



## Research article

# Combining metabolomics and transcriptomics to reveal the potential medicinal value of rare species *Glycyrrhiza squamulose*

Bin Ma<sup>a,b,1</sup>, Siru Wang<sup>c,1</sup>, Haonan Li<sup>c</sup>, Qinyue Wang<sup>c</sup>, Yaqi Hong<sup>c</sup>, Yang-mei Bao<sup>d</sup>, Hua Liu<sup>a</sup>, Ming Li<sup>a</sup>, Yucheng Zhao<sup>c,e,\*</sup>, Lan-ping Guo<sup>b,\*\*</sup>

<sup>a</sup> Institute of Forestry and Grassland Ecology, Ningxia Academy of Agricultural and Forestry Sciences, Yinchuan, 75000, China

<sup>b</sup> State Key Laboratory for Quality Ensurance and Sustainable Use of Dao-di Herbs, National Resource Center for Chinese Materia Medica, China Academy of Chinese Medical Sciences, Beijing, 100700, China

<sup>c</sup> Department of Resources Science of Traditional Chinese Medicines, School of Traditional Chinese Pharmacy, China Pharmaceutical University, Nanjing, 211198, China

<sup>d</sup> School of Chinese Materia Medica, Beijing University of Chinese Medicine, Beijing, China

<sup>e</sup> Medical Botanical Garden, China Pharmaceutical University, Nanjing, 211198, China

## ARTICLE INFO

## Keywords:

*Glycyrrhiza squamulose*  
Isoflavonoid biosynthesis  
Flavonoid biosynthesis  
Metabolomics analysis  
Transcriptomics analysis

## ABSTRACT

Licorice is a well-known Chinese medicinal plant that is widely used to treat multiple diseases and process food; however, wild licorice is now facing depletion. Therefore, there is an urgent need to identify and protect licorice germplasm diversity. In this study, metabolomic and transcriptomic analyses were conducted to investigate the biodiversity and potential medicinal value of the rare wild *Glycyrrhiza squamulose*. A total of 182 differentially accumulated metabolites and 395 differentially expressed genes were identified by comparing *Glycyrrhiza uralensis* and *Glycyrrhiza squamulose*. The molecular weights of the chemical component of *G. squamulose* were comparable with those of *G. uralensis*, suggesting that *G. squamulose* may have medicinal value. Differentially accumulated metabolites (DAMs), mainly flavonoids such as kaempferol-3-O-galactoside, kaempferol-3-O-(6"malonyl) glucoside, and hispidulin-7-O-glucoside, showed potential vitality in *G. squamulose*. Comparative transcriptomics with *G. uralensis* showed that among the 395 differentially expressed genes (DEGs), 69 were enriched in the isoflavonoid biosynthesis pathway. Multiomics analysis showed that the distinction in flavonoid biosynthesis between *G. squamulose* and *G. uralensis* was strongly associated with the expression levels of IF7GT and CYP93C. In addition to identifying similarities and differences between *G. squamulose* and *G. uralensis*, this study provides a theoretical basis to protect and investigate rare species such as *G. squamulose*.

## 1. Introduction

Licorice (Gancao in Chinese) is the dried root and rhizome of the leguminous plants, *Glycyrrhiza uralensis* Fisch, *Glycyrrhiza inflata* Bat, and *Glycyrrhiza glabra* L [1]. It is bestowed with the title of "the senior statesman of the traditional Chinese medicine (TCM)

\* Corresponding author. Medical Botanical Garden, China Pharmaceutical University, Nanjing 211198, China, Department of Resources Science of Traditional Chinese Medicines, School of Traditional Chinese Pharmacy, China Pharmaceutical University, Nanjing 211198, China.

\*\* Corresponding author.

E-mail addresses: [zhaoyucheng@126.com](mailto:zhaoyucheng@126.com) (Y. Zhao), [glp01@126.com](mailto:glp01@126.com) (L.-p. Guo).

<sup>1</sup> These authors contributed equally to this study.

<https://doi.org/10.1016/j.heliyon.2024.e30868>

Received 7 March 2024; Received in revised form 1 May 2024; Accepted 7 May 2024

Available online 10 May 2024

2405-8440/© 2024 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

kingdom [2].” *Glycyrrhiza Linn* is distributed in temperate and subtropical zones and contains 30 species, eight of which are found in China [3]. Licorice expels toxic substances, moistens the lungs to relieve coughing, invigorates the spleen and enhances the harmony of herbal medicines in a single prescription to achieve better therapeutic effects [4,5]. These properties make it one of the most frequently used herbs in TCM [1]. In addition to TCM, licorice is used worldwide [6]. For example, *G. uralensis* was included in the 14th edition of the Russian Pharmacopoeia. Additionally, the European Medicines Agency released a community herbal monograph on *G. glabra*, *G. inflata*, and *G. uralensis* in 2012, which stated that licorice could be used as a traditional herbal medicine to relieve digestive symptoms, including burning sensations, dyspepsia, and coughing associated with colds ([www.ema.europa.eu](http://www.ema.europa.eu)). In recent years, constituent analysis of licorice has led to the isolation of several useful pharmacological compounds that show promising therapeutic potential, owing to their anti-inflammatory, antiviral, antimicrobial, antioxidative, anticancer, immunomodulatory, hepatoprotective, and cardioprotective effect [7]. Recent research employing metabolic and pharmacological approaches revealed that the aerial and underground parts of *G. uralensis* could potentially be utilized as valuable medicinal resources [8]. However, owing to overexploitation and high demand, wild licorice resources are on the verge of depletion, and numerous efforts have been made to protect these resources and explore alternative avenues, including artificial cultivation, utilization of aerial parts, and the development of new licorice varieties. Despite the maturity of artificially cultivated licorice, the shortage of germplasm resources remains a limitation [9].

*G. squamulose* is one of the rarest *Glycyrrhiza Linn* species and has been poorly studied because it is rarely collected from the wild. Between 2013 and 2018, *G. squamulose* specimens were not collected from the Fourth Census of Chinese Medicinal Resources. Compared to *G. uralensis*, the corolla of *G. uralensis* is purple, white, and yellow, and that of *G. squamulose* is white. The leaves also have different colors and shapes (Fig. S1). Due to the scarcity of *G. squamulose*, only a few studies on its chemical composition have been published. Additionally, the lack of omics analysis makes public awareness of *G. squamulose* insufficient, limiting its conservation and usage [10–12]. Hence, protecting and cultivating *G. squamulose* is critical for preserving the diversity of licorice germplasms.

Few studies have focused on *G. squamulose*, a member of the same family as *Glycyrrhiza Linn*, which suggests that it may have a chemical composition similar to that of *G. uralensis*. Flavonoids were among the major constituents, with licorice having the highest content [13]. Although flavonoids are important pharmacological components of licorice, total flavonoid content varies among licorice breeds [14]. According to studies, *G. inflata* had the greatest total flavonoid content, whereas *G. uralensis* had the lowest among the three species defined as licorice in TCM [15]. Although not listed in the TCM resources, *G. korshinskyi* contains a greater percentage of total flavonoids than *G. uralensis* and has demonstrated certain medical value. Based on chemical composition analysis and classical classification, *G. inflata* and *G. korshinskyi* have the second-closest relative relationship with *G. squamulose* [16]. Therefore, studying *G. squamulose* is appealing because of the lack of knowledge about its constituents, their proportion, and potential utility. Therefore, it is imperative to understand the potential similarities and variations in flavonoid content and biosynthetic pathways across different licorice species. This knowledge will facilitate a comprehensive evaluation of diverse licorice species.

Flavonoid biosynthetic pathways consist of a large network that includes eight branches and four important intermediate metabolites. The isoflavone biosynthesis pathway, which is primarily distributed in leguminous plants, is unique among the eight major branches [17]. In this pathway, biosynthesis begins with flavanone production, which is catalyzed by isoflavone synthase (IFS) [18]. Hydroxyisoflavanone dehydratase (HID) converts the flavanones to isoflavones [19], genistein, and daidzein. Many other products, such as maackiain and pterocarpan [20], may be produced via reactions involving pterocarpan synthase (PTS). Biosynthetic pathways of flavonoids and their metabolites have been studied for several decades [21]. Various genetic, morphogenetic, and environmental factors affect plant physiology [22].

However, the principal medicinal tissues of *G. squamulose* and the mechanisms underlying flavonoid accumulation and regulation in these organs remain unclear. Plant metabolites may provide insight into the complex backgrounds of poorly researched species [23, 24]. Transcriptomics can also be utilized to analyze variations in gene expression levels and metabolic processes, which can serve as a foundation for the regulation of secondary metabolism in plants [25]. By combining metabolomic and transcriptomic analyses, a deeper understanding of the characteristics of both the metabolites and transcripts can be obtained. We selected *G. uralensis*, a traditional licorice in the Pharmacopoeia of the People’s Republic of China, as a reference substance, and *G. squamulose* for metabolomic and transcriptomic analyses of their roots, stems, and leaves. These analyses identified differences in flavonoid accumulation among the three parts of both species as well as differentially expressed genes (DEGs) involved in flavonoid biosynthesis. Additionally, Kyoto Encyclopedia of Genes and Genomes (KEGG) analysis was used to elucidate the regulatory relationships between the genes and metabolic pathways involved in flavonoid synthesis.

Overall, the similarity and specificity of flavonoid content and the identified genes involved in the biosynthesis pathways of the three organs of *G. squamulose* and *G. uralensis* were investigated in this study. These findings have significant implications for the comprehensive evaluation and utilization of *G. squamulose* resources, and provide theoretical support for the conservation of this precious plant species to maintain biodiversity.

## 2. Materials and methods

### 2.1. Plant materials and sample collection

For this study, *G. squamulose* and *G. uralensis* were cultivated in Shizuishan, Ningxia, China. The leaf, stem, and root tissues of the two licorice species were separated into six groups, each comprising three biological replicates. For the metabolomic analysis, the leaves, roots, and stems samples of *G. uralensis* were designated as W-M-L1-3, W-M-R1-3, and W-M-S1-3, and those of *G. squamulose* were designated as Y-M-L1-3, Y-M-R1-3, and Y-M-S1-3, respectively. The same samples were used for transcriptome analyses. All samples were preserved at the Medical Botanical Garden of China Pharmaceutical University (sample names: GU2021W-M-L1-3,

GU2021W-M-R1-3, GUW-M-S1-3, GS2021Y-M-L1-3, GS2021Y-M-R1-3, and GS2021Y-M-S1-3). After collection, samples from *G. squamulose* and *G. uralensis* were immediately freeze-dried using a vacuum freeze-dryer. The freeze-dried samples were ground using zirconia beads in a mixer mill (MM 400, Retsch) for 1.5 min at 30 Hz. Lyophilized powder (100 mg) of each sample was dissolved in 1.2 mL of 70 % methanol solution, vortexed 30 s every 30 min for 6 times in total, and samples were stored at 4 °C overnight in the refrigerator. The extracts were filtered (SCAA-104, 0.22 μm pore size) before UPLC-MS/MS analysis after centrifugation at 12000 rpm for 10 min.

## 2.2. Metabolite measurements

The sample extracts were analyzed using a UPLC-ESI-MS/MS system. The metabolites were analyzed using UPLC with an Agilent SB-C18 column (1.8 μm, 2.1 mm \* 100 mm). The mobile phase consisted of solvent A (pure water with 0.1 % formic acid) and solvent B (acetonitrile with 0.1 % formic acid). The gradient program for the mobile phases was: 95 % A and 5 % B initially, followed by a linear gradient of 5 % A and 95 % B within 9 min, which was held for 1 min. Subsequently, the composition was adjusted to 95 % A and 5.0 % B within 1.1 min and held for 2.9 min. The flow velocity was set at 0.35 mL per minute; the column oven was set to 40 °C; the injection volume was 4 μL. The effluent was alternatively connected to an ESI-triple quadrupole-linear ion trap (QTRAP)-MS, and the conditions were as follows: LIT and triple quadrupole (QQQ) scans were acquired on a triple quadrupole-linear ion trap mass spectrometer (QTRAP), AB4500 Q TRAP UPLC/MS System, equipped with an ESI Turbo Ion-Spray interface, operating in positive and negative ion modes and controlled by Analyst 1.6.3 software (AB Sciex). The ESI source operation parameters were as follows: ion source, turbo spray; source temperature 550 °C; ion spray voltage (IS) 5500 V (positive ion mode)/-4500 V (negative ion mode); ion source gas I (GSI); gas II (GSII) and curtain gas (CUR) were set at 50, 60, and 25psi respectively; the collision-activated dissociation (CAD) was high. Instrument tuning and mass calibration were performed with 10 and 100 μmol/L polypropylene glycol solutions in QQQ and LIT modes, respectively. QQQ scans were acquired in the MRM experiments with a collision gas (nitrogen) in the medium. The DP and CE for individual MRM transitions were further optimized.

A specific set of MRM transitions were observed for each period based on the metabolites eluted during that time. Analyst 1.6.3 was used for the UPLC/MS/MS data. The function of the MRM metabolite ion current intensity maps versus retention time was compared with that of self-built metabolite databases. Calibration and integration of the chromatographic peaks were accomplished using MultiaQuant. The area of each chromatographic peak represents the relative content of the metabolite. Finally, the integral data for all chromatographic peak areas was derived and stored.

## 2.3. Data processing and analysis of metabolites

Unsupervised principal component analysis (PCA) was used to identify significantly regulated metabolites ( $VIP \geq 1$  and absolute  $|\text{Log}_2\text{Fold Change}| \geq 1$ ) between groups using statistics function `prcomp` within R. VIP values extracted from OPLS-DA data, which also included score and permutation plots, were generated using R package `MetaboAnalystR`. The data were log-transformed ( $\text{Log}_2$ ) and mean-centered before OPLS-DA. A permutation test (200 permutations) was used to avoid overfitting. The identified metabolites were annotated using the Kyoto Encyclopedia of Genes and Genomes (KEGG) compound database [26], and the annotated metabolites were mapped to the KEGG pathway database [26].

## 2.4. Total RNA detection, library preparation, transcriptome clustering and sequencing

RNA degradation and contamination were monitored using 1 % agarose gel electrophoresis. RNA purity was assessed using a nanophotometer® spectrophotometer (IMPLEN, CA, USA). The RNA concentration was measured using the Qubit® RNA Assay Kit and the Qubit®2.0 Fluorometer (Life Technologies, CA, USA). RNA integrity was assessed using an RNA Nano 6000 Assay Kit on a Bioanalyzer 2100 system (Agilent Technologies, CA, USA). A total of 1 μg RNA per sample was used as input material for the RNA sample preparations. Sequencing libraries were created using NEBNext® Ultra™ RNA Library Prep Kit for Illumina® (NEB, USA) following the manufacturer's recommendations, and index codes were used to assign sequences to each sample. The library fragments were purified using the AMPure XP system (Beckman Coulter, Beverly, USA) to select cDNA fragments ranging preferentially from 250 to 300 bp in length. The size-selected, adaptor-ligated cDNA was treated with 3 μL of USER enzyme (NEB, USA) at 37 °C for 15 min, followed by 5 min at 95 °C before PCR. Phusion High-Fidelity DNA polymerase, Universal PCR primers, and Index (X) primer were used for PCR. Finally, the PCR products were purified (AMPure XP system), and library quality was assessed using the Agilent Bioanalyzer 2100 system. The index-coded samples were clustered using a cBot Cluster Generation System and the TruSeq PE Cluster Kit v3-cBot-HS (Illumina) according to the manufacturer's instructions. After cluster creation, the library preparations were sequenced on an Illumina HiSeq platform, and 125 bp/150 bp paired-end reads were generated.

## 2.5. Data processing, transcriptome assembly and functional annotation

For data quality control, `Fastp` [27] v. 0.19.3 was used to filter the original data, mainly to remove reads with adapters. The paired reads were removed when the N content in any sequencing reads exceeded 10 % of the base number of the reads and any sequencing reads with the number of low-quality ( $Q \leq 20$ ) bases contained in reads exceeded 50 % of the reads' bases. All subsequent analyses were performed using clean reads. The clean reads were produced for the index and compared to the reference genome (*G. uralensis*) using `HISAT` [28,29] v2.1.0. `StringTie` [30] v1.3.4d was used to predict the new genes. `DIAMOND` [31] `blastx` was used to compare the

new genes with the sequences in the KEGG, GO, NR, Swiss-Prot, trEMBL, and KOG databases to obtain the annotation results. We used featureCounts [32] v1.6.2 to compute gene alignment, and then calculated the FPKM of each gene based on its gene length.

## 2.6. Differential expression analysis and functional enrichment

DESeq2 [33,34] v1.22.1 was used to analyze differential expression between the two groups, and the  $p$ -value was corrected using the Benjamini and Hochberg method. The corrected  $p$ -value and  $|\text{Log}_2\text{Fold Change}|$  were used as thresholds for significantly different expressions. An enrichment analysis was performed using a hypergeometric test. For KEGG [26], a hypergeometric distribution test was performed with the pathway as the unit. Gene Ontology (GO) [35] analysis was performed based on GO terms. DEGs were identified and commented on using the KOG database.

## 2.7. Combined metabolomic and transcriptomic analysis method

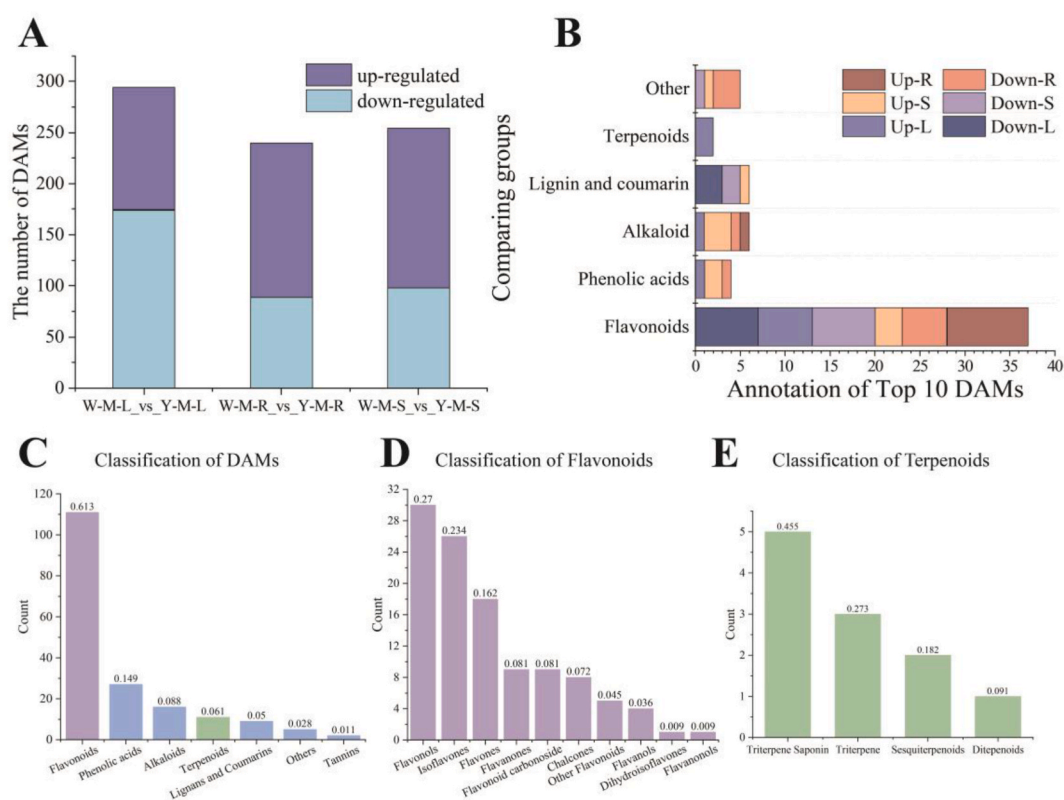
The transcriptome and metabolome were first analyzed using PCA. DEGs and DAMs in the same comparison group were mapped to the KEGG pathway map simultaneously, and the pathways enriched by the two omics were identified. The quantitative values of genes and metabolites in all samples were used for correlation analysis. The correlation method uses the `cor` function in R to calculate the Pearson correlation coefficient of genes and metabolites, select differentially related genes and metabolites, and construct diagrams.

Differential metabolites and genes in the pathways were analyzed using CCA. Finally, O2PLS [36] analysis was used to determine the differential variables between the different groups.

## 3. Results

### 3.1. Metabolomics analyses indicate isoflavonoids vary significantly between *G. squamulose* and *G. uralensis*

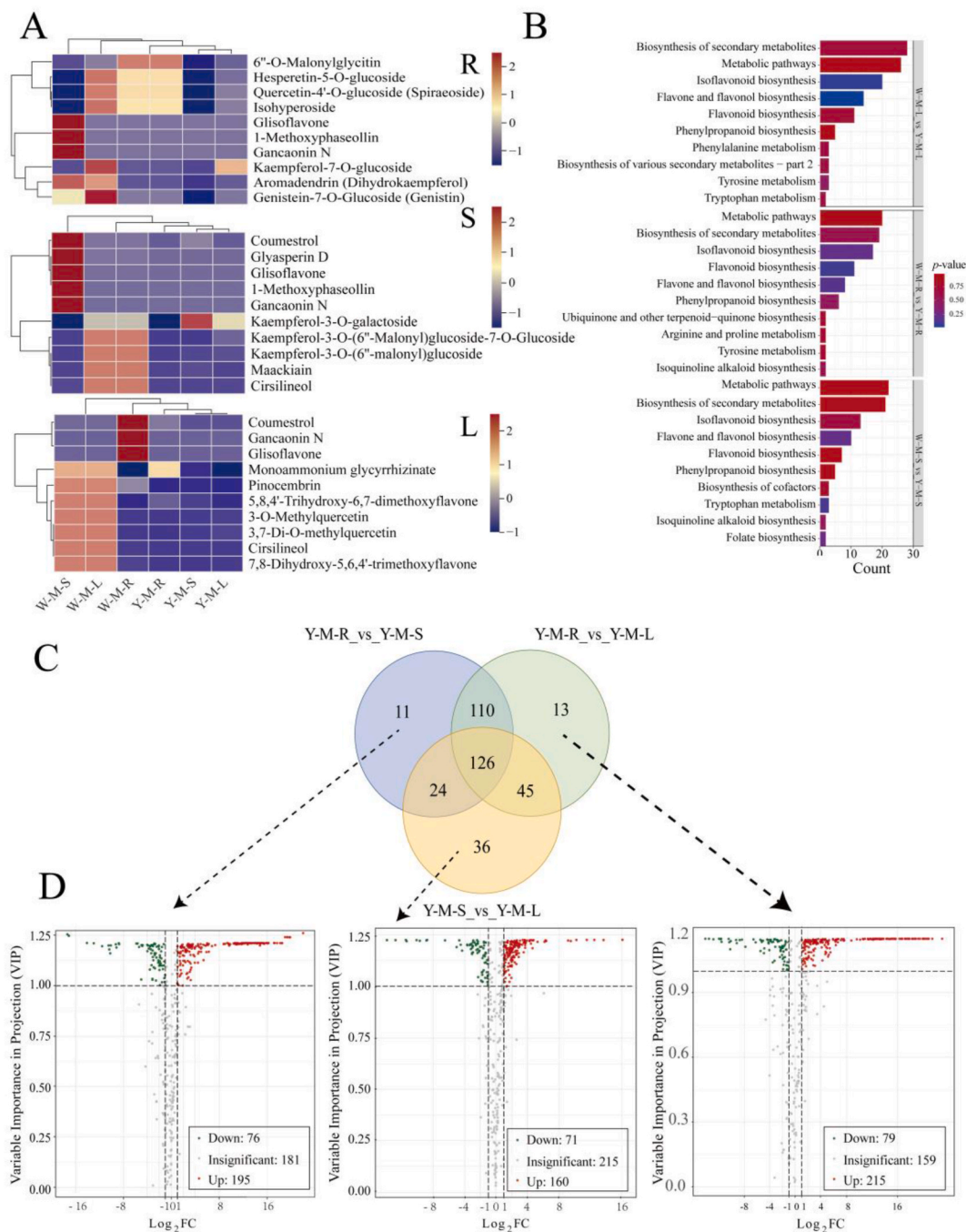
Metabolomic analysis was used to identify the similarities and differences between *G. squamulose* and *G. uralensis*. Based on the



**Fig. 1.** The quantitative analysis of differentially accumulated metabolites (DAMs) between different parts of *G. squamulose* and *G. uralensis*. (A) Numbers of DAMs in the same plant part of different species. W-M represents the metabolomics of *G. uralensis*, while Y-M represents the metabolomics of *G. squamulose*. L, S, and R stand for leaf, stem, and root, respectively. (B) Annotation of the top10 DAMs. Up represents upregulated, and down represents downregulated. (C) Classification of DAMs. Remarkably, the purple bar in the histogram represents flavonoids, and the green bar represents terpenoids. (D) Classification of flavonoids. (E) Classification of terpenoids.



UPLS-MS/MS data, the metabolites were qualitatively analyzed using a secondary spectrum matching method. A total of 453 metabolites were identified, as shown in Fig. S2 as total ion current (TIC) in negative and positive modes. PCA and OPLS-DA were used to obtain a preliminary view of the overall metabolic differences between groups and the deviation of samples within their groups. In the




**Fig. 2.** The visual analysis of DAMs between different parts of *G. squamulose* and *G. uralensis*. (A) Heatmap shows the differential expression patterns of partial DAMs, the values of selected DAMs' |Log<sub>2</sub>Fold Change| are top 10 in each part. Cell colors are the chromatographic peak area (symbolizing relative content) obtained after the relative content standardization treatment (darker red cells indicate higher content, darker blue cells indicate lower content). Three heatmaps present the differential expression patterns of all DAMs between *G. squamulose* and *G. uralensis* in roots, stems, and leaves, respectively. W-M represents the metabolomics of *G. uralensis*, whereas Y-M represents the metabolomics of *G. squamulose*. (B) The annotation results of the cDAMs were classified according to the pathways in the Kyoto Encyclopedia of Genes and Genomes (KEGG). (C) Venn plot shows shared and unique DAMs between each group. (D) The Volcano plots of DAMs correspond to the group shown in the Venn plot. Down, insignificant, and up represent downregulated, insignificant, and upregulated DAMs, respectively.

**Table 1**  
Identified flavonoids with Log2FC between *G. squamulose* and *G. uralensis* and their *p*-values in three organs.

NO.	Metabolites	Formula	Log2FC	<i>P</i> -value	Organism
1	Kaempferol-3- <i>O</i> -galactoside	C <sub>21</sub> H <sub>20</sub> O <sub>11</sub>	-15.86	*	S
2	Kaempferol-3- <i>O</i> -(6"-malonyl) glucoside	C <sub>24</sub> H <sub>22</sub> O <sub>14</sub>	-15.86	*	S
3	Hispidulin	C <sub>16</sub> H <sub>12</sub> O <sub>6</sub>	14.53	**	S
4	Glisoflavone	C <sub>21</sub> H <sub>20</sub> O <sub>6</sub>	-13.65	*	S
5	Kaempferol-3- <i>O</i> -(6"-Malonyl) glucoside-7- <i>O</i> -Glucoside	C <sub>30</sub> H <sub>32</sub> O <sub>19</sub>	-6.83	*	S
6	Epicatechin	C <sub>15</sub> H <sub>14</sub> O <sub>6</sub>	6.78	*	S
7	Cirsilineol	C <sub>18</sub> H <sub>16</sub> O <sub>7</sub>	-6.65	*	S
8	Licoflavonol	C <sub>20</sub> H <sub>18</sub> O <sub>6</sub>	6.46	*	S
9	7,8-Dihydroxy-5,6,4'-trimethoxyflavone	C <sub>18</sub> H <sub>16</sub> O <sub>7</sub>	-6.35	*	S
10	Apigenin-8- <i>C</i> -(2"-glucosyl) arabinoside	C <sub>26</sub> H <sub>28</sub> O <sub>14</sub>	5.86	**	S
11	Pinocembrin	C <sub>13</sub> H <sub>12</sub> O <sub>4</sub>	-5.44	**	S
12	Quercetin-7- <i>O</i> -(6"-malonyl) glucosyl-5- <i>O</i> -glucoside	C <sub>30</sub> H <sub>32</sub> O <sub>20</sub>	-5.39	**	S
13	Puerarin	C <sub>21</sub> H <sub>20</sub> O <sub>9</sub>	5.24	*	S
14	3,7-Di- <i>O</i> -methylquercetin	C <sub>17</sub> H <sub>14</sub> O <sub>7</sub>	-5.19	*	S
15	Glabridin	C <sub>20</sub> H <sub>20</sub> O <sub>4</sub>	-5.02	**	S
16	Afromosin	C <sub>17</sub> H <sub>14</sub> O <sub>5</sub>	19.5	***	L
17	Pinocembrin	C <sub>13</sub> H <sub>12</sub> O <sub>4</sub>	-17.69	**	L
18	5,6,7-Tetrahydroxy-8-methoxyflavone	C <sub>16</sub> H <sub>12</sub> O <sub>6</sub>	11.41	***	L
19	3- <i>O</i> -Methylquercetin	C <sub>16</sub> H <sub>12</sub> O <sub>7</sub>	-10.7	**	L
20	Maackiain	C <sub>16</sub> H <sub>12</sub> O <sub>5</sub>	-8.03	***	L
21	Isorhamnetin	C <sub>16</sub> H <sub>12</sub> O <sub>7</sub>	-6.73	**	L
22	Diosmetin	C <sub>16</sub> H <sub>12</sub> O <sub>6</sub>	-6.58	*	L
23	6,7,8-Tetrahydroxy-5-methoxyflavone	C <sub>16</sub> H <sub>12</sub> O <sub>6</sub>	-6.57	***	L
24	Hispidulin	C <sub>16</sub> H <sub>12</sub> O <sub>6</sub>	-6.55	*	L
25	Rhamnocitrin	C <sub>16</sub> H <sub>12</sub> O <sub>6</sub>	-6.54	**	L
26	Pratensein	C <sub>16</sub> H <sub>12</sub> O <sub>6</sub>	-6.52	***	L
27	Epicatechin	C <sub>15</sub> H <sub>14</sub> O <sub>6</sub>	5.46	*	L
28	Pterocarpine	C <sub>17</sub> H <sub>14</sub> O <sub>5</sub>	5.41	**	L
29	5-Hydroxy-6,7-dimethoxyflavone	C <sub>17</sub> H <sub>14</sub> O <sub>5</sub>	5.25	*	L
30	3',7-dihydroxy-4'-methoxyflavone	C <sub>16</sub> H <sub>12</sub> O <sub>5</sub>	-4.59	**	L
31	Hispidulin-7- <i>O</i> -Glucoside	C <sub>22</sub> H <sub>22</sub> O <sub>11</sub>	17.09	*	R
32	Isorhamnetin-3- <i>O</i> -(6"-malonyl) glucoside	C <sub>25</sub> H <sub>24</sub> O <sub>15</sub>	15.04	**	R
33	6"- <i>O</i> -Malonylglucitin	C <sub>25</sub> H <sub>24</sub> O <sub>13</sub>	-13.78	**	R
34	Kaempferol-7- <i>O</i> -glucoside	C <sub>21</sub> H <sub>20</sub> O <sub>11</sub>	-13.74	**	R
35	Diosmetin-7- <i>O</i> -(6"-malonyl) glucoside	C <sub>25</sub> H <sub>24</sub> O <sub>14</sub>	13.16	**	R
36	Quercetin-4'- <i>O</i> -glucoside (Spiraeoside)	C <sub>21</sub> H <sub>20</sub> O <sub>12</sub>	-13.01	*	R
37	Hesperetin-5- <i>O</i> -glucoside	C <sub>22</sub> H <sub>24</sub> O <sub>11</sub>	-12.92	*	R
38	Aromadendrin	C <sub>15</sub> H <sub>12</sub> O <sub>6</sub>	-11.93	***	R
39	Genistein-7- <i>O</i> -Glucoside	C <sub>21</sub> H <sub>20</sub> O <sub>10</sub>	-11.51	*	R
40	Gancaonin N	C <sub>21</sub> H <sub>20</sub> O <sub>6</sub>	-10.95	**	R
41	Glisoflavone	C <sub>21</sub> H <sub>20</sub> O <sub>6</sub>	-10.73	***	R
42	Kaempferol-3- <i>O</i> -(6"-Malonyl) glucoside-7- <i>O</i> -Glucoside	C <sub>30</sub> H <sub>32</sub> O <sub>19</sub>	-10.5	*	R
43	Naringenin-4'- <i>O</i> -glucoside	C <sub>21</sub> H <sub>22</sub> O <sub>10</sub>	-9.91	**	R
44	6,4'-Dimethoxyisoflavone-7- <i>O</i> -glucoside	C <sub>23</sub> H <sub>24</sub> O <sub>10</sub>	7.8	**	R
45	Isoliquiritin	C <sub>21</sub> H <sub>22</sub> O <sub>9</sub>	-7.71	**	R

*P*-value: \*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$ . Log2FC: Log2Fold Change; the average of three parallel data points from the same plant species was calculated. The accumulation of *G. squamulose* was compared with that of *G. uralensis*. In the 'metabolites' column, metabolites that show significant differences in multiple species at the same time are marked by color, and the same metabolites are given the same color. In the 'Plant species' column, S, L, and R represent the stem, leaf, and root, respectively.

Color-coded according to the fold change using a color bar 

PCA score plot, the sample points swarmed together with those from the same groups, but were separated across other groups, revealing significant differences in metabolites between various organs of the two species of licorice (Fig. S3). S-plots and fold-change plots based on the OPLS-DA results revealed significantly different metabolites in the common parts of the two licorice species, including 2',4'-dihydroxy-4-methoxydihydrochalcone, pterocarpine, quercetin-3-O-(2"-O-galactosyl) glucoside, and 6"-O-malonyl-glycitin (Fig. S4).

Samples from the same parts of different species were compared to identify the relationships between DAMs. The number of DAMs in the same organs of different species is shown in a bar chart (Fig. 1A), which indicates the fewest differences in metabolites in the roots of the two plants, whereas the greatest differences were observed in the leaves. Previous studies have identified the therapeutic potential of various licorice species based on their terpenes, saponins, and flavonoids. Licorice glycosides and glycyrrhizin are particularly noteworthy since they are recognized as crucial bioactive compounds [37]. After analysis, 181 of the 453 DAMs had a substance identification level of 1 (the sample substance's secondary mass spectrometry, RT, and database substance matching scores were more than 0.7 points), among which flavonoids accounted for the majority (61.3%), followed by alkaloids, lignins, coumarins, and terpenoids (Fig. 1C), which is consistent with previous studies. Considering that flavonoids and terpenoids are mainly responsible for anti-inflammatory effects, we continued to classify them into secondary classes. The results showed that flavonols and isoflavonoids were the metabolite categories with significant differences (Fig. 1D), while triterpene saponins, glycyrrhizic acid, and glycyrrhetic acid, which have anti-inflammatory activity, did not appear as DAMs; therefore, terpenoids were not discussed as the main component (Fig. 1E). To investigate the accumulated metabolites in specific plant parts, a heat map of all the obtained metabolites was plotted for all group samples (Fig. 2A), which indicated that the concentration of flavonoids varied significantly across different plant parts. In addition, K-means clustering was performed (Fig. S5). In subclasses 4 and 5, which showed higher metabolite content in the root tissues, flavonoids were the most abundant, verifying the importance of flavonoids in this study. We then focused more on flavonoids, and the top 15 differentially accumulated flavonoids in the three organs between *G. squamulose* and *G. uralensis* were filtered with the cut-offs set at adjusted *p*-values <0.05, and the top 15 |Log2Fold Change| (Table 1). Among the 45 labeled metabolites, five were not only present in one organ, but the same metabolites in the table were also labeled with the same color. Twenty of the forty-five annotated metabolites differed significantly between *G. squamulose* and *G. uralensis*. Subsequently, we focused on variations in metabolites within individual plants. Samples from different organs of each species were investigated, and a Venn plot was constructed to show the relationship between the DAMs (Fig. 2C). A total of 126 and 170 cDAMs were found in *G. squamulose* and *G. uralensis*, respectively, while DAMs between plant parts were comparatively distributed in these two species. Volcano plots showed the significance of the upregulated and downregulated compounds between the two groups in *G. squamulose*, notably, the root and leaf groups had the most DAMs (Fig. 2D). Moreover, DAMs were significantly upregulated in root tissues.

### 3.2. KEGG pathway analysis

DAMs were annotated and displayed in the KEGG (Kyoto Encyclopedia of Genes and Genomes) database. KEGG annotation was used to identify the relevant pathways, and the corresponding DAMs are shown in Fig. 2B. In terms of metabolism, the metabolic pathway and the biosynthesis of secondary metabolites were the two classifications with the highest proportions among the three comparison groups. In the KEGG enrichment analysis, isoflavonoid, flavone, and flavonol biosynthesis were the most significant pathways in the leaves, whereas flavone, flavonol, and isoflavonoid biosynthesis were the most significant pathways in the stems. Isoflavonoid and flavonoid biosynthesis were the most significant pathways in roots. Because flavonoids are one of the main anti-inflammatory compounds in licorice, the roots and stems were predicted to be the medicinal parts of *G. squamulose*, which is consistent with the above results (Fig. 1B).

### 3.3. Transcriptomics analysis of *G. squamulose*

Transcriptomic analysis was performed to explore the similarities and differences between *G. squamulose* and *G. uralensis* at the genetic level. Using RNA-seq and filtration, a total of 118.9 Gb of clean data were obtained, with the average clean data exceeding 5 Gb and Q30 bases scoring >93% for every sample. The clean reads were aligned to the reference genome using HISAT2, and the total number of mapped reads was >70%, indicating satisfactory compliance and cleanliness. The alignment visualized as the GC content was relatively consistent (approximately 41–45%) across all samples. Based on the alignment data, StringTie and *de novo* alignments were performed on the splice transcripts.

The transcripts underwent BLAST analysis for genomic annotation using GffCompare in order to detect novel transcripts or genes. A total of 8651 novel transcripts were screened, and 1797 new genes were annotated using at least two databases: KEGG, Gene Ontology (GO), NCBI non-redundant proteins (NR), Swiss-Prot, TrEMBL, and EuKaryotic Orthologous Groups (KOG). Among them, 49 of the 1797 novel genes were annotated in flavonoid-associated KEGG pathways. We used iTAK to predict transcription factors in plants and identified 44 new genes related to transcription factors (TFs). Combined with other information from the transcripts, these genes were selected for transcriptomic analysis.

### 3.4. Differentially expressed genes (DEGs) between the two licorice species in different organs

Fragments per kilobase per million (FPKM) were used as the metric to measure expression levels and quantities in various samples. DESeq2/edgeR, the Benjamini-Hochberg test and featureCounts/StringTie, were used to acquire DEGs between samples by applying a filter threshold of |Log2Fold Change| ≥ 1 and FDR <0.05. Compared with FPKM, the lowest number of DEGs was expressed between

W-T-R and Y-T-R. A total of 12,466, 11,715, and 10,444 DEGs were identified in leaf (7246 downregulated and 5220 upregulated), stem (6898 downregulated and 4817 upregulated), and root samples (5798 downregulated and 4646 upregulated), respectively. The root samples of the two species contained the fewest number of DEGs, whereas the leaf samples contained the highest number, which corresponded to the distribution pattern of the DAMs to some extent. All DEGs were studied, and a heatmap was generated to present the different expression levels of candidate genes in the identified groups (Fig. 3A). The hierarchical cluster analysis was consistent with this grouping. In addition, we performed K-means cluster analysis on the DEGs; however, the results were unclear (Fig. S6).

For the KEGG enrichment analysis, we screened the top 20 most significantly enriched pathways using the rich factor and  $q$ -value. KEGG bubble pictures showed that candidate genes involved in metabolic pathways and the biosynthesis of secondary metabolites were the most common (Fig. 3B). Biosynthesis of secondary metabolites, plant-pathogen interactions, and plant hormone signal transduction were the most prominent pathways identified in the three comparison groups. The pathways related to flavonoids and isoflavonoids in the leaves and stems were distinct, whereas their absence in the roots indicates similarities between the two species. The DEGs were annotated using several databases, including GO, KEGG, and KOG (Fig. S7), to clarify their functions. The GO classification histogram showed that the metabolic and cellular processes contained most of the genes involved in biological processes among the groups (Fig. 3C). The top 15 GO terms with the lowest  $q$ -values were screened to show the results of the enrichment and pathway enrichment analyses. The analysis of biological processes revealed that the leaf and stem samples had the greatest percentages of secondary metabolite biosynthetic processes. However, the root samples differed as their secondary metabolite biosynthetic processes were excluded from the top 15, indicating that metabolites in the two species vary largely, even from the same parts (leaves and stems), whereas roots bear some similarities. The KEGG annotation histogram shows the relevant pathways and corresponding DEGs. For metabolism, metabolic pathways and the biosynthesis of secondary metabolites were the two most prominent pathways in the three comparison groups, supporting the conclusions of the GO annotation. KOG was used to analyze the clustering of eukaryotic homologous proteins suitable for the annotation of eukaryotic genomes. The classification of KOG annotations showed that general function prediction had the most significant number of DEGs among all the three groups, followed by post-translational modification, protein turnover, chaperones, and signal transduction mechanisms, indicating that the production and modification of enzymes and signal transport may differ between *G. uralensis* and *G. squamulose*. GO terms were used to analyze genes that were not significantly differentially expressed to the extent that they were biologically significant. These genes are involved in plant-pathogen interactions, starch and sucrose metabolism, developmental processes, and cellular component organization, or biogenesis.

### 3.5. Comparison of gene expression patterns of flavonoid and isoflavone biosynthesis pathways between the root group and stem group

Based on metabolomic and transcriptomic analyses, flavonoids were the most significant DAMs. According to a previous study, flavonoids are among the major chemical components of licorice [38]. Hence, we mainly focused on the DEGs associated with flavonoid biosynthesis. We compared the groups containing roots and stems because licorice flavonoids were mainly extracted from these parts [7].

To explore the significantly different metabolic pathways involved in flavonoid biosynthesis in the two species, we mapped the DEGs and DAMs of the same comparison groups using a KEGG pathway map. KEGG enrichment analysis of the metabolites and transcriptomics revealed that KEGG pathways were co-enriched among the three groups. Isoflavonoid and flavonoid biosynthesis pathways were relatively apparent in the root and stem groups. Because of elevated levels of phenylpropanoid biosynthesis, it was considered an earlier step in the synthesis pathway (Fig. 5A and B).

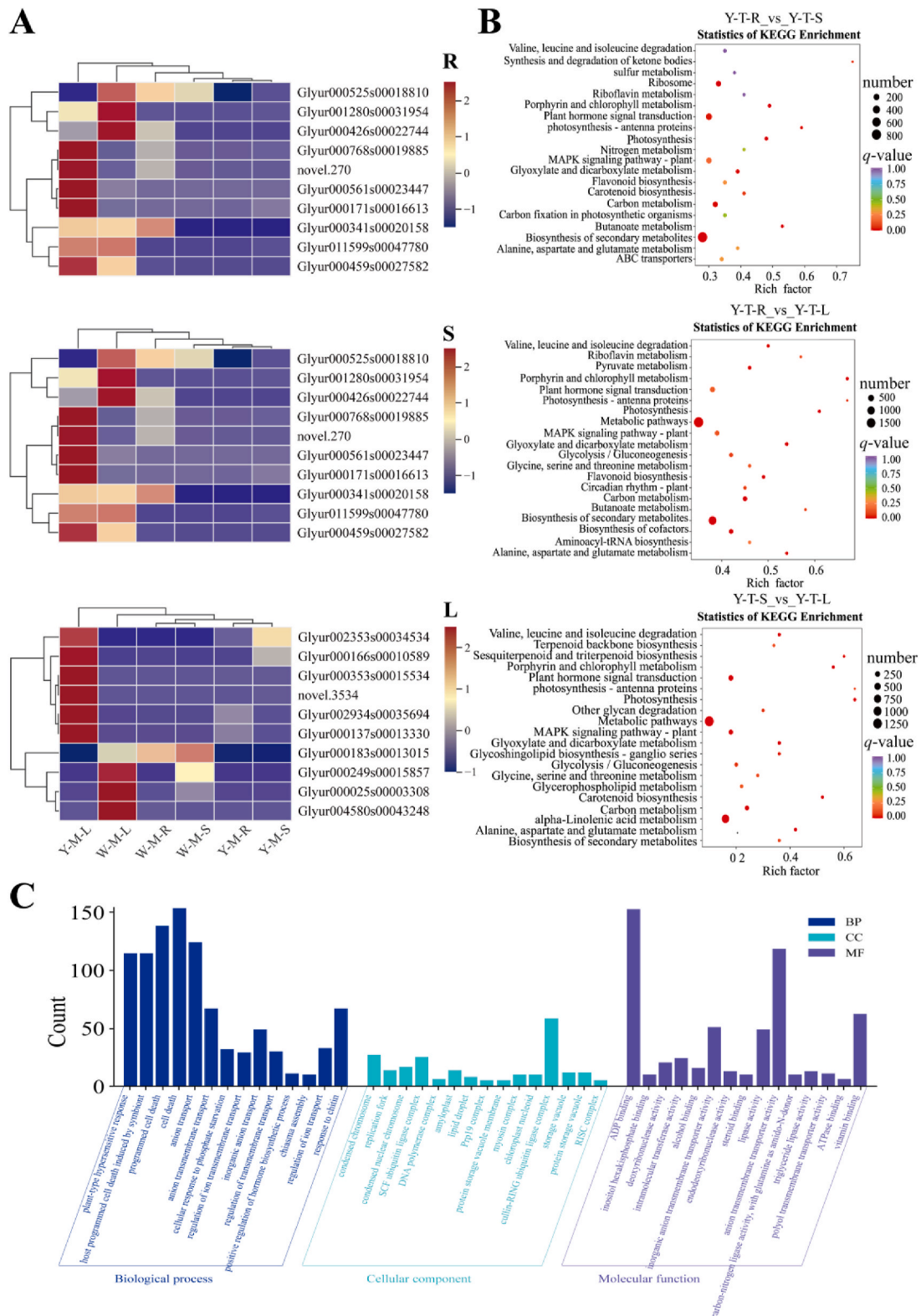
To better understand the interactions between genes and metabolites, the transcript and metabolite levels associated with flavonoid biosynthesis were comprehensively analyzed. We selected three isoflavonoids with significant differences, which were considered important endpoints for our proposed biosynthetic pathway: malonylglycitin, malonyldaidzin, and malonylgenistin. Combined with KEGG pathways, 22 metabolites were identified, and 27 participating DEGs (Table S1), whose  $\text{Log}_2|\text{Fold Change}| > 3$  (Fig. 4) were selected.

In the proposed isoflavonoid synthesis pathway, 11 key enzymes were involved, which had specific upregulated or downregulated relationships with DEGs, resulting in compound up- or down-regulation. The majority of enzyme-encoding genes were regulated in both the roots and stems of *G. squamulose* and *G. uralensis*. However, the vast majority of metabolites were not significantly different. Remarkably, malonylglycitin and malonylgenistin were significantly downregulated in stems, whereas malonylglycitin and malonyldaidzin were downregulated in roots. We conjectured that this phenomenon is related to the downregulation of genes (*Glyur000788s00032588* and *Glyur000488s00030345*) encoding IF7GT, which catalyzes the first step of glycosylation. However, malonylgenistin was upregulated in the root group, which was related to the mixed regulation of genes encoding IF7MAT. In addition, the gene (*Glyur000721s00027409*) encoding CYP93C was not significantly differentially expressed in the roots but was significantly downregulated in the stems. This may explain why the roots and stems contain different amounts of malonylglycitin, malonyldaidzin, and malonylgenistin. The results revealed differences in the metabolomic and transcriptomic levels between *G. squamulose* and *G. uralensis* and that some important isoflavonoids were upregulated in the roots and stems of *G. squamulose*.

### 3.6. Transcription factors (TFs) encoded by DEGs

Studies have reported that Transcriptional control plays a central role in the modulation of flavonoid biosynthesis [17]. As the number of DEGs encoding MYB was the highest, we focused on the MBW complex (Fig. 5C). The MBW complex, composed of MYB, bHLH, and WD40, is the main transcriptional regulator of flavonoid biosynthesis. MBW transcription factors operate mostly as repressors and are mainly considered.



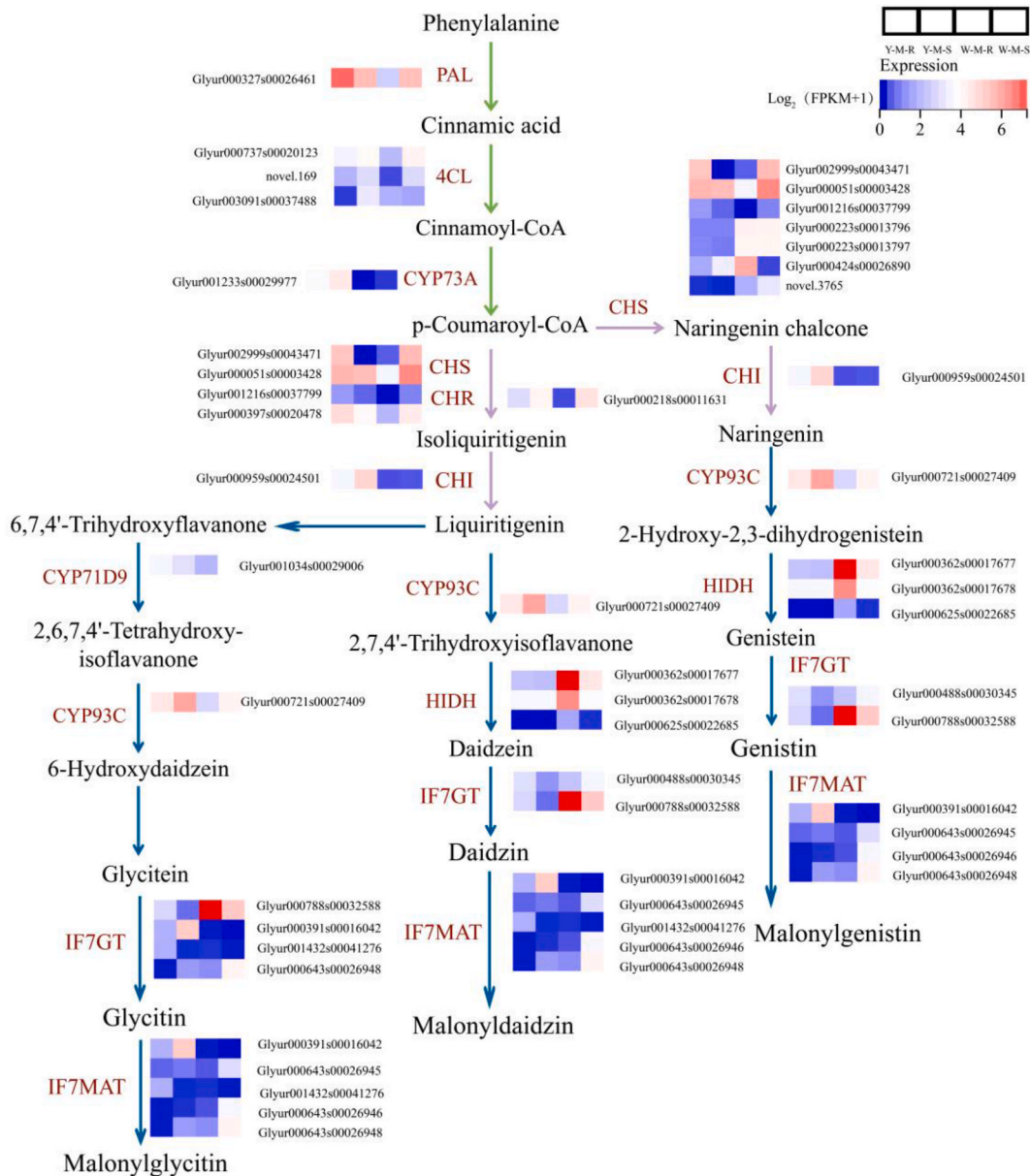


**Fig. 3.** Heat maps and KEGG enrichment bubble diagrams of all differentially expressed genes (DEGs). (A) Differentially expressed gene clustering heat maps. Heatmaps show the differential expression patterns of partial DEGs, the values of selected DEGs'  $|\text{Log}_2\text{Fold Change}|$  are included in the top 10 in each part. Cell colors are the fragments per kilobase per million (FPKM) values obtained after the relative content standardization treatment (darker red cells indicate high content, darker blue cells indicate low content). Three heatmaps present the differential expression patterns



of all DAMs between *G. squamulose* and *G. uralensis* in roots, stems, and leaves, respectively. (B) KEGG enrichment bubble diagrams. The color represents KEGG classification. The x-coordinate represents the enrichment factors of the pathway in different omics, and the y-coordinate represents the names of KEGG pathway. The gradient of red to yellow to blue represents the change of enrichment significance from high to medium to low, which is represented by the *q*-value. And the size of the points indicates the number of enriched DEGs. Y-T represents the transcriptome of *G. squamulose*. (C) GO term enrichment results for genes from the root.

To determine the influence of TFs on flavonoid biosynthesis, we annotated the DEGs in the roots and stems of *G. squamulose* and *G. uralensis*. There were 38,762 DEGs, of which 2303 (5.94 %) were TFs (Fig. 5D). In addition, 92 DEGs encoded MYB, of which 48 were upregulated and 44 were downregulated. Notably, one remarkably downregulated DEG (*Glyur001184s00030523*) was annotated as an



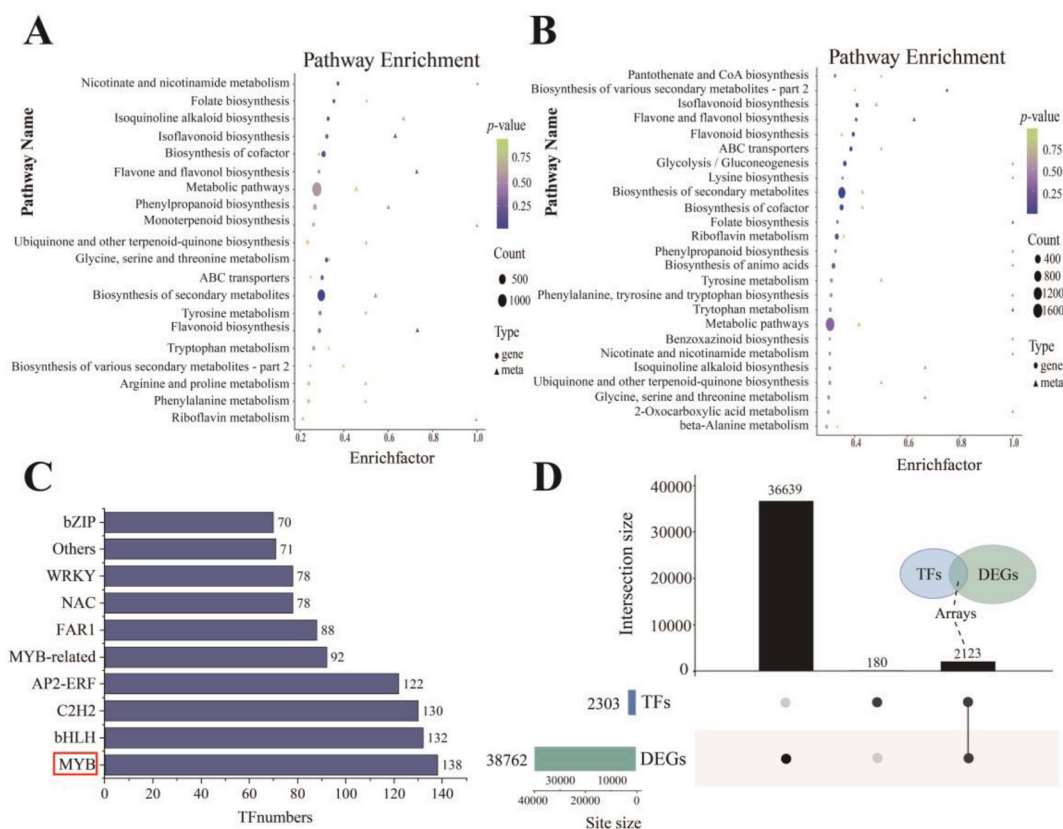
**Fig. 4.** The partial isoflavonoid biosynthesis in plants relevant to our research. The green arrows refer to steps derived from the phenylpropanoid biosynthesis pathway, while the purple and blue steps are derived from the flavonoid biosynthesis and isoflavonoid biosynthesis pathways, respectively. Enzymes encoded by the differentially expressed genes were marked in red, and the heatmaps showed Log<sub>2</sub>(FPKM+1) values for each DEG. PAL, phenylalanine ammonia-lyase; 4CL, 4-coumarate-CoA ligase; CYP73A, *trans*-cinnamate 4-monooxygenase; CHS, chalcone synthase; CHR, chalcone reductase; CHI, chalcone isomerase; CYP93C, 2-hydroxyisoflavone synthase; HIDH, 2-hydroxyisoflavone dehydratase; F6H, flavonoid 6-hydroxylase; IF7GT, isoflavone 7-*O*-glucosyltransferase; IF7MAT, isoflavone 7-*O*-glucoside-6''-*O*-malonyltransferase.

MYB transcription factor in the stem group, indicating that flavonoid expression in the stems of *G. squamulose* was higher. However, most DEGs were not annotated using the KEGG pathway, and the specific enzymes that they regulated were not identified.

#### 4. Discussion

Owing to its significance, the need for licorice as a TCM and food resource is increasing worldwide, and the main source of supply is from wild plants of traditional licorice resources, such as *G. uralensis*. However, the yield of wild *G. uralensis* has been decreasing, emphasizing an urgent need to protect the germplasm of wild licorice resources. Therefore, we integrated metabolomics and transcriptomics of wild *G. squamulose* to study and protect rare licorice species. The identification of differentially accumulated flavonoid content and flavonoid biosynthetic genes in *G. squamulose* and *G. uralensis* is necessary to investigate the production of medicinal bioactive constituents of *G. squamulose* and may provide additional information to guide the protection of this species.

Previous studies have focused on the bioactivity, clinical effects [39], and mechanisms of action of the flavonoids in licorice [40, 41]. Furthermore, research on licorice has mainly focused on the compounds from *G. glabra* and *G. uralensis* [42,43]. However, studies of the bioactive medicinal constituents of *G. squamulose* using metabolomic and transcriptomic analyses are lacking. To the best of our knowledge, this is the first time such a database has been constructed. Flavonoids such as liquiritin are markers for the quality evaluation of *G. uralensis*. Enhancing the accumulation of the principal bioactive constituents has become a target for improving the medicinal value of *G. squamulose*. According to the People's Republic of China Pharmacopoeia, glycyrrhizic acid and liquiritin are essential components [10]. We used metabolomics to identify the difference between *G. squamulose* and *G. uralensis*, and compared the relative content of metabolites (estimated based on the chromatographic peak area) in *G. squamulose* and *G. uralensis* (Fig. S8). The glycyrrhizic acid content of *G. squamulose* was lower than that of *G. uralensis*. Liquiritin was upregulated in the stems and down-regulated in the leaves and roots of *G. squamulose*. Therefore, the complex metabolic pathways and gene regulatory networks were investigated to identify the DAMs and the respective candidate genes in *G. squamulose* and *G. uralensis*. In this study, metabolites and sequenced transcripts from the roots, stems, and leaves of *G. uralensis* and *G. squamulose* were measured. Isoflavonoid regulation



**Fig. 5.** KEGG enrichment analysis and the analysis of transcription factors (TFs). (A) KEGG enrichment bubble diagram between W-M-R and Y-M-R. (B) KEGG enrichment bubble diagram between W-M-S and Y-M-S. The x-coordinate represents the enrichment factors of the pathway in different omics, and the y-coordinate represents the names of KEGG Pathway. The gradient of blue to purple to green represents the change of enrichment significance from high to medium to low, which is represented by  $p$ -value. The shape of the bubble represents different omics, and the size of the bubble represents the number of different metabolites or genes. (C) The number of DEGs encoding TFs. And MYB is marked in a red box. (D) The number of DEGs encoding TFs displayed by the upset diagram.

patterns in *G. uralensis* and *G. squamulose* were compared using a combination of metabolite content and gene expression to identify the regulation of genes involved in the metabolism of isoflavonoids, such as malonyglycitin, malonyldaidzin, and malonylgenistin. In this study, combined metabolic and transcriptomic data were analyzed. The major anti-inflammatory flavonoid compounds in *G. uralensis* were determined. Plant flavonoids with medicinal properties include flavones, isoflavones, chalcones, flavanones, and flavonols [7]. Partial isoflavones were significantly upregulated in *G. squamulose*, and several genes related to isoflavonoid biosynthesis were found to be more strongly expressed in *G. squamulose* than in *G. uralensis*. We conjectured that the total accumulation of isoflavonoids in *G. squamulose* is greater than that in *G. uralensis*. Combined metabolomics and transcriptomics analyses revealed two metabolic pathways that were significantly different between *G. squamulose* and *G. uralensis*. These were the flavonoid and isoflavone synthesis pathways, which were selected using KEGG enrichment analysis. The DEGs of these two metabolic pathways were associated with DAMs via enzymes involved in these synthetic pathways. The results showed that 27 DEGs were significantly differentially expressed and highly correlated with enzymes involved in metabolic pathways. The 27 DEGs encoded 11 enzymes (PAL, 4CL, CYP73A, CHR, CHI, HIDH, CHS, IF7GT, IF7MAT, CYP93C, and CYP71D9), among which CYP93C and IF7GT were significantly differentially expressed. IF7GT was significantly downregulated in both roots and stems, and the combined action of IF7MAT resulted in different levels of malonyglycitin, malonyldaidzin, and malonylgenistin in both *G. squamulose* and *G. uralensis*. A preliminary stage in the biosynthesis pathway, CYP93C, was not significantly differentially expressed in the root group, but was significantly expressed in the stem group, which, to some extent, explained the differences in metabolites between the root and stem. A study of the relationship between DEGs and TFs revealed that 92 DEGs encoded MBW complexes, which are the main transcriptional regulators of flavonoid biosynthesis [44, 45]. Among these, one remarkably downregulated gene (*Glyur001184s00030523*) was possibly associated with the higher expression of flavonoids in the stems of *G. squamulose*. However, the majority of genes were not annotated to any metabolic pathway. In summary, we reported the first metabolome and full-length transcriptome of the herbal medicinal plant *G. squamulose*. However, it must be emphasized that a substantial fraction of the cumulative variation in metabolites remains unexplained by transcriptomic analyses, implying that additional research, such as genome sequencing, is required.

Finally, through the analysis of metabolomic and transcriptomic data, we demonstrated the medicinal value of *G. squamulose*, contributing to filling a gap in omics research on *G. squamulose*, and demonstrating the need to protect wild licorice to ensure its species diversity. However, the outdoor habitats of *G. squamulose* are extremely limited. Suitable regions to grow *Glycyrrhiza* are mostly located in northern China, including the Gansu Province, Inner Mongolia Province, and Northeastern China. The limitations of *G. squamulose* in these areas may be related to climatic similarities. The climate in the referred areas is usually temperate continental or temperate monsoon. Climatic conditions affect flavonoid biosynthesis, indicating that temperature, precipitation, and UV light may regulate its biosynthesis. The specific mechanisms of flavonoid metabolic pathways contribute to climate and location choices. Previous studies have shown that flavonoids act as phytoalexins or antioxidants with reactive oxygen species-scavenging ability and protect plants from damage caused by biotic and abiotic [46,47] stresses, including UV irradiation and cold stress [48]. Quercetin-7-O-(6"-malonyl) glucosyl-5-O-glucoside, a type of glycosylated quercetin that acts as an antioxidant flavonoid and was upregulated in *G. squamulose* (Table 1), can be a candidate. To protect the precious *G. squamulose*, we suggest that tissue culture be introduced to cultivate seedlings so that they can be grown in appropriate test regions or laboratories that imitate suitable climatic conditions. Such studies on *G. squamulose* are important to preserve its biodiversity and licorice germplasm. They can also serve as a guide for the preservation of other rarely encountered plant species and for the enhancement and assessment of the therapeutic values of other plants.

## Data availability

The data presented in the study were deposited in National Center for Biotechnology Information (accession number PRJNA1020643).

## CRediT authorship contribution statement

**Bin Ma:** Methodology, Investigation, Conceptualization. **Siru Wang:** Writing – original draft, Formal analysis, Data curation. **Haonan Li:** Writing – review & editing, Visualization, Formal analysis. **Qinyue Wang:** Writing – review & editing, Formal analysis. **Yaqi Hong:** Writing – review & editing. **Yang-mei Bao:** Methodology, Investigation. **Hua Liu:** Investigation, Data curation. **Ming Li:** Investigation. **Yucheng Zhao:** Writing – review & editing, Validation, Supervision. **Lan-ping Guo:** Writing – review & editing, Validation, Supervision.

## Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Acknowledgements

This study was funded by Natural Science Foundation of Ningxia Hui Autonomous Region (2022AAC03421), China Agriculture Research System (CARS-21), Ningxia Hui Autonomous Region Agricultural Science and Technology Independent Innovation Funding Project (NGSB-2021-16), and the sixth batch of autonomous region youth science and technology talents lift project. This work was

also supported by the open foundation of Shaanxi University of Chinese Medicine state key laboratory of R&D of Characteristic Qin Medicine Resources (SUCM-QM202202) and the Fund of Traditional Chinese Medicine Institute of Anhui Dabie Mountain (TCMADM-2023-18).

## Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.heliyon.2024.e30868>.

## Supplementary information

Supplementary data is available at online.

## References

- [1] M.Y. Jiang, S.J. Zhao, S.S. Yang, X. Lin, X.G. He, X.Y. Wei, Q. Song, R. Li, C.M. Fu, J.M. Zhang, Z. Zhang, An "essential herbal medicine"-licorice: a review of phytochemicals and its effects in combination preparations, *J. Ethnopharmacol.* 249 (2020) 112439, <https://doi.org/10.1016/j.jep.2019.112439>.
- [2] I. Kitagawa, Licorice root. A natural sweetener and an important ingredient in Chinese medicine, *Pure Appl. Chem.* 74 (7) (2002) 1189–1198, <https://doi.org/10.1351/pac200274071189>.
- [3] B. Yan, J. Hou, W. Li, L. Luo, M. Ye, Z. Zhao, W. Wang, A review on the plant resources of important medicinal licorice, *J. Ethnopharmacol.* 301 (2023) 115823, <https://doi.org/10.1016/j.jep.2022.115823>.
- [4] R. Yang, B.C. Yuan, Y.S. Ma, S. Zhou, Y. Liu, The anti-inflammatory activity of licorice, a widely used Chinese herb, *Pharmaceut. Biol.* 55 (1) (2017) 5–18, <https://doi.org/10.1080/13880209.2016.1225775>.
- [5] T.C. Kao, C.H. Wu, G.C. Yen, Bioactivity and potential health benefits of licorice, *J. Agric. Food Chem.* 62 (3) (2014) 542–553, <https://doi.org/10.1021/jf404939f>.
- [6] C. Fiore, M. Eisenhut, E. Ragazzi, G. Zanchin, D. Armanini, A history of the therapeutic use of liquorice in Europe, *J. Ethnopharmacol.* 99 (3) (2005 Jul 14) 317–324, <https://doi.org/10.1016/j.jep.2005.04.015>.
- [7] Z.F. Wang, J. Liu, Y.A. Yang, H.L. Zhu, A Review: the anti-inflammatory, anticancer and antibacterial properties of four kinds of licorice flavonoids isolated from licorice, *Curr. Med. Chem.* 27 (12) (2014) 1997–2011, <https://doi.org/10.2174/0929867325666181001104550>.
- [8] L. Jiang, W. Akram, B. Luo, S. Hu, M.O. Faruque, S. Ahmad, N.A. Yasin, W.U. Khan, A. Ahmad, A.N. Shikov, J. Chen, X. Hu, Metabolomic and pharmacologic insights of aerial and underground parts of *Glycyrrhiza uralensis* Fisch. Ex DC. For maximum utilization of medicinal resources, *Front. Pharmacol.* 12 (2021 Jun 1) 658670, <https://doi.org/10.3389/fphar.2021.658670>.
- [9] S.J. Ma, G.W. Zhu, F.L. Yu, G.H. Zhu, D. Wang, W.Q. Wang, J.L. Hou, Effects of manganese on accumulation of Glycyrrhizic acid based on material ingredients distribution of *Glycyrrhiza uralensis*, *Ind. Crop. Prod.* 112 (2018) 151–159, <https://doi.org/10.1016/j.indcrop.2017.09.035>.
- [10] Chinese Pharmacopoeia Commission, *Pharmacopoeia of People's Republic of China*, 2020 ed., 2020. Beijing.
- [11] H. Liang, R.Y. Zhang, Study of the chemical composition of *G. squamulosa*, *Journal of Beijing Medical University* 24 (5) (1992) 399–400, <https://doi.org/10.3321/j.issn:1001-5302.2000.01.013>.
- [12] P.F. Xiong, Z.F. Chen, Q. Yang, J.J. Zhou, H. Zhang, Z. Wang, B.C. Xu, Surface water storage characteristics of main herbaceous species in semiarid Loess Plateau of China, *Ecology* 12 (8) (2019), <https://doi.org/10.1002/eco.2145>. Article e2145.
- [13] C. Wang, Y.S. Han, X.Y. Zhang, Y. Peng, Study of the Chemical Composition of *Glycyrrhiza squamulosa* Franch. (I), *Journal of Henan Normal University (Natural Science)*, 1994, pp. 48–50, <https://doi.org/10.16366/j.cnki.1000-2367.1994.03.013>, 03.
- [14] Y. Fu, J. Chen, Y.J. Li, Y.F. Zheng, P. Li, Antioxidant and anti-inflammatory activities of six flavonoids separated from licorice, *Food Chem.* 141 (2) (2013) 1063–1071, <https://doi.org/10.1016/j.foodchem.2013.03.089>.
- [15] Z.W. Zhang, N.N. Zhang, Y.M. Li, Content of total flavonoids in different varieties of licorice, *China Pharm.* 16 (1) (2013) 49–51, <https://doi.org/10.3969/j.issn.1008-049X.2013.01.016>.
- [16] H.B. Zhou, Theoretical classification and kinship distinguish of eight species of *Glycyrrhiza* originated from China, *World Chinese Medicine* 16 (3) (2021) 509–515, <https://kns.cnki.net/kcms/detail/11.5529.R.20210304.1544.026.html>.
- [17] W.X. Liu, Y. Feng, S.H. Yu, Z.Q. Fan, X.L. Li, J.Y. Li, H.F. Yin, The flavonoid biosynthesis network in plants, *Int. J. Mol. Sci.* 22 (23) (2021) 12824, <https://doi.org/10.3390/ijms222312824>.
- [18] S.S. Qin, K.H. Wei, Z.H. Cui, Y. Liang, M.J. Li, L. Gu, C.Y. Yang, X.L. Zhou, L.X. Li, W. Xu, C. Liu, J.H. Miao, Z.Y. Zhang, Comparative genomics of *spatholobus suberectus* and insight into flavonoid biosynthesis, *Front. Plant Sci.* 11 (2020) 528108, <https://doi.org/10.3389/fpls.2020.528108>.
- [19] J.T. Yang, H.Y. Yan, Y. Liu, L.L. Da, Q.Q. Xiao, W.Y. Xu, Z. Su, GURFAP: a platform for gene function analysis in *Glycyrrhiza uralensis*, *Front. Genet.* 13 (2022) 823966, <https://doi.org/10.3389/fgene.2022.823966>.
- [20] K. Uchida, T. Aoki, H. Suzuki, T. Akashi, Molecular cloning and biochemical characterization of isoflav-3-ene synthase, a key enzyme of the biosyntheses of (+)-pisatin and coumestrol, *Plant Biotechnol.* 37 (3) (2020) 301–310, <https://doi.org/10.5511/plantbiotechnology.20.0421a>, 2020.
- [21] N. Shen, T.F. Wang, Q. Gan, S. Liu, L. Wang, B. Jin, Plant flavonoids: classification, distribution, biosynthesis, and antioxidant activity, *Food Chem.* 383 (2022) 132531, <https://doi.org/10.1016/j.foodchem.2022.132531>.
- [22] L. Yang, K.S. Wen, X. Ruan, Y.X. Zhao, F. Wei, Q. Wang, Response of plant secondary metabolites to environmental factors, *Molecules* 23 (4) (2018), <https://doi.org/10.3390/molecules23040762>.
- [23] D. Xu, H. Lin, Y. Tang, L. Huang, J. Xu, S. Nian, Y. Zhao, Integration of full-length transcriptomics and targeted metabolomics to identify benzyloisoquinoline alkaloid biosynthetic genes in *Corydalis yanhusuo*, *Horticulture research* 8 (2021) 16, <https://doi.org/10.1038/s41438-020-00450-6>.
- [24] V. Kumari, D. Kumar, R. Bhardwaj, D. Kumar, Metabolome analysis, nutrient and antioxidant potential of aerial and underground parts of *Ajuga parviflora* Benth, *Microchem. J.* 187 (2023) 108451, <https://doi.org/10.1016/j.microc.2023.108451>.
- [25] C. Zhong, C. Chen, X. Gao, C. Tan, H. Bai, K. Ning, Multi-omics profiling reveals comprehensive microbe-plant-metabolite regulation patterns for medicinal plant *Glycyrrhiza uralensis* Fisch, *Plant Biotechnol. J.* 20 (10) (2022) 1874–1887, <https://doi.org/10.1111/pbi.13868>.
- [26] M. Kanehisa, M. Araki, S. Goto, M. Hattori, M. Hirakawa, M. Itoh, T. Katayama, S. Kawashima, S. Okuda, T. Tokimatsu, Y. Yamanishi, KEGG for linking genomes to life and the environment, *Nucleic Acids Res.* 36 (Database issue) (2008) D480–D484, <https://doi.org/10.1093/nar/gkm882>.
- [27] S. Chen, Y. Zhou, Y. Chen, J. Gu, fastp: an ultra-fast all-in-one FASTQ preprocessor, *Bioinformatics* 34 (17) (2018) i884–i890, <https://doi.org/10.1093/bioinformatics/bty560>.
- [28] D. Kim, B. Langmead, S.L. Salzberg, HISAT: a fast spliced aligner with low memory requirements, *Nat. Methods* 12 (4) (2015) 357–360, <https://doi.org/10.1038/nmeth.3317>.
- [29] K. Mochida, T. Sakurai, H. Seki, T. Yoshida, K. Takahagi, S. Sawai, H. Uchiyama, T. Muranaka, K. Saito, Draft genome assembly and annotation of *Glycyrrhiza uralensis*, a medicinal legume, *Plant J.* 89 (2017) 181–194, <https://doi.org/10.1111/tj.13385>.

- [30] M. Pertea, G.M. Pertea, C.M. Antonescu, T.C. Chang, J.T. Mendell, S.L. Salzberg, StringTie enables improved reconstruction of a transcriptome from RNA-seq reads, *Nat. Biotechnol.* 33 (3) (2015) 290–295, <https://doi.org/10.1038/nbt.3122>.
- [31] B. Buchfink, C. Xie, D.H. Huson, Fast and sensitive protein alignment using DIAMOND, *Nat. Methods* 12 (1) (2015) 59–60, <https://doi.org/10.1038/nmeth.3176>.
- [32] Y. Liao, G.K. Smyth, W. Shi, featureCounts: an efficient general purpose program for assigning sequence reads to genomic features, *Bioinformatics* 30 (7) (2014) 923–930, <https://doi.org/10.1093/bioinformatics/btt656>.
- [33] M.I. Love, W. Huber, S. Anders, Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2, *Genome Biol.* 15 (12) (2014) 550, <https://doi.org/10.1186/s13059-014-0550-8>.
- [34] H. Varet, L. Brillet-Gueguen, J.Y. Coppee, M.A. Dillies, SARTools: a DESeq2- and EdgeR-Based R pipeline for comprehensive differential analysis of RNA-Seq data, *PLoS One* 11 (6) (2016) e0157022, <https://doi.org/10.1371/journal.pone.0157022>.
- [35] M. Ashburner, C.A. Ball, J.A. Blake, D. Botstein, H. Butler, J.M. Cherry, A.P. Davis, K. Dolinski, S.S. Dwight, J.T. Eppig, M.A. Harris, D.P. Hill, L. Issel-Tarver, A. Kasarskis, S. Lewis, J.C. Matese, J.E. Richardson, M. Ringwald, G.M. Rubin, G. Sherlock, Gene ontology: tool for the unification of biology. The Gene Ontology Consortium, *Nat. Genet.* 25 (1) (2000) 25–29, <https://doi.org/10.1038/75556>.
- [36] S.E. Bouhaddani, J. Houwing-Duistermaat, P. Salo, M. Perola, G. Jongbloed, H.W. Uh, Evaluation of O2PLS in omics data integration, *BMC Bioinform.* 17 (Suppl 2) (2016) 11, <https://doi.org/10.1186/s12859-015-0854-z>. Suppl 2.
- [37] M.A. Farag, A. Porzel, L.A. Wessjohann, Comparative metabolite profiling and fingerprinting of medicinal licorice roots using a multiplex approach of GC-MS, LC-MS and 1D NMR techniques, *Phytochemistry* 76 (2012 Apr) 60–72, <https://doi.org/10.1016/j.phytochem.2011.12.010>.
- [38] Y. Wu, Z. Wang, Q. Du, Z. Zhu, T. Chen, Y. Xue, Y. Wang, Q. Zeng, C. Shen, C. Jiang, L. Liu, H. Zhu, Q. Liu, Pharmacological effects and underlying mechanisms of licorice-derived flavonoids, *Evid. base Compl. Alternative Med.* (2022) 9523071, <https://doi.org/10.1155/2022/9523071>.
- [39] J. He, Y. Deng, L. Ren, Z. Jin, J. Yang, F. Yao, Y. Liu, Z. Zheng, D. Chen, B. Wang, Y. Zhang, G. Nan, W. Wang, R. Lin, Isoliquiritigenin from licorice flavonoids attenuates NLRP3-mediated pyroptosis by SIRT6 in vascular endothelial cells, *J. Ethnopharmacol.* 303 (2023), <https://doi.org/10.1016/j.jep.2022.115952>. Article 115952.
- [40] L. Yang, Y. Jiang, Z. Zhang, J. Hou, S. Tian, Y. Liu, The anti-diabetic activity of licorice, a widely used Chinese herb, *J. Ethnopharmacol.* 263 (2020) 113216, <https://doi.org/10.1016/j.jep.2020.113216>.
- [41] X. Yu, Y. Bao, X. Meng, S. Wang, T. Li, X. Chang, W. Xu, G. Yang, T. Bo, Multi-pathway integrated adjustment mechanism of licorice flavonoids presenting anti-inflammatory activity, *Oncol. Lett.* 18 (5) (2019) 4956–4963, <https://doi.org/10.3892/ol.2019.10793>.
- [42] T. Li, S. Hua, J. Ma, L. Dong, F. Xu, X. Fu, Spectrum-effect relationships of flavonoids in *Glycyrrhiza uralensis* Fisch, *Journal of Analytical Methods in Chemistry* (2020) 8838290, <https://doi.org/10.1155/2020/8838290>.
- [43] S. Wahab, S. Annadurai, S.S. Abullais, G. Das, W. Ahmad, M.F. Ahmad, G. Kandasamy, R. Vasudevan, M.S. Ali, M. Amir, *Glycyrrhiza glabra* (licorice): a comprehensive review on its phytochemistry, biological activities, clinical evidence and toxicology, *Plants* 10 (12) (2021) 2751, <https://doi.org/10.3390/plants10122751>.
- [44] S. Li, Transcriptional control of flavonoid biosynthesis: fine-tuning of the MYB-bHLH-WD40 (MBW) complex, *Plant Signal. Behav.* 9 (1) (2014) e27522, <https://doi.org/10.4161/psb.27522>.
- [45] Y.H. Liao, J.H. Lu, N. Li, J.Z. Zhang, X.Y. Li, Anatomical Structure, Total flavonoids histochemical localization and content comparization of vegetative organs in *Glycyrrhiza glabra* L., *Acta Bot. Boreali Occident. Sin.* 12 (2010) 2406–2411, <https://doi.org/10.3724/SP.J.1142.2010.40521>.
- [46] G. Agati, E. Azzarello, S. Pollastri, M. Tattini, Flavonoids as antioxidants in plants: location and functional significance, *Plant Sci.* 196 (2012) 67–76, <https://doi.org/10.1016/j.plantsci.2012.07.014>.
- [47] M.L.F. Ferreyra, P. Serra, P. Casati, Recent advances on the roles of flavonoids as plant protective molecules after UV and high light exposure, *Physiol. Plantarum* 173 (3) (2021) 736–749, <https://doi.org/10.1111/ppl.13543>.
- [48] C. Martin, Y. Zhang, R. De Stefano, M. Robine, E. Butelli, K. Bulling, L. Hill, M. Rejzek, H.J. Schoonbeek, Different ROS-scavenging properties of flavonoids determine their abilities to extend shelf life of tomato, *Free Radic. Biol. Med.* 86 (2015) S11, <https://doi.org/10.1016/j.freeradbiomed.2015.07.048>.