Genetic diversity and population structure among 133 elite genotypes of sugarcane (*Saccharum* spp.) for use as parents in sugarcane varietal improvement

T. S. Sarath Padmanabhan¹ · G. Hemaprabha¹

Received: 7 January 2018 / Accepted: 18 July 2018 / Published online: 25 July 2018
© Springer-Verlag GmbH Germany, part of Springer Nature 2018

Abstract

A measure of genetic diversity of genotypes to be used as parents is imperative to use them prudently in crop improvement. In this study, genetic diversity and population structure of 133 sugarcane hybrid derivatives were quantified using 20 sequence-tagged microsatellite sites (STMS) primers. The number of alleles ranged from 9 to 27 with the average of 17.95 alleles per primer, while the polymorphism information content values of the primers ranged from 0.29 to 0.78. Cophenetic correlation coefficient value observed as 0.84 by STMS markers revealed that the cluster result was acceptable for the calculation of genetic similarity matrix. Principal component analysis showed that 133 genotypes fell in two groups, first and second components associated 8.34 and 3.22% with eigen values of 5.61 and 2.17, respectively. Similar trend was observed with principal coordinate analysis, wherein, the first and second component accounted to 8.34 and 3.22% with eigen values of 741.29 and 286.11. The similarity index values ranged from 0.50 to 0.87 for the possible 8778 combinations from 133 genotypes, of which 8069 combinations exhibited less/moderate genetic similarity indicating the availability of sufficient genetic diversity in the experimental material and hence their value in the genetic improvement of sugarcane. Dissimilarity analysis using DARwin of 133 genotypes could distinguish two major clusters and into five subclusters and the results matched with those of the population structure which also showed five subpopulations. The bigger group SP1 was predominantly comprised of clones developed at the main sugarcane-breeding place in India, located at Coimbatore. The subpopulation SP4 was formed largely with clones from research stations other than at Coimbatore and interspecific hybrids, while SP5 comprised of clones of early origin. These observations were similar to the radial tree based on the DARwin software in that 81.95% of the genotypes of each cluster were similar in the two analyses. The results thus showed that location and time of origin were two major factors that contributed to diversity. Based on analysis of molecular variance, subpopulations SP2 and SP4 were more variable from the rest. SP2 (comprising of Co 99008, Co 99006, Co 94012, Co 93023, CoC 671, Co 89034, Co 91003, Co 06022, Co 98017, Co 87044, Co 06018, Co 89003, Co 98014, and Co 86032) exhibited maximum genetic variation, the least gene flow, and the lowest heterozygosity value and would serve as the best group for utilization in genetic improvement. Graphical genotyping (GGT) image of each genotype was distinctly different, indicating the genetic uniqueness of sugarcane genotypes under study as revealed through STMS technology. A core set of 40 genotypes was identified using GGT 2.0 software program for the easiness of harnessing the available genetic diversity of 133 genotypes, through hybridization in sugarcane improvement programs.

Keywords Sugarcane · Sequence-tagged microsatellite markers · Genetic diversity · Population structure · Analysis of molecular variance · Fixation index

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s13205-018-1364-2>) contains supplementary material, which is available to authorized users.

✉ G. Hemaprabha
ghemaprabha1@gmail.com

¹ Division of Crop Improvement, ICAR-Sugarcane Breeding Institute, Coimbatore, Tamil Nadu 641 007, India

Introduction

Sugarcane (*Saccharum* sp.), a commercial crop of great significance caters to about 75% of world sucrose production, is grown in more than 100 countries (<http://www.ilovo.co.za/worldofsugar>). Genetic improvement of sugarcane pioneered with research primarily at Java and India has now become

a major focus in several countries. A precision breeding strategy is yet to be implemented in the crop due to its complex polyploid and aneuploid genetic nature. Breeding is integrated with biotechnology, so that maximum advantage can be taken of new opportunities and that the current level of production is sustained (Hogarth et al. 1997). A readily usable application of molecular approach would be in estimating the genetic diversity of parental lines to harness the best of diversity existing in the gene pool. Molecular markers were used in quantifying the genetic diversity available for crop improvement program including basic germplasm of sugarcane (Jannoo et al. 2001; Hemaprabha and Sree Rangasamy 2001). Microsatellites have been found to be the marker of choice and the results of microsatellite analysis in sugarcane have indicated the importance of this marker system in characterize cane varieties and as genetic markers in sugarcane (Jannoo et al. 2001). Due to its high specificity, sequence-tagged microsatellite markers (STMS) can discriminate even closely related germplasm lines (Hemaprabha et al. 2005, 2013).

Sugarcane improvement is accelerated by the development and utilization of new breeding materials. In this study, we took 133 elite and advanced selections for estimating genetic diversity and to identify genetically more diverse genotypes and combinations making use of 20 sugarcane-specific STMS primers developed from Genbank database. We attempted to effectively use a few computational tools, as mentioned in the methods, to identify the diverse materials for further use in breeding. In view of the reported application of population structure analysis for getting a clear picture on gene flow and to assign individuals to populations (Pritchard et al. 2000), the study was aimed in this direction as well. Apart, the extent of variation as well as a graphical fingerprint of each clone was developed adopting graphical genotyping (GGT) following (Van Berloo 2007) who developed this important tool to facilitate the process of selection and evaluation of plant material. Sugarcane breeding is one of the toughest and time-consuming activities (Heinz 1987). In many cases, the size of germplasm collections limits their utilization in plant breeding. We intended to downsize the number of genotypes from 133, based on the concept of a core collection that would contain as much genetic diversity as the original set (Vaughan 1991) for harnessing the best out of a smaller and manageable parental pool.

Materials and methods

Plant material

The plant materials for the study were obtained from the fields at ICAR- Sugarcane Breeding Institute, Coimbatore.

A total of 133 clones were used for the study as given in (Supplementary Table 1).

DNA extraction and quality check

Plant sample was collected from the fields of ICAR—Sugarcane Breeding Institute, Coimbatore. DNA was isolated by C-TAB method (Doyle and Doyle 1987) and purified. The quantity and quality were checked using Nanodrop Spectrophotometer (ND-1000).

Primers used

About 20 primers were used to analyze and study the genetic diversity among the clones. Among 20 primers used, two primers were derived from EST sequences and 18 primers were from Genomic sequences. Primers used for this study are mentioned on (Supplementary Table 2).

Polymerase chain reaction (PCR) and electrophoresis

The DNA samples were diluted to get a final concentration of 20 ng/μl for PCR amplification. The PCR reaction was performed in a thermal cycler (Eppendorf, PRO-S) using a 10 μl reaction mix consisting of final concentration of 20 ng template DNA, 25 pmol each of forward and reverse primers, 0.3-unit Taq polymerase (Merck, India), 1x Taq Buffer (Merck, India), and 100 μM dNTPs (Merck, India). The basic cycling profile was 2 min at 95 °C followed by 35 cycles of 1 min at 94 °C, 40 s at the annealing temperature standardized for each primer (ranging from 55 to 58 °C), 40 s extension at 72 °C, and a final extension of 7 min at 72 °C. PCR products were resolved on 8% non-denaturing polyacrylamide gels in a vertical electrophoresis apparatus (Scie-Plas) and stained with Silver nitrate. The size of the amplified fragment was estimated using 100 bp DNA ladder. The gel documentation was done using Alpha Imager (Alpha Innotech).

Data analysis

All the genotypes listed in (Supplementary Table 1) were analyzed using 20 sugarcane-specific STMS primers as given in (Supplementary Table 2). The amplified bands were scored for presence (1) or absence (0) in all 133 genotypes and the data were used to calculate the Jaccard's similarity coefficient (JSCs) using the NTSYS- pc software (Version 2.0) (Rohlf 1990). Using molecular marker data, the degree of dissimilarity between two genotypes was derived (Nei and Li 1979) which would describe the allelic variation. Genetic similarity (GS) or Similarity Index (SI) that represents the proportion of molecular

markers common between any two individuals being compared was performed by comparing the number of rows of marker fragments of (apparently) similar size separated by electrophoresis. Genetic distances between each pair of lines were estimated as $GD = 1 - GS$. For genetic diversity analysis, Jaccard's coefficient (1908) was used considering its appropriateness with co-dominant markers like SSR. The pairwise Jaccard index is calculated by the following equation: $GS_{ij} = N_{ij}/[N_i + N_j + N_{ij}]$, where N_{ij} is the total number of bands common to lines i and j , and N_i and N_j are the number of bands only present in i and j , respectively.

A distance-based unweighted neighbour-joining cluster tree was constructed based on pairwise distance matrix and dissimilarity index. Dendrogram was constructed using the DARwin software with 133 sugarcane genotypes (Singh et al. 2013; Perrier and Jacquemoud-Collet 2006). Polymorphic information content (PIC) was calculated using the formula, $PIC = 1 - \sum P_{ij}^2$ (Anderson et al. 1993), where P_{ij} is the frequency of the j th allele for i th locus summed across all alleles for the locus.

To validate the correspondence between the cluster analysis and genetic similarity matrix cophenetic correlation coefficient (CCC) was calculated using the COPH and MXCOMP modules of NTSYS-pc (2.10) (Oliveira et al. 2009). Population distribution was analyzed using principal component analysis (PCA) which was performed on the basis of genotypic data. Principal coordinate analysis (PCoA) and PCA was carried out using the PAST 2 software (Hammer et al. 2001).

Analysis of the population structure of 133 sugarcane genotypes was carried out using STRUCTURE v.2.3.4, for which a Bayesian approach is used to infer population structure (Pritchard et al. 2000). Gene flow (N_m), or the number of migrants entering a population in each generation, was estimated using the formula given below, as suggested by Zhou et al. (2015). $N_m = 0.25 \times (1 - F_{st})/F_{st}$ (five independent runs with K value ranging from 2 to 8 and five iterations for each value of K was set).

To determine the actual K value, the following structure parameter was set with the possibility of admixture and allele frequency correlated: the length of the burning period was 50,000 iterations followed by 200,000 Monte Carlo Markov Chain replicates. To obtain the optimal K value, it was plotted against the mean estimate of log probability of the data $L(K)$. The actual number of subpopulations was identified using the maximum $L(K)$ value. The final population structure was calculated with ΔK , based on the second-order rate of change of likelihood distribution mean $L''(K)$ and with respect to K estimated using STRUCTURE Harvester, it showed a clear peak at optimal K value (Evanno et al. 2005; Earl and vonHoldt 2012). Analysis of molecular variance (AMOVA) was conducted to detect the genetic variance within and

among subpopulation using GenAlEx Ver 6.5 (Peakall and Smouse 2012; Kaur et al. 2016).

A combined analysis was done with all the genotypes from the given sets based on results with common primers using Jaccard's coefficient and genotyping was done using the graphical genotyping software GGT 2.0 (Van Berloo 2007).

Using the graphical genotyping software GGT 2.0 was used to analyze the marker data of 133 genotypes (Van Berloo 2007). GGT data graphically represented a GGT image that showed the first linkage group of all individuals next to each other. In the GGT image, the presence and absence of bands entered were shown in red and blue colour, respectively.

Results and discussion

The availability of genetic variability is an invaluable source for plant breeders for realizing further improvement in the form of new varieties and hybrids. Sugarcane is no exception and the early genetic gains are sustained through genetic base broadening (Heinz 1987). Genetic variation also holds the key to the ability of populations and species to persist over evolutionary period of time through changing environments (Narain 2000) for which a correct estimation of genetic variability is required. One of the ways to estimate genetic variability in a population is to analyze the different alleles at the various gene loci with the use of molecular markers (Solomon et al. 2010). In the present study, we analyzed the genetic variability among 133 elite genetic material of sugarcane using genomic and EST based STMS primers. Among the primers used, 70% were dinucleotide, 25% were trinucleotide, and 5% were tetranucleotide sequences and were earlier reported to be efficient in DNA fingerprinting and genetic diversity analysis in sugarcane (Hemaprabha et al. 2005).

Genomic and EST primers amplified distinct banding pattern among 133 elite sugarcane clones, which exhibited varying degrees of polymorphism. Twenty markers detected 359 alleles with a mean 17.95 primer, reflecting the genetic complexity and polyploidy genome of sugarcane (Fig. 1a, b). The polymorphism percentage ranged from 69.2 (NKS1) to 100% with (NKS 5, NKS 28, NKS 34, and NKS 52) (Table 1). The PIC values among the primers ranged from 0.29 (NKS 3) to 0.78 (NKS 52) with the mean of 0.62. The presence of bands NKS 9_{607,844} identified the clone Co 8013, while NKS 15₁₃₈ was specifically present in three clones viz CoA 7602, CoS 88216 and CoSe 92423. NKS 38₁₃₀ was unique to Co 13004, Co 13006 and CoS 92423 by its presence. The marker NKS 22₂₇₃ was unique to the clones Co 0216, CoC 671, and CoC 8201 by the absence of the band. NKS 57₁₂₀ was unique to the clone Co 0209 and

Fig. 1 **a** Polymorphism among 133 clones based on the STMS primer NKS 22. **b** Polymorphism among 133 clones based on the STMS primer NKS 57

Table 1 Polymorphism information content for each primer

S. no.	Primer name	PIC	Polymorphism %	Size range (bp)	No. of alleles
1	NKS 1	0.53	69.2	219–594	13
2	NKS 3	0.29	77.8	183–375	9
3	NKS 5	0.76	100.0	153–473	20
4	NKS 8	0.69	77.8	171–364	9
5	NKS 9	0.75	95.0	135–425	20
6	NKS 12	0.66	100.0	136–387	20
7	NKS 15	0.61	95.2	118–361	21
8	NKS 20	0.55	93.3	145–706	15
9	NKS 21	0.52	88.9	169–431	18
10	NKS 22	0.48	89.5	121–717	19
11	NKS 28	0.68	100.0	121–528	25
12	NKS 31	0.36	70.0	204–344	10
13	NKS 34	0.75	100.0	121–337	17
14	NKS 38	0.70	96.3	140–773	27
15	NKS 48	0.63	85.7	132–438	21
16	NKS 49	0.66	96.3	130–667	27
17	NKS 52	0.78	100.0	116–295	16
18	NKS 56	0.69	90.0	147–499	20
19	NKS 57	0.64	85.7	101–728	14
20	NKS 61	0.60	77.8	202–735	18
	Overall mean	0.62	89.43		17.95

NKS 57₁₁₄ was unique to Co 0209 and Co 93023, again by their absence. The absence of markers could be explained with the possibility of null alleles with microsatellite-based

markers. It has been reported that null alleles fail to amplify due to primer site variation, thereby a visible amplicon is not produced. In addition, these have been shown to be present

in large populations with an unusually high mutation rate (Chapuis and Estoup 2006). Sugarcane, being an ancient crop maintained by clonal propagation and the buffering of polyploidy preserving all the genetic changes over generations, is an ideal candidate, where one can expect null alleles in large frequencies. Null alleles for the particular locus appear as failed amplifications, in contrast to individuals that are heterozygous (for a null allele) appearing to be homozygous. This would explain the absence of bands to be unique to a genotype. The unique markers could be useful in clonal identification and hybridity analysis. More specifically, GGT images could serve as fingerprints of each of 133 genotypes, as discussed later.

Principal component analysis (PCA) is a dimension-reduction tool that can be used to reduce a large set of variables to a small set that still contains most of the information in the large set. PCA is a mathematical procedure that transforms a number of possibly correlated variables into a smaller number of uncorrelated variables called principal components (Price et al. 2006). In this study, PCA analysis of 133 genotypes based on variant-co variant matrix showed that the first and second component accounted 8.34 and 3.22% variance with eigen values of 5.61 and 2.17, respectively. Among 133 genotypes, 17 types showed eigen value above 1 as a sign of their high genetic variability. PCA scatter plot showed the Jolliffe cutoff value of 0.3569 and 79 genotypes fell above that and hence are more variable. PCA scatter plot revealed two main clusters, while the clones CoS

92263 and CoS 510 fell away from the main clusters, as shown in Fig. 2. Majority of the commercial canes developed at Sugarcane Breeding Institute located at Coimbatore fell in one cluster and the second cluster contained interspecific hybrids and commercial cultivars from other sugarcane research centers.

PCoA is based on Euclidean similarity index. First and the second coordinate axis showed 8.34 and 3.22% with eigen value 741.29 and 286.11, respectively. PCoA scatter plot showed two main clusters. In line with PCA, here also most of the commercial canes fell in cluster one, while the second cluster contained ISH clones and commercial cultivars from other centers (Fig. 3).

Cophenetic analysis was done to evaluate the goodness of fit of the resulting phylogenetic tree. This analysis based on SSR marker data showed consistency of the inferred relationship with a high CCC value of 0.84. CCC results obtained in this study were similar to the results of (Oliveira et al. 2009) and indicated that the cluster results were good and suitable for the calculation of genetic similarity matrix.

Similarity index of all possible combinations were calculated. A total of 8778 combinations were possible with 133 clones. SI values of the combinations showed that 1812 combinations were more diverse with SI between 0.5 and 0.6. Vast majority of combinations (8069) exhibited moderate similarity (SI between 0.60 and 0.70) and 693 combinations were in the range of SI of 0.70–0.80. In contrast, only 16 combinations were with SI between 0.81 and 0.90

Fig. 2 PCA scatter diagram analysis showing the distributions of 133 sugarcane genotypes

base broadening and genetic enrichment in India. We have the practice of excluding the poor performers and addition of new selections from breeding and pre-breeding trials to maintain enough genetic diversity in the parental pool used in commercial breeding programs. The results are clearly being translated into development of new varieties. Since 2000 AD, 55 new varieties have been released in India, of which 29 were from ICAR Sugarcane Breeding Institute (Shukla et al. 2018) and this accounted to 52.7% of the total releases. The institute also facilitated in the development of 24 location-specific varieties making use of the gene pool maintained at the National Hybridization Garden of the Institute.

Different clustering methods are adopted to estimate genetic relatedness of populations or accessions of a species. A radial tree was constructed using the DARwin software (V6) in which 133 genotypes clustered in two major clusters (Fig. 5). Cluster 1 with 81 genotypes formed a larger cluster,

which was sub clustered as subcluster 1, 2 and 5 with 49, 16, and 16 genotypes, respectively (Table 2). Among the subclusters, subcluster 5 largely contained the earlier bred clones as well as short duration clones. The cluster 2 with 52 clones contained two subclusters with 40 and 12 individuals. It could be seen that the subclusters 3 and 4 predominantly contained genotypes developed at other research stations, particularly from subtropical India and interspecific hybrids. Thus, there was a trend of clustering based on time and region of origin.

The genetic architecture of a set of germplasm lines can be precisely estimated by assessing the STRUCTURE of the population using molecular markers viz., SSRs or SNPs etc. (Horst and Wenzel 2007; Varshney et al. 2007). STRUCTURE, being a model-based approach, could discriminate 133 sugarcane genotypes based on SSR polymorphism. Bayesian genetic structure harvester computed at $K=2$, formed two major groups (result not shown). It clarifies the relationship

Fig. 5 Radial tree showing genetic relationship between 133 genotypes

Table 2 Cluster analysis of 133 genotypes of sugarcane based on the DARwin software

S. no.	Clusters	Subcluster	Clones
1	Cluster 1	Sub cluster 1	Co 06027, Co 06022, Co 06020, Co 06007, Co 06018, 2007-209, Co 06017, Co 95020, Co 95005, Co 91010, Co 0238, Co 0237, Co 0233, Co 0212, Co 775, Co 0209, Co 06002, Co 05011, Co 99004, Co 05009, Co 409, Co 06015, Co 0218, Co 0216, Co 0240, Co 0239, ISH 176, Co 93009, Co 13003, Co 11015, Co 11012, Co 86011, Co 0314, Co 06031, Co 06011, Co 06034, Co 2001-13, Co 98010, Co 06030, Co 06010, Co 13020, CoC 92061, Co 8371, 1148-13-11-2-253, Co 97009, Co 92002, Co 09013, Co 07010, Co 07004
		Sub cluster 2	Co 99008, Co 99006, CoC 671, Co 94012, Co 93023, Co 91002, Co 89034, CoM 0265, Co 89003, Co 86032, Co 87044, Co 87023, Co 98017, Co 98014, Co 06003, Co 94008
		Sub cluster 5	Co 8016, Co 7704, Co 8017, Co 8022, Co 8020, Co 8018, Co 740, Co 8341, Co 6806, Co 8347, Co 8338, 2007-172, Co 6304, SBI 2007-291, Co 403, Co 06012
2	Cluster 2	Sub cluster 3	Co 13006, CoS 94270, Co 13018, Co 62198, CoLk 8102, CoS 96268, Co 13004, CoSe 92423, CoC 8201, CoS 8436, Co86250, Co 7314, Co 86002, Co 13005, Co 13007, Co 97015, Co 2001-15, Co 12007, Co 7201, CoSnk 03-44, ISH 175, Co 1148, Co 8363, Co 86249, Co 13013, Q 63, Co 12011, Co 88025, Co 92020, CoPant 97222, CoT 8201, CoPant 84212, Co 13017, CoPant 90223, CoS 88216, CoA 7602, CoS 92263, BO 130, CoJ 64, CoS 510
		Sub cluster 4	BO 91, Co 0232, Co 8013, Co 7706, Co 8213, ISH 09, ISH 100, Co 13014, 85 R 186, ISH 176, Co 0118, NCo 310

obtained through STRUCTURE analysis with that of PCA and PCoA.

However, the best value of K at 5 (maximum ΔK of 5) divided 133 genotypes into five subpopulations viz. SP1, SP2, SP3, SP4, and SP5 (Fig. 6). Among the diverse groups, SP1 contained more number of genotypes (55), while SP2, SP3, SP4, and SP5 contained 14, 41, 10, and 13 genotypes, respectively (Fig. 7). A careful observation of the groups indicated that five subgroups are due to differences in genetic background, ecological environments of their origin as well as time periods of origin. The bigger group SP1 was largely comprised of the clones developed at the main sugarcane-breeding place in India, located at Coimbatore. Similar observation was made for SP2, while SP3 was mainly made up of varieties developed after 1985 AD. Of the remaining groups, SP4 was formed largely with clones from research stations other than at Coimbatore and interspecific hybrids, while SP5 comprised of clones of early origin. Clones from other stations and time periods were present in lesser proportions in all the five groups. These observations were similar to the radial tree based on the DARwin software as 81.95% of the genotypes of each cluster were similar in the two analyses. The results thus showed that location and

time of origin were two major factors that contributed to diversity. The presence of interspecific hybrids in different clusters could be explained by the fact that sugarcane varieties are interspecific hybrids, selected for economically valuable traits and the interspecific hybrids are no exception, though these were not given commercial status (as Co- or Co-allied clones). The observation of clustering based on the time of origin is significant in that their development is dependent on the available genetic sources at the point of their development and hence divergence. The advancement in genetic gain could also be attributed to difference in molecular diversity, which invites a detailed study of marker diversity in relation to generations of breeding.

AMOVA based on population STRUCTURE attributed 84% within group variation and 16% between group variation (Table 3). Fixation index (F_{st}) which indicate the variation within the population showed subpopulations SP4 ($F_{st} = 0.155$) and SP2 (0.152) to be more variable, while among the pairs of groups, SP2 and SP4 (0.266), SP4 and SP5 (0.229) and SP2 and SP 3 (0.226) exhibited more variability (Table 4). Analyzing the gene flow among populations, SP2 showed the least gene flow value of 0.22, the highest being in SP3 (0.76), while the average gene flow was

Fig. 6 Population structure analysis showing 133 sugarcane genotypes grouped into five subpopulations (SPs)

Fig. 7 Subpopulation analysis of 133 sugarcane genotypes

Table 3 Analysis of molecular variance among STRUCTURE detected population (GenAlEx 6.5)

Source	Degree of freedom (df)	Sum of squares (SS)	Mean sum of squares (MS)	Estimated variance	Percentage of variation (%)
Among population	4	2217.990	554.498	10.589	16
Within population	261	14658.386	56.162	56.162	84
Total	265	16876.376		66.752	100
Fixation index	0.159				

Table 4 Fixation index among subpopulation (GenAlEx 6.5)

Subpopulation	SP1	SP2	SP3	SP4	SP5
SP1	0.000				
SP2	0.094	0.000			
SP3	0.157	0.226	0.000		
SP4	0.189	0.266	0.089	0.000	
SP5	0.106	0.175	0.186	0.229	0.000
Mean	0.109	0.152	0.132	0.155	0.139

Table 5 Fixation index (F_{st}) and gene flow (N_m) estimation based on subpopulations of 133 sugarcane genotypes

Subpopulation (SP)	Number of genotypes	Proportion of genotypes	Mean F_{st} value	Gene flow (N_m) value	Average heterozygosity
1	55	0.330	0.2507	0.7472	0.2838
2	14	0.139	0.5284	0.2231	0.1890
3	41	0.277	0.2464	0.7671	0.3069
4	10	0.110	0.4046	0.3679	0.2644
5	13	0.145	0.4065	0.3650	0.2322

0.50. The highest heterozygosity was observed in SP3 and the lowest in SP2 (Table 5). Natural selection, gene mutation, and gene flow are the main causes of changes in genetic structure. It has been reported that gene flow value greater than 1.00 would resist the effect of genetic drift and prevent the differentiation of populations (Hamrick and Godt 1990; Slatkin 1981). N_m value obtained from the subpopulation reflected diversity among 133 genotypes. In light of these findings it could be seen that there existed enough potential among the genotypes under study for their successful

Fig. 8 GGT Image for 133 Sugarcane genotypes using 20 STMS primers

Table 6 Core set obtained from GGT

Core set	Co 775, Co 0232, Co 0233, Co 0237, Co 0239, Co 06002, Co 06007, Co 06012, Co 06022, Co 08013, Co 11012, Co 11015, Co 13003, Co 13020, Co 6806, Co 8016, Co 8017, Co 8338, Co 86011, Co 87023, Co 89003, Co 89034, Co 92002, Co 93009, Co 94008, Co 94012, Co 95005, Co 98014, Co 98017, Co 99004, Co 99006, Co 99008, CoC 671, CoM 0265, ISH 09, ISH 100, Co 7314, Co 7706, Co 88025, CoS 94270
----------	---

utilization in sugarcane improvement. For instance, SP2 (comprising of Co 99008, Co 99006, Co 94012, Co 93023, CoC 671, Co 89034, Co 91003, Co 06022, Co 98017, Co 87044, Co 06018, Co 89003, Co 98014 and Co 86032) with maximum variation, and the least gene flow and heterozygosity values would serve as the best group for utilization in breeding. Intercrossing of individuals from the subpopulations SP2, SP4, and SP5 also would be beneficial.

GGT images of 133 genotypes were constructed as chromosome drawings based on the presence of PCR amplified products and it could be noted that each image was distinct from one another (Fig. 8). This observation has two practical applications. The first is in demonstrating the power of STMS primers in individually discriminating each genotype of sugarcane in spite of common origin of some of the genotypes used for the study. The second is in the use of GGT images of sugarcane genotypes to serve as their unique identity. The heterozygous and polyploidy nature of sugarcane genome led to the amplification of more number of markers per locus and that could give rise to a large number of polymorphic markers to lead to the distinctiveness of the GGT images.

Core set

A core collection would serve as a good starting point when searching for new traits (Vaughan 1991) and could be used for in-depth evaluation and for increasing the knowledge of the entire collection (Knupffer and van Hintum 1995). In this study, forty among the 133 genotypes were shortlisted using GGT that almost represented the available genetic diversity. The diversity value of core selection was 2922.32 against 2469.29 of the original set (Table 6). Hence, this set of clones could be conveniently used in breeding program to cover the genetic diversity available in 133 clones.

Conclusion

The study could quantify genetic diversity among a set of elite genotypes for their judicious use in sugarcane improvement in India. Diverse genotypes, combinations and groups were identified subjecting the data on molecular marker polymorphism to analysis using robust statistical tools. The studied population exhibited moderately high genetic diversity which can be successfully exploited by hybridization

and selection. The clustering based on PCA, PCoA, radial tree, and population structure was in agreement with one another and the populations clustered largely based on the region and time of their development. GGT images clearly distinguished each clone with unique images and would be useful as their molecular signatures for their unambiguous identification. To simplify an otherwise tedious process of hybridization in sugarcane, a core set of 40 genotypes was identified that could satisfactorily represent the diversity among the 133 genotypes. This study thus made use of the molecular marker information for immediate benefit to breeders to strategically harness the genetic variability through breeding and to further diversify the parental pool through pre-breeding activities for faster genetic gains.

Acknowledgements The authors are grateful to the Indian Council of Agricultural Research and the Sugarcane Breeding Institute, Coimbatore for the funding and infrastructure.

Author contributions SPTS performed the experiments and wrote the manuscript and GH designed the work and revised the manuscript.

Compliance with ethical standards

Conflict of interest The authors have declared that no competing interests exist.

References

- Anderson A, Churchill GA, Autrique JE, Tanksley SD, Sorrells ME (1993) Optimizing parental selection for genetic linkage maps. *Genome* 36:181–186. <https://doi.org/10.1139/g93-024>
- Chapuis MP, Estoup A (2006) Microsatellite null alleles and estimation of population differentiation. *Mol Biol Evol* 24(3):621–631. <https://doi.org/10.1093/molbev/msl191>
- Doyle JJ, Doyle JL (1987) A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochem Bull* 19:11–15
- Earl DA, vonHoldt BM (2012) STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conserv Genet Resour* 4:359–361. <https://doi.org/10.1007/s12686-011-9548-7>
- Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol Ecol* 14:2611–2620. <https://doi.org/10.1111/j.1365-294X.2005.02553.x>
- Hammer OD, Harper AT, Paul DR (2001) PAST: paleontological statistics software package for education and data analysis. *Palaeontol Electron* 4:9
- Hamrick JL, Godt MJW (1990) Allozyme diversity in plant species. In: Brown AHD, Clegg MT, Kahler AL, Weir BS (eds) *Plant*

- population genetics, breeding and genetic resources. Sinauer Associates, Sunderland, pp 43–63
- Heinz DJ (1987) Sugarcane improvement through breeding. Elsevier, Amsterdam
- Hemaprabha G, Sree Rangasamy SR (2001) Diversity in sugarcane species and hybrids for peroxidase isozyme. *Sugar Tech* 3:40–44. <https://doi.org/10.1007/BF02945529>
- Hemaprabha G, Govindaraj P, Balasundaram N, Singh NK (2005) Genetic diversity analysis of Indian sugarcane breeding pool based on sugarcane specific STMS markers. *Sugar Tech* 7(2–3):9–14. <https://doi.org/10.1007/BF02945251>
- Hemaprabha G, Priji PJ, Padmanabhan TSS (2013) Molecular fingerprinting of recently notified sugarcane (*Saccharum* L.) varieties using STMS markers. *J Sugarcane Res* 3(2):107–117
- Hogarth DM, Cox MC, Bull JK (1997) Sugarcane improvement: past achievements and future prospects. In: Kang MS (ed) *Crop improvement for the 21st century*, 1st edn. Research Signpost, Trivandrum, pp 29–56
- Horst L, Wenzel G (2007) *Molecular marker systems in plant breeding and crop improvement*. Springer, Berlin
- Jaccard P (1908) Nouvelles recherches sur la distribution florale. *Bull Soc Vaud Sci Nat* 44:223–270
- Jannoo N, Forget L, Dookun A (2001) Contribution of microsatellites to sugarcane breeding program in Mauritius. *Proc Intl Soc Sugar Cane Technol* 24:637–639
- Kaur K, Sharma V, Singh V, Wani MS, Gupta RC (2016) Development of novel SSR markers for evaluation of genetic diversity and population structure in *Tribulus terrestris* L. (*Zygophyllaceae*). *3 Biotech* 6(2):156. <https://doi.org/10.1007/s13205-016-0469-8>
- Knupffer H, van Hintum T (2003) Summarized diversity—the barley core collection. In: von Bothmer R, van Hintum TH, Knupffer H, Sato K (eds) *Diversity in barley (*Hordeum vulgare*)*. Elsevier Science B.V, Amsterdam, pp 259–267
- Narain P (2000) Genetic diversity-conservation and assessment. *Curr Sci Bangalore* 79(2):170–175
- Nei M, Li WH (1979) Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proc Natl Acad Sci* 76:5269–5273. <https://doi.org/10.1073/pnas.76.10.5269>
- Oliveira KM, Pinto LR, Marconi TG, Mollinari M, Ulian EC, Chabregas SM, Souza AP (2009) Characterization of new polymorphic functional markers for sugarcane. *Genome* 52:191–209. <https://doi.org/10.1139/G08-105>
- Peakall R, Smouse PE (2012) GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research—an update. *Bioinformatics*. <https://doi.org/10.1093/bioinformatics/bts460> (in press)
- Perrier X, Jacquemoud-Collet JP (2006) DARwin software: dissimilarity analysis and representation for windows. <http://darwin.cirad.fr/darwin>. Accessed 12 Dec 2017
- Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D (2006) Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet* 38:904–909. <https://doi.org/10.1038/ng1847>
- Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics* 155:945–959
- Rohlf FJ (1990) NTSYS-pc Numerical taxonomy and multivariate analysis system, version 2.0 Exeter software. Setauket, New York
- Shukla SK, Zubair A, Awasthi SK, Pathak AD (2018) Sugarcane varieties identified by AICRP(S) in India. ICAR—All India Co-ordinated Research Project on Sugarcane, IISR, Lucknow, pp 1–111
- Singh N, Choudhury DR, Singh AK, Kumar S, Srinivasan K, Tyagi RK, Singh R (2013) Comparison of SSR and SNP markers in estimation of genetic diversity and population structure of Indian rice varieties. *PLoS One* 8(12):e84136. <https://doi.org/10.1371/journal.pone.0084136>
- Slatkin M (1981) Estimating levels of gene flow in natural populations. *Genetics* 99:323–335
- Solomon KF, Martin I, Zeppa A (2010) Temporal genetic structure patterns in tropical maize populations under reciprocal recurrent selection. *Euphytica* 176(2):239–249. <https://doi.org/10.1007/s10681-010-0238-4>
- Tena E, Mekbib F, Ayana A (2014) Analysis of genetic diversity and population structure among exotic sugarcane (*Saccharum* spp.) cultivars in Ethiopia using simple sequence repeats (SSR) molecular markers. *Afr J Biotech* 13:46. <https://doi.org/10.5897/AJB2014.13849>
- Van Berloo R (2007) GGT: user manual Version 2.0. Wageningen (The Netherlands): Wageningen University. http://www.plantbreeding.wur.nl/Software/ggt/ggt2_manual.pdf
- Varshney RK, Marcel TC, Ramsay L, Russell J, Roder MS, Stein N, Waugh R, Langridge P, Niks RE, Graner A (2007) A high-density barley microsatellite consensus map with 775 SSR loci. *Theor Appl Genet* 114(6):1091–1103. <https://doi.org/10.1007/s00122-007-0503-7>
- Vaughan DA (1991) Choosing rice germplasm for evaluation. *Euphytica* 54:147–154. <https://doi.org/10.1007/BF00039601>
- Zhou R, Wu Z, Jiang FL, Liang M (2015) Comparison of gSSR and EST-SSR markers for analyzing genetic variability among tomato cultivars (*Solanum lycopersicum* L.). *Genet Mol Res* 14:13184–13194. <https://doi.org/10.4238/2015.October.26.14>