

Metabolite profiling and transcriptome analysis explains difference in accumulation of bioactive constituents in Taxilli Herba from different hosts

Jiahuan Yuan

Nanjing University of Chinese Medicine

Nan Wu

Nanjing University of Chinese Medicine

Zhichen Cai

Nanjing University of Chinese Medicine

Cuihua Chen

Nanjing University of Chinese Medicine

Yongyi Zhou

Nanjing University of Chinese Medicine

Haijie Chen

Nanjing University of Chinese Medicine

Jia Xue

Nanjing University of Chinese Medicine

Li Li

Guangxi University of Chinese Medicine

Xunhong Liu (✉ 300100@njucm.edu.cn)

Nanjing University of Chinese Medicine

Wenxin Wang

Nanjing University of Chinese Medicine

Jianming Cheng

Jiangsu Province Engineering Research Center of Classical Prescription

Research Article

Keywords: Taxilli Herba, Hosts, Transcriptomics, Metabolomics, Biosynthesis network

Posted Date: November 30th, 2022

DOI: <https://doi.org/10.21203/rs.3.rs-2257226/v1>

Abstract

Background: Taxilli Herba (TH) is a semi-parasitic therapeutic herb and the host plant is a key factor affecting its quality. *Morus alba* L. (SS) and *Liquidambar formosana* Hance (FXS) are the two most frequent hosts. The purpose of this study was to elucidate the underlying molecular mechanism causing the variation in the accumulation of bioactive constituents in TH from SS and FXS.

Results: In this paper, 3319 differentially expressed genes (DEGs) comprising the transcription factor families AP2/ERF and MYB super_family were identified. In addition, 81 compounds were identified and their relative levels were compared between the SS and FXS groups using data from the ultra-fast liquid chromatography coupled with triple quadrupole-time of flight tandem mass spectrometry (UFLC-Triple TOF-MS/MS) analysis. Putative biosynthesis networks of flavonoid and phenylpropanoid were created combined with structural genes. The expression patterns of genes were mostly consistent with the variation of bioactive constituents in TH. Notably, the expression trend of UDP-glycosyltransferase genes in TH suggested that postmodification of glycosyltransferase may participate in downstream flavonoid synthesis.

Conclusions: The findings of this work would shed light on the relationship between gene regulation and the flavonoid and phenylpropanoid biosynthesis in TH, providing fundamental information on the difference in molecular mechanisms for the quality formation of TH from various hosts.

Introduction

Taxilli Herba (TH) is derived from the dried stems and branches with leaves of *Taxillus chinensis* (DC.) Danser, which belongs to the Ichoranthaceae family [1]. TH is a well-known traditional Chinese herb that has been used for the treatment of symptoms such as soreness and weakness of waist and knees [2]. It is reported that TH contains numerous types of chemical components such as flavonoids [3], organic acids [4], and volatile oils [5]. Furthermore, people adore TH tea as a traditional healthcare product because of its distinctly sweet, refreshing, and precise health care function [6].

As a semi-parasitic plant medicine, extensive range of host plants is an essential biological characteristic. It was found that the host plants are distributed in several families according to previous investigations on the status of its host plants [7]. Chinese Pharmacopoeia has no clear provisions on the host plants of TH [1]. At present, most of the medicinal materials circulating and used in the market are of multi-host sources, with uneven quality. Among them, *L. Morus alba* L. (SS) and *Liquidambar formosana* Hance (FXS) are the two most widely circulated in the market and clinically used sources. The influence of different hosts on the quality of TH has been adequately documented in the herbology of different historical periods, and the differences in toxicity and efficacy of TH from different hosts have been noted [8]. The current research on TH from different host sources is mainly divided into three parts. The first part focuses on the chemical components, including the content determination of total flavonoids and major flavonoids such as quercetin and quercitrin [9–11]. It also includes the chemical components analysis of TH from certain characteristic hosts [12, 13]. The second part focuses on

the pharmacological effects, and scholars have used zebrafish acute toxicity model, liver injury model [14], spontaneously hypertensive rat model [15], and adjuvant arthritis mice model [16] to evaluate the pharmacological effects of TH from different hosts. The third part is the identification and distinction of TH from different hosts using various techniques including HPLC fingerprint and infrared spectroscopy methods [17, 18], but these methods are difficult to use effectively in practical situations. The results of the studies available so far have found significant differences in the quality of TH from different hosts. Nevertheless, the differences in the synthesis and accumulation of chemical components that may result from differences in hosts, it is essential to explore the underlying molecular mechanisms. Only by thoroughly understanding the formation of bioactive constituents of TH can we provide a rational basis for the host selection of TH. Therefore, it is extremely vital to analyze them at the transcriptional level to identify differences and to investigate the genetic quality associated with the biosynthesis of these constituents in TH from two hosts.

RNA-sequencing (RNA-Seq) is a modern method for transcriptome research based on next-generation sequencing technology that has emerged in recent years with the advantages of large throughput, wide coverage, and high precision [19]. It has been widely used in the field of molecular biology research and is a convenient and rapid technical tool. RNA-Seq can sequence cDNA directly [20] and plays an important role in the mining of species resilience genes [21], biosynthesis, and identification of regulatory major genes [22–24], etc. To the best of our knowledge, elucidation of the mechanism of molecular change underlying the accumulation of bioactive constituents in TH remains vacant due to the paucity of transcriptomic information. A combination of metabolomics and transcriptomics could provide comprehensive insight into biological events at the “gene-metabolite” level. In the present study, we investigated the molecular mechanisms involved in the biosynthesis of the chemical components of TH from SS and FXS, using an integrated transcriptomic analysis and metabolomic approach. Differentially expressed genes (DEGs) were screened out and analyzed using bioinformatics analysis. Genes involved in flavonoid and phenylpropanoid biosynthesis were selected and putative biosynthesis networks were created. This work would not only facilitate the elucidation of flavonoid and phenylpropanoid biosynthesis in TH from different hosts, but also contribute to quality improvement of TH.

Materials And Methods

Plant materials

Two parallel TH samples from *Morus alba* L. (No.2022011901–2022011903) and *Liquidambar formosana* Hance (No.2022011904–2022011906) were collected from Wuzhou of Guangxi Province in China, each with three biological replicates. The botanical origins of the materials were authenticated by Professor Xun-hong Liu (Nanjing University of Chinese Medicine). Voucher specimens were deposited in the laboratory of Chinese medicine identification, Nanjing University of Chinese Medicine. Stems and branches with leaves were collected and washed with phosphate-buffered saline (PBS), frozen in liquid nitrogen for further experiments and analysis.

Metabolite profiling analysis of bioactive constituents in Taxilli Herba

The samples were powdered and sieved through 50 mesh. Approximately 0.5 g of powder was extracted by sonication with 15 mL of 50% v/v methanol for 30 min at room temperature. After centrifugation at 13000 r/min for 10 min, the supernatant was passed through a 0.22 µm membrane (Jinteng laboratory equipment Co., Ltd., Tianjin, China). The solutions were stored at 4 °C before ultra-fast liquid chromatography coupled with triple quadrupole-time of flight tandem mass spectrometry (UFLC-Triple TOF-MS/MS) analysis [25].

All samples were analyzed using a UFLC system (Shimadzu, Kyoto, Japan) with the separation conducted by an Agilent ZORBAX SB-C₁₈ column (4.6 mm × 250 mm, 5 µm) at 30 °C. The mobile phase composed of methanol: acetonitrile (1:1, v/v) (A) and 0.4% (v/v) formic acid aqueous with the flow rate at 1.0 mL/min was set as follows: 0–5 min, 2–6% A; 5–6 min, 6–10% A; 6–8 min, 10–15% A; 8–12 min, 15–18% A; 12–18 min, 18–21% A; 18–21 min, 21–23% A; 21–26 min, 23–25% A; 26–30 min, 25–27% A; 30–33 min, 27–40% A; 33–38 min, 40–50% A; 38–40 min, 50–2% A; 40–45 min, 2–2% A. The injection volume was 10 µL [25].

The Triple TOF™ 5600-MS/MS system (AB SCIEX, Framingham, MA, USA) equipped with an ESI source was used to obtain MS data in negative mode. The optimized MS parameters were as follows: the scanning range was from *m/z* 50–1500, the ion source temperature was 600 °C, the curtain gas was 40 psi, the nebulization gas was 60 psi, the collision energy was – 10 V, and the declustering potential voltage was 100 V [25].

The UFLC-Triple TOF-MS/MS data were analyzed by PeakView 1.2 software. The precise molecular mass was determined by MS¹ and the fragment information was obtained by MS². For the identification of compounds, one method was to compared with previously established chemical composition database, and verified with by the retention time of and the mass spectrometry data of the standards. The second method was to speculate by combining the database and relevant literature. At the same time, the relative peak area of metabolites was extracted, and the average value of three samples was taken as the relative content of one metabolite.

Transcriptome analysis of Taxilli Herba

RNA preparation

Total RNA was extracted from TH samples using Plant RNA Purification Reagent (Invitrogen) according to the manufacturer's instructions and genomic DNA was removed using DNase I (TaKara). Then the integrity and purity of the total RNA were determined by 2100 Bioanalyser (Agilent Technologies, Inc., Santa Clara CA, USA) and quantified using the Nanodrop 2000 (NanoDrop Thermo Scientific, Wilmington,

DE, USA). Only high-quality RNA sample ($OD_{260/280} = 1.8-2.2$, $OD_{260/230} \geq 2.0$, $RIN \geq 8.0$, $28S:18S \geq 1.0$, $> 1\mu g$) was used to construct a sequencing library.

Library preparation and sequencing

RNA purification, reverse transcription, library construction, and sequencing were conducted by Shanghai Bio-pharm Biotechnology Co., Ltd (Shanghai, China) following the manufacturer's instructions. Firstly, Illumina TruSeq™ RNA sample preparation Kit (Illumina, USA) was used to prepare the transcriptome library using $1\mu g$ of total RNA. Shortly, mRNA was isolated and enriched from total RNA by A-T base pairing with polyA using magnetic beads with Oligo (dT), and then was fragmented randomly by adding fragmentation buffer, and small fragments of about 300bp were isolated by magnetic bead screening. Subsequently, using these small fragments as templates, a strand of cDNA was synthesized in reverse by adding random primers under the action of reverse transcriptase, followed by two-strand synthesis to form stable double-stranded cDNA. On this base, End Repair Mix was added to complement the double-stranded cDNA into flat ends and used to join the adaptor. Finally, library electrophoresis was performed on 2% Low Range Ultra Agarose with 200–300 bp cDNA as target fragments, which was amplified for 15 cycles using Phusion DNA polymerase (New England Biolabs, Boston, USA). After quantified by TBS380, Illumina NovaSeq 6000 sequencer (2×150 bp read length) was used to sequence the paired-end RNA-seq sequencing library.

De novo assembly and function annotation

First, the bases quality, error rate and distribution were analyzed using fastX_toolkit (http://hannonlab.cshl.edu/fastx_toolkit/). Raw data were filtered by Sickle (<https://github.com/najoshi/sickle>) and SeqPrep (<https://github.com/najoshi/sickle>) to remove low-quality sequences and contaminated adaptors. High-quality clean data was acquired by fastp (<https://github.com/OpenGene/fastp>) followed by de novo assembly with Trinity (<https://github.com/trinityrnaseq/trinityrnaseq>). NCBI protein non-redundant (NR, <ftp://ftp.ncbi.nlm.nih.gov/blast/db/>), Swiss-Prot (http://web.expasy.org/docs/swiss-prot_guideline.html), Pfam (<http://pfam.xfam.org/>), Clusters of Orthologous Groups of proteins (COG, <http://www.ncbi.nlm.nih.gov/COG/>), Gene ontology (GO, <http://www.geneontology.org>) and Kyoto Encyclopedia of Genes and Genomes (KEGG, <http://www.genome.jp/kegg/>) databases were used to acquire function annotation of assembled transcripts by BLAST+ (<ftp://ftp.ncbi.nlm.nih.gov/blast/executables/blast+/2.9.0/>). BLAST2GO (version 2.5.0) was used for GO function annotation, and KOBAS (version 2.1.1) was used to identify related metabolic pathways.

Identification and analysis of differentially expressed genes

The expression level of each transcript was calculated according to the fragments per kilobases per million reads (FPKM) and the abundance was quantified using RSEM software. Furthermore, genes with $|\log_2 FC| \geq 1$ and $p\text{-value} < 0.05$ were considered as DEGs. In addition, the functional and enrichment analysis of GO and KEGG were performed on DEGs to identify the biological functions and pathways

involved. Meanwhile, the genes related to phenylpropanoid and flavonoid biosynthesis were screened out and further analyzed in combination with metabolomics.

Quantitative real-time polymerase chain reaction (qRT-PCR) analysis

Quantitative real-time polymerase chain reaction (qRT-PCR) was applied to quantify the relative mRNA expression patterns of the 9 genes (CCoAOMT, CYP98A, bglB, CAD, CSE, HCT, FLS, 4CL, F3H), which were involved in secondary metabolites synthesis. Total RNA was extracted from samples using the Trizol total RNA isolation kit according to the manufacturer's instructions. The primers were designed using Primer 3.0 (<https://bioinfo.ut.ee/primer3-0.4.0/>). cDNA pools from the total RNA were synthesized for the qRT-PCR using HiScript Q RT SuperMix for qPCR (Vayme Biotech Co., Ltd., Nanjing, China). The 20 μ L reaction volume including 10 μ L 2 $\times$ ChamQ SYBR Color qPCR Master Mix, 0.8 μ L of each primer (μ M), 0.4 μ L 50 $\times$ ROX Reference Dye 1, and 2 μ L cDNA. The PCR procedure was set as follows: 95 $^{\circ}$ C for 5 min, followed by 40 cycles at 95 $^{\circ}$ C for 5 s, primer annealing at 55 $^{\circ}$ C for 30 s, and extension at 72 $^{\circ}$ C for 40 s. Actin was used as the reference standard and the relative gene expression level was calculated using a real-time PCR system (ABI 7500 real-time PCR system) using the $2^{-\Delta\Delta C_t}$ method. Primers for the PCR were shown in Table S1.

Results

The change of metabolic profiling and metabolite content in Taxilli Herba

The metabolic profiles of TH from SS and FXS were investigated according to the previously established UFLC-Triple TOF-MS/MS method. The differences in their metabolic profiles and the changes in metabolites relative contents were further explored [25]. The representative base peak chromatograms of the two groups were shown in Fig. S1. A total of 81 compounds were identified based on retention times, precise molecular weights, and MS² characteristic fragments, combined with comparisons with standards, database searches, and reports in the relevant literature (Table S2). The relative peak area of the identified metabolites was extracted and the average of the relative peak area of three independent samples in each group was calculated as the relative content of a metabolite (Table S3).

As shown in Fig. 1A/B, there was some fluctuation in the relative levels of metabolites in the two TH groups. Since flavonoids are the main active constituents of TH, we explored the total accumulation of flavonol aglycones and flavonol glycosides in two groups. The results showed that there was difference in the total accumulation of both. Among them, flavan-3-ol and flavonol glycosides occupied a higher proportion, respectively, and both compounds accumulated higher in TH from SS (Fig. 1C/D). The accumulation of other constituents showed the same trend (Fig. 1E) as well. In detail, dihydroflavone, dihydroflavanol, flavonol, flavone, flavane, isoflavone, and their corresponding glycosides were obtained and analyzed in this experiment. The heatmap of the metabolite profile (Fig. S2) showed that the content

of most aglycones and glycosides in samples from SS was slightly higher than that in samples from FXS.

De novo transcriptome assembly and sequence analysis

In the present study, the original sequence of TH from SS ranged from 51,805,610–57,061,100, which were optimized to high-quality sequence ranging from 50,022,130–55,502,034, the mapped ratio was 78.15%; the original sequence of TH from FXS ranged from 45,975,326–50,792,870, which were optimized to high-quality sequence ranging from 44,058,022–48,557,952, the mapped ratio was 78.08%. The clean data of each transcriptome sample was as high as 6.32 GB. Q30 (sequences with a sequencing error rate of less than 0.1%) base percentage was 93.37%, Q20 base percentage was 97.55% and the average content of GC was 48.82% (Table S4). Overall, the quality of RNA-seq data was high, which could be used for further analysis. De novo assembly of all clean data was performed using Trinity (Version v2.8.5) and the assembly results were optimized. A total of 57,203 unigenes and 103,690 transcripts were obtained with an average of N50 length of 2,035 bp and 2,195 bp, respectively. In general, the sequencing results were good enough to be used for subsequent analysis.

Functional annotation and classification

After the assembly, the annotation for the assembled unigenes was carried out in six databases (GO, KEGG, COG, NR, Swiss-Prot, and Pfam), a total of 21,197, 10,215, 21,644, 25,791, 17,828 and 17,499 were aligned, respectively (Table S5).

The GO terms were distributed into 52 functional groups (Fig. S3), which were further classified into three categories, namely, biological process, cellular component, and molecular function. The top GO terms were cellular process (9584) and metabolic process (8554) in the biological process category. In the cellular component category, cell part (10391) and membrane part (7528) were major GO terms. In the molecular function category, binding (11864) and catalytic activity (10319) ranked at the top of the terms. As shown in the KEGG annotation (Fig. S4), 10215 (17.86%) genes were significantly matched in the KEGG pathway database and were assigned to five main categories. Metabolism was the biggest category (2739), followed by genetic information processing (1937) and environmental information processing (375). Among them, 146 genes were classified and annotated for biosynthesis of other secondary metabolites. Pathways related to secondary metabolites were listed in Table S6, where phenylpropanoid biosynthesis (79), flavonoid biosynthesis (17), flavone and flavonol biosynthesis (3), isoflavonoid biosynthesis (1) were the main pathways in TH.

3319 DEGs were identified including 1726 up-regulated genes and 1593 down-regulated genes in TH from FXS according to $p < 0.05$, false discovery rate (FDR) < 0.01 and $|\log_2 \text{FC}| \geq 1$ (Fig. 2). All DEGs were mapped against GO database and subjected to enrichment analysis. Figure 3A showed that many genes related to ribosomal subunit were enriched. Among the top 20 items of KEGG enrichment analysis (Fig. 3B), most DEGs were classified into ribosome (79), plant-pathogen interaction (29), biosynthesis of cofactors (21), glyoxylate and dicarboxylate metabolism (18), plant hormone signal transduction (18).

And the enrichment of secondary metabolite pathways was phenylpropanoid biosynthesis (9) and flavonoid biosynthesis (1).

Analysis of genes and metabolites involved in phenylpropanoid and flavonoid biosynthesis in TH

It has been reported that flavonoids are the main active constituents in TH that play an important role in traditional functions. A total of 93 genes involved in phenylpropanoid and flavonoid biosynthesis were screened out (Tables S7 and Fig. 4). Here, the basic synthesis pathways of phenylpropanoid and flavonoid are plotted against the pathways enriched in the KEGG database. The genes annotated in this study in the above two pathways were mapped to the specific positions of active constituents. Simultaneously, the metabolic profiles of TH from two hosts were analyzed, and the relative content of metabolites and the expression of genes in the synthetic pathway were compared and analyzed jointly (Fig. 5).

Phenyl ammonium lyase (PAL) is the first rate-limiting enzyme in the phenylpropanoid biosynthesis pathway. In this study, two PALs were annotated (TRINITY_DN7053_c0_g1, TRINITY_DN7053_c0_g2). Although the expression of one was up-regulated and the expression of the other was down-regulated in TH from SS, the FC value fluctuated around 1 compared to those in TH from FXS, with nearly no difference. 4-coumarate-CoA ligase (4CL) participates in the biosynthesis of flavonoids in the downstream steps of the pathway, converting p-coumaric acid into p-coumaroyl CoA, which is a substrate for the synthesis of various flavonoids constituents. Nine putative 4CLs (TRINITY_DN13193_c0_g2, TRINITY_DN29174_c0_g1, TRINITY_DN3204_c0_g1, TRINITY_DN37921_c0_g1, TRINITY_DN5631_c0_g1, TRINITY_DN5734_c0_g1, TRINITY_DN6526_c0_g1, TRINITY_DN6526_c0_g2 and TRINITY_DN6526_c0_g4) were identified in our study. The differences between the two groups showed almost the same as PAL. Six 4CLs were up-regulated trend and the others were down-regulated, which could reasonably explain the slightly higher relative content of phenylpropanoids and organic acids in the metabolic group results for TH from SS than those from FXS (Fig. 1E). Similarly, the 3 annotated chalcone isomerase (CHI) s also showed up-regulated and down-regulated expression. In Table S7, one caffeic acid 3-O-methyltransferase (COMT, TRINITY_DN14175_c0_g2) and one cinnamoyl-CoA reductase (CCR, TRINITY_DN20850_c0_g2) were DEGs. However, when combining metabolomic results with other DEGs that were primarily annotated at the downstream end of the synthetic pathway, differences in expression levels had little effect on this pathway.

Dihydroflavonoids can generate dihydroflavonol compounds such as dihydroquercetin and dihydrokaempferol under the action of flavanone-3-hydroxylase (F3H). The expression of F3H (TRINITY_DN3827_c0_g1) annotated in this study was up-regulated in TH from SS, which is consistent with the fact that the relative contents of dihydroflavone and dihydroflavonol in TH from SS were higher than those in TH from FXS. Dihydroflavonols can form flavonols again by the action of flavonol synthase (FLS). As a key enzyme for flavonol biosynthesis, FLS can convert dihydroflavonol to kaempferol and quercetin. In the present study, only two FLSs were identified, one of which was significantly up-regulated

in TH from SS with a high $|\text{Log}_2 \text{FC}|$ of 3.77. Based on the fact that the relative content of quercetin in TH from two hosts differed little, it was speculated that the causes of this phenomenon can be divided into two groups as follows. One of the catalytic enzymes involved in the upstream pathway for the formation of quercetin was expressed in highly close amounts in two groups and the other was that the same genes show two opposite expression trends. The combination of the two factors eventually leads to a minor difference in the relative content of quercetin. Anthocyanidin synthase (ANS, TRINITY_DN6705_c0_g1), anthocyanidin reductase (ANR, TRINITY_DN4530_c0_g1), dihydroflavonol 4-reductase (DFR, TRINITY_DN11549_c0_g1) and leucoanthocyanidin reductase (LAR, TRINITY_DN32506_c0_g1) involved in the catechin synthesis pathway were all down-regulated in TH from SS, which contradicted the high relative levels of metabolites detected in this pathway. Meanwhile, the corresponding glycoside contents of flavonoids (including flavonoid glycosides, flavonol glycosides, and isoflavone glycosides) were largely consistent with the changes in gene expression. The changes in UDP-glycosyltransferase expression were shown in Fig. S5 and Table S8. 22 UDP-glycosyltransferases were up-regulated and 28 were down-regulated in TH from SS, but none of the differences in expression were significant, which reasonably explains the slight differences in the overall content of flavonoid glycosides though.

Confirmation of related genes using qRT-PCR

The qRT-PCR results showed that compared with the FXS group, the expression of three of the nine candidate genes (HCT, FLS, and 4CL) were consistent with the transcriptome results while the rest showed the opposite trend (Fig. 6). The possible reason was that the expression levels of the selected genes between two groups were not significantly different (Table S7). Among them, FLS and 4CL were the key rate limiting enzyme genes in the upstream of the flavonoid synthesis pathway.

Analysis of transcription factor

Transcription factors (TFs) are widely found in living organisms as a class of proteins that bind to specific DNA sequences, recognize and bind to cis-acting elements in the upstream regulatory regions of genes through specific functional domains and have an activating or blocking effect on gene expression. TFs are usually composed of a DNA-binding domain (DBD), a trans-activating domain (TAD), and other signal sensing domains (SSD). Depending on the functional domains, TFs can be grouped into different transcription factor families. TF prediction and analysis were performed for DEGs and a total of 113 TFs were predicted (Table S9), with the highest percentage being the AP2/ERF family (19.47%), NAC family (19.47%), followed by the WRKY family (14.16%), MYB superfamily (9.73%), and bHLH family (5.31%). AP2/ERF is involved in the regulation of plant growth and development and plant primary metabolism, but it is also effective in the regulation of plant secondary metabolism, especially in the synthesis of major medicinal active ingredients (artemisinin, paclitaxel, etc.) in medicinal plants [26]. The bHLH family, on the other hand, is a key factor in the regulatory network of flavonoid biosynthesis. NAC, MYB and C2H2 families are generally considered to be associated with the regulation of metabolic and secondary metabolic biosynthesis in plants as well [27–29].

Discussion

According to the records of classical herb books, TH from SS is the most widely used and has better efficacy [30–32], which is also consistent with our previous findings [16]. Meanwhile, results from other researchers showed that there were differences in the synthesis and accumulation of major active components such as flavonoids and phenolic acids in TH from different hosts. However, the underlying molecular mechanism leading to differences in the biosynthesis and accumulation of these components have not been clarified, which severely limits the effective application of TH from other host sources and the cultivation of high-quality TH. In this work, we chose the most commonly used TH from SS and FXS at present and then performed an integrated transcriptomic and metabolite profiling analysis.

The annotated results showed that 54.77% of the genes had no annotation information, indicating that the genetic information of TH needs to be mined and fully investigated. As a non-model plant, there is relatively limited genetic information available for reference. At the same time, since various types of bioinformatics databases are still being updated and improved, there may still be some specific new genes that have not been discovered, therefore further in-depth studies are needed.

In metabolite analysis, a total of 81 constituents were identified including 37 flavonoids, 7 organic acids, 5 tannins, 4 phenylpropanoids, and 18 others, among which flavonoids and phenylpropanoids contained a higher proportion and primarily contributed to the bioactive benefits of TH. Therefore, we chose these two types of compounds for analysis in the further transcriptomic study. It has been shown that differences in the expression of genes based on metabolic pathways can be used to investigate the regulation of secondary metabolite synthesis and accumulation at the molecular level during plant development [33].

Genes identified were involved in the synthesis of phenylpropanoid from steps involving the formation of chemical scaffolds to subsequent modifications, including PAL, 4CL, and CHI. Phenylalanine, a prerequisite for the biosynthesis of phenylpropane, was used to obtain cinnamic acid and coumaric acid by PAL. Coumaric acid is produced by the action of 4CL to form coumaroyl coenzyme A, which is then transferred to the flavonoid biosynthesis pathway by the action of CHS [34]. CHI is an influential bridge for the production of other flavonoids and serves as a key enzyme in the flavonoid metabolic pathway that converts naringenin chalcone to naringenin. It plays a crucial role in biosynthesis involving various defense products, plant antibacterial mechanisms, stress resistance, cell development and differentiation, pigment accumulation, and expression of exogenous genes [35]. The differences in the expression of the above biosynthetic genes and the late pathway enzymes FLS and F3H, etc, were not significant between the two groups. Furthermore, the corresponding glycosides of the flavonoids exhibited similar expression patterns. The synthesis and accumulation of flavonoid components in plants are synergistically regulated by multiple factors such as plant species, developmental stage, tissue site, and growth environment. Studies have shown that transcription factors such as R2R3-MYB and b HLH and transcription factor families such as NAC, WRKY, and b ZIP regulate the expression of key genes at the transcriptional level, which in turn affect the synthesis and accumulation of flavonoids [35, 36]. On the

whole, the expression of genes annotated to the upstream and downstream of the flavonoid biosynthesis pathway did not differ significantly for the most part, with significant differences only in the downstream of the phenylpropanoid biosynthesis pathway. Up-regulation or down-regulation of enzymes involved in these two pathways might regulate downstream product synthesis either positively or negatively. This could be used to reasonably explain that the TH from SS and FXS are undoubtedly not significantly different in terms of secondary metabolites.

Conclusions

In summary, genes annotated to phenylpropanoid and flavonoid biosynthesis in TH were identified by comparative transcriptomics. The pattern of gene expression in both groups was basically consistent with the trend of accumulation of related compounds in the metabolome. However, the non-significant difference in gene expression in the two groups combined with metabolomics further illustrated that the differences between TH from two hosts may indeed be tiny. Unraveling the molecular mechanism of bioactive constituent accumulation in TH from different hosts could contribute to the quality control and selection of hosts of TH. This study reveals the relationship between gene regulation and flavonoid and phenylpropanoid biosynthesis in TH, providing basic information on the difference in molecular mechanism for the quality formation of TH from different hosts.

Abbreviations

TH

Taxilli Herba

SS

Morus alba L.

FXS

Liquidambar formosana Hance

DEGs

Different expressed genes

UFLC-Triple TOF-MS/MS

Ultra-fast liquid chromatography coupled with triple quadrupole-time of flight tandem mass spectrometry

RNA-Seq

RNA-sequencing

PBS

Phosphate-buffered saline

NR

NCBI protein non-redundant

COG

Clusters of Orthologous Groups of proteins

GO

Gene ontology
KEGG
Kyoto Encyclopedia of Genes and Genomes
FPKM
Fragments per kilobases per million reads
FDR
False discovery rate
qRT-PCR
Quantitative real-time polymerase chain reaction
PAL
Phenyl ammonium lyase
4CL
4-coumarate-CoA ligase
CHI
Chalcone isomerase
F3H
Flavanone-3-hydroxylase
FLS
Flavonol synthase
ANS
Anthocyanidin synthase
ANR
Anthocyanidin reductase
DFR
Dihydroflavonol 4-reductase
LAR
Leucoanthocyanidin reductase
TFs
Transcription factors
DBD
DNA-binding domain
TAD
Trans-activating domain
SSD
Signal sensing domain.

Declarations

Supplementary information

The online version contains supplementary material available at

Supplementary Material: **Table S1; Table S2; Table S3; Table S4; Table S5; Table S6; Table S7; Table S8; Table S9; Figure S1; Figure S2; Figure S3; Figure S4; Figure S5.**

Acknowledgments

Not applicable

Author's contributions

YJH, LXH and CZC designed and supervised the project. LL,CJM and LXH contributed to the experimental materials. YJH, WN, ZYY, CHJ,WWX and XJ performed the experiment. YJH, CCH and CZC carried out the data analyses. YJH wrote the manuscript. CZC and LXH revised the manuscript. All of the authors discussed the results and commented on the manuscript. The authors read and approved the final manuscript.

Funding

This work was supported by Guangxi Key Laboratory of Zhuang and Yao Ethnic Medicine [grant number GXZYKF2020-2].

Availability of data and materials

The datasets generated and/or analysed during the current study are available in the National Center for Biotechnology Information (NCBI) Sequence Read Archive Database repository, [<https://www.ncbi.nlm.nih.gov/bioproject/PRJNA904042>].

Ethics approval and consent to participate

The use of plant materials in the paper fully complied with our institutional guidelines and legislation.

Consent for publication

Not applicable

Competing interests

The authors declare that they have no competing interests.

Statement

The authors confirmed that the collection of plant material complies with relevant institutional, national, and international guidelines and legislation.

References

1. State Pharmacopeia Committee of China. Pharmacopoeia of the People's Republic of China. Bei Jing: Chemical Industry Press, 2020; chapter1. p. 312.
2. Li XF, Fang RZ, Feng H, Wang XP. Study on the mechanism of *Taxillus Chinensis* (DC.) Danser of tonifying liver and kidney, strengthening muscles and bones based on network pharmacology. Chin J Ethnomed Ethnopharm. 2021;30(06):16–
3. Lv L. Zhu, YM, Xu MD. Study on flavones from *Taxillus Chinensis* (DC.) Danser and assay of its quercetin. Chin. Tradit. Pat. Med. 2004;26(12):68–
4. Liang Y, Li L, Cai Y, Xu LB, Xie FF, Liang DL, et al. Analysis of chemical constituents in ethyl acetate extract of Taxilli Herba by UPLC-Q-Exactive-MS and screening of potential xanthine oxidase inhibitors. China J Chin Mater Med. 2022;47(04):972–979.
5. Huang FY, Liu RY, Chai ZS, Su BW, Zhu KX, Lu HL, et al. Study on aromatic components in essential oils of *Taxillus Chinensis* from different cold/hot properties hosts by GC–MS. Asia–Pac Trad Med. 2019;15(10):61–66.
6. Huang FY, Qin ZM, Li YH. Relationship between *Taxillus Chinensis* Tea and the culture of south of of the five ridges. Asia-Pac Trad Med. 2018;14(09):83–84.
7. Zhu KX, Lu D, Pei HH, Zhao MH, Li YH. A survey on the distribution of Taxilli Herba and its host status in Guangxi. Guangxi J Tradit Chin Med. 2010;33(02):59–61.
8. Zhang H, Huang FY, Su BW, Zhu KX, Lu HL, Yin SG, et al. Impacts of different host trees on the quality of *Taxillus Chinensis*. Mod Tradit Chin Med Mat Medica-World Sci Tech. 2016;18(07):1182–1187.
9. Wu N, Yuan JH, Wang WX, Liu XH, Li L, et al. Simultaneous determination of multiple active constituents in Taxilli Herba by HPLC-QTRAP-MS/MS. J Instr Anal. 2022;41(08):1153–1162.
10. Huang FY, Liu RY, Lu HL, Li YH. Effect of different medicinal hosts on the total flavonoid content of Taxilli Herba. J Guangxi Uni Chin Med. 2017;20(03):34–36.
11. Su BW, Wang H, Li YH, Pei HH, Zhu KX, Lu D. Content analysis of avicularin, quercitrin and quercetin in Herba Taxilli from different host plants. J Int Pharm Res. 2017;44(07):738–742.
12. Liu RY, Huang FY, Guo M, Lu HL, Zhu KX, Qin WH, et al. Correlation study on chemical constituents of cardiac glycosides in *Taxillus chinensis* and its *Nerium indicum* host by UPLC-Q-TOF-MS/MS. China J Chin Mater Med. 2019;44(11):2283–2291.
13. Zhang XJ, Su BW, Li J, Li YH, Lu D, Zhu KX, et al. Analysis by RP-HPLC of mangiferin component correlation between medicinal *Loranthus* and their mango host trees. J Chromatogr Sci. 2014;52(1):1–4.
14. Xia YP, He CH, Chai ZS, Qin WH, Wu WX, Chen LY, et al. Acute toxicity and hepatotox- icity of aqueous extracts of Taxilli Herba from different hosts in zebrafish. Chin J Exp Tradit Med Form. 2021;27(21): 91–97.
15. Zhang H, Huang FY, Liu RY, Lu HL, Su BW, Zhu KX, et al. Effect of Taxilli Herba from different hosts on lowering blood pressure of spontaneously hypertensive rats. Chin Tradit Pat Med. 2018;40(02):249–254.

16. Yuan, JH, Cai ZC, Chen CH, Wu N, Yin SX, Wang WX, et al. A study for quality evaluation of *Taxilli* Herba from different hosts based on fingerprint-activity relationship modeling and multivariate statistical analysis. *Arabian J. Chem.* 2022;15:103933.
17. Lu SH, Lu YY, Tang L, Zhang FY, Su BW. Evaluation of the quality of *Taxillus chinensis* from different hosts sources by HPLC fingerprint combined with chemometrics. *China Pharm.* 2020;31(07):794–799.
18. Yin SG, Liu RY, Huang FY, Li YH, Li JY. Study on infrared spectra of *Taxillus chinensis* (DC.) Danser from different host trees. *Lishizhen Med Mater Med Res.* 2018;29(11):2629–2631.
19. Guo PY, Zeng CB, Liu CL, Wei SS, Huang JQ, Tang H, et al. Differential expression analysis of genes related to flesh color in *Hylocereu polyrhizus* and *Hylocereu undatus*. *Molecular Plant Breeding. Mol Plant Breed.* 2021;19(13):4311–
20. Zhou H, Zhang X, Liu TY, She FX. Data processing and gene discovery of high-throughput transcriptome sequencing. *Jiangxi Sci.* 2012;30(5):607–
21. Zhu D, Li LH, Gao H. Identification of genes in response to drought and high temperature stress in *Mentha haplocalyx* seedlings based on transcriptome sequencing. *Shengming De Huaxue.* 2020;40(12):2270–
22. Yang ZM, Wu YP, Dai ZG, Chen XJ, Wang HQ, Yang S, et al. Comprehensive transcriptome analysis and tissue-specific profiling of gene expression in jute (*Corchorus olitorius* L.). *Ind Crops Prod.* 2020;146:112101.
23. Terol J, Nueda MJ, Ventimilla D, Tadeo F, Talon M. Transcriptomic analysis of *Citrus clementina* mandarin fruits maturation reveals a MADS-box transcription factor that might be involved in the regulation of earliness. *BMC Plant Biol.* 2019;19(1):47–66.
24. Zhang FS, Zhang X, Luo YY, Li HJ, Qin XM. Biosynthetic mechanisms of isoflavone accumulation affected by different growth patterns in *Astragalus mongholicus* products. *BMC Plant Biol.* 2022;22(1):410–428.
25. Yuan JH, Li L, Cai ZC, Wu N, Chen CH, Yin XS, et al. Qualitative analysis and componential differences of chemical constituents in *Taxilli* Herba from different hosts by UFLC-Triple TOF-MS/MS. *Molecules.* 2021;26(21):6373–6397.
26. Hong L, Yang L, Yang HJ, Wang W. Research Advances in AP2/ERF Transcription Factors in Regulating Plant Responses to Abiotic Stress. *Chin. Bull. Bot.* 2020;55(04):481–496.
27. Duan LP, Xie LQ, Ma YM, He J. Regulation of secondary metabolism by transcription factors in medicinal plants. *J Chin Me. Mater.* 2021;44(04):1002–1007.
28. Wang XP, Niu YL, Zheng Y. Multiple Functions of MYB Transcription Factors in Abiotic Stress Responses. *Int J Mol Sci.* 2021;22(11):6125.
29. Liu MQ, Sun W, Meng XX, Wang HH, Liu TX, Sun JY, Wang Z, et al. Identification and expression analysis of the C2H2 gene family in *Cannabis sativa* L. *Acta Pharm Sin.* 2021;56(05):1486–1496.
30. Tao HJ. Annotation of *Materia Medica* [M]. Beijing: People's Medical Publishing House, 1994;254–255.

31. Tang SW. Classified Materia Medica [M]. Beijing: China Press, 1993;359–360.
32. Li SZ. Compendium of Materia Medica: Vol [M]. Beijing: People's Medical Publishing House, 1982;2158–2159.
33. Wang CC, Chen LH, Cai ZC, Chen CH, Liu ZX, Liu SJ, et al. Metabolite Profiling and Transcriptome Analysis Explains Difference in Accumulation of Bioactive Constituents in Licorice (*Glycyrrhiza uralensis*) Under Salt Stress. *Front Plant Sci.* 2021;12:727882.
34. Chen X, Liu YC, Chen Y, Long YQ, Tong QZ, Liu XD, et al. Cloning and expression analysis of phenylalanine ammonia-lyase gene in *Lonicera macranthoides*. *Chin Tradit Herb Drugs.* 2019;50(01):178–187.
35. Heldt HW, Piechulla B. Phenylpropanoids Comprise a Multitude of Plant-Specialized Metabolites and Cell Wall Components. *Plant Biochem.* 2021;15:411–429.
36. Wang Y, Yang X, Yang RJ, Wang YX, Yang YX, Xia PF, et al. Advances in research of MYB transcription factors in regulating phenylpropane biosynthesis. *J Anhui Agric Univ.* 2019;46(05):859–864.

Figures

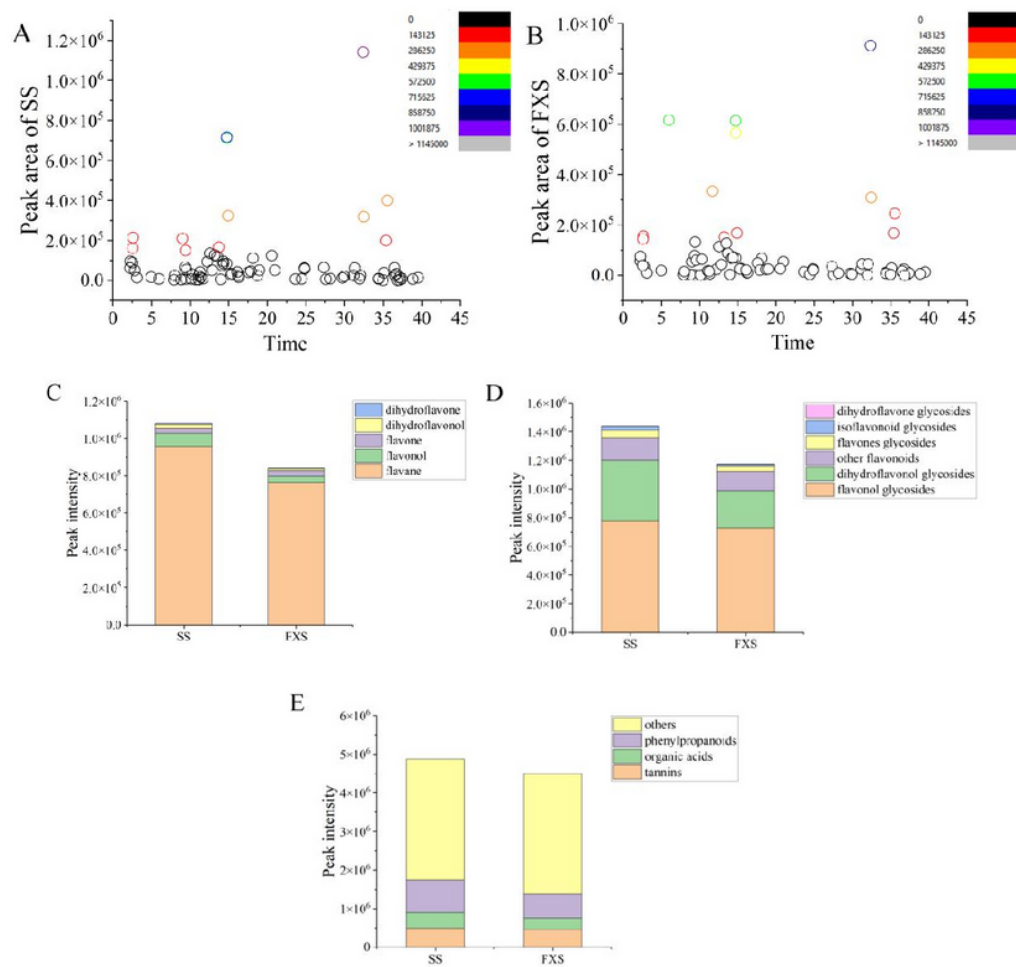


Figure 1

Results of metabolomics analysis. **A** Distribution of constituents identified in Taxilli Herba from SS. **B** Distribution of constituents identified in Taxilli Herba from FXS. **C** Total content changes in flavonoids. **D** Total content changes in flavonoid glycosides. **E** Total content changes in phenylpropanoids, organic acids, tannins and others between two groups.

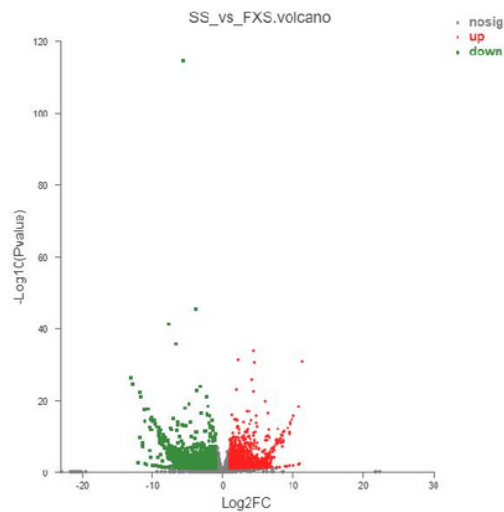


Figure 2

Volcano map analysis of DEGs (red points represent up-regulated genes, green points represent down-regulated genes and unchanged genes are nosing gray.)

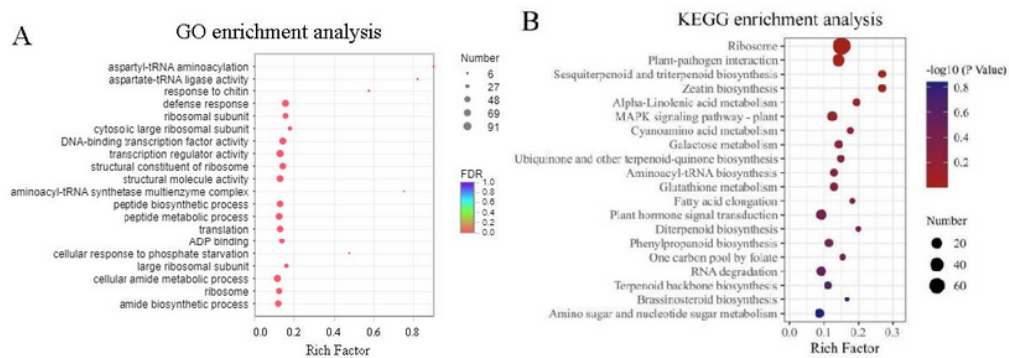


Figure 3

Functional annotation and enrichment analysis. **A** Gene ontology (GO) enrichment of top 20 terms between SS and FXS groups. **B** Kyoto of Genes and Genomes (KEGG) enrichment of top 20 pathways SS and FXS groups.

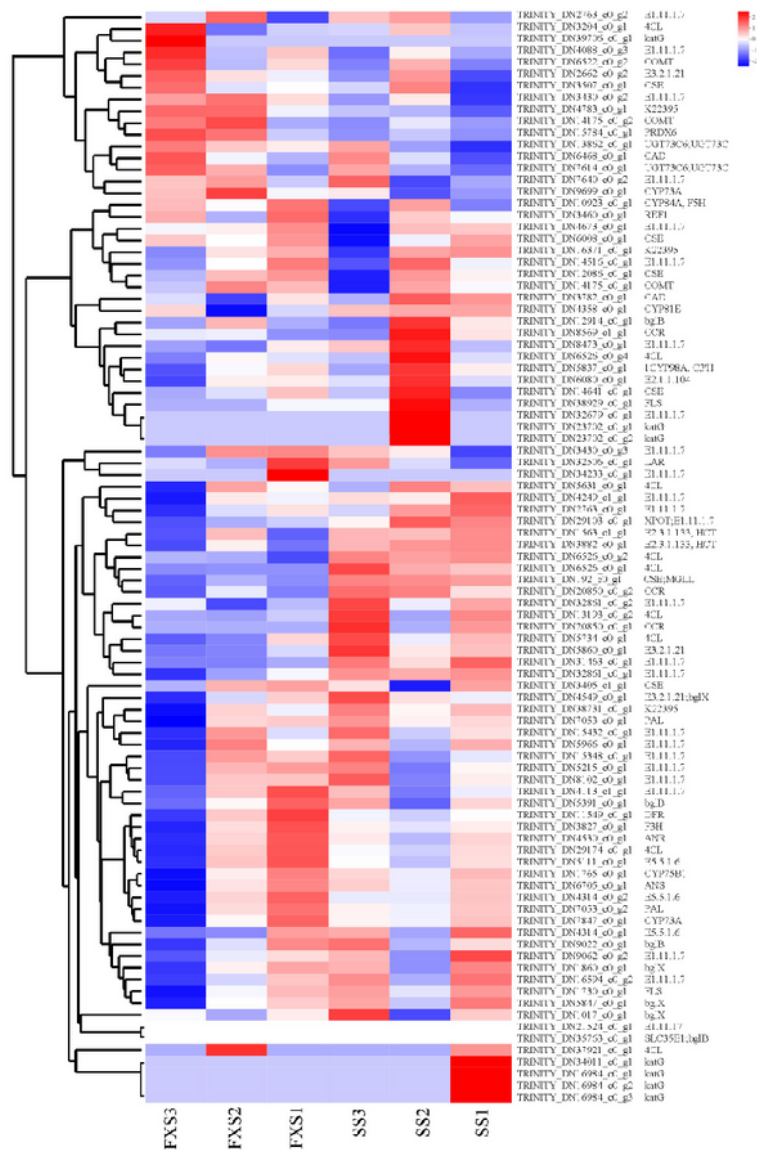


Figure 4

Heatmap of genes related to phenylpropanoid and flavonoid biosynthesis pathways in TH.

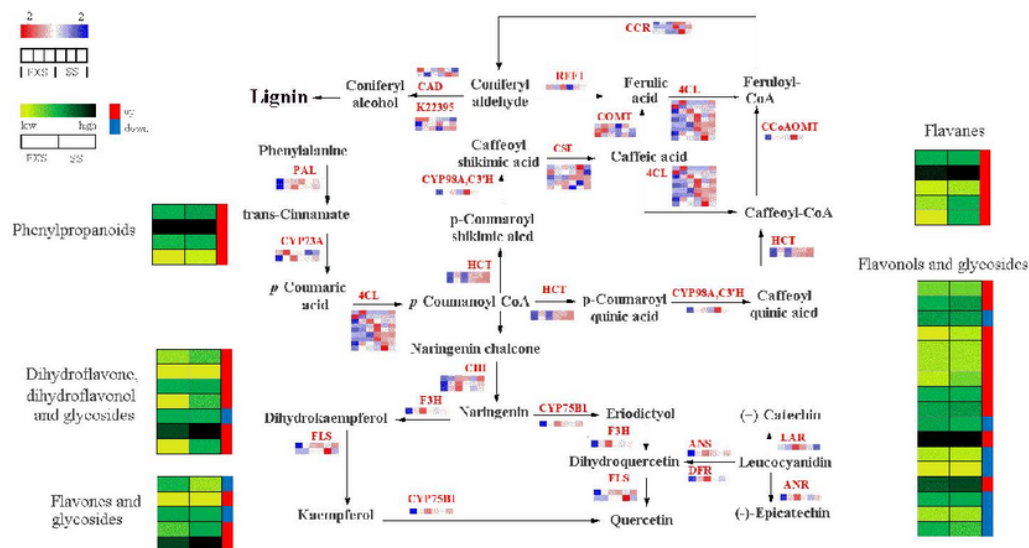


Figure 5

Correlation analysis of expression pattern of genes and metabolite profiles of phenylpropanoid and flavonoid in TH.

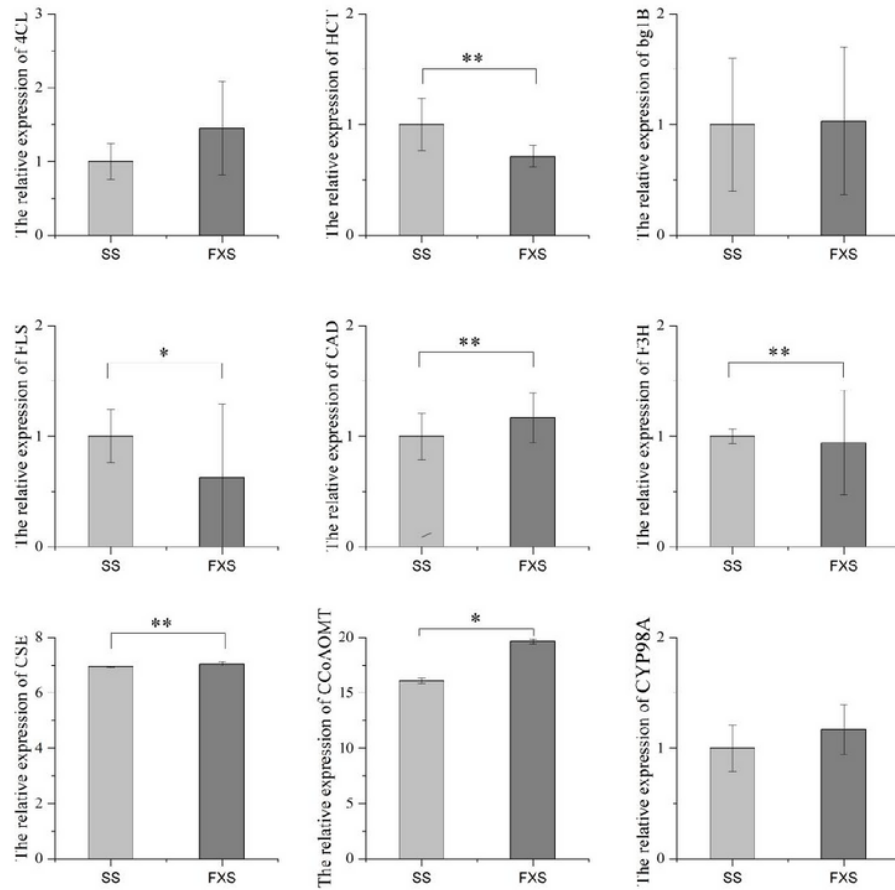


Figure 6

Expression patterns of genes (4CL, HCT, bg1B, FLS, CAD, F3H, CSE, CCoAOMT, CYP98A) between SS and FXS groups (The means $\pm$ SD; n=3; ** $p < 0.01$; * $p < 0.05$)

Supplementary Files

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

- [SupplementaryMaterials.pdf](#)