

# A near-complete genome assembly of the allotetrapolyploid *Cenchrus fungigraminus* (JUJUNCAO) provides insights into its evolution and C<sub>4</sub> photosynthesis

Huakun Zheng<sup>1,5</sup>, Baiyu Wang<sup>2,3,5</sup>, Xiuting Hua<sup>2,5</sup>, Ruiting Gao<sup>2,5</sup>, Yuhao Wang<sup>3,5</sup>, Zixin Zhang<sup>1</sup>, Yixing Zhang<sup>3</sup>, Jing Mei<sup>3</sup>, Yongji Huang<sup>4</sup>, Yumin Huang<sup>3</sup>, Hui Lin<sup>1</sup>, Xingtang Zhang<sup>3</sup>, Dongmei Lin<sup>1</sup>, Siren Lan<sup>1</sup>, Zhongjian Liu<sup>1</sup>, Guodong Lu<sup>1</sup>, Zonghua Wang<sup>1,\*</sup>, Ray Ming<sup>3,\*</sup>, Jisen Zhang<sup>2,\*</sup> and Zhanxi Lin<sup>1,\*</sup>

<sup>1</sup>National Engineering Research Center of JUNCAO Technology, College of Life Science, Fujian Agriculture and Forestry University, Fuzhou 350002, China

<sup>2</sup>State Key Laboratory for Conservation and Utilization of Subtropical Agro-bioresources, Guangxi Key Laboratory of Sugarcane Biology, Guangxi University, Nanning 530004, Guangxi, China

<sup>3</sup>Center for Genomics, Fujian Provincial Key Laboratory of Haixia Applied Plant Systems Biology, Key Laboratory of Genetics, Breeding and Multiple Utilization of Crops, Ministry of Education, Fujian Agriculture and Forestry University, Fuzhou 350002, China

<sup>4</sup>Fuzhou Institute of Oceanography, Minjiang University, Fuzhou 350108, China

<sup>5</sup>These authors contributed equally to this article.

\*Correspondence: Zonghua Wang ([zonghuaw@163.com](mailto:zonghuaw@163.com)), Ray Ming ([rayming@illinois.edu](mailto:rayming@illinois.edu)), Jisen Zhang ([zjisen@126.com](mailto:zjisen@126.com)), Zhanxi Lin ([lzxjuncao@163.com](mailto:lzxjuncao@163.com))

<https://doi.org/10.1016/j.xplc.2023.100633>

## ABSTRACT

JUJUNCAO (*Cenchrus fungigraminus*;  $2n = 4x = 28$ ) is a *Cenchrus* grass with the highest biomass production among cultivated plants, and it can be used for mushroom cultivation, animal feed, and biofuel production. Here, we report a nearly complete genome assembly of JUJUNCAO and reveal that JUJUNCAO is an allopolyploid that originated  $\sim 2.7$  million years ago (mya). Its genome consists of two subgenomes, and subgenome A shares high collinear synteny with pearl millet. We also investigated the genome evolution of JUJUNCAO and suggest that the ancestral karyotype of *Cenchrus* split into the A and B ancestral karyotypes of JUJUNCAO. Comparative transcriptome and DNA methylome analyses revealed functional divergence of homeologous gene pairs between the two subgenomes, which was a further indication of asymmetric DNA methylation. The three types of centromeric repeat in the JUJUNCAO genome (*CEN137*, *CEN148*, and *CEN156*) may have evolved independently within each subgenome, with some introgressions of *CEN156* from the B to the A subgenome. We investigated the photosynthetic characteristics of JUJUNCAO, revealing its typical C<sub>4</sub> Kranz anatomy and high photosynthetic efficiency. NADP-ME and PEPCK appear to cooperate in the major C<sub>4</sub> decarboxylation reaction of JUJUNCAO, which is different from other C<sub>4</sub> photosynthetic subtypes and may contribute to its high photosynthetic efficiency and biomass yield. Taken together, our results provide insights into the highly efficient photosynthetic mechanism of JUJUNCAO and provide a valuable reference genome for future genetic and evolutionary studies, as well as genetic improvement of *Cenchrus* grasses.

**Key words:** genome assembly, allotetrapolyploid, centromere architecture, evolutionary trajectory, photosynthetic efficiency, *Cenchrus* grass

Zheng H., Wang B., Hua X., Gao R., Wang Y., Zhang Z., Zhang Y., Mei J., Huang Y., Huang Y., Lin H., Zhang X., Lin D., Lan S., Liu Z., Lu G., Wang Z., Ming R., Zhang J., and Lin Z. (2023). A near-complete genome assembly of the allotetrapolyploid *Cenchrus fungigraminus* (JUJUNCAO) provides insights into its evolution and C<sub>4</sub> photosynthesis. *Plant Comm.* 4, 100633.

Published by the Plant Communications Shanghai Editorial Office in association with Cell Press, an imprint of Elsevier Inc., on behalf of CSPB and CEMPS, CAS.

## INTRODUCTION

*Cenchrus* is a widespread genus of the Poaceae family. *Cenchrus* grasses are C<sub>4</sub> plants characterized by high photosynthetic efficiency, massive biomass production, and high stress tolerance. Most species are important for animal forage, cultivation material for edible and medicinal mushrooms (so called JUNCAO 菌草), and biofuel production. JUJUNCAO (*Cenchrus fungigraminus*; named JUJUNCAO for giant JUNCAO, 巨菌草; [Supplemental Figure 1](#)) is the most productive *Cenchrus* species, with an average height of ~4.0 m and a green-matter yield of ~400 ton/ha annually ([Lin, 2013](#); [Lin et al., 2022](#)), which is about twice that of the two elephant grasses *Cenchrus purpureus* ([Yan et al., 2020](#)) and *Pennisetum purpureum* ([Zhang et al., 2022b](#)) and other high-biomass C<sub>4</sub> grasses ([Mullet, 2017](#)). JUJUNCAO is able to grow in arid and semi-arid regions as well as ecologically fragile areas ([Lin, 2013](#); [Xu et al., 2014](#); [He et al., 2017](#); [Hayat et al., 2020](#); [Jia et al., 2020](#)). Cultivation of *Cenchrus* grasses such as JUJUNCAO is one of the most efficient strategies for using marginal land that is not optimal for traditional crops and satisfying increasing demands for food, including meat, milk, and mushrooms ([Samson et al., 2005](#); [Lin, 2013](#); [Eisler et al., 2014](#); [Gu et al., 2019](#); [Xie and Xu, 2019](#); [Kuang et al., 2022](#)).

JUJUNCAO is morphologically similar to another *Cenchrus* grass cultivar (*P. purpureum* × *Pennisetum americanum* cv. Reyan No. 4), which is a triploid offspring of elephant grass ([Luo et al., 2016](#)). JUJUNCAO can be distinguished from other *Cenchrus* grasses by its stem, leaf, flower, or color phenotypes ([Lin et al., 2022](#)). These other species include pearl millet (*Cenchrus americanus*; syn. *Pennisetum glaucum*) ([Varshney et al., 2017](#)), *C. purpureus* ([Yan et al., 2020](#)) and *P. purpureum* ([Zhang et al., 2022b](#)), three grasses with assembled genome sequences. Although many efforts have been made to decipher the molecular biology of JUJUNCAO ([Lin et al., 2015](#); [Ye et al., 2015](#); [Zhu et al., 2015](#); [Chen et al., 2016](#); [Li et al., 2020](#); [Zhou et al., 2021](#)), little is known about the molecular basis of its agronomic traits and evolutionary trajectory, owing to the lack of a high-quality reference genome.

Polyploidization is an important evolutionary force in angiosperm plants and provides diverse genetic resources to benefit the domestication of crops ([Adams and Wendel, 2005](#); [Soltis et al., 2009](#)). Polyploidization may also result in high sequence similarity between homologous/homeologous chromosomes and thus in frequent rearrangements between these chromosomes ([Wendel et al., 2016](#)), a feature that poses major challenges in genomic biology, such as deciphering the architecture of telomeres and centromeres. Centromeres are essential for genome integrity because they mediate the junction between sister chromatids and the proper separation of chromosomes during mitosis and meiosis ([Liu et al., 2020b](#)). Plant centromeres often consist of satellite arrays interrupted by long terminal repeat (LTR) retrotransposons ([Zhong et al., 2002](#); [Hall et al., 2004](#); [Comai et al., 2017](#)), making it challenging to understand centromere structure. Assembly of a highly contiguous genome is therefore essential for resolving centromere structure and mechanisms of chromosome fusion and fission in the allotetraploid JUJUNCAO.

Recent advances in DNA sequencing technology and assembly methodology have enabled gapless, telomere-to-telomere (T2T) chromosome/genome assemblies for human cells ([Nurk et al., 2022](#)), barley ([Navrátilová et al., 2022](#)), banana ([Belser et al., 2021](#)), and maize ([Liu et al., 2020a](#)), which enable in-depth investigation of the structure and evolutionary trajectory of chromosomes, especially in regions with complex repetitive sequences such as telomeres and centromeres. Here, we generated a T2T gapless genome of the allotetraploid JUJUNCAO, performed bisulfite whole-genome sequencing, and analyzed its genome evolution and the genomic basis of its C<sub>4</sub> characteristics. These results provide insights into the evolution of JUJUNCAO, particularly its centromere architecture, and the molecular basis of its high photosynthetic efficiency.

## RESULTS

### Assembly of a T2T genome of JUJUNCAO

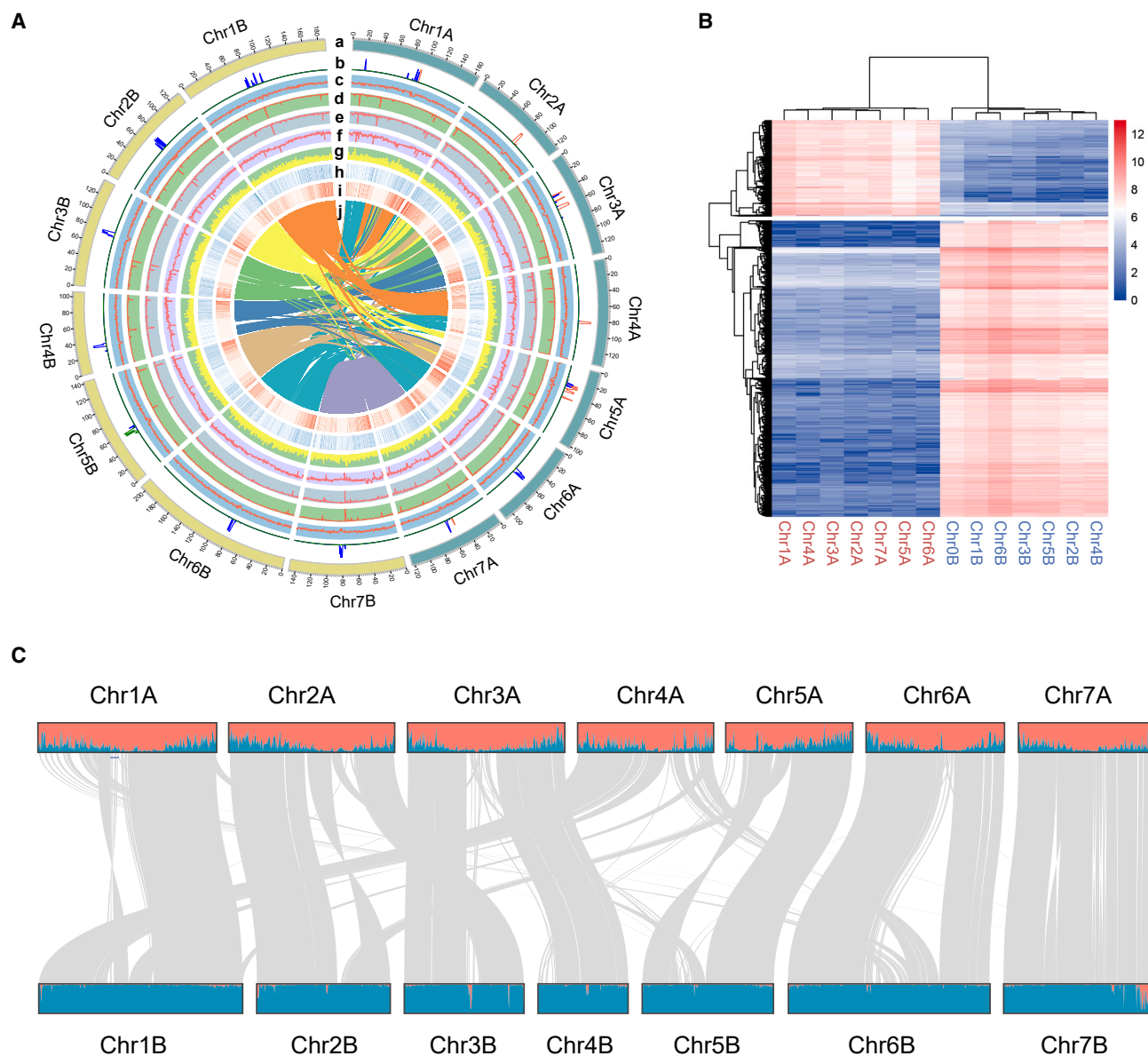
To decipher the genome structure of JUJUNCAO, we first estimated its genome size to be 2.05–2.09 Gb ([Supplemental Table 1](#)) using flow cytometry analysis ([Dolezel and Bartos, 2005](#)). A total of 120 Gb (~60× coverage of the estimated genome size) of PacBio long high-fidelity (HiFi) circular consensus sequencing (CCS) data were generated and assembled into 2.13 Gb of contigs using the Hifiasm assembler ([Cheng et al., 2021](#)). A final total of 116 contigs were obtained after removal of 145 Mb of heterozygous sequences, resulting in a total assembly size of 1.99 Gb, with 1.97 Gb (~99.0%) anchored into 14 pseudo-chromosomes based on Hi-C data ([Figure 1](#); [Supplemental Figure 2](#); [Table 1](#) and [Supplemental Table 2](#)). The assembled JUJUNCAO genome contains 24 annotated telomeres and 13 T2T chromosomes, eight of which are gap-free (Chr2B, 3A, 3B, 4B, 5A, 5B, 6A, and 7A), four of which have only one gap (Chr1B, 2A, 6B, and 7B), and one of which (Chr4A) has three gaps at the telomeric region of the short arm ([Supplemental Figure 2C](#)).

We used benchmarking universal single-copy orthologs (BUSCO; [Simao et al., 2015](#)) and the LTR retrotransposon assembly index (LAI; [Ou et al., 2018](#)) to assess the completeness and contiguity of the assembly. The total complete BUSCO score was 98.4%, and the LAI of the JUJUNCAO assembly was 15.9 ([Table 1](#) and [Supplemental Tables 3](#) and [4](#)).

### Genome annotation

We predicted 68 562 gene models using *ab initio* gene prediction and homology-based methods. Subgenomes A and B harbor 34 630 and 33 779 genes, respectively. BUSCO (v5.4.1) analysis was performed to evaluate the gene annotations ([Simao et al., 2015](#)), and 93.0% of the conserved genes from the BUSCO database were present in the JUJUNCAO genome ([Table 1](#) and [Supplemental Tables 4](#) and [5](#)).

A total of 1.16 Gb of repeat sequences were found in the JUJUNCAO genome, accounting for 68.87% of the assembly. The total proportion of transposable elements (TEs) in JUJUNCAO was comparable to that in *Sorghum bicolor*, higher than that in *Setaria italica* and *Panicum miliaceum*, but lower than that in *Zea mays* ([Supplemental Figure 3](#); [Supplemental Table 6](#)). We also investigated the LTR insertion times and



**Figure 1. Allotetraploid genome of JUJUNCAO (*C. fungigraminus*).**

**(A)** Circos plot of the JUJUNCAO genome assembly. Quantitative tracks (b to j) are aggregated in 500-kbp bins, and independent y axis labels are as follows: (a) chromosomes; (b) centromeric repeat sequences; (c) GC content; (d) CG methylation; (e) CHG methylation; (f) CHH methylation; (g) LTRs; (h) SNPs; (i) gene expression; (j) gene density; (k) syntenic relationships.

**(B)** Clustering of counts of 13-mers that differentiate homeologous chromosomes enabled consistent partitioning of the genome into two subgenomes. Red and blue chromosome names correspond to the A and B subgenomes, respectively.

**(C)** Distribution of subgenome-specific 13-mer sequences (red for subgenome A and blue for subgenome B) for 14 chromosomes and synteny between the two subgenomes.

found a smaller wave of LTR bursts in JUJUNCAO at  $\sim 1.2$  mya, similar to that in *Z. mays*. By contrast, the other species investigated, such as *C. americanus*, *P. miliaceum*, *S. bicolor*, and *S. italica*, had a stronger LTR burst within 1 mya (Supplemental Figure 3B). LTR retrotransposons were the most abundant TEs in the JUJUNCAO genome, accounting for 35.69% of the genome, followed by unknown repeats (28.67%), and DNA transposons occupied only 4.51% of the genome. The Ty3/Gypsy superfamily was the most abundant type of LTR retrotransposon, making up more than half of the retrotranspo-

sons and 22.40% of the JUJUNCAO genome (Supplemental Figure 3C; Supplemental Table 6).

### Disentangling the subgenomes

To further confirm the chromosome phasing results, we identified 32 629 chromosome-specific 13-bp sequences (13-mers), which clustered into two distinct groups, indicative of the two subgenomes in JUJUNCAO (Figure 1B and 1C). These results are consistent with the allopolyploidy features revealed by

Assembly	Initial assembly	Purge haplotigs	Hi-C
Total assembly size of contigs (bp)	2 131 895 497	1 986 757 443	–
Number of contigs	815	116	–
Contig N50 (bp)	131 346 082	134 074 806	–
Contig N90 (bp)	94 754 406	99 768 082	–
Longest contig (bp)	168 262 481	162 376 048	–
GC content (%)	47.11	47.11	47.11
BUSCO completeness of assembly (%)	–	98.40	–
LAI <sup>a</sup>	–	15.90	–
Length of chromosomes (bp)	–	–	1 966 989 277
Number of chromosomes	–	–	14
Number of gaps	–	–	13
Anchored rate (%)	–	–	99.00
Annotation	–	–	–
Percentage of repeat elements	73.38	–	–
Total number of genes	68 526	–	–
BUSCO completeness of annotation (%)	93.00	–	–
Subgenome	A subgenome	B subgenome	–
Length of chromosomes (bp)	899 693 551	1 067 295 726	–
Number of chromosomes	7	7	–
Number of genes	34 630	33 779	–

**Table 1. Genome assembly and annotation statistics of JUJUNCAO**

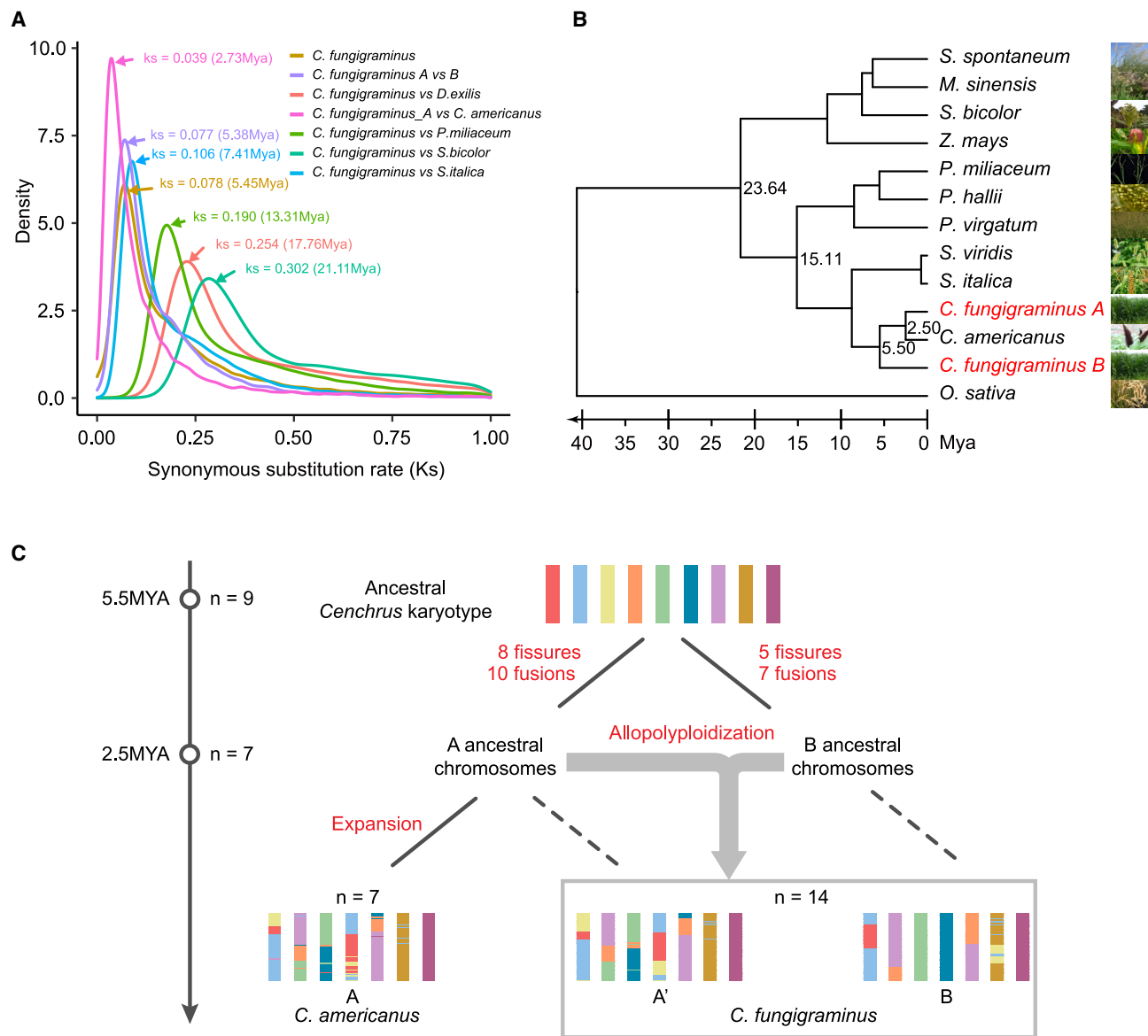
<sup>a</sup>LAI was used to assess genome assembly quality.

our karyotyping results (Supplemental Figure 2A; Zhu et al., 2015). The average nucleotide identity of homologous regions in the two subgenomes was 90.1%, and chromosome rearrangements occurred frequently after the split of the common ancestor of the two subgenomes (Figure 1C). Except for Chr07A/Chr07B, in which large segment arrangements were observed in the other six chromosome pairs, subgenome A of JUJUNCAO exhibited highly conserved collinear synteny with the genome of pearl millet, with 90.7% average nucleotide identity. By contrast, subgenome B of JUJUNCAO and the genome of pearl millet showed more genome rearrangements and had 86.4% average nucleotide identity (Supplemental Figure 4).

### Genome evolution

The near-T2T complete assembly enabled us to investigate the evolution of the JUJUNCAO genome. To establish the timeline of paleo-tetraploidy in JUJUNCAO, we performed inter- and intra-subgenomic comparisons between JUJUNCAO and other Panicoideae species using protein sequences of the primary transcripts from 10 representative Panicoideae species (*Saccharum spontaneum*, *Miscanthus sinensis*, *S. bicolor*, *Z. mays*, *Panicum hallii*, *P. miliaceum*, *Panicum virgatum*, *C. americanus*, *S. italica*, and *Setaria viridis*) and the outgroup species *Oryza sativa* (Figure 2A). Our results showed that JUJUNCAO diverged from *S. italica*, *P. miliaceum*, and *S. bicolor* approximately 7.4 mya, 13.3 mya, and 21.1 mya, respectively (Figure 2A). The allotetraploid origin of JUJUNCAO was estimated at ~2.7 mya based on the Ks value of 0.039.

We then traced the evolutionary trajectory of JUJUNCAO and representative species from different genera of Panicoideae, including *Z. mays*, *S. bicolor*, *Echinochloa haploclada*, *Digitaria exilis*, *P. miliaceum*, *P. hallii*, and *S. italica* (Supplemental Figure 5). According to the number and structure of chromosomes in these species, we inferred that the Paniceae ancestor originated at ~15.1 mya and had nine chromosomes. Most of the Paniceae species evolved from the ancestral grass karyotype (AGK) through 10–18 chromosome rearrangement events, including fusions and fissions. For example, *E. haploclada* evolved with the fewest chromosomal rearrangement events (four fusions and six fissions) and most closely resembled the AGK. Ancestors of JUJUNCAO subgenomes A and B are estimated to have diverged ~5.5 mya from *Cenchrus* ancestors, following the divergence of *Setaria* and *Cenchrus* at ~8.1 mya (Figure 2B and Supplemental Figure 5). Our results showed that each of the two JUJUNCAO subgenomes contained one chromosome (Chr6A and Chr4B) that resembled the AGK, suggesting that the two subgenomes of JUJUNCAO may have evolved independently from the same ancestral *Cenchrus* karyotype. Based on the chromosome synteny of Paniceae species, we propose the ancestral *Cenchrus* karyotype (ACK) (Figure 2C; Supplemental Figure 6). Compared with the ACK, 18 large rearrangements (eight fissions and 10 fusions) were found in subgenome A and 12 large rearrangements (five fissions and seven fusions) in subgenome B. We also found that JUJUNCAO and *P. purpureum* diverged ~0.3 mya and displayed similar karyotypes compared with the AGK, but the sequence similarity between homologous genes



**Figure 2. Phylogenetic relationships between JUJUNCAO and species of Panicoideae.**

**(A)** Distributions of synonymous substitution rate (Ks) between *C. fungigraminus* and seven representative species of Panicoideae. The Ks peaks for pairwise comparisons are indicated by arrows.

**(B)** Phylogenetic tree of 11 species and the A/B subgenomes of *C. fungigraminus*.

**(C)** Chromosome evolution of JUJUNCAO and *C. americana* from ancestral chromosomes. The nine chromosomes of the ancestral *Cenchrus* karyotype (ACK) were reconstructed from syntenic blocks of related species, and the origin of the two JUJUNCAO subgenomes was simulated. Nine colors indicate the composition of the ACK from two subgenomes with distinct origins. These chromosomes are for illustrative purposes of fissure and fusion only and do not represent the actual dimensions.

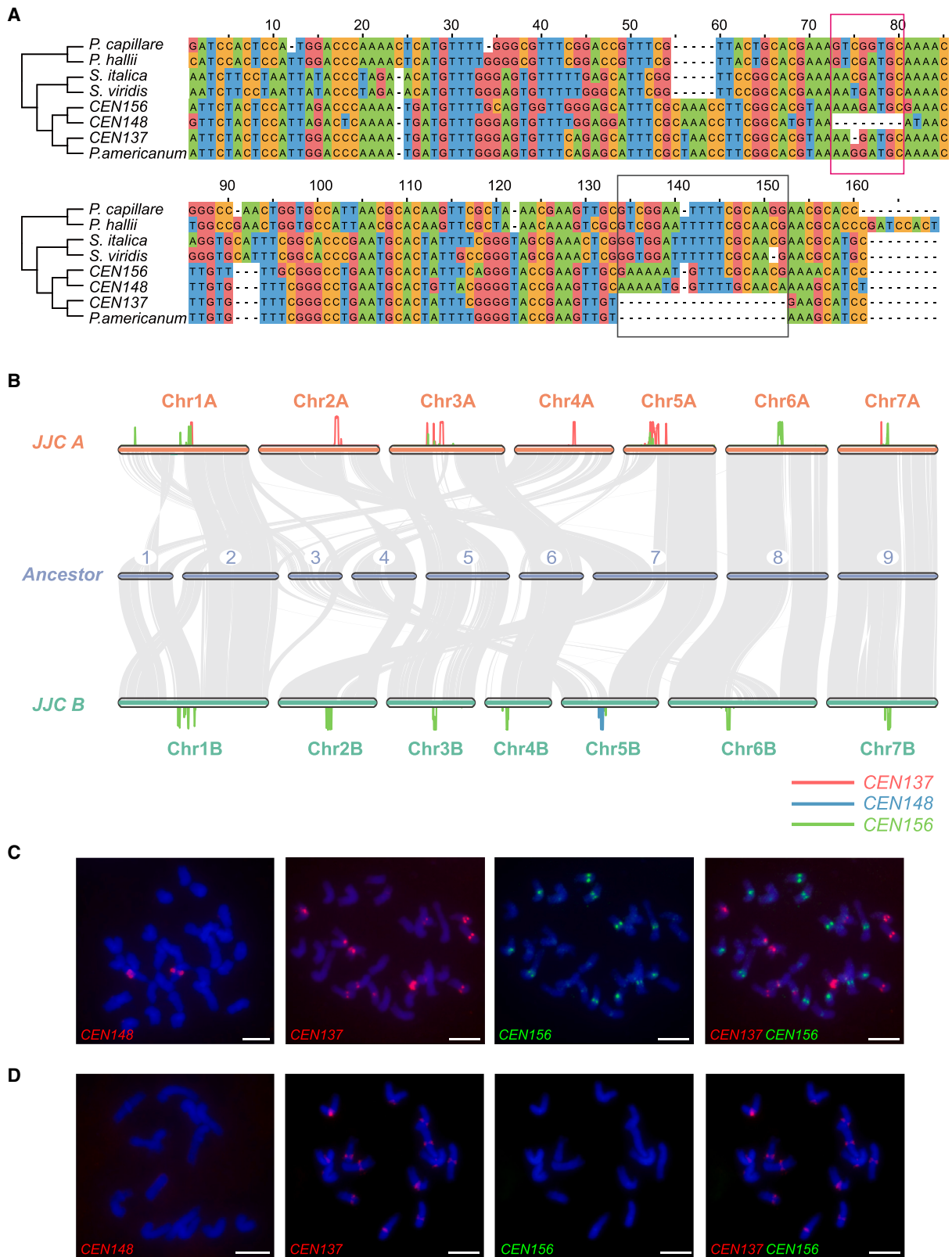
differed among different chromosome pairs (Supplemental Figure 7).

### Centromere architecture in JUJUNCAO

Achieving high contiguity of centromeres is one of the major challenges in T2T genome assembly. Here, the complete genome enabled us to annotate the centromeres of all 14 chromosomes. We identified three types of centromeric repeats with high sequence similarity in the JUJUNCAO genome, *CEN137*, *CEN148*, and *CEN156*, each composed of satellite repeats with

lengths of ~137 bp, ~148 bp, and ~156 bp, respectively (Supplemental Table 7). These satellite repeats shared high sequence similarity and were relatively conserved among Paniceae species (identity: 59%–98%) such as *S. viridis*, *S. italyca*, *P. hallii*, and *Panicum capillare* (Figure 3A). Among them, *CEN156* showed the highest sequence similarity to centromeric monomers from other species, whereas *CEN137* and *CEN148* showed 19-bp and 8-bp deletions, respectively (Figure 3A).

We used a k-mer method to accurately locate the three similar satellite repeats and found that they were distributed in distinct



**Figure 3. Characteristics of centromere sequences in JUJUNCAO.**

(A) Multiple sequence alignment of the three centromeric satellite sequences from JUJUNCAO and their homologous sequences from *S. viridis*, *S. italica*, *P. hallii*, *P. capillare*, and *P. americanum*. Green, red, orange, and blue represent A, G, C, and T bases, respectively. The red and blue boxes represent variant loci within the three centromeric satellites (*CEN137*, *CEN156*, and *CEN148*).

(legend continued on next page)

patterns. *CEN156* is the most abundant centromeric array in the JUJUNCAO genome and is enriched on almost all chromosomes, with the exception of Chr2A and Chr4A. *CEN137* is enriched mainly on chromosomes from subgenome A, and *CEN148* is found specifically on Chr5B (Figure 3B). We performed fluorescence *in situ* hybridization (FISH) to confirm these results in JUJUNCAO and the diploid *P. americanum* cultivar MZL using probes generated from repeat-specific oligomers of the three centromeric repeats. There were two strong signals representing the Chr5B-specific localization of *CEN148* in JUJUNCAO, and these were absent in MZL. *CEN137* signals were present on 12 of the chromosomes of JUJUNCAO and MZL, with the strongest signal on Chr5A. For *CEN156*, there were 11 signals representing six of the chromosomes of JUJUNCAO, and these were absent in MZL. There were also weak signals for *CEN156* in the JUJUNCAO genome, some of which co-localized with that of *CEN137* (Figure 3C). Taken together, our results revealed high sequence similarity and a distinct distribution pattern of the three centromeric satellites identified in the JUJUNCAO genome. The similarity in centromeric structure between the JUJUNCAO A subgenome and the diploid *P. americanum* MZL genome suggest that they are derived from common A ancestral chromosomes (Figure 2C).

### Expression bias of homeologous gene pairs

Polyploidization may cause expression bias between homeologous gene pairs (Yang et al., 2016; Liang and Schnable, 2018). To investigate homeolog expression bias in allotetraploid JUJUNCAO, we defined 22 260 homeologous gene pairs between the A and B subgenomes. Comparative transcriptome analysis using RNA sequencing (RNA-seq) data generated from leaves, stems, and roots showed that A-subgenome genes are expressed at a slightly lower level than those of the B subgenome, but there is no major overall bias for one subgenome (Figure 4A and Supplemental Figure 8). We identified 5668 and 5708 expression-biased genes in subgenomes A and B, respectively (Figure 4B). Subgenome B-biased homeologs were enriched in Chr1B, the subtelomeric region of the long arm of Chr4B, and the subtelomeric region of the short arms of Chr5B and Chr6B. Subgenome A-biased homeologs were enriched in Chr1A, the centromere proximal region, and the short arm of Chr7A (Supplemental Figure 9). In total, 1124 differentially expressed homeologous gene pairs (DEHGs) were identified: 557 showed subgenome-A bias and 567 showed subgenome-B bias. Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) functional classification showed that DEHGs from the two subgenomes were enriched in different metabolism, biosynthesis, and signaling pathways in roots, stems, and leaves. Carboxylic acid biosynthesis, RNA metabolic process, and mitogen-activated protein kinase (MAPK)-signaling pathways were specifically enriched in subgenome A-biased DEHGs, whereas hydrogen peroxide metabolic process, response to oxidative stress, and AMPK-signaling pathways were enriched in subgenome B-biased DEHGs (Supplemental

Figures 10 and 11). These results revealed functional divergence between the two subgenomes of JUJUNCAO.

### DNA methylation of homeologous gene pairs

Allele-specific DNA methylation is one of the mechanisms associated with expression bias of homeologs in allopolyploid plants (Edger et al., 2017; Bird et al., 2021). To investigate whether the expression bias of homeologous gene pairs in JUJUNCAO was determined by DNA methylation, we performed high-throughput bisulfite sequencing using libraries generated from leaves, stems, and roots, with a genome coverage of about 30× and a cytosine depth ranging from 15.1× to 19.1× in the three tissues (Supplemental Table 8). On average, about 67% of the reads generated from JUJUNCAO tissues could be mapped to the assembled genome, and the de-repetitive retention rate was ~75% (Supplemental Table 9). The percentage methylation of CG, CHG, and CHH contexts was highest in roots (73.3%, 53%, and 2.4%, respectively), followed by stems (63.5%, 47.2%, and 2.3%) and leaves (63.0%, 44.3%, and 2.1%; Supplemental Table 10). Cytosine methylation was mainly enriched in regions around centromeres, although the methylation patterns of centromeres differed from each other. These results are consistent with the hypermethylation of repetitive sequences (Supplemental Figure 12).

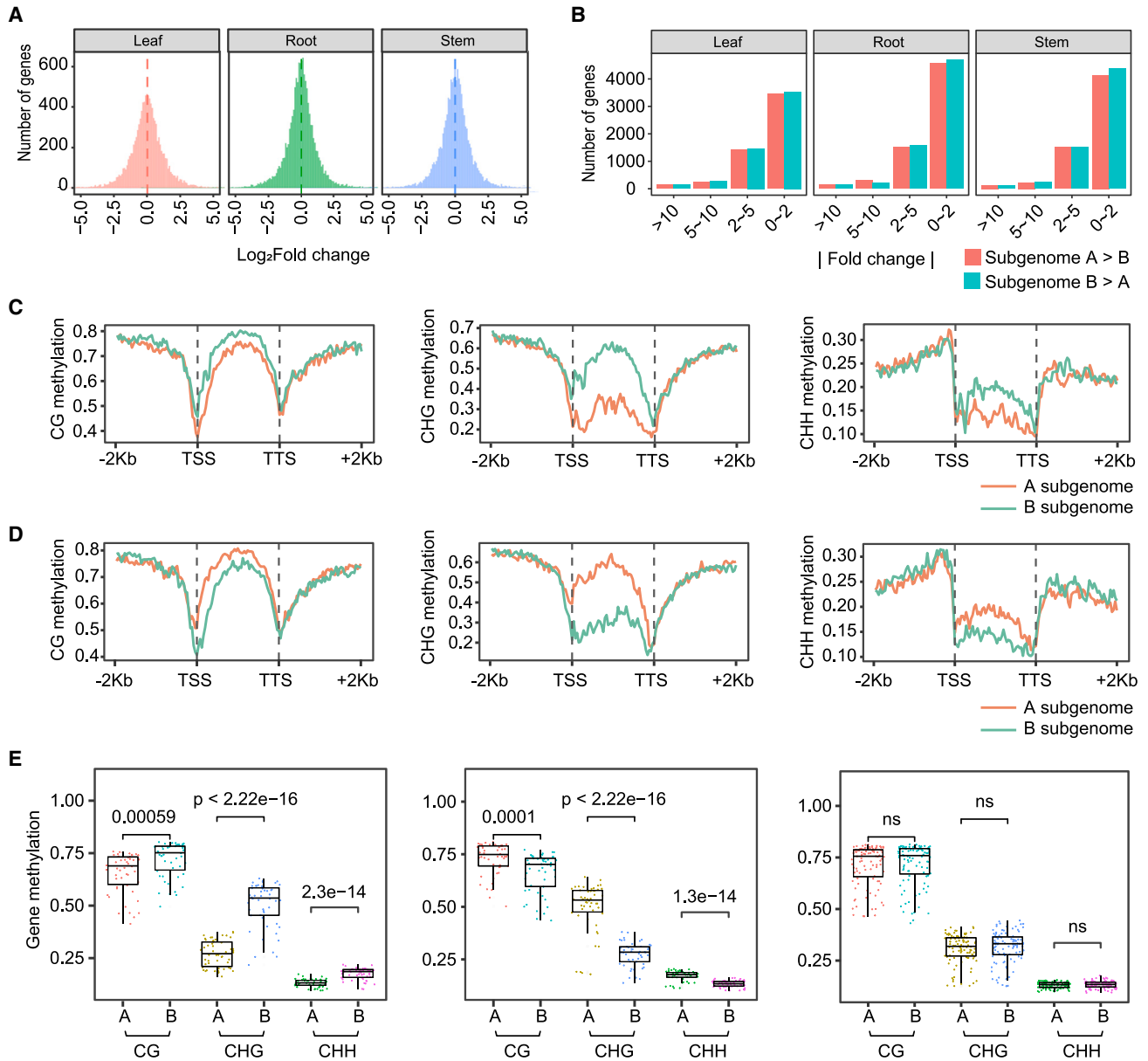
We next compared overall DNA methylation patterns between the two subgenomes in the examined tissues. In upstream and downstream regions, methylation of all CG, CHG, and CHH contexts was significantly higher in subgenome B than in subgenome A for leaves ( $p < 1.9e-05$ ) and stems ( $p < 0.037$ ) (Supplemental Figure 13). In the gene body, only CHH methylation was significantly higher in subgenome B than in subgenome A for all tissues ( $p < 8.5e-05$ ). To further explore the association between DNA methylation and gene expression, we compared DNA methylation patterns of subgenome-biased genes and found that CG and CHG methylation rates in the gene body were significantly lower in subgenome-biased genes (whose expression was more than twice that of their homeologs) than in their subgenome homeologs. By contrast, CHH methylation showed the opposite pattern in the three tissues. We found no significant differences in methylation of gene pairs that showed a lower level of differential expression (0–2 fold change), with the exception of CHH methylation in leaves ( $p = 0.0084$ ) (Figure 4C–4E, Supplemental Figures 14 and 15).

### C<sub>4</sub> photosynthetic characteristics of JUJUNCAO

To explore the C<sub>4</sub> photosynthesis features of JUJUNCAO, we first performed a histological analysis and demonstrated that JUJUNCAO has typical Kranz anatomy, with two mesophyll (M) cells separated by consecutive enlarged bundle sheaths (BS) (BS-M-M-BS) (Figure 5A), similar to other C<sub>4</sub> plants such as maize and sugarcane (Langdale, 2011; Sage et al., 2014). We next investigated the photosynthetic response of JUJUNCAO to light and CO<sub>2</sub> concentration. At 400 μmol mol<sup>-1</sup> CO<sub>2</sub>, JUJUNCAO

**(B)** Syntenic blocks between the A and B subgenomes of JUJUNCAO (JJC A and B) and the expected Paniceae ancestral genome (ACK). Red, blue, and green lines show the distributions of different centromeric sequences in the JUJUNCAO genome.

**(C and D)** FISH assays of JUJUNCAO **(C)** and the *C. americanum* cultivar MZL **(D)** using repeat-specific probes generated from the three centromeric satellites (*CEN137*, *CEN148*, and *CEN156*). Scale bars, 10 μm.



**Figure 4. Expression and DNA methylation divergence in the allotetraploid genome of JUJUNCAO.**

(A) The overall distribution of the  $\log_2$  (fold change) of expression between homeologous gene pairs in leaves, stems, and roots. (B) The number of homeologous gene pairs between the A and B subgenomes with different degrees of expression bias. The expression bias (absolute value of fold change) was divided into four bins: 0–2, 2–5, 5–10, >10. (C) The methylation profiles of homeologous gene pairs with subgenome A-biased expression in the stem (fold change >2). (D) The methylation profiles of homeologous gene pairs with subgenome B-biased expression in the stem (fold change >2). (E) Comparison of three types of methylation level in the gene body of A and B subgenomes in the stem. From left to right panel: the methylation levels of subgenome A-biased homeologous gene pairs (fold change >2), subgenome B-biased homeologous gene pairs (fold change >2), and differentially expressed gene pairs (with 0–2 fold change). All pairwise comparisons were performed by t test.

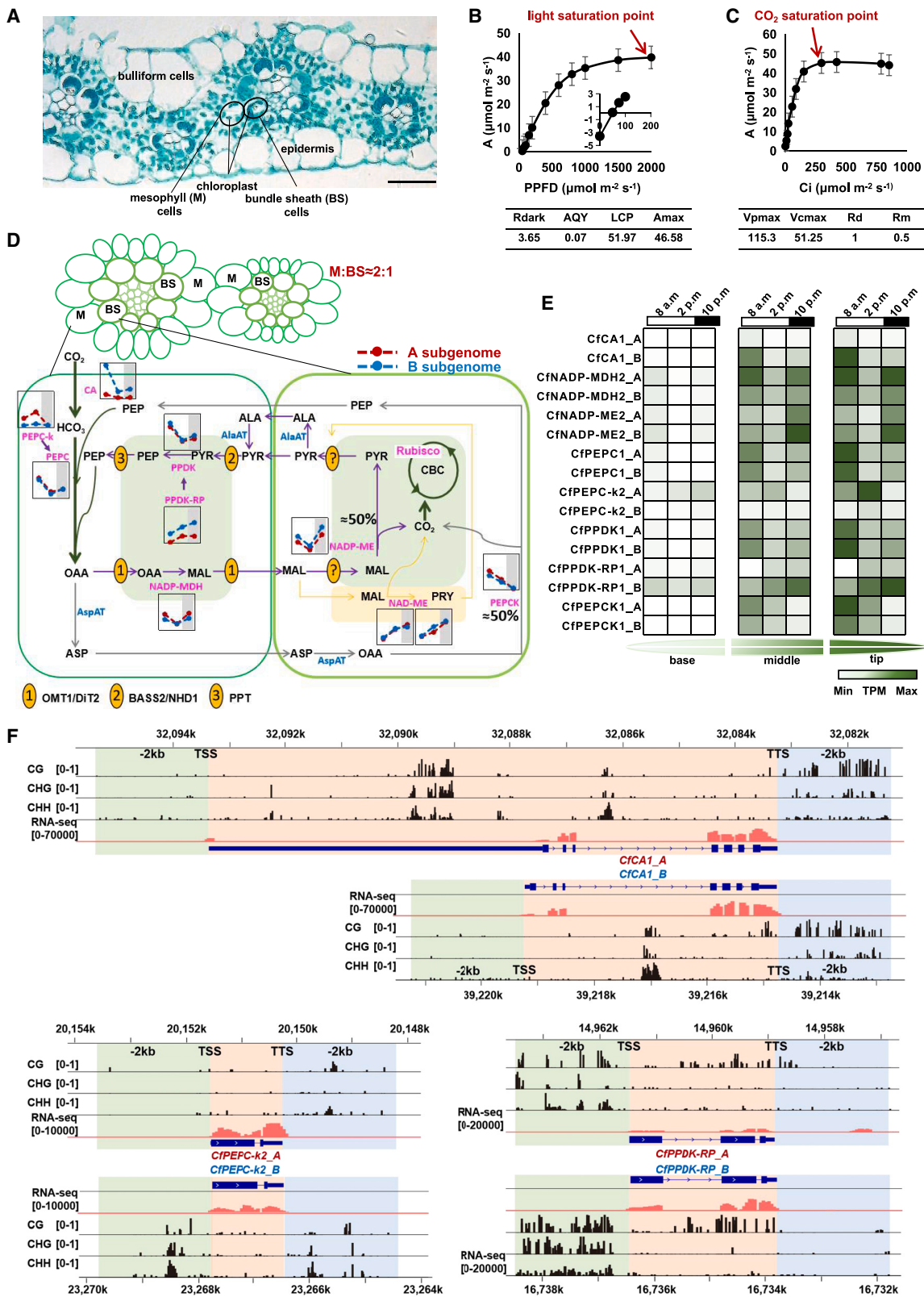
had a light saturation point (LSP) near 2000  $\mu\text{mol mol}^{-1}$  (Figure 5B) and a light-saturated net  $\text{CO}_2$  assimilation rate (A) of 40  $\mu\text{mol m}^{-2} \text{s}^{-1}$ . The apparent quantum yield (AQY) was 0.07, the light compensation point (LCP) was 51.97  $\mu\text{mol photons m}^{-2} \text{s}^{-1}$ , and the  $A_{\text{max}}$  was 46.58  $\mu\text{mol m}^{-2} \text{s}^{-1}$  (Figure 5B). Investigation of the primary method of carbon fixation in JUJUNCAO showed that at a light intensity of 2000  $\mu\text{mol photons m}^{-2} \text{s}^{-1}$ , the  $\text{CO}_2$  response curve with a low  $\text{CO}_2$  compensation point (154  $\mu\text{mol mol}^{-1}$ ) presents a steep slope typical of  $\text{C}_4$  spe-

cies (Figure 5C) (Sage et al., 2014). After fitting, the maximal PEP carboxylation activity ( $V_{\text{pmax}}$ ) was 115.30  $\mu\text{mol m}^{-2} \text{s}^{-1}$ , and the maximal Rubisco activity ( $V_{\text{cmax}}$ ) was 51.25  $\mu\text{mol m}^{-2} \text{s}^{-1}$  (Figure 5C). These results confirmed that JUJUNCAO is a  $\text{C}_4$  grass.

**Genes related to the  $\text{C}_4$  photosynthetic pathway of JUJUNCAO**

We identified  $\text{C}_4$  pathway genes in JUJUNCAO based on their  $\text{C}_3$  and  $\text{C}_4$  orthologs from sugarcane, maize, sorghum, and rice





**Figure 5.  $C_4$  pathway in JUJUNCAO.**

(A) Cross-sections of JUJUNCAO leaves. Horizontal bars indicate 20  $\mu\text{m}$ .

(B) The light response of net  $\text{CO}_2$  assimilation rate ( $n = 5$ ) (mean  $\pm$  SD). PPFD, photosynthetic photon flux density; A, net photosynthetic rate; Rdark, dark respiration; AQY, apparent quantum yield; LCP, light compensation point; Amax, maximum  $\text{CO}_2$  net assimilation rate.

(legend continued on next page)

(Supplemental Figure 16; Supplemental Table 11). The JUJUNCAO genome contained 44 putative genes involved in the C<sub>4</sub> carbon fixation module, including those encoding the key enzymes carbonic anhydrase (CA), NAD-malic enzyme (NAD-ME), NADP-malate dehydrogenase (NADP-MDH), NADP-malic enzyme (NADP-ME), phosphoenolpyruvate carboxylase (PEPC), phosphoenolpyruvate carboxykinase (PEPCK), PEPC kinase (PEPC-k), pyruvate orthophosphate dikinase (PPDK), and PPDK regulatory protein (PPDK-RP). These genes were evenly distributed in the two subgenomes (Supplemental Table 11). To further investigate the photosynthetic and diel expression patterns of C<sub>4</sub> pathway genes, we collected RNA-seq samples at 08:00 h, 14:00 h, and 22:00 h from the immature zone (base), transition-zone (middle), and mature zone (tip) of field-grown JUJUNCAO leaves (Figure 5D and 5E). In the gradient leaf tissues, we were able to distinguish gene family members involved in carbon fixation from non-C<sub>4</sub>-related genes involved in other processes. Eight of the putative C<sub>4</sub> pathway genes, namely *CfCA1*, *CfPEPC1*, *CfPEPC-k2*, *CfNADP-MDH2*, *CfNADP-ME2*, *CfPEPCK*, *CfPPDK1*, and *CfPPDK-RP1*, displayed the expected expression pattern of genes involved in photosynthesis: their expression increased from the leaf base to the tip (Li et al., 2010), and they had a diurnal expression pattern in the mature zone, with little or no expression in the immature zone (Figure 5D and 5E). In addition, expression of these genes was higher in leaves than in stems and roots (Supplemental Figure 17). These results suggested that these eight genes are related to C<sub>4</sub> photosynthesis in JUJUNCAO. Notably, *CfPEPC-k2* displayed a C<sub>4</sub> expression pattern in JUJUNCAO leaves but was clustered with non-C<sub>4</sub> genes of sugarcane, maize, and sorghum (Figure 5D and 5E; Supplemental Figure 17). In the daytime, expression of *CfPEPC-k2* and *CfPPDK-RP1* increased from 08:00 h to 14:00 h (Figure 5E), a pattern opposite to that of the other C<sub>4</sub> genes. These results indicated that a neofunctionalization of these two genes occurred in JUJUNCAO after its divergence from other Poaceae species.

The genes *CfPEPC-k2*, *CfPPDK-RP1*, and *CfCA1* in the A and B subgenomes were differentially expressed at different time points (Figure 5D). *CfCA1* and *CfPPDK-RP1* in subgenome B (*CfCA1\_B* and *CfPPDK-RP1\_B*) displayed enhanced expression (>2.75-fold) in leaves compared with those of subgenome A (*CfCA1\_A* and *CfPPDK-RP1\_A*), and expression of *CfPEPC-k2* was 2.88-fold higher in subgenome A (*CfPEPC-k2\_A*) than in subgenome B (*CfPEPC-k2\_B*) (Supplemental Figure 17). To investigate the

potential influence of gene methylation on gene expression, we examined differences in CG, CHG, and CHH methylation in the 2-kb upstream region, 5' untranslated region (UTR), exon region, intron region, 3' UTR, and 2-kb downstream region of these three genes. *CfPEPC-k2*, *CfPPDK-RP1*, and *CfCA1* in subgenomes A and B had different methylation levels in the three methylation contexts (Figure 5F). *CfCA1\_B* and *CfPPDK-RP1\_B* had higher CG methylation levels in the intron region than *CfCA1\_A* and *CfPPDK-RP1\_A*. In addition, *CfCA1\_A*, with a significantly longer 5' UTR, contained high CG, CHG, and CHH methylation levels (Figure 5F). By contrast, *CfPEPC-k2\_A* had the lowest CG methylation levels in the intron region. CG, CHG, and CHH methylation levels were lower in *CfPEPC-k2\_A* than in *CfPEPC-k2\_B*.

### C<sub>4</sub> decarboxylation is mainly dependent on NADP-ME with the cooperation of PEPCK

There are three subtypes of C<sub>4</sub> plants, each of which utilizes one of the major decarboxylating enzymes: NADP-ME, NAD-ME, and PEPCK (Maier et al., 2011). Comparative transcriptome analysis revealed that genes encoding NADP-ME (*CfNADP-ME2*) and PEPCK (*CfPEPCK*) showed comparable expression levels during the daytime (Figure 5D and 5E); the TPM of *CfNADP-ME2\_A* and *CfNADP-ME2\_B* was 648 and 965, respectively, and that of *CfPEPCK1\_A* and *CfPEPCK1\_B* was 1060 and 799, indicative of a consistent rhythmic trend between the two enzymes. However, the NAD-ME genes *CfNAD-ME1* and *CfNAD-ME2* showed a rhythmic trend opposite to that of *CfNADP-ME2* and *CfPEPCK* (Figure 5E). Therefore, we speculated that NADP-ME and PEPCK are the two major decarboxylating enzymes and contribute equally to the C<sub>4</sub> decarboxylation reaction of JUJUNCAO, with NAD-ME acting in a complementary manner to the two major enzymes.

### Phylogenetic analysis of *Cenchrus* grasses

Although many *Cenchrus* grasses have been identified to date, only a few genetic resources are available for them. To understand the phylogenetic relationships of other *Cenchrus* grasses, we recently performed genome sequencing of 19 *Cenchrus* grasses, including the *Cenchrus* cultivar Bajra, the *P. americanum* cultivar MZL, elephant grass (*P. purpureum* (Schumacher), ZXC), RY4HWC (*P. purpureum* × *P. americanum* cv. Reyan No. 4), ZJLWC (*P. americanum* × *P. purpureum*), G1Z (*P. americanum* × *P. purpureum* cv. Guimu 1), AXC (*P. purpureum* cv. dwarf), TXC (TAIXUCAO), DSXC

(C) The response of net CO<sub>2</sub> assimilation rate to intercellular CO<sub>2</sub> concentration (Ci) (n = 5) (mean ± SD). Vpmax, maximal PEP carboxylation activity; Vcmax, maximal Rubisco activity; Rd, leaf mitochondrial respiration; Rm, mesophyll mitochondrial respiration.

(D) Schematic of the C<sub>4</sub> pathway in JUJUNCAO. The NADP-ME (purple) and PEPCK (grey) pathway predominates, whereas the NAD-ME (orange) pathways function as subsidiary pathways. CA, carbonic anhydrase; PEPC, phosphoenolpyruvate carboxylase; PEPC-k, PEPC kinase; NADP-MDH, NADP-malate dehydrogenase; NADP-ME, NADP-malic enzyme; NAD-ME, NAD-malic enzyme; PEPCK, phosphoenolpyruvate carboxykinase; PPDK, pyruvate orthophosphate dikinase; PPDK-RP, PPDK regulatory protein. Transporters of C<sub>4</sub> metabolites, including BASS2 (probable sodium/metabolite cotransporter 2), NHD1 (sodium/proton antiporter 1), OMT1 (2-oxoglutarate/malate transporter 1), and DiT2 (dicarboxylate transporter 2), and accessory proteins, including AspAT (aspartate transaminase) and AlaAT (alanine transaminase), were also identified (Supplemental Table 11). The figure shows the expression of genes at different time points (08:00 h, 14:00 h, and 22:00 h). Shaded boxes indicate the time points for the night period (22:00 h). Red indicates genes in the A subgenome, and blue indicates genes in the B subgenome.

(E) Expression patterns of JUJUNCAO carbon fixation genes across the leaf gradient at different time points.

(F) IGV tracks showing methylation and RNA-seq levels of *CfCA1*, *CfPEPC-k2*, and *CfPPDK-RP1*. The patterns of CG, CHG, and CHH methylation in C<sub>4</sub> genes are shown from top to bottom. From left to right, the green background represents the 2-kb region upstream of the gene, the red background represents the gene region, and the blue background represents the 2-kb region downstream of the gene.

(DUOSUIXIANGCAO), HNXC (Huanan XIANGCAO), and HNDS (Hainan DUOSUI) (Lin et al., 2022). Three *S. italica* varieties (SRR13414290, SRR13414328, and SRR13414313) were used as outgroups. These data comprise about 3.842 billion clean reads (~0.5 Tbp) (Supplemental Table S12). We aligned the filtered reads to the JUJUNCAO genome and used the GATK pipeline for variant calling. A total of 149 319 SNPs shared by all the tested grasses were used for principal-component analysis (PCA) and phylogenetic analysis. Both results clustered the grasses into three clades: the three *S. italica* varieties were an outgroup, Bajra and MZL comprised group II, and JUJUNCAO (*C. fungigraminus*) and the other grasses comprised group I. This classification also coincides with their matching rates (Supplemental Figure 18; Supplemental Table S12). Our results also demonstrated that the two species closely related to JUJUNCAO are Bajra and MZL. We next estimated the population dynamics of *Cenchrus* using the A and B subgenomes of JUJUNCAO as ancestor genomes. The effective population of the *Cenchrus* genus was estimated to be ~10 000 at 0.1 kya and more than 10 000 earlier. Our results demonstrated an abrupt decrease in the population size of *Cenchrus* grasses between 200 and 400 kya in both simulations. An additional decrease in population size at 20–50 kya was also observed using the B subgenome of JUJUNCAO as the ancestor genome. These results suggested the elimination of large numbers of *Cenchrus* individuals during the two stages, probably caused by extremely low temperature.

## DISCUSSION

A high-quality genome is very important for evolutionary and genetic studies. Although genomes have been assembled for many plant species, particularly economically important crops, only recently have a few genomes with T2T chromosomes been published. Here, we report a gapless, near-complete genome of the high-biomass grass JUJUNCAO. The contig N50 is 134.1 Mb, which is about 7000 times, 73 times, and 46 times that of pearl millet (Varshney et al., 2017), *C. purpureus* (Yan et al., 2020), and *P. purpureum* (Zhang et al., 2022b), respectively. To the best of our knowledge, this assembly is the highest-quality published Panicoideae genome. Because of the high biomass, stress tolerance, and broad adaptation of JUJUNCAO, its highly complete allopolyploid genome not only provides a high-quality reference genome for a *Cenchrus* grass but also serves as a valuable genetic resource for studying the mechanisms that underlie its robust performance and improving high-biomass grasses.

### Genome evolution

Using the near-complete assembly, we discriminated the two subgenomes of JUJUNCAO and traced the evolutionary trajectory of this paleo-tetraploid. Comparison of genome structures revealed that the two subgenomes of JUJUNCAO were united and doubled ~2.7 mya, much more recently than those of the two allotetraploid elephant grasses *C. purpureus* (6.61 mya; Yan et al., 2020) and *P. purpureum* Schum (15 mya; Zhang et al., 2022b). We proposed the ACK and suggested that the two subgenomes of JUJUNCAO evolved independently from the same ACK. This conclusion differs from that of the previous study by Zhang et al. (2022b), which suggested that after its

divergence from pearl millet, the ancestral A genome split into the ancestors of the A' and B subgenomes of *P. purpureum*. Frequent chromosomal rearrangements may have occurred between the two subgenomes after the allotetraploidization of JUJUNCAO. We found that subgenome A of JUJUNCAO is much more closely related to the two *Cenchrus* cultivars Bajra and MZL than to pearl millet (see also Lin et al., 2022). However, our results also suggested that these two *Cenchrus* cultivars are unlikely to be the parental species of JUJUNCAO. The relatives of JUJUNCAO subgenome B and the two elephant grasses remain unknown (Yan et al., 2020; Zhang et al., 2022b), and more *Cenchrus* species should be collected to resolve the phylogenetic relationships within this genus.

Polyploids generally benefit from whole-genome duplication that combines complementary pathways to confer adaptive advantages and robust agronomic performance. This has been attributed in part to subgenome-biased expression of homeologous genes in polyploids, particularly allopolyploids (Wang et al., 2011; Bertoli et al., 2019). Consistent with previous studies, subgenome-biased genes in JUJUNCAO demonstrated biological functional divergence between the two subgenomes, which was to some extent negatively associated with DNA methylation. We inferred from these results that DNA methylation may lead to divergent expression in the two subgenomes of JUJUNCAO, which is thought to have occurred after the divergence of the two subgenomes.

### Centromere evolution

Centromeres in many plants and animals are often composed of a single satellite repeat, typically with a length ranging from 150 bp to 180 bp (Jiang et al., 2003; Yang et al., 2018). Although the function of centromeres is highly conserved among different species, the centromeric satellites are highly variable, and changes in sequences can be detected within and among species and even among different chromosomes (Cheng et al., 2002; Hall et al., 2003; Zhang et al., 2014; Presting, 2018; Yang et al., 2018; Ahmad et al., 2020). In this study, we revealed the dynamic architecture and evolution of JUJUNCAO centromeres and suggested that centromeric satellite repeats were conserved in Paniceae. Among the three centromeric satellite repeats identified in the JUJUNCAO genome, *CEN156* shared the highest similarity with homologs from closely related species and was thought to be the founding sequence of ancestral *Cenchrus* chromosomes. It diverged into *CEN137* in the common ancestral genome of subgenome A and *P. americanum* MZL and into *CEN148* in the ancestral subgenome B after the divergence of the ancestral A and B subgenomes of JUJUNCAO. The subgenome A-specific *CEN137* and subgenome B-specific *CEN148* suggest that the three types of satellite repeats evolved independently within the two subgenomes. However, introgressions of *CEN156* from subgenome B into subgenome A may have occurred after the paleo-polyploidization of the JUJUNCAO genome. This would be a reasonable explanation for the presence of *CEN156* in subgenome A of JUJUNCAO and its absence from the genome of the diploid *P. americanum* MZL, which is phylogenetically close to subgenome A of JUJUNCAO. We also examined overall DNA methylation patterns in the centromeric regions and showed enrichment of cytosine methylation in the peri/centromeric

regions, with different segments in different centromeres or the same centromere being either hypomethylated or hypermethylated. These results were consistent with those reported for other plants such as rice (Yan et al., 2010), *Arabidopsis*, and maize (Zhang et al., 2008; Koo and Jiang, 2011).

### Mechanisms underlying high photosynthetic efficiency

C<sub>4</sub> plants are among the most productive and efficient biomass producers. The *Pennisetum* grasses were identified as C<sub>4</sub> plants decades ago on the basis of their low compensation point and Kranz anatomy (Coombs and Baldry, 1972; Coombs et al., 1973a, 1973c, 1973b; Huber and Sankhla, 1973). C<sub>4</sub> plants are classically grouped into three biochemical subtypes according to whether they contain high levels of NADP-ME, PEPCK, or NAD-ME for decarboxylation of C<sub>4</sub> acids (malate and/or aspartate) (Dengler and Nelson, 1999; Edwards and Voznesenskaya, 2011). Like other *Pennisetum* grasses, JUJUNCAO is also a C<sub>4</sub> plant, with anatomical and physiological adaptations that optimize CO<sub>2</sub> fixation. The centrifugal chloroplast arrangement of bundle sheath cells observed in JUJUNCAO was consistent with that of classic NADP-ME species such as sugarcane and maize (Sage et al., 2014) but different from the evenly distributed pattern of *P. purpureum*, which is typical of the PEPCK subtype. The anatomical variation between *P. purpureum* and JUJUNCAO could be influenced by environmental factors, such as light (Sales et al., 2018). Currently, *Pennisetum* grasses such as *P. purpureum*, *P. americanum*, and *Pennisetum setaceum* are classified as the NADP-ME subtype (Neto and Guerra, 2019). Many studies have suggested that more than one decarboxylation enzyme may co-exist in a single C<sub>4</sub> plant; that the PEPCK pathway could be complementary in NADP-ME subtype species such as sugarcane, maize, and sorghum; and that the contribution of PEPCK to total decarboxylase activity was lower than that of NADP-ME and differed among species (Wang et al., 2014). Both NADP-ME and PEPCK enzymes may also be present in *Pennisetum* grasses (*P. purpureum*), owing to their use of malate and/or aspartate as the C<sub>4</sub> translocated acid (Coombs et al., 1973c). Interestingly, JUJUNCAO *CfPEPCK* reached a transcript level comparable to that of *CfNADP-ME*. This is similar to results from sugarcane, which is categorized as an NADP-ME species, although *PEPCK* transcript abundance was higher than that of *NADP-ME* when measured by the serial analysis of gene expression (SAGE) technique (Calsa and Figueira, 2007). The rise in PEPCK activity can be modulated by various environmental factors (Furbank, 2011; Sales et al., 2018) such as shading and limited water supply (Sales et al., 2018; Cacefo et al., 2019). Therefore, we propose that high *CfPEPCK* transcription may potentially enable the maintenance of greater photosynthetic efficiency in regions where JUJUNCAO thrives.

Allopolyploid-derived subgenomes begin with distinct, global differences that would be expected to lead to global differences in gene expression levels. We observed no expression bias in genes encoding decarboxylases between the A and B subgenomes. However, the C<sub>4</sub> genes *CfCA1*, *CfPPDK-RP1*, and *CfPEPC-k2* showed some evidence of subgenome dominance, with *CfCA1* and *CfPPDK-RP1* exhibiting subgenome B dominance and *CfPEPC-k2* exhibiting subgenome A dominance. Subgenome dominance led to divergence with respect to expression levels.

The accurate regulation of gene expression in the A and B subgenomes might be indispensable for allopolyploid plant development, and DNA methylation may serve as a regulator of C<sub>4</sub> gene expression (Reeves et al., 2017). Higher CG methylation levels in the intron region may confer higher expression of *CfCA1* and *CfPPDK-RP1* from subgenome B, given that gene-body methylation is positively correlated with gene expression (Jones, 2012; Li et al., 2012). However, the transcript level of *CfPEPC-k2* was higher in subgenome A than in subgenome B. DNA methylation can repress gene expression (Chan et al., 2005). Thus, gene expression of *CfPEPC-k2* in subgenome B might be repressed by DNA methylation of the downstream 2-kb region. In this study, we found that the relationship between gene expression and DNA methylation was complicated by genic regions, methylation contexts, and subgenomes in the allopolyploid.

## METHODS

### Plant materials and sequencing

The JUJUNCAO (*C. fungigraminus*, 2n = 28) and *C. americanus* plants used in this study were grown in the greenhouse at Fujian Agriculture and Forestry University. For Illumina short-read sequencing, genomic DNA extracted from leaf tissue was used for construction of a 280-bp paired-end library with the NEBNext Ultra DNA Library Prep Kit. The library was sequenced using the Illumina NovaSeq 6000 platform. For PacBio HiFi library construction and sequencing, the g-TUBE was used to shear gDNA to ~20-kb fragments for construction of SMRTbell libraries using the SMRTbell Express Template Prep Kit 2.0. The libraries were sequenced using the PacBio Sequel system at Novogene Company. Hi-C libraries were generated and sequenced from young leaves of JUJUNCAO at Novogene as described previously (Xie et al., 2015). In brief, after fixation with formaldehyde, the fresh leaves were lysed for DNA extraction. The cross-linked DNA was digested overnight with *MboI* to generate sticky ends for biotinylation and then proximity ligated to form chimeric junctions. The enriched DNA was then physically sheared to a size of 200–600 bp. Chimeric fragments representing the original cross-linked long-distance physical interactions were used to construct paired-end sequencing libraries, which were sequenced using the Illumina HiSeq 2500 platform (PE 125 bp).

### Genome size estimation

Flow cytometry analysis was performed to estimate the C-value of JUJUNCAO using fresh leaves, and tomato was used as a reference genome control (Michaelson et al., 1991; Dolezel and Bartos, 2005).

### Genome assembly and scaffolding

About 117 Gb of clean PacBio CCS reads were used for *de novo* assembly of the JUJUNCAO genome with default settings of HiFiasm v0.16.1-r375 (Cheng et al., 2021), and the primary output contigs with long stretches of phased blocks were sent to NextPolish v1.3.1 (Hu et al., 2020) for polishing using Illumina short reads. Purge\_haplotigs v1.1.1 (Roach et al., 2018) was used to remove heterozygous redundant sequences with the parameters “purge\_haplotigs contigcov -l 10 -m 80 -h 190.” The total span of the final filtered assembly was 1 986 757 443 bp in 116 contigs with a contig N50 size of

134 074 806 bp. To build a chromosome-level assembly, a Hi-C library was constructed and sequenced as described previously (Zhang et al., 2022a). A total of 233 Gb of Hi-C reads were mapped to the contig-level assembly using BWA v0.7.17 with default parameters (Li, 2013). We corrected the contig-level assembly based on the chromatin contact signals. All of the corrected contigs were remapped with Hi-C reads and reordered and scaffolded using the ALLHiC pipeline (<https://github.com/tangerzhang/ALLHiC>), and the resulting assembly was manually corrected according to the visualization of chromatin contact patterns. Finally, we generated a pseudo-chromosome assembled genome that included 14 chromosomes.

### Gene annotation

In brief, we integrated sequence homology, *de novo* prediction, and transcriptome data to build consensus gene models using the GETA pipeline (<https://github.com/chenlianfu/geta>). Protein sequences from closely related grass species (*C. americanus*, *S. bicolor*, and *O. sativa*) were used to perform homology-based prediction. The GETA pipeline randomly selected 1000 homology-based genes to train AUGUSTUS v3.4.0 (Stanke et al., 2004) for *de novo* prediction on the pre-masked genome sequences. We provided the pipeline with RNAs isolated from root, stem, and leaf tissues of JUJUNCAO. To assemble the transcriptomes, trimmomatic v0.39 (Bolger et al., 2014) was used with default parameters to filter the RNA-seq reads. The filtered reads were aligned to the reference genome with HISAT2 v2.1.0 (Kim et al., 2015). The reliable intron and optimal transcript information were identified based on RNA-seq alignment. We next assembled and filtered the transcripts with the “sam2transfrag” function in the GETA pipeline, which retains the isoform of each gene with the highest RNA-seq read depth as the representative transcript. Finally, gene models from these three methods were integrated into a non-redundant high-confidence set of gene models.

### Repeat sequence annotation

A repeat library of the JUJUNCAO genome was constructed *ab initio* using RepeatModeler v2.0.2 with a combination of *de novo* and homology strategies, including two *de novo* repeat-finding programs, RECON and RepeatScout, which we imported into RepeatMasker v4.1.2 (<http://www.repeatmasker.org/>) for identification and clustering of repetitive elements. We also integrated results from LTR\_FINDER (Xu and Wang, 2007) and LTRharvest (Ellinghaus et al., 2008) and removed false positives from the initial predictions using the LTR\_retriever pipeline (Ou and Jiang, 2018) to obtain the final full-length LTR-RTs.

### Telomere and centromere detection

Tandem repeats were identified in the JUJUNCAO genome with Tandem Repeat Finder v4.09 (TRF; parameters: 2 7 7 80 10 50 500 -m -f -d) (Benson, 1999), and the telomere and three types of centromere sequences were found. To precisely investigate the distributions of the three centromere repeat monomers (*CEN137*, *CEN148*, *CEN156*), we divided the genome into adjacent 5-kbp windows, defined 12-mers, and exactly matched these sequences to the window. For each region, we calculated peak values of the distances between 13-mers to test for periodic repeats. For example, if a window contained *CEN137* repeats (monomer size of 137 bp), the most frequent 13-mer

distance would be 137. We merged the same kind of windows, and *CEN137*, *CEN148*, and *CEN156* sequences were matched back to corresponding windows using matchPattern (max.mismatch = 3/peakvalues, with.indels = T) (<https://github.com/Bioconductor/Biostrings>). Finally, we identified 159 310 bp of *CEN137* repeats, 24 075 bp of *CEN148* repeats, and 216 886 bp of *CEN156* repeats in all 14 chromosomes, and several were aligned by mafft.v 7.4 (Katoh et al., 2009) to generate the consensus sequence (Supplemental table 7).

### Subgenome identification

We resolved the JUJUNCAO genome into subgenomes A and B using previously described methods (Session et al., 2016; Mitros et al., 2020). First, homeologous chromosomes were determined based on their conserved synteny to each other and to *C. americanus*. Then, Jellyfish v2.2.10 (Marçais and Kingsford, 2011) was used to scan 13-bp sequences (13-mers), and 32 629 subgenome-specific 13-mers were identified using the following criteria: (1) occurred >1000 times across the genome, and (2) were more than two-fold enriched in one member of each homeologous chromosome pair (setting aside the fusion-related chromosomes). These 13-mers could be clearly divided into two groupings (A 13-mers and B 13-mers) by hierarchical clustering. Similarly, chromosome clustering partitioned the genome into subgenomes A and B.

### Phylogenetic analysis, genomic comparison, and divergence time estimation

Protein sequences of *S. spontaneum*, *M. sinensis*, *S. bicolor*, *Z. mays*, *P. hallii*, *P. miliaceum*, *P. virgatum*, *C. americanus*, *C. fungigraminus*, *S. italica*, *S. viridis*, and *O. sativa* were used to identify gene family clusters with Orthofinder v2.5.4. We used the option “-M msa” to obtain maximum likelihood trees from multiple sequence alignments. Divergence times were estimated using r8s v1.81. Calibration times were obtained from the TimeTree database (<http://www.timetree.org/>). Expansion and contraction of Orthofinder-derived gene clusters were determined using CAFÉ v4.2 (<https://github.com/hahnlab/CAFE>) and were based on changes in gene family size in the inferred phylogenetic history. Collinearity between genomes was analyzed using JCVI (<https://github.com/tanghaibao/jcvi>) with default parameters, and all orthologous and paralogous gene pairs were identified on the basis of syntenic blocks. Paired synonymous substitution rates (Ks) were calculated using the Nei-Gojobori method ([https://github.com/tanghaibao/bio-pipeline/tree/master/synonymous\\_calculation](https://github.com/tanghaibao/bio-pipeline/tree/master/synonymous_calculation)).

### Karyotype evolution analysis of Paniceae

The distribution of seven ancestral eudicot chromosomal lineages for each chromosome in each species was depicted by syntenic blocks between the ancestral chromosomes of rice (Murat et al., 2017) and those of the detected species.

### SNP identification and population genomics analysis

Paired-end raw reads were filtered by removing reads with adapters, reads with >10% N content, and reads with more than 50% low-quality bases ( $Q \leq 5$ ). Filtered paired-end reads were mapped to the assembled JUJUNCAO genome using Bowtie2 v2.4.4 (Langdon, 2015). We then used Picard (v1.129)

(<https://github.com/broadinstitute/picard>) to mark the duplicated reads. SNPs were identified using the GATK toolkit v.3.5 (<https://github.com/broadinstitute/gatk>) with the HaplotypeCaller module. Next, we merged the GCVF files using the GenotypeGVCFs tool in GATK to generate raw SNP files. The SNP filter application VCFtools v0.1.13 (Danecek et al., 2011) was used with “-max-missing 0.8 -maf 0.05 -max-alleles 2 -min-meanDP 4 -minQ 200.” Vcf2phylip software (<https://github.com/edgardomortiz/vcf2phylip>) was used to convert the format, and RaxML v8.12 (Stamatakis, 2014) was used to construct a maximum likelihood tree. PCA was performed using PLINK v1.90 (Purcell et al., 2007) to transform the VCF file into a PLINK file for input.

### Demographic analysis

To calculate the site-frequency spectrum (SFS) of *Pennisetum* populations, we used the “-doSaf” parameter of ANGSD (Korneliussen et al., 2014) to calculate the site allele frequency likelihood based on individual genotype likelihoods assuming HWE from filtered BAM files. Then the “-realSFS” parameter was used to obtain a maximum likelihood estimate of the SFS by an expectation maximization (EM) algorithm. The folded SFS was used to estimate population demographic history by stairway plots (Liu and Fu, 2015) with 200 bootstrap iterations. We used  $6.5e-9$  as the mutation rate and a generation time of 1 year for the analysis.

### Bisulfite whole-genome sequencing and bioinformatic analysis

The JUJUNCAO reference genome and the clean reads generated from bisulfite sequencing were transformed into bisulfite-converted sequences (C-to-T and G-to-A converted). The clean reads were then aligned to the reference genome using BISMARCK with default parameters, and methylation information was extracted using BISMARCK\_METHYLATION\_EXTRACTOR with the following options: -comprehensive -bedGraph -counts -CX\_context -cytosine\_report. The methylation level of an individual cytosine was calculated from the number of methylated cytosines divided by the total cytosine depth (i.e., mC/(mC + non-mC)). Only sites that covered more than three mapped reads were used for subsequent analysis. Sites with more than one methylated cytosine mapping were considered methylation sites.

### FISH

FISH experiments were performed as described previously (Huang et al., 2021). Oligomers specific for each centromeric satellite repeat (Supplemental Table 13) were designed according to their sequence alignment results (Figure 3). The oligomers were labeled via nick-translation using digoxigenin-11-dUTP or biotin-16-dUTP (Roche Diagnostics). Digoxigenin-labeled probes and biotin-labeled probes were detected using rhodamine-conjugated anti-digoxigenin (Roche Diagnostics) and Alexa Fluor 488 streptavidin (Thermo Fisher Scientific), respectively. 4',6-Diamidino-2-phenylindole (DAPI) was used to stain the chromosomes. Images of the samples were captured with an Olympus BX63 fluorescence microscope using an Olympus DP80 CCD camera (Olympus).

### Histological analysis

One-centimeter strips were cut crosswise to the leaf and fixed for 12 h in a solution of 50% ethanol, 5% acetic acid, and 37% form-

aldehyde at 4°C. The strips were then processed in a series of ethanol solutions for 1 h each (50%, 60%, 70%, 85%, 95%, 100%), followed by mixtures of xylene and ethanol for 30 min each (25% xylene + 75% ethanol, 50% xylene + 50% ethanol, 25% ethanol + 75% xylene, 100% xylene) and finally paraffin for 12 h. The tissue was then oriented in paraffin and sectioned at 8  $\mu\text{m}$  using a rotary microtome (Leica, RM2255). Tissue was stained with 1% safranin and fast green solution and viewed under a microscope (Leica, BX3-CBH).

### Identification of genes related to key characteristics of JUJUNCAO

$C_4$  pathway genes from sorghum and *S. spontaneum* were used as query sequences for Basic Local Alignment Search Tool (BLASTP) searches against the predicted sequences of JUJUNCAO. Sequences with E-value  $<1e-5$  were selected for further analysis.

### Phylogenetic analysis of $C_4$ pathway genes

Based on protein sequence alignment, a phylogenetic tree of the  $C_4$  pathway gene family was constructed by the neighbor-joining (NJ) method. The tree was constructed using MEGAX (Kumar et al., 2018) with the “pairwise deletion” option and the “Poisson correction” model, and the reliability of internal branches was assessed using 1000 bootstrap replicates. The results were imported into the interactive Tree of Life (iTOL) (Letunic and Bork, 2016) to create the phylogenetic tree.

### Gas exchange measurements

Leaf gas exchange was measured in the +1 leaf (the first fully expanded leaf from the top to bottom) using a portable infrared gas analyzer (LI-COR 6800XT, LI-COR, Lincoln, NE, USA). Measurements were made at a leaf temperature of 30°C, humidity of 65%, and flow rate of 500  $\mu\text{mol s}^{-1}$ . The light-response curves were made at a  $\text{CO}_2$  concentration of 400  $\mu\text{mol mol}^{-1}$  and photosynthetic photon flux densities (PPFDs) of 2000, 1500, 1000, 800, 600, 400, 200, 150, 100, 75, 50, and 0  $\mu\text{mol m}^{-2} \text{s}^{-1}$ . The Excel sheet used for curve fitting was downloaded from <http://landflux.org/tools>.  $\text{CO}_2$  response curves were measured at a PPFD of 2000  $\mu\text{mol m}^{-2} \text{s}^{-1}$  and  $\text{CO}_2$  concentrations of 400, 300, 200, 120, 70, 40, 20, 0, 400, 400, 400, 600, 800, 1200, and 1500  $\mu\text{mol mol}^{-1}$ . The  $\text{CO}_2$  response curves were fitted according to von Caemmerer.

### DATA AVAILABILITY

The whole-genome sequencing data (including Illumina short reads, HiFi reads, and Hi-C interaction reads), bisulfite whole-genome sequencing data, and transcriptomes of different tissues used in this study have been deposited at the National Genomics Data Center under accession number PRJCA013426. The genome assembly and gene annotation data for *C. fungigraminus* have been deposited at the Genome Warehouse (GWH) under accession number GWHBWDX00000000.1 and are publicly accessible at <https://ngdc.cnpc.ac.cn/gwh>.

### SUPPLEMENTAL INFORMATION

Supplemental information is available at *Plant Communications Online*.

### FUNDING

This work was supported by grants from the Major Special Project of Fujian Province (2021NZ029009) and the Natural Science foundation of Fujian Province (2019J01665).

## AUTHOR CONTRIBUTIONS

J.Z., Z.L., and Z.W. conceived the project. X.H., H.Z., Y.H., Z.Z., and H.L. collected the samples and performed the experiments. B.W., X.H., R.G., Y.W., Y.Z., J.M., and Y.H. contributed to genome assembly and data analysis. J.Z., H.Z., X.H., and B.W. wrote the manuscript. J.Z. and R.M. revised the manuscript. All authors reviewed the manuscript.

## ACKNOWLEDGMENTS

We would like to thank Prof. Xinlong Liu at Yunnan Key Laboratory of Sugarcane Genetic Improvement, Sugarcane Research Institute, Yunnan Academy of Agricultural Sciences, Kaiyuan, China for kindly providing the *Cenchrus* cultivars. No conflict of interest is declared.

Received: January 28, 2023

Revised: April 7, 2023

Accepted: June 1, 2023

Published: June 3, 2023

## REFERENCES

- Adams, K.L., and Wendel, J.F. (2005). Polyploidy and genome evolution in plants. *Curr. Opin. Plant Biol.* **8**:135–141. <https://doi.org/10.1016/j.pbi.2005.01.001>.
- Ahmad, S.F., Singchat, W., Jehangir, M., Suntronpong, A., Panthum, T., Malaivijitnond, S., and Srikulnath, K. (2020). Dark matter of primate genomes: satellite DNA repeats and their evolutionary dynamics. *Cells* **9**:e122714. <https://doi.org/10.3390/cells9122714>.
- Belser, C., Baurens, F.C., Noel, B., Martin, G., Cruaud, C., Istace, B., Yahiaoui, N., Labadie, K., Hřibová, E., Doležel, J., et al. (2021). Telomere-to-telomere gapless chromosomes of banana using nanopore sequencing. *Commun. Biol.* **4**:1047. <https://doi.org/10.1038/s42003-021-02559-3>.
- Benson, G. (1999). Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res.* **27**:573–580. <https://doi.org/10.1093/nar/27.2.573>.
- Bertioli, D.J., Jenkins, J., Clevenger, J., Dudchenko, O., Gao, D., Seijo, G., Leal-Bertioli, S.C.M., Ren, L., Farmer, A.D., Pandey, M.K., et al. (2019). The genome sequence of segmental allotetraploid peanut *Arachis hypogaea*. *Nat. Genet.* **51**:877–884. <https://doi.org/10.1038/s41588-019-0405-z>.
- Bird, K.A., Niederhuth, C.E., Ou, S., Gehan, M., Pires, J.C., Xiong, Z., VanBuren, R., and Edger, P.P. (2021). Replaying the evolutionary tape to investigate subgenome dominance in allopolyploid *Brassica napus*. *New Phytol.* **230**:354–371. <https://doi.org/10.1111/nph.17137>.
- Bolger, A.M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**:2114–2120. <https://doi.org/10.1093/bioinformatics/btu170>.
- Cacefo, V., Ribas, A.F., Zilliani, R.R., Neris, D.M., Domingues, D.S., Moro, A.L., and Vieira, L.G.E. (2019). Decarboxylation mechanisms of C<sub>4</sub> photosynthesis in *Saccharum* spp.: increased PEPCK activity under water-limiting conditions. *BMC Plant Biol.* **19**:144. <https://doi.org/10.1186/s12870-019-1745-7>.
- Calsa, T., Jr., and Figueira, A. (2007). Serial analysis of gene expression in sugarcane (*Saccharum* spp.) leaves revealed alternative C<sub>4</sub> metabolism and putative antisense transcripts. *Plant Mol. Biol.* **63**:745–762. <https://doi.org/10.1007/s11103-006-9121-z>.
- Chan, S.W., Henderson, I.R., and Jacobsen, S.E. (2005). Gardening the genome: DNA methylation in *Arabidopsis thaliana*. *Nat Rev Genet* **6**:351–360.
- Chen, B., Ye, J., Luo, Z., and Lin, J. (2016). ISSR Analysis of genetic diversity of *Pennisetum* germplasm resource. *Heilongjiang Anim. Sci. Vet. Med.* **129**–133.
- Cheng, H., Concepcion, G.T., Feng, X., Zhang, H., and Li, H. (2021). Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. *Nat. Methods* **18**:170–175. <https://doi.org/10.1038/s41592-020-01056-5>.
- Cheng, Z., Dong, F., Langdon, T., Ouyang, S., Buell, C.R., Gu, M., Blattner, F.R., and Jiang, J. (2002). Functional rice centromeres are marked by a satellite repeat and a centromere-specific retrotransposon. *Plant Cell* **14**:1691–1704. <https://doi.org/10.1105/tpc.003079>.
- Comai, L., Maheshwari, S., and Marimuthu, M.P.A. (2017). Plant centromeres. *Curr. Opin. Plant Biol.* **36**:158–167. <https://doi.org/10.1016/j.pbi.2017.03.003>.
- Coombs, J., and Baldry, C.W. (1972). C-4 pathway in *Pennisetum purpureum*. *Nat. New Biol.* **238**:268–270. <https://doi.org/10.1038/newbio238268a0>.
- Coombs, J., Baldry, C.W., and Bucke, C. (1973a). The C-4 pathway in *Pennisetum purpureum* : I. The allosteric nature of PEP carboxylase. *Planta* **110**:95–107. <https://doi.org/10.1007/bf00384832>.
- Coombs, J., Baldry, C.W., and Bucke, C. (1973b). The C-4 pathway in *Pennisetum purpureum* : II. Malate dehydrogenase and malic enzyme. *Planta* **110**:109–120. <https://doi.org/10.1007/bf00384833>.
- Coombs, J., Baldry, C.W., and Brown, J.E. (1973c). The C-4 pathway in *Pennisetum purpureum* : III. Structure and photosynthesis. *Planta* **110**:121–129. <https://doi.org/10.1007/bf00384834>.
- Danecek, P., Auton, A., Abecasis, G., Albers, C.A., Banks, E., DePristo, M.A., Handsaker, R.E., Lunter, G., Marth, G.T., Sherry, S.T., et al. (2011). The variant call format and VCFtools. *Bioinformatics* **27**:2156–2158. <https://doi.org/10.1093/bioinformatics/btr330>.
- Dengler, N.G., and Nelson, T. (1999). 5 - leaf structure and development in C<sub>4</sub> Plants. In *C4 Plant Biology*, R.F. Sage and R.K. Monson, eds. (Academic Press), pp. 133–172. <https://doi.org/10.1016/B978-012614440-6/50006-9>.
- Doležel, J., and Bartos, J. (2005). Plant DNA flow cytometry and estimation of nuclear genome size. *Ann. Bot.* **95**:99–110. <https://doi.org/10.1093/aob/mci005>.
- Edger, P.P., Smith, R., McKain, M.R., Cooley, A.M., Vallejo-Marin, M., Yuan, Y., Bewick, A.J., Ji, L., Platts, A.E., Bowman, M.J., et al. (2017). Subgenome dominance in an interspecific hybrid, synthetic allopolyploid, and a 140-year-old naturally established neo-allopolyploid monkeyflower. *Plant Cell* **29**:2150–2167. <https://doi.org/10.1105/tpc.17.00010>.
- Edwards, G.E., and Voznesenskaya, E.V. (2011). Chapter 4 C<sub>4</sub> photosynthesis: Kranz forms and single-cell C<sub>4</sub> in terrestrial plants. In *C4 Photosynthesis and Related CO2 Concentrating Mechanisms*, A.S. Raghavendra and R.F. Sage, eds. (Springer Netherlands), pp. 29–61. [https://doi.org/10.1007/978-90-481-9407-0\\_4](https://doi.org/10.1007/978-90-481-9407-0_4).
- Eisler, M.C., Lee, M.R.F., Tarlton, J.F., Martin, G.B., Beddington, J., Dungait, J.A.J., Greathead, H., Liu, J., Mathew, S., Miller, H., et al. (2014). Agriculture: steps to sustainable livestock. *Nature* **507**:32–34. <https://doi.org/10.1038/507032a>.
- Ellinghaus, D., Kurtz, S., and Willhoeft, U. (2008). LTRharvest, an efficient and flexible software for de novo detection of LTR retrotransposons. *BMC Bioinf.* **9**:18. <https://doi.org/10.1186/1471-2105-9-18>.
- Furbank, R.T. (2011). Evolution of the C(4) photosynthetic mechanism: are there really three C(4) acid decarboxylation types? *J. Exp. Bot.* **62**:3103–3108. <https://doi.org/10.1093/jxb/err080>.
- Gu, B., Zhang, X., Bai, X., Fu, B., and Chen, D. (2019). Four steps to food security for swelling cities. *Nature* **566**:31–33.
- Hall, A.E., Keith, K.C., Hall, S.E., Copenhaver, G.P., and Preuss, D. (2004). The rapidly evolving field of plant centromeres. *Curr. Opin. Plant Biol.* **7**:108–114. <https://doi.org/10.1016/j.pbi.2004.01.008>.

- Hall, S.E., Kettler, G., and Preuss, D.** (2003). Centromere satellites from *Arabidopsis* populations: maintenance of conserved and variable domains. *Genome Res.* **13**:195–205. <https://doi.org/10.1101/gr.593403>.
- Hayat, K., Zhou, Y., Menhas, S., Bundschuh, J., Hayat, S., Ullah, A., Wang, J., Chen, X., Zhang, D., and Zhou, P.** (2020). *Pennisetum giganteum*: an emerging salt accumulating/tolerant non-conventional crop for sustainable saline agriculture and simultaneous phytoremediation. *Environ. Pollut.* **265**, 114876. <https://doi.org/10.1016/j.envpol.2020.114876>.
- He, K., Huang, Y., Jiang, F., Lin, J., Cheng, Z., Zhang, K., Xie, W., and Huang, Z.** (2017). Effects of two types of herb plants' roots on soil moisture in the alluvial soil in Changning County. *Science of Soil and Water Conservation* **15**:25–34.
- Hu, J., Fan, J., Sun, Z., and Liu, S.** (2020). NextPolish: a fast and efficient genome polishing tool for long-read assembly. *Bioinformatics* **36**:2253–2255. <https://doi.org/10.1093/bioinformatics/btz891>.
- Huang, Y., Ding, W., Zhang, M., Han, J., Jing, Y., Yao, W., Hasterok, R., Wang, Z., and Wang, K.** (2021). The formation and evolution of centromeric satellite repeats in *Saccharum* species. *Plant J.* **106**:616–629. <https://doi.org/10.1111/tpj.15186>.
- Huber, W., and Sankhla, N.** (1973). Eco-physiological studies on Indian arid zone plants : II. Effect of salinity and gibberellin on the activity of the enzymes of amino-acid metabolism in leaves of *Pennisetum typhoides*. *Oecologia* **13**:271–277. <https://doi.org/10.1007/bf00360516>.
- Jia, Y., Liao, Z., Chew, H., Wang, L., Lin, B., Chen, C., Lu, G., and Lin, Z.** (2020). Effect of *Pennisetum giganteum* z.x.lin mixed nitrogen-fixing bacterial fertilizer on the growth, quality, soil fertility and bacterial community of pakchoi (*Brassica chinensis* L.). *PLoS One* **15**, e0228709. <https://doi.org/10.1371/journal.pone.0228709>.
- Jiang, J., Birchler, J.A., Parrott, W.A., and Dawe, R.K.** (2003). A molecular view of plant centromeres. *Trends Plant Sci.* **8**:570–575. <https://doi.org/10.1016/j.tplants.2003.10.011>.
- Jones, P.A.** (2012). Functions of DNA methylation: islands, start sites, gene bodies and beyond. *Nat. Rev. Genet.* **13**:484–492. <https://doi.org/10.1038/nrg3230>.
- Katoh, K., Asimenos, G., and Toh, H.** (2009). Multiple alignment of DNA sequences with MAFFT. *Methods Mol. Biol.* **537**:39–64. [https://doi.org/10.1007/978-1-59745-251-9\\_3](https://doi.org/10.1007/978-1-59745-251-9_3).
- Kim, D., Langmead, B., and Salzberg, S.L.** (2015). HISAT: a fast spliced aligner with low memory requirements. *Nat. Methods* **12**:357–360. <https://doi.org/10.1038/nmeth.3317>.
- Koo, D.H., and Jiang, J.** (2011). Super-stretched pachytene chromosomes for plant molecular cytogenetic mapping. *Methods Mol. Biol.* **701**:239–245. [https://doi.org/10.1007/978-1-61737-957-4\\_13](https://doi.org/10.1007/978-1-61737-957-4_13).
- Korneliusson, T.S., Albrechtsen, A., and Nielsen, R.** (2014). ANGSD: analysis of next generation sequencing data. *BMC Bioinf.* **15**:356. <https://doi.org/10.1186/s12859-014-0356-4>.
- Kuang, W., Liu, J., Tian, H., Shi, H., Dong, J., Song, C., Li, X., Du, G., Hou, Y., Lu, D., et al.** (2022). Cropland redistribution to marginal lands undermines environmental sustainability. *Natl. Sci. Rev.* **9**:nwab091. <https://doi.org/10.1093/nsr/nwab091>.
- Kumar, S., Stecher, G., Li, M., Knyaz, C., and Tamura, K.** (2018). Mega X: molecular evolutionary genetics analysis across computing platforms. *Mol. Biol. Evol.* **35**:1547–1549. <https://doi.org/10.1093/molbev/msy096>.
- Langdale, J.A.** (2011). C<sub>4</sub> cycles: past, present, and future research on C<sub>4</sub> photosynthesis. *Plant Cell* **23**:3879–3892. <https://doi.org/10.1105/tpc.111.092098>.
- Langdon, W.B.** (2015). Performance of genetic programming optimised Bowtie2 on genome comparison and analytic testing (GCAT) benchmarks. *BioData Min.* **8**:1. <https://doi.org/10.1186/s13040-014-0034-0>.
- Letunic, I., and Bork, P.** (2016). Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Res.* **44**:242–245. <https://doi.org/10.1093/nar/gkw290>.
- Li, H.** (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. Preprint at arXiv. <https://doi.org/10.48550/arXiv.1303.3997>.
- Li, P., Ponnala, L., Gandotra, N., Wang, L., Si, Y., Tausta, S.L., Kebrom, T.H., Provart, N., Patel, R., Myers, C.R., et al.** (2010). The developmental dynamics of the maize leaf transcriptome. *Nat. Genet.* **42**:1060–1067. <https://doi.org/10.1038/ng.703>.
- Li, Q., Xiang, C., Xu, L., Cui, J., Fu, S., Chen, B., Yang, S., Wang, P., Xie, Y., Wei, M., and Wang, Z.** (2020). SMRT sequencing of a full-length transcriptome reveals transcript variants involved in C18 unsaturated fatty acid biosynthesis and metabolism pathways at chilling temperature in *Pennisetum giganteum*. *BMC Genom.* **21**:52. <https://doi.org/10.1186/s12864-019-6441-3>.
- Li, X., Zhu, J., Hu, F., Ge, S., Ye, M., Xiang, H., Zhang, G., Zheng, X., Zhang, H., Zhang, S., et al.** (2012). Single-base resolution maps of cultivated and wild rice methylomes and regulatory roles of DNA methylation in plant gene expression. *BMC Genom.* **13**:300. <https://doi.org/10.1186/1471-2164-13-300>.
- Liang, Z., and Schnable, J.C.** (2018). Functional divergence between subgenomes and gene pairs after whole genome duplications. *Mol. Plant* **11**:388–397. <https://doi.org/10.1016/j.molp.2017.12.010>.
- Lin, X., Mao, Y., Qi, Q., Zhang, C., Tian, Y., Chen, Y., and Lin, Z.** (2015). Sequence analysis on ITS and chloroplast *matK* gene of six *Pennisetum* JUNCAO. *Diagn. Pathol.* **10**:174–180.
- Lin, Z.** (2013). *Juncao Science* (Beijing: National School of Administration Press).
- Lin, Z., Lin, D., Liu, Z., and Siren, L.** (2022). *Cenchrus fungigraminus* Z. X. Lin & D. M. Lin & S. R. Lan sp. nov., a new species of Panicoideae (Poaceae) : evidence from morphological, nuclear and plastid genome data. *J. For. Environ.* **42**:514–520.
- Liu, J., Seetharam, A.S., Chougule, K., Ou, S., Swentowsky, K.W., Gent, J.I., Llaca, V., Woodhouse, M.R., Manchanda, N., Presting, G.G., et al.** (2020a). Gapless assembly of maize chromosomes using long-read technologies. *Genome Biol.* **21**:121. <https://doi.org/10.1186/s13059-020-02029-9>.
- Liu, X., and Fu, Y.X.** (2015). Exploring population size changes using SNP frequency spectra. *Nat. Genet.* **47**:555–559. <https://doi.org/10.1038/ng.3254>.
- Liu, Y., Su, H., Zhang, J., Shi, L., Liu, Y., Zhang, B., Bai, H., Liang, S., Gao, Z., Birchler, J.A., and Han, F.** (2020b). Rapid birth or death of centromeres on fragmented chromosomes in maize. *Plant Cell* **32**:3113–3123. <https://doi.org/10.1105/tpc.20.00389>.
- Luo, Z.-z., Lin, J.-r., Lin, Z.-k., Chen, B.-c., Mei, L., and Lin, Z.-x.** (2016). Karyotype analysis of *Pennisetum purpureum* × *P. americanum* cv. Reyan No.4 and *P. purpureum* cv. Guiminyin. *Pratacult. Sci.* **33**:1711–1717.
- Maier, A., Zell, M.B., and Maurino, V.G.** (2011). Malate decarboxylases: evolution and roles of NAD(P)-ME isoforms in species performing C(4) and C(3) photosynthesis. *J. Exp. Bot.* **62**:3061–3069. <https://doi.org/10.1093/jxb/err024>.
- Marçais, G., and Kingsford, C.** (2011). A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinformatics* **27**:764–770. <https://doi.org/10.1093/bioinformatics/btr011>.
- Michaelson, M.J., Price, H.J., Ellison, J.R., and Johnston, J.S.** (1991). Comparison of plant DNA contents determined by Feulgen microspectrophotometry and laser flow cytometry. *Am. J. Bot.* **78**:183–188.



- Mitros, T., Session, A.M., James, B.T., Wu, G.A., Belaffif, M.B., Clark, L.V., Shu, S., Dong, H., Barling, A., Holmes, J.R., et al. (2020). Genome biology of the paleotetraploid perennial biomass crop *Miscanthus*. *Nat. Commun.* **11**:5442. <https://doi.org/10.1038/s41467-020-18923-6>.
- Mullet, J.E. (2017). High-biomass C(4) grasses-Filling the yield gap. *Plant Sci.* **261**:10–17. <https://doi.org/10.1016/j.plantsci.2017.05.003>.
- Murat, F., Armero, A., Pont, C., Klopp, C., and Salse, J. (2017). Reconstructing the genome of the most recent common ancestor of flowering plants. *Nat. Genet.* **49**:490–496. <https://doi.org/10.1038/ng.3813>.
- Navrátilová, P., Toegelová, H., Tulpová, Z., Kuo, Y.T., Stein, N., Doležel, J., Houben, A., Šimková, H., and Mascher, M. (2022). Prospects of telomere-to-telomere assembly in barley: analysis of sequence gaps in the MorexV3 reference genome. *Plant Biotechnol. J.* **20**:1373–1386. <https://doi.org/10.1111/pbi.13816>.
- Neto, M.A.M., and Guerra, M.P. (2019). A new method for determination of the photosynthetic pathway in grasses. *Photosynth. Res.* **142**:51–56. <https://doi.org/10.1007/s11220-019-00646-5>.
- Nurk, S., Koren, S., Rhie, A., Rautiainen, M., Bizakadze, A.V., Mikheenko, A., Vollger, M.R., Altemose, N., Uralsky, L., Gershman, A., et al. (2022). The complete sequence of a human genome. *Science (New York, N.Y.)* **376**:44–53. <https://doi.org/10.1126/science.abj6987>.
- Ou, S., and Jiang, N. (2018). LTR\_retriever: a highly accurate and sensitive program for identification of long terminal repeat retrotransposons. *Plant Physiol.* **176**:1410–1422. <https://doi.org/10.1104/pp.17.01310>.
- Ou, S., Chen, J., and Jiang, N. (2018). Assessing genome assembly quality using the LTR Assembly Index (LAI). *Nucleic Acids Res.* **46**:e126. <https://doi.org/10.1093/nar/gky730>.
- Presting, G.G. (2018). Centromeric retrotransposons and centromere function. *Curr. Opin. Genet. Dev.* **49**:79–84. <https://doi.org/10.1016/j.gde.2018.03.004>.
- Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A.R., Bender, D., Maller, J., Sklar, P., de Bakker, P.I.W., Daly, M.J., and Sham, P.C. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**:559–575. <https://doi.org/10.1086/519795>.
- Reeves, G., Grangé-Guermente, M.J., and Hibberd, J.M. (2017). Regulatory gateways for cell-specific gene expression in C<sub>4</sub> leaves with Kranz anatomy. *J. Exp. Bot.* **68**:107–116. <https://doi.org/10.1093/jxb/erw438>.
- Roach, M.J., Schmidt, S.A., and Borneman, A.R. (2018). Purge Haplotigs: allelic contig reassignment for third-gen diploid genome assemblies. *BMC Bioinf.* **19**:460. <https://doi.org/10.1186/s12859-018-2485-7>.
- Sage, R.F., Peixoto, M.M., and Sage, T.L. (2014). *Photosynthesis in Sugarcane: Physiology, Biochemistry and Functional Biology*, 1st ed. (New York: Wiley), p. 121.
- Sales, C.R., Ribeiro, R.V., Hayashi, A.H., Marchiori, P.E., Silva, K.I., Martins, M.O., Silveira, J.A., Silveira, N.M., and Machado, E.C. (2018). Flexibility of C<sub>4</sub> decarboxylation and photosynthetic plasticity in sugarcane plants under shading. *Environ. Exp. Bot.* **149**:34–42. <https://doi.org/10.1016/j.envexpbot.2017.10.027>.
- Samson, R., Mani, S., Boddey, R., Sokhansanj, S., Quesada, D., Urquiaga, S., Reis, V., and Ho Lem, C. (2005). The potential of C<sub>4</sub> perennial grasses for developing a global. BIOHEAT industry **24**:461–495.
- Session, A.M., Uno, Y., Kwon, T., Chapman, J.A., Toyoda, A., Takahashi, S., Fukui, A., Hikosaka, A., Suzuki, A., Kondo, M., et al. (2016). Genome evolution in the allotetraploid frog *Xenopus laevis*. *Nature* **538**:336–343. <https://doi.org/10.1038/nature19840>.
- Simão, F.A., Waterhouse, R.M., Ioannidis, P., Kriventseva, E.V., and Zdobnov, E.M. (2015). BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **31**:3210–3212. <https://doi.org/10.1093/bioinformatics/btv351>.
- Soltis, D.E., Albert, V.A., Leebens-Mack, J., Bell, C.D., Paterson, A.H., Zheng, C., Sankoff, D., Depamphilis, C.W., Wall, P.K., and Soltis, P.S. (2009). Polyploidy and angiosperm diversification. *Am. J. Bot.* **96**:336–348. <https://doi.org/10.3732/ajb.0800079>.
- Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**:1312–1313. <https://doi.org/10.1093/bioinformatics/btu033>.
- Stanke, M., Steinkamp, R., Waack, S., and Morgenstern, B. (2004). AUGUSTUS: a web server for gene finding in eukaryotes. *Nucleic Acids Res.* **32**:309–312. <https://doi.org/10.1093/nar/gkh379>.
- Varshney, R.K., Shi, C., Thudi, M., Mariac, C., Wallace, J., Qi, P., Zhang, H., Zhao, Y., Wang, X., Rathore, A., et al. (2017). Pearl millet genome sequence provides a resource to improve agronomic traits in arid environments. *Nat. Biotechnol.* **35**:969–976. <https://doi.org/10.1038/nbt.3943>.
- Wang, X., Tang, H., and Paterson, A.H. (2011). Seventy million years of concerted evolution of a homoeologous chromosome pair, in parallel, in major Poaceae lineages. *Plant Cell* **23**:27–37. <https://doi.org/10.1105/tpc.110.080622>.
- Wang, Y., Bräutigam, A., Weber, A.P.M., and Zhu, X.G. (2014). Three distinct biochemical subtypes of C<sub>4</sub> photosynthesis? A modelling analysis. *J. Exp. Bot.* **65**:3567–3578. <https://doi.org/10.1093/jxb/eru058>.
- Wendel, J.F., Jackson, S.A., Meyers, B.C., and Wing, R.A. (2016). Evolution of plant genome architecture. *Genome Biol.* **17**:37. <https://doi.org/10.1186/s13059-016-0908-1>.
- Xie, Q., and Xu, Z. (2019). Sustainable agriculture: from sweet sorghum planting and ensiling to ruminant feeding. *Mol. Plant* **12**:603–606. <https://doi.org/10.1016/j.molp.2019.04.001>.
- Xie, T., Zheng, J.F., Liu, S., Peng, C., Zhou, Y.M., Yang, Q.Y., and Zhang, H.Y. (2015). De novo plant genome assembly based on chromatin interactions: a case study of *Arabidopsis thaliana*. *Mol. Plant* **8**:489–492. <https://doi.org/10.1016/j.molp.2014.12.015>.
- Xu, L., Zhou, J., Liang, J., Cui, H., Tao, M., Zhihui, T., Zhu, Z., and Lin, H. (2014). The remediation potential of *Pennisetum* sp on Cu, Cd contaminated soil. *Acta Ecol. Sin.* **34**:5342–5348.
- Xu, Z., and Wang, H. (2007). LTR\_FINDER: an efficient tool for the prediction of full-length LTR retrotransposons. *Nucleic Acids Res.* **35**:W265–W268. <https://doi.org/10.1093/nar/gkm286>.
- Yan, H., Kikuchi, S., Neumann, P., Zhang, W., Wu, Y., Chen, F., and Jiang, J. (2010). Genome-wide mapping of cytosine methylation revealed dynamic DNA methylation patterns associated with genes and centromeres in rice. *Plant J.* **63**:353–365. <https://doi.org/10.1111/j.1365-3113X.2010.04246.x>.
- Yan, Q., Wu, F., Xu, P., Sun, Z., Li, J., Gao, L., Lu, L., Chen, D., Muktar, M., Jones, C., et al. (2021). The elephant grass (*Cenchrus purpureus*) genome provides insights into anthocyanidin accumulation and fast growth. *Mol. Ecol. Resour.* **21**:526–542. <https://doi.org/10.1111/1755-0998.13271>.
- Yang, J., Liu, D., Wang, X., Ji, C., Cheng, F., Liu, B., Hu, Z., Chen, S., Pental, D., Ju, Y., et al. (2016). The genome sequence of allopolyploid *Brassica juncea* and analysis of differential homoeolog gene expression influencing selection. *Nat. Genet.* **48**:1225–1232. <https://doi.org/10.1038/ng.3657>.
- Yang, X., Zhao, H., Zhang, T., Zeng, Z., Zhang, P., Zhu, B., Han, Y., Braz, G.T., Casler, M.D., Schmutz, J., and Jiang, J. (2018).

- Amplification and adaptation of centromeric repeats in polyploid switchgrass species. *New Phytol.* **218**:1645–1657. <https://doi.org/10.1111/nph.15098>.
- Ye, J., Lin, J., Jiao, W., and Chen, B.** (2015). RAPD Analysis of Genetic Diversity of *Pennisetum* Germplasm Resources in Fujian (Guangdong Agricultural Sciences), pp. 128–132.
- Zhang, H., Koblížková, A., Wang, K., Gong, Z., Oliveira, L., Torres, G.A., Wu, Y., Zhang, W., Novák, P., Buell, C.R., et al.** (2014). Boom-bust turnovers of megabase-sized centromeric DNA in Solanum species: rapid evolution of DNA sequences associated with centromeres. *Plant Cell* **26**:1436–1447. <https://doi.org/10.1105/tpc.114.123877>.
- Zhang, Q., Qi, Y., Pan, H., Tang, H., Wang, G., Hua, X., Wang, Y., Lin, L., Li, Z., Li, Y., et al.** (2022a). Genomic insights into the recent chromosome reduction of autopolyploid sugarcane *Saccharum spontaneum*. *Nat. Genet.* **54**:885–896. <https://doi.org/10.1038/s41588-022-01084-1>.
- Zhang, S., Xia, Z., Li, C., Wang, X., Lu, X., Zhang, W., Ma, H., Zhou, X., Zhang, W., Zhu, T., et al.** (2022b). Chromosome-scale genome assembly provides insights into speciation of allotetraploid and massive biomass accumulation of elephant grass (*Pennisetum purpureum* Schum.). *Mol. Ecol. Resour.* **22**:2363–2378. <https://doi.org/10.1111/1755-0998.13612>.
- Zhang, W., Lee, H.R., Koo, D.H., and Jiang, J.** (2008). Epigenetic modification of centromeric chromatin: hypomethylation of DNA sequences in the CENH3-associated chromatin in *Arabidopsis thaliana* and maize. *Plant Cell* **20**:25–34. <https://doi.org/10.1105/tpc.107.057083>.
- Zhong, C.X., Marshall, J.B., Topp, C., Mroczek, R., Kato, A., Nagaki, K., Birchler, J.A., Jiang, J., and Dawe, R.K.** (2002). Centromeric retroelements and satellites interact with maize kinetochore protein CENH3. *Plant Cell* **14**:2825–2836. <https://doi.org/10.1105/tpc.006106>.
- Zhou, J., Chen, S., Shi, W., David-Schwartz, R., Li, S., Yang, F., and Lin, Z.** (2021). Transcriptome profiling reveals the effects of drought tolerance in Giant Juncao. *BMC Plant Biol.* **21**:2. <https://doi.org/10.1186/s12870-020-02785-7>.
- Zhu, D., Wang, P., Lin, X., Lin, H., Su, D., and Lin, Z.** (2015). Karyotype analysis of *Pennisetum purpureum* and *Pennisetum giganteum*. *Guizhou Agricultural Sciences* **43**:14–18.

**Supplemental information**

**A near-complete genome assembly of the allotetraploid *Cenchrus fungigraminus* (JUJUNCAO) provides insights into its evolution and C4 photosynthesis**

**Huakun Zheng, Baiyu Wang, Xiuting Hua, Ruiting Gao, Yuhao Wang, Zixin Zhang, Yixing Zhang, Jing Mei, Yongji Huang, Yumin Huang, Hui Lin, Xingtian Zhang, Dongmei Lin, Siren Lan, Zhongjian Liu, Guodong Lu, Zonghua Wang, Ray Ming, Jisen Zhang, and Zhanxi Lin**

**A near complete allotetraploid genome provides insight into the evolution and C<sub>4</sub> photosynthesis characteristics of *Cenchrus fungigraminus* (JUNCAO)**

Huakun Zheng<sup>1,#</sup>, Baiyu Wang<sup>2,3,#</sup>, Xiuting, Hua<sup>2,#</sup>, Ruiting Gao<sup>2,#</sup>, Yuhao Wang<sup>3,#</sup>, Zixin Zhang<sup>1</sup>, Yixing Zhang<sup>3</sup>, Jing Mei<sup>3</sup>, Yongji Huang<sup>4</sup>, Yumin Huang<sup>3</sup>, Hui Lin<sup>1</sup>, Xingtian Zhang<sup>3</sup>, Dongmei Lin<sup>1</sup>, Siren Lan<sup>1</sup>, Zhongjian Liu<sup>1</sup>, Guodong Lu<sup>1</sup>, Zonghua Wang<sup>1,\*</sup>, Ray Ming<sup>3,\*</sup>, Jisen Zhang<sup>2,\*</sup>, Zhanxi Lin<sup>1,\*</sup>

<sup>1</sup>National Engineering Research Center of JUNCAO Technology, College of Life Science, Fujian Agriculture and Forestry University, Fuzhou 350002, China.

<sup>2</sup> State Key Laboratory for Conservation and Utilization of Subtropical Agro-bioresources, Guangxi Key Laboratory of Sugarcane Biology, Guangxi University, Guangxi 530004, China.

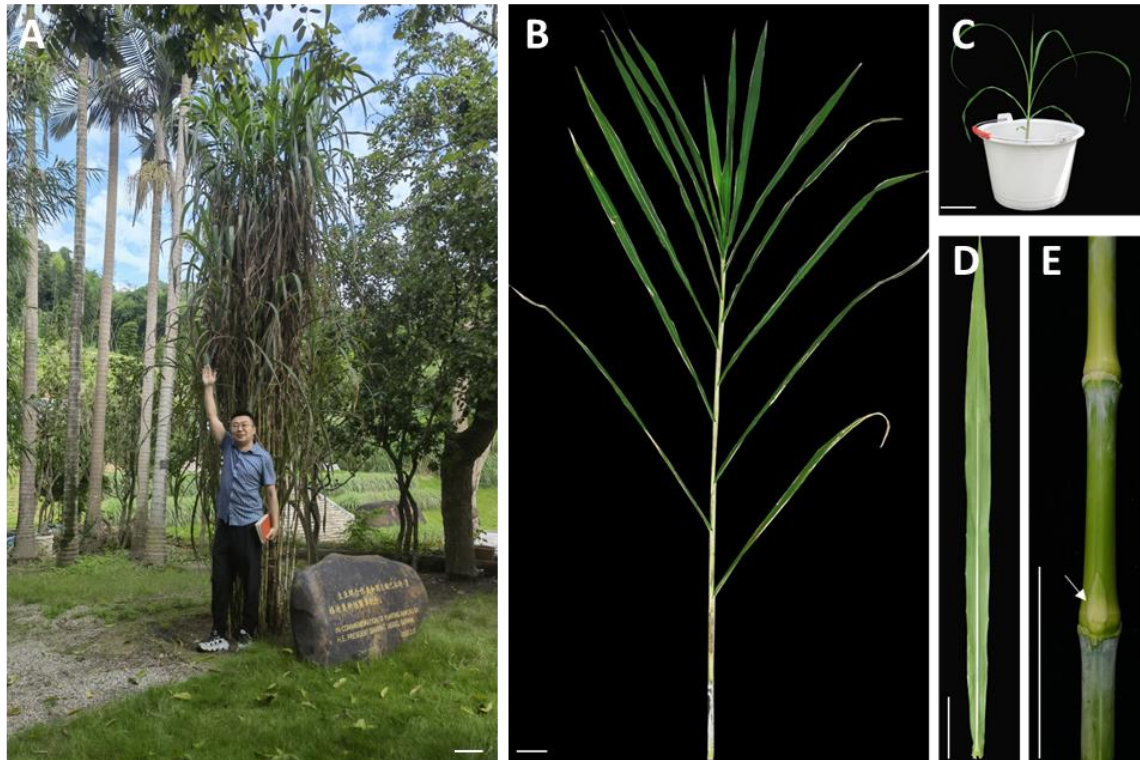
<sup>3</sup> Center for Genomics, Fujian Provincial Key Laboratory of Haixia Applied Plant Systems Biology, Key Laboratory of Genetics, Breeding and Multiple Utilization of Corps, Ministry of Education, Fujian Agriculture and Forestry University, Fuzhou 350002, China.

<sup>4</sup>Fuzhou Institute of Oceanography, Minjiang University, Fuzhou 350108, China.

#These authors contributed to this work equally.

\*To whom correspondence should be addressed. Email:

lzxjuncao@163.com ; zjisen@fafu.edu.cn; rayming@illinois.edu;  
zonghuaw@163.com.

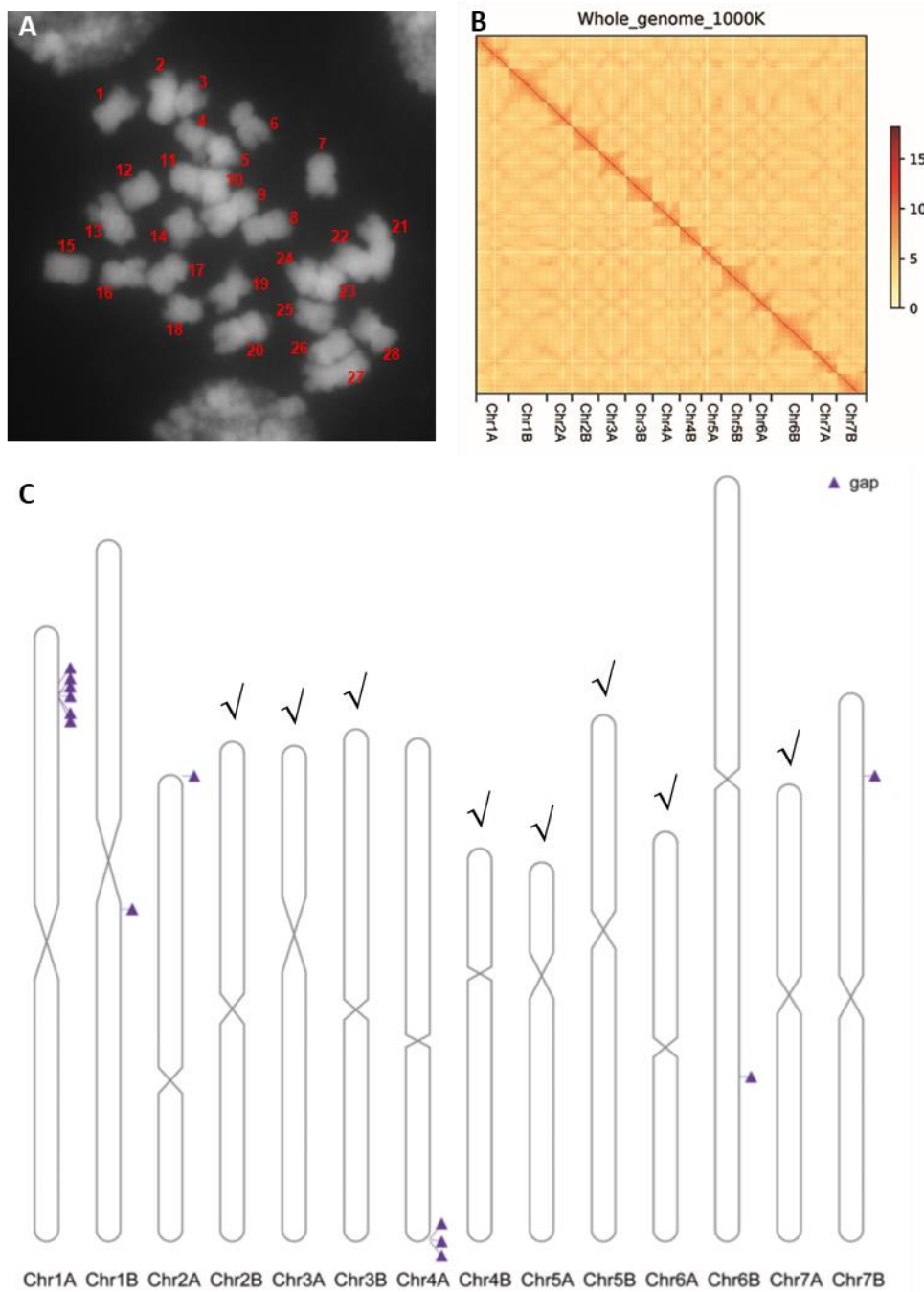


**Figure S1. *Cenchrus fungigraminus* (JUJUNCAO).**

**(A)** The JUJUNCAO plants in the field. Scale bar: 30 cm.

**(B)** Model of JUJUNCAO. Scale bar: 10 cm.

**(C-E)** Seedling (C), leaf (D) and stem (E) of JUJUNCAO. The arrow in (E) indicates the bud on a stem. Scale bar: 10 cm.

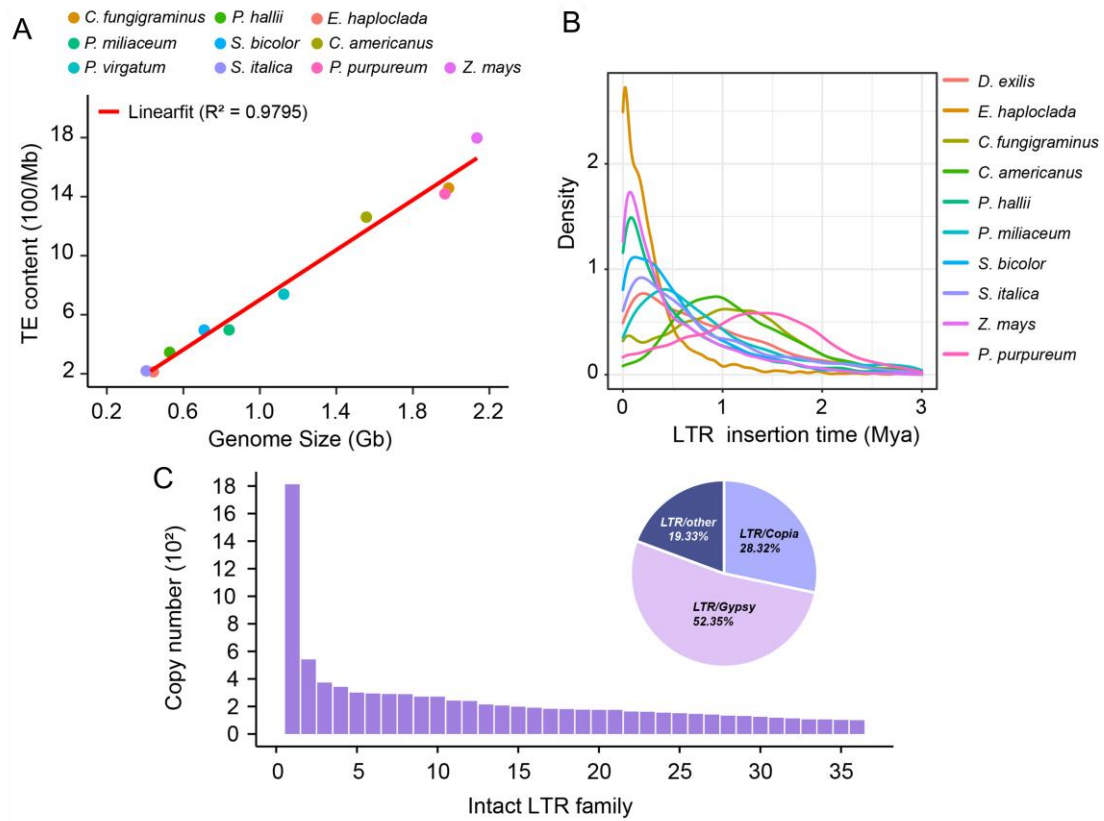


**Figure S2. The JUJUNCAO genome consists of 14 chromosomes.**

**(A)** The microscopic karyotype of JUJUNCAO.

**(B)** The Hi-C chromatin interaction map for the 14 pseudo-chromosome of JUJUNCAO genome.

**(C)** The 14 assembled chromosomes of the JUJUNCAO genome. “✓” indicates the telomere to telomere chromosomes; “▲” indicates the gap on assembled chromosomes.

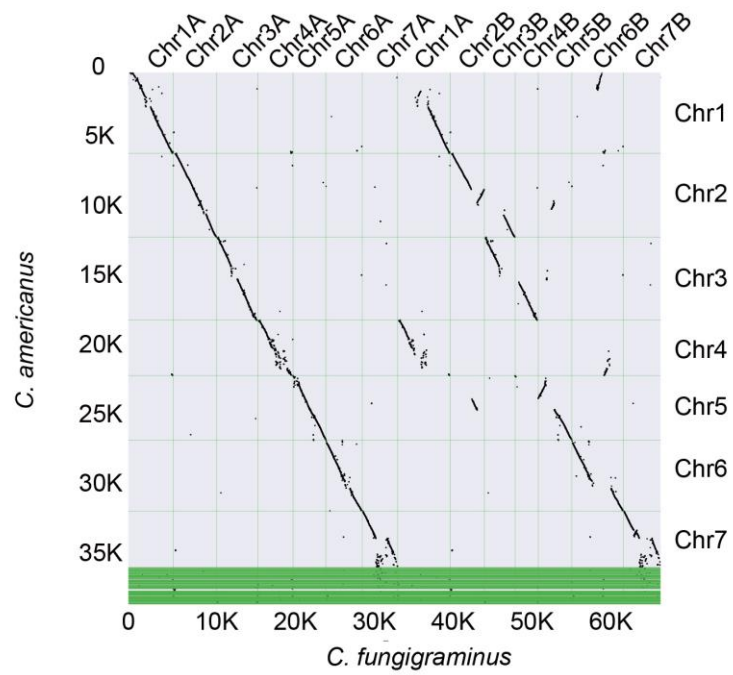
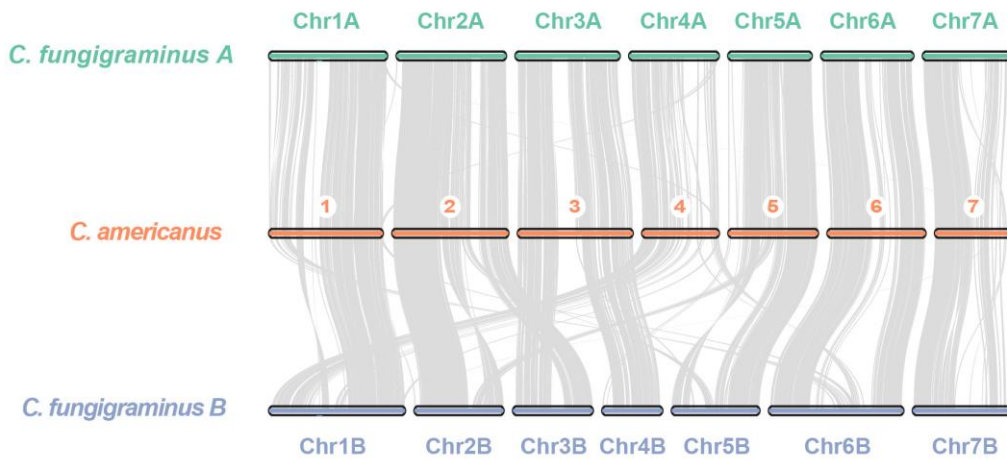


**Figure S3. Characteristics of TE sequences in JUJUNCAO genome.**

**(A)** Genome size expansion is highly correlated with TE amplification bursts ( $R^2 = 0.9795$ ). The red line shows the linear relationship between genome size and TE content.

**(B)** Intact LTR retrotransposon insertion time of *D. exilis*, *E. haploclada*, *C. fungigraminus*, *C. americanus*, *P. hallii*, *P. miliaceum*, *S. bicolor*, *S. italica*, *Z. mays*, *P. purpureum*, X-axis represents LTR retrotransposon insertion time, Y-axis represents density.

**(C)** Classification of intact LTR retrotransposons in the JUJUNCAO genome, LTR families with more than 100 copies are shown.

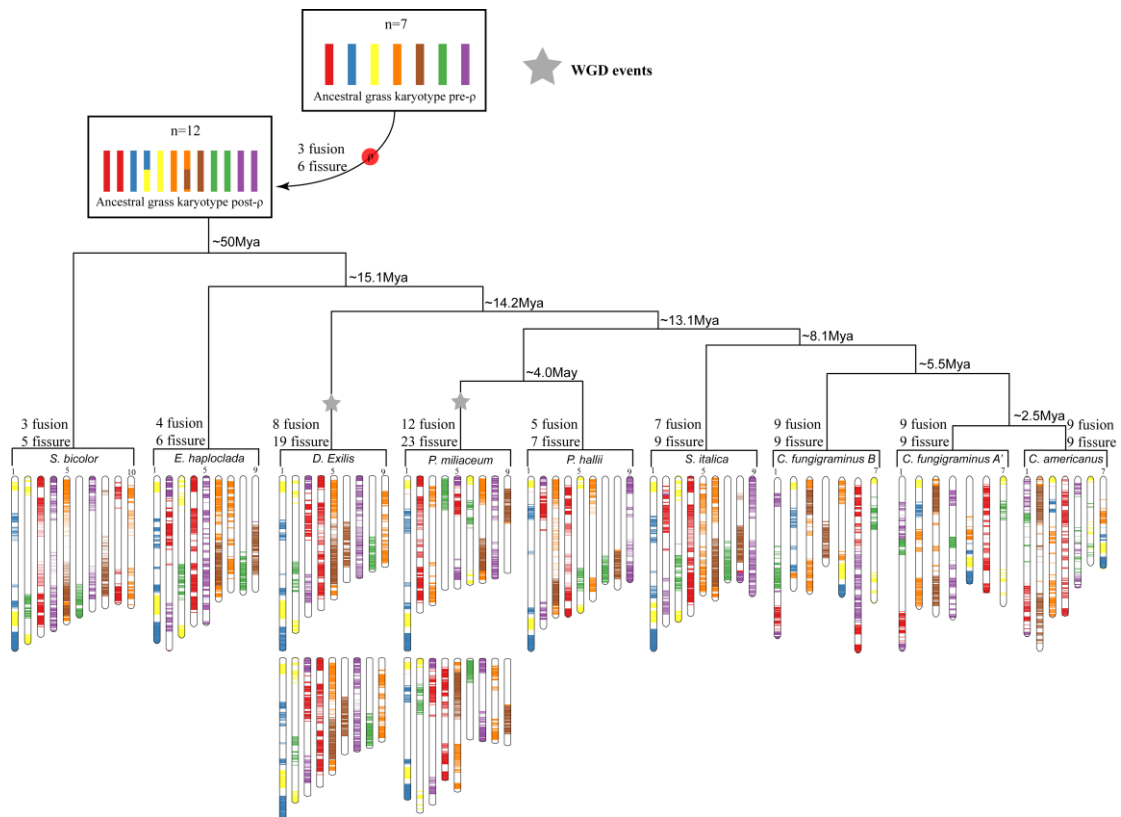
**A****B**

**Figure S4. Comparative genomic analysis of the A and B subgenomes of *C. fungigraminus* and *C. americanus*.**

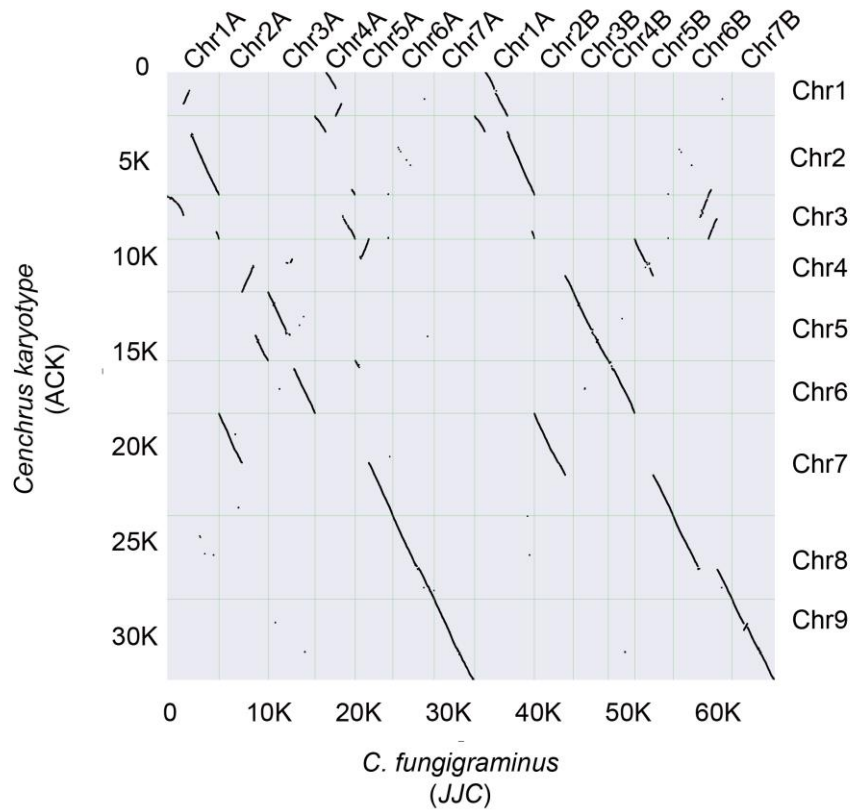
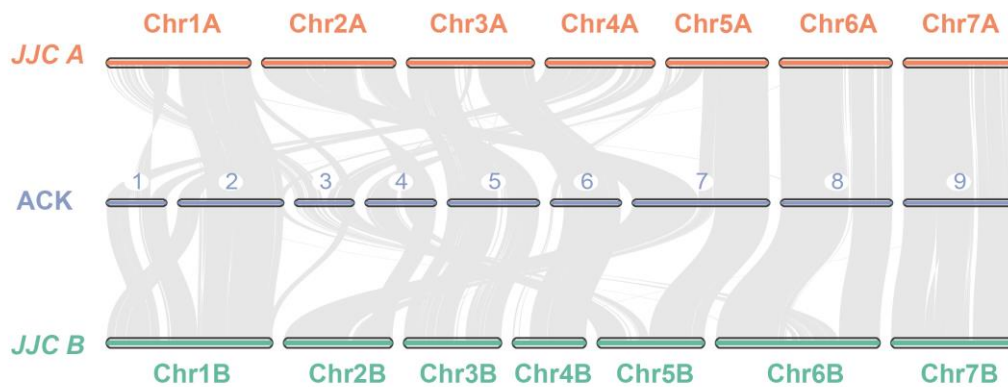
(A) Dot-plot of alignments.

(B) Syntenic blocks





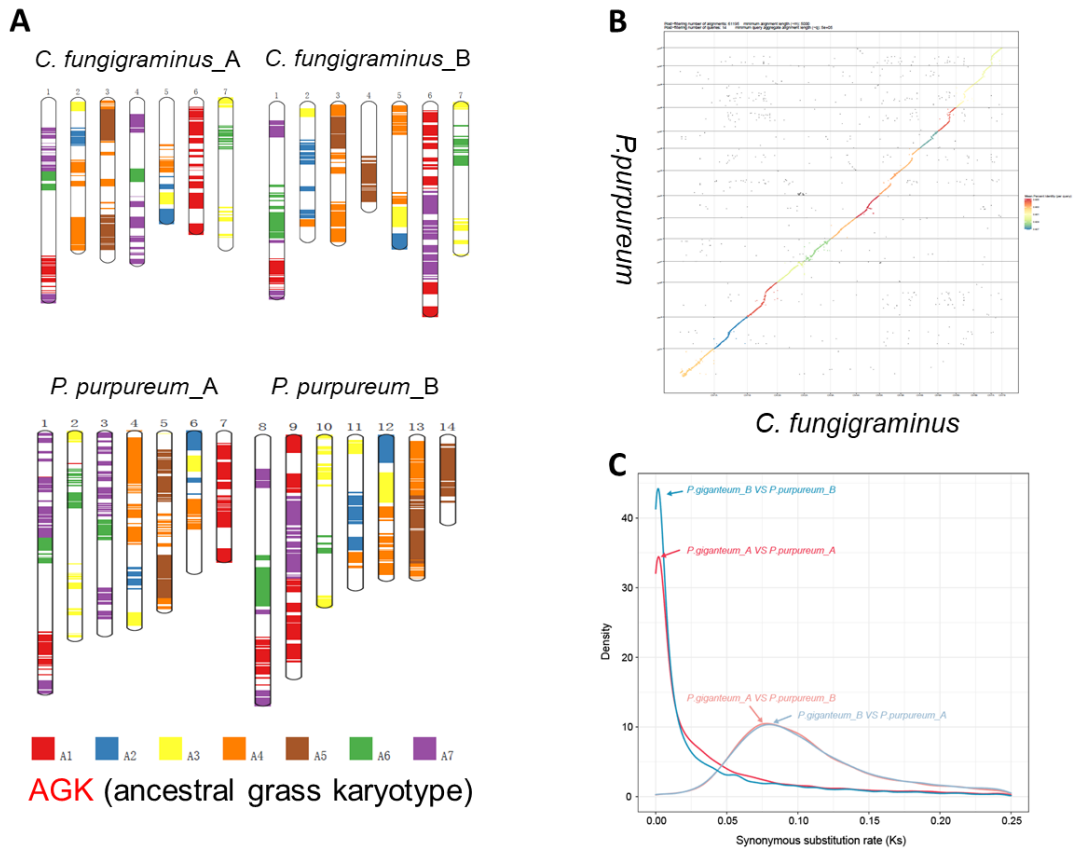
**Figure S5. Chromosome evolution of JUJUNCAO and species of Panicoideae.** Evolutionary scenario of chromosomes in *Z. mays*, *S. bicolor*, *E. haploclad*, *D. exilis*, *P. miliaceum*, *P. hallii*, *S. italica*, *C. americanus*, *C. fungigraminus* A and B, from the seven chromosomes of ancestral grass karyotype (AGK) pre-p and the 12 protochromosomes of the AGK post-p. Polyploidization events are indicated by stars, along with shuffling events (fusions and fissions).

**A****B**

**Figure S6. Comparative genomic analysis of the A and B subgenomes of *C. fungigraminus* and ancestral *Cenchrus* karyotype (ACK).**

(A) Dot-plot of alignments.

(B) Syntenic blocks

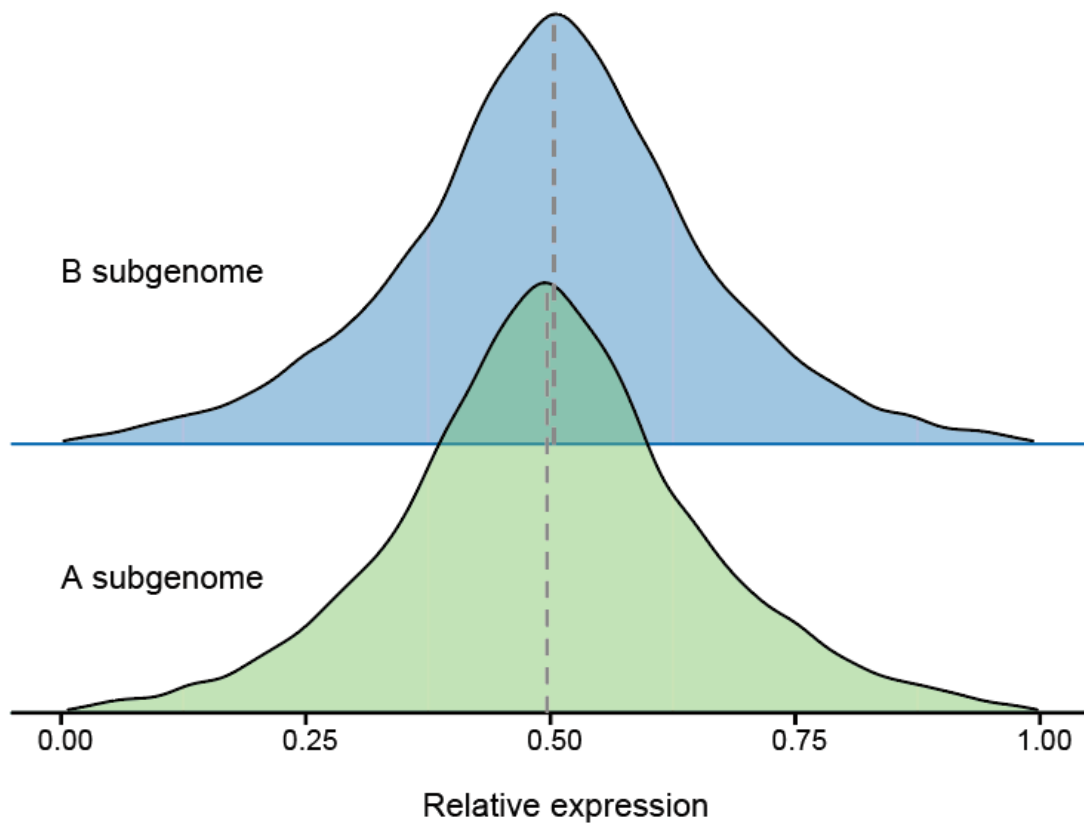


**Figure S7. The chromosome structure and evolution between JUJUNCAO and *P. purpureum*.**

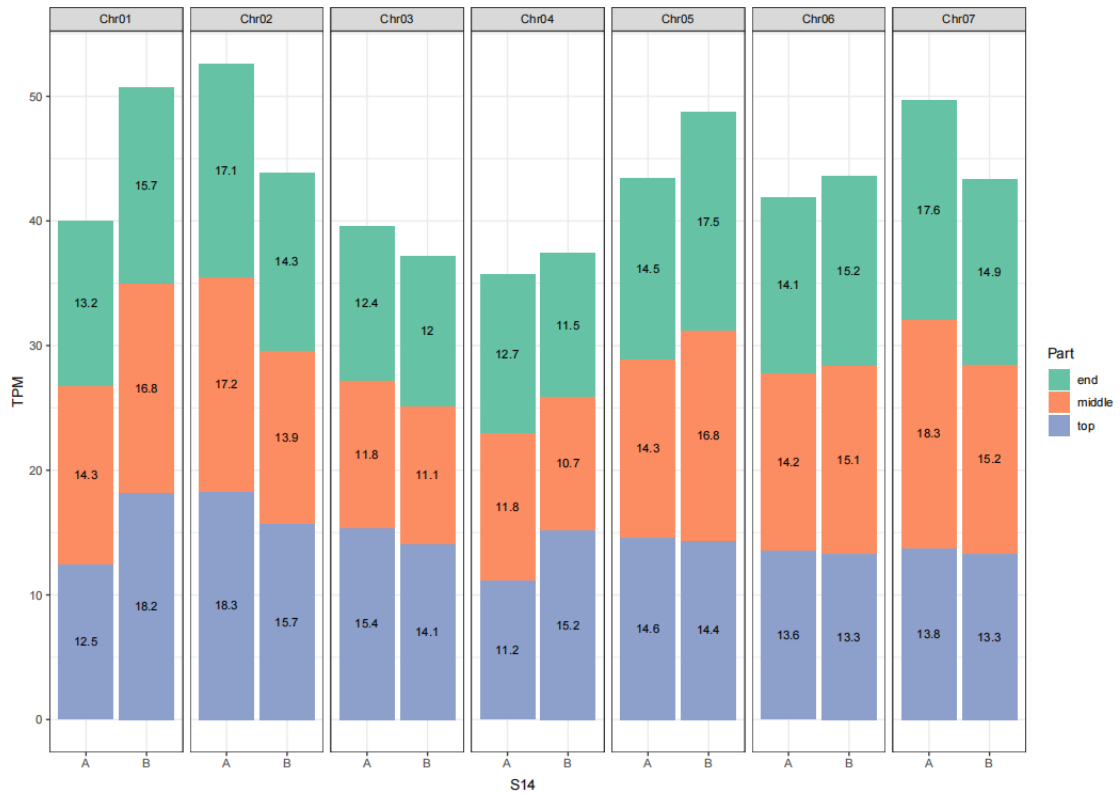
(A) Comparison of chromosome structure between JUJUNCAO and *P. purpureum*.

(B) Graphical alignment of JUJUNCAO chromosomes with *P. purpureum* chromosomes.

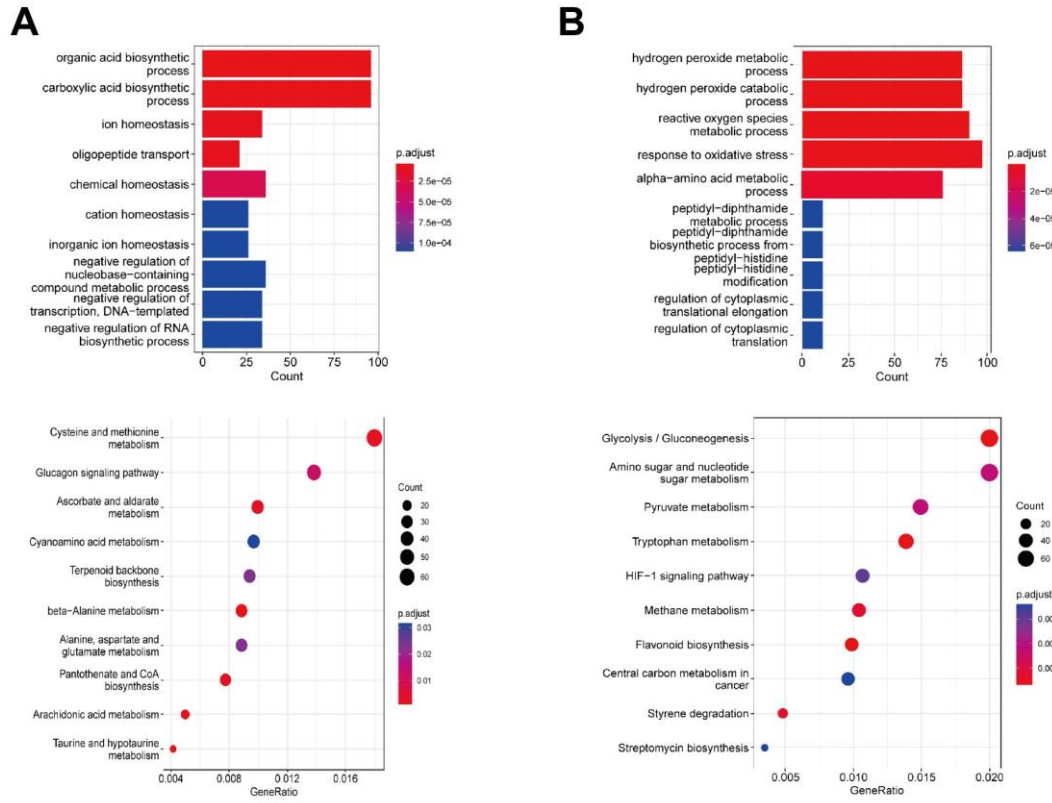
(C) Distributions of Ks between JUJUNCAO and *P. purpureum*.



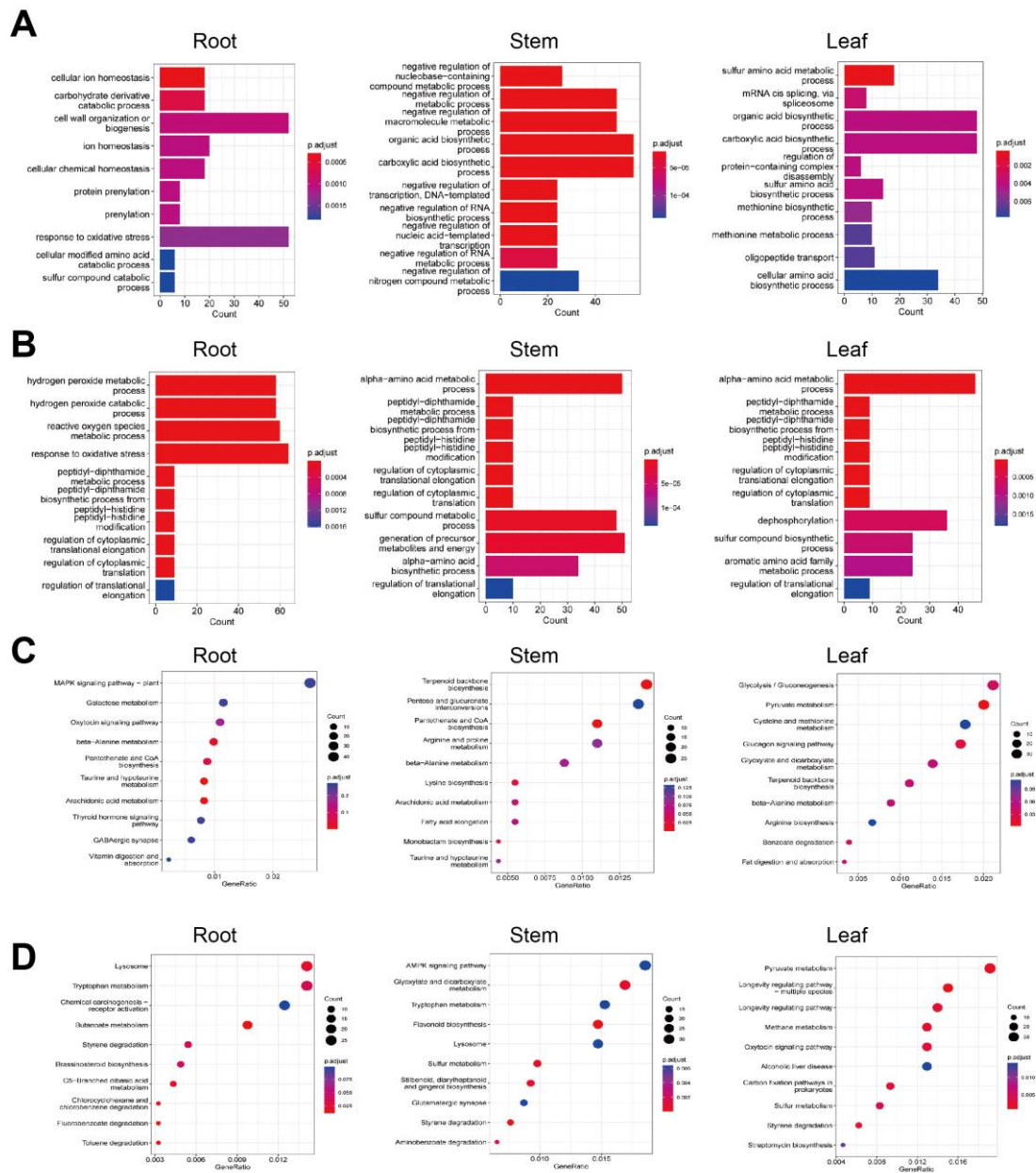
**Figure S8. Comparison of the overall expression level and natural selection feature between the two subgenomes.** The overall expression level between the two subgenomes of JUJUNCAO.



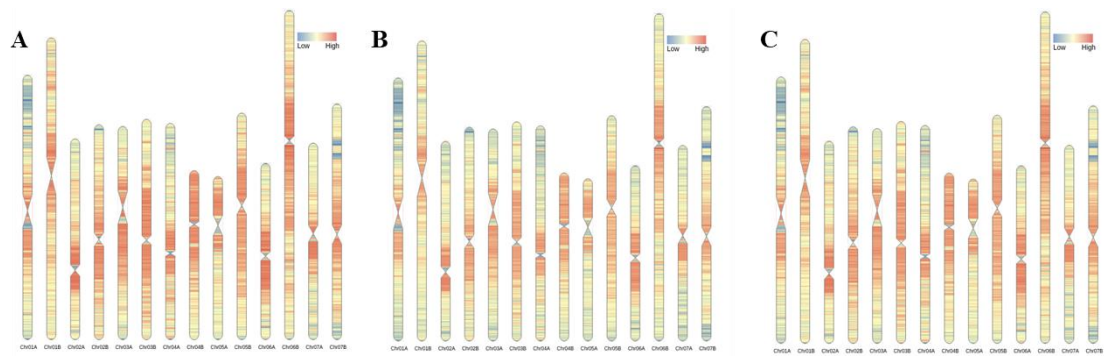
**Figure S9. The distribution of differentially expressed homeologous genes between A and B subgenomes. TPM (Transcript per million).**



**Figure S10. GO (top) and KEGG (bottom) enrichment of subgenome A-biased genes (A) and subgenome B-biased genes (B).**



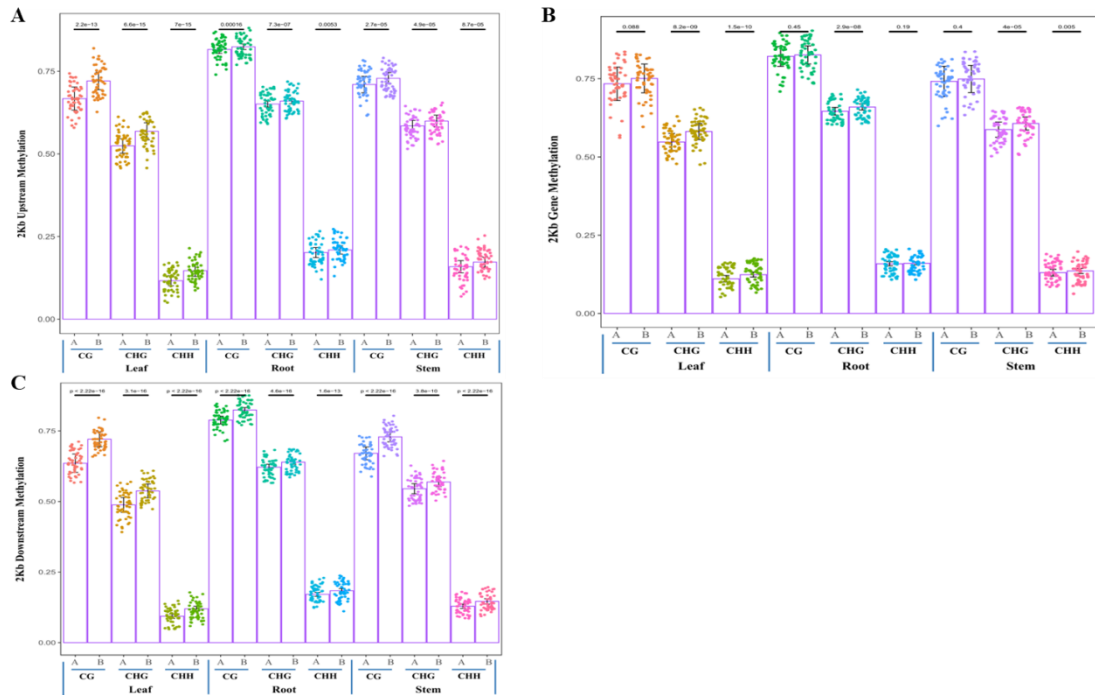
**Figure S11. GO (A,B) and KEGG (C,D) enrichment of A subgenome-biased genes (A,C) and B subgenome biased genes (B,D).**



**Figure S12. Distribution of methylation through the chromosomes from different tissues**

- (A) The distribution of all methylation sites on chromosomes in the leaf.
- (B) The distribution of all methylation sites on chromosomes in the stem.
- (C) The distribution of all methylation sites on chromosomes in the root.



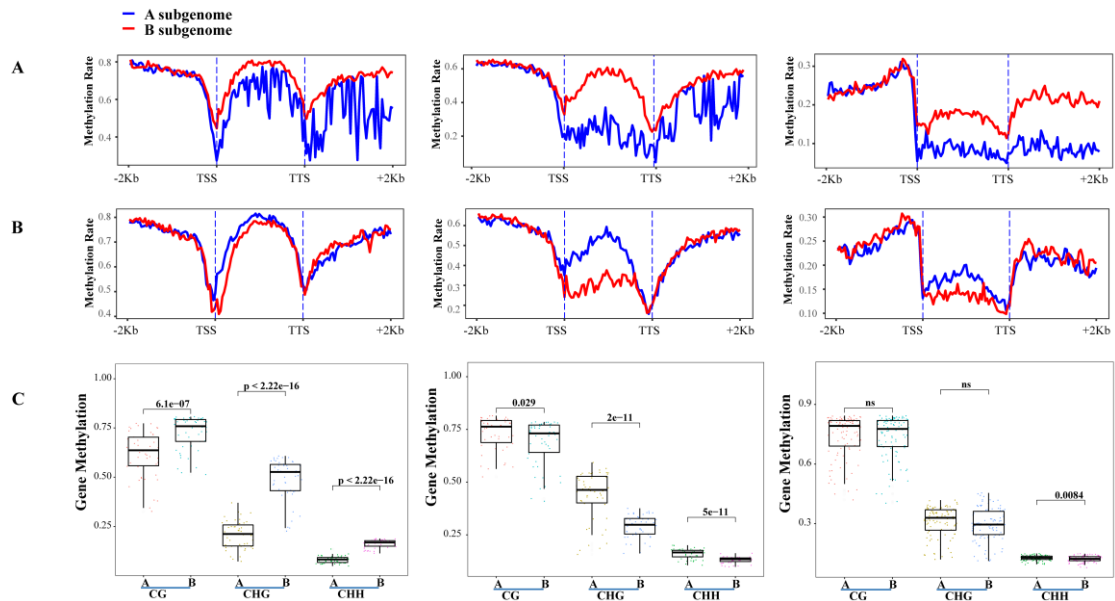


**Figure S13. The comparison of different methylation types of subgenomes A and B in different tissues.** The 2 Kb upstream, gene-body and 2 Kb downstream of A and B subgenomic genes were divided into 50 bins on average. Each point represents the average methylation ratio in a bin.

(A) Comparison of 2 Kb upstream methylation ratio of the genes of A and B subgenomes.

(B) Comparison of gene-body methylation ratio of the genes of A and B subgenomes.

(C) Comparison of 2 Kb downstream methylation ratio of the genes of A and B subgenomes.

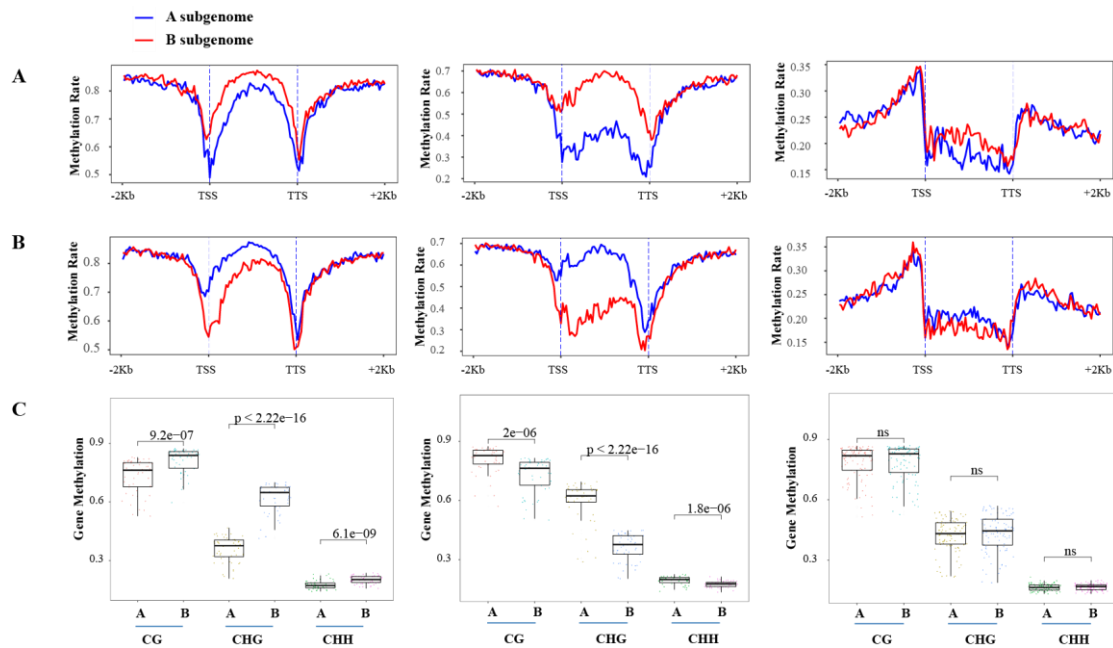


**Figure S14. DNA methylation divergence of the allotetraploid in the leaf.**

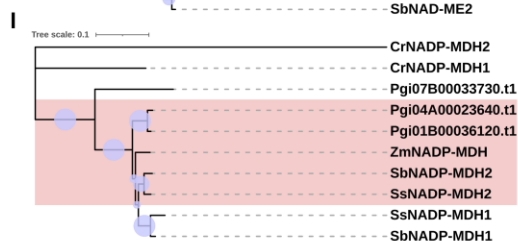
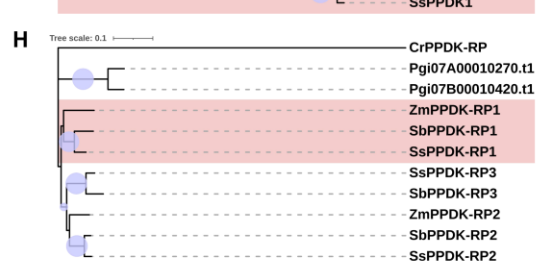
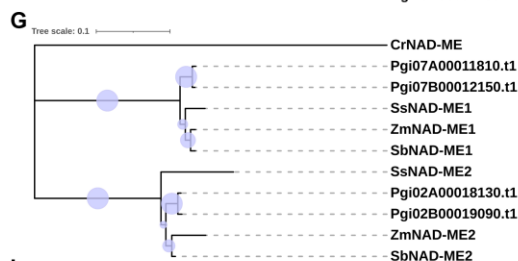
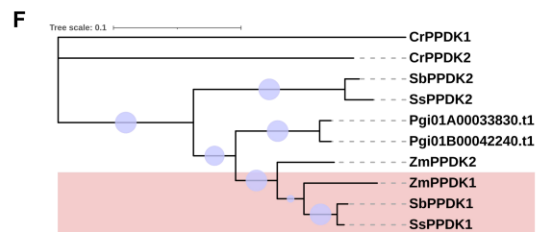
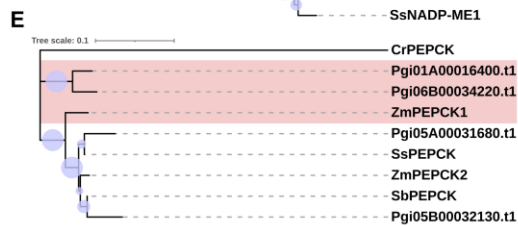
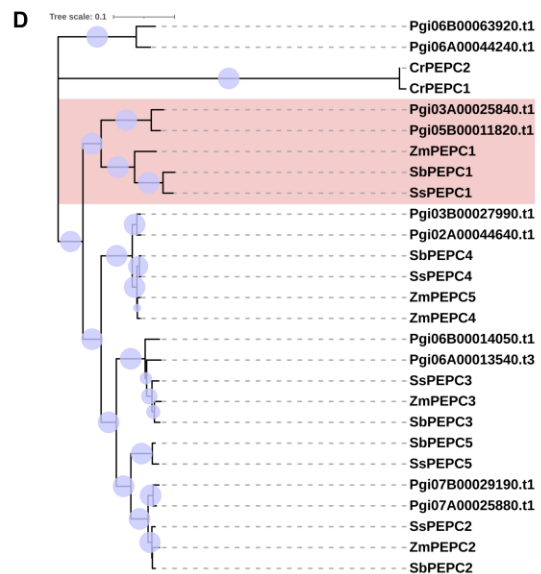
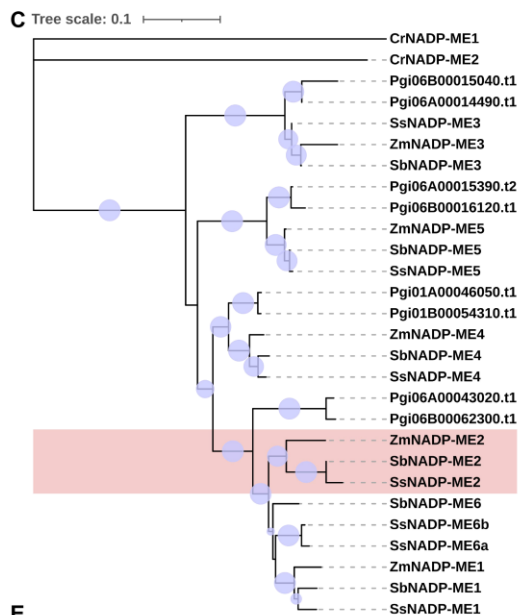
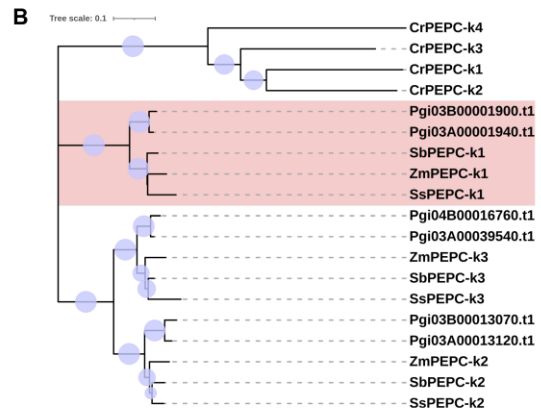
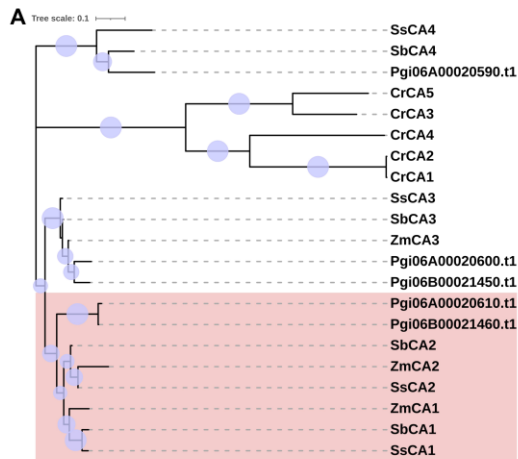
(A) The methylation rates of subgenome A-biased homeologous gene pairs (Fold>2) between the two subgenomes in the leaf.

(B) The methylation rates of subgenome B-biased homeologous gene pairs (Fold>2) between the two subgenomes in the leaf.

(C) Comparison of methylation rate in gene-body of A and B subgenomes in the leaf. The left panel shows the methylation rate of subgenome A-biased homeologous gene pairs between the two subgenomes (fold >2); The middle panel shows the methylation rate of subgenome B-biased homeologous gene pairs between the two subgenomes; The right panel shows the methylation rate of inconspicuously differentially expressed gene pairs (with 0-2 times changes) from the two subgenomes.



**Figure S15. DNA methylation divergence of the allotetraploid in the root.**  
 (A) The methylation rates of subgenome A-biased homeologous gene pairs (Fold>2) between the two subgenomes in the root.  
 (B) The methylation rates of subgenome B-biased homeologous gene pairs (Fold>2) between the two subgenomes in the root  
 (C) Comparison of methylation rate in gene-body of A and B subgenomes in the root. The left panel shows the methylation rate of subgenome A-biased homeologous gene pairs between the two subgenomes (fold >2); The middle panel shows the methylation rate of subgenome B-biased homeologous gene pairs between the two subgenomes; The right panel shows the methylation rate of inconspicuous differentially expressed gene pairs (with 0-2 times changes) from the two subgenomes.



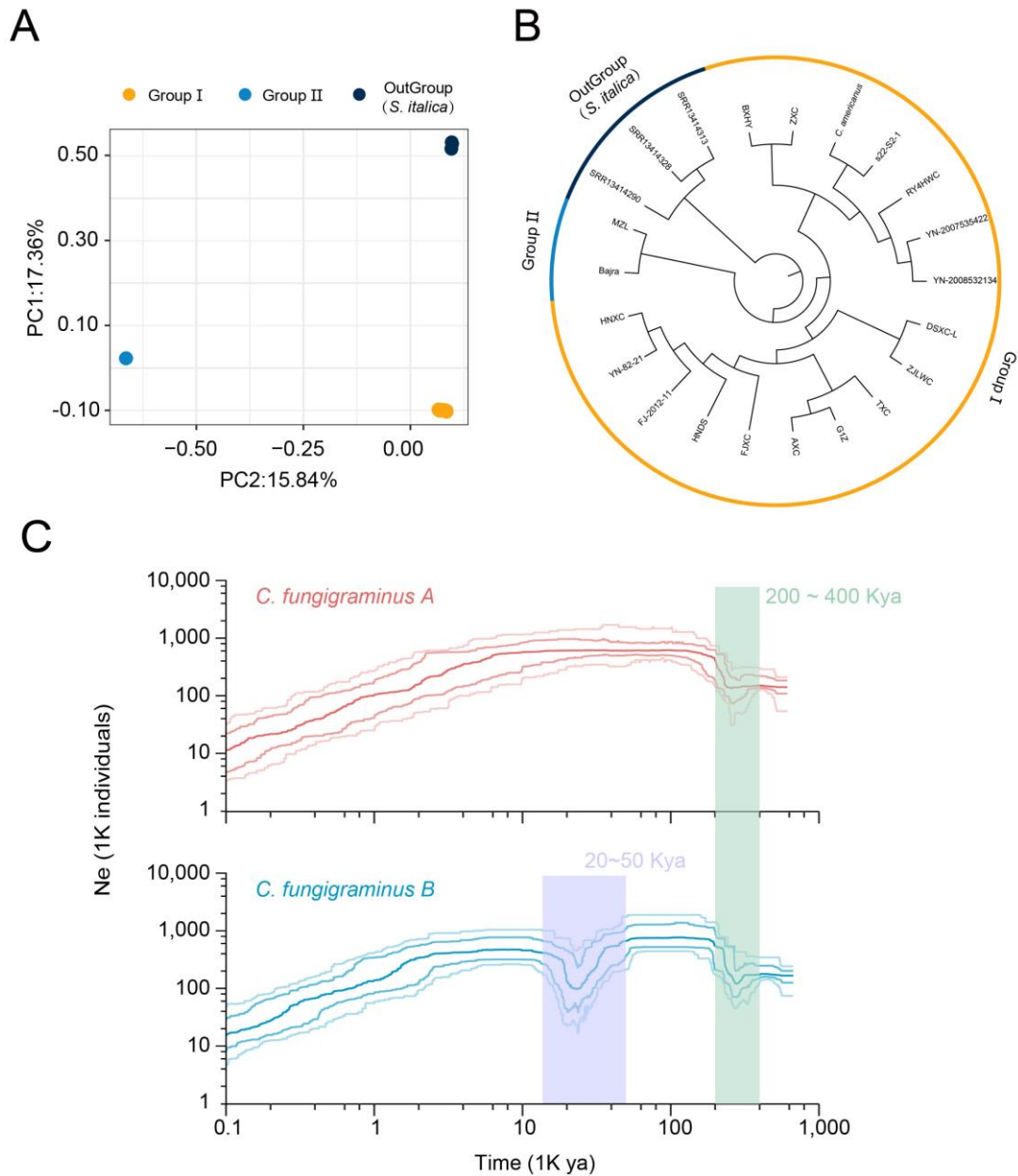
**Figure S16. Phylogenetic analysis of genes encoding C<sub>4</sub> enzymes and their non-C<sub>4</sub> isoforms.** Cf indicates *C. fungigraminus*, Ss indicates *Saccharum spontaneum*, Sb indicates *Sorghum bicolor*, Zm indicates *Zea mays*, Os indicates *Oryza sativa*, and Cr indicates *Chlamydomonas reinhardtii*. Purple circles indicate bootstraps. Genes in the pink clade encode proteins in the C<sub>4</sub> pathway. Tree scale: 0.1.

Phylogenetic analysis of CA **(A)**; PEPC-k **(B)**; NADP-ME **(C)**; PEPC **(D)**; PEPCK **(E)**; PPK (F); NAD-ME **(G)**; PPK-RP **(H)**; and NADP-MDH**(I)** proteins.

	root	stem	leaf
CfCA1_A	1	17	71
CfCA1_B	11	28	196
CfCA3_A	9	13	5
CfCA3_B	8	51	9
CfCA4_A	0	0	0
CfNAD-ME1_A	4	3	3
CfNAD-ME1_B	4	5	6
CfNAD-ME2_A	3	3	2
CfNAD-ME2_B	6	4	1
CfNADP-MDH2_A	0	13	34
CfNADP-MDH2_B	0	8	32
CfNADP-MDH1_B	0	8	34
CfNADP-ME2_A	4	11	20
CfNADP-ME2_B	4	16	34
CfNADP-ME3_A	0	0	0
CfNADP-ME3_B	0	0	0
CfNADP-ME4_A	3	0	0
CfNADP-ME4_B	8	3	0
CfNADP-ME5_A	12	3	2
CfNADP-ME5_B	15	13	20
CfPEPC1_A	0	69	476
CfPEPC1_B	0	137	769
CfPEPC2_A	18	15	7
CfPEPC2_B	67	38	38
CfPEPC3_A	11	9	6
CfPEPC3_B	17	13	23
CfPEPC4_A	66	20	9
CfPEPC4_B	63	23	16
CfPEPC5_A	0	28	207
CfPEPC5_B	0	0	15
CfPEPC-k1_A	12	10	21
CfPEPC-k1_B	14	36	21
CfPEPC-k2_A	86	138	459
CfPEPC-k2_B	39	36	159
CfPEPC-k3_A	26	127	112
CfPEPC-k3_B	50	145	100
CfPPDK1_A	0	26	203
CfPPDK1_B	0	25	252
CfPPDK-RP1_A	2	19	79
CfPPDK-RP1_B	7	68	236
CfPEPCK1_A	0	1	43
CfPEPCK1_B	1	1	8
CfPEPCK2_A	8	1	0
CfPEPCK2_B	5	1	2



**Figure S17. Expression pattern of carbon fixation genes in root, stem and leaf tissue of JUJUNCAO.**



**Figure S18. Population characteristics and evolution of JUJUNCAO.**

**(A)** Principal component analysis with PC1 and PC2.

**(B)** The phylogenetic tree was constructed based on genome-wide SNP, with different colors representing different groups.

**(C)** Historical effective population sizes ( $N_e$ ) for *Cenchrus* population estimated using *C. fungigraminus A* and *B* as the ancestor genome. Nucleotide substitution rate  $\mu = 6.5e-9$  and generation time  $gt = 1$  were used for analysis.

**Table S1. Estimation of the JUJUNCAO genome using flow cytometry**

<b>Sample</b>	<b>Reference (Ref)</b>	<b>Signal of Ref</b>	<b>Signal of Sample</b>	<b>Ratio</b>	<b>Estimated Genome size (Gb)</b>
JUJUNCAO-1	Tomato	24.72	58.76	2.38	2.09
JUJUNCAO-2	Tomato	25.23	59.86	2.37	2.09
JUJUNCAO-3	Tomato	25.25	58.81	2.33	2.05



**Table S2. Statistics of Hi-C mapping of the JUJUNCAO genome**

<b>Statistics of mapping</b>	
Clean Paired-end Reads	815,147,702
Clean Paired-end Reads Rates (%)	100%
Unmapped Paired-end Reads	58,095,120
Unmapped Paired-end Reads Rates	7.12%
Paired-end Reads with singleton	293,148,296
Paired-end Reads with singleton Rates (%)	35.96%
Multi Paired-end Reads	0
Multi Paired-end Reads Rates	0.00%
Unique Paired-end Reads	222,332,443
Unique Paired-end Reads Rates (%)	27.27%
<b>Statistics of valid reads</b>	
Unique Paired-end Reads	222,332,443
Dangling End Paired-end Reads	35,538,486
Dangling End Paired-end Reads Rates (%)	15.98%
Self Circle Paired-end Reads	67,508
Self Circle Paired-end Reads Rates (%)	0.03%
Dumped Paired-end Reads	12,745
Dumped Paired-end Reads Rates (%)	0.01%
Valid interaction Paired-end Reads	174,777,070
Valid interaction Paired-end Reads Rate (%)	78.61%
Lib Valid interaction Paired-end Reads	166,760,242
Lib Valid interaction Paired-end Reads Rate (%)	75.00%
Lib Dup (%)	25.00%

**Table S3. BUSCO analysis of the completeness of the *C. fungigraminus* genome.**

<b>Description</b>	<b>Number</b>	<b>Percentage</b>
<b>Complete BUSCOs (C)</b>	<b>1588</b>	<b>98.4%</b>
<b>single-copy BUSCOs (S)</b>	<b>322</b>	<b>20.0%</b>
<b>duplicated BUSCOs (D)</b>	<b>1266</b>	<b>78.4%</b>
<b>Fragmented BUSCOs (F)</b>	<b>7</b>	<b>0.4%</b>
<b>Missing BUSCOs (M)</b>	<b>17</b>	<b>1.2%</b>
<b>Total BUSCO groups searched</b>	<b>1614</b>	<b>100.0%</b>

**Table S4. Summary of assembly, and annotation statistics for four *Cenchrus* Species**

<b>Species</b>	<b><i>C. fungigraminus</i></b>	<b><i>C. americanus</i></b>	<b><i>C. purpureus</i> cv. <i>Purple</i></b>	<b><i>P. purpureus</i></b>
Ploidy	2n=4x=28	2n=2x=14	2n=4x=28	2n=4x=28
Assembly size (Gb)	1.98	1.58	1.97	2.07
Number of contig	815	208,873	2,059	1956
Contig N50 (bp)	134,074,806	18,180	1,829,308	2,900,000
Longest contig (bp)	162,376,048	282,900	15,071,384	20,190,000
BUSCO assessment	98.40%	-	97.83%	97.80%
Percentage of repeat elements	73.38%	77.2%	66.32%	60.47%
GC content	47.11%	47.90%	46.95%	46.95

**Table S5. BUSCO analysis of the annotation of the *C. fungigraminus* genome.**

<b>Description</b>	<b>Number</b>	<b>Percentage</b>
<b>Complete BUSCOs (C)</b>	<b>1501</b>	<b>93.0%</b>
<b>single-copy BUSCOs (S)</b>	<b>212</b>	<b>13.1%</b>
<b>duplicated BUSCOs (D)</b>	<b>1289</b>	<b>79.9%</b>
<b>Fragmented BUSCOs (F)</b>	<b>10</b>	<b>0.6%</b>
<b>Missing BUSCOs (M)</b>	<b>103</b>	<b>6.4%</b>
<b>Total BUSCO groups searched</b>	<b>1614</b>	<b>100.0%</b>

**Table S6. Repeat contents of JUJUNCAO were compared with other grass species**

	<i>Z.mays</i>	<i>S.italica</i>	<i>P. miliaceum</i>	<i>S.bicolor</i>	<i>C. fungigraminus</i>		
					<i>AB</i>	<i>Sub A</i>	<i>Sub B</i>
Total Repeat Fractions	81.75	48.63	55.08	69.67	68.87	66.39	70.75
Class I :	52.50	18.56	23.35	45.59	35.69	33.14	38.45
Retrotransposon							
LTR							
Retrotransposon							
Ty3/Gypsy	31.98	11.60	19.07	38.78	22.40	19.33	25.37
Ty1/Copia	19.86	5.31	3.26	5.70	11.25	11.92	10.86
Other	0	0.07	0.04	0.18	0.26	0.23	0.3
Non-LTR							
Retrotransposon							
LINE	0.66	1.58	0.98	0.93	1.78	1.66	1.92
SINE	0	0	0	0	0	0	0
Class II :	2.43	5.47	1.99	8.68	4.51	4.25	4.79
DNA transposon							
CACTA	1.58	3.37	1.21	3.56	2.73	2.60	2.89
Mutator	0.32	0.79	0.29	0.68	1.27	1.14	1.40
hAT	0.19	0.53	0.23	0.51	0.22	0.23	0.2
PIF/Harbinger	0.18	0.56	0.06	2.79	0.14	0.13	0.15
Helitron	0.16	0.18	0.20	0.28	0.14	0.14	0.14
TcMar-							
Stowaway	0	0.04	0	0	0.01	0.01	0.01
Other	0	0	0	0.86	0	0	0
Tandem Repeat	0	0	0	2.56	0	0	0
Unknown	26.82	24.60	29.74	4.46	28.67	29.00	27.51

**Table S7. The centromeric repeats identified from JUJUNCAO genome**

<b>Chr</b>	<b>Start</b>	<b>End</b>	<b>Length (bp)</b>	<b>Monomer</b>
Chr01A	18,253,776	19,139,182	885,406	<i>CEN156</i>
Chr01A	75,007,454	95,409,560	20,402,106	<i>CEN137, CEN156</i>
Chr01B	75,121,515	98,234,246	23,112,731	<i>CEN156</i>
Chr02A	79,096,377	86,197,020	7,100,643	<i>CEN137</i>
Chr02B	68,090,932	76,337,386	8,246,454	<i>CEN156</i>
Chr03A	40,718,343	61,191,030	20,472,687	<i>CEN137, CEN156</i>
Chr03B	73,050,604	78,621,099	5,570,495	<i>CEN156</i>
Chr04A	80,088,409	83,465,161	3,376,752	<i>CEN137</i>
Chr04B	32,009,996	35,621,487	3,611,491	<i>CEN156</i>
Chr05A	24,377,613	36,892,929	12,515,316	<i>CEN137, CEN156</i>
Chr05A	45,359,011	46,301,987	942,976	<i>CEN137</i>
Chr05B	52,902,785	63,299,058	10,396,273	<i>CEN148, CEN156</i>
Chr06A	55,670,148	61,016,392	5,346,244	<i>CEN156</i>
Chr06B	79,034,047	84,664,505	5,630,458	<i>CEN156</i>
Chr07A	51,781,816	62,044,859	10,263,043	<i>CEN137, CEN156</i>
Chr07B	76,324,440	87,934,803	11,610,363	<i>CEN156</i>

**Table S8. Depth of the cytosine among three sequence contexts**

<b>Tissue</b>	<b>Depth of all CpG</b>	<b>Depth of all CHG</b>	<b>Depth of all CHH</b>
Leaf	16.9217	18.3554	19.0562
Stem	16.9264	18.4428	19.1824
Root	15.1407	16.2506	17.3339

**Table S9. The comparison and de-repetition of BS-seq**

<b>Tissue</b>	<b>Mapping efficiency (%)</b>	<b>De-repetition retention rate (%)</b>
Leaf	69.60	74.46
Stem	68.70	75.09
Root	63.60	75.63



**Table S10. Methylation information of the three tissue**

---

<b>Tissue</b>	<b>methC of all CpG (%)</b>	<b>methC of all CHG (%)</b>	<b>methC of all CHH (%)</b>
Leaf	63	44.3	2.1
Stem	63.5	47.2	2.3
Root	73.3	53	2.4

---

**Table S11. Members of the C<sub>4</sub> related gene**

<b>Name</b>	<b>Gene ID</b>	<b>Abbreviation ID</b>
	Pgi06A00020610	CfCA1_A
	Pgi06B00021460	CfCA1_B
<b>Carbonic anhydrase</b>	Pgi06A00020600	CfCA3_A
	Pgi06B00021450	CfCA3_B
	Pgi06A00020590	CfCA4_A
	Pgi07A00011810	CfNAD-ME1_A
<b>NAD-malic enzyme</b>	Pgi07B00012150	CfNAD-ME1_B
	Pgi02A00018130	CfNAD-ME2_A
	Pgi02B00019090	CfNAD-ME2_B
	Pgi04A00023640	CfNADP-MDH2_A
<b>NADP-malate dehydrogenase</b>	Pgi01B00036120	CfNADP-MDH2_B
	Pgi07B00033730	CfNADP-MDH1_B
	Pgi06A00043020	CfNADP-ME2_A
	Pgi06B00062300	CfNADP-ME2_B
	Pgi06A00014490	CfNADP-ME3_A
<b>NADP-malic enzyme</b>	Pgi06B00015040	CfNADP-ME3_B
	Pgi01A00046050	CfNADP-ME4_A
	Pgi01B00054310	CfNADP-ME4_B
	Pgi06A00015390	CfNADP-ME5_A
	Pgi06B00016120	CfNADP-ME5_B
	Pgi03A00025840	CfPEPC1_A
	Pgi05B00011820	CfPEPC1_B
	Pgi07A00025880	CfPEPC2_A
	Pgi07B00029190	CfPEPC2_B
<b>Phosphoenolpyruvate carboxylase</b>	Pgi06A00013540	CfPEPC3_A
	Pgi06B00014050	CfPEPC3_B
	Pgi02A00044640	CfPEPC4_A
	Pgi03B00027990	CfPEPC4_B
	Pgi06A00044240	CfPEPC5_A
	Pgi06B00063920	CfPEPC5_B
	Pgi03A00001940	CfPEPC-k1_A
	Pgi03B00001900	CfPEPC-k1_B
<b>PEPC kinase</b>	Pgi03A00013120	CfPEPC-k2_A
	Pgi03B00013070	CfPEPC-k2_B
	Pgi03A00039540	CfPEPC-k3_A
	Pgi04B00016760	CfPEPC-k3_B
	Pgi01A00033830	CfPPDK1_A
<b>Pyruvate orthophosphate dikinase</b>	Pgi01B00042240	CfPPDK1_B
	Pgi07A00010270	CfPPDK-RP_A
<b>PPDK regulatory protein</b>	Pgi07B00010420	CfPPDK-RP_B
<b>Phosphoenolpyruvate carboxykinase</b>	Pgi01A00016400	CfPEPCK1_A

	Pgi06B00034220	CfPEPCK1_B
	Pgi05A00031680	CfPEPCK2_A
	Pgi05B00032130	CfPEPCK2_B
<b>Aspartate transaminase</b>	Pgi03A00002400	CfAspAT_A
	Pgi03B00002380	CfAspAT_B
<b>Alanine transaminase</b>	Pgi02A00024740	CfAlaAT_A
	Pgi02B00025890	CfAlaAT_B
<b>Probable sodium/metabolite cotransporter 2</b>	Pgi06A00020290	CfBASS2_A
	Pgi06B00021230	CfBASS2_B
<b>2-oxoglutarate/malate transporter 1</b>	Pgi04A00009500	CfOMT1_A
	Pgi01B00009320	CfOMT1_B
<b>Dicarboxylate transporter 2</b>	Pgi04A00029510	CfDiT2_A
	Pgi01B00029590	CfDiT2_B
<b>Triose/phosphate transmembrane antiporter</b>	Pgi01A00026190	CfPPT_A
	Pgi01B00023890	CfPPT_B
<b>Sodium/proton antiporter 1</b>	Pgi07A00029340	CfNHD1_A
	Pgi07B00032820	CfNHD1_B

---

**Table S12. Quality control and mapping statistics of resequencing data**

genus names	sample	clean reads	Clean Base (bp)	GC Content (%)	Mapping rate		
					Sub A	Sub B	AB
<i>Pennisetum</i>	YN-2007535422	156,605,216	23,490,782,400	47	72.70%	80.45%	98.01%
<i>Pennisetum</i>	YN-2008532134	150,295,676	22,544,351,400	47	72.53%	80.24%	97.92%
<i>Pennisetum</i>	RY4HWC	219,117,450	32,867,617,500	47	72.18%	79.90%	97.68%
<i>Cenchrus</i>	<i>Cenchrus fungigraminus</i> (JUJUNCAO)	35,610,000	5,341,500,000	47	71.02%	79.05%	97.63%
<i>Pennisetum</i>	G1Z	253,614,512	38,042,176,800	47	70.74%	78.39%	95.66%
<i>Pennisetum</i>	BXHY	146,492,930	21,973,939,500	47	70.88%	78.95%	95.60%
<i>Pennisetum</i>	ZXC	251,433,182	37,714,977,300	46	70.71%	78.80%	95.45%
<i>Pennisetum</i>	DSXC-L	229,195,698	34,379,354,700	47	70.57%	78.83%	95.40%
<i>Pennisetum</i>	ZJLWC	216,187,006	32,428,050,900	47	70.28%	78.67%	95.06%
<i>Pennisetum</i>	FJXC	159,663,548	23,949,532,200	47	69.17%	77.60%	93.78%
<i>Pennisetum</i>	YN-82-21	158,911,116	23,836,667,400	47	69.32%	77.49%	93.74%
<i>Pennisetum</i>	HNXC	248,139,874	37,220,981,100	47	68.99%	77.49%	93.58%
<i>Pennisetum</i>	TXC	258,405,408	38,760,811,200	46	68.96%	77.51%	93.55%
<i>Pennisetum</i>	FJ-2012-11	144,233,242	21,634,986,300	47	69.01%	77.37%	93.53%
<i>Pennisetum</i>	HNDS	205,998,212	30,899,731,800	47	68.48%	76.75%	92.70%
<i>Pennisetum</i>	AXC	225,487,608	33,823,141,200	47	67.87%	76.33%	91.89%
<i>Cenchrus</i>	Bajra	227,346,622	34,101,993,300	48	72.69%	48.43%	74.41%
<i>Pennisetum</i>	s22-S2-1	50,240,552	7,536,082,800	54	62.48%	62.38%	73.48%
<i>Pennisetum</i>	MZL	265,395,788	39,809,368,200	48	71.90%	47.05%	73.35%
<i>Setaria</i>	SRR13414328	76,141,500	11,421,225,000	46	9.16%	8.72%	18.25%
<i>Setaria</i>	SRR13414313	81,423,842	12,213,576,300	47	14.14%	14.67%	16.95%
<i>Setaria</i>	SRR13414290	81,800,876	12,270,131,400	45	13.49%	14.16%	15.83%

**Table S13. Sequence of probes for the three centromeric satellite repeats**

<b>Repeat</b>	<b>Oligomer sequence</b>
<b><i>CEN137</i></b>	ATTCGGGGTACCAAAGTTGTGAAGCATCC
<b><i>CEN148</i></b>	ACCTTCGGCATGTAAATAACTTGTGTTTCG
<b><i>CEN156</i></b>	TTGCGAAAATGGTTTCGCAACAAAACATCC