Molecular polymorphism in SSR loci of the RB92579 variety: perspectives for use in sugarcane improvement and industrial sector

Rodrigo Desordi[1], Claudete Aparecida Mangolin[2], Gustavo Barizon Maranho[3], Rone Charles Maranho[3], Maria de Fátima Pires da Silva Machado[2*]

[1]Genetics and Breeding, State University of Maringá, 87020-900 Maringá, PR, Brazil
[2]Department of Biotechnology, Genetics and Cell Biology, State University of Maringá, 87020-900 Maringá, PR, Brazil
[3]Agronomy, State University of Maringá, 87020-900 Maringá, PR, Brazil

*Corresponding author: mfpsmachado@uem.br

Abstract

The sugarcane variety RB92579 has excellent agricultural productivity, very low flowering, efficient water use, and a high content of sucrose. Despite its excellent agricultural productivity, the RB92579 has not been used as a direct parent in sugarcane improvement. The main goal of the present study was to investigate polymorphisms at the SSR and EST-SSR loci of the RB92579 sugarcane variety to evaluate its potential for breeding and generating new varieties and to guide better use by the industrial sector. A total of 92 samples of the RB92579 variety were collected from plants in the fourth cutting stage grown in two Brazilian states: Paraná (PR; South region) and Mato Grosso do Sul (MS; South-Central region). Four primers for DNA simple sequence repeats (SSRs) and eight primers for expressed sequence tags for simple sequence repeats (EST-SSR) were used for DNA amplification. The polymorphism occurrence in the 12 SSR loci was 28% in the PR and MS populations, with a total of 25 alleles and an average of 2.08 alleles/loci. High values for mean observed heterozygosity, a high value for genetic identity and a low level of population differentiation was found in samples from the PR and MS states. The number of polymorphisms in the EST-SSR and noncoding SSR loci as well as the genetic divergence was low. However, the high heterozygosity in both populations indicates that the RB92579 variety can be used as a parent to generate new cultivars. On the other hand, the low coefficient of genetic divergence and high identity coefficient indicate that there is genetic uniformity; therefore, there is no need for differential industrial adaptations for pretreatment or enzymatic hydrolysis of the sugarcane bagasse from RB92579 at the same cutting stage and planted in the two regions (PR and MS).

Keywords: *Saccharum* spp., molecular polymorphism, microsatellite markers, EST-SSR markers, genetic divergence.
Abbreviations: SSR_simple sequence repeats; EST-SSR_ expressed sequence tags for simple sequence repeats; RIDESA_Inter-University Network for the Development of Sugarcane Industry; SUCEST_sugarcane expressed sequence tag database; EDTA_Ethylenediamine tetraacetic acid); CTAB_Cetyl trimethyl ammonium bromide; PVP-40_Polyvinylpyrrolidone (40 KDa).

Introduction

The use of sugarcane bagasse (*Saccharum* spp.) as a source of electricity and production of second-generation ethanol (2G) has stimulated interest in the development of new varieties of sugarcane. The establishment of two companies for the production of second generation ethanol (2G ethanol) in the states of Alagoas and São Paulo has now alerted sugarcane breeding programs to the need for new varieties with increased biomass (fiber and straw) and deep-rooted plants with higher capacities to produce rhizomes adapted to heterogeneous regions of Brazil (http://www.granbio.com.br/en/wp-content/uploads/sites/2/2014/09/partida_english.pdf).
Straw and the bagasse are the raw materials for 2G ethanol production. "Restart the improvement of sugarcane" is currently a slogan for the production of 2G ethanol.
In addition to characteristics of agronomic and industrial interest, it is important for breeding programs to understand the genetic diversity of sugarcane varieties to choose the parental lines that are promising for the generation of new varieties. Breeding programs frequently adopt microsatellite markers as a molecular diagnosis to investigate the genetic diversity in sugarcane (Cordeiro et al., 2000; 2001; Cordeiro and Henry, 2001; Pan et al., 2003; Pinto et al., 2004; Chen et al., 2009; Duarte-Filho et al., 2010; Hameed et al., 2012; You et al., 2013; Chandra et al., 2014; Maranho et al., 2014; Augusto et al., 2015). Microsatellite markers also denoted as simple sequence repeat loci (SSR loci), have been used to investigate the genetic diversity of RB varieties (Duarte-Filho et al., 2010; Maranho et al., 2014; 2016; Augusto et al., 2015) and have contributed to the development of a breeding program within the Inter-University Network for the Development of Sugarcane Industry (Rede Interuniversitária para o Desenvolvimento do setor Sucroenergético; RIDESA). The research program for the genetic improvement of RIDESA is the largest in the world (Oliveira et al., 2015). The comprehensive selection of superior varieties with agro-industrial value to meet the

diverse environmental conditions of the Brazilian regions has always been a goal of RIDESA.

In studies examining genetic divergence among RB varieties, microsatellite markers have revealed contrasting results for genetic diversity within the RB92579 variety. For example, the sugarcane variety RB92579 has remarkable features, including excellent agricultural productivity, very low flowering, efficient water use and a high content of sucrose. According to the census of the CONAB 2018 for the 2018/2019 harvest, the RB92579 variety occupies the most cultivated area (35.9%), in the states of northeastern Brazil (Alagoas, Bahia, Maranhão, Pernambuco, Paraíba, Rio Grande do Norte, and Piauí). Production of the RB92579 variety in the 2018/2019 harvest was approximately 48,032 ton/h in the North/Northeast regions of Brazil. In the Center-South regions of Brazil, the RB92579 variety occupies the 4th most cultivated area (6.1%). This region includes the states of Parana with 1.7% (8.857 hectares) and Mato Grosso do Sul state with 1.9% (11.300 hectares) of their area being used to cultivate the variety RB92579 (Tonielo et al., 2018). Production of the RB92579 variety in the 2018/2019 harvest was approximately 587,467 ton/h in the Central/Southern region of Brazil (CONAB, 2018).

Based on an analysis of 17 SSR loci, a high level of polymorphism and a high value for mean observed heterozygosity was found for the RB92579 variety (Maranho et al., 2014), in contrast with the lower number of polymorphisms and lower values of mean observed heterozygosity observed using 10 SSRs developed from expressed sequence tags (EST-SSR loci; Augusto et al., 2015). We suspect that the difference in estimates for the genetic diversity within the variety RB92579 is due to the different SSR loci that were used (noncoding SSR and EST-SSR loci). Our premise in the present study is that the analysis of a relatively larger number of plants of the variety RB92579 will provide more reliable values for genetic diversity at both SSR loci and EST-SSRs. Despite its excellent agricultural productivity, the RB92579 has not been used as a direct parent in sugarcane improvement. Estimates of genetic diversity and population structure of the variety RB92579 are important to build a database of RB varieties to evaluate the potential of this variety for breeding to generate new varieties of sugarcane and to guide its better use by the industrial sector. Our goal in this study was to investigate polymorphisms in SSR and EST-SSR loci to assess the genetic diversity and population structure within the sugarcane variety RB92579 using larger number of plants to know if the RB92579 variety may be used as a parent to generate new cultivars.

Results

Polymorphism in EST-SSR and SSR loci of the RB92579 variety of sugarcane

Analysis of the 12 SSR loci in the 92 samples of the RB92579 variety of sugarcane in the fourth cutting stage revealed three alleles in loci *Smc2017FL* (*225/244/276*), *Smc226CG* (*138/147/174*), *EstC45* (*92/113/138*), and *EstC80* (*190/213/245*), and two alleles in loci *mSscir52* (*132/144*), *EstB58* (*184/200*), *EstC69* (*156/172*), *EstC84* (*142/200*), and *EstC119* (*150/163*), whereas the *Smc477CG* (*134*), *EstB149* (*135*), and *EstC113* (*295*) loci were monomorphic. The polymorphism in the 12 SSR loci was 28% in the PR and MS populations with a total of 25 alleles, and an average of 2.08

alleles/loci. Lower number of alleles was detected at simple sequence repeats in DNA sequences that are expressed (EST-SSR) than at simple sequence repeats that are not expressed (2.25 alleles per locus). The percentage of polymorphic loci was also lower (25%) at EST-SSR loci in the PR and MS populations than the polymorphisms in the 12 SSR loci.

Genetic diversity within and between the RB92579 populations

High values for mean observed heterozygosity (H_o) were detected in the RB92579 variety cultivated in PR ($H_o = 0.849$) and MS ($H_o = 0.820$). The Nei identity (Nei, 1972) value calculated from analysis of the 12 microsatellite loci in RB92579 samples from two sugarcane fields showed high genetic identity (I = 0.997). Low population differentiation (G_{ST}) was observed in samples from PR and MS states (G_{ST} = 0.0135). The G_{ST} value indicated that only 1.35% of the total variance in allele frequencies at the SSR and EST-SSR loci was due to genetic differences among the PR and MS populations. A low level of population differentiation (G_{ST} = 0.0176) was also detected in samples from PR and MS states from analysis of the 8 EST-SSR loci in RB92579. Nei´s original measure of genetic identity in the 8 EST-SSR loci was also high (I = 0.9968). AMOVA also showed greater genetic variation within (96%) than among (4%) samples cultivated in PR and MS states.

Genetic relationships between the RB92579 plants from both PR and MS states

The dendrogram, constructed from 12 SSR loci according to the Jaccard coefficient and drawn from a reference tree by the TreeView program (Page, 2001), revealed three large groups formed by plants from both PR and MS states (Figure 2). In 66.6% of the branches, the number of identical plants ranged of three to eleven, which reflects the high estimated value of genetic identity among plants from PR (samples 1 to 46) and MS (samples 47 to 92). On the other hand, the dendrogram constructed from 8 EST-SSR loci identified two larger groups, also formed by plants from both PR and MS states (Figure 3). In 100% of the branches, the number of identical plants ranged from two to 23, which represents the highest estimated value of genetic identity among plants from PR and MS states based on EST-SSR loci. Two larger groups formed by plants from both PR and MS states were also identified in the dendrogram constructed from the 4 SSR loci that are not expressed (Figure 4).

The number of identical plants ranged from two to 55; in one of the branches, 60% of the plants from both PR and MS states were identical at the *Smc226CG*, *Smc477CG*, *Smc2017FL*, and *mSscir52* loci that are not expressed.

Figure 5 shows a three-dimensional projection obtained by the subroutine Simint and Eigen procedures of the NTSYS-pc software package (Rohlf, 1998). The plants from PR (samples 1-46) and MS (samples 47-92) were separated into two well-defined groups with some plants in isolated groups. However, the samples from PR were not distinguished from the samples from MS. Plants from PR and MS were clustered in the same groups. The near genetic distance among samples from the two states is evident in the 3D projection. These results were consistent with the dendrogram generated based on Jaccard coefficients.

Level of genetic admixture between the samples of RB92579 variety

The allele frequencies in the 12 SSR and EST-SSR loci in the 92 RB92579 samples distributed the plants into three ancestral groups according model-based Bayesian statistics. Clustering of the 92 plants according to Bayesian statistics is shown in Figure 6. Each bar in the graph represents a plant and its inferred proportion of SSR and EST-SSR marker admixtures for each locus on their genome. The optimal K value determined by the Bayesian analysis indicated that the plants were grouped into 3 clusters (ΔK2 = 0.00; ΔK3 = 2.1836; and ΔK4 = 0.00). The bar plot obtained for the K value (K = 3; ΔK3 = 2.1836) and the results were consistent with the interpretation that the plants show little mixture of group 2 and 3 alleles, and plants have alleles predominantly from one of the groups.

The bar graph shows that 28.4% of the samples are in the red group, 28% are in the green group, and 43.6% are in the blue group. The bar plot also shows that more than 80% of plants have one of three groups of dominant alleles. The RB92579 variety seems to have been formed by a set of plants containing alleles of three ancestral groups at the loci *mSscir52*, *Smc2017FL*, *Smc226CG*, *Smc477CG*, *EstB58*, *EstB149*, *EstC45*, *EstC69*, *EstC80*, *EstC84*, *EstC113*, and *EstC119*, but there is little mixing of alleles (in mean: 8.68% in PR and 10.66% in MS) in the same plant.

Discussion

The mean observed heterozygosity of the microsatellite markers in the 92 plants of the RB92579 variety (Ho = 0.849 in plants from PR and Ho =0.82 in plants from MS) showed consistent results with the high observed heterozygosity found by Maranho et al. (2014; Ho = 0.808) in the EST-SSR loci and noncoding SSR loci. However, there were fewer polymorphisms in the *EstB58*, *EstB149*, *EstC45*, *EstC69*, *EstC80*, *EstC84*, *EstC113*, *EstC119*, *Smc226CG*, *Smc477CG*, *Smc2017FL*, and *mSscir52* loci of the RB92579 variety cultivated in the PR and MS states (28%) than the polymorphisms at 17 other EST-SSR and SSR loci of DNA sequences that are not expressed (93%) from RB92579 cultivated in the PR and Pernambuco (PE) states. These results suggest that the highest or lowest polymorphism in the RB92579 sugarcane variety dependent on the locus analyzed.

The number of alleles in the RB92579 plants grown in PR and MS (2.25 alleles/locus) was also lower than the number of alleles detected by Maranho et al. (2014) in 17 other EST-SSR and noncoding SSR loci from the RB92579 variety cultivated in PR and PE (2.59 alleles/locus). The higher number of alleles detected in the RB92579 variety cultivated in the PR and PE states may be due to greater number of noncoding SSR loci (12) than EST-SSR loci (four) that were analyzed by Maranho et al. (2014). The present study on the RB92579 variety grown in PR and MS identified fewer alleles in the eight EST-SSR loci (2.0 alleles/locus) than in the four SSR loci of DNA sequences that are not expressed (2.25 alleles/locus). In this way, our interpretation that there is less genetic diversity in the EST-SSR loci of the RB92579 variety seems consistent. Studies with Chinese varieties of sugarcane have also revealed lower genetic diversity in EST-SSR loci (You et al., 2013). Fewer polymorphisms are also seen using EST-SSR markers in species from the *Saccharum* complex obtained from the Sugarcane Breeding Institute, in Coimbatore, India (Parthiban et al., 2018) and sugarcane varieties maintained at Indian Institute of Sugarcane Research (Siraree et al., 2018).

The fixed heterozygosity (100%) at *EstC84* (*142/200*), *EstC119* (*150/163*) and *Smc2017FL* (*225/244/276*) loci in the 92 samples of the RB92579 variety conduce to high values of Ho and also reflect the relative genetic stability of the RB92579 variety. The alleles *EstC84*[142], *EstC84*[200], *EstC119*[150], and *EstC119*[163], as well as *Smc2017FL*[225], *Smc2017FL*[244], and *Smc2017FL*[276], may have been established in the selection processes of the RB92579 variety, and presumably may be now useful to fingerprint the RB92579 variety. The potential utility of the five alleles as traits to assist the process of identification of the RB92579 variety, in addition to morpho-physiological descriptors, should be tested by comparing the genotype of the *EstC84*, *EstC119* and *Smc2017FL* loci with the genotypes of these same loci in other varieties of sugarcane.

Identification of the alleles and genotypes at the *EstC84* and *EstC119* loci may be of interest for the RB variety database because the *EstC84* and *EstC119* loci are in DNA sequences that are expressed and encode proteins involved in important metabolic processes in plants (Table 3). According the homology identified by the SUCEST database, the *EstC84* locus is associated with a DNA sequence that encodes a putative leucine-rich repeat receptor protein kinase (LRR-RLK; EC 2.7.11.1). The LRR-RLKs are transmembrane proteins that have important biological functions in almost all developmental processes throughout the life of the plant and in defense/resistance against a large range of pathogens (Diévart et al., 2011). Homology analysis also identified a putative acetyl-CoA-C-acetyltransferase (ACA; EC 2.3.1.9) as product of the DNA sequence that contains the *EstC119* loci. ACA is involved in the biosynthesis of hormones, cholesterols, and in the production of several phenolic secondary metabolites (Dyer et al., 2009; Chedgy et al., 2015). As the *Est-SSR* loci are associated with DNA expressed sequence tags, analysis of *Est-SSR* loci permits association between gene and traits of agronomic interest.

In contrast, with the relative genetic stability at the *EstC84*, *EstC119*, and *Smc2017FL* loci (100% heterozygosity), high genetic divergence was detected in RB92579 samples in the second cut stage grown in the states of PR and PE due to different numbers and frequencies of alleles at loci *EstA68*, *EstC66*, *EstC67*, *EstC69* and *EstC91* (Augusto et al., 2015). For example, at the locus *EstC67* from RB92574 cultivated in PR, 93.3% of the samples were homozygous for the allele *EstC67*[2], whereas 100% of samples of the same variety grown in PE were homozygous for the other allele, *EstC67*[1]. Thus, the estimative for the genetic stability of RB92574 seems to also be dependent on the analyzed locus. The genetic stability of the RB92574 variety is supported by the low Nei coefficient (G_{ST} = 0.0135) of genetic divergence determined for samples from PR and MS and their high identity coefficient (I = 0.9968). According to Wright (1978), values of G_{ST} lower than 0.05 indicate a minimum difference between the compared populations.

Table 1. Soil and climate condition in the two areas of collection of the RB92579 variety of *Saccharum* spp. and the management at these two sites.

Information	Mandaguaçu (PR)	Angélica (MS)
Stage of cut[1]	4° (third ratoon)	4° (third ratoon)
Propagation[1]	Colmo	Colmo
Heat treatment[1]	No	No
Production environment[1]	E	C
Average rainfall (mm)[2]	1302	1501
Average temperature (°C)[2]	18.6	23.1
Frost damage[1]	No	No
Harvest[1]	Mechanized	Mechanized
Straw burning[1]	Yes	No

[1]Information forwarded by then responsible staff of each mills [2]Source: CLIMATE.DATA.ORG, 2015.
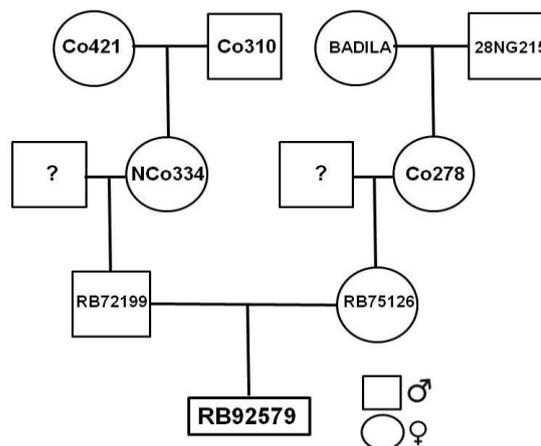

Fig 1. Genealogy of the RB92579 variety of sugarcane. Source: RIDESA.

Table 2. Primers used in the amplification of the genomic DNA of sugarcane, variety RB92579.

Primer	Motif	pb	Sequence
EstB58	$(CCA)_6$	127-205	F: GACGGTCCCGCTCTTGTTCTTGTA
			R: TCCGTGAGCAAAGACAACTGACCT
ESTB149	$(GGA)_6$	105-149	F: CGAGAGGAGCTACCGGTCAGTGTC
			R: CGAGCTCGTTGCGGTTGTGG
ESTC45	$(ATTG)_5$	106-168	F: GCCGCCGTCGCTGGATTG
			R: GATGGATCCCCGCCTACCCTACAC
ESTC69	$(AAAC)_4$	155-162	F: GAACGACGAGCAAGGGAAGGAATG
			R: CAATAAGGCAGGACGGAACAGATG
ESTC80	$(ATTC)_3$	219-277	F: ATTCTTTCTCCCCCTGTTGTGC
			R: GTCGCCAGATCGCTTTCGTT
ESTC84	$(GCCT)_4$	292-371	F: AAGCCGGGTTCCAGTCCAG
			R: GCAACCAAAAGGCTCAGAACAG
ESTC113	$(ACAT)_3$	-	F: CGTCCGCTAGGGCCGATTCCACTCT
			R: GACTAATAGCTGAAGCCGACCAACT
ESTC119	$(AAGC)_4$	84-326	F: GGAATTAAGCTTTGCCGACACCAC
			R: GGCAGCACCTCCCCTTCACC
mSSCIR52	$(GT)_8GC(GT)_6GC(GT)_8$	120-170	F: ACAAGGGAAGACAAATCAG
			R: ACCAAACCACAAAGCAAA
SMC2017Fl	$(CTT)_{13}$	219-258	F: CACAAGTGAAGATAATAGTGTCCCT
			R: GATCCCAAATCCCTTGATCTC
SMC226CG	$(CA)_{10}$	131-168	F: GAGGCTCAGAAGCTGGCAT
			R: ACCCTCTATTTCCGAGTTGGT
SMC477CG	$(CA)_{31}$	90-130	F: CCAACAACGAATTGTGCATGT
			R: CCTGGTTGGCTACCTGTCTTCA

Source: Cordeiro et al. (2000), Singh et al. (2008) and Oliveira et al. (2009).
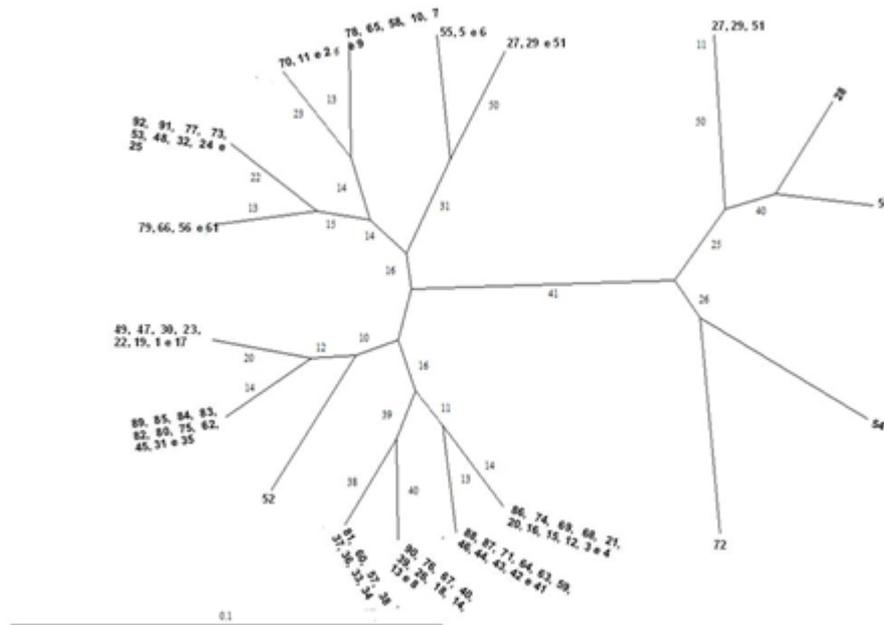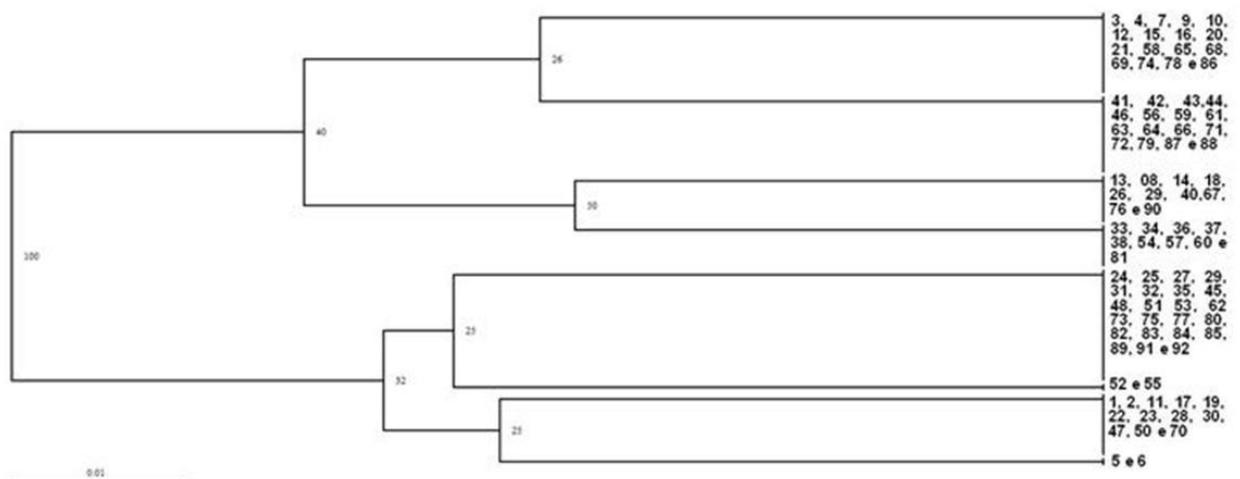
Fig 2. Dendrogram generated by the Jaccard coefficient from analysis of individual plants of sugarcane RB92579 variety obtained from two Brazilian states: Paraná (PR; South region: samples 1-46) and Mato Grosso do Sul (MS; South-Central region: samples 47-92) based on EST-SSR markers and SSR of no-expressed DNA sequences. Three large groups were formed by plants from both PR and MS states. In 66.6% of the branches, the number of identical plants ranged of three to eleven, which reflects the high estimated value of genetic identity among plants from PR (samples 1 to 46) and MS (samples 47 to 92). Numbers beside nodes indicate relative bootstrap frequencies (%).

Table 3. Homologies identified by database SUCEST for the eight EST-SSR loci of the RB92754 variety of sugarcane.
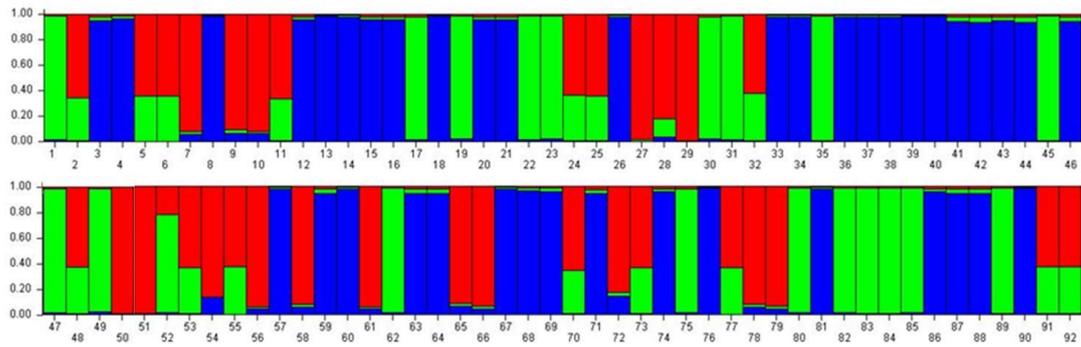
EST-SSR loci	GenBank identification	Identified homology
EstB58	33087065	Putative myb protein
EstB149	9622886	Cellulose synthase-7
EstC45	71067066	SbPCL1
EstC69	9622884	Cellulose synthase-6
EstC80	63087728	Glycosyltransferase
EstC84	47848479	Putative acetyl-CoA C-acyltransferase
EstC113	50726592	Putative thaumatin-like protein
EstC119	50905839	Putative leucine-rich repeat transmembrane protein kinase

Source: Oliveira et al. (2009)

Fig 3. Dendrogram generated by the Jaccard coefficient from analysis of individual plants of sugarcane RB92579 variety obtained from two Brazilian states: Paraná (PR; South region) and Mato Grosso do Sul (MS; South-Central region) based on EST-SSR markers. Numbers beside nodes indicate relative bootstrap frequencies (%).

Fig 4. Dendrogram generated by the Jaccard coefficient from analysis of individual plants of sugarcane RB92579 variety obtained from two Brazilian states: Paraná (PR; South region) and Mato Grosso do Sul (MS; South-Central region) based on SSR of no-expressed DNA sequences. Numbers beside nodes indicate relative bootstrap frequencies (%).

Fig 5. A three-dimensional projection created for plants of the sugarcane RB92579 variety obtained from two Brazilian states: Paraná (PR; South region: samples 1-46) and Mato Grosso do Sul (MS; South-Central region: samples 46-92).

734

Fig 6. Bar plot based on simple sequence repeats and simple sequence repeat-expressed tag sequence markers for 92 samples of the RB92579 variety of sugarcane within the K clusters. Each plant is represented by a single vertical bar that is broken into three colored segments (red, green, and blue). Each color represents the proportion of ancestral alleles in the genome for each individual, which is represented by a vertical bar.

The dendrograms constructed from EST-SSRs and SSRs of DNA sequences that are not expressed, as well as from 8 EST-SSR loci identified four and two large groups formed by plants from the PR and MS states, respectively, illustrating the high degree of identity between plants from the two states. This high degree of identity is also evident in the two dendrogram groups constructed from 4 noncoding SSRs. The fact that more clusters were formed by EST-SSRs indicates that more genetic differences among the sugarcane plants of RB92579 variety can be detected using both molecular markers types simultaneously.

The clustering of the 92 plants according with Bayesian statistics, grouped into 3 clusters, was consistent with the evidence of plants with very little mixture of alleles from groups 2 and 3, and of plants with alleles predominantly of one of the groups. The bar plot also shows that the processes for obtaining the RB92579 variety selected alleles from only three ancestral groups for the 12 SSR loci, despite the genealogy described for the RB92579 released in 2003 (15 years ago) that shows the variety was formed by at least 10 different genotypes.

The hypothesis that a differential estimate for the genetic diversity within the variety RB92579 may be due to different SSR loci (noncoding SSR and EST-SSR loci) was supported in the present study. Polymorphisms at the *EstB58*, *EstB149*, *EstC45*, *EstC69*, *EstC80*, *EstC84*, *EstC113*, *EstC119*, *Smc226CG*, *Smc477CG*, *Smc2017FL*, and *mSscir52* loci from the RB92579 variety cultivated in the PR and MS states as well as the genetic divergence were low. However, the high heterozygosity detected in both populations of the PR and MS indicates that the RB92579 variety can be recommended to be used as a parent to generate new cultivars.

Despite its remarkable features and excellent agricultural productivity, the RB92579 variety seems not to have been used as a direct parent in the formation of 16 new varieties (RB clones) of sugarcane more recently launched by RIDESA (Oliveira et al., 2015). The 16 new varieties, combined with others previously released by RIDESA, total 94 improved cultivars in 45 years of research wherein the RB92579 variety was not used directly a parent. The microsatellite markers used in the present study that allow a molecular diagnosis of the RB92579 variety support their prospective use in sugarcane improvement.

On the other hand, the low Nei coefficient of genetic divergence and high identity coefficient found for samples from PR and MS states indicates the absence of genetic divergence at the *EstB58*, *EstB145*, *EstC45*, *EstC69*, *EstC80*, *EstC84*, *EstC113*, *EstC119*, *Smc226CG*, *Smc477CG*, *Smc2017FL*, and *mSscir52* loci of the RB92579 variety grown in the South and South-Midwest of Brazil. Structure-based grouping and PCA did not separate genotypes according to their geographical location. These findings provide important information for the industrial sector. The samples of RB92579 grown in PR and MS are genetically uniform at the 12 microsatellites loci analyzed and may be mixed in industrial processes, particularly those involving the activity of enzymes encoded by the EST-SSR loci. The low genetic divergence in SSR at the cellulose synthase locus (*EstC69* and *EstB149*, e.g.) indicates there is no need for differential industrial adaptations related to pretreatment and enzymatic hydrolysis of sugarcane bagasse from RB92579 at the same cutting stage and planted in either PR or MS regions.

Materials and Methods

RB92579 variety of sugarcane

The 92 samples of RB92579 were collected from plants in the fourth cutting stage growing in two Brazilian states: Paraná (PR; South region) and Mato Grosso do Sul (MS; South-Central region). The sugarcane variety RB92579 has excellent agricultural productivity and it is important to know if it has the potential to be used as parents in sugarcane improvement. Forty-six samples were collected in the municipal district of Mandaguaçu (PR: Latitude 23°19'12.58"S and Longitude 52°8'24.30"W) and 46 samples were collected in Angélica (MS: Latitude 21°59'14.63"S and Longitude 53°54'24.95"W). Some contrasting characteristics of soil and climate condition as well as the management in the two areas of collection of the RB92579 variety are observed in the two states. The genealogy of the RB92579 variety developed by the RIDESA is shown in Figure 1, and the different aspects of cultivation, differences in the edafoclimatic conditions and management strategies conducted at the experimental stations (mills) in the two regions are shown in Table 1. Samples were collected randomly from different plants (100 plants from each location) according the Fukuda and Otsubo (2003) model, making a zigzag route to cover a homogeneous area of the

735

field. The youngest leaves of each clump were selected (with fewer fibers to facilitate DNA extraction). The samples were packed in aluminum foil and stored on ice, then transferred to the laboratory, where they were kept at -80°C until the DNA extraction process.

Isolation and amplification of DNA

DNA was isolated according to the protocol described by Aljanabi et al. (1999), with the modification of increasing the NaCl concentration to 5 mol·L^{-1}. The youngest leaves (250 mg) were ground in liquid nitrogen, transferred to a 2 mL microtube and homogenized in 300 μL extraction buffer (Tris-HCl 200 mmol·L^{-1} and 50 mmol·L^{-1} EDTA, pH 8.0) and 100 μL of NaCl 5 mol L^{-1}, CTAB 2%, sodium sulfite 0.06%, N-lauryl sarcosine 5% and PVP-40 10%. The mixture was incubated for 60 min at 65°C, and after this incubation, the procedures were followed according to protocol of Aljanabi et al. (1999).

Four primers for DNA simple sequence repeats (SSR: mSSCIR52, SMC2017FL, SMC226CG, and SMC477CG) and eight primers for expressed sequence tag simple sequence repeats (EST-SSR: ESTB58, *EstB149*, ESTC45, ESTC69, ESTC80, ESTC84, ESTC113, and ESTC119) were used for DNA amplification (Table 2). The four SSR primers previously mapped to sugarcane by the International Sugarcane Microsatellite Consortium (Cordeiro et al., 2000; Singh et al., 2008; Oliveira et al., 2009) were synthesized by Invitrogen Technologies Corporation (USA) and used to amplify DNA samples. The EST-SSR sequences were based on libraries of expressed sequence (EST-SSR) developed by Oliveira et al. (2009). Polymerase chain reaction (PCR) was performed using a Techne TC-512 thermal cycler. The amplifications were performed using a Touchdown (Td) PCR program (Don et al., 1991).

For Td-PCR, 20 μL volumes containing 13.1 μL of Milli-Q water (Millipore Corporation), 1.5 μL of genomic DNA (10 ng·μL^{-1}), 0.4 μL of each primer (forward and reverse primers; 10 μM) (Invitrogen, Life Technologies Corporation), 0.8 μL of each dNTP (dATP, dGTP, dCTP, and dTTP; 0.1 mM), 1.6 μL of MgCl$_2$ (2 mM), 0.2 μL (1 U) of Platinum® Taq DNA polymerase (Invitrogen) and 2.0 μL 10X reaction buffer (Invitrogen) were used.

The PCR conditions were as follows: initial denaturation at 94°C for 1 min; 10 cycles of 1 min at 94°C, 1 min with an initial temperature of 65°C and a reduction of 1°C per cycle, and 2 min at 72°C; followed by 20 cycles of 1 min at 94°C, 1 min at 55°C, and 2 min at 72°C. The final extension was 5 min at 72°C. The PCR program using specific temperatures for primer annealing was 94°C for 5 min followed by 30 cycles of 94°C for 1 min and the TA of each primer for 1 min, and later cycles with 72°C for 1 min. The final extension was 15 min at 72°C.

After amplification, 4 μL loading buffer (0.25% bromophenol blue and 30% glycerol) was added to every amplification product and 20 μL of each sample was separated by electrophoresis on a 4% agarose gel (50% agarose and 50% agarose Metaphor™, CAMBREX) containing 0.5× TBE buffer (0.5× TBE buffer (44.5 mM Tris-borate and 1 mM EDTA). All 92 samples amplified by a single SSR primer were run on the same gel at 60 V for 4 h. A 1-kb ladder (Invitrogen) was used as the weight molecular marker. Gels were stained using 0.5 μg·mL^{-1} ethidium bromide, and images were captured using Molecular Image Loccus L-PIX-HE and the Picasa 3 program.

Microsatellite analysis

Homozygous and heterozygous phenotypes for the different alleles were scored within each SSR (*mSscir52, Smc2017FL, Smc226CG*, and *Smc477CG*) and EST-SSR (*EstB58, EstB145, EstC45, EstC69, EstC80, EstC84, EstC113*, and *EstC119*) loci in 92 samples of RB92579. Polyploid plants may to show heterozygous phenotypes, indicated by more than two bands that are products of three or more alleles in loci located on different chromosomes in the polyploid genome. Thus, is possible to detect the proportion of observed homozygous plants as well as the proportion of heterozygous plants containing two or more alleles, so that it is possible to estimate the mean observed heterozygosity for each locus.

For analysis of the polymorphisms and genetic structure of RB92579, 92 plants were scored for the presence or absence of SSR and EST-SSR markers (a score of 1 was assigned for presence and 0 for absence of the homologous band), and the data were fed into a binary data matrix as discrete variables. Polymorphisms from SSR (*mSscir52, Smc2017FL, Smc226CG*, and *Smc477CG*) and EST-SSR (*EstB58, EstB145, EstC45, EstC69, EstC80, EstC84, EstC113*, and *EstC119*) loci were analyzed as dominant markers [(1) presence and (0) absence of amplified DNA segments] because sugarcane plants are polyploids. Using PoPGENE 1.32 software (Yeh et al., 1999), polymorphisms from SSR and EST-SSR loci were analyzed in the two sugarcane populations. The genetic divergence represented by Nei´s (1973) genetic differentiation was also estimated for the PR and MS sugarcane populations. FreeTree software (Pavlícek et al., 1999) was used to perform bootstrap analyses to compare the plants from PR and MS states. The distance similarity matrix was computed with UPGMA (Sneath and Sokal, 1973), followed by Jaccard´s clustering method, with resampling analysis using 1000 replications. A dendrogram was constructed based on a reference tree using the TreeView program (Page, 2001). Principal coordinate analysis (PCA) was also performed as an alternative means of determining and visualizing the structure of the two sugarcane fields (PR and MS). Axes of genetic variation were inferred by Subroutine Simint, Eigen and 3D plot procedures of the NTSYS-pc software package (Rohlf, 1998). Analysis of molecular variance (AMOVA, GenAlEx 6.2; Peakall and Smouse, 2006) was performed to explore the hierarchical partitioning of genetic variation within and between the sugarcane populations from two states.

Polymorphisms of the SSR and EST-SSR markers were also analyzed using STRUCTURE software 2.0 (Pritchard and Wen, 2003), which evaluated the level of genetic admixture between the samples of RB92579 variety. The genotypes were clustered, with the number of clusters (K) ranging from 2 to 5 and tested using the admixture model with a burn-in period of 10,000 iterations followed by 100,000 Markov Chain Monte Carlo (MCMC) iterations, considering the presence or absence of SSR and EST-SSR markers (bands) across the samples. The true number of populations (K) is often identified using the maximal value of Δ (K) returned by the software. The most likely number (K) of subpopulations was identified as described by Evanno et al. (2005). The graphical output display of the STRUCTURE results was taken as input data for STRUCTURE HARVESTER, web-based software

for visualizing the STRUCTURE output and implementing the Evanno method (Earl and von Holdt, 2012) to obtain a graphical representation.

Conclusion

The RB92579 sugarcane variety grown in PR and MS are genetically uniform at the 12 microsatellites loci analyzed and may be mixed in industrial processes, particularly those involving the activity of enzymes encoded by the EST-SSR loci. On the other hand, the high heterozygosity detected in both populations of the PR and MS indicates that the RB92579 variety can be recommended to be used as a parent to generate new cultivars.

Acknowledgements

References

Aljanabi SMM, Forget L, Dookun A (1999) An improved and rapid protocol for the isolation of polysaccharide- and polyphenol-free sugarcane DNA. Plant Mol Biol Rep. 17: 1-8.

Augusto R, Maranho RC, Mangolin CA, Machado MFPS (2015) High polymorphism in *Est-SSR* loci for cellulose synthase and b-amylase of sugarcane varieties (*Saccharum* spp.) used by the industrial sector for ethanol production. Appl Biochem Biotech. 175: 965-973.

Chandra A, Grisham MP, Pan Y-B (2014) Allelic divergence and cultivar-specific SSR alleles revealed by capillary electrophoresis using fluorescence-labeled SSR markers in sugarcane. Genome 57: 363-372.

Chedgy RJ, Köllner TG, Constabel CP (2015) Functional characterization of two acyltransferases from *Populus trichocarpa* capable of synthesizing benzyl benzoate and salicyl benzoate, potential intermediates in salicinoid phenolic glycoside biosynthesis. Phytochemistry 113: 149-159.

Chen PH, Pan Y-B, Chen R-K, Xu L-P, Chen Y-Q (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: 347-354.

CONAB (2018) Companhia Nacional de Abastecimento - Acompanhamento da Safra Brasileira de Cana-de-Açúcar – Terceiro Levantamento – Safra 2018/19, Brasilia. https://www.conab.gov.br/info-agro/safras/cana

Cordeiro GM, Henry RJ (2001) Sugarcane microsatellites; their use and characteristics. The Plant and Animal Genome IX Conference, San Diego.

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

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.

Diévart A, Gilbert N, Droc G, Attard A, Gourgues M, Guiderdoni E, Périn C (2011) Leucine-rich repeat receptor kinases are sporadically distributed in eukaryotic genomes. BMC Evol Biol. 11:367-376.

Don RH, Cox PT, Wainwright BJ, Baker K, Mattick JS (1991) 'Touchdown' PCR to circumvent spurious priming during gene amplification. Nucl Acids Res. 19: 4008.

Duarte-Filho LSC, Silva PP, Santos JM, Barbosa GVS, Ramalho-Neto CE, Soares L, Andrade JCF, Almeida C (2010) Genetic similarity among genotypes of sugarcane estimated by SSR and coefficient of parentage. Sugar Tech 12: 145-149.

Dyer JH, Maina A, Gomez ID, Cadet M, Oeljeklaus S, Schiedel AC (2009) Cloning, expression and purification of an acetoacetyl CoA thiolase from sunflower cotyledon. Int J Biol Sci. 5:736-744.

Earl DA, Von Holdt BM (2012) STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. Conserv Genet Res. 4: 359-361.

Evanno G, Regnalt S, Goudet J (2005) Detecting the number of clusters of individuals using the software structure. A simulation study. Mol Ecol. 14:2611-2620.

Fukuda C, Otsubo AA (2003) Cultivo da mandioca na região centro sul do Brasil. In: EMBRAPA. Sistemas de produção. Brasília, DF.

Hameed U, Pan Y-B, Muhammad K, Afghan S, Iqbal J (2012) Use of simple sequence repeat markers for DNA fingerprinting and diversity analysis of sugarcane (*Saccharum* spp.) cultivars resistant and susceptible to red rot. Genet Mol Res. 11: 1195-1204.

Maranho GB, Maranho RC, Desordi R, Neves AF, Mangolin CA, Machado MFPS (2016) Genetic divergence and admixture of ancestral genome groups in the sugarcane variety 'RB867515' (*Saccharum* spp.) Genet Mol Res. 15: gmr15049209. http://dx.doi.org/10.4238/gmr15049209

Maranho RC, Augusto R, Mangolin CA, Machado MFPS (2014) Use of differential levels of mean observed heterozygosity in microsatellite loci of commercial varieties of sugarcane (*Saccharum* spp.) Genet Mol Res. 13: 10130-10141.

Nei M (1972) Identity of genes by descent within and between populations under mutation and migration pressures. Theor Pop Biol. 3: 460-465.

Nei M (1973) Analysis of gene diversity in subdivided populations. Proc Nat Acad Sci USA. 70: 3321-3323.

Oliveira KM, Pinto LR, Marconi TG, Mollinari M, Ulian EC, Chabregas SM, Falco MC, Burnquist W, Garcia AA, Souza AP (2009) Characterization of new polymorphic functional markers for sugarcane. Genome 52: 191-209.

Oliveira RA, Daros E, Hoffmann HP (2015) Liberação Nacional de Variedades RB de Cana-de-açúcar. Editora Graciosa, Curitiba.

Page RDM (2001) TreeView (Win 32). Version 1.6.6. http://taxonomy.zoology.gla.ac.uk/rod/rod.html.

Pan Y-B, Cordeiro GM, Richard EP, Henry RJ (2003) Molecular genotyping of sugarcane clones with microsatellite DNA markers. Maydica 10:319–329.

Parthiban S, Govindaraj P, Senthilkumar S (2018) Comparison of relative efficiency of genomic SSR and EST-SSR markers in estimating genetic diversity in sugarcane. 3 Biotech 8:144-156.

Pavlícek A, Hrdá S, Flegr J (1999) Free-Tree–freeware program for construction of phylogenetic trees on the basis of distance data and bootstrap/jackknife analysis of the tree robustness. Application in the RAPD analysis of genus *Frenkelia*. Folia Biol. 45: 97–99.

Peakall R, Smouse PE (2006) GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. Mol Ecol Notes 6: 288-295.

Pinto LR, Oliveira KM, Ulian EC, Garcia AA, Souza AP (2004) Survey in the sugarcane expressed sequence tag database (SUCEST) for simple sequence repeats. Genome 47:795-804.

Pritchard JK, Wen W (2003) Documentation for Structure software: Version 2. Available from http://pritch.bsd.uchicago.edu.

Rohlf FJ (1998) NTSYSpc: Numerical taxonomy and multivariate analysis system. Version 2.02. Exeter Software, Setauket, New York.

Singh RK, Srivastava S, Singh SP, Sharma ML, Mohopatra T, Singh NK, Singh SB (2008) Identification of new microsatellites DNA markers for sugar and related traits in sugarcane. Sugar Tech 10: 327-333.

Siraree A, Banerjee N, Kumar S, Khan MS, Singh PK, Kumar S, Sharma S, Singh RK, Singh J (2018) Agro-morphological description, genetic diversity and population structure of sugarcane varieties from sub-tropical India. 3 Biotech 8:469-480.

Sneath PH, Sokal RR (1973) Numerical taxonomy. W.H. Freeman and Company, San Francisco.

Tonielo AE, Paixão ACS, Vanzella CA, Ortolan MCA, Sicchieri MS, Bisson O (2018) Censo varietal IAC no Brasil – Safra 2016/17 e na região Centro-Sul – Safra 2017/18. Rev Canavieiros 140: 40-57.

Wright S (1978) Variability within and among natural populations. Chicago: University of Chicago Press. 580 pp.

Yeh FC, Yang R, Boyle T (1999) POPGENE version 1.32: Microsoft Window-based freeware for population genetics analysis. Edmonton: University of Alberta, Center for International Forestry Research.

You Q, Xu L, Zheng Y, Que Y (2013) Genetic diversity analysis of sugarcane parents in Chinese breeding programs using gSSR markers. The Scient World J. Article ID 613062.