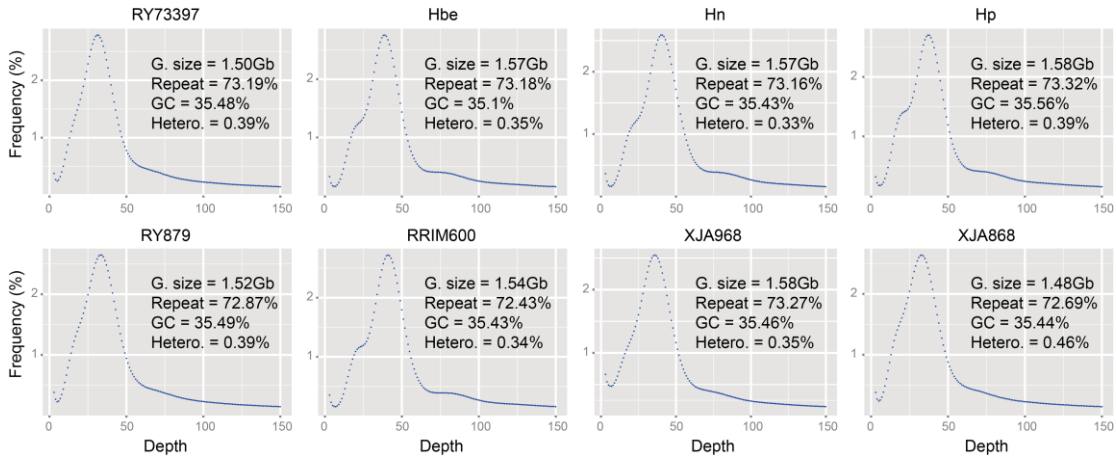
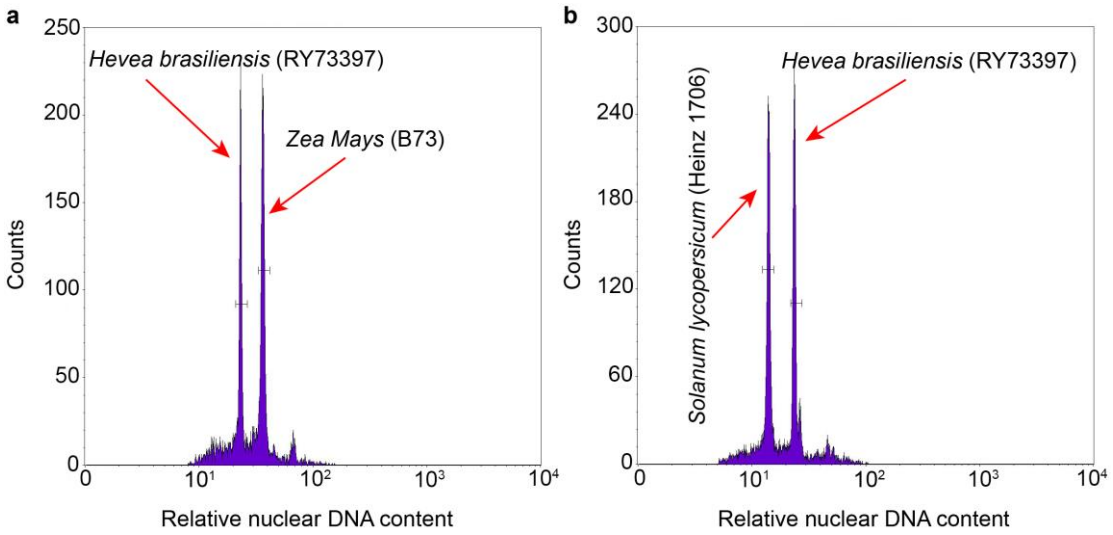
Pan-genome and phylogenomic analyses highlight *Hevea* species delineation and rubber trait evolution

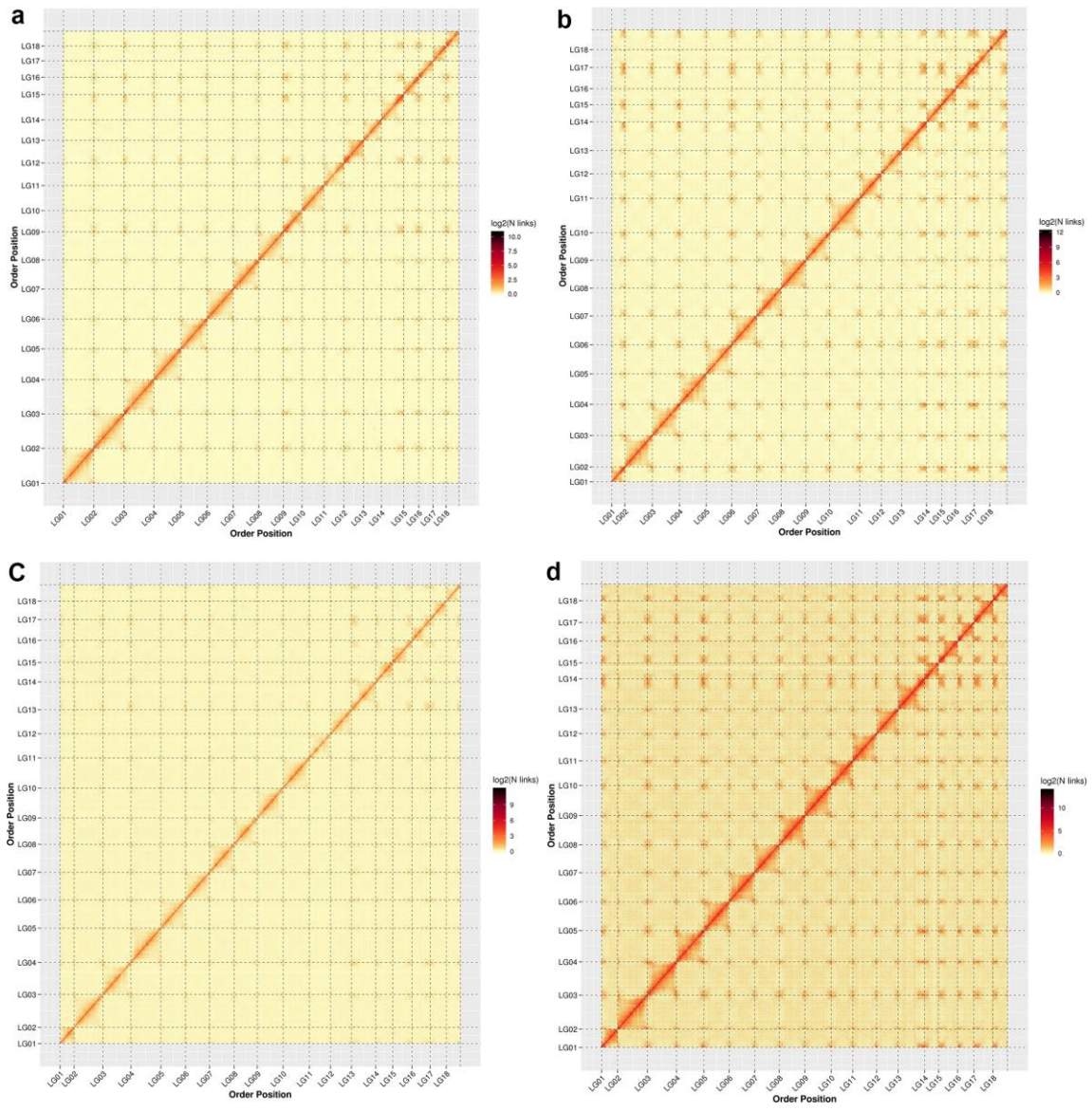
Fang *et al.*



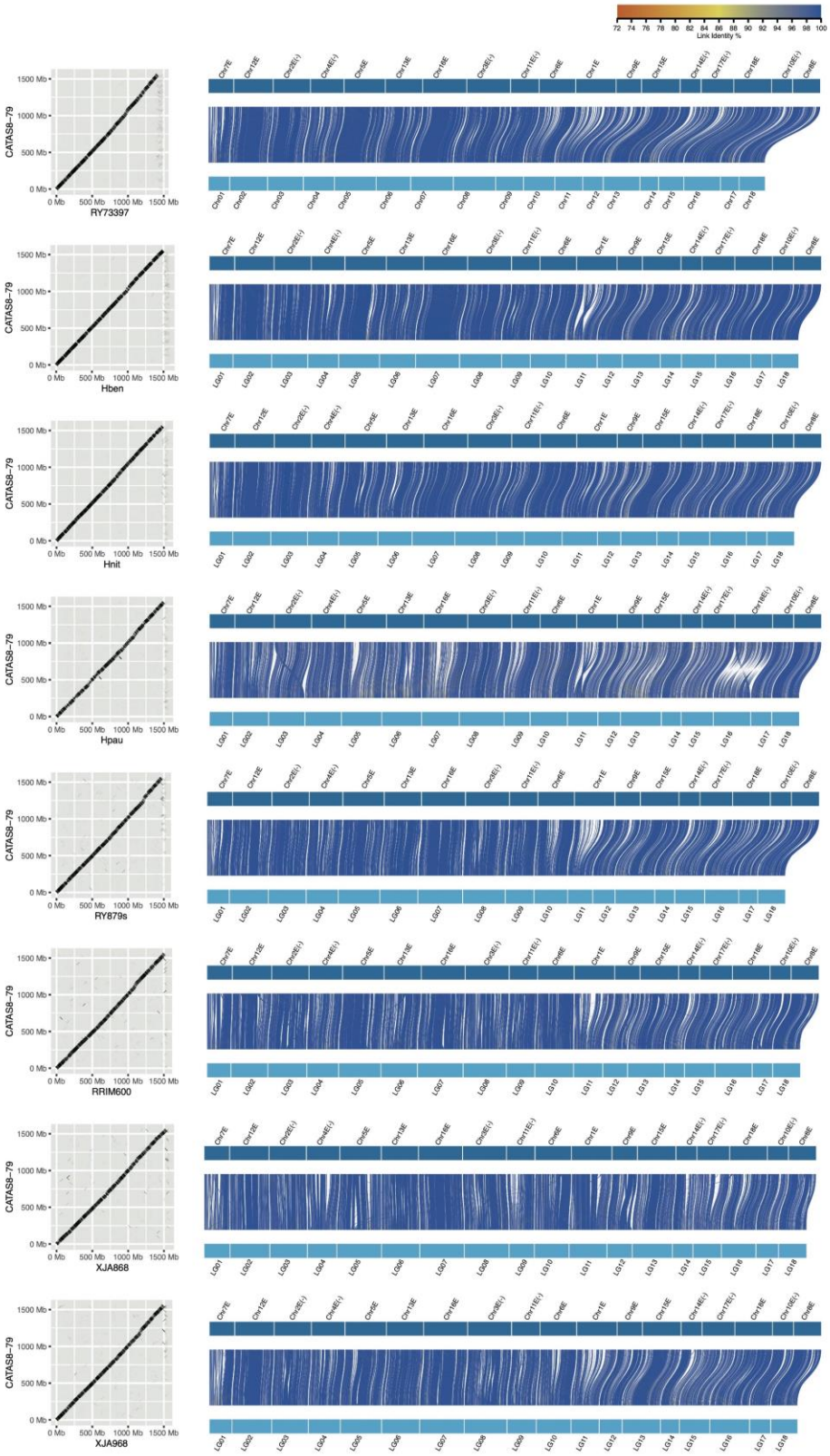
Supplementary Fig. 1. K-mer analysis for estimating genome size, repeat content and heterozygosity of eight *Hevea* accessions. Based on Hiseq reads of the 180-bp small-insert libraries and using the distribution of 17-mers, the eight *Hevea* accessions were found to have genome sizes (G. sizes) ranging from 1.48 to 1.58 Gb, GC contents from 35.1% to 35.6%, heterozygosity (Hetero.) levels from 0.33 to 0.46, and repetitive sequence contents of approximately 73%. Hbe, *H. benthamiana*; Hn, *H. nitida*; Hp, *H. pauciflora*; RY73397, RY879 and RRIM600, three *H. brasiliensis* cultivars; XJA968 and XJA868, two *H. brasiliensis* wild germplasms.



Supplementary Fig. 2. Estimate of *H. brasiliensis* RY73397 genome size by flow cytometry. The genome size of RY73397 was estimated to be 1.49 and 1.50 Gb, respectively, when the genomes of *Zea mays* (B73; 2.3 Gb) (**a**) and *Solanum lycopersicum* (Heinz1706; 0.88 Gb) (**b**) were used as internal standards. Four-month-old tissue-cultured RY73397 plantlets were used for flow cytometry in the Molecular Biology Experiment Center, Germplasm Bank of Wild Species in Southwest China, Kunming Institute of Botany, Chinese Academy of Sciences. Source data are provide as a Source Data file.

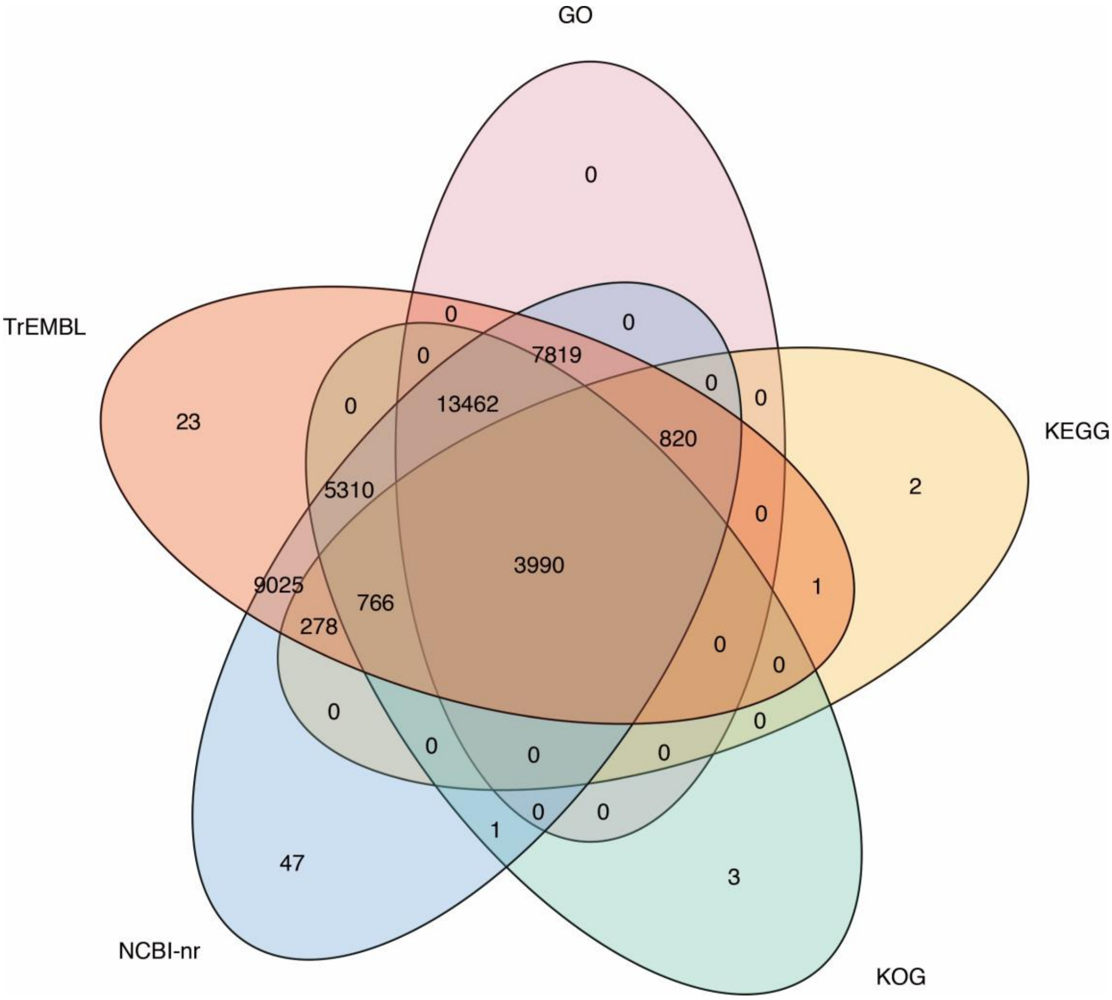


Supplementary Fig. 3. Genome-wide all-by-all high-throughput chromosome conformation capture (Hi-C) interaction heatmap. The heat map was drawn for the 18 linkage groups in *Hevea brasiliensis* RY73397 (a), *H. pauciflora* (b), *H. benthamiana* (c) and *H. nitida* (d), showing the density of the Hi-C interactions between chromosomes. The genomes were partitioned into bins of 100 Kb, and the number of Hi-C read pairs covering each pair of bins was taken as the strength of interaction. The 18 chromosome groups are clearly distinguished. Within each chromosome group, the interaction strength at diagonal positions is much higher than that at non-diagonal positions, indicating a higher interaction between adjacent sequences (diagonal positions) than between non-adjacent sequences (non-diagonal positions).

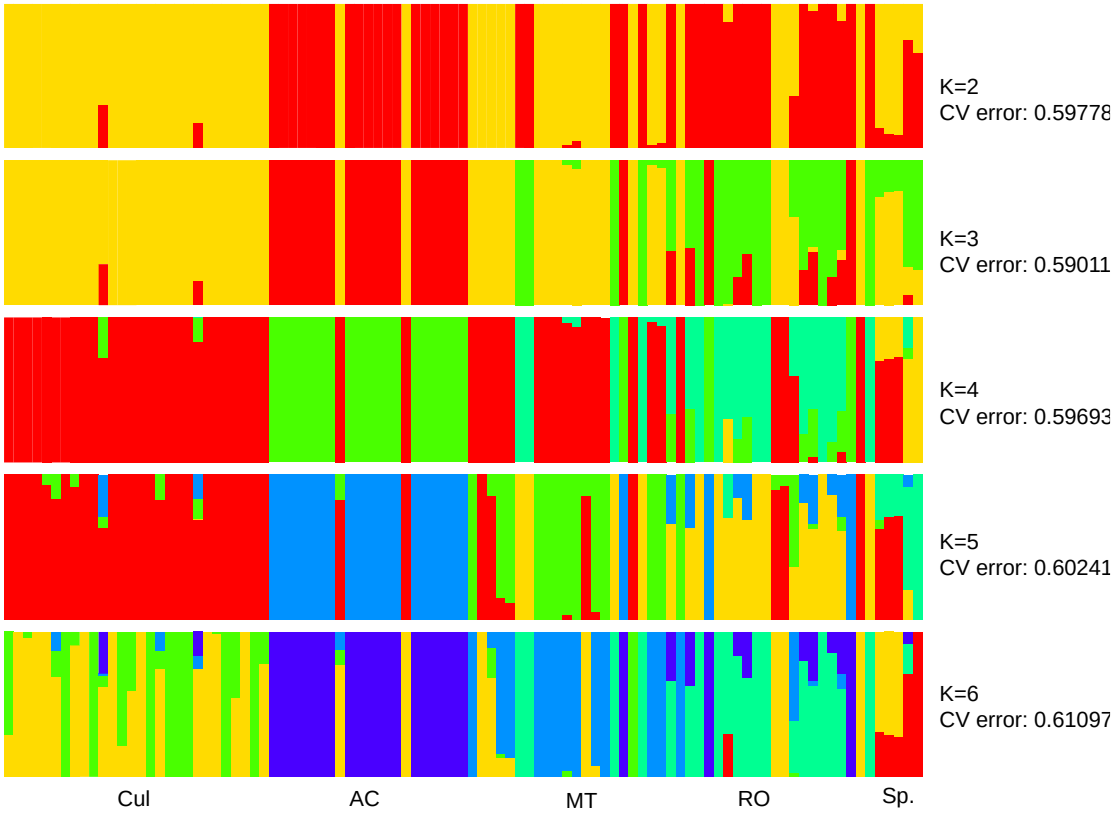


Supplementary Fig. 4. Collinearity between our eight *Hevea* genomes and the latest *H. brasliensis* CATAS8-79 genome[1]. Our eight *Hevea* genome assemblies are three *H. brasiliensis* cultivars of RY879, RY73397 and RRIM600, two wild

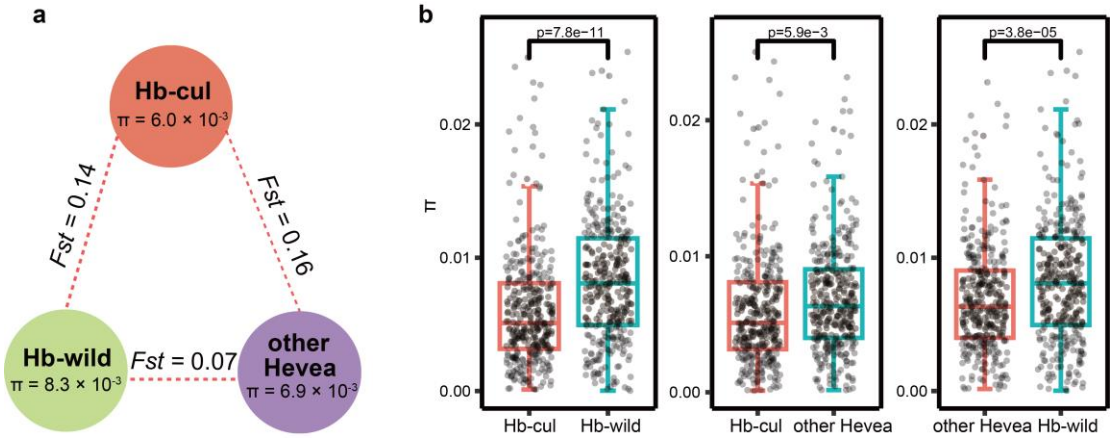
germplasms of XJA868 and XJA968, and three other *Hevea* species of Hbe, Hn and Hp. The eight genome assemblies were aligned, respectively, with the CATAS8-79 genome using the minimap2 algorithm with default parameters. The comparative visualizations were generated using the 'pafr' R package and AliTV tools with length cuts of 20 kb and 200 kb, respectively.



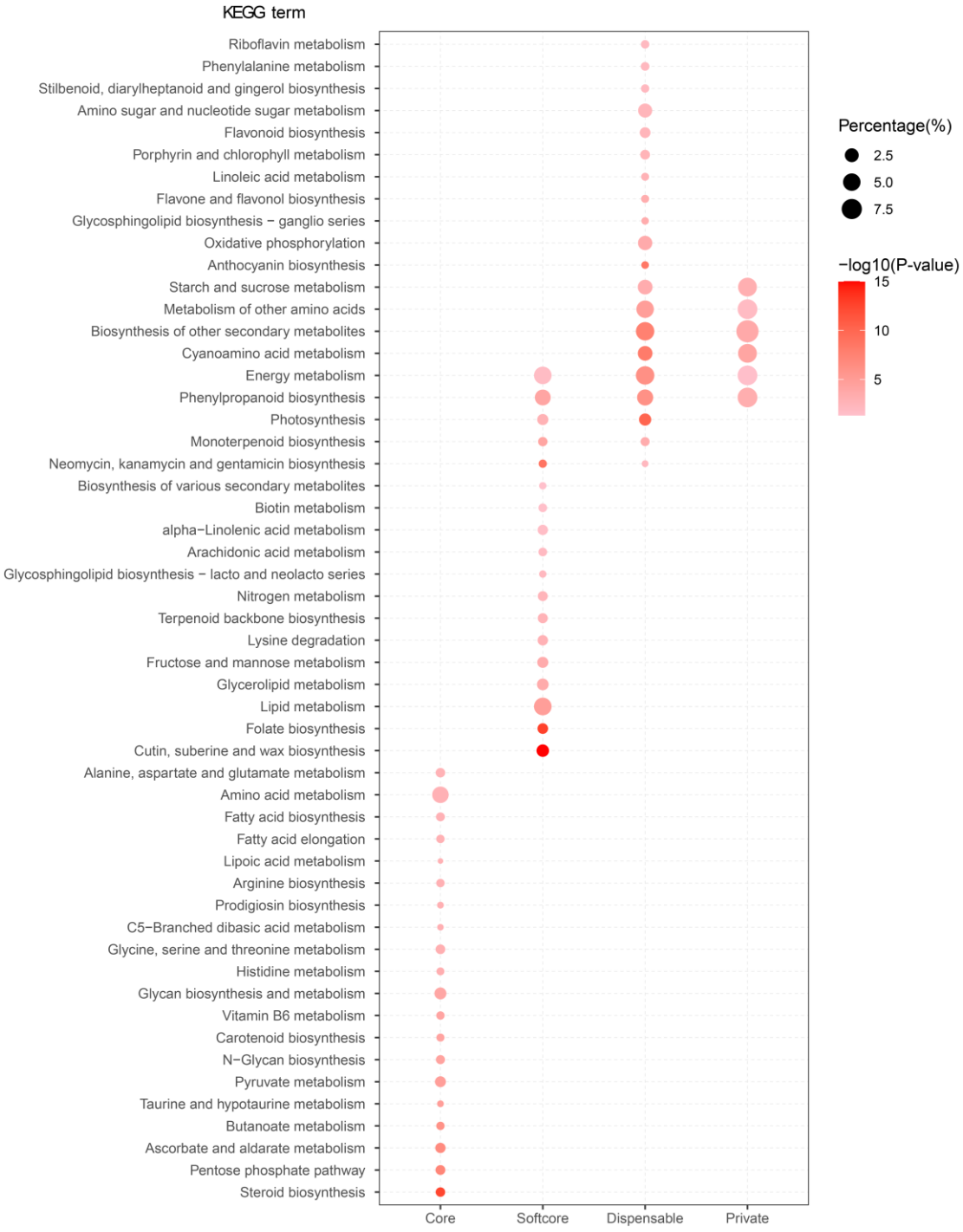
Supplementary Fig. 5. Venn diagram for functional annotation of protein-coding genes in the RY73397 genome. All 42,897 protein-coding genes were subjected to BLAST searches against the five functional databases of NCBI-nr, KOG, GO, KEGG, and TrEMBL, with a threshold set at e-value < 1e-5, for gene functional annotation. Ultimately, 41,546 genes, over 97% of the total, were annotated. The Venn diagram illustrates the annotation status of genes across the five databases. Source data are provided as a Source Data file.



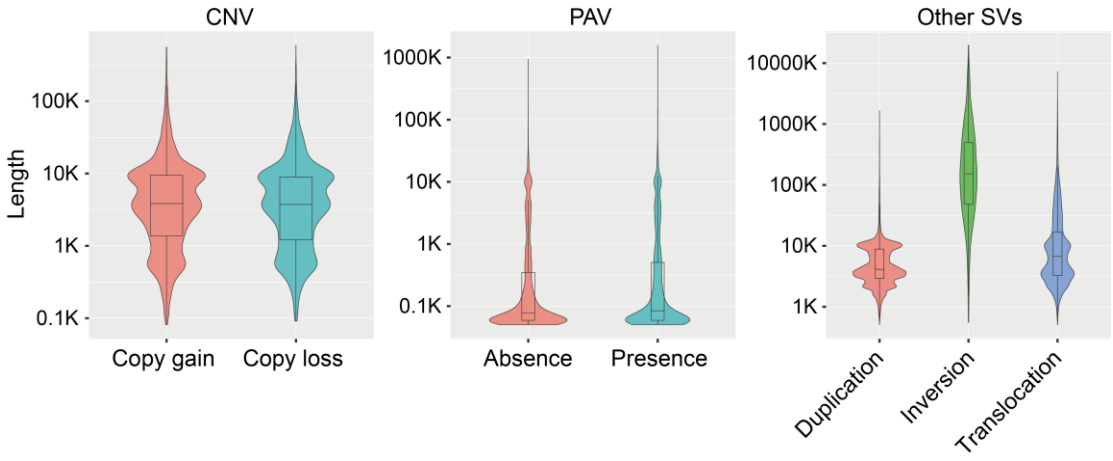
Supplementary Fig. 6. Population structure of 94 *Hevea* accessions inferred by Admixture. The length of each colored segment represents the proportion of each genome inferred from ancestral populations (K = 2 to 6). The 94 accessions are: 25 *H. brasiliensis* cultivars (Cul), 62 *H. brasiliensis* wild germplasms from three Brazilian states (21 AC, 21 MT and 20 RO), and 7 other *Hevea* species (*H. benthamiana*, e; *H. nitida*, n; *H. nitida* Mart var., v; *H. pauciflora*, p; *H. spruceana*, s) as in Supplementary Data 2. AC, MT and RO, the wild accessions from the Brazilian states of Acre, Mato Grosso and Rondônia, respectively; Sp., other *Hevea* species; CV error, cross-validation error.



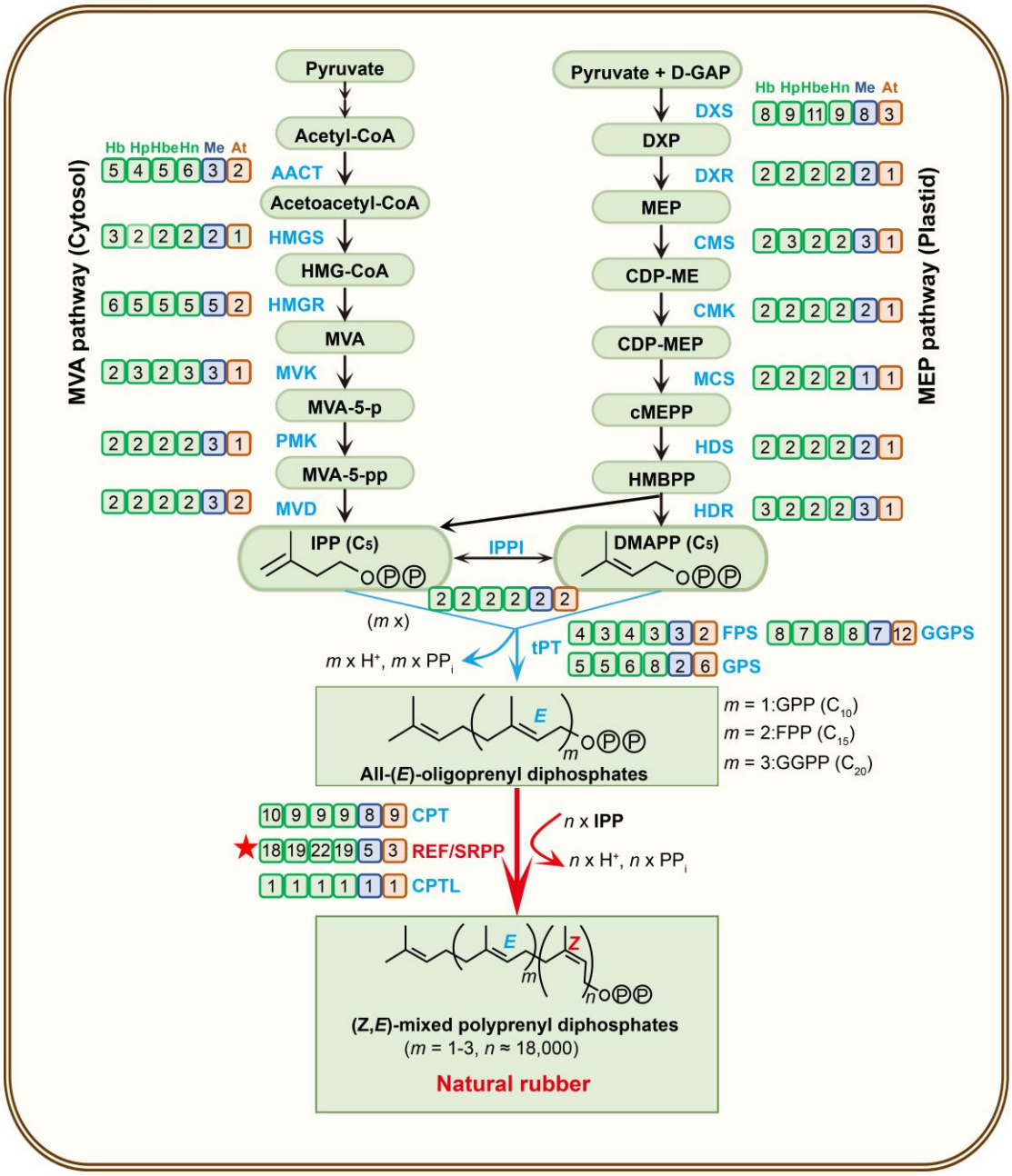
Supplementary Fig. 7. π values and t-test between different groups of *Hevea* germplasms. Based on nuclear SNPs, population genetic diversity (π) (a), and the significance levels (t-test, *p* values) of population differences (b) were calculated. Hb-cul, *H. brasiliensis* cultivars; Hb-wild, *H. brasiliensis* wild germplasms; other *Hevea* species, four species (Hbe, Hn, Hp, Hs) and three Hn variants (HnV5, 7, 8). See details for these germplasms in Supplementary Data 2.



Supplementary Fig. 8. KEGG pathway analyses of core, softcore, dispensable and private genes. The top 20 most enriched pathways under the class of metabolism shown for the four types of gene families. The enriched pathways have overlaps among the three non-core composition genes, i.e., softcore, dispensable and private, but do not have overlaps between these non-core genes with the core genes. Source data are provided as a Source Data file.

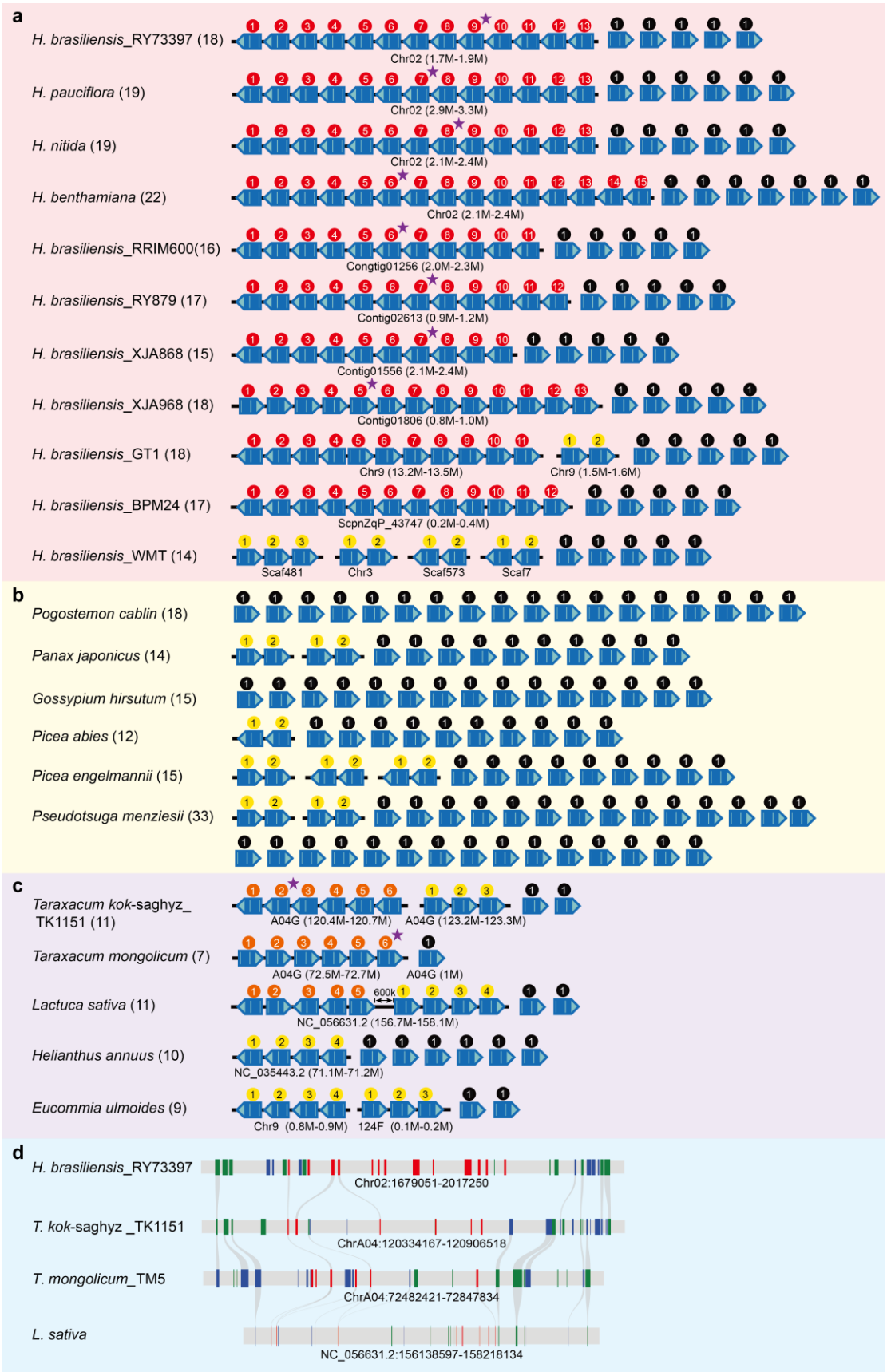


Supplementary Fig. 9. Characterization of large structure variations in *Hevea* genomes. We conducted whole-genome alignment using MUMmer and subsequently employed SyRI (Synteny and Rearrangement Identifier) to detect large structure variations (SVs) including copy number variation (CNV) and presence/absence variation (PAV), and analyzed their length ranges in Kb in the three non-reference genomes, i.e., Hbe, Hp and Hn. Source data are provided as a Source Data file.



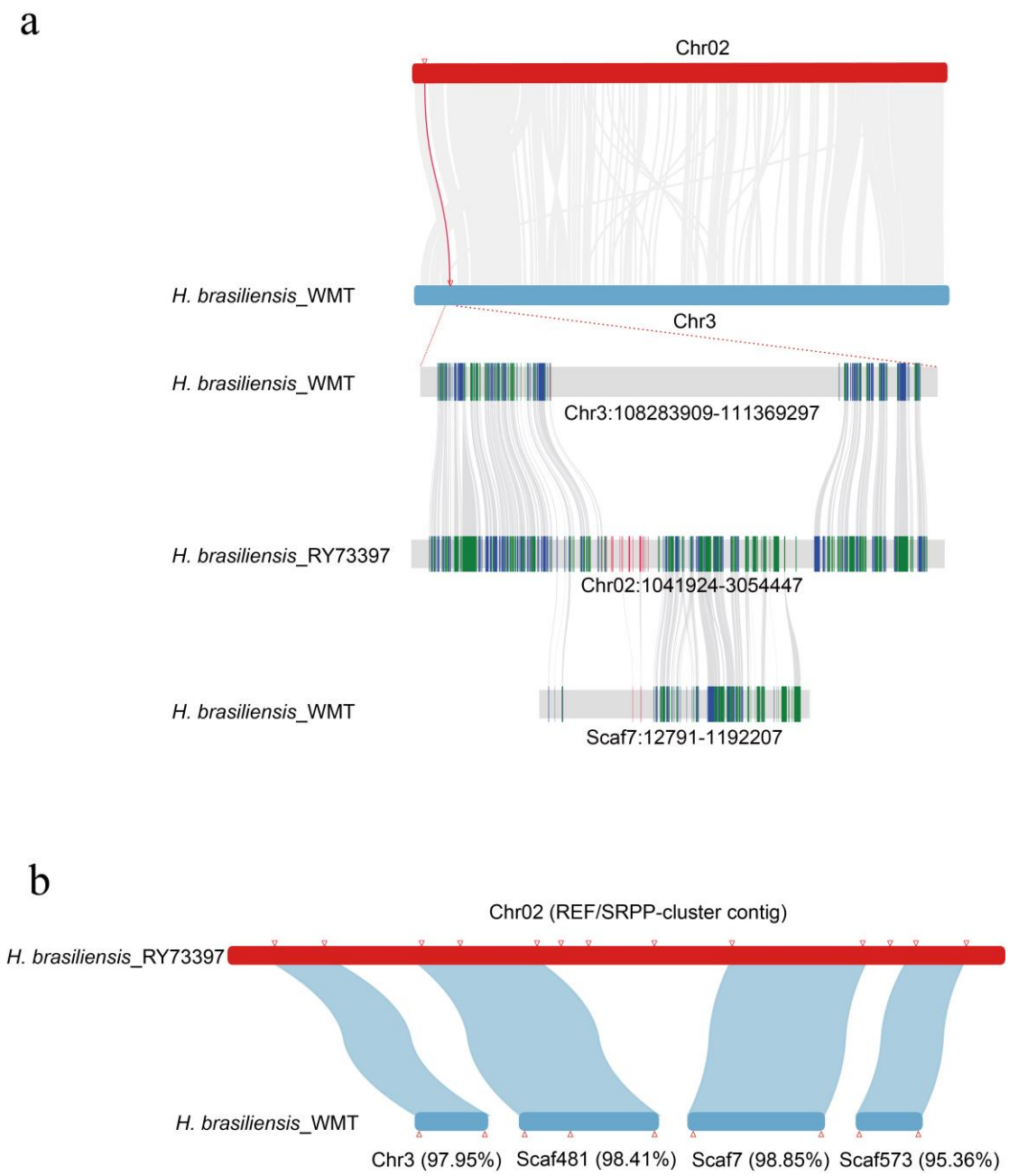
Supplementary Fig. 10. Diagram of the MVA and MEP rubber biosynthesis pathway genes in four *Hevea* species (Hb, Hp, Hbe and Hn) compared to cassava (Me) and Arabidopsis (At). The numbers of encoding genes are shown in squares under respective plant species for all the enzymes or proteins so far reported in the rubber biosynthesis pathway. Enzymes or proteins: AACT, acetyl-CoA acetyltransferase; HMGS, HMG-CoA synthase; HMGR, HMG-CoA reductase; MVK, mevalonate kinase; PMK, phosphomevalonate kinase; MVD, mevalonate diphosphate decarboxylase; IPPI, IPP isomerase; DXS, 1-deoxy-D-xylulose-5-phosphate synthase; DXR, 1-deoxy-D-xylulose-5-phosphate synthase reductoisomerase; CMS, 2-C-

methyl-D-erythritol 4-phosphate cytidylytransferase; CMK, 4-(cytidine 5'-diphospho)-2-C-methyl-D-erythritol kinase; MCS, 2-C-methyl-D-erythritol 2,4-cyclodiphosphate synthase; HDS, 4-hydroxy-3-methylbut-2-en-1-yl diphosphate synthase; HDR, 4-hydroxy-3-methylbut-2-en-1-yl diphosphate reductase; IPPI, IPP isomerase; tPT, trans-prenyltransferase; cPT, cis-prenyltransferase. Metabolites: HMG-CoA, 3-hydroxy-3-methylglutaryl-CoA; MVA, mevalonate; IPP, isopentenyl diphosphate; DMAPP, dimethylallyl diphosphate; D-GAP, D-glyceraldehyde 3-phosphate; DXP, 1-deoxy-D-xylulose; MEP, 2-C-methyl-D-erythritol 4-phosphate; CDP-ME, 4-(cytidine 5'-diphospho)-2-C-methyl-D-erythritol; CDP-MEP, 2-phospho-4-(cytidine 5'-diphospho)-2-Cmethyl-D-erythritol; cMEPP, 2-C-methyl-D-erythritol 2,4-cyclodiphosphate; HMBPP, 1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate.



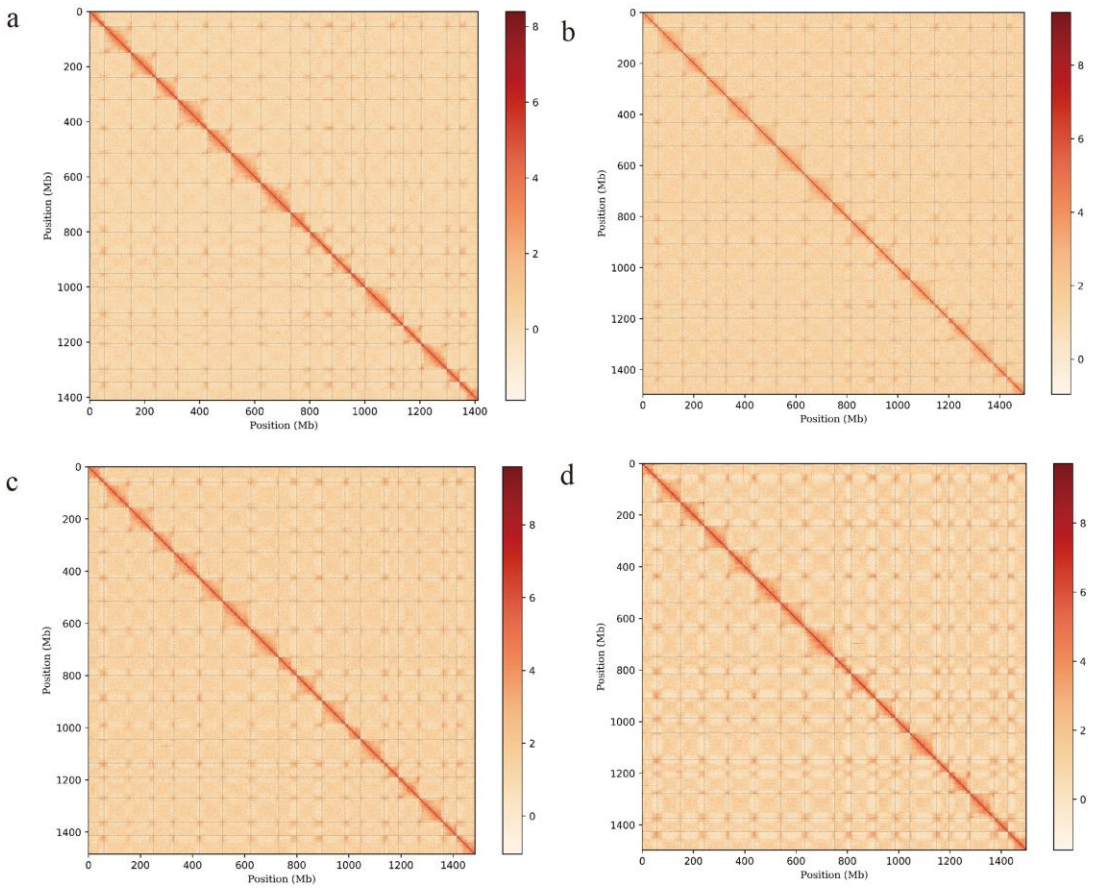
Supplementary Fig. 11. Genomic distribution of the *REF/SRPP* genes in some

selected plant genomes. a, Genomic distribution of the *REF/SRPP* genes in eleven *Hevea* genomes as described in Fig. 4a, including five cultivars and three wild accessions of *H. brasiliensis*, and three other *Hevea* species (*H. pauciflora*, *H. nitida*, *H. benthamiana*). **b**, Genomic distribution of the *REF/SRPP* genes in six species with copy number comparable to that of the *Hevea* genomes. The six species, with *REF/SRPP* copy number of twelve and above, are among the 141 non-*Hevea* genomes as described in Fig. 4a. **c**, Genomic distribution of the *REF/SRPP* genes in five other rubber-producing plant species. In **a** to **c**, the copy number of *REF/SRPP* genes is shown in the brackets beside the species. The *REF/SRPP* genes in a cluster with copies of 10 and above are highlighted in red background, 5 to 9 copies in orange background, 2 to 4 copies in yellow, and scattered genes in black. Expressions of the *REF/SRPP* genes were examined in the latex where RNAseq data were accessible, and the most latex-predominant isoforms are highlighted by a purple asterisk (★). **d**, Collinearity plots of the *REF/SRPP* cluster locus between *H. brasiliensis* RY73397 and three other rubber-producing plant species, *T. kok-saghyz* and *T. mongolicum*[45] and *Lactuca sativa*[50].

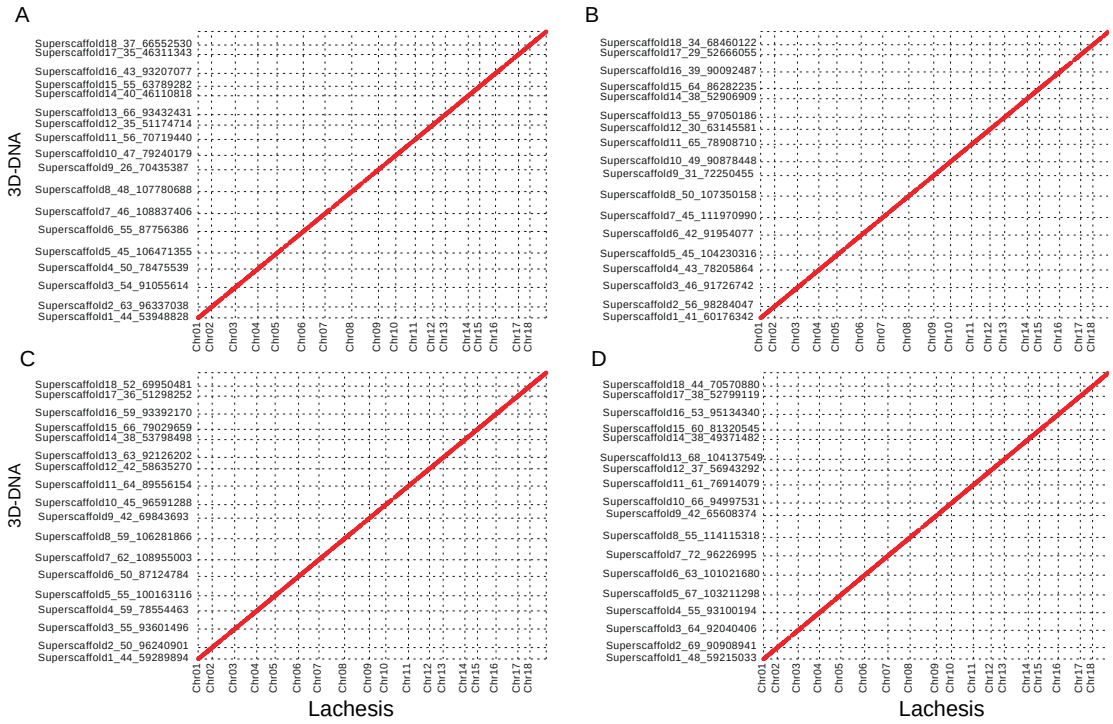


Supplementary Fig. 12. Collinearity between the *REF/SRPP* cluster-harboring chromosomes of RY73397 and WMT. The gene collinearity plots were drawn for the *REF/SRPP* cluster-harboring chromosome (chr02) of *H. brasiliensis* RY73397 and the corresponding chromosome (chr3) in the wild *H. brasiliensis* accession (WMT). The close-up views are shown for the collinearity between the genomic region flanking the *REF/SRPP* cluster in RY73397 and the counterpart in WMT (red triangles) as well as a non-chromosome anchored scaffold of WMT. Red boxes indicate the *REF/SRPP*

genes. Green and blue boxes denote other genes. Grey lines identify syntenic gene pairs. (b) Three scaffolds (not anchored to the 18 chromosomes) and one Chr3 fragment from WMT (Supplementary Fig. 11a), each with 2-3 *REF/SRPP* genes, displayed 95.4%-98.9% collinearity with the large *REF/SRPP*-cluster containing contig from the Chr02 of RY73397. The approximate positions of *REF/SRPP* genes are indicated by red triangles, and the collinearity percentages are shown in brackets after the names of scaffolds and Chr3 fragment.



Supplementary Fig. 13. Heatmap of 3D-DNA-assisted Hi-C assemblies for four *Hevea* genomes. The heat map was drawn for the 18 linkage groups in *Hevea brasiliensis* RY73397 (a), *H. benthamiana* (b), *H. nitida* (c), and *H. pauciflora* (d), showing the density of the Hi-C interactions between chromosomes.



Supplementary Fig. 14. Collinearity comparison between 3D-DNA- and Lachesis-assisted *Hevea* genome assemblies. Collinearity comparisons were conducted using the nucmer (Ver 4.0.0rc1) with the parameters of --mum -l 10000 -c 10000 for the genome assemblies of *Hevea brasiliensis* RY73397 (a), *H. benthamiana* (b), *H. nitida* (c), and *H. pauciflora* (d).

Supplementary reference

1. Chao, J. et al. Genomic insight into domestication of rubber tree. *Nat. Commun.* **14**, 4651 (2023).