

RESEARCH

Open Access



Integrated metabolomics and transcriptomics study of traditional herb *Astragalus membranaceus* Bge. var. *mongolicus* (Bge.) Hsiao reveals global metabolic profile and novel phytochemical ingredients

Xueting Wu^{1†}, Xueting Li^{1,2†}, Wei Wang^{3†}, Yuanhong Shan¹, Cuiting Wang¹, Mulan Zhu^{1,4}, Qiong La⁵, Yang Zhong^{3,5}, Ye Xu^{6*}, Peng Nan^{3*} and Xuan Li^{1,2*}

From The 18th Asia Pacific Bioinformatics Conference
Seoul, Korea. 18-20 August 2020

Abstract

Background: *Astragalus membranaceus* Bge. var. *mongolicus* (Bge.) Hsiao is one of the most common herbs widely used in South and East Asia, to enhance people's health and reinforce vital energy. Despite its prevalence, however, the knowledge about phytochemical compositions and metabolite biosynthesis in *Astragalus membranaceus* Bge. var. *mongolicus* (Bge.) Hsiao is very limited.

(Continued on next page)

* Correspondence: xu.shirley021@gmail.com; nanpeng@fudan.edu.cn; lixuan@sippe.ac.cn

[†]Xueting Wu, Xueting Li and Wei Wang contributed equally to this work.

⁶Department of Colorectal Surgery, Fudan University Shanghai Cancer Center, Shanghai, China

³Ministry of Education Key Laboratory for Biodiversity Science and Ecological Engineering, School of Life Sciences, Fudan University, Shanghai 200438, China

¹Key Laboratory of Synthetic Biology, CAS Center for Excellence in Molecular Plant Sciences, Institute of Plant Physiology and Ecology, Chinese Academy of Sciences, Shanghai 200032, China

Full list of author information is available at the end of the article



© The Author(s). 2020 **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 Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

(Continued from previous page)

Results: An integrated metabolomics and transcriptomics analysis using state-of-the-art UPLC-Q-Orbitrap mass spectrometer and advanced bioinformatics pipeline were conducted to study global metabolic profiles and phytochemical ingredients/biosynthesis in *Astragalus membranaceus* Bge. var. *mongolicus* (Bge.) Hsiao. A total of 5435 metabolites were detected, from which 2190 were annotated, representing an order of magnitude increase over previously known. Metabolic profiling of *Astragalus membranaceus* Bge. var. *mongolicus* (Bge.) Hsiao tissues found contents and synthetic enzymes for phytochemicals were significantly higher in leaf and stem in general, whereas the contents of the main bioactive ingredients were significantly enriched in root, underlying the value of root in herbal remedies. Using integrated metabolomics and transcriptomics data, we illustrated the complete pathways of phenylpropanoid biosynthesis, flavonoid biosynthesis, and isoflavonoid biosynthesis, in which some were first reported in the herb. More importantly, we discovered novel flavonoid derivatives using informatics method for neutral loss scan, in addition to inferring their likely synthesis pathways in *Astragalus membranaceus* Bge. var. *mongolicus* (Bge.) Hsiao.

Conclusions: The current study represents the most comprehensive metabolomics and transcriptomics analysis on traditional herb *Astragalus membranaceus* Bge. var. *mongolicus* (Bge.) Hsiao. We demonstrated our integrated metabolomics and transcriptomics approach offers great potentials in discovering novel metabolite structure and associated synthesis pathways. This study provides novel insights into the phytochemical ingredients, metabolite biosynthesis, and complex metabolic network in herbs, highlighting the rich natural resource and nutritional value of traditional herbal plants.

Keywords: *Astragalus membranaceus* Bge. Var. *mongolicus*, Non-targeted metabolomics, Phytochemical compositions, Secondary metabolites, Flavonoid derivative, Biosynthesis

Background

As a member of the legume family, *Astragalus membranaceus* Bge. var. *mongolicus* (Bge.) Hsiao (AMM for short) is a perennial herbaceous plant grown in northwestern China, having a plant-to-harvest cycle of 2 to 3 years [1, 2]. AMM is one of the common herbs that is widely used in the South and East Asia, to enhance people's health and reinforce vital energy. The dried root of AMM, the main herbal material, was reported to have anti-perspirant, anti-diuretic, antitumor, anti-oxidation, and anti-inflammation effect [1, 3–6]. The major bioactive components of AMM are a variety of flavonoids, terpenoids, polysaccharides, amino acids, etc. Up-to date, about 150 different metabolites were identified from AMM, including saponins, flavonoids, polysaccharides, and amino acids [7]. Calycosin, calycosin-7-O-b-D-glucoside (CG), and astragaloside IV (ASI) are considered major bioactive constituents from AMM [8–10]. Calycosin and CG are isoflavones that were reported to have anti-inflammatory, anti-radiation, anti-cancer, and anti-microbial activities, and act as antioxidant and adjuvant agents [11–13]. The effect of CG was also reported as a hyaluronidase inhibitory component that was used in anti-osteoarthritis treatment [12, 13]. ASI, a natural triterpenoid saponins from AMM, was reported to have effects of anti-fatigue, anti-cancer, anti-coxsackie B virus, etc. [10, 14]. Indeed, these ingredients were used as standard compositions to evaluate the quality of AMM harvests.

Despite being one of the most important herbs, knowledge about AMM phytochemical compositions and metabolite biosynthesis is very limited. Metabolomics seeks

to provide a comprehensive profile of all metabolites present in a biological sample under certain conditions [15, 16]. With the advance of mass spectrum (MS) technology, plant metabolomics presents a new perspective to understand the chemical compositions and functions of the traditional herb like AMM. In the past decades, the study of metabolomics of AMM was based on targeted analysis [17, 18] focusing on some of the active ingredients of AMM, including astragalosides, flavones, and polysaccharides. Recently, liquid chromatography-mass spectrum (LC-MS) has been widely applied in metabolomics studies, with technical advance in wider range of separation, higher selectivity, and higher sensitivity [19]. Metabolomics analyses using advanced LC-MS systems were performed on a number of plants, including *Arabidopsis* [20, 21], *Oryza sativa* [22, 23], Tomato [24, 25], which was proven effective for metabolite identification and pathway elucidation in plants [26, 27]. In metabolomics analysis, tandem MS generates informative fragment peaks, forming fingerprints specific to the detected molecules [28]. Annotation of detected molecules from their fingerprints has been a major challenge for MS studies, because of the limited data from reference mass spectra. Recently, computational approach to simulate compound fragmentation was developed for annotating MS data. A number of in silico tools, like MetFrag [28], FingerID [29], CFM-ID [30], and MetFusion [31], were launched to allow users to query broader chemical structure databases, like KEGG, PubChem, etc. [32].

A comprehensive metabolomics and transcriptomics analysis using advanced LC-MS combined with

transcriptome analysis is desired to study AMM and understand their chemical compositions and metabolite biosynthesis in vivo. In current study, we performed a non-targeted metabolomics analysis on three AMM tissues: leaf, root, and stem. An unprecedented number of stable metabolites (5435) were discovered in AMM, and 2190 metabolites were annotated by our integrated bioinformatics pipeline. Metabolic profiling found that metabolites have different accumulation patterns in AMM tissues. We also performed transcriptome analysis to explore the mechanism underlying the metabolic regulation among three tissues in AMM. The three tissues also presented different expression patterns, and pointed to importance of all the tissues for the production of bioactive compounds. Furthermore, we inferred the complete pathways of Phenylpropanoid, Flavonoid, and Isoflavonoid biosynthesis in AMM by integrating its metabolomics and transcriptomics data, including critical enzymes and metabolites. We further discovered novel derivatives for secondary metabolites, like isoflavonoids, from modifications with various chemical groups. This study reveals the comprehensive profile of metabolic activities on herb AMM, providing novel insights into phytochemical ingredients and metabolite biosynthesis in traditional herbs.

Results and discussion

Metabolomics analysis revealing abundant metabolites in AMM tissues using an integrated informatics pipeline

While root is the main herbal material from AMM, it was found that secondary metabolites was produced in various AMM tissues [2, 33]. We carried out a deep metabolomics study on the different tissues of AMM, i.e. leaf, root and stem, using non-targeted MS-analysis protocol with state-of-art UPLC-Q-Orbitrap mass spectrometer and an integrated informatics pipeline [34] (*Methods*, Additional file 1). Mass spectral data were processed by alignment of all data-sets from the AMM tissues (each with three biological replicates) and controls (*Methods*). After removing redundancy and noise signals, which are of poor quality or non-biological origin [35], 11,101 and 9250 spectra signals were retained for positive and negative mode respectively. The spectra signals for positive and negative mode were merged to obtain a total of 5435 significant peaks from AMM tissues, for which 3985 were tagged with MS2 spectrum (Additional file 2).

Annotation of metabolites with MS2 tags from AMM was performed by comparison of the accurate m/z values and the fragmentation patterns based on the metabolites above [36]. We annotated the metabolites by matching their spectra data against experimental MS reference databases (similarity score threshold 0.8; *Methods*). Those do not have any match were subsequently annotated with

virtual MS reference databases. Taken together, 2190 metabolites from AMM were annotated by our integrated bioinformatics pipeline (Additional file 3). Among them, the most abundant are Terpenoids (254), followed by Flavonoids (101), Alkaloids (97), Phenylpropanoids (57), and Fatty acids related compounds (35) (Table 1). We found an unprecedented number of metabolites from AMM, for which we have increased the number of annotated metabolites by an order of magnitude over previous studies. Note many metabolites previously identified from AMM were included in our list (Additional file 4).

To understand the metabolic activities in AMM, we mapped the annotated metabolites to KEGG pathways. 1330 metabolites were mapped to KEGG and assigned to 247 reference canonical KEGG pathways, which including 144 metabolites in the pathway of 'Biosynthesis of secondary metabolites' (Additional file 5). These metabolites cover most of central metabolism and reflect the physiological state and the edible and nutritive value of AMM.

Transcriptomics analysis of gene expression in AMM tissues

Transcriptomics analysis was conducted by RNA-seq on AMM tissues, leaf, stem, and root with three replicates for each tissue (Additional file 1). A total of 3.60, 3.62 and 3.59 Gb RNA-seq data were generated for leaf, stem, and root, respectively (Fig. 1a). Clean reads were mapped to reference transcriptome [2] with an average mapping ratio ~ 89.0%, validating that our RNA-seq data has high sequencing precision. There were on average 65,000 uni-genes expressed (FPKM ≥ 0.5) in three AMM tissues. Although each tissue had similar distribution of gene expression levels, denoted by their FPKM values (Fig. 1b), we found leaf and stem had differentially expressed genes from root, particularly in some secondary metabolic pathways, which points to important roles of AMM leaf and stem in production of bioactive compounds.

Table 1 Classification of annotated metabolites from AMM

category	counts
Terpenoids	254
Flavonoids	101
Alkaloids	97
Phenylpropanoids	57
Fatty acids related compounds	35
Skimate / acetate-malonate pathway derived compounds	11
Polyketides	10
Amino acid related compounds	4

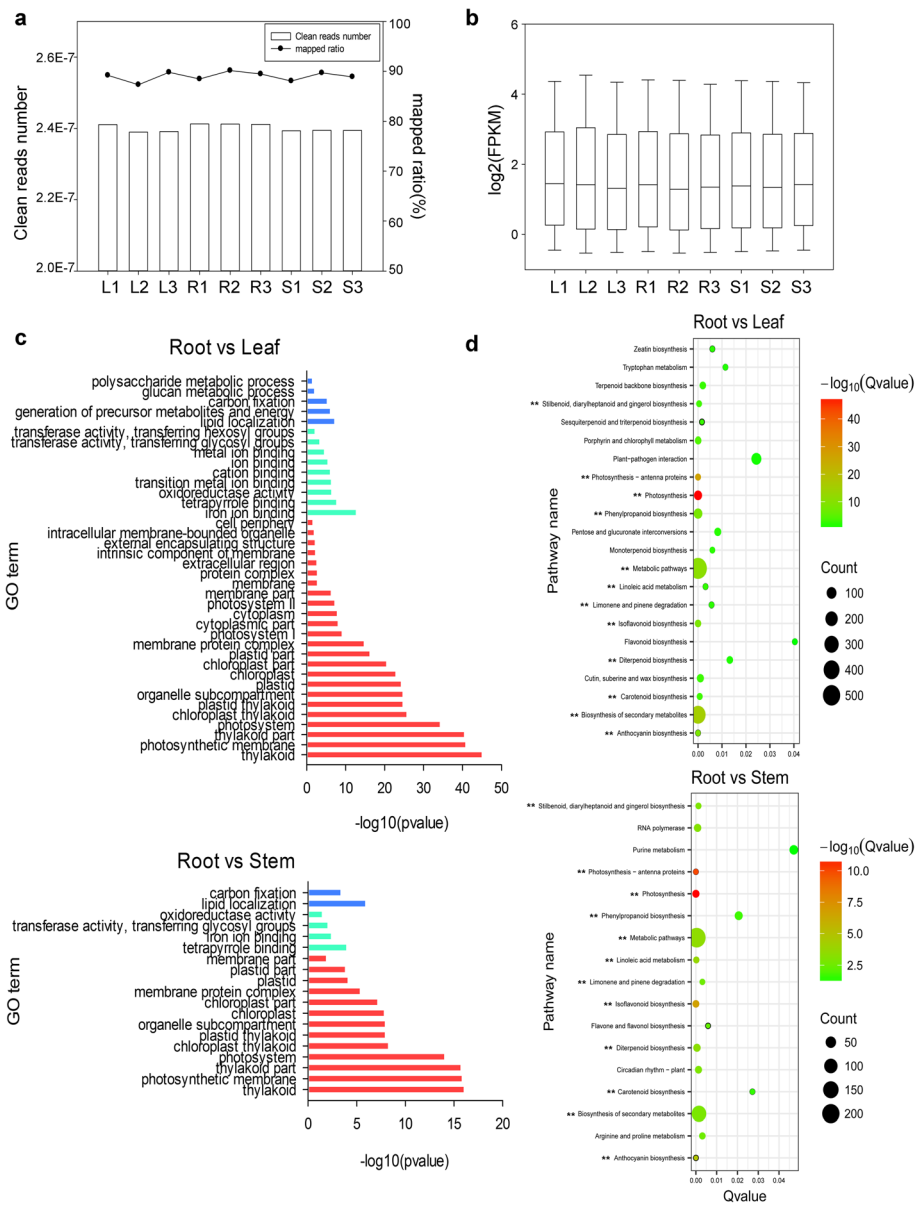


Fig. 1 Functional annotation and classification of differentially expressed unigenes between different tissues of AMM. **a** The reads number and mapped ratio to the reference transcriptome of different tissues (L1-L3:leaf; R1-R3:root; S1-S3: stem). The bar plot represents the clean reads number in different tissues, and the dots plot represents mapped ratio of different tissues to the reference transcriptome in different tissues. **b** The expression level of unigenes in different tissues from AMM (FPKM ≥ 0.5). We took the logarithm of each FPKM value of all unigenes. **c** Gene Ontology (GO) functional classifications of differentially expressed unigenes (Only list significantly enriched GO terms, P value ≤ 0.05). Red: CC Cellular Component; Green: MF Molecular Function; Blue: BP Biological Process. **d** Significantly enriched KEGG pathways between differentially expressed unigenes. The Q value denoted the corrected P -value (Significant pathways were identified by Q value ≤ 0.05). Count denoted the number of differentially expressed unigenes mapped to a certain pathway according to KEGG database. The pathways with ^{***} denoted the significantly enriched pathways that included in both between the root vs leaf and root vs stem analysis

Compared to root transcriptome, there were 3629, and 1358, unigenes differentially expressed in leaf and stem, respectively (Additional file 6, Additional file 7A and B). To understand their functions in leaf and stem, GO analysis were performed to assign them to biological process, cell component, or molecular function.

Compared to root, the most significant differences in leaf and stem were related to chloroplast/thylakoid structure and functions (Fig. 1c). Again, KEGG analysis indicated leaf and stem have significantly altered activities in the biosynthesis pathways of many secondary metabolites, like isoflavonoids and terpenoids (Fig. 1d).

Profiling metabolic activities of AMM tissues by integrated metabolomics and transcriptomics analysis

The metabolic activities of AMM tissues were analyzed by combining metabolomics and transcriptomics data. First, the levels of the 5435 metabolic features in AMM were defined for leaf, root and stem. Surprisingly we found leaf had the most abundant metabolites, whereas root had the least (Additional file 8A). This was consistent with the study that showed leaf had the highest number of expressed unigenes in many metabolic pathways [2]. The relative intensity of the metabolites displayed similar pattern, in which root had significant lower quantities of metabolites than either leaf or stem (Mann-Whitney Rank Sum Test, $P < 0.001$) (Additional file 8A).

Principal Component Analysis (PCA) was carried out as an unsupervised analysis on the metabolomic profiles for AMM tissues (*Methods*). The three AMM tissues were well separated by the first and second major components, for which 46.1 and 21.9% of the variance were explained by the first two main principal components, respectively (Additional file 8B). The result indicated systematic difference in metabolites among the AMM tissues. To understand the major differences in metabolite levels between leaf and root, and stem and root, we then conducted orthogonal partial least squares discriminate analysis (OPLS-DA) [37] and determined the featured metabolites for them. For differences between leaf and root, the tissue samples were well separated in the model (Additional file 8C). The R^2X , R^2Y (goodness-of-fit parameter) and Q^2 (predictive ability parameter) of the OPLS-DA model are 0.814, 1, and 0.995, respectively, indicating good quality and high confidence of our model. S-plot was used to find the featured metabolites distinguishing leaf and root (Additional file 8D). We then performed VIP (variable importance in the projection) prediction with a permutation test ($n = 200$) to validate the models' reliability (Additional file 8E), and obtained 2020 significantly differential metabolites (Additional file 7C) with threshold VIP value > 1 . The differential metabolites between leaf and root included Chikusetsusaponin V, Astragaloside IV, Isoastragaloside II, Formononetin, etc. Similar analysis was also carried out between stem and root (Additional file 9, 2,050 significantly differential metabolites, Additional file 7D). The differential metabolites between stem and root included Chikusetsusaponin V, Astragaloside IV, Quillaic acid, Soyasapogenol B, Oleanolic acid, etc.

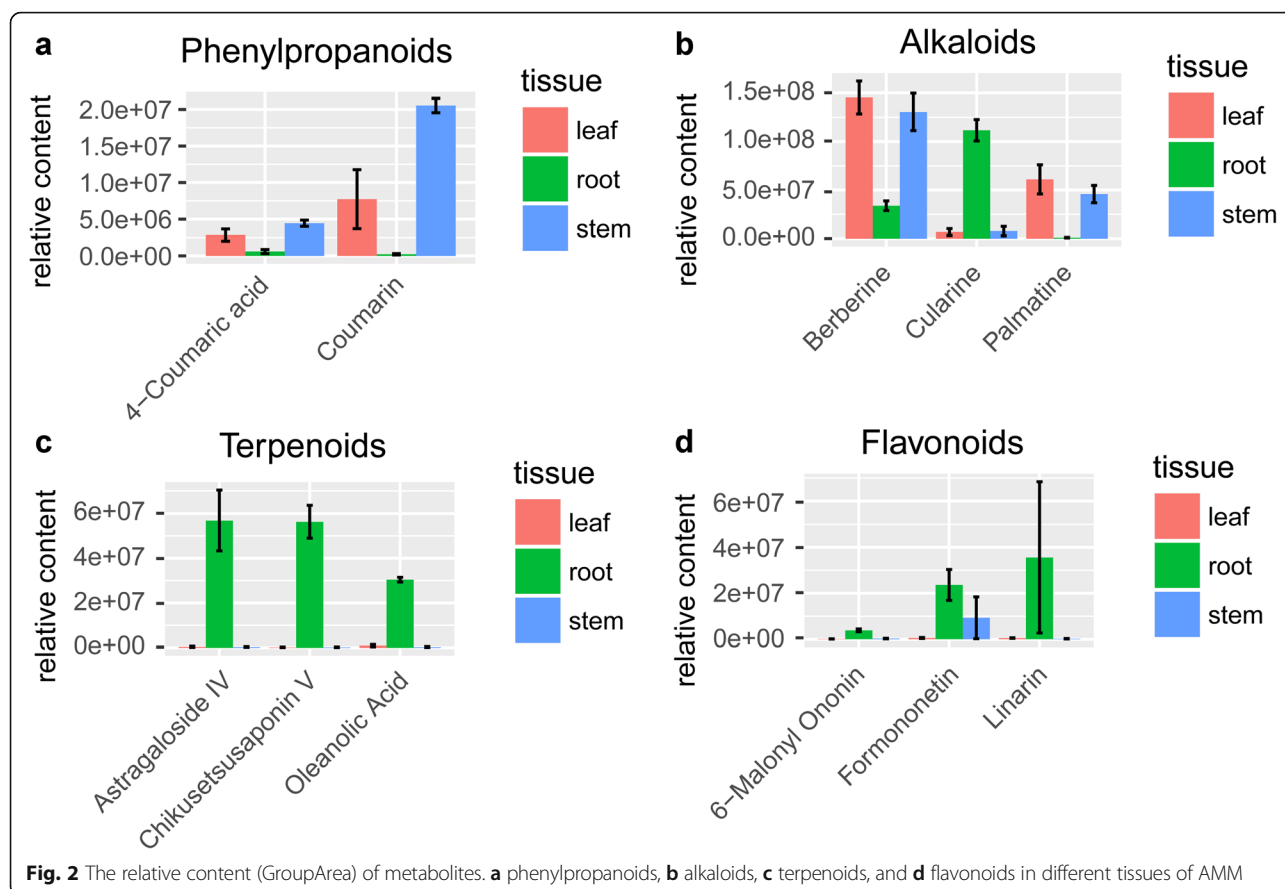
Phenylpropanoids

The phenylpropanoid biosynthesis is upstream of the flavonoid biosynthesis pathway [38]. There are 57 metabolites from AMM that were annotated as phenylpropanoids (Additional file 3). In general, the average content of phenylpropanoid in the root was significantly lower than in

the stem (Mann-Whitney Rank Sum Test $P < 0.001$ (root vs stem)), meanwhile the average content of phenylpropanoid in the root was significantly lower than in the leaf, although statistical significance was not reached (Additional file 10A). Among these phenylpropanoid metabolites, 22 were mapped to the phenylpropanoid biosynthesis pathway. In phenylpropanoid biosynthesis pathway, Coumaric acid, sinapic acid, and cinnamic acid, are precursors for coumarin synthesis [39]. Coumarins are a large family of naturally occurring substances of fused benzene and α -pyrone rings that were found primarily in popular medicines [40]. Coumarins are regarded as phytoalexins [41], which are notable for their role as anticancer, anti-inflammatory, antimicrobial, anti-oxidant and anti-coagulant bioactive compounds [42–44]. The contents of cinnamic acid and 4-Coumaric acid were much more abundant in the leaf and stem than that in the root of AMM (Fig. 2a). Cinnamic acid is transformed into Cinnamoyl-CoA and coumaryol-CoA, which are direct precursors for flavonoid biosynthesis.

Alkaloids

There are 97 alkaloids annotated for AMM (Additional file 3). These alkaloids account for different types of alkaloids: Isoquinoline alkaloids (21), Amines (14), Pyridine alkaloids (7), Piperidine alkaloids (11), etc. They were mapped to KEGG pathways, including biosynthesis of alkaloids derived from shikimate pathway (19), Tropane, piperidine and pyridine alkaloid biosynthesis (18), isoquinoline alkaloid biosynthesis (16), etc. In general, the relative contents of alkaloids in root were lower than the other two tissues, leaf or stem (Additional file 10B), although statistical significance was not reached. Despite, the relative content of Cularine in the root was 16.73 and 33.96 times the amount of that in the leaf and stem, respectively (Fig. 2b). Cularine was reported to have non-specific antispasmodic activity on guinea-pig and human airways [45]. On the other hand, for Palmatine, a protoberberine alkaloids (isoquinoline alkaloid), the relative content in the leaf and stem were 79.37 and 45.24 times the amount of that in the root (Fig. 2b). Palmatine exerts a wide range of functions in vivo, with sedative effect, broad-spectrum antibacterial property, antioxidant activities [46], immunity enhancement, and relief of diabetic neuropathic pain and depression [46–48]. Another protoberberine alkaloids, Berberine, which also had a significantly higher relative content in the leaf than in the root and stem (Fig. 2b), demonstrated efficacy in anti-diabetic, anti-inflammatory, anti-dyslipidemia [49], and antioxidant activities [46]. These results suggested that other tissues besides AMM root may have potential nutritive value in functions like blood circulation, glucose regulation, and immunity enhancement [50].



Terpenoids

There are 254 terpenoids annotated for AMM in our study. They consisted of 12 monoterpenoids, 57 sesquiterpenoids, 81 diterpenoids, and 104 triterpenoids. They were mapped mainly to the monoterpenoid biosynthesis, diterpenoid biosynthesis, sesquiterpenoid and triterpenoid biosynthesis, and biosynthesis of terpenoids and steroids pathways. Their average content in the root was significantly lower than in leaf or stem (Mann-Whitney Rank Sum Test: $p < 0.001$ (root vs leaf); $p < 0.001$ (root vs stem)) (Additional file 10C). These results agree with the previous study that leaf had a higher content of a few terpenoids (saponins) than the root [50]. Our analysis further identified that at least 23 terpenoids accumulated significantly higher in root than in leaf or stem (VIP > 1 and fold change ≥ 2), respectively.

Astragaloside IV (AMM5250) was known as the main active components of AMM and plays a role in anti-inflammatory, antioxidant, regulating energy metabolism, protectionnervous, anti-cancer [51], treating diabetes, etc. [52]. Astragaloside IV in the root, representing the highest content among differential terpenoids, was 219.49 and 194.77 times the amount of that in leaf and stem, respectively (Fig. 2c). Chikusetsusaponin V (AMM5423) in the root, not reported in AMM, was the second highest

content among differential terpenoids, which was 343.74 and 308.81 times the amount of that in leaf and stem, respectively (Fig. 2c). Chikusetsusaponin V was reported to exhibit neuroprotective function [53], inhibit inflammatory responses, and reduce blood lipid [54, 55]. In addition, Oleanolic acid (AMM4067) in root was 32.90 and 205.79 times the amount of that in leaf and stem, respectively (Fig. 2c). Oleanolic acid was widely used for treating hepatopathy, as it can protect the liver from liver damage [56]. It is also an ingredient in skin care products to repair damaged cells and promote cell regeneration [56]. Oleanolic acid and its derivatives were also suggested to have bioactivities of antiosteoporosis, antidiabetes, antibacterial, anticancer and hemolytic effects [56]. The triterpenoids in AMM root were found to mostly belong to Oleananes and Protostanes, which we found to have mainly hexose group for glycosylation modification. Oleananes and Protostanes were reported to have function in regulating lymphocyte proliferation [57], in inflammatory response [58], and anticancer activities [58, 59].

Terpenoids biosynthetic pathway

In higher plants, the biosynthesis of triterpenoid saponins remains unsolved. The terpenoid backbone biosynthesis pathway was the base of the biosynthesis of

triterpenoid saponins. Hence, we first investigated the AMM genes involved in the terpenoid backbone biosynthesis pathway (Fig. 3a), and found all the predicted enzymes involved in the pathway. Additionally, we found different patterns of MVA and MEP/DOXP pathway genes in AMM tissues. While four enzymes in the MVA pathway were highly expressed in root and stem, eight in the MEP/DOXP pathway were highly expressed in leaf, consistent with the action of non-mevalonate (MEP/DOXP) pathway in chloroplasts (Fig. 3b) [2].

We found some high levels of astragalosides in root and stem, where MVA pathway was dominant. Downstream the terpenoid backbone biosynthesis, astragalosides (triterpenoid) were synthesized via the sesquiterpenoid and triterpenoid biosynthesis pathway (Fig. 3a). The sesquiterpenoid and triterpenoid biosynthesis pathway includes cycloartenol synthase (CAS1), cytochromeP450 (P450), and glycosyl transferase(GT). We discovered four isoforms for CAS1, p48468.0, p62650.0, p69652.1 and p75841.1, for which three are new ones. Among these, p48468.0 and p62650.0 were expressed specifically in root. p69652.1 was highly expressed in root and stem, whereas p75841.1 was similarly expressed in three tissues. On other hand, the content of ~20 monoterpenoids and diterpenoids are higher in leaf than in root and stem (VIP > 1 and fold change ≥ 2). This suggested that monoterpenoids and diterpenoids are synthesized through the MEP/DOXP pathway, which is dominant in leaf.

Flavonoids

Flavonoids, especially its subclass isoflavonoids, were suggested to possess antibacterial and antioxidant functions [2]. They have a wide range of clinical usage, e.g. prevention of cancer and neurodegenerative diseases [60]. There are 101 flavonoids annotated for AMM, including flavones, isoflavone, flavanones, chalcones, isoflavanones, flavonols, and anthocyanins. We found glycosylation is the most frequent modification in flavonoids, and the glycosylation groups are mainly glucoside, and rhamnoside. The average content of flavonoids in root, in general, was significantly lower than in leaf and stem (Mann-Whitney Rank Sum Test: $p < 0.001$ (root vs leaf), $p < 0.001$ (root vs stem)) (Additional file 10D).

Despite the general lower contents of flavonoid in the root, we found five flavonoids accumulated significantly higher in root (VIP > 1 and fold change ≥ 2) than in leaf and stem. Formononetin [61] had the highest relative content in differential flavonoid metabolites, and was more abundant in root, which has 39.09 and 65.49 times the amount of that in leaf and stem, respectively (Fig. 2d). Formononetin was reported to have anti-angiogenic effect in treatment of colon cancer cells in vitro and in vivo [62].

Formononetin 7-O-glucoside-6''-O-malonate (6-malonyl ononin), a derivative of formononetin, in root was 24.24 and 8.98 times the amount of that in leaf and stem (Fig. 2d). Linarin (AMM4734) in root was 12.52 and 23.84 times the amount of that in leaf and stem, respectively (Fig. 2d). It was reported to have functions of anti-aging, anti-hypoxia, inhibition of hyperglycemia, anti-acetylcholinesterase and neuroprotective [63, 64].

Phenylpropanoid, flavonoid, and Isoflavonoid biosynthetic pathways

Although some of the flavonoids were sporadically reported for AMM, we sought to assay and present the complete distribution of flavonoids synthesis in different AMM tissues. Flavonoids are synthesized from condensation of phenylpropanoid derivatives with malonyl-CoA. By integrating the metabolomic and the transcriptomic data, we inferred the complete pathways for flavonoids biosynthesis in AMM, and filled in many gaps in previous works. Downstream of Coumaroyl-CoA production, we for the first time, found all three branches of isoflavonoids biosynthesis in AMM, completing the picture of biosynthesis of Calycosin, Biochanin A, and Glycitein in AMM (Fig. 4a). The latter two were reported for the first time in AMM.

First, in the 'Calycosin' branch that we defined previously in AMM [2], in addition to CG, we identified (-)-Maackiain-3-O-glucoside (also named Trifolirhizin) that was not reported in AMM before. It confirmed the activity of I'2H predicted in AMM. Second, the 'Biochanin A' branch that we newly defined in AMM, started with Naringenin chalcone, and completed with Biochanin A and its various derivatives. Isoflavone 7-O-glucosyltransferase (IF7GT), the critical enzyme in the pathway, was first identified for AMM by our current study (Fig. 4a). The relative contents of metabolites in this branch were mostly higher in leaf than in other two tissues. Third, the 'Glycitein' branch started with liquiritigenin and completed with Glycitein 7-O-glucoside (also known as glycitin). The pathway was completed characterized in AMM with metabolite contents defined for all tissues and newly identified F6H and IF7GT for AMM. We found the relative intensities of metabolites were, in general, consistent with the expression level of enzymes (Fig. 4b).

However, we observed more complex situations, like I3'H and UCGT that synthesize Calycosin and CG. They have many isoforms with different expression patterns in AMM tissues. Two isoforms of I3'H, p69684.0 and p74367.1, and one isoform of UCGT, p60303.0, were highly expressed in root, whereas other I3'H isoforms, p48074.0 and p72589.1, and UCGT isoforms, p74367.0 and p68729.0, were highly

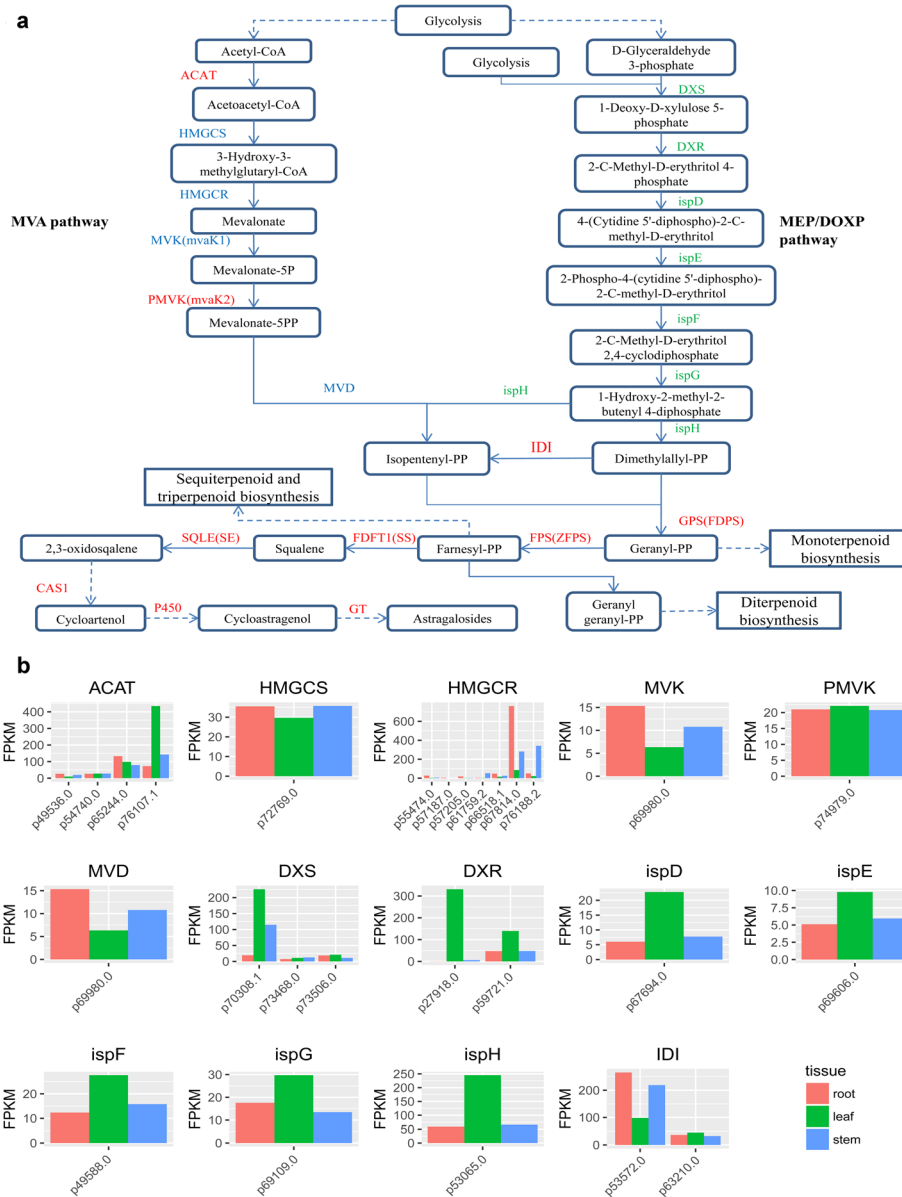


Fig. 3 The inferred pathways for Terpenoids biosynthesis in Mongolicus. **a** The Terpenoids biosynthesis pathway. The enzymes confirmed by transcriptomics data are shown in red. Blue indicate the highly expressed enzymes in root and stem, while green indicate the highly expressed in leaf. ACAT: acetyl-CoA C-acetyltransferase; HMGCS: hydroxymethylglutaryl-CoA synthase; HMGCR: hydroxymethylglutaryl-CoA reductase; MVK: mevalonate kinase; PMVK: phosphomevalonate kinase; MVD: diphosphomevalonate decarboxylase; DXS: 1-deoxy-D-xylulose 5-phosphate synthase; DXR: 1-deoxy-D-xylulose-5-phosphate reductoisomerase; ispD: 2-C-methyl-D-erythritol 4-phosphate cytidyltransferase; ispE: 4-diphosphocytidyl-2-C-methyl-D-erythritol kinase; ispF: 2-C-methyl-D-erythritol 2,4-cyclodiphosphate synthase; ispG: (E)-4-hydroxy-3-methylbut-2-enyl-diphosphate synthase; ispH: 4-hydroxy-3-methylbut-2-en-1-yl diphosphatereductase; IDI: isopentenyl-diphosphate Delta-isomerase. FDPS:farnesylidiphosphate synthase; ZFPS:(2Z,6Z)-farnesylidiphosphate synthase; FDFTH1: farnesyl-diphosphatefarnesyltransferase; SQLE: squalenemonooxygenase. CAS1: cycloartenol synthase; P450: cytochromeP450; GT: glycosyltransferase. **b** The expression of the genes (enzymes) in the pathway measured by FPKM value

expressed in leaf. They might contribute to the synthesis of Calycosin and CG differently in root or leaf due to unequal enzymatic activities. Another likely explanation for the relative higher intensity of Calycosin and CG in root is because of the transportation and storage mechanism in root [65].

Identification of novel flavonoid derivatives and their inferred pathways

Biochanin A was glycosylated to produce Biochanin A 7-O -D-glucoside (also known as Sissotrin) (Fig. 5). Glycitein was subjected to similar modifications in AMM, generating Glycitein 7-O-glucoside (Fig. 5a).

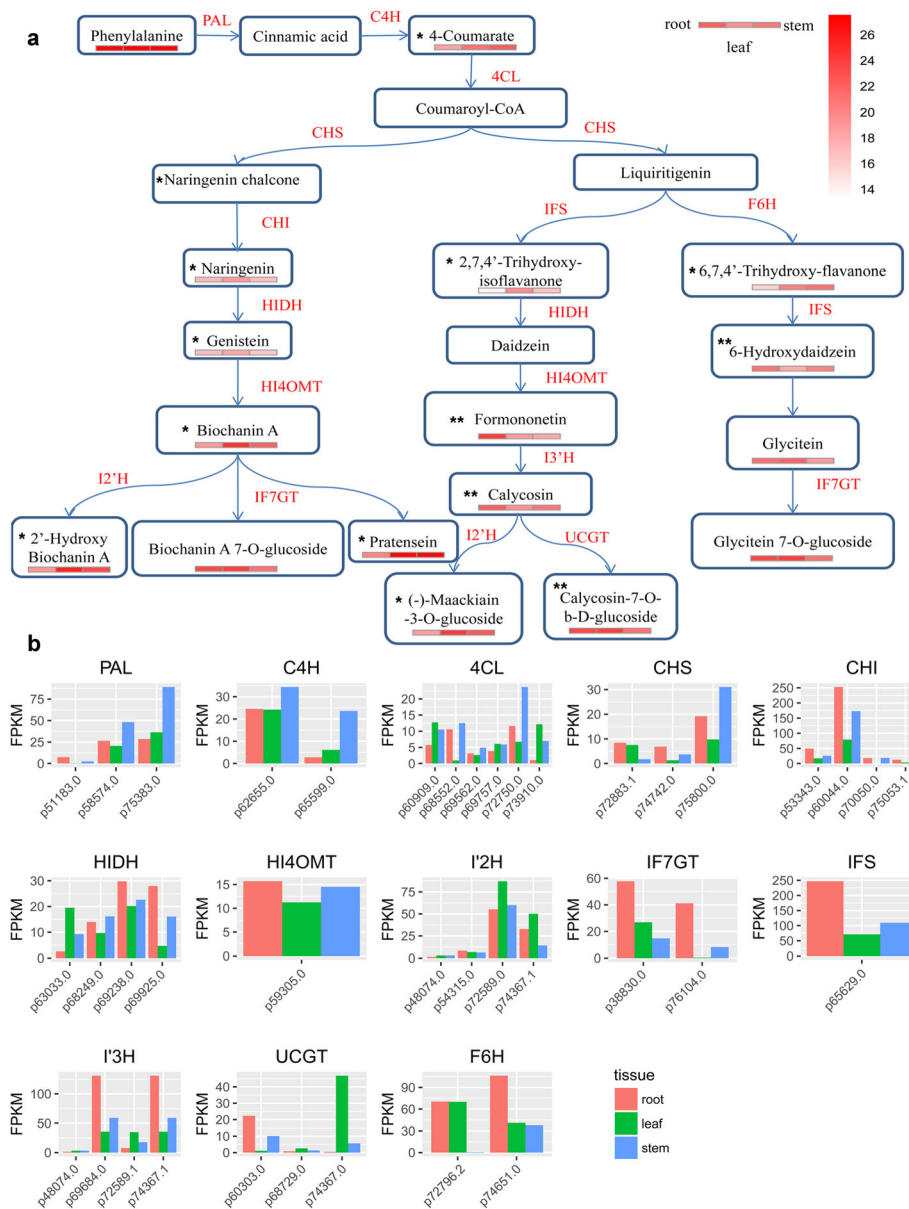
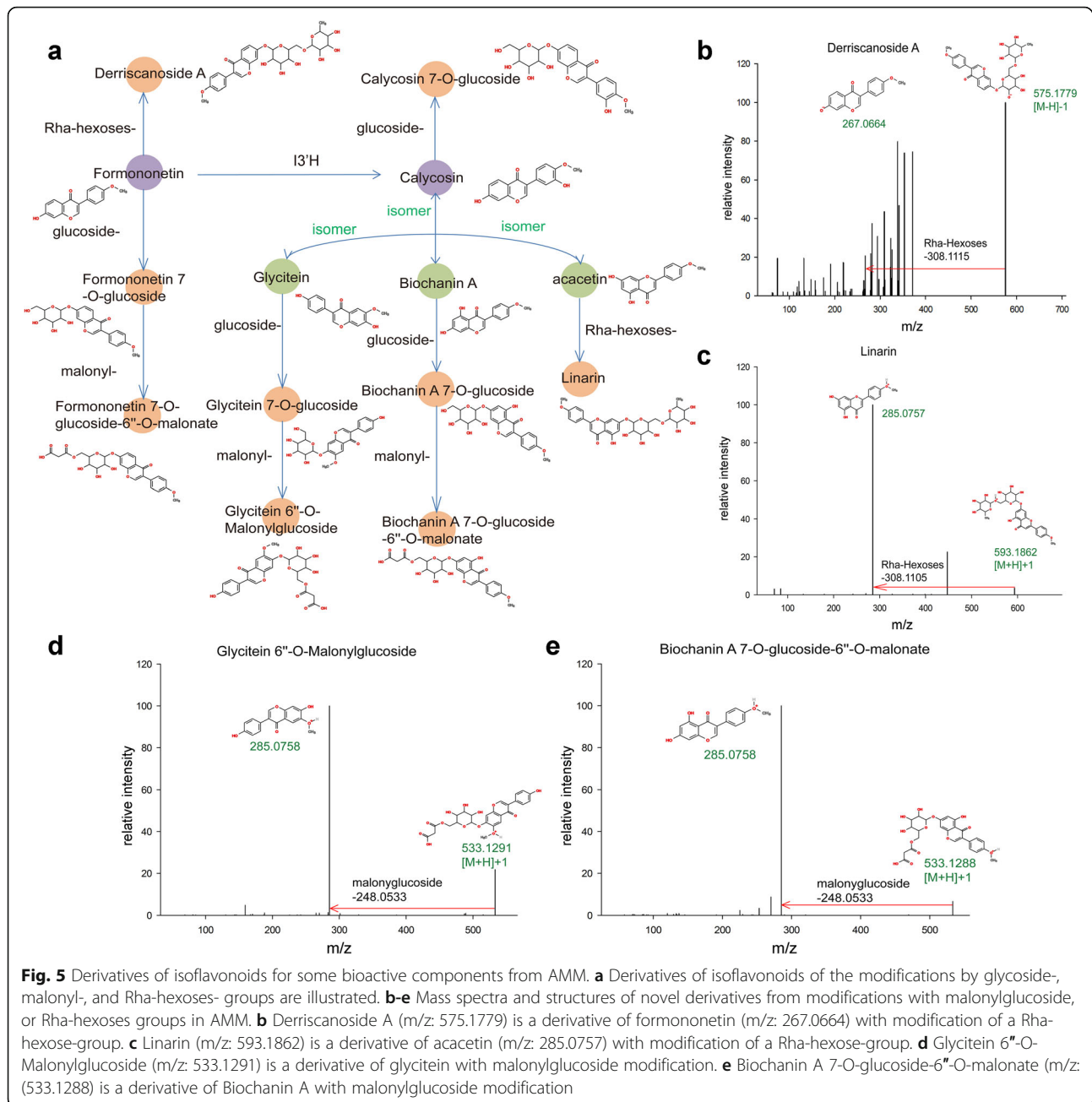


Fig. 4 The inferred pathways for phenylpropanoid biosynthesis, flavonoids biosynthesis, and isoflavonoids biosynthesis in AMM. **a** The phenylpropanoid, flavonoids and isoflavonoids biosynthesis pathway. The relative contents (log2GroupArea) for each metabolite in the three tissues were denoted with juxtaposed colored boxes (left: root; middle: leaf; and right: stem). The metabolites that significantly higher in root than the other two tissues were marked by “**”, while the metabolites that significantly lower in root than the other two tissues were marked by “*”. The enzymes confirmed by transcriptomics data are shown in red. PAL, phenylalanine ammonia-lyase; C4H, cinnamic acid 4-hydroxylase; 4CL, 4-coumarate CoA ligase; CHS, chalcone synthase; CHI, chalcone isomerase; CHR, chalcone reductase; IFS, isoflavone synthase; IOMT, Isoflavone O-methyltransferase; I3'H, isoflavone 3'-hydroxylase; I2'H, isoflavone 2'-hydroxylase; UCGT, under calycosin 7-O-glucosyltransferase [2]; HIDH, 2-hydroxyisoflavanone dehydratase; HI4OMT, isoflavone 4'-O-methyltransferase; IF7GT, isoflavone 7-O-glucosyltransferase. **b** The expression of the genes (enzymes) in the pathway measured by FPKM value

Interestingly, Calycosin not only underwent glycosylation to form (-)-Maackiain-3-O-glucoside (also known as Trifolirhizin) and Calycosin 7-O-glucoside, its precursor Formononetin, and isomer Acacetin were found to undergo novel modifications with Rha-hexose- group to form novel derivatives, Derriscanaside A and Linarin,

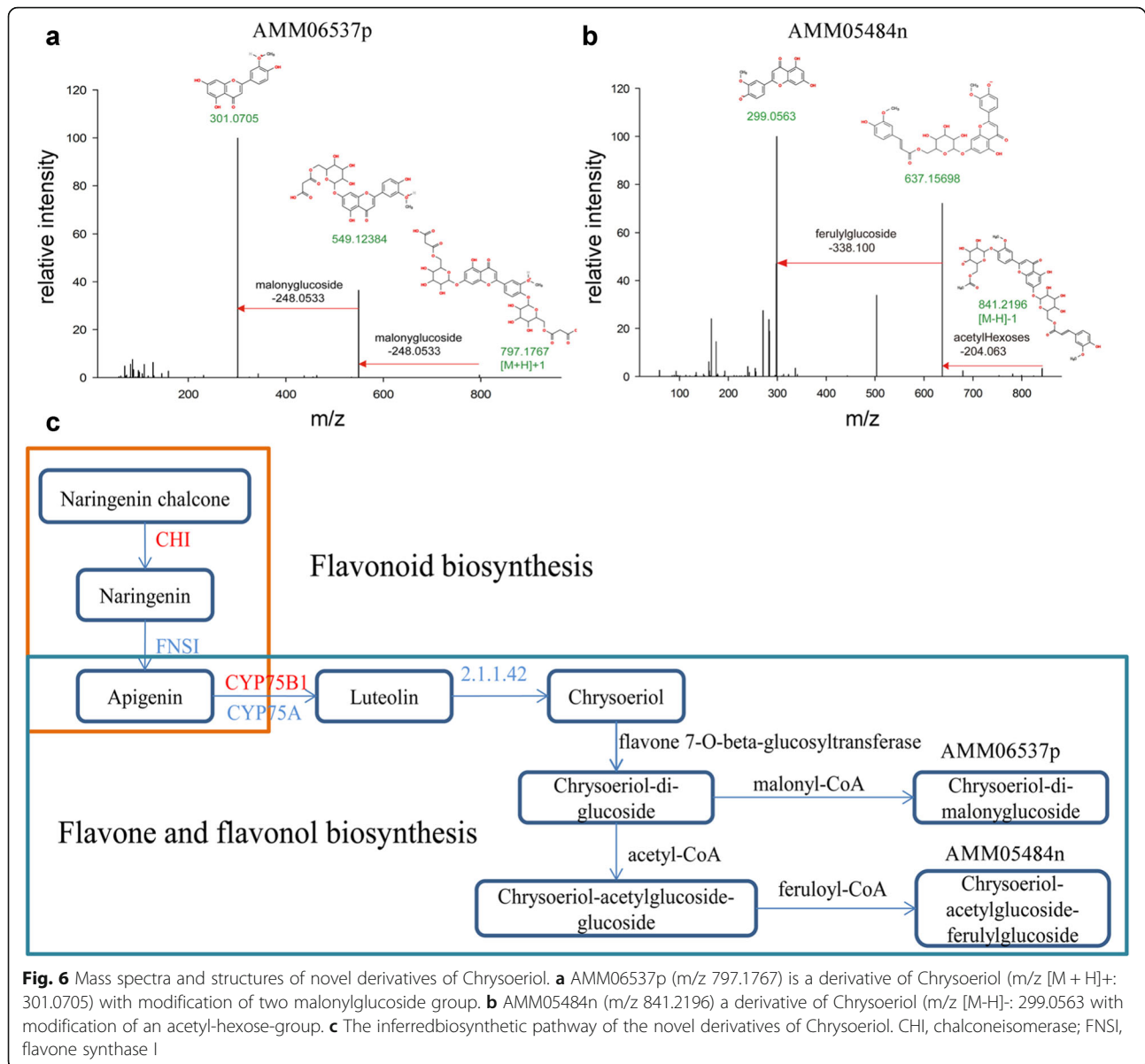
first identified by our MS data (Fig. 5b, c). In addition, glycosylated flavonoids can be further acylated, such as the novel modifications with malonyl- group, leading to production of Formononetin 7-O-glucoside-6''-O-malonate, Glycitein 7-O-glucoside - 6''-O-Malonylglucoside, and Biochanin A 7-O-glucoside-6''-O-malonate



(Fig. 5d, e). Our work reveals the formation of complex derivatives in AMM that were synthesized by extending the isoflavonoids biosynthesis pathway, in which some are first reported by current study.

Flavonoids are a class of metabolites that have a core diphenylpropane backbone (C6-C3-C6) (aglycone) with various modifications. In our bioinformatics pipeline [34], we explored the pattern for the fragmentation of flavonoids based on the backbone in order to detect other novel flavonoids with similar MS2 signature of known flavonoids. We documented a number of signature MS2 profiles for flavonoids, including Calycosin, Formononetin, Kaempferol, Apigenin,

Chrysoeriol, etc. In addition, we collected a list of neutral losses on flavonoids, which including hexoses, pentoses, malonyl, acetyl, etc. We searched for the presence of above precursor ions and neutral loss in the MS2 spectra of unknown metabolites. As a result, many novel flavonoid derivatives were discovered in our study. One example, AMM06537p (Chrysoeriol-di-malonylglucoside) was detected at RT 6.236 min, and, had the precursor ion at m/z 797.1767 ($[M+H]^+$). Its MS2 spectra displayed the fragment ions of Chrysoeriol that was observed at m/z 301.0705 ($[M+H]^+$) due to the losses of the two -malonylglucoside groups (Fig. 6a). Another example, AMM05484n (Chrysoeriol-acetylglucoside-



ferulylglucoside) was detected at RT 7.489 min, and had the precursor ion at m/z 841.2196 ($[M-H]^-$). Its MS2 spectra displayed the fragment ions at m/z 637.15698 (the neutral loss of a -acetylhexoses group from precursor), and at m/z 299.0563 (fragment ion of Chrysoeriol ($[M-H]^-$), the neutral loss of a -ferulylglucoside group from precursor) (Fig. 6b). Chrysoeriol is a dietary methoxyflavonoid with antioxidant, anti-inflammatory, lipase inhibitory, anti-cancer activities [66–68]. While the synthesis pathways of the novel Chrysoeriol derivatives was unknown, we inferred their likely synthesis steps based on their modification groups, i.e. malonylglucoside, acetylglucoside and ferulylglucoside (Fig. 6c). Accordingly, AMM06537p synthesis can be accomplished by consecutive actions of flavone 7-O-beta-glucosyltransferase and malonyl-CoA. For AMM05484n synthesis,

the consecutive actions of flavone 7-O-beta-glucosyltransferase, acetyl-CoA, and feruloyl-CoA are required (Fig. 6c). Taken together, we demonstrated our integrated metabolomics and transcriptomics approach offers great potentials in discovering novel metabolite structure and associated synthesis pathways, which helps provide a comprehensive insight into the complex metabolic network in AMM.

Conclusions

An integrated metabolomics and transcriptomics analysis approach was taken to study the metabolic profile and synthesis of bioactive compounds in traditional herb *A. membranaceus* Bge. var. *mongolicus* (Bge.) Hsiao. An unprecedented number of metabolites from AMM was discovered and annotated, representing an increase by an

order of magnitude over previous studies. Metabolic profiling found the contents of the main bioactive ingredients were significantly enriched in root, whereas contents and synthesis activity of other phytochemicals were significantly higher in leaf and stem. Using integrated metabolomics and transcriptomics data, we illustrated the complete pathways of phenylpropanoid biosynthesis, flavonoids biosynthesis, and isoflavonoids biosynthesis, for which some are first reported in AMM. More importantly, by combing metabolomics and transcriptomics analysis, we discovered novel flavonoids derivatives, and inferred likely synthesis mechanism. The current study represents the most comprehensive metabolomics analysis on traditional herb AMM, and provides novel insights into the diversity and biosynthesis of metabolites in herbs in general.

Methods

Plant materials

AMM were grown for 100 days at 22 °C under a 14 h light/10 h dark photoperiod regime in pots that contained a mixture of vermiculite, black soil and perlite (9:3:0.5). The plants were irrigated with water and fertilized once a week. The 100-day-old plants were harvested, and separated into root, stem and leaf tissues before being frozen immediately in liquid nitrogen and stored at -80 °C until experimentation [2, 69].

RNA isolation and transcriptome sequencing

Total RNA was extracted from AMM tissues with RNA prep Pure Plant Kit (Tiangen Biotech, Beijing). Transcriptome sequencing was performed using BGISEQ-500 at BGI (Shenzhen, China) according to the manufacturer's protocol. Briefly, RNA-seq library construction started with enrichment of mRNA with polyA tail using Oligo (dT) magnetic beads. Then mRNA was fragmented and used to synthesize double-strand cDNA (dscDNA) by reverse transcription with N6 random primer. dscDNA was end-repaired with phosphate at 5' end and stickiness 'A' at 3' end, before ligated with adaptor with stickiness 'T' at 3' end. The ligation products were amplified by PCR using two specific primers before they were heat-denatured. The single-strand DNA was cyclized by splint oligo and DNA ligase. The cDNA library was sequenced on a BGISEQ-500 machine to generate 50 bp single-end reads.

Sequence data mapping and transcriptome analysis

Raw sequencing data were first cleaned by removing reads with adaptor sequences, with more than 10% 'N' bases, or with over 50% low quality bases (with quality value < 5). Bowtie2 [70] (version 2.1.0) were used to align clean reads to the reference unigenes, and HISAT [71] were used to align clean reads to the AMM

reference transcriptome [2]). Default parameters were used except the SE reads parameter was chosen when it was appropriate.

Gene expression abundance was estimated using the FPKM (Fragments Per Kilobase of exon model per Million mapped reads) value of reads that were mapped to each gene. The calculation of FPKM for each gene was performed with RSEM, which computes Maximum likelihood abundance estimates using the Expectation Maximization (EM) algorithm for its statistical model [72]. NOISeq was used to analyze the differentially expressed genes (DEGs) among different tissues (Fold change ≥ 2 and diverge probability ≥ 0.8) [73]. DEGs were further analyzed by gene ontology (GO) enrichment using GSeq, and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment using KOBAS software.

Sample preparation and LC-MS/MS analysis

For AMM tissues, root, stem or leaf (three biological replicates samples for each tissue), 150 mg of sample was ground into powder in liquid nitrogen, followed by extraction with 1 mL 70% (v/v) aqueous methanol solution in an ultrasonic-assisted extraction machine. The sample was reversely blended every 10 min for 3 times before standing at 4 °C for 24 h. The supernatant was transferred into a new 5 ml centrifuge tube after centrifuging at 12,000 g for 10 min at 4 °C. The extraction was repeated for two more times. All the supernatant of the same sample was pooled into the same tube and subsequently treated with a Nitrogen blowing apparatus to volatilize methanol before vacuum dried. Dried extract was re-suspended in 150 μ L 70% aqueous methanol solution (containing 1 mg L⁻¹ capsaicin and lincomycin as internal standard). The sample extract was filtered using a 0.22 μ m filters (ANPEL, Shanghai) and transferred into a LC glass vial with fused glass insert for analyses.

Chromatographic separation was performed on a Waters Acquity Ultra Performance LC (UPLC) using an ACQUITY UPLC BEH C18 column (pore size 1.7 μ m, length, 2.1*100 mm). The mobile phase consisted of (A) water and (B) acetonitrile with 0.04% acetic acid respectively. The following gradient was used for separation with flow rate at 0.25 mL min⁻¹: 95:5 A/B at 0 min, 5:95 A/B at 20.0 min, 5:95 A/B at 24.0 min, 95:5 A/B at 24.1 min, 95:5 A/B at 30.0 min. The sample injection volume was 5 μ L, and column temperature was maintained at 40 °C.

Non-targeted metabolomics analysis was performed on Q-Exactive™ Hybrid Quadrupole-Orbitrap High Resolution Mass Spectrometer (Thermo Fisher Scientific) coupled to the UPLC system. Three biological replicates for each AMM tissue were analyzed, which were subjected to the same LC-Q-Orbitrap-MS system and operated under identical instrument conditions. MS/

MS-acquisition was performed in both positive and negative ionization in Full MS/dd-MS2 mode, in which the MS2 data of the most abundant ions could be automatically obtained. Heated electrospray ionization (HESI) parameters were as follows: Spray voltage (+), 4000 V; Spray voltage (-), 3500 V; Capillary temperature, 320 °C; Sheath gas, 35 arb; Aux gas, 8 arb; Max spray current, 100 mA; Probe heater temperature, 350 °C; S-Lens RF level, 50. MS full scan mass resolution was set to 70,000 at m/z 200, and the scan range was 100–1000 m/z. Normalized HCD energies were 15 eV and 40 eV, and average MS/MS spectrum was retained. Blank control samples were used periodically to monitor the stability of the analytical conditions.

Integrated informatics pipeline for MS data processing and identification of metabolites

An informatics pipeline [34] for MS data analysis was created by integrating the following tools/libraries. Compound Discovery software (CD v2.0; Thermo Fisher Scientific) was used to align MS data using its automatic untargeted metabolomics workflow. The following parameters were used: Min Peak Intensity: 5×10^5 ; Signal/Noise ratio (S/N) Threshold: 10; Mass Tolerance: 5 ppm; Max shift: 1 min; RT Tolerance: 0.1 min. For the others default parameters were used. All the spectra signals were grouped to merged features for positive and negative mode, respectively. After removing redundancy from multi-ion adducts and isotopes, in-source fragmentation, and dimerization, metabolite features with [H]⁺ and [H]⁻ were retained for the positive and negative mode respectively. The GroupArea that assigned by the Compound Discovery software was used to represent the relative content of the metabolites, which displays the median chromatographic peak area for the compound in the sample group.

Raw metabolite features were processed by removing redundancy from multi-ion adducts and isotopes, in-source fragmentation, and dimerization. Then MS signals were also filtered with Group Coefficient of variation (Group CV) < 50%, and Group area > $1E5$, to remove artificial signals that have poor quality or non-biological origin. The fragmentation patterns of the remaining high quality metabolite features were extracted using Xcalibur software (v2.2.0). Then metabolite features from positive and negative mode ([H]⁺ mode or [H]⁻) were merged to eliminate redundancy using parameters: mass value (error < 5 ppm) and retention time (error < 0.5 min).

Metabolites were annotated first using the experimental spectral data from public databases (m/z tolerance < 10 ppm): MassBank [74](<http://www.massbank.jp/>) and METLIN [75](<http://metlin.scripps.edu/index.php>). Compounds with similarity score > 0.8 were retained for

annotation of the query features. Those that had no match were then annotated using MetFrag with virtual reference compound databases: KEGG and Bio-pubchem [28, 76]. Bio-pubchem was a subset of pubchem, that contained 35,954 compounds of biological origin. MetFrag tool (v2.3) [76] was downloaded from <http://c-ruttkies.github.io/MetFrag/> and was used to generate in silico MS/MS spectra of compounds from in KEGG or Bio-pubchem databases and calculate the similarity score between the query MS and reference compounds' MS. For a query MS, a ranked list of reference compounds was generated (mass error < 5 ppm). The compound classification was carried out based on KEGG Phytochemical compounds library (br08003.keg).

Statistical data analysis

Principal Component Analysis (PCA) and orthogonal partial least squares discriminate analysis (OPLS-DA) was carried out using SIMCA-P (version 14). Metabolites with significant differences in content were determined with threshold of VIP (variable importance in the projection) > 1 and fold change ≥ 2 or ≤ 0.5 . Other data analysis carried out by in-house scripts were almost performed with R or Perl, such as removing redundancy, assigning the annotations to the metabolites, and some statistical tests. The figures in the manuscripts were almost drawn by R.

Supplementary information

Supplementary information accompanies this paper at <https://doi.org/10.1186/s12864-020-07005-y>.

Additional file 1. The workflow of the experiment and bioinformatics pipeline of AMM.

Additional file 2. The metabolites from AMM, containing 5435 metabolites with 3985 tagged with MS2 data.

Additional file 3. The metabolites from AMM with annotation.

Additional file 4. Metabolites previously identified in AMM that are included in our study.

Additional file 5. Mapped AMM metabolites in pathways for biosynthesis of secondary metabolites.

Additional file 6. Unigenes differentially expressed in leaf and stem in comparison of root transcriptome.

Additional file 7. The differential metabolites and genes in different tissues of AMM.

Additional file 8. The distribution of metabolites and the PCA and OPLS-DA analysis of metabolites in AMM tissues.

Additional file 9. OPLS-DA analysis of metabolites between stem and root.

Additional file 10. The distribution of the relative content (log2GroupArea) for different metabolite groups.

Abbreviations

AMM: *Astragalus membranaceus* Bge. var. *mongolicus* (Bge.) Hsiao; CG: Calycosin-7-O-b-D-glucoside; ASI: Astragaloside IV; MS: Mass spectrum; LC-MS: Liquid chromatography-mass spectrum; MS2: MS-MS, the secondary mass spectrum; PCA: Principal Component Analysis; OPLS-DA: Orthogonal partial least squares discriminate analysis; VIP: Variable importance in the

projection; CAS: Cycloartenol synthase; P450: CytochromeP450; GT: Glycosyl transferase; IF7GT: Isoflavone 7-O-glucosyltransferase

Acknowledgements

Not applicable.

About this supplement

This article has been published as part of *BMC Genomics Volume 21 Supplement 10, 2020: Selected articles from the 18th Asia Pacific Bioinformatics Conference (APBC 2020): genomics*. The full contents of the supplement are available online at <https://bmccgenomics.biomedcentral.com/articles/supplements/volume-21-supplement-10>

Authors' contributions

X.Li, P.Nan and Y.Xu conceived and designed the study, and drafted the manuscript. X.T.Wu, X.T.Li, C.T.Wang, Y.H.Shan, and W.Wang prepared samples, performed experiments and analysis. M.L.Zhu, Q.La, Y.Zhong, and Y.Xu directed on experiments and data analysis. All authors approved the final manuscript.

Funding

This work was supported in part by grants from the National Key Research and Development Program of China [2018YFA0900701], the National Science and Technology Major Projects [2018ZX08020-003, 2018ZX09711003], the National Natural Science Foundation of China [31900470, 31701137], Shanghai Municipal Commission of Science and Technology [19511121202], Biological Resources Programme, Chinese Academy of Sciences [KFJ-BRP-009]. The publication costs are funded by National Science and Technology Major Projects [2018ZX09711003].

Availability of data and materials

The datasets analyzed during the current study available from the corresponding author on reasonable request.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

¹Key Laboratory of Synthetic Biology, CAS Center for Excellence in Molecular Plant Sciences, Institute of Plant Physiology and Ecology, Chinese Academy of Sciences, Shanghai 200032, China. ²University of Chinese Academy of Sciences, Beijing 100049, China. ³Ministry of Education Key Laboratory for Biodiversity Science and Ecological Engineering, School of Life Sciences, Fudan University, Shanghai 200438, China. ⁴Shanghai Chenshan Plant Science Research Center, Shanghai Chenshan Botanical Garden, Shanghai 201602, China. ⁵Research Institute of Biodiversity & Geobiology, Department of Life Science, Tibet University, Lhasa China 850000, China. ⁶Department of Colorectal Surgery, Fudan University Shanghai Cancer Center, Shanghai, China.

Published: 18 November 2020

References

1. Cho WC, Leung KN. In vitro and in vivo immunomodulating and immunorestorative effects of *Astragalus membranaceus*. *J Ethnopharmacol*. 2007;113(1):132–41.
2. Chen J, Wu XT, Xu YQ, Zhong Y, Li YX, Chen JK, Li X, Nan P. Global transcriptome analysis profiles metabolic pathways in traditional herb *Astragalus membranaceus* Bge. var. *mongolicus* (Bge.) Hsiao. *BMC Genomics*. 2015;16(Suppl 7):S15.
3. Cui R, He J, Wang B, Zhang F, Chen G, Yin S, Shen H. Suppressive effect of *Astragalus membranaceus* Bunge on chemical hepatocarcinogenesis in rats. *Cancer Chemother Pharmacol*. 2003;51(1):75–80.
4. Kuo YH, Tsai WJ, Loke SH, Wu TS, Chiou WF. *Astragalus membranaceus* flavonoids (AMF) ameliorate chronic fatigue syndrome induced by food intake restriction plus forced swimming. *J Ethnopharmacol*. 2009;122(1):28–34.
5. Tseng A, Yang CH, Chen CH, Hsu SL, Lee MH, Lee HC, Su LJ. An in vivo molecular response analysis of colorectal cancer treated with *Astragalus membranaceus* extract. *Oncol Rep*. 2016;35(2):659–68.
6. Ren S, Zhang H, Mu Y, Sun M, Liu P. Pharmacological effects of Astragaloside IV: a literature review. *J Tradit Chin Med*. 2013;33(3):413–6.
7. Mamedova RP, Isaev MI. TRITERPENOIDES FROM *ASTRAGALUS* PLANTS. *Chem Nat Compd*. 2004;40(4):4.
8. Ma X, Tu P, Chen Y, Zhang T, Wei Y, Ito Y. Preparative isolation and purification of two isoflavones from *Astragalus membranaceus* Bge. var. *mongolicus* (Bge.) Hsiao by high-speed counter-current chromatography. *J Chromatogr A*. 2003;992(1–2):193–7.
9. Wu T, Annie Bligh SW, Gu LH, Wang ZT, Liu HP, Cheng XM, Branford-White CJ, Hu ZB. Simultaneous determination of six isoflavonoids in commercial *Astragalus* by HPLC-UV. *Fitoterapia*. 2005;76(2):157–65.
10. Zhang Q, Gao WY, Man SL: [chemical composition and pharmacological activities of astragali radix]. *Zhongguo Zhong Yao Za Zhi*. 2012;37(21):3203–7.
11. Fan Y, Wu DZ, Gong YQ, Zhou JY, Hu ZB. Effects of calycosin on the impairment of barrier function induced by hypoxia in human umbilical vein endothelial cells. *Eur J Pharmacol*. 2003;481(1):33–40.
12. S C, SR P, TR H. Inhibitory effect of astragali radix on matrix degradation in human articular cartilage. *J Microbiol Biotechnol*. 2005;15(6):1258–66.
13. YM L, Lee Y-m, Si C, Choi S-I, JW L L, Jae-won SMJ, et al. Isolation of hyaluronidase inhibitory component from the roots of *Astragalus membranaceus* Bunge (*Astragalus Radix*). *Food Sci Biotechnol*. 2005;14(2):263–7.
14. Wang S, Li J, Huang H, Gao W, Zhuang C, Li B, Zhou P, Kong D. Anti-hepatitis B virus activities of astragaloside IV isolated from radix *Astragali*. *Biol Pharm Bull*. 2009;32(1):132–5.
15. Fiehn O, Kopka J, Dormann P, Altmann T, Trethewey RN, Willmitzer L. Metabolite profiling for plant functional genomics. *Nat Biotechnol*. 2000; 18(11):1157–61.
16. Taylor J, King RD, Altmann T, Fiehn O. Application of metabolomics to plant genotype discrimination using statistics and machine learning. *Bioinformatics*. 2002;18(Suppl 2):S241–8.
17. Liu Y, Liu J, Wu KX, Guo XR, Tang ZH. A rapid method for sensitive profiling of bioactive triterpene and flavonoid from *Astragalus mongolicus* and *Astragalus membranaceus* by ultra-pressure liquid chromatography with tandem mass spectrometry. *J Chromatogr B Analyt Technol Biomed Life Sci*. 2018;1085:110–8.
18. Duan LX, Chen TL, Li M, Chen M, Zhou YQ, Cui GH, Zhao AH, Jia W, Huang LQ, Qi X. Use of the metabolomics approach to characterize Chinese medicinal material Huangqi. *Mol Plant*. 2012;5(2):376–86.
19. Dettmer K, Aronov PA, Hammock BD. Mass spectrometry-based metabolomics. *Mass Spectrom Rev*. 2007;26(1):51–78.
20. Bottcher C, von Roepenack-Lahaye E, Schmidt J, Schmotz C, Neumann S, Scheel D, Clemens S. Metabolome analysis of biosynthetic mutants reveals a diversity of metabolic changes and allows identification of a large number of new compounds in *Arabidopsis*. *Plant Physiol*. 2008; 147(4):2107–20.
21. Matsuda F, Yonekura-Sakakibara K, Niida R, Kuromori T, Shinozaki K, Saito K. MS/MS spectral tag-based annotation of non-targeted profile of plant secondary metabolites. *Plant J*. 2009;57(3):555–77.
22. Gong L, Chen W, Gao Y, Liu X, Zhang H, Xu C, Yu S, Zhang Q, Luo J. Genetic analysis of the metabolome exemplified using a rice population. *Proc Natl Acad Sci U S A*. 2013;110(50):20320–5.
23. Chen W, Gong L, Guo Z, Wang W, Zhang H, Liu X, Yu S, Xiong L, Luo J. A novel integrated method for large-scale detection, identification, and quantification of widely targeted metabolites: application in the study of rice metabolomics. *Mol Plant*. 2013;6(6):1769–80.
24. Moco S, Bino RJ, Vorst O, Verhoeven HA, de Groot J, van Beek TA, Vervoort J, de Vos CH. A liquid chromatography-mass spectrometry-based metabolome database for tomato. *Plant Physiol*. 2006;141(4):1205–18.
25. Zhu G, Wang S, Huang Z, Zhang S, Liao Q, Zhang C, Lin T, Qin M, Peng M, Yang C, et al. Rewiring of the fruit Metabolome in tomato breeding. *Cell*. 2018;172(1–2):249–61 e212.
26. Kusano M, Tabuchi M, Fukushima A, Funayama K, Diaz C, Kobayashi M, Hayashi N, Tsuchiya YN, Takahashi H, Kamata A, et al. Metabolomics data reveal a crucial role of cytosolic glutamine synthetase 1;1 in coordinating metabolic balance in rice. *Plant J*. 2011;66(3):456–66.

27. Matsuda F, Hirai MY, Sasaki E, Akiyama K, Yonekura-Sakakibara K, Provart NJ, Sakurai T, Shimada Y, Saito K. AtMetExpress development: a phytochemical atlas of Arabidopsis development. *Plant Physiol.* 2010;152(2):566–78.
28. Wolf S, Schmidt S, Muller-Hannemann M, Neumann S. In silico fragmentation for computer assisted identification of metabolite mass spectra. *BMC Bioinformatics.* 2010;11:148.
29. Heinonen M, Shen H, Zamboni N, Rousu J. Metabolite identification and molecular fingerprint prediction through machine learning. *Bioinformatics.* 2012;28(18):2333–41.
30. Allen F, Pon A, Wilson M, Greiner R, Wishart D. CFM-ID: a web server for annotation, spectrum prediction and metabolite identification from tandem mass spectra. *Nucleic Acids Res.* 2014;42(Web Server issue):W94–9.
31. Gerlich M, Neumann S. MetFusion: integration of compound identification strategies. *J Mass Spectrom.* 2013;48(3):291–8.
32. Kim S, Thiessen PA, Bolton EE, Chen J, Fu G, Gindulyte A, Han L, He J, He S, Shoemaker BA, et al. PubChem substance and compound databases. *Nucleic Acids Res.* 2016;44(D1):D1202–13.
33. Pan H, Fang C, Zhou T, Wang Q, Chen J. Accumulation of calycosin and its 7-O-beta-D-glucoside and related gene expression in seedlings of *Astragalus membranaceus* Bge. var. *mongholicus* (Bge.) Hsiao induced by low temperature stress. *Plant Cell Rep.* 2007;26(7):1111–20.
34. Li X, Zhou H, Xiao N, Wu X, Shan Y, Chen L, Wang C, Wang Z, Huang J, Li A. Expanding the Coverage of the Metabolic Landscape in Cultivated Rice with Integrated Computational Approaches. *bioRxiv* 2020.
35. Duan L, Molnar I, Snyder JH, Shen GA, Qi X. Discrimination and quantification of true biological signals in metabolomics analysis based on liquid chromatography-mass spectrometry. *Mol Plant.* 2016;9(8):1217–20.
36. Wang S, Tu H, Wan J, Chen W, Liu X, Luo J, Xu J, Zhang H. Spatio-temporal distribution and natural variation of metabolites in citrus fruits. *Food Chem.* 2016;199:8–17.
37. Bylesjo M, Rantalainen M, Cloarec O, Nicholson JK, Holmes E, Trygg J. OPLS discriminant analysis: combining the strengths of PLS-DA and SIMCA classification. *J Chemom.* 2006;20(8–10):341–51.
38. Wang Z, Cui Y, Vainstein A, Chen S, Ma H. Regulation of Fig (*Ficus carica* L.) Fruit Color: Metabolomic and Transcriptomic Analyses of the Flavonoid Biosynthetic Pathway. *Front Plant Sci.* 2017;8:1990.
39. Reen FJ, Gutierrez-Barranquero JA, Parages ML. F OG: Coumarin: a novel player in microbial quorum sensing and biofilm formation inhibition. *Appl Microbiol Biotechnol.* 2018;102(5):2063–73.
40. Stefanachi A, Leonetti F, Pisani L, Catto M, Carotti A. Coumarin: A Natural, Privileged and Versatile Scaffold for Bioactive Compounds. *Molecules.* 2018; 23(2):250.
41. Yang L, Li S, Qin X, Jiang G, Chen J, Li B, Yao X, Liang P, Zhang Y, Ding W. Exposure to Umbelliferone reduces *Ralstonia solanacearum* biofilm formation, Transcription of Type III Secretion System Regulators and Effectors and Virulence on Tobacco. *Front Microbiol.* 2017;8:1234.
42. Joubert J, Foka GB, Repsold BP, Oliver DW, Kapp E, Malan SF. Synthesis and evaluation of 7-substituted coumarin derivatives as multimodal monoamine oxidase-B and cholinesterase inhibitors for the treatment of Alzheimer's disease. *Eur J Med Chem.* 2017;125:853–64.
43. Ren QC, Gao C, Xu Z, Feng LS, Liu ML, Wu X, Zhao F. Bis-coumarin Derivatives and Their Biological Activities. *Current topics in medicinal chemistry.* 2018;18(2):101–13.
44. Kapp E, Visser H, Sampson SL, Malan SF, Streicher EM, Foka GB, Warner DF, Omoruyi SI, Enogieru AB, Ekpo OE, et al. Versatility of 7-Substituted Coumarin Molecules as Antimycobacterial Agents, Neuronal Enzyme Inhibitors and Neuroprotective Agents. *Molecules.* 2017;22(10):1644.
45. Canden ML, Naline E, D'Ocon MP, Cortes D, Advenier C. Effects of coumarin and other isoquinoline alkaloids on Guinea-pig trachea and human bronchus. *J Pharm Pharmacol.* 1990;42(2):102–7.
46. Lyamzaev KG, Pustovidko AV, Simonyan RA, Rokitskaya TI, Domnina LV, Ivanova OY, Severina II, Sumbatyan NV, Korshunova GA, Tashlitsky VN, et al. Novel mitochondria-targeted antioxidants: plastoquinone conjugated with cationic plant alkaloids berberine and palmatine. *Pharm Res.* 2011;28(11): 2883–95.
47. Dhingra D, Bhancker A. Behavioral and biochemical evidences for antidepressant-like activity of palmatine in mice subjected to chronic unpredictable mild stress. *Pharmacol Rep.* 2014;66(1):1–9.
48. Shen Y, Guan S, Ge H, Xiong W, He L, Liu L, Yin C, Liu H, Li G, Xu C, et al. Effects of palmatine on rats with comorbidity of diabetic neuropathic pain and depression. *Brain Res Bull.* 2018;139:56–66.
49. Funk RS, Singh RK, Winefield RD, Kandel SE, Ruisinger JF, Moriarty PM, Backes JM. Variability in potency among commercial preparations of Berberine. *J Diet Suppl.* 2018;15(3):343–51.
50. Yu QT, Qi LW, Li P, Yi L, Zhao J, Bi Z. Determination of seventeen main flavonoids and saponins in the medicinal plant Huang-qi (*Radix astragali*) by HPLC-DAD-ELSD. *J Sep Sci.* 2007;30(9):1292–9.
51. Zhu J, Wen K. Astragaloside IV inhibits TGF- β 1-induced epithelial-mesenchymal transition through inhibition of the PI3K/Akt/NF- κ B pathway in gastric cancer cells. *Phytotherapy research: PTR.* 2018;32(7):1289–96.
52. You LZ, Lin YX, Fang ZH, Shen GM, Zhao JD, Wang TT. [research advances on astragaloside-IV in treatment of diabetes mellitus and its complications pharmacological effects]. *Zhongguo Zhong Yao Za Zhi.* 2017;42(24):4700–6.
53. Wan J, Deng L, Zhang C, Yuan Q, Liu J, Dun Y, Zhou Z, Zhao HY, Liu C, Yuan D, et al. Chikusetsu saponin V attenuates H₂O₂-induced oxidative stress in human neuroblastoma SH-SY5Y cells through Sirt1/PGC-1 α /Mn-SOD signaling pathways. *Can J Physiol Pharmacol.* 2016;94(9):919–28.
54. Dai YW, Zhang CC, Zhao HX, Wan JZ, Deng LL, Zhang Y, Dun YY, Liu CQ, Yuan D, Wang T. Chikusetsusaponin V attenuates lipopolysaccharide-induced liver injury in mice. *Immunopharmacol Immunotoxicol.* 2016;38(3): 167–74.
55. Wang T, Dai Y, Dun Y, Zhang C, Wan J, Deng L, Zhou Z, Liu C, Yuan D. Chikusetsusaponin V inhibits inflammatory responses via NF- κ B and MAPK signaling pathways in LPS-induced RAW 264.7 macrophages. *Immunopharmacol Immunotoxicol.* 2014;36(6):404–11.
56. Xu K, Chu F, Li G, Xu X, Wang P, Song J, Zhou S, Lei H. Oleanolic acid synthetic oligoglycosides: a review on recent progress in biological activities. *Pharmazie.* 2014;69(7):483–95.
57. Perez AJ, Pecio L, Kowalczyk M, Kontek R, Gajek G, Stopinsek L, Mirt I, Oleszek W, Stochmal A. Triterpenoid components from oak heartwood (*Quercus robur*) and their potential health benefits. *J Agric Food Chem.* 2017;65(23):4611–23.
58. Petronelli A, Pannitteri G, Testa U. Triterpenoids as new promising anticancer drugs. *Anti-Cancer Drugs.* 2009;20(10):880–92.
59. Salvadori JAR, Leal AS, Valdeira AS, Goncalves BMF, Alho DPS, Figueiredo SAC, Silvestre SM, Mendes VIS. Oleanane-, ursane-, and quinone methide friedelane-type triterpenoid derivatives: recent advances in cancer treatment. *Eur J Med Chem.* 2017;142:95–130.
60. Choi SI, Heo TR, Min BH, Cui JH, Choi BH, Park SR. Alleviation of osteoarthritis by calycosin-7-O-beta-D-glucopyranoside (CG) isolated from *Astragalus radix* (AR) in rabbit osteoarthritis (OA) model. *Osteoarthr Cartil.* 2007;15(9):1086–92.
61. Zhang YZ, Xu F, Liang J, Tang JS, Shang MY, Wang X, Cai SQ. Isoflavonoids from roots of *Astragalus membranaceus* var. *mongholicus*. *Zhongguo Zhong Yao Za Zhi.* 2012;37(21):3243–8.
62. Auyeung KK, Law PC, Ko JK. Novel anti-angiogenic effects of formononetin in human colon cancer cells and tumor xenograft. *Oncol Rep.* 2012;28(6): 2188–94.
63. Lou H, Fan P, Perez RG. Neuroprotective effects of linarin through activation of the PI3K/Akt pathway in amyloid-beta-induced neuronal cell death. *Bioorg Med Chem.* 2011;19(13):4021–7.
64. PP O, JK J, Al H, PM V. Linarin, a selective acetylcholinesterase inhibitor from *Mentha arvensis*. *Fitoterapia.* 2006;77(6):429–34.
65. Gillissen B, Burkle L, Andre B, Kuhn C, Rentsch D, Brandl B, Frommer WB. A new family of high-affinity transporters for adenine, cytosine, and purine derivatives in *Arabidopsis*. *Plant Cell.* 2000;12(2):291–300.
66. Ramirez G, Zamilpa A, Zavala M, Perez J, Morales D, Tortoriello J. Chrysoeriol and other polyphenols from *Tecoma stans* with lipase inhibitory activity. *J Ethnopharmacol.* 2016;185:1–8.
67. Zeng X, Shi J, Zhao M, Chen Q, Wang L, Jiang H, Luo F, Zhu L, Lu L, Wang X, et al. Regioselective Glucuronidation of Diosmetin and Chrysoeriol by the interplay of Glucuronidation and transport in UGT1A9-overexpressing HeLa cells. *PLoS One.* 2016;11(11):e0166239.
68. Cha BY, Shi WL, Yonezawa T, Teruya T, Nagai K, Woo JT. An inhibitory effect of chrysoeriol on platelet-derived growth factor (PDGF)-induced proliferation and PDGF receptor signaling in human aortic smooth muscle cells. *J Pharmacol Sci.* 2009;110(1):105–10.
69. Hu J, Anderson B, Wessler SR. Isolation and characterization of rice R genes: evidence for distinct evolutionary paths in rice and maize. *Genetics.* 1996; 142(3):1021–31.

70. Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* 2009;10(3):R25.
71. Kim D, Langmead B, Salzberg SL. HISAT: a fast spliced aligner with low memory requirements. *Nat Methods.* 2015;12(4):357–60.
72. Li B, Dewey CN. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics.* 2011;12:323.
73. Tarazona S, Garcia-Alcalde F, Dopazo J, Ferrer A, Conesa A. Differential expression in RNA-seq: a matter of depth. *Genome Res.* 2011;21(12):2213–23.
74. Horai H, Arita M, Kanaya S, Nihei Y, Ikeda T, Suwa K, Ojima Y, Tanaka K, Tanaka S, Aoshima K, et al. MassBank: a public repository for sharing mass spectral data for life sciences. *J Mass Spectrom.* 2010;45(7):703–14.
75. Smith CA, O'Maille G, Want EJ, Qin C, Trauger SA, Brandon TR, Custodio DE, Abagyan R, Siuzdak G. METLIN: a metabolite mass spectral database. *Ther Drug Monit.* 2005;27(6):747–51.
76. Ruttkies C, Schymanski EL, Wolf S, Hollender J, Neumann S. MetFrag relaunched: incorporating strategies beyond in silico fragmentation. *J Cheminform.* 2016;8:3.

Publisher's Note

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

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

