

DATABASE

Open Access



# DhuFAP: a platform for gene functional analysis in *Dendrobium huoshanense*

Qiaoqiao Xiao<sup>1</sup>, Qi Pan<sup>1</sup>, Jun Li<sup>1</sup>, Jinqiang Zhang<sup>1\*</sup> and Jiaotong Yang<sup>1\*</sup>

## Abstract

**Background** *Dendrobium huoshanense*, a traditional medicinal and food plant, has a rich history of use. Recently, its genome was decoded, offering valuable insights into gene function. However, there is no comprehensive gene functional analysis platform for *D. huoshanense*.

**Result** To address this, we created a platform for gene function analysis and comparison in *D. huoshanense* (DhuFAP). Using 69 RNA-seq samples, we constructed a gene co-expression network and annotated *D. huoshanense* genes by aligning sequences with public protein databases. Our platform contained tools like Blast, gene set enrichment analysis, heatmap analysis, sequence extraction, and JBrowse. Analysis revealed co-expression of transcription factors (C2H2, GRAS, NAC) with genes encoding key enzymes in alkaloid biosynthesis. We also showcased the reliability and applicability of our platform using Chalcone synthases (CHS).

**Conclusion** DhuFAP ([www.gzybioinformatics.cn/DhuFAP](http://www.gzybioinformatics.cn/DhuFAP)) and its suite of tools represent an accessible and invaluable resource for researchers, enabling the exploration of functional information pertaining to *D. huoshanense* genes. This platform stands poised to facilitate significant biological discoveries in this domain.

**Keywords** Platform, *Dendrobium huoshanense*, Gene function analysis, Gene Set enrichment analysis

## Background

*Dendrobium huoshanense*, a traditional medicinal and food homologous plant, is a member of the Orchidaceae family and has a rich history of medicinal use [1]. It is commonly employed for its beneficial effects on the stomach, fluid production, heat clearance, and yin nourishment [2, 3]. Previous studies have demonstrated the diverse activities of *D. huoshanense*, including immunoregulation, anti-oxidation, anti-cataract, anti-glycation, anti-aging, anti-tumor, anti-rheumatoid arthritis,

anti-atherosclerosis, anti-inflammation, hypoglycemic, and liver protection activities [4–6]. The therapeutic effects are primarily attributed to active substances such as flavonoids, alkaloids, sesquiterpenes, and especially polysaccharides, which serve as the quality evaluation index for *D. huoshanense* [7, 8].

The advancement of high-throughput sequencing technology has significantly expanded research methods in the field of life sciences. This technology not only enhances the efficiency of scientific research but also drives the progress of basic research. Whole genome sequencing has been successfully accomplished in model plants and crops, with many species now possessing gene function analysis platforms that integrate multiple omics data. For instance, Tian et al. developed the MCENet platform [9], which included extensive Zea mays gene co-expression networks constructed from transcriptomic

\*Correspondence:

Jinqiang Zhang  
552450374@qq.com

Jiaotong Yang  
y\_jiaotong@163.com

<sup>1</sup>Resource Institute for Chinese and Ethnic Materia Medica, Guizhou University of Traditional Chinese Medicine, 550025 Guizhou, China



© The Author(s) 2024. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

data, as well as gene function analysis tools, facilitating the study of gene function and interactions between different genes. More recently, Wang et al. analyzed genomics data from 13 species in 9 genera of the Malvaceae family and established a functional genomic hub for Malvaceae plant [10], including genome-wide association analysis (GWAS) and single nucleotide mutation site (SNP) information, along with 374 sets of transcriptomic and proteomic data.

Currently, there are few analysis platforms that include the genetic information and functionality of *D. huoshanense*. The IMP provides genome information for *D. huoshanense* [11]. However, it lacks information such as expression data, co-expression networks, and other transcriptome-related details. Many databases are not suitable for gene functional analysis of *D. huoshanense*. For instance, essential plant databases like Phytozome [12] do not include the genome and transcriptome data for *D. huoshanense*. As *D. huoshanense* possesses active ingredients with significant pharmacological effects, exploring the genes regulating these active components is crucial for researchers to obtain detailed gene information using existing platforms. Therefore, it is essential to develop a gene function analysis platform for *D. huoshanense* by integrating various annotations. Such a platform will contribute to deeper gene function analysis and exploration in this species.

In 2020, the whole genome sequencing of *D. huoshanense* was successfully completed [13]. This achievement has led to the accumulation of valuable transcriptome data for *D. huoshanense*. To fully utilize and leverage this data, we curated transcriptome data obtained from the Sequence Read Archive (SRA) at the National Center for Biotechnology Information (NCBI) and the Genome Sequence Archive (GSA) at the National Genomics Data Center (NGDC) [14]. We constructed a comprehensive co-expression network of *D. huoshanense*. Additionally, we have developed a gene function analysis platform for *D. huoshanense*, named DhuFAP. This platform incorporates various analysis tools, including BLAST, GSEA, and JBrowse, and so on. These tools are designed to facilitate the exploration of novel gene functions in *D. huoshanense* and enable researchers to delve deeper into the molecular mechanisms underlying its unique characteristics.

## Materials and methods

### Data resource

The genomic data were sourced from the CNSA database's FTP public service, which included genome sequences, gene structure annotation files, protein sequences, and transcript sequences. Transcriptomic data were obtained from SRA and NGDC. Protein sequences from public platforms were downloaded from NCBI, Uniprot, and TAIR databases. KEGG and GO

annotation information was sourced from the KEGG database and agriGO v2. The EAR protein sequences, CAZy protein sequences, and transporter protein sequences in gene families were obtained from the Plant-EAR, CAZy database, and TransportDB, respectively.

### Function annotation

By utilizing the Diamond Blastp algorithm (v2.0.14.152) with the parameters "--evaluate 1E-3" and "--top 1" [15], the protein sequences of *D. huoshanense* were aligned to those present in public databases, including NR, Uniprot, SwissProt, and TAIR. The resulting annotation information was obtained from the best match identified in these databases. KEGG annotation was performed using the GhostKOALA website [16], and the predicted KEGG numbers were employed to retrieve annotation information from the KEGG database [17]. GO annotation and pfam domain information was accomplished through the InterProScan software [18], enabling the acquisition of GO numbers. Subsequently, the corresponding annotation information was downloaded from agriGO v2.0 [19] based on the obtained GO numbers.

### Co-expression network construction

Downloaded transcriptome samples were mapped to the reference genome of *D. huoshanense* using the Hisat2 software [20], resulting in alignment SAM files. Subsequently, SAM files were converted to BAM file and sorted using sam tools [21]. The stringtie software [22] was then employed to obtain the expression values for each transcriptome sample, enabling the construction of an expression matrix. Using the PCC algorithm, we calculated the correlation between gene expressions for every pair of genes. The gene correlations were subsequently ranked using the MR algorithm. The formula is as follows:

$$PCC = \frac{\sum(X - \bar{X})(Y - \bar{Y})}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2} \sqrt{\sum_{i=1}^n (Y_i - \bar{Y})^2}}$$

$$MR(AB) = \sqrt{Rank(AB) \times Rank(BA)}$$

In the given formulas, 'n' represents the total number of samples in the RNA-seq data, while 'x' and 'y' represent the TPM values. The term 'Rank' refers to the order of PCC values, where 'AB' signifies the ranking of gene A among all genes with gene B, and 'BA' indicates the reverse ranking.

Gene pairs in co-expression networks have similar expression patterns and may therefore have similar functions. Similar functionality can be evaluated by GO. The more similar the GO between co-expressed gene pairs,

the more reliable the co-expression network will be. We used co-expressed genes to assess whether the GO can be accurately predicted. If an accurate prediction is true, it cannot be accurately predicted to be false. We took these predictions as input to a binary classifier and calculated the true positive rate (TPR) and false positive rate (FPR), and then plotted the ROC curve. The greater the area under the curve (AUC) values, the better the prediction effect and the more robust the co-expression network. We identified Gene Ontology (GO) terms associated with biological processes, with a particular focus on those exhibiting gene counts ranging from 4 to 20. We evaluated the areas under the ROC curve (AUC) at different thresholds. By comparing the AUC values, we determined the property PCC and MR thresholds.

#### Protein-protein interaction (PPI) network

The construction of the PPI network for *D. huoshanense* involved the use of the OrthoFinder software [23] to predict orthologous relationships between *Arabidopsis* and *D. huoshanense*. Subsequently, the PPI network was mapped from *Arabidopsis* to *D. huoshanense*, establishing the PPI network in *D. huoshanense*.

#### Gene family identification

Initially, a hidden Markov model obtained from iUUCD 2.0 [24] successfully identified ubiquitin families in *D. huoshanense*. The log-odds likelihood scores parameter was from the threshold recommended by the iUUCD 2.0 [24]. OrthoFinder [23] was employed for the prediction of orthologous relationships between *Arabidopsis* and *D. huoshanense* with the default parameter. Following this, identification of proteins with TP, CAZy and proteins with EAR motifs were carried out utilizing the established orthologous relationship. The iTAK software [25] was utilized to identify transcription factors and protein kinases in *D. huoshanense* and the command was “iTAK.pl+protein\_sequence”. The complete genome underwent KEGG pathway annotation through the utilization of GhostKOALA [16]. Moreover, an analysis of the functional annotations for CYP450 genes was carried out, utilizing the information provided by KEGG annotations.

#### Construction of DhuFAP

The platform was built using the LAMP (Linux, Apache, MySQL, PHP) technical stack as its foundation. A MySQL database was created by importing various results and data analyses, such as gene structure annotation, co-expression network, gene functional annotation, PPI network, and gene family information. To enhance data visualization, responsive websites were created by employing a combination of HTML, PHP, JavaScript, and CSS programming languages.

#### Toolkit for gene function analysis

We integrated Gene Set Enrichment Analysis (GSEA) [26], building upon previous descriptions [27–29]. We also incorporated JBrowse software [30], a tool developed by Buels et al., to display transcriptome data and blast tools [31] to find similar sequences. Furthermore, we introduced a sequence extraction tool using a Perl script and implemented a Heatmap analysis tool based on Highchart Javascript. These additions expanded the capabilities of the platform and improved the visualization and analysis of data.

### Result

#### Gene functional annotation

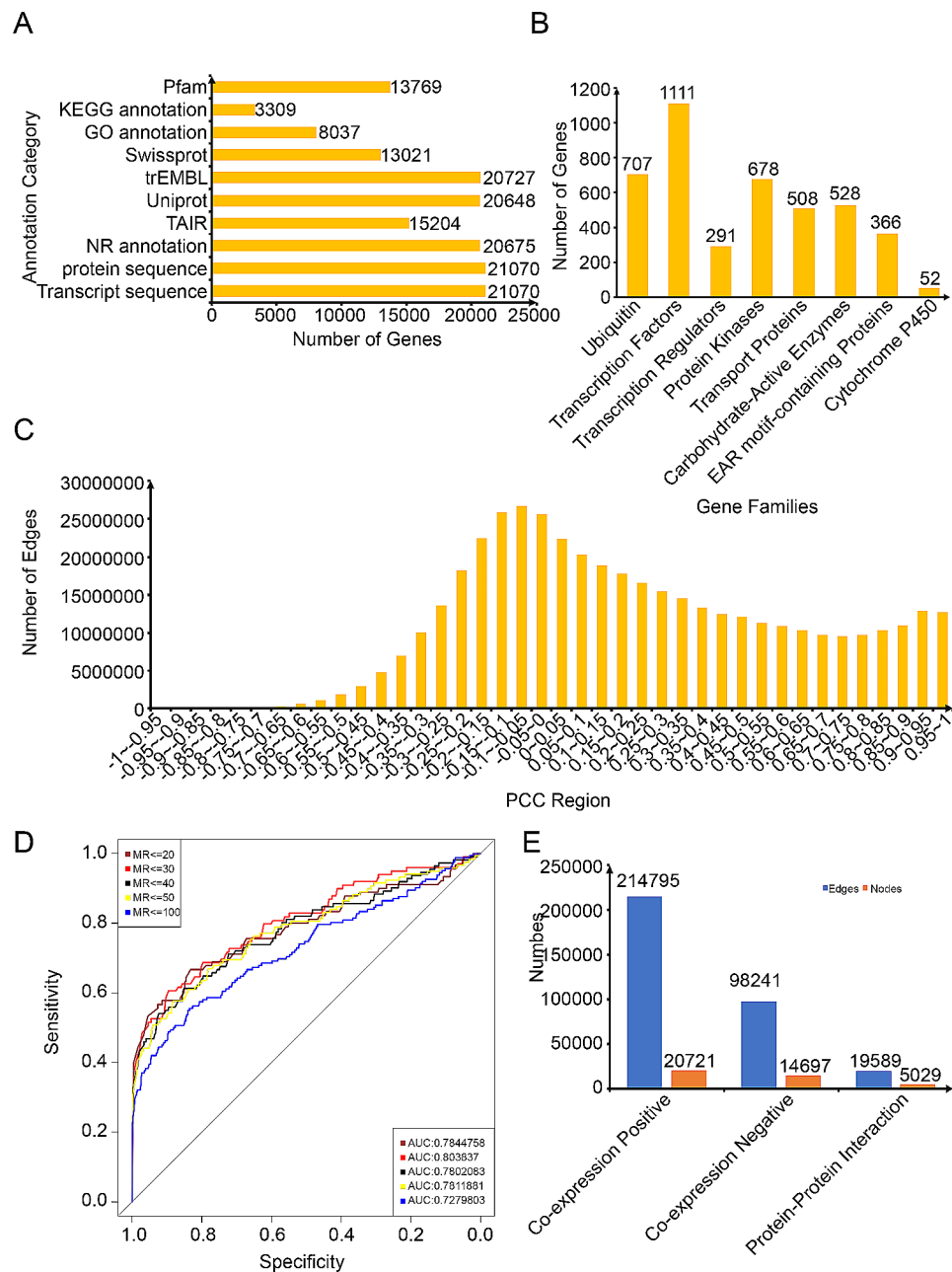
We obtained the genome data of *D. huoshanense* from NGDC, which included a comprehensive dataset of 21,070 transcripts and 21,070 proteins. To ensure accurate annotation, we subjected these resources to alignment with the protein sequences against well-known databases such as NR, Uniprot, TAIR, trEMBL, and Swissprot. Consequently, we annotated a total of 20,675, 20,648, 15,204, 20,727, and 13,021 genes, respectively. Furthermore, we utilized InterProScan software to conduct Gene Ontology (GO) annotations on a total of 8,037 genes [18]. For a comprehensive understanding of functional pathways, we utilized the GhostKOLAL [16] online tools to map KEGG annotation onto a set of 3,309 genes. Lastly, we conducted functional characterization of protein domains using the PfamScan software [32] to provide a comprehensive understanding of proteins (Fig. 1A).

#### Gene family classification

Initially, using iTAK software, we identified 1,111 transcription factors (TFs), 291 transcription regulators (TRs), and 678 protein kinases (PKs) in *D. huoshanense* respectively. Subsequently, utilizing the HMM profile derived from the ubiquitin-proteasome dataset within the iUUCD v2.0 database, we predicted 707 genes accountable for encoding components within the ubiquitin-proteasome system. Additionally, through gene alignment with databases like PlantEAR, TransprotDB, and CAZy, we effectively pinpointed 366 EAR genes, 508 genes to the Transprot family, and 528 genes categorized under the CAZy family. In addition, KEGG annotation allowed us to anticipate the existence of 52 Cytochrome P450 genes (Fig. 1B). These analyses provided valuable insights into the transcriptional regulation, protein kinase activity, ubiquitin-proteasome system, and gene families present in *D. huoshanense*.

#### Construction of co-expression network

We collected 69 transcriptome samples from SRA and NGDC, encompassing data from diverse tissues (roots,



**Fig. 1** Related statistical information of DhuFAP. **(A)** Gene function annotation information provided by DhuFAP. **(B)** Gene family classification information available. **(C)** The relationship between Pearson correlation coefficient (PCC) and the number of edges in the co-expression network. **(D)** Distribution of Area Under the Curve (AUC) values at different Matural Rank (MR) thresholds. **(E)** Statistical analysis of nodes and edges in the positive co-expression network, negative co-expression network, and Protein-Protein Interaction (PPI) network

stems, leaves) under normal growth conditions and various treatments (drought, low temperature, MeJA) under environment stress. Then we constructed a co-expression network by utilizing these transcriptome data, which were subsequently mapped to the reference genome with a mapping ratio exceeding 60% (Table S1). We analyzed the Pearson correlation coefficient (PCC) values obtained from expression profiles to identify co-expressed gene pairs. Many gene pairs showed no or weak correlation

in their expression patterns (Fig. 1C). To pinpoint gene pairs closely linked within each other's network, we used the MR (Matural Rank) method based on their PCC ranking values.

To ensure the reliability of our constructed network, we selected GO terms associated with similar biological activities, resulting in 120 terms with varying gene counts ranging from 4 to 20. We compared the area under the curve (AUC) values for different PCC (0.6, 0.7, 0.8, 0.9),

considering the overlap between positively co-expressed genes and the previously selected GO gene sets. We observed non-significant differences in AUC values among the PCC networks. To encompass a broader set of genes, we opted for a PCC threshold of  $>0.6$  (Figure S1). We further examined the area under the curve (AUC) values across various MR thresholds with the constraint of  $\text{PCC} > 0.6$ . This analysis led us to establish a network threshold of  $\text{MR} < 30$  for the positive co-expression network. The thresholds for the negative co-expression network were set at  $\text{PCC} < -0.5$  and  $\text{MR} < 30$  (Fig. 1D). The resulting co-expression network for *D. huoshanense* consisted of 313,036 co-expression gene pairs. This contained 214,795 gene pairs in the positive co-expression network and 98,241 gene pairs in the negative co-expression network (Fig. 1E).

### Protein–protein interaction network

By predicting the orthologous genes between *Arabidopsis* and *D. huoshanense*, we mapped the protein-protein interaction (PPI) network of *Arabidopsis* onto *D. huoshanense*. This resulted in the identification of 19,589 pairs of PPI relationships, involving a total of 5,029 genes (Fig. 1E).

### DEGs in different transcriptome

In order to incorporate gene co-expression and protein-protein interaction (PPI) networks with gene expression data, we performed differential expression analysis on the transcriptome data (student's t-test ( $p < 0.05$ ) and fold change [ $|\log_2(\text{foldchange})| > 1$ ]), resulting in the identification of differentially expressed genes (DEGs) across five sets of data. Through this process, we obtained a total of 35 distinct groups of DEGs (Table S2).

### Platform content

To enhance gene functional analysis in *D. huoshanense*, a comprehensive platform called DhuFAP has been developed. DhuFAP comprises seven sections—Home, Network, Pathway, Tools, Gene Family, Download, and Help—aimed at enhancing user-friendliness and delivering valuable insights to researchers (Fig. 2). Within the Network section, users can access both protein-protein interaction (PPI) and co-expression networks, offering comprehensive insight into the intricate molecular interactions within *D. huoshanense*. Pathway section consists of gene annotations from the KEGG database. The Gene Family section contains diverse protein families like CYP450, TF, TR, PK, TP, Ubiquitin, GAZy, and EAR motif-containing proteins. The Tools section offers various helpful features. The Blast tool screens nucleic acid or protein sequences for similarities within our database. GSEA enables comprehensive gene set enrichment analysis. The Extract Sequence tool retrieves gene sequences

using accession numbers and locations. Moreover, the Heatmap Analysis tool visually displays gene expression data. The inclusion of JBrowse provides an intuitive visualization of genomic and transcriptomic features. Download section provides convenient access to relevant information, ensuring easy retrieval of necessary resources. Furthermore, the Help section offers a comprehensive user manual, guiding researchers through the platform's functionalities and optimizing their usage of DhuFAP.

### Function application

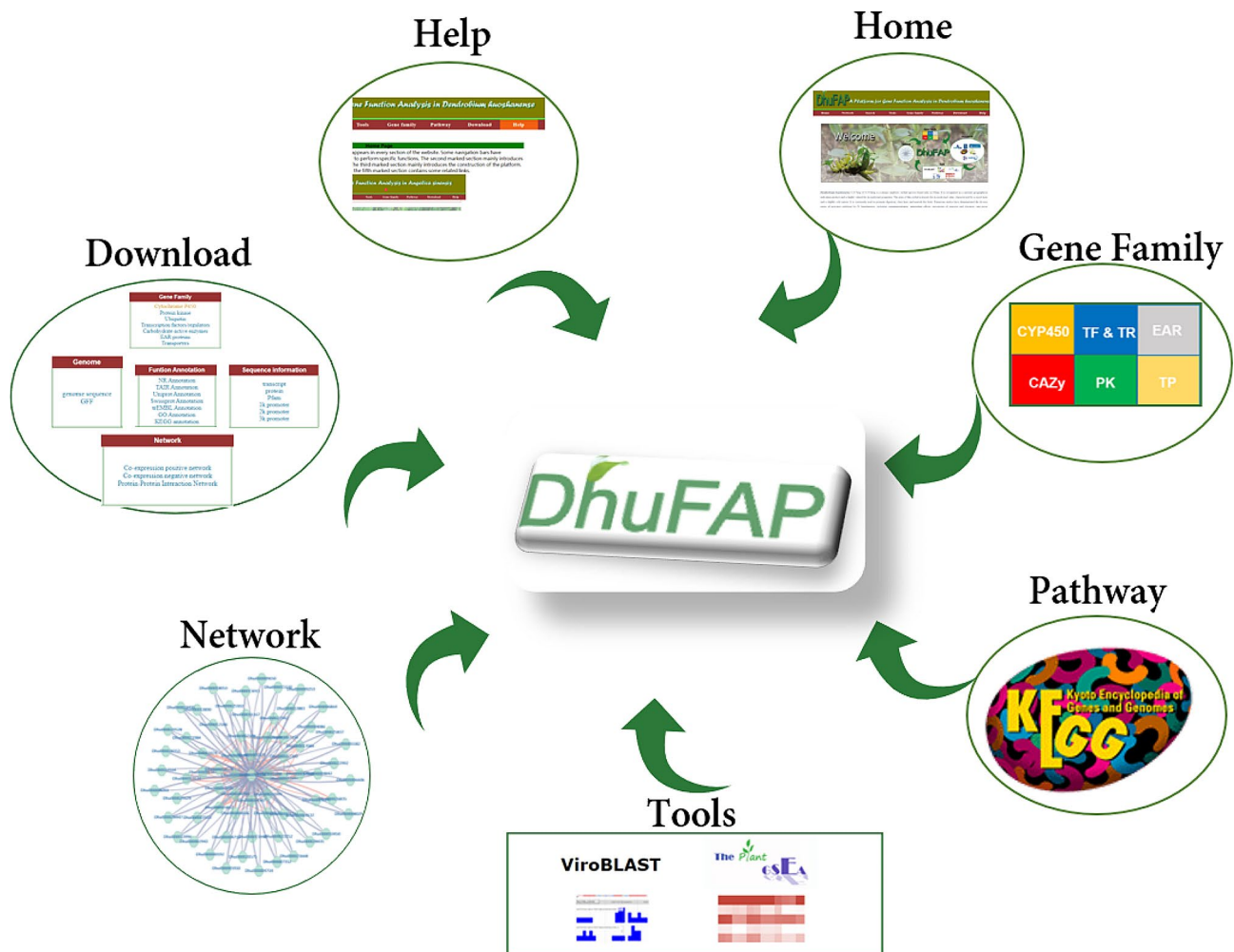
#### *Analysis of key enzyme genes in alkaloid biosynthesis pathway*

The stem contains an alkaloid that is the primary bioactive component in *D. huoshanense*. According to KEGG annotation in DhuFAP, there were 34 genes associated to alkaloid biosynthesis pathways were screened (Table S3). In order to better understand the relationship between key enzyme genes in alkaloid biosynthesis and TFs, co-expression analysis was conducted to identify the TFs which expressions were correlated with the key enzyme genes. The result demonstrated that C2H2, GRAS, NAC and other TFs were co-expressed with these key enzyme genes (Figure S2 and Table S4). Co-expressed genes have the same expression pattern, and may be regulated by the same upstream transcription factors. To explore transcription factors that may bind to the key enzyme gene promoter regions, we analyzed the co-expression relationships among key enzymes. The result showed that there were four pairs of co-expression between the key enzymes (Figure S3). We extracted 3000 bp sequences from the promoter region of each co-expression module to predict transcription factor binding sites. Many transcription factor binding sites were found, including MYB, GRAS and C2H2. This suggests that these transcription factors may bind to the key enzyme promoter regions. Therefore, these transcription factors may play a crucial regulatory role in the biosynthesis of alkaloids.

#### *Characteristic and functional analysis of CHS gene*

Chalcone synthases (CHS) are key enzyme that catalyzes alkaloid biosynthesis [33]. In our platform, the gene *Dhu000001149* was identified as a member of chalcone synthase family (Fig. 3A), spans from 7,498,800 to 7,500,079 bp on chromosome 11 (Fig. 3B). Additionally, co-expression network connections were also furnished (Fig. 3C). It was found that Chalcone and stilbene synthases domain was located at N-terminal and C-terminal in the protein sequence (Fig. 3C). KEGG annotation suggested that enzyme (Fig. 3D and E) involved in tropane, piperidine and pyridine alkaloid biosynthesis and flavonoid biosynthesis. Previous studies have identified CHS involved in the potential accumulation of alkaloid in *D.*





**Fig. 2** The structure of DhuFAP framework consisted of seven primary sections. The Home section served as an introduction to this platform. The Network section encompassed the co-expression network and PPI network. The Gene Family section comprised various gene families such as CYP450 family genes, transcription factors, transcription regulators, protein kinases, ubiquitin proteasomes, CAZy genes, Transport Proteins, and EAR motif-containing proteins. The Tools section offered functionalities like Search, BLAST, JBrowse, Sequence extraction, and GSEA toolkit. Pathway, Download, and Help were presented as separate sections

*huoshanense* [34]. Through expression profiling analysis, we found that the expression level of this gene was higher in stem and leaf compared to root (Fig. 3F). The display of reads mapping using JBrowse also revealed higher expression in leaf and stem (Fig. 4A). Furthermore, the accumulation of alkaloid significantly higher in stem, leaf than root [7]. The expression of this gene showed a similar trend to the synthesis and accumulation of active compounds. Therefore, the analysis results suggest that the gene may be involved in the accumulation of alkaloid.

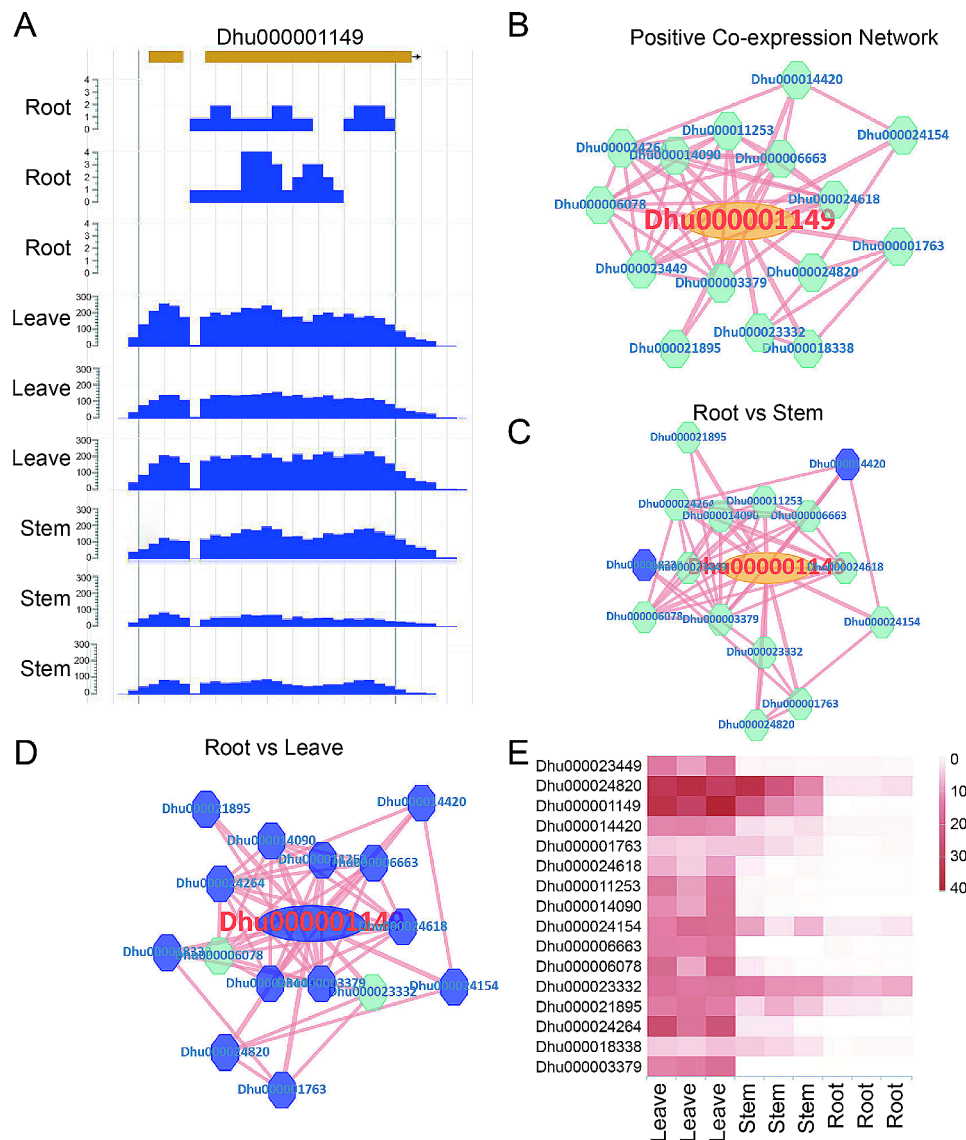
Furthermore, we conducted a co-expression analysis of CHS with its expression profiles. Network analysis revealed 15 genes that showed positive co-expression with CHS (Fig. 4B, Table S4). Additionally, many genes in the co-expression network were significantly upregulated in the leaves and stem (Fig. 4C and D). Analysis of co-expressed genes with CHS through heatmap analysis also

revealed similar results (Fig. 4E). Therefore, our analysis suggests that the CHS gene plays an important role in regulating biosynthesis of alkaloid.

#### Comparative transcriptome analysis

In order to uncover potential key regulatory factors involved in alkaloid biosynthesis, this study conducted an analysis on a transcriptome dataset (CRA000551), including leaves and root with 3 replicates. Using student's t-test ( $p < 0.05$ ) and fold change  $[|\log_2(\text{foldchange})| > 1]$ , we identified 1633 up-regulated and 4387 down-regulated genes by comparing the transcriptomes of roots and leaves (Table S2). Using the GSEA analysis tool provided by the platform, we performed GO enrichment analyses for the up-regulated and down-regulated genes. For the up-regulated genes, there was a significant enrichment of genes related to protein folding and





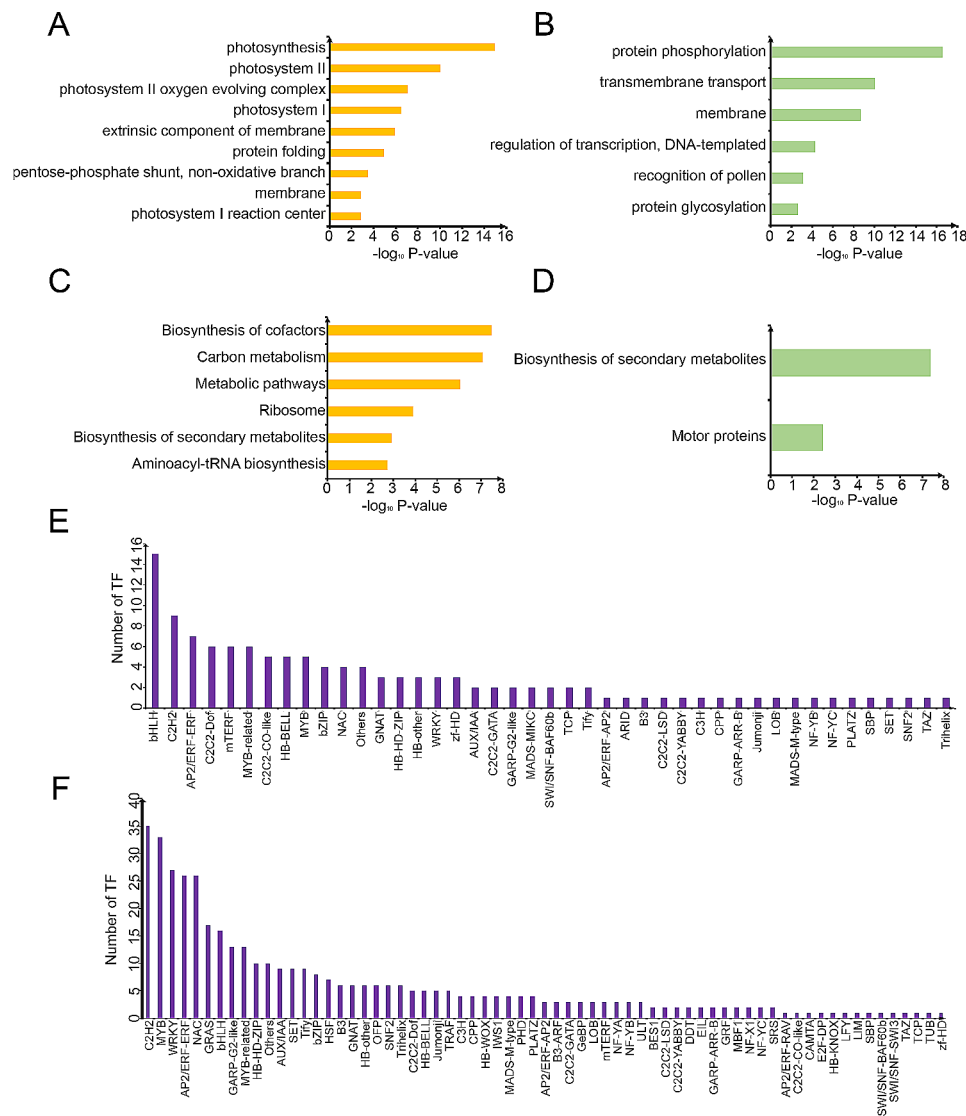
**Fig. 4** Expression and co-expression network analysis of CHS. **(A)** Presentation of CHS gene expression using JBrowse. **(B)** Positive co-expression gene network of CHS. **(C)** Analysis of gene differential expression in the positive co-expression network when comparing root and stem transcriptomes. **(D)** Analysis of gene differential expression in the positive co-expression network when comparing root and leaves transcriptomes. **(E)** Comparative analysis of CHS co-expressed genes' expression in different transcriptome samples using heatmap analysis tool

genes in DhuFAP. These examples showcase key steps such as filtering out important genes from the co-expression network, performing gene enrichment analysis, and interpreting regulatory networks. They not only highlight the capabilities of the platform but also offer practical guidance for users to conduct their own analyses.

DhuFAP serves as a powerful tool for researchers to delve into the functional genes and related biological processes of *D. huoshanense*. By integrating co-expression networks and offering various analysis tools, along with detailed usage examples, we are committed to advancing *D. huoshanense* research and providing valuable resources for scientists in related fields.

While DhuFAP offers valuable features and tools, we acknowledge potential limitations and areas for improvement. We predicted the protein interaction network of *D. huoshanense*. The more protein interaction pairs that have been experimentally confirmed, the more accurate our PPI predictions will be. Compared the predicted interaction protein with those have been reported, the more overlapping the predicted interaction protein is, the more reliable predicted protein interaction network is. We investigated the literature of *D. huoshanense* and found that no protein interaction research of *D. huoshanense* had been reported. If relevant research reported in the future studies, we will further evaluate the predicted protein interaction network.





**Fig. 5** Analysis of root and leaf samples from a set of transcriptome data (CRA000551). **(A)** Results of GO enrichment analysis for significantly upregulated genes. **(B)** Results of GO enrichment analysis for significantly downregulated genes. **(C)** Results of KEGG enrichment analysis for significantly upregulated genes. **(D)** Results of KEGG enrichment analysis for significantly downregulated genes. **(E)** Presence of transcription factors in significantly upregulated genes. **(F)** Presence of transcription factors in significantly downregulated genes

Currently, the platform relies on existing gene expression datasets, and ensuring data quality and coverage remains a challenge. In the future, we plan to expand the scale and diversity of the dataset to provide more comprehensive and accurate analysis results. Additionally, we aim to refine and expand the analysis tools and functionalities of the platform. This involves continuous improvement, introducing new analysis methods and algorithms, and staying updated on the latest research advancements in the field of *D. huoshanense*.

**Supplementary Information**

The online version contains supplementary material available at <https://doi.org/10.1186/s12864-024-10220-6>.

Supplementary Material 1

Supplementary Material 2

**Author contributions**

The project was designed by J.Y., with primary contributions in project completion and drafting by Q.X., Q.P and J.L. were involved in framework construction, while J.Z. provided technical guidance and assisted in paper revisions. All authors have consented to the main content of the article.

**Funding**

This work was supported by the Research Project of Guizhou Dendrobium Industry Development Center (Qian ShikêHé [2019006]), National Natural Science Foundation of China (NO.32160139 and NO.32260140), Guizhou University of Traditional Chinese Medicine Undergraduate Innovation and Entrepreneurship Training Program Project [2021]72, the University Science and Technology Innovation Team of the Guizhou Provincial Department of Education ([2023]071), the Guizhou Provincial Science and Technology

Projects (ZK[2022]505), and the National and Provincial Scientific and Technological Innovation Talent Team of the Guizhou University of Traditional Chinese Medicine (GZYTDHZ[2022]003)

#### Data availability

All the data we used are sourced from public platforms. The genome sequences analysed during the current study are public available in China National GeneBank (CNCB) (<https://ftp.cngb.org/pub/CNSA/data3/CNP0000830/CNS0251991/CNA0014590/>). Transcriptome data are publicly available in Sequence Read Archive (SRA) database (Accession no: SRP122499, SRP151171, SRP225982, SRP268245, SRP291861 and SRP406621) and GSA database in National Genomics Data Center (NGDC) (Accession no: CRA000551, CRA005817 and CRA006607).

#### Declarations

##### Ethics approval and consent to participate

Not applicable.

##### Consent for publication

Not applicable.

##### Competing interests

The authors declare no competing interests.

Received: 8 December 2023 / Accepted: 13 March 2024

Published online: 04 April 2024

#### References

- Gao L, Wang F, Hou T, Geng C, Xu T, Han B, Liu D. *Dendrobium huoshanense* C.Z.Tang et S.J.Cheng: a review of its traditional uses, Phytochemistry, and Pharmacology. *Front Pharmacol*. 2022;13:920823.
- Hsieh YS, Chien C, Liao SK, Liao SF, Hung WT, Yang WB, Lin CC, Cheng TJ, Chang CC, Fang JM, et al. Structure and bioactivity of the polysaccharides in medicinal plant *Dendrobium huoshanense*. *Bioorg Med Chem*. 2008;16(11):6054–68.
- Wang Y, Luo JP, Wei ZJ, Zhang JC. Molecular cloning and expression analysis of a cytokinin oxidase (DhCKX) gene in *Dendrobium huoshanense*. *Mol Biol Rep*. 2009;36(6):1331–8.
- Zhu Y, Kong Y, Hong Y, Zhang L, Li S, Hou S, Chen X, Xie T, Hu Y, Wang X. Huoshanmycins A–C, New Polyketide dimers produced by Endophytic *Streptomyces* sp. HS-3-L-1 from *Dendrobium huoshanense*. *Front Chem*. 2021;9:807508.
- Zhang CC, Gao Z, Luo LN, Liang HH, Xiang ZX. [Comparative analysis of active components and transcriptome between autotetraploid and diploid of *Dendrobium huoshanense*]. *Zhongguo Zhong Yao Za Zhi*. 2020;45(23):5669–76.
- Hao JW, Liu XQ, Zang YJ, Chen ND, Zhu AL, Li LF, Shi MZ. Simultaneous determination of 16 important biologically active phytohormones in *Dendrobium huoshanense* by pressurized capillary electrochromatography. *J Chromatogr B Analyt Technol Biomed Life Sci*. 2021;1171:122612.
- Yuan Y, Yu M, Jia Z, Song X, Liang Y, Zhang J. Analysis of *Dendrobium huoshanense* transcriptome unveils putative genes associated with active ingredients synthesis. *BMC Genomics*. 2018;19(1):978.
- Hao JW, Chen Y, Chen ND, Qin CF. Rapid Detection of Adulteration in *Dendrobium huoshanense* using NIR Spectroscopy coupled with chemometric methods. *J AOAC Int*. 2021;104(3):854–9.
- Li H, Yang L, Miao J, Yu P, Ge F. MCE-Net: polyp segmentation with multiple branch series-parallel attention and channel interaction via edge distribution guidance. *Phys Med Biol* 2023, 68(13).
- Wang D, Fan W, Guo X, Wu K, Zhou S, Chen Z, Li D, Wang K, Zhu Y, Zhou Y. MaGenDB: a functional genomics hub for Malvaceae plants. *Nucleic Acids Res*. 2020;48(D1):D1076–84.
- Chen T, Yang M, Cui G, Tang J, Shen Y, Liu J, Yuan Y, Guo J, Huang L. IMP: bridging the gap for medicinal plant genomics. *Nucleic Acids Res*. 2024;52(D1):D1347–54.
- Goodstein DM, Shu S, Howson R, Neupane R, Hayes RD, Fazo J, Mitros T, Dirks W, Hellsten U, Putnam N, et al. Phytozome: a comparative platform for green plant genomics. *Nucleic Acids Res*. 2012;40(Database issue):D1178–1186.
- Han B, Jing Y, Dai J, Zheng T, Gu F, Zhao Q, Zhu F, Song X, Deng H, Wei P, et al. A Chromosome-Level Genome Assembly of *Dendrobium huoshanense* using long reads and Hi-C Data. *Genome Biol Evol*. 2020;12(12):2486–90.
- Chen T, Chen X, Zhang S, Zhu J, Tang B, Wang A, Dong L, Zhang Z, Yu C, Sun Y, et al. The genome sequence Archive Family: toward Explosive Data Growth and Diverse Data types. *Genomics Proteom Bioinf*. 2021;19(4):578–83.
- Buchfink B, Xie C, Huson DH. Fast and sensitive protein alignment using DIAMOND. *Nat Methods*. 2015;12(1):59–60.
- Kanehisa M, Sato Y, Morishima K. BlastKOALA and GhostKOALA: KEGG Tools for functional characterization of genome and metagenome sequences. *J Mol Biol*. 2016;428(4):726–31.
- Kanehisa M, Goto S, Sato Y, Kawashima M, Furumichi M, Tanabe M. Data, information, knowledge and principle: back to metabolism in KEGG. *Nucleic Acids Res*. 2014;42(Database issue):D199–205.
- Jones P, Binns D, Chang HY, Fraser M, Li W, McAnulla C, McWilliam H, Maslen J, Mitchell A, Nuka G, et al. InterProScan 5: genome-scale protein function classification. *Bioinformatics*. 2014;30(9):1236–40.
- Tian T, Liu Y, Yan H, You Q, Yi X, Du Z, Xu W, Su Z. agriGO v2.0: a GO analysis toolkit for the agricultural community, 2017 update. *Nucleic Acids Res*. 2017;45(W1):W122–9.
- Kim D, Paggi JM, Park C, Bennett C, Salzberg SL. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat Biotechnol*. 2019;37(8):907–15.
- Etherington GJ, Ramirez-Gonzalez RH, MacLean D. Bio-samtools 2: a package for analysis and visualization of sequence and alignment data with SAMtools in Ruby. *Bioinformatics*. 2015;31(15):2565–7.
- Pertea M, Pertea GM, Antonescu CM, Chang TC, Mendell JT, Salzberg SL. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat Biotechnol*. 2015;33(3):290–5.
- Emms DM, Kelly S. OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol*. 2019;20(1):238.
- Zhou J, Xu Y, Lin S, Guo Y, Deng W, Zhang Y, Guo A, Xue Y. iUUCD 2.0: an update with rich annotations for ubiquitin and ubiquitin-like conjugations. *Nucleic Acids Res*. 2018;46(D1):D447–53.
- Zheng Y, Jiao C, Sun H, Rosli HG, Pombo MA, Zhang P, Banf M, Dai X, Martin GB, Giovannoni JJ, et al. iTAK: a program for genome-wide prediction and Classification of Plant Transcription Factors, transcriptional regulators, and Protein Kinases. *Mol Plant*. 2016;9(12):1667–70.
- Yi X, Du Z, Su Z. PlantGSEA: a gene set enrichment analysis toolkit for plant community. *Nucleic Acids Res*. 2013;41(Web Server issue):W98–103.
- Yang J, Yan H, Liu Y, Da L, Xiao Q, Xu W, Su Z. GURFAP: a platform for gene function analysis in *Glycyrrhiza Uralensis*. *Front Genet*. 2022;13:823966.
- Yu J, Zhang Z, Wei J, Ling Y, Xu W, Su Z. SFGD: a comprehensive platform for mining functional information from soybean transcriptome data and its use in identifying acyl-lipid metabolism pathways. *BMC Genomics*. 2014;15:271.
- Yang J, Li P, Li Y, Xiao Q. GelfAP v2.0: an improved platform for Gene functional analysis in *Gastrodia Elata*. *BMC Genomics*. 2023;24(1):164.
- Buels R, Yao E, Diesh CM, Hayes RD, Munoz-Torres M, Helt G, Goodstein DM, Elsik CG, Lewis SE, Stein L, et al. JBrowse: a dynamic web platform for genome visualization and analysis. *Genome Biol*. 2016;17:66.
- Deng W, Nickle DC, Learn GH, Maust B, Mullins JI. ViroBLAST: a stand-alone BLAST web server for flexible queries of multiple databases and user's datasets. *Bioinformatics*. 2007;23(17):2334–6.
- El-Gebali S, Mistry J, Bateman A, Eddy SR, Luciani A, Potter SC, Qureshi M, Richardson LJ, Salazar GA, Smart A, et al. The pfam protein families database in 2019. *Nucleic Acids Res*. 2019;47(D1):D427–32.
- Yuan F, Yin X, Zhao K, Lan X. Transcriptome and metabolome analyses of *Codonopsis Convolvulacea Kurz* Tuber, Stem, and Leaf reveal the Presence of important metabolites and Key pathways Controlling their biosynthesis. *Front Genet*. 2022;13:884224.
- Feng WM, Liu P, Yan H, Yu G, Zhang S, Jiang S, Shang EX, Qian DW, Duan JA. Investigation of enzymes in the Phthalide Biosynthetic Pathway in *Angelica Sinensis* using Integrative Metabolite profiles and Transcriptome Analysis. *Front Plant Sci*. 2022;13:928760.

#### Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.