Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Integrative analyses of flowering-associated variation in wild Saccharum spontaneum

Sicheng Li
Guangxi University
Fengzhen Wu
Guangxi University
Lijun Zhang
Guangxi University
Zhen Wang
Guangxi University
Weixing Duan
Guangxi Academy of Agricultural Science
Xiping Yang
xipingyang@gxu.edu.cn
Guangxi University

Research Article

Keywords: Saccharum spontaneum, flowering, transcriptome, GWAS

Posted Date: October 8th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11068175/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Abstract

Flowering time is a key agronomic trait influencing environmental adaptation, reproductive success, and breeding efficiency in plants. However, the genetic basis of flowering variation in *Saccharum spontaneum*, a wild progenitor of cultivated sugarcane with rich natural diversity, remains poorly understood. Here, a core germplasm population comprising 147 accessions was characterized by integrating phenotypic evaluation, population genetics, genome-wide association study (GWAS), and transcriptomic profiling. The results revealed abundant natural variation in flowering-related traits and a genetically diverse population with extensive gene flow. GWAS identified multiple flowering-associated loci, while transcriptomic analyses across flowering developmental stages revealed coordinated biological processes associated with the vegetative-to-reproductive transition. The functional consistency between GWAS candidate genes and transcriptome-enriched pathways highlighted a set of candidate flowering-associated loci and expression signatures. These findings provide valuable genomic resources for understanding flowering variation in wild sugarcane and establish a foundation for evolutionary studies, molecular marker development, and future pre-breeding applications.

1 Introduction

Sugarcane (*Saccharum* spp.) is a globally important crop, serving as the primary source of sugar and a key raw material for ethanol production [1]. However, modern sugarcane breeding has faced challenges in recent decades due to its complex polyploid genome and narrow genetic background [2, 3]. With the release of high-quality sugarcane genomes, omics-based approaches now provide powerful tools to explore and improve agronomic traits at the molecular level [4–6]. Yet, the effective incorporation of elite traits from wild *S. spontaneum* populations into cultivated varieties represents a critical convergence point—where genomic insights meet practical breeding, bridging natural diversity and targeted crop improvement [7, 8].

Flowering is a pivotal developmental transition in sugarcane, exhibiting a clear “double-edged sword” effect on production. On one hand, it is indispensable for sexual reproduction in breeding programs, enabling the introduction of new genetic variation; on the other hand, the shift from vegetative to reproductive growth reduces sucrose accumulation and biomass, thus diminishing economic yield [9]. Sugarcane flowering is a complex trait regulated by photoperiod, temperature, and plant developmental status, with critical induction occurring under specific daylength conditions [10, 11]. Consequently, understanding and controlling flowering—through both environmental manipulation and molecular approaches—has become a central focus in sugarcane research and breeding.

Flowering time in plants is shaped not only by classical regulatory pathways, but also by long-term ecological adaptation to diverse environmental conditions. Previous studies have shown that environmental factors such as photoperiod, temperature, altitude, and seasonal climate variation can drive flowering-time divergence among natural populations through local adaptation and selection [12–17]. In addition, endogenous regulatory processes, including hormonal signaling, developmental stage

transitions, and carbohydrate metabolism, further coordinate flowering responses under different ecological conditions [18–22]. As a wild progenitor species with broad geographic and ecological distribution, *S. spontaneum* exhibits extensive variation in flowering behavior across environments, suggesting that flowering-time differentiation may represent an adaptive strategy shaped by both genetic background and environmental heterogeneity [23, 24]. However, the genetic basis underlying the relationship between environmental adaptation and flowering variation in natural populations of *S. spontaneum* remains poorly understood.

To address this gap, our study integrated flowering-time GWAS based on a core *S. spontaneum* population with developmental-stage transcriptome analyses of representative early-flowering accessions. By combining population-level genetic variation with dynamic transcriptional profiling, this multi-level approach was designed to identify flowering-associated genetic variants and expression patterns, providing a resource for candidate marker discovery relevant to flowering-time variation. This framework offers potential insights for future pre-breeding applications and the utilization of wild sugarcane germplasm in breeding programs.

2 Materials and methods

2.1 Phenotypic evaluation and genome-wide association analysis of flowering-associated traits

The core germplasm collection of *S. spontaneum* established in this study comprises a total of 147 accessions, including 16 early-flowering accessions and 131 mid- and late-flowering accessions, accounting for 10.9% and 89.1% of the total, respectively. The spatial distribution exhibits a distinct latitudinal clustering pattern. Phenotypic evaluation was performed in the core *S. spontaneum* population during the mature growth stage. Seven traits were measured, including stalk length, stalk diameter, internode number, number of leaves, single stem weight, Brix, and flowering time. All traits were evaluated under the same field conditions using representative healthy plants.All plant materials used in this study, including the 147 core germplasm accessions and the cultivar Hainan 92 − 3, were formally identified by Zhen Wang, and maintained as living germplasm at the Fusui sugarcane experimental base of Guangxi University, Chongzuo, China; no voucher specimens were deposited in a public herbarium, as this study used maintained germplasm accessions rather than newly collected wild plant material. The use of plant material complied with relevant institutional, national and international guidelines and legislation.

The filtered VCF files were converted into binary files using PLINK [25], after which principal components analysis (PCA) were calculated using PLINK and scatter plots were generated using ggplot2 [26]. Population genetic structure was comprehensively analyzed using the Admixture (v1.3.0) [27]. The VCF files were converted to MAP format using PLINK, and chromosome-specific data were extracted using the “--allow-extr-chr” command, before being converted to BED files for subsequent calculations. To assess the impact of different K values on model performance, a 10-fold cross-validation method was

employed. Based on the results of these 10 cross-validation runs, the cross-validation error rates corresponding to each K value were calculated and presented as a line chart to facilitate an intuitive comparison of model performance across different K values. Genetic diversity parameters, including fixation index (F_{st}), nucleotide diversity (π), and Tajima's D, were calculated using VCFtools [28] for different subpopulations. Phylogenetic relationships were reconstructed using the neighbor-joining (NJ) method. VCF files were converted to PHYLIP format using vcf2phylip.py, and phylogenetic trees were constructed using IQ-TREE and subsequently visualized with iTOL v5.

GWAS was performed using the 5,715,473 SNPs obtained after data cleaning. The MLM module of the EMMEA software was employed to conduct GWAS on flowering phenotype in sugarcane, with a significance threshold of $p < 1/N$ to identify significantly associated loci. Manhattan plots and Q-Q plots were generated using CMplot in R version 4.2.1. For each significant locus, the region extending 100 kb upstream and downstream was defined as the candidate interval for locating associated genes.

2.2 Transcriptomic profiling during flowering development in early-flowering *S. spontaneum*

Samples of the sugarcane cultivar Hainan 92 – 3 (an early flower accession) were collected from Guangxi University, Nanning, China (108°19'E, 22°51'N), and the Fusui sugarcane experimental base of Guangxi University, Chongzuo, China (107°54'E, 22°52'N). Samples included early-flowering non-flowering leaves (ENL), early-flowering stage leaves (EEL), mature-flowering stage leaves (EML), early-flowering stage inflorescences (EEF), and mature-flowering stage inflorescences (EMF). In the abbreviations, the first letter "E" indicates the early-flowering population, the second letter represents the developmental stage (N, non-flowering; E, early flowering; M, mature flowering), and the final letter denotes the tissue type (L, leaf; F, inflorescence). All samples were immediately frozen in liquid nitrogen and stored at -80°C until RNA extraction. The early flowering stage was defined as the stage when *S. spontaneum* initiated panicle emergence and developed the flag leaf, whereas the mature flowering stage corresponded to the rupture of the flag leaf and the onset of flowering.

Total RNA was extracted using TRIzol reagent (Invitrogen, USA) according to the manufacturer's instructions. RNA quality was assessed prior to library construction. Sequencing libraries were prepared and sequenced on the BGI platform, generating approximately 6 GB of raw data per sample. Raw reads were filtered using Fastp to remove adapters and low-quality sequences. Clean reads were aligned to the *S. spontaneum* reference genome (Np-X) [29] using HISAT2 with default parameters. Gene expression levels were quantified using Salmon (v1.9.0), and gene-level count matrices were obtained.

Differential expression analysis was performed using the DESeq2 package in R (v 4.2.1). Genes with $|\log_2\text{FoldChange}| \geq 1$ and $p \leq 0.05$ were considered as differentially expressed genes (DEGs). Visualization of differential expression results was conducted using the ggplot2 package. Functional enrichment analysis of DEGs, including Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways, was carried out using the ClusterProfiler package in R, with $p < 0.05$ considered as the threshold for significant enrichment. In addition, OmicShare tools

(https://www.omicshare.com/tools/) were used for supplementary enrichment analysis and visualization.

3 Results

3.1 Phenotypic trait correlations in a core *S. spontaneum* population

Correlation analysis revealed significant associations among several phenotypic traits evaluated in the core *S. spontaneum* population (147 accessions) **(**Fig. 1A**)**. Stalk length was positively correlated with internode number (r = 0.66) and leave number (r = 0.62), indicating that taller plants tended to have more internodes and leaves. Stalk diameter showed a strong positive correlation with single stem weight (r = 0.81), suggesting that stem thickness plays a substantial role in biomass accumulation. In addition, internode number was positively correlated with number of leaves (r = 0.67), reflecting coordinated vegetative growth across these traits. Brix exhibited moderate correlations with several growth-related traits, including stalk length (r = 0.32) and internode number (r = 0.29). In contrast, flowering time displayed only weak correlations with most other traits.

Principal component analysis (PCA) was further conducted to summarize the phenotypic variation **(**Fig. 1B**)**. The first principal component (PC1) accounted for 97.9% of the total phenotypic variance, while the second principal component (PC2) explained 1.6%. Growth-related traits, including stalk length, stalk diameter, internode number, number of leaves, single stem weight, and Brix, clustered closely, indicating strong associations among these traits. In contrast, flowering time was separated from these traits, consistent with the weak correlations observed in the correlation analysis.

3.2 Population structure, genetic diversity and geographic distribution in a core *S. spontaneum* population

Phylogenetic analysis showed no clear correspondence between geographic origin and genetic clustering **(**Fig. 2A**)**, and extensive intrapopulation hybridization was observed. The core germplasm population of *S. spontaneum* constructed in this study consisted of 147 accessions, including 16 early-flowering accessions (approximately 10.9%, including the extreme early-flowering accession Hainan 92 − 3) and 131 normal- and late-flowering accessions (approximately 89.1%). Geographic distribution analysis indicated that early-flowering accessions were mainly distributed in major sugarcane-producing regions of southern China, including Guangxi (6 accessions, 37.5%), Guangdong (3 accessions), Yunnan (3 accessions), Hainan (2 accessions), Fujian (1 accession), and Sichuan (1 accession), showing a clear latitudinal aggregation pattern **(**Fig. 2B**)**. Notably, 84.6% of the early-flowering accessions were located within the 18.5°–19.0°N latitude range, where the annual average photoperiod is 12.8–13.2 hours, suggesting that this light regime may favor their reproductive growth.

Population structure analysis based on ADMIXTURE revealed that the cross-validation error reached a minimum at K = 2, indicating that the core *Saccharum spontaneum* population can be optimally divided into two genetic subgroups **(**Fig. 2C**)**. At K = 2, most accessions were clearly assigned to either Group 1 or Group 2, whereas higher K values (K = 3 and K = 4) resulted in increased admixture, suggesting a complex genetic background and potential gene flow among subpopulations. Principal component analysis further supported this grouping pattern **(**Fig. 2D–E**)**. The two subgroups were partially separated along the first principal component, although some overlap was observed, indicating moderate genetic differentiation. The distribution of samples within each group also revealed substantial within-group variation.

Genetic diversity analysis showed that nucleotide diversity (π) differed slightly between the two subgroups, with Group 2 exhibiting a marginally higher diversity than Group 1 **(**Fig. 2F-G**)**. This suggests that Group 2 may harbor greater genetic variation, potentially contributing to phenotypic diversity within the population.

3.3 Genome-wide association analysis of flowering traits in core *S. spontaneum* population

Table 1
Flowering time associated gene of *S. spontaneum*

Gene name	Functional effects	Species
Sspon.04G0008130-1A	xyloglucan endotransglucosylase/hydrolase	[*Saccharum hybrid* cultivar ROC22]
Sspon.04G0008140-1A	Putative SET-domain containing protein family	[*Zea mays*]
Sspon.04G0008150-1T	/	/
Sspon.04G0008150-1A	/	/
Sspon.04G0008160-1A	auxin-responsive protein SAUR32-like	[*Miscanthus floridulus*]
Sspon.04G0008170-1A	E3 ubiquitin-protein ligase PRT1	[*Saccharum officinarum*]
Sspon.04G0008180-1A	/	/
Sspon.04G0020300-1A	protein trichome birefringence-like 12	[*Miscanthus floridulus*]
Sspon.04G0020310-1A	autophagy protein 5-like	[*Miscanthus floridulus*]
Sspon.04G0020320-1A	UTP–glucose-1-phosphateuridylyltransferase	[*Sorghum bicolor*]
Sspon.04G0020330-1A	peptidyl-prolyl cis-trans isomerase FKBP18	[*Miscanthus floridulus*]
Sspon.04G0020340-1A	glutamate receptor 3.1	[*Sorghum bicolor*]
Sspon.04G0020350-1A	/	/
Sspon.04G0004820-1P	/	/
Sspon.06G0007860-1A	auxin transport protein BIG-like	[*Miscanthus floridulus*]

Flowering time showed a non-normal distribution in the core *S. spontaneum* population due to the presence of extreme early-flowering accessions. A total of 5,715,473 high-quality SNP loci were obtained from the consolidated VCF file generated using bcftools, which were used for subsequent analysis. To accommodate the assumptions of the MLM model and ensure reliable detection of associated loci, flowering time was analyzed as a categorical trait (grouped as early vs. normal/late). A total of four SNPs significantly associated with flowering time were detected on chromosomes 4 and 6 **(**Fig. 3**)**. To

identify candidate genes associated with flowering time, genes located within 100 kb upstream and downstream of the four significant SNPs were examined using NCBI database. Of the four significant loci, the two SNPs on chromosome 4 each mapped to seven candidate genes, while the SNP at 36,630,817 bp on chromosome 6 did not map to any gene, and the SNP at 39,545,717 bp on chromosome 6 mapped to two candidate genes.

The candidate genes identified by GWAS were functionally grouped into several biologically relevant categories, including cell wall modification, carbohydrate metabolism, hormone signaling, and protein regulation. Genes involved in cell wall remodeling, such as those encoding xylan endotransglucosylase/hydrolase and xylan O-acetyltransferase, suggest a potential role of structural dynamics in flowering transition. The identification of UTP-glucose-1-phosphate uridylyltransferase highlights a possible role of carbohydrate metabolism and energy allocation in regulating flowering time. In addition, the presence of auxin-related genes, including *SAUR32-like* and the auxin transporter *BIG-like*, indicates that hormone signaling pathways may contribute to the control of flowering. Furthermore, regulatory proteins such as E3 ubiquitin ligase and FKBP18 implied that post-translational modification and protein turnover could be involved in fine-tuning flowering-related processes.

3.4 Comparative transcriptomics reveals stage-specific and continuously regulated genes during flowering transition in *Saccharum spontaneum*

We performed comparative transcriptomic analyses across three developmental transitions in *S. spontaneum*: (1) early flowering leaves (EEL) vs. early mature flowering leaves (EML), (2) early inflorescences (EEF) vs. mature inflorescences (EMF), and (3) non-flowering leaves (ENL) vs. early flowering leaves (EEL). The results revealed distinct differentially expressed genes (DEGs) patterns among the comparisons, with the most extensive transcriptional reprogramming observed during inflorescence development (EEF vs. EMF), indicating intensive molecular activity associated with reproductive organ maturation.

Volcano plot analysis revealed clear differences in DEG distributions among developmental stages **(Fig. 4A–C)**. In the ENL vs. EEL comparison, 888 significantly upregulated and 1,000 significantly downregulated genes were identified. In the EEL vs. EML comparison, 1,624 genes were significantly upregulated and 577 genes were significantly downregulated. The EEF vs. EMF comparison exhibited the largest number of DEGs, including 5,247 upregulated and 4,833 downregulated genes.

Integrated GO and KEGG enrichment analyses revealed dynamic stage-specific transcriptional changes during flowering development **(Fig. 4E–F)**. In the ENL vs. EEL comparison, DEGs were mainly enriched in secondary metabolism and amino acid-related pathways, including flavonoid biosynthesis, phenylpropanoid biosynthesis, and cysteine and methionine metabolism. These pathways were accompanied by stress-responsive and metabolic GO categories, indicating coordinated metabolic adjustment during the transition from vegetative to reproductive growth.

In the EEL vs. EML comparison, enriched pathways were primarily associated with central carbon metabolism and energy production, including starch and sucrose metabolism, glycolysis, oxidative phosphorylation, and amino acid biosynthesis. GO categories related to hormone responses and signal transduction were also enriched, suggesting increased energetic and metabolic demands during flowering development. In the EEF vs. EMF comparison, DEGs were predominantly enriched in hormone signaling, MAPK signaling, and reproductive development-related pathways. GO terms associated with cell wall organization, pollen development, and cellular structural processes were significantly enriched, together with KEGG pathways related to plant hormone signal transduction and secondary metabolism. These results indicate active reproductive development accompanied by hormone-responsive and structural remodeling processes.

Flowering development involves coordinated transcriptional changes across multiple biological pathways. Venn analysis identified genes showing continuous expression changes across developmental stages, including continuously downregulated genes shared between ENL vs. EEL and EEL vs. EML comparisons, as well as continuously upregulated genes shared among flowering-related comparisons. In total, 14 continuously downregulated and 154 continuously upregulated genes were identified and subjected to KEGG enrichment analysis. Continuously upregulated genes were mainly enriched in pathways associated with secondary metabolism and signal transduction, including terpenoid backbone biosynthesis, isoflavonoid biosynthesis, α-linolenic acid metabolism, and plant hormone signal transduction. Representative genes included an isopentenyl transferase gene (Sspon.02G0026290-1A), an isoflavone synthase gene (*Sspon.07G0004270-1A*), a lipoxygenase gene (*Sspon.01G0049070-1T*), and an auxin response factor gene (*Sspon.01G0011920-1A*). In contrast, continuously downregulated genes were enriched in pathways related to cutin, suberin, and wax biosynthesis, tryptophan metabolism, glyoxylate metabolism, and peroxisome-associated processes.

However, several pathways and representative genes identified during flowering development, including hormone signaling, lipid metabolism, cell wall-related processes, and MAPK-associated pathways, were functionally consistent with candidate genes detected in the environmental association analyses. These results suggest that flowering development in wild *S. spontaneum* involves coordinated metabolic and stress-responsive processes that partially overlap with pathways associated with environmental adaptation.

4 Discussion

4.1 Genetic and phenotypic patterns of flowering-associated variation in a core *S. spontaneum* population

Phylogenetic trees based on sampling locations suggest that the Chinese population may currently be undergoing expansion. Low genetic differentiation between the two groups of *S. spontaneum* and extensive gene flow indicate that population structure has limited influence on phenotypic variation in *S. spontaneum*. The presence of rare alleles and frequent hybridization likely maintain diversity in flowering

time. Together with the latitudinal distribution pattern, these results suggest that flowering time variation is mainly driven by environmental factors (e.g., photoperiod) and specific genetic variants rather than overall genomic background.

Flowering is a complex developmental process initiated by environmental cues such as photoperiod and temperature, which are perceived and transduced into endogenous signaling pathways [30, 31]. These signals subsequently trigger large-scale molecular reprogramming, as well as continuous transcriptional changes in genes associated with metabolism and development. At the phenotypic level, the core *S. spontaneum* population exhibits a clear and structured trait architecture. Growth-related traits, including stalk length, stalk diameter, internode number, number of leaves, single stem weight, and Brix, show strong positive correlations, forming a tightly coordinated growth network that collectively contributes to biomass accumulation and yield formation [32]. In contrast, flowering time displays weak correlations with these growth traits, indicating that it is largely independent of the coordinated vegetative growth system.

Despite this relative independence, a trade-off between vegetative growth and reproductive transition is evident [33]. Early-flowering accessions tend to undergo a shortened vegetative phase, limiting biomass accumulation and resulting in reduced growth-related trait performance. Conversely, late-flowering accessions maintain a prolonged vegetative growth period, allowing greater resource accumulation and potentially higher yield. Consistent with this phenotypic divergence, genome-wide association analysis identified four significant SNPs associated with flowering time, providing preliminary genetic evidence underlying flowering-related variation in the core *S. spontaneum* population.

4.2 Integrated GWAS and transcriptomic analyses reveal biological processes associated with flowering development in *S. spontaneum*

To further characterize flowering-associated biological signatures, GWAS candidate genes were integrated with transcriptomic analyses across flowering developmental stages. Flowering development in *S. spontaneum* was associated with coordinated changes across multiple biological processes, including metabolism, structural remodeling, hormone-related pathways, and protein regulatory processes. The combined analyses indicated that flowering-related GWAS loci and stage-specific transcriptomic patterns converge on partially shared biological processes during the transition from vegetative to reproductive growth. By integrating these two datasets, several candidate loci identified from flowering-associated regions showed correspondence with expression-enriched pathways, highlighting their potential as flowering-associated genetic markers. Candidate genomic regions and expression signatures emerging from this integration may be useful for future pre-breeding efforts and the development of molecular markers for flowering-time variation in wild sugarcane germplasm.

Genes associated with carbohydrate metabolism, such as *UTP–glucose-1-phosphate uridylyltransferase*, corresponded with transcriptomic enrichment in starch and sucrose metabolism as well as glycolysis

pathways, indicating active carbon reallocation and energy metabolism during flowering development [34–36]. Similarly, candidate genes related to cell wall organization, including xyloglucan endotransglucosylase/hydrolase and trichome birefringence-like proteins, were consistent with the enrichment of polysaccharide metabolism and cell wall remodeling pathways, suggesting that structural modification may accompany reproductive tissue differentiation and inflorescence development [30, 31]. Auxin-related genes identified by GWAS, including SAUR32-like and BIG-like, together with additional loci associated with hormone-related and protein regulatory pathways, represent potential genomic regions underlying flowering-related trait variation in *Saccharum spontaneum*. Rather than focusing on specific functional interpretations, these loci provide valuable molecular markers reflecting natural allelic variation associated with flowering-time differentiation across diverse genetic backgrounds.

From a broader perspective, these flowering-associated variants may serve as informative resources for population-level genetic diversity assessment, evolutionary studies of adaptive divergence, and marker-assisted selection in sugarcane pre-breeding programs. The integration of GWAS-derived loci with transcriptomic datasets further enhances their potential utility by linking genetic variation with developmental-stage expression patterns, thereby providing a foundational resource for the identification of breeding-relevant markers in wild germplasm.

Wild *S. spontaneum* populations are distributed across diverse ecological regions in China and are subject to substantial seasonal climatic heterogeneity. Previous studies have shown that extensive gene flow may weaken isolation-by-distance patterns, whereas climatic selection can contribute to adaptive differentiation across heterogeneous environments [37–39]. In addition, seasonal temperature and precipitation fluctuations are important environmental factors influencing sugarcane growth and flowering regulation [32, 40–42]. Consistent with these observations, pathways related to hormone signaling, stress responses, and cell wall-associated processes were enriched during flowering development in this study and were also detected in our preliminary, unpublished environmental association analyses. This overlap indicates that flowering-related variation and environmental responsiveness in *S. spontaneum* may converge on partially shared biological signatures across these two analytical frameworks, suggesting a degree of coordinated patterning between developmental and environmental dimensions of variation. These findings provide valuable candidate genes and biological insights for understanding flowering-associated developmental processes and for the future utilization of wild sugarcane germplasm in breeding programs.

Conclusions

Integrating phenotypic characterization, GWAS and transcriptomic profiling identified genetic loci and biological processes associated with flowering variation in wild *Saccharum spontaneum*. Analysis of 147 accessions identified four SNPs significantly associated with flowering time, while transcriptomic profiling of the early-flowering accession Hainan 92 − 3 revealed developmental changes in metabolism, hormone signaling and cell wall organization. Functional correspondence between GWAS candidate genes and transcriptome-enriched pathways suggests that these processes contribute to flowering-

associated variation. These findings provide candidates for functional investigation and future marker development in sugarcane breeding. Validation in independent populations and functional characterization are needed to establish the roles of these candidates and assess their breeding utility.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

All authors approve the manuscript and consent to the publication of the work.

Funding

This work was financially supported by the Science and Technology Major Project of Guangxi (Grant No. GK AD25069107), the Guangxi Natural Science Foundation (Grant Nos. 2025GXNSFAA069100 and GK AD20297064), the State Key Laboratory for Conservation and Utilization of Subtropical Agro-Bioresources (Grant No. SKLCUSA-a202402), and the "One Hundred Person" Project of Guangxi Province.

Competing interests

The authors declare no competing interests.

Authors' information

Authors and Affiliations

Guangxi Sugarcane Bio-breeding Laboratory, State Key Laboratory for Conservation and Utilization of Subtropical Agro-bioresources, College of Agriculture, Guangxi University, Nanning, Guangxi, 530004,China

Sicheng Li, Fengzhen Wu, Lijun Zhang, Zhen Wang & Xiping Yang

Sugarcane Research Institute, Guangxi Academy of Agricultural Sciences/Guangxi Key Laboratory of Sugarcane Genetic Improvement/Key Laboratory of Sugarcane Biotechnology and Genetic Improvement (Guangxi), Ministry of Agriculture & Rural Affairs, Nanning, Guangxi, 530007, China

Weixing Duan

Corresponding authors

Correspondence to Weixing Duan (duanweixing84@126.com) or Xiping Yang (xipingyang@gxu.edu.cn).

Data availability

The raw data used in this work were deposited at China National Center for Bioinformation (CNCB), BioProject: PRJCA028665.

Author contributions

Sicheng Li: Formal analysis, visualization, writing—original draft, writing—review and editing. Zhen Wang: Formal analysis, visualization, writing—original draft. Lijun Zhang: Resources. Fengzhen Wu: Visualization, writing—review and editing. Xiping Yang: Supervision, resources, funding acquisition, writing—review and editing.Weixing Duan:Supervision, resources.

References

1. Goldemberg J, Coelho S, Guardabassi P. The sustainability of ethanol production from sugarcane. ENERGY POLICY. 2008;36(6):2086–97.
2. Shabbir R, Javed T, Sun S-R, Wang Z-Q, Zhang W, Gao S-J, Wang Q-N. Functional genomics in sugarcane breeding: key challenges and strategies. Crit Rev Biotechnol 2026:1–21.
3. Thirugnanasambandam PP, Hoang NV, Henry RJ. The challenge of analyzing the sugarcane genome. Front Plant Sci. 2018;9:616.
4. Zhang J, Zhang X, Tang H, Zhang Q, Hua X, Ma X, Zhu F, Jones T, Zhu X, Bowers J. Allele-defined genome of the autopolyploid sugarcane Saccharum spontaneum L. Nat Genet. 2018;50(11):1565–73.
5. Zhang J, Qi Y, Hua X, Wang Y, Wang B, Qi Y, Huang Y, Yu Z, Gao R, Zhang Y. The highly allo-autopolyploid modern sugarcane genome and very recent allopolyploidization in Saccharum. Nat Genet. 2025;57(1):242–53.
6. Healey A, Garsmeur O, Lovell J, Shengquiang S, Sreedasyam A, Jenkins J, Plott C, Piperidis N, Pompidor N, Llaca V. The complex polyploid genome architecture of sugarcane. Nature. 2024;628(8009):804–10.
7. Wang H, Pan Y, Wu M, Liu J, Yang S, Wu Q, Que Y. Sugarcane genetics: Underlying theory and practical application. CROP J. 2025;13(2):328–38.
8. Mahadevaiah C, Appunu C, Aitken K, Suresha G, Vignesh P, Swamy H, Valarmathi R, Hemaprabha G, Alagarasan G, Ram B. Genomic Selection in Sugarcane: Current Status and Future Prospects. Front Plant Sci 2021, 12.
9. Babu K, Janakiraman V, Palaniswamy H, Kasirajan L, Gomathi R, Ramkumar T. A short review on sugarcane: its domestication, molecular manipulations and future perspectives. Genet Resour Crop Evol. 2022;69(8):2623–43.
10. Durai AA, Govindaraj P, Pazhany AS. Flowering behaviour of sugarcane genotypes from different agro climatic zones of india. Sugar Tech. 2014;16:157–63.
11. Manechini JRV, Santos PHS, Romanel E, Brito MS, Scarpari MS, Jackson S, Pinto LR, Vicentini R. Transcriptomic analysis of changes in gene expression during flowering induction in sugarcane

under controlled photoperiodic conditions. Front Plant Sci. 2021;12:635784.

12. Quiroz S, Yustis J, Chávez-Hernández E, Martínez T, Sanchez M, Garay-Arroyo A, Alvarez-Buylla E, García-Ponce B. Beyond the Genetic Pathways, Flowering Regulation Complexity in Arabidopsis thaliana. Int J Mol Sci 2021, 22(11).
13. Robles P, Pelaz S. Flower and fruit development in Arabidopsis thaliana. Int J Dev Biol. 2005;49(5–6):633–43.
14. Benlloch R, Kim M, Sayou C, Thévenon E, Parcy F, Nilsson O. Integrating long-day flowering signals: a LEAFY binding site is essential for proper photoperiodic activation of APETALA1. Plant J. 2011;67(6):1094–102.
15. Gan E, Xu Y, Wong J, Goh J, Sun B, Wee W, Huang J, Ito T. Jumonji demethylases moderate precocious flowering at elevated temperature via regulation of FLC in Arabidopsis. Nat Commun 2014, 5.
16. Xu S, Chong K. Remembering winter through vernalisation. Nat PLANTS. 2018;4(12):997–1009.
17. Luo X, He Y. Experiencing winter for spring flowering: A molecular epigenetic perspective on vernalization. J Integr Plant Biol. 2020;62(1):104–17.
18. Searle I, He Y, Turck F, Vincent C, Fornara F, Kröber S, Amasino R, Coupland G. The transcription factor FLC confers a flowering response to vernalization by repressing meristem competence and systemic signaling in Arabidopsis. Genes Dev. 2006;20(7):898–912.
19. Immink R, Posé D, Ferrario S, Ott F, Kaufmann K, Valentim F, de Folter S, van der Wal F, van Dijk A, Schmid M, et al. Characterization of SOC1's Central Role in Flowering by the Identification of Its Upstream and Downstream Regulators. Plant Physiol. 2012;160(1):433–49.
20. Lin S, Wang J, Poon S, Su C, Wang S, Chiou T. Differential regulation of expression by vernalization FLOWERING LOCUS C in cabbage and Arabidopsis. Plant Physiol. 2005;137(3):1037–48.
21. Wang J. Regulation of flowering time by the miR156-mediated age pathway. J Exp Bot. 2014;65(17):4723–30.
22. Gawarecka K, Ahn J. Isoprenoid-Derived Metabolites and Sugars in the Regulation of Flowering Time: Does Day Length Matter? Front Plant Sci 2021, 12.
23. Hassan A, Malik K, Naqvi SAM, Khan K, Sadia H. A comprehensive review of Saccharum spontaneum, its traditional uses, phytochemistry and pharmacology. Ethnobotany Res Appl. 2024;29:1–13.
24. Li S, Wang Z, Xu C, Zhang L, Zhang B, Huang Y, Lu X, Duan W, Yang X. Chloroplast genome diversity and phylogenetic insights of Saccharum spontaneum across Southeast Asia. Plants People Planet. 2025;7(6):1834–44.
25. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, Maller J, Sklar P, de Bakker PIW, Daly MJ, et al. PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. Am J Hum Genet. 2007;81(3):559–75.
26. Wickham H. ggplot2. WIRE Comput Stat. 2011;3(2):180–5.

27. Alexander DH, Lange K. Enhancements to the ADMIXTURE algorithm for individual ancestry estimation. BMC Bioinformatics. 2011;12:1–6.
28. Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST, et al. The variant call format and VCFtools. Bioinformatics. 2011;27(15):2156–8.
29. Zhang Q, Qi Y, Pan H, Tang H, Wang G, Hua X, Wang Y, Lin L, Li Z, Li Y. Genomic insights into the recent chromosome reduction of autopolyploid sugarcane Saccharum spontaneum. Nat Genet. 2022;54(6):885–96.
30. Osnato M, Cota I, Nebhnani P, Cereijo U, Pelaz S. Photoperiod control of plant growth: flowering time genes beyond flowering. Front Plant Sci. 2022;12:805635.
31. Mouradov A, Cremer F, Coupland G. Control of flowering time: interacting pathways as a basis for diversity. Plant Cell. 2002;14(suppl1):S111–30.
32. Zhang B, Huang Y, Zhang L, Zhou Z, Zhou S, Duan W, Yang C, Gao Y, Li S, Chen M. Genome-wide association study unravels quantitative trait loci and genes associated with yield-related traits in sugarcane. J Agric Food Chem. 2023;71(44):16815–26.
33. Reekie EG, Avila-Sakar G. The shape of the trade-off function between reproduction and growth. Reproductive allocation in plants. Elsevier; 2005. pp. 189–214.
34. Wenqi Z. An overview of UDP-glucose pyrophosphorylase in plants. Trop Plant Biology. 2025;18(1):10.
35. MacNeill GJ, Mehrpouyan S, Minow MA, Patterson JA, Tetlow IJ, Emes MJ. Starch as a source, starch as a sink: the bifunctional role of starch in carbon allocation. J Exp Bot. 2017;68(16):4433–53.
36. Cho L-H, Pasriga R, Yoon J, Jeon J-S, An G. Roles of sugars in controlling flowering time. J Plant Biology. 2018;61(3):121–30.
37. Lasky JR, Josephs EB, Morris GP. Genotype–environment associations to reveal the molecular basis of environmental adaptation. Plant Cell. 2023;35(1):125–38.
38. Sexton JP, Hangartner SB, Hoffmann AA. Genetic isolation by environment or distance: which pattern of gene flow is most common? *Evolution* 2014, 68(1):1–15.
39. Orsini L, Vanoverbeke J, Swillen I, Mergeay J, De Meester L. Drivers of population genetic differentiation in the wild: isolation by dispersal limitation, isolation by adaptation and isolation by colonization. Mol Ecol. 2013;22(24):5983–99.
40. Wickramasinghe KP, Kong C-y, Lin X-q, Zhao P-f, Mehdi F, Li X-j, Liu X-l, Mao J, Lu X. Photoperiodic and lighting treatments for flowering control and its genetic regulation in sugarcane breeding. Heliyon 2024, 10(7).
41. Abu-Ellail FF, McCord PH. Temperature and relative humidity effects on sugarcane flowering ability and pollen viability under natural and seminatural conditions. Sugar Tech. 2019;21(1):83–92.
42. Luhring TM, Holdo RM. Trade-offs between growth and maturation: the cost of reproduction for surviving environmental extremes. Oecologia. 2015;178(3):723–32.

Figures

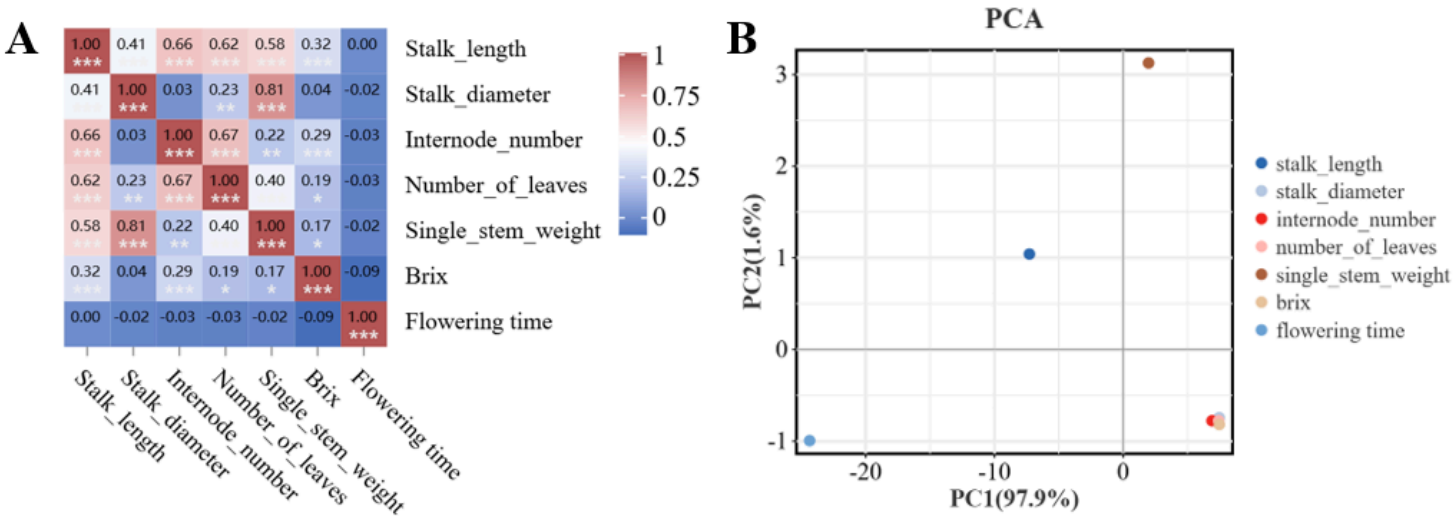


Figure 1

Correlation analysis and principal component analysis (PCA) of phenotypic traits in the core *S. spontaneum* population.

(A) Pearson correlation heatmap illustrating relationships among seven traits, including stalk length, stalk diameter, internode number, number of leaves, single stem weight, Brix, and flowering time. Color intensity represents the strength and direction of correlation, and asterisks denote significance levels.

(B) PCA plot based on the seven traits.

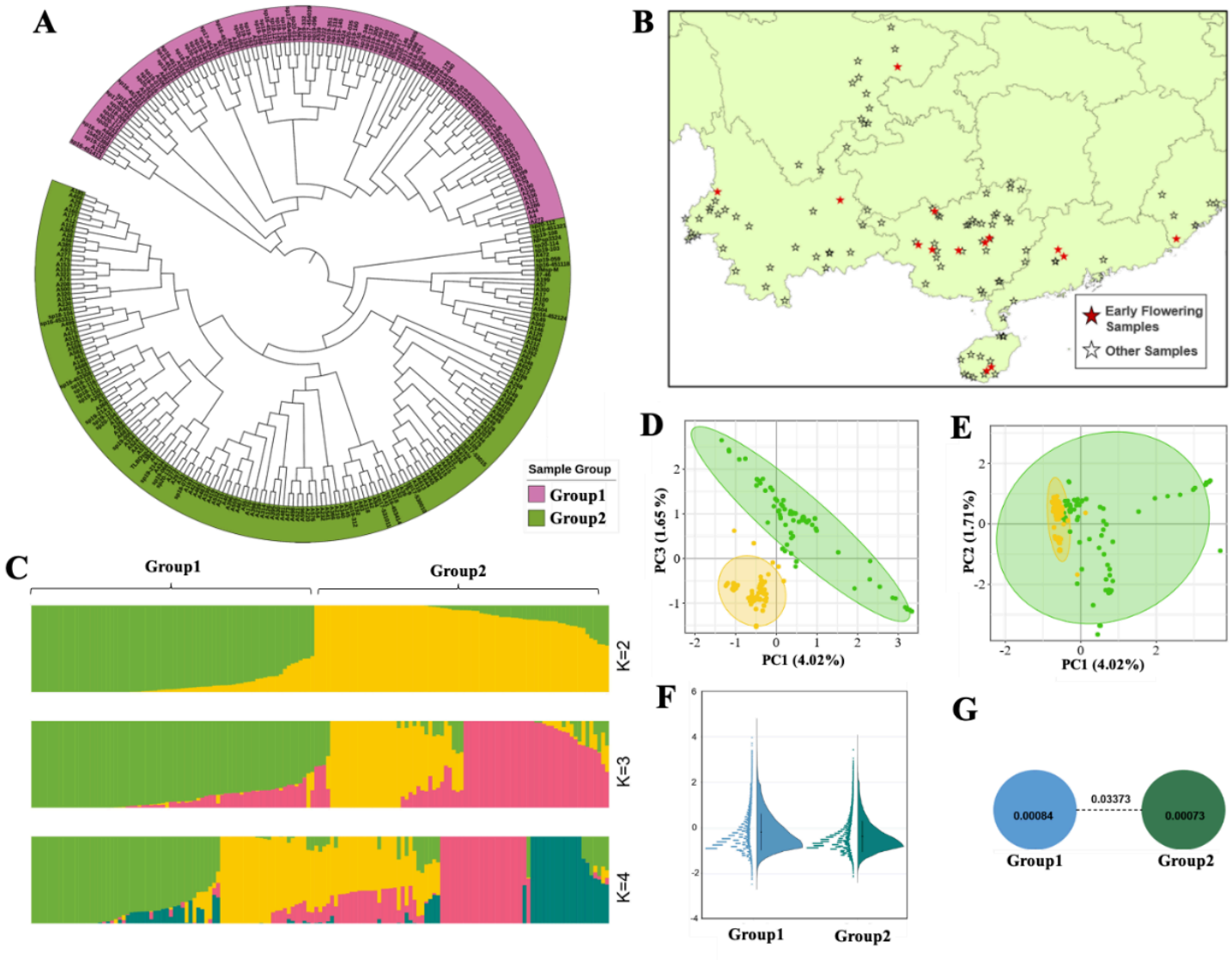


Figure 2

Population structure and PCA of *S. spontaneum* population.

(A) Phylogenetic tree of 147 *S. spontaneum* accessions;

(B) Sampling regions and the distribution of early-flowering accession;

(C) Population structure of subpopulation in K = 2-4;

(D-E) PCA of subpopulation in K = 2;

(F) F_{st} and π of the two subpopulations of *S. spontaneum*; (G) Tajimas'D of the two subgroups of *S. spontaneum*.

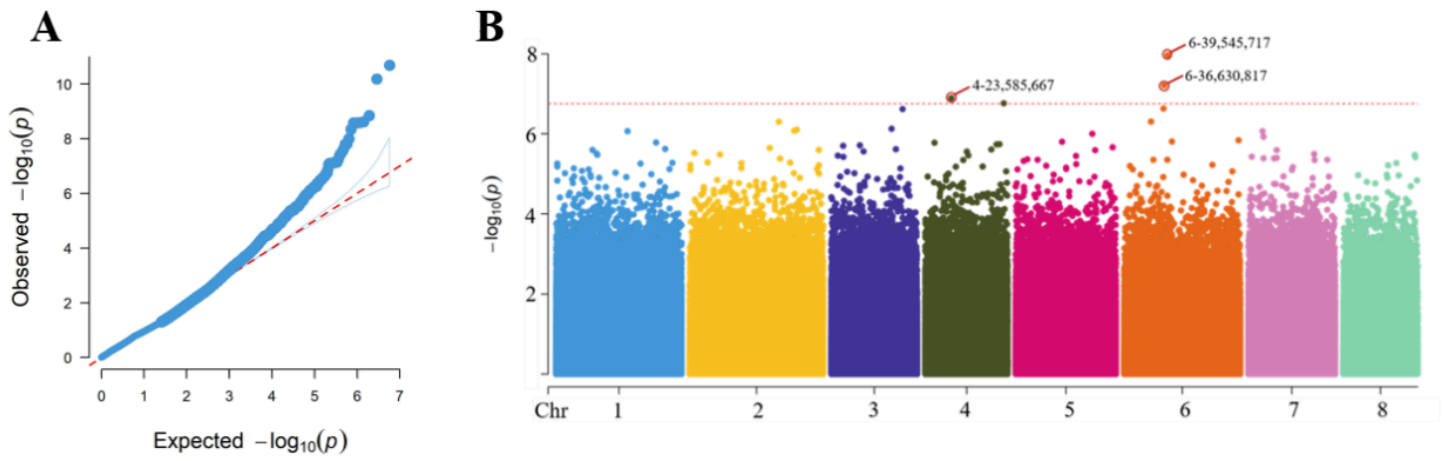


Figure 3

Genome-wide association study of flowering traits. (A) Quantile-quantile plot for the MLM; (B) Genome-wide association scan for flowering time in *S. spontaneum*.

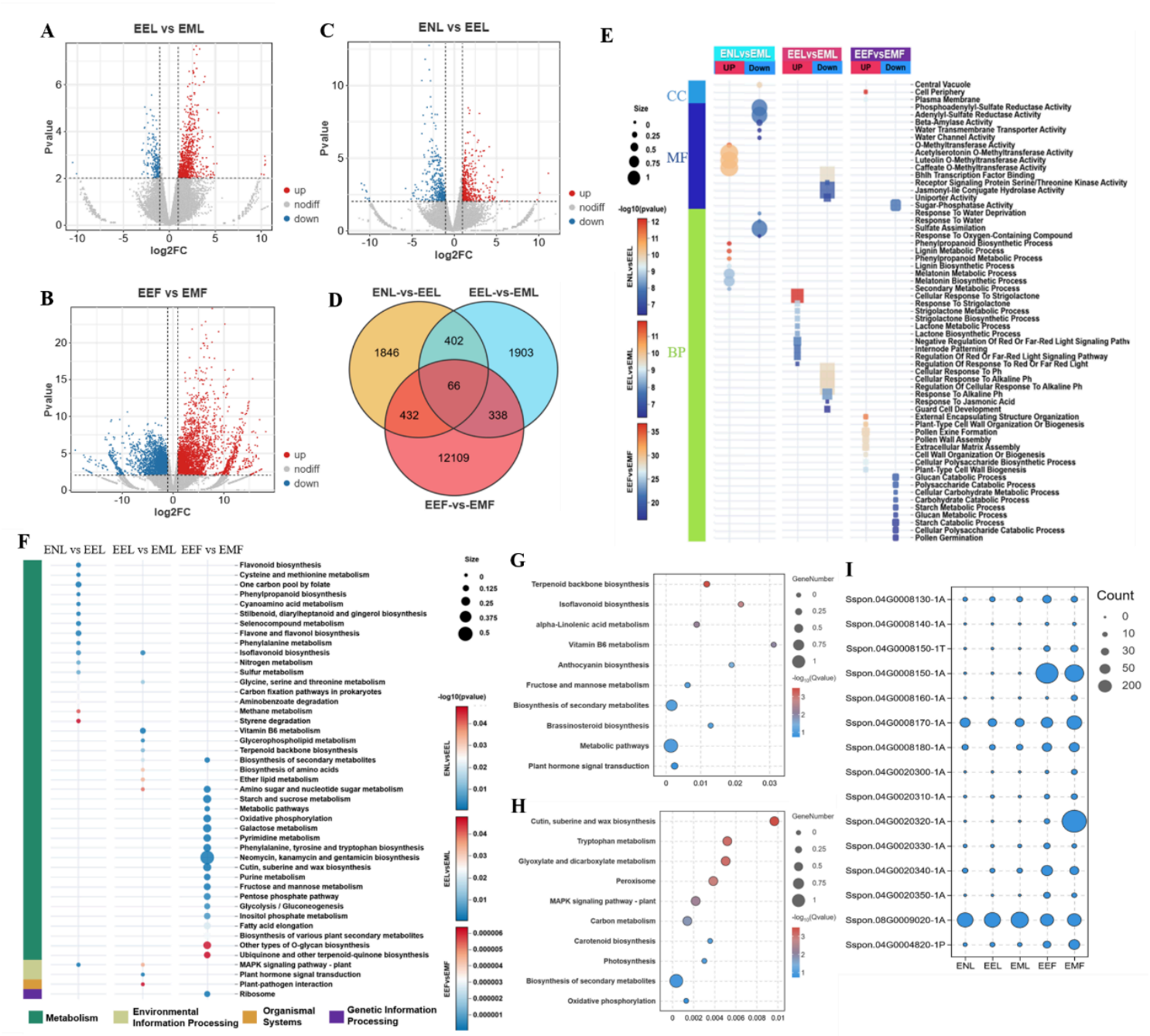


Figure 4

Identification and functional characterization of continuously regulated genes during flowering transition in *S. spontaneum*.

(A–C) Volcano plots showing differentially expressed genes (DEGs) in pairwise comparisons: (A) EEL vs. EML, (B) EEF vs. EMF, and (C) ENL vs. EEL. Red dots indicate upregulated genes, blue dots indicate downregulated genes, and gray dots represent non-significant genes. (D) Venn diagram showing the overlap of DEGs among the ENL vs. EEL, EEL vs. EML, and EEF vs. EMF comparisons. The intersecting regions represent continuously regulated genes across developmental stages.

(E) Gene Ontology (GO) enrichment analysis of DEGs. Dot size represents gene count, and color indicates statistical significance.

(F) KEGG pathway enrichment analysis of DEGs across different comparisons.

(G–H) KEGG enrichment analysis of continuously upregulated (G) and continuously downregulated (H) genes identified from the intersection analysis. The size of dots represents gene number, and color indicates statistical significance. (I) Expression patterns of representative continuously regulated genes across developmental stages (ENL, EEL, EML, EEF, and EMF). Bubble size indicates expression abundance, and color represents relative expression levels.