

RESEARCH ARTICLE

# *De novo* characterization of the *Baphicacanthus cusia*(Nees) Bremek transcriptome and analysis of candidate genes involved in indican biosynthesis and metabolism

Wenjin Lin<sup>1,2</sup>, Wei Huang<sup>1</sup>, Shuju Ning<sup>3</sup>, Xiaohua Wang<sup>1</sup>, Qi Ye<sup>1</sup>, Daozhi Wei<sup>1\*</sup>

**1** School of Life science, Fujian Agriculture and Forestry University, Fuzhou, China, **2** Fujian Key Laboratory of Medical Measurement, Fujian Academy of Medical Sciences, Fuzhou, China, **3** School of Crop Science, Fujian Agriculture and Forestry University, Fuzhou, China

\* [weidz888@sohu.com](mailto:weidz888@sohu.com)



**OPEN ACCESS**

**Citation:** Lin W, Huang W, Ning S, Wang X, Ye Q, Wei D (2018) *De novo* characterization of the *Baphicacanthus cusia*(Nees) Bremek transcriptome and analysis of candidate genes involved in indican biosynthesis and metabolism. PLoS ONE 13(7): e0199788. <https://doi.org/10.1371/journal.pone.0199788>

**Editor:** Xiang Jia Min, Youngstown State University, UNITED STATES

**Received:** December 21, 2017

**Accepted:** June 13, 2018

**Published:** July 5, 2018

**Copyright:** ©2018 Lin et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** All data generated or analysed during this study are available in the NCBI SRA repository: SRA628524 (SRP124081 : PRJNA415260), [https://trace.ncbi.nlm.nih.gov/Traces/sra\\_sub/sub.cgi?acc=SRP124081](https://trace.ncbi.nlm.nih.gov/Traces/sra_sub/sub.cgi?acc=SRP124081). All other relevant data are included in this published article and its Supporting Information files.

**Funding:** This work was funded by 81573517, National Natural Science Foundation of China, <https://isn.nsf.gov.cn/>; and 2017, high-level

## Abstract

*Baphicacanthus cusia* (Nees) Bremek is an herb widely used for the clinical treatment of colds, fever, and influenza in Traditional Chinese Medicine. The roots, stems and leaves can be used as natural medicine, in which indigo and indirubin are two main active ingredients. In this study, quantification of indigo, indirubin, indican and adenosine among various tissues of *B. cusia* was conducted using HPLC-DAD. Leaves have significantly higher contents than stems and roots (380.66, 315.15, 20,978.26, 4323.15 µg/g in leaves, 306.36, 71.71, 3,056.78, 139.45 µg/g in stems, and 9.31, 7.82, 170.45, 197.48 µg/g in roots, respectively). *De novo* transcriptome sequencing of *B. cusia* was performed for the first time. The sequencing yielded 137,216,248, 122,837,394 and 140,240,688 clean reads from leaves, stems and roots respectively, which were assembled into 51,381 unique sequences. A total of 33,317 unigenes could be annotated using the databases of Nr, Swiss-Prot, KEGG and KOG. These analyses provided a detailed view of the enzymes involved in indican backbone biosynthesis, such as cytochrome P450, UDP-glycosyltransferase, glucosidase and tryptophan synthase. Analysis results showed that tryptophan synthase was the candidate gene involved in the tissue-specific biosynthesis of indican. We also detected sixteen types of simple sequence repeats in RNA-Seq data for use in future molecular mark assisted breeding studies. The results will be helpful in further analysis of *B. cusia* functional genomics, especially in increasing biosynthesis of indican through biotechnological approaches and metabolic regulation.

## Introduction

*Baphicacanthus cusia* (Nees) Bremek (*B. cusia*) also named *Strobilanthes cusia* is a perennial medicinal herb distributed broadly in Fujian, Zhejiang, Guangdong, Hainan, Guangxi, Hong

hospital grants from Fujian Provincial Hospital, Fujian province, China, <http://www.fjssl.com.cn/>.

**Competing interests:** The authors have declared that no competing interests exist.

Kong, Taiwan and other southern areas of China. It is also distributed across Bangladesh, India, and Myanmar, from the Himalayas to Indo-China [1]. It has been cultivated in Fujian Province for more than 1000 years since the Song Dynasty. Indigo Naturalis is the leaf and stem extract of *B. cusia*, whose commercial name is “Qingdai” [2], and Baphicacanthus cusiae Rhizome et Radix is derived from its root [3], with the commercial name of “Nanbanlangen”. Previous studies have shown that Qingdai as an active ingredient was used to treat dental ulcers [4], ulcerative colitis [5–7], and psoriasis [8–10], and that Nanbanlangen has many pharmacological applications, including antibacterial [11], antiviral [12], and anti-inflammatory effects [13, 14].

Chemical research has indicated that Qingdai and Nanbanlangen both contain many active ingredients, such as indigo [15] and indirubin [16]. Indican has been known to be a precursor of indigo and indirubin [17–19], the content of which was different in the leaf and root of *B. cusia* [15, 20]. To better understand the biosynthesis of indigo, gene expression possibly involved in indigo biosynthesis have been analysed by RNA-Seq in *Indigofera tinctoria* and *Polygonum tinctorium* [21], species closely related to *B. cusia*. However, these organisms belong to different genera and species classifications and differ in morphological traits.

The Qingdai produced in Fujian Province of China, named Jianqingdai, is a well-known genuine regional drug in China because of its good quality, but its effective constituent biosynthesis mechanism *in vivo* is still unclear. Until now, only nine proteins have been identified in the *B. cusia* leaf [22, 23], and 98 nucleotide sequences are available for *Baphicacanthus cusia*, but none of the nucleotide sequences of *B. cusia* can be found in NCBI GenBank, it is hard to clarify its molecular mechanisms due to the lack of information of genomic sequences, especially sequences of indican biosynthesis and metabolism genes. Therefore, our primary work is to obtain and characterize the transcriptome of *B. cusia*.

RNA-Seq next-generation sequencing (NGS) has been widely used in the analysis of transcriptomes in crops or vegetables, such as rice [24], barley [25], *Sorghum bicolor* [26], *Withania somnifera* [27], *Brassica rapa* [28] and Chinese cherry [29]. It has also been used to illustrate active components of biosynthetic and metabolic pathways in traditional Chinese medicine or identify differentially expressed genes (DEGs) between distinct sample groups such as *Isatis indigotica* [30–34], *Salvia miltiorrhiza* [35, 36], *Taxillus chinensis* [37], *Lonicera japonica* [38], *Toona sinensis* Roem [39], *Pseudostellariae radix* [40], *Eucommia ulmoides* Oliver [41], *Cistanche deserticola* [42], *Lycium chinense* Mill [43] and *Gentiana macrophylla* [44].

In this study, we measured the amounts of indigo, indirubin, adenosine and indican in the leaves, stems and roots of *B. cusia*. To understand the biosynthesis and metabolism pathways of indican in root, stem and leaf tissues of *B. cusia*, we generated multi-tissue transcriptomic data of *B. cusia* using RNA-Seq and annotated the multi-tissue transcriptome using publicly available databases and tools; we then ascertained genes that may be related to indican biosynthesis and metabolism pathways. Here, we report simple sequence repeats (SSR) in leaves, stems and roots from *B. cusia* and a number of differentially expressed genes (DEGs), particularly, cytochrome P450 (CYP450), UDP-glycosyltransferase (UGT), glucosidase and tryptophan synthase, that may be related to differential expression in indican biosynthesis and metabolism among leaves, stems and roots. Based on our results, the functional characterization of the *B. cusia* genes mentioned above should be further investigated.

## Results

### Quantitative analysis with HPLC

The calibration curves of indigo, indirubin, adenosine and indican were formulated with five different concentration references. The regression equation was  $Y = aX + b$ , in which X and Y are the concentration and peak area, respectively. The regression equations and correlation

coefficients ( $r^2$ ) of indigo, indirubin, adenosine and indican were as follows:  $Y = 25561X + 283.01$  (linear range from 0.188 to 23.5  $\mu\text{g}\cdot\text{ml}^{-1}$ ),  $r^2 = 1.0000$  for indigo;  $Y = 39667X - 1296.9$  (linear range from 0.092 to 11.5  $\mu\text{g}\cdot\text{ml}^{-1}$ ),  $r^2 = 1.0000$  for indirubin,  $Y = 15817X + 28642$  (linear range from 1.1 to 275  $\mu\text{g}\cdot\text{ml}^{-1}$ ),  $r^2 = 0.9993$  for adenosine, and  $Y = 6020.5X - 5001$  (linear range from 1.08 to 270  $\mu\text{g}\cdot\text{ml}^{-1}$ ),  $r^2 = 1.0000$  for indican. The relative quantity of four ingredients in the leaf, stem and root samples ( $\mu\text{g/g}$ ) was calculated using the above equations (Fig 1).

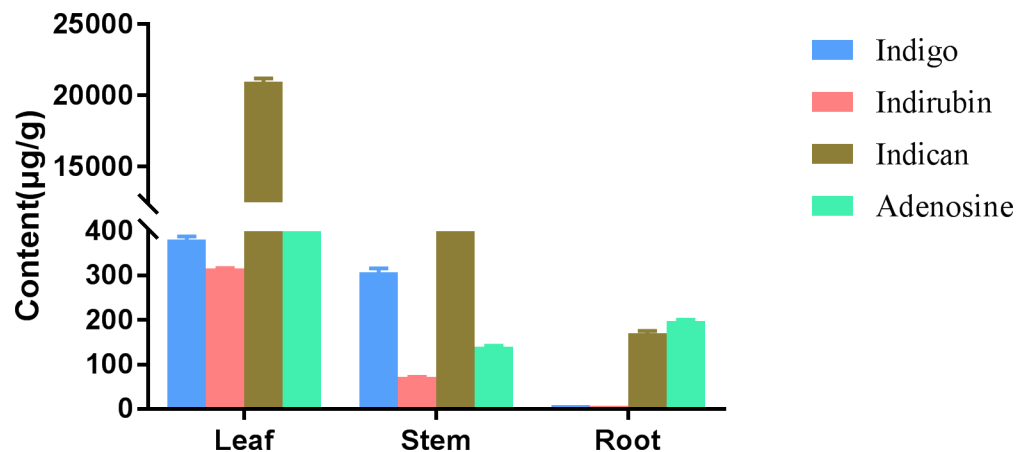
### RNA sequencing and de novo assembly

After total RNA was extracted, mRNA isolated from leaves, stems and roots of *B. cusia* (three-year-old plants cultivated) was enriched by oligo (dT) beads in order to acquire a general set of *B. cusia* transcripts. After removing reads containing adaptors or low-quality paired-end reads, clean and high-quality reads were analysed. A total of 140,240,688 high quality reads were obtained for root, 137,216,248 for leaf and 122,837,394 for stem. Contig and unigene de novo assembly was carried out with the short read assembling program Trinity using the high-quality reads. 63,327,716 bases, 51,381 unigenes and 1,932 bp N50, and average length 1,232 bp were assembled. The GC percentage was 42.4176.

### Annotation of the unigenes

51,381 unigenes were widely searched against various databases in order to determine the potential functions of the assembled unigenes. To annotate the unigenes, we used the BLASTx program with an E-value threshold of  $10^{-5}$  and a  $k$ -mer of 25 with the NCBI non-redundant protein (Nr) database, the Swiss-Prot protein database, the Kyoto Encyclopaedia of Genes and Genomes (KEGG) database, and the COG/KOG database. Unigene functional annotations were obtained according to the best alignment results. The results showed that 32,898, 25,809, 20,751 and 13,232 unigenes matched with the Nr, SwissProt, KOG and KEGG databases, respectively. 33,317 unigenes matched with one of the above databases (Fig 2).

GO functional annotations of *B.cusia* unigenes were obtained from Nr annotation results. Based on their GO terms, the unigenes could be categorized into three main GO groups and



**Fig 1. Contents of four ingredients in leaves, stems and roots of *B. cusia*.** The experiment was repeated three times. The amounts of indigo, indirubin, indican and adenosine in leaf tissue (380.66  $\mu\text{g/g}$ , 315.15  $\mu\text{g/g}$ , 20,978.26  $\mu\text{g/g}$  and 1224.12  $\mu\text{g/g}$ , respectively) were higher than those in the stem tissue (306.36  $\mu\text{g/g}$ , 71.71  $\mu\text{g/g}$ , 3056.78  $\mu\text{g/g}$  and 139.45  $\mu\text{g/g}$ , respectively,  $p = 0.000$ ), the amounts of indigo, indirubin and indican in stem tissue were higher than those in the root tissue (9.31  $\mu\text{g/g}$ , 7.82  $\mu\text{g/g}$  and 170.45  $\mu\text{g/g}$ , respectively,  $p = 0.000$ ), and the amount of adenosine in stem tissue was lower than that in root tissue (197.48  $\mu\text{g/g}$ ), but the difference was not significant ( $p = 0.350$ ).

<https://doi.org/10.1371/journal.pone.0199788.g001>

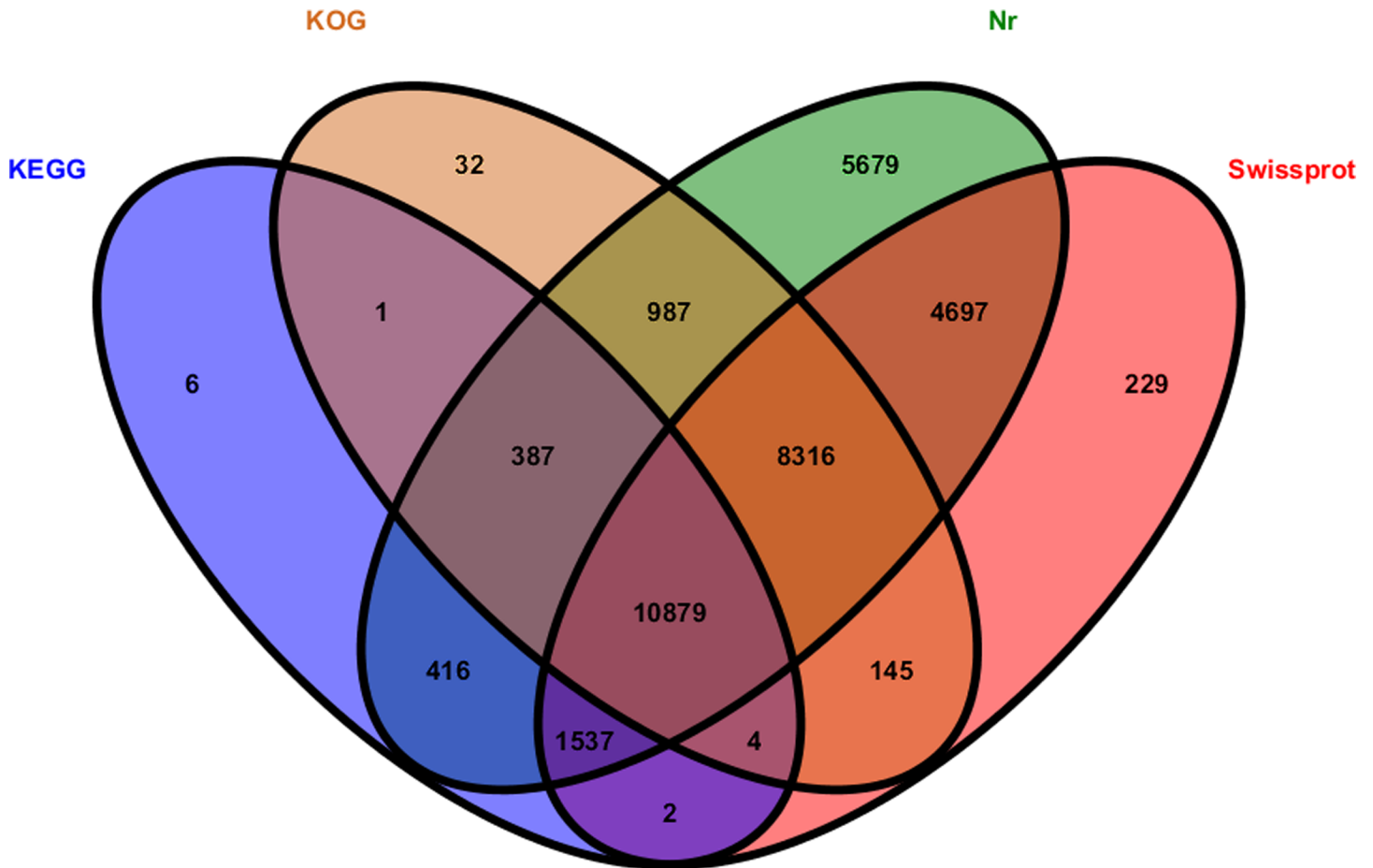


Fig 2. Venn diagram of unigene number annotated by BLASTx with the Nr, SwissProt, KOG and KEGG databases.

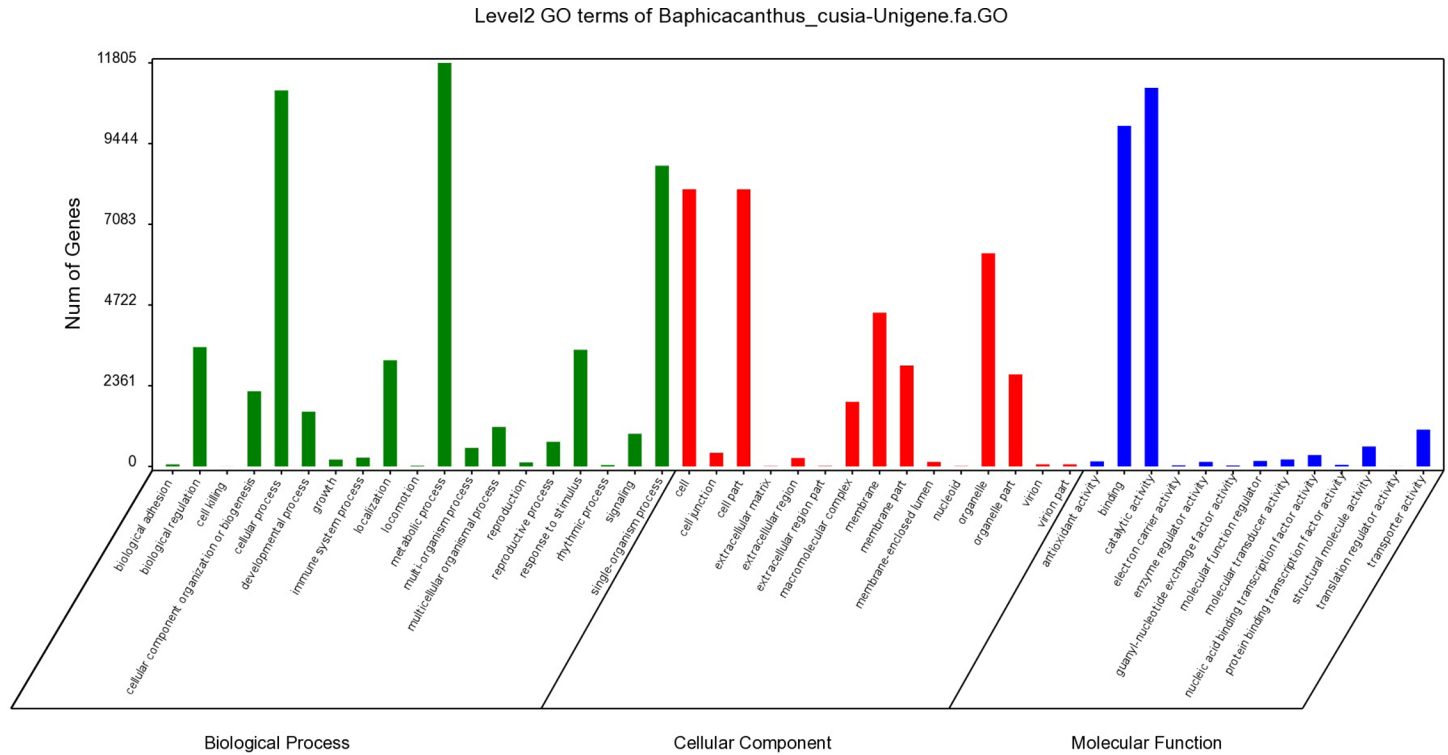
<https://doi.org/10.1371/journal.pone.0199788.g002>

47 functional sub-groups (Fig 3). The three most frequent GO terms in biological process were metabolic processes (11,803 unigenes), cellular processes (10,998 unigenes), and single-organism processes (8,797 unigenes). The three most frequent GO terms in cellular components were cells (8,107 unigenes), cell parts (8,106 unigenes), and organelles (6,236 unigenes). The three most frequent GO terms in molecular function groups were catalytic activity (11,075 unigenes), binding (9,963 unigenes), and transporter activity (1,076 unigenes).

35,671 of the total 51,381 *B. cusia* unigenes from analysis of EuKaryotic Orthologous Groups (KOG) could be classified into 25 clusters. Among the classified unigenes, the most part of the cluster was “general function prediction” (6,624; 18.57%), which was followed by clusters of “signal transduction mechanisms” (4,143; 11.61%), “posttranslational modification, protein turnover, chaperones” (4,042; 11.33%), “transcription” (2,009; 5.43%), “RNA processing and modification” (1,830; 5.13%) and intracellular trafficking, secretion, and vesicular transport (1,791; 5.02%). The cluster of the least number was “cell motility” (14; 0.04%) (Fig 4).

### KEGG pathway mapping

To better understand the biological functions and pathways represented by the unigenes assembled in this study, 7,350 unigenes were classified into 127 pathways. The pathways included metabolism, genetic information processing, cellular processes, environmental



**Fig 3. Histogram of level 2 GO terms of the *B. cusia* unigenes.**

<https://doi.org/10.1371/journal.pone.0199788.g003>

information processing, and organismal systems. The top 20 pathways, which have predominant sequences, are indicated in Fig 5 (S1–S3 Tables), and the most represented pathways which were enriched in the DEG sets compared to the total transcriptome were plant hormone signal transduction and phenylpropanoid biosynthesis. Rich factors were 0.57, 0.40, 0.41 for plant hormone signal transduction (Q-value 3.36E-10, 3.92E-03, 7.03E-13) and 0.61, 0.49, 0.52 for phenylpropanoid biosynthesis (Q-value 1.20E-9, 3.66E-06, 3.70E-18) between root and leaf, root and stem, and stem and leaf, respectively. The main purpose of cultivating *B. cusia* has been to obtain the Indigo Naturalis, which has active ingredients of indole alkaloids and their derivatives. In this study, the annotated results suggest 63 unigenes related to nitrogen metabolism (S4 Table) and 58 unigenes connected to phenylalanine, tyrosine and tryptophan biosynthesis (biosynthesis of indole and its derivatives) (S5 Table).

### Identification of simple sequence repeats

To find SSR markers, MISA software was employed for microsatellite mining in the whole *B. cusia* transcriptome. Based on the MISA results, three pairs of PCR primers have been designed by Primer 3 in the flanking regions of SSRs (S6 Table). Within the 51,381 examined unigenes, 6,782 unigenes were identified containing 8,471 SSR markers, and 1,350 unigenes contained more than one SSR. Tri-nucleotides were the most abundant SSR (3,046; 35.96%), followed by di-nucleotides (3,021; 35.66%), tetra-nucleotides (1,065; 12.57%), hexa-nucleotides (861; 10.16%), and penta-nucleotides (478; 5.64%) were the least frequent SSR (S7 Table). Among the confirmed SSRs, 547 showed a compound formation. There were 16 types of SSR identified in *B. cusia* (Fig 6). Until now, there have been no publicly available SSRs in *B. cusia*, and thus the SSRs identified in this study will be extremely helpful for research on genotyping applications in *B. cusia*.

### KOG Function Classification of *Baphicacanthus\_cusia*-Unigene.fa Sequence

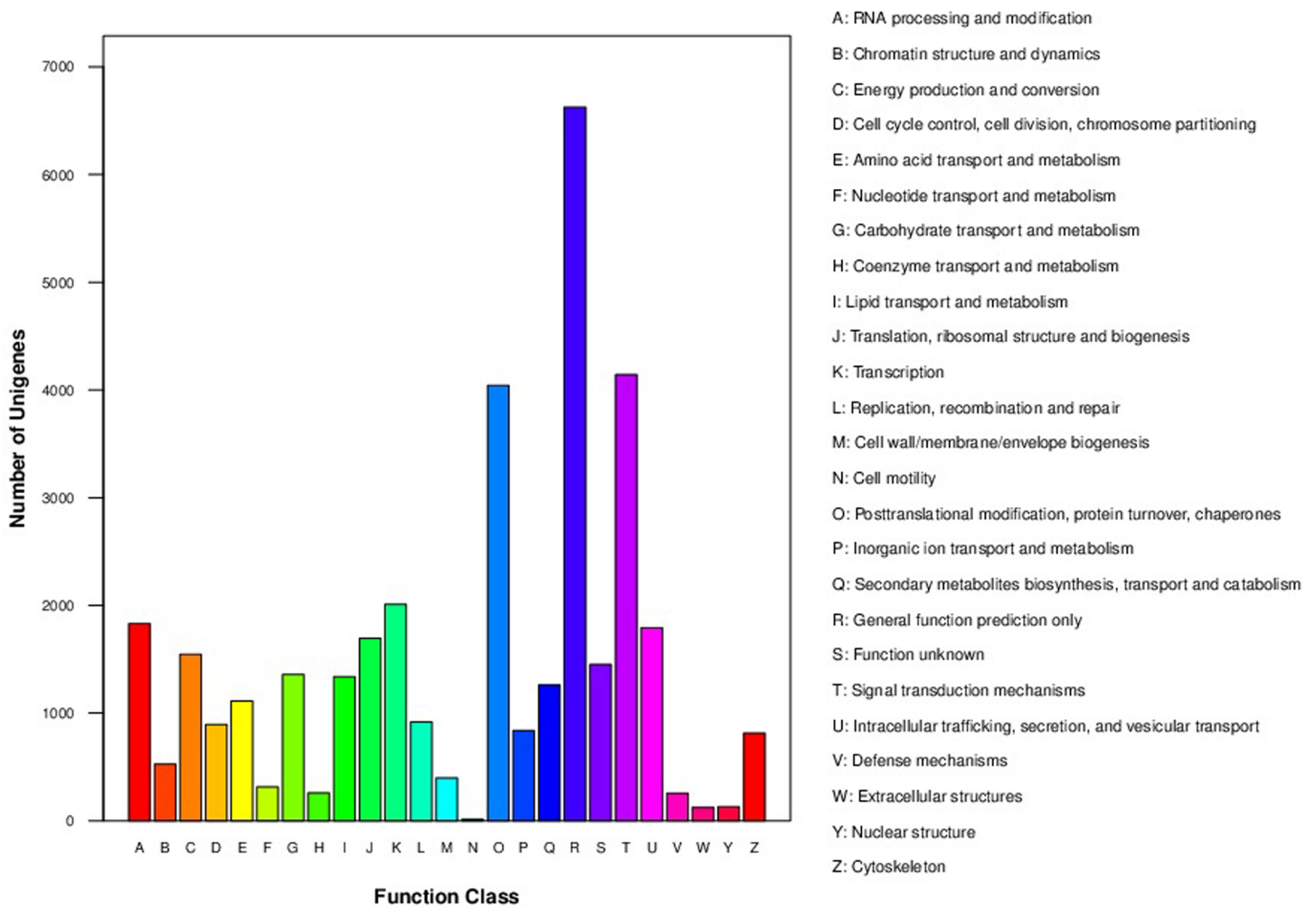


Fig 4. Histogram of KOG function classification of *B. cusia* unigenes.

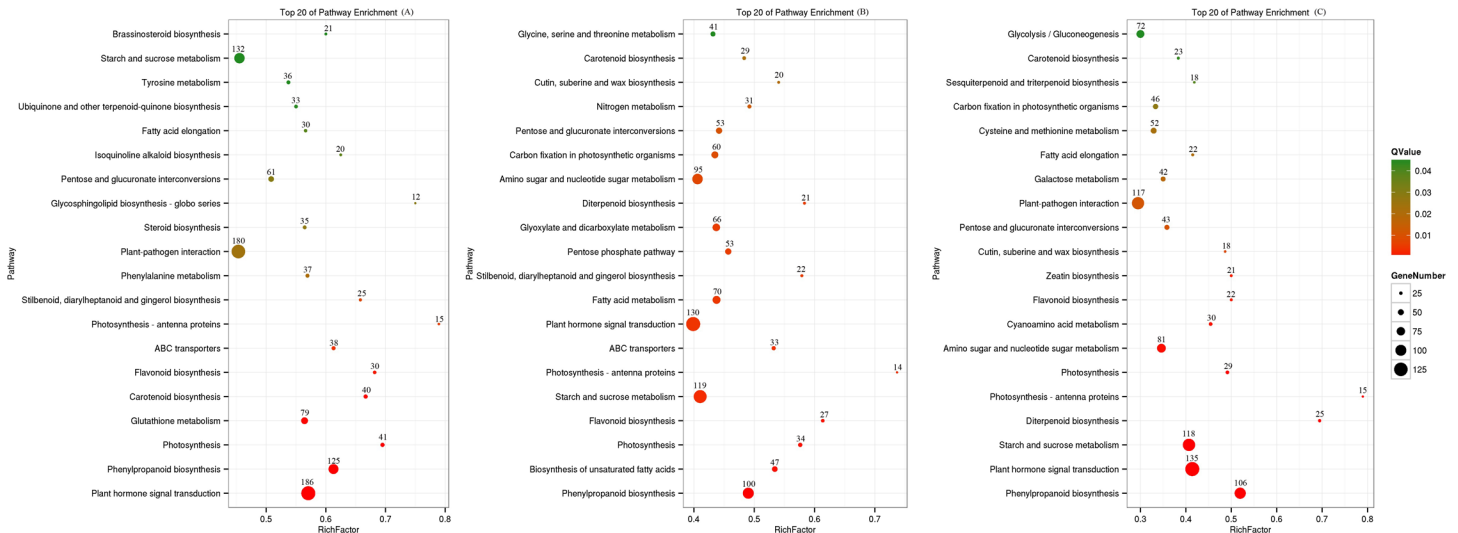
<https://doi.org/10.1371/journal.pone.0199788.g004>

### Analysis of differentially expressed genes

There were 13,836 genes expressed differentially between roots and stems. Among these differentially expressed unigenes, 6,633 were up-regulated and 7,203 were down-regulated in roots compared to stems. There were 20,083 unigenes differentially expressed between roots and leaves; among these differentially expressed unigenes, 5,383 were up-regulated and 14,700 were down-regulated in roots compared to leaves. There were 11,532 unigenes differentially expressed between stems and leaves; among these differentially expressed unigenes, 2,491 were up-regulated and 9,041 were down-regulated in stems compared to leaves.

In light of the putative biosynthetic pathways of indole derivatives in *Isatis indigotica* [34], analyse of the differentially expressed genes among various tissues were carried out for biosynthesis of indican, cytochrome P450 (CYP450) [34], UDP-glycosyltransferase (UGT) [34], glucosidase [45] and tryptophan synthase; these were chosen to be analysed after annotation.

There were 121 CYP450, 34 UGT, 57 glucosidase, and three tryptophan synthase DEGs that were differentially expressed between roots and stems (S8 Table). There were 154 CYP450, 52 UGTs, 71 glucosidase, four tryptophan synthase DEGs between roots and leaves (S9 Table).



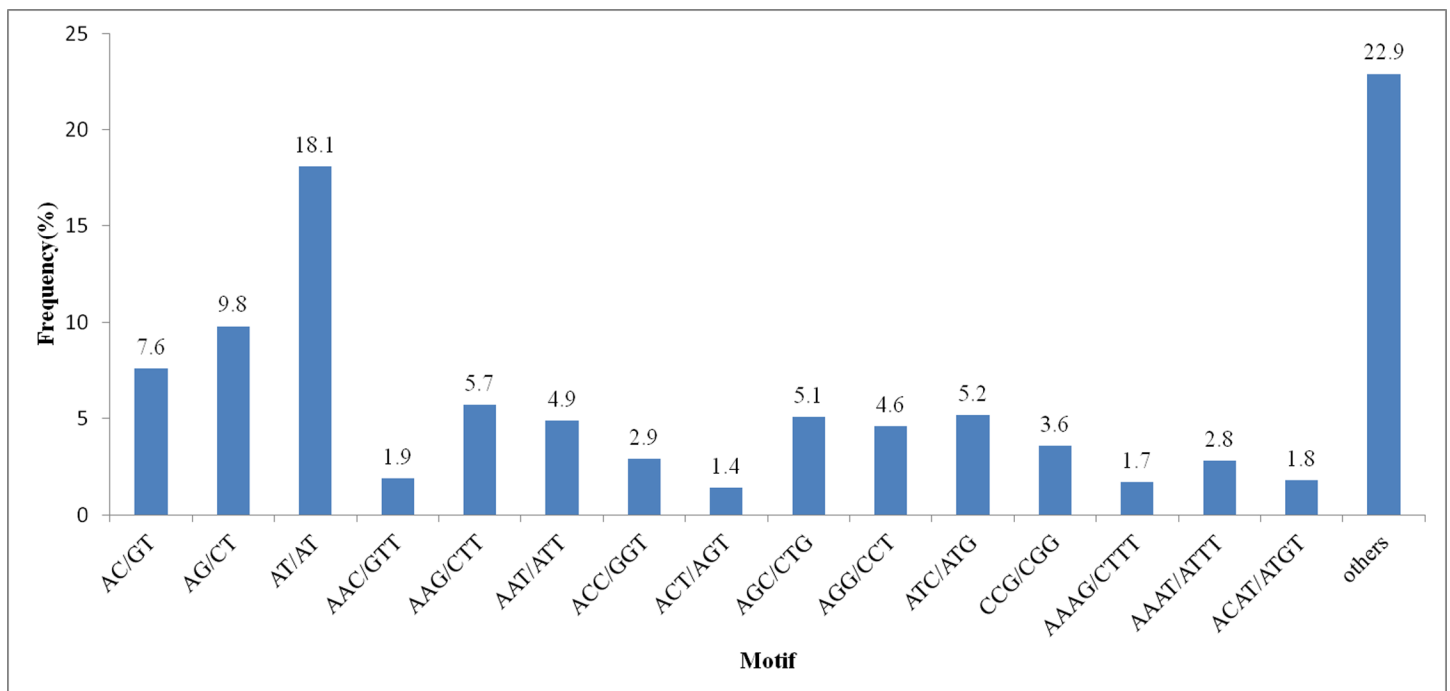
**Fig 5. Top 20 pathways enriched among various tissues of *B. cusia*.** (A) The top 20 pathways enriched between root and leaf; (B) the top 20 pathways enriched between root and stem; (C) the top 20 pathways enriched between stem and leaf. The top 20 pathways enriched from top to bottom are arranged by decreasing Q-value. Data over the dots show the actual gene number.

<https://doi.org/10.1371/journal.pone.0199788.g005>

There were 75 CYP450, 39 UGT, 58 glucosidase, and one tryptophan synthase DEG between stems and leaves (S10 Table and Fig 7); the DEGs of stems were more similar to leaves than those of roots.

### Validation of DEG data by qRT-PCR

We performed qRT-PCR on 16 representative unigenes related to the biosynthesis and metabolism of indican to verify the DEGs acquired by Illumina RNA-Seq. We found that



**Fig 6. Distribution of various types of SSRs identified in *B. cusia* unigenes.**

<https://doi.org/10.1371/journal.pone.0199788.g006>

while three DEGs (two CYP450 unigenes and one glucosidase unigene) did not show consistent expression between qRT-PCR and RNA-Seq data sets, most of the DEGs from qRT-PCR exhibited similarity to RNA-Seq samples using qRT-PCR (Fig 8 and S11 Table). Compared with leaves, tryptophan synthase expression was down-regulated in roots and stems, which agreed with the results of quantitative determination of indican among various tissues of *B. cusia*.

### Discussion

In this study, the amounts of four active ingredients, such as indigo, indirubin, indican and adenosine, were found to differ substantially in root, stem and leaf tissues from *B. cusia*. The amounts of indigo, indirubin, indican and adenosine in leaf tissue were 380.66µg/g, 315.15 µg/g, 20,978.26 µg/g and 1224.12 µg/g; these components in stem tissue were 306.36 µg/g, 71.71µg/g, 3056.78 µg/g and 139.45 µg/g, in root tissue were 9.31 µg/g, 7.82 µg/g, 170.45 µg/g

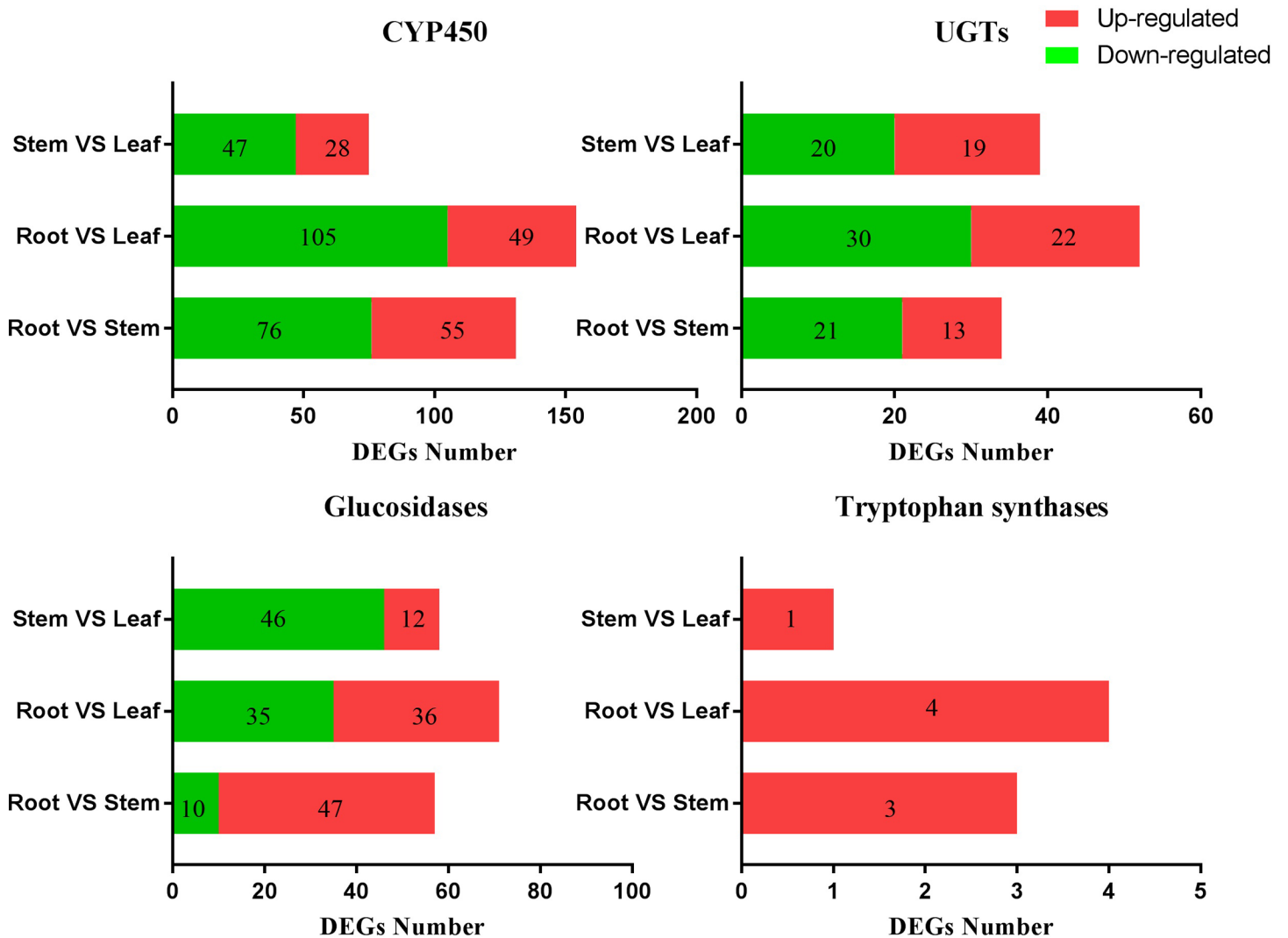
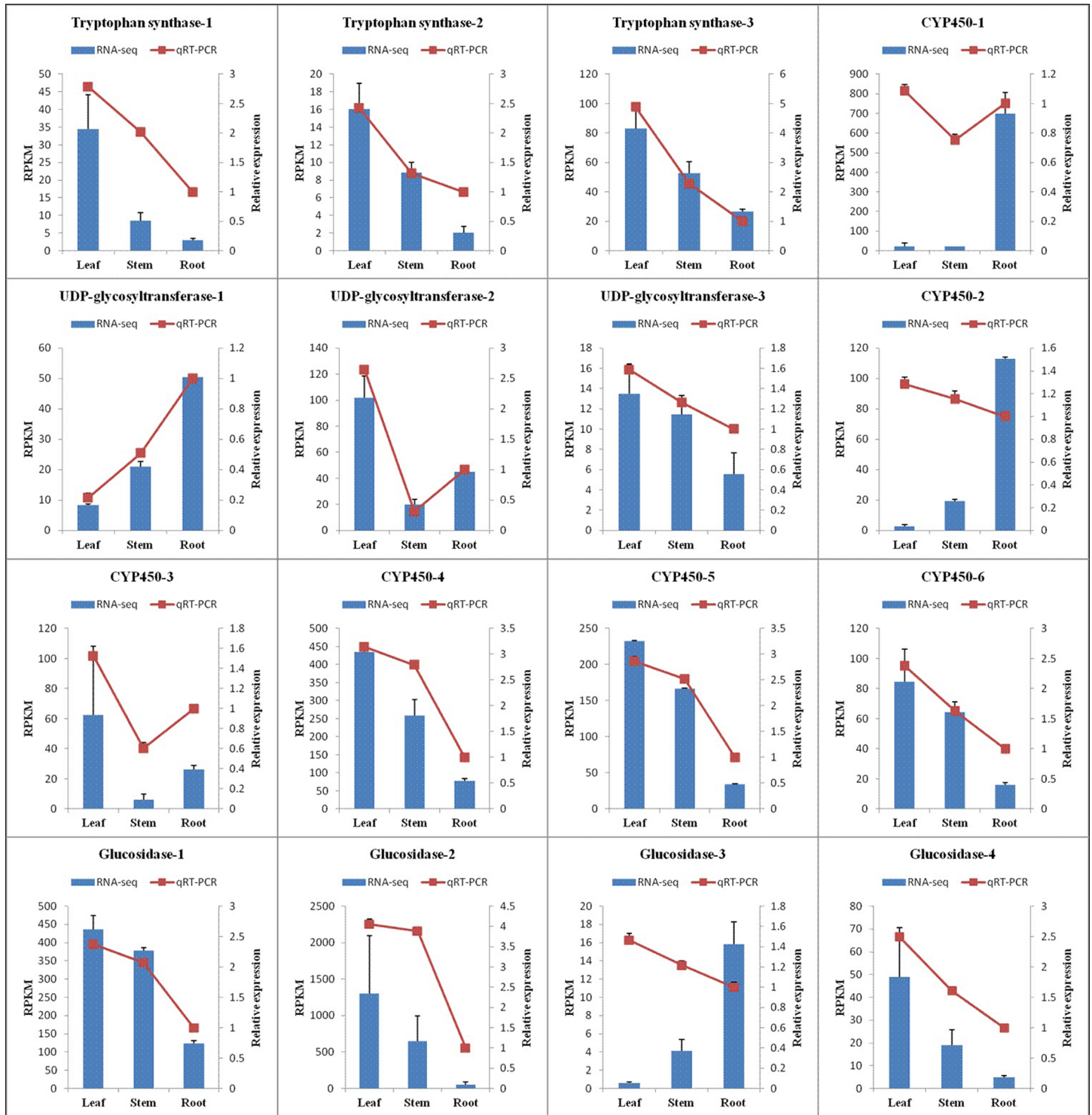


Fig 7. Annotated CYP450, UGT, glucosidase, and tryptophan synthase DEG numbers among various tissues of *B. cusia*. The data in green bars show the down-regulated gene number and the data in red bars show the up-regulated gene number.

<https://doi.org/10.1371/journal.pone.0199788.g007>





**Fig 8. qRT-PCR validation of DEGs identified by RNA-Seq.** Sixteen DEGs involved in the biosynthesis and metabolism of indican were randomly selected for qRT-PCR confirmation. The blue bars and red lines represent RNA-Seq RPKM data and qRT-PCR relative expression levels, respectively. The beta-actin gene was used as an internal control, and every experiment was repeated with three biological samples. All the values shown are the means  $\pm$  SE.

<https://doi.org/10.1371/journal.pone.0199788.g008>

and 197.48  $\mu\text{g/g}$  respectively. The amounts of indigo and indirubin in leaf tissue were higher than those in the root tissue ( $p < 0.01$ ), which was consistent with a previous report [20]. This is the first study of the amounts of adenosine and indican simultaneously in the leaf and root tissues of *B. cusia*. Our work provides novel information with respect to the mechanism of different tissues of *B. cusia* used as natural medicine, since this may be related to the distribution of indican and the metabolites indigo and indirubin.

A total of 63 unigenes related to nitrogen metabolism pathways and 58 unigenes involved in the biosynthesis pathway of indole and its derivatives were found. Further studies of these unigenes will help to understand regulation of secondary metabolism in *B. cusia*. In addition, 154 CYP450, 52 UGT, 71 glucosidase and four tryptophan synthase genes were differentially expressed between leaves and roots, 75 CYP450, 39 UGT, 58 glucosidase and one tryptophan synthase gene were differentially expressed between leaves and stems, 121 CYP450, 34 UGT, 57 glucosidase and three tryptophan synthase genes were differentially expressed between stems and roots. Tryptophan synthase DEGs may be the candidate genes contributing to the production of indigo and indirubin in *B. cusia* leaves.

In addition, there were 125 DEGs involved in phenylpropanoid biosynthesis, 30 in flavonoid biosynthesis, and 35 in the terpenoid backbone biosynthesis pathway between roots and leaves; there were 106 DEGs involved in phenylpropanoid biosynthesis, 22 in flavonoid biosynthesis, and 23 in the terpenoid backbone biosynthesis pathway between stems and leaves, and there were 100 DEGs involved in phenylpropanoid biosynthesis, 27 in flavonoid biosynthesis, and 25 in the terpenoid backbone biosynthesis pathway between roots and stems. The investigation of these unigenes may shed light on the molecular mechanisms of secondary metabolism in *B. cusia*.

In summary, there were 51,381 unigenes from the leaf, stem and root transcriptomes of *B. cusia* using RNA sequencing in this study, but only 64.84% of the unigenes could be annotated by the Nr, SwissProt, KOG or KEGG databases. There were a great many unannotated unigenes and their biological pathways remain to be identified. Until now, this has been the first report of the transcriptome of *B. cusia*; these unigenes could be useful for further investigating genes related to abiotic stress in *B. cusia*, accelerating the research on the molecular mechanisms of secondary metabolism in stress responses and differential metabolic patterns of compounds such as indican, indigo and indirubin in different components of *B. cusia*.

## Materials and methods

### Ethics statement

All necessary permits were acquired for these studies. Samples were collected from our medicinal garden for scientific research at the Fujian Agricultural and Forestry University. We confirm that endangered or protected species were not contained in the species collections.

### Plant materials and RNA sequencing

Three batches of root, stem and leaf samples were collected from three-year-old plants cultivated in our medicinal garden, frozen in liquid nitrogen for 15 min immediately, and then stored at  $-80^{\circ}\text{C}$  refrigerator for future use. The EASYspin plant RNA kit (Aidlab Biotechnologies Co.Ltd, Beijing, China) was used to isolate total RNA, and contaminated DNA was removed by DNase I. An Agilent Bioanalyzer 2100 (Agilent Technologies) was used to analyse the quality and integrity of RNA. cDNA fragments were purified using the QiaQuick PCR extraction kit, end repaired, poly (A) was added, and then the fragments were ligated to

Illumina sequencing adapters. Sequenced were obtained using the Illumina HiSeq™ 4000 by Gene Denovo Biotechnology Co. (Guangzhou, China).

### Analysis with high performance liquid chromatography

Reverse-phase high performance liquid chromatography was performed on an LC-20AT HPLC system (Shimadzu, Japan), which included a LC-20AT pump, DGU-20A degasser, SIL-20A autosampler, SPD-20A detector and CBM-20A central controller, using a C<sub>18</sub> reverse phase chromatographic column (Inertsil ODS-SP, 4.6×250 mm, 5 μm) at 25 °C, with wavelengths monitored at 292 nm for indigo and indirubin and 260 nm for adenosine and indican. The mobile phase included methanol and water, with a volume ratio of 70: 30 for both indigo and indirubin and volume ratio of 15: 85 for adenosine and indican with a mobile phase flow rate of 1.0 ml/min. To prepare the sample and reference solutions, three batches of leaf, stem and root samples were extracted twice by a Soxhlet extractor. Indigo and indirubin mixed reference solutions, in triplicate, were dissolved in N, N-dimethyl formamide, and adenosine and indican mixed reference solutions, in triplicate, were dissolved in 30% methanol. After filtration through a 0.45 μm filter membrane, 10 μl of each sample and reference solution was injected in duplicate. The chromatographic peak retention times of indigo and indirubin and adenosine and indican were approximately 15 min and 20 min, respectively.

### Transcriptome *de novo* assembly and bioinformatic analysis

Raw reads generated from HiSeq 4000 were further filtered to remove adapters and low-quality sequences. The *de novo* assembling program Trinity consists of three components: Inchworm, Chrysalis and Butterfly; *de novo* assembly was generated with the clean reads. In brief, we first removed the wrong k-mers, constructed a k-mer library from the clean reads and then generated a collection of linear contigs by a greedy extension based on (k-1)-mer overlaps. Next, we clustered related contigs into components and built a de Bruijn graph for each cluster of related contigs. Finally, we simplified the de Bruijn graph, then unlinked the de Bruijn graph using real reads and reconstructed the transcript sequences.

All assembled unigenes were annotated by BLASTx with the NCBI non-redundant (Nr) protein database, Swiss-Prot, the Kyoto Encyclopaedia of Genes and Genomes (KEGG) database and EuKaryotic Orthologous Groups (KOG). Sequences with BLASTx hits were determined according to Gene Ontology (GO) terms by Blast2GO software. The functional classification of unigenes was performed by WEGO software. SSRs were scanned by MISA (MicroSatellite identification tool) software (<http://pgrc.ipk-gatersleben.de/misa/>) with the default software parameter settings. To study the differentially expressed genes, the log<sub>2</sub>-converted values of each contig were subjected to a Benjamini-Hochberg program. The differential expression of genes was counted with the FDR value and FDR fold change (fc). If the FDR value < 0.05 and |log<sub>2</sub>fc| > 1, then the gene was considered differential. Unigene expression was counted and normalized to RPKM (reads per kb per million reads) [46]. The formula is as follows:  $RPKM = (1000000 * C) / (N * L / 1000)$ , given RPKM (A) as the expression of Unigene A, C is the number of reads mapped to Unigene A uniquely, N is the total number of reads mapped to all unigenes uniquely, L is the base number length of Unigene A.

Pathway enrichment analysis identified significantly enriched metabolic pathways or signal transduction pathways in DEGs compared to the whole genome background. The formula to

calculate the  $P$ -value is:

$$P = 1 - \sum_{i=0}^{m-1} \frac{\binom{M}{i} \binom{N-M}{n-i}}{\binom{N}{n}} \quad (1)$$

Here,  $N$  is the number of all genes with KEGG annotation,  $n$  is the number of DEGs in  $N$ ,  $M$  is the number of all genes annotated to specific pathways, and  $m$  is the number of DEGs in  $M$ . The calculated  $p$ -value was determined through FDR correction, taking  $FDR \leq 0.05$  as a threshold. Pathways meeting this condition were defined as significantly enriched pathways in DEGs. Rich factor refers to the ratio of the number of genes located in the pathway entry in the differentially expressed genes to the total number of genes located in the pathway entry in all genes.  $Q$ -value is the  $P$ -value after multi-hypotheses test correction. The value range is 0 to 1.

## Supporting information

**S1 Table. Top 20 pathways enriched between root and leaf.**

(XLSX)

**S2 Table. Top 20 pathways enriched between root and stem.**

(XLSX)

**S3 Table. Top 20 pathways enriched between stem and leaf.**

(XLSX)

**S4 Table. KEGG pathway mapping and unigenes involved in nitrogen metabolism.**

(XLSX)

**S5 Table. KEGG pathway mapping and unigenes involved in phenylalanine, tyrosine and tryptophan biosynthesis.**

(XLSX)

**S6 Table. Primers for SSRs identified in the *B. cusia* unigenes.**

(TXT)

**S7 Table. Statistics of the SSRs identified in the *B. cusia* unigenes.**

(XLSX)

**S8 Table. Cytochrome P450, UDP-glycosyltransferase, glucosidase, tryptophan synthase DEGs between root and stem tissues of *B. cusia* unigenes.**

(XLSX)

**S9 Table. Cytochrome P450, UDP-glycosyltransferase, glucosidase, tryptophan synthase DEGs between root and leaf tissues of *B. cusia* unigenes.**

(XLSX)

**S10 Table. Cytochrome P450, UDP-glycosyltransferase, glucosidase, tryptophan synthase DEGs between stem and leaf tissues of *B. cusia* unigenes.**

(XLSX)

**S11 Table. Nucleotide sequences of primers specific to indican metabolic related genes used for qRT-PCR.**

(XLSX)

## Acknowledgments

We thank Gene Denovo Biotechnology Co. (Guangzhou, China) for technical assistance with the RNA sequencing and data analysis.

## Author Contributions

**Conceptualization:** Shuju Ning, Daozhi Wei.

**Data curation:** Wei Huang, Xiaohua Wang.

**Funding acquisition:** Daozhi Wei.

**Investigation:** Wenjin Lin.

**Methodology:** Qi Ye.

**Project administration:** Shuju Ning, Daozhi Wei.

**Resources:** Wenjin Lin.

**Supervision:** Shuju Ning.

**Validation:** Wei Huang, Xiaohua Wang, Qi Ye.

**Writing – original draft:** Wenjin Lin.

**Writing – review & editing:** Daozhi Wei.

## References

1. Committee E. Flora Reipublicae Popularis Sinicae. Beijing: Science Press; 2002 2002. 73 p.
2. NP C. Pharmacopoeia of the People's Republic of China. Beijing: China Medical Science and Technology Press; 2015. 199–200 p.
3. NP C. Pharmacopoeia of the People's Republic of China. Beijing: China Medical Science and Technology Press; 2015. 245 p.
4. Zhao X, He X, Zhong X. Anti-inflammatory and in-vitro antibacterial activities of Traditional Chinese Medicine Formula Qingdaisan. BMC Complement Altern Med. 2016; 16(1):503. Epub 2016/12/07. <https://doi.org/10.1186/s12906-016-1475-4> PMID: 27919254; PubMed Central PMCID: PMC5139090.
5. Sugimoto S, Naganuma M, Kiyohara H, Arai M, Ono K, Mori K, et al. Clinical Efficacy and Safety of Oral Qing-Dai in Patients with Ulcerative Colitis: A Single-Center Open-Label Prospective Study. Digestion. 2016; 93(3):193–201. Epub 2016/03/10. <https://doi.org/10.1159/000444217> PMID: 26959688.
6. Sugimoto S, Naganuma M, Kanai T. Indole compounds may be promising medicines for ulcerative colitis. J Gastroenterol. 2016; 51(9):853–61. Epub 2016/05/11. <https://doi.org/10.1007/s00535-016-1220-2> PMID: 27160749.
7. Suzuki H, Kaneko T, Mizokami Y, Narasaka T, Endo S, Matsui H, et al. Therapeutic efficacy of the Qing Dai in patients with intractable ulcerative colitis. World J Gastroenterol. 2013; 19(17):2718–22. Epub 2013/05/16. <https://doi.org/10.3748/wjg.v19.i17.2718> PMID: 23674882; PubMed Central PMCID: PMC3645393.
8. Lin YK, Chang YC, Hui RC, See LC, Chang CJ, Yang CH, et al. A Chinese Herb, Indigo Naturalis, Extracted in Oil (Lindioil) Used Topically to Treat Psoriatic Nails: A Randomized Clinical Trial. Jama Dermatol. 2015; 151(6):672–4. Epub 2015/03/05. <https://doi.org/10.1001/jamadermatol.2014.5460> PMID: 25738921.
9. Lin YK, See LC, Huang YH, Chang YC, Tsou TC, Lin TY, et al. Efficacy and safety of Indigo naturalis extract in oil (Lindioil) in treating nail psoriasis: a randomized, observer-blind, vehicle-controlled trial. Phytomedicine. 2014; 21(7):1015–20. Epub 2014/04/01. <https://doi.org/10.1016/j.phymed.2014.02.013> PMID: 24680615.
10. Li JH, Wang ZF, Xie YM, Zhao W. [Clinical characteristics and combined use of medicine analysis of 2 991 hospitalized patients with psoriasis based on real world database]. Zhongguo Zhong Yao Za Zhi. 2014; 39(18):3442–7. Epub 2014/12/24. PMID: 25532375.
11. Gu W, Zhang Y, Hao XJ, Yang FM, Sun QY, Morris-Natschke SL, et al. Indole alkaloid glycosides from the aerial parts of *Strobilanthes cusia*. J Nat Prod. 2014; 77(12):2590–4. Epub 2014/11/27. <https://doi.org/10.1021/np5003274> PMID: 25427242.

12. Mak NK, Leung CY, Wei XY, Shen XL, Wong RN, Leung KN, et al. Inhibition of RANTES expression by indirubin in influenza virus-infected human bronchial epithelial cells. *Biochem Pharmacol.* 2004; 67(1):167–74. Epub 2003/12/12. PMID: [14667939](#).
13. Zhou B, Yang Z, Feng Q, Liang X, Li J, Zanin M, et al. Aurantiamide acetate from *baphicacanthus cusia* root exhibits anti-inflammatory and anti-viral effects via inhibition of the NF-kappaB signaling pathway in Influenza A virus-infected cells. *J Ethnopharmacol.* 2017; 199:60–7. Epub 2017/01/26. <https://doi.org/10.1016/j.jep.2017.01.038> PMID: [28119097](#).
14. Ho YL, Kao KC, Tsai HY, Chueh FY, Chang YS. Evaluation of antinociceptive, anti-inflammatory and antipyretic effects of *Strobilanthes cusia* leaf extract in male mice and rats. *Am J Chin Med.* 2003; 31(1):61–9. Epub 2003/05/02. <https://doi.org/10.1142/S0192415X03000783> PMID: [12723755](#).
15. Liao BC, Jong TT, Lee MR, Chen SS. LC-APCI-MS method for detection and analysis of tryptanthrin, indigo, and indirubin in daqingye and banlangen. *J Pharm Biomed Anal.* 2007; 43(1):346–51. Epub 2006/08/04. <https://doi.org/10.1016/j.jpba.2006.06.029> PMID: [16884885](#).
16. Hu Z, Tu Y, Xia Y, Cheng P, Sun W, Shi Y, et al. Rapid Identification and Verification of Indirubin-Containing Medicinal Plants. *Evid Based Complement Alternat Med.* 2015; 2015:484670. Epub 2015/06/20. <https://doi.org/10.1155/2015/484670> PMID: [26089942](#); PubMed Central PMCID: [PMC4451998](#).
17. Warzecha H, Frank A, Peer M, Gillam EM, Guengerich FP, Unger M. Formation of the indigo precursor indican in genetically engineered tobacco plants and cell cultures. *Plant Biotechnol J.* 2007; 5(1):185–91. Epub 2007/01/09. <https://doi.org/10.1111/j.1467-7652.2006.00231.x> PMID: [17207267](#).
18. Tozzi S, Lercari B, Angelini LG. Light quality influences indigo precursors production and seed germination in *Isatis tinctoria* L. and *Isatis indigotica* Fort. *Photochem Photobiol.* 2005; 81(4):914–9. Epub 2005/03/05. <https://doi.org/10.1562/2004-08-03-RA-258> PMID: [15745425](#).
19. Gilbert KG, Maule HG, Rudolph B, Lewis M, Vandenburg H, Sales E, et al. Quantitative analysis of indigo and indigo precursors in leaves of *Isatis* spp. and *Polygonum tinctorium*. *Biotechnol Prog.* 2004; 20(4):1289–92. Epub 2004/08/07. <https://doi.org/10.1021/bp0300624> PMID: [15296465](#).
20. Hou HC, Liang SZ. [Determination of indirubin and indigo in *Baphicacanthus cusia* (Nees) Bremek by HPLC]. *Zhong Yao Cai.* 2006; 29(7):681–2. Epub 2006/10/25. PMID: [17059007](#).
21. Sarangi BK, Minami Y, Thul ST. RNA-Seq analysis for indigo biosynthesis pathway genes in *Indigofera tinctoria* and *Polygonum tinctorium*. *Genom Data.* 2015; 6:212–3. Epub 2015/12/24. <https://doi.org/10.1016/j.gdata.2015.09.021> PMID: [26697377](#); PubMed Central PMCID: [PMC4664758](#).
22. Xiang XL, Ning SJ, Huang YL, Zhang YJ, Zhu RL, Wei DZ. Effects of exogenous salicylic acid on protein expression level in *Baphicacanthus cusia* (Nees) Bremek leaves. *Chinese Journal of Applied Ecology.* 2010; 21(3):689–93. Epub 2010/06/22. PMID: [20560326](#).
23. Marcinek H, Weyler W, Deus-Neumann B, Zenk MH. Indoxyl-UDPG-glucosyltransferase from *Baphicacanthus cusia*. *Phytochemistry.* 2000; 53(2):201–7. Epub 2000/02/19. PMID: [10680172](#).
24. Zhou J, Li X, Chen Y, Dai CC. De novo Transcriptome Assembly of *Phomopsis liquidambari* Provides Insights into Genes Associated with Different Lifestyles in Rice (*Oryza sativa* L.). *Front Plant Sci.* 2017; 8:121. Epub 2017/02/22. <https://doi.org/10.3389/fpls.2017.00121> PMID: [28220138](#); PubMed Central PMCID: [PMC5292412](#).
25. Betts NS, Berkowitz O, Liu R, Collins HM, Skadhauge B, Dockter C, et al. Isolation of tissues and preservation of RNA from intact, germinated barley grain. *Plant J.* 2017; 91(4):754–65. Epub 2017/05/17. <https://doi.org/10.1111/tpj.13600> PMID: [28509349](#).
26. Scully ED, Gries T, Palmer NA, Sarath G, Funnell-Harris DL, Baird L, et al. Overexpression of *SbMyb60* in *Sorghum bicolor* impacts both primary and secondary metabolism. *New Phytol.* 2017. Epub 2017/09/26. <https://doi.org/10.1111/nph.14815> PMID: [28944535](#).
27. Senthil K, Jayakodi M, Thirugnanasambantham P, Lee SC, Duraisamy P, Purushotham PM, et al. Transcriptome analysis reveals in vitro cultured *Withania somnifera* leaf and root tissues as a promising source for targeted withanolide biosynthesis. *Bmc Genomics.* 2015; 16:14. Epub 2015/01/23. <https://doi.org/10.1186/s12864-015-1214-0> PMID: [25608483](#); PubMed Central PMCID: [PMC4310147](#).
28. Braynen J, Yang Y, Wei F, Cao G, Shi G, Tian B, et al. Transcriptome Analysis of Floral Buds Deciphered an Irregular Course of Meiosis in Polyploid *Brassica rapa*. *Front Plant Sci.* 2017; 8:768. Epub 2017/05/30. <https://doi.org/10.3389/fpls.2017.00768> PMID: [28553302](#); PubMed Central PMCID: [PMC5427127](#).
29. Zhu Y, Li Y, Xin D, Chen W, Shao X, Wang Y, et al. RNA-Seq-based transcriptome analysis of dormant flower buds of Chinese cherry (*Prunus pseudocerasus*). *Gene.* 2015; 555(2):362–76. Epub 2014/12/03. <https://doi.org/10.1016/j.gene.2014.11.032> PMID: [25447903](#).
30. Tang X, Xiao Y, Lv T, Wang F, Zhu Q, Zheng T, et al. High-throughput sequencing and De Novo assembly of the *Isatis indigotica* transcriptome. *Plos One.* 2014; 9(9):e102963. Epub 2014/09/27. <https://doi.org/10.1371/journal.pone.0102963> PMID: [25259890](#); PubMed Central PMCID: [PMC4178013](#).

31. Zhang L, Chen J, Zhou X, Chen X, Li Q, Tan H, et al. Dynamic metabolic and transcriptomic profiling of methyl jasmonate-treated hairy roots reveals synthetic characters and regulators of lignan biosynthesis in *Isatis indigotica* Fort. *Plant Biotechnol J*. 2016; 14(12):2217–27. Epub 2016/05/08. <https://doi.org/10.1111/pbi.12576> PMID: 27155368; PubMed Central PMCID: PMC5103230.
32. Zhang L, Chen J, Li Q, Chen W. Transcriptome-wide analysis of basic helix-loop-helix transcription factors in *Isatis indigotica* and their methyl jasmonate responsive expression profiling. *Gene*. 2016; 576(1 Pt 1):150–9. Epub 2015/10/10. <https://doi.org/10.1016/j.gene.2015.09.083> PMID: 26449398.
33. Zhou Y, Kang L, Liao S, Pan Q, Ge X, Li Z. Transcriptomic analysis reveals differential gene expressions for cell growth and functional secondary metabolites in induced autotetraploid of Chinese woad (*Isatis indigotica* Fort.). *Plos One*. 2015; 10(3):e0116392. Epub 2015/03/05. <https://doi.org/10.1371/journal.pone.0116392> PMID: 25739089; PubMed Central PMCID: PMC4349453.
34. Chen J, Dong X, Li Q, Zhou X, Gao S, Chen R, et al. Biosynthesis of the active compounds of *Isatis indigotica* based on transcriptome sequencing and metabolites profiling. *Bmc Genomics*. 2013; 14:857. Epub 2013/12/07. <https://doi.org/10.1186/1471-2164-14-857> PMID: 24308360; PubMed Central PMCID: PMC3890716.
35. Song Z, Guo L, Liu T, Lin C, Wang J, Li X. Comparative RNA-Sequence Transcriptome Analysis of Phenolic Acid Metabolism in *Salvia miltiorrhiza*, a Traditional Chinese Medicine Model Plant. *Int J Genomics*. 2017; 2017:9364594. Epub 2017/02/15. <https://doi.org/10.1155/2017/9364594> PMID: 28194403; PubMed Central PMCID: PMC5282420.
36. Zhang X, Dong J, Liu H, Wang J, Qi Y, Liang Z. Transcriptome Sequencing in Response to Salicylic Acid in *Salvia miltiorrhiza*. *Plos One*. 2016; 11(1):e0147849. Epub 2016/01/26. <https://doi.org/10.1371/journal.pone.0147849> PMID: 26808150; PubMed Central PMCID: PMC4726470.
37. Wei S, Ma X, Pan L, Miao J, Fu J, Bai L, et al. Transcriptome Analysis of *Taxillus chinensis* (DC.) Dancer Seeds in Response to Water Loss. *Plos One*. 2017; 12(1):e0169177. Epub 2017/01/04. <https://doi.org/10.1371/journal.pone.0169177> PMID: 28046012; PubMed Central PMCID: PMC5207531.
38. Rai A, Kamochi H, Suzuki H, Nakamura M, Takahashi H, Hatada T, et al. De novo transcriptome assembly and characterization of nine tissues of *Lonicera japonica* to identify potential candidate genes involved in chlorogenic acid, luteolosides, and secoiridoid biosynthesis pathways. *J Nat Med*. 2017; 71(1):1–15. Epub 2016/09/16. <https://doi.org/10.1007/s11418-016-1041-x> PMID: 27629269; PubMed Central PMCID: PMC5214891.
39. Zhang X, Song Z, Liu T, Guo L, Li X. De Novo Assembly and Comparative Transcriptome Analysis Provide Insight into Lysine Biosynthesis in *Toona sinensis* Roem. *Int J Genomics*. 2016; 2016:6735209. Epub 2016/07/05. <https://doi.org/10.1155/2016/6735209> PMID: 27376077; PubMed Central PMCID: PMC4914729.
40. Hua Y, Wang S, Liu Z, Liu X, Zou L, Gu W, et al. Transcriptomic analysis of *Pseudostellariae Radix* from different fields using RNA-seq. *Gene*. 2016; 588(1):7–18. Epub 2016/04/30. <https://doi.org/10.1016/j.gene.2016.04.043> PMID: 27125225.
41. Feng Y, Zhang L, Fu J, Li F, Wang L, Tan X, et al. Characterization of Glycolytic Pathway Genes Using RNA-Seq in Developing Kernels of *Eucommia ulmoides*. *J Agric Food Chem*. 2016; 64(18):3712–31. Epub 2016/04/14. <https://doi.org/10.1021/acs.jafc.5b05918> PMID: 27074598.
42. Li Y, Wang X, Chen T, Yao F, Li C, Tang Q, et al. RNA-Seq Based De Novo Transcriptome Assembly and Gene Discovery of *Cistanche deserticola* Fleshy Stem. *Plos One*. 2015; 10(5):e0125722. Epub 2015/05/06. <https://doi.org/10.1371/journal.pone.0125722> PMID: 25938435; PubMed Central PMCID: PMC4418726.
43. Wang G, Du X, Ji J, Guan C, Li Z, Josine TL. De novo characterization of the *Lycium chinense* Mill. leaf transcriptome and analysis of candidate genes involved in carotenoid biosynthesis. *Gene*. 2015; 555(2):458–63. Epub 2014/12/03. <https://doi.org/10.1016/j.gene.2014.10.058> PMID: 25445268.
44. Hua W, Zheng P, He Y, Cui L, Kong W, Wang Z. An insight into the genes involved in secoiridoid biosynthesis in *Gentiana macrophylla* by RNA-seq. *Mol Biol Rep*. 2014; 41(7):4817–25. Epub 2014/04/08. <https://doi.org/10.1007/s11033-014-3352-x> PMID: 24706059.
45. Minami Y, Takao H, Kanafuji T, Miura K, Kondo M, Hara-Nishimura I, et al. beta-Glucosidase in the indigo plant: intracellular localization and tissue specific expression in leaves. *Plant Cell Physiol*. 1997; 38(9):1069–74. Epub 1997/11/14. PMID: 9360324.
46. Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat Methods*. 2008; 5(7):621–8. Epub 2008/06/03. <https://doi.org/10.1038/nmeth.1226> PMID: 18516045.