INTERNATIONAL JOURNAL OF AGRICULTURE & BIOLOGY
ISSN Print: 1560–8530; ISSN Online: 1814–9596
19–0482/2019/22–4–793–800
DOI: 10.17957/IJAB/15.1132
http://www.fspublishers.org

Full Length Article

De Novo Transcriptome Sequencing of *Ilex cornuta* and Analysis of Genes Involved in Triterpenoid Biosynthesis

Hui Zeng, Li Zhu, Liang-Qiong Ma, Wei-Wei Zhang*, Feng Xu and Yong-Ling Liao
College of Horticulture and Gardening, Yangtze University, Jingzhou, 434025, Hubei, China
*For correspondence: wwzhangchn@163.com; wwzhang@yangtzeu.edu.cn

Abstract

Ilex cornuta is a highly valuable medicinal plant rich in triterpenoids and flavonoids. Although many terpene medicinal ingredients have been isolated and identified from *I. cornuta*, genomic resources for investigating the biosynthesis of these components are lacking. Therefore, in this paper, we executed transcriptome sequencing in leaves, fruit and roots of *I. cornuta* which obtained 124,003 unigenes with an average length of 661 bp by *de novo* assembly. A total of 81,691 (65.9%) unigenes were functionally annotated through the Gene Ontology and Cluster of Orthologous Groups databases, and pathway analysis was performed by the Kyoto Encyclopedia of Gene and Genomes to understand the biological functions of genes. A total of 8966 differentially expressed genes were identified after comparing three organs. The qRT-PCR analysis found that the expression level of seven unigenes in roots was visibly higher than that in other organs. The HPLC results also indicated that triterpenoid (Total ursolic acids and oleanolic acids) was the highest in roots among the three organs. This study provides the first public transcriptome sequencing of *I. cornuta*, and offers a theoretical basis for molecular mechanism research on the triterpenoid biosynthesis. © 2019 Friends Science Publishers

Keywords: Transcriptome; Differentially expressed gene; Triterpenoid; qRT-PCR

Introduction

Chinese holly (*Ilex cornuta* Lindl. ex Part.) is a dioecious evergreen thorny shrub belonging to Aquifoliaceae that is suitable for garden landscape. It is widely distributed in China and commonly decorated during Christmas in Europe and the U.S. (Miller, 1998; Sibley and Wilson, 2002). In China, this plant not only holds cultural significance, but the aqueous extracts from the leaf, root, stem, and fruit have been directed at traditional Chinese medicine for contraception, calming the liver, clearing heat, and tonifying the kidney (Wei, 1988; Kuang *et al.*, 2011). The plant's remarkable pharmacological property is dependent on its secondary metabolites.

I. cornuta is rich in triterpenoids and flavonoids, and a growing number of triterpenoids are being discovered (Li *et al.*, 2006; Young *et al.*, 2012; Zhou *et al.*, 2012). The natural compounds found in *I. cornuta* have been considered to be antibacterial, hypotensive, anti-obesity, cardioprotective and immunosuppressive due to which the plant's potential for commercial use has also been noted (Nishimura *et al.*, 1999; Lin *et al.*, 2006; Li *et al.*, 2006). The terpene skeleton is synthesized by the mevalonate (MVA) pathway from acetyl-CoA to isopentenyl pyrophosphate (IPP). This process involves the enzymes acetoacetyl-CoA thiolase (AACT), 3-hydroxy-3-methylglutaryl-CoA synthase (HMGS), 3-hydroxy-3-methylglutaryl-CoA reductase, mevalonate kinase (MVK), phosphomevalonate kinase (PMVK) and mevalonate diphosphate decarboxylase (MVD). Then, the pathway is continued by farnesyl pyrophosphate, squalene synthase, and squalene epoxidase (SE) toward sterol and triterpenoid biosynthesis. The various triterpene skeletons are determined by SE (Abe *et al.*, 1993). However, the key genes involved in this secondary metabolite biosynthesis in *I. cornuta* have not been investigated. To accelerate further study on triterpenoid biosynthesis in *I. cornuta*, we performed the transcriptome sequencing of three organs by using an Illumina platform.

In this study, an RNA-Seq project for *I. cornuta* was performed by sequencing the roots, leaves and fruit using Illumina deep sequencing. Our study is the first to explore the *I. cornuta* transcriptome. The dataset of transcriptome sequencing from *I. cornuta* may help accelerate future molecular studies on the triterpenoid biosynthesis.

Materials and Methods

Plant Materials and RNA Extraction

20-year-old *I. cornuta* from one population in Yangtze University (Jingzhou, Hubei Province, China) was used in

To cite this paper: Zeng, H., L. Zhu, L.Q. Ma, W.W. Zhang, F. Xu and Y.L. Liao, 2019. *De novo* transcriptome sequencing of *Ilex cornuta* and analysis of genes involved in triterpenoid biosynthesis. *Intl. J. Agric. Biol.*, 22: 793–800

this study. After the collection of roots, leaves and fruit of plants, these parts were immediately frozen in liquid nitrogen and stored in a refrigerator at −80°C before further analysis. Total RNA was extracted by the Plant RNA Extraction Kit (TaKaRa, MiniBEST). The OD260/280 ratio and agarose gel electrophoresis were used to assess RNA quality.

Library Construction and RNA-Seq

Sequencing library construction and RNA-Seq were conducted by Annoroad Gene Technology Co., Ltd., Beijing, China. Sequence libraries of roots, leaves, and fruit were created using the NEBNext®Ultra™RNA Library Prep Kit for Illumina® (#E7530L, NEB, USA) in accordance with the instructions, and index codes were added to generate sequences to each sample. The libraries were sequenced on an Illumina Hiseq 4000 platform, and 150 bp paired-end reads were produced.

***De novo* Assembly, Unigene Annotation and Functional Classification**

The low-quality raw reads from mRNA-Seq having a Phred Quality Score of less than 19; the polluted adapter and with ambiguous "N" bases of more than 5% were deleted. Remaining reads were then identified as clean reads which were further assembled into contigs by using Trinity software (V20140717) (Grabherr *et al.*, 2011). Candidate coding regions within transcript sequences were identified by TransDecoder (V20140717). The functional annotation of unigenes and ORFs were performed by Trinotate (V20140717). These ORFs were functionally annotated by seven major databases including homology searches (BLAST+/SwissProt), protein domain identification (HMMER/PFAM), protein signal prediction (SingalP/TmHMM), and comparison with annotation databases (Uniprot/eggNOG/GO/KEGG).

Gene ontology (GO; http://www.geneontology.org/) terms of biological processes, molecular functions and cellular components were further used to annotate the assembled sequences. Furthermore, GO annotations were assigned using Blast2GO (Conesa *et al.*, 2005).

The Cluster of Orthologous Groups (COG) method proposed by Tatusov *et al*. (2000) was applied to predict the gene function. We filtered the results using the BLAST software, and removed the alignment sequence with an E-value of more than 1e-5. Afterward, the sequences were repeatedly aligned in COG database and then annotated the sequence.

The Kyoto Encyclopedia of Genes and Genomes (KEGG) database was then used to determine the functions, significant enrichment metabolic pathways and signal transduction pathways of differential expression genes of the biological system. Pathways with P-value < 0.05 were considered to be an apparently enriched entry. With q < 0.05 as the standard, each pathway in KEGG was subjected to super-geometric test to determine the significant enrichment pathways of the differentially expressed genes.

Differentially Expressed Unigene Analysis

Differential gene expression analysis was performed for the transcriptome data from three organs by DEG Seq package for read counts (Wang *et al.*, 2010). Clean reads were mapped back onto the assembled transcriptome and the sequence counts of each gene were calculated from the mapping results by HTSeq v0.6.0, and RPKM (Reads Per Kilobase Millon Mapped Reads) was then calculated to assess the expression level of genes in each organ (Mortazavi *et al.*, 2008). Further, to minimize the false positive rate, we used the threshold value of $q \leqslant 0.05$ and $|\log2\_ratio| \geqslant 1$ to identify the differentially expressed genes (DEGs).

Quantitative Real-Time PCR (qRT-PCR) Analysis

qRT-PCR analysis was performed on selected nine unigenes (Table 6). The cDNA was synthesized with PrimeScript™ RT reagent Kit with gDNA Eraser (TaKaRa, Dalian, China) from total RNA, which was obtained from fruit, roots and leaves. And then qRT-PCR was completed using the SYBR Premix Ex Taq™ II Kit (TaKaRa, Dalian, China). The qRT-PCR amplifications were performed in quantitative Real-Time PCR system (LineGene 9600 plus) with a total volume of 20 μL containing 2 μL of diluted cDNA, 0.4 μL of each primer, 10 μL of 2×SYBR Green Mix, and 7.2 μL of ddH_2O. The qRT-PCR amplification parameters were as follows: 5 min at 95°C; 40 cycles of 10 s at 95°C, 30 s at 55°C; 15 s at 95°C, 1 min at 60°C, 15 s at 95°C. Each sample was repeated in three replicates and the analysis was based on the Ct values of the PCR results. The $2^{-\Delta\Delta Ct}$ method was used to normalize the relative expression with glyceraldehyde-3-phosphate dehydrogenase (GAPDH) (Livak and Schmittgen, 2001).

Determination of Ursolic Acid (UA) and Oleanolic Acid (OA)

A 0.5 g dry sample per organ were dried and grounded into powder, then sonicated for 30 min with 25 mL methanol, filtered and volatilized to dryness. The residue was redissolved in 1 mL methanol and filtered through a 0.45 μm micro-filter. Quantifications were performed on an HPLC system (Thermo Scientific Dionex UltiMate 3000 BioRS, USA) equipped with an ODS C_{18} column (250 × 04.6, 5 μm). OA and UA standards were purchased from Jiangxi Baicaoyuan Biotechology Co. Ltd. (Jiangxi, China). Methanol and acetic acid were purchased from Aladdin Industrial Corporation (Shanghai, China). All the reagents used in the experiment were graded by HPLC. The mobile

phase consisted of 1% AcOH/H_2O (A) and 1% AcOH/MeOH (B) with a ratio of 90:10 (A: B, v/v). The OA and UA were detected at 292 nm with an eluent flow rate of 1.0 mL/min.

Results

RNA-Seq and *De novo* Assembly

Total RNA of *I. cornuta* was extracted from the roots, fruit, and leaves for cDNA library construction. Reads have been deposited in the National Center of Biotechnology Information (NCBI) SRA under BioProject ID PRJNA399054. Transcriptome sequencing of the *I. cornuta* from roots (SRR5947252), fruit (SRR5947253), and leaves (SRR5947245) generated 52,148,906, 54,525,622, and 51,027,138 raw reads, which resulted in 46,512,616, 47,652,144, and 44,978,224 clean reads, respectively (Table 1). After filtering, Q30 reached >96.49% and GC content were approximately 40% in each of the sequenced transcriptomes.

The clean reads obtained from the sequenced transcriptome of three organs were assembled into 176,044 transcripts using Trinity software (V20140717) with an average length of 809.0 bp. Approximately 83.75% of clean reads were mapped to assembled transcriptome. Furthermore, 124003 unigenes were assembled from the clean reads with N50 of 1042 nt (Table 2). In current study, the unigenes of *I. cornuta* exhibited the same length distribution (Fig. 1) as for previous Illumina and transcriptome studies (Li *et al.*, 2014), indicating the significant high quality of unigenes.

Additionally, for a more intuitive understanding of the annotation results, the genes that were annotated to Uniprot, NR, and NT were counted and used to generate Venn diagrams (Fig. 2). The unigenes were then compared with the NCBI non-redundant nucleotide (Nt) and nonredundant protein (Nr) databases using the BLASTN and BLASTX algorithms with an E-value of $<10^{-5}$. These analyses indicated that 45,464 (94.9%) unigenes revealed homologous proteins in the Nr protein database, and 22,760 (47.5%) attained significant matches in the Nt database. Interestingly, 13,285 (27.7%) unigenes matched to both databases.

The Nr homologous species distribution of *I. cornuta* is listed in Fig. 3. Of 49,603 unigenes, 7,509 (15.1%) unigenes had high sequence similarity to *Vitis vinifera*, and 3,061 (6.2%) and 2,426 (4.9%) unigenes matched to *Coffea canephora* and *Sesamum indicum*, respectively. In addition, 4,167 (8.4%) unigenes did not match to any species probably because the sequence is not conserved or has not been identified. The cause of this phenomenon may be that these species are rich in secondary metabolites and have completed genome sequencing, with a large proportion of known functional memory sequences in each database, such as *Vitis vinifera* (Jaillon *et al.*, 2007), *C. canephora* (Denoeud *et al.*, 2014), and *S. indicum* (Wang *et al.*, 2014).

Table 1: Quality parameters of Illumina transcriptome sequencing of *I. cornuta*

Sample	Read length	Raw reads	Clead reads	Clean Q30 (%)
IcLeaf	150+150 bp	51,027,138	44,978,224	96.49
IcFruit	150+150 bp	54,525,622	47,652,144	96.40
IcRoot	150+150 bp	52,148,906	46,512,616	96.41

IcLeaf, IcFruit and IcRoot indicate leaves, fruit and roots from *Ilex cornuta*, respectively

Table 2: Length distribution of transcripts and unigenes for transcriptome assembly of *I. cornuta*

Category	Min length (bp)	Max length (bp)	Mean length (bp)	N50 (nt)	N90 (nt)	Count	Pecent GC (%)
Transcript	201	14550	809.00	1368	311	176044	40.43
Unigene	201	14550	661.26	1042	265	124003	40.07

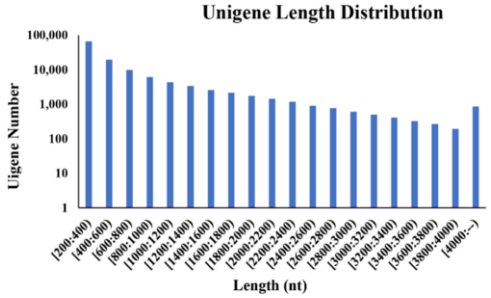

Fig. 1: Length distribution of the assembled unigenes of *I. cornuta*

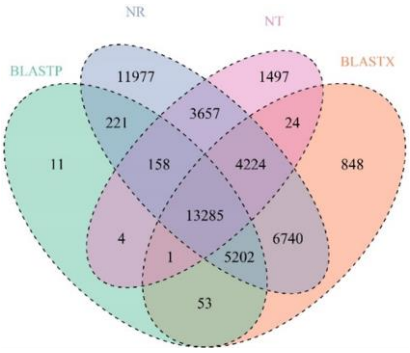

Fig. 2: Venn diagram of functional annotation of predicted open reading frame (ORF) of *I. cornuta*. The annotations were obtained by comparing the predicted ORFs with sequences from Nr, Nt and UniProt of public databases

Gene Ontology Classification

GO analysis revealed that the 29,165 (94.96%) assessed genes were mapped to 67 functional groups among, which 24291 (79.10%), 25332 (82.48%) and 25613 (83.40%) were derived from the biological process, cellular component and molecular function. Results showed (Fig. 4) that the major groups in the biological process were cellular, metabolic, and single-organism processes. In the cellular component group, many sequences were classified to cell part and organelle. In the molecular function group, many sequences were annotated to binding and catalytic categories.

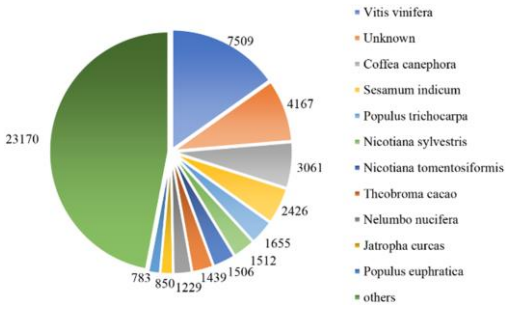

Fig. 3: Species distribution of the BLAST hits in Nr database

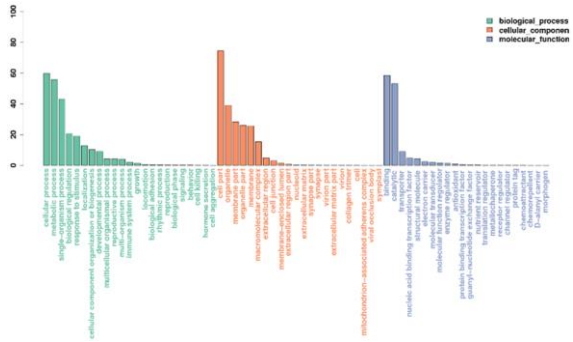

Fig. 4: Functional annotation of assembled sequenced of *I. cornnta* based on gene ontology (GO) categorization. The results are summarized into three main categories: biological process, cellular component and molecular function. The x-axis indicates specific GO category. The y-axis indicates the percentage of sequences in category

COG Classification

A total of 124,003 unigene sequences were searched against the COG database to predict and classify functions (Fig. 5). Only 12,386 of the unigene sequences were bunched into 24 groups. Among these R was the largest group (1,966; 15.87%) which stand for general function followed by group J (1,475; 11.91%) for translation, ribosomal structure, and biogenesis and group O (1,212; 9.79%) for the post-translational modification, protein turnover, and chaperones. Whereas, group I belonging to the lipid transport and metabolism was the smallest group (574; 4.63%). In addition, fewer unigenes were assigned to secondary metabolite biosynthesis/transport and catabolism, coenzyme transport and metabolism, cell wall/membrane/envelope biogenesis, cell cycle control/cell division/chromosome portioning, and cytoskeleton, respectively). Cell motility and nuclear structure were annotated for few unigenes (17 and 2 unigenes). Meanwhile, no one was annotated to extracellular structure. A total of 243 unigenes were allocated to nucleotide transport and metabolism, 159 unigenes to RNA processing and modification, 157 unigenes to defense mechanisms, and 132 unigenes to chromatin structure and dynamics.

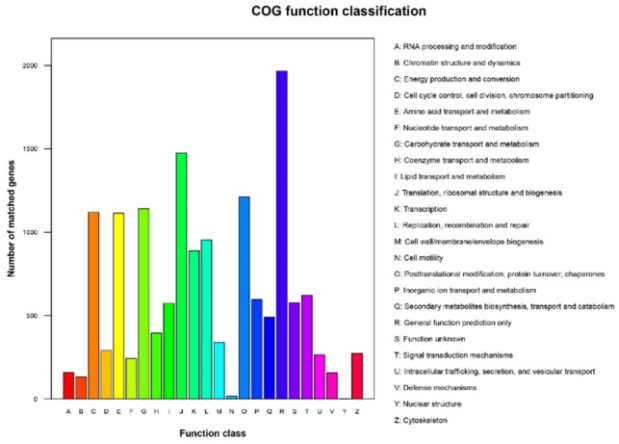

Fig. 5: Classification of the *I. cornuta* transcriptome based on the clusters of orthologous groups (COG)

KEGG Classification

The KEGG Pathway Tools can be used to classify gene functions and thoroughly understand the biochemical pathways. Among the 45,486 differentially expressed genes in IcFruit_IcLeaf of *I. cornuta*, 7319 were associated with 299 pathways of the KEGG database. Moreover, 7858 and 7222 in IcFruit_IcRoot, and IcRoot_IcLeafwere associated with 293 and 291 pathways, respectively. The top 20 significantly KEGG pathways enriched differential gene expression are listed in Table 4. The most enriched differential gene expression pathway is the "Microbial metabolism in diverse environments" and "Metabolic pathway", followed by the "Ribosome" and "Biosynthesis of secondary metabolites".

Identification of Differentially Expressed Genes

To estimate the differential expression pattern of genes in the fruit, roots and leaves, the expression of genes was calculated in FPKM. Among the total 45,486 differentially expressed genes of fruit and leaves (Fig. 6), 19,829 were upregulated and 25,657 were downregulated (Table 3). Between the fruit and roots, we found 49,132 differentially expressed genes, of which 21,424 were upregulated and 27,708 were downregulated. Among the 53,438 differentially expressed genes between the roots and leaves, 27,528 were up-regulated and 25,910 were downregulated. A Total of 8966 genes were differentially expressed among these three organs.

Putative Triterpenoid Biosynthesis Genes in different Organs

In current study, several triterpenoids biosynthesis genes were identified during the functional annotation and KEGG analysis of unigenes among which genes related to the MVA

Table 3: Summary of differential gene expression (DGE) in fruit, root and leaf of *I. cornuta*

Differential comparisons	IcFruit_IcLeaf	IcFruit_IcRoot	IcRoot_IcLeaf
Number of up-regulated genes	19829	21424	27528
Number of down-regulated genes	25657	27708	25910
Number of DGEs	45486	49132	53438

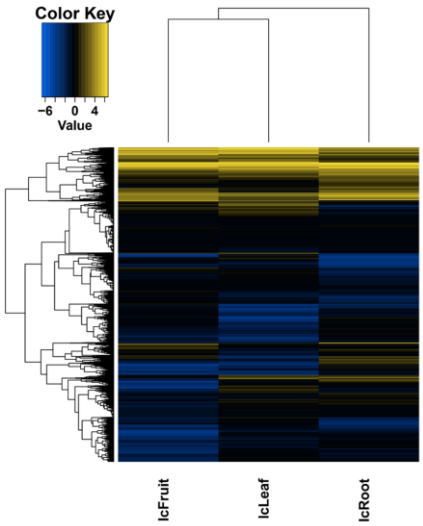

Fig. 6: Differentially expressed genes (DEGs) in different *I. cornuta* organs. Yellow colors indicate high expression, blue colors indicate low expression, and each horizontal bar represents a single gene

and MEP pathways in "Terpenoid backbone biosynthesis" and "Sesquiterpenoid and triterpenoid biosynthesis" were considered as candidate and discussed in detail. As shown in Fig. 7, some enzymes such as FPPS, DXR, CMK, and HDS were identified as single members; while other identified enzymes had more than one unigene from the transcriptome database of *I. cornuta*. For example, the four members of AACT, HMGS, HMGR, MVD have been found in MVA pathway, and five DXS, six HDR in MEP pathway. In addition, three SE were identified in the triterpenoid biosynthetic pathway (Fig. 7). These results are beneficial to the study of triterpenoid biosynthesis in *I. cornuta*.

Validation of Gene Expression by qRT-PCR

To confirm the accuracy of the calculated FPKM results, qRT-PCR analysis was performed on nine unigenes that were potentially involved in triterpenoid biosynthesis (Fig. 8). Results showed that the expression of the six genes (*HMGS*, *DXS*, *HDR*, *IPPI*, *SS*, *SE*) were well in accordance with the expression data obtained by RNA-Seq.

Ursolic acid (UA) and Oleanolic Acid (OA) Contents in Three Organs

As the expression levels of genes involved in triterpenoid biosynthesis were different in various organs of *I. cornuta*. Therefore, leaves, roots and fruit were analyzed for their oleanolic acid (OA) and ursolic acid (UA) content, which were observed significantly different in these organs (Table 5). The results showed that the OA contents were highest in roots (113.14 ± 0.17 μg/g), and the lowest in fruit (6.29 ± 0.77 μg/g). In addition, the UA had highest contents in leaves (64.56 ± 6.20 μg/g), and lowest contents in fruit (9.92 ± 1.20 μg/g). However, the highest total contents of OA and UA was observed in roots, followed by leaves and fruit.

Discussion

Transcriptome sequencing occupies an important position in the functional genomics because it can reflect the differences in the gene expression level of plants in different organs or at different times. This technology also identifies candidate genes associated with specific physiologic and metabolic functions, which build the foundation for the further study of the function, expression and regulation of genes (Zhang *et al.*, 2012; Wang *et al.*, 2015). This article reported the first public transcriptome sequencing of *I. cornuta*. Prior to this work, a low amount of sequence data of *I. cornuta* was available. In the present study, we used RNA-Seq technology to determine the *I. cornuta* transcriptome by the Illumina Hiseq 4000 platform. A Total of 139, 142, 984 clean reads were obtained and assembled into 124, 003 unigenes. Among these reads, 65.9% were successfully annotated with GO, COG and KEGG databases which indicated their relative conservation. The average length of unigene was similar to other plants on the same platform such as *Portunustritu berculatus* (Yang *et al.*, 2015), *Boehmeria nivea* L. (Chen *et al.*, 2014), and *Chlamys farreri* (Deng *et al.*, 2014). This observation indicates the reliability of current assembly which serves as a sequence source for future studies to identify the biochemical processes involved in secondary metabolism in *I. cornuta*.

The sesqui-, di-, and triterpenoids of terpene synthases were found in *I. cornuta*, where the number of diterpenes was more than five times the sum of the two other types of terpenes. At present, numerous terpenoid synthases have been found in plants with genomic sequences. Mono- and sesquiterpene synthases corresponding to dominant essential oils have been found in *Litsea cubeba* (Han *et al.*, 2013). Meanwhile, mono-, sesqui-, di-, and triterpenoids have been found in *Ilex asprella*, among which monoterpenoids covering the majority of the content. Such monoterpenoids exhibit anti-cancer and anti-virus activities (Zheng *et al.*, 2014; Wen *et al.*, 2017). A total of 69 putatively functional genes including those for sesqui-, mono-, and diterpenoids have been found in *Vitis vinifera*. These substances are important to grapevine biology and wine flavor and quality (Martin *et al.*, 2010). Meanwhile, 40 terpenoid synthase genes including those for sesqui-, mono-, and diterpenoids have been found in *Arabidopsis thaliana* (Aubourg *et al.*, 2002). Finally, 26 triterpenoid saponins genes have been isolated from *Eleutherococcus senticosus*, which is a

Table 4: The top twenty significantly enriched Kyoto encyclopedia of genes and genomes (KEGG) pathways

KEGG pathway	Pathway ID	*P*-Value	Number of differentially expressed genes
Microbial metabolism in diverse environments	ko01120	4.9E-17	363
Metabolic pathways	ko01100	2.41E-14	1120
Ribosome	ko03010	1.26E-12	331
Biosynthesis of secondary metabolites	ko01110	6.12E-12	585
Galactose metabolism	ko00052	2.56E-08	53
Glycolysis / Gluconeogenesis	ko00010	3.69E-08	124
Starch and sucrose metabolism	ko00500	0.000000495	141
Glyoxylate and dicarboxylate metabolism	ko00630	0.000000543	94
Degradation of aromatic compounds	ko01220	0.000000651	43
Chlorocyclohexane and chlorobenzene degradation	ko00361	0.00000131	19
Toluene degradation	ko00623	0.00000289	18
Fluorobenzoate degradation	ko00364	0.00000289	18
Phenylalanine metabolism	ko00360	0.00000278	60
Tyrosine metabolism	ko00350	0.00000284	55
Carbon fixation in photosynthetic organisms	ko00710	0.0000059	78
Pentose phosphate pathway	ko00030	0.00000653	71
Benzoate degradation	ko00362	0.0000111	26
Phenylpropanoid biosynthesis	ko00940	0.0000273	63
Pentose and glucuronate interconversions	ko00040	0.0000384	53
Photosynthesis	ko00195	0.0000582	40

Table 5: OA and UA content in three organs of *I. cortuna* (n = 3)

Compound	Leaf	Fruit	Root
OA (*μ*g/g DW)	25.20 ± 3.14b	6.29 ± 0.77c	113.14 ± 0.17a
UA (*μ*g/g DW)	64.56 ± 6.20a	9.92 ± 1.20c	26.64 ± 0.22b

Different letters within each row indicate significant differences (p <0.01)

Table 6: Primer sequences for Quantitative Real-Time PCR

Unigene ID	Unigene	Primer	Sequence (5'-3')
c89809_g1	*IcAACT*	IcAACT-RTS	ATGAAGCCTTTGCGGTCGTG
		IcAACT-RTA	GCTCTACAACAAGGGCTGAT
c88729_g1	*IcHMGS*	IcHMGS-RTS	GGCACGTACTATCTCACTGAAG
		IcHMGS-RTA	TTGTCACAAGTGGTGCTAGG
c75847_g1	*IcHMGR*	IcHMGR-RTS	GTGGGACTCAACTTGCATCT
		IcHMGR-RTA	CCAGCTATAATGGTGGCTAAGAG
c95459_g1	*IcMVD*	IcMVD-RTS	GGTACTAATTGCACGGAACAGA
		IcMVD-RTA	CACCGAGAAGGTAACTGTTGAA
c94033_g1	*IcDXS*	IcDXS-RTS	TTGCTAGTGCAGGGCTGA
		IcDXS-RTA	AAAGGAGGGCTTCACGAGT
c87271_g1	*IcHDR*	IcHDR-RTS	TCGAGAGTACACCAGTGACAT
		IcHDR-RTA	CCATACGCTTCTGCGAGTTT
c83362_g1	*IcIPPI*	IcIPPI-RTS	AAAGCTGATGCTGGTGAGG
		IcIPPI-RTA	CTCAACATTGTCCCACCACT
c93933_g1	*IcSS*	IcSS-RTS	ATAAGAACGATCCCAGCGCCACA
		IcSS-RTA	TGCCAGGACAATGAAGATTAGAACA
c89591_g1	*IcSE1*	IcSE1-RTS	TTAGACTGATTACGAGTGCATCTG
		IcSE1-RTA	CATGGCTTGCTTGAATATTTCTCC
c84824_g1	*IcGAPDH*	IcGAPDH- RTS	TATCAACGGCTTCGGTCGCA
		IcGAPDH- RTA	GGACGGAGTCGTACTTGAGCAT

common pharmacological active ingredient (Hwang *et al.*, 2015). Although the type and content of terpenoids vary widely among different species, these compounds considerably add economic and scientific value to humans.

In this present, the qRT-PCR was performed to further validate the expression patterns of the triterpenoid biosynthetic genes identified by RNA sequencing. The qRT-PCR results of *HMGS*, *DXS*, *HDR*, *IPPI*, *SS* and *SE* were well in accordance with the expression data obtained by RNA-Seq. These results prove that current transcriptome sequence assembly of *I. cornuta* is more accurate and useful. As can be seen that the expression levels of *AACT*, *HMGR*, *MVD*, *DXS*, *HDR*, *SS* and *SE* were significantly higher in roots than in other organs, and the HPLC results showed that the content of triterpenoid (Total OA and UA) was also highest in roots among the three organs. These results were in agreement with earlier findings that roots of *Ilex* species contained higher concentrations of total triterpenoid. For instance, Qiao *et al.* (2018) and Wang *et al.* (2012) found

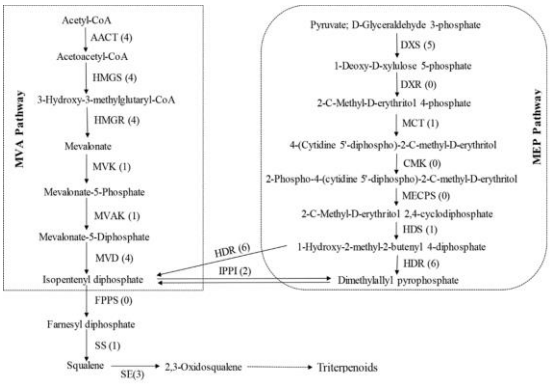

Fig. 7: Triterpenoids biosynthetic pathway in *I.cornuta* (Gahlan *et al*., 2012; Han *et al*., 2013; Zheng *et al.*, 2014). Each enzyme name is followed in parentheses by the Enzyme Code and the number of different unigenes. AACT: EC:2.3.1.9, acetoacetyl-CoA thiolase; HMGS: EC:2.3.1.10, 3-hydroxy-3-methylglutaryl-CoA synthase; HMGR: EC:1.1.1.34, 3-hydroxy-3-methylglutaryl- CoA reductase; MVK: EC:2.7.1.36, mevalonate kinase; MVAK: EC:2.7.4.2, phosphomevalonate kinase; MVD: EC:4.1.1.33, mevalonate diphosphate decarboxylase; FPPS: EC:2.5.1.10, farnesyl diphosphate synthase; SS: EC:2.5.1.21, squalene synthase; SE: EC:1.14.99.7, squalene epoxidase; DXS: EC:2.2.1.7, 1-deoxy-D-xylulose-5-phosphate synthase; DXR: EC:1.1.1.267, 1-deoxy-D-xylulose-5-phosphate reductoisomerase; MCT: EC:2.7.7.60, 2-C-methyl-D-erythritol 4-phosphate cytidylyltransferase; CMK: EC:2.7.1.148, 4-diphosphocytidyl-2-C-methyl-D-erythritol kinase; MECPS: EC:4.6.1.12, 2-C-methyl-D-erythritol 2,4-cyclodiphosphate synthase; HDS: EC:1.17.7.1, (E)-4-hydroxy-3-methylbut-2-enyl-diphosphate synthase; HDR: EC:1.17.7.4, 4-hydroxy-3-methylbut-2-enyl diphosphate reductase; IPPI: EC:5.3.3.2, isopentenyl-diphosphate isomerase

seven new triterpenoid saponins of *I. pubescens* and eight new triterpenoid glycosides *I. asprella* in the roots, respectively. These results indicated that total triterpenoid biosynthesis processes happen more intensively in the root than in other tissues of *I. cornuta*. However, the expression level of *HMGS* and *IPPI* genes are highest in fruit rather than in root, which is inconsistent with the tissue distribution of the total triterpenoid content. This may be because these two genes also play other different roles in *I. cornuta*.

At present, most studies on the biosynthesis of pentacyclic triterpenoid in Aquifoliaceae plants focus on roots or leaves such as in *I. asprella* (Wang *et al.*, 2012; Zhou *et al.*, 2012) and *I. kudincha* (Zuo *et al*., 2011). Screening key genes is the basis for understanding the metabolic pathway of triterpenoid. Similar to previous studies, the results of this study suggest that nine genes (*AACT*, *HMGR*, *MVD*, *HMGS*, *DXS*, *HDR*, *IPPI*, *SS* and *SE*) are involved in triterpenoid biosynthesis in *I. cornuta*. Among them, there are more reports on the *SS* and *SE* enzymes which are the key regulatory enzymes in the triterpenoid pathway (Ryder, 1991). This study showed that the expression levels of *SS* and *SE* were much higher in

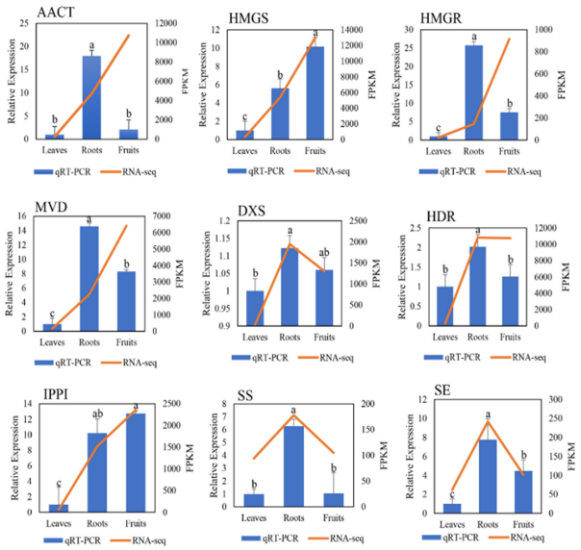

Fig. 8: qRT-PCR validation of the nine candidates unigenes involved in triterpenoid biosynthesis in the *I. cornuta*. Columns indicate relative expression obtained by qRT-PCR. Data represents means ± SD from three independent experiments. Means with different letters are significantly different at $P < 0.05$. Lines indicate relative expression calculated by FPKM method

roots than in leaves of *I. cornuta*, which displaying a similar pattern to that total triterpenoid content. In A*chyranthus bidentate* (Li and Hu, 2009) and *Dioscorea zingiberensis* (Cao *et al.*, 2004), the expression level of the SE gene was higher in roots than in other organs, which are consistent with our results.

This study demonstrates that organ-specific expression of these genes is associated with triterpenoid content in various organs of the *I. cornuta*. These data will enrich our understanding and provide a theoretical basis for further exploration of the molecular mechanism in *I. cornuta.*

Conclusion

In this study genes regulating the triterpenoids metabolism by transcriptome sequencing were identified in *I. cornuta*. Likewise, genes expression in different tissues was analyzed which will offer the new insights for isolation of the genes involved in triterpenoids metabolism. Moreover triterpenoid was the highest in roots indicating that the tissue-specific expression of genes is closely related to their function in regulating the synthesis of triterpenoid in *I. cornuta*.

Acknowledgements

This study was supported by the Doctor Foundation of Yangtze University (801190010127), the Natural Science Foundation of Hubei Province of China (2018CFB448), and the National Natural Science Foundation of China (31500546).

References

Abe, I., M. Rohmer and G.D. Prestwich, 1993. Enzymatic cyclization of squalene and oxidosqualene to sterols and triterpenes. *Chem. Rev*., 93: 2189–2206

Aubourg, S., A. Lecharny and J. Bohlmann, 2002. Genomic analysis of the terpenoid synthase (*AtTPS*) gene family of *Arabidopsis thaliana*. *Mol. Genet. Genomics*, 267: 730–745

Cao, Y.F., T.X. Wang and Z.H. Hu, 2004. Studies on the dynamic trends of diosgenin content in vegetative organs of *Dioscorea zingiberensis*. *Shi Yan Sheng Wu Xue Bao*, 37: 221–226

Chen, J., Z. Pei, L. Dai, B. Wang, L. Liu, X. An and D. Peng, 2014. Transcriptome profiling using pyrosequencing shows genes associated with bast fiber development in ramie. *BMC Genomics*, 15: 919

Conesa, A., S. Götz, J.M. García-Gómez, J. Terol, M. Talón and M. Robles, 2005. Blast2GO: A universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformation*, 21: 3674–3676

Deng, X.X., L.Q. Pen, J.J. Miao, Y.F. Cai and F.X. Hu, 2014. Digital gene expression analysis of reproductive toxicity of benzo[a]pyrene in male scallop chlamys farreri. *Ecotoxicol. Environ. Saf.*, 110: 190–196

Denoeud, F., L. Carreteropaulet, A. Dereeper, G. Droc, R. Guyot, M. Pietrella, C. Zheng, A. Alberti, F. Anthony, G. Apera, J.M. Aury, P. Bento, M. Bernard, S. Bocs, C. Campa, A. Cenci, M.C. Combes, D. Crouzillat, C.D. Silva, L. Daddiego, F.D. Bellis, S. Dussert, O. Garsmeur, T. Gayraud, V. Guignon, K. Jahn, V. Jamilloux, T. Joët, K. Labadie, T. Lan, J. Leclurcq, M. Lepelly, T. Leroy, L.T. Li, P. Libardo, P. Lopez, A. Muñoz, B. Noel, A. Pallavicini, G. Perrotta, V. Poncet, D. Pot, D. Priyono, M. Rigoreau, M. Rouard, J. Rozas, C. Tranchant-Dubreuil, R. VanBuren, Q. Zhang, A.C. Andrade, X. Argout, B. Bertrand, A.D. Kochko, G. Graziosi, R.J. Henry, M. Jayarama, R. Ming, C. Nagai, S. Rounsley, D. Sankoff, G. Giuliano, V.A. Albert, P. Wincker and P. Lashermes, 2014. The coffee genome provides insight into the convergent evolution of caffeine biosynthesis. *Science*, 345: 1181–1184

Gahlan, P., H.R. Singh, R. Shankar, N. Sharma, A. Kumari, V. Chawla, P.S. Ahuja and S. Kumar, 2012. *De novo* sequencing and characterization of *Picrorhiza kurrooa* transcriptome at two temperatures showed major transcriptome adjustments. *BMC Genomics*, 13: 126

Grabherr, M.G., B.J. Hass, M. Yassour, J.Z. Levin, D.A. Thompson, I. Amit, X. Adiconis, L. Fan, R. Raychowdhury, Q. Zeng, Z. Chen, E. Mauceli, N. Hacohen, A. Gnirke, N. Rhind, F.D. Palma, B.W. Birren, C. Nusbaum, K. Lindblad-Toh, N. Friedman and A. Regev, 2011. Full-Length Transcriptome assembly from RNA-Seq Data without a reference genome. *Nat. Biotechnol.*, 29: 644–652

Han, X.J., Y.D. Wang, Y.C. Chen, L.Y. Lin and Q.K. Wu, 2013. Transcriptome sequencing and expression analysis of terpenoid biosynthesis genes in *Litsea cubeba*. *PLoS One*., 8: e76890

Hwang, H.S., H. Lee and Y.E. Choi, 2015. Transcriptomic analysis of Siberian ginseng (*Eleutherococcus senticosus*) to discover genes involved in saponin biosynthesis. *BMC Genomics*, 16: 180

Jaillon, O., J.M. Aury, B. Noel, A. Policriti, C. Clepet, A. Casagrande, N. Choisne, S. Aubourg, N. Vitulo, C. Jubin, A. Vezzi, F. Legeai, P. Hugueney, C. Dasilva, D. Horner, E. Mica, D. Jublot, J. Poulain, C. Bruyère, A. Billault, B. Segurens, M. Gouyvenoux, E. Ugarte, F. Cattonaro, V. Anthouard, V. Vico, C.D. Fabbro, M. Alaux, G.D. Gaspero, V. Dumas, N. Felice, S. Paillard, I. Juman, M. Moroldo, S. Scalabrin, A. Canaguier, I.L. Clainche, G. Malacrida, E. Durand, G. Pesole, V. Laucou, P. Chatelet, D. Merdinoglu, M. Delledonne, M. Pezzotti, A. Lecharny, C. Scarpelli, F. Artiguenave, M.E. Pè, G. Valle, M. Morgante, M. Caboche, A.F. Adam-Blondon, J. Weissenbach, F. Quétier and P. Wincker, 2007. The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. *Nature*, 449: 463–467

Kuang, C.T., X.Z. Li, Y.L. Han and Y.J. Wu, 2011. Determination method and extraction process of total flavonoids from leaves of *Ilex Cornuta* L. *Adv. Mater. Res.*, 236: 216–219

Li, C.Y., J.T. Li, Y.Y. Kuang, R. Xu, Z.X. Zhao, G.Y. Hou and X.W. Sun, 2014. The transcriptomes of the crucian carp complex (*Carassius auratus*) provide insights into the distinction between unisexual triploids and sexual diploids. *Intl. J. Mol. Sci.*, 15: 9386–9406

Li, J. and Z. Hu, 2009. Accumulation and dynamic trends of triterpenoid saponin in vegetative organs of *Achyranthus bidentata*. *J. Integr. Plant Biol.*, 51: 122–129

Li, Y., T. Wu, Z.H. Chen and Z.T.Wang, 2006. New triterpene saponins and flavonol glycosides from the leaves of *Ilex cornuta*. *Chin. J. Chem.*, 24: 577–579

Lin, C., Y.B. Tan, J. Zhang, W.Z. Shen, Y.Z. Chen and Z.Y. Jiang, 2006. Effect of chemical extraction agents from *Ilex latifolia* Thunb on mouse T lymphocytes activation and proliferation *in vitro*. *J. Jinan Univ.*, 27: 199–203

Livak, K.J. and T.D. Schmittgen, 2001. Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta CT}$ Method. *Methods*, 25: 402–408

Martin, D.M., S. Aubourg, M.B. Schouwey, L. Daviet, M. Schalk, O. Toub, S.T. Lund and J. Bohlman, 2010. Functional annotation, genome organization and phylogeny of the grapevine (*Vitis vinifera*) terpene synthase gene family based on genome assembly, FLcDNA cloning, and enzyme assays. *BMC Plant Biol*., 10: 226

Miller, C.B., 1998. Holly days. *Horticulture*, 246: 759

Mortazavi, A., B.A. Williams, K. McCue, L. Schaeffer and B. Wold, 2008. Mapping and quantifying mammalian transcriptomes by RNA-seq. *Nat. Meth.*, 5: 621–628

Nishimura, K., T. Miyase and H. Noguchi, 1999. Triterpenoid saponins from *Ilex kudincha*. *J. Nat. Prod.*, 62: 1128–1133

Qiao, X., M. Ji, Y. Yao, L. Ma, J. Wu, G. Liao, H. Zhou, Z. Liu and P. Wu, 2018. Pubescenosides E-K, seven new triterpenoid saponins from the roots of *Ilex pubescens* and their anti-inflammatory activity. *Molecule*, 23: 1–13

Ryder, N.S., 1991. Squalene epoxidase as a target for the allylamines. *Biochem. Soc. Trans.*, 19: 774–777

Sibley, J.L. and J.C. Wilson, 2002. *Ilex cornuta* 'O.Spring'. *Amer. Nurs.*, 195: 86

Tatusov, R.L., E.V. Koonin and D.J. Lipman, 2000. A genomic perspective on protein families. *Nucl. Acids Res.*, 28: 33–36

Wang, K., S. Jiang, C. Sun, Y. Lin, R. Yin, Y. Wang and M. Zhang, 2015. The spatial and temporal transcriptomic landscapes of ginseng, *Panax ginseng* CA Meyer. *Sci. Rep.*, 5: 1–12

Wang, L., S. Yu, C. Tong, Y. Zhao, Y. Liu, C. Song, Y. Zhang, X. Zhang, Y. Wang, W. Hua, D. Li, D. Li, F. Li, J. Yu, C. Xu, X. Han, S. Huang, S. Tai, J. Wang, X. Xu, Y. Li, S. Liu, R.K. Varshney, J. Wang and X. Zhang, 2014. Genome sequencing of the high oil crop sesame provides insight into oil biosynthesis. *Genome Biol.*, 15: 39

Wang, L., Y. Cai, X.Q. Zhang, C.L. Fan, Q.W. Zhang, X.P. Lai and W.C. Ye, 2012. New triterpenoid glycosides from the roots of *Ilex asprella*. *Carbohydr. Res.*, 349: 39–43

Wang, L.K., Z.X. Feng, X. Wang, X.W. Wang and X.G. Zhang, 2010. DEGseq: an R package for identifying differentially expressed genes from RNA-seq data. *Bioinformatics*, 26: 136–138

Wei, C.W., 1988. Antifertility effects of *Ilex cornura Lindl*. *Zhong Yao Tong Bao*, 13: 48

Wen, L., X. Yun, X. Zheng, H. Xu, R. Zhan, W. Chen, Y. Xu, Y. Chen and J. Zhang, 2017. Transcriptomic comparison reveals candidate genes for triterpenoid biosynthesis in two closely related *Ilex* Species. *Front. Plant Sci.*, 8: 1–14

Yang, Y.X., J.T. Wang, T. Han, T. Liu, C.L. Wang, J. Xiao and H.L. Shi, 2015. Ovarian transcriptome analysis of *Portunus trituberculatus* provides insights into genes expressed during phase III and IV development. *PLoS One*, 10: e0138862

Young, L.S., K.H. Kyung and L.K. Ro, 2012. Four new triterpenes from *Ilex cornuta* Lindley. *Can. J. Chem.*, 91: 382–386

Zhang, J., S. Liang, J. Duan, J. Wang, S. Chen, Z. Cheng, Q. Zhang, X. Liang and Y. Li, 2012. *De novo* assembly and characterisation of the transcriptome during seed development, and generation of genic-SSR markers in Peanut (*Arachis hypogaea* L.). *BMC Genomics*, 13: 90

Zheng, X., H. Xu, X. Ma, R. Zhan and W. Chen, 2014. Triterpenoid saponin biosynthetic pathway profiling and candidate gene mining of the *Ilex asprella* root using RNA-Seq. *Intl. J. Mol. Sci.*, 15: 5970–5987

Zhou, S.X., Z.R. Yao, L.I. Jun and T.U. Peng-Fei, 2012. Flavonoids from the leaves of *Ilex cornuta*. *Chin. J. Nat. Med.*, 10: 84–87

Zuo, W.J., H.F. Dai, J. Chen, H.Q. Chen, Y.X. Zhao, W.L. Mei, X. Li and J.U. Wang, 2011. Triterpenes and triterpenoid saponins from the leaves of *Ilex kudincha*. *Planta Med.*, 77: 1835–1840

[Received 22 Mar 2019; Accepted 10 May 2019; Published (online) 20 Aug 2019]