Telomere-to-telomere Phragmites australis reference genome assembly with a B chromosome provides new insights into its evolution and polysaccharide biosynthesis

Suxia Cui (sxcui@cnu.edu.cn)
Capital Normal University
Jipeng Cui
https://orcid.org/0000-0002-3016-1257
Rui Wang
Ruoqing Gu
Minghui Chen
Ziyao Wang
Li Li
Jianming Hong

Article

Keywords: Phragmites australis, B chromosome, Polysaccharide, Rhizome

Posted Date: January 18th, 2024

DOI: https://doi.org/10.21203/rs.3.rs-3855709/v1

Additional Declarations: There is **NO** Competing Interest.

Telomere-to-telomere *Phragmites australis* reference genome assembly with a B chromosome provides new insights into its evolution and polysaccharide biosynthesis

Jipeng Cui[1, 2], Rui Wang[1, 2], Ruoqing Gu[1, 2], Minghui Chen[1], Ziyao Wang[1, 2], Li Li[1, 2], Jianming Hong[1], Suxia Cui[1, 2, *]

1 College of Life Sciences, Capital Normal University, Haidian District, Beijing, China.

2 Beijing Key Laboratory of Plant Gene Resources and Biotechnology for Carbon Reduction and Environmental Improvement, Beijing, 100048, China.

* Corresponding author: Suxia Cui, E-mail address: sxcui@cnu.edu.cn.
Telephone: 010-68901858

E-mail address:
Jipeng Cui (2210801009@cnu.edu.cn)
Rui Wang (13693533528@163.com)
Ruoqing Gu (18614221560@163.com)
Minghui Chen (chenmh0220@163.com)
Ziyao Wang (juliajuii@163.com)
Li Li (lely@cnu.edu.cn)
Jianming Hong (1660261741@qq.com)

Keywords: *Phragmites australis,* B chromosome, Polysaccharide, Rhizome

Abstract

Phragmites australis, widely distributed worldwide, has a vast biomass and a strong ability to adapt to the environment. Its rhizome (known as "Lugen" in Chinese medicine) has been used as a traditional Chinese medicine for over 2,000 years. In this study, we assembled a chromosome-level reference genome of *Phragmites australis* containing one B chromosome. An explosion of LTR-RTs, centered on the Copia family, occurred during the late Pleistocene, driving the expansion of reeds genome size and subgenomic differentiation. Comparative genomic analysis showed that P. australis underwent two whole gene duplication events, was segregated from *Cleistogenes songorica* at 34.6 Mya, and that 41.26% of the gene families underwent expansion. Based on multi-tissue transcriptomic data, we identified structural genes in the biosynthetic pathway of pharmacologically active Phragmites polysaccharides that have important roles in rhizome development. The assembly of the *P. australis* genome contributes to our understanding of Arundiaceae evolution and polysaccharide biosynthesis in *phragmitis rhizoma*.

Introduction

Phragmites australis are widely distributed in rivers, lakes, dunes, alkaline salt flats, and other habitats worldwide, with strong environmental adaptability and colossal biomass. It has essential ecological and economic values in ecological protection, animal fodder, and traditional Chinese medicine. *Rhizoma phragmitis* (fresh or dried rhizome of *P. australis*), which has been used clinically in China for over 2,000 years, is known as Lugen in Chinese medicine, and its pharmacological effects have been documented in several ancient medical books, such as the Invaluable Prescriptions for Emergencies (Bei Ji Qian Jin Yao Fang), Yu Qiu's Exegesis for Materia Medica (Yu Qiu Yao Jie), and so on[1]. In modern medical research, Lugen is effective in antimicrobial[2], anti-inflammation[3,4], antioxidative and hepatoprotective[5], and antiviral[6]. Lugen has been widely used in the novel coronavirus pneumonia (COVID-19) outbreak with significant efficacy[7,8]. The Lugen polysaccharides have been shown to be the significant medicinal constituents of *Phragmites rhizoma* (PR), in which the acidic polysaccharide PRP-2 has been well characterized in terms of its pharmacological mechanism and structure[2-5]. PRP-2 consists mainly of galactose (34.70%), fucose (36.15%), and rhamnose (0.88%) in the form of →3)-β-D-GalpA-(1→, →2, 3)-α-L-Fucp-(1→ and α-L-Fucp (4SO3-) -(1→[3]. However, due to the complexity genome ploidy and fewer genetic resources of *P. australis*, which seriously limits molecular mechanistic studies and germplasm improvement in *P. australis.*

Non-essential supernumerary or accessory chromosomes that do not follow the typical Mendelian rule of equal segregation, also known as B chromosomes (Bs), have been found in many species[9,10]. It is commonly believed that Bs do not have any function, but current studies have shown that Bs have unexpected functions. For example, Bs can preferentially attach to the spindle on one side of the egg, segregate and produce unequal gametes[11-13]; in some haploid reproducing species, Bs known as PSR (paternity sex ratio) can eliminate the paternal genome during embryonic mitosis[14,15]; and active genes or non-coding RNAs in Bs can also affect maize A-chromosome (As) gene transcriptional profiles or phenotypes[16-18].

The rapid asexual reproduction of *P. australis* utilizing an extensive rhizome network can provide an absolute advantage in interspecific competition and is a crucial organ for *P. australis* to achieve perenniality in complex habitats[19,20]. In *Oryza longistaminata*, Fan et al. found that

the hydrolyzed monosaccharides of sucrose could affect rhizome elongation and orientation by increasing the osmotic pressure of rhizome cells; meanwhile, the sucrose could promote the development of axillary buds into secondary rhizomes and retard the upward growth of axillary buds in rhizomes[21-23]. In addition, the adequate sucrose supply and increased sucrose availability in axillary buds can reduce competition for sucrose from the terminal buds and help rhizome axillary buds break through the inhibition of apical dominance[24-28].

With the continuous improvement and development of High fidelity reads sequencing (HiFi) and high-throughput chromosome conformation capture (Hi-C) technologies, a series of high-quality medicinal plant genome studies have been promoted, such as *Morinda officinalis*[29], *Rheum tanguticum*[30], and *Isatis indigotica*[31], etc. The assembly of these complex genomes provides usable genetic data for identifying biosynthesis pathways of active ingredients in medicinal plants.

Here, we report a heterotetraploid genome at the telomere-to-telomere chromosome level in *P. australis*, which was successfully disassembled into two sets of subgenomic haplotypes. Based on FISH and Hi-C results, we unexpectedly discovered and assembled the B chromosome in *P. australis*. With the full-length transcriptome data from multi-tissue hybrid sequencing, we obtained more accurate genome annotation information. Through phylogenetic and evolutionary analyses, we determined that *P. australis* and *Cleistogenes songorica* shared a common ancestor approximately 34.6 million years ago and inferred a chromosomal evolutionary trajectory in *P. australis* that contained two whole genome duplication (WGD) events. Combined with tissue-specific transcriptome WGCNA analysis, we identified a series of genes associated with polysaccharide biosynthesis in Lugen. Our work deepens our understanding of plant evolution in the Arundiaceae. It provides a vital genetic resource for exploring tissue-specific Lugen polysaccharide biosynthesis and investigating the regulation of rhizome development by carbohydrates.

Results

P. australis genome assembly and annotation

We found a reed strain (*Phragmites australis* subsp. Cuiplus) with an above-ground portion of stem height > 350 cm, basal diameter > 1.5 cm, and highly developed underground rhizomes (Figure 1a), with a genome size predicted by flow cytometry to be approximately 1.8 Gb[32]. Karyotyping and Fluorescence in situ hybridization (FISH) showed that this *P. australis* had 50 chromosomes, mainly proximal or mid-stranded chromosomes. The hybridization signals of the 5SrDNA and 18SrDNA repeat sequences probes were characterized by one pair of weak and one pair of strong hybridization signals (Figure 1a). We used genome survey data (48.31 Gb) for K-mer analysis (k = 17) to further characterize the *P. australis* genome. The size of this *P. australis* genome was 1685.39 Mb, the heterozygosity rate was 0.61%, and the proportion of repetitive sequences was 50.40%, making it a complex genome with high assembly difficulty (Figure S1a). In addition, the GenomeScope results showed that aaab < aabb and Smudgeplot showed that this *P. australis* genome was more inclined to the AB haplotype genome (Figure S1b). Therefore, we hypothesized that this *P. australis* is a heterozygous aneuploid tetraploid strain.

We obtained 482.94 Gb Raw read data and 32.53 Gb HiFi reads data using PacBio Sequel II platform genome sequencing. Using hifiasm for the primary assembly of the genome, we obtained a *P. australis* genome sketch with a size of 874.62 Mb, containing 312 Contig

sequences, N50 of 33.94 Mb, and GC content of 44.21% (Table 1). The genome sketch features were similar to the flow cytometry and genome survey results. Subsequently, we used Hi-C data (113.26 Gb) to visualize Contig's alignment order and orientation for error correction (Table S1). Finally, we anchored 99.46% of the contigs sequences to 25 Chromosomes with a total length of 852.12 Mb, N50 of 34 Mb, and L50 of 10, with only one gap on Chr21 and Chr25, respectively (Table 1). Next, we searched telomeric canonical repeat units (5-12 bp in length) in the *P. australis* genome. Finally, we identified the AAACCCT motif as a *P. australis* telomere repeat monomer with a repeat number 40,863 (Table S2). All chromosomal telomeres were successfully assembled, with 19 chromosomes assembled to telomeres at both ends (Table S3).

Using various genomics data and tools, we identified three candidate *P. australis* centromere repeat sequences in the *P. australis* genome with 108 bp, 216 bp, and 324 bp. Meanwhile, we obtained the genome-wide 5-methylcytosine (5-mC) locus information for *P. australis* from HiFi data using pb-CpG-tools. We found that regions of reduced gene density and increased density of repetitive sequences on *P. australis* chromosomes highly overlapped with areas of dense 5-mC loci (Figure S2). There is growing evidence that the chromosomal centromere region contains fewer genes and denser repetitive sequences and that DNA in this region tends to be highly methylated[33-38]. Finally, we successfully localized the candidate regions of centromere in *P. australis* chromosomes based on the 5-mC distribution (Figure S3).

We combined multiple methods to assess the quality of our assembled *P. australis* genome exhaustively. The genome-wide Hi-C interaction mapping showed that our assembly results conformed to inter- and intra-chromosomal interaction requirements (Figure S4a). The GC content versus sequencing depth distribution map showed that our genome had no significant contaminating sequences during sequencing and assembly (Figure S4b). The CEGMA and BUSCO assessments were 95.97% and 99.30%, respectively (Figure S4c-d). By mapping the survey and iso-seq data to the genome, the mapping rate and coverage exceeded 98%, indicating that our assembled genome has high integrity and uniformity of sequencing (Table S4, Figure S4e). Merqury concordance assessment showed that all chromosomes' consensus quality value (QV) values exceeded Q40, indicating that the accuracy and confidence of our genome assembly exceeded 99.99% (Table S3). The LTR Assembly Index (LAI) reached 11.54, which has reached the reference level genome (Figure S4f). The pure single nucleotide polymorphism (SNP) and insertion/deletion (InDel) rates in the *P. australis* genome were 0.97% and 0.06%, respectively (Table S5). We obtained the *P. australis* genome assembly (PaCui.No1) with ultra-high continuity and integrity from telomere to telomere by Hi-C-assisted assembly, which provided a genetic basis for subsequent in-depth studies of centromeres and highly repetitive regions (Figure 1b).

We combined homology-based and ab initio prediction to identify repetitive sequences in the *P. australis* genome (Figure S5). 64.98% (553,687,485 bp) of the sequences were annotated as transposable elements (TEs), of which 394,357,364 bp were identified as long terminal Repeat transposons (LTRs), which accounted for 46.28% of the *P. australis* genome (71.22% of TEs); 10.77% of the genome sequences were annotated as DNA transposons (16.58% of TEs); and 3.02% (25,707,169 bp) of the sequences were annotated as Tandem Repeat, wherein, 2,965,766 bp were identified as simple sequence repeats (SSRs) containing si- (56.47%), di- (29.61%), tri- (11.68%), tetra- (1.63%), penta- (0.37%), and hexa- (0.23%) repeat units.

We predicted 41,008 protein-coding genes in the *P. australis* genome, with an average gene length and coding DNA sequence (CDS) length of 4,758 bp and 1,242 bp, respectively. The BUSCO assessment of these genes amounted to 98.40% (Table 1). Interestingly, compared to our selected gramineous relatives, we found that *P. australis* has a lower percentage of genes with lengths < 2000 bp but has more genes with lengths > 5000 bp and those containing more than two exons or intron (Figure S6). Gene function annotation of the protein-coding genes annotated within the genome of *P. australis* was performed based on Swiss Prot, KEGG, InterPro, GO, TrEMBL, and NR gene function databases. The combined results showed that 40.527 (98.83%) of the 41,008 protein-coding genes in *P. australis* could be annotated in at least one of the databases (Figure S7). Meanwhile, we identified non-coding RNAs in the *P. australis* genome, including 229 miRNAs, 836 tRNAs, 1744 rRNAs, and 459 snRNAs. We identified three miRNAs (Chr7B, Chr7A, and Chr5B), two rRNAs (Chr6B and Chr4B), and two tRNAs (ChrB and Chr11A) densely populated regions in the *P. australis* genome (Figure S8).

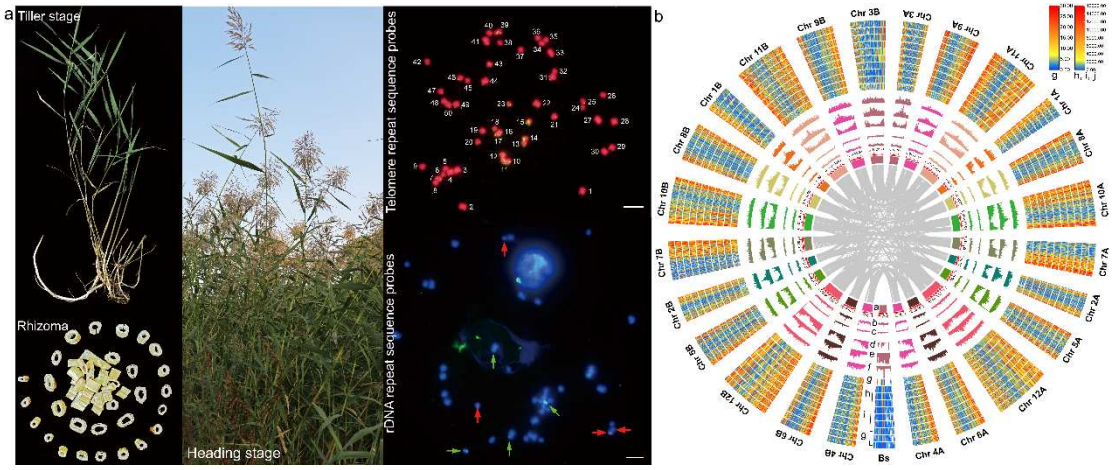


Figure 1. Genomic characterization of *P. australis*. **a** Morphology and Fluorescence in situ hybridization (FISH) of the sequenced plant. Rhizomes of *P. australis* after removal of fibrous roots, cleaning, sectioning, and drying. The heading stage of P. australis was grown at the experimental site of the College of Life Sciences, Capital Normal University. The complete chromosome was identified by telomere repeat sequence probes (green). The 5SrDNA probe (red arrow) showing purple fluorescence and the 18SrDNA probe (green arrow) showing green fluorescence. Scale bars 5 μm. **b** Circos plot of the *P. australis* genome. a. Chromosome length. b. GC content. c. GC skew. d. Gene density. e. LTR-Gypsy density. f. LTR-Copia density. g. Distribution of the coverage of multi-tissue mixed full-length transcriptome sequencing data. h - g. Distribution of leaf, stem (aboveground) coverage, and rhizome RNAseq data, respectively. The inner lines indicate covariate blocks. All densities were calculated in a 100,000 bp window.

Table 1 *P. australis* genome assembly statistics

Contig	PaCui.No1	Draft genome [29]	LpPhrAust1.1
Total size (bp)	874,619,212	1,139,927,050	-
Number of contigs	312	13,411	-
Number of contigs ≥ 50,000 bp	152	4,617	-
Largest contig (bp)	54,964,116	3,219,705	-
GC content (%)	44.21	44.04	-

N50 length (bp)	33,936,801	194,574	-
L50 count	11	1370	-
BUSCO (%)	99.3	93.3	-
Chromosomes			
Number of chromosomes+Unchr	25 +50	-	24+13
Total size (bp)	852,117,256	-	849,281,002
GC content (%)	44.27	-	44.17
N50 length (bp)	34,052,747	-	35,123,032
Gaps	2		59
L50	10	-	10
BUSCO (%)	99.3	-	-
CEGMA (%)	99.46	-	-
LTR Assembly Index (LAI)	11.54	-	-
Second-generation Data Mapping	98.62%	-	-
Third-generation Data Mapping	99.99%	-	-
Annotation			
Number of genes loci	41,008	64,857	-
BUSCO (%)	98.40	-	-
Functional annotation (%)	98.83	-	-
Percentage of repeat sequences (%)	65.62	56.19%	-
Percentage of TEs (%)	64.98	-	-
Number of predicted SSRs	160,796	-	-
Number of rRNAs	1744	-	-
Number of tRNAs	836	-	-
Number of miRNAs	229	-	-
Number of snRNAs	459	-	-

Subgenome	Number of chromosomes	Length (bp)	Number of genes
Subgenome A	12	399249732	20290
Subgenome B	12	427051217	20425
B Chromosome	1	21198758	277

Subgenome sorting and B chromosome assembly

Due to the lack of genetic information on the genomes of *P. australis* parents, subgenome sorting is highly difficult. By intra-species covariance mapping constructed from 13,788 paralogous homologous gene pairs in *P. australis*, we found 12 pairs of very strong 1-to-1 covariance between 25 reed chromosomes (Figure 2a). Combining the Oryza sativa and Panicum virgatum genomes, we extracted homologous chromosome pairs with *P. australis*, respectively, and used their single-copy genes to calculate genetic distances (Figure 2b, Figure S9). We classified the chromosome with closer affinity to the homologous chromosome of O. sativa into subgenome A (LAI: 12.24), while the other one was categorized into subgenome B (LAI: 12.47). We successfully sorted out two haplotype subgenomes that reached the reference genome level (Ordered by chromosome length from largest to smallest) (Table S3). The subgenomes showed an overall 1:1 covariance but frequent chromosomal structural variations between some chromosomes, such as chromosomal inversions between Chr 9 and Chr 12, Chr

19 and Chr 20, and chromosomal translocations between Chr 12 and Chr 13, and Chr 10 and Chr 11 (Figure S10).

We found no significant colinear blocks of chromosome 25 within *P. australis* or between related species (Figure 2a, Figure S9). It has a significantly reduced density and number of protein-coding genes, with only 277 genes present (only 9.65% of the length of this chromosome). Homologs of the 268 genes in Chr25 are widely dispersed in the 24 normal chromosomes of *P. australis* (E value ≤ 1e-5) (Figure 2c). Moreover, 85% of the sequences in the Chr25 chromosome are composed of transposable elements homologous to normal chromosomes. In addition, we found many rRNA sites, multiple significant tRNA density peaks widely distributed in this chromosome, and the presence of multiple possible centromere regions (Figure S8). To investigate whether these retained genes have certain specific functions, we performed GO enrichment of genes on chromosome 25 (Figure 2e). We found that these genes are mainly involved in biological processes related to sister chromatid movement in meiosis, pollen sperm cell differentiation, and spindle. In the Molecular Function classification, these gene products are mainly involved in "histone H3-methyl-lysine-4 demethylase activity", "histone demethylase activity", "histone H3-methyl-lysine-4 demethylase activity", and "telomeric DNA binding". Meanwhile, by RNAseq analysis, we found that the genes expressed on chromosome 25 of *P. australis* (FPKM mean > 1) were mainly involved in the production or regulation of germ cells in *P. australis* and were involved in the molecular functions of several enzymes related to the activity of the electron transport chain (Figure 2d). In summary, this unique chromosome 25 is a B chromosome. We finally determined that this *P. australis* genome is of type AABB (2n = 4x = 48 + 2) and contains two B chromosomes.

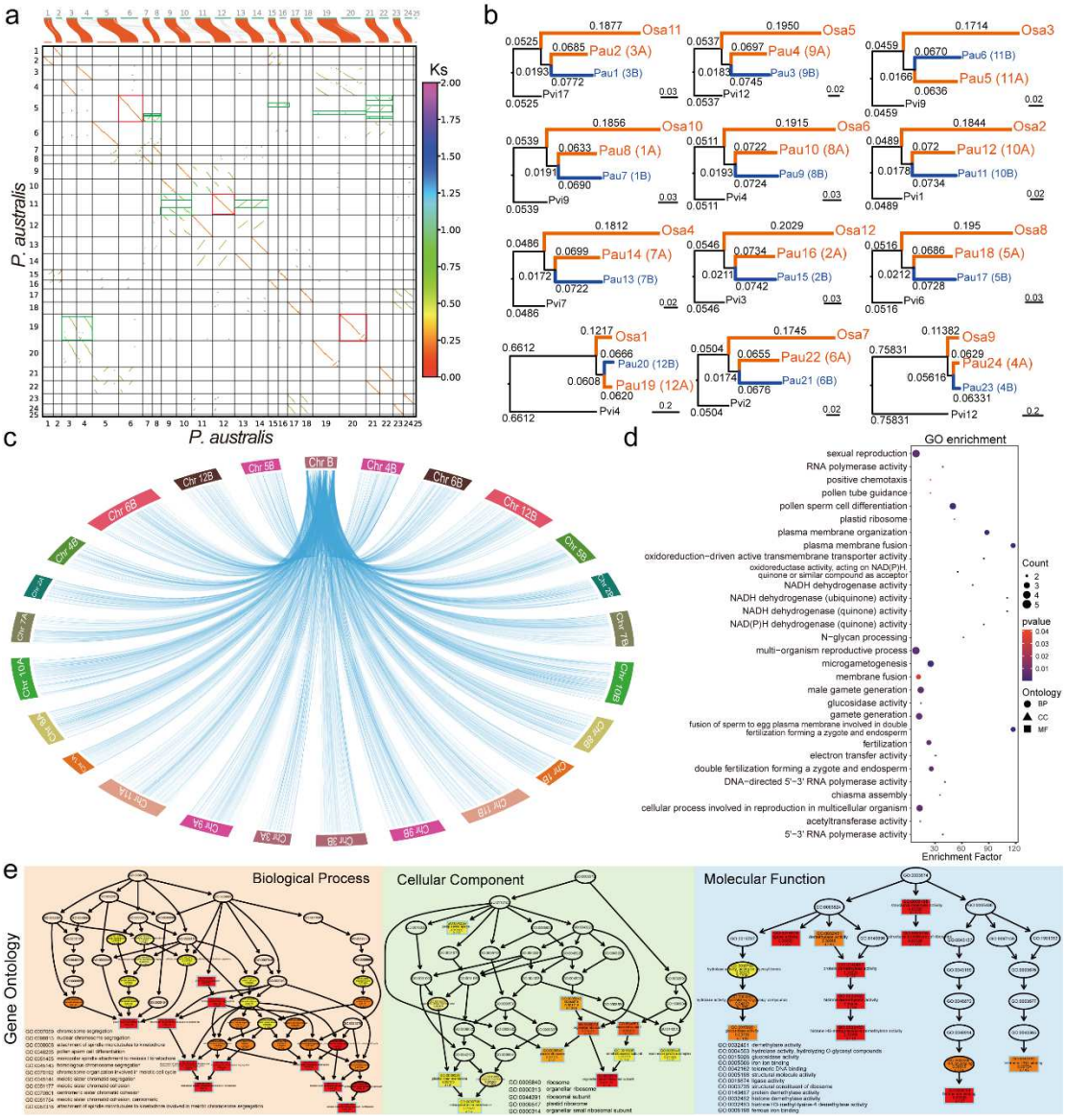


Fig. 2 Subgenomic isolation and B chromosome identification. a Dotplot of co-orthologs genes and Ks within *P. australis* genome. The color of the dot indicates the Ks of the gene pair. **b** Phylogenetic tree of genetic distances constructed based on single-copy homologous sequences. **c** Distribution of genes in the *P. australis* B chromosome in normal chromosomes. The blue connecting line in the figure indicates homology between the genes and those in the B chromosome with an E-value < 1e-10. **d** GO enrichment of genes expressed in the *P. australis* B chromosome. Genes with FPKM mean > 1 in chromosome B were retained for GO enrichment using clusterProfiler 2.0. **e** Directed acyclic map of GO enrichment of genes in the *P. australis* B chromosome. The more specific the molecular function, the lower the hierarchy. The ten most significant nodes are indicated by the boxes in the figure. The enrichment significance of each node is differentiated by color, from highest to lowest: red > orange > yellow.

Two transposon insertion events drive *P. australis* genome expansion

Many long terminal repeat-retrotransposons (LTR-RTs) exist in plant and animal genomes, especially in plant genomes. In *P. australis*, the transposons were mainly LTRs, of which the Gypsy and Copia families accounted for 34.73% and 21.65% of the total LTR length, respectively (Figure S5). The distribution density of the Gypsy family on *P. australis*

chromosomes was strongly negatively correlated with the gene arrangement, and it was mainly distributed near the centromere. Still, the Copia family was relatively uniformly distributed at both ends of *P. australis* chromosomes, which was consistent with the trend of gene distribution (Figure 1b, Figure S2). By calculating the timing of LTR-RT insertions, we found that, except for the *C. songorica* genome, the genomes of *P. australis* and the other gramineous species experienced LTR-RT insertion events of varying magnitude between 0.10 and 0.16 Mya and all of them had higher average chromosome lengths than *C. songorica* (Figure 3a). The amplification densities and peak insertion times of LTR-RTs in these genomes differed considerably, and both *P. australis* genomes underwent at least two significant transposon rapid insertion events (Figure 3b, Figure S11).

P. australis underwent an insertion event of LTR-RTs centered on Gypsy families 1 to 2 million years ago, and we define the LTR-RTs inserted in this event as ancient LTR-RTs (Figure 3b). We found that these ancient Gypsy families were concentrated near the centromere of certain chromosomes, e.g., Chr 4, Chr6, Chr9, Chr16, etc., which not only increased the length of repetitive sequences within the centromere region, but also drove centromere evolution (Figure S12a). Subsequently, a more intensive and rapid insertion event of LTR-RTs occurred 0-500,000 years ago, in which the Copia family was significantly dominant, and we define the LTR-RTs from this event as new LTR-RTs (Figure 3b). These newly inserted Copia family members accounted for 62.05% of all Copia and were widely distributed in or near gene regions (Figure S12b). The large number of insertions of these elements may affect gene sequences in the vicinity of TEs to a certain extent, providing abundant raw material for variation in genome evolution. Interestingly, we found that the content of Copia family sequences in these genomes conformed to a linear normal distribution between the genome size and the Copia family sequences (Figure 3c), and there was a very strong positive correlation (Pearson $r = 0.9618$, $P < 0.001$) (Figure 3d).

Meanwhile, we found that LTR-RT insertion events and extant transposon content occurring between the two sets of subgenes in *P. australis* showed significant differences (Figure 3b and e). Compared to subgenome A, the Gypsy family in subgenome B dominated ancient transposon insertion events and showed a higher density of unknown transposon insertion types in recent events (Figure 3b). In addition, longer Gypsy family transposon sequences were identified in subgenome B. The differences between these insertion events led to the differences in transposon types and lengths in the two sets of subgenomes (Figure 3e). To some extent, the LTR-RT drove the differentiation of the subgenome in *P. australis*.

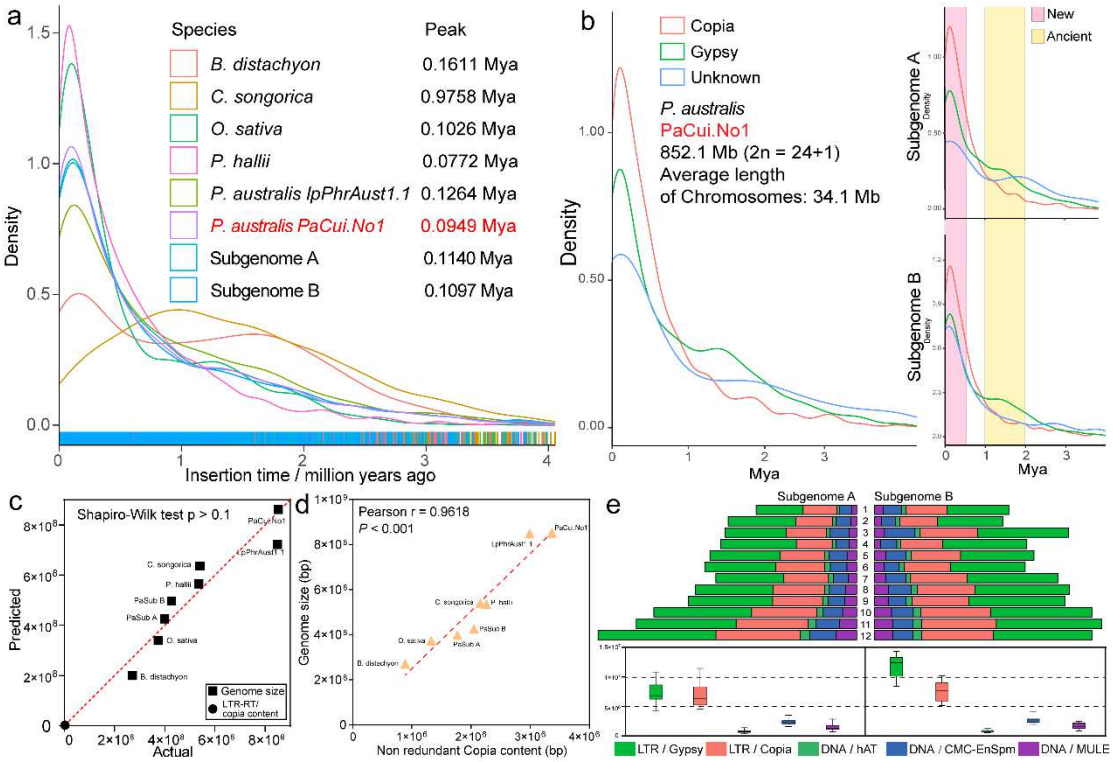


Fig. 3 Analysis of LTR-RTs in *P. australis* genomes. a Insertion times of LTR-RTs in eight genomes, including two *P. australis* subgenomes. **b** Insertion times of significant types of *P. australis* LTR-RTs. The right panel indicates the significant types of LTR-RT insertion times for subgenome A and subgenome B, respectively. **c** Normal QQ plot of Copia family content versus genome size for the *P. australis* genome. The normal distribution test using the Shapiro-Wilk (SW) method. **d** Pearson correlation analysis between Copia family content and genome size in the *P. australis* genome. **e** Differences in transposon content between two sets of subgenes in *P. australis*.

Analysis of the gene family

To understand the patterns of gene family divergence during *P. australis* evolution, we analyzed the *P. australis* genome for gene family clustering with 13 Gramineae species, including *Aegilops tauschii*, *Brachypodium distachyon*, *Cleistogenes songorica*, *Dendrocalamus latiflorus* Munro, *Oryza sativa*, *Panicum hallii*, *Panicum virgatum*, *Pennisetum purpureum* Schum, *Setaria italica*, *Setaria viridis*, *Sorghum bicolor*, *Triticum aestivum*, *Zea mays*, and *Arabidopsis thaliana* as outgroup species. We identified 28,025 gene families in 15 species and 17,494 gene families in *P. australis*, with multiple-copy ortholog genes accounting for 32.21% of all gene families (Figure 4a). All species shared 5507 gene families, and 297 unique gene families were present in *P. australis*, containing 871 genes (Figure S13a). We found that these *P. australis* unique family members are mainly involved in biological processes related to protein or macromolecule depalmitoylation, "protein dephosphorylation", "lipoprotein catabolic process" and "phosphatidic acid metabolic process". The molecular functions were also enriched for the terms "palmitoyl hydrolase activity", "phosphoprotein phosphatase activity", "protein serine/threonine phosphatase activity" and "phospholipase A2 activity". Moreover, many genes related to vesicular transport, such as Golgi and vesicles, were enriched in the Cellular Component category. In addition, some genes were also enriched in Terms related to oleoresin lactone synthesis and reactive oxygen species biosynthetic process. The GO

functional enrichment of these *P. australis* unique genes seems to predict the existence of special competence in depalmitoylation and dephosphorylation with other species (Figure S13b).

We were surprised that 41.26% of gene families in *P. australis* underwent expansion involving 7217 gene families (20,142 transcripts), a much higher percentage than in the other 14 species. These expanded gene families were mainly associated with genes related to telomere maintenance (GO:0032200, GO:0010833, GO:0007004, and GO:0000723), DNA binding (GO:1990837, GO:0043565, GO:0000976, and GO:0003690), and starch synthesis (GO. 0004556 and GO:0016160) are functionally related (Figure S13c). Gene family expansion in *P. australis* has resulted in a significantly increased proportion of Multiple copy orthologs, while providing *P. australis* with more members of genes with potential functions that can help *P. australis* to undergo rapid adaptation and evolution in the face of unfavorable environments and stresses. Meanwhile, we found 288 positively selected genes in 1,039 single-copy direct homologs in *P. australis*. Many of these genes are present in genes involved in telomere maintenance (GO:0032204, GO:0000723, and GO:0032200) as well as DNA damage checking (GO:0007095, GO:0031572) (Figure S13d). These genes subjected to positive selection related to telomere or DNA repair play an essential role in the perennial mechanism of *P. australis*.

***P. australis* evolutionary position and whole genome duplication analyses**

We have constructed a phylogenetic tree to investigate the evolutionary origins and species-related relationships of *P. australis* using 302 single-copy genes identified in 15 species (Figure 4a). The results showed that Arundiaceae is more closely related to Chloridoideae. Their common ancestor appeared earlier than the ancestor of Panicoideae and later than the ancestor of Oryzoideae, and separated from the ancestor of Mibiaceae 38.1 million years ago (Figure 4a). The synteny analysis showed strong colinearity between *P. australis* chromosomes and *O. sativa* and *C. songorica* chromosomes (Figure 4b). Meanwhile, we found 246 syntenic blocks containing 19,384 collinear gene pairs between *P. australis* and *O. sativa*, and 614 syntenic blocks containing 25,732 collinear gene pairs with *C. songorica*. Colinear genes with C. songorica accounted for 62.75% of all genes in *P. australis*, much higher than the collinear gene pairs with *O. sativa* (Table S6). These results suggest that *P. australis* and *C. songorica* are evolutionarily closely related sister groups that diverged about 34.6 million years ago. Colinear blocks of some chromosomes in *P. australis* mapped to more than two chromosomes in *O. sativa* and *C. songorica* simultaneously (Figure 4b).

We found in the genome dot plot that each chromosome of *P. australis*, except Bs, had one best-matched chromosome and two sub-matched chromosomes where more chromosomal rearrangement events occurred, suggesting that two WGD events occurred in *P. australis* (Figure 2a). The depth of colinearity between the *P. australis* genome and O. sativa (1: 2) and *C. songorica* (2: 2), respectively, predicted that *P. australis* experienced a shared WGD event with *O. sativa* and an independent WGD event with *C. songorica* (Figure S14, Table S6). Two significant peaks in the *P. australis* genome, corresponding to the two WGD events, were analyzed by synonymous substitution (Figure 4c). Five Gramineae genomes, including two subgenomes, share an ancient ρ-WGD event between ~60-80 Mya[39-42]. Following this, *P. australis* was successively separated from *O. sativa* (Ks = 0.470) and *C. songorica* (Ks = 0.305) and underwent a polyploidy event unique to *P. australis* at ~23.9 Mya (Ks = 0.211) (Figure 4c). The four-fold degenerate sites (4DTv) analysis was consistent with the Ks results (Figure S15).

Meanwhile, we observed significant peaks at both Ks = 0.628 and Ks = 0.205 in the comparison between subgenes A and B, corresponding to the ancient ρ-WGD event and the polyploidy event unique to the *P. australis* genome, respectively. In summary, we hypothesize that the parental ancestor of this *P. australis* diverged from successive *O. sativa* and *C. songorica* after undergoing the ρ-WGD event shared by most gramineous species and formed a heterotetraploid *P. australis* through interspecific hybridization and an exclusive heterologous tetraploidization event (WGD-IV).

We used WGDI based on the Ancestral monocot karyotype except for Acoraceae (AMK-A) to construct a karyotype evolutionary process that includes the major species of Gramineae. Accompanied by two WGD events, the eight chromosomes of AMK-A underwent at least 356 chromosome breaks, 340 chromosome fusions, and eventual integration into the 24 haplotype chromosomes now found in *P. australis* (Figure 4d). Subgenome A, more closely related to *O. sativa*, experienced a higher frequency of ancestral chromosome splits and fusions and has more collinear gene pairs with other species. These results suggest that subgenome A appears to be the more ancient parental genome. (Figure 4d, Table S6).

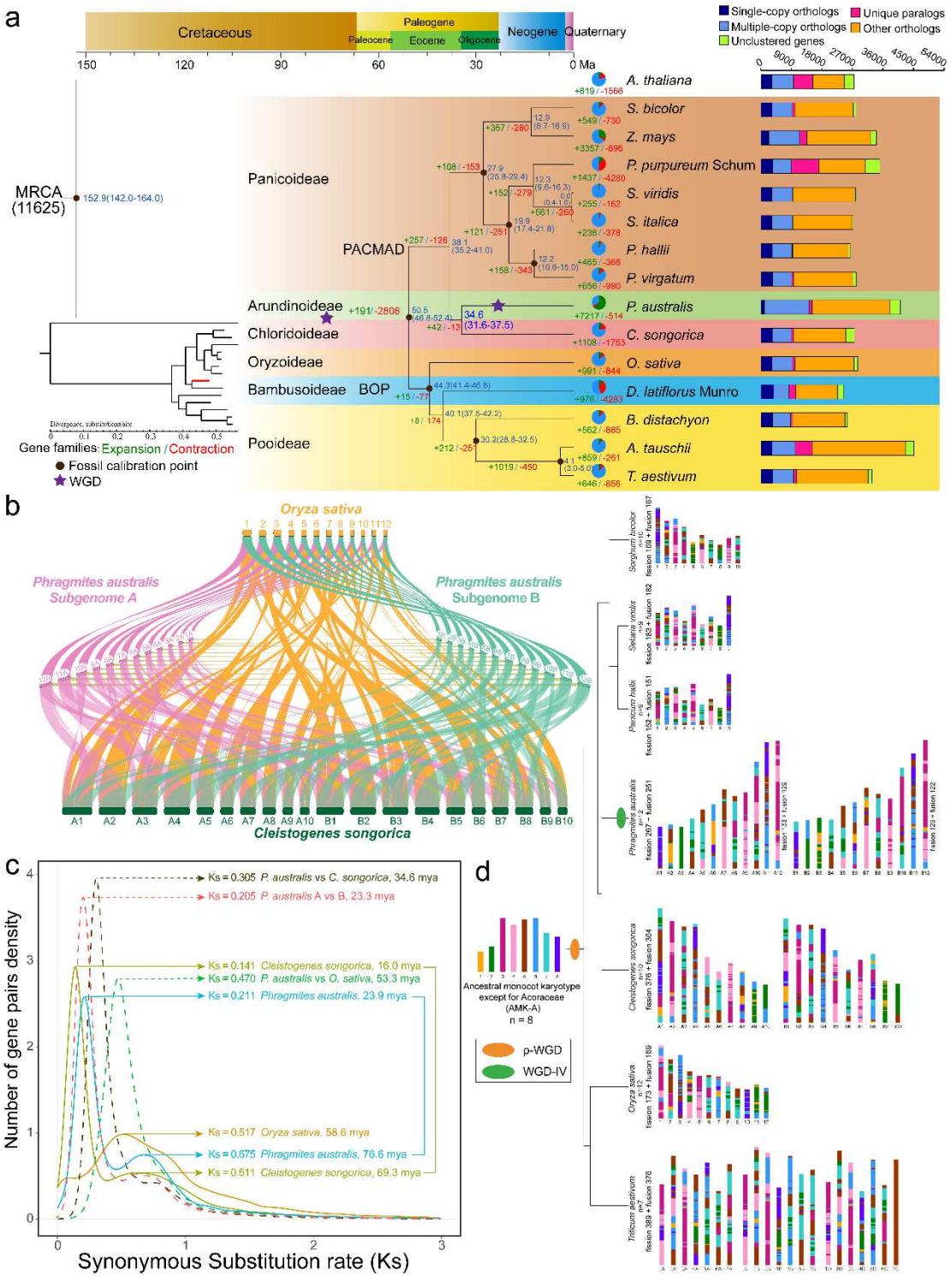


Fig. 4 Evolutionary and comparative genomic analysis. a Phylogenetic analysis, divergence time estimation, and gene family expansion/contraction analysis. A phylogenetic tree was constructed with Arabidopsis thaliana as an outgroup using the genomes of 14 Poaceae, including reed. Purple stars indicate genome-wide replication events, The branch length of a phylogenetic tree represents the amount of cumulative evolution or cumulative mutation; The blue numbers indicate divergence times; and the red and green numbers indicate gene families for expanded and contracted gene families, respectively; Classification of orthologous and lineage-specific gene families in reed and other representative plants

are shown on the right. **b** Collinear relationship of two subgenomes of reed, rice, and awnless cryptomeria. Different color lines connect matched gene pairs between different genomes. **c** Distribution of Ks between Reed and the other two species. The lines indicate the distribution of Ks within genomes (continuous) and between genomes (dashed lines). **d** Ancestral karyotype evolution in *Sorghum bicolor*, *Setaria viridis, Panicum hallii, Phragmites australis, Cleistogenes songorica, Oryza sativa,* and *Triticum aestivum*. Different colors represent chromosome segments of the Ancestral monocot karyotype except for Acoraceae (AMK-A), and different color combinations represent the chromosomal recombination events that the modern karyotype of each species underwent.

The WGCNA analysis of the multi-tissue transcriptome of *P. australis* and the identification of genes involved in the Lugen polysaccharide biosynthesis pathway

To understand the biosynthetic pathway of polysaccharides in *P. australis* and the transcription factors (TFs) that play critical regulatory roles, we performed the weighted correlation network analysis (WGCNA) on the transcriptome data of three tissue samples of *P. australis*, containing leaves, aerial stems, and rhizomes (Figure 5f). Principal component analysis revealed high homogeneity among similar tissue transcriptome data and showed significant differences overall (Figure 5a). We categorized genes with an average FPKM >1 into 14 co-expression modules containing 22,285 genes (softpower = 12) (Figure 5b). We also identified three modules (MEturquoise, MEbrown, and MEblue) that were highly correlated with *P. australi's* unique tissues (Pearson cor > 0.95) (Figure 5c). The GS-MM results showed that these genes highly correlated with the tissue specificity and were also important in the module (Figure 5d). GO enrichment analysis showed that the genes in MEturquoise, which is highly associated with leaf height, are mainly involved in photosynthetic pigment synthesis, electron transport chain, and light-responsive related biological processes. The MEbrown associated with aboveground stem tissues contained genes of sucrose or monosaccharide transporter protein families, and the expression of these genes was specifically increased in aerial stems (Figure 5e-g). Interestingly, the MEblue module significantly associated with underground rhizomes of *P. australis* was enriched with a large number of genes related to plant polysaccharide synthesis and cell wall formation, which play essential roles in polysaccharide synthesis and rhizome development of *P. australis*.

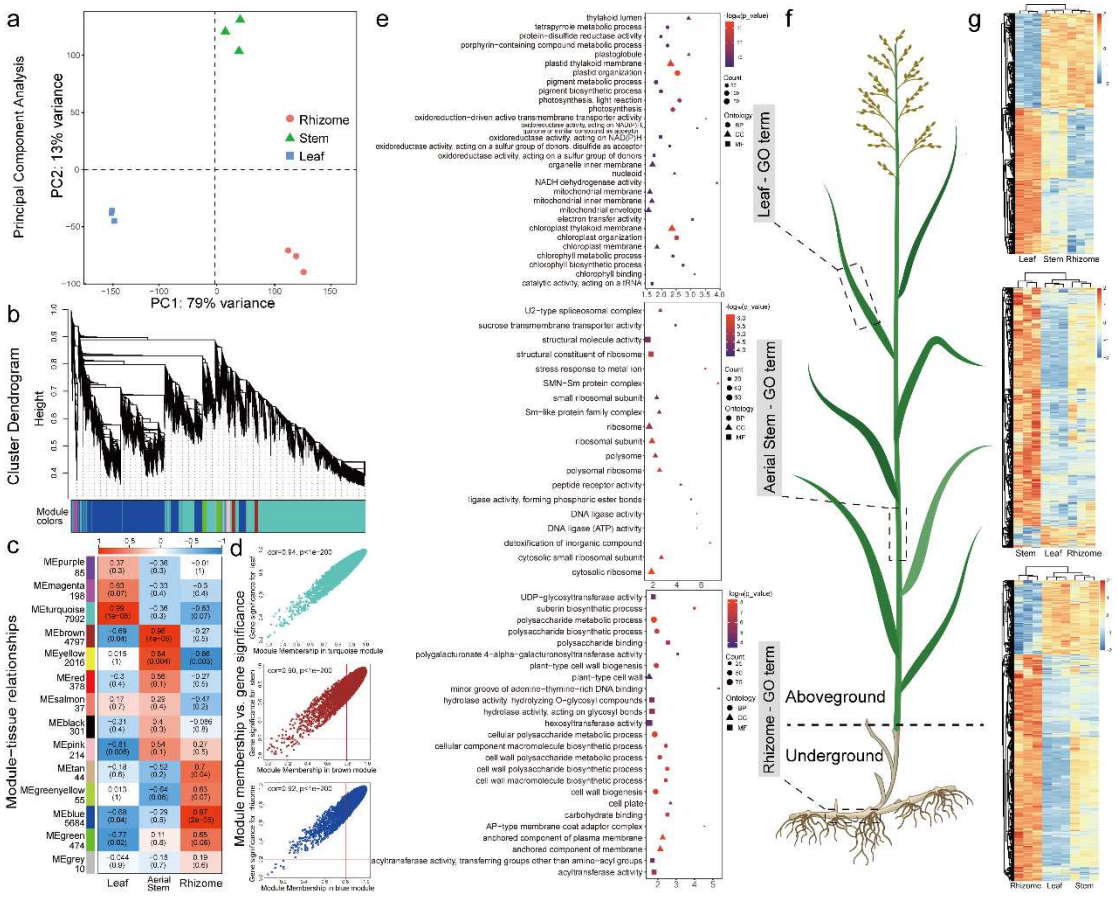


Fig. 5 Construction of the reed co-expression network. a Score scatter plots for the PCA model with nine samples. **b** Clustering dendrogram of gene based on the topological overlay. WGCNA analysis was performed after screening based on gene expression FPKM mean > 1. Each branch of the gene dendrogram corresponds to one module; different colors represent different modules. **c** Heatmap of gene co-expression network module-tissue association. Each row corresponds to a module characteristic gene (eigengene), and each column corresponds to a specific tissue. Each cell contains the corresponding correlation value and P-value, and indicates the strength of the correlation according to the color. **d** A scatterplot of Gene Signifificance (GS) for weight vs. Module Membership (MM) in three module eigengene. There is a highly significant correlation between GS and MM in this module， illustrating that genes highly significantly associated with a trait are often also the most critical elements of modules associated with the trait. The red lines indicate the thresholds for |MM| > 0.8 and |GS| > 0.2, respectively. **e** Bubble diagram showing the results of GO enrichment of genes contained in the module eigengene. **f** Pattern map of reed plants (Created by Biorender, https://www.biorender.com). **g** The heatmap shows the changes in FPKM based on log2 in the module eigengene.

Based on our functional annotation information, we identified structural genes in the Lugen polysaccharide (PRP2) biosynthesis pathway that encompasses the sucrose synthesis (SUS), sucrose transporters protein (SUC), and Lugen polysaccharide monomer (GDP-Fucose, UDP-Rhamnose, UDP-Galactose and UDP-Galacturonate) pathways, involving a total of 182 transcripts (69 transcripts with average FPKM expression >1). Most of these gene families involved in polysaccharide synthesis underwent varying degrees of expansion (Figure 6a). The significantly high expression of Sucrose-6F-phosphate phosphohydrolase (SPP) and sucrose phosphate synthase (SPS), two critical enzymes for sucrose synthesis, in leaves increased the

amount of sucrose produced via photosynthesis. Sucrose transporter proteins (SUC) are not only related to the mobility and availability of sucrose in plants but also essential for plant-specific tissue development regulation. We identified nine *PaSUC* genes involving 12 transcripts in *P. australis*, among which *PaSUC1.1*, *PaSUC1.2*, and *PaSUC4.1* were significantly overexpressed in aerial stem tissues. They may assume the function of a long-distance sucrose transporter for transporting sucrose synthesized in leaves down through above-ground stems to rhizomes. Sucrose synthase (SUS), as a key rate-limiting enzyme in the sucrose synthesis pathway, mediates the reversible conversion of sucrose and ADP (or UDP) to fructose and ADPG (or UDPG). *SUS* (rna-Pau33356.1) expression was significantly increased in *P. australis* rhizomes, which promoted the accumulation of carbohydrates such as sucrose, fructose, and UDP-glucose in the rhizomes. Significantly increased expression of GDP-mannose pyrophosphorylase (GMPP), GDP-mannose 4,6-dehydratase (GMD), and dTDP-4-dehydrorhamnose reductase (UER1) in the polysaccharide monomer synthesis pathway in aboveground stems or rhizomes also led to the accumulation of GDP-Fucose and UDP-Rhamnose monomers in these tissues. In addition to the high expression of specific genes that may be associated with galactose stress response in leaves, most of the UDP-glucose-4-epimerase (UGE) and UDP-glucose-4-epimerase (GALE) were significantly overexpressed in rhizomes, and these genes led to the accumulation of UDP-Galactose and UDP-Galacturonate in rhizomes. The high expression of UDP-D-galactose dehydrogenase (UGD) and UTP-glucose-1-phosphate uridylyltransferase (UGP2) in rhizomes may also play an essential role in polysaccharide synthesis in Lugen.

We analyzed cis-acting elements within the 2000 bp region upstream of the *SUC* gene identified in *P. australis* to predict transcription factors regulating sucrose transporter proteins. We identified multiple cis-acting regulatory elements associated with phytohormone (abscisic acid and MeJA) or stress response stress (defense and stress, light, and low temperature) as well as binding sites for the transcription factor MYB (including MYB binding site involved in drought-inducibility, MYB binding site involved in light responsiveness, and MYBHv1 binding site, involving a total of nine PaSUC members) within the promoter regions of these *SUC* genes, suggesting that the MYB transcription factor may play a crucial role in the regulation of *P. australis* sucrose transporter proteins (Figure 6b). Since the *PaSUC* genes with an average FPKM > 1 in *P. australis* were all categorized in the brown module, we inferred that the genes in this module were associated with sucrose transport in stems. These Hub genes are mainly involved in sucrose transport and response to heat (Figure S16) and contain nine transcription factors and a sucrose transporter protein (*PaSUC1.1*) (Figure 6d). Ultimately, combining the co-expression network and the trend of expression of these genes, we screened for a PaMYB transcription factor (rna-Pau22479.1) that may regulate the expression of the sucrose transporter protein *PaSUC1.1* in stems (Figure 6c-d, Table S7). Our results provide data support and a research basis for studies exploring the regulatory mechanisms of multi-tissue coordinated long-range sucrose transport in *P. australis*.

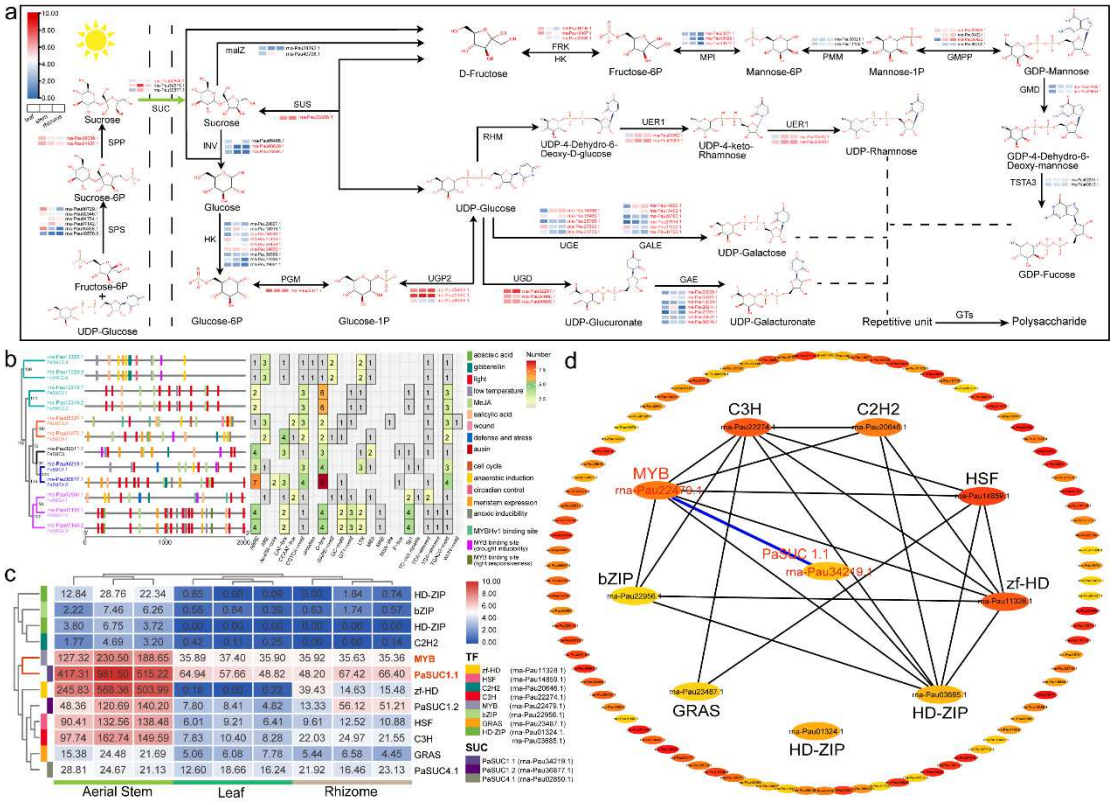


Fig. 6 Lugen polysaccharide biosynthesis pathway and sugar transporter protein transcription factor identification. a Gene expression analysis of genes involved in polysaccharide biosynthesis in "Lugen". Different color blocks represent the expression levels of the encoded genes in different tissues, the squares from left to right correspond to leaves, aerial stems, and rhizomes, and the red font represents the gene family expansion in Reed. Compound structure diagrams were obtained from EMBL-EBI (https://www.ebi.ac.uk/). **b** Prediction of cis-acting elements in the promoter region of the 12 identified *PaSUC* genes. A maximum likelihood phylogenetic tree (left) of 12 PaSUC family members was constructed using IQ-TREE (v 1.6.12) with 1,000 bootstrap replications. **c** Heatmap of sucrose transporters protein expression with transcription factors in Hub gene. The expression level of each gene is represented by log2 (FPKM), with redder colors indicating higher expression levels. **d** Hub genes co-expression network in the brown module. The connecting line indicates the co-expression relationship between TFs and *PaSUC1.1* in hub genes.

3 Discussion

Assembly and annotation of the *P. australis* T2T genome

With the rapid development of genome sequencing and assembly technologies, more and more complex genomes are being assembled and published, especially the genomes of medicinal plants, which are essential for identifying biosynthetic pathways of actives and genetic studies of adversity response in these species[29-31]. The Arundiaceae family is widely distributed worldwide and has enormous biomass and economic value, but only a few genomic resources have been published. In this study, we successfully assembled the high-quality chromosome-level telomere-to-telomere heterozygous tetraploid *P. australis* reference genome that was successfully disassembled into two sets of reference genome-level haplotype subgenomes (Figure 1b). The genome contig N50 (33,936,801bp) and BUSCO (99.3%) assessments were much higher than those of the *P. australis* draft genome (contig N50: 3,219,705bp, BUSCO: 93.3%)[39]. The genome completeness at the chromosome level was far superior to that of the recently publicized LpPhrAust1.1 (https://www.ncbi.nlm.nih.gov/datasets/genome/GCA_958298935.1), and our genome is the most complete reference genome among published *P. australis* genomes (Table 1). The complexity of heterochromatin regions and repetitive sequences has resulted in centromere and telomere sequences being the most challenging regions in genome assembly[40-42]. We successfully identified telomeric and centromere sequences and candidate regions in our *P. australis* genome through the association of multi-omics data, which provided a genetic basis for the subsequent in-depth study of centromere and highly repetitive regions (Figure S3). Meanwhile, we combined multiple strategies to annotate structural genes, repetitive sequences, non-coding RNAs, and gene functions in the *P. australis* genome (Table 1, Figure S5). The high-quality genome assembly of *P. australis* provides a valuable genetic resource for exploring the molecular mechanisms underlying *P. australis*' large biomass, strong environmental adaptation and Lugen pharmacology.

B chromosome (Bs) analysis

As B chromosomes have been discovered and assembled in more species, their functions have been more deeply analyzed[11-18]. We assembled and characterized the Bs in *P. australis*, the first reported Bs in *P. australis*. Only 277 genes are present in this Bs, with 85% of the sequence consisting of transposable elements homologous to the A chromosome and multiple candidate centromere regions. Functional enrichment revealed that genes in Bs are involved in life processes related to chromosome segregation in meiosis, spindle formation, pollen sperm cell differentiation, and modifications related to histone methylation (Figure 2e). A fusion of chromosomal centromere breaks probably formed this Bs in *P. australis*. After passing through frequent gene exchange and gene rearrangement with the A chromosome group, genes related to the movement of Bs in cytokinesis or to the maintenance of Bs stability were retained. Due to the lack of genetic data on the *P. australis* genome, we believe that when more genomes of *P. australis* strains are assembled, the origin and function of *P. australis* Bs will be more accurately characterized.

LTR-RT insertion events have driven *P. australis* genome evolution and expansion.

It is well known that frequent climatic oscillations and glacial movements during the late Pleistocene (0.129-0.0117 Mya) brought dramatic changes to the distribution and genetic structure of plants and animals on a global scale and drove speciation and the creation of new

species[43-45]. Meanwhile, TE outbursts can bring more evolutionary raw materials for species, which is one of the important factors driving species evolution and new species divergence[46]. LTR-RTs can move through the genome and insert into new sites by a "copy and paste" transposition mechanism[47]. TEs with a large number of repetitive sequences can lead to gene inactivation, translocation, pseudogenes, and even chromosomal rearrangements due to their "active insertion" characteristics[48,49]. However, the genome size expansion driven by transposons counteracts the genome size reduction caused by the deletion of mutated genes to a certain extent. It maintains the stability of gene size[50]. The *P. australis* LTR-RTs were mainly composed of the Gypsy family, enriched in the *P. australis* centromere region, and the Copia family, which is consistent with the trend of gene distribution (Figure 1b, Figure S2). Further analysis revealed a strong positive correlation between the Copia family sequence content in the genome and genome size, and the LTR-RTs insertion events were species-specific (Figure 3c-d). This suggests that LTR-RT insertion events are widespread in Gramineae and strongly suggests that the Copia family plays a vital role in driving the rapid expansion of genome size and evolution of Gramineae species. Two LTR-RTs insertion events of different magnitude and type in *P. australis* during the Pleistocene drove the genome size expansion and subgenomic differentiation of *P. australis* on the one hand, and active TEs on the other hand, provided the raw material for a large number of gene mutations for rapid adaptation to the environment. The recent high density of LTR-RTs rapid insertion events in *P. australis* accounted for 55.42% of the total, and these uneliminated transposons may have contributed to the increase in the proportion of genes > 5000 bp in length and containing 2-3 introns in *P. australis* (Figure S6).

To reduce the impact of frequent transposon "jumps" on the genome, plants can not only directly inhibit the activation and mobilization of TE activity through DNA methylation[51-55] but also indirectly regulate heterochromatin ratios through histone methylation[56,57]. In this study, we found that the 5-methylcytosine site in *P. australis* was significantly enriched in the region around the centromere (Figure S3). Therefore, we hypothesize that after two widespread, high-density TE insertion events, LTR-RTs may have been modified by DNA methylation (5-mC) of transposons concentrated in the centromere region. This prevented the expression and "jumping" of which transposons had not yet been eliminated, thus maintaining the stability of the centromeric region of the genome.

Gene family expansion and WGD events

As we all know, telomere length and stability are essential in the biological life cycle. As the most difficult-to-repair, highly repetitive region in the genome, the existence and length maintenance of the telomere sequence are related to genome stability and cellular lifespan[58-61]. In *P. australis*, 41.26% of the gene families expanded during long-term evolution, mainly involving gene families related to telomere maintenance and lengthening or DNA repair, and these genes were subjected to positive selection simultaneously (Figure S13). This provided for telomere integrity and stability and played an essential role in the evolution of *P. australis*' extreme environmental adaptability and perennial characteristics.

Comparative genomic and phylogenetic analyses showed that *P. australis* of the Arundiaceae subfamily is sister to *C. songorica* of the Chloridoideae, consistent with the same previous study[39]. The *P. australis* genome undergoes two WGD events, a ρ event shared with most gramineous species and a heterologous quadruplication event (WGD-IV) that occurs at 23.9 Mya exclusive to *P. australis*. Large-scale chromosome fission and fusion events in the *P.*

australis genome have resulted in large segments of chromosome alterations and increases in chromosome number, while contributing to some extent to the evolution of the *P. australis* genome (Figure 4d).

Tissue-specific WGCNA analysis and identification of biosynthetic pathways of Lugen polysaccharides

The phragmitis rhizoma (known as "Lugen" in Chinese medicine) is a medicinal plant used clinically for more than 2000 years, and Lugen polysaccharides are its main medicinally active molecule[2-5]. Here, we constructed the Lugen polysaccharide (PRP-2) biosynthesis pathway involving the 182 transcripts (Figure 6a). By multi-tissue RNAseq analysis, we found that *P. australis* increased the expression of genes related to sucrose synthesis in the leaves and of sucrose transporter proteins responsible for the long-distance downward transport of sucrose in the aerial stems. Sugar content fluctuations resulting from the transport of plant carbohydrates over long distances to the roots can act as signal-regulating hormones or other signaling fluxes to induce the activation of apical meristematic tissues (axillary buds) and promote the development and extension of lateral and branching roots[62-68].The high expression of these genes not only promoted the accumulation of sucrose and *P. australis* polysaccharides in rhizomes but also significantly reduced the apical dominance inhibition of rhizome axillary buds and facilitated the development of the complex rhizome network of *P. australis*. In addition, the high expression of genes in the monomer synthesis pathway of Lugen polysaccharides increased the accumulation of raw materials for cell wall synthesis, laying a rich energy and material foundation for the rapid differentiation and development of perennial rhizomes in *P. australis*.

Meanwhile, we identified multiple action elements related to stress or hormone response and a MYB transcription factor binding site related to drought response within the promoter region of PaSUC1.1 (Figure 6b). A candidate MYB transcription factor that can regulate sucrose transporter protein was identified based on WGCNA analysis (Figure 6c-d). In conclusion, we hypothesize that MYB affects sucrose fluctuation signaling and rhizome development by regulating the expression of sucrose transporter proteins. This regulatory mechanism contributes to the rapid occupation of ecological niches by *P. australis* in various harsh environments and interspecific competition.

Materials and methods

Plant material

In this study, young leaf tissues of biennial *Phragmites australis* (Cav.) var. Cuiplus was collected for genome survey sequencing, HiFi sequencing, and Hi-C sequencing from a sample plot planted with reeds at Capital Normal University. At the same time, different tissues are used for full-length transcriptome sequencing (Nine types of tissues including flower, stem apical meristem, above-ground stems, leaves, above-ground stem buds, rhizome internodal tissues, rhizome nodal meristem, rhizome buds, and fibrous roots) and transcriptome sequencing (including mature leaves, above-ground stems, and rhizome tissues, in three biological replicates) were collected.

FISH

We first subjected root tip meristematic tissues to laughing gas treatment and obtained dispersed mid-stage chromosome cells using glacial acetic acid fixation and enzymatic digestion with a mixed enzyme solution (cellulase and pectinase 3:1). Subsequently, DAPI staining was used to obtain more explicit chromosome images and accurate numbers. Finally, the samples were subjected to fluorescence in situ hybridization based on a fluorescent probe for telomere-conserved repeats, 5SrDNA, and 18SrDNA universal probes, and photographed under an Olympus BX70 fluorescence microscope.

DNA extraction and genome sequencing

High-quality genomic DNA was extracted from young leaves of Phragmites australis using a modified CTAB method. The quality and quantity of the extracted DNA were examined using a NanoDrop 2000 spectrophotometer (NanoDrop Technologies, Wilmington, DE, USA), Qubit dsDNA HS Assay Kit on a Qubit 3.0 Fluorometer (Life Technologies, Carlsbad, CA, USA) and electrophoresis on a 0.8% agarose gel, respectively. This was followed by BGI short reads, PacBio subreads, and Hi-C interactive reads, which were sequenced. First, we performed paired-end reads with an insert size of 150 bp sequencing using the BGI T7 sequencing platform and obtained 47.74Gb clean reads data after filtering and cleaning by SOAPnuke[69] software. The basic genomic information such as genome size, heterozygosity, the proportion of repetitive sequences, and other genomic information was obtained by K-mer frequency distribution analysis with k = 17 by jellyfish[70,71] and genomescope software (https://github.com/tbenavi1/genomescope2.0). Subsequently, the SMRTbell library was constructed using the SMRTbell Express Template Prep kit 2.0 (Pacific Biosciences). Pacbio High-fidelity (HiFi) reads were performed using SMRT Cell on the Sequel II System with Sequel II Sequencing Kit by Frasergen Bioinformatics Co., Ltd. (Wuhan, China). Finally, the Hi-C library was constructed according to previous studies[72]. The quality of the library was ensured by using the Q-PCR method. The Hi-C libraries were quantified and sequenced on the BGISEQ-500 platform (BGI, China). After dejointing the raw sequencing data using Trimmomatic software and filtering the low-quality Reads, we ended up with 113.26 GB of clean data for genome-assisted assembly.

The consensus reads they were generated using ccs software (https://github.com/pacificbiosciences/unanimity) with the parameter '-minPasses 3'. After quality control, we obtained 32.53 Gbps of HiFi reads data. These long (~15 kb) and highly accurate (>99%) HiFi reads were assembled using hifiasm v0.16.1[73] with the parameter '-l3', yielding a total of 312 contigs containing the initial assembled genome. Then, we applied 3D-

DNA to order and orient the clustered contigs. The Juicer[71] was used to filter and cluster the sequences, and the Juicebox was applied to adjust chromosome construction manually. We formed a Hi-C-assisted pre-assembly genome sketch sequence containing 79 Contigs after heterozygous and redundant sequence filtering based on Hi-C interaction signals and NT comparison results. Next, by examining allelic interactions in the Hi-C heatmap, the order and orientation of alleles on the pseudochromosomes were evaluated and adjusted. Finally, we anchored 79 contigs to 25 chromosomes, and the effective mounting rate of Contig was 99.46%.

Identification of telomeres, centromere and 5-methylcytosine

We used tidk (0.2.3)[74] to search the genome for telomeric repeat sequences from length 5 to length 12, and the most abundant repeat sequence units were counted and visualized. The identification of centromere was based on the method described by Shi et al[35], where we filtered the *P. australis* genome for repetitive sequences contained in the genome (filtered condition: period >= 50, copies >= 2.0 To obtain 5-methylcytosine information from Pacbio HiFi sequencing data, we first generated reads with 5mC tags from HiFi reads using jasmine (v 2.0.0, https://github.com/PacificBiosciences/jasmine) and used IGV to find centromere sequences[75]. Next, soft clip mapped reads with methylation tags using pbmm2 (v 1.13.1, https://github.com/PacificBiosciences/pbmm2). Finally, information on the 5mC locus in the *P. australis* genome was analyzed using pb-CpG-tools (v2.3.2, https://github.com/PacificBiosciences/pb-CpG-tools) and visualized using IGV[75].

allopolyploid subgenome phasing

First, we used blast[76] (-evalue ≤1e-10, -num_alignments = 10) to identify paralogous homologous genes within the Phragmites australis genome, orthologous homologous genes between Phragmites australis and Oryza sativa, and orthologous genes between Phragmites australis and Panicum virgatum. Next, we identified homologous chromosome pairs in the genomes of different species by JCVI (v0.5.7, https://github.com/tanghaibao/jcvi/)[77] and WGDI (v0.6.1)[78] and determined the blocks of colinearity between genomes. Based on the colinearity results, we extracted single-copy genes in the homologous chromosomes between Phragmites australis-Oryza sativa and Phragmites australis-Panicum virgatum, respectively, for sequence splicing and genetic distance estimation. Finally, with the Panicum virgatum outgroup, we used RAxML (v.8.2.X)[79] software to construct phylogenetic trees for the chromosomes of the three species. We classified the Phragmites australis chromosome, which is closer to the homologous chromosome of Oryza sativa, to subgenome A and the other homologous chromosome to subgenome B.

Evaluation of assembled genomes

We used multiple software and assembly metrics to assess our assembled genomes' completeness, accuracy, and consistency. First, we evaluated the assembly quality of our genome by calculating Length, N50, L50, GC content, and other metrics using QUAST (Quality Assessment Tool for Genome Assemblies)[80]. Next, we mapped the genome survey and Iso-seq sequencing data to the assembled genome using minimap2 software and assessed the genome integrity[81]. The uniformity of sequencing coverage and contamination in the sequencing data are based on the ratio of reads, coverage, and the distribution of the GC content with the sequencing depth. Subsequently, we evaluated the accuracy and consistency of the genomes by calculating the genome consensus quality value (QV) values using Merqury[82] software. We used BWA[83] to map the second-generation sequencing data onto the assembled reed genome,

followed by GATK (GATK: https://github.com/broadinstitute/gatk) for SNP calling and filtering, and evaluated the accuracy of the genome by counting the number of homozygous and heterozygous SNPs. Finally, we used BUSCO (Benchmarking Universal Single-Copy Orthologs)[84,85] based on the Embryophyta_odb10 gene set and CEGMA (Core eukaryotic genes mapping approach)[86] based on 248 ultra-conserved core eukaryotic genes (CEGs) to assess genomic accuracy and completeness. To more accurately assess the coherence of our assembled genomes, we first used the combination of LTR_Finder[87] (-D 15000 -d 1000 -L 7000 -l 100 -p 20 -C -M 0.9) and LTR_harvest[88] (-similar 90 -vic 10 -seed 20 -seqids yes -minlenltr 100 -maxlenltr 7000 -mintsd 4 -maxtsd 6 -motif TGCA -motifmis 1) to identify long terminal repeats (LTR) sequences in our assembled whole genomes and two sets of subgenomes, respectively, and then used LTR_retreiver[89] (-u parameter using the evolutionary rate of rice: 1.3e-8) to integrate the results and to calculate the LTR Assembly Index (LAI)[90] and LTR density, and finally the genomic LAI was visualized by ggplot2[91].

Gene prediction and functional annotation

To identify the repeat contents in the Phragmites australis genome more accurately and comprehensively, we combined two strategies of homology-based prediction and ab initio prediction to identify repetitive sequences in the *P. australis* genome.

Homolog-based approach: Based on RepBase[92,93], a database of known repetitive sequences (http://www.girinst.org/repbase), we used RepeatMasker (v4.0.7, www.repeatmasker.org) and RepeatProteinMask software to predict sequences that are similar to known repetitive sequences. Sequences that are similar to known repeats. Ab initio approach: first, an ab initio repetitive sequence library (http://www.repeatmasker.org/RepeatModeler/) was constructed using RepeatModeler software and LTR-FINDER[87], and subsequently, an ab initio repetitive sequence library was constructed by the RepeatMasker software for repeat sequence prediction. In addition, we predicted the Tandem Repeat in the reed genome using TRF software[94] and identified the SSR sites present in the reed genome by MISA[95].

We accurately predicted the gene structure of the *P. australis* genome by combining three strategies. Firstly, we compared the coding protein sequence of *P. australis* related species (Panicum virgatum, Setaria italica, Setaria viridis, Sorghum bicolor) with the genome sequence, and obtained the gene structure information of homology-based prediction using Exonerate[96]. Next, we obtained gene structure information for transcriptome-based prediction by mapping the Iso-Seq data to the genome and using PASA[97] to determine the gene's shear sites and exon regions. Finally, we performed de novo prediction of gene structures by AUGUSTUS[98] and GlimmHMM[99] software. Integrating the above prediction results, we used MAKER2[100] to combine the gene structure information predicted by various methods into a non-redundant and more complete gene set. We used PASA[97] to update the gene structures with the transcriptome data.

Functional annotation information of the genes was obtained by homologous search of the predicted gene set in several databases, including SwissProt[101], NR (ftp://ftp.ncbi.nlm.nih.gov/blast/db/FASTA/nr.gz), PFAM[102,103], GO[104], KEGG[105], InterPro[106], TrEMBL[107].

Non-coding RNA prediction: based on the structural characteristics of tRNAs, tRNAscan-SE[108] was used to find tRNA sequences in the genome; based on the highly conserved nature of rRNAs, rRNA sequences of closely related species can therefore be selected as reference

sequences and rRNAs in the genome can be found by BLASTN comparison; in addition, using the Rfam[109] family of covariance models to predict miRNA and snRNA sequence information using INFERNAL[110].

Gene family and phylogenetic analysis

We collected and organized protein sequences from Phragmites australis and 14 species, including Arabidopsis thaliana, Aegilops tauschii, Brachypodium distachyon, Cleistogenes songorica, Dendrocalamus latiflorus Munro, Oryza sativa, Panicum hallii, Panicum virgatum, Pennisetum purpureum Schum, Setaria italica, Setaria viridis, sorghum bicolor, Triticum aestivum, and Zea mays, for gene family identification and evolutionary analysis.

First, we performed a comprehensive protein sequence blastp analysis (E-value ≤ 1e-5) using diamond[111]. Subsequently, the blastp results were clustered using OrthoFinder2[112] for immediate homologous gene finding and Orthogroup construction, and single-copy and multicopy gene families were obtained. To perform an accurate multi-species phylogenetic relationship analysis, we filtered the single-copy immediate homologous gene families shared by all species. We retained only those genes with amino acid lengths ≥100. Next, multiple sequence comparisons were performed separately for genes within each single-copy homologous gene family using MUSCLE[113]. Next, multiple sequence alignment results were integrated and converted into super-gene alignment in phylip format. Finally, a phylogenetic tree was constructed by Maximum Likelihood using RAxML[79] with Arabidopsis thaliana as the outgroup.

Based on obtaining time-corrected points based on fossil evidence at the TimeTree[114] website and in the literature, we used r8s[115] and the mcmctree program in the PAML[116] package (based on the Bayesianrelaxed molecular clock approach) to estimate divergence times for species. According to the results of gene families and phylogenetic trees, CAFÉ[117] was used to predict the expansion or contraction of gene families in different species in each evolutionary branch. Based on obtaining organisms each sharing a single-copy orthologous homologous gene family member, we used the Codeml program in the PAML[116] package (using the branch-site model) to test whether the gene was under positive selection.

Whole genome doubling (WGD) and karyotype evolution analysis

We performed blastp[76] comparisons between multiple species genomes to obtain orthologous pairwise (best comparison results for mutual blastp). Subsequently, The MCscan (https://github.com/tanghaibao/jcvi/)[77] was utilized to search for covariate segments between species genomes, and 4dTv values and Ks values were calculated for gene pairs contained in the covariate segments. We used the time of divergence of the evolutionary trees of *P. australis* and Cleistogenes songorica (Ks = 0.305, T = 34.6 mya) to infer the evolutionary rate μ according to the formula Ks = 2 μT, after which we calculated the time of divergence between species and the time of occurrence of the species WGDs based on μ.

We updated the ancestral genome sequence of the Ancestral monocot karyotype except for Acoraceae (AMK-A) obtained from WGDI using the genetic information of *Sorghum bicolor, Triticum aestivum, Oryza sativa, Cleistogenes songorica, Phragmites australis, Panicum hallii, and Setaria viridis* by the parameter "-akr", and the genetic information in reed by the parameter "-km" visualization of genome karyotypes in *P. australis*.

Transcriptome sequencing and WGCNA analysis

Sequencing libraries were constructed from RNA extracted from different tissues of *P. australis*

(mature leaves, aerial stems, and rhizome tissues) and sequenced using the BGI T7 platform for PE150 short reads. After filtering by fastp (v0.21.0)[118], we mapped different tissue clean reads to the *P. australis* genome using HISAT2 (v2.2.1)[119] and quantified reads by featureCounts (v2.0.3)[120]. We used FPKM to indicate the expression value of each gene. We used FPKM to indicate the expression value of each gene. Subsequently, we performed a weighted gene co-expression network analysis of the average FPKM > 1 gene in all tissues using the WGCNA package (v1.72-1) in R software[121]. Tissue-specific gene clusters were identified based on module-tissue correlations (Pearson |cor| > 0.95). Genes in the obtained tissue-specific modules were analyzed for GO and KEGG enrichment using the clusterProfiler (v4.0) package[122].

Identification of Lugen polysaccharide biosynthetic pathway genes and prediction of transcription factors regulating *PaSUC* genes

We screened the structural genes related to polysaccharide biosynthetic pathways by filtering them in the functional annotation results of structural genes. Then, we reconfirmed the screening results by identifying specific structural domains in these structural genes by HMMER[123]. To obtain the Hub genes in the tissue-specific module, the genes with weight > 0.4 were further analyzed by the Radiality algorithm in the CytoHubba plugin. We defined the Top 100 genes as Hub genes and then visualized the Hub gene co-expression network using Cytoscape. Transcription factor prediction was performed by submitting the protein sequences of the genes in the module to the Plant TFDB (v5.0) database[124]. Cis-acting element prediction was performed by uploading the sequences within 2000 bp upstream of the *PaSUC* gene into the PlantCARE database[125]. Genomic circos and gene expression and cis-acting elements were visualized by TBtools (v2.03)[126].

Acknowledgments

This study was accomplished under the financial support of the National Natural Science Foundation of China (31972934, 31170784) and the Capital Normal University's Capacity Building of Science and Technology Innovation Service-Basic Scientific Research Operating Expenses (No.19530050183). We thank Xikun Wu and Zhiqiang Wang for their help in material collection. Thanks to Wuhan Fraser Genetic Information Co. for help in genome sequencing and assembly.

Data availability statement

Whole genome and transcriptome raw sequencing data used in this study have been deposited at the National Center for Biotechnology Information (NCBI) under accession number PRJNA1055898 (Reviewer link : https://dataview.ncbi.nlm.nih.gov/object/PRJNA1055898?reviewer=f69iak9ecmtpjpqq8514vruin4). The genome assembly and gene annotation data of *P. australis* have been deposited in the China National Center for Bioinformation (CNCB) under project number PRJCA022478.

Conflict of interests

The authors declare no conflict of interest.

Authors' Contributions

JPC participated in the experimental design and was responsible for bioinformatics data analysis and manuscript writing. Manuscript proofreading and reference organization by RW, RQG and MHC. RW, RQG, and ZYW were responsible for the management and sampling of

the plant samples. LL and JMH participated in genomic survey analysis. SXC was responsible for this study's supervision, design, writing review, and editing. All authors read and approved the final manuscript.

References

1 Ren, Y. et al. Traditional Uses, Phytochemistry, Pharmacology and Toxicology of Rhizoma phragmitis: A Narrative Review. Chinese Journal of Integrative Medicine 28, 1127-1136, doi:10.1007/s11655-022-3572-1 (2022).

2 Sohaib, M., Al-Barakah, F. N. I., Migdadi, H. M. & Husain, F. M. Comparative study among Avicennia marina, Phragmites australis, and Moringa oleifera based ethanolic-extracts for their antimicrobial, antioxidant, and cytotoxic activities. Saudi J Biol Sci 29, 111-122, doi:10.1016/j.sjbs.2021.08.062 (2022).

3 Zhou, R. et al. Isolation, structure identification and anti-inflammatory activity of a polysaccharide from Phragmites rhizoma. International Journal of Biological Macromolecules 161, 810-817, doi:10.1016/j.ijbiomac.2020.06.124 (2020).

4 Cui, M. et al. Phragmites rhizoma polysaccharide-based nanocarriers for synergistic treatment of ulcerative colitis. International Journal of Biological Macromolecules 220, 22-32, doi:10.1016/j.ijbiomac.2022.07.245 (2022).

5 Jung, T. W. et al. The aqueous extract of Phragmites rhizome improves hepatic steatosis in obese mice via the AMPK-mediated inhibition of ER stress. J Funct Foods 95, doi:ARTN 10516410.1016/j.jff.2022.105164 (2022).

6 Zhu, L. et al. Anti-Inflammatory and antiviral effects of water-soluble crude extract from Phragmites australis in vitro. Pak J Pharm Sci 30, 1357-1362 (2017).

7 Zhou, M. E. et al. Xuanfei Baidu Decoction regulates NETs formationCXCL2/CXCR2 signaling pathway that is involved in acute lung injury. Biomed Pharmacother 161, doi:ARTN 11453010.1016/j.biopha.2023. 114530 (2023).

8 Pan, X., Dong, L., Yang, L., Chen, D. & Peng, C. Potential drugs for the treatment of the novel coronavirus pneumonia (COVID-19) in China. Virus Res 286, 198057, doi:10.1016/j.virusres.2020.198057 (2020).

9 D'Ambrosio, U. et al. B-chrom: a database on B-chromosomes of plants, animals and fungi. New Phytol 216, 635-642, doi:10.1111/nph.14723 (2017).

10 Ramos, É. et al. The repetitive DNA element BncDNA, enriched in the B chromosome of the cichlid fish Astatotilapia latifasciata, transcribes a potentially noncoding RNA. Chromosoma 126, 313-323, doi:10.1007/s00412-016-0601-x (2017).

11 Banaei-Moghaddam, A. M. et al. Nondisjunction in favor of a chromosome: the mechanism of rye B chromosome drive during pollen mitosis. Plant Cell 24, 4124-4134, doi:10.1105/tpc.112.105270 (2012).

12 Camacho, J. P. M. Non-Mendelian segregation and transmission drive of B chromosomes. Chromosome Res 30, 217-228, doi:10.1007/s10577-022-09692-7 (2022).

13 Akera, T. et al. Spindle asymmetry drives non-Mendelian chromosome segregation.

Science 358, 668-+, doi:10.1126/science.aan0092 (2017).
14 Aldrich, J. C., Leibholz, A., Cheema, M. S., Ausió, J. & Ferree, P. M. A 'selfish' B chromosome induces genome elimination by disrupting the histone code in the jewel wasp Nasonia vitripennis. Sci Rep-Uk 7, doi:ARTN 4255110.1038/srep42551 (2017).
15 Dalla Benetta, E. et al. Genome elimination mediated by gene expression from a selfish chromosome. Sci Adv 6, eaaz9808, doi:10.1126/sciadv.aaz9808 (2020).
16 Huang, W., Du, Y., Zhao, X. & Jin, W. B chromosome contains active genes and impacts the transcription of A chromosomes in maize (Zea mays L.). BMC Plant Biol 16, 88, doi:10.1186/s12870-016-0775-7 (2016).
17 Banaei-Moghaddam, A. M. et al. Genes on B chromosomes: old questions revisited with new tools. Biochim Biophys Acta 1849, 64-70, doi:10.1016/j.bbagrm.2014.11.007 (2015).
18 Martis, M. M. et al. Selfish supernumerary chromosome reveals its origin as a mosaic of host genome and organellar sequences. P Natl Acad Sci USA 109, 13343-13346, doi:10.1073/pnas.1204237109 (2012).
19 Santin-Montanya, M. I., Jimenez, J., Vilan, X. M. & Ocana, L. Effects of size and moisture of rhizome on initial invasiveness ability of giant reed. J Environ Sci Health B 49, 41-44, doi:10.1080/03601234.2013.836881 (2014).
20 Zhai, S. S. et al. Effect of Rhizome Severing on Survival and Growth of Rhizomatous Herb Is Regulated by Sand Burial Depth. Plants-Basel 11, doi:ARTN 319110.3390/plants11233191 (2022).
21 Fan, Z., Huang, G., Fan, Y. & Yang, J. Sucrose Facilitates Rhizome Development of Perennial Rice (Oryza longistaminata). Int J Mol Sci 23, doi:10.3390/ijms232113396 (2022).
22 Bessho-Uehara, K., Nugroho, J. E., Kondo, H., Angeles-Shim, R. B. & Ashikari, M. Sucrose affects the developmental transition of rhizomes in Oryza longistaminata. J Plant Res 131, 693-707, doi:10.1007/s10265-018-1033-x (2018).
23 Fan, Z. Q., Cai, Z. Q., Shan, J. W. & Yang, J. Y. Letter to the Editor: Bud Position and Carbohydrate Play a More Significant Role than Light Condition in the Developmental Transition between Rhizome Buds and Aerial Shoot Buds of. Plant Cell Physiol 58, 1281-1282, doi:10.1093/pcp/pcx061 (2017).
24 Mason, M. G., Ross, J. J., Babst, B. A., Wienclaw, B. N. & Beveridge, C. A. Sugar demand, not auxin, is the initial regulator of apical dominance. Proc Natl Acad Sci USA 111, 6092-6097, doi:10.1073/pnas.1322045111 (2014).
25 Henry, C. et al. Regulation of RhSUC2, a sucrose transporter, is correlated with the light control of bud burst in Rosa sp. Plant Cell Environ 34, 1776-1789, doi:10.1111/j.1365-3040.2011.02374.x (2011).
26 Kebrom, T. H. et al. Inhibition of Tiller Bud Outgrowth in the tin Mutant of Wheat Is Associated with Precocious Internode Development. Plant Physiology 160, 308-318, doi:10.1104/pp.112.197954 (2012).

27 Yuan, Y., Khourchi, S., Li, S., Du, Y. & Delaplace, P. Unlocking the Multifaceted Mechanisms of Bud Outgrowth: Advances in Understanding Shoot Branching. Plants (Basel) 12, doi:10.3390/plants12203628 (2023).
28 Guo, W. J., Pommerrenig, B., Neuhaus, H. E. & Keller, I. Interaction between sugar transport and plant development. J Plant Physiol 288, doi:ARTN 15407310.1016/j.jplph.2023.154073 (2023).
29 Wang, J. et al. A high-quality genome assembly of Morinda officinalis, a famous native southern herb in the Lingnan region of southern China. Hortic Res 8, 135, doi:10.1038/s41438-021-00551-w (2021).
30 Li, Y., Wang, Z., Zhu, M. et al. A chromosome-scale Rhubarb (Rheum tanguticum) genome assembly provides insights into the evolution of anthraquinone biosynthesis. Commun Biol 6, 867. https://doi.org/10.1038/s42003-023-05248-5.(2023)
31 Kang, M. et al. A chromosome-scale genome assembly of Isatis indigotica, an important medicinal plant used in traditional Chinese medicine: An Isatis genome. Hortic Res 7, 18, doi:10.1038/s41438-020-0240-5 (2020).
32 Xi Zhang, T. Q., Anan Wang, Huajian Zhou, Min Yuan, Li Li, Sulan Bai, Suxia Cui*. Morphology and Genetic Diversity of Phragmites australis in Beijing. Chinese Bulletin of Botany, doi:10.11983/CBB20006 (2020).
33 Mo, W. et al. Single-molecule targeted accessibility and methylation sequencing of centromeres, telomeres and rDNAs in Arabidopsis. Nat Plants 9, 1439-1450, doi:10.1038/s41477-023-01498-7 (2023).
34 Naish, M. et al. The genetic and epigenetic landscape of the Arabidopsis centromeres. Science 374, eabi7489, doi:10.1126/science.abi7489 (2021).
35 Shi, X. et al. The complete reference genome for grapevine (Vitis vinifera L.) genetics and breeding. Hortic Res 10, uhad061, doi:10.1093/hr/uhad061 (2023).
36 Zhang, L. et al. A near-complete genome assembly of Brassica rapa provides new insights into the evolution of centromeres. Plant Biotechnol J, doi:10.1111/pbi.14015 (2023).
37 Cho, A. et al. An improved Raphanus sativus cv. WK10039 genome localizes centromeres, uncovers variation of DNA methylation and resolves arrangement of the ancestral Brassica genome blocks in radish chromosomes. Theor Appl Genet 135, 1731-1750, doi:10.1007/s00122-022-04066-3 (2022).
38 Chang, X. et al. High-quality Gossypium hirsutum and Gossypium barbadense genome assemblies reveal the landscape and evolution of centromeres. Plant Commun, 100722, doi:10.1016/j.xplc.2023.100722 (2023).
39 Oh, D. H. et al. Novel genome characteristics contribute to the invasiveness of Phragmites australis (common reed). Mol Ecol 31, 1142-1159, doi:10.1111/mec.16293 (2022).
40 Huang, Y. et al. The formation and evolution of centromeric satellite repeats in Saccharum species. Plant J 106, 616-629, doi:10.1111/tpj.15186 (2021).
41 Li, Y. et al. Centromeric DNA characterization in the model grass Brachypodium

distachyon provides insights on the evolution of the genus. Plant J 93, 1088-1101, doi:10.1111/tpj.13832 (2018).
42 Zhang, H. et al. Boom-Bust Turnovers of Megabase-Sized Centromeric DNA in Solanum Species: Rapid Evolution of DNA Sequences Associated with Centromeres. Plant Cell 26, 1436-1447, doi:10.1105/tpc.114.123877 (2014).
43 Xu, X. X. et al. Late Pleistocene speciation of three closely related tree peonies endemic to the Qinling-Daba Mountains, a major glacial refugium in Central China. Ecol Evol 9, 7528-7548, doi:10.1002/ece3.5284 (2019).
44 Dong, L., Heckel, G., Liang, W. & Zhang, Y. Phylogeography of Silver Pheasant (Lophura nycthemera L.) across China: aggregate effects of refugia, introgression and riverine barriers. Mol Ecol 22, 3376-3390, doi:10.1111/mec.12315 (2013).
45 Cohen, K. M., Finney, S. C., Gibbard, P. L. & Fan, J. X. The ICS International Chronostratigraphic Chart. Episodes 36, 199-204, doi:DOI 10.18814/epiiugs/2013/v36i3/002 (2013).
46 Galindo-Gonzalez, L., Mhiri, C., Deyholos, M. K. & Grandbastien, M. A. LTR-retrotransposons in plants: Engines of evolution. Gene 626, 14-25, doi:10.1016/j.gene.2017.04.051 (2017).
47 Mc, C. B. The origin and behavior of mutable loci in maize. Proc Natl Acad Sci USA 36, 344-355, doi:10.1073/pnas.36.6.344 (1950).
48 Schulman, A. H. Retrotransposon replication in plants. Curr Opin Virol 3, 604-614, doi:10.1016/j.coviro.2013.08.009 (2013).
49 Li, Q. et al. A D-genome-originated Ty1/Copia-type retrotransposon family expanded significantly in tetraploid cottons. Molecular Genetics and Genomics 293, 33-43, doi:10.1007/s00438-017-1359-4 (2018).
50 Zhou, S. S. et al. A comprehensive annotation dataset of intact LTR retrotransposons of 300 plant genomes. Sci Data 8, 174, doi:10.1038/s41597-021-00968-x (2021).
51 Cui, X. & Cao, X. Epigenetic regulation and functional exaptation of transposable elements in higher plants. Curr Opin Plant Biol 21, 83-88, doi:10.1016/j.pbi.2014.07.001 (2014).
52 Lisch, D. Epigenetic regulation of transposable elements in plants. Annu Rev Plant Biol 60, 43-66, doi:10.1146/annurev.arplant.59.032607.092744 (2009).
53 Hirochika, H., Okamoto, H. & Kakutani, T. Silencing of retrotransposons in Arabidopsis and reactivation by the ddm1 mutation. Plant Cell 12, 357-369, doi:10.1105/tpc.12.3.357 (2000).
54 Miura, A. et al. Mobilization of transposons by a mutation abolishing full DNA methylation in Arabidopsis. Nature 411, 212-214, doi:10.1038/35075612 (2001).
55 Liu, B. B. & Zhao, M. X. How transposable elements are recognized and epigenetically silenced in plants? Current Opinion in Plant Biology 75, doi:ARTN 10242810.1016/j.pbi.2023.102428 (2023).

56 Peters, A. H. & Schubeler, D. Methylation of histones: playing memory with DNA. Curr Opin Cell Biol 17, 230-238, doi:10.1016/j.ceb.2005.02.006 (2005).

57 Takatsuka, H. & Umeda, M. Epigenetic Control of Cell Division and Cell Differentiation in the Root Apex. Front Plant Sci 6, doi:ARTN 117810.3389/fpls.2015.01178 (2015).

58 Aguado, J., d'Adda di Fagagna, F. & Wolvetang, E. Telomere transcription in ageing. Ageing Res Rev 62, 101115, doi:10.1016/j.arr.2020.101115 (2020).

59 Kalmykova, A. Telomere Checkpoint in Development and Aging. Int J Mol Sci 24, doi:10.3390/ijms242115979 (2023).

60 Stroik, S. & Hendrickson, E. A. Telomere replication-When the going gets tough. DNA Repair (Amst) 94, 102875, doi:10.1016/j.dnarep.2020.102875 (2020).

61 Vaiserman, A. & Krasnienkov, D. Telomere Length as a Marker of Biological Age: State-of-the-Art, Open Issues, and Future Perspectives. Front Genet 11, doi:ARTN 63018610.3389/fgene.2020.630186 (2021).

62 Valifard, M. et al. Vacuolar fructose transporter SWEET17 is critical for root development and drought tolerance. Plant Physiology 187, 2716-2730, doi:10.1093/plphys/kiab436 (2021).

63 Stevenson, C. C. & Harrington, G. N. The impact of supplemental carbon sources on Arabidopsis thaliana growth, chlorophyll content and anthocyanin accumulation. Plant Growth Regul 59, 255-271, doi:10.1007/s10725-009-9412-x (2009).

64 Xiong, Y. et al. Glucose-TOR signalling reprograms the transcriptome and activates meristems. Nature 496, 181-186, doi:10.1038/nature12030 (2013).

65 Okooboh, G. O. et al. Overexpression of the vacuolar sugar importer BvTST1 from sugar beet in Camelina improves seed properties and leads to altered root characteristics. Physiol Plantarum 174, doi:ARTN e1365310.1111/ppl.13653 (2022).

66 Riou-Khamlichi, C., Menges, M., Healy, J. M. & Murray, J. A. Sugar control of the plant cell cycle: differential regulation of Arabidopsis D-type cyclin gene expression. Mol Cell Biol 20, 4513-4521, doi:10.1128/MCB.20.13.4513-4521.2000 (2000).

67 Fichtner, F. et al. Trehalose 6-phosphate is involved in triggering axillary bud outgrowth in garden pea (Pisum sativum L.). Plant Journal 92, 611-623, doi:10.1111/tpj.13705 (2017).

68 Salam, B. B. et al. Sucrose promotes stem branching through cytokinin. Plant Physiology 185, 1708-1721, doi:10.1093/plphys/kiab003 (2021).

69 Chen, Y. et al. SOAPnuke: a MapReduce acceleration-supported software for integrated quality control and preprocessing of high-throughput sequencing data. Gigascience 7, 1-6, doi:10.1093/gigascience/gix120 (2018).

70 Zimin, A. V. et al. The MaSuRCA genome assembler. Bioinformatics 29, 2669-2677, doi:10.1093/bioinformatics/btt476 (2013).

71 Marcais, G. & Kingsford, C. A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. Bioinformatics 27, 764-770, doi:10.1093/bioinformatics/btr011 (2011).

72 Padmarasu, S., Himmelbach, A., Mascher, M. & Stein, N. In Situ Hi-C for Plants: An Improved Method to Detect Long-Range Chromatin Interactions. Methods Mol Biol 1933, 441-472, doi:10.1007/978-1-4939-9045-0_28 (2019).
73 Cheng, H., Concepcion, G. T., Feng, X., Zhang, H. & Li, H. Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. Nat Methods 18, 170-175, doi:10.1038/s41592-020-01056-5 (2021).
74 Brown, M., González De la Rosa, P. M. and Mark, B. A Telomere Identification Toolkit. Zenodo, doi:10.5281/zenodo.10091385 (2023).
75 Thorvaldsdottir, H., Robinson, J. T. & Mesirov, J. P. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. Brief Bioinform 14, 178-192, doi:10.1093/bib/bbs017 (2013).
76 Camacho, C. et al. BLAST+: architecture and applications. BMC Bioinformatics 10, 421, doi:10.1186/1471-2105-10-421 (2009).
77 Tang, H. et al. Synteny and collinearity in plant genomes. Science 320, 486-488, doi:10.1126/science.1153917 (2008).
78 Sun, P. et al. WGDI: A user-friendly toolkit for evolutionary analyses of whole-genome duplications and ancestral karyotypes. Mol Plant 15, 1841-1851, doi:10.1016/j.molp.2022.10.018 (2022).
79 Stamatakis, A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30, 1312-1313, doi:10.1093/bioinformatics/btu033 (2014).
80 Gurevich, A., Saveliev, V., Vyahhi, N. & Tesler, G. QUAST: quality assessment tool for genome assemblies. Bioinformatics 29, 1072-1075, doi:10.1093/bioinformatics/btt086 (2013).
81 Li, H. Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics 34, 3094-3100, doi:10.1093/bioinformatics/bty191 (2018).
82 Rhie, A., Walenz, B. P., Koren, S. & Phillippy, A. M. Merqury: reference-free quality, completeness, and phasing assessment for genome assemblies. Genome Biol 21, 245, doi:10.1186/s13059-020-02134-9 (2020).
83 Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25, 1754-1760, doi:10.1093/bioinformatics/btp324 (2009).
84 Seppey, M., Manni, M. & Zdobnov, E. M. BUSCO: Assessing Genome Assembly and Annotation Completeness. Methods Mol Biol 1962, 227-245, doi:10.1007/978-1-4939-9173-0_14 (2019).
85 Simao, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V. & Zdobnov, E. M. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics 31, 3210-3212, doi:10.1093/bioinformatics/btv351 (2015).
86 Parra, G., Bradnam, K. & Korf, I. CEGMA: a pipeline to accurately annotate core genes in eukaryotic genomes. Bioinformatics 23, 1061-1067,

doi:10.1093/bioinformatics/btm071 (2007).
87 Xu, Z. & Wang, H. LTR_FINDER: an efficient tool for the prediction of full-length LTR retrotransposons. Nucleic Acids Res 35, W265-268, doi:10.1093/nar/gkm286 (2007).
88 Ellinghaus, D., Kurtz, S. & Willhoeft, U. LTRharvest, an efficient and flexible software for de novo detection of LTR retrotransposons. BMC Bioinformatics 9, 18, doi:10.1186/1471-2105-9-18 (2008).
89 Ou, S. & Jiang, N. LTR_retriever: A Highly Accurate and Sensitive Program for Identification of Long Terminal Repeat Retrotransposons. Plant Physiol 176, 1410-1422, doi:10.1104/pp.17.01310 (2018).
90 Ou, S., Chen, J. & Jiang, N. Assessing genome assembly quality using the LTR Assembly Index (LAI). Nucleic Acids Res 46, e126, doi:10.1093/nar/gky730 (2018).
91 Ito, K. & Murphy, D. Application of ggplot2 to Pharmacometric Graphics. CPT Pharmacometrics Syst Pharmacol 2, e79, doi:10.1038/psp.2013.56 (2013).
92 Bao, W., Kojima, K. K. & Kohany, O. Repbase Update, a database of repetitive elements in eukaryotic genomes. Mob DNA 6, 11, doi:10.1186/s13100-015-0041-9 (2015).
93 Jurka, J. et al. Repbase Update, a database of eukaryotic repetitive elements. Cytogenet Genome Res 110, 462-467, doi:10.1159/000084979 (2005).
94 Benson, G. Tandem repeats finder: a program to analyze DNA sequences. Nucleic Acids Res 27, 573-580, doi:10.1093/nar/27.2.573 (1999).
95 Thiel, T., Michalek, W., Varshney, R. K. & Graner, A. Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (Hordeum vulgare L.). Theor Appl Genet 106, 411-422, doi:10.1007/s00122-002-1031-0 (2003).
96 Slater, G. S. & Birney, E. Automated generation of heuristics for biological sequence comparison. BMC Bioinformatics 6, 31, doi:10.1186/1471-2105-6-31 (2005).
97 Haas, B. J. et al. Improving the Arabidopsis genome annotation using maximal transcript alignment assemblies. Nucleic Acids Res 31, 5654-5666, doi:10.1093/nar/gkg770 (2003).
98 Stanke, M. et al. AUGUSTUS: ab initio prediction of alternative transcripts. Nucleic Acids Res 34, W435-439, doi:10.1093/nar/gkl200 (2006).
99 Majoros, W. H., Pertea, M. & Salzberg, S. L. TigrScan and GlimmerHMM: two open source ab initio eukaryotic gene-finders. Bioinformatics 20, 2878-2879, doi:10.1093/bioinformatics/bth315 (2004).
100 Holt, C. & Yandell, M. MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects. BMC Bioinformatics 12, 491, doi:10.1186/1471-2105-12-491 (2011).
101 Bairoch, A. & Apweiler, R. The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000. Nucleic Acids Res 28, 45-48, doi:10.1093/nar/28.1.45 (2000).
102 El-Gebali, S. et al. The Pfam protein families database in 2019. Nucleic Acids Res 47, D427-D432, doi:10.1093/nar/gky995 (2019).

103 Mistry, J. et al. Pfam: The protein families database in 2021. Nucleic Acids Res 49, D412-D419, doi:10.1093/nar/gkaa913 (2021).
104 Ashburner, M. et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat Genet 25, 25-29, doi:10.1038/75556 (2000).
105 Kanehisa, M. & Goto, S. KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res 28, 27-30, doi:10.1093/nar/28.1.27 (2000).
106 Blum, M. et al. The InterPro protein families and domains database: 20 years on. Nucleic Acids Res 49, D344-D354, doi:10.1093/nar/gkaa977 (2021).
107 O'Donovan, C. et al. High-quality protein knowledge resource: SWISS-PROT and TrEMBL. Brief Bioinform 3, 275-284, doi:10.1093/bib/3.3.275 (2002).
108 Lowe, T. M. & Eddy, S. R. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res 25, 955-964, doi:10.1093/nar/25.5.955 (1997).
109 Griffiths-Jones, S. et al. Rfam: annotating non-coding RNAs in complete genomes. Nucleic Acids Res 33, D121-124, doi:10.1093/nar/gki081 (2005).
110 Nawrocki, E. P. & Eddy, S. R. Infernal 1.1: 100-fold faster RNA homology searches. Bioinformatics 29, 2933-2935, doi:10.1093/bioinformatics/btt509 (2013).
111 Buchfink, B., Xie, C. & Huson, D. H. Fast and sensitive protein alignment using DIAMOND. Nat Methods 12, 59-60, doi:10.1038/nmeth.3176 (2015).
112 Emms, D. M. & Kelly, S. OrthoFinder: phylogenetic orthology inference for comparative genomics. Genome Biol 20, 238, doi:10.1186/s13059-019-1832-y (2019).
113 Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res 32, 1792-1797, doi:10.1093/nar/gkh340 (2004).
114 Hedges, S. B., Dudley, J. & Kumar, S. TimeTree: a public knowledge-base of divergence times among organisms. Bioinformatics 22, 2971-2972, doi:10.1093/bioinformatics/btl505 (2006).
115 Sanderson, M. J. r8s: inferring absolute rates of molecular evolution and divergence times in the absence of a molecular clock. Bioinformatics 19, 301-302, doi:10.1093/bioinformatics/19.2.301 (2003).
116 Yang, Z. PAML: a program package for phylogenetic analysis by maximum likelihood. Comput Appl Biosci 13, 555-556, doi:10.1093/bioinformatics/13.5.555 (1997).
117 De Bie, T., Cristianini, N., Demuth, J. P. & Hahn, M. W. CAFE: a computational tool for the study of gene family evolution. Bioinformatics 22, 1269-1271, doi:10.1093/bioinformatics/btl097 (2006).
118 Chen, S., Zhou, Y., Chen, Y. & Gu, J. fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics 34, i884-i890, doi:10.1093/bioinformatics/bty560 (2018).
119 Kim, D., Paggi, J. M., Park, C., Bennett, C. & Salzberg, S. L. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. Nat Biotechnol 37, 907-915, doi:10.1038/s41587-019-0201-4 (2019).

120 Liao, Y., Smyth, G. K. & Shi, W. The Subread aligner: fast, accurate and scalable read mapping by seed-and-vote. Nucleic Acids Res 41, e108, doi:10.1093/nar/gkt214 (2013).
121 Langfelder, P. & Horvath, S. WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics 9, 559, doi:10.1186/1471-2105-9-559 (2008).
122 Wu, T. Z. et al. clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. Innovation-Amsterdam 2, doi:ARTN 10014110.1016/j.xinn.2021.100141 (2021).
123 Mistry, J., Finn, R. D., Eddy, S. R., Bateman, A. & Punta, M. Challenges in homology search: HMMER3 and convergent evolution of coiled-coil regions. Nucleic Acids Res 41, e121, doi:10.1093/nar/gkt263 (2013).
124 Tian, F., Yang, D. C., Meng, Y. Q., Jin, J. & Gao, G. PlantRegMap: charting functional regulatory maps in plants. Nucleic Acids Res 48, D1104-D1113, doi:10.1093/nar/gkz1020 (2020).
125 Lescot, M. et al. PlantCARE, a database of plant cis-acting regulatory elements and a portal to tools for in silico analysis of promoter sequences. Nucleic Acids Res 30, 325-327, doi:10.1093/nar/30.1.325 (2002).
126 Chen, C. et al. TBtools-II: A "one for all, all for one" bioinformatics platform for biological big-data mining. Mol Plant 16, 1733-1742, doi:10.1016/j.molp.2023.09.010 (2023).

Supplementary Table

Table S1 *P. australis* genome Hi-C assisted assembly statistics.

	Total (bp)	Contig Number	Contigs N50 (bp)	Scaffold Number	Scaffold N50 (bp)
Primary assembly	874,619,212	312	33,936,801	-	-
Hi-C assisted pre-assembly sketch	852,116,256	79	34,052,747	-	-
Chromosomes after Hi-C-assisted assembly	847,498,707	27	34,052,747	25	34,052,747

Table S2 Telomere Repeat Unit Finding Statistics.

canonical repeat unit	count	canonical repeat unit	count	canonical repeat unit	count
AAACCCT	40863	AACCC	9127	AAACC	1582
AAGAGAAGAG	25960	AAGAG	8398	ACCCT	1116
AAAGAAAG	16600	AGGGG	4796	AAACCTAAACCT	1018
AAAACCCT	16307	ACATCCTG	4752	AAACCT	997
AAAAG	16122	CCCCGGG	4606	AAACCCAACCCT	942
AACTC	13178	AGATAGAT	3734	AGAGCC	844
AACCT	13151	AAGAC	3690	CCGCG	760
AACCCT	11364	AAACCCCT	1929	AGCGG	726
AAAAACCCT	10192	AACCTG	1608		

Table S3 Genomic statistics of *P. australis* at chromosome level.

Chr	Supersc affolds	Length (bp)	Num of contigs	Num of genes	GC %	QV	Telomere status	Num of repetitions on the left	Num of repetitions on the right
1A	8	20,193,544	1	953	44.45%	48.08	both	1438	1975
1B	7	25,746,373	1	1076	44.02%	48.79	both	709	1291
2A	16	23,702,067	1	1008	43.98%	47.41	both	1889	1106
2B	15	24,664,423	1	1110	44.08%	48.26	right	0	781
3A	2	24,084,666	1	956	43.91%	48.99	both	1103	3221
3B	1	33,936,801	1	1159	44.15%	48.2	both	2020	1697
4A	24	25,221,131	1	1132	45.14%	49.34	left	2008	0
4B	23	27,472,612	1	1116	45.12%	48.15	left	1194	0
5A	18	28,937,882	1	1340	44.12%	47.17	left	1571	0
5B	17	31,241,098	1	1431	44.05%	48.33	both	2286	503
6A	22	30,183,679	1	1459	43.99%	47.47	both	543	2419
6B	21	30,499,992	2	1512	44.41%	48.36	both	451	2801
7A	14	31,468,791	1	1815	44.07%	47.98	both	1197	1492
7B	13	37,131,324	1	1849	44.39%	49.03	both	1464	1433
8A	10	33,085,515	1	1649	44.11%	47.91	both	801	2315
8B	9	37,281,543	1	1702	44.11%	49.74	both	2879	1689
9A	4	34,052,747	1	1746	44.20%	48.79	both	1669	1327
9B	3	35,764,372	1	1590	44.15%	49.59	both	2384	2130
10A	12	43,967,752	1	2434	44.09%	48.58	both	1119	2073
10B	11	44,949,043	1	2304	44.02%	49.77	both	905	469
11A	5	49,387,842	1	2881	44.19%	49.22	left	1662	0
11B	6	48,350,371	1	2642	43.92%	48.59	both	1604	1041
12A	19	54,964,116	1	2917	44.05%	49.1	both	1703	2128
12B	20	50,013,265	1	2934	44.03%	48.37	both	2228	1509
B	25	21,198,758	2	277	44.73%	57.28	left	3067	0

Note: Length of Contigs: the total length of the contigs that make up the chromosome; Length of Superscaffold: the total length of the contigs that are linked together with 500 'N' inserted between each two contigs. QV (Quality Value): the quality value of the bases, for example, Q30 means 99.9% consistency accuracy, Q40 is 99.99%.

Table S4 Survey and full-length transcriptome data mapping statistics.

Data type	Mapping rate (%)	Average sequencing depth	Coverage (%)	Coverage (>= 5X, %)	Coverage (>= 10X, %)	Coverage (>= 20X, %)
Survey data	98.62	54.4	99.91	99.78	99.5	97.67
Iso_Seq data	99.99	36.65	99.99	99.91	99.62	92.27

Table S5 Genomic SNP and Indel detection.

	SNP			Indel		
	Number	Percentage of indel (%)	Percentage of genome (%)	Number	Percentage of indel (%)	Percentage of genome (%)
All	8,240,383	100	0.967	477,839	100	0.0561
Heterozygosis	8,234,289	99.93	0.9663	475,777	99.57	0.0558
Homology	6,094	0.07	0.0007	2,062	0.43	0.0002

Table S6 Genomic Synteny depth statistics.

	Number of syntenic blocks	Number of collinear gene pairs
P. australis vs *O. sativa*	246	19384
P. australis vs *C. songorica*	614	25732
P. australis A vs *O. sativa*	87	14311
P. australis B vs *O. sativa*	93	13972
P. australis A vs *C. songorica*	216	19090
P. australis B vs *C. songorica*	241	18727

Table S7 Cis-acting element statistics for the promoter of PaSUC1.1 (rna-Pau34219.1).

GeneID	Name of cis acting element	Cis acting element motif	Functional classification of cis acting elements
rna-Pau34219.1	TC-rich repeats	GTTTTCTTAC	cis-acting element involved in defense and stress responsiveness
rna-Pau34219.1	LTR	CCGAAA	cis-acting element involved in low-temperature responsiveness
rna-Pau34219.1	LTR	CCGAAA	cis-acting element involved in low-temperature responsiveness
rna-Pau34219.1	LTR	CCGAAA	cis-acting element involved in low-temperature responsiveness
rna-Pau34219.1	ABRE	ACGTG	cis-acting element involved in the abscisic acid responsiveness
rna-Pau34219.1	ABRE	ACGTG	cis-acting element involved in the abscisic acid responsiveness
rna-Pau34219.1	ABRE	ACGTG	cis-acting element involved in the abscisic acid responsiveness
rna-Pau34219.1	ARE	AAACCA	cis-acting regulatory element essential for the anaerobic induction
rna-Pau34219.1	G-box	CACGAC	cis-acting regulatory element involved in light responsiveness
rna-Pau34219.1	G-box	CACGTC	cis-acting regulatory element involved in light responsiveness
rna-Pau34219.1	G-box	TACGTG	cis-acting regulatory element involved in light responsiveness
rna-Pau34219.1	G-box	CACGTC	cis-acting regulatory element involved in light responsiveness
rna-Pau34219.1	CGTCA-motif	CGTCA	cis-acting regulatory element involved in the MeJA-responsiveness
rna-Pau34219.1	CGTCA-motif	CGTCA	cis-acting regulatory element involved in the MeJA-responsiveness
rna-Pau34219.1	CGTCA-motif	CGTCA	cis-acting regulatory element involved in the MeJA-responsiveness
rna-Pau34219.1	TGACG-motif	TGACG	cis-acting regulatory element involved in the MeJA-responsiveness
rna-Pau34219.1	TGACG-motif	TGACG	cis-acting regulatory element involved in the MeJA-responsiveness
rna-Pau34219.1	TGACG-motif	TGACG	cis-acting regulatory element involved in the MeJA-responsiveness
rna-Pau34219.1	Sp1	GGGCGG	light responsive element
rna-Pau34219.1	MBS	CAACTG	MYB binding site involved in drought-inducibility

Supplementary Figure

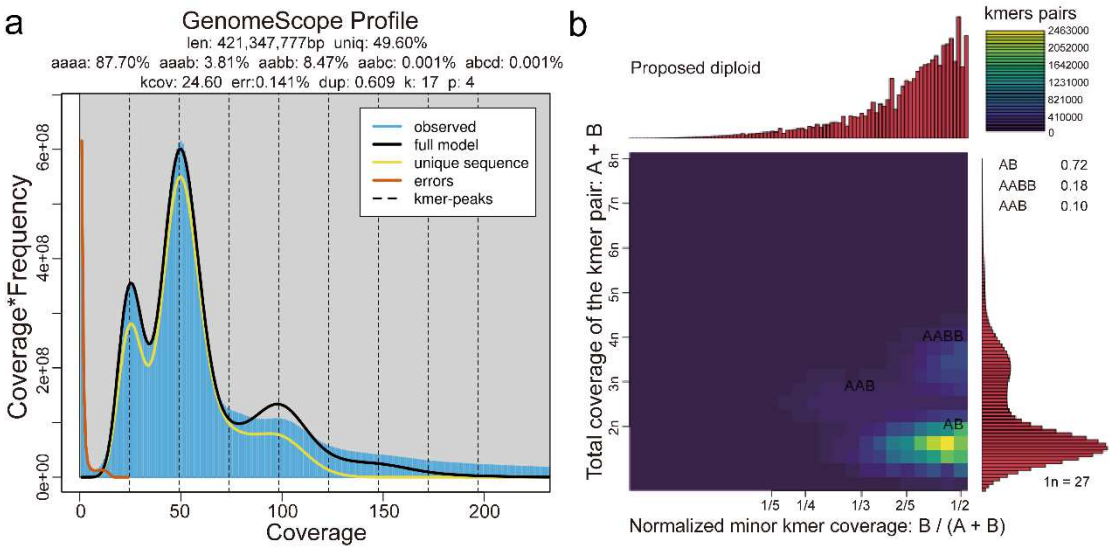


Figure S1. Survey analysis of the *P. australis* genome. a The 17-mer depth and number frequency distribution. Analyzed and visualized using GenomeScope 2.0. **b** Smudgeplot of the *P. australis* genome with relative coverage (CovB / (CovA + CovB)) in the horizontal, total coverage (CovA + CovB) in the vertical, and colors indicating the frequency of k-mer pairs. Each possible haplotype structure is presented as a "Smudge" on the graph, and the heat of the "Smudge" indicates the frequency of the haplotype structure in the genome.

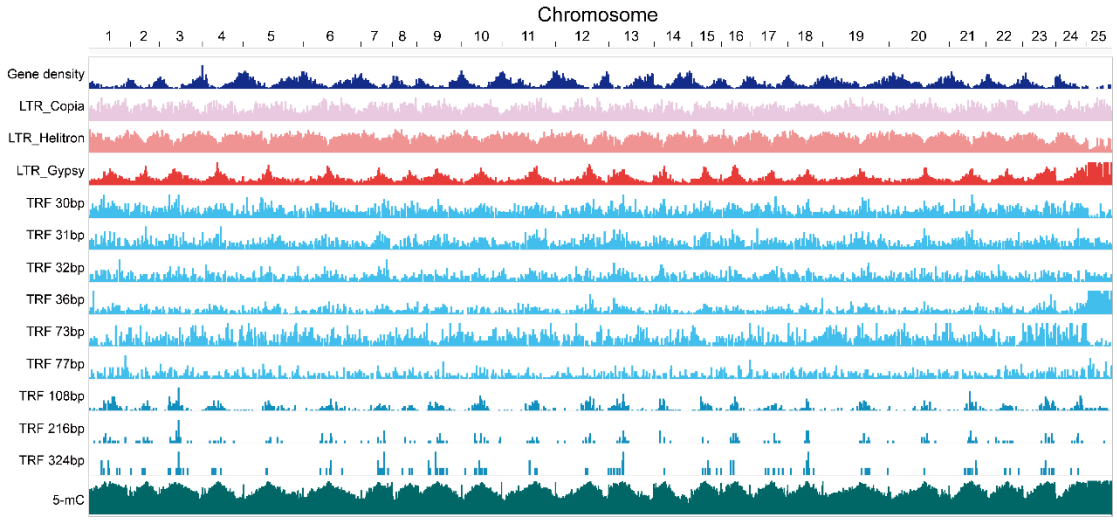


Figure S2. The distribution of gene density, different types of TE and TRF, and 5-methylcytosine in the chromosomes. The distribution of these data was visualized using IGV.

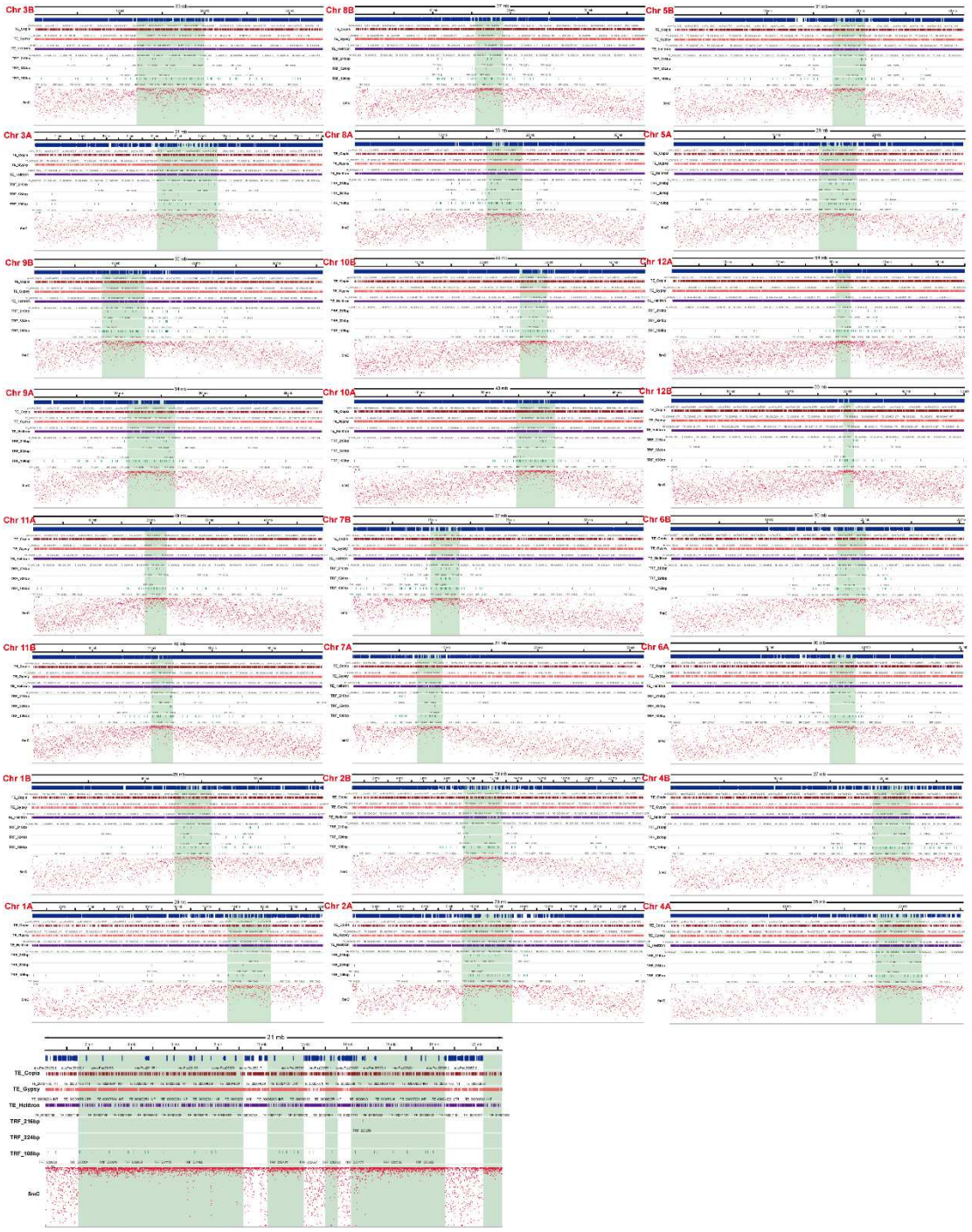


Figure S3. 5-Methylcytosine distribution density distribution identifies candidate locations for *P. australis* centromere. The green background indicates candidate centromere locations.

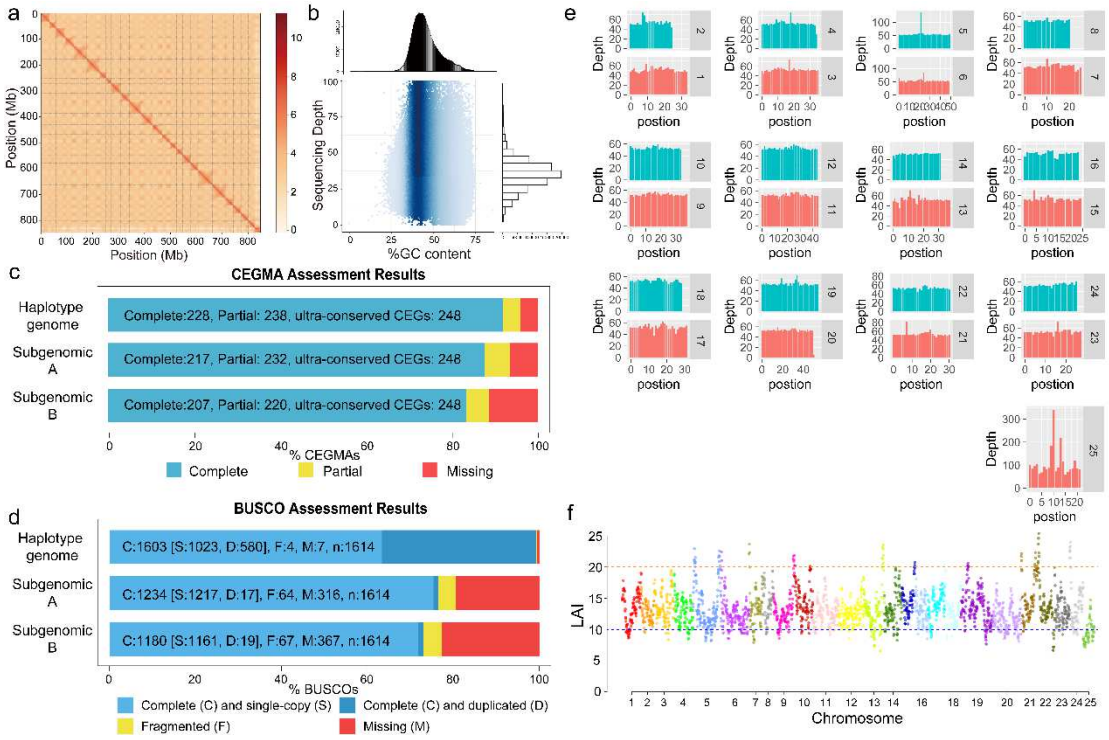


Figure S4. Genome assembly quality assessment. (A) Genome-wide Hi-C interactions heatmap of the *P. australis* genome. Colors from light to dark in the plot indicate increasing strength of interactions, with darker colors showing stronger interactions. The 25 squares on the diagonal of the plot are the 25 chromosomes of *P. australis*, and there is no obvious clustering error (3C). (B) Density plot of GC content distribution and sequencing depth distribution in the genome of *P. australis*, with darker dots representing a higher number of dots here. (C) CEGMA assessment of gene region integrity in the genome. (D) BUSCO assessment. The BUSCO single-copy homozygous gene set is embryophyta_odb10. (E) Assessment of assembly integrity and uniformity of sequencing coverage by mapping second-generation data into the assembled genome, with the vertical coordinate indicating second-generation data coverage. (F) Genetic integrity on the 25 chromosomes assembled was assessed by LAI, with the horizontal coordinate indicating the chromosome number of the assembly and the vertical coordinate indicating the LAI of a particular chromosome.

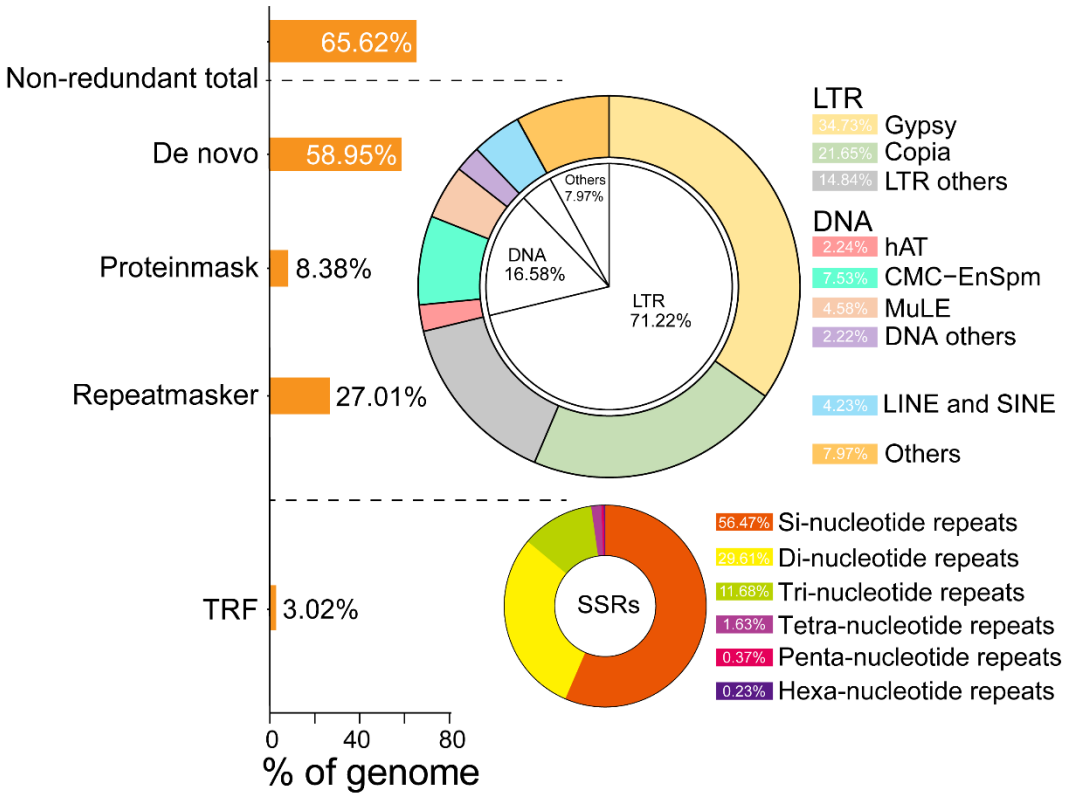


Figure S5. Repetitive sequence annotation in the P. australis genome. The bar graph on the left indicates the results of repeated sequence prediction in the P. australis genome by different strategies.

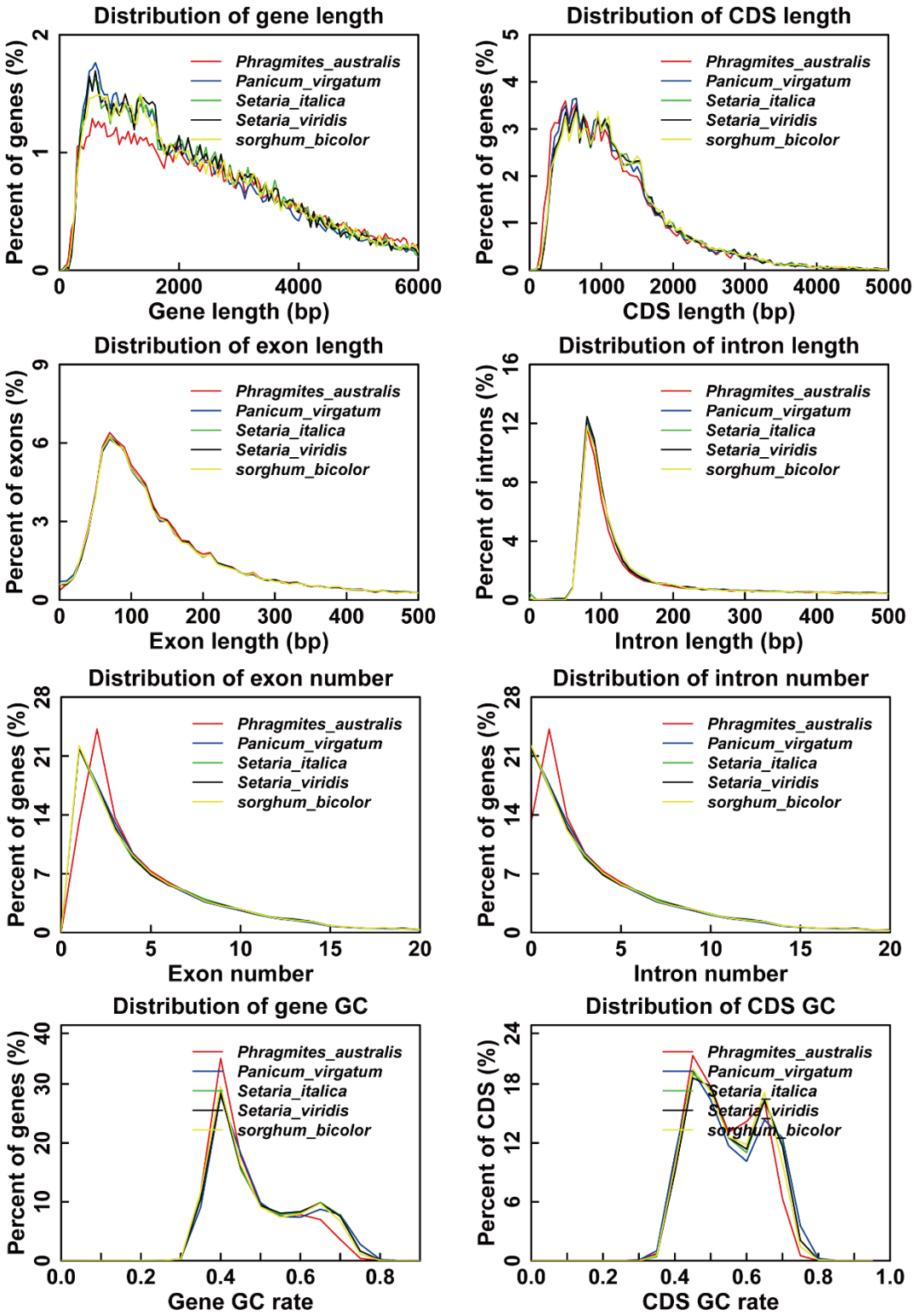
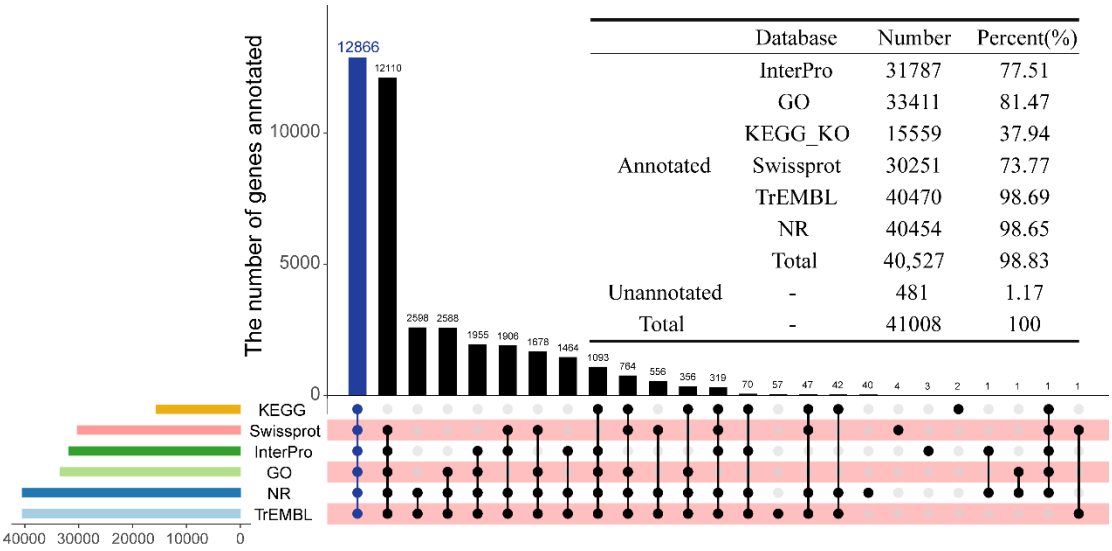


Figure S6. Comparison of five Gramineae gene elements. This includes gene length distribution, CDS length distribution, length and number distribution of exons and introns, and GC rate distribution of genes and CDS.



	Database	Number	Percent(%)
Annotated	InterPro	31787	77.51
	GO	33411	81.47
	KEGG_KO	15559	37.94
	Swissprot	30251	73.77
	TrEMBL	40470	98.69
	NR	40454	98.65
	Total	40,527	98.83
Unannotated	-	481	1.17
Total	-	41008	100

Figure S7. Upset plots of the functional annotation of the *P. australis* genome. The blue bars in the figure indicate the number of genes for which the best matches exist in six databases simultaneously. Visualization was performed using the UpSetR (https://github.com/hms-dbmi/UpSetR) packages in R.

Figure S8. Distribution of non-coding RNAs on chromosomes in the P. australis genome.

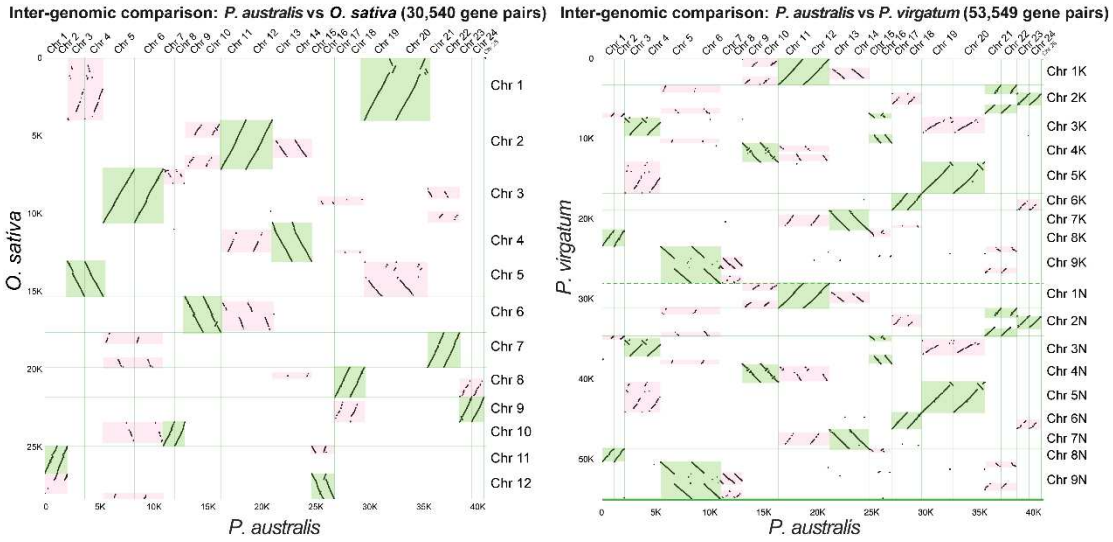


Figure S9. Chromosomal synteny of *P. australis* with *Oryza sativa* and *Panicum virgatum*, respectively. All syntenic genes with C-score ≥ 0.7 are shown using the jcvi ortholog algorithm.

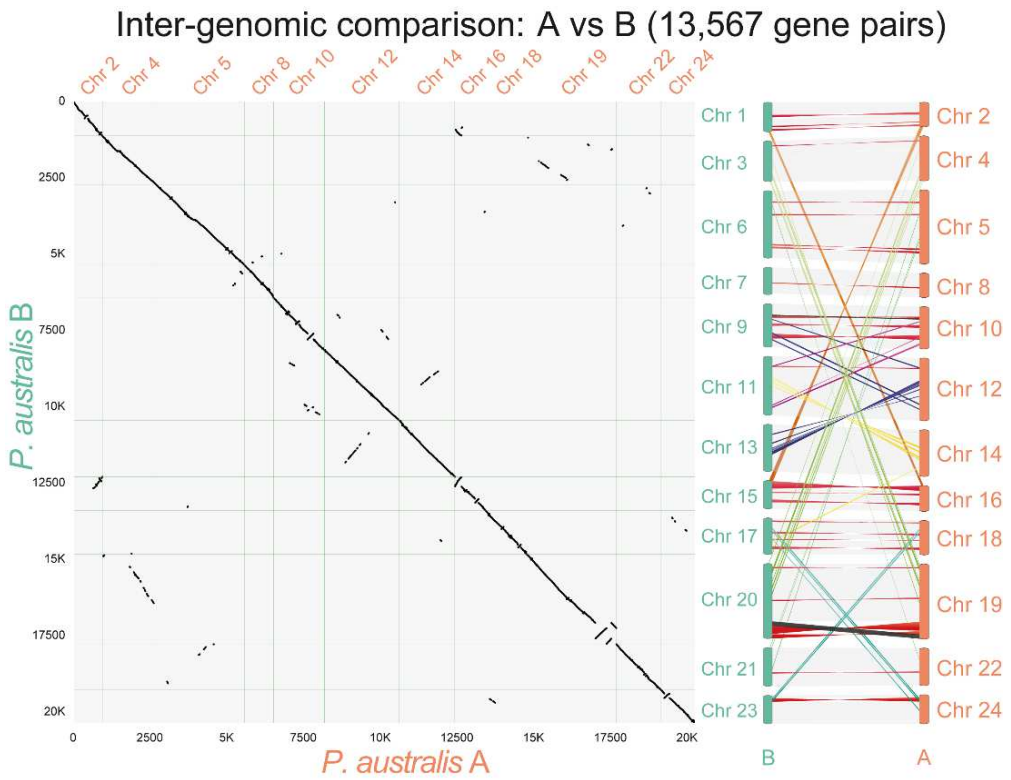


Figure S10. Dot plot of colinearity analysis of the two subgenomes of *P. australis*. The right side indicates syntenic blocks between subgenomes A and B, and red crosshairs indicate chromosome inversions.

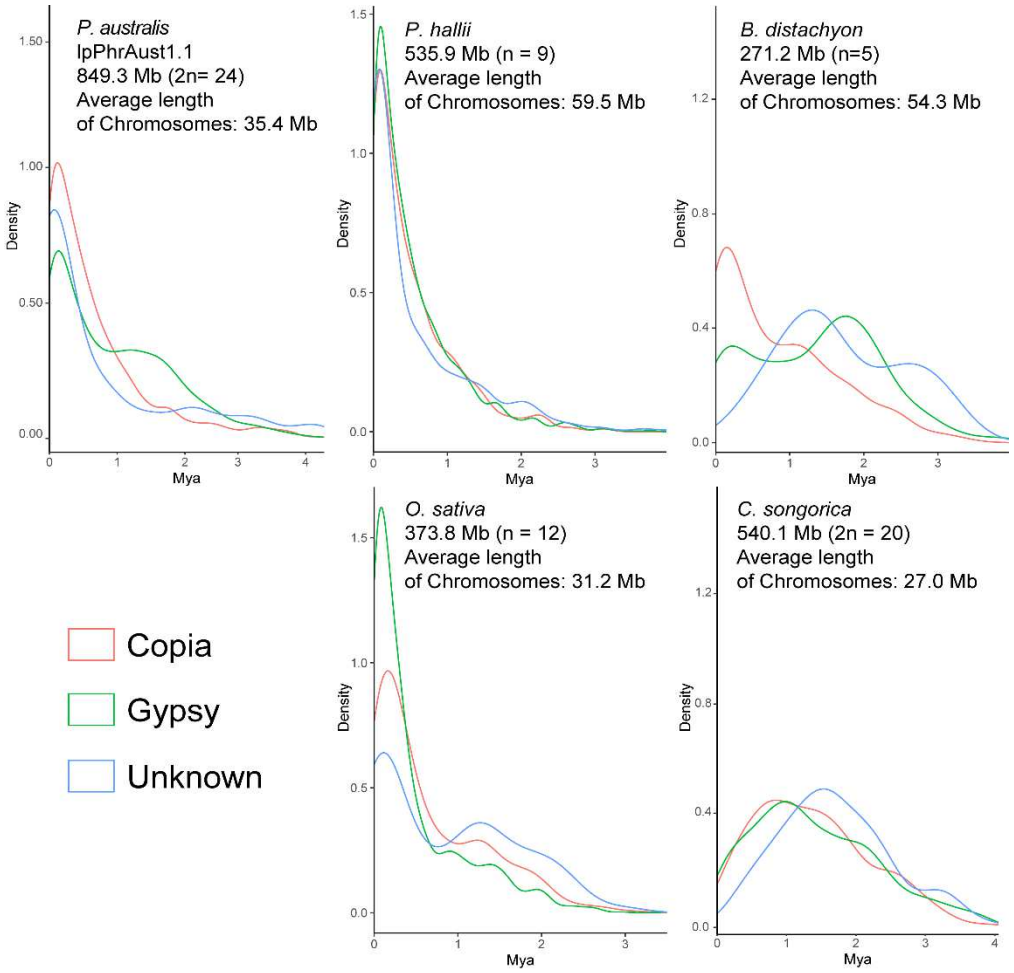
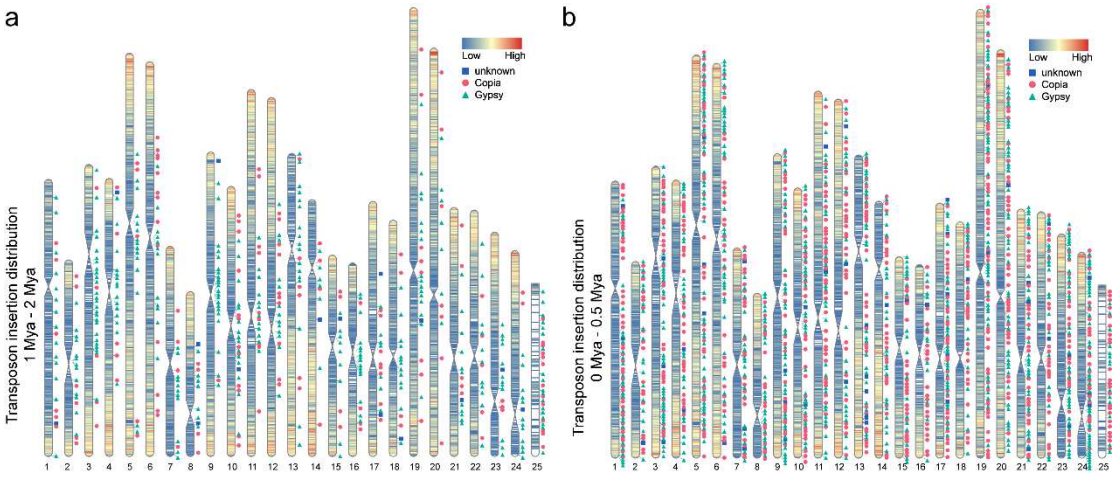


Figure S11. Distribution of insertion densities of different types of LTR-RTs in five Graminae plants. Including the *P. australis* (IpPhrAust1.1), *Panicum hallii, Brachypodium distachyon, Oryza sativa,* and *Cleistogenes songorica.*



1222

1223 **Figure S12. Distribution of ancient (A) and new (B) inserted LTR-RTs on chromosomes.**

1224

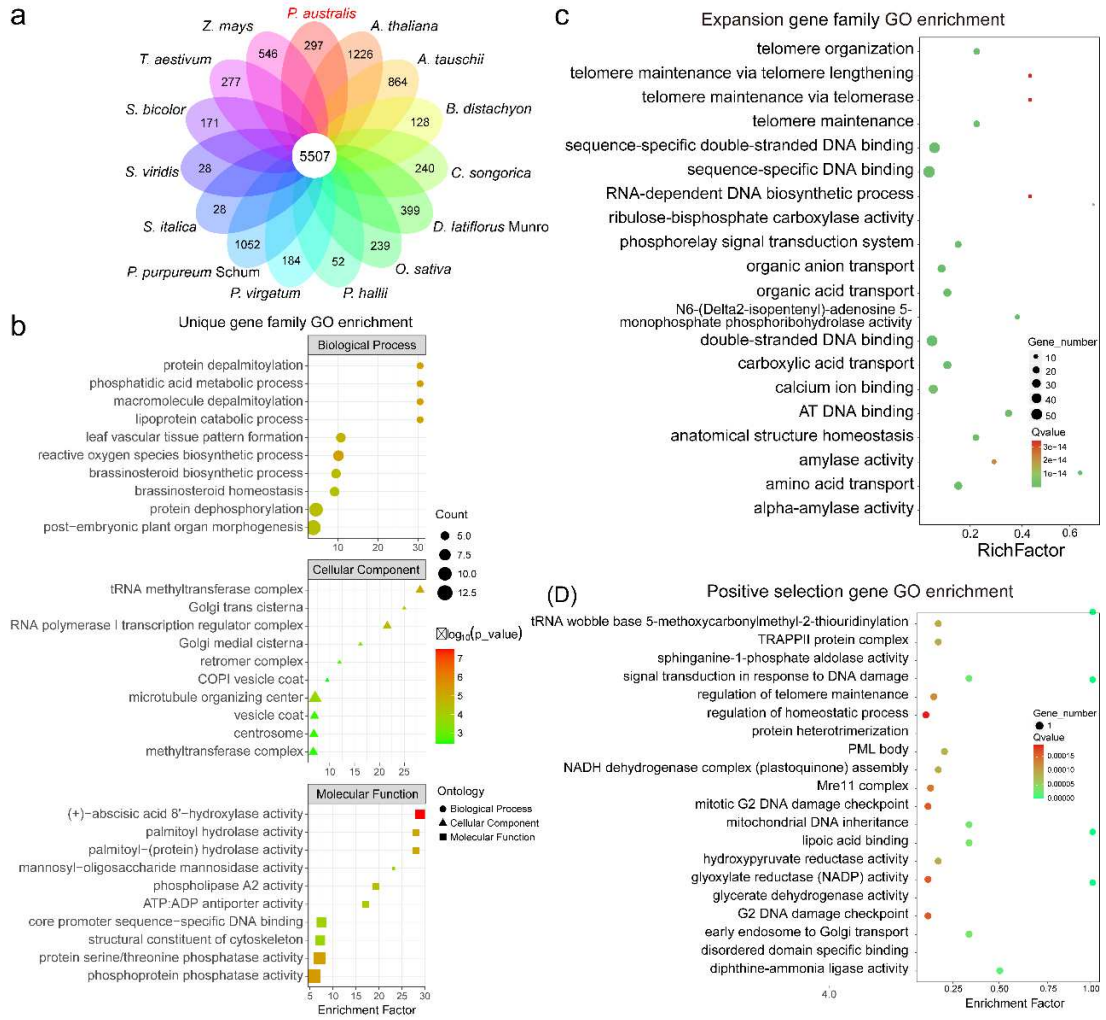


Figure S13. *P. australis* gene family enrichment analysis. (A) Venn diagram of 15 species sharing or unique gene family clustering. The numbers in the diagram indicate the number of species gene families. (B) GO enrichment analysis of individual gene families in *P. australis.* (C) GO enrichment analysis of gene family expansions occurring in *P. australis*. (D) GO enrichment of the positive selection genes in *P. australis*.

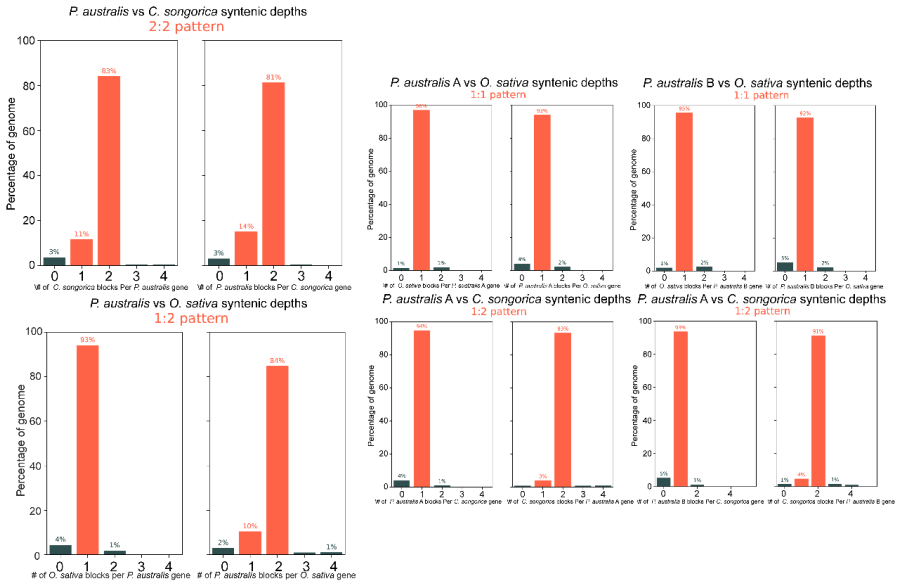


Figure S14. The syntenic depth of *P. australis* genomes/subgenomes with *Cleistogenes songorica* and O. sativa, respectively.

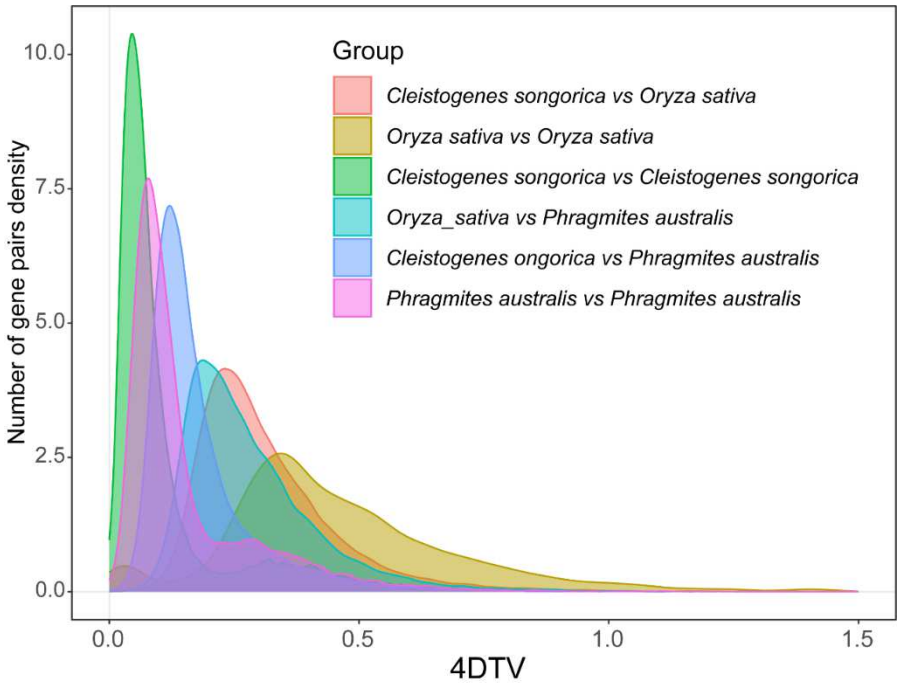


Figure S15. Distribution of Four-fold synonymous third-codon transversion rate (4Dtv) distances for gene pairs in the syntenic blocks of *P. australis*, *O. sativa*, and *C. songorica* genomes.

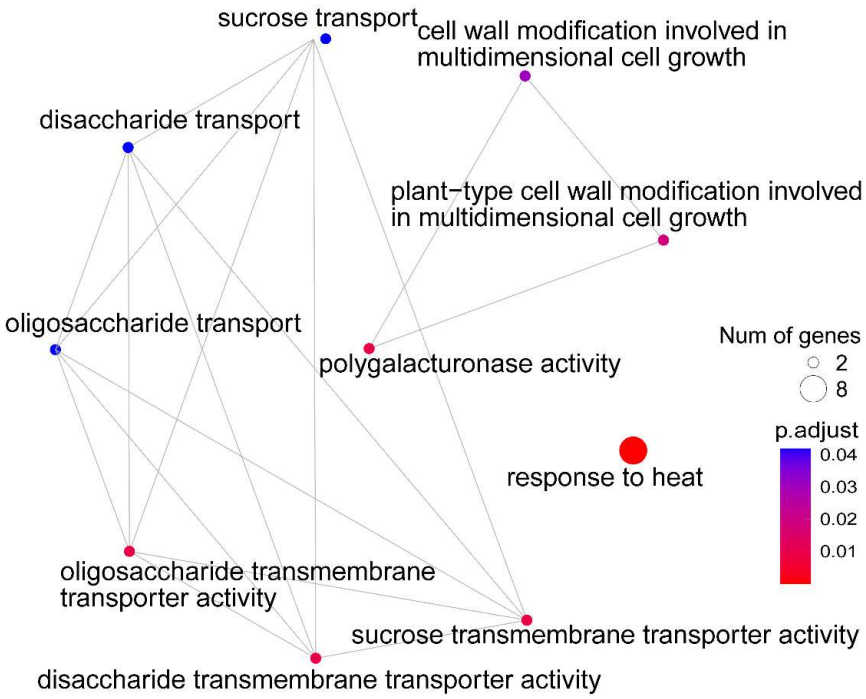


Figure S16. GO enrichment of Hub genes in the brown module.