JIPB Journal of Integrative Plant Biology

www.jipb.net 植物学报(英)
VOLUME 64, ISSUE 3
MARCH 2022

WILEY

JIPB Journal of Integrative Plant Biology

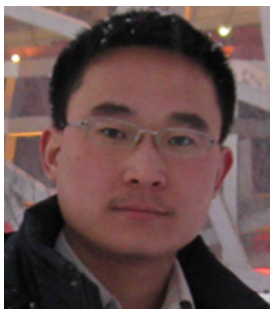
New Resources
https://doi.org/10.1111/jipb.13217

Allele-aware chromosome-scale assembly of the allopolyploid genome of hexaploid Ma bamboo (*Dendrocalamus latiflorus* Munro)[∞]

Yushan Zheng[1]*, Deming Yang[1], Jundong Rong[1], Liguang Chen[1], Qiang Zhu[2], Tianyou He[2], Lingyan Chen[2], Jing Ye[1], Lili Fan[1], Yubang Gao[3], Hangxiao Zhang[3] and Lianfeng Gu[3]*

1. College of Forestry, Fujian Agriculture and Forestry University, Fuzhou 350002, China
2. College of Landscape Architecture, Fujian Agriculture and Forestry University, Fuzhou 350002, China
3. Basic Forestry and Proteomics Research Center, College of Forestry, Fujian Provincial Key Laboratory of Haixia Applied Plant Systems Biology, Fujian Agriculture and Forestry University, Fuzhou 350002, China
*Correspondences: Yushan Zheng (zys1960@163.com); Lianfeng Gu (lfgu@fafu.edu.cn, Dr. Gu is fully responsible for the distribution of the materials associated with this article)

Yushan Zheng



Lianfeng Gu

ABSTRACT

Dendrocalamus latiflorus Munro is a woody clumping bamboo with rapid shoot growth. Both genetic transformation and clustered regularly interspaced short palindromic repeats (CRISPR)/CRISPR-associated protein 9 (Cas9) gene editing techniques are available for *D. latiflorus,* enabling reverse genetic approaches. Thus, *D. latiflorus* has the potential to be a model bamboo species. However, the genome sequence of *D. latiflorus* has remained unreported due to its polyploidy and large genome size. Here, we sequenced the *D. latiflorus* genome and assembled it into three allele-aware subgenomes (AABBCC), representing the largest genome of a major bamboo species. We assembled 70 allelic chromosomes (2, 737 Mb) for hexaploid *D. latiflorus* using both single-molecule sequencing from the Pacific Biosciences (PacBio) Sequel platform and chromosome conformation capture sequencing (Hi-C). Repetitive sequences comprised 52.65% of the *D. latiflorus* genome. We annotated 135 231 protein-coding genes in the genome based on transcriptomes from eight different tissues. Transcriptome sequencing using RNA-Seq and PacBio single-molecule real-time long-read isoform sequencing revealed highly differential alternative splicing (AS) between non-abortive and abortive shoots, suggesting that AS regulates the abortion rate of bamboo shoots. This high-quality hexaploid genome and comprehensive strand-specific transcriptome datasets for this Poaceae family member will pave the way for bamboo research using *D. latiflorus* as a model species.

Keywords: alternative splicing, bamboo shoots, chromosome conformation capture sequencing, *D. latiflorus*, genome assembly, single-molecular sequencing

Zheng, Y., Yang, D., Rong, J., Chen, L., Zhu, Q., He, T., Chen, L., Ye, J., Fan, L., Gao, Y., Zhang, H., and Gu, L. (2022). Allele-aware chromosome-scale assembly of the allopolyploid genome of hexaploid Ma bamboo (*Dendrocalamus latiflorus* Munro). J. Integr. Plant Biol. **64:** 649–670.

INTRODUCTION

Bamboo is one of the fastest-growing non-timber plant resources, is distributed widely in the subtropics and tropics, and accounts for approximately 1.0% of the total forest area worldwide (Zhou et al., 2011). The market value of bamboo is more than $2.5 billion per year (Lobovikov et al., 2007). In addition to its economic value, bamboo forests have high potential for carbon fixation, providing a carbon sink to combat rising atmospheric carbon dioxide concentrations (Xu et al., 2011;

Song et al., 2013; Li et al., 2015b). Bamboo shows unique characteristics due to its long flowering intervals (Liu et al., 2012; Ge et al., 2017) and rapid growth (Wei et al., 2017; Li et al., 2018; Guo et al., 2019a). An analysis of dynamic changes in gene expression revealed that hormone-associated genes are involved in regulating bamboo shoot growth (Li et al., 2018). A recent study on the rapid elongation of a single bamboo internode revealed that the cell division and cell elongation zones occupied two successive ~1 cm sections located at the bottom of the internode (Wei et al., 2019). By studying a thick-walled variant with a narrow pith cavity, the same group revealed the important role of pith tissue in promoting the primary growth of underground shoots (Wei et al., 2017).

Although genome sequencing (Peng et al., 2013; Zhao et al., 2018) in *Phyllostachys edulis* has been reported, *Agrobacterium*-mediated transformation has not been established for this monopodial (running) bamboo, making it difficult to use this species as a model for bamboo research. Bamboo's unique features, including its rapid growth rate and prolonged vegetative phase, are also present in the sympodial (clumping) bamboo *Dendrocalamus latiflorus* Munro (Liu et al., 2012). Methods for transgenic expression based on *Agrobacterium*-mediated transformation (Qiao et al., 2014; Ye et al., 2017) and clustered regularly interspaced short palindromic repeats (CRISPR)/CRISPR-associated protein 9 (Cas9) mediated genome editing (Ye et al., 2020) are available for *D. latiflorus*. Thus, the availability of the genome sequence of *D. latiflorus* would shed light on the genomic organization of this plant, paving the way for molecular and functional studies of bamboo.

During the early stages, the Illumina platform was used for genome assembly in trees such as desert poplar (*Populus euphratica*) (Ma et al., 2013) and Norway spruce (*Picea abies*) (Nystedt et al., 2013). More recently, single-molecule, real-time (SMRT) sequencing has been shown to produce high-quality genomes because of its long reads and has been used to generate genome assemblies for plants such as the desiccation-tolerant grass *Oropetium thomaeum* (VanBuren et al., 2015), apple (*Malus domestica* Borkh) (Daccord et al., 2017), and *Casuarina equisetifolia* (Ye et al., 2019).

To date, chromosome-level assemblies of autopolyploid genomes have been reported for cultivated alfalfa (*Medicago sativa* L.) (Chen et al., 2020) and sugarcane (*Saccharum spontaneum*) (Zhang et al., 2018). Due to the reduced fertility of autopolyploid plants, more plant species are allopolyploid (Ming and Man Wai, 2015). Allopolyploid genomes have been assembled for the allohexaploid wheat (*Triticum aestivum*; $2n = 6x = 42$, AABBDD) (International Wheat Genome Sequencing, 2014) and the allotetraploids mustard (*Brassica juncea*; $2n = 4x = 19$, $A_nA_nC_nC_n$) (Yang et al., 2016), peanut (*Arachis hypogaea*) (Chen et al., 2019), cotton (*Gossypium hirsutum*; $2n = 4x = 52$, AtAtDtDt) (Li et al., 2015a), and oilseed rape (*Brassica napus* L.; $A_nA_nC_nC_n$) (Chalhoub et al., 2014). Allopolyploids (from closely related species) have advantages over autopolyploids (from the same species) due to intergenomic heterosis or interspecific crosses (hybrid vigor) (Ming and Wai, 2015). However, an allele-aware chromosome-level genome for an allopolyploid forestry tree species has not been reported.

Here, we report the *de novo* whole-chromosome assemblies of the hexaploid bamboo species *D. latiflorus* using the PacBio Sequel platform (Kyriakidou et al., 2018). We took advantage of a combination of PacBio SMRT sequencing technology and high-throughput chromosome conformation capture (Hi-C) to distinguish the three closely related subgenomes and generate an allele-aware chromosome-level genome (2,737 Mb) for *D. latiflorus*. Moreover, by performing PacBio isoform sequencing (Iso-Seq) RNA-Seq of eight different tissues, we obtained comprehensive annotations of *D. latiflorus* genes and uncovered their post-transcriptional regulation. The availability of valuable chromosome-level genome and transcriptome datasets generated in this study, together with an overexpression system (Ye et al., 2017) and a genome editing protocol (Ye et al., 2020), make *D. latiflorus* an excellent model for investigating the molecular mechanisms underlying the extraordinarily rapid growth and other distinctive features of bamboo.

RESULTS

Genome survey analysis of *D. latiflorus* Munro

Sympodial (clumping) *D. latiflorus* differs morphologically from the monopodial (running) bamboo *P. edulis*. *D. latiflorus* has wider leaves than *P. edulis* (Figure 1A), which is consistent with the heterosis advantage of allopolyploids. Karyotypic studies (Figure S1) revealed that *D. latiflorus* has higher chromosomal complexity than the first reported bamboo genome (*P. edulis*), which comprises 48 chromosomes ($2n = 4x = 48$) (Peng et al., 2013; Zhao et al., 2018). Using flow cytometry (FCM) and diploid cultivated rice (*Oryza sativa* ssp. *japonica*; 430 Mbp) (Eckardt, 2000) as an internal reference, we estimated the genome size of hexaploid *D. latiflorus* to be approximately 1,547.431 ± 96 Mbp (1C) (Figure S2), which is significantly larger than that of hexaploid *Bonia amplexicaulis* (Guo et al., 2019b). We extracted genomic DNA from *D. latiflorus* leaf tissue and performed a genome survey and whole-genome sequencing using both Illumina and PacBio SMRT technology. We performed a genome survey using 416,576,411 paired reads from the Illumina platform (150 bp long) to estimate the genome size of *D. latiflorus* using Genomescope (Vurture et al., 2017) with 17 *k*-mers (Figure S3). The estimated genome size was 1,324 Mb (1,324,331,250 bp). The heterozygosity was estimated to be 2.26% by calculating the ratio of heterozygous sites in the sequence. The calculated GC content of the genome was 45.02% (Figure S4).

De novo assembly using PacBio single molecule sequencing

We utilized PacBio SMRT technology DNA sequencing with a 20 kb insert size to generate a high-quality *de novo* assembly of the *D. latiflorus* genome. Deep long read sequencing of *D. latiflorus* produced 318.99 Gb of Pacific Bioscience SMRT long reads, representing ~110-fold coverage of the genome (Table S1). To facilitate genome

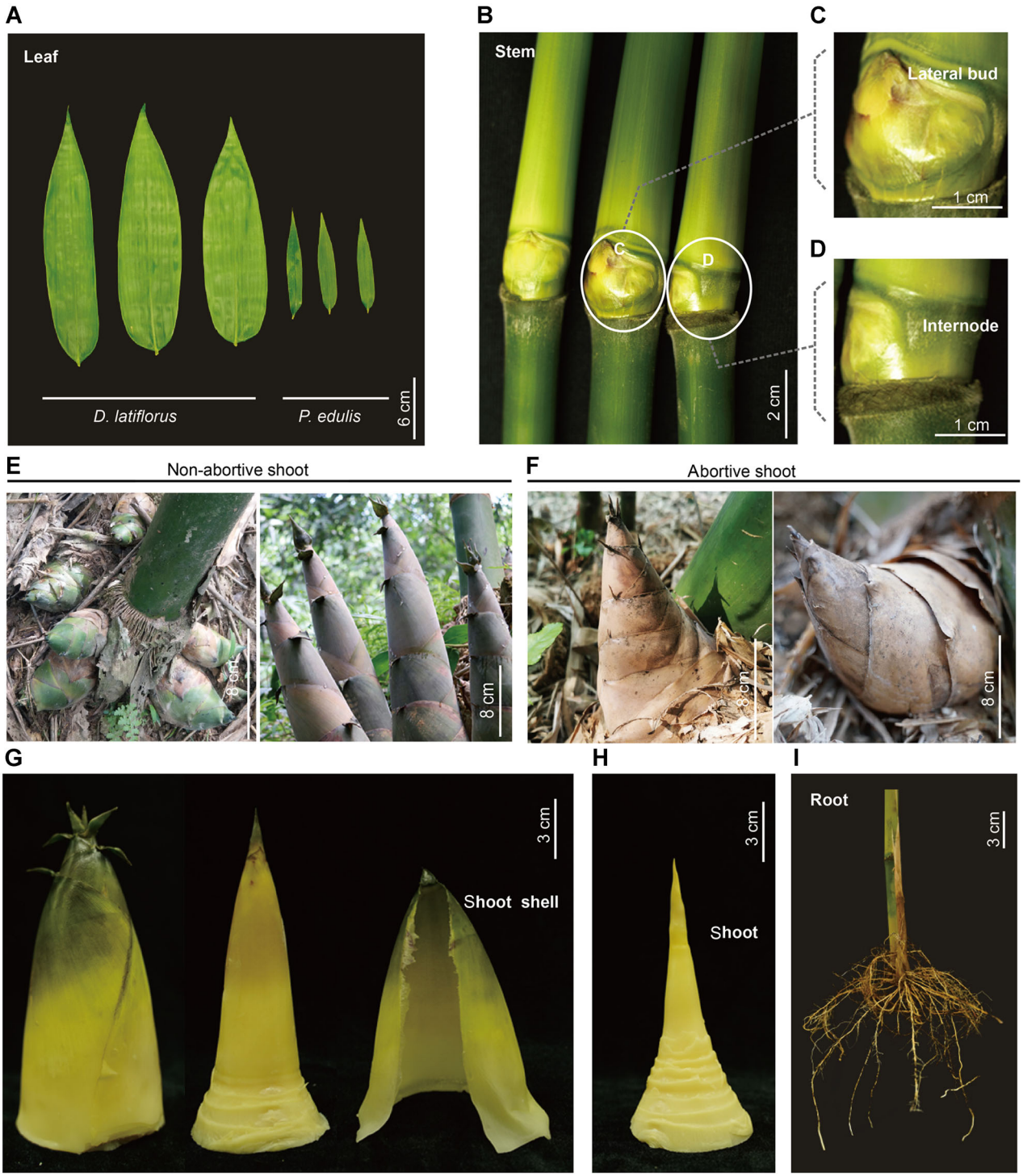


Figure 1. Materials used for genome and transcriptome sequencing

(**A**) Leaf material from *Dendrocalamus latiflorus* used for genome sequence. For comparison, leaves from moso bamboo (*Phyllostachys edulis*) are shown on the right. (**B**) *D. latiflorus* stem, (**C**) lateral bud, (**D**) internode, (**E**) non-abortive shoot, (**F**) abortive shoot, (**G**) shoot shell, (**H**) shoot, and (**I**) root tissue used for transcriptome sequencing.

assembly, we also used 10X Genomics, which produced approximately 343.80 Gb of Illumina short reads, representing a 119-fold coverage of the genome. We constructed the initial genome assembly by single molecule long-read sequencing data using the FALCON assembler (Chin et al., 2016). We aligned 10X Genomics linked reads to the consensus sequence of the PacBio assembly to obtain the superscaffold, since 10X Genomics reads provide long-range information from each fragment of DNA. Only the consensus sequences with linked-reads support were used for

subsequent assembly. Following combination assembly using SSPACE (Boetzer et al., 2011), the size of the final merged genome of *D. latiflorus* was upgraded to 2,621.24 Mb.

The Contig N50 and scaffold N50 were 2.57 and 4.5 Mb, respectively. The largest scaffold was 21.73 Mb, which was obtained using ultra-long PacBio reads to overcome the high level of repetitiveness and allopolyploidy. Using a 10 kb sliding window, we calculated the GC content of each window and the average depth by aligning the Illumina reads to the assembled genome. As shown in the GC scatter plot, the GC content of the *D. latiflorus* genome was 44.65%, with uniform GC composition, suggesting that no foreign contamination was present in the sample (Figure S5). The average depth of the scatter distribution is consistent with the sequencing depth.

High-throughput chromosome conformation capture assisted genome assembly

High-throughput chromosome conformation capture, a technique involving massive parallel DNA sequencing after the purification of ligation products, is used to determine chromatin interactions at the megabase scale by measuring the frequency of contact between pairs of loci (Lieberman-Aiden et al., 2009). In this study, we used PacBio SMRT sequencing and Hi-C to resolve the assembly of hexaploid *D. latiflorus*. In total, our Hi-C library generated 368.28 Gb of purified Hi-C fragments (~131× sequence coverage) using paired-end sequencing on the Illumina HiSeq platform. More than 11,611,419 processed reads were mapped to the scaffold sequence with split alignments using Burrows-Wheeler Aligner (BWA)'s bwa-sw mode to generate 3,760,845 valid mapped paired-reads (termed di-tags). Finally, we obtained 2,988,739 unique di-tags, which corresponded to one restriction fragment from a capture target location and its ligated interacting partner. We assigned unique di-tags to the assembled scaffold at the chromosome level. The effect rate (unique di-tags/total reads processed) was approximately 25.74% (Table S1).

Several haplotype-resolved genome assemblies have been reported in potato (*Solanum tuberosum* L.) (Zhou et al., 2020), Oolong tea (*Camellia sinensis*) (Zhang et al., 2021), and apple (*M. domestica* cv. Gala) (Sun et al., 2020), respectively. To generate an allele-aware chromosome-level genome, we used the ALLHiC algorithm (Zhang et al., 2019), which was designed to scaffold polyploid genomes and allelic haplotypes. Finally, we incorporated the scaffolds into 70 allele-aware chromosome-level subgenomes (AABBCC), which correspond to 35 chromosome pairs (Figure 2). Contig assembly or ALLHiC phasing may generate switch errors, which is discussed in a previous study (Zhang et al., 2021). Following a previous method for estimation of switch error (Zhang et al., 2021), we found that the phase switch errors were 25.28%, 23.67%, and 25.34% for subgenomes A, B, and C, respectively. However, the percentage of switch errors is comparable with published haplotype-resolved genome assemblies, which included 15.35% and 29.62% switch errors in diploid potato (Zhou et al., 2020) and apple (Sun et al., 2020), respectively. Different heterozygosity rates from different species may affect the genome assembly and haplotype phasing algorithm. Both the apple (Sun et al., 2020) and potato (Zhou et al., 2020) genomes show high heterozygosity rates. Moreover, complex genomes, including autopolyploid and allopolyploid genomes, might be substantially more difficult to analyze during haplotype phasing due to similar sequences from the homologous chromosome. In consideration of complexity for phasing of an allopolyploid genome, current accuracy between haplotypes is acceptable for the haplotype-resolved genome of *D. latiflorus*. Further improvement of haplotype phasing algorithms for polyploids is required to generate a near-complete haplotype-resolved genome (Zhou et al., 2020).

Synteny and variations between two haplotypes

High genome-wide synteny was observed based on a comparison of alignments between two haplotypes (Figure S6A). Clear diagonal distribution was observed in a synteny dot plot from the JCVI, which showed 15,336, 14,528, and 13,880 gene pairs in gene-level collinearity between haplotypes for subgenomes A, B, and C, respectively. Pairwise alignments between haplotypes revealed 104,345, 98,017, and 78,872 alignment blocks for subgenomes A, B, and C, respectively. Alignment blocks between haplotypes are in synteny (Figure S6B).

In total, we identified 9,792,585, 8,898,779, and 7,167,029 single nucleotide variations (SNVs) in subgenomes A, B, and C, respectively (Figure S7A). We also identified 614,183, 573,834, and 473,273 deletions (Figure S7B) and 615,070, 574,519, and 474,429 insertions in subgenomes A, B, and C, respectively (Figure S7C). The sequence variants between the two haplotypes were widely distributed across all chromosomes (Figure S8).

This assembly generated 2,737,182,539 bp of genome sequence. The largest haplotype chromosomes were 61 Mb (chr1.2) and 64 Mb (chr1.1). The *D. latiflorus* genome generated in this study represents the largest genome among the currently available sequences for the major bamboo lineages. High-density synteny was also observed among the subgenomes of *D. latiflorus* at the chromosome level (Figure 2), providing a valuable resource for investigating subgenomes and polyploidy in bamboo lineages. Both *Raddia guianensis* (626 Mb) and *Olyra latifolia* (646 Mb) are diploid bamboos with similar genome sizes. The tetraploid genome sizes of *P. edulis* (1.8 Gb) and *Guadua angustifolia* (1,614 Mb) are also roughly similar. *D. latiflorus* (1,368 Mb, 1C) has a larger genome than the hexaploid *B. amplexicaulis* (848 Mb). Currently, only the genomes of the hexaploid sympodial bamboo *D. latiflorus* (this study) and the tetraploid running bamboo *P. edulis* (Peng et al., 2013; Zhao et al., 2018) have been assembled at the chromosome-level, providing valuable resources for bamboo research.

The *D. latiflorus* subgenomes share homologous syntenic blocks with the 12 *O. sativa* chromosomes and show a 1:6 pattern between *O. sativa* and *D. latiflorus* (Figure S9); this is

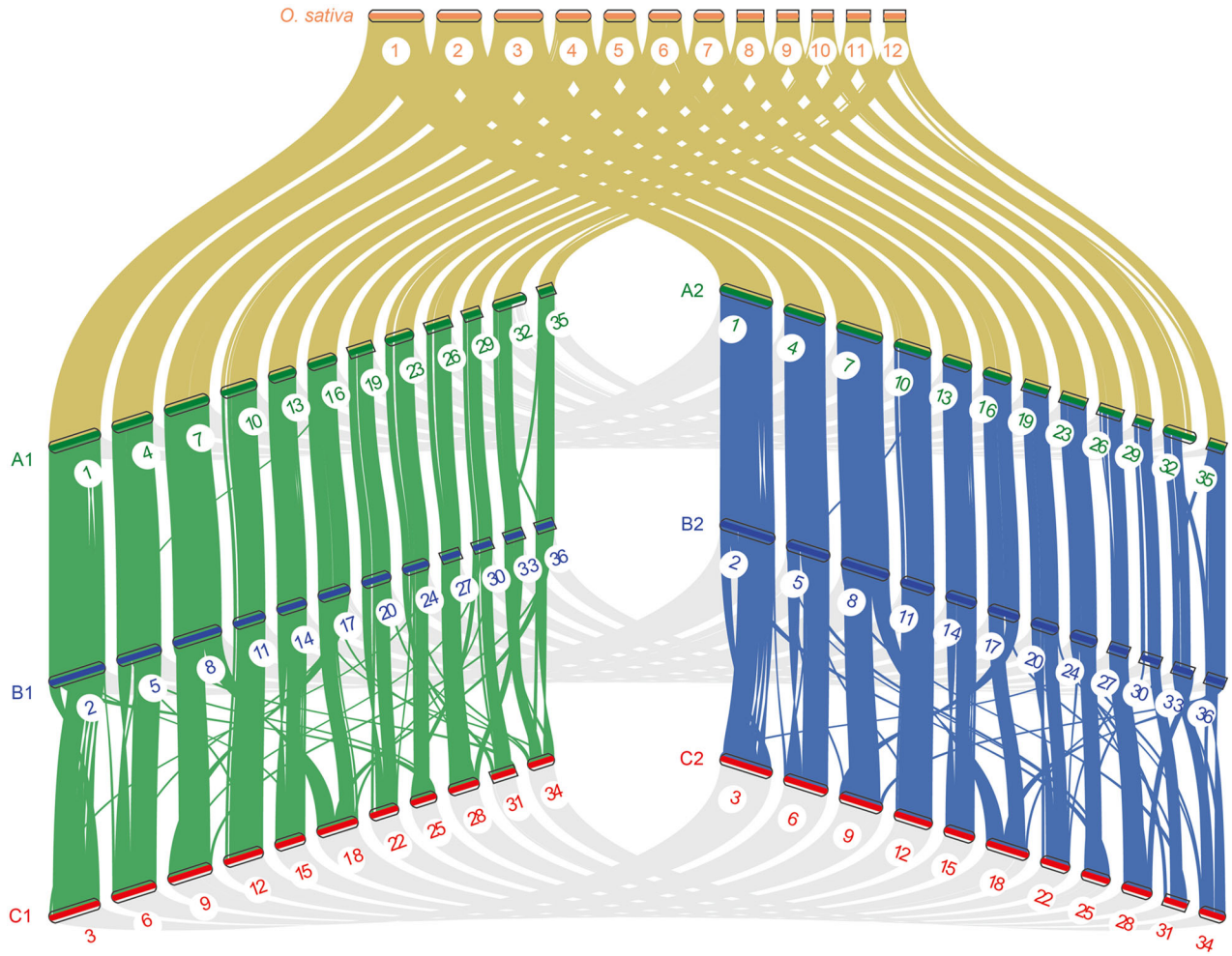


Figure 2. Chromosome-scale assembly of the hexaploid genome of *Dendrocalamus latiflorus*

Karyotype figure representing the hierarchy of synteny between *D. latiflorus* and *Oryza sativa*. Green, blue, and red represent subgenomes A, B, and C, respectively. Yellow lines represent homologous syntenic blocks between the *O. sativa* genome and the subgenomes of *D. latiflorus*. Green (A1B1C1) and blue (A2B2C2) lines represent homologous syntenic blocks among the subgenomes of *D. latiflorus*. Gray lines represent synteny between two allelic chromosomes. *D. latiflorus* chromosomes in this study are numbered Chr1 to Chr36 according to the number of 36 bamboo linkage groups (D1–D36).

consistent with the observation that the basic chromosome number of hexaploid *D. latiflorus* is 12. We aligned a high-density genetic map of bamboo (Guo et al., 2019b) with our assembly and identified 35 linkage groups (LGs), supporting our chromosomal assignment (Figure S10).

Evaluation of the completeness of the genome assembly

To assess the completeness of the assembled *D. latiflorus* genome, we performed BUSCO version 5.2.1 (Benchmarking Universal Single-Copy Orthologs) analysis by searching against the 1,440 highly conserved embryophyta universal benchmarking single-copy orthologs (SCOs). Assessment of BUSCO notation showed that 99.7% of SCOs were complete BUSCOs, confirming that the *D. latiflorus* genome assembled in this study is relatively complete (Table S1). We also assessed the completeness of conserved genes in the *D. latiflorus* genome with Core Eukaryotic Genes Mapping Approach (CEGMA) using 248 core genes. We found that 96.37% of core genes from six eukaryotic model organisms were complete in the *D. latiflorus* genome, as revealed by a combination of tblastn, genewise, and geneid analyses.

We also collected samples of different organs and tissues (Figure 1A) for transcriptome sequencing using RNA-Seq. We aligned all reads from the RNA-Seq data including non-abortive shoot, abortive shoot, leaf, root, stem, internode, and bud tissue using the *D. latiflorus* genome sequence. More than 80% of the transcriptome reads could be aligned in most of these libraries, further confirming that the assembled sequence also represents the majority of the functional transcriptome.

Annotation of protein-coding genes

Using three methods to annotate protein-coding genes, including *de novo* prediction, homology-based searches, and transcriptome-based alignment (Figure S11), we identified 49

571,45,421, and 40,239 protein-coding genes in the AA, BB, and CC subgenomes, respectively (Figure 3A, B). The *D. latiflorus* genome contains 135,231 protein-coding genes, which account for ~180 Mb of the total genome assembly. In total, more than 91% of the annotated genes encode proteins with at least 100 amino acid residues (aa), with an average transcript length of 3,952 bp and a mean of 4.52 exons per gene. The average length of exons and introns is 239 and 817 bp, respectively. A comparison with several close relatives, including *P. edulis* (Peng et al., 2013; Zhao et al., 2018), *O. latifolia*, *G. angustifolia*, *B. amplexicaulis*, and *R. guianensis* (Guo et al., 2019b) indicated that the average transcript length in *D. latiflorus* is similar to that in other species (Figure S12).

Annotation of repetitive sequences

In addition to coding genes, we also identified 1,676 transfer RNAs (tRNAs), 343 ribosomal RNAs (rRNAs), and 3,070 small nuclear RNAs (snRNAs) (Figure S13). By integrating repetitive sequences from *de novo* prediction with the Repbase database and annotation with RepeatMasker software, we showed that 52.65% of the *D. latiflorus* genome could be classified as repetitive regions (Table S1). To generate annotations for the transposable elements (TEs), we identified DNA transposons and retrotransposons in the *D. latiflorus* genome, which account for 2.42% and 50.96% of the total genome sequence, respectively (Figure 3B). Long terminal repeats (LTRs) were the most common TEs in the *D. latiflorus* genome, accounting for 50.42% of the total genome sequences. Finally, 0.93% of the total assembly represents uncharacterized repeats, which were not assigned to any type of TE or repetitive sequence. The percentage of TEs is similar to that in major bamboo lineages and lower than that in *Zea mays*.

Kimura distance-based copy divergence analysis confirmed the strong predominance of LTRs compared to DNA transposons or long interspersed nuclear elements (LINEs) (Figure S14). Using Kimura distance-based copy divergence analysis, we deduced the ages and transposition histories of the TEs. Low and high *K*-values indicate similar (recent) and divergent (ancient) copies, respectively. We detected three major transposition bursts in *D. latiflorus*. An initial burst (high *K*-values) involving LTR, LINE, and DNA transposons was followed by a second burst (moderate *K*-values) of LTR and DNA transposons. The third, most recent burst (low *K*-values) was due to the expansion of LTRs only, with no associated increase in DNA transposons or LINEs (Figure S14).

Comparison of transcriptomes and post-transcriptional regulation among different tissues in *D. latiflorus*

To generate a comprehensive transcriptome atlas, we collected eight types of tissue, including non-abortive shoot, abortive shoot, shoot shell, leaf, root, stem, internode, and bud tissue, and analyzed their transcriptomes by RNA-Seq (Figure 1). We generated multi-dimensional scaling (MDS) plots to illustrate the similarity between repeats and tissues (Figure 4A). Stem, internode, and lateral bud tissue are closer to each other in the plots, which is consistent with their position in the plant. Non-abortive shoot and abortive shoot tissues are closer to each other than to the other tissues (Figure 4A). Leaf, root, and shoot shell tissue are highly dispersed in the MDS plots. A pair-wise comparison revealed the fewest differentially expressed genes (DEGs) (false discovery rate (FDR) < 0.01, fold change > 3) among stem, internode, and lateral bud tissue (Figure 4B), which is consistent with the results in the MDS plots. Shoot shell tissue had more DEGs than internode or lateral bud tissue. We subjected 10 DEGs to quantitative reverse transcription polymerase chain reaction (qRT-PCR), which validated the expression patterns of 70% of the DEGs (Figure S15; Table S1).

Both non-abortive and abortive shoots derived from the stumps of mature bamboo appear quite different from shoot shells in terms of color (Figure 1B). The rate of shoot abortion determines the production of bamboo shoots, but the molecular mechanisms underlying shoot abortion remain mysterious. A pair-wise comparison of non-abortive and abortive shoots showed that the Gene Ontology (GO) terms overrepresented among the DEGs included those related to nutrition, such as cellular glucan metabolic process (p-adjust = 2.10E-05), glycogen biosynthetic process (p-adjust = 0.00075), and polysaccharide catabolic process (p-adjust = 0.0046). Thus, these DEGs might underlie the different levels of energy metabolism between non-abortive and abortive shoots. Interestingly, functional terms related to structural carbohydrates, including cellulose biosynthetic process (p-adjust = 0.033), were also enriched among these DEGs, which might result in the altered structural growth of abortive shoots (Figure 4C). For example, one of these DEGs, encoding xyloglucan endotransglycosylase 4 (*XTR4*), plays an essential role in cellulose biosynthesis and was downregulated in abortive versus normal non-abortive shoots (Figure 4D).

We also identified spermine biosynthetic process (p-adjust = 4.02E-05) and spermidine biosynthetic process (p-adjust = 4.90E-05) among the overrepresented GO terms. The biosynthesis of polyamines, including spermidine, spermine, and putrescine, is associated with chilling stress (Zhao et al., 2017). Thus, perhaps unsuitable temperatures during the germination stage alter polyamine biosynthesis activity in shoots, leading to abortion. Kyoto Encyclopedia of Genes and Genomes enrichment analysis revealed multiple enriched metabolic pathways, including energy metabolism pathways and hormone response pathway (Figure S16). The expression patterns in these specific pathways could be successfully verified by qRT-PCR (Figure S17; Table S1).

Multiple isoforms can be generated from a single locus by alternative splicing (AS), a post-transcriptional mechanism that increases the fitness of the plant (Liu et al., 2017). However, AS has not been reported in *D. latiflorus* due to the challenges in investigating AS without a reference genome. In this study, genes involved in messenger RNA (mRNA) splicing via the spliceosome (GO:0000398) showed obvious

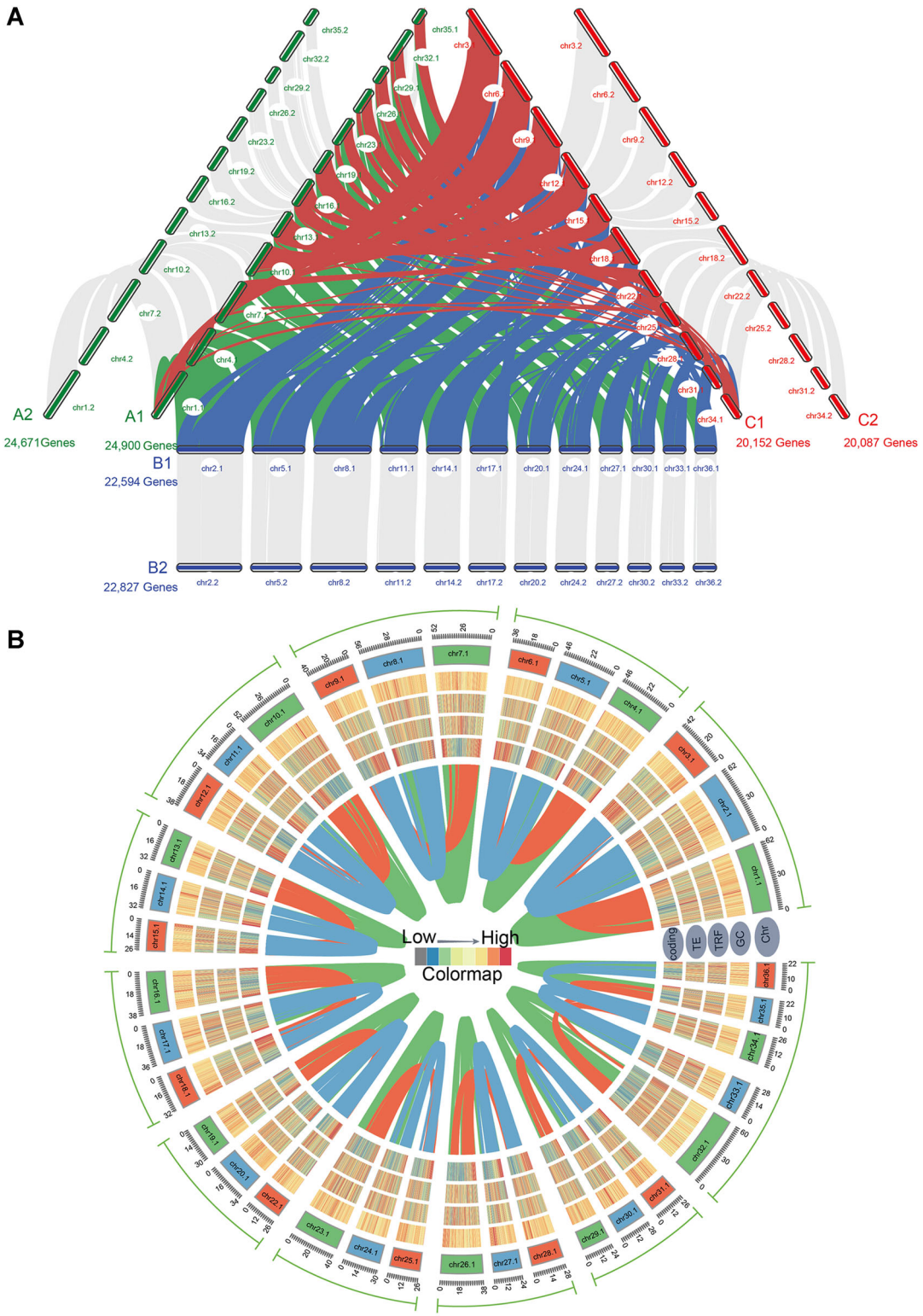


Figure 3. Annotation of the hexaploid genome of *Dendrocalamus latiflorus*

(**A**) Annotated coding genes for 35 chromosome pairs with allele-aware chromosome-level subgenomes (AABBCC). The annotated coding genes for each subgenome are labeled. Green, blue, and red lines represent homologous gene pairs among subgenomes A, B, and C, respectively. (**B**) Circular graph representing the distribution of chromosomal features of the *D. latiflorus* genome (A1B1C1) with integration of coding genes and transposable element (TE) annotation. The outer track indicates each chromosome, which is divided into three subgenomes (A, B, and C), as indicated by green, blue, and red blocks, respectively. For each 100 kb window size, the number of TEs, tandem repeats, and protein-coding genes were calculated. Heatmaps (from the inner to outer circle) represent gene density and TE, tandem repeat (TRE), and GC content. The inner connecting lines indicate homeologous connections among subgenomes.

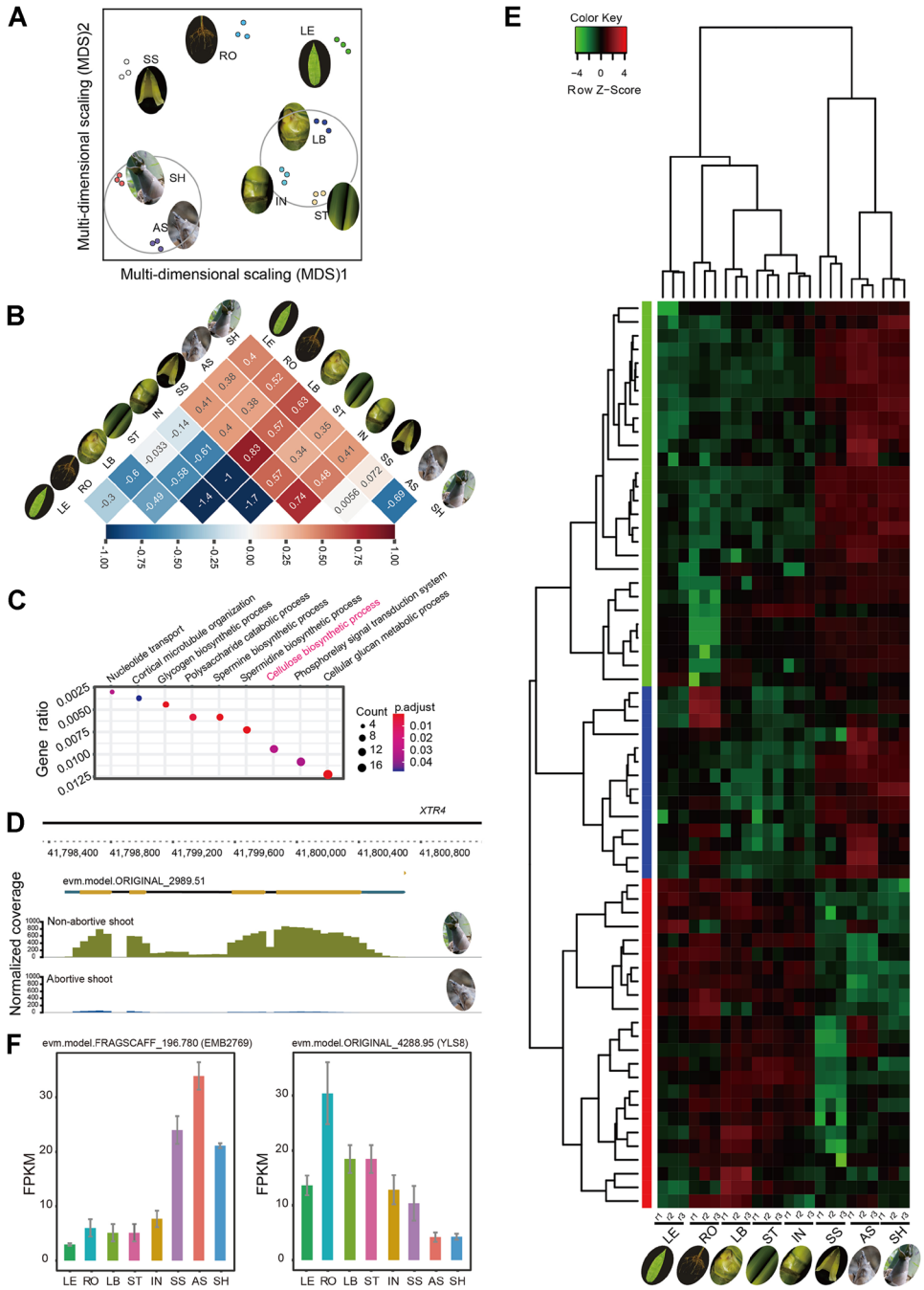


Figure 4. Transcriptome profiling in different tissues

(**A**) Multi-dimensional scaling (MDS) plots to analyze the variation among eight tissues by sklearn.manifold.MDS using $\log_2$(fragments per kilobase of transcript per million mapped reads (FPKM) + 1). (**B**) Heatmap of the number of differentially expressed genes (DEGs) using pair-wise comparisons. Each pair-wise comparison was plotted using values calculated with the following formula: (number of DEGs in each pair-wise comparison − average number of DEGs)/*SD*. (**C**) Gene Ontology (GO) enrichment analysis of DEGs between non-abortive and abortive shoots. The gene ratio represents the percentage of genes related to a GO term/total number of DEGs. (**D**) Wiggle plot showing the downregulated expression of *XTR4* in abortive versus non-abortive shoots. The two wiggle tracks represent the gene expression patterns in non-abortive shoots and abortive shoots, as indicated using the phasing of RNA-Seq data. (**E**) Heatmap of the expression of 67 genes enriched in the GO term messenger RNA splicing, via spliceosome (GO:0000398) based on FPKM values. (**F**) Bar plot showing differential gene expression in non-abortive shoots, abortive shoots, and shoot shells compared with other tissues. (AS, abortive shoot; IN, internode; LB, lateral bud; LE, leaf; RO, root; SH, non-abortive shoot; SS, shoot shell; ST, stem).

differences in expression in non-abortive shoots, abortive shoots, and shoot shells compared to other tissues (Figure 4E). For example, *EMB2769* and *YLS8* encode RNA binding proteins involved in mRNA *cis* splicing via the spliceosome. Both genes were differentially expressed in shoots compared to other tissues (Figure 4F), suggesting that AS might play a role in shoot development. There was little difference in the expression levels of genes involved in RNA processing between non-abortive shoots and abortive shoots. However, different splicing events between non-abortive and abortive shoot tissue occurred in eight splicing processing-related genes, which is consistent with the finding that most splicing factors are regulated by AS of their own transcripts.

To uncover potential AS-related regulation during shoot development, we also used our transcriptome atlas to identify differential AS events in *D. latiflorus*. We classified each read based on SNP calling using the phasing pipeline (Feng et al., 2021) (Figure 5A). AS events, including a retained intron (RI), skipped exon (SE), alternative 5' splice site (A5SS), and alternative 3' splice site (A3SS), were identified using phasing reads (Figure 5B). Based on the results of Iso-Seq phasing, AS events were enriched in RNA processing, nonsense-mediated decay, and negative regulation of mRNA polyadenylation (Figure 5C). The enrichment of epigenetic factors suggests that AS affects histone methylation and chromatin remodeling genes (Figure 5C). Following RNA-Seq phasing, we identified different alternative splicing events among the eight tissues using replicate multivariate analysis of transcript splicing (rMATs). Shoots and abortive shoots showed the most differential AS events compared with other tissues (Figure 5D). This result suggests that non-abortive and abortive shoots show differential AS events, which is consistent with the identification of DEGs encoding splicing factors in these tissues (Figure 4E). For RI events, abortive shoots showed more intron inclusion isoforms than non-abortive shoots. However, non-abortive shoots showed more exon inclusion isoforms in SE events than abortive shoots (Figure 5E). For example, evm.model.ORIGINAL_4797.35, which is involved in pseudouridine synthesis, shows an obvious retained intron in abortive versus non-abortive shoots (Figure 5F).

These genes with differential AS patterns were enriched in GO terms related to mRNA splice site selection (Figure 5G); this is consistent with the observation that genes encoding splicing factors are highly regulated by AS (Reddy, 2007). Previous comparative transcriptome analysis of naturally occurring slow-growing bamboo mutants with abnormal internode development revealed significantly downregulated genes related to auxin, vesicle transport, and cytoskeleton organization (Wei et al., 2018). Interestingly, we found a similar enrichment of DEGs involved in responses to hormones between non-abortive and abortive shoots. In addition to the above GO terms, several other terms were enriched, such as circadian rhythm and regulation of signal transduction (Figure 5G). These differential AS events provide valuable information that could serve as a basis to further explore the molecular mechanisms leading to shoot abortion.

Structural variations, expressional bias, and AS between allelic genes

We selected 11,630, 11,134, and 10,670 allelic gene pairs and identified 1,001,491, 937,692, and 809,164 sequence variations (including SNV and InDels) in subgenomes A, B, and C, respectively. These sequence variations (located within allelic genes) provided basic allelic variants for isolating allelic transcripts between two haplotypes. To accurately identify allele-specific expressed (ASE) genes, we calculated the normalized expression (reads per million reads (RPM)) of all allelic genes based on phased PacBio Iso-Seq to analyze the differences in gene expression in two parental genomes. We visualized allele-specific gene expression in an MA (log ratio/mean average) plot (Figure S18A). Finally, we selected ASE genes using a cutoff RPM $>$ 100 to avoid loci with low-abundance transcripts. In total, we identified 554, 534, and 578 ASE genes in subgenomes A, B and C, respectively.

The ASE genes in subgenome A (Figure S18B) were enriched in the GO terms photosynthesis, cortical microtubule organization, and adenosine triphosphate (ATP) hydrolysis coupled proton transport, whereas the ASE genes in subgenome B (Figure S18C) were enriched in vesicle-mediated transport, proteolysis involved in cellular protein catabolic process, ubiquitin-dependent protein catabolic process, and protein N-linked glycosylation. The ASE genes in subgenome C (Figure S18D) were enriched in the GO terms mitochondrial electron transport, ATP synthesis coupled proton transport, photosynthesis, and mRNA splicing. There was no obvious difference in the percentage of AS genes between allelic gene pairs (Figure S19A). However, we identified 5 358 and 7 437 AS genes that were only detected in the haplotypes A and haplotypes B genomes, respectively. GO enrichment analysis of these AS genes revealed that the alleles from haplotypes A were enriched in the terms of microtubule-based movement, response to hormone, and histone methylation (Figure S19B), whereas alleles from haplotypes B were enriched in protein deubiquitination, mRNA processing, cellular response to nitrate, and cytoskeleton organization (Figure S19C).

Evolutionary analysis of *D. latiflorus* and homologous species

Analysis of orthologs in the sequenced genomes of 15 other species revealed 44,665 gene families (Figure 6A). Among these, the 15 sequenced genomes share a core set of 3,750 gene families. Compared with *P. edulis*, *O. latifolia*, *R. guianensis*, *G. angustifolia*, and *B. amplexicaulis*, we identified 6,919 specific gene families (including 20,163 genes) in *D. latiflorus* (Figure 6B). *D. latiflorus* and *B. amplexicaulis*, which evolved from paleotropical woody bamboos, were clustered together in a branch of the phylogenetic tree. The divergence time between *D. latiflorus* and *B. amplexicaulis* was estimated to be 10.5–16.5 million years ago (Figure 6C).

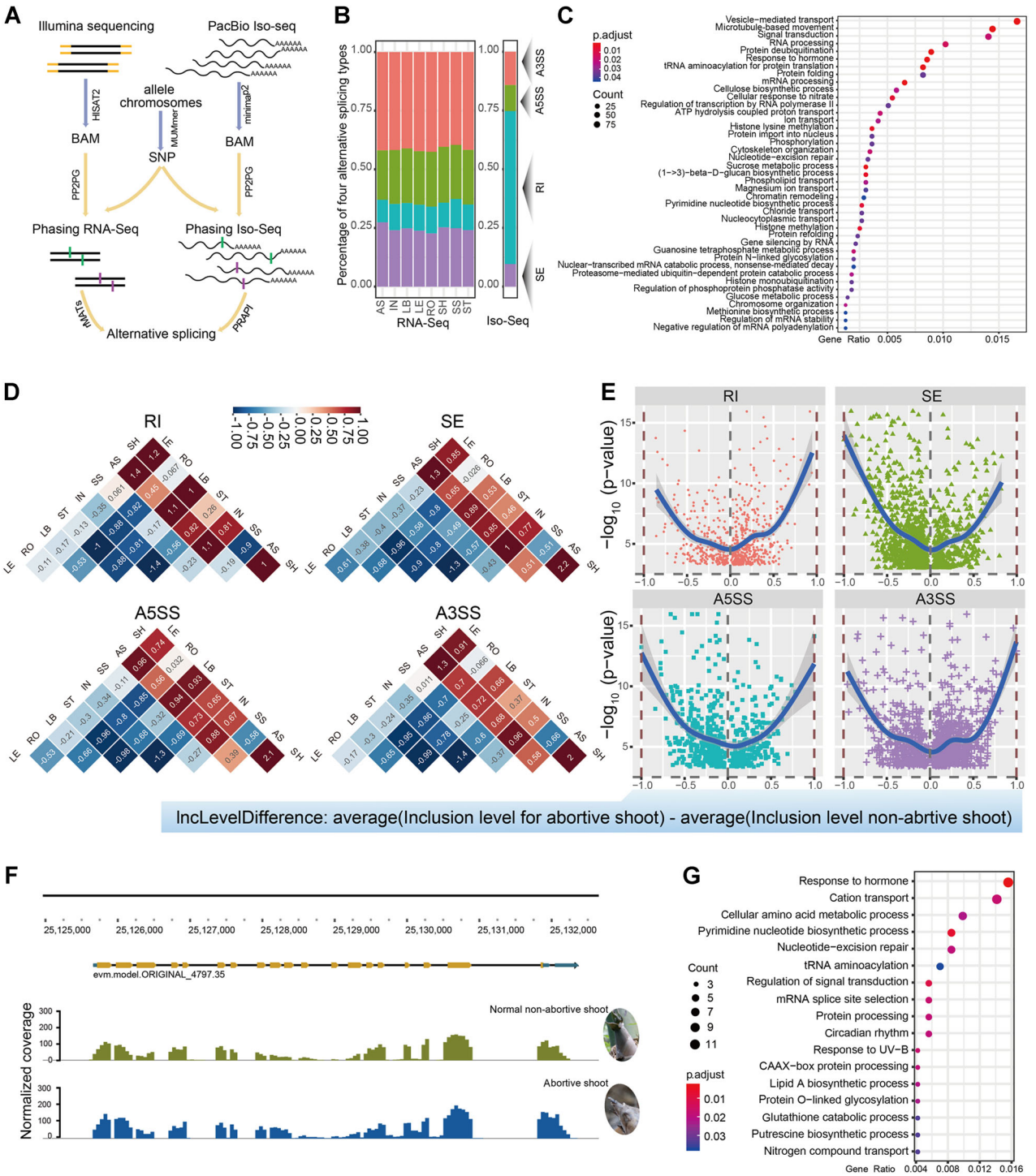


Figure 5. Differential splicing events between non-abortive and abortive shoots

(**A**) Flowchart for the phasing of PacBio isoform sequencing (Iso-Seq) and Illumina RNA-Seq data, respectively. (**B**) Percentages of the four types of alternative splicing (AS). SE, skipped exon; RI, retained intron; A5SS, alternative 5' splice site; A3SS, alternative 3' splice site. (**C**) Enriched Gene Ontology (GO) terms for AS events based on Iso-Seq. (**D**) Differential splicing events among pair-wise comparisons. Each pair-wise comparison was plotted using values calculated according to the following formula: (number of differential AS genes in each pair-wise comparison − average number of differential AS genes)/*SD*. (**E**) Volcano plot of differential splicing events between non-abortive and abortive shoots. *x*-axis and *y*-axis represent the subtracted value of average inclusion level between non-abortive and abortive shoots and $-\log_{10}$ (*P*-value), respectively. (**F**) Example of a differentially spliced gene, encoding a protein involved in pseudouridine synthesis. Two wiggle tracks represent the differential intron retention events between normal non-abortive shoots and abortive shoots based on RNA-Seq phasing data. (**G**) GO enrichment of differentially spliced genes. The gene ratio represents the percentage of genes related to a GO term/total number of differentially spliced genes.

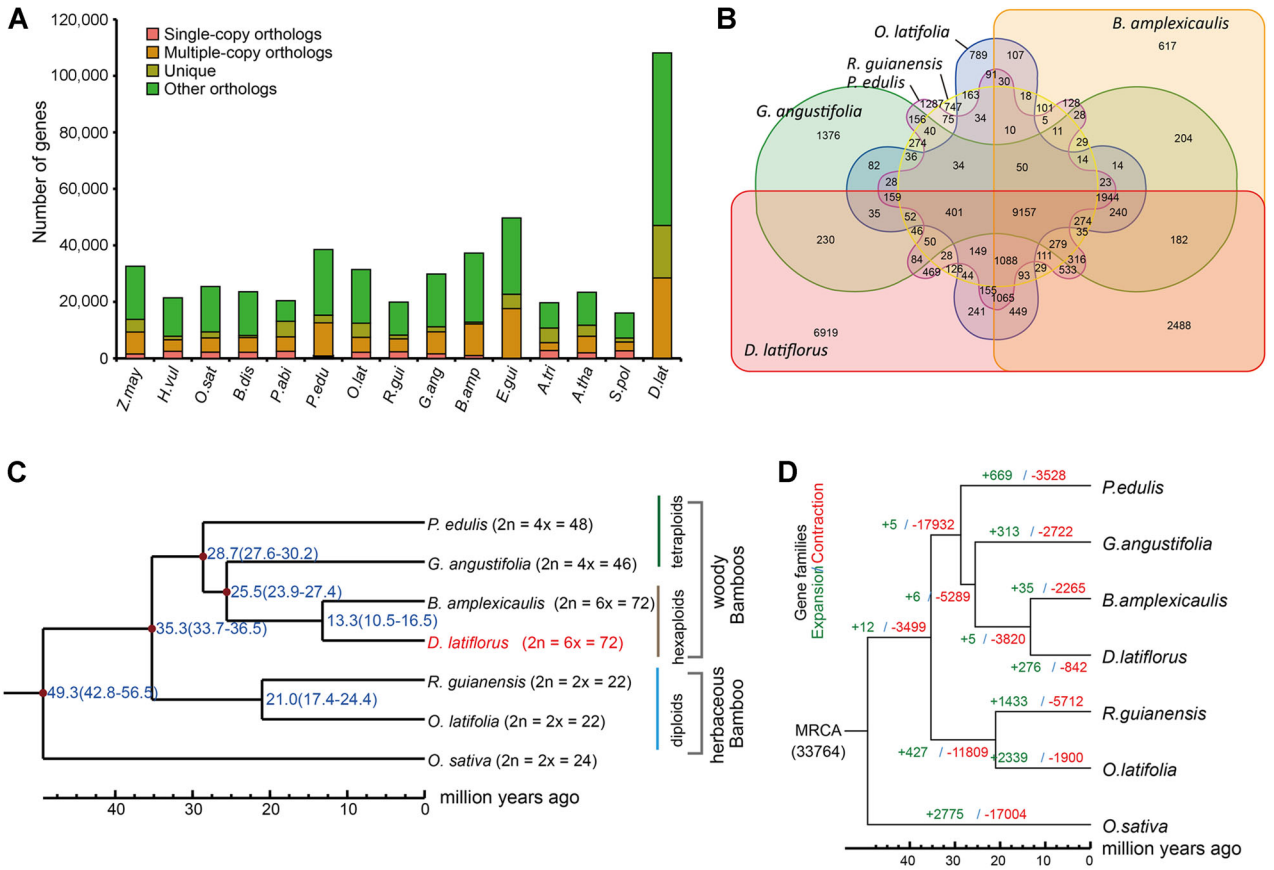


Figure 6. Evolutionary analysis of *Dendrocalamus latiflorus* and homologous species

Cluster and evolutionary analysis of gene families in the 15 species examined in this study: *Zea mays*, *Hordeum vulgare*, *Oryza sativa*, *Brachypodium distachyon*, *Picea abies*, *Phyllostachys edulis*, *Olyra latifolia*, *Raddia guianensis*, *Guadua angustifolia*, *Bonia amplexicaulis*, *Elaeis guineensis*, *Amborella trichopoda*, *Arabidopsis thaliana*, *Spirodela polyrhiza*, and *D. latiflorus*. (**A**) Bar plot showing the distribution of single-copy and multiple-copy orthologs. *x*-axis and *y*-axis represent species and gene numbers in the four categories, respectively. (**B**) Venn diagram showing the number of overlapping and unique gene families. (**C**) Estimation of divergence time. The numbers on the nodes represent the divergence times. (**D**) The expansion and contraction of gene families. The numbers in green and red separated by a slash indicate the expanded and contracted gene families, respectively.

We estimate that 276 gene families (10,057 genes) have undergone expansion and 842 (4,221 genes) have undergone contraction (Figure 6D). The enriched GO terms for the contracted families include ATPase activity, trehalose biosynthetic process, cell communication, response to light stimulus, recognition of pollen, and so on. The contraction of gene families involved in the reproductive process might have affected the flower physiology of *D. latiflorus*. The enriched GO terms for the expanded families include cysteine-type peptidase activity, protein dimerization activity, hydrolase activity, acting on ester bonds, zinc ion binding, transposition, telomere maintenance, peptidase activity, hydrolase activity, DNA helicase activity, DNA binding, proteolysis, helicase activity, cellular response to stress, chromosome organization, DNA repair, and so on. These results suggest that the expansion of genes involved in telomere maintenance and DNA repair likely plays an important role in the long-lived vegetative growth of *D. latiflorus*.

D. latiflorus is a woody sympodial bamboo, with a distinct morphology from the running bamboo *P. edulis*. Both overexpression and CRISPR-Cas9 gene editing are available to elucidate gene function in *D. latiflorus*. However, the transformation of *P. edulis* remains a challenge. Transformation efficiency might be affected by multiple developmental genes, including *WUSCHEL* (*WUS*), *BABY BOOM* (*BBM*), *SHOOT MERISTEMLESS* (*STM*), *B-type ARABIDOPSIS RESPONSE REGULATOR* (*ARR*), *WOUND-INDUCED DEDIFFERENTIATION* (*WIND*), and so on (Nagle et al., 2018). To obtain clues about the transformation bottleneck in *P. edulis*, we focused on the synteny of these developmental genes. First, we compared the 24 chromosomes of *P. edulis* with *D. latiflorus* by whole-genome alignment and identified 467 560, 435,836, and 437,516 alignment blocks for subgenomes A, B, and C, respectively. We then analyzed the syntenic blocks between *P. edulis* and *D. latiflorus* using annotated genes, revealing 22,184, 12,080, and 20,463 gene pairs in subgenomes A, B, and C, respectively (Figure 7A). Above dot plot synteny comparison divided the 24 chromosomes of tetraploid *P. edulis* into two groups (Figure 7B). Analysis of global chromosomal synteny revealed that chromosomes 22, 5, and 2 of group 1 from *P. edulis* are syntenic to

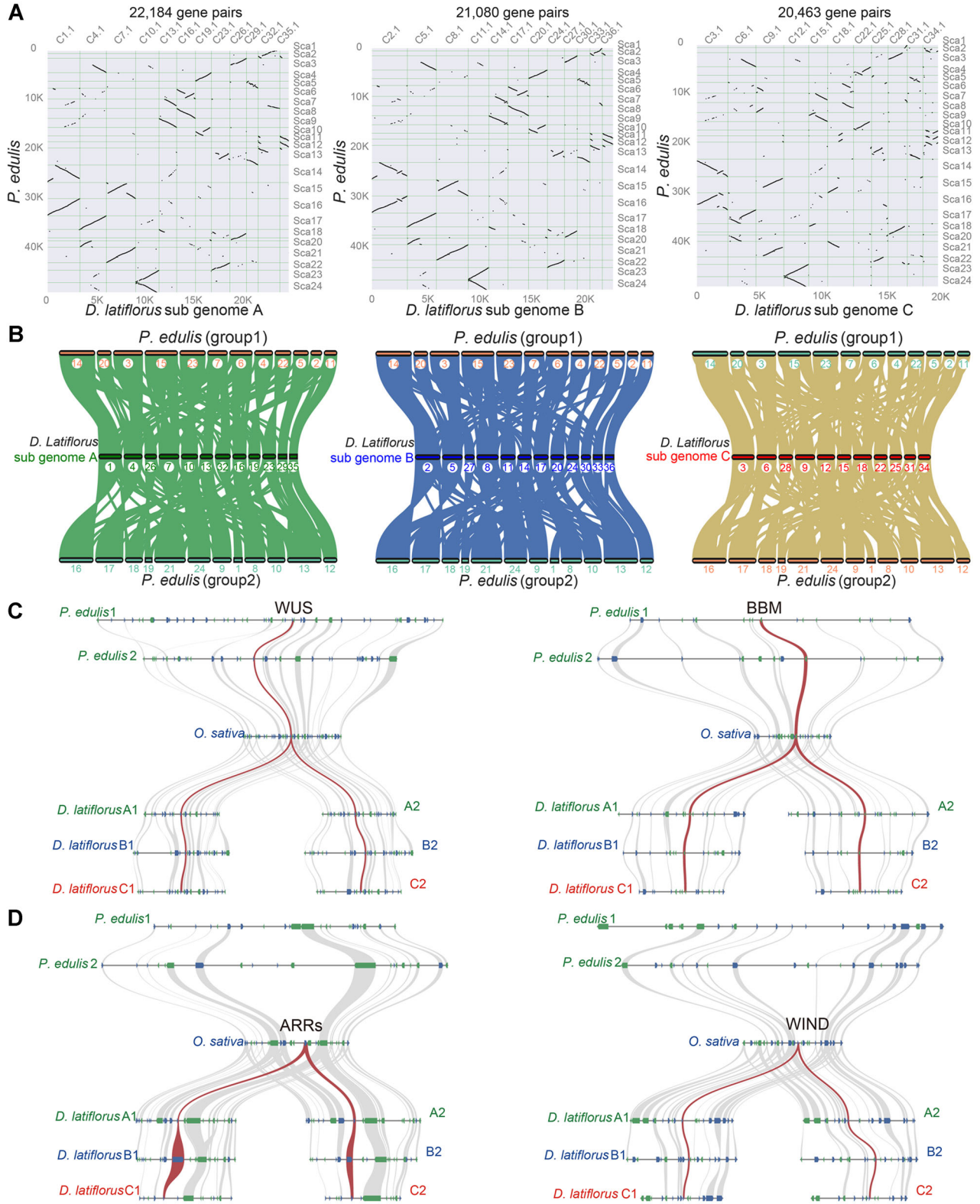


Figure 7. Synteny comparison between *Phyllostachys edulis* and *Dendrocalamus latiflorus*

(**A**) Dot plot showing the synteny comparison between *P. edulis* and *D. latiflorus*. (**B**) Macro-synteny between *D. latiflorus* and *P. edulis*. The green, blue, and yellow lines represent the syntenic relationships of *P. edulis* with subgenomes A, B, and C, respectively. (**C**) Micro-synteny analysis of *D. latiflorus* with *Oryza sativa* and *P. edulis*. Red lines in the left and right panels represent homologous *WUS* and *BBM* genes, respectively, from *P. edulis, O. sativa*, and *D. latiflorus*. Gray lines are flanked micro-syntenic genes. (**D**) Red lines in the left and right panel represent *ARRs* and *WIND*, respectively, from *O. sativa* and *D. latiflorus*.

chromosomes 25, 31, and 34 from subgenome C in *D. latiflorus*. However, chromosome 13 of group 2 from *P. edulis* shares synteny with the above three chromosomes (22, 5, and 2) from *P. edulis*, suggesting a possible fusion/fission event (Figure 7B). Most of the synteny blocks show global collinearity with *D. latiflorus* (Figure 7B).

For developmental genes related to transformation efficiency, micro-synteny visualization revealed that the orthologous pairs *WUS* and *BBM* share collinearity in gene order between *P. edulis* and *D. latiflorus* (Figure 7C). This type of synteny relationship was also observed in other members, including *STM*, *LEC2*, *HD3*, and *ESR* (Figure S20). However, *ARRs* and *WIND* did not show syntenic relationships in *P. edulis* (Figure 7D). ARR activates *WUS* transcription, and WIND1 enhances callus formation (Nagle et al., 2018). Thus, the lack of syntenic relationships of *ARR* and *WIND1* in *P. edulis* after its divergence from *D. latiflorus* might affect the transformation.

DISCUSSION

The draft genomes of four bamboo species including *O. latifolia*, *R. guianensis*, *G. angustifolia*, and *B. amplexicaulis* were recently obtained using Illumina short reads with different insert sizes (Guo et al., 2019b). However, the genome of *D. latiflorus* had not been made available due to its large genome size and its hexaploid nature, which pose many challenges to high-quality genome assembly and analysis. However, *D. latiflorus* is the only bamboo species for which overexpression (Ye et al., 2017) and genome editing (Ye et al., 2020) protocols are currently available. Thus, the genome sequence of *D. latiflorus* was required to enable its use as a model bamboo species.

In this study, we used PacBio third generation long reads to overcome the disadvantage of Illumina short reads assembly and the high error rates of Oxford Nanopore Technologies (ONT) (Tyler et al., 2018). We generated three allele-aware chromosome-level subgenome assemblies of *D. latiflorus* (2 737 182 539 bp), which is larger than that of the hexaploid species *B. amplexicaulis* (848 Mb of scaffolds) (Guo et al., 2019b). The *D. latiflorus* genome contains 135 231 annotated genes, a number comparable to those of other allopolyploid crops, such as *A. hypogaea* (108 604 gene loci) (Chen et al., 2019), *T. aestivum* (124 201 gene loci) (International Wheat Genome Sequencing, 2014), and *B. napus* L. (101 040 gene loci) (Chalhoub et al., 2014).

Allele-aware chromosome-level genome assemblies have been reported for the autopolyploid sugarcane (*S. spontaneum*) (Zhang et al., 2018) and the autotetraploid cultivated *M. sativa* (Chen et al., 2020), providing useful information for introducing tetra-allelic mutations using CRISPR/Cas9-based genome editing. We previously established a CRISPR/Cas9-mediated genome editing protocol for *D. latiflorus* and generated mutants of the phytoene synthase gene *PSY1* and *GA-responsive gene 1* (*GRG1*), which exhibited an albino phenotype and increased plant height, respectively (Ye et al., 2020). Here, we constructed an allele-aware chromosome-level genome assembly using ALLHiC (Zhang et al., 2019) based on Hi-C data, representing a valuable resource that could be used to identify heterozygous or homozygous mutants in all six alleles in the hexaploid species *D. latiflorus*. The newly generated allele-aware chromosome-level genome assembly lays the foundation for CRISPR-Cas9-mediated parental (paternal/maternal) allele-specific targeting (Wu et al., 2020). In this study, we identified 135 231 annotated genes from our allele-aware annotation (Figure 3), increasing our understanding of the allele-specific transcriptome of *D. latiflorus* (Figure S18).

In total, DNA transposons and retrotransposons accounted for 52.65% of the total genome sequence of *D. latiflorus*, similar to the percentage for Moso bamboo (Peng et al., 2013). Among these TE sequences, a previous study revealed that 982 LTR-retroelement families accounted for more than 40% of the genome in Moso bamboo (Zhou et al., 2017). Consistent with this finding, we determined that LTR retroelements were the most common TEs in the *D. latiflorus* genome, accounting for 50.42% of the total genome. It would be interesting to investigate the impact of TEs on the polyploidy of *D. latiflorus*. Our findings provide a valuable resource to investigate the roles of the epigenetic regulation of these TEs in flanking gene expression.

Counting of mitotic metaphase chromosomes revealed 72 chromosomes in *D. latiflorus* seedling samples (Qiao et al., 2013). In the current study, we assembled 35 chromosome pairs but did not detect one chromosome pair (D21), the smallest chromosome pair based on genetic linkage mapping (Guo et al., 2019b). To exclude the possibility that D21 was fused with other chromosomes during whole-genome assembly, we examined the Hi-C interaction matrix, which did not present obvious breakpoint regions. However, we cannot exclude the possibility that the strong homology of D21 with other chromosomes might prevent the aligner from detecting the potential artifact interaction signal of the smallest chromosome due to sequence similarity.

The progenitor genomes of woody bamboos contained four distinct ancestral subgenomes (subA to subD) (Guo et al., 2019b). It appears that the ancestral B/C subgenomes hybridized to generate allotetraploid lineages (BBCC), which further hybridized to the ancestral subgenome A to generate allohexaploid woody bamboos (AABBCC) (Guo et al., 2019b). Our subgenome assembly supports the presence of a distinct progenitor of subgenome C, which showed a different arrangement from subgenomes A and B. Allohexaploid wheat contains three closely related subgenomes (Ling et al., 2013). A recent study revealed that subgenome-divergent regulation is associated with the dynamic modulation of epigenetic regulatory elements (epiREs) in wheat (Wang et al., 2021a). The assembly of subgenomes A, B, and C generated in this study provides a valuable resource for future investigations of the role of epiREs among the three subgenomes (AABBCC) in determining the long flowering interval and rapid growth of bamboo.

Based on pairwise alignment of allele chromosomes using MUMmer (Delcher et al., 2003), we identified SNPs in subgenomes A, B, and C (Figure S7). Allelic SNPs can be used for phasing analysis of the transcriptome. Iso-Seq phasing revealed that RI comprised the majority of AS events in *D. latiflorus* (Figure 5B), which is consistent with a previous report for Moso bamboo (Wang et al., 2017). However, RNA-Seq phasing tended to underestimate the most prevalent RI type (Figure 5B). Long reads from Iso-Seq have the advantage of being able to produce higher separation ratios than short reads from RNA-Seq during the phasing process (Feng et al., 2021). Reads that could not be assigned to any allelic chromosome were discarded for downstream analysis, which might lead to discontinuous regions and a low percentage of RI identification. With the improvement of sequencing depth using the PacBio or Nanopore platform (Zhao et al., 2019), transcriptomes with long reads could be generated, providing advantages during the phasing process.

In conclusion, a transformation system has not yet been established for sequenced bamboo species, such as Moso bamboo; thus, there is an urgent need to develop a model species for bamboo research. Here, the generation of a high-quality reference genome sequence for *D. latiflorus* provided interesting insights into the genomic and transcriptome structure of this hexaploid species, which will be invaluable for subsequent functional studies. Furthermore, we conducted transcriptome analysis to investigate how differential expression or splicing affects shoot growth in *D. latiflorus*. Our data represent valuable resources for designing breeding strategies within the Poaceae family, which includes agronomically important cereal crops such as rice and maize.

MATERIALS AND METHODS

Analysis of chromosome number in *D. latiflorus*

A fresh, tender *D. latiflorus* root tip was incubated in 0.04% 8-hydroxyquinoline solution at 4°C for 4 h in the dark. After washing with distilled water, the sample was transferred to Carnoy's solution (anhydrous ethanol : glacial acetic acid = 3:1) and incubated at 4°C for 72 h in the dark. The sample was washed with distilled water, incubated at room temperature for 1 h, transferred to 0.25 mol HCl, and hydrolyzed for 10 min at room temperature. The sample was then washed with distilled water, incubated for 10 min, transferred to a centrifuge tube containing mixed enzyme solution (1.5% pectinase and 3% cellulose), and hydrolyzed in a 37°C water bath for 2.0–2.5 h. After removing the enzyme solution, the sample was washed with distilled water. A small amount of distilled water was added to the tube to achieve hypotonicity and incubated at room temperature for 1 h. The root tip was aspirated and dripped onto a glass slide. The sample was dried by blotting with filter paper, and 1–2 drops of freshly prepared Carnoy's solution were added the sample. After the solution was dried, a drop of Carbol fuchsin was added. Finally, the sample was examined by microscopy.

DNA FCM of *D. latiflorus*

Nuclear DNA content in *D. latiflorus* was examined by FCM as described previously (Loureiro et al., 2007). In brief, fresh leaves from seedlings were placed in a glass Petri dish. After adding 1.5 mL woody plant buffer (WPB), the leaves were cut into a paste using a sharp blade. During this process, the materials were immersed in the dissociation solution to better dissociate the cell nuclei. The sample was filtered through 400-screen mesh nylon into a 1.5 mL centrifuge tube, and the filtrate was incubated at 4°C for 2 min. The sample were centrifuged at 1000 r/min (89.46 *g*) for 2 min at 4°C, the supernatant was discarded, and 600 μL WPB buffer and 60 μL of pre-chilled propidium iodide staining solution were added to the pellet. The sample was incubated at 4°C in the dark for 15 min and subjected to FCM to detect nuclear DNA content.

Genomic DNA extraction

D. latiflorus primarily reproduces by asexual reproduction, and floral tissues are seldom observed due to the long juvenile stage. For genome sequencing, we used young leaf tissue from cultivated *D. latiflorus* from the Botanical Garden of Bamboo at Fujian Agriculture and Forestry University (E119°140; N26°050) rather than tissue culture material, thereby avoiding potential artifacts generated during tissue culture (Qiao et al., 2017). All samples were immediately frozen in liquid nitrogen. Genomic DNA was extracted from leaf samples using a Plant Mini kit (Qiagen, Hilden, Germany). DNA concentration was quantified using a Qubit fluorometer (Thermo Fisher Scientific). Agarose gel electrophoresis was used to confirm the purity and integrity of the DNA.

DNA sequencing using the Illumina HiSeq X Ten platform

Paired-end Illumina libraries with 350 bp insert sizes were constructed using Illumina TruSeq Nano DNA Library Prep Kits according to the standard protocol provided by Illumina (San Diego, CA, USA) and sequenced on the Illumina HiSeq X Ten sequencer.

DNA sequencing using the PacBio Sequel platform

PacBio sequencing libraries with an insert size of 20 kb were constructed using SMRTbell template Prep Kits according to the manufacturer's recommendations. Sequencing was performed using a PacBio Sequel instrument with P6-C4 sequencing reagent. Libraries for SMRT PacBio genome sequencing were constructed following the manufacturer's standard protocols. Briefly, high molecular weight genomic DNA was sheared to ~20 kb target size, followed by damage repair and end repair, blunt-end adaptor ligation, and size selection. Finally, the libraries were sequenced on the PacBio Sequel platform (Continuous Long reads (CLR) sequencing mode).

Genome survey

SOAPdenovo (Li et al., 2010) was used to generate the surveyed genome using *k*-mer 41 for assembly. We calculated

the genome size and heterozygosity using Genomescope (Vurture et al., 2017) with 17 *k*-mers.

De novo assembly

Before assembling, "daligner" executed by the main script of the FALCON assembler (Chin et al., 2016) was used to correct PacBio long reads errors with PacBio short reads <7 000 bp to generate consensus sequences. After error correction, the consensus sequences can achieve accuracies up to 99.999%. FALCON then identified the overlaps between all pairs of the preassembled error-corrected reads. The read overlaps were used to construct a directed string graph following the Myers' algorithm. Contigs were constructed by finding the paths from the string graph (falcon_sense_option = – output_multi –min_idt 0.70 –min_cov 4 –max_n_read 300 –n_core 22; overlap_filtering_setting = –max_diff 200 –max_cov 200 –min_cov 2 –n_core 30). Error correction of the preceding assembly was performed using the consensus-calling algorithm Quiver (Chin et al., 2013). The Illumina reads were also used to correct the contigs with Pilon (Walker et al., 2014). Heterozygosity was then removed from the error-corrected contigs using purge_haplotigs with the Illumina clean reads.

10X Genomics scaffold extension was performed with fragScaff software (Adey et al., 2014). The procedure was as follows. (1) Linked-reads generated using the 10X Genomics library were aligned to the consensus sequences of the PacBio assembly results to obtain the superscaffold using BOWTIE v2.2 (Langmead and Salzberg, 2012). When the actual distance of consensus sequences increases, the number of linked reads that support this connection decreases. Consensus sequences without linked-reads support were filtered out, and only consensus sequences with linked-reads support were used for subsequent assembly. FragScaff uses the parameters "-fs1 ′-m 3000 -q 30 -E 30000 -o 60000′ -fs2 ′-C 1′ -fs3 ′-j 1.25 -u 2."

Repeat annotation

For repetitive sequence annotation, a combined strategy based on homology alignment and *de novo* searches to identify whole genome repeats was applied with the repeat annotation pipeline. Tandem repeats were extracted using Tandem Repeats Finder (TRF) by *ab initio* prediction (Benson, 1999). Homologous TEs were predicted using the Repbase (http://www.girinst.org/repbase) database (Jurka et al., 2005) with default parameters employing RepeatMasker (http://www.repeatmasker.org/) software (Tempel, 2012) and its in-house scripts (RepeatProteinMask) with default parameters to the extracted repeat regions. *Ab initio* prediction was performed with a *de novo* repetitive elements database with LTR_FINDER (Xu and Wang, 2007) (http://tlife.fudan.edu.cn/ltr_finder/), RepeatScout (http://www.repeatmasker.org/) (Price et al., 2005), followed by RepeatModeler (http://www.repeatmasker.org/RepeatModeler.html) (Chen, 2004) with default parameters. All repeat sequences >100 bp long with gap "*N*" < 5% constituted the raw TE library. A custom library (a combination of Repbase and our *de novo* TE library, which was processed by UCLUST algorithm to yield a non-redundant library) were analyzed using RepeatMasker for DNA-level repeat identification.

Structural annotation

Three methods were employed for structural annotation: *de novo* prediction, homology-based predication, and transcriptome-assisted prediction. For *de novo* prediction, we used five tools, including Augustus (v3.2.3) (Stanke and Morgenstern, 2005), Geneid (v1.4) (Guigo et al., 1992), Genescan (v1.0), GlimmerHMM (v3.04), and SNAP (2013-11-29) with default options. For homolog prediction, homologous protein sequences were downloaded from the Ensembl/National Center for Biotechnology Information (NCBI)/other databases and aligned to the *D. latiflorus* genome using TblastN (v2.2.26; *E*-value ≤ 1e – 5). The matching proteins were aligned to the homologous genome sequences for accurate spliced alignments with GeneWise (v2.4.1) software (Birney et al., 2004), which was used to predict gene structure contained in each protein region.

For transcriptome-based prediction, the transcriptome reads from different tissues were aligned to *D. latiflorus* genome sequences using HISAT (v2.0.4) with default parameters to identify exon and splice junction regions (Kim et al., 2019). The Binary Alignment Map (BAM) results were used as input for StringTie (v1.3.3) (Pertea et al., 2015) with default parameters for genome-based transcript assembly. Unigene assemblies were also generated with Trinity (v2.1.1) (Grabherr et al., 2011) using the transcriptome reads. The non-redundant reference gene structure was generated and integrated by merging genes predicted based on the above methods with EVM (Haas et al., 2008) using PASA (Haas et al., 2008).

Functional annotations

The putative protein sequences were aligned to the Swiss-Prot (Bairoch and Apweiler, 2000) and DIAMOND (v0.8.22) databases using BLASTP with *E*-value threshold ≤ 1e – 5 to assign gene function according to the best match. Protein domains were annotated using InterProScan70 (v5.31) (Mulder and Apweiler, 2008) by searching against publicly available databases, including ProDom, PRINTS, Pfam, SMRT, PANTHER, and PROSITE. The GO IDs for each gene were assigned according to the corresponding InterPro entry.

Annotation of ncRNAs

tRNAs were predicted using the program tRNAscan-SE (http://lowelab.ucsc.edu/tRNAscan-SE) (Lowe and Eddy, 1997). For rRNAs (which are highly conserved), rRNA sequences of related species were used as references, and rRNA sequences were predicted using BLAST. Other ncRNAs, including miRNAs and snRNAs, were identified by searching against the Rfam database with default parameters using infernal software (http://infernal.janelia.org/).

Construction and sequencing of Hi-C libraries

Leaves were fixed in 1% formaldehyde solution in MS buffer (10 mmol/L potassium phosphate, pH 7.0; 50 mmol/L NaCl;

0.1 mol/L sucrose) at room temperature for 30 min under a vacuum. After fixation, the leaves were incubated in MC buffer (0.15 m glycine, 10 mM potassium phosphate, pH 7.0, 50 mM NaCl; 0.1 m sucrose) with 0.15 mol/L glycine at room temperature for 5 min under a vacuum. Approximately 2 g fixed tissue was homogenized in liquid nitrogen, resuspended in nuclei isolation buffer, and filtered through a 40 nm cell strainer. The procedures for enriching nuclei from flow-through and subsequent denaturation were based on a 3C protocol established for maize; the chromatin extraction step was similar to that used in the DNase I digestion experiment. The procedures were similar to those described previously. Briefly, chromatin was digested for 16 h with 400 U of the restriction enzyme HindIII (NEB) at 37°C. DNA ends were labeled with biotin and incubated at 37°C for 45 min, and the enzyme was inactivated with 20% sodium dodecyl sulfate solution. DNA ligation was performed by adding T4 DNA ligase (NEB) and incubating at 16°C for 4–6 h. After ligation, proteinase K was added to reverse cross-linking, and the sample was incubated at 65°C overnight. The DNA fragments were purified and dissolved in 86 μL of water. Unligated ends were then removed, and the purified DNA was fragmented to a size of 300–500 bp and end repaired. The biotin-labeled DNA fragments were separated on Dynabeads® M-280 Streptavidin (Life Technologies). The Hi-C libraries were subjected to quality control and sequenced on an Illumina HiSeq X Ten sequencer.

Hi-C analysis

The paired-end Hi-C clean data were mapped to our assembly with HiCUP (Wingett et al., 2015) using the following option: –format sanger –longest 800 –shortest 150 –nofill N. Filtered data from HiCUP were used as input for ALLHiC (Zhang et al., 2019). ALLHIC was used to reconstruct each chromosome via five steps with default options: prune, partition, rescue, optimize, and build. Genetic LGs from *D. latiflorus* (Guo et al., 2019b) were aligned to the final assembly using minimap2 (Li, 2018). Synteny was inferred using JCVI utility libraries with the default option (Tang et al., 2008).

Evaluation of haplotype accuracy

To examine the accuracy between haplotypes, the haplotype-resolved genome assembly was divided into haplotypes A and B **(**Figure S21**)**. We generated consistently phased SNPs in PacBio long reads and 10X Genomics Linked reads for haplotype A from the phased genome as a reference. First, we mapped Illumina reads to the haplotype A genome using BWA (Li and Durbin, 2009) and performed SNP calling using the Genome Analysis Tool Kit (GATK) pipeline to generate Variant Call Format (VCF) files. For PacBio long reads, we mapped PacBio Iso-Seq reads to the haplotype A genome using minimap2 (Li, 2018) with the option -t 20 -ax map-pb –secondary = no to generate a BAM alignment. We then combined the VCF and BAM alignments to identify SNP phasing for PacBio long reads using the whatshap program (Martin et al., 2016): whatshap phase –ignore-read-groups. For 10X Genomics linked reads, we processed the raw 10X Genomic reads using proc10xGenomics.py and mapped the processed reads to the haplotype A genome using BWA (Li and Durbin, 2009): bwa mem -t 40 -p -C to generate a BAM alignment for SNP phasing using the whatshap program. We then compared haplotype B to haplotype A using nucmer (nucmer –mum -l 1000 -c 200 -g 200 -t 12) to generate a delta file, which was used to obtain SNP information for haplotype chromosome groups with show-snps. Finally, we calculated potential switch error in the phased genome assembly based on SNP phasing information using pipeline for switch error estimation (Zhang et al., 2021). For comparison with other haplotype-resolved genomes, we also calculated switch error for published potato (Zhou et al., 2020) and apple (Sun et al., 2020) using the same method.

Haplotype-to-haplotype comparison of collinearity

Synteny and collinearity of gene pairs between haplotypes were identified with JCVI using the following option: python3 -m jcvi.compara.catalog ortholog haplotype A haplotype B –cscore = .99 –no_strip_names. The chromosome comparison between haplotypes was generated using GSAlign (Lin and Hsu, 2020) with default options, which used the Burrows-Wheeler transform. We identified syntenic gene pairs between haplotypes with JCVI using cutoff: –cscore = .99 as described previously (Wang et al., 2021b) to generate 1 versus 1 pairs between haplotypes. We further selected gene pairs located in alignment blocks between haplotypes using GSAlign (Lin and Hsu, 2020).

Identification of sequence variations between haplotypes

We aligned haplotype A to haplotype B using nucmer (Marçais et al., 2018) with the option –mum –maxgap = 500 –mincluster = 100 to generate delta alignment between haplotypes. SNVs between haplotypes were identified with show-snps using the option -C -H -I -T -r -l, based on the delta encoded alignment file. To identify InDels (insertions/deletions), we used GSAlign (Lin and Hsu, 2020) to identify InDels between haplotypes.

Phylogenetic analysis of *D. latiflorus* and homologous species

Single-copy and multiple-copy orthologs were identified from 15 sequenced genomes by cluster analysis of gene families using OthoMCL (Li et al., 2003). Single-copy gene families were clustered with MUSCLE (Edgar, 2004) to construct a super alignment matrix, which was used to construct a phylogenetic tree with RAxML (Yang, 2007) based on the maximum likelihood method with 100 replications. The divergence time was estimated using MCMCTREE (v4.4) from the PAML package (He et al., 2013) to perform Bayesian divergence time estimation with the following options: burn-in = 10,000, sample-number = 100,000, sample-frequency = 2. The expansion and contraction of *D. latiflorus* gene families were identified using CAFE (De Bie et al., 2006).

RNA extraction and library construction for RNA-Seq

Tissues including shoots, abortive shoots, shoot shells, leaves, and roots were collected from *D. latiflorus* seedlings at the Botanical Garden of Bamboo at Fujian Agriculture and Forestry University, China (E119°140; N26°050) in June 2019. Seeds, stems, internodes, and buds were also collected from seedlings. The samples were snap frozen in liquid nitrogen for total RNA extraction with an RNAprep Pure Plant Kit (Polysaccharides & Polyphenolics-rich). The integrity of the RNA samples was evaluated using an Agilent 2100 Bioanalyzer, and samples with RNA integrity number values >8 were used for downstream RNA-Seq library construction using the deoxyuridine triphosphate method as described previously (Wang et al., 2017) on the Illumina NovaSeq platform as 150-nt paired-end reads.

Bioinformatics analysis of RNA-Seq reads

RNA-Seq reads were mapped to the assembled *D. latiflorus* genome using TopHat (v2.0.11) with an anchor length >8-nt and further assembled using Cufflinks v2.1.1 (Trapnell et al., 2012) with the following option: -F 0.05 -A 0.01 -I 100000 –min-intron-length 30. The assembled transcripts were combined using the Cuffmerge utility with the following option: –min-isoform-fraction 0.01. The expression levels were measured and normalized as fragments per kilobase of transcript per million mapped reads (FPKM) (Mortazavi et al., 2008; Trapnell et al., 2012). A fold change of FPKM > 3 and FDR < 0.01 were considered to be the thresholds for the identification of DEGs. *P*-values were calculated using the statistical package edgeR (Robinson et al., 2010).

SMRT Iso-Seq using the PacBio Sequel platform

We pooled the above samples as a single library for PacBio full-length isoform sequencing and constructed Iso-Seq libraries as described previously (Wang et al., 2017). The full-length transcriptome sequencing experiment included three steps: sample detection, library construction, and sequencing. In brief, the Iso-Seq libraries were constructed using 1 μg of equally mixed RNAs. The integrity and concentration of total RNAs were detected using the Agilent 2100 system. The poly (A) RNAs were enriched using magnetic beads with Oligo (dT). The RNA was transcribed into cDNA using a SMARTer™ PCR cDNA Synthesis Kit. Full-length cDNA was identified using BluePippin and used to construct libraries with inserts of different lengths. Following end repair, the full-length cDNAs were connected to SMRT dumbbell-shaped adapters to obtain SMRT bell libraries. Finally, 3–6 and 5–10 kb SMRT bell PacBio transcriptome Iso-Seq libraries were further screened using BluePippin. The libraries were quantified using Qubit 2.0. The insert size of the library was detected with an Agilent 2100 Bioanalyzer. Following calculations using PacBio Calculator, the sequencing primers and sequencing enzymes were mixed in equal proportions with the SMRT template. Sequencing was performed with the PacBio Sequel sequencing system using SMRT Cell as the sequencing carrier in CLR sequencing mode.

For bioinformatics analysis, the workflow in SMRT Link was used to generate highly accurate consensus sequences with default options. The raw reads were cleaned by poly(A) tail removal and concatemer detection using the refine tool from IsoSeq v3. PacBio BAM files were converted into fasta files using bam2fasta from BAM2fastx tools. LoRDEC was used for Iso-Seq long reads correction using the following option: lordec-correct -T 40 -k 19 -s 3 (Salmela and Rivals, 2014).

SNP-based phasing of Iso-Seq and RNA-Seq data

PacBio single molecule long reads were mapped to the genome with the "-ax splice -uf –secondary = no -C5 -O6,24 -B4" option using minimap2 (2.15-r905) (Li, 2018). Illumina short RNA-Seq reads were mapped to the genome with the "–dta -k 1" option using HISAT2 (Version:2.1.0) (Kim et al., 2019). SNPs among allelic chromosomes were identified using a distinct utility program from MUMmer (4.0.0beta2) (Delcher et al., 2003), which included nucmer (–mum –maxgap = 500 –mincluster = 100), delta-filer (-1 -q -r), and show-snps (-C -H -I -T -r -l). Following this three-step process, polymorphisms were created between allelic chromosomes. The phasing of Iso-Seq and RNA-Seq data was extracted using PP2PG (phasing pipeline based on two parental genomes) (Feng et al., 2021) based on the SNP calling and mapping results. After phasing of the RNA-Seq data, the aligned short reads were loaded into rmats-turbo-4.1.0 (Shen et al., 2014) using default parameters to identify differential AS events. In total, AS events with FDR < 0.05 were regarded as differential AS events. Coverage tracks (bigWig) were generated using bamCoverage from deeptools (Ramírez et al., 2016). Finally, we used PRAPI (Gao et al., 2018) with default options to identify and analyze splicing variants based on aligned long reads after phasing of the Iso-Seq data. GO enrichment analysis of DEG and differential AS events were carried out using the clusterProfiler package (Yu et al., 2012). GO terms with adjusted $P < 0.05$ were considered to be significantly enriched.

Availability of supporting data

Raw data from this study have been deposited in the NCBI Sequence Read Archive (SRA) under BioProject PRJNA600661. The assembly genome is available at NCBI under accession number JACBGG000000000 and forestry.fafu.edu.cn/pub/Dla. Transcriptome sequencing data were deposited under accession number GSE155494.

ACKNOWLEDGEMENTS

We thank Prof. Rosa Lozano-Durán (Shanghai Center for Plant Stress Biology, Chinese Academy of Sciences) for helpful suggestions about the manuscript. We thank Prof Xingtan Zhang, Huiyuan Wang, and Yongkang Yang from Fujian Agriculture and Forestry University for assisting with analysis of switch error, rMATs, and JCVI, respectively. We thank Prof. Ling-Ling Chen, Huan Li, and Jia-Wu Feng (Huazhong Agricultural University) for their helpful suggestions about phasing analysis. This work was

supported by the National Key Research and Development Program of China (2018YFD0600104), the National Natural Science Foundation of China Grant (31971734), the Natural Science Foundation of Fujian Province (Grant No. 2021J02027), the Distinguished Young Scholar Program of Fujian Agriculture and Forestry University (Grant No. xjq202017), the Technological Innovation Team at the University of Fujian province, and the Forestry Peak Discipline Construction Project from Fujian Agriculture and Forestry University.

CONFLICTS OF INTEREST

The authors declare they have no conflicts of interest associated with this work.

AUTHOR CONTRIBUTIONS

Y.Z., Q.Z., L.G. conceived and designed the study; D.Y., J.R., L.C., T.H., L.C., J.Y., L.F. performed the experiments; L.G., Y. G., D.Y., H.Z. analyzed the high-throughput sequencing data; L.G., D.Y., Y.Z. wrote the manuscript. All authors have read and approved the final version of the manuscript.

Edited by: Xuehui Huang, Shanghai Normal University, China

Received Nov. 22, 2021; **Accepted** Jan. 4, 2022; **Published** Jan. 6, 2022

OO: OnlineOpen

REFERENCES

Adey, A., Kitzman, J.O., Burton, J.N., Daza, R., Kumar, A., Christiansen, L., Ronaghi, M., Amini, S., Gunderson, K.L., Steemers, F.J., and Shendure, J. (2014). *In vitro*, long-range sequence information for de novo genome assembly via transposase contiguity. Genome Res. **24:** 2041–2049.

Bairoch, A., and Apweiler, R. (2000). The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000. Nucleic Acids Res. **28:** 45–48.

Benson, G. (1999). Tandem repeats finder: A program to analyze DNA sequences. Nucleic Acids Res. **27:** 573–580.

Birney, E., Clamp, M., and Durbin, R. (2004). GeneWise and genomewise. Genome Res. **14:** 988–995.

Boetzer, M., Henkel, C.V., Jansen, H.J., Butler, D., and Pirovano, W. (2011). Scaffolding pre-assembled contigs using SSPACE. Bioinformatics **27:** 578–579.

Chalhoub, B., Denoeud, F., Liu, S., Parkin, I.A., Tang, H., Wang, X., Chiquet, J., Belcram, H., Tong, C., Samans, B., Correa, M., Da Silva, C., Just, J., Falentin, C., Koh, C.S., Le Clainche, I., Bernard, M., Bento, P., Noel, B., Labadie, K., Alberti, A., Charles, M., Arnaud, D., Guo, H., Daviaud, C., Alamery, S., Jabbari, K., Zhao, M., Edger, P.P., Chelaifa, H., Tack, D., Lassalle, G., Mestiri, I., Schnel, N., Le Paslier, M.C., Fan, G., Renault, V., Bayer, P.E., Golicz, A.A., Manoli, S., Lee, T.H., Thi, V.H., Chalabi, S., Hu, Q., Fan, C., Tollenaere, R., Lu, Y., Battail, C., Shen, J., Sidebottom, C.H., Wang, X., Canaguier, A., Chauveau, A., Berard, A., Deniot, G., Guan, M., Liu, Z., Sun, F., Lim, Y.P., Lyons, E., Town, C.D., Bancroft, I., Wang, X., Meng, J., Ma, J., Pires, J.C., King, G.J., Brunel, D., Delourme, R., Renard, M., Aury, J.M., Adams, K.L., Batley, J., Snowdon, R.J., Tost, J., Edwards, D., Zhou, Y., Hua, W., Sharpe, A.G., Paterson, A.H., Guan, C., Wincker, P. (2014). Plant genetics. Early allopolyploid evolution in the post-Neolithic *Brassica napus* oilseed genome. Science. **345:** 950–953.

Chen, H., Zeng, Y., Yang, Y., Huang, L., Tang, B., Zhang, H., Hao, F., Liu, W., Li, Y., Liu, Y., Zhang, X., Zhang, R., Zhang, Y., Li, Y., Wang, K., He, H., Wang, Z., Fan, G., Yang, H., Bao, A., Shang, Z., Chen, J., Wang, W., and Qiu, Q. (2020). Allele-aware chromosome-level genome assembly and efficient transgene-free genome editing for the autotetraploid cultivated alfalfa. Nat. Commun. **11:** 2494.

Chen, N. (2004). Using RepeatMasker to identify repetitive elements in genomic sequences. Curr. Protoc. Bioinformatics **5:** 4–10.

Chen, X., Lu, Q., Liu, H., Zhang, J., Hong, Y., Lan, H., Li, H., Wang, J., Liu, H., Li, S., Pandey, M.K., Zhang, Z., Zhou, G., Yu, J., Zhang, G., Yuan, J., Li, X., Wen, S., Meng, F., Yu, S., Wang, X., Siddique, K.H. M., Liu, Z.J., Paterson, A.H., Varshney, R.K., and Liang, X. (2019). Sequencing of cultivated peanut, *Arachis hypogaea*, yields insights into genome evolution and oil improvement. Mol. Plant **12:** 920–934.

Chin, C.S., Alexander, D.H., Marks, P., Klammer, A.A., Drake, J., Heiner, C., Clum, A., Copeland, A., Huddleston, J., Eichler, E.E., Turner, S.W., and Korlach, J. (2013). Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. Nat. Methods **10:** 563–569.

Chin, C.S., Peluso, P., Sedlazeck, F.J., Nattestad, M., Concepcion, G. T., Clum, A., Dunn, C., O'Malley, R., Figueroa-Balderas, R., Morales-Cruz, A., Cramer, G.R., Delledonne, M., Luo, C., Ecker, J. R., Cantu, D., Rank, D.R., and Schatz, M.C. (2016). Phased diploid genome assembly with single-molecule real-time sequencing. Nat. Methods **13:** 1050–1054.

Daccord, N., Celton, J.M., Linsmith, G., Becker, C., Choisne, N., Schijlen, E., van de Geest, H., Bianco, L., Micheletti, D., Velasco, R., Di Pierro, E.A., Gouzy, J., Rees, D.J.G., Guerif, P., Muranty, H., Durel, C.E., Laurens, F., Lespinasse, Y., Gaillard, S., Aubourg, S., Quesneville, H., Weigel, D., van de Weg, E., Troggio, M., and Bucher, E. (2017). High-quality de novo assembly of the apple genome and methylome dynamics of early fruit development. Nat. Genet. **49:** 1099–1106.

De Bie, T., Cristianini, N., Demuth, J.P., and Hahn, M.W. (2006). CAFE: A computational tool for the study of gene family evolution. Bioinformatics **22:** 1269–1271.

Delcher, A.L., Salzberg, S.L., and Phillippy, A.M. (2003). Using MUMmer to identify similar regions in large sequence sets. Curr. Protoc. Bioinformatics **1:** 10.3.

Eckardt, N.A. (2000). Sequencing the rice genome. Plant Cell **12:** 2011–2017.

Edgar, R.C. (2004). MUSCLE: Multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. **32:** 1792–1797.

Feng, J.-W., Lu, Y., Shao, L., Zhang, J., Li, H., and Chen, L.-L. (2021). Phasing analysis of the transcriptome and epigenome in a rice hybrid reveals the inheritance and difference in DNA methylation and allelic transcription regulation. Plant Commun. **2:** 100185.

Gao, Y., Wang, H., Zhang, H., Wang, Y., Chen, J., and Gu, L. (2018). PRAPI: Post-transcriptional regulation analysis pipeline for Iso-Seq. Bioinformatics **34:** 1580–1582.

Ge, W., Zhang, Y., Cheng, Z., Hou, D., Li, X., and Gao, J. (2017). Main regulatory pathways, key genes and microRNAs involved in flower formation and development of moso bamboo (*Phyllostachys edulis*). Plant Biotechnol. J. **15:** 82–96.

Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., and Zeng, Q. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat. Biotechnol. **29:** 644–652.

Guigo, R., Knudsen, S., Drake, N., and Smith, T. (1992). Prediction of gene structure. J. Mol. Biol. **226:** 141–157.

Guo, L., Sun, X., Li, Z., Wang, Y., Fei, Z., Jiao, C., Feng, J., Cui, D., Feng, X., Ding, Y., Zhang, C., and Wei, Q. (2019a). Morphological

dissection and cellular and transcriptome characterizations of bamboo pith cavity formation reveal a pivotal role of genes related to programmed cell death. Plant Biotechnol. J. **17:** 982–997.

Guo, Z.H., Ma, P.F., Yang, G.Q., Hu, J.Y., Liu, Y.L., Xia, E.H., Zhong, M. C., Zhao, L., Sun, G.L., Xu, Y.X., Zhao, Y.J., Zhang, Y.C., Zhang, Y. X., Zhang, X.M., Zhou, M.Y., Guo, Y., Guo, C., Liu, J.X., Ye, X.Y., Chen, Y.M., Yang, Y., Han, B., Lin, C.S., Lu, Y., and Li, D.Z. (2019b). Genome sequences provide insights into the reticulate origin and unique traits of woody bamboos. Mol. Plant **12:** 1353–1365.

Haas, B.J., Salzberg, S.L., Zhu, W., Pertea, M., Allen, J.E., Orvis, J., White, O., Buell, C.R., and Wortman, J.R. (2008). Automated eukaryotic gene structure annotation using EVidenceModeler and the Program to Assemble Spliced Alignments. Genome Biol. **9:** R7.

He, N., Zhang, C., Qi, X., Zhao, S., Tao, Y., Yang, G., Lee, T.H., Wang, X., Cai, Q., Li, D., Lu, M., Liao, S., Luo, G., He, R., Tan, X., Xu, Y., Li, T., Zhao, A., Jia, L., Fu, Q., Zeng, Q., Gao, C., Ma, B., Liang, J., Wang, X., Shang, J., Song, P., Wu, H., Fan, L., Wang, Q., Shuai, Q., Zhu, J., Wei, C., Zhu-Salzman, K., Jin, D., Wang, J., Liu, T., Yu, M., Tang, C., Wang, Z., Dai, F., Chen, J., Liu, Y., Zhao, S., Lin, T., Zhang, S., Wang, J., Wang, J., Yang, H., Yang, G., Wang, J., Paterson, A.H., Xia, Q., Ji, D., and Xiang, Z. (2013). Draft genome sequence of the mulberry tree Morus notabilis. Nat. Commun. **4:** 2445.

International Wheat Genome Sequencing, C. (2014). A chromosome-based draft sequence of the hexaploid bread wheat (*Triticum aestivum*) genome. Science **345:** 1251788.

Jurka, J., Kapitonov, V.V., Pavlicek, A., Klonowski, P., Kohany, O., and Walichiewicz, J. (2005). Repbase Update, a database of eukaryotic repetitive elements. Cytogenet. Genome. Res. **110:** 462–467.

Kim, D., Paggi, J.M., Park, C., Bennett, C., and Salzberg, S.L. (2019). Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. Nat. Biotechnol. **37:** 907–915.

Kyriakidou, M., Tai, H.H., Anglin, N.L., Ellis, D., and Stromvik, M.V. (2018). Current strategies of polyploid plant genome sequence assembly. Front. Plant Sci. **9:** 1660.

Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat. Methods **9:** 357–359.

Li, F., Fan, G., Lu, C., Xiao, G., Zou, C., Kohel, R.J., Ma, Z., Shang, H., Ma, X., Wu, J., Liang, X., Huang, G., Percy, R.G., Liu, K., Yang, W., Chen, W., Du, X., Shi, C., Yuan, Y., Ye, W., Liu, X., Zhang, X., Liu, W., Wei, H., Wei, S., Huang, G., Zhang, X., Zhu, S., Zhang, H., Sun, F., Wang, X., Liang, J., Wang, J., He, Q., Huang, L., Wang, J., Cui, J., Song, G., Wang, K., Xu, X., Yu, J.Z., Zhu, Y., and Yu, S. (2015a). Genome sequence of cultivated Upland cotton (*Gossypium hirsutum* TM-1) provides insights into genome evolution. Nat. Biotechnol. **33:** 524–530.

Li, H. (2018). Minimap2: Pairwise alignment for nucleotide sequences. Bioinformatics **34:** 3094–3100.

Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics **25:** 1754–1760.

Li, L., Cheng, Z., Ma, Y., Bai, Q., Li, X., Cao, Z., Wu, Z., and Gao, J. (2018). The association of hormone signalling genes, transcription and changes in shoot anatomy during moso bamboo growth. Plant Biotechnol. J. **16:** 72–85.

Li, L., Stoeckert, C.J., and Roos, D.S. (2003). OrthoMCL: Identification of ortholog groups for eukaryotic genomes. Genome Res. **13:** 2178–2189.

Li, P., Zhou, G., Du, H., Lu, D., Mo, L., Xu, X., Shi, Y., and Zhou, Y. (2015b). Current and potential carbon stocks in Moso bamboo forests in China. J. Environ. Manage. **156:** 89–96.

Li, R., Zhu, H., Ruan, J., Qian, W., Fang, X., Shi, Z., Li, Y., Li, S., Shan, G., and Kristiansen, K. (2010). *De novo* assembly of human genomes with massively parallel short read sequencing. Genome Res. **20:** 265–272.

Lieberman-Aiden, E., van Berkum, N.L., Williams, L., Imakaev, M., Ragoczy, T., Telling, A., Amit, I., Lajoie, B.R., Sabo, P.J., Dorschner, M.O., Sandstrom, R., Bernstein, B., Bender, M.A., Groudine, M., Gnirke, A., Stamatoyannopoulos, J., Mirny, L.A., Lander, E.S., and Dekker, J. (2009). Comprehensive mapping of long-range interactions reveals folding principles of the human genome. Science **326:** 289–293.

Lin, H.N., and Hsu, W.L. (2020). GSAlign: An efficient sequence alignment tool for intra-species genomes. BMC Genomics **21:** 1–10.

Ling, H.-Q., Zhao, S., Liu, D., Wang, J., Sun, H., Zhang, C., Fan, H., Li, D., Dong, L., and Tao, Y. (2013). Draft genome of the wheat A-genome progenitor Triticum urartu. Nature **496:** 87–90.

Liu, M., Qiao, G., Jiang, J., Yang, H., Xie, L., Xie, J., and Zhuo, R. (2012). Transcriptome sequencing and de novo analysis for Ma bamboo (*Dendrocalamus latiflorus* Munro) using the Illumina platform. PLoS One **7:** e46766.

Liu, X., Mei, W., Soltis, P.S., Soltis, D.E., and Barbazuk, W.B. (2017). Detecting alternatively spliced transcript isoforms from single-molecule long-read sequences without a reference genome. Mol. Ecol. Resour. **17:** 1243–1256.

Lobovikov, M., Ball, L., Paudel, S., Guardia, M., Piazza, M., Wu, J., Ren, H., and Russo, L. (2007). *World Bamboo Resources: A Thematic Study Prepared in the Framework of the Global Forest Resources Assessment 2005*. Food and Agriculture Org. **18:** 1–71.

Loureiro, J., Rodriguez, E., Dolezel, J., and Santos, C. (2007). Two new nuclear isolation buffers for plant DNA flow cytometry: A test with 37 species. Ann. Bot. **100:** 875–888.

Lowe, T.M., and Eddy, S.R. (1997). tRNAscan-SE: A program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res. **25:** 955–964.

Ma, T., Wang, J., Zhou, G., Yue, Z., Hu, Q., Chen, Y., Liu, B., Qiu, Q., Wang, Z., Zhang, J., Wang, K., Jiang, D., Gou, C., Yu, L., Zhan, D., Zhou, R., Luo, W., Ma, H., Yang, Y., Pan, S., Fang, D., Luo, Y., Wang, X., Wang, G., Wang, J., Wang, Q., Lu, X., Chen, Z., Liu, J., Lu, Y., Yin, Y., Yang, H., Abbott, R.J., Wu, Y., Wan, D., Li, J., Yin, T., Lascoux, M., Difazio, S.P., Tuskan, G.A., Wang, J., and Liu, J. (2013). Genomic insights into salt adaptation in a desert poplar. Nat. Commun. **4:** 2797.

Marçais, G., Delcher, A.L., Phillippy, A.M., Coston, R., Salzberg, S.L., and Zimin, A. (2018). MUMmer4: A fast and versatile genome alignment system. PLoS Comp. Biol. **14:** e1005944.

Martin, M., Patterson, M., Garg, S., Fischer, S., Pisanti, N., Klau, G.W., Schöenhuth, A., and Marschall, T. (2016). WhatsHap: Fast and accurate read-based phasing. BioRxiv 085050.

Ming, R., and Wai, C.M. (2015). Assembling allopolyploid genomes: No longer formidable. Genome Biol. **16:** 27.

Mortazavi, A., Williams, B.A., McCue, K., Schaeffer, L., and Wold, B. (2008). Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nat. Methods **5:** 621–628.

Mulder, N.J., and Apweiler, R. (2008). The InterPro database and tools for protein domain analysis. Curr. Protoc. Bioinformatics **2:** 2–7.

Nagle, M., Déjardin, A., Pilate, G., and Strauss, S.H. (2018). Opportunities for innovation in genetic transformation of forest trees. Front. Plant Sci. **9:** 1443.

Nystedt, B., Street, N.R., Wetterbom, A., Zuccolo, A., Lin, Y.C., Scofield, D.G., Vezzi, F., Delhomme, N., Giacomello, S., Alexeyenko, A., Vicedomini, R., Sahlin, K., Sherwood, E., Elfstrand, M., Gramzow, L., Holmberg, K., Hallman, J., Keech, O., Klasson, L., Koriabine, M., Kucukoglu, M., Kaller, M., Luthman, J., Lysholm, F., Niittyla, T., Olson, A., Rilakovic, N., Ritland, C., Rossello, J.A., Sena, J., Svensson, T., Talavera-Lopez, C., Theissen, G., Tuominen, H., Vanneste, K., Wu, Z.Q., Zhang, B., Zerbe, P., Arvestad, L., Bhalerao, R., Bohlmann, J., Bousquet, J., Garcia Gil, R., Hvidsten, T.R., de Jong, P., MacKay, J., Morgante, M., Ritland, K., Sundberg, B., Thompson, S.L., Van de Peer, Y., Andersson, B., Nilsson, O., Ingvarsson, P.K., Lundeberg, J., Jansson, S. (2013). The Norway spruce genome sequence and conifer genome evolution. Nature. **497:** 579–584.

Peng, Z., Lu, Y., Li, L., Zhao, Q., Feng, Q., Gao, Z., Lu, H., Hu, T., Yao, N., and Liu, K. (2013). The draft genome of the fast-growing non-timber

forest species moso bamboo (*Phyllostachys heterocycla*). Nat. Genet. **45:** 456–461.

Pertea, M., Pertea, G.M., Antonescu, C.M., Chang, T.C., Mendell, J.T., and Salzberg, S.L. (2015). StringTie enables improved reconstruction of a transcriptome from RNA-Seq reads. Nat. Biotechnol. **33:** 290–295.

Price, A.L., Jones, N.C., and Pevzner, P.A. (2005). *De novo* identification of repeat families in large genomes. Bioinformatics **21:** i351–i358.

Qiao, G., Li, H., Liu, M., Jiang, J., Yin, Y., Zhang, L., and Zhuo, R. (2013). Callus induction and plant regeneration from anthers of *Dendrocalamus latiflorus* Munro. In Vitro Cell. Dev. Biol. Plant **49:** 375–382.

Qiao, G., Liu, M., Song, K., Li, H., Yang, H., Yin, Y., and Zhuo, R. (2017). Phenotypic and comparative transcriptome analysis of different ploidy plants in *Dendrocalamus latiflorus* Munro. Front. Plant Sci. **8:** 1371.

Qiao, G., Yang, H., Zhang, L., Han, X., Liu, M., Jiang, J., Jiang, Y., and Zhuo, R. (2014). Enhanced cold stress tolerance of transgenic *Dendrocalamus latiflorus* Munro (Ma bamboo) plants expressing a bacterial CodA gene. In Vitro Cell Dev. Biol. Plant **50:** 385–391.

Ramírez, F., Ryan, D.P., Grüning, B., Bhardwaj, V., Kilpert, F., Richter, A.S., Heyne, S., Dündar, F., and Manke, T. (2016). deepTools2: A next generation web server for deep-sequencing data analysis. Nucleic Acids Res. **44:** W160–W165.

Reddy, A.S. (2007). Alternative splicing of pre-messenger RNAs in plants in the genomic era. Annu. Rev. Plant Biol. **58:** 267–294.

Robinson, M.D., McCarthy, D.J., and Smyth, G.K. (2010). edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics **26:** 139–140.

Salmela, L., and Rivals, E. (2014). LoRDEC: Accurate and efficient long read error correction. Bioinformatics **30:** 3506–3514.

Shen, S., Park, J.W., Lu, Z.X., Lin, L., Henry, M.D., Wu, Y.N., Zhou, Q., and Xing, Y. (2014). rMATS: Robust and flexible detection of differential alternative splicing from replicate RNA-Seq data. Proc. Natl. Acad. Sci. U.S.A. **111:** E5593–E5601.

Song, X.Z., Peng, C.H., Zhou, G.M., Jiang, H., Wang, W.F., and Xiang, W.H. (2013). Climate warming-induced upward shift of Moso bamboo population on Tianmu Mountain, China. J. Mountain Sci. **10:** 363–369.

Stanke, M., and Morgenstern, B. (2005). AUGUSTUS: A web server for gene prediction in eukaryotes that allows user-defined constraints. Nucleic Acids Res. **33:** W465–W467.

Sun, X., Jiao, C., Schwaninger, H., Chao, C.T., Ma, Y., Duan, N., Khan, A., Ban, S., Xu, K., and Cheng, L. (2020). Phased diploid genome assemblies and pan-genomes provide insights into the genetic history of apple domestication. Nat. Genet. **52:** 1423–1432.

Tang, H., Bowers, J.E., Wang, X., Ming, R., Alam, M., and Paterson, A.H. (2008). Synteny and collinearity in plant genomes. Science **320:** 486–488.

Tempel, S. (2012). Using and understanding RepeatMasker. Methods Mol. Biol. **859:** 29–51.

Trapnell, C., Roberts, A., Goff, L., Pertea, G., Kim, D., Kelley, D.R., Pimentel, H., Salzberg, S.L., Rinn, J.L., and Pachter, L. (2012). Differential gene and transcript expression analysis of RNA-Seq experiments with TopHat and Cufflinks. Nat. Protoc. **7:** 562–578.

Tyler, A.D., Mataseje, L., Urfano, C.J., Schmidt, L., Antonation, K.S., Mulvey, M.R., and Corbett, C.R. (2018). Evaluation of Oxford Nanopore's MinION sequencing device for microbial whole genome sequencing applications. Sci. Rep. **8:** 10931.

VanBuren, R., Bryant, D., Edger, P.P., Tang, H., Burgess, D., Challabathula, D., Spittle, K., Hall, R., Gu, J., and Lyons, E. (2015). Single-molecule sequencing of the desiccation-tolerant grass *Oropetium thomaeum*. Nature **527:** 508.

Vurture, G.W., Sedlazeck, F.J., Nattestad, M., Underwood, C.J., Fang, H., Gurtowski, J., and Schatz, M.C. (2017). GenomeScope: Fast reference-free genome profiling from short reads. Bioinformatics **33:** 2202–2204.

Walker, B.J., Abeel, T., Shea, T., Priest, M., Abouelliel, A., Sakthikumar, S., Cuomo, C.A., Zeng, Q., Wortman, J., Young, S.K., and Earl, A.M. (2014). Pilon: An integrated tool for comprehensive microbial variant detection and genome assembly improvement. PLoS One **9:** e112963.

Wang, M., Li, Z., ZhangYe., Zhang, Y., Xie, Y., Ye, L., Zhuang, Y., Lin, K., Zhao, F., and Guo, J. (2021a). An atlas of wheat epigenetic regulatory elements reveals subgenome divergence in the regulation of development and stress responses. Plant Cell **33:** 865–881.

Wang, P., Yu, J., Jin, S., Chen, S., Yue, C., Wang, W., Gao, S., Cao, H., Zheng, Y., and Gu, M. (2021b). Genetic basis of high aroma and stress tolerance in the oolong tea cultivar genome. Hortic. Res. **8:** 1–15.

Wang, T., Wang, H., Cai, D., Gao, Y., Zhang, H., Wang, Y., Lin, C., Ma, L., and Gu, L. (2017). Comprehensive profiling of rhizome-associated alternative splicing and alternative polyadenylation in moso bamboo (*Phyllostachys edulis*). Plant J. **91:** 684–699.

Wei, Q., Guo, L., Jiao, C., Fei, Z., Chen, M., Cao, J., Ding, Y., and Yuan, Q. (2019). Characterization of the developmental dynamics of the elongation of a bamboo internode during the fast growth stage. Tree Physiol. **39:** 1201–1214.

Wei, Q., Jiao, C., Ding, Y., Gao, S., Guo, L., Chen, M., Hu, P., Xia, S., Ren, G., and Fei, Z. (2018). Cellular and molecular characterizations of a slow-growth variant provide insights into the fast growth of bamboo. Tree Physiol. **38:** 641–654.

Wei, Q., Jiao, C., Guo, L., Ding, Y., Cao, J., Feng, J., Dong, X., Mao, L., Sun, H., Yu, F., Yang, G., Shi, P., Ren, G., and Fei, Z. (2017). Exploring key cellular processes and candidate genes regulating the primary thickening growth of Moso underground shoots. New Phytol. **214:** 81–96.

Wingett, S., Ewels, P., Furlan-Magaril, M., Nagano, T., Schoenfelder, S., Fraser, P., and Andrews, S. (2015). HiCUP: Pipeline for mapping and processing Hi-C data. F1000Res. **4:** 1310.

Wu, J., Tang, B., and Tang, Y. (2020). Allele-specific genome targeting in the development of precision medicine. Theranostics **10:** 3118.

Xu, Y., Wong, M., Yang, J., Ye, Z., Jiang, P., and Zheng, S. (2011). Dynamics of carbon accumulation during the fast growth period of bamboo plant. Bot. Rev. **77:** 287–295.

Xu, Z., and Wang, H. (2007). LTR_FINDER: An efficient tool for the prediction of full-length LTR retrotransposons. Nucleic Acids Res. **35:** W265–W268.

Yang, J., Liu, D., Wang, X., Ji, C., Cheng, F., Liu, B., Hu, Z., Chen, S., Pental, D., Ju, Y., Yao, P., Li, X., Xie, K., Zhang, J., Wang, J., Liu, F., Ma, W., Shopan, J., Zheng, H., Mackenzie, S.A., and Zhang, M. (2016). The genome sequence of allopolyploid *Brassica juncea* and analysis of differential homoeolog gene expression influencing selection. Nat. Genet. **48:** 1225–1232.

Yang, Z. (2007). PAML 4: Phylogenetic analysis by maximum likelihood. Mol. Biol. Evol. **24:** 1586–1591.

Ye, G., Zhang, H., Chen, B., Nie, S., Liu, H., Gao, W., Wang, H., Gao, Y., and Gu, L. (2019). De novo genome assembly of the stress tolerant forest species *Casuarina equisetifolia* provides insight into secondary growth. Plant J. **97:** 779–794.

Ye, S., Cai, C., Ren, H., Wang, W., Xiang, M., Tang, X., Zhu, C., Yin, T., Zhang, L., and Zhu, Q. (2017). An efficient plant regeneration and transformation system of Ma bamboo (*Dendrocalamus latiflorus* Munro) started from young shoot as explant. Front. Plant Sci. **8:** 1298.

Ye, S., Chen, G., Kohnen, M.V., Wang, W., Cai, C., Ding, W., Wu, C., Gu, L., Zheng, Y., Ma, X., Lin, C., and Zhu, Q. (2020). Robust CRISPR/Cas9 mediated genome editing and its application in manipulating plant height in the first generation of hexaploid Ma bamboo (*Dendrocalamus latiflorus* Munro). Plant Biotechnol. J. **18:** 1501–1503.

Yu, G., Wang, L.G., Han, Y., and He, Q.Y. (2012). clusterProfiler: An R package for comparing biological themes among gene clusters. OMICS: J. Integr. Biol. **16:** 284–287.

Zhang, J., Zhang, X., Tang, H., Zhang, Q., Hua, X., Ma, X., Zhu, F., Jones, T., Zhu, X., Bowers, J., Wai, C.M., Zheng, C., Shi, Y., Chen, S., Xu, X., Yue, J., Nelson, D.R., Huang, L., Li, Z., Xu, H., Zhou, D., Wang, Y., Hu,

W., Lin, J., Deng, Y., Pandey, N., Mancini, M., Zerpa, D., Nguyen, J.K., Wang, L., Yu, L., Xin, Y., Ge, L., Arro, J., Han, J.O., Chakrabarty, S., Pushko, M., Zhang, W., Ma, Y., Ma, P., Lv, M., Chen, F., Zheng, G., Xu, J., Yang, Z., Deng, F., Chen, X., Liao, Z., Zhang, X., Lin, Z., Lin, H., Yan, H., Kuang, Z., Zhong, W., Liang, P., Wang, G., Yuan, Y., Shi, J., Hou, J., Lin, J., Jin, J., Cao, P., Shen, Q., Jiang, Q., Zhou, P., Ma, Y., Zhang, X., Xu, R., Liu, J., Zhou, Y., Jia, H., Ma, Q., Qi, R., Zhang, Z., Fang, J., Fang, H., Song, J., Wang, M., Dong, G., Wang, G., Chen, Z., Ma, T., Liu, H., Dhungana, S.R., Huss, S.E., Yang, X., Sharma, A., Trujillo, J.H., Martinez, M.C., Hudson, M., Riascos, J.J., Schuler, M., Chen, L.Q., Braun, D.M., Li, L., Yu, Q., Wang, J., Wang, K., Schatz, M.C., Heckerman, D., Van Sluys, M.A., Souza, G.M., Moore, P.H., Sankoff, D., VanBuren, R., Paterson, A.H., Nagai, C., and Ming, R. (2018). Allele-defined genome of the autopolyploid sugarcane *Saccharum spontaneum* L. Nat. Genet. **50:** 1565–1573.

Zhang, X., Chen, S., Shi, L., Gong, D., Zhang, S., Zhao, Q., Zhan, D., Vasseur, L., Wang, Y., and Yu, J. (2021). Haplotype-resolved genome assembly provides insights into evolutionary history of the tea plant *Camellia sinensis*. Nat. Genet. **53:** 1250–1259.

Zhang, X., Zhang, S., Zhao, Q., Ming, R., and Tang, H. (2019). Assembly of allele-aware, chromosomal-scale autopolyploid genomes based on Hi-C data. Nat. Plants **5:** 833–845.

Zhao, H., Gao, Z., Wang, L., Wang, J., Wang, S., Fei, B., Chen, C., Shi, C., Liu, X., Zhang, H., Lou, Y., Chen, L., Sun, H., Zhou, X., Wang, S., Zhang, C., Xu, H., Li, L., Yang, Y., Wei, Y., Yang, W., Gao, Q., Yang, H., Zhao, S., and Jiang, Z. (2018). Chromosome-level reference genome and alternative splicing atlas of moso bamboo (*Phyllostachys edulis*). Gigascience **7:** giy115.

Zhao, H., Zhang, K., Zhou, X., Xi, L., Wang, Y., Xu, H., Pan, T., and Zou, Z. (2017). Melatonin alleviates chilling stress in cucumber seedlings by up-regulation of CsZat12 and modulation of polyamine and abscisic acid metabolism. Sci. Rep. **7:** 4998.

Zhao, L., Zhang, H., Kohnen, M.V., Prasad, K.V., Gu, L., and Reddy, A. S. (2019). Analysis of transcriptome and epitranscriptome in plants using PacBio Iso-Seq and nanopore-based direct RNA sequencing. Front. Genet. **10:** 253.

Zhou, G., Meng, C., Jiang, P., and Xu, Q. (2011). Review of carbon fixation in bamboo forests in China. Bot. Rev. **77:** 262.

Zhou, M., Hu, B., and Zhu, Y. (2017). Genome-wide characterization and evolution analysis of long terminal repeat retroelements in moso bamboo (*Phyllostachys edulis*). Tree Genet. Genom. **13:** 43.

Zhou, Q., Tang, D., Huang, W., Yang, Z., Zhang, Y., Hamilton, J.P., Visser, R.G., Bachem, C.W., Buell, C.R., and Zhang, Z. (2020). Haplotype-resolved genome analyses of a heterozygous diploid potato. Nat. Genet. **52:** 1018–1023.

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article: http://onlinelibrary.wiley.com/doi/10.1111/jipb.13217/suppinfo

Figure S1. Chromosome number of *Dendrocalamus latiflorus*

Figure S2. Flow cytometry (FCM) analysis comparing *Oryza sativa* ssp. *japonica* and *Dendrocalamus latiflorus*

Figure S3. GenomeScope profile for the estimation of genome size and heterozygosity*x*-axis and *y*-axis represent the coverage and frequency of 17 *k*-mers, respectively

Figure S4. GC content and distribution of sequencing depth based on short reads from the Illumina platform

The *x*-axis indicates the GC content; the *y*-axis indicates the sequencing depth. The right panel represents the contig coverage depth distribution; the upper panel represents the GC content distribution. The higher density region of dots in the scatter plot is highlighted in red.

Figure S5. GC content and distribution of sequencing depth based on long reads from the PacBio platform

The main image shows the GC content distribution; the *x*-axis represents the GC content; the *y*-axis represents the sequencing depth. The top histogram represents GC content distribution; the histogram on the right represents sequencing depth distribution.

Figure S6. Haplotype-to-haplotype collinearity comparison

(**A**) Dot plot of chromosomal synteny comparison between two haplotypes. (**B**) Collinearity between two haplotypes. Subgenomes A, B, and C are shown in green, blue, and red, respectively. Macro-syntenic blocks between haplotype chromosomes are shown in gray.

Figure S7. Single nucleotide variations (SNVs) and InDels (insertions/deletions) between the two haplotypes in the three subgenomes

The quantity of SNVs/InDels was generated from pairwise comparison between the two haplotypes (A1 vs. A1, B1 vs. B2, and C1 vs. C2). (**A**) SNVs. (**B**) Deletions. (**C**) Insertions.

Figure S8. Sequence variations between two haplotypes

The *x*-axis indicates the number of sequencing variants in 1 kb windows, and the *y*-axis indicates the chromosome coordinate. Blue, red, and green represent sequence variations in subgenomes **A**, **B**, and **C**, respectively.

Figure S9. Visualization of pairwise synteny

Dot plot showing a 1:6 synteny pattern between *Oryza sativa* and *Dendrocalamus latiflorus*. The 1:6 synteny pattern is also validated by the histogram.

Figure S10. High-density genetic map of bamboo supports the chromosomal assignments

Figure S11. Flowchart for the annotation of protein-coding genes and transposable elements

Figure S12. Comparison of gene structures among closely related species

Figure S13. Distribution intervals for non-coding genes

Figure S14. Graphs showing the transposable element (TE) accumulation history in *Dendrocalamus latiflorus* using Kimura distance-based copy divergence analysis

The *x*-axis represents the Kimura distances to their corresponding consensus sequences; the *y*-axis represents the percentage of the genome occupied by each type of TE, including DNA, long interspersed nuclear element (LINE), long terminal repeats (LTR), and short interspersed nuclear element (SINE).

Figure S15. Quantitative reverse transcription polymerase chain reaction validation of differentially expressed genes

Figure S16. Kyoto Encyclopedia of Genes and Genomes enrichment of differentially expressed genes between non-abortive and abortive shoots

The gene ratio represents the percentage of differentially expressed genes related to KEGG terms/total number of differentially expressed genes.

Figure S17. Quantitative reverse transcription polymerase chain reaction validation of the differential expression of metabolic pathway genes

Figure S18. Allele-specific expressed (ASE) genes

(**A**) MA plot of allele expression. (**B**), (**C**), and (**D**) show the enriched Gene Ontology terms of ASE genes in subgenomes **A**, **B**, and **C**, respectively.

Figure S19. Allele-specific genes showing alternative splicing

(**A**) Bar plot of the percentage of alternative splicing (AS) types from allelic gene pairs. The four colors represent different AS events. (**B**) and (**C**) show the enriched Gene Ontology terms of allele-specific AS genes in the haplotypes A and haplotypes B, respectively.

Figure S20. Micro-synteny analysis of developmental genes related to the transformation efficiency of *P. edulis, Oryza sativa*, and *Dendrocalamus latiflorus*

Figure S21. Flowchart for assessing the accuracy of the haplotype-resolved genome assembly

Table S1. Statistics of genome assembly in *Dendrocalamus latiflorus*

Table S2. Statistics of Contig and Scaffold for genome assembly of *Dendrocalamus latiflorus*

Table S3. Statistics of percentage of different type tags that the high-throughput chromosome conformation capture (Hi-C) generated for chromosome-scale assembly of *Dendrocalamus latiflorus*

Table S4. Overview of the chromosome-level genome assemblies and annotations

Table S5. Assessment of genome assembly and annotation completeness using BUSCO (Benchmarking Universal Single-Copy Orthologs)

Table S6. Assessment of genome assembly and annotation completeness using CEGMA (Core Eukaryotic Genes Mapping Approach)

Table S7. Length, GC content, number of genes for subgenome A

Table S8. Length, GC content, number of genes for subgenome B

Table S9. Length, GC content, number of genes for subgenome C

Table S10. Annotation of the hexaploid genome of *Dendrocalamus latiflorus*

Table S11. Functional assignment of the hexaploid genome of *Dendrocalamus latiflorus*

Table S12. Overview of non-coding RNA (ncRNA) genes in *Dendrocalamus latiflorus*
Table S13. Statistics of transposable element in *Dendrocalamus latiflorus*
Table S14. Quantitative reverse transcription polymerase chain reaction validation of differentially expressed genes
Table S15. Quantitative reverse transcription polymerase chain reaction validation of genes in differential metabolism pathway