

Chromosome-scale reference genome of Finger corn, an eight-rowed maize landrace retaining primitive traits

Jieying Yao

The Shenzhen Campus of Sun Yat-sen University, Sun Yat-sen University

Yi Jiang

The Shenzhen Campus of Sun Yat-sen University, Sun Yat-sen University

Kaiwen Tan

The Shenzhen Campus of Sun Yat-sen University, Sun Yat-sen University

Kesen Zhu

The Shenzhen Campus of Sun Yat-sen University, Sun Yat-sen University

Qingyun Li

The Shenzhen Campus of Sun Yat-sen University, Sun Yat-sen University

Junpeng Shi

The Shenzhen Campus of Sun Yat-sen University, Sun Yat-sen University

Zechao Yin

yinzch5@mail.sysu.edu.cn

The Shenzhen Campus of Sun Yat-sen University, Sun Yat-sen University

data-descriptor

Keywords:

Posted Date: October 20th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-11424299/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

Finger corn (*Zea mays* ssp. *mays*) is a maize landrace from Yunnan, China, characterized by 8 kernel rows and several teosinte-like morphological traits. Here, we present an integrated genomic, epigenomic, and three-dimensional chromatin resource for Finger corn using PacBio Fiber-seq and Hi-C data. The final assembly spanned 2.21 Gb, exhibiting high sequence continuity (contig N50 = 50.07 Mb), completeness (BUSCO = 99.1%, LAI = 26.75), and base-level accuracy (QV = 72.75). Genome annotation identified 39,413 protein-coding genes, with repetitive elements accounting for 87.35% of the genome. Fiber-seq data further enabled genome-wide profiling of chromatin accessibility and DNA methylation, identifying 238,443 accessible chromatin regions (ACRs). Hi-C data from female ears provided chromatin contact maps and annotations of A/B compartments and 3,053 topologically associating domains (TADs). Together, these genomic, epigenomic, and chromatin conformation datasets provide a resource for comparative genomics and studies of maize domestication, genome regulation, and agronomic trait variation.

Background & Summary

Maize is one of the most important cereal crops worldwide. As global food demand continues to rise, unlocking the full yield potential of maize has been a long-standing core goal for breeders [1, 2]. Kernel row number (KRN) is a key determinant of grain yield, and also represents one of the most representative selection targets during crop domestication and genetic improvement. Maize was domesticated from teosinte, the ears of wildtype only 2–4 kernel rows, whereas modern cultivated maize after millennia of selection often has more than 16 kernel rows, representing a tremendous improvement in yield potential [3–8]. However, most current studies on the genetic regulation of KRN have focused on modern cultivated maize, while research on wild relatives and local landraces remains relatively scarce, limiting our comprehensive understanding of the evolutionary trajectory and genetic basis of this key domestication trait.

Finger corn (*Zea mays* ssp. *mays*) is a unique local landrace endemic to Yunnan Province in southwestern China, derived from a cross between SiluNuo and a local small-kernel waxy maize. Its most prominent agronomic feature is a stable kernel row number of 8 (Fig. 1). Furthermore, phenotypic characterization in this study revealed that Finger corn retains a large set of traits typical of teosintes, such as tillering, multiple ears and abundant tassel branches. As a local landrace that has not undergone intensive modern breeding selection, the genome of Finger corn likely preserves some genetic variants from the early stages of maize domestication. These genetic resources hold unique and irreplaceable value for tracing maize evolutionary history and mining favorable alleles that have not been utilized in modern breeding.

Nevertheless, the lack of a high-quality reference genome for Finger corn has restricted its application in gene mining for key evolutionary traits and molecular breeding. To fill this gap, we performed de novo chromosome-level genome assembly of Finger corn, using PacBio Fiber-seq single-molecule long-read

sequencing combined with Hi-C chromosome conformation capture technology. The reference genome has a total length of 2.21 Gb, with a contig N50 exceeding 50 Mb, and exhibits high base accuracy and sequence completeness. In total, 39,413 protein-coding genes and repetitive elements accounting for 87.35% of the genome were annotated. The genomic resources generated will facilitate the use of Finger corn as a genetic model for dissecting the evolution of KRN and other traits in maize. Meanwhile, this work also has important scientific significance and application value for maize genetic diversity research, and promoting the conservation, development and utilization of local characteristic germplasm resources.

Methods

Materials and sample collection

The genetic material used in this study was Finger corn, a genetically stable local landrace. B73 as a widely used modern cultivated inbred line, was included as a phenotypic reference. All materials were grown under consistent field conditions and uniform agronomic management at the experimental station of Sun Yat-sen University, Guangming district, Shenzhen (113.91°E, 22.78°N). Major agronomic traits were investigated and statistically analyzed at maturity. The results showed that B73 exhibited no tillering, 3–6 tassel branches, 1–2 ears per plant, and 14–18 kernel rows per ear (Fig. 1). In contrast, Finger corn displayed typical tillering characteristics, with 8–18 tassel branches, 3–5 ears per plant, and a stable KRN of 8 (Fig. 1). Compared with the modern cultivar B73, Finger corn had more tassel branches, ears per plant whereas fewer kernel rows. Notably, tillering is a typical plant architecture trait of wild ancestors, suggesting that Finger corn may retain a genetic background more closely related to teosinte for these evolutionary traits (Fig. 1) [7, 8].

Sequencing

PacBio Fiber-seq sequencing

Young and fully expanded leaves at the 4-leaf stage of Finger corn were collected, snap-frozen in liquid nitrogen, and stored at – 80°C for PacBio Fiber-seq sequencing. After extraction and quality control of genomic DNA, Fiber-seq library construction and sequencing were performed by Berry Genomics Co., Ltd. (Beijing, China). DNA treated with m⁶A-MTase was fragmented into long fragments using the Megaruptor system, ligated with single-molecule real-time (SMRT) bell adapters to construct SMRTbell libraries, and sequenced on the PacBio Revio platform [9, 10]. In total, 60.26 Gb (~ 27× coverage) of HiFi reads were generated, with a read N50 of 15.43 kb (Table 1) [11–14]. The HiFi data from Fiber-seq were used for subsequent de novo genome assembly, chromatin accessibility, and DNA 5-methylcytosine (5mC) methylation analyses.

Hi-C sequencing

Ear primordia of Finger corn were collected at the primordium differentiation stage, sampled and preserved in the same manner for subsequent Hi-C library construction. Following the standard plant Hi-C protocol in our laboratory, ear tissue was cross-linked with formaldehyde to fix chromatin conformation, and the reaction was terminated with glycine followed by thorough grinding in liquid nitrogen. Intact nuclei were isolated using a sucrose gradient nuclear isolation buffer [15]. Chromatin within nuclei was digested with the restriction enzyme *Mbol*, followed by biotin labeling, proximity ligation, and Hi-C library construction [16, 17]. Qualified Hi-C libraries were sent to Wuhan FraserGen Bioinformatics Co., Ltd. (Wuhan, China) for high-throughput sequencing on the BGI platform using the PE150 strategy. A total of 279.33 Gb (~ 126× coverage) of raw Hi-C data were generated, with 95.29% of bases achieving a quality score of Q30 or higher (Table 1) [12–14, 18].

RNA-seq sequencing

Three developmental stages of early ears (2 mm, 5 mm, and 10 mm) were selected for RNA-seq analysis, with three biological replicates per stage and 6–10 developmentally synchronized individual plants per replicate. Sample collection, total RNA extraction, and quality control for each stage were performed in our laboratory. mRNA enrichment, library construction, and sequencing were conducted by Novogene Co., Ltd. (Beijing, China). mRNA was enriched using Oligo(dT) magnetic beads, and fragmented mRNA was used as template for cDNA synthesis. After end repair, A-tailing, amplification, purification, and quality inspection, double-stranded cDNA libraries were sequenced on the Illumina NovaSeq 6000 platform with the PE150 strategy [19]. Raw reads were filtered using fastp with default parameters [20], yielding 63.58 Gb (~ 29× coverage) of clean data, with Q30 values ranging from 96.62% to 97.1% and GC content from 47.37% to 49.63%. RNA-seq data were used to profile gene expression during ear development and to characterize their distribution patterns in chromatin compartments and TADs.

Genome estimation and assembly

To construct the Finger corn genome, we integrated two types of data: PacBio HiFi reads and Hi-C reads (Table 1). De novo primary assembly of PacBio HiFi reads was performed using Hifiasm (version 0.25.0-r726) to generate initial contig sequences [21]. Heterozygous redundant contigs were identified and removed using Purge_dups (version 1.0; `purge_dups -T ${cutoffs} -l 1 -m 10 -u 60`) based on sequencing coverage and sequence homology [22]. Subsequently, RagTag (version 2.1.0) was applied for sequence correction and reference-guided scaffolding to generate chromosome-level assemblies from the optimized contigs [23, 24].

The preliminary chromosome-level assembly was further refined using Hi-C data. Hi-C reads were aligned to the scaffold-level genome using HiC-Pro (version 3.1.0) [25]. After filtering, valid interaction pairs were extracted to construct contact matrices at resolutions of 5 kb, 10 kb, 40 kb, and 100 kb, which were then normalized using the ICE method. Based on the normalized matrices, chromosome clustering,

ordering, and orientation were performed using HapHiC (version 1.0.7; -remove_allelic_links 2) [26]. Final manual inspection and curation were conducted using Juicebox (version 1.11.08) [27], and chromosome-level sequences were generated via juicer post-processing. Hi-C interaction heatmaps showed uniform inter-chromosomal contact frequencies with no obvious gaps (Fig. 2A) [28, 29]. To assess genome structural conservation, whole-genome alignment between the Finger corn assembly and the B73 reference genome was performed using minimap2 (version 2.30; -x asm10), and synteny visualization was generated using the D-GENIES online platform (<https://dgenies.toulouse.inra.fr/run>) [30, 31], revealing highly conserved chromosome structure between the two genomes (Fig. 2B).

The quality of the Finger corn genome assembly was evaluated using QUAST (version 5.3.0; -split-scaffolds -eukaryote) [32]. The total assembly length was 2.21 Gb, comparable to the genome size of the reference maize inbred line B73 (Table 2) [33]. The assembly contained 218 contigs with a contig N50 of 50.07 Mb. Genome completeness was assessed using BUSCO (version 6.0.0; busco -l embryophyta_odb10), yielding a completeness score of 99.1% (Table 2) [34]. Full-length LTR retrotransposons were identified using LTR_retriever (version 3.0.4), and the LTR Assembly Index (LAI) was calculated. An LAI value of 26.75 indicated excellent assembly continuity in repetitive regions (Table 2) [35, 36]. Single-base accuracy was evaluated using Merqury (version 1.3; meryl count k = 21) based on the k-mer distribution of HiFi reads [37], resulting in a QV value of 72.75 (Table 2). Additionally, based on the conserved plant telomeric tandem repeat motif (5'-TTTAGGG-3'), a total of 10 telomeric sequences were identified across the terminal regions of the 10 chromosomes., confirming the assembly completeness of chromosome terminal regions. Taken together, the Finger corn genome assembled in this study exhibits high quality in terms of continuity, completeness, and accuracy.

Genome annotation

Protein-coding gene annotation

De novo prediction and multi-database functional annotation of protein-coding genes were performed on the Finger-corn genome sequences. Gene structures were predicted using Helixer (version 0.3.6; Helixer.py -lineage land_plant) [38]. Meanwhile, TE-sorter (version 1.0.0; -dbrexdb -plant -stprot) was used to filter and classify transposon-encoded proteins from the Helixer-predicted protein sequences [39]. The filtered sequences were subjected to homology alignment and functional annotation against the NR, Swiss-Prot, InterProScan, and eggNOG-mapper databases, yielding 39,413 protein-coding genes (E-value < 1e-5). All encoded proteins contained functional domains supported by at least one database (Fig. 3). Further validation was performed using RNA-seq data from three developmental stages of ears. The results showed that 31,184 genes were stably expressed during ear development (raw count > 1, TPM > 1), corroborating the accuracy and reliability of the gene annotation at the transcriptomic level.

Transposable element annotation

Transposable elements (TEs) are fundamental components of the maize genome. High-quality TE annotation is essential for comprehensive dissection in maize [40]. For repetitive sequence annotation and TE analysis, a comprehensive maize-specific repeat sequence library was constructed in a hierarchical manner. The curated maize TE sequence library (maizeTE02052020) from the Maize TE Consortium (MTEC, <https://github.com/oushujun/MTEC>) was used as the core reference [41], supplemented with non-TE repetitive elements including 5S rRNA monomer (Genebank ID: DQ351339.1), its1_5.8S rRNA_its2 sequences (Genebank ID: AF019817.1), sub-telomere_U39642 (Genebank ID: U39642.1), subtelomere_4-12-1 (GeneBank ID: CL569186.1), as well as SINEs, 25S rRNA, 18S rRNA, Cent4, and tRNASAT_ZM from Repbase [42–44]. Meanwhile, Tandem Repeats Finder (TRF, v4.09.1) was used for genome-wide scanning to identify tandem repeat satellite sequences, which were also incorporated into the library to form a comprehensive maize repeat sequence database [45]. Finally, RepeatMasker (version 4.1.5) was used to perform systematic genome-wide annotation of repetitive sequences and classification of TE composition based on this comprehensive database [46, 47].

In the Finger corn genome, repetitive sequences accounted for 1,931.80 Mb (87.35% of the genome), of which LTR retrotransposons comprised 73.98% and DNA transposons accounted for 8.95%, respectively (Table 3). Repeat annotation of the B73 reference genome was performed using the same method, and the results were highly consistent with previously published studies [33]. Specifically, the proportions of satellite sequences and rDNA sequences in the Finger corn genome were higher than those in B73 (Table 3), which may be attributed to the better assembly continuity, especially for repetitive regions in the Finger corn genome.

Chromatin Accessibility and DNA Methylation

Fiber-seq performs PacBio sequencing of native chromatin DNA treated with m⁶A-MTase. Since m⁶A-MTase preferentially methylates adenines in accessible chromatin regions, analysis of Fiber-seq data can reveal accessible chromatin regions as well as their association with transposons and genes in the corresponding tissue [48, 49]. Firstly, genome-wide ACRs were identified using FIRE (version 0.1) [50]. A total of 238,443 ACRs were detected, ranging from 66 bp to 12,132 bp in length, with an average length of approximately 363 bp and a median length of 313 bp (Fig. 4A, B). The number of ACRs per chromosome ranged from 15,968 (Chr10) to 34,486 (Chr1), and ACR number was positively correlated with chromosome length. Meanwhile, single-base resolution DNA methylation sites were identified using pb-CpG-tools (version 3.0.0) [51]. The results showed that highly methylated regions largely overlapped with TE-dense regions, accompanied by reduced density of genes and ACRs; conversely, hypomethylated regions were enriched for genes and accessible chromatin (Fig. 4A).

To further characterize the distribution of ACRs, they were classified into three categories based on their positions relative to protein-coding genes: approximately 18.99% were located within genic regions (from transcription start to termination sites), 7.93% in the 2 kb flanking regions upstream and downstream of genes (Proximal), and 73.09% in distal intergenic regions, suggesting that most accessible chromatin in

the maize genome corresponds to distal cis-regulatory elements (Fig. 4C). To explore the association between open chromatin and transposons, we further analyzed the repeat composition within ACR regions. At the whole-genome level, TEs accounted for 87.35% of the genome, dominated by LTR retrotransposons. However, the TE proportion in ACR regions was significantly lower than the genome-wide average, especially for genic ACRs (gACRs), consistent with the evolutionary constraint on TE insertion in coding regions [52]. The TE proportion in distal ACRs (dACRs) was similar to the genome-wide level, with LTR retrotransposons accounting for 83.76% of the sequence, indicating that distal accessible chromatin is mostly distributed in TE-enriched regions (Fig. 4D) [52]. Motif enrichment analysis of transcription factor binding sites (TFBSs) was performed using AME (version 5.0.2; -verbose 1 -scoring avg -method fisher -hit-lo-fraction 0.25 -evalue-report-threshold 10.0 -control -shuffle -kmer 2) [53]. The results revealed significant enrichment of binding motifs for transcription factor families including ERF, C2H2, AP2, MADS, Dof, and bZIP in ACRs, indicating that these transcription factor families play important roles in transcriptional regulation in Finger corn (Fig. 4E).

Three-dimensional Chromatin Structure

To further elucidate the three-dimensional chromatin structure of Finger corn, we identified chromatin A/B compartments and topologically associating domains (TADs) using Hi-C data, and analyzed the compartment distribution of expressed genes during early ear development [54, 55]. Chromatin compartment analysis was performed using HOMER (version 5.1; runHiCpca.pl -res 100000 -superRes 100000), dividing the genome into transcriptionally active A compartments and transcriptionally repressed B compartments at 100 kb resolution [56]. Genome-wide partition analysis revealed that A compartments accounted for 50.79% of the whole genome, while B compartments accounted for 49.21%. TEs showed a slight enrichment in B compartments, consistent with the feature that TEs are mainly located in heterochromatic regions and subjected to epigenetic transcriptional silencing (Fig. 5A) [57]. In contrast, 70.52% of all genes were located in A compartments, a proportion far higher than the genome-wide A compartment ratio and the TE proportion in A compartments. The distribution pattern directly supports the functional definition of A compartments as transcriptionally permissive chromatin environments (Fig. 5A).

We then validated the functional differentiation between A and B compartments from both transcriptomic and epigenetic perspectives. Genes residing in A compartments exhibited significantly higher expression levels than those in B compartments across all three developmental stages (2 mm, 5 mm, and 10 mm), and all comparisons reached extremely significant levels (Fig. 5B). In parallel, genome-wide DNA methylation levels were consistently lower in A compartments than in B compartments (Fig. 5C). The combination of higher transcriptional activity and lower DNA methylation further confirms the open, active state of A compartments and the compact, silenced state of B compartments.

Subsequently, TAD identification was performed using HiCExplorer (version 3.6), including matrix format conversion, ICE correction, and TAD boundary calling. A total of 3,053 TAD domains were identified [58].

Each TAD was divided into 50 equal subsets along its relative position. Gene density peaked at the two edges of TADs and remained at a relatively uniform low level across the central regions (Fig. 5C). We defined the sequences spanning 10% of the length at each end of a TAD as boundary regions, and the intervening 80% as internal regions. Statistical comparison confirmed a significant enrichment of protein-coding genes at TAD boundaries relative to internal domains (Fig. 5D), supporting the notion that TAD boundaries represent transcriptionally permissive chromatin compartments preferentially populated with functional genetic elements.. DNA methylation exhibited the opposite spatial trend, with significantly lower methylation levels in boundary regions than in internal regions (Fig. 5E).

Data Records

All sequencing data and de novo genome assemblies generated in this study have been deposited for public access in two repositories: the National Center for Biotechnology Information (NCBI) and the National Genomics Data Center (NGDC). At NCBI, all datasets are organized under BioProject accession number PRJNA1519011. Specifically, raw sequencing reads are available in the Sequence Read Archive (SRA) under accession numbers SRR40363285 (Fiber-seq library) [11] and SRR40363286 (Hi-C library) [18]. The GenBank accession number for the corresponding de novo genome assembly is: JCDFTM000000000 [28]. At NGDC, all datasets are organized under BioProject accession number PRJCA074292. Raw sequencing reads have been deposited in the Genome Sequence Archive (GSA) under accession number CRA081484 [12]. The Genome Warehouse (GWH) accession number for the corresponding de novo genome assembly is GWHKDOA000000000.1 [29].

Technical Validation

Assembly quality was assessed using contiguity (Contig N50), repeat length, BUSCO completeness and base-level accuracy. Contig N50 lengths is 50.07 Mb. The complete BUSCO values exceeded 99.1%, and the LAI score was 26.75. The base-level accuracy of the assembly was represented by a quality value (QV) of 72.75

Declarations

Competing interests

The authors declare no conflicts of interest.

Author Contribution

Jieying Yao: Data curation, Formal analysis, Writing– original draft. Yi Jiang: Data curation, Formal analysis, Writing– review & editing. Kaiwen Tan: Data curation, Writing– review & editing. Kesen Zhu:

Data curation. Qingyun Li: Data curation. Junpeng Shi: Supervision, Funding acquisition, Project administration, Writing– review & editing. Zechao Yin: Conceptualization, Supervision, Visualization, Writing– original draft, Writing– review & editing.

Acknowledgement

We want to thank the colleagues from Berry Genomics Co., Ltd.(China, Beijing), Wuhan Frasergen Bioinformatics Co.,Ltd(China, Wuhan) and Novogene Co.,Ltd.(China, Beijing) for their assistance inpreparing the Fiber-seq, Hi-C and RNA-seq librariesas well as The High-performance Computing Public Platform (Shenzhen Campus) of SUN YAT-SEN UNIVERSITY for providing computing resources. This research was funded by grants from the National Natural Science Foundation of China (32372115), the Shenzhen Outstanding Youth Science Fund project (RCYX20231211090225038), the Shenzhen Science and Technology Program for Sustainable Development (KCXFZ20240903093007010) and the Guangdong Provincial College Students' Innovation Training Program Supported Projects (S105582026067) to J.S.

Data Availability

All sequencing data and de novo genome assemblies generated in this study have been deposited for public access in two repositories: the National Center for Biotechnology Information (NCBI) and the National Genomics Data Center (NGDC). At NCBI, all datasets are organized under BioProject accession number PRJNA1519011. Specifically, raw sequencing reads are available in the Sequence Read Archive (SRA) under accession numbers SRR40363285 (Fiber-seq library) and SRR40363286 (Hi-C library) . The GenBank accession number for the corresponding de novo genome assembly is: JCDFTM000000000 . At NGDC, all datasets are organized under BioProject accession number PRJCA074292. Raw sequencing reads have been deposited in the Genome Sequence Archive (GSA) under accession number CRA081484 . The Genome Warehouse (GWH) accession number for the corresponding de novo genome assembly is GWHKDOA000000000.1 .

Code availability

All programs and pipelines were executed following official manuals orhelp pages. Version and parameter information that we used for our analysis have been described in the Methods section.

References

1. Duvick, D. N. Genetic progress in yield of United States maize (*Zea mays L.*). *Maydica* 50, 193–202 (2005).

2. Rosegrant, M., Ringler, C., Sulser, T.B., Ewing, M., Palazzo, A., & Zhu, T. Agriculture and food security under global change: prospects for 2025/2050. *International Food Policy Research Institute, Washington, DC* 89 (2009).
3. Iltis, H. H. From teosinte to maize: the catastrophic sexual transmutation. *Science* 222, 886–894 (1983).
4. Matsuoka, Y., Vigouroux, Y., Goodman, M. M., Sanchez, G. J., Buckler, E., & Doebley, J. A single domestication for maize shown by multilocus microsatellite genotyping. *Proceedings of the National Academy of Sciences* 99(9), 6080–6084 (2002).
5. Piperno, D. R., Ranere, A. J., Holst, I., Iriarte, J., Dickau, R. Starch grain and phytolith evidence for early ninth millennium BP maize from the Central Balsas River Valley, Mexico. *Proceedings of the National Academy of Sciences* 106, 5019–5024 (2009).
6. Wills, D. M., Whipple, C. J., Takuno, S., Kursel, L. E., Shannon, L. M., & Ross-Ibarra, J. From many, one: genetic control of prolificacy during maize domestication. *PLoS genetics* 9, e1003604 (2013).
7. Doebley, J., Stec, A., Hubbard, L. The evolution of apical dominance in maize. *Nature* 386, 485–488 (1997).
8. Doebley, J. The genetics of maize evolution. *Annual Review of Genetics* 38, 37–59 (2004).
9. Stergachis, A. B., Debo, B. M., Haugen, E., Churchman, L. S., Stamatoyannopoulos, J. A. Single-molecule regulatory architectures captured by chromatin fiber sequencing. *Science* 368(6498), 1449–1454 (2020).
10. Li, L., Chen, G., Zhu, G., Wu, J., Kui, X., & Zhang, Z. Single-molecule views of chromatin accessibility and structure during photomorphogenesis. *Proceedings of the National Academy of Sciences* 122(48), e2516708122 (2025).
11. NCBI Sequence Read Archive <https://dataview.ncbi.nlm.nih.gov/object/SRR40363285> (2026).
12. NGDC Genome Sequence Archive <https://ngdc.cncb.ac.cn/gsub/submit/gsa/subCRA081484> (2026)
13. Zhang, S., Chen, X., Jin E., Wang A., Chen, T., & Zhang, X. The GSA Family in 2025: A Broadened Sharing Platform for Multi-Omics and Multimodal Data. *Genomics, Proteomics & Bioinformatics* 23(4):qzaf072 (2025).
14. CNGB-GDC Members and Partners, Database resources of the National Genomics Data Center, China National Center for Bioinformation in 2026. *Nucleic Acids Research* 54(D1), 28–47 (2026).
15. Belton, J. M., McCord, R. P., Gibcus, J. H., Naumova, N., Zhan, Y., & Dekker, J.. Hi-C: a comprehensive technique to capture the conformation of genomes. *Methods* 58(3), 268–276 (2012).
16. Rao, S. S., Huntley, M. H., Durand, N. C., Stamenova, E. K., Bochkov, I. D., & Robinson, J. T. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* 159(7), 1665–1680 (2014).
17. Padmarasu, S., Himmelbach, A., Mascher, M., Stein, N. In situ Hi-C for plants: an improved method to detect long-range chromatin interactions. *Methods in Molecular Biology* 1933, 441–472 (2019).
18. NCBI Sequence Read Archive <https://dataview.ncbi.nlm.nih.gov/object/SRR40363286> (2026)

19. Mortazavi, A., Williams, B. A., McCue, K., Schaeffer, L., Wold, B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nature methods*, 5(7), 621–628 (2008).
20. Chen, S., Zhou, Y., Chen, Y., Gu, J. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics*, 34(17), i884-i890 (2018).
21. Cheng, H., Concepcion, G. T., Feng, X., Zhang, H. Li, H. Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. *Nature Methods* 18, 170–175 (2021).
22. Guan, D., McCarthy, S. A., Wood, J., Howe, K., Wang, Y., & Durbin, R. Identifying and removing haplotypic duplication in primary genome assemblies. *Bioinformatics* 36(9), 2896–2898 (2020).
23. Alonge, M., Soyk, S., Ramakrishnan, S., Wang, X., Goodwin, S., & Sedlazeck, F. J. RaGOO: fast and accurate reference-guided scaffolding of draft genomes. *Genome biology* 20(1), 224 (2019).
24. Alonge, M. Lebeigle, L., Kirsche, M., Aganezov, S., Wang, X., & Lippman, Z. B. Automated assembly scaffolding elevates a new tomato system for high-throughput genome editing. *BioRxiv* 2021-11 (2021).
25. Servant, N., Varoquaux, N., Lajoie, B. R., Viara, E., Chen, C., & Vert, J. P. HiC-Pro: an optimized and flexible pipeline for Hi-C data processing. *Genome biology* 16(1), 259 (2015).
26. Zeng, X., Yi, Z., Zhang, X., Du, Y., Li, Y., & Zhou, Z. Chromosome-level scaffolding of haplotype-resolved assemblies using Hi-C data without reference genomes. *Nature Plants* 10(9), 1184–1200 (2024).
27. Robinson, J. T., Turner, D., Durand, N. C., Thorvaldsdóttir, H., Mesirov, J. P., & Aiden, E. L. Juicebox. js provides a cloud-based visualization system for Hi-C data. *Cell systems* 6(2), 256–258 (2018).
28. NCBI GenBank <https://identifiers.org/ncbi/insdc:JCDFTM0000000000> (2026)
29. NGDC Genome Warehouse <https://ngdc.cncb.ac.cn/gwh/submit/submission> (2026)
30. Li, H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* 34(18), 3094–3100 . (2018)
31. Cabanettes, F., Klopp, C. D-GENIES: dot plot large genomes in an interactive, efficient and simple way. *PeerJ* 6, e4958 (2018).
32. Gurevich, A., Saveliev, V., Vyahhi, N., Tesler, G. QUAST: quality assessment tool for genome assemblies. *Bioinformatics* 29(8), 1072–1075 (2013).
33. Hufford, M. B., Seetharam, A. S., Woodhouse, M. R., Chougule, K. M., Ou. S., & Liu, J. De novo assembly, annotation, and comparative analysis of 26 diverse maize genomes. *Science* 373(6555), 655–662 (2021).
34. Tegenfeldt, F., Kuznetsov, D., Manni, M., Berkeley, M., Zdobnov, E.M., & Kriventseva, E.V. OrthoDB and BUSCO update: annotation of orthologs with wider sampling of genomes. *Nucleic Acids Research* 53(D1), 516–522 (2025).
35. Manni, M., Berkeley, M. R., Seppey, M., Simão, F. A., Zdobnov, E. M. BUSCO update: novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of

- eukaryotic, prokaryotic, and viral genomes. *Molecular biology and evolution* 38(10), 4647–4654 (2021).
36. Ou, S., Jiang, N. LTR_retriever: a highly accurate and sensitive program for identification of long terminal repeat retrotransposons. *Plant physiology* 176(2), 1410–1422 (2018).
 37. Rhie, A., Walenz, B. P., Koren, S., Phillippy, A. M. Merqury: reference-free quality, completeness, and phasing assessment for genome assemblies. *Genome biology* 21(1), 245 (2020).
 38. Holst, F., Bolger, A. M., Kindel, F., Günther, C., Ma, J., & Triesch, S. Helixer: ab initio prediction of primary eukaryotic gene models combining deep learning and a hidden Markov model. *Nature Methods* 23, 732–739 (2026).
 39. Zhang, R., Li, G., Wang, X., Dainat, J., Wang, Z., & Ou, S. TESorter: an accurate and fast method to classify LTR-retrotransposons in plant genomes. *Horticulture Research* 9, uhac017 (2022).
 40. Stitzer, M. C., Anderson, S. N., Springer, N. M., Ross-Ibarra, J. The genomic ecosystem of transposable elements in maize. *PLoS genetics* 17(10), e1009768 (2021).
 41. Ou, S., Su, W., Liao, Y., Chougule, K., Agda, J., & Hellinga, A. Benchmarking transposable element annotation methods for creation of a streamlined, comprehensive pipeline. *Genome Biology* 20, 275 (2019).
 42. Tan, K., Liu, X., Wang, Z., Zhang, Z., Huang, W., & Liu, S. Near-complete genome assembly of a transformation-efficient elite inbred line LH244 and its comparison with B73. *Journal of Integrative Plant Biology* 68, 366–382 (2026).
 43. Chen, J., Wang, Z., Tan, K., Huang, W., Shi, J., & Li, T. A complete telomere-to-telomere assembly of the maize genome. *Nature Genetics* 55, 1221–1231 (2023).
 44. Jurka, J., Kapitonov, V. V., Pavlicek, A., Klonowski, P., Kohany, O., & Walichiewicz, J. Repbase update, a database of eukaryotic repetitive elements. *Cytogenetic and Genome Research*, 110(1–4), 462–467 (2005).
 45. Benson, G. Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Research* 27(2), 573–580 (1999).
 46. Flynn, J. M., Hubley, R., Goubert, C., Rosen, J., Clark, A. G., Feschotte, C., & Smit, A. F. RepeatModeler2 for automated genomic discovery of transposable element families. *Proceedings of the National Academy of Sciences* 117, 9451–9457 (2020).
 47. Tarailo-Graovac, M., Chen, N. Using Repeat Masker to identify repetitive elements in genomic sequences. *Current protocols in bioinformatics* 5, 4.10.1–14.10.14 (2004).
 48. Noshay, J. M., Anderson, S. N., Zhou, P., Ji, L., Ricci, M., & Lu, Z. Monitoring the interplay between transposable element families and DNA methylation in maize. *PLoS Genetics* 15(9), e1008291 (2019).
 49. Xu, G., Lyu, J., Li, Q., Liu, H., Wang, D., & Zhang, M. Evolutionary and functional genomics of DNA methylation in maize domestication and improvement. *Nature communications* 11(1), 5539 (2020).

50. Bohaczuk, S. C., Amador, Z. J., Li, C., Mallory, B. J., Swanson, E. G., & Ranchalis, J. Resolving the chromatin impact of mosaic variants with targeted Fiber-seq. *Genome Research* 34(12), 2269–2278 (2024).
51. Ni, P., Nie, F., Zhong, Z., Xu, J., Huang, N., & Zhang, J. DNA 5-methylcytosine detection and methylation phasing using PacBio circular consensus sequencing. *Nature communications* 14(1), 4054 (2023).
52. Bubb, K. L., Hamm, M. O., Tullius, T. W., Min, J. K., Ramirez-Corona, B., & Mueth, N. A. The regulatory potential of transposable elements in maize. *Nature Plants* 11(6), 1181–1192 (2025).
53. McLeay, R. C., Bailey, T. L. Motif Enrichment Analysis: a unified framework and an evaluation on ChIP data. *BMC bioinformatics* 11(1), 165 (2011).
54. Lieberman-Aiden, E., Van Berkum, N. L., Williams, L., Imakaev, M., Ragoczy, T., & Telling, A. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* 326(5950), 289–293 (2009).
55. Tian, L., Ku, L., Yuan, Z., Wang, C., Su, H., & Wang, S. Large-scale reconstruction of chromatin structures of maize temperate and tropical inbred lines. *Journal of Experimental Botany* 72(10), 3582–3596 (2021).
56. Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y. C., & Laslo, P. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Molecular cell* 38(4), 576–589 (2011).
57. Dong, P., Tu, X., Liang, Z., Kang, B. H., Zhong, S. Plant and animal chromatin three-dimensional organization: similar structures but different functions. *Journal of Experimental Botany* 71(17), 5119–5128 (2020).
58. Wolff, J., Bhardwaj, V., Nothjunge, S., Richard, G., Renschler, G., & Gilsbach, R. Galaxy HiCExplorer: a web server for reproducible Hi-C data analysis, quality control and visualization. *Nucleic acids research* 46(W1), W11-W16 (2018).

Tables

Tables 1 to 3 are available in the supplementary files section.

Figures

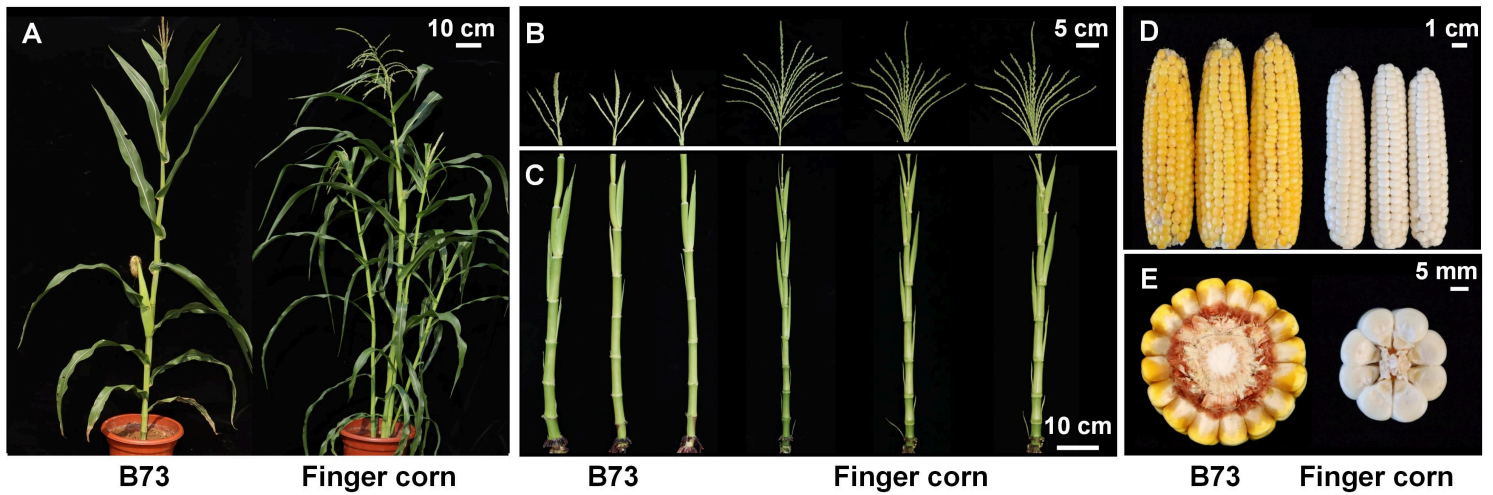


Figure 1

Comparison of multiple agronomic traits between Finger corn and B73. (A) The whole plant showed multiple tillerings of Finger corn compared with no tillering of B73. (B-C) The number of tassel branches (B) and ears (C) was significantly higher in Finger corn. (D) Finger corn produces markedly smaller ears with white kernels (E) Cross-sections of mature ears, illustrating the stable 8-kernel-row phenotype of Finger corn, in contrast to the higher kernel row count in B73.

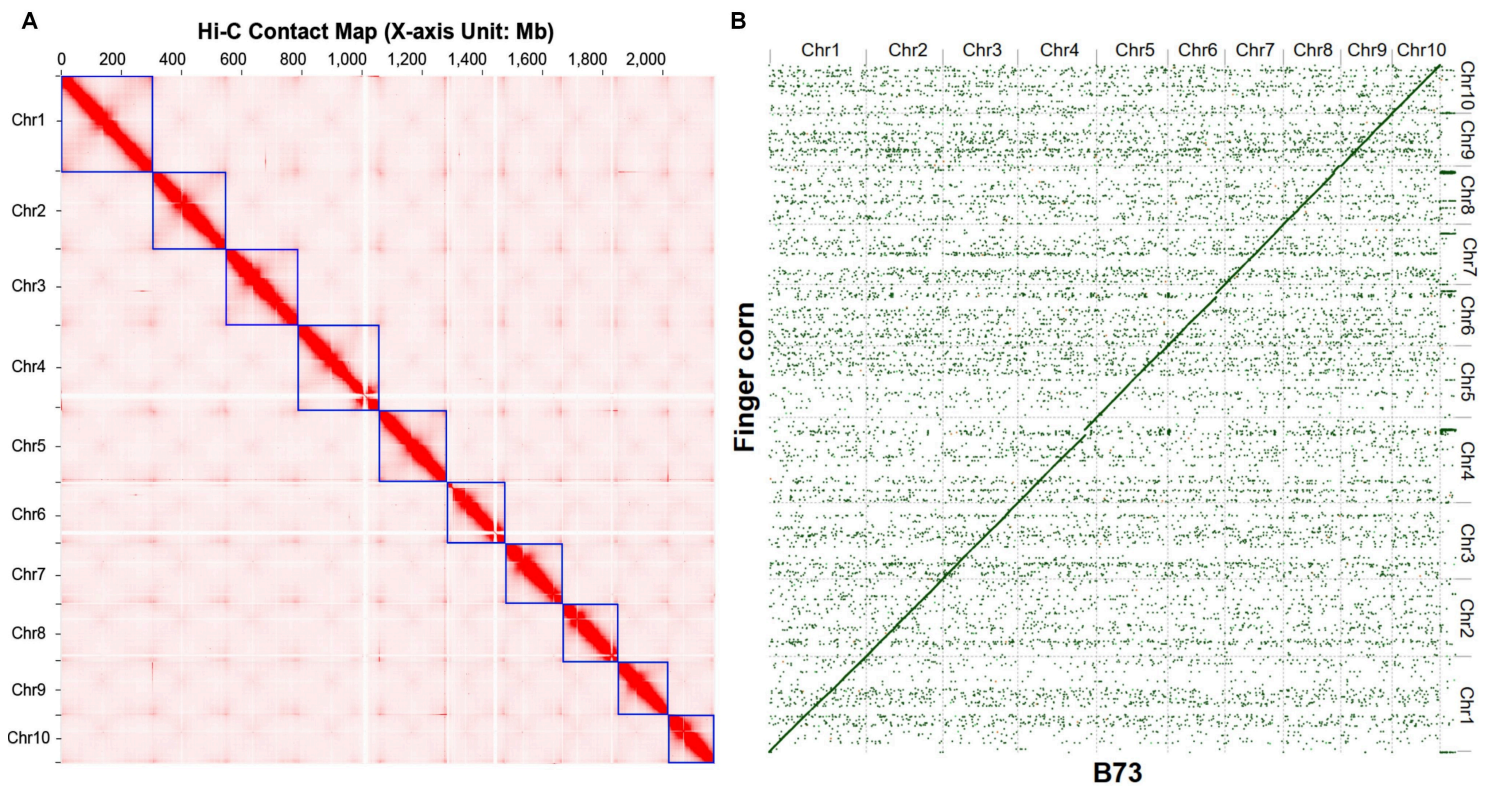


Figure 2

Genomic characteristics of the Finger corn genome. (A) Genome-wide Hi-C interactions among 10 pseudochromosomes of Finger corn. (B) The colinearity between 10 chromosomes of Finger corn and B73.

the reference genome of B73.

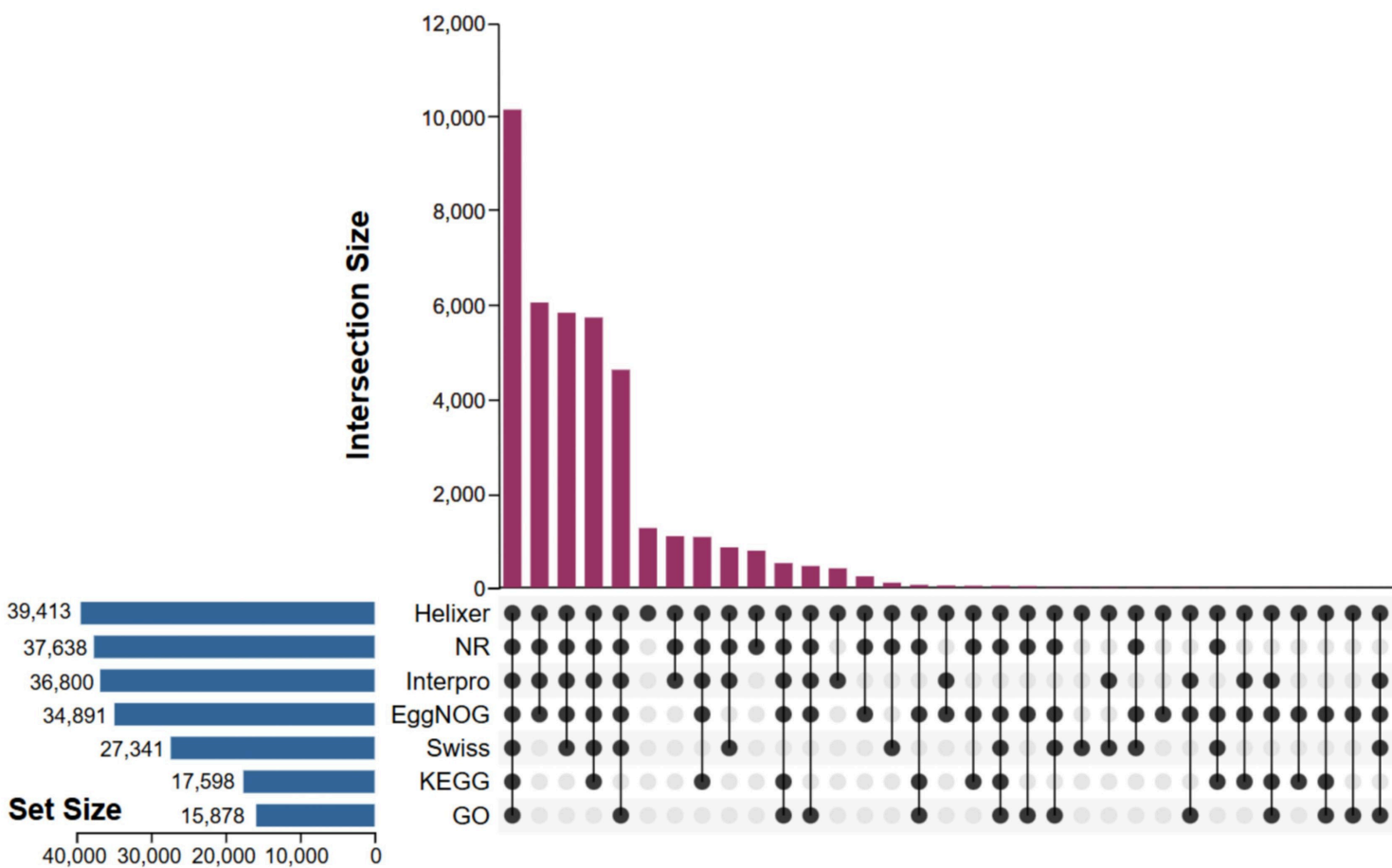


Figure 3

Functional annotations of predicted genes in Finger corn. The Upset plot shows the overlaps of genes across six annotation databases. Most predicted genes are supported by multiple databases, confirming the high reliability of the gene annotation.

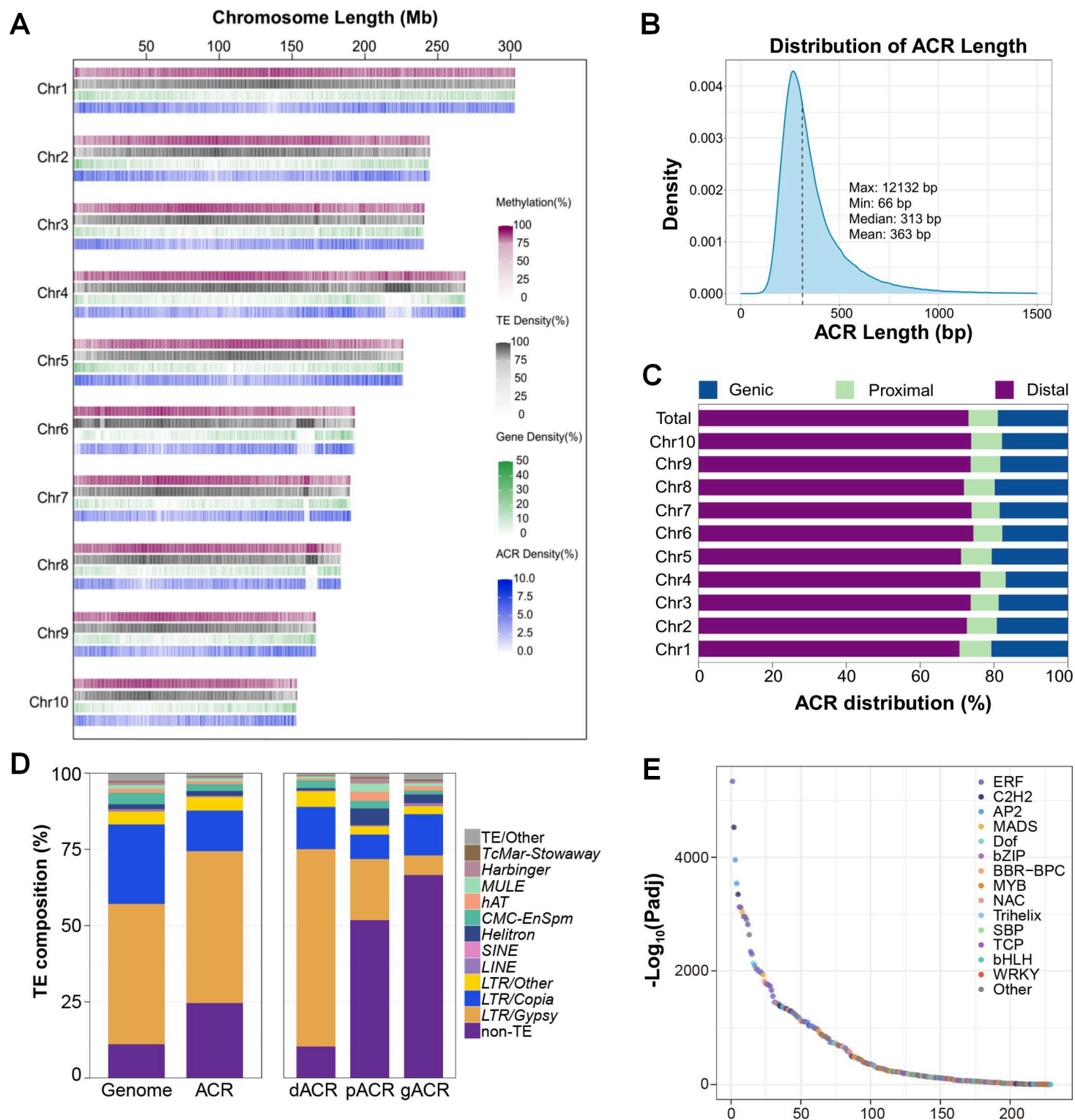


Figure 4

Genome-wide characterization of accessible chromatin regions in Finger Corn. (A) Chromosomal landscape of DNA methylation, transposable element (TE) density, gene density and accessible chromatin region (ACR) density across the 10 chromosomes. **(B-C)** Length and positional distribution of ACR. **(D)** Comparison of TE sequence composition among the whole genome and three subsets of ACRs. **(E)** The enrichment of transcription factor binding motifs in ACRs.

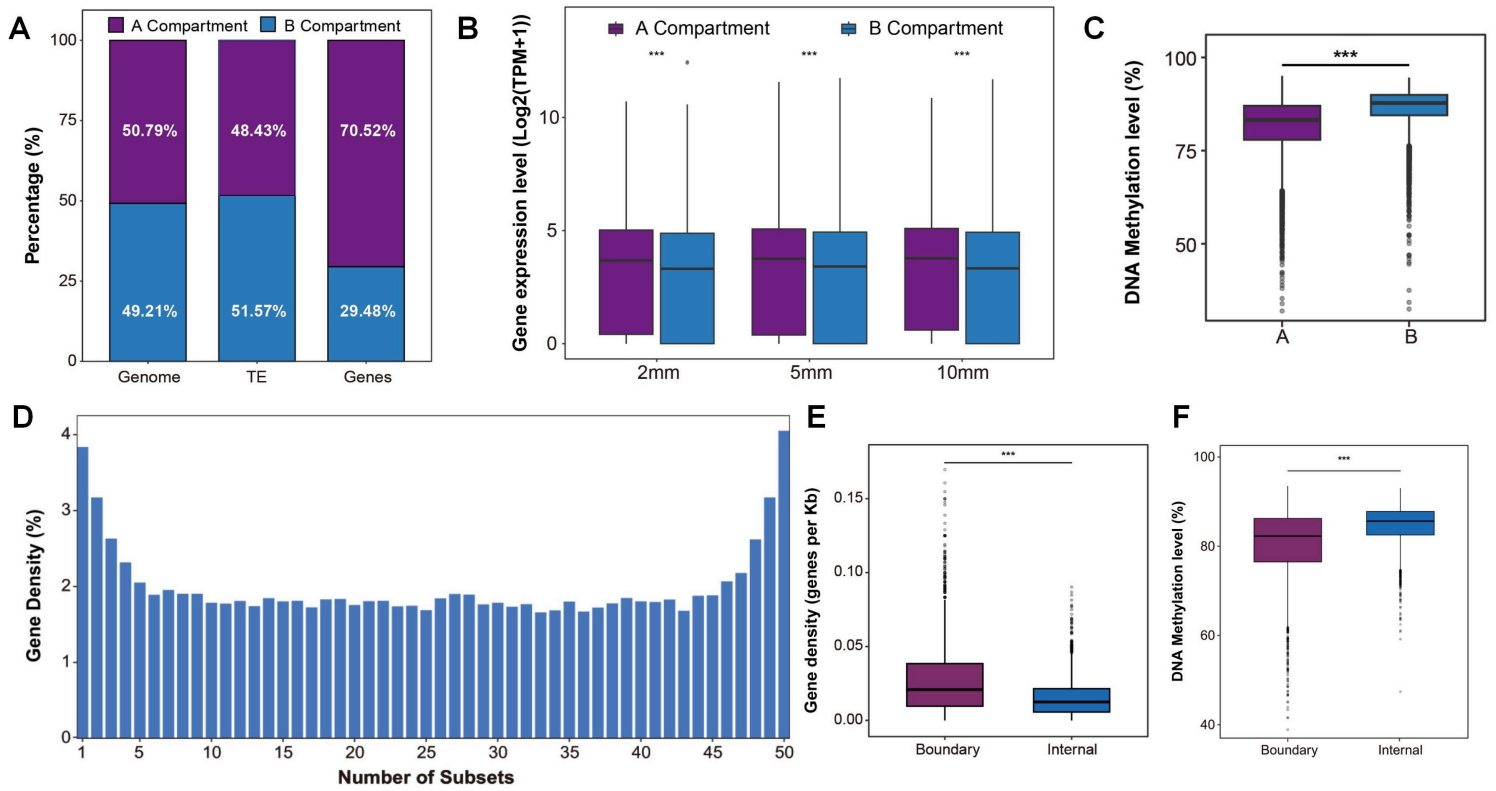


Figure 5

Chromatin compartments and TAD structures in the Finger corn genome. (A) Genome-wide proportions of A and B compartments, and their distribution patterns for TEs and genes. (B) Gene expression levels ($\log_2(\text{TPM}+1)$) of genes located in A and B compartments across three ear developmental stages (2 mm, 5 mm, and 10 mm). (C) Genome-wide DNA methylation levels between A and B compartments. (D) Distribution of gene density across 50 equal-distance subsets extending from TAD boundaries to internal regions. (E) Comparison of gene density between TAD boundary regions and internal regions. (F) Comparison of DNA methylation levels between TAD boundary regions and internal regions.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Table1.xlsx](#)
- [Table2.xlsx](#)
- [Table3.xlsx](#)