

## Flavonoid accumulation and identification of flavonoid biosynthesis genes in *Dimocarpus longan* Lour. by transcriptome sequencing

Wei ZHENG<sup>1\*</sup>, Ziwei ZHANG<sup>1</sup>, Xuefei YU<sup>2</sup>,  
Xueming DONG<sup>1</sup>, Qiuying ZHANG<sup>1</sup>

<sup>1</sup>Harbin University of Commerce, School of Pharmacy, Research Center of Pharmaceutical Engineering Technology, No. 138, Tongda Street, Daoli District, Harbin, 150076, China; [wei2013zheng@163.com](mailto:wei2013zheng@163.com) (\*corresponding author); [zww19951216@163.com](mailto:zww19951216@163.com); [xm\\_2019Dong@163.com](mailto:xm_2019Dong@163.com); [Qiuying88Zhang@163.com](mailto:Qiuying88Zhang@163.com)

<sup>2</sup>Harbin University of Commerce, Library, No. 1, Xuehai Street, Songbei District, Harbin, 150028, China; [yuxuefei83@163.com](mailto:yuxuefei83@163.com)

### Abstract

*Dimocarpus longan* Lour. is widely cultivated and is very popular around the world. Its by-products such as roots and leaves have been used as traditional Chinese medicines due to their content of important secondary metabolites, especially flavonoids. However, the economic value and application of *D. longan* roots and leaves are limited because they accumulate relatively low levels of flavonoids. Therefore, it is important to find key genes that regulate the accumulation of the predominant flavonoid compounds in *D. longan* roots and leaves. Here, we have used RNA-sequencing to describe the transcriptome of *D. longan*. We obtained 75,229,529 raw reads and 15.04 GB of clean data, generating 56,055 unigenes (N50 = 1,583 nt, mean length = 829.61 nt). Next, we annotated these unigenes using the various available bioinformatics databases. By this approach, we identified 6,684 genes differentially expressed between root and leaf tissues, of which thirteen were identified as flavonoid biosynthesis genes. Of these, eight genes were much highly expressed in roots (*DIC4H*, *DIHCT*, *DIDFR*, *DLANS*, *DLANR*, *DICH5*, *DIF3'H*, and *DIF3H*), and two were much highly expressed in leaves (*DILAR* and *DIFLS*). The contents of thirteen flavonoids in *D. longan* roots and leaves were measured by LC-MS, and epicatechin was found to be the predominant flavonoid in both tissues, which was significantly higher than the other flavonoids measured in the study. Its contents were 213,773.65 ng/g in roots and 22,388.71 ng/g in leaves. Our findings will facilitate efforts to increase the economic value and expand the applications of *D. longan* roots and leaves by means of genetic engineering.

**Keywords:** flavonoid; gene expression; LC-MS; *D. longan*; transcriptome sequencing

### Introduction

*Dimocarpus longan* Lour., a member of the Sapindaceae family, is a well-known tropical and subtropical fruit tree that has been widely cultivated in Southeast Asia, especially in China, Vietnam, and Thailand (Lin *et al.*, 2020). Due to its delicate flavour and sweet taste, the fruit of *D. longan* has been extensively accepted by consumers and has enjoyed great popularity around the world (Tseng *et al.*, 2014). Most tissues of *D. longan*

(fruit, leaf, flower, root, and seed) contain abundant bioactive compounds such as flavonoids, phenolic acids, and polysaccharides, and they have been used as traditional Chinese medicines for more than a century (Tang *et al.*, 2019). The use of *D. longan* as a traditional Chinese medicine has been associated with improved blood metabolism, increased immunity, relief of insomnia, and improvements in learning and memory (Yang *et al.*, 2008; Park *et al.*, 2010).

Flavonoids are an abundant class of plant secondary metabolites. More than 10,000 plant flavonoids have been identified, and antiviral, antibacterial, estrogenic, and anti-obese properties have been described for many of these compounds (Ferrer *et al.*, 2008; Tang *et al.*, 2009; Wilson *et al.*, 2009; Fang, 2014). In plants, flavonoids contribute to plant growth and development, reproduction, and stress responses. They are categorized into six subclasses based on their structural properties: flavones, flavanones, flavonols, isoflavones, proanthocyanidins, and anthocyanins (Dixon and Pasinetti, 2010; Yao *et al.*, 2018). Mechanisms of flavonoid accumulation have been characterized in numerous plants, including *Salvia miltiorrhiza*, *Dendrobium officinale*, and *Syringa oblata* (Deng *et al.*, 2018; Liu *et al.*, 2019; Yuan *et al.*, 2020). Because of the functional importance of flavonoids, much research effort is being aimed at cataloguing the genes responsible for plant flavonoid biosynthesis. Numerous enzymes encoded by plant flavonoid biosynthesis genes have been characterized (Petrucci *et al.*, 2013; Saito *et al.*, 2013; Dastmalchi and Dhaubhadel, 2015; Deng *et al.*, 2018; Ni *et al.*, 2018; Yuan *et al.*, 2020). However, flavonoid biosynthesis genes from *D. longan* have not yet been studied.

Because of its accuracy, throughput, and reproducibility, RNA sequencing (RNA-Seq) is widely used for the study of transcriptomes (Lei *et al.*, 2018). In plants, RNA-Seq data has been used to greatly advance our understanding of gene transcription patterns, posttranscriptional modification and mutation, functional analysis, and gene regulatory networks (Wang *et al.*, 2010; Glenn, 2011). To date, we still know little about the expression of flavonoid biosynthesis genes in its roots and leaves. Here, we use RNA-Seq to identify genes differently expressed between the roots and leaves of *D. longan*. We also identified genes associated with flavonoid biosynthesis: *DIC4H*, *DIHCT*, *DIDFR*, *DLANS*, *DLANR*, *DICHS*, *DIF3' H*, and *DIF3H*. Genes identified as differentially expressed by RNA-Seq were then validated by quantitative real time PCR (qRT-PCR). Also, we measured the flavonoid contents of roots and leaves by LC-MS in order to study the relationship between flavonoid accumulation and biosynthesis gene expression. To our knowledge, no other study has applied RNA-Seq to investigate flavonoid biosynthesis gene expression in the roots and leaves of *D. longan*. Our results will aid future efforts to increase flavonoid accumulation in *D. longan* through genetic engineering.

## Materials and Methods

### *Plant materials*

*D. longan* was cultivated in a greenhouse at 25 °C and 50% relative humidity. Leaves from the upper peripheral branches and roots were harvested from ten 2-month-old plants. Upon collection, samples were frozen in liquid nitrogen and stored at -80 °C.

### *RNA isolation, library construction, and RNA sequencing*

First, for each sample, RNA was extracted using a commercial kit (RN38 EASYspin plus Plant RNA Kit, Aidlab Biotech, Beijing, China). The recovered RNA was assessed (The concentration and integrity) using a UV-Vis spectrophotometer and Agilent 2100 Bioanalyzer. High-quality RNA samples were then pooled to achieve the optimal diversity of transcriptional units.

A commercially available kit (TruSeq RNA Sample Preparation Kit, Illumina Inc, San Diego, CA, USA) was then used to construct mRNA-Seq libraries. Poly-A mRNA was then isolated using magnetic beads

carrying poly-T oligos. The recovered poly-A mRNA was then fragmented using divalent cations at high temperature. First strand cDNA synthesis was done using a reverse transcriptase enzyme and random primers, and second-strand cDNA synthesis was done using DNA Polymerase I. Next, the purified short fragments were ligated to a single 'A' base and ligated to sequencing adapters. Fragments ( $\sim 200 \pm 25$  bp) were then separated by agarose gel electrophoresis and used as sequencing templates for PCR. This generated a cDNA library that was sequenced on the Illumina HiSeq™ 2000 platform.

#### *Sequence analysis and assembly*

To ensure that the data used for assembly was of sufficient quality, adaptor sequences, low-quality reads, and reads where > 10% of bases had a Q-value < 20 were removed from the raw reads (Wang *et al.*, 2010; Xie *et al.*, 2012). Sequences shorter than 60 bases were also discarded (because small reads are likely to be sequencing artifacts) (Meyer *et al.*, 2009). Trinity (<http://trinityrnaseq.sourceforge.net/>) was used to assemble clean reads into contigs, recovering full-length transcripts (Grabherr *et al.*, 2011). These contigs were then combined into transcripts based on paired-end information. The resulting transcripts were then clustered based on sequence identity. The longest transcripts classified as unigenes were then combined, giving the final assembly that could be used for functional annotation.

#### *Functional annotation*

We used multiple databases [NCBI Nr, Gene Ontology (GO), Clusters of Orthologous Group (COG), Kyoto Encyclopaedia of Genes and Genomes (KEGG), and SwissProt] to annotate the assembled sequences. Gene names were assigned to unigenes based on the top BLAST hit. Genes without descriptions were excluded. As an additional step, we submitted predicted protein sequences to BLASTX (searching against GenBank, NCBI Nt, PDB, and RefSeq). The GetORF function of EMBOSS was used to predict open reading frames (ORFs) and the longest of there were extracted for each unigene (Hernández *et al.*, 2018).

The Swiss-Prot BLAST results were imported into Blast2GO to assign GO terms and putative gene functions (Conesa *et al.*, 2005; Conesa and Gotz, 2008). ANNEX was then applied to enrich and refine these annotations (Ye *et al.*, 2006; Myhre *et al.*, 2006).

To further annotate gene functions, unigenes were aligned to COG database (Galperin *et al.*, 2019), and the KEGG pathways were assigned using the online server (<http://www.genome.jp/kegg/kaas/>). KEGG Orthology (KO) assignments were made using the bi-directional best hit method (Moriya *et al.*, 2007). The KEGG analysis outputs included KO assignments and KEGG pathways.

#### *Detection of simple sequence repeat (SSR) markers*

MISA software (<http://pgrc.ipk-gatersleben.de/misa/>) was used to detect SSR markers in the assembled sequences (those longer than 1 kb). The parameters were optimized to identify perfect dinucleotide motifs (minimum of six repeats, as well as tri-, tetra-, penta- and hexa-nucleotide motifs with a minimum of five repeats) (Wei *et al.*, 2011).

#### *Quantitative Real Time PCR analysis (qRT-PCR)*

Total RNA was extracted from *D. longan* root and leaf tissues using the cetyltrimethylammonium bromide (CTAB) method (Jordon-Thaden *et al.*, 2015). The cDNA Synthesis SuperMix (TransGen Biotech, Beijing, China) and TransScript® One-Step gDNA Removal kits were used to synthesize first strand cDNA. All qRT-PCR reactions were done using TransStart® Top Green qPCR SuperMix (TransGen Biotech, Beijing, China) according to the manufacturer's instructions. The primers used for qRT-PCR are given in Table 1, and the *D. longan* tubulin gene was selected as a reference gene. The qRT-PCR cycling parameters were: 95 °C for 1 min, followed by 95 for 5 s, 60 °C for 30 s, and 72 °C for 30 s. All experiments were conducted in triplicate.

Relative gene expression was calculated using the relative quantification ( $2^{-\Delta\Delta CT}$ ) method (Schmittgen and Livak, 2008).

**Table 1.** Primer sequences for qRT-PCR

Gene	Primer sequence	Gene	Primer sequence
<i>DIC4H</i>	QF: GTCTTGTCATCCCCTGTGA QR: TGCCACCCAAGTTCTACACA	<i>DICH5</i>	QF: ACAGCTGTGATCTCGGAACA QR: CAAGCGCCTCATGATGTACC
<i>DIHCT</i>	QF: TGACCCGTGACCAGCTAAAT QR: ATGCAGACCTCCACACATGA	<i>DIF3'H</i>	QF: AAATCGATGGTGGTGGAGGT QR: GCTTCTTCATCTTAGCCGCC
<i>DIDFR</i>	QF: TCGTTTTTCAGGGTCCTTGGA QR: GCTGATTTGGCTGATGAGGG	<i>DIFLS</i>	QF: TGGAGAGTGTACAAGCCGTT QR: TTGCAGGCTGTCTCTCTCA
<i>DLANS</i>	QF: CGATGCTTGTGTGACAGCTT QR: TAGGGGCAGAAAGGAAGCAA	<i>DICCoAOMT</i>	QF: TTGGAGAAGGCTTTTGTGGC QR: ACGCTCAACACACAACCAAA
<i>DLANR</i>	QF: AGTGTGAACAGCATAGCCCT QR: TGTAGTTGGTGGCACTGGAT	<i>DICYP</i>	QF: TCGGAAACCTCTACGACGTC QR: CCGAAATGATTGGGCGGTAG
<i>DILAR</i>	QF: TTGTGACAGAGGCCAGCATA QR: AAGGGTTGCTGGTGTGGTA	<i>DIF3H</i>	QF: CGATGAGGACCCGAGACTAC QR: CCTCCTTCTCCAACCCATT
<i>DICHI</i>	QF: CAACTCCCCTCAGTGTCACT QR: TAGTGAGACCGAGGTGAGGA	<i>Tubulin</i>	QF: CTCATGTATGCCAAGCGTGC QR: CTCTGCAGACTCAGACCAA

#### *Identification and quantification of individual flavonoid by Liquid Chromatography-Mass Spectrometry (LC-MS)*

Root and leaf tissues from *D. longan* were freeze-dried at -80 °C for 48 h, then ground into a powder using a mortar and pestle. An aliquot of 50 mg of dried sample was transferred to a microcentrifuge tube (1.5 mL), and two steel balls added to the tube. Then 80% methanol solution (1.2 mL, containing 0.01 mol L<sup>-1</sup> BHT and 0.1% formic acid) was added to each tube, and the tissue was ground at 60 Hz for 2 min in a grinder. The mixture was ultrasonicated at 50 °C for 2 h, then centrifuged at 12,000 rpm for 10 min at room temperature. The supernatant was transferred to a glass sampling vial. Later, 1 mL of 80% methanol solution (containing 0.01 mol L<sup>-1</sup> BHT and 0.1% formic acid) was added to the residue, and the mixture was ultrasonicated at 50 °C for 30 min, followed by centrifugation at 12,000 rpm for 10 min at ambient temperature. The two supernatants were combined and dried at room temperature. An aliquot of 200 µL of 35% methanol/ 65% water were added to dissolve the dried supernatants. The resulting mixture was vortexed vigorously and then centrifuged for 10 min. The supernatants were placed at room temperature for 30 min before LC-MS analysis. Flavonoids were separated on a Waters ACQUITY UPLC HSS T3 (100 × 2.1 mm, 1.8 µm) with a Thermo UHPLC system. The mobile phase A consisted of 0.1% formic acid and water, and the mobile phase B consisted of methanol. The flow rate was maintained at 0.2 mL min<sup>-1</sup>, and the injection volume was 5 µL. The concentrations of flavonoid compounds were determined from a standard curve. All samples were analysed in triplicate. Values are displayed as means ± SDs.

## Results

### *D. longan* RNA-Seq and transcriptome assembly

To comprehensively analyse transcript diversity and gene expression, RNA samples from *D. longan* leaves and roots (three replicate samples of each tissue) were sequenced on the Illumina HiSeq™ 2000 platform. We obtained 75,229,529 raw reads and 15.04 GB of clean data. The mean GC content was 45.84%. Q value ≥ 20 was the basic requirement for all reads, and greater than 85.03% of raw reads showed Q ≥ 30, indicating that the reads were of acceptable quality for assembly. After a stringent filtering step to remove contaminated, low-quality, and short reads, 15,397,494 contigs were assembled (Li *et al.*, 2010). Table 2 summarizes the sequencing pipeline. The high-quality reads have been submitted to the NCBI SRA database

(SRP155595). Using Trinity, short-read sequences were assembled into 106,991 transcripts (N50, 2100 nt; mean length, 1249.38 nt). All assembled transcripts were clustered to produce 56,055 unigenes with an N50 of 1,583 nt and a mean length of 829.61 nt. Among these unigenes, 14,260 were longer than 1000 nt. Figure 1 shows the length distributions of contigs, transcripts, and unigenes.

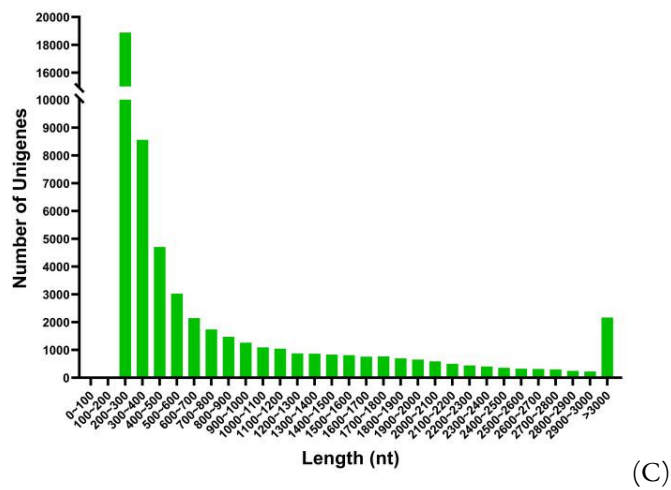
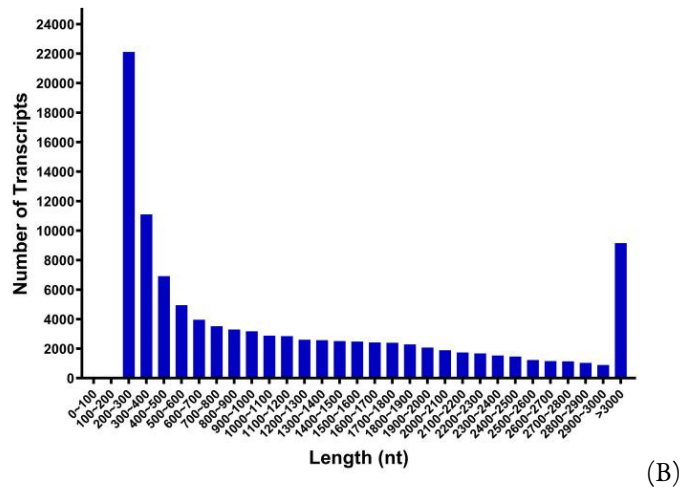
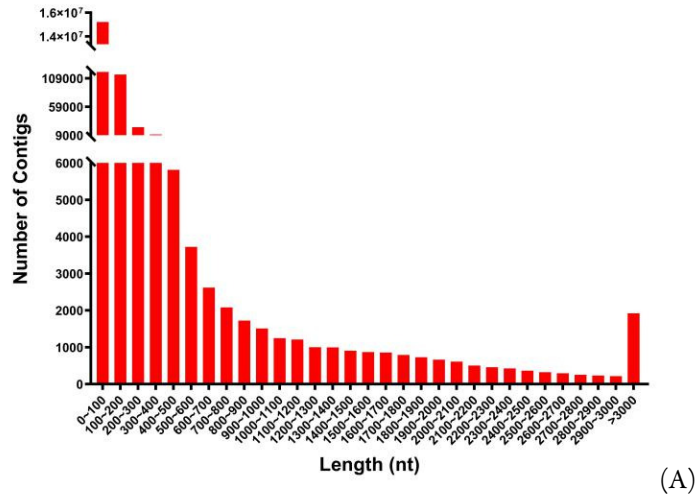
**Table 2.** Summary of sequencing and assembly of the *D. longan* transcriptome

Length Range	Contig	Transcript	Unigene
200-300	15,354,791(99.72%)	22,111(20.67%)	18,887(33.69%)
300-500	16,186(0.11%)	18,019(16.84%)	13,260(23.66%)
500-1000	11,655(0.08%)	18,895(17.66%)	9,648(17.21%)
1000-2000	9,270(0.06%)	25,084(23.44%)	8,392(14.97)
2000+	5,592(0.04%)	22,882(21.39%)	5,868(10.47%)
Total Number	15,397,494	106,991	56,055
Total Length	580,611,316	133,672,267	46,503,766
N50 Length	36	2,100	1,583
Mean Length	37.71	1249.38	829.61

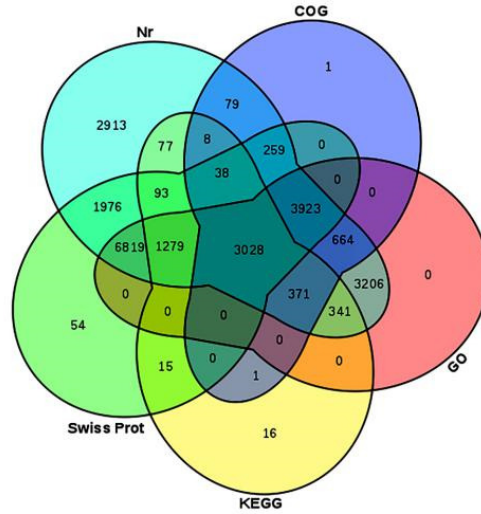
#### *Functional annotation*

The assembled sequences were annotated in multiple ways. In total, 44.9% of the unigenes (25,161) were annotated based on similarity to sequences deposited in diverse databases, as described in the Methods. Of these, 3,028 unigenes were successfully annotated by all the databases used in this study (Figure 2 and Table 3). First, the Basic Local Alignment Search Tool (BLAST) tool was used to search the NCBI Nr and Swiss-Prot protein databases ( $E \leq 10^{-5}$ ). In total, 25,074 and 17,484 unigenes were homologous to sequences in the Nr and Swiss-Prot databases ( $E \leq 10^{-5}$ ), accounting for 44.7% and 31.2% of the total assembled unigenes, respectively. Additionally, 8,372 (14.9%), 19,631 (35.0%), and 5,267 (9.4%) of the unigenes matched entries in the COG, GO, and KEGG databases, respectively.

Among the matches to the Nr database, 62.0% of the aligned unigenes had an E-value of  $< 1.0E^{-60}$ , indicating strong sequence homology. The remaining aligned unigenes had E-values ranging between  $1.0E^{-5}$  and  $1.0E^{-60}$  (Figure 3A). The similarities of the unigenes to their top Blast hits in the Nr database was also assessed: 74.5% of the unigenes had a similarity  $> 60\%$ , 21.5% had a similarity between 40% and 60%, and 4.0% had a similarity lower than 40% (Figure 3B). Over 80% of the unigenes had top Blast hits from the following seven species: *Fragaria vesca* subsp. *vesca*, *Glycine max*, *Populus trichocarpa*, *Prunus persica*, *Ricinus communis*, *Theobroma cacao*, and *Vitis vinifera* (Figure 3C).



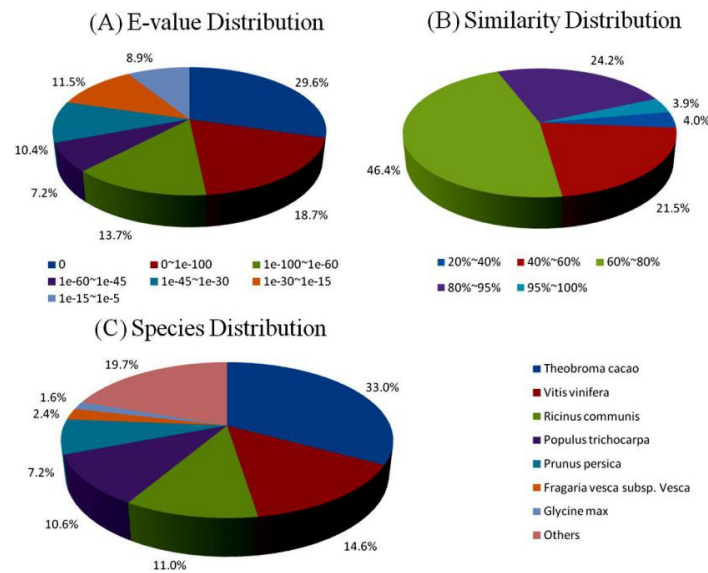
**Figure 1.** Overview of the *D. longan* transcriptome assembly. (A) Size distribution of contigs; (B) Size distribution of transcripts; (C) Size distribution of unigenes



**Figure 2.** Venn graph of the five database annotation results

**Table 3.** Functional annotation of the *D. longan* transcriptome

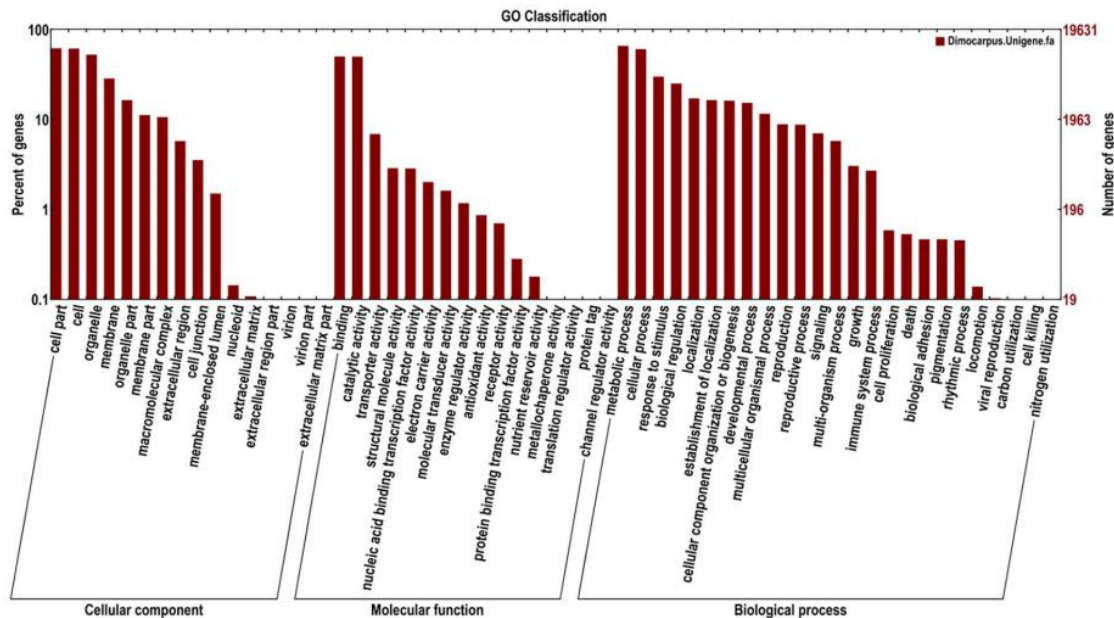
Annotated databases	Unigenes	Percentage of unigenes	≥ 300nt	≥ 1000nt
COG	8,372	14.9%	7,701	5,889
GO	19,631	35.0%	17,091	11,353
KEGG	5,267	9.4%	4,659	3,275
Swiss-Prot	17,484	31.2%	15,267	9,986
Nr	25,074	44.7%	21,487	13,341
all	25,161	44.9%	21,532	13,343



**Figure 3.** Characteristics of sequence homology of *D. longan* genes blasted against the NCBI non-redundant (Nr) database. (A) E-value distribution of BLAST hits for matched unigene sequences. (B) Similarity distribution of top BLAST hits for each unigene. (C) Species distribution of the top BLAST hits

*GO classification*

Gene Ontology (GO) analysis was used to classify the unigenes into functional groups (Hillier *et al.*, 2009). Of the 56,066 assembled unigenes, 19,631 unigenes were categorized into 57 functional groups and three major categories: 49.6% (n = 2,849) in biological process, 23.7% (n = 540) in cellular component, and 26.7% (n = 2,236) in molecular function (Figure 4). Among the biological process terms, metabolic process, cellular process, and response to stimulus were relatively well represented. Among the molecular function terms, binding, catalytic activity, and transporter activity were common. Among the cellular component terms, cell part, cell, and organelle were most abundant.

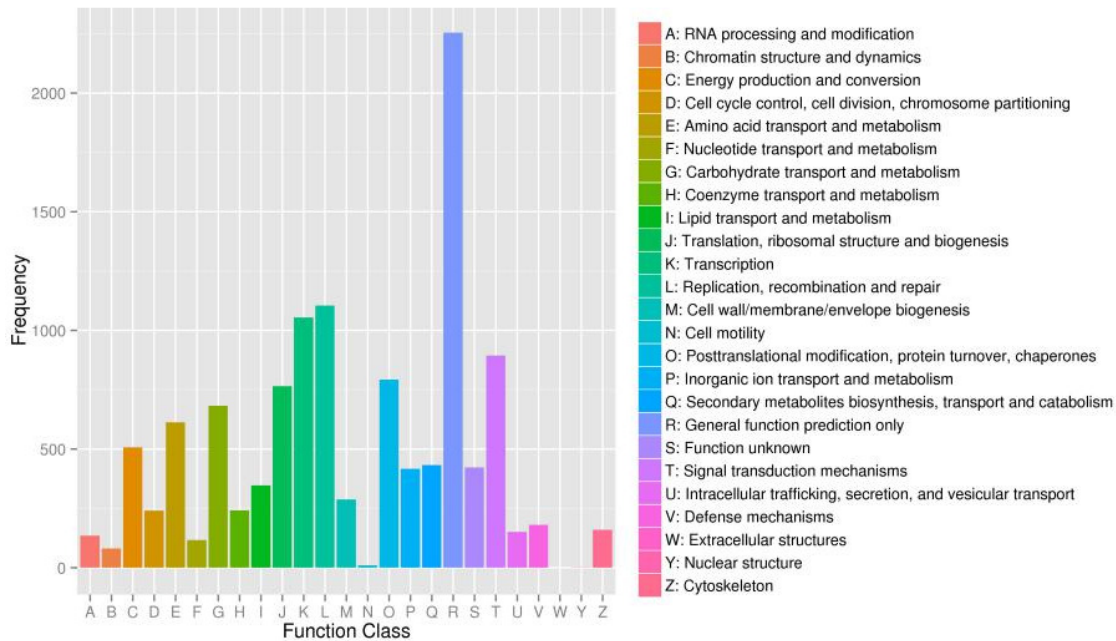


**Figure 4.** Functional annotation of assembled sequences based on gene ontology (GO) terms. The unigenes were classified into three main categories: Cellular component, Molecular function, and Biological process

*COG classification*

The COG database was used to assign potential functions to the assembled unigenes. Of the 56,055 unigenes, 8,372 (14.9%) were clustered into functional categories based on their Nr annotation (Figure 5). The largest cluster was general function prediction (2254; 18.95%), followed by replication, recombination, and repair (1105; 9.29%); transcription (1055; 8.87%); signal transduction mechanisms (894; 7.52%); posttranslational modification, protein turnover, and chaperones (793; 6.67%); translation, ribosomal structure, and biogenesis (765; 6.43%); carbohydrate transport and metabolism (682; 5.74%); amino acid transport and metabolism (613; 5.16%); energy production and conversion (507; 4.26%); and secondary metabolite biosynthesis, transport, and catabolism (432; 3.63%). Only one unigene was assigned to nuclear structure and ten to cell motility. No unigene was assigned to extracellular structures. In addition, 417 unigenes (3.51%) were assigned to inorganic ion transport and metabolism, and 347 (2.92%) were assigned to lipid transport and metabolism.





**Figure 5.** Cluster of orthologous groups (COG) classification

#### *KEGG classification*

Based on a BLASTX search of assembled unigenes against the KEGG protein database, 3,655 unigenes were assigned to 111 pathways. The most common KEGG pathway was ribosome (ko03010), containing 349 (9.55%) of the unigenes, followed by protein processing in the endoplasmic reticulum (ko04141) (186, 5.09%), glycolysis/gluconeogenesis (ko00010) (174, 4.76%), oxidative phosphorylation (ko00190) (171, 4.68%), plant hormone signal transduction (ko04075) (163, 4.46%), RNA transport (ko03013) (163, 4.46%), spliceosome (ko03040) (149, 4.08%), purine metabolism (ko00230) (133, 3.64%), plant-pathogen interaction (ko04626) (130, 3.57%), starch and sucrose metabolism (ko00500) (128, 3.50%), cysteine and methionine metabolism (ko00270) (119, 3.26%), pyruvate metabolism (ko00620) (103, 2.82%), and carbon fixation in photosynthetic organisms (ko00710) (101, 2.76%).

#### *Identification of SSRs*

The MISA web tool (MIcroSATellite, <http://pgrc.ipk-gatersleben.de/misa/>) was used to identify SSRs in 14,260 unigenes. We confirmed the presence of 9,066 putative microsatellites in 6,042 (42.37%) of the examined unigenes. Six types of SSRs were observed, and 2,030 unigenes contained more than one SSR. The mononucleotide repeat motif was most common, accounting for 52.78% of the SSRs, followed by dinucleotide (26.72%), trinucleotide (19.67%), and tetranucleotide (0.68%) motifs. Penta- and hexanucleotide motifs each accounted for 0.15% of the SSRs (Table 4).

#### *Differential gene expression in D. longan root and leaf tissues*

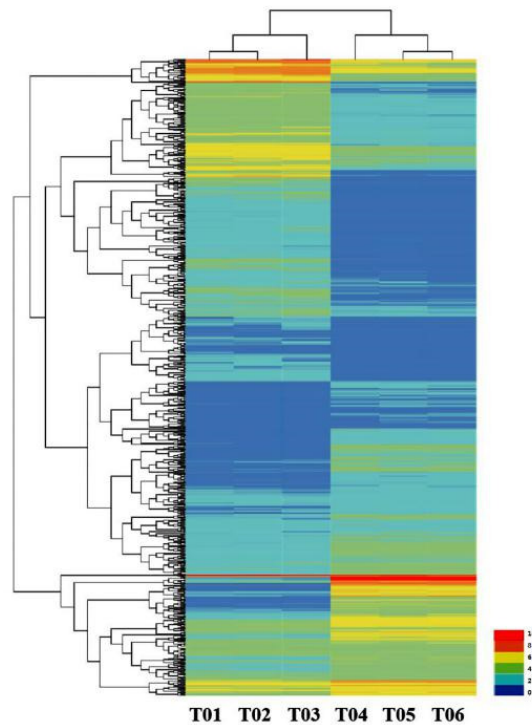
In total, 6,684 genes were differentially expressed between root and leaf tissues, and their expression levels are presented as a clustered heat map in Figure 6. Among the differentially expressed genes (DEGs), 3,550 genes had higher expression in roots, and 3,134 genes had higher expression in leaves (Figure 7).

The DEGs were analyzed based on their functions in diverse pathways. These pathways could be divided into five categories: genetic information processing, organismal systems, cellular processes, environmental information processing, and metabolism (Figure 8).

**Table 4.** Summary of simple sequence repeat (SSR) types in *D. longan*

Searching Item	Numbers
Total number of sequences examined	14,260
Total size of examined sequences (bp)	30,117,747
Total number of identified SSRs	9,066
Number of SSR containing sequences	6,042
Number of sequences containing more than 1 SSR	2,030
Number of SSRs present in compound formation	896
Mono nucleotide	4,785
Di nucleotide	2,422
Tri nucleotide	1,783
Tetra nucleotide	62
Penta nucleotide	7
Hexa nucleotide	7

One-hundred-thirty-one DEGs were related to ribosomes in genetic information processing, accounting for the largest percentage in this category. Thirty-four DEGs were related to plant-pathogen interactions in the organismal systems category. In the cellular process's category, 20 and 16 DEGs were associated with the peroxisome and the phagosome, respectively. Only one function (plant hormone signal transduction) was identified in the environmental information processing category (n = 63 DEGs). In the metabolism category, 53 DEGs were annotated as having a role in glycolysis/gluconeogenesis, followed by fewer genes related to purine metabolism, pyrimidine metabolism, carbon fixation in photosynthetic organisms, and other functions. Given that flavonoids are a principal group of secondary metabolites in *D. longan*, we next studied putative flavonoid biosynthesis genes and their relationships with flavonoid contents in root and leaf tissues.



**Figure 6.** Hierarchical clustering of transcript levels in root and leaf tissues. T01, T02, and T03 are 3 replicates from *D. longan* roots, and T04, T05, and T06 are 3 replicates from *D. longan* leaves

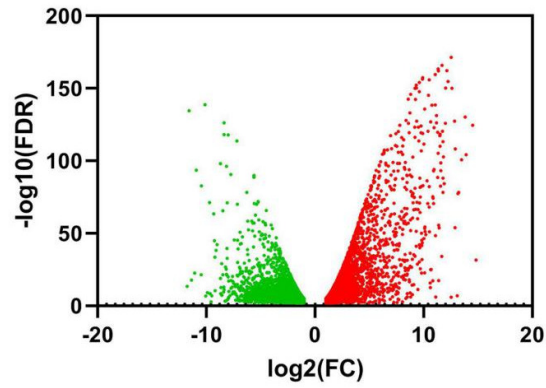


Figure 7. Volcano plot of differentially expressed genes in root and leaf tissues of *D. longan*

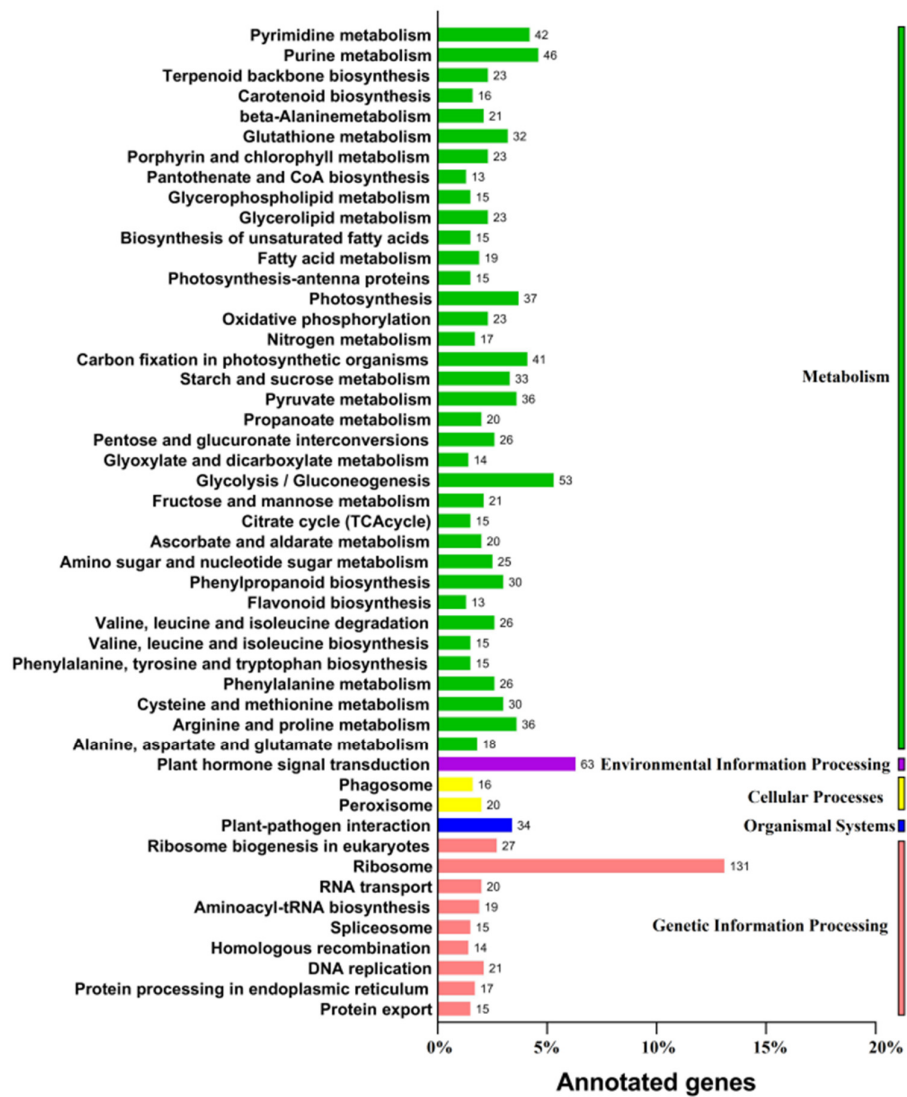


Figure 8. Percentage of annotated genes from various pathways and processes that were differentially expressed between root and leaf tissues

*Putative flavonoid biosynthesis genes*

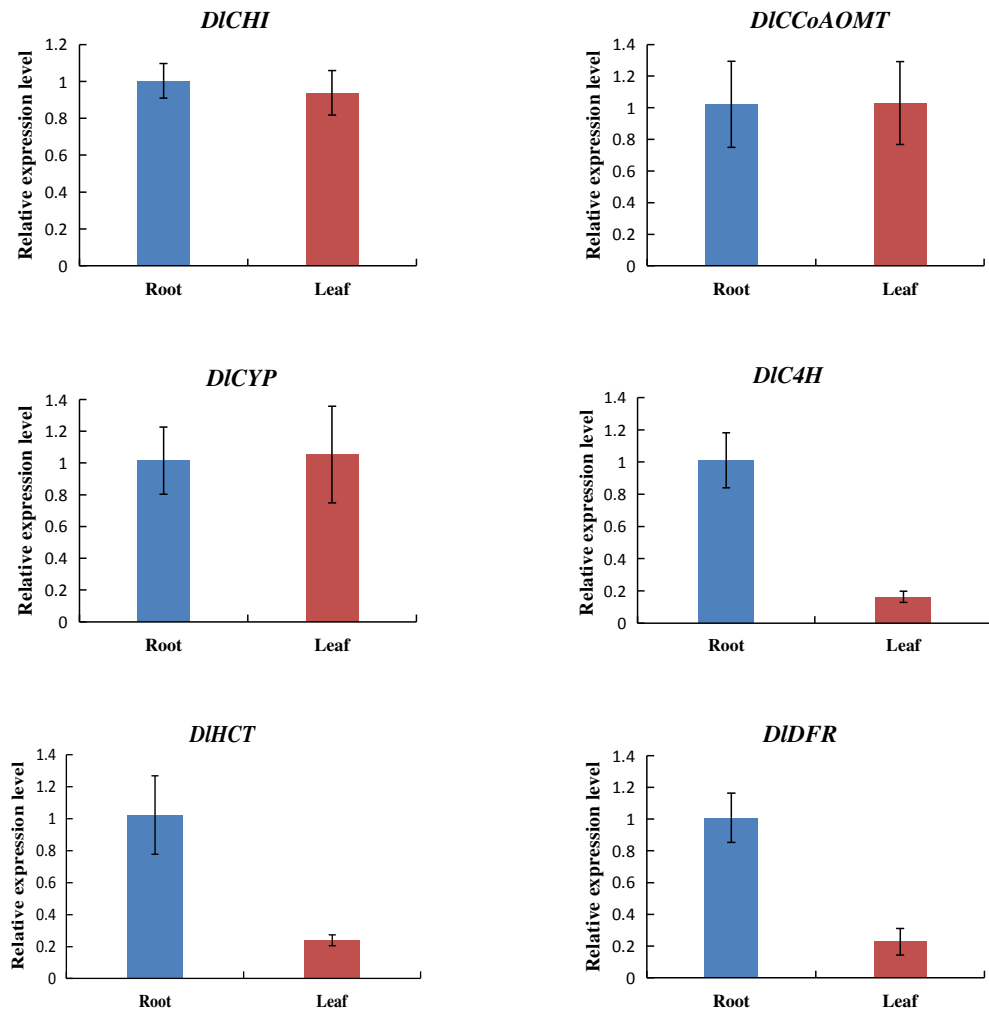
Using the NCBI Nr and KEGG annotations, 13 flavonoid biosynthesis genes were identified: *DIC4H* (536 aa), *DIHCT* (430 aa), *DIDFR* (353 aa), *DLANS* (357 aa), *DLANR* (336 aa), *DILAR* (343 aa), *DICHI* (244 aa), *DICHS* (393 aa), *DIF3' H* (515 aa), *DIFLS* (335 aa), *DICCoAOMT* (247 aa), *DICYP* (508 aa), and *DIF3H* (365 aa). The sequence analysis results shown in Table 5 indicate that flavonoid biosynthesis genes from *D. longan* exhibit high percentage identities with other orthologous genes.

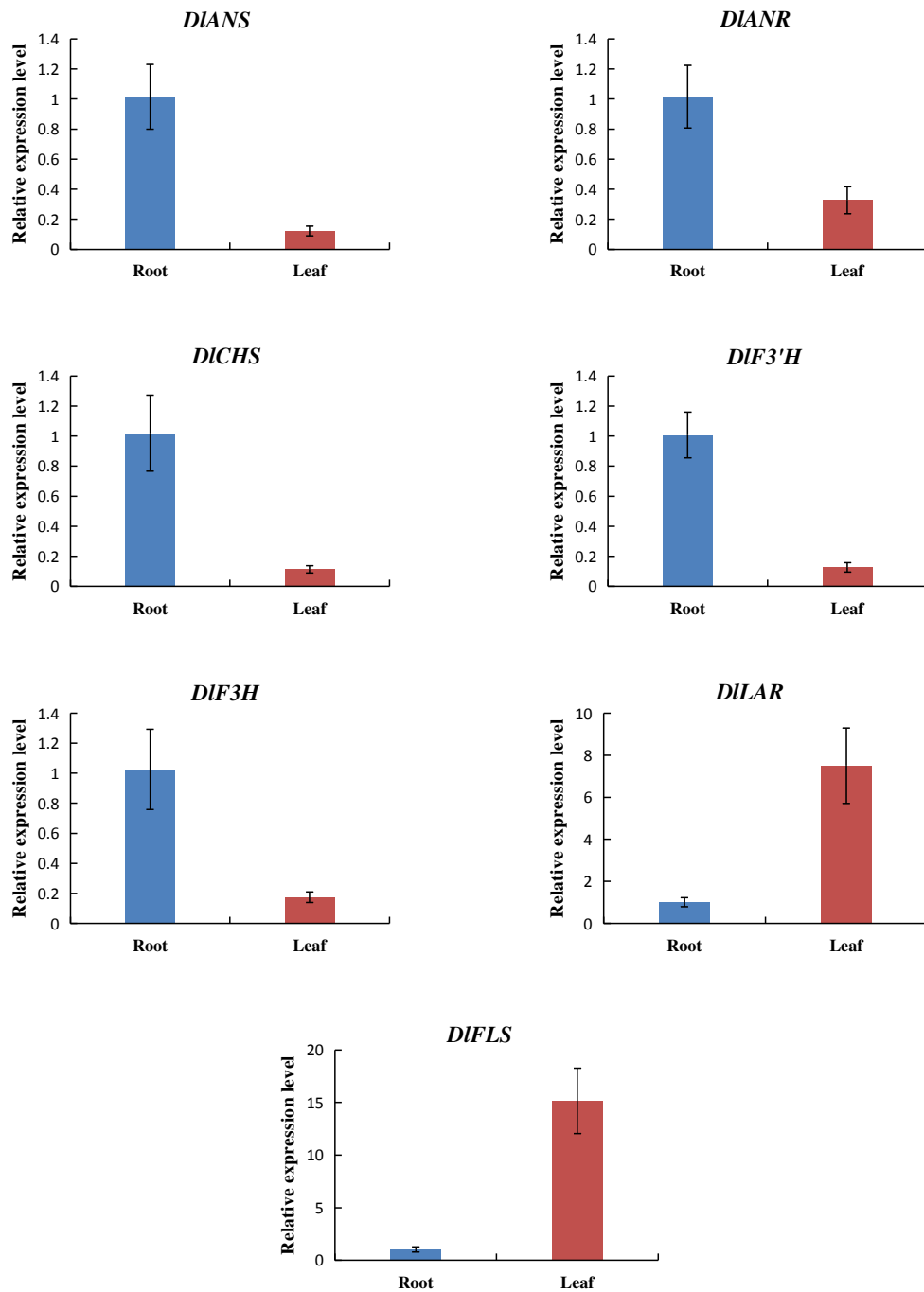
**Table 5.** Comparison of flavonoid biosynthetic genes of *D. longan* with the most orthologous genes

Gene	Length (aa)	Orthologous genes (Accession No.)	Identity (%)
<i>DIC4H</i>	536	<i>Pistacia vera C4H</i> (XM_031412360.1)	82.77
		<i>Citrus clementina C4H</i> (XM_024178714.1)	81.25
		<i>Citrus sinensis C4H1</i> (NM_001288840.1)	81.25
<i>DIHCT</i>	430	<i>Citrus clementina HCT</i> (XM_006427596.2)	82.07
		<i>Citrus sinensis HCT</i> (XM_006493322.3)	81.77
		<i>Pistacia vera HCT</i> (XM_031417710.1)	82.43
<i>DIDFR</i>	353	<i>Acer palmatum DFR</i> (MF460441.1)	86.69
		<i>Populus euphratica DFR</i> (XM_011030570.1)	78.40
		<i>Populus trichocarpa DFR</i> (XM_002300723.3)	77.80
<i>DLANS</i>	357	<i>Litchi chinensis ANS</i> (HQ402913.1)	96.32
		<i>Acer palmatum ANS</i> (MF460440.1)	86.79
		<i>Pistacia vera ANS</i> (XM_031428804.1)	83.65
<i>DLANR</i>	336	<i>Pistacia vera ANR</i> (XM_031427444.1)	83.90
		<i>Pistacia vera ANR</i> (XM_031391609.1)	84.24
		<i>Mangifera indica ANR1-2</i> (MG322128.1)	83.97
<i>DILAR</i>	343	<i>Carya illinoensis LAR</i> (MH613771.1)	80.44
		<i>Populus trichocarpa LAR</i> (XM_024611241.1)	77.08
		<i>Hevea brasiliensis LAR</i> (XM_021826359.1)	77.52
<i>DICHI</i>	244	<i>Acer x freemanii CHI</i> (KC255248.1)	89.40
		<i>Acer x freemanii CHI</i> (KC121346.1)	91.58
		<i>Canarium album</i> cultivar 'Changying' <i>CHI</i> (JF728821.1)	84.54
<i>DICHS</i>	393	<i>Litchi chinensis CHS</i> (GU288819.1)	96.78
		<i>Litchi chinensis CHS</i> (GU288820.1)	97.01
		<i>Dysoxylum binectariferum CHS</i> (KR703480.1)	83.42
<i>DIF3'H</i>	515	<i>Litchi chinensis F3'H</i> (KY302362.1)	95.70
		<i>Pistacia vera F3'H</i> (XM_031402999.1)	79.20
		<i>Pistacia vera F3'H</i> (XM_031419916.1)	79.13
<i>DIFLS</i>	335	<i>Pistacia vera FLS</i> (XM_031431135.1)	78.97
		<i>Citrus unshiu FLS</i> (AB011796.1)	76.09
		<i>Citrus clementina FLS</i> (XM_006426379.2)	75.99
<i>DICCoAOMT</i>	247	<i>Citrus sinensis CCoAOMT</i> (XM_006486982.3)	86.40
		<i>Citrus clementina CCoAOMT</i> (XM_006422921.2)	87.52
		<i>Pistacia vera CCoAOMT</i> (XM_031400822.1)	86.75
<i>DICYP</i>	508	<i>Pistacia vera CYP</i> (XM_031416202.1)	83.86
		<i>Citrus clementina CYP</i> (XM_006430432.2)	83.70
		<i>Citrus sinensis CYP</i> (XM_006481963.2)	82.11
<i>DIF3H</i>	365	<i>Litchi chinensis F3H</i> (HQ402912.1)	95.03
		<i>Pistacia vera F3H</i> (XM_031425041.1)	85.23 85.05

*Gossypium arboreum* F3H (XM\_017782525.1)*Expression analysis of flavonoid biosynthesis genes in roots and leaves of D. longan*

Based on RNA-Seq data from root and leaf tissues, we found that the thirteen putative flavonoid biosynthesis genes had different expression levels. qRT-PCR was used to verify the accuracy of the transcriptome data for the 13 flavonoid biosynthetic genes. The qRT-PCR results (Figure 9) were broadly in line with the RNA-Seq results. All 13 flavonoid biosynthesis genes were expressed in root and leaf tissues but with different expression patterns. *DICHI*, *DICCoAOMT*, and *DICYP* were expressed similarly in root and leaf tissues. Eight genes (*DIC4H*, *DIHCT*, *DIDFR*, *DLANS*, *DLANR*, *DICH5*, *DIF3' H*, and *DIF3H*) had relatively higher expression in roots (vs. leaves), and two genes (*DILAR* and *DIFLS*) had much higher expression levels in leaves (vs. roots).





**Figure 9.** Expression of flavonoid biosynthesis genes in root and leaf tissues of *D. longan*. Each value is the mean of 3 replicates, and error bars indicate standard deviations (SDs)

*Analysis of flavonoid contents of D. longan root and leaf tissues*

The same plant materials used for qRT-PCR were also used to analyse the contents of fifteen flavonoids in *D. longan* roots and leaves by LC-MS. Eight flavonoids were not detected in roots or leaves (genistein, isoliquiritigenin, isorhamnetin, formononetin, (S)-pinocembrin, chrysin, galangin, and daidzein). Among the

seven flavonoids detected, vitexin was detected only in roots. The remaining six flavonoids (catechin, epicatechin, rutin, morin, luteolin, and apigenin) were detected in both roots and leaves, but at different levels. Of these, epicatechin was the predominant ingredient in both root and leaf tissues of *D. longan*, and epicatechin content differed by approximately an order of magnitude between roots (213,773.65 ng/g) and leaves (22,388.71 ng/g). By comparison, apigenin content in roots (5.31 ng/g) and leaves (6.06 ng/g) was very low and did not differ significantly between the tissues. The flavonoid composition and contents in roots and leaves of *D. longan* were shown in Table 6. In general, roughly half of the flavonoids we tested were not detected in either root or leaf tissue. Among the remaining flavonoids, the most abundant one was epicatechin in the roots. It was present at much higher levels than any other flavonoid and deserves further investigation for future development and utilization.

**Table 6.** Flavonoid composition and contents in roots and leaves of *D. longan*

Flavonoids	Roots (ng/g)	Leaves (ng/g)
Catechin	993.80 ± 14.19	251.42 ± 12.70
Epicatechin	213773.65 ± 21813.84	22388.71 ± 2242.88
Vitexin	35.82 ± 14.87	NF
Rutin	239.97 ± 31.79	890.37 ± 236.04
Morin	75.68 ± 7.50	594.11 ± 338.63
Luteolin	760.66 ± 39.19	424.47 ± 127.91
Apigenin	5.31 ± 0.50	6.06 ± 0.66
Genistein	NF	NF
Isoliquiritigenin	NF	NF
Isorhamnetin	NF	NF
Formononetin	NF	NF
(S)-Pinoembrin	NF	NF
Chrysin	NF	NF
Galangin	NF	NF
Daidzein	NF	NF

Note: NF indicates not found

## Discussion

Flavonoids are one of the most important groups of phenolic compounds and the most common secondary metabolites in plants. Up to the present, more than 10,000 flavonoids have been identified in plant, and they are divided into six subclasses including flavones, flavanones, flavonols, isoflavones, proanthocyanidins, and anthocyanins (Dixon and Pasinetti, 2010; Yao *et al.*, 2018). Flavonoids have numerous pharmacological activities. They can not only be beneficial to the plant itself but also can be used to treat human diseases such as hypertension, dysentery, liver and urinary problems, diarrhoea, and cerebral disorders (Azuma *et al.*, 2012; Jennings *et al.*, 2012; Chang *et al.*, 2013).

*D. longan* is one of the most famous fruits in China, and China accounts for more than 50% of its global production. The secondary metabolites, especially flavonoids, found in *D. longan* fruits, leaves, flowers, and seeds are thought to have medicinal properties, and *D. longan* has long been used in traditional Chinese medicine (Thitiratsakul and Anprung, 2014). However, the leaves and roots of *D. longan*, while rich in resources, nonetheless accumulate relatively small amounts of flavonoids, limiting the scope of their application and their economic value. We therefore hope to enhance flavonoid accumulation in *D. longan* roots and leaves by means of biotechnology.

Transcriptome sequencing is a powerful biological technology for discovering molecular markers and identifying novel genes. Compared with the SOLiD, Roche 454, and HeliScope platforms, the Illumina HiSeq platform used in this study has been widely adopted because of its many advantages, in particular its capacity to analyse the expression levels of transcripts and to identify previously undiscovered transcripts (Ronaghi *et al.*, 1998; Harris *et al.*, 2008; Smith *et al.*, 2008). In this study, we used RNA-Seq to profile *D. longan* transcript levels. A total of 15.04 GB of clean data were obtained and assembled into 56,055 unigenes, 25,161 (44.9%) of which were successfully annotated with the GO, COG, KEGG, Swiss-Prot, or NCBI Nr databases, suggesting that they may have relatively conserved functions. The remaining unannotated unigenes (55.1%) could not be matched to known sequences in public databases, likely because of the presence of poorly conserved regions, such as untranslated regions (UTRs) (Hou *et al.*, 2011). These unannotated unigenes may play specific roles in *D. longan* which have not yet been identified and therefore deserve further investigation. The mean *D. longan* unigene length (829.61 nt) was greater than that of *Litchi chinensis* Sonn (687 nt for phase 1 and 669 nt for phase 2) (Zhang *et al.*, 2014). N50 length is commonly used to evaluate the quality of a transcriptome assembly, with a high number indicating a higher quality assembly (Li and Dewey, 2011). The N50 of the *D. longan* transcriptome was 1,583 nt, significantly longer than that of *Litchi chinensis* Sonn (811 nt), suggesting that the assembly quality was adequate for subsequent transcriptome analysis (Li *et al.*, 2013). We also identified DEGs when comparing the root and leaf transcriptomes of *D. longan*. In total, 6,684 genes differed in expression between roots and leaves: 3,550 genes were much highly expressed in roots, and 3,134 genes were much highly expressed in leaves. Expression patterns analysed by RNA-Seq and qRT-PCR were broadly consistent, suggesting that the RNA-Seq data will be very useful for subsequent studies of *D. longan*.

Thirteen putative flavonoid biosynthesis genes were differentially expressed between *D. longan* leaf and root tissues: *DIC4H*, *DIHCT*, *DIDFR*, *DLANS*, *DLANR*, *DICH3*, *DIF3' H*, and *DIF3H* were much highly expressed in roots, and *DILAR* and *DIFLS* were much highly expressed in leaves. Based on known flavonoid biosynthetic pathways (Vogt, 2010; Tohge *et al.*, 2017), all of the thirteen identified genes were predicted to function in flavonoid biosynthesis. To ascertain the relationship between gene expression and flavonoid accumulation, we measured the contents of fifteen flavonoids by LC-MS. In total, seven flavonoids were detected, including catechin, epicatechin, rutin, morin, luteolin, apigenin, and vitexin. Vitexin was only detected in roots. The other eight flavonoids (genistein, isoliquiritigenin, isorhamnetin, formononetin, (S)-pinocembrin, chrysin, galangin, and daidzein) were detected in neither roots nor leaves. The epicatechin contents were 213,773.65 ng/g in roots and 22,388.71 ng/g in leaves, both significantly higher than the contents of the other six flavonoids. Thus, epicatechin was clearly the predominant bioactive flavonoid in root and leaf tissues of *D. longan* and deserves further investigation into means of increasing its accumulation. Epicatechin has been demonstrated to have multiple functions including anticancer, antilipidemic, antineoplastic, cardio protective, anti-inflammatory, and antioxidant properties (Shanmugam *et al.*, 2017). In addition, the natural epicatechin was verified to regulate insulin secretion from pancreatic  $\beta$  cells (Yang and Chan, 2018). ANR is thought to be the gene involved in catalysing plant epicatechin production (Han *et al.*, 2012; Tan *et al.*, 2018; Yuan *et al.*, 2020). The expression level of *DLANR* was higher in roots than that in leaves, consistent with the high accumulation of epicatechin in roots. Therefore, we speculate that protein encoded by *DLANR* catalyses the biosynthesis of epicatechin in *D. longan*. We could increase epicatechin content of *D. longan* roots and leaves by means of genetic engineering in future. As a result, the medicinal value of roots and leaves, by-products of *D. longan* fruit cultivation, will be enhanced, turning waste into treasure.



## Conclusions

In this study, RNA-Seq technology was used to profile the transcriptome of *D. longan* on the Illumina HiSeq™ 2000 platform. In total, 56,055 unigenes were obtained, with an N50 of 1,583 nt and a mean length of 829.61 nt. Unigenes were annotated using NCBI Nr, Gene Ontology (GO), Clusters of Orthologous Group (COG), and Kyoto Encyclopaedia of Genes and Genomes (KEGG) databases. A total of 6,684 genes were differentially expressed between root and leaf tissues, of which thirteen were identified as flavonoid biosynthesis genes. Of these, eight genes were much highly expressed in roots (*DIC4H*, *DIHCT*, *DIDFR*, *DLANS*, *DLANR*, *DICH5*, *DIF3' H*, and *DIF3H*), and two were much highly expressed in leaves (*DILAR* and *DIFLS*). The contents of thirteen flavonoids in *D. longan* roots and leaves were measured by LC-MS, and epicatechin was found to be the predominant flavonoid in both tissues, which was significantly higher than the other flavonoids measured in the study. Its contents were 213,773.65 ng/g in roots and 22,388.71 ng/g in leaves. The results of this study will help to increase the economic value and expand the applications of *D. longan* roots and leaves by means of genetic engineering.

## Authors' Contributions

ZW conceived and designed the study, as well as wrote the manuscript. ZW, ZZW, and DXM performed the experiments. YXF reviewed and edited the manuscript. DXM and ZQY analyzed the data. All authors read and approved the final manuscript.

## Acknowledgements

This research was funded by the Doctoral Science Foundation (14LG06) and Youth Project (2019CX39) at Harbin University of Commerce.

## Conflict of Interests

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

## References

- Azuma A, Yakushiji H, Koshita Y, Kobayashi S (2012). Flavonoid biosynthesis-related genes in grape skin are differentially regulated by temperature and light conditions. *Planta* 236(4):1067-1080.
- Chang JJ, Hsu MJ, Huang HP, Chung DJ, Chang YC, Wang CJ (2013). Mulberry anthocyanins inhibit oleic acid induced lipid accumulation by reduction of lipogenesis and promotion of hepatic lipid clearance. *Journal of Agricultural and Food Chemistry* 61(25):6069-6076.
- Conesa A, Gotz S (2008). Blast2GO: A comprehensive suite for functional analysis in plant genomics. *International Journal of Plant Genomics* 2008:619832-619843.
- Conesa A, Gotz S, Garcia-Gomez JM, Terol J, Talon M, Robles M (2005). Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* 21(18):3674-3676.
- Dastmalchi M, Dhaubhadel S (2015). Proteomic insights into synthesis of isoflavonoids in soybean seeds. *Proteomics* 15(10):1646-1657.
- Deng Y, Li C, Li H, Lu S (2018). Identification and characterization of flavonoid biosynthetic enzyme genes in *Salvia miltiorrhiza* (Lamiaceae). *Molecules* 23(6):1467-1486.

- Dixon RA, Pasinetti GM (2010). Flavonoids and isoflavonoids: from plant biology to agriculture and neuroscience. *Plant Physiology* 154:453-457. <https://doi.org/10.1104/pp.110.161430>
- Fang J (2014). Bioavailability of anthocyanins. *Drug Metabolism Reviews* 46(4):508-520. <https://doi.org/10.3109/03602532.2014.978080>
- Ferrer JL, Austin MB, Stewart CJ, Noel JP (2008). Structure and function of enzymes involved in the biosynthesis of phenylpropanoids. *Plant Physiology and Biochemistry* 46(3):356-370.
- Galperin MY, Kristensen DM, Makarova KS, Wolf YI, Koonin EV (2019). Microbial genome analysis: the COG approach. *Briefings in Bioinformatics* 20(4):1063-1070. <https://doi.org/10.1093/bib/bbx117>
- Glenn TC (2011). Field guide to next-generation DNA sequencers. *Molecular Ecology Resources* 11(5):759-769.
- Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, ... Regev A (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnology* 29(7):644-652.
- Han Y, Vimolmangkang S, Soria-Guerra RE, Korban SS (2012). Introduction of apple *ANR* genes into tobacco inhibits expression of both *CHI* and *DFR* genes in flowers, leading to loss of anthocyanin. *Journal of Experimental Botany* 63(7):2437-2447.
- Harris TD, Buzby PR, Babcock H, Beer E, Bowers J, Braslavsky I, ... Xie Z (2008). Single-molecule DNA sequencing of a viral genome. *Science* 320(5872):106-109.
- Hernández Y, Bernstein R, Pagan P, Vargas L, McCaig W, Ramrattan G, ... Qiu WG (2018). BpWrapper: BioPerl-based sequence and tree utilities for rapid prototyping of bioinformatics pipelines. *BMC Bioinformatics* 19(1):76. <https://doi.org/10.1186/s12859-018-2074-9>
- Hillier LW, Reinke V, Green P, Hirst M, Marra MA, Waterston RH (2009). Massively parallel sequencing of the polyadenylated transcriptome of *C. elegans*. *Genome Research* 19(4):657-666.
- Hou R, Bao Z, Wang S, Su H, Li Y, Du H, ... Hu X (2011). Transcriptome sequencing and *de novo* analysis for Yesso scallop (*Patinopecten yessoensis*) using 454 GS FLX. *PLoS One* 6(6):e21560.
- Jennings A, Welch AA, Fairweather-Tait SJ, Kay C, Minihane AM, Chowienczyk P, ... Cassidy A (2012). Higher anthocyanin intake is associated with lower arterial stiffness and central blood pressure in women. *American Journal of Clinical Nutrition* 96(4):781-788.
- Jordon-Thaden IE, Chanderbali AS, Gitzendanner MA, Soltis DE (2015). Modified CTAB and TRIzol protocols improve RNA extraction from chemically complex Embryophyta. *Applications in Plant Sciences* 3(5):1400105. <https://doi.org/10.3732/apps.1400105>
- Lei Z, Zhou C, Ji X, Wei G, Huang Y, Yu W, ... Qiu Y (2018). Transcriptome analysis reveals genes involved in flavonoid biosynthesis and accumulation in *Dendrobium catenatum* from different locations. *Scientific Reports* 8(1):6373-6388.
- Li B, Dewey CN (2011). RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* 12:323. <https://doi.org/10.1186/1471-2105-12-323>
- Li C, Wang Y, Huang X, Li J, Wang H (2013). *De novo* assembly and characterization of fruit transcriptome in *Litchi chinensis* Sonn and analysis of differentially regulated genes in fruit in response to shading. *BMC Genomics* 14(552):552-567.
- Lin Y, Li N, Lin H, Lin M, Chen Y, Wang H, ... Lin Y (2020). Effects of chitosan treatment on the storability and quality properties of longan fruit during storage. *Food Chemistry* 306:125627-125636.
- Li R, Zhu H, Ruan J, Qian W, Fang X, Shi Z, ... Wang J (2010). *De novo* assembly of human genomes with massively parallel short read sequencing. *Genome Research* 20(2):265-272.
- Liu YY, Chen XR, Wang JP, Cui WQ, Xing XX, Chen XY, ... Li YH (2019). Transcriptomic analysis reveals flavonoid biosynthesis of *Syringa oblata* Lindl. in response to different light intensity. *BMC Plant Biology* 19(1):487-502.
- Meyer E, Aglyamova GV, Wang S, Buchanan-Carter J, Abrego D, Colbourne JK, ... Matz MV (2009). Sequencing and *de novo* analysis of a coral larval transcriptome using 454 GSFLX. *BMC Genomics* 10(219):219-235.
- Moriya Y, Itoh M, Okuda S, Yoshizawa AC, Kanehisa M (2007). KAAS: an automatic genome annotation and pathway reconstruction server. *Nucleic Acids Research* 35:W182-W185.
- Myhre S, Tveit H, Mollestad T, Laegreid A (2006). Additional gene ontology structure for improved biological reasoning. *Bioinformatics* 22(16):2020-2027.

- Ni J, Dong L, Jiang Z, Yang X, Chen Z, Wu Y, Xu M (2018). Comprehensive transcriptome analysis and flavonoid profiling of Ginkgo leaves reveals flavonoid content alterations in day-night cycles. *PLoS One* 13(3):e0193897.
- Park SJ, Park DH, Kim DH, Lee S, Yoon BH, Jung WY, ... Ryu JH (2010). The memory-enhancing effects of *Euphoria longan* fruit extract in mice. *Journal of Ethnopharmacology* 128(1):160-165.
- Petrussa E, Braidot E, Zancani M, Peresson C, Bertolini A, Patui S, Vianello A (2013). Plant flavonoids-biosynthesis, transport and involvement in stress responses. *International Journal of Molecular Sciences* 14(7):14950-14973.
- Ronaghi M, Uhlen M, Nyren P (1998). A sequencing method based on real-time pyrophosphate. *Science* 281(5375):363-365.
- Saito K, Yonekura-Sakakibara K, Nakabayashi R, Higashi Y, Yamazaki M, Tohge T, Fernie AR (2013). The flavonoid biosynthetic pathway in *Arabidopsis*: structural and genetic diversity. *Plant Physiology and Biochemistry* 72:21-34.
- Schmittgen TD, Livak KJ (2008). Analyzing real-time PCR data by the comparative C(T) method. *Nature Protocols* 3(6):1101-1108.
- Shanmugam B, Shanmugam KR, Ravi S, Subbaiah GV, Ramakrishana C, Mallikarjuna K, Reddy KS (2017). Exploratory studies of (-)-epicatechin, a bioactive compound of *Phyllanthus niruri*, on the antioxidant enzymes and oxidative stress markers in D-galactosamine-induced hepatitis in rats: a study with reference to clinical prospective. *Pharmacognosy Magazine* 13(1):S56-S62. <https://doi.org/10.4103/0973-1296.203973>
- Smith DR, Quinlan AR, Peckham HE, Makowsky K, Tao W, Woolf B, ... Richardson PM (2008). Rapid whole-genome mutational profiling using next-generation sequencing technologies. *Genome Research* 18(10):1638-1642.
- Tang NP, Zhou B, Wang B, Yu RB, Ma J (2009). Flavonoids intake and risk of lung cancer: a meta-analysis. *Japanese Journal of Clinical Oncology* 39(6):352-359.
- Tang YY, He XM, Sun J, Li CB, Li L, Sheng JF, ... Ling DN (2019). Polyphenols and alkaloids in byproducts of longan fruits (*Dimocarpus Longan* Lour.) and their bioactivities. *Molecules* 24(6):1186-1201.
- Tan L, Wang M, Kang Y, Azeem F, Zhou Z, Tuo D, ...Pan Z (2018). Biochemical and Functional characterization of anthocyanidin reductase (ANR) from *Mangifera indica* L. *Molecules* 23(11):2876-2895. doi: [10.3390/molecules23112876](https://doi.org/10.3390/molecules23112876)
- Thitiratsakul B, Anprung P (2014). Prebiotic activity score and bioactive compounds in longan (*Dimocarpus longan* Lour.): influence of pectinase in enzyme-assisted extraction. *Journal of Food Science and Technology* 51(9):1947-1955.
- Tohge T, de Souza LP, Fernie AR (2017). Current understanding of the pathways of flavonoid biosynthesis in model and crop plants. *Journal of Experimental Botany* 68(15):4013-4028.
- Tseng HC, Wu WT, Huang HS, Wu MC (2014). Antimicrobial activities of various fractions of longan (*Dimocarpus longan* Lour. Fen Ke) seed extract. *International Journal of Food Sciences and Nutrition* 65(5):589-593.
- Vogt T (2010). Phenylpropanoid biosynthesis. *Molecular Plant* 3(1):2-20.
- Wang Z, Fang B, Chen J, Zhang X, Luo Z, Huang L, ... Li Y (2010). De novo assembly and characterization of root transcriptome using Illumina paired-end sequencing and development of cSSR markers in sweet potato (*Ipomoea batatas*). *BMC Genomics* 11:726-739.
- Wei W, Qi X, Wang L, Zhang Y, Hua W, Li D ... Zhang X (2011). Characterization of the sesame (*Sesamum indicum* L.) global transcriptome using Illumina paired-end sequencing and development of EST-SSR markers. *BMC Genomics* 12:451-463.
- Wilson RT, Wang J, Chinchilli V, Richie JP, Virtamo J, Moore LE, Albanes D (2009). Fish, vitamin D, and flavonoids in relation to renal cell cancer among smokers. *American Journal of Epidemiology* 170(6):717-729.
- Xie F, Burklew CE, Yang Y, Liu M, Xiao P, Zhang B, Qiu D (2012). De novo sequencing and a comprehensive analysis of purple sweet potato (*Ipomoea batatas* L.) transcriptome. *Planta* 236(1):101-113.
- Yang C, He N, Ling X, Ye M, Zhang C, Shao W, ... Li Q (2008). The isolation and characterization of polysaccharides from longan pulp. *Separation and Purification Technology* 63(1):226-230. <https://doi.org/10.1016/j.seppur.2008.05.004>
- Yang K, Chan CB (2018). Epicatechin potentiation of glucose-stimulated insulin secretion in INS-1 cells is not dependent on its antioxidant activity. *Acta Pharmacologica Sinica* 39:893-902.

- Yao Q, Lin MT, Zhu YD, Xu HL, Zhao YZ (2018). Recent trends in potential therapeutic applications of the dietary flavonoid didymin. *Molecules* 23(10):2547. <https://doi.org/10.3390/molecules23102547>
- Ye J, Fang L, Zheng H, Zhang Y, Chen J, Zhang Z, ... Wang J (2006). WEGO: a web tool for plotting GO annotations. *Nucleic Acids Research* 34:W293-W297.
- Yuan Y, Zhang J, Liu X, Meng M, Wang J, Lin J (2020). Tissue-specific transcriptome for *Dendrobium officinale* reveals genes involved in flavonoid biosynthesis. *Genomics* 112(2):1781-1794.
- Zhang HN, Wei YZ, Shen JY, Lai B, Huang XM, Ding F, ... Chen HB (2014). Transcriptomic analysis of floral initiation in litchi (*Litchi chinensis* Sonn.) based on de novo RNA sequencing. *Plant Cell Reports* 33(10):1723-1735.



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



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