Tropical Plants

ARTICLE

Open Access

https://doi.org/10.48130/tp-0024-0020

Tropical Plants **2024**, 3: e014

Design and application of the HbGBTS80K liquid chip in rubber tree

Authors

Jinquan Chao#, Yan Li#, Shuguang Yang, Xingming Chen, Yilin He, ..., Qiang Gao*, Weimin Tian*

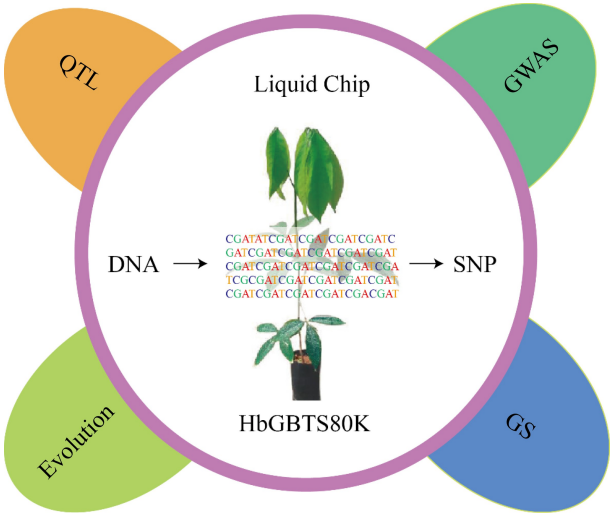
Correspondence

qgao@qi-biodesign.com; wmtian@163.com

In Brief

Development of high-quality liquid chip in rubber tree.

Graphical abstract



Highlights

- A high-quality HbGBTS80K liquid chip was designed based on the 335 re-sequenced germplasms.
- HbGBTS80K performed well in rubber tree population genetic diversity analysis.
- HbGBTS80K performed well in rubber tree GWAS analysis.

Citation: Chao J, Li Y, Yang S, Chen X, He Y, et al. 2024. Design and application of the HbGBTS80K liquid chip in rubber tree. *Tropical Plants* 3: e014 https://doi.org/10.48130/tp-0024-0020

Tropical Plants

ARTICLE

Open Access

https://doi.org/10.48130/tp-0024-0020

Tropical Plants **2024**, 3: e014

Design and application of the HbGBTS80K liquid chip in rubber tree

Jinquan Chao[1,2#], Yan Li[1#], Shuguang Yang[1], Xingming Chen[3], Yilin He[3], Jianan Zhang[3], Zongliang Yue[4], Qiang Gao[5*] and Weimin Tian[1*]

[1] *Ministry of Agriculture and Rural Affairs Key Laboratory of Biology and Genetic Resources of Rubber Tree, State Key Laboratory Breeding Base of Cultivation and Physiology for Tropical Crops, Rubber Research Institute, Chinese Academy of Tropical Agricultural Sciences, Haikou 571101, China*

[2] *National Key Laboratory for Tropical Crop Breeding, Sanya Research Institute, Chinese Academy of Tropical Agricultural Sciences, Sanya 572024, China*

[3] *Molbreeding Biotechnology Company, Shijiazhuang 050000, China*

[4] *Health Outcome Research and Policy Department, Harrison College of Pharmacy, Auburn University, Auburn 36849, USA*

[5] *Qi Biodesign, Beijing 100000, China*

[#] Authors contributed equally: Jinquan Chao, Yan Li

* Corresponding authors, E-mail: qgao@qi-biodesign.com; wmtian@163.com

Abstract

The single nucleotide polymorphism (SNP) chip provides a convenient platform for accelerating breeding progress and promoting basic research in the field of crop science. Given the long and inefficient traditional breeding of the rubber tree, it is urgent to develop an SNP chip to promote changes in breeding patterns. Here, we developed and validated a liquid SNP chip named 'HbGBTS80K'. The SNPs were selected from whole-genome resequencing data containing 335 diverse rubber tree accessions. After quality control of 69 additional accessions, the final SNP chip included 80,080 SNPs evenly distributed across 18 chromosomes. Population genetic diversity analysis showed that the HbGBTS80K was effective in distinguishing rubber accessions into four groups, which was highly consistent with previous findings based on resequencing data. Genome-wide association study (GWAS) on the reported genotypic data of the number of laticifer rings (NLR) could also detect the major gene *HbPSK5* in the HbGBTS80K platform. In summary, the HbGBTS80K liquid SNP chip is a valuable tool that will facilitate functional studies and molecular breeding of the rubber tree.

Citation: Chao J, Li Y, Yang S, Chen X, He Y, et al. 2024. Design and application of the HbGBTS80K liquid chip in rubber tree. *Tropical Plants* 3: e014 https://doi.org/10.48130/tp-0024-0020

Introduction

Molecular markers, also known as genetic markers, refer to specific DNA fragments that can reflect the differences among biological individuals. Developing molecular markers closely related to the target traits is the prerequisite for molecular marker-assisted breeding. Traditional molecular markers such as restriction fragment length polymorphism (RFLP)[1], random amplified polymorphic DNA (RAPD)[2], amplified fragment length polymorphism (AFLP)[3], and simple sequence repeats (SSR)[4] are limited in their application in modern breeding due to their relatively low coverage of the genome. Single nucleotide polymorphisms (SNPs) refer to the DNA sequence polymorphisms caused by the variation of a single nucleotide at the genomic level. As SNP is the most common type of genetic variation, it has evolved into one of the most important molecular markers in the study of plant genetic variation in recent years[5–7].

SNP-based gene chips are a useful platform for genetic diversity, evolutionary analysis, target gene identification, and more. Traditional gene chips mainly refer to solid-phase chips, which work by in-situ synthesizing oligonucleotides in an ordered manner on a solid support, and then hybridizing with labeled samples to be tested. A fluorescence detection system scans the chip and genetic information of the sample to be tested is obtained through software analysis. Solid-phase chips have achieved significant development in the field of molecular breeding in the past few decades[6,8–10]. However, due to their own shortcomings, such as the high cost of developing and applying loci and the physical fixation model of target loci, their application is increasingly limited. Recently, researchers have developed liquid-phase chips represented by SNP-based genotyping by targeted sequencing (GBTS)[11–13]. Compared with solid-phase chips, liquid-phase chips are more flexible, more cost-effective, and have lower requirements on the detection platform, and have been applied in cotton[7,12], soybeans[14], cucumbers[6], rice[15,16], etc.

Natural rubber (NR) is an essential industrial raw material. Over 2,000 plant species worldwide can produce natural rubber[17]. Among them, the rubber tree (*Hevea brasiliensis*) is the primary source of NR, supplying more than 98% of the global demand[18]. The laticifer located in the bark of the rubber tree is the site for NR biosynthesis and storage[19]. The rubber tree is a perennial crop of the Euphorbiaceae family and is native to the Amazon Basin in South America[20]. Unlike other crops (such as rice, wheat, and soybean) that have undergone domestication for hundreds or thousands of years, the domestication of the rubber tree is less than 150 years[21,22]. As a perennial tree with a long juvenile period, traditional rubber tree breeding is very time-consuming, usually taking 30-40 years. There is an urgent need to develop a breeding chip to accelerate the efficiency of rubber tree breeding.

Based on the high-quality rubber tree genomic data and the advantages of the GBTS approach, we developed the first rubber tree liquid SNP chip, termed 'HbGBTS80K'. Research on genetic diversity and target gene identification demonstrated

the high efficiency and accuracy of the HbGBTS80K chip, indicating its vast potential in rubber tree molecular breeding.

Method and material

Plant materials

In this study, a total of 256 wild germplasms and 148 cultivars planted in the field of the National Tropical Plants Germplasm Resource Center-Rubber Tree (Danzhou, Hainan Province) were utilized (Supplemental Table S1). Among them, 335 accessions (including 208 wild germplasms and 148 cultivars) were re-sequenced previously and utilized for the design of the liquid-phase chip, while the remaining 69 accessions (including 48 wild germplasms and 21 cultivars) were used for the quality control of the liquid-phase chip.

DNA isolation

The genomic DNA of 69 accessions was extracted using the DNAsecure Plant Kit (Tiangen Biotech, China). The integrity and quality of the total DNA were evaluated in 1.5% agarose gel and by NanoDrop 2000 (Thermo Fisher Scientific, USA), respectively.

Liquid-phase chip design and quality control

The VCF files containing variation information of 335 re-sequenced accessions were used for the design of the liquid-phase chip. Briefly, 96,044 SNPs were selected based on the criteria of minor allele frequency > 0.05, deletion rate < 0.2, and heterozygosity rate < 0.3. The probes were designed using GenoBaits Probe Designer (Molbreeding Biotech., Shijiazhuang, China) and used for hybridization capture sequencing of 69 test accessions according to the protocol of the GenoBaits DNA seq Library Prep kit (Molbreeding Biotech., Shijiazhuang, China). The captured sequencing results were sorted in the order of priority of gene region > gene promoter region > gene downstream region > gene intergenic region. Finally, 80,080 high-confidence SNPs were retained to obtain the HbGBST80K liquid-phase chip (Supplemental Table S2).

Bioinformatics analysis

A phylogenetic tree with 100 bootstrap replicates was constructed using RAxML (version 8.2.12), PCA was performed using SNPRelate (version 1.16.0), and the population structure of the rubber tree accessions was determined using fastStructure (version 1.0). The genotypic data of NLR reported previously was used herein. For GWAS analysis, the Q-matrix corrected linear regression model (GLM-Q) of the PLINK software was applied. The genome-wide significance threshold of the GWAS corresponding to the raw P values was set to $-\log10P > 4$.

Results

Design of the HbGBTS80K chip

We previously assembled a high-quality rubber tree cultivar 'CATAS8-79' genome and re-sequenced 335 accessions with an average depth of ~20×[19]. After aligning the re-sequencing data with the 'CATAS8-79' genome, a total of 5,323,701 SNPs were generated and used as a panel for the design of the liquid chip (Fig. 1). Based on the criteria of minor allele frequency > 0.05, deletion rate < 0.2, and heterozygosity < 0.3, 96,044 SNPs were selected for the design of capture probes. All probes were evaluated using an additional 69 accessions. Finally, 80,080 high-confidence SNP sites were retained (Supplemental Table S2), referred to as the HbGBST80K liquid-phase chip.

Characteristics of the HbGBTS80K chip

All 80,080 SNPs were distributed across the genome, with an average of 4,449 SNPs per chromosome (Fig. 2a). Among them, Chr6 had the maximum number of SNPs (7,128), while Chr1 had the minimum number of SNPs (2,642). Gene annotation revealed that 35.20% (28,191) of the SNPs were located in the intergenic region, 30.80% (24,667) were in the exonic region, 20.46% (16,386) were in the intronic region, 6.93% (5,550) were in the upstream region, and 6.60% (5,286) were in the downstream region (Fig. 2b).

HbGBTS80K chip application to population genetic diversity analysis

ADMIXTURE analysis showed that all accessions could be clearly classified into four groups by 80,080 SNPs (Fig. 3). Group I was mainly composed of cultivars, while Groups II, III, and IV were mainly composed of the wild germplasms from the Acre (WAC), Rondonia (WRO), and Mato Grosso (WMG) States of Brazil, respectively. In addition, the principal component analysis and the structure analysis obtained similar results, suggesting that the HbGBTS80K chip exhibited high accuracy in the genotyping of rubber tree germplasms.

HbGBTS80K chip application to genome-wide association study

The number of laticifer rings (NLR) is the major trait for NR yield. Using re-sequencing data, a major gene related to NLR was identified by GWAS on 208 individuals[19]. To evaluate the accuracy of the HbGBTS80K chip, the genotypic data of NLR reported previously was reanalyzed for GWAS analysis. The results showed that a strong association signal was detected in Chr15 (Fig. 4). Around the signal peak, the major gene controlling NLR, *HbPSK5* was also detected, suggesting high consistency with that from resequencing.

Discussion

Liquid-phase chip based on GBTS represents the next-generation biochip. Compared with the solid-phase chip, the liquid-phase chip has more advantages[13, 23]. (1) Higher accuracy: GBTS can perform multiple amplifications and sequencing of the target sites to ensure accuracy. (2) Lower price: Without the complicated chip hybridization work, the price is only one-fifth that of the solid-phase chip. (3) More convenient writing loci: Due to the use of PCR amplification technology, the targeted sites can be added or removed conveniently. In the past few years, the liquid-phase chip has been widely used in the biological breeding of more than 100 animals and plants[24,25].

Due to the lack of genomic data, the development of molecular markers for rubber trees has been slow for a long time, seriously restricting breeding innovation. With the development of next-generation sequencing technology, several rubber tree genomes have been published[19,26–30], and hundreds of germplasms from different sources have also been re-sequenced, providing sufficient data for the design of liquid-phase chips. Using the high-quality CATAS8-79 genome and the high-depth re-sequencing data of 335 accessions[19], we designed the first rubber tree liquid-phase chip, HbGBTS80K. The selection of loci avoids highly repetitive sequences and

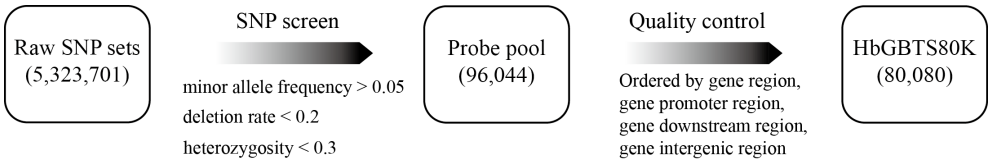


Fig. 1 Design pipeline of the rubber tree HbGBTS80K chip.

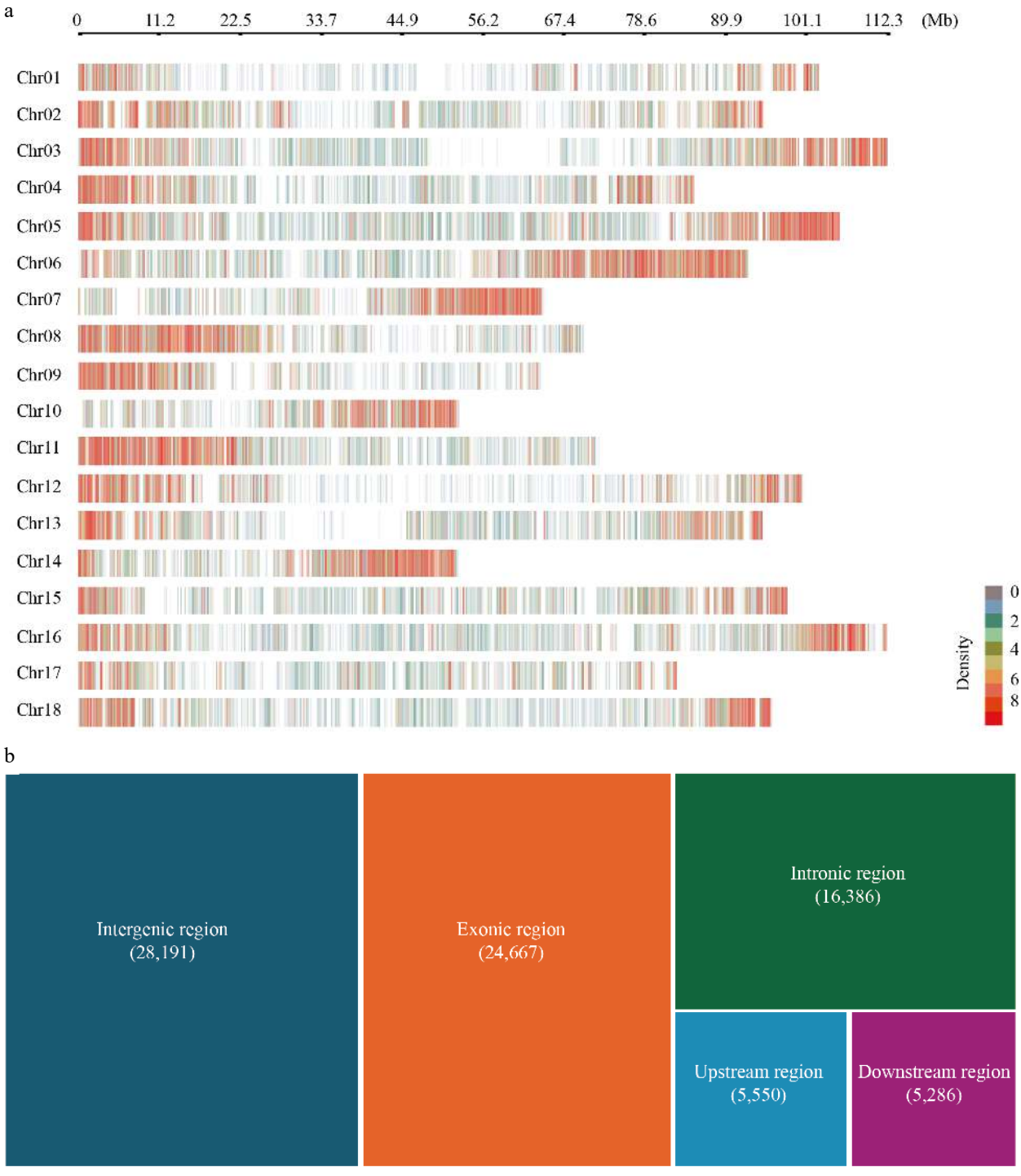


Fig. 2 Characteristics of the HbGBTS80K chip. (a) SNP marker density throughout the genome. (b) SNP locations within the genome.

leans more towards gene-rich regions. As evidence, 64.80% of the loci of HbGBTS80K are located in the gene body region (exon, intron, up-/down-stream), which will help to identify functional genes in subsequent use.

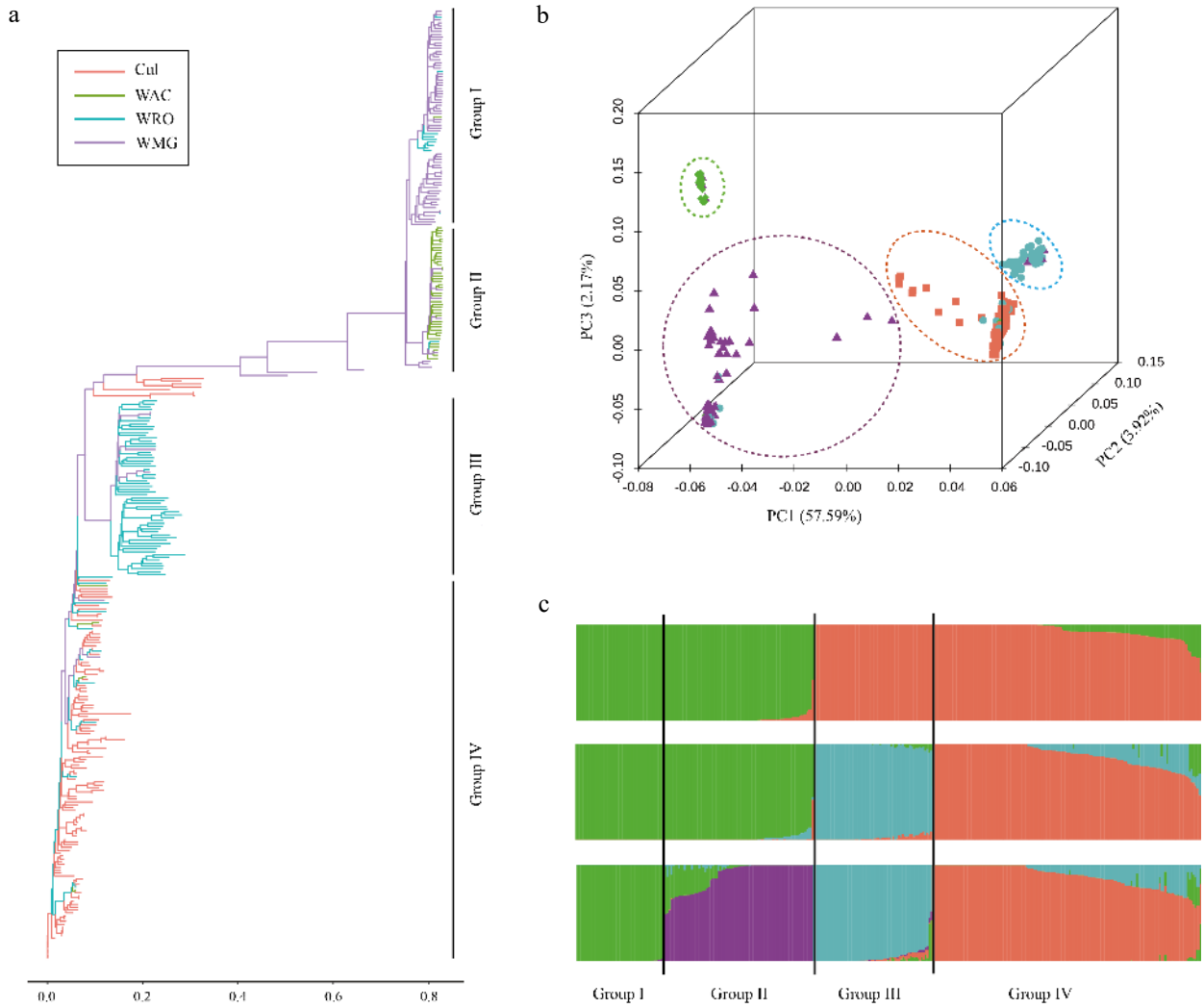


Fig. 3 Application in population genetic diversity analysis. (a) ADMIXTURE analysis. (b) Principal component analysis. (c) Structure analysis.

Population diversity analysis is a prerequisite step for evolutionary analysis and the identification of trait-related genes. By using RAPD markers and SSR markers, researchers have shown that wild germplasms have diverged to some extent[31]. Using 336,200 SNPs, Chao et al. demonstrated that wild germplasms can be classified into three groups, and the cultivars are derived from the wild germplasms in the Mato Grosso State[19]. In the present study, HbGBTS80K can also clearly divide wild germplasms into three groups, suggesting that the chip achieves the accuracy of high-depth resequencing in genetic diversity, while greatly reducing the number of SNPs.

Unlike other crops such as rice, soybean, and wheat, the mechanism of rubber tree yield formation is very complex. Tapping, which involves cutting laticifer rings to obtain latex, is the only way to harvest NR[19]. This process is not only influenced by genetic factors such as laticifer differentiation and NR biosynthesis, but also by environmental factors such as temperature and tapping method. Although attempts have been made to use GWAS to identify NR yield-related genes, the results have not been satisfactory. Recently, we have shown that NLR is the major trait of NR yield and identified a major gene HbPSK5[19]. In the present study, performing GWAS on the HbGBTS80K platform can also identify the HbPSK5 gene, reinforcing the fact that HbPSK5 is the major gene for NLR and that the HbGBTS80K chip has high sensitivity in target gene identification by GWAS. Additionally, because most SNPs of HbGBTS80K are located in the gene body, it is easier to obtain functional genes by GWAS compared to high-depth resequencing.

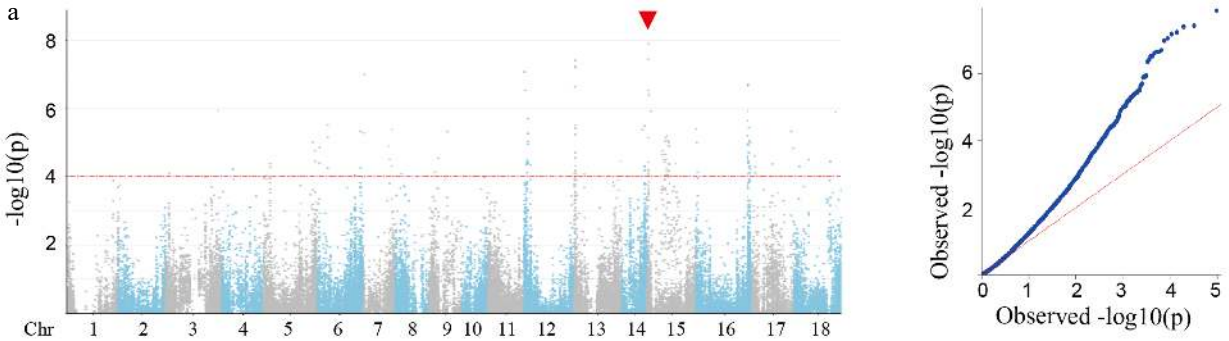
Conclusions

In this study, a liquid SNP chip named 'HbGBTS80K' was developed. This SNP chip comprised 80,080 SNPs evenly distributed across 18 chromosomes. The application of HbGBTS80K in population genetic diversity analysis effectively distinguished 404 rubber accessions into four groups. GWAS analysis of NLR could also detect the major gene HbPSK5 by using the HbGBTS80K platform. This newly developed HbGBTS80K liquid SNP chip is a valuable tool that will facilitate functional studies and molecular breeding of rubber tree.

Author contributions

The authors confirm contribution to the paper as follows: study conception and design: Chao J, Zhang J, Tian W; data

a



b

Marker	*p* value	Ref.	Mut.	Type	Gene	Annotation
Chr15_1115673	6.94E-05	A	T	Downstream	Hb15g001160	Hypothetical protein
Chr15_1189260	7.93E-05	T	C	Exonic	Hb15g001240	Arogenate dehydratase
Chr15_1200714	7.06E-06	G	A	Upstream	Hb15g001250	Phytosulfokines
Chr15_1452899	6.22E-05	G	T	Downstream	Hb15g001470	DNA repair protein
Chr15_1526979	1.29E-08	A	T	Exonic	Hb15g001510	UDP-D-apiose
Chr15_1542654	3.69E-08	T	C	Exonic	Hb15g001520	Diacylglycerol acyltransferase
Chr15_1566529	3.78E-05	A	G	Downstream	Hb15g001520	Diacylglycerol acyltransferase
Chr15_1115673	6.94E-05	A	T	Intronic	Hb15g001550	Magnesium transporter
Chr15_1786054	9.51E-05	T	C	Intronic	Hb15g001730	Vacuolar protein

Fig. 4 Application in GWAS. (a) Manhattan plots and Q-Q plots for NLR. (b) Significant markers around the peak signal.

analysis: Li Y, Chen X, He Y, Yue Z; sample collection: Li Y, Yang S; draft manuscript preparation: Chao J, Li Y. All authors reviewed the results and approved the final version of the manuscript.

Data availability

The genome assembly that support the findings of this study are available in the Zenodo repository: https://doi.org/10.5281/zenodo.7123623. The raw sequencing data used in this study are available in the National Genomics Data Center (NGDC, https://ngdc.cncb.ac.cn/) under project number PRJCA004986.

Acknowledgments

The research was supported by the Earmarked Fund for China Agriculture Research System (CARS-33-YZ1).

Conflict of interest

The authors declare that they have no conflict of interest. Weimin Tian is the Editorial Board member of *Tropical Plants* who was blinded from reviewing or making decisions on the manuscript. The article was subject to the journal's standard procedures, with peer-review handled independently of this Editorial Board member and the research groups.

Supplementary information accompanies this paper at (https://www.maxapress.com/article/doi/10.48130/tp-0024-0020)

Dates

Received 18 April 2024; Revised 27 April 2024; Accepted 29 April 2024; Published online 17 May 2024

References

1. Dai S, Long Y. 2015. Genotyping analysis using an RFLP assay. *Methods in Molecular Biology* 1245:91–99
2. Thilaga S, Rahul Nair R, Rajesh Kannan M, Ganesh D. 2017. RAPD markers for screening shoot gall maker (*Betousa stylophora* Swinhoe) tolerant genotypes of amla (*Phyllanthus emblica* L.). *Journal of Genetic Engineering and Biotechnology* 15:323–30
3. Ranamukhaarachchi DG, Kane ME, Guy CL, Li QB. 2000. Modified AFLP technique for rapid genetic characterization in plants. *BioTechniques* 29:858–59
4. Li X, Qiao L, Chen B, Zheng Y, Zhi C, et al. 2022. SSR markers development and their application in genetic diversity evaluation of garlic (*Allium sativum*) germplasm. *Plant Diversity* 44:481–91
5. Hiraoka Y, Ferrante SP, Wu GA, Federici CT, Roose ML. 2024. Development and assessment of SNP genotyping arrays for citrus and its close relatives. *Plants* 13:691
6. Zhang J, Yang J, Zhang L, Luo J, Zhao H, et al. 2020. A new SNP genotyping technology target SNP-seq and its application in genetic analysis of cucumber varieties. *Scientific Reports* 10:5623
7. Si Z, Jin S, Li J, Han Z, Li Y, et al. 2022. The design, validation, and utility of the "ZJU CottonSNP40K" liquid chip through genotyping by target sequencing. *Industrial Crops & Products* 188:115629
8. Li A, Acevedo-Rocha CG, Sun Z, Cox T, Xu JL, et al. 2018. Beating bias in the directed evolution of proteins: combining high-fidelity on-chip solid-phase gene synthesis with efficient gene assembly for combinatorial library construction. *Chembiochem* 19:221–228
9. Jin X, Fu R, Du W, Shan X, Mao Z, et al. 2022. Rapid, highly sensitive,

and label-free pathogen assay system using a solid-phase self-interference recombinase polymerase amplification chip and hyperspectral interferometry. *Analytical Chemistry* 94:2926−2933
10. Wöhrle J, Krämer SD, Meyer PA, Rath C, Hügle M, et al. 2020. Digital DNA microarray generation on glass substrates. *Scientific Reports* 10:5770
11. Bremer A, Mittag T, Heymann M. 2020. Microfluidic characterization of macromolecular liquid-liquid phase separation. *Lab on Chip* 20:4225−4234
12. Chen H, Han Z, Ma Q, Dong C, Ning X, et al. 2022. Identification of elite fiber quality loci in upland cotton based on the genotyping-by-target-sequencing technology. *Frontiers in Plant Science* 13:1027806
13. Xiang M, Liu S, Wang X, Zhang M, Yan W, et al. 2023. Development of breeder chip for gene detection and molecular-assisted selection by target sequencing in wheat. *Molecular Breeding* 43:13
14. Yuan D, Kong J, Li X, Fang X, Chen Q. 2018. Colorimetric LAMP microfluidic chip for detecting three allergens: peanut, sesame and soybean. *Scientific Reports* 8:8682
15. Kim KW, Nawade B, Nam J, Chu SH, Ha J, et al. 2022. Development of an inclusive 580K SNP array and its application for genomic selection and genome-wide association studies in rice. *Frontiers in Plant Science* 13:1036177
16. Morales KY, Singh N, Perez FA, Ignacio JC, Thapa R, et al. 2020. An improved 7K SNP array, the C7AIR, provides a wealth of validated SNP markers for rice breeding and genetics studies. *PLoS One* 15:e0232479
17. Rousset A, Amor A, Punvichai T, Perino S, Palu S, et al. 2021. Guayule (*Parthenium argentatum* A. Gray), a renewable resource for natural polyisoprene and resin: composition, processes and applications. *Molecules* 26:664
18. Priyadarshan PM, Goncalves PDS. 2003. *Hevea* gene pool for breeding. *Genetic Resources and Crop Evolution* 50:101−14
19. Chao J, Wu S, Shi M, Xu X, Gao Q, et al. 2023. Genomic insight into domestication of rubber tree. *Nature Communications* 14:4651
20. Onokpise OU. 2004. Natural rubber, *Hevea brasiliensis* (Willd. ex a. Juss.) Müll. Arg., germplasm collection in the Amazon basin, Brazil: a retrospective. *Economic Botany* 58:544−55
21. Schultes RE. 1977. The odyssey of the cultivated rubber tree. *Endeavour* 1:133−38
22. Supriya R, Priyadarshan PM. 2019. Genomic technologies for *Hevea* breeding. In *Advances in Genetics*, ed. Dhavendra K. San Diego, USA: Academic Press. vol. 104. pp. 1−73. https://doi.org/10.1016/bs.adgen.2019.04.001
23. Yang Q, Zhang J, Shi X, Chen L, Qin J, et al. 2023. Development of SNP marker panels for genotyping by target sequencing (GBTS) and its application in soybean. *Molecular Breeding* 43(4):26
24. Wang R, Xing S, Bourke PM, Qi X, Lin M, et al. 2023. Development of a 135K SNP genotyping array for *Actinidia arguta* and its applications for genetic mapping and QTL analysis in kiwifruit. *Plant Biotechnology Journal* 21:369−80
25. Sun C, Dong Z, Zhao L, Ren Y, Zhang N, et al. 2020. The Wheat 660K SNP array demonstrates great potential for marker-assisted selection in polyploid wheat. *Plant Biotechnology Journal* 18:1354−60
26. Liu J, Shi C, Shi CC, Li W, Zhang QJ, et al. 2020. The chromosome-based rubber tree genome provides new insights into spurge genome evolution and rubber biosynthesis. *Molecular Plant* 13:336−50
27. Pootakham W, Sonthirod C, Naktang C, Ruang-Areerate P, Yoocha T, et al. 2017. *De novo* hybrid assembly of the rubber tree genome reveals evidence of paleotetraploidy in *Hevea* species. *Scientific Reports* 7:41457
28. 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
29. Lau NS, Makita Y, Kawashima M, Taylor TD, Kondo S, et al. 2016. The rubber tree genome shows expansion of gene family associated with rubber biosynthesis. *Scientific Reports* 6:28594
30. Rahman AYA, Usharraj AO, Misra BB, Thottathil GP, Jayasekaran K, et al. 2013. Draft genome sequence of the rubber tree *Hevea brasiliensis*. *BMC Genomics* 14:75
31. Hussein MAA, Eid M, Rahimi M, Filimban FZ, Abd El-Moneim D. 2023. Comparative assessment of SSR and RAPD markers for genetic diversity in some Mango cultivars. *Peer J* 11:e15722