEGEN-REML/BLUP: Sistema Estatístico e Seleção Genética Computadorizada via Modelos Lineares Mistos. Embrapa Florestas.

Santchurn D, Ramdoyal K, Badaloo MGH and Labuschagne M (2012). From sugar industry to cane industry: investigations on multivariate data analysis techniques in the identification of different high biomass sugarcane varieties. *Euphytica* 185: 543-558.

Santos JM, Duarte Filho LSC, Soriano ML and Silva PP (2012). Genetic diversity of the main progenitors of sugarcane from the RIDESA germplasm bank using SSR markers. *Ind. Crop. Prod.* 40: 145-150.

SAS Institute (2002). Software Version 9.1.3 of the SAS System for Windows. SAS Institute Inc, Cary, NC.

Silva PP, Soares L, Costa JG and Silva Viana L (2012). Path analysis for selection of drought tolerant sugarcane genotypes through physiological components. *Ind. Crop. Prod.* 37: 11-19.

Sindhu R, Govindaraj P, Balamurugan A and Appunu C (2011). Genetic diversity in sugarcane hybrids (*Saccharum* spp complex) grown in tropical India based on STMS markers. *J. Plant Biochem. Biot.* 20: 118-124.

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.

Waclawovsky AJ, Sato PM, Lembke CG, Moore PH, et al. (2010). Sugarcane for bioenergy production: an assessment of yield and regulation of sucrose content. *Plant Biotechnol. J.* 8: 263-276.