Comparative proteomics analysis reveals the molecular mechanisms underlying the accumulation difference of bioactive constituents in Taxilli Herba from two hosts

Jiahuan Yuan[1], Nan Wu[1], Zhichen Cai[1], Cuihua Chen[1], Yongyi Zhou[1], Haijie Chen[1], Jia Xue[1], Xunhong Liu[1], Wenxin Wang[1], Jianming Cheng[2], and Li Li[3]

[1]Nanjing University of Chinese Medicine School of Pharmacy
[2]Jiangsu Province Engineering Research Center of Classical Prescription
[3]GuangXi University of Chinese Medicine

March 30, 2023

Abstract

Taxilli Herba (TH) is famous for its semi-parasitic feature, with hundreds of host species. The molecular mechanisms governing the accumulation of bioactive constituents in TH from different hosts have yet to be completely understood. The purpose of this study was to investigate the differences in synthesis and accumulation of bioactive constituents in TH from Morus alba L. and Liquidambar formosana Hance. The proteome of TH from two hosts were compared, and proteins were identified using tandem mass tag (TMT) technology. Sixteen bioactive constituents were quantitatively determined and analyzed in conjunction with the results of proteomics. A total of 5681 proteins were detected, of which 259 were significantly up-regulated and 274 significantly down-regulated in expression levels. Differentially expressed proteins (DEPs) were shown to be mostly engaged in cellular process and metabolic process by GO analysis. DEPs were considerably abundant in amino acid metabolic pathways, according to KEGG pathway analysis. The changes in the contents of 16 bioactive constituents were mostly compatible with the expression patterns of relevant structural proteins. The results of this study will provide basic information for the role of protein in the secondary metabolism biosynthesis pathway and the molecular mechanism of the quality formation of TH.

Comparative proteomics analysis reveals the molecular mechanisms underlying the accumulation difference of bioactive constituents in Taxilli Herba from two hosts

Jiahuan Yuan[1], Nan Wu[1], Zhichen Cai[1], Cuihua Chen[1], Yongyi Zhou[1], Haijie Chen[1], Jia Xue[1], Xunhong Liu[1,2,*], Wenxin Wang[1], Jianming Cheng[1,2], Li Li[3]

[1] College of Pharmacy, Nanjing University of Chinese Medicine, Nanjing, Jiangsu, China

[2] Jiangsu Province Engineering Research Center of Classical Prescription, Nanjing, Jiangsu, China

[3] College of Pharmacy, Guangxi University of Chinese Medicine, Nanning, Guangxi, China

yuanjiahuan1027@163.com (J.Y.); wunan7272@163.com (N.W.); caizhichen2008@126.com (Z.C.); cuihuachen2013@163.com (C.C.); zyy18851091328@163.com (Y.Z.); chenhaijie039X@163.com (H.C.); 17860505214@163.com (J.X.); 300100@njucm.edu.cn (X.L.); wangwenxin66666666@163.com (W.W.); cjm7895@163.com (J.C.); Lil2014@gxtcmu.edu.cn (L.L.)

A list of abbreviations

AGC: Automatic Gain Control

BP: Biological Process

BBP: Benzyl Butyl Phthalate

CC: Cellular Component

DDA: Data Dependent Acquisition

DEPs: Differentially Expressed Proteins

FXS: *Liquidambar formosana* Hance

HCD: High Energy Collisional Dissociation

KEGG: Kyoto Encyclopedia of Genes and Genomes

MF: Molecular Function

SS: *Morus alba* L.

TEBA: Triethylammonium Bicarbonate Buffer

TH: Taxilli Herba

TMT: Tandem Mass Tag

UFLC-QTRAP-MS/MS: ultrafast liquid chromatography simultaneously coupled with triple quadrupole/linear ion trap tandem mass spectrometry

Keywords: bioactive constituents, comparative proteomics, hosts, molecular mechanisms, metabolomics, Taxilli Herba

Total number of words: 5962 words

Abstract

Taxilli Herba (TH) is famous for its semi-parasitic feature, with hundreds of host species. The molecular mechanisms governing the accumulation of bioactive constituents in TH from different hosts have yet to be completely understood. The purpose of this study was to investigate the differences in synthesis and accumulation of bioactive constituents in TH from *Morus alba* L. and *Liquidambar formosana* Hance. The proteome of TH from two hosts were compared, and proteins were identified using tandem mass tag (TMT) technology. Sixteen bioactive constituents were quantitatively determined and analyzed in conjunction with the results of proteomics. A total of 5681 proteins were detected, of which 259 were significantly up-regulated and 274 significantly down-regulated in expression levels. Differentially expressed proteins (DEPs) were shown to be mostly engaged in cellular process and metabolic process by GO analysis. DEPs were considerably abundant in amino acid metabolic pathways, according to KEGG pathway analysis. The changes in the contents of 16 bioactive constituents were mostly compatible with the expression patterns of relevant structural proteins. The results of this study will provide basic information for the role of protein in the secondary metabolism biosynthesis pathway and the molecular mechanism of the quality formation of TH.

Statement of significance of the study

Taxilli Herba (TH) is a kind of widely used traditional Chinese medicine with the characteristic of semi-parasitism. Host is an important factor affecting the quality of TH. However, most of the medicinal materials circulating and used in the market are of multi-host sources, with uneven quality. Meanwhile, the study on the molecular mechanism on the biosynthesis and accumulation of bioactive constituents in TH from different hosts remain unclear, which seriously hinders the comprehensive quality evaluation, development and utilization of TH. Therefore, it is essential to explore the the underlying molecular mechanisms. We integrated TMT-base proteomics and metabolomics to investigate the trends in the changes of proteins and

metabolites involved in flavonoid pathway, and constructed a putative network of flavonoid biosynthesis in TH.

Our results provide a theoretical foundation for the selection of TH on hosts by examining the protein differences between the two host-derived TH and the functions of proteins in the biosynthesis pathway of bioactive constituents.

1 Introduction

Traditional Chinese medicine has traditionally utilized Taxilli Herba (TH), which is composed of the dried stems and branches with leaves of*Taxillus chinensis* (DC.) Danser [1]. It is distributed in Guangxi, Yunnan, Sichuan, Gansu, and other provinces in China [2]. Shennong Materia Medica classified TH as a top grade medication and listed it as being used to treat rheumatic arthralgia, waist and knee weakness, among other symptoms [3]. Meanwhile, multiple types of chemical constituents have been identified, including flavonoids, phenolic acids, and tannins, etc. These constituents act synergistically to exert multiple pharmacological effects such as anti-inflammatory and analgesic [4,5]. Additionally, TH tea has a number of health advantages as well as significant commercial growth potential [6].

The diversity of hosts is an influential biological feature of TH. The hosts of TH used are mainly *Morus alba* L. (SS) and*Liquidambar formosana* Hance (FXS) [7,8]. It is worth noting that the relationship between hosts and the quality of TH has attracted the attention of experts and scholars in recent years, and relevant studies have been carried out one after another in terms of chemical constituents and pharmacology [9–15]. Since TH takes water, inorganic salts, and nutrients from its hosts, the host has a significant influence on how it grows and develops [16]. To the best of our knowledge, molecular changes and elucidating the mechanisms underlying the accumulation of bioactive constituents in TH remain vacant. Therefore, it is necessary to analyze the differences at protein level and investigate the proteins associated with the biosynthesis of these constituents in TH from two hosts.

The proteomics of medicinal plants is a useful tool for understanding mechanisms, metabolic pathways, and functional gene mining [17,18]. Studies are gradually expanding to include non-model plants instead of only model plants such Arabidopsis, rice, and soybean [19–23]. In this study, mapping of proteins and related bioactive constituents in metabolic pathways was investigated using tandem mass tag (TMT)-based quantitative proteomics in TH from SS and FXS. First, a proteome database of TH was established by annotating protein sequences according to the related existing database. A total of 5681 proteins were detected, of which 533 were significantly differentially expressed proteins (DEPs). The DEPs were enriched in cellular process and secondary metabolite biosynthesis. Candidate proteins involved in flavonoid biosynthesis were selected for further validation. Meanwhile, ultrafast liquid chromatography simultaneously coupled with triple quadrupole/linear ion trap tandem mass spectrometry (UFLC-QTRAP-MS/MS) was applied to determine the content of sixteen bioactive constituents. Ultimately, parallel reaction monitoring (PRM) was carried out to verify the abundance of associated proteins. This work aimed to provide a theoretical foundation for the selection of TH on hosts by examining the protein differences between the two host-derived TH and the functions of proteins in the biosynthesis pathway of bioactive constituents.

2 Materials and methods

2.1 Plant materials

TH samples from SS and FXS were collected from Wuzhou of Guangxi Province in China. Three duplicates of TH samples from two hosts were examined. The botanical origins of the materials were authenticated by Professor Xun-Hong Liu (Department for Authentication of Chinese Medicines, School of Pharmacy, Nanjing University of Chinese Medicine, Nanjing, China). Voucher specimens were deposited in the laboratory of Chinese medicine identification, Nanjing University of Chinese Medicine. The medicinal parts were gathered, washed with PBS, frozen in liquid nitrogen, and refrigerated at –80 °C.

2.2 Protein extraction, digestion, and TMT labelling

1% PVPP and appropriate amount of BPP solution were added to the samples, and then agitated 3 times for 40 s each using a high-throughput tissue grinder. The supernatant was centrifuged at 1200 g for 20 min followed by tris-aturated phenol and BPP solution added and vortexed for 10 min at 4 . Protein was precipitated overnight at –20 degC in 5 times volumes of pre-chilled ammonium acetate methanol solution, and the following day it was centrifuged at 12000 g for 20 min at 4 degC. The precipitate was mixed with 90% pre-cooled acetone, and the supernatant was discarded and repeated twice. The obtained precipitate was dissolved with protein lysis solution. The protein concentrations were determined by the BCA method [24].

100 μg of protein from each sample was diluted with an appropriate amount of TEBA. Following the addition of TCEP, which was allowed to react at 37 °C for 60 min, 1AM was added and allowed to react at room temperature for 40 min. The alkylated proteins were precipitated by adding six times the volume of pre-cooled acetone at a temperature of –20 °C for 4 hours, followed by centrifugation at 10000 g for 20 min. 100 mL of 100 mM TEBA was used to completely dissolve the precipitate. In order to complete the enzymatic digestion, trypsin was added at a ratio of 1:50 (enzyme: protein) and left overnight at 37 °C.

TMT Reangent kit (Thermo, USA) was used to label the resultant peptides in accordance with the manufacturer's protocol (Thermo Fisher Scientific, Waltham, MA, USA). Briefly, 100 μg of peptide were mixed with a tube of TMT reagent and incubated at room temperature for 2 h. After mixed with hydroxylamine, the derivatization reaction was burst for 30 min. The TMT-labeled samples were then mixed and vacuum centrifuged dried. Six TMT-labeled samples were ultimately examined utilizing a Thermo Scientifir Vanquish F UHPLC system in combination with a Q-Exactive hybrid quadrupole orbital trap mass spectrometer (Thermo Fisher Scientific, Waltham, MA, USA).

2.3 High-pH reversed-phase separation

UPLC loading buffer was used to redissolve the digested peptides, an ACQUITY UPLC BEH C_{18} Column (1.7 μm, 2.1 mm × 150 mm, Waters, USA) was used to separate them at a flow rate of 200 $\mu L \cdot min^{-1}$. The mobile phase contained 2% acetonitrile (A, pH=10) and 80% acetonitrile (B, pH=10) with a gradient elution: 0–2 min, 0–5% solvent B; 2–17 min, 5% solvent B; 17–18 min, 5–10% solvent B; 18–35.5 min, 10–30% solvent B; 35.5–28 min, 30–36% solvent B; 38–39 min, 36–42% solvent B; 39–40 min, 42–100% solvent B. Finally, 14 fractions were created by combing a total of 28 fractions were collected according to their peak type and time.

2.4 Mass spectrometry analysis

The fraction was then centrifuged at 12000 g for 20 min at 4 °C after being lyophilized and redissolved with 0.1% formic acid in 2% acetonitrile (solvent A). A C_{18} (150 μm × 15 cm, Evosep) was used in conjunction with the Orbitrap Exploris 480 (Thermo, USA) and Evosep One (Evosep). Data were acquired by Thermo Xcalibur 4.0 (Thermo, USA). The peptide mixture underwent gradient elution using mobile phases of 0.1% formic acid (A) and 100% acetonitrile 0.1% formic acid (B) at a flow rate of 300 $nL \cdot min^{-1}$. The specific gradients are as follows: 0–2 min, 5–5% solvent B; 2–30 min, 5–38% solvent B; 30–40 min, 38–90% solvent B; 40–44 min, 90–100% solvent B. The mass spectrum parameters were set as follows: scanning range, 350–1500m/z for full scan; acquisition mode, data dependent acquisition (DDA); MS1 resolution, 60000; maximum injection time, 25 ms; fragmentation mode, high energy collisional dissociation (HCD); MS/MS resolution, 15000; maximum injection time, 22 ms; fixed first mass, 110m/z ; minimum automatic gain control (AGC), 8 e3; intensity threshold, 8.3 e4; dynamic exclusion time, 30 s.

2.5 Database search and quantitative proteomic analysis

The raw data were processed by Proteome Discoverer^{TM}(version 2.4, Thermo Fisher, Waltham, MA, USA) and then identified by Protein Pilot Software 5.0 (AB SCIEA, USA) in combination with SwissProt (version 2021.06)/UniProt (version 2021.09) databases. The search parameters are as follows: cys alkylation, iodoacetamide; static modification, carbamidomethyl; dynamic modification, oxidation of acetyl; proteolytic enzyme, Trypsin (Full); maximum number of missed cleavage allowed, 2; precursor mass tolerance, 20 ppm;

fragment mass tolerance, 0.02 Da; quantitative method, TMT 6Plex. For the quantification of TMT, the ratio of ion intensities reported by TMT in the MS/MS spectra of the original data set was used to calculate the fold change (FC) between samples. FC [?] 1.2 or [?] 0.83 and p -value < 0.05 were set as cut-off values to designate significantly differentially expressed proteins.

2.6 Bioinformatic analysis

Protein sequence alignment was performed by comparison with the databases Uniprot (version 2021.09), NCBI_NR (version 2021.09) and Swissprot (version 2021.06) databases. Principal component analysis (PCA) was carried out to distinguish the samples from two hosts. Gene Ontology (GO, version 2021.0918) was used for functional annotation to learn the functions of proteins in biological process (BP), cellular component (CC) and molecular function (MF) by BLAST2GO (version 2.5.0). Kyoto Encyclopedia of Genes and Genomes (KEGG, version 2021.09) was used to explain the involvement of proteins in biological pathways by KOBAS (version 2.1.1). Goatools (version 0.6.5) and Python (version 3.11.2) were applied for enrichment analysis. Hierarchical clustering analysis of protein was performed by Cluster (version 3.0) based on Pearson's correlations.

2.7 Parallel reaction monitoring analysis

Majorbio Bio-pharm Technology Co., Ltd. utilized PRM to validate the abundance of 17 candidate proteins related to the synthesis of secondary metabolites (Shanghai, China). For the PRM analysis, three biological replicates from each group were considered. Based on the TMT data, the signature peptides of the targeted proteins were identified, and only distinct peptide sequences were chosen. The chromatographic column, orbit trap and electrospray voltage, protein identification library, and raw data processing software used for the LC-MS analysis in this investigation were identical to those used for the TMT analysis.

2.8 Bioactive constituents analysis

The sample preparation and analysis procedures were carried out in accordance with prior studies' descriptions [14]. UFLC-QTRAP-MS/MS (AB SCIEX, Framingham, USA), operating in negative and positive modes with multiple reaction monitoring, was used to evaluate the samples (MRM). The previous conditions were also present, along with others.

2.9 Statistical analysis

The t-test was used for statistical comparisons across groups by SPSS (version 21.0), and the findings were expressed as the the mean +- standard deviation of three biological replicates, with p -values < 0.05 denoting statistical significance.

3 Results

3.1 Protein detection, identification, and primary data

According to TMT data, a total of 33204 spectrum and 22138 unique peptides were identified, and 5681 proteins were detected (Figure S1A and Table S1). The functional annotation of proteins was displayed in Figure S1B. The majority of the identified peptides' mass error-matching distribution within were 10 ppm (Figure S2A), indicating good accuracy and reliability for the identification of peptides. After sufficient proteolysis, peptides typically contained 6 to 25 amino acids (Figure S2B), showing that the lengths of the peptides was reasonable. Additionally, over 50% of the identified proteins had a molecular weight distribution of 20–80 kDa (Figure S3A), while less than 30% of the identified protein had sequence coverage (Figure S3B). The isoelectric point of identified proteins generally fell between 4.5 and 10. The overall expression trends of all identified proteins are shown in Figure S4.

Three biological replicates were clustered after the PCA analysis, which demonstrated that TH samples from two hosts could be differentiated successfully and had acceptable repeatability (Figure 1A). Figure 1B displayed the correlation between the two sample groups. 533 proteins in total were significantly differentially

expressed, 259 of which were up-regulated and 274 of which were down-regulated in TH from FXS (Figure 1C and Table S2).

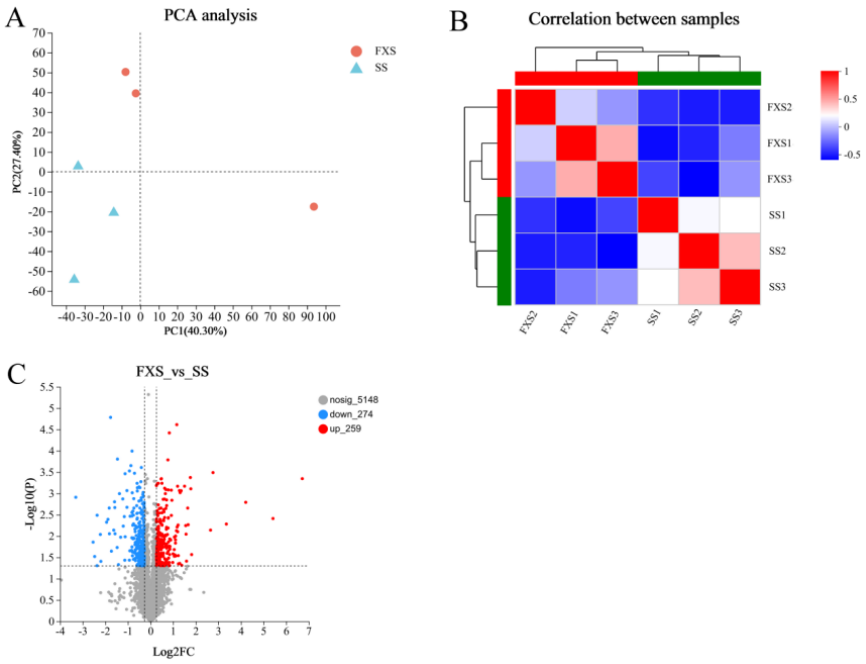

FIGURE 1 (A) Principal component analysis. (B) Correlation heat map of STH samples from SS and FXS groups. (C) Volcano map analysis of significantly differentially expressed proteins (red point represents up-regulated proteins, blue point represents down-regulated proteins, and unchanged proteins are in nosing gray).

3.2 GO functional annotation classification and enrichment analysis of DEPs

The DEPs was annotated into BP, CC, and MF by GO annotation analysis. The top 20 GO terms were calculated on the basis of ascending$p$ -value ranking (Figure 2A and Table S3). The DEPs were mainly involved in BP, of which cellular process and metabolic process accounted for a higher proportion. Apart from this, biological regulation and response to stimulus account for the remaining portion. It indicated that the cells of TH from two hosts differed in physiological status somewhat. Furthermore, under the CC category, proteins were exclusively connected to cellular anatomical entity and protein-containing complex. Most DEPs participated in catalytic activity while others were related to binding and transporter activity during MF. It was evident that in TH from SS and FXS, the proteins with catalytic and binding functions undergo changes in major functions such as metabolism and cellular processes in cellular anatomical entity and protein-containing complex. In Figure 2B and Table S4, the top 20 GO enrichment histogram of the DEPs was presented. There were eight items enriched in BP, among which GO: 0006952 (defense response) and GO: 0042440 (pigment metabolic process), and GO: 0033013 (tetrapyrrole metabolic process) were enriched in more proteins (23, 13, and 11, respectively). There were 11 items that were enriched in MF, two of which had lower p -values than the others (GO: 0003824 (catalytic activity and GO: 0090729 (toxin activity) were lower. There was one entry enriched in CC, with GO: 0005887 (integral component of plasma membrane).

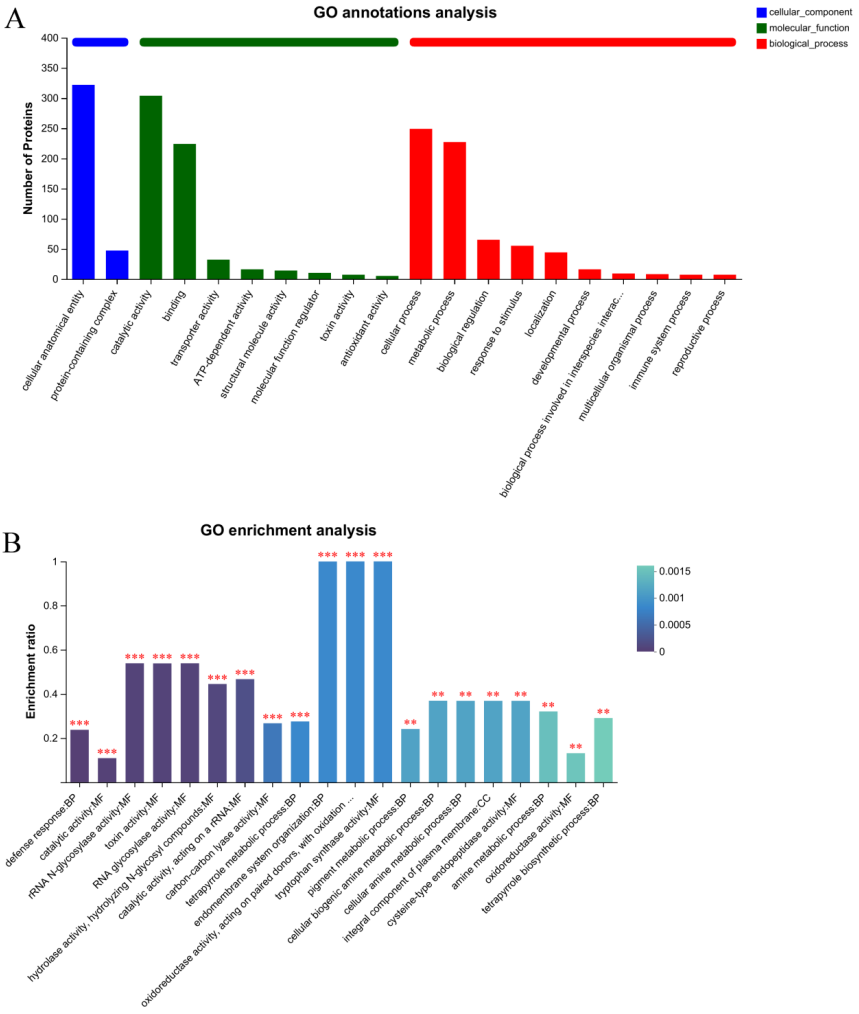

FIGURE 2 (A) GO terms of DEPs (biological process, cellular component, and molecular function). (B) Enrichment of DEPs. (** $p< 0.01$; *** $p < 0.001$).

3.3 KEGG functional annotation classification and enrichment analysis of DEPs

The DEPs were also investigated using the KEGG database. Figure 3A revealed that the DEPs were annotated on five primary metabolic pathways, including metabolism, genetic information processing, environmental information processing, cellular process and organismal systems (Table S5). A large number of proteins were annotated to pathways such as carbohydrate metabolism, amino acid metabolism, metabolism of cofactors and vitamins. As can be seen in Figure 3B, in terms of primary metabolism, proteins were enriched in various amino acids metabolism including glycine, serine and threonine metabolism and so on, as well as glycolysis/gluconeogenesis, indicating that the physiological activity of TH from FXS was vigorous. For the secondary metabolism, a total of 9 DEPs (6 down-regulated and 3 up-regulated) were associated with flavonoids biosynthesis with no significant protein enrichment in pathway. Overall, the quality disparities in TH from different hosts were primarily caused by DEPs implicated in biosynthetic pathways of bioactive constituents (Table S6).

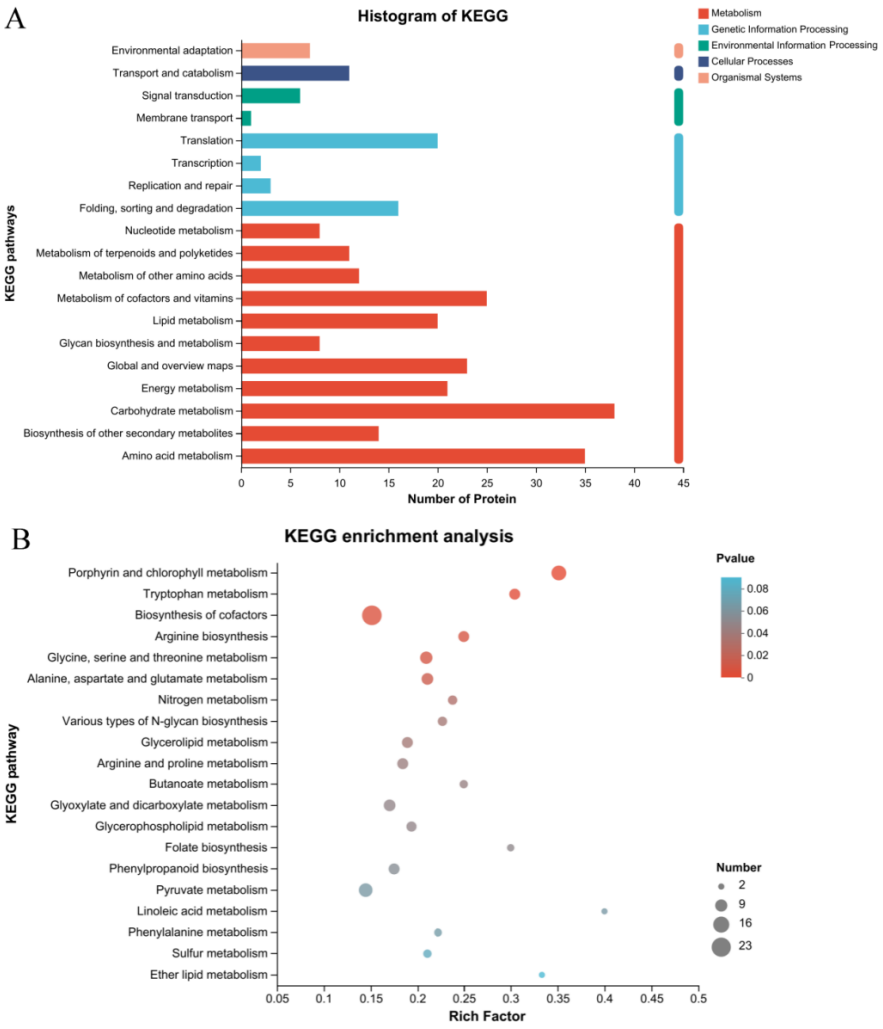

FIGURE 3 (A) KEGG pathway classification. (B) KEGG enrichment of DEPs.

3.4 Proteins involved in flavonoid biosynthesis

Proteins related to flavonoid biosynthesis were clustered to better explore the molecular mechanisms underlying the changes in the accumulation of secondary metabolites. Figure 4 and Figure S5 showed the clustering results of the proteins involved in the above pathway, and the specific positions were mapped, respectively. A total of 51 proteins were identified, of which 32 were up-regulated and 19 were down-regulated in TH from FXS (Table S7).

PAL (EC: 4.3.1.24) and 4CL (EC: 6.2.1.12) are common precursors in the flavonoid biosynthsis pathway. PAL functions as a key rate-limiting enzyme that catalyzes the phenylalanine to trans-cinnamic acid, which was up regulated in TH. 4CL is the third key enzyme in dihydroflavone biosynthesis as well. Four 4CLs (TRINITY_DN29174_c0_g1_i1_orf1, TRINITY_DN3204_c0_g1_i2_orf1, TRINITY_DN5631_c0_g1_-i2_orf1, TRINITY_DN6526_c0_g1_i1_orf1) were identified in TH, where two and two were up-regulated and down-regulated, respectively. CHI (EC: 5.5.1.6) is the second key enzyme in this pathway, which is important for the regulation of the synthesis of the entire flavonoid secondary metabolites, and the expression of CHI gene can provide sufficient precursors for the synthesis of flavonoids [25]. 3 CHIs (TRINITY_DN5111_-c0_g1_i4_orf1, TRINITY_DN4314_c0_g2_i2_orf1, TRINITY_DN4314_c0_g2_i1_orf1) were identified, while the expression trend was not consistent. The midstream enzymes were annotated to FLS (EC: 1.14.20.6), HCT

(EC: 2.3.1.133), F3H (EC: 1.14.11.9), CYP75B1 (EC: 1.14.14.82), and C3'H (EC: 1.14.14.96, CYP98A). FLS directly determines the content of flavonols in plant tissues and serves as a bridge between the flavonoid synthesis pathway and the catechin synthesis pathway as well [26]. The downstream enzymes exhibited not only increasing trends but also decreasing trends. Other enzymes along this pathway have been annotated, including ANS (EC: 1.14.20.4), ANR (EC: 1.3.1.77), LAR (EC: 1.17.1.3), DFR (EC: 1.1.1.219), and UGT73C6. The expression of upstream enzymes in phenylpropanoid biosynthesis was both up and down regulated, while the expression of enzymes in the downstream pathway was predominantly down-regulated in TH from SS. The differences in the expression of proteins annotated to the flavonoid biosynthesis were not significant, they still deserved more attention because they might act synergistically with enzymes in the upstream phenylpropanoid biosynthesis to regulate flavonoid synthesis and accumulation.

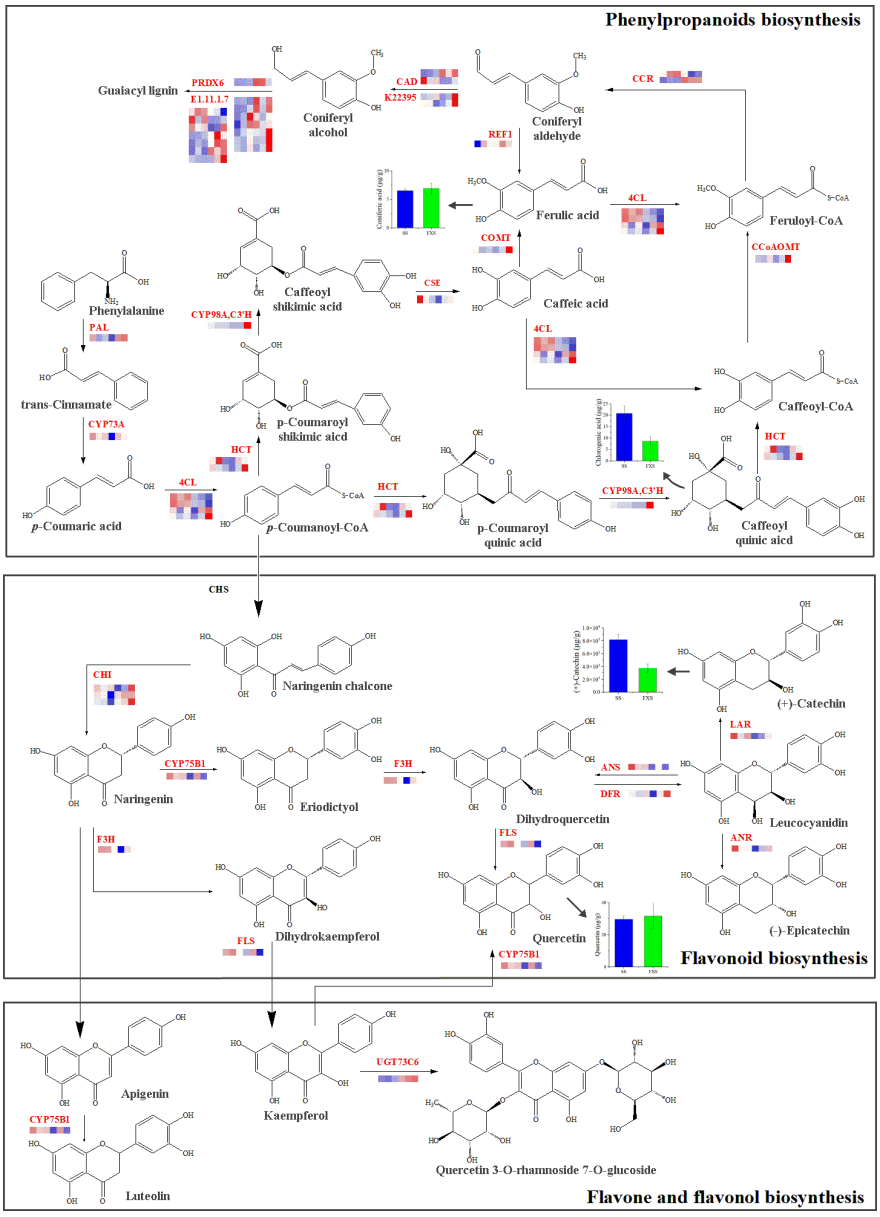

FIGURE 4 Putative biosynthesis pathways and expression of proteins involved in flavonoid.

3.5 Determination of multiple bioactive constituents

The main bioactive components in TH are reported to be phenolic acids and flavonoids [27,28]. To investigate the potential relationship between metabolite accumulation and protein expression, the content of main active constituents was determined simultaneously by UFLC-QTRAP-MS/MS. The condition of UFLC-QTRAP-MS/MS and method validation studies were reported in a previous study [14]. A total of 16 constituents including 12 flavonoids and 4 phenolic acids were quantitated in TH from two SS and FXS. The contents of major flavonoid glycosides (quercitrin, hyperoside, auicularin, quercetin-3-*O* -β -*D* -glucuronide, and rutin), and (+)-Catechin in TH from SS were higher than those in TH from FXS. There was no significant difference between the two groups in the content of other constituents (Figure 5 and Table S8).

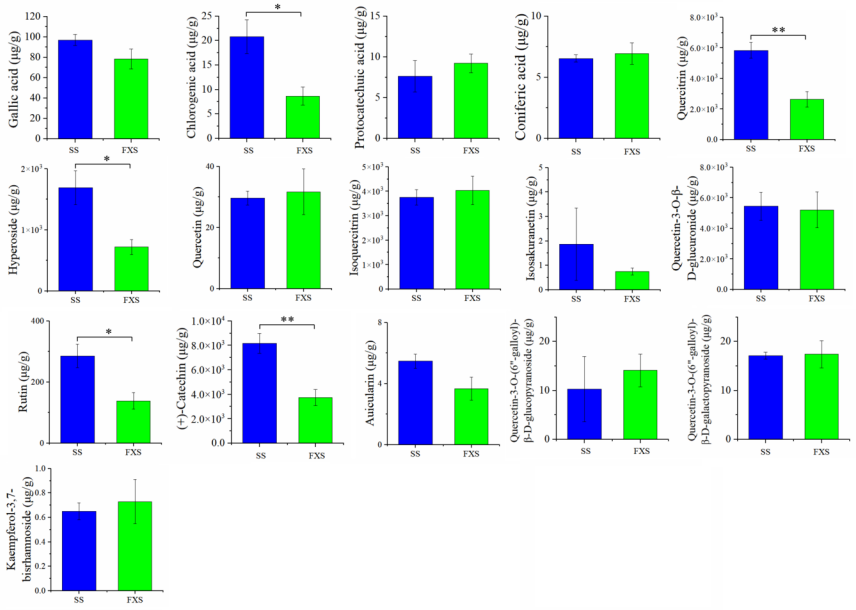

FIGURE 5 Variations in 16 bioactive constituents in the Taxilli Herba from SS and FXS. (mean ± SD; * $p < 0.05$; ** $p < 0.01$)

3.6 Parallel reaction monitoring verification

To verify the abundance of candidate proteins, PRM was used to conformed validate 17 proteins (COMT, HCT, PRDX6, FLS, REF1, F3H, ANR, C3'H, PAL, CYP75B1, LAR, CHI, bglB, CCoAOMT, ANS, CYP73A, and E1.11.1.7), as determined by TMT data. Only the PRM results of PAL, COMT, HCT, PRDX6, REF1 and C3'H were accordance with the proteomic data (Figure 6). However, most of the fold changes in protein abundance detected by PRM were larger than those detected by TMT data.

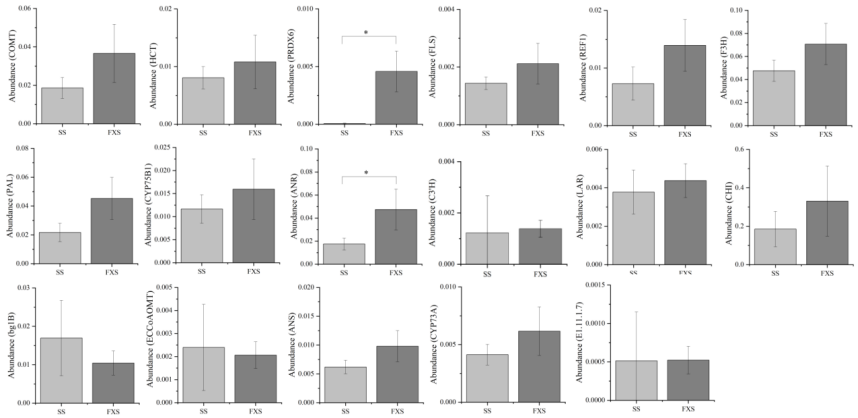

FIGURE 6 Profiles of protein level (TMT data) involved in bioactive constituent accumulation in TH between SS and FXS groups. (mean ± SD; * $p < 0.05$)

4 Discussion

TH has abundant resources in China, and the growth and development of TH is closely related to its hosts. It not only absorbs water, nutrients, etc. from the hosts, but also contains some characteristic components of the hosts [29–31]. Proteomic analysis of medicinal plants can contribute to our understanding of the changes that occur at the protein level in TH from different hosts and the molecular mechanisms by which these changes further lead to the accumulation of secondary metabolites. In this study, proteomics was used to analyze TH samples from common hosts of SS and FXS. The expression of proteins related to several pathways, such as amino acid metabolism, glycolysis/gluconeogenesis, phenylpropanoid biosynthesis, flavonoid biosynthesis, and terpenoid biosynthesis, have been altered.

Enzymes associated with abiotic stress, such as bglB and bglX, play an important role in maintaining the normal physiological functions of the organism [32]. 14 peroxidases were identified in this study, and their expression was not consistent in TH from SS, so we hypothesized that the hosts had influence on the activity and quantity of the enzymes. Most proteins took part in the process from chemical backbone to modification of flavonoids. PAL, C3'H and 4CL are key enzymes in the phenylpropanoid biosynthesis. With the exception of one 4CL (TRINITY_DN6526_c0_g1_i1_orf1), the expression of these proteins was not noticeably altered. However, minor changes might result in variations in the accumulation of coumaroyl-CoA in the upstream flavonoid synthesis in TH from two hosts. Coumaroyl-CoA is a prerequisite for the formation of flavonoid mother nucleus structure. The 4CL coding gene has been successfully cloned from SS [33]. According to the research, different 4CL plays the role of rate-limiting enzyme in lignin monomer biosynthesis pathway, but its specific function varies with plant species [34]. Most of the differences in protein expression identified upstream of the pathway in this study were not significant, while significant differences were mainly found in the downstream proteins, including CAD (TRINITY_DN6468_c0_g1_i11_orf1), 2 CCRs (TRINITY_DN20850_c0_g2_i1_orf1, TRINITY_DN8569_c1_g1_i5_orf1), and bglX (TRINITY_DN1860_c0_g1_-i15_orf1). Based on the expression of CAD in roots, collaterals, leaves and shoots of Pinus massoniana [35], we speculated that there was a similar phenomenon in TH.

The content of ferulic acid in TH from SS was lower than that of TH from FXS, presumably as a result of the synergistic effect of forming multiple enzymes on the ferulic acid pathway that were mostly under-expressed. In contrast, the chlorogenic acid content showed a completely opposite trend. Three pathways of chlorogenic acid synthesis have been reported [24]. Only one of these pathways was activated in this study, and the content level might be reflected in multiple pathways jointly. The expression of F3H and FLS were up-regulated in TH from SS. Meanwhile, based on the content of quercetin, the change trend of protein level and product content is not consistent. The content of (+)-Catechin from SS was significantly higher than

that in FXS, while the expression levels of related proteins in the upstream of catechin synthesis pathway were not different. Therefore, the reasons for this phenomenon are worth exploring. In terms of the content of flavonoid aglycones ((+)-Catechin and isosakuranetin) were higher in TH from SS than those from FXS, which indicated that the formation of flavonoid aglycones was positively correlated with the expression of upstream catalytic enzymes. The synthesis and accumulation of flavonols in plant tissues are synergistically regulated by various factors such as plant species, developmental stage, tissue site, and growth environment. The metabolism of compounds in plants is influenced by plant genetic and environmental factors, thus the amount and content of bioactive components can reflect intra-species variation to some extent [36], which also leads to large differences in the same herb species in terms of yield and quality.

Proteomics approaches have been demonstrated to be effective in elucidating the biological changes that accompany plant growth or in response to environmental stress in a variety of main crop species. In general, the overall content of flavonoid aglycones in TH from SS was higher than these from FXS, and the content of flavonoid glycosides was nearly 7 times that of the latter. The abundance of enzymes encoded by particular genes in biosynthesis pathways was correlated with the production and accumulation of these metabolites [37]. Since only one glycosidase was discovered in our work, we hypothesized that the ultimate cause of the difference in the content of bioactive constituents is not only the expression of the relevant synthase, but also the involvement of glycosyltransferases [38].

5 Conclusion

In conclusion, analysis was done on the protein differences of TH from two hosts. Functional annotation and enrichment analysis were performed for DEPs. Several key enzymes in the biosynthetic pathway of flavonoid such as PAL, CHI, 4CL, C3'H were identified. At the same time, the content of 16 bioactive constituents were quantitatived. The findings demonstrated that the regulated patterns of the most proteins generally agreed with the content changes of bioactive constituents between the two groups of TH. In conclusion, our study integrated protemic and metabolic data, revealing the molecular mechanisms underling the the differences in the accumulation of secondary metabolites and providing novel insights into the proteomics in TH from different hosts.

Acknowledgments

This work was supported by the Guangxi Key Laboratory of Zhuang and Yao Ethnic Medicine [grant number GXZYKF2020-2, the principal funding recipients Xunhong Liu]; Natural Science Foundation of Jiangsu Province Basic Research Program [grant number BK20220479, the principal funding recipients Zhichen Cai]; Basic Science (Natural Science) Research Project of Higher Education of Jiangsu Province [grant number 22KJB360007, the principal funding recipients Zhichen Cai].

Declaration of Competing Interest

The authors declare no conflict of interest.

Availability of data

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Supplementary figure

Figure S1. (A) Protein information in TH. (B) Functional annotation of all proteins in COF, GO, KEGG, Pfam and SubCell-Location.

Figure S2. (A)Peptide matching error distribution. (B) Peptide length distribution.

Figure S3. (A) The molecular weight distribution of the identified protein. (B) Distribution of peptide sequences coverage.

Figure S4. Trend analysis of all protein expression annotated in this study (10 trends).

Figure S5. Metabolic pathway maps of identified proteins involved in phenylpropanoid biosynthesis in KEGG enrichment (green for annotation, blue font for DEPs).

Supplementary data

Table S1. The information of 5861 proteins.

Table S2. The information of significantly differentially expressed proteins.

Table S3. GO annotation analysis.

Table S4. GO enrichment analysis.

Table S5. KEGG pathway classification.

Table S6. KEGG enrichment analysis.

Table S7. Proteins related to phenylpropanoid and flavonoid biosynthesis.

Table S8. Contents of 16 bioactive constituents (μg/g, mean±SD, n=3).

A list of abbreviations

AGC: Automatic Gain Control

BP: Biological Process

BBP: Benzyl Butyl Phthalate

CC: Cellular Component

DDA: Data Dependent Acquisition

DEPs: Differentially Expressed Proteins

FXS: *Liquidambar formosana* Hance

HCD: High Energy Collisional Dissociation

KEGG: Kyoto Encyclopedia of Genes and Genomes

MF: Molecular Function

SS: *Morus alba* L.

TEBA: Triethylammonium Bicarbonate Buffer

TH: Taxilli Herba

TMT: Tandem Mass Tag

UFLC-QTRAP-MS/MS: ultrafast liquid chromatography simultaneously coupled with triple quadrupole/linear ion trap tandem mass spectrometry

References

1. Chinese Pharmacopoeia Commission. (2020). Pharmacopoeia of the People's Republic of China, Part I. Beijing: Chemical Industry Press. p. 312.
2. Deng, J.G., & Wei, S.G. (2007). Genuine and well-reputed medicinal materials in Guangxi. Beijing: China Press of Traditional Chinese Medicine. p. 350.
3. Tao, H. J. (1994). Annotation of Materia Medica. Beijing: People's Medical Publishing House. pp. 254–255.
4. Zhuo, Z. M., Fan, Z. C., & Guo, X. (2020). Effect of Herba Taxilli extracts combined with miR-375 on viability and apoptosis of osteoarthritic chondrocytes. *Chinese Journal of Pathophysiology* , *36* (6), 1082–1088.

5. Meng, T. X., Yuan, X. L., Liang, F., Yuan, L. F., & Gong, Z. Q. (2021). Hypoglycemic effect of total flavonoids from mistletoe on diabetic mice induced by streptozotocin. *Journal of Shandong University of Traditional Chinese Medicine* , *44* (6), 55–60.
6. Liu, S. S., Liu, Q., He, C. N., Peng, Y., & Xiao, P. G. (2019). Application history and modern research progress of Taxillus chinensis herba tea. *Modern Traditional Chinese Medicine* , *21* (2), 147–153.
7. Tang, S. W. (1993). Classified Materia Medica. Beijing: Huaxia Publishing House. p. 359–360.
8. Li, S. Z. (1982). Compendium of Materia Medica. Beijing: People's Medical Publishing House. pp. 2158–2159.
9. Chai, Z. S., Liu, R. Y., Li, L. Z., Xia, Y. P., & Li, Y. H. (2020). Study on the problems and countermeasures in the application of traditional Chinese medicine Taxilli Herba. *China Journal of Traditional Chinese Medicine* , *35* (4), 2066–2069.
10. Lu, S. H., Gan, Y. Y., Tang, L., Zhang, F. Y., & Su, B. W. (2020). Evaluation of the quality of Taxillus chiueusis from different hosts sources by HPLC fingerprint combined with chemomtrics. *China Pharmacy* , *31* (07), 794–799.
11. Xia, Y. P., He, C. H., Chai, Z. S., Qin, W. H., Wu, W. X., Chen, L.Y., ... Li, Y. H. (2021). Acute toxicity and reprotoxicity of aqueous of Taxlli Herba from different hosts in zebrefish model. *Chinese Journal of Experimental Traditional Medical Formulae* , *27* (21), 91–97.
12. Zhang, H., Huang, F. Y., Liu, R. Y., Lu, H. L., Su, B. W., Zhu, K. X., & Li, Y. H. (2018). Effects of Taxilli Herba from different hosts on lowering blood pressure of spontaneously hypertensive rats. *Chinese Traditional Patent Medicine* , *40* (2), 249–254.
13. Yin, S. G., Liu, R. Y., Huang, F. Y.; Li, Y. H., & Li, J. Y. (2018). Study on infrared spectra of *Taxillus chinensis* (DC.) Danser from different host trees. *Lishizhen Medicine and Materia Medica Research* , *29* (11), 2629–2631.
14. Wu, N., Li, Li., Cai, Z. C., Yuan, J. H., Wang, W. X., Yin, S. X., ... Li, J. (2021). Quality evaluation of Taxilli Herba from different hosts based on simultaneous determination of multiple bioactive constituents combined with multivariate statistical analysis. *Molecules* , *26* (24), 7490–7505.
15. Yuan, J. H., Cai, Z. C., Chen, C. H., Wu, N., Yin, S. X., Wang, W. X., Chen, H. J., ... Liu, X.H. (2022). A study for quality evaluation of Taxilli Herba from different hosts based on fingerprint-activity relationship modeling and multivariate statistical analysis. *Arabian Journal of Chemistry* , *15* , 103933.
16. Ye, H. Y. (2020). Relationship between water transport and host plants in Taxillus chinensis. PhM Thesis, University of Chinese Academy of Sciences, Beijing.
17. Zhan, Z. G. (2020). Advances in biosynthesis and regulation of the active ingredient of Salvia miltiorrhiza based on multi-omics approach. *Acta Pharmaceutica Sinica* , *55* (12), 2892–2903.
18. Jiao, S. G., Ren, J. G., & Wang, J. L. (2021). Advance on proteomics related to the regulation of accumulation of active ingredients in medicinal plants. *Research and Practice on Chinese Medicines* ,*35* (2), 91–95.
19. Mergner, J., Frejno, M., List, M., Papacek, M., Chen, X., Chaudhary, A., ... Kuster, B. (2020). Mass-spectrometry-based draft of the Arabidopsis proteome. *Nature* , *579* (7799), 409–414.
20. Komatsu, S., Konishi, H., Shen, S.H., & Yang, G. X. (2005). Rice proteome analysis: A step toward functional analysis of the rice genome. *Proteomics* , *5* (4), 938–949.
21. She, C. (2017). Analysis on genetic basis fot yield difference between wild and cultivated soybeans based on proteome, transcriptome and genome. PhD Thesis, Nanchang University, Nanchang.
22. Chen, C. H., Wang, C. C., Liu, Z. X., Cai, Z. C., Hua, Y. J., Mei, Y. Q., ... Liu, X. H. (2020). iTRAQ-based proteomic technique provides insights into salt stress responsive proteins in Apocyni Veneti Folium (*Apocynum venetum* L.). *Environmental and Experimental Botany* , *180* , 104247–104258.
23. Mehmood, S. S., Lu, G. Y., Luo, D., Hussain, M. A., Raza, A., Zafar, Z., ... Lv, Y. (2021). Integrated analysis of transcriptomics and proteomics provides insights into the molecular regulation of cold response in *Brassica napus* . *Environmental and Experimental Botany* , *187* , 104480.
24. Cai, Z. C., Wang, C. C., Chen, C. H., Chen, H., Yang, R., Chen, J. J., ... Liu, X. H. (2021). Omics map of bioactive constituents in Lonicera japonica flowers under salt stress. *Industrial Crops and Products* , *167* , 113526.

25. Ma, X. Y., Nan, C. L., & Xue, Y. C. (2022). Overview of structural and functional studies of chalcone isomerases in plants. *Biology Teaching* , *47* (1), 2–4.
26. Park, S., Kim, D., Park, B. R., Lee, J. Y., & Lim, S. H. (2019). Molecular and functional characterization of Oryza sativa Flavonol Synthase (OsFLS), a bifunctional dioxygenase. *Journal of Agricultural and Food Chemistry* , *67* (26), 7399–7409.
27. Liang, Y., Li, L., Cai, Y., Xu, L., Xie, F. F., Liang, D. L., & Chen, F. L. (2022). Analysis of chemical constituents in ethyl acetate extract of Taxilli Herba by UPLC-Q-Exactive-MS and screening of potential xanthine oxidase inhibitors. *China Journal of Chinese Materia Medica* , *47* (4), 972–979.
28. Zhang, J. (2011). Study on composition analysis of Herba Taxilli and screening active site of anti-leukemia cells. PhM Thesis, Guangzhou University of Chinese Medicine, Guangzhou.
29. Zhang, F. K., Pan, T., Sheng, L., & Ji, K. S. (2014). Cloning and coding protein structure prediction on cDNA encoding cinnamyl alcohol dehydrogenase from Pinus massoniana. *Molecular Plant Breeding* ,*12* (4), 638–645.
30. Li, Y. H., Su, B. W., Zhang, X. J., Zhu, K. X., Pei, H. H., Zhao, M. H., & Lu, D. (2011). Correlation of DNJ between taxilli herba and its host-plants. *China Journal of Chinese Materia Medica* ,*36* (5), 2102–2106.
31. Liu, R. Y., Huang, F. Y., Guo, M., Lu, H. L., Zhu, K. X., Qin, W. H., ... Su, B.W. (2019). Correlation study on chemical constitutes of cardiac glycosides in Taxillus chinensis and its Nerium indicum host by UPLC-Q-TOF-MS/MS. *China Journal of Chinese Materia Medica* ,*44* (11), 2283–2291.
32. Zhuang, D. H., Song, J. J., & Huang, Y. (2003). Study on peroxidase isozyme in several species of Hibiscus. *Journal of Shantou University(Natural Science Edition)* , *18* (4), 9–13+30.
33. Huang, M. F., Wang, H., Fang, R. J., Zhao, W. G., Zhang, L., Pan, G., & Liu, L. (2014). Cloning and differential expression of cinnamic acid 4-hydroxylase gene in different mulberry varieties. *Acta Sericologica Sinica* , *40* (4), 592–600.
34. Ni, Z. Y., Wang, J., Lv, M., Li, B., Bai, Y., & Fan, L. (2010). Cloning and prokaryotic expression of Gh4CL2 gene from *Gossypium hirsutum* L. *Acta Botanica Boreali-Occidentalia Sinica* ,*30* (3), 429–436.
35. Zhang, X. J., Su, B. W., Li, J., Li, Y. H., Lu, D., Zhu, K. X., ... Zhao, M. H. (2014). Analysis by RP-HPLC of mangiferin component correlation between medicinal loranthus and their mango host trees.*Journal Of Chromatographic Science* , *52* (1), 1–4.
36. Khatoon, A., Rehman, S., Oh, M. W., Woo, S. H., & Komatsu, S. (2012). Analysis of response mechanism in soybean under low oxygen and flooding stresses using gel-base proteomics technique. *Molecular Biology Reports* , *39* (12), 10581–10594.
37. Brunetti, C., George, R. M., Tattini. M., Field, K., & Davey, M. P. (2013). Metabolomics in plant environmental physiology. *Journal of Experimental Botany* , *64* (13), 4011–4020.
38. Wang, C. C., Chen, L. H., Cai, Z. C., Chen, C. H., Liu, Z. X., Liu, X. H., ... Mei, Y. Q. (2020). Comparative proteomic analysis reveals the molecular mechanisms underlying the accumulation difference of bioactive constituents in Glycyrrhiza uralensis Fisch under salt stress. *ournal of Agricultural and Food Chemistry* ,*68* (5), 1480–1493.