

Comparative transcriptome profiling provides insights into endocarp lignification of walnut (*Juglans Regia* L.)

Xiaoting Wu

Hebei Agricultural University <https://orcid.org/0000-0003-3133-2801>

Zechao Zhang

Hebei Agricultural University

Mintao Sun

Chinese Academy of Agricultural Sciences Institute of Vegetables and Flowers

Xiuhong An

Hebei Agricultural University

Shugang Zhao

Hebei Agricultural University

Yingbin Qi

Nanjing Agricultural University

Zhijia Zhang (✉ zzh@hebau.edu.cn)

Mountainous Areas Research Institute, Hebei Agricultural University, Baoding, 071001, Hebei, P.R. China; Research Center for Agricultural Engineering Technology of Mountain District, Baoding, 071001, Hebei, P.R. China <https://orcid.org/0000-0003-1457-9773>

Hongxia Wang

Hebei Agricultural University

Research article

Keywords: Transcriptome, Lignin, Differential expressed genes, Phenylpropanoid biosynthesis, Walnut endocarp

Posted Date: April 8th, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-20346/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 *Scientia Horticulturae* on May 1st, 2021. See the published version at <https://doi.org/10.1016/j.scienta.2021.110030>.

Abstract

Background Lignin is the main component of walnut endocarp, although we know little about the molecular mechanism of lignin formation in walnut endocarp. To understand the molecular mechanisms behind the two kinds of walnut phenotype and explore the genes involved into lignin formation, transcriptome sequencing was conducted in the walnut endocarp of the 'Zanmei' (ZM) and 'Liaoning 7' (L7) cultivars, which have different endocarp thicknesses. Compared with L7 walnut endocarp, the endocarp of ZM walnut is thicker, which decreases dehiscent nuts and compromised kernels.

Results There are more differentially expressed genes (DEGs) in the ZM walnut cultivar. The DEGs involved in the phenylpropanoid biosynthesis were significantly upregulated in both cultivars 45 days after full bloom (DAFB), but more genes were upregulated in ZM than in L7. Moreover, the same DEGs showed different expression levels in the two cultivars. Most of the key genes in ZM had more different multiples than those in L7. Interestingly, when qRT-PCR was used to determine the expression of the key genes in different development stages of the two varieties, the expression patterns were different from those known in other species. Furthermore, transcription factors regulating secondary cell wall and lignin biosynthesis were identified. Quantitative real-time PCR results were consistent with transcriptome data.

Conclusion In this study, transcriptome analysis was used to understand the molecular mechanisms of lignin formation in two walnut cultivars with different shell thickness. Several important key genes in the phenylpropanoid biosynthesis pathway were significantly different in the two cultivars, which may be the reason for the phenotypic differences. The analysis of transcription factors revealed that the regulation network in endocarp of walnut may be different from that of drupe such as apricot or peach. This study provides important candidate genes for exploring the complicated metabolic processes involved in the formation of walnut lignin.

Background

As an economically important tree, walnut (*Juglans regia* L.) is widely cultivated globally for its nuts and timber [1, 2]. Walnut endocarp plays a vital role in the growth, development, maturation, rinsing, transportation and storage of walnut, and it is also closely related to the commercial quality of the kernels [3]. In recent years, the cultivation area of thin-shell walnut in China has increased rapidly, and problems of incomplete hulls, dehiscent nuts and dark kernels have become increasingly serious in some areas. These problems are caused by shell dysplasia; without complete shell protection, walnuts are vulnerable to pests and other injuries during growth, and they are also vulnerable to contamination during transportation, cleaning or storage after harvesting, thus reducing the quality of the walnuts. The lignin content of walnut shells has been shown to be significantly positively correlated with the thickness of the shell and the tightness of the suture seal [4].

Lignin is the second most abundant plant biopolymer; it is deposited most commonly in the secondary cell walls of plants and is essential for water transport, mechanical support and plant pathogen defence

[5-7]. The lignin biosynthesis pathway has been studied by scholars for many years, and with the progress of science and technology, understanding of this pathway has undergone a major revision in the past decade. Lignin biosynthesis is energy-consuming and irreversible, and is affected by many developmental and environmental factors, such as light, circadian cycles, plant hormones, and injuries [8]. The lignin biosynthesis pathway is complicated and involves many enzymatic reactions. Great progress has been made in cloning new genes using genetic information and biochemistry [9, 10]. The biosynthesis pathway of lignin is very complex and can be divided into three steps: 1) the shikimic acid pathway: a series of processes from the assimilation of plant photosynthesis products to the synthesis of aromatic amino acids; 2) the phenylpropane pathway and lignin synthesis pathway: a natural lignin polymer generally comprised of guaiacyl (G), syringyl (S) and p-hydroxyphenyl (H) units, which are produced by three primary monolignols, coniferyl, sinapyl alcohols and p-coumaryl respectively; and 3) the polymerization of lignin monomers into lignin [11]. Recent studies have found that there may be other types of monomers [12]. Since most enzymes involved in the biosynthesis of lignin monomers have been found and documented, the monolignol biosynthesis pathway is becoming better understood [13-15]. Metabolic engineering of novel lignin is now used to improve forage crop nutritional quality and industrial processing [16].

Many genes encoding key enzymes have been separated and the functions of these genes are gradually being clarified [17-19]. In *Arabidopsis*, Rohde et al. found a significant positive correlation between the *AtPAL* and lignin content, affecting both lignin and flavonoids in the knock-out mutant without the gene [20-22]. Xu et al. inhibited *Pv4CL* in *Panicum virgatum* by RNAi and found that the expression of *4CL* in transgenic switchgrass decreased by 80%, the total lignin content and G lignin content were significantly reduced, and the biomass of the transgenic plants and fibre hydrolysis ability were significantly improved [23]. Wagner et al. inhibited the *4CL* activity of *Pinus tabulaeformis*, and found that the total lignin content of the transgenic plants decreased by 28% and the lignin type was changed, although the plant phenotype remained unchanged [24]. In hard-seeded pomegranate, *CCR* and *CAD* had higher expression levels during the critical period of seed hardening [25]. Although the genes encoding key enzymes in lignin metabolism have been isolated from many plants, the function of these genes in specific species cannot be fully inferred from other species. For example, the genes involved in lignin metabolism vary in different species, and the same gene may have different effects. The removal of *4CL* from *Arabidopsis* had no significant effect on lignin content or monomer composition [26]. There remains much to be learned about lignin metabolism; the transcriptional regulation network of lignin synthesis is also being explored and revised. Related gene family members such as MYB and NAC continue to be discovered in various species. Zhao constructed a transcriptional network model of secondary cell wall biosynthesis in *Arabidopsis*, and he found that the whole secondary cell wall biosynthesis can be activated by either *NST1/2/3* or *MYB46/MYB83* [27]. Chen found that *MYB46* in apples altered plant resistance by affecting lignin content [28].

In recent years, transcriptome sequencing technology has been widely used in plants. Zhang et al. [29] used transcriptome technology to analyse key genes in apricot endocarp and identify the molecular mechanisms in 'Liehe' apricot, the endocarp-cleaving phenotype. Zhang [30] used transcriptome

technology to analyse the sequence of cotton and screened out candidate genes for lignin biosynthesis, which provided a new mechanism for cotton lignin biosynthesis and laid a foundation for further improving the fibre quality. With the help of transcriptome sequencing, more comprehensive understanding of the biological system can be achieved. Using transcriptome sequencing technology to reveal the molecular mechanism of lignin synthesis in the endocarp during development can improve walnut quality and lay a foundation for walnut breeding.

Research on walnut cultivars mainly focuses on the study of the fat, protein, fatty acid and other nutrients of seed kernel [31, 32], or the flavonoids, phenolic acids and quinones in the shoots [33], but few studies have focused on the development of the hard shell, which plays a vital role in the health of the seeds. Yu [34] et al. used transcriptome sequencing technology to screen the differentially expressed genes in the lignification stage of the endocarp of 'Zhipi' walnut and predicted their functions. However, they only studied the lignification stage, and did not explain the differences in different growing stages or varieties. Because of good taste and rich nutrition, 'Zanmei' and 'Liaoning 7' walnut are widely cultivated in China. However, the endocarp in 'Liaoning 7' is relatively thin, which often results in split and exposed kernels, seriously affecting the quality of the nuts. In this study, the mechanisms of endocarp development and phenotype formation were investigated through analysis of transcriptome sequencing and the expression patterns of candidate genes in walnut endocarp development and lignification. It is of great significance to clarify the molecular mechanism of lignin formation to achieve artificial regulation to improve walnut quality and enrich germplasm resources.

Results

Morphological characteristics of walnut endocarp

To identify the critical lignification periods during development of young fruit, the timing and pattern of lignin deposition was studied. Within 30 DAFB, the pericarp was clearly stratified, and the endocarp appeared milky white and the Wiesner reaction revealed that no lignification occurred during this period (Fig.1 A, C, E, G). Only the vascular tissue at the junction of the mesocarp and endocarp was stained. At 45 DAFB, the endocarp near the outer pericarp began to lignify. The Wiesner reaction showed that lignification gradually occurred out from the outside to the inside of the endocarp (Fig.1 B, D, F, H). The apex (blossom end) and the junction of the endocarp and mesocarp were most stained. Pink gradually diffused inward from the junction and became shallow. Little or no staining was observed in tissues other than the endocarp with the exception of scattered small vascular elements in the mesocarp, trichomes present on the fruit surface and residual floral structure. Seed kernels were still water-like at this time. The key point at which the lignification of the endocarp begins to develop is between 30 and 45 DAFB, suggesting that the lignification of the endocarp begins relatively early in the development of the fruit. To ensure there was no lignin deposition in the whole endocarp, the endocarp of ZM and L7 of 20 DAFB was collected for transcriptome sequencing, when the endocarp and mesocarp had differentiated, ensuring the accuracy of sampling. Forty-five DAFB is a time of rapid lignin formation in the endocarp. Therefore, samples of 45 DAFB were also collected for transcriptome sequencing. As seen from the fresh fruit

pictures in each dimension, the mesocarp of ZM is thicker and oval in shape, while the mesocarp of L7 is thinner and round in shape. The difference in the endocarp between the two cultivars, however, was not obvious.

Transcriptome sequencing and quality control

In total, 501,163,896 raw reads were obtained from the transcriptome sequencing of all samples, and 423,289,427 clean reads were mapped to the walnut reference genome ([https://www.ncbi.nlm.nih.gov/genome/?term=txid51240\[orgn\]](https://www.ncbi.nlm.nih.gov/genome/?term=txid51240[orgn])) after filtering out the low-quality reads (Table 1). The results of the mapping rate, Q30 rate and GC content suggested the high quality of our data.

Table 1 Data output quality list

Sample name	Clean reads number	HQ clean reads number	HQ clean data(bp)	Q20(%)	Q30(%)	GC(%)	Total mapped	Multiple mapped	Uniquely mapped
T-ZM-1	35615318	35374570	5291868071	96.13	90.61	45.78	35285776(85.50%)	1264394 (3.58%)	28905769 (81.92%)
T-ZM-2	25050350	24890318	3723545540	96.20	90.69	45.76	24773372(85.70%)	888828 (3.59%)	20341281 (82.11%)
T-ZM-3	27408692	27205416	4068797895	96.00	90.30	45.88	27113566(85.18%)	964252 (3.56%)	22131934 (81.63%)
F-ZM-1	28753730	28587726	4277543832	96.39	91.09	45.32	28516602(86.08%)	1134118 (3.98%)	23413987 (82.11%)
F-ZM-2	24711764	24530414	3671111290	96.01	90.33	45.42	24472276(85.18%)	932310 (3.81%)	19911981 (81.37%)
F-ZM-3	26865506	26702752	3995147175	96.22	90.71	45.33	26633542(85.61%)	1045948 (3.93%)	21754542 (81.68%)
T-L7-1	25572784	25419584	3801704427	96.29	90.90	45.83	25336008(86.08%)	897638 (3.54%)	20910521 (82.53%)
T-L7-2	30463302	30261640	4525656120	96.10	90.52	46.05	30188378(85.55%)	1061016 (3.51%)	24764898 (82.03%)
T-L7-3	22644096	22460148	3359569046	95.76	89.82	45.76	22395556(84.57%)	755580 (3.37%)	18184535 (81.20%)
F-L7-1	26090952	25933396	3878988211	96.18	90.73	45.48	25875950(85.23%)	994914 (3.84%)	21058716 (81.38%)
F-L7-2	27939918	27764000	4153061801	96.03	90.35	45.55	27676730(85.16%)	1065578 (3.85%)	22503546 (81.31%)
F-L7-3	24382998	24230820	3625142180	96.28	90.96	45.39	24122168(85.48%)	952856 (3.95%)	19665864 (81.53%)

Clean reads: the reads removing low-quality reads. HQ clean reads: the reads from the second filter. Q20 and Q30: the mistake rate lower than 1% and 0.1%, respectively. GC content: the GC ratio of total base number. Total mapped: the number of reads which can be mapped to the reference genome. Multiple mapped: the number of reads with multiple alignment positions on the reference sequence. Uniquely mapped: the number of reads with unique alignment positions on the reference sequence.

Supplementary Fig. S1 showed the Pearson's correlation coefficient results of all groups. The correlation of each group was higher than 0.95, suggesting the high quality of the sequencing data. PCA is largely used to show the structure and relationship of the samples; the results of PCA in this study showed good

correlation in each group and the difference between the two periods in each cultivar (Supplementary Fig. S2).

Analysis of genes between groups

There were 36,861 reference genes in total, 30,276 of which were known. The new gene number of all the samples through sequencing was 1,870. The expressed gene number of each group was clearly illustrated in Table 2.

Table 2 Gene statistics for each group

Group name	Known gene number	New gene number	All gene number
T-ZM	27494 (74.59%)	1762	29256
F-ZM	27663 (75.05%)	1774	29437
T-L7	26899 (72.97%)	1744	28643
F-L7	27453 (74.48%)	1768	29221

To identify the overall difference, differential expression analysis was performed for the different periods and cultivars. The (T-ZM vs F-ZM) and (T-L7 vs F-L7) groups revealed the shells' DEGs in different developmental stages and showed that they were related to lignin biosynthesis, which had the most DEGs. As a result, 3502 and 3663 genes were individually up- and downregulated between T-ZM and F-ZM while there were 3418 and 2727 genes between T-L7 and F-L7 (Fig. 2a). Compared with L7, ZM had more DEGs during the lignin formation period. In the (T-ZM vs F-ZM) group, there were more downregulated genes than upregulated genes. But in the (T-L7 vs F-L7) group, there were more upregulated genes than downregulated genes. The two groups had 4,537 DEGs in common (Fig. 2b) from which the same regulatory genes controlling the growth and development of the two varieties can be found. Apart from these common differences, the (T-ZM vs F-ZM) and (T-L7 vs F-L7) group had 2,628 and 1,608 genes, respectively, which may hold the key to the differences between the two varieties. The (T-L7 vs T-ZM) and (F-L7 vs F-ZM) groups showed the shells' DEGs in different cultivars at the same development stages, in which 1672 and 881 genes were individually up and downregulated between T-L7 and T-ZM, respectively, and 1,233 and 1,025 genes were up and downregulated between F-L7 and F-ZM. Compared with L7, ZM had more upregulated genes in each period. The (T-ZM vs F-ZM), (T-L7 vs F-L7) and (F-L7 vs F-ZM) groups had 516 common DEGs, including DEGs that play a role in the different varieties and at different stages of development. Since there was no lignin at 20 DAFB, we did not compare the (T-L7 and T-ZM) group with the other groups. The above results showed that the number of DEGs in different developmental stages of ZM was higher than that of L7, indicating that the developmental regulation of ZM was more complicated than that of L7.

Function prediction

To learn the potential functions of the DEGs between the different groups, GO (<http://www.geneontology.org/>) and KEGG (<https://www.genome.jp/>) enrichments were conducted for functional classification. GO analysis showed that all DEGs can be classified into biological processes

(BP), cell components (CC) and molecular functions (MF) and each of them contains several exact functions and GO items (Fig. 3).

The overall GO analysis showed that the differential genes for CC were mainly concentrated in the cell parts, organelles and membranes, the MF included catalytic activity and binding; and the BP were metabolic processes, cellular processes and single-organism processes, etc. The proportion of upregulated genes was significant in (T-L7 vs T-ZM) and (F-L7 vs F-ZM) groups, while the proportion of downregulated genes was significant in the other two groups, and the total number of DEGs was much larger. In comparing the (T-ZM vs F-ZM) and (T-L7 vs F-L7) groups, the number of downregulated genes of ZM in each component was significantly higher than that of L7, while the upregulated genes were not significantly different. After $FDR < 0.05$ (adjusted p value) condition screening and refinement classification, further functional analysis could be performed (Supplementary Table S1). Except for the (F-L7 vs F-ZM) group, the lignin metabolic process pathway and phenylpropanoid metabolic processes were significantly enriched in the rest of the groups. The lignin metabolic process pathway was the most significant in the (F-L7 vs F-ZM) group. No lignin was found in the endocarp at 30 DAFB (Fig. 1), nor 20 DAFB. The enrichment of the lignin metabolic process pathway and phenylpropanoid metabolic process in the (T-L7 and T-ZM) group may have been due to the growth and development of cell walls in the endocarp. It may also have been due to the development and formation of vascular tissue. There were more different pathways in the (T-ZM vs F-ZM) and (T-L7 vs F-L7) groups, and the cell wall related pathways and DEGs were also significantly increased. Secondary metabolic processes, responses to hormone and hormones transport were also significantly enriched. A large number of hormones regulated and participated in regulating the changes in the different developmental stages, indicating that the growth and development at this time were due to secondary metabolism.

The top 20 KEGG pathway enrichment analyses of four groups are shown in Fig. 4. In the (T-ZM vs F-ZM) and (T-L7 vs F-L7) groups, most DEGs were enriched in plant hormone signal transduction, phenylpropanoid biosynthesis, tryptophan metabolism and starch and sucrose metabolism (Fig. 4A, B). A large number of hormones were involved in the transformation of the endocarp from growth to secondary metabolism. The differences in these metabolic pathways were consistent with the results in GO enrichment. In the (T-ZM vs F-ZM) group, fatty acid elongation and DNA replication were also significantly enriched. In the (T-L7 vs F-L7) group, photosynthesis, ascorbate and aldarate metabolism, thiamine metabolism and ABC transporters were significantly enriched. These differences might explain the developmental status and nutrient supply differences between the two varieties. Glutathione metabolism and taurine and hypotaurine metabolism showed significant differences in the (F-L7 vs F-ZM) group, while DNA replication, glutathione metabolism and purine metabolism showed significant differences in the (T-L7 and T-ZM) group (Fig. 4C, D).

Expression analysis of DEGs related to lignin biosynthesis and validation of DEG expression by RT-qPCR

The continuous deposition of lignin in walnut endocarp and its eventual lignification is complicated. A series of biological processes are involved, including the synergistic action of many genes to regulate

various pathways, of which the phenylpropanoid biosynthesis pathway is one of the most important.

The relative expression of the DEGs in the (T-ZM vs F-ZM) and (T-L7 vs F-L7) groups in the phenylpropanoid biosynthesis pathway are intuitively showed in Fig.5. The expression of the genes corresponding to the key enzymes in the metabolic pathway at different stages of ZM and L7 are also visually illustrated. In total, seventy-six DEGs were annotated in this pathway in ZM and forty DEGs were highly expressed, including, *4CL* (6), *CAD* (3), *CCR* (1), *C4H* (2), *HCT* (2), *C3'H* (2), *COMT* (2), *CCoAOMT* (1), *F5H* (2), *CSE* (1) and *POD* (18). Most of the genes were upregulated and the most downregulated gene was *POD*. A total of sixty-five DEGs in L7 were enriched in this pathway, of which thirty-two were highly expressed, including *PAL* (1), *4CL* (5), *CAD* (3), *CCR* (1), *C4H* (2), *HCT* (1), *C3'H* (2), *COMT* (1), *CCoAOMT* (1), *F5H* (1) and *POD* (14). As in the case of ZM, most of the DEGs in L7 were upregulated, although there were fewer DEGs than in ZM. The expression of most of the DEGs was consistent in the two varieties, and the differences were mainly concentrated in *PAL*, *HCT*, *POD*, *CSE*, etc. *PAL* is located at the entrance of the phenylpropanoid pathway and is involved not only in the synthesis of lignin monomers but also in the synthesis of other phenolic substances [34]. Therefore, there was no expression difference between the two periods in ZM. *PAL*, *4CL* and *C4H* often form multi-enzyme complexes to efficiently regulate the phenylpropanoid metabolism pathway [35]. Both *4CL* and *C4H* showed an upregulated trends. *HCT* controls the biosynthesis of lignin monomers and is an important control point for the biological conversion of H monomer and G/S monomer [36, 37]. During the period of lignin synthesis, the downregulation of the gene expression might be related to the regulation of lignin monomer type. *POD* enriched the most DEGs in both ZM and L7. *POD* is a key enzyme involved in the last step of lignin synthesis, which catalyses the dehydrogenation of the monomers metabolized by phenylpropane to form lignin. *POD* is encoded by multiple gene families. Due to the differences in gene structure and function, *POD* also participates in a variety of metabolic pathways, such as auxin metabolism, cell wall extension and thickening, metabolism of reactive oxygen species and reactive nitrogen, and various resistance processes of plants [38, 39]. *CSE* has been demonstrated to have the ability to catalyse the conversion of caffeoyl shikimate into caffeate in *Arabidopsis thaliana* [40], and Wout [41] confirmed its role in poplar lignin synthesis. However, due to a lack of further evidence, its function has not been fully explored.

The transcriptome results narrowed the scope for further screening. RT-qPCR was performed on some identified genes related to the lignin pathway. The specific gene primers involved in the RT-qPCR analysis and corresponding gene information are listed in Supplementary Table S2. Eighteen genes encoding key enzymes in the phenylpropanoid biosynthesis pathway were selected. Flavonoids are important secondary metabolites, and involved in plant growth and development, disease resistance and other important processes [42]. The synthesis of flavonoids is derived from the phenylpropanoid pathway, of which chalcone synthase (*CHS*) is the first step. Thus, a gene encoding *CHS* was also being examined.

The results showed that the expression profile of the pathway related to DEGs determined by RT-qPCR was consistent with the corresponding \log_2 (F/T) value of the RNA-seq analysis (Table 3).

The expression level of genes involved in the phenylpropanoid pathway was downregulated in L7 compared with the ZM cultivar in both the development stages and lignin synthesis. The two *COMT* genes showed the same expression pattern in ZM and L7. They were highly expressed at the growth stage in ZM, and they continuously increased in the lignin formation stage until they reached the peak value of 65DAFB. This gene showed the same change trend in L7, but the difference multiple was much lower than that of ZM (Fig.6). Except for *CCR*, the expression levels of other genes in the phenylpropanoid biosynthesis pathway were all greater in ZM than in L7. *C3'H*(gene11203) showed the same expression trend in ZM and L7, but the peak value of L7 appeared later (70DAFB). *C3'H*(gene21068) was always highly expressed in ZM, reaching its maximum value at 65DAFB, which was much higher than that in L7. *HCT* (gene 11284), *CAD* (gene39775), and *C4H*(gene40342) had a similar situation with *C3'H* (gene21068). *CCR1* was highly expressed at the beginning of lignin formation in L7 and reached its peak at 50 DAFB. The trend of *CCoAOMT* (gene 32982) was similar. The expression levels of *POD* varied greatly. The large differences in these key genes may be the key reason for the differences in endocarp lignification between the two varieties.

Table 3 Expression trends of 18 genes by real-time PCR and sequencing

Gene ID	F-ZM	F-L7	Note
gene10640	1.2	1.34	RNA-seq
	9.26	1.13	qPCR
gene11284	1.66	0.98	RNA-seq
	2.48	2.61	qPCR
gene12497	-3.43	-2.93	RNA-seq
	0.12	0.11	qPCR
gene12948	2.04	1.96	RNA-seq
	2.87	4.85	qPCR
gene16314	-3.38	-2.24	RNA-seq
	0.53	0.32	qPCR
gene21068	2.36	1.58	RNA-seq
	24.54	9.92	qPCR
gene22810	3.99	4.08	RNA-seq
	2.42	2.39	qPCR
gene32982	4.79	5.97	RNA-seq
	7.10	7.08	qPCR
gene35880	8.21	8.92	RNA-seq
	8039.50	256.15	qPCR
gene37179	6.99	7.01	RNA-seq
	29.61	28.32	qPCR
gene39775	1.88	1.93	RNA-seq
	1.54	1.38	qPCR
gene40342	1.27	1.33	RNA-seq
	1.68	1.26	qPCR
gene40352	1.02	0.82	RNA-seq
	11.22	1.26	qPCR
gene42443	9.70	8.13	RNA-seq
	82.97	629.17	qPCR
gene8744	-1.95	-0.49	RNA-seq
	0.15	0.13	qPCR
gene35863	-1.72	-1.03	RNA-seq
	0.81	0.45	qPCR
Gene11203	2.54	1.95	RNA-seq
	0.10	0.25	qPCR
gene32581	-1.63	-0.86	RNA-seq
	0.15	0.01	qPCR

Note: The data in RNA-seq is the difference multiple of $\log_2(F/T)$ value, and the data in qPCR is the relative expression of $2^{-\Delta\Delta CT}$ value. The endocarp of 20 DAFB was used as control group.

According to the previous studies, the lignin biosynthetic genes are generally regulated by different families of transcription factors, such as the MYB, NAC, LBD and NST families etc. [27, 43]. We made a heat map analysis of all the screened genes in these families (Supplementary Figure S3). The specific gene information is shown in the Supplementary Table S3. Under the given conditions (FPKM > 5, $|\log_2 F/T| > 1$), the lignin-related transcription factors concerned in some studies were not screened out, such as *MYB46* and *MYB83*. *NST1* and *LBD41* were highly expressed in the lignin generation stage of both varieties, but they were significantly higher in ZM than in L7, which is note worthy. Several

transcription factors whose differential multiples did not meet the criteria were also screened because their expression might be involved in lignin metabolism. Most transcription factors were upregulated during lignin formation. Several groups of different transcription factor families had similar expression patterns, which were related to the lignin formation law. *NAC100* and *NAC12* had significant high expression in F-L7.

Discussion

It takes approximately 130 days from siphonogamy to ripening for walnut, which can be roughly divided into three stages, including rapid growth, shell hardening and ripening. The period of 35-40 days after fertilization is a rapid growth period, and the next 35-40 days is the shell hardening stage. During the shell hardening stage, the endocarp of the walnut gradually becomes lignified, forming a sclereid layer, which is the endocarp of the nut. Studies have shown that lignin is the most important component of walnut endocarp [44]. In lignified cells, lignin firstly forms in cell corners, then synthesizes between cells and gradually becomes primary and secondary cell walls [45]. With the further development of mature lignification of plant tissues, lignin deposits inside and outside cell walls and connects cells [9]. Shell structures are formed in the cell walls of plants due to the deposition of lignin [46]. The content and composition of lignin vary not only between plant taxa but also between different cell types within the same tissue and even within the same cell wall [47]. Dardick believes that the lignin in peach endocarp begins to deposit in the parenchyma layer radiating out from the flower end [48], while Zhang believes that the lignin deposition process of apricot begins at the top of the endocarp [29]. Walnut fruits, however, do not belong to the true drupe, as they develop from bracts, perianth and ovaries. Wen et al. believed that the deposition of lignin in walnut endocarp begins at the top corner [49], while Yu et al. believed that the lignification of 'Zhipi' walnut endocarp might occur almost simultaneously throughout the entire endocarp [34]. In this study, dyed walnut samples were collected 30 and 45 days after full bloom, since lignin surely began to form between these two points. Unfortunately, due to the large time span, the exact time of lignin formation cannot be determined in this study. The vertex, junction and vascular bundle were stained deeply. However, it was not sufficiently accurate to judge the position of lignin only by the depth of colour. Dense sampling can be conducted near the key points of lignin synthesis, and the application of microsection technology is conducive to observing the initial site of lignin synthesis more directly. Approximately 45 days after flowering is a period of rapid lignin production in the endocarp, but it is not the turning point. Therefore, transcriptome sequencing was performed on 20 DAFB, a time when there was no lignin at all, and on 45 DAFB, a time at which lignin is produced rapidly.

Lignin can be used as a source of pulp, animal feed, or as fuel resource, and it serves as an important supporting and protective material in plants. Fruit endocarp lignification occurs in many angiosperms and plays an important role in seed dispersal and protection. Lignified endocarp can protect seed kernels from foodborne pathogens and is also an important symbol of fruit ripening [50]. In recent years, increasing attention has been paid to it. As the two main walnut cultivars in China, ZM and L7 differ significantly in their development of endocarp. The endocarp of ZM is thick and hard after ripening, while the endocarp of L7 is thin and brittle after ripening. Phenylpropanoid biosynthesis is vital in endocarp

lignification in both ZM and L7. Analysis of transcriptome sequencing revealed a series of differentially expressed genes involved in the phenylpropanoid pathway, such as *C3'H*, *CCoAOMT*, *COMT*, *HCT*, *CCR*, *POD*, and *CAD*. Among these genes, the genes that differ between the two cultivars at different periods may better reveal the molecular mechanism of lignin synthesis in walnut endocarp, and their different expression patterns may be the reason for the phenotypic differences between the two cultivars. We therefore measured the expression of these DEGs over more time periods. *CCoAOMT* and *COMT* are methylases at different substrate levels in the lignin biosynthesis pathway, that form two parallel pathways in the lignin biosynthesis pathway [51]. A previous study found that a reduction in *COMT* reduced the lignin content in transgenic alfalfa and affected the structure of S-type lignin [52]. After inhibiting *CCoAOMT* expression in transgenic tobacco, lignin content decreased significantly [53]. *COMT* in pomegranate endocarp was highly expressed in the early stage of development, and the expression in soft seeds was significantly higher than that in hard seeds at the same time. The *CCoAOMT* was also highly expressed at the beginning of lignin biosynthesis and then gradually decreased [25]. The expression of *COMT* in apricot endocarp also decreased gradually from the initial high expression [29]. The qRT-PCR results in this study demonstrated that the expression level of *COMT* and *CCoAOMT* in the two varieties was similar at the early stage of development, but the expression level of *COMT* in ZM was significantly higher than that in L7 at the middle and late stages of lignin formation. When T-DNA was inserted into *C3'H* in *Arabidopsis thaliana*, the mutant plant was severely stunted in growth and development, and the lignin content was reduced by 20%-40% compared with the wild type, while the G-type and S-type monolignol content was significantly decreased [54]. The xylem development and lignin deposition mode of *HCT1*-*rnai* transgenic poplar were changed, and the wall of vessel and wood fibre in the xylem decreased [55]. The relationship between *HCT* and *C3'H* also affected the proportion of different lignin monomers [56-59]. In this study, both *C3'H* and *HCT* in ZM were significantly highly expressed at 65 days after flowering, while only some of the expression levels in L7 were consistent with those in ZM. When there was no lignin at the early stage of growth and development, *C3'H* (*gene21068*), *HCT* (*gene11284*) and *POD72* (*gene16314*) were highly expressed in both varieties. Whether the high expression of these genes was due to the development of vascular tissue, preparation for subsequent lignin synthesis, or the need for cell proliferation is uncertain. *C4H*, which catalyses the second step in the phenylpropanoid pathway, showed high expression at the stage of lignin biosynthesis in ZM, while L7 had high expression at 85 DAFB. *CAD* and *COMT* had similar expression patterns. The expression of *CCR* in hard seed pomegranate was significantly higher than that in soft seed [25], but *CCR* was significantly lower in ZM during the lignin formation period. Several *POD* showed different expression patterns in different varieties at different periods. These results suggested that difference in endocarp thickness cannot be the result of one specific phenylpropanoid pathway gene. These genes in the phenylpropanoid pathway may be responsible for the differences in endocarp development and lignification in the two cultivars. The monolignol biosynthesis pathway is still being revised as new genes and regulation mechanism continue to be discovered [43].

The synthesis of flavonoids shares the same substrate, p-coumaroyl-CoA, with lignin (Fig.5). There is no doubt that the flavonoid synthesis pathway competes with the lignin pathway, and the coordination of

this process may be the key to controlling some important agronomic traits in fruits and nuts. CHS is the first key enzyme in plants that shifts from p-coumaroyl-CoA in the phenylpropanoid metabolic pathway to flavonoids biosynthesis [60, 61]. We speculate that the relationship between *CHS* and *HCT* may determine the direction of the two major metabolic pathways. The expression of *CHS2* increased in the late period in L7 and *HCT* decreased. *CHS2* in ZM did not show the same trend. Perhaps the competition between the flavonoid pathway and the phenylpropanoid pathway does not start when sufficient substrates are accumulated in the metabolic pathway in the beginning. Maybe when the substrate is insufficient, the ratio of *CHS* to *HCT* determines the next flow direction. The induction of the lignin pathway is variable and the induction of flavonoids is conservative in Rosaceae, which may also be the case in walnuts. There may be a common transcription factor upstream of these two enzymatic control genes that regulates at the overall level of the plant. This conjecture still requires further experiments to verify. The competition model, however, would be more complicated. These expression data reveal a complex relationship between lignin and flavonoid pathways regulating the development of walnut fruit. This may be partly responsible for the difference in endocarp between the two varieties.

SHP and *NST* are transcription factors that mediate the development of the lignified endocarp layer (enb) in *Arabidopsis*. *NST1* promotes enb lignification after tissue identity has been established [62]. The expression of *SHP* and *STK* decreased before the lignin deposition, and then induced the expression of *NST1* in peaches [48]. *NST1* was found in DEGs, but *SHP* and *STK* were not found in our analysis. The expression of *NST1* in the lignin generation stage was higher than that in the early stage. Walnuts differ significantly from *Arabidopsis* and peach. In *Arabidopsis thaliana*, *MYB46* and *MYB83* are considered to be key regulators of lignin, while *NST1* is the switch that regulates *MYB46* and *MYB83* [27]. However, in our transcriptome data, both *MYB46* and *MYB83* were filtered out due to low FPKM values. We screened out *MYB4* and *MYB7* downstream of *MYB46* and *MYB83* regulation. *AtMYB4*, a lignin inhibitor, was downregulated in both the wounded leaves and in systemically non-wounded leaves of the same rosette. Therefore, we speculate that the reduction in *AtMYB4* levels upregulate the monolignol pathway, enabling lignin to be produced in damaged tissues as a defense strategy. However *MYB4* was highly expressed in the lignin synthesis stage in this study and the expression level was higher in L7, which has a thinner shell. NAC transcription factors can control the formation of secondary walls in woody tissues. The positive feedback regulation network formed by LBD and NAC protein regulates the secondary growth of *Arabidopsis* by controlling the differentiation of xylem cells [63]. The *LBD* family has several biological functions such as plant development, secondary growth, pathogen response, biological stress and hormone signaling [64-68]. *LBD25* can control the growth of the hypocotyl of *Arabidopsis* by regulating the morphogenesis of light, resulting in the failure of the hypocotyl of *LBD25* mutant under dark culture conditions [69]. In the transcriptome sequencing results, *LBD41* was highly expressed in the lignin synthesis stage and the value in F-ZM was significantly higher than in L7. In the KEGG enrichment pathway, hormone signal transduction enriched the most differentially expressed genes, so it can be seen that the initiation and development of endocarp lignification is necessary to mobilize a large number of hormones, and the significant changes in *LBD41* may be related to this. *LBD25* was highly expressed at 20DAFB, during the growth and development stage of walnut. *AtLBD6/AtAS2* downregulates the

expression of *KNOX*, thus participating in leaf morphogenesis [69]. In our results, *LBD6* was highly expressed at 45 DAFB, which may be related to the location of lignin deposition. Walnut endocarp showed some ways of transcriptional regulation that were different from those of *Arabidopsis* regulatory network, and perhaps the function of these genes was also different from the existing known functions; hence the regulation of walnut endocarp still requires further study. The lignification of walnut endocarp is a complex process under the combined action of hormones, enzymes, related genes and regulatory networks.

Conclusions

The results showed that the lignification of the endocarp in both of ZM and L7 walnut started between 30 DAFB and 45 DAFB, the early stage of fruit growth and development. Transcriptome sequencing results and analysis showed that there were differences in endocarp development and lignification between ZM and L7 cultivar. The analysis of sequencing and qPCR identified DEGs involved in phenylpropanoid pathway including *C3'H*, *CCoAOMT*, *COMT*, *HCT*, *CCR*, *POD*, and *CAD*. The different expression patterns of these key genes in the phenylpropanoid pathway may be responsible for the differences in endocarp phenotypes in the two cultivars. The analysis of transcription factors revealed that the regulation network in endocarp of walnut may be different from that of drupe such as apricot or peach. Meanwhile some transcription factors, such as MYB, NAC and LBD, were also significantly different in the two varieties, which may be the reason of discrepancies in endocarp development and lignification of ZM and L7 walnut.

Methods

Plant materials

'Zanmei' walnut was selected from Taihang Mountain population by Hebei Agricultural University and collected in the experimental base of Hebei Agricultural University in Zanhuang County, Hebei Province, China (37°69' N, 114°43'E). The identification of 'Zanmei' cultivar was done by the Hebei Province forest variety Examination and Approval Committee in 2012, with the accession number S-SV-JR-006-2011[70]. 'Liaoning 7' walnut was the hybrid of Xinjiang 'Zhipi' walnut 21102 × Liaoning 'Dama' walnut, which was introduced into Hebei province from Liaoning Economic Forest Research Institute by Hebei Forestry Research Institute in 1995. The identification of 'Liaoning 7' cultivar was done by the Hebei Province forest variety Examination and Approval Committee in 2005, with the accession number S-ETS-JR-010-2005. It was introduced in experimental base of Hebei Agricultural University through the project "The promotion of Liaoning 7 and other three excellent walnut varieties" by Hebei Forestry Research Institute[71, 72]. The fruit of 'Zanmei' (ZM) and 'Liaoning 7' (L7) used in this study were picked in experimental base of Hebei Agricultural University from 20 DAFB to 100 DAFB with the permission of the person in charge of the base, Mr. Zhigui Bai. The materials at 20 DAFB (T) and 45 DAFB (F) were used for transcriptome sequencing. ZM had a thicker shell than L7; the shell's average thickness of ZM was 1.30mm, the average shell thickness in L7 was 0.97mm. Walnut endocarp tissue was mashed and mixed

in liquid N then stored at -80°C for follow-up experiments. Three biological replicates of walnut endocarp were collected in each treatment and some samples were collected for slice observation.

Specimen preparation for light microscopy

The endocarp was cut into cubes of approximately 5 mm^3 and fixed with FAA. The tissues were dehydrated with 70%, 85%, 95% and 100% ethanol at room temperature (1 hour for each step). Gradient vitrification was performed from 100% ethanol to 100% xylene; then the samples were permeated and embedded in paraffin. Sections of $8\text{ }\mu\text{m}$ each were then made using a microtome (SLEE-CUT4062) and subsequently mounted on adhesive microscope slides that were gradually dewaxed to expose the pure tissue. The sections were stained with 1% phloroglucinol-HCl and then were quickly observed under a microscope. Phloroglucinol-HCl staining is the fastest and simplest method to detect the presence and content of lignin. Under acidic conditions, lignin is dyed pink by phloroglucinol, and the higher the lignin content, the darker the colour is.

Sequencing library construction

Samples in this part

The endocarp of ZM and L7 samples at 20 DAFB (T) and 45 DAFB (F) were used for transcriptome sequencing. Samples were named T-ZM, F-ZM, T-L7 and F-L7 (shown in Table 4).

Table 4 List of samples

Material name	20 DAFB	45 DAFB
ZM	T-ZM	F-ZM
L7	T-L7	F-L7

RNA isolation and qualification

Total RNA was extracted from the samples using RNAPrep Pure Plant Plus Kit (Tiangen Biotech, Beijing, China). RNA degradation and contamination were monitored on 1% agarose gels and RNA purity and concentration were measured using the BIOSPEC-NANO[®] (230V) spectrophotometer (SHIMADZU). Three biological replicates were used for each sample. After total RNA was extracted, mRNA was enriched using Oligo (dT) beads.

cDNA synthesis and amplification

The enriched mRNA was fragmented into short fragments using fragmentation buffer and reverse transcribed into first strand of cDNA with random primers using cDNA first strand synthesis kit (Takara, Dalian, China). Second-strand cDNA were synthesized by DNA polymerase I, RNase H, dNTP and buffer.

Then the cDNA fragments were purified with QiaQuick PCR Purification kit (Qiagen, Shanghai, China), end repaired, poly (A) added, and ligated to Illumina sequencing adapters. The ligation products were size selected by agarose gel electrophoresis, PCR amplified, and sequenced using Illumina HiSeq 2500 by Gene Denovo Biotechnology Co. (Guangzhou, China).

Sequencing data analysis

Reads obtained from the sequencing machines included raw reads containing adapters or low-quality bases which should be filtered out to get high quality clean reads. The rRNA removed reads of each sample were then mapped to the reference genome by TopHat2 (version 2.0.3.12). The reconstruction of the transcripts was performed using Cufflinks software, which together with TopHat2, allows biologists to identify new genes and new splice variants of known genes. To identify the new gene transcripts, all of the reconstructed transcripts were aligned to the reference genome and were divided into twelve categories by using Cuff compare. Gene abundances were quantified by RSEM software. The gene expression level was normalized using the FPKM method. Next, the relationship analysis of the samples was performed using correlation analysis and principal component analysis (PCA). PCA was performed using R package models (<http://www.r-project.org/>). To identify differentially expressed genes across samples or groups, the edge R package (<http://www.r-project.org/>) was used. We identified genes with a fold change ≥ 2 and a false discovery rate (FDR) < 0.05 in a comparison as significant DEGs. DEGs were then subjected to enrichment analysis of GO functions (<http://www.geneontology.org/>) and KEGG pathways (<https://www.kegg.jp/>). Transcription factor analysis was based on the HMM model HMMER v3.1 TFs predict b2 program (<http://hmmer.org/>), while the configuration of the database for plant transcription factor was PlantTFDB (<http://planttfdb.cbi.pku.edu.cn/>).

Gene expression analysis by quantitative RT-PCR

Eighteen genes were chosen from the differentially expressed genes (DEGs) that control the key enzymes in the phenylpropanoid biosynthesis pathway, which can catalyse phenylalanine turn into lignin. Total RNA was separately extracted from T-ZM, F-ZM, T-L7 and F-L7. Gene-specific primers were designed by Primer Premier 5.0. The first strand of cDNA was synthesized with reverse transcription following the TransScript One-Step gDNA Removal and cDNA Synthesis SuperMix Kit's instructions (Takara, Dalian, China). It was used as a template for qRT-PCR to detect the expression level of candidate genes. Real-time quantification was performed on the Lightcycler 96 (Roch, USA) with SYBR Green Master (Roche). The expression levels of the candidate genes were calculated using the $2^{-\Delta\Delta CT}$ method. All PCR experiments were performed on three independent biological samples, each including three technical replicates. *Actin* was used as the reference gene in walnut.

Abbreviations

DAFB, days after full bloom; PAL, phenylalanine ammonia-lyase; 4CL, 4-coumarate-CoA ligase; C3'H, p-coumaroyl shikimate 3'-hydroxylase; C4H, cinnamate 4-hydroxylase; CAD, cinnamyl alcohol

dehydrogenase; DEG: differentially expressed gene; RT-qPCR: Quantitative real-time Polymerase Chain Reaction; CCR, cinnamoyl-CoA reductase; COMT, caffeic acid 3-O-methyltransferase; CCoAOMT, caffeoyl-CoA O-methyltransferase; F5H, ferulate-5-hydroxylase; HCT, shikimate O-hydroxycinnamoyl transferase; POD, peroxidase; CSE, caffeoyl shikimate esterase; FPKM, fragments per kilo base of transcript per million mapped reads.

Declarations

Ethics approval and consent to participate

No specific permit was required for the present study in P. R. China.

Consent for publication

Not applicable.

Availability of data and materials

The raw RNA-seq data of the samples is uploading into the NCBI with accession number SUB7208520.

Competing interest

We declare that we do not have any commercial or associative interest that represents a conflict of interest in connection with the work submitted.

Funding

This work was supported by the Natural Science Foundation of Hebei Province (C2019204270), the Science and Technology Research Project of the Universities of Hebei Province (QN2019159), and Research Foundations for Returned Scholars from Overseas of the Human Resources Dept. of Hebei Province (C20190345). The funders played no role in the design or operation of the study or in the analysis of data or in writing the manuscript.

Author' contributions

XTW conceived and designed the experiments. XTW and YBQ collected the material. XTW and ZCZ performed the experiments and analyzed the data. XTW wrote the manuscript. MTS XHA and SGZ participated in the revision of manuscripts and gave suggestions. All authors read and approved the manuscript.

Acknowledgements

We thank the Elsevier Language Editing Service for corrections of English on the present manuscript.

References

1. Aradhya M, Woeste K, Velasco D: Genetic diversity, structure and differentiation in cultivated walnut (*Juglans regia* L.). *Acta Horticulturae* 2012, 861(861):127-132.
2. Chen LN, Ma QG, Chen YK, Wang BQ, Pei D: Identification of major walnut cultivars grown in China based on nut phenotypes and SSR markers. *Scientia Horticulturae* 2014, 168(168):240-248.
3. KazemArzani, Mansouri-Ardakan H, Vezvaei A, Roozban MR: Morphological variation among Persian walnut (*Juglans regia*) genotypes from central Iran. *New Zealand Journal of Experimental Agriculture* 2008, 36(3):159-168.
4. Li XB, Zhao SG, Wang HX, Gao Y, Zhang ZH: Relationship between Walnut Shell Structures and the Content of Lignin and Cellulose. *Hubei Agricultural Sciences* 2012, 51(22):5076-5079.
5. Vanholme R, Demedts B, Morreel K, Ralph J, Boerjan W: Lignin Biosynthesis and Structure. *Plant physiology* 2010, 153(3):895-905.
6. Vanholme R, Storme V, Vanholme B, Sundin L, Christensen JH, Goeminne G, Halpin C, Rohde A, Morreel K, Boerjan W: A Systems Biology View of Responses to Lignin Biosynthesis Perturbations in *Arabidopsis*. *The Plant cell* 2012, 24(9):3506-3529.
7. Denness L, McKenna JF, Segonzac C, Wormit A, Madhou P, Bennett M, Mansfield J, Zipfel C, Hamann T: Cell Wall Damage-Induced Lignin Biosynthesis Is Regulated by a Reactive Oxygen Species- and Jasmonic Acid-Dependent Process in *Arabidopsis*. *Plant physiology* 2011, 156(3):1364-1374.
8. Vojtěch D, Phil J, Jan H: Hormonal regulation of secondary cell wall formation. *Journal of experimental botany* 2015, 66(16):5015-5027.
9. Boerjan W, Ralph J, Baucher M: Lignin Biosynthesis. *Annual Review of Plant Biology* 2003, 54(1):519-546.
10. Lee Y, Chen F, Gallego-Giraldo L, Dixon RA, Voit EO: Integrative analysis of transgenic alfalfa (*Medicago sativa* L.) suggests new metabolic control mechanisms for monolignol biosynthesis. *PLoS computational biology* 2011, 7(5):e1002047.
11. Barakat A, Bagniewska-Zadworna A, Choi A, Plakkat U, Diloreto DS, Yellanki P, Carlson JE: The cinnamyl alcohol dehydrogenase gene family in *Populus*: phylogeny, organization, and expression. *BMC plant biology* 2009, 9(1):26.
12. Vanholme R, Barbara DM, John R, Wout B: Lignin biosynthesis and its integration into metabolism. *Current Opinion in Biotechnology* 2019, 56:230-239.
13. Jaclyn J S, Takuya A, Clint C, John R, Shawn D M: The effects on lignin structure of overexpression of ferulate 5-hydroxylase in hybrid poplar. *Plant physiology* 2009, 150(2):621-635.
14. Voxeur A, Wang Y, Sibout R: Lignification: different mechanisms for a versatile polymer. *Current opinion in plant biology* 2015, 23:83-90.
15. C G W, S D M, F L, S W, J-Y P, S D K, E G-V, D P, F U, J R: Monolignol ferulate transferase introduces chemically labile linkages into the lignin backbone. *Science* 2014, 344(6179):90-93.

16. Ruben V, Kris M, Chiarina D, Paula O, Grabber JH, John R, Wout B: Metabolic engineering of novel lignin in biomass crops. *New Phytologist* 2012, 196(4):978-1000.
17. Wang JP, Matthews ML, Williams CM, Shi R, Chiang VL: Improving wood properties for wood utilization through multi-omics integration in lignin biosynthesis. *Nature Communications* 2018, 9(1).
18. Goujon T, Ferret V, Mila I, Pollet B, Ruel K, Burlat V, Joseleau J-P, Barrière Y, Lapierre C, Jouanin L: Down-regulation of the *AtCCR1* gene in *Arabidopsis thaliana*: effects on phenotype, lignins and cell wall degradability. *Planta* 2003, 217(2):218-228.
19. Gallego-Giraldo L, Jikumaru Y, Kamiya Y, Tang Y, Dixon RA: Selective lignin downregulation leads to constitutive defense response expression in alfalfa (*Medicago sativa* L.). *New Phytologist* 2010, 190(3):627-639.
20. Rohde A, Morree K, Ralph J, Goeminne G, Hostyn V, Rycke RD, Kushnir S, Doorselaere JV, Joseleau J-P, Vuylsteke M *et al*: Molecular Phenotyping of the *pal1* and *pal2* Mutants of *Arabidopsis thaliana* Reveals Far-Reaching Consequences on Phenylpropanoid, Amino Acid, and Carbohydrate Metabolism. *The Plant cell* 2004, 16(10):2749-2771.
21. Sookyung O, Sunchung P, Kyung-Hwan H: Transcriptional regulation of secondary growth in *Arabidopsis thaliana*. *Journal of experimental botany* 2003, 54(393):2709-2722.
22. Huang J, Gu M, Lai Z, Fan B, Shi K, Zhou Y-H, Yu J-Q, Chen Z: Functional Analysis of the *Arabidopsis* PAL Gene Family in Plant Growth, Development, and Response to Environmental Stress. *Plant physiology* 2010, 153(4):1526-1538.
23. Xu B, Escamilla-Treviño LL, Sathitsuksanoh N, Shen Z, Shen H, Zhang YHP, Dixon RA, Zhao B: Silencing of 4-coumarate:coenzyme A ligase in switchgrass leads to reduced lignin content and improved fermentable sugar yields for biofuel production. *New Phytologist* 2011, 192(3):611-625.
24. Wang DD, Bai H, Chen W-Q, Lu H, Jiang X-N: Identifying a Cinnamoyl Coenzyme A Reductase (CCR) Activity with 4-Coumaric Acid: Coenzyme A Ligase (4CL) Reaction Products in *Populus tomentosa*. *Journal of Plant Biology* 2009, 52(5):482-491.
25. Zarei A, Zamani Z, Fatahi R, Mousavi A, Salami SA, Avila C, Cánovas FM: Differential expression of cell wall related genes in the seeds of soft- and hard-seeded pomegranate genotypes. *Scientia Horticulturae* 2016, 205:7-16.
26. Costa MA, Bedgar DL, Moinuddin SGA, Kim KW, Lewis NG: Characterization in vitro and in vivo of the putative multigene 4-coumarate:CoA ligase network in *Arabidopsis*: syringyl lignin and sinapate/sinapyl alcohol derivative formation. *Phytochemistry* 2005, 66(17):2072-2091.
27. Qiao Z, Dixon RA: Transcriptional networks for lignin biosynthesis: more complex than we thought? *Trends in plant science* 2011, 16(4):227-233.
28. Chen K, Song M, Guo Y, Liu L, Xue H, Dai H, Zhang Z: *MdMYB 46* Could Enhance Salt and Osmotic Stress Tolerance in Apple by Directly Activating Stress Responsive Signals. *Plant Biotechnology Journal* 2019, 17(12):2341-2355.
29. Zhang X, Zhang L, Zhang Q, Xu J, Liu W, Dong W: Comparative transcriptome profiling and morphology provide insights into endocarp cleaving of apricot cultivar (*Prunus armeniaca* L.). *BMC*

- plant biology* 2017, 17(72).
30. Zhang S, Jia T, Zhang Z, Zou X, Fan S, a KL, Jiang X, Niu D, Yuan Y, Shang H: Insight into the relationship between S-lignin and fiber quality based on multiple research methods. *Plant Physiology and Biochemistry* 2020, 147:251-261.
 31. Szemilao KWC, Sathe SK: Walnuts (*Juglans regia* L): Proximate composition, protein solubility, protein amino acid composition and protein in vitro digestibility. *Journal of the Science of Food & Agriculture* 2000, 80(9):1393-1401.
 32. Prado ACPD, Manion BA, Seetharaman K, Deschamps FC, Barrera Arellano D, Block JM: Relationship between antioxidant properties and chemical composition of the oil and the shell of pecan nuts [*Carya illinoensis* (Wangenh) C. Koch]. *Industrial Crops and Products* 2013, 45(45):64-73.
 33. Solar A, Colaric[˘] M, Usenik V, Stampar F: Seasonal variations of selected flavonoids, phenolic acids and quinones in annual shoots of common walnut (*Juglans regia* L.). *Plant Science* 2006, 170(3):453-461.
 34. Shangqi Y, Changlu J, Yan S, Chunhua L, Yongcui G, Wentao Z, Liping C, Rui Z: Screening and functional prediction of differential expression genes at lignification stage of endocarp in 'Zhipi'walnut. *Journal of Fruit Science* 2019, 36(04):20-30.
 35. Winkel-Shirley B: Evidence for enzyme complexes in the phenylpropanoid and flavonoid pathways. *Cheminform* 2010, 31(13):142-149.
 36. Hoffmann L, Besseau S, Geoffroy P, Ritzenthaler C, Meyer D, Lapierre C, Pollet B, Legrand M: Silencing of Hydroxycinnamoyl-Coenzyme A Shikimate/Quinate Hydroxycinnamoyltransferase Affects Phenylpropanoid Biosynthesis. *The Plant cell* 2004, 16(6):1446-1465.
 37. Hoffmann L, Maury Sp, Martz Fo, Geoffroy P, Legrand M: Purification, Cloning, and Properties of an Acyltransferase Controlling Shikimate and Quinate Ester Intermediates in Phenylpropanoid Metabolism. *Journal of Biological Chemistry* 2002, 278(1):95-103.
 38. Tognolli M, Penel C, Greppin H, Simon P: Analysis and expression of the class III peroxidase large gene family in *Arabidopsis thaliana*. *Gene* 2002, 288(1-2):129-138.
 39. Hiraga S, Yamamoto K, Ito H, Sasaki K, Ohashi Y: Diverse expression profiles of 21 rice peroxidase genes. *Febs Letters* 2000, 471(2-3):245-250.
 40. Vanholme R, Cesarino I, Rataj K, Xiao Y, Sundin L, Goeminne G, Kim H, Cross J, Morreel K, Araujo P *et al*: Caffeoyl Shikimate Esterase (CSE) Is an Enzyme in the Lignin Biosynthetic Pathway in *Arabidopsis*. *Science* 2013, 341(6150):1103-1106.
 41. Saleme MDLS, Cesarino I, Vargas L, Kim H, Vanholme R, Goeminne G, Van Acker R, Fonseca FCdA, Pallidis A, Voorend W: Silencing CAFFEYOYL SHIKIMATE ESTERASE affects lignification and improves saccharification. *Plant physiology* 2017, 175:1040-1057.
 42. Taylor LP, Grotewold E: Flavonoids as developmental regulators. *Current opinion in plant biology* 2005, 8(3):317-323.
 43. Zhao Q: Lignification: Flexibility, Biosynthesis and Regulation. *Trends in plant science* 2016, 21(8):713-721.

44. Zheng ZC, Zou JC, Hua B, Zhang HJ, Wang R: Study on the Constituents of Walnut Shell. *JOURNAL OF SOUTHWEST FORESTRY COLLEGE* 2006, 26(02):35-38.
45. Hatfield R, Vermerris W. Lignin formation in plants. The dilemma of linkage specificity. *Plant physiology* 2001, 126(4):1351-1357.
46. Baucher M, Halpin C, Petitconil M, Boerjan W: Lignin: genetic engineering and impact on pulping. *Crc Critical Reviews in Biochemistry* 2003, 38(4):305-350.
47. Chabannes M, Ruel K, Yoshinaga A, Chabbert B, Jauneau A, Joseleau J-P, Boudet A-M: In situ analysis of lignins in transgenic tobacco reveals a differential impact of individual transformations on the spatial patterns of lignin deposition at the cellular and subcellular levels. *The Plant Journal* 2001, 28(3):271-282.
48. Dardick CD, Callahan AM, Chiozzotto R, Schaffer RJ, Scorza R: Stone formation in peach fruit exhibits spatial coordination of the lignin and flavonoid pathways and similarity to Arabidopsis dehiscence. *Bmc Biology* 2010, 8(1):1-17.
49. Shu-gang Z, Jing W, Hong-xia W, Zhi-hua Z, Xi-bo L: Changes in Lignin Content and Activity of Related Enzymes in the Endocarp during the Walnut Shell Development Period. *Horticultural Plant Journal* 2016, 2(3):141-146.
50. Lee M-H, Jeon HS, Kim SH, Chung JH, Roppolo D, Lee H-J, Cho HJ, Tobimatsu Y, Ralph J, Park OK: Lignin-based barrier restricts pathogens to the infection site and confers resistance in plants. *The EMBO journal* 2019, 38(23):1-17.
51. Sewalt V, Ni WT, Blount JW, Jung HG, Dixon RA: 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* 1997, 115(1):41-50.
52. Guo D, Chen F, Inoue K, Blount JW, Dixon RA: Downregulation of caffeic acid 3-O-methyltransferase and caffeoyl CoA 3-O-methyltransferase in transgenic alfalfa. impacts on lignin structure and implications for the biosynthesis of G and S lignin. *The Plant cell* 2001, 13(1):73-88.
53. Zhao H, Wei J, Zhang J, Liu H, Wang T, Song Y: Lignin biosynthesis by suppression of two O-methyltransferases. *Chinese Science Bulletin(English version)* 2002, 47(13):1092-1095.
54. Abdulrazzak N, Pollet B, Ehling Jr, Kim Larsen, Carole Asnaghi, Ronseau S, Proux C, Erhardt M, Seltzer V, Renou J-P *et al*: Corrections: A coumaroyl-ester-3-hydroxylase Insertion Mutant Reveals the Existence of Nonredundant meta-Hydroxylation Pathways and Essential Roles for Phenolic Precursors in Cell Expansion and Plant Growth. *Plant physiology (on line)* 2007, 143(4):1982-1982.
55. Peng XP, Sun S-L, Wen J-L, Yin W-L, Sun R-C: Structural characterization of lignins from hydroxycinnamoyl transferase (HCT) down-regulated transgenic poplars. *Fuel* 2014, 134:485-492.
56. Laurent H, Sébastien B, Pierrette G, Christophe R, Denise M, Catherine L, Brigitte P, Michel L: Silencing of hydroxycinnamoyl-coenzyme A shikimate/quinic acid hydroxycinnamoyltransferase affects phenylpropanoid biosynthesis. *The Plant cell* 2004, 16(6):1446-1465.
57. Pattathil S, Saffold T, Gallego-Giraldo L, O'Neill M, York WS, Dixon RA, Hahn MG: Changes in Cell Wall Carbohydrate Extractability Are Correlated with Reduced Recalcitrance of HCT Downregulated Alfalfa

- Biomass. *Industrial Biotechnology* 2012, 8(4):217-221.
58. Tong Z, Li H, Zhang R, Lei M, Dong J, Tao W: Co-downregulation of the hydroxycinnamoyl-CoA:shikimate hydroxycinnamoyl transferase and coumarate 3-hydroxylase significantly increases cellulose content in transgenic alfalfa (*Medicago sativa* L.). *Plant Science* 2015, 239:230-237.
59. Adams ZP, Ehling J, Edwards R: The regulatory role of shikimate in plant phenylalanine metabolism. *Journal of theoretical biology* 2018, 462:158-170.
60. Kushwaha C, Chand R, Singh AK, Rai R, Srivastava CP, Singh BD, Mohapatra C: Lignification and early abortive fungal colonies as indicators of partial resistance to rust in pea. *Tropical Plant Pathology* 2016, 41(2):1-7.
61. Zhang X-H, Zheng X-T, Sun B-Y, Peng C-L, Chow WS: Over-expression of the CHS gene enhances resistance of Arabidopsis leaves to high light. *Environmental and Experimental Botany* 2018, 154:33-43.
62. Mitsuda N, Ohme-Takagi M: NAC transcription factors *NST1* and *NST3* regulate pod shattering in a partially redundant manner by promoting secondary wall formation after the establishment of tissue identity. *Plant Journal* 2008, 56(5):768-778.
63. Yamaguchi M, Ohtani M, Mitsuda N, Kubo M, Ohme-Takagi M, Fukuda H, Demura T: VND-INTERACTING2, a NAC Domain Transcription Factor, Negatively Regulates Xylem Vessel Formation in Arabidopsis. *The Plant cell* 2010, 22(4):1249-1263.
64. Fan M, Xu C, Xu K, Hu Y: LATERAL ORGAN BOUNDARIES DOMAIN transcription factors direct callus formation in *Arabidopsis* regeneration. *Cell Research* 2012, 22(7):1169-1180.
65. Lu Q, Shao F, Macmillan C, Wilson IW, van der Merwe K, Hussey SG, Myburg AA, Dong X, Qiu D: Genomewide analysis of the lateral organ boundaries domain gene family in *Eucalyptus grandis* reveals members that differentially impact secondary growth. *Plant Biotechnology Journal* 2018, 16(1):124-136.
66. Fernando GJ, Sebastian L, Heike T, Britt-Louise L, Thomas K, David G, Edgar M, Mark V, I. SE: Massive up-regulation of LBD transcription factors and EXPANSINs highlights the regulatory programs of rhizomania disease. *Molecular Plant Pathology* 2018, 19(10):2333-2348.
67. Yang T, Yue FG, Hua H, Jianghua C, Meng-xiang S: Genome-Wide Identification, Evolutionary Analysis and Expression Profiles of LATERAL ORGAN BOUNDARIES DOMAIN Gene Family in *Lotus japonicus* and *Medicago truncatula*. *Plos One* 2016, 11(8):e0161901.
68. Nihal Dharmasiri, Dharmasiri S, Estelle M: The F-box protein TIR1 is an auxin receptor. *Nature* 2005(435):441-445.
69. Mangeon A, Bell EM, Lin W-c, Jablonska B, S.Springer P: Misregulation of the LOB domain gene *DDA1* suggests possible functions in auxin signalling and photomorphogenesis. *Journal of experimental botany* 2011, 62(1):221-233.
70. Shu-gang Z, Hong-xia W, Yi G, Fa-chao C, Zhi-hua Z, Jian-chao G, Ling L: A New Precocious Walnut Cultivar 'Zanmei'. *Acta Horticulturae Sinica* 2013, 40(6):1207-1208.

71. Shi-hai W, Qi C, Bao-jun Z, Wan-sheng L: Breeding report of new walnut cultivar 'Liaoning 5' and 'Liaoning 7'. *China Fruits* 1997(02):7-9.
72. Zhen-hua X: The project of Central Finance Forestry Promotion "The promotion of Liaoning 7 and other three excellent walnut varieties" passed the inspection of experts. *China Forestry Science and Technology* 2013(1):62.

Figures

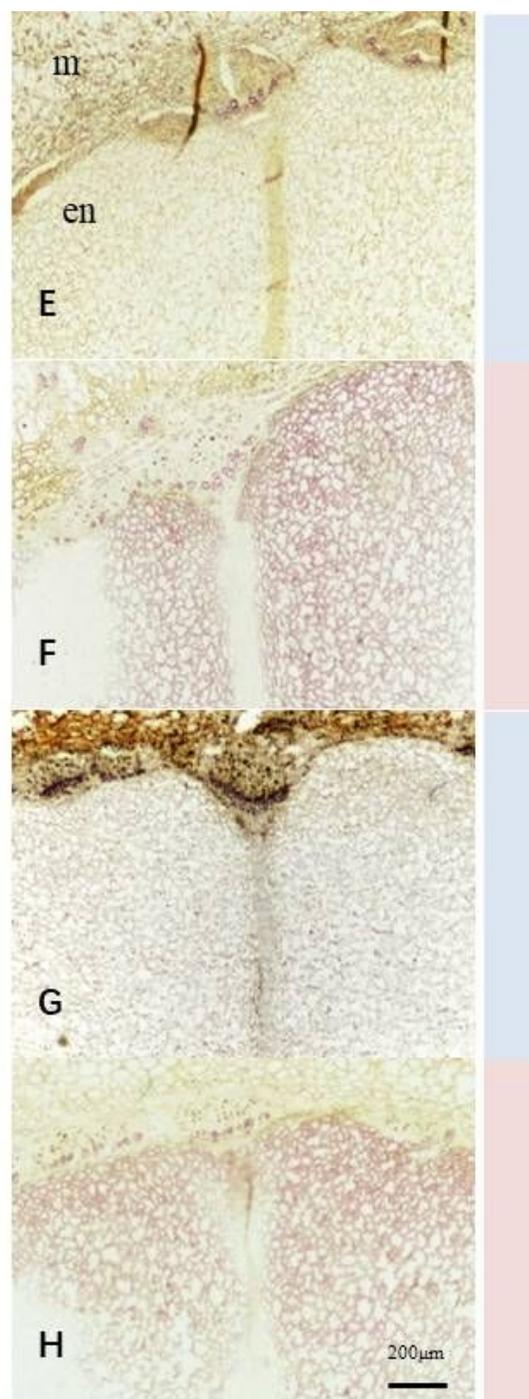
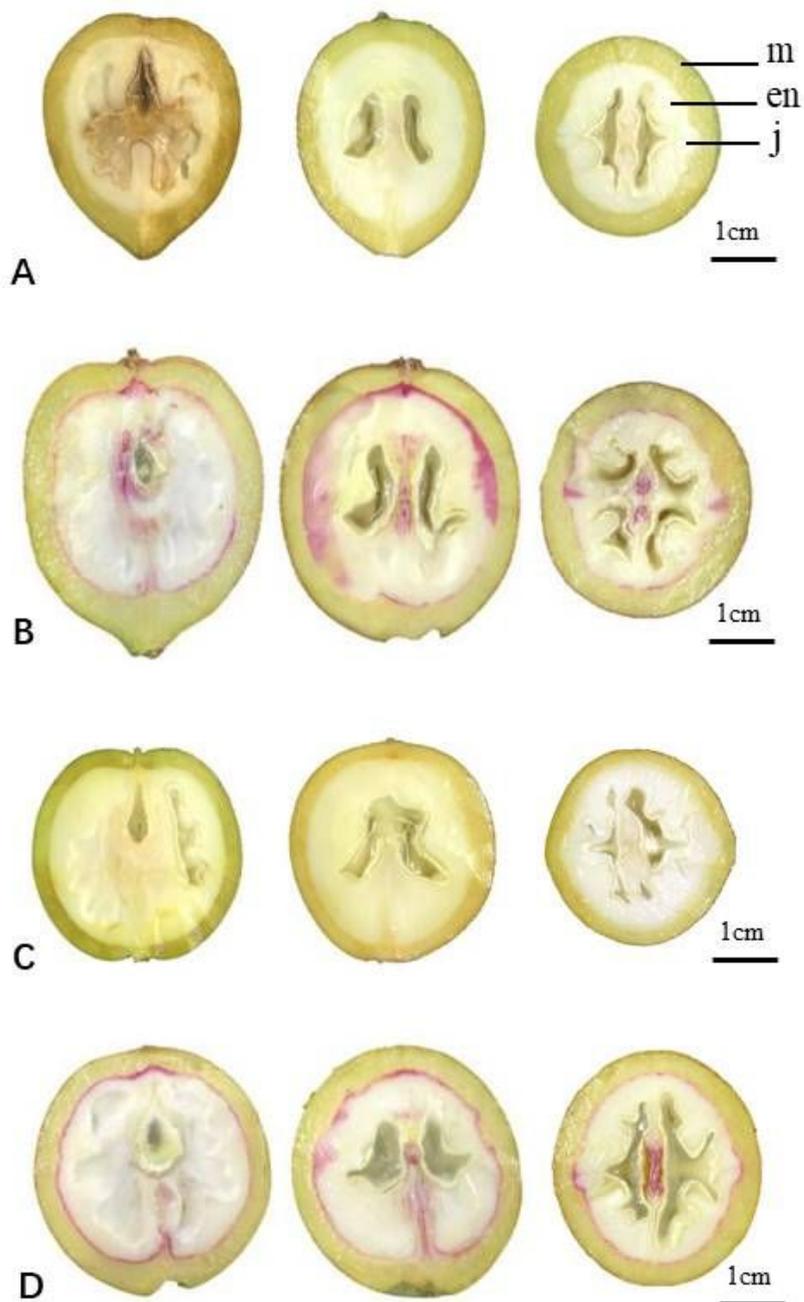


Figure 1

Morphological characteristics of walnut endocarp in 'Zanmei' (A, B, E, F) and 'Liaoning 7' (C, D, G, H). A-D showed the fresh cut walnut stained with phloroglucinol-HCl. Walnuts from left to right in the picture are longitudinal section of suture line, longitudinal section and cross section. E-H showed the sample slice with phloroglucinol-HCl stain. A, C, E, G 30 DAFB. B, D, F, H 45 DAFB. m: mesocarp, en: endocarp, j: junction of endocarp.

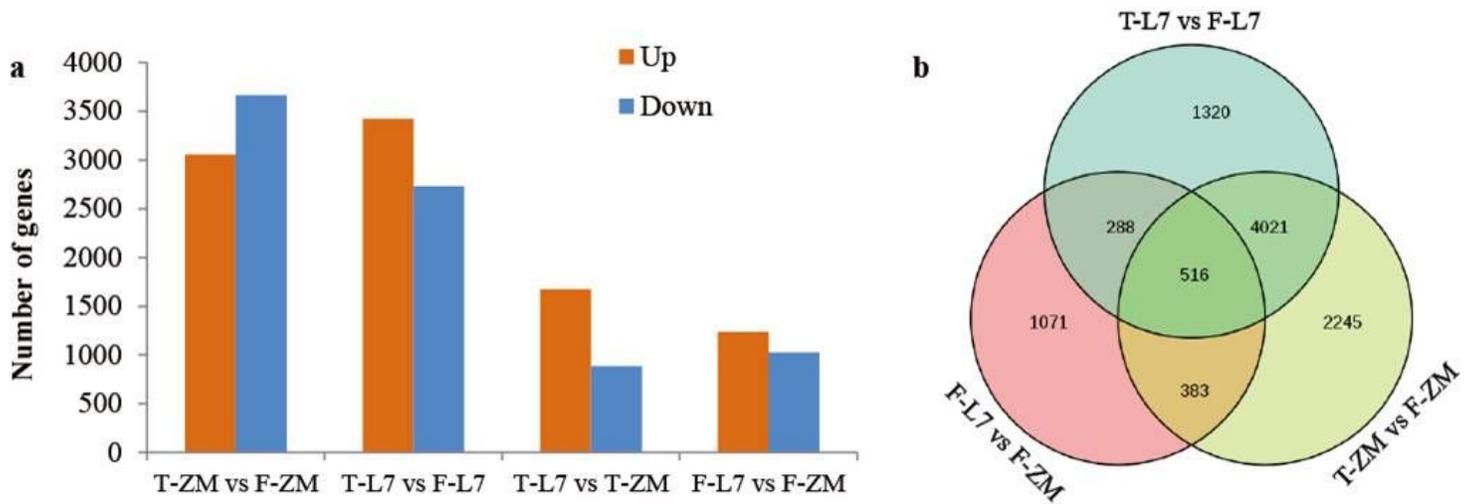


Figure 2

DEGs statistics (a) and Venn diagrams between different groups (b). a: Red color indicates up-regulated expression of DEGs; Blue color indicates down-regulated expression of DEGs. b: The number represents the number of DEGs in the region.

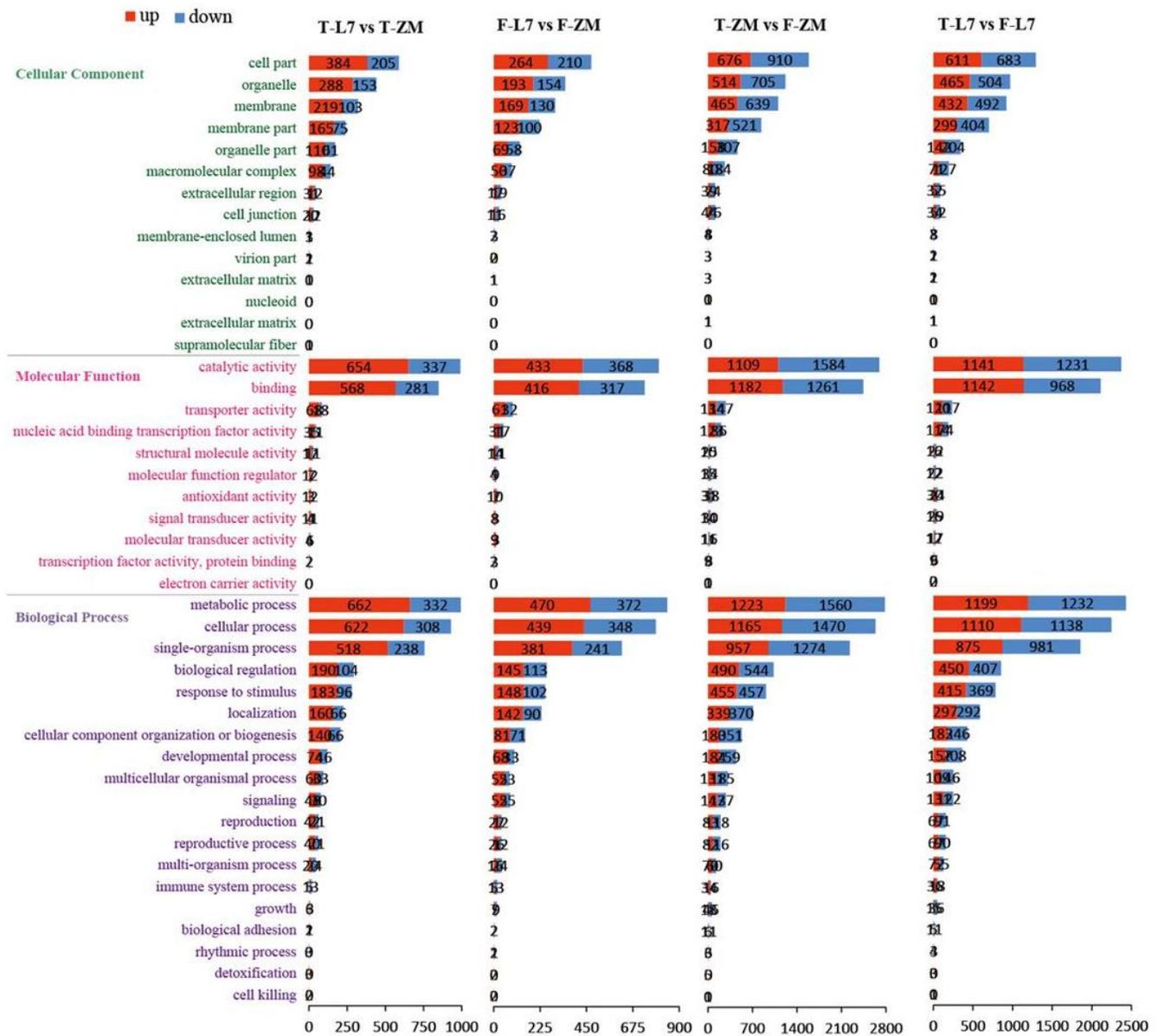


Figure 3

Gene ontology (GO) enrichment of DEGs in group (T-L7 and T-ZM), (F-L7 vs F-ZM), (T-ZM vs F-ZM) and (T-L7 vs F-L7). Level two GO terms belonged to three categories, cellular component, molecular function, and biological process. They were represented along the y-axis. Numbers indicate the number of unigenes assigned to each term.



Figure 4

Top 20 of KEGG pathway enrichment of groups. A: (T-ZM vs F-ZM) group; B: (T-L7 vs F-L7) group; C: (F-L7 vs F-ZM) group; D: (T-L7 vs T-ZM) group. The circle size represents the number of different genes. The more DEGs, the larger the circle. Richfactor refers to the ratio of the number of genes in the pathway entry in DEGs to the total number of genes in the pathway entry in all genes. The larger the rich factor, the

higher the degree of enrichment. Qvalue is a pvalue corrected by multiple hypothesis tests. The value ranges from 0 to 1. The closer it approaches zero, the more significant the enrichment is. The graph is plotted using pathways in the top 20 rankings of qvalue from small to large.

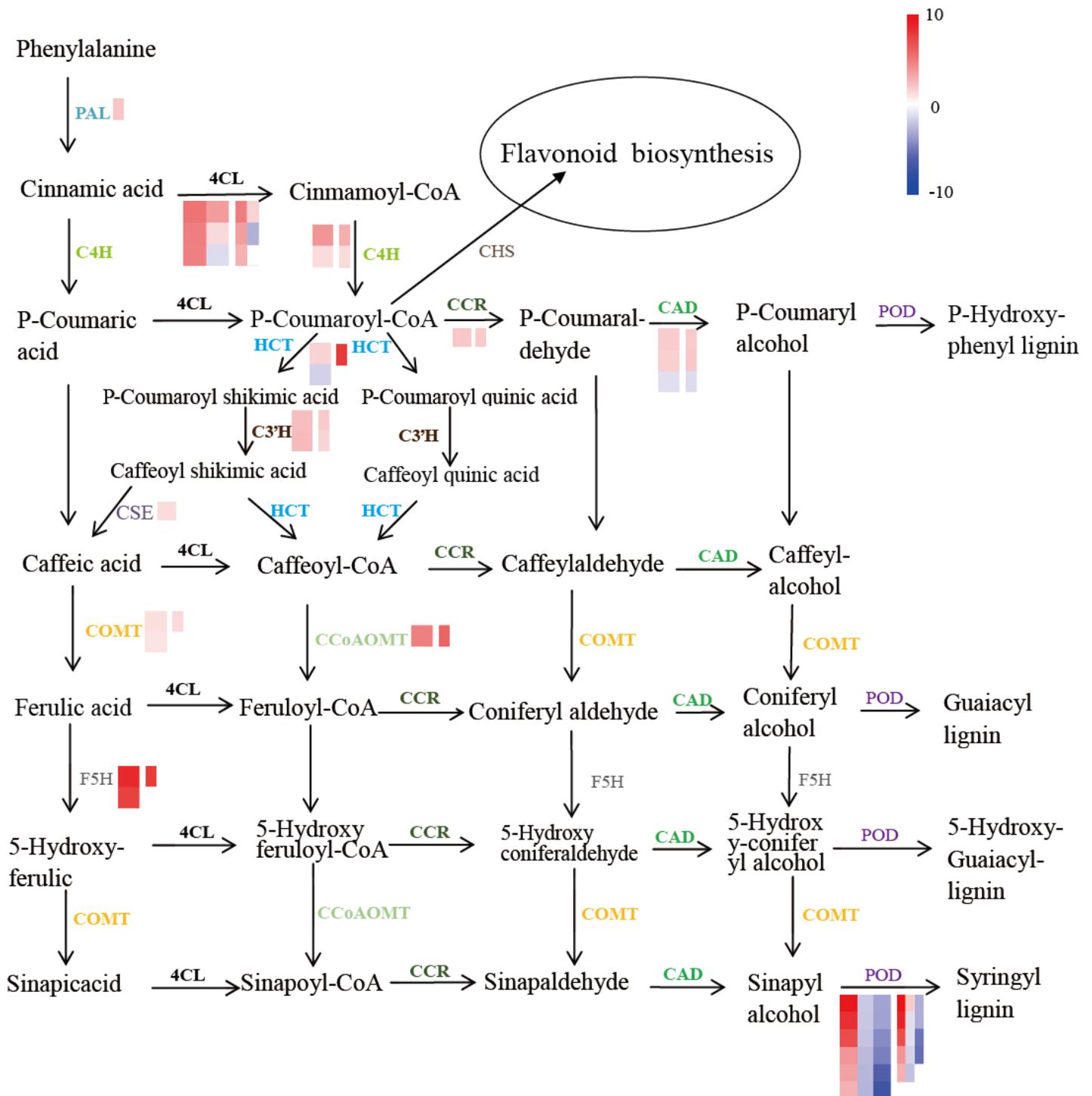


Figure 5

Relative expression of DEGs in phenylpropanoid biosynthesis pathway of group (T-ZM vs F-ZM) and (T-L7 vs F-L7). A square represented a DEG in group (T-ZM vs F-ZM), and a rectangle represented a DEG in group (T-L7 vs F-L7). The colors of the squares and rectangles indicated the relative expression of the gene. The deeper the red, the greater the positive difference multiples; The darker the blue, the greater the

negative difference multiple. Expression pattern in this figure is from $\log_2(F/T)$. All the genes shown in the figure had been screened, and the screening condition was: FPKM value > 5, FDR < 0.05, $|\log_2(F/T)| > 1$. The same enzyme was shown in the same color.

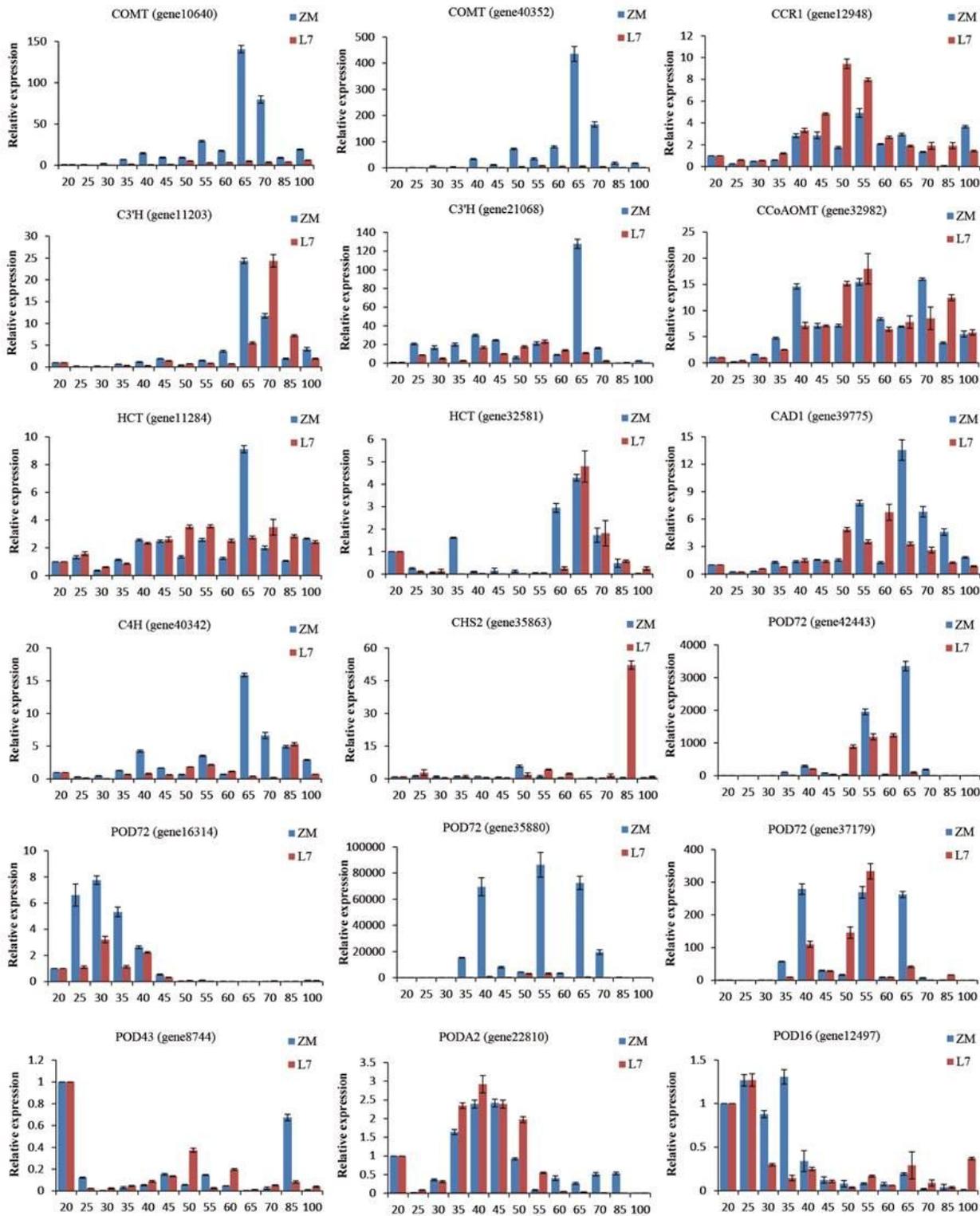


Figure 6

qPCR analyses of selected candidate genes involved in this study. Blue bars indicate ZM walnut; red bars indicate L7 walnut. Numbers under the x-axis indicate the days after full bloom.

Supplementary Files

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

- [AdditionalFiles.zip](#)