



OPEN

Comparative analysis of tuberous root metabolites between cultivated and wild varieties of *Rehmannia glutinosa* by widely targeted metabolomics

Yanqing Zhou^{1✉}, Luying Shao^{1,2}, Jialin Zhu^{1,2}, Huimin Li¹ & Hongying Duan¹

Differential metabolites between tuberous roots from cultivated variety (ZP) and wild variety (YS) of *Rehmannia glutinosa* were analyzed by widely targeted metabolomics, and annotated to KEGG pathways. 228 secondary metabolites (SM) in ZP and YS were detected, of which 58 were differential metabolites (DM), including 41 flavonoids, 10 phenolic acids, 3 terpenoids, 2 alkaloids and 2 others, and 170 were unchanged; Among 58 DMs, 44 (75.9%) were up-regulated in YS, of which 30 were unique to YS, while 14 (24.1%) were down-regulated in YS, of which 10 were unique to ZP; Among flavonoids, 33 (80.5%) were more highly expressed in YS than in ZP; Among phenolic acids, 7 (70%) were more highly expressed in YS than in ZP; 12 of 58 DMs were annotated into 17 types of KEGG pathways. Among them, benzoic acid and p-Coumaryl alcohol were up-regulated in YS, and annotated into 10 pathways (58.8%) and 4 pathways (23.5%), respectively. In addition, much of DMs possess various pharmacological effects. These results indicated better quality of YS than ZP and the necessity of YS domestication. Taken together, this study will provide a reference for the scientific introduction, comprehensive development and utilization of wild *Rehmannia glutinosa*.

Rehmannia glutinosa Libosch. (*R. glutinosa*) is a perennial herb belonging to the genus *Rehmannia* (Scrophulariaceae), and has important economic value as medicinal and food materials. Its tuberous root is a commonly used bulk Chinese herbal medicine. The pharmaceutical effects of *R. glutinosa* are closely related to its secondary metabolites. Many metabolites in its tuberous root have been identified up to date. For example, 1049 metabolites were identified from the developing tuberous root of its variety Jinjiu¹. They are composed of catalpol, acteoside, saccharides, terpene glycosides, amino acids, trace elements and other components, which possesses some bioactivities such as anti-cardiovascular diseases, nerve protection, hypertension resistance, immunity enhancement and so on^{2–4}. *R. glutinosa* is mainly distributed in some provinces, such as Henan, Shanxi, Shandong and others, in China. Among them, better *Rehmanniae Radix* is from cultivated *R. glutinosa* in Huaiqing region including Wen County, Wuzhi County, Boai County and Jiaozuo city, Henan, which has higher contents of bioactive components such as catalpol and verbascoside, and lower clinical dosage than that from other places⁵. Its commonly cultivated varieties are Beijing No.3, Wen 85-5, Jinjiu, etc. In recent years, with its increasing consumption and loss of its farmland, its resources become increasingly exhausted. Moreover, its variety complexity and long-term vegetative propagation make its varieties degenerated, which lead to its poor quality, low yield and narrow genetic basis⁶. Therefore, it urgently requires new germplasms. To our knowledge, one practical solution to its new germplasms is the introduction, comprehensive development and utilization of wild *R. glutinosa*.

Wild *R. glutinosa* resources with many good traits and genes are very important and significant for enriching *R. glutinosa* germplasm resources and improving the yield and quality of cultivated *R. glutinosa* varieties, but their qualities vary. As a result, they need be identified and evaluated before their development and utilization. SNP-based *R. glutinosa* germplasms analysis indicated that there were more significant differences of its cultivated varieties from its wild resources⁷. Moreover, compared with wild *Rehmannia* germplasms, the qualities of cultivated *R. glutinosa* varieties decrease in that their some genes were lost via long-term artificial selection. Accurate chemical composition analysis of *R. glutinosa* varieties is crucial to their quality evaluation. Therefore,

¹College of Life Sciences, Henan Normal University, Xinxiang 453007, Henan, People's Republic of China. ²These authors contributed equally: Luying Shao and Jialin Zhu. ✉email: yqzhou@htu.cn

Variety	Wild type	Beijing No.3
Elevation	184 m	97.10 m
Root number per plant	2–4	4.00
Fresh weight of the heaviest root per plant (g)	2.10	100
Growing period (year)	Natural growth for many years	Vegetative reproduction for many years
Growing environment	Natural mountain soil	Cultivated soil
Tuberous root color	Dark yellow	Bright yellow

Table 1. Phenotypic analysis of two varieties.

the study of chemical composition of wild and cultivated *R. glutinosa* will help to solve the lack of its new germ-plasms. Metabolomics, an ongoing and practical technology, has been widely used in medicine development, medicine toxicity and mechanism research, medicine screening and efficacy evaluation, mining new secondary metabolites, plant metabolism and response mechanisms, microbial interactions, gene function elucidation, effective metabolic pathways and related regulatory mechanisms^{1,8–14}. At present, metabolomic technologies include targeted metabolomics, untargeted metabolomics and widely targeted metabolomics. So far, the first two metabolomic technologies have widely been used in plants, including *R. glutinosa*^{15–20}, but both have advantages and disadvantages. Widely targeted metabolomics is next generation metabolomics, combining the advantages of untargeted metabolomics and targeted metabolomics, and possesses some advantages such as qualitative and quantitative accuracy, high throughput, high sensitivity and wide coverage. Using Q-TRAP mass spectrometry based on MRM mode, hundreds of known metabolites and nearly a thousand unknown metabolites can be simultaneously quantified, and the detection and identification of highly sensitive and widely targeted metabolites come true^{21,22}. In recent years, this technology has been successfully applied in *Sesame*²³, *Rice*^{22,24,25}, *Vanilla*²⁶ and *Chrysanthemum morifolium*²⁷, respectively.

In our study, widely targeted metabolomics was used to compare the metabolic profiles between ZP and YS of *R. glutinosa* for the first time in order to qualitatively and quantitatively identify differential metabolites between ZP and YS. The results of this study will provide a reference for the scientific introduction, comprehensive development and utilization of wild *R. glutinosa*.

Results

Phenotypic analysis. There were some significant differences between ZP and YS in phenotype before the propagation of wild *R. glutinosa* (Table 1). However, the metabolite profiles of these two varieties were unknown and needed further analysis in this study.

Multivariate analysis. Multivariate analysis is the analysis of three or more variables. It is used to deal with the relationship between variables, including different methods such as Principal components analysis (PCA). Orthogonal Partial Least Squares-Discriminant Analysis (OPLS-DA) and so on. PCA and OPLS-DA were used in this study.

PCA. PCA was used to reveal the internal structure of several variables by a few principal components. In the PCA plot (Fig. 1), three mix samples as QC were grouped together, suggesting QC samples with similar metabolic profiles and the entire analysis with stability and repeatability. Meanwhile, it could be seen from Fig. 1 that YS samples were clustered together, located to the left side of the QC samples, and that ZP samples gathered together, located on the right side to the QC samples. These results indicated that our analysis was stable and repeatable.

OPLS-DA. OPLS-DA is a multivariate statistical analysis method with supervised pattern recognition, and can solve the problem that PCA is not sensitive to the variables with little correlation. According to the differential variables, the score plot of every component (Fig. 2a) was formed to further show the differences between the components²⁸. It was seen from Fig. 2a that R^2Y and $Q^2 Y(Q^2)$ were 1, while R^2X equals 0.976, suggesting that OPLS-DA should be stable and reliable. Because Q^2 equals 1, more than 0.9, OPLS-DA is excellent. The OPLS-DA model is verified using 200 alignment experiments. The horizontal line corresponds to the R^2Y and Q^2 of the original model, while the red and blue dots represent R^2Y and Q^2 after replacement, respectively. R^2Y (0.81) and Q^2 (0.52) in Fig. 2b were smaller than R^2Y (1) and Q^2 (1) of the original model, suggesting that the corresponding points should not exceed the corresponding lines. Therefore, OPLS-DA is meaningful, from which variable influence in projection (VIP) values are obtained. VIP value is used to screen differential metabolites.

Identification of differential metabolites. A total of 228 SMs between YS and ZP were detected by HMDB, MWDB and METLIN databases (see Supplementary Table S1 online). Using both $FC \geq 2$ or ≤ 0.5 and $VIP \geq 1$ as the screening standards, 58 DMs were screened and identified (Table 2).

Display of the difference data between groups by Volcano plot. Based on $\log_2 FC$ value and p-value, DMs were displayed in Volcano plot (Fig. 3a). In Fig. 3a, there were 44 up-regulated metabolites indicating that their expression contents in YS were higher than that in ZP, 14 down-regulated metabolites indicat-

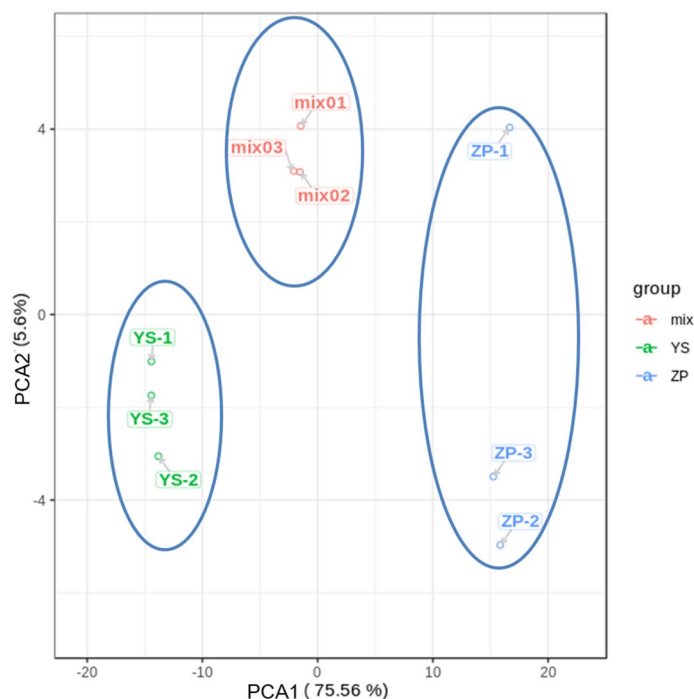


Figure 1. PCA score chart based on mass spectrum data of ZP, YS and QC samples. Ordinate: the second principal component, Abscissa: the first principal component. PCA Plot visualized in ggplot2 toolkit of R v3.5.0 (<https://www.r-project.org/>).

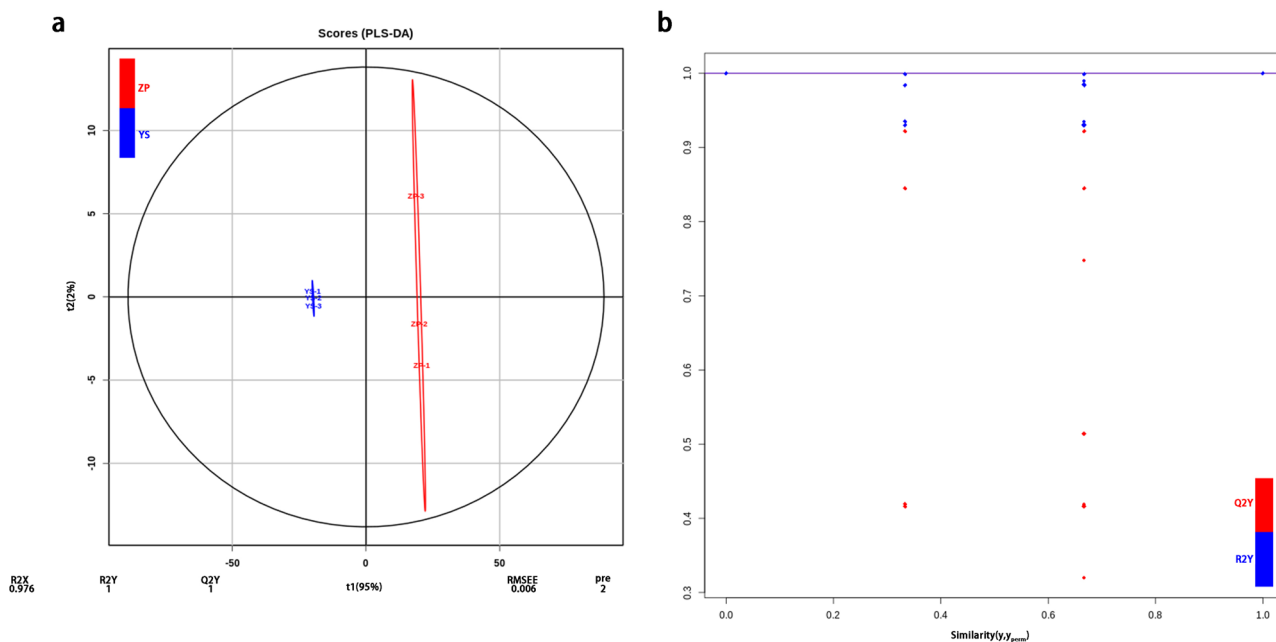


Figure 2. (a) OPLS-DA score Plot. t1: Predicted principal components-score value of main components and difference between observation groups, t01: Orthogonal principal component-score value of orthogonal components and difference in observation group, R²Y: Percentage of Y matrix information that can be explained by OPLS-DA, Q²Y (Q²): Prediction ability of OPLS-DA, Pre: Predicted principal component number, RMSEE: Root mean squared error, Red: ZP, Blue: YS. (b) OPLS-DA permutation verification plot. Red: Q²Y, Blue: R²Y. OPLS-DA score Plot visualized in ggplot2 toolkit of R v1.0.1 (<https://www.r-project.org/>).

No	Formula	Compounds	VIP	Fold change	p-value	Type
1	C ₂₉ H ₃₄ O ₁₆	Limocitrin-O-rhamnoside-O-rhamnoside	1.329	102.477	0.002	Up
2	C ₃₁ H ₄₈ O ₇	Phytolaccagenin	1.929	0.000	0.000	Down
3	C ₁₄ H ₁₉ NO ₇	Ehretioside	1.895	12,259.259	0.000	Up
4	C ₁₆ H ₁₂ O ₆	Aracarpene 1	1.952	21,777.778	0.004	Up
5	C ₂₇ H ₂₈ O ₁₇	Kaempferol 3-glucuronide-7-glucoside	1.857	8529.630	0.008	Up
6	C ₂₇ H ₂₆ O ₁₇	Apigenin-7-O-diglucuronide	2.074	79,666.667	0.005	Up
7	C ₁₇ H ₁₄ O ₇	Tricin	1.681	1644.444	0.001	Up
8	C ₂₇ H ₃₀ O ₁₅	Kaempferol glc-rha	1.926	16,814.815	0.007	Up
9	C ₁₆ H ₁₂ O ₆	6,7,8-Tetrahydroxy-5-methoxyflavone	1.852	8066.667	0.004	Up
10	C ₂₁ H ₁₈ O ₁₁	Baicalin	1.166	34.653	0.004	Up
11	C ₁₆ H ₁₂ O ₆	Diosmetin	1.950	21,444.444	0.002	Up
12	C ₂₇ H ₃₁ ClO ₁₅	Pelargonin chloride	1.804	0.000	0.007	Down
13	C ₉ H ₁₀ O ₂	p-Coumaryl alcohol	1.028	15.936	0.000	Up
14	C ₂₀ H ₃₉ NO ₂	N-Oleylethanolamine	1.035	16.587	0.000	Up
15	C ₃₃ H ₄₀ O ₁₉	Robinin(kaempferol-3-O-gal-rham-7-O-rham)	1.832	0.000	0.170	Down
16	C ₃₁ H ₃₈ O ₁₆	2'-Acetylacteoside	1.195	0.023	0.029	Down
17	C ₁₆ H ₂₂ O ₁₀	Geniposidic acid	1.329	0.010	0.001	Down
18	C ₂₈ H ₃₂ O ₁₅	Diosmin	1.234	47.068	0.000	Up
19	C ₉ H ₁₂ O ₃	Homovanillic alcohol	1.683	1718.519	0.017	Up
20	C ₂₂ H ₂₃ O ₁₁ +	Peonidin O-hexoside	1.804	5107.407	0.006	Up
21	C ₂₃ H ₂₄ O ₁₃	Syringetin 3-O-hexoside	1.125	0.036	0.002	Down
22	C ₂₈ H ₃₂ O ₁₆	Chrysoeriol O-hexosyl-O-hexoside	1.836	6937.037	0.004	Up
23	C ₂₆ H ₂₆ O ₁₅	Tricin O-malonylhexoside	1.896	12,407.407	0.000	Up
24	C ₂₃ H ₂₄ O ₁₂	Tricin 7-O-hexoside	1.164	33.997	0.001	Up
25	C ₁₇ H ₂₂ O ₁₀	1-O-β-D-Glucopyranosyl sinapate	1.910	0.000	0.012	Down
26	C ₂₈ H ₃₂ O ₁₅	Chrysoeriol 7-O-rutinoside	1.902	13,111.111	0.002	Up
27	C ₂₂ H ₂₀ O ₁₂	Chrysoeriol O-glucuronic acid	1.440	228.916	0.005	Up
28	C ₂₃ H ₂₂ O ₁₃	Tricin O-glucuronic acid	2.081	85,814.815	0.006	Up
29	C ₂₂ H ₂₈ O ₁₃	3-O-p-coumaroyl quinic acid O-hexoside	1.604	859.259	0.006	Up
30	C ₂₂ H ₂₆ O ₁₂	5-O-p-Coumaroyl shikimic acid O-hexoside	1.749	3048.148	0.005	Up
31	C ₁₅ H ₁₀ O ₆	Luteolin	1.521	452.222	0.037	Up
32	C ₇ H ₆ O ₂	Benzoic acid	1.937	18,703.704	0.002	Up
33	C ₂₁ H ₂₀ O ₁₀	Kaempferol 7-O-rhamnoside	1.708	0.000	0.196	Down
34	C ₇ H ₆ O ₃	Protocatechuic aldehyde	1.918	15,407.407	0.003	Up
35	C ₂₇ H ₃₀ O ₁₄	Kaempferol 3,7-dirhamnoside (kaempferitrin)	2.059	0.000	0.151	Down
36	C ₂₂ H ₂₃ O ₁₁ +	Peonidin 3-O-glucoside	1.797	4762.963	0.003	Up
37	C ₂₇ H ₃₀ O ₁₆	Bioquercetin	1.571	0.001	0.046	Down
38	C ₂₇ H ₃₀ O ₁₄	Kaempferol-3,7-O-α-L-rhamnoside	1.872	0.000	0.108	Down
39	C ₂₃ H ₃₂ O ₁₅	β-D-Furanofructosyl-α-D-(3-mustard cyl)glucoside	1.029	15.965	0.000	UP
40	C ₁₆ H ₂₄ O ₇	3-Hydroxy-4-isopropylbenzylalcohol 3-glucoside	1.986	31,074.074	0.001	Up
41	C ₂₁ H ₁₈ O ₁₁	Apigenin-7-O-β-D-glucuronide	1.220	49.046	0.006	Up
42	C ₂₁ H ₁₈ O ₁₂	Tetrahydroxy-flavone-7-O-β-D-glucuronide	2.006	38,074.074	0.000	Up
43	C ₁₆ H ₁₂ O ₆	Hispidulin	1.854	8237.037	0.003	Up
44	C ₁₇ H ₁₄ O ₇	Jaceosidin	1.632	1081.111	0.002	Up
45	C ₂₁ H ₁₈ O ₁₂	Scutellarin	2.008	39,037.037	0.001	Up
46	C ₂₄ H ₂₂ O ₁₃	Malonylgenistin	1.505	40.364	0.000	Up
47	C ₁₆ H ₁₂ O ₆	Pratensein	1.856	8407.407	0.004	Up
48	C ₂₅ H ₁₈ O ₉	Luteolin-7-O-glucuronide	1.608	60.303	0.003	Up
49	C ₂₂ H ₂₂ O ₁₁	Diosmetin-7-O-galactoside	1.016	14.960	0.001	Up
50	C ₂₂ H ₂₀ O ₁₂	Diosmetin-7-O-glucuronide	1.388	156.084	0.004	Up
51	C ₂₅ H ₂₄ O ₁₄	Diosmetin-7-O-(6'-O-malonyl)-β-D-glucoside	1.672	1533.333	0.003	Up
52	C ₂₇ H ₃₀ O ₁₅	Luteolin-7-O-rutinoside	1.873	0.000	0.165	Down
53	C ₂₂ H ₃₀ O ₁₄	3'-O-D-glucosylgentiopicoside	1.694	1862.963	0.003	Up
54	C ₁₅ H ₂₂ O ₈	Bartsioside	1.977	28,333.333	0.002	Up
55	C ₃₀ H ₃₈ O ₁₅	Cistanoside C	1.044	0.056	0.038	Down
56	C ₂₂ H ₃₄ NO ₁₀ +	Sinapine glucoside	1.828	0.000	0.315	Down

Continued

No	Formula	Compounds	VIP	Fold change	p-value	Type
57	C ₂₈ H ₂₈ O ₁₈	Chrysoeriol-7-O-[[β-D-glucuronopyranosyl-(1→2)-O-β-D-glucuronopyranoside]	2.067	74,185.185	0.013	Up
58	C ₂₇ H ₂₆ O ₁₇	Apigenin-7-O-[[β-D-glucuronopyranosyl(1→2)-O-β-D-glucuronopyranoside]	2.109	116,407.407	0.002	Up

Table 2. DMs from 228 SMs. Up: compared with ZP, the corresponding metabolite was up-regulated in YS. Down: compared with ZP, the corresponding metabolite was down-regulated in YS.

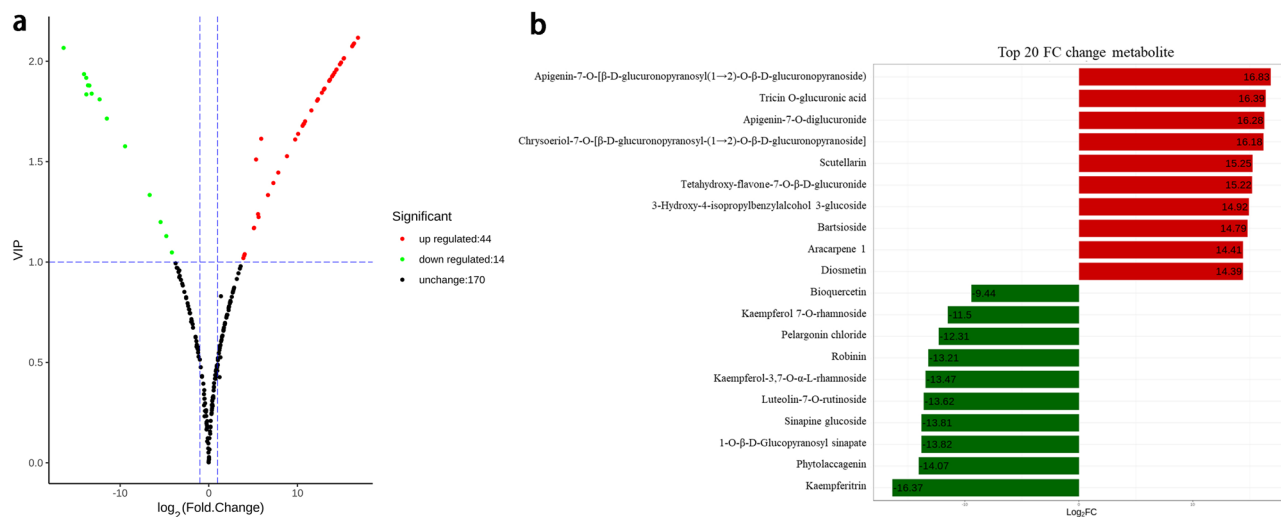


Figure 3. (a) Volcano Plot of differential metabolites. Abscissa: FC value, Ordinate: VIP value, The dots: differential metabolites, Green dots: down regulated metabolites, Red dots: up regulated metabolites, Black dots: detected metabolites without significant difference. Volcano Plot visualized in ggplot2 toolkit of R v3.5.0 (<https://www.r-project.org/>). (b) Top 20 FC change metabolites. Ordinate: metabolite, Blackish green color: down regulated metabolites, Red: up regulated metabolites.

ing that their expression contents in ZP were higher than that in YS, and 170 unchanged metabolites indicating that their expression contents did not vary between YS and ZP. Meanwhile, Top 20 FC change metabolites were presented in Fig. 3b. This result was consistent with that based on the VIP values and FCs (Table 2).

Heatmap clustering. In order to show the varying law of 58 DMs with significant differences, their heatmap was drawn (Fig. 4). The results showed that three YS repeats were grouped into one category and three ZP repeats were grouped into the other category. Because different metabolites had different accumulation trends in different samples, and the closer the accumulation trends, the closer the distances, 58 DMs had obvious expression differences between ZP and YS. 58 DMs were grouped into 5 clusters: (1) flavonoids including 41 ones, which were dominant in the SM of *R. glutinosa* tuberous roots. Among them, 33 (80.5%) metabolites were expressed higher in YS than in ZP. (2) 10 phenolic acids, of which 7 (70%) metabolites were expressed higher in YS than in ZP. (3) 3 terpenoids, 1 (33.33%) of which was expressed higher in YS than in ZP. (4) 2 alkaloids, half of which was higher expressed in YS than in ZP. (5) 2 metabolites that were higher expressed in YS than in ZP. In total, 44 differential metabolites (75.9%) were higher expressed in YS than in ZP, which was consistent with the screening of DMs based on VIP and FC as well as by Volcano plot (Table 2, Fig. 3b).

Pharmaceutical activities of DMs suggesting quality changes. It was seen from Fig. 4 that DMs were grouped into 5 types such as flavonoids, phenolic acids, terpenoids, alkaloids and and others. Among them, flavonoids including luteolin, protocatechuic aldehyde and triclin, were a large family of plant secondary metabolites and a subdivision of polyphenols, a versatile class of natural compounds that represented secondary metabolites from higher plants. Flavonoids were an effective ingredient of many Chinese herbal medicines and had various medicinal effect, including antibacterial, anti-inflammatory, anti-oxidant^{29,30}. Phenolic acids including benzoic acid and p-Coumaryl alcohol, had been proved to have a variety of pharmacological activities, such as cardiovascular and cerebrovascular effects, anti-tumor, anti-oxidation, anti-inflammation, anti-fibrosis, etc³¹. Terpenoids including phytolaccagenin, geniposidic acid, were also a kind of important compounds in Chinese herbal medicine and played an important role in plant growth and development, resistance and defense, etc³². Alkaloids including ehretioside were a kind of nitrogen-containing basic organic compounds existing in nature (mainly plants). Alkaloids had many pharmacological activities, such as analgesia, spasmolysis, anti-inflammation, anti-tumor, etc. In addition, some compounds in alkaloids could interact with chemical components such

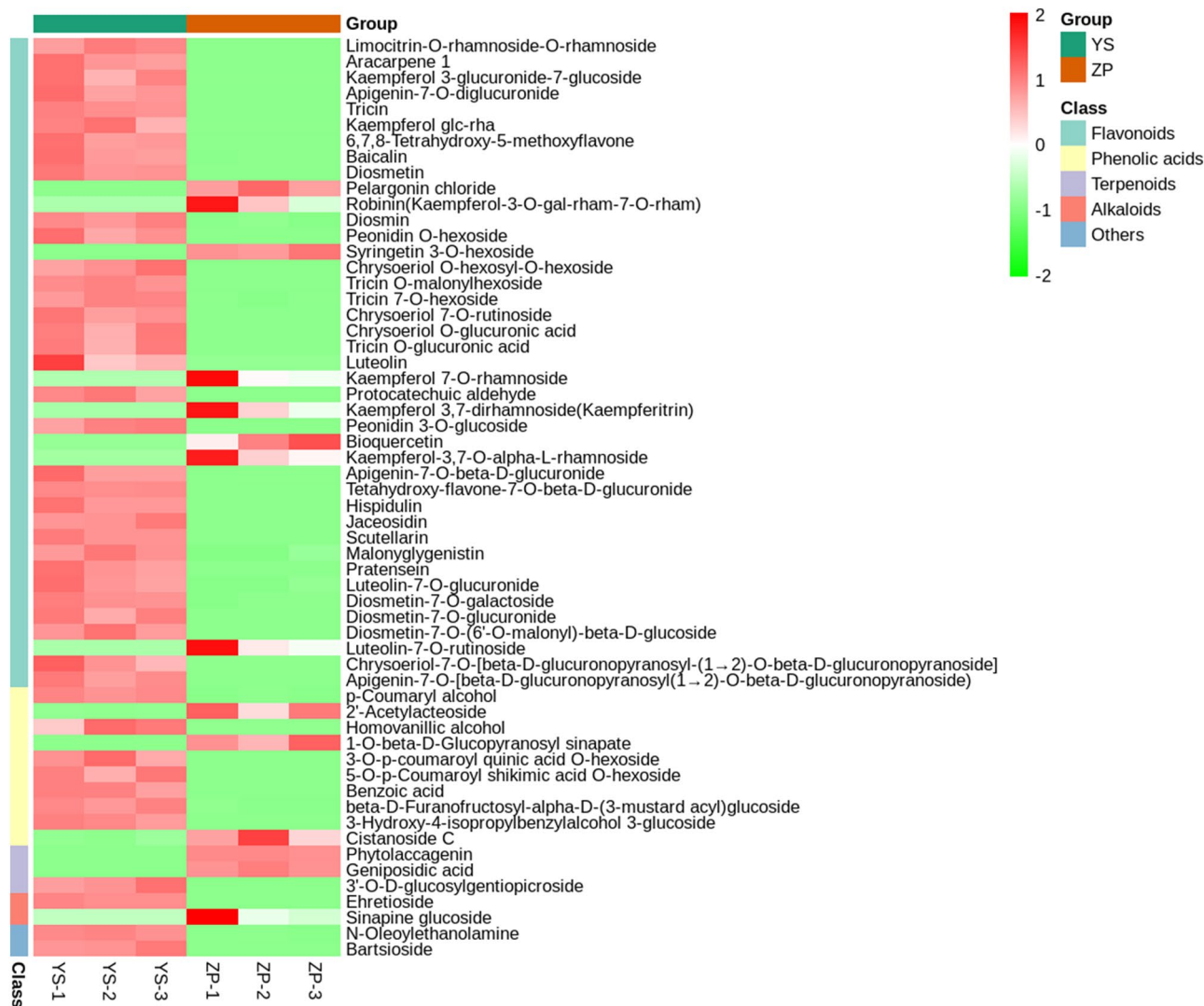


Figure 4. Heat map of 58 DMs. Ordinate: clustering after normalization of relative contents of different metabolites, 5 kinds of colors stand for 5 classes of metabolites, Abscissa: grouping of samples, 2 kinds of colors stand for 2 groups of samples, Color scale from green to red indicated the contents of differentially expressed metabolites vary from low to high. Heatmap visualized in ggplot2 toolkit of R v1.0.12 (<https://www.r-project.org/>).

as saponins and terpenoids, which could achieve the pharmacological effects of relieving cough, eliminating phlegm and relieving asthma³³.

Differential metabolic pathways. When 228 SMs were annotated to KEGG pathways by KEGG database (<https://www.kegg.jp/kegg/kegg1.html>)³⁴, 38 SMs were annotated to KEGG pathways 202 times. Their annotation times ranged from 1 times per metabolite to 33 times per metabolite (succinic acid). After some duplicate KEGG pathways were removed, there were still 77 KEGG pathways (see Supplementary Table S2 online). Take Phenylpropanoid biosynthesis of KEGG pathway as an example (Fig. 5). In Fig. 5, p-Coumaric acid, p-coumaryl alcohol, caffeic acid, ferulic acid, coniferyl aldehyde, coniferyl alcohol, p-coumaroyl quinic acid and coniferin were annotated in Phenylpropanoid biosynthesis, of which p-coumaryl alcohol was a differential metabolite up-regulated in YS, compared with ZP, but p-Coumaric acid, caffeic acid, ferulic acid, coniferyl aldehyde, coniferyl alcohol, p-coumaroyl quinic acid and coniferin were unchanged.

Among 58 DMs, 12 were annotated to KEGG database, 7 of which were annotated to KEGG pathways 23 times (see Supplementary Table S3 online). After some duplicate KEGG pathways were removed, there were still 17 KEGG pathways (Fig. 5). The seven metabolites were pelargonin chloride, peonidin 3-O-glucoside, p-Coumaryl alcohol, luteolin, benzoic acid, protocatechuic aldehyde and pratensein, respectively. These 17 KEGG pathways were categories anthocyanin biosynthesis, phenylpropanoid biosynthesis, biosynthesis of phenylpropanoids, metabolic pathways, biosynthesis of secondary metabolites, flavonoid biosynthesis, flavone and flavonol biosynthesis, phenylalanine metabolism, benzoate degradation, dioxin degradation, toluene degradation, aminobenzoate degradation, biosynthesis of alkaloids derived from shikimate pathway, microbial metabolism in

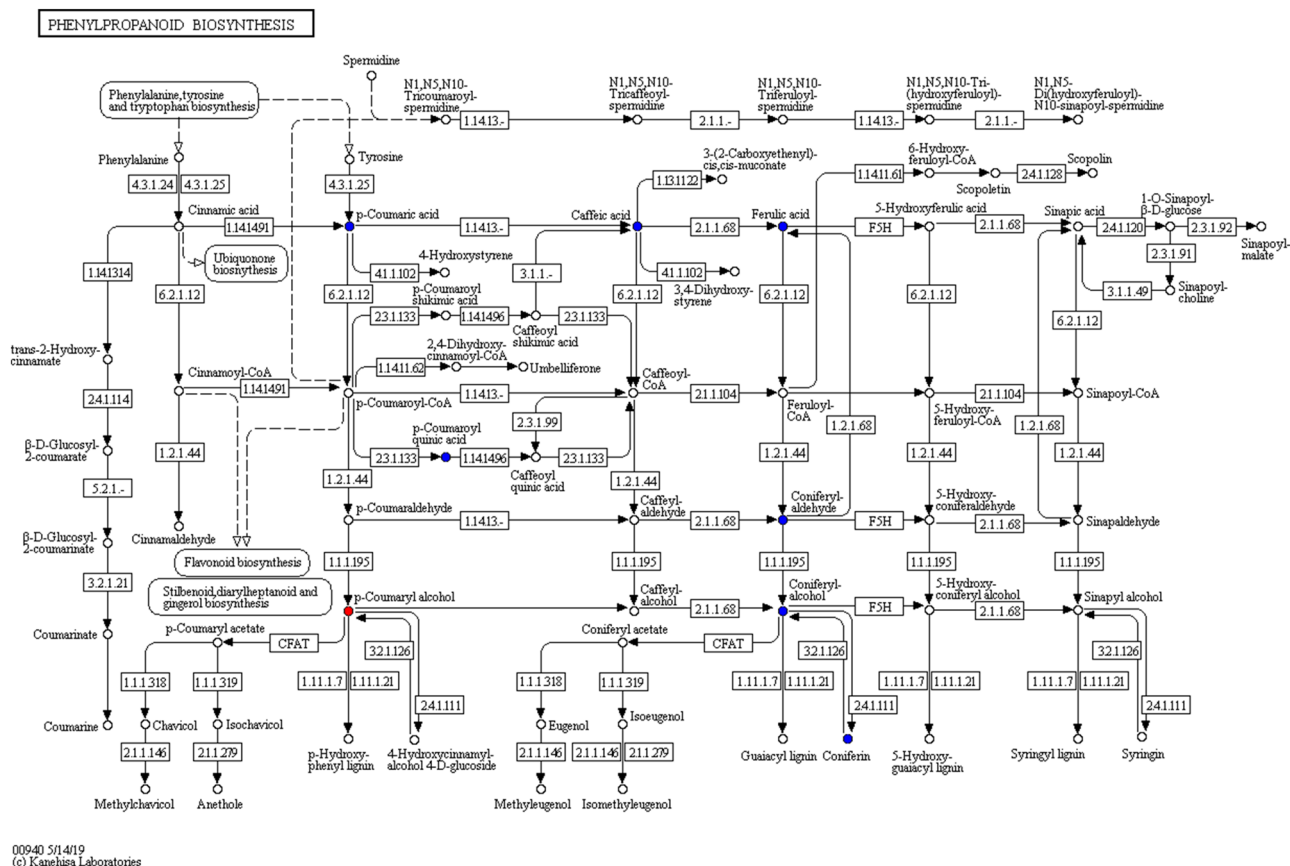


Figure 5. Phenylpropanoid biosynthesis. The red dots: the differentially expressed metabolites that are increased, the blue dots: the detected metabolites, but there is no significant difference.

diverse environments, degradation of aromatic compounds, isoquinoline alkaloid biosynthesis and isoflavonoid biosynthesis. Among them, phenylpropanoid biosynthesis and biosynthesis of phenylpropanoids were different metabolic pathway names in KEGG, which had similar names, but corresponded to two different pathway IDs (Table 3).

Discussion

Rehmannia glutinosa is an important perennial herb, and its tuberous roots is clinically used to treat fever, nervous conditions, diabetes, and hypertension, and to increase liver function, hematopoietic function and immune defense and so on in different ways¹. The great medicinal market demand for *R. glutinosa* leads to its overexploitation and germplasm resource scarcity, and long-term vegetative propagation makes the variety degenerate, quality poor, yield low and genetic basis narrow, so it is urgent to enrich and improve *R. glutinosa* by using candidate germplasm resources. Wild *R. glutinosa* is a kind of important candidate resources for this objective. Before we make good use of its wild resources, we used widely targeted metabolomics to compare the SM between its cultivated and wild varieties in that its SM are vital for its important pharmacological activities.

On the one hand, we compared the phenotypic characteristics of ZP and YS. It was found that both differences were very obvious, especially in both sizes and weights (Table 1). This was consistent with that of Zuo et al.³⁵. On the other hand, 228 SMs were detected in both YS and ZP using widely targeted metabolomics technology coupled with MRM and public databases (see Supplementary Table S1 online). They were divided into 170 unchanged metabolites and 58 DMs (44 upregulated, 14 downregulated) (Table 2, Fig. 3b). 58 DMs were divided into 5 types such as flavonoids (41), phenolic acids (10), terpenoids (3), alkaloids (2) and others (2) (Table 2).

It is known that plant metabolomes are composed of over 200,000 metabolites that control plant development, and even *Arabidopsis* contains some 5000 metabolites¹, so *R. glutinosa* should contain more than these 228 SMs in YS and ZP of *R. glutinosa* Libosch. Moreover, 1049 metabolites were ever identified in the developing tuberous roots of *R. glutinosa* variety Jinjiu³⁶. Therefore, the high throughout identification of widely targeted metabolomics used in this study was limited, compared with untargeted metabolomics³⁶ and others^{37,38}. Its limitation should be attributed to the lack of large herbal medicine public metabolite databases.

The SMs of medicinal plants are very important substances in their life activities, closely related to their defense against diseases and insect pests and environmental stress, and important pharmacological activities. Based on these, our DMs may reflect the differences between wild and cultivated varieties of *R. glutinosa*. In this study, (1) the number of upregulated metabolites (44) were bigger than that of downregulated metabolites (14) in YS, compared to ZP (Table 2, Fig. 3b), Among 41 flavonoids, 33 (80.5%) were more highly expressed in

KEGG pathway	Ko-ID	Number	DM	Cpd-ID
Anthocyanin biosynthesis	ko00942	2	Pelargonin chloride; peonidin 3-O-glucoside	C08725 + C12141
Phenylpropanoid biosynthesis	ko00940	1	p-Coumaryl alcohol	C02646
Biosynthesis of phenylpropanoids	ko01061	1	p-Coumaryl alcohol	C02646
Metabolic pathways	ko01100	3	Luteolin; p-coumaryl alcohol; benzoic acid	C01514 + C02646 + C00180
Biosynthesis of secondary metabolites	ko01110	3	Benzoic acid; luteolin; p-coumaryl alcohol	C00180 + C01514 + C02646
Flavonoid biosynthesis	ko00941	1	Luteolin	C01514
Flavone and flavonol biosynthesis	ko00944	1	Luteolin	C01514
Phenylalanine metabolism	ko00360	1	Benzoic acid	C00180
Benzoate degradation	ko00362	1	Benzoic acid	C00180
Dioxin degradation	ko00621	1	Benzoic acid	C00180
Toluene degradation	ko00623	1	Benzoic acid	C00180
Aminobenzoate degradation	ko00627	1	Benzoic acid	C00180
Biosynthesis of alkaloids derived from shikimate pathway	ko01063	2	Benzoic acid; protocatechuic aldehyde	C00180 + C16700
Microbial metabolism in diverse environments	ko01120	1	Benzoic acid	C00180
Degradation of aromatic compounds	ko01220	1	Benzoic acid	C00180
Isoquinoline alkaloid biosynthesis	ko00950	1	Protocatechuic aldehyde	C16700
Isoflavonoid biosynthesis	ko00943	1	Pratensein	C10520

Table 3. The categories of 12 DMs-annotated KEGG pathways. Ko-ID = ID of KEGG pathway, Number = the number of metabolites that can be annotated to the corresponding KEGG pathways, Cpd-ID = Number of compound in KEGG.

YS than ZP, while among 10 phenolic acids, 7 (70%) were more highly expressed in YS than ZP. The number of unique metabolites to YS (30) was bigger than that of unique metabolites to ZP (10). Because these DMs possess important effects, these results suggested that the quality of YS is better than that of ZP. For example, flavonol had antioxidation; luteolin had a variety of pharmacological effects such as anti-tumor, antibacterial, anti-inflammatory, antiviral and analgesic effects³⁹. Anthocyanin, a water-soluble flavonoid compound, had a variety of biological functions such as protecting plants from UV damage, scavenging reactive oxygen species, resisting adversity and changing the color of plants, as well as health efficacy such as anti-aging, anti-obesity and prevention of cardiovascular diseases⁴⁰. According to investigation and research, flavonoids were synthesized by using intermediate products from phenylalanine converted via phenylpropane route as synthetic precursors, which were formed by different synthetic routes. Due to their various pharmacological effects such as cardiovascular protection, anti-cancer, anti-oxidation, anti-inflammation, liver protection and anti-tumor etc. flavonoids had become hot spots in the development and research of natural medicine at home and abroad^{27,41,42}. According to our results, the expression of flavonoids and phenolic acids in YS was much higher; (2) Among quality control standards of *R. glutinosa* such as catalpol and rehmannioside D (now) or verbascoside (ever) in the Pharmacopoeia of the People's Republic of China⁴³, rehmannioside D, catalpol and verbascoside were all contained in unchanged. These result showed that YS could be used to modify ZP, of which catalpol' result is consistent with a previous report⁴⁴; (3) In this study, as the autotoxic metabolites such as ferulic acid, benzoic acid, protocatechuic aldehyde, 4-Hydroxybenzoic acid and so on, their the contents in YS were also much higher than that in ZP. Among them, benzoic acid had antibacterial and antiseptic effects⁴⁵, ferulic acid could inhibit obesity and improve the steady state of blood lipid and blood sugar^{46,47}, and protocatechuic aldehyde had cardiovascular and cerebrovascular protective effects⁴⁸. Moreover, 30 kinds of DMs such as protocatechuic aldehyde, luteolin, tricin, diosmetin, homovanillic alcohol, jaceosidin, pratensein, hispidulin, malonyglygenistin, bartsioside, scutellarin and other compounds were unique to YS and reported for the first time in *R. glutinosa*. These results indicate ZP and YS contained unique DMs, unique ones to YS were much more than that to ZP, and that artificial breeding increased the contents of main active metabolites in cultivated variety of *R. glutinosa* and selected out many metabolite in its wild variety.

In addition, KEGG database helps researchers to study genes, expression information and metabolite content as a whole network, and provides integrated metabolic pathways involved in such as the pathways of carbohydrate, nucleoside and amino acids as well as biodegradation of organic compounds with enzymes. Therefore, it is a powerful tool for metabolism analysis and metabolic network research in vivo^{1,49}. In the present study, 12 of 58 DMs between ZP and YS were annotated to 17 non-repetitive KEGG pathways (Fig. 6, Table 3). The main differential metabolic pathways between ZP and YS included metabolic pathways and biosynthesis of secondary metabolites. Among them, benzoic acid and p-Coumaryl alcohol up-regulated in YS were annotated into 10 (58.8%) and 4 (23.5%), respectively (Table 3). These results provided a clue for analyzing the metabolism of these metabolites and DMs, and their metabolic networks in *R. glutinosa*. In conclusion, based on the phenotypic differences between YS and ZP, we detected 228 SMs of YS and ZP by using widely targeted metabolomics, and

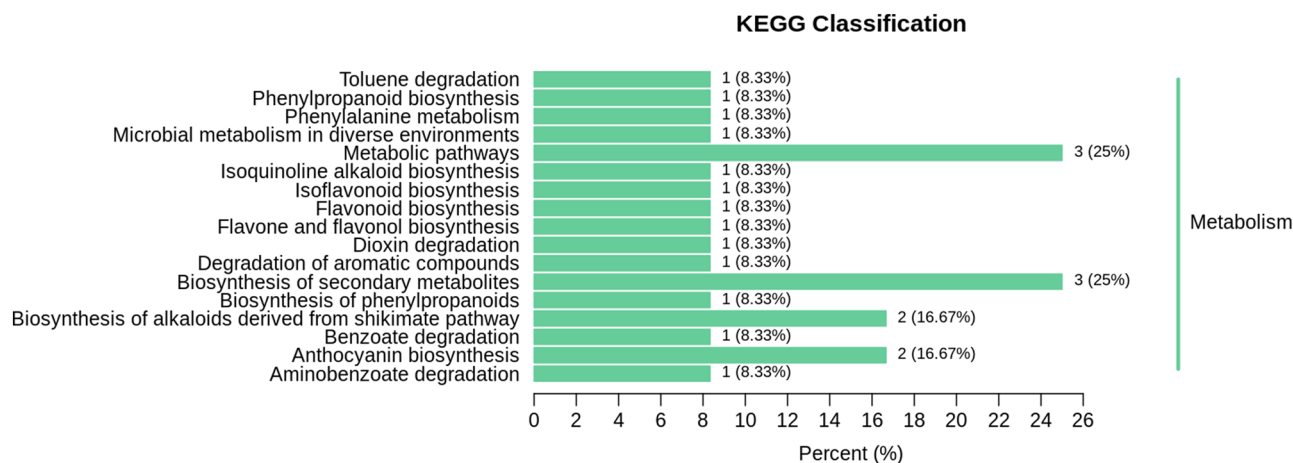


Figure 6. KEGG pathway classification map of DMs. Ordinate: KEGG metabolic pathway types. Abscissa: the number of DMs annotated to the KEGG pathway type and the proportion of this number to the total number of all the DMs annotated to all the KEGG pathways.

then identified 58 DMs between ZP and YS via multivariate analysis. It was found that the metabolites of YS were more unique, and some of them were quality control metabolites in YS instead of ZP. Our results indicated better quality of YS than that of ZP and the necessity of YS domestication, and will provide a reference for the scientific introduction, comprehensive development and utilization of wild *R. glutinosa*. In addition, the related metabolic pathways will provide a theoretical basis for the subsequent exploration of biosynthesis of related metabolites in *R. glutinosa*.

Materials and methods

Plant materials. The plant materials of *R. glutinosa* were provided by Four huaiyao research institutes in Wuzhi county, Jiaozuo city, Henan province on November 10, 2019. Among materials, the cultivated variety (ZP) of *R. glutinosa* was collected from germplasm resource nursery of Four huaiyao research institutes and the wild variety (YS) of *R. glutinosa* was collected from Sumen mountain located in Wuzhi county, Jiaozuo city, Henan province. Relevant Permissions had been provided. Then, two varieties were planted in experimental field of Henan Normal University. All samples were collected with approval and permission. Both tuberous roots were collected and freeze-dried for metabolites profiling. Three samples were used, of which each was composed of 6 different tuberous roots from 6 different mature plants in their harvest time. In each sample, they were pooled in equal parts after collection. These samples were metabolically profiled at Wuhan Maitville Biotechnology Co., Ltd (<http://www.metware.cn>).

Statement about the wild plant collected. Collection of *Rehmannia glutinosa* in this research material conforms to and complies with the IUCN Policy Statement on Research Involving Species at Risk of Extinction and the Convention on the Trade in Endangered Species of Wild Fauna and Flora. In addition, according to the List of National Key Protected Wild Plants issued by the State Forestry and Grassland Bureau of China, *Rehmannia glutinosa*, the experimental material of this study, is neither a national key protected wild plant nor an endangered plant species.

Sample preparation and extraction. Related experiments such as sample preparation and extraction of metabolites were conducted by Wuhan Maitville Biotechnology Co., Ltd. Gradient grades of methanol, acetonitrile, and acetic acid were purchased from Merck Company, Germany (<http://www.merck-chemicals.com>). The internal standard lidocaine was bought from Shanghai New Asiatic Pharmaceuticals Co., Ltd (<http://www.xinyapharm.com/>). Authentic standards of phenylpropanoids were purchased from BioBioPha Co., Ltd (<http://www.biobiopha.com/>) and all other standards were from Sigma-Aldrich, USA (<http://www.sigmaaldrich.com/united-states.html>). The vacuum freeze-dried samples (three biological replicates) were ground into powder using a mixer mill (MM 400, Retsch) with a zirconia bead for 1.5 min at 30 Hz. 100 mg powder was weighed and extracted overnight at 4 °C with 1.0 ml pure methanol (or 70% aqueous methanol) containing 0.1 mg/L lidocaine for lipid-solubility metabolites (or water-solubility metabolites). After centrifuged at 10,000g for 10 min, the supernatants were absorbed, and filtrated by millipore filter (SCAA-104, 0.22 µm pore size; ANPEL, Shanghai, China, <http://www.anpel.com.cn/>) and stored in vials for UPLC-MS/MS analysis³⁸.

Widely targeted metabolomics conditions. Data acquisition instrument system: Ultra Performance Liquid Chromatography (UPLC) (Shim-pack UFLC SHIMADZU CBM30A, <http://www.shimadzu.com.cn/>) and Tandem mass spectrometry, MS/MS (Applied Biosystems 4500 QTRAP, <http://www.appliedbiosystems.com.cn/>). Liquid chromatographic conditions: Column- Waters ACQUITY UPLC HSS T3 C18 (1.8 µm, 2.1 mm * 100 mm); solvent system, water (0.04% acetic acid): acetonitrile (0.04% acetic acid); gradient program, 100:0 V/V at 0 min, 5:95 V/V at 11.0 min, 5:95 V/V at 12.0 min, 95:5 V/V at 12.1 min, 95:5 V/V at 15.0 min; flow

rate, 0.40 mL/min⁵⁰; temperature, 40 °C; injection volume: 5 µL. The effluent was alternatively connected to an ESI-triple quadrupole-linear ion trap (Q TRAP)-MS.

LIT and triple quadrupole (QQQ) scans were acquired on a triple quadrupole-linear ion trap mass spectrometer (Q TRAP), API 4500 Q TRAP LC/MS/MS System, equipped with an ESI Turbo Ion-Spray interface, operating in a positive ion mode 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; ion source gas I (GSI), gas II(GSII), curtain gas (CUR) were set at 55, 60, and 25.0 psi, respectively; the collision gas(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 as MRM experiments with collision gas (nitrogen) set to 5 psi. DP and CE for individual MRM transitions were done with further DP and CE optimization. A specific set of MRM transitions were monitored for each period according to the metabolites eluted within this period. Similar but inconsistent experimental procedures had been successfully applied and implemented by Zhang et al.⁵¹ before.

Qualitative and quantitative analyses of metabolites. Metabolite structure analysis referred to existing mass spectrum public databases such as MassBank (<http://www.massbank.jp/>), KNAPSACK (<http://kanaya.naist.jp/KNAPSACK/>), HMDB (<http://www.hmdb.ca/>)⁵², MoTo DB (<http://www.ab.wur.nl/moto/>) and METLIN (<http://metlin.scripps.edu/index.php>)⁵³ and others. The primary and secondary spectra detected by mass spectrometry were analyzed qualitatively, and isotopic signals were removed during the analysis of some substances, including repeated signals of K⁺ ions, Na⁺ ions, NH₄⁺ ions, and fragment ions that were themselves other larger molecular weight substances repeating signal; Metabolites were quantified using MRM mode for mass spectrum peaks of metabolites, a peak per metabolite (mass spectrum file). Mass spectrum file was processed by MultiaQuant software for integration and correction of chromatographic peaks, a chromatographic peak per metabolite⁵⁴.

Statistical data analysis. Metabolite data were log₂-transformed for statistical analysis to improve normality and normalized. In order to explore the metabolites of cultivated and wild *R. glutinosa*, the 228 SMs had been used for cluster analysis by R v3.5.0 (<http://www.r-project.org/>). The cluster heat map were obtained using the agglomeration method of 'complete linkage' based on the Euclidean distances of 228 SMs between accessions. The color scale indicates the intensity of the metabolites (log₂-transformed). Peak areas were integrated using the IntelliQuan algorithm⁵⁰. Differences in the metabolites of root tissue between ZP and YS were determined using Welch's t-test (P < 0.01). The significantly changed (P < 0.01) metabolites were used for subsequent PCA. In parallel, unsupervised PCA was carried out by R v3.5.0 (<https://www.r-project.org/>). The supervised OPLS-DA was carried out by R v1.0.1, MetaboAnalystR (<https://www.r-project.org/>)⁵¹. Data were processed using Analyst 1.6.3 software.

KEGG function annotation. DMs were annotated to KEGG pathways by KEGG database³⁴. These KEGG pathways were classified according to their types.

Compliance with ethical standards. The conducted experiment complies with the laws of China.

Received: 10 September 2020; Accepted: 17 May 2021

Published online: 01 June 2021

References

- Zhou, Y. Q. *et al.* Metabolite accumulation and metabolic network in developing roots of *Rehmannia glutinosa* reveals its root developmental mechanism and quality. *Sci. Rep.* **8**, 14127–14137 (2018).
- Li, H. W. & Meng, X. L. Research progress on chemical constituents and pharmacological activities of *Rehmannia glutinosa*. *Drug Eval. Res.* **38**, 218–228 (2015).
- Li, H. X. & Wei, S. J. Studies on chemical constituents and pharmacology of *Rehmannia glutinosa*. *Chin. Med. Mod. Distance Educ. China.* **10**, 116–117 (2012).
- Alipieva, K., Korkina, L., Orhan, I. E. & Georgiev, M. I. Verbascoside—A review of its occurrence, (bio)synthesis and pharmacological significance. *Biotechnol. Adv.* **32**, 1065–1076 (2014).
- Quan, Y. P., Wang, Y. S., Xin, Z. H. & Ji, S. D. The genuine medicinal materials of four-major Huai-yao and its developmental prospects. *Hubei Agric. Sci.* **50**, 3097–3101 (2011).
- Wang, F. Q., Xie, C. X., Sun, R. B. & Zhang, Z. Y. Progress on germplasm enhancement and breeding of *Rehmannia glutinosa*. *China J. Chin. Materia Med.* **43**, 4203–4209 (2018).
- Guo, M. M., Zhou, Y. Q., Duan, H. Y., Yang, K. & Shao, L. Y. Development of SNP marker based on the *Rehmannia glutinosa* transcriptome database and construction of DNA fingerprint in *Rehmannia*. *Biotechnol. Bull.* **35**, 224–230 (2019).
- Lea, G., Vessela, A. P., Sylvain, C. & Florence, R. F. Metabolomics to decipher the chemical defense of cereals against *Fusarium graminearum* and deoxynivalenol accumulation. *Int. J. Mol. Sci.* **16**, 24839–24872 (2015).
- Steinmeyer, S., Lee, K., Jayaraman, A. & Alaniz, R. C. Microbiota metabolite regulation of host immune homeostasis: A mechanistic missing link. *Curr. Allergy Asthma Rep.* **15**, 24 (2015).
- Chen, Z. M. *et al.* Overview on application of metabonomics in study on processing mechanism of traditional Chinese materia medica. *Chin. Arch. Tradit. Chin. Med.* **36**, 275–279 (2018).
- Wang, C. C. *et al.* Thoughts of omics research on quality formation in Dao-di herbs. *China J. Chin. Materia Med.* **43**, 2307–2412 (2018).
- Lin, S., Wu, H. Y. & Li, K. P. Advances in metabonomics of medicinal plants based on XC-MS and NMR. *J. Chin. Med. Mater.* **38**, 2651–2654 (2015).

13. Liu, X. & Wang, Y. H. The application of NMR-based metabolomics technique to medicinal plants. *Acta Pharmaceutica Sinica*. **52**, 541–549 (2017).
14. Xue, S. & Chen, S. Q. Analysis on change rule of carbohydrates in *Rehmanniae radix* praeparata with different processing times based on metabolomics. *Chin. J. Exp. Tradit. Med. Formulae* **24**, 1–5 (2018).
15. Chen, W. K. Study of targeted metabolome and transcriptome in Grape berries grown under double cropping viticulture system. Doctoral thesis, China Agricultural University, Beijing (2018).
16. Tian, W. J. Targeted metabolomics profiling of Loquat during fruit development and postharvest ripening. Master thesis, Zhejiang University, Zhejiang (2017).
17. Dai, W. D., Hu, Z. Y., Xie, D. C., Tan, J. F. & Lin, Z. A novel spatial-resolution targeted metabolomics method in a single leaf of the tea plant (*Camellia sinensis*). *Food Chem.* **311**, 1–4 (2020).
18. Li, P. L. *et al.* Metabolomic analysis reveals the composition differences in 13 Chinese tea cultivars of different manufacturing suitabilities. *J. Sci. Food Agric.* **98**, 1153–1161 (2018).
19. Tan, J. F. *et al.* Study of the dynamic changes in the non-volatile chemical constituents of black tea during fermentation processing by a non-targeted metabolomics approach. *Food Res. Int.* **79**, 106–113 (2016).
20. Li, J. *et al.* Phytochemical comparison of different tea (*Camellia sinensis*) cultivars and its association with sensory quality of finished tea. *LWT*. **117**, 108595 (2020).
21. Sawada, Y. *et al.* Widely targeted metabolomics based on large-scale MS/MS data for elucidating metabolite accumulation patterns in plants. *Plant Cell Physiol.* **50**, 37–47 (2009).
22. Chen, W. *et al.* A novel integrated method for large-scale detection, identification, and quantification of widely targeted metabolites: Application in the study of rice metabolomics. *Mol. Plant* **6**, 1769–1780 (2013).
23. Wang, D. D. *et al.* Identification of nutritional components in black sesame determined by widely targeted metabolomics and traditional Chinese medicines. *Molecules* **23**, 1180–1192 (2018).
24. Chen, W. *et al.* Comparative and parallel genome-wide association studies for metabolic and agronomic traits in cereals. *Nat. Commun.* **7**, 12767 (2016).
25. Fang, C. Y. *et al.* Control of leaf senescence by an MeOH-jasmonates cascade that is epigenetically regulated by OsSRT1 in rice. *Mol. Plant* **9**, 1366–1378 (2016).
26. Cai, Y. Y., Chen, X. X., Gu, F. L. & Xu, F. Vanilla beans forming different processing stages with widely targeted metabolomic technique. *Chin. J. Trop. Crops*. **40**, 1325–1335 (2019).
27. Wang, T. *et al.* Widely targeted metabolomics analysis reveals the effect of flooding stress on the synthesis of flavonoids in *Chrysanthemum morifolium*. *Molecules* **24**, 3695 (2019).
28. Thévenot, E. A., Roux, A., Xu, Y., Ezan, E. & Junot, C. Analysis of the human adult urinary metabolome variations with age, body mass index, and gender by implementing a comprehensive workflow for univariate and OPLS statistical analyses. *J. Proteome Res.* **14**, 3322–3335 (2015).
29. Iker, H., Leonor, A., Frank, V. B. & Sergi, M. B. How relevant are flavonoids as antioxidants in plants?. *Trends Plant Sci.* **14**, 125–132 (2009).
30. Dalia, S., Vanessa, P. & Ramiro, M. M. Review of flavonoids: A diverse group of natural compounds with anti-Candida albicans activity in vitro. *Arch. Oral Biol.* **76**, 76–83 (2017).
31. Liang, W. Y. *et al.* Research progress on salvanlic acids of *Salvia miltiorrhiza*. *China J. Chin. Mater. Med.* **41**, 806–812 (2016).
32. Zhao, H. Y. *et al.* Cloning and bioinformatics analysis of promoters of three terpene synthase genes and their terpenoid regulation in *Amomum villosum*. *Chin. Tradit. Herb. Drugs*. **52**, 1117–1127 (2021).
33. Zhang, Q. L., Zhao, H. B., Wang, J. & Cui, Z. J. Research progress on chemical constituents and pharmacological activities of *Aconitum gymnanodrum*. *West China J. Pharm. Sci.* **36**, 99–102 (2021).
34. Kanehisa, M. & Goto, S. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* **28**, 27–30 (2020).
35. Zuo, M. & Zuo, J. J. Domestication, cultivation and application of wild *Rehmannia glutinosa*. *Agric. Technol.* **39**, 93–94 (2019).
36. Yang, K. Metabolite accumulation and metabolic network in developing roots of *Rehmannia glutinosa* reveal its root developmental mechanism, quality and verbascoside biosynthesis pathway. Master thesis, Henan Normal University, Xinxiang (2019).
37. Maldini, M. *et al.* Metabolomic study of wild and cultivated caper (*Capparis spinosa* L.) from different areas of Sardinia and their comparative evaluation. *J. Mass Spectrom.* **51**, 716–728 (2016).
38. Zhu, G. *et al.* Rewiring of the fruit metabolome in tomato breeding. *Cell* **172**, 249–261 (2018).
39. Huang, S. & Cai, T. Research progress of flavonoid eutectic compound. *J. Mod. Med. Health.* **35**, 3791–3795 (2019).
40. Qiu, Z. M. *et al.* Research progress on anthocyanins in Lettuce. *China Veg.* **12**, 25–30 (2019).
41. Zou, L. Q., Wang, C. X., Kuang, X. J., Li, Y. & Sun, C. Advance in flavonoids biosynthetic pathway and synthetic biology. *China J. Chin. Mater. Med.* **41**, 4124–4128 (2016).
42. Kang, Y. L. *et al.* Research progress on flavonoid metabolic synthesis pathway and related function genes in medicinal plants. *Chin. Tradit. Herb. Drugs*. **45**, 1336–1341 (2014).
43. Pharmacopoeia of the People's Republic of China: Part I [S]. 2015 Edition 156–157 (China Medical Science Press, 2015).
44. Li, X. E. Physiological characterization of *Rehmannia glutinosa* cultivar and wild type varieties and its relationship with the tuberous root development. Doctoral thesis, Nanjing Agricultural University, Jiangsu (2013).
45. Xie, Q. *et al.* Research progress on chemical constituents and pharmacological effects of benzoin plants. *J. Chin. Med. Mater.* **1**, 243–248 (2020).
46. Wang, W. *et al.* Ferulic acid suppresses obesity and obesity-related metabolic syndromes in high fat diet-induced obese C57BL/6J mice. *Food Agric. Immunol.* **29**, 1116–1125 (2018).
47. Naowaboot, J. *et al.* Ferulic acid improves lipid and glucose homeostasis in high-fat diet induced obese mice. *Clin. Exp. Pharmacol. Physiol.* **43**, 242–250 (2016).
48. Ren, Y. *et al.* Study on components efficacy of *Salviae Miltiorrhizae Radix et Rhizoma* based on systematic traditional Chinese medicine. *China J. Chin. Mater. Med.* **45**, 3251–3258 (2020).
49. Zhou, Y. Q., Wang, X. N., Wang, W. S. & Duan, H. Y. D. *nov*o transcriptome sequencing-based discovery and expression analyses of verbascoside biosynthesis-associated genes in *Rehmannia glutinosa* tuberous roots. *Mol. Breed.* **36**, 139–149 (2016).
50. Qian, R., Ren, X. Y., Liu, Y. X. & Jiang, X. L. Widely-targeted metabolomics analysis of *Shiraiia bambusicola* bioactive components. *Res. Pap.* **38**, 631–645 (2019).
51. Zhang, S. S. *et al.* Identification of potential metabolites mediating bird's selective feeding on *Prunus mira* flowers. *BioMed Res. Int.* 1–8 (2019).
52. Wishart, D. S. *et al.* HMDB 3.0—The human metabolome database in 2013. *Nucleic Acids Res.* **41**, 801–807 (2013).
53. Zhu, Z. J. *et al.* Liquid chromatography quadrupole time-of-flight mass spectrometry characterization of metabolites guided by the METLIN database. *Nat. Protoc.* **8**, 451–460 (2013).
54. Fraga, C. G., Clowers, B. H., Moore, R. J. & Zink, E. M. Signature-discovery approach for sample matching of a nerve-agent precursor using liquid chromatography–mass spectrometry, XCMS, and chemometrics. *Anal. Chem.* **82**, 4165–4173 (2010).

Acknowledgements

This research was funded by National Natural Science Foundation of China (31870312), Natural Science Foundation of Henan province (No. 182300410018), Key R & D and promotion projects in Henan Province (No. 212102110405) and Fund of Henan Normal University (No. 2019JQ01). The authors would like to thank Kanehisa Laboratories for providing the image of Phenylpropanoid biosynthesis, and thank Director Huojin Li, Four Huaiyao Institute of Qinhuai, Wuzhi County, Henan, China for his provision of *Rehmannia glutinosa* materials.

Author contributions

Z.Y.Q. conceptualized the study, carried out the conceptualization, validation, data curation, project administration, reviewed and edited manuscript. S.L.Y. and Z.J.L. performed the formal analysis, data curation and original draft preparation. L.H.M. and D.H.Y. helped improve the content of the manuscript. All authors gave final approval for publication.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-021-90961-6>.

Correspondence and requests for materials should be addressed to Y.Z.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021