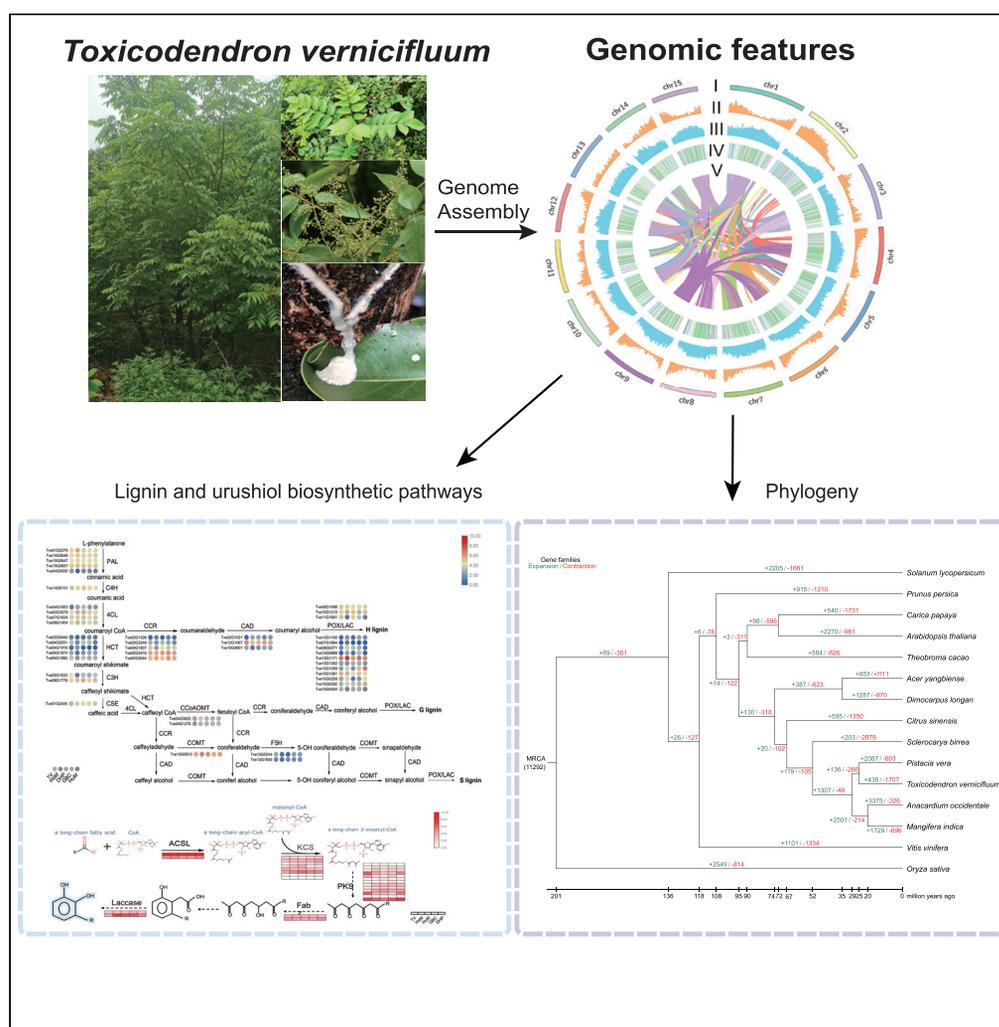


Article

The chromosome-level genome for *Toxicodendron vernicifluum* provides crucial insights into Anacardiaceae evolution and urushiol biosynthesis



Guoqing Bai,  
Chen Chen,  
Chenxi Zhao, ...,  
Xiaofeng Cong,  
Yun Jia, Sifeng Li

lisf60@sina.com

Highlights

We provide the first chromosome-level genome for *Toxicodendron vernicifluum*

We explore the phylogenetic position of lacquer tree

We identify genes involved in the urushiol and lignin biosynthetic pathways

Analysis of the unique active metabolites in cultivar lacquers by LC-MS/MS is reported

Bai et al., iScience 25, 104512  
July 15, 2022 © 2022 The Author(s).  
<https://doi.org/10.1016/j.isci.2022.104512>



## Article

The chromosome-level genome for *Toxicodendron vernicifluum* provides crucial insights into Anacardiaceae evolution and urushiol biosynthesisGuoqing Bai,<sup>1,7</sup> Chen Chen,<sup>1,7</sup> Chenxi Zhao,<sup>2,6,7</sup> Tao Zhou,<sup>3,7</sup> Dan Li,<sup>4,7</sup> Tianhua Zhou,<sup>5</sup> Weimin Li,<sup>1</sup> Yuan Lu,<sup>1</sup> Xiaofeng Cong,<sup>1</sup> Yun Jia,<sup>1</sup> and Sifeng Li<sup>1,8,\*</sup>

## SUMMARY

The lacquer tree (*Toxicodendron vernicifluum* (Stokes) F.A. Barkley) is an important tree with economic, industrial, and medicinal values. Here, we generated the reference genome of *T. vernicifluum* at the chromosome level with 491.93 Mb in size, in which 98.26% of the assembled contigs were anchored onto 15 pseudo-chromosomes with the scaffold N50 of 32.97 Mb. Comparative genomic analysis revealed the gene families related to urushiol biosynthesis were expanded, contributing to the ecological fitness and biological adaptability of the lacquer tree. We combined multi-omics data to identify genes that encode key enzymes in the *T. vernicifluum* urushiol and lignin biosynthetic pathways. Furthermore, the unique active metabolites, such as butin and fisetin, in cultivar lacquers were identified by metabolism profiling. Our work would provide crucial insights into metabolite synthesis such as urushiol and lignin, meanwhile offer a basis for further exploration of the cultivation and breeding of *T. vernicifluum* and other Anacardiaceae members.

## INTRODUCTION

*Toxicodendron vernicifluum* (Stokes) F. A. Barkley belongs to Anacardiaceae family, also known as the Chinese lacquer tree (Wang et al., 2020a). Like rubber trees, the lacquer tree can produce a sap called “raw lacquer” by cutting its phloem, which is an excellent adhesive and paint material with multiple properties such as anticorrosion, antirust, non-oxidation, and resistance to acid, alcohol, and high temperature. The lacquer tree is native to China and the Indian subcontinent, and it has been farmed for thousands of years in East Asian nations, such as Korea and Japan (Fu et al., 2007; Nie et al., 2009). Historically, raw lacquer was traditionally produced of lacquerwares that were used as tableware and sacrificial utensils in Asian cultures (Li et al., 2016; Ma et al., 2015). Currently, Chinese raw lacquer manufacturing accounts for 85% of the global total output value, and plentiful lacquer tree germplasm resources were available in China, such as “Dahongpao”, “Gaobachi”, “Huangmao Guizhou”, and “Hongmao Guizhou” etc.

The most common product obtained from the lacquer tree is raw lacquer, which has a wide range of applications in aesthetic decoration, petroleum, nautical, and other industrial fields (Lu et al., 2014; Xu et al., 2012; Yang et al., 2015). Raw lacquer has been shown to be a natural polymer made up of various components, including urushiol, laccol, thitsiol, laccase, polysaccharides, and glycoprotein (Honda et al., 2008; Yang et al., 2018). Urushiol is the fundamental component for quick drying and film formation of raw lacquer, and it has a long history as a natural performance. Many novel coating materials based on the structure of urushiol have been created recently (Watanabe et al., 2016; Xue et al., 2020). Urushiol, in particular, has been shown to represent a collection of closely related chemicals that can cause contact dermatitis. (Bonnekoh et al., 2018; Llanchezian et al., 2012; Rozas-Muñoz et al., 2012). Nonetheless, Urushiol has been studied for their pharmacological properties, which include anticancer, antibacterial, and antioxidant properties. (Cho et al., 2015; Kim et al., 1997; Xie et al., 2016; Zhao et al., 2009).

Urushiol is a mixture of phenols (catechol and resorcinol) with a lengthy side chain of 15–17C and contains multiple terminal vinyl structures (Kim et al., 2019a, 2019b). It has been shown that fatty acid metabolism intermediates, such as hexadecanoyl-CoA, alkyl tetraketone are significant in the formation of urushiol (Bai et al., 2018; Weisberg, 2014). The phenylpropanoid metabolic pathway might also produce catechol

<sup>1</sup>Xi'an Botanical Garden of Shaanxi Province, Shaanxi Province Qinling-Bashan Mountains Engineering Research Centre of Conservation and Utilization of Biological Resources, Xi'an 710061, China

<sup>2</sup>BGI Genomics, BGI-Shenzhen, Shenzhen 518083, China

<sup>3</sup>School of Pharmacy, Xi'an Jiaotong University, Xi'an 710061, China

<sup>4</sup>SDIC Biotech Investment Co., Ltd., Shanghai 200082, China

<sup>5</sup>College of Biological Science and Engineering, Shaanxi University of Technology, Hanzhong 723001, China

<sup>6</sup>College of Life Sciences, University of Chinese Academy of Sciences, Beijing 100049, China

<sup>7</sup>These authors contributed equally

<sup>8</sup>Lead contact

\*Correspondence: lisf60@sina.com

<https://doi.org/10.1016/j.isci.2022.104512>



and resorcinol. According to transcriptome and enzyme functional investigations, type III polyketide synthase (PKS) was thought to be the first enzyme to catalyze urushiol production, with *OXSM/FabF*, *FabZ*, *FATB*, and *ACSL* all taking part. However, the biosynthesis routes of urushiol remain largely uncharacterized.

Thanks to the emergence and advancement of high-throughput sequencing technology, almost 600 complete plant genome assemblies are now accessible in public sources (Kersey, 2019). The high-quality genomes of commercial plants in Anacardiaceae, such as the “king of fruits” Mango, and the “important nut crops” Pistachio, have been sequenced recently (Wang et al., 2020b; Zeng et al., 2019). These genomes provide a foundation for studying genomic-based trait development and specialized Anacardiaceae biochemistry underneath. Furthermore, whole-genome duplication (WGD) or polyploidization is major drivers of specialization and the creation of new features and functions. These duplicates make them available for the evolution of a range of activities, such as producing insect resistance, boosting tolerance to stress, and encouraging plants to acquire adaptive benefits, all of which contribute to improved plant fitness (Jiang et al., 2021; Kang et al., 2020; Ma et al., 2021; Tang et al., 2016). Terpenoid biosynthesis genes, for example, are clustered and co-expressed, lowering the risk of incomplete transcription in Lamiaceae due to the loss of a single gene participating in the process (Consortium, 2018; Li et al., 2021; Xu et al., 2016; Zhao et al., 2019). Furthermore, the mango genome’s increased cluster of chalcone synthase genes (CHS) has co-syteny expression, which is linked to specific phenolic compounds (Wang et al., 2020b). Those specific biochemistries are extremely important for plant evolution and fitness.

Despite the considerable importance of the lacquer tree, genetic information on the species is scarce, which has hindered its study and utilization. Here, we reported the chromosome level of the lacquer genome, which is the first sequenced species in the genus *Toxicodendron*. Using genomic, transcriptomic, and metabolomic techniques, we thoroughly screen and discover the candidate genes responsible for urushiol and lignin production in the lacquer tree. The availability of our genomic data will be helpful for exploration of urushiol biosynthesis and molecular breeding and engineering of lacquer trees.

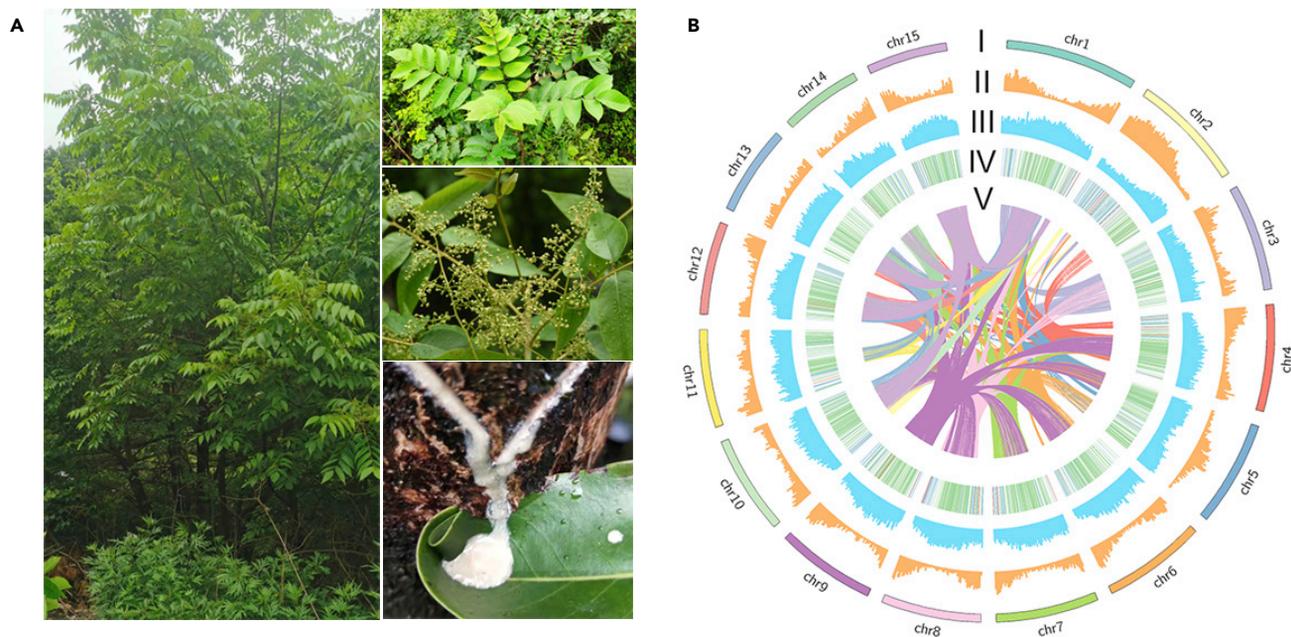
## RESULTS

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

A total of 130.8 Gb of short reads and 106.28 Gb of ONT long reads were obtained based on DNBSEQ and the Oxford Nanopore platform, respectively. After quality control and trimming, there were 124.4 Gb of DNBSEQ clean reads and 105.28 Gb of ONT high-quality reads. Based on 17-mer frequency distribution analysis with short reads, the *T. vernicifluum* (Figure 1A) genome was estimated in size of 560.06 Mb (Figure S1, Tables S1) with relatively high heterozygosity (0.56%) and repetitive sequence content (69.02%) (Table S1). The predicted genome size based on 17-mer analysis was close to the genome size (~540 Mb) determined experimentally by flow cytometry (Figure S2).

After assembling, we obtained a 491.81 Mb assembly with a contig N50 length of 5.26 Mb by using SMART *denovo* and polished it by *racon*, *medaka*, and *pilon* (Figure 1D and Table 1). The Benchmarking Universal Single-Copy Orthologs (BUSCO) evaluation score was used to assess the completeness and quality of the genome assembly, which resulted in 98.9% of the single-copy orthologs’ completeness (Table S3), indicating a very complete and high-quality genome assembly. To assist the chromosome-level assembly, 130.74 × Hi-C data were generated on the DNBSEQ platform. Finally, 491.93 Mb super-scaffolds with N50 of 32.97 Mb was generated based on ~73.22 Gb of high-throughput chromatin conformation capture (Hi-C) data using the hierarchical clustering strategy. In total, 98.26% of the assembled sequence was anchored and oriented onto 15 pseudochromosomes with lengths ranging from 25.95 to 42.48 Mb in total (Table S4). The genome of *T. vernicifluum* had a GC content of 35.7% (Figure S4 and Table S5).

The full-length transcriptome of a mixed library of leaves and phloem was generated and sequenced using the PacBio platform to aid with genome annotation (Table S6). We identified 303.34 Mb (61.66% of the genome length) of repetitive sequences in the genome assemblies of *T. vernicifluum* based on a combination of homology-based and *de novo* approaches (Table S7). Among them, 266.81 Mb of long terminal repeat (LTR) retrotransposons accounted for the largest proportion and made up 54.24% of the genome (Table S8, Figure S5). Totally, we predicted 32,682 protein-coding genes in the genome of *T. vernicifluum* with an average length of 3993.57 bp and an average exon number of 10.17 for each gene (Table 1) based on the homology annotation, *de novo* prediction, transcript mapping, and further



**Figure 1. Genomic landscape of lacquer tree**

(A) Tree, leaf, flower, and raw lacquer bleeding of *Toxicodendron vernicifluum*.

(B) Overview of lacquer genome assembly. (I) Chromosomes of *T. vernicifluum*; (II) gene density of the genome (500-kb sliding windows); (III) TE density of the genome (500-kb sliding windows); (IV) non-coding RNA content of the genome (red lines represent miRNA, blue lines represent tRNA, green lines represent snRNA, purple lines represent rRNA); (V) paralogous genes on different chromosomes.

filtering. 96.16% (31,428) of the protein-coding genes in the *T. vernicifluum* genome were assigned to multiple functions based on the analysis via various protein database, including NR, SwissProt, Kyoto Encyclopedia of Genes and Genomes (KEGG), KOG, TrEMBL, and Inter-Pro. The gene distribution and GC content along each chromosome were calculated and showed in Figure 1B. In addition, the *T. vernicifluum* genome contained 104 miRNAs, 510 tRNAs, 498 rRNAs, and 4,096 small nuclear RNAs (snRNA) (Figure 1B and Tables 1 and S9).

### Phylogenomics and divergence time estimation

To better comprehend the phylogenetic placements of Anacardiaceae among angiosperms, 1,914 single-copy genes shared by *T. vernicifluum* and other 14 plant taxa were recovered for the phylogenomic interference, with *Oryza sativa* was set as an outgroup. The phylogeny confirmed the monophyly of malvids. All Anacardiaceae, Rutaceae, and Sapindaceae species were discovered to be clustered in the same clade, showing close relationships of these three families. In particular, all Anacardiaceae species (*Anacardium occidentale*, *T. vernicifluum*, *Mangifera indica*, *Sclerocarya birrea*, and *Pistacia vera*) were clustered into one monophyletic group (Figure 2A) with a high bootstrap support value, and *T. vernicifluum* was closely related to *P. vera*. The phylogenetic tree indicated that the malvid clade diverged ~95.5 million years ago (Mya) (88.9–101.4 Mya). The splits in the crown group Sapindales were estimated to have occurred 74.3 million years ago (Mya) (67.7–81.7 Mya), while the Anacardiaceae divergence time was estimated to be 52.3 Mya (42.5–61.9 Mya). The divergence between the Trib. *Rhoideae* clade and the Trib. *Anacardieae* clade occurred 29.6 Mya (24.7–35.4 Mya) for four species, and it was also demonstrated that *T. vernicifluum* and *P. vera* belonging to Trib. *Rhoideae* diverged 25.2 Mya (13.5–27.8 Mya).

### WGD and collinearity analyses

Here, the density distribution of synonymous substitution rates per gene (Ks) between collinear paralogous genes was utilized to detect the WGD events. After calculating the Ks values of duplicate gene pairs, we found the values of Ks values for the collinear gene pairs peaked at 1.8, corresponding to an ancient WGD event that occurred around 47.3 Mya. We found that the peak value of orthologs between *T. vernicifluum* and *P. vera* (Ks = 0.1) was lower than the value of Ks = 0.4 between *T. vernicifluum* and *S. birrea*, indicating that the speciation between *T. vernicifluum* and *P. vera* occurred later, and this result

**Table 1. Summary of genome assembly and annotation of *T. vernicifluum* genome**

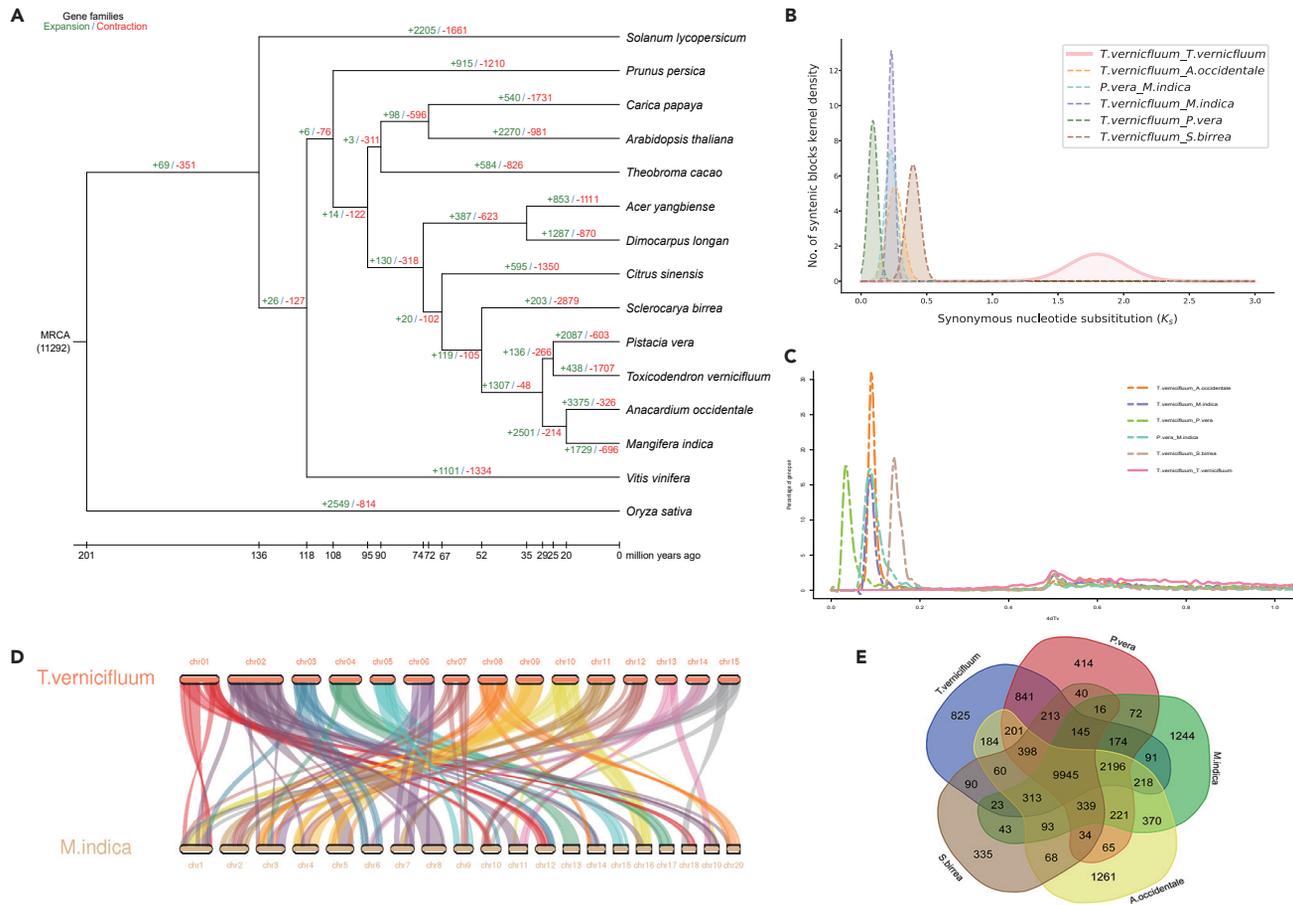
| Parameter                             | <i>T. vernicifluum</i> |
|---------------------------------------|------------------------|
| Estimate of genome size (Survey)      | 560.06 M               |
| Estimate of genome size (Hi-C)        | 491.93 M               |
| Total length of contigs               | 491.81M                |
| Total number of contigs               | 866                    |
| N50 of contigs (bp)                   | 5.26 M                 |
| Largest contig (bp)                   | 19.36 M                |
| Total length of super-scaffolds       | 491.93 M               |
| N50 of super-scaffolds                | 32.97 M                |
| GC content                            | 35.77%                 |
| Complete BUSCOs %                     | 98.9                   |
| <b>Annotation features</b>            |                        |
| No. of genes                          | 32,682                 |
| No. of predicted protein-coding genes | 31,428                 |
| Average gene length (bp)              | 3,993.57               |
| Average exons per gene                | 10.17                  |
| miRNAs 2307                           | 104                    |
| tRNAs 817                             | 510                    |
| rRNAs 2281                            | 498                    |
| snRNAs 1334                           | 4096                   |
| Percentage of repeat sequences (%)    | 61.66                  |

was supported by the phylogeny. Based on the distribution of  $K_s$ , we found that *T. vernicifluum* may have no recent specific WGD, and it probably shared a WGD event with other Anacardiaceae species before the diversification of Anacardiaceae. The sharp peaks in 4-fold synonymous third codon transversion (4DTV) values ( $K_s = 0.5$ ) also suggest the *T. vernicifluum* genome shared a WGD event with *M. indica* (Figures 2B and 2C).

Syntenic block analysis is often used to determine chromosome evolution among related species. Here, we analyzed the aligned protein sequences of *T. vernicifluum* in comparison with its close relatives *M. indica*. A total of 19,987 syntenic blocks, containing 210,759 pairs of collinear genes, were identified in the *T. vernicifluum* genome, and there were 64,444 colinear gene pairs from 4,867 collinear blocks detected between *T. vernicifluum* and *M. indica*. The comparative genome structure between *T. vernicifluum* and *M. indica* showed relatively high collinearity (Figure 2D). For several collinear regions, the chromosomes of *M. indica* showed a one-to-one syntenic relationship with *T. vernicifluum*. For instance, MiChr8, MiChr15, MiChr16, and MiChr20 of *M. indica* corresponded to TvChr2, TvChr5, TvChr10, and TvChr8. Interestingly, TvChr4 of *T. vernicifluum* corresponded to MiChr13 and of MiChr17 *M. indica* (Figure 2D), indicating that these two chromosomes of *M. indica* might be derived from the chromosome duplication. In addition, we also found that most chromosomes of *T. vernicifluum* matched to more than one chromosome of *M. indica*. For instance, TvChr1 corresponded to MiChr1, MiChr10, MiChr12, and MiChr18; TvChr2 corresponded to MiChr1, MiChr3, MiChr4, MiChr5, MiChr7, MiChr8, MiChr11, MiChr14, MiChr16, MiChr19, and MiChr20; TvChr3 corresponded to MiChr1, MiChr6, MiChr10, and MiChr12. Therefore, it can be inferred that most chromosomes in *M. indica* might have formed by fragmentation and recombination after the divergence between *T. vernicifluum* and *M. indica*, and our analyses also indicated that there was a recent WGD for *M. indica*, but not for *T. vernicifluum*.

### Expansion and contraction of gene families for *T. vernicifluum*

As a result, a total of 15,917 gene families comprising 27,163 genes were homologously identified in the lacquer genome (Table S11), which included 9,945 gene families shared among the four species (*A. occidentale*, *M. indica*, *P. vera*, and *S. birrea*), and 539 gene families containing 1,559 genes were specific to *T. vernicifluum* (Figure 2E). GO enrichment analysis of unique gene families revealed that these

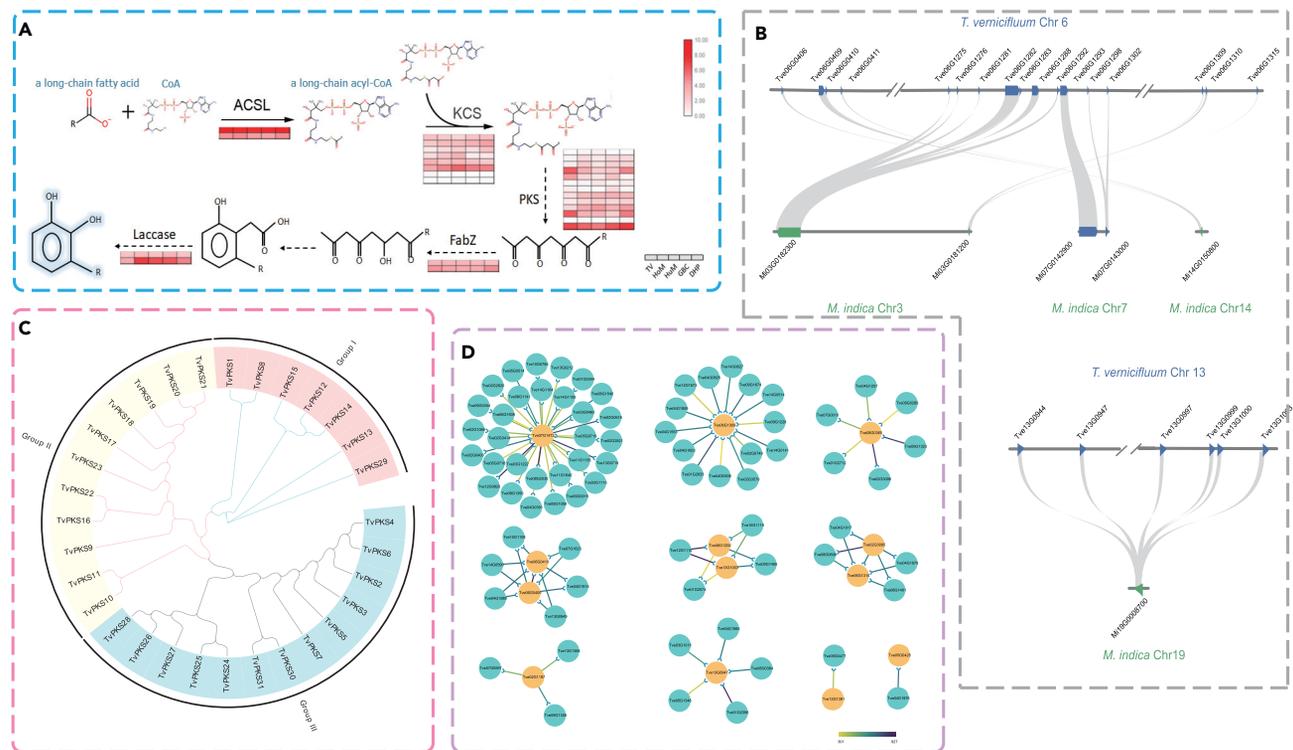


**Figure 2. Comparative genomic analysis of the lacquer tree genome**

- (A) Gene family expansion and contraction among 15 plants. Green and red numbers on the branches represent the expansion and contraction gene families in each species.
- (B) WGD analysis of *T. vernicifluum*.
- (C) 4dTv distributions of syntenic blocks for *T. vernicifluum* paralogs and orthologs with other species of family Anacardiaceae are represented.
- (D) Synteny and microsynteny among *M. indica* and *T. vernicifluum*.
- (E) Venn diagram of *A. occidentale*, *M. indica*, *P. vera*, *S. birrea*, and *T. vernicifluum*. Each number in the diagram is the number of gene family within a group.

specific gene families were considerably enriched in cysteine-type peptidase activity (GO: 0008234), peptidase activity (GO: 0070011), defense response (GO: 0006950), and response to stress (GO: 0006950) (Table S12). A KEGG pathway analysis of *T. vernicifluum* specific gene families revealed significantly enrichment in genes involved in endocytosis, metabolic pathways, phenylpropanoid biosynthesis, and plant-pathogen interaction, possibly contributing to the ecological fitness and biological adaptability of the lacquer tree (Table S13, Figure S8).

To further explore the significant expansion and contraction of gene families in the evolution process, an OrthoFinder analysis of *T. vernicifluum* and other 14 species was performed. Our results showed that 438 families exhibited significant expansion, while 1,707 gene families exhibited significant contraction (Figure 2A). KEGG pathway analysis of the expanded gene families showed marked enrichment in genes involved in functions, such as phenylpropanoid biosynthesis, biosynthesis of unsaturated fatty acids, metabolism of terpenoids and polyketides, and plant-pathogen interaction (Table S14, Figure S9). And the GO enrichment analysis revealed that the expanded gene families were enriched in binding, defense, oxygen oxidoreductase activity, and lignin catabolic process (Table S15). On the other hand, the contracted gene families were enriched in functions related to the phosphotransferase activity, transporter activity, and catalytic activity (Table S16). Furthermore, the contracted gene families showed the marked enrichment in



**Figure 3. Insights into urushiol biosynthesis**

(A) Putative biosynthetic pathway of urushiols in lacquer tree. Enzymes involved at each step of the volatile terpenoid biosynthesis pathway are shown in black. Relative expression profiles of genes implicated in volatile lipid acid and derivative chalcone biosynthesis among lacquer cultivars (TV, the wild lacquer; HoM, Hongmao Guizhou; DHP, Dahongpao; GBC, Gaobachi; HuM, Huangmao Guizhou) are presented as heatmaps. (B) Syntenic analysis among *T. vernicifluum* and *M. indica*. Top: Syntenic analysis of *T. vernicifluum* chr.6, and mango chr. 3, chr. 7, and chr.14. Bottom: Syntenic analysis of *T. vernicifluum* chr.13 and mango chr. 19. Co-syteny regions are linked with gray belts. (C) Phylogenetic relationships of TvPKSs of *T. vernicifluum*. The protein sequences were aligned by Clustal X and generated the phylogenetic tree using MEGA 11.0 by the maximum likelihood method. The PKS family members were classified into three groups, which are denoted by differently colored branches. (D) The protein-protein interaction network of TvPKS by cytoscape.

pathways of phosphatidylinositol signaling system, MAPK signaling pathway, and RNA degradation (Table S17, Figure S10).

### Determination of genes involved in urushiol biosynthesis and a candidate polyketide synthase family

The urushiol synthesis pathway is illustrated in Figure 3A, where the enzymes and their orthologs found in *T. vernicifluum* genome are indicated according to the expression profile of genes related to urushiol biosynthesis. In total, we found 79 orthologs participating in urushiol biosynthesis process (Tables S18 and S19). There are enzymes associated with the synthesis of long-chain fatty acid and their derivatives, including long-chain acyl-CoA synthetase (ACSL), long-chain 3-ketoacyl-CoA synthase (KCS), acetyl-CoA carboxylase biotin carboxyl carrier protein (*accB*), fatty acyl-ACP thioesterase A (*FAT*), fatty acyl-ACP thioesterase B (*FATB*), and 3-oxoacyl-[acyl-carrier-protein] synthase II (*FabF*), as well as enzymes related to polyketide production, such as polyketide synthase (PKS), reductases such as 3-hydroxyacyl-[acyl-carrier-protein] dehydratase (*FabZ*), and laccase (*LAC*). Moreover, RNA-seq was also used to screen the expression levels of urushiol-biosynthesis-related genes from four cultivars and a wild lacquer tree. As a result, we discovered that seven urushiol-biosynthesis-related genes, including *KCS11*, *KCS4*, and *TvPKS9*, 16, 19, 22, 23, were expressed lower in the wild type while they were expressed higher in the cultivars (Figure S11). Our metabolite profile in the five varieties of lacquer trees supports this conclusion. The results showed that urushiol contents, including 3-[8'Z,12'E,15'Z-pentadecatrienyl]-catechol, 3-[8'Z,11'Z,14'-pentadecatrienyl]-catechol, 3-(9E,11E,13Z)-pentadecatrienyl-1,2-catechol and 3-(8,12-heptadecadienyl)-1,2-catechol, 3-(10Z,13E)-10,13-pentadecadienyl-1,2-catechol, urushiol III, and 3-(8Z,11E)-8,11-pentadecadienyl-1,2-catechol, were significantly greater in cultivars than in wild lacquer tree (Table S21).

PKS played a vital role in the initial phase of urushiol biosynthesis and facilitated to yield the key intermediate. We identified 33 orthologs of PKS clustered in Chr 2, Chr 6–8, Chr 12, and Chr 13 using the conserved domain searched in *T. vernicifluum* genome (Table S19). Tve06G1275, Tve06G1276, Tve06G1281, Tve06G1282, Tve06G1283, Tve06G1288, and Tve06G1292 were closely connected to Mi03G0182300 on Chr 3 of the *M. indica* genome. Similarly, the clusters of Tve13G0944, Tve13G0947, Tve13G0997, Tve13G0999, Tve13G1000, and Tve13G1003 were very closely related to Mi19G0008700 in Chr 9 of *M. indica* (Figure 3B). Besides, to evaluate the evolutionary relationships of 33 PKS genes derived from *T. vernicifluum*, all the protein sequences were aligned to construct an unrooted ML phylogenetic tree. PKS genes were found to be divided into three distinct subgroups, implying that their structure and function are divergent (Figure 3C). To estimate who interacted with PKS proteins, we analyzed the PKS's protein interaction network. It showed that the gene pairs formed as TvPKS13 and TvPKS14, TvPKS23 and TvPKS7, and TvPKS9 and TvPKS29 interacted with phenylpropanoid pathway enzymes, such as reductase (Tv09G1323), CHS (Tv02G3096), chalcone isomerase (Tv01G2833), and 4CL (Tv09G1674, Tv14G0514), which are involved in phenylpropanoid pathways (Figure 3D). Quantitative real-time PCR (qRT-PCR) analysis was conducted to validate the expression levels of enzyme genes involved in urushiol pathway in different cultivars, such as *ACSL1*, *KCS3*, *KCS6*, *PKS4*, and *PKS11*, and compared with their abundance from sequencing data. The qRT-PCR analysis showed that *ACSL1* was highly expressed in GBC and HoM, and *KCS3* has significantly high expression in DHP, HoM, and HuM, which is generally agreed well with the sequencing result (Figure S13).

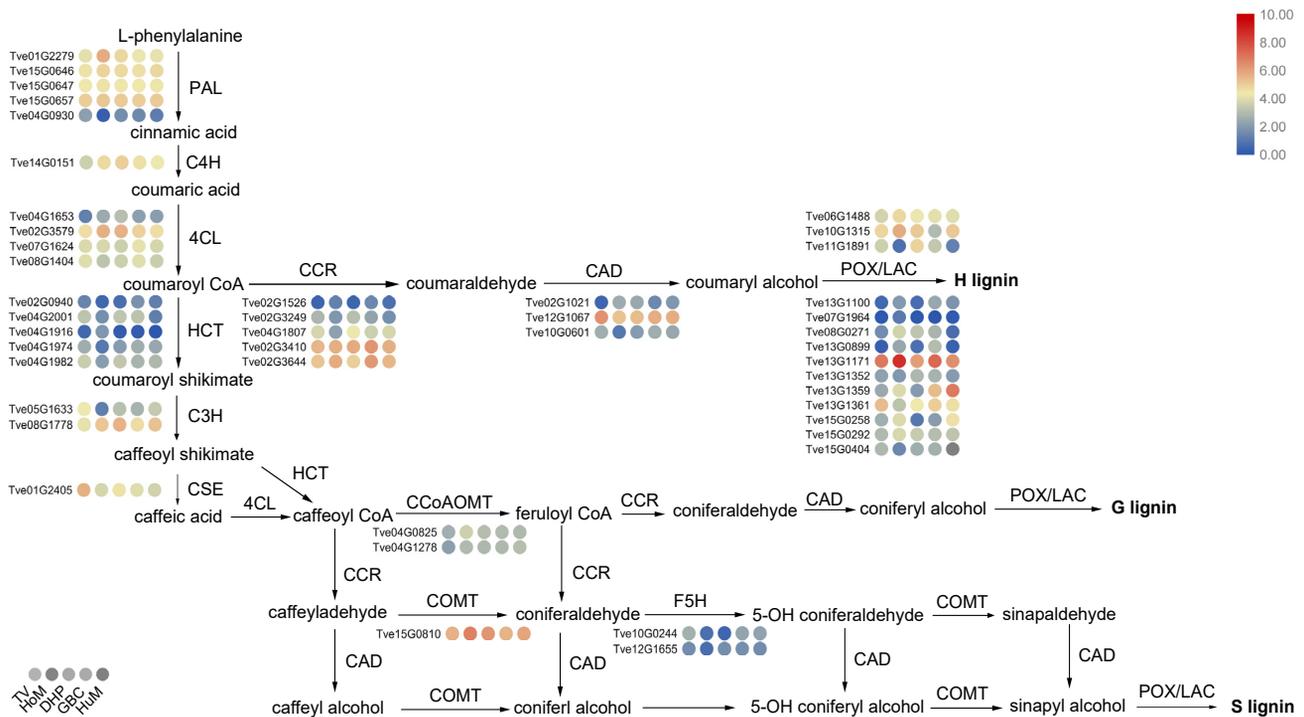
### Screening for genes encoding enzymes of the lignin biosynthetic pathway

Clarifying lacquer's lignin production route is highly beneficial in understanding the growth law and the distinctions between natural tree and artificially produced types. Using the Basic Local Alignment Search Tool (BLAST), we identified 177 orthologous genes involved in lignin biosynthesis based on our lacquer's genome database (Table S20), including phenylalanine ammonia-lyase (*PAL*), cinnamate 4-hydroxylase (*C4H*), 4-coumarate: CoA ligase (*4CL*), hydroxycinnamoyl-CoA: shikimate hydroxycinnamoyl transferase (*HCT*), cinnamoyl-CoA reductase (*CCR*), *p*-coumarate 3-hydroxylase (*C3H*), cinnamyl alcohol dehydrogenase (*CAD*), caffeoyl-CoA O-methyltransferase (*CCoAOMT*), caffeic acid O-methyltransferase (*COMT*), ferulate 5-hydroxylase (*F5H*), caffeoylshikimate esterase (*CSE*), lignin-forming anionic peroxidase (*POX*), and laccases (*LAC*). Among them, *CCR* catalyzes the coumarin production, plays an important role in regulating the carbon flow of lignin biosynthesis pathway. In total, we found 31 *CCR* homology genes in lacquer genome, while 31 are present in pear (*Pyrus bretschneideri*) (Cheng et al., 2017), 14 are present in soybean (Cui et al., 2021), and six are present in mulberry (Chao et al., 2021). The TvCCR phylogenetic tree constructed with CCRs from *Arabidopsis thaliana*, *O. sativa*, *Populus*, as well as *P. vera* and *S. birrea* revealed that CCRs are clustered into four groups (Figure S12), which indicated that CCRs had structural and functional divergence. Additionally, our transcriptome data revealed diverse expression of lignin biosynthesis genes in both wild and cultivars of lacquer tree (Figure 4), which will provide fundamental insights into engineering of lacquer tree for lignin production. Similarly, qRT-PCR analysis was carried out to analyze the expression level of four genes related to lignin biosynthesis pathway to confirm the accuracy of the Illumina expression profiles. It showed that *CCR4*, *CCR5*, and *LAC5* were significantly highly expressed in GBC, and *LAC5* were expressed higher in HoM, which is agreed well with the sequencing data (Figure S13).

### Metabolite profiling for characterization of bioactive chemicals in lacquer trees

The extractions from lacquer tree are also used as a traditional medicinal herb in China for its various flavonoids, unsaturated fatty acids, and polysaccharides. Notably, comparative genomics revealed that 295 gene families were expanded in the category of "Biosynthesis of secondary metabolites", and 10 for "Biosynthesis of unsaturated fatty acids", 23 for "Flavonoid biosynthesis", and 31 for "Metabolism of terpenoids and polyketides", respectively (Figure S9). These enriched genes may contribute to enhancing the biosynthesis of the important bioactive metabolites such as flavonoids, unsaturated fatty acids, and urushiol biosynthesis in the lacquer tree.

The phloem of these five samples was collected for metabolite profiling by UPLC-MS/MS to examine the difference of lacquer tree cultivars "Dahongpao (DHP)", "Hongmao Guizhou (HoM)", "Huangmao Guizhou (HuM)", "Gaobachi (GBC)", and wild type (TV) in metabolism levels (Figures 5A and 5B). In the results, 12 classifications containing 862 metabolites were detected in lacquer tree phloem, including flavonoids (178), phenolic acids (166), lipids (100), organic acids (82), lignans and coumarins (31), etc (Table S21). We found differences in metabolites between cultivars and wild trees. There are 165 metabolites including flavonoid, flavanols, lignans, tannin, alkaloids, and intermediates or derivatives in urushiol biosynthesis,



**Figure 4. Diagram of the lignin biosynthetic pathway**

Heatmaps showing the expression levels of lignin-biosynthesis-related genes at lacquers (TV, the wild lacquer; HoM, Hongmao Guizhou; DHP, Dahongpao; GBC, Gaobachi; HuM, Huangmao Guizhou). The expression values (FPKM) at the row scale were normalized. Higher mRNA expression levels are indicated by red, and lower mRNA expression levels are indicated by blue.

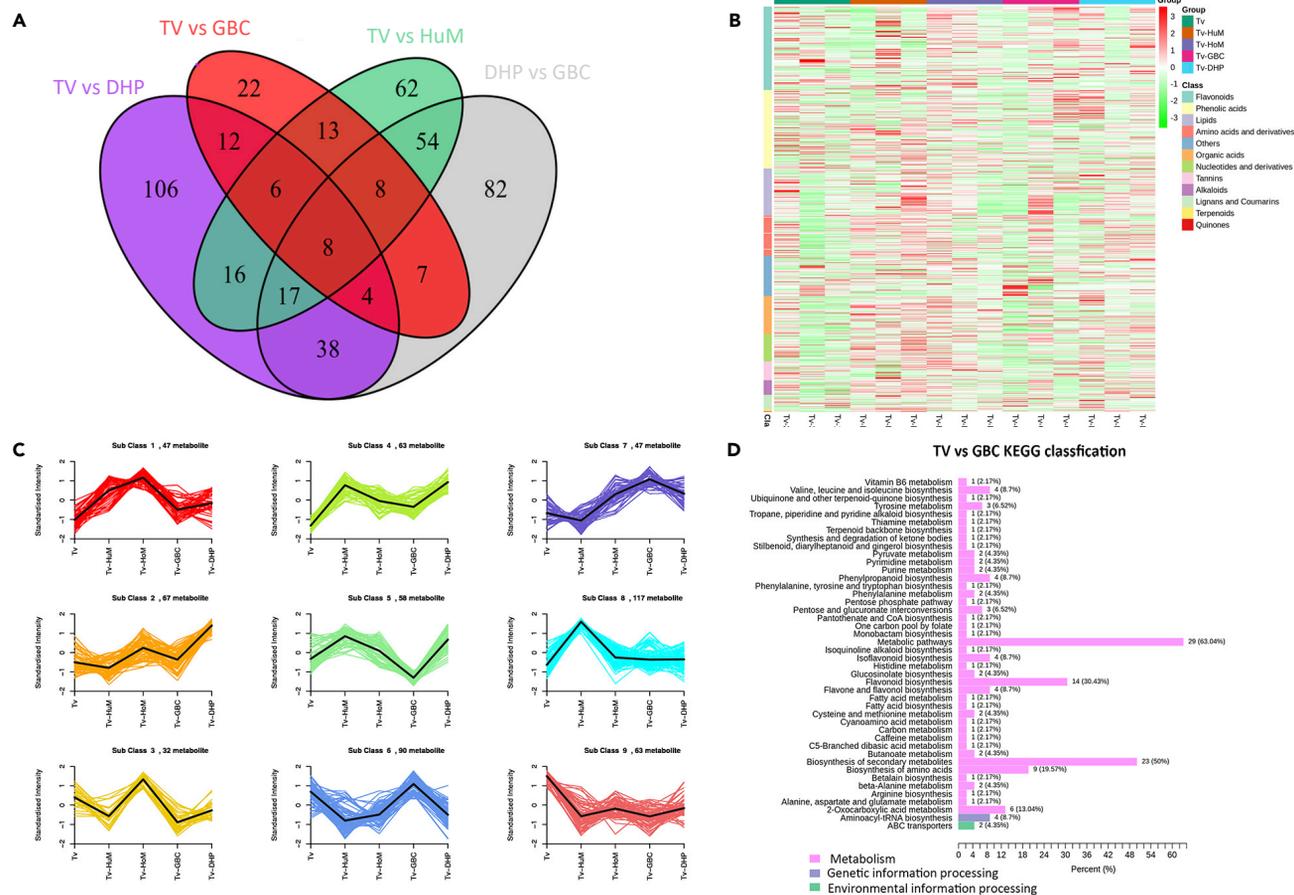
such as 3-(10E)-10-pentadecenyl-1,2-benzenediol, 3-(10E)-10-pentadecenyl-1,2-benzenediol, 3-(10,13,16-heptadecatrienyl)-(Z,Z)-1,2-benzenediol, 3-(8,10,12-heptadecatrienyl)-1,2-benzenediol, 3-(8,11,14-heptadecatrienyl)-(Z,Z)-1,2-benzenediol, 3-(8,11-heptadecadienyl)-1,2-benzenediol, and 2-(10Z,13E)-10,13-heptadecadienyl-1,4-benzenediol, were found to be more abundant than TV (Figures 5D and S14). And we also found the chemicals related to urushiol biosynthesis, such as long-chain fatty acids, urushiols, and their analogs like pentadecatrienyl/pentadecadienyl-benzenediols (Table S22). In subcluster 7, it is shown that 24 flavonoids including isoliquiritigenin, butein, and scutellarin were significantly upregulated in GBC compared to wild lacquer tree (Figure 5C and Table S23).

Compared to wild lacquer tree, 231 metabolites including flavonoids, phenolic acids, lipids, organic acids, alkaloids, lignans, coumarins, amino acids and derivatives, and terpenoids were significantly changed in cultivar DHP. Similarly, there were 218 special metabolites in HoM and 287 significantly different metabolites in HuM compared with wild lacquer tree. Most phenolic acids, alkaloids, lipids, organic acids, and terpenoids were found to be abundant in cultivars than in the wild. Notably, 1-*O-p*-coumaroylquinic acid, cirsimaritin, eupatorin, persicoside, limocitrin-3-*O*-sophoroside, limocitrin-3,7-di-*O*-glucoside, amentoflavone-7''-*O*-glucoside, quercetin-3-*O*-(2'''-Feruoyl) sophoroside, syringaresinol-4'-*O*-(6''-acetyl) glucoside, and jasmonic acid were exclusively detected in the cultivars and not in the wild lacquer tree. Accordingly, there is also an apparent deficiency of metabolites in the cultivars. For example, we did not detect sexangularetin-3-*O*-glucoside-7-*O*-rhamnoside in HuM, and we did not detect protocatechuic acid ethyl ester in HoM, which existed in the wild lacquer tree (Table S21). This metabolic profile among wild and cultivar lacquer trees will provide data for excellent variety breeding and comprehensive utilization of lacquer trees.

## DISCUSSION

### Lacquer genome is the first one at chromosome level in *Toxicodendron*

A high-quality reference genome for lacquer tree can facilitate lacquer molecular breeding and evolutionary research of Anacardiaceae, and it is also valuable resource for studying the molecular basis of specialized metabolites diversity and how their biosynthetic pathways evolved. Here, we generated the



**Figure 5. Metabolite profiling of lacquer cultivars**

(A) The Venn showed the different metabolites among cultivars (TV, the wild lacquer; HoM, Hongmao Guizhou; DHP, Dahongpao; GBC, Gaobachi; HuM, Huangmao Guizhou).  
 (B) The metabolite profiling of lacquer cultivars Tv, DHP, GBC, HoM, and HuM.  
 (C) Metabolites cluster analysis by k-mer among the cultivars.  
 (D) The KEGG pathway enrichment of different metabolites between TV and GBC.

first reference genome of the lacquer tree with 491.93 Mb in size at chromosome level. The contig N50 and scaffold N50 sizes are 5.26 and 32.97 Mb, respectively, and 98.9% of the single-copy orthologs based on the BUSCO analysis, which suggested the *T. vernicifluum* genome has good assembly and the largest genome completeness (Table 1). In addition, the heterozygosity of the lacquer tree (0.56%) is lower than that of its related species, which is 0.68% and 1.72% in *M. indica* and *P. vera*, respectively (Wang et al., 2020b; Zeng et al., 2019). Owing to the lack of a genome background, the advancement of the exploration on urushiol biosynthesis pathway, gene editing, and molecular breeding of lacquer tree were hindered. In this study, the chromosome-level genome of lacquer tree not only provides a critical resource for *T. vernicifluum* functional genomic studies, particularly the investigation of the urushiol biosynthesis pathway, but also contributes to the understanding of *T. vernicifluum* biology and evolution.

### Evolution of lacquer genome

The monophyly of malvids and close phylogenetic relationships of three Sapindales order families (Anacardiaceae, Rutaceae, and Sapindaceae) are well supported by phylogenomic analysis of 1,914 single-copy genes from 14 typical seed plant genomes. The phylogenetic signals generated in this work are consistent with the plastid genome phylogenies (Li et al., 2019). Besides, within the Anacardiaceae, we found that *T. vernicifluum* was more closely related to *P. vera*. This result is in agreement with the phylogeny based on nuclear DNA and plastid markers (Jiang et al., 2019; Wang et al., 2020a). Therefore, it means that phylogenomics is effective to rebuilt the robust phylogeny for angiosperm or even for specific taxa. However,

future phylogenomic studies should recover sufficient lineage sampling and gene sampling as well as comprehensive analytical methods to provide more convincing phylogenetic evidence for research taxa.

WGD event or polyploidization seems to have occurred throughout the routes of most plant evolution, which can facilitate the emergence and specialization of novel traits and functions (Jiao et al., 2011; Wolfe, 2001). The genome of *T. vernicifluum* obtained in this study improved our understanding of the timing of the WGD event in Anacardiaceae. Previous studies indicated that there was no lineage-specific whole genome duplication occurred in *P. vera* (Guo et al., 2017), and our results also verified no lineage-specific WGD event for *T. vernicifluum* which might occur after the split of *P. vera*. The WGD time of *T. vernicifluum* genome (~47.3 Ma) occurred before the split of Trib. Anacardiaceae (29.6 Mya). In addition, 4DTv analyses also confirmed no specific WGD event in the *T. vernicifluum*. It is worth mentioning that *M. indica* genome shared high collinearity with *T. vernicifluum*, and it shared ancient WGD with other species belong to Trib. Anacardiaceae (Wang et al., 2020b). Unexpectedly, *M. indica* genome shared high collinearity with *T. vernicifluum* experienced a recent specific recent WGD (~33 MYA) which after the ancient WGD event time for *T. vernicifluum* (Wang et al., 2020b), and it had been found that some retained genes from WGD of *M. indica* were involved in the biosynthesis of secondary metabolisms. It is appealing to hypothesize that the WGD in *T. vernicifluum* genome may also impact the different metabolism categories with prominent contributions.

### Draft maps of urushiol and lignin biosynthesis pathway

Genome sequencing is a powerful tool for studying various aspects of physiology and genetics in non-model plants, particularly in medical herbs and economical important species (Li et al., 2021; Ma et al., 2021; Tang et al., 2016; Zeng et al., 2019; Zhao et al., 2019). For example, waterlily, *Magnolia officinalis*, and persimmon draft genomes provides information into their evolutionary process (Yin et al., 2021; Zhang et al., 2020; Zhu et al., 2019), and *Cucumis* and *Vitis* genus genomes provide an understanding of important agronomic traits and stress response (Patel et al., 2020; Wang et al., 2021; Yu et al., 2021). Furthermore, many reference genomes, such as *Scutellaria baicalensis*, *Panax Notoginseng*, *Senna tora*, *Lithospermum erythrorhizon*, and garlic (*Allium sativum*) shed light on genes involved in the biosynthesis of specific flavonoids, terpenes, alkaloids, lignan, and a number of other secondary metabolites (Auber et al., 2020; Fan et al., 2020; Kang et al., 2020; Sun et al., 2020; Yin et al., 2021).

Phenolic compounds such as flavonoids, coumarins, and phenolic acids and their derivatives exert ecological functions as protectors against the attack of herbivores and microorganisms, attractors of pollinating species, and in plant-plant communication (Marchiosi et al., 2020). Lignin is a complex heteropolymer of simple phenolic acids that gives the plant cell wall mechanical resistance and is one of the most abundant biopolymers on planet (Maeda, 2016). Lignin participates in cell and tissue physiology and cell-to-cell communication for the adaptation of plants. Economically, lignin controls the wood quality and impacts the growth rate and harvest years for lacquer trees. The most genes encoding the enzymes involved in the biosynthesis of lignin have been cloned and characterized in many plants (Goujon et al., 2003; Maeda, 2016; Shen et al., 2013; Xu et al., 2009). It is still a rare report on functional identification of lignin-related genes in lacquer trees. In this study, we identified 177 orthologous genes involved in lignin biosynthesis according to homological search on our lacquer genome database (Figure 4, Table S20). These genes were expanded in the lacquer genome, indicating more genes participated in lignin biosynthesis. Cinnamoyl-CoA reductase (CCR) is the first enzyme in the monolignol-specific branch of the lignin biosynthesis pathway, where it converts feruloyl-CoA to coniferaldehyde (Leple et al., 2007). Downregulation of CCR in tobacco significantly reduces lignin content and the rate of plant development, as characterized by xylem that collapses and turns an orange-brown color (Piquemal et al., 2002). CCR in plant differs structurally and functionally. For example, AtCCR1 is expressed in the flowers, leaves, and stems, undergoing a lignification process, whereas AtCCR2 is lower expressed in normal stages but highly triggered by pathogen invasion (Lauvergeat et al., 2001). CCR genes in poplar have similar properties, such as selective expression in leaves, bark, and xylem (Barakat et al., 2011). Our finding revealed that the 30 TvCCRs were clustered into four groups, whereas the most CCRs were divided into two subgroups separately (Figure S12), which indicates that TvCCRs could exist with structural and functional divergence from CCRs in other plants. In addition, in a combined transcriptome and metabolic profile, the mediates, such as cinnamic acid, *p*-coumaryl alcohol, and caffeic acid, along with the related genes, such as *C4H*, *4CL*, and *CAD*, were significantly expressed higher in cultivars than in the wild, implying that artificial selection or cultivation has altered lacquer lignin synthesis and improved its

woody properties. The lignin biosynthesis pathway of the lacquer tree has been identified, laying the groundwork for future molecular breeding and cultivation studies.

Urushiols are produced in many Anacardiaceae plants, particularly those in the genus *Toxicodendron*, such as *T. vernicifluum* (lacquer tree), *T. radicans* (poison ivy), and *T. diversilobum* (poison oak), and often cause contact dermatitis in sensitive individuals (Mina et al., 2017). Urushiol is alk-(en)-yl catechol derivatives with a 15- or 17-carbon side chain that are probable insect, fungus, or herbivorous defense compounds (Kim et al., 2019a, 2019b). Urushiol has been shown to have substantial antioxidant, immune-enhancing, neuroprotective, anti-inflammatory, and antitumor properties in numerous studies (Cho et al., 2012; Kim et al., 2015c, 2015d; Lee et al., 2004, 2015, 2017). Kadokura et al. also found two novel urushiols with HIV-1 reverse transcriptase inhibitory activity, indicating that urushiol could be a potential bioactive agent in AIDS treatment (Kadokura et al., 2015). Since 1982, many analytical chemistry techniques have been utilized successively to isolate and identify urushiol from *Toxicodendron* plants, such as gas chromatography-mass spectrometry (GC-MS), LC-MS-MS, matrix-assisted laser desorption/ionization-mass spectrometry imaging (MALDI-MSI), two-dimensional (2D)  $^1\text{H}$ - $^{13}\text{C}$  heteronuclear multiple quantum coherence (HMQC), and heteronuclear multiple bond coherence (HMBC) NMR spectroscopic techniques (Draper et al., 2002; ElSohly et al., 1982; Harigaya et al., 2007; Mina et al., 2017). However, little is currently known about the biosynthetic pathway of urushiol in the *Toxicodendron* genus. The urushiol biosynthesis pathway has been proposed to be involved in fatty-acid metabolism with polyketide synthase being the first step (Fu et al., 2007). Wang et al. proposed that the early genes in phenolic biosynthesis pathways like chalcone synthase (CHS) have extensive expansion in the mango genome, highlighting the importance to the metabolic flow of phenols (Wang et al., 2020b). So far, there is still a lack of details and more genes involved in this synthetic framework. In our data, we found 79 orthologs containing 33 TvPKS genes participated in urushiol biosynthesis according to their co-expression profile and homological structure. Based on our data, we drew the skeleton of the urushiol biosynthesis pathway and provided more details at whole genome level. Moreover, the expression levels of urushiol-related genes (*KCS11*, *KCS4*, and *TvPKS9*, 16, 19, 22, 23) were linked to the amount of urushiol in the lacquer cultivars, further indicating those genes might participate in the urushiol biosynthesis. TvPKS genes also preferred to cluster together on the chromosome, showing that they evolved from a common ancestor gene or co-transcript to perform the same function (Figure 3B). Besides, TvPKSs could interact with enzymes in phenylpropanoid pathways, such as CHS, CHI, and 4CL (Figure 3D). This may be due to the broad promiscuity of the substrates of PKS, which enables TvPKSs to catalyze reactions on the phenylpropanoid pathway, either competitively or synergistically interacting with the 4CL, CHS, etc. This interaction might allow for fine switching of metabolic flow between different metabolic pathways. However, the speculation still requires further confirmation by enzymology or crystal structural and catalytic machinery analysis. In this study, we identified urushiol-related genes genome-wide and described generally the synthetic pathway of urushiol, updating the understanding of urushiol biosynthesis and providing candidate key enzyme genes for further genetic and enzymatic verification.

### Metabolic variations in lacquer cultivars

The metabolomic profiling of four cultivars and the wild type of lacquer tree depicted the presence of bioactivities, such as flavonoids, phenolic acids, terpinene, lignans, and coumarins (Table S21). And we also found the chemicals related to urushiol biosynthesis, such as long-chain fatty acids, urushiols, and their analogs like pentadecatrienyl/pentadecadienyl-benzenediols (Table S22). The isolation and pharmacological identification of bioactive compounds from lacquer tree has been a popular field of research for decades, and several researchers have explored those bioactivities have antioxidant, antitumor, anti-inflammatory, antibacterial, antiviral, and neuroprotective potential (Hong et al., 2013; Kim et al., 2019a, 2019b; Choi et al., 2013; Lee et al., 2010; Cho et al., 2012; Kang et al., 2012). Fisetin, a natural flavonoid found in lacquer tree has been proved to process the excellent multiple pharmacological activities, such as antioxidant, anti-inflammatory, and anticancer effects in various cell types (Higa et al., 2003; Kim et al., 2015b; Sabarwal et al., 2017). It is reported that fisetin could induce autophagy via the crosstalk between the AMPK/mTOR and p8-dependent pathways in pancreatic cancer cells (Jia et al., 2019). And Chamcheu et al. suggested recently that fisetin has a great potential to be developed as an effective and inexpensive agent for the treatment of psoriasis and other related inflammatory skin disorders (Chamcheu et al., 2019). In addition, another flavonoid butin is potent antioxidant against oxidative stress-related diseases, such as cancer, aging, liver diseases, and diabetes (Brusselmans et al., 2005; Patil et al., 2003; Kuzu et al., 2008; Shu et al., 2009). Many other major flavonoids isolated from lacquer tree, such as sulfuretin, naringin, and pinocembrin, are known to have anti-inflammatory effects, and may have therapeutic value in preventing or delaying the

progression of rheumatoid arthritis, diabetes mellitus, and other inflammatory disease (Lee et al., 2012; Kandhare et al., 2014; Gu et al., 2017). In addition to flavonoid active components, the lacquer tree includes phenolic acids and terpenoids, such as caffeic acid and ursolic acid, with strong antioxidant, antibacterial, and antiviral properties, which are extremely healthy for us (Jayanthi and Subash, 2010; Tohmé et al., 2019). Although we detected many active ingredients in this study, unfortunately, intermediates in the urushiol biosynthesis pathway could not be found possibly due to structural instability or low content in the sample.

The metabolic profiling is often inferred that metabolic information reflects biological endpoints more accurately than transcript or protein analysis. The cultivars have their own metabolite compositions during developmental processes or seasons. For example, the bioactive contents and the antioxidant potential of leaves of olive cultivars (Arbequina, Manzanilla, and Picual) collected in four periods of the year showed differences (Lorini et al., 2021). And the metabolism analysis of two varieties revealed that grape cultivars undergo several changes in primary metabolite concentration during berry developmental progression (Cuadros-Inostroza et al., 2016). Therefore, metabolomics is the key technology to screen the metabolic indicators of plants. Meanwhile, the metabolic markers can be used as biomarkers to identify plant species or varieties, and even the population (Sarrou et al., 2017; Sawada et al., 2019).

In our data, compared with wild lacquer, the cultivars process different contents and compositions of urushiol, and they also process their unique metabolites, such as cirsimaritin, eupatorin, persicoside, etc. Among the cultivars, we found that limocitrin-3-*O*-sophoroside and persicoside are the specific chemicals in HuM, and scutellarin is the unique metabolite in HoM. 7,8-Dihydroxy-5,6,4'-trimethoxyflavone and eriodictyol-7-*O*-(6''-malonyl) glucoside are the specific metabolites in DHP, while limocitrin-3,7-di-*O*-glucoside is specific to GBC. These specific metabolites could be used as metabolic markers to distinguish the four lacquer cultivars. And the special chemicals they made could be formed by long-term domestication, mutagenesis, breeding, and selection. Otherwise, the unique flavonoids in lacquer trees (such as butin, fisetin, fustin, and sulfuretin) as the pharmacological foundation could produce beneficial and safe functional diets for preventing neurological disorders (Afzal et al., 2021; Kim et al., 2015b; Zhang et al., 2011). These metabolome data provide a basis for the comprehensive utilization of the lacquer tree.

### Limitations of the study

We sequenced the genome, transcriptome data, and metabolome data to provide the genetic background for the lacquer tree. We identified the candidate genes associated with the urushiol and lignin biosynthesis pathways only by analyzing the expression profile and homology blast analysis, with no experimental validation. Therefore, the function of these candidate genes should be further verified using genetic and biochemical technologies. Moreover, the metabolome data should be further developed into a stable and reliable method of identifying more lacquer cultivars and applied in the lacquer planting industry.

### STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
- METHOD DETAILS
  - Genome sequencing and assembly
  - Gene prediction and annotation
  - Assembly and gene set evaluation
  - Phylogenetic analyses
  - Genome synteny and whole-genome duplication analysis
  - Identification of polyketide synthases genes from genome sequences
  - Metabolomic profile of lacquer cultivars
  - qRT-PCR analysis
- QUANTIFICATION AND STATISTICAL ANALYSIS
- ADDITIONAL RESOURCES



- flavonoids isolated from *Rhus verniciflua* in neuronal HT22 and microglial BV2 cell lines. *Food Chem. Toxicol.* 50, 1940–1945. <https://doi.org/10.1016/j.fct.2012.03.052>.
- Choi, H.S., Seo, H.S., Kim, S.R., Choi, Y.K., Jang, B.H., Shin, Y.C., and Ko, S.G. (2014). Anti-inflammatory and anti-proliferative effects of *Rhus verniciflua* Stokes in RAW264.7 cells. *Mol. Med. Rep.* 9, 311–315. <https://doi.org/10.3892/mmr.2013.1775>.
- Boachon, B., Buell, C.R., Crisovan, E., Dudareva, N., Garcia, N., Godden, G., Henry, L., Kamileen, M.O., Kates, H.R., Kilgore, M.B., et al. (2018). Phylogenomic mining of the mints reveals multiple mechanisms contributing to the evolution of chemical diversity in Lamiaceae. *Mol. Plant* 11, 1084–1096. <https://doi.org/10.1016/j.molp.2018.06.002>.
- Cuadros-Inostroza, A., Ruiz-Lara, S., González, E., Eckardt, A., Willmitzer, L., and Peña-Cortés, H. (2016). GC-MS metabolic profiling of Cabernet Sauvignon and Merlot cultivars during grapevine berry development and network analysis reveals a stage- and cultivar-dependent connectivity of primary metabolites. *Metabolomics* 12, 39. <https://doi.org/10.1007/s11306-015-0927-z>.
- Cui, J., Lu, Z., Wang, T., Chen, G., Mostafa, S., Ren, H., Liu, S., Fu, C., Wang, L., Zhu, Y., et al. (2021). The genome of *Medicago polymorpha* provides insights into its edibility and nutritional value as a vegetable and forage legume. *Hortic. Res.* 8, 47. <https://doi.org/10.1038/s41438-021-00483-5>.
- De Bie, T., Cristianini, N., Demuth, J.P., and Hahn, M.W. (2006). CAFE: a computational tool for the study of gene family evolution. *Bioinformatics* 22, 1269–1271. <https://doi.org/10.1093/bioinformatics/btl097>.
- De Coster, W., D’Hert, S., Schultz, D.T., Cruets, M., and Van Broeckhoven, C. (2018). NanoPack: visualizing and processing long-read sequencing data. *Bioinformatics* 34, 2666–2669. <https://doi.org/10.1093/bioinformatics/bty149>.
- Draper, W.M., Wijekoon, D., McKinney, M., Behniwal, P., Perera, S.K., and Flessel, C.P. (2002). Atmospheric pressure ionization LC-MS-MS determination of urushiol congeners. *J. Agric. Food Chem.* 50, 1852–1858. <https://doi.org/10.1021/jf011364t>.
- Dudchenko, O., Batra, S.S., Omer, A.D., Nyquist, S.K., Hoeger, M., Durand, N.C., Shamim, M.S., Machol, I., Lander, E.S., Aiden, A.P., and Aiden, E.L. (2017). De novo assembly of the *Aedes aegypti* genome using Hi-C yields chromosome-length scaffolds. *Science* 356, 92–95. <https://doi.org/10.1126/science.aal3327>.
- Durand, N.C., Shamim, M.S., Machol, I., Rao, S.S., Huntley, M.H., Lander, E.S., and Aiden, E.L. (2016). Juicer provides a one-click system for analyzing loop-resolution Hi-C experiments. *Cell Syst.* 3, 95–98. <https://doi.org/10.1016/j.cels.2016.07.002>.
- Edgar, R.C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 32, 1792–1797. <https://doi.org/10.1093/nar/gkh340>.
- Edgar, R.C., and Myers, E.W. (2005). PILER: identification and classification of genomic repeats. *Bioinformatics* 21, i152–i158. <https://doi.org/10.1093/bioinformatics/bti1003>.
- ElSohly, M.A., Adawadkar, P.D., Ma, C.Y., and Turner, C.E. (1982). Separation and characterization of poison ivy and poison oak urushiol components. *J. Nat. Prod.* 45, 532–538. <https://doi.org/10.1021/np50023a004>.
- Emms, D.M., and Kelly, S. (2015). OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biol.* 16, 157. <https://doi.org/10.1186/s13059-015-0721-2>.
- Fan, G., Liu, X., Sun, S., Shi, C., Du, X., Han, K., Yang, B., Fu, Y., Liu, M., Seim, I., et al. (2020). The chromosome level genome and genome-wide association study for the agronomic traits of *Panax Notoginseng*. *iScience* 23, 101538. <https://doi.org/10.1016/j.isci.2020.101538>.
- Cao, X., Li, S., Wang, C., and Lu, M. (2007). Distributed regularity and anatomic feature of laticiferous canals of *Toxicodendron vernicifluum* (Stokes) F. A. Barkley. *Acta Bot. Boreali-Occid. Sin.* 23, 885–890. [https://doi.org/10.1016/S1872-2075\(07\)60055-7](https://doi.org/10.1016/S1872-2075(07)60055-7).
- Goujon, T., Sibout, R., Eudes, A., MacKay, J., and Jouanol, L. (2003). Genes involved in the biosynthesis of lignin precursors in *Arabidopsis thaliana*. *Plant Physiol. Biochem.* 41, 677–687. [https://doi.org/10.1016/S0981-9428\(03\)00095-0](https://doi.org/10.1016/S0981-9428(03)00095-0).
- Griffiths-Jones, S., Moxon, S., Marshall, M., Khanna, A., Eddy, S.R., and Bateman, A. (2004). Rfam: annotating non-coding RNAs in complete genomes. *Nucleic Acids Res.* 33, D121–D124. <https://doi.org/10.1093/nar/gki081>.
- Gu, X., Zhang, Q., Du, Q., Shen, H., and Zhu, Z. (2017). Pinocembrin attenuates allergic airway inflammation via inhibition of NF- $\kappa$ B pathway in mice. *Int. Immunopharm.* 53, 90–95. <https://doi.org/10.1016/j.intimp.2017.10.005>.
- Guo, W., Wu, H., Zhang, Z., Yang, C., Hu, L., Shi, X., Jian, S., Shi, S., and Huang, Y. (2017). Comparative analysis of transcriptomes in rhizophoraceae provides insights into the origin and adaptive evolution of mangrove plants in intertidal environments. *Front. Plant Sci.* 8, 795. <https://doi.org/10.3389/fpls.2017.00795>.
- Harigaya, S., Honda, T., Rong, L., and Miyakoshi, T. (2007). Enzymatic dehydrogenative polymerization of urushiols in fresh exudates from the lacquer tree, *Rhus vernicifera* DC. *J. Agr. Food Chem.* 55, 2201–2208. <https://doi.org/10.1021/jf063161g>.
- Higa, S., Hirano, T., Kotani, M., Matsumoto, M., Fujita, A., Suemura, M., Kawase, I., and Tanaka, T. (2003). Fisetin, a flavonol, inhibits th2-type cytokine production by activated human basophils. *J. Allergy Clin. Immunol.* 111, 1299–1306. <https://doi.org/10.1067/mai.2003.1456>.
- Honda, T., Lu, R., Sakai, R., Ishimura, T., and Miyakoshi, T. (2008). Characterization and comparison of Asian lacquer saps. *Prog. Org. Coat.* 61, 68–75. <https://doi.org/10.1016/j.porgcoat.2007.09.003>.
- Hong, S.H., Suk, K.T., Choi, S.H., Lee, J.W., Sung, H.T., Kim, C.H., Kim, E.J., Kim, M.J., Han, S.H., Kim, M.Y., et al. (2013). Anti-oxidant and natural killer cell activity of Korean red ginseng (*Panax ginseng*) and urushiol (*Rhus vernicifera* Stokes) on non-alcoholic fatty liver disease of rat. *Food Chem. Toxicol.* 55, 586–591. <https://doi.org/10.1016/j.fct.2013.01.022>.
- Jayanthi, R., and Subash, P. (2010). Antioxidant effect of caffeic acid on oxytetracycline induced lipid peroxidation in albino rats. *Indian J. Clin. Biochem.* 25, 371–375. <https://doi.org/10.1007/s12291-010-0052-8>.
- Jia, S., Xu, X., Zhou, S., Chen, Y., Ding, G.P., Cao, L., and Gao, L.P. (2019). Fisetin induces autophagy in pancreatic cancer cells via endoplasmic reticulum stress- and mitochondrial stress-dependent pathways. *Cell Death Dis.* 10, 142. <https://doi.org/10.1038/s41419-019-1366-y>.
- Jiang, T., Zhang, M., Wen, C., Xie, X.L., Tian, W., Wen, S.Q., Lu, R.K., and Liu, L.D. (2020). Integrated metabolomic and transcriptomic analysis of the anthocyanin regulatory networks in *Salvia miltiorrhiza* Bge. flowers. *BMC Plant Biol.* 20, 349. <https://doi.org/10.1186/s12870-020-02553-7>.
- Jiang, Y., Gao, M., Meng, Y., Wen, J., Ge, X.J., and Nie, Z.L. (2019). The importance of the North Atlantic land bridges and eastern Asia in the post-Boreotropical biogeography of the Northern Hemisphere as revealed from the poison ivy genus (*Toxicodendron*, Anacardiaceae). *Mol. Phylogenet. Evol.* 139, 106561. <https://doi.org/10.1016/j.ympev.2019.106561>.
- Jiang, Z., Tu, L., Yang, W., Zhang, Y., Hu, T., Ma, B., Lu, Y., Cui, X., Gao, J., Wu, X., et al. (2021). The chromosome-level reference genome assembly for *Panax notoginseng* and insights into ginsenoside biosynthesis. *Plant Commun.* 2, 100113. <https://doi.org/10.1016/j.xplc.2020.100113>.
- Jiao, Y., Wickett, N.J., Ayyampalayam, S., Chanderbali, A.S., Landherr, L., Ralph, P.E., Tomsho, L.P., Hu, Y., Liang, H., Soltis, P.S., et al. (2011). Ancestral polyploidy in seed plants and angiosperms. *Nature* 473, 97–100. <https://doi.org/10.1038/nature09916>.
- Kadokura, K., Suruga, K., Tomita, T., Hiruma, W., Yamada, M., Kobayashi, A., Takatsuki, A., Nishio, T., Oku, T., and Sekino, Y. (2015). Novel urushiols with human immunodeficiency virus type 1 reverse transcriptase inhibitory activity from the leaves of *Rhus verniciflua*. *J. Nat. Med.* 69, 148–153. <https://doi.org/10.1007/s11418-014-0871-7>.
- Kandhare, A.D., Ghosh, P., and Bodhankar, S.L. (2014). Naringin, a flavanone glycoside, promotes angiogenesis and inhibits endothelial apoptosis through modulation of inflammatory and growth factor expression in diabetic foot ulcer in rats. *Chem. Biol. Interact.* 219, 101–112. <https://doi.org/10.1016/j.cb.2014.05.012>.
- Kanehisa, M., Furumichi, M., Tanabe, M., Sato, Y., and Morishima, K. (2017). KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Res.* 45, D353–D361. <https://doi.org/10.1093/nar/gkw1092>.
- Kang, S.Y., Kang, J.Y., and Oh, M.J. (2012). Antiviral activities of flavonoids isolated from the bark of *Rhus verniciflua* Stokes against fish pathogenic viruses *in Vitro*. *J. Microbiol.* 50, 293–300. <https://doi.org/10.1007/s12275-012-2068-7>.

- Kang, S.H., Pandey, R.P., Lee, C.M., Sim, J.S., Jeong, J.T., Choi, B.S., Jung, M., Ginzburg, D., Zhao, K., Won, S.Y., et al. (2020). Genome-enabled discovery of anthraquinone biosynthesis in *Senna tora*. *Nat. Commun.* 11, 5875. <https://doi.org/10.1038/s41467-020-19681-1>.
- Keilwagen, J., Hartung, F., and Grau, J. (2019). GeMoMa: homology-based gene prediction utilizing intron position conservation and RNA-seq data. *Methods Mol. Biol.* 1962, 161–177. [https://doi.org/10.1007/978-1-4939-9173-0\\_9](https://doi.org/10.1007/978-1-4939-9173-0_9).
- Kent, W.J. (2002). BLAT—the BLAST-like alignment tool. *Genome Res.* 12, 656–664. <https://doi.org/10.1101/gr.229202>.
- Kersey, P.J. (2019). Plant genome sequences: past, present, future. *Curr. Opin. Plant Biol.* 48, 1–8. <https://doi.org/10.1016/j.pbi.2018.11.001>.
- Kim, D.H., Kim, M.J., Kim, D.W., Kim, G.Y., Kim, J.K., Gebru, Y.A., Choi, H.S., Kim, Y.H., and Kim, M.K. (2019a). Changes of phytochemical components (urushiol, polyphenols, Gallotannins) and antioxidant capacity during fomitella fraxinea-mediated fermentation of *Toxicodendron vernicifluum* bark. *Molecules* 24, 683. <https://doi.org/10.3390/molecules24040683>.
- Kim, D., Langmead, B., and Salzberg, S.L. (2015a). HISAT: a fast spliced aligner with low memory requirements. *Nat. Methods* 12, 357–360. <https://doi.org/10.1038/nmeth.3317>.
- Kim, J.A., Lee, S., Kim, D.E., Kim, M., Kwon, B.M., and Han, D.C. (2015b). Fisetin, a dietary flavonoid, induces apoptosis of cancer cells by inhibiting HSF1 activity through blocking its binding to the hsp70 promoter. *Carcinogenesis* 36, 696–706. <https://doi.org/10.1093/carcin/bgv045>.
- Kim, K.H., Moon, E., Choi, S.U., Pang, C., Kim, S.Y., and Lee, K.R. (2015c). Identification of cytotoxic and anti-inflammatory constituents from the bark of *Toxicodendron vernicifluum* (Stokes) F.A. Barkley. *J. Ethnopharmacol.* 162, 231–237. <https://doi.org/10.1016/j.jep.2014.12.071>.
- Kim, M.J., Kim, C.J., and Kwak, S.S. (1997). Antifungal activity of urushiol components in the sap of Korean lacquer tree (*Rhus vernicifera* Stokes). *Korean J. Polar Res.* 10, 231–234.
- Kim, M.O., Yang, J., Soo Kwon, Y., and Jo Kim, M. (2015d). Antioxidant and anticancer effects of fermented *Rhus verniciflua* stem bark extracts in HCT-116 cells. *Sci. Asia* 41, 322. <https://doi.org/10.2306/scienceasia1513-1874.2015.41.322>.
- Kim, M.S., Lee, C.W., Kim, J.H., Lee, J.C., and An, W.G. (2019b). Extract of *Rhus verniciflua* Stokes induces p53-mediated apoptosis in MCF-7 breast cancer cells. *Evid. Based Complement Alternat. Med.* 2019, 1–9. <https://doi.org/10.1155/2019/9407340>.
- Korf, I. (2004). Gene finding in novel genomes. *BMC Bioinf.* 5, 59. <https://doi.org/10.1186/1471-2105-5-59>.
- Kuzu, N., Bahcecioglu, I.H., Dagli, A.F., Ozercan, I.H., Ustündag, B., and Sahin, K. (2008). Epigallocatechin gallate attenuates experimental non-alcoholic steatohepatitis induced by high fat diet. *J. Gastroenterol. Hepatol.* 23, e465–e470. <https://doi.org/10.1111/j.1440-1746.2007.05052.x>.
- Lauvergeat, V., Lacomme, C., Lacombe, E., Lasserre, E., Roby, D., and GrimaPettenati, J. (2001). Two cinnamoyl-CoA reductase (CCR) genes from *Arabidopsis thaliana* are differentially expressed during development and in response to infection with pathogenic bacteria. *Phytochemistry* 57, 1187–1195. [https://doi.org/10.1016/S0031-9422\(01\)00053-X](https://doi.org/10.1016/S0031-9422(01)00053-X).
- Lee, D.S., Jeong, G.S., Li, B., Park, H., and Kim, Y.C. (2010). Anti-inflammatory effects of sulfuretin from *Rhus verniciflua* Stokes via the induction of heme oxygenase-1 expression in murine macrophages. *Int. Immunopharm.* 10, 850–858. <https://doi.org/10.1016/j.intimp.2010.04.019>.
- Lee, E.J., Lee, G., Sohn, S.H., and Bae, H. (2017). Extract of *Rhus verniciflua* Stokes enhances Th1 response and NK cell activity. *Mol. Cell. Toxicol.* 12, 399–407. <https://doi.org/10.1007/s13273-016-0044-8>.
- Lee, J.C., Lee, K.Y., Kim, J., Na, C.S., Jung, N.C., Chung, G.H., and Jang, Y.S. (2004). Extract from *Rhus verniciflua* Stokes is capable of inhibiting the growth of human lymphoma cells. *Food Chem. Toxicol.* 42, 1383–1388. <https://doi.org/10.1016/j.fct.2004.03.012>.
- Lee, J.H., Kim, M., Chang, K.H., Hong, C.Y., Na, C.S., Dong, M.S., Lee, D., and Lee, M.Y. (2015). Antiplatelet effects of *Rhus verniciflua* Stokes heartwood and its active constituents-fisetin, butein, and sulfuretin-in rats. *J. Med. Food* 18, 21–30. <https://doi.org/10.1089/jmf.2013.3116>.
- Lee, Y.R., Hwang, J.K., Koh, H.W., Jang, K.Y., Lee, J.H., Park, J.W., and Park, B.H. (2012). Sulfuretin, a major flavonoid isolated from *Rhus verniciflua*, ameliorates experimental arthritis in mice. *Life Sci.* 90, 799–807. <https://doi.org/10.1016/j.lfs.2012.04.015>.
- Leple, J.C., Dauwe, R., Morreel, K., Storme, V., Lapierre, C., Pollet, B., Naumann, A., Kang, K.Y., Kim, H., Ruel, K., et al. (2007). Downregulation of cinnamoyl-coenzyme A reductase in poplar: multiple-level phenotyping reveals effects on cell wall polymer metabolism and structure. *Plant Cell* 19, 3669–3691. <https://doi.org/10.1105/tpc.107.054148>.
- Letunic, I., Doerks, T., and Bork, P. (2012). Smart 7: recent updates to the protein domain annotation resource. *Nucleic Acids Res.* 40, D302–D305. <https://doi.org/10.1093/nar/gkr931>.
- Li, B., and Dewey, C.N. (2011). RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinf.* 12, 323. <https://doi.org/10.1186/1471-2105-12-323>.
- Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. Preprint at arXiv. <https://doi.org/10.48550/arXiv.13033997>.
- Li, H. (2018). Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* 34, 3094–3100. <https://doi.org/10.1093/bioinformatics/bty191>.
- Li, H.T., Yi, T.S., Gao, L.M., Ma, P.F., Zhang, T., Yang, J.B., Gitzendanner, M.A., Fritsch, P.W., Cai, J., Luo, Y., et al. (2019). Origin of angiosperms and the puzzle of the *Jurassic* gap. *Nat. Plants* 5, 461–470. <https://doi.org/10.1038/s41477-019-0421-0>.
- Li, J., Wang, Y., Dong, Y., Zhang, W., Wang, D., Bai, H., Li, K., Li, H., and Shi, L. (2021). The chromosome-based lavender genome provides new insights into Lamiaceae evolution and terpenoid biosynthesis. *Hortic. Res.* 8, 53. <https://doi.org/10.1038/s41438-021-00490-6>.
- Li, X.Y., Wu, X.Y., Zhao, Y., Wen, Q.Y., Xie, Z.B., Yuan, Y.H., Tong, T., Shen, X., Shena, X.Y., and Tong, H. (2016). Composition/structure and lacquering craft analysis of Wenzhou Song dynasty lacquerware. *Anal. Methods* 8, 6529–6536. <https://doi.org/10.1039/C6AY01694D>.
- Liu, H., Wu, S., Li, A., et al. (2021). SMARTdenovo: a de novo assembler using long noisy reads[J]. *Gigabyte* 2021, 1–9.
- Llanchezhan, R., Joseph, C.R., Fau-Rabinarayan, A., and Rabinarayan, A. (2012). Urushiol-induced contact dermatitis caused during Shodhana (Purificatory measures) of *Bhallataka* (*Semecarpus anacardium* Linn.) fruit. *Ayu* 33, 270–273. <https://doi.org/10.4103/0974-8520.105250>.
- Lorini, A., Aranha, B.C., Antunes, B.d.F., Otero, D.M., Jacques, A.C., and Zambiasi, R.C. (2021). Metabolic profile of olive leaves of different cultivars and collection times. *Food Chem.* 345, 128758. <https://doi.org/10.1016/j.foodchem.2020.128758>.
- Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 15, 550. <https://doi.org/10.1186/s13059-014-0550-8>.
- Lowe, T.M., and Eddy, S.R. (1997). tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res.* 25, 955–964. <https://doi.org/10.1093/nar/25.5.955>.
- Lu, R., Ebata, N., Zhang, F.I., and Miyakoshi, T. (2014). Development of a new type lacquer based on *Rhus vernicifera* sap with chitosan. *Prog. Org. Coat.* 77, 439–443. <https://doi.org/10.1016/j.porgcoat.2013.11.006>.
- Ma, D., Dong, S., Zhang, S., Wei, X., Xie, Q., Ding, Q., Xia, R., and Zhang, X. (2021). Chromosome-level reference genome assembly provides insights into aroma biosynthesis in passion fruit (*Passiflora edulis*). *Mol. Ecol. Resour.* 21, 955–968. <https://doi.org/10.1111/1755-0998.13310>.
- Ma, X., Shi, Y.L., Khanjian, H., Schilling, M., Li, M., Fang, H., Cui, D.Y., Kakoulli, I., and Kakoulli, L. (2015). Characterization of early imperial lacquerware from the luozhuang Han tomb, China. *Archaeometry* 59, 121–132. <https://doi.org/10.1111/arc.12226>.
- Maeda, H.A. (2016). Lignin biosynthesis: Tyrosine shortcut in grasses. *Nat. Plants* 2, 16080. <https://doi.org/10.1038/nplants.2016.80>.
- Marcais, G., and Kingsford, C. (2012). Jellyfish: a fast k-mer counter. *Tutorialis e Manuais* 1, 1–8.
- Marchiosi, R., dos Santos, W.D., Constantin, R.P., de Lima, R.B., Soares, A.R., Finger-Teixeira, A., Mota, T.R., de Oliveira, D.M., Foletto-Felipe, M.d.P., Abrahão, J., and Ferrarese-Filho, O. (2020). Biosynthesis and metabolic actions of

- simple phenolic acids in plants. *Phytochem. Rev.* 19, 865–906. <https://doi.org/10.1007/s11101-020-09689-2>.
- Aziz, M., Sturtevant, D., Winston, J., Collakova, E., Jelesko, J., and Chapman, K. (2017). MALDI-MS imaging of urushiols in poison ivy stem. *Molecules* 22, 711. <https://doi.org/10.3390/molecules22050711>.
- Nie, Z.L., Sun, H., Meng, Y., and Wen, J. (2009). Phylogenetic analysis of *Toxicodendron* (Anacardiaceae) and its biogeographic implications on the evolution of north temperate and tropical intercontinental disjunctions. *J. Syst. Evol.* 47, 416–430. <https://doi.org/10.1111/j.1759-6831.2009.00045.x>.
- Patil, C.S., Singh, V.P., Satyanarayan, P.S., Jain, N.K., Singh, A., and Kulkarni, S.K. (2003). Protective effect of flavonoids against aging- and lipopolysaccharide-induced cognitive impairment in mice. *Pharmacology* 69, 59–67. <https://doi.org/10.1159/000072357>.
- Patel, S., Robben, M., Fennell, A., Londo, J.P., Alahakoon, D., Villegas-Diaz, R., and Swaminathan, P. (2020). Draft genome of the native American cold hardy grapevine *Vitis riparia* Michx. ‘Manitoba 37’. *Hortic. Res.* 7, 92. <https://doi.org/10.1038/s41438-020-0316-2>.
- Pertea, M., Pertea, G.M., Antonescu, C.M., Chang, T.-C., Mendell, J.T., and Salzberg, S.L. (2015). StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat. Biotechnol.* 33, 290–295. <https://doi.org/10.1038/nbt.3122>.
- Piquemal, J., Lapiere, C., Myton, K., O’Connell, A., Schuch, W., GrimaPettenati, J., and Boudet, A.M. (2002). Downregulation of cinnamoyl CoA reductase induces significant changes of lignin profiles in transgenic tobacco plants. *Plant J.* 13, 71–83. <https://doi.org/10.1046/j.1365-313X.1998.00014.x>.
- Price, A.L., Jones, N.C., and Pevzner, P.A. (2005). De novo identification of repeat families in large genomes. *Bioinformatics* 21, i351–i358. <https://doi.org/10.1093/bioinformatics/bt11018>.
- Rozas-Muñoz, E., Lepoittevin, J.P., Pujol, R.M., and Giménez-Arnau, A. (2012). Allergic contact dermatitis to plants: understanding the chemistry will help our diagnostic approach. *Actas Dermo Sifiliogr.* 103, 456–477. <https://doi.org/10.1016/j.adengl.2012.07.006>.
- Sabarwal, A., Agarwal, R., and Singh, R.P. (2017). Fisetin inhibits cellular proliferation and induces mitochondria-dependent apoptosis in human gastric cancer cells. *Mol. Carcinog.* 56, 499–514. <https://doi.org/10.1002/mc.22512>.
- Sarrou, E., Ganopoulos, I., Xanthopoulou, A., Masuero, D., Martens, S., Madesis, P., Mavromatis, A., and Chatzopoulou, P. (2017). Genetic diversity and metabolic profile of *Salvia officinalis* populations: implications for advanced breeding strategies. *Planta* 246, 201–215. <https://doi.org/10.1007/s00425-017-2666-z>.
- Sawada, Y., Sato, M., Okamoto, M., Masuda, J., Yamaki, S., Tamari, M., Tanokashira, Y., Kishimoto, S., Ohmiya, A., Abe, T., and Hirai, M.Y. (2019). Metabolome-based discrimination of chrysanthemum cultivars for the efficient generation of flower color variations in mutation breeding. *Metabolomics* 15, 118. <https://doi.org/10.1007/s11306-019-1573-7>.
- Servant, F., Bru, C., Carrère, S., Courcelle, E., Gouzy, J., Peyruc, D., and Kahn, D. (2002). ProDom: automated clustering of homologous domains. *Brief. Bioinform.* 3, 246–251. <https://doi.org/10.1093/bib/3.3.246>.
- Shen, H., Mazarei, M., Hisano, H., Escamilla-Trevino, L., Fu, C., Pu, Y., Rudis, M.R., Tang, Y., Xiao, X., Jackson, L., et al. (2013). A genomics approach to deciphering lignin biosynthesis in switchgrass. *Plant Cell* 25, 4342–4361. <https://doi.org/10.1105/tpc.113.118828>.
- Shu, X.S., Lv, J.H., Tao, J., Li, G.M., Li, H.D., and Ma, N. (2009). Antihyperglycemic effects of total flavonoids from polygonatum odoratum in STZ and alloxan-induced diabetic rats. *J. Ethnopharmacol.* 124, 539–543. <https://doi.org/10.1016/j.jep.2009.05.006>.
- Simão, F.A., Waterhouse, R.M., Ioannidis, P., Kriventseva, E.V., and Zdobnov, E.M. (2015). BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* 31, 3210–3212. <https://doi.org/10.1093/bioinformatics/btv351>.
- Sigrist, C.J.A., Cerutti, L., de Castro, E., Langendijk-Genevaux, P.S., Bulliard, V., Bairoch, A., and Hulo, N. (2010). PROSITE, a protein domain database for functional characterization and annotation. *Nucleic Acids Res.* 38, D161–D166. <https://doi.org/10.1093/nar/gkp885>.
- Stamatakis, A. (2006). RAXML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* 22, 2688–2690. <https://doi.org/10.1093/bioinformatics/btl446>.
- Stanke, M., Keller, O., Gunduz, I., Hayes, A., Waack, S., and Morgenstern, B. (2006). AUGUSTUS: ab initio prediction of alternative transcripts. *Nucleic Acids Res.* 34, W435–W439. <https://doi.org/10.1093/nar/gkl200>.
- Sun, P., Jiao, B., Yang, Y., Shan, L., Li, T., Li, X., Xi, Z., Wang, X., and Liu, J. (2021). WGDl: a user-friendly toolkit for evolutionary analyses of whole-genome duplications and ancestral karyotypes. Preprint at bioRxiv. <https://doi.org/10.1101/2021.04.29.441969>.
- Sun, X., Zhu, S., Li, N., Cheng, Y., Zhao, J., Qiao, X., Lu, L., Liu, S., Wang, Y., Liu, C., et al. (2020). A chromosome-level genome assembly of garlic (*Allium sativum*) provides insights into genome evolution and allicin biosynthesis. *Mol. Plant* 13, 1328–1339. <https://doi.org/10.1016/j.molp.2020.07.019>.
- Tarailo-Graovac, M., and Chen, N. (2009). Using RepeatMasker to identify repetitive elements in genomic sequences. *Curr. Protocols Bioinf.* 25, 4. <https://doi.org/10.1002/0471250953.bi0410s25>.
- Tang, C., Yang, M., Fang, Y., Luo, Y., Gao, S., Xiao, X., An, Z., Zhou, B., Zhang, B., Tan, X., et al. (2016). The rubber tree genome reveals new insights into rubber production and species adaptation. *Nat. Plants* 2, 16073. <https://doi.org/10.1038/nplants.2016.73>.
- Thomas, P.D., Campbell, M.J., Kejarawal, A., Mi, H., Karlak, B., Daverman, R., Diemer, K., Muruganujan, A., and Narechania, A. (2003). PANTHER: a library of protein families and subfamilies indexed by function. *Genome Res.* 13, 2129–2141. <https://doi.org/10.1101/gr.772403>.
- Tohmé, M., Giménez, M., Peralta, A., Colombo, M.I., and Delgui, L.R. (2019). Ursolic acid: a novel antiviral compound inhibiting rotavirus infection in vitro. *Int. J. Antimicrob. Agents* 54, 601–609. <https://doi.org/10.1016/j.ijantimicag.2019.07.015>.
- Vaser, R., Sović, I., Nagarajan, N., and Šikić, M. (2017). Fast and accurate de novo genome assembly from long uncorrected reads. *Genome Res.* 27, 737–746. <https://doi.org/10.1101/gr.214270.116>.
- Vurture, G.W., Sedlazeck, F.J., Nattestad, M., Underwood, C.J., Fang, H., Gurtowski, J., and Schatz, M.C. (2017). GenomeScope: fast reference-free genome profiling from short reads. *Bioinformatics* 33, 2202–2204. <https://doi.org/10.1093/bioinformatics/btx153>.
- Walker, B.J., Abeel, T., Shea, T., Priest, M., Abouelliel, A., Sakthikumar, S., Cuomo, C.A., Zeng, Q., Wortman, J., Young, S.K., and Earl, A.M. (2014). Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One* 9, e112963. <https://doi.org/10.1371/journal.pone.0112963>.
- Wang, L., He, N., Li, Y., Fang, Y., and Zhang, F. (2020a). Complete chloroplast genome sequence of Chinese lacquer tree (*Toxicodendron vernicifluum*, Anacardiaceae) and its phylogenetic significance. *BioMed Res. Int.* 2020, 9014873. <https://doi.org/10.1155/2020/9014873>.
- Wang, P., Luo, Y., Huang, J., Gao, S., Zhu, G., Dang, Z., Gai, J., Yang, M., Zhu, M., Zhang, H., et al. (2020b). The genome evolution and domestication of tropical fruit mango. *Genome Biol.* 21, 60. <https://doi.org/10.1186/s13059-020-01959-8>.
- Wang, Y., Tang, H., DeBarry, J.D., Tan, X., Li, J., Wang, X., Lee, T.-h., Jin, H., Marler, B., Guo, H., et al. (2012). MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Res.* 40, e49. <https://doi.org/10.1093/nar/gkr1293>.
- Wang, Y., Xin, H., Fan, P., Zhang, J., Liu, Y., Dong, Y., Wang, Z., Yang, Y., Zhang, Q., Ming, R., et al. (2021). The genome of Shanputao (*Vitis amurensis*) provides a new insight into cold tolerance of grapevine. *Plant J.* 105, 1495–1506. <https://doi.org/10.1111/tpj.15127>.
- Watanabe, H., Fujimoto, A., Nishida, J., Ohishi, T., and Takahara, A. (2016). Biobased polymer coating using catechol derivative urushiol. *Langmuir* 32, 4619–4623. <https://doi.org/10.1021/acs.langmuir.6b00484>.
- Weisberg, A.J. (2014). *Investigations into the Molecular Evolution of Plant Terpene, Alkaloid, and Urushiol Biosynthetic Enzymes* (Virginia Polytechnic Institute and State University).
- Wolfe, K.H. (2001). Yesterday’s polyploids and the mystery of diploidization. *Nat. Rev. Genet.* 2, 333–341. <https://doi.org/10.1038/35072009>.
- Wu, T.D., and Watanabe, C.K. (2005). GMAP: a genomic mapping and alignment program for

- mRNA and EST sequences. *Bioinformatics* 21, 1859–1875. <https://doi.org/10.1093/bioinformatics/bti310>.
- Xie, Y., Zhang, J., Liu, W., Xie, N., Feng, F., and Qu, W. (2016). New urushiols with platelet aggregation inhibitory activities from resin of *Toxicodendron vernicifluum*. *Fitoterapia* 112, 38–44. <https://doi.org/10.1016/j.fitote.2016.05.001>.
- Xu, H., Song, J., Luo, H., Zhang, Y., Li, Q., Zhu, Y., Xu, J., Li, Y., Song, C., Wang, B., et al. (2016). Analysis of the genome sequence of the medicinal plant *Salvia miltiorrhiza*. *Mol. Plant* 9, 949–952. <https://doi.org/10.1016/j.molp.2016.03.010>.
- Xu, Y., Chen, Q., Bai, W., and Lin, J. (2012). Preparation and properties of raw lacquer/multihydroxyl polyacrylate/organophilic montmorillonite nanocomposites. *Polym. Bull.* 68, 983–992. <https://doi.org/10.1007/s00289-011-0593-x>.
- Xu, Z., Zhang, D., Hu, J., Zhou, X., Ye, X., Reichel, K.L., Stewart, N.R., Syrenne, R.D., Yang, X., Gao, P., et al. (2009). Comparative genome analysis of lignin biosynthesis gene families across the plant kingdom. *BMC Bioinform.* 10, S3. <https://doi.org/10.1186/1471-2105-10-S11-S3>.
- Xu, Z., and Wang, H. (2007). LTR\_FINDER: an efficient tool for the prediction of full-length LTR retrotransposons. *Nucleic Acids Res.* 35, W265–W268. <https://doi.org/10.1093/nar/gkm286>.
- Xue, H., Zhang, Y., Li, X., Xia, J., and Lin, Q. (2020). Degradation urushiol coating composition crosslinked by diallyl trisulfide. *J. Polym. Environ.* 28, 3192–3200. <https://doi.org/10.1007/s10924-020-01845-1>.
- Yang, J., Deng, J., Zhu, J., Shen, Q., and Li, D. (2015). Lacquer sap with reactive maleic hemiester surfactant-modified phase interface and its properties. *Prog. Org. Coating* 87, 138–145. <https://doi.org/10.1016/j.porgcoat.2015.05.027>.
- Yang, J., Zhu, J., Shen, F., Cai, J., and Zhou, M. (2018). Promotion by copper (II)-modified montmorillonite of the drying property of oriental lacquer sap. *Prog. Org. Coat.* 118, 72–81. <https://doi.org/10.1016/j.porgcoat.2018.01.023>.
- Yang, Z. (2007). Paml 4: phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* 24, 1586–1591. <https://doi.org/10.1093/molbev/msm088>.
- Yin, Y., Peng, F., Zhou, L., Yin, X., Chen, J., Zhong, H., Hou, F., Xie, X., Wang, L., Shi, X., et al. (2021). The chromosome-scale genome of *Magnolia officinalis* provides insight into the evolutionary position of magnoliids. *iScience* 24, 102997. <https://doi.org/10.1016/j.isci.2021.102997>.
- Yu, X., Wang, P., Li, J., Zhao, Q., Ji, C., Zhu, Z., Zhai, Y., Qin, X., Zhou, J., Yu, H., et al. (2021). Whole-genome sequence of synthesized allopolyploids in cucumis reveals insights into the genome evolution of allopolyploidization. *Adv. Sci.* 8, 2004222. <https://doi.org/10.1002/adv.202004222>.
- Yu, X.J., Zheng, H.K., Wang, J., Wang, W., and Su, B. (2006). Detecting lineage-specific adaptive evolution of brain-expressed genes in human using rhesus macaque as outgroup. *Genomics* 88, 745–751. <https://doi.org/10.1016/j.ygeno.2006.05.008>.
- Zdobnov, E.M., and Apweiler, R. (2001). InterProScan—an integration platform for the signature-recognition methods in InterPro. *Bioinformatics* 17, 847–848. <https://doi.org/10.1093/bioinformatics/17.9.847>.
- Zeng, L., Tu, X.L., Dai, H., Han, F.M., Lu, B.S., Wang, M.S., Nanaei, H.A., Tajabadipour, A., Mansouri, M., Li, X.L., et al. (2019). Whole genomes and transcriptomes reveal adaptation and domestication of pistachio. *Genome Biol.* 20, 79. <https://doi.org/10.1186/s13059-019-1686-3>.
- Zhang, L., Chen, F., Zhang, X., Li, Z., Zhao, Y., Lohaus, R., Chang, X., Dong, W., Ho, S.Y.W., Liu, X., et al. (2020). The water lily genome and the early evolution of flowering plants. *Nature* 577, 79–84. <https://doi.org/10.1038/s41586-019-1852-5>.
- Zhang, R., Lee, I.K., Piao, M.J., Kim, K.C., Kim, A.D., Kim, H.S., Chae, S., Kim, H.S., and Hyun, J.W. (2011). Butin (7, 3', 4'-trihydroxydihydroflavone) reduces oxidative stress-induced cell death via inhibition of the mitochondria-dependent apoptotic pathway. *Int. J. Mol. Sci.* 12, 3871–3887. <https://doi.org/10.3390/ijms12063871>.
- Zhao, Q., Yang, J., Cui, M.Y., Liu, J., Fang, Y., Yan, M., Qiu, W., Shang, H., Xu, Z., Yidiresi, R., et al. (2019). The reference genome sequence of *Scutellaria baicalensis* provides insights into the evolution of wogonin biosynthesis. *Mol. Plant* 12, 935–950. <https://doi.org/10.1016/j.molp.2019.04.002>.
- Zhao, W., Zhu, L., Srinivasan, S., Damodaran, C., and Rohr, J. (2009). Identification of urushiols as the major active principle of the siddha herbal medicine semecarpus lehyam: anti-tumor agents for the treatment of breast cancer. *Pharm. Biol.* 47, 886–893. <https://doi.org/10.1080/13880200902942410>.
- Zhu, Q.G., Xu, Y., Yang, Y., Guan, C.F., Zhang, Q.Y., Huang, J.W., Grierson, D., Chen, K.S., Gong, B.C., and Yin, X.R. (2019). The persimmon (*Diospyros oleifera* Cheng) genome provides new insights into the inheritance of astringency and ancestral evolution. *Hortic. Res.* 6, 138. <https://doi.org/10.1038/s41438-019-0227-2>.

## STAR★METHODS

### KEY RESOURCES TABLE

| REAGENT or RESOURCE  | SOURCE                         | IDENTIFIER  |
|--|--------------------------------|---|
| <b>Biological samples</b>  |                                |   |
| The total DNA and RNA of wild <i>T. vernicifluum</i> and the varieties | This study                     | NA  |
| <b>Critical commercial assays</b>                                      |                                |   |
| TRIzol™  | Invitrogen                     | 10296010  |
| DNA isolation  | CTAB                           | -   |
| NEBNext Ultra II End Repair/A-Tailing Module                           | NEB                            | E7546L  |
| <b>Deposited data</b>  |                                |   |
| Assembly genome  | This paper                     | GenBank: PRJNA836312  |
| <b>Software and algorithms</b>   |                                |   |
| SOAPnuke v.1.5.6   | Chen et al., 2018              | <a href="https://github.com/BGI-flexlab/SOAPnuke">https://github.com/BGI-flexlab/SOAPnuke</a>   |
| NanoFilt   | De Coster et al., 2018         | <a href="https://github.com/wdecoster/nanofilt">https://github.com/wdecoster/nanofilt</a>   |
| GenomeScope v1.0   | Vurture et al., 2017           | <a href="https://github.com/schatzlab/genomescope">https://github.com/schatzlab/genomescope</a>   |
| Jellyfish v2.0   | Marcais and Kingsford, 2012    | <a href="https://github.com/gmarcais/Jellyfish">https://github.com/gmarcais/Jellyfish</a>   |
| SMARTdenovo  | Liu et al., 2021               | <a href="https://github.com/ruanjue/smartdenovo">https://github.com/ruanjue/smartdenovo</a>   |
| Racon v1.3.3   | Vaser et al., 2017             | <a href="https://github.com/isovic/racon">https://github.com/isovic/racon</a>   |
| Pilon v.1.225  | Walker et al., 2014            | <a href="https://github.com/broadinstitute/pilon">https://github.com/broadinstitute/pilon</a>   |
| Juicer v.1.5   | Durand et al., 2016            | <a href="https://github.com/aidenlab/juicer">https://github.com/aidenlab/juicer</a>   |
| 3D-DNA, v.180922   | Dudchenko et al., 2017         | <a href="https://github.com/aidenlab/3d-dna">https://github.com/aidenlab/3d-dna</a>   |
| GMAP   | Wu and Watanabe, 2005          | <a href="https://github.com/juliangehring/GMAP-GSNAP">https://github.com/juliangehring/GMAP-GSNAP</a>   |
| HISAT v2.1.1   | Kim et al., 2015a              | <a href="http://daehwankimlab.github.io/hisat2/">http://daehwankimlab.github.io/hisat2/</a>   |
| StringTie v1.3.3b  | Pertea et al., 2015            | <a href="https://github.com/gpertea/stringtie">https://github.com/gpertea/stringtie</a>   |
| RSEM v1.2.12   | Li and Dewey, 2011             | <a href="https://github.com/deweylab/RSEM">https://github.com/deweylab/RSEM</a>   |
| DESeq2   | Love et al., 2014              | <a href="https://bioconductor.org/packages/release/bioc/html/DESeq2.html">https://bioconductor.org/packages/release/bioc/html/DESeq2.html</a> |
| RepeatMasker v.4.0.7 and RepeatProteinMask v.4.0.7                     | Tarailo-Graovac and Chen, 2009 | <a href="http://www.repeatmasker.org/cgi-bin/RepeatProteinMaskRequest">http://www.repeatmasker.org/cgi-bin/RepeatProteinMaskRequest</a>       |
| Repbase database v21.12  | Bao et al., 2015               | <a href="http://www.girinst.org/repbase">http://www.girinst.org/repbase</a>   |
| LTR-FINDER   | Xu and Wang, 2007              | <a href="http://tlife.fudan.edu.cn/ltr_finder/">http://tlife.fudan.edu.cn/ltr_finder/</a>   |
| Piler  | Edgar and Myers, 2005          | <a href="http://www.drive5.com/piler">http://www.drive5.com/piler</a>   |
| RepeatScout  | Price et al., 2005             | <a href="http://bix.ucsd.edu/repeatscout/">http://bix.ucsd.edu/repeatscout/</a>   |
| TRF v.4.09   | Benson, 1999                   | <a href="https://tandem.bu.edu/trf/trf.html">https://tandem.bu.edu/trf/trf.html</a>   |
| Augustus   | Stanke et al., 2006            | <a href="https://github.com/Gaius-Augustus/Augustus">https://github.com/Gaius-Augustus/Augustus</a>   |
| SNAP   | Korf, 2004                     | <a href="https://github.com/KorfLab/SNAP">https://github.com/KorfLab/SNAP</a>   |
| MAKER pipeline   | Cantarel et al., 2008          | <a href="https://www.yandell-lab.org/software/maker.html">https://www.yandell-lab.org/software/maker.html</a>                                 |
| GeMoMa v1.6  | Keilwagen et al., 2019         | <a href="http://www.jstacs.de/index.php/GeMoMa">http://www.jstacs.de/index.php/GeMoMa</a>   |
| tRNAscan-SE v1.3.1   | Lowe and Eddy, 1997            | <a href="http://lowelab.ucsc.edu/tRNAscan-SE">http://lowelab.ucsc.edu/tRNAscan-SE</a>   |
| Rfam v12.0 database  | Griffiths-Jones et al., 2005   | <a href="https://rfam.xfam.org/">https://rfam.xfam.org/</a>   |
| BUSCO v4.0.6   | Simão et al., 2015             | <a href="https://busco.ezlab.org/">https://busco.ezlab.org/</a>   |
| BLASTp   | Kent, 2002                     | <a href="https://blast.ncbi.nlm.nih.gov/Blast.cgi">https://blast.ncbi.nlm.nih.gov/Blast.cgi</a>   |

(Continued on next page)

**Continued**

| REAGENT or RESOURCE | SOURCE                       | IDENTIFIER  |
|---------------------|------------------------------|---|
| OrthoFinder         | Emms and Kelly, 2015         | <a href="https://github.com/davidemms/OrthoFinder">https://github.com/davidemms/OrthoFinder</a>               |
| RaxML               | Stamatakis, 2006             | <a href="https://github.com/stamatak/standard-RAxML">https://github.com/stamatak/standard-RAxML</a>           |
| MCMCtree            | Yang, 2007                   | <a href="http://abacus.gene.ucl.ac.uk/software/paml.html">http://abacus.gene.ucl.ac.uk/software/paml.html</a> |
| MCSanX v1.5.2       | Wang et al., 2012            | <a href="https://github.com/wyp1125/MCSanX">https://github.com/wyp1125/MCSanX</a>                             |
| WGDI                | Sun et al., 2021             | <a href="https://github.com/SunPengChuan/wgdi">https://github.com/SunPengChuan/wgdi</a>                       |
| GeneWise v2.4       | Birney et al., 2004          | <a href="https://bio.tools/Wise">https://bio.tools/Wise</a>   |
| Muscle v3.8.31      | Edgar, 2004                  | <a href="http://www.drive5.com/muscle">http://www.drive5.com/muscle</a>                                       |
| <b>Other</b>        |                              |   |
| PacBio sequencing   | Pacific Biosciences          | Sequel II platform  |
| Nanopore sequencing | Oxford Nanopore Technologies | Oxford Nanopore platform  |
| DNBSEQ sequencing   | MGI Tech Co., Ltd.           | MGISEQ-2000 platform  |

**RESOURCE AVAILABILITY****Lead contact**

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Sifeng Li ([lif60@sina.com](mailto:lif60@sina.com)).

**Materials availability**

This study did not generate new unique reagents.

**Data and code availability**

- The accession number for the genome assembly and raw reads reported in this paper is GenBank: PRJNA836312.
- This paper does not report original code.
- Any additional information required to reanalyze the data reported in this paper is available from the [Lead contact](#) upon request.

**EXPERIMENTAL MODEL AND SUBJECT DETAILS**

This study does not include experiments model or subjects.

**METHOD DETAILS****Genome sequencing and assembly***Plant materials and genome sequencing*

We extracted high-quality genomic DNA from fresh leaves of a diploid female *T. vernicifluum* 'Gaobachi' grown in Xi'an Botanical Garden, Xi'an, Shanxi Province (E109°1'43.65", N34°12'31.32"). The libraries with short insert sizes of 350 bp and long insert sizes of about 20 kb were constructed and sequenced according to the standard protocols on the MGISEQ-2000 platform and Oxford Nanopore platform, respectively. DNBSEQ clean data was filtered by SOAPnuke v.1.5.6 (Chen et al., 2018), and we were trimmed and filtered ONT reads by NanoFilt (De Coster et al., 2018), with base quality score < 7 and reads which shorter than 5 kb.

For building a Hi-C library, genomic DNAs were fixed with formaldehyde and were sheared by a restriction enzyme (Mbol). The library was then sequenced by MGISEQ-2000 platform and filtering by SOAPnuke v.1.5.6. A total of 488,159,058 read pairs with 73.22 Gb of 150PE Hi-C data were generated to acquire a chromosome-level assembly of the genome.

### De novo assembly

Genome size, repeat fraction and heterozygosity of *T. vernicifluum* were estimated by using GenomeScope v1.0 (Vurture et al., 2017) based on the k-mer histograms provided by Jellyfish v2.0 (Marcais and King-sford, 2012) with 17-mer. After corrected and trimmed by Canu v1.7, the ONT long reads were *de novo* assembled by using SMARTdenovo (<https://github.com/ruanjue/smartdenovo>) with the parameters '-k 26 -z 10 -Z 16 -U -1 -m 0.6 -A 1000'. Next, three rounds of minimap2 (Li, 2018)-Racon v1.3.3 (Vaser et al., 2017) and one round of MEDAKA (<https://github.com/nanoporetech/medaka>) were applied to polish the assembly with uncorrected ONT sequence data, and finally Pilon v.1.225 (Walker et al., 2014) was employed to calibrate the consensus contigs using DNBSEQ short reads.

Then, Hi-C reads were mapped to draft genome assembly and processed to generate interaction matrices by Juicer v.1.5 (Durand et al., 2016). 3D *de novo* assembly (3D-DNA, v.180922) pipeline (Dudchenko et al., 2017) was performed to anchor primary contigs into chromosome-level assembly by using the output of Juicer pipeline.

### Transcriptome sequencing and gene expression analysis

RNAs from leaves, roots, stems, flowers and drupes from *T. vernicifluum* 'Gaobachi' were extracted for full-length transcriptome sequencing using the PacBio Sequel II platform. PacBio ISO-Seq3 pipeline (<https://github.com/PacificBiosciences/IsoSeq>) was used to obtain full-length non-chimeric (FLNC) transcripts via ccs, classify, cluster and polish stages. The high-quality transcripts were mapped against the assembled genome using GMAP (Wu and Watanabe, 2005), and collapsed to reduce redundancy using Cupcake ToFU ([https://github.com/Magdoll/cDNA\\_Cupcake/wiki/Cupcake-ToFU](https://github.com/Magdoll/cDNA_Cupcake/wiki/Cupcake-ToFU)) scripts.

Five varieties of *T. vernicifluum* were sampled for RNA-seq analysis, include 'Gaobachi' (GBC), 'Dahong-pao' (DHP), 'Huangmaoguizhou' (HuM), 'Hongmaoguizhou' (HoM) and wild variety, each has three biological repeats, libraries of insert size 300–400 bp were prepared according to the manufacturer's protocol of DNBSEQ sequencing platform and paired-end sequenced by MGISEQ-2000 platform. The reads filtering by SOAPnuke were mapped to assembled genome using HISAT v2.1.1 (Kim et al., 2015a) and assembled by StringTie v1.3.3b (Pertea et al., 2015). Gene expression levels in terms of fragments per kilobase per million reads mapped (FPKM) values were calculated by RSEM v1.2.12 (Li and Dewey, 2011) and differentially expressed genes (DEGs) were detected by DEseq2 (Love et al., 2014).

### Gene prediction and annotation

Repetitive elements in the *T. vernicifluum* genome were identified through a combination of homolog-based and *de novo* approaches. RepeatMasker v.4.0.7 and RepeatProteinMask v.4.0.7 (Tarailo-Graovac and Chen, 2009) were used to detect repeats by aligning against Repbase database v21.12 (Bao et al., 2015). LTR-FINDER (Xu and Wang, 2007), Piler (Edgar and Myers, 2005) and RepeatScout (Price et al., 2005) were employed to build transposable element (TE) consensus sequences as a *de novo* TE library, and TRF v.4.09 (Benson, 1999) was used to obtain tandem repetitive sequences. RepeatMasker v.4.0.7 was then used to discover and identify repetitive sequences with the combined library of the *de novo* TEs.

MAKER pipeline (Cantarel et al., 2008) was employed twice to annotate protein-coding genes in genome with comprehensive evidence from *ab initio* prediction, homology-based and transcriptome-based prediction. Protein sequences of *Acer yangbiense* (GigaDB), *Anacardium occidentale* (Phytozome v13), *Arabidopsis thaliana* (TAIR 10), *Citrus sinensis* (Phytozome v13), *Dimocarpus longan* (GigaDB) and *Pistacia vera* were used as homology-based evidence. Non-redundant isoforms were used as transcriptome-based evidence. At first round, the genome was annotated to gain genes by Maker pipeline with proteins and transcripts evidence. The above result was used to train the parameters for Augustus (Stanke et al., 2006) and SNAP (Korf, 2004). The second round, we combined the models trained by Augustus and SNAP with protein sequences of five species and isoforms to predict the gene set. Homologous genes from *Pistacia vera* was predicted by GeMoMa v1.6 (Keilwagen et al., 2019) and compared with the gene set generated by MAKER. Then the gene models not predicted by MAKER were integrated into the final gene set.

For the annotation of non-coding RNA, we detected tRNA by tRNAscan-SE v1.3.1 (Lowe and Eddy, 1997), and identified other noncoding RNA, including miRNA, rRNA and snRNA, by searching Rfam v12.0 database (Griffiths-Jones et al., 2005).

Gene functional annotation was performed based on consensus of sequence and domain. The protein sequences were aligned to NCBI Non-Redundant Protein Sequence (NR) databases, Kyoto Encyclopedia of Genes and Genomes (KEGG v89.0; (Kanehisa et al., 2017)), SwissProt and TrEMBL (Uniprot release 2020-06) (Boeckmann et al., 2003) with BLASTp. The domains were searched and predicted by using InterProScan v5.11-55.0 (Zdobnov and Apweiler, 2001) with publicly available databases including PANTHER (Thomas et al., 2003), Pfam (Bateman et al., 2004), PRINTS (Attwood et al., 2000), ProDom (Servant et al., 2002), PROSITE profiles (Sigrist et al., 2010), and SMART (Letunic et al., 2012). Gene ontology (GO) terms (Ashburner et al., 2000) for each gene were predicted from the InterPro descriptions.

### Assembly and gene set evaluation

DNBSEQ clean reads and ONT corrected reads were mapped against the *T. vernicifluum* genome with BWA-MEM (Li, 2013) and minimap2 (Li, 2018) to evaluate the quality of our assembly. BUSCO v4.0.6 (Simão et al., 2015) with parameter '-l embryophyta\_odb10' was applied to evaluate completeness of the assembly and gene set.

### Phylogenetic analyses

We collected *T. vernicifluum* and other fourteen species, including the species used for gene annotation as well as *Carica papaya*, *Oryza sativa*, *Prunus persica*, *Solanum lycopersicum*, *Theobroma cacao*, *Vitis vinifera* downloaded from Phytozome v13, *Sclerocarya birrea* (GigaDB) and *Mangifera indica* (Wang et al., 2020a, 2020b) to identify gene families. After removing alternative splicing sequences, we performed Orthofinder (Emms and Kelly, 2015) to obtain orthologous and paralogous genes of these species. Then RaxML (Stamatakis, 2006) was applied to construct the phylogenetic tree by using the single copy orthologous gene families with the GTRGAMMA model.

MCMCtree program in PAML v4.9j (Yang, 2007) was performed to estimate the divergence time between fifteen species in the phylogenetic tree with the REV substitution model. Four calibration time points based on the TimeTree database (<http://www.timetree.org>) were used as references, including *A. thaliana* and *A. yangbiense* (96- 104 Mya), *A. yangbiense* and *D. longan* (53- 104 Mya), *A. yangbiense* and *C. sinensis* (81- 103 Mya), and *A. occidentale* and *C. sinensis* (62- 84 Mya).

CAFÉ (De Bie et al., 2006) was employed to investigate gene family expansion and contraction under a maximum likelihood framework, single copy orthologous gene families and estimated divergence time between different species were used as input files. Families with p-value < 0.05 were regarded as significant expansion or contraction and were used for enrichment analysis.

### Genome synteny and whole-genome duplication analysis

We applied MCScanX v1.5.2 (Wang et al., 2012) to detect synteny among *T. vernicifluum* and other species in Anacardiaceae, and calculated the 4DTv of syntenic blocks. The WGD (Sun et al., 2021) pipeline also used to estimate synonymous substitution per site (Ks) of these species. Then, the WGD events were evaluated with the 4DTv and Ks distribution.

### Identification of polyketide synthases genes from genome sequences

Reference sequences of Polyketide synthases (PKS) genes were download from NCBI. Sequences of PKS genes were aligned from *A. occidentale*, *A. thaliana*, *M. indica*, *P. vera*, *O. sativa*, *S. birrea* and *T. vernicifluum* by using tBLASTn with an E-value of  $1 \times 10^{-5}$ , and the blasted hits were conjoined by SOLAR v0.9 (Yu et al., 2006). We predicted the structure of PKS genes by GeneWise v2.4 (Birney et al., 2004), and extended the 3' and 5' direction in potential PKS genes that incomplete. PKS genes with interrupting stop codons or frameshifts were classified as pseudogenes, and the genes with a partially intact were classified as partial genes. The PKS domain numbered PF00195 and PF02797 obtained from the Pfam database was used as a standard sequence, and then all candidate sequences were analyzed in the Pfam database by InterProScan. The potential genes from start codon to stop codon and contain N-terminal domain and C-terminal domain were considered as intact PKS genes. we converted coding sequences (CDS) to protein sequences and used Muscle v3.8.31 (Edgar, 2004) to perform multiple sequence alignments. RaxML was employed to construct gene family phylogenetic tree with the PROTGAMMAUTO model.

### Metabolomic profile of lacquer cultivars

We collected the phloem of the wild lacquer (TV) and cultivars HoM, DHP, GBC, and HuM for metabolomic analysis. Weighed 100-mg samples and added 1.2 mL of 70% methyl alcohol extraction solvent to each sample, and vortexed every 30 min for 30 s, and stored the extractions at 4°C overnight. Then, the samples were centrifuged for 10 min at 12,000 rpm and filtrated by 0.22 μM filter membrane. Next, 4 μL of supernatant was injected into an Agilent SB-C18 column (1.8 μm, 2.1 × 100 mm) for ultraperformance liquid chromatography-electrospray ionization-mass spectrometry (UPLC-MS/MS) analysis. For qualitative analysis, we used the MVDB V2.0 database of Wuhan Maiteville Biotechnology Co., Ltd. (Wuhan, China). Base on the stepwise MIM-EPI (multiple ion monitoring-enhanced product ions) to build the commercially available standard Metabolites Database (Metware Biotechnology Co., Ltd. Wuhan, China) and the public metabolite databases such as MassBank, KNAPSACK, HMDB, MoToDB, and METLIN, qualitative analysis of the metabolite data was performed.

### qRT-PCR analysis

Nine genes were involved in urushiol and lignin pathways randomly chosen to validate the RNA-seq data. Total RNA samples were same as the RNA-seq analysis. The First-strand cDNA was reverse transcribed using the PrimeScript™ RT reagent Kit with gDNA Eraser (Takara, Dalian, China). qRT-PCR was performed using SYBR Premix Ex Taq™ (Takara, Dalian, China). All samples were normalized to TvActin to determine the isoform expression level. qRT-PCR primers were designed with PRIMER PREMIER 6 (PREMIER Biosoft, <http://www.premierbiosoft.com/primerdesign/>), and their specificity was verified by PCR (Table S24). The conditions for the qRT-PCR amplification were as follows: incubation at 95°C for 30 sec, 35 cycles of 95°C for 5 sec and 60°C for 30 sec. The specificity of the primer amplicons was tested using a melting curve analysis. The amplification was carried out in triplicate with LightCycler96® instrument (Roche Diagnostics, Switzerland) and the expression levels of the candidate genes were calculated using the  $2^{-\Delta\Delta C_t}$  method.

### QUANTIFICATION AND STATISTICAL ANALYSIS

The quantitative analysis of metabolites used multiple reaction monitoring (Chen et al., 2013). Significantly different metabolites between groups were determined by variable importance in projection (VIP) ≥ 1 and an absolute Log<sub>2</sub>FC (fold change) ≥ 1 (Jiang et al., 2020). Linear discriminate analysis (LDA) effect size analysis of top 10 differential metabolites among groups and LDA scores and significance of  $p \leq 0.05$  as determined by Wilcoxon's signed-rank test.

### ADDITIONAL RESOURCES

This study does not include additional resources.