Genetic diversity and coefficient of parentage between clones and sugarcane varieties in Brazil

B.P. Brasileiro[1], C.D. Marinho[1], P.M.A. Costa[1], L.A. Peternelli[1], M.D.V. Resende[2], D.E. Cursi[3], H.P. Hoffmann[3] and M.H.P. Barbosa[4]

[1]Departamento de Estatística, Universidade Federal de Viçosa,
Viçosa, MG, Brasil
[2]Embrapa Floresta, Colombo, PR, Brasil
[3]Departamento de Biotecnologia e Produção Vegetal e Animal,
Universidade Federal de São Carlos, São Carlos, SP, Brasil
[4]Departamento de Fitotecnia, Universidade Federal de Viçosa,
Viçosa, MG, Brasil

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

ABSTRACT. The success of the development of new sugarcane varieties is associated with the ability to correctly select the genitor. The aim of this study was to evaluate the genetic diversity between 113 clones and sugarcane varieties using the Ward-modified location model procedure with added information about the coefficient of parentage and endogamy. In this study, data was used from 100 experiments that evaluated clones; the experimental phase was conducted in 70 places between the years 2002 and 2009 on the outlining in random blocks. According to the diversity analysis, 3 groups formed: G1, G2, and G3, which were composed of 58, 8, and 47 genotypes, respectively. The clones of groups G1 and G3 were the most outstanding. Thus, biparental crossbreeding involving clones and varieties of these 2

groups can efficiently obtain transgressive genotypes. Knowledge of the heterotypic groups indicated by the Ward-modified location model method, along with the parentage information, will make it a lot easier to define the desirable and undesirable crossbreeds for public and private breeding programs that develop sugarcane varieties.

Key words: Sugarcane varieties; Genetic diversity; Joint analysis; Coefficient of parentage; Germplasm characterization

INTRODUCTION

Because of the increasing worldwide demand for the production of biofuels, the production of sugarcane in Brazil has increased considerably in the past few years. New agricultural areas have been used, including regions with adverse edaphoclimatic conditions (Endres et al., 2010; Silva et al., 2012). In order to increase production without having to incorporate new areas, clones that are more productive and adapted to different productive regions should be indicated by the breeding programs and incorporated in the fieldwork in a short amount of time.

The success of the development of new sugarcane varieties is associated with the ability to correctly choose the genitors. Therefore, a better understanding of the genetic diversity between clones that are used as genitors becomes essential to define new crossbreeding strategies (Alwala et al., 2006; Santos et al., 2012).

There is no consensus on the best length measurement to be used in studies of diversity. The kind of length to be applied in the analysis is dependent on the nature of the variable evaluated (Crossa and Franco, 2004).

The Euclidean distance and the generalized distance by Mahalanobis are commonly used for quantitative variables (Gonçalves et al., 2008). For binary variables, several coefficients of association have been proposed. The Sorensen-Dice coefficient and the complement of the index of Jaccard are the favorite ones for the diversity analysis of molecular data (Mohammadi and Prasanna, 2003). For multicategory variables, the dissimilarity measurement proposed by Cole-Rodgers et al. (1997) is the most appropriate one that is used in studies of genetic diversity (Cruz et al., 2012).

Gower (1971) proposed an algorithm that is able to generate a dissimilarity measurement from the joint analysis of continuous, binary, and multicategory variables, which can combine different kinds of variables to obtain a single matrix of distance. Franco et al. (1998) proposed the Ward-modified location model (Ward-MLM) method. In this case, the grouping is based upon the matrix of distance by Gower and allows the definition of a great number of groups and the calculation of an average of groups with high precision by using all of the information that is available about the genotypes (Crossa and Franco, 2004).

Recently, the Ward-MLM strategy has been widely used to study diversity in cultures, like the physic nut (Brasileiro et al., 2013), castor-oil plant (Oliveira et al., 2013), banana (Pestanana et al., 2011; Pereira et al., 2012), bean (Barbé et al., 2010; Cabral et al., 2010), corn (Ortiz et al., 2008), and tomato (Gonçalves et al., 2008).

Additional important information that can help the management of germplasm associated with the breeding programs is the calculation of the coefficient of parentage and endogamy (Peternelli et al., 2009). This information allows efficient decisions to be made about

the choice of a genitor that enables the offspring to have a large heterotype.

The aim of this study was to evaluate the genetic diversity among 113 clones and sugarcane varieties selected by Rede Interuniversitária para o Desenvolvimento do Setor Sucroenergético (RIDESA) breeding program (Barbosa et al., 2012) using the Ward-MLM procedure (Franco et al., 1998). The results will add to the information about the coefficient of parentage and endogamy, which will assist the selection of genitors and the obtainment of new varieties.

MATERIAL AND METHODS

Vegetal material

This study used data from 100 experiments that evaluated clones to obtain the genotype average of 113 clones and varieties. The experiments were conducted between 2002 and 2009 on the outlining of random blocks. Each experiment included 20 or 22 clones, with 3 to 4 repetitions per experiment, and the number of harvests varied from 1 to 4. Each plot was composed of 4 or 5 rows that were 8 or 10 m in length and were spaced 1.4 m apart. Of the 70 sites of experimental procedures, 18 were in mills in the State of São Paulo, and the others in 52 places in the State of Minas Gerais, Brazil. A total of 769 sugarcane clones were evaluated in the 100 experiments.

From the 769 clones, the data of 113 were used. These 113 clones were selected because they were previously characterized and introduced in the Germplasm Active Bank (BAG) at Universidade Federal de Alagoas (UFAL), a place where crossbreeding is evaluated by the genetic improvement program of RIDESA (Table 1) (Barbosa et al., 2002).

Phenotypic characterization

The multicategory and binary characteristics were quantified from the information observed in the clones in several experiments, multiplication fields, and seedling production. With this information, the average concept or the most frequent one was attributed for the clone introduced to the BAG of UFAL. This way, the database has only 1 piece of information for each characteristic.

The multicategory characteristics evaluated were maturation, flowering, pith, adoption, development, growth habit, and sprouting. The binary characteristics related to the plague and disease resistance were resistant to the mosaic, smut disease, red rot, leaf rot, leaf scald, eye spot, brown spot, ratoon stunting disease, rust, yellow spot, and nematodes. The continuous characteristics evaluated were tons of stalks per hectare (TSH), fiber percent (Fiber), and sucrose content of the sugarcane (SSC) (Table 2).

Data analysis

The continuous characteristics (TSH, Fiber, and SSC) were obtained at the plot level in 100 experiments installed in random blocks. This database was analyzed via restricted maximum likelihood (REML) and best linear unbiased prediction (BLUP) through the following statistical model: $Y = Xf + Za_1 + Wa_2 + Ta_3 + Qa_4 + e$, in which y is the vector data, f is the vector of effects assumed as fixed (measurement) added to the general average, a_1 is the vector

Table 1. Identification of the approaches, genitors, and endogamy coefficient (F) from the 113 clones of sugarcane belonging to the Germplasm Active Bank at Rede Interuniversitária para o Desenvolvimento do Setor Sucroenergético (RIDESA) Brazil.

Clone	Female	Male	F	Clone	Female	Male	F
Group 1[+]				Group 2			
RB008041	SP84-2025	SP80-3280	0.000584	RB008098	RB845197	?	0.000032
RB008133	SP81-3250	SP80-1842	0.000007	RB765418	M253/48	?	0
RB008293	RB855113	?	0.000032	RB835486	L60-14	?	0.002187
RB008296	SP80-1816	RB855589	0.000177	RB855575	SP70-1143	RB72454	0.000147
RB008309	SP80-1842	?	0	RB955980	RB855206	RB855035	0.005588
RB835019	RB72454	NA56-79	0.003323	RB965906	RB835486	RB855536	0.000709
RB835054	RB72454	NA56-79	0.003323	RB987965	RB72454	RB739359	0.004075
RB835089	RB72454	NA56-79	0.003323	RB998369	SP82-6108	IAC86-2210	0.00452
RB845197	RB72454	SP70-1143	0.000147	Group 3			
RB845210	RB72454	SP70-1143	0.000147	RB008004	SP80-3480	SP80-1836	0.004502
RB845239	RB72454	SP70-1143	0.000147	RB008026	RB845197	?	0.000032
RB855002	SP70-1143	RB72454	0.000147	RB008304	SP80-1816	RB855589	0.000177
RB855035	L60-14	SP70-1284	0.002362	RB008310	SP80-1842	?	0
RB855046	SP70-1143	TUC71-7	0.000824	RB008342	SP80-3280	?	0
RB855156	RB72454	TUC71-7	0.004711	RB008348	SP80-3280	?	0
RB855357	RB72454	?	0.000147	RB845257	RB72454	SP70-1143	0.000147
RB855453	TUC71-7	?	0.000824	RB855036	RB72454	SP70-1143	0.000147
RB855563	TUC71-7	SP70-1143	0.000824	RB855113	SP70-1143	RB72454	0.000147
RB865230	SP70-1143	RB72454	0.000147	RB855536	SP70-1143	RB72454	0.000147
RB865513	RB72454	?	0.000147	RB975949	RB825548	RB855206	0.002409
RB865547	CP69-1062	H69-9018	0.001687	RB975950	RB825548	RB835486	0.000933
RB925211	RB855206	?	0.00101	RB975952	RB835486	RB825548	0.000933
RB925230	RB855511	?	0.000008	RB977508	SP80-1842	?	0
RB925268	RB855511	?	0.000008	RB977619	RB72454	RB806043	0.001685
RB925298	RB855589	?	0.000177	RB977625	RB83102	RB855002	0.018676
RB928064	SP70-1143	?	0	RB987649	RB72454	RB739359	0.004075
RB988113	RB72454	RB83102	0.001685	RB987905	RB72454	NA56-79	0.003323
RB988137	RB72454	RB83102	0.001685	RB987915	RB72454	CB45-3	0.001662
RB997627	SP80-180	SP84-7017	0.00328	RB987935	RB72454	RB83102	0.001685
RB997751	SP84-7017	SP80-185	0	RB987957	RB72454	RB739359	0.004075
RB937570	SP70-1143	RB72454	0.000147	RB988067	RB83102	RB72454	0.001685
RB945067	RB805004	?	0.000038	RB988082	RB83102	RB72454	0.001685
RB945099	RB835486	?	0.000469	RB988105	RB72454	RB83102	0.001685
RB945961	RB855206	?	0.00101	RB997671	SP80-185	SP80-3280	0
RB945962	RB855206	?	0.00101	RB997810	SP80-1816	SP88-721	0
RB945965	RB855589	?	0.000177	RB997984	RB86552	RB855584	0.000032
RB947520	RB72454	SP70-1143	0.000147	RB998025	SP80-1816	SP88-721	0
RB947625	RB765418	RB72454	0.000706	RB998118	RB835486	RB835205	0.001297
RB947663	RB72454	SP70-1143	0.000147	RB855546	SP70-1143	RB72454	0.000147
RB955430	RB72454	SP71-6949	0.001685	RB855595	SP70-1143	TUC71-7	0.000824
RB955469	RB855595	?	0.000177	RB867515	RB72454	?	0.000147
RB955970	RB845197	RB835486	0.000709	RB877603	F150	?	0.000213
RB957610	RB83102	?	0.000038	RB925345	H59-1966	?	0
RB965517	RB835089	?	0.000712	RB945040	RB855113	RB835632	0.000032
RB965518	RB855536	NA73-1454	0.005064	RB945063	RB855113	?	0.000032
RB965586	RB835486	RB855536	0.000709	RB945065	RB855113	?	0.000032
RB965743	RB855063	RB855536	0.019	RB945273	R855206	?	0
RB965902	RB855536	RB855453	0.001143	RB945275	RB845197	?	0.000032
RB965909	RB855511	RB855156	0.001842	RB945276	RB835486	RB845239	0.000709
RB965916	RB855589	?	0.000177	RB957689	RB751194	RB83102	0.000038
RB965918	RB855511	RB855002	0.000415	RB965505	SP79-1011	RB845210	0.004577
RB965920	RB855574	RB845210	0.019	RB965550	RB855589	?	0.000177
RB975939	RB845197	TUC71-7	0.002726	RB965560	RB845197	?	0.000032
RB975944	RB855563	RB735200	0.000665	RB965699	RB855156	RB855511	0.001842
RB977662	RB855453	?	0.000177	RB965911	RB855546	?	0.000032
RB987580	RB72454	RB83102	0.001685	RB965917	RB855453	RB855536	0.001143
RB987932	RB72454	RB83102	0.001685				
RB987933	RB72454	RB83102	0.001685				

[+]Groups were defined by the Ward-modified location model (results below).

Table 2. Categorical, binary, and continuous characteristics used to characterize 113 clones of sugarcane from the Germplasm Active Bank at Universidade Federal de Alagoas/RIDESA, Brazil.

Variable	Classes
Categorical[a]	
MAT	(1 = very premature, 2 = premature 3 = intermediate, 4 = late)
FLO	(1 = low, 2 = average, 3 = high, 4 = rare)
PITH	(1 = low, 2 = average, 3 = high, 4 = rare)
ADO	(1 = low, 2 = average, 3 = high)
DEV	(1 = regular, 2 = good, 3 = great)
GRH	(1 = erect, 2 = semi-decumbent , 3 = decumbent)
SPR	(1 = regular, 2 = good, 3 = great)
Binary[b]	
RMO	(0 = susceptible, 1 = resistant)
RSM	(0 = susceptible, 1 = resistant)
RRR	(0 = susceptible, 1 = resistant)
RLR	(0 = susceptible, 1 = resistant)
RSC	(0 = susceptible, 1 = resistant)
RES	(0 = susceptible, 1 = resistant)
RBS	(0 = susceptible, 1 = resistant)
RRS	(0 = susceptible, 1 = resistant)
RRU	(0 = susceptible, 1 = resistant)
RYS	(0 = susceptible, 1 = resistant)
RNE	(0 = susceptible, 1 = resistant)
Continuous[c]	
TSH	Tons of stalks per hectare
Fiber	Fiber percent
SSC	Sugar content (%)

[a]MAT = maturation; FLO = flowering; PITH = pith; ADO = adoption; DEV = development; GRH = growth habit; SPR = sprouting. [b]RMO = mosaic resistance; RSM = resistance to smut; RRR = resistance to red rot; RLR = resistance to leaf rot; RSC = resistance to leaf scald; RES = resistance to eye spot; RBS = resistance to brown spot; RRS = resistance to ratoon stunting disease; RRU = resistance to rust; RYS = resistance to yellow spot; RNE = resistance to nematodes. [c]TSH = tons of stalks per hectare; Fiber = fiber percent; SSC = sucrose content of the sugarcane.

of random effects of genotype, a_2 is the vector of random effects of the local-block interaction, a_3 is the vector of random effects of the local-block-genotype interaction, a_4 is the vector of random effects of genotype-measurement interactions, and e is the vector of random errors. The capital letters represent the matrices of incidence for the referred effects. The analysis was performed with the use of the SELEGEN-REML/BLUP software (Resende, 2007).

The diversity analysis was performed using the Ward-MLM procedure (Franco et al., 1998) and the SAS version 9.1.3 software (SAS Institute, 2002) using the genotypic averages in a joint way of the quantitative characteristics and the class of the qualitative characteristics of 113 clones and sugarcane varieties.

Parentage coefficient

The parentage analysis was done considering the sugarcane as octoploid and considering all of the parentage generations in the pedigree. In order to perform this analysis in the R software (R Development Core Team, 2013), functions developed by Peternelli et al. (2009) were used based on the generalized expressions developed by Kempthorne (1973).

The pedigree integrality (PI) was calculated by the method proposed by MacCluer et al. (1983), where the PI corresponds to the average proportion of the complete pedigree; for each generation, the values vary from 0 to 1. This way, if all of the ancestors of an organism

in a specific generation are known, PI = 1, otherwise, $0 \leq PI < 1$.

RESULTS

Based on the matrix of distance by Gower, we used the criteria pseudo-F to define the *a priori* ideal number of groups, which was 3 (Figure 1).
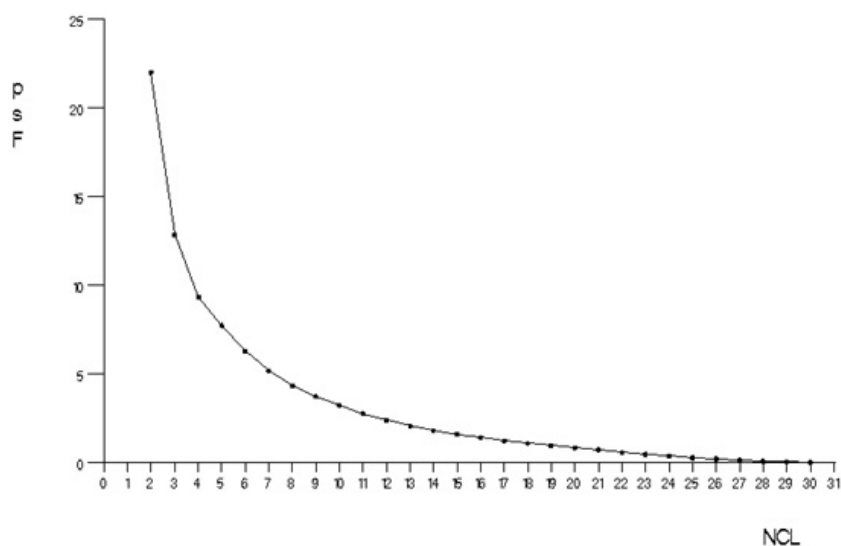

Figure 1. Pseudo-F criteria (psF) used to define the number of groups (NCL) *a priori* and based on the matrix of distance of Gower.

The *a posteriori* risk profile associated with the verosimilarity also showed the largest increment of the logarithm function of probability when considering 3 groups, with an increment of 45.1884 (Table 3). According to Crossa and Franco (2004), the point of the largest increment of the verosimilarity function must be used as a criterion to define the ideal number of groups.

Table 3. Number of *a posteriori* groups according to the increment of the logarithmic function of verosimilarity (Log-likelihood).

Number of groups	Log-likelihood	Increment
1	-1107.4161	0
2	-1063.1852	44.2309
3	-1017.9968	45.1884*
4	-994.5522	23.4446
5	-962.6031	31.9491
6	-937.929	24.6741
7	-923.1247	14.8043

*Largest increment.

In this study, the first 2 canonical variables obtained with the Ward-MLM method explained 100% of the variation observed, allowing a satisfactory comprehension of the genetic

variability between the groups and between the clones of the same group. Groups G1, G2, and G3 were composed of 58, 8, and 47 genotypes, respectively. The formation of the 3 groups can be observed in the graphic representation of the first 2 canonical variables (Figure 2).
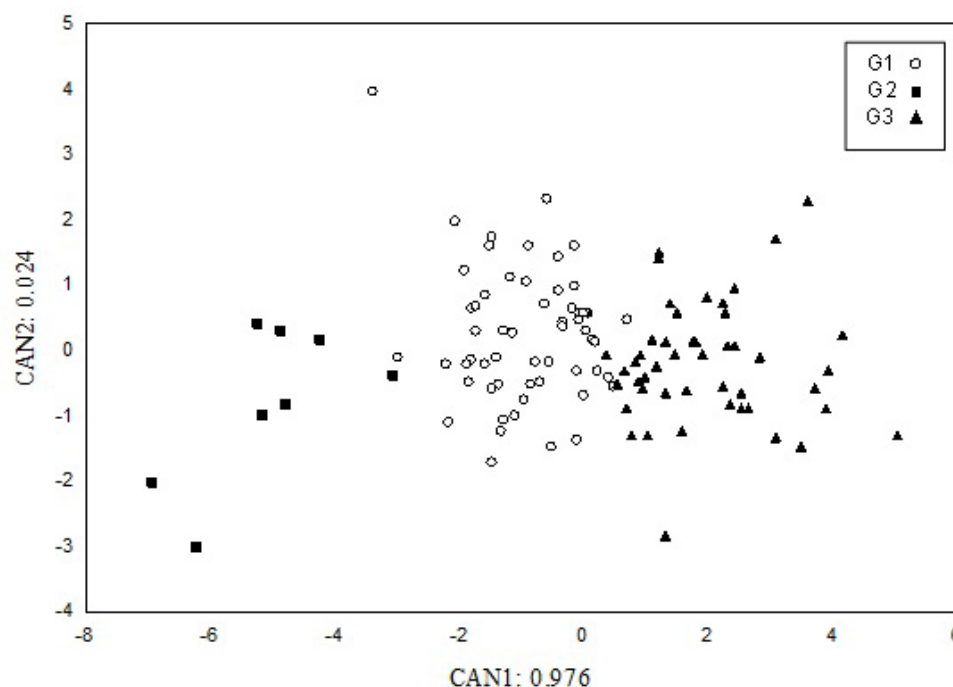

Figure 2. Representative graphic of the first 2 canonical variables for the 3 groups formed by the Ward-modified location model (MLM) method.

Group G1 was composed of 12 commercial varieties (RB835019, RB835054, RB835089, RB845197, RB845210, RB855035, RB855156, RB855563, RB925268, RB937570, RB965902, and RB965917) and 44 additional clones. Of the 8 clones belonging to the G2 group, 3 are commercial varieties: RB765418, RB835486, and RB855453. These varieties were launched in 1988, 1992, and 1995, respectively, and they correspond to the oldest varieties among those evaluated. In the G3 group, 10 varieties were allocated (RB845257, RB855036, RB855113, RB855546, RB855536, RB925345, RB865230, RB867515, RB925211, and RB928064) along with the other 39 clones (Table 1).

In all of the groups, the clones with upright growth predominated, followed by semi-decumbent material and a few decumbent genotypes. The clones with high adoption occurred only in the G1 and G3 groups, and most of the clones evaluated showed average adoption (Figure 3).

Only group G1 included clones with increased physiologic pith, which occurs in the absence of flowering. Most of the genotypes in this group present late maturation and low flowering. There are also many clones of average flowering, good sprouting, average adoption, and upright growth. Of the clones, 95% are resistant to eye spot, 92% to yellow spot, 91% to red rot, 90% to leaf rot, 88% to mosaic, 83% to rust, and 62% to nematodes and ratoon stunting disease. However, 50% of the clones are susceptible to brown spot (Figure 3).
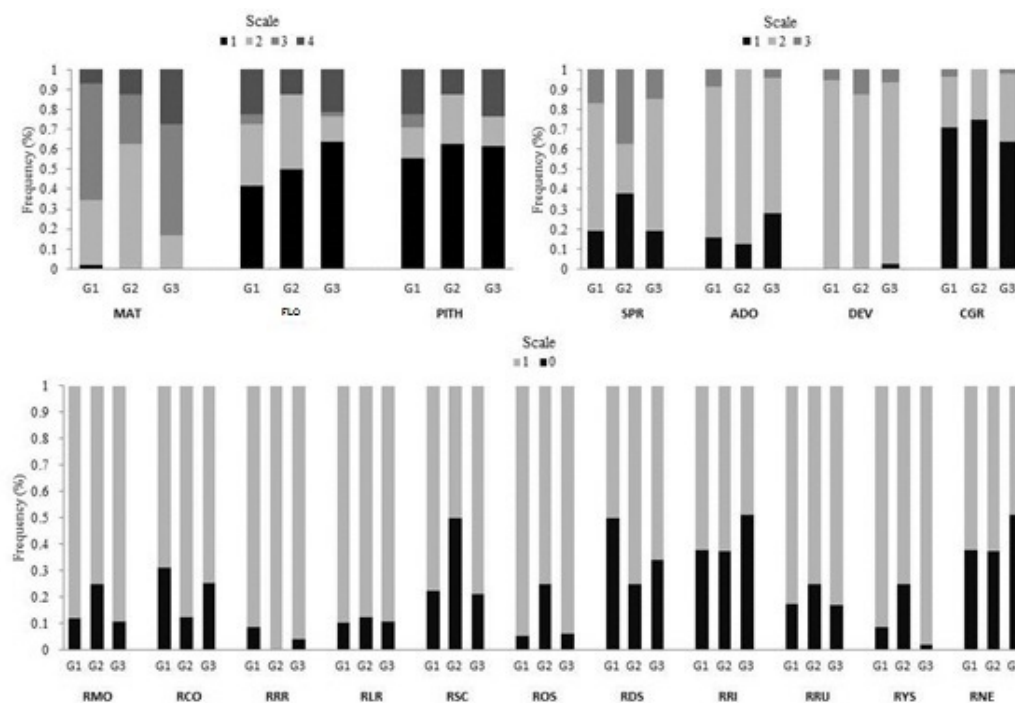
Figure 3. Frequency of the classes of each of the 7 multicategory variables and 11 binary variables evaluated in the 3 groups (G1, G2, and G3) formed by 113 clones of sugarcane. See classes (Scale) in Table 2. MAT = maturation; FLO = flowering; PITH = pith; ADO = adoption; DEV = development; GRH = growth habit; SPR = sprouting; bRMO = mosaic resistance; RSM = resistance to smut; RRR = resistance to red rot; RLR = resistance to leaf rot; RSC = resistance to leaf scald; RES = resistance to eye spot; RBS = resistance to brown spot; RRS = resistance to ratoon stunting disease; RRU = resistance to rust; RYS = resistance to yellow spot; RNE = resistance to nematodes.

In group G2, most clones had intermediate maturation, low flowering, regular sprouting, average adoption, and upright growth. All of the clones are resistant to red rot; 87% to smut and leaf rot; 75% to mosaic, brown spot, yellow spot, eye spot, and rust; and 62% are resistant to nematodes and ratoon stunting disease, whereas, 50% are susceptible to leaf scald (Figure 3).

In the G3 group, most of the clones were characterized by late maturation, low flowering, average adoption, and good sprouting. Of the clones, 98% are resistant to yellow spot, 95% to red rot, 93% to eye spot, 90% to leaf rot and mosaic, 83% to rust, 79% to leaf scald, 75% to smut, and 65% to brown spot (Figure 3).

Considering the average of the continuous variables in each group determined by the Ward-MLM strategy, it is possible to observe that the highest averages for TSH and Fiber occurred in group G3. This group has the highest number of clones with superior averages for the 2 characteristics. On the other hand, this was the group with the lowest average SSC. Group G2 presented the lowest averages for TSH and Fiber and the highest average for the SSC. The G1 group presented intermediate averages for the 3 measured quantitative characteristics (Figure 4).
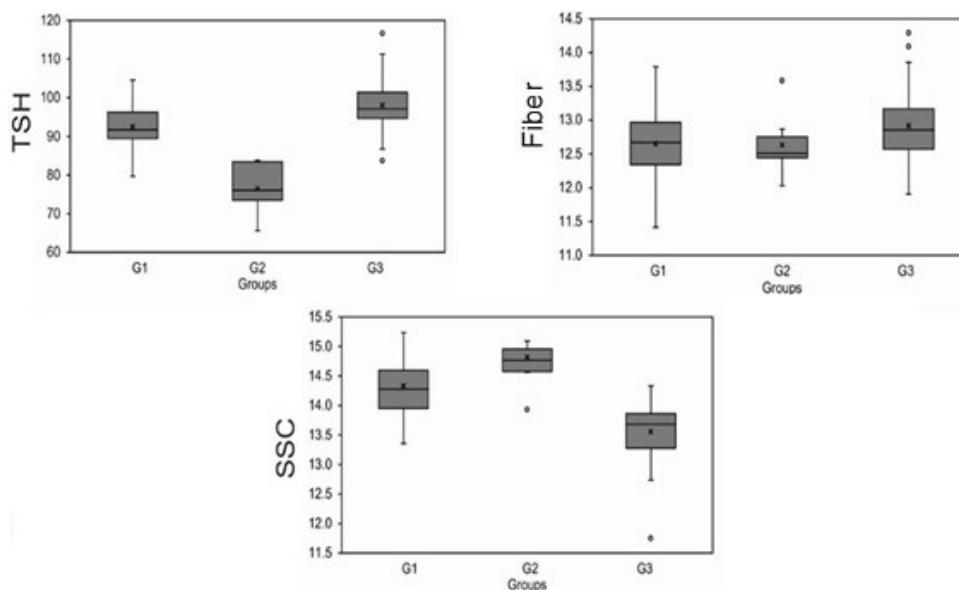
Figure 4. Box-plots of the tons of stalks per hectare (TSH), fiber (Fiber), and sucrose content of the sugarcane (SSC) variables evaluated in the 113 accessions of sugarcane comprising 3 groups by the Ward-MLM method. X = average; o = outliers.

The 20 genotypes with the highest averages for TSH, Fiber, and SSC are presented in Table 4. The accuracy of the estimation of the genotype values for the 20 best clones varied from 0.69 to 0.97 for TSH, from 0.88 to 0.98 for Fiber, and from 0.90 to 0.99 for SSC (Table 4).

Table 4. Genotype values (GV) and accuracy (Ac) of the 20 best clones among the 113 evaluated clones for the TSH, Fiber, and SSC characteristics in 100 experiments that evaluated clones in the states of Minas Gerais and São Paulo, Brazil.

Order	Clone	GV	Ac	Clone	GV	Ac	Clone	GV	Ac
		TSH			Fiber			SSC	
1	[3]RB988082	116.65	0.95	[3]RB997810	14.29	0.96	[1]**RB945965**	16.31	0.93
2	[3]RB987935	111.28	0.95	[3]RB997671	14.09	0.96	[2]**RB955980**	15.91	0.96
3	[3]RB997984	107.81	0.92	[3]RB977508	13.85	0.97	[1]RB835089*	15.23	0.90
4	[3]RB988105	105.65	0.94	[3]RB975950	13.80	0.96	[1]**RB845239**	15.21	0.91
5	[3]RB965911	105.29	0.96	[1]**RB945067**	13.79	0.95	[1]RB845197*	15.13	0.95
6	[3]RB945276	105.04	0.91	[3]RB945273	13.77	0.94	[2]RB855575	15.09	0.90
7	[1]RB987932	104.53	0.95	[3]RB998025	13.75	0.96	[1]RB855002	15.09	0.95
8	[3]RB965917*	104.05	0.96	[1]RB997751	13.72	0.94	[1]RB855046	15.07	0.99
9	[3]RB965550	103.34	0.86	[3]RB987957	13.65	0.92	[1]**RB965916**	15.04	0.96
10	[3]RB867515*	102.78	0.97	[3]RB008310	13.63	0.94	[1]RB965918	14.95	0.96
11	[3]RB945063	102.73	0.93	[2]**RB955980**	13.58	0.94	[2]RB008098	14.91	0.97
12	[1]RB925230	102.57	0.88	[1]RB835019*	13.55	0.95	[2]RB965906	14.88	0.98
13	[1]**RB945965**	102.41	0.89	[3]RB008304	13.49	0.96	[1]RB965920	14.86	0.98
14	[1]RB987580	102.29	0.69	[1]RB855563*	13.43	0.94	[3]**RB925345***	14.83	0.99
15	[3]RB855546*	101.73	0.91	[1]RB008293	13.42	0.95	[1]RB855156*	14.76	0.99
16	[3]RB987915	101.63	0.91	[3]RB998118	13.41	0.96	[1]**RB945067**	14.75	0.97
17	[1]**RB845239**	101.45	0.85	[1]RB008309	13.33	0.96	[1]RB865547	14.71	0.93
18	[1]RB957610	101.23	0.88	[1]RB008296	13.28	0.88	[1]RB937570*	14.67	0.99
19	[3]RB855536*	101.18	0.95	[3]**RB925345***	13.27	0.98	[1]RB988137	14.66	0.98
20	[1]RB965743	101.14	0.90	[1]**RB965916**	13.27	0.94	[2]RB987965	14.65	0.96

[1]Group 1; [2]Group 2; [3]Group 3; *commercial varieties. The ranked clones that are among the 20 best for more than 1 characteristic are written in boldface.

Of the 20 best genotypes for TSH, 13 belonged to group G3; of these 13 clones, 4 were commercial varieties (RB965917, RB867515, RB855546, and RB855536). The other 7 clones belonged to group G1. It is worth mentioning that none of the 12 commercial varieties belonging to group G1 were among the 20 best genotypes.

Of the 20 best genotypes for Fiber, 11 clones and 1 commercial variety belonged to group G3. Another 5 clones and 2 varieties belonged to group G1, and only the RB955980 clone belonged to group G2. Of the 20 best genotypes for SSC, 10 clones and 4 varieties belonged to the G1 group, 5 clones belonged to the G2 group, and 1 variety belonged to the G3 group (Table 4).

The RB945965 and RB845239 clones were ranked among the best 20 genotypes for TSH and SSC. The RB955980, RB965916, and RB945067 clones and the RB925345 variety were prominent for both Fiber and SSC. This result shows the potential of these genotypes as future genitors to obtain varieties that stand out in more than 1 characteristic of economic interest, for example, Fiber and SSC (Table 4).

Six genotypes were prominent in more than 1 quantity characteristic. A total of 54 clones showed superior genotype averages for at least 1 of the quantitative variables that were evaluated. The genotypes of the G1 and G3 groups stood out the most (Table 4). Therefore, biparental crossbreeding involving clones and varieties of these 2 groups can be efficient to obtain transgressive genotypes. As a complement to this study, the specific capacity of a combination is suggested to better define the heterotic groups to select the populational reciprocal recurring individual.

The pedigree used to calculate the parentage coefficient (COP) and endogamy shows the genealogy among the organisms up to the sixth generation (Table 5). The total number of genitors involved in the crossbreeding that resulted in the 113 clones is 168 genitors that are known and up to 159 unknown genitors. Because only 1 generation is considered, this number drops to 45 known genitors and 38 unknown genitors. These 45 genitors come from crossbreeding that involved another 39 known genitors and 25 unknown genitors. The number of genitors identified along the generations increases in an effective way to the first generation. However, the number of unknown genitors remains high in all of the generations. In the sixth generation, while only 5 new genitors are included in the pedigree, 20 other unknown genitors are involved in obtaining 13 clones of the fifth generation (Table 5).

Table 5. Number of known and unknown genitors (UG) in each generation according to the pedigree data of 113 clones and the pedigree integrality (PI) values for each generation.

Generation	1	2	3	4	5	6	UG	PI
1	45	45	45	45	45	45	38	0.54
2		39	39	39	39	39	25	0.60
3			37	37	37	37	22	0.62
4				29	29	29	20	0.59
5					13	13	34	0.27
6						5	20	0.20
Total						168	159	-

The endogamy coefficients were low for all of the 113 clones. These values varied between 0 and 0.019 (Table 1), showing high heterozygosis in the clones and varieties of sugarcane that were evaluated.

The Pearson's correlation coefficient (r) between the genetic similarity (GS) and the COP was only 0.15 ($P < 0.01$), showing a low association between the 2 strategies of estimating the genetic variability (Figure 5).
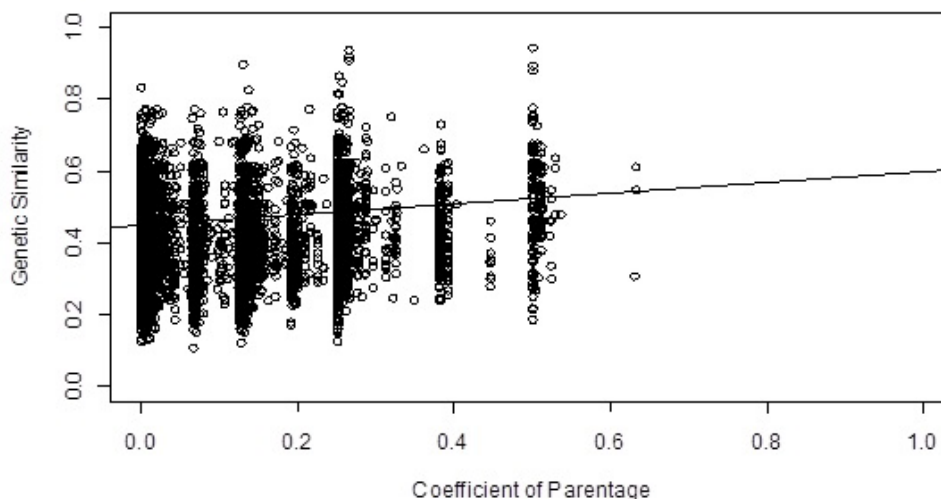
Figure 5. Diagram of the dispersion of the parentage coefficient (2 x r_{xy}) and the genetic similarity (1 - distance of Gower) among the 6328 pairs of sugarcane genotypes.

According to what was observed in the graph of the dispersion of the COP and GS values, of all the possible genotype pairs (6328) involving the 113 clones (Figure 5), only 13 pairs of genotypes presented a GS above 0.8, and the highest GS (0.94) was detected among the RB845257 and RB855113 clones that showed a COP of 0.50. The RB855113 clone also showed a high GS (0.92) with RB965505, but the COP for the 2 genotypes was only 0.26.

Generally speaking, the COPs were low, and there were only 3 pairs of genotypes with a COP above 0.6 (Figure 5). The highest COP values occurred among clones RB965743 and RB855536 (0.63), RB965920 and RB845210 (0.63), and RB977625 and RB855002 (0.63). However, the GS values among these pairs of genotypes were 0.61, 0.55, and 0.31, respectively, indicating a high correlation between the GS and COP values for the first 2 pairs of clones.

DISCUSSION

According to Gonçalves et al. (2009), the logarithmic function of probability defines the number of groups with high precision, resulting in less subjective grouping. Everitt (1981) suggests using this method only when the relationship between the number of observations (n) and the number of variables (p) is higher than 5 (n/p > 5) and n > 50, a condition observed in the data in this study, where n/p = 113/21 = 5.38.

In sugarcane, studies evaluating the genetic diversity have been conducted using quantitative and qualitative characteristics or molecular markers (Silva et al., 2008; Duarte Filho et al., 2010; Dutra Filho et al., 2011; Sindhu et al., 2011; Perera et al., 2012; Santchurn et al., 2012; Santos et al., 2012). In these studies, high GS between the genotypes was verified, resulting in the formation of fewer groups.

Just like genetic divergence, pedigree data are extremely important to define a better combination of crossbreeding between genotypes. On the basis of the pedigree, necessary information is created to avoid endogamy depression due to crossbreeding among parental organisms, increasing the efficiency of improvement programs because the heterozygosis on

the progeny is maintained, especially in cases of intra-populational recurrent selection.

In future crossbreeding, it is necessary to consider the parental relationships among the organisms. This way, the unweighted pair-group method with arithmetic mean (UPGMA) grouping (Sneath and Sokal, 1973), which uses the matrix of parentage obtained from the pedigree data, contributes to identifying the relationships among the 113 evaluated genotypes (Figure 6). Besides, the COP and endogamy estimates were more precise when the analysis considered many crossbreeding generations.
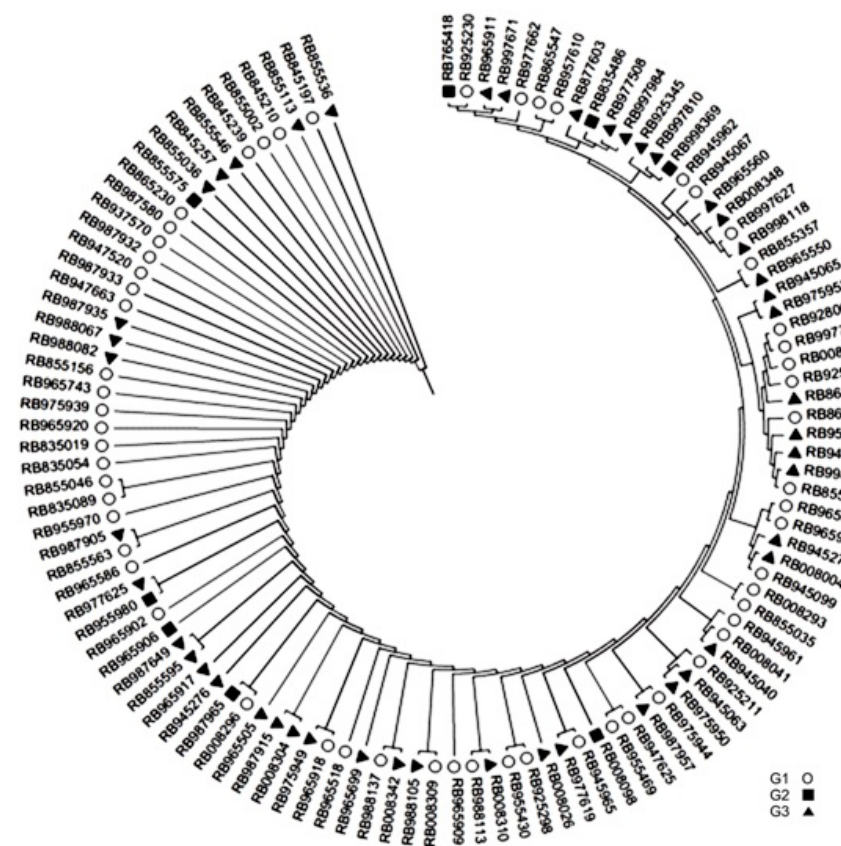
Figure 6. Dendrogram created by the unweighted pair-group method with arithmetic mean from the parentage matrix of the 113 sugarcane clones of Rede Interuniversitária para o Desenvolvimento do Setor Sucroenergético (RIDESA).

It is noticed that the average PI values increased from the sixth to the second generation, with a slight reduction in the first generation. However, the values obtained indicate a tendency of improved genealogic information quality with time (Table 5). The PI estimate is important because the endogamy coefficient of an organism depends on how much of its ascendency is known. This way, the better this understanding, the more reliable its endogamy coefficient will be estimated related to the populational base studied (Faria et al., 2010).

The COP was calculated considering many generations of genitor information, which makes evaluating the parentage degree more efficient among the clones and varieties. A good

example is the case of 2 of the best genotypes for Fiber (RB997810 and RB997671) that have different genitors (Table 1). However, according to the dendrogram created from the parentage matrix, it is possible to observe the narrow relationship between these clones (Figure 6).

The UPGMA allowed precise visualization of the parentage among the clones. Yet, there is evident disagreement in the parentage matrix among the grouping generated by the Ward-MLM analysis of diversity and by the UPGMA analysis (Figure 6). Nevertheless, the concomitant use of these 2 procedures must be applied to gain information about the evaluated genotypes.

A low correlation between the GS and COP was also observed by Lima et al. (2002) in sugarcane. Although these authors used amplified fragment length polymorphism markers, the correlation found between the GS and COP was only 0.42. Duarte Filho et al. (2010) also found a low correlation between the GS and COP (r = 0.17) evaluating sugarcane clones using simple sequence repeat markers.

Whereas most of the GS values were between 0.3 and 0.6, most of the COP values were between 0 and 0.2 (Figure 5). The low COP among the genotypes can be related to the lack of information involving a great portion of the crossbreeding in the evaluated population. Approximately half of the genitors involved in the crossbreeding were unknown. This caused low PI values in all of the generations (Table 5).

The results presented in this study are very important to define future crossbreeding among the best clones for TSH, Fiber, and SSC. Understanding of the heterotic groups indicated by the Ward-MLM method, together with the parentage information, will help define desirable and undesirable crossbreeding schemes for the genetic improvement program conducted by RIDESA. This information will also be useful for other companies that develop sugarcane varieties.

With these analyses, we were able to evaluate the genetic diversity among sugarcane clones and varieties. The analysis strategies allowed the proper grouping of genotypes, creating an efficient classification of the evaluated clones. Besides, clones were identified that are promising for TSH, Fiber, and SSC.

ACKNOWLEDGMENTS

REFERENCES

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

Barbé TD, Amaral Júnior AT, Gonçalves LSA, Rodrigues R, et al. (2010). Association between advanced generations and genealogy in inbred lines of snap bean by the Ward-Modified Location Model. *Euphytica* 173: 337-343.

Barbosa GVS, Cruz MM, Soares L, Rocha AMC, et al. (2002). A brief report on sugarcane breeding program in Alagoas, Brazil. *Crop Breed. Appl. Biotechnol.* 2: 613-616.

Barbosa MHP, Resende MDV, Dias LAS, Barbosa GVS, et al. (2012). Genetic improvement of sugar cane for bioenergy: the Brazilian experience in network research with RIDESA. *Crop Breed. Appl. Biotechnol.* S2: 87-98.

Brasileiro BP, Silva SA, Souza DR, Santos PA, et al. (2013). Genetic diversity and selection gain in the physic nut (*Jatropha curcas*). *Genet. Mol. Res.* 12: 2341-2350.

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

Cole-Rodgers P, Smith DW and Bosland PW (1997). A novel statistical approach to analyze genetic resource evaluations using *Capsicum* as an example. *Crop Sci.* 37: 1000-1002.

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

Cruz CD, Regazzi AJ and Arneiro PCS (2012). Modelos Biométricos Aplicados ao Melhoramento Genético. Editora UFV, Viçosa.

Duarte Filho LSC, Silva PP, Santos JM, Barbosa GVS, et al. (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, Anunciação Filho CJ, et al. (2011). Aplicação de técnicas multivariadas no estudo da divergência genética em cana-de-açúcar. *Rev. Cienc. Agron.* 42: 185-192.

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

Everitt BS (1981). A Monte Carlo investigation of the likelihood ratio test for the number of components in a mixture of normal distributions. *Multivar. Behav. Res.* 16: 171-180.

Faria LC, Queiroz SA, Vozzi PA, Lôbo RB, et al. (2010). Variabilidade genética da raça Brahman no Brasil detectada por meio de análise de pedigree. *Pesq. Agropec. Bras.* 45: 1133-1140.

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

Gonçalves LS, Rodrigues R, Amaral AT Jr, Karasawa M, et al. (2008). Comparison of multivariate statistical algorithms to cluster tomato heirloom accessions. *Genet. Mol. Res.* 7: 1289-1297.

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.

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

Kempthorne O (1973). An Introduction to Genetic Statistics. Iowa State University Press, Ames.

Lima ML, Garcia AA, Oliveira KM, Matsuoka S, et al. (2002). Analysis of genetic similarity detected by AFLP and coefficient of parentage among genotypes of sugar cane (*Saccharum* spp.). *Theor. Appl. Genet.* 104: 30-38.

MacCluer JW, Boyce AJ, Dyke B and Weitkamp LR (1983). Inbreeding and pedigree structure in Standardbred horses. *J. Hered.* 74: 394-399.

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

Oliveira RS, Silva SA, Brasileiro BP, Medeiros EP, et al. (2013). Genetic divergence on castor bean using the Ward-MLM strategy. *Rev. Cienc. Agron.* 44: 564-570.

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 Evol.* 55: 1055-1064.

Pereira VM, Borges CV, Brandão LP, Oliveira LS, et al. (2012). Genetic diversity between improved banana diploids using canonical variables and the Ward-MLM method. *Pesq. Agropec. Bras.* 47: 1480-1488.

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

Pestanana RKN, Amorim EP, Ferreira CF, Amorim VB, et al. (2011). Agronomic and molecular characterization of gamma ray induced banana (*Musa* sp.) mutants using a multivariate statistical algorithm. *Euphytica* 178: 151-158.

Peternelli LA, Ferreira FM, Rocha RB, Barros WS, et al. (2009). Análise dos coeficientes de endogamia e de parentesco para qualquer nível de ploidia usando o pacote estatístico R. *Bragantia* 68: 849-855.

R Development Core Team (2013). R: A Language and Environment for Statistical Computing. Available at [http://www.R-project.org]. Accessed January 20, 2013.

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

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, Silva PP, et al. (2012). Genetic diversity of the main progenitors of sugarcane from the RIDESA germplasm bank using SSR markers. *Ind. Crops Prod.* 40: 145-150.

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

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.

Silva PP, Soares L, Costa JG, Viana LS, et al. (2012). Path analysis for selection of drought tolerant sugarcane genotypes through physiological components. *Ind. Crops 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. Biotechnol.* 20: 118-124.

Sneath PHA and Sokal RR (1973). Numerical Taxonomy. Freeman, San Francisco.