nature portfolio

Corresponding author(s): Xingtan Zhang

Last updated by author(s): Apr 3, 2026

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☒	☐	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection

Pair-end re-sequencing was performed using BGI DNBSEQ-T7 platform;
HIFI reads and Isoform sequences were generated using PacBio Revio platform;
Ultra-long Nanopore sequencing was performed using Promethion platform;
Pore-C sequencing was performed using Oxford Nanopore Technologies;
The parenchyma cell statistics were obtained using our custom code (https://github.com/Chaiyuangungun/Automated-Parenchyma-Cell-Analysis-in-Sugarcance-Stems).

Data analysis	Genome assembly, assessment, and annotation: hifiasm(0.19.1-r559), C-Phasingv(v0.2.0. r273), Juicebox (v2.17.00), BUSCO(v5.4.3), Meryl (v1.4.1), Merqury (v1.3), LTR_FINDER_parallel (v1.2), LTR_HARVEST_parallel (v1.1), LTR_retriever (v2.9.6), blastn (v2.9.0), Mosdepth (v0.3.9), quarTeT (v1.2.2), GETA (https://github.com/chenlianfu/geta), AUGUSTUS (v.3.3.3), GENEWISE (v.2.4.1), EDTA (2.3.0), RepeatMasker (http://www.repeatmasker.org), RepeatModeler2(https://github.com/Dfam-consortium/RepeatModeler), DIAMOND (v0.7.10), JCVI (v1.4.18), KMC(v3.2.4), Broccoli (https://github.com/rderelle/Broccoli). Allele-specific expression analysis: ccs tool (https://github.com/PacificBiosciences/ccs), skera (https://github.com/PacificBiosciences/skera), lima (https://github.com/PacificBiosciences/barcoding), isoseq refine (https://github.com/PacificBiosciences/IsoSeq), mmseqs (v15.6f452), minimap2 (v2.24-r1122), bedtools (v2.31.1), Allele-Express (https://github.com/xiaochen-eden/Allele-Express). Population analysis: GTX (v2.1.0) (http://www.gtxlab.com/en/product/cat), VCF2Dis (v1.52), Admixture (v1.3.0), PopLDdecay (v3.42), VCF2PCACluster (v1.40), IBDseq (r1206), SweeD (v4.0.0), RAiSD (v2.9), VCFtools (v0.1.16), XP-CLR (https://github.com/hardingnj/xpclr), bcftools (v1.10.2), SnpEff (v5.2c). GWAS: KMERIA (https://github.com/Sh1ne111/KMERIA), CAVIAR (http://genetics.cs.ucla.edu/caviar/). Parenchyma cell statistics: Cellpose, custom code (https://github.com/Chaiyuangungun/Automated-Parenchyma-Cell-Analysis-in-Sugarcance-Stems).

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

The raw sequencing data has been deposited in NCBI database under accession number of PRJNA1303125. The genome assembly and annotation are available at figshare database (https://figshare.com/s/e79249d6ce11539fa105) or NCBI database under accession number of PRJNA1312687.

Research involving human participants, their data, or biological material

Policy information about studies with human participants or human data. See also policy information about sex, gender (identity/presentation), and sexual orientation and race, ethnicity and racism.

Reporting on sex and gender	Not applicable
Reporting on race, ethnicity, or other socially relevant groupings	Not applicable
Population characteristics	Not applicable
Recruitment	Not applicable
Ethics oversight	Not applicable

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	In this study, a total of 981 sugarcane accessions were collected from 19 countries or regions, representing a broad range of genetic and phenotypic variation. This collection comprised 613 hybrid cultivars, 78 Saccharum officinarum, and 290 Saccharum spontaneum, constituting the largest known sugarcane population sufficient for comprehensive genetic analysis. To assess the contribution of POJ2878 to sugarcane breeding both within and outside China, we selected 573 cultivated varieties with well-defined geographic origins, including 327 Chinese cultivars (bred in China) and 246 non-Chinese cultivars (bred outside China). Due to the difficulty in cross-sectioning of hard and thick sugarcane stalks, a subset of 265 hybrid cultivars was randomly selected for detailed sugar content analysis, parenchyma cell statistics, and subsequent GWAS analysis. 43 S. robustum accessions used for the selective sweep analysis of S. officinarum were retrieved from the public database NCBI database under BioProjects PRJNA1228676 and PRJNA1209834.

Data exclusions	No data was excluded from the analyses.
Replication	We confirmed that all attempts at replication were successful.
Randomization	For the filed-based sugarcane accessions trials, planting plots were randomized using a complete block design to minimize environmental heterogeneity. For experiments involving stem cross-sectioning and Brix measurement, no experimental group were defined; rather, a representative set of plants at identical developmental stages and physiological ages was selected from the population to ensure phenotypic consistency. Because these experiments involved systematic phenotyping across the population (or genotype) rather than a comparative study between randomized experimental groups, group allocation was not applicable. To control for potential covariates, all selected plants were grown under identical field management practices, and measurements were performed using standardized sampling protocols (e.g., harvesting the same internode position).
Blinding	Blinding was applied during parenchyma cell analysis, Brix measurements, KMERIA-based GWAS, and independent replications of phenotypic. For transgenic experiments, blinding was not applicable because material identification and strict segregation of distinct genetic lines were required. Blinding was not possible for transgenic experiments, as the strict segregation and tracking of distinct genetic lines were required.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☐	☒ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Clinical data
☒	☐ Dual use research of concern
☐	☒ Plants

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Antibodies

Antibodies used	Anti-PAT/bar (Genesino, China)
Validation	Antibody dilution: Not applicable. We used a commercial lateral flow assay (test strip), not a Western blot. The manufacturer pre-fixes the conjugated antibodies to the strips, making them ready-to-use. The Methods section has been updated to clarify this. Validation statements and experiments can be obtained from the following websites: https://www.genesino.cn/pd.jsp?id=9&fromColId=0#_pp=0_722_2

Dual use research of concern

Policy information about dual use research of concern

Hazards

Could the accidental, deliberate or reckless misuse of agents or technologies generated in the work, or the application of information presented in the manuscript, pose a threat to:

No	Yes
☒	☐ Public health
☒	☐ National security
☒	☐ Crops and/or livestock
☒	☐ Ecosystems
☒	☐ Any other significant area

Experiments of concern

Does the work involve any of these experiments of concern:

No	Yes	
☒	☐	Demonstrate how to render a vaccine ineffective
☒	☐	Confer resistance to therapeutically useful antibiotics or antiviral agents
☒	☐	Enhance the virulence of a pathogen or render a nonpathogen virulent
☒	☐	Increase transmissibility of a pathogen
☒	☐	Alter the host range of a pathogen
☒	☐	Enable evasion of diagnostic/detection modalities
☒	☐	Enable the weaponization of a biological agent or toxin
☒	☐	Any other potentially harmful combination of experiments and agents

Plants

Seed stocks	The 981 sugarcane accessions were provided by the Yunnan Academy of Agricultural Sciences and are currently stored in the National Sugarcane Germplasm Repository, Kaiyuan City, Yunnan Province (23°30′–23°58′ N, 103°04′–103°43′ E, altitude: 1051 m).
Novel plant genotypes	Overexpression (OE) lines for both sugarcane (ROC22) and rice (ZH11) were generated using Agrobacterium tumefaciens (EHA105)-mediated transformation of embryogenic calli. For sugarcane, OE lines for ShSUT2, ShTIP1, ShCBL1, and ShATG18a were developed. For rice (ZH11), OE lines for ShTIP1, ShCBL1, and ShATG18a were developed.
Authentication	Verification of the overexpression lines was performed through quantitative RT-PCR, Semi-quantitative RT-PCR, or immunoblotting using Basta antibodies.