

Identification key genes involved in lignin and flavonoid accumulation during Tilia tuan seed maturation

Lei Liu

Beijing Forestry University

Cui Long

Beijing Forestry University

Xuri Hao

Beijing Forestry University

Rui Zhang

Beijing Forestry University

Chenqi Li

Beijing Forestry University

Yuepeng Song

yuepengsong@bjfu.edu.cn

Beijing Forestry University <https://orcid.org/0000-0001-8946-6437>

Research Article

Keywords: flavonoid, high-throughput sequencing, lignin, seed development, Tilia tuan, TO-GCN

Posted Date: June 21st, 2024

DOI: <https://doi.org/10.21203/rs.3.rs-4509655/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Version of Record: A version of this preprint was published at Plant Cell Reports on August 1st, 2024.
See the published version at <https://doi.org/10.1007/s00299-024-03287-8>.

Abstract

Tilia tuan Szyszyl. (*T. tuan*) is a timber tree species with important ecological and commercial value. However, its highly lignified pericarp results in a low seed germination rate and a long dormancy period. In addition, it is unknown whether there is an interaction between the biosynthesis of flavonoids and lignin as products of the phenylpropanoid pathway during seed development. To explore the molecular regulatory mechanism of lignin and flavonoid biosynthesis, *T. tuan* seeds were harvested at five stages (30, 60, 90, 120, and 150 days after pollination) for lignin and flavonoid analyses. The results showed that lignin accumulated rapidly in the early and middle stages (S1, S3, and S4), and rapid accumulation of flavonoids during the early and late stages (S1 and S5). High-throughput RNA sequencing analysis of developing seeds identified 50,553 transcripts., including 223 phenylpropanoid biosynthetic pathway genes involved in lignin accumulation grouped into three clusters, and 106 flavonoid biosynthetic pathway genes (FBPGs) grouped into two clusters. Subsequent WGCNA and time-ordered gene co-expression network (TO-GCN) analysis revealed that 24 TFs (e.g., *TtARF2* and *TtWRKY15*) were involved in lignin and flavonoid biosynthesis regulation. The transcriptome data were validated by qRT-PCR to analyze the expression profiles of key enzyme-coding genes. This study revealed that there existed a competitive relationship between the flavonoid biosynthesis pathway and the lignin biosynthesis pathway, both of which are downstream pathways of phenylpropane that provide a foundation for the further exploration of molecular mechanisms underlying lignin and flavonoid accumulation in *T. tuan* seeds.

Key message

Transcriptomics and phenotypic data analysis identified 24 transcription factors (TFs) that play key roles in regulating the competitive accumulation of lignin and flavonoids.

1. Introduction

Tilia is a genus in the subfamily Tiliodeae (family Malvaceae) that comprises 23 species of deciduous trees distributed in temperate areas (Yan et al. 2022; Mabberley 2008; Pigott et al. 2012). *Tilia* species are cultivated worldwide for their ecological and commercial value in honey and timber production, and as ornamental garden trees (Cai et al. 2015; Negri et al. 2013; Shahabi and Yassa 2015). However, the highly lignified pericarp of most *Tilia* species makes it difficult for water and oxygen to penetrate, resulting in a low germination rate, particularly in *T. tuan* (Magherini and Nin 1994; Maastricht 1993).

Lignin is a complex macromolecular insoluble aromatic polymer in vascular plant tissues, which provides mechanical support, physical strength, and rigidity for cell structures. (Tribot et al. 2019; Zhao 2016). The lignin biosynthetic process involves monolignol biosynthesis and monolignol polymerization (Banu et al. 2019). A study explored the role of phenylpropanoid pathway enzymes in the lignin biosynthesis pathway in transgenic *Populus* (Wang et al. 2018). The phenylpropanoid pathway provides precursors for the synthesis of secondary metabolites, that flavonoids, lignin, and coumarins (Fraser and

Chapple 2011). Lignin originates from the deamination of phenylalanine in the phenylpropane synthetic pathway, which is catalyzed by PHENYLALANINE AMMONIA-LYASE (PAL) (Ayyachamy et al. 2013; Ragauskas et al. 2014; Lacombe et al. 1997). PAL deaminates phenylalanine to produce cinnamic acid, which is converted to secondary metabolites such as coumarin and flavonoids (Winkel-Shirley et al. 2001). Other enzymes that are essential to the lignin pathway include TCYP73A, 4CL, HCT, C3'H, CCR, CAD, CCM, PRDX6/POD, F5H, and COMT etc. (Jansen et al. 2014). The hard, lignified pericarp is a main factor contributing to the low germination rate of *T. tuan* seeds. Various seed treatments before sowing improve the germination rate (Heit 1977). Lignin quality and accumulation can be regulated by controlling the expression levels of lignin biosynthetic pathway genes (LBPGs) (Boudet and Grima-Pettenati 1996). However, the regulatory mechanism of lignin accumulation in *T. tuan* pericarp has not been studied.

Flavonoids are plant secondary metabolites with several functions including signaling, protection against biotic and abiotic stressors, and as pigments. Flavonoid biosynthesis originates with the phenylpropanoid pathway (Dixon and Paiva 1995; Weissshaar and Jenkins 1998; Winkel-Shirley et al. 2001). Flavonoids can control the development process of seeds by interacting with various hormones, such as seed size and fertility (Allan et al. 2008; Shi and Xie 2014; Zhang et al. 2014). CHLCONE SYNTHASE (CHS) is the first catalytic enzyme in the flavonoid pathway; its substrate is 4-coumaroyl-CoA, which is catalyzed by CHS to naringenin chalcone (Wade et al. 2003; Winkel-Shirley et al. 2001). Most enzymes involved in pathways leading to the major flavonoid classes have been determined, such as CHI, F3H, DFR, ANS, and FGT etc. (Shih et al. 2008), and most of the key enzymes involved in these pathways are encoded by a single gene in *Arabidopsis thaliana* (Shih et al. 2008; Winkel et al. 2006).

However, the accumulation pattern of flavonoids in *T. tuan* seeds has not been studied, and previous reports have mainly focused on their medicinal functions. In addition, as the main component of thick peel in *T. tuan* seeds, the biosynthetic pattern of lignin in the *T. tuan* seeds has not been reported yet. Therefore, analyzing the accumulation patterns of lignin and flavonoids in *T. tuan* seeds can not only reveal the reasons for the thick peel produced by *T. tuan* seeds but also provide new insights to reveal the interaction between the flavonoid biosynthesis pathway and the lignin biosynthesis pathway in *T. tuan* seeds. Previous studies shown that the lignin/flavonoid yield can be controlled by regulating one or some genes related to lignin/flavonoid biosynthesis. However, the important genes that regulate the biosynthesis of lignin/flavonoids in *T. tuan* seeds still need to be fully studied.

In this study, we explored the biosynthesis and regulatory mechanisms of lignin and flavonoids during the development of *T. tuan* seeds at the transcriptional level. The first, we measured lignin and flavonoid synthetic rates during *T. tuan* seed development and analyzed their variation. Then, we analyzed the expression patterns of genes encoding enzymes associated with lignin and flavonoid biosynthesis, and proposed a *T. tuan* lignin and flavonoid regulatory network. We identified lignin/flavonoid-specific clusters or modules through WGCNA and Mfuzz clustering (Langfelder and Horvath 2008) to construct a transcription factor (TF)–gene regulatory network controlling the accumulation of lignin/flavonoids, and analyzed the correlation of their co-expression profiles with genes in the lignin and flavonoid biosynthetic

pathways. Revealed the relationship between the flavonoid biosynthesis pathway and the lignin biosynthesis pathway which both as phenylpropanoid downstream pathway. We also utilized a time-course transcriptomics analysis to construct a TO-GCN. By analyzing time series transcriptome data to analyze the dynamic response of plants to environmental changes over time, which is an effective method for studying gene regulatory networks. (Chang et al. 2018). Nevertheless, to date, no study has applied TO-GCN to specifically study seed development. Therefore, we constructed a TO-GCN to explore the response of *T. tuan* seeds during development in a time series, establish the potential regulatory networks of lignin and flavonoid biosynthesis, and establish the potential regulatory relationship between TFs and non-TFs in each subnetwork. This is the first study to apply time-course transcriptomics analysis to *Tilia* seed development. Finally, we conducted a preliminary analysis of the *T. tuan* seed development response network.

Based on previous studies, *WRKY15* (Ge et al., 2020) and *ARF2* (Jiang et al., 2022) were able to regulate the biosynthesis of lignin and flavonoids in plants, respectively. However, it is currently unclear whether these two TFs have a regulatory effect on lignin and flavonoid biosynthesis during the development of *T. tuan* seeds. We hypothesize that these two TFs have similar functions during the development of *T. tuan* seeds and conducted the following research analysis. Based on the analysis of the synthesis rate of lignin and flavonoids and the changes of pathway gene expression levels, it was determined that there had competitive relationship between the two pathways. At the same time, special attention is paid to the screening of TFs with regulatory effects on LPPGs and FBPGs. In summary, the results of this analysis explored the relationship between the two biosynthetic pathways of lignin and flavonoids during the development of *T. tuan* seeds and obtained TFs that have regulatory effects on both pathways, providing new insights for the biosynthesis of the two types of compounds.

2. Materials and methods

2.1 Plant materials

T. tuan seeds were collected from mature plants in the natural growth state in Beijing Forestry University, Beijing, China (Lat. 116°20'E, 40°00'N, and Alt. 55m). To simulate the natural growth state of *T. tuan* maximally. The sampling method was random sampling, Mature flowers were pollinated, and the globose fruits were harvested at 30-day intervals from June 1st to October 1st after pollination for five developing stages (S1-S5, refer to June 1st July 1st August 1st September 1st October 1st after pollination, respectively). The collected seeds were immediately frozen in liquid nitrogen and stored at temperature – 80°C for further analysis.

2.2 Analysis of seed lignin content and flavonoid content

The lignin content and flavonoid content of *T. tuan* seeds from five different developmental stages were determined using the method described by the Assay Kit (<https://www.bboxbio.cn/>). Three biological replicates in each stage were used for content determination. Briefly, the sample was dried at 80°C until

it reached a constant weight, and then homogenized with a MiniBeadbeater-96, crushed, and sieved through a 30–50 mesh sieve, which was the sample to be tested.

Approximately 3 mg of seeds were used for each analysis placed in a 1.5 ml centrifuge tube, empty tube was used as control, and then add reagents in sequence according to the product instructions, finally obtain the supernatant, and suck 200 µl the clear liquid was put into a 96 well UV plate, and the absorbance value at 280 nm was measured by readmax1200 light absorption full wavelength microplate reader, and then the result was obtained according to the calculation formula of lignin content.

Approximately 100 mg of seeds were used for each analysis placed in a 1.5 ml brown centrifuge tube, add 1 ml of extraction solution, conduct ultrasonic extraction (program setting: 60°C, power 300 W, ultrasonic for 5 s, intermittent for 8 s, total time 30min), centrifuge 12000r at room temperature for 10min, and take the supernatant as the sample to be tested. Set measuring tube, control tube, standard tube (used to establish standard curve) and blank tube, and then add reagents in sequence according to the product instructions, finally obtain the supernatant, and suck 200 µL the clear liquid was put into a 96 well UV plate, and the absorbance value at 510 nm was measured by readmax1200 light absorption full wavelength microplate reader. The standard equation ($y = b + kx$) was obtained by using the standard curve, and then the result was obtained according to the calculation formula of flavonoid content (flavonoid content (mg/g) = x/w , w is sample quality).

2.3 RNA extraction and sequencing

Selected seeds from five stages (S1-S5) of *T. tuan* tree seed development process for RNA extraction and transcriptome analysis. Three biological replicates in each stage were used for the transcriptome sequencing. Used the RNA prep Pure Plant Plus Kit (Polysaccharides and Polyphenolics-rich) (TIANGEN BIOTECH, BEIJING) to extracted u total RNA. A NanoPhotometer spectrophotometer (NanoDrop Technologies, USA) and an Agilent 2100 Bioanalyzer (Agilent) were used for the quantification. A total of 1 µg RNA per sample was used as input for RNA sample preparation, and sequencing libraries were generated using the NEB Next ®Ultra TMRNA Library Prep Kit for Illumina® (NEB, USA). The library fragments were purified using the AM Pure XP System (Beckman Coulter, Beverly, USA) and screened for cDNA fragments around 200bp to finalise the library. Preliminary quantification was performed using a Qubit2.0 Fluorometer, and the libraries were diluted to 1.5 ng/ul, followed by evaluation of library quality using an Agilent 2100 bioanalyzer system, and accurately determine the effective concentration of the library by qRT-PCR (the effective concentration was higher than 2 nm). After the libraries were tested, the different libraries were pooled according to target data volume requirements and the effective concentration of libraries and sequenced on the Illumina Novaseq platform, and 150bp paired-end reads were generated.

2.4 Transcriptome sequencing data analysis and gene set enrichment analysis (GSEA)

Raw sequencing data from high-throughput sequencing platforms were filtered to exclude aptamers, ploy-N, and low-quality reads contained in raw Illumina sequencing reads to obtain pure data (clean reads). Clean data for each sample were analysed using Trinity software with default parameters for assembly generation with a minimum linkage of 150bp. Short read data were analysed using RSEM to determine transcript levels using full length transcripts as reference sequences. Gene expression levels were quantified as the number of fragments per kilobase of transcript per million mapped reads (FPKM). Meanwhile, Q20, Q30 and GC content of raw reads were calculated. Transcribed sequences (unigenes) were annotated using BLAST searches of the NR, SwissProt, GO, Pfam and KEGG protein databases. differentially expressed genes (DEGs) were screened using the DESeq2 R software package (1.16.1). $p\text{-value} < 0.05$ and $|\log_2| > 2$ was used as thresholds to identify significantly DEGs. The differentially expressed gene sequences were then used for enrichment analysis on the official website of Kyoto Encyclopedia of Genes and Genomes (KEGG) (<https://www.genome.jp/kegg/pathway.html>) to screen for transcripts encoding enzymes related to lignin and flavonoid synthesis pathways. For gene set enrichment analyses, we used pathway genes annotated by KEGG enrichment analyses as gene sets of known function and performed GSEA studies on lignin/flavonoid biosynthesis pathway gene sets according to previously published methods and available software (Subramanian et al. 2005).

2.5 WGCNA and visualization of co-expressed gene networks

Gene co-expression networks were constructed based on all genes from the transcriptome data analysis by resolving co-expression relationships between genes using R software (Langfelder and Horvath 2008). Used the automatic network construction function (blockwise Modules) to build co-expression modules with a soft threshold of NA, in addition, mergeCutHeight, deepSplit, and minModuleSize were 0.25, 4, and 30, respectively. All other parameters are default values. The gene co-expression module was constructed by sequentially normalising the FPKM values, constructing the adjacency matrix and constructing the topological overlap matrix (TOM). The correlation between each gene module and lignin accumulation in unit time (LAUT) and flavonoid accumulation in unit time (FAUT) during fruit development was calculated. The expression matrix was visualised using Cytoscape v3.8.2 to characterise the correlation between paired transcripts.

2.6 Expression profiles analysis

Differentially expressed genes were verified using RNA samples provided by Beijing Novogene Technology Co., Ltd. 15 candidate genes were selected and their expression profiles were determined by qRT-PCR. The CdNAs were synthesised using the TB Green® Premix Ex Taq™ RR420A kit (TaKaRa, Japan) and then subjected to qRT-PCR on the QuantStudio 6 Flex Real-Time PCR system (ThermoFisher Scientific, USA). The PEROXIDASE6-2 (POD6-2) was used as the internal control. All reactions were repeated four times. The relative abundance of 15 candidate genes was calculated using the $2^{-\Delta\Delta C_p}$ method. qPCR primers are listed in Supplemental Table S1.

2.7 Time-ordered gene co-expression network (TO-GCN) analysis

Co-expression relationships and multi-level regulatory relationships between TFs and pathway genes were explored by constructing TO-GCN between TFs and lignin/flavonoid biosynthesis pathway genes (Chang et al. 2018). A gene with the highest expression in S1 and down-regulated expression in both S2–S5 was set as a bait gene. The Pearson's coefficient (PCC) between TF genes was calculated and the PCC threshold was set at 0.8. The breadth-first search (BFS) algorithm (Knuth 1997) was then used to assign temporal order levels to all genes in the major GCNs. Regulatory networks were mapped using Cytoscape v3.8.2.

3. Results

3.1 Temporal patterns of phenotypic and lignin and flavonoid content changes

The temporal pattern of the phenotypic changes in *T. tuan* seeds during five developmental stages (S1–S5) was analyzed. The 1,000-grain weight of the seeds increased rapidly from S1 to S4, and then slowed in mature seeds at S5. Similarly, the diameter of the seeds increased rapidly from S1 to S3, whereas it declined slightly with seed development at S5 (Fig. 1A and Table S2). The color of the pericarp remained light green at S1 and S2, and then turned yellow-green at S3, and yellowish-brown from S3 to S5. These significant changes in pericarp color probably indicated rapid lignin accumulation.

Mature *T. tuan* seeds have a solid ligneous pericarp (Fig. 1A). The temporal patterns of LAUT and FAUT of *T. tuan* seeds were analyzed at five developmental stages (S1–S5) (Fig. 1B). The amount of lignin produced at S1 was 9.032 mg/g. Lignin content increased gradually from S1 to S5. LAUT was highest at S4 (0.507 mg/g). The amount of flavonoids produced at S1 was 8.187 mg/g. Flavonoid content increased gradually from S1 to S5. The greatest increase (3.754 mg/g) was detected at S2. Thus, the accumulation in unit time (AUT) of both compounds appeared to exhibit a competitive relationship during the early stages of seed development (S2 and S3), and complex interactions during the late stages of seed development (S4 and S5).

3.2 Differential expression of lignin and flavonoid biosynthetic pathway genes during seed development

The phenylpropanoid biosynthetic pathway is the main pathway for lignin biosynthesis. To fully understand the differential expression patterns of genes associated with lignin production, we determined their expression by assembling and quantifying clean sequencing data from RNA samples. The longest transcript in each gene was identified as a unigene, and 50,553 unigenes were identified. The principal component analysis results showed differential gene expression among the stages (Fig. S1A). The correlation values of most samples at the same stage were generally high (Fig. S1B),

indicating a high level of reproducibility in the gene expression patterns. A total of 223 of the 50,553 unigenes were mapped to the phenylpropanoid biosynthetic pathway (Table S3). A total of 223 LBPGs encoded 12 key biosynthetic enzymes (Table S3). Significant differences were observed in the number of members in the gene families. The number of *PEROXIDASE*-like members was the largest, which was 52 genes, whereas *FERULATE-5HYDROXYLASE*-like had only one gene. To fully understand the differential expression patterns of genes associated with flavonoid production, 106 of the 50,553 unigenes were mapped to the flavonoid biosynthetic pathway (Table S4). In total, 106 FBPGs belonged to 10 gene families that regulate the biosynthesis of 10 key enzymes (Table S4). The numbers of genes differed significantly among gene families; *CHS* had the most members (41 genes), whereas *FLAVONOL SYNTHASE* (*TtFLS*) had only one.

3.3 Lignin/flavonoid biosynthetic pathway genes and TF hierarchical cluster analysis

A total of 223 genes encoding key synthases in the lignin biosynthetic pathway were identified during *T. tuan* seed development: *TtPAL* (23 genes), *TtCYP73A* (17), *Tt4CL* (28), *TtHCT* (6), *TtC3'H* (5), *TtCCR* (9), *TtCAD* (37), *TtCCM* (16), *TtPRDX6* (15), *TtPOD* (52), *TtF5H* (1), and *TtCOMT* (14). The expression levels of most LBPGs decreased during seed development and tended to be consistent with the accumulation of lignin metabolites (Fig. 2A). Hierarchical cluster analysis was performed based on the normalized FPKM values of the transcripts during the five developmental stages. The results showed that the 223 LBPGs were grouped into three clusters (Fig. S2A). The expression levels of 62 genes in cluster 1 increased gradually and continuously (Fig. S2B). Among these, there were 21 *TtCAD*, 15 *TtPRDX6*, 11 *Tt4CL*, five *TtPOD*, four *TtCCM*, two *TtHCT*, and one *TtCOMT* gene, respectively. *TtPLA*, *TtCCR*, and *TtC3'H* each had only one gene (Table S3). Cluster 2 consisted of 56 genes. Expression levels increased slowly from S1 to S3 and then rapidly decreased during S4 and S5. Notably, a few genes had upregulated expression at S5, including one *TtCYP73A*, two *Tt4CL*, 3 *TtCAD*, and four *TtPOD*. The expression levels were decreased continuously from the 105 genes in cluster 3, and fastest from S1 to S2 (Fig. S2B). The expression levels of these genes in cluster 3 were positively correlated with the amount of lignin synthesized during S1–S2, suggesting that genes may play a significant role regulating lignin biosynthesis during S1–S2.

A total of 106 genes encoding key synthases in the flavonoid biosynthetic pathway were identified during *T. tuan* seed development: *TtCHS* (41), *TtCHI* (2), *FLAVONOID 3',5'-HYDROXYLASE* (5 *TtCYP75A*), *FLAVONOID 3'-MONOOXYGENASE* (10 *TtCYP75B1*), *TtDFR* (7), *TtF3H* (16), *TtFLS* (1), *LEUCOANTHOCYANIDIN REDUCTASE* (4 *TtLAR*), *ANTHOCYANIDIN SYNTHASE* (15 *TtANS*), and *ANTHOCYANIDIN REDUCTASE* (5 *TtANR*). The expression of most of the FBPGs decreased during seed development (Fig. 2B). Hierarchical cluster analysis was performed based on the normalized FPKM values of the transcripts. The results showed that 106 FBPGs were grouped into two clusters (Fig. S3A). The expression of eight genes in cluster 1 showed a downward–upward–downward trend. Expression levels were rapidly downregulated during the S1–S2 and S3–S4 stages, upregulated during the S2–S3

stages and slowly downregulated during S4–S5. This group included three *TtDFR* and two *TtANR*. In addition, the expression levels of 98 genes in cluster 2 decreased continuously (Fig. S3B and Table S4).

In summary, metabolite and transcript profiling of *T. tuan* seeds during development revealed a unique rapid accumulation of lignin and flavonoids at S1 and upregulation of most genes related to both metabolites (Fig. 1B and Fig. 2A, B). Gene set enrichment analysis (GSEA) was applied to determine whether the lignin/flavonoid biosynthetic pathway gene set was highly expressed during S1. The GSEA results showed that both terms were significantly enriched during the S1 phase compared with S2–S5 (Fig. 2C, D).

A total of 2,173 differentially expressed putative TFs, classified into 76 families, were detected in *T. tuan* seeds according to the PlantTFDB database (Fig. 2E and Table S5). Among them, the MYB, C3H, C2H2, bZIP, and NAC TF families were the highest level of expression, containing 160, 115, 112, 107, and 99 members, respectively. Cluster analysis of all TFs indicated that most of TF families were highly expressed during S1–S3. In addition, the FPKM values of some TF families were higher than those of other TF families., including CSD, S1Fa-like, MBF1, HMG, and C2C2-YABBY (Fig. 2E). In addition, Mfuzz clustering results showed that the expression patterns of group 2 TFs (Fig. 2F) were consistent with those of LBPGs (group 2; Fig. S2) and FBPGs (group 1; Fig. S3), including 107 bZIPs, 100 HBs, 99 NACs, 80 GRASs and 48 EILs. The expression pattern of TFs in group 3 (Fig. 2F) was consistent with that of LBPGs (group 3; Fig. S2) and FBPGs (group 2; Fig. S3), which included 160 MYBs, 68 bHLHs, 56 ARFs, 49 AUX/IAAs and 41 TUBs. The above results suggest that these TFs may play an important role in the regulation of lignin and flavonoid biosynthesis in *T. tuan* (Fig. 2F).

3.4 Co-expression network analyses identified flavonoid and lignin biosynthetic pathway genes

By constructing a co-expression network for all expressed genes WGCNA, we further understood the biosynthesis patterns of lignin and flavonoids in *T. tuan* seeds. After the deletion of outlier values, 45,399 genes remained. These genes were clustered into 14 large modules, each represented by a different color, with grey modules containing genes of low relevance discarded. (Fig. 3A, Table S6). Based on gene expression levels, 45,399 genes were clustered into 12 clusters using the Mfuzz R package (Fig. S4). Analysis of the relationships between gene modules and lignin and flavonoid biosynthesis showed that the green module (417 genes) had the highest positive correlation coefficients with lignin; however, there were no lignin or flavonoid pathway genes in the green module. Therefore, we selected the tan module with the second highest correlation coefficients for analysis. The tan module (17,106 genes) also had the highest positive correlation coefficients with flavonoids (Fig. 3B), suggesting strong synergy between lignin and flavonoid biosynthesis during *T. tuan* seed development. The brown (4,433 genes) and black (723 genes) modules were significantly negatively correlated with flavonoids and lignin, respectively (Fig. 3B). Therefore, we selected tan (17106 genes), brown (4433 genes), and black (723 genes) modules for further research. Genes preferentially expressed during the early stages of *T. tuan* seeds development (S1) mainly clustered in the tan module (Fig. 3C), whereas the number of highly expressed

genes in the tan module decreased and then increased during development, indicating a strongly positive correlation with LAUT and FAUT. However, the number of genes preferentially expressed in the brown module first increased (S2–S4) and then decreased (S4–S5), indicating a significant negative correlation with FAUT. The number of genes preferentially expressed in the black module decreased gradually (S2–S5), which was positively correlated with the gradual decrease in LAUT during S4–S5 and may be related to the few LBPGs (3 genes) in the black module.

Although these three modules were strongly correlated with lignin and flavonoids, the genes in all three modules had differential expression patterns, indicating that these genes may perform different functions in the biosynthesis of lignin and flavonoids. Furthermore, the Mfuzz clustering analysis results with the patterns of gene expression changes in the three modules were consistent, in which 4,440 genes (Fig. S4, cluster 1) were highly expressed during the early stage of fruit growth (S1), 2,190 genes (cluster 9) were highly expressed during S3–S4, and 3,779 genes (cluster 7) were highly expressed from S2 to S3 (Fig. 3C–E). Further analysis found that most of the genes in these three modules correspond to the genes in the selected Mfuzz cluster. Although the algorithms of WGCNA and Mfuzz were different, the results of these two clustering methods were consistent.

3.5 Identification of candidate central genes associated with flavonoid and lignin biosynthesis

Module hub genes were representative of given modules in the biological networks. In the tan module, 84 genes were strongly correlated (edge weight ≥ 0.626) (Table S7 and Fig. S6A). Among these, chaperone protein DnaJ 10-like, upf0591 membrane protein, mitochondrial-like, and actin-depolymerizing factor 2-like had the highest degree of connectivity. The network graph generated by the brown module contained 187 co-expressed genes (edge weight ≥ 0.599), and genes with the highest degree of connectivity included phospholipase D alpha 2 isoform 1 and hypothetical protein GOBAR_DD18655 (Table S8 and Fig. S6B). The network graph generated by the black module contained 81 co-expressed genes (edge weight ≥ 0.599), and genes with the highest degree of connectivity included plant peroxidase and glycosyl hydrolase protein families (Table S9 and Fig. S6C). The above findings suggest that these genes may be involved in flavonoid and lignin biosynthesis. However, this finding requires further verification.

3.6 Construction of biosynthesis regulatory network for flavonoid and lignin

By screened genes involved in flavonoid and lignin biosynthesis pathways in tan, brown, and black modules. 105 LBPGs and 71 FBPGs (weight > 0.35) were found in the brown module, and their expression levels were closely related to the accumulation of lignin and flavonoids (Tables S3 and S4). We identified 110 TFs through associating the expression patterns of these LBPGs (Table S10), and the expression levels were found to be strongly correlated with the 27 LBPGs in the tan module (Fig. S7A). We identified 105 TFs (Table S11) in which the expression levels were strongly correlated with the 17 FBPGs in the tan module (Fig. S7A). Among these, 44 TFs were associated with genes from both

pathways (Table S12), including AUX/IAA (AUX/IAAPB.1133.1, AUX/IAAPB.864.1), NAC (NACPB.9926.1, NACPB.9973.1), MYB (MYBPB.5506.1), WRKY (WRKYPB.13481.1, WRKYPB.22195.1), and SBP (SBPPB.2437.1, SBPPB.700.1), which generated related networks (Fig. 4A).

We identified 20 LBPGs and one FBPG in the brown module (weight > 0.35). The expression patterns were highly correlated with flavonoid and lignin accumulation (Tables S3 and S4). We identified 61 (Table S13) and 46 TFs (Table S14), the expression levels were highly correlated with LBPG and FBPG in the brown module, respectively (Figure S7B). 31 TFs were associated with genes from both pathways (Table S15), including bHLH, C2H2, ARF, and bZIP; thus, we generated a correlated network (Fig. 4B). The above results suggest that these TFs corresponded to the putative regulators regulating lignin and flavonoid biosynthesis in *T. tuan*. Due to the identification of only three candidate lignin biosynthesis genes in the black module, no corresponding TF-gene pairs were detected.

3.7 Quantitative real-time PCR validated candidate differentially expressed genes (DEGs)

To assess the quality of the transcriptome data, we used qRT-PCR to analysed the changing patterns of relative expression levels of 15 DEGs candidate genes (10 LBPGs and 5 DETFs) associated with lignin accumulation. The analysis showed that the RNA-seq and qRT-PCR results of the candidate genes had the same trend of change (Fig. 5), and there was a significant correlation between the FPKM values and the $2^{-\Delta\Delta Ct}$ values of 12 (80%) candidate genes ($r_p > 0.8$) (Table S16), indicating the validity of the transcriptome data.

3.8 Temporal gene co-expression networks during *T. tuan* seed development

We performed time-ordered comparative transcriptome analyses detected 45555 genes (2,150 TFs and 43,405 structural genes), and their FPKM > 0.01 in at least one sample (Table S17). As the initial node, The TF gene (PB.34500.1) (Fig. 6A), which was highly expressed at S1 but not later, was selected to generate a TO-GCN (Chang et al. 2018). A large, major TO-GCN was reconstructed that containing 2,009 TFs, 216 LBPGs, and 106 FBPGs (L1–L10) using the suggested positive/negative cutoff values (0.83; -0.51; Fig. S8) (Chang et al. 2018), and 10 time-series expression levels of 1,983 TFs, 215 LBPGs, and 106 FBPGs (26 TFs and one LBPG were cleaned) (Fig. 6B). The remaining 26 TFs formed 13 GCNs, each GCN < 10 TFs (Fig. S8). The TFs in the TF TO-GCN were assigned to 10 levels (Fig. 6B). These constructed levels matched the expression time order of the TFs over the five-time points during seed development, as shown in the diagonal red square of the heat map constructed by averaging the normalized FPKM values (z-scores) (Fig. 6B, C). Time-course transcriptomes were further distinguished for the initial (S1), transitional (S2 and S3), and terminal stages (S4 and S5) of lignin and flavonoid pigmentation (Fig. 6C).

Genes assigned to S1 (corresponding to time-series levels L1–L2) had the highest FPKM values, followed by those assigned to the transitional stages (S2–S3, corresponding to L3–L4) and those expressed at the terminal stages (S4–S5, corresponding to L5–L10). We obtained the basis for the upstream and downstream gene/trait regulatory relationships at the network level by coupling gene expression peaks between continuous levels (L1–L10) and time-series samples (S1–S5). Periods of high-expression overlapped at successive levels, suggesting a potential hierarchical structure of regulatory pathways that may be formed by upper-level genes through the regulation of same level or lower-level co-expressed genes. The network, including genes contained in L1 and L2 (618 TFs, 105 FBPGs, and 83 LBPGs), was associated with S1, this includes 31 TFs co-expressed with LBPGs and FBPGs identified in WGCNA (Fig. 6D), suggesting that these genes may be downregulated during seed development. The genes (895 TFs, 34 FBPGs, and 68 LBPGs) of L3–L4 corresponded to those of S2/S3, such that this stage represents the middle period of flavonoid and lignin accumulation, this includes 8 TFs co-expressed with LBPGs and FBPGs identified in WGCNA (Fig. 6D). The remaining 644 genes (571 TFs, 9 FBPGs, and 64 LBPGs) were associated with L7–L10, corresponding to S4/S5, and were assigned to a later period of flavonoid and lignin accumulation, this includes 8 TFs co-expressed with LBPGs and FBPGs identified in WGCNA (Fig. 6D).

3.9 Regulatory dynamics of lignin and flavonoid accumulation in *T. tuan*

Based on the TO-GCN, we retrieved eight LBPGs and two FBPGs in L1 respectively, and constructed a subnetwork (Fig. 7A). The subnetwork included PAL, CYP73A, CCR, CCM, COMT, POD, LBPGs, DFR, and F3H and their potential upstream regulatory TFs. Classifying all the TFs in the sub-network, the results show that the FBPGs in L1 were mainly regulated by the TF gene families, including MYB (7), AUX/IAA (7), SBP (7), NAC (5), and bZIP (4). The LBPGs in L1 were mainly regulated by TF gene families, including MYB (14), AUX/IAA (12), HB (15), AP2-EREBP (8), and PHD (7). A heatmap constructed based on the FPKM values showed that the expression peaks of all genes in L1 occurred at S1, except that of *TtF3H* (FPB.274700.1), and these genes were significantly downregulated during seed development (Fig. 7A, Table S18).

The lignin biosynthesis regulator *WRKY15* (Ge et al. 2020) was detected in L2, and the same was true for the WGCNA. Therefore, we retrieved 12 LBPGs and one FBPG in L2 co-expressed with *TtWRKY15*, respectively, and constructed a related network (Fig. 7B). The subnetwork mainly included CAD (4), CCR (2), 4CL (2), CYP73A (1), CCM, HCT, F3H, and 285 TFs with upstream regulatory potential. A heatmap constructed based on the FPKM values showed that the expression peaks of all genes in the subnetwork occurred at S1, and these genes were gradually downregulated during seed development (Fig. S9, Table S19).

Many TFs related to seed development or the regulation of lignin and flavonoid biosynthesis appeared at earlier levels in the TF TO-GCN. For example, *ABI3* plays key roles during seed development (Lalanne et al. 2021), *ARF2-1/2* plays a key role during flavonoid biosynthesis (Jiang et al. 2022), *WRKY15* (Ge et al.

2020) plays a key roles during lignin biosynthesis and belongs to L2, and *WRKY32* plays a key role during flavonoid biosynthesis (Song et al. 2022) and belongs to L3. In addition, several TFs closely related to seed development and stress regulation have been reported, including *TtERF72*, *TtSPL7*, and *TtbZIP44-1/2* (Liu et al. 2018; Gielen et al. 2016; Lglesias-Fernández et al. 2013). This result indicates that our method effectively revealed key regulators of lignin and flavonoid biosynthesis.

A joint analysis of the WGCNA (Fig. 4A) and TO-GCN (Fig. 7A) networks revealed 3 and 12 common TFs in Brown, TAN, and L1, respectively. Eight and 38 common TFs were observed in Brown, TAN, and L2, respectively (Fig. 7C). In addition, the joint analysis revealed the conserved regulatory network of lignin and flavonoid accumulation during the development of *T. tuan* seeds. There were 24 TFs (including *TtARF2* and *TtWRKY15*, which play important roles in flavonoid and lignin biosynthesis), 10 LBPGs, and four FBPGs (Fig. 7D). Expression level changes in these TFs were similar to the lignin accumulation pattern (Fig. 7E and 1B).

4 Discussion

Tilia tuan is woody species with important ecological and commercial value (Cai et al. 2015; Shahabi and Yassa 2015). However, the large quantity of accumulated lignin produces a hard pericarp in this species, which contributes to its low seed germination rate. Our comprehensive analysis of the lignin-related transcriptome and lignin accumulation phenotypes at different stages of *T. tuan* seed development and maturation generated a regulatory network related to LBPGs. We also analyzed the FBPG regulatory network, which was downstream of the phenylpropanoid pathway. Using our dataset, we identified TFs that may play key roles in lignin and flavonoid production and provide a valuable resource for improving the quality of *T. tuan*.

Quantitative analyses of differences in lignin and flavonoids at different growth stages of *T. tuan* increased our understanding of changes in LAUT and FAUT in *T. tuan* seeds. Both lignin and flavonoids are products of the shikimate pathway (D'Auria and Gershenzon 2005; Vogt 2010). A previous study of *Paphiopedilum armeniacum* seeds showed that LAUT is low at the early developmental stage and gradually increases during later seed developmental stages (Fang et al. 2020), which was consistent with most of our metabolic data (Fig. 1B). Interestingly, *P. armeniacum* also has a highly lignified pericarp. Lignin deposition in the seed coat protects the seed from adversity (Liu et al. 2018). Decreased activity of *PAL*, the entry enzyme into the phenylpropanoid pathway and *C4H*, decreases lignin content in transgenic plants (Elkind et al. 1990; Bate et al. 1994; Sewalt et al. 1997), perhaps due to the decreased expression or activity of *PAL* or *C4H*, which reduce flux in the lignin pathway during the early stage of lignin biosynthesis, leading to lower lignin content (Sewalt et al. 1997). We identified 23 members of the *TtPAL* family and 17 members of the *TtC4H* family in this study (Fig. 2A, Table S3). Most of these genes (21 *TtPALs* and 15 *TtC4Hs*) were expressed at very high levels during seed development stage S1, which was consistent with LAUT at this stage (Fig. 1). The expression of two *TtPALs* during S2, including LPB.13022.1 (*TtPAL 1*), was upregulated (Table S3), but the lignin biosynthetic rate did not increase. *PAL1* is necessary for lignin biosynthesis in *Arabidopsis thaliana* in response to temperature and nitrogen

stress (Olsen et al. 2008). Our result also shows that the upregulation of *TtPAL 1* expression failed to increase the lignin synthetic rate in *T. tuan* seeds, which contrasts with its regulatory role in *A. thaliana*. Given the almost identical transcriptional content among members of the *PtPAL* family, and that the expression of the *PtPAL 1* translation protein was 20% of that of other members, *PtPAL 1* was determined not to be the main functional gene for lignin formation in *Populus trichocarpa* (Shi et al. 2013). In summary, unique regulatory mechanisms for lignin synthesis may be present in perennial woody plants. The decrease in the lignin synthetic rate during maturation of *T. tuan* seeds observed in this study may be more closely related to the overall downregulation of downstream genes in the lignin biosynthetic pathway, rather than solely determined by *TtPAL 1*.

In Glycine max, exogenous C4H, as an allelochemical, not only increases the total tissue lignin content, but also alters the total amount and ratio of H -, G - and S - lignin monomers in the tissue, leading to increased cell wall stiffness. (Salvador et al. 2013). We identified 17 genes that control C4H, of which 15 *TtC4H* were continuously downregulated during S1–S5, consistent with the changes in LAUT, indicating that C4H activity gradually decreases during seed development. Transgenic *Populus tremuloides* with the antisense *Pt4CL* gene has 30–40% less lignin (Li et al. 2003). Downregulation of *Zm4CL* expression during maize seed germination results in a significant decrease in lignin content (Zhao et al. 2022). Our results were consistent with this finding. Unlike the S4 stage, 17 of 28 members in the *Tt4CL* family displayed varied downregulated expression during S5, leading to a decrease in LAUT. In addition, 11 members displayed varied upregulated expression during S5. This result indicates that, consistent with the *PAL 1* analysis, upregulated genes cannot upregulate LAUT, as these 11 genes may be redundant in lignin biosynthesis. Cinnamyl alcohol dehydrogenase (CAD) catalyzes the last step in the biosynthesis of monolignols, which are lignin precursors. CAD downregulation in transgenic poplar using antisense and co-suppression strategies did not significantly change either the amount of lignin or the lignin monomeric composition (syringyl/guaiacyl) (Baucher et al. 1996). Only five (13.51%) of the 37 *TtCADs* identified in this study were highly expressed in S1; 21 (56.76%) *TtCADs* were highly expressed in S5 (Fig. 2A), which was negatively correlated with LAUT at this stage. This finding is similar to those of Baucher et al. (1996), who showed that the downregulation of *TtCADs* does not decrease in LAUT during *T. tuan* seed development.

Ferulic acid-5-hydroxylase (F5H) is a key enzyme in the synthesis of the monolignol sinapyl alcohol, and an S-lignin monomer (Meyer et al. 1998). S-lignin is almost nondetectable in *A. thaliana* F5H mutants (Marita et al. 1999). In addition, overexpression of *F5H* failed to detect a significant increase in S-lignin, as COMT was unable to keep pace with the increase of 5-hydroxy coniferyl aldehyde generated by *F5H* overexpression (Ralph et al. 2001). The reduced expression levels of *COMT* and *CAD*, which are late enzymes in lignin monomer synthesis, affect the composition of lignin but not its content (Tetreault et al. 2018). *F5H* was overexpressed in poplar, and a significant increase in S-lignin was detected, because the final methylation step catalyzed by COMT catalyzed an increase in 5-hydroxyconiferol (Li et al. 2000; Li et al. 2003). We identified a *TtF5H* with continuously low expression levels during S1–S5, but with continuously increasing LAUT during S2–S4, indicating that the composition of the lignin monomer synthesized during S2–S4 may have changed. In our study, some genes of other lignin biosynthesis-

related enzymes were highly expressed during S5, whereas LAUT declined, suggesting that F5H acts as a control valve in the process of lignin accumulation in *T. tuan* seeds. However, this hypothesis requires further research.

The *A. thaliana* stems and anthers of *AtCCR1*, *AtCAD1*, and *AtCAD2* gene mutants exhibit low lignin content compared to the increased level of flavonoid components (flavonol glycosides) in stems (Thévenin et al. 2011). In addition, the downregulation of *AtHCT* gene expression in *A. thaliana* decreases lignin levels and increases flavonoid levels (Besseau et al. 2007). These studies demonstrate that changes in the expression of the *CCR*, *CAD*, and *HCT* genes affect the entire flavonoid and lignin metabolic pathway. Thus, our analysis suggests that complex changes in the expression levels of *TtCCR*, *TtCAD*, and *TtHCT* at different stages may explain the series of changes in LAUT and FAUT. Flavonoids are important plant secondary metabolites that play various roles (Winkel-Shirley et al. 2001). PAL deaminates phenylalanine to cinnamic acid, which is transformed into secondary metabolites such as coumarins and flavonoids, by a variety of central enzymes (Watts et al. 2006). In our study, *T. tuan* seeds developed rapidly, with rapid increases in quality and volume during S1, and the color of the pericarp changed most significantly during S1–S2, probably due to the large quantity of flavonoids synthesized in the seeds. Both of these findings are consistent, as FAUT and almost all FBPGs were significantly enriched during S1 (Figs. 1 and 2B). Putative unigenes in flavonoid biosynthesis, i.e., *CHS*, *CHI*, *DFR*, *F3H*, *FLS*, *LAR*, *ANS*, and *ANR*, had their highest expression levels during S1 (Fig. 2B). Therefore, we conclude that flavonoid accumulation occurred mainly during S1 of seed development. Enzymes encoded by different members of the 4CL family have different metabolic functions and exhibit different substrate affinities (Wang et al. 2022). Upregulated expression of 4CL in tobacco increases flavonoid accumulation (Li et al. 2021; Li et al. 2020). Eleven members of the *Tt4CL* gene family from *T. tuan* were upregulated at S5, which was positively correlated with the significant increase of FAUT at this stage. Considering that almost all FBPGs were minimally expressed during S2–S5, we speculate that *Tt4CL* plays a unique role in the later stage of flavonoid accumulation. Additionally, LAUT decreased during S5, possibly because coumaroyl-CoA produced by 4CL catalysis entered the flavonoid pathway. This indicates a potential competitive relationship between the flavonoid biosynthesis pathway and the lignin biosynthesis pathway for coumaroyl CoA from the phenylpropanoid pathway. The increase of FAUT and the decrease of LAUT during S5 confirms this conjecture (Fig. 1B).

In this study, co-expression analysis of LBPGs, FBPGs and potential TFs was performed. TFs, as key regulators in metabolic networks, can regulate the expression of multiple genes and can be regulated by multiple TFs. (Grotewold et al. 2008). TFs binds directly to key genes in the flavonoid and lignin biosynthetic pathway to regulate the accumulation of flavonoids and lignin in several species, such as *CmMYB8* (Zhu et al. 2020) *PtoMYB156* (Yang et al. 2017), *AtWRKY12* (Wang et al. 2010), *HvWRKY23* (Karre et al. 2019), *PtrbHLH186* (Liu et al. 2022), and *PtNAC101* (Qu et al. 2023). Transcriptome analysis of *T. tuan* seeds showed that various TF families were expressed at high levels and were co-expressed with some lignin and FBPGs, such as *WRKYPB.22195.1* (*TtWRKY15*), *WRKYPB.13481.1* (*TtWRKY32*), *NACPB.9926.1* (*TtNAC62-1*), *bHLHPB.3725.1* (*TtbHLH62*), and *AUX/IAAPB.1133.1* (*TtARF2*) (Fig. 4,

Table S12). This result suggests that these TFs may play a role in regulating the accumulation of lignin and flavonoids in *T. tuan* seeds.

We conducted TO-GCN using the full set of genes to explore the regulatory relationship between TFs and the lignin/flavonoid biosynthetic pathway during *T. tuan* seed development. TO-GCN has been used extensively in studies of other species. For example, entire metabolic pathways for anthocyanins and carotenoids in azaleas and their potential regulatory networks have been reconstructed based on TO-GCN. (Yang et al. 2020). Furthermore, this analytical method was used to determine the biological regulation process of auxins and carrigin in *Pinus tabulaeformis* Carr. by promoting flavonoid biosynthesis and enhancing antioxidant capacity to resist UV-B. (Xu et al. 2021). We divided the gene clusters during seed development into 10 expression levels, L1–L10 (Fig. 6B, C), and constructed sub-networks that included the lignin and flavonoid biosynthetic pathways (Fig. 7A) to explore the mechanism of lignin/flavonoid biosynthesis during *T. tuan* seed development. The TO-GCN analysis identified many TFs that may be related to the accumulation of lignin and flavonoids in *T. tuan* (Fig. 9A). These TFs also appeared in the WGCNA results (Fig. 4A); therefore, we reconstructed the co-expression network of these TFs with LBPGs and FBPGs through joint analysis of the WGCNA and TO-GCN networks that included *TtWRKY15* and *TtARF2* (Fig. 7D). These two TFs have been demonstrated to participate in the lignin or/and flavonoid biosynthetic pathways. *AtARF2* acts as a positive regulator to fine-tune the spatial and temporal accumulation of flavonoids (mainly flavonols in *A. thaliana*) (Jiang et al. 2022). In our study, *TtARF2* expression levels were consistently downregulated, concurrent with FAUT at S1–S3, and had a negative regulatory relationship at S4–S5, suggesting that *TtARF2* exhibits different regulatory effects at different *T. tuan* seed developmental stages. Overexpression of *AtWRKY15* results in reduced lignification, and downregulation of *AtWRKY15* results in increased lignification in *A. thaliana* (Ge et al. 2020). In our study, the expression level of *TtWRKY15* had different regulatory relationships with changes in LAUT and FAUT at different stages of seed development in *T. tuan*. Both were downregulated simultaneously during S1–S2, and both exhibited a negative regulatory relationship at S3–S4. This result suggests that the *WRKY15* regulatory mode differs among woody and herbaceous plants, or in different plant organs. In summary, our analysis suggests that *TtARF2* and *TtWRKY15* may be involved in the regulation of flavonoid and lignin accumulation during the development of *T. tuan* seeds. However, decreased activity of the *TtWRKY15* transcript activates the biosynthesis of callose in pollen (Li et al. 2020); therefore, *TtWRKY15* may have a regulatory effect on fertility of *T. tuan* seeds. In addition, NAC and MYB family members were identified (Fig. 7D), and several members of these two families participate in the regulation of lignin biosynthesis (Wang et al. 2022). However, the regulatory effect of these TFs on the accumulation of lignin/flavonoids during the development of *T. tuan* seeds requires further verification.

Overall, this study was the first to systematically study the biosynthesis patterns of lignin and flavonoid in *T. tuan* seeds under natural conditions and their relationship with the expression patterns of pathway genes. The regulatory network of pathway genes-TFs was constructed for the first time by joint WGCNA and TO-GCN analysis, and TFs (*TtWRKY15* and *TtARF2*) that regulate lignin and flavonoid biosynthesis were obtained based on network analysis. These research results have opened the research of biological

changes during the maturation and development of *T. tuan* seeds, provided new information to overcome the problem of difficult germination of *T. tuan* seeds, and promoted the regional breeding of *T. tuan*.

5. Conclusions

In conclusion, this study explored the biosynthesis and regulation of lignin and flavonoids in *T. tuan* seeds at the transcriptional level. A joint WGCNA and TO-GCN analysis showed that the accumulation of lignin and flavonoids involved a variety of factors, including TFs and lignin- and flavonoid-related enzymes, and identified important genes involved in lignin and flavonoid accumulation. By conducting functional analysis of candidate genes, we will further expand our understanding of the biosynthesis and regulation of lignin and flavonoids in *T. tuan* seeds. However, FAUT and LAUT in *T. tuan* seeds changed similarly during S1–S4, and showed the opposite developmental trend during S5. Our analysis identified 24 TFs with potential regulatory roles in the biosynthesis of lignin and flavonoids. Potential competition for carbon flow between flavonoid and lignin biosynthesis during different stages of *T. tuan* seed development requires further research. These results provide valuable information needed to deepen our understanding of seed development in *T. tuan*, and provide a basis for elucidating the formation mechanism of lignified pericarp in this seed type.

Declarations

Acknowledgements

This work was supported by the Major Science and Technology Projects of Inner Mongolia Autonomous Region(2021ZD0008), the 111 Project (No. B20050). We are grateful for the sequence information produced by the U.S. Department of Energy Joint Genome Institute (<http://www.jgi.doe.gov>).

Author Contributions

SY and LL conceived and designed research. LC contributed analytical methods. HX, ZR and LC provided assistance in analyzing data. LL analyzed the data and wrote a manuscript. SY provided guidance throughout the process.

Conflict of interests

The authors declare that the contents of the manuscript have no conflicting interest

Data Availability

The Illumina RNA-seq data of *T. tuan* seeds were uploaded to the National Center for Biotechnology Information Sequence Read Archive (NCBI: <http://www.ncbi.nlm.gov/>

sra/;accessed on 23 July 2022) with BioProject accession number: PRJNA861431, PRJNA861958 and PRJNA878938.

References

1. Allan AC, Hellens RP, Laing WA (2008) MYB transcription factors that colour our fruit. *Trends Plant Sci.* 13, 99–102. <https://doi.org/10.1016/j.tplants.2007.11.012>.
2. Ayyachamy M, Cliffe FE, Coyne JM, Collier J, Tuohy MG (2013) Lignin: untapped biopolymers in biomass conversion technologies. *Biomass Convers. Biorefinery.* 3, 255–269. <https://doi.org/10.1007/s13399-013-0084-4>.
3. Bate NJ, Orr J, Ni W, Meroni A, Nadler-Hassar T, Doerner P.W, Dixon RA, Lamb CJ, Elkind Y (1994) Quantitative relationship between phenylalanine ammonia-lyase levels and phenylpropanoid accumulation in transgenic tobacco identifies a rate determining step in natural product synthesis. *Proc. Natl. Acad. Sci. U.S.A.* 91, 7608-7612. <https://doi.org/10.1073/pnas.91.16.7608>.
4. Baucher M, Chabbert B, Pilate G, Doorselaere JV, Tollier MT, Petit-Conil M, Cornu D, Monties B, Montagu MV, Inze D, Jouanin L, Boerjan W (1996) Red Xylem and Higher Lignin Extractability by Down-Regulating a Cinnamyl Alcohol Dehydrogenase in Poplar. *Plant Physiology.* 112(4), 1479–1490. <https://doi.org/10.1104/pp.112.4.1479>.
5. Boudet AM, Grima-Pettenati J (1996) Lignin genetic engineering. *Mol. Breeding.* 2, 25-39. <https://link.springer.com/article/10.1007/BF00171349>.
6. Besseau S, Hoffmann L, Geoffroy P, Lapierre C, Pollet B, Legrand M (2007) Flavonoid accumulation in *Arabidopsis* repressed in lignin synthesis affects auxin transport and plant growth. *The Plant cell.* 19(1), 148–162. <https://doi.org/10.1105/tpc.106.044495>.
7. Barros J, Serrani-Yarce JC, Chen F, Baxter D, Venables BJ, Dixon RA (2016) Role of bifunctional ammonia-lyase in grass cell wall biosynthesis. *Nature Plants.* 2, 16050. <https://doi.org/10.1038/nplants.2016.50>.
8. Banu JR, Kavitha S, Kannah RY, Devi TP, Gunasekaran M, Kim SH, Kumar G (2019) A review on biopolymer production via lignin valorization. *Bioresource technol.* 290, 121790. <https://doi.org/10.1016/j.biortech.2019.121790>.
9. Cai J, Ma PF, Li HT, Li DZ (2015) Complete Plastid Genome Sequencing of Four *Tilia* Species (Malvaceae): A Comparative Analysis and Phylogenetic Implications. *Plos One* 10(11), e0142705. <https://doi.org/10.1371/journal.pone.0142705>.
10. Chang YM, Lin HH, Liu WY, Li WH (2018) Comparative transcriptomics method to infer gene co-expression networks and its applications to maize and rice leaf transcriptomes. *Proc. Natl. Acad. Sci. U.S.A.* 116(8) 3091-3099. <https://doi.org/10.1073/pnas.1817621116>
11. Dixon RA, Paiva NL (1995) Stress-induced phenylpropanoid metabolism. *The plant cell.* 7(7), 1085-1097. <https://doi.org/10.1105/tpc.7.7.1085>.
12. D'Auria JC, Gershenzon J (2005) The secondary metabolism of *Arabidopsis thaliana*: growing like a weed. *Current opinion in plant biology.* 8(3), 308–316. <https://doi.org/10.1016/j.pbi.2005.03.012>.
13. De Baerdemaeker NJF, Salomón RL, Roo LD, Steppe K (2017) Sugars from woody tissue photosynthesis reduce xylem vulnerability to cavitation. *New Phytol.* 216, 720–727.

<https://doi.org/10.1111/nph.14787>.

14. Elkind Y, Edwards R, Mavandad M, Hedrick SA, Ribak O, Dixon RA, Lamb CJ (1990) Abnormal plant development and down regulation of phenylpropanoid biosynthesis in transgenic tobacco containing a heterologous phenylalanine ammonia-lyase gene. *Proc. Natl. Acad. Sci. U.S.A.* 87, 9057-9061. <http://www.jstor.org/stable/2355687>.
15. Fraser C, Chapple C (2011) The phenylpropanoid pathway in *Arabidopsis*. *Arabidopsis book*. 9, e0152. <https://doi.org/10.1199/tab.0152>.
16. Fang L, Xu X, Li J, Zheng F, Li M, Yan J, Li Y, Zhang X, Li L, Ma G, Zhang A, Lv F, Wu K, Zeng S (2020) Transcriptome analysis provides insights into the non-methylated lignin synthesis in *Paphiopedilum armeniacum* seed. *BMC Genomics* 21, 524. <https://doi.org/10.1186/s12864-020-06931-1>.
17. Guillet G, Poupart J, Basurco J, Luca VD (2000) Expression of Tryptophan Decarboxylase and Tyrosine Decarboxylase Genes in Tobacco Results in Altered Biochemical and Physiological Phenotypes. *Plant Physiology*. 122(3), 933–944. <https://doi.org/10.1104/pp.122.3.933>.
18. Grotewold E (2008) Transcription factors for predictive plant metabolic engineering: are we there yet? *Current opinion in biotechnology*. 19(2), 138–144. <https://doi.org/10.1016/j.copbio.2008.02.002>.
19. Gielen H, Remans T, Vangronsveld J, Cuypers A (2016) Toxicity responses of Cu and Cd: the involvement of miRNAs and the transcription factor SPL7. *BMC Plant Biol.* 16(1), 1-16. <https://doi.org/10.1186/s12870-016-0830-4>.
20. Ge S, Han X, Xu X, Shao Y, Zhu Q, Liu Y, Du J, Xu J, Zhang, S (2020) WRKY15 Suppresses Tracheary Element Differentiation Upstream of VND7 During Xylem Formation. *The Plant Cell*. 32(7), 2307–2324. <https://doi.org/10.1105/tpc.19.00689>.
21. Heit C (1977) Propagation from seed. part 27. collecting testing and growing linden species. *Amer. Nurs.* 146(7), 100-110. <http://digital.casalini.it/10.1400/14299> - Casalini id: 2208995.
22. Jansen F, Gillesen B, Mueller F, Commandeur U, Fischer R, Kreuzaler F (2014) Metabolic engineering for p-coumaric alcohol production in *Escherichia coli* by introducing an artificial phenylpropanoid pathway. *Biotechnol. Appl. Biochem* 61(6), 646–654. <https://doi.org/10.1002/bab.1222>.
23. Jiang W, Xia Y, Su X, Pang Y (2022) ARF2 positively regulates flavonols and proanthocyanidins biosynthesis in *Arabidopsis thaliana*. *Planta*. 256, 44. <https://doi.org/10.1007/s00425-022-03936-w>.
24. Knuth DE (1997) The art of computer programming (Vol. 3). Pearson Education.
25. Karre S, Kumar A, Yogendra K, Kage U, Kushalappa A, Charron JB (2019) *HvWRKY23* regulates flavonoid glycoside and hydroxycinnamic acid amide biosynthetic genes in barley to combat *Fusarium* head blight. *Plant. Mol. Biol.* 100(6), 591–605. <https://doi.org/10.1007/s11103-019-00882-2>.
26. Lacombe E, Hawkins S, Doorselaere JV, Piquemal J, Goffner D, Poeydomenge O, Boudet AM, Grima-Pettenati J (1997) Cinnamoyl CoA reductase, the first committed enzyme of the lignin branch

- biosynthetic pathway: cloning, expression and phylogenetic relationships. *Plant J.* 11(3), 429-41. <https://doi.org/10.1046/j.1365-313X.1997.11030429.x>.
27. Li L, Popko JL, Umezawa T, Chiang VL (2000) 5-hydroxyconiferyl aldehyde modulates enzymatic methylation for syringyl monolignol formation, a new view of monolignol biosynthesis in angiosperms. *The Journal of biological chemistry.* 275(9), 6537–6545. <https://doi.org/10.1074/jbc.275.9.6537>.
 28. Li L, Zhou Y, Cheng X, Sun J, Marita J, Ralph J, Chiang VL (2003) Combinatorial modification of multiple lignin traits in trees through multigene contrasformation. *Proc. Natl. Acad. Sci. U.S.A.* 100(8), 4939–4944. <https://doi.org/10.1073/pnas.0831166100>.
 29. Langfelder P, Horvath S (2008) WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics.* 9, 559. <https://doi.org/10.1186/1471-2105-9-559>.
 30. Lglesias-Fernández R, Barrero-Sicilia C, Carrillo-Barral N, Oñate-Sánchez L, Carbonero P (2013) *Arabidopsis thaliana* bZIP44: a transcription factor affecting seed germination and expression of the mannanase-encoding gene AtMAN7. 74(5) 767-780. <https://doi.org/10.1111/tpj.12162>
 31. Liu K, Li Y, Chen X, Li L, Liu K, Zhao H, Zhao WY, Han S (2018) ERF72 interacts with ARF6 and BZR1 to regulate hypocotyl elongation in *Arabidopsis*. *Journal of Experimental Botany.* 69 (16) 3933–3947. <https://doi.org/10.1093/jxb/ery220>.
 32. Liu Q, Luo L, Zheng L (2018) Lignins: Biosynthesis and Biological Functions in Plants. *Int. J. Mol. Sci.* 19(2), 335. <https://doi.org/10.3390/ijms19020335>.
 33. Li SS, Chang Y, Li B, Shao SL, Zhang ZZ (2020) Functional analysis of 4-coumarate: CoA ligase from *Dryopteris fragrans* in transgenic tobacco enhances lignin and flavonoids. *Genetics and molecular biology*, 43(2), e20180355. <https://doi.org/10.1590/1678-4685-GMB-2018-0355>.
 34. Li Y, Li L, Wang Y, Wang YC, Wang NN, Lu R, Wu YW, Li XB (2020) Pollen-Specific Protein PSP231 Activates Callose Synthesis to Govern Male Gametogenesis and Pollen Germination. *Plant Physiology.* 184(2), 1024–1041. <https://doi.org/10.1104/pp.20.00297>.
 35. Li S, Chang Y, Teoh PL, Wang D, Mo J, Li B, Shao S (2021) Overexpression of *Df4CL1* from *Dryopteris fragrans* Enhances Flavonoids and Lignin Production in Transgenic Tobacco. *Russ. J. Plant Physiology.* 68, 110–117. <https://doi.org/10.1134/S1021443721010118>.
 36. Lalanne D, Malabarba J, Ly-Vu J, Hundertmark M, Delahaie J, Leprince O, Buitink J, Verdier J (2021) *Medicago ABI3* Splicing Isoforms Regulate the Expression of Different Gene Clusters to Orchestrate Seed Maturation. *Plants.* 10(8), 1710. <https://doi.org/10.3390/plants10081710>.
 37. Liu H, Gao J, Sun J, Li S, Zhang B, Wang Z, Zhou C, Sulis D B, Wang JP, Chiang VL, Li W (2022) Dimerization of PtrMYB074 and PtrWRKY19 mediates transcriptional activation of *PtrbHLH186* for secondary xylem development in *Populus trichocarpa*. *New phytologist.* 234(3), 918–933. <https://doi.org/10.1111/nph.18028>.
 38. Maastricht WI (1993) WOCMAP I - Medicinal and Aromatic Plants Conference: part 3 of 4.
 39. Magherini R, Nin S (1994) Propagation of selected *Tilia* spp. by seed and semihardwood cuttings. *Adv. Horticul. Sci.* 8(2), 91–96. <https://www.jstor.org/stable/42881744>.

40. Meyer K, Shirley AM, Cusumano JC, Bell-Lelong DA, Chapple C (1998) Lignin monomer composition is determined by the expression of a cytochrome P450-dependent monooxygenase in *Arabidopsis*. Proc. Natl. Acad. Sci. U.S.A. 95(12), 6619–6623. <https://doi.org/10.1073/pnas.95.12.6619>.
41. Marita J, Ralph J, Hatfield RD, Chapple C (1999) NMR characterization of lignins in *Arabidopsis* altered in the activity of ferulate-5-hydroxylase. Proc. Natl. Acad. Sci. U.S.A. 96. 12328–12332. <https://doi.org/10.1073/pnas.96.22.12328>.
42. Mabberley DJ (2008) A portable dictionary of plants, their classifications, and uses. Cambridge University Press., Cambridge, pp. 148.
43. Negri G, Santi D, Tabach R (2013) Flavonol Glycosides Found in Hydroethanolic Extracts from *Tilia Cordata*, a Species Utilized as Anxiolytics. Rev. Bras. plantas Med. 15 (2), 217–224. <https://doi.org/10.1590/S1516-05722013000200008>.
44. Olsen KM, Lea US, Slimestad R, Verheul M, Lillo C (2008) Differential expression of four *Arabidopsis* PAL genes: PAL1 and PAL2 have functional specialization in abiotic environmental triggered flavonoid synthesis. Plant Physiology. 165, 1491–1499. <https://doi.org/10.1016/j.jplph.2007.11.005>.
45. Pigott CD (2012) Lime-trees and basswoods. A biological monograph of the genus *Tilia*. Cambridge University Press., Cambridge, pp. 197.
46. Qu D, Wu F, Yang J, Li M, Yang L, Xie R, Zhou J, Yang J, Wang L, Su H (2023) Transcription factor PtNAC101 negatively regulates the lignin synthesis and salt tolerance in *Populus trichocarpa*. Environmental and Experimental Botany. 205, 105149. <https://doi.org/10.1016/j.envexpbot.2022.105149>.
47. Ralph J, Lapierre C, Marita J, Kim H, Lu F, Hatfield RD, Ralph SA, Chapple C, Franke R, Hemm MR, Doorselaere JV, Sederoff RR, O'Malley DM, Scott JT, MacKay JJ, Yahiaoui N, Boudet AM, Pean M, Pilate G, Jouanin L, Boerjan W (2001) Elucidation of new structures in lignins of CAD- and COMT-deficient plants by NMR. Phytochemistry. 57(6), 993–1003. [https://doi.org/10.1016/S0031-9422\(01\)00109-1](https://doi.org/10.1016/S0031-9422(01)00109-1).
48. Ragauskas AJ, Beckham GT, Bidy MJ, Chandra R, Chen F, Davis MF, Davison BH, Dixon RA, Gilna P, Keller M (2014) Lignin valorization: improving lignin processing in the biorefinery. Science. 344 (6185), 1246843. <https://www.science.org/doi/10.1126/science.1246843>.
49. Sewalt VJ, Ni W, Blount JW, Jung HG, Masoud SA, Howles PA, Lamb C, Dixon RA (1997) Reduced Lignin Content and Altered Lignin Composition in Transgenic Tobacco Down-Regulated in Expression of L-Phenylalanine Ammonia-Lyase or Cinnamate 4-Hydroxylase. Plant Physiology. 115(1), 41-50. <https://doi.org/10.1104/pp.115.1.41>.
50. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES, Mesirov JP (2005) Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. Proc. Natl. Acad. Sci. U.S.A. 102(43), 15545–15550. <https://doi.org/10.1073/pnas.0506580102>.
51. Shih CH, Chu H, Tang LK, Sakamoto W, Maekawa M, Chu IK, Wang W, Lo C (2008) Functional characterization of key structural genes in rice flavonoid biosynthesis. Planta. 228, 1043–1054.

<https://link.springer.com/article/10.1007/s00425-008-0806-1>

52. Salvador VH, Lima RB, dos Santos WD, Soares AR, Böhm PA, Marchiosi R, Ferrarese MLL, Ferrarese-Filho O (2013) Cinnamic acid increases lignin production and inhibits soybean root growth. *PLoS One*. 8, e69105. <https://doi.org/10.1371/journal.pone.0069105>
53. Shi R, Shuford CM, Wang JP, Sun YH, Yang Z, Chen HC, Tunlaya-Anukit S, Li Q, Liu J, Muddiman DC, Sederoff RR, Chiang VL (2013) Regulation of phenylalanine ammonia-lyase (PAL) gene family in wood forming tissue of *Populus trichocarpa*. *Planta*. 238, 487–497. <https://doi.org/10.1007/s00425-013-1905-1>.
54. Shi MZ, Xie DY (2014) Biosynthesis and metabolic engineering of anthocyanins in *Arabidopsis thaliana*. *Recent Pat. Biotechnol*. 8, 47–60.
<https://www.ingentaconnect.com/content/ben/biot/2014/00000008/00000001/art00007#>
55. Shahabi M, Yassa N (2015) Flavonoids from the Leaves of Iranian Linden, *Tilia Rubra* Subsp. *Caucasica*. *Res. J. Pharm*. 2(3), 17–22.
https://www.rjpharmacognosy.ir/article_9949_07c1e10763667241dec404c2fd97908b.pdf.
56. Song Q, Ji K, Yu X, Chen L, Wang LK, Gong W, Yuan D (2022) Dynamic metabolic and transcriptomic profiling reveal synthetic characters and regulators of flavonoid biosynthesis in *Camellia oleifera* seeds. *Industrial Crops and Products*. 186, 115295. <https://doi.org/10.1016/j.indcrop.2022.115295>.
57. Thévenin J, Pollet B, Letarnec B, Saulnier L, Gissot L, Maia-Grondard A, Lapierre C, Jouanin L (2011) The simultaneous repression of CCR and CAD, two enzymes of the lignin biosynthetic pathway, results in sterility and dwarfism in *Arabidopsis thaliana*. *Molecular plant*. 4(1), 70–82.
<https://doi.org/10.1093/mp/ssq045>.
58. Tetreault HM, Scully ED, Gries T, Palmer NA, Funnell-Harris DL, Baird L, Seravalli J, Dien BS, Sarath G, Clemente TE, Sattler SE (2018) Overexpression of the *Sorghum bicolor* *SbCCoAOMT* alters cell wall associated hydroxycinnamoyl groups. *PloS One*. 13(10), e0204153.
<https://doi.org/10.1371/journal.pone.0204153>.
59. Tribot A, Amer G, Alio MA, Baynast DH, Delattre C, Pons A, Mathias JD, Callois JM, Vial C, Michaud P, Dussap CG (2019) Wood-lignin: supply, extraction processes and use as bio-based material. *Eur Polym J*. 112, 228–240. <https://doi.org/10.1016/j.eurpolymj.2019.01.007>.
60. Weisshaar B, Jenkins GI (1998) Phenylpropanoid biosynthesis and its regulation. *Curr Opin Plant Biol*. 1(3), 251–257. [https://doi.org/10.1016/S1369-5266\(98\)80113-1](https://doi.org/10.1016/S1369-5266(98)80113-1).
61. Winkel-Shirley B (2001) Flavonoid biosynthesis. A colorful model for genetics, biochemistry, cell biology, and biotechnology. *Plant Physiology*. 126(2), 485–493.
<https://doi.org/10.1104/pp.126.2.485>.
62. Wade HK, Sohal AK, Jenkins GI (2003) *Arabidopsis* ICX1 is a negative regulator of several pathways regulating flavonoid biosynthesis genes. *Plant Physiology*. 131(2), 707–715.
<https://doi.org/10.1104/pp.012377>.
63. Winkel BSJ (2006) The biosynthesis of favonoids. In: Grotewold E (ed) *The Science of Flavonoids*. Springer Sciences & Business Media, New York. pp 71–96.

64. Watts KT, Mijts BN, Lee PC, Manning AJ, Schmidt-Dannert C (2006) Discovery of a substrate selectivity switch in tyrosine ammonia-lyase, a member of the aromatic amino acid lyase family. *Chemistry & biology*. 13(12), 1317–1326. <https://doi.org/10.1016/j.chembiol.2006.10.008>.
65. Vogt T (2010) Phenylpropanoid biosynthesis. *Molecular plant*, 3(1), 2–20. <https://doi.org/10.1093/mp/ssp106>.
66. Wang H, Avci U, Nakashima J, Hahn MG, Chen F, Dixon RA (2010) Mutation of WRKY transcription factors initiates pith secondary wall formation and increases stem biomass in dicotyledonous plants. *Proc. Natl. Acad. Sci. U.S.A.* 107(51), 22338–22343. <https://doi.org/10.1073/pnas.1016436107>.
67. Wang X, Lin L, Dong J, Ling J, Wang W, Wang H, Zhang Z, Yu X (2018) Simultaneous improvements of *Pseudomonas* cell growth and Polyhydroxyalkanoate production from a lignin derivative for lignin-consolidated bioprocessing. *Appl. Environ Microbiol.* 84, e01469–18. <https://doi.org/10.1128/AEM.01469-18>.
68. Wang Y, Guo L, Zhao Y, Zhao X, Yuan Z (2022) Systematic Analysis and Expression Profiles of the 4-Coumarate: CoA Ligase (4CL) Gene Family in Pomegranate (*Punica granatum* L.). *International journal of molecular sciences*. 23(7), 3509. <https://doi.org/10.3390/ijms23073509>.
69. Wang Y, Hou Y, Wang J, Zhao H (2022) Analyzing lignin biosynthesis pathways in rattan using improved co-expression networks of NACs and MYBs. *BMC Plant Biol.* 22, 411. <https://doi.org/10.1186/s12870-022-03786-4>.
70. Xu J, Nie S, Xu CQ, Liu H, Jia KH, Zhou SS, Zhao W, Zhou XQ, A El-Kassaby Y, Wang XR, Porth L, Mao JF (2021) UV-B-induced molecular mechanisms of stress physiology responses in the major northern Chinese conifer *Pinus tabulaeformis* Carr. *Tree Physiology*, 41(7), 1247–1263, <https://doi.org/10.1093/treephys/tpaa180>.
71. Yang L, Zhao X, Ran L, Li C, Fan D, Luo K (2017) *PtoMYB156* is involved in negative regulation of phenylpropanoid metabolism and secondary cell wall biosynthesis during wood formation in poplar. *Scientific reports*. 7, 41209. <https://doi.org/10.1038/srep41209>.
72. Yang FS, Nie S, Liu H, Shi TL, Tian XC, Zhou SS, Bao YT, Jia KH, Guo JF, Zhao WZ, An N, Zhang RG, Yun QZ, Wang XZ, Mannapperuma C, Porth L, Aly El-Kassaby Y, Robert SN, Wang XR, Van de Peer Y, Mao JF (2020) Chromosome-level genome assembly of a parent species of widely cultivated azaleas. *Nat Commun.* 11(1), 5269. <https://doi.org/10.1038/s41467-020-18771-4>.
73. Yan L, Wang H, Huang X, Li Y, Yue Y, Wang Z, Tang S (2022) Chloroplast Genomes of Genus *Tilia*: Comparative Genomics and Molecular Evolution. *Frontiers in Genetics*. 13, 925726. <https://doi.org/10.3389/fgene.2022.925726>.
74. Zhang Y, Butelli E, Martin C (2014) Engineering anthocyanin biosynthesis in plants. *Current opinion in plant biology*. 19, 81–90. <https://doi.org/10.1016/j.pbi.2014.05.011>
75. Zhao Q (2016) Lignification: Flexibility, Biosynthesis and Regulation, *Trends Plant Sci.* 21(8), 713–721. <https://doi.org/10.1016/j.tplants.2016.04.006>.

76. Zhu L, Guan Y, Zhang Z, Song A, Chen S, Jiang J, Chen F (2020) *CmMYB8* encodes an R2R3 MYB transcription factor which represses lignin and flavonoid synthesis in chrysanthemum. *Plant Physiology and Biochemistry*. 149, 217–224. <https://doi.org/10.1016/j.plaphy.2020.02.010>.
77. Zhao X, Niu Y, Bai X, Mao T (2022) Transcriptomic and Metabolic Profiling Reveals a Lignin Metabolism Network Involved in Mesocotyl Elongation during Maize Seed Germination. *Plants*. 11(8), 1034. <https://doi.org/10.3390/plants11081034>

Figures

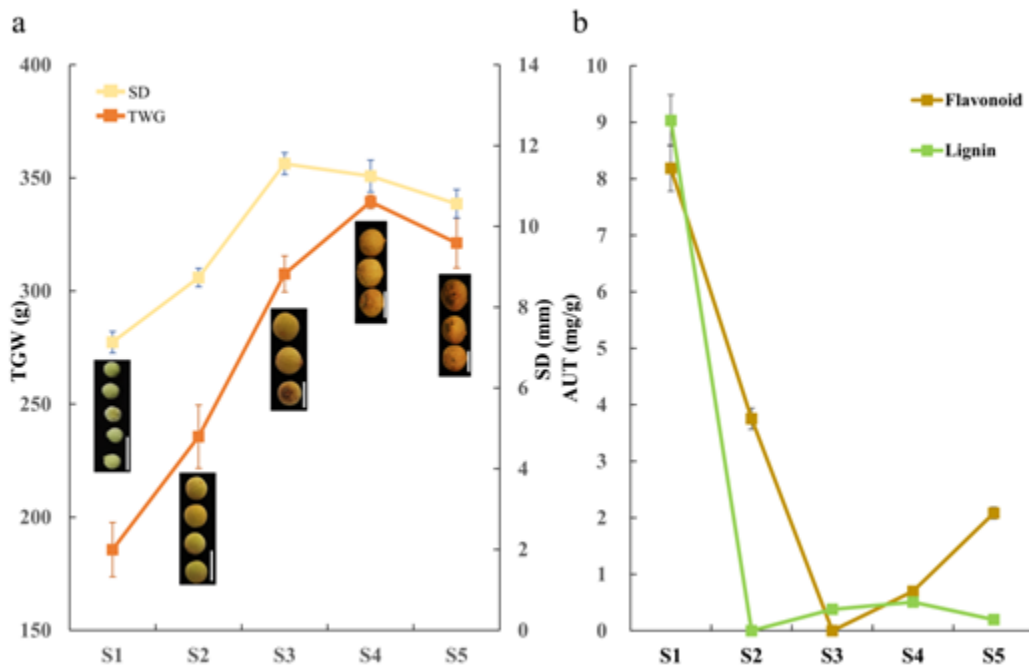


Figure 1

Temporal pattern of phenotypic changes in *Tilia tuan* seeds during five developmental stages (S1–S5). (a) Orange line represents the change in 1,000-grain weight (TGW) during seed development (stages S1–S5). Blue line represents changes in seed diameter (SD) during seed development. Five photographs illustrate the developmental progress of *T. tuan* seeds. The cone was harvested at 30 days after pollination (S1), and then every 30 days until 150 days after pollination (S5, mature seeds). (b) Green line represents the lignin accumulation in unit time (LAUT) for *T. tuan* seeds at different developmental stages (i.e., the increase in the lignin value between two neighboring measurement stages). Brown line represents the flavonoid accumulation in unit time (FAUT) of *T. tuan* seeds at different developmental stages (i.e., the increase in the flavonoid value between two neighboring measurement stages).

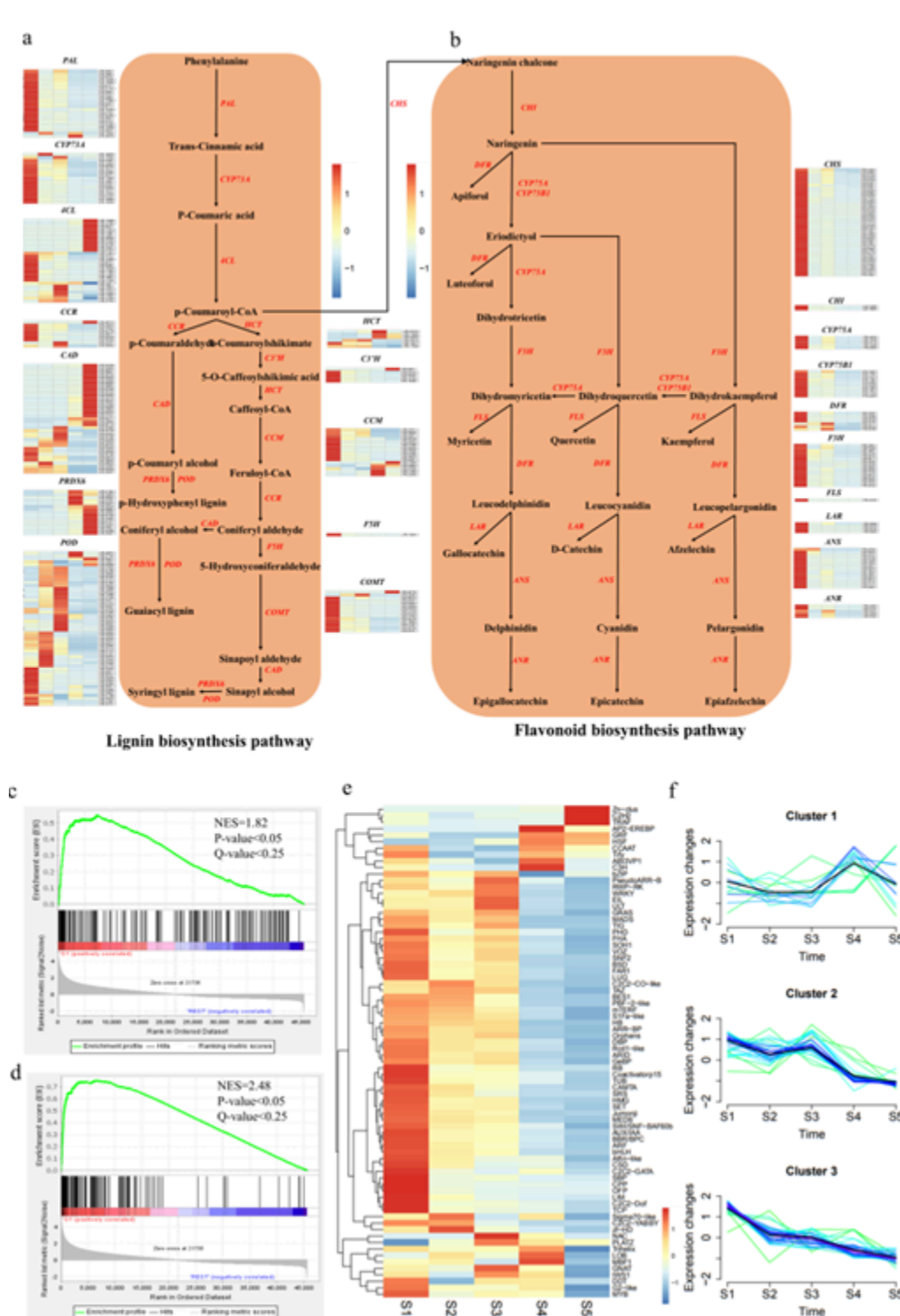


Figure 2

Biosynthetic Pathways of Lignin and Flavonoids in *T. tuan* and gene set enrichment analysis (GSEA) of the lignin and flavonoid biosynthetic pathway. Expression levels of LBPBs (a) and FBPGs(b) are shown, blue and red represent low expression and high expression, respectively. Z-scores were standardized to -2 to 2. (c, d) Results of the (c) lignin and (d) flavonoid biosynthesis pathway gene sets for GSEA analysis of stage S1 enrichment. Green line (above) indicates gene enrichment scores (ES, vertical axis). The peaked line indicates the ES of the gene set; the genes before the peak are identified as the core

genes of the gene set. One black barcode vertical line on the horizontal axis corresponds to one gene in the gene set. A plot of the rank value distribution of the background gene set is shown, where the vertical coordinate is a ranked list metric, i.e., fold-change values after formula processing, and the horizontal coordinate is the rank in the ordered dataset. Red and blue regions of the heatmap indicate genes highly expressed during S1 and other stages, respectively. The corresponding signal-to-noise ratio for each gene is shown as a grey area plot. (e) Expression levels of TF families. Different colours indicate the sum of the expression levels of all members of a particular TF family. Higher and lower levels of gene expression are indicated by red and blue, respectively. Z-scores are standardized from -2 to 2. (f) TFs were clustered into three sets based on total FPKM values using the Mfuzz R package, and the expression profiles of all TFs are also shown.

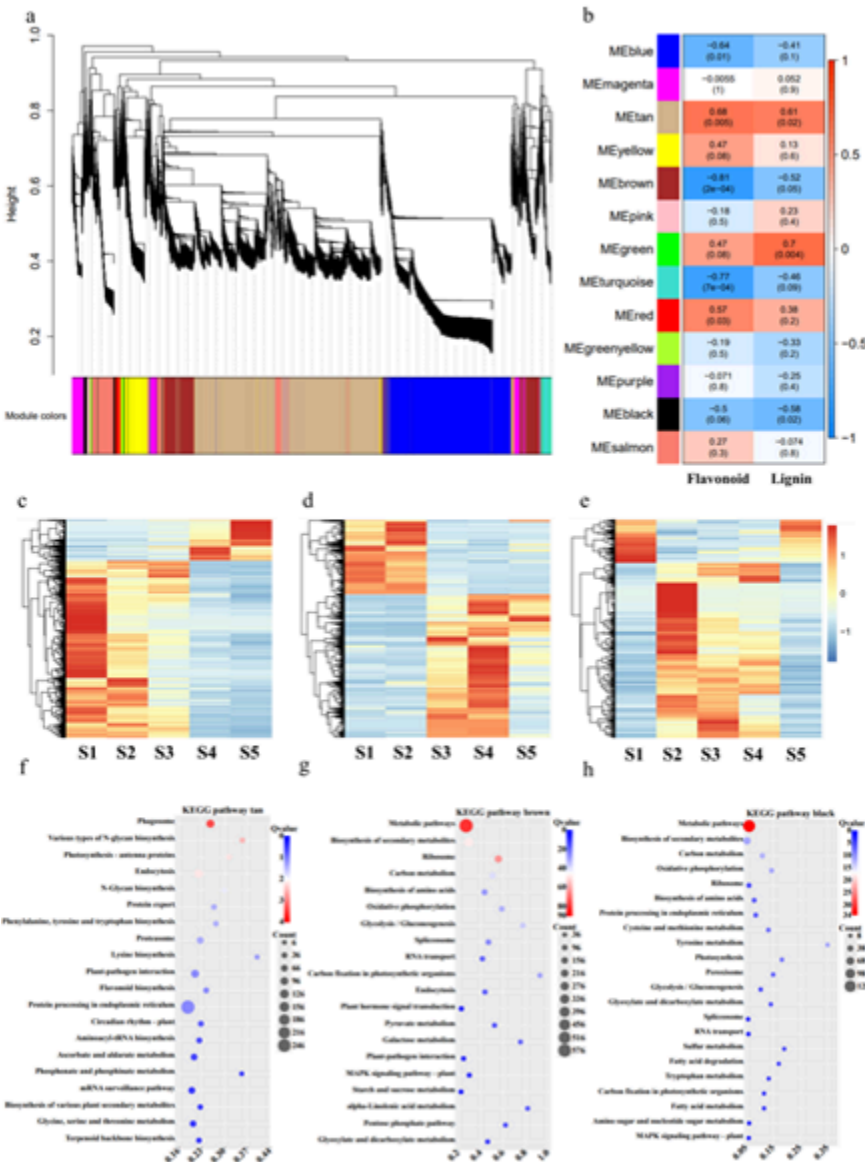


Figure 3

Transcriptomic and LAUT/FAUT correlation analysis during *T. tuan* fruit developmental stages. (a) Through weighted correlation network analysis (WGCNA), a total of 13 co-expression modules (clusters) were obtained. Mark the modules with different colors, and each leaf on the dendrogram corresponds to

a gene. (b) Correlation between FAUT/LAUT and modules. Each row represents a module. Columns 1 and 2 represent FAUT and LAUT, respectively (see Table S6). Values outside of parentheses in each cell at the row–column intersection are correlation coefficients between the module and FAUT and LAUT. Values in parentheses in each cell represent P values. The color scale expresses module–trait correlation (–1 to 1). (c–e) Expression levels of genes in the (c) tan, (d) brown, and (e) black modules. Z-scores were standardized to –2 to 2. (F–H) Selected Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analysis results for co-expressed genes in the (f) tan, (g) brown, and (h) black modules. Q-values are transformed as $-\log_{10}Q$ -values.

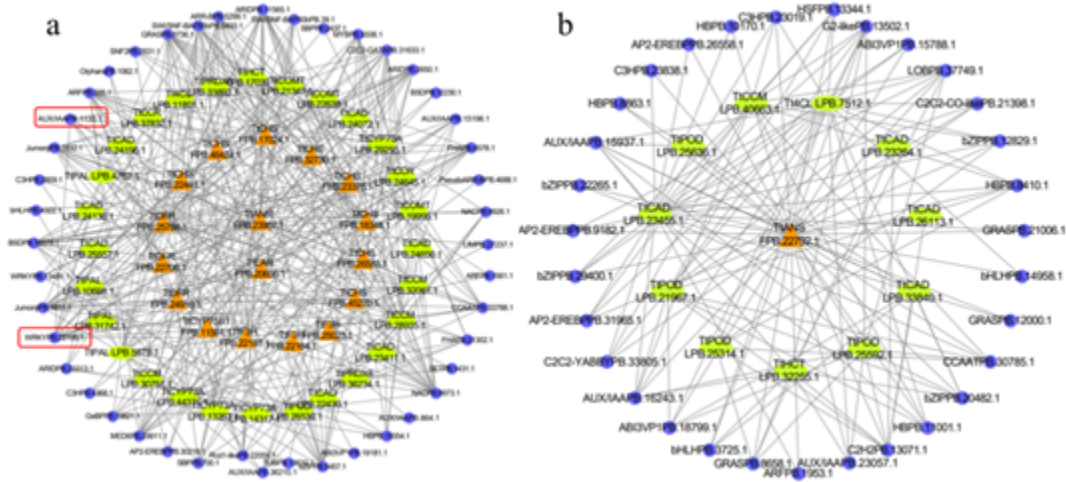


Figure 4

Regulatory network for lignin and flavonoid biosynthesis in *T. tuan*. (a) Tan module. (b) Brown module. Triangles and ellipses indicate pathway genes involved in flavonoid and lignin biosynthesis during fruit development and ripening, respectively. Circles represent TFs associated with flavonoid and lignin pathway gene expression.

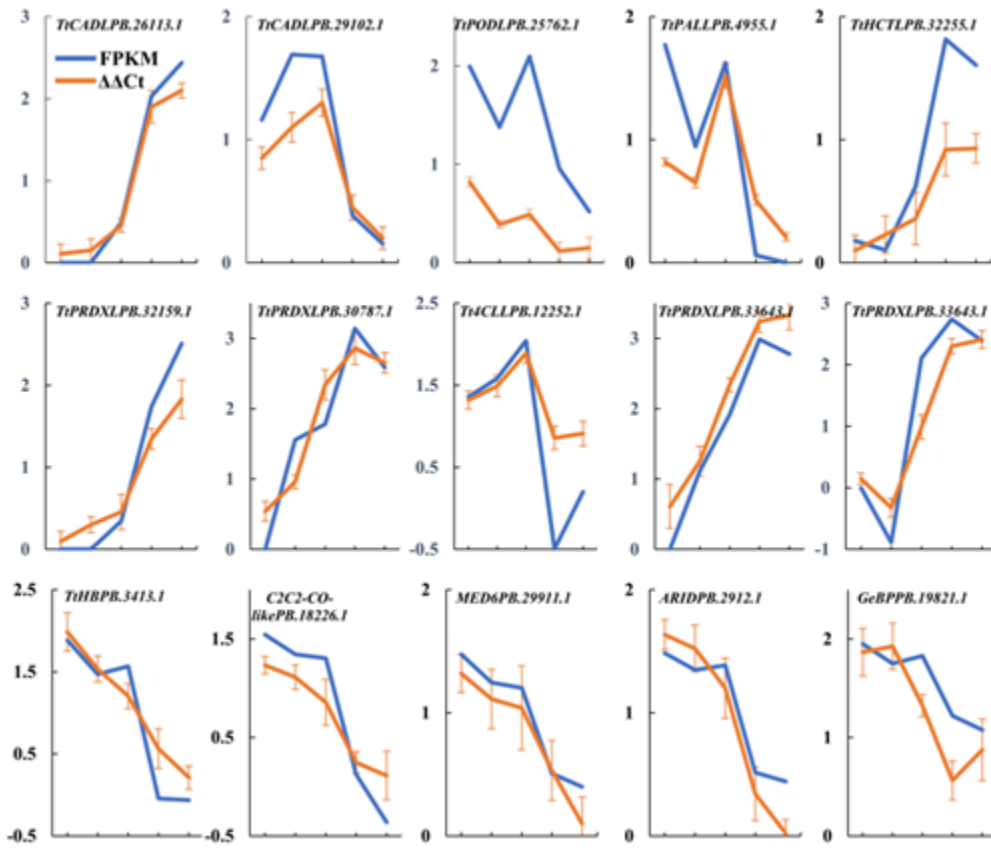


Figure 5

The validity of the transcriptome data was verified using qRT-PCR. The $\log_{10}$ (FPKM) and $2^{-\Delta\Delta C_t}$ values from S1 phase were normalised and used as controls, and the results were averaged over three biological replicates (mean $\pm$ SD, n = 3).

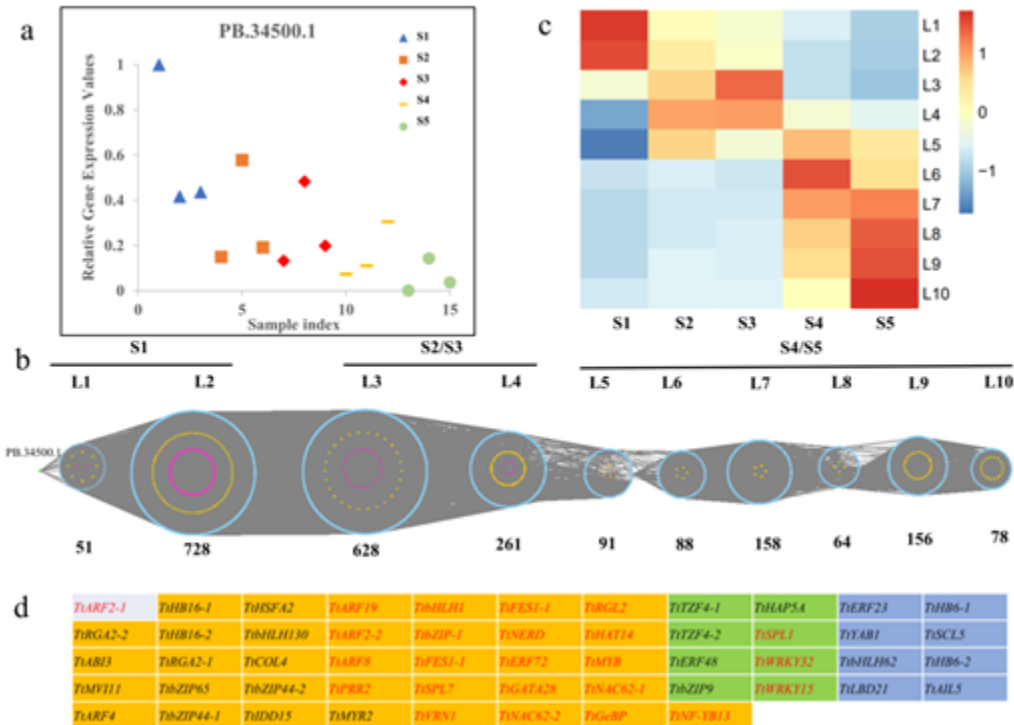


Figure 6

Time-ordered gene co-expression network (TO-GCN) during seed development of *T. tuan*. (a) Relative expression values of the bait gene (PB.34500.1). Each node represents three biological replicates. (b) TO-GCN, green nodes represent the bait gene (PB.34500.1); blue nodes represent TFs; red and yellow nodes represent flavonoid and lignin biosynthetic pathway genes, respectively. The number below each GCN represents the number of nodes. (c) Gene expression levels identified in the gene regulatory network at each time point and temporal order during seed development, blue and red represent low expression and high expression, respectively. Z-scores were standardized to -2 to 2. (d) Distribution of TFs from flavonoid and lignin biosynthetic regulatory networks in the TO-GCN network. Grey, yellow, green, and blue squares represent TFs distributed in L1, L2, L3, and L8, respectively. Red and black text represent TFs in the tan and brown modules, respectively.

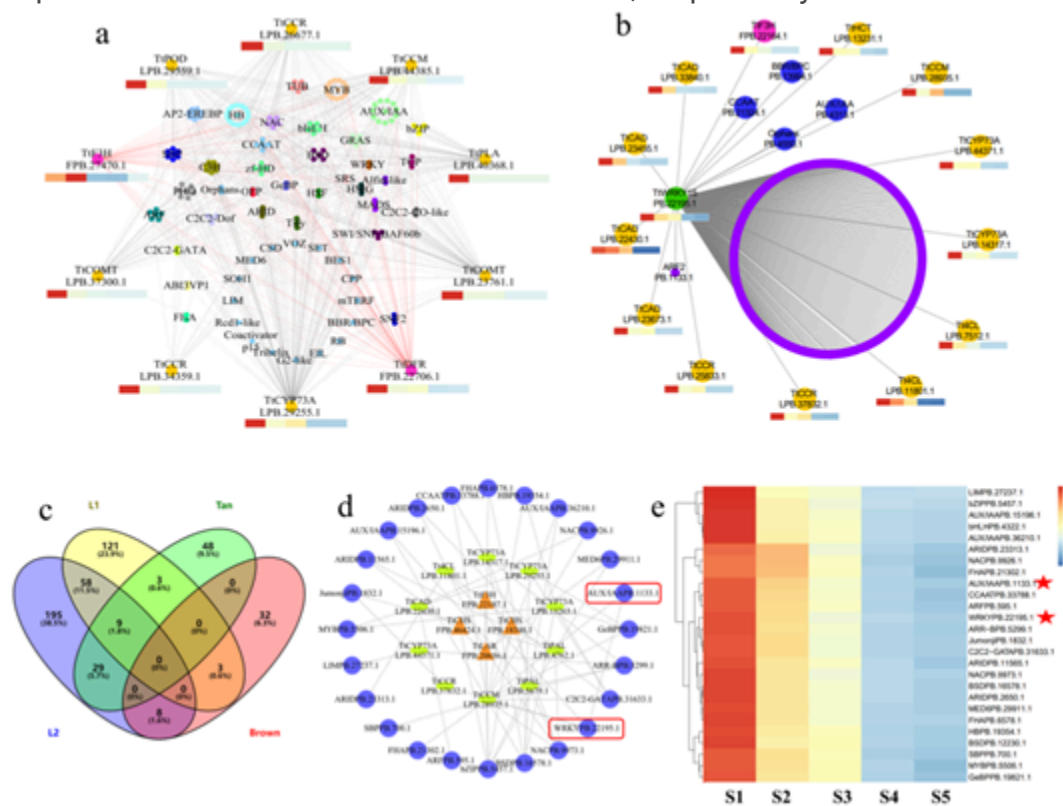


Figure 7

Analysis of the regulation relationship between key transcription factors and pathway gene. (a) Regulatory lignin and flavonoid biosynthetic subnetwork. (b) Regulatory subnetwork of *TtWRKY15* (PB.22195.1). Purple and blue circles represent TFs, and yellow and pink circles represent pathway genes. (c) Venn diagrams of all nodes among the four comparisons (L1 vs. TAN, L1 vs. Brown, L2 vs. TAN, and L2 vs. Brown). (d) Joint analysis of WGCNA and TO-GCN networks for the conserved regulatory network of lignin and flavonoid accumulation during *T. tuan* seed development. Blue circles represent TFs; orange and green nodes represent pathway genes. Lines between genes represent potential co-expression between TF and non-TF. (e) Expression levels of TFs in the conservative regulatory network

associated with lignin and flavonoid accumulation. Low to high expression levels are indicated by a blue to red color gradient. Z-scores were standardized to -2 to 2.

Supplementary Files

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

- [AppendixA.Supplementarytables.xlsx](#)
- [AppendixB.Supplementaryfigures.docx](#)