ORIGINAL ARTICLE

Transcriptome analysis of *Akebia quinata* (Thunb.) Decne. and discovery of key enzyme genes in the phenylpropanoid biosynthesis pathway

Chengcheng Qian[1] · Yatian Yang[1] · Rui Wu[1] · Bangxing Han[2] · Deling Wu[1] · Jinmei Ou[1,3,4]

Abstract

Akebia quinata is a dry vine stem of *Akebia quinata* (Thunb.) Decne. in the Lardizabalaceae Akebia Decne, which has antibacterial, anti-tumor, and diuretic properties. Calceolarioside B is a known effective component of *Akebia quinata*. Moreover, calceolarioside B is a phenylpropanoid, the current understanding of the molecular mechanisms of phenylpropanoid biosynthetic pathway is not clear. In this study, we performed RNA-Seq analysis of the flowers, leaves, roots, and stems of *Akebia quinata* using the BGISEQ-500 platform. The assembly of transcripts from all four types of tissues generated 123,576 unigenes, of which 90,273 were mapped to several public databases for functional annotation. We identified 182 genes encoding 13 key enzymes involved in phenylpropanoid biosynthesis through their KEGG annotation. DEGs were identified via the comparison of gene expression levels between stems and other tissues (flowers, leaves, and roots). We also verified the expression levels of some unigenes encoding key enzymes using qRT-PCR. This analysis of the phenylpropanoid biosynthesis pathway and its crucial enzymes in *Akebia quinata* lays a foundation for uncovering the regulatory mechanism of phenylpropanoid biosynthesis. The expression patterns of genes involved in the phenylpropanoid biosynthesis pathway were verified using qRT-PCR. We suggest that transcripts of 4CL may be key regulators in the phenylpropanoid biosynthetic pathway. Our results greatly extend the public transcriptome data set of *Akebia quinata* and provide valuable information for the identification of candidate genes involved in phenylpropanoid biosynthesis.

Keywords *Akebia quinata* (Thunb.) Decne. · RNA-Seq · Phenylpropanoid biosynthesis · DEGs · qRT-PCR

Abbreviations

RNA-Seq	RNA sequencing
KEGG	Kyoto Encyclopedia of Genes and Genomes
DEGs	Differentially expressed genes
qRT-PCR	Quantitative real-time polymerase chain reaction
NR	NCBI non-redundant protein sequence
NT	NCBI nucleotide sequence
KOG	Clusters of euKaryotic Orthologous Groups
PFAM	Protein families database
GO	Gene ontology
PAL	Phenylalanine ammonia-lyase
C4H	Cinnamate 4-hydroxylase
4CL	4-Coumarate–CoA ligase
HCT	Shikimate O-hydroxycinnamoyltransferase
C3′H	5-O-(4-coumaroyl)-D-quinate 3′-monooxygenase
CCoAOMT	Caffeoyl-CoA O-methyltransferase
CCR	Cinnamoyl-CoA reductase
F5H	Ferulate-5-hydroxylase
COMT	Caffeic acid 3-O-methyltransferase
CAD	Cinnamyl-alcohol dehydrogenase
CHS	Chalcone synthase
CHI	Chalcone isomerase
F3H	Naringenin 3-dioxygenase

Communicated by V. P. Singh.

✉ Jinmei Ou
  toojm9319@163.com

1 College of Pharmacy, Anhui University of Chinese Medicine, Hefei 230012, China

2 College of Biological and Pharmaceutical Engineering, West Anhui University, Lu'an 237012, China

3 Anhui Key Laboratory of New Technology of Chinese Herbs Pieces Manufacturing, Hefei 230012, China

4 Institute of Conservation and Development of Traditional Chinese Medicine Resources, Anhui Academy of Chinese Medicine, Hefei 230012, China

ORF   Open reading frame
TFs   Transcription factors

Introduction

Akebia quinata (Thunb.) Decne., a dicotyledonous Lardizabalaceae Akebia Decne plant (Yang et al. 2022), is a traditional Chinese herbal medicine; the medicinal part is its dry stem, which is widely distributed in Shanxi, Anhui, Hubei, and Hunan (Liu and Wang 2011; Fu 2000; Li et al. 2010). The picking method has developed from wild picking to artificial cultivation (Zheng and Jin 2021). There are many medicinal active ingredients in *Akebia quinata* plants, such as calceolarioside B (Ou et al. 2015), These compounds have a variety of pharmacological effects, such as antibacterial, anti-inflammatory, anti-tumor, antioxidant, and immunomodulatory effects (Jiang et al. 2017; Julião Lde et al. 2009; Chu et al. 2009; Zhu et al. 2009; Ablise et al. 2008; Chen et al. 2009).

Phenylethanoid glycosides (PhGs) are glycosides composed of caffeic acid, phenylethanol aglycone, and glycosyl; hence, PhGs are also called phenylpropanoid glycosides (Wang et al. 2022). Calceolarioside B is a phenylpropanoid (Wang et al. 2014); moreover, there are other phenylpropanoids in *Akebia quinata* plants, such as caffeic acid and esculetin (Liu et al. 2020). Phenylpropanoids are naturally occurring compounds composed of benzene ring and three straight-chain carbon (C6-C3) groups; generally, they have a phenol structure and, thus are phenolic substances, including simple phenylpropanoid, coumarins, flavonoids, lignans, and lignins (Gu et al. 2012). The predominant parent sugar nucleus of phenylethanoid compounds is β-D-glucopyranose (Gao et al. 2015); during biosynthesis, most of these compounds are transformed from aromatic amino acids such as phenylalanine and tyrosine derived from shikimic acid, through a series of reactions such as deamination and hydroxylation. Phenylpropanoid compounds mainly exist in dicotyledonous plants such as Magnoliaceae and Rosaceae, and are found in traditional medicinal plants such as *Rehmannia glutinosa* (Wang et al. 2022), *Scrophularia ningpoensis* (Li et al. 2020), and *Forsythia suspense* (Peng et al. 2021). Phenylpropanoid metabolism is the most common pathway in plant secondary metabolism. Moreover, 4CL is a key enzyme in the biosynthesis of plant flavonoids and lignin, which is involved in the synthesis of a variety of secondary metabolites; it plays an important role in plant growth and development.

In recent years, with the development and application of high-throughput sequencing technology, RNA-Seq has been gradually applied in the study of a variety of medicinal plants. In the early stage, our research group completed transcriptome sequencing and database construction of *Akebia trifoliate* (Qian et al. 2022) and *Langmei* (Ou et al. 2020), mined the key enzyme genes of *Akebia trifoliata* triterpenoid saponins synthesis, and analyzed the expression of key enzyme genes of citrate synthesis in *Langmei* fruits. Presently, *Akebia quinata* research is mostly conducted in terms of quality evaluation and pharmacological action (Li 2018; Liu et al. 2012); nevertheless, the molecular understanding of *Akebia quinata* phenylpropanoid biosynthesis pathway has not been reported.

Based on the four tissue samples of *Akebia quinata* (flowers, leaves, roots, and stems), we applied high-throughput sequencing technology for transcriptome analysis and used qRT-PCR for data verification to mine the related genes and key enzymes of the *Akebia quinata* phenylpropanoid biosynthesis pathway, which laid a theoretical foundation for further exploration and development of phenylpropanoid biosynthesis and biology.

Materials and methods

Plant material and RNA extraction

In this study, the fresh *Akebia quinata* was selected as the experimental material and was harvested from Dashushan, Hefei City, Anhui Province on March 26, 2021. Four tissues (flowers, leaves, roots, and stems) were collected and identified as Akebia Decne Lardizabalaceae by Associate Professor Jinmei Ou of Anhui University of Chinese Medicine. Fresh plant samples were washed several times with ultra-pure water and then dried using a filter paper. Four tissue types were placed into 10 mL centrifuge tubes, instantly frozen in liquid nitrogen, and then stored in an ultra-low temperature refrigerator at −80 °C until their RNA extractions. The flowers, leaves, roots, and stems from three replicates were pooled together. The total RNA was extracted from each tissue using an RNA kit.

cDNA library construction and sequencing

The mRNA enrichment method was used to purify qualified total RNA with magnetic beads using Oligo (dT) and enrich the mRNA with a polyA tail. An appropriate amount of breaking reagent was added to the resulting mRNA sample and segmented under high temperature conditions. A strand of cDNA was synthesized using the broken mRNA as a template. Subsequently, the two-strand cDNA was synthesized by configuring the two-strand synthesis reaction system, and the two-strand cDNA was purified and recovered using the kit; the sticky end was repaired, base "A" was added to the 3′ end of the cDNA and connected the sequencing connector, and fragment size was selected. Finally, the appropriate fragment length for PCR amplification was selected to obtain

the cDNA library. Constructed libraries were examined using Agilent 2100 Bioanalyzer and ABI StepOnePlus real-time PCR System (Zhao et al. 2020; Shan et al. 2020a; Shi et al. 2019). After passing the test, Illumina HiSeq sequencing platform (BGI, Shenzhen, China) was used to conduct high-throughput sequencing analysis on the complete cDNA libraries of the four tissues of the qualified *Akebia quinata*.

De novo transcriptome assembly and gene functional annotation

After completing the transcriptome sequencing, Trimmomatic 0.36 was used to filter out reads containing joints (joint contamination), unknown base N content greater than 5%, and low quality reads to obtain high-quality reads. Clean reads with high quality were assembled using Trinity 2.0.6 software (Kim et al. 2017), and unigenes were obtained using Tgicl software for clustering and redundancy removal (Pertea et al. 2003), which were used as reference sequences for subsequent analysis. Using NCBI Blast+ (Pirooznia et al. 2008) software, unigenes NCBI nucleic acid sequence database (NT, https://www.ncbi.nlm.nih.gov/blast/db), NCBI non-redundancy protein sequence databases (NR, https://www.ncbi.nlm.nih.gov/blast/db), protein orthologous database (KOG, http://www.ncbi.nlm. nih.gov/KOG), the Swiss-Prot protein sequence databases (SwissProt, http://ftp.ebi.ac.uk/pub/databases/swissprot), the Kyoto encyclopedia of genes and genomes (KEGG, http://www.genome.jp/kegg), multi-sequence alignment were matched with the annotation database of protein family domain constructed by hidden Markov model (PFAM, http://pfam.Xfam.org). According to the annotation information of the NR library, the similarity between the transcription sequences of species and related species, as well as the functional information of homologous sequences were checked, and the genetic ontology (GO, http://geneontology.org) information of related unigenes was matched with Blast2GO 2.5.0. $E < 1 \times 10^{-5}$ was set to obtain the corresponding functional annotations and classification information of unigenes (Shan et al. 2020b).

Analysis of the structural characteristics of 4CL

The 4CL gene was translated using Translate tool online software, and the complete opening reading frame (ORF) sequence was screened. The protein multiple sequence alignment was implemented using DNAMAN and Clustalx software; the 4CL secondary structure was predicted by SOPMA, and NCBI-CDD database was used to predict conservative domains of 4CL. The SWISS-MODEL (https://swissmodel.expasy.org/) online system and PyMOL software were used to simulate 4CL tertiary structure.

Differentially expressed gene analysis

High-quality readings for each sample were checked using Bowtie2 (version 2.2.5) software (Langmead and Salzberg 2012) against the genome sequences of *Akebia quinata*, and the gene and transcript expression levels of each organ of *Akebia quinata* were calculated using the RSEM (version 1.2.8) software and TPM formula (Li and Dewey 2011). The standard expression of each unigene (fragments per kilobase of exon model per million mapped fragments, FPKM) was obtained. Differentially expressed genes (DEGs) were analyzed and compared between the stem and other tissues (flowers, leaves, and roots), using Poisson distribution method. Functional annotation and KEGG pathway enrichment analysis were performed on DEGs. By default, DEGs were those with false discovery rate (FDR) ≤ 0.001 and multiple of difference > 2.

Analysis of transcription factors (TFs) encoding genes

Genes with the ability of encoding TFs were predicted, and the TF families of genes were classified. The ORF of unigenes was detected by Getorf (EMBOSS version 6.5.7.0) (Rice et al. 2000); the ORF was aligned to the TF protein domain using Hmmsearch (version 3.0); and the competent unigenes were then identified based on the TF family features described by PlantTFDB (Mistry et al. 2013).

Analysis of the key gene expression levels in phenylpropanoid biosynthesis using qRT-PCR

To validate the *Akebia quinata* transcriptome data sets, the expression of key enzyme genes for phenylpropanoid biosynthesis was determined by quantitative PCR. According to the principles of primer design, genes involved in phenylpropanoid biosynthesis and primers for reference genes (actin) were designed using Primer Premier (version 5.0). After reverse transcription PCR, each PCR tube was prepared in a 15 μL mixture volume containing diluted cDNA (2 μL), 2.5 μM gene primer (upstream + downstream) (1.5 μL), 2 × qPCR mixture (7.5 μL), and Water Nuclease-Free (4 μL), under the following conditions: pre-denaturation at 95 °C for 30 s, then 40 cycles of denaturation at 95 °C for 15 s, and annealing or extension at 60 °C for 30 s; three technical replicates were performed for each sample. The relative expression level of each selected unigene was data-normalized with the actin gene (Un 4699), and the relative expression levels of each unigene were statistically analyzed using the $2^{-\Delta\Delta Ct}$ calculation method (Livak and Schmittgen 2001).

Results

Transcriptome sequencing and data assembly

Through transcriptome sequencing of *Akebia quinata*, 56.49 Gb of data were obtained. The clean reads Q20 and Q30 values for each sample were greater than 98.38% and 94.54%, respectively, indicating that the transcriptome data were highly reliable. A total of 123,576 unigenes were obtained after assembly and de-redundancy using Trinity software, with an average sequence length of 1328 bp; the length of N50 was 2083 bp, and the content of GC was 40.65% (Table 1). The unigenes obtained from the assembled *Akebia quinata* sequence were mainly distributed in 200–500 bp (Fig. 1), accounting for 69.23% of the total this indicated that the integrity of unigenes was good, and thus the unigenes could be functionally annotated.

Functional classification and expressed overview of unigenes

Among the 123,576 unigenes, 86,789, 72,122, 67,481, 69,642, 69,196, 64,795, and 69,624 unigenes were finally annotated by the seven functional databases NR, NT, Swiss-Prot, KEGG, KOG, PFAM, and GO, respectively. Moreover,

Table 1 Quality evaluation of unigenes from the transcriptome data of *Akebia quinata*

Sample	Raw reads (million)	Clean reads (million)	Total length (bp)	Average length (bp)	N50 (bp)	GC (%)
Flower 1(AF1)	48.49	36.94	49,185,670	1084	1700	41.52
Flower 2(AF2)	40.37	29.26	43,053,010	1029	1631	41.73
Flower 3(AF3)	42.59	32.28	44,337,417	1068	1684	41.55
Leaf 1(AL1)	44.20	34.68	56,776,139	1109	1723	41.09
Leaf 2(AL2)	37.47	27.91	49,544,476	1077	1684	41.47
Leaf 3(AL3)	42.63	32.38	53,271,956	1091	1698	41.26
Root 1(AR1)	43.70	32.79	61,894,373	1107	1763	40.73
Root 2(AR2)	51.90	31.81	56,666,475	943	1496	41.31
Root 3(AR3)	38.92	28.10	52,069,019	1014	1591	41.34
Stem 1(AS1)	36.68	27.63	53,955,299	1073	1674	41.19
Stem 2(AS2)	44.69	25.38	42,100,495	947	1425	41.53
Stem 3(AS3)	47.57	37.44	62,390,744	1144	1776	41.00

Fig. 1 Unigenes length distribution of *Akebia quinata* assembly
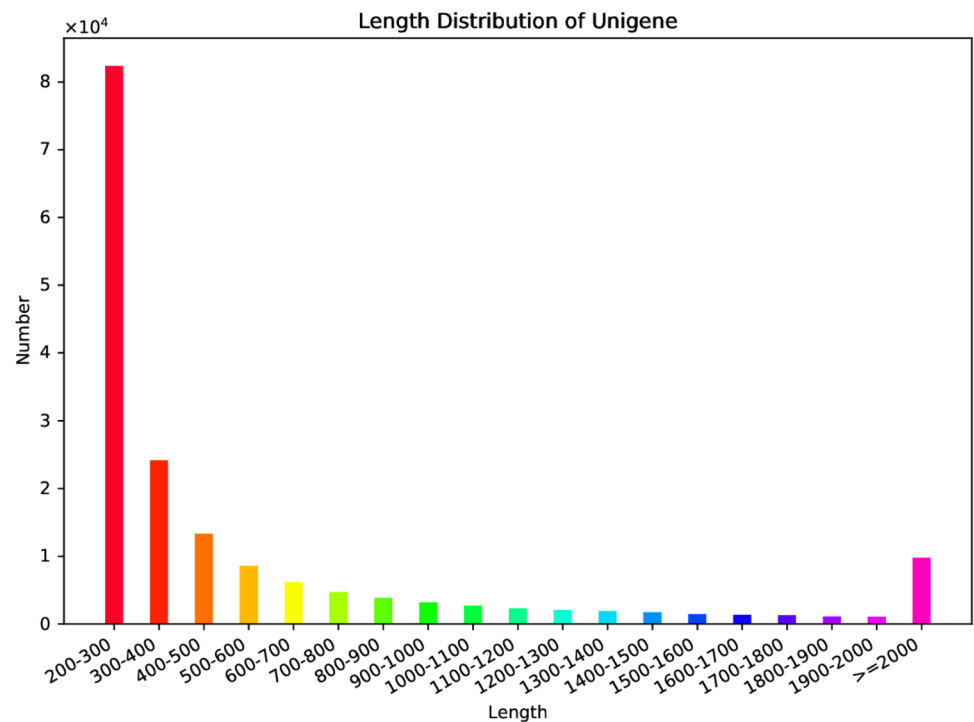
38,486 unigenes were annotated by all databases, accounting for 31.14%; 90,273 unigenes were annotated by any database, accounting for 73.05% (Table 2).

Data related to *Akebia quinata* were compared with those of the KOG database, and their functions were classified and annotated. The results showed that 69,196 unigenes were annotated to the KOG database. Furthermore, 13,666 unigenes were annotated in "general function prediction only", followed by 8304, 6027, 4407, and, 3428 unigenes annotated in "signal transduction mechanisms", "posttranslational modification, protein turnover, and chaperones", "transcription", and "carbohydrate transport and metabolism", respectively (Fig. 2). In addition, based on the 86,789 unigenes annotated in the NR database, the proportion of different species related to the annotated information was calculated; the species distribution map is shown in Fig. 3. The results showed that 28,939 (33.34%) unigenes were mapped to *Macleaya cordata* (Lin et al. 2018), which had the highest degree of similar sequence matching; moreover, 17,447
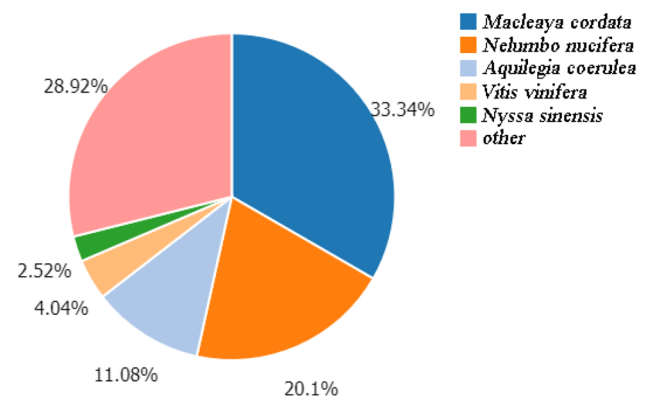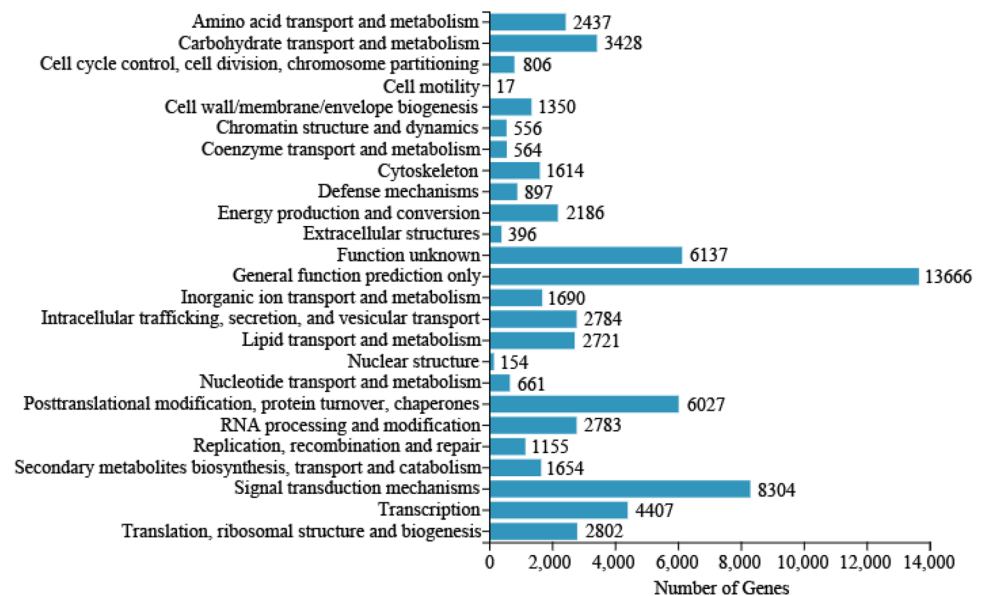

Fig. 3 Species comparison of homologous sequences of *Akebia quinata*

(20.10%), 9,614 (11.08%), 3,506 (4.04%), 2,186 (2.52%), and 25,097 (28.92%) unigenes were mapped to *Nelumbo nucifera* (Chen et al. 2019), *Aquilegia coerulea* (Singh et al. 2020), *Vitis vinifera* (Nassiri-Asl and Hosseinzadeh 2016), *Nyssa sinensis* (Yang and Ji 2017) and others, respectively. These species provide a reference for the sequence annotation of *Akebia quinata*.

The unigene results of all comparisons from the NR database were further annotated in the GO database, and statistically annotated to the classification map of three aspects of GO: biological process, cellular component, and molecular function (Fig. 4). According to the information in the GO database, 69,624 unigenes were annotated: in the biological process, 40,916 unigenes are mainly concentrated in "cellular process" and 32,079 unigenes are concentrated in "metabolic process"; in the cellular component, 56,303 unigenes focused on "cellular anatomical entity" and 28,792 unigenes

Table 2 Unigenes functional annotation results

Database	Annotated unigenes (bp)	Ratio (%)
NR	86,789	70.23
NT	72,122	58.36
Swiss-Prot	67,481	54.61
KEGG	69,642	56.36
KOG	69,196	55.99
PFAM	64,795	52.43
GO	69,624	56.34
Intersection	38,486	31.14
Overall	90,273	73.05

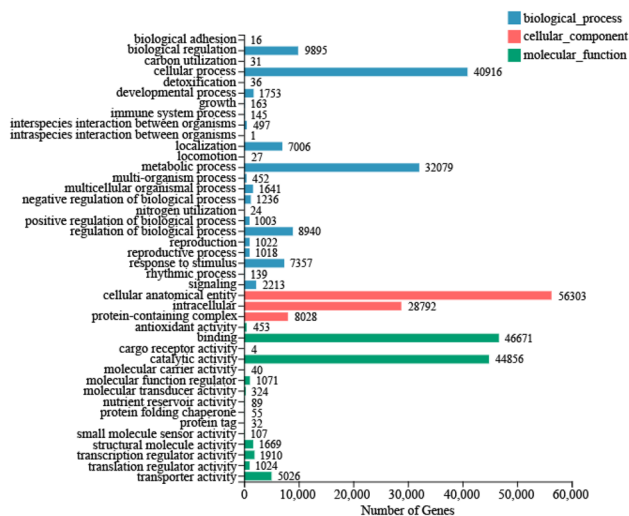**Fig. 2** KOG functional classification of unigenes of *Akebia quinata*

Fig. 4 Classification of GO function of unigenes of *Akebia quinata*

focused on "intracellular"; in the molecular function, 46,671 unigenes are mainly clustered in "binding" and 44,856 unigenes clustered in "catalytic activity".

In the transcriptome data of flowers, leaves, roots, and stems, there were 85,866, 90,701, 91,838, and 91,624 unigenes, respectively, of which 47,630, 45,680, 42,297 and 43,776 unigenes had FPKM ≤ 1 with low expression; 29,349, 34,707, 39,225 and 36,137 unigenes had FPKM = 1–10 with medium expression; and 8887, 10,314, 10,316 and 11,711 unigenes had FPKM ≥ 10 with high expression (Fig. 5A). According to the analysis of Graphpad Prism 8.0.2 and FPKM gene expression level, there were significant differences in gene expression levels between stems and flowers of *Akebia quinata*; there were no significant differences between stems and leaves and between stems and roots. The
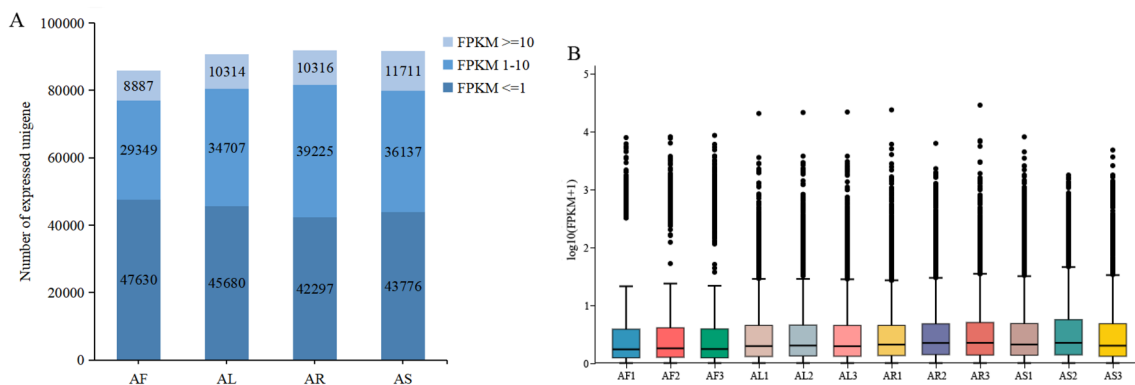
overall expression level of unigenes was highest in stems, followed by roots, leaves, and flowers (Fig. 5B).

Identification of genes related to phenylpropanoid biosynthesis by KEGG enrichment analysis

The gene sequence of *Akebia quinata* involved in the metabolic pathway was analyzed using the KEGG database. The data showed that 69,642 unigenes were annotated in the *Akebia quinata* transcriptome, including 3036 unigenes in cellular processes, mainly concentrated in transport and catabolism; 4255 unigenes in environmental information processing, mainly concentrated in signal transduction; 14,273 unigenes in genetic information processing, mainly in translation; 39,377 unigenes in metabolism, mainly concentrated in global and overview maps; and 2942 unigenes in organismal systems, mainly concentrated in environmental adaptation (Fig. 6). In the transcriptome data of *Akebia quinata*, 1805 unigenes were located in other pathways related to secondary metabolic biosynthesis. KEGG pathway enrichment results showed that 1174 unigenes were located in phenylpropanoid biosynthesis, the highest proportion of unigenes (Fig. 7).

By analyzing the transcriptome data of *Akebia quinata*, we found 13 enzymes involved in phenylpropanoid biosynthesis, including PAL, C4H, 4CL, HCT, C3'H, CCOAOMT, CCR, F5H, COMT, CAD, CHS, CHI, and F3H; of these, 45 and 27 unigenes encoded CCR and CAD, respectively (Table 3). The complete protein sequences of 13 key enzyme genes are shown in the appendix, as detailed in the supplementary material. Based on the screening conditions (FPKM of unigenes > 1 and distance from the complete coding frame < 50), 10, 5, 12, 8, 1, 5 unigenes were found to encode PAL, C4H, 4CL, HCT,

Fig. 5 Gene expression in four tissues of *Akebia quinata*. **A** The expression distribution of unigenes in four tissues (flowers, leaves, roots and stems). The *X*-axis represents the sample name, and the *Y*-axis represents the number of genes. Different gene expression levels are expressed by color shade: FPKM ≤ 1 is low expression,

FPKM 1–10 is medium expression, FPKM ≥ 10 is high expression. **B** The degree of dispersion of the data distribution can be observed by showing the unigenes expressed in the four organizations through the box chart. The *X*-axis are subtissue types of *Akebia quinata*, the *Y*-axis is log10 (FPKM + 1)

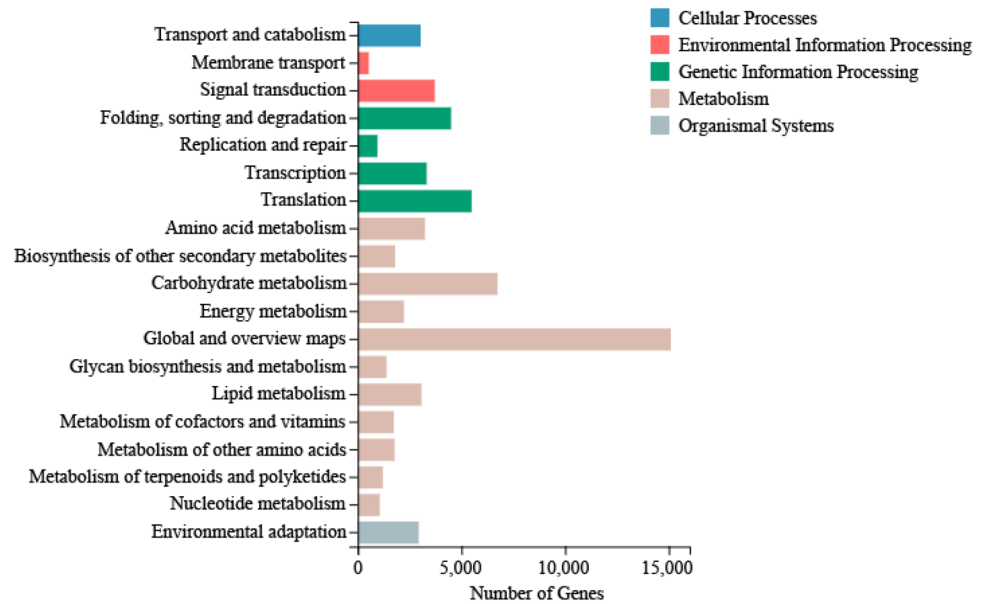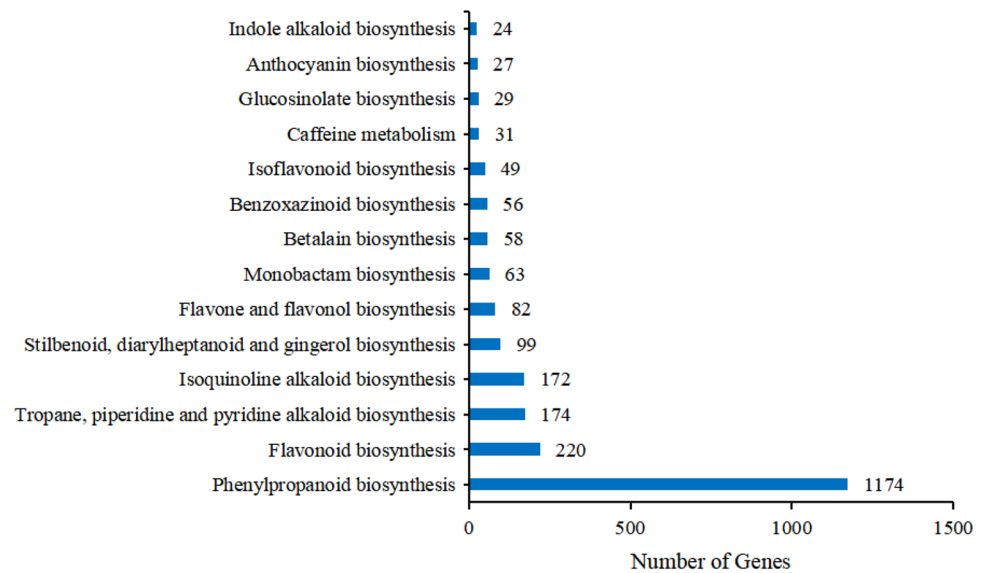**Fig. 6** KEGG functional classification of unigenes of *Akebia quinata*

Fig. 7 Classification of other secondary metabolic biosynthetic pathways in *Akebia quinata*

C3'H, CCOAOMT enzymes, respectively in the simple phenylpropanoid pathway; furthermore, 29, 2, 8, 13, 8, 7, 3 unigenes encoded CCR, F5H, COMT, CAD, CHS, CHI, F3H enzymes, respectively. The relative expression levels of these key enzyme genes in various tissues were shown in the form of heat map; in the phenylpropanoid biosynthesis pathway of *Akebia quinata*, C4H, CCR, and F3H were mainly highly expressed in flowers; PAL and CHI were mainly highly expressed in leaves; CHS was mainly highly expressed in roots; and 4CL, CCOAOMT, F5H, COMT, and C3'H were mainly highly expressed in stems (Fig. 8).

Expression analysis of structural characteristics of 4CL, a key enzyme

Based on the analysis of *Akebia quinata* transcriptome data, 12 unigenes encoding 4CL enzyme were found, and their nucleotide sequences were translated into protein sequences and compared with those of the NCBI Blast database; 10 unigenes were confirmed to have complete ORF. The self-sequences of the 10 unigenes were compared using DNAMAN, and the results showed a sequence identity of 72.97%, which indicated a high sequence homology of 4CL; in addition, the spatial structure
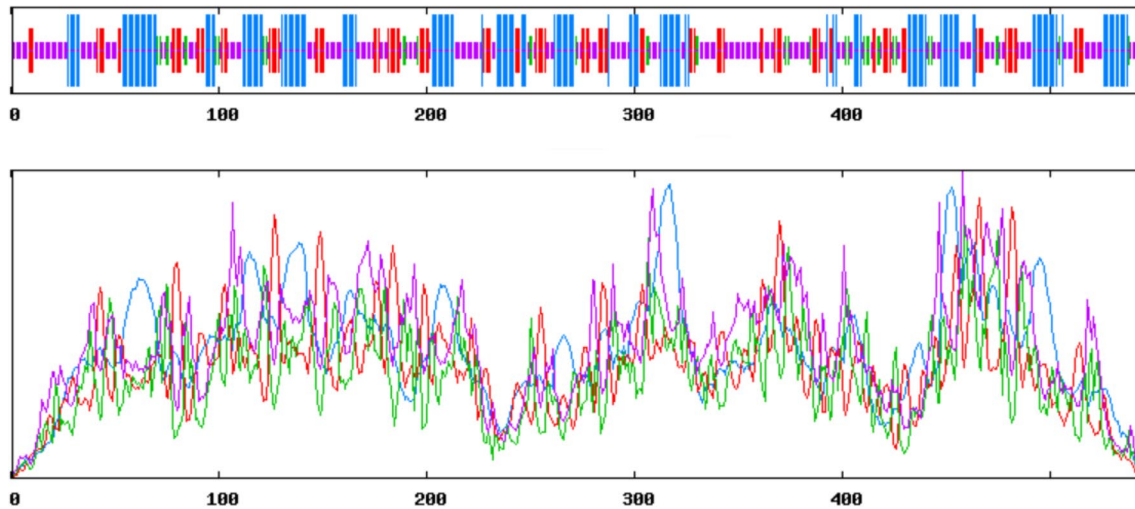
Table 3 Distribution of unigenes in four tissues during phenylpropanoid biosynthesis of *Akebia quinata*

Key enzyme	EC	Unigene number	No.in flower	No.in leaf	No.in root	No.in stem
PAL	4.3.1.24	14	13	13	11	13
C4H	1.14.14.91	6	6	5	5	4
4CL	6.2.1.12	16	15	15	15	16
HCT	2.3.1.133	14	13	14	13	13
C3'H	1.14.14.96	1	1	1	1	1
CCoAOMT	2.1.1.104	9	7	9	9	9
CCR	1.2.1.44	45	43	35	38	34
F5H	1.14.-.-	2	2	1	2	2
COMT	2.1.1.68	18	17	16	15	17
CAD	1.1.1.195	27	26	21	22	23
CHS	2.3.1.74	13	9	8	12	13
CHI	5.5.1.6	12	11	11	6	10
F3H	1.14.11.9	5	5	3	4	5

Fig. 8 Biosynthetic route of phenylpropanoid in *Akebia quinata*. F, flowers; L, leaves; R, roots; S, stems. The expression level of key enzyme genes, green means low expression of genes, on the contrary, high expression of red. Cluster of transcripts and Unigene are abbreviated as "CL" and "Un", respectively

similarity of unigenes constructed using SWISS MODEL was high. Among them, the CL 12789-2 unigene with the highest expression level of complete reading frame (length of ORF = 1629 bp) encoded 542 amino acids. Our team selected the 4CL (Un 64258, 76.27%) enzyme with the highest sequence homology to the template protein to describe its secondary and spatial structure. Secondary structure prediction showed that the 4CL enzyme consisted of 29.78%, 20.59%, 6.99%, and 42.65% alpha helix, extension chain, beta turn, and irregular crimp,

Fig. 9 Prediction of the secondary structure of 4CL protein. Blue: alpha helix; red: extension main chain; green: beta turn; purple: irregular crimp

respectively (Fig. 9). Based on the prediction of the spatial structure of the 4CL enzyme, two conserved peptide motifs were found in its amino acid sequence, including the conserved AMP binding domain (P187-V199, PYSS-GTTGLPKGV) and the conserved GEICIRG domain (386–392 amino acid sequence) (Fig. 10). The peptide motif BoxII, the conservative region of GEICIRG, is reported to be absolutely conserved in the 4CL enzyme, and its central C residue is considered to be directly involved in the catalytic process (Stuible and Kombrink 2001).

Analysis of DEGs

The results showed that the four tissues of *Akebia quinata* co-expressed 76,419 unigenes, and the specific expressions of flowers, leaves, roots, and stems were 10,048, 4350, 12,802, and 4451 unigenes, respectively (Fig. 11A). Using flowers as control, 36,494 DEGs were found in stems, with 18,102 and 18,392 up- and down-regulated genes, respectively. Using leaves as control, 11,285 DEGs were found in stems, including 5264 and 6021 up- and down-regulated genes, respectively. Moreover, 19,771 DEGs were found in stems, when roots were used as control, with 8225
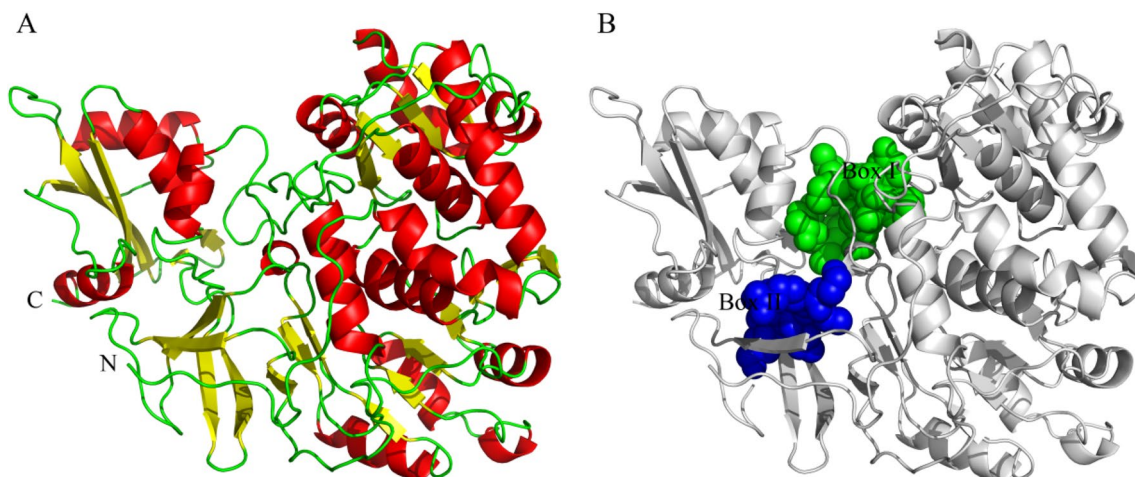

Fig. 10 4CL enzyme spatial structure model of *Akebia quinata*. **A** The spatial structure model of 4CL (Un 64258; template: 3a9u.1.A; sequence identity: 76.27%); **B** peptide motif Box I is conserved AMP binding domain, expressed by green sphere; peptide motif Box II is GEICIRG conservative domain, expressed by blue sphere
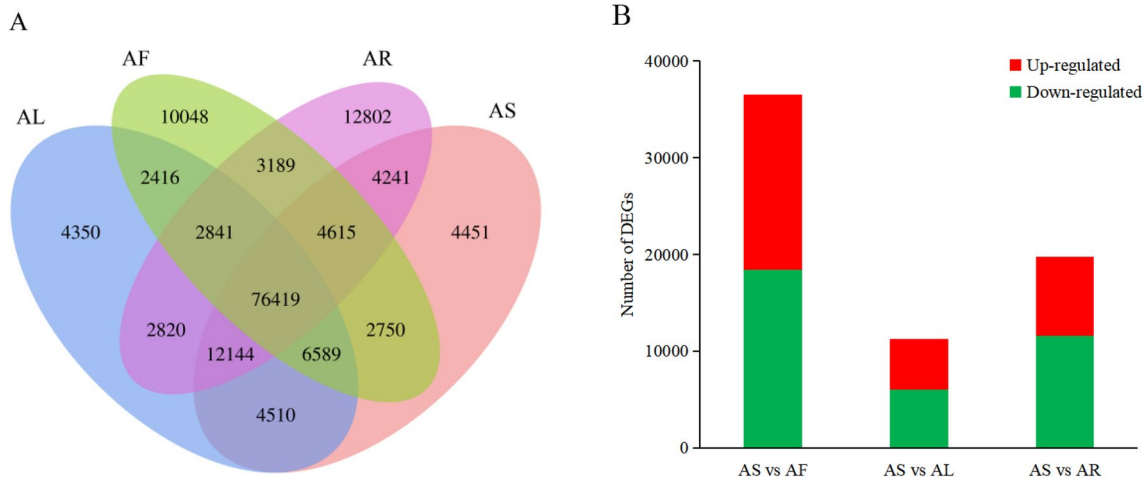
A

B

Fig. 11 Gene expression and number of DEGs in four tissues of *Akebia quinata*. **A** unigenes Venn diagram of four tissues; **B** number of DEGs: compared with flowers, leaves and roots, the number of DEGs with high or low expression levels in stems was defined as "up-regulated" and "down-regulated", respectively. a: Starch and sucrose metabolism; b: pyruvate metabolism; c: porphyrin and chlorophyll metabolism; d: photosynthesis—antenna proteins; e: phenylpropanoid biosynthesis; f: pentose and glucuronate interconversions; g: glycoly-

sis/gluconeogenesis; h: glycine, serine and threonine metabolism; i: glycerophospholipid metabolism; j: galactose metabolism; k: flavonoid biosynthesis; l: fatty acid metabolism; m: cysteine and methionine metabolism; n: cyanoamino acid metabolism; o: citrate cycle (TCA cycle); p: carbon metabolism; q: carbon fixation in photosynthetic organisms; r: biosynthesis of amino acids; s: amino sugar and nucleotide sugar metabolism; t: 2-oxocarboxylic acid metabolism

and 11,546 up- and down-regulated genes, respectively (Fig. 11B).

The results of KEGG enrichment analysis of DEGs in *Akebia quinata* showed that 8213 DEGs in flowers versus stems were annotated to 134 metabolic pathways; 3075 DEGs in leaves versus stems were annotated to 127

metabolic pathways; 4971 DEGs in roots versus stems were annotated to 128 metabolic pathways; and the first three metabolic pathways with more enriched genes were the same, including carbon metabolism, biosynthesis of amino acids, and starch and sucrose metabolism (Fig. 12A–C).
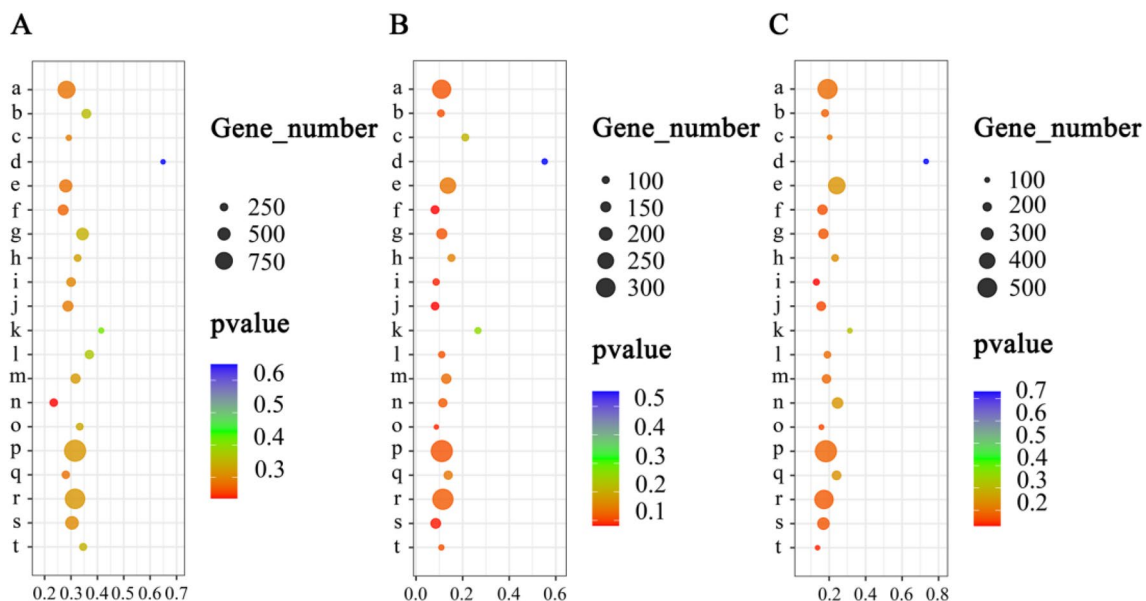

Fig. 12 KEGG pathway enrichment of differential genes. The ordinate is the first 20 metabolic pathways of enrichment of differential genes, and the abscissa represents the enrichment factors; **A–C** repre-

sent the significant enrichment pathways of the number of differential genes in stem compared with flowers, leaves and roots, respectively

In *Akebia quinata*, 506 DEGs in stems versus flowers were annotated to the phenylpropanoid biosynthesis pathway (ko00940), of which 250 DEGs were specifically up-regulated in stems; 247 DEGs in stems versus leaves were annotated to the phenylpropanoid biosynthesis pathway, of which 109 DEGs were specifically up-regulated in stems; and 436 DEGs in stems versus roots were annotated to the phenylpropanoid biosynthesis pathway, of which 185 DEGs were specifically up-regulated in stems.

Analysis of TFs related to phenylpropanoid biosynthesis

TFs have been shown to regulate various secondary metabolic processes. In the *Akebia quinata* transcriptome database, 30 unigenes involved in the phenylpropanoid biosynthesis pathway were identified, including 24, 2, 2, 1, and 1 unigenes belonging to the MYB, Tify, NAC, FHA, and GRAS families, respectively.

Validation of RNA-Seq data with qRT-PCR

In order to verify the differential gene expression and transcriptome data in different tissues, six unigenes involved in phenylpropanoid biosynthesis were selected for qRT-PCR analysis. The expression level of CL 12789-2 (4CL) and CL 14341-24 (CCR) were highest in flowers; similarly, the expression level of Un 17571 (COMT) was highest in leaves, whereas those of CL 4383-2 (COMT), Un 14954 (CAD), and CL 233-17 (PAL) were highest in stems. The qRT-PCR

results were consistent with those from the transcriptome data (Fig. 13).

Discussion

Phenylpropanoids, one of the most important bioactive components of *Akebia quinata,* have good pharmacological properties. In this study, we identified genes involved in *Akebia quinata* phenylpropanoid biosynthesis, thereby facilitating the understanding of phenylpropanoid biosynthesis and regulatory mechanism, which are important in promoting the comprehensive development and utilization of *Akebia quinata* plants.

Furthermore, we performed transcriptome sequencing on four tissues of *Akebia quinata* using the BGISEQ platform, and 56.49 Gb data were obtained. After reassembly and de-redundancy, 123,576 unigenes were derived. The total length, average length, N50, and GC content were 164,180,390 bp, 1328 bp, 2083 bp, and 40.65%, respectively. Subsequently, the obtained unigene was compared to those of seven functional databases NR, NT, SwissProt, KOG, KEGG, GO, and PFAM; finally, 86,789 (70.23%), 72,122 (58.36%), 67,481 (54.61%), 69,196 (55.99%), 69,642 (56.36%), 69,624 (56.34%), and 64,795 (52.43%) unigenes obtained functional annotations in the abovementioned databases, respectively. The obtained RNA-Seq data provide high-quality annotation assembly for the following analysis results, and the clean reads Q30 of each sample was greater
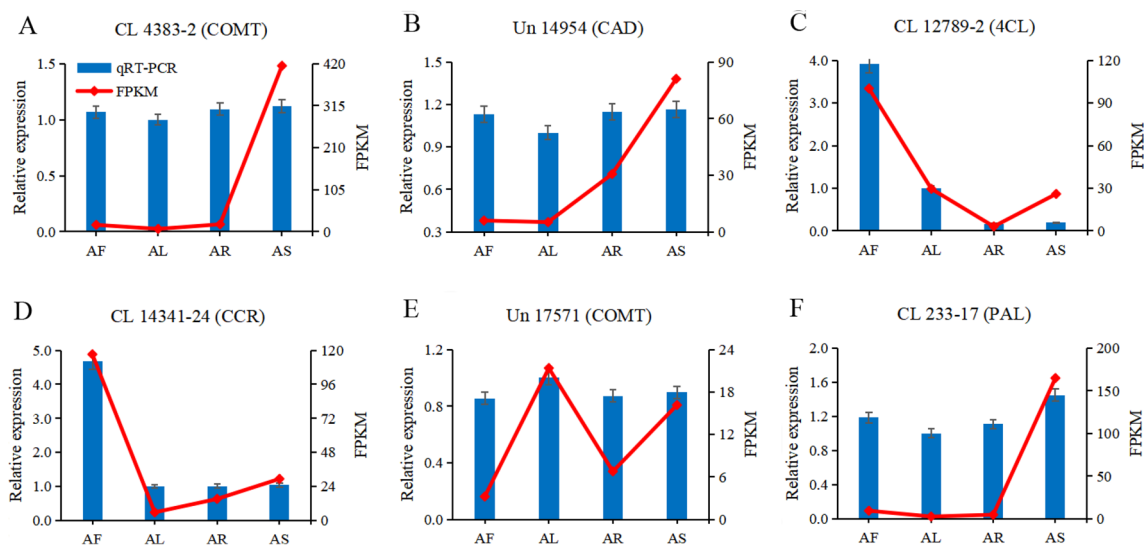

Fig. 13 Expression analysis of six unigenes encoding enzymes involved in phenylpropanoid biosynthesis. Relative expression of **A** CL 4383-2 (COMT), **B** Un 14954 (CAD), **C** CL 12789-2 (4CL), **D** CL 14341-24 (CCR), **E** Un 17571 (COMT), and **F** CL 233-17 (PAL) was analyzed by qRT-PCR using the actin gene (Un 4699) as the reference gene for normalization. Blue bars represent qRT-PCR data, red lines represent FPKM values. The data represents the average ± standard error of three repetitions. The left Y-axis is the relative expression level of genes detected by qRT-PCR, and the right Y-axis is the FPKM values of RNA-Seq data

than 94.54%, indicating that the *Akebia quinata* transcriptome data were reliable.

The annotation results of the NR database showed that *Akebia quinata* without reference genome had the highest genetic relationship with *Macleaya cordata* of Papaveraceae, and a large number (69,624) of unigenes were assigned and annotated to the GO database, which indicated that there was significant diversity in the transcriptome of different *Akebia quinata* tissues. A total of 69,642 unigenes were annotated in the KEGG database, and 2267 unigenes were involved in other secondary metabolic biosynthesis pathways, of which 1174 unigenes were enriched in the phenylpropanoid biosynthesis pathway. Our analysis showed that 13 key enzymes were involved in the phenylpropanoid biosynthesis pathway, and 5 key enzymes were mainly highly expressed in stems. Based on the expression levels of genes encoding key enzymes, we found that 4CL (CL 12789-2) and CCR (CL 14341-24) had the highest expression levels in flower tissue; COMT (Un 17571) had the highest expression level in leaf tissue; and COMT (CL 4383-2), CAD (Un 14954), and PAL (CL 233-17) had the highest expression level in stem tissue. Six genes—4CL (CL 12789-2), CCR (CL 14341-24), COMT (Un 17571), COMT (CL 4383-2), CAD (Un 14954), and PAL (CL 233-17)—were detected using the qRT-PCR method, and the authenticity and reliability of our transcriptional data were verified.

The 4CL gene exists in many plants, such as *Lycium chinense* (Zhang et al.) and *Arabidopsis thaliana* (Zhai et al. 2017). The ORF length of the 4CL gene of *Akebia quinata* is 1629 bp, which encodes 542 amino acids. Through BLAST comparative analysis, the amino acid encoded by this gene is closely related to that in *Thalictrum thalictroides* of Ranunculaceae. Based on the prediction and analysis of the secondary structure of *Akebia quinata* protein, it was found that the secondary structure of *Akebia quinata* 4CL protein was composed of 29.78%, 20.59%, 6.99%, and 42.65% alpha helix, extension chain, beta turn, and irregular crimp, respectively. The polypeptide chain of 4CL protein is further coiled or folded on the basis of various secondary structures to form a regular three-dimensional spatial structure. The 4CL amino acid sequence of *Akebia quinata*, akin to that of other plants, contains the AMP binding domain "PYSSGTTGLP-KGV" and the conserved sequence "GEICIRG" (Gao et al. 2020; Tian et al. 2017). The expression of the 4CL gene, which has a certain specificity in different tissues of plants, is highest in stems, followed by flowers, roots, and leaves; the relative expression of 4CL in roots, leaves, and flowers of *Osmanthus* sp. showed that flowers were larger than roots and leaves (Mu et al. 2016).

Gene expression level analysis showed a significant difference in gene expression between stems and flowers, and the overall expression level of unigene was highest in stems. Compared with leaves and roots, stems showed the highest number of up-regulated genes compared with flowers; hence, stems demonstrated higher gene expression levels than flowers. There were 250 up-regulated genes enriched in phenylpropanoid biosynthesis in DEGs of stems versus flowers, which was more than those in stems versus leaves and stems versus roots. These up-regulated genes may be more related to the higher value of traditional Chinese medicine in *Akebia quinata* stems.

Conclusion

We analyzed the transcriptome for four *Akebia quinata* tissues (flower, leaf, root, and stem), and identified many genes related to phenylpropanoid biosynthesis using RNA-Seq; moreover, we excavated the key enzymes involved in phenylpropanoid biosynthesis and analyzed the DEGs. Some key enzyme genes were verified using qRT-PCR, and the results were consistent with the RNA-Seq data. The study findings have deepened the understanding of the metabolic pathways related to *Akebia quinata* phenylpropanoid biosynthesis, and thus are helpful to further analyze the molecular mechanism of regulating *Akebia quinata* phenylpropanoid biosynthesis

Author contribution statement CQ, YY, and RW contributed to the sample collection and sequencing. CQ performed the transcriptome data analysis and wrote the manuscript. BH, DW and JO revised the manuscript. JO provided financial support. All authors read and approved the manuscript.

Data availability In this study, the RNA-Seq raw data reads have been deposited in NCBI Sequence Read Archive (SRA) database. Reviewer link: https://dataview.ncbi.nlm.nih.gov/object/PRJNA860796?reviewer=2gocus2ljundflnokhdusp9s6u.

Declarations

Conflict of interest The authors have declared that no competing interests exist.

References

Ablise M, Cartier A, Siest G, Visvikis S, Loppinet V (2002) Molecular pharmacophore determination of lipid lowering drugs with the receptor mapping method. Mini Rev Med Chem 2(2):97–102

Chen CH, Song TY, Liang YC, Hu ML (2009) Acteoside and 6-O-acetylacteoside downregulate cell adhesion molecules induced by IL-1beta through inhibition of ERK and JNK in human vascular endothelial cells. J Agric Food Chem 57(19):8852–8859

Chen GL, Zhu MZ, Guo MQ (2019) Research advances in traditional and modern use of *Nelumbo nucifera*: phytochemicals, health promoting activities and beyond. Crit Rev Food Sci Nutr 59(sup1):S189–S209

Chu HB, Tan NH, Peng CS (2009) Progress in research on Pedicularis plants. China J Chin Mater Med 34(19):2536–2546

Fu LG (2000) Chinese higher plants. Qingdao Publishing House, Qingdao, p 586

Gao W, Xin WN, Ao WC, Xu LC, Gong C, Peng YY (2015) Phenylethanoid glycosides B content of *Akebia* plants determined by HPLC. South China for Sci 43(05):48–51

Gao RG, Li H, Xiang QZ, Li M (2020) Cloning and protein expression analysis of 4-coumarate coenzyme A ligase gene *Cs4CL1* in tea tree. Shandong Agric Sci 52(03):8–12

Gu XW, Chen Y, Wang M, Dong YF, Feng X, Liang JY (2012) Progress in the research on the phenylpropanoids and its related chemical constituents of Malvaceae. Chin Wild Plant Resour 31(01):17–21

Jiang X, Lu LM, Zhang CL, Wang XY, Sun Q, Wang JT, Yuan GJ (2017) Antibacterial phenylethanoid glycosides from fruits of *Forsythia suspensa*. Mod Chin Med 19(05):642–647

Julião Lde S, Piccinelli AL, Marzocco S, Leitão SG, Lotti C, Autore G, Rastrelli L (2009) Phenylethanoid glycosides from *Lantana fucata* with in vitro anti-inflammatory activity. J Nat Prod 72(08):1424–1428

Kim CS, Winn MD, Sachdeva V, Jordan KE (2017) K-mer clustering algorithm using a MapReduce framework: application to the parallelization of the Inchworm module of trinity. BMC Bioinform 18(01):467

Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. Nat Methods 9(04):357–359

Li L (2018) Study on the resource and quality evaluation of *Akebia quinata* in Anhui province. Anhui University of Chinese Medicine, Hefei

Li B, Dewey CN (2011) RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinform 12:323

Li L, Chen XZ, Yao XH, Tian H, Huang HW (2010) Geographic distribution and resource status of three important Akebia species. J Wuhan Bot Res 28(04):497–506

Li LX, Chen DL, Zhou XJ (2020) Research progress on chemical constituents, pharmacological activity and Q-marker analysis and prediction of *Scrophularia ningpoensis* Hemsl. Chin Tradit Patent Med 42(09):2417–2426

Lin L, Liu YC, Huang JL, Liu XB, Qing ZX, Zeng JG, Liu ZY (2018) Medicinal plants of the genus *Macleaya* (*Macleaya cordata*, *Macleaya microcarpa*): a review of their phytochemistry, pharmacology, and toxicology. Phytother Res 32(01):19–48

Liu H, Wang DQ (2011) Research the medicinal plant resources of *Akebia* plants in Anhui Province. Mod Chin Med 13(09):9–11

Liu YT, Hou XJ, Xie Y, Feng YL, Ou YS, Yang SL (2012) Overview of research on chemical constituents and pharmacological activities of *Akebia* Dence. J Jiangxi Univ Chin Med 24(04):87–93

Liu YC, Li HL, Zeng XH, Wang XB (2020) Research progress on chemical components and pharmacological effects of *Akebia trifoliata* (Thunb.) koidz. stem. Acad J Shanghai Univ Tradit Chin Med 34(03):99–106

Livak KJ, Schmittgen TD (2001) Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) method. Methods 25(04):402–408

Mistry J, Finn RD, Eddy SR, Bateman A, Punta M (2013) Challenges in homology search: HMMER3 and convergent evolution of coiled-coil regions. Nucleic Acids Res 41(12):e121

Mu HN, Sun TZ, Xu C, Li HG, Yang XL, Wang LG (2016) Gene clone and expression analysis of 4-coumarate-CoA ligase in sweet Osmanthus (*Osmanthus fragrans* Lour.). Mol Plant Breed 14(03):536–541

Nassiri-Asl M, Hosseinzadeh H (2016) Review of the pharmacological effects of *Vitis vinifera* (grape) and its bioactive constituents: an update. Phytother Res 30(09):1392–1403

Ou JM, Chu XQ, Zhang H (2015) Determination of three active components of Akebia from Anhui Province, China, by high-performance liquid chromatography. J Anhui Univ Chin Med 34(01):70–73

Ou JM, Yang X, Shan CM, Zhang SX, Shi YY, Wu JW, Huang LQ, Wang R (2020) Transcriptome analysis of "*Langmei*" fruits and key enzyme genes structure and function prediction involved in citric acid biosynthesis. China J Chin Mater Med 45(19):4606–4616

Peng SP, Dong CM, Zhu YH (2021) Screening and identification of R2R3-MYB transcription factor related to lignan synthesis in *Forsythia suspensa*. J Chin Med Mater 44(07):1590–1596

Pertea G, Huang XQ, Liang F, Antonescu V, Sultana R, Karamycheva S, Lee Y, White J, Cheung F, Parvizi B, Tsai J, Quackenbush J (2003) TIGR Gene Indices clustering tools (TGICL): a software system for fast clustering of large EST datasets. Bioinformatics 19(05):651

Pirooznia M, Perkins EJ, Deng Y (2008) Batch Blast Extractor: an automated blastx parser application. BMC Genom 9(Suppl 2):S10

Qian CC, Zhao LQ, Yang YT, Han BX, Wu JW, Ou JM (2022) Analysis of the transcriptome and discovery of key enzyme genes of the triterpenoid saponin biosynthesis pathway in *Akebia trifoliata* (Thunb.) Koidz. Plant Sci J 40(03):378–389

Rice P, Longden I, Bleasby A (2000) EMBOSS: the European molecular biology open software suite. Trends Genet 16(06):276–277

Shan CM, Wang CK, Shi YY, Zhang SX, Zhao LQ, Wu JW (2020a) Identification of key enzyme genes involved in biosynthesis of steroidal saponins and analysis of biosynthesis pathway in *Polygonatum cyrtonema*. China J Chin Mater Med 45(12):2847–2857

Shan CM, Wang CK, Zhang SX, Shi YY, Ma KL, Yang QS, Wu JW (2020b) Transcriptome analysis of *Clinopodium gracile* (Benth.) Matsum and identification of genes related to Triterpenoid Saponin biosynthesis. BMC Genom 21(01):49

Shi YY, Zhang SX, Peng DY, Wang CK, Zhao DR, Ma KL, Wu JW, Huang LQ (2019) Transcriptome analysis of *Clinopodium chinense* (Benth.) O. Kuntze and identification of genes involved in triterpenoid saponin biosynthesis. Int J Mol Sci 20(11):2643

Singh S, Bhatt V, Kumar V, Kumawat S, Khatri P, Singla P, Shivaraj SM, Nadaf A, Deshmukh R, Sharma TR, Sonah H (2020) Evolutionary understanding of aquaporin transport system in the basal eudicot model species *Aquilegia coerulea*. Plants (Basel) 9(06):799

Stuible HP, Kombrink E (2001) Identification of the substrate specificity-conferring amino acid residues of 4-coumarate: coenzyme A ligase allows the rational design of mutant enzymes with new catalytic properties. J Biol Chem 276(29):26893–26897

Tian XM, Yan LH, Xiang GF, Jiang LY (2017) Research progress on 4-coumarate: coenzyme A ligase (4CL) in plants. Biotechnol Bull 33(04):19–26

Wang J, Zhou ZY, Xu QL, Tan JW (2014) Phenylpropanoids from the stems of *Akebia trifoliata*. J Trop Subtrop Bot 22(05):511–515

Wang FQ, Yang X, Zuo X, Miao CY, Zhang ZY (2022) Full-length transcriptome sequence and identification of genes involved in phenylethanol glycoside biosynthesis in *Rehmannia glutinosa*. Acta Pharm Sin 57(03):831–838

Yang ZY, Ji YH (2017) Comparative and phylogenetic analyses of the complete chloroplast genomes of three Arcto-Tertiary relicts: *Camptotheca acuminata*, *Davidia involucrata*, and *Nyssa sinensis*. Front Plant Sci 8:1536

Yang YN, Liu HC, Gao EG, Li YY, Chen SS, Xin XH (2022) Phenological investigation and variation analysis of yield and quality of *Akebia* Decne. germplasm resources. J South Agric 53(01):134–145

Zhai Y, Liu QS, Shi DW, Wu YK, Zhang NQ, Qi QS (2017) cloning and sequence analysis of a 4-coumaryl: CoA ligase gene from *Arabidopsis thaliana*. Biol Chem Eng 3(03):1–8

Zhang KM, Geng GG, Qiao F (2022) 4CL Activity and gene expression during fruit development of *Lycium chinense*. Mol Plant Breed. http://kns.cnki.net/kcms/detail/46.1068.S.20220328.1531.014.html

Zhao LQ, Shan CM, Zhang SX, Shi YY, Ma KL, Wu JW (2020) Identification of key enzyme genes involved in anthocyanin synthesis pathway in *Clinopodium gracile* by transcriptome analysis. Bull Bot Res 40(06):886–896

Zheng Y, Jin XF (2021) The research progress on development and utilization of family Lardizabalaceae plant resources. Chin Wild Plant Resour 40(10):83–108

Zhu F, Cai YZ, Sun M, Ke J, Lu D, Corke H (2009) Comparison of major phenolic constituents and in vitro antioxidant activity of diverse Kudingcha genotypes from *Ilex kudingcha*, Ilex cornuta, and Ligustrum robustum. J Agric Food Chem 57(14):6082–6089