scientific data

Check for updates

OPEN

DATA DESCRIPTOR

Chromosome-scale genome assembly of *Zoysia japonica* uncovers cold tolerance candidate genes

Liangying Shen[1,4], Zewen Qi[1,2,4], Xiuru Dai[2], Ye Ai[1], Jiabao Chen[1], Yuehui Chao[1], Hang He[2], Liebao Han[1,3] ✉ & Lixin Xu[1] ✉

Zoysiagrass stands out as a crucial native turfgrass due to its exceptional abiotic stress tolerance, extensive adaptability, and high ornamental value. In this study, we generated a high-quality chromosome-level genome assembly of Compadre (COM) zoysiagrass, leveraging PacBio SMRT sequencing and Hi-C scaffolding technologies. The resulting genome assembly (312.42 Mb) is anchored on 20 chromosomes, with a Scaffold N50 of 18.72 Mb. In total, 49,074 genes and 306,768 repeat sequences were annotated in the assembled genome. The first chromosome-scale genome of *Zoysia japonica* 'Compadre' provides a critical genetic resource for cold-tolerant turfgrass breeding through identifying stress-responsive candidate genes. Additionally, we have successfully established a cell nucleus extraction and library construction protocol tailored for zoysiagrass ATAC-seq technology, and a total of 80 low temperature tolerance candidate genes were preliminarily identified via ATAC-seq and RNA-seq profiling, thereby initiating the exploration of turfgrass epigenomics.

Background & Summary

Zoysia japonica Steud., commonly known as zoysiagrass. Taxonomically, it belongs to the *Poaceae* family, the subfamily *Chloridoideae*, and the *Zoysia Wild* (*Zoysia*). *Zoysia* are perennial C4 allotetraploid plants (2n = 4x = 40) within the *Poaceae* family, typically possessing genome sizes ranging from 300–400 Mb[1]. Zoysiagrass showcases exceptional characteristics, including drought tolerance, heat resilience, cold hardiness, salt-alkali tolerance, trampling durability, adaptability to barren soil, disease resistance, and broad environmental adaptability. These qualities contribute to its widespread use in diverse lawn construction projects, as well as in soil and water conservation efforts and greening initiatives. Notably, zoysiagrass ranks among the most extensively employed warm-season turfgrass in China[2,3]. Compadre (COM) variety of zoysiagrass, characterized by its high ecotype purity and robust resistance, is a widely used commercial cultivar and is particularly well-suited for genome mapping endeavors. However, among the four major genera in the subfamily *Chloridoideae* (*Zoysia*, *Cynodon*, *Eleusine*, and *Eragrostris*), *Zoysia* remains the only genus lacking a publicly available chromosome-level genome assembly[4,5]. While Tanaka *et al.*[1] generated a draft genome of *Zoysia japonica* 'Nagirizaki' using Illumina short-read sequencing technology, resulting in a genome size of 334 Mb and an N50 of 2.3 Mb. Guan *et al.*[6] later provided a full-length transcriptome of *Z. japonica* using PacBio single-molecule long-read sequencing. Building on this, Yang *et al.*[7] generated a PacBio long-read assembly for the *Z. japonica* 'Yaji' cultivar, yet functional annotations remained incomplete. Therefore, obtaining detailed annotations for the zoysiagrass genome is essential to fully understand the genetic mechanisms underlying the development of its valuable traits.

Here we generated a chromosome-scale genome of COM zoysiagrass employing PacBio SMRT (Single Molecule Real Time) and Illumina sequencing techniques, along with Hi-C (High-throughput chromosome conformation capture) assisted assembly methods. Genomic DNA was extracted from the leaf tissue, and various libraries were constructed and sequenced using a suitable platform. After quality filtering and

[1]School of Grassland Science, Beijing Forestry University, Beijing, 100083, China. [2]Institute of Advanced Agricultural Sciences, Peking University, Shandong, 261325, China. [3]Engineering and Technology Research Center for Sports Field and Slope Protection Turf, National Forestry and Grassland Administration, Beijing, 100083, China. [4]These authors contributed equally: Liangying Shen, Zewen Qi. ✉e-mail: hanliebao@163.com; lixinxu@bjfu.edu.cn

trimming of the raw data, LACHESIS software was employed to assemble the genome. The integrity and quality of the chromosome-scale genome of COM zoysiagrass were evaluated from three distinct perspectives. Simultaneously, this study used multi-level assembly technology to perform repetitive sequence annotation, gene structure prediction, and gene function annotation on *Zoysia japonica*. The assembled genome for zoysiagrass spans 312.42 Mb with a Scaffold N50 of 18.72 Mb. Genome annotation identified 49,074 genes and characterized 306,768 repeat sequences. The assembled genome demonstrated an alignment rate 98% at the single-base level. Moreover, the genome's integrity was confirmed by high assessment scores from the CEGMA (98.69%) and BUSCO (97.65%) analyses. Taken together, this superior reference genome serves as an invaluable foundation for the conservation and utilization of zoysiagrass germplasm resources, as well as for the discovery of gene functions and breeding advancements.

Methods

Sample collection. The genomic sequencing data presented in this study pertains to the COM ecotype of zoysiagrass, with mature zoysiagrass seeds sourced from Beijing Forestry University. Prior to planting, the COM zoysiagrass seeds were immersed in a sterile solution for 20 min, followed by a thorough wash with sterile water. Subsequently, the seeds were sown in pots containing sterilized soil and cultivated in a light incubator. The photoperiod was maintained at a 16-hour light and 8-hour dark cycle for 90 days, with temperature controls set at 28 °C for the light phase and 23 °C for the dark phase, designated as the control group (CK). The multi-omics dataset was generated from six distinct zoysiagrass accessions: Qingdao(QD), Compadre (COM), Zenith(ZEN), and Mayer (M), representing commercially cultivated varieties, along with X1 and X4, which were originally collected from Northeast China by Beijing Forestry University as wild-type materials. For low-temperature stress group (L), zoysiagrasses were initially cultivated under normal cultivation conditions for 70 days, followed by a 20-day stress period with day and night temperature maintained at 9 °C and 4 °C, respectively. From each pot, 10 mature and fully unfolded leaves of these accessions were randomly selected for subsequent DNA, RNA-seq and ATAC-seq analysis.

DNA, RNA extraction, and ATAC libraries construction. DNA was extracted from the young zoysiagrass leaves after three months of cultivation utilizing the CTAB method[8]. The resulting purified DNA samples were dispatched to Beijing Biomarker Technologies Co, LTD. for sequencing. RNA extraction from zoysiagrass leaves was using the RNeasy Plant Mini Kit from QIAGEN. The experiment followed specific operating steps according to the kit instructions. After determining the concentration and quality of the RNA solution using the Fragment Analyzer 5400 (Agilent Technologies, CA, USA), the process included mRNA purification, fragmentation, first and second strand cDNA synthesis, end adaptation, size selection using the AMPure XP system (Beckman Coulter, Beverly, USA) for optimal fragment length, and PCR amplification. The library was then clustered using TruSeq PE Cluster Kit v3-cBot-HS (Illumina) and sequenced on an Illumina Novaseq 6000 platform by Novogene Technology Co, LTD to generate 150 bp paired-end reads.

Referring to the nuclear extraction reagent formulas of various *Gramineae* plants[9], zoysiagrass nuclei were extracted from 24 samples of zoysiagrass, categorized into control, low temperature, and freeze damage groups, and the library was constructed using Vazyme's V2TruePrep® DNA Library Prep-Kit V2 for Illumina®-TD501 reagent kit. The concentration of the zoysiagrass library was detected using Qubit, and preliminary detection was performed using Agilent 2100, followed by sequencing by Novogene Technology Co, LTD. For the formal genome sequencing, PacBio third-generation sequencing was conducted using P6/C4 chemistry inPacBio Sequel II, While DNA second-generation library sequencing and Hi-C library sequencing were performed using Illumina Novaseq 6000. A total of 63.72 Gb of high-quality data was generated using the Illumina platform. *K*-mer analysis suggested that the genome size of COM zoysiagrass was approximately 324.25 Mb, with an estimated repeat sequence content of around 32.24%. Sequencing was conducted Utilizing the Pacbio SMRT sequencing platform (PacBio Sequel II), we yielded approximately 163.97 Gb of long-read sequence data. The overall sequencing depth achieved was approximately 499.14 X, with a read N50 of 29.02 Kb (Table S1).

Contig assembly of COM zoysiagrass. After subjecting sequencing data to quality control using fastQC (v0.11.9), low-quality fragments were eliminated using Trimmomatic (v0.36)[10]. To ensure data quality, a second round of quality control was conducted using fastQC. Subsequently, the filtered data underwent error correction using Canu (v1.4)[11] software, followed by three iterative rounds of Pilon (v1.23)[12] software refinement with second-generation sequencing data by Smartdenovo (v1.0.0). Upon completion of the assembly and error correction processes, the COM zoysiagrass genome size of 313.28 Mb was obtained, encompassing 190 contigs exceeding 1 Kb in length, with a Contig N50 of 6.13 Mb and a GC content of 43.96% (Fig. 1A, Table 1).

Hi-C data analysis and pseudochromosome construction. The adapter sequences in the raw reads were excised, and low-quality paired-end reads were excluded to obtain clean data. The clean Hi-C reads, constituting 87.13% of the zoysiagrass genome, were initially truncated at the presumed Hi-C junctions. The Hi-C sequencing data yielded a total of 100 Gb, and subsequent to the Hi-C scaffolding process, the final reference genome size was determined to be 328.5 Mb. The sequencing depth, calculated based on the total data and the genome size with a coverage of 304X. Subsequently, the resulting trimmed reads were aligned to the assembly results using the BWA aligner. Paired reads were retained only if they were uniquely alignable with a mapping quality score (MAPQ) $\geq$20. Invalid read pairs—including dangling ends, self-ligation, re-ligation, and unmapped reads—were filtered using HiC-Pro (v2.8.1)[13]. Ultimately, 97.10% of unique mapped read pairs were deemed valid interaction pairs, which were then used for correction of scaffolds. These pairs were also utilized in the clustering, ordering, and orientation scaffolds onto chromosomes by LACHESIS (v1.0)[5].

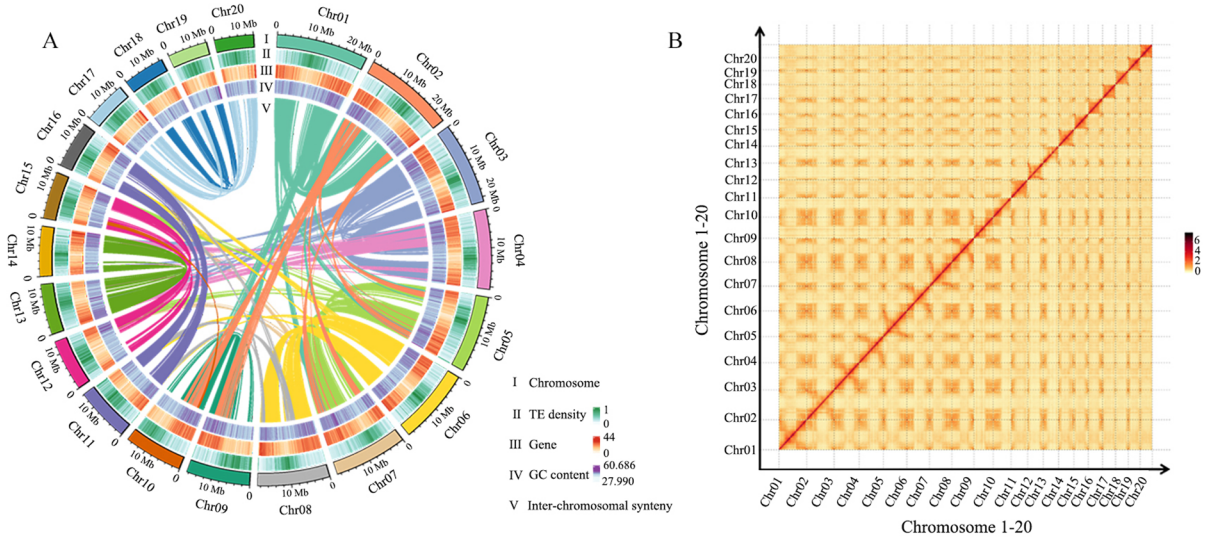

Fig. 1 Features of the COM zoysiagrass genome and Hi-C-assisted genome assembly. (**A**) The tracks from outermost to innermost indicate the following: I. 20 chromosomes with thick mark labeling each 2 Mb; II. transposable element (TE) density; III. Gene density; IV. GC content; and V. Whole-genome duplication (WGD) events shown by inter-chromosomal synteny. (**B**) Hi-C interaction heatmap showing 100-kb resolution superscaffolds.

Data type	PacBio + Illumina (Contig)	PacBio + Illumina + Hi-C (Scaffold)
Assembly size (Mb)	313.28	312.42
Scaffold N50 (Mb)	5.73	18.72
Number of scaffolds	190	81
GC content of the genome	43.96%	43.96%
Total length (Mb)	328.49	328.51
Minimum length (Mb)	6.83	6.83
Maximum length (Mb)	18.10	23.74
Mean length (Mb)	17.29	40.56
Median length (Mb)	3.90	1.65
Total low case counts	0	0
Total N counts	0	10900

Table 1. Genome assembly statistics of zoysiagrass 'Compadre'.

Prior to chromosome assembly, a pre-assembly stage was executed to correct errors within the scaffolds, necessitating the division of scaffolds into segments averaging 50 kb. BWA (was employed to map the Hi-C data to these segments. Uniquely mapped data were preserved for subsequent assembly using LACHESIS. Any two segments displaying inconsistent connections with information from the raw scaffold were manually verified. The corrected scaffolds were then assembled with LACHESIS. Finally, utilizing the LACHESIS software for genome sequence grouping, sorting, and orientation, a total of 312.42 Mb of genome sequences was successfully mapped onto the 20 chromosomes of COM zoysiagrass, with scaffold N50 reaching 18.72 Mb (Fig. 1, Table 1). The heatmap demonstrates that interaction intensities observed at diagonal positions within each group surpasses those at non-diagonal positions. This finding indicates an increased intensity of interaction between adjacent regions of the chromosome sequence facilitated by Hi-C assembly, thereby substantiating the efficacy of genome assembly (Fig. 1B). Furthermore, it confirms that zoysiagrass is a tetraploid plant, with a chromosome count of 2n = 40.

Repeat sequences and gene annotation. The transposable elements (TEs) of the zoysiagrass were constructed using structural prediction and *de novo* prediction principles. LTR_FINDER[14] and RepeatScout[15] software were employed to establish a TEs database for zoysiagrass. PASTE Classifier[16] was then utilized to classify this database. The resulting zoysiagrass database was amalgamated with Repbase[17] to create the final TEs database for zoysiagrass. RepeatMask[18] software was employed to predict TEs within the constructed database. A cumulative sum of 131.66 Mb repeat sequences has been annotated within the genome of zoysiagrass. Among these, TEs account for 117.04 Mb, with the long terminal repeat (LTR) comprises 82.59 Mb, predominantly of Class I/LTR/Gypsy type, followed by Class I/LTR/Copia type. Upon extraction of flanking sequences surrounding the

Metric	COM zoysiagrass
Total number of repeat sequences	306,768
Total size of repeat sequences (Mb)	131.66
Number of TEs	252,482
Size of TEs (Mb)	117.04
Size of LTR (Mb)	82.59
Number of SSR	2,077
Number of genes	49,074
Mean gene length (bp)	3,022.15
Number of exons	49,074
Mean exons length (bp)	1,269.02
Number of introns	190,566
Mean introns length (bp)	1,753.13
Number of CDS	234,874
Mean CDS length (bp)	1,080.45

Table 2. Statistics of COM zoysiagrass genome annotation.

LTR of zoysiagrass for comparative analysis and subsequent distance calculations using the Kimura model, two LTR insertion events in the zoysiagrass genome were identified, occurring approximately 0.9 and 2 Mya. We also identified 2,077 simple sequence repeats (SSRs) for the assembled genome (Table 2).

Gene prediction. To predict the genes within the zoysiagrass genome, three primary strategic methods were employed: *de novo* prediction (Ab initio), homologous species prediction, and unigene prediction (RNA-seq). *De novo* prediction involved the use of Genscan[19], Augustus[20], GlimmerHMM[21], GeneID[22], and SNAP[23] software. Homologous species prediction relied on the GeMoMa[24] software. In case where transcriptome data lacked a reference assembly, PASA[25] software was employed for Unigene prediction. Finally, EVM[26] software was used to integrate gene prediction results, which were further refined with PASA. Finally, the integration of results from these three methods yielded a total of 49,074 genes, comprising 239,640 exons and 190,566 introns, with 234,874 coding sequences (CDS). The average gene length was 3,022.15 bp, with each gene containing 4.88 exons. The average lengths of CDS, exons, and introns were 1,080.45, 1,269.02, and 1,753.13 bp, respectively. Gene density, GC content, and TE density on each pseudomolecule are illustrated in a circular diagram (Fig. 1A). Additionally, 49 RNA families were annotated as noncoding RNAs, including 181 miRNAs, 497 tRNAs, and 374 rRNAs.

Gene functional annotation. Gene functional annotation of predicted gene sequences in zoysiagrass was conducted using functional databases such as NR[27], KOG[28], GO[29], KEGG[30], and TrEMBL[31]. Homology comparisons with the database were performed using BLAST[32] software. This analysis included comprehensive gene KEGG pathway annotation, KOG functional annotation, and GO functional annotation. A total of 36,67 genes were annotated, accounting for 74.71% of the total genes.

Comparative genomics analysis. To elucidate the evolutionary dynamics and functional diversity of gene regulation mechanisms in zoysiagrass, we conducted comparative genomics studies by retrieving the sequences from *Triticum aestivum*[33], *Oryza sativa*[34], *Setaria viridis*[35], *Brachypodium distachyon*[36], *Setaria italica*[37], and *Dioscorea Cayenensis*[38]. Orthofinder[39] was used for family clustering (E-value of 0.001), MAFFT[40] software was used to analyze the gene copy number of each gene family, IQ-TREE[41] was used to obtain the optimal model (JTT + F + I + G4), and to construct evolutionary tree by using the maximum likelihood (ML) method with a bootstrap count of 1000, followed by Gblocks[42] to retain well-aligned regions. GO and KEGG enrichment analysis on the expansion and contraction gene families was performed using Cluster Profile[43]. CAFE (v.4.2.1) software[44] was used to estimate the contraction and expansion of zoysiagrass gene family, PAML[45] was used for positive selection analysis, and WGD software[46] was used to draw the distribution of Ks values and analyzed the time and type of gene replication events.

The analysis yielded a total of 36,205 gene families, including 2,258 single-copy genes (Fig. 2A). Using the sequences of these 2,258 single-copy genes, we constructed an evolutionary tree with *Dioscorea Cayenensis* as an outgroup (Fig. 2B). Positive selection analysis across zoysiagrass gene families identified 674 genes under positive selection ($P < 0.05$). Comparative genomic analyses revealed that 194 expanded gene families and 11 contracted families in zoysiagrass. Functional enrichment analysis of expanded families demonstrated significant associations with stress resistance and adaptability, including calcium-mediated signaling, photomorphology regulation, hypotonic salinity response, and DNA integration processes. Additionally, the genes were found to be enriched in biological processes related to the physiological and biochemical responses of organisms to mechanical stimuli, suggesting a potential connection to zoysiagrass's tolerance to pruning (Fig. 2C). The expansion and contraction of gene families typically resulted from WGD, which was concomitant with the phenomenon of genome duplication doubling. To enhance comprehension of the evolutionary history of plants and the evolution of functional genes, a whole-genome replication event analysis was conducted on zoysiagrass. The timing of the WGD was computed based on differentiation time (Fig. 2B), revealing that the event occurred approximately 15–20 million years ago (Fig. 2D)

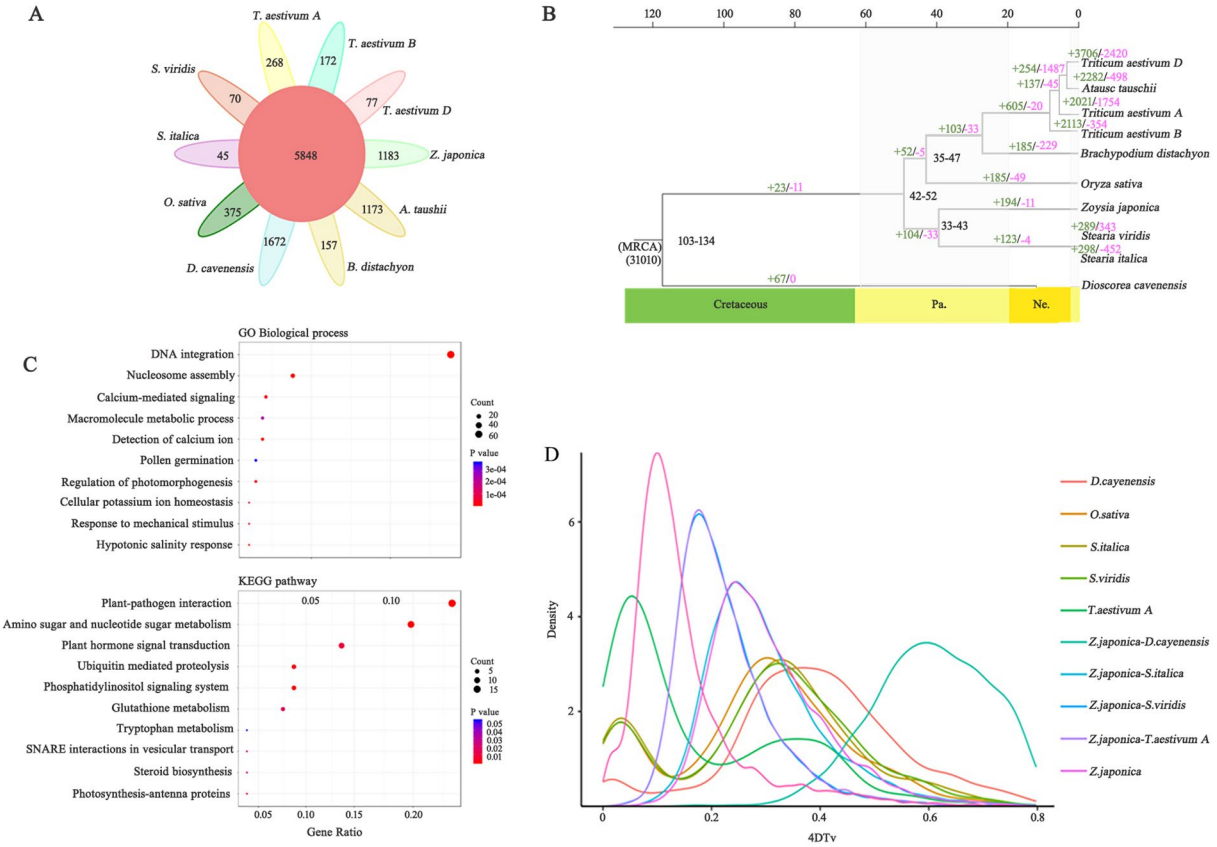

Fig. 2 Gene family and phylogenetic tree analyses of COM zoysiagrass and other representative plant genomes. (**A**) Venn diagram of the number gene families within Z. *japonica, T. aestivum, O. sativa, S.viridis, B. distachyon, A. tauschii, S. italica* and *D. cayenensis*. (**B**) Gene family clustering in Z. *japonica* and other plant genomes, gene family expansions and contractions among them and a phylogenetic tree based on single-copy gene families. (**C**) Functional analysis of gene family expansions in *Z. japonica*. (**D**) Distribution of fourfold degenerate sites of the third codons (4DTv) for syntenic genes within Z. *japonica, T. aestivum, O. sativa, S.viridis, B. distachyon, A. tauschii, S. italica* and *D. cayenensis*.

ATAC-seq and RNA-seq data analysis. ATAC-seq sequencing was conducted on 23 materials. Among these, the 'COM' accessions included two biological replicates and two technical replicates for both the low-temperature treatment and the CK treatment. The remaining included only biological replicates. Additionally, the 'ZEN' ATAC-seq material was missing for the low-temperature treatment, resulting in 13 valid SRR accessions and achieving a single sample sequencing depth of 10 G, which is approximately 19 X coverage. After quality control of ATAC-seq data, BWA (v0.7.10-r789)[47] and Bowtie2(v2.4.5)[47] were used for comparison. SAMtools (v0.1.19) compressed and sorted SAM format files and stored them as BAM format files. The multiBamSummary command in deeptools (v3.5.1) software was used to calculate the coverage of sequencing data. The plotCorrelation or plotPCA command was used to calculate the correlation of zoysia ATAC-seq data. The MarkDuplicates function in Picard (v2.20.3) was used to remove PCR duplication, SAMtools was used for filtering, and grep-v chrM and grep-v chrP were used to remove mitochondrial chloroplasts. The macs2 (v2.1.0) was used to find the calling peak, and the DiffBind (v2.14.0) toolkit in R(v4.2.3) was used to process the peak. The findings indicated a strong correlation among the data from the same treatment group, while there was a significant divergence in the data across different treatment groups. Additionally, the analysis revealed a high number of peaks with robust signal strength (Fig. 3A,B).

For the 24 RNA-seq datasets obtained from six zoysiagrass accessions, each treatment includes two biological replicates and two technical replicates. Raw paired reads quality was initially assessed using FastQC (v0.11.9). Low-quality sequences were filtered using Trimmomatic(v0.36), followed by secondary quality control to ensure data reliability. Quality-controlled reads were then aligned to the COM reference genome with orientation mode using HISAT2(v2.2.1), generating SAM files that were subsequently converted to sorted BAM files using Picard and SAMtools. Genome coverage was calculated using multiBamSummary from deeptools (v3.5.1) with a default 10 kb interval, while sample correlation analysis was performed using either plotCorrelation or plotPCA functions. Read quantification was conducted using HTseq-count (v0.11.2) to generate count matrices, followed by differential gene expression analysis with DESeq. 2 (v1.26.0) using a significance threshold of $P < 0.01$. Functional annotation of differentially expressed genes was carried out using MapMan and GO databases. Statistical significance was evaluated using Fisher's exact test in R, with results visualized through ggplot2. The constructed gene co-expression network clustered into 24 distinct modules (Fig. 3C–E).

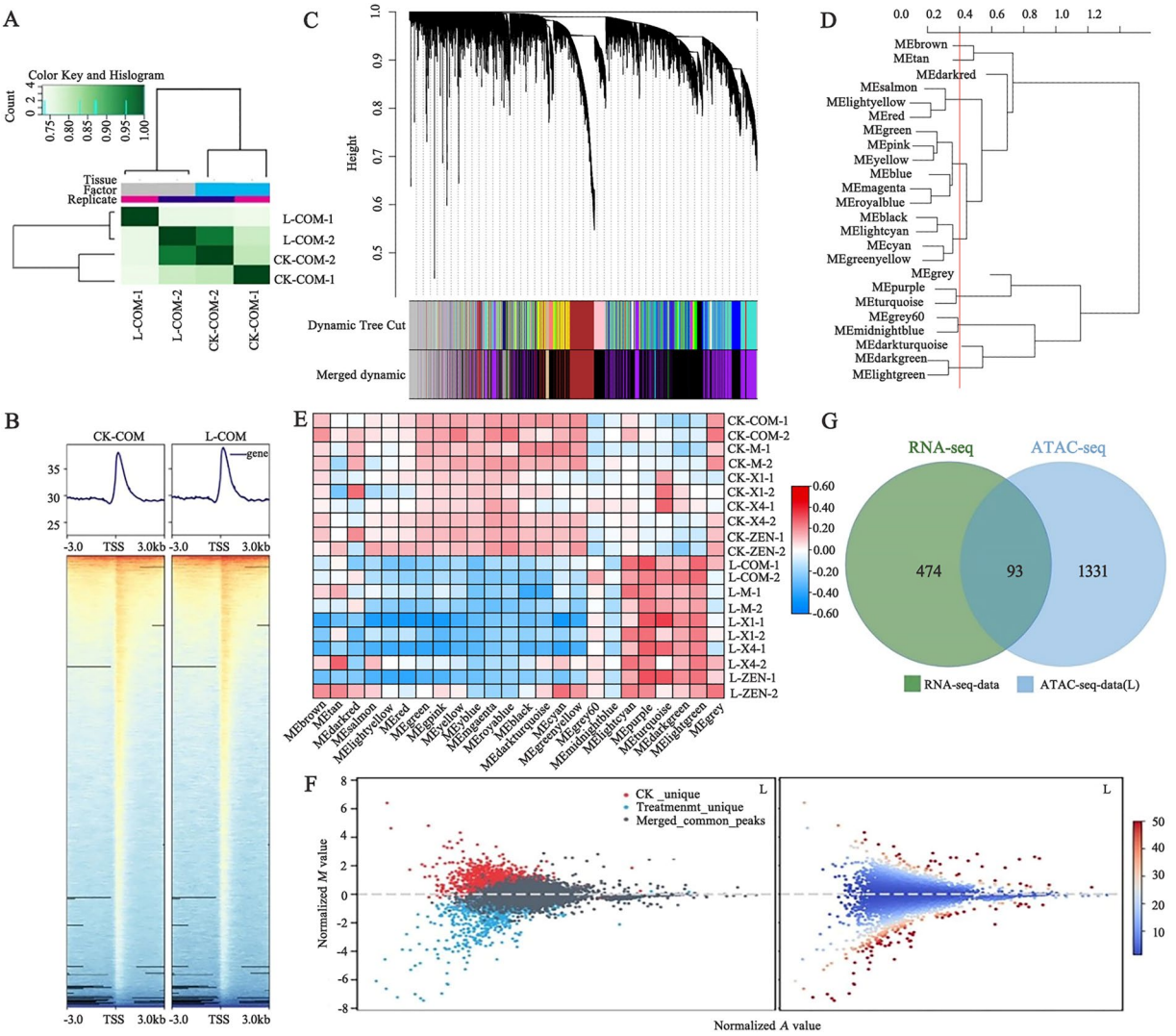

Fig. 3 Analysis of ATAC-seq and RNA-seq data. (**A**) The repeatability evaluation of zoysia ATAC-seq data. (**B**) The signal of calling peak. (**C**) Zoysiagrass gene expression and clustering. (**D**) Module phylogenetic tree analysis. (**E**) Module gene co-expression heatmap. (**F**) Differential peak analysis of zoysiagrass. The red solid dot is the peak of the C, the blue solid dot is the peak of the treatment group, and the gray solid dot is the common peak. 'L' is the low temperature treatment group. (**G**) The joint analysis of RNA-seq and ATAC-seq.

Functional enrichment analysis of the 24 modules revealed 108 significantly enriched pathways, including 35 associated with low-temperature tolerance (Fig. S1). The integrative analysis identified four cold-responsive modules (MEpurple, MEturquoise, MEdarkturquoise, and MElightgreen) significantly enriched in key biological processes such as ubiquitin-dependent protein catabolism, chlorophyll degradation, carotenoid biosynthesis, glutathione metabolism, and negative regulation of catalytic activity. These processes are linked to cold adaptation and chlorophyll degradation in zoysiagrass, suggesting the modules' potential roles in winter coloration and cold acclimation.

MAnorm (v1.3.0)[48] was utilized to analyze and count the difference in peak expression. To more clearly visualize the differences between the datasets, two types of visualization maps were constructed. The distribution of the differential peak expression was observed following data normalization, and peaks with significant differences were identified through the variation in peak expression (Fig. 3F). Gene association and annotation for the regulation of differential genes were conducted based on the location information of the differential peaks. As a result, 1,424 genes regulated by differential peaks and associated with low temperature treatment were identified (Fig. 3F,G). Subsequently, yielding 93 candidate genes associated with low temperature (Fig. 3G). KEGG enrichment analysis of cold-responsive genes was performed using TBtools software (v2.154), followed by visualization of the enriched pathways with the ggplot2 package in R. KEGG analysis revealed key pathways associated with cold adaptation in zoysiagrass, including photosynthesis, plant hormone signaling, protein family metabolism, environmental stress response, and glutaredoxin biosynthesis (Fig. S2). We performed RT-qPCR on the discovered genes, and the results were largely consistent with the transcriptome data trend, leading to the identification of 80 related candidate genes (Table S2). These pathways enhance cold tolerance by modulating

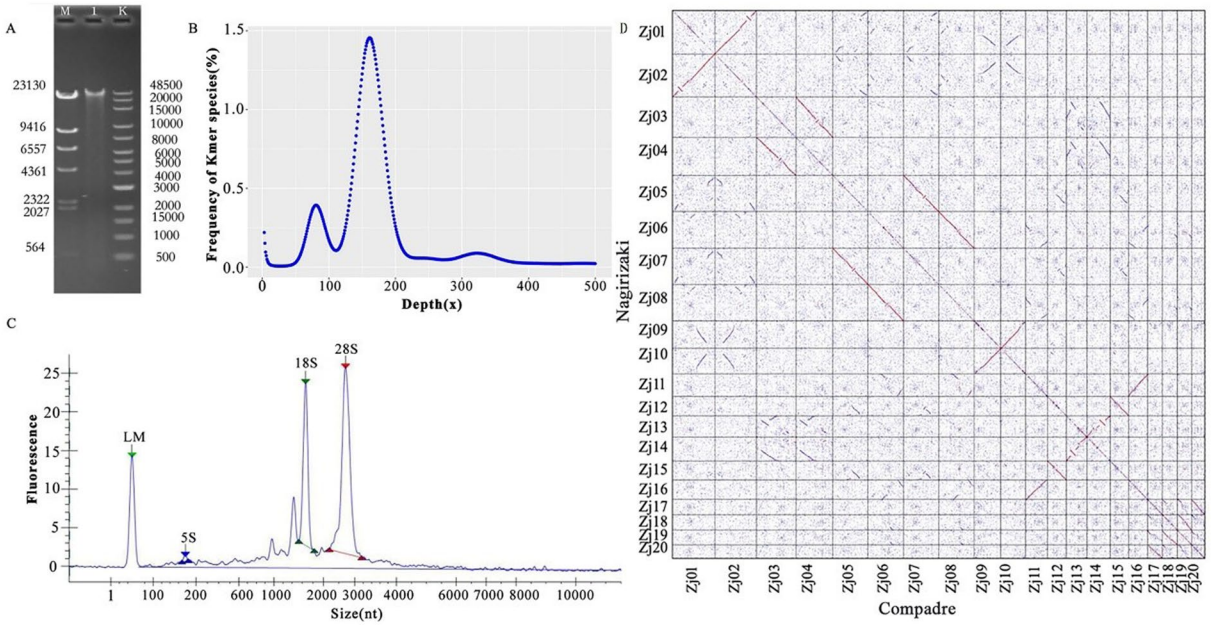

Fig. 4 Evaluation of zoysiagrass genome characteristics. (**A**) Zoysiagras genomic DNA agarose gel electrophoresis. (**B**) K-mer frequency distribution. (**C**) rRNA size distribution. (**D**) Genome collinearity analysis.

photosynthetic efficiency, hormone signaling, protein homeostasis, and cellular stability. Cold-responsive genes were identified within these pathways, such as the photosynthesis-related expansion genes *Zja04G017410* and *Zja14G019500*, whose downregulation is linked to leaf chlorosis under cold stress. The positively selected gene *Zja09G017200*, involved in protein family metabolism, exhibits expression levels inversely correlated with cold tolerance. In contrast, *Zja06G019230* (a homolog of *LHCB*) and several CONSTANS-like genes (e.g., *Zja05G026720*, *Zja08G010730*, *Zja11G006650*, *Zja10G012190*, *Zja20G003530*) exhibit expression levels positively associated with enhanced cold tolerance (Fig. S3). This approach allowed for a more targeted identification of genes that were not only differentially expressed under the specified conditions but also had associated changes in chromatin accessibility, providing a deeper understanding of the regulatory mechanisms at play.

Data Records

The assembled Compadre zoysiagrass genome (GCA_040438285.1)[49], Pacbio (SRX22497642)[50], Hi-C (SRX22497643)[51], Illumina (SRX22497644)[52] and WGS data have been deposited at the NCBI under project PRJNA1036829. Additionally, the RNA-seq and ATAC-seq (SRX22552165–SRX22552188, SRX22496703–SRX22496715)[53] data are publicly available under PRJNA1037028. For broader accessibility, we have deposited the assembled genome, annotation and protein sequences data in the figshare repository[54].

Technical Validation

Evaluation of genome size and complexity. Assessment of sequencing data contamination was performed using BLAST, while SOAP[55] was utilized to evaluate nuclear DNA content. KmerGenie[56] was engaged to discern gene size, repeat sequence ratio, and heterozygosity. The Illumina sequencing platform's sequencing data has a Q20 ratio of over 98.23% and and a Q30 ratio of over 94.88%, indicating that the sequencing data quality is very high with low error rates. Upon comparison of the 350 bp library with the NT database, no contamination was found in the samples with an aligned percentage of 1.89%, which ensures the accuracy and reliability of the sequencing data. Additionally, when assessing the extranuclear DNA content of the 350 bp library, it was found that the paired-read accounted for only 0.01%, indicating a low content. This low level of extranuclear DNA makes it suitable for evaluating genome size. Notably, no distinct heterozygous peak was observed, and the heterozygosity was estimated to be about 0.60%. The low heterozygosity of COM zoysiagrass proved advantageous for genome sequencing assembly (Fig. 4A–C). The assembled genome size is 312.42 Mb, and the genome's N50 reaches 18.72 Mb, marking a triple increase compared to Nagirizaki zoysiagrass. The COM zoysiagrass genome attains the chromosome level, exhibiting superior integrity, continuity, making it more suitable for functional genomics and molecular breeding research.

The analysis of zoysiagrass genome assemblies revealed distinct characteristics among Nagirizaki, Yaji, and Compadre varieties. The Yaji variety, sequenced with PacBio-RSII technology, had a genome size of 373 Mb with an N50 value of 3.96 Mb, indicating good genomic continuity despite a lower sequencing depth compared to Compadre. Compadre excelled in chromosome-level genome assembly, achieving the highest CEGMA score of 98.69% and a chromosome coverage of 96.3%, significantly outperforming Nagirizaki. Yaji's gene prediction accuracy was high, with 50,140 predicted genes matching the number of annotations, while Compadre had slightly fewer gene annotations at 40,082. Yaji's genome coverage was 91.8%, higher than Nagirizaki's 85.6%

Type	Nagirirzaki	Yaji	Compadre
assembly genome size	334.38 Mb	373 Mb	313.29 Mb
sequencing technology	Illumina	PacBio RSII	Pacbio Sequel II Illumina
sequencing depth	49.3 Gb	38.6 Gb	163.97 Gb
N50	2.37 Mb	3.96 Mb	18.72 Mb
chromosome-level genome	No	No	Yes
GC content	44.1%	44.17%	43.96%
repeat sequence	35.6%	/	40.8%
predicted gene	49,103	50,140	49,074
gene annotation	/	50,140	40,082
CEGMA evaluation	97.6%	96.50%	98.69%
genome coverage	85.6%	91.8%	96.6%
chromosome coverage	70.2%	/	96.3%

Table 3. Comparative analysis of genome assembly of Nagirizaki, Yaji and Compadre.

but lower than Compadre's 96.6%. These insights are valuable for gene function research and the development of molecular markers for zoysiagrass. The Yaji variety exhibited improvements in genomic continuity, gene prediction accuracy, and genome coverage, while the Compadre variety showed significant advantages in chromosome-level genome assembly and overall genome assembly quality (Table 3). These findings offer valuable insights for gene function research and the development of molecular markers for zoysiagrass.

Quality assessment of the genome assembly. The quality and integrity of the COM zoysiagrass genome assembly were assessed from multiple methods. The alignment rate was assessed for the integrity of the assembled genome. BWA[57] software was employed to align second-generation sequencing sequences with the assembled zoysiagrass genome, achieving a mapping ratio of 98.66%. The completeness of the assembled zoysiagrass genome was evaluated using the default parameters of BUSCO (v4.0.5)[58] and CEGMA (v2.5)[59] software. BUSCO evaluation of the zoysiagrass genome revealed a 97.65% proportion of complete BUSCOs, indicating excellent integrity in the assembled zoysiagrass genome. As for CEGMA, a total of 4525 conserved core genes (98.69%) and 235 highly conserved genes (94.76%) were identified. Meanwhile, the strong interactions between proximal chromosomal sequences, as revealed by Hi-C, corroborate superior genome assembly outcomes (Fig. 1). Furthermore, the COM zoysiagrass genome has achieved reference quality, with a LTR Assembly Index (LAI) score of 13.54. Collinearity analysis revealed high similarity between the COM and Nagirizaki genomes (Fig. 4D). Upon employing Merqury(v1.3)[60] for the assessment of the chromosome-level genome assembly, in conjunction with all second-generation survey map data, the resulting QV value of 33.29 was observed to be comparatively favorable, indicating a high level of confidence in the assembly's accuracy (Fig. S4). In conclusion, the validation outcomes confirmed the high accuracy and completeness of the *de novo* assembled COM zoysiagrass genome. In contrast to the previously assembled genome, the COM zoysiagrass genome sequencing employs third-generation sequencing technology, providing a substantial advancement in obtaining raw data.

Quality assessment of the ATAC-seq and RNA-seq data. The nuclei, exemplified by the control group COM zoysiagrass, were observed to be spherical and intact (Fig. S5A), with minimal cellular disruption (Fig. S5B), indicating that the quality of the extracted nuclei was superior and well-suited for subsequent library construction. Upon completion of library construction, the quality of the libraries was evaluated, revealing optimal size of library bands and peak detection values when using the control group COM zoysiagrass as a reference. The libraries demonstrated high quality, making them suitable for further sequencing processes (Fig. S5C,D). These findings confirm the reliability of the data, deeming it appropriate for the subsequent identification of resistance genes in zoysiagrass. Furthermore, following the filtration of raw data from 24 zoysiagrass RNA-seq samples, alignment was performed against the previously assembled COM zoysiagrass genome using the HISAT2 software(v2.2.1). Post-alignment, the data underwent redundancy analysis. Anomalies and outliers within the dataset were identified and eliminated through the application of Relative Log Expression (RLE) and Variance Stabilizing Transformation (VST) methodologies, ensuring the precision and reliability of the data. The results showed strong correlations among data from the same treatment group and significant differences between various treatments, thereby validating the data's reliability and its utility for the subsequent exploration of zoysiagrass tolerance genes.

Code availability

All data processing was conducted using software tools selected based on the bioinformatic methods detailed in the methods section. Each software tool was strictly implemented according to the manufacturer's protocol. In instances where specific parameter settings were not explicitly detailed in the manual, the software's default settings were applied. Below, we provide a detailed description of the software versions, settings, and parameters utilized in our analysis, with default parameters being applied unless otherwise noted.

KmerGenie v 1.7051: -m 260 -x 440
LACHESIS v1: cluster_min_re_sites = 52, cluster_max_link_density = 2, order_min_n_res_in_trun = 54, order_min_n_res_in_shreds = 55

BUSCO v4.0.5:–E-value 1e-03 (E-value cutoff for BLAST searches) and -sp (reference species for gene prediction)
CAFE v.4.2.1: -p 0.05 -t 10 -r 10000
RepeatMask v1.331: -nolow -noIs -norna with the WUBlast engine
PASA v2.3.3: -align parameters_tools gmap and -maxIntronLen 20000
BLAST v2.15.0: -e value 1e-5
IQ-TREE v1.6.12: -m MFP -nt AUTO -cmac15 -redo -bb 1000
Gblocks v0.91b: - b5 = h
Macs2 v2.1.0: -f BAMPE, -keep up all, -q 0.01, and -cutoff analysis
Trimmomatic v0.36: --phred33 --leading:3 --trailing:3 --minlen:51

Received: 24 April 2024; Accepted: 13 March 2025;
Published online: 03 April 2025

References

1. Tanaka, H. *et al*. Sequencing and comparative analyses of the genomes of zoysiagrasses. *DNA Res* **23**, 171–180 (2016).
2. Guo, H. *et al*. Genetic linkage map construction and qtl mapping of salt tolerance traits in zoysiagrass (*Zoysia japonica*). *PLoS ONE* **9**, e107249 (2014).
3. Zhang, D. *et al*. Physiological and molecular responses of *Zoysia japonica* to rust infection. *International Journal of Molecular Sciences* **23**, 4185 (2022).
4. Cui, F. *et al*. The genome of the warm-season turfgrass African bermudagrass (*Cynodon transvaalensis*). *Horticulture Research* **8**, 93 (2021).
5. Burton, J. N. *et al*. Chromosome-scale scaffolding of *de novo* genome assemblies based on chromatin interactions. *Nat Biotechnol* **31**, 1119–1125 (2013).
6. Guan, J. *et al*. Single-molecule long-read sequencing analysis improves genome annotation and sheds new light on the transcripts and splice isoforms of *Zoysia japonica*. *BMC Plant Biol* **22**, 263 (2022).
7. Yang, D.-H., Jeong, O.-C., Sun, H.-J., Kang, H.-G. & Lee, H.-Y. Genome analysis of Zoysia japonica 'Yaji' cultivar using PacBio long-read sequencing. *Plant Biotechnol Rep* **17**, 275–283 (2023).
8. Kuo, P., Henderson, I. R. & Lambing, C. CTAB DNA extraction and genotyping-by-sequencing to map meiotic crossovers in plants. *Methods in Molecular Biology* **2484**, 43–53 (2022).
9. Qi, Z. A method of nuclei extraction and library construction for chromatin transposase accessibility sequencing in gramineous plants. *Acta agronomica sinica* **49**, 1799–1807 (2023).
10. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
11. Koren, S. *et al*. Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Res* **27**, 722–736 (2017).
12. Walker, B. J. *et al*. Pilon: An integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS ONE* **9**, e112963 (2014).
13. Servant, N. *et al*. HiC-Pro: an optimized and flexible pipeline for Hi-C data processing. *Genome Biol* **16**, 259 (2015).
14. Xu, Z. & Wang, H. LTR_FINDER: an efficient tool for the prediction of full-length LTR retrotransposons. *Nucleic Acids Research* **35**, W265–W268 (2007).
15. Price, A. L., Jones, N. C. & Pevzner, P. A. *De novo* identification of repeat families in large genomes. *Bioinformatics* **21**, i351–i358 (2005).
16. Hoede, C. *et al*. PASTEC: An automatic transposable element classification tool. *PLoS ONE* **9**, e91929 (2014).
17. Jurka, J. *et al*. Repbase Update, a database of eukaryotic repetitive elements. *Cytogenet Genome Res* **110**, 462–467 (2005).
18. Tarailo-Graovac, M. & Chen, N. Using repeatmasker to identify repetitive elements in genomic sequences. *CP in Bioinformatics* **25**, (2009).
19. Burge, C. & Karlin, S. Prediction of complete gene structures in human genomic DNA. *Journal of Molecular Biology* **268**, 78–94 (1997).
20. Stanke, M. & Waack, S. Gene prediction with a hidden Markov model and a new intron submodel. *Bioinformatics* **19**, ii215–ii225 (2003).
21. Majoros, W. H., Pertea, M. & Salzberg, S. L. TigrScan and GlimmerHMM: two open source *ab initio* eukaryotic gene-finders. *Bioinformatics* **20**, 2878–2879 (2004).
22. Blanco, E., Parra, G. & Guigó, R. Using geneid to Identify Genes. *CP in Bioinformatics* **00**, (2003).
23. Korf, I. Gene finding in novel genomes. *BMC Bioinformatics* **5**, 59 (2004).
24. Keilwagen, J. *et al*. Using intron position conservation for homology-based gene prediction. *Nucleic Acids Research* **44**, e89–e89 (2016).
25. Campbell, M. A., Haas, B. J., Hamilton, J. P., Mount, S. M. & Buell, C. R. Comprehensive analysis of alternative splicing in rice and comparative analyses with Arabidopsis. *BMC Genomics* **7**, 327 (2006).
26. Haas, B. J. *et al*. Automated eukaryotic gene structure annotation using EVidenceModeler and the program to assemble spliced alignments. *Genome Biol* **9**, R7 (2008).
27. Marchler-Bauer, A. *et al*. CDD: a Conserved Domain Database for the functional annotation of proteins. *Nucleic Acids Research* **39**, D225–D229 (2011).
28. Koonin, E. V. *et al*. A comprehensive evolutionary classification of proteins encoded in complete eukaryotic genomes. *Genome Biol* **5**, R7 (2004).
29. Dimmer, E. *et al*. The UniProt-GO annotation database in 2011. *Nucleic Acids Research* **40**, D565–D570 (2011).
30. Kanehisa, M. KEGG: Kyoto Encyclopedia of genes and genomes. *Nucleic Acids Research* **28**, 27–30 (2000).
31. Boeckmann, B. The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003. *Nucleic Acids Research* **31**, 365–370 (2003).
32. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *Journal of Molecular Biology* **215**, 403–410 (1990).
33. Zhu, T. *et al*. Optical maps refine the bread wheat *Triticum aestivum* cv. Chinese Spring genome assembly. *Plant J* **107**, 303–314 (2021).
34. Kawahara, Y. *et al*. Improvement of the *Oryza sativa* Nipponbare reference genome using next generation sequence and optical map data. *Rice (N Y)* **6**, 4 (2013).
35. Wang, S. & Gao, L.-Z. Complete chloroplast genome sequence of green foxtail (*Setaria viridis*), a promising model system for C4 photosynthesis. *Mitochondrial DNA A DNA Mapp Seq Anal* **27**, 3707–3708 (2016).
36. International Brachypodium Initiative. Genome sequencing and analysis of the model grass Brachypodium distachyon. *Nature* **463**, 763–768 (2010).
37. Bennetzen, J. L. *et al*. Reference genome sequence of the model plant Setaria. *Nat Biotechnol* **30**, 555–561 (2012).
38. Mariac, C. *et al*. Cost-effective enrichment hybridization capture of chloroplast genomes at deep multiplexing levels for population genetics and phylogeography studies. *Mol Ecol Resour* **14**, 1103–1113 (2014).
39. Emms, D. M. & Kelly, S. OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol* **20**, 238 (2019).
40. Katoh, K., Asimenos, G. & Toh, H. Multiple alignment of DNA sequences with MAFFT. *Methods Mol Biol* **537**, 39–64 (2009).

41. Nguyen, L.-T., Schmidt, H. A., von Haeseler, A. & Minh, B. Q. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol* **32**, 268–274 (2015).
42. Talavera, G. & Castresana, J. Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Syst Biol* **56**, 564–577 (2007).
43. Yu, G., Wang, L.-G., Han, Y. & He, Q.-Y. ClusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS* **16**, 284–287 (2012).
44. Han, M. V., Thomas, G. W. C., Lugo-Martinez, J. & Hahn, M. W. Estimating gene gain and loss rates in the presence of error in genome assembly and annotation using CAFE 3. *Mol Biol Evol* **30**, 1987–1997 (2013).
45. Yang, Z. PAML: a program package for phylogenetic analysis by maximum likelihood. *Comput Appl Biosci* **13**, 555–556 (1997).
46. Zwaenepoel, A. & Van de Peer, Y. wgd-simple command line tools for the analysis of ancient whole-genome duplications. *Bioinformatics* **35**, 2153–2155 (2019).
47. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**, 357–359 (2012).
48. Shao, Z., Zhang, Y., Yuan, G.-C., Orkin, S. H. & Waxman, D. J. MAnorm: a robust model for quantitative comparison of ChIP-Seq data sets. *Genome Biol* **13**, R16 (2012).
49. *NCBI GenBank* https://identifiers.org/ncbi/insdc.gca:GCA_040438285.1 (2024).
50. *NCBI Sequence Read Archive* https://identifiers.org/ncbi/insdc.sra:SRX22497642 (2024).
51. *NCBI Sequence Read Archive* https://identifiers.org/ncbi/insdc.sra:SRX22497643 (2024).
52. *NCBI Sequence Read Archive* https://identifiers.org/ncbi/insdc.sra:SRX22497644 (2024).
53. *NCBI Sequence Read Archive* https://identifiers.org/ncbi/insdc.sra:SRP471413 (2024).
54. Shen, L. The genome assembly and annotation of Zoysia japonica. *figshare* https://doi.org/10.6084/m9.figshare.25610391 (2024).
55. Li, R., Li, Y., Kristiansen, K. & Wang, J. SOAP: short oligonucleotide alignment program. *Bioinformatics* **24**, 713–714 (2008).
56. Chikhi, R. & Medvedev, P. Informed and automated *k* -mer size selection for genome assembly. *Bioinformatics* **30**, 31–37 (2014).
57. Li, H. *et al*. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
58. Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V. & Zdobnov, E. M. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **31**, 3210–3212 (2015).
59. Parra, G., Bradnam, K. & Korf, I. CEGMA: a pipeline to accurately annotate core genes in eukaryotic genomes. *Bioinformatics* **23**, 1061–1067 (2007).
60. Rhie, A., Walenz, B. P., Koren, S. & Phillippy, A. M. Merqury: reference-free quality, completeness, and phasing assessment for genome assemblies. *Genome Biol* **21**, 245 (2020).

Acknowledgements

This work was supported by National Key Research and Development Program of China (2023YFD1200302), 5·5 Engineering Research & Innovation Team Project of Beijing Forestry University (BLRC2023A04), and Key R&D Program of Shandong Province, China (ZR202211070163). Bioinformatics analyses were performed at the High-Performance Computing Facility of Peking University Institute of Advanced Agricultural Sciences. We thank Beijing Biomarker Technologies Co, LTD. for their assistance with the PacBio sequencing data, Illumina sequencing data and Hi-C data.

Author contributions

L.S. and Z.Q. wrote the draft manuscript and performed the data analysis. Z.Q. assembled and annotated the genome. X.D. analyzed the data and drew pictures. Y.A. and J.C. collected and grew the plant material. L.X. and Y.C. improved the manuscript and offered the critical suggestion. L.H. and H.H. designed and coordinated the study. All authors approved the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41597-025-04827-x.

Correspondence and requests for materials should be addressed to L.H. or L.X.