

Deep sequencing and transcriptome analyses to identify genes involved in iridoid biosynthesis in the medicinal plant *Valeriana jatamansi* Jones

Chenshu WANG, Shuang ZHAO*

Yunnan University of Chinese Medicine, College of Pharmaceutical Sciences, NO.1076, Yuhua Road, Kunming 650500, China; 107376820@qq.com; zbaoshuang0871@hotmail.com (*corresponding author)

Abstract

Valeriana jatamansi Jones is utilized for medicinal purposes in China, and is also an important substitute for European *Valeriana officinalis*. The major active principles are generally called valepotriates, which belong to iridoids compounds. To better understand the iridoid biosynthesis pathway in *V. jatamansi*, we generated transcriptome sequences from the leaf and root tissues, and performed de novo sequence assembly, a total of 183,524,060 transcripts and 61,876 unigenes for *V. jatamansi* were obtained from 13.28 Gb clean reads. 56,641 unigenes were annotated by public databases, while 5,235 unigenes remained unannotated. Different unigenes in *V. jatamansi* were identified by MISA analysis, and 5,195 unigenes containing Simple Sequence Repeat (SSR) were identified. When examining the annotation of transcriptome contigs against the KEGG database, we identified 24 unigenes that could be classified into 24 enzyme categories associated with three metabolic pathways leading to iridoid biosynthesis, 6 genes of MVA pathways, 9 genes of MEP pathways and 9 genes of iridoids pathways. We selected 9 genes encoding key enzymes in the iridoid pathway of *V. jatamansi* to examine their organ specificity of expression using quantitative real-time PCR (qPCR). In conclusion, we generated a comprehensive transcriptome assembly representing the gene space in *V. jatamansi*, and the genomic dataset and analyses presented here lay the foundation for further research on this important medicinal plant.

Keywords: gene expression; iridoid biosynthesis; transcriptome; *Valeriana jatamansi*

Introduction

Valeriana jatamansi Jones, belonging to the family Valerianaceae, is utilized for medicinal purposes in China, and Valerianae Jatamansi Rhizoma et Radix is indexed in Chinese Pharmacopeia (Part 1) in the 1977, 2010, and 2015 editions as a traditional Chinese medicine for gastrointestinal diseases and anxiety (He *et al.*, 2018) and *V. jatamansi*, as an important substitute for European *V. officinalis*, has also been traditionally employed for treatment of various diseases including sleep disorder, obesity, nervous disorders, epilepsy, insanity, snake poisoning, eye trouble, and skin diseases (Xu *et al.*, 2011). The major active principles of *V. jatamansi* are generally held to be a number of epoxy iridoid esters called valepotriates, the principal component of which is valtrate (Lin *et al.*, 2010) and Valtrate illustrates the structural relationship to loganin, though these

compounds contain additional ester functions, frequently isovaleryl, and the hemiacetal is fixed as an ester, rather than as a glycoside (Dewick, 2009). Besides valepotriates and other iridoids, sesquiterpenoids and essential oils have also been identified from this species (Bhatt *et al.*, 2012; Li *et al.*, 2013).

Iridoids take the classical cytosolic mevalonate (MVA) / plastidial methylerythritol phosphate (MEP) pathways of terpene biosynthesis, and they originate from geraniol, which is biosynthesized from geranyl diphosphate (GPP), a condensation product of Isopentenyl Diphosphate (IPP) and dimethylallyl diphosphate (DMAPP) (Miettinen *et al.*, 2014; Padhan *et al.*, 2015). The iridoid pathway is thought to comprise several enzymes catalysing successive oxidation, reduction, glycosylation and methylation reactions (Salim *et al.*, 2014) and loganin is a key intermediate in the biosynthesis of many other iridoid structures (Dewick, 2009).

The biosynthesis of iridoids has been studied in *Catharanthus roseus*, and many key enzymatic genes have been exploited (Miettinen *et al.*, 2014; Kumar *et al.*, 2015). However, little research has been done on the related genes in the plants of the genus *Valeriana* (Valerianaceae), especially in *V. jatamansi*. To harness the biotechnological potential of iridoids from *V. jatamansi*, the enzymes involved in the biosynthetic pathway must be elucidated.

With the rapid development of next-generation sequencing technologies, RNA Sequencing (RNA-Seq) has quickly become a powerful approach for studying transcriptional regulation systematically (Wang *et al.*, 2011). In this study, RNA-seq technology was used to identify genes involved in iridoid biosynthetic pathway of *V. jatamansi*. Our study is not only useful to analyse the iridoid biosynthetic molecular mechanism of *V. jatamansi* but also helpful for its resource development.

Materials and Methods

Plant material

Fresh roots and leaves were collected from *V. jatamansi* plants in their wild habitat located in Shizhong, County of Yunnan Province, China. The samples were immediately frozen in liquid nitrogen and stored at -80 °C prior to RNA extraction.

RNA isolation and transcriptome sequencing

Total RNA from the two tissues from *V. jatamansi* plants was extracted separately. The quality of the RNA samples was evaluated using a Bioanalyzer (Agilent Technologies, USA). To reduce the cost of sequencing, three RNA samples from the leaves of three plants were mixed equally, and three RNA samples from the roots of three plants were mixed equally. RNA-seq libraries were constructed with insert sizes of 200 to 500 bp, and were sequenced using the Illumina HiSeq 2500 platform with paired-end (PE) reads of 100 bp.

De novo assembly and sequence processing

The raw data of the two RNA-seq reads were quality trimmed and then used for sequence assembly using CLC Genomics Workbench (v8.0.2) with the default parameters. For the removal of the redundant sequences, the CD-HIT (Fu *et al.*, 2012) package with an identity parameter of 95% was employed to generate a set of non-redundant contig sequence files. Transcripts shorter than 300 bp were not included in the final *V. jatamansi* transcriptome assembly dataset for further analyses.

Gene function annotation

The consensus contig sequence files were annotated against the Nr (NCBI non-redundant protein sequences), Pfam (Protein family), and Swissprot (A manually annotated and reviewed protein sequence database) databases using local BLASTX program with an E-value threshold of 1e-10. GO (Gene Ontology

Consortium) analysis was performed using Blast2GO with the same E-value cut off. To identify the transcription factors in the *V. jatamansi* transcript dataset, we also applied the BLASTX search program against the plant transcription factor database (Perez-Rodriguez *et al.*, 2010). Annotated sequences were further characterized based on KOG (EuKaryotic Orthologous Groups) / COG (Cluster of Orthologous Groups of proteins) classifications. Metabolic pathway mapping of the transcripts in the Nr Unigene set was performed using the KEGG (Kyoto Encyclopedia of Genes and Genomes) Automatic Annotation Server.

Identification of SSRs

The MicroSatellite identification tool (MISA, <http://pgrc.inpk-gatersleben.de/misa/>) has been used to identify SSRs (Zhang *et al.*, 2015). The parameters were adjusted to identify perfect dinucleotide, trinucleotide, tetranucleotide, pentanucleotide, and hexanucleotide motifs with a minimum of 6, 5, 5, 4, and 4 repetitions, respectively. Primer pairs were designed using the Primer 5 software and selected according to the following criteria: (1) primers with SSRs were eliminated; (2) primers aligned with unigene sequences were allowed three mismatches at the 50 site and one mismatch at the 30 site; (3) primers that aligned to more than one unigene were eliminated; and (4) SSRs were identified using the *ssr_finder* (<http://www.fresnostate.edu/ssrfinder/>). We kept the products whose results from *ssr_finder* and MISA were the same.

qPCR analysis

To verify the quality of the sequences assembled in this study, 9 unigenes related to the iridoid biosynthesis were validated to examine their expression patterns across different tissues (Petiole, Leaf, Rhizome, Root) by quantitative real-time polymerase chain reaction (qPCR). qPCR was performed using the StepOne™ Real-Time PCR Systems (ABI, USA) with 2SYBR Green mix (Fermentas, DE, USA) according to the manufacturer's instructions. First-strand cDNA was synthesized from 1 µg of total RNA with reverse transcriptase (Takara, Japan) and oligo (dT) primer, and the resulting products were used as templates for qPCR. The specific primers used for qPCR are listed in Table 1.

Table 1. Sequences of primers for qPCR

| Gene | Primer | Primer sequence (5'-3') |
|-------|---------|--------------------------|
| GES | GES-F | CGTTTCCATGTCGTCTTTACC |
| | GES-R | CTCCATTATTCTGCTGCTTTCA |
| G10H | G10H-F | GCAGCACAGTAGATTGGACGAT |
| | G10H-R | ATGGCGCAGCAGGGTATAA |
| 10HGO | 10HGO-F | CATACTTCCCACACCGACT |
| | 10HGO-R | GGTCCGTTGATTGGATTGTT |
| IS | IS-F | AATCCCACGAACCTTTAGACC |
| | IS-R | CAGTTTGATGAACATTGCGAGT |
| 7DLS | 7DLS-F | TACCGAGAATGACCAACCAAG |
| | 7DLS-R | AGAACCAAACGGAAGTAACGAG |
| 7DLGT | 7DLGT-F | TGACCCATAAGAACCGAACC |
| | 7DLGT-R | AATCCGTGATCTACGTCAGCTT |
| DL7H | DL7H-F | ATCATTATGTGGGTGGGAAGA |
| | DL7H-R | TATGCCGCAGAGCAAGTGTC |
| LAMT | LAMT-F | AAAAGCAGTCTTGATGGGTTG |
| | LAMT-R | TAAGTGAAGAGAAAGTGGACCTGT |
| SLS | SLS-F | CAATTGAGCTGGACGCACA |
| | SLS-R | GATACCCATTACCAAAGCAGTTAC |
| GAPDH | GAPDH-F | CTTCTGAGTGGCAGTAATGGAG |
| | GAPDH-R | CCGATGTTTGTGTGGGTG |

The qPCR thermal cycling condition for all reactions was 95 °C for 3 min, followed by 45 cycles at 95 °C for 7 sec and 57 °C for 10 sec. All reactions were conducted in triplicates, and the results were expressed as relative expression levels to the GAPDH gene. The Ct values obtained were used as the original data to calculate the relative expression level of different genes to GAPDH gene by the $2^{-\Delta\Delta C_t}$ method (Huang *et al.*, 2015; Ye *et al.*, 2016). Each sample was analysed in triplicate.

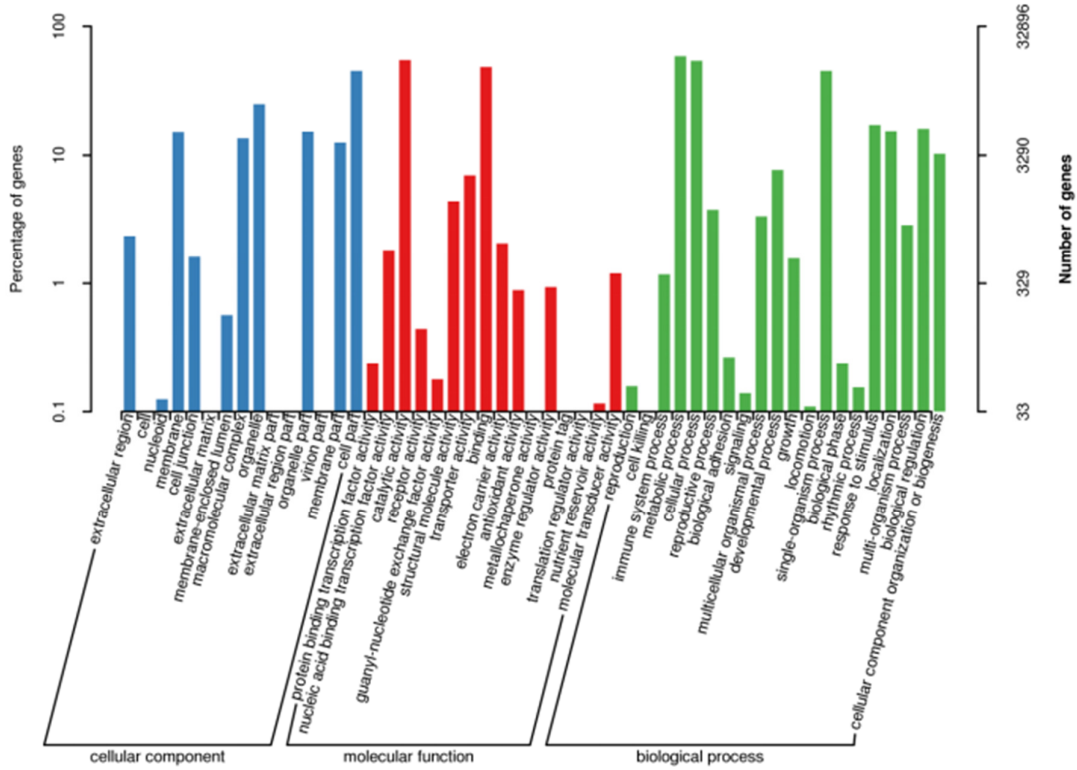
Results and Discussion

Illumina sequencing, read assembly and functional annotation

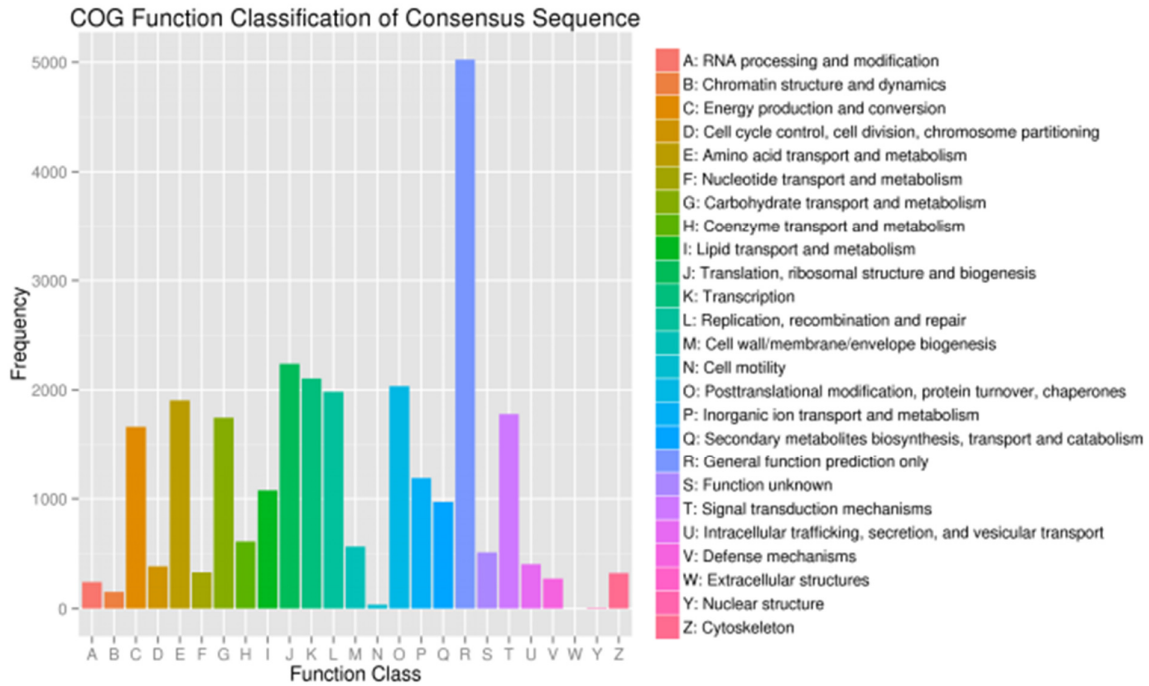
To generate a transcriptome database for *V. jatamansi*, two RNA-seq libraries were constructed from the root and leaf tissues collected from *V. jatamansi* plants. The total clean data of sequencing was approximately 13.28 Gb under the quality control. Moreover, raw data were deposited in NCBI with accession number SRP10331536. The final assembly dataset contained 183,524,060 transcripts with an N50 of 1,128 bp, and 61,876 unigenes with an N50 of 1,099 bp. There were 61,876 unigenes (78.50%) in the size range of 300-1,000 bp, 10,973 (13.92%) at 1,001–2,000 bp, and 5,976 (7.58%) > 2,000 bp.

All the unigenes were annotated by public databases, including COG, GO, KEGG, KOG, Pfam, Swissprot and Nr; 56,641 unigenes were annotated, while 5,235 unigenes remained unannotated. A total of 55,118 unigenes matched sequences in the Nr databases. 32,896 unigenes were annotated by GO database and classified into the biological process, cellular component, and molecular function categories (Figure 1A). Within the cellular component domain, the three most enriched categories were “cell part” (14,868, 45.20%), “organelle” (8,181, 24.87%) and “organelle part” (5,032, 15.30%). In the molecular function domain, the three most matched categories were “binding” (15,924, 48.41%), “catalytic activity” (18,076, 54.95%), and “transporter activity” (2,304, 7.00%). In the biological process domain, the three most common categories were “cellular process” (17,771, 54.02%), “metabolic process” (19,325, 58.75%), and “single-organism process” (14,895, 45.28%). A total of 27,119 unigenes were annotated by COG database and grouped into 25 functional classifications (Figure 1B). The most frequently identified classes were “general function” (5,029, 18.54%), followed by translation, ribosomal structure and biogenesis (2,151, 7.93%), transcription (2,049, 75.56%), post-translational modification, protein turnover, and chaperones (2,026, 74.71%), Replication, recombination and repair (1,991, 7.34%), Amino acid transport and metabolism (1,812, 6.68%), carbohydrate transport and metabolism (1,685, 6.21%), the nuclear structure (12, 0.04%) and extracellular structure (0, 0%) categories showed the fewest corresponding genes.

The high percentage of unigenes involved in the function of catalytic activity indicated the variety of the secondary metabolites produced by *V. jatamansi*. To investigate the biological function of unigenes further, a total of 21,452 unigenes were assigned to the metabolic pathways described in the KEGG database, including metabolic pathways, biosynthesis of secondary metabolites, ubiquitin-mediated proteolysis, mitogen-activated protein kinase (MAPK) signalling pathway, and regulation of autophagy RNA transport.



(A)



(B)

Figure 1. GO function and COG function classification of the *V. jatamansi* transcriptome. (A) GO functional classification of the *V. jatamansi* transcriptome. (B) COG functional classification of the *V. jatamansi* transcriptome

Development and characterization of cDNA-Derived SSR markers

Different unigenes in *V. jatamansi* were identified by MISA analysis, and 5,195 unigenes containing SSR were identified, among which, 924 sequences contained more than one SSR. Moreover, 295 SSRs were present in the compound formation. The distribution and frequency of the mono-, di-, tri-, tetra-, and hexa-nucleotide repeats were analysed. The most abundant repeat motifs were mononucleotides (2,023, 38.94%), followed by dinucleotides (1,402, 26.99%), trinucleotides (1,362, 26.22%), tetra-nucleotides (104, 2.00%), penta- (17, 0.33%) and hexa-nucleotide (11, 0.23%) (Figure 2). Among the mononucleotide's repeats. The most frequent was A/T (1,992, 38.34%), followed by C/G (31, 0.60%). Among the di-nucleotide repeats, AT/TA was present at the highest (821, 15.80%) frequency, followed by AG/CT (323, 6.22%) and AC/GT (257, 4.95%). For tri-nucleotide repeats, AAG/CTT showed the highest occurrence (410, 7.89%), followed by AAT/ATT (134, 2.58%) and ACC/GGT (108, 2.08%). The availability of this large number of SSRs can greatly enhance large-scale genotyping studies for various applications in *V. jatamansi*, such as genetic diversity assessment and association mapping for important traits.

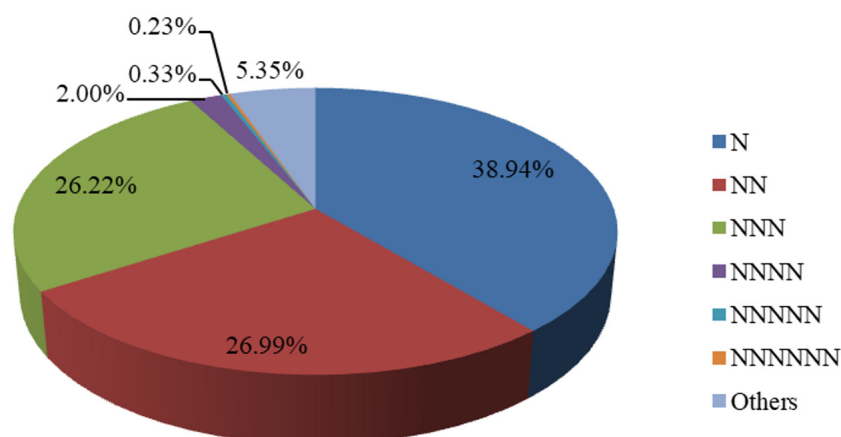


Figure 2. Distribution of different classes of SSR in the *V. jatamansi* transcriptome

Putative genes involved in the iridoids biosynthesis pathway

Iridoids take the classical MVA/MEP route, and the prenyl-transfer of IPP on DMAPP catalysed by geranyl diphosphate synthase (GPPS) forms geranyl diphosphate (GPP) (Rai *et al.*, 2013), which is followed by the formation of the monoterpene geraniol by geraniol synthase (GES). All the steps of the *C. roseus* secoiridoid pathway have been characterized successfully, it starts with geraniol and is thought to comprise 9 enzymes catalysing successive oxidation, reduction, glycosylation and methylation reactions (Salim *et al.*, 2014).

When examining the annotation of transcriptome contigs against the KEGG database, we identified 24 transcripts that could be classified into 24 enzyme categories associated with three metabolic pathways leading to iridoid biosynthesis. Table 2 shows that there are 6 genes of the MVA pathways, such as ACCT, HMGS, HMGR, MVK, PMK, MVD; there are 9 genes of the MEP pathways, such as DXS, DXR, ISPD, ISPE, ISPF, GCPE, ISPH, IDI, GPPS; there are 9 of the iridoid pathway, such as G10H, 10HGO, IS, 7DLS, 7DLGT, DL7H, LAMT, SLS. The analysis of the metabolic pathways with transcriptomic data help us identify the coding sequences of the genes involved in specific processes of the pathway in targeted plant species and can aid in uncovering their family members and expression patterns in different organs.

Table 2. Summary of transcripts in *V. jatamansi* encoding enzymes involved in the iridoid pathway

| Pathway | Gene | Gene name | E.C. number | KO | Gene ID |
|---------|----------------------|--|-------------|----------------------|-----------------------|
| MVA | ACCT | acetyl-CoA C-acetyltransferase | 2.3.1.9 | K00626 | c21891.graph_c0_seq1 |
| | HMGs | hydroxymethylglutaryl-CoA synthase | 2.3.3.10 | K01641 | c41531.graph_c0_seq1 |
| | HMGR | hydroxymethylglutaryl-CoA reductase | 1.1.1.34 | K00021 | c43230.graph_c0_seq1 |
| | MVK | mevalonate kinase | 2.7.1.36 | K00869 | c48658.graph_c0_seq12 |
| | PMK | phosphomevalonate kinase | 2.7.4.2 | K00938 | c38258.graph_c0_seq1 |
| | MVD | diphosphomevalonate decarboxylase | 4.1.1.33 | K01597 | c45193.graph_c0_seq1 |
| MEP | DXS | 1-deoxy-D-xylulose-5-phosphate synthase | 2.2.1.7 | K01662 | c44604.graph_c0_seq1 |
| | DXR | 1-deoxy-D-xylulose-5-phosphate reductoisomerase | 1.1.1.267 | K00099 | c46256.graph_c0_seq13 |
| | ISPD | 2-C-methyl-D-erythritol 4-phosphate cytidyltransferase | 2.7.7.60 | K00991 | c44696.graph_c0_seq1 |
| | ISPE | 4-diphosphocytidyl-2-C-methyl-D-erythritol kinase | 2.7.1.148 | K00919 | c43601.graph_c0_seq1 |
| | ISPF | 2-C-methyl-D-erythritol 2,4-cyclodiphosphate synthase | 4.6.1.12 | K01770 | c5341.graph_c0_seq1 |
| | GCPE | (E)-4-hydroxy-3-methylbut-2-enyl-diphosphate synthase | 1.17.7.1 | K03526 | c49626.graph_c0_seq2 |
| | ISPH | 4-hydroxy-3-methylbut-2-en-1-yl diphosphate reductase | 1.17.7.4 | K03527 | c48041.graph_c0_seq1 |
| | IDI | isopentenyl-diphosphate delta-isomerase | 5.3.3.2 | K01823 | c77009.graph_c0_seq1 |
| | GPPS | geranyl diphosphate synthase | 2.5.1.1 | K14066 | c51297.graph_c0_seq1 |
| Iridoid | GES | geranyl diphosphate diphosphatase | 3.1.7.11 | K20979 | c48823.graph_c0_seq1 |
| | G10H | geraniol 10-hydroxylase | 1.14.13.152 | K15099 | c51475.graph_c1_seq4 |
| | 10HGO | 10-hydroxygeraniol oxidoreductase | 1.1.1.324 | | c41113.graph_c0_seq1 |
| | IS | iridoid synthase | 1.3.1.99 | K20144 | c52233.graph_c0_seq1 |
| | 7DLS | 7-deoxyloganetic acid synthase | | | c23708.graph_c0_seq1 |
| | 7DLGT | 7-deoxyloganetic acid glucosyl transferase | 2.4.1.323 | | c43645.graph_c0_seq1 |
| | DL7H | 7-deoxyloganic acid hydroxylase | | | c47117.graph_c0_seq1 |
| | LAMT | loganic acid O-methyltransferase | 2.1.1.50 | | c35974.graph_c0_seq3 |
| SLS | secologanin synthase | 1.3.3.9 | K13400 | c45910.graph_c0_seq1 | |

qPCR validation of different expression patterns

We selected 9 transcripts encoding key enzymes in the iridoid pathway of *V. jatamansi* to examine their organ specificity of expression using qPCR (Figure 3). The first key step in the biosynthesis is the generation of geraniol from geranyl diphosphate (GPP) in the iridoid pathway. Analyses of GES transcript levels by qPCR showed that the relative expression of GES was highest in the leaves of *V. jatamansi*. And the expression of CrGES is also highest in the leaf of *Madagascar periwinkle* (Kumar *et al.*, 2015).

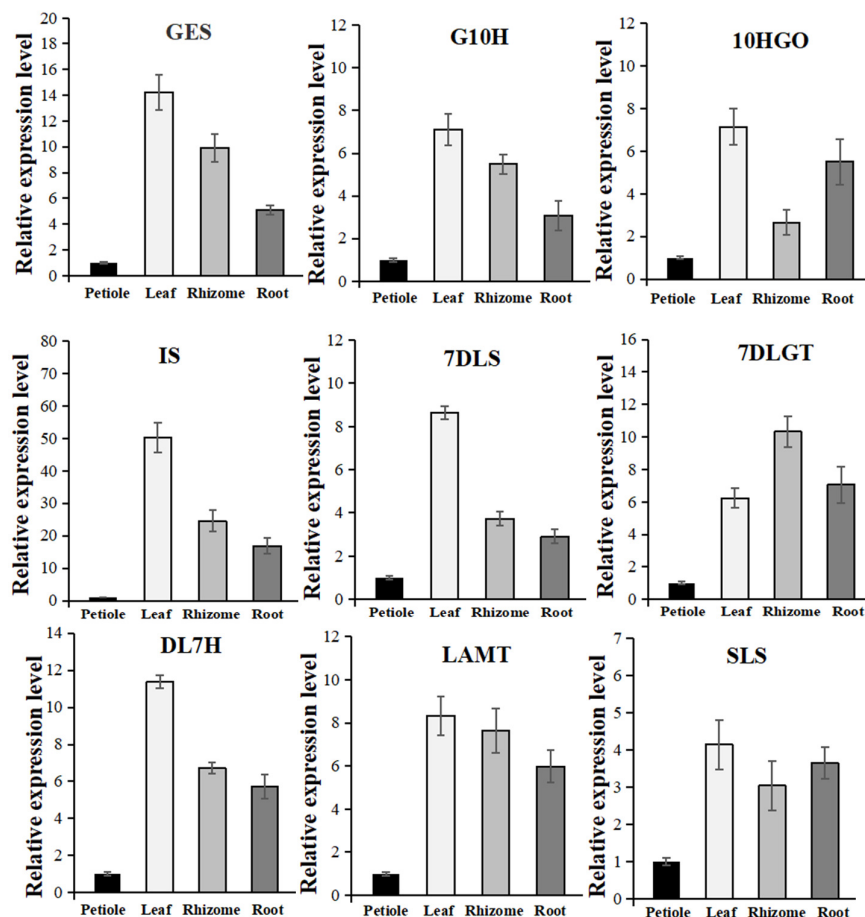


Figure 3. The expression pattern of 9 genes in iridoid biosynthesis pathway across different tissues (petiole, leaf, rhizome, root)

The enzyme geraniol 10-hydroxylase (G10H) is a cytochrome P450 monooxygenase, which hydroxylates the monoterpene geraniol at the C-10. G10H belongs to the CYP76B subfamily, and was designated CYP76B6. And it has G10H activity when expressed in *C. roseus* and yeast cells (Collu *et al.*, 2001). Analyses of G10H transcript levels by qPCR showed that the relative expression of G10H was highest in the leaves of *V. jatamansi*, and the higher expression of G10H in leaf tissue than root tissue has also been reported for *Gentiana rigescens* (Zhang *et al.*, 2015) and *G. macrophylla* (Hua *et al.*, 2014).

10-hydroxygeraniol oxidoreductase (10 HGO) is active with the substrates 10-OH-geraniol, 10-OH-geranial and 10-oxogeraniol in the presence of NADt, yielding mixtures of the three compounds and 10-oxogeranial in varying relative amounts depending on the combination and the incubation time (Miettinen *et al.*, 2014). Analyses of 10 HGO transcript levels by qPCR showed that the relative expression of 10 HGO was highest in the leaves of *V. jatamansi*.

Iridoid synthase (IS) first catalyses a reaction that parallels the 1,4-reduction of progesterone by P5 β R (Progesterone-5 β -reductase). The analogous 1,4-reduction of 10-oxogeranial would yield an enol species that is poised for an inverse electron demand oxo-Diels-Alder reaction, a rare occurrence in enzymology (Geu-Flores *et al.*, 2012) so IS is responsible for the reductive cyclization step. Analyses of IS transcript levels by qPCR showed that the relative expression of IS was highest in the leaves of *V. jatamansi*, and the higher expression of IS in leaf tissue than root tissue has also been reported for *Swertia mussotii* (Liu *et al.*, 2017).

CYP76A26 was responsible for the oxidation steps leading to the formation of 7-deoxyloganetic acid, and it was consequently named 7-deoxyloganetic acid synthase (7DLS). And 7DLS can catalyse the 3-step oxidation of iridodial-nepetalactol to form 7-deoxyloganetic acid in *C. roseus* iridoid biosynthesis (Salim *et al.*, 2014). Because 7DLS converted both iridodial and iridotrial into 7-deoxyloganetic acid and was also named iridoid oxidase (IO) (Miettinen *et al.*, 2014). Analyses of 7DLS transcript levels by qPCR showed that the relative expression of 7DLS was highest in the leaves of *V. jatamansi*, and the higher expression of 7DLS in leaf tissue than root tissue has also been reported for *C. roseus* (Salim *et al.*, 2014).

7-deoxyloganetic acid glucosyl transferase (7DLGT) could catalyse the glucosylation of 7-deoxyloganetic acid to form 7-deoxyloganic acid using UDP-glucose as the sugar donor (Miettinen *et al.*, 2014). The qPCR results showed that the expression pattern of 7DLGT was quite different from the other transcripts in the pathway, and 7DLGT was the only gene with its highest expression in the rhizome tissue; similar to what was reported for *S. chirayita* (Padhan *et al.*, 2015) and *S. mussotii* (Liu *et al.*, 2017).

CYP72A224 catalyses the conversion of 7-deoxyloganic acid into loganic acid in yeast microsomes. The aglycone derivative of 7-deoxyloganic acid was not a substrate for CYP72A224, confirming that glycosylation precedes hydroxylation of the cyclopentane ring. CYP72A224 was thus named 7-deoxyloganic acid hydroxylase (DL7H) (Miettinen *et al.*, 2014). Analyses of DL7H transcript levels by qPCR showed that the relative expression of DL7H was also highest in the leaves of *V. jatamansi*.

Loganic acid O-methyltransferase (LAMT) can catalyse loganic acid to form loganin. The recombinant LAMT showed striking specificity for loganic acid since it is not active with other similar iridoids, including deoxy loganic, dehydrologanic, epiloganic, loganetic acid, or with the reaction product, loganin (Murata *et al.*, 2008). Analyses of LAMT transcript levels by qPCR showed that the relative expression of LAMT was also highest in the leaves of *V. jatamansi*.

Secologanin synthase (SLS) is also a cytochrome P450 monooxygenase, which can catalyse an unusual ring-opening reaction using loganin in the biosynthesis of secologanin in the secoiridoid pathway. And CYP72A1 from *Madagascar periwinkle* has secologanin synthase activity (Irmeler *et al.*, 2000). The qPCR indicated that the highest expression of SLS was in the leaf and lowest in the petiole of *V. jatamansi*.

The 4 genes G10H, 7DLS, DL7H and SLS were P450 monooxygenase, during 9 transcripts encoding key enzymes in the iridoid pathway. Although G10H and 7DLS belong to the same P450 family CYP76, interestingly *Catharanthus* G10H (CYP76B6) sequence had lower amino acid sequence identities with *Catharanthus* 7DLS (CYP76A26) sequence (39%). The substrate specificity of 7DLS (CYP76A26) for nepetalactol shows the distinct properties of this enzyme in catalysing a successive three step oxidation compared to G10H (CYP76B6) that accepts a broader range of substrates including geraniol, its oxidized derivative, nerol, and flavonoids (Hofer *et al.*, 2013; Sung *et al.*, 2011).

DL7H (CYP72A224) and SLS (CYP72A1) belong to the same P450 subfamily CYP72A, and both of these P450s use glycosylated substrates and have similar regiospecificities, suggesting that SLS evolved from DL7H thus extending the iridoid pathway to the secoiridoids. The overlap in their substrate specificities has been investigated, as SLS showed no DL7H activity and vice versa, proposed evolutionary process appears to have yielded specific and mutually exclusive catalytic activities for each enzyme (Miettinen *et al.*, 2014).

Conclusions

RNA-Seq technology has been a reliable tool to elucidate new genes and their involvement in biochemical pathways in non-model plants lacking whole-genome sequencing data. In this study, the transcriptome of *V. jatamansi* was first sequenced by an Illumina HiSeq 2500 platform, and the deep transcriptome sequence data was performed to generate comprehensive information of the gene space in this medical plant and genes associated with iridoid biosynthesis pathways. A total of 21,452 unigenes were assigned

to the metabolic pathways described in the KEGG database, and 5,195 unigenes containing SSR were identified. Our *V. jatamansi* transcriptome results revealed that 24 unigenes, which encoded key enzymes in the iridoid biosynthesis, including 6 genes of the MVA pathways, 9 genes of the MEP pathways, and 9 genes of the iridoid pathway. qPCR analysis was employed to validate the expression levels in different organs of 9 genes in the iridoid pathway, including G10H, 10HGO, IS, 7DLS, 7DLGT, DL7H, LAMT, SLS. qPCR analysis results revealed that 7DLGT gene showed the highest expressed level in the rhizome tissue, while all the other 8 genes showed the highest expressed level in the leaf tissue. It was supposed that 7DLGT gene may be the rate-limiting enzyme in the iridoid biosynthesis pathway, and this gene could be a potential target for a biotechnological approach to engineer an improved the pathway for iridoid compounds yield in *V. jatamansi* in the future. These results support the hypothesis that iridoid compounds may be synthesized in leaves and then be transported to rhizomes for storage. Clearly, further corroboration through molecular, biochemical, enzymatic, and physiological studies are needed to gain a better understanding of the underlying regulatory mechanisms involved. Nevertheless, the transcriptome data and analyses represent the first genomics resource for *V. jatamansi* and lay the foundation for future research towards the improvement of the important medical plant through genetics, genomics, and biotechnological approaches.

Data availability

The raw Illumina data generated in this study were deposited in the NCBI Sequence Read Archive (SRA) under the accession number SRP10331536.

Acknowledgements

This work was supported by the National Natural Science Foundation of China, grant number: 81160502, and the Application and Basis Research Program of Yunnan Province-Chinese Medicine Joint Project China, grant number: 2017FF116(-008).

Conflict of Interests

The authors declare that there are no conflicts of interest related to this article.

References

- Bhatt ID, Dauthal P, Rawat S, Gaira KS, Jugran A, Rawal RS, Dhar U (2012). Characterization of essential oil composition, phenolic content, and antioxidant properties in wild and planted individuals of *Valeriana jatamansi* Jones. *Scientia Horticulturae* 136:61-68.
- Collu G, Unver N, Peltenburg-Looman AM, Heijden R, Verpoorte R, Memelink J (2001). Geraniol 10-hydroxylase, a cytochrome P450 enzyme involved in terpenoid indole alkaloid biosynthesis. *FEBS Letters* 508(2):215-220.
- Dewick PM (2009). *Medicinal natural products: a biosynthetic approach* (3rd edition). John Wiley & Sons, Ltd., West Sussex.
- Fu L, Niu B, Zhu Z, Wu S, Li W (2012). CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics* 28: 3150-2.
- Geu-Flores F, Sherden NH, Courdavault V, Burlat V, Glenn WS, Wu C, ... O'Connor SE (2012). An alternative route to cyclic terpenes by reductive cyclization in iridoid biosynthesis. *Nature* 492(7427):138-142.

- He X, Wang S, Shi J, Sun Z, Lei Z, Yin Z, ... Xie H (2018). Genotypic and environmental effects on the volatile chemotype of *Valeriana jatamansi* Jones. *Frontiers in Plant Science* 9:1003-1012.
- Huang L, Li G, Mo Z, Xiao P, Li J, Huang J (2015). *De novo* assembly of the Japanese flounder (*Paralichthys olivaceus*) spleen transcriptome to identify putative genes involved in immunity. *PLoS ONE* 10(2): e0117642.
- Hua W, Zheng P, He Y, Cui L, Kong W, Wang Z (2014). An insight into the genes involved in secoiridoid biosynthesis in *Gentiana macrophylla* by RNA-seq. *Molecular Biology Reports* 41(7):4817-25.
- Kumar K, Kumar SR, Dwivedi V, Rai A, Shukla AK, Shanker K, Nagegowda DA (2015). Precursor feeding studies and molecular characterization of geraniol synthase establish the limiting role of geraniol in monoterpene indole alkaloid biosynthesis in *Catharanthus roseus* leaves. *Plant Science* 239:56-66.
- Li YD, Wu ZY, Li HM, Li HZ, Li RT (2013). Iridoids from the roots of *Valeriana jatamansi*. *Helvetica Chimica Acta* 96(3):424-430.
- Lin S, Chen T, Liu XH, Shen YH, Li HL, Shan L, ... Wang H (2010). Iridoids and lignans from *Valeriana jatamansi*. *Journal of Natural Products* 73(4):632-638.
- Liu Y, Wang Y, Guo F, Zhan L, Mohr T, Cheng P, ... Gu YQ (2017). Deep sequencing and transcriptome analyses to identify genes involved in secoiridoid biosynthesis in the Tibetan medicinal plant *Swertia mussotii*. *Scientific Reports* 7:43108.
- Miettinen K, Dong L, Navrot N, Schneider T, Burlat V, Pollier J, ... Werck-Reichhart D (2014). The seco-iridoid pathway from *Catharanthus roseus*. *Nature Communications* 5:1-12.
- Rai A, Smita SS, Singh AK, Shanker K, Nagegowda DA (2013). Heteromeric and homomeric geranyl diphosphate synthases from *Catharanthus roseus* their role in monoterpene indole alkaloid biosynthesis. *Molecular Plant* 6(5):1531-1549.
- Salim V, Wiens B, Masada-Atsumi S, Yu F, De Luca V (2014). 7-Deoxyloganetic acid synthase catalyzes a key 3 step oxidation to form 7-deoxyloganetic acid in *Catharanthus roseus* iridoid biosynthesis. *Phytochemistry* 101:23-31.
- Tamura K, Stecher G, Peterson D, Filipski A, Kumar S (2011). MEGA6: Molecular evolutionary genetics analysis version 6.0. *Molecular Biology and Evolution* 30(12):2725-272.
- Wang L, Si Y, Dedow LK, Shao Y, Liu P, Brutnell TP (2011). A low-cost library construction protocol and data analysis pipeline for illumina-based strand-specific multiplex RNA-Seq. *PLoS ONE* 6(10): e26426.
- Xu J, Yang B, Guo Y, Jin DQ, Guo P, Liu C, ... Sun Z (2011). Neuroprotective bakkenolides from the roots of *Valeriana jatamansi*. *Fitoterapia* 82(6):849-853.
- Ye W, Wu H, He X, Wang L, Zhang W, Li H, ... Gao X (2016). Transcriptome sequencing of chemically induced *Aquilaria sinensis* to identify genes related to agarwood formation. *PLoS ONE* 11(5):1-16.
- Zhang X, Allan AC, Li C, Wang Y, Yao Q (2015). *De novo* assembly and characterization of the transcriptome of the Chinese medicinal herb, *Gentiana rigescens*. *International Journal of Molecular Science* 16(5):11550-11573.
- Zhang MF, Jiang LM, Zhang DM, Jia GX (2015). *De novo* transcriptome characterization of *Lilium* 'Sorbonne' and key enzymes related to the flavonoid biosynthesis. *Molecular Genetics and Genomics* 290(1):399-412.



The journal offers free, immediate, and unrestricted access to peer-reviewed research and scholarly work. Users are allowed to read, download, copy, distribute, print, search, or link to the full texts of the articles, or use them for any other lawful purpose, without asking prior permission from the publisher or the author.

License - Articles published in *Notulae Botanicae Horti Agrobotanici Cluj-Napoca* are Open-Access, distributed under the terms and conditions of the Creative Commons Attribution (CC BY 4.0) License.
 © Articles by the authors; UASVM, Cluj-Napoca, Romania. The journal allows the author(s) to hold the copyright/to retain publishing rights without restriction.