

Comparative genome analyses highlight transposon-mediated genome expansion shapes the evolutionary architecture of 3D genomic folding in cotton

Maojun Wang^{1,3*}, Jianying Li^{1,3}, Pengcheng Wang¹, Fang Liu², Zhenping Liu¹, Guannan Zhao¹, Zhongping Xu¹, Kunbo Wang^{2*}, and Xianlong Zhang^{1*}

¹National Key Laboratory of Crop Genetic Improvement, Huazhong Agricultural University, Wuhan 430070, Hubei, China.

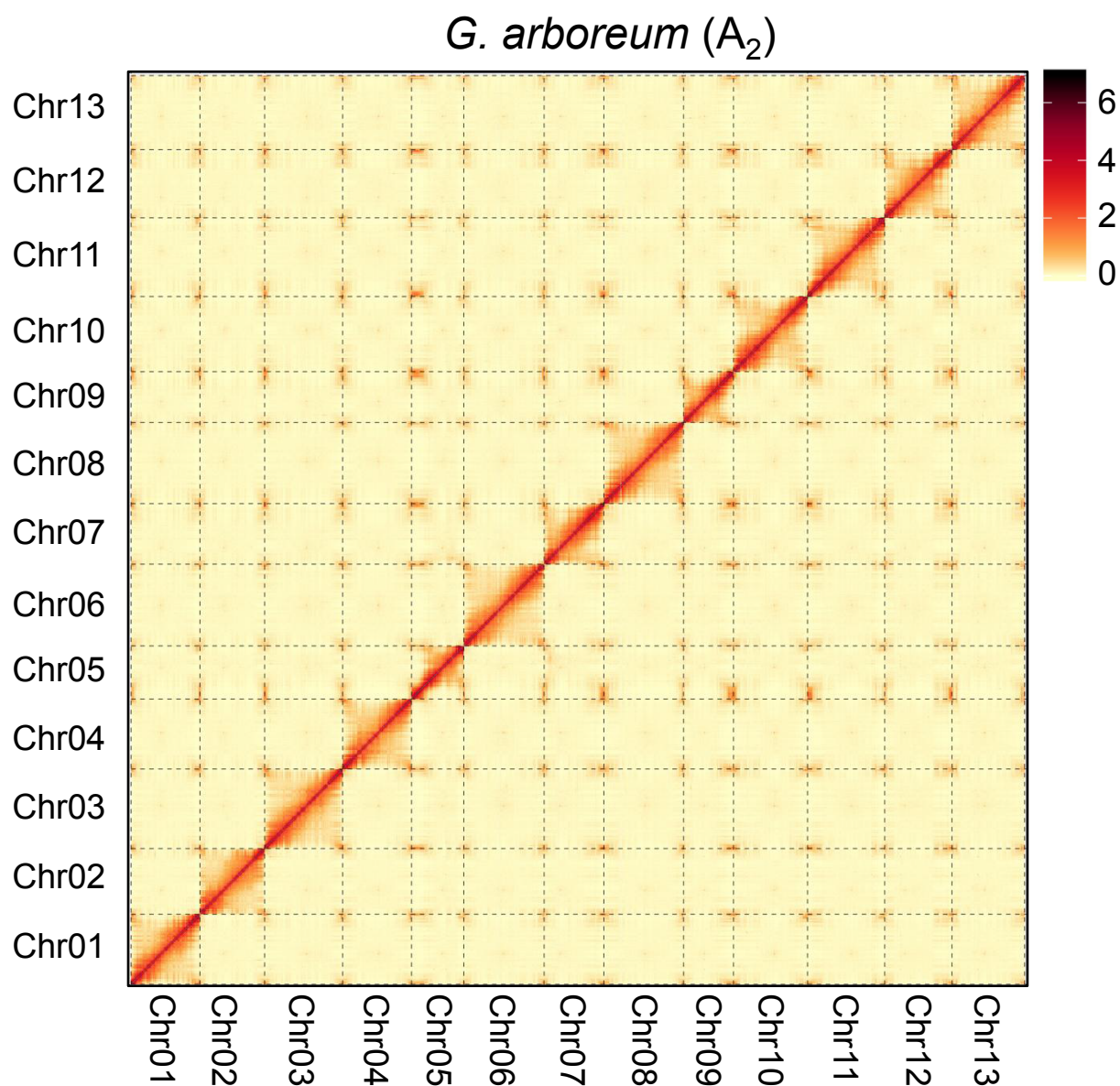
²State Key Laboratory of Cotton Biology, Institute of Cotton Research, Chinese Academy of Agricultural Sciences, Anyang 455000, Henan, China.

³These authors contributed equally to this work.

Correspondence should be addressed to M.W. (mjwang@mail.hzau.edu.cn), K.W. (wangkunbo@caas.cn) or X.Z. (xlzhang@mail.hzau.edu.cn).

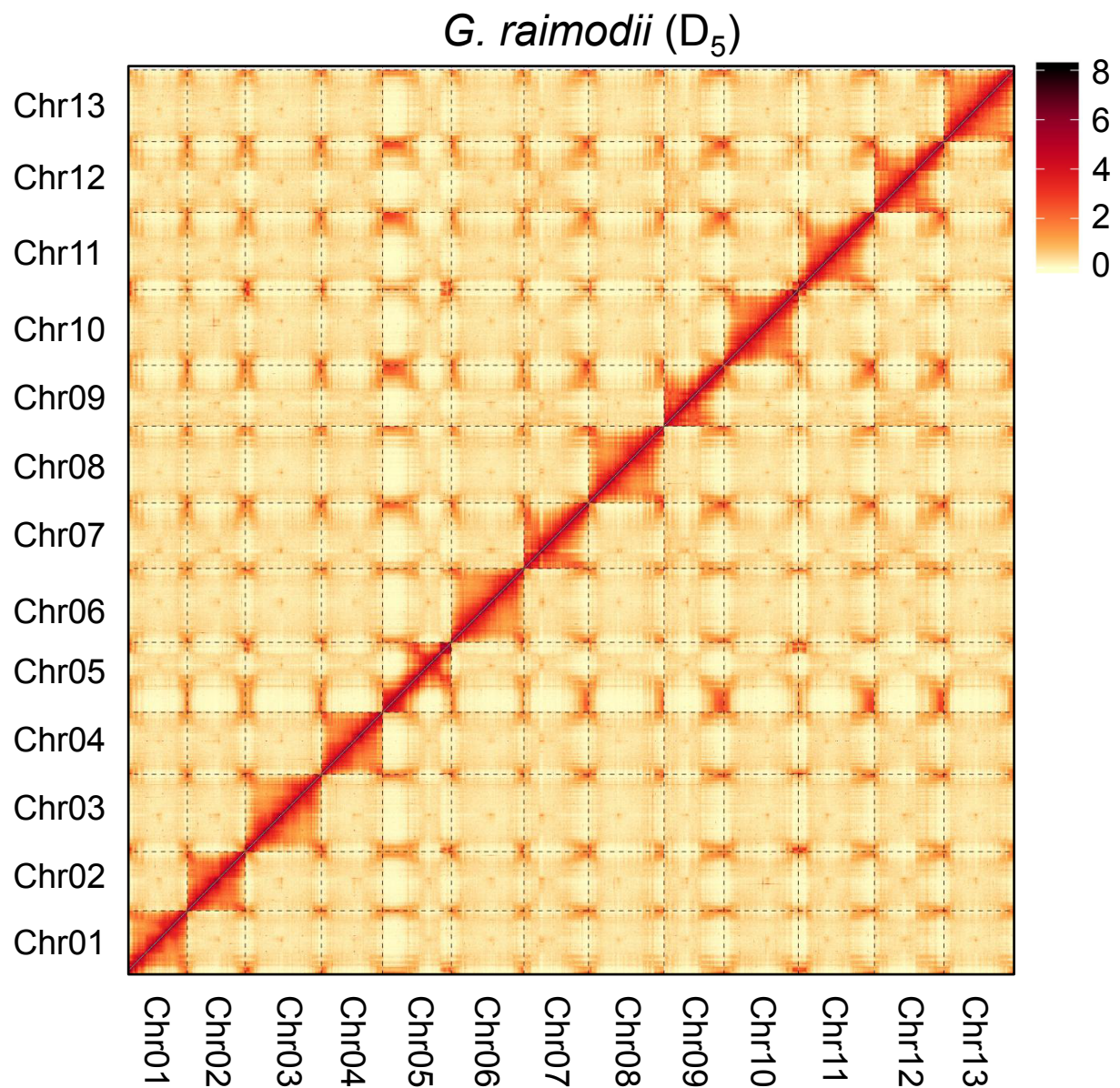
Supplementary Figures

Supplementary Figures 1–19



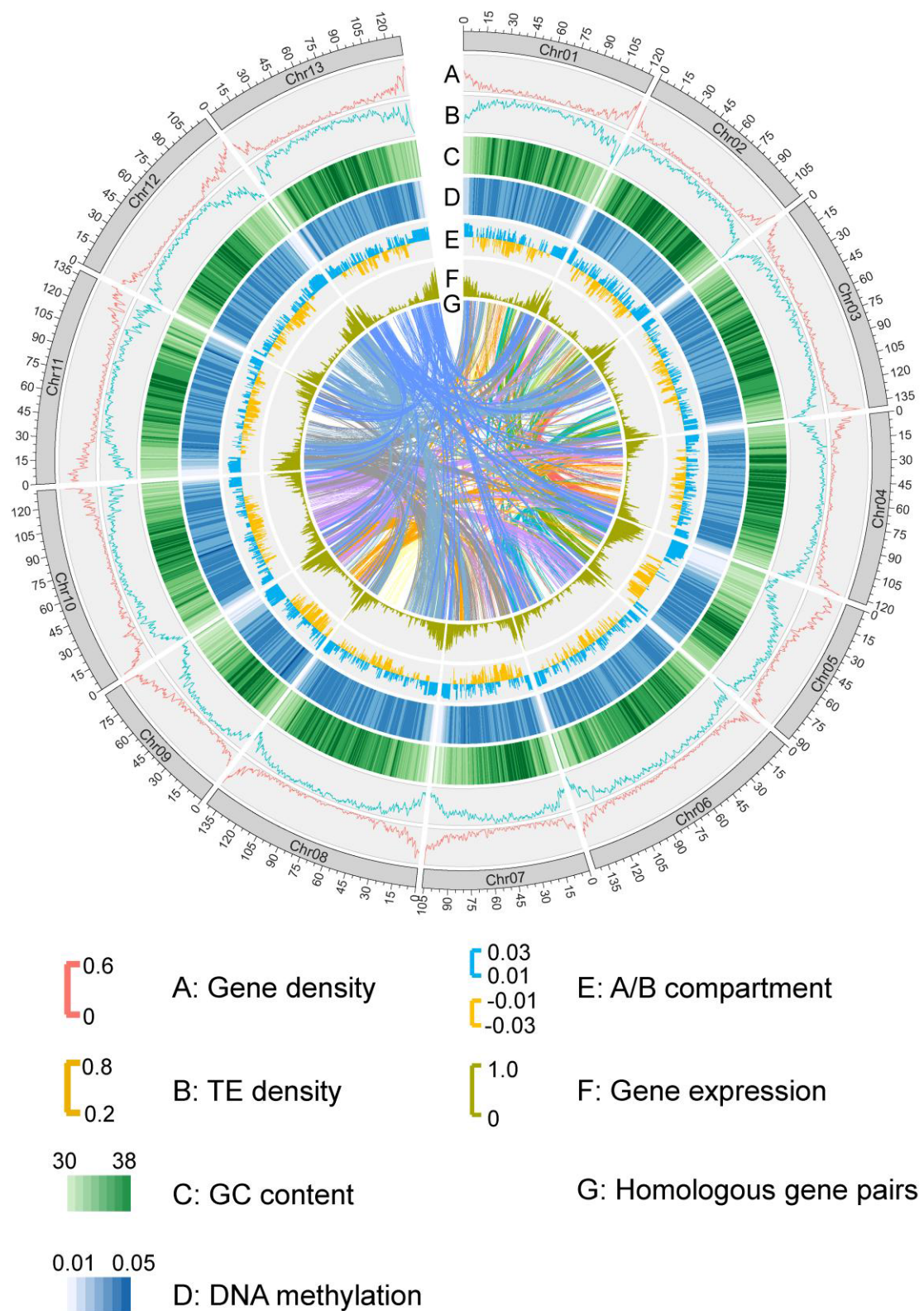
Supplementary Figure 1

Hi-C heatmap showing chromatin interactions at 100 Kb resolution in *G. arboreum* (A₂) .



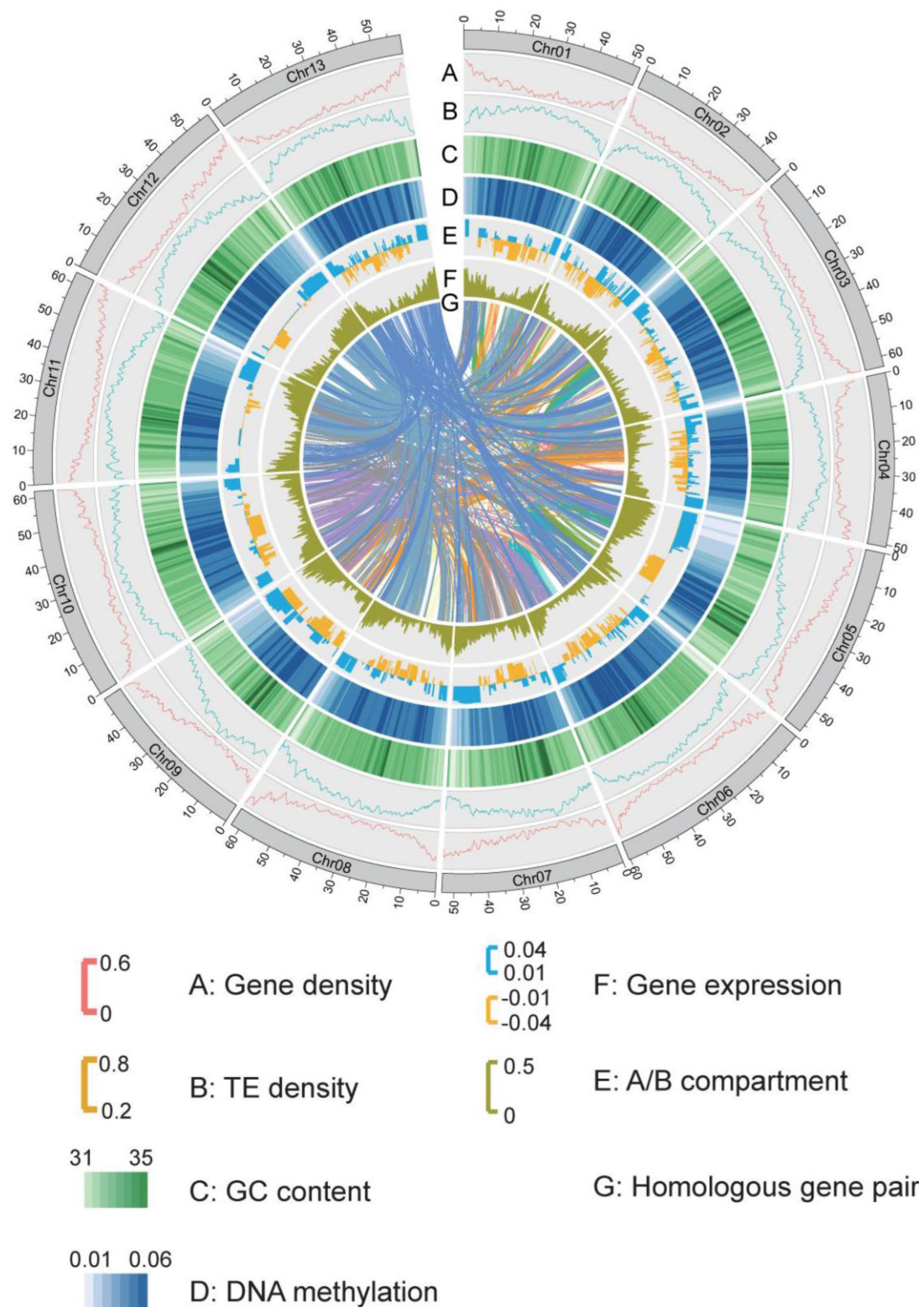
Supplementary Figure 2

Hi-C heatmap showing chromatin interactions at 100 Kb resolution in *G. raimondii* (D₅).



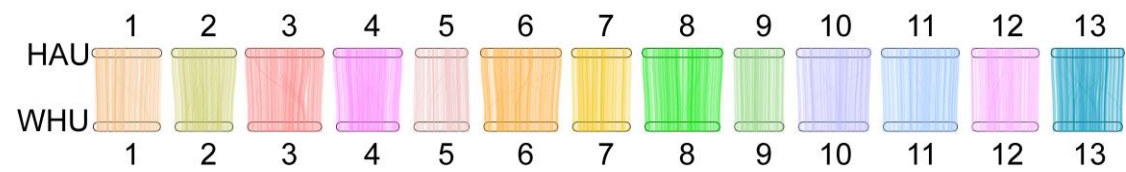
Supplementary Figure 3

Chromosomal characteristics of *G. arboreum* (A₂) genome. The tracks represent gene density (A), transposable element (TE) density (B), GC content (C), DNA methylation (D), A/B compartment (E), gene expression (F) and homologous gene pairs (G) in each chromosome. In each track (A-D, F), the feature data is shown in 1 Mb windows sliding 200 Kb.



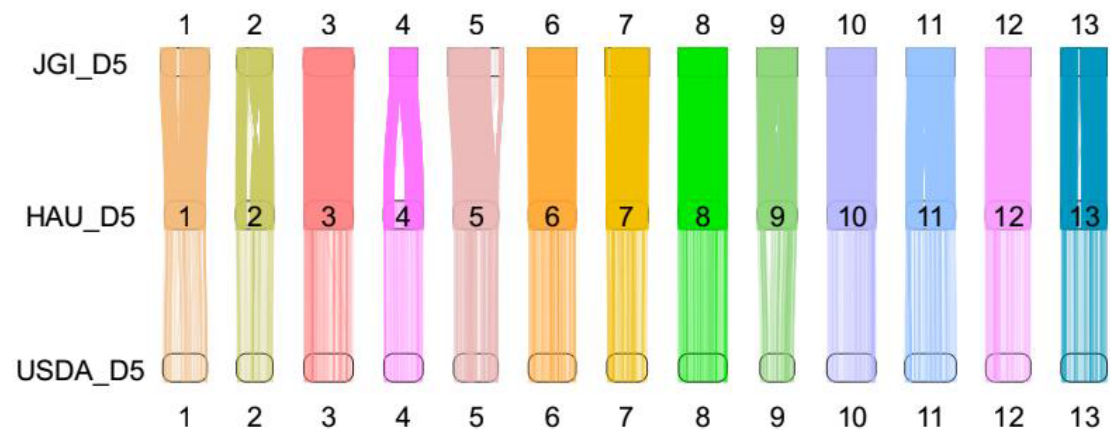
Supplementary Figure 4

Chromosomal characteristics of *G. raimondii* (D₅) genome. The tracks represent gene density (A), transposable element (TE) density (B), GC content (C), DNA methylation (D), A/B compartment (E), gene expression (F) and homologous gene pairs (G) in each chromosome. In each track (A-D, F), the feature data is shown in 1 Mb windows sliding 200 Kb.



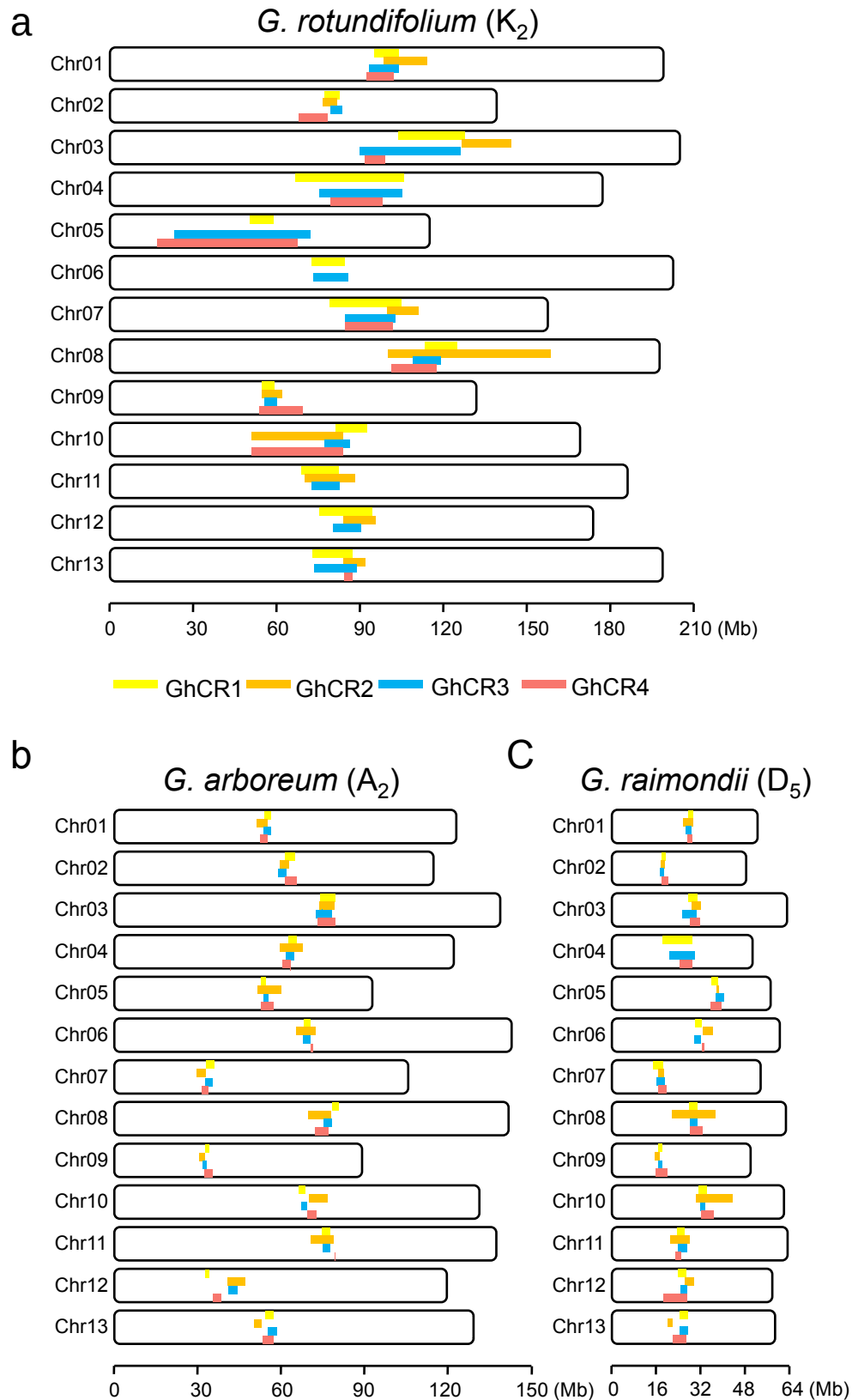
Supplementary Figure 5

Gene syntenic alignment of *G. arboreum* (top panel) in this study with previously published genome (below panel) (Huang et al., 2020). Alignment blocks are shown for each chromosome in lines with different colors.



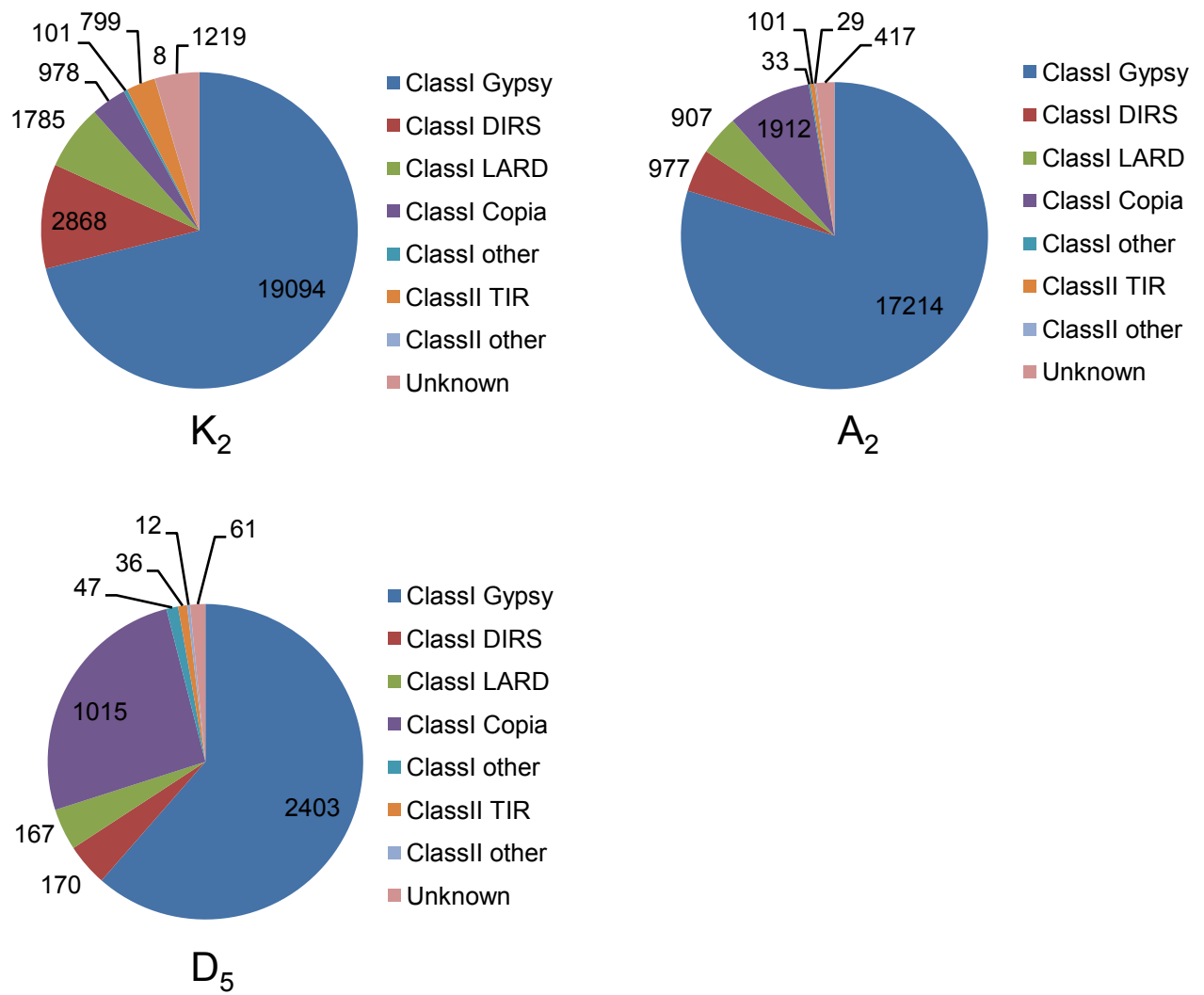
Supplementary Figure 6

Gene syntenic alignment of *G. raimondii* (D5) in this study (HAU) with previously published genomes (USDA, (Udall et al., 2019); JGI, (Paterson et al., 2012)). Alignment blocks are shown for each chromosome in lines with different colors.



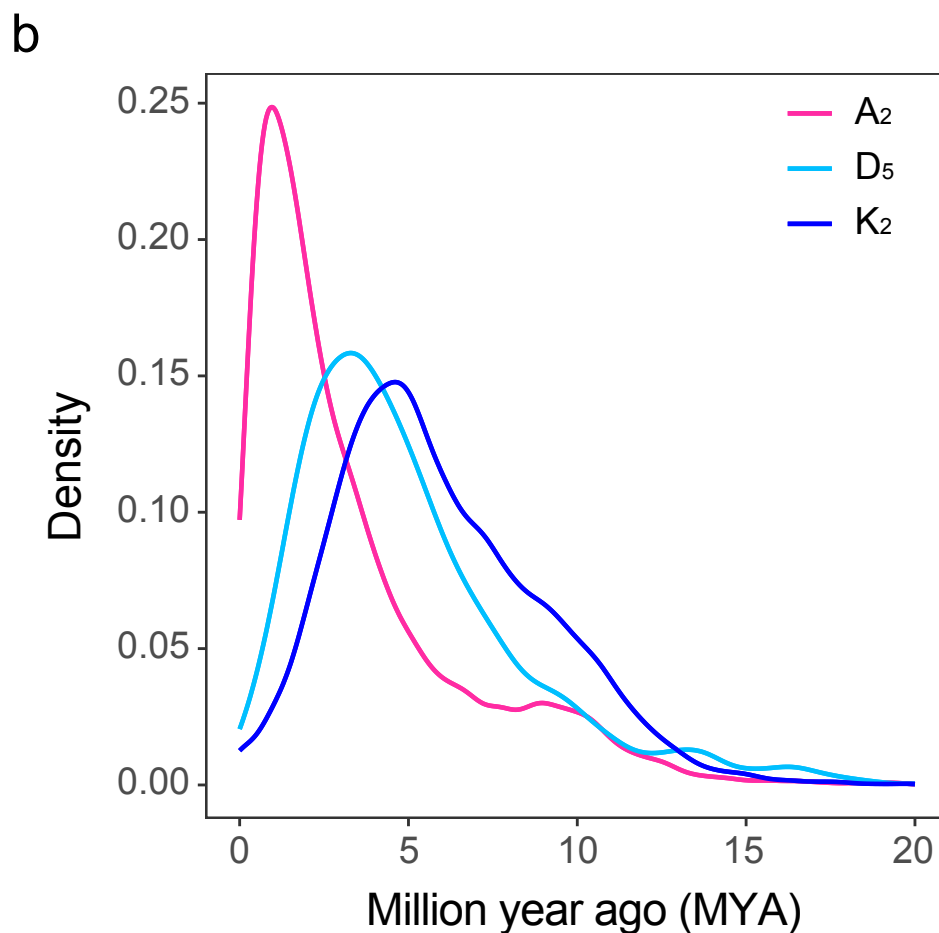
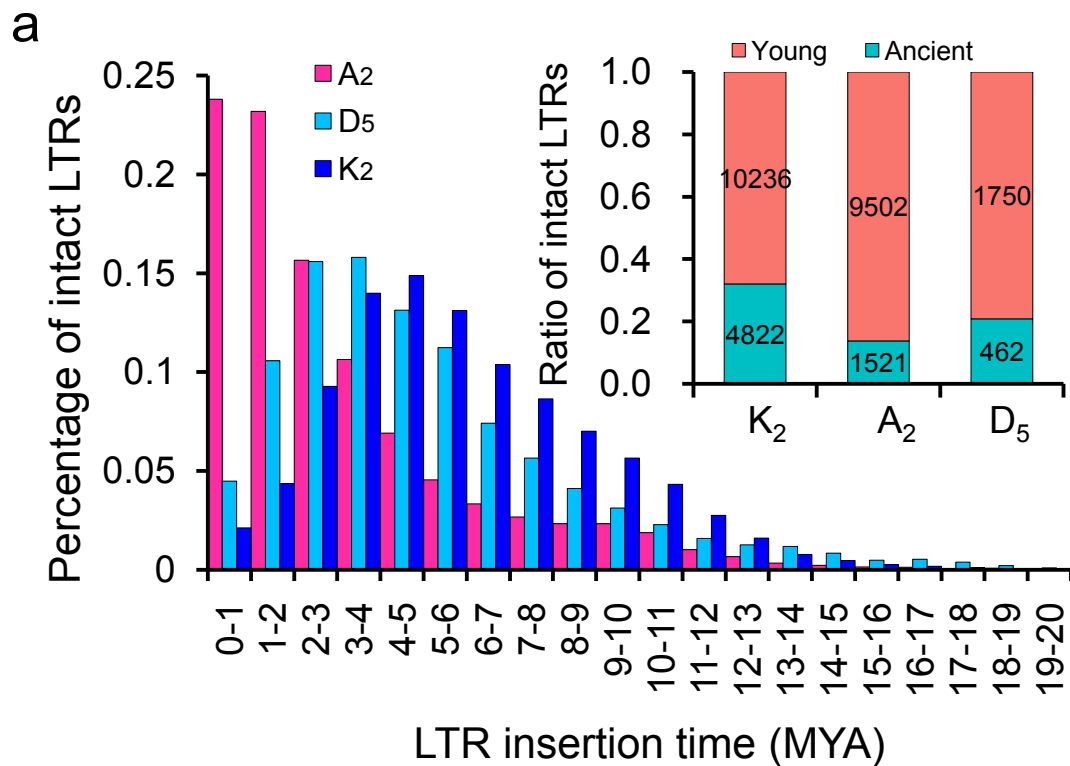
Supplementary Figure 7

Identification of centromeric regions in *G. rotundifolium* (K_2), *G. arboreum* (A_2) and *G. raimondii* (D_5) genomes. The homologous centromeric regions from our recently published TM-1 genomes are shown with boxes in different colors for three genomes (Wang et al., 2019).



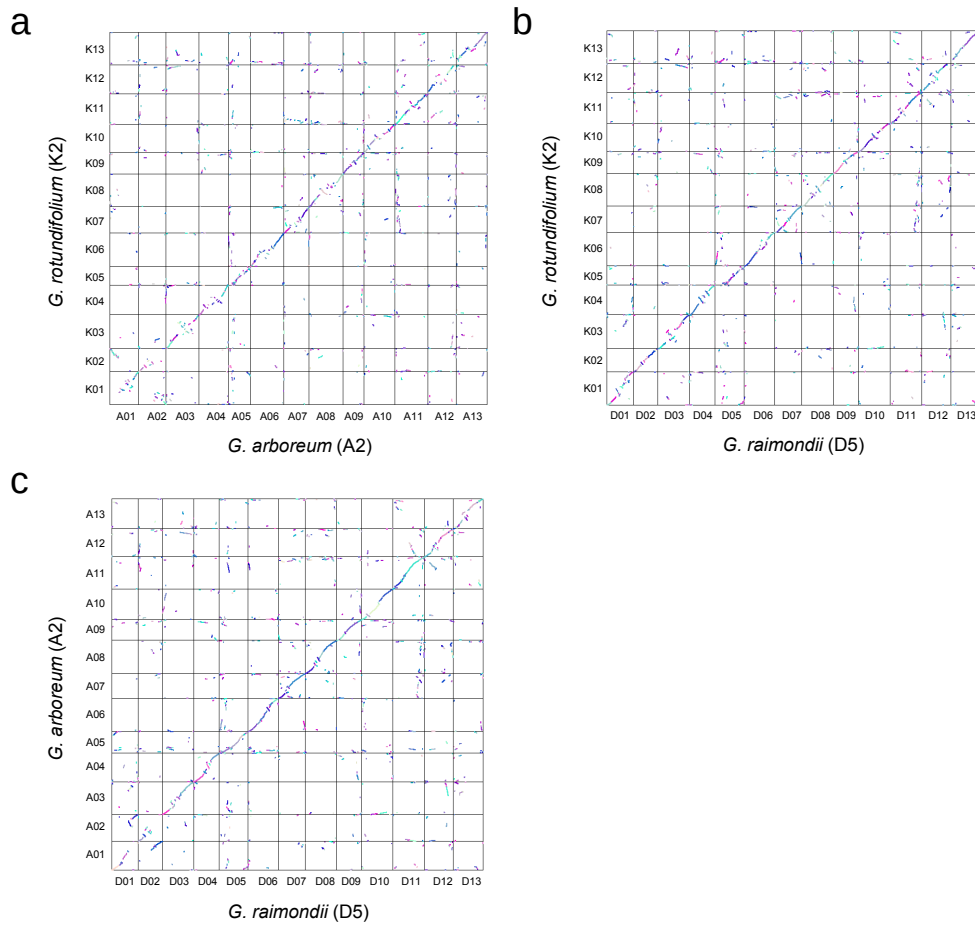
Supplementary Figure 8

Identification of intact LTRs in three genomes.



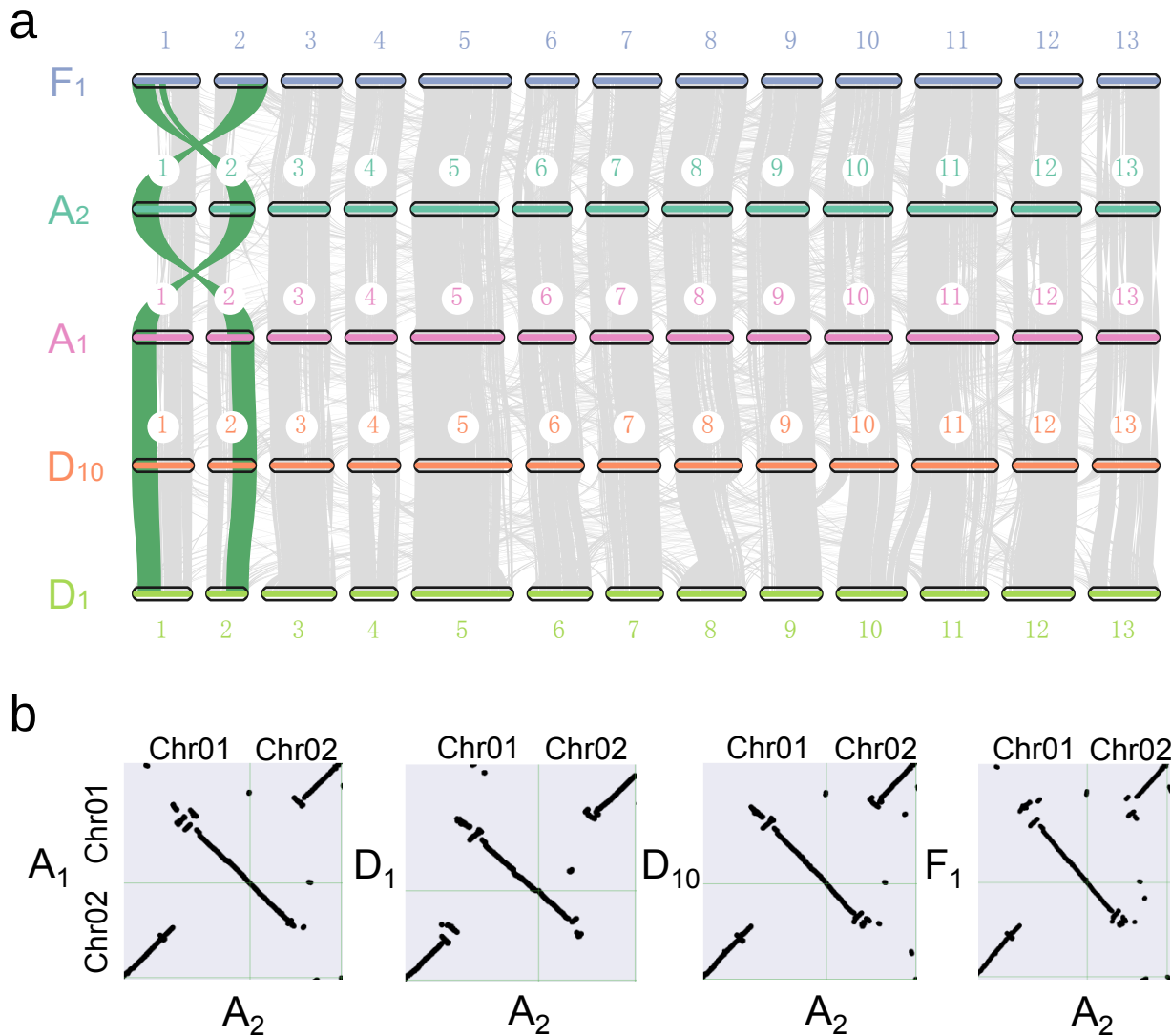
Supplementary Figure 9

Full-length LTR retrotransposon burst time in *G. rotundifolium* (K₂), *G. arboreum* (A₂) and *G. raimondii* (D₅) genomes. **a**, Analysis of the number of LTRs and insertion time in *G. rotundifolium* (K₂), *G. arboreum* (A₂) and *G. raimondii* (D₅). The y axis indicate the number of LTRs, whereas x axis represent the estimated insertion time. **b**, The density of LTR burst in three species.



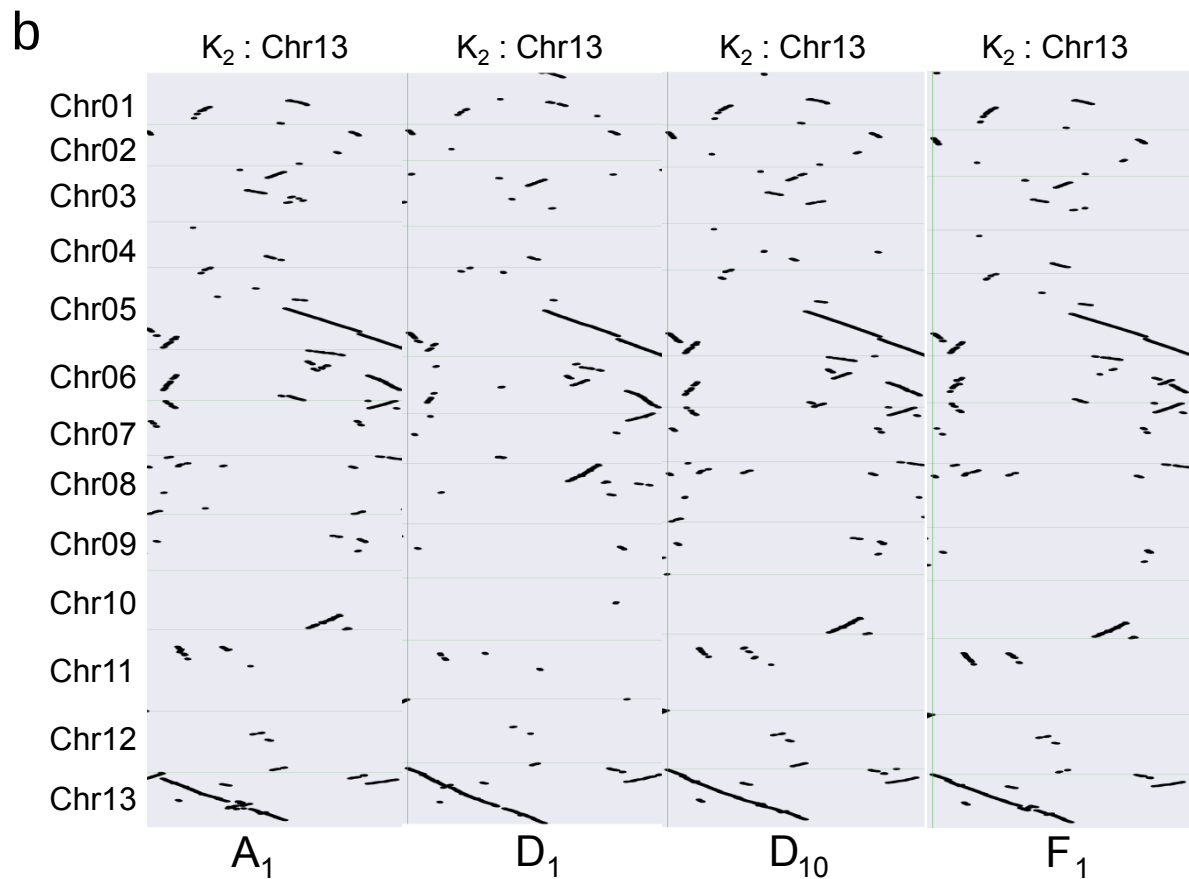
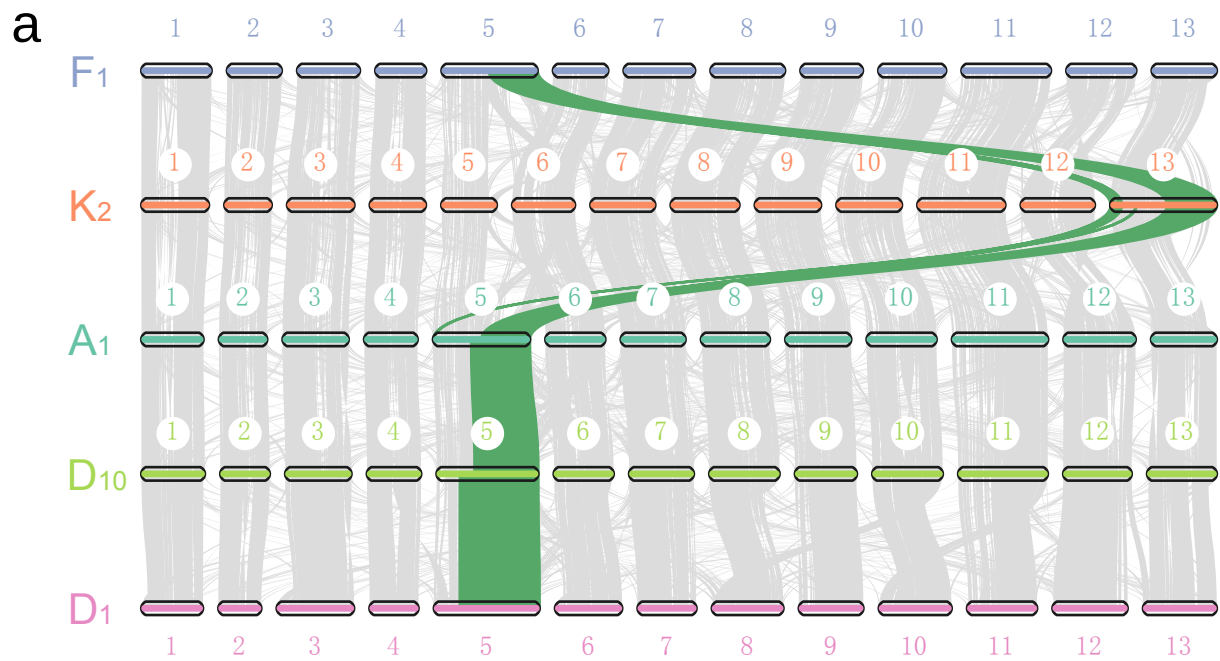
Supplementary Figure 10

Gene synteny between *G. rotundifolium* (K₂), *G. arboreum* (A₂) and *G. raimondii* (D₅) genomes. **a**, Dot plot showing gene synteny between A₂ (y-axis, chromosome A01-A13) and K₂ (x-axis, chromosome K01-K13). **b**, Dot plot showing gene synteny between A₂ (y-axis, chromosome A01-A13) and K₂ (x-axis, chromosome K01-K13). **c**, Dot plot showing gene synteny between D₅ (x-axis, chromosome D01-D13) and A₂ (y-axis, chromosome A01-A13). Genome alignment was carried out using the MCScanX package.



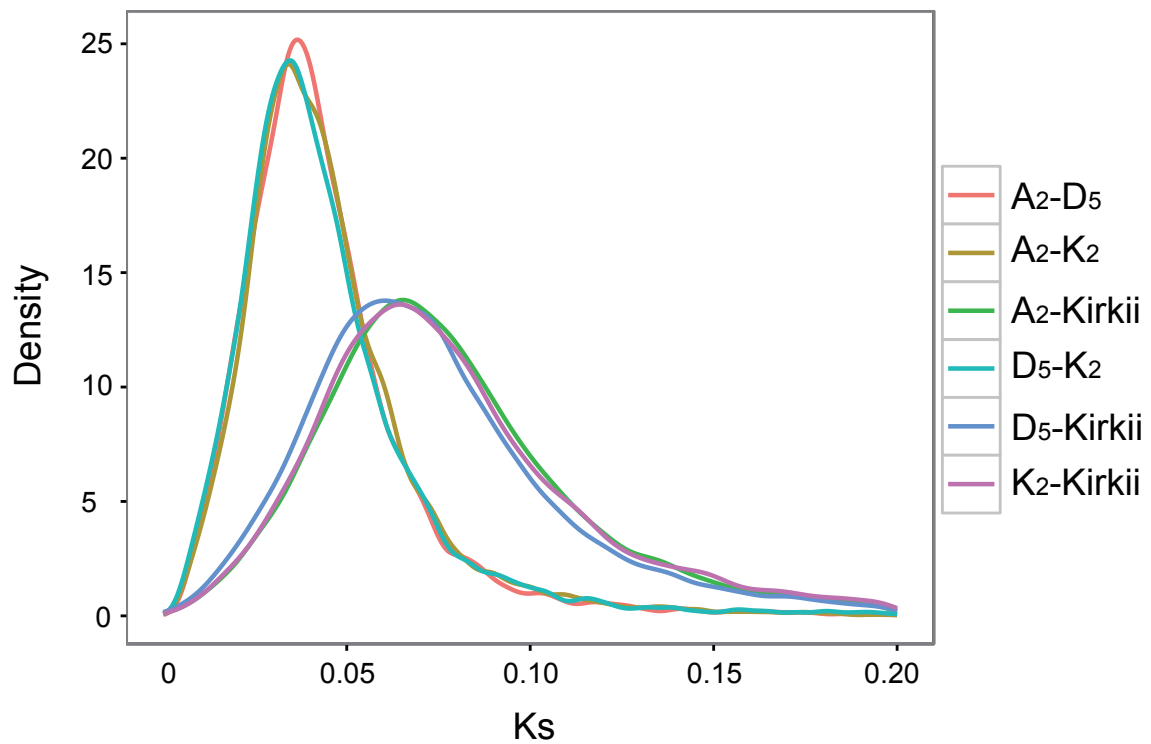
Supplementary Figure 11

Characterization of A_2 -specific rearrangement event (Chr01-Chr02 translocation in other diploid *Gossypium* genomes. (A) The genomic synteny blocks of A_2 and *G. herbaceum* (A_1), *G. thurberi* (D_1), *G. turneri* (D_{10}), and *G. longicalyx* (F_1) were identified using MCScanX. (B) Dot plot showing genome collinearity of Chr01-Chr02 translocation between A_2 and A_1 , D_1 , D_{10} and F_1 . These genomes were downloaded from the CottonGen website (<https://www.cottongen.org/data/download>).



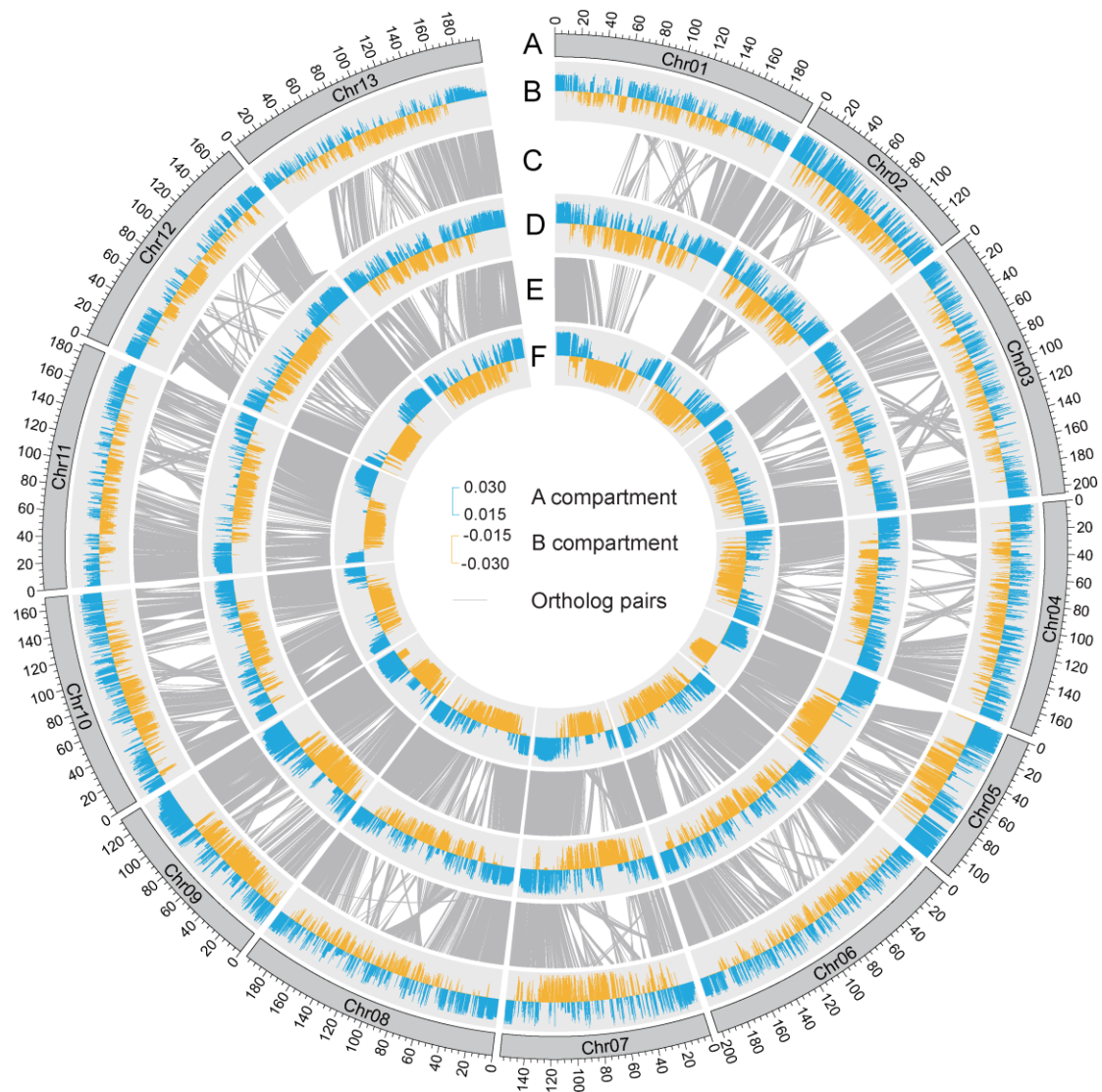
Supplementary Figure 12

Characterization of K_2 -specific rearrangement event (Chr13-Chr05 translocation) in other diploid *Gossypium* genomes. **a**, The genomic synteny blocks of A_2 and A_1 , D_1 , D_{10} , and F_1 were identified using MCScanX. **b**, Dot plot showing genome collinearity of Chr01-Chr02 translocation between A_2 and A_1 , D_1 , D_{10} , and F_1 .



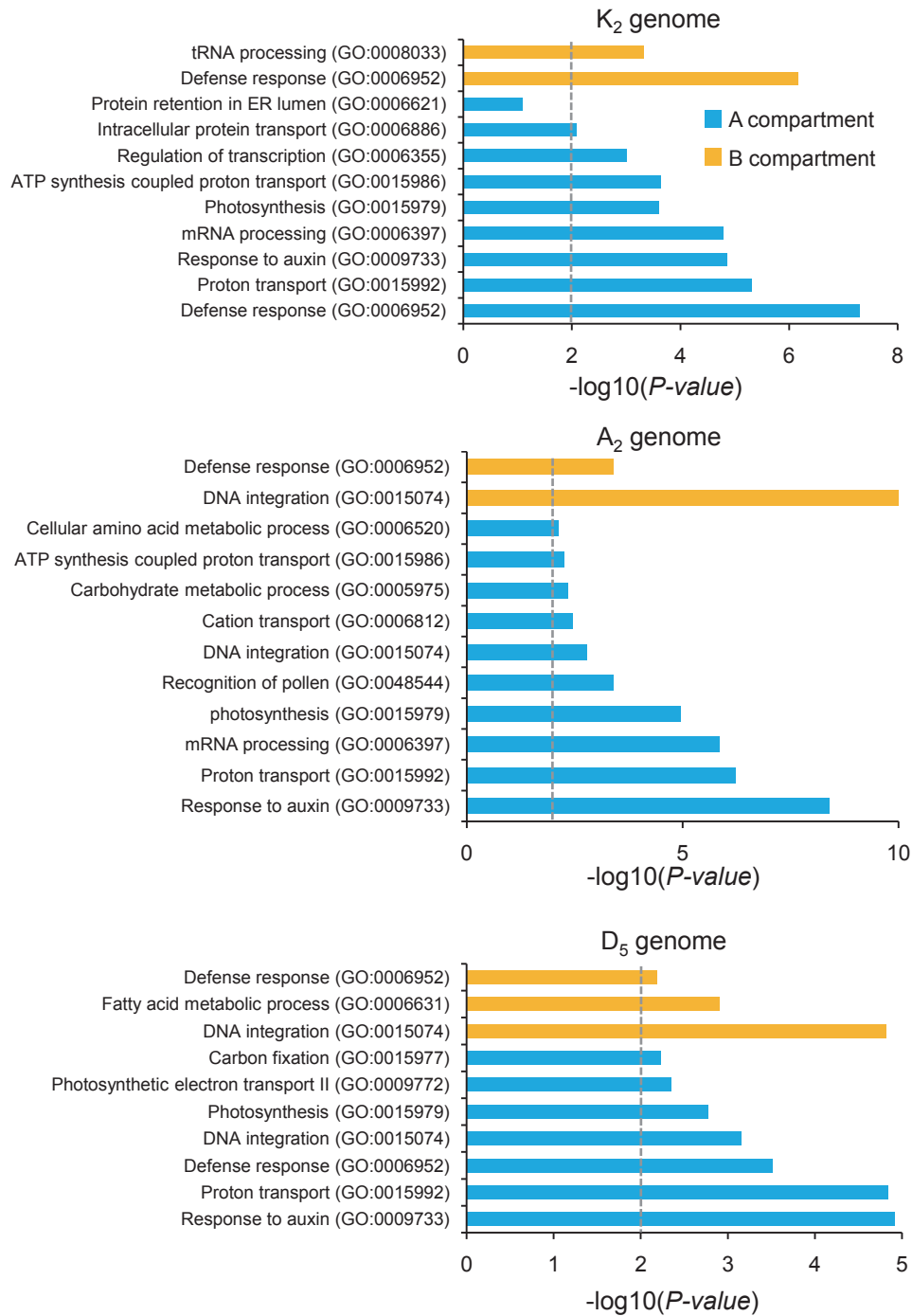
Supplementary Figure 13

Ks distribution of syntenic genes between *Gossypium* and their relative *G. kirkii* species. The peak for *Gossypium* divergence is estimated at ~5 million years ago (MYA), whereas the *Gossypium* and relative species *G. kirkii* divergence time is estimated at 8.5 ~ 10 MYA.



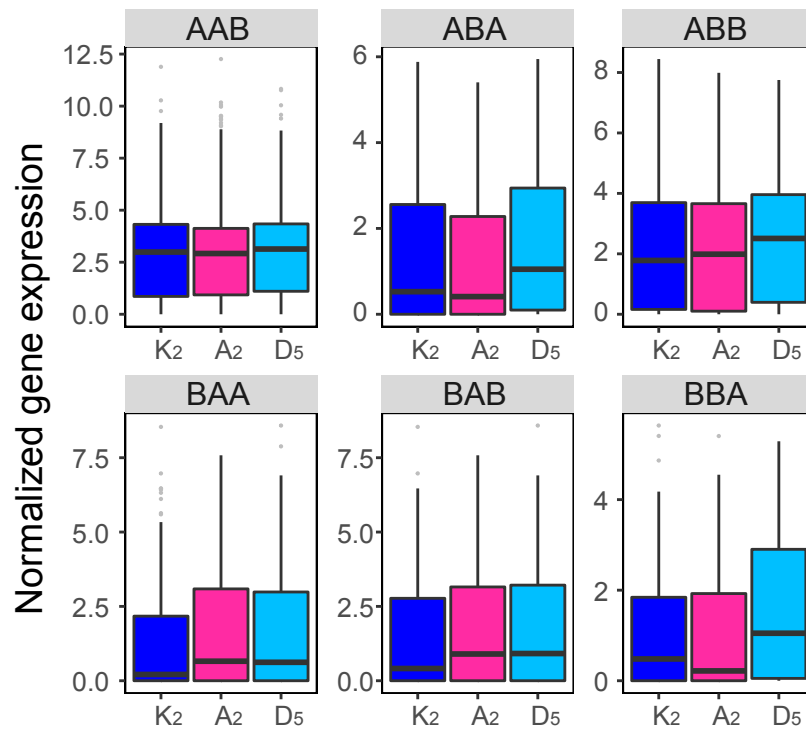
Supplementary Figure 14

The orthologous gene pairs and A/B compartment in *G. rotundifolium* (K₂), *G. arboreum* (A₂) and *G. raimondii* (D₅) genomes. The tracks represent chromosome length of K₂ genome (A), A/B compartment regions in K₂ genome (B), orthologous gene pairs between K₂ and A₂ (C), A/B compartment regions in A₂ (D), orthologous gene pairs between A₂ and D₅ (E), A/B compartment regions in A₂ genome (F). In track A, the chromosome karyotype is shown based on K₂ genome in 1 Mb windows sliding 200 Kb. The each chromosome length of D₅ and A₂ was normalized based on K₂ genome.



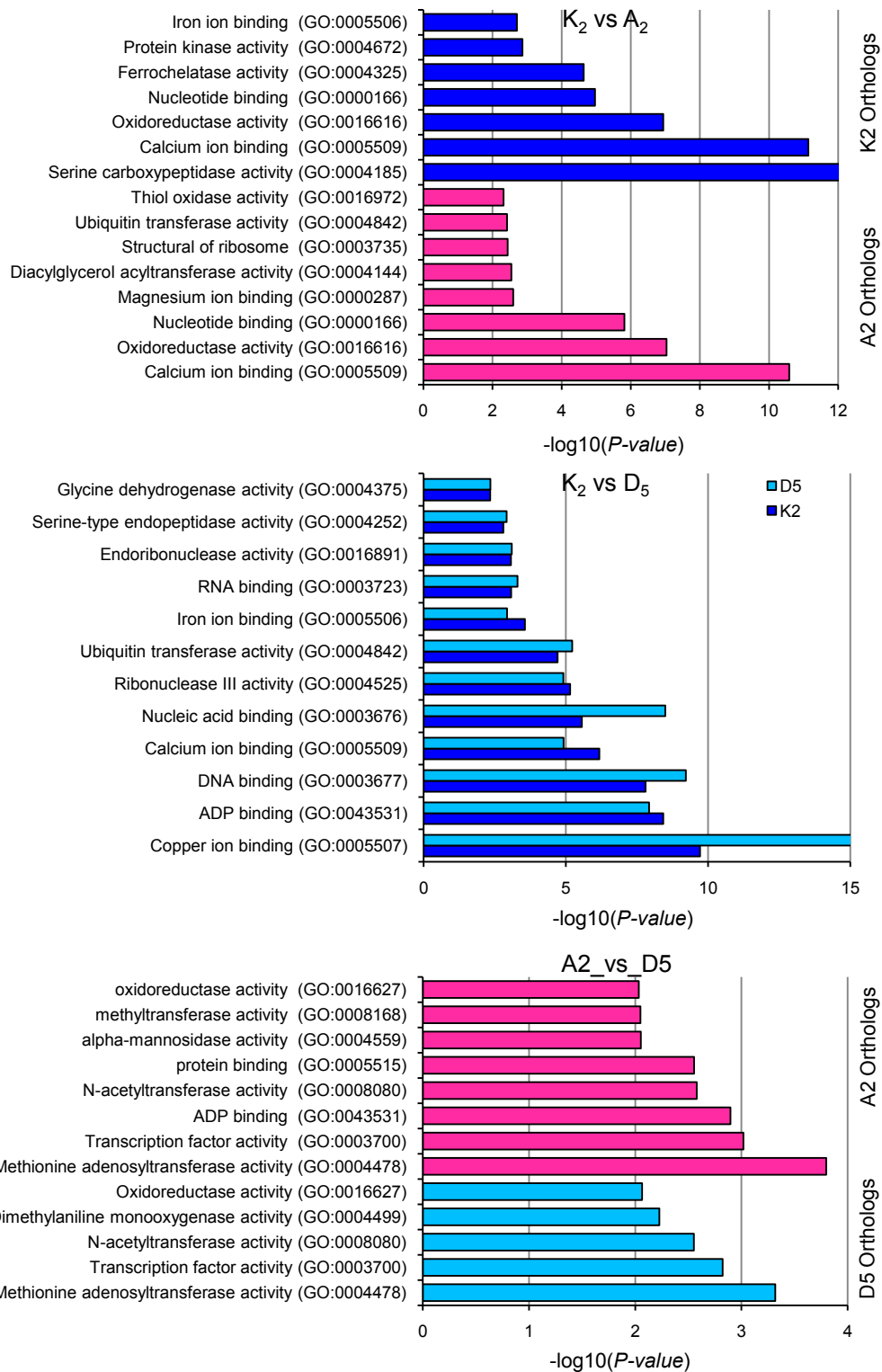
Supplementary Figure 15

GO enrichment analysis of genes in A and B compartments in *G. rotundifolium* (K₂), *G. arboreum* (A₂) and *G. raimondii* (D₅).



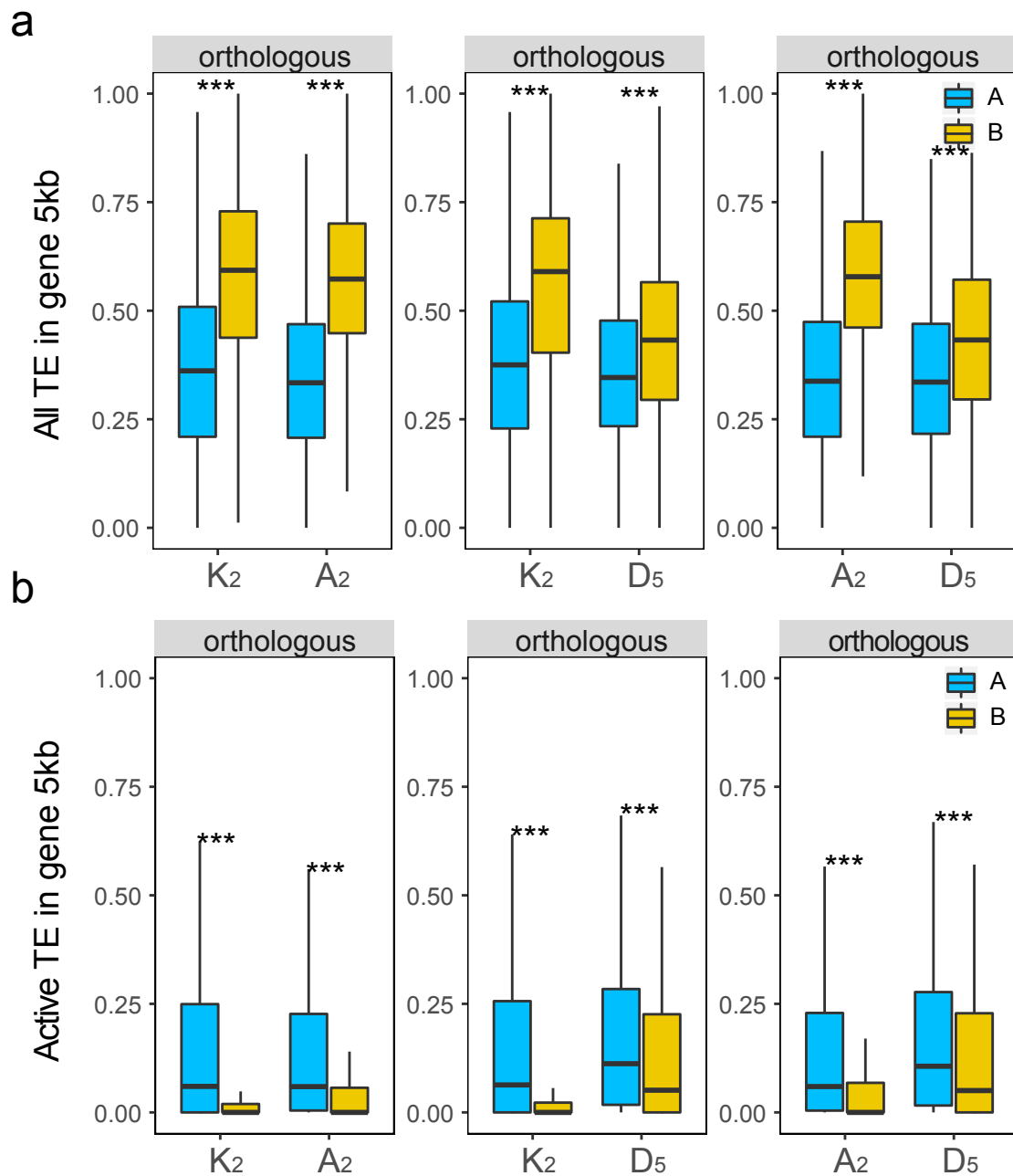
Supplementary Figure 16

Expression levels of orthologous gene pairs of *G. rotundifolium* (K₂), *G. arboreum* (A₂) and *G. raimondii* (D₅) genomes in six groups.



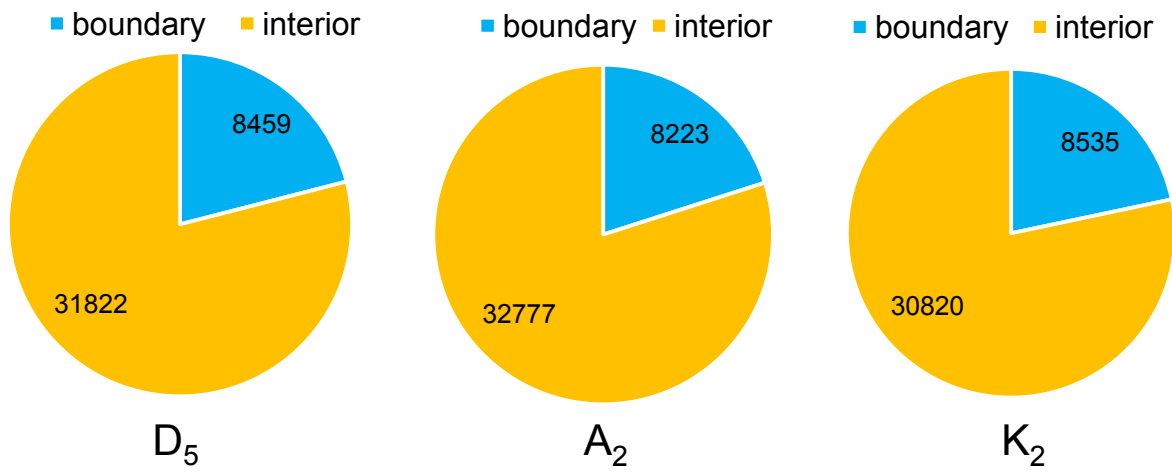
Supplementary Figure 17

GO enrichment analysis of orthologous genes with A/B compartment transition in three comparisons.



Supplementary Figure 18

TE expression residing in orthologous genes and gene flanking 5 Kb regions. All TE (**a**) and active TE (**b**) ratio in orthologous genes and their flanking 5 Kb region.



Supplementary Figure 19

The number of genes in TAD boundaries and interior regions in K_2 , A_2 , and D_5 genomes.

Supplementary Tables

Supplemental Table 1. Summary of Oxford Nanopore sequencing reads.

Supplemental Table 2. Mapping ratio of paired-end sequencing reads to initial contigs.

Supplemental Table 3. Statics of Core Eukaryotic Genes Mapping Approach (CEGMA) and Benchmarking Universal Single-Copy Orthologs (BUSCO) estimating genome assemblies.

Supplemental Table 4. Comparing A₂ genome with previously published genome version.

Supplemental Table 5. Comparing D₅ genome with previously published genome version.

Supplemental Table 6. Gene annotation against annotation database.

Supplemental Table 7. Non-coding RNAs annotation in three reference genomes.

Supplemental Table 8. Identification of centromeres in *G. rotundifolium* (K₂), *G. arboreum* (A₂) and *G. raimondii* (D₅) genomes.

Supplemental Table 9. Repeat sequences annotation in three reference genomes.

Supplemental Table 10. Summary of A/B compartment switching genes in K₂ versus A₂ genome.

Supplemental Table 11. Summary of A/B compartment switching genes in K₂ versus D₅ genome.

Supplemental Table 12. Summary of A/B compartment switching genes in A₂ versus D₅ genome.

Supplemental Table 13. The A-B-A and B-A-B compartment transition genes of *G. rotundifolium* (K₂), *G. arboreum* (A₂) and *G. raimondii* (D₅) genomes.

Supplemental Table 14. Identification of TADs in three genomes.

Supplemental Table 15. Summary of motifs of TAD boundaries identified in three genomes.

References

1. Huang, G. *et al.* Genome sequence of *Gossypium herbaceum* and genome updates of *Gossypium arboreum* and *Gossypium hirsutum* provide insights into cotton A-genome evolution. *Nat. Genet* **52**, 516–524 (2020).
2. Paterson, A.H., *et al.* Repeated polyploidization of *Gossypium* genomes and the evolution of spinnable cotton fibres. *Nature* **492**, 423–427 (2012).
3. Udall, J.A. *et al.* De Novo Genome Sequence Assemblies of *Gossypium raimondii* and *Gossypium turneri*. *G3* (Bethesda) **9**:3079-3085(2019)..
4. Wang, M. *et al.* Reference genome sequences of two cultivated allotetraploid cottons, *Gossypium hirsutum* and *Gossypium barbadense*. *Nat. Genet* **51**, 224–229 (2019).