Regular Article

Highlighted Paper selected by Editor-in-Chief

Transcriptome Analysis of Nine Tissues to Discover Genes Involved in the Biosynthesis of Active Ingredients in *Sophora flavescent*

Rongchun Han, Hiroki Takahashi, Michimi Nakamura, Somnuk Bunsupa, Naoko Yoshimoto, Hiromi Yamamoto, Hideyuki Suzuki, Daisuke Shibata, Mami Yamazaki, and Kazuki Saito*

a Graduate School of Pharmaceutical Sciences, Chiba University; Chuo-ku, Chiba 260-8675, Japan; b Pharmacy College, Liaoning University of Traditional Chinese Medicine; DD-Port, Dalian 116600, People’s Republic of China; c Medical Mycology Research Center, Chiba University; Chuo-ku, Chiba 260-8673, Japan; d Faculty of Life Sciences, Toyo University; Ora-gun, Gunma 374-0193, Japan; and *Kazusa DNA Research Institute; 2-6-7 Kazusa-kamata, Kisarazu, Chiba 292-0818, Japan.

Received December 8, 2014; accepted March 11, 2015

*Sophora flavescent* AITON (kurara, 苦参) has long been used to treat various diseases. Although several research findings revealed the biosynthetic pathways of its characteristic chemical components as represented by matrine, insufficient analysis of transcriptome data hampered in-depth analysis of the underlying putative genes responsible for the biosynthesis of pharmaceutical chemical components. In this study, more than 200 million fastq format reads were generated by Illumina’s next-generation sequencing approach using nine types of tissue from *S. flavescent*, followed by CLC de novo assembly, ultimately yielding 83325 contigs in total. By mapping the reads back to the contigs, reads per kilobase of the transcript per million mapped reads values were calculated to demonstrate gene expression levels, and overrepresented gene ontology terms were evaluated using Fisher’s exact test. In search of the putative genes relevant to essential metabolic pathways, all 1350 unique enzyme commission numbers were used to map pathways against the Kyoto Encyclopedia of Genes and Genomes. By analyzing expression patterns, we proposed some candidate genes involved in the biosynthesis of isoflavonoids and quinolizidine alkaloids. Adopting RNA-Seq analysis, we obtained substantially credible contigs downstream work. The preferential expression of the gene for putative lysine/ornithine decarboxylase committed in the initial step of matrine biosynthesis in leaves and stems was confirmed in semi-quantitative polymerase chain reaction (PCR) analysis. The findings in this report may serve as a stepping-stone for further research into this promising medicinal plant.

**Key words** *Sophora flavescent*; RNA-Seq analysis; biosynthesis; quinolizidine alkaloid; isoflavonoid

*Sophora flavescent* AITON (kurara, 苦参) has been recorded and used for more than 1800 years.1 As a widely distributed and effective herbal medicine, it severed as a cure for asthma, sores, gastrointestinal hemorrhage, diarrhea, allergy, inflammation in eastern Asian countries.2,3 Main chemical components of *S. flavescent* include flavonoids (1.5%), alkaloids (3.3%), alkyxanthones, quinones, triterpene glycosides, fatty acids as well as essential oils and recently, several clinical studies reported that alkaloids of *S. flavescent* were efficacious in treating various types of solid tumors (including breast, lung, liver and gastrointestinal tract cancers), which drew close attention to this traditional herbal plant.4,5

In 1889, the characteristic compound matrine was isolated from the dry roots of *S. flavescent* by Naga5 and then in 1966, Okuda et al. confirmed the absolute structure of (+)-matrine.6 Due to its notable medicinal efficacy, attempts to synthesize and biosynthesize matrine were conducted.7–9 In 1995, Saito et al. proposed the biosynthetic pathway of the carbon framework of matrine.10 Although the first steps of quinolizidine alkaloids biosynthesis have been elucidated recently in *S. flavescent*,11 more effort needs to be done to puzzle out the practical biosynthetic pathway of matrine.

*S. flavescent* also contain series of flavonoids and isoflavonoids such as kuraridin, kurarinone, isokurarinine, daidzein, maackiain.12 Much effort has been done to elucidate the biosynthetic pathway of flavonoids and many of the related genes are clear now in other species,13 but the ones involved in the biosynthesis of flavonoids in *S. flavescent* have not yet been discovered. In spite of several research findings cracking relevant quinolizidine alkaloids biosynthetic pathway and membrane-bound prenyltransferase,11,14,15 to our best knowledge, analysis using next generation sequencing approach for *S. flavescent* was not found up to date.

More and more genomes of model organisms have been sequenced.16,17 Nevertheless, for those non-model plants, lack of reference genome information jeopardizes studies on the underlying genes which are involved in biological processes related to vital plant physiology and drug development, etc. With this regard, transcriptome sequencing plays an essential role in apprehending the genetic diversity of organisms.18 Furthermore, such approaches help to obtain overall insights into whole gene sets associated to the protein diversity.19 Armed with well-established approaches such as Short Oligonucleotide Analysis Package (SOAP de novo),20 Assembly by Short Oligonucleotide (AbySS),21 Trinity,22 transcriptome profiling accelerates its pace in processing tremendous amount of data generated from large-scale sequencing projects. Despite the intensive challenges including library construction, reducing

This paper is dedicated to Professor Emeritus Isamu Murakoshi on the occasion of his 85th birthday in recognition of his great contribution to the study of quinolizidine alkaloids.

* To whom correspondence should be addressed. e-mail: ksaito@faculty.chiba-u.jp

© 2015 The Pharmaceutical Society of Japan
errors in image analysis and removal of low-quality reads, mass spectrometry (MS) offers a more precise measurement of levels of transcripts and their isoforms than other methods.23)

In this study, 203598590 fastq format reads from 9 tissues of *S. flavescens* were generated by Illumina's next-generation sequencing approach. CLC Genomics Workbench (CLC Bio, Denmark) was subsequently applied to conduct de novo assembly. Based on the findings provided by Gene Ontology and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway mapping, the candidate genes that may be involved in the biosynthesis of key chemical components were identified. By studying expression pattern of genes related to the biosynthesis of quinolizidine alkaloids, we propose some promising contigs for future research.

**MATERIALS AND METHODS**

**Plant Materials and Total RNA Extraction** Except for leaf sample collected in May 2012, all the rest fresh tissues and organs were obtained from healthy *S. flavescens* plants growing in Chiba, Japan in June 2013. Callus tissue whose origin and subculturing were described previously,24) and other eight parts of the plant were sampled, including leaf, flower, stem, young bud, mature bud, bud right before blossoming (BBB), pedicel while bud stage (PBS), pedicel while blossom (PWB) (Supplementary 1). The materials were dipped into RNA stabilization solution (RNAlater, Life Technologies, U.S.A.) immediately after removal from the field. Then RNAlater solution was gently wiped with a Kimwipe and the remaining sample was frozen by liquid nitrogen and subsequently powdered by Multi Beads Shocker (Yasui Kikai, Japan). By using TRIzol Reagent (Invitrogen, U.S.A.), total RNA was extracted from powdered tissues of *S. flavescens*, respectively. The obtained RNA was then cleaned up by RNeasy Mini Kit (Qiagen, U.S.A.).

For semi-quantitative reverse transcription-polymerase chain reaction (RT-PCR) analysis, in November 2014, the tissues of root, stem and leaf of kurara were sampled in the same location, where the other tissues were collected for deep-transcriptome sequencing, and subjected to the extraction of total RNA as described above.

**cDNA Library Preparation and Sequencing** TruSeq RNA Sample Prep Kit v2 (Illumina, U.S.A.) was applied according to the manufacturer’s recommendation. Once the mRNA in total RNA had been polyA-selected and fragmented, double-stranded cDNA was prepared for cDNA library construction. With the creation of blunt-end fragments and the indexed adaptor ligation, the samples were hybridized to flow cells. Cluster amplification was performed using the cBot Cluster Generation System (Illumina, U.S.A.) and then sequenced by Illumina’s next-generation sequencing instrument. CLC Workbench for de Novo Assembly Prior to assembly, the original fastq format data of *S. flavescens* were subjected to CLC trimming process in order to eliminate reads with poor quality and as a result, clean reads were obtained. The CLC approach (version 6.5) was then used to process the clean reads and all contigs over 300 bp were taken into consideration for downstream work. Since the assembly process may result in duplicate contigs due to sequencing error, CD-HIT-EST was utilized with representative sequences lacking technical duplicates, the non-parametric approach for the identification of differentially expressed (DE) genes, NOISEq-sim,31) was adopted to analyze 36 independent pairwise sample comparisons.

**Annotation Pipeline and Data Mining** After applying CD-HIT-EST, All 83325 de novo contigs were used as query sequences for the basic local alignment search tool (BLASTx) sequence similarity search against the non-redundant (NR) protein database at NCBI and the Universal Protein resource (UniProt) at UniProt consortium.25) The e-value threshold was set to 1e−10; the upper limit on the number of subject sequences from databases to show alignment was limited to 20. As to the large BLASTx output, only percent identities over 40% and e-values less than 1e−30 were taken into consideration. After eliminating redundancies, all unique gene identifiers in fasta format were uploaded to the UniProt ID mapping website for online data processing (http://www.uniprot.org). By consolidating the returned target list and the UniProtKB accession numbers (ACs) obtained from the above-mentioned BLASTx output against the UniProt database, we applied the redundancy-free ACs to annotation using the same online facilities. Out of the huge number of annotation results, we examined the reviewed findings from UniProtKB/Swiss-Prot as well as UniProtKB/TrEMBL for data mining. The sequences with Gene Ontology (GO) terms at the protein level were classified. Ultimately, 1350 enzyme commission (EC) numbers were applied to map pathways against the KEGG,33) and the enzymes related to flavonoids and isoflavonoids biosynthesis were studied.

**Semi-quantitative RT-PCR for a Putative Lysine/Oryzidine Decarboxylase (L/ODC) Gene** By using the total RNA extracted from the tissues harvested in 2014, cDNA was prepared according to the manufacturer’s instructions, SuperScript VILO kit (Invitrogen, U.S.A.). A 477 bp fragment of *L/ODC* was amplified by PCR using Ex tag DNA polymerase (TaKaRa, Japan) and specific primers (L/ODC-F: 5’-GAC ATT GGT GGC GGT TTC AC-3’, L/ODC-R: 5’-AGT GCT AAA GCC ATT GAA GTT GG-3’). The PCR for *L/ODC* cDNA was performed with an initial denaturation at 94°C for 2 min, then 26, 28 or 30 cycles each at 94°C for 30 s, at 54°C for 30 s, and 72°C for 50 s. For normalization of the different RNA preparations, a 571 bp fragment of *S. flavescens* β-actin was amplified with the following primers: (Act-F: 5’-AACAGGC CAAC AGA GAG AAG ATG AC-3’, Act-R: 5’-ACC CACCA CTA AGC CGA TATTT-3’). The PCR for *β-actin* cDNA was performed with an initial denaturation at 94°C for 2 min, then 22 cycles each at 94°C for 30 s, at 53°C for 30 s, and 72°C for 50 s. The PCR products were separated by electrophoresis

**Reads per Kilobase of the Transcript per Million Mapped Reads (RPKM) Calculation and Expression-Based Analysis** The standard formula for RPKM is as follows: RPKM=(10^9×C)/(N×L), where *C* is the number of reads mapped to the gene’s exons, *N* the total number of mapped reads in the experiment and *L* the total length of the exons in base pairs.20) In order to estimate the expression level of the contigs, with the aid of Burrows–Wheeler Aligner,27) Sequence Alignment/Map tools28) and High Throughput Sequencing (HTSeq),29) we applied the Mortazavi’s approach30) to map all the fastq format reads back to the contigs and calculated the RPKM values. Because the *S. flavescens* samples lacked technical replicates, the non-parametric approach for the identification of differentially expressed (DE) genes, NOISeq-sim,31) was adopted to analyze 36 independent pairwise sample comparisons.

**Semi-quantitative RT-PCR for a Putative Lysine/Oryzidine Decarboxylase (L/ODC) Gene** By using the total RNA extracted from the tissues harvested in 2014, cDNA was prepared according to the manufacturer’s instructions, SuperScript VILO kit (Invitrogen, U.S.A.). A 477 bp fragment of *L/ODC* was amplified by PCR using Ex tag DNA polymerase (TaKaRa, Japan) and specific primers (L/ODC-F: 5’-GAC ATT GGT GGC GGT TTC AC-3’, L/ODC-R: 5’-AGT GCT AAA GCC ATT GAA GTT GG-3’). The PCR for *L/ODC* cDNA was performed with an initial denaturation at 94°C for 2 min, then 26, 28 or 30 cycles each at 94°C for 30 s, at 54°C for 30 s, and 72°C for 50 s. For normalization of the different RNA preparations, a 571 bp fragment of *S. flavescens* β-actin was amplified with the following primers: (Act-F: 5’-AACAGGC CAAC AGA GAG AAG ATG AC-3’, Act-R: 5’-ACC CACCA CTA AGC CGA TATTT-3’). The PCR for *β-actin* cDNA was performed with an initial denaturation at 94°C for 2 min, then 22 cycles each at 94°C for 30 s, at 53°C for 30 s, and 72°C for 50 s. The PCR products were separated by electrophoresis
using 1.5% gel at 100 V, and the gel was further stained by Ethidium Bromide Solution (Nacalai Tesque Inc., Japan).

RESULTS

Plant RNA Extraction and cDNA Library Preparation
The principal bioactive constituents of *S. flavescens* are the major quinolizidine alkaloids matrine and oxymatrine. The contents of these two components in *S. flavescens* are also the main valuation criteria of this plant. Meanwhile, considerable quantity of flavonoids and isoflavonoids are also accumulated in *S. flavescens*. In this study, we aimed to collect information about the nature of the genes responsible for the biosynthesis of matrine and oxymatrine, and study the related genes involved in the biosynthesis of flavonoids and isoflavonoids in *S. flavescens*. We extracted total RNA from the 9 tissues of this plant, resulting in 9 distinct cDNA libraries. We will refer to the libraries in the following manner: Lib 1 (callus), Lib 2 (leaf), Lib 3 (flower), Lib 4 (stem), Lib 5 (young bud), Lib 6 (mature bud), Lib 7 (bud right before blossom), Lib 8 (pedicel while bud stage), and Lib 9 (pedicel while blossom).

Illumina Sequencing and de Novo Assembly
All 9 libraries were processed using the Illumina HiSeq platform. Empty reads, reads of low quality and those containing unknown bases were trimmed using CLC software. Nucleotide fastq sequences of Illumina reads were deposited at DDBJ Sequence Read Archive (DRA) with the accession number DRA003182. To practice de novo assembly, the fastq reads from all 9 libraries were combined together to generate fasta format contigs. After the entire run was assembled, the quality of the assembly (e.g., the ratio of aligned reads, average contig size, N75, N50 and N25 contig length) was identified to be better than the de novo results from individual libraries (data not shown). The assembly of the 9 libraries was utilized in the following analysis to discover their genetic information.

85054 contigs were generated. By applying CD-HIT-EST with a threshold of 0.9, duplicates were identified and removed, leaving 83325 non-redundant contigs. An overview of the experimental pipeline is shown in Fig. 1. Table 1 summarizes sequencing and assembly results.

**Table 1. Overview of *S. flavescens* Transcriptome Assembly**

<table>
<thead>
<tr>
<th>Items</th>
<th>Numbers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total bases</td>
<td>20044281186</td>
</tr>
<tr>
<td>Average length of reads (bp)</td>
<td>98.4</td>
</tr>
<tr>
<td>No. of reads</td>
<td>203598590</td>
</tr>
<tr>
<td>Average length of contigs (bp)</td>
<td>664</td>
</tr>
<tr>
<td>Maximum contig length (bp)</td>
<td>15827</td>
</tr>
<tr>
<td>N75, N50, N25 (bp)</td>
<td>431, 969, 1947</td>
</tr>
<tr>
<td>No. of contigs over 300bp</td>
<td>85054</td>
</tr>
<tr>
<td>Non-redundant contigs</td>
<td>83325</td>
</tr>
</tbody>
</table>

**Guanine–Cytosine (GC) Content Analysis**
The reported GC content for unigene sequences in soybean and *Arabidopsis* was 43% and 44%, respectively. The average GC content of *S. flavescens* transcripts was found to be 39.3% (Supplementary 2). In eukaryotes, mean GC content varies from ca. 20 to 60%. Our values are in the middle of this range, slightly lower than those reported for *Glycine max* (43%) but very close to those reported for *Medicago truncatula* (40%).

**Transcriptome Information and the Differential Accumulation of Transcripts**
RPKM calculation was performed as the first step of transcript expression analysis after RNA-Seq reads from every library were aligned to all contigs. RPKM calculation considers gene length variation and the number of total mapped reads, which allows this normalized output to be used directly for the comparison of gene expression. To perform DE analysis, 36 pair-wise comparisons of the nine libraries were conducted by applying NOISeq-sim, which is a non-parametric approach for the identification of DE genes from count data or previously normalized count data.
By running NOISeq-sim on the R language platform with a given threshold \((q=0.9)\) for selecting differentially expressed features, the resultant number of DE transcripts varied across comparisons. The highest value obtained was 1061 differences between callus and mature bud transcripts; the lowest value obtained was 0 between mature bud and BBB (Fig. 2 and Supplementary 3).

**Protein Function Annotations and GO Classification**

Functional annotations according to sequence similarity are often the initial step in studying the role and biological functions of gene products.\(^{39}\) BLAST program was utilized to scan nucleotide query sequences against protein databases (NR, UniProt) to identify similar subject sequences. When the threshold \(E\)-value for BLASTx searches was set to \(1e^{-10}\) and if possible, the top 20 subject sequences for each query sequence were taken into consideration, we obtained 570611 subject sequences for all 83325 query sequences. To obtain reliable results while reducing redundancy, we set stricter requirements (refer to Annotation Pipeline and Data Mining) for retrieving the candidate genes. With this approach, significant matches were assigned to 27909 contigs.

Gene Ontology (GO), consisting of three main domains (cellular components, molecular functions and biological processes), is a useful instrument with which to study the nature of annotated genes.\(^{40}\)
Based on NCBI NR BLAST results, with the aid of Web Gene Ontology Annotation Plot (WEGO) software, 4, 12, 25921 contigs yielded corresponding GO terms that could be further classified into 50 sub-categories: 14 related to cellular components, 13 to molecular function and 23 to biological processes (Fig. 3). The limitation of evaluating gene classification by directly counting the number of GO terms which possess the same or very similar functions is that the expression level of these query sequences varies, which grants distinct weight to the same GO term as it corresponds to different query sequences. With this concern, overrepresented GO terms were identified by Fisher’s exact test. The one-tailed Fisher’s exact p-values corresponding to overrepresented categories were calculated according to the counts in 2 × 2 contingency tables. Counts \(n_{11}, \ldots, n_{22}\) in each table stand for: \(n_{11}\), number of observations of a specific category in the first gene set; \(n_{12}\), number of other categories in the first gene set; \(n_{21}\), number of observations of a category in the second gene set; and \(n_{22}\), number of observations of other categories in the second gene set. 4, 12, 25

\[ p\text{-Values were corrected using the false discovery rate (FDR) method with the threshold set at 0.05.} \]

For each \(S. \text{ flavescens}\) library, contigs with RPKM value over 30.0 (the top ca. 10% of all transcripts) were regarded as highly expressed genes and extracted respectively. Then the merged 9867 contigs were used to perform Fisher’s exact test. The overrepresented GO terms (Supplementary 4, GO terms with \(p < 1 \times 10^{-30}\) are listed) indicate expectedly cellular component like plasma membrane (GO: 0005886) and cytoplasm (GO: 0005737) plays crucial roles in all aspects regarding \(S. \text{ flavescens}\). Meanwhile, biological process including isopentenyl diphosphate biosynthetic process, methylenylthritol 4-phosphate pathway (GO: 0019288) is also highlighted in the list, which will provide information for future research.

**KEGG Pathway Retrieval** KEGG \(^4\) provides a robust instrument for biological pathway assignment as well as the functional annotation of gene products. Based on ID mapping results, we obtained 1350 unique enzymes and subsequently retrieved pathways using KEGG. These EC numbers were assigned to 154 biological pathways with the largest number of enzymes (707) involved in metabolic pathways. Given the remarkable reputation of \(S. \text{ flavescens}\) with regard to the ability to accumulate functional flavonoids, 18 flavonoid biosynthetic and 13 isoflavonoid biosynthetic enzymes are presented in Supplementary 5.

**Putative Genes Involved in Isoflavonoid and Quinolizidine Alkaloids Biosynthesis in \(S. \text{ flavescens}\)** Flavonoids are a group of polyphenolic compounds distributed widely throughout the plant kingdom. These compounds modulate the activity of enzymes to benefit the entire organism. As an important subgroup of flavonoids, isoflavonoids are mainly produced in legumes and affect oxidative stress markers, immune function and adipogenesis. 45 In the phenylpropanoid pathway, the synthesis of flavonoids is initialized by transforming phenylalanine into \(p\text{-coumaroyl-CoA}\). To initiate flavonoid biosynthesis, chalcone synthase catalyzes the formation of chalcone scaffolds, from which all flavonoids derive. 4, 12, 19

Based on our functional annotation findings, 34 contigs were predicted to represent seven enzymes critical to the biosynthesis of daidzein in \(S. \text{ flavescens}\). The number of contigs corresponding to each enzyme and the biosynthesis procedure are presented in Fig. 4. In our annotation result, contig c_86767 had 97.54% similarity to the L/ODC which catalyzes lysine...
into cadaverine and thus initiates the first steps toward the final product quinolizidine alkaloids including matrine and oxymatrine.\textsuperscript{11)} By far, despite the sustained effort to crack the biosynthetic pathway regarding quinolizidine alkaloids, the identified genes underlying this process are very few. We focused on co-expression pattern related to c_86767 in order to hunt for candidate genes. According to RPKM values from all the studied organs, we calculate the Pearson correlation coefficients. The heatplot (Fig. 5) demonstrates some promising candidates including c_55306 (MLP-like protein), c_87044 (cytochrome P450) and c_87047 (ripening related protein). Such contigs shared very similar expression pattern across all the 9 tissues, suggesting their potential relationship in the process of quinolizidine alkaloid synthesis.

For contig c_86767 annotated as \textit{L/ODC} gene, RPKM values of leaf and stem were as high as 553 and 321, respectively. Its expression pattern was further confirmed by semi-quantitative RT-PCR of \textit{L/ODC} gene in the samples from leaf, stem and root of \textit{S. flavescens}, using \(\beta\)-actin as the internal control gene,\textsuperscript{46)} as shown in Supplementary 6. PCR amplification was carried out with three different cycle numbers (30, 28 and 26 cycles) to verify the linearity of semi-quantitative analysis. The highest expression was seen in leaf followed by stem. No detectable expression was observed in root in this condition. These RT-PCR results verified the expression pattern deduced from RPKM values provided by RNA-Seq analysis. Furthermore, it was also consistent with the preferential \textit{L/ODC} expression in the leaf of \textit{Lupinus angustifolius}.\textsuperscript{11)} In all, accumulation of the final product such as matrine and oxymatrine would be high in other organs including root, but the initial steps for quinolizidine alkaloid biosynthesis feature in green parts of \textit{S. flavescens}, such as leaf and stem as suggested previously.\textsuperscript{5,10)}

**DISCUSSION**

\textit{S. flavescens} has remarkable pharmaceutical attributes, including treating viral hepatitis, viral myocarditis, gastrointestinal hemorrhage and skin diseases (such as psoriasis and eczema). Numerous studies have reported the isolation and pharmacological action of the bioactive components in \textit{S. flavescens}. Despite the importance and wide application of this medicinal plant, very few studies focused on its genetic profile. To generate overall deep transcriptome data for this plant will be helpful in discovering its underlying mechanism for the production of quinolizidine alkaloids. In this study, we utilized more than 200 million fastq format reads resulted from 9 tissues of \textit{S. flavescens} to perform RNA-Seq analysis. 83325 representative contigs were obtained and the key enzymes involved in the biosynthetic pathways of active compounds were retrieved (Supplementary 5).

As to differentially expressed genes, NOISeq-sim output
(Supplementary 3) showed DE genes identified by pair-wise comparison between callus and the rest 8 tissues were more than that of any other individual comparison. This is quite reasonable because in contrast with other organs of \textit{S. flavescens}, callus consists of undifferentiated photoautotrophic cells. And in the slightly different situations as mature bud and bud right before blossom, not a single DE gene could be detected. Callus of \textit{S. flavescens} does not accumulate quinolizidine alkaloids,\(^{5}\) which allows for the responsible genes to pop up by comparing it with other capable organs. By measuring the changes of matrine and oxymatrine in different growth stages\(^{73}\) and organs\(^{48}\) of \textit{S. flavescens} by HPLC, the concentration of above-mentioned compounds in root and seeds was 3- to 6-fold higher than that of leaf, stem and flower. Nevertheless, such reports also showed leaf, stem and flower did accumulate considerable quantity of matrine and oxymatrine. These observations suggest that the gene(s) responsible for the formation of such characteristic compounds should lie in the list of differentially expressed genes (Fig. 2) and the alkaloids may translocate from the site of de novo biosynthesis to the different sites of final accumulation.

Feeding studies illustrated the very first step for quinolizidine alkaloids biosynthesis. It concerns the decarboxylation of L-lysine into cadaverine by catalysis of L\textit{ODC} (EC 4.1.1.18). With the presence of copper amine oxidase (CuAO, EC 1.4.3.22), oxidative deamination of cadaverine produces 5-aminopentanal that spontaneously cyclizes to \(\Delta^2\)-piperidine Schiff base.\(^{49,50}\) The following steps to the final product matrine and oxymatrine still remain unknown. Based on co-expression pattern across the studied samples, we analyzed some candidate genes which were clustered into the same clade with the identified \textit{L/ODC} gene. What’s more, the genes suggested by the RNA-Seq data of \textit{S. flavescens} are also the ones indicated by the deep transcriptome data of \textit{Lupinus angustifolius}, a species of legume that accumulates considerable quinolizidine alkaloids (Bunsupa \textit{et al.}, unpublished).

Using enzyme commission numbers identified in \textit{S. flavescens} dataset to search KEGG for possible biosynthetic pathways retrieved 379 enzymes involved in biosynthesis of secondary metabolites. In addition to flavonoid biosynthetic pathway, valuable information about other pathways were also presented, such as monoterpenoid biosynthesis and steroid biosynthesis, which will contribute to the better understanding of the related mechanisms in plant kingdom. In this experiment, the obtained putative novel genes which may underlie the pharmacological function of \textit{S. flavescens} and the findings on several essential biosynthetic pathways may serve as a stepping-stone for further studies on this promising and time-honored medicinal plant.

\textbf{Acknowledgments} This study was supported, in part, by Grants-in-Aid for Scientific Research from the Ministry of Education, Culture, Sports, Science and Technology (MEXT) of Japan and by Health Labour Sciences Research Grant from the Ministry of Health, Labour and Welfare of Japan.

\textbf{Conflict of Interest} The authors declare no conflict of interest.

\textbf{Supplementary Materials} The online version of this article contains supplementary materials.

\textbf{REFERENCES}

5) Nagai N. Study on \textit{Sophora flavescens}. Yakugaku Zasshi, 84, 54–87 (1889).
17) Michael TP, Jackson S. The first 50 plant genomes. Plant Genome-Ux, 6, 03.0001in (2013).


