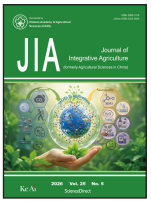
Journal of Integrative Agriculture 2026, 25(5): 1939–1948

Available online at ScienceDirect

Journal of Integrative Agriculture



RESEARCH ARTICLE

Genome-wide association study of sucrose content and stem diameter in sugarcane (*Saccharum* spp.)

Fenggang Zan[1, 2*], Zhuandi Wu[1, 2*], Chengcai Xia[3, 4*], Long Zhao[3, 4], Qi Liu[3, 4], Zihao Wang[3, 4], Yanjie Lu[3, 4], Meiling Zou[3, 4], Yong Zhao[1, 2], Peifang Zhao[1, 2], Xuan Luo[3, 4], Jiayong Liu[1, 2#], Zhiqiang Xia[1, 3, 4#]

[1] *National Key Laboratory for Tropical Crop Breeding, Kunming 650205, China*
[2] *Sugarcane Research Institute, Yunnan Academy of Agricultural Sciences/Yunnan Key Laboratory of Sugarcane Genetic Improvement, Kaiyuan 661699, China*
[3] *School of Breeding and Multiplication (Sanya Institute of Breeding and Multiplication), Hainan University, Sanya 572025, China*
[4] *School of Tropical Agriculture and Forestry, Hainan University, Danzhou 571737, China*

Highlights

- Sucrose content and stem diameter of 292 sugarcane germplasm resources were determined.
- A total of 46 SNPs associated with the phenotypic traits were identified through genome-wide association analysis.
- Twenty-six candidate genes related to the target traits were detected.

Abstract

Increasing the sucrose content of sugarcane, a major sugar crop, is a key breeding objective. However, the complex genetic background of sugarcane affects development of sugarcane hybrids. In this study, we sequenced 292 sugarcane germplasm accessions and identified 2,542,965 single nucleotide polymorphisms (SNPs) and insertions/deletions (InDels). We performed a genome-wide association study (GWAS) for two important sugarcane traits: sucrose content and stem diameter. Both traits followed a normal distribution and showed typical characteristics of quantitative traits. Population structure analysis revealed four subpopulations with an average genetic distance of 0.236. GWAS of the sucrose content detected 27 SNPs. After annotating genes at or near significant loci, 17 candidate genes were screened. For stem diameter, GWAS revealed 19 SNPs, from which 9 candidate genes were identified. These results improve our understanding of genetic mechanisms affecting sucrose content in sugarcane, and identify important genetic resources to accelerate breeding of new sugarcane varieties with high sucrose content.

Keywords: sugarcane, sucrose content, stem diameter, genome-wide association study (GWAS)

1. Introduction

Sugarcane (*Saccharum* spp.) is the most important sugar crop globally, accounting for about 80% of world sugar and 40% of biofuel production (Waclawovsky *et al*. 2010; Chandra *et al.* 2015; Verma *et al.* 2022). Increasing sucrose content, an important index in sugarcane breeding (Chen *et al.* 2019; Khanbo *et al*. 2021), is a target of sugarcane breeding programs (Waclawovsky *et al.* 2010; Chandra *et al.* 2015). Therefore, evaluating and exploring high-sucrose germplasm resources is important for sugarcane breeding programs. However, conventional ways to achieve this have been inefficient and resulted in relatively low use efficiency of high-sucrose parents, breeding on a large scale, and unsatisfactory effects (Brown *et al.* 1968; Mirzawan *et al.* 1994; Atkin *et al.* 2009; Wang *et al.* 2024; Li *et al.* 2025). Breeding of high-sucrose varieties is mainly performed using traditional methods based on sucrose content, and the process is relatively slow (Silva *et al.* 2016; Cursi *et al.* 2021).

With rapid development of sugarcane genomics, high-throughput sequencing technology, and GWAS, new opportunities have developed for molecular breeding and germplasm evaluation. For example, the development of

Received 27 November 2024; Received in revised form 25 June 2025; Accepted 29 August 2025; Available online 26 December 2025
[#]Correspondence Jiayong Liu, E-mail: lljjyy1976@163.com; Zhiqiang Xia, E-mail: zqiangx@gmail.com
[*] These authors contributed equally to this study.

© 2026 CAAS. Publishing services by Elsevier B.V. on behalf of KeAi Communications Co. Ltd. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/). Peer review under responsibility of Editorial Board of *Journal of Integrative Agriculture*.
doi: 10.1016/j.jia.2025.12.044

molecular markers for sucrose content traits can improve the efficiency of early selection and resolve problems associated with selecting phenotypic traits (Aitken 2022). Genome selection can also shorten the breeding cycle of sugarcane varieties, reduce the size of the breeding population, and assist in the selection of ideal parents and desirable sugarcane varieties (Job 2002; Sabzehzar *et al.* 2019).

During the evolution of sugarcane cultivars, their ancestors underwent two genome-wide replication events. They became octaploid (Jannoo *et al.* 2007), the genome doubled and the number of alleles increased, increasing the possibility of adaptation by cultivars (de Souza *et al.* 2013). Most cultivated sugarcane originated from hybridizing *S. officinarum* (2n=8x=80; x=10), *S. spontaneum* (2n=5x–16x=40–128; x=8), and wild relatives, which is a highly polyploid and aneuploid hybrid crop (de Souza *et al.* 2013; Wang *et al.* 2022). Most of these sugarcane varieties have 110–130 chromosomes (Jannoo *et al.* 2004; de Souza *et al.* 2013) and very large genomes (D'Hont *et al.* 1998; Zhang J *et al.* 2018; Piperidis and Hont 2020).

With progress in sugarcane genome sequencing (Garsmeur *et al.* 2018; Zhang J *et al.* 2018) and accumulation of genomic data resources, research on sugarcane molecular-assisted breeding has become increasingly feasible. Development and application of high-throughput sequencing technology (Battenfield *et al.* 2018; Li *et al.* 2019) and QTL analysis have provided new opportunities for molecular breeding and germplasm evaluation (Garsmeur *et al.* 2018), rendering marker-assisted breeding of sugarcane has become a desirable and powerful tool. Simplified genome sequencing technology reduces the cost of next-generation sequencing technology, generates thousands of single nucleotide polymorphisms (SNPs) and insertions/deletions (InDels), and is an ideal method to improve the density and coverage of molecular markers (Liu *et al.* 2022). Commonly used simplified genome sequencing technologies include restriction site-associated DNA, gene sequencing typing, site-specific amplified fragment sequencing (Zhang *et al.* 2022; Zhou and Pan 2023), amplified fragment SNP and methylation sequencing (AFSM) and Hyper-seq (a novel extremely low-cost, effective, flexible, and high-throughput DNA sequencing library preparation and genotyping approach) (Xia *et al.* 2014, 2018a, b; Zou *et al.* 2017; Zou and Xia 2022).

AFSM-seq is commonly used to construct plant genetic linkage maps and QTL mapping. This methodology has been applied to various plant species (Wen *et al.* 2010; Xia *et al.* 2014, 2018ab; Zou *et al.* 2017; Zhang S *et al.* 2018; Wang F *et al.* 2021), and excellent linkage molecular markers have been obtained. Hyper-seq is a low-cost, effective, flexible and high throughput DNA sequencing library preparation and genotyping approach, which significantly reduces the time and cost associated with library preparation and sequencing (Zou and Xia 2022).

Deciphering the sugarcane genome provides considerable data for the development of large-scale molecular markers. The haploid genome of *S. spontaneum* accession AP85-441 (1n=4x=32, 3.5 Gb) was sequenced in 2018 (Zhang J *et al.* 2018). Moreover, using the collinearity between the sugarcane and sorghum genomes, the mosaic haploid sequence of R570 (530.7 Mb) was obtained based on the bacterial artificial chromosome (BAC) library (Garsmeur *et al.* 2018), and a 3.13 Gb genome was finally assembled. The publication of RNA-Seq data on sugarcane varieties has also brought new impetus and opportunities for sugarcane molecular breeding.

GWAS can determine precise chromosomal locations, and their resolution can define polymorphisms within specific genes and coding regions (Korte and Farlow 2013). Genetic studies on sucrose content and the development and use of molecular markers can improve efficiency in use of sugarcane germplasm and parents, and accelerate the breeding process of high-sucrose varieties (Mirajkar *et al.* 2016). We use the Hyper-seq to analyze the entire genome of sucrose content traits in 292 sugarcane germplasm resources from different countries. We aim to identify agronomic traits of desirable sugarcane germplasm, and specific germplasm for use to breed high-sucrose varieties. This study utilizes Hyper-seq approach to perform GWAS on 292 sugarcane germplasm accessions, aiming to rapidly identify key genetic loci associated with sucrose content and stem diameter, thereby providing a genetic basis for accelerating the breeding of high-sucrose sugarcane varieties.

2. Materials and methods

2.1. Sample collection

A total of 292 sugarcane resources (China (113), Brazil (19), Philippines (23), the United States of America (60), France (16), Australia (51), Reunion (1), India (5), and Indonesia (4)) were obtained from the China National Sugarcane Germplasm Resource Nursery (Appendix A).

2.2. Experimental design, and determining sucrose content and stem diameter

Each of the 292 germplasm materials was planted in March 2017 at the Sugarcane Research Institute, Yunnan Academy of Agricultural Sciences, Kaiyuan, Yunnan Province, China. The experimental setup involved a random block design, replicated twice. During the growing seasons of 2017–2018, the sucrose content of each material was determined from November of the same year to March of the following year using a one-step optical rotation method (once a month, with two repeated measurements each time, and three healthy sugarcane plants were taken from each material for detection each time). The stem diameter should be measured once a year during the mature period (divided into repeated measurements, with 3 plants measured in each repetition).

In this study, we conducted precise measurements of the sucrose content and stalk diameter of sugarcane. To ensure the scientificity and rationality of the grading standards, we conducted a comprehensive analysis based on data collected over three consecutive years and formulated a detailed grading system for sucrose content and stalk diameter accordingly. In terms of sucrose content grading (based on the three-year average), the specific grading criteria are as follows: Grade 1 sugarcane has a sucrose content of 16% or above; Grade 2 sugarcane has a sucrose content

between 15 and 16% (inclusive); Grade 3 sugarcane has a sucrose content within the range of 14% (inclusive) to 15%; Grade 4 sugarcane has a sucrose content between 13% (inclusive) and 14%; Grade 5 sugarcane has a sucrose content from 12% (inclusive) to 13%; and sugarcane with a sucrose content below 12% is classified as Grade 6. For stalk diameter grading, also using the three-year average as a reference, the grading standards are as follows: Grade 1 sugarcane has an average stalk diameter of 3.0 cm or above; Grade 2 sugarcane has an average stalk diameter between 2.8 cm (inclusive) and 3.0 cm; Grade 3 sugarcane has an average stalk diameter within the range of 2.6 cm (inclusive) to 2.8 cm; Grade 4 sugarcane has an average stalk diameter between 2.4 cm (inclusive) and 2.6 cm; Grade 5 sugarcane has an average stalk diameter from 2.2 cm (inclusive) to 2.4 cm; and sugarcane with an average stalk diameter below 2.2 cm is classified as Grade 6. Currently, the sucrose content and stalk diameter data of all sugarcane samples have been strictly categorized according to the above grading standards. This is aimed at providing a clear and accurate data foundation for subsequent analysis of sugarcane growth characteristics and comparative studies among different varieties. An optimal linear unbiased prediction of sucrose was performed using Genstat software (Emebiri *et al.* 2005); results were used as phenotypes for association analysis.

2.3. DNA preparation and sequencing

An improved cetyltrimethylammonium bromide method (Murray and Thompson 1980) was used to extract genomic DNA from sugarcane leaves. After detecting and quantifying DNA concentrations by 1% agarose gel electrophoresis, the working DNA solution was diluted to 100 ng μL^{-1} and stored at −20°C. A Hyper-seq approach (Zou and Xia 2022) was used to construct a library of 292 sugarcane DNA samples. The library underwent high-throughput sequencing in an Illumina HiSeq2000 apparatus (Illumina Inc., San Diego, CA, USA).

2.4. SNP extraction and annotation

Custom Perl Scripts2 was used to process raw Illumina sequence reads. Processing was performed to optimize read numbers and reduce artifacts within data. Barcodes were used to assign obtained sequences to individual samples; these sequences were also reduced to 2×132 bp. Distinct tags in more than five lines were determined in the complete set of reads. Bowtie 2 (Langmead and Salzberg 2012) was used to align filtered sequence reads to the *S. spontaneum* AP85-441 reference genome (Zhang J *et al.* 2018), from https://www.life.illinois.edu/ming/downloads/Spontaneum_genome/ (Li *et al.* 2009) and VCFtools (Danecek *et al.* 2011) were used to identify SNPs and InDels, which were annotated based on the *S. spontaneum* AP85-441 genome through SnpEff (Cingolani *et al.* 2012). SNPs were categorized as intergenic regions, untranslated regions (3′ and 5′ UTR), intronic, and coding sequences. SNPs in coding sequences were classified as synonymous or nonsynonymous. InDels in exons were categorized according to their frame shift effect.

2.5. Population genetic analysis

Phylogenetic trees were constructed using the neighborhood linkage method and distance matrix calculated in PHYLIP4. All SNPs were used to analyze genetic relationships between individuals. Principal Component Analysis (PCA) and Genome-wide Complex Trait Analysis evaluated genetic relationships among sugarcane individuals (Jolliffe and Cadima 2016; Ferrara *et al.* 2019). The eigenvector decomposition of the matrix in R was calculated using eigenfunctions, and the PCA diagram was drawn using these values. Population structure analysis used ADMIXTURES (Alexander *et al.* 2009), with PLINK 1.9 Beta (Purcell *et al.* 2007) used to obtain required data files. The input parameter *K* ranged 1–12 and represented the hypothetical population of a simulated population. We used VCFtools to calculate the fixed index (FST); the selected sweep analysis was performed with 10 kb windows sliding in 5 kb steps. TreeMix (Fitak 2021) was used to determine historical relationships between the four groups, and gene flow between them.

2.6. Genome-wide association study (GWAS)

We identified 66,699 high-quality SNPs and 96,680 InDels filtered by MAF>0.05 and HWE>0.001 (Pongpanich *et al.* 2010). GWAS analyzed the sucrose content of each of the 292 germplasm materials. The Compressed Mixed Linear Model (Compressed MLM) and General Linear Model (GLM) of TASSEL 5.0 software (Bradbury *et al.* 2007) were used for correlation analysis. Relatedness was observed in all SNPs, with a significant correlation threshold of 1/N, where N represents the total number of SNPs and Indels. We selected associated site genes as candidate genes based on SNP annotation and linkage disequilibrium (LD) decay. If the locus was upstream and downstream of other genes, these upstream and downstream genes were also considered candidate genes. If the locus was within the interval, the nearest upstream and downstream gene was selected as the candidate gene. Using the physical location of the SNP in the sugarcane genome, and its *P*-value, the qqman package in R software was used to draw Manhattan and QQ plots. Samtools was used to manually verify regions significantly correlated with the reordered read results of the sugarcane reference genome.

3. Results

3.1. Phenotypic analysis of 292 sugarcane materials

Phenotypic data (Appendix B) were normally distributed (Fig. 1-A and C). Boxplots of sucrose content and stem diameter (Fig. 1-B and D) reveal a median of 3, upper quartile of 4, and lower quartile of 2, indicating that the data are relatively consistent and show little variation. This distribution is typical of quantitative traits and suitable for correlation analysis.

3.2. Genotyping of 292 sugarcane materials

We obtained 920 GB of sequence data from 292 sugarcane germplasm resources. Sequencing readings from each sample were compared with the reference genome of tetraploid *S. spontaneum* AP85-441, resulting in more than

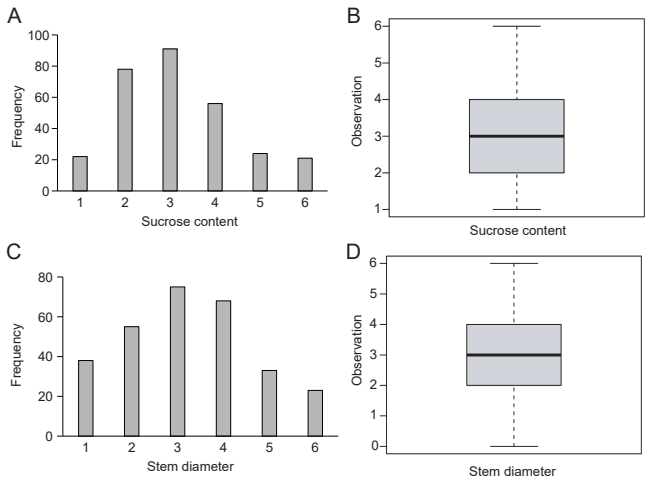


Fig. 1 Sucrose content and stem diameter in 292 sugarcane germplasm. A and B, frequency distribution and boxplot of sucrose content, respectively. C and D, frequency distribution and boxplot of stem diameter, respectively.

6.172 billion 150 bp paired-end reads. Genetic map results of these germplasm materials were used to identify all SNPs and InDels using Samtools (Li *et al.* 2009) and VCFtools, and SNPs of germplasm materials were identified. In total, 2,542,965 SNPs and InDels were detected (comprising 2,257,929 SNPs and 285,036 InDels). For all SNPs and InDels, two filter conditions were set: secondary allele frequency (MAF)>0.05 and HWE>0.001 (Pongpanich *et al.* 2010). Ultimately, 163,379 high-quality variation loci were selected for subsequent analysis.

3.3. Population structure analysis of sugarcane

ADMIXTURE 1.3 software was used to analyze all 163,379 high-quality variation loci, infer the maximum cluster subgroup number (K) as 1–12, and calculate the cross-validation error (CV error) rate under each *K* value. When *K*=4, the CV error was the smallest, 0.45806, and the change was most significant (Fig. 2-A). That is, the entire sugarcane population could be divided into four subgroups. Each germplasm was classified into one of four subgroups with the largest Q value.

All high-quality SNP and InDel molecular markers were analyzed and plotted by PCA using R software; the four subgroups were grouped by ADMIXTURE software. PCA clearly separated groups I–III, but group IV was mixed among groups I–III (Fig. 2-B). Seven subgroups were distinguished on PC1; clustering results were consistent with population structure division.

The neighbor-joining method was used to construct a phylogenetic tree; to explore genetic relationships between the 292 sugarcane germplasms, the tree was drawn with iTOL software. Overall, clustering results were consistent with population structure division, with groups I–III clustering together, with samples from group IV mixed in with group III, and with some crossover among samples (Fig. 2-C). No significant correlation was observed between geographic origin and genetic structure (Fig. 2-D), indicating little regional differentiation of sugarcane, possibly because samples from different countries had descended from common or similar parents. ADMIXTURE analysis of germplasm revealed groups I–IV to mainly come from four genetic backgrounds, and for the genetic background of each group to be mixed.

To further explore differences between groups, we estimated their nucleotide diversity by calculating PI (the probability that two randomly selected homologous sequences in a population were identical). Group II had the highest (4.18×10^{-5}) average nucleotide diversity (π), followed by groups I (4.17×10^{-5}), III (4.09×10^{-5}), and IV (4.16×10^{-5}) (Fig. 3-A). Genetic differentiation between groups III and I was greatest (FST=0.005), that between groups III and II was second only to that between III and I (FST=0.0049), and that between groups III and IV was lowest (FST=0.0038) (Fig. 3-A). Group III was more closely related to IV. TreeMix phylogenetic network analysis indicated groups IV and I had a common genetic lineage (Fig. 3-B).

3.4. GWAS for sucrose contents

The mixed linear model analysis performed correlation analysis on sucrose contents. A total of 163,379 high quality SNPs and InDels were used for association analysis with sucrose content using a GLM and compressed MLM (Fig. 4-A). The QQ map revealed SNP sites associated with significant correlation to be reliable (Fig. 4-B). $P<0.05$ was set as the threshold to determine significant loci for sucrose contents. Through GWAS, sites of significant variation were mined to analyze reasonably and to identify candidate genes. SNPs in annotated loci were obtained by combining SNP genomic positions. A total of 27 SNP sites (or those adjacent to them) associated with sucrose levels in sugarcane were identified (Appendix C), as were specific mutation positions of different SNPs on different chromosomes and their *P*-value. Among these sites, we screened 17 candidate genes (Appendix D). Genes on chromosome 6 were mainly related to potassium transporter proteins. In rice research, it has been discovered that neither the mutation of OsHAK18 nor the enhancement of its expression using its own promoter affects the K^+ uptake rate in roots. However, the mutation of OsHAK18 reduces the concentrations of K^+ and soluble sugars in phloem sap as well as the K^+ content in roots, while increasing the relative accumulation of K^+ in the shoots. In contrast, the enhanced expression of OsHAK18 exhibits opposite effects (Peng *et al.* 2023). On chromosome 5, genes were mainly related to the major facilitator superfamily, and zinc-induced protein. Genes on chromosome 4 were more associated with reverse transcriptase, SNARE-associated Golgi protein, and the ABC transporter transmembrane region. Genes on chromosome 2 were mainly associated with the CRM1 C terminal and 60S ribosomal protein. The candidate gene on chromosome 8 was related to kinase-like and RNA recognition motifs. Additionally, three genes with unknown functions were found on chromosomes 1 and 6, requiring further verification. Candidate genes on chromosome 8 were related to early nodulin 93 (ENOD93) protein. Studies have shown that *ENOD93* plays a crucial role in growth, development, and responses to abiotic stresses. For instance, in rice, the expression of *OsENOD93-6* and *OsENOD93-7* is significantly induced and upregulated

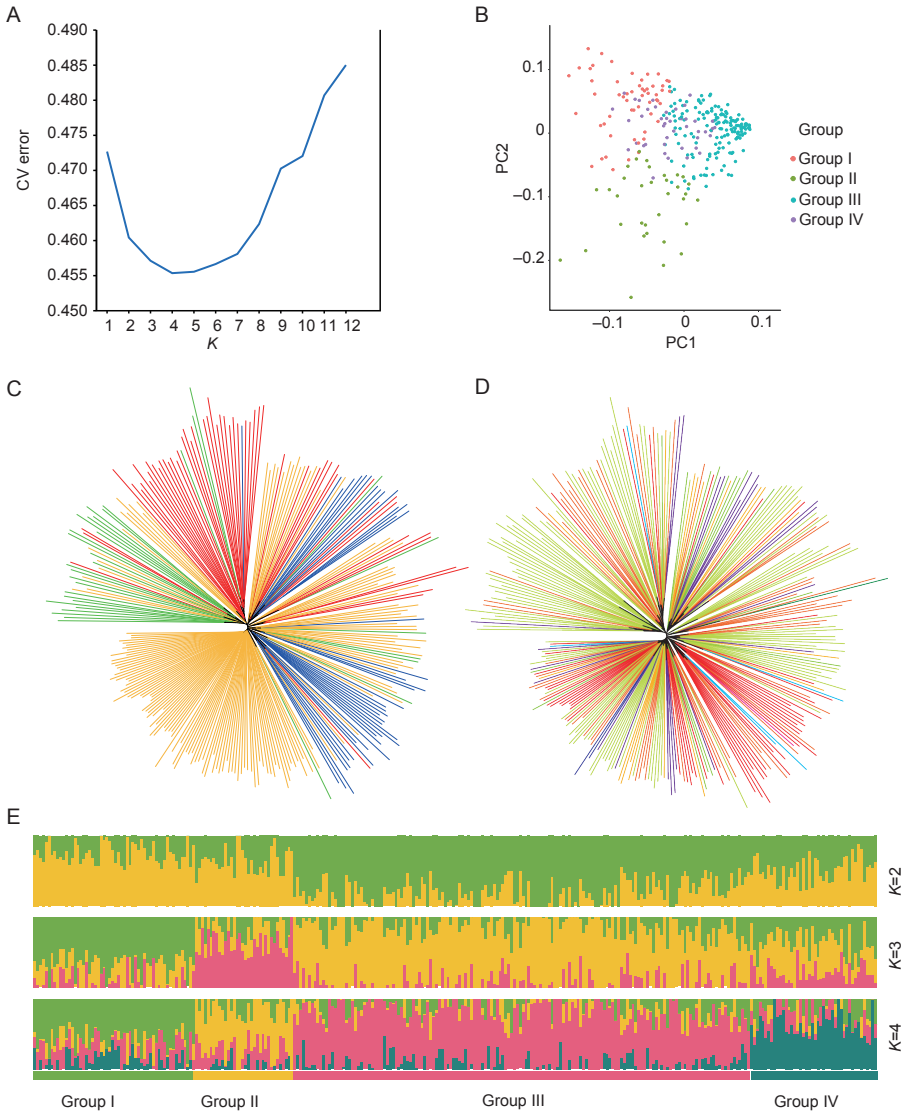


Fig. 2 Population structure of 292 resequenced sugarcane germplasm accessions. A, cross-validation (CV) error from ADMIXTURE analysis across *K*=1–12. B, population structure based on *K*=4. The entire 292 sugarcane germplasms were divided into four subgroups. C, neighbor-joining phylogenetic tree, with subgroups colored: red (Group I), green (Group II), yellow (Group III), blue (Group IV). D, geographic origins of accessions mapped onto the phylogenetic tree: red (Australia), orange (Brazil), green-yellow (China), lawn green (France), cyan (India), blue (Indonesia), indigo (Philippines), deep pink (Reunion), orange-red (USA). E, ADMIXTURE plot of 292 sugarcane germplasms.

during the infection process of *Xanthomonas oryzae* pv. *oryzae* strain PXO99A. WANG *et al.* (2015) previously concluded through their research that the ENOD93 gene is a candidate gene for enhancing resistance to the brown planthopper (*Nilaparvata lugens*) in Bt transgenic rice. It was found that the expression of *AtENOD93* is upregulated under hypoxic conditions in *Arabidopsis thaliana* (Branco-Price 2005).

3.5. GWAS for stem diameter in sugarcane

We employed the same method as that used for the GWAS of sucrose content in sugarcane to conduct the GWAS of stem diameter in sugarcane (Fig. 4-C and D). A total of 19 SNP sites (or those adjacent to them) were identified to be associated with stem diameter (Appendix E), as were specific mutation positions of different SNPs on different chromosomes (including their starting and ending points) and their p-value values. Among these sites, we screened nine candidate genes (Appendix F). Candidate genes on chromosome 2 were related to potassium transporter, transposition, RNA-mediated, and domain of unknown function (DUF4283) proteins. However, genes related to chromosome 4 were mainly related to CLIP-associated proteins. Genes on chromosome 5 were mainly involved in protein translocase subunit secA and disease resistance protein (RPM1). *RPM1*, as a *CC-NBS-LRR* type protein kinase, specifically recognizes effector proteins released by pathogens, thereby triggering effector-triggered immunity and the hypersensitive response to defend against diseases. Genes on chromosome 7 belong to the major facilitator superfamily and dienelactone hydrolase family. Additionally, one gene of unknown function occurred on chromosome 5, requiring further verification.

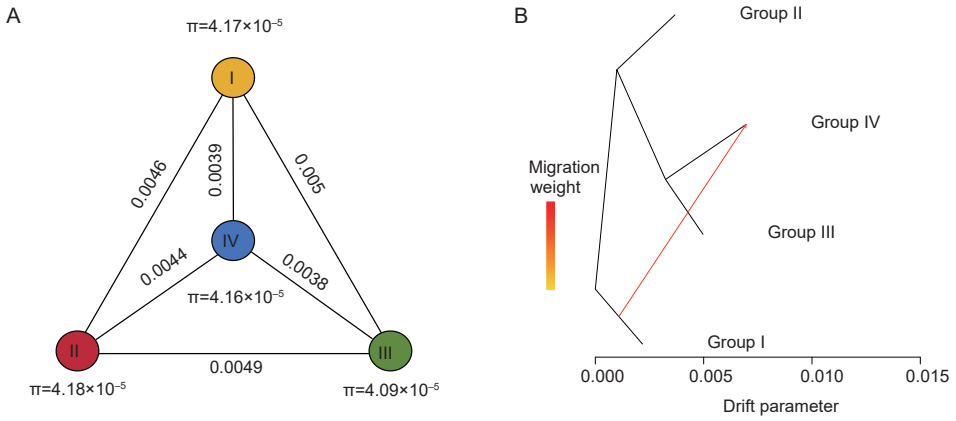


Fig. 3 Analysis of genetic diversity in 292 sugarcane germplasm. A, FST and π values of groups I–IV. B, TreeMix population graph of groups I–IV.

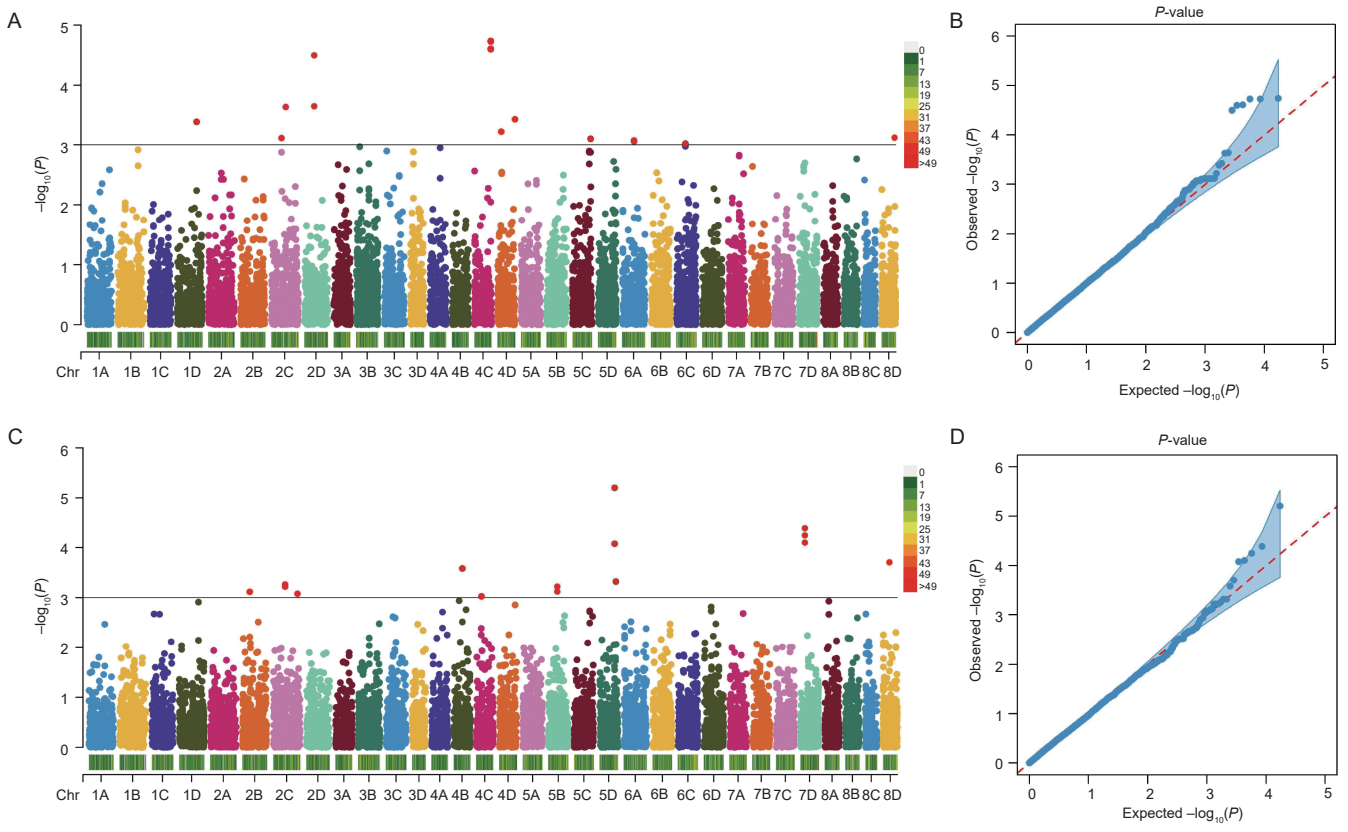


Fig. 4 Genome-wide association study (GWAS) for sucrose content and stem diameter in 292 sugarcane germplasm. A and B, Manhattan plot (A) and quantile-quantile (Q-Q) plot (B) of GWAS for sucrose content. C and D, Manhattan plot (C) and Q-Q plot (D) of GWAS for stem diameter.

4. Discussion

4.1. Analysis of genetic characteristics of stem diameter traits in sugarcane

Molecular marker-assisted selection significantly enhances breeding efficiency. Hyper-seq — a next-generation genome resequencing technique — simplifies the complexity of genomes. It is a powerful tool for identifying evolutionary origins, genotyping, genetic mapping, and quantitative trait loci identification (Zou and Xia 2022). We report stem diameter traits measured from 292 germplasm resources over a two-year period to conform to a normal distribution, and have a genetic mode characteristic of quantitative traits similar to that reported in other studies (Aitken *et al.* 2008; Barbosa *et al.* 2017). Similar results have also been reported in genetic studies on crops such as sorghum (Kong *et al.* 2020), cucumber (Yang *et al.* 2022), and broad bean (Sun *et al.* 2021). Furthermore, we report stem diameter to be mainly influenced by genetic factors, and that it can be

genetically improved by selective breeding using superior parental plants (Aitken *et al.* 2008).

By categorizing 192 sugarcane germplasms into seven subgroups we provide a valuable reference for the cross-utilization of sugarcane varieties. However, we observed no clear differentiation among sugarcane samples from different regions, nor a significant correlation between germplasm and geographical origin. This result is consistent with earlier reports that most sugarcane germplasm is generated through intervarietal hybridization and backcrossing involving a limited number of original species (Zhang J *et al.* 2018). Most cultivated sugarcane cultivars are offspring of *S. officinarum* and *S. spontaneum* hybrids; modern sugarcane cultivars are obtained through multi-generation backcrosses with tropical species (Hu *et al.* 2018; Zhang *et al.* 2021). Therefore, the 292 sugarcane germplasm resources from nine countries that we examined may include genetic lineages derived from ancestral cultivars and tropical species. This may explain why achieving significant increases in yield, sucrose content, and disease resistance has proven challenging in recent breeding endeavors.

4.2. Analysis of sugarcane population structure and genetic diversity

As an important sugar and energy crop, sugarcane provides ~80% of the world's sugar, and more than 92% of China's sugar industry. Modern cultivated sugarcane is an aneuploid allopolyploid crop, an artificially hybridized cross between *S. officinarum* and *S. spontaneum* more than 100 years ago (~100–130 chromosomes, of which 70–80% are from *S. officinarum* and 10–20% are from *S. spontaneum*). The genetic background and genetic relationships between sugarcane cultivars are complex, with up to 12 sets of chromosomes, a large genome (~10 Gb) (D'Hont 2005), and high heterozygosity. Yang *et al.* (2017, 2019) selectively sequenced most sugarcane gene-coding regions to a sequencing depth of 99X and allelic dose resolution, distinguished most genotypes, and speculated that as many as 13 genotypes existed with sequence variation of biallelic genes. However, development and use of sugarcane-related genome resources remain limited, with considerable potential for development.

We performed simplified resequencing on 292 sugarcane varieties, and obtained 163,379 high quality SNPs and InDels through comparison with reference genomes. Based on these SNPs, phylogenetic analysis, PCA, and population structure analysis, these germplasms were divided into four groups (I–IV). The genetic diversity index of the sugarcane population (4.15×10^{-5}) is rich. Most FST between groups ranged 0.038–0.05, and most differentiation between subgroups was low, caused by the relative closeness of samples. After long-term natural selection and artificial domestication of crops, natural populations of different ecological types have formed complex population structures, which are critical in association population analysis and are an essential factor causing false positives in association analysis.

Linkage imbalance is an important parameter in genome-wide association and population evolutionary selection analysis, determining the accuracy of marker density and results. Influencing factors include genetic and non-genetic factors. Mutation and recombination are the most important, including reproduction mode, genetic drift, and selection. We report an attenuation distance in the sugarcane population of 18.9 kb, with the LD of each subgroup at 17.5 and 35.6 kb, indicating that the attenuation distance between subgroups is quite different. While this attenuation distance of sugarcane is smaller than that of cultivated rice (123 kb) (Huang *et al.* 2010) and soybean (133 kb) (Zhou *et al.* 2015) and slightly smaller than that of cultivated maize (30 kb) (Hufford *et al.* 2012), it is greater than a maize inbred line population (1.5 kb) (Andersen *et al.* 2005), potato (1 kb) (Simko 2004), and *Vitis amurensis* (10 kb) (Wang Y *et al.* 2021).

4.3. Candidate genes

The sugarcane AP85-441 genome was used as the reference in this study. Through GWAS, significant variation sites are mined to analyze reasonably and identify candidate genes. SNPs in annotated loci were obtained by combining SNP genomic positions. Because sugarcane is a first-generation alternative energy crop to produce ethanol biofuel, studying genes related to sucrose content in sugarcane is of commercial significance. GWAS identified 17 candidate genes associated with sucrose and 9 candidate genes associated with stem diameter. Candidate genes on chromosome 8 are related to early nodulin 93 (ENOD93) protein, and those on chromosome 6 are mainly related to potassium transporter proteins and others; those on chromosome 2 are related to potassium transporter, transposition, RNA-mediated, and domain of unknown function (DUF4283) proteins. Genes on chromosome 4 are mainly related to CLIP-associated proteins. Genes encoding the protein translocase subunit SecA and the disease resistance protein RPM1 are located on chromosome 5.

In modern sugarcane varieties, their genetic background exhibits a complex composition. Approximately 70–80% of the genetic information originates from *S. officinarum*, 10–23% comes from *S. spontaneum*, and another 5–17% is the result of recombination between these two sugarcane species. Previous studies have demonstrated that an increase in the genetic component of *S. officinarum* in sugarcane cultivars is closely associated with an improvement in sugar-related traits and a reduction in fiber-related traits. Specifically, the excellent gene resources carried by *S. officinarum* play a crucial role in enhancing sucrose content and improving sucrose quality in sugarcane. On the other hand, *S. spontaneum* has contributed a wealth of genes related to disease resistance and stress tolerance traits to sugarcane hybrids. These genes endow sugarcane with stronger adaptability in the face of various biotic and abiotic stresses. Researchers discovered a significant candidate gene, Sh_Ss10A0001506, in their research on the genetic variation of sucrose-related traits. This gene encodes an NB-ARC domain protein, and its uniqueness lies in the fact that it originates from *S. spontaneum* but is absent in *S. officinarum*. This finding contradicts our common

understanding that *S. spontaneum* mainly contributes disease-resistance and stress-tolerance traits to modern sugarcane varieties (Zhang *et al.* 2025). Moreover, there is a significant correlation between sucrose accumulation and viral infection. This discovery has sparked in-depth thinking about the genetic patterns of resistance-related genes localized in traits such as sucrose content and stem diameter. Whether these genes follow genetic rules similar to those of the *Sh_Ss10A0001506* gene warrants further systematic research. In the study of sucrose contents, we have localized multiple ENOD93 genes. Numerous reports have indicated that these ENOD93 genes play important roles in the plant's response to abiotic stress. This suggests that these genes may, when sugarcane is exposed to adverse environmental conditions, maintain its physiological balance by regulating their own expression levels, thereby influencing sucrose synthesis and accumulation. Meanwhile, we have also identified a resistance gene, RPM1, which is significantly associated with stem diameter. This gene may influence stem diameter by regulating processes such as cell division and elongation during sugarcane stem development. Additionally, it may also be involved in sugarcane's defense responses, enhancing its resistance to pests and diseases.

In the research on sucrose content and stem diameter, we observed common characteristics exhibited by the aforementioned genes. This phenomenon reinforces a research paradigm: in the highly polyploid system of sugarcane, defense-related genes may have pleiotropic effects on core agronomic traits. In other words, a single gene may be involved in the regulation of multiple traits simultaneously, or there may exist a complex interaction network among these genes that collectively influences the growth, development, yield, and quality of sugarcane. On the other hand, this phenomenon may also imply that there is a trade - off in physiological resource allocation in sugarcane. For instance, when sugarcane allocates more resources to defense responses, it may, to a certain extent, affect sucrose synthesis and accumulation. Conversely, when sugarcane focuses on increasing sucrose yield, its stress resistance may be somewhat compromised. In summary, our research findings further strengthen the view that balancing the relationship between defense-related genes and core agronomic traits, as well as reasonably regulating the allocation of physiological resources, are the core challenges in sugarcane genetic improvement. In the future, we need to comprehensively employ various research approaches to deeply uncover the intrinsic connections between sugarcane genes and traits, thereby providing a solid theoretical foundation and technical support for the genetic improvement of sugarcane varieties.

5. Conclusion

Analysis of genetic loci controlling sucrose content will benefit breeding programs for developing new high-sucrose sugarcane varieties. We used high-quality sugarcane germplasm from various countries to establish a diverse germplasm population. We then performed GWAS on this population using high-density Hyper-seq based on DNA markers. We successfully identified 46 SNP loci and 26 candidate genes that exhibited significant associations with sucrose content and stem diameter, respectively. Furthermore, we categorized 292 sugarcane germplasms into 4 subgroups, providing a valuable reference for marker-assisted selection and the utilization of sugarcane resources. We report the performance of agronomic traits in elite sugarcane germplasm and provide information that will aid future parental selection in breeding programs. Associated markers could be further used after validation to facilitate the selection of desirable sugarcane parental materials and progeny.

Acknowledgements

This work was supported by the National Natural Science Foundation of China (32060505, 31660418), the Project of National Key Laboratory for Tropical Crop Breeding, China (NKLTCB202318), the Yunnan Seed Laboratory, China (202205AR070001-13), the Technology Innovation Talents in Yunnan Province, China (202305AD160041), the National Sugar Industry Technology System, China (CARS-170101) and the High-performance Computing Platform of YaZhou Bay Science and Technology City Advanced Computation Center, China.

Declaration of competing interest

The authors declare that they have no conflict of interest.

Declaration of generative AI and AI-assisted technologies in the writing process

During the preparation of this work, the authors used DeepSeek to assist with language polishing and conceptual clarification. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the published article.

Appendices associated with this paper are available at https://doi.org/10.1016/j.jia.2025.12.044

References

Aitken K S. 2022. History and development of molecular markers for sugarcane breeding. *Sugar Tech*, **24**, 341–353.

Aitken K S, Hermann S, Karno K, Bonnett G D, Mcintyre L C, Jackson P A. 2008. Genetic control of yield-related stalk traits in sugarcane. *Theoretical and Applied Genetics*, **117**, 1191–1203.

Alexander D H, Novembre J, Lange K. 2009. Fast model-based estimation of ancestry in unrelated individuals. *Genome Research*, **19**, 1655–1664.

Andersen J R, Schrag T, Melchinger A E, Zein I, Lubberstedt T. 2005. Validation of Dwarf8 polymorphisms associated with flowering time in elite european inbred lines of maize (*Zea mays* l.). *Theoretical and Applied Genetics*, **111**, 206–217.

Atkin F C, Dieters M J, Stringer J K. 2009. Impact of depth of pedigree and inclusion of historical data on the estimation of additive variance and breeding values in a sugarcane breeding program. *Theoretical and Applied Genetics*, **119**, 555–565.

Barbosa R P, Neto F A, Gravina L M, Gravina G A, Portela M G , Bezerra A A. 2017. Early selection of sugarcane using path analysis. *Genetics and Molecular Research*, **16**, 1–8.

Battenfield S D, Sheridan J L, Silva L, Miclaus K J, Dreisigacker S, Wolfinger R D , Pena R J, Singh R P, Jackson E W, Fritz A K, Guzman C, Poland J A. 2018. Breeding-assisted genomics:

Applying meta-GWAS for milling and baking quality in CIMMYT wheat breeding program. *PLoS ONE*, **13**, e0204757.
Bradbury P J, Zhang Z, Kroon D E, Casstevens T M, Ramdoss Y, Buckler E S. 2007. Tassel: Software for association mapping of complex traits in diverse samples. *Bioinformatics*, **23**, 2633–2635.
Branco-Price C, Kawaguchi R, Ferreira R B, Bailey-Serres J. 2005. Genome-wide analysis of transcript abundance and translation in Arabidopsis seedlings subjected to oxygen deprivation. *Annals of Botany*, **96**, 647–660.
Brown A H, Daniels J, Latter B D. 1968. Quantitative genetics of sugarcane: I. Analysis of variation in a commercial hybrid sugarcane population. *Theoretical and Applied Genetics*, **38**, 361–369.
Chandra A, Verma P K, Islam M N, Grisham M P, Jain R, Sharma A, Roopendra K, Singh K, Singh P, Verma I, Solomon S. 2015. Expression analysis of genes associated with sucrose accumulation in sugarcane (*Saccharum* spp. Hybrids) varieties differing in content and time of peak sucrose storage. *Plant Biology*, **17**, 608–617.
Chen Z L, Qin C X, Wang M, Liao F, Liao Q, Liu X H, Li Y R, Lakshmanan P, Long M H, Huang D L. 2019. Ethylene-mediated improvement in sucrose accumulation in ripening sugarcane involves increased sink strength. *BMC Plant Biology*, **19**, 1–17.
Cingolani P, Platts A, Wang L L, Coon M, Nguyen T, Wang L, Land S J, Lu X, Ruden D M. 2012. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of drosophila melanogaster strain w1^{118}; Iso-2; Iso-3. *Fly*, **6**, 80–92.
Cursi D E, Gazaffi R, Hoffmann H P, Brasco T L, Do A L, Dourado N D. 2021. Novel tools for adjusting spatial variability in the early sugarcane breeding stage. *Frontiers in Plant Science*, **12**, 749533.
Danecek P, Auton A, Abecasis G, Albers C A, Banks E, Depristo M A, Handsaker R E, Lunter G, Marth G T, Sherry S T, Mcvean G, Durbin R. 2011. The variant call format and VCFtools. *Bioinformatics*, **27**, 2156–2158.
D'Hont A. 2005. Unraveling the genome structure of polyploids using fish and gish; Examples of sugarcane and banana. *Cytogenetic and Genome Research*, **109**, 27–33.
D'Hont A, Ison D, Alix K, Roux C, Glaszmann J C. 1998. Determination of basic chromosome numbers in the genus *Saccharum* by physical mapping of ribosomal RNA genes. *Genome*, **41**, 221–225.
Emebiri L C, Matassa V, Moody D B. 2005. GENSTAT programs for performing Muir's alternative partitioning of genotype-by-environment interaction. *Journal of Heredity*, **96**, 78–79.
Ferrara C, Martella F, Vichi M. 2019. Probabilistic disjoint principal component analysis. *Multivariate Behavioral Research*, **54**, 47–61.
Fitak R R. 2021. Optm: Estimating the optimal number of migration edges on population trees using treemix. *Biology Methods & Protocols*, **6**, bpab017.
Garsmeur O, Droc G, Antonise R, Grimwood J, Potier B, Aitken K, Jenkins J, Martin G, Charron C, Hervouet C, Costet L, Yahiaoui N, Healey A, Sims D, Cherukuri Y, Sreedasyam A, Kilian A, Chan A, Van S M, Swaminathan K, *et al.* 2018. A mosaic monoploid reference sequence for the highly complex genome of sugarcane. *Nature Communications*, **9**, 2638.
Hu W, Hua X, Zhang Q, Wang J, Shen Q, Zhang X, Wang K, Yu Q, Lin Y, Ming R, Zhang J, 2018. New insights into the evolution and functional divergence of the sweet family in *Saccharum* based on comparative genomics. *BMC Plant Biology*, **18**, 270.
Huang X, Wei X, Sang T, Zhao Q, Feng Q, Zhao Y, Li C, Zhu C, Lu T, Zhang Z, Li M, Fan D, Guo Y, Wang A, Wang L, Deng L, Li W, Lu Y, Weng Q, Liu K, *et al.* 2010. Genome-wide association studies of 14 agronomic traits in rice landraces. *Nature Genetics*, **42**, 961–967.
Hufford M B, Xu X, van Heerwaarden J, Pyhajarvi T, Chia J M, Cartwright R A, Elshire R J, Glaubitz J C, Guill K E, Kaeppler S M, Lai J, Morrell P L, Shannon L M, Song C, Springer N M, Swanson-Wagner R A, Tiffin P, Wang J, Zhang G, Doebley J, *et al.* 2012. Comparative population genomics of maize domestication and improvement. *Nature Genetics*, **44**, 808–811.
Jannoo N, Grivet L, Chantret N, Garsmeur O, Glaszmann J C, Arruda P, D'Hont A. 2007. Orthologous comparison in a gene-rich region among grasses reveals stability in the sugarcane polyploid genome. *Plant Journal*, **50**, 574–585.
Jannoo N, Grivet L, David J, D'Hont A, Glaszmann J C. 2004. Differential chromosome pairing affinities at meiosis in polyploid sugarcane revealed by molecular markers. *Heredity*, **93**, 460–467.
Job D. 2002. Plant biotechnology in agriculture. *Biochimie*, **84**, 1105–1110.
Jolliffe I T, Cadima J. 2016. Principal component analysis: A review and recent developments. *Philosophical Transactions of the Royal Society A*, **374**, 20150202.
Khanbo S, Tangphatsornruang S, Piriyapongsa J, Wirojsirasak W, Punpee P, Klomsa-Ard P, Ukoskit K. 2021. Candidate gene association of gene expression data in sugarcane contrasting for sucrose content. *Genomics*, **113**, 229–237.
Kong W, Jin H, Goff V H, Auckland S A, Rainville L K, Paterson A H. 2020. Genetic analysis of stem diameter and water contents to improve sorghum bioenergy efficiency. *G3*: *Genes Genomes Genetics*, **10**, 3991–4000.
Korte A, Farlow A. 2013. The advantages and limitations of trait analysis with GWAS: A review. *Plant Methods*, **9**, 29.
Langmead B, Salzberg S L. 2012. Fast gapped-read alignment with Bowtie 2. *Nature Methods*, **9**, 357–359.
Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. 2009. The sequence alignment/map format and samtools. *Bioinformatics*, **25**, 2078–2079.
Li Q, Song X, Meng X, Zhang J, Zhang M, Chen L, Li J, Wang B. 2025. Shaping future sugarcane: Ideal plant architecture and breeding strategies. *Molecular Plant*, **18**, 725–728.
Li X, Singh J, Qin M, Li S, Zhang X, Zhang M, Khan A, Zhang S, Wu J. 2019. Development of an integrated 200K SNP genotyping array and application for genetic mapping, genome assembly improvement and genome wide association studies in pear (*Pyrus*). *Plant Biotechnology Journal*, **17**, 1582–1594.
Liu Y, Song H, Zhang M, Yang D, Deng X, Sun H, Liu J, Yang M. 2022. Identification of qtls and a putative candidate gene involved in rhizome enlargement of Asian lotus (*Nelumbo nucifera*). *Plant Molecular Biology*, **110**, 23–36.
Mirajkar S J, Suprasanna P, Vaidya E R. 2016. Spatial distribution and dynamics of sucrose metabolising enzymes in radiation induced mutants of sugarcane. *Plant Physiology and Biochemistry*, **100**, 85–93.
Mirzawan P D, Cooper M, Delacy I H, Hogarth D M. 1994. Retrospective analysis of the relationships among the test environments of the southern queensland sugarcane breeding programme. *Theoretical and Applied Genetics*, **88**, 707–716.
Murray M G, Thompson W F. 1980. Rapid isolation of high molecular weight plant DNA. *Nucleic Acids Research*, **8**, 4321–4325.
Peng L, Xiao H, Li R, Zeng Y, Gu M, Moran N, Yu L, Xu G. 2023. Potassium transporter OsHAK18 mediates potassium and sodium circulation and sugar translocation in rice. *Plant Physiology*, **193**, 2003–2020.
Piperidis N D, Hont A. 2020. Sugarcane genome architecture decrypted with chromosome-specific oligo probes. *The Plant Journal*, **103**, 2039–2051.
Pongpanich M, Sullivan P F, Tzeng J Y. 2010. A quality control algorithm for filtering SNPs in genome-wide association studies. *Bioinformatics*, **26**, 1731–1737.
Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira M, Bender D, Maller J, Sklar P, De Bakker P, Daly K. 2007. PLINK: a tool set for whole-genome association and population-based linkage analyses. *The American Journal of Human Genetics,* **81**, 559–575.
Sabzehzari M, Naghavi M R. 2019. Phyto-miRNA: A molecule with beneficial abilities for plant biotechnology. *Gene*, **683**, 28–34.
Silva L A, Resende R T, Ferreira R A, Silva G N, Kist V, Barbosa M H, Nascimento M, Bhering L L. 2016. Selection index using the graphical area applied to sugarcane breeding. *Genetics and Molecular Research*, doi: 10.4238/gmr.15038711.
Simko I. 2004. One potato, two potato: Haplotype association mapping in autotetraploids. *Trends in Plant Science*, **9**, 441–448.
de Souza M C P, Silva J N, Almeida C. 2013. Differential detection of transposable elements between *Saccharum* species. *Genetics and Molecular Biology*, **36**, 408–412.
Sun C Y, Yang Y M, Jia L, Liu X Q, Xu H Q, Lv H Y, Huang Z W, Zhang D. 2021. QTL mapping of the genetic basis of stem diameter in soybean. *Planta*, **253**, 1–12.
Verma K K, Song X P, Yadav G, Degu H D, Parvaiz A, Singh M, Huang H R, Mustafa G, Xu L, Li Y R. 2022. Impact of agroclimatic variables on proteogenomics in sugar cane (*Saccharum* spp.) Plant productivity. *ACS Omega*, **7**, 22997–23008.
Waclawovsky A J, Sato P M, Lembke C G, Moore P H, Souza G M.

2010. Sugarcane for bioenergy production: An assessment of yield and regulation of sucrose content. *Plant Biotechnology Journal*, **8**, 263–276.

Wang F, Ning D, Chen Y, Dang C, Han N S, Liu Y E, Ye G Y. 2015. Comparing gene expression profiles between Bt and non-Bt rice in response to brown planthopper infestation. *Frontiers in Plant Science*, **6**, 1181.

Wang F, Zou M, Zhao L, Xia Z, Wang J. 2021. Genome-wide association mapping of late blight tolerance trait in potato (*Solanum tuberosum* l.). *Frontiers in Genetics*, **12**, 714575.

Wang K, Cheng H, Han J, Esh A, Liu J, Zhang Y, Wang B. 2022. A comprehensive molecular cytogenetic analysis of the genome architecture in modern sugarcane cultivars. *Chromosome Research*, **30**, 29–41.

Wang H, Pan Y B, Wu M, Liu J, Yang S, Wu Q, Que Y. 2024. Sugarcane genetics: Underlying theory and practical application. *The Crop Journal*, **13**, 328–338.

Wang Y, Xin H, Fan P, Zhang J, Liu Y, Dong Y, Wang Z, Yang Y, Zhang Q, Ming R, Zhong G Y, Li S, Liang Z. 2021. The genome of Shanputao (*Vitis amurensis*) provides a new insight into cold tolerance of grapevine. *Plant Journal*, **105**, 1495–1506.

Wen M, Wang H, Xia Z, Zou M, Lu C, Wang W. 2010. Developmenrt of EST-SSR and genomic-SSR markers to assess genetic diversity in *Jatropha curcas* L. *BMC Research Notes*, **3**, 1–8.

Xia Z, Liu K, Zhang S, Yu W, Zou M, He L, Wang W. 2018a. An ultra-high density map allowed for mapping QTL and candidate genes controlling dry latex yield in rubber tree. *Industrial Crops and Products*, **120**, 351–356.

Xia Z, Zhang S, Wen M, Lu C, Sun Y, Zou M, Wang W. 2018b. Construction of an ultrahigh-density genetic linkage map for *Jatropha curcas* L. and identification of QTL for fruit yield. *Biotechnology for Biofuels*, **11**, 1–10.

Xia Z, Zou M, Zhang S, Feng B, Wang W. 2014. AFSM sequencing approach: A simple and rapid method for genome-wide SNP and methylation site discovery and genetic mapping. *Scientific Reports*, **4**, 7300.

Yang X, Song J, Todd J, Peng Z, Paudel D, Luo Z, Ma X, You Q, Hanson E, Zhao Z, Zhao Y, Zhang J, Ming R, Wang J. 2019. Target enrichment sequencing of 307 germplasm accessions identified ancestry of ancient and modern hybrids and signatures of adaptation and selection in sugarcane (*Saccharum* spp.), a 'sweet' crop with 'bitter' genomes. *Plant Biotechnology Journal*, **17**, 488–498.

Yang X, Song J, You Q, Paudel D R, Zhang J, Wang J. 2017. Mining sequence variations in representative polyploid sugarcane germplasm accessions. *BMC Genomics*, **18**, 1–16.

Yang Y, Dong S, Miao H, Liu X, Dai Z, Li X, Gu X, Zhang S. 2022. Genome-wide association studies reveal candidate genes related to stem diameter in cucumber (*Cucumis sativus* L.). *Genes*, **13**, 1095.

Zhang H, Lin P, Liu Y, Huang C, Huang G, Jiang H, Xu L, Zhang M, Deng Z, Zhao X. 2022. Development of SLAF-sequence and multiplex SNaPshot panels for population genetic diversity analysis and construction of DNA fingerprints for sugarcane. *Genes*, **13**, 1477.

Zhang J, Qi Y, Hua X, Wang Y, Wang B, Qi Y, Huang Y, Yu Z, Gao R, Zhang Y, Wang T, Wang Y, Mei J, Zhang Q, Wang G, Pan H, Li Z, Li S, Liu J, Qi N, *et al.* 2025. The highly allo-autopolyploid modern sugarcane genome and very recent allopolyploidization in *Saccharum*. *Nature Genetics*, **57**, 242–253.

Zhang J, Zhang X, Tang H, Zhang Q, Hua X, Ma X, Zhu F, Jones T, Zhu X, Bowers J, Wai C M, Zheng C, Shi Y, Chen S, Xu X, Yue J, Nelson D R, Huang L, Li Z, Xu H, *et al.* 2018. Allele-defined genome of the autopolyploid sugarcane *Saccharum spontaneum* L. *Nature Genetics*, **50**, 1565–1573.

Zhang Q, Hua X, Liu H, Yuan Y, Shi Y, Wang Z, Zhang M, Ming R, Zhang J. 2021. Evolutionary expansion and functional divergence of sugar transporters in *Saccharum* (*S. spontaneum* and *S. officinarum*). *The Plant Journal*, **105**, 884–906.

Zhang S, Chen X, Lu C, Ye J, Zou M, Lu K, Feng S, Pei J, Liu C, Zhou X, Ma P, Li Z, Liu C, Liao Q, Xia Z, Wang W. 2018. Genome-wide association studies of 11 agronomic traits in cassava (*Manihot esculenta* Crantz). *Frontiers in Plant Science*, **9**, 503.

Zhou Y, Pan H. 2023. Specific-locus amplified fragment sequencing (SLAF-Seq). *Methods in Molecular Biology*, doi: 10.1007/978-1-0716-3024-2_11.

Zhou Z, Jiang Y, Wang Z, Gou Z, Lyu J, Li W, Yu Y, Shu L, Zhao Y, Ma Y, Fang C, Shen Y, Liu T, Li C, Li Q, Wu M, Wang M, Wu Y, Dong Y, Wan W, *et al.* 2015. Resequencing 302 wild and cultivated accessions identifies genes related to domestication and improvement in soybean. *Nature Biotechnology*, **33**, 408–414.

Zou M, Lu C, Zhang S, Chen Q, Sun X, Ma P, Hu M, Peng M, Ma Z, Chen X, Zhou X, Wang H, Feng S, Fang K, Xie H, Li Z, Liu K, Qin Q, Pei J, Wang S, *et al.* 2017. Epigenetic map and genetic map basis of complex traits in cassava population. *Scientific Reports*, **7**, 41232.

Zou M, Xia Z. 2022. Hyper-seq: A novel, effective, and flexible marker-assisted selection and genotyping approach. *The Innovation*, **3**, 100254.

Executive Editor-in-Chief Sanwen Huang
Managing Editor Lingyun Weng