



Electronic Journal of Plant Breeding

Research Article

Mining and characterization of genic-microsatellites associated with sucrose synthesis and their utility in population structure analysis in sugarcane

R. Vigneshwari* and R. M. Shanthi

ICAR-Sugarcane Breeding Institute, Coimbatore-641007, Tamil Nadu, India.

*E-Mail: vignesh182019@gmail.com

Abstract

Sucrose content is a highly valued trait in sugarcane breeding that is regulated by several genes. Tracking of multiple genes involved in the expression of a complex trait in a highly heterozygous crop like sugarcane during the breeding programme is a huge challenge. Trait tagging through expressed sequence tags (ESTs) derived molecular markers has been considered as powerful tool because they are encoded by functionally relevant elements of the genome and directly linked with the target traits. Functional markers related to sucrose content are very limited in sugarcane, which hindered the genetic improvement of this trait. Seventy-three sucrose rich and fifteen low sucrose clones chosen from twelve full-sib families were used for the study on the development and characterization of genic microsatellites associated with sucrose synthesis. A total of 186 alleles were obtained, with an average of 6.20 alleles per microsatellite marker. Polymorphism Information Content (PIC) values were ranged from 0.31 to 0.83, with a mean value of 0.61. The average values of gene diversity (GD) and resolving power (RP) were 0.60 and 5.14, respectively. The present study identified genic SSR markers for four major enzymes viz., CA076844₁₈₇ (FK), CA155160₂₉₅ (SREBF), CA163325_{191,213} (SREBF), TA38405-4547₂₃₂ (PPFK), and CA135596_{178,225} (DFPP) which could be useful to identify sucrose rich clones in the initial testing stages of the selection programme in sugarcane. Analysis of molecular variance (AMOVA) showed 74 per cent of the total variation within population and 26 per cent among population. The EST-SSR primers showed a high allelic polymorphism, as demonstrated by high genetic diversity estimates ($I = 1.17$, $H_o = 0.78$). The low mean values of fixation index ($F_{ST} = 0.078$) and the high estimates of gene flow ($N_m = 4.166$) indicates the low level of population differentiation and maximum genetic exchange between two populations. Based on SSR marker data, eighty-eight clonal accessions were classified broadly into high and low sugar population which corresponded to the molecular diversity of the sucrose specific alleles through structure analysis and principal coordinate analysis.

Key words

Sugarcane, sucrose content, genic-microsatellites, AMOVA, population structure

INTRODUCTION

Sucrose is the major end product of photosynthesis in source tissues and an important component of yield in sugarcane. Sucrose metabolic pathway enzymes, such as invertases, sucrose phosphate synthase (SPS) and sucrose synthase (SS), hexokinase (HK), and phosphofructokinase (PFK) are differentially expressed in leaves and stalks among different genotypes, developmental stages, and cellular locations (Zhu *et al.*, 1997). Improving

sucrose yield in sugarcane can be accomplished by down/up regulation of the enzymes that are associated with sucrose biosynthesis (Chandra *et al.*, 2012). However, QTL mapping and marker assisted breeding in sugarcane have been challenging because of their complex interspecific aneuployploidy genome that comprises of 100 to 130 chromosomes (Hoarau *et al.*, 2006). Identifying functional markers closely associated with sucrose content are well

suited for this task since they derived from gene-rich areas of genome. Microsatellite markers are considered unique and efficient compared to other marker systems because of their high degree of polymorphism, multi-allelic character, co-dominant inheritance and less expensive with the availability of public databases (Agarwal *et al.*, 2008).

Polymorphism in the sugar specific alleles of first clonal hybrids is likely to provide vital information for sugarcane breeders to choose best parents to cross and select better progenies to evaluate in developing new cultivars. Leena Lavanya and Hemaprabha (2012) adopted a MAS approach using a mapping population involving the high sucrose parent (CoC 671) and a low sucrose parent (BC 6) to identify STMS markers associated with sucrose content. Marker-trait associations of two STMS markers viz., NKS45₁₈₀ and NKS46₂₂₀ for sucrose content were reported in high and low sucrose commercial clones. In another study by Shanthi *et al.* (2016), SSR markers developed for three major enzymes involved in sucrose metabolism viz., neutral invertase (TA31929-4547_{151,167}), pyrophosphate-dependent phosphofructokinase (TA32256-4547_{466,525}) and cell wall invertase (AF050129₂₂₄) was reported in a set of high sucrose Co canes used as parental breeding stocks in sugarcane breeding programme. In view of the above, the present study was conducted to i) develop novel genic microsatellite markers from sugarcane transcript sequences linked to sucrose synthesis ii) evaluate the polymorphic potential of these SSR markers iii) estimate the utility of these recently developed polymorphic markers through genetic structure analysis.

MATERIALS AND METHODS

Using the FLIP software (Brossard, 1997) available through the Organelle Genome Megasequencing Project (OGMP), microsatellite sequences containing di, tri, and tetra repeat motifs were mined in 157190 DNA sequences deposited in Cotton Microsatellite Database (CMD) (<http://www.cottonssr.org>). The results showed a large number of microsatellite repeats in the mined ESTs, resulting in 16,129 SSR repeat motifs. Tri-nucleotides accounted for 53.00 per cent, followed by di-nucleotides (45.00%) and tetra-nucleotides (2.00%) (Fig. 1a). Open reading frames (ORFs) had the most microsatellite motifs (71.88%), followed by 5'UTRs (15.66%), 3'UTRs (8.24%), and introns (4.22%). There were 187 different SSR motifs found, 76 of them occurring at least four times and remaining repeats present from two to five times. On the basis of distribution frequency, top twenty repeat motifs in the coding region of genes closely linked with sucrose content were identified (Fig. 1b). Primer3 software (Rozen and Skaletsky, 1999) was used to design suitable primers from the flanking sequences, based on the following optimum criteria: primer length of 20 nucleotides (nt) (18-26 nt), melting temperature of 50°C (45-55°C), and G/C content of 50 per cent. EST-SSR details such as repeat motifs, location,

primer sequence with annealing temperature (T_m) and size of alleles amplified are presented in Table 1.

Hybridization was carried out involving nineteen Indian commercial Co cane varieties (sucrose % ranged between 18.00–21.00) and the seedling progenies were raised from the fluff (hybrid sugarcane seeds) collected during 2015 flowering season. After initial screening in ground nursery (120 seedlings/cross) for H.R.brix (%), 195 genotypes performing above cross average were advanced to the first clonal selection stage in 2016-17. These selections were laid out in a randomized block design with two replications along with five checks (Co 85004, CoC 671, Co 86032, Co 99004 and Co 0403) to evaluate juice quality parameters. From this, a representative set of clones numbering 88 for sucrose per cent with different ranges (recording above 19.0% sucrose and more than 85.0% purity were categorized as high sucrose and with less than 17.0 % sucrose and below 80% purity as low sucrose) were selected and served as the mapping population for progeny survey with genic-microsatellites linked with sucrose synthesis. Juice analysis was conducted at 360 days of crop age using Sucrolyser (Anton Paar) and the following parameters viz., brix (%), sucrose (%), purity (%) and CCS (%) were recorded using five millable stalks per genotype. The list of HS and LS clones and their parentage with sucrose (at 360 days of crop age) are presented in Table 2.

The CTAB protocol was used to isolate genomic DNA from shoot apical tissue (Murray and Thompson, 1980). Spectrophotometer (ND-1000) at absorbance 260/280 nm was used to check the concentration and consistency of extracted DNA and diluted to 25 ng/μl. In a total volume of 10 μL containing 50 ng of genomic DNA, 0.33 mM dNTPs, 2.5mM MgCl₂, 0.5U of Taq DNA polymerase and 0.15 μM of primer PCR reactions were performed. PCR amplifications were carried out in a Bio-Rad thermal cycler using the following cycle profile: 95°C for 2 minutes, then 35 cycles of 94°C for 1 minute, primer melting temperature for 40 seconds, then 72°C for 40 seconds, then 72°C for 7 minutes. Silver staining was used to visualise PCR products on an eight per cent polyacrylamide gel electrophoresis. A 100 bp ladder was used to estimate band size (GelStar). Each primer's loci were assigned using polymorphic bands, which were scored as present (1) or absent (0).

SSR amplicons were manually scored and translated into a binary data matrix, with 1 indicating the presence of alleles and 0 indicating the absence of alleles. The PIC depicts the usefulness of loci and their ability to detect genotype variations, while gene diversity (GD) refers to the likelihood that two randomly selected alleles from the population are different, was calculated using Power Marker v3.25 (Liu and Muse, 2005). The resolving power of the marker indicates the potential of SSR markers to discriminate the accessions was computed as

$RP = \sum lb$, where lb explains allele information, which can be calculated by using the formula $lb = 1 - (2 \times (0.5 - P))$. P value is the proportion of the total 88 genotypes containing the allele (Prevost and Wilkinson, 1999). Using GenAlEx v. 6.503 (Peakall and Smouse 2012), allele scores of polymorphic primers were compiled to estimate AMOVA. GenAlEx v. 6.503 was also used to calculate genetic diversity estimates such as the number of different alleles (N_a), the number of effective alleles (N_e), observed (H_o) and expected heterozygosity (H_e) and Shannon's information index (I). The population structure of the genotypes was determined using a Bayesian model-based approach with Structure software ver. 2.2

(Pritchard *et al.*, 2000). Five independent runs were used to evaluate population clusters (K), with K values varying from 1 to 10. Each run was given a 50,000 burn-in-time and run duration of 100,000 Markov Chain Monte Carlo (MCMC) generations. The genotypes were divided into sub-clusters based on their maximum likelihood values (L_{nPD}). The Delta K (ΔK) approach was used to reassess the sub-population using Structure Harvester (<http://taylor0.biology.ucla.edu/structureHarvester/>). The molecular diversity produced by the EST-SSR markers for each genotype was visualised using principal coordinate analysis (PCoA) using PAST software ver. 4.03 (Hammer *et al.*, 2001).

Table 1. Newly developed and characterized EST-SSR markers for major sucrose metabolic pathway enzymes

S. No.	SSR loci	Enzyme	Repeat motif	Location	Forward	Reverse	Ta (°C)	Number of alleles	Size (bp)
1	TA31929-4547	NI1	(GGA) ₆	ORF	AAGTTTGGTGAGACCCAAGG	GCGTGAGCTCTAAGTCGTCC	53	9	139 - 303
2	TA31932-4547	NI2	(GGA) ₅	ORF	AAGTTTGGTGAGACCCAAGG	GAGTACATGCCGTCGAAGC	56	5	133 - 335
3	CA268640-2	SAI 1	(TTC) ₆	ORF	ACATGGTCACGGACATCTGA	ATTTGAGAGGCAAGCAAGGA	53	4	136 - 204
4	CA268640	SAI 2	(TTC) ₁₀	ORF	CCATAATTTTGGGGGCTTTT	CTGGCGTCCTCCACTGTAAT	53	4	136 - 204
5	SA16124	SAI 3	(ATT) ₅	ORF	GTGCTCATCTGCATTGCTGT	CTTGTGCCAATTGTTGTGG	56	5	235 - 425
6	SA16123	SAI4	(ATT) ₆	ORF	GTGCTCCTGCTGCTCCTC	AGCCAACAACCTACCGTTTCG	58	6	220 - 358
7	AF050129	CWI1	(CTT) ₅	ORF	CCGTCTTCTTCAGGGTGTTT	CGTAGAGGTGAGCGTCCTTC	51	4	153 - 224
8	AF050128	CWI2	(CTT) ₆	ORF	CTAAAGGCTGGGCTGGAAT	GTGACCGACTTTGCTCTCAG	53	5	145 - 315
9	TA32256-4547	PPFK 1	(GCC) ₅	ORF	CTGAGTTCTCCAATGGCGTC	GTATCCGTAGTCGCCCTCCT	58	7	296 - 525
10	CA172223	PPFK 2	(AGG) ₆	5'UTR	AGTTCTCCAATGGCATCGTC	GTATCCGTAGTCGCCCTCCT	56	6	150 - 374
11	CA139117	PPFK 3	(GCC) ₆	ORF	CAAGAGAGATGAACGCCGAC	GACTGCCTCCCAGAGAACAC	56	5	295 - 525
12	TA38405-4547	PPFK 4	(CAC) ₅	ORF	CATCTGCCTCCTCTCGTGAT	CGTCTTCCAATACGTACCCG	58	8	137 - 443
13	CA118742	PPFK5	(AGG) ₆	5'UTR	GGCTTCCAGAGAGGAAATCA	ACGTGGTCGAGAACGACAG	53	6	170 - 385
14	CA155160	SREBF1	(GT) ₅	ORF	ATCTCATCTCACACCGCCAT	CCCTTTATCCTGTCGCACTC	59	13	127 - 557
15	CA163325	SREBF2	(GT) ₅	ORF	CGATCGAGCCAATCTCATCT	CCCTTTATCCTGTCGCACTC	53	11	115 - 412
16	CA082782	SREBF3	(GT) ₅	ORF	GTTCCAGTCGAGCCAATCTC	CCCTTTATCCTGTCGCACTC	56	6	165 - 225
17	CA246373	FK1	(CT) ₅	ORF	CCTCTCCTCTAGTCTCCCC	CGAAACTCACCACCAGCTCT	56	11	127 - 593
18	CA076844	FK 2	(CT) ₅	5'UTR	GCCTCCTCTCTCTAGTCC	AAACTCACCACCAGCTCTCG	56	5	126 - 296
19	TA32931-4547	PFK	(AT) ₅	3'UTR	AACTGCAATGGAGCAGTGTG	CTCTACAACGACAGCACCCA	60	4	127 - 187
20	CA201178	SPS 1	(CG) ₅	ORF	CCTCCCCAAGGAGTCCC	CAGCTGCTCCAGCTTGTTT	50	6	135 - 499
21	SP22298	SPS 2	(GCG) ₅	ORF	AAGGTGATGTCCGTGTGTG	TTACGCACCAAATCCTCTCC	56	5	130 - 425
22	SS22299	FBP1	(GCC) ₆	ORF	AAAAACTGGGATGGGCCTAC	GATCTCCTCCACGTCATCGT	60	6	258 - 330
23	CA069535	FBP2	(GCC) ₆	ORF	GTGCCACCAGCAGCAAT	TCTCGTAGCTGCTCGACTTC	58	6	122 - 246
24	CA135596	DFPP 1	(CGA) ₆	ORF	ACGAGAACCACAACCACCAT	GGAGGTAGTCGGTGAAGTGC	60	10	117 - 349
25	CA096889	DFPP 2	(CGA) ₆	ORF	AGAACCACAACCACCAGGAG	CCCTCGGAGATCTTGACGTA	56	6	145 - 259
26	CA173915	DFPP 3	(GA) ₅	3'UTR	ACTAGCGCTGCTAAACCTCG	GGAGGTTGCGGTACGAGTC	58	4	155 - 268
27	SS22297	SS1	(ATT) ₅	ORF	ATGGTATTCTCCGCAAGTGG	CCTGCGATTCTTGGAATGT	51	5	257 - 582
28	SS22296	SS2	(GTT) ₅	ORF	CCTGTCCATCTACTTCCCGTA	GATTGGCTTGTTCTGTCTGT	56	6	213 - 356
29	SS02400	SS3	(ATT) ₅	3'UTR	CACGACGATGATGTTGAATGACC	GAGAGCAAGTGGAGTGTGTC	58	4	157 - 320
30	TA23722-4547	SS4	(GTT) ₅	5'UTR	GCGTTCGTTGCTACTCTTCC	ACCAACATCCCATCCAGAAA	56	4	225 - 442

6-phospho-1-fructokinase (PFK), fructokinase-1 (FK), pyrophosphate-dependent phosphofructo-1-kinase (PPFK), alkaline/neutral invertase (NI), Soluble acid invertase (SAI), D-fructose-6-phosphate 1-phosphotransferase (DFPP), Sucrose-phosphate synthase (SPS), sucrose-responsive element binding factor (SREBF), Sucrose synthase (SS)

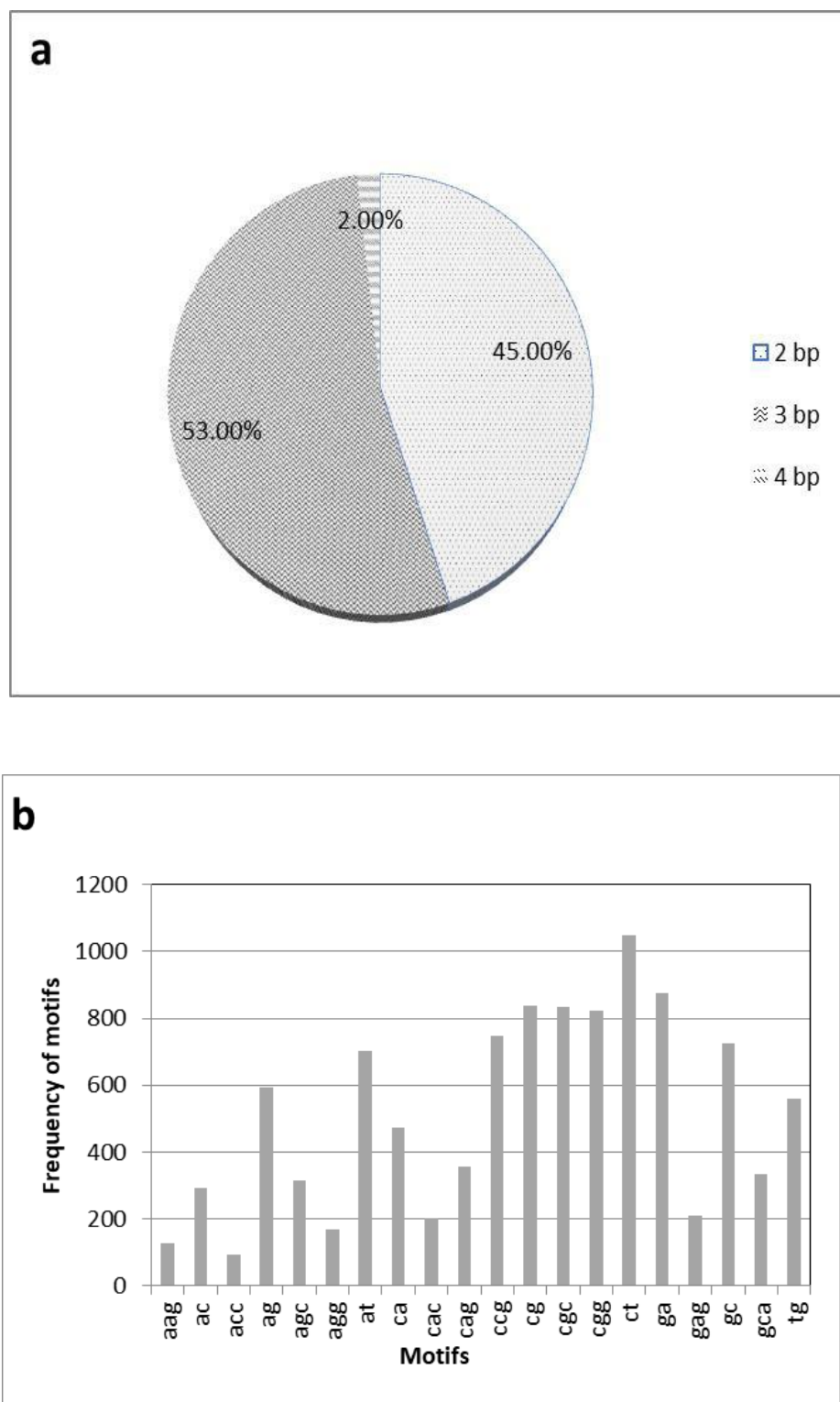


Fig. 1. Frequency of repeat motifs and distribution of EST derived microsatellites related to sucrose metabolic pathway enzymes. a) Frequency of major SSR motifs b) Top twenty SSR repeat motifs in sucrose related ESTs

Table 2. Details of sugarcane clones involved in polymorphism survey of EST derived SSRs linked to sucrose content

Clones	Juice sucrose (%)	Parentage	Clones	Juice sucrose (%)	Parentage	Clones	Juice sucrose (%)	Parentage
High sucrose								
C2-71	20.35	Co 85002 x Co 8209	C5-508	20.35	Co 86011 x CoT 8201C11-964	20.41	Co 11015 x Co 775	
C2-83	20.03	Co 85002 x Co 8209	C6-337	20.46	Co 8371 x Co 86011 C11-1002	21.61	Co 11015 x Co 775	
C2-92	20.50	Co 85002 x Co 8209	C6-360	20.83	Co 8371 x Co 86011 C11-1007	20.90	Co 11015 x Co 775	
C2-374	21.16	Co 85002 x Co 8209	C6-708	20.00	Co 8371 x Co 86011 C11-1261	22.83	Co 11015 x Co 775	
C2-381	20.74	Co 85002 x Co 8209	C7-771	20.83	CoC 671 x Co 94019 C11-1268	21.50	Co 11015 x Co 775	
C2-382	21.17	Co 85002 x Co 8209	C8-785	21.76	Co 86002 x CoC 671 C11-1271	22.20	Co 11015 x Co 775	
C2-392	22.17	Co 85002 x Co 8209	C8-792	21.69	Co 86002 x CoC 671 C11-1280	20.75	Co 11015 x Co 775	
C2-396	21.05	Co 85002 x Co 8209	C8-795	23.85	Co 86002 x CoC 671 C11-1284	20.72	Co 11015 x Co 775	
C2-401	20.64	Co 85002 x Co 8209	C8-803	22.81	Co 86002 x CoC 671 C11- 1287	21.75	Co 11015 x Co 775	
C2-404	20.45	Co 85002 x Co 8209	C8-809	23.41	Co 86002 x CoC 671 C11-1297	21.75	Co 11015 x Co 775	
C2-410	20.99	Co 85002 x Co 8209	C8-827	22.15	Co 86002 x CoC 671 C11-1298	21.19	Co 11015 x Co 775	
C2-413	20.95	Co 85002 x Co 8209	C8-839	20.16	Co 86002 x CoC 671 C11-1312	20.25	Co 11015 x Co 775	
C4-193	20.15	Co 86002 x Co 99006	C8-842	21.38	Co 86002 x CoC 671 C11-1318	20.06	Co 11015 x Co 775	
C4-196	20.06	Co 86002 x Co 99006	C8-851	20.80	Co 86002 x CoC 671	Low sucrose		
C4-202	20.15	Co 86002 x Co 99006	C8-856	22.61	Co 86002 x CoC 671 C1-8	16.41	Co 10033 x Co 13007	
C4-214	20.50	Co 86002 x Co 99006	C8-989	21.93	Co 86002 x CoC 671 C1-13	16.00	Co 10033 x Co 13007	
C4-221	20.89	Co 86002 x Co 99006	C8-996	22.83	Co 86002 x CoC 671 C1-36	16.49	Co 10033 x Co 13007	
C4-601	20.10	Co 86002 x Co 99006	C8-1108	20.77	Co 86002 x CoC 671 C1-41	15.00	Co 10033 x Co 13007	
C4-642	20.03	Co 86002 x Co 99006	C8-1136	20.80	Co 86002 x CoC 671 C1-52	16.13	Co 10033 x Co 13007	
C5-241	21.71	Co 86011 x CoT 8201	C8-1137	21.86	Co 86002 x CoC 671 C1-54	15.13	Co 10033 x Co 13007	
C5-244	20.94	Co 86011 x CoT 8201	C8-1138	20.81	Co 86002 x CoC 671 C3-146	16.06	Co 87044 x Co 86011	
C5-253	21.71	Co 86011 x CoT 8201	C8-1140	20.16	Co 86002 x CoC 671 C3-441	16.40	Co 87044 x Co 86011	
C5-261	21.15	Co 86011 x CoT 8201	C8-1142	20.29	Co 86002 x CoC 671 C3-448	15.00	Co 87044 x Co 86011	
C5-268	20.50	Co 86011 x CoT 8201	C10-910	21.87	Co 8371 x Co 94008 C9-893	16.03	CoM 0265 x Co 0314	
C5-274	21.77	Co 86011 x CoT 8201	C10- 918	20.67	Co 8371 x Co 94008 C9-898	15.33	CoM 0265 x Co 0314	
C5-288	20.44	Co 86011 x CoT 8201	C10-949	21.31	Co 8371 x Co 94008 C12-1028	16.00	Co 62198 x Co 8353	
C5-485	21.80	Co 86011 x CoT 8201	C10-915	21.93	Co 8371 x Co 94008 C12-1349	16.23	Co 62198 x Co 8353	
C5-499	21.74	Co 86011 x CoT 8201	C10-941	20.90	Co 8371 x Co 94008 C12-1355	16.03	Co 62198 x Co 8353	
C5-501	20.81	Co 86011 x CoT 8201	C10-944	22.88	Co 8371 x Co 94008 C12-1369	16.17	Co 62198 x Co 8353	
C5-504	20.86	Co 86011 x CoT 8201	C10- 950	21.71	Co 8371 x Co 94008			

C1 to C12: Twelve full-sib families

RESULTS AND DISCUSSION

Improvements in sucrose content were reported in SI (Sucrose Isomerase) transgenic sugarcane clones under greenhouse conditions, but not in field grown conditions (Basnayake *et al.*, 2012). Sucrose accumulation is a multi-enzyme mediated complex trait, therefore, developing sugarcane clones with altered enzyme activity to improve sucrose yield has been more challenging (Moore *et al.*, 1997). The results on SSR polymorphism, informativeness

and its utility in genetic structure analysis corresponding to nine major enzymes of sucrose metabolism are discussed in this paper.

Sucrose is cleaved into UDG-glucose and fructose for further sugar metabolism by sucrose synthase or invertases. Fructokinases (FK) have a high affinity and substrate specificity for phosphorylating free fructose. Higher expression of FK genes reported earlier in matured

culm tissues than in younger culm tissues implies the significance of this enzyme in sucrose accumulation (Chen *et al.*, 2017). They also reported FK gene family (SsFRK1–SsFRK7) with seven genes in *Saccharum* spp. through transcriptome analysis. In this study, two primers CA246373 and CA076844 were designed for the FK1 and FK2 genes. FK1 was found to be highly polymorphic and amplified 11 SSR alleles, of which two alleles of 166 bp and 187 bp size were observed in all high sucrose genotypes otherwise absent in all low sucrose genotypes.

SREBF is a DNA binding protein and functions as repressor in the presence of high sucrose levels. Grierson *et al.* (1994) reported two sucrose responsive elements (SUREs) in the potato patatin promoter region. For the SREB factor, we designed three EST-SSR primers: CA155160, CA163325, and CA082782. Of these, CA155160 and CA163325 had the highest number of amplicons, with 127 bp -557 bp and 115 bp - 412 bp, respectively. CA155160 amplified a 295 bp fragment; while CA163325 amplified two SSR alleles 191 bp and 213 bp in high sugared genotypes but not in low sugared genotypes.

Pyrophosphate-dependent phosphofructokinase (PFP) is one of the key enzymes involved in the phosphorylation of fructose. PFP activity and sucrose levels were found to be negatively correlated in both commercial types and in a segregating F_1 sugarcane population (Whittaker and Botha, 1999). Groenewald and Botha (2001) correlated low level of PFP activity with an increase of sucrose accumulation in transgenic sugarcane clones. Four microsatellite primers for SSR motifs found in the ORF and 5'UTR regions of PFK genes were developed. Using the primer TA38405-4547, a SSR amplicon with a size of 232 bp possibly associated with high sucrose content was established. Similarly, Merwe *et al.* (2010) reported a negative association between DFPP (D-fructose-6-phosphate 1-phosphotransferase) activity and sucrose concentration in immature internodes of transgenic sugarcane. For repeat motifs found in the ORF and 3'UTR regions of D-fructose-6-phosphate 1-phosphotransferase (DFPP), three SSR primers were designed. The primer CA135596, which was designed to detect SSR repeat motifs on the ORF, was found to be successful in discrimination of high and low sucrose genotypes. Two SSR alleles of 178 bp and 225 bp in size were specific to high sucrose genotypes and absent in the low sucrose progenies.

As part of the present study, attempts were made to validate the recently identified markers among thirty high and eleven low sucrose commercial clones in order to enhance their applicability in identifying sucrose rich clones during early phases of selection in a sugarcane breeding programme. The analysis showed that seven transcript markers namely, CA076844¹⁸⁷ (FK), CA155160²⁹⁵ (SREBF), CA163325^{191,213} (SREBF), TA38405-4547²³² (PPFK) and CA135596^{178, 225} (DFPP) was specifically

present in all the high sugared clones, while absent in the low sucrose genotypes. Thus, transcript markers developed in this present study can be used in the screening of a large seedling and clonal population in the early selection stages and also expected to improve marker-trait association of the QTLs linked with sucrose trait. Polymorphism of high and low sucrose commercial genotypes using TA38405- 4547 and CA135596 are presented in Fig. 2.

Compiled allele scores of thirty transcript SSRs detected a total of 186 alleles, of these 135 alleles (72.58%) were polymorphic. The number of amplicons per locus ranged from four to thirteen, with an average of 6.20 alleles per SSR locus. The allele size ranged from 115 bp to 582 bp. From 30 SSR markers screened, 15 markers which exhibited clear and repeatable polymorphic bands were selected for developing sucrose specific markers. Allelic diversity estimates on 15 polymorphic SSR markers are presented in Table 3. The PIC values ranged from 0.31 to 0.83, with an average of 0.61 per marker locus. The average PIC value in this investigation was similar to the previous reports of 0.87 in *Saccharum* hybrids and 0.86 in *S. spontaneum* (Ali *et al.*, 2019). Ten microsatellite markers were found to be highly informative (PIC > 0.5), while five SSR loci moderately informative (0.5 < PIC < 0.75) as described by Botstein *et al.* (1980). The probability that two randomly selected alleles from the population are different is measured by gene diversity (GD). The GD ranged from 0.44 (SS22299) to 0.76 (CA163325), with a mean value of 0.60. The distribution frequency of alleles within genotypes determines the loci's resolving power (RP). The RP ranged from 2.65 (CA172223) to 9.14 (CA163325), with a mean value of 5.14. Kalwade *et al.* (2012) reported a mean value of 4.20 utilizing 17 sugarcane genotypes and 27 ISSR markers, which is similar to the observation made in the present study. The number of alleles and effective alleles were in the range of 4 to 13 and 2.52 (CA268640-2) to 7.05 (CA155160) respectively. Expected heterozygosity (H_e) ranged from 0.42 (CA172223) to 0.85 (CA163325), with an average of 0.66. Further, three SSR loci namely TA38405-4547, CA163325, and CA135596 produced high value of observed heterozygosity (H_o) of 1.00. Shannon's information index (I) ranged from 0.53 (SS22299) to 2.07 (CA155160). An average Shannon's index (I) was calculated as 1.17, which indicated the high polymorphic potential of transcript SSRs in this study.

The average value of F_{ST} was 0.078 and the AMOVA results revealed that it was mostly related to within population variation (74%) rather than variation among population (26%). (Table 4). Gene flow (vN_m) is used to describe the transfer of genetic variation between populations, high value of gene flow indicates the greater genetic exchange and *vice versa*. High gene flow estimate (4.166) obtained in this study did show an evidence for maximum genetic exchange in the studied population. In a previous investigation, which involved 139 accessions of the genus

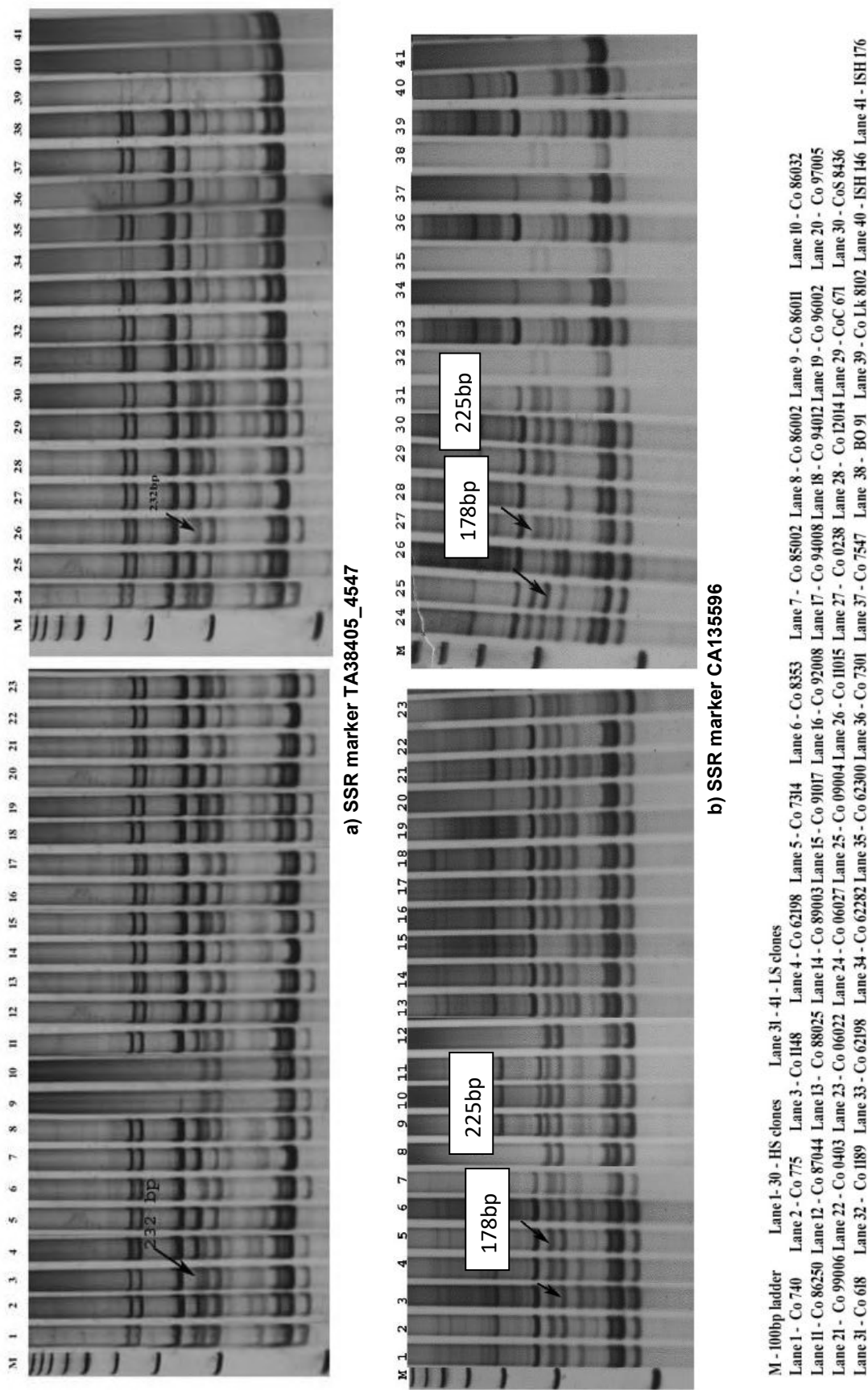


Fig. 2. a) Polymorphism among the high and low sucrose commercial clones/varieties using TA38405-4547 b) Marker pattern of CA135596 SSR loci among the high and low sugar clones

Table 3. Genetic diversity estimates of fifteen polymorphic microsatellite markers

SSR loci	PIC	GD	RP	N _a	N _e	H _o	H _e	I
TA31932-4547	0.61	0.53	3.26	5	3.82	0.90	0.73	1.40
CA268640-2	0.43	0.50	3.19	4	2.52	0.64	0.49	0.66
CA172223	0.47	0.55	2.65	6	4.33	0.68	0.42	0.58
TA38405-4547	0.83	0.74	8.48	7	4.26	1.00	0.75	1.53
CA155160	0.79	0.71	6.94	13	7.05	0.86	0.84	2.07
CA163325	0.81	0.76	9.14	9	6.80	1.00	0.85	2.02
CA082782	0.69	0.63	5.30	6	4.89	0.76	0.62	0.79
CA246373	0.72	0.74	7.85	11	6.23	1.00	0.83	1.92
CA076844	0.69	0.61	4.86	5	3.87	0.82	0.74	1.43
TA32931-4547	0.42	0.51	3.11	4	3.82	0.69	0.53	0.71
CA135596	0.78	0.70	8.69	10	6.71	1.00	0.84	1.99
CA201178	0.51	0.48	2.69	6	4.50	0.59	0.53	0.56
SP22298	0.44	0.49	3.19	5	3.20	0.59	0.47	0.61
SS22299	0.31	0.44	3.15	4	2.56	0.54	0.51	0.53
CA069535	0.68	0.62	4.66	6	3.86	0.71	0.76	0.87
Mean	0.61	0.60	5.14	6.33	4.56	0.78	0.66	1.17

MAF – Major allele frequency, PIC – Polymorphic information content, GD – Gene diversity, RP – Resolving power, N_a – Number of alleles, N_e – Number of effective alleles, H_o – Observed heterozygosity, H_e – Expected heterozygosity, I – Shannon's information index

Table 4. Analysis of molecular variance (AMOVA) for 88 sugarcane clones based on fifteen polymorphic SSR markers

Source of variation	DF	Sum of squares	Mean sum of squares	Estimated variance	Percentage of variation
Between	1	236.345	236.345	3.517	26
Within	86	1006.311	11.701	28.015	74
Total	87	1,242.656		31.532	100
Fixation Index	F _{ST} = 0.078				
Gene Flow	Nm = 4.166				

Saccharum, allied genera of Poaceae, and sugarcane cultivars, 83 per cent of the total variation was attributed to variation within the population while variation among population accounted for the remaining 17 per cent and this observation demonstrated greater exchange of gene pool across regions of origin (Singh *et al.*, 2020).

Compiled allele scores of fifteen polymorphic SSR markers were used for population structure analysis. Bayesian model-based structure analysis, divides clones into various groups which corresponded to allelic heterogeneity in sucrose-specific genes. When we examined the distribution of genotypes from $K = 1$ to $K = 10$ using the model established by Evanno *et al.* (2005), the ΔK value was relatively high when $K = 3$. (Fig.3a). The sub-population I (red group) had 17.04 per cent of membership proportion (15 out of the 88 genotypes) and consisted of the low sucrose genotypes. The sub-

population II (green group) and III (blue group) contained the high sucrose genotypes with membership proportion of 71.59% (63 genotypes out of 88 genotypes) and 11.36 per cent (10 genotypes out of 88 genotypes), respectively. The genotypes with 0.8 per cent or more membership proportion was considered as pure and other genotypes considered as admixture, all three sub-populations I, II and III possessed lower proportion of admixture 13.33, 6.25 and 10.00 per cent (Fig.3b). Population structure analysis showed that the genotypes with high sucrose content belonged to the sub-population II and III, while most of the genotypes with the low sucrose content belonged to the sub-population I because of the extensive similarity in their allelic combination in the each estimated sub-population. In a similar study, 61 polymorphic SSR markers grouped 139 clones of the genus *Saccharum* and allied genera of Poaceae into three distinct sub-populations, but the sub-grouping had no significant association with the geographic origin of the region (Singh *et al.*, 2020).

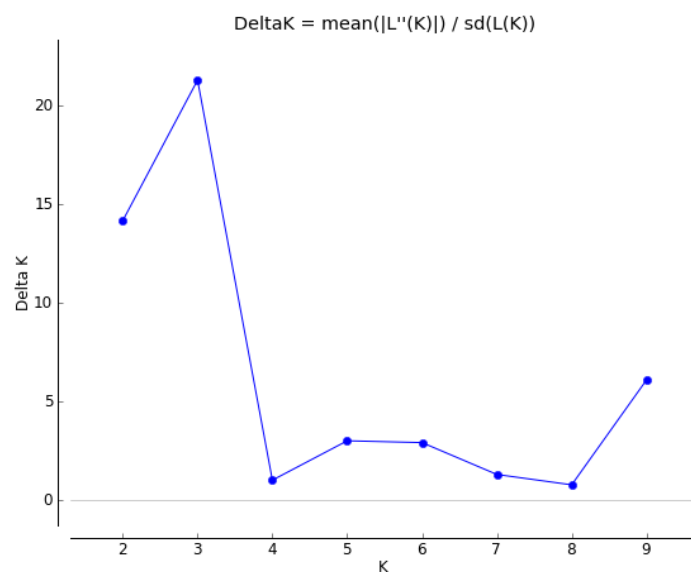


Fig. 3a. The relationship between number of K and ΔK

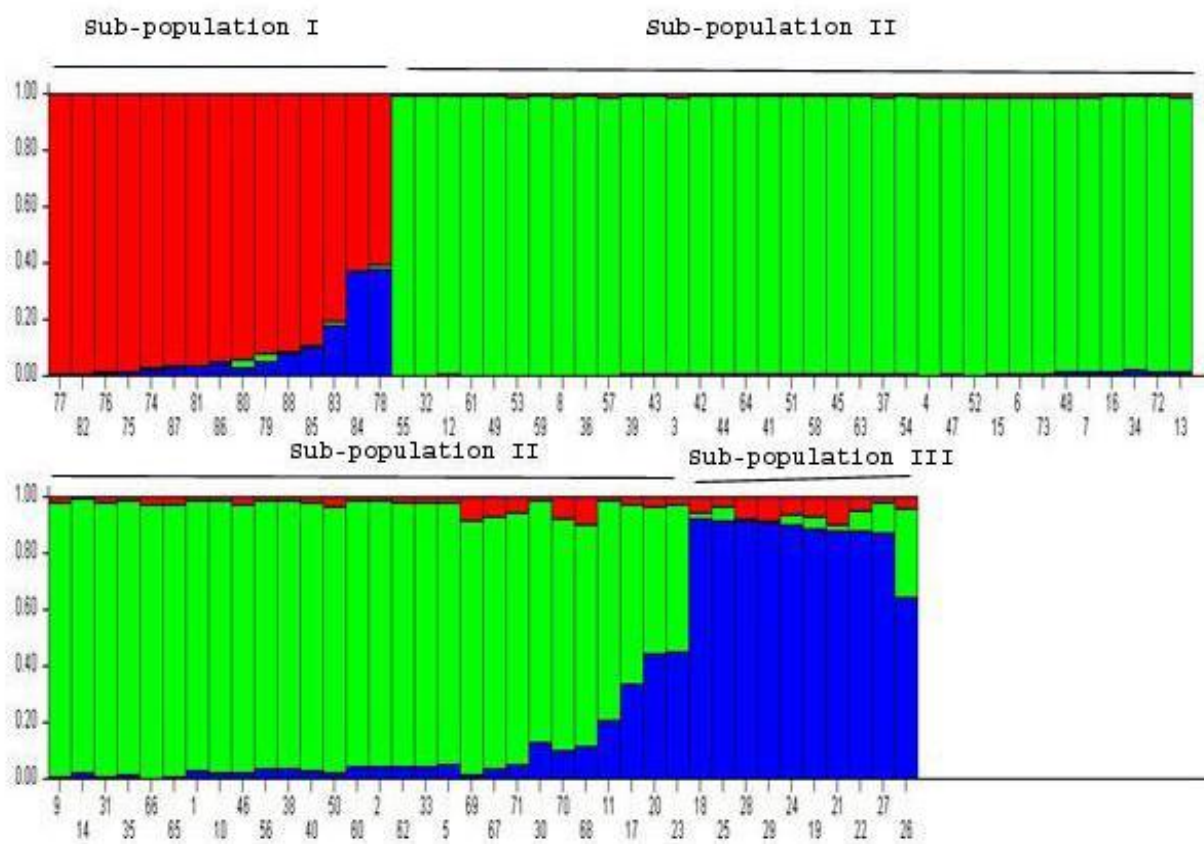


Fig. 3b. Assignment of 88 sugarcane genotypes into three sub-population (I, II and III) using structure. Each bar represents one genotype, the Y axis representing the proportion of each genotype genome assigned into three sub-population and the numbers at the X axis representing clones as listed in Table 2.

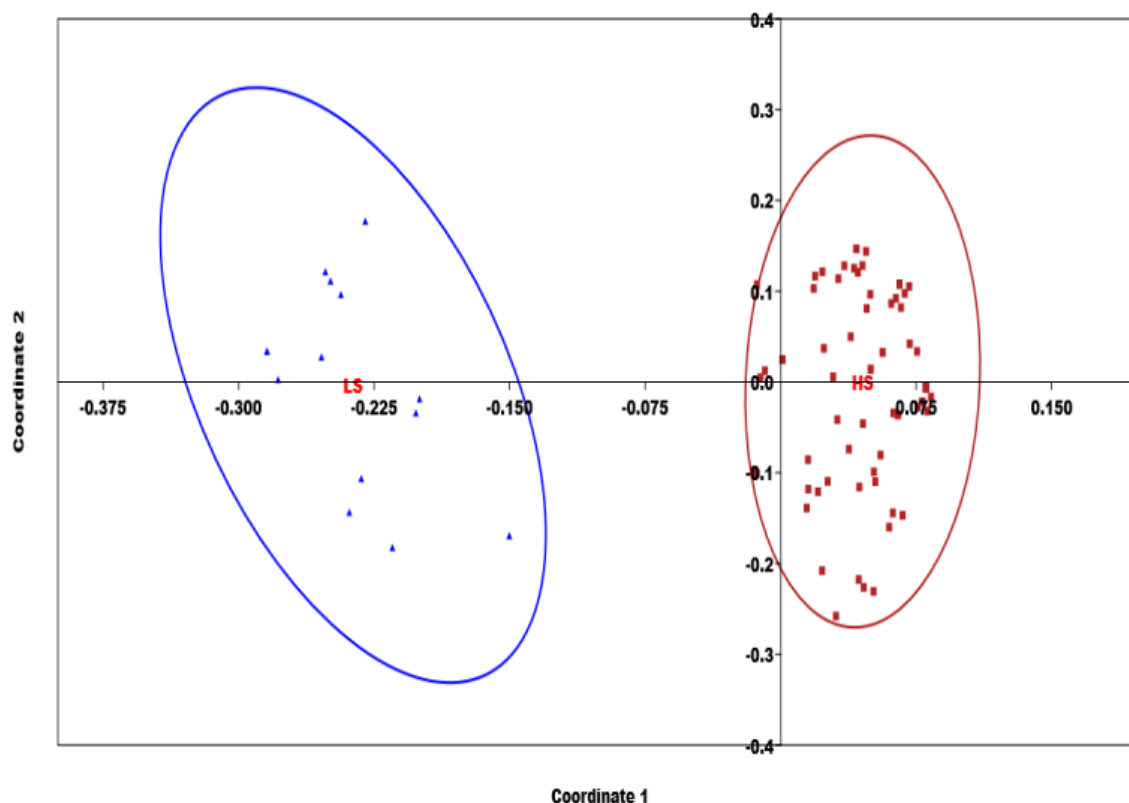


Fig. 4. Principal coordinate analysis (PCoA) based on eigenvalues calculated from 15 polymorphic SSR markers. Red squares represent the high sucrose and blue triangle represents the low sucrose genotypes

The results on principal coordinate analysis (PCoA) are presented in **Fig.4**. A clear separation was observed between the high and low sucrose sugarcane genotypes. The first three axes explained 60.02 per cent of total variation, while the first and second coordinates each explained 48.20 per cent of total variation. The eighty-eight sugarcane genotypes were divided into two groups: the high and low sucrose (HS and LS). This demonstrates that the two populations (HS and LS) are distinct, when it comes to the molecular diversity of genes regulating sucrose metabolism as determined by the study. Eleven distinct groups were reported in an earlier study involving 79 accessions from five species of *Saccharum*, six accessions from *Erianthus arundinaceus* and 30 *Saccharum* spp. hybrids using 21 polymorphic SSR primers (Ali *et al.*, 2019).

The present study identified novel polymorphic SSR markers for four key enzymes associated with sucrose synthesis namely CA076844₁₈₇ (FK), CA155160₂₉₅ (SREBF), CA163325_{191,213} (SREBF), TA38405-4547₂₃₂ (PPFK), and CA135596_{178,225} (DFPP) would be a valuable tool in the marker assisted selection programme (MAS) for genetic improvement of sucrose trait. The transcript SSRs with high polymorphic potential, high value of

effective number of alleles and observed heterozygosity developed in this study will be useful in the germplasm characterization and genetic diversity studies across *Saccharum* spp. The results on PCoA and structure analysis indicated a substantial association between sub-grouping with that of sucrose content, such genetic heterogeneity is directly correlated with allelic diversity in the QTLs influencing sucrose content, a property that could be beneficial in sugarcane breeding programmes.

ACKNOWLEDGEMENT

The authors are grateful to Dr. Bakshi Ram, Director, Sugarcane Breeding Institute for providing all facilities to conduct this experiment. Authors would like to thank Dr. G. Hemaprabha, Head, Division of Crop Improvement for providing necessary facilities during the study period. Thanks are also due to Mr. M. Kannaian in managing the field experiment and assisting with data collection.

REFERENCES

- Agarwal, M, Shrivastava, N. and Padh, H. 2008. Advances in molecular marker techniques and their applications in plant sciences. *Plant Cell Reports.*, vol. 27, no. 4: 617–631. [[Cross Ref](#)]

- Ali, A, Pan, Y.B, Wang, Q.N, Wang, J.D, Chen, J.L. and Gao, S.J. 2019. Genetic diversity and population structure analysis of *Saccharum* and *Erianthus* genera using microsatellite (SSR) markers. *Scientific Reports* **9:395**: 1 - 10. [\[Cross Ref\]](#)
- Basnayake, S.W, Morgan, T.C, Wu, L. and Birch, R.G. 2012. Field performance of transgenic sugarcane expressing isomaltulose synthase. *Plant Biotechnol J.*, **10**: 217 - 225. [\[Cross Ref\]](#)
- Botstein, D, White, R, Skolnick, M. and Davis, R.W. 1980. Construction of a genetic map in man using restriction fragment length polymorphisms. *Am.J.Hum. Genet.*, **32**: 314 – 331.
- Brossard, N. 1997. FLIP: a Unix Program used to find/translate orfs. bionet software. <http://www.bch.umontreal.ca/ogmp/manlinks/flip.txt>.
- Chandra, A, Jain, R. and Solomon, S. 2012. Complexities of invertases controlling sucrose accumulation and retention in sugarcane. *Curr. Sci.*, **102**: 857 – 866.
- Chen, Y, Zhang, Q, Hu, W, Zhang, X, Wang, L, Hua, X, Yu, Q, Ming, R. and Zhang, J. 2017. Evolution and expression of the fructokinase gene family in *Saccharum*. *BMC Genomics.*, **18**, 197. [\[Cross Ref\]](#)
- Evanno, G, Regnaut, S. and Goudet, J. 2005. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol. Ecol.*, **14**: 2611 – 2620. [\[Cross Ref\]](#)
- Grierson, C, Du, J.S, de Torres Zabala M, Beggs, K, Smith, C, Holdsworth, M. and Bevan, M. 1994. Separate cis sequences and trans factors direct metabolic and developmental regulation of a potato tuber storage protein gene. *Plant J.*, **5**: 815 – 826. [\[Cross Ref\]](#)
- Groenewald, J.H. and Botha, F.C. 2001. Down-regulation of pyrophosphate: fructose 6-phosphate 1-phosphotransferase (PFP) activity in sugarcane enhances sucrose accumulation in immature internodes. *Transgenic Res.*, **17**: 85 - 92. [\[Cross Ref\]](#)
- Hammer, Harper, D.A.T. and Ryan, P.D. 2001. PAST: Paleontological statistics software package for education and data analysis. *Palaeontologia Electronica* **4**(1): 9pp. http://palaeo-electronica.org/2001_1/past/issue1_01.htm
- Hoarau, J.Y, Souza, G. and D'Hont, A. 2006. Sugarcane, A Tropical Crop with a Highly Complex Genome, Plant Genomics, Sciences Publishers, Enfield, NH, USA.
- Hoepfner, S. W. and Botha, F.C. 2003. Expression of fructokinase isoforms in the sugarcane culm. *Plant Physiol. Biochem.*, **41**: 741–747. [\[Cross Ref\]](#)
- Kalwade, S.B, Devarumath, R.M, Kavar, P.G. and Sushir, K.V. 2012. Genetic profiling of sugarcane genotypes using inter simple sequence repeat markers. *Electron. J. Plant Breed.*, **3**(1): 621 – 628.
- Leena Lavanya, D. and Hemaprabha, G. 2012. Identification of a STMS marker for sucrose content in sugarcane (*Saccharum* spp.). *Indian J. Genet.*, **72**(3): 343 – 348.
- Liu, K. and Muse, S.V. 2005. Power marker: Integrated analysis environment for genetic marker data. *Bioinformatics.*, **21**(9): 2128 – 2129. [\[Cross Ref\]](#)
- Merwe, M.J, Groenewald, J.H, Stitt, M, Kossmann, J. and Botha, F.C. 2010. Downregulation of pyrophosphate: D-fructose-6-phosphate 1-phosphotransferase activity in sugarcane culms enhances sucrose accumulation due to elevated hexose phosphate levels. *Planta.*, **231**: 595 – 608. [\[Cross Ref\]](#)
- Moore, P.H, Botha, F.C, Furbank, K.R. and Grof, C. 1997. Potential for overcoming physio-biochemical limits to sucrose accumulation. In: (eds. Keating BA, Wilson JR) *Intensive Sugarcane Production: Meeting the Challenges Beyond 2000*. CAB International, Wallingford, UK. pp. 141 – 155.
- Murray, M.G. and Thompson, W.F. 1980. Rapid isolation of high molecular weight plant DNA. *Nucleic Acid Res.*, **8**: 4321 - 4325. [\[Cross Ref\]](#)
- Prevost, A. and Wilkinson, M.A. 1999. A new system of comparing PCR primers applied to ISSR fingerprinting of potato cultivars. *Theor. Appl. Genet.*, **98**: 107 – 112. [\[Cross Ref\]](#)
- Peakall, R. and Smouse, P.E. 2012. GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research - an update. *Bioinformatics.*, **28**: 2537 - 2539. [\[Cross Ref\]](#)
- Pritchard, J.K, Stephens, M. and Donnelly, P. 2000. Inference of population structure using multilocus genotype data. *Genetics.*, **155**: 945 – 959. [\[Cross Ref\]](#)
- Rozen, S. and Skaletsky, H. 1999. Primer3 on the WWW for General Users and Biologist Programmers. *Methods in Molecular Biology, Bioinformatics Methods and Protocols* **132**: 365 - 386. [\[Cross Ref\]](#)
- Shanthi, R. M, Hemaprabha, G, Vigneshwari, R, Sarath Padmanabhan, T.S. and Mohanraj, K. 2016. New genic SSR markers identified for three major enzymes associated with sucrose synthesis in sugarcane. *Journal of Sugarcane Research.*, **6**(1): 1- 10.

- Singh Kushwah R.B, Mahendrakar, M.D, Jugran, A. and Singh, R.K. 2020. Assessing genetic diversity and population structure of sugarcane cultivars, progenitor species and genera using microsatellite (SSR) markers. *Gene.*, Vol. **753**: 144808. [\[Cross Ref\]](#)
- Whittaker, A. and Botha, F.C. 1999. Pyrophosphate: D-fructose-6-phosphate 1-phosphotransferase activity patterns in relation to sucrose storage across sugarcane varieties. *Plant Physiol.*, **107**:379 - 386. [\[Cross Ref\]](#)
- Zhu, Y.J, Komor, E. and Moore, P.H. 1997. Sucrose Accumulation in the Sugarcane Stem Is Regulated by the Difference between the Activities of Soluble Acid Invertase and Sucrose Phosphate Synthase. *Plant Physiol.*, **115**: 609 - 616. [\[Cross Ref\]](#)