



OPEN ACCESS

EDITED BY

Ting Peng,
Henan Agricultural University, China

REVIEWED BY

Milind B. Ratnaparkhe,
ICAR Indian Institute of Soybean Research,
India
Juliano Lino Ferreira,
Embrapa Pecuária Sul, Brazil

*CORRESPONDENCE

Ashutosh Singh

✉ singh.ashutosh026@gmail.com

RECEIVED 07 August 2023

ACCEPTED 09 October 2023

PUBLISHED 27 October 2023


CITATION

Divakar S, Jha RK, Kamat DN and Singh A (2023) Validation of candidate gene-based EST-SSR markers for sugar yield in sugarcane. *Front. Plant Sci.* 14:1273740. doi: 10.3389/fpls.2023.1273740

COPYRIGHT

© 2023 Divakar, Jha, Kamat and Singh. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Validation of candidate gene-based EST-SSR markers for sugar yield in sugarcane

S. Divakar¹, Ratnesh Kumar Jha², D. N. Kamat³ and Ashutosh Singh ^{2*}

¹Department of AB&MB, CBSH, Dr. Rajendra Prasad Central Agricultural University (RPCAU), Samastipur, Bihar, India, ²Centre for Advanced Studies on Climate Change, Dr. Rajendra Prasad Central Agricultural University (RPCAU), Samastipur, Bihar, India, ³Sugarcane Research Institute, Dr. Rajendra Prasad Central Agricultural University (RPCAU), Samastipur, Bihar, India

Sugarcane (*Saccharum* spp.) is a widely cultivated crop that fulfils approximately 75% of the sucrose demand worldwide. Owing to its polyploidy and complex genetic nature, it is difficult to identify and map genes related to complex traits, such as sucrose content. However, association mapping is one of the alternatives for identifying genes or markers for marker-assisted selection. In the present study, EST-SSR primers were obtained from *in silico* studies. The functionality of each primer was tested using Blast2Go software, and 30 EST-SSR primers related to sugar content were selected. These markers were validated using association analysis. A total of 70 F1 diverse genotypes for sugar content were phenotypes with two check lines. All parameters related to sugar content were recorded. The results showed a significant variation between the genotypes for sugar yield traits such as Brix value, purity, and sucrose content, etc. Correlation studies revealed that the Brix%, sucrose content, and sucrose recovery were significantly correlated. An association analysis was performed using mixed linear model to avoid false positive associations. The association analysis revealed that the SEM 407 marker was significantly associated with Brix% and sucrose content. The SEM 407 primers are putatively related to diphosphate-fructose-6-phosphate 1-phosphotransferase which is associated with Brix% and sucrose content. This functional marker can be used for marker-assisted selection for sugar yield traits in sugarcane that could accelerate the sugarcane breeding program.

KEYWORDS

sugarcane, EST-SSR, candidate genes, sugar content, association mapping, marker-assisted selection

Introduction

Sugarcane (*Saccharum* spp.) belongs to the Poaceae/Gramineae family of the Andropogoneae tribe. *Saccharum* and its species were commercially used for sugar production owing to their high biomass and sucrose accumulation (Godshall and Legendre, 2003). The sugarcane genome is complex polyploid in nature, with *S. officinarum* having a basic chromosome number of $x = 10$ ($2n = 80$) and *S. spontaneum* $x = 8$ ($2n = 40-128$). Thus, there are two distinct chromosomes that coexist in modern cultivars (D'Hont et al., 1994; Zhang et al., 2018).

Sucrose is a commercial component of sugarcane, and the improvement of sugar recovery is the primary focus of any crop improvement program. The identification of genes or marker for sugar yield is an important strategy for the improvement of sugarcane. The mapping of genes is a promising tool for characterizing genetic architecture such as yield component traits, such as sucrose yield, cane yield, stalk diameter, stalk height, stalk number, and stalk weight, as well as resistance to diseases, pests, and abiotic stresses (Aitken et al., 2008; Welham et al., 2010; Singh et al., 2013; Gazaffi et al., 2014; Margarido et al., 2015; Balsalobre et al., 2016; Balsalobre et al., 2017; Yang et al., 2018). The complex polyploid and highly heterogeneous genetic nature of sugarcane association mapping could establish the QTL from linkage disequilibrium between the markers and the trait. Surveying a large number of genotypes in the existing germplasm of sugarcane can be helpful in finding associations between the markers and traits, using association mapping (Wei et al., 2006; Banerjee et al., 2015). To avoid spurious associations, the population structure and kinship of the association map population were employed to elucidate inferences (Lander and Schork, 2006). Validation of all those markers linked with QTL will have been identified by means of association mapping in a diverse population (Korir et al., 2013; Picañol et al., 2013; Ukoskit et al., 2019).

The expressed sequence tag (EST) database was used to identify the targeted SSR markers because ESTs are considered effective for the direct association with the trait of interest (Dudhe and Sarada, 2012). The interspecific transferability of expressed sequence tags derived from simple sequence repeats (EST-SSRs) and genomic SSRs is well established (Wen et al., 2010). EST-SSR primers are more beneficial than anonymous SSRs from untranslated regions (UTRs) or non-coding sequences, being frequently more transferrable between closely related genera (Pashley et al., 2006; Chapman et al., 2009). Due to the primer target sequences' location in the expressed DNA regions, which are predicted to be reasonably well preserved, there is a higher likelihood that the marker will be transferable across species borders (Varshney et al., 2005). EST-SSRs appear to disclose comparable amounts of polymorphism compared to SSRs found in UTRs despite their potential to reflect selectively harmful frame-shift mutations in coding areas. This is most likely because these coding regions have evolved to contain tri-nucleotide repeats (Ellis and Burke, 2007). Since EST-SSRs are physically connected to expressed genes, they constitute potentially useful markers. EST-SSR markers have a greater average rate of transferability between species than genomic SSRs because they reflect the expressed regions of a genome (Gupta et al., 2003). EST-SSR markers have been effectively used in

gene tagging, linkage map construction, and QTL mapping (Qiu et al., 2010). Varietal crop improvement in various crops has become more feasible with the establishment of EST-SSR markers (Qiu et al., 2010; Ukoskit et al., 2019). EST-SSRs are highly regarded as a tool for breeding practices, perhaps because of their direct association with the genes of interest. It is also used in the identification of candidate genes in breeding and conservation input and population genetics studies (Yu et al., 2011). A mapping population obtained from a cross between commercial cultivars indulges in the introduction of EST-SSR markers into sugarcane linkage mapping (Oliveira et al., 2007; Palhares et al., 2012). There are some published reports of association mapping in sugarcane for traits such as biotic stress, cane yield, and sugar content (Wei et al., 2006; Wei et al., 2010; Débibakas et al., 2014; Banerjee et al., 2015; Gouy et al., 2015; Singh et al., 2016; Barreto et al., 2019; Fickett et al., 2019; Ukoskit et al., 2019; Coutinho et al., 2022). There are only a few studies that have investigated the identification of markers or genes for sugar yield traits using interspecific crosses (Reffay et al., 2005; Ukoskit et al., 2019). However, an association analysis requires a large population size and numerous EST-SSR primers. Moreover, these limitations could be avoided by choosing a diverse population with candidate genes for sugar yield for the validation of markers using an association study.

Therefore, the present study was conducted to identify different candidate gene-based EST-SSR markers from *in silico* studies, and these markers were validated using the association analysis of a diverse collection of sugarcane genotypes.

Materials and methods

Plant material

A total of 70 F1 diverse sugarcane genotypes for sugar yield traits were obtained from 14 different crosses, and five genotypes were chosen from each cross. The parents of all crosses were developed by crossing of *Saccharum officinarum* and *Saccharum spontaneum*. A few genotypes (BO102GC, BO137GC, and BO139GC) were also developed by general crosses (GC) for more variability. Seventy genotypes with two check lines (CoP16437 and CoP2061) were used in this study (Supplementary Table S2). These parents had contrasting natures for sugar and fiber yields. Because of limited seeds, these 70 genotypes were planted in an augmented complete block designed in the year 2021 at the Pusa farm of Dr. Rajendra Prasad Central Agricultural University, Samastipur, India. All genotypes were grown in seven different blocks, and each block had 10 genotypes with two check lines. The check lines were planted randomly in each block. The plot size of each block was 5.4 m². All standard agronomic practices were followed to raise the crops.

Phenotypic data and field data analysis

The population of 70 F1 genotypes was a phenotype for sugar-related traits after harvesting all genotypes. Brix, polarization (pol), sugar yield, and purity were recorded from four random stalks taken from each plot.

(1) Brix value (%): Brix value was measured using a hand-held refractometer. One degree of Brix is equal to the presence of 1 g sucrose in 100 g of the solution, and $1^{\circ}\text{Brix} = 1\% \text{ Brix}$. The strength of the solution was measured as a percentage of its mass. The reading was recorded using a refractometer with a sharp needle pierced through the stalk, and the collected substance was placed on a refractometer glass. The reading was recorded by the angle of the refractive index.

(2) Sucrose (%): Sucrose percentage was calculated using the following formula (Mehareb and Abazied, 2017):

$$\text{Sucrose (\%)} = \frac{\text{Brix \%} \times \text{purity \%}}{100}$$

(3) Purity (%): The percentage purity of the juice was calculated using the following formula (Mehareb and Abazied, 2017):

$$\text{Purity (\%)} = \frac{\text{Mass of the pure substance}}{\text{Mass of the impure sample}}$$

where pure sample: sucrose (%) and impure sample: Brix (%).

(4) Sugar recovery (CCS %): It was calculated using the following formula (Mehareb and Abazied, 2017):

$$\text{CCS \%} = [\text{Sucrose \%} - \{\text{Brix \%} - \text{Sucrose \%}\} \times 0.4] \times 0.73$$

(5) Sugar yield (t/ha)

Sugar yield was calculated using the following formula (Mehareb and Abazied, 2017):

$$\text{Sugar yield} = \frac{\text{Cane yield} \times \text{CCS \%}}{100}$$

Phenotypic data analysis

Pearson's correlation coefficients (r) between traits were calculated using the SAS CORR procedure based on trait means. Traits that were distributed normally with the Shapiro-Wilk test were considered normal data (Weber and Moorthy, 1952). Morphological data were used for the principal component analysis (PCA) using R studio. The data were imported in Excel format, and the eigenvalue must be greater than 1 for the variables. A bi-plot analysis was conducted using eigenvalues.

Extraction of DNA

A total of 500 mg of young leaf tissue was collected for marker analysis, and DNA was extracted using the cetyltrimethylammonium bromide method of Srivastava and Gupta (2008). DNA quantity and quality were determined using 1.0% agarose gel electrophoresis and nanodrop spectrophotometry, respectively.

Identification of a suitable EST-SSR marker and its validation by association study

Sugarcane is a polyploid crop, and its genome size and structure vary from genotype to genotype. Therefore, EST-SSR markers are

best suited for tagging complex traits such as sugar yield. A total of 213 EST-SSR primers of sugarcane were identified in an *in silico* study (Supplementary Table S1). Furthermore, the functionality of these primers was tested using Blast2Go software. A total of 30 EST-SSR primers related to sugar yield were selected for this study based on their functionality (Supplementary Table S3). The PCR products were separated at 3.5–4.0% agarose gel. Out of 30, a total of 25 primers were amplified by PCR and used for further analysis. All 25 primers were scored based on the presence (1) or absence (0) of bands in the 70 genotypes of the mapping population.

Validation of EST-SSR primers using association mapping

The similarity coefficient among the genotypes was calculated using the genetic distance (Nei and Li, 1979). A neighbor-joining dendrogram was constructed using Past3 software. N-J analysis was performed using multivariate clustering, and the tree was constructed by Euclidean genetic distance with bootstrap replications of 100. The population structure was analyzed using STRUCTURE software to estimate the number of groups/subpopulations by setting the burning period length to 100,000, and each value of K was run three times with the K value varying from 1 to 10. Furthermore, a Q value below 0.9 was described as an admixture. An association analysis was performed using a mixed linear model as described by Yu and Buckler (2006). It was performed using TASSEL incorporating the Q matrix and K matrix to avoid false positives. The significant threshold for association was set at $P < 0.05$.

Results and discussion

Sugarcane is polyploid crop with a complex genome, and it is difficult to interpret genome data. Less information about the sugarcane genome makes gene manipulation very difficult, but the identification of the functional genes responsible for sucrose accumulation makes it possible. EST-SSR markers were considered to be a highly regarded breeding tool as they are able to localize the functional gene by marker association (Palhares et al., 2012) since they may be directly associated with the gene expressing a particular trait.

Phenotypic yield data analysis

The germplasms used in this study were produced using an inter-varietal cross to validate primers for sugar yield traits in sugarcane. The yield distribution showed that the selected genotypes varied for sugar yield traits (Supplementary Figure S1). These germplasms were derived by crossing the sugarcane genotypes with contrasting sugar yield, and a few crosses were produced by general cross. The maximum, minimum, and mean values of all five phenotypic traits showed variation in the population for sugar yield traits. The Brix (%) values vary from

19.27% to 22.72%, with a mean of 20.65%. Similar trends were recorded for purity, sucrose content, sugar recovery, and sugar yield. This indicates that the selected germplasm shows variability and is appropriate for the study. The phenotypic correlations between the five traits were significant. The highest phenotypic correlation was found between sucrose content and sugar recovery, while the lowest phenotypic correlation was found between Brix value and purity (Table 1).

Population structure and genetic relationship

Many studies were conducted on sugarcane considering its complex polyploid genome, and several assumptions are not fulfilled for its complex structure; therefore, the applicability of this algorithm may be limited in sugarcane (Wei et al., 2006). Analyzing the sugarcane subpopulation using Structure software and mixed linear model provides an opportunity to track the gene related to complex traits of sugar yield in sugarcane. In the present study, minimum population size with a maximum variation was used, which is the most favorable for association analysis (Wei et al., 2006). Spurious associations were controlled, while the power to detect true associations was maximized using PCA as a random component to control for population structure (Pastina et al., 2012). PCA, as a random component, is included in the analysis, and the large population structure is captured with the first few axes that account for most of the variation, while the more subtle relationships among individuals are captured by the remaining significant axes. The population showed a clear and continuous variation in its structure, PCA, and neighbor-joining dendrogram (Fickett et al., 2019; Ukoskit et al., 2019). Furthermore, most of the structures found in these genotypes seem to originate from subtle kinship relationships rather than a large-scale population structure. The mapping population was diverse and highly heterozygous, and environmental conditions

play a major role in the formation of sugar. The PCA results showed that Brix%, sucrose content, and CCS% were highly positively correlated, and the CCS and cane yield are highly negatively correlated (Figure 1). The results of the PCA and neighbor-joining dendrogram were coherent and showed no disjuncture in the population. The biplot of PCA overlays both individuals and the variables in a single graph. The loading range was varied from -5 to 5. The high absolute loadings were directed to either positively or negatively describe the variable that strongly influences the component, and a value less than that of the high loading indicates that they had a weak influence on the component. Sucrose content, Brix, and CCS% were highly positively correlated, as identified by PCA, and they showed 49% of the total variation in the phenotypic data (Figure 1). Hence, the traits directly related with sugar yield showed significant variability and indicate the diverse genotypic nature of germplasms.

The markers were identified by surveying the sugarcane database SUCEST, with their functionality scored by BLAST to determine the homology and putative function of the marker sequence (Supplementary Table S1). EST-SSR markers are derived from the expression regions of the genome and have greater potential for the direct association of the trait. The data Blast2Go showed that, at every 18.60 kb, one SSR motif was found to be very similar to cotton and wheat (Cardle et al., 2000). Pinto et al. (2004) reported the density of SSR in sugarcane at every 16.90 kb.

The neighbor-joining dendrogram is a bottom-up clustering method designed to provide a single tree and may be able to produce more than one dendrogram from the same data (Saitou and Nei, 1987). This method provides faster and better results than UPGMA, and most implementations provide a single tree (Page and Holmes, 2009). The dendrogram of the neighbor-joining relationship based on EST-SSR allele frequencies separated the populations into three differentiated clusters: A, B, and C. Cluster A has 33 genotypes, cluster B has 31 genotypes, and cluster C has 13 genotypes (Figure 2).

TABLE 1 Mean performance and the correlations between traits at the time of harvesting of sugarcane.

Statistics	Brix	Purity	Sucrose content	Sugar recovery	Sugar yield
Mean	20.65	85.88	18.12	14.33	10.4
Maximum	22.72	87.79	20.06	15.97	16.41
Minimum	19.27	83.45	16.9	13.2	6.76
Correlation ^a					
Brix					
Purity	0.127 ^{ns}				
Sucrose content	0.952*	0.139 ^{ns}			
Sugar recovery	0.947*	0.208*	0.981*		
Sugar yield	0.272*	0.345*	0.267*	0.301*	
Significant* $P < 0.05$					

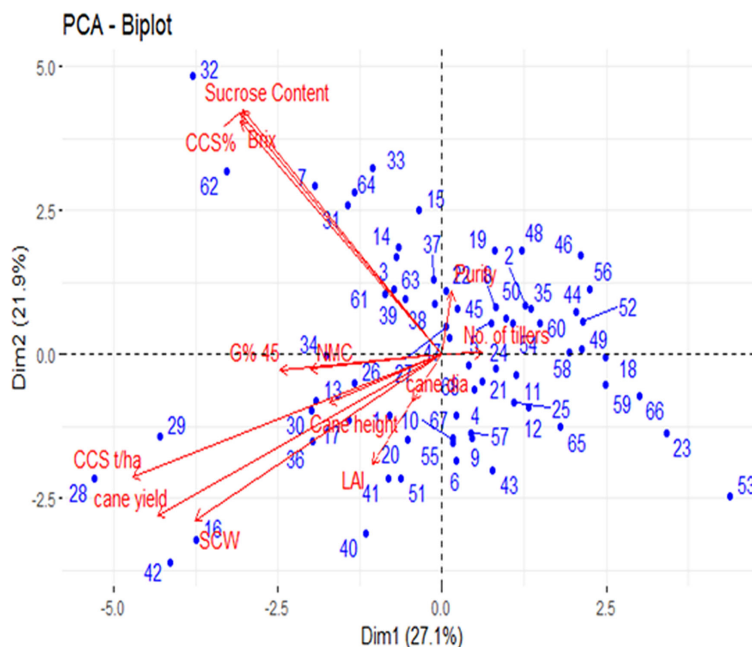


FIGURE 1 Biplot of principal component analysis PC1 (Dim1) and PC2 (Dim2); Where the dot represents the genotypes (70); Arrow represents the traits.

The Q matrix was calculated, and grouping of the population was performed using genotypic data. The result showed the marker-based kinship of the population. The number of subpopulations (ΔK) was identified by structure and maximized at $K = 5$. Populations with a single color were not shared by any other group, indicating that they were genetically distinct. The populations showed that the sharing of colors was similar to that of another group of individuals in the population (Figures 3, 4). The marker-based kinship matrix underlying the study of Yu et al. (2006) was determined based on the definition

that random pairs of genotypes are unrelated, whereas Zhao et al. (2007) defined pairs of genotypes that do not share any allele as unrelated. The results clearly indicated that most of the genotypes were different.

Validation of EST-SSR markers

Sugarcane crops are polyploid and exhibit a high level of variation in the F1 generation. Therefore, validation of primers

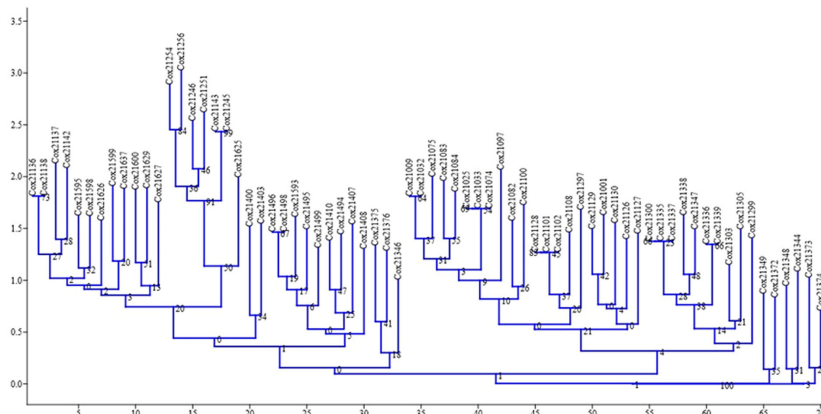


FIGURE 2 Dendrogram showing the relationship among 70 genotypes of sugarcane using EST-SSR markers.

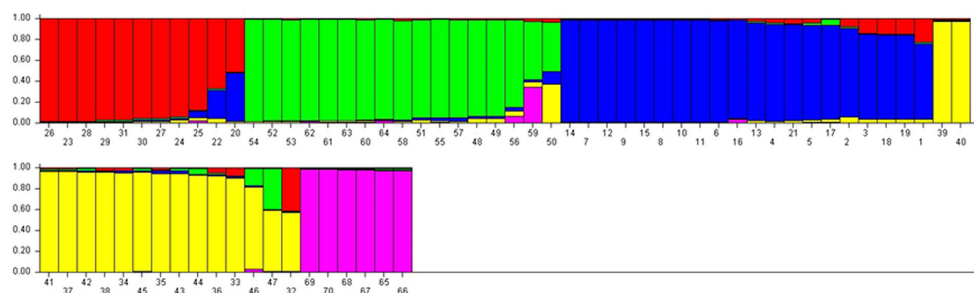


FIGURE 3

STRUCTURE analysis of bar plot. Populations with one solid color that is not shared by another group are genetically distinct. Populations that share colors are more similar. Bar graphs for five sub-populations are indicated by different colors. The vertical coordinates indicate the membership coefficient of each individual and the horizontal represents the genotypes.

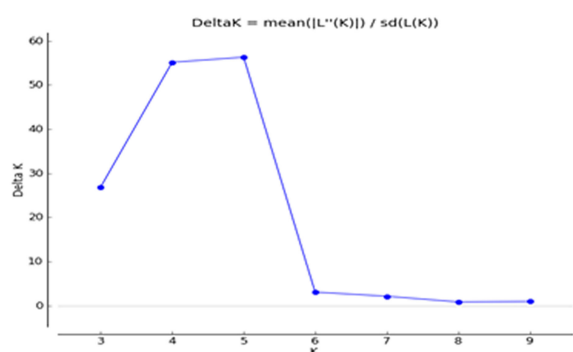


FIGURE 4

Delta K of STRUCTURE software using Evanno's criterion.

using a simple chi-square is not possible. Hence, the primers were validated using association mapping. If the number of primers is a major limitation of the study, then the candidate gene approach for association mapping is best suited. A total of 30 EST-SSR markers were used for the association analysis of sugar yield traits. Of the 30 primers, 25 were amplified, and all these primers were tested for association studies. EST-SSR markers are a tool for association studies (Ukoskit et al., 2019; Coutinho et al., 2022). The EST-SSR markers related to sucrose content could be more effective than markers that focus on the varied functions of the gene. The linkage disequilibrium (LD) decay plot for the r^2 values between the markers was plotted against the genetic distance. The highest frequency of loci pair in LD is mapped less than 3 bp. The lowest frequency of loci pair was more than 20 bp, indicating that the probability of LD is low between distinct loci pairs. The majority of the loci pairs in LD with $r^2 > 0.01$ at $P < 0.05$ were found in ≤ 20 bp (Figure 5). The values decreased as the genetic distance between the loci pairs increased. EST-SSR marker represented by the different regions of the genome was associated with the trait of interest at a P -value of 0.05. SEM407 was significantly associated with Brix%, sucrose content, and sugar recovery (CCS%) (Table 2). Except for SEM407, other EST-SSR markers did not show any association with

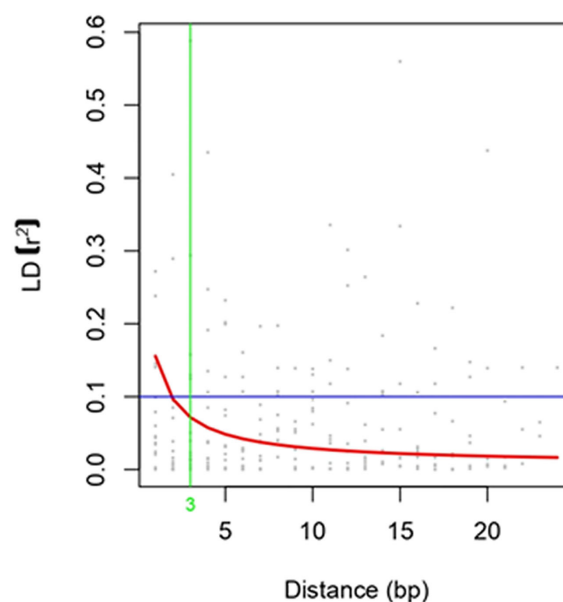


FIGURE 5

Linkage Disequilibrium Decay Plot.

sugar yield traits because the functional allele discovered in the mapping population might not be recognized in plants because it is rare in the larger germplasm. Compared to the genes tagged in the mapping population, the sugarcane accessions in the association population may have various trait-related alleles of various genes at various sites. Additional functional alleles that are absent from the mapping population may be found by validating the marker-trait relationship (Peace and Norelli, 2009). Although the number of markers used in this study is relatively low, the marker-trait association of SEM407 was significant. This marker was found to be significant for the sugar-related yield traits of genotypes. These results suggest that the association approach used in this study is consistent with the detection of QTL associated with sugar yield traits. The marker identified to the respective QTLs or genes should

TABLE 2 Associations study between EST-SSRs and sugar-related traits at $P < 0.05$.

Markers	Association mappin2				
	Brix%	Purity	Sucrose content	Sugar recovery (CCS %)	Sugar yield
	P value	P value	P value	P value	P value
SEM2	0.07	0.47	0.08	0.1	0.2
SEM58	0.29	0.44	0.3	0.3	0.38
SEM112	0.3	0.2	0.29	0.3	0.2
SEM117	0.70	0.68	0.5	0.47	0.25
SEM159	0.70	0.77	0.64	0.62	0.22
SEM168	0.72	0.39	0.79	0.82	0.74
SEM191	0.99	0.26	0.84	0.90	0.56
SEM199	–	–	–	–	–
SEM203	0.43	0.91	0.41	0.43	0.96
SEM358	0.89	0.5	0.77	0.82	0.39
SEM368	0.12	0.85	0.3	0.27	0.48
SEM369	0.95	0.86	0.59	0.57	0.76
SEM407	0.002*	0.70	0.002*	0.002*	0.70
SEM425	0.29	0.40	0.2	0.1	0.55
SEM428	0.92	0.77	0.95	0.95	0.97
SEM430	0.58	0.61	0.56	0.58	0.38
SEM432	0.91	0.25	0.98	0.93	0.92
SEM433	0.63	0.46	0.7	0.63	0.56
SEM435	0.83	0.1	0.77	0.86	0.57
SEM436	0.73	0.87	0.62	0.63	0.41
SEM437	0.15	0.86	0.12	0.13	0.28
SEM439	0.76	0.12	0.74	0.80	0.11
SEM440	0.92	0.43	0.98	0.93	0.43
SEM454	0.46	0.96	0.59	0.56	0.30
SEM456	0.99	0.79	0.87	0.85	0.23

The experiment-wise threshold was based on the Bonferroni corrected method. The thresholds for marker significance were at $*P < 0.05$.

be used for marker-assisted selection (MAS). MAS for simply inherited traits are gaining increasing importance in breeding programs, allowing the acceleration of the breeding process of sugarcane (Francia et al., 2005). This study would be helpful for the plant breeders in marker-assisted selection in the prospect of achieving higher sugar yields while designing their crossing program.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary Material, further inquiries can be directed to the corresponding author/s.

Author contributions

SD: Writing – original draft, Data curation, Formal Analysis, Methodology. RJ: Writing – review & editing, Funding acquisition, Supervision. DK: Writing – review & editing, Data curation. AS: Writing – original draft, Conceptualization, Funding acquisition, Supervision.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This research was funded by Bihar government under Climate Resilient Agriculture Program (India).

Acknowledgments

The authors thank to Centre for Advanced Studies on Climate Change, RPCAU, Pusa and Department of Molecular Biology and Biotechnology, CBSH, RPCAU, Pusa Samastipur, Bihar, for providing the necessary infrastructure for us to carry out research work.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

- Aitken, K. S., Hermann, S., Karno, K., Bonnett, G. D., McIntyre, L. C., and Jackson, P. A. (2008). Genetic control of yield related stalk traits in sugarcane. *Theor. Appl. Genet.* 117, 1191–1203. doi: 10.1007/s00122-008-0856-6
- Balsalobre, T. W. A., da Silva Pereira, G., Margarido, G. R. A., Gazaffi, R., Barreto, F. Z., Anoni, C. O., et al. (2017). GBS-based single dosage markers for linkage and QTL mapping allow gene mining for yield-related traits in sugarcane. *BMC Genomics* 18 (1), 1–19. doi: 10.1186/s12864-016-3383-x
- Balsalobre, T. W., Mancini, M. C., Pereira, G. D. S., Anoni, C. O., Barreto, F. Z., Hoffmann, H. P., et al. (2016). Mixed modeling of yield components and brown rust resistance in sugarcane families. *Agro. J.* 108 (5), 1824–1837. doi: 10.2134/agronj2015.0430
- Banerjee, N., Siraree, A., Yadav, S., Kumar, S., Singh, J., Kumar, S., et al. (2015). Marker-trait association study for sucrose and yield contributing traits in sugarcane (*Saccharum* spp. hybrid). *Euphytica* 205, 185–201. doi: 10.1007/s10681-015-1422-3
- Barreto, F. Z., Rosa, J. R. B. F., Balsalobre, T. W. A., Pastina, M. M., Silva, R. R., Hoffmann, H. P., et al. (2019). A genome-wide association study identified loci for yield component traits in sugarcane (*Saccharum* spp.). *PLoS One* 14 (7), e0219843. doi: 10.1371/journal.pone.0219843
- Cardle, L., Ramsay, L., Milbourne, D., Macaulay, M., Marshall, D., and Waugh, R. (2000). Computational and experimental characterization of physically clustered simple sequence repeats in plants. *Genetics* 156 (2), 847–854. doi: 10.1093/genetics/156.2.847
- Chapman, M. A., Hvala, J., Strever, J., Matvienko, M., Kozik, A., Michelmore, R. W., et al. (2009). Development, polymorphism, and cross-taxon utility of EST-SSR markers from safflower (*Carthamus tinctorius* L.). *Theor. Appl. Genet.* 120, 85–91. doi: 10.1007/s00122-009-1161-8
- Coutinho, A. E., da Silva, M. F., Perecin, D., Carvalheiro, R., Xavier, M. A., de Andrade Landell, M. G., et al. (2022). Association mapping for sugarcane quality traits at three harvest times. *Sugar Tech.* 24 (2), 448–462. doi: 10.1007/s12355-021-01056-5
- D'Hont, A., Lu, Y. H., León, D. G. D., Grivet, L., Feldmann, P., Lanaud, C., et al. (1994). A molecular approach to unraveling the genetics of sugarcane, a complex polyploid of the Andropogoneae tribe. *Genome* 37 (2), 222–230. doi: 10.1139/g94-031
- Débibakas, S., Rocher, S., Garsmeur, O., Toubi, L., Roques, D., D'Hont, A., et al. (2014). Prospecting sugarcane resistance to sugarcane yellow leaf virus by genome-wide association. *Theor. Appl. Genet.* 127, 1719–1732. doi: 10.1007/s00122-014-2334-7
- Dudhe, M. Y., and Sarada, C. (2012). Comparative assessment of microsatellite identification tools available in public domain. *DOR News Lett.* 18 (2), 8–9.
- Ellis, J. R., and Burke, J. M. (2007). EST-SSRs as a resource for population genetic analyses. *Heredity* 99 (2), 25–132. doi: 10.1038/sj.hdy.6801001
- Fickett, N., Gutierrez, A., Verma, M., Pontif, M., Hale, A., Kimbeng, C., et al. (2019). Genome-wide association mapping identifies markers associated with cane yield components and sucrose traits in the Louisiana sugarcane core collection. *Genomics* 111 (6), 1794–1801. doi: 10.1016/j.ygeno.2018.12.002
- Francia, E., Tacconi, G., Crosatti, C., Barabaschi, D., Bulgarelli, D., Dall'Aglio, E., et al. (2005). Marker assisted selection in crop plants. *Plant Cell Tissue Organ Culture* 82, 317–342. doi: 10.1007/s11240-005-2387-z
- Gazaffi, R., Margarido, G. R., Pastina, M. M., Mollinari, M., and Garcia, A. A. F. (2014). A model for quantitative trait loci mapping, linkage phase, and segregation pattern estimation for a full-sib progeny. *Tree Genet. Genomes* 10, 791–801. doi: 10.1007/s11295-013-0664-2
- Godshall, M. A., and Legendre, B. L. (2003). *SUGAR| Sugarcane Encyclopedia of Food Sciences and Nutrition* (Second Edition), Elsevier.
- Gouy, M., Rousselle, Y., Thong Chane, A., Anglade, A., Royaert, S., Nibouche, S., et al. (2015). Genome wide association mapping of agro-morphological and disease resistance traits in sugarcane. *Euphytica* 202, 269–284. doi: 10.1007/s10681-014-1294-y
- Gupta, P. K., Rustgi, S., Sharma, S., Singh, R., Kumar, N., and Balyan, H. S. (2003). Transferable EST-SSR markers for the study of polymorphism and genetic diversity in bread wheat. *Mol. Genet. Genomics* 270, 315–323. doi: 10.1007/s00438-003-0921-4
- Korir, N. K., Han, J., Shangguan, L., Wang, C., Kayesh, E., Zhang, Y., et al. (2013). Plant variety and cultivar identification: advances and prospects. *Crit. Rev. biotech.* 33 (2), 111–125. doi: 10.3109/07388551.2012.675314
- Lander, E. S., and Schork, N. J. (2006). Genetic dissection of complex traits. *Focus* 265 (3), 2037–2458.
- Margarido, G. R. A., Pastina, M. M., Souza, A. P., and Garcia, A. A. F. (2015). Multi-trait multi-environment quantitative trait loci mapping for a sugarcane commercial cross provides insights on the inheritance of important traits. *Mol. Breed.* 35, 1–15. doi: 10.1007/s11032-015-0366-6
- Mehareb, E. M., and Abazied, S. R. (2017). Genetic variability of some promising sugarcane varieties (*Saccharum* spp.) under harvesting ages for juice quality traits, cane and sugar yield. *Open Access J. Agric. Res.* 2 (2), 1–14. doi: 10.23880/OAJAR-16000127
- Nei, M., and Li, W. H. (1979). Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proc. Nat. Aca. Sci.* 76 (10), 5269–5273. doi: 10.1073/pnas.76.10.5269
- Oliveira, K. M., Pinto, L. R., Marconi, T. G., Margarido, G. R., Pastina, M. M., Teixeira, L. H. M., et al. (2007). Functional integrated genetic linkage map based on EST-markers for a sugarcane (*Saccharum* spp.) commercial cross. *Mol. Breed.* 20, 189–208. doi: 10.1007/s11032-007-9082-1
- Page, R. D., and Holmes, E. C. (2009). *Molecular evolution: a phylogenetic approach*. (John Wiley & Sons).
- Palhares, A. C., Rodrigues-Morais, T. B., Van Sluys, M. A., Domingues, D. S., Maccheroni, W., Jordão, H., et al. (2012). A novel linkage map of sugarcane with evidence for clustering of retrotransposon-based markers. *BMC Genet.* 13 (1), 1–16. doi: 10.1186/1471-2156-13-51
- Pashley, C. H., Ellis, J. R., McCauley, D. E., and Burke, J. M. (2006). EST databases as a source for molecular markers: lessons from Helianthus. *J. Hered.* 97 (4), 381–388. doi: 10.1093/jhered/esl013
- Pastina, M. M., Malosetti, M., Gazaffi, R., Mollinari, M., Margarido, G. R. A., Oliveira, K. M., et al. (2012). A mixed model QTL analysis for sugarcane multiple-harvest-location trial data. *Theor. Appl. Genet.* 124, 835–849. doi: 10.1007/s00122-011-1748-8
- Peace, C., and Norelli, J. (2009). Genomics approaches to crop improvement in the Rosaceae. *Genet. Genomics Rosaceae* 6, 19–53. doi: 10.1007/978-0-387-77491-6_2
- Picañol, R., Eduardo, I., Aranzana, M. J., Howad, W., Batlle, I., Iglesias, I., et al. (2013). Combining linkage and association mapping to search for markers linked to the flat fruit character in peach. *Euphytica* 190, 279–288. doi: 10.1007/s10681-012-0844-4
- Pinto, L. R., Oliveira, K. M., Ulian, E. C., Garcia, A. A. F., and De Souza, A. P. (2004). Survey in the sugarcane expressed sequence tag database (SUCEST) for simple sequence repeats. *Genome* 47 (5), 795–804. doi: 10.1139/g04-055
- Qiu, L., Yang, C., Tian, B., Yang, J. B., and Liu, A. (2010). Exploiting EST databases for the development and characterization of EST-SSR markers in castor bean (*Ricinus communis* L.). *BMC Plant Biol.* 10, 1–10. doi: 10.1186/1471-2229-10-278
- Reffay, N., Jackson, P. A., Aitken, K. S., Hoarau, J. Y., D'Hont, A., Besse, P., et al. (2005). Characterisation of genome regions incorporated from an important wild

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2023.1273740/full#supplementary-material>

relative into Australian sugarcane. *Mol. Breed.* 15, 367–381. doi: 10.1007/s11032-004-7981-y

Saitou, N., and Nei, M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* 4 (4), 406–425. doi: 10.1093/oxfordjournals.molbev.a040454

Singh, R. K., Banerjee, N., Khan, M. S., Yadav, S., Kumar, S., Duttamajumder, S. K., et al. (2016). Identification of putative candidate genes for red rot resistance in sugarcane (*Saccharum species hybrid*) using LD-based association mapping. *Mol. Genet. Genomics* 291, 1363–1377. doi: 10.1007/s00438-016-1190-3

Singh, R. K., Singh, S. P., Tiwari, D. K., Srivastava, S., Singh, S. B., Sharma, M. L., et al. (2013). Genetic mapping and QTL analysis for sugar yield-related traits in sugarcane. *Euphytica* 191, 333–353. doi: 10.1007/s10681-012-0841-7

Srivastava, S., and Gupta, P. S. (2008). Inter simple sequence repeat profile as a genetic marker system in sugarcane. *Sugar Tech.* 10, 48–52. doi: 10.1007/s12355-008-0008-y

Ukoskit, K., Posudsavang, G., Pongsiripat, N., Chatwachirawong, P., Klomsa-Ard, P., Poomipant, P., et al. (2019). Detection and validation of EST-SSR markers associated with sugar-related traits in sugarcane using linkage and association mapping. *Genomics* 111 (1), 1–9. doi: 10.1016/j.ygeno.2018.03.019

Varshney, R. K., Graner, A., and Sorrells, M. E. (2005). Genic microsatellite markers in plants: features and applications. *Trends Biotech.* 23 (1), 48–55. doi: 10.1016/j.tibtech.2004.11.005

Weber, C. R., and Moorthy, B. R. (1952). Heritable and nonheritable relationships and variability of oil content and agronomic characters in the F2 generation of soybean crosses 1. *Agr. J.* 44 (4), 202–209. doi: 10.2134/agronj1952.00021962004400040010x

Wei, X., Jackson, P. A., Hermann, S., Kilian, A., Heller-Uszynska, K., and Deomano, E. (2010). Simultaneously accounting for population structure, genotype by environment interaction, and spatial variation in marker–trait associations in sugarcane. *Genome* 53 (11), 973–981. doi: 10.1139/G10-050

Wei, X., Jackson, P. A., McIntyre, C. L., Aitken, K. S., and Croft, B. (2006). Associations between DNA markers and resistance to diseases in sugarcane and effects of population substructure. *Theor. Appl. Genet.* 114, 155–164. doi: 10.1007/s00122-006-0418-8

Welham, S. J., Gogel, B. J., Smith, A. B., Thompson, R., and Cullis, B. R. (2010). A comparison of analysis methods for late-stage variety evaluation trials. *Aust. New Z. J. Stat* 52 (2), 125–149. doi: 10.1111/j.1467-842X.2010.00570.x

Wen, M., Wang, H., Xia, Z., Zou, M., Lu, C., and Wang, W. (2010). Development of EST-SSR and genomic-SSR markers to assess genetic diversity in *Jatropha Curcas L.* *BMC Res. Notes* 3 (1), 1–8. doi: 10.1186/1756-0500-3-42

Yang, X., Islam, M. S., Sood, S., Maya, S., Hanson, E. A., Comstock, J., et al. (2018). Identifying quantitative trait loci (QTLs) and developing diagnostic markers linked to orange rust resistance in sugarcane (*Saccharum spp.*). *Front. Plant Sci.* 9 350. doi: 10.3389/fpls.2018.00350

Yu, J., and Buckler, E. S. (2006). Genetic association mapping and genome organization of maize. *Curr. Opin. biotech.* 17 (2), 155–160. doi: 10.1016/j.copbio.2006.02.003

Yu, J., Pressoir, G., Briggs, W. H., Vroh Bi, I., Yamasaki, M., Doebley, J. F., et al. (2006). A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. *Nat. Genet.* 38 (2), 203–208. doi: 10.1038/ng1702

Yu, Y., Yuan, D., Liang, S., Li, X., Wang, X., Lin, Z., et al. (2011). Genome structure of cotton revealed by a genome-wide SSR genetic map constructed from a BC1 population between *Gossypium hirsutum* and *G. barbadense*. *BMC Genomics* 12 (1), 1–14. doi: 10.1186/1471-2164-12-15

Zhang, J., Zhang, X., Tang, H., Zhang, Q., Hua, X., Ma, X., et al. (2018). Allele-defined genome of the autopolyploid sugarcane *Saccharum spontaneum L.* *Nat. Genet.* 50 (11), 1565–1573. doi: 10.1038/s41588-018-0237-2

Zhao, K., Aranzana, M. J., Kim, S., Lister, C., Shindo, C., Tang, C., et al. (2007). An Arabidopsis example of association mapping in structured samples. *PLoS Genet.* 3 (1), e4. doi: 10.1371/journal.pgen.0030004