African Journal of Biotechnology Vol. 11(21), pp. 4722-4728, 13 March, 2012
Available online at http://www.academicjournals.org/AJB
DOI: 10.5897/AJB11.2614
ISSN 1684–5315 © 2012 Academic Journals

Full Length Research Paper

DNA fingerprinting based on simple sequence repeat (SSR) markers in sugarcane clones from the breeding program RIDESA

Dennis Crystian Silva[1], Luiz Sérgio Costa Duarte Filho[1], João Messias dos Santos[2], Geraldo Veríssimo de Souza Barbosa[2] and e Cícero Almeida[1*]

[1]Laboratory of Genetics Resources, Campus Arapiraca, Federal University of Alagoas, Brazil.
[2]Center for Agricultural Sciences, Federal University of Alagoas, PMGCA/RIDESA/UFAL, Brazil.

Accepted 3 November, 2011.

New varieties of sugarcane are protected using morphological descriptors, which have limitations in identifying morphologically similar cultivars. Development of a reliable DNA fingerprint system for identification of new varieties would contribute greatly to the breeding of these species. Microsatellite markers are tools with great potential to generate unique genetic profiles that identify individuals without the presence of errors. The aim of this study was to analyze molecular markers to assist in the protection of new varieties of sugarcane from RIDESA (RB varieties). DNA sequences containing microsatellites were extracted from the expressed sequence tag (EST) database of sugarcane and analyzed in 15 accessions. Markers showed a high number of alleles and low nonspecific amplification. For the three markers analyzed, the polymorphic information content showed high values and the genetic identity of all markers was 1×10^{-6}. Accessions showed an average genetic similarity of 0.46 and consisted of four distinct groups. Primers used in this experiment were able to discriminate all accessions of sugarcane analyzed, demonstrating the ability of the methodology to determine unique genetic profiles. Thus, DNA fingerprinting is a useful tool to aid in the protection of new RB varieties developed by the breeding program RIDESA.

Key words: Varieties protection, genetic identity, SSR molecular markers, DNA fingerprinting, sugarcane.

INTRODUCTION

The breeding program RIDESA (Rede Interuniversitária para o Desenvolvimento do Setor Sucroenergético) has developed RB varieties (República do Brasil) for the different crop environments in Brazil, contributing to the current levels of crop productivity. Currently, RIDESA's varietal collection consists of 78 commercial varieties that occupy 58% of the crop area in Brazil. However, there is no reliable system to aid correct identification and protection of these varieties leading frequently to identification errors and facilitating the piracy of protected biological material.

Morphological descriptions are used routinely to determinethe varietal identity; however, such descriptions are inaccurate due to environmental influence on character expression. In addition, the reduced number of descriptors makes it difficult to identify the large number of varieties, requiring alternative tools to help varietal identification. Genome sequencing is a tool capable of differentiating similar organisms; however, this approach is prohibitively expensive and impractical given the large number of sugarcane varieties. A great number of fingerprinting techniques are in currency, which are variations and combinations of restriction fragmentation and polymerase chain reaction (PCR). Using PCR technology, a number of simpler techniques have become available for molecular characterization. These include, random amplified polymorphic DNA (RAPD) and its several variants, inter simple sequence repeats (ISSR), amplified fragment length polymorphism (AFLP) and simple sequence repeats (SSR), also known as

*Corresponding author. E-mail: cicerocarlos@hotmail.com.

Table 1. Identification of the 15 sugarcane accessions used for DNA fingerprinting with pedigree and their origin.

Genotype	Pedigree	Origem
RB835560	(H49-4566 x ?) US59-16-1 x RB72454 (CP53-76 x ?)	Brazil
TUC73-518	L66-48 x US66-569	Argentine
RB991562	(BO17 x IAC 50/134) SP80-185 x ?	Brazil
RB845286	(CP57-76 x ?) RB72454 x SP70-1143 (IAC 48/65 x ?)	Brazil
RB991511	(NCo310 x 54N7096) Q135 x RB835054 (NA56-79 x RB72454)	Brazil
SP80-4967	SP71-3079 x ?	Brazil
RB961	(RB72454 x B3337) RB83594 x B4362 (B37161 x POJ2878)	Brazil
RB72454	(F36-819 x CP36-46) CP53-76 x ?	Brazil
SP79-1011	(Co419 x SELF) NA56-79 x Co775 (POJ2878 x Co371)	Brazil
RB971765	(NA56-79 x Q47) RB8533 x RB835019 (RB72454 x NA56-79)	Brazil
RB961552	(B37161 x POJ2878) B4362 x IAC68/12 (Co419 x IAC52/179)	Brazil
RB9629	(IAC48/65 x RB705096) RB80457 x ?	Brazil
RB991532	(C278 x ?) RB75126 x RB855536 (SP70-1143 x RB72454)	Brazil
RB971754	(NCo310 x PT43-52) F150 x RB739359 (IANE55-34 x ?)	Brazil
RB971741	(H53-3989 x ?)RB825336 x RB91524 (Co6806 x RB72454)	Brazil

microsatellites. The SSR is certainly the most efficient way of resolving issues of identity, because they are typically codominant and multiallelic and due to the specificity of the PCR assay and elevated information content, it as well allows the determination of identity between individual based in formal estimates derived from allele frequencies (Kirst et al., 2005).

SSRs are recommended by UPOV (International Union for the Protection of New Varieties of Plants) as an auxiliary tool to minimum morphological descriptors because it is PCR-based, has high reproducibility and repeatability, high discriminatory power, allows storage in a database, and has accessible methodology (Xu and Crouch, 2008; Chen et al., 2009). SSRs can be obtained through genomic libraries or sequences deposited in the EST (Expressed Sequence Tag) database. The latter allows the development of markers with low cost and eliminates the need for DNA sequencing (Pinto et al., 2004).

SSR markers are repeated DNA sequences in tandem, ranging from one to six nucleotides in length (Kalia et al., 2011). These markers are frequently found in the genomes of eukaryotic organisms, and are usually highly polymorphic and co-dominant (Xu and Crouch, 2008; Chen et al., 2009). Several studies have used SSR markers for genetic analysis in sugarcane (Cordeiro et al., 2000, 2001, 2003; Pinto et al., 2004; Duarte et al., 2010); however, the breeding program RIDESA lacks studies on the implementation of a DNA fingerprinting system to identify protected varieties. The aim of this study was to develop a robust system of DNA fingerprinting for sugarcane varieties from RIDESA based on SSR markers.

MATERIALS AND METHODS

DNA extraction

Fifteen accessions (Table 1) of sugarcane were obtained from the germplasm bank of the Sugarcane Breeding Program, Federal University of Alagoas Brazil. These accessions are modern sugarcane, derived from interspecific crosses between noble cane *Saccharum officinarum* (x = 10, 2n = 8x = 80) and the wild specie *Saccharum spontaneum* (x = 8, 2n = 5 -16x = 40 - 128), which is estimated that 70 to 80% are derived from *S. officinarum*, 10 to 20% from *S. spontaneum* and approximately 10% are recombinants between the two genomes, which are aneuploid and polyploidy, with 2n = 100 - 130 chromosomes (D'Hont et al., 1996).

The genomic DNA of genotypes was extracted from young leaf tissue using the procedure as described by Saghai-Marrof et al. (1984). Quality and integrity of genomic DNA was assessed by agarose gel electrophoresis (1%) revealed with ethidium bromide.

SSR markers and DNA amplification

Microsatellite sequences were searched from the EST database of sugarcane deposited in the Gene Index Project (http://compbio.dfci.harvard.edu/tgi/) using the software TRF (Tandem Repeats Finder) (Benson, 1999). The main parameters for obtaining primers were defined according to the following characteristics: primer length between 18 to 22 nucleotides, PCR product size between 100 to 300 bp, 60°C optimum annealing temperature and 40 to 65% GC content. Primers were designed using the software Primer 3 (http:// primer3.sourceforge.net) and called SCC.

For PCR amplification, a final volume of 50 µl was used containing: 60 ng genomic DNA, 10X enzyme buffer, 1.5 mM $MgCl_2$

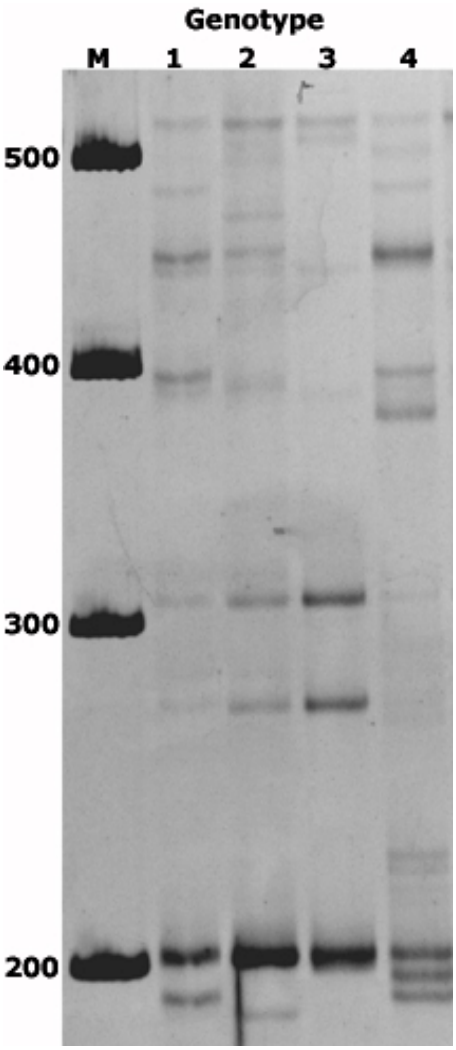

Figure 1. Polyacrylamide gel banding patterns for microsatellite markers (SCC82) in four genotypes (Lane M -100 pb, lane 1 - RB835560, lane 2 - SP80-4967, lane 3 – RB961552, lane 4 – RB991532).

(Gibco BRL), 0.2 mM dNTP (Gibco BRL), 1U Taq DNA polymerase (Gibco BRL), and 30 pmol of each primer. Amplification was performed in a thermocycler (PTC-100, MJ Research, Inc.) with 35 cycles of 94°C for 1 min, 60°C for 1 min and 72°C for 1 min and a final extension of 72°C for 10 min. The PCR amplification product was separated on a 6% polyacrylamide gel, using a 100 bp marker (Fermentas Life Science) for molecular weight determination of the respective amplified DNA fragments.

Data analysis

Despite being considered co-dominant SSR markers, in this study they were considered dominant markers, because in highly polyploid genomes such as that of sugarcane microsatellites have difficulty in distinguishing the alleles of homologous chromosomes, making it difficult to determine the heterozygosity or homozygosity at any particular locus (Cordeiro et al., 2003; Oliveira et al., 2009). From this assumption, all possible alleles detected in the progenies were converted to a binary system. For each genotype, clear and distinct alleles were classified as absent (0) or present. To form a binary matrix that was used to estimate the following variables:

1. Number of alleles with absent or present among genotypes (n_p).
2. Number of alleles without absent among genotypes (n_{np}).
3. Confusion probability (C_j) of the *j*th primer (Tessier, 1999):

$$C_j = \sum_{i=1}^{I} p_i \frac{(N.p_i - 1)}{N - 1}$$

Where, p_i is the frequency of the *i*th pattern and *N* is the number of individuals.

4. Discriminatory power (D_j) of the *j*-th primer (Tessier, 1999):

$$D_j = 1 - C_j \ .$$

5. Polymorphism Information Content (PIC) obtained by the expression:

$$PIC = 1 - \sum pi^2$$

Where, p_i^2 is the squared frequency of allele *i* for the *j*th marker (Powell et al., 1996).

6. The probability of identity $P_{(ID)}$ or probability that two individuals in a population have the same genotype at multiple loci was computed considering the microsatellites as dominant markers, with expression $P_{(ID)dom} = p^2 + (2pq)^2 + (q^2)^2$ as described by Waits et al. (2001), where *p* is the frequency of the allele presence and *q* the frequency of allele absence.
7. The genetic similarity (GS) between pairs genotypes was calculated using the Jaccard coefficient, obtained by the expression GS = *a/(a + b + c),* where GS is the measure of genetic similarity between *i* and j individuals. For each pair of accesses, *a* represents the number of coincidences of type 1-1, , *b,* number of dislocations of type 1-0 and c, number of dislocations of type 0-1(Reif et al., 2005).

All analyses were performed using Microsoft Excel 2007 and genetic similarity values were used to generate a cluster using the UPGMA method using the software NTSYS 2.1.

RESULTS

Microsatellites

Ninety-eight microsatellite markers were amplified, and then the presence and quality of amplification products was verified on acrylamide gels. Of these, three primers were selected because they had banding patterns of high quality and low nonspecific amplification (Figure 1). The three SSR loci amplified a total of 31 alleles in fifteen genotypes, with size ranging from 192 to 505 bp and the number of alleles ranging from 2 to 8 per individual, showing the polyploid nature of sugarcane (Table 2). The number of alleles with absent or present among genotypes ranged from 6 to 12, and the number of alleles without absent among genotypes was low with two in the

Table 2. Microsatellite primer sequences, the confusion probabilities (C_j), discriminating power (Dj), probability of identity ($P_{(ID)dom}$), polymorphism information content (PIC); Number of alleles with absent or present among genotypes (n_p); Number of alleles without absent among genotypes (n_{np}); Allelic range (bp) and no. of alleles per individual.

SSR	GenBank number	Motif	Primer (5'- 3')	C*j*	D*j*	PIC	$P_{(ID)dom}$	N_p	N_{np}	Allelic range (bp)	Number of allelic per individual
SCC82	CA205346	$CTC_{(10)}$	F: CTATCCCATCCCGGAAAAA. R: CCGACTTGAACACCACCAG	0.134	0.866	0.875	0.008	10	2	231-505	4-8
SCC89	CA207738	$CCT_{(9)}$	F:AGTGTTGCGAGAAGCAGCAG R:CCCATGGATCACATGACAGA	0.177	0.823	0.835	0.006	12	0	192-276	4-8
SCC93	CA210595	$GCA_{(9)}$	F: AATCCCAGCCCCGATGAT R: AGCCACACCTTGACCTTGAC	0.267	0.733	0.752	0.027	6	1	216-291	2-7
Mean	-	-	-	0.193	0.807	0.821	0.014	9.3	1	313	3.8
Total	-	-	-	-	-	-	$1 . 10^{-6}$	28	3	-	-

primer SCC 82 and one in the primer SCC93. The primers used in this work showed good distribution allele, with allele more frequent and rare alleles. Among the primers, the SCC82 showed more uniform distribution, while the SCC93 showed two alleles with greater frequency and five alleles with a lower incidence (Figure 2). Note that the primer SCC82 showed high PIC (Table 2), reflecting the large number of alleles observed and similar allele frequency distribution when compared with other primers.

Identification of varieties

The PIC average value among the SSR primers was 0.821, with a minimum of 0.752 (SCC93) and 0.875 maximum (SCC82). The discriminatory power of primers showed 0.807 mean and the primer SCC82 showed the highest value (0.866). The probability of confusion ranged from 0.134 for the marker SCC82 to 0.267 in the SCC93 initiator. However, the genetic identity, the probability of two genotypes being identical in a population, ranged from 0.006 (SCC89) to 0.027 (SCC93); the combined value for all the three primers was 1×10^{-6}, that is, one in a million accessions may have the same genotype (Table 2).

The genetic similarity among genotypes ranged from 0.76 between RB835560 and TUC73-518 to 0.25 between varieties RB971741 and RB991562, with 0.46 average genetic similarities. Using the average genetic similarity, the clustering (Figure 3) divided the accessions into four distinct groups, group I and II formed by five cultivars each, group III consisting of two cultivars and group IV containing three cultivars.

DISCUSSION

The presence of microsatellite molecular markers in the EST sequences of sugarcane is relatively low (Pinto et al., 2004; Duarte et al., 2010); 20 times lower than the frequency of SSRs derived from enriched libraries (Cordeiro et al., 2000, 2001). However, given that EST-SSR markers are derived from transcribed DNA regions, development of these markers will enable better efficiency in plant breeding when used for assisted selection (Wang et al., 2007). In addition, the increased transferability of these markers due to their presence in the conserved regions is another important advantage, especially in crops with a complex genome such as sugarcane (Swapna et al., 2010).

With regards to the level of polymorphism, EST-SSR markers showed less variation compared with markers derived from genomic libraries (Cho et al., 2000), possibly due to greater conservation of DNA sequences in transcribed regions. Microsatellite markers have become highly useful in the genetic analysis of sugarcane because they enable the genetic discrimination of individuals (Cordeiro et al., 2001; Pinto et al., 2004, 2006; Oliveira et al. 2009 Duarte et al., 2010).

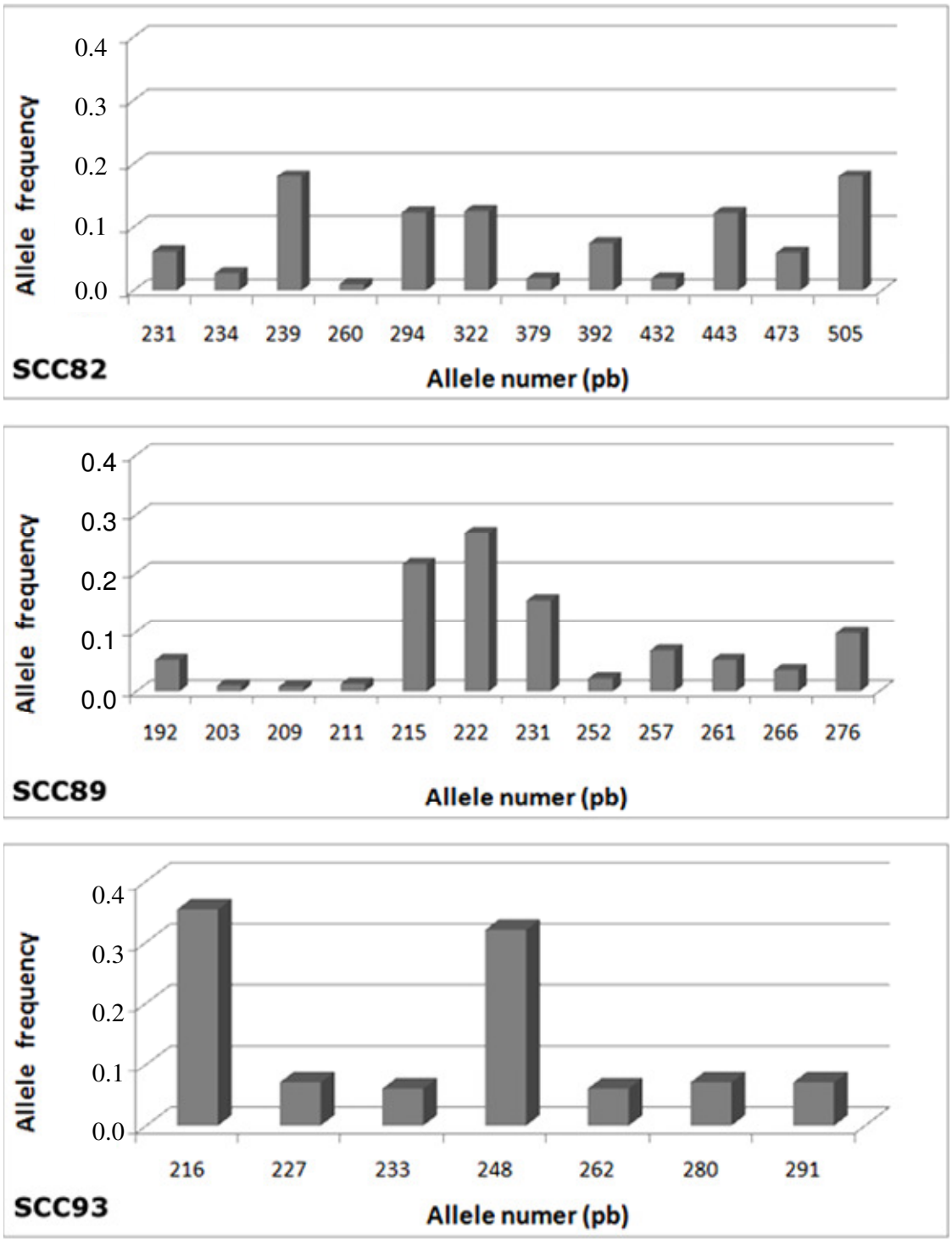

Figure 2. Allele frequency distributions of three EST-SSR loci in the 15 accessions of sugarcane.

Microsatellite markers exhibited high PIC values in other studies of sugarcane (Cordeiro et al., 2003; Pinto et al., 2004, 2006; Pan et al., 2006; Oliveira et al., 2009), suggesting they were suitable for use in the protection of new varieties. The ability of a few markers to generate unique genetic profiles in sugarcane accessions is due to the polyploid nature of the genome, which permitted the detection of several allelic types in a single accession (Cordeiro et al., 2001; Pinto et al., 2004; Duarte et al., 2010; Maccheroni et al., 2009). However, polyploidy also makes it difficult to calculate population genetic statistics and to score complex genotypes, making it more

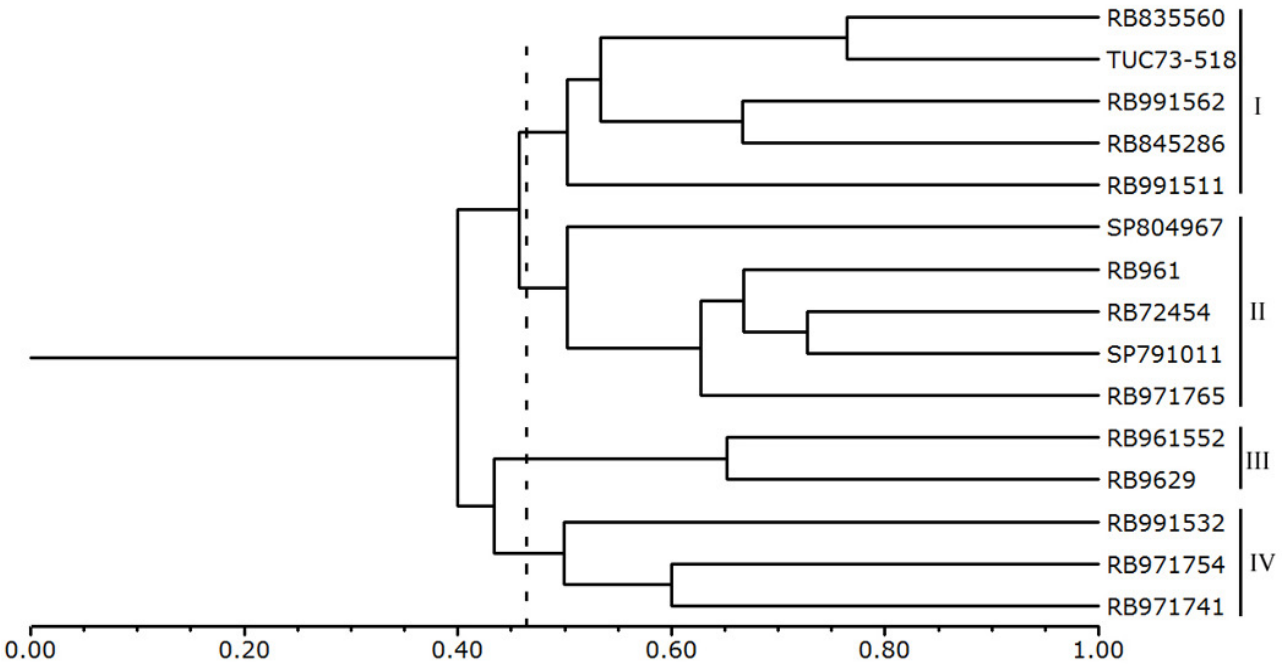

Figure 3. Dendrogram representing the relationship among the 15 varieties of sugarcane, based on Jaccard's similarity, using UPGMA cluster analysis. I, II, III and IV – groups of the sugarcane.

appropriate to analyze the SSR markers as dominant.

The high PIC and D_j values obtained by this study are due mainly to the presence of unique alleles in certain accessions that genetically differentiate them from other accessions (Figure 2 and Table 2). Similar results were obtained by Oliveira et al. (2009), where high values of D_j and PIC were obtained in 135 SSR of sugarcane.

The utility of specific markers for identifying individuals can be quantified by the probability of genetic identity P ($_{ID}$). The present study showed low probabilities of genetic identity (1×10^{-6}) for all markers, corroborating the results obtained by De-Lucas et al. (2008), where P ($_{ID}$) was 1.18×10^{-9} using SSR markers in *Populus*. Similarly, the results obtained by Pollegioni et al. (2009), in *Nogueira*, showed P ($_{ID}$) values close to 1×10^{-4}. In sugarcane, Maccheroni et al. (2009), analyzing more than 1,200 cultivated individuals of *Saccharum* (commercial cultivar, *S. officinarum*, *S. spontaneum*, *S. barberi*, *S. sinense*, *S. edule* and *Erianthus bengalensis*), found the value 1×10^{-7}. It should be noted that accessions of sugarcane share many of the same genomic regions (Cordeiro et al., 2001, 2003; Pinto et al., 2004, 2006; Glynn et al., 2009; Duarte et al., 2010), which may contribute to a reduced discriminatory power of molecular markers.

Determining the genetic profile of varieties has great importance in breeding, allowing for the identification of duplicates in the germplasm bank and the organization of crosses leads to a better breeding result (Ruiz et al., 2011). In addition, it enables the establishment of criteria for the protection of varieties, with the genetic profile included as a unique feature of each access. The identification of varieties based on molecular markers is important to establish distinctness, uniformity and stability of protected cultivars (Swapna et al., 2010). The microsatellites analyzed in this study are highly polymorphic, allowing for the identification of individual clones and the generation of unique genetic profiles of sugarcane that can be used by the RIDESA breeding program to assist the protection of new RB varieties.

REFERENCES

Benson G (1999). Tandem repeats finder: a program to analyze DNA sequences. Nucleic Acids Res. 27: 573-580.

Chen PH, Pan YB, Chen RK Xu LP, Chen YQ (2009). SSR marker-based analysis of genetic relatedness among sugarcane cultivars (*Saccharum* spp. hybrids) from breeding programs in China and other countries. Sugar Tech. 11(4): 347-354.

Cho YG, Ishii T, Temnykh S, Chen X, Lipovich L, Mccouch SR, Park WD, Ayres N, Cartinhour S (2000). Diversity of microsatellites derived from genomic libraries and GenBank sequences in rice (*Oryza sativa* L.). Theor. Appl. Genet. 100(5): 0713-0722.

Cordeiro GM, Casu R, Mcintyre CL, Manners JM, Henry RJ (2001). Microsatellite markers from sugarcane (*Saccharum* spp.) ESTs cross transferable to *Erianthus* and sorghum. Plant Sci. 160: 1115-1123.

Cordeiro GM, Pan YB, Henry RJ (2003). Sugarcane microsatellites for the assessment of genetic diversity in sugarcane germplasm. Plant Sci. 165: 181-189.

Cordeiro GM, Taylor GO, Henry JR (2000). Characterization of microsatellite markers from Sugarcane (*Saccharum* sp.) a highly polyploid species. Plant Sci. 155: 161-168.

De-Lucas AI, Santana JC, Recio P, Hidalgo E (2008). SSR-based tool for identification and certification of commercial *Populus* clones in Spain. Ann. For. Sci. 65: p. 107.

Duarte LSC, Silva PP, Santos JM, Barbosa GVS, Ramalho CE, Soares

L, Andrade JCF, Almeida C (2010). Genetic similarity among genotypes of sugarcane estimated by SSR and coefficient of parentage. Sugar Tech. 12(2): 145-149.
Glynn NC, Mccorkle K, Comstock C (2009). Diversity among mainland USA sugarcane cultivars examined by SSR genotyping. J. Am. Soc. Sugar Cane Technol. 29: 36-52.
Kalia RK, Rai MK, Kalia S, Singh R, Dhawan AK (2011). Microsatellite markers: an overview of the recent progress in plants. Euphytica, 177: 309-334.
Kirst M, Cordeiro CM, Rezende GDSP, Grattapaglia D (2005). Power of Microsatellite Markers for Fingerprinting and Parentage Analysis in *Eucalyptus grandis* Breeding Populations. J. Hered. 96: 161-166.
Maccheroni W, Jordão H, De Gaspari R, De Moura GL, Matsuoka S (2009). Development of a dependable microsatellite-based fingerprinting system for sugarcane. Sugarcane Int. 27(2): 8-13.
Oliveira KM, Pinto LR, Marconi TG, Mollinari M, Ulian EC, Chabregas SM, Falco MC, Burnquist W, Garcia AAF, Souza AP (2009). Characterization of new polymorphic functional markers for sugarcane. Genome, 52: 191-209.
Pan Y (2006). Highly Polymorphic Microsatellite DNA Markers for Sugarcane Germplasm Evaluation and Variety Identity Testing. Sugar Tech. 8: 246-256.
Pinto LR, Oliveira KM, Marconi T, Garcia AAF, Ulian EC, de Souza AP (2006). Characterization of novel sugarcane expressed sequence tag microsatellites and their comparison with genomic SSRs. Plant Breed. 125: 378-384.
Pinto LR, Oliveira KM, Ulian EC, Garcia AAF, De Souza AP (2004). Survey in the sugarcane expressed sequence tag database (SUCEST) for simple sequence repeats. Genome, 47: 795-804
Pollegioni P, Woeste K, Mugnozza GS, Malvolti ME (2009). Retrospective identification of hybridogenic walnut plants by SSR fingerprinting and parentage analysis. Mol. Breed. 24: 321-335.
Powell WM, Morgante C, Andre M, Hanafey J, Vogel Tingey, Rafalski A (1996). The comparison of RFLP, RAPD, AFLP, and SSR (microsatellite) markers for germplasm analysis. Mol. Breed. 2: 225-238.
Reif JC, Melchinger AE, Frisch M (2005) Genetical and mathematical properties of similarity and dissimilarity coefficients applied in plant breeding and seed bank management. Crop Sci. 45: 1-7.
Ruiz M, Aguiriano E, Giraldo P, Carrillo JM (2011). Genetic redundancy among durum wheat accessions as assessed by SSRs and endosperm proteins. Span. J. Agric. Res. 9(1): 156-165.
Saghai-Marrof MA, Soliman KM, Jorgensen RA, Allard RW (1984). Ribosomal DNA Spacer-length polymorphism. Proc. Natl. Acad. Sci. 81: 8014-8018
Swapna M, Sivaraju K, Sharma RK, Singh NK, Mohapatra T (2010). Single-Strand Conformational Polymorphism of EST-SSRs: A Potential Tool for Diversity Analysis and Varietal Identification in Sugarcane. Plant Mol. Biol. Rep. 29(3): 505-513.
Tessier C, David J, This P, Boursiquot JM, Charrier A (1999). Optimization of the choice of molecular markers for varietal identification in *Vitis vinifera* L. Theor. Appl. Genet. 98:171-177.
Waits L, Luikart G, Taberlet P (2001). Estimating the probability of identity among genotypes in natural populations: cautions and guidelines. Mol. Ecol. 10: 249-256.
Wang H, Wei Y, Yan ZE, Zheng Y (2007). EST-SSR DNA polymorphism in durum wheat (*Triticum durum* L.) collections. J. Appl. Genet. 48(1): 35-42.
Xu YB, Crouch JH (2008). Marker-assisted selection in plant breeding: From publications to practice. Crop Sci. 48: 391-407.