



OPEN ACCESS

EDITED BY

Collins Kimbeng,
Louisiana State University Agricultural Center,
United States

REVIEWED BY

Gurjeet Singh,
Texas A and M University, United States
Baoqing Zhang,
Guangxi Academy of Agricultural Science,
China

*CORRESPONDENCE

Sukhwinder Singh
✉ Sukhwinder.singh@usda.gov
Gul Shad Ali
✉ Gul.ali@usda.gov

RECEIVED 09 June 2025

ACCEPTED 19 September 2025

PUBLISHED 30 October 2025

CITATION

Eltaher S, DeStefano D, Abuasbeh Y,
Sandhu HS, Rouse M, Ali GS and Singh S
(2025) Genetic variation and preliminary
marker-trait associations for cane
quality traits in a diverse world
collection of sugarcane (*Saccharum*
spp.) and wild relatives.
Front. Plant Sci. 16:1643469.
doi: 10.3389/fpls.2025.1643469

COPYRIGHT

© 2025 Eltaher, DeStefano, Abuasbeh, Sandhu,
Rouse, Ali and Singh. This is an open-access
article distributed under the terms of the
[Creative Commons Attribution License \(CC BY\)](#).
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.

Genetic variation and preliminary marker-trait associations for cane quality traits in a diverse world collection of sugarcane (*Saccharum* spp.) and wild relatives

Shamseldeen Eltaher^{1,2}, Douglas DeStefano¹,
Yasmeen Abuasbeh³, Hardev S. Sandhu⁴, Matthew Rouse⁵,
Gul Shad Ali^{1*} and Sukhwinder Singh^{1*}

¹United States Department of Agriculture, Agricultural Research Service (USDA-ARS), Subtropical Horticulture Research Station (SHRS), Miami, FL, United States, ²Department of Plant Biotechnology, Genetic Engineering and Biotechnology Research Institute (GEBRI), University of Sadat City (USC), Sadat City, Egypt, ³Florida Sugar Cane League Inc., Clewiston, FL, United States, ⁴Institute of Food and Agricultural Sciences, Everglades Research and Education Center, University of Florida, Belle Glade, FL, United States, ⁵Sugarcane Field Station, United States Department of Agriculture, Agricultural Research Service (USDA-ARS), Canal Point, FL, United States

Introduction: Modern sugarcane cultivars originate from a limited genetic pool, primarily comprising *Saccharum spontaneum* and *S. officinarum*, which restricts yield improvements and stress resilience. It's vital to conserve and utilize genetic diversity from the World Collection of Sugarcane and Related Grasses (WCSRG) and is a key resource for future advancements.

Methods: The Cane Presentation System measures essential parameters, including Brix, Polarity, moisture, and fiber content, which are critical for evaluating sugar extraction potential and processing efficiency. Molecular markers hold significant value in sugarcane breeding, and genome-wide association studies have been conducted to identify genetic loci associated with these target traits.

Results: The study assessed variations in cane quality traits (Brix, polarity, fiber, and moisture content), finding that hybrids and *S. robustum* outperformed in these characters. Trait correlation analysis indicated independent genetic control, forming a basis for future research. Genome-wide association studies identified 40 significant SNPs across chromosomes 2 to 8, with markers on chromosomes 4, 6, and 7 consistently associated with Brix, and markers on chromosomes 2 and 5 linked to fiber. Additionally, marker AX-171243917-4651 on chromosome 6D was associated with both Brix and polarity.

Discussion: This study examines the genetic diversity and trait associations in sugarcane, emphasizing the application of stable SNP markers linked to key characteristics, such as Brix, Polarity, and fiber content, for enhancing multiple traits through marker-assisted selection. The findings reveal significant phenotypic variation among *Saccharum* species, advocating for a broad

genetic base in breeding programs. Due to sugarcane's polyploid nature, the study calls for further validation through fine mapping, gene expression analysis, and multi-location testing. Future research should enhance marker density, expand populations, and adopt new genomic approaches for effective crop improvement.

KEYWORDS

sugarcane germplasm, marker traits association, brix, qualitative trait loci, genetics

Introduction

Sugarcane (*Saccharum* spp.) is a high-yielding perennial grass cultivated mainly for its sucrose-containing stalks. The stalks are utilized in the production of sugar and bioethanol. A significant cash crop, nearly 80% of the world's sugar and 60% of the world's biofuel is produced from sugarcane (Dahlquist, 2013). The genus *Saccharum* consists of several species, including *S. officinarum*, *S. spontaneum*, *S. robustum*, *S. sinense*, and *S. barberi*. Among them, *S. officinarum* is a high-sugar containing species with large stalks, while *S. spontaneum* is valued for characteristics such as disease resistance and tolerance to a broad spectrum of environmental conditions (Ming et al., 2006). Modern sugarcane cultivars are highly polyploid and aneuploid hybrids with 100–130 chromosomes, which are products of interspecific hybridization among these species, predominantly *S. officinarum* ($2n = 80$, $x = 10$) and *S. spontaneum* ($2n = 40$ – 128 , $x = 8$) leading to complex genomic compositions that are difficult for genetic analysis and breeding (Hoarau et al., 2022). The initial interspecific F_1 hybrids were backcrossed extensively to *S. officinarum* clones or other hybrids to recover high sugar content, a process named “nobilization” (Daniels and Roach, 1987; Hoarau et al., 2022).

Modern sugarcane cultivars trace back to only a few *S. spontaneum* and *S. officinarum* ancestor clones (Deren, 1995; Lima et al., 2002; Yang et al., 2020). This narrow genetic base limited sugarcane cultivar improvement for yield and tolerance to various stresses. Therefore, conservation, characterization, and utilization of genetic diversity in sugarcane germplasm collections are essential to widen the genetic base for its improvement. The World Collection of Sugarcane and Related Grasses (WCSR), is a precious resource for future sugarcane cultivar improvement (Nayak et al., 2014; Todd et al., 2014; Yang et al., 2020). The collection comprises ~900 accessions from 45 countries, comprising *Saccharum* germplasm and related grass species, the most frequent of which are *S. spontaneum*, *S. officinarum*, and interspecific hybrids. These countries, however, do not represent the center of origin of the genus. The collection holds gene resources that can potentially be used to improve yield, fiber, and abiotic and biotic stress tolerance in sugarcane breeding programs. A core germplasm collection was developed by selecting representative

accessions from the WCSR (Nayak et al., 2014; Todd et al., 2014; Yang et al., 2020), which showed sufficient natural phenotypic diversity and are valuable resources for the identification of desirable alleles controlling yield and stress tolerance for sugarcane improvement.

The Cane Presentation System (CPS) measures key parameters for harvested sugarcane that reflect its potential to meet industrial standards for sugar extraction. CPS evaluates several important characteristics of cane quality, including Brix (total soluble solids, predominantly sucrose), Polarity (proportion of sucrose in the juice), moisture (which affects processing efficiency), and fiber (affects milling ability). Brix and Polarity are key indicators of sugar yield potential, and moisture and fiber content are significant parameters for processing efficiency and biomass conversion. A precise evaluation of these characteristics is essential for maximum sugar recovery and refining breeding selection criteria in breeding programs (Jackson, 2005). Variation in sucrose content between the different *Saccharum* species is extreme and assumes a key function in sugarcane breeding programs. *Saccharum officinarum*, also known as the “Noble cane,” contains high sucrose content with a high vulnerability to disease. This species constitutes approximately 80% of the genome of modern commercial cultivars and is the principal donor of genes responsible for sucrose accumulation (Dinesh Babu et al., 2022; Khan et al., 2023). Sucrose content in sugarcane exhibits considerable diversity in different tissues of the plant and varies with the progression of plant age. Major determinants, such as genetic constitution, environmental conditions, and developmental stage, have a significant influence on sucrose accumulation in sugarcane internodes. During the growth of the plant, every internode behaves independently, and the lower internodes reach maturity first in relation to the top ones, which are in the growing phase. The plant goes through a ripening process with age, where sucrose storage is enhanced in a greater number of internodes (de Moraes et al., 2015; Jackson, 2005; Khan et al., 2023).

Molecular markers have become useful tools in sugarcane breeding, significantly improving knowledge of complex genetics and assisting breeders in the genetic improvement of varieties. Their application is especially important in sugarcane breeding due to the crop's complicated genetic makeup, high ploidy levels (ranging

from 8 to 13 X), incidences of aneuploidy, variable flowering depending on environmental conditions, and a long growth period that usually exceeds a year (Wang et al., 2024). Single Nucleotide Polymorphisms (SNPs) have become the marker of choice for mapping and marker-assisted selection (MAS) in sugarcane breeding. This is not only because SNP markers are codominant, and highly abundant in the genome, but also because of the constant lowering of SNP genotyping cost and evolution of more efficient SNP genotyping techniques (Khanbo et al., 2023; Xiong et al., 2023). Over the last few years, hundreds of genetic loci and single nucleotide polymorphisms (SNPs) related to important agronomic traits in sugarcane have been uncovered using different methods, such as quantitative trait loci (QTL) mapping, gene mapping, and genome-wide association studies (GWAS) (Khanbo et al., 2023; Wang et al., 2024; Xiong et al., 2023). Genome-wide association studies (GWAS) have been utilized as a efficient tool for disassembling the genetic structure of sugarcane complex quality traits. GWAS has been particularly effective in mapping genetic variants associated with economically important traits such as Brix, polarity, moisture, fiber content, and components of yield that are critical for both sugar production as well as bioethanol conversion efficiency. Recent GWAS research has succeeded in identifying some marker-trait associations (MTAs) for significant quality traits, and studies have reported the identification of 23 MTAs for traits like soluble solid content, stalk height, stalk number, stalk weight, and cane yield (Barreto et al., 2019). Fiber and sucrose concentrations, being the two most significant quantitative traits deserving multi-year and multi-location testing, have been studied

widely using GWAS approaches for creating molecular breeding strategies like marker-assisted breeding and genomic selection (Xiong et al., 2023). Identified 13 and 9 markers associated with fiber and sucrose contents, using 237 self-pollinated progenies of LCP 85-384, a popular Louisiana cultivar. Also, the application of multi-locus mixed models in GWAS has boosted the detection ability to recognize substantial associations, and top-ranked markers explained up to 15% of the overall phenotypic variation for traits of sucrose concentration. The objective of this study was to investigate (1) phenotypic diversity for cane quality traits in a diverse collection of 397 *Saccharum* accessions, (2) genotypic diversity among this diverse set of genotypes using SNP markers, and (3) marker trait associations for important cane quality traits in this diverse set of accessions.

Materials and methods

Plant materials

The sugarcane accessions used in this study are maintained at the United States Department of Agriculture, Agricultural Research Service (USDA-ARS), Subtropical Horticulture Research Station (SHRS) in Miami, Florida. The number of accessions for each species, along with the country or institution from which the accession was acquired, is documented in the SHRS USDA-ARS (WCSRGS) sugarcane germplasm database, as illustrated in Figure 1. The list of all accessions, species, country or institution from which

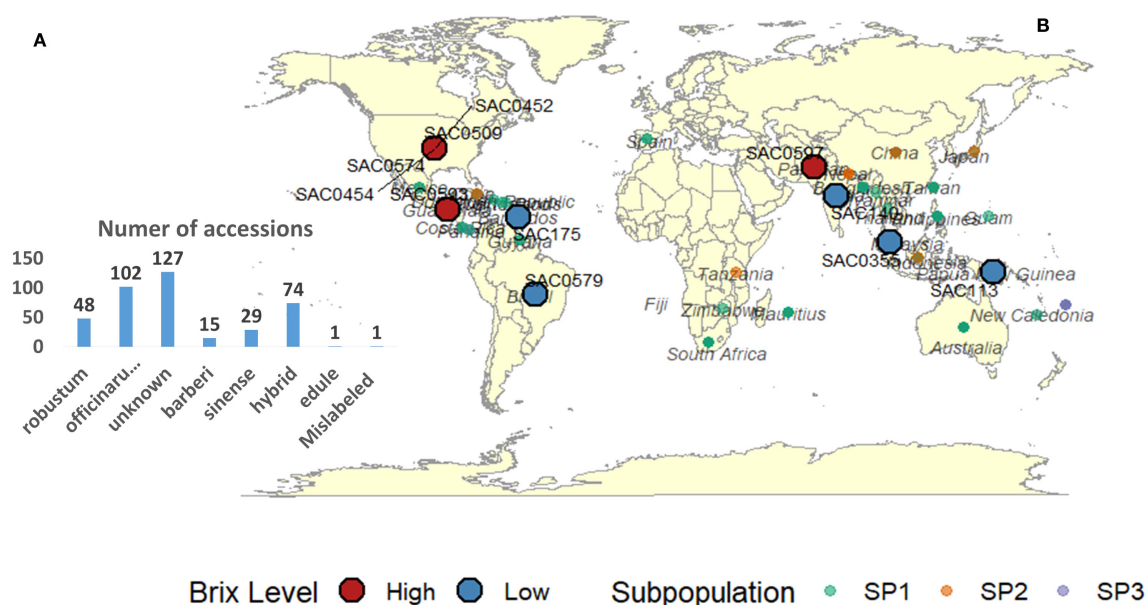


FIGURE 1

(A) Number of accessions per *Saccharum* species included in this study. (B) Country or institution from which each accession was obtained, as recorded in the SHRS USDA-ARS germplasm database. These records do not represent the true evolutionary or domestication origins of the 397 sugarcane accessions from the World Collection of Sugarcane and Related Grasses (WCSRGS); accessions with unknown origin were excluded. Colors indicate subpopulation assignments from STRUCTURE analysis: forest green (Subpopulation 1), light red (Subpopulation 2), and light blue (Subpopulation 3). Red circles mark accessions with the highest Brix values, and blue circles mark those with the lowest.

the accession was obtained, subpopulation based on STURCURE analysis and the values of 4 cane quality traits were presented in (Supplementary Table S1).

Sample collection

The cultivation of 397 sugarcane clones at the USDA-ARS Subtropical Horticulture Research Station (SHRS) in Miami, Florida, represents a significant step in our understanding of sugarcane genetics and cultivation. All accessions were planted as plant cane (first crop cycle) in single rows, following standard agronomic practices for the subtropical region of Miami. Ratoon crops were excluded from this study to reduce variability caused by environmental and physiological factors that disproportionately influence ratoon performance. Unlike plant crops, ratoon yields are highly sensitive to prior harvest conditions, soil nutrient depletion, pest pressure, and age-related decline in vigor. Including ratoon data would have confounded genotype comparisons and reduced the reliability of trait assessments across accessions. To ensure uniformity and maximize the precision of phenotypic evaluations, only plant crop data were considered (Dlamini and Zhou, 2024; Xu et al., 2021). The sugarcane stalks collected over a period spanning from June 2024 to December 2024 in five distinct batches, corresponding to the physiological maturity of each accession rather than a fixed calendar schedule. Due to the genotypic differences in growth rates and accessions' diversity, physiological maturity was established by consistent visual and physical symptoms, including internode swelling, hardening of the rind, and yellowing of green leaves. These have been widely accepted as effective markers for sucrose accumulation and harvest preparedness in sugarcane, particularly where flowering is nil or irregular. After the removal of leaves and tops, six randomly selected stalks from each sugarcane variety were meticulously processed in the cane presentation systems (CPS). The weight of each bundle was recorded before the samples were analyzed for sucrose content using the CPS near-infrared analysis system (Bruker). The samples were then shredded with a Dedini shredder, and the resulting shredded material was processed through CPS to quantify Brix (%), juice Polarity (%), and total fiber content (%) (Legendre, 1992).

It must be understood that this study was achieved in the form of a non-replicated field trial. Replicated multi-year trials involving all 397 accessions were not possible owing due to limitations in land availability, plant materials, and finances at the Miami station, and high cost and space requirements of holding large germplasm collections. Instead, our aim was to provide extensive overview of genetic diversity and make preliminary marker–trait associations in this unique germplasm panel. These non-replicated designs have been applied to sugarcane germplasm characterization at the discovery phase with the awareness that subsequent multi-environment and replicated tests are necessary to validate marker–trait associations. Therefore, we explicitly acknowledge the limitations and have reported our results as exploratory association and diversity analyses, providing valuable hypotheses

and data for breeders in addition to apparent directions for future confirmation studies.

DNA isolation and SNP calling

As described Park et al. (2024), DNA was isolated from young harvested leaf from each sugarcane accession using BioArk leaf Kit supplied by LGC, Biosearch Technologies (<https://www.biosearchtech.com/>). A total of 2,000 SNPs were randomly selected from the Axiom Sugarcane 100K SNP array. Based on minor allele frequency (MAF) and linkage disequilibrium (LD) and 400 SNP were ultimately retained. Another filtering was performed using different criteria such as (1) retaining biallelic SNPs, (2) removal of SNPs with >10% missing values and (3) removal of SNPs with < 5% MAF. Finally, 357 high quality SNPs across all sugarcane accessions were used in this study. The sequences of each marker of the 357 SNPs are presented in Supplementary Table S2.

The chromosomal positions of the 357 SNP markers were established by aligning their sequences with the reference genome of *Saccharum spontaneum* (Nascimento et al., 2019). This species is utilized as a model for sugarcane due to its relatively high-quality genome assembly and the significant contribution of its ancestral genomes to contemporary sugarcane varieties (J. Zhang et al., 2018). *Saccharum spontaneum* comprises eight chromosomes (chr1 to chr8), and its entire genome is organized into sets of homeologous chromosomes identified as A, B, and D. These sets reflect the polyploid and aneuploid nature of sugarcane, a key aspect of the species' genetic makeup.

The statistical analysis

A one-way analysis of variance (ANOVA) was conducted to estimate the genetic variability among the sugarcane accessions for the cane quality traits (Brix, polarity, moisture and fiber). Since our data included unreplicated genotypic observation, the analysis was done based on the genotype values of each trait and the genetic variability was tested among the subpopulation resulted from the structure finding. The model used for this analysis is expressed as follows.

$$F = \frac{\sum [n_i(\mu_i - \mu)^2]}{K - 1} \div \frac{\sum \sum (x_{ij} - \mu_i)^2}{n - K}$$

where F is the calculated F-statistic, Σ is summation, n_i = sample size in subpopulation I, μ = overall mean, μ_i = mean of group I, x_{ij} = j-th observation in subpopulation I, K = number of subpopulation and n = total number of observations.

Pearson's correlation analyses were performed among these traits, and the resultant correlation coefficients (r values) were calculated using the "cormat" function in R-4.4.1 (Team, R. C et al., 2016) software and the upper triangle heatmap was visualized using the "ggplot2" package (Wickham, 2016). The genetic distance analysis was performed using PowerMarker software V 3.25 (Liu

and Muse, 2005) and the dendrogram cluster was visualized using MEGA 11 (Tamura et al., 2021).

Analysis of population structure

A model-based (Bayesian) method with the 357 SNPs was utilized to evaluate the possible number of subpopulations in the accessions used in this study. The analysis of population structure was performed using STRUCTURE 3.4.0 (Pritchard et al., 2000). Structure was analyzed by means of k-values (an assumed fixed number of subpopulations K) from 1 to 10 in each K. For each K, three independent runs were performed using an admixture ancestry model with correlated allele frequencies. Each run included 50,000 burn-in iterations followed by 100,000 Markov Chain Monte Carlo (MCMC) replications and sampling frequency every 1,000 reps. These parameters were chosen based on the protocol of (Park et al., 2024). The optimal number of model components (K) was determined based on delta k (Evanno et al., 2005). The best k for the current population was determined using STRUCTURE SELECTOR (Li and Liu, 2018). The analysis of molecular variance (AMOVA) and the calculation of fixation index were conducted as described before (Peakall and Smouse, 2006). All allelic diversity pattern including number of private allele (PAL), number of different allele (Na), number of effective allele (Ne), number of common allele (N_{com}) and the Shannon information diversity index (H) were calculated using R 4.41 (Team, R. C et al., 2016).

Genome-wide association study for cane quality traits

Genome-wide association studies (GWAS) were conducted for all traits utilizing the rMVP R package (Yin et al., 2021). This analysis employed three distinct GWAS models: (1) the Fixed and Random Model Circulation Probability Unification (FarmCPU), (2) the Generalized Linear Model (GLM), and (3) the Mixed Linear Model (MLM). To address population structuring and kinship (Kin), principal component analysis (PCA) and PCA combined with Kin were incorporated into FarmCPU and MLM each individually. FarmCPU integrates the advantages of mixed linear models and stepwise regression (fixed effect models), employing them in an iterative manner to adjust for balancing population structure and overfitting the model (Liu X. et al., 2016).

Results

The descriptive analysis for cane quality traits

Polarity, Brix, fiber and moisture analyses were conducted using CPS near infrared to evaluate sugarcane accessions. A summary of these analyses is presented in Table 1. The Brix values ranged from 1.13 to 21.96 with a mean of 14.83 and a standard deviation of 3.32. Polarity had a mean of 46.07 and SD = 16.00, with values ranging from 2.42 to 83.54. The mean moisture content was 61.38 and SD = 19.20, and the median was 69.77, indicating a left-skewed distribution because the mean was lower than the median. Fiber showed moderate variability, with a mean of 17.23, SD = 4.92 and a range of 2.22 to 42.03.

The distributions of various traits within the germplasm collection provide significant insights into their variability, as illustrated in Figure 2. The histogram for Brix and Polarity reveals a bell-shaped normal distribution, indicating that most sugarcane samples exhibit moderate to high sugar content and a wide variation in sugar levels. The germplasm is generally well-suited for sugar production, as most samples exhibit moderate Brix values, with only a few showing extremely low or high levels. Conversely, the moisture histogram illustrates a left-skewed distribution, reflecting a predominance of samples with elevated moisture content.

Figure 3 illustrates box plots that depict the distribution of four essential cane quality traits (Brix, Fiber, Moisture, and Polarity) across six accessions of *Saccharum* species, after excluding one incorrectly labeled genotype and one *S. edule* species from the dataset. The data uncovers significant patterns in trait variability that are relevant for sugarcane breeding and processing applications. Upon examining the distributions, the Brix measurements indicate relatively stable performance across most species, with medians clustering around 15–17 units and comparable interquartile ranges. However, contrary to initial expectations, the hybrid species do not consistently exhibit the highest values across all traits. In terms of fiber content, *S. robustum* stands out with the highest median values, approximately 25 units, which significantly exceeds other species that range between 15–20 units. The moisture content displays the most distinct variation, with *S. officinarum* showing an exceptionally wide distribution from nearly 0 to about 80 units, while other species maintain more uniform levels around 65–75 units. Polarity measurements reveal moderate variation among species, with most distributions centered

TABLE 1 Descriptive statistics for the cane traits of *Saccharum* species, including brix, polarity, moisture, and fiber content.

Cane Quality Traits	Mean	SD*	Median	Min*	Max*	Range	SE*
Brix	14.83	3.32	15.42	1.13	21.96	20.83	0.16
Polarity	46.07	16.00	47.9	2.42	83.54	81.12	0.80
Moisture	61.38	19.20	69.77	7.14	79.8	72.66	0.96
Fiber	17.23	4.91	16.76	2.22	42.03	39.81	0.24

*SD, stander deviation; min, minimum; max, maximum and SE, stander error.

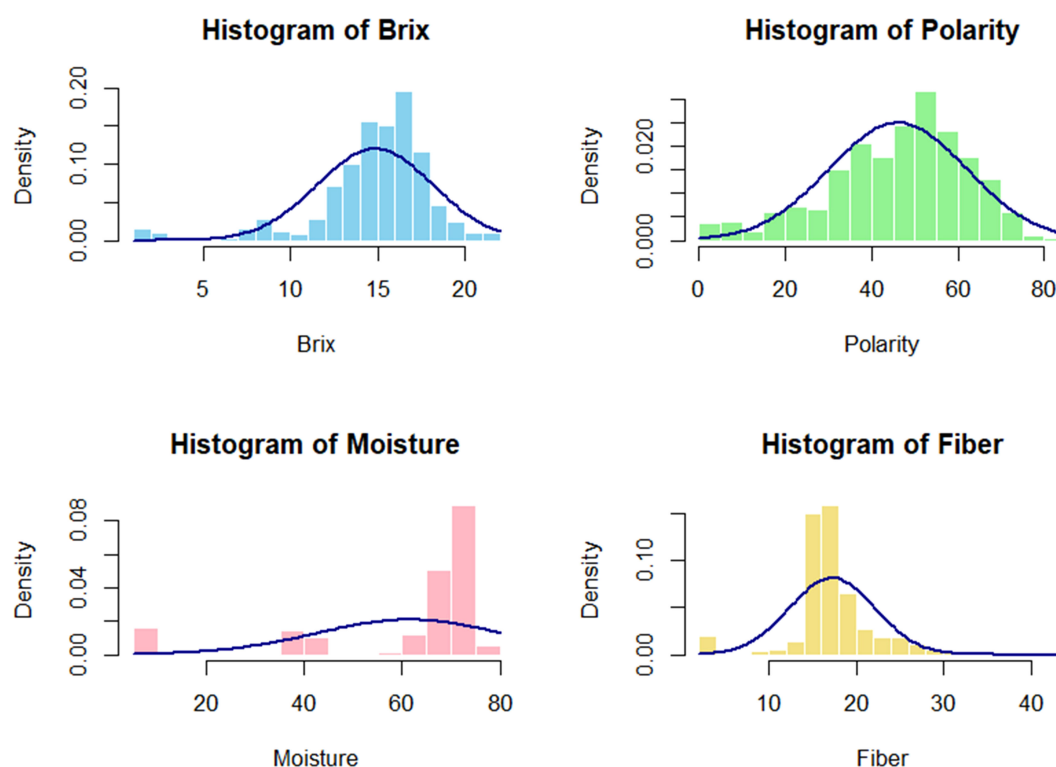


FIGURE 2

The phenotypic distribution of cane quality traits in the 397 sugarcane accessions exhibited a visually normal frequency distribution for all characteristics, with some showing skewness to the right and left.

between 40–60 units, yet exhibiting different spreads and notable outliers. These trait distributions highlight considerable variability both within and among species, offering valuable insights for breeding programs aimed at enhancing agricultural practices and formulating processing strategies. The intricate patterns observed suggest that various species may provide unique advantages for specific applications, with *S. robustum* demonstrating potential for high-fiber applications and the necessity for careful selection within the *officinarum* group due to its significant moisture variability.

Correlation coefficient analysis

The Pearson correlation matrix illustrates the interrelationships among Brix, polarity, moisture, and fiber as depicted in Figure 4. A moderate positive correlation ($r = 0.39^{***}$) indicates that an increase in Brix is associated with a slight increase in Polarity. On the other hand, a weak negative and non-significant correlation was observed between the remaining traits.

Population structure and relationships

The population structure of 397 sugarcane genotypes was examined with the STRUCTURE analysis software (Figure 5). To determine the fitting value of the best K, the number of clusters (K) was plotted against ΔK , revealing a strong peak at $k = 3$ (Figure 5A).

However, the log likelihood [$\ln P(D)$] increased continuously and gradually as K increased (Figure 5B). The optimal K value was 3, suggesting that the 397 sugarcane genotypes can be grouped into three subpopulations (SP). With 292 belonging to SP1, 53 to SP2, and 52 to SP3 subgroups (Figure 5C). The results of PCA were consistent with STRUCTURE analyses (Figure 5D).

Analysis of variance among subpopulations identified by population structure

The one-way ANOVA results are presented in Table 2, revealing significant differences among subpopulations defined by the population structure analysis. Polarity and fiber content showed highly significant variation ($F = 17.27$, $p = 3.97 \times 10^{-5}$; and $F = 34.73$, $p = 8.12 \times 10^{-9}$, respectively). In contrast, Brix and moisture content did not differ significantly among subpopulations ($F = 2.208$, $p = 0.138$; and $F = 0.949$, $p = 0.33$, respectively), indicating that these traits were relatively consistent across groups.

Comprehensive assessment of genetic diversity and differentiation in the WCSRG collection

Subpopulations showed considerable genetic divergence, and the average distance (expected heterozygosity) of each

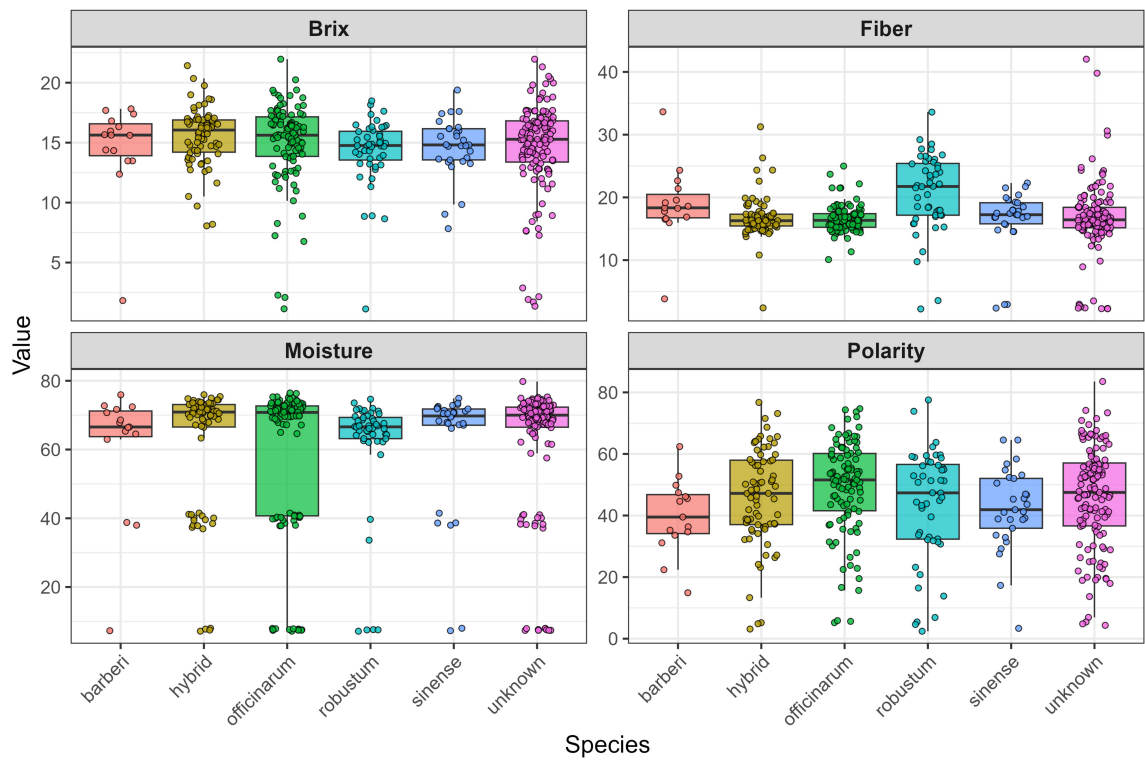


FIGURE 3 Box plots illustrating the distribution of Brix, Polarity, Moisture, and Fiber for *Saccharum* ssp. Accessions after removing one genotype mislabeled and one *S. edule* species.

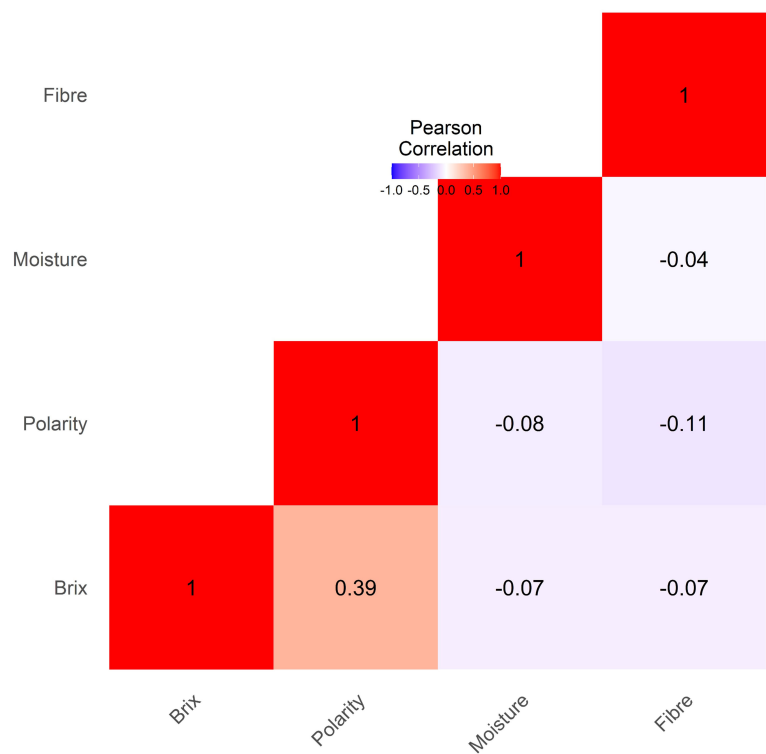
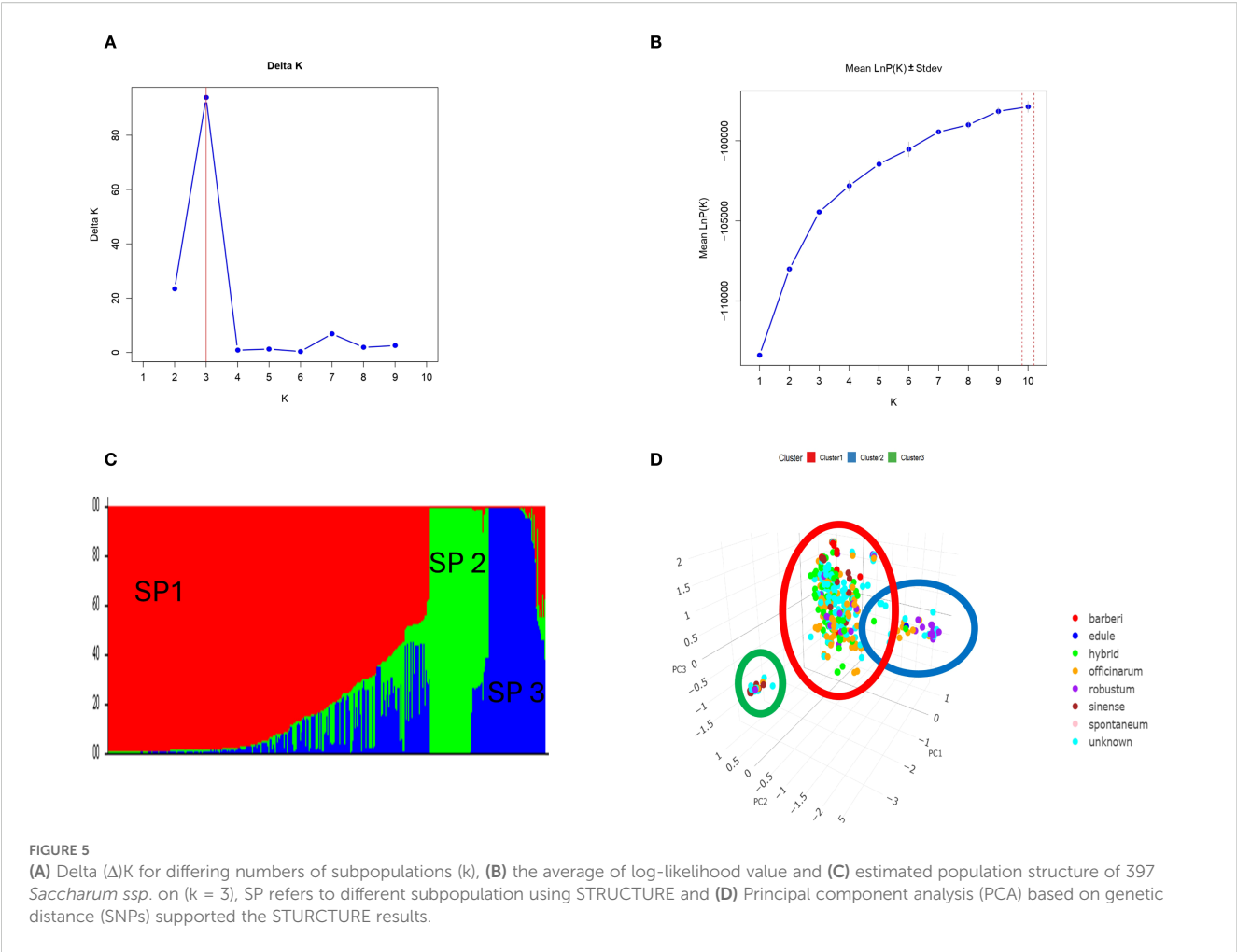


FIGURE 4 Phenotypic correlation analysis among the cane traits (Brix, Polarity, Moisture and Fiber) showed the high correlation between Brix and Polarity.



subpopulation is shown in Table 3. With an average of 0.26, SP1 had the highest expected heterozygosity, while SP2 had moderate value at 0.21 and SP3 had the lowest at 0.11. To examine genetic differentiation due to population substructure, the Fixation index (F_{st}) was computed, which is good indicator for evaluating total genetic variation among subpopulations. The F_{st} values for SP1, SP3, and SP3 were 0.06, 0.46, and 0.56, respectively. Utilizing the

three subpopulations found in the STRUCTURE study, the AMOVA, and number of migrant (N_m) were computed, which are presented in Table 4. Approximately 99% of the variation was observed across individuals with only 1% occurring among subpopulations. The high haploid N_m value (29.66) suggests significant historical gene exchange and genetic connectivity among species or groups, based on allele frequency patterns.

TABLE 2 The analysis of variance (ANOVA) conducted among subpopulations identified by population structure for the cane quality traits of Brix, polarity, moisture, and fiber.

Trait	Source of Variance	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Brix	Subpopulation	2	24	24.38	2.208	0.138ns
	Residuals	394	4361	11.04		
Polarity	Subpopulation	2	4252	4252	17.27	3.97E-05 ***
	Residuals	394	97224	246		
Moisture	Subpopulation	2	350	350	0.949	0.33ns
	Residuals	394	145634	368.7		
Fiber	Subpopulation	2	774	774.3	34.73	8.12E-09***
	Residuals	394	8807	22.3		

ns refer to nonsignificant and *** Significant at p -value < 0.001.

TABLE 3 The STRUCTURE analysis results of 397 sugarcane accessions included the fixation index (F_{st}), average distances (expected heterozygosity), and the number of genotypes in each subpopulation.

Subpopulation	Fst ¹	Ex. Hetro ²	No. of genotypes
SP1	0.06	0.26	292
SP2	0.46	0.21	53
SP3	0.56	0.11	52

¹ F_{st} is a measure of genetic differentiation among three subpopulations: ²Expected Heterozygosity.

These findings showed that there was considerable genetic differentiation within subpopulations and low genetic differentiation between subpopulations.

Allelic diversity pattern across *Saccharum* species

To further evaluate species-specific genetic contributions in the sugarcane germplasm panel, we calculated some parameters of allelic diversity, including the number of private alleles (PAL), number of different alleles (Na), number of common alleles (N_{com}), number of effective alleles (Ne), and Shannon's diversity index (H). Among the species, *S. robustum*, *S. officinarum*, and *S. barberi* possessed the maximum number of private alleles (3, 2, and 3, respectively) followed by hybrid and unknown types, each of which carried very few private alleles. These species also indicated higher average values of Na, Ne, and H, which indicate a broader allelic range. On the other hand, *S. sinense*, *S. edule*, and the mislabeled accession which had no private allele and had the least allelic diversity ($Na = 1$, $Ne = 1$, $H \approx 0$). These findings provide a quantitative overview of the genetic distinctness and diversity retained in each species group. The values for the complete allelic diversity pattern are given in Table 5. The WCSRG germplasm collection, which includes accessions obtained from 36 countries crossing diverse geographical regions across the globe, reflects this complexity. The broad representation enhances our ability to detect meaningful patterns of genetic differentiation and provides a strong foundation for future breeding strategies aimed at improving sugarcane performance under varied environmental conditions.

TABLE 4 Analysis of molecular variance (AMOVA) for 397 sugarcane accessions grouped into three clusters.

Source	df	SS	MS	Est. Var.	%
Among Subpopulations	2	393.787	196.894	0.485	1%
Among Individuals	394	45379.996	115.178	57.589	99%
Within Individuals	397	0.000	0.000	0.000	0%
Total	793	45773.783		58.074	100%
Nm		29.66			

Nm refers to gene exchange rate or number haploid migrants .

TABLE 5 Mean of allelic diversity pattern including number of private allele (PAL), number of different alleles (Na), number of common allele (N_{com}), number of effective allele (Ne) and Shannon's index (H), in each *Saccharum* species.

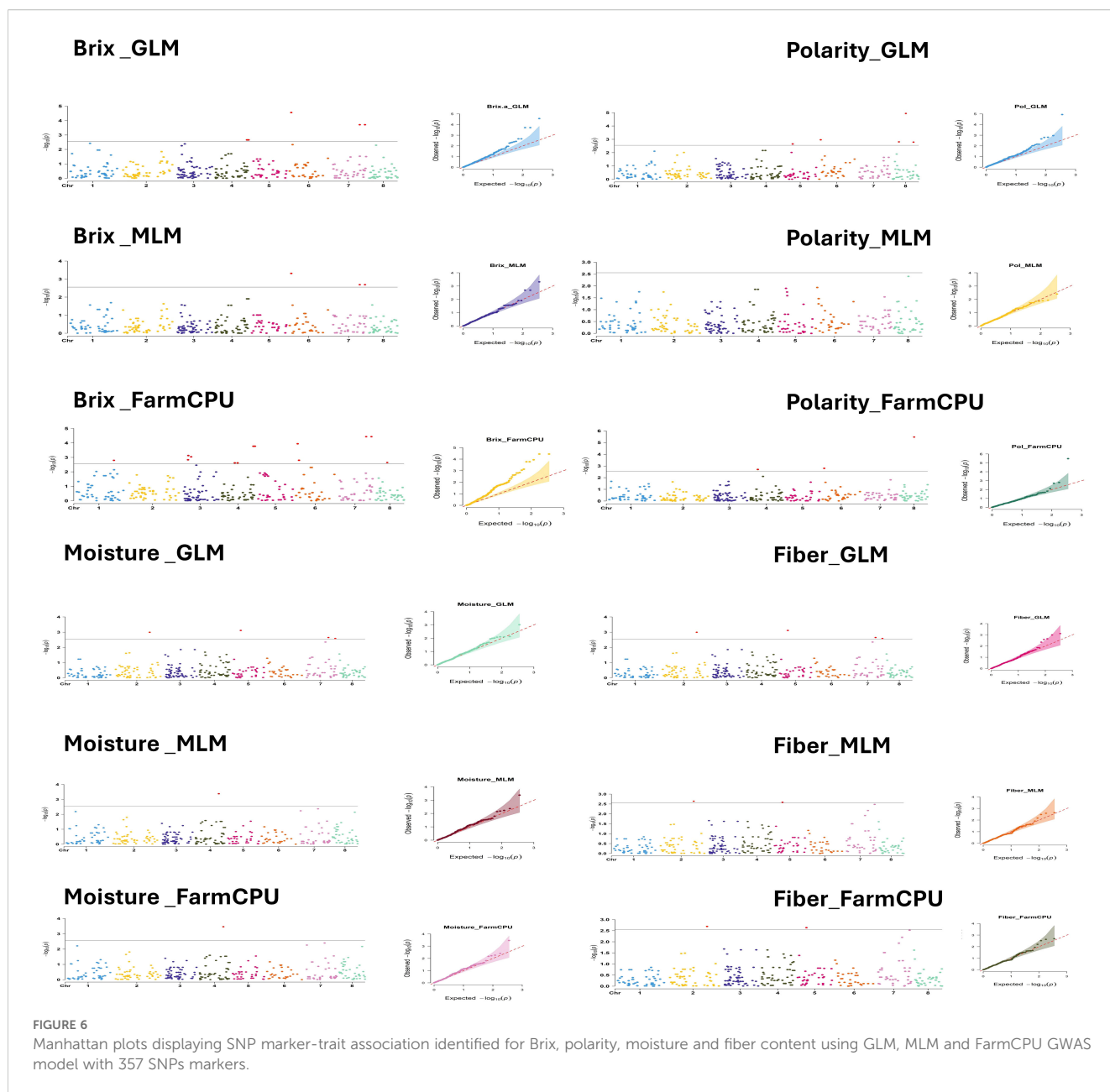
Species	PAL	Na	N_{com}	Ne	H
<i>officinarum</i>	2	2.02	2.02	1.52	0.49
<i>unknown</i>	2	2.04	2.04	1.52	0.49
<i>hybrid</i>	1	1.99	1.99	1.54	0.48
<i>robustum</i>	3	2.01	2.01	1.40	0.40
<i>barberi</i>	3	1.79	1.79	1.38	0.36
<i>sinense</i>	0	1.79	1.79	1.32	0.33
<i>edule</i>	0	1	1	1	-1E-10
Mislabeled	0	1	1	1	-1E-10

Genome-wide association study of cane traits

Marker-trait associations were investigated using three different GWAS models with 357 SNP markers. These analyses revealed 15, 6 and 19 significant SNPs markers using GLM, MLM and FarmCPU models, respectively (Supplementary Table S3). Significant markers were associated with Brix, fiber and moisture in GLM, MLM and FarmCPU. Significant markers were found for Polarity in two of the three models. The Manhattan and Q-Q plots for the significant traits of Brix, polarity, moisture and fiber are illustrated in Figure 6. For Brix, 21 markers were found significant in all GWAS models. All markers were in chromosomes 3, 4, 6 and 7. Three markers (AX-171243917-4651; chr. 6D, AX-171305424-5045; chr.7B and AX-171305424-5356; chr. 7D) were found common in the three GWAS models and two markers (AX-171270569-3055; chr.4B and AX-171270569-3274; chr.4C) were supported with GLM and MLM. The GLM and FarmCPU also identified 8 significant markers linked to polarity, distributed across chromosomes 4, 5, 6 and 8. Two markers (AX-171243917-4651; chr. 6D and AX-171247985-5866; chr. 8D) were confirmed with GLM and FarmCPU. One marker AX-171269335-3422; chr.4D was found significantly associated with moisture and supported with the three GWAS models. Eight markers were significantly associated with the fiber content and all these markers resided in three different chromosomes 2, 5 and 7. Two markers (AX-171287224-1157; chr. 2A and AX-171363600-3972; chr. 5D) were validated with three GWAS models. All the common markers in the GWAS models are presented in Supplementary Table S4.

Allelic effects and their relationship to known sucrose QTLs

To additionally validate the associations uncovered by the GWAS, we examined the allelic effects of five top-ranked SNPs on Brix values (Figure 7; Table 6). Significant associations were identified between Brix content and several SNP loci. For AX-

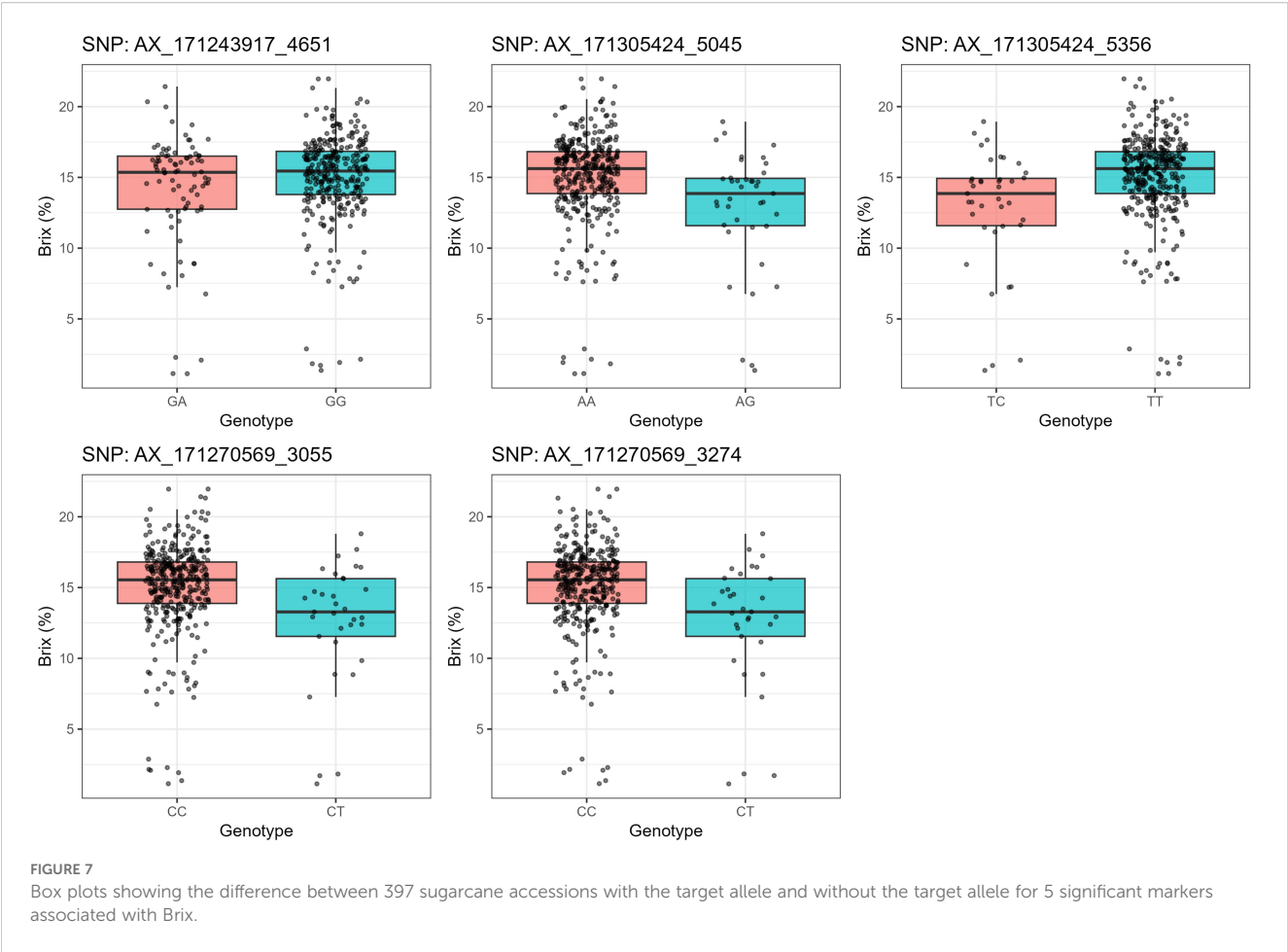


171243917-4651 (Chr6D:4287042–4287343), the GA genotype had a mean Brix of 14.15 ± 0.43 (SE) compared with 15.03 ± 0.18 (SE) for GG ($p = 0.0296$). At AX-171305424-5045 (Chr7B:80181102–80181403), the AA genotype showed a higher mean Brix of 15.06 ± 0.17 compared with 12.76 ± 0.68 for AG ($p = 3.26 \times 10^{-5}$). A similar pattern was observed for AX-171305424-5356 (Chr7D:67813880–67814181), where TT exhibited a higher Brix (15.06 ± 0.17) relative to TC (12.76 ± 0.68) with the same significance level ($p = 3.26 \times 10^{-5}$). For AX-171270569-3055 (Chr4B:75471756–75472057), the CC genotype had a mean Brix of 15.05 ± 0.16 compared with 12.55 ± 0.76 for CT ($p = 3.06 \times 10^{-5}$). Likewise, AX-171270569-3274 (Chr4C:79075825–79076126) showed higher Brix in CC (15.05 ± 0.16) compared with CT (12.55 ± 0.76), with a similar level of significance ($p = 3.06 \times 10^{-5}$). Overall, these results indicate that favorable homozygous genotypes consistently conferred higher Brix

values than their heterozygous counterparts, suggesting strong marker–trait associations for sugar content in sugarcane.

Identification of promising accessions for sugarcane breeding programs based on brix content

Selection of the 10 highest and 10 lowest sugarcane accessions, based on Brix values, holds great promise for sugarcane breeding programs. **Figure 1**, depicting the global distribution map of the selected sugarcane accessions, is a testament to this potential due to most high Brix values accessions belonging to tropical and subtropical regions. The Brix values for the highest group ranged from 19.91 to 21.96, while the lowest group exhibited values



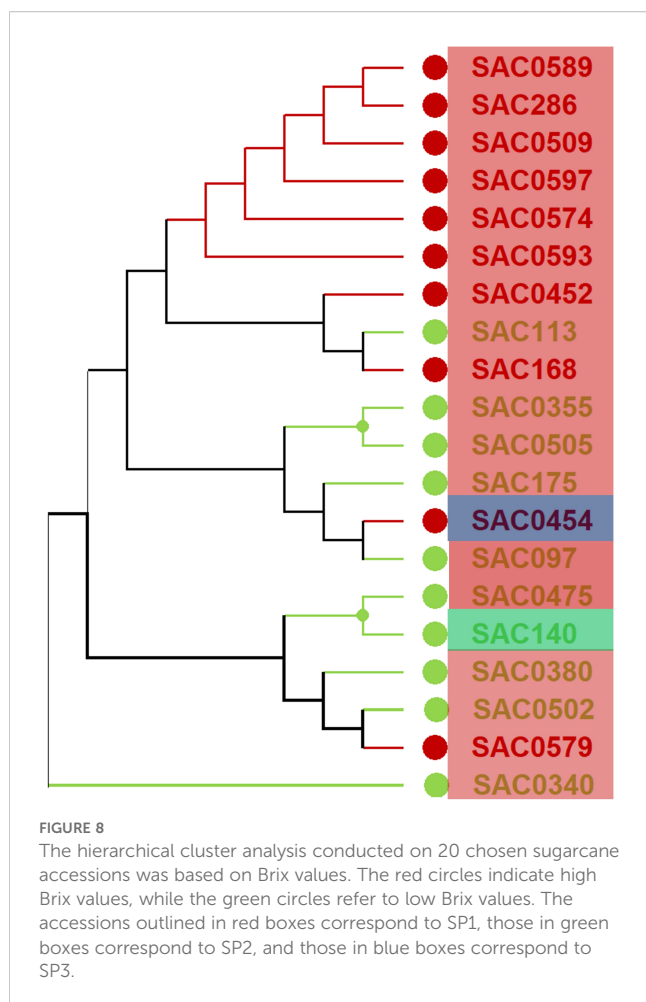
between 1.13 and 2.88. These extremes underscore the significant phenotypic diversity within the WCSRG germplasm, a diversity that we are poised to harness for the betterment of sugarcane breeding programs (Supplementary Table S5).

Among the accessions with the highest Brix values, two notable clones are “SAC0454” (belonging to *S. officinarum*, of US, SP3) and “SAC0168” (considered an unknown in both species and country,

SP1), both of which recorded a Brix content of 21.96. These accessions represent ideal candidates for direct selection or use as parental lines in hybridization programs. Moreover, noteworthy Brix values of 21.42 and 20.35 were recorded in “SAC0593” (SP1, a hybrid from Guatemala) and “SAC0452” (a hybrid from the US), respectively. This observation indicates that accessions with hybrid backgrounds, particularly those from Guatemala and US breeding

TABLE 6 Genotypic grouping and statistical association of significant SNPs with Brix content across homozygous and heterozygous alleles.

Significant SNP	Genotype	Mean_Brix	SD_Brix	SE_Brix	No. of genotypes	P_value
AX-171243917-4651: Chr6D:4287042-4287343	GA	14.15	3.99	0.43	86	0.0296
	GG	15.03	3.08	0.18	309	0.0296
AX-171305424-5045: Chr7B:80181102-80181403	AA	15.06	3.11	0.17	356	3.26E-05
	AG	12.76	4.27	0.68	39	3.26E-05
AX-171305424-5356: Chr7D:67813880-67814181	TC	12.76	4.27	0.68	39	3.26E-05
	TT	15.06	3.11	0.17	356	3.26E-05
AX-171270569-3055: Chr4B:75471756-75472057	CC	15.05	3.12	0.16	362	3.06E-05
	CT	12.55	4.38	0.76	33	3.06E-05
AX-171270569-3274: Chr4C:79075825-79076126	CC	15.05	3.12	0.16	362	3.06E-05
	CT	12.55	4.38	0.76	33	3.06E-05



programs, may possess allelic advantages for high sugar accumulation due to their selection history and adaptation to tropical and subtropical environments. Notably, eight of the 10 highest Brix accessions were classified under SP1, suggesting that this group is particularly enriched with beneficial alleles for high sucrose content. Additionally, six top-performing accessions were categorized as unknown species, which raises the possibility that they may represent hybrids or belong to *S. officinarum*. Hybrid backgrounds may indicate genetic exchange with *S. officinarum*, *S. spontaneum*, or *S. robustum*, potentially contributing to improved Brix values.

Conversely, the lowest Brix accessions, specifically “SAC0175” (belonging to *S. robustum*, from Barbados, with a Brix of 1.13) and “SAC0113” (belonging to *S. officinarum*, from Papua New Guinea, with a Brix of 1.14), were also found within SP1. These accessions may represent unselected or ancestral lines with reduced sucrose accumulation capabilities. Nevertheless, low Brix content accessions could retain value for traits such as stress resistance, biomass production, and disease resistance qualities that are typically associated with wild *Saccharum* species preserved in ancestral collections. Importantly, the countries or institutions listed for each accession refer not to their geographic origin, but to the source from which the accession was obtained, as recorded in the

Subtropical Horticulture Research Station (SHRS) at the USDA-ARS germplasm database.

Utilizing the selected accessions characterized by high and low Brix values is a key strategy to enhance the genetic diversity of the WCSRG germplasm (Supplementary Table S6). An analysis of the genetic distance among these chosen accessions was conducted, and the resultant dendrogram cluster is illustrated in Figure 8. The selected genotypes encompass three distinct subpopulations, with genetic distances among these accessions ranging from 0.07 to 0.12. Such low genetic distances suggest a high degree of genomic similarity, likely reflecting a shared breeding history or common genetic background among the accessions. By evaluating the dendrogram cluster, it is possible to identify the most suitable parent plants for sugarcane crossbreeding. The optimal parent combination, as determined by distinct subpopulations, Brix values, and significant genetic distance, is (“SAC0454”, *S. officinarum*, US, SP3 with a Brix of 21.96) crossed with (“SAC0175”, *S. robustum*, SP1, Barbados, with a Brix of 1.13). This crossbreeding has the potential to generate F1 offspring that exhibit high sucrose content inherited from *S. officinarum*, alongside resistance traits conferred by the wild species *S. robustum*, both of which are well-adapted to tropical and subtropical environments.

Discussion

Phenotypic variation and trait distributions

A comprehensive phenotypic evaluation of cane quality traits (Brix, Polarity, Moisture, and Fiber) in 397 sugarcane genotypes revealed substantial variability, as evidenced by broad trait ranges, high standard deviations, and distinct distribution patterns within and among species and subpopulations. This variation could potentially be valuable for sugarcane improvement, particularly in breeding programs aimed at enhancing sugar yield and quality. Descriptive statistics indicated that Brix and Polarity exhibited intermediate to high mean values, representing an overall positive sugar content profile of the germplasm. The broad range in the values of Brix measured (1.13–21.96) was also noted by (Aitken et al., 2006), who also noted the same ranges when they performed a survey of various accessions of sugarcane. The broad range measured on Polarity (2.42–83.54) and Fiber (2.22–42.03) agrees with what (Zhou and Gwata, 2015) reported as considerable variation among collections of sugarcane germplasm. Normal distribution of Brix and Polarity indicates that most of the genotypes have moderate to high sugar and therefore are ideal for cane quality traits and other industrial purposes (Jackson, 2005; Srinivasan et al., 2021; Todd et al., 2014). A comparison of trait variation among species reveals Hybrid and *S. robustum* genotypes with high sugar and fiber content and superior agronomic potential. These findings are consistent with previous reports that recognized interspecific hybrids as elite germplasm based on their increased biomass and sucrose yield (Srinivasan et al., 2021). Conversely, species like *S. barberi* and *S. edule* recorded lower values for sugar-

related traits in this study; however, this does not necessarily reflect their overall breeding value, as they may possess unique alleles contributing to stress tolerance, disease resistance, or other agronomic traits not assessed here. Variance analysis indicated that there were significant differences among subpopulations for Polarity and fiber, but not for Brix or moisture. The correlation analysis revealed a moderate positive association between Polarity and Brix, which aligns with their shared involvement in sugar metabolism and accumulation within the stalk. However, the generally weak correlations observed among the other traits suggest complex genetic regulation, likely involving multiple loci with small effects. This observation is consistent with the polygenic nature of quantitative traits in sugarcane, where traits such as sucrose content, fiber, and moisture are influenced by diverse genomic regions, potentially with interacting or pleiotropic effects (Bhatt et al., 2025; Liu P. et al., 2016; Perlo et al., 2020).

Population structure and genetic differentiation

STRUCTURE analysis of 397 sugarcane genotypes identified three distinct subpopulations (SP1, SP2, SP3), a finding verified by principal component analysis (PCA). Notably, SP1 included most genotypes, indicating a broad genetic base, while SP2 and SP3 were specialized subpopulations. The grouping agrees with the observations in a previous study where sugarcane germplasm exhibited clear genetic stratification by species and countries and institutions was obtained (Xiong et al., 2022). In contrast to the findings of, Saavedra-Díaz et al. (2024) who revealed four subpopulations in 220 genotypes from the Cenicaña's diverse panel, findings of this study indicate population structure can be very different based on geographical distribution and breeding history. Park et al. (2024) also revealed two distinct subgroups within the *S. spontaneum* cluster, while the remaining species formed a separate, unified cluster. To examine the substructure within these two clusters, they conducted a STRUCTURE analysis of each cluster using the same procedure and parameters. Hierarchical analysis revealed that two were the most likely number of genetically differentiated groups (K) for the *S. spontaneum* cluster, and four for the other species.

The highest expected value of heterozygosity was in SP1 at 0.26, revealing a higher rate of genetic diversity than in SP2 at 0.21 and SP3 at 0.11. The genetic diversity of variation among the subpopulations agrees with Wei et al. (2010) who calculated the same heterogeneity rate in sugarcane germplasm. The fixation Index (F_{st}) also reflected the genetic differentiation of the subpopulation. SP2 and SP3 reflected high F_{st} values of 0.46 and 0.56, respectively, indicating high genetic divergence, whereas SP1 reflected low F_{st} of 0.06, indicating greater genetic admixture. These results are consistent with the work of Park et al. (2024), which indicated high genetic differentiation among *Saccharum* species, *S. robustum* and *S. sinense* ($F_{st} = 0.323$). Using single-dose SNP marker analysis done by them, the highest mean F_{st} value was 0.239 in *S. sinense*, followed by 0.224 for *S. robustum*, 0.178 for *S.*

barberi, 0.176 for *S. spontaneum*, and 0.123 for *S. officinarum*, showing various levels of genetic divergence among species. In this study a low level of minor genetic divergence between subpopulations, estimated as 1% difference based on AMOVA, and extensive gene exchange rate ($N_m = 29.66$) indicate extensive gene exchange at the large scale, as in findings by (You et al., 2016) in evaluating global sugarcane genetic resources. High levels of haploid migrants ($N_m = 29.66$) confirm the possibility of large-scale gene exchange between subpopulations. Likewise, Park et al. (2024) found that intra-species differences accounted for 85% of the entire genetic diversity and only 15% of the diversity was realized between species.

The allelic variation study within *Saccharum* species revealed contrasting patterns of genetic diversity, emphasizing the species-specific contribution to the sugarcane germplasm pool. *S. robustum*, *S. officinarum*, and *S. barberi* possessed the highest value of private alleles, and also high values for unique alleles (N_a), effective alleles (N_e), and Shannon's index of diversity (H), which indicate their broader allelic range and genetic differentiation (Berkman et al., 2014; Xiong et al., 2022). These findings are consistent with previous studies that have highlighted the rich genetic diversity retained in these species, particularly *S. robustum*, which has been proposed as a progenitor of cultivated sugarcane and contributes significantly to its genetic base (Wu et al., 2019). In contrast, species such as *S. sinense*, *S. edule*, and the misidentified group showed minimal allelic diversity and lacked private alleles, likely due to limited sample representation or narrower genetic backgrounds (Wu et al., 2019). Private alleles as well as diversity indices such as Shannon's index have also proven helpful in the estimation of genetic diversity and identification of unique genetic materials among sugarcane germplasm collections (Xiong et al., 2022). These results are relevant to guiding breeding schemes for expanding the genetic makeup of modern cultivars and maximizing trait introgression across different species. Such information regarding population structure and genetic differentiation will be an important component in breeding, agricultural practice improvement, and in adapting processing technologies for enhancing yield and quality of sugarcane. The genetic structure observed among subpopulations corresponds with known evolutionary relationships: *S. officinarum* is believed to have evolved from *S. robustum*, while *S. sinense* and *S. barberi* are recognized as hybrids of *S. officinarum* and *S. spontaneum* (Daniels and Roach, 1987). These relationships provide context for the clustering patterns observed in STRUCTURE and PCA analyses.

Genome wide association mapping for cane traits

Application of three other GWAS models General Linear Model (GLM), Mixed Linear Model (MLM), and Fixed and Random Model Circulating Probability Unification (FarmCPU) is a practically corroborating approach in detecting true associations under confounding effect (Yang et al., 2019). The application of principal components as covariates in GLM effectively circumvented population structure, which was required to avoid

false positives in sets of heterogeneous germplasm (Pritchard et al., 2000). The additional incorporation of a kinship matrix accounts for population structure and family relatedness, which further enhances association detection (Yu et al., 2006). The varying number of significant SNPs detected by each model (15 by GLM, 6 by MLM, and 19 by FarmCPU) agrees with previous research demonstrating that various statistical models possess differing stringency and power (Liu P. et al., 2016). MLM will generally call fewer markers due to being more conservative in handling population structure and kinship, thereby lowering false positives but maybe increasing false negatives (Z. Zhang et al., 2010). The enhanced marker provided by the FarmCPU model suggests its increased statistical power employing its unique two-step iterative algorithm with effective confounding effects control but retaining statistical power (Liu X et al., 2016).

The discovery of 21 large-effect markers for Brix on chromosomes 3, 4, 6, and 7 is most interesting. Discovering three Brix markers (AX-171243917-4651, AX-171305424-5045, and AX-171305424-5356) with all three models provides strong evidence for their existence (Tam et al., 2019). Previous studies have placed QTLs for Brix on sugarcane chromosomes 3 and 6 (Singh et al., 2013; Wei et al., 2010). Using the GLM and FarmCPU models, we identified eight markers significantly associated with polarity, located on chromosomes 4, 5, 6, and 8. Notably, the marker AX-171243917-4651 on chromosome 6D was significantly associated with both Brix and polarity, suggesting a potential link to genes involved in sucrose regulation. This is physiological because Polarity has a very close association with sucrose content, which also influences Brix values (Ming et al., 2006). The lack of significant markers for Polarity in the MLM model suggests that this trait may be strongly influenced by population structure and kinship relationships. The comprehensive correction for these factors within the MLM likely reduced spurious associations, resulting in no markers surpassing the significance threshold (Segura et al., 2012). This emphasizes the benefit of having multiple models for the analysis of complex traits in crops with highly complex genetic structures, like sugarcane. The single marker AX-171269335-3422 (chromosome 4D) was consistently detected by all three models for moisture content, indicating a robust and reliable association. Sugarcane moisture content determines processing efficiency and sugar recovery, so this marker would be highly useful in marker-assisted selection (Waclawovsky et al., 2010). The few important markers for this trait indicate that moisture content might be under less complex genetic control than other cane traits or that our marker density was too low to detect all contributing loci.

The identification of eight significant markers associated with fiber content on chromosomes 2, 5, and 7 is very informative regarding the genetic control of this economically important trait. Those two markers (AX-171287224-1157; chr. 2A and AX-171363600-3972; chr. 5D) identified by all three GWAS models to be significantly associated with fiber content, provides high confidence of the validity of these markers warranting further followed up studies of these markers. Previous QTLs for fiber content have been described on chromosomes 2 and 5 (Gouy et al., 2015), consistent with these findings, lending support to the

argument that these regions of the chromosome contain genes responsible for cell wall biosynthesis or regulation. To further explore the biological significance of significant SNPs, BLAST searches of their sequences were performed. Several aligned with sugar metabolism-related contigs in public databases (Supplementary Tables S3, S4).

The multi-model and multi-trait approach employed in this study has also identified several possible markers for the improvement of sugarcane. Most beneficial are the established markers that correspond to more than a single model since these indicate stronger associations less likely to be false positives (Balding, 2006). The three markers for Brix, as well as the two makers for fiber content markers, are strong potential candidates for validation and utilization within marker-assisted selection programs. The fact that AX-171243917-4651; chr. 6D has been found significant for both Brix and Polarity suggests the value of being able to enhance various quality traits parallel, which will enhance breeding efficiency (Chen and Lübberstedt, 2010). While this GWAS analysis identified several encouraging marker-trait associations, the complex polyploid nature of sugarcane presents some unique challenges for genetic analysis (D'Hont et al., 2008). The variation in the number of significant markers identified across models demonstrates the contribution of methodology to GWAS in polyploid crops.

The consistent allelic effects observed in this study align with previous reports of marker-trait associations for sucrose accumulation in sugarcane and related grasses. For example, QTLs for Brix and Pol have been mapped in sugarcane breeding populations (Aitken et al., 2005; Refay et al., 2005), and several GWAS studies have also identified SNPs linked to sucrose content and fiber traits (Fickett et al., 2019; Xiong et al., 2023; Yang et al., 2019). Similarly, studies in sorghum and maize have revealed allelic variants in sucrose metabolism and transport genes that explain significant variation in sugar accumulation (Chhabra et al., 2021; Liu et al., 2014). Our findings therefore provide complementary evidence that the identified here may represent reliable candidates for marker-assisted selection, even though replication across environments will be required to validate their predictive utility in breeding programs. These findings, though preliminary, suggest possible application in genomic selection schemes, where identified SNPs could serve as predictors for cane quality traits. Future work in replicated trials will be essential to validate and refine these associations.

The recommendation of this study that following experiments will have to utilize a larger number of SNPs marker sets, larger population sizes, and incorporation of haplotype-based approaches to further define the genetic structure of these complex traits (Garcia et al., 2013). Also, the functional characterization of these markers through molecular mapping, gene expression, and ultimately transformation experiments will be necessary to translate these statistical associations into biologically meaningful data (Huang et al., 2010). We acknowledge that the absence of replication and multi-year data limits the robustness of the marker-trait associations. Therefore, the markers identified in this study should be considered preliminary and require validation in

replicated, multi-location breeding trials or segregating populations. Although fiber and sucrose are traits typically requiring multi-year, multi-location testing, this study provides a first exploratory analysis across 397 diverse accessions. Future work will evaluate these traits under replicated, multi-year trials to confirm the stability and breeding utility of the associated markers.

Conclusion

This study offers an insight into genetic diversity and marker-trait associations in traits of sugarcane, with tools available for marker-assisted selection. The identification of stable and pleiotropic SNP markers, particularly for Brix, Polarity and fiber content, indicates their potential utility in improving more than one trait simultaneously. Adequate control of population structure and kinship in the GWAS models rendered the associations more robust. The phenotypic variation seen among *Saccharum* species and the genetic admixture underscore the importance of wide-based germplasm in breeding. However, the polyploid character of sugarcane demands further verification using fine mapping, gene expression, and multi-location testing. Future efforts need to increase marker density, increase population size, and follow haplotype-based and functional genomic approaches to fully value and utilize the genetic basis of cane traits for crop improvement.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/[Supplementary Material](#).

Author contributions

SHE: Formal Analysis, Software, Writing – original draft, Data curation, Methodology, Supervision, Writing – review & editing. SS: Funding acquisition, Writing – original draft, Conceptualization, Data curation, Resources, Writing – review & editing. DD: Resources, Writing – review & editing, Data curation. YA: Resources, Writing – review & editing, Data curation. HS: Methodology, Resources, Data curation, Writing – review & editing. MR: Supervision, Writing – review & editing. GA: Writing – review & editing, Resources, Conceptualization.

Funding

The author(s) declare financial support was received for the research and/or publication of this article. This research was funded by the appropriated funds from the United States Department of Agriculture, Agricultural Research Service (USDA-ARS) under Project Number 6038-21000-026-00D.

Acknowledgments

We extend our gratitude to the technicians, research team, retired scientists, and administrative staff for their contributions.

Conflict of interest

Author YA was employed by the company Florida Sugar Cane League Inc.

The remaining 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.

Generative AI statement

The author(s) declare that no Generative AI was used in the creation of this manuscript.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

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.

Author disclaimer

Mention of a trademark, warranty, proprietary product or vendor does not constitute a guarantee by the U.S. Department of Agriculture and does not imply approval or recommendation of the product to the exclusion of others that may be suitable.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2025.1643469/full#supplementary-material>

SUPPLEMENTARY TABLE 1

The list of 397 sugarcane accessions, species, STRUCTURE subpopulation and the mean of 4 cane quality traits.

SUPPLEMENTARY TABLE 2

The sequences of each marker of 357 SNPs used in this study.

SUPPLEMENTARY TABLE 3

Detailed GWAS results for the 4 cane quality traits (Brix, polarity, moisture and fiber) using GLM, MLM and FarmCPU model with 357 SNPs.

SUPPLEMENTARY TABLE 4

List of Common significant markers associated with cane quality traits (Brix, polarity, moisture and fiber) and identified using 3 different GWAS models.

SUPPLEMENTARY TABLE 5

The highest and lowest ten sugarcane accessions with brix values across homozygous alleles (blue cells) and heterozygous alleles (red cells).

SUPPLEMENTARY TABLE 6

Pairwise genetic distance among 20 selected sugarcane accessions with extreme Brix values (10 highest and 10 lowest) based on 357 SNP markers.

References

- Aitken, K. S., Jackson, P. A., and McIntyre, C. (2005). A combination of AFLP and SSR markers provides extensive map coverage and identification of homo (eo) logous linkage groups in a sugarcane cultivar. *Theor. Appl. Genet.* 110, 789–801. doi: 10.1007/s00122-004-1813-7
- Aitken, K. S., Jackson, P. A., and McIntyre, C. (2006). Quantitative trait loci identified for sugar related traits in a sugarcane (*Saccharum* spp.) cultivar *Saccharum officinarum* population. *Theor. Appl. Genet.* 112, 1306–1317. doi: 10.1007/s00122-006-0233-2
- Balding, D. J. (2006). A tutorial on statistical methods for population association studies. *Nat. Rev. Genet.* 7, 781–791. doi: 10.1038/nrg1916
- 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, e0219843. doi: 10.1371/journal.pone.0219843
- Berkman, P. J., Bundock, P. C., Casu, R. E., Henry, R. J., Rae, A. L., and Aitken, K. S. (2014). A survey sequence comparison of *Saccharum* genotypes reveals allelic diversity differences. *Trop. Plant Biol.* 7, 71–83. doi: 10.1007/s12042-014-9139-3
- Bhatt, B., Jeena, A. S., Aggarwal, G., Edhigalla, P., Parihar, P., Chaudhary, D., et al. (2025). "Empowering Sugarcane Breeding with New Technologies," in *Revolutionizing Sugarcane Molecular Breeding and Biotechnological Approaches: Current Status and Future Strategies* (Springer, Berlin, Heidelberg), 25–66.
- Chen, Y., and Lübberstedt, T. (2010). Molecular basis of trait correlations. *Trends Plant Sci* 15, 454–461. doi: 10.1016/j.tplants.2010.05.004
- Chhabra, R., Muthusamy, V., Gain, N., Katral, A., Prakash, N. R., Zunjare, R. U., et al. (2021). Allelic variation in sugary1 gene affecting kernel sweetness among diverse-mutant and wild-type maize inbreds. *Mol. Genet. Genomics* 296, 1085–1102. doi: 10.1007/s00438-021-01807-9
- D'Hont, A., Souza, G. M., Menossi, M., Vincenz, M., Van-Sluis, M.-A., Glaszmann, J. C., et al. (2008). "Sugarcane: a major source of sweetness, alcohol, and bio-energy," in *Genomics of tropical crop plants* (Springer, Berlin, Heidelberg), 483–513.
- Dahlquist, E. (2013). ed. Biomass as energy source: resources, systems and applications. CRC Press.
- Daniels, J., and Roach, B. T. (1987). "Taxonomy and evolution," in *Developments in crop science*, vol. 11. (Elsevier, Amsterdam), 7–84.
- de Moraes, L. K., de Aguiar, M. S., de Albuquerque e Silva, P., Câmara, T. M. M., Cursi, D. E., Júnior, A. R. F., et al. (2015). Breeding of sugarcane. *Ind. Crops: Breed. Bioprocess Bioproducts* 9, 29–42. doi: 10.1007/978-1-4939-1447-0_2
- Deren, C. W. (1995). Genetic base of US mainland sugarcane. *Crop Sci* 35, 1195–1199. doi: 10.2135/cropsci1995.0011183X003500040047x
- Dinesh Babu, K. S., Janakiraman, V., Palaniswamy, H., Kasirajan, L., Gomathi, R., and Ramkumar, T. R. (2022). A short review on sugarcane: its domestication, molecular manipulations and future perspectives. *Genet. Resour. Crop Evol.* 69, 2623–2643. doi: 10.1007/s10722-022-01430-6
- Dlamini, N. E., and Zhou, M. (2024). Predicting ratooning ability of sugarcane varieties in selection trials. *Sugar Tech* 26, 52–62. doi: 10.1007/s12355-023-01319-3
- 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. doi: 10.1111/j.1365-294X.2005.02553.x
- 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, 1794–1801. doi: 10.1016/j.ygeno.2018.12.002
- Garcia, A. A. F., Mollinari, M., Marconi, T. G., Serang, O. R., Silva, R. R., Vieira, M. L. C., et al. (2013). SNP genotyping allows an in-depth characterisation of the genome of sugarcane and other complex autopolyploids. *Sci. Rep.* 3, 3399. doi: 10.1038/srep03399
- 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
- Hoarau, J.-Y., Dumont, T., Wei, X., Jackson, P., and D'hont, A. (2022). Applications of quantitative genetics and statistical analyses in sugarcane breeding. *Sugar Tech* 24, 320–340. doi: 10.1007/s12355-021-01012-3
- Huang, X., Wei, X., Sang, T., Zhao, Q., Feng, Q., Zhao, Y., et al. (2010). Genome-wide association studies of 14 agronomic traits in rice landraces. *Nat. Genet.* 42, 961–967. doi: 10.1038/ng.695
- Jackson, P. A. (2005). Breeding for improved sugar content in sugarcane. *Field Crops Res.* 92, 277–290. doi: 10.1016/j.fcr.2005.01.024
- Khan, Q., Qin, Y., Guo, D.-J., Yang, L.-T., Song, X.-P., Xing, Y.-X., et al. (2023). A Review of the diverse genes and molecules involved in sucrose metabolism and innovative approaches to improve sucrose content in sugarcane. *Agronomy* 13, 2957. doi: 10.3390/agronomy13122957
- Khanbo, S., Somyong, S., Phetchawang, P., Wirosirak, W., Ukoskit, K., Klomsa-Ard, P., et al. (2023). A SNP variation in the Sucrose synthase (SoSUS) gene associated with sugar-related traits in sugarcane. *PeerJ* 11, e16667. doi: 10.7717/peerj.16667
- Legendre, B. L. (1992). The core/press method for predicting the sugar yield from cane for use in cane payment. *Sugar Journal* 54(9), 2–7.
- Li, Y.-L., and Liu, J.-X. (2018). StructureSelector: A web-based software to select and visualize the optimal number of clusters using multiple methods. *Mol. Ecol. Resour.* 18, 176–177. doi: 10.1111/men.2018.18.issue-1
- Lima, M. L. A., Garcia, A. A. F., Oliveira, K. M., Matsuoka, S., Arizono, H., de Souza, C. L. Jr., et al. (2002). Analysis of genetic similarity detected by AFLP and coefficient of parentage among genotypes of sugar cane (*Saccharum* spp.). *Theor. Appl. Genet.* 104, 30–38. doi: 10.1007/s001220200003
- Liu, P., Chandra, A., Que, Y., Chen, P.-H., Grisham, M. P., White, W. H., et al. (2016). Identification of quantitative trait loci controlling sucrose content based on an enriched genetic linkage map of sugarcane (*Saccharum* spp. hybrids) cultivar 'LCP 85-384'. *Euphytica* 207, 527–549. doi: 10.1007/s10681-015-1538-5
- Liu, X., Huang, M., Fan, B., Buckler, E. S., and Zhang, Z. (2016). Iterative usage of fixed and random effect models for powerful and efficient genome-wide association studies. *PLoS Genet.* 12, e1005767. doi: 10.1371/journal.pgen.1005767
- Liu, K., and Muse, S. V. (2005). PowerMarker: an integrated analysis environment for genetic marker analysis. *Bioinformatics* 21, 2128–2129. doi: 10.1093/bioinformatics/bti282
- Liu, Y., Nie, Y.-D., Han, F.-X., Zhao, X.-N., Dun, B.-Q., Lu, M., et al. (2014). Allelic variation of a soluble acid invertase gene (SAI-1) and development of a functional marker in sweet sorghum [*Sorghum bicolor* (L.) Moench. *Mol. Breed.* 33, 721–730. doi: 10.1007/s11032-013-9988-8
- Ming, R., Moore, P. H., Wu, K.-K., D'hont, A., Glaszmann, J. C., Tew, T. L., et al. (2006). Sugarcane improvement through breeding and biotechnology. In J. Janick (Ed.), *Plant Breeding Reviews* 27, 15–118. Wiley-Blackwell, Berlin, Heidelberg. doi: 10.1002/9780470650349.ch2
- Nascimento, L. C., Yanagui, K., Jose, J., Camargo, E. L. O., Grassi, M. C. B., Cunha, C. P., et al. (2019). Unraveling the complex genome of *Saccharum spontaneum* using Polyploid Gene Assembler. *DNA Res.* 26, 205–216. doi: 10.1093/dnares/dsz001
- Nayak, S. N., Song, J., Villa, A., Pathak, B., Ayala-Silva, T., Yang, X., et al. (2014). Promoting utilization of *Saccharum* spp. genetic resources through genetic diversity analysis and core collection construction. *PLoS One* 9, e110856. doi: 10.1371/journal.pone.0110856
- Park, S., Zhang, D., and Ali, G. S. (2024). Assessing the genetic integrity of sugarcane germplasm in the USDA-ARS National Plant Germplasm System collection using single-dose SNP markers. *Front. Plant Sci* 14, 1337736. doi: 10.3389/fpls.2023.1337736
- Peakall, R. O. D., and Smouse, P. E. (2006). GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. *Mol. Ecol. Notes* 6, 288–295. doi: 10.1111/j.1471-8286.2005.01155.x
- Perlo, V., Botha, F. C., Furtado, A., Hodgson-Kratky, K., and Henry, R. J. (2020). Metabolic changes in the developing sugarcane culm associated with high yield and early high sugar content. *Plant Direct* 4, e00276. doi: 10.1002/pld3.v4.11
- Pritchard, J. K., Stephens, M., and Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics* 155, 945–959. doi: 10.1093/genetics/155.2.945
- 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 relative into Australian sugarcane. *Mol. Breed.* 15, 367–381. doi: 10.1007/s11032-004-7981-y

- Saavedra-Díaz, C., Trujillo-Montenegro, J. H., Jaimes, H. A., Londoño, A., Villareal, F. A. S., López, L. O., et al. (2024). Genetic association analysis in sugarcane (*Saccharum* spp.) for sucrose accumulation in humid environments in Colombia. *BMC Plant Biol.* 24, 570. doi: 10.1186/s12870-024-05233-y
- Segura, V., Vilhjálmsson, B. J., Platt, A., Korte, A., Seren, Ü., Long, Q., et al. (2012). An efficient multi-locus mixed-model approach for genome-wide association studies in structured populations. *Nat. Genet.* 44, 825–830. doi: 10.1038/ng.2314
- 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
- Srinivasan, A., Elumalai, K., Ranganathan, N., Manickavasagam, R. S., Viswanathan, R., Suganya, A., et al. (2021). Molecular identification and genetic diversity analysis of sugarcane clones by SSR markers. *Journal of Sugarcane Research* 10(2), 140–151. doi: 10.37580/JSR.2020.2.10.140-151
- Tam, V., Patel, N., Turcotte, M., Bossé, Y., Paré, G., and Meyre, D. (2019). Benefits and limitations of genome-wide association studies. *Nat. Rev. Genet.* 20, 467–484. doi: 10.1038/s41576-019-0127-1
- Tamura, K., Stecher, G., and Kumar, S. (2021). MEGA11: molecular evolutionary genetics analysis version 11. *Mol. Biol. Evol.* 38, 3022–3027. doi: 10.1093/molbev/msab120
- Team, R. C., Chambers, J. M., Ihaka, R., Gentleman, R., Bates, D., and Venables, W. N. (2016). *R: A language and environment for statistical computing* (Vienna, Austria: R Foundation for Statistical Computing). Available online at: <http://www.R-Project.org/> (accessed September 24, 2025).
- Todd, J., Wang, J., Glaz, B., Sood, S., Ayala-Silva, T., Nayak, S. N., et al. (2014). Phenotypic characterization of the Miami World Collection of sugarcane (*Saccharum* spp.) and related grasses for selecting a representative core. *Genet. Resour. Crop Evol.* 61, 1581–1596. doi: 10.1007/s10722-014-0132-3
- Waclawovsky, A. J., Sato, P. M., Lembke, C. G., Moore, P. H., and Souza, G. M. (2010). Sugarcane for bioenergy production: an assessment of yield and regulation of sucrose content. *Plant Biotechnol. J.* 8, 263–276. doi: 10.1111/j.1467-7652.2009.00491.x
- Wang, H., Pan, Y.-B., Wu, M., Liu, J., Yang, S., Wu, Q., et al. (2024). Sugarcane genetics: Underlying theory and practical application. *Crop J* 13(2), 328–338. doi: 10.1016/j.cj.2024.11.005
- 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, 973–981. doi: 10.1139/G10-050
- Wickham, H. (2016). “Data analysis,” in *ggplot2: elegant graphics for data analysis* (Springer, Berlin, Heidelberg), 189–201.
- Wu, J., Wang, Q., Xie, J., Pan, Y.-B., Zhou, F., Guo, Y., et al. (2019). SSR marker-assisted management of parental germplasm in sugarcane (*Saccharum* spp. hybrids) breeding programs. *Agronomy* 9, 449. doi: 10.3390/agronomy9080449
- Xiong, H., Chen, Y., Gao, S.-J., Pan, Y.-B., and Shi, A. (2022). Population structure and genetic diversity analysis in sugarcane (*Saccharum* spp. hybrids) and six related *Saccharum* species. *Agronomy* 12, 412.
- Xiong, H., Chen, Y., Pan, Y.-B., and Shi, A. (2023). A genome-wide association study and genomic prediction for fiber and sucrose contents in a mapping population of LCP 85–384 sugarcane. *Plants* 12, 1041. doi: 10.3390/plants12051041
- Xu, F., Wang, Z., Lu, G., Zeng, R., and Que, Y. (2021). Sugarcane ratooning ability: Research status, shortcomings, and prospects. *Biology* 10, 1052. doi: 10.3390/biology10101052
- Yang, X., Luo, Z., Todd, J., Sood, S., and Wang, J. (2020). Genome-wide association study of multiple yield traits in a diversity panel of polyploid sugarcane (*Saccharum* spp.). *Plant Genome* 13, e20006. doi: 10.1002/tpg2.20006
- Yang, X., Sood, S., Luo, Z., Todd, J., and Wang, J. (2019a). Genome-wide association studies identified resistance loci to orange rust and yellow leaf virus diseases in sugarcane (*Saccharum* spp.). *Phytopathology* 109, 623–631. doi: 10.1094/PHYTO-08-18-0282-R
- Yang, X., Todd, J., Arundale, R., Binder, J. B., Luo, Z., Islam, M. S., et al. (2019b). Identifying loci controlling fiber composition in polyploid sugarcane (*Saccharum* spp.) through genome-wide association study. *Ind. Crops Products* 130, 598–605. doi: 10.1016/j.indcrop.2019.01.023
- Yin, L., Zhang, H., Tang, Z., Xu, J., Yin, D., Zhang, Z., et al. (2021). rMVP: a memory-efficient, visualization-enhanced, and parallel-accelerated tool for genome-wide association study. *Genomics Proteomics Bioinf.* 19, 619–628. doi: 10.1016/j.gpb.2020.10.007
- You, Q., Pan, Y.-B., Xu, L.-P., Gao, S.-W., Wang, Q.-N., Su, Y.-C., et al. (2016). Genetic diversity analysis of sugarcane germplasm based on fluorescence-labeled simple sequence repeat markers and a capillary electrophoresis-based genotyping platform. *Sugar Tech* 18, 380–390. doi: 10.1007/s12355-015-0395-9
- 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, 203–208. doi: 10.1038/ng1702
- Zhang, Z., Ersoz, E., Lai, C.-Q., Todhunter, R. J., Tiwari, H. K., Gore, M. A., et al. (2010). Mixed linear model approach adapted for genome-wide association studies. *Nat. Genet.* 42, 355–360. doi: 10.1038/ng.546
- 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, 1565–1573. doi: 10.1038/s41588-018-0237-2
- Zhou, M. M., and Gwata, E. T. (2015). Location effects and their implications in breeding for sugarcane yield and quality in the midlands region in South Africa. *Crop Sci* 55, 2628–2638. doi: 10.2135/cropsci2015.02.0101