RESEARCH ARTICLE

De novo leaf and root transcriptome analysis to identify putative genes involved in triterpenoid saponins biosynthesis in *Hedera helix* L.

Huapeng Sun[1], Fang Li[2], Zijian Xu[3], Mengli Sun[3], Hanqing Cong[1], Fei Qiao[1]*, Xiaohong Zhong[2]*

1 Key Laboratory of Crop Gene Resources and Germplasm Enhancement in Southern China, Ministry of Agriculture / Tropical Crops Genetic Resources Institute, Chinese Academy of Tropical Agricultural Sciences, Danzhou, China, 2 Horticulture & Landscape College, Hunan Agricultural University, Changsha, China, 3 Institute of Tropical Agriculture and Forestry, Hainan University, Haikou, China

* qiaofeixn@gmail.com (FQ); xh-zhong@163.com (X-hZ)

Abstract

Hedera helix L. is an important traditional medicinal plant in Europe. The main active components are triterpenoid saponins, but none of the potential enzymes involved in triterpenoid saponins biosynthesis have been discovered and annotated. Here is reported the first study of global transcriptome analyses using the Illumina HiSeq™ 2500 platform for *H. helix*. In total, over 24 million clean reads were produced and 96,333 unigenes were assembled, with an average length of 1385 nt; more than 79,085 unigenes had at least one significant match to an existing gene model. Differentially Expressed Gene analysis identified 6,222 and 7,012 unigenes which were expressed either higher or lower in leaf samples when compared with roots. After functional annotation and classification, two pathways and 410 unigenes related to triterpenoid saponins biosynthesis were discovered. The accuracy of these *de novo* sequences was validated by RT-qPCR analysis and a RACE clone. These data will enrich our knowledge of triterpenoid saponin biosynthesis and provide a theoretical foundation for molecular research on *H. helix*.

Introduction

Hedera helix L. is a perennially evergreen climbing vine that belongs to the Araliaceae family and is collected in the European Pharmacopoeia and British Pharmacopoeia as a traditional medicinal plant [1,2]. It occurs in Europe and is widely cultivated around the world. Functional components in *H. helix* with high medicinal value are extracted from fresh leaves and stems (European Pharmacopoeia 7.0, 2010). Clinical studies have revealed its pharmacological effect for treating cough [3], asthma [4], bronchitis [3,5,6] and other respiratory diseases. A recent pharmacology study also showed that the extracts have anti-inflammatory, antimicrobial, anti-oxidative, antitumour, anti-mutagenicity, antispasmodic, anti-leishmania and

hepatoprotective activities [1]. Due to its efficacy, it had been approved by many European countries for clinical applications [7]. The extracts of *H. helix* can be made into several medicinal dosage forms, such as syrups and tablets. In Poland, preparations with its extracts as the main components are commercially available, including Hedelix syrups, Prospan, PiniHelix, Helical, Hederasal and Hederoin tablets [1,8].

Using modern biotechnology to dissect biosynthesis pathways and manipulate the functional genes involved in biosynthesis of the active components in medicinal plants are two main development directions for the modernization of herbal medicines. The main active components in *H. helix* are hederacoside C [9,10] and α-hederin [11,12], which belong to the oleanane-type triterpenoid saponins [13]. Many studies describe the biosynthesis pathway of triterpenoid saponins in three parts [14,15]. Part I is upstream isopentenyl diphosphate (IPP) and dimethylallyl pyrophosphate (DMAPP) biosynthesis [16,17], part II is the midstream carbocyclic of triterpenoid biosynthesis [18], and part III is the downstream modification pathway of complex functional groups in triterpenoid carbocyclic [19]. Genes upstream and midstream of the pathway have been intensively studied [20,21], but the specific downstream genes involved in catalysing triterpenoid saponin biosynthesis have yet to be revealed. Since genomic information for *H. helix* is lacking and no genes in the triterpenoid saponin pathway in this plant have been reported, transcriptome analysis using next-generation sequencing technology provides a convenient and economical tool. It is also powerful for identifying various transcripts and functional genes and for providing quantitative estimates of gene expression. Next-generation sequencing technology can also generate forty to eighty million bp of high-quality sequences per sample. It has been widely used in medicinal plants such as *Panax ginseng* [20], *Panax notoginseng* [21], *Asparagus racemosus* [22], *Cephalotaxus hainanensis* [23], *Pseudostellaria heterophylla* [24], *Sinopodophyllum hexandrum* [25], *Solanum elaeagnifolium* [26].

In this study, we aimed to screen the functional genes involved in triterpenoid saponin biosynthesis in *H. helix* using *de novo* transcriptome sequencing. Since hederacoside C, hederacoside B, hederacoside D and α-hederin are rich in leaves and hardly detectable in roots (S1 Fig), the transcriptomes between leaves and roots were compared to dissect the molecular basis of this tissue-specific accumulation. Interestingly, 2 pathways and 410 unigenes related to triterpenoid saponins biosynthesis were discovered in the present study, which enriched the database of molecular information for *H. helix*.

Materials and methods

Plant materials

One-year-old *H. helix* were cultivated from scions, and scion woods were collected from the Hunan Research Institute of Vine Plant and the green house in National Center for Citrus Improvement, Hunan Agricultural University, Hunan Province, China on April 11, 2104. Leaf and root tissues were collected randomly from these 1-year-old plants for transcriptome analysis on May 15, 2015. Tissues were rinsed in water, cut into small pieces, frozen in liquid nitrogen immediately and stored at −80°C for further analyses.

RNA extraction, cDNA library construction and transcriptome sequencing

Leaf and root samples were harvested from five plants for RNA extraction and three biological replicates (L1 and R1, L2 and R2, L3 and R3) were performed. First, total RNA was extracted using RNA plant Plus Reagent (Tiangen, Beijing, China) according to the manufacturer's

instructions for polysaccharides & polyphenolics-rich samples. The total RNA purity was checked using a NanoDrop®2000 (Thermo, CA, USA) and the total RNA concentration and integrity were assessed using the RNA Nano 6000 Assay Kit of the Agilent Bioanalyzer 2100 system (Agilent Technologies, CA, USA).

The cDNA libraries were generated using the NEB Next® Ultra™ RNA Library Prep Kit for Illumina® (NE, USA) following the manufacturer's recommendations, and index codes were added to attribute sequences to each sample. Poly-(A) mRNA was isolated from total RNA using Oligo-(dT) magnetic beads and fragmented in fragmentation buffer. First-strand cDNA was synthesized using a random hexamer primer and M-MuLV Reverse Transcriptase (RNase H-). Second-strand cDNA was synthesized using buffer, dNTPs, RNaseH, and DNA polymerase I. The remaining overhangs were converted into blunt ends via exonuclease/polymerase activities. After adenylation of the 3' ends of DNA fragments, NEBNext Adaptor with a hairpin loop structure was ligated to prepare for hybridization. To select cDNA fragments preferentially 250–300 bp in length, the library fragments were purified with the AMPure XP system (Beckman Coulter, Beverly, USA). Then, 3 µL of USER Enzyme (NEB, USA) was used with size-selected, adaptor-ligated cDNA at 37˚C for 15 min. PCR was performed with Q5 Hot Start HiFi DNA polymerase, Universal PCR primers and Index (X) Primer. PCR products were purified (AMPure XP system) and library quality was assessed on an Agilent Bioanalyzer 2100 system.

Clustering of the index-coded samples was performed using a cBot Cluster Generation System using the TruSeq PE Cluster Kit v4-cBot-HS (Illumina) according to the manufacturer's instructions. After cluster generation, the library preparations were sequenced on an Illumina Hiseq 2500 platform, and paired-end reads were generated.

De novo assembly, unigene annotation and functional classification

Raw reads obtained from HiSeq-2500 sequencing were filtered to exclude reads containing adaptors, reads with more than 5% unknown nucleotides, and low-quality reads. At the same time, the Q20, Q30 and GC-content of the clean data were calculated. All downstream analyses were based on clean data of high quality. *De novo* assembly of the transcriptome was performed with Trinity [27]. After assembly of clean reads with Trinity and removal of redundant sequences using the TGI Clustering Tool (TGICL) [28], clusters (prefixed with CL) and singletons (prefixed with Unigene) were finally obtained.

All assembled unique gene sequences were aligned against the non-redundant (Nr) protein database (http://www.ncbi.nlm.nih.gov/), SwissProt (http://www.expasy.ch/sprot/), Kyoto Encyclopedia of Genes and Genomes (KEGG) [29] (http://www.genome.jp/kegg/) and Clusters of Orthologous Groups (COG) (http://www.ncbi.nlm.nih.gov/cog/) databases using BLASTx algorithms with a threshold of E<1.0E-5, and the protein functional annotation information was searched against the Nt database using BLASTn algorithms with a threshold of E-value<0.00001 [30]. Gene ontology (GO) (http://www.geneontology.org) terms functional annotation was performed based on the best hits from Nr annotation using Blast2go software (http://www.blast2go.de/), and WEGO (http://wego.genomics.org.cn/cgi-bin/wego/index.pl) software was used for further resultant GO ID and classification [31,32]. All assembled unigenes were searched using the Gene ID listed in S1 Table.

Differentially expressed unigene analysis

The expression levels of unigenes were established by in-house Perl scripts for each leaf and root sample; clean data were mapped back onto the assembled transcriptome and read-counts for each gene were obtained from the mapping results. To calculate the gene expression levels,

the Fragments Per kb per Million fragments (FPKM) method was used [33]. The expression difference was identified by Perl scripts, in which Fisher's exact test and the likelihood ratio were proposed to identify differentially expressed genes, and the P-value and false discovery rate (FDR) for each gene were calculated. Differentially expressed genes were required to have thresholds of "log2 ratio≥1" and "FDR<0.001" [34]. Next, GO and KEGG analysis were again performed on the DEGs.

Quantitative real-time PCR analysis and RACE clone

The total RNA for RT-qPCR analysis was extracted using a RNAprep Pure Kit (Tiangen, Beijing, China), and 1.0 μg RNA was used for reverse transcription with a Fast Quant RT Kit (Tiangen, Beijing, China) in a 20-mL reaction volume according to the manufacturer's instructions. Gene-specific primer pairs (S2 and S3 Tables) were designed with Beacon Designer 8 software based on transcriptome-assembled data and synthesized by a commercial supplier (Sangon, Shanghai, China). The F-box gene was used as an internal control [35]. RT-qPCR was performed in 96-well plates in a Bio-Rad CFX96 real-time PCR system (Bio-Rad, CA, USA) with a SYBR Green-based PCR assay. The final volume for each reaction was 20 mL with the following components: 2 mL diluted cDNA template (1 mg/mL), 10 mL SYBR Green Mix (Bio-Rad, CA, USA), 2.5 mL forward primer (2.5 mM), 2.5 mL reverse primer (2.5 mM) and 3 mL ddH$_2$O. The reaction was conducted under the following conditions: 95˚C for 3 min, followed by 40 cycles of denaturation at 95˚C for 10 s and annealing/extension at 56˚C for 30 s. The melting curve was obtained by heating the amplicon from 65˚C to 95˚C at increments of 0.5˚C per 5 s. Each RT-qPCR analysis was performed with three biological replicates. The relative quantification of gene expression was computed using the $2^{-\Delta\Delta Ct}$ method. The full-length cDNAs of the two genes were cloned by RACE with a SMARTer RACE 5'/3' Amplification kit (Clontech, CA, USA) according to the user manual. The RACE fragments were amplified by specific primers, ligated into the pRACE vector (Clontech, CA, USA) and sequenced. Then, the RACE fragments were used for the subsequent verification and analysis of full-length cDNAs.

Results

Reads generation and De novo assembly

Six cDNA libraries prepared from three leaves and three roots tissue of *H. helix* were sequenced using the Illumina HiSeq™2500 platform. A total of 138,709,902 and 139,062,838 raw reads were generated, and after removing adapters and filtering the low-quality sequences, 123,162,610 and 122,589,364 clean reads were generated for the leaf (L1, L2, L3) and root (R1, R2, R3) libraries (Table 1). Since there is no reference genome sequence in *H. helix*, all clean reads from these six libraries were *de novo* assembled into contigs using the Trinity software, and reads were mapped back to contigs, redundancy was removed and the reads were assembled further using TGICL [28]. Finally, 96,333 unigenes were obtained with a mean length of 1385 nucleotides (nt) (N50 1927 nt). A detailed summary of the sequencing and assembly results is shown in Table 2 and the length distribution of all unigenes is shown in Fig 1.

Functional annotation and classification

All assembled unigenes were first searched against the Nr protein, Swissprot protein, KEGG, GO and COG databases using the Blastx program with E-value threshold of 1E-5. Of the 96,333 unique sequences, 79,085 unigenes (82.1%) had at least one significant match to an existing gene model. The detailed results are shown in Table 3 and S1 Table. Based on the Nr database,

Table 1. Summary of data output quality of various libraries.

Library	Raw Reads	Clean Reads	Error (%)	Q20 (%)	Q30 (%)	GC (%)
L1	45,630,708	39,909,334	0.1	97.08	95.40	45.05
L2	45,480,900	40,761,120	0.1	97.14	95.71	44.72
L3	47,598,294	42,492,156	0.1	97.10	95.46	44.89
R1	46,941,114	40,930,938	0.1	96.65	95.76	44.73
R2	46,855,462	41,306,100	0.1	97.16	95.52	44.54
R3	45,266,262	40,352,326	0.1	97.10	95.68	44.58

https://doi.org/10.1371/journal.pone.0182243.t001

the E-value distribution showed that approximately 54.5% of the mapped unigenes ranged from 1E-5 to 1E-100 (Fig 2A). As shown in the similarity distribution, approximately 24.9% unigene sequences have a similarity above 80%, while 67.2% of the hits have a similarity from 40% to 80% (Fig 2B). Furthermore, 31.9% of the *H. helix* unigenes sequences are homologous to genes of *Vitis vinifera*; the others are 8.3% to *Theobroma cacao*, 7.8% to *Solanum tuberosum*, 5.1% to *Lycopersicon esculentum*, 5.0% to *Amygdalus persica*, 4.7% to *Populus balsamifera* subsp. trichocarpa and 4.3% to *Mimulus guttatus*, respectively (Fig 2C).

GO analysis is an international standard system of gene function classification, that has three main ontologies to describe molecular functions, cellular components and biological processes. Based on the Nr annotation, we used GO analysis to classify functions and understand the general distribution of the unigenes of *H. helix*. Among the 75,773 annotated unigenes, 50,479 sequences were categorized into 56 GO functional groups, as shown in Fig 3 and S4 Table. The biological process category has the highest number of unigenes among the three categories, followed by the cellular components category and molecular function category. From the overall analysis, there are 15 items clustering more than 5,047 unigenes. Within the biological processes category, "cellular process" and "metabolic process" have the highest number of unigenes. Meanwhile, "cell", "cell part" and "organelle" were enriched in the cellular components category. Genes encoding "binding" proteins and proteins related to "catalytic activity" were the largest proportion in the molecular function category.

Using COG classification to further evaluate the completeness and effectiveness of the *H. helix* annotation, a total of 32,443 annotated sequences were assigned to 25 COG categories. As shown in Fig 4 and S5 Table, the cluster of "General function prediction" represents the

Table 2. Summary of assembly results of *H. helix*.

Samples	Total Number	Total Length (nt)	Mean Length (nt)	N50 (nt)	Distinct Clusters	Distinct Singletons
L1-Contig	116,550	99,922,035	857	1339	-	-
L2-Contig	117,656	101,212,619	860	1341	-	-
L3-Contig	114,891	97,770,836	851	1323	-	-
R1-Contig	136,187	111,354,934	818	1271	-	-
R2-Contig	140,439	115,262,101	821	1280	-	-
R3-Contig	128,613	104,187,419	810	1267	-	-
L1-Unigene	71,734	69,222,524	965	1511	34,970	36,764
L2-Unigene	70,985	69,495,387	979	1526	35,544	35,441
L3-Unigene	68,637	66,692,896	972	1514	34,305	34,332
R1-Unigene	81,825	75,839,021	927	1462	38,521	43,304
R2-Unigene	85,302	79,392,368	931	1474	39,032	46,270
R3-Unigene	78,940	71,597,869	907	1440	35,493	43,447
All-Unigene	96,333	133,417,819	1385	1927	55,721	40,612

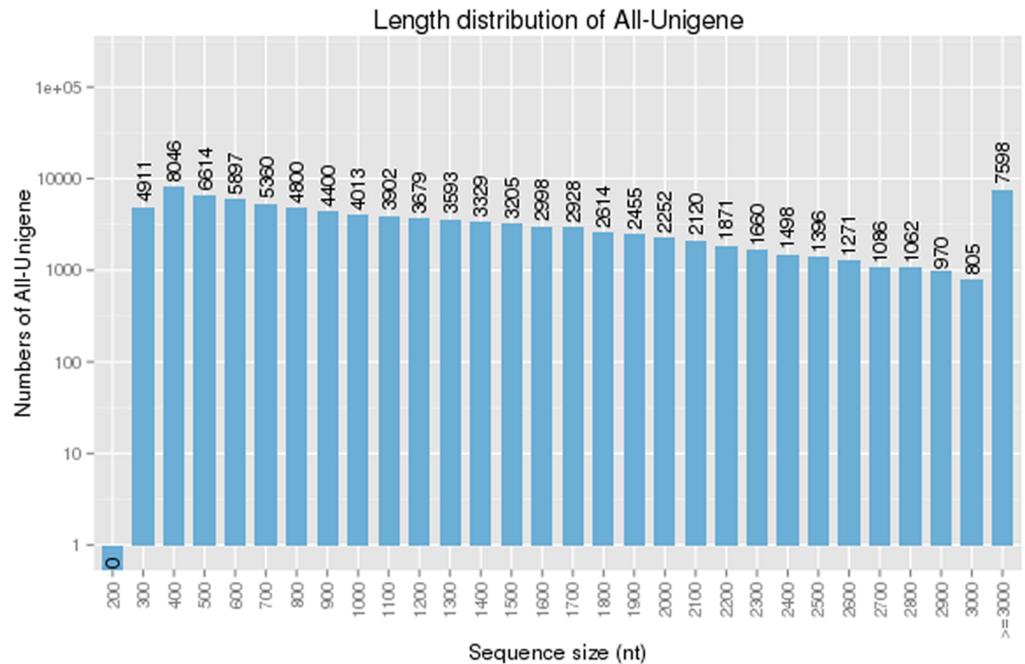https://doi.org/10.1371/journal.pone.0182243.t002

Fig 1. Length distribution frequency of unigenes in *H. helix*.

largest group (11,547, 35.6%), followed by "Transcription" (6,459, 19.9%), "Replication, recombination and repair" (5,560, 17.1%), "Signal transduction mechanisms" (4,791 14.8%) and "Posttranslational modification, protein turnover, chaperones" (4,643 14.3%). The smallest group was "Nuclear structure", with only 8 sequences.

For KEGG analysis, a total of 96,333 annotated unigenes were mapped to identify active pathways in *H. helix* with a cut-off of E-value<0.00001. Of these, 47,100 with significant matches from the database were assigned to 128 KEGG pathways (S6 Table). The largest category was metabolic pathways (10,484, 22.3%), followed by biosynthesis of secondary metabolites (4,795, 10.2%), plant-pathogen interactions (2,805, 6.0%), plant hormone signal transduction (2,623, 5.6%) and spliceosome (2,074, 4.4%). These annotations provide a further understanding of the transcriptome data and their functions and pathways in *H. helix*.

Differentially Expressed Gene analysis

The Differentially Expressed Genes (DEG) of the six transcriptome libraries were used to discover the unigenes with significant differences in expression. In this study, the expression of

Table 3. Summary of functional annotations of *H. helix* unigenes.

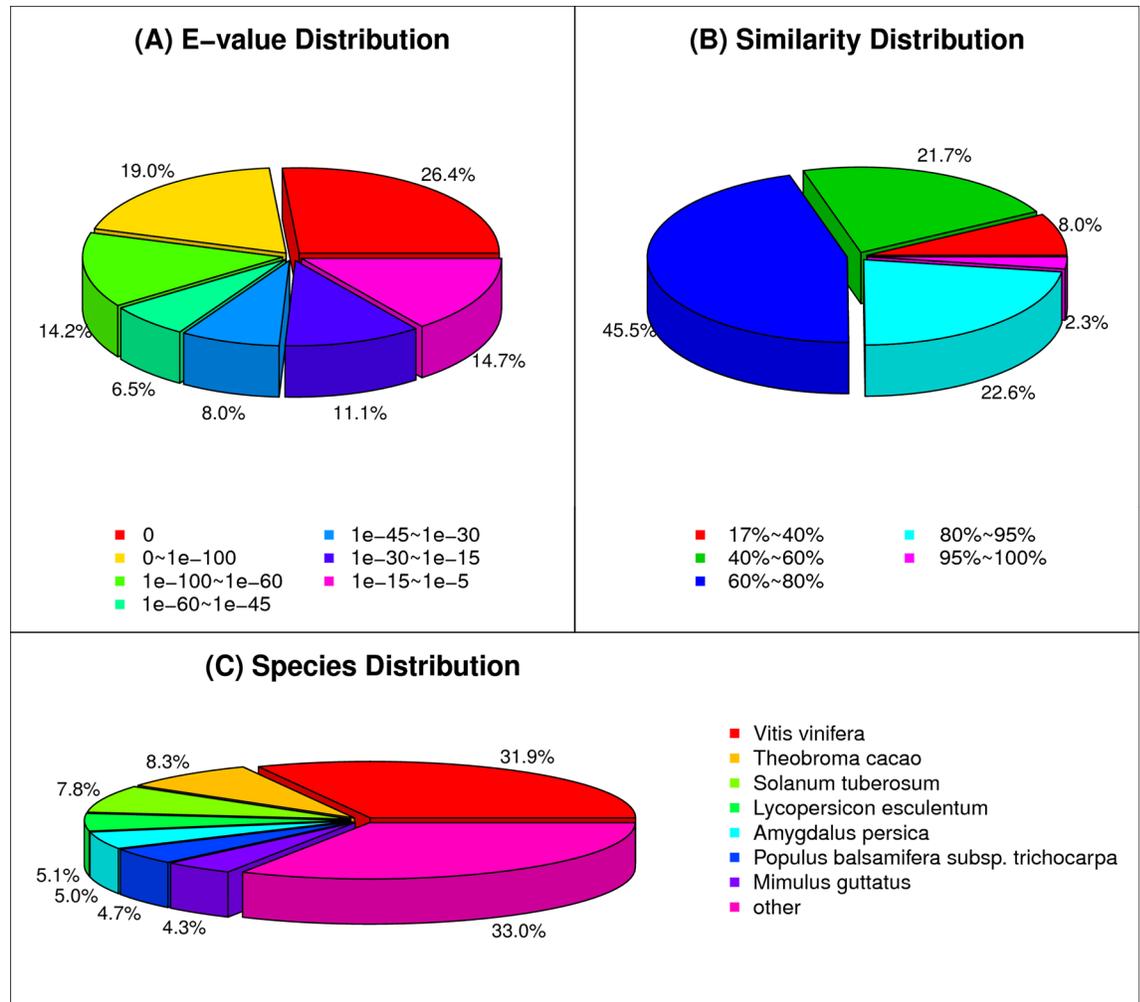| Public database	Number of Unigenes	Percentage (%)
Annotated in Nr	75,773	78.7
Annotated in Nt	70,728	73.4
Annotated in SwissProt	51,320	53.3
Annotated in KEGG	47,100	48.9
Annotated in COG	32,443	33.7
Annotated in GO	50,479	52.4
All annotated Unigenes	79,085	82.1
Total Unigenes	96,333	100

Fig 2. Gene similarity of unigenes against the Nr database. (A) E-value distribution of top BLAST hits for each unigene (E-value of 1.0E-5). (B) Similarity distribution of best BLAST hits for each unigene. (C) Distribution of BLAST results by species shown as percentage of total homologous sequences (E-value ≤1.0E-5). All plant proteins in the NCBI Nr database were used for homology search and the best hit of each sequence was used for analysis.

unigenes was calculated by FPKM. The different analysis methods were as follows: R1 and L1, R2 and L2, R3 and L3, were compared, and the DEGs found were used to comment on all three replicates for GO classification and KEGG pathway analysis. In total, 6,222 unigenes were more expressed higher and 7,012 unigenes were expressed lower in leaf samples than root samples (S7 and S8 Tables). According to the expression differences between the roots (R) and Leaves (L) shown in Fig 5, GO and KEGG analysis were performed again based on the Nr annotation. Of the 13,234 DEGs, 8,771 were assigned to 52 GO categories, and of these unigenes, 4,243 of the 6,222 expressed higher unigenes and 4,528 of the 7,012 down-expressed lower unigenes were assigned to at least one of the GO terms (biological process, cellular component, or molecular function). After KEGG analysis was performed again, 3,657 of the 6,222 expressed higher unigenes were annotated into 123 pathways and 4,324 of the 7,012 expressed lower unigenes were annotated into 122 pathways. Meanwhile, the top 20 statistically concentrated pathways were analysed based on the false discovery rate (FDR)≤0.001 and these results
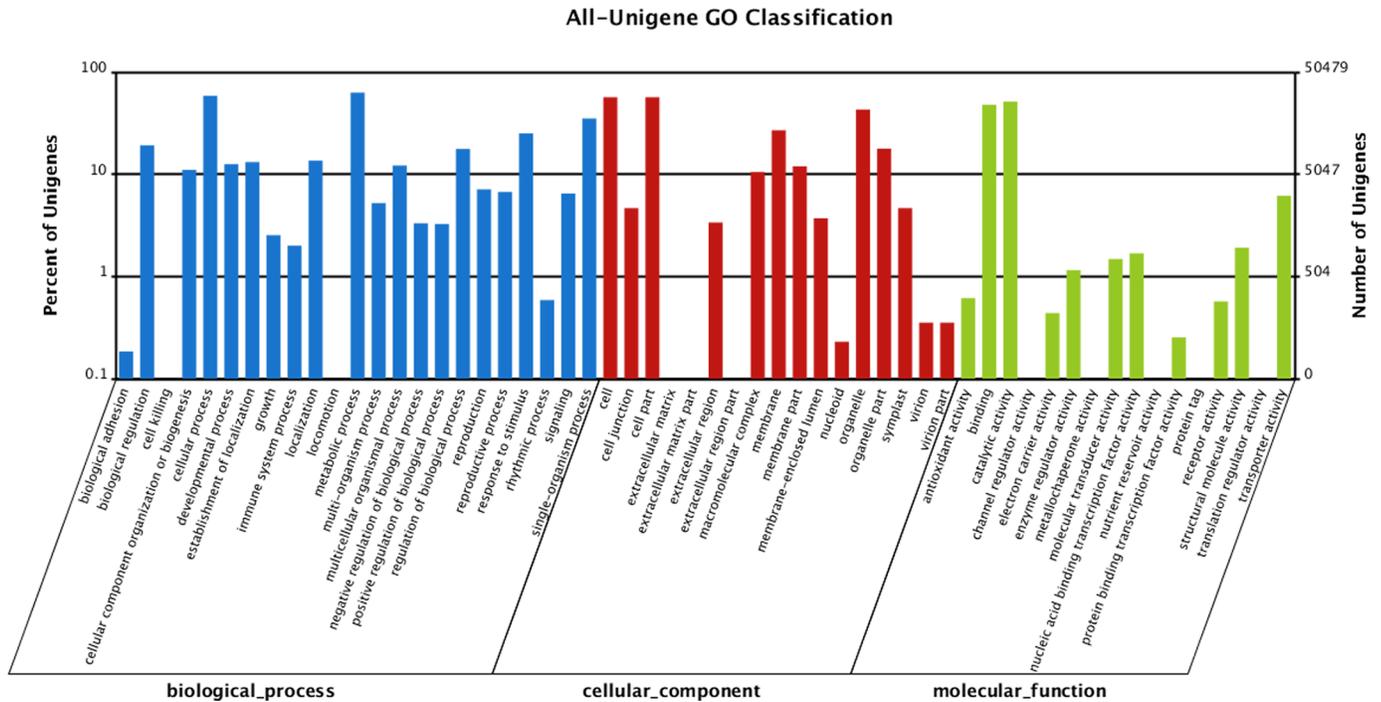
Fig 3. Comparison of Gene ontology (GO) classifications of *H. helix*. Results are summarized into three main GO categories (biological process, cellular component, molecular function) and 44 sub-categories. The x-axis indicates subcategories; right y-axis indicates number of genes in a category; and left y-axis indicates percentage of a specific category of genes in the main category.

can identify the main biochemical metabolic pathways and signal transduction pathways that the DEGs attended, as seen in S2 Fig.

Identification of triterpene saponin biosynthesis genes

In addition to the KEGG analysis and unigenes functional annotation results, several triterpene saponin biosynthesis genes were discovered and identified. In the KEGG analysis results of 96,333 unigenes, two pathways related to triterpene saponin biosynthesis in *Hedera helix* L., i.e., the terpenoid backbone biosynthesis and sesquiterpenoid and triterpenoid biosynthesis pathways, were assigned; these two pathways contained the MVA and MEP pathways upstream and carbocyclic biosynthesis midstream. From the annotation results, 339 unigenes were mapped to the terpenoid backbone biosynthesis pathway and 71 unigenes were mapped to the sesquiterpenoid and triterpenoid biosynthesis pathway. Among these unigenes, the genes encoding key enzymes involved in triterpene saponin biosynthesis are listed in full in Table 4: 6 putative genes (*AACT*, *HMGS*, *HMGR*, *MVK*, *PMVK*, *MVD*) and 8 putative genes (*DXS*, *DXR*, *ispD*, *ispE*, *ispF*, *ispG*, *ispH*, *IDI*) were discovered in the MVA and MEP pathway, respectively, and 6 putative genes (*GPS*, *GGPS*, *FPS*, *SS*, *SE*, *β-AS*) were discovered for carbocyclic biosynthesis. In most cases, more than one unigene can be annotated to the same enzyme encoded gene or gene family [36,37]; this phenomenon was observed in this study. As shown in Table 4, most genes had multiple alleles or paralogues in the transcriptome, such as *1-deoxy-D-xylulose-5-phosphate synthase* (*DXS*, EC: 2.2.1.7), *geranylgeranyl diphosphate synthase* (*GGPS*, EC: 2.5.1.29) and *squalene synthase* (*SS*, EC: 2.5.1.21), these genes were annotated with nine paralogues. Meanwhile, in the above pathways related to triterpene saponin biosynthesis, we also identified genes with different expression levels in leaf compared to roots using

COG Function Classification of All–Unigene.fa Sequence

A: RNA processing and modification
B: Chromatin structure and dynamics
C: Energy production and conversion
D: Cell cycle control, cell division, chromosome partitioning
E: Amino acid transport and metabolism
F: Nucleotide transport and metabolism
G: Carbohydrate transport and metabolism
H: Coenzyme transport and metabolism
I: Lipid transport and metabolism
J: Translation, ribosomal structure and biogenesis
K: Transcription
L: Replication, recombination and repair
M: Cell wall/membrane/envelope biogenesis
N: Cell motility
O: Posttranslational modification, protein turnover, chaperones
P: Inorganic ion transport and metabolism
Q: Secondary metabolites biosynthesis, transport and catabolism
R: General function prediction only
S: Function unknown
T: Signal transduction mechanisms
U: Intracellular trafficking, secretion, and vesicular transport
V: Defense mechanisms
W: Extracellular structures
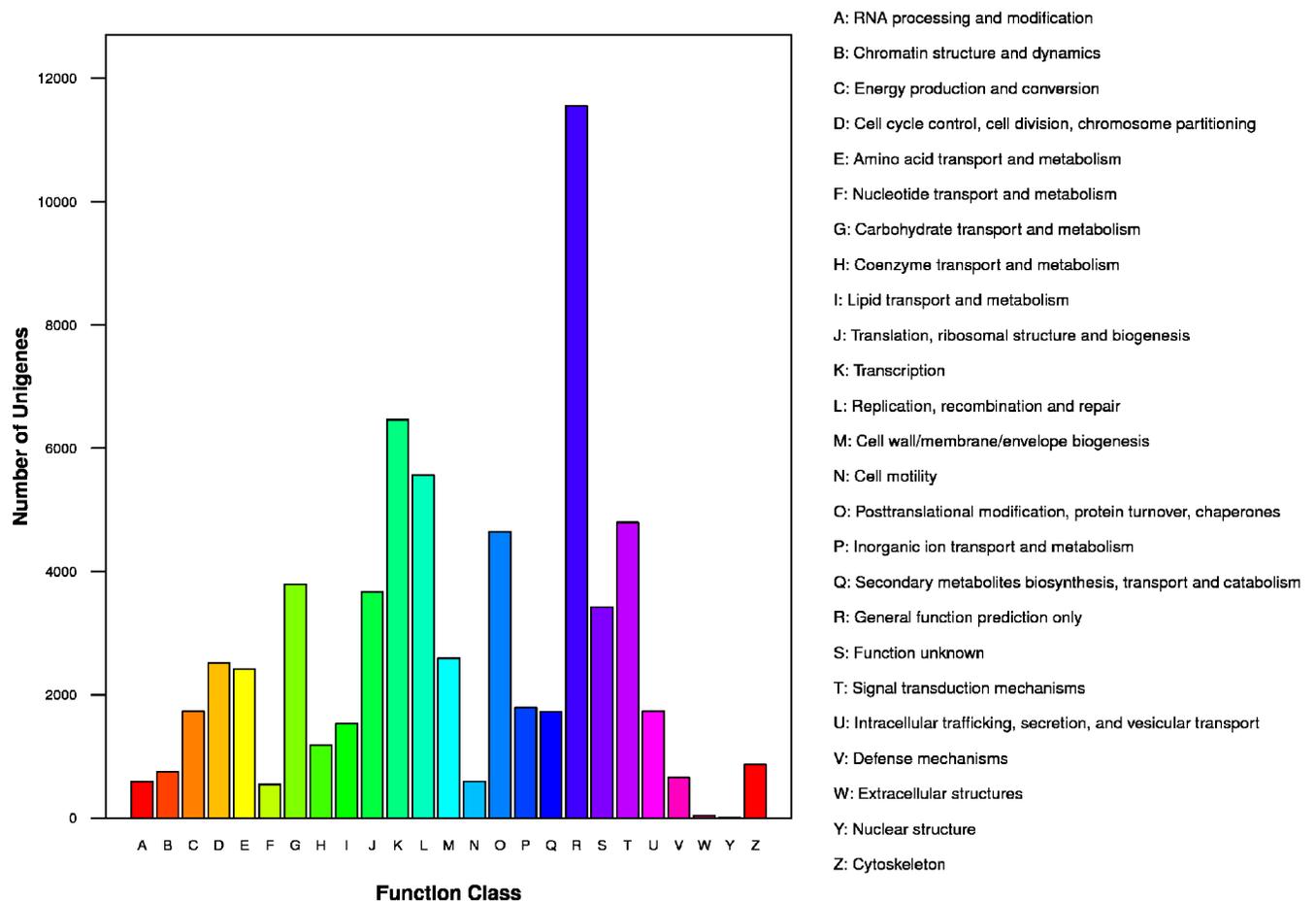Y: Nuclear structure
Z: Cytoskeleton

Fig 4. COG function classification of *H. helix* transcriptome. A total of 33,205 unigenes showed significant homology (E-value $\leq$1.0E-5) to genes in one of the 25 categories (A-W, Y and Z) in the NCBI COGs database.

combined DEG analysis and FPKM calculation results. There are 13 unigenes belonging to 8 putative genes expressed higher and 8 unigenes belonging to 4 putative genes expressed lower in leaves (Table 4). In addition, the modification of the triterpenoid carbocyclic is a putative downstream pathway of triterpene saponin biosynthesis in *H. helix*; specifically, CYP450s and GTs played a major role in this part. In this study, a total of 269 and 197 unigenes were identified for the CYP450 family and GT family, respectively (S9 and S10 Tables).

Validation by RT-qPCR and RACE Clone

To confirm the accuracy of the Illumina paired-end sequencing and FPKM calculated results, we selected 12 unigenes and used RT-qPCR to determine their relative expression level in the leaf and root tissues of *H. helix*. All 12 unigenes were triterpene saponin biosynthesis genes and contained 5 higher-expressed unigenes unigenes (*DXS*_CL1741, *iSPF*_Unigene5248, *iSPH*_CL3923, *SS*_CL11265, *SE*_CL6504), 3 lower-expressed unigenes unigenes (*HMGR*_CL84, *MVK*_CL10135, *DXR*_CL4453) and 4 unchanged unigenes (*AACT*_CL10734, *HMGS*_CL4843, *iSPE*_Unigene29532) as calculated by FPKM. The RT-qPCR and FPKM results were compared and are presented in Fig 6, the expression levels are similar.
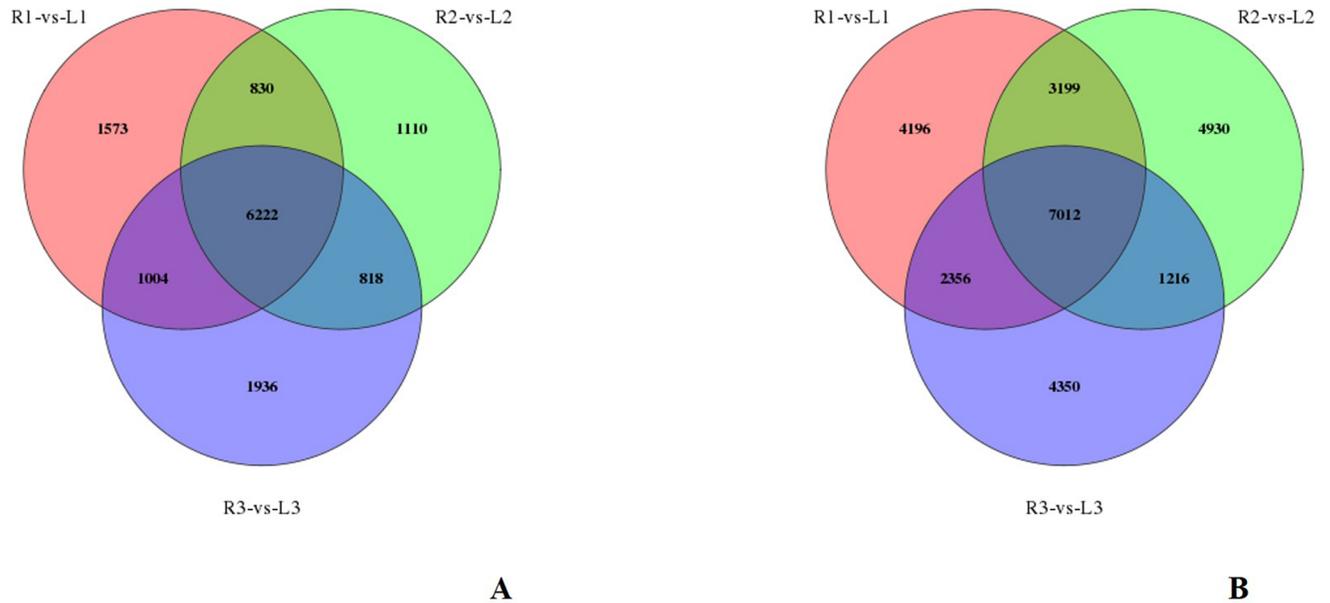
Fig 5. Differentially expressed gene analysis of six libraries in *H. helix*. (A) Expressed higher unigenes in leaf samples. (B) Expressed lower unigenes in leaf samples.

To further validate the *de novo* assembly results, we also cloned the full-length cDNA of 2 genes (*HMGR accession No.* KX056076, *SE accession No.* KU942524) from the *H. helix* leaf using the RACE method according to fragments of the transcriptome (*HMGR*_CL84, *SE*_CL6504). The full-length cDNA sequence of *HMGR* is 2215 bp and includes a complete open reading frame of 1731 bp, encoding a polypeptide of 577 amino acids (S3 Fig). A phylogenetic tree shows that the HMGR gene and SE gene in *H. helix* are highly homologous with other species. The predicted molecular weight of the HMGR putative protein is 61.98kD, with a theoretical pI of 7.52. Similarly, the full-length cDNA sequence of *SE* is 2040 bp and includes a complete open reading frame of 1615 bp, encoding a polypeptide of 537 amino acids (S4 Fig). A phylogenetic tree was built to investigate the evolutionary relationship with other plants and found higher homology between them. The putative protein of SE had a 58.69-kD predicted molecular weight with a theoretical pI of 8.95. Both the full-length cDNA sequences of *HMGR* and *SE* had a Poly (A) signal sequence.

Discussion

As it is a traditional medicinal plant in Europe, previous studies of *H. helix* have mainly focused on its pharmacology, efficacy and active components [1,3,7]; molecular biology research is less common, and genomic and transcriptomic data are still not available in NCBI database for this plant. As the first study using next-generation sequencing technology to obtain transcript coverage of *H. helix*, our deep sequencing provides important, fundamental molecular data. In the current study, the clean reads and *de novo* assembled unigenes were obtained using the Illumina HiseqTM 2500 platform (3 replicates of leaf and root tissues, 6 libraries in total). Their N50 length and average length indicated that all sequencing results were assembled effectively and with high quality [37,38]. In transcriptome analysis without a reference genome, unigenes were mapped in a public database, and unigenes with a high matching rate by the homology principle were annotated and classified. The results indicate

Table 4. Discovery and expression of unigenes involved in triterpenesaponin biosynthesis in *Hedera helix* L.

Enzymes name	Abbreviation	EC number	Putative ortholog	Expressed higher	Expressed lower
Acetyl-CoA acetyl transferase	AACT	EC:2.3.1.9	CL10734, CL8643		
Hydroxymethylglutaryl CoA synthase	HMGS	EC:2.3.3.10	CL4883, Unigene7122		
3-hydroxy-3-methylglutaryl-coenzyme A reductase	HMGR	EC:1.1.1.34	CL84, Unigene12643, Unigene12948, Unigene9993,	Unigene9993	CL84, Unigene12643
Mevalonate kinase	MVK	EC:2.7.1.36	CL10135, Unigene31942, Unigene37001		CL10135
Phosphomevalonate kinase	PMVK	EC:2.7.4.2	CL2755		
Mevalonate diphosphosphate decarboxylase	MVD	EC:4.1.1.33	CL12343		
1-deoxy-D-xylulose-5-phosphate synthase	DXS	EC:2.2.1.7	CL1741, CL6964, CL7506, Unigene12956, Unigene20321, Unigene26465, Unigene26467, Unigene26469, Unigene7001	CL1741	
1-deoxy-D-xylulose-5-phosphate reductoisomerase	DXR	EC:1.1.1.267	CL4453, Unigene20397, Unigene27513, Unigene32536, Unigene5694		CL4453, Unigene5694
2-C-methyl-D-erythritol 4-phosphate cytidylyltransferase	ispD	EC:2.7.7.60	CL177, Unigene28885		
4-diphosphocytidyl-2-C-methyl-D-erythritol kinase	ispE	EC:2.7.1.148	Unigene15885, Unigene29532		
2-C-methyl-D-erythritol 2,4-cyclodiphosphate synthase	ispF	EC:4.6.1.12	Unigene5248	Unigene5248	
(E)-4-hydroxy-3-methylbut-2-enyl-diphosphate synthase	ispG	EC:1.17.7.1	CL4323, Unigene18905, Unigene21207, Unigene39208		
4-hydroxy-3-methylbut-2-enyl diphosphate reductase	ispH	EC:1.17.7.4	CL3923, Unigene11112, Unigene39239, Unigene40179	CL3923	
Isopentenyl-diphosphate Delta-isomerase	IDI	EC:5.3.3.2	Unigene9222		
Geranyl diphosphate synthase	GPS	EC:2.5.1.1	CL3895, Unigene23216, Unigene31228		
Geranylgeranyl diphosphate synthase	GGPS	EC:2.5.1.29	CL12915, CL1497, CL3631, CL5060, CL8765, Unigene20086, Unigene26473, Unigene6117, Unigene7971		Unigene26473, CL1497, Unigene20086
Farnesyl diphosphate synthase	FPS	EC:2.5.1.10	CL8585, Unigene7386, Unigene7389, Unigene256	CL8585	
Squalene synthase	SS	EC:2.5.1.21	CL11265, Unigene15662, Unigene17742, Unigene18488, Unigene28187, Unigene30735, Unigene35625, Unigene7507, Unigene14696	CL11265, Unigene35625, Unigene7507, Unigene15662	
Squaleneepoxidase	SE	EC:1.14.14.17	CL6504, CL9719, CL9981, Unigene14156, Unigene17245	CL6504,CL9981	
β-amyrin synthase	β-AS	EC:5.4.99.39	CL5897, CL1580, CL11721, Unigene29516, Unigene32344	CL11721, CL1580	

that the transcriptomes of *H. helix* have great accuracy and integrity and can be used for functional genes studies or other molecular biology studies on *H. helix*.

As triterpenoid saponins are the main functional components in *H. helix* [4,9,10], here we aimed to expand knowledge of their biosynthesis pathway and screen related functional genes
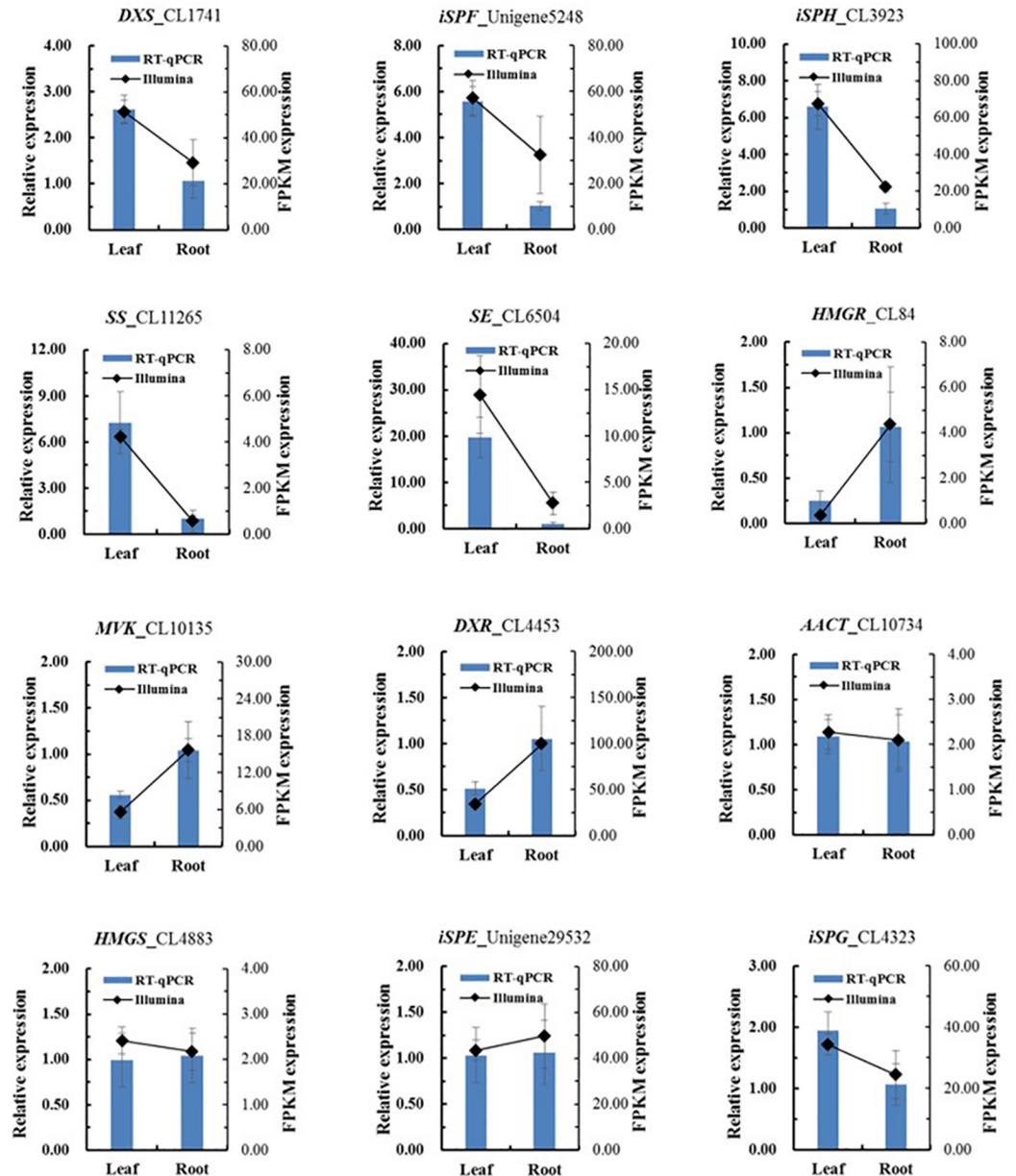
Fig 6. RT-qPCR validation of selected unigenes involved in triterpene saponin biosynthesis. Columns indicate relative expression obtained by RT-qPCR (left y-axis); lines indicating the expression level were calculated by FPKM method (right y-axis). All data are presented as mean value of three repeats.

https://doi.org/10.1371/journal.pone.0182243.g006

using transcriptome sequencing. The HPLC results clearly show that triterpenoid saponin contents are significantly different in leaves and roots (S1 Fig); therefore, choosing the leaf and root for comparative transcriptome analysis will greatly facilitate dissection of the genes involved in organ-specific secondary metabolite biosynthesis of triterpenoid saponins. This approach is widely used for mining and identifying novel genes in secondary metabolite biosynthesis [22, 39]. In this study, with *de novo* assembling and deciphering, the terpenoid biosynthesis backbone and sesquiterpenoid and triterpenoid biosynthesis pathway were

addressed, and 20 functional genes were found. Although the two pathways are common pathways in the biosynthesis of terpenoids and sterols reported in *Panax notoginseng* [36] and *Solanum elaeagnifolium* [26], our results provide the first accurate and comprehensive gene information for *H. helix.*

CYP450s [19, 40, 41] and GTs [42,43], which modify the triterpene carbon ring, play key roles in the biosynthesis of various natural products. Moreover, CYP450s catalyse an irreversible oxygenating reaction and are multifunctional enzymes with a very complicated functional classification. Therefore, they have attracted special attention in metabolic engineering. However, in the synthesis pathway of terpenoid-based natural products, only 19 CYP450s gene functions have been validated [40].

Based on our transcriptome sequencing, we annotated 4 CYP450s genes in terpenoid biosynthesis: CYP716A52v2 with β-Amyrin 28-oxidase function, which produces oleanane-type ginsenosides [44]; CYP716A47 with protopanaxadiol synthase function that produces protopanaxadiol [45], CYP72A1 with secologanin synthase function that produces secologanin [46], CYP71D55 with premnaspirodiene oxygenase function that produces solavetivone [47]. Due to the versatile function of CYP450s, the exact function of the annotated CYP450s in terpenoid biosynthesis must still be illustrated. Glycosyltransferase (GT) participates in the last step in triterpenoid saponin biosynthesis; the GT family also has a complicated function as a member of the CYP450 family. Lahoucine Achine *et al* accurately validated that Ugt73k1 and Ugt71g1 participates in triterpene saponin biosynthesis [48]. In our transcriptome, there are 197 unigene with GT activity; however, we did not annotate these two GT genes but only annotated the obtained unigenes belonging to the Ugt73 sub-family. This indicates that GTs are not conserved in different species, even when catalysing the same reaction.

Testing and verification of the accuracy of *de novo* assembled sequences is one of the most crucial steps for no-reference transcriptome research. In this study, we calculated the relative expression of 12 unigenes using an RT-qPCR method and cloned full-length putative *HMGR* and *SE* genes for validation. These two-part results mutually confirmed that our *de novo* assembled sequences were accurate and useful. The twelve chosen unigenes all belong to the triterpenoid saponins biosynthesis pathway and represent three types of expression patterns (higher, lower, or no change) in leaf transcripts by FPKM calculation. *HMGR* and *SE* were selected as examples to validate the accuracy because they represent the genes with highest and lowest expression level in leaf samples. In addition, *HMGR* is the first rate-limiting enzyme in the terpenoid biosynthesis pathway, and many reports have presented *HMGR* as an important gene in the terpenoid biosynthesis pathway [49, 50]; *SE* participates in the first oxidation reaction in the triterpenoid saponins biosynthesis pathway and promoting squaleneep oxidation to produce 2,3-oxidosqualene. *SE* is also an important regulation gene, as indicated by many studies [51, 52]. In summary, this transcriptome is successful and reliable and can provide data to support further molecular research on *H. helix.*

Conclusions

In the current study, comparative transcriptomes between leaf and root tissues in H. helix were implemented using the Illumina HiseqTM 2500 platform. After *de novo* assembly and sequence annotation, a total of 96,333 unigenes were obtained with a mean length of 1,385 nt; 32,443 annotated sequences were assigned to 25 COG categories, 50,479 sequences were categorized into 56 GO functional groups and 47,100 unigenes were assigned to 128 KEGG pathways. After differentially expressed gene analysis, 6,222 unigenes were more highly expressed and 7,012 unigenes were less expressed in leaf samples. The RT-qPCR analysis and RACE clone results indicated that *de novo* assembled sequences were accurate and valuable. Two pathways and 20

putative genes related to triterpenoid saponin biosynthesis were discovered. These data will enrich our knowledge and provide a theoretical foundation for molecular research on *H. helix*.

Supporting information

S1 Fig. HPLC chromatogram of different tissue samples of *H. helix*.
(TIF)

S2 Fig. Scatter plot of top 20 KEGG pathway enrichment for DEGs.
(TIF)

S3 Fig. The amino acid sequence of *HMGR* gene in *H. helix*.
(TIF)

S4 Fig. The amino acid sequence of *SE* gene in *H. helix*.
(TIF)

S1 Table. Annotation summary.
(XLS)

S2 Table. Primers for RT-qPCR analysis.
(DOCX)

S3 Table. Primer sequences for RACE clone.
(DOCX)

S4 Table. All-Unigene GO analysis.
(XLS)

S5 Table. All-Unigene COG classification.
(XLS)

S6 Table. All-Unigene KEGG analysis.
(XLS)

S7 Table. Expressed higher DEGs in leaf.
(XLS)

S8 Table. Expressed lower DEGs in leaf.
(XLS)

S9 Table. Annotation_CYP450s.
(XLS)

S10 Table. Annotation_GTs.
(XLS)

Author Contributions

Conceptualization: Fei Qiao, Xiaohong Zhong.

Data curation: Huapeng Sun, Fang Li, Xiaohong Zhong.

Formal analysis: Huapeng Sun, Zijian Xu, Mengli Sun, Hanqing Cong, Fei Qiao, Xiaohong Zhong.

Funding acquisition: Fei Qiao, Xiaohong Zhong.

Methodology: Huapeng Sun, Fang Li.

Project administration: Huapeng Sun, Fei Qiao, Xiaohong Zhong.

Resources: Huapeng Sun.

Validation: Huapeng Sun, Fang Li.

Writing – original draft: Huapeng Sun, Fei Qiao, Xiaohong Zhong.

Writing – review & editing: Huapeng Sun, Fang Li, Zijian Xu, Mengli Sun, Hanqing Cong, Fei Qiao, Xiaohong Zhong.

References

1. Lutsenko Y, Bylka W, MatlawskaI, Darmohray R. (2010) *Hedera helix* as a medicinal plant. Herba. Polonica 56: 83–96.

2. Landgrebe H, Matusch R, Marburg FR, Hecker M. (1999) Effectiveness and use of an old medicinal plant. Pharm Zeitung Nr 35: 11–15.

3. Fazio S, Pouso J, Dolinsky D, Fernandez A, Hernandez M, Clavier G, et al. (2009) Tolerance, safety and efficacy of *Hedera helix* extract in inflammatory bronchial diseases under clinical practice conditions: a prospective, open, multicentre post marketing study in 9657 patients. Phytomedicine 16: 17–24. https://doi.org/10.1016/j.phymed.2006.05.003 PMID: 16860549

4. Hofmann D, Hecker M, Völp A. (2003) Efficacy of dry extract of ivy leaves in children with bronchial asthma-a review of randomized controlled trials. Phytomedicine 10: 213–220. https://doi.org/10.1078/094471103321659979 PMID: 12725580

5. Marzian O. (2007) Treatment of acute bronchitis in children and adolescents. Non-interventional post marketing surveillance study confirms the benefit and safety of a syrup made of extracts from thyme and ivy leaves. MMW Fortschr Med 149: 69–74. PMID: 17619603

6. Cwientzek U, Ottillinger B, Arenberger P. (2011) Acute bronchitis therapy with ivy leaves extracts in a two-arm study. A double-blind, randomised study vs. another ivy leaves extract. Phytomedicine 18: 1105–1109. https://doi.org/10.1016/j.phymed.2011.06.014 PMID: 21802921

7. Holzinger F, Chenot JF. (2011) Systematic review of clinical trials assessing the effectiveness of Ivy leaf (*Hedera Helix*) for acute upper respiratory tract infections. Evid-based Compl Alt 1: 1–9.

8. Stauss-Grabo M, Atiye S, Warnke A, Wedemeyer R, Donath F, Blume H. (2011) Observational study on the tolerability and safety of film-coated tablets containing ivy extract (Prospan® Cough Tablets) in the treatment of colds accompanied by coughing. Phytomedicine 18: 433–436. https://doi.org/10.1016/j.phymed.2010.11.009 PMID: 21211950

9. Khdair A, Mohammad MK, Tawaha K, Al-Hamarsheh E, AlKhatib HS, Al-Khalidi B, et al. (2010) A validated RP HPLC-PAD method for the determination of hederacoside C in Ivy-Thyme cough syrup. Int J Anal Chem PMID: 20862201

10. Hussien SA and Awad ZJ. (2014) Isolation and characterization of triterpenoid saponin Hederacoside C present in the leaves of *Hedera helix* L. Cultivated in Iraq. Iraq J Pharm Sci 23: 33–41.

11. Prescott AK, Rigby LP, Veitch NC, Simmonds SJ. (2014) The haploinsufficiency profile of a-hederin suggests a caspofungin-like antifungal mode of action. Phytochemistry 101: 116–120. https://doi.org/10.1016/j.phytochem.2014.01.022 PMID: 24569176

12. Mendel M, Chłopecka M, Dziekan N, Wiechetek M. (2013) Participation of extracellular calcium in α-hederin-induced contractions of rat isolated stomach strips. J Ethnopharmacol 146: 423–426. https://doi.org/10.1016/j.jep.2012.12.023 PMID: 23274745

13. Vincken JP, Heng L, de Groot A, Gruppena H. (2007) Saponins, classification and occurrence in the plant kingdom. Phytochemistry 68: 275–297. https://doi.org/10.1016/j.phytochem.2006.10.008 PMID: 17141815

14. Haralampidis K, Trojanowska M, Osbourn AE. (2002) Biosynthesis of triterpenoid saponins in plants. AdvBiochemEngBiot 75: 31–49.

15. Choi DW, Jung JD, Ha YI, Park HW, In DS, Chung HJ, et al. (2005) Analysis of transcripts in methyl jasmonate-treated ginseng hairy roots to identify genes involved in the biosynthesis of ginsenosides and other secondary metabolites. Plant cell rep 23: 557–566. https://doi.org/10.1007/s00299-004-0845-4 PMID: 15538577

16. Rohdich F, Lauw S, Kaiser J, Feicht R, Kohler P, Bacher A, et al. (2006) Isoprenoid biosynthesis in plants-2C-methyl-D-erythritol-4-phosphate synthase (IspC protein) of Arabidopsis thaliana. Febs J 273: 4446–4458. https://doi.org/10.1111/j.1742-4658.2006.05446.x PMID: 16972937

17. Leopoldini M, Malaj N, Toscano M, Sindona G, Russo N. (2010) On the inhibitor effects of bergamot juice flavonoids binding to the 3-hydroxy-3-methylglutaryl- CoA reductase (HMGR) enzyme. J Agr Food Chem 58: 10768–10773.

18. Aharoni A, Jongsma MA, Kim TY, Ri MB, Giri AP, Verstappen FWA, et al. (2006) Metabolic engineering of terpenoid biosynthesis in plants. Phytochem Rev 5: 49–58.

19. Carelli M, Biazzi E, Panara F, Tava A, Scaramelli L, Porceddu A, et al. (2011) *Medicago truncatula* CYP716A12 is a multifunctional oxidase involved in the biosynthesis of hemolytic saponins. Plant Cell 23: 3070–3081. https://doi.org/10.1105/tpc.111.087312 PMID: 21821776

20. Jayakodi M, Lee SC, Park HS, Jang WJ, Lee YS, Choi BS, et al. (2014) Transcriptome profiling and comparative analysis of *Panax ginseng* adventitious roots. J Ginseng Res 38: 278–288. https://doi.org/10.1016/j.jgr.2014.05.008 PMID: 25379008

21. Liu MH, Yang BR, Cheung WF, Yang KY, Zhou HF, Kwok JSL, et al. (2015) Transcriptome analysis of leaves, roots and flowers of *Panax notoginseng* identifies genes involved in ginsenoside and alkaloid biosynthesis. BMC Genomics 16: 265. https://doi.org/10.1186/s12864-015-1477-5 PMID: 25886736

22. Upadhyay S, Phukan UJ, Mishra S, Shukla RK. (2014) De novo leaf and root transcriptome analysis identified novel genes involved in Steroidal sapogenin biosynthesis in *Asparagus racemosus*. BMC Genomics 15: 746. https://doi.org/10.1186/1471-2164-15-746 PMID: 25174837

23. Qiao F, Cong HQ, Jiang XF, Wang RX, Yin JM, Qian D, et al. (2014) De Novo Characterization of a *Cephalotaxus hainanensis* Transcriptome and Genes Related to Paclitaxel Biosynthesis. PLoS ONE https://doi.org/10.1371/journal.pone.0106900 PMID: 25203398

24. Li J, Zhen W, Long DK, Ding L, Gong AH, Xiao CH, et al. (2016) De Novo Sequencing and Assembly Analysis of the *Pseudostellaria heterophylla* Transcriptome. PLoS ONE https://doi.org/10.1371/journal.pone.0164235 PMID: 27764127

25. Kumari A, Singh HR, Jha A, Swarnkar MK, Shankar R, Kumar S. (2014) Transcriptome sequencing of rhizome tissue of *Sinopodophyllum hexandrumat* two temperatures. BMC Genomics 15: 871. https://doi.org/10.1186/1471-2164-15-871 PMID: 25287271

26. Tsaballa A, Nikolaidis A, Trikka F, Ignea C, Kampranis SC, Makris AM, et al. (2015) Use of the de novo transcriptome analysis of silver-leaf nightshade (*Solanum elaeagnifolium*) to identify gene expression changes associated with wounding and terpene biosynthesis. BMC Genomics 16: 504. https://doi.org/10.1186/s12864-015-1738-3 PMID: 26149407

27. Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, et al. (2011) Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat Biotechnol 29: 644–652. https://doi.org/10.1038/nbt.1883 PMID: 21572440

28. Perteal G, Huang XQ, Liang F, Antonescu V, Sultanal R, Karamycheval S, et al. (2003) TIGR gene indices clustering tools (TGICL): a software system for fast clustering of large EST datasets. Bioinformatics 19: 651–652. PMID: 12651724

29. Kanehisal M, Araki M, Goto S, Hattori M, Hirakawal M, Itoh M, et al. (2008) KEGG for linking genomes to life and the environment. Nucleic Acids Res 36: 480–484.

30. Altschul SF, Madden TL, Schäfferl AA, Zhang JH, Zhang Z, Miller W, et al. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25: 3389–3402. PMID: 9254694

31. Gotz S, Garcia-Gomez JM, Terol J, Williams TD, Nagaraj SH, Nueda MJ, et al. (2008) High-throughput functional annotation and data mining with the Blast2GO suite. Nucleic Acids Res 36: 3420–3435. https://doi.org/10.1093/nar/gkn176 PMID: 18445632

32. Young MD, Wakefield MJ, Smyth GK, Oshlack A. (2010) Gene ontology analysis for RNA-seq: accounting for selection bias. Genome Biology 11: R14. https://doi.org/10.1186/gb-2010-11-2-r14 PMID: 20132535

33. Mortazavil A, Williams BA, McCuel K, Schaeffer L, Wold B. (2008) Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nat. Methods 5: 621–628. https://doi.org/10.1038/nmeth.1226 PMID: 18516045

34. Audic S, Claverie JM. (1997) The significance of digital gene expression profiles. Genome Res 7: 986–995. PMID: 9331369

35. Sun HP, Li F, Ruan QM, Zhong XH. (2016) Identification and validation of reference genes for quantitative real-time PCR studies in *Hedera helix* L. Plant PhysiolBioch 108: 286–294.

36. Liu MH, Yang BR, Cheung WF, Yang KY, Zhou HF, Kwok JSL, et al. (2015) Transcriptome analysis of leaves, roots and flowers of *Panax notoginseng* identifies genes involved in ginsenoside and alkaloid biosynthesis. BMC Genomics 16: 265. https://doi.org/10.1186/s12864-015-1477-5 PMID: 25886736

37. Zhang XD, Allan AC, Li CX, Wang YZ, Yao QY. (2015) De Novo assembly and characterization of the transcriptome of the Chinese medicinal herb, *Gentiana rigescens*. Int J MolSci 16: 11550–11573.

38. Tian XJ, Long Y, Wang J, Zhang JW, Wang YY, Li WM, et al. (2015) De novo transcriptome assembly of common wild rice (*Oryza rufipogon* Griff.) and discovery of drought-response genes in root tissue based on transcriptomic data. PLoS ONE https://doi.org/10.1371/journal.pone.0131455 PMID: 26134138

39. Gurkok T, Turktas M, Parmaksiz I,Unver T. (2015) Transcriptome Profiling of Alkaloid Biosynthesis in Elicitor Induced Opium Poppy. Plant Mol Biol Rep 33(3): 673–688.

40. Rasool S, Mohamed R. (2016) Plant cytochrome P450s: nomenclature and involvement in natural product biosynthesis. Protoplasma 253(5):1197–1209. https://doi.org/10.1007/s00709-015-0884-4 PMID: 26364028

41. Weitzel C, Simonsen HT. (2015) Cytochrome P450-enzymes involved in the biosynthesis of mono-and sesquiterpenes. Phytochem Rev 14: 7–24.

42. Achnine L, Huhman DV, Farag MA, Sumner LW, Blount JW, Dixon RA. (2005) Genomics based selection and functional characterization of triterpene glycosyltransferases from the model legume *Medicago truncatula*. Plant J 41: 875–887. https://doi.org/10.1111/j.1365-313X.2005.02344.x PMID: 15743451

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

44. Han JY, Kim MJ, Ban YW, Hwang HS, Choi YE. (2013) The involvement of β-amyrin 28-oxidase (CYP716A52v2) in oleanan-type ginsenoside biosynthesis in Panax ginseng. Plant Cell Physiol 54: 2034–2046. https://doi.org/10.1093/pcp/pct141 PMID: 24092881

45. Han JY, Kim HJ, Kwon YS, Choi YE. (2011) The Cyt P450 enzyme CYP716A47 catalyzes the formation of protopanaxadiol from dammarenediol-II during ginsenoside biosynthesis in Panax ginseng. Plant Cell Physiol 52: 2062–2073. https://doi.org/10.1093/pcp/pcr150 PMID: 22039120

46. Irmler S, Schroder G, St-Pierre B, Crouch NP, Hotze M, Schmidt J, et al. (2000) Indole alkaloid biosynthesis in Catharanthus roseus: new enzyme activities and identification of cytochrome P450 CYP72A1 as secologanin synthase. Plant J 24: 797–804. PMID: 11135113

47. Takahashi S, Yeo YS, Zhao Y, O'Maille PE, Greenhagen BT, Noel JP, et al. (2007) Functional characterization of premnaspirodiene oxygenase a cytochrome P450 catalysing region and stereo-specific hydroxylation of diverse sesquiterpene substrates. J Biol Chem 282: 31744–31754. https://doi.org/10.1074/jbc.M703378200 PMID: 17715131

48. Achnine L, Huhman DV, Farag MA, Sumner LW, Blount JW, Dixon RA. (2005) Genomics based selection and functional characterization of triterpene glycosyltransferases from the model legume Medicago truncatula. Plant J 41: 875–887. https://doi.org/10.1111/j.1365-313X.2005.02344.x PMID: 15743451

49. Wang GY, Keasling JD. (2002) Amplification of HMG-CoA reductase production enhances carotenoid accumulation in *Neurospora crassa*. MetabEng 4: 193–201.

50. Dai Zb, Cui Gh, Zhou SF, Zhang XN, Huang LQ. (2011) Cloning and characterization of a novel 3-hydroxy-3-methylglutaryl coenzyme A reductase gene from Salvia miltiorrhiza involved in diterpenoid tanshinone accumulation. J Plant Physiol 168: 148–157. https://doi.org/10.1016/j.jplph.2010.06.008 PMID: 20637524

51. Suzuki H, Achnine L, Xu R, Matsuda SPT, Dixon RA. (2002) A genomics approach to the early stages of triterpene saponin biosynthesis in *Medicago truncatula*. Plant J 32: 1033–1048. PMID: 12492844

52. Han JY, In JG, Kwon YS, Choi YE. (2010) Regulation of ginsenoside and phytosterol biosynthesis by RNA interferences of squalene epoxidase gene in *Panax ginseng*. Phytochemistry 71: 36–46. https://doi.org/10.1016/j.phytochem.2009.09.031 PMID: 19857882