

# Genome-wide association study identifies candidate genes responsible for inorganic phosphorus and sucrose content in rubber tree latex

## Authors

Yitong Yan, Cuili Liang, Xing Liu, Yingchao Tan, Yilong Lu, ..., Chaorong Tang\*, Kaiye Liu\*

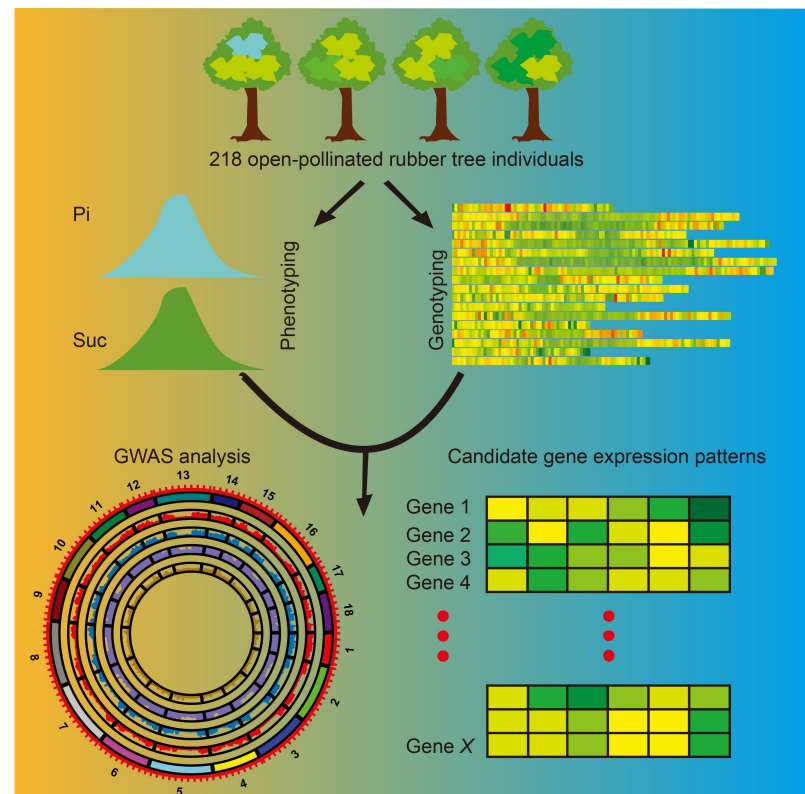
## Correspondences

[chaorongtang@hainanu.edu.cn](mailto:chaorongtang@hainanu.edu.cn);  
[kaiyeliu@hainanu.edu.cn](mailto:kaiyeliu@hainanu.edu.cn)

## In Brief

Candidate genes regulating Pi and Suc in rubber tree latex were initially identified through a genome-wide association study. This finding provides valuable insights for the utilization of latex diagnosis in breeding high-yielding and stress-adapted rubber tree clones.

## Graphical abstract



## Highlights

- Candidate genes determining Pi and Suc are first identified by GWAS in the rubber tree.
- A total of 29 and 57 candidates for Pi and Suc, respectively, are detected.
- Five Pi and three Suc candidate genes exhibit specific expression in latex.

**Citation:** Yan Y, Liang C, Liu X, Tan Y, Lu Y, et al. 2023. Genome-wide association study identifies candidate genes responsible for inorganic phosphorus and sucrose content in rubber tree latex. *Tropical Plants* 2:24 <https://doi.org/10.48130/TP-2023-0024>

# Genome-wide association study identifies candidate genes responsible for inorganic phosphorus and sucrose content in rubber tree latex

Yitong Yan<sup>1,2,3#</sup>, Cuili Liang<sup>1,2,3#</sup>, Xing Liu<sup>3#</sup>, Yingchao Tan<sup>1,2,3</sup>, Yilong Lu<sup>3</sup>, Yuanyuan Zhang<sup>4</sup>, Hongli Luo<sup>2</sup>, Chaozu He<sup>2</sup>, Jie Cao<sup>1,2,3</sup>, Chaorong Tang<sup>1,2,3,5\*</sup> and Kaiye Liu<sup>1,2,3\*</sup>

<sup>1</sup> National Key Laboratory for Biological Breeding of Tropical Crops, Hainan University, Haikou 570228, China

<sup>2</sup> Sanya Nanfan Research Institute of Hainan University, Hainan Yazhou Bay Seed Laboratory, Hainan University, Haikou 570288, China

<sup>3</sup> Natural Rubber Cooperative Innovation Center of Hainan Province and Ministry of Education of PR China, Hainan University, Haikou 570228, China

<sup>4</sup> Rubber Research Institute, Chinese Academy of Tropical Agricultural Sciences (CATAS), Danzhou, Hainan 571737, China

<sup>5</sup> Yunnan Institute of Tropical Crops, Xishuangbanna 666100, China

# These authors contributed equally: Yitong Yan, Cuili Liang, Xing Liu

\* Corresponding authors, E-mail: [chaorongtang@hainanu.edu.cn](mailto:chaorongtang@hainanu.edu.cn); [kaiyeliu@hainanu.edu.cn](mailto:kaiyeliu@hainanu.edu.cn)

## Abstract

Rubber tree (*Hevea brasiliensis*) provides the sole commercial source of natural rubber in the world. Over the past three decades, a number of latex physiological parameters has been applied to construct tapping systems, enhance natural rubber production, monitor tapping panel dryness, and evaluate rubber clone performance. Inorganic phosphorus and sucrose content (Pi and Suc) are two well-interpreted parameters to describe the latex metabolic status of rubber trees, whereas the candidate genes responsible for regulating Pi and Suc have never been reported. Here, a natural population consisting of 218 rubber tree individuals was exploited to identify candidate genes associated with Pi and Suc. The phenotype of this population was measured over a period of two years. Genome-wide association study (GWAS) was conducted using mixed linear models and 7,835,736 single-nucleotide polymorphisms (SNP). A total of 26 and 11 SNPs were significantly associated with Pi and Suc, respectively. By combining the GWAS results with genomic annotation data, 57 and 29 candidate genes within 50 kb of the significant SNPs were identified for Pi and Suc, respectively. The expression patterns of these candidates were further investigated, and five for Pi and three for Suc were identified to be specifically expressed in the latex. A DOF transcription factor and a histone H2A gene have been interpreted, while the annotations for the other candidate genes remain limited. These findings contribute to understanding the genetic foundation of Pi and Suc in the latex as well as the breeding of high-yielding rubber clones by employing latex physiological parameters.

**Citation:** Yan Y, Liang C, Liu X, Tan Y, Lu Y, et al. 2023. Genome-wide association study identifies candidate genes responsible for inorganic phosphorus and sucrose content in rubber tree latex. *Tropical Plants* 2:24 <https://doi.org/10.48130/TP-2023-0024>

## Introduction

Natural rubber (NR), primarily composed of cis-1,4-polyisoprene, is a natural polymer compound. It is considered the only renewable industrial material and strategic resource, and is often grouped with steel, petroleum, and coal as one of the four major industrial raw materials<sup>[1]</sup>. High molecular weight natural rubber possesses excellent characteristics such as elasticity, abrasion and impact resistance, efficient heat dissipation, insulation, impact resistance, and fatigue resistance<sup>[2]</sup>. As a result, it finds extensive applications in agriculture, industry, transportation, machinery manufacturing and healthcare, and cannot be replaced by synthetic alternatives<sup>[3]</sup>. Although more than 2,500 plant species are known to biosynthesize rubber-like hydrocarbon polymers, the Para rubber tree (*Hevea brasiliensis*) is the only plant widely cultivated in agricultural<sup>[4,5]</sup>. Several alternative rubber-producing crops, including *Parthenium argentatum* Gray, *Taraxacum kok-saghyz*, *T. brevicorniculatum*, and *Lactuca sativa*, have been developed, while the *Hevea* tree is the almost sole commercial source of natural rubber due to its high latex production, good quality, and long economic lifespan<sup>[3,6]</sup>. To date, over 98% of the global total natural rubber

production originates from the *Hevea* tree. Given the escalating global demand for NR, it is imperative to acquire a comprehensive understanding of the influence of harvesting systems on latex physiology to effectively optimize NR production<sup>[7]</sup>. Several latex physiological parameters, encompassing latex pH, total solid content (TSC), sucrose content (Suc), inorganic phosphorus content (Pi), thiols content (RSH), redox potential (RP), bursting index (BI), and magnesium content (Mg<sup>2+</sup>), were identified as key factors associated with both latex yield and tapping panel dryness<sup>[8]</sup>. This knowledge has facilitated the development of latex diagnosis.

Similar to a human blood test, latex diagnosis (LD) utilizes a series of physiological parameters of latex to gain valuable insights into the metabolism status of the rubber-producing laticifers. A number of LD parameters, i.e., TSC, Suc, Pi, RSH, BI, RP, and Mg<sup>2+</sup>, have been developed to effectively determine and evaluate the diagnosis of latex<sup>[8]</sup>. However, in daily practice, four main parameters, namely TSC, Suc, Pi, and RSH, are commonly employed. These parameters represent reliable indicators of the latex regeneration level, carbohydrate availability, metabolic activity, and antioxidant status<sup>[7–9]</sup>. Among them, TSC is primarily focused on converting fresh yield to dry yield,

## Pi and Suc are identified by GWAS in rubber tree

whereas RSH is commonly used to assess the level of stress encountered by the plant. Therefore, some studies have exclusively focused on the physiological parameters of Suc and Pi in latex<sup>[10,11]</sup>. A meta-analysis indicates that TSC, Suc, and RSH show a decreasing trend as stimulation intensity increases, whereas Pi demonstrates a reverse parabolic pattern<sup>[12]</sup>. After analyzing a total of 158 scientific papers, it was observed that Suc is the parameter most frequently analyzed, followed by Pi and RSH, whilst TSC was the least frequently examined parameter<sup>[13]</sup>. LD parameters are extensively employed for yield optimization in agronomy, physiological mechanism comprehension in physiology studies, as well as investigation of growth and yield-related traits in breeding studies<sup>[13]</sup>. For example, the exploration of new tapping systems, e.g., double cutting, ultrasound, and low-frequency tapping accompanied by stimulation, has been evaluated using LD parameters<sup>[10,14,15]</sup>. LD parameters are developed to detect the production of free radicals and scavengers, and to monitor the stress level in laticifers to prevent tapping panel dryness<sup>[16]</sup>. Besides, LD parameters are also utilized for the characterization of clonal typology and performance assessment of rubber clones<sup>[11,17]</sup>. At present, our understanding of the regulatory genes that determine the content of latex physiological parameters remains limited.

QTL mapping studies of rubber trees are dominated by the South American Leaf Blight (SALB), growth, and rubber yield related traits. For instance, one major QTL and four minor QTLs associated with SALB resistance were identified in the F1 population derived from the rubber tree cultivars MDF 180 and PB 260<sup>[18]</sup>. Eighteen QTLs related to growth traits, such as plant height and girth, were detected in a full-sib population resulting from the crossing of PB 217 and PR 225<sup>[19]</sup>. Additionally, 11 QTLs for stem growth and 12 QTLs for latex yield have been detected from a high-density genetic map, developed by SLAF-seq, in the F1 population crossed between the cultivated variety CATAS 879 and wild germplasm MT/C/119/67<sup>[20]</sup>. With the development of efficient and low-cost molecular markers, genome-wide association studies are conducted in rubber trees to detect growth and rubber yield traits. For example, four significant SNPs related to stem diameter are identified using GWAS in a rubber tree population genotyped by genotyping-by-sequencing<sup>[21]</sup>. In a separate study, Cheng et al.<sup>[22]</sup> performed a genome-wide association study using 147 rubber tree accessions and successfully detected 155 markers significantly associated with latex yield. The *HbPSK5*, responsible for regulating laticifer differentiation in rubber tree, was discovered through GWAS conducted on a comprehensive collection of 335 rubber tree accessions<sup>[23]</sup>.

Over the past three decades, extensive research has highlighted the profound significance of LD parameters in rubber production, with particular emphasis on Suc and Pi<sup>[13]</sup>. However, there is a dearth of literature exploring the candidate genes that determine Suc and Pi in latex. Here, a natural population, comprising 218 open-pollinated progenies, was constructed and employed whole-genome resequencing to obtain high-density SNP markers. We performed a genome-wide association study for Suc and Pi, leading to the identification of 29 and 57 candidate genes, respectively. Furthermore, the expression patterns analysis revealed that five candidate genes for Pi and three candidate genes for Suc are latex-specific. Our findings provide valuable insights for the future utilization of LD in breeding high-yielding and stress-adapted rubber tree clones.

## Materials and methods

### Population materials

In 2014, more than 2,000 seeds were randomly collected from a field garden that planted 400 Whickham *Hevea* germplasms at National Tropical Plants Germplasm Resource Center-Rubber Tree (Danzhou, Hainan province, China). Subsequently, the seeds were germinated in the nursery garden for one month and transplanted to the experimental farm of Hainan University (Danzhou, Hainan province, China). The row and plant spacing were 1.5 m and 3.0 m, respectively. In 2020, adhering to the technical code of practice for rubber tree breeding (NY/T 607-2018), the tapping panel was opened at 50 cm above the ground, and tapped by the half spiral cut, third-daily (S/2 d/3) system<sup>[24]</sup>. A total of ten tappings were conducted annually from October to November, and the latex samples were harvested and examined for the last five tappings. In this study, 218 accessions exhibiting continuous variations in rubber yield, growth, Pi, and Suc were selected for genotyping (Supplemental Table S1) and the latex physiological parameters were measured in 2021 and 2022 (seventh and eighth years after planting).

### Phenotyping

To quantify the contents of inorganic phosphorus and sucrose, 1 mL of fresh latex was mixed with 9 mL of 2.5% trichloroacetic acid (2.5% w/v) for each tapping, centrifuged at 5,000 g for 10 min at 4 °C, and the aqueous solution was then separated and stored at -20 °C before testing. Inorganic phosphate was quantified using the binding principle of ammonium molybdate, followed by reduction with FeSO<sub>4</sub> in an acidic reaction. The samples underwent a color change to blue and were quantitatively analyzed utilizing the INFINITE 200 PRO TECAN (TECAN, Switzerland) spectrophotometer. The absorbance was measured at  $\lambda$ 627 nm, following the established methodology outlined by Tausky & Shorr<sup>[25]</sup>. The sucrose undergoes dehydration upon being heated in concentrated sulfuric acid (70% H<sub>2</sub>SO<sub>4</sub>), resulting in the formation of a furfural derivative. This derivative then reacts with anthrone, resulting in a blue reaction. Subsequently, the absorbance at  $\lambda$ 627 nm was measured using the INFINITE 200 PRO TECAN (TECAN, Switzerland), following Dische's anthrone method<sup>[26]</sup>. All measures were conducted at least three replicates for each tapping, and the phenotypic values were averaged from five tappings. The BLUP (best linear unbiased prediction) values across two environments (2021 and 2022 years) were conducted using the R package 'lme4'<sup>[27]</sup>. Histogram plots were depicted using the 'hist' function in R software to show the phenotypic distributions under each environment.

### Sequencing and polymorphism analysis

Genomic DNA was extracted from young leaves of six-year-old plants of each accession using the cetyl trimethyl ammonium bromide method<sup>[28]</sup>. The quality and quantity of DNA were examined using the INFINITE 200 PRO TECAN (TECAN, Switzerland) and agarose gel electrophoresis. High-quality DNA with an absorbance ratio of 260/280 between 1.8 and 2.0 and a concentration greater than 200 ng/ $\mu$ L were utilized for library construction. Whole-genome resequencing was performed by using the DNBSEQ-T7 sequencing platform to generate 150-bp paired-end reads following the manufacturer's instructions. The resequencing data of each individual was filtered using Trimmomatic V.0.39 to eliminate low-quality reads and aligned to

the reference genome of RY 73397 using BWA-MEM software with the following parameters:  $-t\ 40\ -M\ -R^{[29-31]}$ . The BAM files were sorted, indexed, and duplicates removed using the SAMtools software<sup>[32]</sup>. Variant calling was conducted utilizing HaplotypeCaller, CombineGVCFs, GenotypeGVCFs, and SelectVariant from the Genome Analysis Toolkit (GATK) version 4.1.8.1<sup>[33]</sup>. SNPs were assessed using the following parameters: 'QD < 2.0, MQ < 40.0, FS > 60.0, SOR > 3.0, MQRankSum < -12.5, ReadPosRankSum < -8.0'. The resulting variants were further filtered using VCFtools with the parameters 'max-missing 0.8, maf 0.05'<sup>[34]</sup>. The SNP density map was generated utilizing the R package 'CMplot'<sup>[35]</sup>.

### Genome-wide association study

GWAS was conducted on 218 open-pollinated rubber tree progenies, utilizing a dataset consisting of 7,835,736 high-confidence SNPs. A genome-wide efficient mixed-model analysis was carried out using the GEMMA software to assess the associations between the traits and SNPs<sup>[36]</sup>. To minimize false-positive rates, the population structure was appropriately adjusted by incorporating the first three principal components and a kinship matrix as covariates. Manhattan and quantile-quantile (QQ) plots were constructed using the R package 'CMplot', and the genome-wide significance threshold in this study was set at  $p < 1e-6$ . Linkage disequilibrium (LD) and haplotype maps for the most significant SNP associated with Pi and Suc were generated using the 'LDBlockShow' software<sup>[37]</sup>.

### Expression analysis by RNA-seq

A 50 kb region upstream and downstream of the significant SNPs was selected, and the genes encompassed within this interval being meticulously designated as candidates. Total RNA was extracted from the bark and xylem of RY 73397 using the TRNzol Universal Reagent (Catalog Number 4992730) from TIANGEN Biotech Co., Ltd. (Beijing, China) according to the manufacturer's instructions. The quality and integrity of RNA

were evaluated using the INFINITE 200 PRO (TECAN, Switzerland) and Agilent 2100 Bioanalyzer (Agilent Technologies, CA, USA). The mRNA libraries were constructed by using the NEBNext Ultra RNA Library Prep Kit following the manufacturer's instructions, and sequenced by the Illumina NovaSeqTM 6000 platform to generate 150-bp paired-end reads. The RNA-seq data from leaf and latex of RY 73397 were collected from our previous published paper<sup>[31]</sup>. Transcriptomic analysis was conducted by employing the HISAT2-Stringtie pipeline, while gene expression levels were normalized as FPKM (fragments per kilobase of transcript per million mapped reads). The heatmap of gene expression was visualized using the R package 'pheatmap'.

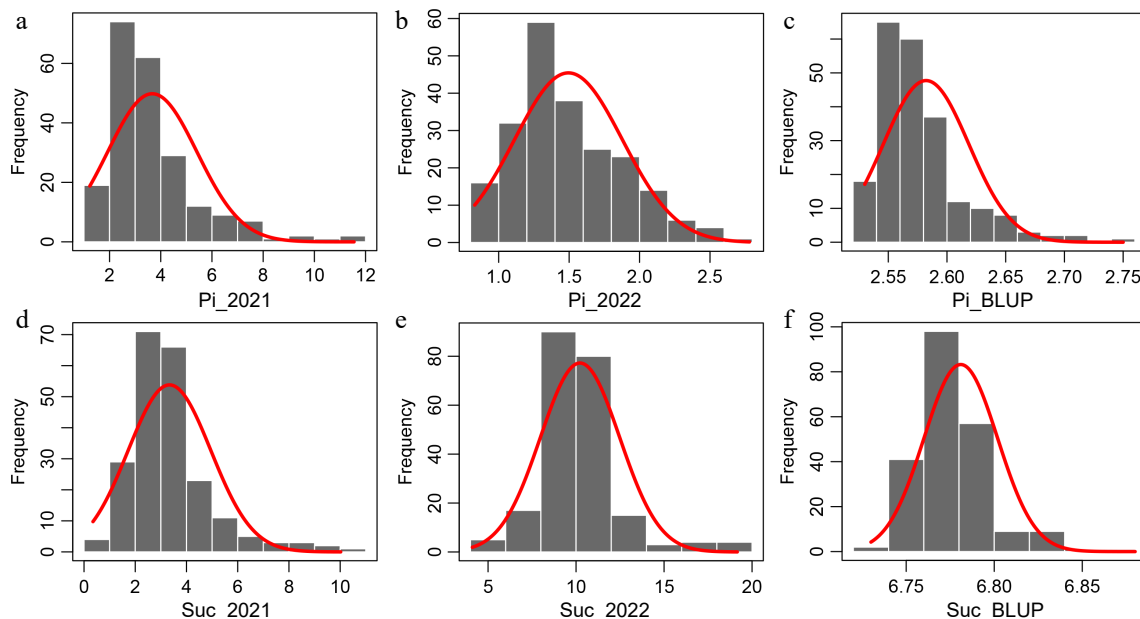
## Results

### Phenotypic variation of inorganic phosphorus and sucrose content

The Pi and Suc for this natural population in 2021 and 2022 are displayed in Fig. 1 and Supplemental Table S2. The Pi ranged from 0.83 to 11.55 mM, while the Suc ranged from 0.35 to 19.16 mM. Upon comparing the data spanning two years, we observed a significant increase in the mean value of Suc in 2022 compared to 2021. Conversely, the mean value of Pi significantly diminished in the same period (Supplemental Table S2). The statistical results of skewness and kurtosis suggested a non-normal distribution for both Suc and Pi (except in 2022) in this natural population (Supplemental Table S2). Moreover, the frequency distribution histograms of Pi and Suc also revealed non-normal distribution patterns (Fig. 1). The BLUP values of Pi and Suc were computed across two environments for subsequent analysis (Supplemental Table S2).

### Distribution of SNPs

The resequencing data were genotyped using GATK toolkit and subsequently filtered using VCFtools. An extensive analysis



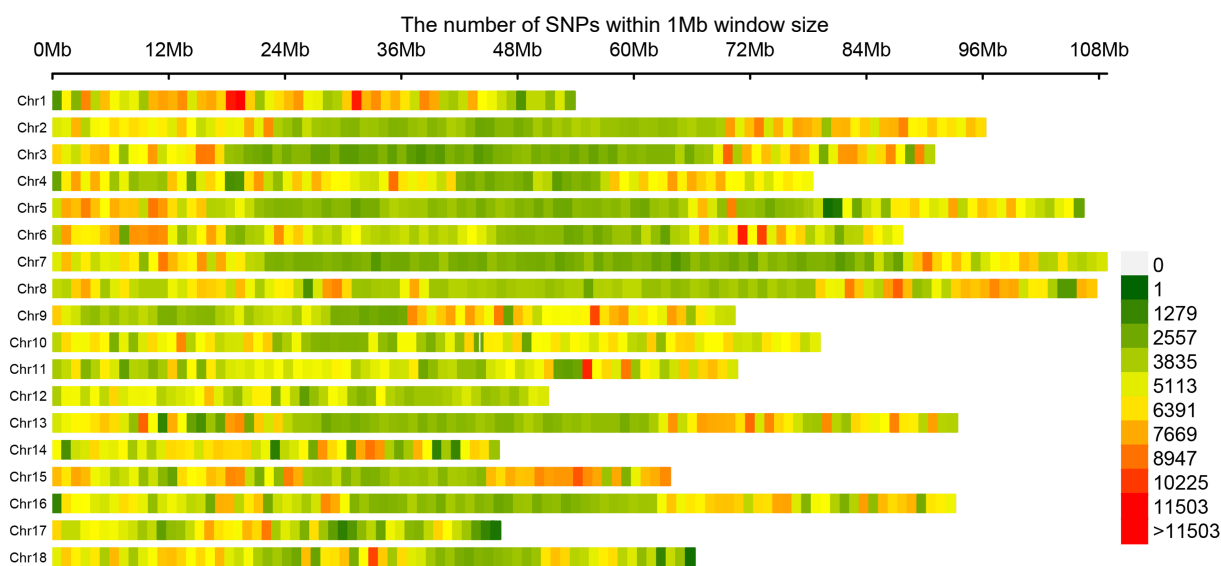
**Fig. 1** Frequency distributions of Pi and Suc in the natural population. (a), (b) Frequency distribution of Pi in 2021 and 2022. (c) Frequency distribution of BLUP data for Pi. (d), (e) Frequency distribution of Suc in 2021 and 2022. (f) Frequency distribution of BLUP data for Suc. The curves represent a fit of a normal distribution to the measured distribution.

## Pi and Suc are identified by GWAS in rubber tree

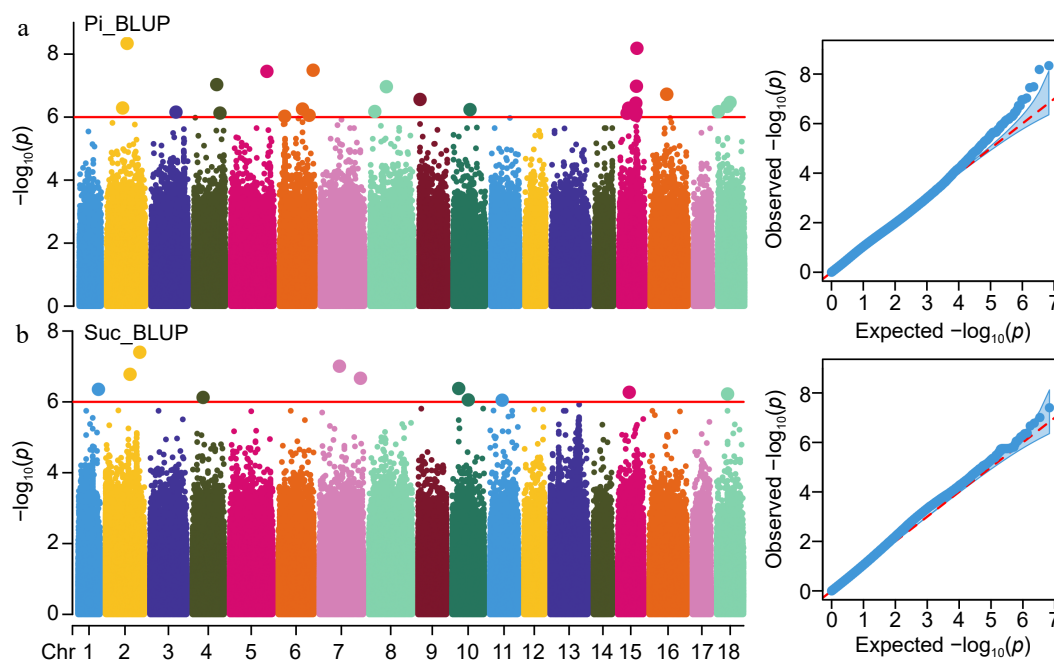
of population structure and genetic diversity using various methods, including linkage disequilibrium decay, principal component analysis, phylogenetic tree analysis, nucleotide diversity ( $\pi$ ), and Tajima's D values, were conducted by Liang et al. (Unpublished). Finally, a total of 7,835,736 high-confidence SNPs were identified from this natural population of 218 individuals. We quantified the frequency of SNPs per mega-base (Mb) interval and visually represented it through a heatmap illustrating SNP density (Fig. 2). We observed a non-uniform distribution of SNPs across chromosomes, with chromosomes 1, 14, and 17 displaying no discernible distribution pattern. However, other chromosomes exhibited a noticeable trend of high SNP density at both ends and lower density towards the middle (Fig. 2).

## GWAS for Pi and Suc

The association between the SNP markers and the two latex physiological parameters traits was investigated using a mixed linear model (MLM) that controlled for the population structure and kinship coefficient of varieties. In total, 37 SNPs significantly correlated with the Pi and Suc traits were detected using the BLUP data. They were scattered across 14 of the 18 chromosomes in the genome of RY 73397. Significant marker-trait associations were identified by negative  $\log_{10}$  ( $p$  value) and the results have been presented in form of Manhattan and QQ plots. A total of 34 and two SNPs were found significantly associated with Pi in 2021 and 2022, respectively (Supplemental Fig. S1). Additionally, 26 SNPs displayed a significant association with Pi by employing the BLUP data (Fig. 3a). In the years 2021



**Fig. 2** Genome-wide profiling of SNP density in the natural population. The number of SNPs is counted in 1 Mb intervals, and the color is based on the counted values.



**Fig. 3** Manhattan and QQ plots depict the GWAS results of Pi and Suc. (a), (b) GWAS results of Pi and Suc traits using BLUP data. Red lines represent a  $p$ -value of  $1e-6$  and the dot size of significant SNPs is magnified two-fold.

and 2022, we identified 14 and 44 SNPs, respectively, that exhibited significant correlations with Suc (Supplemental Fig. S2). Meanwhile, 11 SNPs were found to be significantly correlated with Suc by using BLUP data (Fig. 3b).

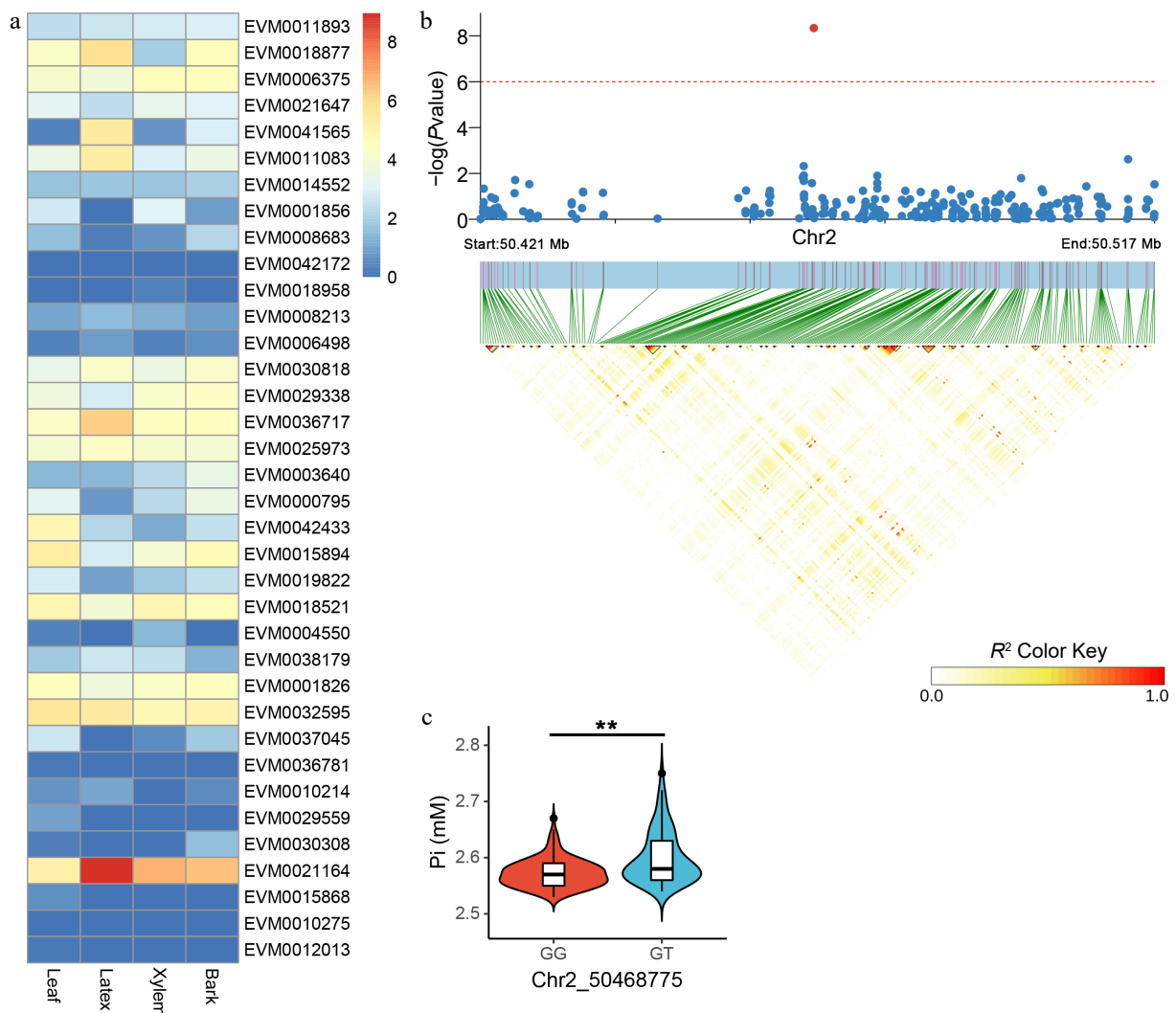
**Candidate genes identification and expression analysis**

In this study, the significant SNPs identified from the BLUP data were utilized to conduct a more in-depth analysis of candidate genes. The genomic region spanning a 100 kb interval of the significant SNPs was precisely captured and compared against the genome annotation file. The expression levels of these candidate genes were quantified using RNA sequencing in four *Hevea* tissues, including leaf, latex, xylem, and bark. Simultaneously, candidate genes that exhibited no expression across the four aforementioned tissues were excluded.

We further performed gene annotation using the association analysis results obtained from the BLUP data, and a total of 57 candidate genes for Pi were found (Supplemental Table S3). Five candidate genes (*EVM0018877*, *EVM0041565*, *EVM0011083*, *EVM0036717*, and *EVM0021164*), which are specifically and

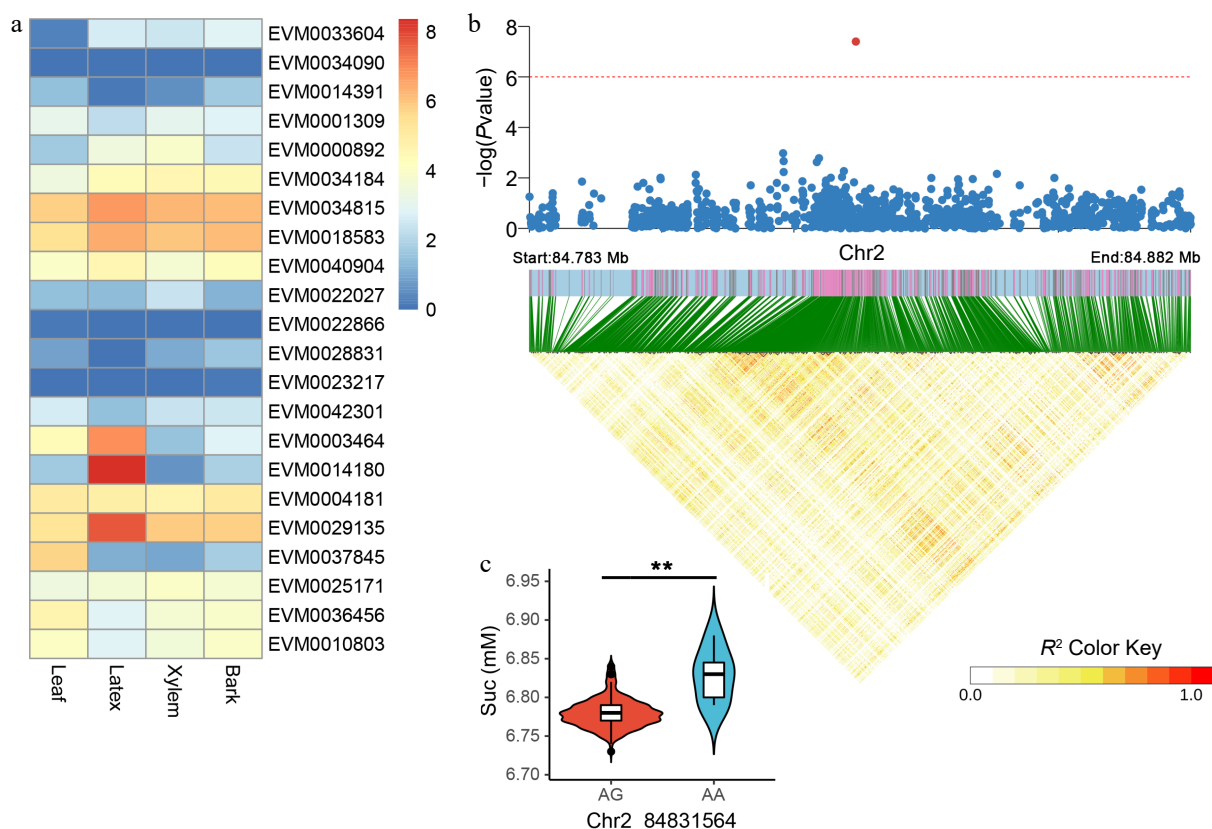
highly expressed in latex and implicated in the regulation of Pi, were identified (Fig. 4a). The SNP most significantly associated with Pi was found on chromosome 2, and the linkage disequilibrium block analysis revealed the absence of a robust linkage region for this specific SNP (Fig. 4b). Nevertheless, in this population, the latex Pi of G/T heterozygous was significantly higher than that of G/G homozygous (Fig. 4c). Furthermore, a candidate gene (*EVM0003477*) was identified for this SNP, whereas its expression was not detected in any of the four tissues examined. Functional annotation of latex-specific candidates revealed that the DOF transcription factor (*EVM0041565*) was the sole gene with a well-defined function, whereas multiple other genes lacked evident homologs.

We identified a total of 29 candidate genes that are correlated with Suc, corresponding to 11 significantly associated SNPs. Of which, *EVM0003464*, *EVM0014180*, and *EVM0029135* exhibited a notable and specific expression pattern in latex (Fig. 5a). The Suc-associated SNP on chromosome 2 displayed the highest level of significance, and the linkage disequilibrium block analysis revealed a closely linked region for this specific



**Fig. 4** Candidate gene expression patterns and LD block analysis for Pi-associated variants. (a) Quantification of candidate gene expression levels related to Pi in the leaf, latex, xylem, and bark tissues of RY73397. The color is based on the  $\log_2(\text{FPKM} + 1)$  values. (b) LD block analysis of the most significant SNP correlated with Pi. The red line represents a  $p$ -value of  $1e-6$ . (c) Comparison of Pi between GG ( $n = 143$ ) and GT ( $n = 75$ ) haplotypes. \*\* indicates  $p < 0.01$ .

## Pi and Suc are identified by GWAS in rubber tree



**Fig. 5** Candidate gene expression patterns and LD block analysis for Suc-associated variants. (a) Quantification of candidate gene expression levels related to Suc in the leaf, latex, xylem, and bark tissues of RY73397. The color is based on the  $\log_2(\text{FPKM} + 1)$  values. (b) LD block analysis of the most significant SNP correlated with Suc. The red line represents a  $p$ -value of  $1e-6$ . (c) Comparison of Suc between AG ( $n = 212$ ) and AA ( $n = 6$ ) haplotypes. \*\* indicates  $p < 0.01$ .

SNP (Fig. 5b). The latex Suc of A/A homozygous was significantly higher than that of A/G heterozygous in this natural population (Fig. 5c). Three candidate genes (*EVM0028831*, *EVM0021953*, and *EVM0035051*) were identified for this SNP, of which two were observed to have no expression (Fig. 5a, Supplemental Table S3). Functional annotation of the outstanding candidates for latex expression revealed that *EVM0029135* corresponds to a histone H2A gene, while the functions of the other candidate genes remain unclear.

## Discussion

The LD was utilized to check the physiological status at one moment of the rubber tree, and frequently used in physiological and breeding studies. Several physiological parameters correlated with latex yield and tapping panel dryness have been developed. By systematically collecting and analyzing a total of 158 articles published from 1970 to 2020, it was detected that Suc and Pi stand out as the two most extensively investigated parameters<sup>[13]</sup>. Moreover, they found a wide range of variation in these two parameters. For the Pi, the minimum value was 0.15 mM and the maximum was 36.3 mM, the average was 16.3 mM<sup>[13,38]</sup>. The Suc displayed a minimum value of approximately 0.4 mM, while the non-outlier maximum observation reached 20.2 mM, with an average of 9.2 mM<sup>[13,39]</sup>. In this study, we measured Suc and Pi for two consecutive years. Notably, both Pi and Suc had significantly changed over the two-year period. In 2022, the Pi exhibited a substantial

decrease whereas the Suc demonstrated a significant increase, compared to the data of 2021 (Supplemental Table S2). These results suggest that the physiological parameters of Pi and Suc serve as direct indicators of the growth state of rubber trees, and their levels may change significantly and rapidly in response to various factors such as tree age, climate conditions, tapping stimulation, and panel management.

GWAS aims to identify associations of genotypes with phenotypes by testing hundreds of thousands of genetic variants among individuals who share ancestral similarities but exhibit phenotypic differences<sup>[40]</sup>. Population structure, marker density, and population size are crucial indicators that play a significant role in enhancing the resolution of GWAS. With the advancement of techniques such as next-generation sequencing (NGS), a substantial number of SNP markers have been identified, resulting in a notable improvement in marker density. In recent years, there has been a growing interest in conducting GWAS on rubber trees, primarily focused on investigating yield and growth-related traits<sup>[22,23]</sup>. Our study represents the pioneering effort to identify candidate genes associated with the regulation of Pi and Suc using GWAS. The ultimate objective is to provide valuable insights for optimizing the tapping systems of rubber trees through the effective modulation of latex physiological parameters. In this study, we obtained high-density SNPs by resequencing the whole genome with an average depth of approximately 15x. Finally, 7,835,736 SNPs were obtained after quality control and filtering (Fig. 2). The abundance of SNP markers presents an

exceptional foundation for conducting comprehensive genome-wide association study. Moreover, the genome sequences of various rubber tree varieties are consistently updated and released. Several candidate genes, including *COBL4* for girth and *HbPSK5* for laticifer differentiation, have been identified through GWAS conducted on rubber trees<sup>[23,41]</sup>. The GWAS results for Pi over a span of two years revealed a notable lack of reproducibility. Specifically, in 2022, only two significant loci were detected on chromosome 4 (Supplemental Fig. S1). Meanwhile, 14 and 44 SNPs were found to be significantly associated with Suc in 2021 and 2022, respectively, (Supplemental Fig. S2). These findings imply that the substantial variations in phenotypes across different years led to significant disparities in the outcomes of the association analysis. Hence, a BLUP analysis on the phenotypes across two environments was performed to facilitate subsequent association analysis.

In this study, we determined a range of candidate genes by annotating of SNPs that exhibited significant associations with the BLUP data and conducted a comprehensive transcriptome analysis to further screen these candidate genes. Among the 57 candidate genes identified as responsible for regulating Pi in the natural population, only 36 genes were found to be expressed in at least one of the four tissues (Supplemental Table S3). Meanwhile, 22 out of the 29 candidate genes involved in Suc exhibited detectable expression levels. Eventually, five genes for Pi and three genes for Suc were identified with a predominant expression in latex by analyzing the expression levels of these candidate genes in leaf, latex, bark and xylem (Figs 4 & 5). Laticifer cells are a type of special cell that has evolved in rubber-producing plants and have obvious difference and differentiation from other cell types<sup>[42]</sup>. We speculated that the genes governing the regulation of physiological parameters in latex predominantly manifest their expression and functionality within the latex. Therefore, we identified these candidate genes with latex-specific expression to provide significant insights for further verification. Gene function annotation provides further evidence that these genes, which are specifically expressed in latex, do not exhibit significant homology in other species, and only two candidate genes have annotated information, while the remaining genes lack any available annotations. These findings offer valuable insights for harnessing the potential of LD in the development of high-yielding and stress-resistant rubber tree clones.

## Author contributions

The authors confirm contribution to the paper as follows: study conception and design: Liu K, Tang C; data collection: Yan Y, Liang C, Tan Y, Lu Y, Zhang Y, Luo H, He C; analysis and interpretation of results: Liu K, Yan Y, Cao J; draft manuscript preparation: Tang C, Liu K. All authors reviewed the results and approved the final version of the manuscript.

## Data availability

The raw sequencing data have been deposited in the Genome Sequence Archive in National Genomics Data Center (<https://ngdc.cncb.ac.cn/>) under accession number PRJCA020685.

## Acknowledgments

This work was supported by the National Natural Science Foundation of China (Grant Nos 32160383, 32201450 and 31825007), the Hainan Provincial Natural Science Foundation of China (Grant No. 322RC577), the Yunnan Academician and Expert Workstation Project (202105AF150017), and the Starting Research Fund of Hainan University. We would like to thank the other staff in the Laboratory of Rubber Production Biology (LRPB) at Hainan University for their technical assistance.

## Conflict of interest

The authors declare that they have no conflict of interest.

**Supplementary Information** accompanies this paper at (<https://www.maxapress.com/article/doi/10.48130/TP-2023-0024>)

## Dates

Received 4 September 2023; Accepted 13 November 2023; Published online 22 November 2023

## References

1. Cherian S, Ryu SB, Cornish K. 2019. Natural rubber biosynthesis in plants, the rubber transferase complex, and metabolic engineering progress and prospects. *Plant Biotechnology Journal* 17:2041–61
2. Cornish K. 2001. Biochemistry of natural rubber, a vital raw material, emphasizing biosynthetic rate, molecular weight and compartmentalization, in evolutionarily divergent plant species. *Natural Product Reports* 18:182–89
3. Salehi M, Cornish K, Bahmankar M, Naghavi MR. 2021. Natural rubber-producing sources, systems, and perspectives for breeding and biotechnology studies of *Taraxacum kok-saghyz*. *Industrial Crops and Products* 170:113667
4. van Beilen JB, Poirier Y. 2007. Establishment of new crops for the production of natural rubber. *Trends in Biotechnology* 25:522–29
5. Metcalfe CR. 1967. Distribution of latex in the plant kingdom. *Economic Botany* 21:115–27
6. Yang N, Yang D, Yu X, Xu C. 2023. Multi-omics-driven development of alternative crops for natural rubber production. *Journal of Integrative Agriculture* 22:959–71
7. d'Auzac J, Jacob JL, Chrestin HE. 1989. *Physiology of Rubber Tree Latex*. Boca Raton: CRC Press. pp. 345–82.
8. Jacob JL, Eschbach JM, Prévot JL, Roussel D, Lacroette R, et al. 1985. Physiological basis for latex diagnosis of the functioning of the laticiferous system in rubber trees. *International Rubber Conference, Kuala Lumpur, Malaysia, 1985*. pp. 43–68. [https://horizon.documentation.ird.fr/exl-doc/pleins\\_textes/divers20-05/010028013.pdf](https://horizon.documentation.ird.fr/exl-doc/pleins_textes/divers20-05/010028013.pdf)
9. Eschbach JM, Roussel D, Sype HVD, Jacob JL, D'auzac J. 1984. Relationship between yield and clonal physiological characteristics of latex from *Hevea brasiliensis*. *Physiologie végétale* 22:295–304
10. Chantuma P, Lacote R, Leconte A, Gohet E. 2011. An innovative tapping system, the double cut alternative, to improve the yield of *Hevea brasiliensis* in Thai rubber plantations. *Field Crops Research* 121:416–22
11. Phearun P, Chetha P, Bunthuon H, Lacote R, Chhek C, et al. 2016. Early clonal selection of *Hevea brasiliensis* based on latex physiological parameters in Cambodia. *International Rubber Conference, Siem Reap, Cambodia, 2016*. pp. 286–97. <https://agritrop.cirad.fr/582388/>
12. Gohet E, Cauchy T, Soumahoro M, Kotochi C, Chegbene P, et al. 2019. Meta-analysis of a large industrial latex diagnosis database



## Pi and Suc are identified by GWAS in rubber tree

- provides insight on *Hevea brasiliensis* clonal adaptation and site-specific yield potential in Western Africa. *International Rubber Conference, Nay Pyi Taw, Myanmar, 2019*. pp. 1–19. <https://agritrop.cirad.fr/593887/1/Gohet%20et%20al.%202019.pdf>
13. Junaidi J, Clément-Vidal A, Nuringtyas TR, Gohet E, Subandiyah S, Montoro P. 2022. A meta-analysis of latex physiology studies reveals limited adoption and difficulties to interpret some latex diagnosis parameters in *Hevea brasiliensis*. *HAYATI Journal of Biosciences* 30:358–71
  14. She F, Zhu D, Kong L, Wang J, An F, et al. 2013. Ultrasound-assisted tapping of latex from Para rubber tree *Hevea brasiliensis*. *Industrial Crops and Products* 50:803–8
  15. Sainoi T, Sdoodee S, Lacote R, Gohet E, Chantuma P. 2017. Stimulation affecting latex physiology and yield under low frequency tapping of rubber (*Hevea brasiliensis*) clone RRIM 600 in southern Thailand. *Australian Journal of Crop Science* 11:220–27
  16. Tistama R, Siregar P, Ade-Fipriani L, Junaidi J. 2019. Physiological status of high and low metabolism *Hevea* clones in the difference stage of tapping panel dryness. *Biodiversitas* 20:267–73
  17. Cardoso SEA, Freitas TA, Silva DdC, Gouvêa LRL, Gonçalves PdS, et al. 2014. Comparison of growth, yield and related traits of resistant *Hevea* genotypes under high South American leaf blight pressure. *Industrial Crops and Products* 53:337–49
  18. Le Guen V, Garcia D, Doaré F, Mattos CRR, Condina V, et al. 2011. A rubber tree's durable resistance to *Microcyclus ulei* is conferred by a qualitative gene and a major quantitative resistance factor. *Tree Genetics & Genomes* 7:877–89
  19. Souza LM, Gazaffi R, Mantello CC, Silva CC, Garcia D, et al. 2013. QTL mapping of growth-related traits in a full-sib family of rubber tree (*Hevea brasiliensis*) evaluated in a sub-tropical climate. *PLoS ONE* 8:e61238
  20. An Z, Zhao Y, Zhang X, Huang X, Hu Y, et al. 2019. A high-density genetic map and QTL mapping on growth and latex yield-related traits in *Hevea brasiliensis* Müll. Arg. *Industrial Crops and Products* 132:440–48
  21. Francisco FR, Aono AH, Da Silva CC, Gonçalves PS, Scaloppi Junior EJ, et al. 2021. Unravelling rubber tree growth by integrating GWAS and biological network-based approaches. *Frontiers in Plant Science* 12:768589
  22. Cheng H, Song X, Hu Y, Wu T, Yang Q, et al. 2023. Chromosome-level wild *Hevea brasiliensis* genome provides new tools for genomic-assisted breeding and valuable loci to elevate rubber yield. *Plant Biotechnology Journal* 21:1058–72
  23. Chao J, Wu S, Shi M, Xu X, Gao Q, et al. 2023. Genomic insight into domestication of rubber tree. *Nature Communications* 14:4651
  24. Gouvêa LRL, Silva GAP, Verardi CK, de Oliveira ALB, Gonçalves ECP, et al. 2013. Rubber tree early selection for yield stability in time and among locations. *Euphytica* 191:365–73
  25. Taussky HH, Shorr E. 1953. A microcolorimetric method for the determination of inorganic phosphate. *The Journal of biological chemistry* 202:675–85
  26. Hedge JE, Hofreiter BT. 1962. Carbohydrate Chemistry. New York: Academic Press
  27. Bates D, Mächler M, Bolker B, Walker S. 2015. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 67:1–48
  28. Stewart Jr. C, Via LE. 1993. A rapid CTAB DNA isolation technique useful for RAPD fingerprinting and other PCR applications. *BioTechniques* 14(5):748–50
  29. Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30:2114–20
  30. Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25:1754–60
  31. Tang C, Yang M, Fang Y, Luo Y, Gao S, et al. 2016. The rubber tree genome reveals new insights into rubber production and species adaptation. *Nature Plants* 2:16073
  32. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, et al. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25:2078–79
  33. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, et al. 2010. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research* 20:1297–303
  34. Danecek P, Auton A, Abecasis G, Albers CA, Banks E, et al. 2011. The variant call format and VCFtools. *Bioinformatics* 27:2156–58
  35. Yin L, Zhang H, Tang Z, Xu J, Yin D, et al. 2021. rMVP: a memory-efficient, visualization-enhanced, and parallel-accelerated tool for genome-wide Association study. *Genomics, Proteomics & Bioinformatics* 19:619–28
  36. Zhou X, Stephens M. 2012. Genome-wide efficient mixed-model analysis for association studies. *Nature Genetics* 44:821–24
  37. Dong SS, He WM, Ji JJ, Zhang C, Guo Y, et al. 2021. LDBlockShow: a fast and convenient tool for visualizing linkage disequilibrium and haplotype blocks based on variant call format files. *Briefings in Bioinformatics* 22:bbaa227
  38. Anushka PVA, Withanage SP, Karunaratne NPSN, Kudaligama KVV, Dahanayake TTD, et al. 2019. Assessment and selection based on girth and yield performance of new *Hevea* genotypes generated from controlled hybridization. *Journal of the Rubber Research Institute of Sri Lanka* 99:126–41
  39. Putranto RA, Herlinawati E, Rio M, Leclercq J, Piyatrakul P, et al. 2015. Involvement of ethylene in the latex metabolism and tapping panel dryness of *Hevea brasiliensis*. *International Journal of Molecular Sciences* 16:17885–908
  40. Uffelmann E, Huang QQ, Munung NS, de Vries J, Okada Y, et al. 2021. Genome-wide association studies. *Nature Reviews Methods Primers* 1:59
  41. Bhusudsawang G, Rattanawong R, Phumichai T, Pootakham W, Tangphatsornruang S, et al. 2021. Identification of candidate gene-based markers for girth growth in rubber trees. *Plants* 10:1440
  42. Castelblanque L, Balaguer B, Martí C, Rodríguez JJ, Orozco M, et al. 2016. Novel insights into the organization of laticifer cells: a cell comprising a unified whole system. *Plant Physiology* 172:1032–44



Copyright: © 2023 by the author(s). Published by Maximum Academic Press on behalf of Hainan University. This article is an open access article distributed under Creative Commons Attribution License (CC BY 4.0), visit <https://creativecommons.org/licenses/by/4.0/>.