Comparative Transcriptome Analysis Reveals Candidate Genes Involved in Isoquinoline Alkaloid Biosynthesis in *Stephania tetrandra*

Authors
Yangyang Zhang[1], Yun Kang[1], Hui Xie[2], Yaqin Wang[1], Yaoting Li[1], Jianming Huang[1]

Affiliations
1 School of Pharmacy, Fudan University, Shanghai, P. R. China
2 Human Phenome Institute, Fudan University, Shanghai, P. R. China

Key words
Menispermaceae, *Stephania tetrandra*, transcriptome, isoquinoline alkaloids, biosynthesis, RT-qPCR

received	February 7, 2020
accepted after revision	June 21, 2020
published online	August 5, 2020

Bibliography
Planta Med 2020; 86: 1258–1268
DOI 10.1055/a-1209-3407
ISSN 0032-0943

Georg Thieme Verlag KG, Rüdigerstraße 14, 70469 Stuttgart, Germany

Correspondence
Dr. Jianming Huang
School of Pharmacy, Fudan University
826 Zhangheng Road, Shanghai 201203, P. R. China
Phone: + 86 21 51 98 01 32, Fax: + 86 21 51 98 01 32
jmhuang@shmu.edu.cn

Supporting information available online at http://www.thieme-connect.de/products

ABSTRACT

The roots of *Stephania tetrandra* are used as a traditional Chinese medicine. Isoquinoline alkaloids are considered to be the most important and effective components in this herb, but little is known about the molecular mechanism underlying their biosynthesis. In this context, this study aimed to reveal candidate genes related to isoquinoline alkaloid biosynthesis in *S. tetrandra*. Determination of tetrandrine and fangchinoline in the roots and leaves of *S. tetrandra* by HPLC showed that the roots had much higher contents of the two isoquinoline alkaloids than the leaves. Thus, a comparative transcriptome analysis of the two tissues was performed to uncover candidate genes involved in isoquinoline alkaloid biosynthesis. A total of 71 674 unigenes was obtained and 31 994 of these were assigned putative functions based on BLAST searches against 6 annotation databases. Among the 79 isoquinoline alkaloid-related unigenes, 51 were differentially expressed, with 42 and 9 genes upregulated and downregulated, respectively, when the roots were compared with the leaves. The upregulated differentially expressed genes were consistent with isoquinoline alkaloid accumulation in roots and thus were deemed key candidate genes for isoquinoline alkaloid biosynthesis in the roots. Moreover, the expression profiles of 10 isoquinoline alkaloid-related differentially expressed genes between roots and leaves were validated by quantitative real-time polymerase chain reaction, which indicated that our transcriptome and gene expression profiles were reliable. This study not only provides a valuable genomic resource for *S. tetrandra* but also proposes candidate genes involved in isoquinoline alkaloid biosynthesis and transcription factors related to the regulation of isoquinoline alkaloid biosynthesis. The results lay a foundation for further studies on isoquinoline alkaloid biosynthesis in this medicinal plant.

Introduction

Stephania Lour. is the largest genus of the family Menispermaceae and includes approximately 60 species distributed in tropical and subtropical areas of the world. Forty species have been identified in China, most of which are medicinal plants containing diverse bioactive alkaloids [1, 2]. The roots of *Stephania tetrandra* S. Moore, which have been recorded in the Chinese Pharmacopoeia, are widely used to treat hypertension, angina pectoris, fibrosis, and tumors [3]. This species is rich in IQAs, of which tetrandrine and fangchinoline are major constituents. These IQAs possess significant antimicrobial, antitumor, anti-inflammatory, and analgesic activities [4, 5]. Although these metabolites are of wide pharmacological importance, their underlying biosynthetic pathways are still unclear.

In recent decades, overexploitation has made the wild resources of this species increasingly rare, and planting yields are rather limited. To sustainably meet the demands for this herb

ABBREVIATIONS

COG	clusters of orthologous groups of proteins
CYP	cytochrome P450
DEG	differentially expressed gene
FPKM	fragments per kilobase of transcript per million mapped reads
GO	gene ontology
IQA	isoquinoline alkaloid
KEGG	Kyoto Encyclopedia of Genes and Genomes
NCBI	National Center for Biotechnology Information
NR	NCBI non-redundant protein sequences
ORF	open reading frame
Pfam	protein family
qRT-PCR	quantitative real-time polymerase chain reaction
SPE	solid-phase extraction
St	*Stephania tetrandra*
TAT	tyrosine amino transferase
TF	transcription factor

and its bioactive components, it is important to study the biosynthesis of IQAs in *S. tetrandra*, which could help to improve metabolic regulation in this plant and the sustainable production of target bioactive metabolites.

IQAs are derived from the conversion of L-tyrosine to dopamine and 4-hydroxyphenylacetaldehyde and then the formation of various structures via intramolecular coupling, reduction, methylation, hydroxylation, and other reactions [6]. The biosynthetic pathways underlying several IQAs have been reported in some plants other than *Stephania* species, such as the biosynthesis of bisbenzylisoquinoline alkaloids in *Berberis stolonifera* and aporphine alkaloids in *Nelumbo nucifera*, and many enzymes related to IQA biosynthesis in these plants have been characterized [6, 7].

However, little is known about the biosynthesis of IQAs in *S. tetrandra*, mainly because of the lack of genome or transcriptome information. Given the fact that plant secondary metabolites are often biosynthesized in a tissue-specific manner [8], transcriptomic comparisons between plant tissues combined with correlation analyses of chemical components represent an efficient approach to discover the key genes involved in secondary metabolism [9–11]. This method has already been successfully used to uncover candidate genes related to the biosynthesis of chemical constituents in some plant species [6, 12–14].

In this study, HPLC analysis of tetrandrine and fangchinoline in the roots and leaves of *S. tetrandra* indicated the tissue specificity of IQA biosynthesis. Then, a comparative transcriptome analysis of the two tissues was conducted to reveal candidate genes involved in the biosynthesis of IQAs and TFs related to the regulation of IQA biosynthesis in *S. tetrandra*. Moreover, 10 IQA-related unigenes were validated by qRT-PCR. This work provides a molecular basis for understanding IQA accumulation in the roots of *S. tetrandra* and lays the foundation for further studies on the molecular mechanism underlying IQA biosynthesis in this medicinal plant.

Results

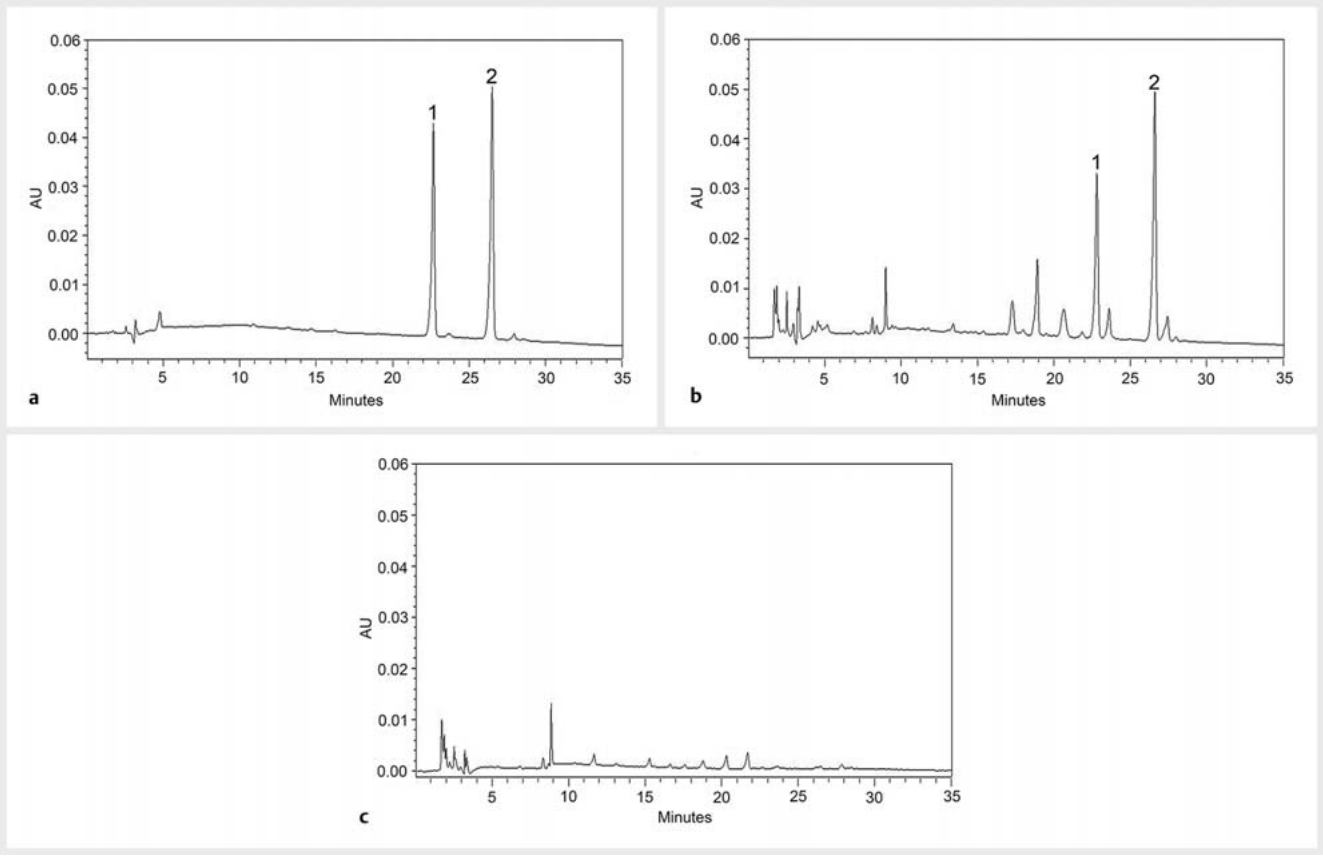
According to previous reports, tetrandrine and fangchinoline are the main IQA constituents of *S. tetrandra* [5, 15, 16]. Thus, an SPE-HPLC method was developed for the determination of the two IQAs in the roots and leaves of *S. tetrandra*. The representative sample chromatograms (▶ **Fig. 1**) and the validation results (**Tables 1S–3S**, Supporting Information) indicated that the proposed method was specific, accurate, and repeatable. The average contents of tetrandrine and fangchinoline in the three root samples were 9.44 and 6.06 mg/g, respectively, showing that the roots were rich in IQAs; in contrast, the two IQAs were almost undetectable in the leaves (▶ **Fig. 2**). This organ-specific distribution of tetrandrine and fangchinoline suggested a root-preferred expression pattern of IQA biosynthetic pathway genes.

To uncover the molecular mechanisms underlying the tissue specificity of IQA biosynthesis in *S. tetrandra*, six RNAseq libraries were constructed from the total RNA of each plant sample. After removing adaptor sequences and ambiguous and low-quality reads, 173, 128, 434 and 150, 248, 762 clean reads were obtained from leaves and roots, respectively, and both data sets were characterized by Q30≥94.24%. These clean reads were used for *de novo* transcriptome assembly. The *S. tetrandra* transcriptome integrated from the clean reads was then constructed and consisted of 71 674 unigenes with an average length of 1044 bp and an N50 length of 1813 bp. The length distribution of all unigenes is shown in **Fig. 1S**, Supporting Information.

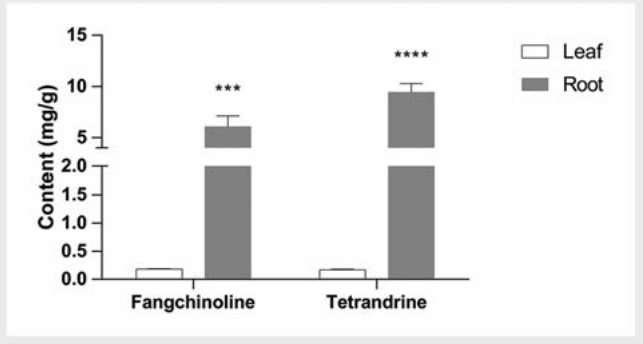
The putative functions of the unigenes were annotated against six public protein databases (▶ **Table 1**). Among the 71 674 unigenes, 31 994 (44.64%) were annotated in at least 1 public database, of which 6964 unigenes (56.24%) were co-annotated in all 6 databases. Based on the NR database, 18.71 and 13.52% of the annotated unigenes were matched to the sequences from *Macleaya cordata* [17] and *N. nucifera* [6], respectively (▶ **Fig. 3**), two species which are also rich in isoquinoline alkaloids. Additionally, 87.82% of the unigenes were distributed within the 60 to 100% similarity interval, indicating high-quality annotation of the transcriptome data.

In the GO analysis, 21 376 unigenes were assigned into 3 main categories, including biological processes, cellular components, and molecular functions (**Fig. 2S**, Supporting Information). Metabolic process and cellular process were the most abundant subcategories among the biological processes, whereas in the molecular functions, binding and catalytic activities were the richest. GO analysis suggested that the identified unigenes were involved in a series of biological processes. There were 6964 unigenes annotated and grouped into 24 functional categories in the COG database, which enables for homologously classifying gene products. Among these categories, the group of "secondary metabolite biosynthesis, transport, and catabolism" (98 unigenes) plays an essential role in the biosynthesis of secondary metabolites of *S. tetrandra* (**Fig. 3S**, Supporting Information).

KEGG pathway analysis is useful to understand the biological function of genes. A total of 10 982 unigenes were annotated in the KEGG database and were assigned to 6 main categories and 131 biological pathways (▶ **Fig. 4** and **Table 4S**, Supporting Information). The category of metabolism, which had the largest num-



▶ **Fig. 1** Representative HPLC chromatograms of (**a**) a standard mixture of tetrandrine and fangchinoline, (**b**) *S. tetrandra* roots, and (**c**) *S. tetrandra* leaves. Peaks: (1) fangchinoline; (2) tetrandrine.



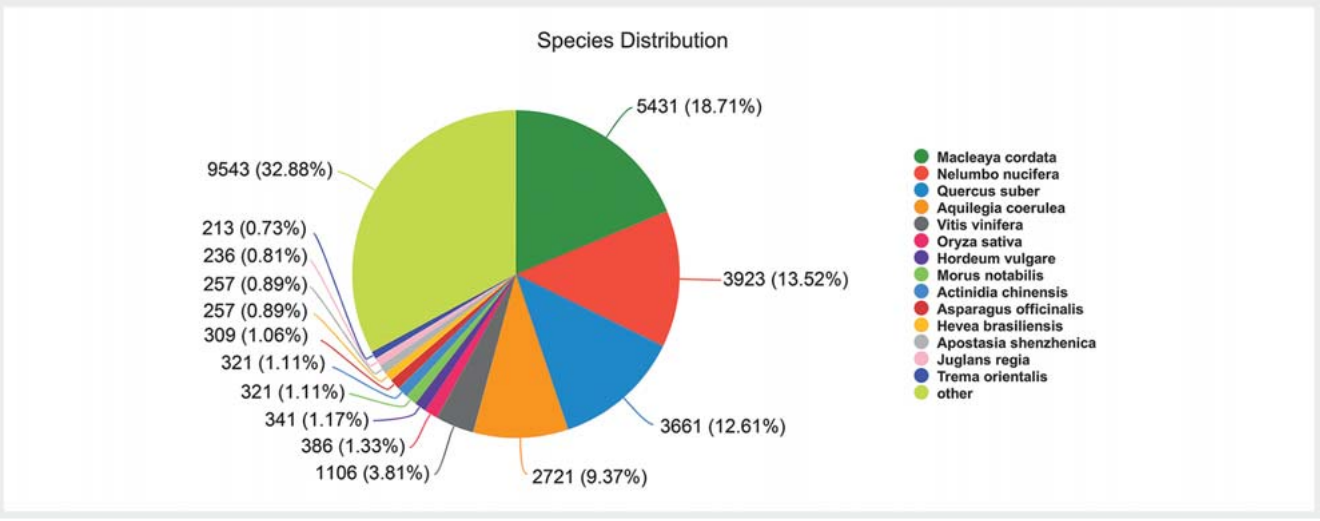
▶ **Fig. 2** Contents of two main IQAs in the leaves and roots of *S. tetrandra*. ***P < 0.001, *****p < 0.0001.

▶ **Table 1** Statistics of annotations in six public databases for *S. tetrandra* unigenes.

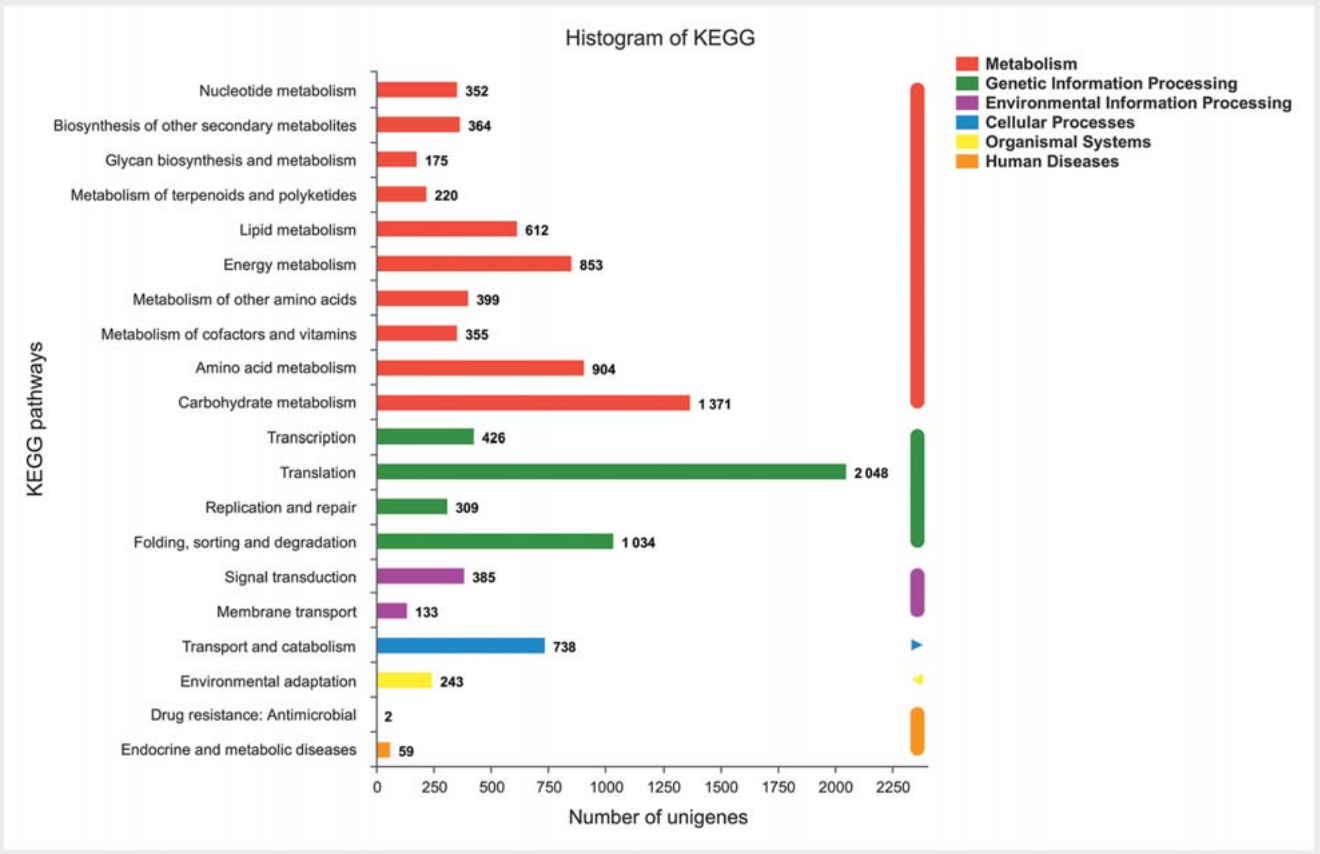
Database	Number of unigenes	Percentage (%)
NR	29228	40.78
Swiss-Prot	23844	33.27
Pfam	22831	31.85
COG	6964	9.72
GO	21376	29.82
KEGG	15017	20.95
Co-annotated in all six databases	3731	5.21
Annotated in at least one database	31994	44.64

ber of unigenes, consisted of 364 unigenes involved in the biosynthesis of other secondary metabolites. These unigenes were distributed over 13 pathways including IQA biosynthesis (79 unigenes, 21.7%) (**Table 5S**, Supporting Information).

To investigate DEGs between roots and leaves in *S. tetrandra*, the FPKM method was adopted. Thus, 37084 unigenes from the roots and 38062 from the leaves were found to have FPKM values above 1. Among them, 33236 unigenes were expressed in both tissues, whereas 4826 and 3848 were specifically expressed in leaves and roots, respectively (**Fig. 4Sa**, Supporting Information). The number of DEGs between the two tissues was 11433, of which 5664 unigenes were upregulated and 5769 unigenes were downregulated in the roots compared to expression in the leaves (**Fig. 4Sb**, Supporting Information).

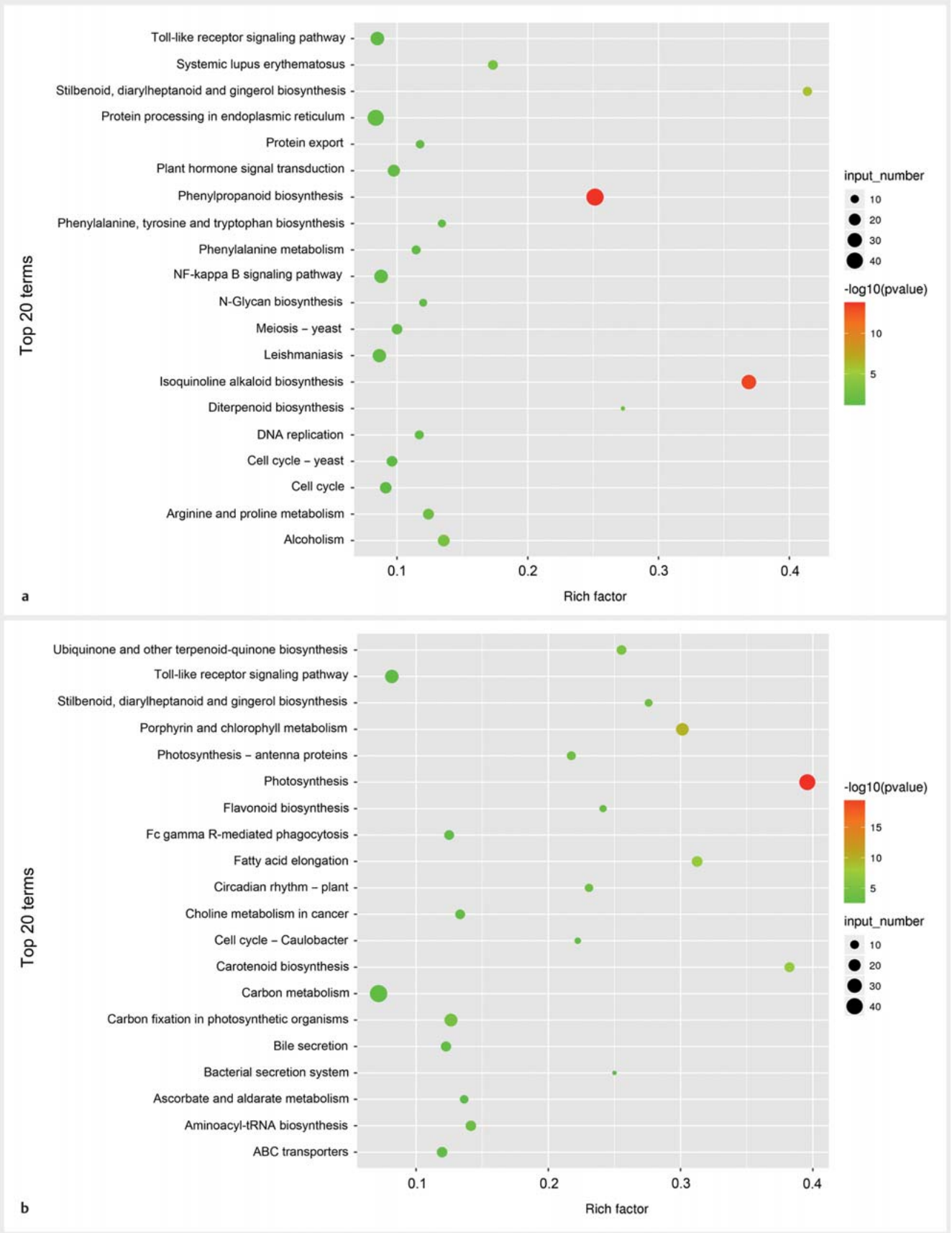


▶ **Fig. 3** Species distribution of the unigenes annotated in non-redundant protein sequences (NR) database.

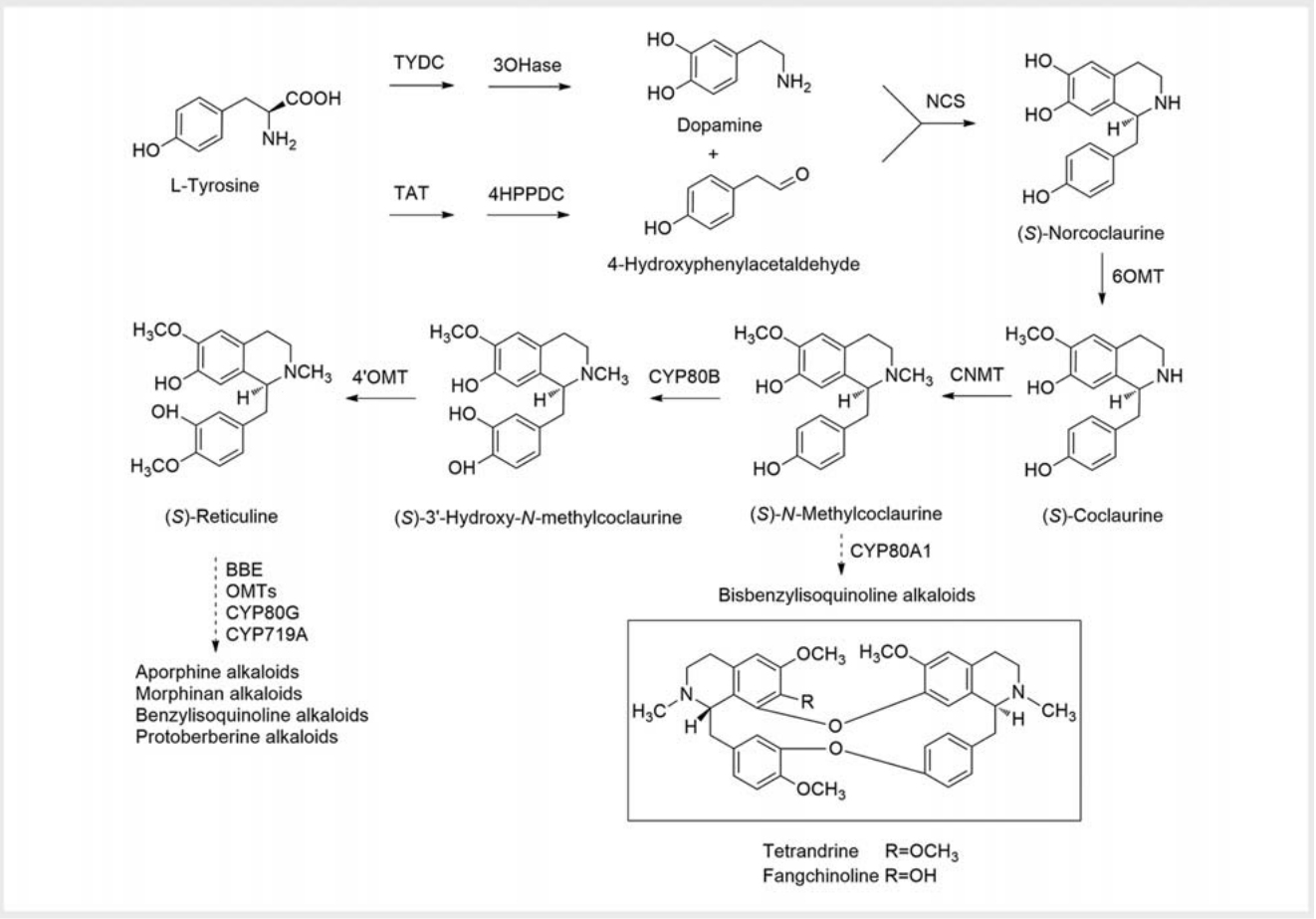


▶ **Fig. 4** Kyoto Encyclopedia of Genes and Genomes (KEGG) annotation of the unigenes in *S. tetrandra*.

These DEGs were further used for GO and KEGG enrichment analyses. GO terms corresponding to organic substance metabolic process, nitrogen compound metabolic process, and primary metabolic process were significantly enriched in the roots of *S. tetrandra* (**Fig. 5S**, Supporting Information). KEGG pathways specifically enriched in the leaves included photosynthesis, carotenoid biosynthesis, and porphyrin and chlorophyll metabolism, whereas phenylpropanoid biosynthesis and isoquinoline biosynthesis were specifically enriched in the roots (▶ **Fig. 5**). These results were consistent with the biological functions of these tissues, indicat-



▶ **Fig. 5** Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analyses of the differentially expressed unigenes in the roots (**a**) and leaves (**b**) of *S. tetrandra*.



▶ **Fig. 6** Proposed biosynthetic pathway for IQAs in *S. tetrandra*. Arrows with dashed lines represent multiple enzymatic reactions.

ing that *S. tetrandra* root is the main tissue for the accumulation of specific secondary metabolites, and that our transcriptome data are suitable for studying the biosynthesis of these metabolites, especially IQAs.

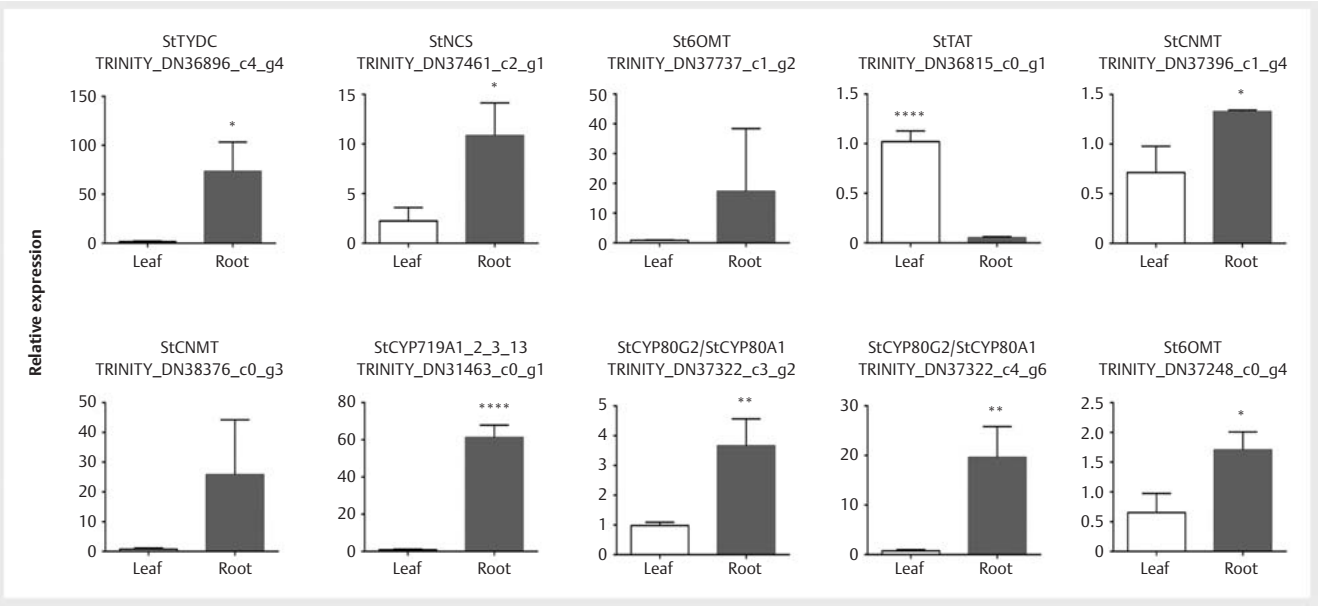
We next focused on the discovery of genes involved in IQA biosynthesis. Among the 79 IQA-related unigenes based on KEGG annotation, 51 were differentially expressed with 42 and 9 genes upregulated and downregulated, respectively, when roots were compared with leaves (**Table 5S** and **Fig. 6S**, Supporting Information). The upregulated DEGs were deemed to be key candidate genes for IQA biosynthesis in the roots. Based on KEGG pathways, we discovered the unigenes encoding most known enzymes for (*S*)-reticuline (a common precursor of benzylisoquinoline alkaloids) biosynthesis, which included TAT, TYDC, NCS, 6OMT, CNMT, 4′OMT, and CYP80B. Moreover, the downstream CYP450 candidate genes, including those encoding CYP80G2, CYP719A, CYP80B1, and CYP80A1, were also found (▶ **Fig. 6**). Most of the identified candidate genes showed a higher expression level in the roots than in the leaves, which was particularly true of those located downstream of IQA biosynthesis. This expression pattern was associated with a higher content of IQAs in the roots of *S. tetrandra*. Thus, these candidate genes were likely to play an important role in IQA accumulation in the roots.

To confirm the transcriptome analysis results and to evaluate the differential expression profiles between the two tissues, 10 DEGs related to IQA biosynthesis were chosen for qRT-PCR analysis. The results showed that the expression patterns of all 10 genes were consistent between qRT-PCR (▶ **Fig. 7**) and DEG analysis (**Table 5S**, Supporting Information), despite some differences in fold changes. Thus, our transcriptome and gene expression profile are reliable for the study of candidate genes related to IQA biosynthesis in *S. tetrandra*.

Discussion

IQAs are the most important active components in *S. tetrandra*. The contents of major IQAs were much higher in the roots than in the leaves, which suggests that this is an appropriate species to study candidate genes involved in IQA biosynthetic pathways.

The biosynthetic pathways underlying the generation of several IQAs have been reported in some other plant species, such as those leading to bisbenzylisoquinoline alkaloids in *B. stolonifera* [7], benzylisoquinoline alkaloids in *Coptis* species, and aporphine alkaloids in *Coptis japonica* [18]. Generally, the biosynthesis of IQAs begins with the conversion of L-tyrosine to dopamine and 4-hydroxyphenyl acetaldehyde, which are then transformed by a se-



▶ **Fig. 7** Expression patterns of 10 unigenes involved in IQA biosynthesis in *S. tetrandra*. The unigenes were analyzed by qRT-PCR with actin as a reference gene.

ries of enzymes, including NCS, 6OMT, CNMT, 4′OMT, and CYP80B, to form the central intermediate (*S*)-reticuline. Then (*S*)-reticuline undergoes intramolecular coupling, reduction, methylation, hydroxylation, and other reactions to form different types of IQAs, including benzylisoquinoline, protoberberine, aporphine, and morphinan alkaloids [6, 19–23]. Bisbenzylisoquinoline alkaloids are produced through another pathway, which involves the (*S*)-reticuline precursor *N*-methylcoclaurine and is catalyzed by CYP80A1 [7, 20].

In *S. tetrandra*, homologs of NCS, 6OMT, CNMT, 4′OMT, CYP80B1, CYP80A1, CYP719A1_2_3_13, and CYP80G2 were found in this study (**Table 5S**, Supporting Information). NCS, which catalyzes the condensation of dopamine and 4-hydroxyphenyl acetaldehyde to (*S*)-norcoclaurine, plays an important role in IQA biosynthesis because of its entry point location in the pathway [19–22, 24].

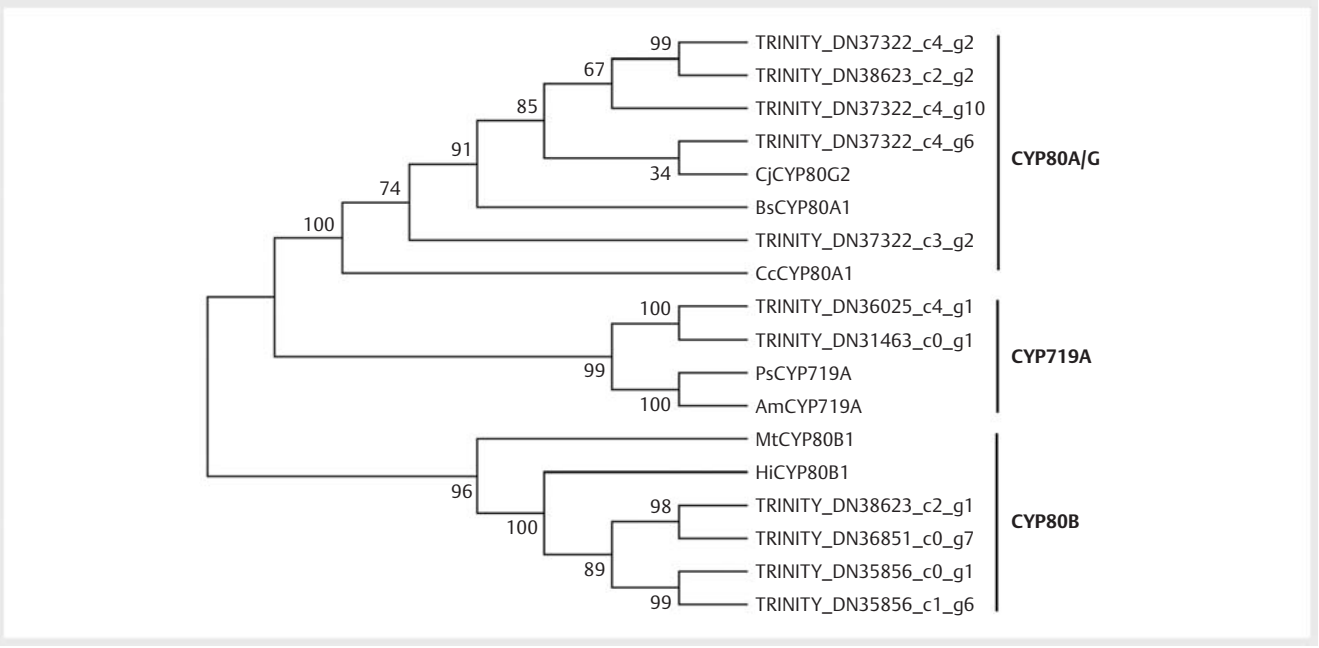
Two NCS differentially expressed unigenes were identified in the *S. tetrandra* transcriptome. The expression levels of both in roots were 3.46- and 3.68-fold higher than those in leaves. These two unigenes showed a significant correlation with alkaloid accumulation in roots.

The homologs of 6OMT, 4′OMT, and CNMT were methyltransferases identified in *S. tetrandra*. Among the nine differentially expressed homologs of 6OMT, six were upregulated and three were downregulated in the roots compared to the expression levels in the leaves. Only one 4′OMT-encoding unigene was found and its expression level was 39-fold higher in the roots than in the leaves. All 12 differentially expressed homologs of CNMT were upregulated in the roots compared to expression levels in the leaves. These results indicated that most methyltransferases were positively correlated with alkaloid accumulation in roots.

CYP450s play a key role in synthesizing plant secondary metabolites. CYP80A, CYP80B, and CYP80G catalyze C–O phenol-coupling, hydroxylation, and C-C phenol-coupling, respectively, whereas CYP719A catalyzes methylenedioxy bridge formation [7, 21–23]. Eleven differentially expressed unigenes encoding IQA-related CYP450 were found in *S. tetrandra*, including CYP80B1, CYP719A1_2_3_13, CYP80A1, CYP80G2, and all were upregulated in the roots compared with expression levels in the leaves, indicating a positive effect on high IQA content in the roots. Phylogenetic analysis (▶ **Fig. 8**) showed that two homologs of CYP719A1 (TRINITY_DN36025_C4_g1 and 31463_C0_g1) clustered with the two known CYP719A genes with a high bootstrap value. Likewise, four homologs of CYP80B were phylogenetically related to previously reported CYP80B1 genes in other species. However, five homologs of CYP80A/G were related to either the CYP80A or CYP80G gene clusters and thus could not be definitively assigned. The reasons for this ambiguous identification could be the high similarity of amino acid sequences between CYP80G and CYP80A [18, 20] and the lack of enough reference genes, which impeded the clarification of their differences. However, the homologs of CYP80A/G would be key candidate genes involved in the biosynthesis of major bisbenzylisoquinoline alkaloids such as tetrandrine and fangchinoline in *S. tetrandra*, because CYP80A has been reported to catalyze the transformation of (*S*)-*N*-methylcoclaurine into bisbenzylisoquinoline alkaloids such as berbamunine [25] and liensinine [26]. Therefore, we will undertake, in a further study, the functional characterization of these CYP450 genes, which are important for IQA biosynthesis in *S. tetrandra*.

Combining the expression pattern with alkaloid content, we predict that the 42 upregulated homologous unigenes in the roots, which are positively correlated with IQA accumulation, are candidate genes involved in IQA biosynthesis in *S. tetrandra* roots.

TFs, which usually exist as gene families, play essential roles in regulating the expression of related genes to control the flux of



▶ **Fig. 8** Phylogenetic analysis of IQA-related CYPs. The phylogenetic tree was constructed based on the amino acid sequences from *S. tetrandra* CYPs and other plant CYPs. Abbreviations and GenBank accession numbers for the sequences from other plants are as follows: BsCYP80A1, *B. stolonifera* cytochrome P-450 CYP80 (AAC48987.1); CcCYP80A1, *Capsicum chinense* Berbamunine synthase (PHU28278.1); CjCYP80G2, *C. japonica* var. Dissecta corytuberine synthase (BAF80448.1); PsCYP719A, *Papaver somniferum* stylopine synthase (ADB89214.1); AmCYP719A, *Argemone mexicana* stylopine synthase (ABR14721.1); MtCYP80B1, putative *Medicago truncatula N*-methylcoclaurine 3′-monooxygenase (RHN63317.1); HiCYP80B1, *Handroanthus impetiginosus N*-methylcoclaurine 3′-monooxygenase (PIN00452.1).

secondary metabolites. TFs that regulate IQA biosynthesis are primarily focused on the WRKY and bHLH families. In *C. japonica*, expression of these two transcription factors causes a significant increase in the expression of several berberine biosynthetic genes [27, 28]. In *California poppy*, the transactivation effect of the regulatory factor WRKY1 results in an increase of up to 60-fold in the expression level of EcCYP80B1 [(*S*)-*N*-methylcoclaurine 3′-hydroxylase] and EcBBE (berberine bridge enzyme) transcripts [29]. Thus, candidate TFs might be used to increase the production of IQAs.

In *S. tetrandra*, 15 DEGs encoding WRKY TFs and 30 DEGs encoding bHLH TFs were identified (▶ **Fig. 9**), of which 13 WRKY and 18 bHLH genes were upregulated in the roots compared with the expression levels in the leaves (**Table 6S**, Supporting Information). These upregulated unigenes are critical candidate genes to further investigate the regulation of IQA biosynthesis in *S. tetrandra*. However, further studies are still needed to verify the function of these candidate TFs in regulating IQA biosynthesis.

In summary, we conducted a comprehensive transcriptome analysis of leaves and roots of *S. tetrandra* and identified candidate genes involved in IQA biosynthesis and TFs related to the regulation of this process. These results will be helpful to explain the accumulation of IQAs in *S. tetrandra* and provide a molecular basis for further studies on the IQA biosynthesis in this medicinal plant.

Materials and Methods

Chemicals

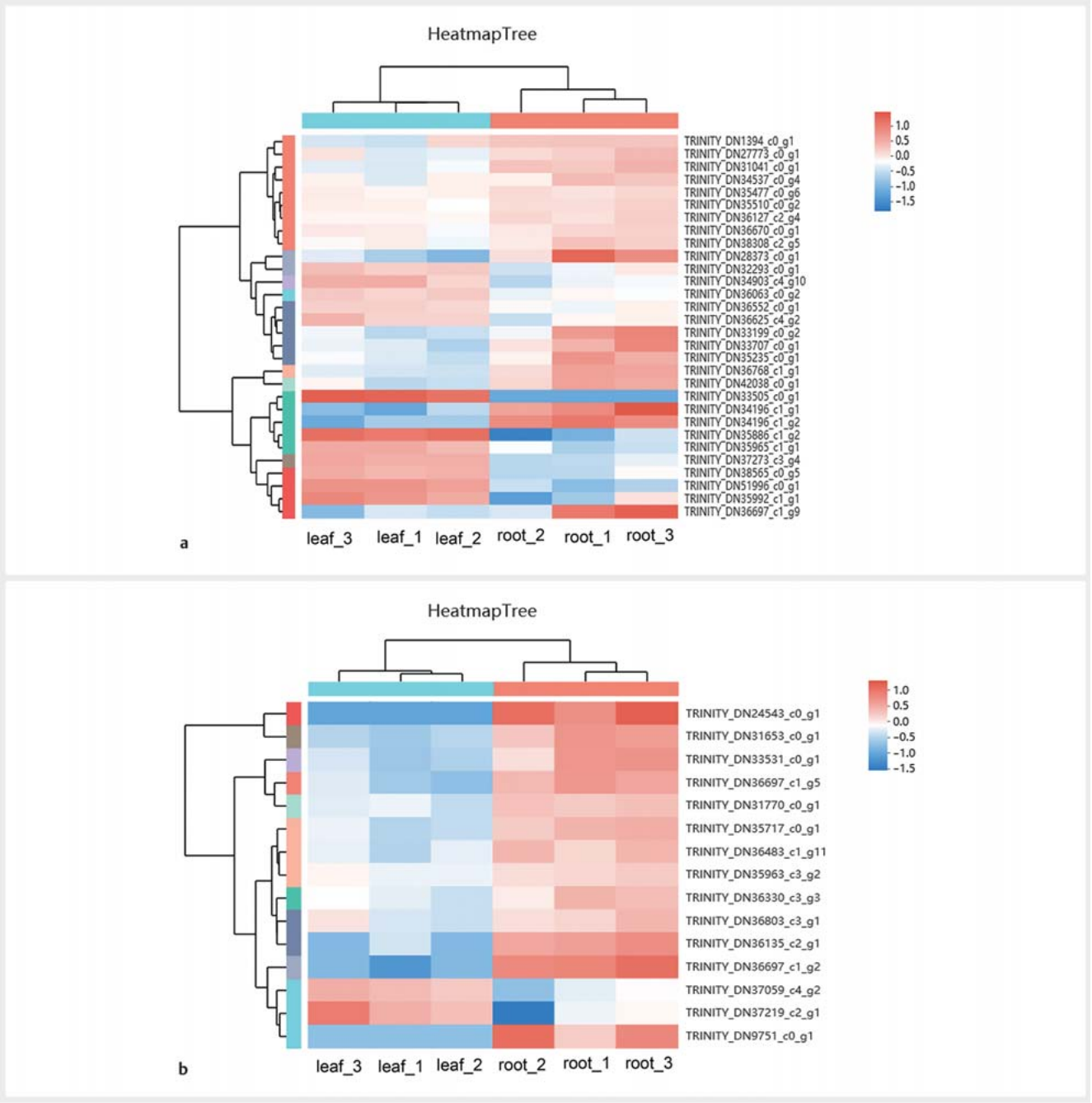
Tetrandrine (purity > 98 %) and fangchinoline (purity > 98 %) were purchased from Chengdu Desite Bio-Technology Co., Ltd. Solvents used for extraction and HPLC analysis were of HPLC grade and were obtained from Shanghai Xingke High Purity Reagent Co., Ltd.

Plant materials

S. tetrandra was collected from Hangzhou Botanical Garden in Zhejiang province, China (30°15′N, 120°7′E and altitude of 42 m) in June 2018. The plant material was identified by Dr. Yun Kang, Fudan University. Voucher specimens (S20180009, S20180010, and S20180020) were deposited in the herbarium of the School of Pharmacy, Fudan University (SHMU). Three biological replicates of the roots and leaves were randomly harvested separately from healthy individual plants. One part of each sample was cut into small pieces, frozen immediately with liquid nitrogen, and then stored at − 80 °C until further RNA extraction; meanwhile, the other sample was desiccated with silica gel immediately after collection and then dried at 40 °C for chemical analysis.

Quantification of tetrandrine and fangchinoline by HPLC

An ultrasonic-assisted extraction method followed by a solid-phase cleanup step was developed for the extraction of major



▶ **Fig. 9** Heat maps of differentially expressed unigenes encoding WRKY TFs (**a**) and bHLH1 TFs (**b**) in the roots and leaves of *S. tetrandra*.

IQAs (tetrandrine and fangchinoline) from *S. tetrandra*. The dried powder of *S. tetrandra* (0.1 g) was ultrasonicated with 6 mL of hydrochloric acid (0.01 mol/L) for 30 min. The extract was filtered, and the filtrate was diluted to 10 mL with 0.01 mol/L hydrochloric acid. An aliquot (5 mL) of the extract solution was then loaded onto a PCX cartridge (Cleanert PCX, 60 mg, 1 mL) preconditioned successively with methanol (3 mL) and 0.01 mol/L hydrochloric acid (3 mL). The loaded cartridge was washed successively with 0.01 mol/L hydrochloric acid and methanol (3 mL) and the retained compounds were eluted with 3 mL of a methanol–25% ammonia solution (97.5 : 2.5, *v*/*v*).

To determine tetrandrine and fangchinoline in *S. tetrandra*, a Waters Alliance 2695 HPLC system with a 996 photodiode array detector was used. Chromatographic separations were carried out on a Promosil C_{18} column (250 mm × 4.6 mm, 5 µm; Aglea) at a column temperature of 40 °C. The mobile phase consisted of 0.5% triethylamine aqueous solution adjusted to pH 9.0 with 10% phosphoric acid (solvent A) and acetonitrile (solvent B). The gradient elution was as follows: 0–30 min, linear from 20% to 95% B; 30–35 min, held at 95% B. The flow rate was 1.0 mL/min and the injection volume was 10 µL. The monitoring wavelength was set at 280 nm.

RNA extraction, RNA-seq library preparation, and Illumina sequencing

Total RNA was extracted from the roots and leaves of *S. tetrandra* using TRIzol Reagent (Invitrogen), and genomic DNA was removed using DNase I (TaKara). RNA purification, reverse transcription, library construction, and sequencing were performed at Shanghai Majorbio Bio-pharm Biotechnology Co., Ltd. using the methods previously reported.

The *S. tetrandra* RNA-seq transcriptome libraries were prepared using the Illumina TruSeqTM RNA sample preparation kit. Six RNA-seq libraries were sequenced in a single lane on an Illumina Hiseq X ten sequencer (Illumina) for 2 × 150 bp paired-end reads. All raw sequencing data were submitted and deposited in NCBI's Gene Expression Omnibus (GEO) repository with the SRA accession number: PRJNA599532

De novo assembly and functional annotation

The raw reads from *S. tetrandra* were trimmed and subjected to quality control using SeqPrep (https://github.com/jstjohn/SeqPrep) and Sickle (https://github.com/najoshi/sickle) with default parameters. Next, clean data were used to perform *de novo* assembly with Trinity software (Version v2.8.5) [30]. All assembled unigenes were annotated functionally via BLAST searches against the following six public annotation databases: NR, SwissProt, Pfam, GO, COG, and KEGG [31, 32].

Differential expression analysis of unigenes

To identify DEGs between different tissues of *S. tetrandra*, the expression level of each transcript was calculated using the FPKM method. RSEM software (Version 1.3.1) [33] was used to quantify gene and subtype abundances. The R statistical package software EdgeR (Version 3.24.3) [34] was used for differential expression analysis. In this work, a |log2 fold-change| ≥ 1 and a p value ≤ 0.05 were set as thresholds to judge the significance of gene expression. The DEGs were then used for functional enrichment analysis including GO and KEGG.

Phylogenetic analysis

Phylogenetic analysis was performed based on the deduced amino acid sequences of CYPs involved in IQA biosynthesis in *S. tetrandra* and other plants. The evolutionary distances were calculated with the Poisson correction method and a neighbor-joining tree was established with MEGA6.0 software. Bootstrap values obtained after 1000 copies are indicated on the branches.

Analysis of genes encoding transcription factors

To find the TF families expressed in the *S. tetrandra* transcriptome, the ORFs of each unigene were tested with the software getorf (EMBOSS:6.5.7.0) [35]. These ORFs were then compared to all TF protein domains using the plant transcription factor database PlantTFDB (Version 4.0) with BLASTX (e value ≤ 1e-5), adopting the hmmsearch method [36].

Verification of gene expression using quantitative real-time polymerase chain reaction

To validate RNA-Seq data, qRT-PCR was performed using an ABI7500 fast Real-Time PCR system (Applied Biosystems) with a HiScript Q RT SuperMix for qPCR (+ gDNA wiper) (Vazyme Biotech). Ten DEGs related to IQA biosynthesis were selected for qRT-PCR analysis. The actin gene (TRINITY_DN37280_c2_g18) was used as a reference to standardize the expression level. The primer sequences used and the conditions of each qRT-PCR reaction are listed in **Table 7S**, Supporting Information. Three biological replicates and three technical repeats were performed for each candidate gene and sample. The relative expression level of each unigene was calculated by the $2^{-\Delta\Delta CT}$ method [37].

Supporting information

Retention times, regression equations, correlation coefficients, LODs, LOQs, linear ranges, precision, accuracy, and repeatability for the two analytes in *S. tetrandra* (**Tables 1S** and **2S**, respectively), the contents of the two analytes in the three batches of *S. tetrandra* (**Table 3S**), KEGG pathways assigned to the unigenes, 79 unigenes related to isoquinoline alkaloid biosynthesis, 15 DEGs encoding WRKY TFs, and 30 DEGs encoding bHLH TFs in *S. tetrandra* (**Tables 4S–6S**, respectively), primer sequences used in qRT-PCR (**Table 7S**), length distribution of *S. tetrandra* unigenes (**Fig. 1S**), classification of the *S. tetrandra* unigenes based on GO and COG analysis (**Figs. 2S** and **3S**, respectively), unigenes expressed in the leaves and roots of *S. tetrandra* (**Fig. 4S**), and GO enrichment analysis of the DEGs and heat map based on the expression level of the IQA-related DEGs between the roots and leaves of *S. tetrandra* (**Figs. 5S** and **6S**, respectively) are available as Supporting Information.

Contributors' Statement

Data collection: Y. Y. Zhang, Y. Kang, H. Xie, Y. Q. Wang, Y. T. Li, J. M. Huang; design of the study: Y. Y. Zhang, Y. Kang, H. Xie, J. M. Huang; statistical analysis: Y. Y. Zhang, Y. Kang, H. Xie, Y. Q. Wang, Y. T. Li, J. M. Huang; analysis and interpretation of the data: Y. Y. Zhang, Y. Kang, H. Xie, Y. Q. Wang, Y. T. Li, J. M. Huang; drafting the manuscript: Y. Y. Zhang; Y. Q. Wang, Y. T. Li; critical revision of the manuscript: J. M. Huang.

Acknowledgements

We thank Shun Liu, Long Lin, Jiayun Xue, Haitian Hao, and Xieli Wang for assistance with the field work.

Conflict of Interest

The authors declare that they have no conflict of interest.

References

[1] Editorial Committee of Flora of China. Flora Reipublicae Popularis Sinicae, 1st Edition, Vol. 30 (1). Beijing: Science Press; 1996: 40–70

[2] Zhang ZG, Meng AP, Wang HC, Li JQ. A new species of *Stephania* (Menispermaceae) from South Guangxi, China. Novon 2013; 22: 379–382

[3] National Pharmacopoeia Commission. Pharmacopoeia of the People's Republic of China, Volume 1. Beijing, China: Medical Science and Technology Press; 2015: 148–149

[4] Semwal DK, Badoni R, Semwal R, Kothiyal SK, Singh GJP, Rawat U. The genus *Stephania* (Menispermaceae): chemical and pharmacological perspectives. J Ethnopharmacol 2010; 132: 369–383

[5] Liu JQ, Zhang YN, Zhao W, Kuang HX, Wang QH. Research progress on chemical constituents and pharmacology of *Stephania tetrandra*. Acta Chin Med Pharmacol 2017; 45: 100–103

[6] Yang M, Zhu LP, Li L, Li JJ, Xu LM, Feng J, Liu YL. Digital gene expression analysis provides insight into the transcript profile of the genes involved in aporphine alkaloid biosynthesis in lotus (*Nelumbo nucifera*). Front Plant Sci 2017; 8: 80

[7] Kraus PFX, Kutchan TM. Molecular cloning and heterologous expression of a cDNA encoding berbamunine synthase, a C–O phenol-coupling cytochrome P450 from the higher plant *Berberis stolonifera*. Proc Natl Acad Sci U S A 1995; 92: 2071–2075

[8] Yang CQ, Fang X, Wu XM, Mao YB, Wang LJ, Chen XY. Transcriptional regulation of plant secondary metabolism. J Integr Plant Biol 2012; 54: 703–712

[9] Metzker ML. Sequencing technologies-the next generation. Nat Rev Genet 2010; 11: 31–46

[10] Tang FC, Barbacioru C, Wang YZ, Nordman E, Lee C, Xu NL, Wang XH, Bodeau J, Tuch BB, Siddiqui A, Lao KQ, Surani MZ. mRNA-Seq whole-transcriptome analysis of a single cell. Nat Methods 2009; 6: 377–382

[11] Zhang H, He L, Cai L. Transcriptome sequencing: RNA-Seq. Methods Mol Biol 2018; 1754: 15–27

[12] Matsuda F, Hirai MY, Sasaki E, Akiyama K, Yonekura-Sakakibara K, Provart NJ, Sakurai T, Shimada Y, Saito K. AtMetExpress development: a phytochemical atlas of Arabidopsis development. Plant Physiol 2010; 152: 566–578

[13] Saito K, Hirai MY, Yonekura-Sakakibara K. Decoding genes with coexpression networks and metabolomics – 'majority report by precogs'. Trends Plant Sci 2008; 13: 36–43

[14] Rai A, Nakaya T, Shimizu Y, Rai M, Nakamura M, Suzuki H, Saito K, Yamazaki M. *De novo* transcriptome assembly and characterization of *Lithospermum officinale* to discover putative genes involved in specialized metabolites biosynthesis. Planta Med 2018; 84: 920–934

[15] Li XN, Yan HX, Sha N, Hua HM, Wu LJ, Guo DA. Isolation and identification of alkaloids from the root of *Stephania tetrandra*. J Shenyang Pharm Univ 2009; 26: 430–433

[16] Liu RM, He FY, Sun AL. Separation and identification of alkaloids in *Stephania tetrandra* by capillary electrophoresis electrospray – MS/MS. Acta Pharm Sin 2005; 39: 45–48

[17] Liu XB, Liu YS, Huang P, Ma YS, Qing ZX, Tang Q, Cao HF, Cheng P, Zheng YJ, Yuan ZJ, Zhou Y, Liu JF, Tang ZS, Zhuo YX, Zhang YC, Yu LL, Huang JL, Yang P, Peng Q, Zhang JB, Jiang WK, Zhang ZH, Lin K, Ro DK, Chen XY, Xiong XY, Shang Y, Huang SW, Zeng JG. The genome of the medicinal plant *Macleaya cordata* provides new insights into benzylisoquinoline alkaloids metabolism. Mol Plant 2017; 10: 975–989

[18] Ikezawa N, Iwasa K, Sato F. Molecular cloning and characterization of CYP80G2, a cytochrome P450 that catalyzes an intramolecular C-C phenol coupling of (*S*)-reticuline in magnoflorine biosynthesis, from cultured *Coptis japonica* Cells. J Biol Chem 2008; 283: 8810–8821

[19] Liscombe DK, Facchini PJ. Evolutionary and cellular webs in benzylisoquinoline alkaloid biosynthesis. Curr Opin Biotechnol 2008; 19: 173–180

[20] Ziegler J, Facchini PJ. Alkaloid biosynthesis: metabolism and trafficking. Annu Rev Plant Biol 2008; 59: 735–769

[21] Diamond A, Desgagné-Penix I. Metabolic engineering for the production of plant isoquinoline alkaloids. Plant Biotechnol J 2016; 14: 1319–1328

[22] Glenn WS, Runguphan W, O'Connor SE. Recent progress in the metabolic engineering of alkaloids in plant systems. Curr Opin Biotechnol 2013; 24: 354–365

[23] Hagel JM, Facchini PJ. Benzylisoquinoline alkaloid metabolism: a century of discovery and a brave new world. Plant Cell Physiol 2013; 54: 647–672

[24] Minami H, Dubouzet E, Iwasa K, Sato F. Functional analysis of norcoclaurine synthase in *Coptis japonica*. J Biol Chem 2007; 282: 6274–6282

[25] Hagel JM, Morris JS, Lee EJ, Desgagné-Penix I, Bross CD, Chang LM, Chen X, Farrow SC, Zhang Y, Soh J, Sensen CW, Facchini PJ. Transcriptome analysis of 20 taxonomically related benzylisoquinoline alkaloid-producing plants. BMC Plant Biol 2015; 15: 227

[26] Deng XB, Zhao L, Fang T, Xiong YQ, Ogutu C, Yang D, Vimolmangkang S, Liu YL, Han YP. Investigation of benzylisoquinoline alkaloid biosynthetic pathway and its transcriptional regulation in lotus. Hortic Res 2018; 5: 29

[27] Kato N, Dubouzet E, Kokabu Y, Yoshida S, Taniguchi Y, Dubouzet JG, Yazaki K, Sato F. Identification of a WRKY protein as a transcriptional regulator of benzylisoquinoline alkaloid biosynthesis in *Coptis japonica*. Plant Cell Physiol 2007; 48: 8–18

[28] Yamada Y, Kokabu Y, Chaki K, Yoshimoto T, Ohgaki M, Yoshida S, Kato N, Koyama T, Sato F. Isoquinoline alkaloid biosynthesis is regulated by a unique bHLH-type transcription factor in *Coptis japonica*. Plant Cell Physiol 2011; 52: 1131–1141

[29] Apuya NR, Park JH, Zhang L, Ahyow M, Davidow P, Fleet JV, Rarang JC, Hippley M, Johnson TW, Yoo HD, Trieu A, Krueger S, Wu CY, Lu YP, Flavell RB, Bobzin SC. Enhancement of alkaloid production in opium and California poppy by transactivation using heterologous regulatory factors. Plant Biotechnol J 2008; 6: 160–175

[30] Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng QD, Chen ZH, Mauceli E, Hacohen N, Gnirke A, Rhind N, Palma F, Birren BW, Nusbaum C, Lindblad-Toh K, Friedman N, Regev A. Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat Biotechnol 2011; 29: 644–652

[31] Conesa A, Gotz S, Garcia-Gomez JM, Terol J, Talon M, Robles M. Blast2-GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics 2005; 21: 3674–3676

[32] Kanehisa M, Goto S. KEGG: Kyoto Encyclopedia of Genes and Genomes. Nucleic Acids Res 2000; 28: 27–30

[33] Li B, Dewey CN. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics 2011; 12: 323

[34] Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 2010; 26: 139–140

[35] Rice P, Longden I, Bleasby A. EMBOSS: the European molecular biology open software suite. Trends Genet 2000; 16: 276–277

[36] Mistry J, Finn RD, Eddy SR, Bateman A, Punta M. Challenges in homology search: HMMER3 and convergent evolution of coiled-coil regions. Nucleic Acids Res 2013; 41: e121

[37] Pfaffl MW. A new mathematical model for relative quantification in real-time RT-PCR. Nucleic Acids Res 2001; 29: 2002–2007