The genome for *Toxicodendron vernicifluum* brings crucial insights into Anacardiaceae evolution and urushiol biosynthesis

The MGI's DNBSEQ-G400 sequencing platform helps decipher the chromosome-level genome for *Toxicodendron vernicifluum* and identify genes involved in urushiol biosynthesis pathway.

The research team led by Li Sifeng at Xi'an Botanical Garden of Shaanxi Province completed the reference genome of *Toxicodendron vernicifluum* and identified genes related to lignin and urushiol biosynthetic pathways based on the MGI's DNBSEQ-G400 sequencing platform. This provides valuable resources for the study of the genome of *T.vernicifluum*, especially the urushiol biosynthetic pathway and the comprehensive utilization of *T. vernicifluum*.

The research was published on *iScience* entitled with "The chromosome-level genome for *Toxicodendron vernicifluum* provides crucial insights into Anacardiaceae evolution and urushiol biosynthesis"[1].

Recommended application: Agricultural Genomics

Recommended model: DNBSEQ-G400RS

- **Data output is efficient and high-quality**

  DNBSEQ sequencing technology has significant features such as high accuracy, low duplication rate, and low index hopping rate, etc.

- **Provide a complete product combination for the experimental process**

  Based on independently developed automatic solutions for experimental processes and data analysis, MGI provides a complete set of product combination from sample input to report output.

Background

Toxicodendron vernicifluum (Stokes) F. A. Barkley, also called the Chinese lacquer tree, is a member of Anacardiaceae family. The lacquer tree, similar to rubber trees, can secrete a sap known as "raw lacquer" after the phloem is cut. Raw lacquer is an excellent adhesive and paint material with numerous properties like anti-corrosion, anti-rust, non-oxidation, and resistance to acid, alcohol and high temperature. The lacquer tree is indigenous to China and the Indian subcontinent, and it has been grown in East Asian nations such as South Korea and Japan for thousands of years[1]. Currently, China accounts for 85% of the global raw lacquer production and has abundant lacquer tree germplasm resources, such as "Dahongpao", "Gaobachi", "Huangmao Guizhou", and "Hongmao Guizhou"[1].

The raw lacquer is a natural polymer composed of numerous substances, containing urushiol, laccol, thitsiol, laccase, polysaccharides, and glycoprotein. The basic ingredient in raw lacquer for quick drying and film formation is urushiol, which also provides the structural basis for the development of various novel coating materials[2]. Although lacquer can induce contact dermatitis, scientists are interested in its pharmacological characteristics including anticancer, antibacterial, and antioxidant effects[2]. Urushiol is a mixture of phenols (catechol and resorcinol) with a side chain length of 15–17C and numerous terminal vinyl structures. Previous studies have shown that fatty acid metabolism intermediates, such as hexadecanoyl-CoA and alkyl tetraketone, play an important role in the formation of urushiol[3]. Based on transcriptome and enzyme functional annotation, Type III polyketide synthase (PKS) is thought to be the first enzyme that catalyzes the production of urushiol. Other enzymes involved in the process include OXSM/FabF, FabZ, FATB, and ACSL, but the biosynthetic pathway of urushiol remains largely unknown[1].

China has abundant lacquer tree germplasm resources, but the breeding of lacquer trees is still stuck in the traditional breeding stage due to the lack of genomic information. The biosynthesis pathway of urushiol in lacquer trees and the key enzyme genes in this pathway have not been deciphered, which seriously restricts the research and utilization of comprehensive lacquer tree resources[4]. With the quickly development of high-throughput sequencing technology, it provides a powerful tool to investigate the genetic information of this species[4].

Study description

The research team completed the first reported chromosome-level high-quality genome sequencing of the lacquer tree and deciphered the candidate genes related to the production of urushiol and lignin in the lacquer tree using genomic, transcriptomic, and metabolomic technologies. This study will provide important support for studies such as exploring the biosynthetic pathway of urushiol and conducting molecular breeding on lacquer trees[1].

Materials and Methods

a.Sample collection

The lacquer trees used in this study were cultivated in Xi'an Botanical Garden, Xi'an City, Shaanxi Province. Fresh leaves of a diploid female "Gaobachi" (*T. vernicifluum* "Gaobachi") were used for genome sequencing. Roots, stems, leaves, flowers and drupes from *T. vernicifluum* "Gaobachi" were used for full-length transcriptome sequencing. Five varieties including "Gaobachi" (GBC), "Dahongpao" (DHP), "Huangmao Guizhou" (HuM), "Hongmao Guizhou" (HoM) and wild lacquer were used for transcriptome sequencing, metabolome detection, and qRT-PCR validation, each experiment was performed with three biological repeats. The phloem of each variety was collected for metabolomic analysis.

b. Genome sequencing and assembly

From the above experimental materials, high-quality genomic DNAs were extracted. Libraries with short insert sizes of 350 bp and long insert sizes of approximately 20 kb were then constructed according to the standard procedures and sequenced on the DNBSEQ-G400 platform and Oxford Nanopore platform, respectively. Data output from the DNBSEQ-G400 platform was filtered using SOAPnuke v.1.5.6, and data from the Oxford Nanopore platform was trimmed and filtered using NanoFilt. Genome assembly was conducted using GenomeScope v1.0, Canu v1.7, SMARTdenovo, and Pilon v.1.225.

After fixing the extracted genomic DNAs with formaldehyde, the DNAs were sheared by the restriction enzyme (MboI) to construct a Hi-C library. Then, a paired-end 150 bp (PE150) sequencing was performed on the DNBSEQ-G400 platform. A total of 488,159,058 paired-end reads with 73.22 Gb sequencing data were generated. The sequencing data was filtered using SOAPnuke v.1.5.6, and contigs were anchored to the chromosome-level assembly using Juicer v.1.5.

c. Transcriptome sequencing and gene expression analysis

Following RNA extraction, samples of five varieties—GBC, DHP, HuM, HoM, and wild variety were taken for RNA-seq analysis with three biological repeats for each variety. RNA libraries with insert sizes of 300–400 bp were constructed in accordance with the instructions for library construction, and then paired-end sequenced on the DNBSEQ-G400 platform. SPAPnuke was used to filter the reads. HISAT v2.1.1 and StringTie v1.3.3b were employed to map and assemble the genome. RSEM v1.2.12 was used to calculate the gene expression level and DEseq2 was employed to detect differentially expressed genes (DEGs).

A full-length transcriptome sequencing was performed on the extracted RNA using the PacBio Sequel II platform, and full-length non-chimeric (FLNC) transcripts were obtained using the PacBio ISO-Seq3 pipeline.

d. Evolution analysis and urushiol biosynthesis pathway analysis

Combining the above sequencing data and existing resources, phylogenetic analysis, genome assembly and whole-genome duplication analysis, identification of polyketide synthase genes, and metabolomic analysis of lacquer tree varieties were conducted to study the evolution of the lacquer tree family and the biosynthesis of urushiol.

Sample collection	Library preparation and sequencing	Bioinformatics analysis	Result analysis
Genome sequencing: *T. vernicifluum* 'Gaobachi' Transcriptome sequencing: 'Gaobachi' (GBC), 'Dahongpao' (DHP), 'Huangmaoguizhou' (HuM), 'Hongmaoguizhou' (HoM) and wild variety	 MGIEasy Universal DNA Library Prep Kit MGIEasy RNA Library Prep Kit  DNBSEQ-G400 Genetic Sequencer	SOAPnuke v.1.5.6 Juicer v.1.5 HISAT v2.1.1 StringTie v1.3.3b RSEM v1.2.12 DEseq2	Phylogenetic analysis, Genome assembly, Whole-genome duplication analysis, Identification of polyketide synthase gene and metabolomic analysis of lacquer tree varieties

Results

Chromosome-level assembly and annotation of *T. vernicifluum* genome

DNBSEQ-G400, Nanopore, PacBio Sequel II, and Hi-C technologies were utilized in this study to sequence the whole genome of the *T. vernicifluum* "Gaobachi". After quality control and trimming, the clean reads on the MGI's DNBSEQ platform were 124.4 Gb, and the ONT high-quality reads were 105.28 Gb. After assembling, the assembled genome size was 491.93 Mb, the contig and scaffold N50 were 5.26 Mb and 32.97 Mb respectively, the chromosome mounting rate was 98.26%, a total of 32,682 genes were annotated, and a high-integrity and high-quality genome assembly was obtained (Figure 1).

Lacquer tree genomic analysis

After conducting a phylogenetic analysis of the 1,914 single-copy genes shared by lacquer tree and 14 other plant taxa, the results showed that all Anacardiaceae, Rutaceae, and Sapindaceae species were clustered in the same clade, indicating a close relationship among these three families.

Evolution analysis showed that the divergence between *T. vernicifluum* and *P. vera* occurred approximately 25.2 million years ago (Mya), and the *T. vernicifluum* experienced an ancient whole-genome duplication (WGD) event 47.3 Mya (Figure 2). The sharp peaks in 4-fold synonymous third codon transversion (4DTV) values (Ks = 0.5) also indicated that the *T. vernicifluum* genome shared a WGD event with *M. indica* , and there was a relative high collinearity between the two.

This study further explored the significant expansion and contraction of gene families during evolution by performing OrthoFinder analysis of *T. vernicifluum* and other 14 species. KEGG enrichment analysis of expanded gene families were enriched in functions related to phenylpropanoid biosynthesis, biosynthesis of unsaturated fatty acids, metabolism of terpenoids and polyketides, and plant-pathogen interaction; while contracted gene families showed significantly enrichment in genes related to functions, such as phosphotransferase activity, transporter activity, and catalytic activity. The gene families related to urushiol biosynthesis were expanded significantly, which may play an important role in the ecological fitness and biological adaptability of lacquer trees.

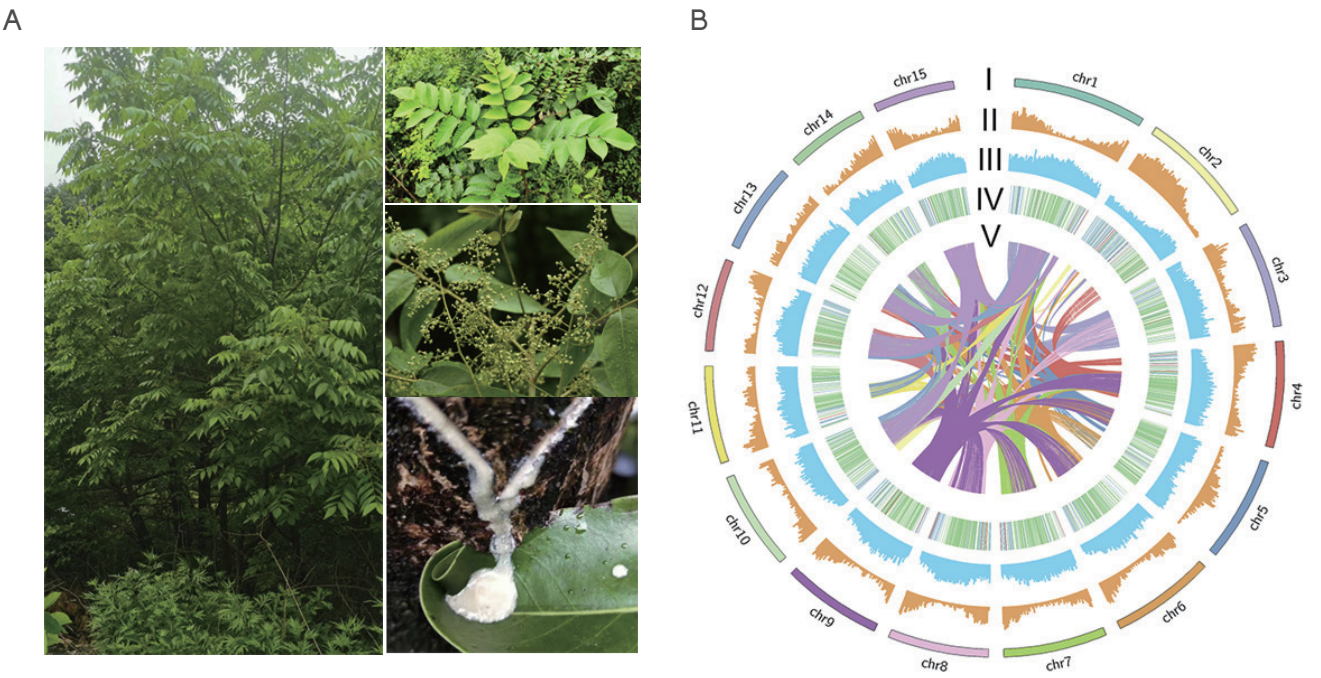

Figure 1. Tree, leaf, flower, and raw lacquer blooding of *T.vernicfluum* and high-quality lacquer genome assembly

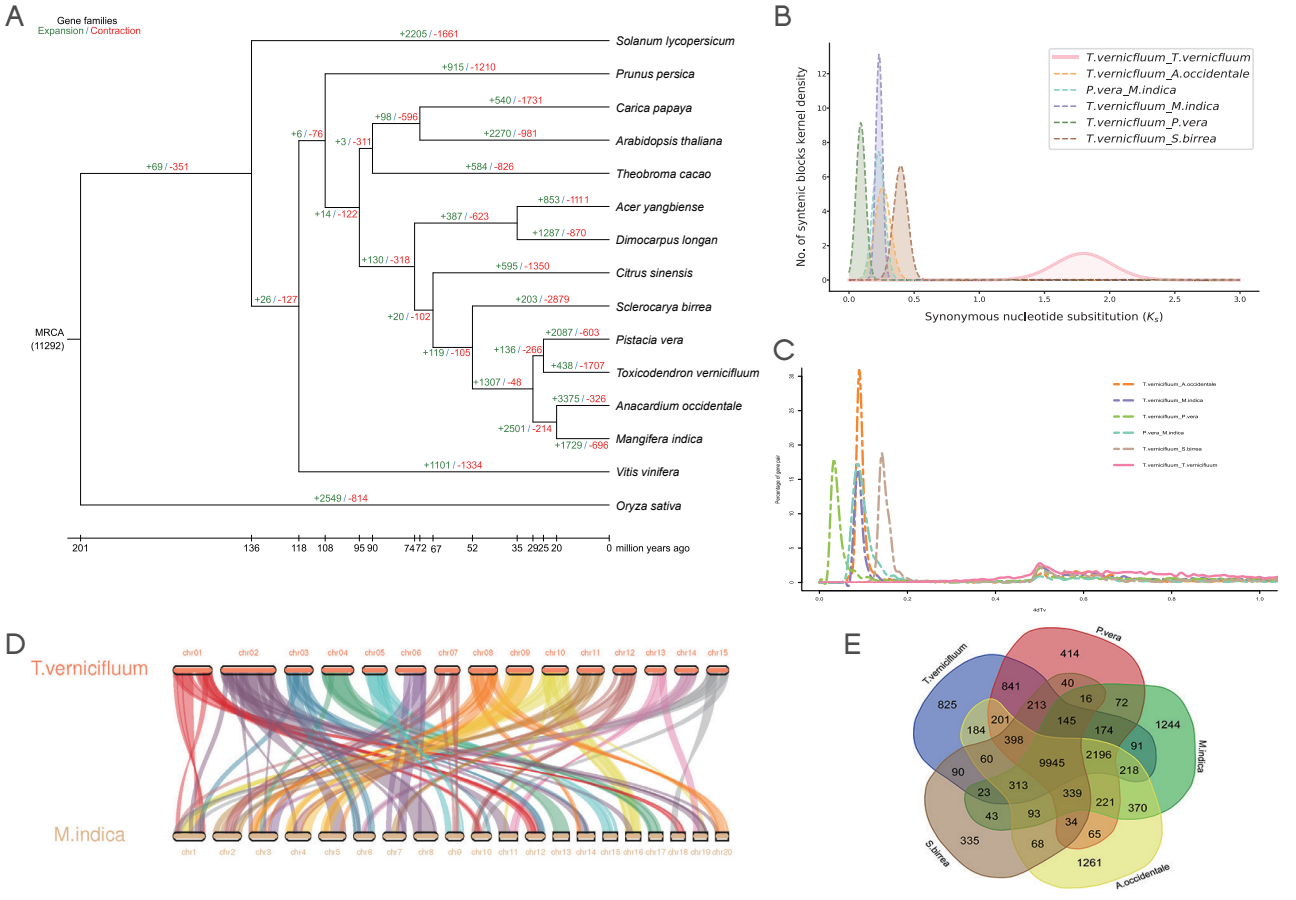

Figure 2. Comparative genomic analysis of the lacquer tree genome

Urushiol and lignin biosynthetic pathway analysis

The researchers identified key enzyme genes involved in the synthesis of urushiol and lignin, and drew a schematic diagram of the biosynthetic pathways of urushiol and lignin combining the genomic, transcriptomic, and metabolomic data of *T. vernicifluum* (Figure 3). 33 PKS genes were identified in *T. vernicifluum*, which were key enzyme genes in the biosynthetic pathway of urushiol. Different varieties have differences in their unique natural products, and flavonoid compounds such as butin and fisetin can be used as metabolic markers for the identification of *T. vernicifluum* varieties.

This study provided the genetic information for lacquer trees, identified candidate genes related to the biosynthesis pathways of urushiol and lignin, provided key enzyme genes for further genetic and enzymatic validation, and laid the groundwork for lacquer tree molecular breeding and cultivation research.

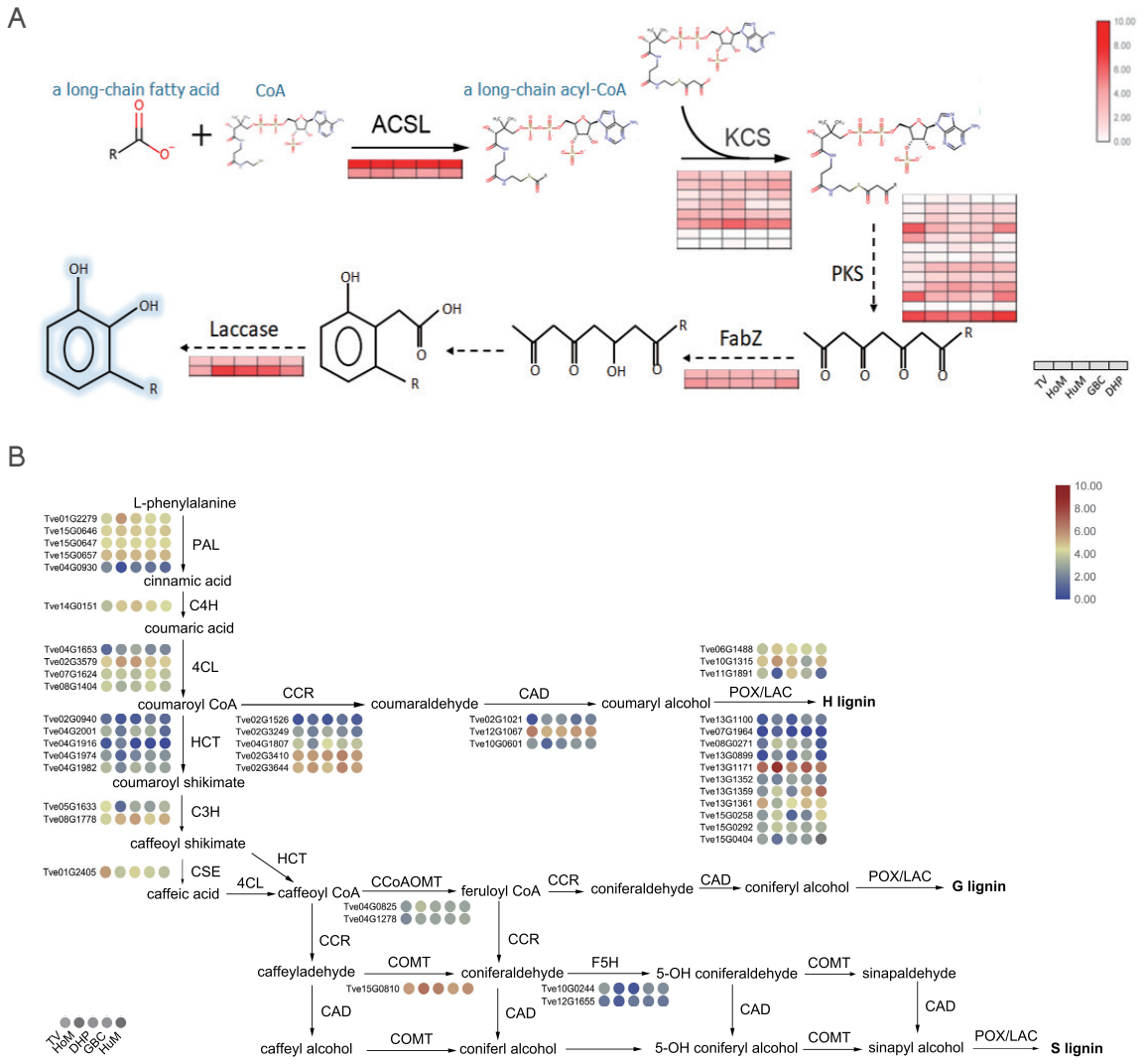

Figure 3. Diagram of the biosynthesis pathway of urushiol and lignin

Summary

MGI's sequencing platform DNBSEQ-G400, based on DNBSEQ™ technology, facilitated the assembly of lacquer tree genome, including generating short reads to correct the error for long reads, producing Hi-C data for chromosome construction of the genome and performing RNA-seq. MGI's sequencing platform enables the first high-quality lacquer tree genome decryption by the Institute of Botany of Shaanxi Province. This study offers a genomic basis for the exploration of lacquer tree molecular breeding and natural products and provides a support for lacquer tree comparative genomics and evolutionary biology studies.

DNBSEQ-G400 Genetic Sequencer

References

1. Bai, G. *et al*. The chromosome-level genome for Toxicodendron vernicifluum provides crucial insights into Anacardiaceae evolution and urushiol biosynthesis. *iScience* **25**, 104512, doi:10.1016/j.isci.2022.104512 (2022).
2. Cho, J. Y., Park, K. Y., Kim, S. J., Oh, S. & Moon, J. H. Antimicrobial activity of the synthesized non-allergenic urushiol derivatives. *Biosci Biotechnol Biochem* **79**, 1915-1918, doi:10.1080/09168451.2015.1061418 (2015).
3. Bai, G. Q. *et al*. [De novo Assembly and Analysis of Sumac (Toxicodendron vernicifluum (Stokes) F.A. Barkley) Transcriptomes Provides Insights into the Biosynthesis of Urushiol]. *Mol Biol (Mosk)* **52**, 764-772, doi:10.1134/S0026898418050051 (2018).
4. Kersey, P. J. Plant genome sequences: past, present, future. *Curr Opin Plant Biol* **48**, 1-8, doi:10.1016/j.pbi.2018.11.001 (2019).

Recommended Ordering Information

Category	Product	Cat. NO.
Instruments	DNBSEQ-G400RS Genetic Sequencer	900-000170-00
	MGISP-100RS Automated Sample Preparation System	900-000206-00
	MGISP-960RS Automated Sample Preparation System	900-000146-00
Software	MegaBOLT Bioinformatics analysis accelerator	900-000555-00
Library Prep	MGIEasy Universal DNA Library Prep Set（16 RXN）	1000006985
	MGIEasy RNA Library Prep Set（16 RXN）	1000006383
Sequencing Reagents	DNBSEQ-G400RS High-throughput Sequencing Set (FCL PE150)	1000016952

MGI Tech Co.,Ltd

Building 11, Beishan Industrial Zone, Yantian District, Shenzhen, CHINA, 518083

Version: August 2023

+86-4000-688-114

en.mgi-tech.com

MGI-service@mgi-tech.com

Author：Jie Chen, Hao Zhang

Editor-in-Charge：Qiwei Wang

Reviewer：Yao Jiang

1. For StandardMPS and CoolMPS: Unless otherwise informed, StandardMPS and CoolMPS sequencing reagents, and sequencers for use with such reagents are not available in Germany, Spain, UK, Sweden, Belgium, Italy, Finland, Czech Republic, Switzerland, Portugal, Austria and Romania. Unless otherwise informed, StandardMPS sequencing reagents, and sequencers for use with such reagents are not available in Hong Kong. No purchase orders for StandardMPS products will be accepted in the USA until after January 1, 2023.
2. For HotMPS sequencers: This sequencer is only available in selected countries, and its software has been specially configured to be used in conjunction with MGI's HotMPS sequencing reagents exclusively.
3. For HotMPS reagents: This sequencing reagent is only available in selected countries.