

DOI: 10.5586/asbp.3574

**Publication history**

Received: 2017-08-20

Accepted: 2018-02-23

Published: 2018-03-28

**Handling editor**

Grzegorz Jackowski, Faculty of Biology, Adam Mickiewicz University in Poznań, Poland

**Authors' contributions**

DP, NY: research designing; RH, DX, WZ, GL: conducting experiments; RH, XT: data analyzing and manuscript writing; RH and DX contributed equally to this study

**Funding**

This study was supported by Education Department of Anhui Province (grant No. gxyq2017016), Scientific Research Innovation Platform for Universities in Anhui Province (grant No. 2015-49), Quality Improvement Project for Anhui Authentic Chinese Herbal Medicine (grant No. 2013-02), and Anhui University of Chinese Medicine (grant No. 2017zrzd011).

**Competing interests**

No competing interests have been declared.

**Copyright notice**© The Author(s) 2018. This is an Open Access article distributed under the terms of the [Creative Commons Attribution License](#), which permits redistribution, commercial and noncommercial, provided that the article is properly cited.**Citation**Han R, Xie D, Tong X, Zhang W, Liu G, Peng D, et al. Transcriptomic landscape of *Dendrobium huoshanense* and its genes related to polysaccharide biosynthesis. *Acta Soc Bot Pol.* 2018;87(1):3574. <https://doi.org/10.5586/asbp.3574>**Digital signature**

This PDF has been certified using digital signature with a trusted timestamp to assure its origin and integrity. A verification trust dialog appears on the PDF document when it is opened in a compatible PDF reader. Certificate properties provide further details such as certification time and a signing reason in case any alterations made to the final content. If the certificate is missing or invalid it is recommended to verify the article on the journal website.

## ORIGINAL RESEARCH PAPER

# Transcriptomic landscape of *Dendrobium huoshanense* and its genes related to polysaccharide biosynthesis

Rongchun Han<sup>1,2</sup>, Dongmei Xie<sup>1</sup>, Xiaohui Tong<sup>3</sup>, Wei Zhang<sup>1</sup>, Gang Liu<sup>1</sup>, Daiyin Peng<sup>1</sup>, Nianjun Yu<sup>1\*</sup><sup>1</sup> School of Pharmacy, Anhui University of Chinese Medicine, Hefei 230012, China<sup>2</sup> Anhui Innovative Team from Colleges for Scientific Research's Platform – The Innovative Team in Researching the Key Technologies Concerning the Integration of Processing Chinese Medicine Decoction Pieces in Producing Area, Hefei 230012, China<sup>3</sup> Institute of Metabolic and Cardiovascular Diseases, Joint Research Unit 1048 INSERM Adipolab Unit, Paul Sabatier University, Toulouse 31432, France\* Corresponding author. Email: [ynj2005288@sina.com](mailto:yjn2005288@sina.com)**Abstract**

*Dendrobium huoshanense* has long been used to treat various diseases in oriental medicine. In order to study its gene expression profile, transcripts involved in the biosynthesis of precursors of polysaccharides, as well as mechanisms underlining morphological differences between wild and cultivated plants, three organs of both wild type and cultivated *D. huoshanense* were collected and sequenced by Illumina HiSeq4000 platform, yielding 919,409,540 raw reads in FASTQ format. After Trinity de novo assembly and quality control, 241,242 nonredundant contigs with the average length of 967.5 bp were generated. qRT-PCR experiment on the selected transcripts showed the transcriptomic data were reliable. BLASTx was conducted against NR, SwissProt, String, Pfam, and KEGG. Gene ontology annotation revealed more than 40,000 contigs assigned to catalytic activity and metabolic process, suggesting its dynamic physiological activities. By searching KEGG pathway, six genes potentially involved in mannose biosynthetic pathway were retrieved. Gene expression analysis for stems between wild and cultivated *D. huoshanense* resulted in 956 genes differentially expressed. Simple sequence repeats (SSRs) analysis revealed 143 SSRs with the unit size of 4 and 3,437 SSRs the size of 3. The obtained SSRs are the potential molecular markers for discriminating distinct cultivars of *D. huoshanense*.

**Keywords***Dendrobium huoshanense*; transcriptomic analysis; RNA-Seq; biosynthesis**Introduction**

*Dendrobium huoshanense* C. Z. Tang et S. J. Cheng has been used in traditional Chinese medicine for centuries with the efficacy of curing asthenic symptoms resulted from various diseases as well as treating loss of body fluid due to fever or other causes. It is embodied in *Chinese pharmacopeia* as *Dendrobii Caulis*, along with several other entries that are from the same genus of *Dendrobium* in Orchidaceae [1]. Its main active ingredients are polysaccharides, terpenoids, alkaloids and flavonoids, etc., with pharmacological effects including oxidation resistance, immune enhancement, liver protection, and antitumor as well [2,3]. Metabolic study on *D. huoshanense* demonstrated that according to accumulating pattern of polysaccharides and related metabolites, its best harvest time should be in the third year [4]. A polysaccharide from this plant showed significant hepatoprotective attributes in mice [5]. And efforts to tackle the structure of such compounds by anion exchange as well as gel permeation chromatography were

also made [6]. On the other hand, a characteristic alkaloid found in *D. huoshanense*, dendrobine, which was first isolated by H. Suzuki in 1932 from *D. nobile* and comprised the sesquiterpene skeleton structure with 15 carbons, acts as an antagonist of taurine and beta-alanine [7]. Moreover, dendrobine also exhibits antipyretic, hypertensive, and antiviral effects [8]. Although several groups have reported racemic and enantioselective syntheses for this compound during the past five decades [9–12], study on its biosynthetic pathway remains limited [13,14].

Due to its rigorous requirement for growth environment, long growth cycle, and illogical development by human beings, resource for *D. huoshanense* in China is now endangered. Employing transcriptomic approaches to collect and analyze bioinformation for understanding its genetic makeup and biosynthetic pathways of important secondary metabolites is a practical way to better protect and utilize this medicinal resource by using biotechnologies to produce products of interest in the future. By far, studies on genomic and transcriptomic level of *Dendrobium* plants have been confined to a few species such as *D. officinal* and *D. nobile* [13,15,16]. To our best knowledge, no transcriptomic report on *D. huoshanense* was retrieved up to date.

Analyses on transcripts before the dawn of microarray usually deployed radioactive DNA probes to hybridize targets in addition to common protocols such as homogenization and nucleic acid extraction. It was slow, expensive, laborious, and environmentally harmful [17]. With the development of DNA microarray, studying transcriptome became possible by fixing custom-designed DNA, cDNA, or oligonucleotides probes on glass surface or silicon chips for hybridization [18]. Shortcomings accompanying this powerful approach are mainly the requirement of existing knowledge regarding genomic information and the limited dynamic range of microarray chips themselves. In the early days of this century, RNA-Seq (RNA sequencing), also known as deep or high-throughput sequencing, offered a brand-new solution for those nonmodel plants, including *D. huoshanense*, which lack reference genomes. Meanwhile, RNA-Seq technologies intensively read amplified cDNA sequences and the coverage showing the times at certain locations where a nucleotide is detected could be as many as several thousand. Main stream software for de novo assembly of the data resulted from sequencers include SOAPdenovo2 [19] (Short Oligonucleotide Analysis Package 2), AbySS [20] (Assembly by Short Sequences), and Trinity [21] with numerous downstream applications to facilitate individual tasks.

In this study, 919,409,540 raw reads in FASTQ format from six organs of *D. huoshanense* were obtained by Illumina HiSeq4000 sequencing platform. De novo assembly was fulfilled by Trinity software. Transcripts possibly involved in the biosynthesis of important secondary metabolites, polysaccharides, and dendrobine, were proposed by analyzing findings from KEGG [22] (Kyoto Encyclopedia of Genes and Genomes) pathway and Gene Ontology databases. Assembled contigs, annotation, as well as differently expressed genes between distinct samples were uploaded to Github (<https://github.com/rongchunhan/Dendrobium-huoshanense>).

## Material and methods

### Plant material and mRNA purification

All fresh samples were collected from healthy *D. huoshanense* plants in Huoshan County, Anhui Province, China in June 2016. Three biological replicates were prepared for wild flower of this species, three for wild leaf and three for wild stem. As far as cultivated type is concerned, the biological replicates for flower, leaf, and stem are two, three and three, respectively. All the subjects were submerged in RNAlater solution (Invitrogen, USA) right after sampling in the fields, with each sample trimmed to approximately 5 mm in every dimension. For total RNA extraction, RNAlater was removed from the sample surface by using a Kimwipe and then materials were frozen in liquid nitrogen. Subsequently, plant RNA purification reagent (Invitrogen) was used according to the manufacturer's instructions. Agilent 2100 Bioanalyzer was also applied to assess RNA integrity number (RIN), which should be over 7.5. mRNA was gathered by magnetic beads with oligo(dT) complementary to polyA tail of mRNA.

### mRNA fragmentation, library preparation

Fragmentation buffer from Truseq RNA Sample Prep Kit (Illumina, USA) helped to randomly break mRNA, which could be several thousand kilobases long, into 300-bp fragments. With the presence of reverse transcriptase and random primers, first strand cDNA was synthesized. After the formation of double-stranded cDNA, end repair mix was added to make blunt end followed by attachment of adaptors.

### High throughput sequencing and de novo assembly

cDNA libraries were enriched through PCR amplification for 15 cycles and then certified low range ultra agarose was used for electrophoresis to recover target fragments. After quantification by TBS380 Picogreen (Invitrogen), samples were loaded to cBot for bridge PCR to form clusters. Finally, original data were generated from Illumina instrument through sequencing by synthesis approach.

By examining A/T/G/C base content distribution, distribution of base quality, and base error rate distribution, raw data were trimmed by using trimmomatic [23] to generate clean data for following processes. We adopted Trinity method to conduct de novo assembly and yielded contigs, which were the basis for downstream biological function analysis.

### Transcripts annotation and KEGG pathway analysis

Prior to gene annotation, we used ORF prediction package provided by Trinity to do gene prediction for all the transcripts obtained. For eukaryotes, each sequence should contain UTR and then best ORF region is predicted according to Markov models [24]. Pfam database serves the purpose of revising predicted results and only protein sequences which can be matched to this database will be retained.

For BLASTx (Basic Local Alignment Search Tool), all the contigs were used to search NCBI NR (nonredundant), String, SwissProt [25], and KEGG databases for functional annotation with E-value set to  $1e-5$ . NCBI NR contains nonredundant information from SwissProt, PIR (Protein Information Resource), PRF (Protein Research Foundation), and PDB (Protein Data Bank) databases and the translated protein data from GenBank as well as RefSeq. By comparing *D. huoshanense* transcriptomic data with NR, information for homologous sequences and similarity between related species could be attained.

By standardizing biological jargons across different databases, gene ontology [26] (GO) aims to categorize the results related to genetic research all over the world. We analyzed GO classification for *D. huoshanense* to study biological meanings and its distribution of all gene products. Currently, information from 66 unicellular organisms in COG (Clusters of Orthologous Groups of proteins) is generally divided into functional classifications [27]. By extracting COG number when comparing *D. huoshanense* sequence information against String database, we studied the functional attributes of this medicinal plant.

In living organisms, gene products do not exist or take effect in isolation. Different factors interact orderly to exhibit biological function, which makes KEGG, a database focusing on tackling genetic information in a systematic view, more valuable. It integrates sophisticated biological functions such as metabolic pathway, genetic information transfer, cellular process, and so forth to solve problems theoretically and practically. By BLASTx against KEGG database [28], we extracted KEGG orthology numbers which were used to retrieve biological pathways.

### Gene expression analysis

RSEM [29] software calculates number of read counts for each transcript in a specific library and then FPKM (fragments per kilobase of exon per million mapped reads) values were obtained to show gene expression level. Three ingredients are essential: total

exon fragments, mapped reads (millions), and exon length (Kb). To perform differential gene expression (GE) analysis, edgeR [30] was used based on RSEM results. For *D. huoshanense*, screening criteria for differential GE were FDR (false discovery rate) < 0.05 and  $\log_2|\text{FC}(\text{fold change})| \geq 1$ . We also analyzed upregulated and downregulated information in terms of GO and be aware that for a single transcript, it may correspond to several GO entries and vice versa.

### Simple sequence repeats extraction

SSRs stands for simple sequence repeats, also called microsatellite or STRs (simple tandem repeats). SSR is widely distributed in eukaryotic genomes and because number and type of repeats vary in different individuals, it is very useful as molecular marker in the form of SSR or ISSR (inter SSR). We applied MISA (ver. 1.0) to detect SSR in transcripts [31].

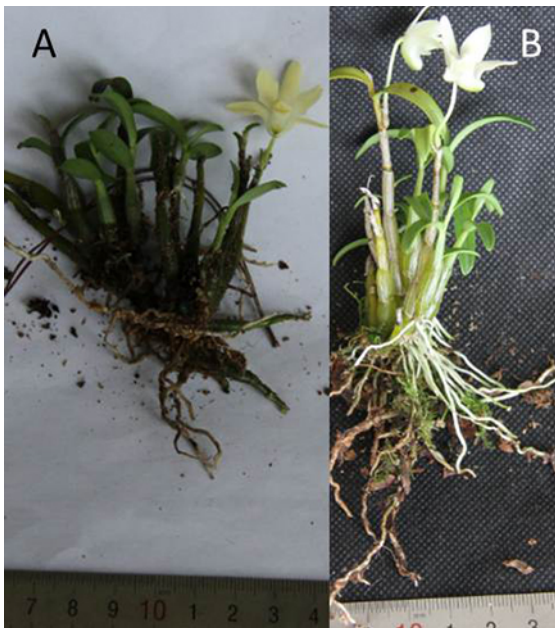
### Gene expression verification using quantitative RT-PCR

Total RNA was extracted from different samples by the same method described in section “Plant materials and mRNA purification”. RNA quality was then evaluated by ScanDrop 200 (Jena, Germany). Corresponding cDNA was synthesized with Superscript III Reverse Transcriptase (Invitrogen). Primers for qRT-PCR were designed by using NCBI/Primer-BLAST and tubulin was adopted as the internal control gene (Tab. S1). Amplification was performed using Stepone Plus Real-Time PCR System (Applied Biosystems, USA) with the method described previously [32].

## Results

### Plant materials and cDNA library construction

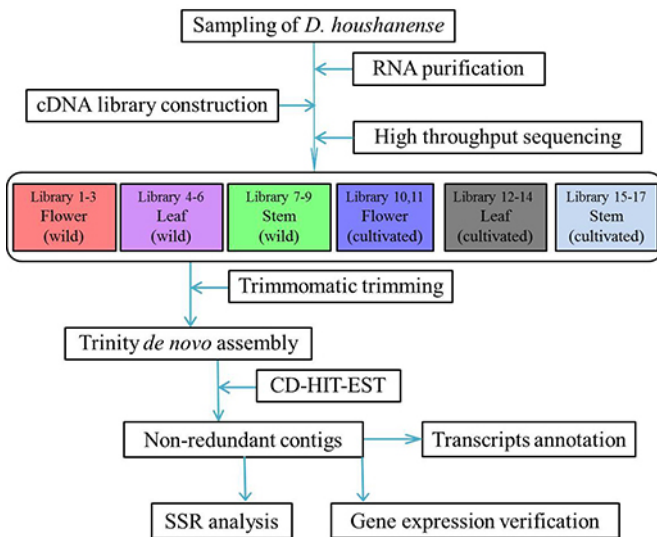
*Dendrobium huoshanense* is perennial herb with succulent stems ranging from 3 to 9 cm long. Basal parts of its stem are typically inflated. The stems of cultivated *D. huoshanense* are commonly longer than that of wild ones (Fig. 1) and after transplanting the cultivated plants to wild environment, this feature is preserved. In this study, we intended to collect information for the mechanisms underlining this morphological difference. After RNA extraction and purification, quality control was conducted to assure values of  $\text{OD}_{260/280}$  and  $\text{OD}_{260/230}$  were in the range of 1.8–2.0. A total of 17 cDNA libraries were constructed and sequenced by Illumina system. Library 1 to 3 belonged to *D. huoshanense* flower (wild), 4–6 leaf (wild), 7–9 stem (wild), 10–11 flower (cultivated), 12–14 leaf (cultivated), and 15–17 stem (cultivated). Fig. 2 illustrates the experimental flowchart for *D. huoshanense* analysis.



**Fig. 1** Plant materials for *D. huoshanense*. Wild (A) and cultivated types (B) were collected in 2016 for RNA-Seq analysis with three biological replicates, respectively.

### Trinity de novo assembly

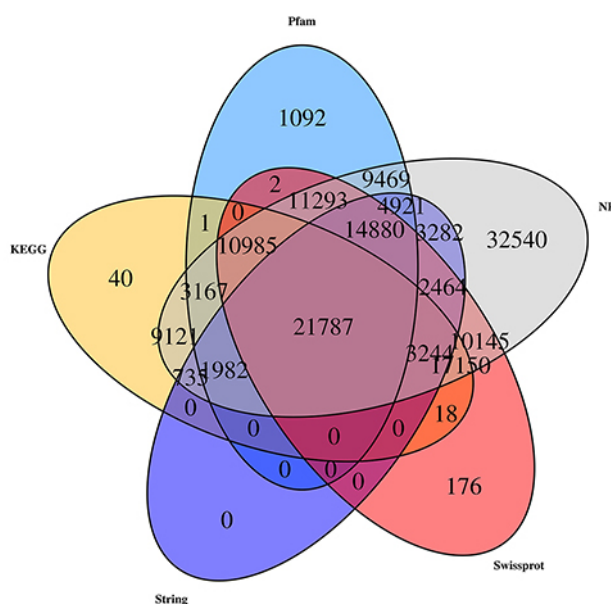
After high-throughput sequencing, trimming as well as quality validation of the raw reads are indispensable because they may include adapters or data with poor quality. The software of Trimmomatic was used to detect and discard adapters, subsequences of ambiguous nucleotides, and low-quality reads. The resultant clean reads were used for downstream de novo assembly. Clean reads from all 17 libraries were put together to be processed by Trinity and the output contigs longer than 200 bp were retrieved. To remove duplicates of contigs, CD-HIT-EST [33] was applied



**Fig. 2** Experimental flowchart for *D. huoshanense* RNA-Seq analysis.

**Tab. 1** Overview of *D. huoshanense* RNA-Seq assembly.

Items	Numbers
Total bases of clean reads	1.32167e+11
No. of clean reads	897,088,840
Average length of contigs (bp)	967.5
Maximum contig length (bp)	59,869
Percent GC	43.77
N50 (bp)	1,469
No. of nonredundant contigs	241,242

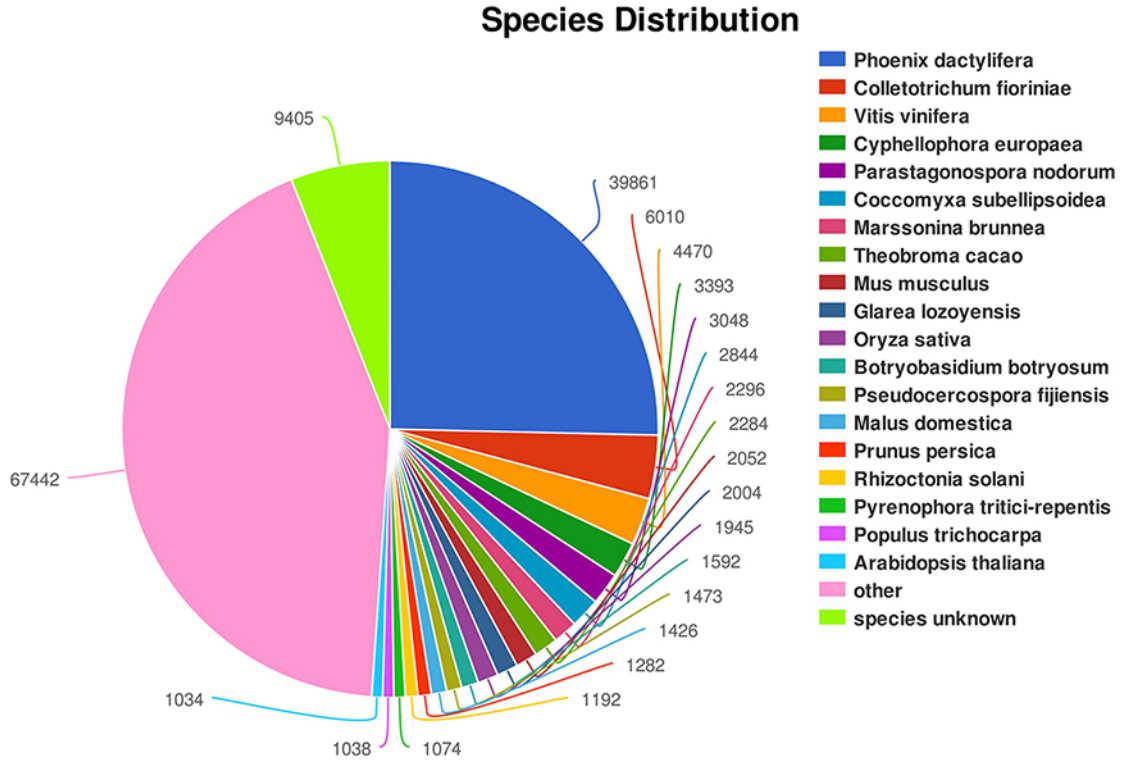


**Fig. 3** BLAST results for *D. huoshanense* transcripts. The Venn diagram demonstrates distribution of annotated contigs against five public protein databases.

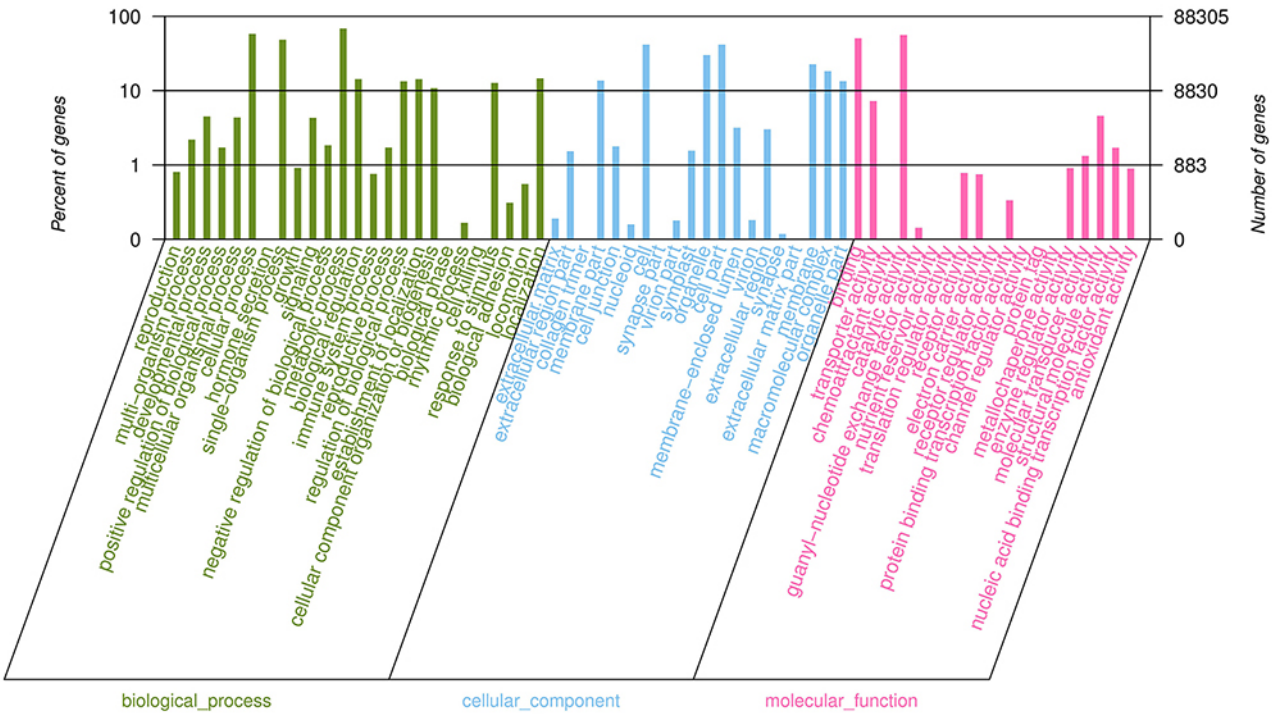
and its threshold of 0.9 was set to reduce noise interference and computational time. Finally, 241,242 nonredundant contigs were obtained. Tab. 1 shows statistics of sequencing and assembly. Percent of GC in eukaryotes varies from 20% to 60% [34]. GC content of *D. huoshanense* was found to be 43.77%, which is in the middle. Sequence length distribution of nonredundant contigs was summarized in Fig. S1. Number of contigs with the length from 201 to 1,000 bp was 161,876, which covered 67.1% of all transcripts.

#### Annotation of assembled transcripts

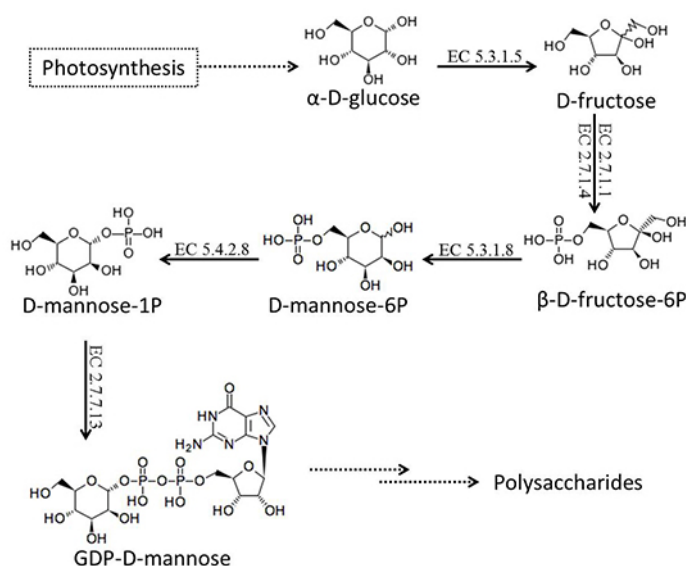
The number of contigs predicted to contain ORF by using ORF prediction package was 142,224 and this equaled 58.9% of the total assembled contigs, suggesting reliable results of our analysis. BLASTx is an algorithm to search a protein database with a translated nucleotide query. We adopted this strategy to do BLASTx against NCBI NR and other four public repositories. As a result, 32.99% of the transcripts had hit(s) in Pfam, 28.28% in KEGG, 38.20% in SwissProt, 22.09% in String, and 65.15% in NR. A detailed breakdown is presented in Fig. 3. By analyzing the hits against NCBI NR, transcripts from *D. huoshanense* that had similar sequences with different species could be studied. To our surprise, 39,861 contigs showed significant homology with *Phoenix dactylifera*, commonly known as date palm from family Arecaceae. And only 1,034 contigs were found closely similar to *Arabidopsis thaliana* (Fig. 4). Gene ontology classification provides overall insights into expression pattern of organisms under certain circumstances. We used Blast2GO online facility (<http://www.blast2go.com/b2ghome>) to conduct GO annotation. Due to the nature of GO categories, one contig may correspond to several GO numbers from the three distinct classes: biological process, cellular component, and molecular function. From Fig. 5, over 40,000 contigs were assigned to catalytic activity and metabolic process, suggesting dynamic physiological properties taking place in *D. huoshanense*, which is responsible for its medicinal efficacy. Nevertheless, the implications by using such strategy to analyze GO have obvious limitations. KEGG database is mostly dedicated for human and animal models and thus GO terms related to animal diseases may not be relevant to the study on plants. We turned to agriGO v2.0 [35], which focused on agricultural species, for more information and compared *D. huoshanense* GO data with that of *A. thaliana*. The result (Fig. S2) showed representative GO terms of the two species had similar distribution pattern. Percent of genes for *A. thaliana* was higher than that for *D. huoshanense* and this may attribute to its abundant short contigs that could not hit corresponding GO terms. We retrieved COG numbers for *D. huoshanense* transcripts by comparing sequence similarity against String database and subsequently acquired their COG function classification. Within



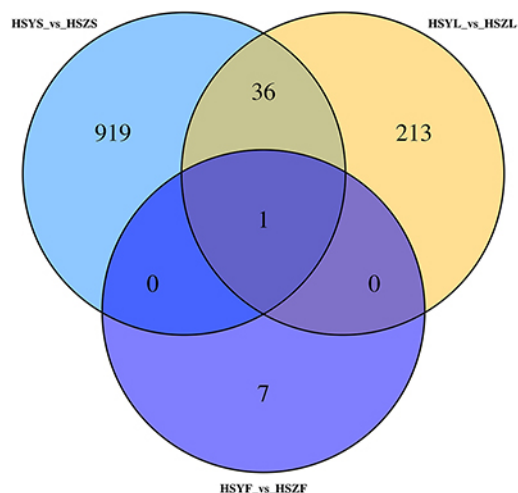
**Fig. 4** Species distribution showing homogenous sequences between *D. huoshanense* and the related organisms.



**Fig. 5** GO classification for overall expression pattern of *D. huoshanense*.



**Fig. 6** Proposed biosynthetic pathway of mannose in *D. huoshanense*. Enzyme commission (EC) numbers are presented in the diagram.



**Fig. 7** Differently expressed genes in comparisons between organs of wild and cultivated plants. HSYS – stem of the wild *D. huoshanense*; HSZS – stem of the cultivated; HSYL – leaf of the wild; HSZL – leaf of the cultivated; HSYF – flower of the wild; HSZF – flower of the cultivated.

the 25 classes, 3,909 transcripts were assigned to the type of information storage and processing with the functional label as translation, ribosomal structure, and biogenesis. Of the 12,472 genes that are related to metabolism, 1,144 transcripts were linked to secondary metabolites biosynthesis, transport, and catabolism (Fig. S3). KEGG pathway enrichment provides information regarding genes' interaction in a metabolic web. We listed 20 biosynthetic pathways in KEGG, which involved the most transcripts from *D. huoshanense* in Fig. S4 and 3,679 genes were assigned to ribosome biosynthesis.

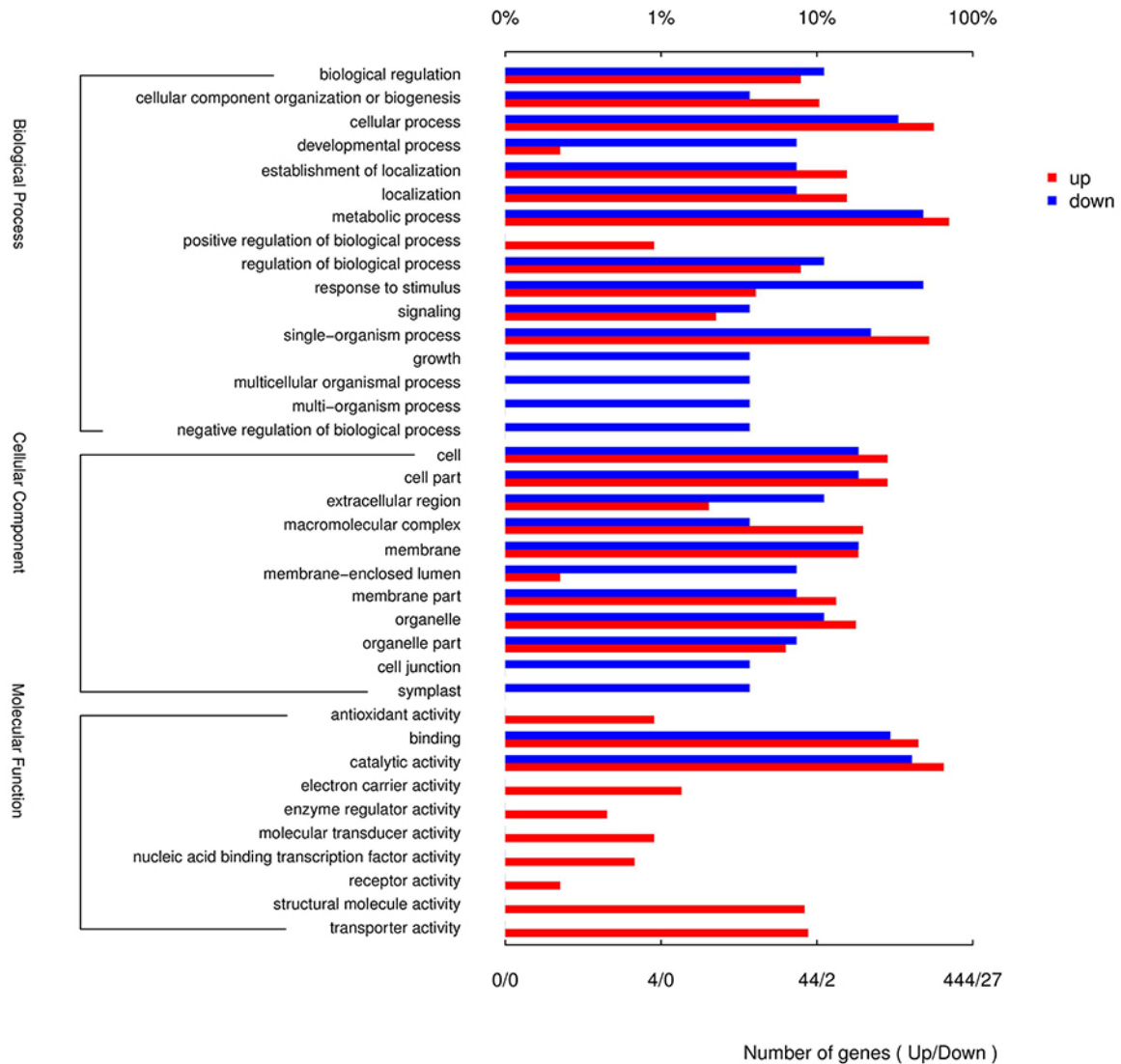
*Dendrobium huoshanense* is famous for its high contents of polysaccharides. By referring to KEGG database, the following is a possible biosynthetic pathway for *D. huoshanense* to produce mannose as the precursor of polysaccharides. After α-D-glucose is synthesized via photosynthesis, xylose isomerase catalyzes it to produce D-fructose, which is then turned into β-D-fructose-6P with the help of hexokinase or fructokinase. Subsequently, mannose-6-phosphate isomerase catalyzes β-D-fructose-6P to yield D-mannose-6P which is catalyzed by phosphomannomutase to form D-mannose-1P. Finally, with the presence of mannose-1-phosphate guanylyltransferase, GDP-D-mannose comes into being. The six above-mentioned enzymes that may be involved in the biosynthesis of polysaccharides were all presented in *D. huoshanense* transcriptomic dataset (Fig. 6).

#### Differential expression and SSR analysis

We were interested in analyzing differentially expressed (DE) genes between the wild and cultivated plants. Fig. 7 demonstrates that 956 genes were differentially expressed between stems of distinct types of *D. huoshanense* and this was a solid foundation to rely on for studying their morphological traits. Compared to leaf and flower, the morphological differences of stem between the wild and cultivated were more obvious. Generally, the stem of *D. huoshanense* has four–seven nodes and the protruding character for the cultivar is its longer

internode compared with the wild type (Fig. 1), whereas it is hard to distinguish the two types solely by morphology of their leaves and flowers. As a new species, *D. huoshanense* was named fairly recently in 1984 [36] and cultivation for this herb started even later. In 2005, the seeds were collected by our lab in the wild and then suitable medium were prepared to help them germinate. After a few generations, the traits from the selected cultivar showed genetic stability. The relatively short period of breeding and inherent stability of the propagative organ may explain the reason why there were only eight DE genes found between wild and cultivated flowers. For DE genes between leaves of cultivated and wild type, there were 250 of them.

Gene ontology differential expression enrichment was also conducted for stems of both wild type and cultivar (Fig. 8). GO terms related to growth, multiorganism process and negative regulation of biological process were downregulated, which meant certain FPKM values in the cultivar were lower. On the contrary, receptor activity, structural molecular activity, and some others were higher.



**Fig. 8** Differently expressed genes in comparisons between organs of wild and cultivated plants. HSYS – stem of the wild *D. huoshanense*; HSZS – stem of the cultivated; HSYL – leaf of the wild; HSZL – leaf of the cultivated; HSYF – flower of the wild; HSZF – flower of the cultivated.

SSR is a practical tool to discriminate different cultivars within a species. By examining a total of 171,111 contigs of *D. huoshanense*, we identified 25,424 SSRs and 4,367 contigs contained more than one SSR. In terms of unit size, it ranged from 1 to 6. SSRs in a number of 143 were retrieved with the unit size of 4 and 3,437 SSRs the size of 3. Fig. S5 demonstrates the distribution of SSRs in sequences.

#### Gene expression validation

To conduct qRT-PCR, three biological replicates were utilized for each reaction and a negative control comprising template without primers was also included. One gene involved in flavonoids biosynthesis, the other related to sugar metabolism, and the third annotated as class-1 low-molecular-weight heat shock protein (LMW HSP) were chosen as the target genes. Tab. 2 shows similar trend as far as the fold change between results of bioinformatics and qRT-PCR is concerned. For LMW HSP, it was differently expressed according to edgeR analysis, which coincided with our qPCR results. This indicated the RNA-Seq output was reproducible and reliable.



**Tab. 2** Evaluation of selected DE genes.

Contig ID	Annotation	Fold change: RNA-Seq and qPCR verification					
		Flower (wild/cultivated)		Stem (wild/cultivated)		Leaf (wild/cultivated)	
		FPKM	2 <sup>-ΔCT</sup>	FPKM	2 <sup>-ΔCT</sup>	FPKM	2 <sup>-ΔCT</sup>
c297732_g1_i1	CHI	2.51	2.03	1.06	1.35	1.18	2.12
c298649_g4_i2	PPM	1.25	1.77	1.13	1.29	0.64	0.27
c298471_g5_i1	LMW HSP	2.05	1.24	6.52*	5.10	8.22*	6.89

Fold change of FPKM and 2<sup>-ΔCT</sup> between wild and cultivated *D. huoshanense*, respectively. \* Differently expressed gene detected by edgeR. CHI – chalcone isomerase; PPM – phosphomannomutase; LMW HSP – class-1 low-molecular-weight heat shock protein.

## Discussion

*Dendrobium huoshanense* possesses remarkable pharmaceutical attributes and according to ancient materia medica books, it has been applied to treat various diseases for a very long time in China. However, due to its high morphological similarity with species from the same genus *Dendrobium*, it was often confused with others and researchers made judgments according to its very narrow distribution area which was recorded in ancient books. It was not until 1984 [36] that its scientific name was published and *Chinese pharmacopeia* included several *Dendrobium* species under the name *Dendrobii Caulis*. Commercial cultivation of *D. huoshanense* began in late twentieth century and the limited time span may explain the reason why gene expression was so similar concerning flower from the cultivar and wild type. Fig. 8 demonstrates downregulated GO terms related to growth in cultivar despite its longer stem and internode. This suggests the underlying mechanisms of *D. huoshanense* morphology are complicated. Despite the effort to tackle genetic information for the medicinal plants in genus *Dendrobium*, RNA-Seq analysis for *D. huoshanense* has not been conducted so far. To provide transcriptomic data for this plant will facilitate the study on the biosynthesis of its abundant polysaccharides and other important secondary metabolites. By collecting 17 samples of wild and cultivated plants, we used Illumina platform to sequence its flower, stem, and leaf. Through qRT-PCR verification, the resultant data were reliable. RNA-Seq analysis provides not only the overall expression profile of this plant, but also detailed information of sequences that are involved in various biosynthetic pathways. The transcripts identified in fructose and mannose biosynthesis is particularly valuable because the genes will help to certify the actual pathway in *D. huoshanense* and by gene expression correlation, new genes related to downstream polysaccharide biosynthesis may pop up. The results of this experiment will serve as a stepping stone for *D. huoshanense* future research.

## Supplementary material

The following supplementary material for this article is available at <http://pbsociety.org.pl/journals/index.php/asbp/rt/suppFiles/asbp.3574/0>:

**Tab. S1** Primers designed for qRT-PCR verification.

**Fig. S1** Sequence length distribution of the assembled contigs.

**Fig. S2** *Dendrobium huoshanense* GO annotation compared with *A. thaliana*.

**Fig. S3** COG functional classification.

**Fig. S4** Top 20 biosynthetic pathways in KEGG that *D. huoshanense* transcripts were involved in.

**Fig. S5** Distribution of different SSR types.

## References

1. Chinese Pharmacopoeia Commission. Chinese pharmacopoeia. Beijing: The Medicine Science and Technology Press of China; 2015.
2. Ng TB, Liu J, Wong JH, Ye X, Wing Sze SC, Tong Y, et al. Review of research on *Dendrobium*, a prized folk medicine. *Appl Microbiol Biotechnol*. 2012;93(5):1795–1803. <https://doi.org/10.1007/s00253-011-3829-7>
3. Fan Y, He X, Zhou S, Luo A, He T, Chun Z. Composition analysis and antioxidant activity of polysaccharide from *Dendrobium denneanum*. *Int J Biol Macromol*. 2009;45(2):169–173. <https://doi.org/10.1016/j.ijbiomac.2009.04.019>
4. Jin Q, Jiao C, Sun S, Song C, Cai Y, Lin Y, et al. Metabolic analysis of medicinal *Dendrobium officinale* and *Dendrobium huoshanense* during different growth years. *PLoS One*. 2016;11(1):e0146607. <https://doi.org/10.1371/journal.pone.0146607>
5. Tian CC, Zha XQ, Luo JP. A polysaccharide from *Dendrobium huoshanense* prevents hepatic inflammatory response caused by carbon tetrachloride. *Biotechnol Biotechnol Equip*. 2015;29(1):132–138. <https://doi.org/10.1080/13102818.2014.987514>
6. Si HY, Chen NF, Chen ND, Huang C, Li J, Wang H. Structural characterisation of a water-soluble polysaccharide from tissue-cultured *Dendrobium huoshanense* C. Z. Tang et S. J. Cheng. *Nat Prod Res*. 2018;32(3):252–260. <https://doi.org/10.1080/14786419.2017.1350670>
7. Kudo Y, Tanaka A, Yamada K. Dendrobine, an antagonist of beta-alanine, taurine and of presynaptic inhibition in the frog spinal cord. *Br J Pharmacol*. 1983;78(4):709–715. <https://doi.org/10.1111/j.1476-5381.1983.tb09424.x>
8. Li R, Liu T, Liu M, Chen F, Liu S, Yang J. Anti-influenza A virus activity of dendrobine and its mechanism of action. *J Agric Food Chem*. 2017;65(18):3665–3674. <https://doi.org/10.1021/acs.jafc.7b00276>
9. Kende AS, Bentley TJ, Mader RA, Ridge D. A simple total synthesis of (plus or minus)-dendrobine. *J Am Chem Soc*. 1974;96(13):4332–4334. <https://doi.org/10.1021/ja00820a052>
10. Yamada K, Suzuki M, Hayakawa Y, Aoki K, Nakamura H. Total synthesis of (+)-dendrobine. *J Am Chem Soc*. 1972;94(23):8278–8280. <https://doi.org/10.1021/ja00778a083>
11. Lynch VM, Li W, Martin SF, Davis BE. The structure of a key intermediate in the total synthesis of dendrobine. *Acta Crystallogr C*. 1991;47(Pt 2):439–440. <https://doi.org/10.1107/S0108270190007594>
12. Kreis LM, Carreira EM. Total synthesis of (–)-dendrobine. *Angew Chem Int Ed Engl*. 2012;51(14):3436–3439. <https://doi.org/10.1002/anie.201108564>
13. Li Q, Ding G, Li B, Guo SX. Transcriptome analysis of genes involved in dendrobine biosynthesis in *Dendrobium nobile* Lindl. infected with mycorrhizal fungus MF23 (*Mycena* sp.). *Sci Rep*. 2017;7(1):316. <https://doi.org/10.1038/s41598-017-00445-9>
14. Yamazaki M, Matsuo M, Arai K. Biosynthesis of dendrobine. *Chem Pharm Bull*. 1966;14(9):1058–1059. <https://doi.org/10.1248/cpb.14.1058>
15. Guo X, Li Y, Li C, Luo H, Wang L, Qian J, et al. Analysis of the *Dendrobium officinale* transcriptome reveals putative alkaloid biosynthetic genes and genetic markers. *Gene*. 2013;527(1):131–138. <https://doi.org/10.1016/j.gene.2013.05.073>
16. Yan L, Wang X, Liu H, Tian Y, Lian J, Yang R, et al. The genome of *Dendrobium officinale* illuminates the biology of the important traditional Chinese orchid herb. *Mol Plant*. 2015;8(6):922–934. <https://doi.org/10.1016/j.molp.2014.12.011>
17. Han R, Rai A, Nakamura M, Suzuki H, Takahashi H, Yamazaki M, et al. De novo deep transcriptome analysis of medicinal plants for gene discovery in biosynthesis of plant natural products. *Methods Enzymol*. 2016;576:19–45. <https://doi.org/10.1016/bs.mie.2016.03.001>
18. Schena M, Shalon D, Davis RW, Brown PO. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science*. 1995;270(5235):467–470. <https://doi.org/10.1126/science.270.5235.467>
19. Luo R, Liu B, Xie Y, Li Z, Huang W, Yuan J, et al. Erratum: SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. *GigaScience*. 2015;4:30. <https://doi.org/10.1186/s13742-015-0069-2>
20. Simpson JT, Wong K, Jackman SD, Schein JE, Jones SJ, Birol I. ABySS: a parallel assembler for short read sequence data. *Genome Res*. 2009;19(6):1117–1123.

<https://doi.org/10.1101/gr.089532.108>

21. Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, et al. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotechnol.* 2011;29(7):644–652. <https://doi.org/10.1038/nbt.1883>
22. Kanehisa M, Goto S, Sato Y, Furumichi M, Tanabe M. KEGG for integration and interpretation of large-scale molecular data sets. *Nucleic Acids Res.* 2012;40:D109–D114. <https://doi.org/10.1093/nar/gkr988>
23. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics.* 2014;30(15):2114–2120. <https://doi.org/10.1093/bioinformatics/btu170>
24. Stanke M, Waack S. Gene prediction with a hidden Markov model and a new intron submodel. *Bioinformatics.* 2003;19(2 suppl):ii215–ii225. <https://doi.org/10.1093/bioinformatics/btg1080>
25. Magrane M, UniProt C. UniProt knowledgebase: a hub of integrated protein data. *Database.* 2011;2011:bar009.
26. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet.* 2000;25(1):25–29. <https://doi.org/10.1038/75556>
27. Tatusov RL, Fedorova ND, Jackson JD, Jacobs AR, Kiryutin B, Koonin EV, et al. The COG database: an updated version includes eukaryotes. *BMC Bioinformatics.* 2003;4:41. <https://doi.org/10.1186/1471-2105-4-41>
28. Conesa A, Gotz S, Garcia-Gomez JM, Terol J, Talon M, Robles M. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics.* 2005;21(18):3674–3676. <https://doi.org/10.1093/bioinformatics/bti610>
29. Li B, Dewey CN. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics.* 2011;12:323. <https://doi.org/10.1186/1471-2105-12-323>
30. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics.* 2010;26(1):139–140. <https://doi.org/10.1093/bioinformatics/btp616>
31. Faircloth BC. MSATCOMMANDER: detection of microsatellite repeat arrays and automated, locus-specific primer design. *Mol Ecol Resour.* 2008;8(1):92–94. <https://doi.org/10.1111/j.1471-8286.2007.01884.x>
32. Han R, Takahashi H, Nakamura M, Yoshimoto N, Suzuki H, Shibata D, et al. Transcriptomic landscape of *Pueraria lobata* demonstrates potential for phytochemical study. *Front Plant Sci.* 2015;6:426. <https://doi.org/10.3389/fpls.2015.00426>
33. Fu L, Niu B, Zhu Z, Wu S, Li W. CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics.* 2012;28(23):3150–3152. <https://doi.org/10.1093/bioinformatics/bts565>
34. Serres-Giardi L, Belkhir K, David J, Glemin S. Patterns and evolution of nucleotide landscapes in seed plants. *Plant Cell.* 2012;24(4):1379–1397. <https://doi.org/10.1105/tpc.111.093674>
35. Tian T, Liu Y, Yan H, You Q, Yi X, Du Z, et al. agriGO v2.0: a GO analysis toolkit for the agricultural community, 2017 update. *Nucleic Acids Res.* 2017;45(W1):W122–W129. <https://doi.org/10.1093/nar/gkx382>
36. Tang Z, Cheng S. A study on the raw plants for the Chinese traditional medicine “Huoshan Shi-hu”. *Bulletin of Botanical Research.* 1984;4(3):141–145.