

1

2

3

4

5 **Allele-aware chromosome-scale assembly of the**

6 **allopolyploid genome of hexaploid Ma Bamboo**

7 **(*Dendrocalamus latiflorus* Munro)**

8

9

10 Yushan Zheng^{1*}, Deming Yang¹, Jundong Rong¹, Liguang Chen¹, Qiang Zhu²,

11 Tianyou He², Lingyan Chen², Jing Ye¹, Lili Fan¹, Yubang Gao³, Hangxiao

12 Zhang³, Lianfeng Gu^{3*}

13

14

15

16

17 **SUPPORTING INFORMATION**

Figure S1

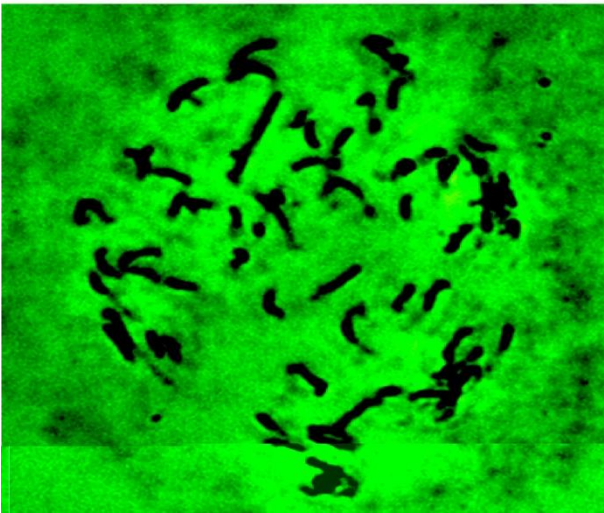


Figure S1. Chromosome number of *D. latiflorus*

Figure S2

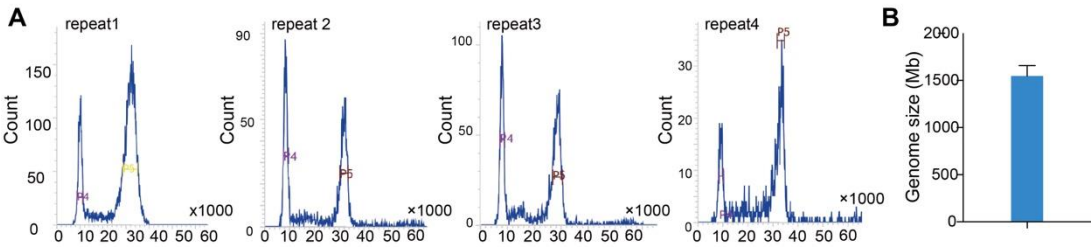


Figure S2. Flow cytometry (FCM) analysis comparing *Oryza sativa* ssp. *japonica* and *D. latiflorus*

Figure S3

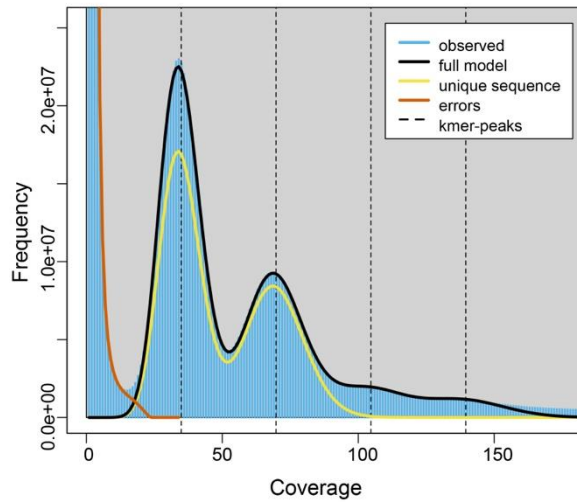


Figure S3. GenomeScope profile for the estimation of genome size and heterozygosity
x-axis and y-axis represent the coverage and frequency of 17 *k*-mers, respectively.

Figure S4

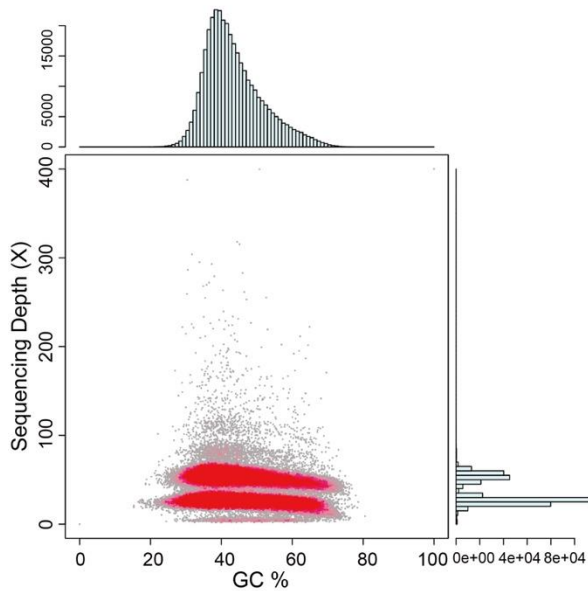


Figure S4. GC content and distribution of sequencing depth based on short reads from the Illumina platform

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

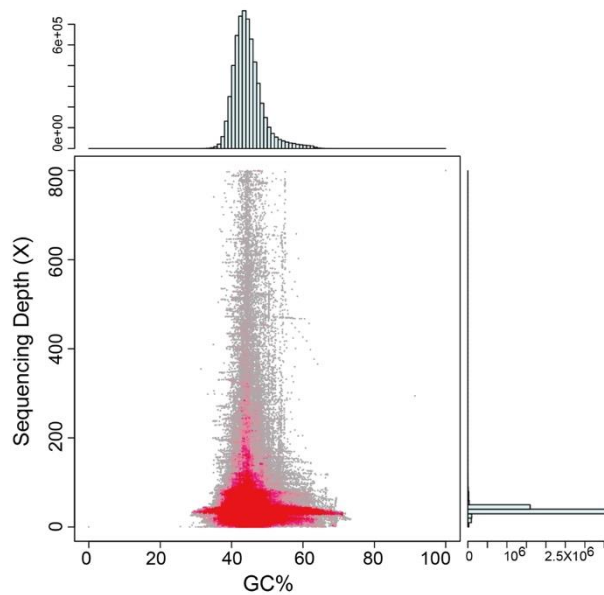


Figure S5. GC content and distribution of sequencing depth based on long reads from the PacBio platform

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

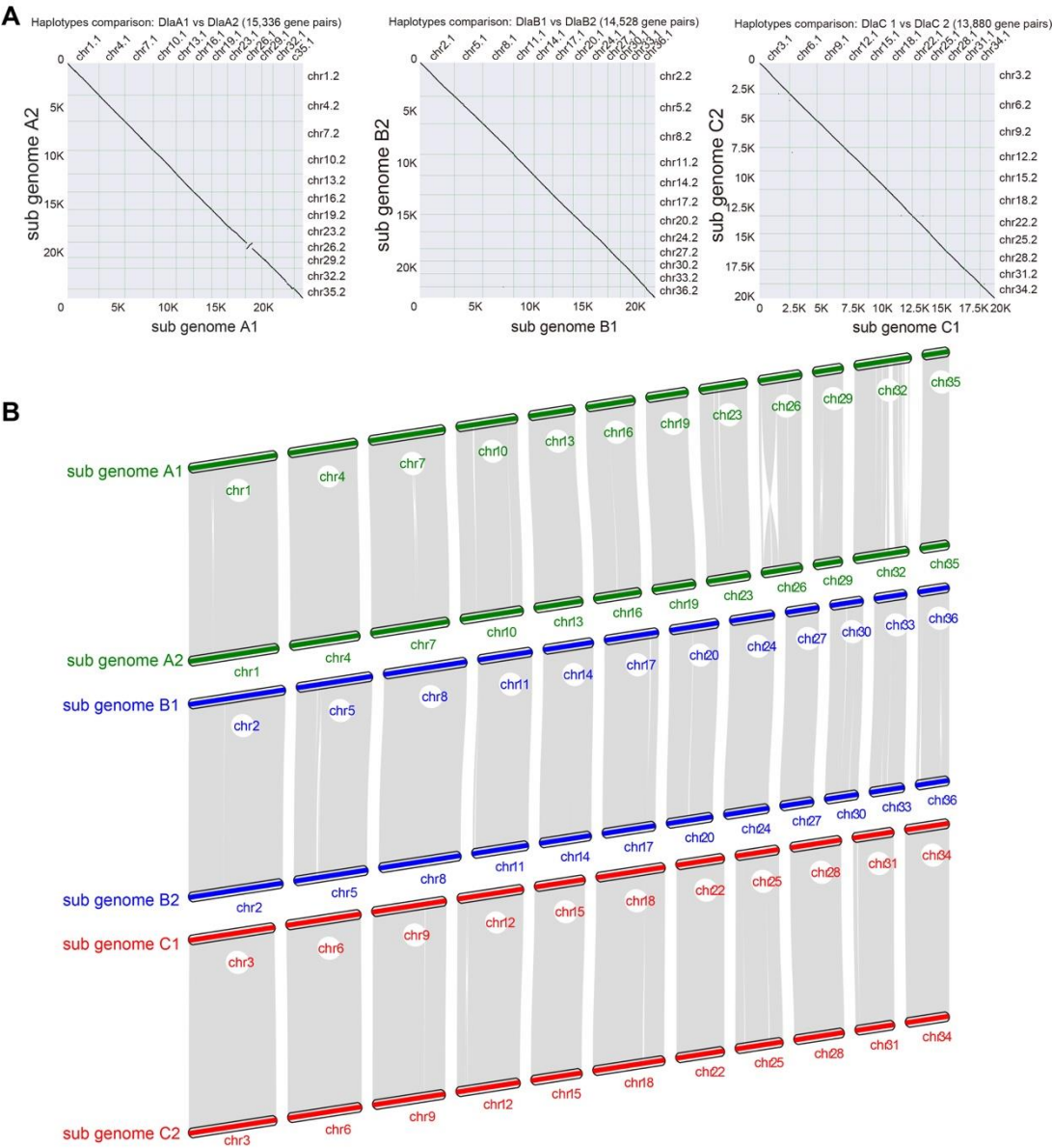


Figure S6. Haplotype-to-haplotype collinearity comparison

(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

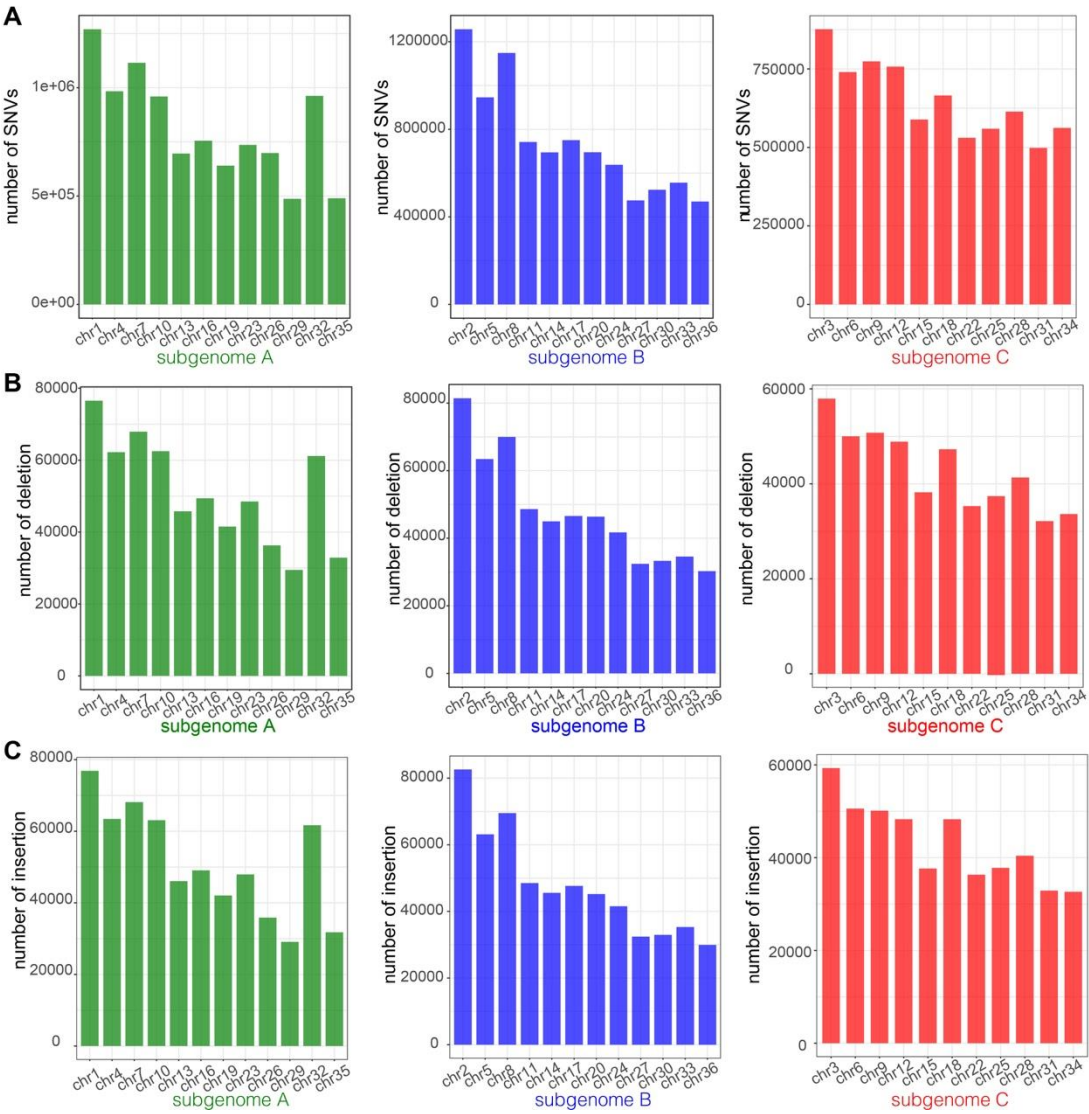


Figure S7. Single nucleotide variations (SNVs) and InDels (insertions/deletions) between the two haplotypes in the three subgenomes. The quantity of SNVs/INDELs was generated from pairwise comparison between the two haplotypes (A1 vs A1, B1 vs B2, and C1 vs C2). **(A)** SNVs. **(B)** Deletions. **(C)** Insertions.

Figure S8

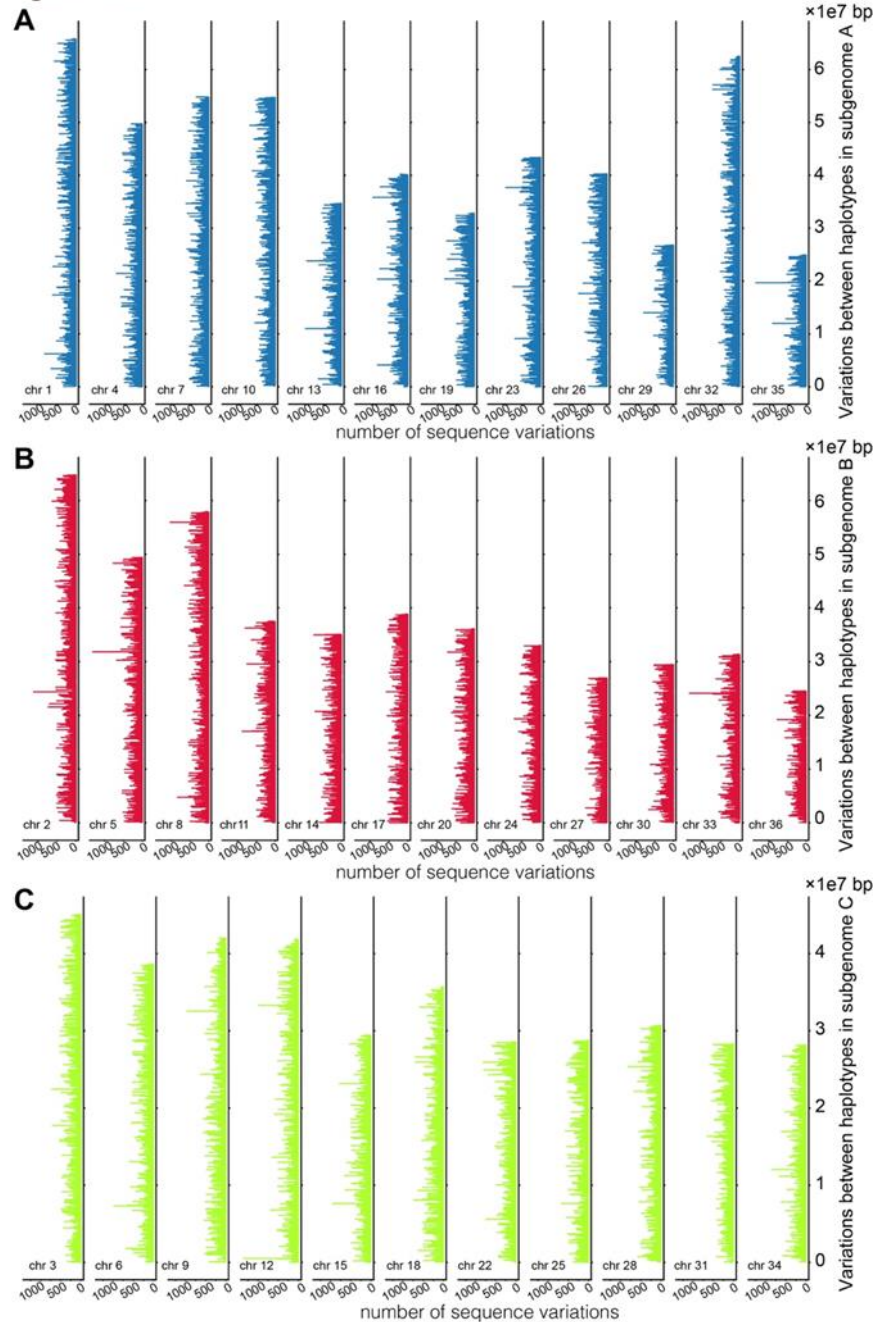


Figure S8. Sequence variations between two haplotypes

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

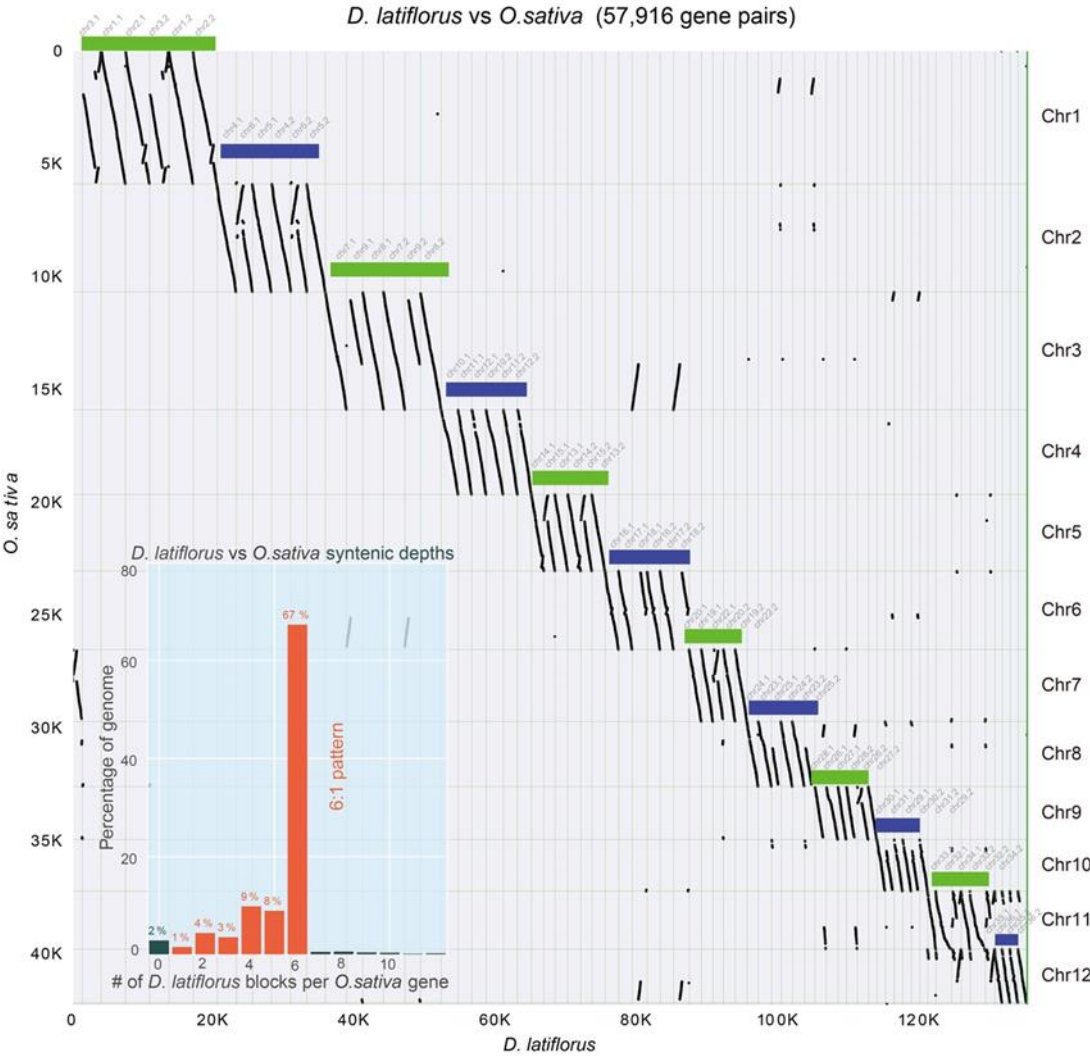


Figure S9. Visualization of pairwise synteny
 Dot plot showing a 1:6 synteny pattern between *O. sativa* and *D. latiflorus*.
 The 1:6 synteny pattern is also validated by the histogram.

Figure S10

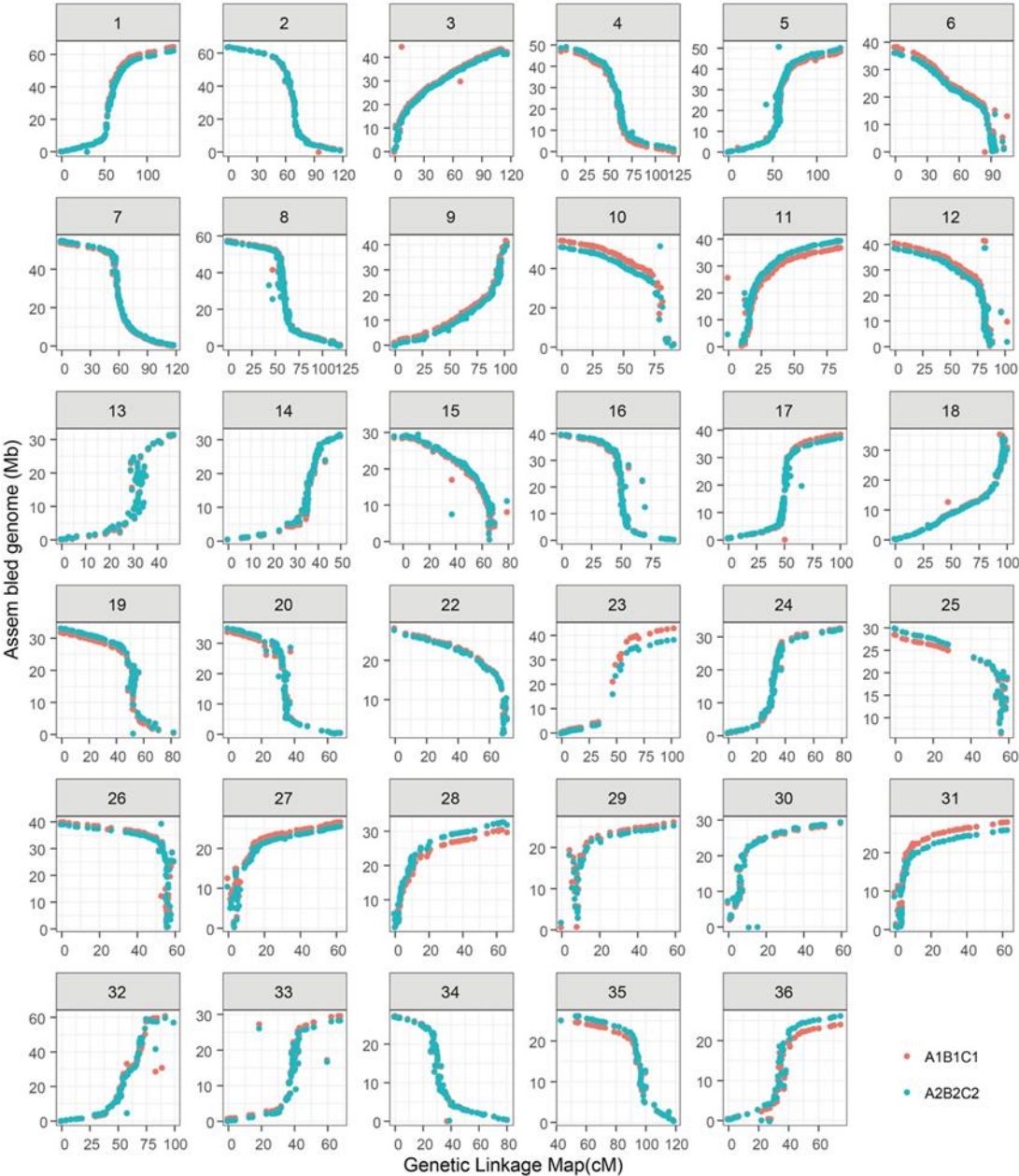
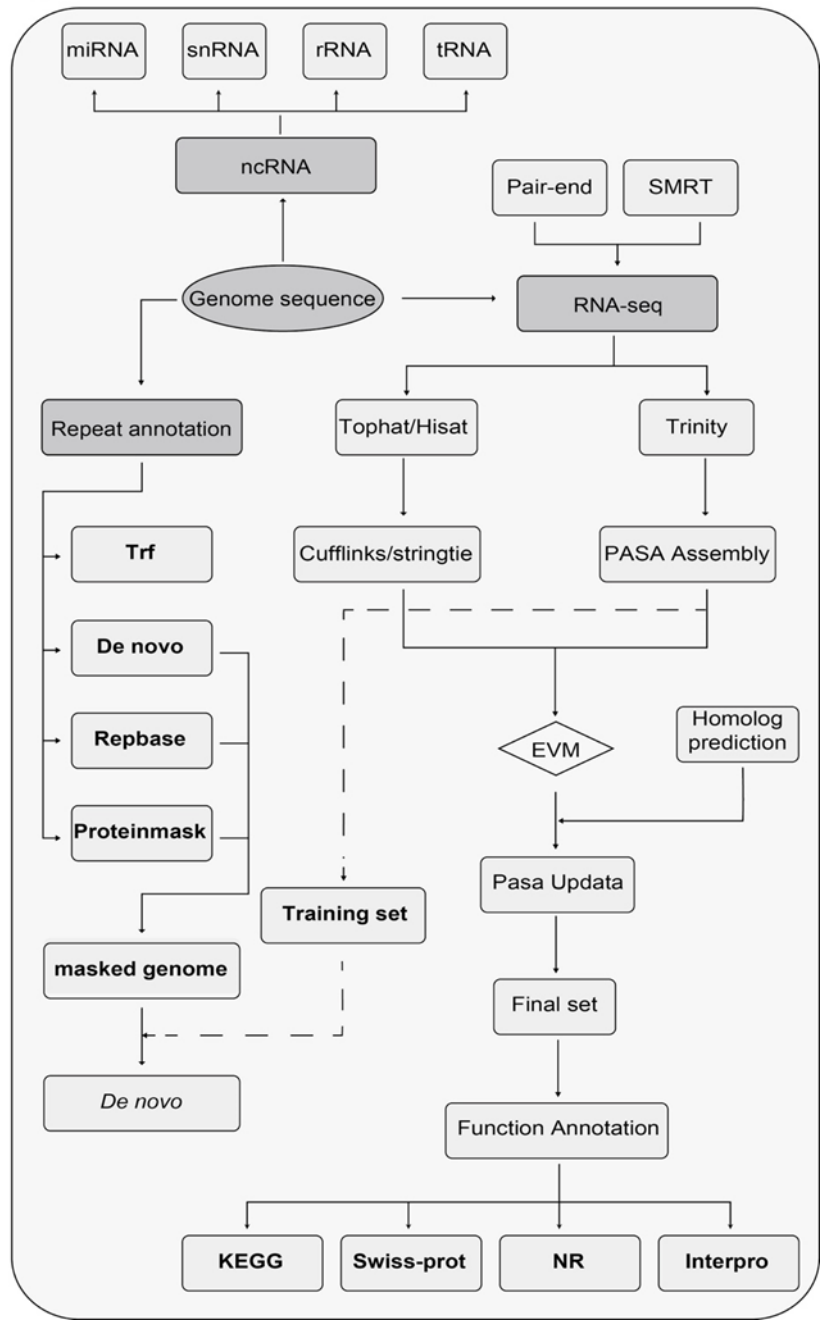


Figure S10. High-density genetic map of bamboo supports the chromosomal assignments

Figure S11



73

74 **Figure S11.** Flowchart for the annotation of protein-coding genes and TEs

Figure S12

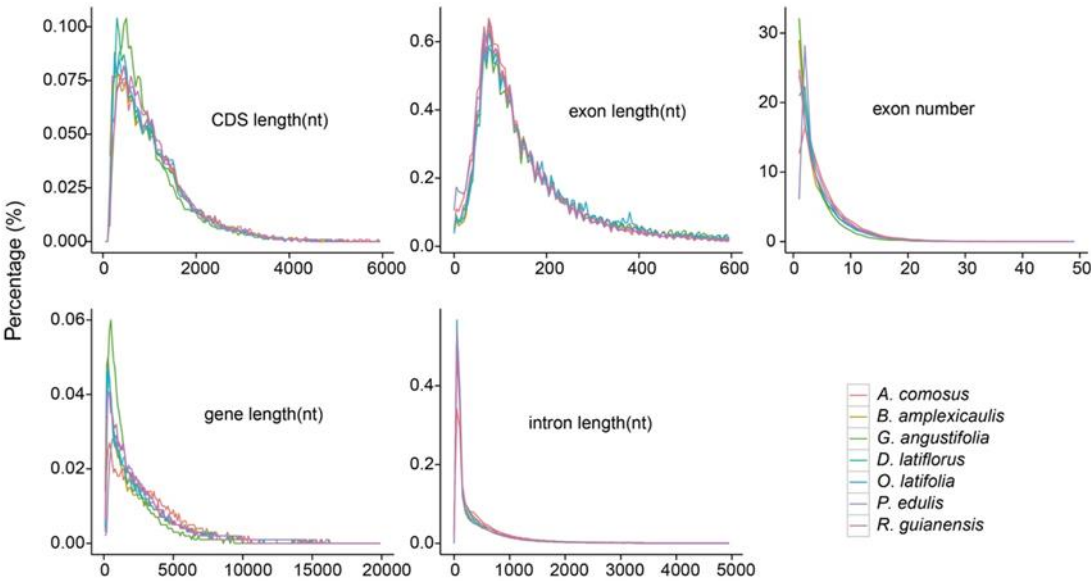


Figure S12. Comparison of gene structures among closely related species

Figure S13

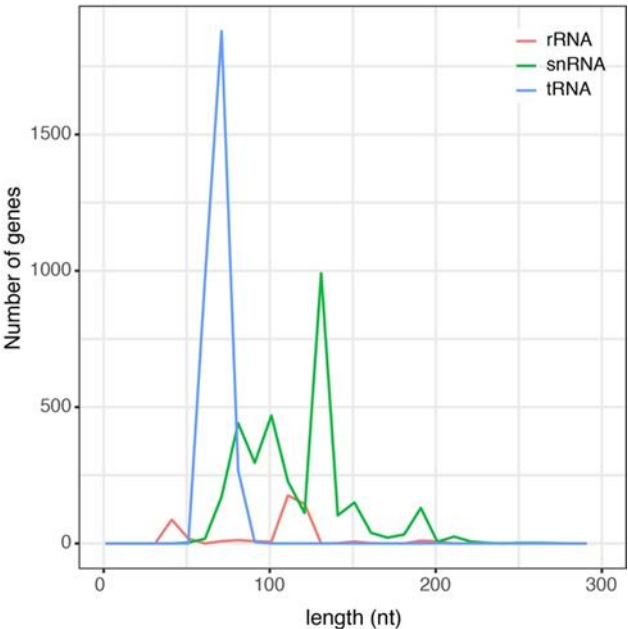


Figure S13. Distribution intervals for non-coding genes

Figure S14

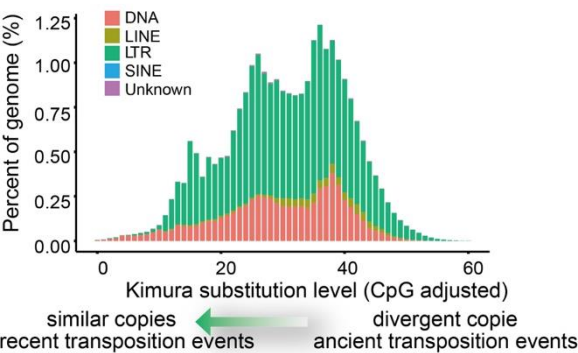


Figure S14. Graphs showing the TE accumulation history in *D. latiflorus* using Kimura distance-based copy divergence analysis. The x-axis represents the Kimura distances to their corresponding consensus sequences; the y-axis represents the percentage of the genome occupied by each type of TE, including DNA, LINE, LTR, and SINE.

Figure S15

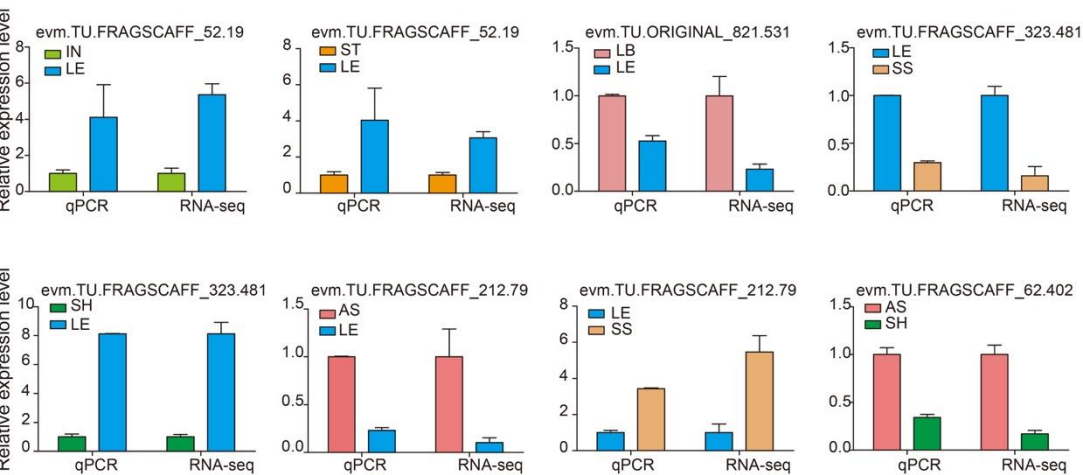


Figure S15. qRT-PCR validation of differentially expressed genes

Figure S16

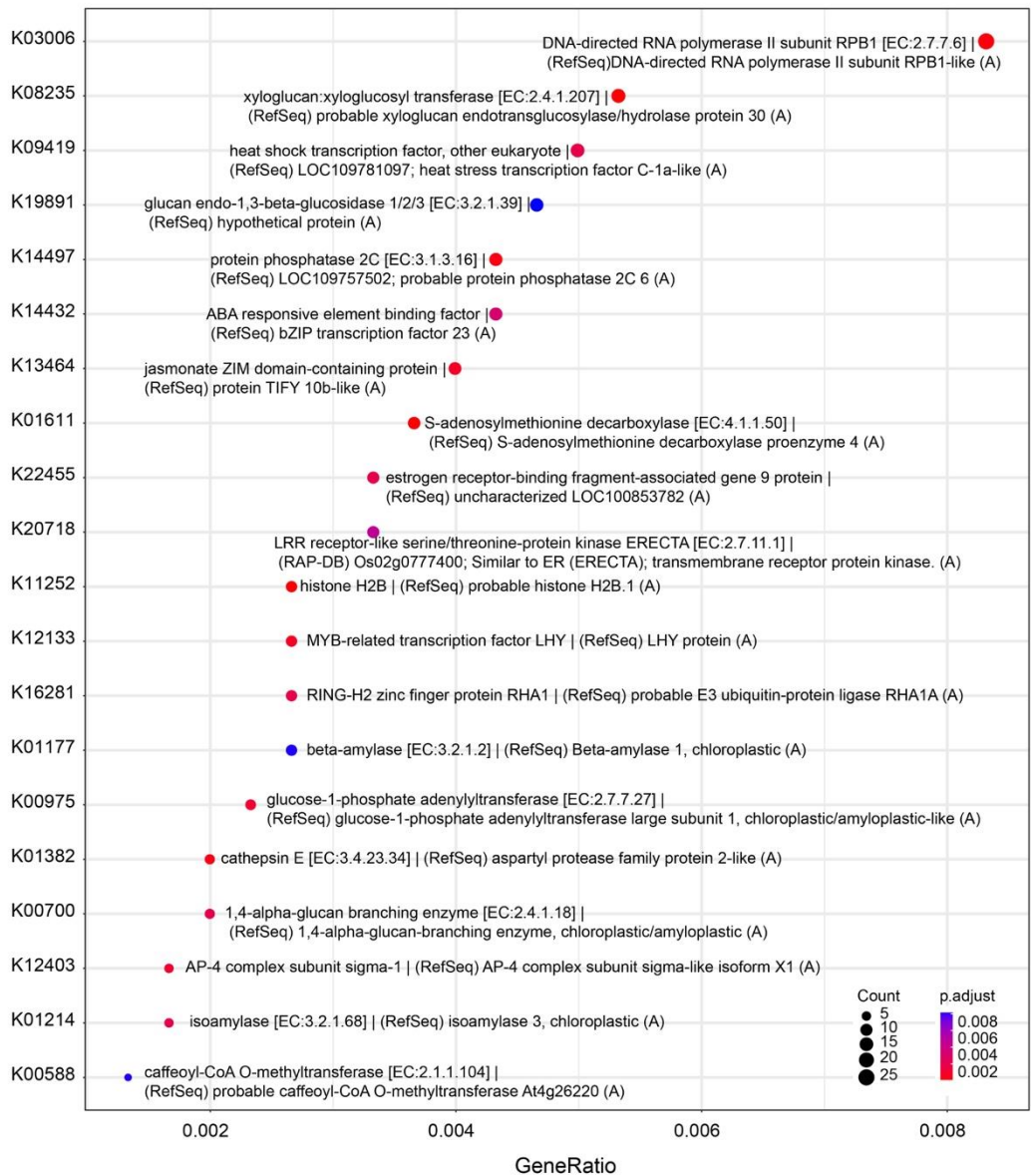


Figure S16. KEGG enrichment of differentially expressed genes between non-abortive and abortive shoots

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

Figure S17

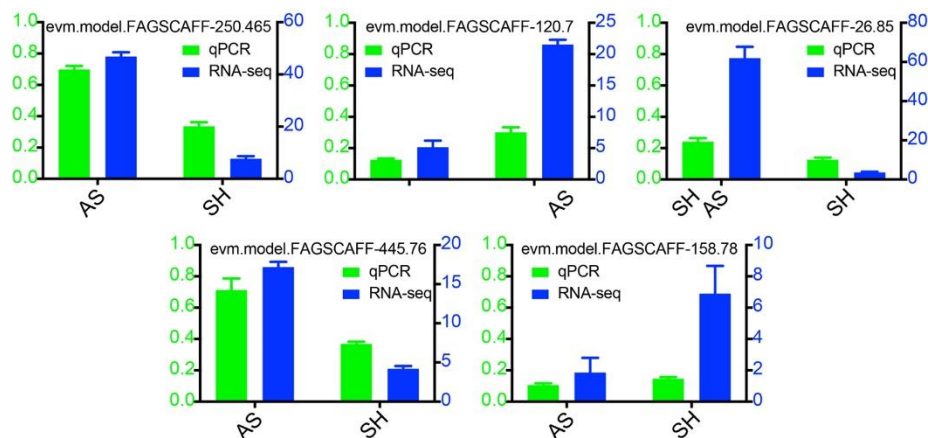


Figure S17. qRT-PCR validation of the differential expression of metabolic pathway genes

Figure S18

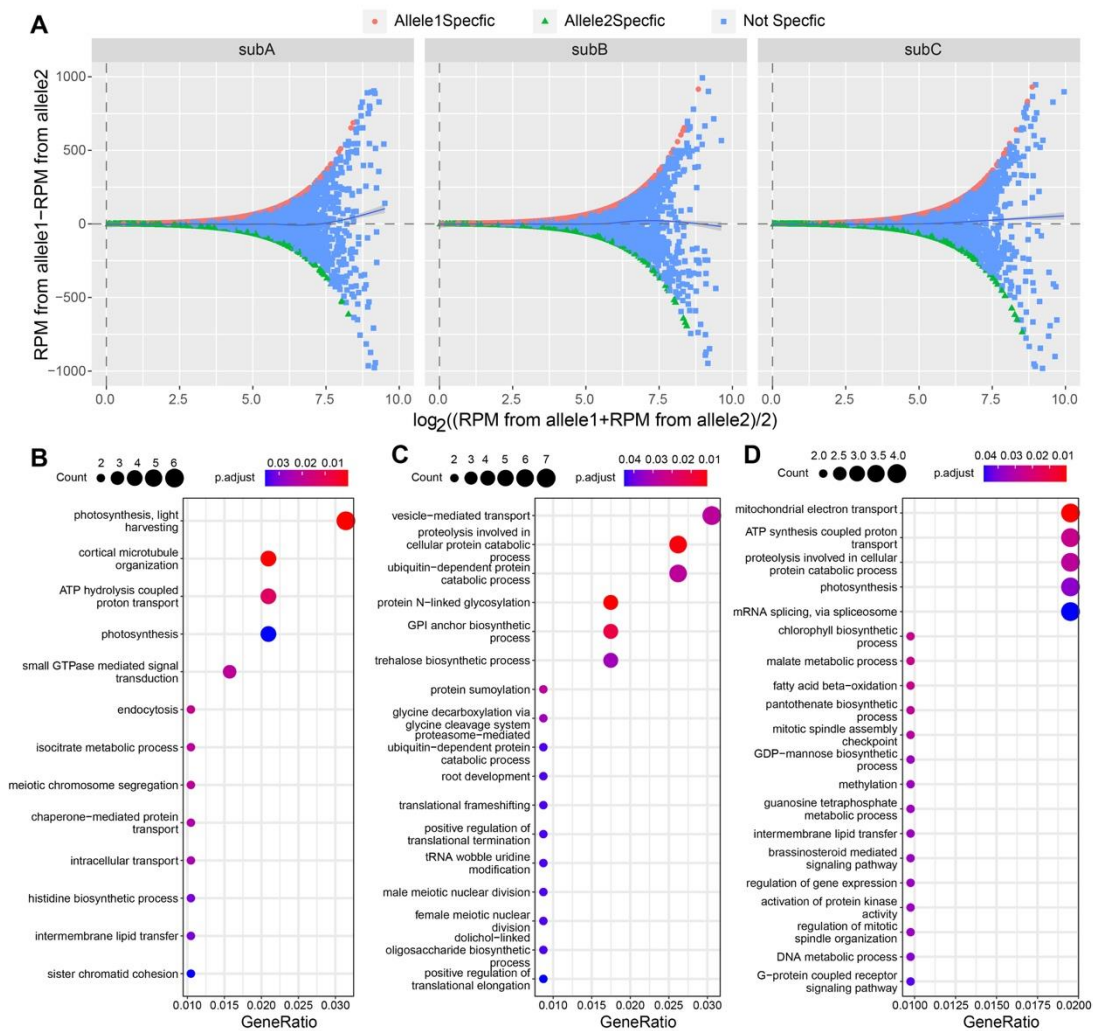


Figure S18. Allele-specific expressed (ASE) genes

(A) MA plot of allele expression. **(B)**, **(C)**, and **(D)** show the enriched GO terms of allele-specific expressed genes in subgenomes **A**, **B**, and **C**, respectively.

Figure S19

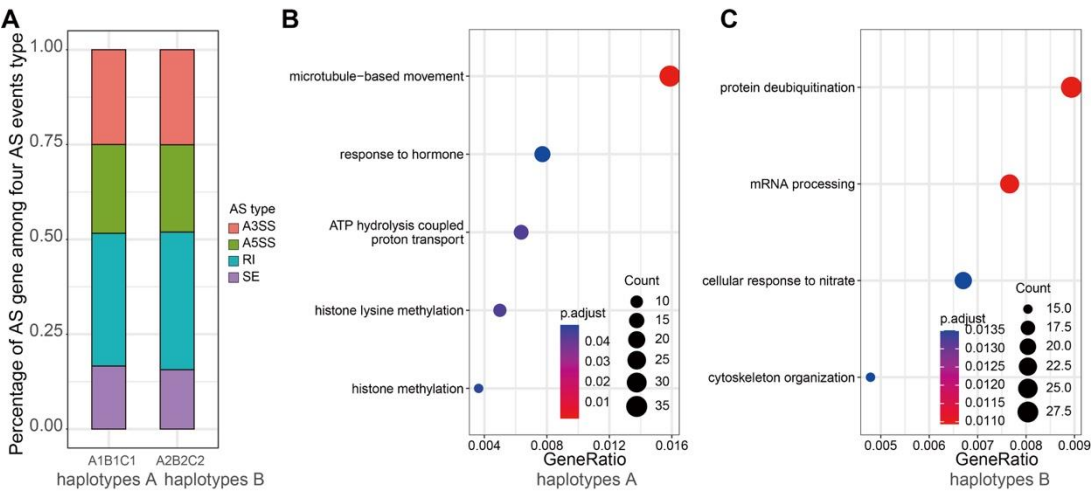


Figure S19. Allele-specific genes showing alternative splicing

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

Figure S20

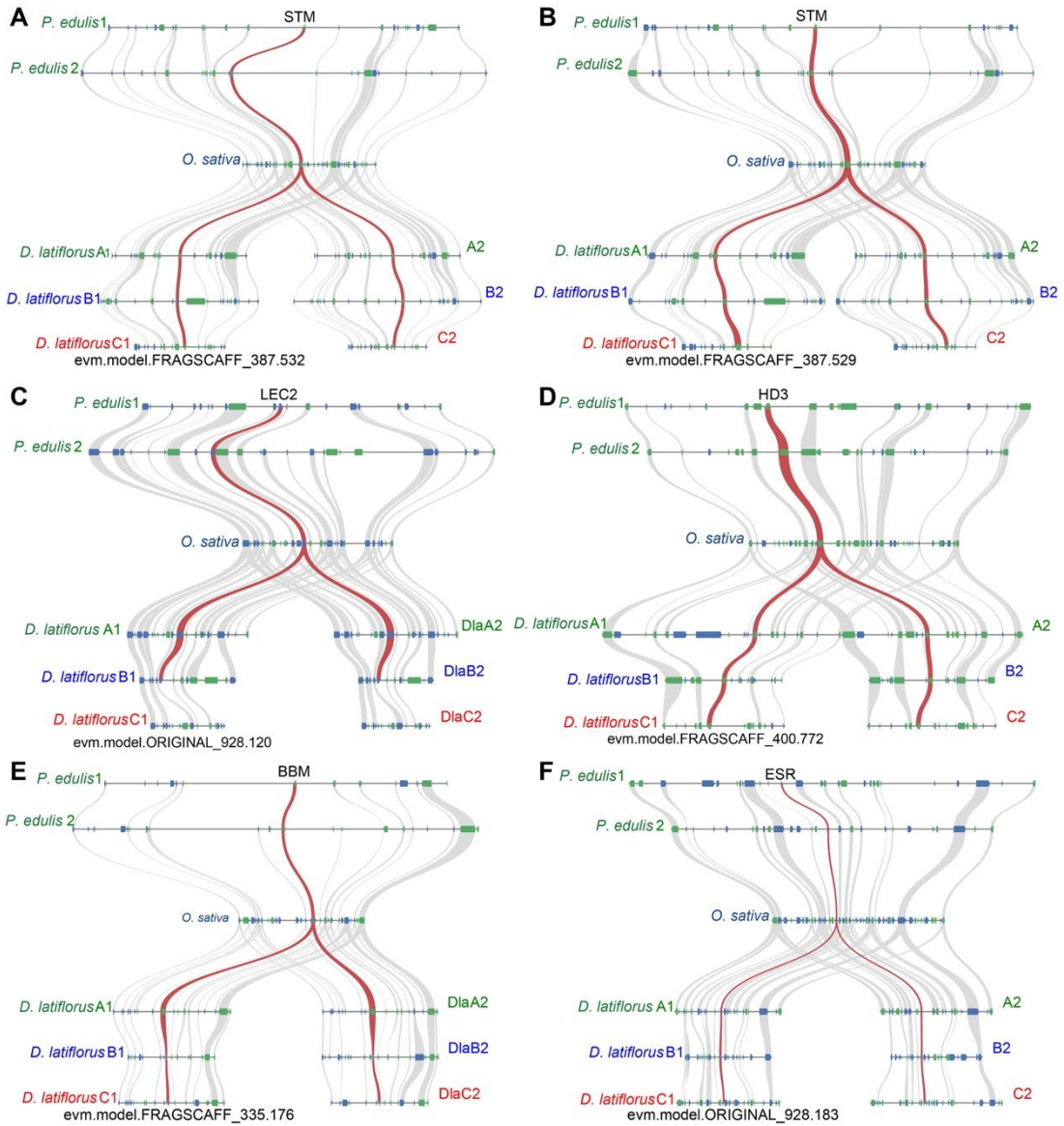


Figure S20. Micro-synteny analysis of developmental genes related to the transformation efficiency of *P. edulis*, *O. sativa*, and *D. latiflorus*

Figure S21

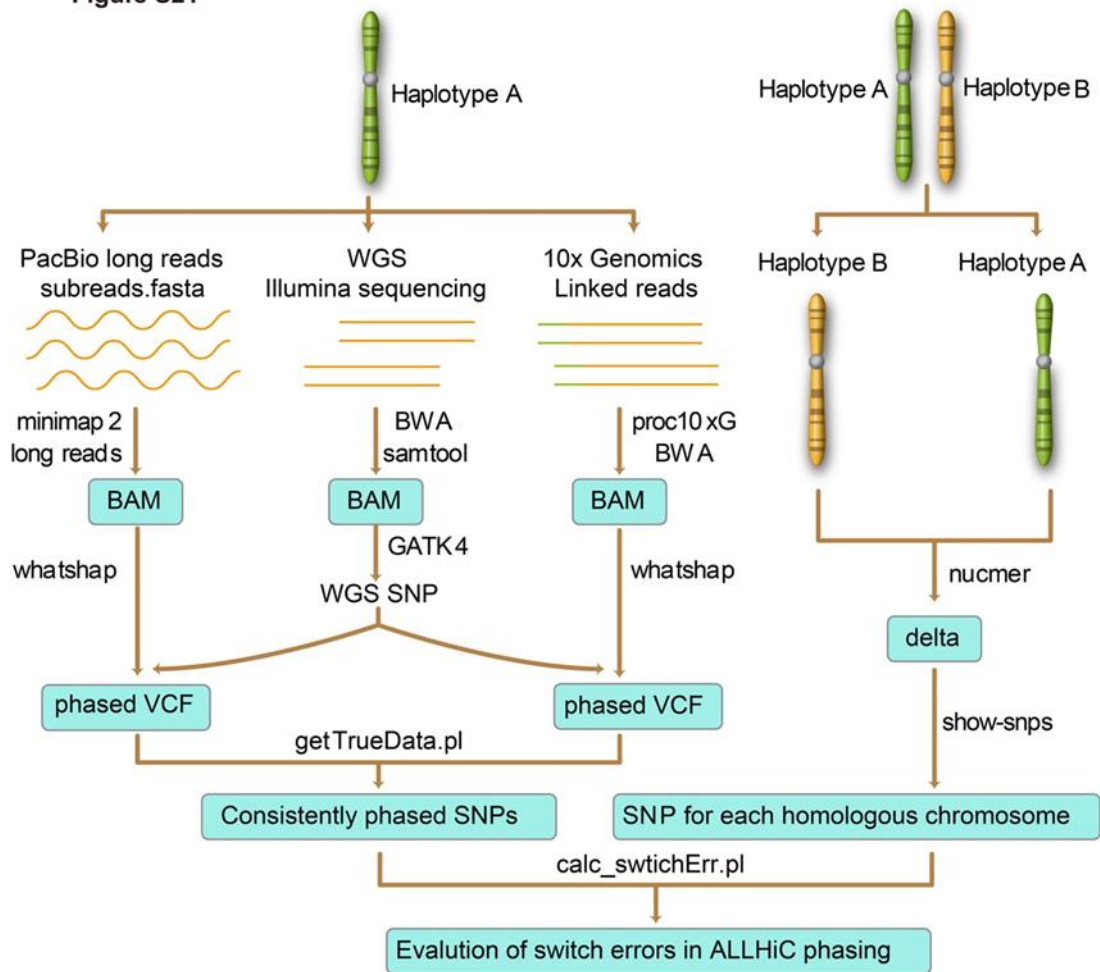


Figure S21. Flowchart for assessing the accuracy of the haplotype-resolved genome assembly

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

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

120

121 **Table S3.** Statistics of percentage of different type tags that the Hi-C
122 generated for chromosome-scale assembly of *D. latiflorus*

123

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

126

127 **Table S5.** Assessment of genome assembly and annotation completeness
128 using BUSCO

129

130 **Table S6.** Assessment of genome assembly and annotation completeness
131 using CEGMA

132

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

134

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

136

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

138

139 **Table S10.** Annotation of the hexaploid genome of *D. latiflorus*

140

141 **Table S11.** Functional assignment of the hexaploid genome of *D. latiflorus*

142 **Table S12.** Overview of Non-coding RNA(ncRNA) genes in *D. latiflorus*

143

144 **Table S13.** Statistics of transposable element in *D. latiflorus*

145

146 **Table S14.** qRT-PCR validation of differentially expressed genes

147

148 **Table S15.** qRT-PCR validation of genes in differential metabolism pathway