



Comparative Transcriptomic Analysis on White and Blue Flowers of *Platycodon grandiflorus* to Elucidate Genes Involved in the Biosynthesis of Anthocyanins

Yanping Lv^{1†}, Xiaohui Tong^{2†}, Pengfei Zhang¹, Nianjun Yu¹, Shuangying Gui¹, Rongchun Han^{1*}, Dezhu Ge^{3*}

¹School of Pharmacy, Anhui University of Chinese Medicine; Hefei 230012, China

²School of Life Sciences, Anhui University of Chinese Medicine; Hefei 230012, China

³Department of Research and Development, Anhui Jiren Pharmaceutical Company; Bozhou 236800, China

*Corresponding author: Dezhu Ge, Department of Research and Development, Anhui Jiren Pharmaceutical Company, No. 1799, Yaodu Rd, Qiaocheng District, Bozhou, Anhui, 236800, China. Tel: +86-5582808341, Fax: +86-5585587510, E-mail: gedezhu@jirenjitian.com

*Corresponding author: Rongchun Han, School of Pharmacy, Anhui University of Chinese Medicine, No. 350, Longzihu Rd, Yaohai District, Hefei, Anhui, 230012, China. Tel: +86-55168129171, Fax: +86-55165169222, E-mail: hanr@ahtcm.edu.cn

† These authors contributed equally to this study.

Background: *Platycodon grandiflorus* has long been used in Northeast Asia as a food and folk medicine to treat various diseases. The intense blue color of *P. grandiflorus* corolla is its characteristic feature.

Objectives: By comparing deep transcriptomic data of *P. grandiflorus* and its white cultivar, we intended to elucidate the molecular mechanisms concerning the biosynthesis of anthocyanins in this plant.

Material and Methods: We sampled blue mature flowers (PgB) and yellow young buds (PgY) of *P. grandiflorus*. Meanwhile, mature flowers (PgW) of *P. grandiflorus* white cultivar were also collected for RNA extraction and next-generation sequencing. After high-throughput sequencing, Trinity software was applied for de novo assembly and the resultant 49934 unigenes were subjected for expression analysis and annotation against NR, KEGG, UniProt, and Pfam databases.

Results: In all, 32.77 Gb raw data were generated and the gene expression profile for the flowers of *P. grandiflorus* was constructed. Pathway enrichment analysis demonstrated that genes involved in flavone and flavonol biosynthesis were differently expressed.

Conclusions: The extremely low expression of flavonoid-3',5'-hydroxylase in PgY and PgW was regarded as the reason for the formation of its white cultivar. Our findings provided useful information for further studies into the biosynthetic mechanism of anthocyanins.

Keywords: Anthocyanin; Biosynthetic pathway; Flavonoid-3',5'-hydroxylase; Next-generation sequencing; *Platycodon grandiflorus*

1. Background

The root of *Platycodon grandiflorus* (Jacq.) A. DC. has been used as a traditional herbal medicine mainly in Northeast Asia for centuries (1). Meanwhile, in the area of Japan, Korea, Mongolia, and China, *P. grandiflorus* was also a popular functional food in addition to its medicinal efficacy to treat sore throat, cough, excessive phlegm, diabetes, inflammatory

diseases, and so forth (2-5). Its pharmacological material basis is the chemical constituents including triterpenoid saponins, flavonoids, phenolic acids, polyacetylenes, sterols, and polysaccharides (6-10). Notably, among the nine isolated flavonoids, six of them were from *P. grandiflorus* flowers or seeds (11). Flavonoids, along with betalains and carotenoids, are essential floral pigments and for plants, flavonoids

play the most significant role as colored pigments. Belonging to polyphenolic compounds, flavonoids can be divided into several subgroups including anthocyanidins, flavanones, flavones, isoflavones, flavonols, and flavan-3-ols (12, 13). The first flavonoid identified from *P. grandiflorus* flower was platyconin, a diacylated anthocyanin that comprises the anthocyanidin called delphinidin and sugar moieties (14). From the six typical anthocyanidins including delphinidin, malvidin, petunidin, cyanidin, peonidin and pelargonidin, hundreds of anthocyanins as their derivatives can be biosynthesized in the plant kingdom (15). On the other hand, cyanidin and peonidin share the same core structure, and malvidin as well as petunidin use delphinidin as the substrate for their biosynthesis. The main groups of anthocyanidins can then be reduced to three: delphinidin, cyanidin, and pelargonidin (16). Typically, delphinidin contributes to the blue color of plant flowers, while cyanidin often produces red color and pelargonidin orange color. Previous studies demonstrated that delphinidin was the key compound for the blue color of *P. grandiflorus* flowers (11, 14). To understand further the biosynthetic pathway of flavonoid and anthocyanin by comparing RNA-Seq data from *P. grandiflorus* and its white cultivar will provide useful information for commercial applications to develop unique flower colors through the creation of transgenic varieties that produce desired anthocyanins resulted from delphinidin.

2. Objectives

The flower of *P. grandiflorus* is blue. For its white cultivar, the color of the corolla in the bud stage is yellow and then eventually turns white when mature (**Fig.1**). The only morphological difference between the two plants is the color of their flowers. Therefore, comparing deep transcriptomic data would shed light

on biosynthetic mechanism of anthocyanins that give the flower characteristic blue color. Blue genes were reported previously from a handful of plants including butterfly pea, gentian, and petunias. Our study will enrich genetic information for a better understanding of the mechanisms of the various blue genes.

3. Material and Methods

3.1. Sample Collection and Total RNA Preparation

The blue flowers of *P. grandiflorus* (PgB), the yellow buds (PgY), and white flowers (PgW) from its white cultivar were collected from the botanical garden of Anhui University of Chinese Medicine in June, 2019 and immediately submerged into liquid nitrogen. The materials were then grinded into fine powders without being thawed. Total RNA extraction and quality evaluation were performed according to the method reported previously (17) with slight modifications as following: 750 μ L extraction buffer were used, along with other designated reagents for the initial step of RNA isolation.

3.2. Sequencing and Data Interpretation

We applied BGISEQ-500 platform to conduct next-generation sequencing. At first, to produce clean reads, all the raw reads in fastq format were subjected to Trimmomatic software to eliminate low-quality reads, adapter contamination, and reads containing excessive ambiguous bases (18). Then Trinity was applied to fulfill *de novo* assembly using clean reads as its inputs (19). The resultant *P. grandiflorus* unigenes were then annotated against four functional databases (NR, SwissProt, KEGG, and Pfam) using BLAST program. By using Bowtie2 and RSEM, clean reads from PgB, PgY and PgW were aligned to unigenes to calculate Fragments Per Kilobase per Million reads (FPKM), respectively (20).

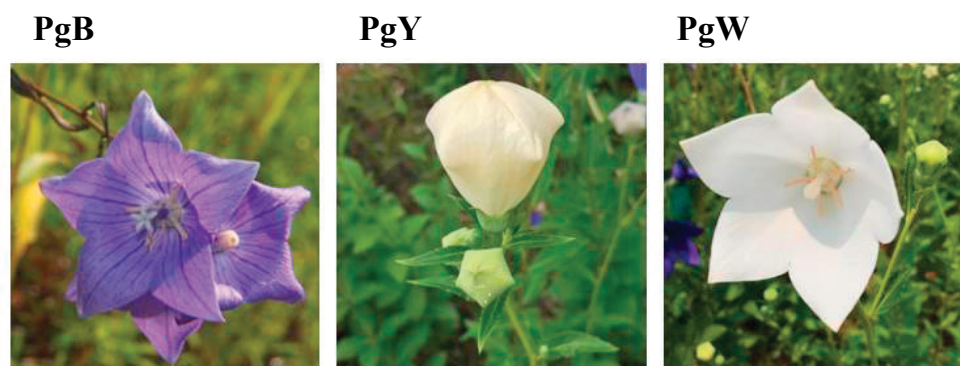


Figure 1. Flowers of *P. grandiflorus* and its white cultivar. PgB Blue flowers of *P. grandiflorus*, PgY Yellow buds of *P. grandiflorus* white cultivar, PgW White flowers of the white cultivar.

Based on KEGG annotation result, we conducted biosynthetic pathway enrichment analysis for *P. grandiflorus* differently expressed genes (DEG) with the aid of phyper function in R program (21). False discovery rate (FDR) was applied to adjust *P* values and the resultant *Q* values less than or equal to 0.05 were considered significant enrichment.

4. Results

4.1 Summary of Data Output

32.77 Gb data with high quality were acquired by BGISEQ platform. Total length, average length, N50, and GC contents of the resultant unigenes were summarized in **Table 1**. 35907 unigenes (71.91% of the total 49934 unigenes) from *P. grandiflorus* were assigned annotations against NR database, and for SwissProt, KEGG, and Pfam, the numbers are 26326 (52.72%), 27916 (55.91%) and 26867 (53.81%). Overall, 36312 (72.72%) unigenes received respective gene annotation, suggesting the high quality of sequenced reads as well as assembled outputs. Gene expression pattern of *P. grandiflorus* fit the expected normal distribution and Box-plot showed dispersion of FPKM values for all the three samples was similar, indicating reliable results for downstream analysis (**Fig. 2A**). Detailed unigene intersections were presented in **Figure 2B** where 19450 out of all 36312 unigenes were annotated unanimously by NR, SwissProt, KEGG, and Pfam. With the help of BLASTx program, up to 35907 unigenes obtained annotations in NCBI NR database, which accounted for approximately 98.9 percent of all annotated unigenes.

4.2 Function Annotation for *P. grandiflorus* Transcripts

To predict coding sequences (CDS) of the unigenes, TransDecoder (22) was utilized and finally, 29830 CDS were obtained. The length distribution of all detected CDS was demonstrated in **Supplementary Figure. S1A**. We also analyzed KEGG function

distribution pattern for all the unigenes and found that 15829 genes were involved in the category of metabolism. Notably, 951 genes were related to the biosynthesis of secondary metabolites, which was worthy of further exploration in terms of blue anthocyanins (**Fig.3**). Transcription factors (TFs) are essential for the biosynthesis of plant metabolites and genes from MYB transcription factor family were reported to regulate the flavonoid biosynthetic pathway and relate to flavonoid accumulation (23, 24). TFs from *P. grandiflorus* RNA-Seq data were screened and for MYB family 151 genes were retrieved (**Supplementary Fig. S1B**).

4.3. Pathway Enrichment, Gene Expression, and qPCR Validation

As far as DEG between PgB and PgW were concerned, ten out of the 132 studied pathways were significant and interestingly, flavone and flavonol biosynthesis was one of them with the *Q* value of 0.028 (**Supplementary Fig. S1C**). We also performed quantitative real-time PCR (qPCR) to validate gene expression profiles by selecting three genes related to flavonoid biosynthesis on a ROCHE Z480 instrument. Three technical replicates were prepared for PgW, PgB and PgY respectively. After total RNAs were extracted and transcribed, the resultant cDNA was used for the quantification of gene expression. We adopted the SYBR green reaction protocol. Reaction conditions included 5 min pre-incubation at 95 °C, followed by 45 cycles of 95 °C for 10 s, 60 °C for 10s and 72 °C for 10s. Finally, melt-curve analysis was conducted to confirm single PCR product amplification. Ubiquitin C was chosen as the internal control gene (25, 26) and we found qPCR output coincided with RNA-Seq data (**Supplementary Table S1 and S2**). Fold changes regarding FPKM and 2-DDCt displayed similar pattern. The *P. grandiflorus* sequences discussed or tested in this study were all listed in **Supplementary Table S3**.

Table 1. Overview of *P. grandiflorus* transcriptomic assembly

Items	Numbers
Unigene total length (bp)	58117764
No. of unigenes	49934
Average length of unigenes (bp)	1163
N50, N70, N90 (bp)	1731, 1217, 573
GC contents (%)	41.90

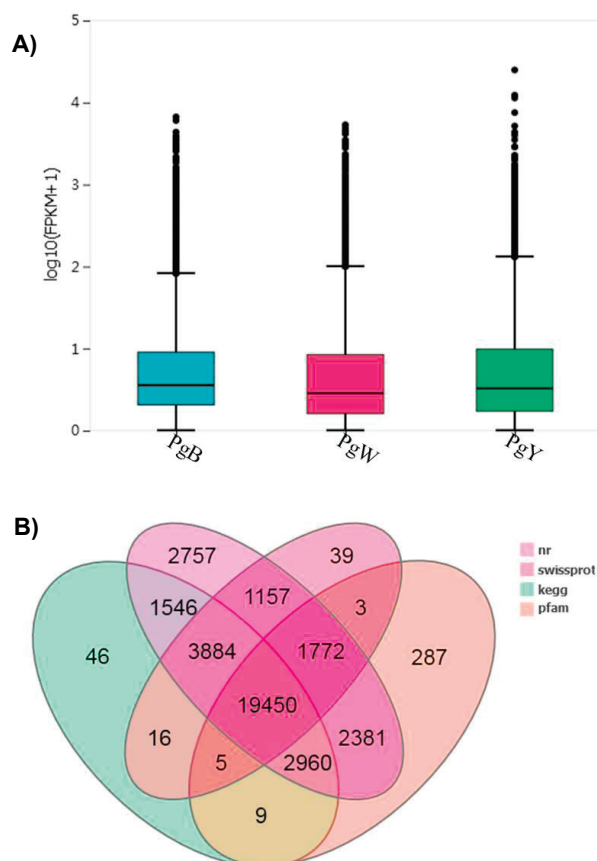


Figure 2. Profile of assembled unigenes and the overall expression distribution. **A)** Expression pattern of the unigenes from three samples. Median values for PgB, PgW and PgY were 0.55, 0.46 and 0.51, with upper quartile 0.96, 0.93 and 0.99. **B)** Annotations for all unigenes against public databases.

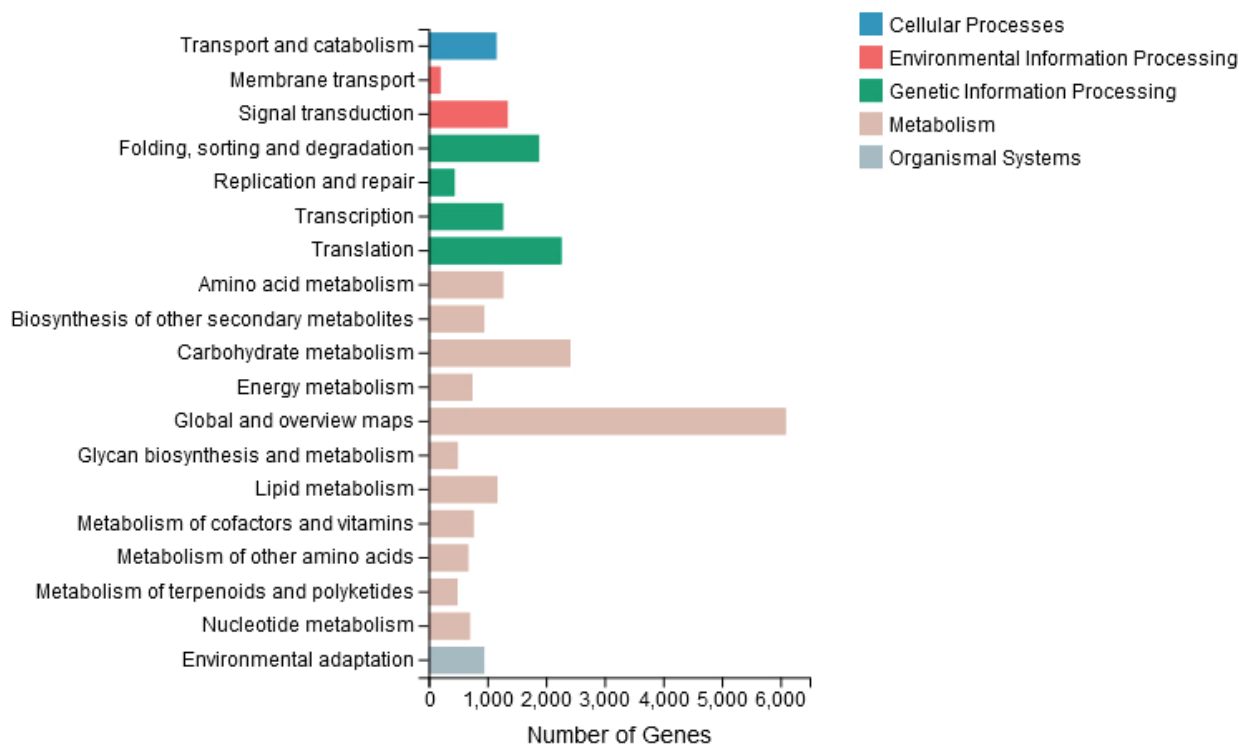


Figure 3. *P. grandiflorus* gene function distribution against KEGG database

5. Discussion

The structural differences among delphinidin, cyanidin, and pelargonidin are the position and number of hydroxyl groups bonded to the B ring of the flavonoid (**Fig. 4**). Flavonoids are derived from phenylalanine, an essential amino acid, and then chalcone is biosynthesized by chalcone synthase (CHS) to be the initial intermediate for all the flavonoids. Based on the information retrieved from *P. grandiflorus* RNA-Seq data, we proposed that the plant utilized the following strategy to produce delphinidin. From naringenin chalcone, naringenin was yielded by chalcone isomerase (CHI). Subsequently, naringenin 3-dioxygenase (F3H) catalyzed the reaction to produce dihydrokaempferol which was then turned into dihydromyricetin by flavonoid 3',5'-hydroxylase (F3'5'H). With the help of bifunctional dihydroflavonol 4-reductase/flavanone 4-reductase (DFR) and anthocyanidin synthase (ANS),

delphinidin finally came into being, which provided the very substrate for the blue pigment platyconin in its flower (**Fig. 4**).

P. grandiflorus and its white cultivar share similar genetic backgrounds. Therefore, comparing their transcriptomic data was worthwhile to pinpoint the key gene(s) responsible for the blue corolla. Over-expression of a petunia F3'5'H and a petunia DFR gene led to the commercialization of blue carnation which was impossible to produce by hybridization breeding (16). Recently, Japanese researchers succeeded in creating blue roses although the color of the rose was not pure blue. The principle in producing *P. grandiflorus* blue flowers and the related genes will facilitate us to create purer blue flowers. On the other hand, this study on the blue gene of *P. grandiflorus* will deepen our understanding concerning the mechanism of such genes by providing more genetic information to the scientific community.

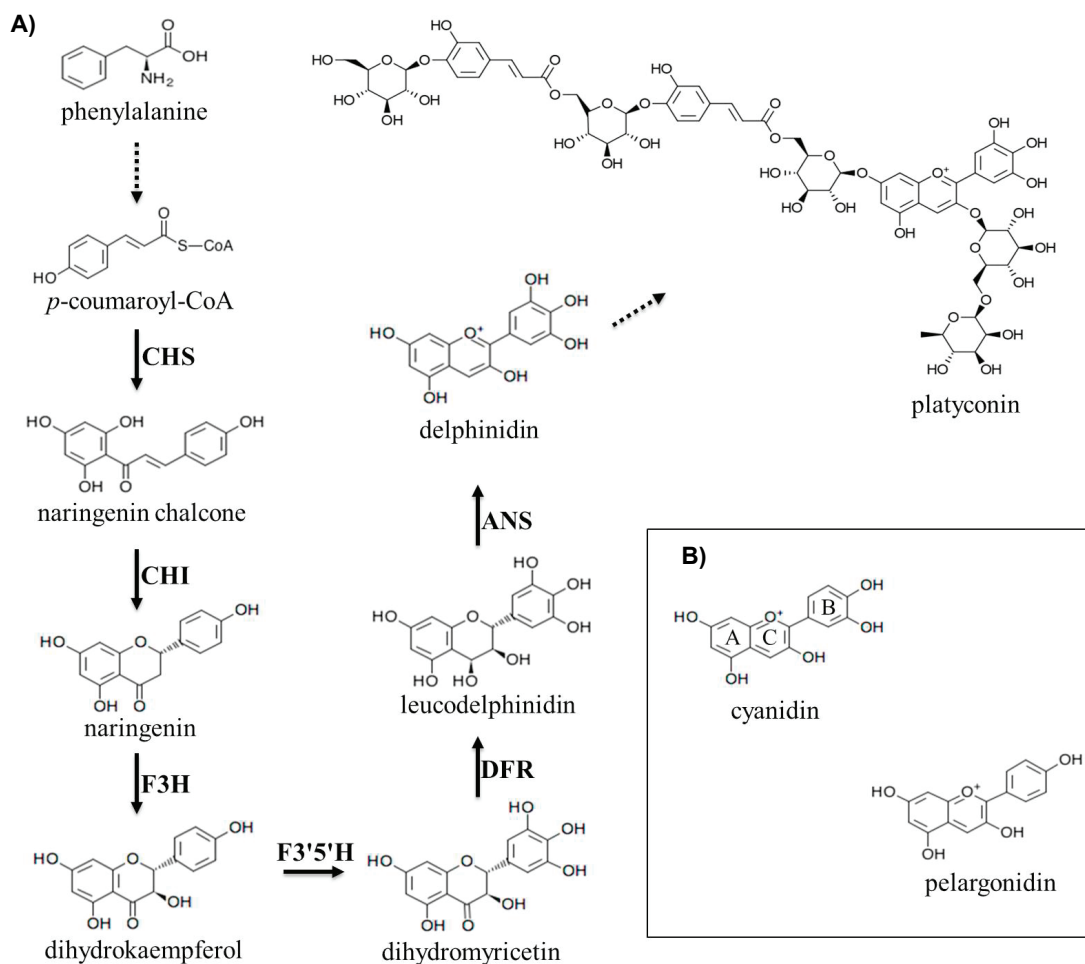


Figure 4. Proposed platyconin biosynthetic pathway in *P. grandiflorus*. **A)** Dotted arrows indicated multistep reactions. Enzymes responsible for the conversion of certain substrate were enumerated next to the respective arrow. **B)** Chemical structures of cyanidin and pelargonidin to show hydroxyl group(s) bonded to B ring of flavonoids, compared with delphinidin.

To analyze the possible reason why the white cultivar of *P. grandiflorus* changed its corolla color, we searched unigenes with annotation related to flavonoid biosynthesis with such inclusion criteria: the length of the unigene must be over 400 bp and FPKM value(s) should be above 1.0 in at least one sample. We gathered 19 putative genes and plotted them adopting the R program (Fig. 5). In general, the gene expression level of PgB was higher than that of PgW and PgY, suggesting PgB maintained a more robust function to produce and accumulate flavonoids. There were more than three paralogs genes for CHS, CHI, F3H, DFR, and ANS, but F3'5'H had only one. FPKM values for PgW and PgY were 0.26 and 1.20, compared to 208.85 in PgB.

6. Conclusion

Based on *P. grandiflorus* transcriptomic data as well

as qPCR validation findings, we hypothesized that the malfunction of F3'5'H gene was the underlying reason for the formation of white cultivars. Young buds of the white cultivar demonstrated yellowish coloration; this may be because of yellow flavonoid substrates accumulation, which could not be converted to delphinidin but then be catalyzed to form colorless compounds.

The color of anthocyanins depends not only on the structure themselves, acyl or glycosyl moieties, metal ions and vacuolar pH also play essential roles (16, 27). Moreover, transcription factors that are responsible for the regulation of flavonoid genes should also be thoroughly studied. Our findings regarding flowers of *P. grandiflorus* and its white cultivar contribute to the understanding of flavonoid biosynthesis and the data obtained will serve as a stepping stone for further studies.

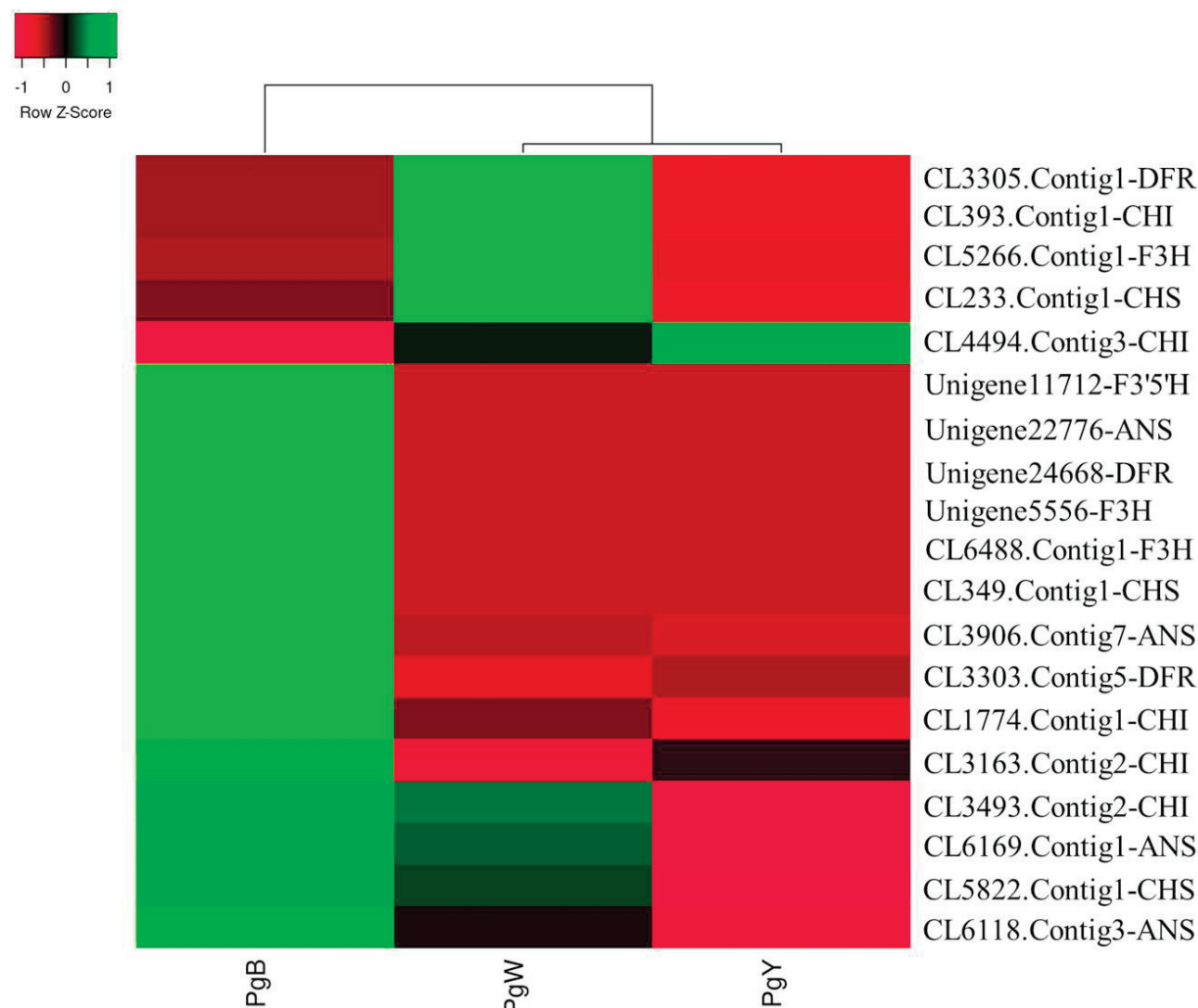


Figure 5. Heatmap for selected genes involved in *P. grandiflorus* flavonoid biosynthesis. FPKM values for respective genes were standardized, with Row Z-Score illustrated in the upper left panel.

Acknowledgments

This study was supported by Education Department of Anhui Province (Grant No. KJ2019A0463, KJ2020A0385), Department of Science and Technology of Anhui Province (Grant No. 18030801128), and Anhui University of Chinese Medicine (Grant No. 2019zrzd02).

References

- Nyakudya E, Jeong JH, Lee NK, Jeong YS. Platycosides from the roots of *Platycodon grandiflorum* and their health benefits. *Prev Nutr Food Sci.* 2014;**19**(2):59-68. doi:10.3746/pnf.2014.19.2.059.
- Chinese Pharmacopoeia Committee. The pharmacopoeia of the people's republic of China 2020 edition. China Medical Science Press, Beijing. 2020.
- Ji MY, Bo A, Yang M, Xu JF, Jiang LL, Zhou BC, *et al.* The pharmacological effects and health benefits of *Platycodon grandiflorus*-A medicine food homology species. *Foods.* 2020;**9**(2):142. doi:10.3390/foods9020142.
- Li W, Liu Y, Wang Z, Han Y, Tian YH, Zhang GS, *et al.* Platycodin D isolated from the aerial parts of *Platycodon grandiflorum* protects alcohol-induced liver injury in mice. *Food Funct.* 2015;**6**(5):1418-1427. doi:10.1039/c5fo00094g.
- Zhang W, Hou J, Yan X, Leng J, Li R, Zhang J, *et al.* *Platycodon grandiflorum* saponins ameliorate cisplatin-induced acute nephrotoxicity through the NF- κ B-mediated inflammation and PI3K/Akt/Apoptosis signaling pathways. *Nutrients.* 2018;**10**(9):1328. doi:10.3390/nu10091328.
- Ma G, Guo W, Zhao L, Zheng Q, Sun Z, Wei J, *et al.* Two new triterpenoid saponins from the root of *Platycodon grandiflorum*. *Chem Pharm Bull (Tokyo).* 2013;**61**(1):101-104. doi:10.1248/cpb.c12-00713.
- Nikaido T, Koike K, Mitsunaga K, Saeki T. Two new triterpenoid saponins from *Platycodon grandiflorum*. *Chem Pharm Bull (Tokyo).* 1999;**47**(6):903-904. doi:10.1248/cpb.47.903.
- Qiu L, Xiao Y, Liu YQ, Peng LX, Liao W, Fu Q. Platycosides P and Q, Two New Triterpene Saponins from *Platycodon grandiflorum*. *J Asian Nat Prod Res.* 2019;**21**(5):419-425. doi:10.1080/10286020.2018.1488835.
- Mazol I, Glensk M, Cisowski W. Polyphenolic compounds from *Platycodon grandiflorum* A. DC. *Acta Pol Pharm.* 2004;**61**(3):203-208.
- Kim M, Hwang IG, Kim SB, Choi AJ. Chemical characterization of Balloon Flower (*Platycodon grandiflorum*) sprout extracts and their regulation of inflammatory activity in lipopolysaccharide-stimulated RAW 264.7 murine macrophage cells. *Food Sci Nutr.* 2019;**8**(1):246-256. doi:10.1002/fsn3.1297.
- Zhang L, Wang Y, Yang D, Zhang C, Zhang N, Li M, *et al.* *Platycodon grandiflorus* - an ethnopharmacological, phytochemical and pharmacological review. *J ethnopharmacol.* 2015;**164**:147-161. doi: 10.1016/j.jep.2015.01.052.
- Saito K, Yonekura-Sakakibara K, Nakabayashi R, Higashi Y, Yamazaki M, Tohge T, *et al.* The flavonoid biosynthetic pathway in Arabidopsis: structural and genetic diversity. *Plant Physiol Biochem.* 2013;**72**:21-34. doi:10.1016/j.plaphy.2013.02.001.
- Panche AN, Diwan AD, Chandra SR. Flavonoids: an Overview. *J Nutr Sci.* 2016;**5**:e47. doi:10.1017/jns.2016.41.
- Goto H, Kondo T, Tamura H, Kawahori K, Hattori H. Structure of platyconin, a diacylated anthocyanin isolated from the Chinese Bell-flower *Platycodon grandiflorum*. *Tetrahedron letters.* 1983;**24**(21):2181-2184. doi:10.1016/S0040-4039(00)81877-8.
- Veitch NC, Grayer RJ. Flavonoids and their glycosides, including anthocyanins. *Nat Prod Rep.* 2011;**28**(10):1626-1695. doi:10.1039/c1np00044f.
- Tanaka Y, Brugliera F, Chandler S. Recent progress of flower colour modification by biotechnology. *Int J Mol Sci.* 2009;**10**(12):5350-5369. doi:10.3390/ijms10125350.
- Liu L, Han R, Yu N, Zhang W, Xing L, Xie D, *et al.* A method for extracting high-quality total RNA from plant rich in polysaccharides and polyphenols using *Dendrobium huoshanense*. *PLoS one.* 2018;**13**(5):e0196592. doi:10.1371/journal.pone.0196592.
- Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics.* 2014;**30**(15):2114-2120. doi:10.1093/bioinformatics/btu170.
- Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J, *et al.* De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat Protoc.* 2013;**8**(8):1494-1512. doi:10.1038/nprot.2013.084.
- Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat methods.* 2012;**9**(4):357-359. doi:10.1038/nmeth.1923.
- Kanehisa M, Sato Y, Furumichi M, Morishima K, Tanabe M. New approach for understanding genome variations in KEGG. *Nucleic Acids Res.* 2019;**47**(D1):D590-D595. doi:10.1093/nar/gky962.
- Kim HS, Lee BY, Won EJ, Han J, Hwang DS, Park HG, *et al.* Identification of xenobiotic biodegradation and metabolism-related genes in the copepod *Tigriopus japonicus* whole transcriptome analysis. *Mar Genomics.* 2015;**3**:207-208. doi:10.1016/j.margen.2015.05.011.
- Huang W, Sun W, Lv H, Luo M, Zeng S, Pattanaik S, *et al.* A R2R3-MYB transcription factor from *Epimedium sagittatum* regulates the flavonoid biosynthetic pathway. *PLoS One.* 2013;**8**(8):e70778. doi:10.1371/journal.pone.0070778.
- Zhang W, Xu F, Cheng S, Liao Y. Characterization and functional analysis of a MYB gene (GbMYBFL) related to flavonoid accumulation in *Ginkgo biloba*. *Genes Genomics.* 2018;**40**(1):49-61. doi:10.1007/s13258-017-0609-5.
- Sinha P, Saxena RK, Singh VK, Krishnamurthy L, Varshney RK. Selection and validation of housekeeping genes as reference for gene expression studies in Pigeonpea (*Cajanus cajan*) under heat and salt stress conditions. *Front Plant Sci.*

- 2015;**6**:1071. doi:10.3389/fpls.2015.01071.
26. 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. doi:10.3389/fpls.2015.00426.
27. Yoshida K, Kawachi M, Mori M, Maeshima M, Kondo M, Nishimura M, *et al.* The involvement of tonoplast proton pumps and Na⁺(K⁺)/H⁺ exchangers in the change of petal color during flower opening of Morning Glory, *Ipomoea tricolor* cv. Heavenly Blue. *Plant Cell Physiol.* 2005;**46**(3):407-415. doi:10.1093/pcp/pci057.