

Transcriptomic Analysis Reveals the Mechanism of Glycyrrhizic acid Biosynthesis in *Glycyrrhiza Uralensis* Fisch.

Junping He

Tianjin University

Lu Yao

Tianjin University

Juan Wang

Tianjin University

Wenyuan Gao (✉ pharmgao@tju.edu.cn)

Tianjin University <https://orcid.org/0000-0003-0050-501X>

Research Article

Keywords: Glycyrrhiza uralensis Fisch, Transcriptome, Glycyrrhizic acid, Transcription factor

Posted Date: February 16th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-231260/v1>

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

[Read Full License](#)

Abstract

In order to better understand the mechanism of glycyrrhizic acid biosynthesis and explore important enzyme gene resources in *Glycyrrhiza uralensis* Fisch., we sequenced the transcriptome of the adventitious roots of *G. uralensis* treated by methyl jasmonate (MJ) and assembled the de novo sequence. 256503 unique transcripts with an average length of 898bp were produced. Transcriptome sequencing and data analysis showed that the key genes of glycyrrhizic acid biosynthesis changed significantly after MJ treatment. 2720 up-regulated genes and 3493 down regulated genes were found. In the process of oxidation and glycosylation of glycyrrhizic acid biosynthesis. A putative CYP450 gene (Cluster-30944.70498) is positively correlated with glycyrrhetic acid. The glycosyltransferase gene (Cluster-30944.25725) is positively correlated with glycyrrhizic acid and glycyrrhetic acid. In addition, we found an AP2-EREBP family transcription factor (Cluster-30944.55070). It had high amino acid sequence similarity with PgERF1. In *Panax ginseng*, PgERF1 was identified as promoting the biosynthesis of triterpenoid saponins. According to the correlation analysis of transcription factors, functional gene expression and component accumulation, we speculated that this transcription factor can positively regulate the expression of farnesyl diphosphate, squalene epoxide and glycosyltransferase (Cluster-30944.25725) genes and ultimately increase the content of glycyrrhizic acid.

1 Introduction

As a traditional herb, *Glycyrrhiza uralensis* Fisch has a history of more than 3000 years. Its roots and stolons are widely used as natural sweeteners and drugs because they contain glycyrrhizic acid (Liu et al. 2019; Ma et al. 2018). Glycyrrhizic acid as a triterpenoid saponin, is 150 times sweeter than sucrose, and has a variety of pharmacological activities, such as anti-inflammatory, immune regulation and so on (Seki et al. 2008). Due to the increasing scarcity of glycyrrhiza plant resources, the production of glycyrrhizic acid from *G. uralensis* adventitious roots (*GCAR*) or hairy roots can solve the resource problem (Yin et al. 2019).

The formation of glycyrrhizic acid is a very complex process, in which many enzymes are involved. At present, 3-hydroxy-3-methylglutaryl-coenzyme A reductase (HMGR), squalene synthase (SS), β -amyrin synthetase (β -AS), lupanol synthase (LUS), cycloatenolase (CAS) and chalcone synthase (CHS) have been cloned from *G. uralensis* (Kim et al. 2011; Wang et al. 2013; Zhang et al. 2009; Zhang et al. 2011). In recent years, cytochrome P450 monooxygenases (CYP450s) and glycosyltransferase in downstream pathway have been gradually found (Seki et al. 2008; Seki et al. 2011; Tamura et al. 2017; Xu et al. 2016). A glucuronyltransferase, *GuUGAT* (c55437_g1) has been discovered, which has unprecedented ability to catalyze two consecutive steps of glucuronidation of glycyrrhetic acid to glycyrrhizic acid (Xu et al. 2016). Multi-site oxidation and glycosylation are the main reasons for the diversity and stability of secondary metabolites, but there are still some similar synthetic genes that have not been discovered. Moreover, the transcription of these functional genes into are regulated by a variety of regulatory factors, such as microRNA and transcription factors (Li et al. 2020a; Li et al. 2020b). The crucial element of transcription regulation is transcription factors (TFs), which bind to the promoter region of target gene or

form complexes with other DNA binding proteins to regulate gene expression (Yang et al. 2012). *GubHLH3*, a bHLH family transcription factor, was identified to regulate the expression of β -AS in *G. uralensis* (Tamura et al. 2018a). Two MYB family transcription factors GIMYB4 and GIMYB88 were found under the MJ treatment, and could regulate the accumulation of flavonoids in *G. uralensis* (Li et al. 2020b). However, the transcription factors related to glycyrrhizic acid synthesis have not been revealed.

Transcriptome research is an important part of functional gene research and an inevitable link between genome genetic information and biological function. Transcriptome sequencing is a high-throughput sequencing technology based on whole gene transcriptome developed in recent years. Compared with eukaryotic genome sequencing, transcriptome sequencing does not contain introns and other non-coding sequences. It provides new ideas and methods for functional gene mining, biosynthesis and regulation of active components of medicinal plants, evaluation and expansion of medicinal plant germplasm resources, and exploration of authentic molecular mechanism of medicinal materials. In the analysis of *Ganoderma lucidum* gene sequence, transcriptome sequencing technology was applied, and the functions of CYP450 enzyme, carbohydrate enzyme and lignin digestive enzyme were studied (Chen et al. 2012). RNA-Seq transcriptome was used to analyze the expression genes profile of genes expressed in five tissues of *Panax japonicus* and compare with other *Panax* plants to identify potential genes involved in saponin biosynthesis (Rai et al. 2016).

In this study, the content of glycyrrhizic acid in *GCAR* was significantly increased after treatment with methyl jasmonate (MJ). Using transcriptome sequencing technology to study the functional genes with high glycyrrhizic acid content from the level of differential genes, it is of great significance to reveal its regulatory mechanism. Therefore we constructed transcriptome libraries of MJ treated and control groups of *G. uralensis* to screen the differential genes. Bioinformatics methods were used for functional classification and metabolic pathway analysis to obtain functional genes. Elucidating the mechanism of glycyrrhizic acid biosynthesis and lay a foundation for further regulation of its accumulation.

2. Materials And Methods

2.1 Experimental materials

The *GCAR* were obtained by tissue culture. Its medium was 1/2 MS containing 1 mg/L indole-3-butyric acid (IBA) and 30 g/L sucrose. The culture period was 35 days. MJ (5.0 mg / L) was added at 24 h and 48 h before harvest.

2.2 Extraction and determination of active components in *G. uralensis*

The samples were crushed and dried, extracted with 80% ethanol at 65°C for 1 h, filtered, dried by rotary evaporation, and dissolved in 1 mL methanol. Then filtered through a 0.22 μ m Millipore filter. The extract was used for the quantification of glycyrrhizic acid, glycyrrhetic acid, isoliquiritin, licochalcone A, liquiritigenin, isoliquiritigenin and glabridin by HPLC (Agilent 1100, USA) using a Kromasil C18 column

(4.6 mm × 250 mm, 5 μm). When detecting glycyrrhizic acid, eluents used were as follows: A-acetonitrile (Concord, China) and B-water. Gradient elution profile was 0→8 min, A: 19%; 8→35 min, A: 19%→50%; 35→36 min, A: 50%→100%; 36→40 min, A: 100%→19%. The flow rate of the mobile phase was 1.0 ml/min. Samples (20 μL) were injected into the HPLC, and the peak was monitored at 237 nm.

When detecting glycyrrhetic acid, isoliquiritin, licochalcone A, liquiritigenin, isoliquiritigenin and glabridin, eluents used were as follows: A-acetonitrile (Concord, China) and B- 0.04% formic acid. Gradient elution profile was 0→4 min, A: 20%; 4→20 min, A: 20%→38%; 20→25 min, A: 38%→55%; 25→38 min, A: 55%→90%; 38→50 min, A: 90%. The flow rate of the mobile phase was 1.0 ml/min. Samples (20 μL) were injected into the HPLC, and the peak was monitored at 250 nm.

2.2 RNA extraction and transcriptome sequencing

RNA of 1.5 μg sample (control group, *GCAR* treated with MJ for 24 h and 48 h) was sequenced. The sequencing library was performed according to the instructions of the kit [™] RNA Library Prep Kit for Illumina® (NEB, USA). PCR products were purified by AMPure XP system and analyzed by Agilent Bioanalyzer 2100 system. The sequencing work was entrusted to Beijing Nuohe Zhiyuan Co., Ltd.

2.3 Gene function annotation

Gene function was annotated based on the following databases: Nr (NCBI non-redundant protein sequences); Nt (NCBI non redundant nucleotide sequence); Pfam (protein family); KOG (homologous protein group); Swiss-Prot (A manually annotated and reviewed protein sequence database); KO (KEGG Ortholog database); GO (Gene Ontology).

2.4 Screening of differential and functional genes

Differential expression analysis was performed using the R package DESeq (Bhandari et al. 2006). The p-values were adjusted by the Benjamini-Hochberg (BH) method. Genes with an adjusted p-value < 0.05 were assigned as DEGs. According to the results of gene function annotation and DEGs, we screened transcription factors, upstream and downstream genes related to glycyrrhizic acid biosynthesis, and further analyzed potential transcription factors and functional genes by phylogenetic analysis and correlation analysis.

2.5 qRT-PCR analysis

0.2 g *GCAR* treated with MJ and the control group were extracted with plant RNA Kit (OMEGA, USA) and HiFiscript first strand cDNA synthesis Kit (CW BIO, China). cDNA was used as a template for qRT-PCR. PCR amplification conditions: pre denaturation 94 °C 2 min, 94 °C 30 s, 35 cycles denaturation, annealing temperature 57 °C 60 s, extension 72 °C 30 s, 72 °C final extension 120 s. The qRT-PCR product was analyzed by 2% agarose gel. The gene expression was measured by $2^{-\Delta\Delta CT}$ and GAPDH was used for reference gene, and each group was repeated 3 times. Significant differences between different samples were tested with IBM SPSS Statistics 17.0 and Heml 1.0.1 (CUCKOO, Wuhan, China) software. qRT - PCR primers are shown in Table 1.

Table 1
qRT - PCR primers

Genes	Primers
GAPDH-F	5'-CCTTCATTGAGACCAAGTACGC-3'
GAPDH-R	5'-GTACTTGAGCATGTAGGCCTGTG-3'
FPS-F	5'-GGAAGCTGAACCGTGGACTGTCA-3'
FPS-R	5'-TCACCCATCATGAGCAATGCAC-3'
SS-F	5'-AATATGAGGAGAACTCCGTGAAAG-3'
SS-R	5'-GCCATGATCTGGGGAATAGC-3'
SE-F	5'-AGCAGCAGTTGACAAAGG-3'
SE-R	5'-GCCACATTCGTTTTGGTGAAGG-3'
β -AS-F	5'-TCACTTACGGTTCTTGGTTCGC-3'
β -AS-R	5'-TACTCCCGTGATTTCTGTTGG-3'
Cluster-30944.55070-F	5'-CTTACGTGAAACAAAATTCAATTCA-3'
Cluster-30944.55070-R	5'-TATCTTAAGTAATAACTGATCGCC-3'
Cluster-30944.70498-F	5'-AAGAGTGACAGATTCCAAAATATCCC-3'
Cluster-30944.70498-R	5'-CTGAATATCTTTACTATTTTCATCAAATAGTTAACC-3'
Cluster-30944.25725-F	5'-TGTTTCATCTTTCTTTCTCTAAGCTCC-3'
Cluster-30944.25725-R	5'-TTTGAAACGAAATATTAGCCTCTCC-3'

2.6 Phylogenetic analysis of genes

In order to analyze the phylogenetic relationship between the putative gene sequences and the identified genes from other plants, amino acid sequences were obtained from sequencing data and GenBank sequence data. The putative amino acid sequences of these genes were compared by ClustalW program, and phylogenetic tree was generated by neighbor joining method of MEGA 5.1 software.

2.7 Statistical analysis

SPSS 17.0 (SPSS Inc., Chicago, USA) and Hemi 1.0.1 (cuckoo, Wuhan, China) were used for statistical analysis. Pearson correlation coefficient was used to evaluate the relationship between putative gene expression and other gene expression levels, compound content. The significance of Pearson correlation coefficient at $P < 0.05$ level was evaluated (Pan et al. 2015). All samples had three independent biological repeats.

3 Result And Discussion

3.1 Effect of MJ on the content of secondary metabolites in the *GCAR*

The effect of MJ on the content of active components in *GCAR* is shown in Table 2. After treatment with MJ, the contents of glycyrrhizic acid, glycyrrhetic acid and flavonoids increased. The highest contents of glycyrrhizic acid and glycyrrhetic acid were 0.18 mg/g and 0.23 mg/g at 48 h, respectively, while the highest contents of total flavonoids were 2.42 mg/g at 24 h.

Table 2
Effect of MJ on secondary metabolites in the *GCAR*

Content (mg/g) Groups	0h	24h	48h
Isoliquiritin	0.18 ± 0.05 ^a	0.17 ± 0.05 ^a	0.21 ± 0.08 ^a
Liquiritigenin	0.26 ± 0.01 ^a	0.20 ± 0.01 ^b	0.19 ± 0.07 ^{ab}
Licochalcone A	0.07 ± 0.03 ^a	0.01 ± 0.00 ^a	0.14 ± 0.02 ^b
Isoliquiritigenin	0.02 ± 0.01 ^a	0.13 ± 0.01 ^b	0.20 ± 0.02 ^c
Glabridin	0.42 ± 0.01 ^a	1.91 ± 0.05 ^b	0.57 ± 0.09 ^c
Total flavonoid	0.96 ± 0.06 ^a	2.42 ± 0.10 ^b	1.31 ± 0.13 ^c
Glycyrrhizic acid	0.12 ± 0.03 ^a	0.16 ± 0.03 ^a	0.18 ± 0.03 ^a
Glycyrrhetic acid	0.03 ± 0.01 ^a	0.06 ± 0.01 ^a	0.23 ± 0.03 ^b
Each value represents mean ± standard error of three replicates			
^a Means followed by different letters within a column is significantly different at $p < 0.05$ according to LSD test			

3.2 Gene function annotation of *GCAR* treated by MJ

The transcriptome of *GCAR* was analyzed by Illumina RNA sequencing. After MJ treatment, 12.90 G clean bases, 13.55 G clean bases and 15.54 G clean bases were obtained in 0 h, 24 h and 48 h treatment groups, respectively. A total of 256,503 transcripts were assembled with an average length of 898 bp.

GO is an international classification system of gene function. According to the GO classification annotation of the transcriptome genes of *GCAR*, three functions were obtained: “Biological Process” (879847), “Cellular Component” (254116) and “Molecular Function” (195196). In “Biological Process”, most genes are related to cell process and metabolic process. In the category of “Cellular Component”,

most genes are related to cell and cell part. For the "Molecular Function" category, more genes are annotated in binding and catalytic activity (Fig. 1a).

KOG is a classification system based on orthologous genes, which have the same function and common ancestor. We used KOG database to divide the functional gene (31648) into 26 groups (Fig. 1b). Most of the genes (3760) were annotated in "Post-translational modification, protein turnover, chaperones", 3614 genes were annotated in "General function prediction only" group, 464 genes and 2001 genes were annotated in "Secondary metabolites biosynthesis, transport and catabolism" and "Signal transduction mechanisms", respectively.

Transcriptome sequencing is an important method to identify functional genes involved in plant secondary metabolic pathways. The quality of transcriptome of *G. uralensis* in this study was better than that reported before (152246 transcripts, average length 891) (Xu et al. 2016). GO analysis showed that *G. uralensis* genes were involved in many biological processes, especially cellular and metabolic processes. This result showed that many genes participate in the biosynthesis of different secondary metabolites, which was consistent with previous reports on *G. uralensis* (Ramilowski et al. 2013).

3.3 Screening differentially expressed genes in *GCAR* treated with MJ

A total of 5632 differentially expressed genes (DEGs) were found in three groups of experiments (*GCAR* were treated with MJ at 0 h, 24 h and 48 h), including 2720 up-regulated genes and 3493 down regulated genes (Fig. 2a, b). Among the DEGs, the most annotated genes are "catalytic activity" and "metabolic process". Among the up-regulated genes, "single organ process" and "cation binding" genes received the most annotation. However, the down regulated genes were mainly "catalytic activity", "metabolic process" and "single organism process".

Compared with 0 h, 6851 DEGs were obtained after MJ treatment for 24 h. Among the differentially expressed genes, "catalytic activity" and "metabolic process" have been most annotated, which indicates that most of them are involved in different biosynthesis processes.

Analysis based pathway provides information on biological function and secondary metabolites synthesis at the molecular level. A total of 36573 genes were annotated to 19 KEGG pathways (Fig. 3). Among them, 3367 genes were involved in the "carbohydrate metabolism" pathway, 3183 genes were involved in the "translation" pathway, 758 genes were involved in the "Metabolism of terpenoids and polyketides" pathway, and 925 genes were involved in the "Biosynthesis of other secondary metabolites" pathway. KEGG is composed of a variety of databases, which is conducive to the annotation of gene products and the distribution of biological processes. In this study, the "carbohydrate metabolism" pathway received the most annotation (3367 genes). Primary metabolism plays an important role in plant growth and development. In adventitious roots culture, primary metabolism is an important factor in plant root proliferation.

3.4 Transcription factors involved in triterpenoids biosynthesis

There are 3 pathways involved in transcription factors (Table 3), including 26 plant circadian rhythms (23 MYB family (up-regulated 3), 1 TCP family and 2 bZIP family), 118 plant hormone signal transduction pathways (83 bZIP family (up-regulated 1), 23 AP2-EREBP family (4 up-regulated), 11 bHLH family (up-regulated 1) and 33 plant pathogen interaction WRKY family (2 up-regulated).

Table 3
Potential transcription factors associated with triterpenoids biosynthesis in *G. uralensis*

Pathway	KO	KO Definition	Unigene Number	TF Family	Up-regulation
Circadian rhythm - plant	K12133	MYB-related transcription factor LHY	23	MYB	Cluster-30944.13985, Cluster-30944.13987, Cluster-30944.97522
	K16221	transcription factor TCP21	1	TCP	0
	K16241	transcription factor HY5	2	bZIP	0
Plant hormone signal transduction	K14431	transcription factor TGA	83	bZIP	Cluster-30944.47102
	K14516	ethylene-responsive transcription factor 1	23	AP2-EREBP	Cluster-30944.55070, Cluster-30944.13653, Cluster-30944.16532, Cluster-30944.76495
	K14517	Ethylene-responsive transcription factor 2	1	AP2-EREBP	0
	K13422	transcription factor MYC2	11	bHLH	Cluster-30944.78787
Plant-pathogen interaction	K18835	WRKY transcription factor 2	5	WRKY	0
	K18834	WRKY transcription factor 1	3	WRKY	0
	K13425	WRKY transcription factor 22	8	WRKY	Cluster-30944.105535
	K13424	WRKY transcription factor 33	15	WRKY	Cluster-30944.54055
	K13426	WRKY transcription factor 29	2	WRKY	0

The evolutionary relationship between the 11 up-regulated transcription factors and other triterpenoids related transcription factors is shown in Fig. 4. The potential transcription factors were located in group 2, group 3 and group 4. Cluster-30944.55070 in group 2 was close to PgERF1. DNAMAN software compared the amino acid sequence similarity of Cluster-30944.55070 and PgERF1 had 28.30% consistency.

PgERF1 can promote the synthesis of triterpenoid saponins, so Cluster-30944.55070 may also contribute to the synthesis of triterpenoid saponins. Cluster-30944.78787 in group 3 was close to NP_001288107.1. The similarity of amino acid sequences between Cluster-30944.78787 and NP_001288107.1 was 26.33% consistent. It is speculated that Cluster-30944.78787 has NP_001288107.1 similar function. Cluster-30944.105535 in group 4 is close to PgMYB5. Compared with PgMYB5, Cluster-30944.105535 had only 11.93% identity with it, indicating that the genetic relationship between Cluster-30944.105535 and PgMYB5 is far away.

qRT - PCR results showed that the expression of Cluster-30944.55070 gene in 24 h and 48 h groups was 1.51 and 1.73 times higher than that in control group, respectively (Fig. 5a). Pearson correlation analysis showed that Cluster-30944.55070 was positively correlated with isoliquiritigenin and glycyrrhetic acid, negatively correlated with liquiritigenin. The correlation coefficient between Cluster-30944.55070 and isoliquiritigenin was as high as 0.951. All these results suggest that Cluster-30944.55070 may have the ability to regulate the biosynthesis of glycyrrhetic acid, isoliquiritigenin and liquiritigenin (Fig. 5b).

Many transcription factors have been isolated and identified from plants, and they play an important role in the biosynthesis of active ingredients (Cao et al. 2018; Deng et al. 2017; Liu et al. 2017; Spyropoulou et al. 2014). In addition to glycyrrhizic acid, glycyrrhizin also belongs to triterpenoids. GubHLH3 preferentially activates the transcription of CYP93E3 and CYP72A566 promoters, which are responsible for C-22 β -hydroxylation (Tamura et al. 2018b). In this study, we found that the putative transcription factor Cluster-30944.55070 had high amino acid sequence similarity with PgERF1. In *Panax ginseng*, PgERF1 was identified as promoting the biosynthesis of triterpenoid saponins (Wu et al. 2015). The homology between Cluster-30944.55070 and PgERF1 suggests that Cluster-30944.55070 may be similar functions in the biosynthesis of triterpenoid saponins. Pearson correlation analysis further showed that the expression of Cluster-30944.55070 gene was correlated with the contents of glycyrrhetic acid, isoliquiritigenin and liquiritigenin, suggesting that Cluster-30944.55070 may have the ability to regulate the biosynthesis of glycyrrhetic acid, isoliquiritigenin and liquiritigenin.

3.5 Analysis of glycyrrhizic acid biosynthesis genes

Through transcriptome sequencing, we found that 60 genes were involved in glycyrrhizic acid biosynthesis pathway. FPS (24) and SE (16) genes were more annotated. Among the differentially expressed genes, 8 genes were up-regulated and 7 genes were down regulated (Table 4).

Table 4
Potential genes associated with triterpenoids biosynthesis in *G. uralensis*

Enzyme name	Abb.	KO	Total unigene	Upregulated	Downregulated
(-)-Germacrene D synthase		K15803	3	0	0
Farnesol dehydrogenase	FPS	K15891	4	0	0
Farnesyl-diphosphate farnesyltransferase		K00801	19	2	1
alpha-farnesene synthase		K14173	1	0	0
Squalene monooxygenase	SE	K00511	16	2	2
β -amyrin synthase	β -AS	K15813	1	1	0
β -amyrin 24-hydroxylase		K15814	1	1	0
Cytochrome P450	CYP450	K09588	4	0	0
		K09832	3	1	1
Dolichyl-diphosphooligosaccharide-protein glycosyltransferase	GT	K07151	8	1	1

When *G. glabra* was under osmotic stress, the expression levels of key genes SS and β -AS increased, which could lead to the increase of glycyrrhizic acid content (Nasrollahi et al. 2014). Previous studies have found that CYP88D6 can catalyze the oxidation of β -amyrin to 11-oxo- β -amyrin, and then CYP72A154 further converts 11-oxo- β -amyrin to glycyrrhetic acid by C-30 oxidation (Seki et al. 2008; Seki et al. 2011). *Aspergillus niger* and *Meyerozyma guilliermondii* isolates also significantly induced the up regulation of key genes such as SE, β -AS, CYP88D6 and CYP72A154 in glycyrrhizic acid biosynthesis (Li et al. 2016a). MJ, as an elicitor, has similar function to other elicitors in the biosynthesis of glycyrrhizic acid.

Transcriptome sequencing indicated that 7 CYP450s (Cluster-30944.83601, Cluster-30944.73694, Cluster-30944.70498, Cluster-30944.73707, Cluster-30944.103306, Cluster-30944.48606, Cluster-30944.61621) may participate in glycyrrhizic acid biosynthesis. 8 UGTs (Cluster-30944.48382, Cluster-30944.25725, Cluster-30944.46636, Cluster-30944.43149, Cluster-30944.65204, Cluster-30944.60730, Cluster-30944.41365, Cluster-30944.53225) may participate in glycyrrhizic acid biosynthesis.

The phylogenetic relationships between 7 putative CYP genes from *GCAR* and other plant CYP genes were analyzed (Fig. 6). Cluster-30944.70498 was highly similar to CYP88D6 (Seki et al. 2008) in *G. uralensis* (28.38%). In addition, Cluster-30944.48606, Cluster-30944.73694, Cluster-30944.73707 had 19.00%, 26.16% and 16.37% amino acid sequence similarity with CYP51, 2DFS and 3MH4, respectively. Thermogram analysis showed the expression of 7 putative CYP genes (Fig. 7a). After treated with MJ

48h, the gene expression level of Cluster-30944.70498 was 1.64 times higher than that of control group. Pearson correlation analysis was used to explore the relationship between CYP gene expression and compound content in glycyrrhizic acid biosynthesis pathway (Fig. 7b). The results showed that Cluster-30944.70498 was positively correlated with glycyrrhetic acid, licochalcone A and isoliquiritigenin, indicating that Cluster-30944.70498 may have the ability to regulate the biosynthesis of glycyrrhetic acid, licochalcone A and isoliquiritigenin.

The evolutionary relationships between 8 potential UGTs and other plant UGTs are shown in Fig. 8. The potential UGT was located in group 3 and group 4. In group 3, Cluster-30944.48382 was close to 5GMV. DNAMAN software compared the amino acid sequence similarity and found that Cluster-30944.48382 had 8.43% identity with 5GMV. Compared with the others, Cluster-30944.25725 in group 4 is closer to 6EZN. The similarity of amino acid sequences between Cluster-30944.25725 and 6EZN related is 44.76%. 6EZN belongs to oligosaccharyltransferase, so Cluster-30944.25725 also has the similar function of 6EZN. The putative expression of UGT gene in *GCAR* treated with MJ was revealed by thermograph analysis (Fig. 9a). After MJ treatment, the expression of Cluster-30944.25725 increased by 2.46 times, which was higher than that of control group. Pearson correlation analysis showed that the expression of Cluster-30944.25725 was positively correlated with glycyrrhizic acid, glycyrrhetic acid, licochalcone A and isoliquiritigenin. The correlation coefficient between Cluster-30944.25725 and glycyrrhetic acid was as high as 0.943 at the significant level of 0.01 (Fig. 9b).

3.6 Regulation mechanism of glycyrrhizic acid biosynthesis

In the biosynthesis of glycyrrhizic acid, CYP450s: CYP88D6 and CYP72A154 have been identified to oxidize β -amyrin, 11-oxo- β -amyrin to form 11-oxo- β -amyrin and glycyrrhetic acid, respectively (Seki et al. 2008; Seki et al. 2011). The up regulation of CYP88D6 and CYP72A154 in glycyrrhizic acid biosynthesis induced by *A. niger* and *M. guilliermondii* protein isolate may lead to a significant increase in glycyrrhizic acid content in *GCAR* (Li et al. 2016b). Transcriptome analysis showed that the key genes of glycyrrhizic acid biosynthesis changed significantly after MJ treatment. The putative CYP450 gene Cluster-30944.70498 was positively correlated with glycyrrhetic acid. The putative UGT gene Cluster-30944.25725 was positively correlated with glycyrrhizic acid and glycyrrhetic acid. The correlation coefficient between Cluster-30944.25725 and glycyrrhetic acid was as high as 0.943 at the significant level of 0.01. These results suggest that Cluster-30944.25725 may play a more important role in the biosynthesis of glycyrrhetic acid after MJ treatment.

The relationship between the putative transcription factor Cluster-30944.55070 and the expression of functional genes in glycyrrhizic acid biosynthesis after MJ treatment was further analyzed (Fig. 10). The Pearson correlation coefficients between Cluster-30944.55070 and FPS, SE, UGT (Cluster-30944.25725) genes were 0.983, 0.667 and 0.748, respectively, indicating that Cluster-30944.55070 was positively correlated with FPS, SE and UGT (Cluster-30944.25725). Pearson correlation analysis showed that FPS and SE were positively correlated with glycyrrhetic acid and glycyrrhizic acid. According to the relationship between transcription factor Cluster-30944.55070 and glycyrrhizic acid biosynthesis gene, it

is speculated that Cluster-30944.55070 can positively regulate the expression of FPS, SE, UGT genes, and ultimately increase the contents of glycyrrhetic acid and glycyrrhizic acid (Fig. 11). Similar results were found in Tamura's study, and the transcription factor GubHLH3 has the ability to positively regulate soybean saponin biosynthesis genes in *Glycyrrhiza* (Tamura et al. 2018b).

4 Conclusion

The transcriptome analysis of the *GCAR* before and after MJ treated showed that 11 transcription factors, 7 CYP450s and 8 glycosyltransferases are involved in glycyrrhizic acid biosynthesis. Cluster-30944.55070, a transcription factor of AP2-EREBP family, may ultimately increase glycyrrhizic acid and glycyrrhetic acid content by regulating FPS, SE and glycosyltransferases (Cluster-30944.25725) genes.

Declarations

5 Acknowledgements

National Key Research and Development Program(2020YFA0907903),The open project of state key laboratory of innovative natural medicine and TCM injections (QFSKL2020015).

6 Conflicts of interest

There are no conflicts to declare.

References

- Bhandari AP, Sukanya DH & Ramesh CR (2006) Application of Isozyme Data in Fingerprinting Napier Grass (*Pennisetum purpureum* Schum.) for Germplasm Management. *Genetic Resources and Crop Evolution* 53(2):253-264. <https://doi.org/10.1007/s10722-004-6120-2>
- Cao W, Wang Y, Shi M, Hao X, Zhao W, Wang Y, Ren J & Kai G (2018) Transcription Factor SmWRKY1 Positively Promotes the Biosynthesis of Tanshinones in *Salvia miltiorrhiza*. *Frontiers in Plant Science* 9:554. <https://doi.org/10.3389/fpls.2018.00554>
- Chen S, Xu J, Liu C, Zhu Y, Nelson DR, Zhou S, Li C, Wang L, Guo X, Sun Y, Luo H, Li Y, Song J, Henrissat B, Levasseur A, Qian J, Li J, Luo X, Shi L, He L, Xiang L, Xu X, Niu Y, Li Q, Han MV, Yan H, Zhang J, Chen H, Lv A, Wang Z, Liu M, Schwartz DC & Sun C (2012) Genome sequence of the model medicinal mushroom *Ganoderma lucidum*. *Nature Communications* 3(1):913. <https://doi.org/10.1038/ncomms1923>
- Deng B, Huang Z, Ge F, Liu D, Lu R & Chen C (2017) An AP2/ERF Family Transcription Factor PnERF1 Raised the Biosynthesis of Saponins in *Panax notoginseng*. *J. Plant Growth Regul.* 36(3):691-701. <https://doi.org/10.1007/s00344-017-9672-z>

Kim D-W, Kim RN, Choi S-H, Kim D-W, Nam S-H, Choi H-S, Koh HD, Kim A, Chae S-H, Ahn JC, Kang A & Park H-S (2011) EST Analysis Predicts Putatively Causative Genes Underlying the Pharmaceutical Application of *Glycyrrhiza uralensis* Fisch. *Plant Molecular Biology Reporter* 29(4):814-824. [10.1007/s11105-011-0290-9](https://doi.org/10.1007/s11105-011-0290-9)

Li J, Wang J, Li J, Li J, Liu S & Gao W (2016a) Protein elicitor isolated from *Escherichia coli* induced bioactive compound biosynthesis as well as gene expression in *Glycyrrhiza uralensis* Fisch adventitious roots. *RSC Advances* 6(112):111622-111631. <https://doi.org/10.1039/C6RA16903A>

Li J, Wang J, Li J, Liu D, Li H, Gao W, Li J & Liu S (2016b) *Aspergillus niger* Enhance Bioactive Compounds Biosynthesis As Well As Expression of Functional Genes in Adventitious Roots of *Glycyrrhiza uralensis* Fisch. *Appl. Biochem. Biotechnol.* 178(3):576-593. <https://doi.org/10.1007/s12010-015-1895-5>

Li Y, Chen C, Xie Z, Xu J, Wu B & Wang W (2020a) Integrated Analysis of mRNA and microRNA Elucidates the Regulation of Glycyrrhizic Acid Biosynthesis in *Glycyrrhiza uralensis* Fisch. *Int J Mol Sci* 21(9). <https://doi.org/10.3390/ijms21093101>

Li Y, Chen X, Wang J, Zou G, Wang L & Li X (2020b) Two responses to MeJA induction of R2R3-MYB transcription factors regulate flavonoid accumulation in *Glycyrrhiza uralensis* Fisch. *Plos One* 15(7). <https://doi.org/10.1371/journal.pone.0236565>

Liu J, Gao F, Ren J, Lu X, Ren G & Wang R (2017) A Novel AP2/ERF Transcription Factor CR1 Regulates the Accumulation of Vindoline and Serpentine in *Catharanthus roseus*. *Frontiers in Plant Science* 8:2082. <https://doi.org/10.3389/fpls.2017.02082>

Liu Y-Y, Yang Y-N, Feng Z-M, Jiang J-S & Zhang P-C (2019) Eight new triterpenoid saponins with antioxidant activity from the roots of *Glycyrrhiza uralensis* Fisch. *Fitoterapia* 133:186-192. <https://doi.org/10.1016/j.fitote.2019.01.014>

Ma S, Zhu G, Yu F, Zhu G, Wang D, Wang W & Hou J (2018) Effects of manganese on accumulation of Glycyrrhizic acid based on material ingredients distribution of *Glycyrrhiza uralensis*. *Industrial Crops and Products* 112:151-159. <https://doi.org/10.1016/j.indcrop.2017.09.035>

Nasrollahi V, Mirzaie-asl A, Piri K, Nazeri S & Mehrabi R (2014) The effect of drought stress on the expression of key genes involved in the biosynthesis of triterpenoid saponins in liquorice (*Glycyrrhiza glabra*). *Phytochemistry* 103:32-37. <https://doi.org/10.1016/j.phytochem.2014.03.004>

Pan Y-J, Liu J, Guo X-R, Zu Y-G & Tang Z-H (2015) Gene transcript profiles of the TIA biosynthetic pathway in response to ethylene and copper reveal their interactive role in modulating TIA biosynthesis in *Catharanthus roseus*. *Protoplasma* 252(3):813-824. <https://doi.org/10.1007/s00709-014-0718-9>

Rai A, Yamazaki M, Takahashi H, Nakamura M, Kojoma M, Suzuki H & Saito K (2016) RNA-seq Transcriptome Analysis of *Panax japonicus*, and Its Comparison with Other *Panax* Species to Identify

Potential Genes Involved in the Saponins Biosynthesis. *Frontiers in Plant Science* 7(481).

<https://doi.org/10.3389/fpls.2016.00481>

Ramilowski JA, Sawai S, Seki H, Mochida K, Yoshida T, Sakurai T, Muranaka T, Saito K & Daub CO (2013) *Glycyrrhiza uralensis* Transcriptome Landscape and Study of Phytochemicals. *Plant and Cell Physiology* 54(5):697-710. <https://doi.org/10.1093/pcp/pct057>

Seki H, Ohyama K, Sawai S, Mizutani M, Ohnishi T, Sudo H, Akashi T, Aoki T, Saito K & Muranaka T (2008) Licorice β -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* 105(37):14204. <https://doi.org/10.1073/pnas.0803876105>

Seki H, Sawai S, Ohyama K, Mizutani M, Ohnishi T, Sudo H, Fukushima EO, Akashi T, Aoki T, Saito K & Muranaka T (2011) Triterpene Functional Genomics in Licorice for Identification of CYP72A154 Involved in the Biosynthesis of Glycyrrhizin. *The Plant Cell* 23(11):4112. <https://doi.org/10.1105/tpc.110.082685>

Spyropoulou EA, Haring MA & Schuurink RC (2014) RNA sequencing on *Solanum lycopersicum* trichomes identifies transcription factors that activate terpene synthase promoters. *BMC Genomics* 15(1):402. <https://doi.org/10.1186/1471-2164-15-402>

Tamura K, Seki H, Suzuki H, Kojoma M, Saito K & Muranaka T (2017) CYP716A179 functions as a triterpene C-28 oxidase in tissue-cultured stolons of *Glycyrrhiza uralensis*. *Plant Cell Rep* 36(3):437-445. <https://doi.org/10.1007/s00299-016-2092-x>

Tamura K, Yoshida K, Hiraoka Y, Sakaguchi D, Chikugo A, Mochida K, Kojoma M, Mitsuda N, Saito K, Muranaka T & Seki H (2018a) The Basic Helix-Loop-Helix Transcription Factor GubHLH3 Positively Regulates Soyasaponin Biosynthetic Genes in *Glycyrrhiza uralensis*. *Plant Cell Physiol* 59(4):778-791. <https://doi.org/10.1093/pcp/pcy046>

Tamura K, Yoshida K, Hiraoka Y, Sakaguchi D, Chikugo A, Mochida K, Kojoma M, Mitsuda N, Saito K, Muranaka T & Seki H (2018b) The Basic Helix-Loop-Helix Transcription Factor GubHLH3 Positively Regulates Soyasaponin Biosynthetic Genes in *Glycyrrhiza uralensis*. *Plant and Cell Physiology* 59(4):783-796. <https://doi.org/10.1093/pcp/pcy046>

Wang D, Pang Y-X, Wang W-Q, Wan C-Y, Hou J-L, Yu F-L, Wang Q-L, Liu F-B & Zhang X-D (2013) Effect of molybdenum on secondary metabolic process of glycyrrhizic acid in *Glycyrrhiza uralensis* Fisch. *Biochem. Syst. Ecol.* 50:93-100. <https://doi.org/10.1016/j.bse.2013.03.045>

Wu B, Long Q, Gao Y, Wang Z, Shao T, Liu Y, Li Y & Ding W (2015) Comprehensive characterization of a time-course transcriptional response induced by autotoxins in *Panax ginseng* using RNA-Seq. *BMC Genomics* 16(1):1010. <https://doi.org/10.1186/s12864-015-2151-7>

- Xu G, Cai W, Gao W & Liu C (2016) A novel glucuronosyltransferase has an unprecedented ability to catalyse continuous two-step glucuronosylation of glycyrrhetic acid to yield glycyrrhizin. *New Phytologist* 212(1):123-135. <https://doi.org/10.1111/nph.14039>
- Yang CQ, Fang X, Wu XM, Mao YB, Wang LJ & Chen XY (2012) Transcriptional Regulation of Plant Secondary Metabolism. *Journal of Integrative Plant Biology* 54(10):703-712. <https://doi.org/10.1111/j.1744-7909.2012.01161.x>
- Yin Y-c, Zhang X-d, Gao Z-q, Hu T, Yang L, Zhang Z-x, Li W-d & Liu Y (2019) Over-expressing root-specific β -amyrin synthase gene increases glycyrrhizic acid content in hairy roots of *glycyrrhiza uralensis*. *Chinese Herbal Medicines* 11(2):192-199. <https://doi.org/10.1016/j.chmed.2019.03.001>
- Zhang H-C, Liu J-M, Lu H-Y & Gao S-L (2009) Enhanced flavonoid production in hairy root cultures of *Glycyrrhiza uralensis* Fisch by combining the over-expression of chalcone isomerase gene with the elicitation treatment. *Plant Cell Reports* 28(8):1205-1213. <https://doi.org/10.1007/s00299-009-0721-3>
- Zhang HC, Liu JM, Chen HM, Gao CC, Lu HY, Zhou H, Li Y & Gao SL (2011) Up-Regulation of Licochalcone A Biosynthesis and Secretion by Tween 80 in Hairy Root Cultures of *Glycyrrhiza uralensis* Fisch. *Mol. Biotechnol.* 47(1):50-56. <https://doi.org/10.1007/s12033-010-9311-4>

Figures

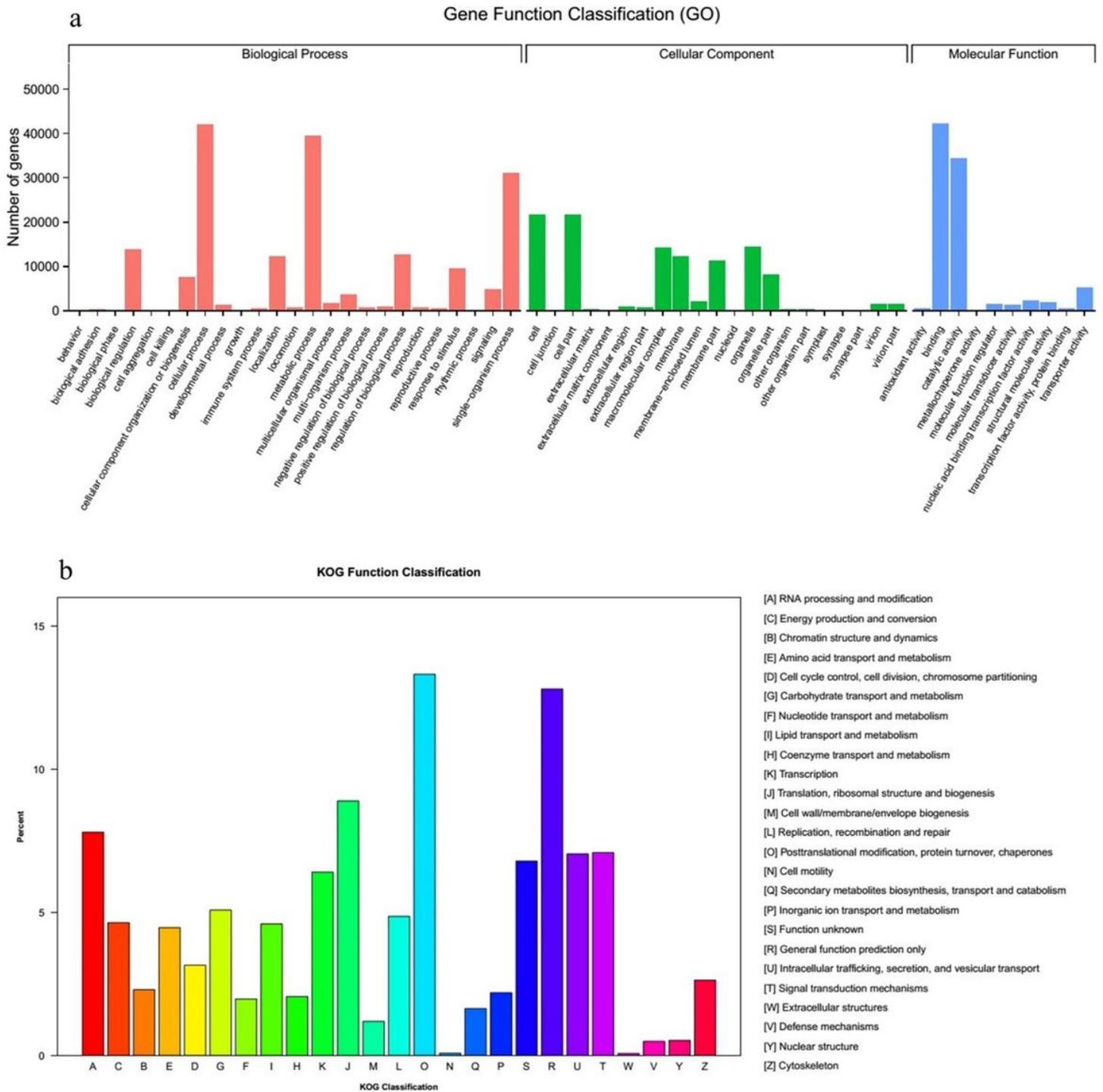


Figure 1

Histogram of gene ontology (a) and unigene KOG (b) classification.

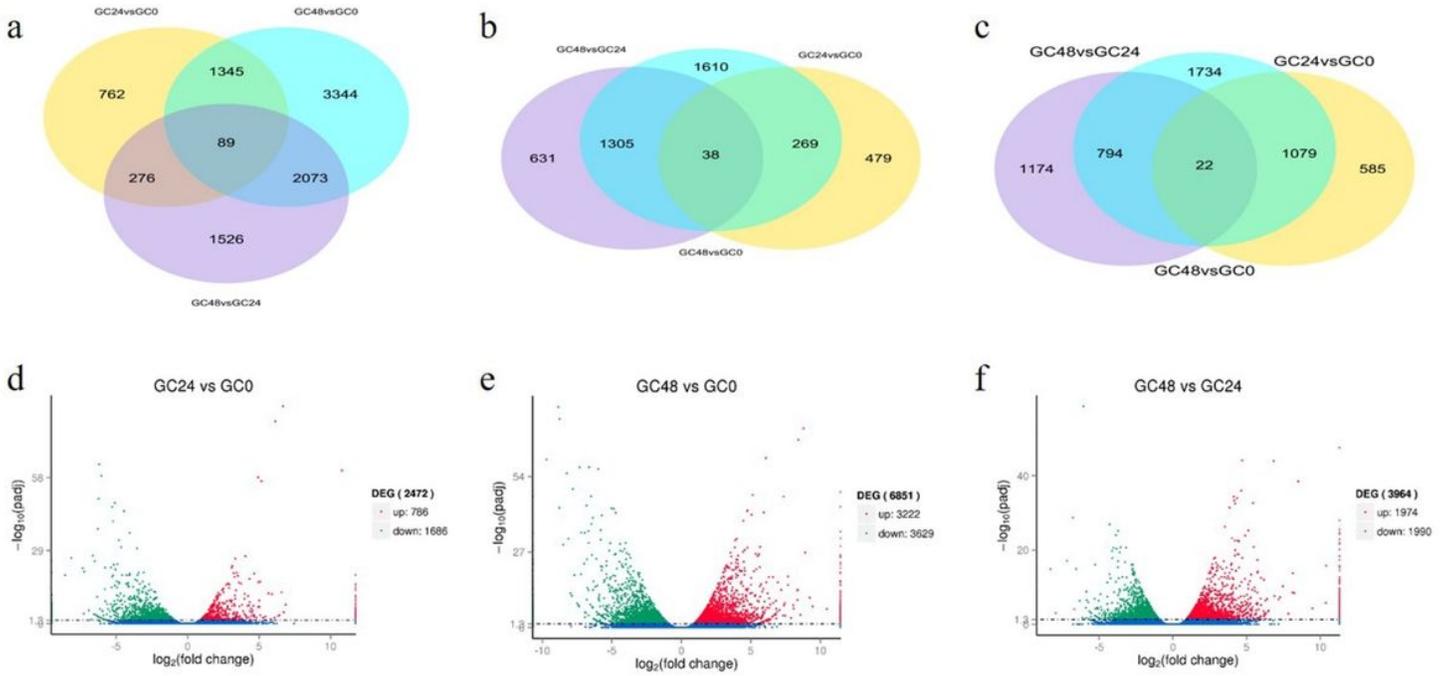


Figure 2

Differentially expressed genes (DEGs) were found in three groups of experiments (GCAR were treated with MJ at 0 h, 24 h and 48 h). a, b, c - Venn diagram at 0 h, 24 h and 48 h. a, differentially expressed genes; b up-regulated gene; c, down-regulated gene. d, e, f- Expression level of DEGs. d, 24h vs 0h; e, 48h vs 0h; f, 48h vs 24h.

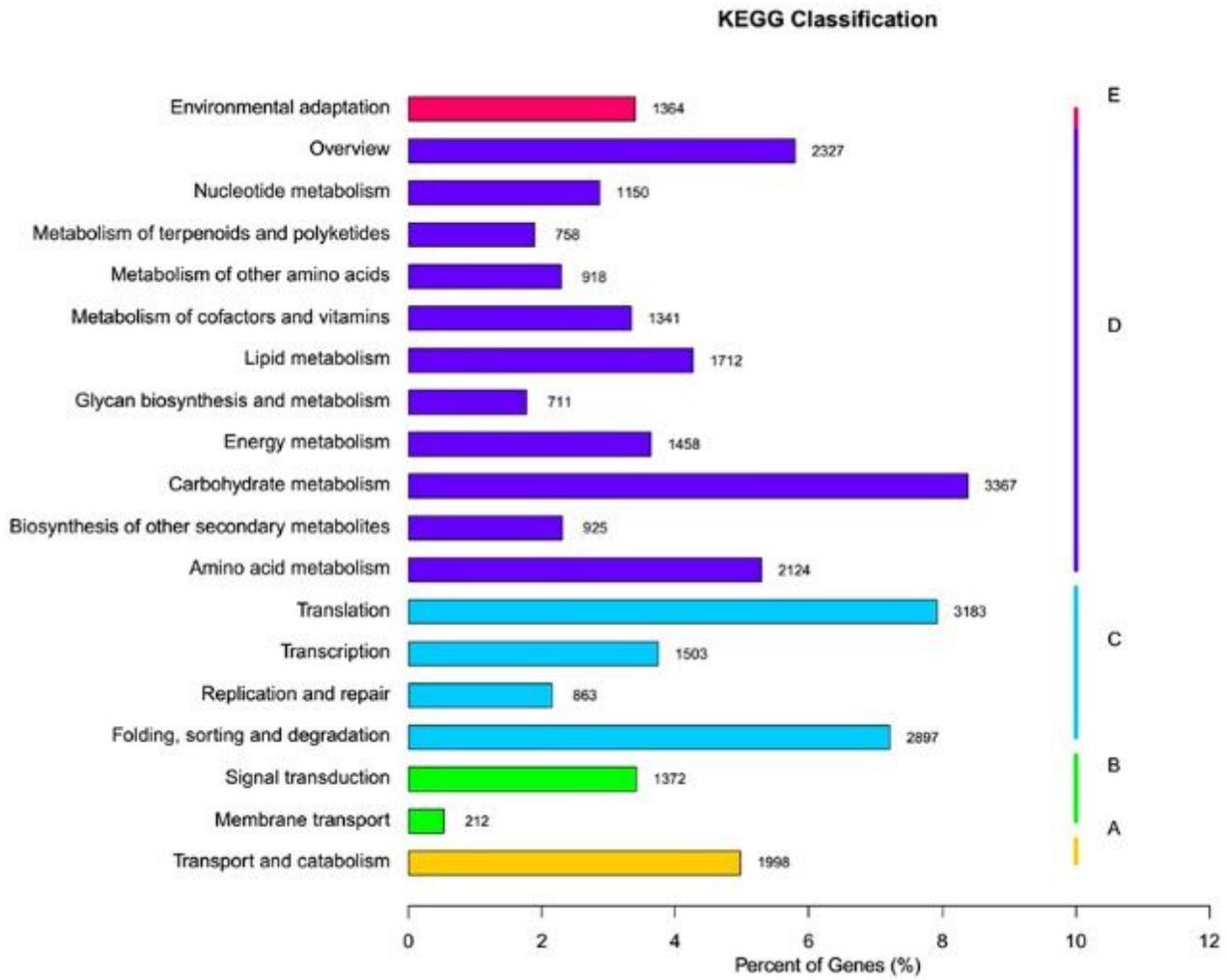


Figure 3

Functional classification and pathway assignment of unigenes by KEGG.

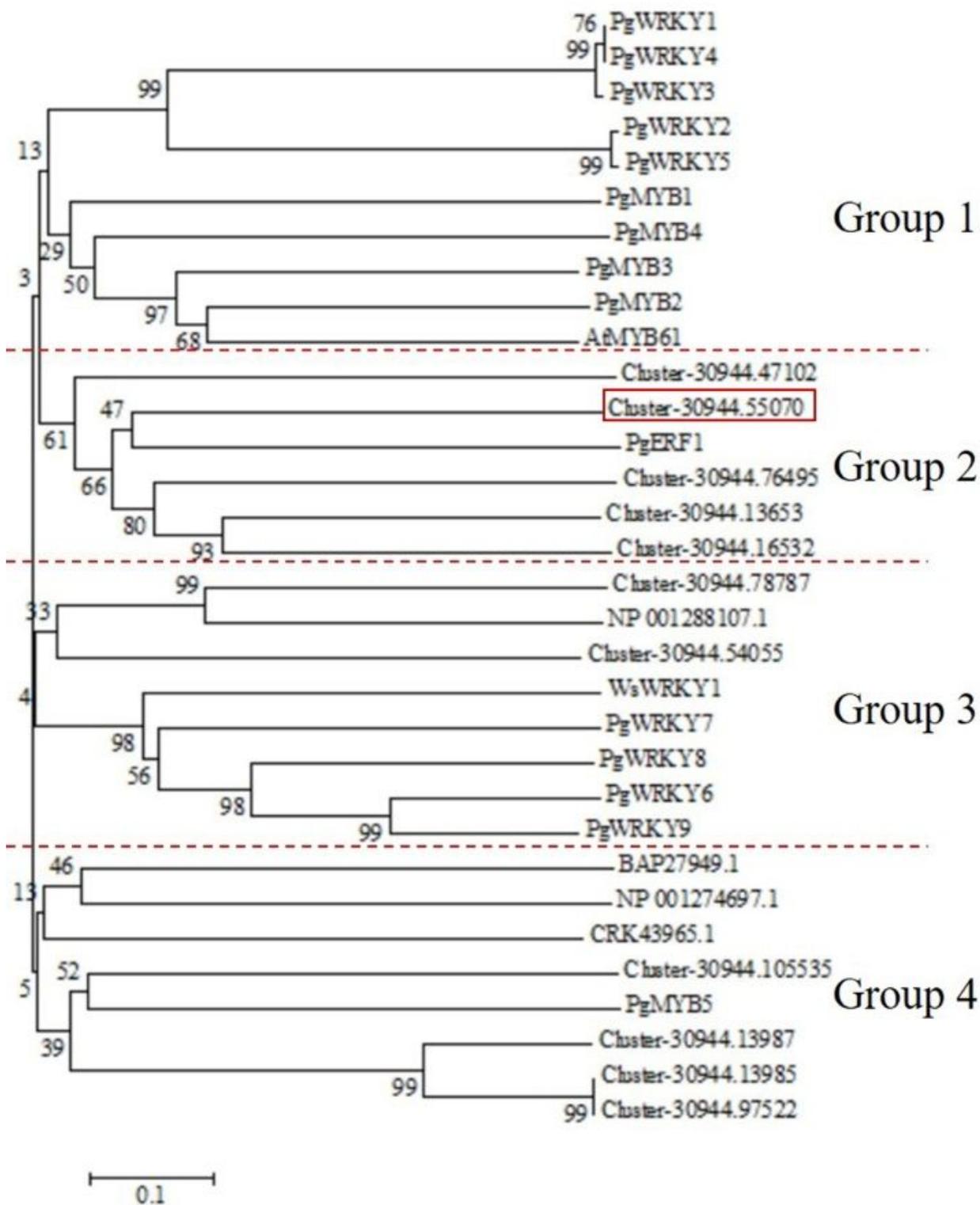


Figure 4

Phylogenetic analysis based on the deduced amino acid sequences of 11 up-regulated putative transcription factors and triterpenoids related TFs from other plants. Bar = 0.2 amino acid substitutions per site. Red rectangle, putative TFs.

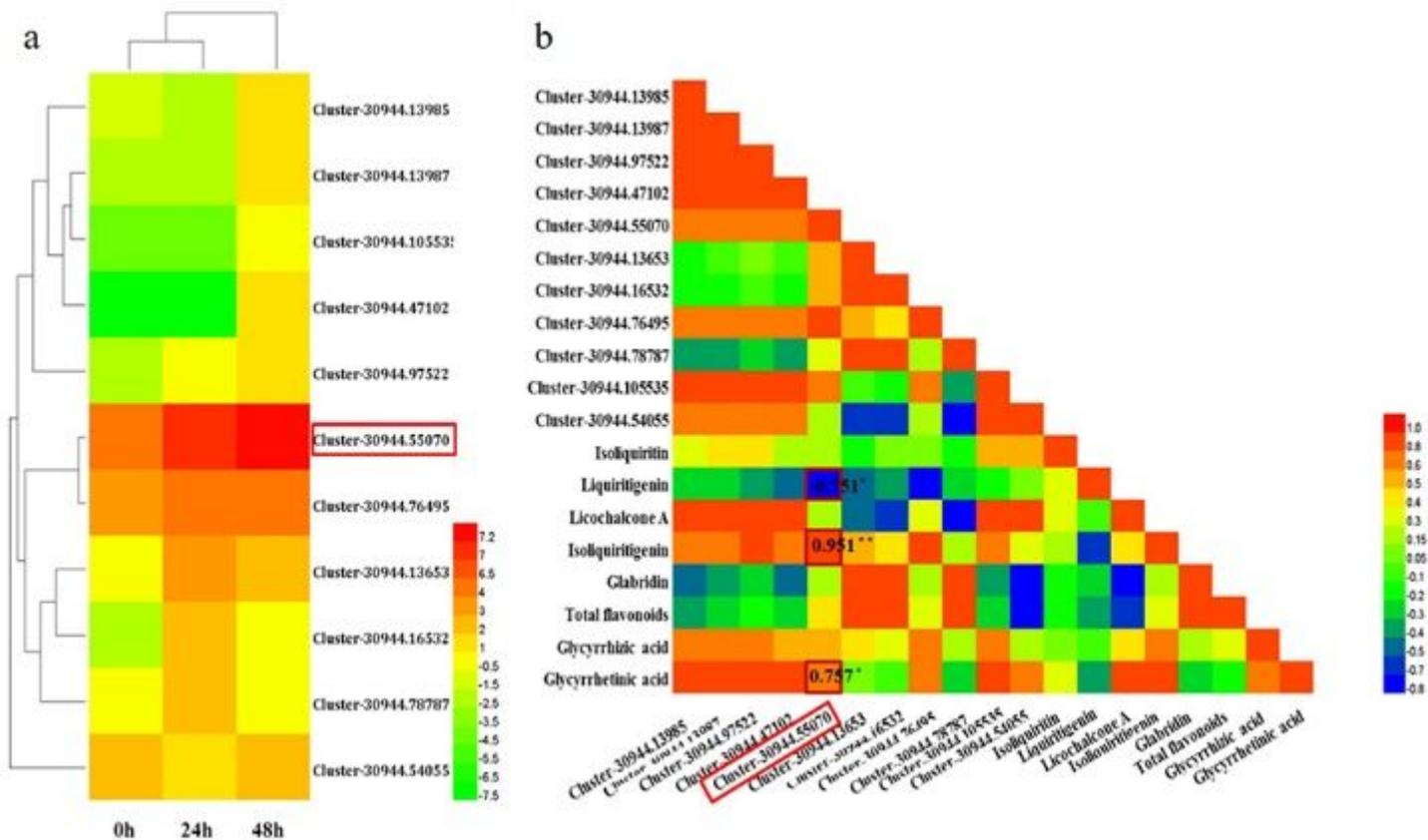


Figure 5

The expression of 11 up-regulated putative TFs at different times (a), and Pearson correlation analysis of putative TFs and contents of active ingredients (b). (a) Heatmap analysis of the expression of 11 up-regulated putative TFs at 0h, 24h, 48h after MJ treatment. (b) Pearson correlation analysis of the expression of 11 up-regulated putative TFs and contents of active ingredients after MJ treatment. The correlation coefficient between Cluster-30944.55070 and liquiritigenin isoliquiritigenin glycyrrhetic acid, was 0.751, 0.951, 0.757, respectively. **Correlation is significant at the 0.01 level. * Correlation is significant at the 0.05 level.

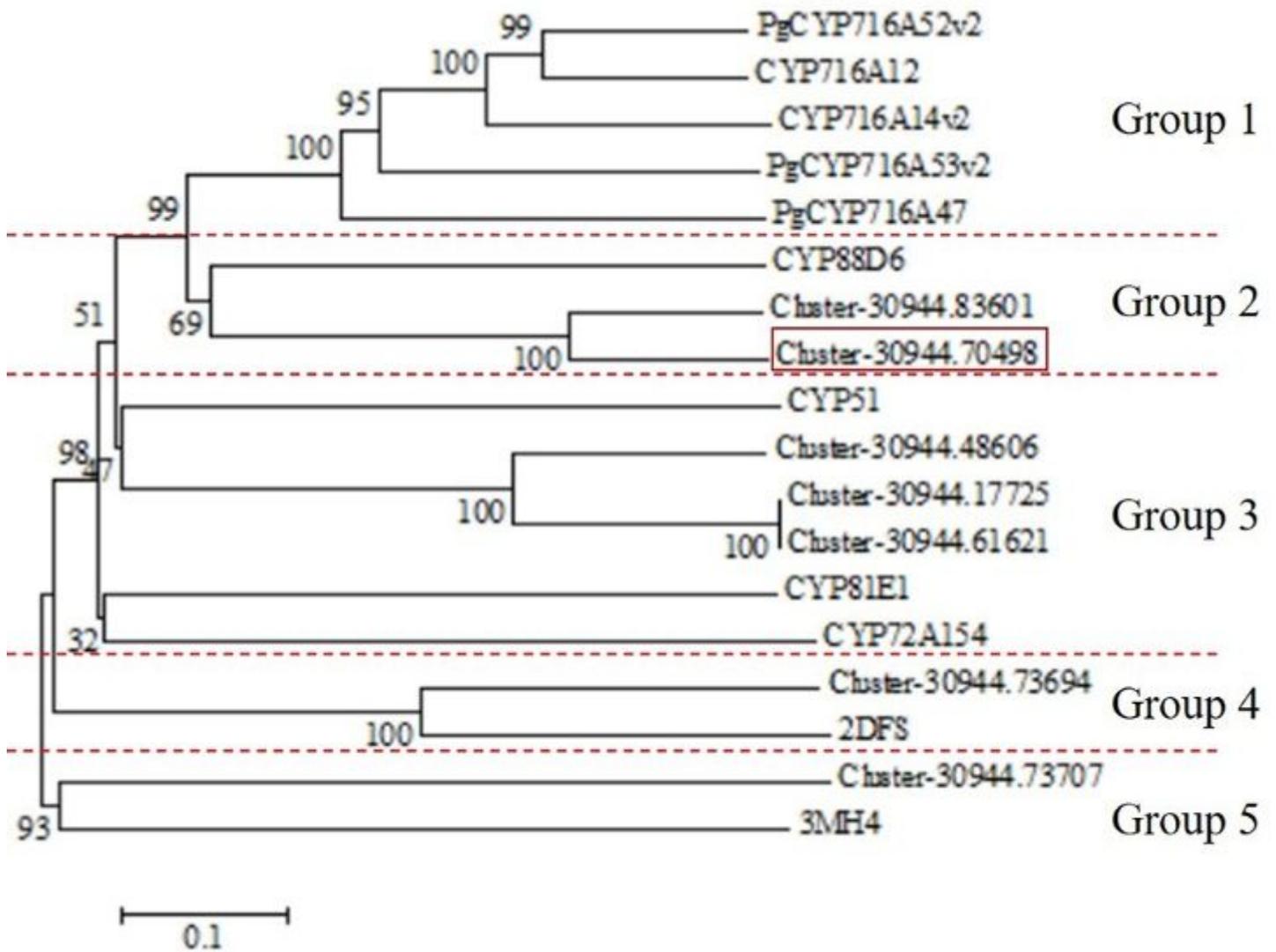


Figure 6

Phylogenetic analysis based on the deduced amino acid sequences of 7 putative CYP450s and triterpenoids related CYP450s from other plants. Bar = 0.2 amino acid substitutions per site.

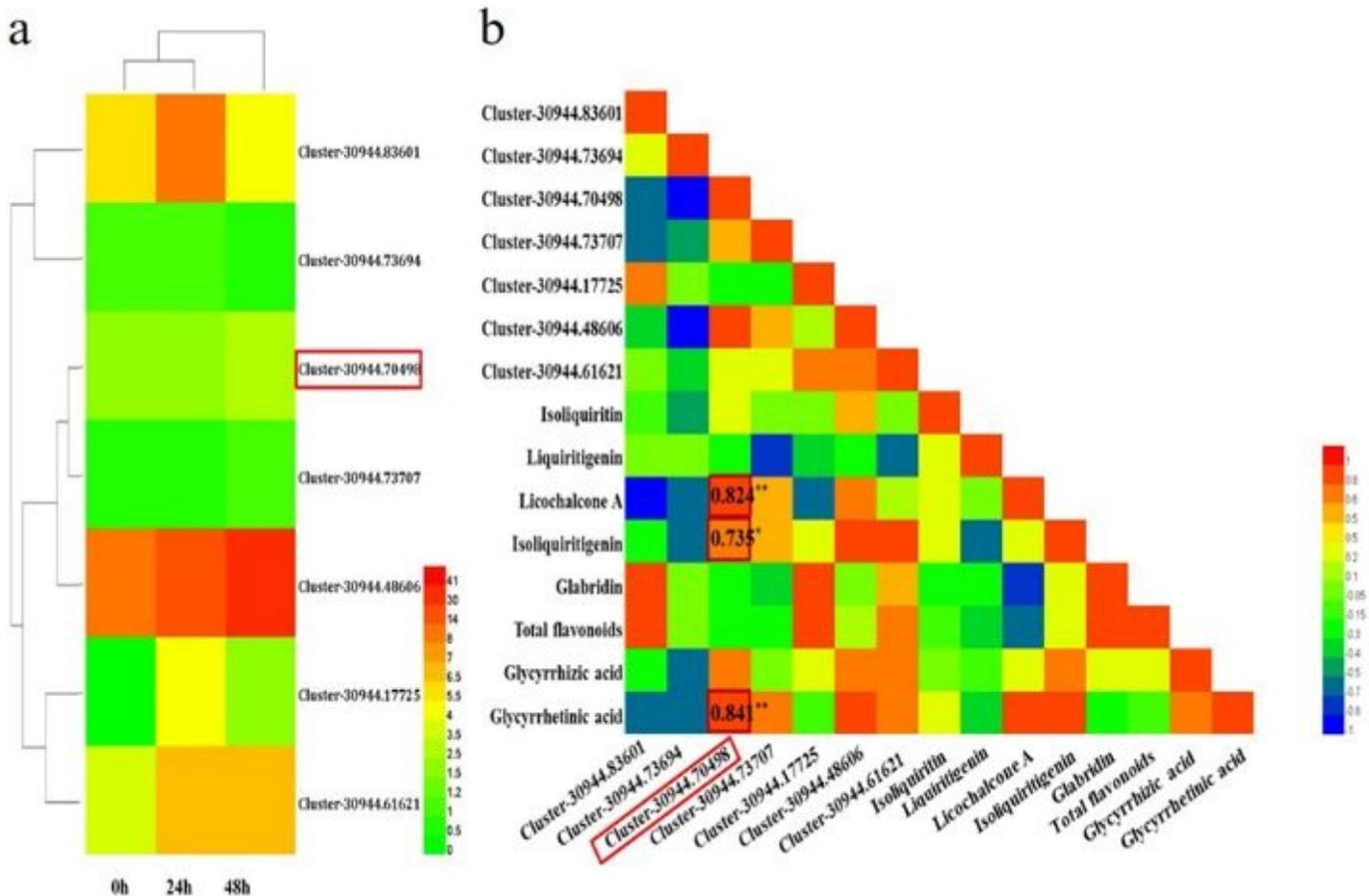


Figure 7

The expression of 7 putative CYP450s at different times (a), and Pearson correlation analysis of putative CYP450s and contents of active ingredients (b). (a) Heatmap analysis of the expression of putative CYP450s and contents of active ingredients (b). (a) Heatmap analysis of the expression of putative CYP450s at 0h, 24h, 48h after MJ treatment. (b) Pearson correlation analysis of the expression of 7 putative CYP450s and contents of active ingredients after MJ treatment. The correlation coefficient between Cluster-30944.70498 and licochalcone A, isoliquiritigenin, glycyrrhetic acid, was 0.824, 0.735, 0.841, respectively. ** Correlation is significant at the 0.01 level. * Correlation is significant at the 0.05 level.

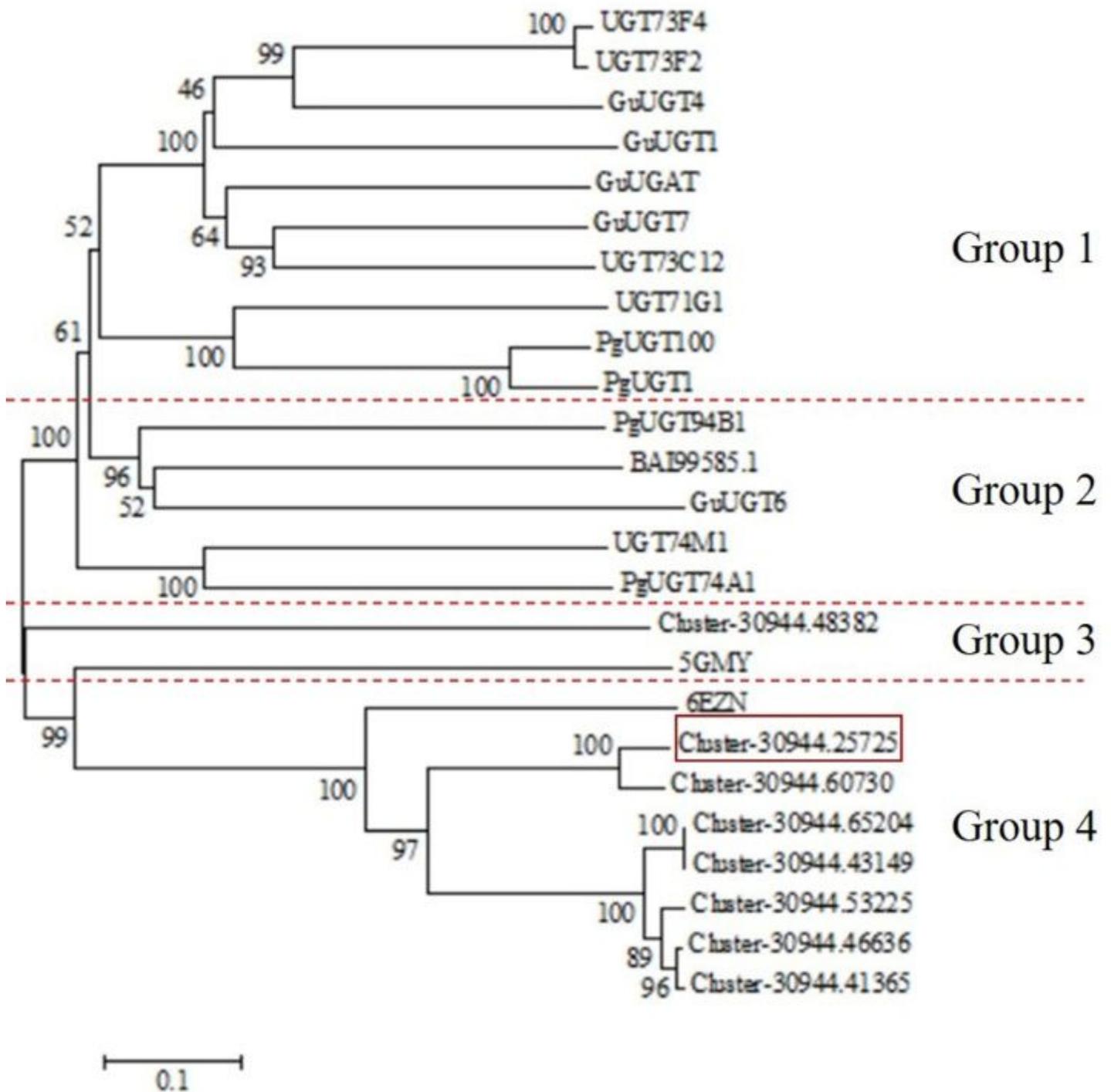


Figure 8

Phylogenetic analysis based on the deduced amino acid sequences of 8 putative UGTs and triterpenoids related UGTs from other plants. Bar = 0.2 amino acid substitutions per site.

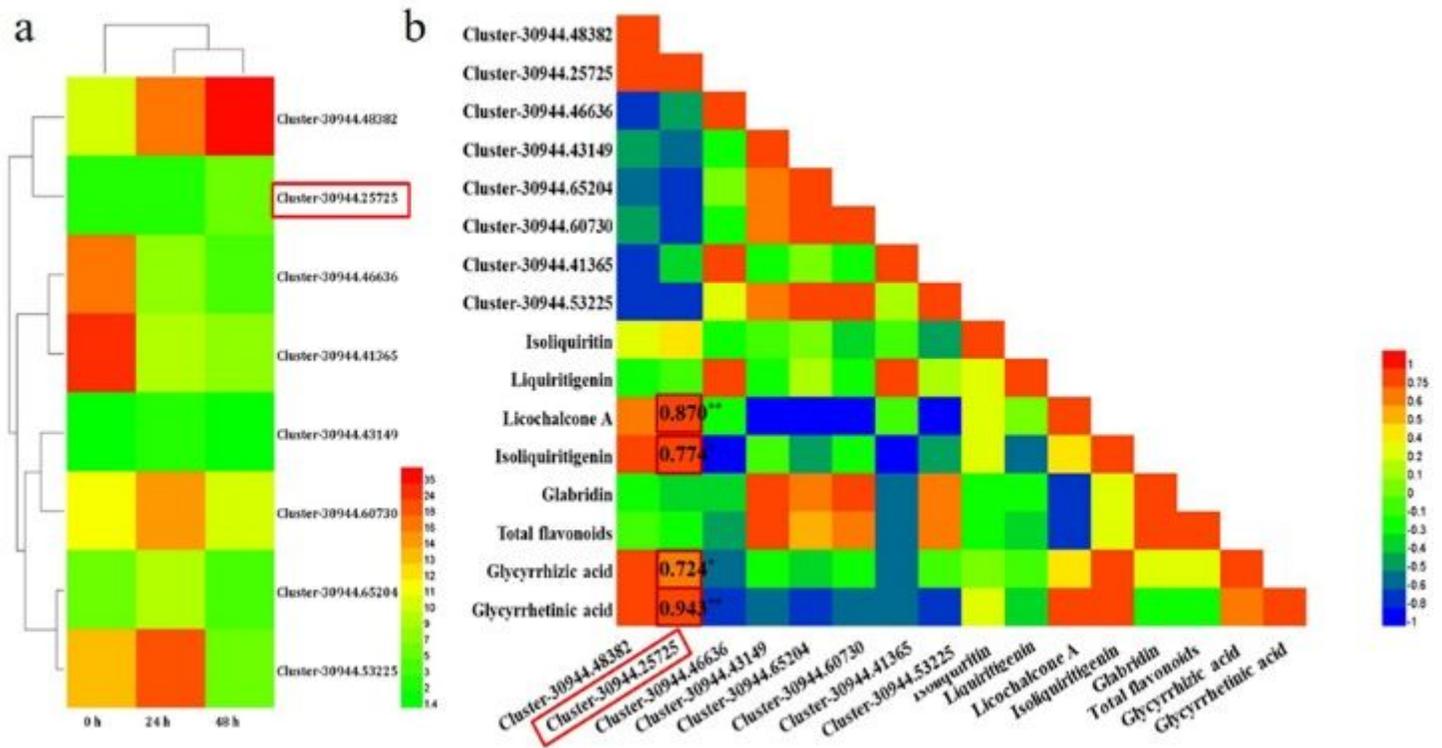


Figure 9

The expression of 8 putative UGTs at different times (a), and Pearson correlation analysis of putative UGTs and contents of active ingredients (b). (a) Heatmap analysis of the expression of 8 putative UGTs at 0h, 24h, 48h after MJ treatment. (b) Pearson correlation analysis of the expression of 8 putative UGTs and contents of active ingredients after MJ treatment. The correlation coefficient between Cluster-30944.25725 and licochalcone A, isoliquiritigenin, glycyrrhizic acid, glycyrrhetic acid, was 0.870, 0.774, 0.724, 0.943, respectively. ** Correlation is significant at the 0.01 level. * Correlation is significant at the 0.05 level.

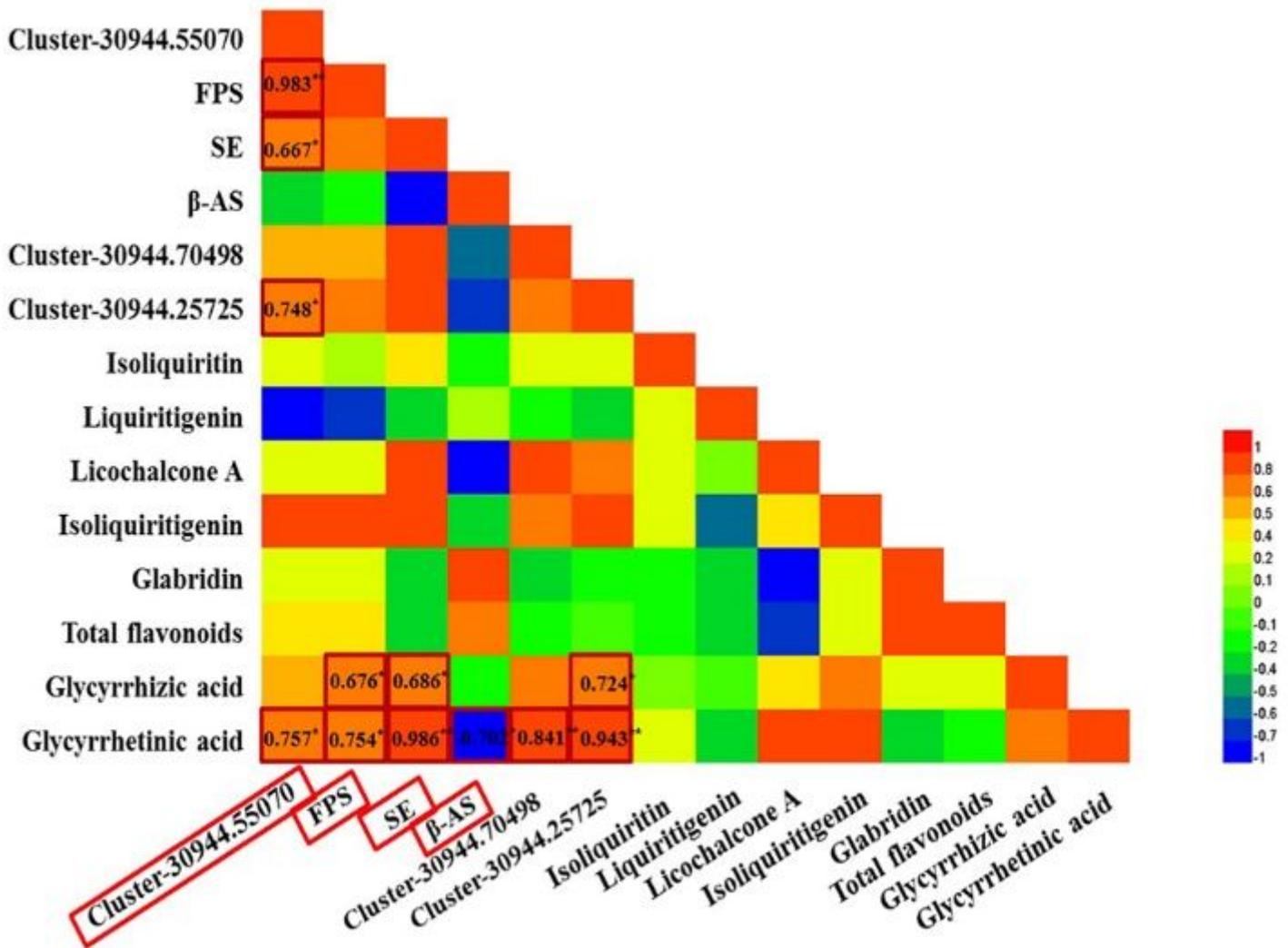


Figure 10

Pearson correlation analysis of the expression of putative TF, Cluster-30944.55070, functional genes in the biosynthesis of glycyrrhizic acid, and contents of active ingredients after MJ treatment. Cluster-30944.55070, FPS, SE, β -AS, Cluster-30944.70498, Cluster-30944.25725 showed correlation with glycyrrhetic acid. FPS, SE, Cluster-30944.25725 showed correlation with glycyrrhizic acid. Cluster-30944.55070 also showed correlation with FPS, SE, Cluster-30944.25725. ** Correlation is significant at the 0.01 level. * Correlation is significant at the 0.05 level.

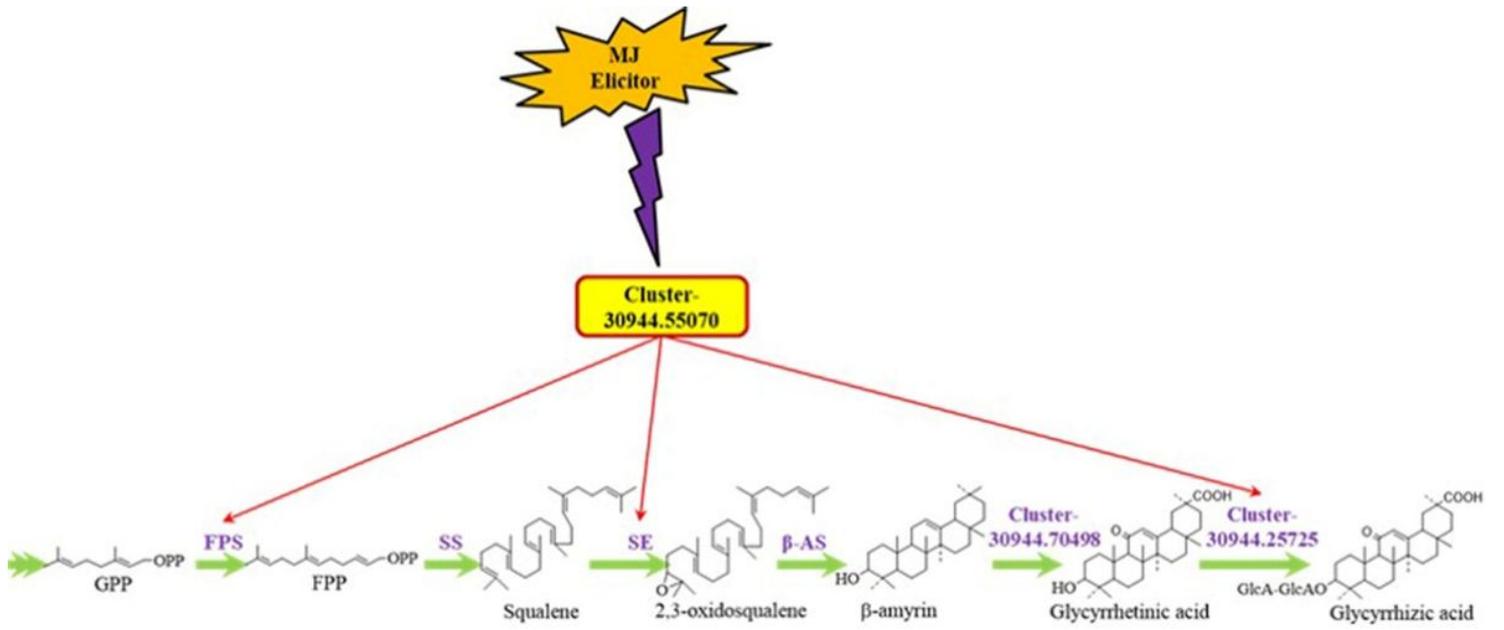


Figure 11

Putative MJ induction mechanism with TF function on the biosynthesis of glycyrrhizic acid.