Integration of a Decrescent Transcriptome and Metabolomics Dataset of *Peucedanum praeruptorum* to Investigate the CYP450 and MDR Genes Involved in Coumarins Biosynthesis and Transport

Yucheng Zhao[1], Tingting Liu[1], Jun Luo[1], Qian Zhang[1], Sheng Xu[2], Chao Han[1], Jinfang Xu[1], Menghan Chen[1], Yijun Chen[1] and Lingyi Kong[1]*

[1] State Key Laboratory of Natural Medicines, Department of Natural Medicinal Chemistry, China Pharmaceutical University, Nanjing, China, [2] Institute of Botany, Jiangsu Province and Chinese Academy of Sciences, Nanjing, China

Peucedanum praeruptorum Dunn is well-known traditional Chinese medicine. However, little is known in the biosynthesis and the transport mechanisms of its coumarin compounds at the molecular level. Although transcriptomic sequence is playing an increasingly significant role in gene discovery, it is not sufficient in predicting the specific function of target gene. Furthermore, there is also a huge database to be analyzed. In this study, RNA sequencing assisted transcriptome dataset and high-performance liquid chromatography (HPLC) coupled with electrospray-ionization quadrupole time-of-flight mass spectrometry (Q-TOF MS)-based metabolomics dataset of *P. praeruptorum* were firstly constructed for gene discovery and compound identification. Subsequently, methyl jasmonate (MeJA)-induced gene expression analysis and metabolomics analysis were conducted to narrow-down the dataset for selecting the candidate genes and the potential marker metabolites. Finally, the genes involved in coumarins biosynthesis and transport were predicted with parallel analysis of transcript and metabolic profiles. As a result, a total of 40,952 unigenes and 19 coumarin compounds were obtained. Based on the results of gene expression and metabolomics analysis, 7 cytochrome-P450 and 8 multidrug resistance transporter unigenes were selected as candidate genes and 8 marker compounds were selected as biomarkers, respectively. The parallel analysis of gene expression and metabolites accumulation indicated that the gene labeled as *23,746*, *228,* and *30,922* were related to the formation of the coumarin core compounds whereas *36,276* and *9533* participated in the prenylation, hydroxylation, cyclization or structural modification. Similarly, *1462*, *20,815,* and *15,318* participated in the transport of coumarin core compounds while *124,029* and *324,293* participated in the transport of the modified compounds. This finding suggested that integration of a decrescent transcriptome and metabolomics dataset could largely narrow down the number of gene to be investigated and significantly improve the efficiency of functional gene predication. In addition, the large amount of transcriptomic data produced from *P. praeruptorum* and the genes discovered in this study would provide useful information in investigating the biosynthesis and transport mechanism of coumarins.

Keywords: *P. praeruptorum*, transcriptomic, metabolomics, coumarins, biosynthesis, transporter

INTRODUCTION

Radix Peucedani (Baihua Qianhu in Chinese), the roots of *Peucedanum praeruptorum* Dunn, is one of the most popular traditional Chinese medicine and has been used for more than 1500 years. It is also listed in the current Pharmacopeia of the Peoples' Republic of China (Commission, 2010). Traditionally, Radix Peucedani is used as a kind of herbal medicine for reducing fevers and resolving phlegm and is generally employed to treat anemopyretic cold, cough with abundant phlegm and congested chest (Zhou et al., 2014). Modern pharmacological studies have indicated that the extracts of *P. praeruptorum* (the main chemical constituents are coumarins) also displayed anti-cancer, anti-inflammatory, anti-hyperglycemic, anti-oxidant, and calcium-channel-blocking properties (Wu et al., 2003; Kumar et al., 2009; Yu et al., 2012). However, little is known in coumarins biosynthetic pathways, even at the biochemical level (Bourgaud et al., 2006). As shown in **Figure 1**, among the three stages of proposed biosynthetic pathways, only the enzyme involved in the first stage are thoroughly studied (Gaid et al., 2012; Karamat et al., 2012; Kim et al., 2012). It is well known that secondary metabolites can accumulate at a high concentration in particular tissues. In addition, the roots of *P. praeruptorum* are usually chosen in medical treatments (Yazaki et al., 2008; Commission, 2010). While, other questions, such as how this kind of compound accumulates and how they maintain at an appropriate concentration in different compartments, arose. Recently, a mountain of reports indicated that a transporter-based transport mechanism was involved in the process of accumulation. In addition, some transporters had also been identified. However, there are no reports on the transport mechanisms of coumarin compounds. Even worse, no coumarins transporter was identified (Li et al., 2002; Yazaki, 2005, 2006; Rea, 2007; Mehrshahi et al., 2013; Yu and De Luca, 2013). Considering their important pharmacological benefits and the uncertainties in coumarins biosynthesis and transport, it is critical to investigate the enzymes and genes relevant to their biosynthesis and transport.

As an example for gene function prediction, "guilt-by-association" or "comparative co-expression" has emerged in the past 10 years and displayed powerful performance (Usadel et al., 2009; Mutwil et al., 2011; Movahedi et al., 2012). However, it seems difficult in predicting gene that has not been reported, especially the gene that no reference conserved co-expression clusters (Movahedi et al., 2012). Another success in studying functional genomics is integrating analysis of comprehensive gene expression and metabolic profiling according to the fact that the enhanced accumulation of metabolites was preceded by coordinated increases in the transcript level of relevant genes (Saito et al., 2008; Saito and Matsuda, 2010; Choi et al., 2012; Gaid et al., 2012). However, these successes depend largely on the availability of genome sequence data and metabolite accumulation database (Saito et al., 2008; Saito and Matsuda, 2010). Hence, it is advantageous to study the functional genomics of model plants such as *Arabidopsis* and rice which have genome data and metabolite database (Saito et al., 2008; Yonekura-Sakakibara et al., 2012), but it is not necessarily straightforward in non-model plants especially *P. praeruptorum*

whose transcriptome data is still unknown. As an important supplement to genomic analysis, transcriptomic analysis is playing an increasingly significant role in the discovery of new genes involved in plant secondary metabolism, particularly in the case of plants for which the full genomic sequences are not currently available. Next-generation-sequencing technology (NGS), such as Roche/454 and Illumina HiSeq platforms, has emerged as a primary tool for high-through-put sequencing which can dramatically improve the efficiency and rapidity of gene discovery (Schuster, 2008; Ansorge, 2009; Wang et al., 2014). However, it is insufficient in gene discovery just using sequence data.

It is generally acceptable and evaluated the assumption of correlation between gene expression and metabolite accumulation and plenty of works have been reported (Hirai et al., 2007; Saito et al., 2008). However, unlike *Arabidopsis*, there is no metabolite database available in *P. praeruptorum* (Saito et al., 2008). Considering the complicated chemical constituents of coumarins in *P. praeruptorum*, methods need to be established to analyze and construct the metabolite dataset. Hence, in this study, HPLC coupled with Q-TOF MS was used to identify the metabolites because it is rapid and can provide accurate mass measurements (Ling et al., 2013). Merging with transcriptome dataset, it could largely improve the rapidity, accuracy and efficiency of gene discovery (Yamazaki et al., 2013; Li et al., 2015). However, correlation analysis of the whole transcriptomic and metabolomics seems intertwined and unnecessary because we mainly concern some certain pathways or genes. Thus, in this study, a decrescent dataset of transcriptome and metabolomics of *P. praeruptorum* was constructed to investigate the target gene involved in our specific pathway. As described in **Figure 2**, the transcriptome sequencing and Q-TOF-MS/MS-based metabolites analysis were first conducted to construct the dataset. Then, expression analysis and metabolomics analysis were conducted to establish the descendant dataset. Finally, the differences in the peak intensities and distributions of the biomarker compounds in the different tissues were analyzed in conjunction with the expression analysis of selected gene to identify the genes involved in the biosynthesis and transport of coumarins.

As a result, putative CYP450 and MDR genes that participated in the biosynthesis and transport of coumarin compounds were predicted at the gene and metabolic level. Additionally, the specific accumulation of coumarin compounds in the roots, stems, and leaves was also investigated, the results of which may facilitate the utilization of portions of medicinal plants. To the best of our knowledge, this is the first study to examine the transcriptome of *P. praeruptorum* and the first attempt to merge transcriptomic and metabolomics to understand the biosynthetic pathways and identify the genes that participate in the biosynthesis and transport of coumarins.

MATERIALS AND METHODS

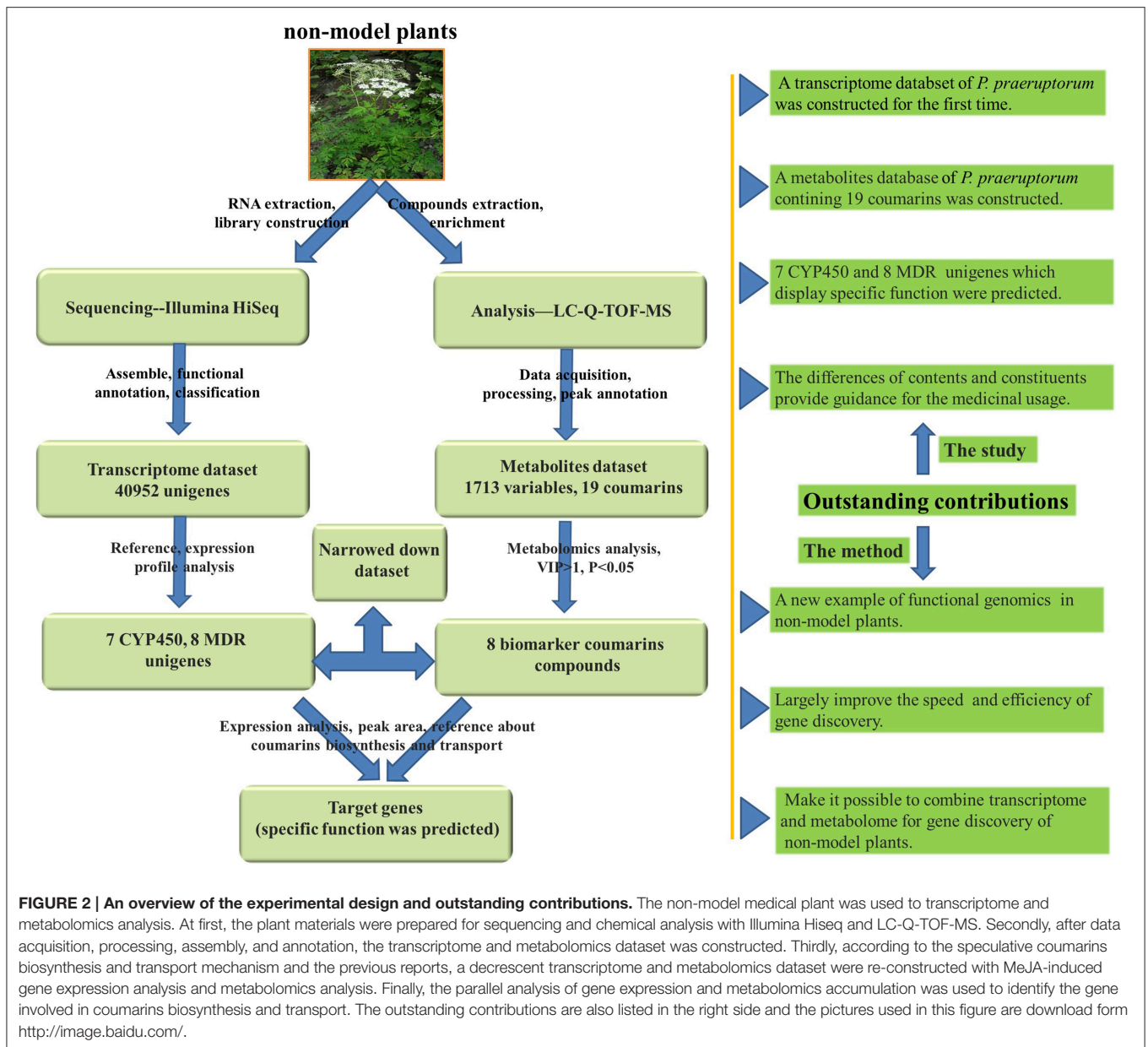### Plant Materials and RNA Preparation

The two-year-old *P. praeruptorum* material was collected from the fields of Ningguo City, Anhui Province, China, which has
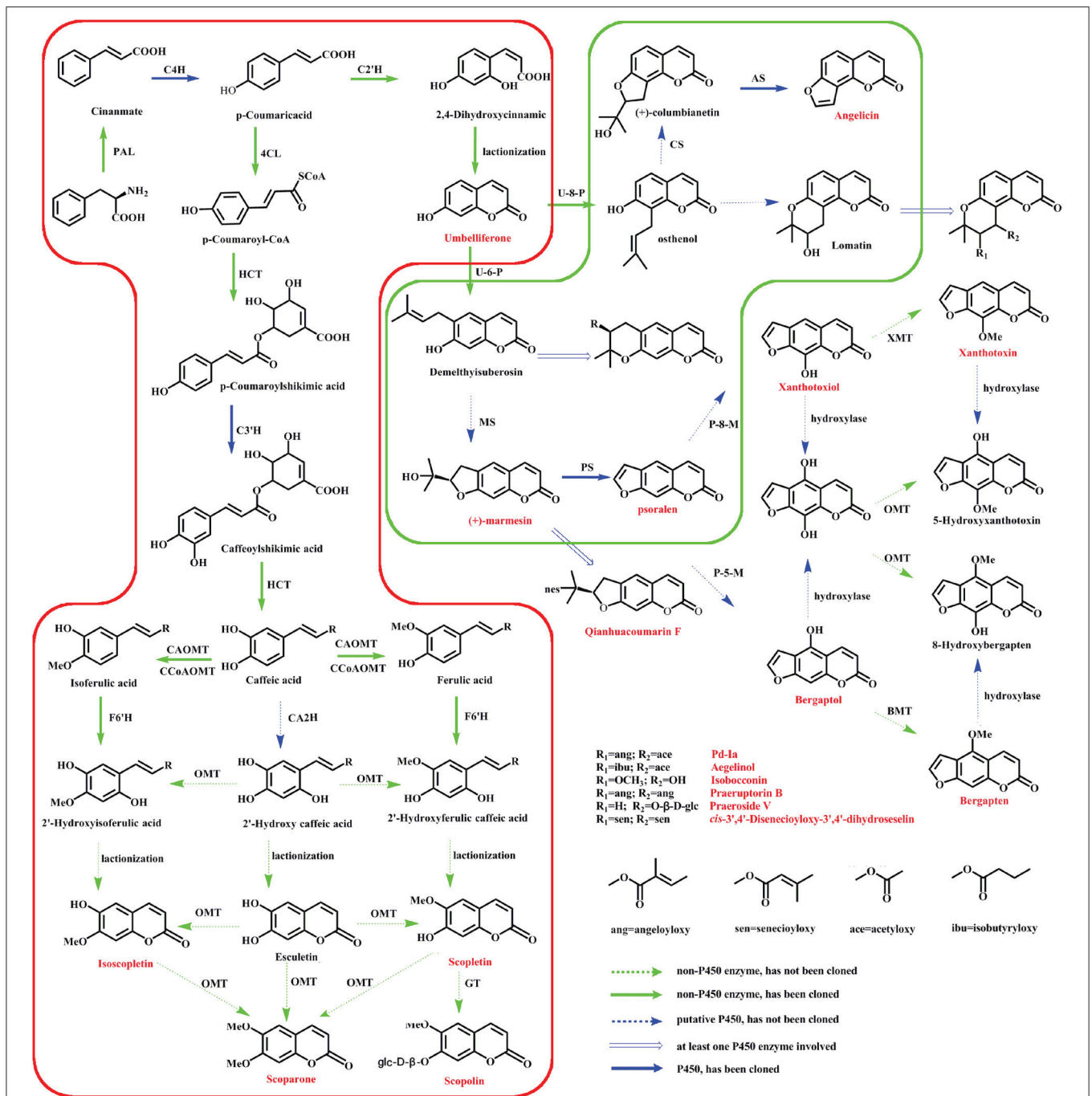
FIGURE 1 | Putative coumarins biosynthetic pathway in *P. praeruptorum*. The three stages of proposed biosynthetic pathways: formation of coumarins core (red box), prenylation, hydroxylation, and cyclization (green box) and structure modifications. PAL, phenylalanine ammonia lyase; C4H, cinnamate 4-hydroxylase; 4CL, 4-coumarate: coenzyme A ligase; HCT, hydroxycinnamoyl CoA shikimate/quinate hydroxycinnamoyl transferase; CCoAOMT, caffeoyl-CoA O-methyltransferase; C3H, p-coumarate 3-hydroxylase; CA2H, caffeic acid 2-hydroxylase; CAOMT, caffeic acid O-methyltransferase; O-MT, O-methyl-transferase; GT, glycosyltransferase; C2'H, Cinnamic acid 2'-hydroxylase; F6'H, feruloyl-CoA 6'-hydroxylase; U-6-P, Umbelliferone 6-prenyltransferase; U-8-P, Umbelliferone 8-prenyltransferase; MS, Marmesin synthase; PS, Psoralen synthase; P-8-M, Psoralen 8-monoooxygenase; P-5-M, Psoralen 5-monoooxygenase; XMT, Xanthotoxiol O-methyltransferase; BMT, Bergaptol O-methyltransferase; CS, Columbianetin synthase; AS, Angelicin synthase. The compounds identified in *P. praeruptorum* are marked with red characters.

the reputation of being "The County of Qianhu in China." After removing the soil and deadwood, the plants were immediately frozen in liquid nitrogen and were stored at −80°C until use

for total RNA isolation. Plants were also transplanted into plastic basins containing with a mixture of vermiculite, perlite, and peat moss at a 1:1:1 ratio and were grown in an environmentally

FIGURE 2 | An overview of the experimental design and outstanding contributions. The non-model medical plant was used to transcriptome and metabolomics analysis. At first, the plant materials were prepared for sequencing and chemical analysis with Illumina Hiseq and LC-Q-TOF-MS. Secondly, after data acquisition, processing, assembly, and annotation, the transcriptome and metabolomics dataset was constructed. Thirdly, according to the speculative coumarins biosynthesis and transport mechanism and the previous reports, a decrescent transcriptome and metabolomics dataset were re-constructed with MeJA-induced gene expression analysis and metabolomics analysis. Finally, the parallel analysis of gene expression and metabolomics accumulation was used to identify the gene involved in coumarins biosynthesis and transport. The outstanding contributions are also listed in the right side and the pictures used in this figure are download form http://image.baidu.com/.

controlled chamber with a long photoperiod (16 h light and 8 h dark) at 25°C, 40–65% relative humidity and 3000 lux of light intensity until use in the MeJA induction experiment. The total RNA was isolated using TransZol Plant reagent (TransGen Biotech, Beijing, China) according to the manufacturer's recommendations, and the quantity and quality of the RNA were determined using a Spectramax plus384 enzyme-labeling instrument (Molecular Devices, Sunnyvale, USA) and 1% agarose gels. To evaluate the differences in tissue-specific expression and accomplish the aim of this study, the total RNA was extracted from a mixture of roots, stems and leaves for construction of a cDNA library. It was also used for tissue-specific expression analysis. All of the samples were treated with DNase I (Takara, Daliang, China) at a concentration of 1 unit/μg of total RNA.

cDNA library Construction and Transcriptome sequence Processing and Assembly

A cDNA library was prepared with a kit provided by Illumina according to the manufacturer's recommendations and previously used methods (Yuan et al., 2015). For details, total RNA was extracted from a mixture of roots, stems and leaves using TransZol Plant reagent (TransGen Biotech, Beijing, China) according to the manufacturer's recommendations. Then, poly (A) mRNA was purified from the total RNA using oligo (dT) beads. After purification, mRNA was sheared into small pieces using fragmentation buffer. First-strand cDNA was annealed with random primers using cleaved mRNA fragments as templates. The second-strand cDNA was synthesized with

DNA polymerase I and RNase H. Subsequently, the cDNA fragments were purified and ligated to index adapters. Finally, the cDNA library was constructed and subject to Illumina HiSeq 2500 system for high-throughput sequencing. The raw data were converted into fastaq format and then compressed as.gz files to be deposit to the National Center for Biotechnology Information (NCBI).

Sequence Read Achieve (SRA) sequence database under project accession number SRX997427. Due to the error rate in the raw data, low-quality-sequence fragments were removed via slip-window sampling using the following parameters: quality threshold of 20 (error rate = 1%), window size of 5 bp and length threshold of 35 bp. To adjust the pollution of the reads, 10^5 sequences were randomly selected for sequence alignment of the nr reads at an E-value of $<1e^{-10}$ and a coverage level of $>80\%$. After the pollution of the reads was cleaned, the good reads was used to assemble transcripts and unigenes using Trinity software (version trinityrnaseq_r2013-02-25) [(http://trinityrnaseq.sf.net)]. The unigenes representing the longest transcripts at each loci (comp*_c*_) were assembled using the Chrysalis cluster module of the Trinity program. To normalize the abundance of the transcripts, a k-mer value of 25 RPKM (reads per kb of an exon model per million mapped reads) (Wagner et al., 2012) was applied and defined in this way:

$$RPKM = \frac{\text{total exon reads}}{\text{mapped reads(millions)} * \text{exon length(KB)}}$$

Functional Annotation and Classification

To find the most descriptive annotation for each transcript sequence, BLAST searches (Altschul et al., 1997) were conducted based on sequence similarities using a series of databases (Kanehisa et al., 2002; Dimmer et al., 2012), with the significance threshold set at an e-value of $\leq 1e^{-5}$. The functional categories of these unique sequences were analyzed using the Gene Ontology (GO: http://www.geneontology.org/) database, AGI codes and the TAIR GO slim program provided by TAIR (Lamesch et al., 2012). Pathway assignments were conducted based on the KEGG mapping results (Du et al., 2014) and enzyme commission (EC) numbers were assigned to the unique sequences. The KOG/COGs (clusters of orthologous groups) of the proteins were aligned to the entries in the EggNOG database to predict and classify the possible functions of the unigene products (http://www.ncbi.nlm.nih.gov/COG/).

MeJA Elicitation and preparation of samples for HPLC-Q-TOF-MS/MS Analysis

The MeJA stock solution was prepared at 200 mM in ethanol and was filter-sterilized. For short period elicitation, plants were placed in deionized water, and then an aliquot of the stock MeJA solution was added to deionized water to yield a final concentration of 200 μM. For treatment, the MeJA solution with a final concentration was sprayed onto the leaves. The treated samples and the untreated control samples (to which an equivalent amount of ethanol was applied) were harvested for analysis at 1, 3, 6, 9, 12, and 24 h after elicitation. The treated samples were divided into three parts, the roots, stems and leaves, and then the samples were dried until reaching a constant weight. A 0.5 g dried sample was sequentially extracted three times using methanol, with ultrasonication, for 30 min at room temperature (3 × 4 ml). The three methanol extracts were combined, and the mixture was concentrated under reduced pressure conditions and then readjusted to a volume of 5 ml. To separate the high-abundance contents (such as praeruptorin A and praeruptorin C) from the low-abundance contents and then concentrate the low-abundance contents, a C18E cartridge was used. Before this separation was accomplished, the cartridge was equilibrated using 20 ml of 50% methanol, after which 0.25 ml (containing approximately 20 mg of compounds) of the volume-readjusted extract was loaded onto the cartridge and was allowed to adsorb for approximately 4 h. Then, 20 ml of 50% methanol was loaded for elution, and the eluate was evaporated to dryness using a stream of nitrogen gas. Finally, 20 ml of 90% methanol was used to elute the high-abundance contents. The residue of the 50% methanol eluate was suspended in 250 μl of methanol, and the 90% methanol eluate was suspended in 1 ml of methanol for further analysis.

Quantitative Real-time PCR Analysis

To determine the expression level of putative CYP450s and MDR transcripts in *P. praeruptorum* after MeJA elicitation, quantitative real-time PCR analysis was performed using the SYBR Green PCR Master Mix (Vazyme, Nanjing, China) with LightCycler 480 instrument (Roche Molecular Biochemicals, Mannheim, Germany). Each reaction mixture contained 10 μl of 2 × SYBR Green Master Mix Reagent, 10 ng of cDNA and 4 μM of gene-specific primers in a total volume of 20 μl. The cycling conditions were as follows: 1 cycle of 95°C for 5 min, 40 cycles of 95°C for 10 s, and then 60°C for 30 s, followed by 1 cycle of 95°C for 15 s, 60°C for 60 s, and 95°C for 15 s. The mean value of three replicates was normalized with glyceraldehyde-3-phosphate dehydrogenase (GAPDH) (*comp23086_c0_seq1* in our transcriptome dataset). PCR amplification was performed using primers specific for the putative CYP450 and MDR transporter transcripts, which are listed in Table S1. Primer 5.0 software was used to design the primers. The relative expression levels were calculated by comparing the CTs (cycle thresholds) of the target genes with that of GAPDH using the $2^{-\Delta\Delta CT}$ method (Pfaffl, 2001).

HPLC-Q-TOF-MS/MS Analysis

LC/electrospray ionization (ESI)-Q-TOF-MS/MS, applied in the positive ionization mode, was used to detect the coumarin compounds. The LC analysis of the selected coumarin compounds was conducted using an Agilent 1290 HPLC system (Agilent Technologies, Santa Clara, CA, USA) equipped with a binary pump, an online degasser, an auto plate sampler and a thermostatically controlled column compartment. Chromatographic separations were performed using an Agilent ZORBAX SB-C18 column (4.6 × 250 mm, 5 μm, Agilent Technologies, Santa Clara, CA, USA) with a solvent flow rate of 1 ml/min at 35°C. The sample injection volume was set to 10 μl, and the diode array detector was operated at 254 and 310 nm. The mobile phase A was 0.1% formic acid in water and the

mobile phase B was methanol. The solvent gradient conditions A:B (v/v) are as follows: 0 min, 90:10; 3 min, 90:10; 7 min, 75:25; 8 min, 70:30; 12 min, 70:30; 13 min, 58:42; 20 min, 58:42; 21 min, 58:42; 25 min, 40:60; 28 min, 40:60; 35 min, 0:100; and 40 min, 0:100. A 5-min post-run equilibration to the initial mobile phase composition was conducted after each analysis was completed. This HPLC system was connected to a quadrupole time-of-flight mass spectrometer (Agilent Technologies, Santa Clara, CA, USA) that was equipped with an electrospray interface. The conditions of the ESI source were as follows: drying gas (N_2) flow rate, 8.0 l/min; drying gas temperature, 300°C; nebulizer, 241 kPa (35 psig); capillary voltage, 4000 V; fragmentor voltage, 150 V; collision energy, 30 eV; skimmer voltage, 60 V, and octopole radio frequency, 250 V. This instrument provided a typical resolution of 9500 ± 500 (m/z 922.0098). All of the operations and analysis of data were controlled using Agilent LC-Q-TOF-MS Mass Hunter workstation software (version B.04.00).

Data Processing and Metabolomics Analysis

The Transomics metabolomics software package of the Waters Corporation was used for data processing because it could use the retention time (RT) and m/z data pairs as identifiers. The raw data regarding the coumarin content of the roots, stems and leaves were converted to common data format files using conversion software and then imported into the SIMCA-P13.0 software package (version 13.0, Umetrics, Umea, Sweden) for multivariate statistical analysis (Yan et al., 2015). As an unbiased statistical method, a principal component analysis (PCA) was first conducted to detect the inherent trends within the data, and then partial least-squares-discriminant analysis (PLS-DA) was conducted to study the variance in the levels in the different tissue compartments. To evaluate the quality of the model, the R^2 and Q^2 values were calculated. To select the potential biomarkers, the loading plot and variable importance in the project (VIP) values were first set at 1 (VIP >1), and then Student's t-test was conducted with the P-value set at 0.05 ($P < 0.05$) to determine the significantly different variables. The selected potential biomarkers were evaluated according to the metabolites identified based on their retention times, exact mass data and fragmentation ions. To analyze the gene expression and metabolites accumulation, three biological and technical replicates were used to obtain the data. Unless the special comments, the data were presented as mean of triplicate experiments ± standard deviation (SD). Graphs were generated using OriginPro 8 (OriginLab Corporation, Northampton, MA, USA).

In general, plant materials were cultured for RNA purification and MeJA treatment. A cDNA library was constructed for transcriptome sequencing and then the sequencing results were assembled, functional annotated and classified based on BLAST searches against public databases. The qPCR was used to analyze the gene expression. HPLC-Q-TOF-MS/MS was conducted for metabolomics analysis and investigating metabolites accumulation. Metabolomics analysis was used to identify the biomarkers.

RESULTS AND DISCUSSION

Transcriptomic Analysis as a Powerful tool for the Discovery of Target Genes

Because there is no transcriptomic data in the coumarin-producing plants, the transcriptomic dataset of *P. praeruptorum* was constructed and the results were summarized in Table S2 and **Figure 3A**. The dataset could also be retrieved from NCBI SRA sequence database under project accession number SRX997427 and the assemblies of the transcriptome constructed in this study are listed in Table S3. For short, 59,346,260 raw reads, 62,386 transcripts, and 40,952 unigenes were obtained. Based on the results of the *de novo* sequencing, assembly and annotation, the unigenes that hit in at least one database were discovered and subjected to functional classification (**Figure 3**). The results showed that the categories related to the biosynthesis and transport of metabolites, such as "metabolic process," "catalytic activity," "secondary metabolite biosynthesis," and "transport" appeared in the three selected annotation and classification databases (**Figures 3B–D**). Thus, we found nearly all of the known genes encoding enzymes involved in the biosynthesis of the coumarin core compounds (**Table 1**). This result indicated that the method provided comprehensive and useful information on the function of the transcripts. Although the genes were annotated, the specific functions need to be verified. In addition, plenty of genes that did not participate in the target pathway need to be eliminated. Hence, a descent and specific dataset need to be constructed for further analysis.

Targeted Metabolic Pathways Analysis to Select Candidate CYP450 and MDR Genes Involved in the Biosynthesis and Transport of Coumarins

The main question resolved in this study was investigating the genes involved in coumarins biosynthesis and transport. Hence, the genes that may participate in this process were selected to narrow down the dataset. As shown in **Figure 1**, the enzymes involved in the monooxygenase reaction and hydroxylation reaction are largely unknown although some reports related to the reaction of core compounds formation had been published (Karamat et al., 2012, 2014; Vialart et al., 2012). Hence, the genes involved in monooxygenase reaction and hydroxylation reaction were selected as candidate genes. For example, cinnamate 4-hydroxylase that converts cinnamic acid to 4-coumaric acid is a CYP450 monooxygenase in the CYP73A family (Teutsch et al., 1993) and psoralen synthase that converts (+)-marmesin to psoralen is also a monooxygenase in the CYP71AJ family (Larbat et al., 2007, 2009). Therefore, the genes annotated as CYP450 genes, particular those in the CYP71AJ, CYP73, CYP84, and CYP98 clans which had been reported to be involved in coumarins biosynthesis were chosen for investigation (Bourgaud et al., 2006; Mizutani and Ohta, 2010; Nelson and Werck-Reichhart, 2011) (**Table 2**, Table S4). Hence, it could narrow down the query coverage (Bourgaud et al., 2006).

As an important complement to secondary metabolites biosynthesis, the transporters involved in the transport of
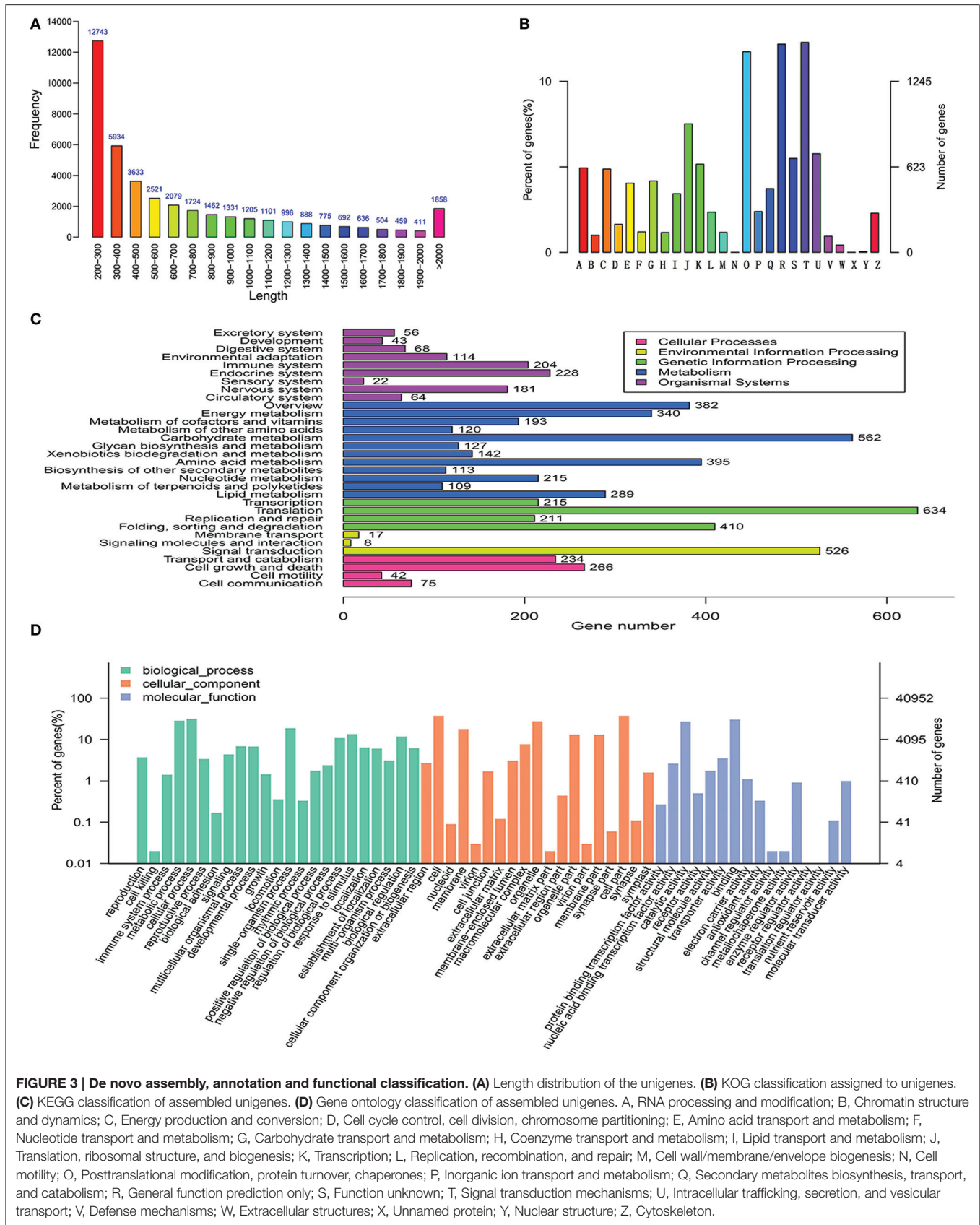
FIGURE 3 | De novo assembly, annotation and functional classification. (A) Length distribution of the unigenes. **(B)** KOG classification assigned to unigenes. **(C)** KEGG classification of assembled unigenes. **(D)** Gene ontology classification of assembled unigenes. A, RNA processing and modification; B, Chromatin structure and dynamics; C, Energy production and conversion; D, Cell cycle control, cell division, chromosome partitioning; E, Amino acid transport and metabolism; F, Nucleotide transport and metabolism; G, Carbohydrate transport and metabolism; H, Coenzyme transport and metabolism; I, Lipid transport and metabolism; J, Translation, ribosomal structure, and biogenesis; K, Transcription; L, Replication, recombination, and repair; M, Cell wall/membrane/envelope biogenesis; N, Cell motility; O, Posttranslational modification, protein turnover, chaperones; P, Inorganic ion transport and metabolism; Q, Secondary metabolites biosynthesis, transport, and catabolism; R, General function prediction only; S, Function unknown; T, Signal transduction mechanisms; U, Intracellular trafficking, secretion, and vesicular transport; V, Defense mechanisms; W, Extracellular structures; X, Unnamed protein; Y, Nuclear structure; Z, Cytoskeleton.

TABLE 1 | The genes involved in coumarins biosynthesis.

Gene name	Enzyme No.	Numbers
PAL	4.3.1.24	3
C4H	1.14.13.11	1
4CL	6.2.1.12	17
HCT	2.3.1.133	3
C3'H	1.14.13.36	1
CCoAOMT	2.1.1.104	3
CAOMT	2.1.1.68	7
CA2H		0
F6'H		2
C2'H		0
PS		5
AS		3
BMT	2.1.1.69	1
XMT	2.1.1.70	0

plant secondary metabolites play an increasingly significant role in biosynthetic studies (Yu and De Luca, 2013). Firstly, the differential subcellular localizations of the enzymes involved in biosynthesis indicate that compounds are synthesized in different organelles or tissues (Yazaki, 2005, 2006; Rea, 2007; Zhao and Dixon, 2010). Secondly, some of the genes responsible for the formation of plant secondary metabolites may be highly expressed in one tissue, whereas the metabolites may mainly accumulate in another tissue or organ (Yazaki, 2005, 2006; Rea, 2007; Zhao and Dixon, 2010). Hence, the pathway intermediates must be transferred between these locations and tissues to allow the enzymes, which displayed different subcellular localizations, to participate in the synthetic pathways (Yazaki, 2005). This just happened in coumarin compounds which were synthesized via phenylalanine in different organelles and then accumulated in different tissues and/or organelles for defense against pathogenic infections or other purposes. However, unlike the case for the genes that participate in coumarins biosynthesis, the transporters involved in coumarins transport have not been identified to date (Karamat et al., 2012, 2014; Vialart et al., 2012). And, no genes were annotated as being associated with coumarins transport despite of a large number of genes hit, annotation and classification as transporters (**Figure 3**). Of all the transcripts annotated for transport, the ATP-binding cassette (ABC) transporter, particularly the MDR (multidrug resistance protein, ABCC) have been reported to be associated with the transport of metabolites or to participate in transmembrane transport and/or the regulation of other transporters (Yazaki, 2005; Rea, 2007). Hence, the transcripts annotated as MDR were selected for investigating the genes involved in the transport of coumarins. In conclusion, 19 of 239 CYP450 genes and 55 of 236 ABC transporter genes were selected as candidate genes involved in the biosynthesis and transport of coumarin compounds according to previous studies (Yazaki, 2005; Mizutani and Ohta, 2010; Nelson and Werck-Reichhart, 2011) (**Table 2**, Table S5).

MeJA-induced Gene Expression Analysis to Narrow Down the Candidate CYP450 and MDR Genes

Although a large number of genes that did not participate in the biosynthesis or transport of coumarin compounds were eliminated, it was not practical to simultaneously analyze all of the genes above. In addition, not all genes interested us. Therefore, MeJA-induced gene expression analysis was conducted to further screen the genes actually involved in the biosynthesis and transport of coumarin compounds based on the fact that the MeJA signaling pathway is generally regarded as a transducer of elicitor-signal transduction that leads to the biosynthesis of plant secondary compounds (Zhao et al., 2005; Mizutani and Ohta, 2010; Nelson and Werck-Reichhart, 2011). For example, the exogenous application of MeJA resulted in elevated levels of active enzymes in soybean cell-suspension cultures and the accumulation of high amounts of phenolics and lignins (Lois et al., 1989; Mandal, 2010). Hence, members of the CYP71AJ, CYP73, CYP84, and CYP98 clans were selected as candidate genes and then subject to qRT-PCR-based expression analysis. As shown in **Figure 4A**, the genes labeled as 2, 3, 8, 9, 13, 17, and 19 (corresponding to *comp228_c0_seq1, comp30922_c1_seq1, comp36276_c0_seq5, comp9533_c0_seq1, comp24646_c0_seq1, comp89584_c0_seq1,* and *comp23746_c0_seq1* in Table S1 and abbreviated to *228, 30,922, 36,276, 9533, 24,646, 89,584,* and *23,746,* respectively) displayed up-regulated expression. Then, the up-regulated genes were further subject to tissue-specific expression analysis (**Figure 4B**) and time-dependent induction experiment (**Figure 5A**). It indicated that *30,922* expressed at the highest level in the leaves, whereas *36,276* expressed at the highest level in the stems, and there was no difference in expression levels of 89,584 in three tissues. The time-dependent induction experiment showed that all of the genes displayed different modes of expression. For instance, MeJA had little effect on the expression level of *9533* but had a strong effect on *36276*. *228* responded immediately after the addition of MeJA, whereas *30,922* did not respond for 3 h (**Figure 5A**). The trend in the responses showed an up-down regulated expression pattern that was consistent with the results of the previous study (Belhadj et al., 2008). It has been reported that the genes in tissues of different states (young or old) displayed a different expression pattern and this pattern was related to the accumulation of compounds (Li et al., 2002; Sun et al., 2011). Thus, the expression levels of the selected genes in tissues with different physiological statuses were analyzed (**Figures 5B,C**). The results showed that high levels of expression occurred in young stems, whereas low levels of expression occurred in mature tissue. However, the results for the leaves were partially opposite of those for the stems, that is, *228, 30,922, 36,276,* and *24,646* were expressed at a higher level in mature leaves than in young leaves. Interestingly, old leaves tend to have low expression levels due to their physiological status of no longer participating in the biosynthesis of metabolites (Li et al., 2002).

Similar to the procedure above, 55 MDR transporters were first selected from the 236 ABC-transporter transcripts and

TABLE 2 | Summary of the CYP families and ABC families in the *P. praeruptorum* transcriptome dataset.

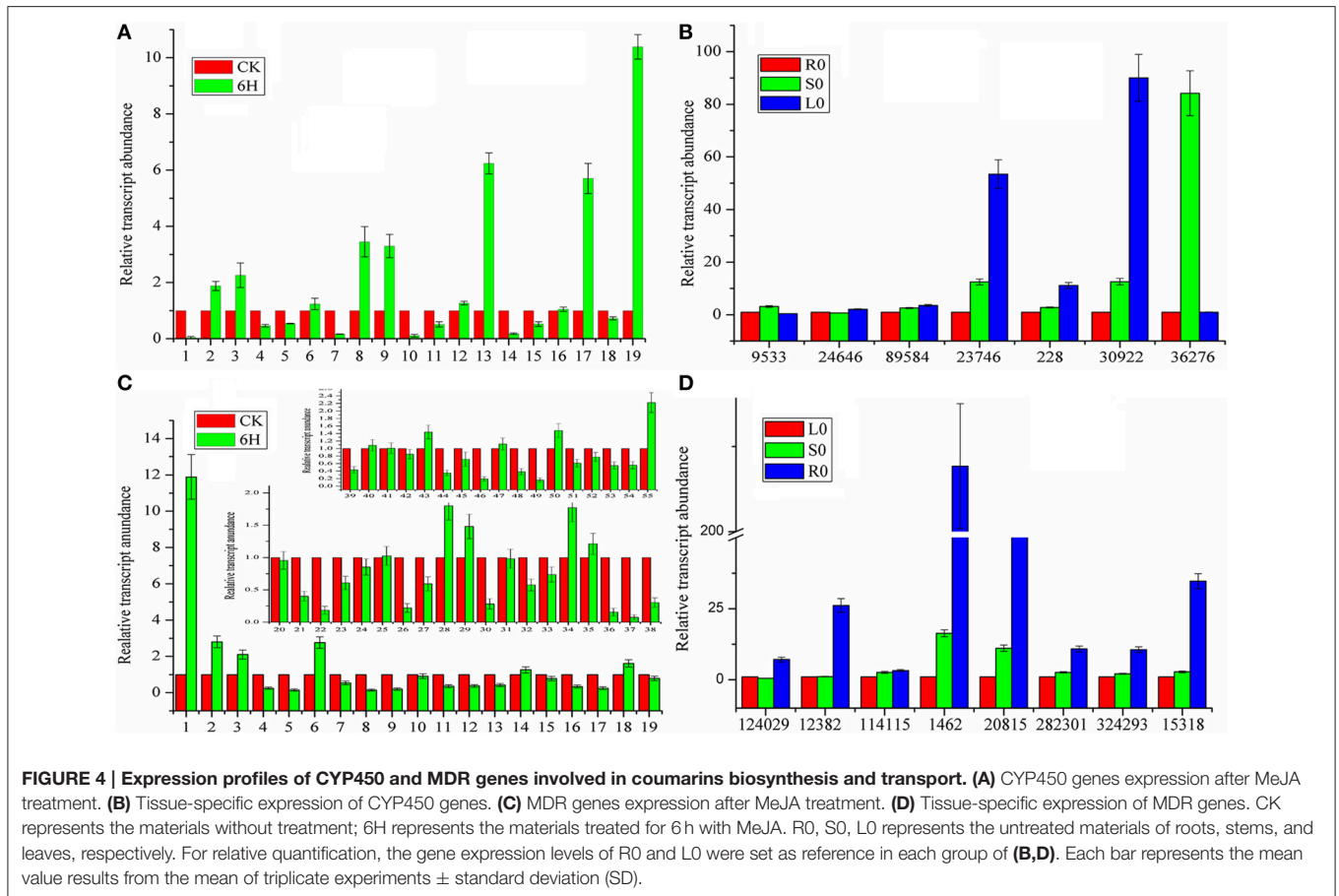| Clans	Gene	Gene No.	Clans	Gene	Gene No.
CYP71	A1,A2,A6,A8,A9,A22,		CYP704	C1	4
	A23,A25,A26,AJ1,AJ2,		CYP707	A1,A3,A4,A7	9
	AJ3,AJ4,B10,B12,B14,		CYP710	A1A1B1,B2B1A1,A2A1,A4A1	2
	B23,B34,B36,B37,BL3,		CYP711	A1	1
	D6,D7,D8,D9,D11,D12,		CYP714	A1	2
	D55,E1	61	CYP716	B1,B2	7
CYP72	A1,C1	12	CYP724	B1	3
CYP73	A10	1	CYP725	A1,A2	2
CYP74	A	1	CYP734	A1,A4	3
CYP75	A2,B1,B2	4	CYP735	A1	2
CYP76	A1,A2,B1,B6,B10, C1,		Others		8
	C2,C3,C4	25	Total		239
CYP77	A1,A3,A4	5			
CYP78	A5,A6,A9	3	ABCA	1,2,4,7,8	9
CYP80	B2	1	ABCB	1,2,4,6,9,10,11,13,15,	47
CYP81	D1,E1,F1	14		19,20,21,25,26,28,29	
CYP82	A1,A3,A4,C2,C4	13	ABCC	1,2,3,5,8,9,10,12,13,	
CYP83	A1,B1	2		14,15	55
CYP84	A1	6	ABCE	2	3
CYP85	A1	5	ABCF	1,3,4,5	9
CYP86	A1,A2,B1	7	ABCG	1,3,4,5,7,8,10,11,12,14,	
CYP89	A2,A9	4		15,16,20,21,22,25,27,	
CYP90	A1,B1,D1,D2	5		28,29,32,35,40,41,43	72
CYP93	A2,A3	4	ABCI	1,6,8,10,11,19,20	8
CYP94	A1,A2	12	ABCPDR	1,2,3,4,7,12,15	28
CYP97	A3,B2,C1	8	Otherstotal		5
CYP98	A2,A3	3	Total		236

then, the corresponding genes were screened using MeJA-induced expression analysis. As shown in **Figure 4C**, 1, 2, 3, 6, 28, 29, 34, and 55 (corresponding to *124,029, 114,115, 12,382, 1462, 20,815, 282,301, 324,293,* and *15,318*) displayed up-regulated expression. The tissue-specific expression analysis showed that all of the genes highly expressed in roots. Interestingly, *1462* and *20,815* had the highest expression levels and their expression levels in stems were also higher than those of the other MDR-transporter genes, whereas, the others showed no difference in the expression levels of the stems and leaves (**Figure 4D**). The time-dependent MeJA-induced expression analysis showed that these genes had the same expression pattern with CYP450s (**Figure 5D**). Analysis of the expression profiles of the selected genes at different physiological statuses revealed that they expressed at a higher level in mature leaves and had a lower expression level in young leaves (**Figure 5F**). However, this trend was partially reversed in the stems and a relatively higher expression level was found in young stems (**Figure 5E**). The differential expression behavior was related to the profiles of the secondary metabolites in different tissues.

Above all, the up-/down-regulated expression patterns were consistent with those observed in the previous study (Belhadj et al., 2008) (**Figures 5A,D**). Upon MeJA treatment, only 7 of

19 CYP450 genes and 8 of 55 MDR transporters displayed an increased expression level, which greatly reduced the number of genes to be evaluated. While, whether all of the genes we investigated participated in the biosynthesis and/or transport of coumarin compounds and whether these genes were associated with the compounds accumulation depended on the quantitative analysis of coumarin contents.

Metabolite Analysis and Structural Identification

To identify the coumarin compounds in *P. praeruptorum*, HPLC coupled with Q-TOF MS was used because this method was rapid and could provide accurate mass measurements (Ling et al., 2013). Additionally, the peak areas could also be used for relative quantification (Zhang et al., 2014). However, in our first attempt to utilize this method, less than 10 peaks were obtained using HPLC. The same problem occurred in LC-MS (data no given). Considering that the peaks corresponding to the high-abundance compounds (such as praeruptorin A and praeruptorin C) may hide those peaks corresponding to the low-abundance compounds, a C18E cartridge was used to separate the two fractions for the first time according to the method described in materials and methods part. After this process of treatment, peaks of low-abundance compounds were
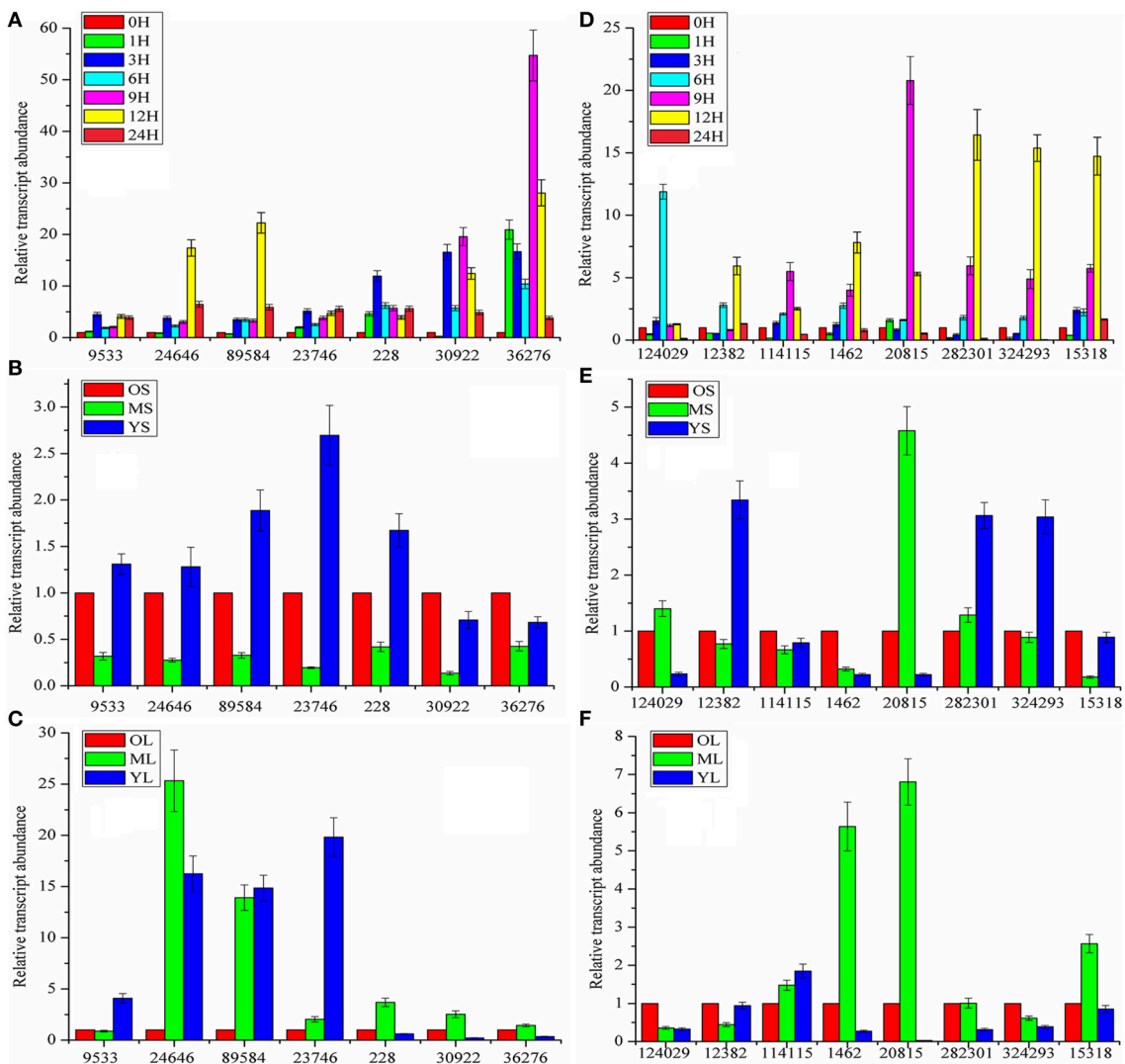
FIGURE 4 | Expression profiles of CYP450 and MDR genes involved in coumarins biosynthesis and transport. (A) CYP450 genes expression after MeJA treatment. **(B)** Tissue-specific expression of CYP450 genes. **(C)** MDR genes expression after MeJA treatment. **(D)** Tissue-specific expression of MDR genes. CK represents the materials without treatment; 6H represents the materials treated for 6 h with MeJA. R0, S0, L0 represents the untreated materials of roots, stems, and leaves, respectively. For relative quantification, the gene expression levels of R0 and L0 were set as reference in each group of **(B,D)**. Each bar represents the mean value results from the mean of triplicate experiments ± standard deviation (SD).

observed in the extracts using HPLC and LC-MS (**Figure 6**), which made it possible to analyze compound contents. **Figure 6** shows the distinct differences in the contents of the compounds and their peak numbers, and it can be easily observed that many low-abundance compounds were identified in the extracts of the roots, stems, and leaves. Interestingly, roots tend to have more numbers of compounds and higher peak intensity than stems and leaves which could account for why people choose roots as medical parts (Commission, 2010). Another phenomenon is that, in a "relative" point of view, the leaves seem have higher ratio of maximal polar components than stems and roots. This could be interpreted as that leaves are the main sits for biosynthesis, hence, the maximal polar components (such as umbelliferone) mainly existed in this tissue (**Figure 6**). **Table 3** shows the 19 coumarin compounds (the structures are listed in Figure S1) identified by their signals in the corresponding extracted-ion chromatograms and their second-order fragment ion. Their identities were confirmed by previous studies or standard compounds. Here, we took (±)-Praeruptorin A (Pd-Ia) as an example to illustrate the process of identification for it constituted the main composition of *P. praeruptorum* (Kong et al., 1996). Firstly, according to the mass spectrometry data, the exact mass of the quasi-molecular ion 404.1704 (m/z, $[M+NH_4]^+$) and the MS/MS fragment ions 245.0808, 227.0699, 175.0385, and 83.0502 (m/z)

were obtained (Figure S2). And then, Agilent LC-Q-TOF-MS MassHunter Qualitative Analysis Software was used to calculate the elemental composition and five different possible element compositions were produced. Thirdly, the MS, MS/MS fragments and elemental composition all compared to available literature information, combined with the deduced fragmentation pathway (Figure S3), (+)-Praeruptorin A (Pd-Ia, $C_{24}H_{26}O_7$) was found to be the most possible compound. For further confirmation, standard compound was used and the result was consistent with the identification outcome. Subsequently, another 18 coumarin compounds were identified.

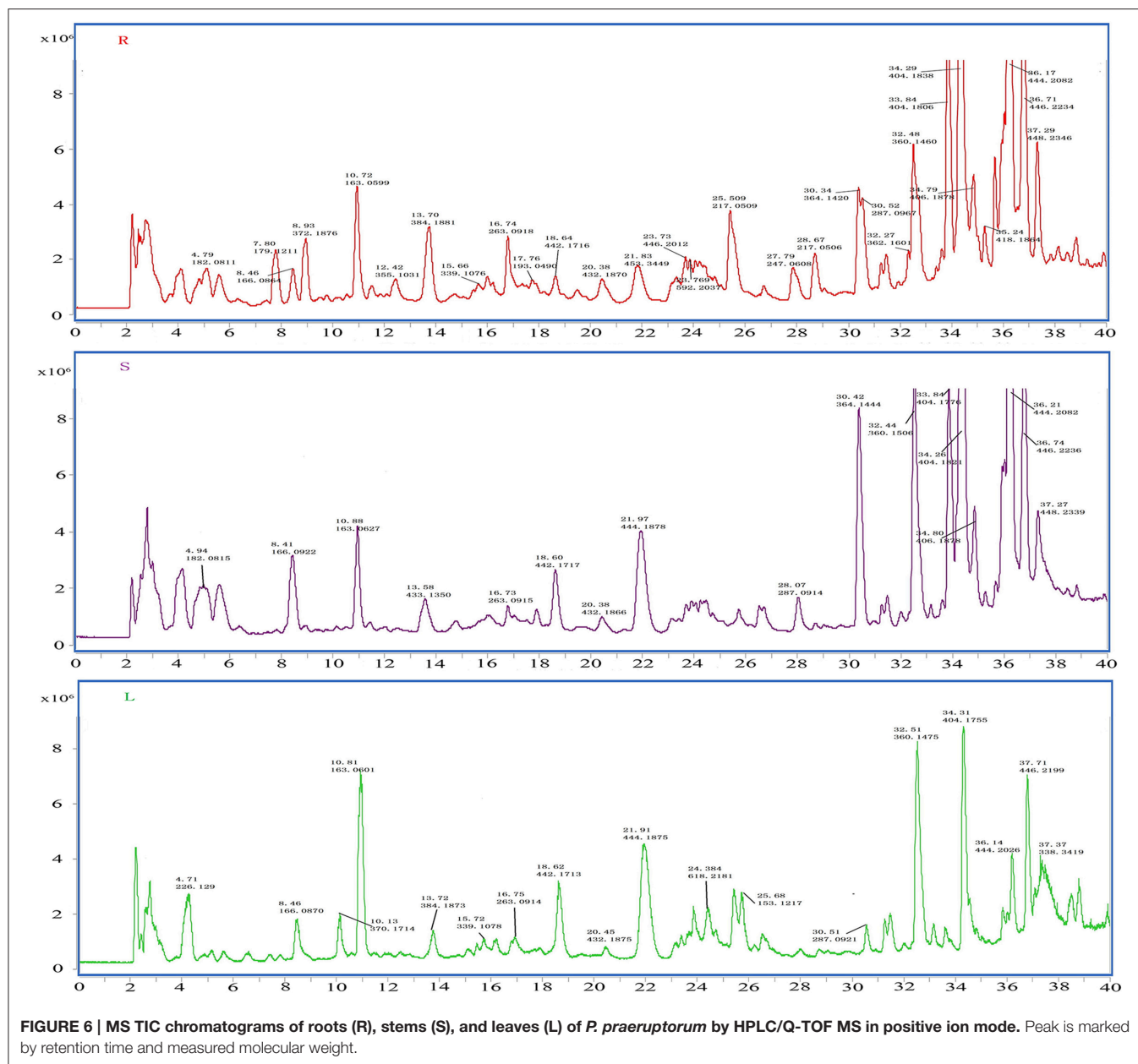## Metabolomics Analysis and Potential Biomarkers Exploration

After the method of separation was established, discovering the significant difference compounds for a correlation study of gene expression and metabolites accumulation was urgent. To solve this problem, metabolomics analysis was conducted due to its powerful performance in natural product discovery (Yan et al., 2015). At first, HPLC combined with Q-TOF MS was first conducted for data collection, and then the data were normalized according to the method described below in materials and methods part. Subsequently, the 1713 normalized variables were imported into the SIMCA-P V13.0 (Umetrics, Sweden) platform for multivariate analysis. The PCA score plots

FIGURE 5 | Time and growth-stage-specific gene expression. (A,D) Time-dependent expression levels of CYP450 and MDR genes. **(B,E)** Expression levels of CYP450 and MDR genes in stems (S) at different growth stages. **(C,F)** Expression levels of CYP450 and MDR in leaves (L) at different growth stages. 0, 1, 3, 6, 9, 12, 24H represent the time interval after MeJA addition; O, M, Y: short for old, mature, and young, respectively. For relative quantitative, the gene expression levels of 0H **(A,D)**, OS **(B,E)**, and OL **(C,F)** was set as reference in each group. Each bar are presented as mean of triplicate experiments±standard deviation (SD).

(**Figure 7A**) showed that the components of the roots, stems and leaves were clearly clustered into three groups. The OPLS-DA scores also showed that the components of the roots, stems and leaves fell into well-distinguished classes (**Figure 7B**, R^2X = 0.957, Q^2 = 0.885, were acceptable), indicating that the study had good levels of predictability and reliability. OPLS-DA loading plots were used to determine which compounds contributed highly to the differences among the three tissues. As shown in **Figure 7C**, 405 of 1713 variables with a VIP value of >1.000 were denoted as potential contributors and they were colored in red. To select potential biomarkers for a correlation study, student's *t*-test was used with the *P*-value set to 0.05 to designate significantly different variables. After two rounds of selection, 65 marker metabolites were selected.

However, not all of the 65 marker metabolites were coumarin compounds which we were mainly interested in. Hence, the metabolites identified in **Table 3** were used as a coumarin compounds reference dataset for further selection. Finally, 8 significantly different coumarin compounds were identified as potential biomarkers for further study. For example, the peak intensities of coumarin compounds present at different levels in the roots, stems and leaves were first investigated, and the results are shown in **Figure 7D**. The result indicated that there was an obvious difference in the contents of the roots, stems, and leaves. These differences were consistent with the results listed in **Figures 4**, **6** in some extent and it may be related to the levels of expression of the genes involved in their biosynthesis and transport.

FIGURE 6 | MS TIC chromatograms of roots (R), stems (S), and leaves (L) of *P. praeruptorum* by HPLC/Q-TOF MS in positive ion mode. Peak is marked by retention time and measured molecular weight.

Merging the Gene Expression and Compound Accumulation to Identify the Genes Involved in Coumarins Biosynthesis and Transport

After the candidate genes and potential biomarkers were discovered, the relationship between gene expression and compound accumulation arose. To address these issues, the differences in the peaks intensity of the coumarin compounds in roots, stems and leaves were investigated, and the results are shown in **Figure 7D**. As shown, the levels of umbelliferone, praeroside VI and scopolin were higher in leaves, whereas those of khellactone, qianhucoumarin J, praeruptorin B, and praeruptorin G were higher in roots and stems. The reason for

these phenomena was that the former compounds synthesized during the core formation and the latter required prenylation, hydroxylation or cyclization or other structural modifications (**Figure 1**). Consistent with the tissue-specific difference in the level of compounds accumulation, the related gene expression level also displayed a tissue-specific pattern (**Figure 4B**). In particular, the expression level of *23,746, 228,* and *30,922* tended toward a leaves-specific expression pattern and those of *36,276* and *9533* tended toward a stems-specific expression pattern. These results indicated that *23,746, 228,* and *30,922* may be related to the formation of coumarin core compounds and *36,276* and *9533* participated in the prenylation, hydroxylation and cyclization or structural modification. Considering the tissue-specific patterns of gene expression, the question is
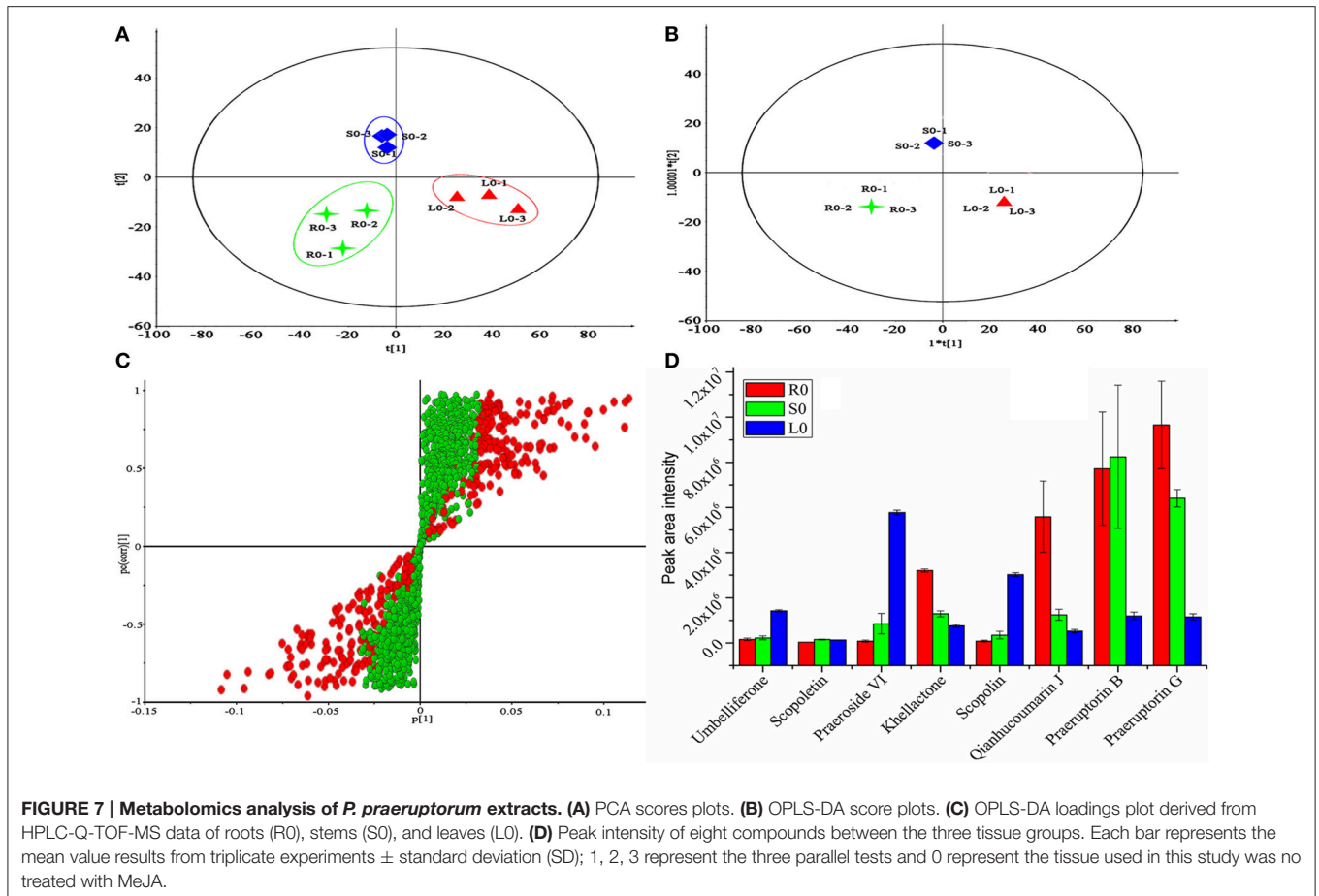
TABLE 3 | Compounds detected in the extracts of *P. praeruptorum* by HPLC-Q-TOF-MS/MS.

Formula	Identification	$t_{R(min)}$	Monitoring ion	Measured (m/z)	Calculated (m/z)	Error (ppm)	Major fragment ions (m/z)
$C_9H_6O_3$	Umbelliferone	10.724	$[M+H]^+$	163.5099	163.1421	0.17	135.0442,119.0493
$C_{16}H_{18}O_9$	Scopolin	12.423	$[M+H]^+$	355.1029	355.1024	−1.62	163.0385,145.0284,135.0440,117.0335
$C_{16}H_{18}O_8$	4-Methylumbelliferyl glucoside	15.669	$[M+H]^+$	339.1073	339.1072	0.47	147.0438,119.0490,91.0548
$C_{14}H_{14}O_5$	Khellactone	16.749	$[M+NH_4]^+$	263.0918	263.0914	−1.68	229.0862,203.0700,175.0391,59.0506
$C_{10}H_8O_4$	Scopoletin	17.662	$[M+H]^+$	193.0493	193.0495	1.43	178.0255,150.0308,133.0281,122.0360,94.0417
$C_{20}H_{24}O_{10}$	Apterin	18.641	$[M+NH_4]^+$	442.171	442.1708	−0.45	245.0799,203.0694,187.0381
$C_{20}H_{26}O_{10}$	Praeroside VI	21.974	$[M+NH_4]^+$	444.1864	444.1871	−1.63	247.0962,229.0858,175.0388
$C_{28}H_{30}O_{13}$	Praeroside I	23.761	$[M+NH_4]^+$	592.2028	592.2025	−0.58	313.0894,263.0912,151.0377
$C_{12}H_8O_4$	5-Methoxypsoralen	25.501	$[M+H]^+$	217.0492	217.0495	1.29	202.0252,174.0306,146.0275,118.0411,90.0466
$C_{12}H_8O_4$	8-Methoxypsoralen	28.677	$[M+H]^+$	217.0501	217.0495	−2.75	202.0256,174.0360,146.0385,118.0412,90. 0468,77.0396
$C_{18}H_{18}O_7$	Qianhucoumarin D	30.344	$[M+NH_4]^+$	364.14	364.1391	−2.61	245.0810,227.0703,175.0389
$C_{19}H_{20}O_6$	Qianhucoumarin A	32.271	$[M+NH_4]^+$	362.1606	362.1598	−2.21	243.0644,213.0694,57.0708
$C_{19}H_{18}O_6$	Pd-Ib	32.463	$[M+NH_4]^+$	360.1447	360.1442	−1.49	243.0655,215.0704,83.0501,55.0556
$C_{21}H_{22}O_7$	(±)-Praeruptorin A(Pd-Ia)	34.296	$[M+NH_4]^+$	404.1705	404.1704	−0.29	245.0808,227.0699,175.0385
$C_{21}H_{24}O_7$	Peucedanocoumarin I	34.792	$[M+NH_4]^+$	406.1863	406.186	−0.68	245.0806,227.0698,175.0383,85.0651,57.0708
$C_{22}H_{24}O_7$	Qianhucoumarin J	35.243	$[M+NH_4]^+$	418.1856	418.186	0.97	227.0699,83.0498,55.0552
$C_{24}H_{26}O_7$	(+)-Praeruptorin B(Pd-II)	36.172	$[M+NH_4]^+$	444.2104	444.2017	0.6	327.1219,227.0688,83.0498,55.0553
$C_{24}H_{28}O_7$	Praeruptorin F	36.713	$[M+NH_4]^+$	446.2176	446.2173	−0.57	245.0801,227.0696,175.0384,83.0499,55.0553
$C_{24}H_{30}O_7$	Praeruptorin G	37.293	$[M+NH_4]^+$	448.2334	448.233	−1.04	245.0800,227.0695,175.0381,85.0659,57.0708

how these processes occur. As is known to all, different compounds were synthetized not only in different tissues but also in different organelles, which was proved by subcellular localization experiments (Karamat et al., 2014). For instance, CYP450 proteins are heme-thiolate membrane-bound proteins that are generally bound to the surface of the endoplasmic reticulum (Kawai et al., 2014), whereas O-methyltransferase tended to express throughout the cytoplasm (Yazaki, 2005) (our unpublished data). Moreover, the leaves and stems are likely to be the main sites of coumarins biosynthesis, but these compounds are localized mainly in the roots which are the medicinal portions of *P. praeruptorum* (Commission, 2010). Given the existence of transporters, these facts appear to be understandable, and there are many reports concerning the transporters (Li et al., 2002; Yazaki, 2005; Rea, 2007; Yu and De Luca, 2013). As shown in **Figure 4D**, the selected transporters were expressed at a higher level in the roots than in other tissues, which was consistent with the differential content levels of the relevant compounds (**Figure 7D**). This phenomenon indicated that the selected genes participated in the transport of coumarin compounds. Does a transporter translocate a specific compound or perhaps two/more compounds? It was widely accepted that ABC transporters have broad substrate specificity (Yazaki, 2006). However, the results of our study indicated that a transporter specifically translocated one kind of compounds. Although all of the selected genes were highly expressed in the roots, *1462,*

20,815, and *15,318* tended to highly express in mature or young leaves and *124,029* and *324,293* tended to highly express in old leaves (**Figures 5E,F**). The results indicated that the former transcripts may participate in the transport of coumarins core compounds and the latter transcripts may participate in the transport of the completed coumarin compounds. This tendency for a certain compound to be transported via a various mechanism was analyzed using vacuolar-membrane vesicles purified from the red beets of *Beta vulgaris*. The results demonstrated that two phenol glucosides, p-hydroxycinnamic acid, and p-hydroxybenzoic acid, were apparently transported via a H^+-gradient-dependent mechanism, the glutathione conjugate of an analog of the herbicide chlorsulfuron appeared to be transported via an ABC transporter (Yazaki, 2006). Transport mechanisms may also function in *P. praeruptorum*, and this is why its roots are generally chosen for medicinal usage (Commission, 2010). As shown in Figure S4, the content of "made-up compound," such as qianhucoumarin J and praeruptorin G, accumulated in a time-dependent pattern but the content of the coumarin core compounds (such as umbelliferone and scopoletin) did not change over time. The results also showed a decreasing content of the 8 selected compounds. This phenomenon was consistent with the fact that this type of plant is generally being harvested at 1 or 2 year after planting, and its quality would affect being grown for a long period particularly after it had bloomed.

FIGURE 7 | Metabolomics analysis of *P. praeruptorum* extracts. (A) PCA scores plots. **(B)** OPLS-DA score plots. **(C)** OPLS-DA loadings plot derived from HPLC-Q-TOF-MS data of roots (R0), stems (S0), and leaves (L0). **(D)** Peak intensity of eight compounds between the three tissue groups. Each bar represents the mean value results from triplicate experiments ± standard deviation (SD); 1, 2, 3 represent the three parallel tests and 0 represent the tissue used in this study was no treated with MeJA.

CONCLUSIONS

In this study, a large amount of transcriptomic data for *P. praeruptorum* was assembled for the first time. After sequence assembly and functional annotation, the putative CYP450 and MDR genes involved in the biosynthesis and transport of coumarin compounds were selected. Based on the results of the MeJA-induced expression analysis, 7 CYP450s and 8 MDR transporter unigenes were selected as candidate transcripts. Then, HPLC-Q-TOF-MS/MS-based metabolomics analysis were merged to identify the differential chemical constituents and quantify their contents in roots, stems and leaves for a correlation study of gene expression and metabolites accumulation in coumarins biosynthesis and transport. The results indicated that *23,746, 228,* and *30,922* may be related to the formation of the coumarin core compounds whereas *36,276* and *9533* participated in the prenylation, hydroxylation, cyclization or structural modification. It appeared that *1462, 20,815,* and *15,318* participated in the transport of coumarin core compounds while *124,029* and *324,293* participated in the transport of the "made-up cumpounds." Additionally, studies of the tissue-specific, growth-stage-specific, and time-dependent accumulation of coumarin compounds were performed for the first time. The results would facilitate investigations of

the biosynthesis and transport of coumarin compounds and promote the medicinal usage of components of *P. praeruptorum*. In addition, the method used in this study would also provide a guideline to gene discovery of other non-model plants.

AUTHOR CONTRIBUTIONS

LK and YZ conceived and designed the experiments and contributed the reagents/materials/analysis tools. QZ and SX analyzed the data and wrote the manuscript. CH, JX, and MC performed the RNA isolation experiment. JL and TL provided the *P. praeruptorum* materials and participated in the design of the study. CH, SX, and JX helped to analyze the data and draft the manuscript. YC coordinated the study and revised the manuscript. SX, CH, JX, and MC participated in picture editing. LK, CY, JL, TL, and QZ conducted in compound identification. All authors read and approved the final manuscript.

ACKNOWLEDGMENTS

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: http://journal.frontiersin.org/article/10.3389/fpls.2015.00996

REFERENCES

Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J. H., Zhang, Z., Miller, W., et al. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 25, 3389–3402. doi: 10.1093/nar/25.17.3389

Ansorge, W. J. (2009). Next-generation DNA sequencing techniques. *New Biotechnol.* 25, 195–203. doi: 10.1016/j.nbt.2008.12.009

Belhadj, A., Telef, N., Saigne, C., Cluzet, S., Barrieu, F., Hamdi, S., et al. (2008). Effect of methyl jasmonate in combination with carbohydrates on gene expression of PR proteins, stilbene and anthocyanin accumulation in grapevine cell cultures. *Plant Physiol. Bioch.* 46, 493–499. doi: 10.1016/j.plaphy.2007.12.001

Bourgaud, F., Hehn, A., Larbat, R., Doerper, S., Gontier, E., Kellner, S., et al. (2006). Biosynthesis of coumarins in plants: a major pathway still to be unravelled for cytochrome P450 enzymes. *Phytochem. Rev.* 5, 293–308. doi: 10.1007/s11101-006-9040-2

Choi, B., Kang, S., Bae, H., Lim, H., Bang, W., and Bae, H. (2012). Transcriptional analysis of the ρ-coumarate 3-hydroxylase (C3H) gene from *Hibiscus cannabinus* L. during developmental stages in various tissues and in response to abiotic stresses. *Res. J. Biotechnol.* 7, 23–33.

Commission, C. P. (2010). *Pharmacopoeia of the People's Republic of China, Vol. 1.* Beijing: People's Medical Publishing House.

Dimmer, E. C., Huntley, R. P., Alam-Faruque, Y., Sawford, T., O'Donovan, C., Martin, M. J., et al. (2012). The UniProt-GO Annotation database in 2011. *Nucleic Acids Res.* 40, D565–D570. doi: 10.1093/nar/gkr1048

Du, J. L., Yuan, Z. F., Ma, Z. W., Song, J. Z., Xie, X. L., and Chen, Y. L. (2014). KEGG-PATH: kyoto encyclopedia of genes and genomes-based pathway analysis using a path analysis model. *Mol. Biosyst.* 10, 2441–2447. doi: 10.1039/C4MB00287C

Gaid, M. M., Sircar, D., Müller, A., Beuerle, T., Liu, B. Y., Ernst, L., et al. (2012). Cinnamate: CoA ligase initiates the biosynthesis of a benzoate-derived xanthone phytoalexin in *Hypericum calycinum* cell cultures. *Plant Physiol.* 160, 1267–1280. doi: 10.1104/pp.112.204180

Hirai, M. Y., Sugiyama, K., Sawada, Y., Tohge, T., Obayashi, T., Suzuki, A., et al. (2007). Omics-based identification of *Arabidopsis* Myb transcription factors regulating aliphatic glucosinolate biosynthesis. *Proc. Natl. Acad. Sci. U.S.A.* 104, 6478–6483. doi: 10.1073/pnas.0611629104

Kanehisa, M., Goto, S., Kawashima, S., and Nakaya, A. (2002). The KEGG databases at GenomeNet. *Nucleic Acids Res.* 30, 42–46. doi: 10.1093/nar/30.1.42

Karamat, F., Olry, A., Doerper, S., Vialart, G., Ullmann, P., Werck-Reichhart, D., et al. (2012). CYP98A22, a phenolic ester 3′-hydroxylase specialized in the synthesis of chlorogenic acid, as a new tool for enhancing the furanocoumarin concentration in *Ruta graveolens*. *BMC Plant Biol.* 12:152. doi: 10.1186/1471-2229-12-152

Karamat, F., Olry, A., Munakata, R., Koeduka, T., Sugiyama, A., Paris, C., et al. (2014). A coumarin-specific prenyltransferase catalyzes the crucial biosynthetic reaction for furanocoumarin formation in *parsley*. *Plant J.* 77, 627–638. doi: 10.1111/tpj.12409

Kawai, Y., Ono, E., and Mizutani, M. (2014). Evolution and diversity of the 2-oxoglutarate-dependent dioxygenase superfamily in plants. *Plant J.* 78, 328–343. doi: 10.1111/tpj.12479

Kim, I. A., Kim, B. G., Kim, M., and Ahn, J. H. (2012). Characterization of hydroxycinnamoyltransferase from rice and its application for biological synthesis of hydroxycinnamoyl glycerols. *Phytochemistry* 76, 25–31. doi: 10.1016/j.phytochem.2011.12.015

Kong, L. Y., Li, Y., Min, Z. D., Li, X., and Zhu, T. R. (1996). Coumarins from *Peucedanum praeruptorum*. *Phytochemistry* 41, 1423–1426. doi: 10.1016/0031-9422(95)00783-0

Kumar, A., Maurya, R., Sharma, S., Ahmad, P., Singh, A., Bhatia, G., et al. (2009). Pyranocoumarins: a new class of anti-hyperglycemic and anti-dyslipidemic agents. *Bioorg. Med. Chem. Lett.* 19, 6447–6451. doi: 10.1016/j.bmcl.2009.09.031

Lamesch, P., Berardini, T. Z., Li, D. H., Swarbreck, D., Wilks, C., Sasidharan, R., et al. (2012). The *Arabidopsis* Information Resource (TAIR): improved gene annotation and new tools. *Nucleic Acids Res.* 40, D1202–D1210. doi: 10.1093/nar/gkr1090

Larbat, R., Hehn, A., Hans, J., Schneider, S., Jugdé, H., Schneider, B., et al. (2009). Isolation and functional characterization of CYP71AJ4 encoding for the first P450 monooxygenase of angular furanocoumarin biosynthesis. *J. Biol. Chem.* 284, 4776–4785. doi: 10.1074/jbc.M807351200

Larbat, R., Kellner, S., Specker, S., Hehn, A., Gontier, E., Hans, J., et al. (2007). Molecular cloning and functional characterization of psoralen synthase, the first committed monooxygenase of furanocoumarin biosynthesis. *J. Biol. Chem.* 282, 542–554. doi: 10.1074/jbc.M604762200

Li, D., Ono, N., Sato, T., Sugiura, T., Altaf-Ul-Amin, M., Ohta, D., et al. (2015). Targeted integration of RNA-seq and metabolite data to elucidate curcuminoid biosynthesis in four curcuma species. *Plant Cell Physiol.* 56, 843–851. doi: 10.1093/pcp/pcv008

Li, S., Yi, Y., Wang, Y., Zhang, Z., and Beasley, R. S. (2002). Camptothecin accumulation and variations in *camptotheca*. *Planta Med.* 68, 1010–1016. doi: 10.1055/s-2002-35652

Ling, Y., Li, Z., Chen, M., Sun, Z., Fan, M., and Huang, C. (2013). Analysis and detection of the chemical constituents of *Radix Polygalae* and their metabolites in rats after oral administration by ultra high-performance liquid chromatography coupled with electrospray ionization quadrupole time-of-flight tandem mass spectrometry. *J. Pharmaceut. Biomed.* 85, 1–13. doi: 10.1016/j.jpba.2013.06.011

Lois, R., Dietrich, A., Hahlbrock, K., and Schulz, W. (1989). A phenylalanine ammonia-lyase gene from parsley: structure, regulation and identification of elicitor and light responsive cis-acting elements. *EMBO J.* 8, 1641–1648.

Mandal, S. (2010). Induction of phenolics, lignin and key defense enzymes in eggplant (*Solanum melongena* L.) roots in response to elicitors. *Afr. J. Biotechnol.* 9, 8038–8047. doi: 10.5897/AJB10.984

Mehrshahi, P., Stefano, G., Andaloro, J. M., Brandizzi, F., Froehlich, J. E., and DellaPenna, D. (2013). Transorganellar complementation redefines the biochemical continuity of endoplasmic reticulum and chloroplasts. *Proc. Natl. Acad. Sci. U.S.A.* 110, 12126–12131. doi: 10.1073/pnas.1306331110

Mizutani, M., and Ohta, D. (2010). Diversification of P450 genes during land plant evolution. *Annu. Rev. Plant Biol.* 61, 291–315. doi: 10.1146/annurev-arplant-042809-112305

Movahedi, S., Van, B. M., Heyndrickx, K. S., and Vandepoele, K. (2012). Comparative co-expression analysis in plant biology. *Plant Cell Environ.* 35, 1787–1798. doi: 10.1111/j.1365-3040.2012.02517.x

Mutwil, M., Klie, S., Tohge, T., Giorgi, F. M., Wilkins, O., Campbell, M. M., et al. (2011). PlaNet: combined sequence and expression comparisons across plant networks derived from seven species. *Plant Cell.* 23, 895–910. doi: 10.1105/tpc.111.083667

Nelson, D., and Werck-Reichhart, D. (2011). A P450-centric view of plant evolution. *Plant J.* 66, 194–211. doi: 10.1111/j.1365-313X.2011.04529.x

Pfaffl, M. (2001). A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Res.* 29:e45. doi: 10.1093/nar/29.9.e45

Rea, P. A. (2007). Plant ATP-binding cassette transporters. *Annu. Rev. Plant Biol.* 58, 347–375. doi: 10.1146/annurev.arplant.57.032905.105406

Saito, K., Hirai, M. Y., and Yonekura-Sakakibara, K. (2008). Decoding genes with coexpression networks and metabolomics-'majority report by precogs'. *Trends Plant Sci.* 13, 36–43. doi: 10.1016/j.tplants.2007.10.006

Saito, K., and Matsuda, F. (2010). Metabolomics for functional genomics, systems biology, and biotechnology. *Annu. Rev. Plant Biol.* 61, 463–489. doi: 10.1146/annurev.arplant.043008.092035

Schuster, S. (2008). Next-generation sequencing transforms today's biology. *Nat. Methods.* 5, 16–18. doi: 10.1038/nmeth1156

Sun, Y. Z., Luo, H. M., Li, Y., Sun, C., Song, J. Y., Niu, Y. Y., et al. (2011). Pyrosequencing of the Camptotheca acuminata transcriptome reveals putative genes involved in camptothecin biosynthesis and transport. *BMC Genomics* 12:533. doi: 10.1186/1471-2164-12-533

Teutsch, H. G., Hasenfratz, M. P., Lesot, A., Stoltz, C., Garnier, J. M., Jeltsch, J. M., et al. (1993). Isolation and sequence of a cDNA encoding the Jerusalem artichoke cinnamate 4-hydroxylase, a major plant cytochrome P450 involved in the general phenylpropanoid pathway. *Proc. Natl. Acad. Sci. U.S.A.* 90, 4102–4106. doi: 10.1073/pnas.90.9.4102

Usadel, B., Obayashi, T., Mutwil, M., Giorgi, F. M., Bassel, G. W., Tanimoto, M., et al. (2009). Co-expression tools for plant biology: opportunities for hypothesis generation and caveats. *Plant Cell Environ.* 32, 1633–1651. doi: 10.1111/j.1365-3040.2009.02040.x

Vialart, G., Hehn, A., Olry, A., Ito, K., Krieger, C., Larbat, R., et al. (2012). A 2-oxoglutarate-dependent dioxygenase from *Ruta graveolens* L. exhibits p-coumaroyl CoA 2'-hydroxylase activity (C2'H): a missing step in the synthesis of umbelliferone in plants. *Plant J.* 70, 460–470. doi: 10.1111/j.1365-313X.2011.04879.x

Wagner, G. P., Kin, K., and Lynch, V. J. (2012). Measurement of mRNA abundance using RNA-seq data: RPKM measure is inconsistent among samples. *Theory Biosci.* 131, 281–285. doi: 10.1007/s12064-012-0162-3

Wang, Y., Wang, H., Fan, R., Yang, Q., and Yu, D. (2014). Transcriptome analysis of soybean lines reveals transcript diversity and genes involved in the response to common cutworm (*Spodoptera litura Fabricius*) feeding. *Plant Cell Environ.* 37, 2086–2101. doi: 10.1111/pce.12296

Wu, J., Fong, W. F., Zhang, J. X., Leung, C. H., Kwong, H. L., Yang, M. S., et al. (2003). Reversal of multidrug resistance in cancer cells by pyranocoumarins isolated from *Radix Peucedani*. *Eur. J. Pharmacol.* 473, 9–17. doi: 10.1016/S0014-2999(03)01946-0

Yamazaki, M., Mochida, K., Asano, T., Nakabayashi, R., Chiba, M., Udomson, N., et al. (2013). Coupling deep transcriptome analysis with untargeted metabolic profiling in *Ophiorrhiza pumila* to further the understanding of the biosynthesis of the anti-cancer alkaloid camptothecin and anthraquinones. *Plant Cell Physiol.* 54, 686–696. doi: 10.1093/pcp/pct040

Yan, B., Deng, Y. P., Hou, J. J., Bi, Q. R., Yang, M., Jiang, B. H., et al. (2015). UHPLC-LTQ-Orbitrap MS combined with spike-in method for plasma metabonomics analysis of acute myocardial ischemia rats and pretreatment effect of Danqi Tongmai tablet. *Mol. Biosyst.* 11, 486–496. doi: 10.1039/C4MB00529E

Yazaki, K. (2005). Transporters of secondary metabolites. *Curr. Opin. Plant. Biol.* 8, 301–307. doi: 10.1016/j.pbi.2005.03.011

Yazaki, K. (2006). ABC transporters involved in the transport of plant secondary metabolites. *FEBS Lett.* 580, 1183–1191. doi: 10.1016/j.febslet.2005.12.009

Yazaki, K., Sugiyama, A., Morita, M., and Shitan, N. (2008). Secondary transport as an efficient membrane transport mechanism for plant secondary metabolites. *Phytochem. Rev.* 7, 513–524. doi: 10.1007/s11101-007-9079-8

Yonekura-Sakakibara, K., Fukushima, A., Nakabayashi, R., Hanada, K., Matsuda, F., Sugawara, S., et al. (2012). Two glycosyltransferases involved in anthocyanin modification delineated by transcriptome independent component analysis in *Arabidopsis thaliana*. *Plant J.* 69, 154–167. doi: 10.1111/j.1365-313X.2011.04779.x

Yu, F., and De Luca, D. (2013). ATP-binding cassette transporter controls leaf surface secretion of anticancer drug components in *Catharanthus roseus*. *Proc. Natl. Acad. Sci. U.S.A.* 110, 15830–15835. doi: 10.1073/pnas.1307504110

Yu, P. J., Jin, H., Zhang, J. Y., Wang, G. F., Li, J. R., Zhu, Z. G., et al. (2012). Pyranocoumarins isolated from *Peucedanum praeruptorum* Dunn suppress lipopolysaccharide-induced inflammatory response in murine macrophages through inhibition of NF-κB and STAT3 activation. *Inflammation* 35, 967–977. doi: 10.1007/s10753-011-9400-y

Yuan, F., Lyu, M. J., Leng, B. Y., Zheng, G. Y., Feng, Z. T., Li, P. H., et al. (2015). Comparative transcriptome analysis of developmental stages of the *Limonium bicolor* leaf generates insights into salt gland differentiation. *Plant Cell Environ.* 38, 1637–1657. doi: 10.1111/pce.12514

Zhang, F., Li, X., Li, Z., Xu, X., Peng, B., Qin, X., et al. (2014). UPLC/Q-TOF MS-Based Metabolomics and qRT-PCR in Enzyme Gene Screening with Key Role in Triterpenoid Saponin Biosynthesis of *Polygala tenuifolia*. *PLoS ONE* 9:e105765. doi: 10.1371/journal.pone.0105765

Zhao, J., Davis, L. C., and Verpoorte, R. (2005). Elicitor signal transduction leading to production of plant secondary metabolites. *Biotechnol. Adv.* 23, 283–333. doi: 10.1016/j.biotechadv.2005.01.003

Zhao, J., and Dixon, R. (2010). The 'ins' and 'outs' of flavonoid transport Trends. *Plant Sci.* 15, 72–80. doi: 10.1016/j.tplants.2009.11.006

Zhou, J., Wang, W., Liu, M., and Liu, Z. (2014). Molecular authentication of the traditional medicinal plant *Peucedanum praeruptorum* and its substitutes and adulterants by DNA - barcoding technique. *Pharmacogn. Mag.* 10, 385–390. doi: 10.4103/0973-1296.141754