Pak. J. Bot., 56(1): 197-205, 2024. DOI: http://dx.doi.org/10.30848/PJB2024-1(26)

HIGH-THROUGHPUT TRANSCRIPTOME SEQUENCING AND ANALYSIS OF THE ENDANGERED ANTICANCER MEDICINAL PLANT *SINOPODOPHYLLUM HEXANDRUM* (ROYLE) T. S. YING

HAONA GAO, ZHENG ZHANG, LU LI, XIUFANG ZHAO AND WEI LIU*

College of Agriculture, Henan University of Science and Technology, Luoyang 471023, China
**Corresponding author's email: 15729111052@163.com*

Abstract

Sinopodophyllum hexandrum (Royle) T. S. Ying is a traditional medicinal plant in China. Podophyllotoxin, a chemical compound contained in its rhizomes, has important anticancer medicinal value for the treatment of cervical cancer, metrocarcinoma, leukemia, and rheumatism. To obtain the information characteristics of the transcriptome of *S. hexandrum*, the Illumina HiSeqTM2000 sequencing system was used as the library sequencing platform. The non-reference transcriptome sequencing analysis of the roots, stems and leaves of *S. hexandrum* was carried out by double-end sequencing method, and 74026 unigenes with high reliability of annotation were obtained. The COG and GO functional classification of unigenes in the transcriptome of *S. hexandrum* showed that the proportion of unigenes associated with metabolic process, catalytic reaction function, and binding function accounted for a larger proportion. The results of the KEGG analysis showed that *S. hexandrum* transcriptome unigenes were annotated for 125 metabolic pathways. The number of unigenes annotated to the metabolic pathway was the highest, up to 2921 (24.96%), and the pathway ID of this pathway was ko01100; followed by the biosynthetic pathway controlling secondary metabolites, with 1425 (12.18%) annotated genes, corresponding to the pathway ID of ko01110; in third place was the pathway controlling ribosome metabolism, with 889 (7.6%) annotated genes, corresponding to the pathway ID of ko03010. High-throughput transcriptome sequencing is the sequencing method selected in this study. Subsequently, the overall transcription of *S. hexandrum* was explored by bioinformatics methods such as COG, GO, and KEGG. These results provide theoretical support for the analysis of the *S. hexandrum* biosynthesis pathway, the mining of key regulatory genes of podophyllotoxin, and the innovative development and utilization of *S. hexandrum* resources.

Key words: *Sinopodophyllum hexandrum*; Podophyllotoxin; High-throughput transcriptome sequencing; Functional classification; Metabolic pathways.

Introduction

Sinopodophyllum hexandrum (Royle) T.S. Ying, a perennial herb monocot, belonging to the genus *Sinopodophyllum* in the family *Berberaceae* (Anon., 2011). The rhizome of *S. hexandrum* is a traditional folk medicinal plant in China and has been recorded in several Chinese herbal-related literature (Anon., 1975, 2002). *S. hexandrum* can be used as medicine for detoxification, pain relief, and blood circulation (Li *et al*., 2005; Zhao *et al*., 2011; Anon., 2020). Its rhizomes, leaves, and fruits contain a variety of chemical components, including lignans, flavonoids, and sterols (Zhao *et al*., 2023b). Podophyllotoxin lignans are important chemical components in *S. hexandrum*. They have strong anti-cancer biological activity and have good effects in the treatment of neuroblastoma, lung, testicular and other cancers (Damayanthi & Lown, 1998). However, due to the toxicity of podophyllotoxin, etoposide (VP-16-213) and teniposide (VM-26) are semi-synthetic derivatives of podophyllotoxin (Jackson & Dewick, 1984; Damayanthi & Lown, 1998; Moraes *et al*., 2000; Canel *et al*., 2001). At present, a variety of podophyllotoxin lignans have been isolated from *S. hexandrum*, such as podophyllotoxin, epipodophyllotoxin, 4'-demethylepipodophyllotoxin, deoxypodophyllotoxin, etc (Huang *et al*., 2012; Yan *et al*., 2020; Yang *et al*., 2022). As early as 1861, some scholars found that podophyllotoxin has anti-tumor activity. Subsequent researchers have verified this discovery. Experiments have proved that podophyllotoxin has the effect of inhibiting tumor growth. (Kelleher, 1978; Tomioka *et al*., 1989). In 1946, King & Sullivan (1946) found that the anti-tumor mechanism of podophyllotoxin is similar to that of colchicine. Both of them inhibit tumor growth by binding to tubulin and preventing the formation of microtubule bundles during mitosis.

A high concentration of podophyllotoxin was found in the roots of *S. hexandrum*, and the content of podophyllotoxin was the highest among some species of podophyllum (Fay & Ziegler, 1985; Stähelin & Albert, 1991; Giri & Lakshmi, 2000). Therefore, the demand for *S. hexandrum* in the field of health care is increasing. Due to commercial interests and medicinal value, *S. hexandrum* is continuously over-excavated, which has a devastating effect on its subsequent reproduction (asexual reproduction and sexual reproduction) and genetic development, greatly destroying the natural wild population of *S. hexandrum* (Yang & Lu, 2022). *S. hexandrum* grows at a high altitude of 3500-5000 m. In order to cope with the harsh plateau environment, its seeds have thick film, dense seed coat, and dormancy characteristics. Such reproductive characteristics led to a low seed germination rate of *S. hexandrum*, which had a direct impact on the natural regeneration of *S. hexandrum* resources (Li *et al*., 2008; Anon., 2011). In addition, with global warming in recent years, the scope of human activities has expanded, and the ecological environment of the plateau has also changed. The suitable habitat of *S. hexandrum* has been reduced year by year, and it has migrated to higher altitudes and higher latitudes, further increasing the difficulty of *S. hexandrum* survival (Guo *et al*., 2014; Lai *et al*., 2022). At present, the number of populations of *S. hexandrum* is decreasing day by day. It is recorded in the《China Biodiversity Red List》, 《China Plant Red Book》, and《The IUCN Red List》.

It is a nationally rare and protected plant of Class II (Anon., 1987; Fu, 1991). In the face of this situation, some scholars have bred *S. hexandrum* by means of root propagation and introduction and cultivation, but *S. hexandrum* has mainly harvested rhizomes and the efficacy has a great relationship with plant age, so it has not been large-scale breeding (Linghu *et al*., 2011). In addition, in terms of tissue culture, although the in vitro culture technology of plants has become increasingly mature, the callus cultivated by *S. hexandrum* is easy to become browning, has no subculture ability, and the rooting ability of aseptic seedlings is weak. It is still difficult to carry out industrial production through tissue culture (Chattopadhyay *et al*., 2001). The resources are endangered, and a variety of factors have led to a downward trend in the population of *S. hexandrum*. However, the demand for podophyllotoxin in the market is still increasing. Researchers need to try more ways to obtain podophyllotoxin with high efficiency, high yield, and no damage to the natural resources of *S. hexandrum*.

At present, the research on *S. hexandrum* mainly focuses on chemical composition extraction (Wang *et al*., 2023), tissue culture condition optimization (Sharma *et al*., 2022), podophyllotoxin biosynthesis mechanism (Guo *et al*., 2023), genetic diversity (Naik *et al*., 2010; Liu *et al*., 2014), transcriptome sequencing (Grabherr *et al*., 2011; Yang *et al*., 2011; Kumar *et al*., 2017) and so on. High-throughput transcriptome sequencing technology can comprehensively obtain transcript information and gene sequences of biological tissues or organs of species in a certain state, so as to study gene expression levels. Therefore, transcriptome analysis can effectively develop and mine functional genes of non-reference genome species in batches (Huang, 2020). Transcriptome high-throughput sequencing technology has become an important means to study the development of medicinal plants and elucidate the key gene mining and transcriptional expression regulation of plant active components and secondary metabolite biosynthesis pathways (Kapoor *et al*., 2021). In recent years, the research on *S. hexandrum* by transcriptome analysis technology has gradually increased. Kumari *et al*., (2014) tested transcriptome of rhizome tissue of *S. hexandrum* at 15℃ and 25°C, revealed temperature response of transcriptome of *S. hexandrum*. Zhao *et al*. (2023a) used the methods of HPLC, proteomic, transcriptomic in light-induced flavonoid biosynthesis in *S. hexandrum*. Guo *et al*., (2023) employed the RNA-seq technology to identify different somatic embryogenesis (SE) stages of *S. hexandrum*, enlightened the key plant hormones in SE stages of *S. hexandrum*. The phylogenomic analyses of *Podophylloideae* between Eastern Asia and Eastern North America by mRNA-Seq indicated *S. hexandrum* was identified as sister to the remainder of *Podophylloideae* (Ye *et al*., 2022). However, the complete genetic information and transcriptome of *S. hexandrum* is very limited, which influences in-depth research and development and utilization of *S. hexandrum*. Therefore, it is urgent to use high-throughput transcriptome sequencing technology at the RNA level to understand the overall transcription level of this characteristic Chinese herbal medicine, so as to provide a theoretical basis and excellent genetic resources for the biosynthesis pathway analysis, molecular directional breeding and innovative development and utilization of podophyllotoxin in *S. hexandrum*.

Material and Methods

Plant materials: *S. hexandrum* is mainly distributed in Nepal, Bhutan, northern India, Pakistan, eastern Afghanistan and Kashmir. In China, it is mainly distributed in Yunnan, Sichuan, Tibet, Gansu, Qinghai, and Shaanxi (Anon., 2011). *S. hexandrum* is located in the rocky gap of alpine grassland, humus-rich mountain podzolic soil, dark gray calcium soil, gray-cinnamonic soil, and mountain brown soil. It usually grows in a wide valley or a valley forest , rock crevices, forest edges, slopes, riverside wetland shrubs with secondary vegetation and good light transmittance by an altitude of 1500 ~ 4500 m, a small number of growing alpine meadows or open grasslands (Zhao *et al*., 2011). The soil is mostly fertile dark humus soil, yellow clay soil, and sandy soil. It is suitable for cold and humid, low temperature and rainy in summer, and dry-cold climates in winter and spring. The minimum temperature is -10°C, and the annual precipitation is 400-900mm, mostly concentrated in June-September(Lv *et al*., 2020).

S. hexandrum used in this study was collected in 2018 from Mingxing Temple (N34^{0}15′52.21″, E107^{0}45′34.75″) in Taibai Mountain, Qinling Mountains, Meixian County, Shaanxi Province, China (Fig. 1.). *S. hexandrum* with the same ages (7 years) and growth status were collected as test samples. Three biological replicates were used, each of which contained three *S. hexandrum* individuals in this study. After sampling, the sample was washed with distilled water, and the water remaining on the surface of the sample was dried with absorbent paper. The leaves, roots, and stems of the samples were sub-packed. After liquid nitrogen quick freezing, the treated samples were stored in refrigerator at -80°C.

Fig. 1. Individual morphology of *S. hexandrum.*

Total RNA extraction and identification: Total RNA was extracted from leaves, stems, and roots of *S. hexandrum* by RNAiso Plus kit (TakaRa Bio, CHN) with the instructions provided of the supplier. The RNA integrity of the samples was detected by non-denaturing agarose gel electrophoresis, and the purity and concentration of RNA were determined by Nanodrop2000 ultramicro spectrophotometer.

RNA library construction and sequencing: Before the construction of the RNA library, the extracted total RNA of leaves, stems and roots of *S. hexandrum* should be mixed in equal volume to remove the genomic DNA in the total RNA. Enriched mRNA was analyzed using magnetic beads with Olige (dT), and PCR data were used to construct a sequencing library. The sequencing was determined by Gideon Biotechnology. The Illumina HiSeqTM2000 Sequencing System was used as the library sequencing platform. The sequencing method used was the ' paired-end' method. The original image data obtained after sequencing was processed by base calling to obtain raw reads. After downloading the raw data, contaminated sequences, and poor-quality reads were eliminated to provide clean reads. Next, the assembly software SOAPdenove was used to assemble the clean reads obtained after screening to obtain unigenes.

Analysis of sequencing results: There is no reference genome in *S. hexandrum* as the research background. In this experiment, the unigenes obtained by splicing in the transcriptome data were compared with three known protein databases of GO, COG, and KEGG to obtain the functional information of these unigenes.

Results

Sequencing yield and assembly: A total of 108654776 reads in the transcriptome of *S. hexandrum* were collected by sequencing, and 74026 unigenes were produced with the use of specialized software assembly. As shown in (Fig. 2.) when the number of reads with a length of 1-10 is 26010, the number of unigenes is close to 27000. On top of that, when the number of reads with a length of 11-100 reads was 25220, the number of unigenes was also close to 27000. Additionally, when the number of reads with a length of 101-200 was 4473, the number of unigenes was close to 5400. Furthermore, when the number of reads with a length of 201-300 was 2246, the number of unigenes was close to 2500, and the number of other unigenes was analyzed in turn. Therefore, it is speculated that there is not a completely proportional relationship between the amount of sequencing throughput and the number of unigenes obtained by splicing.

COG classification of unigenes: By COG classification, the unigenes in the transcriptome of the sample could be divided into 24 categories according to the function (represented by the letters A-Z in the figure), and the number of genes for each function of A-Z was counted (Table 1 and Fig. 3). Showed that the sample unigenes' COG function included the majority of the live body's activities, the number of genes in the overall functional class was the largest, the general function prediction was the largest, and the number of genes related to the nuclear structure was the least at 8 (Table 1 and Fig. 3.).

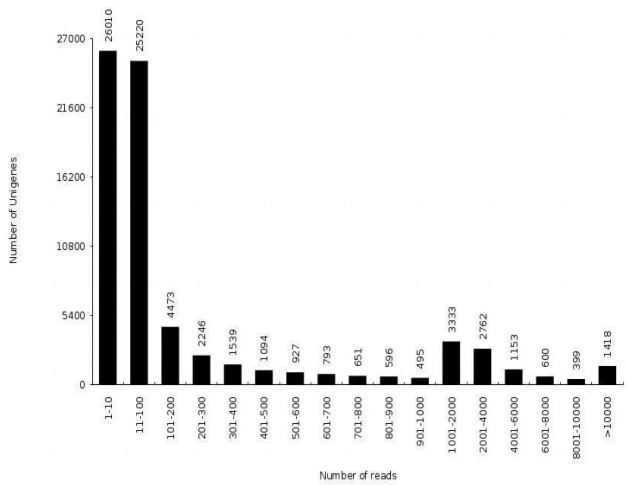

Fig. 2. Results of *S. hexandrum* sequencing.

GO classification of unigenes: According to the GO database, the gene classification function of the sample was classified into 70 functional categories (Fig. 4.). Among them, there are 31 species whose ontology function category is a biological process. Among the 31 gene functions, the top three genes in the number of genes were genes controlling metabolic processes, genes controlling cellular processes, and genes controlling single organism, with the numbers 7751, 7064, and 5444 respectively. There were 14 kinds of cellular components in the ontology function category. Among the 14 kinds of gene functions, the gene that regulated the cell had the most genes, with a number of 7653. The extracellular matrix and extracellular matrix part were controlled by the fewest number of genes, with 2 genes each. There were 25 kinds of molecular functions in the ontology function category. The top three genes were genes that controlled metabolic processes, genes that control catalytic activity, and genes that controlled binding. The number of genes were 8250,7258, and 6639, respectively.

KEGG pathway analysis: Combined with the KEGG database, unigenes obtained from short sequences by specific software, assembly was included in 125 metabolic pathways. The number of unigenes annotated to the metabolic pathway was the largest, with a total of 2921 (24.96 %), and the ID controlling this pathway was ko01100. The pathway, second only to the number of annotated genes in the metabolic pathway is the biosynthetic pathway of secondary metabolites, with a number of 1425 (12.18 %), and the code assigned to it is ko01110. The third is the pathway that controls ribosome metabolism. The number of annotated genes is 889 (7.6%), the ID number of this pathway was ko03010, and the number is also large, accounting for the third place. The fourth was the pathway controlling protein synthesis in the endoplasmic reticulum. The number of annotated genes was 530 (4.53%), and the pathway ID was ko04141. The number of unigenes annotated to Betalain and Benzoxazinoid biosynthesis was the least, each accounting for 0.01%, with ID numbers ko00965 and ko00402, respectively. Since there is no flow chart of metabolic pathways in the database, the secondary metabolite biosynthesis pathway, ribosomal metabolism pathway, and endoplasmic reticulum protein processing pathway were analyzed here. The results of the secondary metabolite biosynthesis pathway, ribosome metabolism pathway, and endoplasmic reticulum protein processing pathway were shown in (Figs. 5-6).

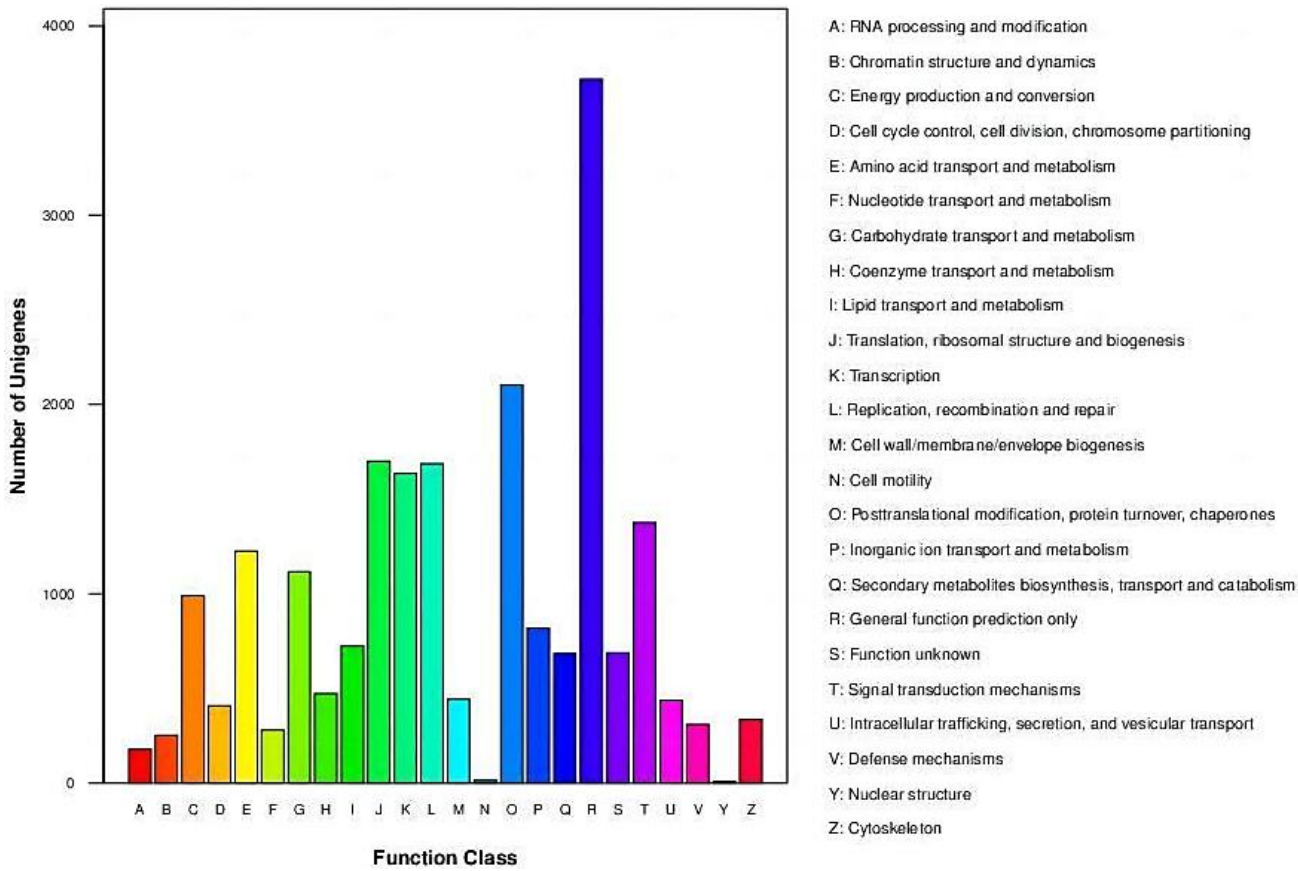

Fig. 3. COG classification of unigenes sample *S. hexandrum* transcriptome.

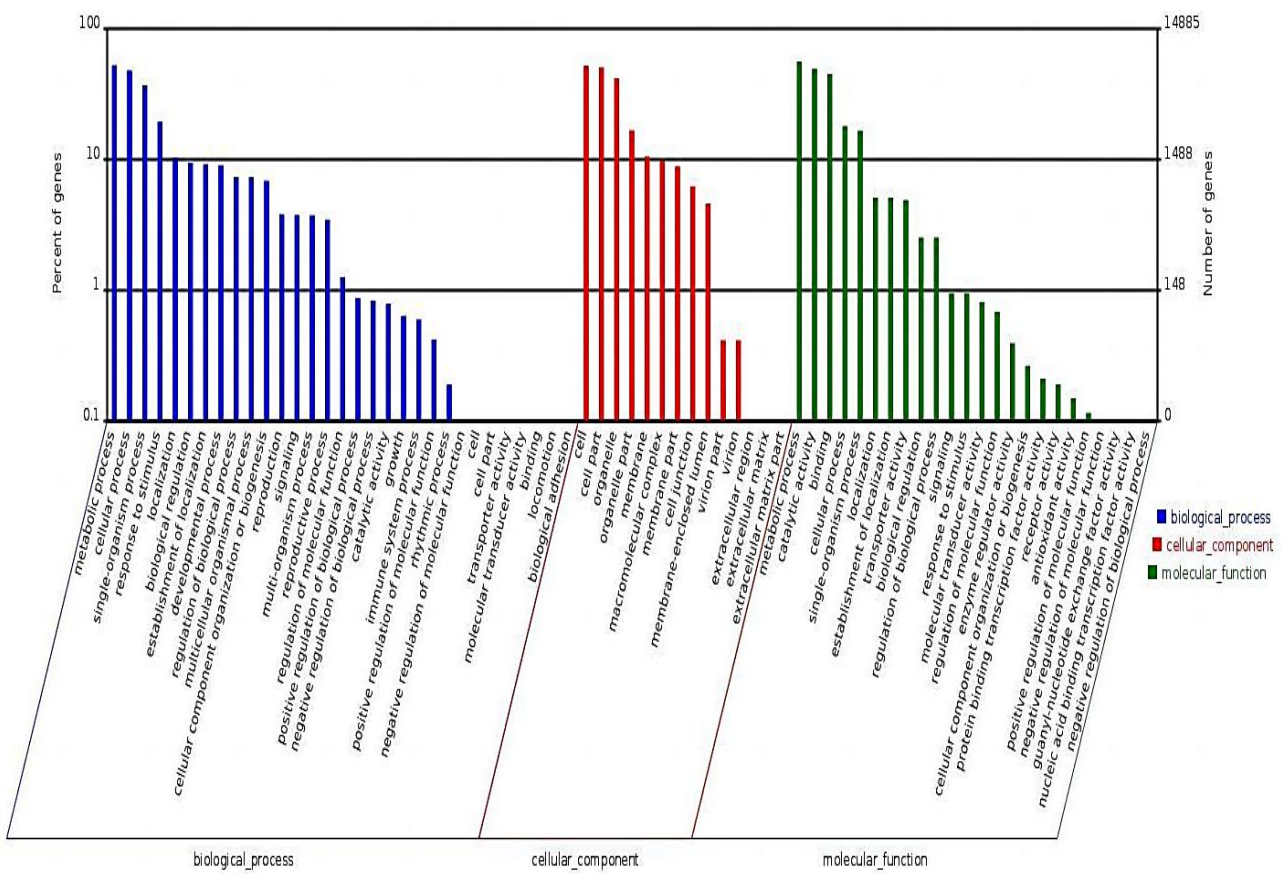

Fig. 4. GO functional classification of unigenes sample *S. hexandrum*.

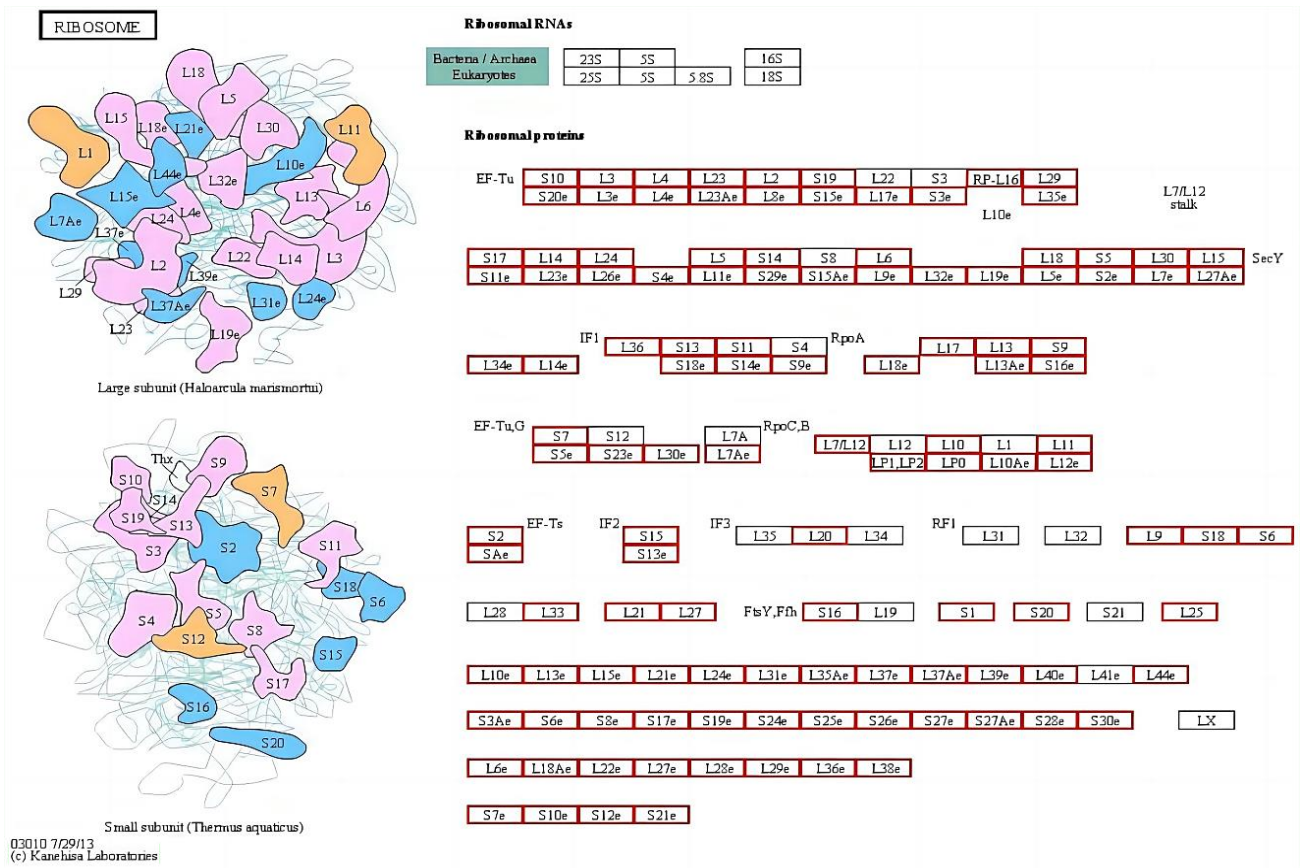

Fig. 5. Ribosome metabolic pathway.

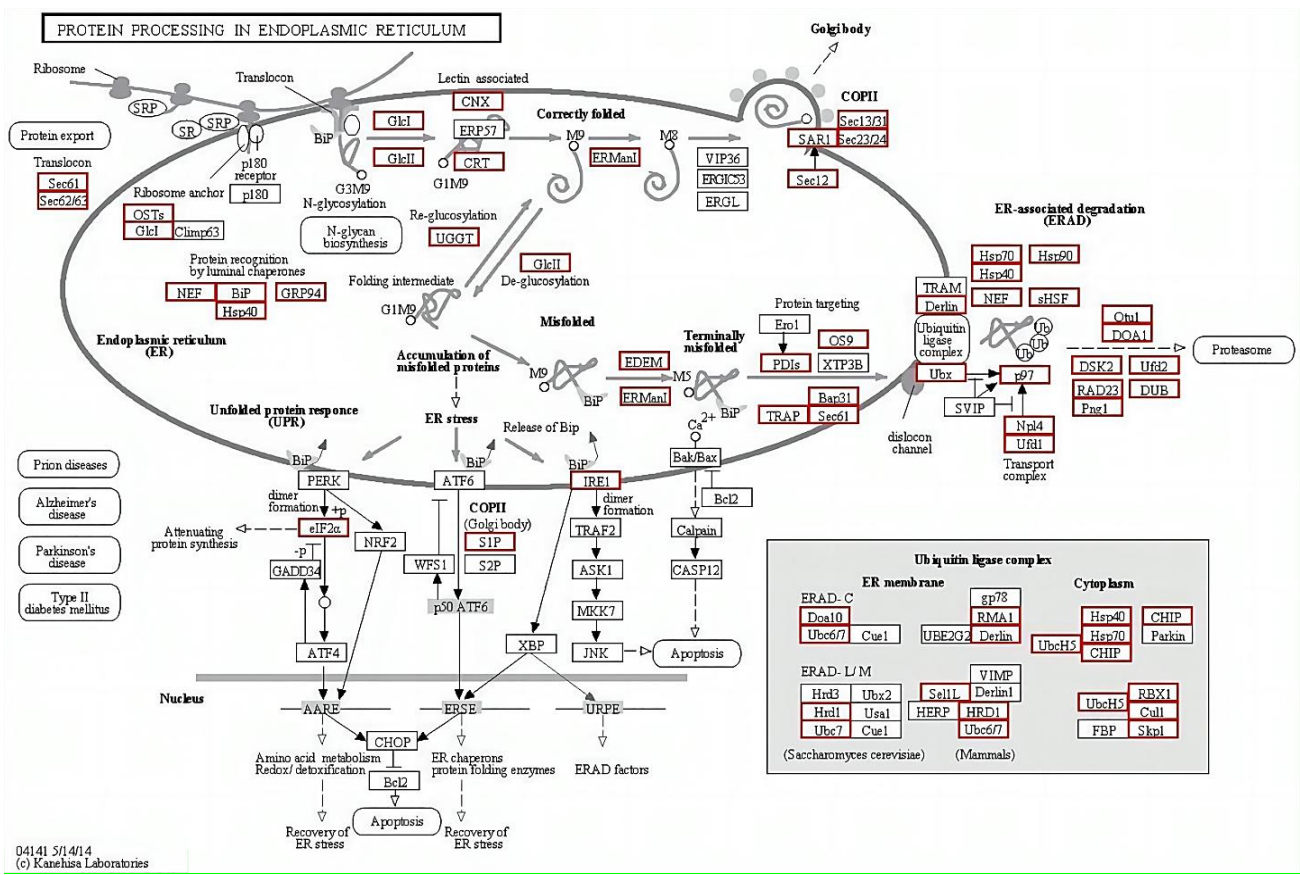

Fig. 6. Endoplasmic reticulum protein processing pathway.

Table 1. COG classification of unigenes in *S. hexandrum* transcriptome.

No.	Functional categories	Gene dosage
A	RNA processing and modification	179
B	Chromatin structure and dynamics	253
C	Energy production and conversion	989
D	Cell cycle control, cell division, chromosome partitioning	408
E	Aminoacid transport and metabolism	1225
F	Nucleotide transport and metabolism	279
G	Carbohydrate transport and metabolism	1117
H	Coenzyme transport and metabolism	472
I	Lipid transport and metabolism	724
J	Translation, ribosomal structure, and biogenesis	1700
K	Transcription	1636
L	Replication, recombination, and repair	1687
M	Cell wall/membrane/envelope biogenesis	443
N	Cell motility	14
O	Posttranslational modification, protein turnover, chaperones	2104
P	Inorganic ion transport and metabolism	818
Q	Secondary metabolites' biosynthesis, transport, and catabolism	684
R	General function prediction	3718
S	Function unknown	688
T	Signal transduction mechanisms	1376
U	Intracellular trafficking, secretion, and vesicular transport	438
V	Defense mechanisms	309
Y	Nuclear structure	8
Z	Cytoskeleton	335

Through KEGG annotation, the results showed that a certain number of unigenes were annotated to the KEGG metabolic pathway where the secondary metabolite biosynthesis is located in the transcriptome of the sample *S. hexandrum*. The number of unigenes was 1425 (12.18 %), and the code given to it was ko01110; secondly, some unigenes were annotated to the KEGG metabolic pathway where ribosome metabolism is located in the transcriptome of sample *S. hexandrum*. The number of unigenes was 889 (7.6 %), and the ID number of this pathway was ko03010 (Fig. 5.); thirdly, unigenes were annotated to the KEGG metabolic pathway where endoplasmic reticulum protein processing was located in the transcriptome of the sample. The number of unigenes was 530 (4.53%), and the ID number of this pathway was ko04141 (Fig. 6.).

Discussion

Transcriptomics is a discipline that studies gene expression and transcriptional regulation at the RNA expression level. RNA sequencing technology can not only identify and annotate the function of genes, but also study the quantitative gene expression level (Wilhelm & Landry, 2009)、identify differentially expressed genes (Camarena *et al*., 2010), and analyze splice variants (Zenoni *et al*., 2010). The concept of sequencing while synthesizing is used by second-generation sequencing technology. High-throughput sequencing (HTS) is another name for it since it can sequence millions of nucleic acid molecules at once and produce tens of billions of base sequences (Tang *et al*., 2019). Second-generation sequencing improves the low throughput of the first-generation sequencing method, maintains high accuracy, and improves the sequencing throughput (Lan *et al*., 2020). Because multiple samples can be sequenced simultaneously and the cost is low, it is used by more and more researchers to analyze a large number of biological problems. Klepikova *et al*., (2021) analyzed the transcriptome of 19 organs at different developmental stages of the orchid *Phalaenopsis equestris*. The obtained transcriptome map lays a theoretical foundation for further study of the unique traits of this *Phalaenopsis equestris* and other orchids. In order to understand the evolutionary and environmental importance of the terpenes in Piper species in black pepper berries by transcriptome sequencing, George *et al*., (2021) studied the whole terpene synthase family. Using methylome and whole transcriptome sequencing profiles, Lyu *et al*., (2022) investigated the molecular regulatory functions in response to drought stress of sea buckthorn leaves at epigenetic and transcriptional levels, which would support genetic breeding for the enhancement of crop drought resistance.

Currently, key gene mining at the transcriptome level and researching the molecular mechanism of production of significant plant secondary metabolites need the use of RNA-Seq (Peng *et al*., 2015). Especially in medicinal plant research, RNA-Seq has been widely used to study the synthesis pathways of key medicinal compounds in different medicinal plants. Ouyang *et al*., (2021) used the Illumina Hiseq 4000 high-throughput sequencing platform to sequence *Coix lacryma-jobi*, and the analysis results provided a data basis for further improving the medicinal value of *Coix lacryma-jobi*. Nett *et al*. (2020) sequenced and analyzed the *Gloriosa superba* transcriptome, and identified 10 methyltransferase candidate genes, which successfully demonstrated the metabolic pathway of the typical tropolone skeleton of colchicine in *N. benthamiana*. Su *et al*., (2021) found that 47 CYPs and 22 TFs were strongly correlated with tanshinone-related metabolites

and were candidate core genes related to tanshinone synthesis by transcriptome analysis of wild-type and mutant *Salvia miltiorrhiza*. In the study of *S. hexandrum*, six candidate genes of the podophyllotoxin synthesis pathway were identified by transcriptome analysis of the public database and the mechanically damaged leaves of *S. hexandrum*. The differential conformation of podophyllotoxin, epipodophyllotoxin, was successfully synthesized by co-expression with known genes in tobacco (Lau & Sattely, 2015). At present, the genome of *S. hexandrum* has not been sequenced, and there is no complete genome sequence for reference. Therefore, RNA-Seq independent of the genome reference sequence provides technical support for studying the molecular mechanism of podophyllotoxin biosynthesis and key gene mining in *S. hexandrum* at the transcriptome level.

In this study, 74026 unigenes with high annotation reliability were classified by COG and GO functions. The Cluster of Orthologous Groups of Proteins (KOG/ COG) database is a phylogenetic relationship based on the complete genome-encoded proteins of eukaryotes, bacteria, and algae. The comparison with the COG database showed that the transcriptome of *S. hexandrum* involved most of the life activities necessary for the normal survival of life. In detail, organisms need energy for normal activities, and the number of genes that control the production and conversion of energy is relatively large, 989; macromolecules such as protein and fat are necessary for the existence of living organisms, so the number of genes that control the transport and metabolism of their monomeric substances such as amino acids, nucleotides, coenzymes, and carbohydrates is as high as 3000. In order to expand the number of populations, organisms must reproduce, hence the number of genes that control cell division, chromosome division, genetic material recombination, and repair of these functions is also relatively large. GO is interpreted as gene ontology. Compared with the GO database, the GO database is mainly divided into three ontology functional categories: biological process, molecular function, and cell component. In the category of biological processes, the number of genes controlling metabolic process is the largest, which is 7751. In the category of molecular function, the number of genes controlling metabolic processes is the largest, 8250. In the cellular component category, the number of genes controlling cell synthesis is the largest, at 7653, which is consistent with the GO classification results of Bhattacharyya *et al*., (2013) by sequencing the *Podophyllum hexandrum* (synonym of *S. hexandrum*) cell culture transcriptome. Indicating that functions such as metabolism, catalysis, and binding play an essential role in the growth and reproduction of a living organism. KEGG is interpreted as a whole genome and metabolic pathway database, which is a relatively systematic database that can be used to analyze gene function. Compared with the KEGG database, it was found that the number of genes annotated to the metabolic pathway (ko01100) was the largest, but did not differ much from the number of genes controlling the synthesis pathway of secondary metabolites and the ribosome metabolic pathway. By comparing the results with these three databases, it can be seen that the number of genes controlling the metabolism of organisms occupies a large proportion. The result of this experiment is consistent with the earlier conclusions of Wang *et al*., (2015).

In this study, the transcriptome of the endangered anticancer medicinal plant *S. hexandrum* was used as the starting point, and the Illumina HiSeqTM2000 Sequencing System was used as the library sequencing platform for high-throughput sequencing of the transcriptome of *S. hexandrum*. By comparing with the three protein databases of COG, GO, and KEGG, the transcriptome data of *S. hexandrum* were comprehensively and systematically compared and analyzed. The results of this study are helpful to understand the overall transcription of *S. hexandrum* at the RNA level and provide a new theoretical basis and excellent gene resources for the analysis of the molecular mechanism of the podophyllotoxin biosynthetic pathway of *S. hexandrum* and its molecular directional breeding. It has important theoretical and practical guiding significance for the efficient development and innovative utilization of *S. hexandrum* resources. At the same time, the results can also provide some reference for the study of the biosynthesis molecular mechanism of other secondary metabolites of *S. hexandrum* and the transcriptome study of other medicinal plants.

Conclusions

High-throughput sequencing technology was used for transcriptome sequencing of *S. hexandrum* in this study. The transcriptome data obtained by sequencing were compared with three known protein databases of GO, COG, and KEGG. A total of 108654776 reads were obtained by sequencing, and 74026 unigenes were obtained after assembly. The COG functional classification of the transcriptome of *S. hexandrum* showed that the unigenes of *S. hexandrum* transcriptome were classified into the most General functional prediction, which was 3718; the genes controlling protein folding, translation, replication, and transcription were 2104, 1700, 1687 and 1636, respectively. These data suggest that protein synthesis plays an important role in biological activities. From the GO classification results chart of the transcriptome data of *S. hexandrum*, it can be found that the number of genes belonging to metabolic process, catalytic activity, and binding function is the largest, 8250, 7258 and 6639, respectively. By analyzing the results of KEGG metabolic pathways, a clear conclusion can be drawn: all unigenes assembled by short sequences through software are classified into 125 metabolic pathways, of which the number of unigenes annotated to the metabolic pathway (ko01100) is the largest, 2921, accounting for 24.96%; the number of unigenes annotated to Betalain biosynthesis (ko00965) and Benzoxazinoid biosynthesis (ko00402) was the least, accounting for 0.01%, respectively.

References

Anonymous. 1975. *National compilation of herbal medicines*. People's Medical Publishing House, Beijing.

Anonymous. 1987. *List of Rare and Endangered Plants in China*. Vol: 1. Science Press, Beijing.

Anonymous. 2002. *Chinese materia medica: Tibetan medicine roll*. Shanghai Science and Technology Press, Shanghai.

Anonymous. 2011. *Flora of China*. Vol: 19. Science Press, Beijing; Missouri Botanical Garden Press, St. louis, p. 783.

Anonymous. 2020. *Pharmacopoeia of the People's Republic of China*. China Medical Science Press, Beijing, p. 48-49.

Bhattacharyya, D., R. Sinha, S. Hazra, R. Datta and S. Chattopadhyay. 2013. De novo transcriptome analysis using 454 pyrosequencing of the Himalayan Mayapple, Podophyllum hexandrum. *BMC Genom.*, 14: 1-13.

Camarena, L., V. Bruno, G. Euskirchen, S. Poggio and M. Snyder. 2010. Molecular mechanisms of ethanol-induced pathogenesis revealed by RNA-sequencing. *PLoS Pathog.*, 6(4): e1000834.

Canel, C., F.E. Dayan, M. Ganzera, I.A. Khan, A. Rimando, C.L.J. Burandt and R.M. Moraes. 2001. High yield of podophyllotoxin from leaves of *Podophyllum peltatum* by in situ conversion of podophyllotoxin 4-O-β-D-glucopyranoside. *Planta Med.*, 67(01): 97-99.

Chattopadhyay, S., A.K. Srivastava, S.S. Bhojwani and V.S. Bisaria. 2001. Development of suspension culture of *Podophyllum hexandrum* for production of podophyllotoxin. *Biotechnol. Lett.*, 23: 2063-2066.

Damayanthi, Y. and J.W. Lown. 1998. Podophyllotoxins: Current status and recent developments. *Curr. Med. Chem.*, 5: 205-252.

Fay, D. and H. Ziegler. 1985. Botanical source differentiation of podophyllum resin by HPLC. *J. Liq. Chromatogr.*, 8(8): 1501-1505.

Fu, L.G. 1991. *Chinese Plant Red Book - Rare and Endangered Plants*. Vol: 1. Science Press, Beijing.

George, J.K., S. Shelvy, A.M. Fayad, P. Umadevi, U.B. Angadi, M.A. Iquebal, S. Jaiswal, A. Rai and D. Kumar. 2021. De novo transcriptome sequencing assisted identification of terpene synthases from black pepper (*Piper nigrum*) berry. *Physiol. Mol. Biol. Plants*, 27(5): 1153-1161.

Giri, A. and N.M. Lakshmi. 2000. Production of podophyllotoxin from *Podophyllum hexandrum*: A potential natural product for clinically useful anticancer drugs. *Cytotechnology*, 34: 17-26.

Grabherr, M.G., B.J. Haas, M. Yassour, J.Z. Levin, D.A. Thompson, I. Amit, X. Adiconis, L. Fan, R. Raychowdhury and Q. Zeng. 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.*, 29(7): 644-652.

Guo, S., Y. Chen, Y. Zhu and M. Tian. 2023. Transcriptome analysis reveals differentially expressed genes involved in somatic embryogenesis and podophyllotoxin biosynthesis of *Sinopodophyllum hexandrum* (Royle) T. S. Ying. *Protoplasma*, 260(4): 1221-1232.

Guo, Y.L., H.Y. Wei, C.Y. Lu, H.L. Zhang and W. Gu. 2014. Predictions of potential geographical distribution of Sinopodophyllum hexandrum under climate change. *Chinese J. Plant Ecol.*, 38(3): 249-261.

Huang, K., W. Jiang, J.F. Zhao, X. Liu, Z.W. Zhang, C.H. Wang, S.Y. Qin and G.Y. Zhong. 2012. Research progress on lignans chemical constituents and their activities in *Sinopodophyllum hexandrum*. *Tradit. Chin. Drug Res. Clin. Pharmacol.*, 23(02): 232-238.

Huang, Q.L. 2020. Transcriptome Analysis of Alpinia officinarum. *Tradit. Chin. Drug Res. Clin. Pharmacol.*, 31(06): 729-734.

Jackson, D.E. and P.M. Dewick. 1984. Aryltetralin lignans from *Podophyllum hexandrum* and *Podophyllum peltatum*. *Phytochem.*, 23(5): 1147-1152.

Kapoor, B., A. Kumar and P. Kumar. 2021. Transcriptome repository of North-Western Himalayan endangered medicinal herbs: a paramount approach illuminating molecular perspective of phytoactive molecules and secondary metabolism. *Mol. Genet. Genom.*, 296(6): 1177-1202.

Kelleher, J.K. 1978. Correlation of Tubulin-Binding and Antitumor Activities of Podophyllotoxin Analogs 1, 2. *Cancer Treat. Rep.*, 62(10): 1443-1447.

King, M.L.S. and M.M. Sullivan. 1946. The similarity of the effect of podophyllin and colchicine and their use in the treatment of *Condyl. Acuminata Sci.*, 104(2698): 244-245.

Klepikova, A.V., A.S. Kasianov, M.A. Ezhova, A.A. Penin and M.D. Logacheva. 2021. Transcriptome atlas of *Phalaenopsis equestris*. *Peer J.*, 9: e12600.

Kumar, P., V. Jaiswal, T. Pal, J. Singh and R.S. Chauhan. 2017. Comparative whole-transcriptome analysis in Podophyllum species identifies key transcription factors contributing to biosynthesis of podophyllotoxin in *P. hexandrum*. *Protoplasma*, 254(1): 217-228.

Kumari, A., H.R. Singh, A. Jha, M.K. Swarnkar, R. Shankar and S. Kumar. 2014. Transcriptome sequencing of rhizome tissue of *Sinopodophyllum hexandrum* at two temperatures. *BMC Genom.*, 15: 1-17.

Lai, W., X. Ye, G. Wen, C. Shi, W. Zhang, L. Ye and G. Zhang. 2022. Analysis of potential suitable regions for the precious Tibetan medicine *Sinopodophyllum hexandrum* based on the optimized MaxEnt model. *J. Fujian Agric. For. Univ. (Nat. Sci. Ed.)*, 51(1): 112-120.

Lan, W., Y. Sun and J. Xie. 2020. Application of High-throughput Sequencing Technology in Aquatic Products during Processing and Storage. *Packag. Eng.*, 41(21): 11-17.

Lau, W. and E.S. Sattely. 2015. Six enzymes from mayapple that complete the biosynthetic pathway to the etoposide aglycone. *Science*, 349(6253): 1224-1228.

Li, C.D., W. Li, M.F. Li, P. Sun, W.T. Wang and Y.Y. Song. 2008. Seed Dor mancy and Ger mination in Endangered Plant *Podophyllum hexandrum* Royle. *Bull. Bot. Res.*, 28(5): 618-621.

Li, Z.C., X.L. Wang and X.J. Ge. 2005. The biology and conservation of *Sinopodophyllum hexandrum* an endangered medicinal herb. *Guihaia*, 25(2): 179-185.

Linghu, Y.W., S.F. Li, B. Li, R.J. Li and Y. Zhang. 2011. Study on the introduction and cultivation of Sinopodophyllum in Xi 'an area and its economic value. *Shanxi J. Agric. Sci.*, 57(6): 5-7.

Liu, W., D. Yin, J. Liu and N. Li. 2014. Genetic diversity and structure of *Sinopodophyllum hexandrum* (Royle) Ying in the Qinling Mountains, China. *PLoS One*, 9(10): e110500.

Lv, R., F. Wei and L. Jin. 2020. Study on Ecological Suitability of Endangered Chinese Materia Medica Sinopodophyllum. *Chin. J. Inf. Tradit. Chin. Med.*, 27(10): 5-8.

Lyu, Z., G. Zhang, Y. Song, S. Diao, C. He and J. Zhang. 2022. Transcriptome and DNA methylome provide insights into the molecular regulation of drought stress in sea buckthorn. *Genomics*, 114(3): 110345.

Moraes, R.M., C. Burandt Jr, M. Ganzera, X. Li, I. Khan and C. Canel. 2000. The American mayapple revisited: *Podophyllum peltatum*: still a potential cash crop? *Econ. Bot.*, 54(4): 471-476.

Naik, P.K., M.A. Alam, H. Singh, V. Goyal, S. Parida, S. Kalia and T. Mohapatra. 2010. Assessment of genetic diversity through RAPD, ISSR and AFLP markers in *Podophyllum hexandrum*: A medicinal herb from the Northwestern Himalayan region. *Physiol. Mol. Biol. Plants*, 16(2): 135-148.

Nett, R.S., W. Lau and E.S. Sattely. 2020. Discovery and engineering of colchicine alkaloid biosynthesis. *Nature*, 584(7819): 148-153.

Ouyang, Y., L. Li, H. Shi, Y. Zhao, F. Zhu, Y. Yin and L. Wang. 2021. Transcriptome sequencing and gene function annotation of *Coix larchryma-jobi* L. var. *ma-yuen* Stapf. *Cent. South Pharm.*, 19(07): 1286-1293.

Peng, H., X. He, J. Gao, H. Ma, Z. Zhang, Y. Shen, G. Pan and H. Lin. 2015. Transcriptomic changes during maize roots development responsive to Cadmium (Cd) pollution using comparative RNAseq-based approach. *Biochem. Biophys. Res. Commun.*, 464(4): 1040-1047.

Sharma, N., M. Thakur, P. Sharma, B. Dutt and Y.P. Sharma. 2022. *In vitro* propagation from seeds and enhanced synthesis of podophyllotoxin from root callus of *Sinopodophyllum hexandrum* (Royle) TS Ying (Himalayan Mayapple)− An endangered medicinal plant. *Ind. Crops Prod.*, 186: 115300.

Stähelin, H.F. and V.W. Albert. 1991. The chemical and biological route from podophyllotoxin glucoside to etoposide: Ninth Cain memorial Award lecture. *Cancer Res.*, 51(1): 5-15.

Su, Y., J. Zhang, Z. Xu, J. Li, P. Wang, Z. Song, G. Tian, L. Li, J. Song and J. Wang. 2021. Integrative analysis of metabolome and transcriptome reveals the mechanism of color formation in white root (*Salvia miltiorrhiza*). *Ind. Crops Prod.*, 170: 113784.

Tang, D., G. Zhang and X. Zhao. 2019. Bioinformatics analysis of transcriptome sequencing based on next generation sequencing. *J. Henan Univ. (Méd. Sci.)*, 38(01): 67-76.

Tomioka, K., Y. Kubota and K. Koga. 1989. Synthesis and antitumor activity of podophyllotoxin aza-analogues. *Tetrahedron Lett.*, 30(22): 2953-2954.

Wang, J., H. Li, Y. Li, A. Yang, S. Bao, Y. Zhang, Q. Du, Z. Zheng and X. Wang. 2023. Sinoflavonoids NJ and NK, anti-inflammatory prenylated flavonoids from the fruits of *Podophyllum hexandrum* Royle. *Nat. Prod. Res.,* 1-5. https://doi.org/10.1080/14786419.2023.2188590.

Wang, X., H. Tan, Z. Chen, L. Meng, W. Wang and S. Fan. 2015. RNA-Seq-based transcriptome assembly and analysis of Forsythia suspensa and development of SSR molecular markers. *Sci. Sin. Vitae*, 45(3): 301-310.

Wilhelm, B.T. and J.R. Landry. 2009. RNA-Seqquantitative measurement of expression through massively parallel RNA-sequencing. *Methods*, 48(3): 249-257.

Yan, S.T., H. Fan, R.L. Li, Y.L. Guo, Q. Liu, F. Gao, L. Ou, L. Chen, M. Li, P.F. Wei and L. Zhang. 2020. Research progress on chemical constituents and pharmachological Activities of *Sinopodophyllum hexandrum*. *Chin. Wild Plant Resour.*, 39(07): 43-50.

Yang, H., Y. Mao, F. Kong, G. Yang, F. Ma and L. Wang. 2011. Profiling of the transcriptome of *Porphyra yezoensis* with Solexa sequencing technology. *Chin. Sci. Bull.*, 56: 2119-2130.

Yang, L. and J. Lu. 2022. Resources Status and Protection of Endangered Tibetan Medicinal Plants in Sejila Mountain. *J. Shandong For. Sci. Technol.*, 52(01): 105-110.

Yang, Z.M., J. Cao, Y.P. Du and Z.Q. Yu. 2022. Study on chemical constituents and medicinal value of *Sinopodophyllum emodi. Contemp. Anim. Husb.*, 2022(12): 44-46.

Ye, W., S. Zhu, H.P. Comes, T. Yang, L. Lian, W. Wang and Y. Qiu. 2022. Phylogenomics and diversification drivers of the Eastern Asian–Eastern North American disjunct Podophylloideae. *Mol. Phylogen. Evol.*, 169: 107427.

Zenoni, S., A. Ferrarini, E. Giacomelli, L. Xumerle, M. Fasoli, G. Malerba, D. Bellin, M. Pezzotti and M. Delledonne. 2010. Characterization of transcriptional complexity during berry development in Vitis vinifera using RNA-Seq. *Plant Physiol.*, 152(4): 1787-1795.

Zhao, J.F., X. Liu, C.H. Wang, Z.W. Zhang, S.Y. Qin and G.Y. Zhang. 2011. Resource investigation of rare and endangered medicinal plant *Sinopodophyllum hexandrum*. *China J. Chin. Mater. Med.*, 36(10): 1255-1260.

Zhao, Q., M. Dong, M. Li, L. Jin and P.W. Paré. 2023a. Light-induced flavonoid biosynthesis in *Sinopodophyllum hexandrum* with high-altitude adaptation. *Plants*, 12(3): 575.

Zhao, Q., M. Li, M. Li, L. Jin and J. Wei. 2023b. Changes in growth characteristics and secondary metabolites in *Sinopodophyllum hexandrum* with increasing age. *Ind. Crops Prod.*, 196(2023): 116509.

(Received for publication 14 October 2022)