The haplotype-resolved telomere-to-telomere genome and OMICS analyses reveal the genetic responses of tapping in rubber tree

Han Cheng

forcheng@gmail.com

Rubber Research Institute, Chinese Academy of Tropical Agricultural Sciences (CATAS)

Chaochao Li

Chinese Academy of Tropical Agricultural Sciences (CATAS)

Yuan Yuan

Rubber Research Institute, Chinese Academy of Tropical Agricultural Sciences (CATAS)

Zhiyuan Wang

Rubber Research Institute, Chinese Academy of Tropical Agricultural Sciences (CATAS)

Zhiyi Nie

Rubber Research Institute, Chinese Academy of Tropical Agricultural Sciences (CATAS)

Tingkai Wu

Rubber Research Institute, Chinese Academy of Tropical Agricultural Sciences (CATAS)

Jiangxia Qiao

Agricultural Genomics Institute at Shenzhen, Chinese Academy of Agricultural Sciences

Zhi Deng

Rubber Research Institute, Chinese Academy of Tropical Agricultural Sciences (CATAS)

Xiaobo Wang

China Agricultural University

Dong Xu

Rubber Research Institute, Chinese Academy of Tropical Agricultural Sciences (CATAS)

Xu Wang

Agricultural Genomics Institute at Shenzhen, Chinese Academy of Agricultural Sciences

Shuo Cao

Chinese Academy of Agricultural Sciences

Bingqin Li

Rubber Research Institute, Chinese Academy of Tropical Agricultural Sciences (CATAS)

Zewei An

Rubber Research Institute, Chinese Academy of Tropical Agricultural Sciences (CATAS)

https://orcid.org/0000-0002-8633-5332

Wenguan Wu
Rubber Research Institute, Chinese Academy of Tropical Agricultural Sciences (CATAS)
Zhongxin Jin
Agricultural Genomics Institute at Shenzhen, Chinese Academy of Agricultural Sciences
Huasun Huang
Rubber Research Institute, Chinese Academy of Tropical Agricultural Sciences (CATAS)
Wei Hu
Institute of Tropical Bioscience and Biotechnology, Chinese Academy of Tropical Agricultural Sciences
https://orcid.org/0000-0002-6796-8046
Sanwen Huang
Agricultural Genomics Institute at Shenzhen https://orcid.org/0000-0002-8547-5309
Yongfeng Zhou
Agricultural Genomics Institute at Shenzhen, Chinese Academy of Agricultural Sciences
https://orcid.org/0000-0003-0780-2973

Article

Keywords: Hevea brasiliensis, natural rubber, T2T genome, rubber biosynthesis, rubber tapping

Posted Date: August 16th, 2024

DOI: https://doi.org/10.21203/rs.3.rs-4798254/v1

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

The haplotype-resolved telomere-to-telomere genome and OMICS analyses reveal the genetic responses of tapping in rubber tree

Chaochao Li[#1,2,3], Yuan Yuan[#1,3], Zhiyuan Wang[#1,2,3], Zhiyi Nie[#1,3,5], Tingkai Wu[1,3], Jiangxia Qiao[2], Zhi Deng[1,3,4], Xiaobo Wang[1,3,4], Dong Xu[1,3,4], Xu Wang[2], Shuo Cao[2], Bingqin Li[1,3], Zewei An[1,3,4], Wenguan Wu[1,3,4], Zhongxin Jin[1,2,3], Huasun Huang[1,3,4], Wei Hu[3,6], Sanwen Huang[1,2], Yongfeng Zhou[*2,5], Cheng Han[*1,3,4]

[1] National Key Laboratory for Tropical Crop Breeding, Ministry of Agriculture and Rural Affairs Key Laboratory of Biology and Genetic Resources of Rubber Tree, Rubber Research Institute, Chinese Academy of Tropical Agricultural Sciences, Haikou, China

[2] National Key Laboratory for Tropical Crop Breeding, Shenzhen Branch, Guangdong Laboratory of Lingnan Modern Agriculture, Key Laboratory of Synthetic Biology, Ministry of Agriculture and Rural Affairs, Agricultural Genomics Institute at Shenzhen, Chinese Academy of Agricultural Sciences, Shenzhen, China

[3] Sanya Research Institute, Chinese Academy of Tropical Agricultural Sciences, Sanya, China

[4] State Key Laboratory Incubation Base for Cultivation and Physiology of Tropical Crops, Rubber Research Institute, Haikou, China

5 National Key Laboratory of Tropical Crop Breeding, Tropical Crops Genetic Resources Institute, Chinese Academy of Tropical Agricultural Sciences, Haikou, China

[6] National Key Laboratory for Tropical Crop Breeding, Key Laboratory of Biology and Genetic Resources of Tropical Crops, Institute of Tropical Bioscience and Biotechnology, Sanya Research Institute of Chinese Academy of Tropical Agricultural Sciences, Haikou, China

[#] These authors contributed equally to this work.

[*] Corresponding author:

forcheng@gmail.com

zhouyongfeng@caas.cn

Abstract

Rubber tree (*Hevea brasiliensis*), as the primary source of natural rubber (NR), holds a significant economic significance. The high-quality genome has been long pursued to understand the rubber production process, genomic structure, and genomics-assisted breeding. Here, we assemble the first haplotype-resolved telomere-to-telomere (T2T), gap-free reference genome for rubber tree (CATAS 7-33-97). Two haplotypes (both 1.56 Gb) reveal all telomere and most centromeric regions. A dramatic amount of variations exist between the two haplotypes, including a 32.71 Mb inversion (*sv33M*) on chromosome 8. Complete assemblies of all 36 chromosomes enabled the exhaustive identification of rubber biosynthesis genes and the revealing of consistent allele specific expression (ASE) profiles. The reconstruction of the natural rubber biosynthesis pathway through transcriptomic and metabolomic profiling unraveled that mevalonate and its derivatives serve as the major carbon reservoir for quick latex replenishment during tapping. Jasmonic acid (JA) was crucial for the consecutive tapping-dependent rubber yield increment by responding to mechanic injuries, and by elevating rubber biosynthesis activity. Finally, we proposed a model of rubber tree's response to tapping, in which JA levels increased after mechanical damage. Subsequently, MYC2 in the JA signaling pathway stimulated the expression of MVK1 gene and the synthesis of mevalonic acid (MVA), enhancing rubber biosynthesis. The assembly of the haplotype-resolved T2T genome is a major step forward to understanding the complexity of the rubber biosynthesis mechanism in latex-producing plants, accelerating rubber tree genetic improvement.

Keywords: *Hevea brasiliensis*, natural rubber; T2T genome; rubber biosynthesis; rubber tapping

Introduction

The rubber tree (*Hevea brasiliensis*) is the source of 99% of natural rubber (NR) globally, and its distinctive physical properties play pivotal roles across various industrial sectors[1]. Rubber plantation is the major income source for the farmers in many tropical countries[2]. It was first introduced to Southeast Asia in 1876 by Wickham, therefore, the descendants were called Wickham germplasm[3]. Rubber breeding was first started in 1915 when its vegetative budding propagation was achieved, through which the agronomic traits could be maintained[4]. After over one century of breeding, the rubber tree clones have undergone a lineage of four generations, resulting in a notable six-fold increase in yield[5]. However, the rubber yield seems hardly to be further elevated after rapid increasing at the early stage of domestication. Breeders attribute the stagnation in yield elevation to limited genetic diversity, as modern rubber clones predominantly descend from only nine ancestral trees[6].

In addressing this challenge, researchers have turned to advanced genomic approaches. High-density genetic maps and genome-wide association studies (GWAS) were used to identify loci and genes related to rubber agronomic traits[7, 8, 9, 10]. The application of marker-assisted selection (MAS) and genomic selection (GS), based on the identification of key genes, represents an efficacious breeding strategy in plant breeding.[11, 12, 13, 14, 15]. Recent genomic studies have revealed the process through which wild rubber trees have been artificially selected to evolve into high-yield clones[9, 16]. These insights, derived from high-quality genomic data, are crucial for precise and efficient breeding. However, the presence of large gaps and unresolved complex regions in the existing genome assemblies limits the accuracy and effectiveness of these genomic tools.

To date, several rubber tree genomes have been published. The first published genome was RRIM 600, a widely planted clone around the world, which was assembled using Illumina data with a 43× coverage. However, this assembly had a contig N50 of only 2.9 Kb[17]. Subsequent assembly of RRIM 600, BMP24, and CATAS 7-33-97 was performed using Illumina, and BAC library data to obtain scaffold-level genomes[18, 19, 20]. In 2019, the first chromosome-level genome of the GT1 clone was published, which increased N50 to 152.7 Kb and reduced scaffolds to 600[21]. Further enhancement of the rubber genome quality was achieved by

advances in long-read sequencing. The use of Oxford Nanopore Technologies (ONT), High throughput chromosome conformation capture (Hi-C), and BioNano facilitated the development of chromosome-level genomes for wild (MT/VB/25A 57/8)[9] and the cultivated clone (CATAS 8-79)[16], which improved chromosomal continuity with contig N50 of 3.51 and 11.2 Mb, respectively. These significant advances offer crucial data resources for genetic study and clone improvements in rubber trees. Currently, all the published genome sequences are not haplotype-resolved. Besides, the large amount of gaps in these assemblies reduces the data integrity and hinders the investigation into the complex genomic regions. Notably, critical regions like telomeres and centromeres remain unresolved in these genomes.

In this study, a haplotype-resolved T2T gap-free genome assembly was successfully constructed in the widely cultivated rubber clone, CATAS 7-33-97. The comprehensive T2T genome marks a major step forward for revealing the complex genome's recalcitrant structural features, and highlights previously overlooked structural variation among haplotypes. Besides, the T2T rubber genome enables the analysis of the evolution of genes related to rubber production and reveals the rubber biosynthesis regulon behind the tapping response.

Results

The haplotype-resolved T2T assembly for the rubber tree reference genome

We generated 193.19 Gb HiFi reads (~122× coverage), 654.04 Gb Hi-C reads (~414× coverage), and 199.24 Gb ultra-long ONT reads (~314× coverage) for assembling CATAS 7-33-97 genome (Fig. 1A, Table1 and Supplementary Tables 1). The initial assembly comprises 256 scaffolds with a size of 3.16 Gb (N50 length: 93.6 Mb). Hi-C reads were employed to rearrange the scaffolds into 36 pseudo-molecules for two haplotypes (Supplementary Fig. 1). Haplotype A (hapA) contained nine gaps (one gap on chromosomes 8, and eight on chromosome 11), while Haplotype B (hapB) displayed seven gaps (one gap on chromosome 4, and three gap on chromosomes 11 and 18 respectively) (Supplementary Tables 2). Then the ONT and HiFi reads were mapped onto the refined pseudo-chromosomes and the existing gaps were manually filled by checking the existence of long reads spanning across the gap junctions (Supplementary Fig. 2). All gaps were successfully filled, resulting in two gap-free haplotypes for the rubber clone CATAS 7-33-97, each with a size of 1.56 Gb. The assembly covers 98.88%

and 99.01% of conserved complete orthologous genes in the Benchmarking Universal Single-Copy Ortholog (BUSCO) embryophyta gene for hapA and hapB, respectively (Supplementary Fig. 3). Total 38,635 protein-coding genes, 4,664 snoRNA, 1,343 rRNA, 736 tRNA, and 250 miRNA were found on hapA, while the numbers were 38,618, 4,694, 1,416, 740, and 247 on hapB (Supplementary Tables 3). About 1.31 Gb (83.97%) of the hapA genome was annotated as repetitive sequences, including 76.56% (1.2 Gb) long terminal repeats (LTRs), of which 48.96% are Gypsy and 13.44% are Copia retroelements.

Telomeres were identified by detecting a 7-base telomeric repeat sequence (CCCTAAA/TTTAGGG) on all chromosomes (Fig. 1B). The centromere regions of chromosomes were preliminarily predicted using quarTeT and CentIER, and then centromeric regions were identified in the haplotype genome using the distribution characteristics of tandem repeats (TRs) and transposable elements (TEs) (Supplementary Fig. 4-5, Supplementary Tables 4-5). Our study successfully identified and included these telomeres and centromeres, resulting in a more complete genome. By comparing the previous genomes, we found that earlier studies lacked the important structures of telomeres and centromeres. For example, the previous MT/VB/25A 57/8 (wild) genome showed significant sequence gaps in the telomeric and centromeric regions of chromosome 16 (Fig. 1B). The centromeric sequences of rubber are predominantly satellite repeats of 61 and 223 bp, in addition to centromeric satellite sequences of different lengths with unit sizes of 60-473 bp.

Variations were detected between the two haplotypes of CATAS 7-33-93 genome, including 28 inversions, 10 transitions, 2,476,725 single nucleotide polymorphisms (SNPs), 169,365 insertions, and 170,240 deletions (Supplementary Tables 6, Supplementary Fig. 6). 71.46% of the SNPs were located in the intergenic region, and only 3.5% were in the exons. To assess the impact of InDel between hapA and hapB on genomic heterozygosity, we screened for structural variants over 100 bp and identified 6,330 InDel across 18 chromosome pairs. Annotation of the InDel indicated that most InDels were located in the non-coding region and only 114 in the exons. Strikingly, a 32.71 Mb multi-structured variant (*sv33M*) region (20,588,014 - 53,302,788) was found between chr8 chromosomal pairs, characterized by numerous structural variant events (Fig. 1B, D). It is noteworthy that high degree of variations was observed in the chr8 corresponding regions in the wild (MT/VB/25A 57/8) and the

cultivated clone (CATAS 8-79) (Fig. 1B). To exclude any sequence misassembly, the long reads of CATAS 7-33-97, the wild, and CATAS 8-79 were mapped to the chr8A. The reads spanning the left and right boundaries of *sv33M* were examined. The read depth decreased rapidly within *sv33M*, with percentages of 82.35% (hapB), 92.31% (wild), and 92.86% (CATAS 8-79) compared to left region outside of *sv33M*. The percentages were 84.62% (hapB), 84.44% (wild), and 86.4% (CATAS 8-79) in comparison to the region outside of the right boundary (Supplementary Fig. 7). In particular, some long reads were found to span across both boundaries, demonstrating the real existing of this large inversion in all these genomes. This is the very first identification of large structural variation between homologous chromosomes in the rubber tree. Due to its extremely large size, *sv33M* may have distinct genetic roles during evolution in *Hevea brasiliensis*.

Comparative genomics of rubber-producing genes in Euphorbiaceae

The T2T rubber genome enables the detailed evolution investigation for the genes related to rubber production. To identify orthologroups of rubber-producing related genes, we collected genomes from eight species (*Arabidopsis thaliana*[22], *Eucommia ulmoides*[23], *Manihot esculenta*[24], *Oryza sativa*[25], *Populus trichocarpa*[26], *Ricinus communis*[27], *Taraxacum kok-saghyz*[28], and *Vitis vinifera*[29]) in addition to rubber tree. Totally 27,311 orthogroups were classified, occupying 89.7% of the overall genes, in which 8,364 orthogroups were present in all these species (Supplementary Tables 7, Supplementary Fig. 8). The rubber clone CATAS 7-33-97 (hapA) has 35,026 genes being divided into 16,584 orthogroups. In comparison to the low-yield wild rubber genome (MT/VB/25A 57/8), the CATAS 7-33-97 exhibits 1,297 expanded orthogroups and 662 contracted orthogroups (Fig. 2A). This represents 95.97% of the families exhibiting significant expansions and contractions among these species (P value $\leq$ 0.01). GO enrichment analysis revealed that the expanded gene family are involved in wounding response, terpene biosynthesis, hydrocarbon metabolism, jasmonic acid response, and lipid metabolism (Fig. 2B). This suggests that these gene family expansions may relate to the rubber yield, as latex is produced after the plant is injured and need to be replenished by synthesizing terpene from sucrose[30].

To further study the evolution of rubber biosynthesis genes in Euphorbiaceae, we collected seven genome of Euphorbiaceae species, including two less rubber-producing plants, *Ricinus*

communis (RC039)[27] and *Manihot esculenta* (SC205)[24], and five rubber tree genomes GT1[21], wild accession MT/VB/25A 57/8, CATAS 8-79, CATAS 7-33-97 (scaffold level), and CATAS 7-33-97 (T2T). There are four classes of genes related to rubber biosynthesis in the plant, MVA pathway genes, Methylerythritol 4-phosphate (MEP) pathway genes, initiator synthesis genes, and rubber elongation genes (Fig. 2C, Supplementary Tables 8). By comparing six genome assemblies with differences in the number of genes in four types, we found that the number of genes for rubber elongation type in rubber was significantly higher than those of *Ricinus communis* and *Manihot esculenta* (Fig. 2D). It could be inferred that the high rubber-producing ability of rubber trees was attributed to the expansion of genes in the rubber elongation type.

The rubber elongation factor (REF) and small rubber particle protein (SRPP) genes are recognized as the most important rubber-producing genes, which share high sequence similarity. The difference lies in the absence of a carboxy-terminus domain in REF protein. A total of 19 REF/SRPP genes were identified in this CATAS 7-33-97 genome, most of which exist in a cluster at edge of chr3, showing a state of dramatic expansion (Supplementary Fig. 9). Compared with the previously published assemblies, this T2T genome revealed an extra SRPP, *HbSRPP11* in rubber tree. In wild rubber tree, we identified 5 REF and 9 SRPP genes in the genome, which are significantly less than other cultivated clones. In *Ricinus communis* and *Manihot esculenta*, no REF but 3 and 5 SRPP genes were identified respectively. These results suggest REF/SRPP gene numbers were correlated to rubber-producing capability. With remarkably high homology between REF and SRPP genes, phylogenetic analysis was conducted on the identified 94 REF and SRPP genes. The SRPP genes were grouped into four clusters, while the REF genes were divided into three clusters (Fig. 2E), indicating the REF and SRPP subgroups may have originated through gene duplication events, followed by divergence and convergent evolution, leading to the formation of distinct subgroups for the REF and SRPP genes, respectively. Moreover, rubber trees have significantly more *cis-prenyltransferase* (*CPT)* genes (11) than *Ricinus communis* (3) and *Manihot esculenta* (7), suggesting its roles on contributing latex production (Fig. 2C).

Consistent allele specific expression (ASE) in rubber tree

Using the collected transcriptome data of CATAS 7-33-97, we identified the differentially expressed bi-alleles in flower (FL), callus (CA), primary embryo (PE), cotyledonary embryo

(CE), mature embryo (ME), and the latex in different tapping periods (T1, T7, TN)[31, 32] (Supplementary Tables 9, Supplementary Fig. 10-11). 10,136 alleles displayed specific expression in all these samples (Fig. 3A), among which 9,546 (94.18%) showed consistent allele specific expression (ASE), i.e., always dominantly expressed in a given haplotype. In these consistent ASE alleles, 4,753 preferred to be highly expressed in hapA chromosomes, while 4,793 alleles were always dominant in hapB (Fig. 3B). It was also surprising that only 590 showed dynamic ASE (dominant allele varied in different samples) pattern. The high percentage of consistent ASE pattern is quite different with that in cassava[24], which is about 33%, indicating a distinguish ASE profiles in rubber tree.

Among these tissues and organs, TN is the latex sample collected from trees tapped for five years, representing the producing natural rubber in mature rubber tree. A total of 4,735 alleles showed showed ASE pattern, in which 2,349 was highly expressed in hapA genome, while 2,386 was dominantly expressed on hapB. These genes' functions were primarily concentrated in biological processes, nucleic acid phosphodiester bond hydrolysis involved in DNA and RNA metabolism (Supplementary Fig. 10-11). These processes represent crucial mechanisms by which cells regulate the stability of DNA and RNA. Interestingly, 17 out 82 genes related to rubber biosynthesis (genes in MVA, MEP, initiator synthesis, and rubber elongation pathway) showed consistent ASE pattern (Fig. 3C, D), leading to expression imbalances between homologous chromosomes that can influence rubber production..

The mevalonate serves as a major carbon reservoir for rubber biosynthesis

Rubber yield is gained by tapping of the bark to collect latex produced in the laticifers (Fig. 1A). When a virgin tree is start to be tapped, the latex yield would increase gradually and reach a stable level after 7-10 tappings. The rubber-producing capability is enhanced during these consecutive taps (repeated mechanical damages). To unravel the mechanism under this phenomenon, we collected latex samples (T1-T10) from unharvested trees (8 years old) from the first to the tenth tapping, as well as latex samples (TN) from trees that have been tapped for five years (13 years old). (Fig. 4A). As expected, the dry rubber yield gradually elevated during the first seven taps (T1-T7, 11.97 ± 1.01 g) and maintained at stable level afterward (T8-TN, Fig. 4B and Supplementary Tables 10). Rubber particle sizes, which is negatively correlated with rubber biosynthesis activity[33, 34, 35], reduced from 1.36 ± 1.11 μm to 0.96 ± 0.01 μm (Fig.

4C, Supplementary Tables 11). Latex sucrose content, a negative indicator of latex metabolic activity[36, 37], was also dropped to lowest level of 5.08 ± 1.27 mM at T7 (Fig. 4D, Supplementary Tables 12), demonstrating latex metabolism reached a very high level after 7 consecutive tappings. The rubber synthetic activity (atom percentage of ^{13}C) measurement showed an increment from 1.09 ± 0.004 APC^{13}C (T1) to 1.16 ± 0.015 APC^{13}C (T7) and reached stable high activity afterward to TN (Fig. 4E, Supplementary Tables 13). We therefore propose that the seventh tapping (T7) represents a critical time point for rubber synthetic activity assessment. Correlation analysis indicates a significant relationship among rubber particle size, sucrose content, dry latex yield, and rubber synthesis rate. Specifically, there is a positive correlation between rubber particle size and sucrose content ($r = 0.7$), as well as between dry latex yield and rubber synthesis rate ($r = 0.69$) (Supplementary Fig. 12).

To elucidate how the rubber synthetic activity is elevated after consecutive tappings, comparative metabolomic and transcriptomic analysis were carried out among T1, T7, and TN (Supplementary Tables 14). Totally 390 differentially expressed metabolites (DEM) were identified among T1, T7, and TN taps. These DEMs were classified into 20 categories with 4 expression patterns (Supplementary Fig. 13). The category of amino acids and derivatives contains the largest number of DEMs (58), reflecting that the cytoplasm (latex) needs to be replenished after evacuation (tapping) in laticifers. Among the top 20 DEMs, the up-regulated metabolites mainly belong to coumarins. Mevalonic acid (MVA), a precursor for isoprene pyrophosphate (IPP) involved in NR biosynthesis, was also up-regulated. The down-regulated ones are primarily dominated by saccharides, which are considered as carbon source for rubber biosynthesis[38] (Fig. 4F).

Rubber biosynthesis is a prolonged terpene biosynthesis pathway, mainly including precursor IPP synthesis and rubber polymer formation. Based on this high-quality genome, all the genes involved in NR biosynthesis were identified and a pathway diagram was reconstructed according to their expression profiles responding to tapping (Fig. 4G). We found that transcript levels of MVA pathway genes are overall higher than those of MEP pathway genes. In the MVA pathway, each enzyme has at least one gene being significantly up-regulated by tapping. Among all genes in the MVA pathway, *mevalonate kinase 1* (*MVK1*) is the most significantly up-regulated one, with over 10-fold transcript increment. Moreover, MVA, the

substrate of MVK, was also found to increase by 14 and 8 folds at T7 and TN tappings (Fig. 4G). Both the transcript and metabolite expression data strongly suggest that MVA accumulation is an indicator for rubber synthetic activity elevation, and this metabolite and its derivatives (mevalonate-5-phosphate and mevalonate-5-pyrophosphate) may serve as a major carbon reservoir for rubber biosynthesis in laticifers.

Consecutive tapping elevates rubber biosynthesis by accumulating JA in latex

A total of 6,446 differential expression genes (DEG) were identified between three tapping periods (Supplementary Fig. 14A, B). The most significant differences were observed between T7 and T1 (4,941 DEGs), in which 2,328 genes were up-regulated and 2,613 down-regulated (Supplementary Fig. 14C). These DEGs could be categorized into four clusters, with 1,256 consecutively up-regulated genes in clusters1 (C1) and 1,476 down-regulated genes in the clusters4 (C4) (Fig. 5A). Gene ontology (GO) enrichment results indicated that terms related to the regulation of RNA biosynthetic processes (GO:2001141), heterocycle biosynthetic processes (GO:0018130), regulation of macromolecule biosynthetic process (GO:0010556) and nitrogen compound metabolic processes (GO:0051171) were significantly enriched in C1 (Fig. 5B), representing genes positively correlated with latex yield ($r = 0.999$) and synthetic activity ($r = 0.951$) in T1, T7 and TN tappings. In contrast, the C4 cluster observed genes with declined expression trends, positively correlating with negative indicators of latex metabolic activity, such as rubber particle size ($r = 0.979$) and sucrose content ($r = 0.982$). The Venn diagram comparison of these four enriched GO terms (GO: 2001141, 0018130, 0010556, and 0051171) revealed an overlap in 92 genes (Supplementary Fig. 14D).

The vast majority of these genes encode plant transcription factors (TF), in which *MYC2* was predicted to be the hub gene in a String network regulating rubber biosynthesis (Fig. 5C). *ERF1B* and *JAZ8* were JA signaling pathways TF, which may be used together with *MYC2* to regulate the increase of rubber synthetic activity during continuous tapping (Fig. 5C). Furthermore, the hormone levels, JA, JA-ILE, and OxIAA, were significantly elevated with consecutive tappings, especially up-regulated by more than 4 folds from T1 to T7. (Fig. 5D-G, Supplementary Tables 15). This may suggest JA's role in regulating rubber synthetic activity. To testify to JA's roles in promoting rubber biosynthesis, we treated the untapped virgin trees with exogenous JA and performed the first tapping (T1) at 24 h after treatment (Supplementary

Tables 16). The atom percentage of ^{13}C was significantly higher in JA-treated trees than the mock trees at T4 and T7 tappings (Fig. 5H), indicating exogenous JA application elevated rubber synthetic activity. These results testified that consecutive tapping activated rubber biosynthesis by accumulating JA in latex. In addition, a regulatory network responsive to auxin was observed in a subset of the TF String network, which is composed of several IAA proteins, indicating that IAA may also play a potential role in rubber tapping. The upregulation of OxIAA levels provides strong evidence for this (Fig. 5C, G).

Discussion

In conclusion, our haplotype-resolved T2T genome facilitated the resolution of the intricate telomeric and centromeric regions of the rubber genome, in addition to structural variation and allelic expression imbalance among rubber haplotypes. Comparative genomic analyses of Euphorbiaceae plants revealed the expansion of the terpene synthesis gene family in cultivated rubber and copy number variations of the rubber elongation gene within the Rubber genus. Multi-omics data from consecutive tappings were employed to propose a model elucidating how tapping activates rubber biosynthesis and improves latex yield. While the detailed regulatory pathways remain to be experimentally validated, this gap-free reference genome is expected to elucidate detail mechanisms soon, and to help in breeding more high yield clones.

Haplotype assembly improves understanding of homologous chromosome variation in rubber tree

Currently, most genomes of highly heterozygous diploid or polyploid species are chimeric, overlooking the differences between haplotypes. However, researchers have recently assembled haplotype genomes in many species, revealing the impact of chromosomal rearrangements, structural variations, and haplotype-specific allele expression on phenotypes[39, 40, 41]. In this study, we assembled the first haplotype-resolved T2T genome in the rubber tree, which has significant advantages over previously published genomes in both structural studies and allele expression comparisons. In addition, haplotype-resolved genome assembly can facilitate allele studies. In rubber tree, new RNA need to be synthesized to assemble ribosomes for protein translation and to initiate latex replenishment after each tapping. During this process, bi-allelic

expression differences were detected in rubber-producing genes, indicating rubber biosynthesis potential may be determined by the parental origin of the allele (Supplementary Fig. 11). The strong allele preferences possibly be attributed to the genetically narrow background with high heterozygosity in rubber trees. This finding may have great significance in rubber breeding by utilizing dominant alleles to accelerate the genome selection strategies.

In terms of genome structure, our T2T genome has identified all telomeres and the majority of centromeric regions across chromosomes. However, there are still 14 chromosomes with unidentified centromeres (Supplementary Tables 5). In most plants, centromeres are composed of large arrays of species-specific satellite repeats interspersed with long-terminal repeat (LTR) retrotransposons. The majority of these centromeres are enriched with DNA repeats ranging in size from 100 to 1000 bp[42, 43]. Additionally, there is a diversity of satellite repeats for different chromosomal centromeres, which suggests a rapid evolution of the centromeric sequence[44, 45, 46, 47]. In the centromere region of rubber genome, the most dominant satellite repeats are 61 and 223 bp (Supplementary Fig. 5). Additionally, the centromere region contains various lengths of repeat units, and the rubber centromere exhibits a high diversity of sequences. However, centromeric regions of the 14 chromosomes were not identified in this study, which is related to the identification method based on repeat sequences. Similar observations have been reported in potato, where some chromosomes lack detectable satellite repeats in their centromere sequences[48]. There are no satellite repeats longer than 30 kb on chromosomes whose the centromere has not been identified (Supplementary Table. 17). This leads us to speculate that these rubber chromosome centromere sequences may also be repeat-free centromeres. Currently, there are no reports on centromere proteins in rubber. Employing a centromere protein-based ChIP-seq approach could facilitate the localization of centromere regions across all rubber chromosomes.

Haplotype resolved T2T genome facilitates the elucidation of yield elevation mechanism

Currently, there are few studies on rubber trees responding to tapping. Our multi-omics data suggest that rubber yield is usually elevated by consecutive tapping and normally reaches a stable high level at T7 (first tapping year). This process takes about 21 days under the current tapping system (1/2S 3d). Sucrose content, rubber particle sizes, and rubber synthetic activity

has been fully activated during this process (Fig. 4B-E). JA and its active derivatives were highly accumulated (4-6 fold) after consecutive injuries in 21 days (Fig. 5D-G). Several regulatory factors involved in JA signaling were significantly up-regulated during tapping, including *MYC2*, *ERF1B* and *JAZ8* (Fig. 5C). *MYC2* functions as a pivotal regulatory hub in the signal transduction processes that follow JA signaling[49]. Meanwhile, *ERF1B* is identified as an essential integrator of ethylene and JA signals by driving *MYC2* transcription[50]. *JAZ8* plays a different role by mediating transcriptional[51, 52]. The JAZ8-MYC2 module, which integrates both ethylene and JA signals, is known to regulate the biosynthesis of anthocyanin in apples[53], and it has been established that *JAZ8* interacts with *MYC2*[52]. In summary, we believe that the *MYC2* may be deeply involved in the regulation of JA signals during rubber tapping process.

In addition, we have discovered high expression of the *MVK1* gene and accumulation of MVA. *MVK1* expression and its substrate associated metabolite MVA rapidly responded to rubber tapping. The MVA levels increased dramatically by about 14 times, indicating that the MVA and its derivatives may serve as a metabolic marker of rubber synthetic activity, and as a carbon reservoir for quick poly-isoprene replenishment in laticifers (Fig. 4G). In rubber tree, studies have reported that JA signaling is involved in regulating rubber biosynthesis within laticifer cells[54]. It is noteworthy that a MYC-responsive element was identified in the upstream region of the *MVK1* (-456 bp), which the homologous gene MVK2 does not contain. We collected the latex processed by JA, and qRT-PCR results showed that JA upregulated the expression of *MVK1* in latex (Supplementary Figure 15). These findings support that JA promotes the transcription of *MVK1* and activates the MVA pathway through a signaling network, providing a carbon reservoir for the rapid biosynthesis of rubber.

Based on these findings, we propose a mechanism by which consecutive tapping affects rubber yield. Mechanical damage inflicted by tapping increases the levels of JA, which subsequently activates genes within the JA signaling pathway. Among these genes, *MYC2* functions as a central transcription factor that enhances the expression of *MVK1*, thereby upregulating the MVA pathway. Carbohydrates provide the necessary substrates for this pathway, ultimately promoting the biosynthesis of rubber (Fig. 5I).

Methods

Plant materials and sequencing

The rubber tree clone, CATAS 7-33-97, is planted in the National Rubber Tree Germplasm Repository (Danzhou, Hainan, 19° 52′ N, 109° 50′ E, South China). Young bronze leaves were collected, immediately frozen in liquid nitrogen, and sent for library construction. For latex transcriptome and metabolome analysis, the trees were tapped half spiral tapping (S/2) with 3 days interval (1/2 S 3d). Fresh latex was collected, immediately frozen in liquid nitrogen, and used for subsequent analysis. The CATAS 7-33-97 WGS data were generated from the PacBio Sequel II platform, producing high-fidelity reads (HIFI) data; the Oxford Nanopore PromethION 48 platform, producing Oxford Nanopore Technologies (ONT) sequences; and the Illumina platform, producing High-throughput Chromosome Conformation Capture (Hi-C) sequences.

Genome assembly and annotation

The HiFi data, ONT data, and Hi-C data underwent self-correction, trimming, assembly, and phasing using the default parameters of the hifiasm (v.0.19.4)[55]. The initial output of hifiasm resulted in two haplotype genome sequences. Ragtag with default parameters was applied to generate homology-based scaffolds by referring to the wild rubber genome[56]. For feature annotation, juicer (v1.6)[57] was employed in conjunction with homology-based scaffolds and Hi-C data. Subsequently, 3D-DNA software was used for scaffolding[58]. The assembly results were manually corrected using Juicebox Assembly Tools (v2.20.00)[59], resulting in a chromosome-level genome. However, gaps persisted in the genome. Finally, minimap2 (v2.24)[60] was used to map ONT and HiFi data to the chromosome GAP regions. The BAM files were observed through IGV[61] to fill the GAP and confirm the sequences in GAP positions. In the end, we obtained a gap-free genome for CATAS 7-33-97.

Repetitive sequences prediction: The tandem repeats within the entire genome were initially annotated using Tandem Repeats Finder (TRF)[62]. Subsequently, transposable elements (TEs) within the rubber tree genome were identified through a combination of ab initio prediction and homology-based methods. To begin, an *ab initio* repeat library specific to the rubber tree was constructed using default parameters with MITE-hunter[63] and RepeatModeler[64], which included LTR_FINDER, ltr_harverst, and LTR_retriver components tailored for plant genomes. This library was then compared against TEclass Repbase

(http://www.girinst.org/repbase) to classify the various types of repeat families. For comprehensive genome-wide identification of repeats, RepeatMasker[65] was employed to search for both known and novel TEs, utilizing both the de novo repeat library and the Repbase TE library. Overlapping TEs belonging to the same repeat class were subsequently merged and analyzed.

Non-coding RNA prediction: For the prediction of non-coding RNAs, a multi-step approach was employed. Initially, the cmscan program within the Infernal (v1.1.4)[66] was utilized to align genome sequences against the Rfam database[67], facilitating the detection of various RNA types including MicroRNA, rRNA, small nuclear RNA (snRNA), and small nucleolar RNA (snoRNA). Subsequently, tRNAscan-SE (v2.0.9)[68] was applied with eukaryote-specific parameters to identify transfer RNA (tRNA) sequences within the rubber tree genome. Lastly, the prediction of ribosomal RNAs (rRNAs) and their subunits was accomplished using models constructed by RNAmmer (v1.2)[69]. This comprehensive approach ensured a thorough exploration of diverse non-coding RNA species within the genome.

Functional gene annotation: The annotated genes were subjected to further annotation by aligning their corresponding translated protein sequences against various databases, including the Non-Redundant Protein Database (NR), Kyoto Encyclopedia of Genes and Genomes (KEGG)[70], Eukaryotic Orthologous Groups of proteins (KOG)[71], Gene Ontology (GO)[72], and Swiss-Prot databases[73], utilizing the blastp program with parameters set to -evalue 1e-5 and -max_target_seqs 1. For the genes related to rubber production, we performed a manual correction.

Annotation evaluation: To evaluate the genome annotation, Benchmarking Universal Single-Copy Orthologs (BUSCO)[74] analysis was employed. The annotated protein sequences were aligned against the corresponding BUSCO databases using HMMER3[75], and the integrity of the alignment results was used to determine the presence of BUSCOs.

Identification of telomeres and centromeres in the Rubber genome

To analyze the telomere repeat units using the TIDK (v0.2.0) (https://github.com/tolkit/telomeric-identifier). Then, we conducted a whole-genome search using the following parameters: tidk search -f hap.fa -s TTTAGGG -o tidk_search --dir telomere_find. Finally, we used TIDK plots and circos (v0.69.6)[76] to visualize telomere peaks.

We used the CentroMiner module of quarTeT (v1.1.5)[77] to predict the centromere regions of rubber with those parameters (n = 150, m = 300, g = 10000, l = 5000). In addition, centromere regions were identified using the default parameters of CentIER[78]. The periods of the repeat units located in the predicted region were then screened, and the distribution of these repeat units on all chromosomes was observed using IGV (tandem repeats smaller than 50 period sizes and copy numbers were filtered out), and finally the centromeric region was screened.

Structural variation analysis

To investigate structural variations among rubber genomes, we compared the 4 haplotype genomes sequentially. First, use minimap2 to compare the two genomes to obtain the bam file, and then use SyRI (v0.23.4)[79] to mine the generated VCF variation information of the bam file. Finally, structural variations were visualized using plotsr (v1.1.0) and NGenomeSyn (v1.41)[80, 81].

On chromosome 8, we observed a large structural variation region to verify the accuracy of the structural variation. We used minimap2[60] to map the SMRT long-reads of CATAS 7-33-97, MT/VB/25A 57/8, and CATAS 8-79 to the hapA genome and then used IGV[61] to observe the read characteristics of the 20 kb region upstream and downstream of this region and verify the structural variation.

Gene family expansion and contraction analysis

The orthogroups in the ten genomes were analyzed using OrthoFinder[82] to infer their orthologous relationships. The species tree alignment sequences were then trimmed using Trimal[83], and a species phylogenetic tree was constructed using raxmlHPC-PTHREADS with the PROTGAMMAJTT model. Finally, CAFE5[84] was employed to analyze the expansion and contraction of orthogroups. Orthogroups exhibiting significant expansion or contraction ($p \leq 0.01$) in cultivated rubber were selected for GO enrichment analysis.

Identification of rubber production-related genes

Firstly, we extracted the CDS sequences and converted them into protein sequences. Then, we used the published CDS sequences of rubber biosynthesis genes as a database[20] and employed Blast+[85] for conducting alignments and searches. Subsequently, we manually checked the alignment and identified the sequences that align with the genes. For SRPP and REF genes, we aligned the database sequences to the hapA genome and identified any

unannotated ones. Additionally, we incorporated NCBI's Conserved Domain Database (https://www.ncbi.nlm.nih.gov/Structure/bwrpsb/bwrpsb.cgi)[86] and MEME (https://meme-suite.org/meme/db/motifs)[87] as complementary tools for aiding in the identification process, thereby enhancing the overall accuracy.

Bi-allelic expression analysis

Firstly, syntenic blocks between hapA and hapB were identified utilizing MCScanX[88]. Subsequently, genes within these syntenic blocks were aligned using BLASTP (e value=1e-10). Gene pairs with the highest bit score and greater than 50% identity for each gene in the block were selected. Ultimately, 30,586 pairs of bi-allelic were obtained.

For expression analysis, RNA-seq data were aligned with the hapA and hapB sequences using HISAT2 (v2.2.0)[89]. The alignment records with unique alignment positions were extracted based on the characteristics of HISAT2 alignment. Subsequently, quantification was performed using the Rsubread (v2.16.1)[90] package. Finally, differential gene expression was determined by analyzing the identified gene pair sequences from previous steps.

Rubber latex physiology diagnosis and rubber biosynthesis activity measurement

5 ml fresh latex from each tree was chilled on ice immediately for 5 min, and the rubber latex physiology diagnosis was carried out when the samples were carried back to the laboratory. Detailed measurements were according to those previously described[91]. For rubber molecular weight determination, 0.2 ml latex was dried and dissolved in 10 ml tetrahydrofuran (THF) for 3 days, then filtered with a 0.45 μm Minisart SRP 25 filter (Sartorius, Gottingen, Germany). The dissolved NR was passed through a Styragel HR5gel permeation chromatography (GPC) column (Waters Corporation) installed on a Waters 1515 isocratic HPLC pump system, and the signal was detected with a 2414 refractive index detector (RID) (Waters Corporation, Milford, MA, USA). GPC was carried out using 30 μl of solution at 40 ℃, and the flow rate was set as 1.0 ml min−1. The latex rubber biosynthesis activity was determined by calculating the ^{13}C content percentage of the total carbon according to Deng et al [54].

To study the correlation between rubber yield-related substances, ggstatsplot (v0.12.2)[92] was used to conduct correlation analysis and statistical significance on the rubber particle size, dry latex content, sucrose content, and synthetic activity of the latex.

Analyse differential expression, gene clustering, functional enrichment

R package corrplot (v0.92)[93] was used to present the correlation coefficient matrix and analyze feature samples. The vegan (v2.6.4)[94] package for PCA analysis was evaluated sample similarity and dispersion. The edgeR (v3.38.1)[95] was used for differential expression analysis. The statistical threshold of DEGs were the following parameters: fold change > 2 and adjusted $P < 0.05$. Tapping rubber displaying the temporal profiles of clusters was detected by k-means using R package Mfuzz (v2.62.0)[96]. For each clustering, genes were performed functional enrichment by TBtools (v2.041)[97]. GO terms of each gene are from genomic functional annotation. The top 5 with P-value representative GO terms were presented for each clustering. Protein interaction relationship was obtained using protein sequences in the STRING[98] database, with Arabidopsis as a reference, and combined with expression differences between T7 and T1 visualized into a PPI network by cytoscape (v3.10.0)[99].

Data availability

The raw data of genome and latex transcriptome generated in this study have been deposited in the Genome Sequence Archive in BIG Data Center, Beijing Institute of Genomics (BIG), Chinese Academy of Sciences, under accession code PRJCA022207. The assembled genome and annotation file of CATAS 7-33-97 are available at HeveaDB (http://hevea.catas.cn/home/index).

Acknowledgments

This research was financially supported by the Central Public-interest Scientific Institution Basal Research Fund for Chinese Academy of Tropical Agricultural Sciences (1630022023010), Chinese Academy of Tropical Agricultural Sciences for Science and Technology Innovation Team of National Tropical Agricultural Science Center (NO. CATASCXTD202401), Hainan Province Science and Technology Talent Innovation Project (KJRC2023C18), the Project of National Key Laboratory for Tropical Crop Breeding (No. NKLTCB202321).

Author contributions

H.C., Y.F.Z. and S.W.H. designed the project. C.C.L. performed a comparative genomic and evolutionary analysis of rubber-producing genes. Z.Y.W., J.X.Q., X.W. and S.C. assembled and annotated the genome. Z.Y.N., T.K.W., W.G.W., H.S.H., Z.W.A., Z.D and D.X. collected the sample in this study. Y.Y., Z.X.J., X.B.W. and W.H were involved in OMICS analyses.

B.Q.L participated in image layout and schematic drawing. C.C.L. and Y.Y. wrote the manuscript with contributions from H.C. and Y.F.Z.

Competing interests

The authors declare no competing interests.

Figure Legends

Table. 1 Genome assembly statistics for the rubber.

Table 1 | Genome assembly statistics for the rubber.

	CATAS 7-33-97		CATAS 7-33-97(V2016)	CATAS8-79	MT/VB/25A 57/8
Illumina (Gb)	-		138	85	348
ONT (Gb)	-		-	-	465
ONT ultralong (Gb)	199		-	-	-
Hi-C (Gb)	654		-	110	119
Pacbio (Gb)	173		-	192	-
Total assembly size (Gb)	3.16		1.37	1.58	1.73
Longest contigs (Mb)	112.6	113.9	0.31	56.83	21.48
Number of contigs	140	116	84,285	1,058	1,573
Contig N50 (Mb)	93.6	94.3	0.03	11.21	3.51
Scaffold N50 (Mb)	96	96	1.28	94.82	47.97
Genome size(Gb)	hapA	hapB	1.35	1.55	1.68
	1.56	1.56			
Genome completeness (BUSCO)	1,596 (98.8%)	1,598 (99.0%)	-	1,546 (95.80%)	1,344(97.75%)
Repeats (%)	76.96%	76.89%	71.18%	-	70.81%
Protein-coding genes	35,152	32,189	35,267	38,595	39,342
Number of telomeres	36	36	-	23	8
Gap number	0	0	-	282	1,373

'-', no data in the original publication.

Fig.1 The phenotypic and genomic features of CATAS 7-33-97. A The tree of cultivated rubber clone, CATAS 7-33-97. **B** The T2T genome of CATAS 7-33-97 is compared with the previously published genomes of wild rubber (MT/VB/25A 57/8) and cultivated rubber (CATAS 8-79). Telomeric regions are marked with red dots, centromeric regions are marked with black blocks, and inverted regions are marked with yellow. **C** The circos plot of CATAS 7-33-97 genome characteristics, the center represents the syntenic regions between the hapA and hapB genomes. **I-VIII** represents the length of the 18 chromosome pairs in CATAS 7-33-97, telomere regions, GC content, gene density, *Copia* density, *Gypsy* density, tandem repeat density, and transposon element density, respectively. **D** Collinearity of chromosome 8 between hapA and hapB genomes.

Fig.2 Evolution and diversity of rubber-producing genes in Euphorbiaceae. A The *Hevea*

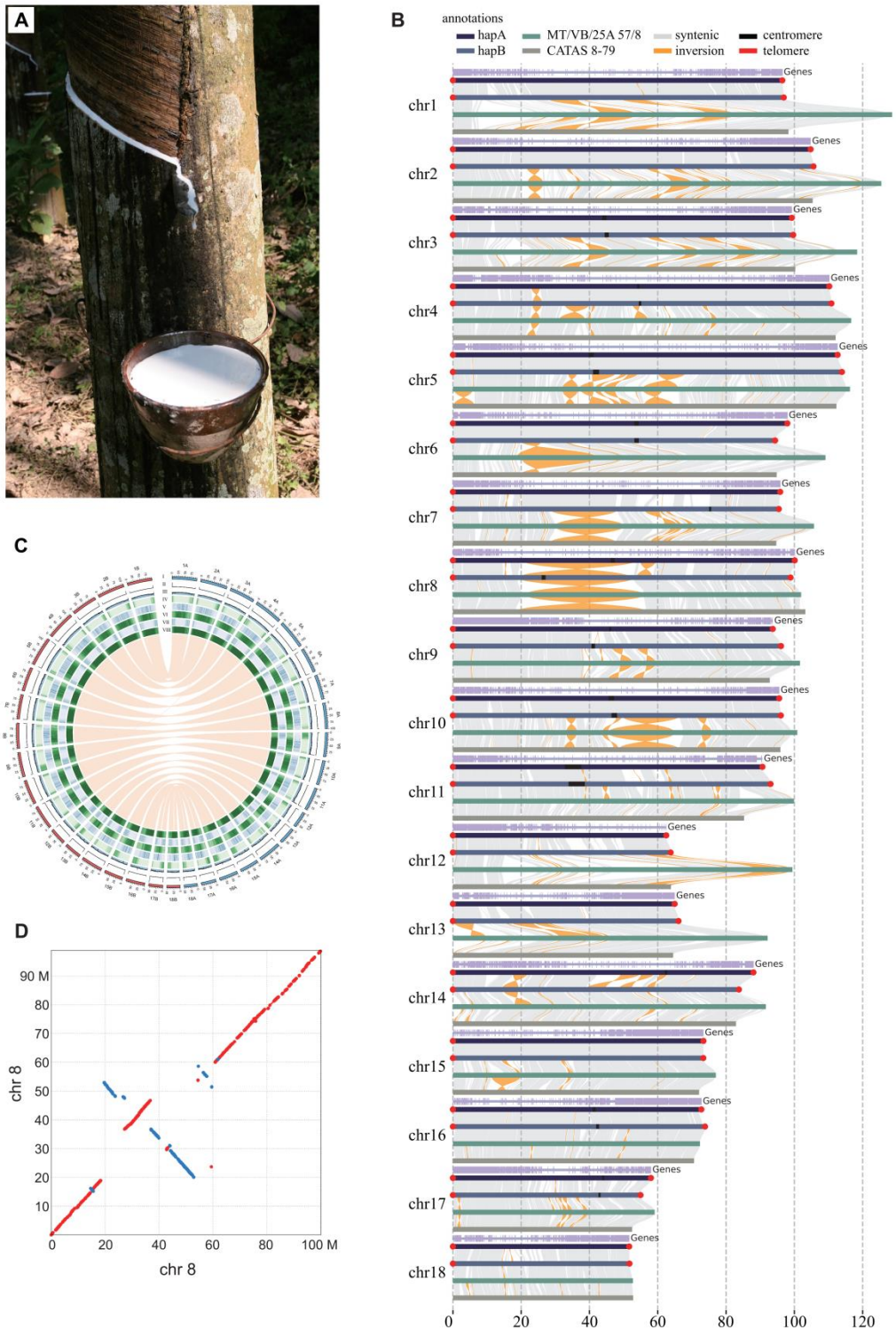
gene cluster expansion and contraction: the pink, red and blue segments of the pie chart represent the maintained, expanded and contracted gene, respectively. **B** GO enrichment results for gene clusters with highly significant expansion and contraction of CATAS 7-33-97. **C** The copy number of rubber-producing genes in Euphorbiaceae plants. The coloured bars represent the functional pathways of the gene. **D** The number of genes involved in the MVA, MEP, initiator synthesis, and rubber elongation pathways in Euphorbiaceae genomes. **E** Phylogenetic tree of REF and SRPP genes in Euphorbiaceae plants. The REF and SRPP genes are labeled in green and blue, respectively.

Fig.3 Allele specific expression (ASE) pattern in rubber. A Differential bi-allelic (DBA) between CATAS 7-33-97 haplotypes (pink) and differentially expressed bi-allelic (DEBA) genes in various samples (green). **B** Consistent ASE pattern in rubber tree. The horizontal bar chart on the left represents the number of highly expressed bi-allelic genes in each haplotype. The vertical bar chart on the top shows the number of ASE genes in each haplotype in the samples connected by solid dots and lines. Note that the majority of ASE genes were highly expressed in a given haplotype in all samples. **C** The expression profile of the ASE genes in hapA and hapB, red dots represent genes involved rubber biosynthesis. **D** Genes involved in the rubber biosynthesis pathways, i.e. MVA, MEP, IS and RE. Genes highlight in orange are in ASE.

Fig.4 Response of rubber biosynthetic pathway metabolites to consecutive tappings. A Rubber tapping system diagram. Rubber trees that are 8 years old and have reached a circumference of 50 cm undergo the first tapping (T1) to collect latex, and the next tapping takes place after 3 days interval. The consecutive tappings were performed until T10. In addition, latex is collected from trees that have been tapped for 5 years (TN). **B-E** Measurement of dry latex content (B), rubber particle size (C), sucrose content (D) and synthetic activity of the latex (E) from each tapping. The *P*-values were calculated by comparing the data with those of the previous tapping. **F** Differentially expressed metabolites (DEM) in T1, T7 and TN latex. **G** Rubber biosynthetic pathway, the heatmaps represent the expression levels of genes and metabolites in T1, T7 and TN latex.

Fig.5 Metabolome and hormone profiling during consecutive tappings. A K-means clustering shows the expression changes over time, with T1, T7, and TN representing key tapping intervals. **B** Functional category of GO enrichment among the four major clusters. **C** Regulatory network of key transcription factors involved in the macromolecular synthesis. **D-G** Jasmonic acid (JA), JA isoleucine conjugate (JA-ILE), JA valine (JA-Val) and indole-3-acetic acid (OxIAA) levels in the latex of T1, T7 and TN. **H** JA application facilitate rubber biosynthesis activity. **I** Mechanism of rubber biosynthesis in response to tapping.

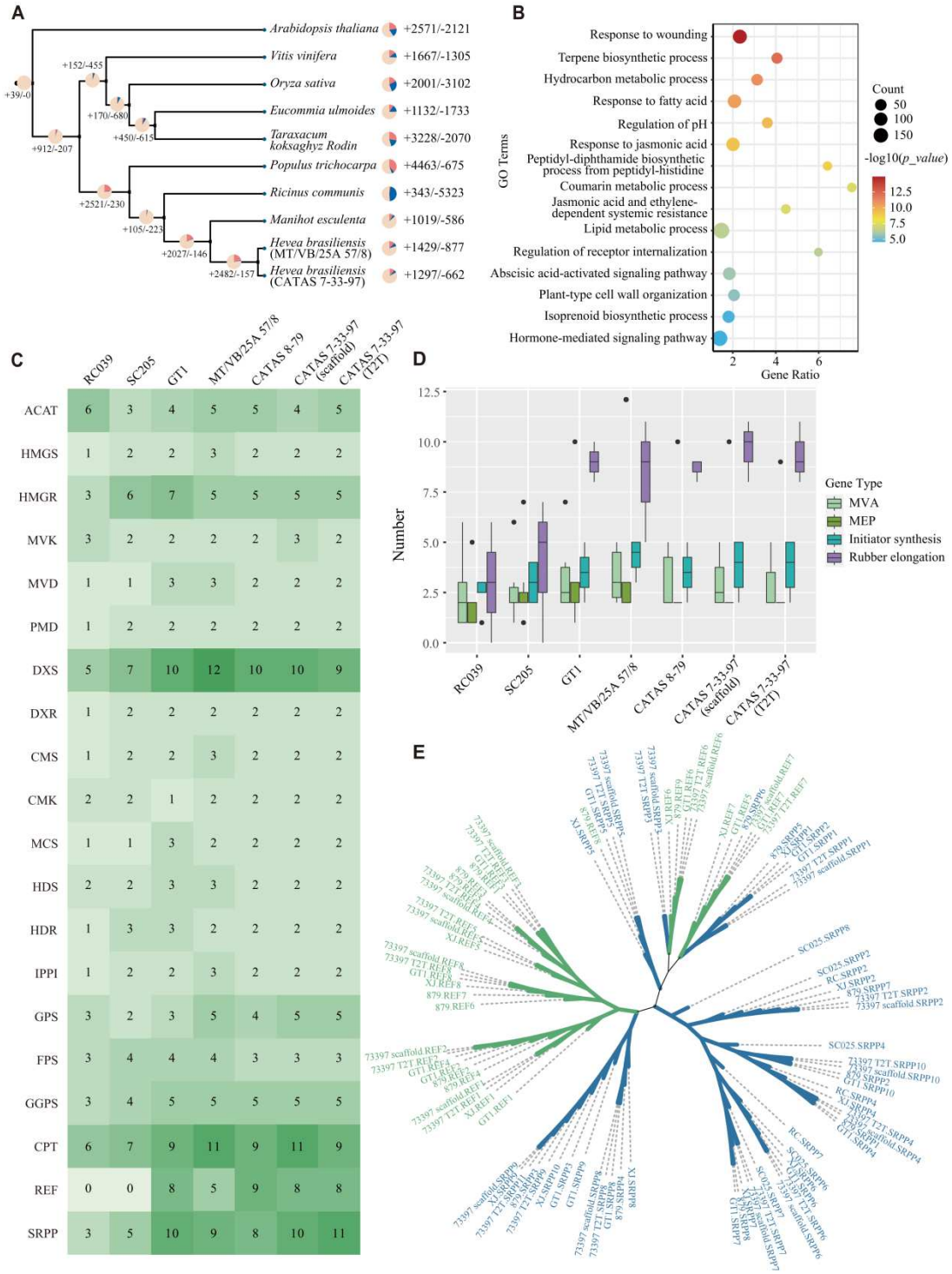
584 **Fig. 1**



585

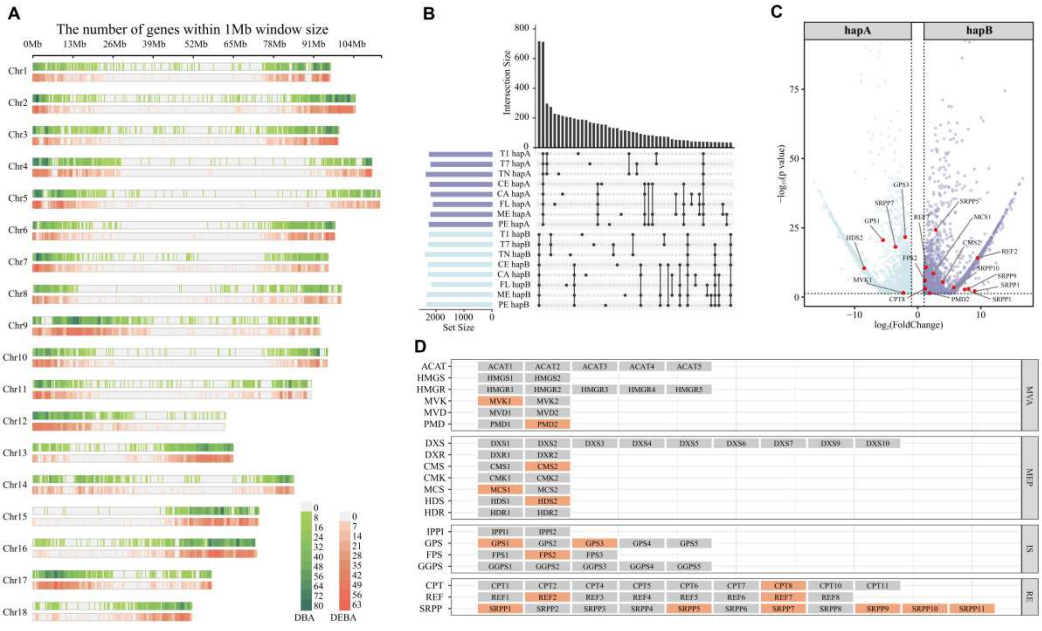
586

587 **Fig. 2**


A
Arabidopsis thaliana +2571/-2121
Vitis vinifera +1667/-1305
Oryza sativa +2001/-3102
Eucommia ulmoides +1132/-1733
Taraxacum koksaghyz Rodin +3228/-2070
Populus trichocarpa +4463/-675
Ricinus communis +343/-5323
Manihot esculenta +1019/-586
Hevea brasiliensis (MT/VB/25A 57/8) +1429/-877
Hevea brasiliensis (CATAS 7-33-97) +1297/-662
+39/-0
+152/-455
+170/-680
+450/-615
+912/-207
+2521/-230
+105/-223
+2027/-146
+2482/-157
B
Response to wounding
Terpene biosynthetic process
Hydrocarbon metabolic process
Response to fatty acid
Regulation of pH
Response to jasmonic acid
Peptidyl-diphthamide biosynthetic process from peptidyl-histidine
Coumarin metabolic process
Jasmonic acid and ethylene-dependent systemic resistance
Lipid metabolic process
Regulation of receptor internalization
Abscisic acid-activated signaling pathway
Plant-type cell wall organization
Isoprenoid biosynthetic process
Hormone-mediated signaling pathway
GO Terms
Gene Ratio
Count
50
100
150
-log10(p_value)
12.5
10.0
7.5
5.0
C
RC039 SC205 GT1 MT/VB/25A 57/8 CATAS 8-79 CATAS 7-33-97 (scaffold) CATAS 7-33-97 (T2T)
ACAT 6 3 4 5 5 4 5
HMGS 1 2 2 3 2 2 2
HMGR 3 6 7 5 5 5 5
MVK 3 2 2 2 2 3 2
MVD 1 1 3 3 2 2 2
PMD 1 2 2 2 2 2 2
DXS 5 7 10 12 10 10 9
DXR 1 2 2 2 2 2 2
CMS 1 2 2 3 2 2 2
CMK 2 2 1 2 2 2 2
MCS 1 1 3 2 2 2 2
HDS 2 2 3 3 2 2 2
HDR 1 3 3 2 2 2 2
IPPI 1 2 2 3 2 2 2
GPS 3 2 3 5 4 5 5
FPS 3 4 4 4 3 3 3
GGPS 3 4 5 5 5 5 5
CPT 6 7 9 11 9 11 9
REF 0 0 8 5 9 8 8
SRPP 3 5 10 9 8 10 11
D
Number
Gene Type
MVA
MEP
Initiator synthesis
Rubber elongation
RC039
SC205
GT1
MT/VB/25A 57/8
CATAS 8-79
CATAS 7-33-97 (scaffold)
CATAS 7-33-97 (T2T)
E

588

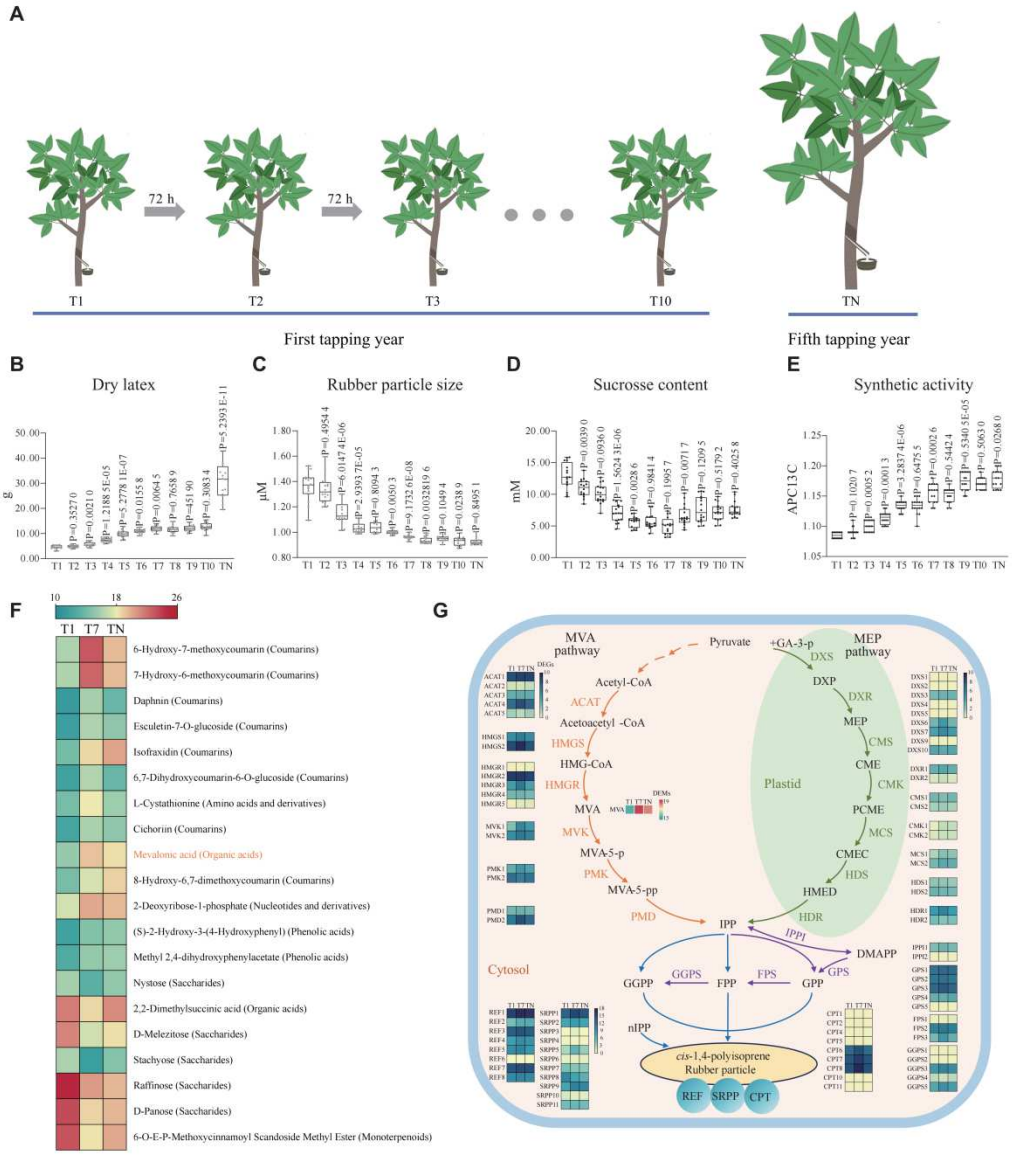
Fig. 3


A
The number of genes within 1Mb window size
0Mb 13Mb 26Mb 39Mb 52Mb 65Mb 78Mb 91Mb 104Mb
Chr1 Chr2 Chr3 Chr4 Chr5 Chr6 Chr7 Chr8 Chr9 Chr10 Chr11 Chr12 Chr13 Chr14 Chr15 Chr16 Chr17 Chr18
DBA DEBA
B
Intersection Size
Set Size
T1 hapA T7 hapA TN hapA CE hapA CA hapA FL hapA ME hapA PE hapA
T1 hapB T7 hapB TN hapB CE hapB CA hapB FL hapB ME hapB PE hapB
C
hapA hapB
−log10(p value)
log2(FoldChange)
D
ACAT HMGS HMGR MVK MVD PMD
DXS DXR CMS CMK MCS HDS HDR
IPPI GPS FPS GGPS
CPT REF SRPP
MVA MEP IS RE

590

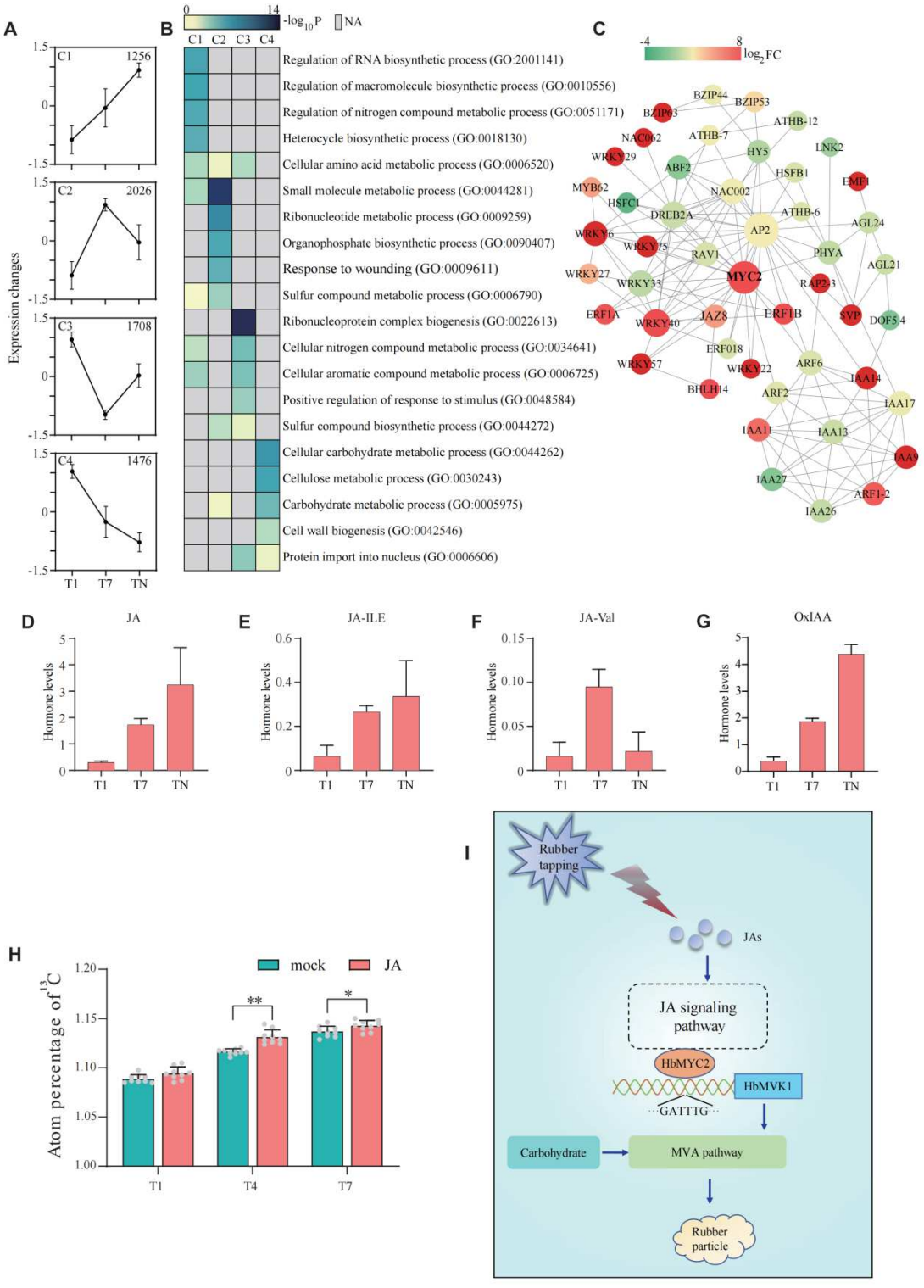
591

592 **Fig. 4**


A
72 h
72 h
T1
T2
T3
T10
TN
First tapping year
Fifth tapping year
B
Dry latex
g
C
Rubber particle size
μM
D
Sucrosse content
mM
E
Synthetic activity
APC13 C
F
10
18
26
T1 T7 TN
6-Hydroxy-7-methoxycoumarin (Coumarins)
7-Hydroxy-6-methoxycoumarin (Coumarins)
Daphnin (Coumarins)
Esculetin-7-O-glucoside (Coumarins)
Isofraxidin (Coumarins)
6,7-Dihydroxycoumarin-6-O-glucoside (Coumarins)
L-Cystathionine (Amino acids and derivatives)
Cichoriin (Coumarins)
Mevalonic acid (Organic acids)
8-Hydroxy-6,7-dimethoxycoumarin (Coumarins)
2-Deoxyribose-1-phosphate (Nucleotides and derivatives)
(S)-2-Hydroxy-3-(4-Hydroxyphenyl) (Phenolic acids)
Methyl 2,4-dihydroxyphenylacetate (Phenolic acids)
Nystose (Saccharides)
2,2-Dimethylsuccinic acid (Organic acids)
D-Melezitose (Saccharides)
Stachyose (Saccharides)
Raffinose (Saccharides)
D-Panose (Saccharides)
6-O-E-P-Methoxycinnamoyl Scandoside Methyl Ester (Monoterpenoids)
G
MVA pathway
MEP pathway
Pyruvate
+GA-3-p
Acetyl-CoA
ACAT
Acetoacetyl -CoA
HMGS
HMG-CoA
HMGR
MVA
MVK
MVA-5-p
PMK
MVA-5-pp
PMD
DXS
DXP
DXR
MEP
CMS
CME
CMK
PCME
MCS
CMEC
HDS
HMED
HDR
IPP
IPPI
DMAPP
GPS
GPP
FPS
FPP
GGPS
GGPP
nIPP
Plastid
Cytosol
cis-1,4-polyisoprene Rubber particle
REF
SRPP
CPT

593

Fig. 5


A
B
C
C1 1256
C2 2026
C3 1708
C4 1476
Expression changes
T1 T7 TN
-log10P NA
C1 C2 C3 C4
Regulation of RNA biosynthetic process (GO:2001141)
Regulation of macromolecule biosynthetic process (GO:0010556)
Regulation of nitrogen compound metabolic process (GO:0051171)
Heterocycle biosynthetic process (GO:0018130)
Cellular amino acid metabolic process (GO:0006520)
Small molecule metabolic process (GO:0044281)
Ribonucleotide metabolic process (GO:0009259)
Organophosphate biosynthetic process (GO:0090407)
Response to wounding (GO:0009611)
Sulfur compound metabolic process (GO:0006790)
Ribonucleoprotein complex biogenesis (GO:0022613)
Cellular nitrogen compound metabolic process (GO:0034641)
Cellular aromatic compound metabolic process (GO:0006725)
Positive regulation of response to stimulus (GO:0048584)
Sulfur compound biosynthetic process (GO:0044272)
Cellular carbohydrate metabolic process (GO:0044262)
Cellulose metabolic process (GO:0030243)
Carbohydrate metabolic process (GO:0005975)
Cell wall biogenesis (GO:0042546)
Protein import into nucleus (GO:0006606)
log2FC
BZIP44
BZIP53
BZIP93
ATHB-12
NAC062
ATHB-7
LNK2
HY5
WRKY29
ABF2
HSFB1
EMF1
MYB62
NAC002
HSFC1
DREB2A
ATHB-6
AGL24
WRKY6
AP2
WRKY75
RAV1
PHYA
AGL21
WRKY27
MYC2
WRKY33
RAP2.3
ERF1A
JAZ8
ERF1B
SVP
DOF5.4
WRKY40
ERF018
WRKY57
WRKY22
ARF6
IAA14
BHLH14
ARF2
IAA17
IAA11
IAA13
IAA9
IAA27
ARF1-2
IAA26
D
JA
E
JA-ILE
F
JA-Val
G
OxIAA
Hormone levels
T1 T7 TN
H
mock
JA
Atom percentage of 13C
T1 T4 T7
**
*
I
Rubber tapping
JAs
JA signaling pathway
HbMYC2
HbMVK1
GATTTG
Carbohydrate
MVA pathway
Rubber particle

Supplementary Data

Supplementary Fig. 1 Hi-C interactions between 18 pairs of chromosomes in CATAS 7-33-97.

Supplementary Fig. 2 Example of using extra-long ONT reads to fill the gap.A Reads alignment result before GAP filling, green box is the GAP region. **B** Reads alignment result after GAP filling.

Supplementary Fig. 3 CATAS 7-33-97 genome quality assessment. A-B The BUSCO assessment results for Haplotype A and Haplotype B, respectively.

Supplementary Fig. 4 Prediction of centromere regions using quarTeT.

Supplementary Fig. 5 Distribution of TR and different repeat units in CATAS 7-33-97 genome, A and B represent hapA and hapB, respectively.

Supplementary Fig. 6 Comparison of CATAS 7-33-97 with wild (MT/VB/25A 57/8) and cultivated (CATAS 8-79) species genomes. A Comparison of chromosome lengths in four haploid genomes. **B** Statistical of structural variation among genomes. **C** Density of SNPs between chromosomes of hapA and hapB.

Supplementary Fig. 7 Chromosome 8 Structural Variation Analysis. A Collinearity of chromosome 8 between hapA and hapB genomes. **B-C** The IGV map shows the results of aligning the CATAS 7-33-97 ONT reads to the left and right of the large structural variation regions on chromosome 8 of hapA. **D** Collinearity of chromosome 8 between hapA and MT/VB/25A 57/8 genomes. **E-F** The IGV map displays the results of aligning MT/VB/25A 57/8 long reads to the left and right region of the huge structural variation regions on chromosome 8 of hap A. **G** Collinearity of chromosome 8 between hapA and MT/VB/25A 57/8 genomes. **H-I** The IGV map displays the results of aligning CATAS 8-79 long reads to the left and right region of the huge structural variation regions on chromosome 8 of hap A.

Supplementary Fig. 8 The overlap of orthogroups in ten species. A Statistics of orthogroups in different species. The horizontal bar chart represents the number of orthogroups in different speciess. The vertical bar graph showing the number of orthogroups shared in different species. The pie chart showing the number of core, unique and accessory genes

Supplementary Fig. 9 The location of REF and SRPP genes on the chromosomes of CATAS 7-33-97, with the gene cluster containing both REF and SRPP genes on chr2.

Supplementary Fig. 10 Differentially expressed bi-allelic in different tissues of CATAS 7-33-97. A Distribution of bi-allelic with similar expression ($-\log_{10}padj < 2$, $\log_2FoldChange < 2$) and differentially expressed bi-allelic ($-\log_{10}padj > 2$, $\log_2FoldChange > 2$). **B-I** Differentially expressed bi-allelic in primary laticifer, secondary laticifer, latex, male flower, primary embryo, cotyledonary embryo, mature cotyledonary embryo, and embryogenic callus.

Supplementary Fig. 11 GO enrichment results of differentially expressed bi-allelic in different tissues of CATAS 7-33-97. A-H Enrichment results of differentially expressed bi-allelic in primary laticifer, secondary laticifer, latex, male flower, primary embryo, cotyledonary embryo, mature cotyledonary embryo, and embryogenic callus. On the X-axis, number greater than 0 represents dominant genes of hapA, conversely number lesser than 0 represents hapB. **I** The Gene Ontology enrichment results of the 686 differentially expressed bi-allelic across all tissues.

Supplementary Fig. 12 Correlation matrix of traits in rubber biosynthesis. The heatmap illustrates Pearson correlation coefficients between dry latex content, rubber particle size, sucrose content, and synthetic activity. Each square represents the correlation between the traits on the vertical and horizontal axes, with the sample data (n = 176) applicable to all correlations. Synthetic activity shows a strong positive correlation with dry latex content (0.69), whereas it is negatively correlated with rubber particle size (-0.8) and sucrose content (-0.56). Conversely, sucrose content is positively correlated with rubber particle size (0.7).

Supplementary Fig. 13 Expression heatmap of differential expression metabolites.

Supplementary Fig. 14 Transcriptomic analysis of rubber under different tapping times. **A** The PCA plot of sample replicability during the T1, T7, and TN time periods. **B** The number of genes upregulated (red) and downregulated (blue) in comparisons T7 vs T1 and TN vs T1. **C** The overlap of differentially expressed genes between the tapping time comparisons T1, T7, and TN. **D** The distribution of genes encoding enzymes involved in RNA biosynthetic processes, heterocycle biosynthetic processes, and regulation of nitrogen compound metabolic processes across the clusters.

Supplementary Fig. 15 Relative expression of MVK1 in latex following JA treatment. A Primer sequences for MVK1 qRT-PCR utilizing YLS8 as the internal reference gene. **B** The relative expression of MVK1 in latex demonstrates a progressive increase post-JA treatment, peaking at 12 hours before rapidly diminishing.

Supplementary Table 1. Sequencing data statistics for genome assembly.

Supplementary Table 2. Number of GAPs on each chromosome in CATAS 7-33-97.

Supplementary Table 3. Non-coding RNA genes of CATAS 7-33-97.

Supplementary Table 4. Centromeric region predicted by CENIER.

Supplementary Table 5. Centromeric region prediction based on CENIER, quarTeT results, and TR distribution.

Supplementary Table 6. Differences in the haploid genome of rubber.

Supplementary Table 7. Statistical results of orthofinder identification of 10 species

orthogroups.

Supplementary Table 8. Rubber production related gene information

Supplementary Table 9. RNA-seq data used in this study.

Supplementary Table 10. Dry latex content of during different tapping times.

Supplementary Table 11. Size of rubber particle during different tapping times.

Supplementary Table 12. Sucrose concent of latex during different tapping times.

Supplementary Table 13. Synthetic activity during different tapping times.

Supplementary Table 14. Metabolite content of latex at different tapping times.

Supplementary Table 15. The hormone levels of latex at three different period.

Supplementary Table 16. Effect of JA treatment on rubber synthetic activity.

Supplementary Table 17 Distribution of tandem repeats on chromosomes whose centromeres have not yet been identified (tandem repeats smaller than 50 period sizes and copy numbers were filtered out).

References

1. Arias M, Van Dijk PJ. What is natural rubber and why are we searching for new sources. *Frontiers for Young Minds* **7**, 100 (2019).

2. Fox J, Castella J-C, Ziegler AD, Westley SB. Rubber plantations expand in mountainous Southeast Asia: what are the consequences for the environment? , (2014).

3. Hurley PE. History of natural rubber. *Journal of Macromolecular Science—Chemistry* **15**, 1279-1287 (1981).

4. Priyadarshan PM. Genesis and Development. In: *Biology of Hevea Rubber*). Springer International Publishing (2017).

5. Junquera V, Meyfroidt P, Sun Z, Latthachack P, Grêt-Regamey A. From global drivers to local land-use change: understanding the northern Laos rubber boom. *Environmental science & policy* **109**, 103-115 (2020).

6. Cheng H. HeveaDB: A Hub for Rubber Tree Genetic and Genomic Resources. In: *The Rubber Tree Genome* (eds Matsui M, Chow K-S). Springer International Publishing (2020).

7. Wu W*, et al.* Ultrahigh-density genetic map construction and identification of quantitative trait loci for growth in rubber tree (Hevea brasiliensis). *Industrial Crops and Products* **178**, 114560 (2022).

8. Chanroj V, Rattanawong R, Phumichai T, Tangphatsornruang S, Ukoskit K. Genome-wide association mapping of latex yield and girth in Amazonian accessions of Hevea brasiliensis grown in a suboptimal climate zone. *Genomics* **109**, 475-484 (2017).

9. Cheng H*, et al.* Chromosome-level wild Hevea brasiliensis genome provides new tools for genomic-assisted breeding and valuable loci to elevate rubber yield. *Plant Biotechnology Journal* **21**, 1058-1072 (2023).

10. Francisco FR*, et al.* Unravelling rubber tree growth by integrating GWAS and biological network-based approaches. *Frontiers in plant science* **12**, 768589 (2021).

11. Degen B, Müller NA. A simulation study comparing advanced marker-assisted selection with genomic selection in tree breeding programs. *G3 Genes|Genomes|Genetics* **13**, (2023).

12. Grattapaglia D. Status and Perspectives of Genomic Selection in Forest Tree Breeding. In: *Genomic Selection for Crop Improvement: New Molecular Breeding Strategies for Crop Improvement* (eds Varshney RK, Roorkiwal M, Sorrells ME). Springer International Publishing (2017).

13. Isik F. Genomic Prediction of Complex Traits Complex traits in Perennial Plants: A Case for

Forest Trees Forest trees. In: *Genomic prediction of complex traits: Methods and Protocols*). Springer (2022).

14. Souza L*, et al.* Genomic Selection in Rubber Tree Breeding: A Comparison of Models and Methods for dealing with G× E. *bioRxiv*, 603662 (2019).

15. Pootakham W, Shearman JR, Tangphatsornruang S. Development of molecular markers in Hevea brasiliensis for marker-assisted breeding. *The rubber tree genome*, 67-79 (2020).

16. Chao J*, et al.* Genomic insight into domestication of rubber tree. *Nature Communications* **14**, 4651 (2023).

17. Rahman AYA*, et al.* Draft genome sequence of the rubber tree Hevea brasiliensis. *BMC genomics* **14**, 1-15 (2013).

18. Lau N-S*, et al.* The rubber tree genome shows expansion of gene family associated with rubber biosynthesis. *Scientific reports* **6**, 28594 (2016).

19. Pootakham W*, et al.* De novo hybrid assembly of the rubber tree genome reveals evidence of paleotetraploidy in Hevea species. *Scientific reports* **7**, 41457 (2017).

20. Tang C*, et al.* The rubber tree genome reveals new insights into rubber production and species adaptation. *Nature plants* **2**, 1-10 (2016).

21. Liu J*, et al.* The chromosome-based rubber tree genome provides new insights into spurge genome evolution and rubber biosynthesis. *Molecular plant* **13**, 336-350 (2020).

22. Sloan DB, Wu Z, Sharbrough J. Correction of persistent errors in Arabidopsis reference mitochondrial genomes. *The Plant Cell* **30**, 525-527 (2018).

23. Li Y*, et al.* High-quality de novo assembly of the Eucommia ulmoides haploid genome provides new insights into evolution and rubber biosynthesis. *Horticulture Research* **7**, (2020).

24. Hu W*, et al.* Allele-defined genome reveals biallelic differentiation during cassava evolution. *Molecular Plant* **14**, 851-854 (2021).

25. Kawahara Y*, et al.* Improvement of the Oryza sativa Nipponbare reference genome using next generation sequence and optical map data. *Rice* **6**, 1-10 (2013).

26. Tuskan GA*, et al.* The genome of black cottonwood, Populus trichocarpa (Torr. & Gray). *science* **313**, 1596-1604 (2006).

27. Lu J*, et al.* A Chromosome-level Genome Assembly of Wild Castor Provides New Insights into its Adaptive Evolution in Tropical Desert. *Genomics, Proteomics & Bioinformatics* **20**, 42-59

(2022).

28. Lin T*, et al.* Genome analysis of Taraxacum kok-saghyz Rodin provides new insights into rubber biosynthesis. *National Science Review* **5**, 78-87 (2018).

29. Shi X*, et al.* The complete reference genome for grapevine (Vitis vinifera L.) genetics and breeding. *Horticulture Research* **10**, uhad061 (2023).

30. Tang C*, et al.* The sucrose transporter HbSUT3 plays an active role in sucrose loading to laticifer and rubber productivity in exploited trees of Hevea brasiliensis (para rubber tree). *Plant, cell & environment* **33**, 1708-1720 (2010).

31. Li F, Wang Y, Gao H, Zhang X, Zhuang N. Comparative transcriptome analysis reveals differential gene expression in sterile and fertile rubber tree varieties during flower bud differentiation. *Journal of Plant Physiology* **265**, 153506 (2021).

32. Wang Y, Li H-L, Zhou Y-K, Guo D, Zhu J-H, Peng S-Q. Transcriptomes analysis reveals novel insight into the molecular mechanisms of somatic embryogenesis in Hevea brasiliensis. *BMC genomics* **22**, 1-18 (2021).

33. Ohya N, Tanaka Y, Koyama T. Activity of rubber transferase and rubber particle size in Hevea latex. *Journal of Rubber Research* **3**, 214-221 (2000).

34. Men X, Wang F, Chen G-Q, Zhang H-B, Xian M. Biosynthesis of natural rubber: current state and perspectives. *International Journal of Molecular Sciences* **20**, 50 (2018).

35. Yamashita S*, et al.* Purification and characterization of small and large rubber particles from Hevea brasiliensis. *Bioscience, Biotechnology, and Biochemistry* **82**, 1011-1020 (2018).

36. Jacob J*, et al. The development of Hevea latex diagnosis*.

37. Jacob JL*, et al.* Yield-Limiting Factors, Latex Physiological Parameters, Latex Diagnosis, and Clonal Topology.) (2018).

38. Tupý J. Some aspects of sucrose transport and utilization in latex producing bark of Hevea brasiliensis Muel. Arg. *Biologia plantarum* **27**, 51-64 (1985).

39. Sun H*, et al.* Chromosome-scale and haplotype-resolved genome assembly of a tetraploid potato cultivar. *Nature genetics* **54**, 342-348 (2022).

40. Han X*, et al.* Two haplotype-resolved, gap-free genome assemblies for Actinidia latifolia and Actinidia chinensis shed light on the regulatory mechanisms of vitamin C and sucrose metabolism in kiwifruit. *Molecular Plant* **16**, 452-470 (2023).

41. Hu W, *et al.* Resequencing of 388 cassava accessions identifies valuable loci and selection for variation in heterozygosity. *Genome biology* **22**, 1-23 (2021).

42. Comai L, Maheshwari S, Marimuthu MP. Plant centromeres. *Current Opinion in Plant Biology* **36**, 158-167 (2017).

43. Melters DP, *et al.* Comparative analysis of tandem repeats from hundreds of species reveals unique insights into centromere evolution. *Genome biology* **14**, 1-20 (2013).

44. Su H, Liu Y, Liu C, Shi Q, Huang Y, Han F. Centromere satellite repeats have undergone rapid changes in polyploid wheat subgenomes. *The Plant Cell* **31**, 2035-2051 (2019).

45. Ma H, *et al.* Centromere plasticity with evolutionary conservation and divergence uncovered by wheat 10+ genomes. *Molecular Biology and Evolution* **40**, msad176 (2023).

46. Liu Y, *et al.* Pan-centromere reveals widespread centromere repositioning of soybean genomes. *Proceedings of the National Academy of Sciences* **120**, e2310177120 (2023).

47. Lv Y, *et al.* A centromere map based on super pan-genome highlights the structure and function of rice centromeres. *Journal of Integrative Plant Biology*, (2024).

48. Gong Z, *et al.* Repeatless and repeat-based centromeres in potato: implications for centromere evolution. *The Plant Cell* **24**, 3559-3574 (2012).

49. Kazan K, Manners JM. MYC2: the master in action. *Molecular plant* **6**, 686-703 (2013).

50. Lorenzo O, Piqueras R, Sánchez-Serrano JJ, Solano R. ETHYLENE RESPONSE FACTOR1 integrates signals from ethylene and jasmonate pathways in plant defense. *The Plant Cell* **15**, 165-178 (2003).

51. Shyu C, *et al.* JAZ8 lacks a canonical degron and has an EAR motif that mediates transcriptional repression of jasmonate responses in Arabidopsis. *The Plant Cell* **24**, 536-550 (2012).

52. Pauwels L, Goossens A. The JAZ proteins: a crucial interface in the jasmonate signaling cascade. *The Plant Cell* **23**, 3089-3100 (2011).

53. Wang S, *et al.* MdERF1B–MdMYC2 module integrates ethylene and jasmonic acid to regulate the biosynthesis of anthocyanin in apple. *Horticulture Research* **9**, (2022).

54. Deng X, *et al.* Jasmonate signalling in the regulation of rubber biosynthesis in laticifer cells of rubber tree, Hevea brasiliensis. *Journal of experimental botany* **69**, 3559-3571 (2018).

55. Cheng H, Concepcion GT, Feng X, Zhang H, Li H. Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. *Nature methods* **18**, 170-175 (2021).

56. Alonge M*, et al.* Automated assembly scaffolding elevates a new tomato system for high-throughput genome editing. *BioRxiv*, 2021.2011. 2018.469135 (2021).

57. Durand NC*, et al.* Juicer provides a one-click system for analyzing loop-resolution Hi-C experiments. *Cell systems* **3**, 95-98 (2016).

58. Dudchenko O*, et al.* De novo assembly of the Aedes aegypti genome using Hi-C yields chromosome-length scaffolds. *Science* **356**, 92-95 (2017).

59. Durand NC*, et al.* Juicebox provides a visualization system for Hi-C contact maps with unlimited zoom. *Cell systems* **3**, 99-101 (2016).

60. Li H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* **34**, 3094-3100 (2018).

61. Thorvaldsdóttir H, Robinson JT, Mesirov JP. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Briefings in bioinformatics* **14**, 178-192 (2013).

62. Benson G. Tandem repeats finder: a program to analyze DNA sequences. *Nucleic acids research* **27**, 573-580 (1999).

63. Han Y, Wessler SR. MITE-Hunter: a program for discovering miniature inverted-repeat transposable elements from genomic sequences. *Nucleic acids research* **38**, e199-e199 (2010).

64. Flynn JM*, et al.* RepeatModeler2 for automated genomic discovery of transposable element families. *Proceedings of the National Academy of Sciences* **117**, 9451-9457 (2020).

65. Chen N. Using Repeat Masker to identify repetitive elements in genomic sequences. *Current protocols in bioinformatics* **5**, 4.10. 11-14.10. 14 (2004).

66. Nawrocki EP, Eddy SR. Infernal 1.1: 100-fold faster RNA homology searches. *Bioinformatics* **29**, 2933-2935 (2013).

67. Kalvari I*, et al.* Rfam 14: expanded coverage of metagenomic, viral and microRNA families. *Nucleic Acids Research* **49**, D192-D200 (2021).

68. Chan PP, Lin BY, Mak AJ, Lowe TM. tRNAscan-SE 2.0: improved detection and functional classification of transfer RNA genes. *Nucleic acids research* **49**, 9077-9096 (2021).

69. Lagesen K, Hallin P, Rødland EA, Stærfeldt H-H, Rognes T, Ussery DW. RNAmmer: consistent and rapid annotation of ribosomal RNA genes. *Nucleic acids research* **35**, 3100-3108 (2007).

70. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic acids research* **28**, 27-30 (2000).

71. Galperin MY, Makarova KS, Wolf YI, Koonin EV. Expanded microbial genome coverage and improved protein family annotation in the COG database. *Nucleic acids research* **43**, D261-D269 (2015).

72. Ashburner M*, et al.* Gene ontology: tool for the unification of biology. *Nature genetics* **25**, 25-29 (2000).

73. Boeckmann B*, et al.* The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003. *Nucleic acids research* **31**, 365-370 (2003).

74. Manni M, Berkeley MR, Seppey M, Simão FA, Zdobnov EM. BUSCO update: novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes. *Molecular biology and evolution* **38**, 4647-4654 (2021).

75. Potter SC, Luciani A, Eddy SR, Park Y, Lopez R, Finn RD. HMMER web server: 2018 update. *Nucleic acids research* **46**, W200-W204 (2018).

76. Krzywinski M*, et al.* Circos: an information aesthetic for comparative genomics. *Genome Res* **19**, 1639-1645 (2009).

77. Lin Y*, et al.* quarTeT: a telomere-to-telomere toolkit for gap-free genome assembly and centromeric repeat identification. *Horticulture Research*, uhad127 (2023).

78. Xu D*, et al.* CentIER: accurate centromere identification for plant genomes with sequence specificity information. *bioRxiv*, 2023.2012. 2024.573279 (2023).

79. Goel M, Sun H, Jiao W-B, Schneeberger K. SyRI: finding genomic rearrangements and local sequence differences from whole-genome assemblies. *Genome biology* **20**, 1-13 (2019).

80. Goel M, Schneeberger K. plotsr: visualizing structural similarities and rearrangements between multiple genomes. *Bioinformatics* **38**, 2922-2926 (2022).

81. He W, Yang J, Jing Y, Xu L, Yu K, Fang X. NGenomeSyn: an easy-to-use and flexible tool for publication-ready visualization of syntenic relationships across multiple genomes. *Bioinformatics* **39**, btad121 (2023).

82. Emms DM, Kelly S. OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome biology* **20**, 1-14 (2019).

83. Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. trimAl: a tool for automated alignment

trimming in large-scale phylogenetic analyses. *Bioinformatics* **25**, 1972-1973 (2009).

84. Mendes FK, Vanderpool D, Fulton B, Hahn MW. CAFE 5 models variation in evolutionary rates among gene families. *Bioinformatics* **36**, 5516-5518 (2020).

85. Camacho C*, et al.* BLAST+: architecture and applications. *BMC bioinformatics* **10**, 1-9 (2009).

86. Marchler-Bauer A*, et al.* CDD/SPARCLE: functional classification of proteins via subfamily domain architectures. *Nucleic acids research* **45**, D200-D203 (2017).

87. Bailey TL, Johnson J, Grant CE, Noble WS. The MEME suite. *Nucleic acids research* **43**, W39-W49 (2015).

88. Wang Y*, et al.* MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic acids research* **40**, e49-e49 (2012).

89. Kim D, Paggi JM, Park C, Bennett C, Salzberg SL. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nature biotechnology* **37**, 907-915 (2019).

90. Liao Y, Smyth GK, Shi W. The R package Rsubread is easier, faster, cheaper and better for alignment and quantification of RNA sequencing reads. *Nucleic acids research* **47**, e47-e47 (2019).

91. Nie Z, Kang G, Duan C, Li Y, Dai L, Zeng R. Profiling ethylene-responsive genes expressed in the latex of the mature virgin rubber trees using cDNA microarray. *PLoS One* **11**, e0152039 (2016).

92. Patil I. Visualizations with statistical details: The'ggstatsplot'approach. *Journal of Open Source Software* **6**, 3167 (2021).

93. Wei T, Simko V, Levy M, Xie Y, Jin Y, Zemla J. Package 'corrplot'. *Statistician* **56**, e24 (2017).

94. Oksanen J*, et al.* The vegan package. *Community ecology package* **10**, 719 (2007).

95. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *bioinformatics* **26**, 139-140 (2010).

96. Kumar L, Futschik ME. Mfuzz: a software package for soft clustering of microarray data. *Bioinformation* **2**, 5 (2007).

97. Chen C*, et al.* TBtools: an integrative toolkit developed for interactive analyses of big biological data. *Molecular plant* **13**, 1194-1202 (2020).

98. Szklarczyk D*, et al.* The STRING database in 2017: quality-controlled protein–protein

association networks, made broadly accessible. *Nucleic acids research*, gkw937 (2016).

99. Shannon P*, et al.* Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* **13**, 2498-2504 (2003).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryMaterial.pdf