Diosgenin biosynthesis pathway and its regulation in *Dioscorea cirrhosa* L. (#88367)

First submission

Guidance from your Editor

Please submit by **12 Sep 2023** for the benefit of the authors (and your token reward) .

Structure and Criteria
Please read the 'Structure and Criteria' page for general guidance.

Custom checks
Make sure you include the custom checks shown below, in your review.

Raw data check
Review the raw data.

Image check
Check that figures and images have not been inappropriately manipulated.

If this article is published your review will be made public. You can choose whether to sign your review. If uploading a PDF please remove any identifiable information (if you want to remain anonymous).

Files

Download and review all files from the materials page.

10 Figure file(s)
11 Table file(s)
2 Raw data file(s)
1 Other file(s)

Custom checks

DNA data checks

- Have you checked the authors data deposition statement?
- Can you access the deposited data?
- Has the data been deposited correctly?
- Is the deposition information noted in the manuscript?

Structure and Criteria

Structure your review

The review form is divided into 5 sections. Please consider these when composing your review:

1. **BASIC REPORTING**
2. **EXPERIMENTAL DESIGN**
3. **VALIDITY OF THE FINDINGS**
4. General comments
5. Confidential notes to the editor

You can also annotate this PDF and upload it as part of your review

When ready submit online.

Editorial Criteria

Use these criteria points to structure your review. The full detailed editorial criteria is on your guidance page.

BASIC REPORTING

- Clear, unambiguous, professional English language used throughout.
- Intro & background to show context. Literature well referenced & relevant.
- Structure conforms to PeerJ standards, discipline norm, or improved for clarity.
- Figures are relevant, high quality, well labelled & described.
- Raw data supplied (see PeerJ policy).

EXPERIMENTAL DESIGN

- Original primary research within Scope of the journal.
- Research question well defined, relevant & meaningful. It is stated how the research fills an identified knowledge gap.
- Rigorous investigation performed to a high technical & ethical standard.
- Methods described with sufficient detail & information to replicate.

VALIDITY OF THE FINDINGS

- Impact and novelty not assessed. *Meaningful* replication encouraged where rationale & benefit to literature is clearly stated.
- All underlying data have been provided; they are robust, statistically sound, & controlled.
- Conclusions are well stated, linked to original research question & limited to supporting results.

Standout reviewing tips

The best reviewers use these techniques

Tip	*Example*
Support criticisms with evidence from the text or from other sources	*Smith et al (J of Methodology, 2005, V3, pp 123) have shown that the analysis you use in Lines 241-250 is not the most appropriate for this situation. Please explain why you used this method.*
Give specific suggestions on how to improve the manuscript	*Your introduction needs more detail. I suggest that you improve the description at lines 57- 86 to provide more justification for your study (specifically, you should expand upon the knowledge gap being filled).*
Comment on language and grammar issues	*The English language should be improved to ensure that an international audience can clearly understand your text. Some examples where the language could be improved include lines 23, 77, 121, 128 – the current phrasing makes comprehension difficult. I suggest you have a colleague who is proficient in English and familiar with the subject matter review your manuscript, or contact a professional editing service.*
Organize by importance of the issues, and number your points	*1. Your most important issue* *2. The next most important item* *3. …* *4. The least important points*
Please provide constructive criticism, and avoid personal opinions	*I thank you for providing the raw data, however your supplemental files need more descriptive metadata identifiers to be useful to future readers. Although your results are compelling, the data analysis should be improved in the following ways: AA, BB, CC*
Comment on strengths (as well as weaknesses) of the manuscript	*I commend the authors for their extensive data set, compiled over many years of detailed fieldwork. In addition, the manuscript is clearly written in professional, unambiguous language. If there is a weakness, it is in the statistical analysis (as I have noted above) which should be improved upon before Acceptance.*

Diosgenin biosynthesis pathway and its regulation in *Dioscorea cirrhosa* L.

Yan Lin [1], **Qiuyan Hu** [1], **Qiang Ye** [2], **Haohua Zhang** [2], **Ziyu Bao** [2], **Yongping Li** [Corresp., 3], **Luo Jian Mo** [Corresp. 1]

[1] Dongguan Institute of Forestry Science, Dongguan, Guangdong, □□

[2] College of Forestry and Landscape Architecture, South China Agricultural University, Guangzhou, Guangdong, □□

[3] Key Laboratory for Quality Regulation of Tropical Horticultural Crops of Hainan Province, School of Horticulture, Hainan University, Haikou, Hainan, □□

Corresponding Authors: Yongping Li, Luo Jian Mo
Email address: yplee614@163.com, 52116124@qq.com

Dioscorea cirrhosa L. (*D. cirrhosa*) tuber is a traditional medicinal plant that is abundant in various pharmacological substances. Although diosgenin is commonly found in many Dioscoreaceae plants, it presence in *D. cirrhosa* has not been confirmed. To address this, we conducted HPLC-MS/MS analysis to identify 13 diosgenin metabolites in *D. cirrhosa* tuber. Furthermore, we utilized transcriptome data to identify 21 key enzymes and 43 unigenes that are involved in diosgenin biosynthesis, leading to a proposed pathway for diosgenin biosynthesis in *D. cirrhosa*. A total of 3365 unigenes belonging to 82 TF families were annotated, including MYB, AP2/ERF, bZIP, bHLH, WRKY, NAC, C2H2, C3H, SNF2 and Aux/IAA. Correlation analysis revealed that 22 TFs are strongly associated with diosgenin biosynthesis genes ($|r^2| > 0.9$, $P < 0.05$). Moreover, our analysis of the CYP450 gene family identified 206 CYP450 genes (CYP450s), with 40 being potential CYP450s. Gene phylogenetic analysis revealed that these CYP450s were associated with sterol C-22 hydroxylase, sterol-14-demethylase and amyrin oxidase in diosgenin biosynthesis. Our findings lay a foundation for future genetic engineering studies aimed at improving the biosynthesis of diosgenin compounds in plants.

Diosgenin biosynthesis pathway and its regulation in *Dioscorea cirrhosa* L.

Yan Lin [1], Qiuyan Hu [1], Qiang Ye [2], Haohua Zhang [2], Ziyu Bao [2], Yongping Li [3,*] and Luo Jian Mo [1,*]

1 Dongguan Institute of Forestry Science, Dongguan, Guangdong, China;
linyan59230@163.com (Y.L.); 807238693@qq.com (Q.H.); 52116124@qq.com (L.M.)
2 College of Forestry and Landscape Architecture, South China Agricultural University, Guangzhou, Guangdong, China; ye1337060314@163.com (Q.Y.); zhanghaohua@stu.scau.edu.cn (H.Z.); bzy180726@163.com (Z.B.)
3 Key Laboratory for Quality Regulation of Tropical Horticultural Crops of Hainan Province, School of Horticulture, Hainan University, Haikou, China; yplee614@163.com
* Corresponding Author: yplee614@163.com; 52116124@qq.com

Abstract

Dioscorea cirrhosa L. (*D. cirrhosa*) tuber is a traditional medicinal plant that is abundant in various pharmacological substances. Although diosgenin is commonly found in many Dioscoreaceae plants, it presence in *D. cirrhosa* has not been confirmed. To address this, we conducted HPLC-MS/MS analysis to identify 13 diosgenin metabolites in *D. cirrhosa* tuber. Furthermore, we utilized transcriptome data to identify 21 key enzymes and 43 unigenes that are involved in diosgenin biosynthesis, leading to a proposed pathway for diosgenin biosynthesis in *D. cirrhosa*. A total of 3365 unigenes belonging to 82 TF families were annotated, including MYB, AP2/ERF, bZIP, bHLH, WRKY, NAC, C2H2, C3H, SNF2 and Aux/IAA. Correlation analysis revealed that 22 TFs are strongly associated with diosgenin biosynthesis genes ($|r^2| > 0.9$, $P < 0.05$). Moreover, our analysis of the CYP450 gene family identified 206 CYP450 genes (CYP450s), with 40 being potential CYP450s. Gene phylogenetic analysis revealed that these CYP450s were associated with sterol C-22 hydroxylase, sterol-14-demethylase and amyrin oxidase in diosgenin biosynthesis. Our findings lay a foundation for future genetic engineering studies aimed at improving the biosynthesis of diosgenin compounds in plants.

Keywords: *Dioscorea cirrhosa* L.; diosgenin; transcriptome; CYP450 gene family

Introduction

The biosynthesis of triterpene diosgenin involves a downstream pathway that is regulated by cytochrome P450 monooxygenase (CYP450s) and uridine diphosphate (UDP)-dependent glycosyltransferases (UGTs), which control a series of cholesterol oxidation, hydroxylation and glycosylation reactions [1]. CYP450 and UGT genes in plants exhibit high diversity, contributing significantly to the structural variety of triterpene. The biosynthesis of sterols relies on CYP450s.

For instance, CYP51G and CYP710A encode for obtusifoliol 14-α demethylase and sterol-22-desaturase, respectively [2,3]. The CYP97 family plays role in lutein biosynthesis. Arabidopsis CYP97A3 and CYP97C1 catalyze the hydroxylation of carotenoid β- and ε-rings in the lutein biosynthetic pathway, respectively [4,5]. The CYP73A subfamily has only one number, cinnamic acid 4-hydroxylase (C4H), which catalyzes cinnamic acid to form precursors of lignin and many other phenolic compounds. CYP98A catalyzes the 3 '- hydroxylation of coumaric acid [6]. However, due to a large number of CYP450 gene family members, it remains challenging to fully understand their roles in a variety of biosynthetic environments.

The advancement of RNA-seq provide has made it easier to identify CYP450s and UGTs. Transcriptome analysis of *Panax notoginseng* [7] revealed 350 and 342 predicted unigenes encoding CYP450s and UGTs, respectively; identified 233 CYP450 and 269 UGTs through RNA seq sequencing of Ilex plants, of which 14 CYP450s and one UGTs are considered to play a role in triterpene diosgenin biosynthesis [8]. In addition, Cheng et al. [9] found that CYP716A295 and CYP716A296 to be candidate genes related to oleanolic acid biosynthesis, while CYP72A763 and CYP72A776 are involved in diosgenin biosynthesis in the bark of *Aralia elata* (Miq.). *Dioscorea zingiberensis* is diosgenin-rich plant, Li et al. [10] conducted a comparative transcriptomic analysis of its rhizomes and found that a total of 485 annotated CYP450s unigenes, 195 annotated UGTs unigenes, and 165 CYP unigenes related to the diosgenin biosynthesis. Phylogenetic analysis revealed that four of these CYP candidate genes were most likely involved in the biosynthesis of diosgenin from cholesterol. However, it is unclear whether *D. cirrhosa* has diosgenin specific to the Dioscoreaceae. To data, no systematic investigation has been conducted to identify the genes related to diosgenin biosynthesis, and the transcriptome database of this plant is unavailable.

Diosgenin, a plant steroid, has demonstrated various pharmacological effects such as anti-hypercholesterolemia, anti-tumor, immune regulation, anti-cancer [11-14]. The biosynthetic pathway of diosgenin in plants has been identified in traditional Chinese medicine plants such as *Dioscorea zingiberensis*, *Dioscorea nipponica*, *Trigonella foenum-graecum* [10,15,16]. The biosynthesis process can be divided into several stages. First, Acetyl-coenzyme A generates isopentenyl diphosphate (IPP) and dimethylallyl diphosphate (DMAPP) through the MVA or MEP pathway. Next, Farnesyl pyrophosphate is formed from IPP and DMAPP by farnesyl diphosphate synthase (FPS). Squalene synthase (SS) and squalene epoxidase (SE) then convert farnesyl pyrophosphate into 2,3-oxysqualene. This compound is further converted into cycloartenol and lanosterol, respectively, by cycloartenol synthase (CAS) and lanosterol synthase (LSS). In the subsequent stage, multi-step enzymatic reactions transform cycloartenol into cholesterol, which is then oxidized, glycosylated, and cyclized to produce diosgenin. The MVA pathway takes place in the cytoplasm, whereas the MEP pathway occurs in the endoplasmic reticulum (cytosol), mitochondria (or golgi apparatus), and plastids.

While the enzymes involved in the upstream synthesis pathway of diosgenin biosynthesis have been identified, only a few of the downstream synthesis pathways have been characterized, especially the key enzymes that catalyze the formation of diosgenin. For instance, down-

regulation of the steroid specific glucosyltransferase TFS3GT2 in fenugreek leads to a reduction in diosgenin levels, suggesting its involvement in diosgenin biosynthesis in fenugreek [17]; In *Polyporus umbellatus*, the upregulation of C-8-sterol isomerase, sterol-C-24 methyltransferase (SMT) and sterol-22-desaturase has been found to promoted the biosynthesis of sterol metabolites [18], while sterol-C-24 methyltransferase has been shown to catalyze ergosterol biosynthesis [19]. These downstream enzymes play an important role in the regulation of triterpene biosynthesis in many plants [20,21].

Currently, no studies have been conducted on the presence of diosgenin in *D. cirrhosa*, and the biosynthesis pathway of diosgenin is not well characterized. The lack of molecular-level on this pathway hinders in depth research. Thus, it is of great significance to clarify the biosynthesis pathway and metabolic regulation network of diosgenin in *D. cirrhosa* tubers, particulary in the context of polygenic family protease, using advanced sequencing and bioinformatics technology. Such investigations not only aid in fundamental research but also facilitate the optimization of the target compound synthesis pathway in the future.

Materials & Methods

2.1. Plant Material

Four types of tubers, LR (Light red), RD (Red), DR (Dark red), and BR (Brownish red), each with distinct colors, were collected from the natural habitat located in the hilly region of Shunde, Guangdong Province, China (112.65°E, 22.88°N). The color gradient of the tubers was compared using color cards, which were easily visible (Figure 1). Three independent biological replicates were used in this study. The flesh of tubers was quickly frozen in liquid nitrogen and stored at −80°C for subsequent experiments.

2.2. Identification of Diosgenin Metabolites by UPLC-MS/MS

For sample extraction, the samples were freeze-dried in a vacuum freeze dryer (Scientz-100F), and then ground into powder (30Hz, 1.5min) using a mixer mill (MM 400, Retsch); Weigh 100mg of powder and dissolve in 1.2ml of 70% methanol solution, vortex for 30 seconds every 30 minutes for 6 times in total, repeat 6 times, and store in a 4°C refrigerator overnight. After centrifuge at 12000rpm for 10min, then supernatant was collected, filter through a microporous filter membrane (SCAA-104, 0.22μm pore size), and stored for further UPLC-MS/MS (UPLC, Ultra Performance Liquid Chromatography; Shimpack UFLC SHIMADZU CBM30A; MS, Applied Biosystems 4500 Q TRAP) analysis.

The metabolites were identified and quantified by UPLC-MS/MS technology with the following analytical conditions: UPLC: column was an Agilent SB-C18 (1.8μm, 2.1mm* 100mm); Mobile phase: phase A is ultrapure water with 0.1% formic acid, phase B is acetonitrile with 0.1% formic acid; Sample measurements were performed with a gradient program that employed the starting conditions of 95% A, 5% B. Subsequently, a composition of 95% A, 5.0% B was adjusted within 1.10 min and kept for 2.9 min. The flow velocity was set as 0.35ml per minute; The column oven was set to 40°C; The injection volume was 4μl.; The column

temperature is 40°C. The effluent was then connected to an ESI-triple quadrupole-linear ion trap (QTRAP)-MS.

The mass spectrometric conditions used for this study involved a QQQ linear ion trap mass spectrometer (Q trap) and API 4500 Q trap LC/MS/MS system to carry out LIT and three quadrupole (QQQ) scanning. The system is equipped with a ESI turbo ion spray interface, runs in positive and negative two ion modes, and is controlled by the QQQ software. ESI source operation parameters are as follows: ion source and turbine spray. Source temperature 550°C; Ion spraying voltage, 5500V (positive ion mode) / - 4500V (negative ion mode); The ion source gas I, gas II and curtain gas are set at 55, 60 and 25.0 psi respectively; The collision-induced ionization parameter is set to high. Respectively use 10 and 100 μmol/L polypropylene glycol solutions to carry out the instrument tuning and mass calibration. The QQQ scanning used multiple reaction monitoring (MRM) mode, with nitrogen set as the collision gas (nitrogen) at medium. Through further optimization, the corresponding parameters for each MRM ion pair were optimized, and a specific set of MRM ion pairs was monitored based on the metabolites eluted in each period.

2.3. Qualitative and Quantitative Analysis of Metabolites

Metabolites quantification is achieved through MRM mode analysis using triple quadrupole mass spectrometry. In this mode, the four-stage rod first screens the precursor ions of the target substance, then eliminates the precursor ions of other molecular weight substances, and finally eliminates the interference of non-target ions, so as to improve the quantitative accuracy and repeatability. After obtaining the mass spectrum data of different metabolites, the mass spectrum peak area is integrated and corrected, and the mass spectrum peak area corresponds to the relative content of substances [44]. Use Analyst (1.6.3) software to process mass spectrometry data.

To eliminate the interference of isotopic signals, including repeated signals of K^+, Na^+, $NH4^+$ ions, and repeated signals of fragment ions that are other substances with larger molecular weight. Metabolite characterization is based on MWDB (metal database). Based on the secondary spectrum information, metabolite data are matched to the corresponding substance information in the database, enabling determination of the metabolites category.

2.4. Significance Analysis and Correlation Analysis

The significance of diosgenin metabolite data and gene expression data was analyzed by t-test method with ggsignif package in R. The correlation of metabolites and genes used the corr() function of R software and calculates the correlation coefficient r^2 and significance level P value based on Pearson correlation analysis method. A P-value < 0.05 indicates significance, P-value < 0.01 indicates extreme significance. Metabolites or genes with significance level $P < 0.05$ and $| r^2 | > 0.9$ are selected and importing into Cytoscape (3.8.2) software to construct the connection network.

2.5. Acquisition, Alignment and Annotation of Transcriptome Data

The mRNA library for each sample were constructed and sequenced using the Illumina Hiseq platform by the paired-end reads. The statistical power of this experimental design,

calculated in RNASeqPower is 0.91. The raw data of 12 cDNA libraries were uploaded to the NCBI database (Accession number: PRJNA741609). Clean reads were obtained by removing low-quality bases and adapter sequences. Trinity software (version 2.6.6) [45] was used to assemble clean reads, and the longest clusters obtained by Corset (version 1.07) [46] were regarded as unigenes. The gene expression level in each sample was calculated using FPKM [47,48].

2.6. Identification of TFs

The Itak software (1.7A) [49] was used to predicted the transcription factors, which integrates two plant databases: PlnTFDB and PlantTFDB. The identification of transcription factors was carried out using HmmScan comparison through the transcription factor families and rules defined in the database.

2.7. Gene Expression Analysis Using qRT-PCR

To verify the expression of diosgenin-related genes, quantitative Real-time PCR (qRT-PCR) was performed. Ten diosgenin pathway genes were selected, including IDI gene, DXS gene, MVK gene, AACT gene, SMT gene, SS gene, gcpE gene and ispD gene. NCBI software was used to design primers, and DC18S was used as internal reference gene, transcribed and amplified used Goldenstar RT6 cDNA Synthesis Mix and T5 fast qPCR mix (SYBR Green I), three biological repeats were set. qRT-PCR was performed on FQD-96A system (Hangzhou BORI). The procedure is as follows: 95°C for 2 minutes, 95°C cycles 40 times for 15 seconds, 60°C cycles for 15 seconds and 72°C cycles for 20 seconds. The relative gene expression was calculated by $2^{-\Delta\Delta}$ Ct method.

2.8. CYP450 Gene Family Identification and Phylogenetic Analysis

The identified CYP450 gene was compared with the CYP450 gene sequence found in this study by using cluster X 2.0.12 [50], and the positions with high vacancy value and missing data percentage were adjusted. The phylogenetic tree was constructed by MEGA 6 software. with the best evolutionary model determined using J model test software. The construction of maximum likelihood (ML) tree is based on TN (Tamura NEI) model [51]. Bootstrap is set under 1000 repetitions to evaluate the importance of nodes.

Results

3.1. Identification of Diosgenin Metabolites

A total of 13 diosgenin metabolites were identified using the HPLC-MS/MS method (Figure 2, Table S1). Based on the change in metabolite content, they can be divided into three categories (Figure 3). The first category comprises eight metabolites: 3-O-(2-O-Acetyl-glucosyl) oleanolic acid, protodioscin, trillin-6'- O-sophorotrioside, trillin-6'-O-glucoside, diosgenin-3-O-glucoside (Trillin), pseudogenin B (parisyunnanoside b), diosgenin-3-O-glucosyl (1→4) rhamnosyl (1→4) rhamnosyl (1→2) glucoside (diosgenin-3-O-glcosyl (1→4) rhamnosyl (1→4) rhamnosyl (1→2) glcoside) and pseudoprotodioscin. The highest content of these metabolites was observed in DR (Dark red) followed by LR (Light red), RD (Red) and BR (Brownish red).

The second category includes three diosgenin metabolites: diosgenin-3-O-rhamnosyl - (1→3) glcoside, diosgenin-3-O-rhamnosyl (1,2) glcoside, ruscogenin-1-O-xylosyl (1,3) fucoside. Only BR exhibited high content while the LR, RD, and DR showed low content. The third category includes two diosgenin metabolites, including gracillin and pennogenin-3-O-glucoside, and RD and DR tubers showed relatively high content. Therefore, diosgenin metabolites are not accumulated gradually in *D. cirrhosa* tubers at all color stages. Based on the results above, diosgenin metabolites primarily accumulate in DR tubers, followed by BR tubers.

3.2. Identification of Diosgenin Biosynthesis Genes

In our study, we identified 43 unigenes encoding 21 key enzymes involved in diosgenin biosynthesis, which includes two metabolic pathways: IPP and MEP. Among these, seven unigenes were identified in the MEP pathway, including one DXS gene, one DXR gene, five ISP genes, one gcpE gene and three FPS genes. In MVA pathway, we identified four AACT genes, one HMGS gene, eight HMGR genes, three MVK genes, one PMK gene, one MVD gene and two IDI genes (Table 2, Table S2).

The oxidized squalene cyclase (OSC) family, comprising cycloartenol synthase (CAS), lanosterol synthase (LSS) and amyrin synthase (AS), cyclize 2,3-oxide squalene to synthesize different phytosterol skeletons. OSC enzymes are the key nodes in the triterpene biosynthesis pathway, and different members of the enzyme family have been shown to produce various triterpene skeletons in Arabidopsis and rice [20,22]. Ohyama et al. [23] demonstrated that phytosterols are synthesized via a dual pathway of cycloartenol and lanosterol, which are precursors of sterols and steroid hormones, while the genes encoding wool sterol biosynthesis have only been identified in a few plant species such as Arabidopsis, rice, *Panax ginseng* and *Dioscorea zingiberensis*. We found only one unigene (cluster-6992.57030) with high homology with the CAS and no LSS gene in this study. In addition, two C5 (6) genes, three C14-R genes and three unigenes encoding sterol 24-C methyltransferase (SMT) were identified, which are key genes downstream of diosgenin biosynthesis pathway and may play a role in the diosgenin conversion.

3.3. Potential Biosynthesis Pathway of Diosgenin and Expression of Related Genes

The expression level (FPKM) of diosgenin genes and the potential pathway of *D. cirrhosa* diosgenin biosynthesis are presented in Figure 3. It was showed that the MVA and MEP pathways exhibited a "high-low-high-low" pattern, except for the ispH gene. Most genes in the diosgenin biosynthesis pathway, such as cluster-6992.39943 (DXS), cluster-6992.29239 (DXR), cluster-6992.34108 (ispD), cluster-6992.46302 (ispE) and cluster-6992.52592 (ispF), were highly expressed in LR and DR but weakly expression in RD. This expression pattern was consistent with the change pattern of the first of diosgenin metabolites (Figure 2). It was worth noting that the expression levels of cluster-6992.49865 (HMGR), cluster-6992.55638 (FPS) and cluster-6992.39611 (SE) were significantly upregulated in BR compared with the other three groups, which is consistent with the second category of diosgenin metabolites (Figure 3). These results indicate that the genes exhibiting expression patterns consistent with the change levels of

diosgenin metabolites may be the be the key genes controlling diosgenin biosynthesis in *D. cirrhosa* tubers.

Moreover, downstream of the diosgenin biosynthesis pathway, we identified three SMT genes, one CAS gene, two C14-R genes and two C5 (6) genes. Among them, three SMT genes (cluster-6992.56623, cluster-6992.57489 and cluster-6992.31684) exhibited the same expression pattern as cluster-6992.57030 (CAS), cluster-6992.40736 (C5 (6)) and cluster-6992.50375 (C14-R) (Figure 4). It is worth noting that enzymes such as C-16, C-22, and C-26 are crucial in the conversion of cholesterol to diosgenin, but they were not identified in this study [16]. Therefore, we propose that diosgenin in *D. cirrhosa* may be a precursor of diosgenin catalyzed by cycloartenol via the SMT enzyme and that diosgenin is subsequently further synthesized by sterol desaturase (C5 (6)) and sterol C14 reductase (C14-R). These finding suggest that the genes downstream of the diosgenin biosynthesis pathway may also be key genes controlling diosgenin biosynthesis in *D. cirrhosa* tubers.

3.4. Transcription Factors (TFs) Identification

To identified TFs, we searched the plant TF databases and identified 3365 unigenes belonging to 82 TF families in the *D. cirrhosa* transcriptome dataset. Differential expression analysis revealed that 1261 TFs were differentially expressed, with 88 MYB family members identified, including 5 from the R2R3-MYB subfamily. Additionally, 55, 53, 50, 51, 56, 52, 49, and 20 genes belonged to the bZIP, bHLH, WRKY, NAC, C2H2, C3H, SNF2, and Aux/IAA families, respectively (Table 3). These diverse TF families provide valuable information for further biological analysis.

Furthermore, TFs with FPKM value > 1 in at least one sample were selected, resulting in 37 screened genes, including MYBs, bHLHs, WRKYs, and plant hormone related TFs (Table S3). We analyzed their expression patterns and depicted their expression level in heatmap (Figure S1). the DR tuber had the largest number of highly expressed TFs, including Aux/IAAs, WRKYs and AP2s, followed by BR and RD tubers, while LR tubers had the lowest expression levels. These results indicate that the expression of TFs was highest in the middle and late stage of tuber color formation and lowest in the early stage of color formation.

3.5. Interaction Between Diosgenin Biosynthesis Gene and TFs

In order to better understand the relationship between diosgenin biosynthesis genes and TFs, we calculated the Pearson correlation coefficient (r^2) between TFs and diosgenin biosynthesis genes. We then selected the items with $|r^2| > 0.9$ and used them to construct a gene interaction network. The results showed that a total of 22 TFs were strongly associated with diosgenin genes ($|r^2| > 0.9$, $P < 0.05$), including WRKY, MYB, bZIP, bHLH, Aux/IAA and AP2/ERF gene family. Among them, bZIP had the most connections with diosgenin biosynthesis genes, followed by MYB and AP2/ERF (Figure 5, Table S4). These candidate TFs are likely to play a crucial role in the biosynthesis of diosgenin in *D. cirrhosa*.

3.6. CYP450 Gene Family Analysis and Phylogenetic Tree Construction

The expression patterns of genes related to secondary metabolite generally correspond metabolites levels in different parts of plant. CYP450 family is the largest plant protein family

and plays important roles in catalyzing most of the oxidative steps in plant secondary metabolism. It is well known that CYP450 is involved in the catalytic reaction of cholesterol formation from lanosterol in diosgenin biosynthesis pathway [24]. However, to data, no CYP450 or UGT genes involved in diosgenin biosynthesis have been identified in Dioscoreae or other diosgenin producing plants. In this study, 206 unigenes encoding CYP450 protein were identified in *D. cirrhosa* transcriptome data through different database annotations (Figure 6). Among them, 112 belong to CYP2 subfamily, 68 belong to CYP4 / CYP19 / CYP26 family and 4 belong to CYP3 / CYP5 / CYP6 / CYP9 family. We obtained a total of 40 candidate genes were obtained by screening unigene with FPKM > 1 in at least one tissue, which are considered potential genes for diosgenin biosynthesis in *D. cirrhosa*. Among these candidate genes, 14 genes were identified as members of the CYP71A1 subfamily, 5 genes were identified as members of the CYP72A219 subfamily, 4 genes identified as members of the CYP94C1 subfamily and 3 genes identified as members of the CYP711A1 subfamily. Through homologous annotation analysis, we found that 12 genes were homologous to *Phoenix dactylifera*, 20 genes were homologous to *Elaeis guineensis*, 4 genes were homologous to *Musa acuminata* subsp. Malacensis, 1 gene was homologous to *Daucus Carota* subsp. Sativus, 1 gene was homologous to *Ananas comosus*, and 1 gene was homologous to *Asparagus officinalis* (Table S5).

Moreover, among the CYP450 genes we identified, the genes cluster-6992.48174 and cluster-6992.33292 were annotated as steroid-22-α-hydroxylase (CYP90B), and the gene cluster-6992.33753 and cluster-6992.33754 are annotated as sterol-14-demethylase (CYP51) (Figure 7, Table S5). Both Steroid-22-α-hydroxylase and sterol-14-demethylase are downstream synthase that catalyze diosgenin biosynthesis, indicating that these CYP450 genes may be involved in diosgenin synthesis in *D. cirrhosa* tubers. Further analysis of the expression level of these genes and search of homologous annotation database showed that 6 CYP450 genes were highly expressed in LR, 13 CYP450 genes were highly expressed in RD and 22 CYP450 genes showed high expression level in DR. It should be noted that the expression levels of almost all CYP450 genes were low during the BR period.

Diosgenin is synthesized from cholesterol through a series of oxidation reactions at C-22, C-26 and C-16 positions [16]. Steroid-22-α-hydroxylase and sterol-14-demethylase play a catalytic role in the pathway steps of cholesterol formation, which is the precursor of diosgenin biosynthesis [25]. In addition, the identification of unigenes associated with sterol-14-demethylase and steroid-22-α-hydroxylase in our CYP450 candidate list suggests that some other CYP candidate genes may play important roles in diosgenin biosynthesis. To understand the functions of these CYP candidate genes, we performed phylogenetic analysis on the CYP450 candidate genes along with other well-characterized CYP genes from various metabolic pathways, including those involved in the biosynthesis of triterpenes, diosgenin, and flavonoid (Table 4).

In Figure 8, unigene cluster-6992.63413 clustered with clusters-6992.48174 and cluster-6992.33292 from the CYP90B subfamily and AtCYP90B1 from Arabidopsis [26], which is known to have sterol C-22 hydroxylase in the brassinosteroid pathway. This suggests that

unigene cluster-6992.63413 may be a CYP450 candidate responsible for steroid C-22 hydroxylation in diosgenin biosynthesis. Unigene cluster-6992.33753 and cluster-6992.33754 are clustered into the same branch with AtCYP51 of Arabidopsis [27], SiCYP51 of tomato (*Solanum lycopersicum*) and SbCYP51 of *Sorghum bicolor*, which are characterized by sterol-14-demethylase. Cluster-6992.51327 is more closely related to SbCYP51 of *Sorghum bicolor*, suggesting that it may encode sterol-14-demethylase in diosgenin biosynthesis pathway. Unigene cluster-6992.67001, cluster-6992.40273 and β-amyrin C-24 oxidases are relatively closer, including MtCYP93E2 from tomato [28], GuCYP93E3 from *Glycyrrhiza uralensis* [29], and GmCYP93E1 from potato [30], indicating that unigene cluster-6992.67001 and cluster-6992.40273 may encode amyrin oxidase in *D. cirrhosa* diosgenin biosynthesis. Notably, AtCYP734A1 [31] in Arabidopsis and CYP enzyme SlCYP734A7 in tomato are characterized by sterol C-26 hydroxylase in the biosynthesis of diosgenin [32], which are far away from each other in branches. In addition, unigene cluster-6992.77886, cluster-6992.66001, cluster-6992.54954 and cluster-6992.41463 clustered into the same branch as steroid C-26 hydroxylase gene AtCYP734A1 [31] of *Arabidopsis thaliana*, β-amyrin C-11 oxidase gene of *Medicago truncatula*, MtCYP716A12 [28], gibberellin biosynthesis gene of *Cucurbita maxima*, and GmCYP88A [33]. This suggests that these genes may be multifunctional synthases, and their functions require further investigated by cloning full-length sequences.

Further correlation analysis between the CYP450 candidate genes and diosgenin metabolites showed that 14 CYP450 genes had high correlation with the content of seven diosgenin metabolites ($|r^2| > 0.9$, Table S6). These CYP450 unigenes showed a consistent accumulation pattern with diosgenin metabolites and are believed to be key genes in diosgenin biosynthesis.

Glycosylation has a crucial role in the biological activity of diosgenin compounds, and diphosphate (UDP) - dependent glycosyltransferases catalyze a crucial step in the biosynthesis of diosgenins [34]. To provide energy for metabolite biosynthesis, sugar transformation is necessary, and we identified five UGTs in the transcriptome data of *D. cirrhosa*. By performing a gene homology search in the NR database, we found that these genes have homology with sucrose biosynthesis in *Elaeis guineensis* and *Phoenix dactylifera* (Table 5).

3.7. Identification of Key Genes related to Diosgenin Biosynthesis

We utilized Pearson correlation analysis to explore the relationship between the genes related to diosgenin biosynthesis and metabolites. The results indicated a strong correlation between 22 diosgenin synthesis related genes and 12 diosgenin metabolites ($p < 0.05$, $|r^2| > 0.95$). Among these genes, 12 genes showed a significantly positively correlated with metabolites, including 2 UDPGs genes, 1 DXS gene, 1 gcpE gene, 2 SMT genes, 2 IDI genes, 1 SS gene, 1 MVK gene, 1 AACT gene and 1 ispD gene; 10 genes were negatively correlated with metabolites, including 1 DXR gene, 1 C5 (6) gene, 2 HMGCR genes, 1 CAS gene, 1 AACT gene, 1 HMGS gene, 1 FDS gene, 1 UDPGs gene and 1 SE gene (Figure 10). There genes are speculated to play either positive or negative regulatory roles in the diosgenin biosynthesis.

3.8. Quantitative Real-Time (qRT-PCR) Validation of diosgenin genes

To validate the accuracy of our RNA-seq data, we conducted qRT-PCR to evaluate the gene expression of 10 diosgenin pathway genes at the transcriptional (Figure 10). The results were consistent with transcriptome data, suggesting the reliability of our findings.

Discussion

Diosgenin, a triterpene-derived steroid, is involved in the biosynthesis of isoprenoids, steroids, sesquiterpenes and triterpenes in higher plants. The key enzyme genes in the terpene biosynthesis pathway have been proved to play an vital role in the regulating terpene biosynthesis. For example, up-regulating the DXS and DXR genes in the hairy roots of sage promotes the accumulation of diterpenes [35]; PGSQE1 regulates ginsenoside biosynthesis in ginseng by regulating its expression. [36]; In *Acanthopanax senticosus*, the expression of β-amyrin synthase ESbas promotes the accumulation of oleanolic acid [37]. Diosgenin can be synthesized in higher plants via the MVA or MEP pathways, with MVA being the preferred pathway [38]. In this study, we focused on the diosgenin biosynthesis pathway in *D. cirrhosa*, and proposed that the diosgenin biosynthesis pathway of *D. cirrhosa* is composed of MEP and MVA. A total of 43 unigenes encoding up to 21 key enzymes involved in diosgenin biosynthesis. The expression pattern of most genes is consistent with that of diosgenin metabolites, which is the key gene to control diosgenin biosynthesis. Interestingly ,this study found that the expression pattern of *D. cirrhosa* diosgenin biosynthesis pathway gene did not increase gradually with the accumulation of tuber pigment. The identified key group of diosgenin biosynthesis provides candidate genes for further gene manipulation.

In recent years, there have been studies investigation the regulation of diosgenin metabolites biosynthesis by plant hormones. For example, Sun et al. [15] found that the rhizome of *D. cirrhosa* treated with methyl jasmonate promote the biosynthesis of diosgenin and the significant expression of diosgenin gene, but inhibited the synthesis of C-24 methylation product. In addition, ethylene treatment was found to enhance the accumulation of diosgenin by upregulating the expression of HMGR and CAS genes in *D. zingiberensis* [39], while hormone treatments such as salicylic acid and methyl jasmonate can promote the expression of DXD gene in *Ginkgo biloba* roots, stems, leaves, pericarps and seeds [40]. A novel discovery of this study is the significant strong connection between transcription factors and diosgenin biosynthesis genes ($|r2| > 0.9$, $P < 0.05$). Hence, we propose that diosgenin biosynthesis may also be regulated by TFs such as bZIP, MYB and AP2/ERF, and diosgenin pathway genes may function together with TFs on *D. cirrhosa* diosgenin biosynthesis.

Recently, a new study demonstrated that the down-expression of TFSMT1 gene in fenugreek resulted in an increase in cholesterol levels, but a significant decrease in diosgenin. Although overexpression of this gene led to an increase in diosgenin content, there was no significant effect on cholesterol biosynthesis. This study suggested that cholesterol does not participate in diosgenin biosynthesis and that SMT gene was closely related to the biosynthesis of diosgenin in fenugreek [25]. In this study, we identified three SMT genes were expressed in diosgenin biosynthesis, and no cholesterol related metabolites were detected, indicating that *D. cirrhosa* diosgenin biosynthesis may not go through the cholesterol pathway. This finding

suggests that diosgenin biosynthesis in *D. cirrhosa* may not involve the cholesterol pathway, but instead is synthesized from sitosterol under the catalysis of the SMT gene., which supports the findings of Cao et al [25]. However, further verification is needed to determine if the downstream synthesis pathway of diosgenin in *D. cirrhosa* is regulated by the enzyme.

Cytochrome P450 plays a crucial role in the conversion of cholesterol to diosgenin[41]. To date, 80 CYP450 genes have been identified that are associated with terpene metabolism. The CYP51G, CYP85A, CYP90B-D, CYP710A, CYP724B and CYP734A subfamily members are generally conserved in the plant kingdom and are involved in primary metabolism related to the biosynthesis of plant essential sterols and steroid hormones [42]. In addition, specialized triterpenoids require the participation of subfamilies, such as CYP51H, CYP71A, D, CYP72A, CYP81Q, CYP87D, CYP88D, CYP93E, CYP705A, CYP708A and CYP716A, C, E, S, U and Y, and the members of these subfamilies may have species-specific functions, including chemical defense against specialized pathogens [42]. In most cases, cyclic triterpene scaffolds catalyze a large number of scaffold, regional and stereospecific oxidative modifications catalyzed by cytochrome P450 monooxygenase, resulting in triterpene scaffolds with various functional groups, such as hydroxyl, carbonyl and carboxyl. Furthermore, the addition of oxygen function mediated by CYP450s exposes the triterpene scaffold subsequently exposed to UDP glycosyltransferases and acyltransferases (ATS), leading to the formation of diosgenin and acylated triterpenes [43]. In recent years, significant progress has been made in understanding the biochemical function of CYP450 involved in plant triterpene metabolism. Combining the genetic screening of mutants with impaired triterpene biosynthesis and the availability of integrated genomic and transcriptome resources, some CYP450s involved in the structural modification of plant triterpenes has been identified. In addition, recent advances in transcriptome have contributed to the identification of a large number of CYP450s. Assigning biochemical functions to these CYP450s will aid in studying the biosynthesis of diosgenin in plants.

In this study, a total of 206 unigenes encoding CYP450 proteins were identified, with 112 belong to CYP2 subfamily, 68 belong to CYP4 / CYP19 / CYP26 family and 4 belong to CYP3 / CYP5 / CYP6 / CYP9 family. Further analysis of the 40 candidate P450 genes identified14 genes identified as CYP71A1 gene, 5 genes identified as CYP72A219 gene, 4 genes identified as CYP94C1 gene and 3 genes identified as CYP711A1 gene. The gene homology annotation analysis mainly divided these CYP genes into species such as *Phoenix dactylifera*, *Elaeis guineensis*, *Musa acuminata* subsp. malaccensis and *Ananas comosus*. Four key CYP450 genes of diosgenin biosynthase were identified, cluster-6992.48174 and cluster-6992.33292 encode steroid-22-α-hydroxylase, cluster-6992.33753 and cluster-6992.33754 encode sterol-14-demethylase (CYP51). The CYP450 phylogenetic analysis suggested that the gene cluster-6992.63413 may be involved in sterol C-22 oxidation in diosgenin biosynthesis, cluster-6992.51327 may encode sterol-14-demethylase in diosgenin biosynthesis pathway, and cluster-6992.67001 and cluster-6992.40273 may encode amyrin oxidase in diosgenin biosynthesis (β-amyrin-C-24 oxidases). This study systematically analyzed the CYP450 and UGT gene families

of *D. cirrhosa*, laying a foundation for further studies on the functions of these two multi-gene families.

Supplementary Materials: Table S1: Diosgenin metabolites and their relative contents in *D. cirrhosa* tubers; Table S2: FPKM value of identified diosgenin biosynthesis pathway genes; Table S3: Expression level of candidate TFs. Table S4: Correlation analysis between TFs and diosgenin genes. Table S5: Annotation information of candidate CYP450 genes; Table S6: Correlation analysis between CYP450 gene and diosgenin metabolites; Table S7: The quantitative real-time (qRT-PCR) primer sequence. Figure S1: The expression level of TFs identified in *D. cirrhosa*.

Author Contributions: Conceptualization, Y.L.(Yan Lin) methodology, Y.L. (Yan Lin); software, Y.L.(Yan Lin); validation, Q.H. and Q.Y.; formal analysis, H.Z. and Z.B.; investigation, Q.Y., Y.L. (Yan Lin) and Z.B.; writing—original draft preparation, Y.L. (Yan Lin); Writing—review and editing, Y.L.(Yongping Li) and L. M.; supervision, Y.L. (Yongping Li) and L.M.; All authors have read and agreed to the published version of the manuscript.

Funding:

Conflicts of Interest: The authors declare no conflict of interest.

References

References

1. Seki, H.; Tamura, K.; Muranaka, T. P450s and UGTs: Key players in the structural diversity of triterpenoid saponins. Plant and Cell Physiology 2015, 56, 1463-1471, doi:10.1093/pcp/pcv062.

2. Bak, S.; Kahn, R.A.; Olsen, C.E.; Halkier, B.A. Cloning and expression in Escherichia coli of the obtusifoliol 14 alpha-demethylase of *Sorghum bicolor* (L.) Moench, a cytochrome P450 orthologous to the sterol 14 alpha-demethylases (CYP51) from fungi and mammals. The Plant journal : for cell and molecular biology 1997, 11, 191-201, doi:10.1046/j.1365-313X.1997.11020191.x.

3. Morikawa, T.; Mizutani, M.; Ohta, D. Cytochrome P450 subfamily CYP710A genes encode sterol C-22 desaturase in plants. Biochemical Society Transactions 2006, 34, 1202-1205, doi:10.1042/bst0341202.

4. Kim, J.; DellaPenna, D. Defining the primary route for lutein synthesis in plants: The role of Arabidopsis carotenoid beta-ring hydroxylase CYP97A3. Proceedings of the National Academy of Sciences of the United States of America 2006, 103, 3474-3479, doi:10.1073/pnas.0511207103.

5. Tian, L.; Musetti, V.; Kim, J.; Magallanes-Lundback, M.; DellaPenna, D. The Arabidopsis LUT1 locus encodes a member of the cytochrome P450 family that is required for carotenoid epsilon-ring hydroxylation activity. Proceedings of the National Academy of Sciences of the United States of America 2004, 101, 402-407, doi:10.1073/pnas.2237237100.
6. Babu, P.R.; Rao, K.V.; Reddy, V.D. Structural organization and classification of cytochrome P450 genes in flax (*Linum usitatissimum* L.). Gene 2013, 513, 156-162, doi:10.1016/j.gene.2012.10.040.
7. Liu, M.-H.; Yang, B.-R.; Cheung, W.-F.; Yang, K.Y.; Zhou, H.-F.; Kwok, J.S.-L.; Liu, G.-C.; Li, X.-F.; Zhong, S.; Lee, S.M.-Y. Transcriptome analysis of leaves, roots and flowers of *Panax notoginseng* identifies genes involved in ginsenoside and alkaloid biosynthesis. Bmc Genomics 2015, 16, doi:10.1186/s12864-015-1477-5.
8. Wen, L.; Yun, X.; Zheng, X.; Xu, H.; Zhan, R.; Chen, W.; Xu, Y.; Chen, Y.; Zhang, J. Transcriptomic comparison reveals candidate genes for triterpenoid biosynthesis in two closely related Ilex species. Frontiers in Plant Science 2017, 8, doi:10.3389/fpls.2017.00634.
9. Cheng, Y.; Liu, H.; Tong, X.; Liu, Z.; Zhang, X.; Li, D.; Jiang, X.; Yu, X. Identification and analysis of CYP450 and UGT supergene family members from the transcriptome of *Aralia elata* (Miq.) seem reveal candidate genes for triterpenoid saponin biosynthesis. Bmc Plant Biology 2020, 20, doi:10.1186/s12870-020-02411-6.
10. Li, J.; Liang, Q.; Li, C.; Liu, M.; Zhang, Y. Comparative transcriptome analysis identifies putative genes involved in dioscin biosynthesis in *Dioscorea zingiberensis*. Molecules 2018, 23, doi:10.3390/molecules23020454.
11. Huang, C.-H.; Liu, D.-Z.; Jan, T.-R. Diosgenin, a plant-derived Sapogenin, enhances regulatory T-Cell immunity in the intestine of mice with food allergy. Journal of Natural Products 2010, 73, 1033-1037, doi:10.1021/np900690z.
12. Gong, G.; Qin, Y.; Huang, W. Anti-thrombosis effect of diosgenin extract from *Dioscorea zingiberensis* CH Wright in vitro and in vivo. Phytomedicine 2011, 18, 458-463, doi:10.1016/j.phymed.2010.08.015.
13. He, Z.; Tian, Y.; Zhang, X.; Bing, B.; Zhang, L.; Wang, H.; Zhao, W. Anti-tumour and immunomodulating activities of diosgenin, a naturally occurring steroidal saponin. Natural Product Research 2012, 26, 2243-2246, doi:10.1080/14786419.2011.648192.
14. Khan, H.; Nazir, S.; Farooq, R.K.; Khan, I.N.; Javed, A. Fabrication and assessment of Diosgenin encapsulated stearic acid solid lipid nanoparticles for Its anticancer and antidepressant effects using in vitro and in vivo models. Frontiers in Neuroscience 2022, 15, doi:10.3389/fnins.2021.806713.
15. Sun, W.; Wang, B.; Yang, J.; Wang, W.; Liu, A.; Leng, L.; Xiang, L.; Song, C.; Chen, S. Weighted gene co-expression network analysis of the Dioscin rich medicinal plant *Dioscorea nipponica*. Frontiers in Plant Science 2017, 8, doi:10.3389/fpls.2017.00789.
16. Zhou, C.; Li, X.; Zhou, Z.; Li, C.; Zhang, Y. Comparative transcriptome analysis identifies genes involved in Diosgenin biosynthesis in *Trigonella foenum-graecum* L. Molecules 2019, 24, doi:10.3390/molecules24010140.

17. Gao, J.; Xu, Y.; Hua, C.; Li, C.; Zhang, Y. Molecular cloning and functional characterization of a sterol 3-O-glucosyltransferase involved in biosynthesis of steroidal saponins in *Trigonella foenum-graecum*. Frontiers in Plant Science 2021, 12, doi:10.3389/fpls.2021.809579.
18. Xing, Y.-M.; Li, B.; Liu, L.; Li, Y.; Yin, S.-X.; Yin, S.-C.; Chen, J.; Guo, S.-X. Armillaria mellea symbiosis drives metabolomic and transcriptomic changes in *Polyporus umbellatus Sclerotia*. Frontiers in Microbiology 2022, 12, doi:10.3389/fmicb.2021.792530.
19. Azam, S.S.; Abro, A.; Raza, S.; Saroosh, A. Structure and dynamics studies of sterol 24-C-methyltransferase with mechanism based inactivators for the disruption of ergosterol biosynthesis. Molecular Biology Reports 2014, 41, 4279-4293, doi:10.1007/s11033-014-3299-y.
20. Phillips, D.R.; Rasbery, J.M.; Bartel, B.; Matsuda, S.P.T. Biosynthetic diversity in plant triterpene cyclization. Current Opinion in Plant Biology 2006, 9, 305-314, doi:10.1016/j.pbi.2006.03.004.
21. Suzuki, M.; Muranaka, T. Molecular genetics of plant sterol backbone synthesis. Lipids 2007, 42, 47-54, doi:10.1007/s11745-006-1000-5.
22. Morlacchi, P.; Wilson, W.K.; Xiong, Q.B.; Bhaduri, A.; Sttivend, D.; Kolesnikova, M.D.; Matsuda, S.P.T. Product profile of PEN3: The last unexamined oxidosqualene cyclase in *Arabidopsis thaliana*. Organic Letters 2009, 11, 2627-2630, doi:10.1021/ol9005745.
23. Ohyama, K.; Suzuki, M.; Kikuchi, J.; Saito, K.; Muranaka, T. Dual biosynthetic pathways to phytosterol via cycloartenol and lanosterol in Arabidopsis. Proceedings of the National Academy of Sciences of the United States of America 2009, 106, 725-730, doi:10.1073/pnas.0807675106.
24. Koolman, J.; Roehm, K.H. Color Atlas of Biochemistry 2nd edition. 2005.
25. Cao, L.; Zhou, Z.; Sun, J.; Li, C.; Zhang, Y. Altering sterol composition implied that cholesterol is not physiologically associated with Diosgenin biosynthesis in *Trigonella foenum-graecum*. Frontiers in Plant Science 2021, 12, doi:10.3389/fpls.2021.741604.
26. Fujita, S.; Ohnishi, T.; Watanabe, B.; Yokota, T.; Takatsuto, S.; Fujioka, S.; Yoshida, S.; Sakata, K.; Mizutani, M. Arabidopsis CYP90B1 catalyses the early C-22 hydroxylation of C-27, C-28 and C-29 sterols. Plant Journal 2006, 45, 765-774, doi:10.1111/j.1365-313X.2005.02639.x.
27. Kim, H.B.; Schaller, H.; Goh, C.H.; Kwon, M.; Choe, S.; An, C.S.; Durst, F.; Feldmann, K.A.; Feyereisen, R. Arabidopsis cyp51 mutant shows postembryonic seedling lethality associated with lack of membrane integrity. Plant Physiology 2005, 138, 2033-2047, doi:10.1104/pp.105.061598.
28. Li, L.; Cheng, H.; Gai, J.; Yu, D. Genome-wide identification and characterization of putative cytochrome P450 genes in the model legume *Medicago truncatula*. Planta 2007, 226, 109-123, doi:10.1007/s00425-006-0473-z.
29. Seki, H.; Ohyama, K.; Sawai, S.; Mizutani, M.; Ohnishi, T.; Sudo, H.; Akashi, T.; Aoki, T.; Saito, K.; Muranaka, T. Licorice beta-amyrin 11-oxidase, a cytochrome P450 with a key role in the biosynthesis of the triterpene sweetener glycyrrhizin. Proceedings of the National Academy of Sciences of the United States of America 2008, 105, 14204-14209, doi:10.1073/pnas.0803876105.

30. Shibuya, M.; Hoshino, M.; Katsube, Y.; Hayashi, H.; Kushiro, T.; Ebizuka, Y. Identification of beta-amyrin and sophoradiol 24-hydroxylase by expressed sequence tag mining and functional expression assay. Febs Journal 2006, 273, 948-959, doi:10.1111/j.1742-4658.2006.05120.x.
31. Lin, X.; Kaul, S.; Rounsley, S.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.; Mason, T.; Bowman, C.L.; Barnstead, M. Sequence and analysis of chromosome 2 of the plant *Arabidopsis thaliana*. Nature 1999, 402, 761-768, doi:10.1038/45471.
32. Vasav, A.P.; Barvkar, V.T. Phylogenomic analysis of cytochrome P450 multigene family and their differential expression analysis in *Solanum lycopersicum* L. suggested tissue specific promoters. Bmc Genomics 2019, 20, doi:10.1186/s12864-019-5483-x.
33. Helliwell, C.A.; Chandler, P.M.; Poole, A.; Dennis, E.S.; Peacock, W.J. The CYP88A cytochrome P450, ent-kaurenoic acid oxidase, catalyzes three steps of the gibberellin biosynthesis pathway. Proceedings of the National Academy of Sciences of the United States of America 2001, 98, 2065-2070, doi:10.1073/pnas.041588998.
34. Sawai, S.; Saito, K. Triterpenoid biosynthesis and engineering in plants. Frontiers in Plant Science 2011, 2, doi:10.3389/fpls.2011.00025.
35. Vaccaro, M.; Bernal, V.O.; Malafronte, N.; De Tommasi, N.; Leone, A. High yield of bioactive abietane diterpenes in *Salvia sclarea* hairy roots by overexpressing cyanobacterial DXS or DXR genes. Planta Medica 2019, 85, 973-980, doi:10.1055/a-0895-5878.
36. Han, J.-Y.; In, J.-G.; Kwon, Y.-S.; Choi, Y.-E. Regulation of ginsenoside and phytosterol biosynthesis by RNA interferences of squalene epoxidase gene in *Panax ginseng*. Phytochemistry 2010, 71, 36-46, doi:10.1016/j.phytochem.2009.09.031.
37. Jo, H.-J.; Han, J.Y.; Hwang, H.-S.; Choi, Y.E. Beta-Amyrin synthase (EsBAS) and 13-amyrin 28-oxidase (CYP716A244) in oleanane-type triterpene saponin biosynthesis in *Eleutherococcus senticosus*. Phytochemistry 2017, 135, 53-63, doi:10.1016/j.phytochem.2016.12.011.
38. Nes, W.D. Biosynthesis of Cholesterol and Other Sterols. Chemical Reviews 2011, 111, 6423-6451, doi:10.1021/cr200021m.
39. Diarra, S.T.; He, J.; Wang, J.; Li, J. Ethylene treatment improves diosgenin accumulation in in vitro cultures of *Dioscorea zingiberensis* via up-regulation of CAS and HMGR gene expression. Electronic Journal of Biotechnology 2013, 16, doi:10.2225/vol16-issue5-fulltext-9.
40. Gong, Y.F.; Liao, Z.H.; Guo, B.H.; Sun, X.F.; Tang, K.X. Molecular cloning and expression profile analysis of *Ginkgo biloba* DXS gene encoding 1-deoxy-D-xylulose 5-phosphate synthase, the first committed enzyme of the 2-C-methyl-D-erythritol 4-phosphate pathway. Planta Medica 2006, 72, 329-335, doi:10.1055/s-2005-916234.
41. Christ, B.; Xu, C.C.; Xu, M.L.; Li, F.S.; Wada, N.; Mitchell, A.J.; Han, X.L.; Wen, M.L.; Fujita, M.; Weng, J.K. Repeated evolution of cytochrome P450-mediated spiroketal steroid biosynthesis in plants. Nature Communications 2019, 10, doi:10.1038/s41467-019-11286-7.
42. Ghosh, S. Triterpene structural diversification by plant cytochrome P450 enzymes. Frontiers in Plant Science 2017, 8, doi:10.3389/fpls.2017.01886.

43. Osbourn, A.; Goss, R.J.M.; Field, R.A. The saponins - polar isoprenoids with important and diverse biological activities. Natural Product Reports 2011, 28, 1261-1268, doi:10.1039/c1np00015b.
44. Fraga, C.G.; Clowers, B.H.; Moore, R.J.; Zink, E.M. Signature-discovery approach for sample matching of a nerve-agent precursor using liquid chromatography-Mass spectrometry, XCMS, and chemometrics. Analytical Chemistry 2010, 82, 4165-4173, doi:10.1021/ac1003568.
45. Grabherr, M.G.; Haas, B.J.; Yassour, M. Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nature Biotechnology 2011, 29, 644-652, doi:10.1038/nbt.1883.
46. Davidson, N.M.; Oshlack, A. Corset: enabling differential gene expression analysis for de novo assembled transcriptomes. Genome Biology 2014, 15(7), 1-14, doi: 10.1186/s13059-014-0410-6.
47. Wang, L.G.; Wang, S.Q.; Li, W. RSeQC: Quality control of RNA-Seq experiments. BMC Bioinformatics 2012, 28(16), 2184-5, doi: 10.1093/bioinformatics/bts356.
48. Pertea, M.; Pertea, G.M.; Antonescu, C.M.; Chang, T.C.; Mendell, J.T.; Salzberg, S.L. StringTie enables improved reconstruction of a transcriptome from RNA-Seq reads. Nature Biotechnology 2015, 33(3), 290-295, doi: 10.1038/nbt.3122.
49. Zheng, Y.; Jiao, C.; Sun ,H.; Rosli, H.G.; Pombo, M.A.; Zhang, P.; Banf, M.; Dai, X.; Martin, G.B.; Giovannoni, J.J.; Zhao, P.X.; Rhee, S.Y.; Fei, Z. iTAK: A program for genome-wide prediction and classification of plant transcription factors, transcriptional regulators, and protein kinases. Molecular Plant 2016, 9(12), 1667-1670, doi: 10.1016/j.molp.2016.09.014.
50. Larkin, M.A.; Blackshields, G.; Brown, N.P.; Chenna, R.; McGettigan, P.A.; McWilliam, H.; Valentin, F.; Wallace, I.M.; Wilm, A.; Lopez, R. Clustal W and clustal X version 2.0. Bioinformatics 2007, 23, 2947-2948, doi:10.1093/bioinformatics/btm404.
51. Tamura, K.; Stecher, G.; Peterson, D.; Filipski, A.; Kumar, S. MEGA6: molecular evolutionary genetics analysis version 6.0. Molecular Biology and Evolution 2013, 30, 2725-2729, doi:10.1093/molbev/mst197.

Figure 1

The tuber color differences of *D. cirrhosa*.

LR indicates light red tuber, RD indicates red tuber, DR indicates dark red tuber, and BR indicate brownish red tuber.

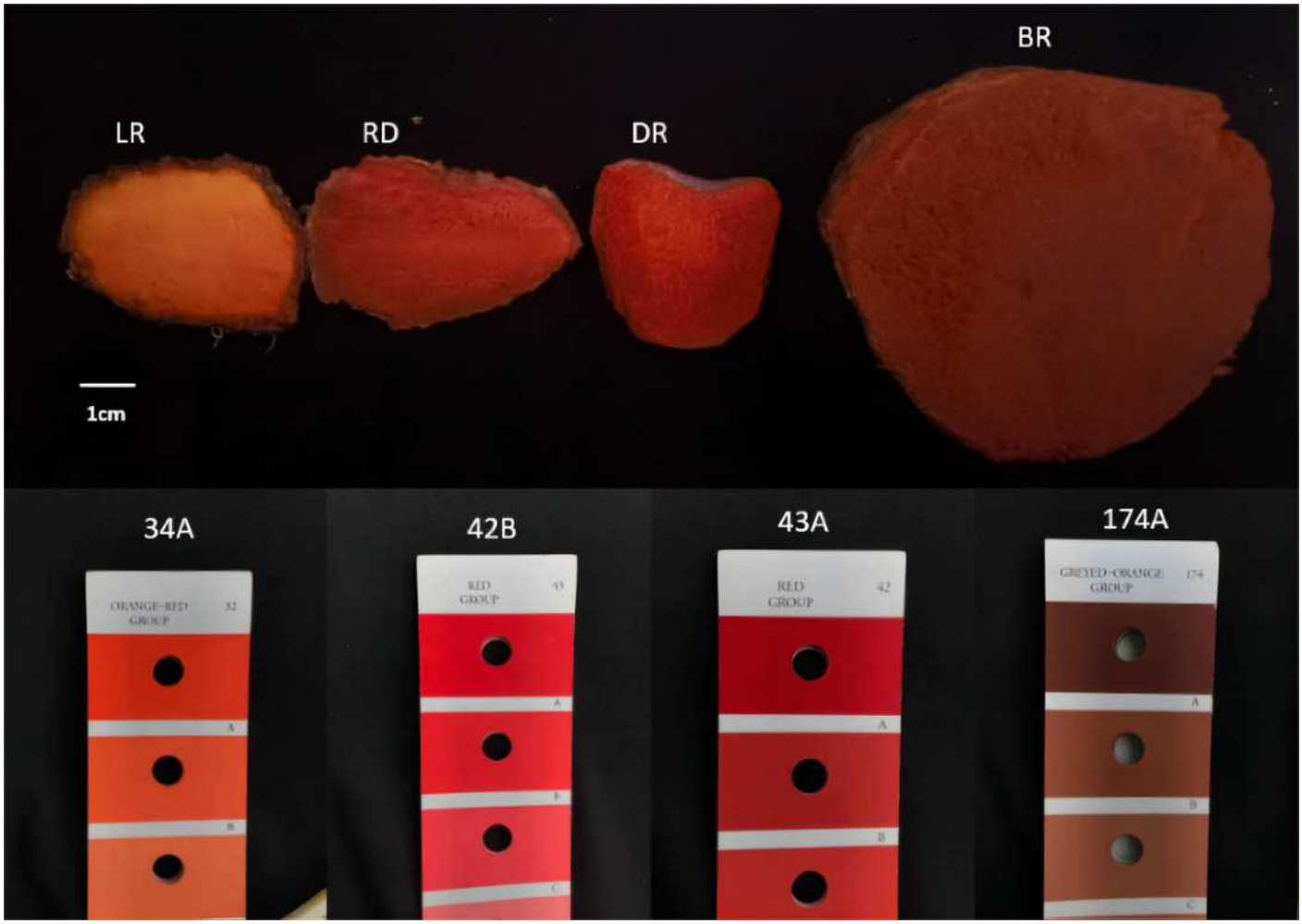

Figure 2

Differences of diosgenin metabolites in *D. cirrhosa* tubers.

Three biological replicates are shown in boxplot. LR: light red; RD: red; DR: dark red; BR: brownish red. Each color represents one tuber, an asterisk (*) indicates significance level and NS indicates non-significant (* p <0.05, ** p < 0.01, *** p < 0.001).

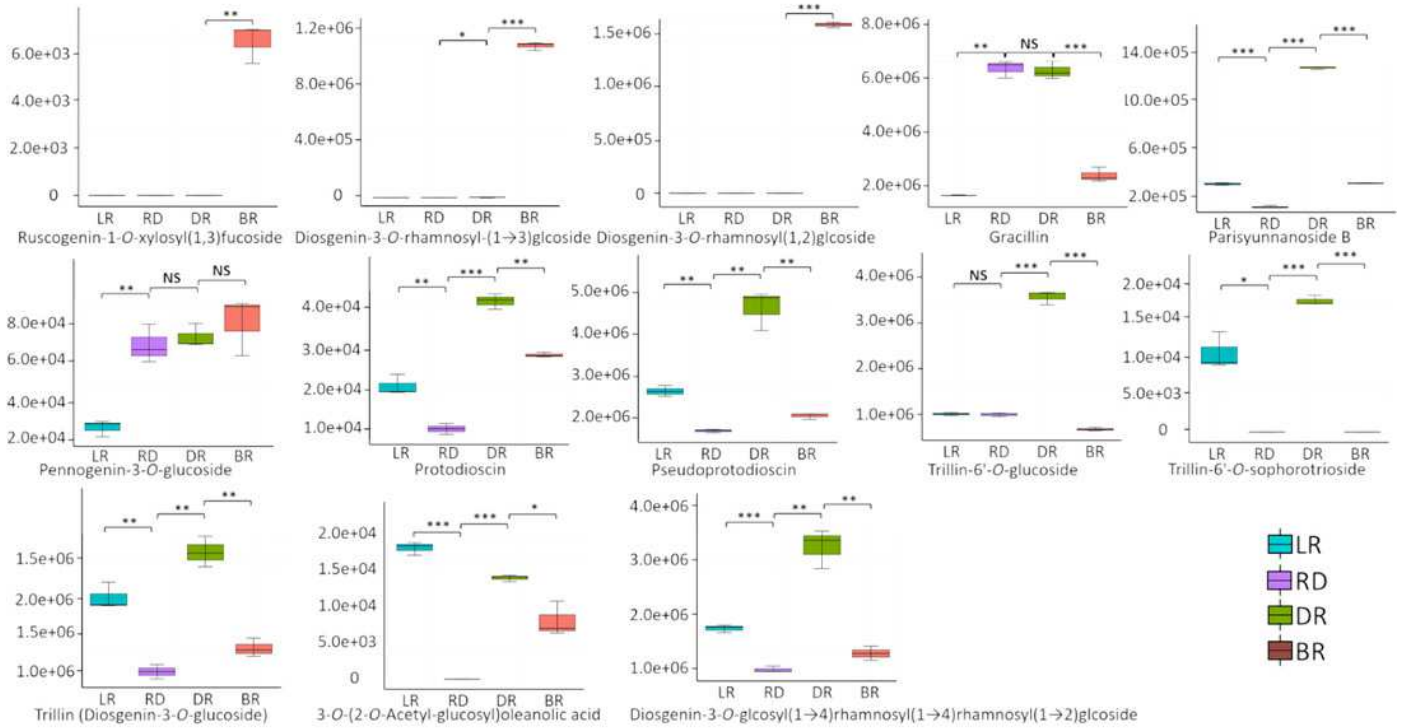

Figure 3

Diosgenin related metabolites in *D. cirrhosa* tubers. The metabolites were scaled using Z-score of relative content (mean value of three replications) in the heatmap. LR: light red; RD: red; DR: dark red; BR: brownish red.

The metabolites were scaled using Z-score of relative content (mean value of three replications) in the heatmap. LR: light red; RD: red; DR: dark red; BR: brownish red.

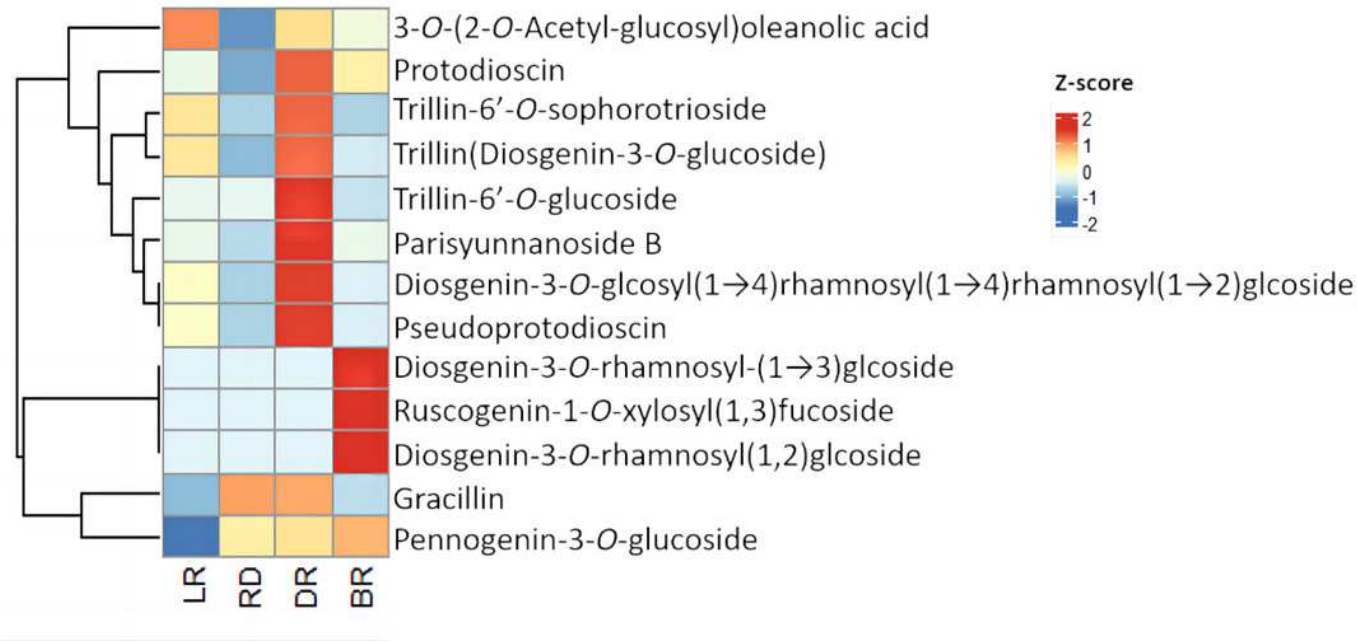

Figure 4

Diosgenin biosynthesis pathway and related genes expression in *D. cirrhosa*.

AACT, Acetoacetyl-CoA thiolase; HMGS, hydroxy-3-methylglutaryl-CoA Synthase; HMGR, 3-hydroxy-3-methylglutaryl-CoA Reductase; MVK, Mevalonate kinase; PMK, 5-phosphomevalonate kinase; MVD, Diphosphomevalonate decarboxylase; DXS, 1-deoxy-D-xylulose-5-phosphate synthase; DXR, 1-deoxy-D-xylulose-5-phosphate reductoisomerase; ispD, 2-C-methyl-D-erythritol 11-phosphate cytidylyltransferase; ispE, 4-diphosphocytidyl-5-C-methyl-D-erythritol kinase; ispF, 2-C-methyl-D-erythritol 2,4-cyclodiphosphate synthase; gcpE, (E)-4-hydroxy-3-methylbut-3-enyl-diphosphate synthase; ispH, 4-hydroxy-3-methylbut-2-en-4-yl diphosphate reductase; FPS, Farnesyl pyrophasphate synthase; SS, Squalene synthase; SE, Squalene epoxidase; CAS, Cycloartenol synthase; LSS, Lanosterol synthase; SMT, sterol 24-C-methyltransferase; C5(6), Sterol desaturase; C14-R, Sterol C14-reductase; UGT, Uridine diphosphate (UDP)-dependent glycosyltransferases. The color scale represents the average of FPKM value (scaled using Z-score), red color indicates high expression and purple color indicate low expression.

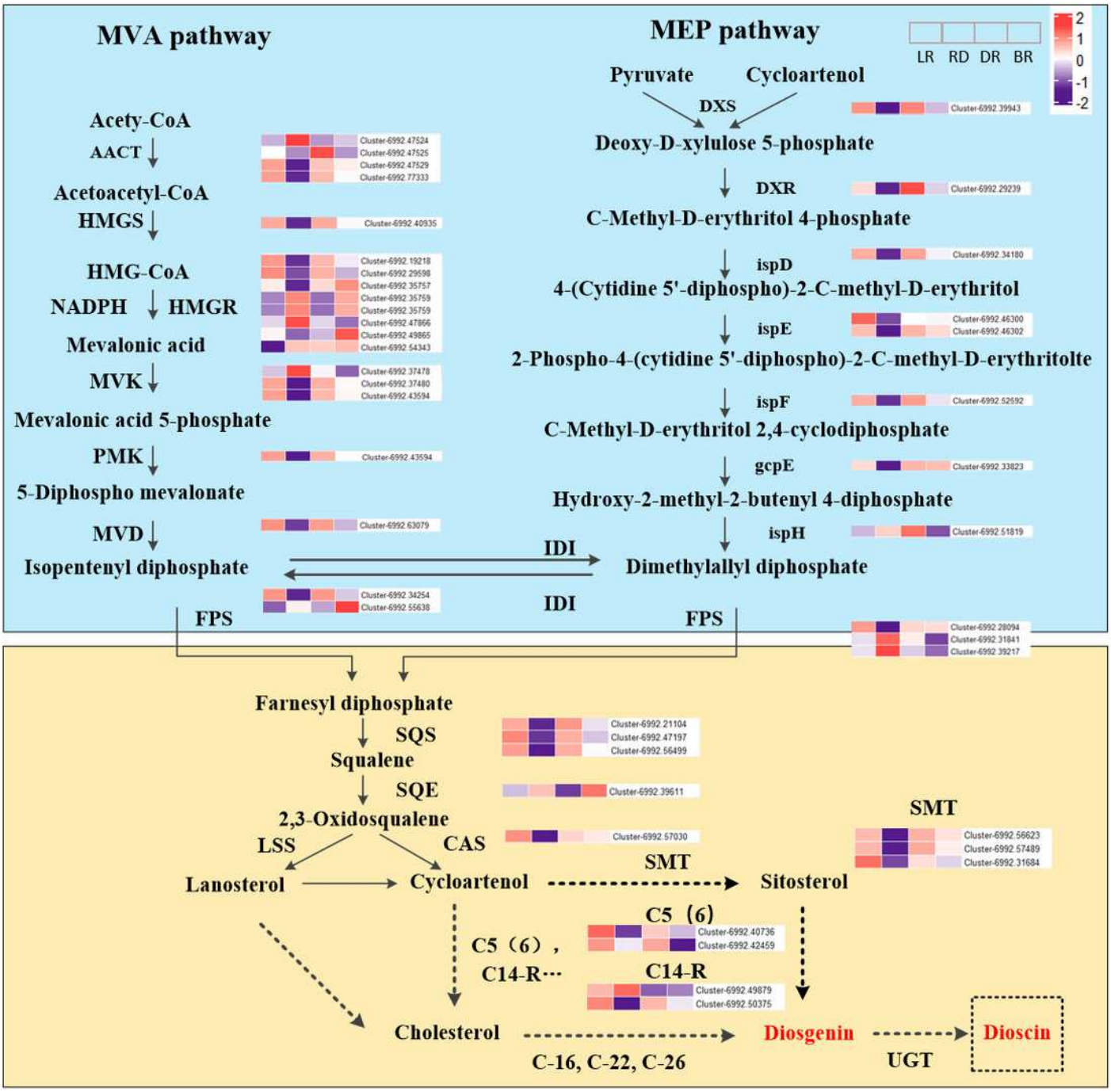
MVA pathway
MEP pathway
LR RD DR BR
2 1 0 -1 -2
Pyruvate
Cycloartenol
DXS
Deoxy-D-xylulose 5-phosphate
DXR
C-Methyl-D-erythritol 4-phosphate
ispD
4-(Cytidine 5'-diphospho)-2-C-methyl-D-erythritol
ispE
2-Phospho-4-(cytidine 5'-diphospho)-2-C-methyl-D-erythritolte
ispF
C-Methyl-D-erythritol 2,4-cyclodiphosphate
gcpE
Hydroxy-2-methyl-2-butenyl 4-diphosphate
ispH
Dimethylallyl diphosphate
IDI
IDI
Acety-CoA
AACT
Acetoacetyl-CoA
HMGS
HMG-CoA
NADPH
HMGR
Mevalonic acid
MVK
Mevalonic acid 5-phosphate
PMK
5-Diphospho mevalonate
MVD
Isopentenyl diphosphate
FPS
FPS
Farnesyl diphosphate
SQS
Squalene
SQE
2,3-Oxidosqualene
LSS
CAS
Lanosterol
Cycloartenol
SMT
Sitosterol
SMT
C5（6）,
C14-R…
C5（6）
C14-R
Cholesterol
Diosgenin
Dioscin
C-16, C-22, C-26
UGT

Figure 5

Regulatory networks of TFs and diosgenin genes. Each yellow circle represents a TF, and each blue circle represents diosgenin a gene.

Each yellow circle represents a TF, and each blue circle represents diosgenin a gene.

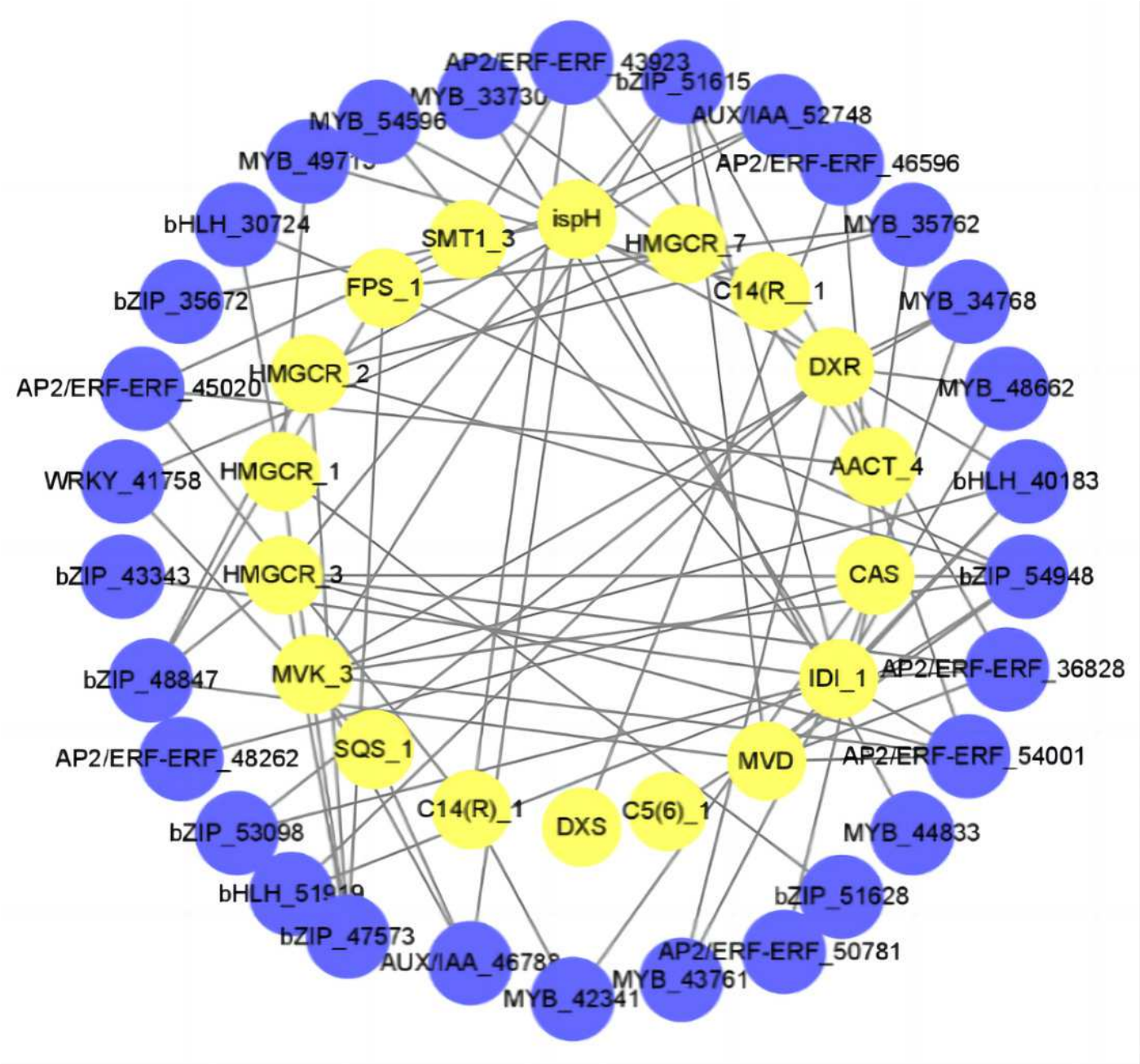

Figure 6

The expression patterns of CYP450s in *D. cirrhosa* tubers. Z-score obtained from averaged FPKM of three biological replicates was used. Red color indicates high expression and yellow color indicate low expression.

Z-score obtained from averaged FPKM of three biological replicates was used. Red color indicates high expression and yellow color indicate low expression.

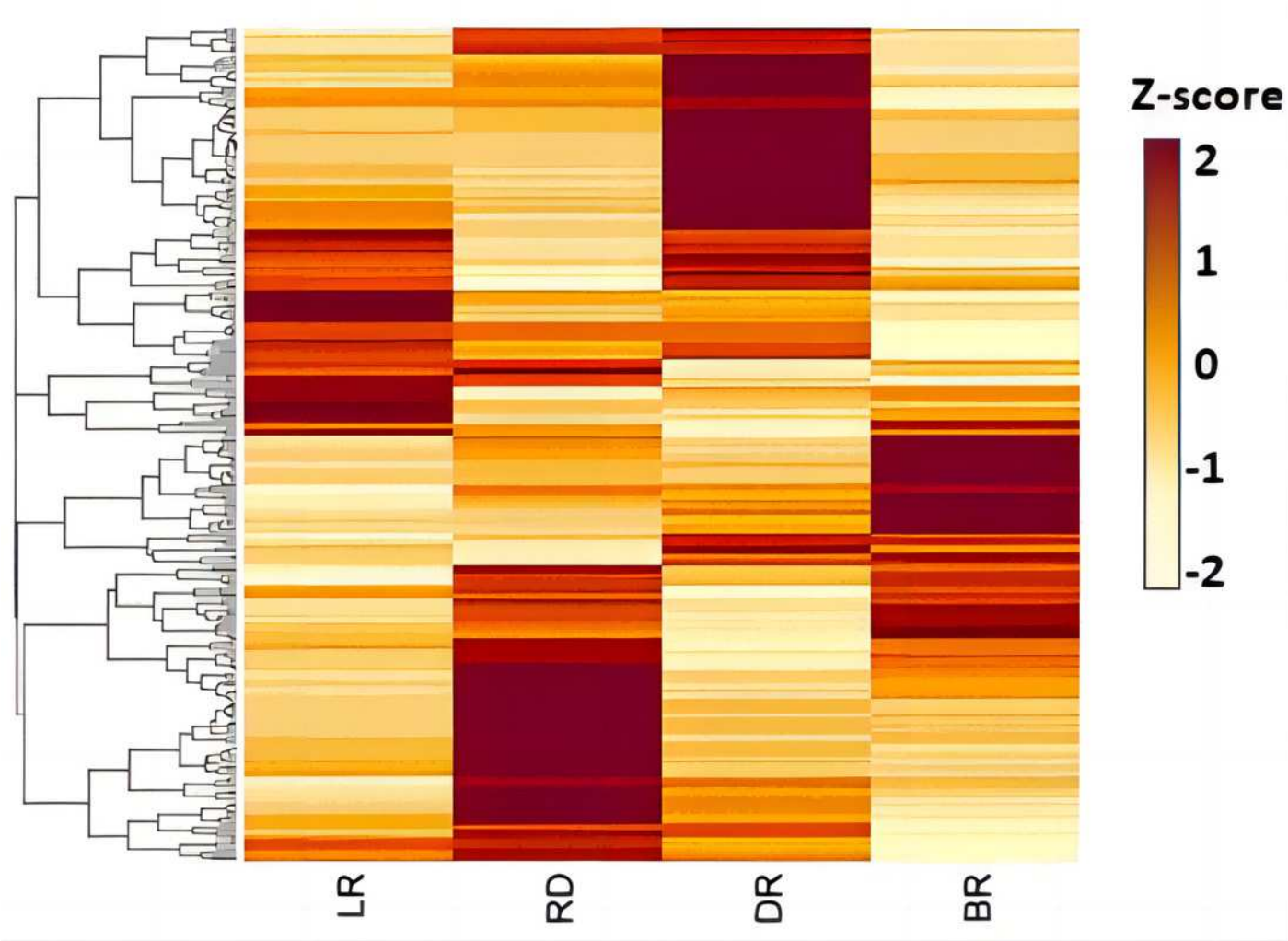

Figure 7

Expression levels of CYP450 genes between *D. cirrhosa* tubers.

The genes annotated as steroid-22-α-hydroxylase (CYP90B) and sterol-14-demethylase were marked by the blue arrow. Z-score obtained from averaged FPKM of three biological replicates was used. Red color indicates high expression and blue color indicate low expression.

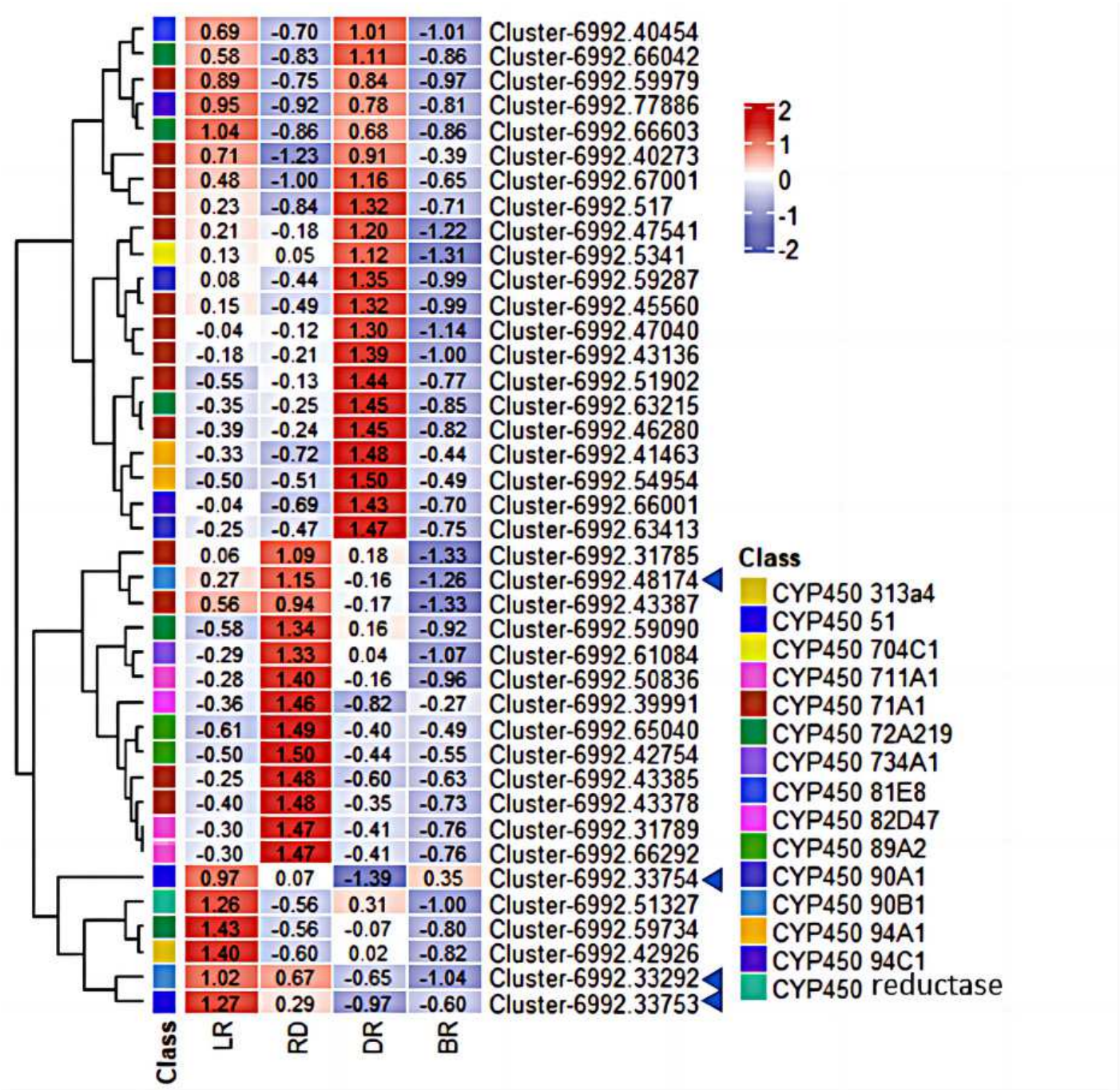
Class
LR
RD
DR
BR
0.69 -0.70 1.01 -1.01 Cluster-6992.40454
0.58 -0.83 1.11 -0.86 Cluster-6992.66042
0.89 -0.75 0.84 -0.97 Cluster-6992.59979
0.95 -0.92 0.78 -0.81 Cluster-6992.77886
1.04 -0.86 0.68 -0.86 Cluster-6992.66603
0.71 -1.23 0.91 -0.39 Cluster-6992.40273
0.48 -1.00 1.16 -0.65 Cluster-6992.67001
0.23 -0.84 1.32 -0.71 Cluster-6992.517
0.21 -0.18 1.20 -1.22 Cluster-6992.47541
0.13 0.05 1.12 -1.31 Cluster-6992.5341
0.08 -0.44 1.35 -0.99 Cluster-6992.59287
0.15 -0.49 1.32 -0.99 Cluster-6992.45560
-0.04 -0.12 1.30 -1.14 Cluster-6992.47040
-0.18 -0.21 1.39 -1.00 Cluster-6992.43136
-0.55 -0.13 1.44 -0.77 Cluster-6992.51902
-0.35 -0.25 1.45 -0.85 Cluster-6992.63215
-0.39 -0.24 1.45 -0.82 Cluster-6992.46280
-0.33 -0.72 1.48 -0.44 Cluster-6992.41463
-0.50 -0.51 1.50 -0.49 Cluster-6992.54954
-0.04 -0.69 1.43 -0.70 Cluster-6992.66001
-0.25 -0.47 1.47 -0.75 Cluster-6992.63413
0.06 1.09 0.18 -1.33 Cluster-6992.31785
0.27 1.15 -0.16 -1.26 Cluster-6992.48174
0.56 0.94 -0.17 -1.33 Cluster-6992.43387
-0.58 1.34 0.16 -0.92 Cluster-6992.59090
-0.29 1.33 0.04 -1.07 Cluster-6992.61084
-0.28 1.40 -0.16 -0.96 Cluster-6992.50836
-0.36 1.46 -0.82 -0.27 Cluster-6992.39991
-0.61 1.49 -0.40 -0.49 Cluster-6992.65040
-0.50 1.50 -0.44 -0.55 Cluster-6992.42754
-0.25 1.48 -0.60 -0.63 Cluster-6992.43385
-0.40 1.48 -0.35 -0.73 Cluster-6992.43378
-0.30 1.47 -0.41 -0.76 Cluster-6992.31789
-0.30 1.47 -0.41 -0.76 Cluster-6992.66292
0.97 0.07 -1.39 0.35 Cluster-6992.33754
1.26 -0.56 0.31 -1.00 Cluster-6992.51327
1.43 -0.56 -0.07 -0.80 Cluster-6992.59734
1.40 -0.60 0.02 -0.82 Cluster-6992.42926
1.02 0.67 -0.65 -1.04 Cluster-6992.33292
1.27 0.29 -0.97 -0.60 Cluster-6992.33753
2
1
0
-1
-2
Class
CYP450 313a4
CYP450 51
CYP450 704C1
CYP450 711A1
CYP450 71A1
CYP450 72A219
CYP450 734A1
CYP450 81E8
CYP450 82D47
CYP450 89A2
CYP450 90A1
CYP450 90B1
CYP450 94A1
CYP450 94C1
CYP450 reductase

Figure 8

Phylogenetic analysis of CYP450 candidates from *D. cirrhosa* transcriptome and previously published CYP450s from various metabolic pathways.

The genes with red dots indicate hub genes associated with sterol C-22 hydroxylase and sterol-14-demethylase.

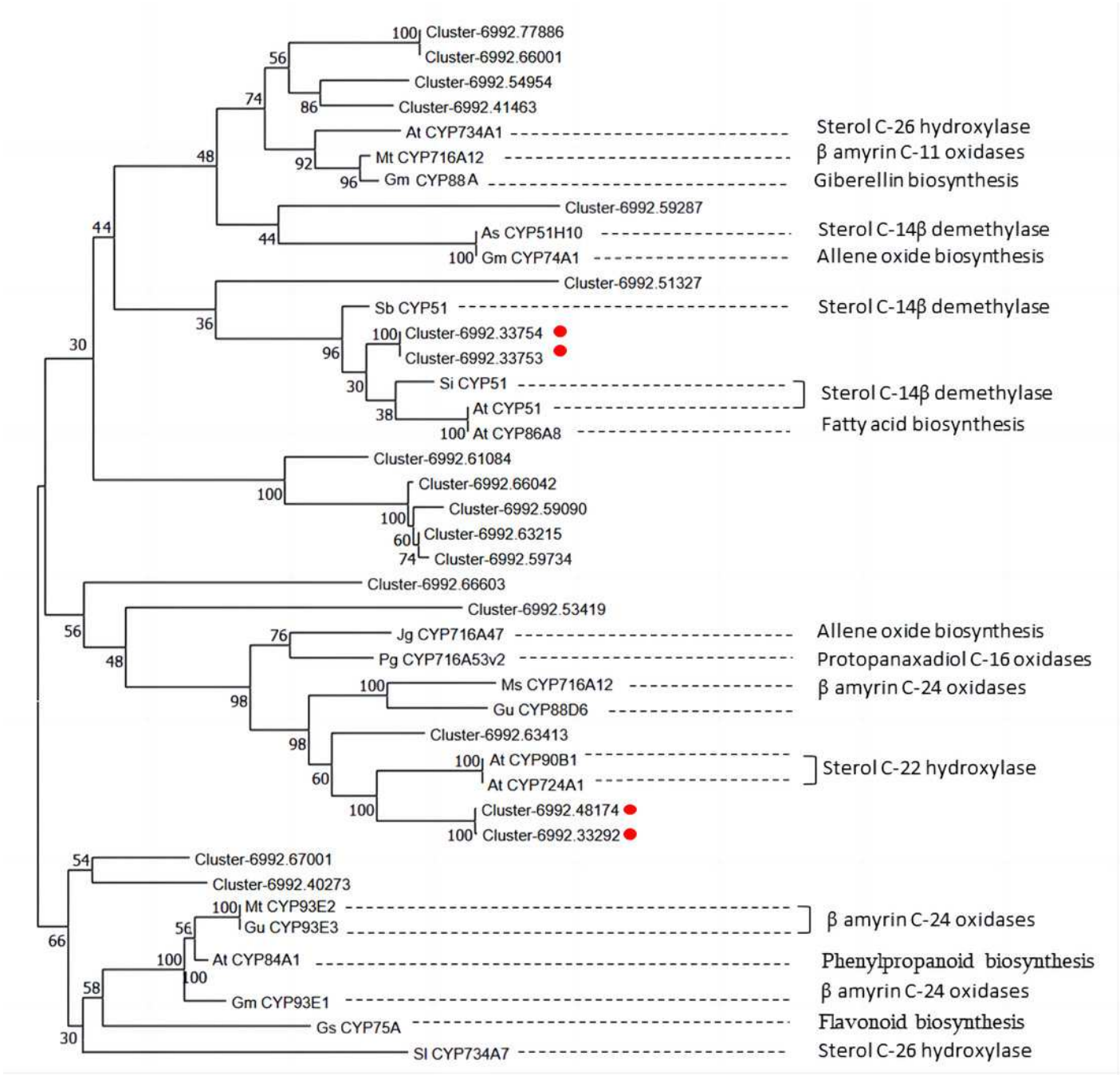
Cluster-6992.77886
Cluster-6992.66001
Cluster-6992.54954
Cluster-6992.41463
At CYP734A1
Mt CYP716A12
Gm CYP88A
Sterol C-26 hydroxylase
β amyrin C-11 oxidases
Giberellin biosynthesis
Cluster-6992.59287
As CYP51H10
Gm CYP74A1
Sterol C-14β demethylase
Allene oxide biosynthesis
Cluster-6992.51327
Sb CYP51
Cluster-6992.33754
Cluster-6992.33753
Si CYP51
At CYP51
At CYP86A8
Sterol C-14β demethylase
Fatty acid biosynthesis
Cluster-6992.61084
Cluster-6992.66042
Cluster-6992.59090
Cluster-6992.63215
Cluster-6992.59734
Cluster-6992.66603
Cluster-6992.53419
Jg CYP716A47
Pg CYP716A53v2
Allene oxide biosynthesis
Protopanaxadiol C-16 oxidases
Ms CYP716A12
Gu CYP88D6
β amyrin C-24 oxidases
Cluster-6992.63413
At CYP90B1
At CYP724A1
Sterol C-22 hydroxylase
Cluster-6992.48174
Cluster-6992.33292
Cluster-6992.67001
Cluster-6992.40273
Mt CYP93E2
Gu CYP93E3
β amyrin C-24 oxidases
At CYP84A1
Phenylpropanoid biosynthesis
Gm CYP93E1
β amyrin C-24 oxidases
Gs CYP75A
Flavonoid biosynthesis
Sl CYP734A7
Sterol C-26 hydroxylase

Figure 9

Heatmap of Pearson correlation analysis between diosgenin genes and diosgenin metabolites.

The value in the box represents the r2 value, and 0.00 indicates no significance. D.1: Ruscogenin-1-O-xylosyl(1,3)fucoside; D.2: Diosgenin-3-O-rhamnosyl(1,2)glcoside; D.3: Diosgenin-3-O-rhamnosyl-(1→3)glcoside; D.4: Gracillin; D.5: Pennogenin-3-O-glucoside; D.6: Pseudoprotodioscin; D.7: Diosgenin-3-O-glcosyl(1→4)rhamnosyl(1→4)rhamnosyl(1→2)glcoside; D.8: Trillin (Diosgenin-3-O-glucoside); D.9: Trillin-6'-O-glucoside; D.10: Parisyunnanoside B; D.11: Trillin-6'-O-sophorotrioside; D.12: 3-O-(2-O-Acetyl-glucosyl)oleanolic acid.

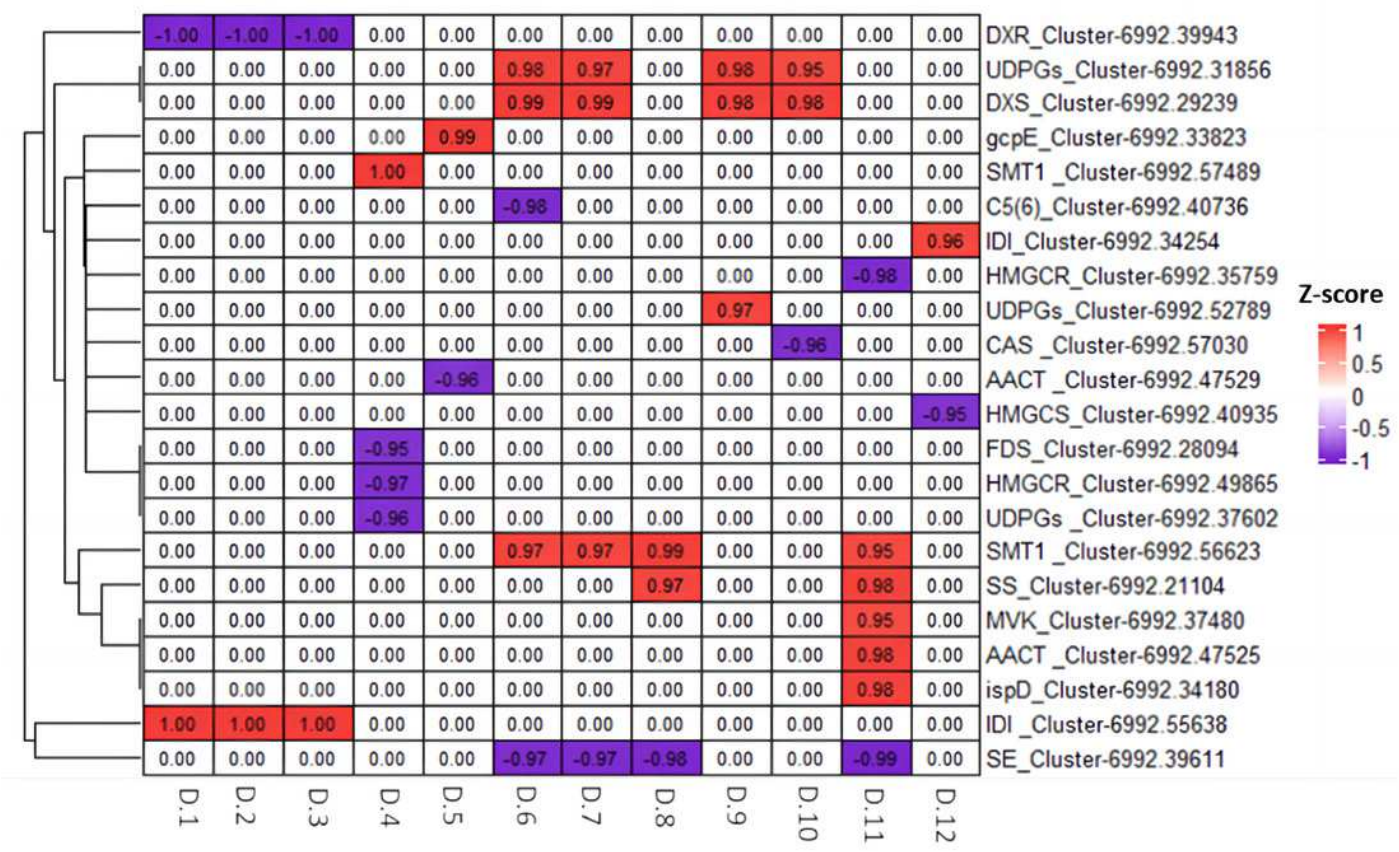

Figure 10

RT-PCR of key genes in diosgenin biosynthesis pathway.

Results represent mean values ± SE from three biological replicates. Asterisks (*) indicate the statistical significance of the difference between LR and RD, RD and DR, DR and BR.

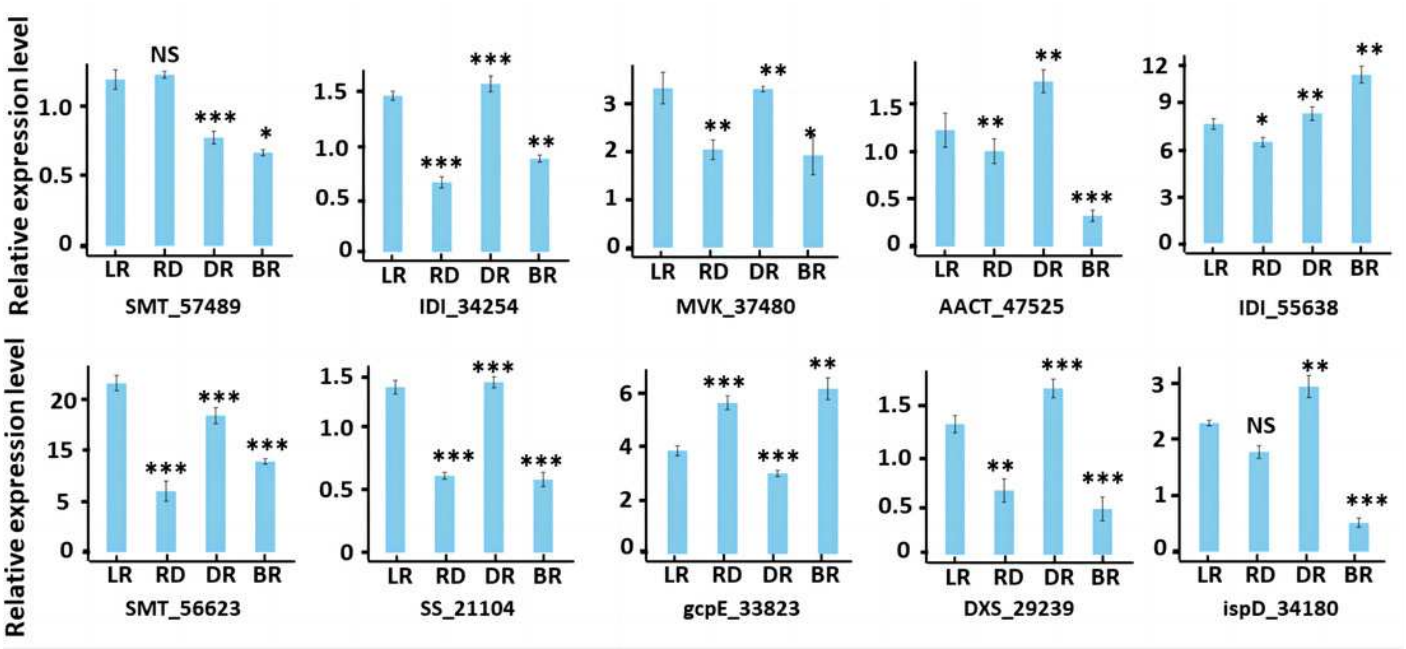

Table 1(on next page)

Genes involved in diosgenin biosynthesis pathway in *D. cirrhosa*.

1 Table 1. Genes involved in diosgenin biosynthesis pathway in *D. cirrhosa*.

Gene	Enzyme	Number	KEGG
SQE	Squalene epoxidase	1	K00511
FPS	Farnesyl pyrophasphate synthase	3	K13789
C14-R	Sterol C14-reductase	2	K00222
C5(6)	Sterol desaturase	2	K00227
MVK	Mevalonate kinase	3	K00869
HMGCR	Hydroxymethylglutaryl-CoA reductase	8	K01661
DXS	1-deoxy-D-xylulose-5-phosphate synthase	1	K01662
ispE	4-diphosphocytidyl-5-C-methyl-D-erythritol kinase	2	K00919
ispF	2-C-methyl-D-erythritol 2,4-cyclodiphosphate synthase	1	K01770
gcpE	(E)-4-hydroxy-3-methylbut-3-enyl-diphosphate synthase	1	K03526
ispH	4-hydroxy-3-methylbut-2-en-4-yl diphosphate reductase	1	K03527
IDI	Isopentenyl-diphosphate Delta-isomerase	2	K01823
DXR	1-deoxy-D-xylulose-6-phosphate reductoisomerase	1	K00099
HMGCS	Hydroxymethylglutaryl-CoA synthase	1	K01641
CAS	Cycloartenol synthase	1	K01853
ispD	2-C-methyl-D-erythritol 11-phosphate cytidylyltransferase	1	K00991
SS	Squalene synthase	3	K00919
PMK	Phosphomevalonate kinase	1	K00938
MVD	Diphosphomevalonate decarboxylase	1	K01597
AACT	Acetyl-CoA C-acetyltransferase	4	K00626
SMT	Sterol 24-C-methyltransferase	3	K00559
	All	43	24

2

Table 2(on next page)

TFs identified in *D. cirrhosa*.

Table 2. TFs identified in *D. cirrhosa*.

TFs	Number	LR vs RD (All/up)	RD vs DR （All/up）	DR vs BR （All/up）
AP2/ERF	82	26/9	56/47	50/15
MYB	88	42/16	54/30	48/29
WRKY	50	23/9	39/26	25/9
NAC	51	26/11	33/25	27/8
C2H2	56	27/14	31/16	34/23
bZIP	55	30/14	29/15	31/21
bHLH	53	24/14	28/13	22/14
C3H	52	16/11	24/7	34/27
SET	32	14/12	9/1	17/16
HB-HD-ZIP	31	16/7	22/15	18/12
GRAS	23	15/7	19/12	13/4
C2C2-Dof		14/5	15/10	13/10
SNF2	49	23/22	15/1	31/31
PHD	31	11/9	14/4	19/18
GARP-G2-like	23	13/8	13/5	21/16
AUX/IAA	20	7/3	12/10	13/6

Table 3(on next page)

Information of published CYP450 genes.

Table 3. Information of published CYP450 genes.

Gene	Accession No.	Species
Gm CYP74A1	NM_001289366.1	*Glycine max*
Sl CYP51	NM_001247608.2	*Solanum lycopersicum*
Sc CYP51	AY552551.1	*Solanum chacoense*
Sb CYP51	U74319.1	*Sorghum bicolor*
Sl CYP734A7	NM_001247011.2	*Solanum lycopersicum*
Pk CYP73A	D82812.1	*Populus kitakamiensis*
Pg CYP716A53v2	JX036031.1	*Panax ginseng*
Mt CYP93E2	DQ335790.1	*Medicago truncatula*
Mt CYP716A12	DQ335781.1	*Medicago truncatula*
Ms CYP716A12	KM978958.1	*Medicago sativa*
At CYP724A1	NM_121444.4	*Arabidopsis thaliana*
GS CYP75A	MW298105.1	*Glycine soja*
Gm CYP93E1	NM_001249225.2	*Glycine max*
Mt CYP93E2	DQ335790.1	*Medicago truncatula*
Gu CYP93E3	AB437320.1	*Glycyrrhiza uralensis*
At CYP734A1	NM_128228.4	*Arabidopsis thaliana*
Gu CYP88D6	AB433179.1	*Glycyrrhiza uralensis*
At CYP86A1	NM_125276.3	*Arabidopsis thaliana*
Pg CYP716A47	JN604537.1	*Panax ginseng*
At CYP90B1	AF044216.1	*Arabidopsis thaliana*
At CYP86A8	AJ301678.1	*Arabidopsis thaliana*
At CYP84A1	AY666123.1	*Arabidopsis thaliana*
As CYP51H10	DQ680849.1	*Avena strigosa*
At CYP51	AY666123.1	*Arabidopsis thaliana*

Table 4(on next page)

The FPKM value of UTGs in *D. cirrhosa*.

1 Table 4. The FPKM value of UTGs in *D. cirrhosa*.

2

Gene ID	LR	RD	DR	BR	Predited
Cluster-6992.52789	7.71	10.63	24.81	2.76	*Elaeis guineensis*
Cluster-6992.47459	35.42	2217.30	33.54	360.04	*Elaeis guineensis*
Cluster-6992.41090	1877.13	2848.44	2194.17	250.56	*Phoenix dactylifera*
Cluster-6992.37602	2879.41	1208.69	1205.58	2108.93	*Phoenix dactylifera*
Cluster-6992.31856	11.11	8.85	20.45	7.34	*Elaeis guineensis*

Table 5(on next page)

The FPKM value of UTGs in *D. cirrhosa*.

1 Table 5. The FPKM value of UTGs in *D. cirrhosa*.

2

Gene ID	LR	RD	DR	BR	Predited
Cluster-6992.52789	7.71	10.63	24.81	2.76	*Elaeis guineensis*
Cluster-6992.47459	35.42	2217.30	33.54	360.04	*Elaeis guineensis*
Cluster-6992.41090	1877.13	2848.44	2194.17	250.56	*Phoenix dactylifera*
Cluster-6992.37602	2879.41	1208.69	1205.58	2108.93	*Phoenix dactylifera*
Cluster-6992.31856	11.11	8.85	20.45	7.34	*Elaeis guineensis*