OPEN ACCESS

COMMENTARY

Sugarcane genomics: Origin, evolution, and domestication

Zihao Zhang,[1,2] Qibin Wu,[1] Dongjiao Wang,[1] Yuanyuan Zhang,[1] Fei Chen,[2,*] and Youxiong Que[1,*]

[1]State Key Laboratory of Tropical Crop Breeding, Sanya Research Institute, Chinese Academy of Tropical Agricultural Sciences, Sanya 572024, China

[2]State Key Laboratory of Tropical Crop Breeding, College of Breeding & Multiplication, Hainan University, Sanya 572025, China

*Correspondence: feichen@hainanu.edu.cn (F.C.); queyouxiong@126.com (Y.Q.)

Received: October 24, 2025; Accepted: November 19, 2025; Published Online: November 20, 2025; https://doi.org/10.59717/j.xinn-life.2025.100175

Citation: Zhang Z., Wu Q., Wang D., et al. (2025). Sugarcane genomics: Origin, evolution, and domestication. *The Innovation Life* **3**:100175.

Sugarcane (*Saccharum* spp.), accounting for approximately 80% of global sugar production, serves as a cornerstone of the global sugar industry and the bioeconomy. However, its genetic gains in sugar yield have stagnated, a bottleneck rooted in the formidable genomic complexity of modern hybrid cultivars.[1] These cultivars are characterized by a highly heterozygous, redundant, and mosaic genome derived from multiple ancestral species, representing a complex polyploid structure.[1] The complexity manifests in pervasive gene copy number variation, interspecific chromosomal mosaicism, and extensive genomic duplication. Collectively, these features a vast genetic labyrinth that severely impedes gene mapping, haplotype resolution, and positional cloning, presenting a fundamental bottleneck to breeding progress.

THE COMPLEX ORIGIN AND EVOLUTIONARY HISTORY OF SUGARCANE

A recent landmark study by Garsmeur et al., published in *Cell*, embarked on a comprehensive overview of sugarcane evolutionary history.[2] By performing whole-genome sequencing on 390 diverse accessions, the team established a multi-dimensional framework to systematically reconstruct the complex pedigree of this crop. They reconstructed the complete domestication trajectory from wild progenitors to cultivated hybrids (Figure 1A). These sophisticated computational inferences were subsequently substantiated by empirical data obtained through chromosomal fluorescence in situ hybridization (FISH) and "in silico" chromosome painting, thus providing both physical and sequence-level validation. This work resolved long-standing taxonomic debates and provided a definitive evolutionary framework for the crop. A key methodological innovation was the integration of in silico chromosome painting with empirical FISH validation, significantly enhancing the reliability of ancestry inferences.

The study confirmed that *S. officinarum* was directly domesticated from *S. robustum* and revealed its genome as a mosaic involving distinct *S. robustum* subgroups. Clarifying this mosaic origin is pivotal for identifying valuable allelic diversity that has been lost during domestication, offering a roadmap for the reintroduction of these traits into modern breeding programs.[2] Furthermore, it identified two independent centers of diversification: one in continental Asia, where *S. officinarum* hybridized with local *S. spontaneum* forms, leading to the emergence of *S. barberi* and *S. sinense*; and the other in the Melanesian and Polynesian islands, where introgression occurred between a previously unknown wild *Saccharum* ancestor and *Miscanthus*, the latter contributing to the formation of *S. maximum*.[2] Notably, this

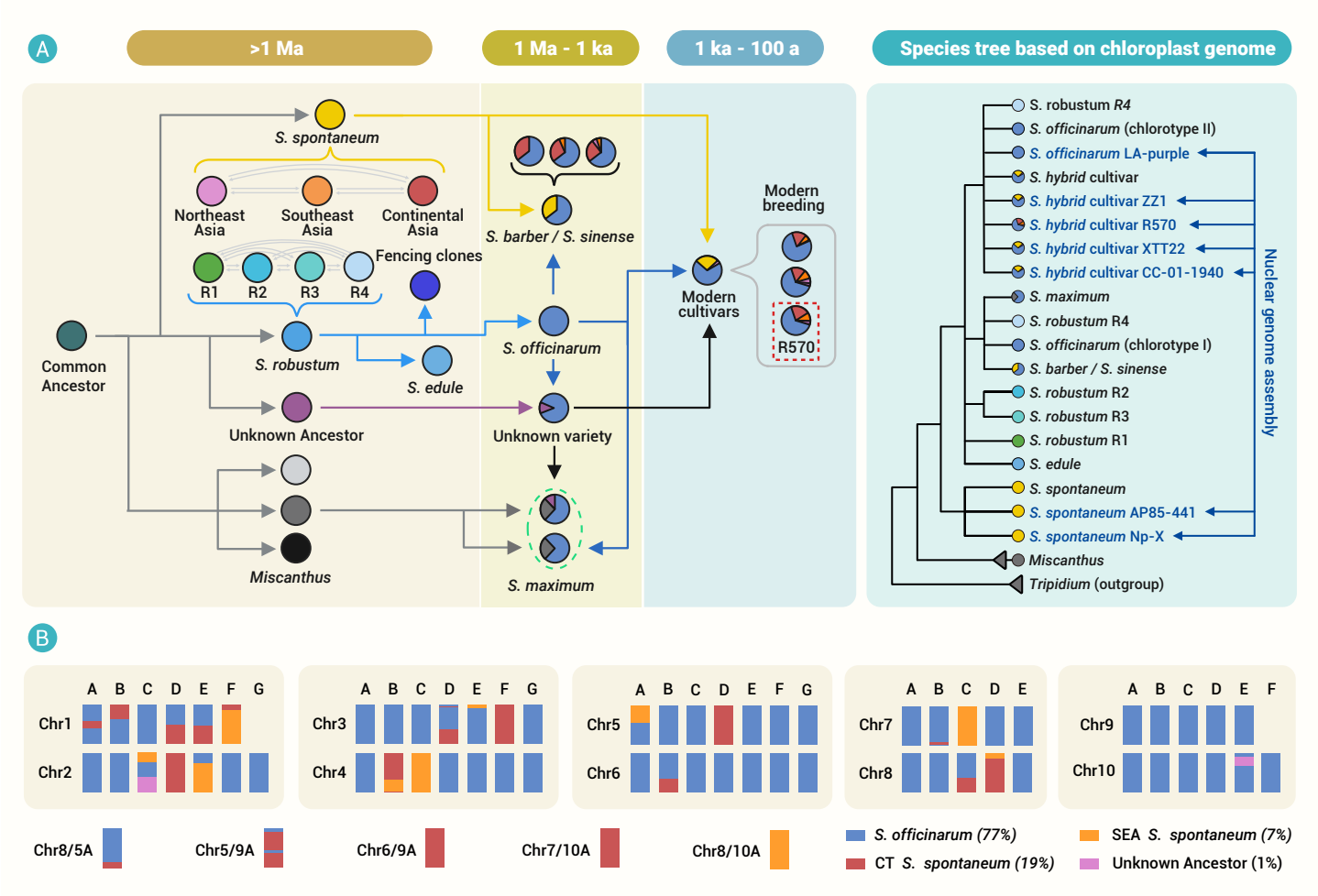

Figure 1. Genomic insights into the origin and domestication of sugarcane (A) Origins and domestication pathway of sugarcane. (B) Overview of ancestral composition in the R570 genome.

unknown ancestor was found to have contributed approximately 1% of the genome to many modern cultivars, likely originating from the East Melanesian biodiversity hotspot, representing a novel and potentially endangered genetic resource that awaits conservation and exploration (Figure 1A). The identification of this lineage underscores the urgent need to conserve wild *Saccharum* germplasm, as uncollected relatives may harbor untapped genetic diversity that is essential for the resilience of future crops.

Building on the aforementioned research, they provided the first precise quantification of ancestral contributions in modern cultivar R570: approximately 74% from *S. officinarum*, about 25% from *S. spontaneum* (further divisible into Southeast Asian and Continental Asian subgroups), and roughly 1% from the unknown ancestor (Figure 1B). This refined ancestral map provides the foundational clarity needed to begin partitioning complex traits to their specific ancestral sources. As expected, the R570 genotype has served as a key reference in sugarcane research, making it an ideal model for this precise dissection.

A NEW PARADIGM FOR SUGARCANE BREEDING IN THE GENOMICS ERA

The primary methodological innovation of this research lies in the distinctive application of repetitive k-mers—short DNA subsequences of length k that are highly abundant across the genome and often correspond to transposable elements or other repetitive sequences. The researchers treated these sequences as "genomic fossils", leveraging their high stability and resistance to changes in sequencing depth to accurately detect and quantify ancient ancestral contributions that are no longer present in a homozygous state within current germplasm collections.[2] They used these genome sequences as reliable and informative markers of ancestral heritage, benefiting from their robustness against variations in sequencing depth. By combining data from chloroplast phylogenomics and nuclear SNP patterns, they established a twofold line of evidence. Notably, the chloroplast genome effectively revealed maternal lineages, while nuclear SNPs clearly unraveled the complex mixed structure of the nuclear genome.[2] This methodology addressed the typical challenges associated with standard SNP analyses in highly polyploid genomes, enabling the researchers to trace ancient hybridization events and uncover previously unidentified ancestors. This k-mer-based approach well addresses the limitations inherent in standard SNP-based analyses, which frequently struggle with highly polyploid genomes due to ambiguous read mapping and complex allele dosage effects. Although this method is powerful, it is important to acknowledge that k-mer accuracy may be challenged in highly repetitive genomic regions, highlighting a consideration for future methodological refinements.

This groundbreaking historical reconstruction is not an isolated endeavor. It converges with the first polyploid reference genome assembled by Healey et al. (2024), together providing a well-rounded framework for deciphering the sugarcane genome.[1,2] The key methodological breakthrough centered on a partial-inbred genome assembly strategy. By employing this approach, this reference genome successfully resolved all unique chromosome haplotypes, thereby establishing an indispensable foundation for precisely tracing the crop's century-long breeding history.[1]

Leveraging this high-quality genome, the authors seamlessly bridged the physical map gap for the broad-spectrum brown rust resistance gene *Bru1*.[1] Through precise annotation, the candidate gene was narrowed down to a tandem kinase-pseudokinase pair originating from *S. spontaneum*, molecularly confirming the systematic contribution of wild ancestral genomes to critical resistance traits.[1] It demonstrates the practical utility of this genomic resource in identifying agronomically critical loci.

It is noteworthy that genomic analyses of the wild ancestor *S. spontaneum* revealed its unique genetic architecture. The chromosomes underwent a basic number reduction from "x=10" to "x=8", accompanied by extensive fragmentation and rearrangement. Particularly important is that these rearranged regions have unexpectedly become hotspots for disease resistance genes, harboring approximately 80% of NBS-type resistance genes and maintaining high allelic diversity through balancing selection in natural populations.[3] These discoveries provide a solid theoretical foundation for systematically and directionally mining the valuable diversity within the approximately 25% of the genome derived from *S. spontaneum* and the approximately 1% from the unknown ancestor. Collectively, these studies establish a powerful genomic resource platform for sugarcane biology.

CHALLENGES AND FUTURE RESEARCH DIRECTIONS

Researchers can use the allele-defined genomes as foundational guides to systematically explore novel genes associated with disease resistance and stress tolerance. To overcome the bottlenecks of traditional breeding, these high-precision haplotype maps will revolutionize quantitative trait locus (QTL) mapping. These maps provide a clear chromosomal framework that facilitates the precise assignment of causal variants, which overcomes the historical challenges of low resolution and ambiguous linkage in polyploid backgrounds.

At the level of precise genomic design, strategies are becoming more diversified. On one hand, priority may be given to targeting key single-copy or low-copy genes (such as the TKP gene for brown rust resistance) located on specific ancestral haplotypes for editing, drawing inspiration from successful examples like the wheat *Yr15* resistance gene.[4] On the other hand, for complex multi-copy gene families that are challenging to manipulate, alternative approaches such as gene silencing or epigenetic editing could be explored to achieve coordinated regulation of gene function.[5] These strategies collectively form a complete technical pathway from allele mining to variety design. However, due to the polyploid nature of sugarcane, a significant challenge often occurs in editing-based strategies, which necessitates the simultaneous modification of multiple homologous gene copies.

Nevertheless, significant challenges remain: These extend from the technical, such as the difficulty of multi-copy gene editing, to the fundamental, for instance, how can functional differences be accurately identified among highly similar haplotypes? And how can superior alleles from different ancestors be optimally integrated into a single genetic background? Addressing these questions requires sustained global scientific collaboration.

Despite these challenges, a new phase is marked by the first-ever possession of a complete evolutionary map and a high-resolution structural blueprint. Sugarcane breeding is progressively transforming from an empirically-driven practice into a predictive science. The convergence of these foundational studies not only provides critical data and tools but also establishes a standardized analytical workflow. This integrated approach, which combines dynamic evolutionary history with high-resolution structural maps, holds promise for unraveling the genomic complexities of other polyploid crops, ultimately laying the foundation for precise design-based breeding. This progress is poised to inject robust and enduring momentum into the sustainable development of the global sugar and bioenergy industries.

REFERENCES

1. Healey A. L., Garsmeur O., Lovell J. T., et al. (2024). The complex polyploid genome architecture of sugarcane. *Nature* **628**:804–810. DOI:10.1038/s41586-024-07231-4
2. Garsmeur O., Rio S., Pompidor N., et al. (2025). The genomic footprints of wild *Saccharum* species trace domestication, diversification, and modern breeding of sugarcane. *Cell* **188**:1–15. DOI:10.1016/j.cell.2025.09.017
3. Zhang J., Zhang X., Tang H., et al. (2018). Allele-defined genome of the autopolyploid sugarcane *Saccharum spontaneum* L. *Nat. Genet.* **50**:1565–1573. DOI:10.1038/s41588-018-0237-2
4. Klymiuk V., Yaniv E., Huang L., et al. (2018). Cloning of the wheat *Yr15* resistance gene sheds light on the plant tandem kinase-pseudokinase family. *Nat. Commun.* **9**:3735. DOI:10.1038/s41467-018-06138-9
5. Zhu H., Li C., Gao C. (2020). Applications of CRISPR-Cas in agriculture and plant biotechnology. *Nat. Rev. Mol. Cell Biol.* **21**:661–677. DOI:10.1038/s41580-020-00288-9

FUNDING AND ACKNOWLEDGMENTS

This work was funded by Chinese Academy of Tropical Agricultural Sciences for Science and Technology Innovation Team of National Tropical Agricultural Science Center (CATASCXTD202402), Project of State Key Laboratory of Tropical Crop Breeding (SKLTCBYWF202503, NKLTCBCXTD24 and NKLTCBCXTD38), and China Agriculture Research System of MOF and MARA (CARS-17).The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

AUTHOR CONTRIBUTIONS

Zihao Zhang: Conceptualization, Writing—Original Draft; Qibin Wu: Supervision, Writing—Review & Editing; Dongjiao Wang: Writing—Review & Editing; Yuanyuan Zhang: Writing—Review & Editing; Fei Chen: Conceptualization, Supervision, Writing—Review & Editing; Youxiong Que: Conceptualization, Supervision, Writing—Review & Editing, Funding acquisition. All authors contributed to the manuscript and approved the final version.

DECLARATION OF INTERESTS

The authors declare no competing interests.