Check for updates

The gastrodin biosynthetic pathway in *Pholidota chinensis* Lindl. revealed by transcriptome and metabolome profiling

Baocai Liu[1,2,3,4,5], Jingying Chen[2]*, Wujun Zhang[2],
Yingzhen Huang[2], Yunqing Zhao[2], Seifu Juneidi[6],
Aman Dekebo[7,8], Meijuan Wang[9], Le Shi[1,3,4,5]
and Xuebo Hu[1,3,4,5]*

[1]Institute for Medicinal Plants, College of Plant Science and Technology, Huazhong Agricultural University, Wuhan, China, [2]Institute of Agricultural Bioresource, Fujian Academy of Agricultural Sciences, Fuzhou, China, [3]Innovation Academy of International Traditional Chinese Medicinal Materials, Huazhong Agricultural University, Wuhan, China, [4]National-Regional Joint Engineering Research Center in Hubei for Medicinal Plant Breeding and Cultivation, Huazhong Agricultural University, Wuhan, China, [5]Medicinal Plant Engineering Research Center of Hubei Province, Huazhong Agricultural University, Wuhan, China, [6]Department of Applied Biology, School of Natural Science, Adama Science and Technology University, Adama, Ethiopia, [7]Applied Chemistry Department, School of Applied Natural Sciences, Adama Science and Technology University, Adama, Ethiopia, [8]Institute of Pharmaceutical Sciences, Adama Science and Technology University, Adama, Ethiopia, [9]Shengnongjia Academy of Forestry, Shengnongjia, Hubei, China

Pholidota chinensis Lindl. is an epiphytic or lithophytic perennial herb of Orchidaceae family used as a garden flower or medicinal plant to treat high blood pressure, dizziness and headache in traditional Chinese medicine. Gastrodin (GAS) is considered as a main bioactive ingredient of this herb but the biosynthetic pathway remains unclear in *P. chinensis*. To elucidate the GAS biosynthesis and identify the related genes in *P. chinensis*, a comprehensive analysis of transcriptome and metabolome of roots, rhizomes, pseudobulbs and leaves were performed by using PacBio SMART, Illumina Hiseq and Ultra Performance Liquid Chromatography Tandem Mass Spectrometry (UPLC-MS/MS). A total of 1,156 metabolites were identified by UPLC-MS/MS, of which 345 differential metabolites were mainly enriched in phenylpropanoid/phenylalanine, flavone and flavonol biosynthesis. The pseudobulbs make up nearly half of the fresh weight of the whole plant, and the GAS content in the pseudobulbs was also the highest in four tissues. Up to 23,105 Unigenes were obtained and 22,029 transcripts were annotated in the transcriptome analysis. Compared to roots, 7,787, 8,376 and 9,146 differentially expressed genes (DEGs) were identified in rhizomes, pseudobulbs and leaves, respectively. And in total, 80 Unigenes encoding eight key enzymes for GAS biosynthesis, were identified. Particularly, glycosyltransferase, the key enzyme of the last step in the GAS biosynthetic pathway had 39 Unigenes candidates, of which, transcript28360/f2p0/1592, was putatively identified as the most likely candidate based on analysis of co-expression, phylogenetic analysis, and

homologous searching. The metabolomics and transcriptomics of pseudobulbs versus roots showed that 8,376 DEGs and 345 DEMs had a substantial association based on the Pearson's correlation. This study notably enriched the metabolomic and transcriptomic data of *P. chinensis*, and it provides valuable information for GAS biosynthesis in the plant.

Introduction

Pholidota chinensis Lindl, a member of the Orchidaceae family, is commonly known as "Shi Xian Tao" in China (Figures 1A–G). It is an epiphytic or lithophytic perennial herb widely distributed in southern China (Want et al., 2010; Dunn et al., 2011). The whole plants or pseudobulbs are used as ornamental flowers or folklore medicine in treating high blood pressure, dizziness and headache (Medincine, 2006). It is also orally administered in treating cough, tuberculosis, scrofula, diuresis, and infantile malnutrition as a traditional medicine by the Maonan tribal minorities in Guangxi province of China (Hong et al., 2015). Researchers have shown that polysaccharides, stilbenoids, dihydrophenanthrenes and triterpenoids are the main bioactive components in *P. chinensis* (Yao et al., 2008). These compounds exhibited multiple therapeutic activities including anti-tumor (Luo et al., 2018), anti-oxidant (Yang et al., 2016), anti-bacterial (Ti et al.,

2020), anti-diabetic (Ren et al., 2020), anti-inflammatory (Wang et al., 2006), anti-pain and inhibit central nervous system (Liu et al., 2007; Rueda et al., 2014; Wang et al., 2016). Extensive chemical and pharmacological studies have laid a solid foundation for further application of these ingredients as medicine (Yang et al., 2016; Ti et al., 2020).

A Chinese patented medicine "Toutongding Syrup" is made up of *P. chinensis* for treating neurological headaches and concussion sequelae, and the gastrodin (4-hydroxymethylphenyl-β-Dglucopyranoside, GAS) and gastrodigenin (4-hydroxybenzyl alcohol, HBA) were shown to be the primary active ingredients (Weng, 2006; Zou et al., 2017). According to the established high performance liquid chromatography (HPLC) fingerprinting of *P. chinensis*, GAS was one of its analytical markers and its content in *P. chinensis* was higher than another traditional Chinese medicine *Gastrodia elata* Blume. (Zhang et al., 2019; Zhang et al., 2020). *G. elata* is the major source of GAS and HBA that is widely used to treat
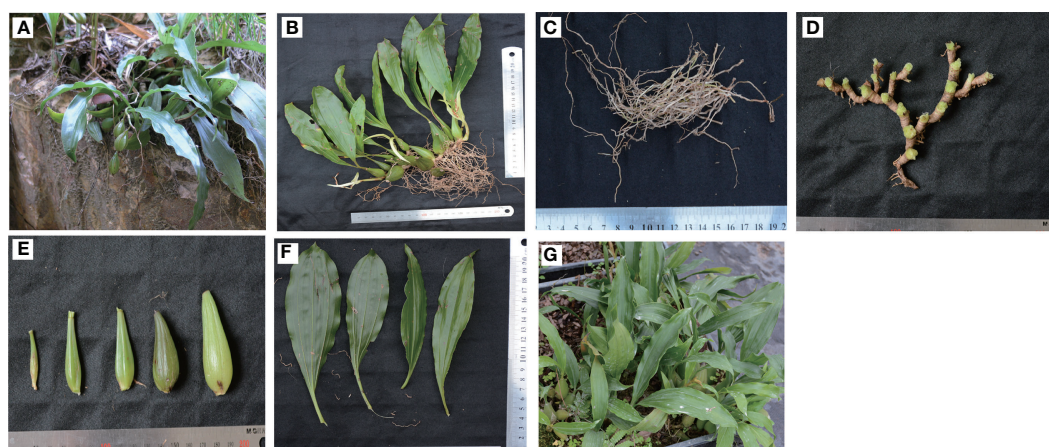
FIGURE 1
The morphological characteristics and growing environment of *P. chinensis*. **(A)**, Wild growth environment and plants growing on stone surfaces; **(B)**, the whole plant; **(C)**, Roots (designated as B1 for the rest of the sample analysis); **(D)**, Rhizomes (designated as B2 for the rest of the sample analysis); **(E)**, pseudobulbs (designated as B3 for the rest of the sample analysis); and **(F)** leaves (designated as B4 for the rest of the sample analysis) were analyzed; **(G)**, Artificially cultivated plants in a garden.

neurological disorders for centuries in China (Yuan et al., 2018; Zhang et al., 2019; Bae et al., 2022).

The GAS, a phenolic glycoside, is widely used to treat sedative, hypnotic, anticonvulsive and neuroprotective diseases in clinics (Liu and Yang, 2022). Synthesis of GAS is accomplished through glycosylation by a glycosyltransferase (GT) which transforms HBA with different glucose donors (Bai et al., 2016). Toluene was considered as the biosynthetic precursor for HBA that catalyzed by monooxygenase of cytochrome P450 (Tsai et al., 2016). The biosynthetic pathway of phenolic components, including GAS and HBA, were synthesized from phenylalanine through the phenylpropanoid pathway, which was speculated in *G. elata* by transcriptome analysis (Shan et al., 2021), and the biosynthetic pathway of 4-hydroxylbenzaldehyde and vanillin had been well studied in *Vanilla* spp. (Gallage et al., 2014; Wang et al., 2018). Interestingly, whether the precursor is toluene or phenylalanine, the last step is a GT that catalyzes the conversion of HBA to GAS in the GAS biosynthetic pathway (Tsai et al., 2016; Yin et al., 2020). However, the full, native biosynthetic pathway of GAS in *P. chinensis* has still not yet been documented.

To date, the transcriptome and metabolome studies provide effective strategies for understanding the molecular mechanisms of active ingredient formation (Hu et al., 2021; Chen et al., 2022). The combination of transcriptome and metabolome makes it possible to identify genes in any complex biological process with high sensitivity and accuracy (Song et al., 2022). The next-generation sequencing merges short reads into longer fragments by computation and it unavoidably affects the accuracy and integrity in fragments assembly (Cheng et al., 2021). In contrast, the third-generation sequencing technology has an advantage of sequencing reads as long as 100 kb but with lower sequencing accuracy (Liu et al., 2022). Therefore, the combination of next-generation sequencing and third-generation sequencing may assist to make up the shortcomings of each sequencing tool.

In the present study, based on multi-omics comparison, the GAS biosynthesis pathway and the genes involved in *P. chinensis* were elucidated. To the best of our knowledge, this study is the first to dissect the genes for GAS biosynthesis in *P. chinensis* and the same genus.

Materials and methods

Plant materials

P. chinensis was collected in June 2018 from Lingxia Village, Dongzhang Town, Fuqing City, Fujian Province of China (with 25°41.221′ N; 119°08.358′E and altitude 259 m). The plant sample was authenticated by Prof. Xuebo Hu (College of Plant Science and Technology, Huazhong Agricultural University, Wuhan, China), and Prof. Jingying Chen (Institute of

Agricultural Bioresource Fujian Academy of Agricultural Sciences Fuzhou, China). The samples were collected from a wild forest (Figures 1A, B), and the plant part subjected to study was immediately separated into roots (B1, Figure 1C), rhizomes (B2, Figure 1D), pseudobulbs (B3, Figure 1E) and leaves (B4, Figure 1F). The samples with six independent biological replicates were washed clean, surface dried, and flash-frozen in liquid nitrogen, and then stored at -80°C until chemical composition analysis and RNA extraction. The rest plants were relocated to a greenhouse (Figure 1G).

UPLC-MS/MS conditions

Liquid chromatography-mass spectrometry (LC-MS) was used to analyze phytochemical constitutes *of P. chinensi*s. The fresh materials of roots, rhizomes, pseudobulbs and leaves (0.1 g) was ground and extracted with 0.5 ml 80% (v/v) MeOH (LC-MS Grade, Thermo Fisher, USA). Samples were sonicated with a Vortex (Kylin-Bell, Jiangsu, China) and centrifuged for 20 min at 15, 000 g. The obtained supernatant was filtered through a 0.22 μm organic nylon needle filter (SCAA-104, ANPEL, Shanghai, China) and stored in a sample bottle (Want et al., 2010; Dunn et al., 2011). The extraction was performed in 6 replicates for statistical analysis. The metabolites were extracted and identified by the Novogene Bioinformatics Technology Co., Ltd.

Metabolite identification and quantification

The raw data of the mass spectrometry detection were imported into Compound Discoverer 3.1 (CD) software (Hao et al., 2018), used for extraction of metabolite feature. The characteristics of metabolites were obtained based on simple screening of data with their retention time, mass-to-charge ratio and peak alignment, molecular weight of the compound, and the mass deviation and adduct ion information. By matching fragment ions, collision energy and other information of each compound in the mzCloud database, the metabolites in the biological system were identified. Then, the QC Compounds with a CV (Coefficient of Variance) value less than 30% (Dai et al., 2017) were selected and used for final identification. Data quality control was performed to ensure the accuracy and reliability of the data. These metabolites were annotated using the Kyoto encyclopedia of genes and genomes (KEGG) database (http://www.genome.jp/kegg/), human metabolome database (HMDB) (http://www.hmdb.ca/) and Lipidmaps database (http://www.lipidmaps.org/). Principal components analysis (PCA) and Partial least squares discriminant analysis (PLS-DA) were performed with metaX (Wen et al., 2017). Metabolites with significant differences in content were identified according to the thresholds of variable importance

in projection (VIP) >1, fold change >2 or <0.5 and P value <0.05. Hierarchical clustering (HCA) and metabolite correlation analysis to reveal the relationship among the samples and metabolites (Chen et al., 2015). The metabolic pathway enrichment of differential metabolites (DEMs) was performed, when the ratio x/n > y/N (x, number of differential metabolites associated with this pathway; y, number of background (all) metabolites associated with this pathway; n: number of differential metabolites annotated by KEGG; N, number of KEGG-annotated background (all) metabolites), metabolic pathway was considered as enriched. When the P-value of metabolic pathway < 0.05, metabolic pathway was considered as statistically significant enrichment.

HPLC analyses of GAS and HBA

The major constituents of *P. chinensi*s, GAS and HBA, were analyzed by HPLC system. GAS and HBA were extracted from dried and fresh *P. chinensi*s tissues (roots and rhizomes, pseudobulbs as well as leaves) and measured, as described previously (Zhang et al., 2019), with slight modifications. Briefly, dried (0.5 g) and fresh powder of each tissue was extracted in 25 mL of 50% (v/v) methanol by ultrasonication for 30 min. Using the following chromatographic conditions, injection volume, 10 uL; column, Agilent SB-aq (5 μm, 4.6 mm × 250 mm); temperature, 30℃; flow rate, 1.0 mL min^{-1}; detector and UV-VIS detector at 220 nm. The mobile phases were containing 99.9% acetonitrile (A) and 0.05% phosphoric acid (B).

RNA extraction and Illumina sequencing

Frozen tissues were transferred to a mortar pre-cooled by liquid nitrogen and ground with a pestle. Total RNA was extracted from roots, rhizomes, pseudobulbs and leaves (4 tissues× 3 biological replications) by using the RNAprep Pure Plant Kit 264 (Tiangen, Beijing, China), following the manufacturer's instructions. The quality and quantity of RNA was checked by agarose gel electrophoresis and spectrophotometry (IMPLEN, CA, USA) and Agilent Bioanalyzer 2100 system (Agilent Technologies, CA, USA), respectively. The RNA samples with A260/A280 of 1.8-2.2 were selected for cDNA synthesis.

An Illumina Hiseq platform was conducted using the NGS sequencing. The sequencing libraries were generated using NEBNext[®] Ultra[™] RNA Library Prep Kit for Illumina[®] (NEB, USA) following manufacturer's recommendations, and index codes were added to attribute sequences to each sample. The RNA-seq experiment was performed at Novogene Bioinformatics Technology Co., Ltd. The raw data were uploaded to Sequence Read Archive (http://www.ncbi.nlm.-nih.gov/) as accession PRJNA841044.

RNA extraction and PacBio ISO-Seq

To obtain a complete information of all transcripts, the full-length transcriptome sequencing was adopted in the present study. In order to reduce experimental error, the best RNA sample of three replicates was selected from each sample used in Illumina sequencing, and then mixed together in an equal quantity, as one sample, for SMRT sequencing. The Iso-Seq library was prepared according to the Isoform Sequencing protocol (Iso-Seq) using the Clontech SMARTer PCR cDNA Synthesis Kit and the BluePippin Size Selection System protocol as described by Pacific Biosciences (PN 100-092-800-03). The generated cDNA was re-amplified by PCR. A Qubit fluorometer (Life Technologies, Carlsbad, CA, USA) was used to determine fragment size distribution. The quality of the libraries was assessed using the Agilent Bioanalyzer 2100 system. The SMRT sequencing was performed using the Pacific Biosciences' real time sequencer using C2 sequencing reagents. The RNA-seq experiment was performed at Novogene Bioinformatics Technology Co., Ltd. The raw data were deposited to Sequence Read Archive (http://www.ncbi.nlm.-nih.gov/) with accession. PRJNA806713.

The sequence data were processed using the SMRTlink 5.0 software (https://www.pacb.com/support/software-downloads/). Circular consensus sequence (CCS) was generated from subread BAM files parameters: min_length 200, max_drop_fraction 0.8, no_polish TRUE, min_zscore -9999, min_passes 1, min_predicted_accuracy 0.8, max_length 18000. The CCS.BAM files were output, which were then classified into full length and non-full length reads using pbclassify.py script, ignore polyA false, minSeq Length 200. Non-full length and full-length fasta files produced were then fed into the cluster step, which does isoform-level clustering, followed by final Arrow polishing, hq_quiver_min_accuracy 0.99, bin_by_primer false, bin_size_kb 1, qv_trim_5p 100, qv_trim_3p 30. Additional nucleotide errors in consensus reads were corrected using the Illumina RNA- seq data with the software LoRDEC (Salmela and Rivals, 2014). After all redundancy corrected, the consensus reads were removed by CD-HIT (Fu et al., 2012), and the final consensus isoforms were obtained for the subsequent analysis.

Functional annotation

Final consensus isoforms were searched used diamond v0.8.36 software against NCBI non-redundant (Nr), Swiss-Prot, euKaryotic Ortholog Groups (KOG)/Cluster of Orthologous Groups and Kyoto Encyclopedia of Genes and Genomes (KEGG) databases with an E value threshold of $1e^{-5}$. The BLAST software with E-value ≤1e−5 was used for NT database analysis. The Hmmscan procedure was used in the Pfam database, and GO function categories were performed by

Blast2 GO v2.5 based on Pfam annotation. We use the confidence protein sequences of R. ferrugineus or closely related species for ANGLE training, and then run the ANGLE predictions for given sequences (Shimzu et al., 2006). Transcription factors (TF) were performed by the iTAK software (Zheng et al., 2016). Coding Potential Calculator (CPC) (Kang et al., 2017), and Pfam-scan (Finn et al., 2016) to predict the coding potential of transcripts.

RNA-seq read mapping and expression analysis

The consensus after de-redundancy correction was used the reference sequence (ref), and the clean reads of each sample obtained by Illumina sequencing were aligned to the ref using RSEM software (Li and Dewey, 2011). Further, RSEM software was used to count the comparison results of bowtie2, obtained the read count value of each sample compared to each gene, performed reads per kilo base of transcript per million mapped reads (FPKM) normalization, and then analyzed the expression level of the gene.

Differential expression analysis

Differential expression analysis of two conditions/groups was performed using the DESeq R package (Love et al., 2014). The DESeq provide statistical routines for determining differential expression in digital gene expression data using a model based on the negative binomial distribution. The resulting P-values were adjusted using the Benjamini and Hochberg's approach for controlling the false discovery rate. Genes with an adjusted P-value <0.05 found by DESeq were assigned as differentially expressed. Gene Ontology (GO) enrichment analysis of differentially expressed genes or lncRNA target genes were implemented by the GOseq R package (http://www.bioconductor.org/packages/release/bioc/html/goseq.html), in which gene length bias was corrected. The KOBAS software (http://kobas.cbi.pku.edu.cn/download.php) was used to test the statistical enrichment of differentially expressed genes or lncRNA target genes in KEGG pathways.

Identification of candidate genes involved in the GAS biosynthesis pathway

Candidate genes belonging to the GAS biosynthetic pathway in P. chinensis were manually identified according to the annotated sequences in the above databases. Protein coding sequences (CDS) were acquired by Angel software (Shimzu et al., 2006), and multiple sequence alignment was carried out by MEGA7.0 (Kumar et al., 2016).

Phylogenetic analysis

Amino acid alignments were performed using Clustal W, and phylogenetic trees were built using MEGA7.0 (Kumar et al., 2016), employing the neighbor joining method with 1000 bootstrap replicates, and applying the default settings for other parameters. The GenBank accession numbers/transcript numbers for all sequences are shown in Supplementary Table S1.
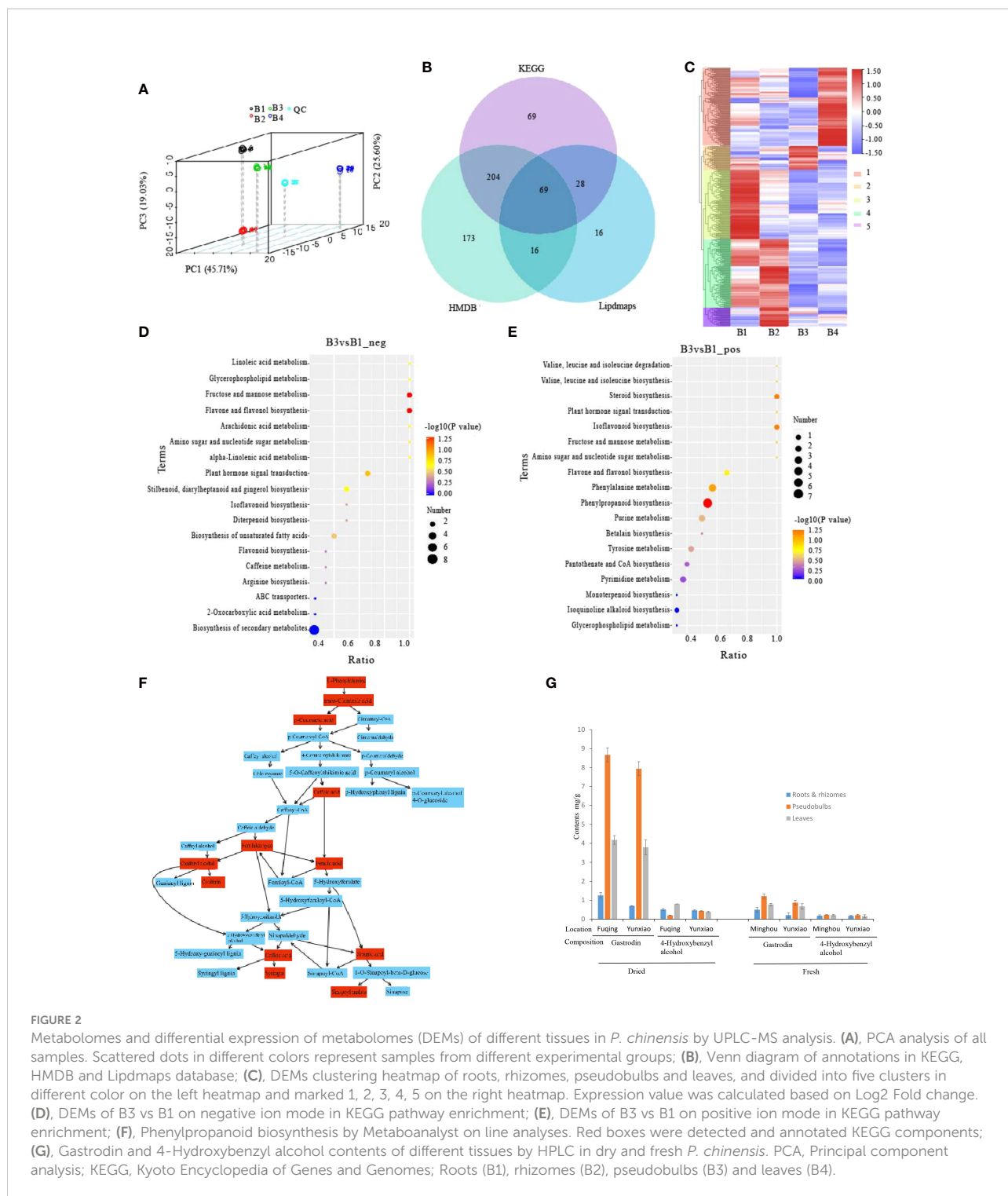
Quantitative Real-Time PCR validation

To verify the accuracy of transcriptomic data, 6 DEGs in roots, rhizomes, pseudobulbs and leaves were selected for qRT-PCR verification. Primers were designed using Primer-BLAST on the NCBI website (Supplementary Table S2). RNA was reverse transcribed using a TransScript® RT/R1 reagent kit according to the manufacturer's instructions. The qRT-PCR was performed on an ABI QuantStudio 3. There were three biological and three technical replicates for each sample. The qRT-PCR reaction system (20 μL) consisted of 10 μL of Universal SYBR Green Fast qPCR Mix SYBR Green Master Mix, 1 μL of cDNA, 0.4 μL each of forward and reverse primer, and 8.2 μL of sterile water. The qRT-PCR procedure included 3 min of initiation, followed by 40 cycles at 95°C for 5 s, 60°C for 30 s, and 72°C for 12 s. Relative expression levels were calculated using the $2^{-\Delta\Delta Ct}$ method and normalized according to the actin gene of β-tubulin.

Results

Tissue specific metabolites analysis

To explore the metabolite differences in roots, rhizomes, pseudobulbs and leaves of fresh P. chinensis, the samples were analyzed by UPLC-MS/MS (Figure 2A). A total of 1,156 (positive: 711, negative: 445) metabolites were identified (Supplementary Table S3). They were subsequently annotated in the KEGG, HMDB and Lipdmaps database, and annotations of 375, 462, and 129 metabolites were obtained, respectively (Figure 2B). The results of HCA showed that the DEMs were significantly varied in different organs, which were divided into five clusters (Figure 2C). The metabolites were comparable in leaves and pseudobulbs as well as roots and rhizomes. The relative content of GAS, Com_2638 in negative metabolites, was the highest in B3 and the lowest in B1 (Supplementary Table S4). Therefore, the comparison group of B3 and B1 was profiled and the results showed (Figures 2D, E) that 345 DEMs were mainly enriched in phenylpropanoid biosynthesis and phenylalanine metabolism (positive ion model), secondary metabolite biosynthetic process and flavone and flavonol

FIGURE 2

Metabolomes and differential expression of metabolomes (DEMs) of different tissues in *P. chinensis* by UPLC-MS analysis. **(A)**, PCA analysis of all samples. Scattered dots in different colors represent samples from different experimental groups; **(B)**, Venn diagram of annotations in KEGG, HMDB and Lipdmaps database; **(C)**, DEMs clustering heatmap of roots, rhizomes, pseudobulbs, and leaves, and divided into five clusters in different color on the left heatmap and marked 1, 2, 3, 4, 5 on the right heatmap. Expression value was calculated based on Log2 Fold change. **(D)**, DEMs of B3 vs B1 on negative ion mode in KEGG pathway enrichment; **(E)**, DEMs of B3 vs B1 on positive ion mode in KEGG pathway enrichment; **(F)**, Phenylpropanoid biosynthesis by Metaboanalyst on line analyses. Red boxes were detected and annotated KEGG components; **(G)**, Gastrodin and 4-Hydroxybenzyl alcohol contents of different tissues by HPLC in dry and fresh *P. chinensis*. PCA, Principal component analysis; KEGG, Kyoto Encyclopedia of Genes and Genomes; Roots (B1), rhizomes (B2), pseudobulbs (B3) and leaves (B4).

biosynthesis (negative ion model). Phenylpropanoid biosynthesis, including C00079 (L-Phenylalanine), C00423 (trans-Cinnamate), C00811 (4-Coumarate), C01197 (Caffeate), C02666 (Coniferyl aldehyde), C00590 (Coniferyl alcohol), C00761 (Coniferin), C01494 (Ferulate), C02325 (Sinapyl alcohol), C01533 (Syringin), C00482 (Sinapate) and C02887 (Sinapoyl malate), were shown by Metaboanalyst on line analyses (Figure 2F). Phenylalanine or its derivatives may be precursors for the biosynthesis of GAS biosynthesis in *P. chinensis*.

GAS and HBA contents

To investigate GAS and HBA contents in roots, rhizomes, pseudobulbs and leaves, crude MeOH extracts of dried or fresh *P. chinensis* samples from different sites were analyzed by HPLC. The results indicated that both in dried and fresh samples, the GAS of pseudobulbs was the highest (of 0.867 and 0.794% in dried samples, of 0.121 and 0.087% in fresh samples), followed by leaves. The lowest content was detected in roots and rhizomes (Figure 2G). However, HBA did not show significant difference among sampled plant organs. This result was consistent with the result of UPLC-MS/MS.

Sequencing and analysis of RNA-Seq

To obtain the transcriptome expression profiles in *P. chinensis*, the RNA was extracted from roots, rhizomes, pseudobulbs and leaves, and mixed together in an equal quantity, as one sample for PacBio Sequel sequencing. As a result, 26.66 Gigabytes Polymerase Read Bases from PacBio Sequel were produced. A total of 506,905 circular consensus sequences (CCS) with an average length of 2,195 bp was obtained after filtration with full passes ≥ 1 and quality > 0.90 (Table 1). To further improve the accuracy, >6 Gb of raw reads were obtained for each sample from Illumina Hiseq platform performed using NGS sequencing (Supplementary Table S5). The redundant and similar sequences were removed using CD-HIT software. Finally, 23,105 Unigenes were obtained with an average length of 2,186 bp. It was taken as the reference transcriptome (Table 1 and Supplementary Figure S1).

A total of 22,029 transcripts were annotated functionally in this analysis by searching against the GO, KEGG, COG/KOG, NT, Pfam, NR, and Swiss-Prot databases with transcripts 15,307 (69.49%), 21,322(96.79%), 14,155 (64.26%), 15,664 (71.11%), 15,307 (69.49%), 21,512 (97.65%), and 18,731 (85.03%), respectively (Figure 3A and Supplementary Table S5). However, 8,577 (38.94%) transcripts were annotated in all seven databases (Figure 3B and Supplementary Table S6). Based on the homologous sequence alignment with NR database and statistical analysis, *Elaeis guineensis* was the most homology species (6,867 transcripts, 31.92%) (Figure 3C and Supplementary Table S7). In KEGG database annotation, the transcripts were grouped into six main categories: Cellular processes (1,406 transcripts), Environmental information processing (1,269 transcripts), Genetic information processing (2,336 transcripts), Human diseases (2,668 transcripts), Metabolism (4,795 transcripts), and Organismal systems (2,215 transcripts). In the metabolism of phenylalanine and terpenoid backbone biosynthesis, 58 and 63 transcripts were involved, respectively. (Figure 3E and Supplementary Table S7). A group of 128 transcripts were matched to phenylpropandoid biosynthesis (ko00940), including: phenylalanine ammonia-lyase (PAL), 4-coumarate-CoA ligase (4CL), trans-cinnamate 4-monooxygenase (CYP73A) (Supplementary Figure S2). GO analysis showed that 15,307 transcripts could be classified into three categories: cellular component, molecular function and biological process. However, the GO terms of metabolic process (7,492 transcripts, 48.94%) were the most annotated transcripts in the Biological process (Figure 3D and Supplementary Table S7). In KOG classifications, the transcripts yielded 26 functional categories (Figure 3F and Supplementary Table S7). Up to 611 transcripts were annotated in amino acid transport and metabolism and 501 transcripts were annotated in secondary metabolites biosynthesis, transport and catabolism. The number of annotated transcripts identified using the NT, Pfam, and Swiss-Prot databases were summarized in Supplementary Table S7. These transcripts involved in amino acid metabolism or secondary metabolism might be partially involved in GAS biosynthesis in *P. chinensis*.

Analysis of differentially expressed genes (DEGs)

To identify genes differently expressed in different tissues of *P. chinensis*, 12 cDNA libraries, were mapped to reference sequence (CD-HIT software de-redundant and corrected consensus sequence). The cDNA libraries were generated with mRNA from roots, rhizomes, pseudobulbs and leaves. The matched rate of all the clean reads was >45% (Supplementary Table S8). The expression level per sample was shown with read count and FPKM in Supplementary Table S9. Compared to roots, the 7,787 DEGs (2,907 up-regulated and 4,880 down-

TABLE 1 The characteristics of transcriptome sequences of *P. chinensis* by PacBio sequencing and Illumina.

| Item	Number	Average length (bp)	N50	Min_Length	Max_Length
Subreads	14573961	1756	2075	51	——
Circular consensus sequences (CCS)	506905	2195	2542	53	14966
Full-Length non-chimeric Read (FLNC)	450366	2074	2431	56	14705
Polished consensus reads	45157	2054	2388	65	8171
Transcripts of after Illumina correction	45157	2054	2388	65	8171
Unigenes of after CD-HIT De-redundancy	23105	2186	2525	109	8171
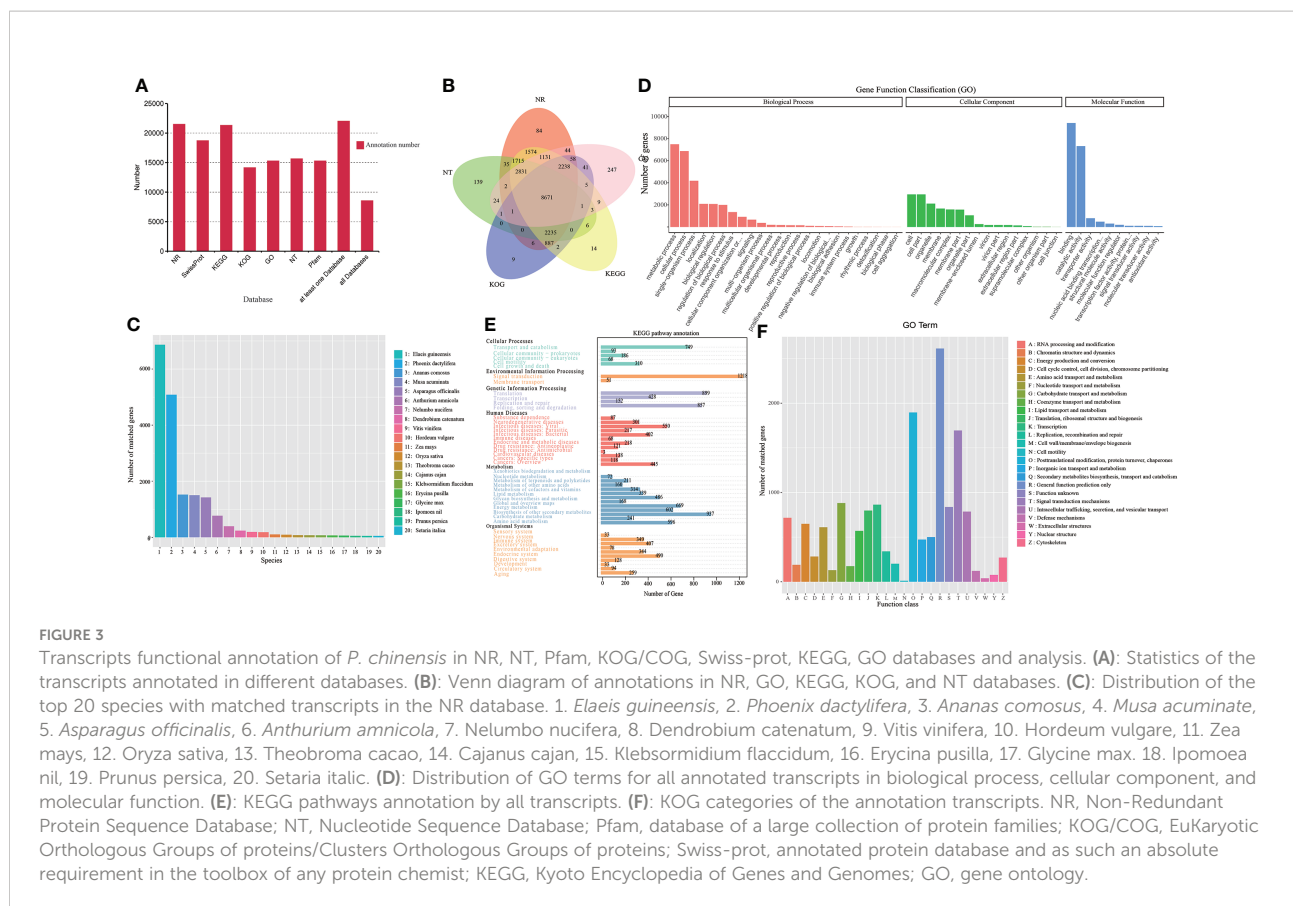
FIGURE 3

Transcripts functional annotation of *P. chinensis* in NR, NT, Pfam, KOG/COG, Swiss-prot, KEGG, GO databases and analysis. **(A)**: Statistics of the transcripts annotated in different databases. **(B)**: Venn diagram of annotations in NR, GO, KEGG, KOG, and NT databases. **(C)**: Distribution of the top 20 species with matched transcripts in the NR database. 1. *Elaeis guineensis*, 2. *Phoenix dactylifera*, 3. *Ananas comosus*, 4. *Musa acuminate*, 5. *Asparagus officinalis*, 6. *Anthurium amnicola*, 7. Nelumbo nucifera, 8. Dendrobium catenatum, 9. Vitis vinifera, 10. Hordeum vulgare, 11. Zea mays, 12. Oryza sativa, 13. Theobroma cacao, 14. Cajanus cajan, 15. Klebsormidium flaccidum, 16. Erycina pusilla, 17. Glycine max. 18. Ipomoea nil, 19. Prunus persica, 20. Setaria italic. **(D)**: Distribution of GO terms for all annotated transcripts in biological process, cellular component, and molecular function. **(E)**: KEGG pathways annotation by all transcripts. **(F)**: KOG categories of the annotation transcripts. NR, Non-Redundant Protein Sequence Database; NT, Nucleotide Sequence Database; Pfam, database of a large collection of protein families; KOG/COG, EuKaryotic Orthologous Groups of proteins/Clusters Orthologous Groups of proteins; Swiss-prot, annotated protein database and as such an absolute requirement in the toolbox of any protein chemist; KEGG, Kyoto Encyclopedia of Genes and Genomes; GO, gene ontology.

regulated), 8,376 DEGs (3,210 up-regulated and 5,166 down-regulated) and 9,146 DEGs (3,581 up-regulated and 5,565 down-regulated) were identified in rhizomes, pseudobulbs and leaves extracts, respectively (Figure 4A). And in total, 16,175 DEGs unigenes in all four tissues were identified. Among different tissues DEGs, only 357 common genes were expressed in all four compared tissues (Figure 4B).
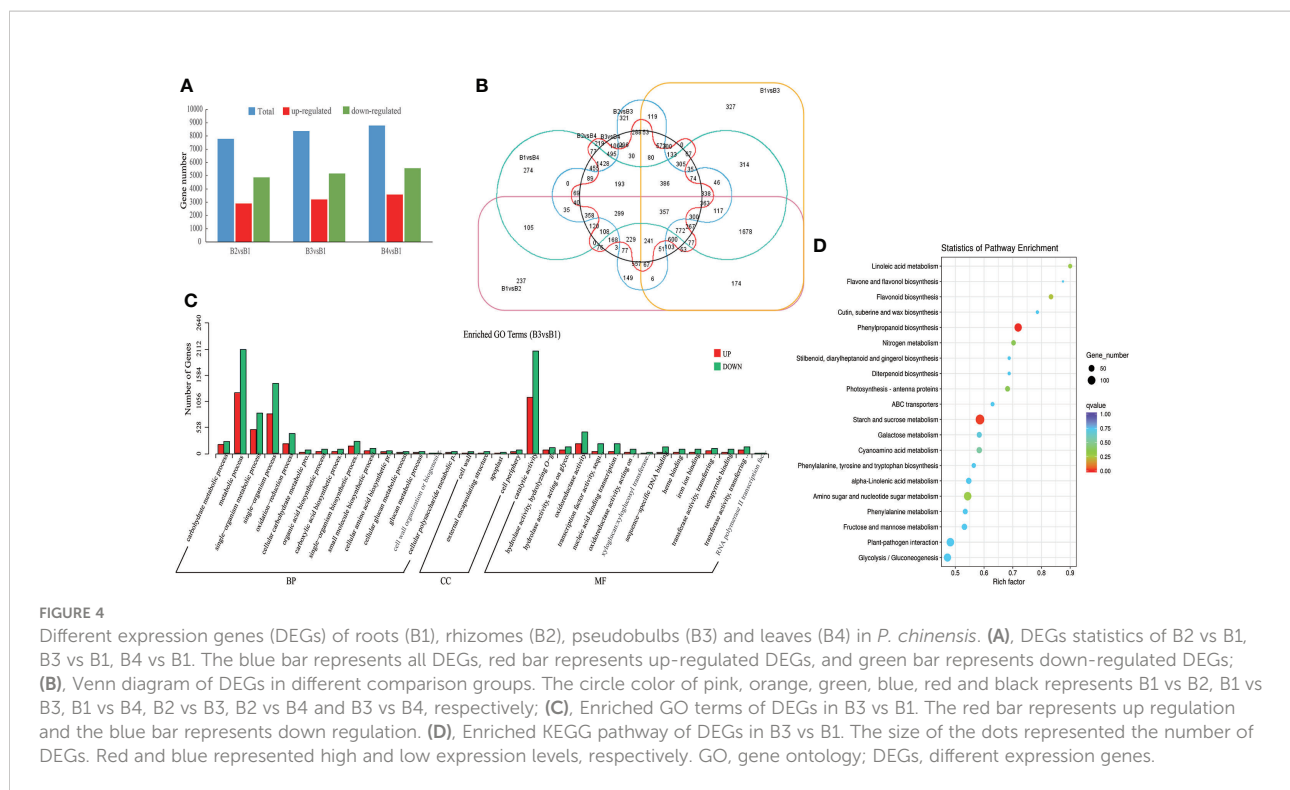
To reveal the biological significance of these DEGs, function annotation and enrichment analysis were performed by GO and KEGG database. The analysis of GO functional classification indicated that all the DEGs of B3 and B1 comparison group were grouped into 34 functional groups, including 15 molecular function categories, 15 biological processes, and 4 cellular components. (Figure 4C). Metabolic process and single-organism in the biological processes, and catalytic activity in the molecular function were the most enriched terms. However, in almost all terms, down-regulated genes were higher than up-regulated genes. To further illustrate the alterations of gene expression between B3 and B1, the KEGG analysis of all the DEGs of B3 and B1 comparison group was made. The DEGs were enriched in linoleic acid metabolism, flavonoid biosynthesis, phenylpropanoid biosynthesis and others (Figure 4D). However, while the up-regulated DEGs of B3 and B1 were mainly enriched in photosynthesis - antenna proteins,

phenylpropanoid biosynthesis; the down-regualted DEGs of B3 and B1 were mainly enriched in flavonoid biosynthesis, linoleic acid metabolism and others terms (Supplementary Figures 3-4.). 28 transcripts were up-regulated in phenylpropanoid biosynthesis of B3 and B1 comparison group, including encoding 4CL, cinnamyl-alcohol dehydrogenase, peroxidase, and beta-glucosidase. These enzymes might be critical for the synthesis of GAS precursors. In addition, the transcripts transcript28360/f2p0/1592, transcript25791/f2p0/1719, etc. had significant expression differences in B3 and B1 comparison group and the *p* value was close to zero.

The candidate genes involved in GAS biosynthesis pathway

Based on the KEGG pathway (map00940, map00996) analysis as reported in *G. elata* (Shan et al., 2021), the putative GAS biosynthetic pathway of *P. chinensis* is shown in Figure 6. The biosynthesis of GAS primarily initiated from the L-phenylalanine, which is derived from the common phenylpropanoid biosynthesis pathway that is broadly distributed in plants (Zhang et al., 2020; Rai et al., 2021). A total of 80 unigenes were identified that encoding eight key

FIGURE 4
Different expression genes (DEGs) of roots (B1), rhizomes (B2), pseudobulbs (B3) and leaves (B4) in *P. chinensis*. **(A)**, DEGs statistics of B2 vs B1, B3 vs B1, B4 vs B1. The blue bar represents all DEGs, red bar represents up-regulated DEGs, and green bar represents down-regulated DEGs; **(B)**, Venn diagram of DEGs in different comparison groups. The circle color of pink, orange, green, blue, red and black represents B1 vs B2, B1 vs B3, B1 vs B4, B2 vs B3, B2 vs B4 and B3 vs B4, respectively; **(C)**, Enriched GO terms of DEGs in B3 vs B1. The red bar represents up regulation and the blue bar represents down regulation. **(D)**, Enriched KEGG pathway of DEGs in B3 vs B1. The size of the dots represented the number of DEGs. Red and blue represented high and low expression levels, respectively. GO, gene ontology; DEGs, different expression genes.

enzymes controlling GAS biosynthesis: phenylalanine ammonia-lyase (PAL), trans-cinnamate 4-monooxygenase (CYP73A), 4CL, shikimate O-hydroxycinnamoyltransferase (HCT), 5-O-(4-coumaroyl)-D-quinate 3'-monooxygenase (C3H), caffeoyl coenzyme A-O-methyltransferase (CCoAOMT), alcohol dehydrogenases (ADH) and GT. The relative expression levels of the DEGs in the different tissues were showed in the heatmap (Figure 5). However, as the last key enzyme, GT, which catalyzed the GAS synthesis from HBA with UDP-glucose, had 39 Unigenes. These unigenes were divided into four types according to the types of encoded enzymes, including GT1 (3 beta-glucosyltransferase (2.4.1.173)), GT2 (cis-zeatin O-glucosyltransferase (2.4.1.215)), GT3 (hydroquinone glucosyltransferase (2.4.1.218)) and GT4 (others glucosyltransferase (2.4.1-)). Some annotated transcripts GTs were differently expressed in targeted tissues of the studied plant: transcript28360/f2p0/1592, transcript16563/f4p0/2237, transcript19586/f3p0/2041 and transcript25251/f2p0/1759 were highly expressed in pseudobulbs, and lower in leaves, least in roots and rhizomes (Figure 5).

To verify the accuracy of RNA-seq data, the quantitative real-time PCR (qRT-PCR) was used to validate differential gene expression levels of roots, rhizomes, pseudobulbs and leaves with gene-specific primers (Supplementary Table S2). The results showed that the gene relative expression profile was almost consistent with the RNA-seq data. It further demonstrated the credibility of the data generated in the present study (Figure 6).

Identification of glucosyltransferase

The previous results strongly suggest that the AsUGT, a serpentwood-derived GT convert HBA to GAS with high catalytic efficiency in yeast, compared with UGT73B6 from *Rhodiola sachalinensis* in *Escherichia coli* (Bai et al., 2016; Yin et al., 2020). In *P. chinensis*, significant homology was found using queries from the encoding sequences of the AsUGT and UGT73B6 genes. The most similar transcripts were transcript17418/f3p0/2174 (55.72% identical), transcript28360/f2p0/1592 (47.38% identical), transcript27824/f2p0/1641 (49.90% identical), and transcript29686/f2p0/1549 (48.86% identical) (Supplementary Table S10). To further characterize the GTs, a phylogenetic tree was constructed based on *P. chinensis* and other plant GT protein sequences, including *Zea mays, Arabidopsis thaliana, Apostasia shenzhenica, Rhodiola sachalinensis, Dendrobium catenatum, Rhodiola sachalinensis* and *Phalaenopsis equestris* (Supplementary Table S1). All GT members in *P. chinensis* were divided into eight phylogenetic groups. Among them, AsUGT (Rse_Q9AR73.1), the reference UGT was clustered in the same clade with transcript17418/f3p0/2174, transcript28360/f2p0/1592, transcript24635/f2p0/1797 and transcript28208/f27p0/1567, whereas UGT73B6 (Rsa_AAS55083.1) was clustered in the same clade with transcript19941/f2p0/2048 (Figure 6). The gene expression level of these transcripts exhibited great tissue-specific tendency. While others were highly expressed in the roots, transcript28360/f2p0/1592 was the only one with high expression in the pseudobulbs. Meanwhile, homologous searching by using
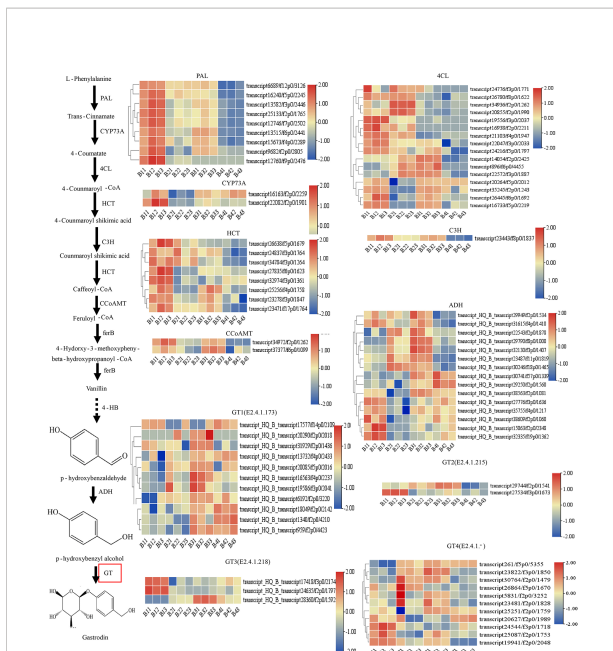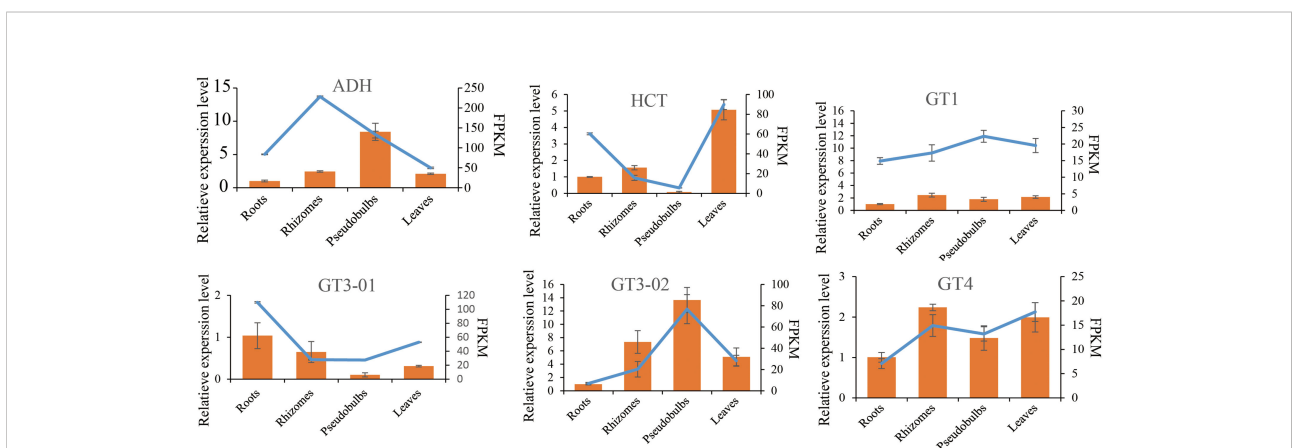
FIGURE 5

Putative gastrodin biosynthesis pathway in *P. chinensis*. This pathway was constructed based on the KEGG pathway (map00940, map01061) and literature references. The expression levels deduced from the RMPK of each Unigene that encodes the relevant enzyme, were shown as heat map, whereas roots (triplicates as B11, B12, B13), rhizomes (triplicates as B21, B22, B23), pseudobulbs (triplicates as B31, B32, B33) and leaves (triplicates as B41, B42, B43) were separately analyzed. The expression value was calculated based on the Log2 Fold change. Red and blue represented high and low expression levels, respectively. Non-dashed line arrows represent identified enzymatic reactions, and dashed line arrows represent multiple enzymatic reactions through multiple steps.
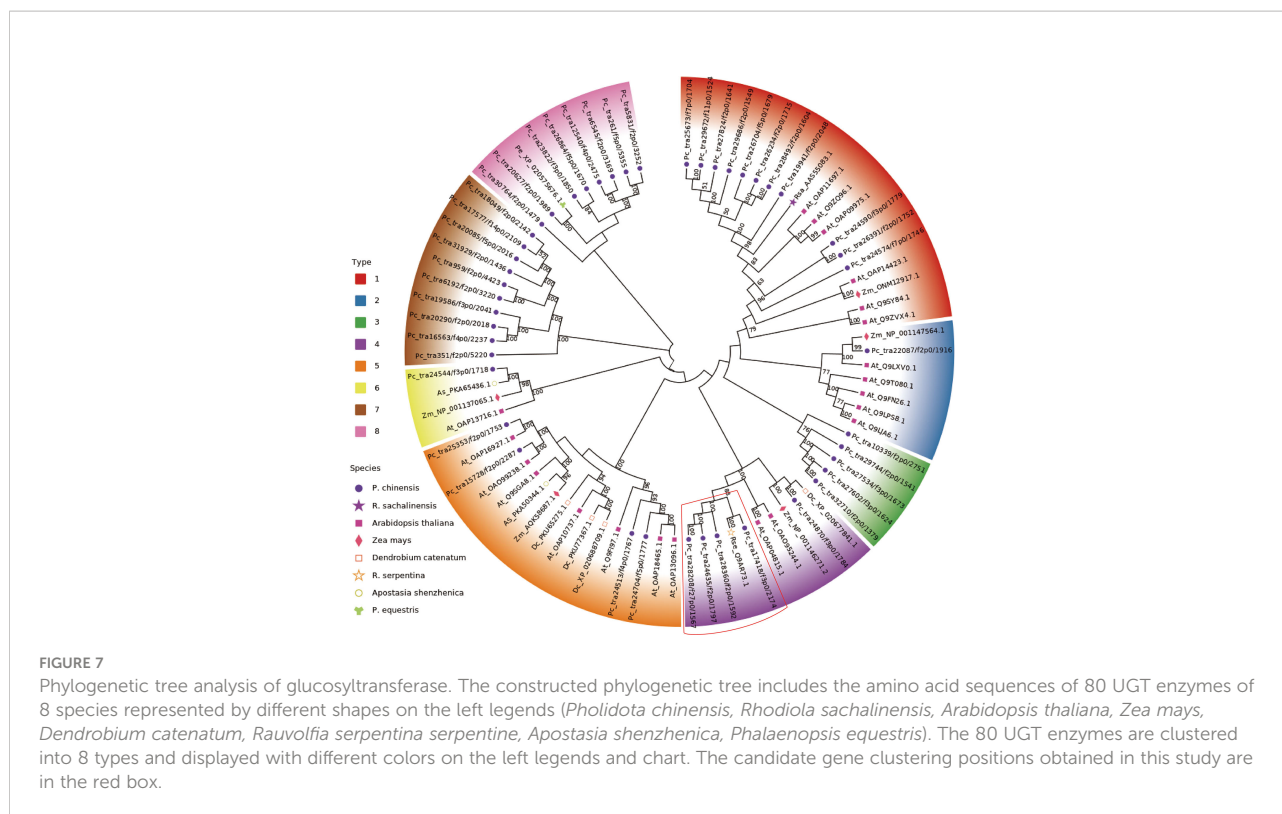
queries from the transcript TRINITY_DN50323_c0_g1(which might participate in glucosylation in the GAS biosynthetic pathway of *G. elata*) (Tsai et al., 2016), showed significant homology with transcript20627/f2p0/1989 (84.29% identical) in *P. chinensis* and XM_020720017.1 (83.4% identical) in *P. equestris*, respectively (Supplementary Table S11). However, the expression level of transcript20627/f2p0/1989 was the highest in leaves and lower in pseudobulbs, which might be not the best candidate gene (Figure 7).

Integrative analysis of transcriptomics and metabolomics

To obtain a deeper understanding, a multi-omics analysis was performed. These analyses integrated the metabolomics with the transcriptomic data. In negative/positive ion mode, top 50 DEMs (sorted by p value from small to large) and top 100 DEGs (sorted by *p* value from small to large) of B2 and B1 comparison group, B3 and B1 comparison group, B4 and B1 comparison group were shown in Supplementary Figures 5–10. These DEMs and DEGs had a stronger positive or negative connection (R>0.9). To identify the major biochemical pathways and signal transduction involved pathways of DEMs and DEGs, all DEMs and DEGs were matched to the KEGG pathway. The results revealed that the DEMs and DEGs were main enriched in phenylpropanoid biosynthesis and linoleic acid metabolism of B2 and B1 comparison group, phenylpropanoid biosynthesis and amino sugar and nucleotide sugar metabolism of B3 and B1 comparison group, phenylpropanoid biosynthesis and flavonoid biosynthesis of B4 and B1 comparison group (Figure 8). The phenylpropanoid biosynthesis may be critical for GAS

FIGURE 6

Verification of six selected DEGs by qRT-PCR. Comparison of RNA-seq data (Blue line chart) with qRT-PCR data (Yellow bar graph). The relative qRT-PCR expression level of selected DEGs is shown on the y-axis to the left. *β-tubulin* (TUB) was used as the internal control. Three biological replicates were used. The normalized expression level (FPKM; expected number of Fragments Per Kilobase of transcript sequence per Millions base pairs sequenced) of RNA-seq is indicated on the *y*-axis to the left.
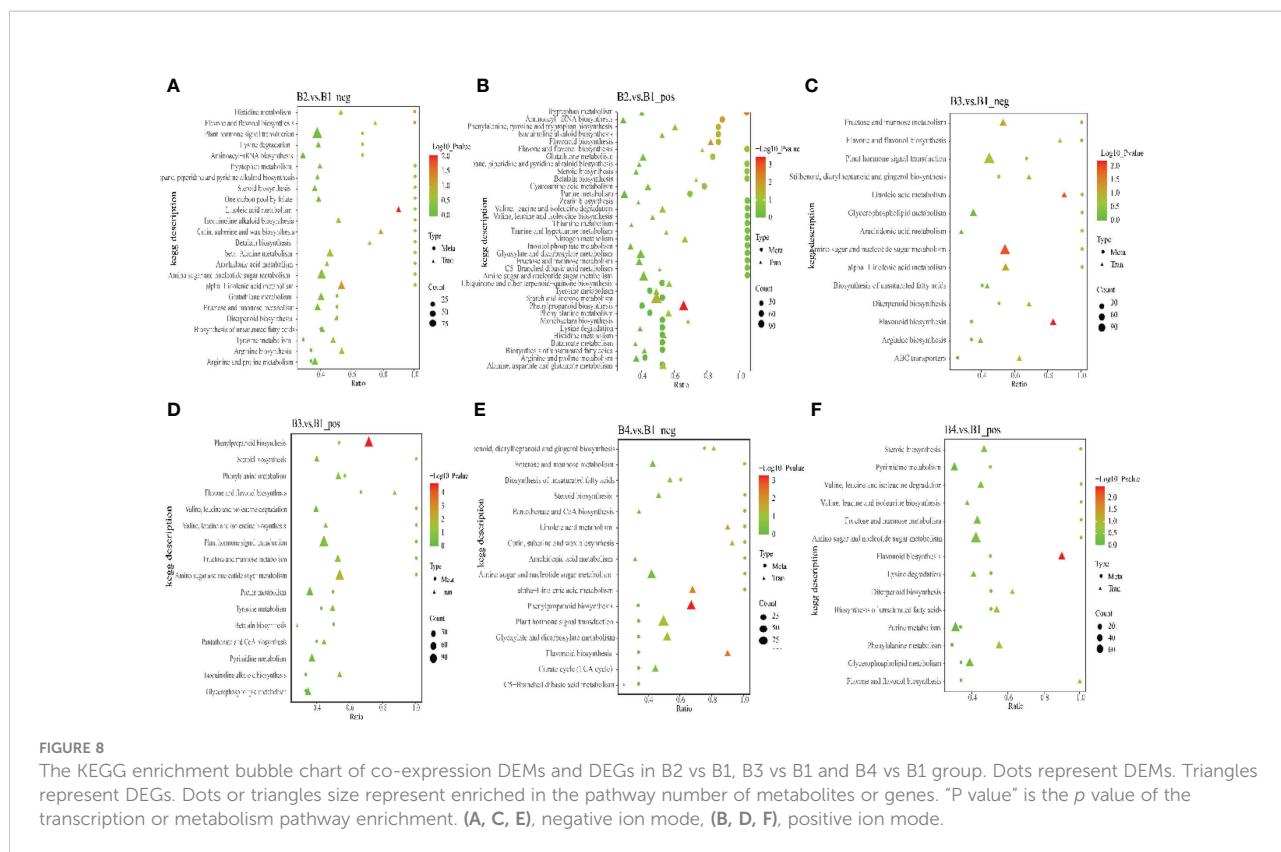
FIGURE 7

Phylogenetic tree analysis of glucosyltransferase. The constructed phylogenetic tree includes the amino acid sequences of 80 UGT enzymes of 8 species represented by different shapes on the left legends (*Pholidota chinensis, Rhodiola sachalinensis, Arabidopsis thaliana, Zea mays, Dendrobium catenatum, Rauvolfia serpentina serpentine, Apostasia shenzhenica, Phalaenopsis equestris*). The 80 UGT enzymes are clustered into 8 types and displayed with different colors on the left legends and chart. The candidate gene clustering positions obtained in this study are in the red box.

biosynthesis in *P. chinensis*. There were a notable association (R>0.9) between 8,376 DEGs and 345 DEMs based on the Pearson's correlation coefficient in B3 and B1 comparison group (Supplementary Table S12). And the results showed that no matter in positive or negative ion mode, there were metabolites that had significant correlation with most genes, such as coniferyl aldehyde, butylparaben, 3_4_5-trimethoxycinnamic acid, monobenzyl phthalate etc. Coniferyl aldehyde, a natural non-toxic and anti-inflammatory phenolic compound extracted from edible and medicinal plants (Wang et al., 2020), might be involved in the biosynthesis of GAS (Huccetogullari et al., 2019) and it had significant associations (R>0.9) with 6266 DEGs (Supplementary Table S13).

To further understand the relationship between metabolites and genes in common pathway, DEGs and DEMs of B3 and B1 comparison group were simultaneously mapped to the KEGG pathway. The results in negative ion mode showed that 2 DEMs and 42 DEGs (fructose and mannose metabolism), 2 DEMs and 7 DEGs (flavone and flavonol biosynthesis), 2 DEMs and 115 DEGs (plant hormone signal transduction), 2 DEMs and 11 DEGs (stilbenoid, diarylheptanoid and gingerol biosynthesis) enriched the corresponding biological processes (Figure 8C). However, the results in positive ion mode showed that 7 DEMs and 92 DEGs, 2 DEMs and 19 DEGs, 4 DEMs and 31 DEGs, 2 DEMs and 7 DEGs were enriched phenylpropanoid biosynthesis, steroid biosynthesis, phenylalanine metabolism, and flavone and flavonol biosynthesis, respectively (Figure 8D). Furthermore, the putative candidate

gene transcript28360/f2p0/1592 in GT3 was significantly negative correlated with coniferylaldehyde and spermidine, but had a significantly positive correlation with GAS, isoeugenol and sinapoyl malate (Supplementary Table S14). These results are consistent with transcriptome or metabolome results.

Discussion

The GAS is the second compound identified from the plant *G. elata* after vanilyalcohol. It is a phenolic glycoside that chemically known as 4-hydroxybenzyl alcohol-4-*O*-β-*D*-glucopyranoside. And it is also the main bioactive constituent of another TCM *Rhizoma Gastrodiae* (Tao et al., 2009; Liu et al., 2018). Being the largest and the most widespread class of plant secondary metabolites, phenolics have been extensively researched due to the diverse health benefits. Some of these include flavonoids, lignans, coumarins, chalcones, and phenolic acids, which participate in the regulation of plant growth, seed germination, and in defense responses (Acosta-Estrada et al., 2014; Naikoo et al., 2019). The GAS content is the most appreciated analytical marker for the quality standardization of *P. chinensis* (Zhang et al., 2019). The mechanism of GAS action is gradually being understood and recognized (Chen et al., 2022; Yang et al., 2022). Several reports have shown that the content of GAS in *P. chinensis* was higher than the content in *G. elata*, one of the major sources of GAS (Zhang et al., 2019; Wang et al., 2022).

FIGURE 8
The KEGG enrichment bubble chart of co-expression DEMs and DEGs in B2 vs B1, B3 vs B1 and B4 vs B1 group. Dots represent DEMs. Triangles represent DEGs. Dots or triangles size represent enriched in the pathway number of metabolites or genes. "P value" is the *p* value of the transcription or metabolism pathway enrichment. **(A, C, E)**, negative ion mode, **(B, D, F)**, positive ion mode.

Metabolomics, especially untargeted metabolomics using LC-MS, is considered to be the best omics technique to represent the phenotypes because its data analysis based on the state of biochemical activity in the living organism (Fraisier-Vannier et al., 2020; Zeki et al., 2020). In this study, 1,156 metabolic differences were identified in roots, rhizomes, pseudobulbs and leaves of fresh *P. chinensis* using the UPLC-MS/MS. Many of the metabolites were phenolic compounds that were mainly enriched in phenylpropanoid biosynthesis, flavone and flavonol biosynthesis, and phenylalanine metabolism. And some of these compounds were known to exhibit antioxidant, antidiabetic, and anti-inflammatory activities (Ren et al., 2020). Simultaneously, HPLC analysis revealed that pseudobulbs were the primary tissue of GAS, followed by leaves, roots and rhizomes.

Some medicinal plants lack genomic information due to the wide variety, limiting some research, such as the integrity of transcriptome assembly (Cheng et al., 2021). The combination of third-generation sequencing and second-generation sequencing is an effective method for gene mining without reference genome (Liu et al., 2022). In the present study, the mean length of predicted unigenes (2,186 bp) the N50 length (2,525 bp) were longer than some other traditional medicine including *G. elata* (Tsai et al., 2016), *Dendrobium officinale* (Li et al., 2021) and *Dendrobium sinense* (Zhang et al., 2021), indicated that the transcriptome assembly were of high reliability and quality.

The GAS was synthesized from HBA with UDP-glucose *via* glycosylation catalyzed through GT, and HBA was synthesized from cresols (toluene) degradation through two steps of hydroxylation *via* monooxygenase in *G. elata* (Tsai et al., 2016). The monooxygenase 1.14.13.-, which were reported in *G. elata* (Tsai et al., 2016), were not discovered in *P. chinensis*. However, it was also reported that GAS could be synthesized *via* the phenylpropanoid pathway, and the PAL, C4H, 4CL and GT are the key genes in biosynthetic pathway of GAS in *G. elata* (Shan et al., 2021). It is worth noting that GAS was synthesized from glucose by an artificial microbial pathway with key genes of ADHs and GT in *Saccharomyces cerevisiae* and *Escherichia coli*, respectively (Bai et al., 2016; Yin et al., 2020). Glycosylation is often the last step in the biosynthesis of natural products in plants and plays an important role in a variety of biosynthetic pathways (He et al., 2022). According to the above analysis, although the starting point of GAS synthesis is different, but the last step is same, that is, GT catalyzes the conversion of HBA to GAS. In this study, putatively 80 unigenes involved in the biosynthetic pathway of GAS in *P. chinensis* were identified including genes for PAL, CYP73A, 4CL, HCT, C3H, CCoAOMT, ADH and GT. The GT (39 unigenes) were divided into four subgroups according to the types of encoded enzymes. Among all transcripts being found, transcript28360/f2p0/1592, transcript16563/f4p0/2237, transcript19586/f3p0/

2041 and transcript25251/f2p0/1759 were highly expressed in pseudobulbs, and lower in other targeted plant parts (leaves, roots and rhizomes), which could be key candidates. Based on phylogenetic tree analysis, the transcript28360/f2p0/1592 in GT3 was deduced as the best candidate gene because it shares a highly homologous sequence with AsUGT, which was identified as the plant-derived GT that converts HBA to GAS with high catalytic efficiency in yeast (Yin et al., 2020).

Integrated analysis of transcriptome and metabolome provides an efficient approach for the research of metabolic networks and key genes. For example, multi-omics were applied for flavonoid biosynthesis in a purple tea plant cultivar (Song et al., 2022), the response of *Zanthoxylum bungeanum* and apple to different stresses (Li et al., 2021; Sun et al., 2021). It is worth noting that the proteome (Camp et al., 2022) is also often analyzed together with the transcriptome or metabolome, but this study has not yet performed a proteome of *P. chinensis*. Coniferylaldehyde, coniferyl alcohol, isoeugenol and sinapoyl malate were members of dominant group of volatile compounds, and those volatile phenyl propene formation might takes two enzymatic steps with lignin, and they were perhaps involved in the synthesis of GAS precursors (Ramya et al., 2020).

Other than our focus on the dissection of transcriptomic and metabolic profiles of the extremely less studied traditional medicine for understanding of the GAS biosynthetic pathway, our study, as the first in the genus, also provides useful data for other basic researches on *P. chinensis* and related species. For example, the plant undergoes a well differentiated developmental stage of pseudobulb and our data could be lent for mining of the molecular mechanisms of pseudobulb development.

Conclusion

The GAS is an important active component of a traditional Chinese medicine, but its biosynthetic pathway in *P. chinensis* is still unclear. In the present study, the biosynthetic pathway of GAS in *P. chinensis* was speculated by combination of transcriptome and metabolome analysis. Unigenes involved in the biosynthetic pathways, as well as the metabolites, were identified. Besides commonly known unigenes in the synthetic pathway for PAL, CYP73A, 4CL, HCT, C3H, CCoAOMT, and ADH, best candidates for the last synthetic step of GAS, the transcript28360/f2p0/1592, were assured by bioinformatics. Since the growth of the plant is extremely slow and it is not practical to cultivate it in large area to gain sufficient yield and profit, the biosynthetic pathway disclosed in this study, especially the last unique GT for GAS, will be extremely useful for possible biosynthetic engineering of the chemical in microbial platforms. To the best of our knowledge, this study is the first exploration of the genes involved in the GAS biosynthesis in *P. chinensis* and plants in the same genus.

Data availability statement

The original contributions presented in the study are publicly available. This data can be found here: NCBI, PRJNA841044 and PRJNA806713.

Author contributions

Funding

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2022.1024239/full#supplementary-material

SUPPLEMENTARY FIGURE 1

Assemble Unigene length distribution of *P. chinensis* by PacBio sequencing and Illumina data correction, Abscissa (length/bp), ordinate (number of genes).

SUPPLEMENTARY FIGURE 2

The transcripts were matched to phenylpropandoid biosynthesis (ko00940)

SUPPLEMENTARY FIGURE 3

Up-regulated DEGS of B3 vs B1. The size of the dots indicates the number of DEGS in this pathway, and the colors of the dots correspond to different q value ranges, which is closer zero (red color) and more significant enrichments.

SUPPLEMENTARY FIGURE 4

Down-regulated DEGS of B3 vs B1. The size of the dots indicates the number of DEGS in this pathway, and the colors of the dots correspond to different q value ranges, which is closer zero (red color) and more significant enrichments.

SUPPLEMENTARY FIGURE 5

The correlation heat map in negative ion mode of co-expression B2 vs B1 DEMs & B2 vs B1 DEGs.

SUPPLEMENTARY FIGURE 6

The correlation heat map in positive ion mode of co-expression B2 vs B1 DEMs & B2 vs B1 DEGs.

SUPPLEMENTARY FIGURE 7

The correlation heat map in negative ion mode of co-expression B3 vs B1 DEMs & B3 vs B1 DEGs.

SUPPLEMENTARY FIGURE 8

The correlation heat map in positive ion mode of co-expression B3 vs B1 DEMs & B3 vs B1 DEGs.

SUPPLEMENTARY FIGURE 9

The correlation heat map in negative ion mode of co-expression B4 vs B1 DEMs & B4 vs B1 DEGs.

SUPPLEMENTARY FIGURE 10

The correlation heat map in positive ion mode of co-expression B4 vs B1 DEMs & B4 vs B1 DEGs.

SUPPLEMENTARY TABLE 3

Species of components in positive and negative ion modes by UPLC-MS/MS.

SUPPLEMENTARY TABLE 4

The GAS relative content of different tissue by UPLC-MS/MS.

SUPPLEMENTARY TABLE 5

Sequencing data quality from Illumina Hiseq.

SUPPLEMENTARY TABLE 6

Annotation and classification of detected unigenes by the seven datasets.

SUPPLEMENTARY TABLE 7

The annotation of NR, GO, KEGG, KOG, NT, Pfam, Swissprot.

SUPPLEMENTARY TABLE 8

The comparison rate of the second-generation data map to the third-generation data.

SUPPLEMENTARY TABLE 9

The expression level of each sample.

SUPPLEMENTARY TABLE 10

Based on amino acid blast results in *P. chinensis* with AsUGT and UGT73B6 genes as query.

SUPPLEMENTARY TABLE 11

Blast results of GT in *P. chinensis* and *Phalaenopsis equestris* with candidate gene TRINITY_DN50323_c0_g1 of *Gastrodia elata* as query.

SUPPLEMENTARY TABLE 12

The correlation in positive and negative ion mode of DEMs & DEGs based on $r^2 > 0.91$ and p<0.01.

SUPPLEMENTARY TABLE 13

Coniferyl aldehyde correlation DEGs based on $r^2 > 0.91$ and p<0.01.

SUPPLEMENTARY TABLE 14

The Correlation putative candidate gene transcript28360/f2p0/1592 in GT3 with metabolites.

References

Acosta-Estrada, B. A., Gutierrez-Uribe, J. A., and Serna-Saldivar, S. O. (2014). Bound phenolics in foods, a review. *Food Chem.* 152, 46–55. doi: 10.1016/j.foodchem.2013.11.093

Bae, E. K., An, C., Kang, M. J., Lee, S. A., Lee, S. J., Kim, K. T., et al. (2022). Chromosome-level genome assembly of the fully mycoheterotrophic orchid gastrodia elata. *G3 Genes|Genomes|Genetics (Bethesda)* 12 (3), 3. doi: 10.1093/g3journal/jkab433

Bai, Y. F., Yin, H., Bi, H. P., Zhuang, Y. B., Liu, T., and Ma, Y. H. (2016). De novo biosynthesis of gastrodin in escherichia coli. *Metab. Eng.* 35, 138–147. doi: 10.1016/j.ymben.2016.01.002

Camp, E. F., Kahlke, T., Signal, B., Oakley, C. A., Lutz, A., Davy, S. K., et al. (2022). Proteome metabolome and transcriptome data for three symbiodiniaceae under ambient and heat stress conditions. *Sci. Data* 9, 153. doi: 10.1038/s41597-022-01258-w

Cheng, Q. Q., Ouyang, Y., Tang, Z. Y., Lao, C. C., Zhang, Y. Y., Cheng, C. S., et al. (2021). Review on the development and applications of medicinal plant genomes. *Front. Plant Sci.* 12. doi: 10.3389/fpls.2021.791219

Chen, J., Tang, Y. J., Kohler, A., Lebreton, A., Xing, Y. M., Zhou, D. Y., et al. (2022). Comparative transcriptomics analysis of the symbiotic germination of d. officinale (Orchidaceae) with emphasis on plant cell wall modification and cell wall-degrading enzymes. *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.880600

Chen, X., Wang, J., He, Z., Liu, X., Liu, H., and Wang, X. (2022). Analgesic and anxiolytic effects of gastrodin and its influences on ferroptosis and jejunal microbiota in complete freund's adjuvant-injected mice. *Front. Microbiol.* 13. doi: 10.3389/fmicb.2022.841662

Chen, X., Xie, C., Sun, L., Ding, J., and Cai, H. (2015). Longitudinal metabolomics profiling of parkinson's disease-related alpha-synuclein A53T transgenic mice. *PloS One* 10, e0136612. doi: 10.1371/journal.pone.0136612

Dai, W., Xie, D., Lu, M., Li, P., Lv, H., Yang, C., et al. (2017). Characterization of white tea metabolome: Comparison against green and black tea by a nontargeted metabolomics approach. *Food Res. Int.* 96, 40–45. doi: 10.1016/j.foodres.2017.03.028

Dunn, W. B., Broadhurst, D., Begley, P., Zelena, E., Francis-McIntyre, S., Anderson, N., et al. (2011). Procedures for large-scale metabolic profiling of serum and plasma using gas chromatography and liquid chromatography coupled to mass spectrometry. *Nat. Protoc.* 6, 1060–1083. doi: 10.1038/nprot.2011.335

Finn, R. D., Coggill, P., Eberhardt, R. Y., Eddy, S. R., Mistry, J., Mitchell, A. L., et al. (2016). The pfam protein families database: towards a more sustainable future. *Nucleic Acids Res.* 44, D279–D285. doi: 10.1093/nar/gkv1344

Fraisier-Vannier, O., Chervin, J., Cabanac, G., Puech, V., Fournier, S., Durand, V., et al. (2020). MS-CleanR: A feature-filtering workflow for untargeted LC-MS based metabolomics. *Anal. Chem.* 92, 9971–9981. doi: 10.1021/acs.analchem.0c01594

Fu, L., Niu, B., Zhu, Z., Wu, S., and Li, W. (2012). CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics* 28, 3150–3152. doi: 10.1093/bioinformatics/bts565

Gallage, N. J., Hansen, E. H., Kannangara, R., Olsen, C. E., Motawia, M. S., Jorgensen, K., et al. (2014). Vanillin formation from ferulic acid in vanilla planifolia is catalysed by a single enzyme. *Nat. Commun.* 5, 4037. doi: 10.1038/ncomms5037

Hao, L., Wang, J., Page, D., Asthana, S., Zetterberg, H., Carlsson, C., et al. (2018). Comparative evaluation of MS-based metabolomics software and its application to preclinical alzheimer's disease. *Sci. Rep.* 8, 9291. doi: 10.1038/s41598-018-27031-x

He, B., Bai, X., Tan, Y., Xie, W., Feng, Y., and Yang, G. Y. (2022). Glycosyltransferases: Mining, engineering and applications in biosynthesis of glycosylated plant natural products. *Synth Syst. Biotechnol.* 7, 602–620. doi: 10.1016/j.synbio.2022.01.001

Hong, l., Guo, Z., Huang, K., Wei, S., Liu, B., Meng, S., et al. (2015). Ethnobotanical study on medicinal plants used by maonan people in China. *J. Ethnobiol. Ethnomed.* 11, 32. doi: 10.1186/s13002-015-0019-1

Huccetogullari, D., Luo, Z. W., and Lee, S. Y. (2019). Metabolic engineering of microorganisms for production of aromatic compounds. *Microb. Cell Fact* 18, 41. doi: 10.1186/s12934-019-1090-4

Hu, H. C., Fei, X. T., He, B. B., Luo, Y. L., Qi, Y. C., and Wei, A. Z. (2021). Integrated analysis of metabolome and transcriptome data for uncovering flavonoid components of zanthoxylum bungeanum maxim. leaves under drought stress. *Front. Nutr.* 8. doi: 10.3389/fnut.2021.801244

Kang, Y. J., Yang, D. C., Kong, L., Hou, M., Meng, Y. Q., Wei, L., et al. (2017). CPC2: a fast and accurate coding potential calculator based on sequence intrinsic features. *Nucleic Acids Res.* 45, W12–W16. doi: 10.1093/nar/gkx428

Kumar, S., Stecher, G., and Tamura, K. (2016). MEGA7: Molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol. Biol. Evol.* 33, 1870–1874. doi: 10.1093/molbev/msw054

Li, B., and Dewey, C. N. (2011). RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinf.* 12 (1), 323. doi: 10.1186/1471-2105-12-323

Li, N., Dong, Y., Lv, M., Qian, L., Sun, X., Liu, L., et al. (2021). Combined analysis of volatile terpenoid metabolism and transcriptome reveals transcription factors related to terpene synthase in two cultivars of dendrobium officinale flowers. *Front. Genet.* 12. doi: 10.3389/fgene.2021.661296

Li, P., Ruan, Z., Fei, Z., Yan, J., and Tang, G. (2021). Integrated transcriptome and metabolome analysis revealed that flavonoid biosynthesis may dominate the resistance of zanthoxylum bungeanum against stem canker. *J. Agric. Food Chem.* 69, 6360–6378. doi: 10.1021/acs.jafc.1c00357

Liu, Y., Gao, J., Peng, M., Meng, H., Ma, H., Cai, P., et al. (2018). A review on central nervous system effects of gastrodin. *Front. Pharmacol.* 9. doi: 10.3389/fphar.2018.00024

Liu, X., Gong, X., Liu, Y., Liu, J., Zhang, H., Qiao, S., et al. (2022). Application of high-throughput sequencing on the Chinese herbal medicine for the data-mining of the bioactive compounds. *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.900035

Liu, S., and Yang, S. (2022). Cardiovascular protective properties of gastrodin. *Asian Pacific J. Trop. Biomed.* 12 (4), 4. doi: 10.4103/2221-1691.340558

Liu, Y., Zhang, Y., Jin, Y., and Chen, Y. (2007). Advance on the chemical and pharmacological studies on plants of pholidota genus. *Lishizhen Med. Mater. Med. Res.* 18, 1631–1633. doi: 1008-0805(2007)-1631-03

Love, M. I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 15, 550. doi: 10.1186/s13059-014-0550-8

Luo, D., Wang, Z., Li, Z., and Yu, X. (2018). Structure of an entangled heteropolysaccharide from pholidota chinensis lindl and its antioxidant and anti-cancer properties *Int. J. Biol. Macromol.* 112, 921–928. doi: 10.1016/j.ijbiomac.2018.02.051

Medincine, N. J. (2006). *Dictionary of traditional Chinese medicine* (Shanghai: Science Press), 837.

Naikoo, M. I., Dar, M. I., Raghib, F., Jaleel, H., Ahmad, B., Raina, A., et al. (2019). Role and regulation of plants phenolics in abiotic stress tolerance. *Plant Signaling Mol.* 9, 157–168. doi: 10.1016/b978-0-12-816451-8.00009-5

Rai, A., Hirakawa, H., Nakabayashi, R., Kikuchi, S., Hayashi, K., Rai, M., et al. (2021). Chromosome-level genome assembly of ophiorrhiza pumila reveals the evolution of camptothecin biosynthesis. *Nat. Commun.* 12, 405. doi: 10.1038/s41467-020-20508-2

Ramya, M, Jang, S, An, HR, Lee, SY, Park, PM, Park, PH, et al. (2020). Volatile Organic Compounds from Orchids: From Synthesis and Function to Gene Regulation. *Int. J Mol. Sci.* 21 (3)

Ren, M., Xua, W., Zhang, Y., Ni, L., Lin, Y., Zhang, X., et al. (2020). Qualitative and quantitative analysis of phenolic compounds by UPLC-MS/MS and biological activities of pholidota chinensis lindl. *J. Pharm. Biomed. Anal.* 187, 1–11. doi: 10.1016/j.jpba.2020.113350

Rueda, D. C., Schoffmann, A., De Mieri, M., Raith, M., Jahne, E. A., Hering, S., et al. (2014). Identification of dihydrostilbenes in pholidota chinensis as a new scaffold for GABAA receptor modulators. *Bioorg. Med. Chem.* 22, 1276–1284. doi: 10.1016/j.bmc.2014.01.008

Salmela, L., and Rivals, E. (2014). LoRDEC: accurate and efficient long read error correction. *Bioinformatics* 30, 3506–3514. doi: 10.1093/bioinformatics/btu538

Shan, T., Yin, M., Wu, J., Yu, H., Liu, M., Xu, R., et al. (2021). Comparative transcriptome analysis of tubers, stems, and flowers of gastrodia elata blume reveals potential genes involved in the biosynthesis of phenolics. *Fitoterapia* 153, 104988. doi: 10.1016/j.fitote.2021.104988

Shimzu, K., Adachi, J., and Muraoka, Y. (2006). ANGLE_ a sequencing errors resistant program for:predicting protein coding regions in unfinished cDNA. *J. Bioinf. Comput. Biol.* 4, 649–664. doi: 10.1142/s0219720006002260

Song, S. S., Tao, Y., Gao, L., Liang, H., Tang, D., Lin, J., et al. (2022). An integrated metabolome and transcriptome analysis reveal the regulation mechanisms of flavonoid biosynthesis in a purple tea plant cultivar. *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.880227

Sun, M., Zhao, Y., Shao, X., Ge, J., Tang, X., Zhu, P., et al. (2021). doi: 10.21203/rs.3.rs-415397/v1

Tao, J., Luo, Z. Y., Msangi, C. I., Shu, X. S., Wen, L., Liu, S. P., et al. (2009). Relationships among genetic makeup, active ingredient content, and place of origin of the medicinal plant gastrodia tuber. *Biochem. Genet.* 47, 8–18. doi: 10.1007/s10528-008-9201-7

Ti, H., Zhuang, Z., Li, Y., Wei, G., and Wang, F. (2020). Three new phenanthrenes from pholidota chinensis lindl. and their antibacterial activity. *Natural Product Res.* 11, 1–7. doi: 10.1080/14786419.2020.1845168

Tsai, C. C., Wu, K. M., Chiang, T. Y., Huang, C. Y., Chou, C. H., Li, S. J., et al. (2016). Comparative transcriptome analysis of gastrodia elata (Orchidaceae) in response to fungus symbiosis to identify gastrodin biosynthesis-related genes. *BMC Genomics* 17, 212. doi: 10.1186/s12864-016-2508-6

Wang, S., Bilal, M., Hu, H., Wang, W., and Zhang, X. (2018). 4-hydroxybenzoic acid-a versatile platform intermediate for value-added compounds. *Appl. Microbiol. Biotechnol.* 102, 3561–3571. doi: 10.1007/s00253-018-8815-x

Wang, Y., Gao, Y. J., Li, X., Sun, X. L., Wang, Z. Q., Wang, H. C., et al. (2020). Coniferyl aldehyde inhibits the inflammatory effects of leptomeningeal cells by suppressing the JAK2 signaling. *BioMed. Res. Int.*, 2020, 4616308, 12. doi: 10.1155/2020/4616308

Wang, X. Y., Li, L., Zhu, H., et al (2016). Advance studies on plants of pholidota chinensis. *Asia-Pacfic Tradit. Med.* 12, 42–44. doi: 10.11954/ytcty.201601016

Wang, J., Matsuzaki, K., and KiItanakaba, T. (2006). Stilbene derivatives from pholidota chinensis and their antiinflammatory activity. *Chem. Pharm. Bull.* 54, 1216–1218. doi: 10.1248/cpb.54.1216

Wang, C., Xu, Q., Chen, Z., Hou, J., Li, H., Zhang, X., et al. (2022). Quality evaluation of gastrodia elata in different producing areas. *Cjhinese Tradit. Patent Med.* 44, 487–492. doi: 10.3969/j.issn.1001-1528.2922.02.028

Want, E. J., Wilson, I. D., Gika, H., Theodoridis, G., Plumb, R. S., Shockcor, J., et al. (2010). Global metabolic profiling procedures for urine using UPLC-MS. *Nat. Protoc.* 5, 1005–1018. doi: 10.1038/nprot.2010.50

Weng, S. (2006). Advances in pholidota chinensis. *Modern Chin. Med.* 8, 35–36. doi: 10.13313/j.issn.1673-4890.2006.06.013

Wen, B., Mei, Z., Zeng, C., and Liu, S. (2017). metaX: a flexible and comprehensive software for processing metabolomics data. *BMC Bioinf.* 18, 183. doi: 10.1186/s12859-017-1579-y

Yang, F., Li, G., Lin, B., and Zhang, K. (2022). Gastrodin suppresses pyroptosis and exerts neuroprotective effect in traumatic brain injury model by inhibiting NLRP3 inflammasome signaling pathway. *J. Integr. Neurosci.* 21, 72. doi: 10.31083/j.jin2102072

Yang, H., Wu, Y., Gan, C., Yue, T., and Yuan, Y. (2016). Characterization and antioxidant activity of a novel polysaccharide from pholidota chinensis lindl. *Carbohydr Polym.* 138, 327–334. doi: 10.1016/j.carbpol.2015.11.071

Yao, S., Tang, C. P., Ye, Y., Tibor, K., Attila, K. S., Sándor, A., et al. (2008) Stereochemistry of atropisomeric 9,10-dihydrophenanthrene dimers from pholidota chinensis *Tetrahedron: Asymm.* 19 (17) doi: 10.1016/j.tetasy.2008.08.013

Yin, H., Hu, T., Zhuang, Y., and Liu, T. (2020). Metabolic engineering of saccharomyces cerevisiae for high-level production of gastrodin from glucose. *Microb. Cell Fact.* 19 (1). doi: 10.1186/s12934-020-01476-0

Yuan, Y., Jin, X., Liu, J., Zhao, X., Zhou, J., Wang, X., et al. (2018). The gastrodia elata genome provides insights into plant adaptation to heterotrophy. *Nat. Commun.* 9, 1615. doi: 10.1038/s41467-018-03423-5

Zeki, O. C., Eylem, C. C., Recber, T., Kir, S., and Nemutlu, E. (2020). Integration of GC-MS and LC-MS for untargeted metabolomics profiling. *J. Pharm. BioMed. Anal.* 190, 113509. doi: 10.1016/j.jpba.2020.113509

Zhang, C., Chen, J., Huang, W., Song, X., and Niu, J. (2021). Transcriptomics and metabolomics reveal purine and phenylpropanoid metabolism response to drought stress in dendrobium sinense, an endemic orchid species in hainan island. *Front. Genet.* 12. doi: 10.3389/fgene.2021.692702

Zhang, M., Chen, L., Zhu, H., Li, L., Da, F., Long, L., et al. (2019). Establishment of HPLC fingerprint of pholidota chinensis and its cluster analysis. *China Pharm.* 30, 1792–1795. doi: 10.6039/j.issn.1001-0408.2019.13.13

Zhang, Z. L., Gao, Y. G., Zang, P., Gu, P. P., Zhao, Y., He, Z. M., et al. (2020). Research progress on mechanism of gastrodin and p-hydroxybenzyl alcohol on central nervous system. *China J. Chin. Mater. Med.* 45, 312–320. doi: 10.19540/j.cnki.cjcmm.20190730.401

Zhang, D., Song, Y. H., Dai, R., Lee, T. G., and Kim, J. (2020). Aldoxime metabolism is linked to phenylpropanoid production in camelina sativa. *Front. Plant Sci.* 11. doi: 10.3389/fpls.2020.00017

Zheng, Y., Jiao, C., Sun, H., Rosli, H. G., Pombo, M. A., Zhang, P., et al. (2016). iTAK: A program for genome-wide prediction and classification of plant transcription factors, transcriptional regulators, and protein kinases. *Mol. Plant* 9, 1667–1670. doi: 10.1016/j.molp.2016.09.014

Zou, Z., Huang, P., Zeng, H., Xu, Q., Li, J., and Liu, Y. (2017). The research progress of Chinese pholidota pseudobulb or herb's biological activity function. *Guangdong Chem. Industry* 44, 61–62. doi: 1007-1865 (2017)02-0061-02