LRF-824
[1-10]

RESEARCH ARTICLE Legume Research- An International Journal

Weighted Correlation Network Analysis Mining Key Metabolites and Genes for Regulation of Root Growth and Secondary Metabolism of *Astragalus membranaceus mongholicus*

Xiaojie Li[1,2], Yingtong Mu[1,2], Junjie Wang[1,2], Xiaoming Zhang[1,2]

10.18805/LRF-824

ABSTRACT

Background: Harvesting period is an important feature that not only affects the yield quality of *Astragalus* herbs but also the content of astragaloside IV and calycosin 7-O-β-D-glucopyranoside which are considered as the core indicators of the quality of *Astragalus* herbs. Herb yield and quality undergo rapid and substantial changes during the harvesting process. Part of enzymes and their encoding genes have been identified as potential candidates for the biosynthesis of herb quality-related triterpenoid saponins and isoflavonoids. However, wide gap still exists in the understanding of the pathways involving growth, development and biosynthesis of these two important active components in the roots of *Astragalus*.

Methods: In this study, we used *Astragalus* roots as material to analyze its metabolome and transcriptome of herbs at harvesting stage.

Result: The results indicated that changes in the harvesting period had a significant effect on the yield and quality of the herbs. Correlation of metabolome and RNA-seq results constructed using weighted gene co-expression network analysis (WGCNA) with herb growth and development indicated that the expression patterns of 19 hub metabolites (Oxamniquine, 4-ene-Valproic acid, Traumatic Acid, L-Histidinol, Salidroside, Temsavir, Retaspimycin, Isosakuranetin, *etc*.) and 16 structural genes with annotations (CNX2, talA, ABCF2, HIBCH, crtZ, *etc*.) had the highest connectivity with root growth and secondary metabolism of *Astragalus*. To search for potential regulatory genes related to root growth and development, we screened 54 relevant genes using 16 hub genes with annotations. The results advance the molecular basis of biosynthesis of secondary metabolites related to the harvesting period of *Astragalus* herbs, which can provide a theoretical basis for determining the regulatory mechanism of yield and quality.

Key words: *Astragalus*, Hub gene and metabolite, Secondary metabolism, WGCNA.

INTRODUCTION

Astragalus membranaceus mongholicus (*Astragalus*) is a plant belonging to the genus *Astragalus* in the family Leguminosae, which is used as a medicine by roots (State Pharmacopeia Committee of China., 2020). It is a homologous plant used for medicine and food and has vast economic significance. *Astragalus* plants are widely used in clinical trials and other applications due to the high value of their medicinal components (Pham, 2015). Chemical composition of *Astragalus* mainly includes polysaccharides, flavonoids and triterpenoids. Its pharmacological effects are wide-ranging, including enhancement of immune function, anti-tumour, regulation of body metabolism, protection of the nervous system (Xiao *et al*., 2020; Chen *et al*., 2022). One of the main effective components of *Astragalus* are triterpenoid saponins (Xiang *et al*., 2022). At present, more than 50 triterpenoid saponins have been obtained from *Astragalus* plants, mainly including astragalosides I-VIII, isoastragalosides I, II, IV, acetyl astragalosides, soyasaponins (Xiao *et al*., 2020). The accumulation of chemical components in *Astragalus* plants showed a dynamic change and the content of effective components varied greatly at different harvest times. Even in the same year and the same harvest time, the content of effective components varied greatly among different individuals (Wang *et al*., 2019; Xin *et al*., 2015). Therefore, it is particularly important to study the synthetic genes related to the growth, development and quantities of secondary metabolites in *Astragalus* roots and their expression regulation mechanisms, which can help in further research on ways to improve the quality of *Astragalus* herbs.

---

[1] Key Laboratory of Grassland Resources, College of Grassland, Resource and Environmental Science, Inner Mongolia Agricultural University, Ministry of Education, Hohhot, 010021, P.R. of China.
[2] Engineering Research Center for the Seed breeding of Chinese and Mongolian Medicinal Materials in Inner Mongolia, Hohhot, 010011, Inner Mongolia, China.

Corresponding Author: Xiaoming Zhang, Key Laboratory of Grassland Resources, College of Grassland, Resource and Environmental Science, Inner Mongolia Agricultural University, Ministry of Education, Hohhot, 010021, P.R. of China.
Email: bagenna123@aliyun.com

How to cite this article: Li, X., Mu, Y., Wang, J. and Zhang, X. (2024). Weighted Correlation Network Analysis Mining Key Metabolites and Genes for Regulation of Root Growth and Secondary Metabolism of *Astragalus membranaceus mongholicus.* Legume Research. 1-10. doi: 10.18805/LRF-824.

Submitted: 03-06-2024 **Accepted:** 13-09-2024 **Online**: 10-01-2025

Non-targeted metabolomics and transcriptome sequencing technology can determine the functional metabolites and genes of individuals at different stages and their expression. It can establish a connection between the genetic information of the biological genome and functional metabolites, which can be helpful for the analysis of functional genes. Currently, metabolome and transcriptome sequencing technologies have been widely used in the fields of plants and animals, which greatly broaden the research methods of biological metabolites and beneficial to the study of the synthesis and accumulation mechanisms of active ingredients in medicinal plants, especially applicable to non-model plants like *Astragalus* (Wang *et al*., 2019). At present, the available research on transcriptome and metabolome of the medicinal parts (roots) of *Astragalus* is very less and inadequate. Construction of co-expression network is an important tool for the efficient use of transcriptomic and metabolomic data. Combining multi-sample data with weighted gene co-expression network analysis (WGCNA) can cluster genes with similar expression patterns to construct a co-expression gene module. Then, the module associated with sample characteristics and the key module with the highest correlation with the research phenotype can be quickly found and the core genes in the key module can be excavated (Ji *et al*., 2022; Aidi *et al*., 2020; Na *et al*., 2024). These core genes and metabolites may be the key factors for the regulation of root phenotype and secondary metabolism in *Astragalus*.

Based on the academic background of the secondary metabolite biosynthesis of *Astragalus*, the objectives of this research aimed to identify the key metabolites and genes involved in the root growth, development, synthesis and accumulation of two important active ingredients in the herb. Key metabolites and key structural genes were identified by metabolomics and transcriptomics studies using Weighted Gene Co-Expression Network Analysis (WGCNA). These results not only filled some gaps in the knowledge regarding regulatory pathways of growth and development of Astragalus, but also identified some key genes and metabolites involved in the regulatory mechanisms of root secondary metabolism. Our findings can provide a basis for growth and development, active ingredient synthesis and breeding research of this plant.

MATERIALS AND METHODS

Plant material and treatments

The *Astragalus* roots harvested during 2022 season from a test base (Inner Mongolia, Hohhot) were used as the test material for this study. We declare that the Astragalus roots in this study complied with institutional, national and international guidelines for the collection and cultivation of any plant materials. At each sampling period, six replicates of root were collected from each batch (period A, Oct.6; period B, Oct.16; period C, Oct.26; period D, Nov.6 respectively). The materials collected were frozen with liquid nitrogen and stored at -80°C. The treatment and sample collection, as well as the related indices (root length, root diameter, root fresh weight, thick lateral root, contents of astragaloside IV and calycosin 7-O-β-D-glucopyranoside) were estimated as described by Li *et al.* (2024).

Untargeted metabolome analysis, WGCNA and metabolite network visualization

xpression profile matrices of metabolites were derived from 24 metabolite expression samples from *Astragalus* roots of different harvesting periods. Co-expression networks were created using WGCNA (v1.29) package in R. A total of 1660 metabolites were used as input to the signed WGCNA network construction. The automatic network construction function block wise was used to build modules. In standard WGCNA networks, soft power was set to 6, Merge Module Threshold was 30 and the merge Cut Height value was 0.25. Metabolite expression adjacency matrices were constructed and used to analyze the network topology. In addition, module correlation analysis was performed on module eigenvalues and phenotypic trait data and Pearson correlation was used to calculate the correlation coefficient between phenotypic trait data and metabolite module eigenvalues and their correlation heat maps were drawn. In order to filter the modules with high correlation of tissue development, the threshold of correlation coefficient was defined as 0.70 and the significant level was 0.005. This means that any module with a correlation coefficient higher than 0.70 and a significant level of 0.005 will be regarded as a tissue development-specific module.

RNA-seq, WGCNA and gene network visualization

Expression profile matrices of genes were derived from 12 gene expression samples of Astragalus roots at different harvesting periods. The details of RNA-seq were described by Li et al (Li *et al*., 2024). The SRA accession numbers are SRR28617061-SRR28617072. Co-expression networks were created using WGCNA (v1.29) package in R. To ensure the specificity of the co-expression network, the data used for WGCNA co-expression network analysis were all RNAseqgenes after discarding none detectable genes (FPKM > 1) (Lin *et al*., 2017). A total of 23851 genes were used as input to the signed WGCNA network construction. The automatic network construction function block wise was used to build modules. In standard WGCNA networks, soft power was set to 5, min Module Size was 30 and the merge Cut Height value was 0.25.

Gene expression adjacency matrices were constructed and used to analyze the network topology. In addition, module correlation analysis was performed on module eigenvalues and phenotypic trait data and Pearson correlation was used to calculate the correlation coefficient between phenotypic trait data and gene module eigenvalues and their correlation heat maps were drawn. In order to filter the modules with high correlation of tissue development, the threshold of correlation coefficient was defined as 0.70 and the significant level was 0.005. This means that any module with a correlation coefficient higher than 0.70 and a significant level of 0.005 will be regarded as a tissue development-specific module (Downs *et al.*, 2013).

Enrichment pathway analysis

Firstly extracted the genes and metabolites in the modules significantly associated with the growth, development and secondary metabolism of *Astragalus* and then analysed by GO and KEGG enrichment using Goatools (Version 0.6.5, https://github.com/tanghaibao/GOatools) and KOBAS 2.1.1 (Version 2.1.1, http://kobas.cbi.pku.edu.cn/home.do) (Diao *et al.*, 2019). Essentially, differential expression analysis was performed using the DESeq2. DEGs with |log2FC| > 1 and FDR < 0.05) were considered to be significantly different expressed genes.

Related modules on growth and secondary metabolic regulatory pathways

According to the topological overlap matrix, the weights of different genes in the module were calculated. The higher the weight value, the higher the degree of association between the genes. The genes with higher correlation coefficients among the modules were selected and at the same time, the genes that have already been annotated in the modules were selected. The genes were used to construct an interactions network, in which each node represented a gene and the two connecting lines were used to connect the two genes. Each node in the network represents a gene and the genes at both ends of the connection line have similar biological functions. As complex agronomic traits, growth development and secondary metabolism are not only related to external environmental signals, but also involve endogenous signal regulation. According to the existing literature, some keywords related to growth and metabolism regulation, such as pentose phosphate pathway, carotenoid biosynthesis, zeatin biosynthesis, plant hormone signal transduction, etc., were used to search for Look for metabolites and genes involved in regulation in tissue-specific modules (Tuan *et al.*, 2015; Li *et al.*, 2022; Zhang *et al.*, 2022). The network of candidate target genes and metabolites (Top 10 percent connectivity) were visualized by Cytoscape (v3.7.1, USA).

qRT-PCR of genes

To further validate the analysis of the Illumina sequencing data, we randomly selected six DEGs for qRT-PCR amplification and specific primers were designed by Primer-BLAST from NCBI. Total root RNA was isolated from three biological replicate samples of each group by QIAzolLysisReagent (Qiagen, Germany). Reverse transcription and polymerase chain reaction were conducted by PrimeScriptTM 1st Strand cDNA Synthesis Kit and SYBR®Premix Ex TaqTM II (Tli RNaseH Plus) (Takara Biomedical Technology (Beijing) Co., Ltd., China). Targeted gene was measured using Heal Force RealTime PCR System (Heal Force CG-05, Hangzhou Jingle Scientific Instrument Co., Ltd., China) as described previously and each sample was measured for three times technical replicates. The reaction system was 20 µL, including 10 µL of Tli RNaseH Plus, 0.4 µL of Primer F (10 µM), 0.4 µL of Primer R (10 µM), 2 µL of cDNA, ROX Reference Dye 0.08 µL and 7.12 µL of ddH_2O. All reactions were conducted in 96-well plates. The qPCR protocol included annealing at 95°C for 60 s, followed by 40 cycles of 95°C for 30 s, primer annealing at 60°C for 30 s and extension at 72°C for 20 s. A negative control without template for each primer pair was included in each qPCR analysis. Expression was normalized to GAPDH and calculated using $2^{-\Delta\Delta Ct}$.

Statistical analysis

The R package from Bioconductor on Majorbio Cloud Platform was used to conduct a WGCNA analysis (https://horvath.genetics.ucla.edu/html/CoexpressionNetwork/Rpackages/WGCNA/faq.html). The statistical significance was calculated with a Student's t-test (*P<0.05, **P<0.01 and ***P<0.001). The figures were prepared using Excel 2019 software. Details of the data analysis are described in support file.

RESULTS AND DISCUSSION

Phenotypes and active ingredients analysis during harvesting

Understanding the formation of secondary metabolites is critical to the quality and marketability of *Astragalus*, especially the production of astragaloside IV and calycosin 7-O-β-D-glucopyranoside, two important metabolites that are major components in the formation of herb quality (Fig 1B,C,D). During the growth of *Astragalus*, secondary metabolite synthesis-related genes and metabolites showed up-regulated or down-regulated conditions with the extension of harvesting time. It was suggested that the accumulation of astragaloside IV and calycosin 7-O-β-D-glucopyranoside during ripening was likely driven by ACAT, HMGCS, HMGCR, MVK, MVAK2, MVD and other genes (Wang *et al.*, 2023; Li, 2020; Zhang, 2020). The production of isofavone, especially calycosin 7-O-β-D-glucopyranoside, appeared to be strongly dependent on formonetin and calycosin metabolites (Zhang *et al.*, 2022). In this study, the content of secondary metabolite astragaloside IV was significantly higher in C and D periods and calycosin 7-O-β-D-glucopyranoside content was significantly higher in D period.

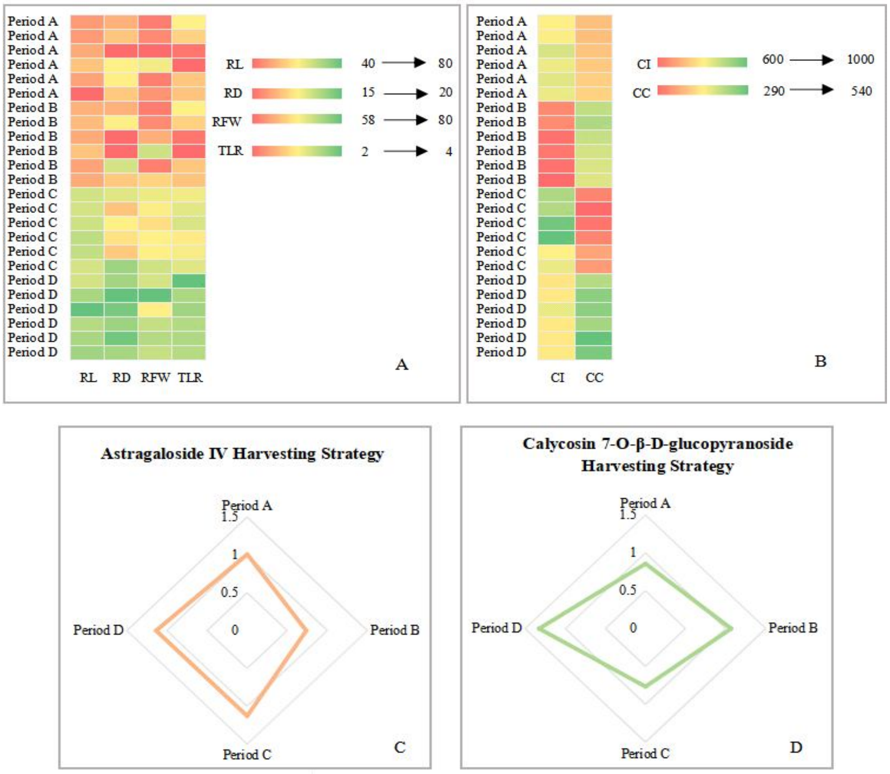
Growth and development indicators and two important secondary metabolites of *Astragalus* roots at different harvesting periods were analysed according to the group's previous research. Indicators related to herb yield and calycosin 7-O-β-D-glucopyranoside content were highest in period D (Fig 1A,B). Astragaloside IV stayed high level in Period C root and identified as the major contributors to *Astragalus* efficacy (Wu *et al.*, 2020; Li *et al.*, 2024). Besides, higher calycosin 7-O-β-D-glucopyranoside content in herbs after harvesting in periods B and D. Analyses of the contents of the two important active ingredients and the accumulation pattern of the herbs showed that the highest yield, astragaloside IV content and calycosin 7-O-β-D-glucopyranoside content of the herbs after harvesting in period D was the best harvesting period. If batch extraction of astragaloside IV or calycosin 7-O-β-D-glucopyranoside is required for subsequent studies, the herbs could be harvested in stages according to the harvesting strategy (Fig 1C,D).

WGCNA of metabolomic data

The WGCNA was performed to investigate the coexpression networks, in which all co-expressed metabolites were connected to each other with varying correlation strengths. Metabolites were partitioned into 9 co-expression modules (Fig 2A,B). Period D were positively correlated with metabolites expression in the 'yellow' module and negatively correlated with metabolites expression in the 'turquoise' and 'turquoise' modules, with a coefficient of 0.91 and -0.82, respectively (Fig 2A). A total of 637 metabolites were identified based on the correlation (r > 0.8) of metabolites with period D in the tow modules.

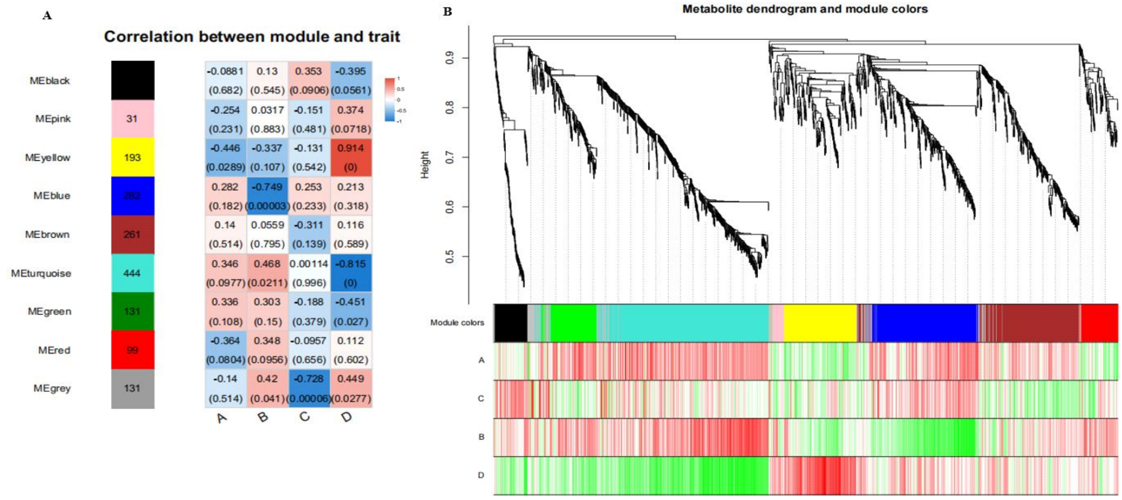
Functional analysis of metabolites in correlated modules

KEGG compounds were classified for the two modules screened in relation to the *Astragalus* roots harvesting period D. Classification of KEGG analysis revealed that metabolites could be summarized in three main functional categories, including lipids, peptides and carbohydrates process (Figure 3A/B). We performed pathway enrichment analyses on the relevant modules and selected significantly enriched metabolic pathways. KEGG enrichment revealed that the modules related to growth and development were mainly enriched in Phenylpropanoid biosynthesis, Flavonoid biosynthesis, Nucleotide metabolism, Stilbenoid, diarylheptanoid and gingerol biosynthesis, Glycerophospholipid metabolism, alpha-Linolenic acid metabolism, Plant hormone signal transduction, Linoleic acid metabolism and Arachidonic acid metabolism (Figure 3C/D). We speculated that these pathways may play an important role in regulating the root growth and development and secondary metabolic processes of *Astragalus.*



(A) Thermograms of Root length-RL, Root diameter-RD, Root fresh weight-RFW and Thick lateral root-TLR at different harvesting periods. (B) Thermograms of Content of astragaloside IV-CI and Content of calycosin 7-O-β-D-glucopyranoside-CC at different harvesting periods. (C) Mode of accumulation of astragaloside IV content and herb yield. (D) Mode of accumulation of calycosin 7-O-β-D-glucopyranoside content and herb yield. Note: Calculation of yield and accumulation patterns of two major active ingredients. Root fresh weight * content in different periods/mean value of weight * mean value of content in four periods

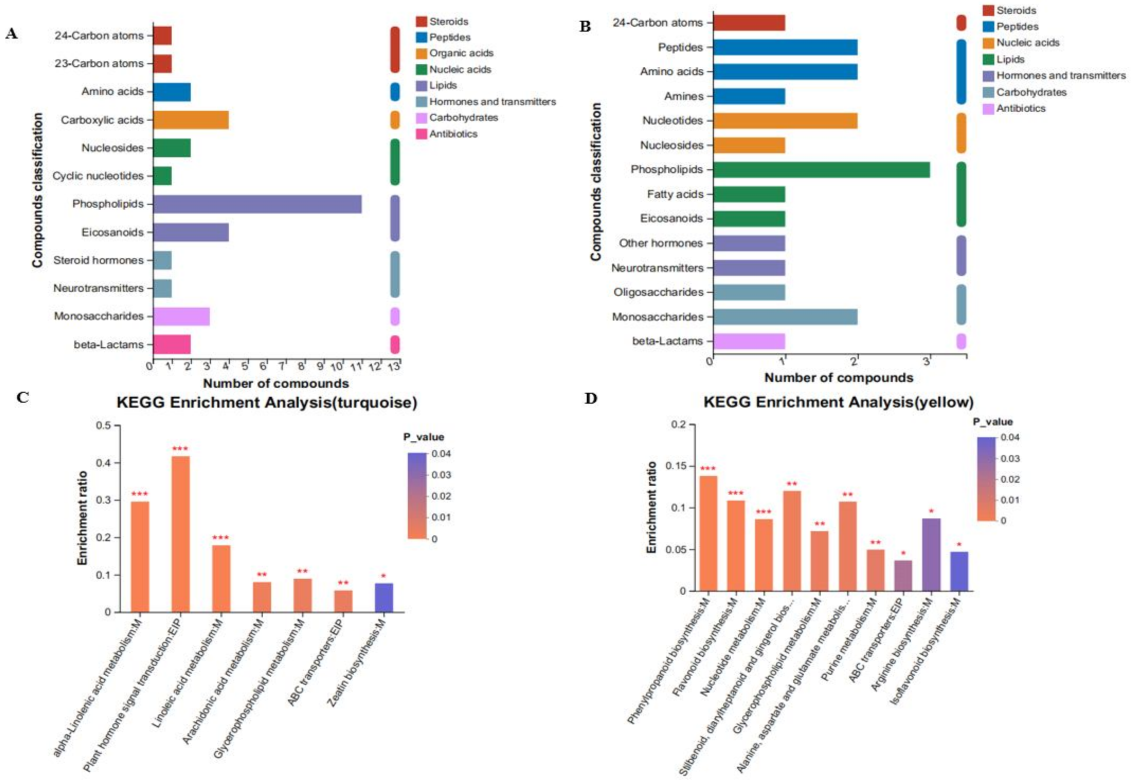
Fig 1: Analysis of phenotypes and active Ingredients in *Astragalus* roots.



		A	B	C	D
MEblack		-0.0881 (0.682)	0.13 (0.545)	0.353 (0.0906)	-0.395 (0.0561)
MEpink	31	-0.254 (0.231)	0.0317 (0.883)	-0.151 (0.481)	0.374 (0.0718)
MEyellow	193	-0.446 (0.0289)	-0.337 (0.107)	-0.131 (0.542)	0.914 (0)
MEblue	282	0.282 (0.182)	-0.749 (0.00003)	0.253 (0.233)	0.213 (0.318)
MEbrown	261	0.14 (0.514)	0.0559 (0.795)	-0.311 (0.139)	0.116 (0.589)
MEturquoise	444	0.346 (0.0977)	0.468 (0.0211)	0.00114 (0.996)	-0.815 (0)
MEgreen	131	0.336 (0.108)	0.303 (0.15)	-0.188 (0.379)	-0.451 (0.027)
MEred	99	-0.364 (0.0804)	0.348 (0.0956)	-0.0957 (0.656)	0.112 (0.602)
MEgrey	131	-0.14 (0.514)	0.42 (0.041)	-0.728 (0.00006)	0.449 (0.0277)

(A) Module-period correlations and corresponding p-values. Each row corresponds to a cluster. Each column corresponds to a period. The left panel shows 9 modules and the right panel is a color scale for module trait correlation from -1 to 1. (B) Hierarchical cluster tree showing 9 modules of co-expressed metabolites. Each leaf represents one gene in the tree. Each row represents a period, each column represents a metabolite within the module and the colour represents the magnitude of the correlation, with red representing a positive correlation and green a negative correlation.

Fig 2: Co-expression network analysis of metabolomic associated with growth, development and secondary metabolism during *Astragalus* roots harvesting.



(A) Classification of KEGG of turquoise module associated with root growth and development. (B) Classification of KEGG of yellow module associated with root growth and development. (C) KEGG enrichment of turquoise module associated with root growth and development. (D) KEGG enrichment of yellow module associated with root growth and development.

Fig 3: Classification of KEGG compounds co-expressing metabolites and analysis of KEGG enrichment of metabolites in target modules.

Regulation of growth and development and two major active components by key metabolites

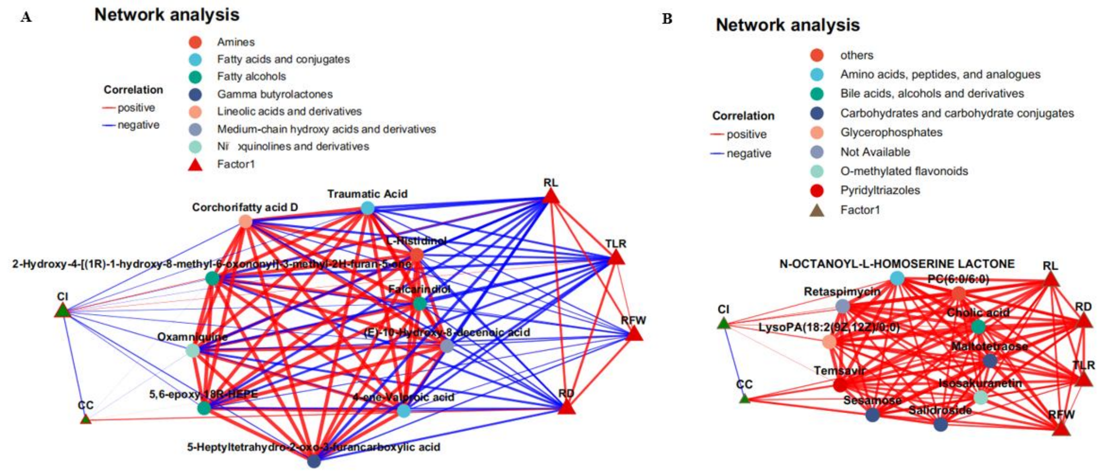
Metabolites networks were visualised and metabolite connectivity analysed for metabolites in modules related to period D in *Astragalus*. To determine the Hub metabolites related to period D, we classified the ten metabolites with the highest kME values in each module as hub metabolites and utilized the hub metabolites and their interacting indicators of phenotype and active ingredient content to map the metabolites co-expression network (Fig 4). 19 highly connected metabolites screened by WGCNA analysis using period D tissue-specific modules (yellow turquoise) for multifactorial correlation were highly correlated with indicators of growth development and active ingredient content of *Astragalus* (Fig 4A,B).

WGCNA of RNA-seq data

The WGCNA was performed to investigate the coexpression networks, in which all co-expressed genes were connected to each other with varying correlation strengths. Genes were partitioned into 36 co-expression modules (Figure 5A/B). Growth and development indicators were positively correlated with gene expression in the ‘blue’ module and negatively correlated with gene expression in the ‘brown’ and ‘turquoise’ modules. The content of CI was negatively correlated with gene expression in the ‘grey60’ module, with a coefficient of -0.76. The content of CC was positively correlated with gene expression in the ‘pink’ module and negatively correlated in the ‘brown’ and ‘yellow’ modules, with a coefficient of 0.71, -0.86 and -0.78, respectively (Fig 5A). A total of 14701 genes were identified based on the correlation ($r > 0.7$) of genes with growth and development and active ingredient content in the six modules. A total of 14,701 genes were identified based on the correlation between genes and growth and development in the 3 modules. A total of 5096 genes were identified based on the correlation between genes and active ingredient content in the 4 modules.

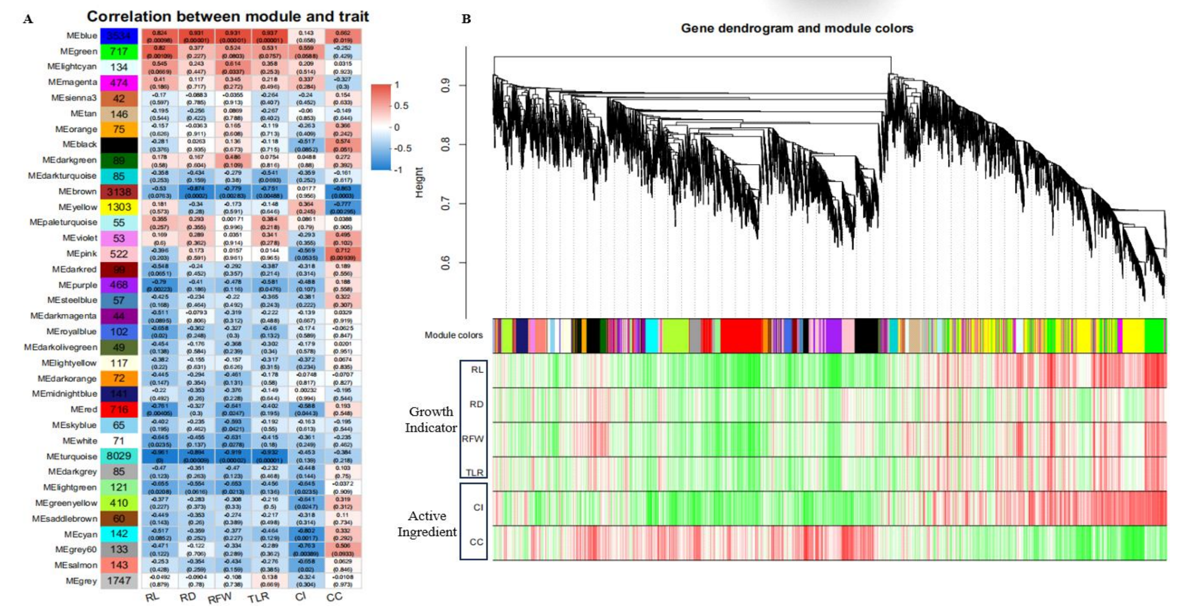
Functional analysis of genes in correlated modules

GO annotation of three types of modules related to root growth and development, astragaloside IV content and calycosin 7-O-β-D-glucopyranoside content of screened *Astragalus* roots, respectively. GO annotation analysis revealed that genes could be summarized in three main functional categories, including cellular component, molecular function and biological process. Five groups, including macromolecule metabolic process, integral component of membrane, intracellular organelle, cellular macromolecule metabolic process and organonitrogen compound metabolic process, were the main classifications for more than 50% of the genes in the GO annotations related to growth and development (Fig 6A). Six groups, including macromolecule metabolic process, integral component of membrane, intracellular organelle, intracellular membrane-bounded organelle, organonitrogen compound metabolic process and cellular macromolecule metabolic process, were the main classifications for more than 50% of the genes in the GO annotations related to calycosin 7-O-β-D-glucopyranoside content (Fig 6B). Molecular function is the main classifications for more than 50% of the genes in the GO annotations related to Astragaloside IV content (Fig 6C).



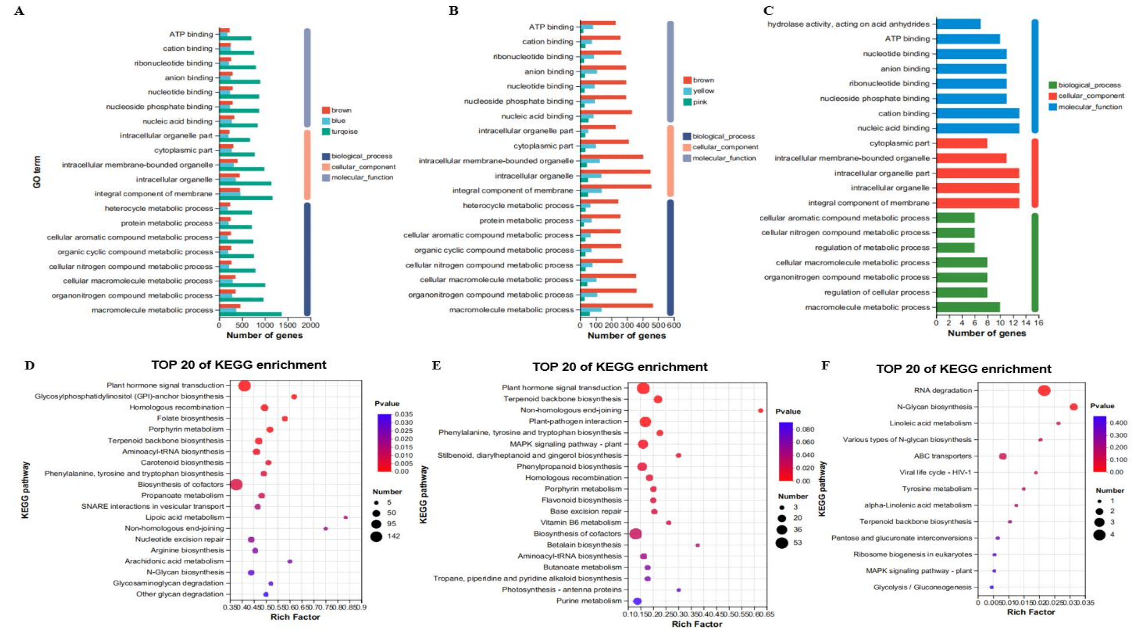
(A) Network interaction analysis of hub metabolites in turquoise module. (B) Network interaction analysis of hub metabolites in yellow module. The multifactorial correlation network diagram mainly reflects the correlation between various indicators and metabolite classification levels in different periods. The size of the nodes in the graph indicates the degree size of the nodes and different colours indicate different metabolites; the colours of the connecting lines indicate positive and negative correlation, red indicates positive correlation and blue indicates negative correlation; the thickness of the lines indicates the size of the correlation coefficient; the thicker the line, the higher the correlation between the species; and the more lines, the closer the connection between the nodes.

Fig 4: Analysis of multifactorial interactions of hub metabolites network interaction in period D significant enrichment module.



(A) Module-Phenotypes and active ingredients correlations and corresponding p-values. Each row corresponds to a cluster. Each column corresponds to a indicator. The left panel shows 36 modules and the right panel is a color scale for module trait correlation from -1 to 1. (B) Hierarchical cluster tree showing 36 modules of co-expressed genes. Each leaf represents one gene in the tree. Each row represents a indicator, each column represents a gene within the module and the colour represents the magnitude of the correlation, with the default designation of red for positive correlation and green for negative correlation.

Fig 5: Co-expression network analysis of transcriptomes associated with growth, development and secondary metabolism during *Astragalus* roots harvesting.



(A) GO annotation of blue, brown and turquoise modules associated with root growth and development. (B) Calycosin 7-O-β-D-glucopyranoside content-related GO annotation in brown, yellow and pink modules. (C) Astragaloside IV content-related GO annotation in grey60 module. (D) KEGG enrichment of blue, brown and turquoise modules associated with root growth and development. (E) Calycosin 7-O-β-D-glucopyranoside content-related KEGG enrichment in brown, yellow and pink modules. (F) Astragaloside IV content-related KEGG enrichment in grey60 module.

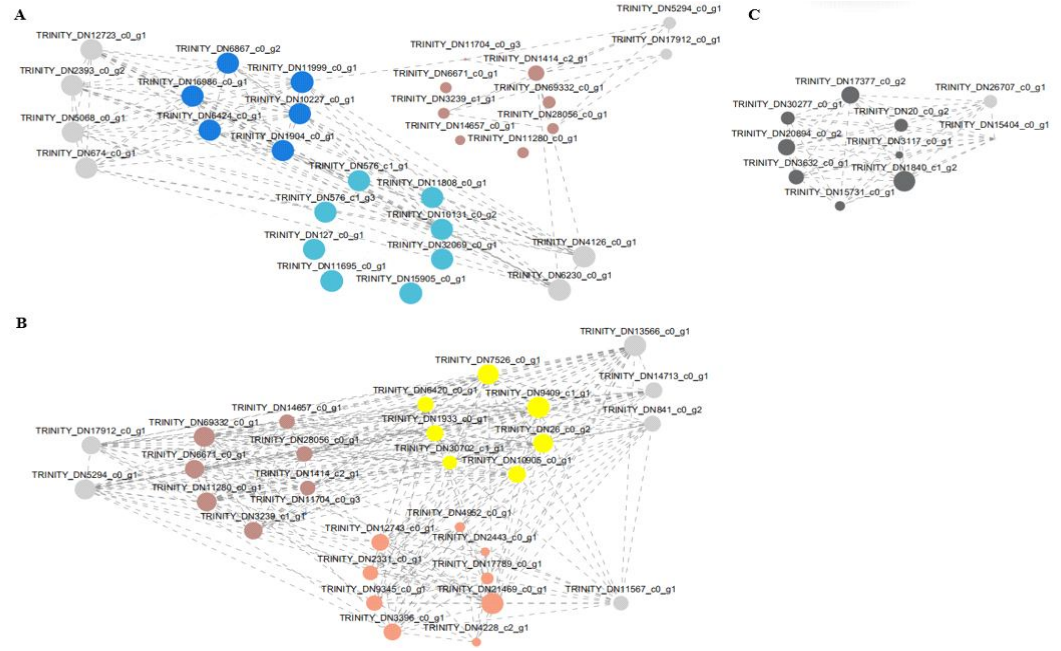
Fig 6: GO annotation of co-expressed genes and KEGG enrichment analysis of genes in target module.

We performed pathway enrichment analysis on growth and development and active ingredient-related modules and selected the top 20 metabolic pathways from each comparison. KEGG enrichment revealed that the modules related to growth and development were mainly enriched in plant hormone signal transduction, Glycosyl phosphatidylinositol (GPI)-anchor biosynthesis, Homologous recombination, Folate biosynthesis, Terpenoid backbone biosynthesis and Aminoacyl-tRNA biosynthesis (Fig 6D). The modules related to the content of calycosin 7-O-β-D-glucopyranoside were mainly enriched in Plant hormone signal transduction, Terpenoid backbone biosynthesis, Non-homologous end-joining, Phenylalanine, tyrosine and tryptophan biosynthesis and Plant-pathogen interaction (Fig 6E). The modules related to astragaloside IV content were mainly enriched in RNA degradation, Linoleic acid metabolism and N-Glycan biosynthesis (Fig 6F). Thus, we speculated that these pathways may play an important role in regulating the growth and development process of *Astragalus.*

Analysis of hub genes interaction network in the module

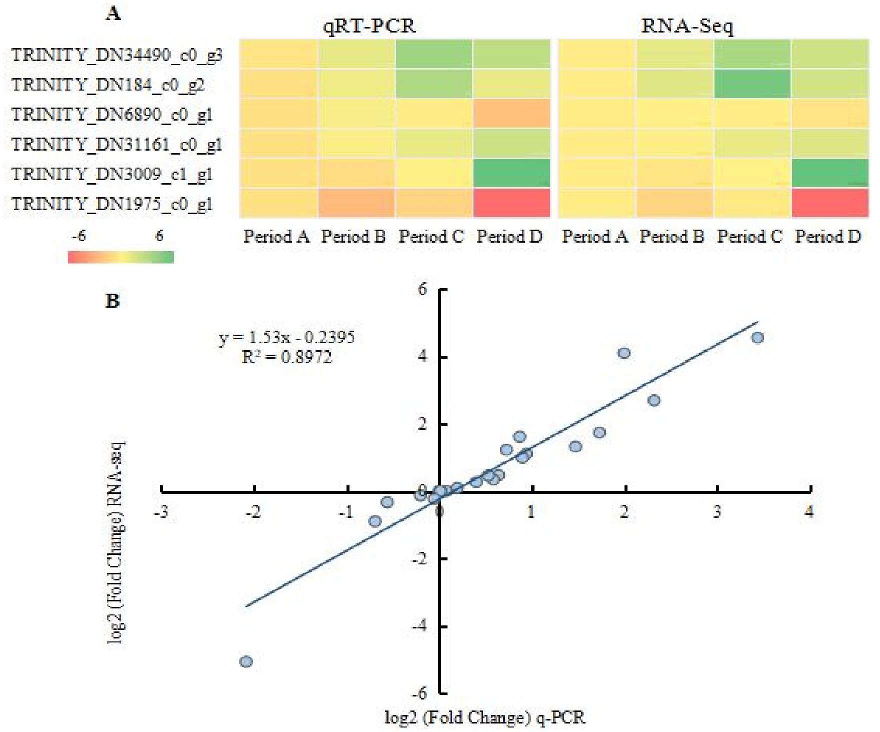
Mining and identification of key genes regulating root growth and development and analysing the regulatory pathways of plant growth and development are important elements of biological breeding of *Astragalus*. Members of the major plant transcription factor families, including the bHLH and AP2/ERF families, have been validated for their roles in the regulation of plant growth and development (Yue, 2022). However, genes and metabolites related to the regulation of root growth and development in *Astragalus* less well studied. Metabolomic and transcriptomic WGCNA analyses in this study indicated that 19 metabolites (Fig 4) and 8 annotated key genes (*CNX2*, *talA*, *ABCF2*, *HIBCH*, *crtZ*, *AXY4*, *TTC5* and *UNKL*) are essential for the regulation of growth development of *Astragalus* roots.

In this study, gene networks were visualised and gene connectivity analysed for genes in modules related to growth and development, calycosin 7-O-β-D-glucopyranoside content and astragaloside IV content in *Astragalus*. To determine the Hub genes related to growth development and secondary metabolism, we classified the ten genes with the highest kME values in each module as hub genes and utilized the hub genes and their interacting genes to map the gene co-expression network (Fig 7). 22 highly connectivity genes were screened by WGCNA analysis using 8 annotated genes (*CNX2*, *talA*, *ABCF2*, *HIBCH*, *crtZ*, *AXY4*, *TTC5*, *UNKL*) in the tissue-specific modules (blue brown turquoise) related to growth and development (Fig 7A). 6 annotated genes (*ETR*, *IPT*, *AP3B*, *talB*, *AXY4*, *CYP707A*) in the tissue-specific modules (brown yellow pink) related to calycosin 7-O-β-D-glucopyranoside content were used to screen 24 candidate genes with high connectivity (Fig 7B). 8 high connectivity candidates were



(A) Network interaction analysis of hub genes in growth and development-related modules (blue brown turquoise). (B) Network interaction analysis of hub genes in modules (brown yellow pink) related to the content of calycosin 7-O-β-D-glucopyranoside. (C) Network interaction analysis of hub genes in module (grey60) related to the content of astragaloside IV. The size gradient of the dots represents the high or low connectivity. Light gray dots represent genes that have been annotated.

Fig 7: Analysis of hub genes network interaction in phenotypic and active ingredients significant enrichment module.



(A) Heat map showing the expression changes (log2-fold change) in different harvesting periods for each candidate gene as measured by RNA-seq and qRT-PCR. (B) Linear regression between qRT-PCR and RNA-Seq.

Fig 8: The expression pattern of 6 selected genes identified by RNA-seq was verified by qRT-PCR.

screened using 2 annotated genes (*copA*, *JAR*) in the tissue-specific module related to astragaloside IV content (Fig 7C). 16 annotated high connectivity candidates and 54 related high connectivity candidates were screened in the seven modules.

Validation genes by qRT-PCR analysis

To evaluate the reliability of the gene expression profile of *Astragalus* at different harvesting periods, qRT-PCR was used to verify the gene expression level. We selected six genes: TRINITY_DN34490_c0_g3, TRINITY_DN184_c0_g2, TRINITY_DN6890_c0_g1, TRINITY_DN31161_c0_g1, TRINITY_DN3009_c1_g1 and TRINITY_DN1975_c0_g1 to be upregulated or down regulated under different harvesting periods. We found that the expression levels of these 6 genes in *Astragalus* herbs roots at different times were consistent with the transcriptome sequencing results, which indicated that our RNA-seq data were reliable (Fig 8).

CONCLUSION

In this study, metabolomic and transcriptomic WGCNA analyses were used to show the roles of 19 highly-connected hub metabolites and 16 highly-connected hub genes with annotations in the growth and development of the roots of *Astragalus* and in the accumulation of the two important active components. The 16 annotated hub genes were also used to screen 54 key genes with high connectivity related to growth development and secondary metabolism. Overall, these findings provide new insights into the regulation of growth, development and secondary metabolism in the roots of *Astragalus*, as well as the roles of specific structural genes and metabolites in the accumulation of herb yield and active ingredient content.

ACKNOWLEDGEMENT

Xiaojie Li and Xiaoming Zhang designed the experiments. Xiaojie Li and Yingtong Mu conducted the experiments. Xiaojie Li wrote the manuscript. Junjie Wang and Xiaoming Zhang reviewed the manuscript. All authors approved the paper. All co-authors reviewed the final version and approved the manuscript before submission. This work was supported by a grant from “Identification of high yielding and-quality germplasm and breeding technology of *Astragalus* Membranaceus Mongholicus” (2023-Society-2), “Efficient Breeding and Processing Technology Integration of Multifunctional Grass Seeds for Grassland” (2022YFDZ0025) and “Scientific and Technological Achievements Transformation Project of Inner Mongolia Autonomous Region” (CGZH2018134).

Conflict of 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.

REFERENCES

Aidi, Z., Qiuyun, Z., Jianzhao, L., Hansheng, G., Xinguang, F., Yanqing, Y., Xiaofen, L., Xueren, Y. (2020). Transcriptome co-expression network analysis identifies key genes and regulators of ripening kiwifruit ester biosynthesis. BMC Plant Biology. 20: 103. https://doi.org/10.1186/s12870-020-2314-9.

Chen, Y., Fang, T., Su, H., Duan, S.F., Ma, R.R., Wang, P., Wu, L., Sun, W.B., Hu, Q.C., Zhao, M.X., Sun, L.J., Dong, X.H. (2022). A reference-grade genome assembly for *Astragalus* mongholicus and insights into the biosynthesis and high accumulation of triterpenoids and flavonoids in its roots, Plant Commun. 4(2): 100469.

Diao, J., Liu, H., Hu, F. (2019). Transcriptome analysis of immune response in fat greenling (*Hexagrammos otakii*) against Vibrio harveyi infection. Fish and Shellfish Immunology. 84: 937-947.

Downs, G.S., Bi, Y.M., Colasanti, J., Wu, W., Chen, X., Zhu, T., Rothstein, S.J., Lukens, L.N. (2013). A developmental transcriptional network for maize defines coexpression modules. Plant Physiol. 161: 1830-1843.

Ji, X., Gao, Q., Zhuang, Z., Wang, Y., Zhang, Y., Peng, Y. (2022). Transcriptome-based weighted correlation network analysis of maize leaf angle regulation by exogenous brassinosteroid. Agronomy. 12: 18-95. https://doi.org/10.3390/agronomy12081895.

Li, W. (2020). Transcriptome Study of *Astragalus* Mongholicus Root and Screening and Expression of Key Genes for Flavonoids Biosynthesis. Shanxi Agricultural University. pp. 30-36.

Li, X.J., Mu, Y.T., Hua, M., Wang, J.J., Wang., Zhang, X.M. (2024). Integrated phenotypic, transcriptomics and metabolomics: Growth status and metabolite accumulation pattern of medicinal materials at different harvest periods of *Astragalus*. BMC Plant Biology. 24: 358.

Lijuan, F., Gongfa, S., Juan, Y. (2022). A protective role of phenylalanine ammonia-lyase from *Astragalus membranaceus* against saline-alkali stress. International Journal of Molecular Sciences. 23(24): 156-86.

Lin, X.Z., Zhang, Z.H., Yang, Q., Huang, S.W. (2017). Identification and characterization analysis of co-expression gene modules in cucumber (*Cucumis sativus* L.). Journal of Agricultural Biotechnology. 23: 1121-1130.

Na, X., Qiqaing, G., Huie, L., Gangyi, Y., Qin, G., Yang, X., Mengyun, L., Lan, Y. (2024). Integrated transcriptomic and WGCNA analyses reveal candidate genes regulating mainly flavonoid biosynthesis in *Litsea coreana* var. Sinensis. 24: 231. https://doi.org/10.1186/s12870-024-04949-1.

Pham, A.T., Eunsook, C., Aye, A.T., Li, X.H., Kim, Y.B., Mariadhas, V.A. (2015). Transcriptional profling and molecular characterization of astragalosides, calycosin and calycosin-7-O-β-D-glucoside biosynthesis in the hairy roots of *Astragalus* membranaceus in response to methyl jasmonate. J. Agric. Food Chem. 63(27): 6231-6240.

State Pharmacopeia Committee of China. (2020). Pharmacopoeia of the People's Republic of China. Bei Jing: Chemical Industry Press. Chapter 1: 315-316.

Tuan, P.A., Chung, E., Thwe, A.A. (2015). Transcriptional profiling and molecular characterization of astragalosides, calycosin and calycosin-7-O-β-D-glucoside biosynthesis in the hairy roots of *Astragalus membranaceus* in response to methyl jasmonate. J. Agric. Food Chem. 63: 40.

Wang, L.L., Yang, L.C., Xiong, F., Nie, X.Q., Xiao, Y.M., Li, C.B., Zhao, X.H., Zheng, Z., Zhou, G.Y. (2019). Effect of different harvest periods and cultivation density on growth, development and yield of *Astragalus mongolica.* Molecular Plant Breeding. 17(23): 7962-7968. doi: 10.13271/j.mpb.017.007962.

Wang, R.X., Li, C. (2019). Application of full-length transcriptome sequencing in non-model plant transcriptome research. Molecular Plant Breeding. 17(2): 502-508. doi: 10.13271/j.mpb.017.000502.

Wang, X., Wang, Y., An, J., Wang, Q., Jin, Y., Huang, J., Zhang, G., Yan, Y. (2023). Transcriptomic analysis of *Astragalus* mongolicum from different years and mining of key genes for synthesis of active components. Chinese Traditional and Herbal Drugs. 54(03): 915-925.

Wu, X., Li, X., Wang, W., Shan, Y., Wang, C., Zhu M, La Q, Li X. (2020). Integrated metabolomics and transcriptomics study of traditional herb *Astragalus membranaceus* Bge. var. *mongolicus* (Bge.) Hsiao reveals global metabolic profile and novel phytochemical ingredients. BMC Genomics. 21: 697. doi: 10.1186/s12864-020-07005-y.

Xiang, L., Zhang, Qiao, Y., Zhao, Q.M., Qin, L.P., Gong, W. (2022). Research progress on chemical constituents, pharmacological effects and clinical applications of Astragali Radix-*Angelicae sinensis* Radix. Chinese Traditional and Herbal Drugs. 53(07): 2196-2213. doi: 10.7501/j.issn.0253-2670.2022.07.030.

Xiao, L.M., Cao, P.H., Luo, Z.H. (2020). Cycloartane-type triterpenoids from the root of *Astragalus membranaceus* var. mongholicus. J. Asian Nat. Prod. Res. 22(10): 905-913.

Xiao, L.M., Cao, P.H., Luo, Z.H. (2020). Cycloartane-type triterpenoids from the root of *Astragalus membranaceus* var. mongholicus. J. Asian Nat. Prod. Res. 22(10): 905-913.

Xie, N., Guo, Q., Li, H., Yuan, G., Gui, Q., Xiao, Y., Liao, M., Yang, L. (2024). Integrated transcriptomic and WGCNA analyses reveal candidate genes regulating mainly flavonoid biosynthesis in *Litsea coreana* var. Sinensis. 24: 231. https://doi.org/10.1186/s12870-024-04949-1.

Xin, B., Xie, J., Wang, W.Q. (2015). Content determination of total polysaccharides and total flavonoids from radix astragali in different growth time. Information on Traditional Chinese Medicine. 32(5): 31-34. doi:10.19656/j.cnki.1002-2406.2015.05.009.

Yue, M.F., Zhang, C., Zheng, D.Y., Zou, H.W., Wu, Z.Y. (2022). Response of maize transcriptional factor ZmbHLH91 to abiotic stress. Acta Agron Sin. 48: 3004-3017.

Zhang, F., Zhang, X., Luo, Y., Li, H., Qin, X. (2022). Biosynthetic mechanisms of isoflavone accumulation affected by different growth patterns in *Astragalus mongholicus* products. BMC Plant Biol. 22: 410. doi:10.1186/s12870-022-03769-5.

Zhang, Y. (2020). Analysis of *Astragalus* transcriptome and expression of key genes in saponin synthesis pathway. Shanxi Agricultural University. pp. 36-43.