Xu et al. *Horticulture Research* (2021)8:16
https://doi.org/10.1038/s41438-020-00450-6

Horticulture Research
www.nature.com/hortres

ARTICLE Open Access

Integration of full-length transcriptomics and targeted metabolomics to identify benzylisoquinoline alkaloid biosynthetic genes in *Corydalis yanhusuo*

Dingqiao Xu[1], Hanfeng Lin[2], Yuping Tang[1], Lu Huang[1], Jian Xu[2], Sihui Nian[3] and Yucheng Zhao[2]

Abstract

Corydalis yanhusuo W.T. Wang is a classic herb that is frequently used in traditional Chinese medicine and is efficacious in promoting blood circulation, enhancing energy, and relieving pain. Benzylisoquinoline alkaloids (BIAs) are the main bioactive ingredients in *Corydalis yanhusuo*. However, few studies have investigated the BIA biosynthetic pathway in *C. yanhusuo*, and the biosynthetic pathway of species-specific chemicals such as tetrahydropalmatine remains unclear. We performed full-length transcriptomic and metabolomic analyses to identify candidate genes that might be involved in BIA biosynthesis and identified a total of 101 full-length transcripts and 19 metabolites involved in the BIA biosynthetic pathway. Moreover, the contents of 19 representative BIAs in *C. yanhusuo* were quantified by classical targeted metabolomic approaches. Their accumulation in the tuber was consistent with the expression patterns of identified BIA biosynthetic genes in tubers and leaves, which reinforces the validity and reliability of the analyses. Full-length genes with similar expression or enrichment patterns were identified, and a complete BIA biosynthesis pathway in *C. yanhusuo* was constructed according to these findings. Phylogenetic analysis revealed a total of ten enzymes that may possess columbamine-O-methyltransferase activity, which is the final step for tetrahydropalmatine synthesis. Our results span the whole BIA biosynthetic pathway in *C. yanhusuo*. Our full-length transcriptomic data will enable further molecular cloning of enzymes and activity validation studies.

Introduction

Corydalis yanhusuo W.T. Wang is a medicinal herb in the *Corydalis* genus of the Papaveraceae family. In traditional Chinese medicine, its dried tubers (called "Yan-Hu-Suo" or "Yuan-Hu") are frequently used as an important therapeutic agent, with well-recorded efficacy in promoting blood circulation, enhancing energy, and providing analgesia[1]. Recent pharmacological research has found that active compounds in *C. yanhusuo* can inhibit tumor cell proliferation[2]; attenuate acute, inflammatory and neuropathic pain[1]; reduce drug addiction[3,4]; and treat cardiovascular diseases[5–7]. Owing to the low productivity of tubers and the short life cycle of *C. yanhusuo*, plant material sources are limited, greatly impeding its further application. Synthetic biology, which harnesses bacteria as chemical synthesis factories, has been working to provide an alternative to circumvent this problem[8,9]. However, little is known about the biosynthetic pathways of active compounds in *C.*

Correspondence: Sihui Nian (niansihui@126.com) or
Yucheng Zhao (zhaoyucheng1986@126.com)
[1]Key Laboratory of Shaanxi Administration of Traditional Chinese Medicine for TCM Compatibility, State Key Laboratory of Research & Development of Characteristic Qin Medicine Resources (Cultivation), Shaanxi Key Laboratory of Chinese Medicine Fundamentals and New Drugs Research, and Shaanxi Collaborative Innovation Center of Chinese Medicinal Resources Industrialization, Shaanxi University of Chinese Medicine, 712046 Xi'an, Shaanxi, China
[2]Department of Resources Science of Traditional Chinese Medicines and State Key Laboratory of Natural Medicines, School of Traditional Chinese Pharmacy, China Pharmaceutical University, 210009 Nanjing, Jiangsu, China
Full list of author information is available at the end of the article
These authors contributed equally: Dingqiao Xu, Hanfeng Lin

© The Author(s) 2021
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 license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license 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 license, visit http://creativecommons.org/licenses/by/4.0/.

Downloaded from https://academic.oup.com/hr/article/doi/10.1038/s41438-020-00450-6/6446576 by guest on 04 August 2026

yanhusuo, especially its protoberberine- and aporphine-type benzylisoquinoline alkaloids (BIAs).

Characterization of the chemical composition of *C. yanhusuo* has been ongoing for nearly a century[10], and a series of compounds have been identified, including sugars, amino acid derivatives, triterpenes, anthraquinones, phenolic acids, steroids and organic acids[11]. Additionally, over 60 kinds of alkaloids have been identified in *C. yanhusuo* tubers[10–14], most of which are BIAs and produce the main pharmacological effects[15,16]. These compounds are all synthesized from tyrosine and undergo a common precursor pathway through the synthesis of (S)-reticuline before diverging into different BIA classes[17]. Among these BIAs, most can be classified as protoberberines and aporphines according to their chemical structures[14], but protopine-type and benzo[c]phenanthridine-type BIAs are also present in *C. yanhusuo* bulb extracts. Currently, thanks to the persistent efforts of biochemists and phytochemists, the labyrinthine BIA biosynthetic pathways have been well elucidated in the model species *Papaver somniferum* (opium poppy)[17,18], but the clarification of the aporphine-type BIA biosynthetic pathway will require substantial additional efforts[19]. Additionally, a few amino acid sequences of key enzymes in protoberberine BIA pathways are still missing, which greatly impedes the molecular cloning and metabolic engineering of protoberberines. Furthermore, little attention has been given to the BIA biosynthetic process in *Corydalis*, which has resulted in a lack of knowledge about the biosynthesis of several pharmacologically active alkaloids in *C. yanhusuo*, such as glaucine, corybulbine[20], and dehydrocorydaline[21].

To date, research has mainly focused on the pharmacological effects of *C. yanhusuo* extracts and the chemicals isolated from these extracts. Despite its medicinal importance, researchers have paid little attention to the genetics of this plant. Although there are reports regarding the transcriptomes of BIA-producing plants[22–26], short-read-length NGS without reference genomes generally depends on de novo assembly, which has relatively poor performance in complex regions with repetitive or heterozygous sequences and thus prevents the retrieval of full-length transcript information[27]. Moreover, although one previous study investigated the time-course of BIA production during bulb development[26], the spatial distribution of transcripts in *C. yanhusuo* remains unknown.

Here, we used single-molecule real-time (SMRT) sequencing on the PacBio Sequel platform, which produces long reads, to fully mine full-length transcriptome information in *C. yanhusuo*. The low raw accuracy of SMRT sequencing can be overcome by the use of NGS short reads to proofread SMRT[28]. SMRT sequencing can offer highly complete transcriptome data for further analysis[29,30], paving the way for reconstructing transcripts, detecting splicing events, and circumventing computational challenges in NGS data assembly[31]. Metabolic profiling or metabolomics can monitor and quantify metabolite variations in response to both endogenous and exogenous factors in biological systems[29,30]. Broad-scope metabolomics encompassing both primary and secondary metabolism has been largely restricted to model plant species. In this study, the spatial variation of *C. yanhusuo* transcripts in leaves and tubers provides evidence for BIA biosynthesis in tubers rather than in leaves, and high-resolution mass spectrometry applied for alkaloid analysis identified 19 metabolites involved in BIA biosynthesis in the leaves and tubers of *C. yanhusuo*. A targeted metabolomics study of 19 alkaloids was conducted for the BIA accumulation pathway, with special focus on the alkaloids whose content in tubers was four times higher than that in leaves. In tandem with our key full-length transcriptome analysis, this work elucidates the major BIA biosynthesis pathways and their transcriptional regulation among different organs in *C. yanhusuo*.

Results

Sequencing and functional annotation

To identify the genes involved in BIA synthesis in *C. yanhusuo*, we obtained the full-length transcriptome by performing SMRT sequencing on the PacBio Sequel platform. The statistics of the SMRT sequencing and comparison with previously reported Illumina sequencing[26] are listed in Table 1. It produced a total of 5,228,902 subreads (12.24 GB), with an average length of 2341 bp, and the N50 length of all subreads was 2619 bp, much longer than that of Illumina sequencing (Fig. 1a). To improve the quality of the subreads, a total of 225,223 circular consensus sequences (CCSs) were found. The FLNC (full-length nonchimeric) reads were later screened by identifying the coexistence of 5′-primers, 3′-primers and poly-A tails. We detected 184,584 FLNC reads (81.96% of CCS) with an average length of 2693 bp (Fig. 1b, c). In addition, six total RNA samples from tubers and leaves were sequenced using second-generation sequencing (SGS) on the Illumina platform, yielding 86,394 proofread consensus sequences ranging from 188 to 13,219 bp with a mean length of 2748 bp. After clustering the redundant and 95% similar transcript sequences, 49,585 unigenes were identified, among which 47,969 unigenes (96.74%) were over 1000 bp, 36,397 unigenes (73.40%) were over 2000 bp and 17,815 unigenes (35.93%) were over 3000 bp in length (Fig. 1d). To analyze the functions of the unigenes obtained from the full-length transcripts, we annotated the unigenes using the data from seven nucleotide and protein databases (NR, NT, Pfam, Swiss-Prot, KOG, KEGG, and GO) (Fig. S1). Out of 49,585 unigenes, 47,352 (95.50%) were annotated in at least one database, a much higher percentage than was

Downloaded from https://academic.oup.com/hr/article/doi/10.1038/s41438-020-00450-6/6446576 by guest on 04 August 2026

Table 1 Statistics of *C. yanhusuo* SMRT sequencing and previously reported Illumina results

	SGS[26]	SMRT
Total clean raw reads	25,013,630	5,228,902
Total nucleotides (bp)	2,251,226,700	236,495,561
GC percentage	44.65%	41.50%
Average read-length (bp)	90	2341
Total contigs	158,101	N.A.[a]
Length range (bp)	60–5,752	N.A.[a]
Mean contig length (bp)	198	N.A.[a]
Total scaffolds	65,839	86,394
Length range (bp)	100–7,977	188–13,219
Mean scaffold length (bp)	340	12,991
Total Unigenes	47,081	49,585
Total nucleotides (bp)	23,016,560	141,504,321
Length range (bp)	150–7,973	188–13,219
Mean unigene length (bp)	489	2853
N50 (bp)	681	3124
No. (%) of annotated unigenes ($e < 10^{-5}$)	30,868 (65.56%)	47,352 (95.50%)
NR	30,660 (65.12%)	46,586 (93.95%)
Swiss-Prot	20,664 (43.89%)	40,861 (82.41%)
KEGG	13,111 (27.85%)	45,821 (92.41%)
COG	9,412 (19.99%)	30,808 (62.13%)
GO	12,455 (26.45%)	31,242 (63.01%)

[a]SMRT sequencing produces full-length transcript reads, therefore contig assembly process is not applicable for SMRT sequencing

obtained for the previous Illumina sequencing annotation result (65.56%). In addition, using full-length transcript data, the gene structures and transcription factor binding sites were also predicted (Fig. S2). KEGG analysis found 530 unigenes related to the biosynthesis of other secondary metabolites (Table S1), but there remains a question that until now, benzylisoquinoline alkaloids were the only alkaloid class reported in *C. yanhusuo*[13,26,32,33]. This seemingly conflicts with the KEGG annotation result that 69 unigenes were annotated to the tropane/piperidine/pyridine alkaloid metabolism pathway. However, upon further inspection, we found that nearly half of these unigenes were primary amine oxidases, and one-third were aspartate aminotransferases, such as GOTs and ASP5, which have a broad range of substrate specificities and participate in various metabolic pathways. The remaining ten unigenes were annotated to tropinone reductase I in the tropane alkaloid pathway, and seven had a high BLAST identity to this enzyme. Considering that no other enzymes in this pathway were annotated, nor were any tropinone alkaloids reported in *C. yanhusuo*, the function of these predicted proteins needs to be carefully evaluated.

Multivariate analysis of Q-TOF-MS data—global metabolomics

To further evaluate the candidate genes, untargeted ultra-performance liquid chromatography quadrupole-time-of-flight mass spectrometry (UPLC-Q-TOF-MS) profiling was conducted to generate mass lists corresponding to a wide variety of metabolites, including alkaloids. Data processing was conducted with Progenesis QI software (Waters, Milford, MA, USA). Compound identification was based not only on accurate mass but also on matching of the MS/MS fragmentation patterns (Fig. S3). Nineteen of these compound annotations were further validated by comparison with authentic standards (Fig. S4). Tuber and leaf groups were analyzed by principal component analysis (PCA) and the orthogonal partial least-squares-discriminant analysis (OPLS-DA) model. In the score plots, the tuber and leaf groups were separated and clustered into two groups (Fig. 2a). The S-plots (Fig. 2b) and color-coded loading plots (Fig. 2c) for OPLS-DA revealed the variation in metabolites in the tuber and leaf compared with the leaf group in the *C. yanhusuo* extracts. The tuber and leaf groups showed clear separation in the OPLS-DA score plots of *C. yanhusuo* extracts with satisfactory goodness of fit and statistical significance. The S-plots and fold-change plots revealed the variation in metabolites in the tuber and leaf groups, including increased levels of scoulerine, tetrahydroberberine, tetrahydrocolumbamine, columbamine, tetrahydropalmatine, sanguinarine, corydaline, protopine, tetrahydrocoptisine, noroxyhydrastinine, dehydrocorydaline, oxoglaucine, 8-oxycoptisine, palmatine, coptisine, berberine, jatrorrhizine, epiberberine, and glaucine. The details of these metabolites are presented in the supporting information (Fig. S4, Fig. S5, and Fig. S6). Thirty-eight of the 53 annotated metabolites were found to be significantly different between leaves and tubers (Table 2). According to the spatial difference in BIA compound accumulation between the leaf and tuber, the total transcripts in each organ are useful sources to perform differential gene expression analysis, which helps in further identification of unigenes involved in BIA biosynthesis.

Differential expression of genes between *C. yanhusuo* tubers and leaves

To investigate the abundance of organ-specific transcripts, leaves and tubers were sampled for RNA-seq on the Illumina platform, and their abundance was evaluated by converting read counts to fragments per kilobase per million (FPKM). The distribution of gene

Downloaded from https://academic.oup.com/hr/article/doi/10.1038/s41438-020-00450-6/6446576 by guest on 04 August 2026

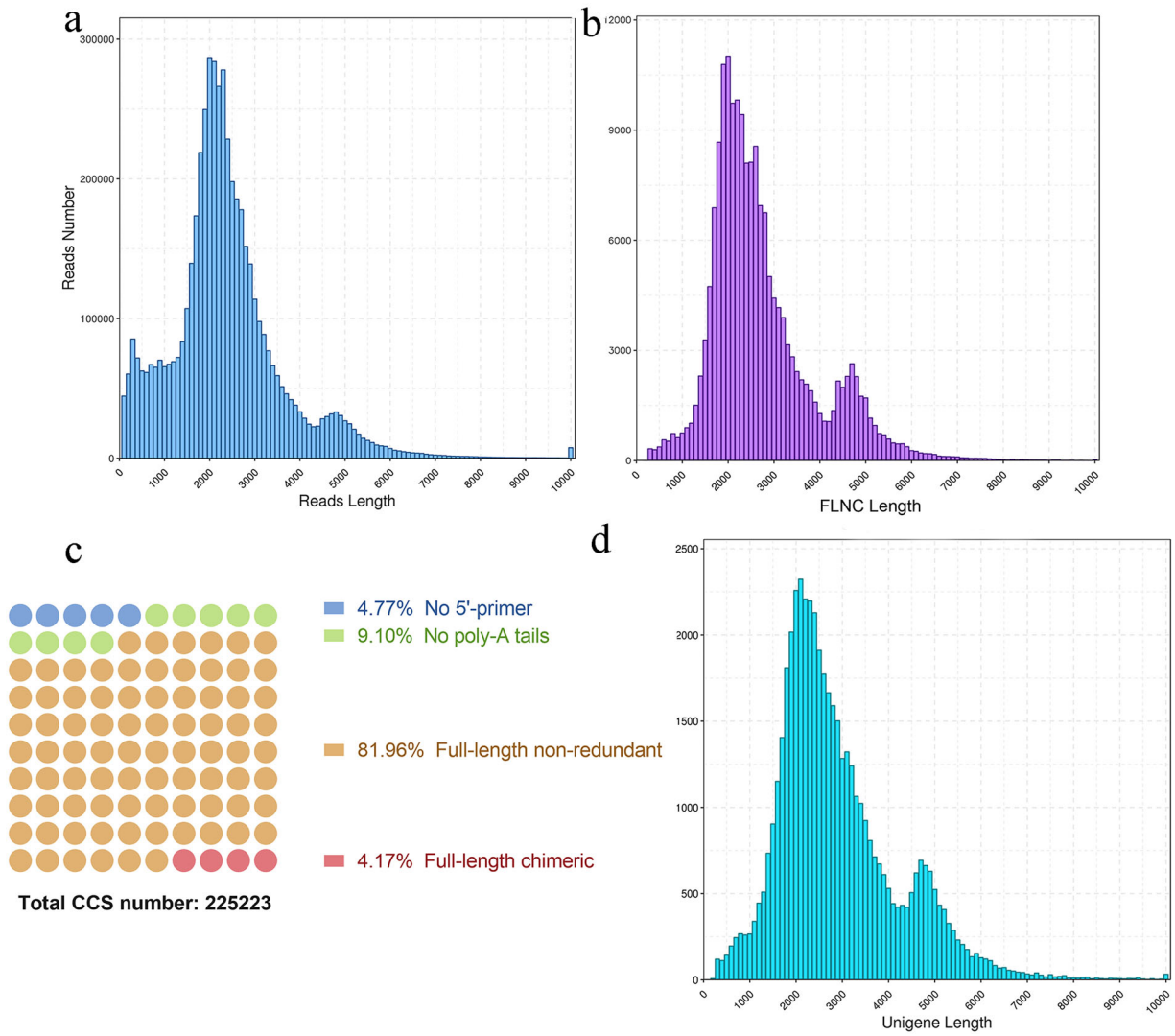

Fig. 1 The quality control data for the full-length transcriptome of *C. yanhusuo* obtained by SMRT sequencing on the PacBio sequel platform. a The distribution of raw data subread length for SMRT sequencing. **b** The length distribution of full-length nonchimeric (FLNC) reads containing 5′/3′-primers and poly-A tails. **c** The quality of circular consensus sequences (CCS) exhibited as the proportion of FLNC reads. **d** The length distribution of high-quality, full-length, and polished unigene sequences that were proofread by SGS reads

expression levels for six samples is plotted in Fig. 3a. Figure 3b shows the overall expression pattern of each unigene and their relative ratio for each FPKM interval in roots and leaves. The differentially expressed genes (DEGs) were filtered with a threshold of adjusted p-values < 0.05 and $|\log_2^{\text{FoldChange}}| > 0$, in which fold-change was defined as the read count in tubers divided by that in leaves. A total of 8794 unigenes (17.74%) were identified as DEGs (Table S2). A volcano plot was used to show the relationship between the adjusted p-value and $\log_2^{\text{FoldChange}}$ (Fig. 3c). GO term enrichment and pathway enrichment analysis presented in the form of a bubble plot provided certain information (Fig. 3d, e). The most enriched pathway was that of the photosynthesis-antenna proteins, followed by flavonoid biosynthesis and carotenoid biosynthesis, both of which are important secondary metabolites as plant pigments. As we predicted, isoquinoline alkaloid biosynthesis had an enrichment factor of 0.38 with a Q-value of 6.02×10^{-4}, revealing that isoquinoline alkaloid biosynthesis genes are mainly enriched in tubers compared with leaves.

Identifying candidate genes in the benzylisoquinoline biosynthetic pathway

KEGG analysis found that 530 unigenes were associated with secondary metabolite synthesis (Table S1). One-hundred thirty-four unigenes were annotated to participate

Downloaded from https://academic.oup.com/hr/article/doi/10.1038/s41438-020-00450-6/6446576 by guest on 04 August 2026

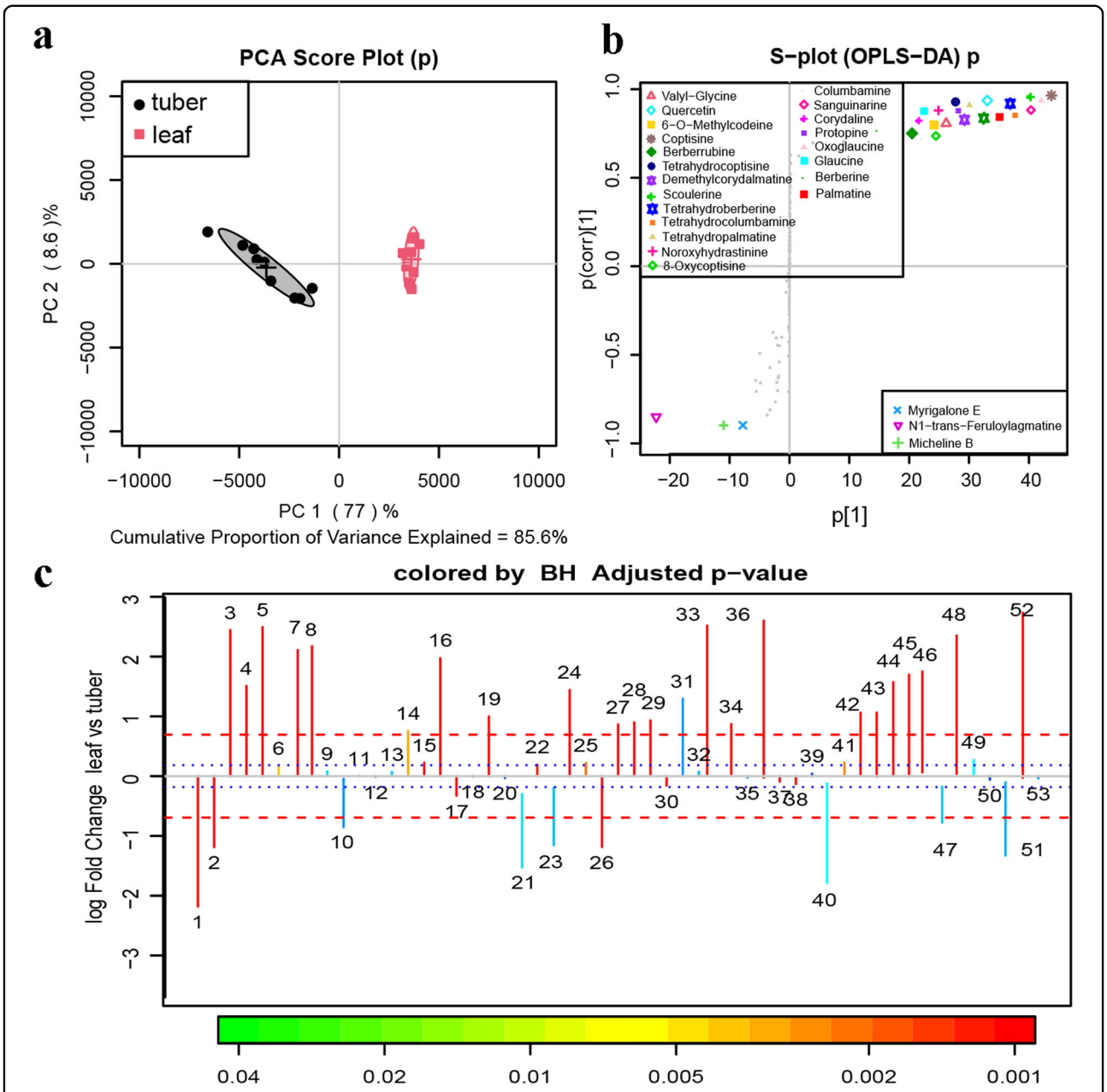

Fig. 2 Metabolomic multivariate analysis of QTOF-MS data of *C. yanhusuo* extracts for the tuber and leaf groups. a PCA score plots, **b** color-coded S-plots, and **c** fold-change plots in positive ion mode ($n = 9$). In the PCA score plots, the tuber and leaf groups were satisfactorily separated and clearly clustered into two groups. Fold-change plots color-coded according to the corresponding *p*-values adjusted by the Benjamini–Hochberg method indicate the significance of the altered metabolites in *C. yanhusuo* extracts. The blue dotted lines and red dashed lines represent an increase or decrease of 20% and 100%, respectively

in isoquinoline alkaloid biosynthesis, which is the most abundant alkaloid synthesis pathway. Additionally, Pfam annotation found 328 unigenes containing the cytochrome P450 conserved domain, which are known to play indispensable roles in BIA biosynthesis. As most of the reported secondary metabolites in *C. yanhusuo* are BIAs, here, to identify the candidate genes involved in the BIA biosynthetic pathway, we BLAST searched the nonredundant full-length sequence we retrieved from *C. yanhusuo* against the reported BIA biosynthetic enzyme homologs. The protein homologs mainly belong to species in the Papaveraceae family, including *Papaver somniferum*, *Macleaya cordata*, and *Eschscholzia californica*. Sequences from *Coptis japonica* and *Berberis stolonifera* were also used because certain enzymes

Downloaded from https://academic.oup.com/hr/article/doi/10.1038/s41438-020-00450-6/6446576 by guest on 04 August 2026

Table 2 Identified metabolites with fold changes between tuber and leaf groups and their *p*-values in *C. yanhusuo*

NO.	Metabolites	*m/z*	Formula	Mass Error (ppm)	Fold Change	*P* value
1	4-Hydroxy-3-methoxybenzoic acid	192.0817	$C_8H_8O_4$	-0.43	0.94	
2	Pontevedrine	382.0830	$C_{21}H_{19}NO_6$	-0.14	0.73	*
3	Valyl-Glycine	175.1088	$C_7H_{14}N_2O_3$	-0.84	0.91	
4	Micheline B	277.1661	$C_{17}H_9NO_3$	0.78	1.52	*
5	(S)-Scoulerine	328.1548	$C_{19}H_{21}NO_4$	0.81	1.25	*
6	N1-trans-Feruloylagmatine	307.1769	$C_{15}H_{22}N_4O_3$	0.72	0.82	
7	Tetrahydroprotopapaverine	330.1702	$C_{19}H_{23}NO_4$	0.61	1.72	**
8	6-O-Methylcodeine	314.1755	$C_{19}H_{23}NO_3$	0.17	1.08	
9	Tetrahydrocolumbamine	342.1705	$C_{21}H_{25}NO_4$	0.52	1.31	*
10	Protopine	354.1346	$C_{20}H_{19}NO_5$	0.09	2.05	**
11	Pseudo protopine	354.3540	$C_{20}H_{19}NO_5$	-0.53	1.91	**
12	Alpha-N-methylcanadine	355.1376	$C_{20}H_{20}NO_5^+$	0.08	1.19	
13	Bulbocapnine	326.1389	$C_{19}H_{19}NO_4$	-0.43	2.13	**
14	Cryptopin	370.1744	$C_{21}H_{23}NO_5$	-0.79	1.90	**
15	Coptisine	320.0918	$C_{19}H_{14}NO_4^+$	0.23	1.20	
16	Berberrubine	358.7875	$C_{19}H_{16}ClNO_4$	0.08	0.99	
17	Capaurine	372.1809	$C_{21}H_{25}NO_5$	0.23	1.31	
18	Tetrahydropalmatine	356.1848	$C_{21}H_{25}NO_4$	-0.84	11.04	***
19	Glauvent	356.4090	$C_{21}H_{25}NO_4$	0.34	1.26	
20	Hyndarin	356.3468	$C_{21}H_{25}NO_4$	-0.23	0.69	**
21	Tetrahydrocoptisine	324.1233	$C_{19}H_{17}NO_4$	0.28	2.08	**
22	Glaucine	354.1624	$C_{21}H_{25}NO_4$	-0.69	2.39	**
23	Demethylcorydalmatine	328.1911	$C_{19}H_{21}NO_4$	-0.24	7.20	***
24	Corydine	343.5000	$C_{20}H_{23}NO_4$	-0.78	1.40	*
25	Dehydronantenine	338.1389	$C_{20}H_{19}NO_4$	-0.23	1.97	*
26	Columbamine	339.1421	$C_{20}H_{20}NO_4^+$	-0.65	2.93	**
27	Nantenin	340.1545	$C_{20}H_{21}NO_4$	-0.80	1.21	
28	Corydaline	370.1579	$C_{22}H_{27}NO_4$	-0.01	4.23	**
29	N-methyllaurotetanine	342.1650	$C_{20}H_{23}NO_4$	-0.33	1.25	
30	(R)-Canadine	340.3708	$C_{20}H_{21}NO_4$	0.61	1.86	*
31	Berberine	336.1232	$C_{20}H_{18}NO_4$	-0.94	2.71	**
32	13-methylpalmatrubine	353.1581	$C_{21}H_{22}NO_4^+$	0.69	3.65	**
33	Palmatine	352.1546	$C_{21}H_{22}NO_4^+$	0.57	4.85	***
34	Dehydrocorydaline	367.1740	$C_{22}H_{24}NO_4$	0.90	1.17	***
35	Epiberberine	366.1688	$C_{20}H_{18}NO_4^+$	0.86	1.34	***
36	Oxoglaucine	352.1191	$C_{28}H_{38}N_2O_4$	0.65	2.54	***
37	Tetrahydrocorysamine	338.4094	$C_{20}H_{19}NO_4$	0.23	1.30	***
38	8-Oxycoptisine	336.3102	$C_{19}H_{13}NO_5$	-2.43	4.52	***
39	Noroxyhydrastinine	192.8362	$C_{10}H_9NO_3$	0.28	5.10	***
40	Dehydroglaucine	354.1660	$C_{21}H_{21}NO_4$	0.63	1.98	*
41	Corynoline	368.1499	$C_{21}H_{21}O_5$	0.17	1.18	***
42	Isocorybulbine	356.4280	$C_{21}H_{25}NO_4$	0.76	1.18	***
43	Scoulerine	328.3744	$C_{19}H_{21}NO_4$	0.77	1.19	*
44	Corynoloxine	366.3729	$C_{21}H_{19}NO_5$	-0.09	0.91	
45	Tetrahydroberberine	340.3947	$C_{20}H_{21}NO_4$	-0.69	3.53	**
46	Dihydrochelerythrine	350.1385	$C_{21}H_{19}NO_4$	0.85	2.46	**
47	Jatrorrhizine	339.3887	$C_{20}H_{20}NO_4$	0.85	1.96	*
48	Sanguinarine	333.3326	$C_{20}H_{14}NO_4^+$	0.97	2.37	**
49	Dihydrosanguinarine	334.1073	$C_{20}H_{15}NO_4$	0.72	1.00	
50	Stigmasterol	413.2665	$C_{29}H_{48}O$	0.57	0.82	
51	Dehydrocorybulbine	353.2662	$C_{21}H_{22}NO_4$	0.85	1.08	
52	Cavidine	354.2707	$C_{21}H_{23}NO_4$	0.88	0.76	*
53	Quercetin	303.2290	$C_{15}H_{10}O_7$	0.01	0.72	*

P-value: *$p < 0.05$, **$p < 0.01$, and ***$p < 0.001$. Color coded according to fold change using color bar

Downloaded from https://academic.oup.com/hr/article/doi/10.1038/s41438-020-00450-6/6446576 by guest on 04 August 2026

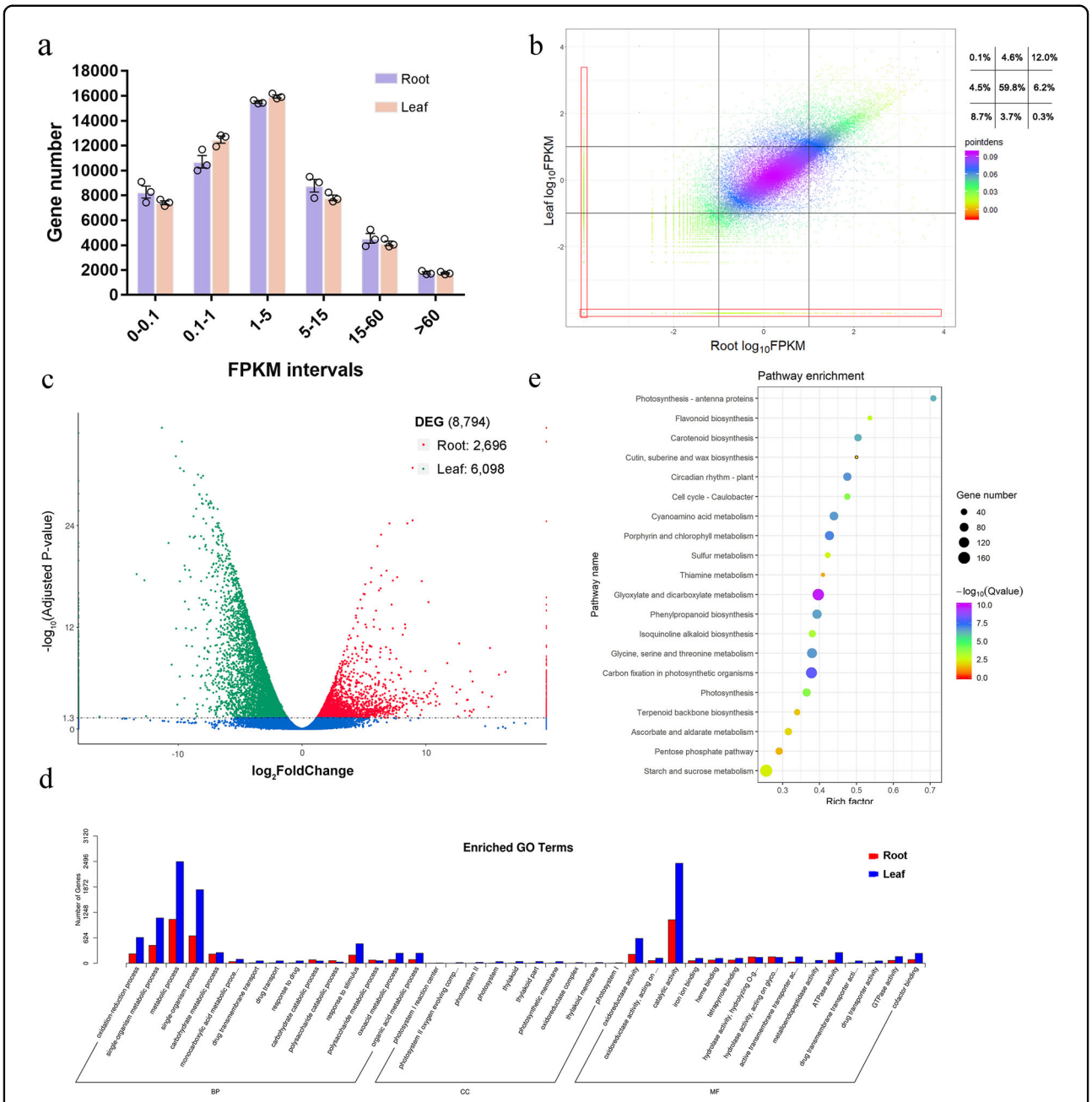

Fig. 3 Differential expression of genes between *C. yanhusuo* tubers and leaves. Leaves and tubers were sampled for RNA-seq on the Illumina platform, and their abundance was evaluated by converting read count to FPKM. The gene expression level distribution of six samples is elaborated. **a** Distribution of the number of genes within certain FPKM intervals. **b** Density scatter plot provides an overview of different expression patterns for each gene in different organs. The ratio number in each well indicates the proportion of unigenes in the corresponding frame. Points within the red boxes have FPKM values of 0 in either leaf or tuber. **c** Volcano plot showing the distribution of differentially expressed genes (DEGs) with a *p*-value threshold of 0.05. **d** GO term enrichment results for genes from the tuber and leaf. **e** Pathway enrichment analysis presented in the form of a bubble plot

(for example, columbamine O-methyltransferase and berbamunine synthase) are found exclusively in these plants[34].

We found a total of 101 nonredundant unigenes related to the enzymes in the BIA biosynthetic pathway according to the tblastn results (Fig. 4 and Table S3). All the enzymes in the common pathway (pathway in yellow in Fig. 4a or Fig. S7) succeeded in finding their corresponding unigenes, which proves that *C. yanhusuo*, like other species, uses a common pathway to provide the fundamental precursor (S)-reticuline for synthesizing downstream BIAs. Every

Downloaded from https://academic.oup.com/hr/article/doi/10.1038/s41438-020-00450-6/6446576 by guest on 04 August 2026

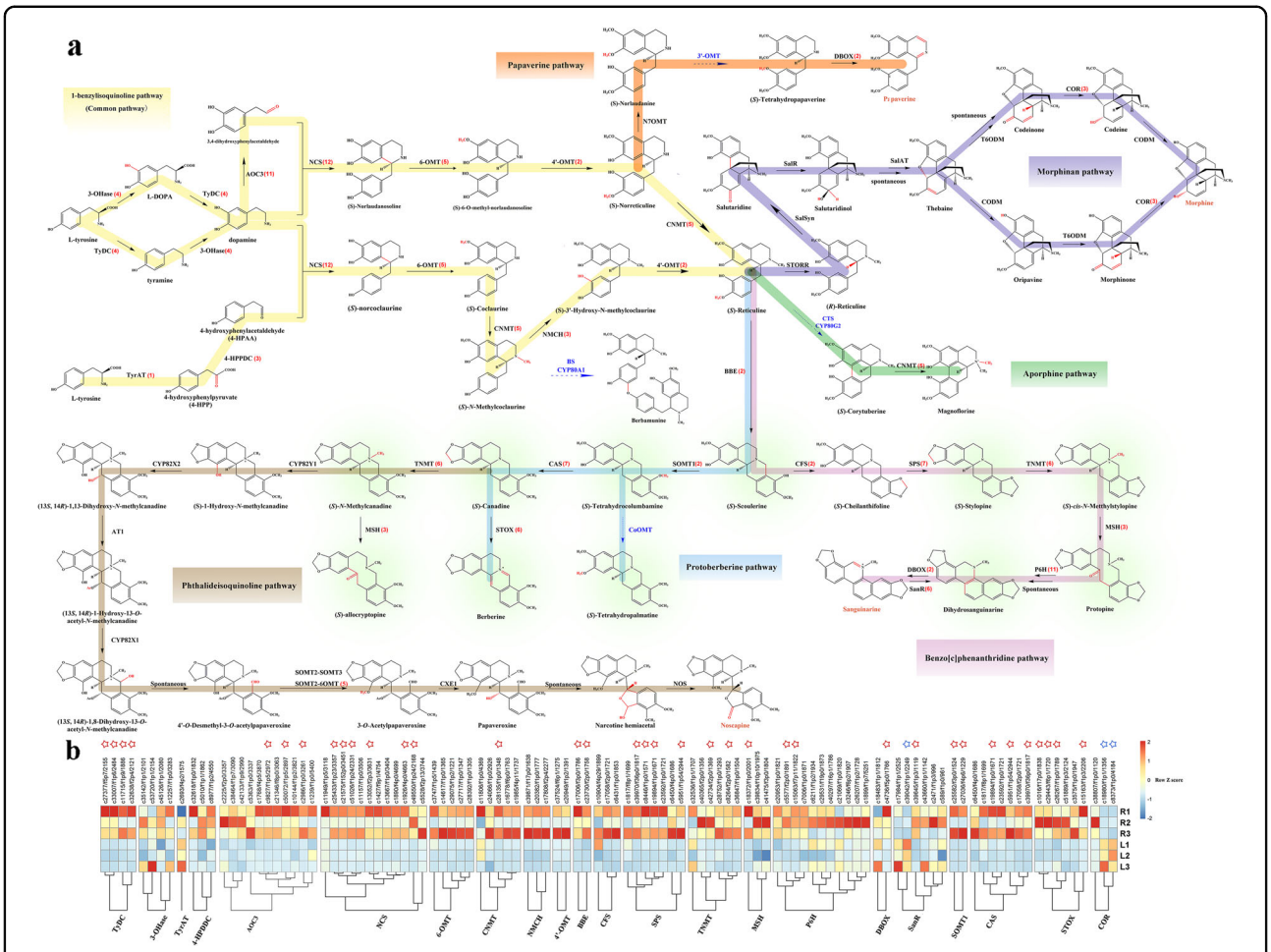

Fig. 4 A proposed benzylisoquinoline biosynthetic pathway. a A total of 101 nonredundant unigenes are involved in the BIA biosynthetic pathway according to the tblastn results. Highlighted compounds are those previously reported in *C. yanhusuo*. Red numbers after each enzyme are the number of unigenes blasted to their homologs with an identity over 60%. Enzymes catalyzing reactions with blue arrows have not been identified in the Papaveraceae family. **b** Expression profile of all BIA biosynthesis candidate genes. Three biological replicates from the leaf (L) and tuber (R) were plotted individually. Unigenes labeled by red asterisks are DEGs predominant in tubers, while those marked by blue asterisks are predominant in leaves. TyDC tyrosine/dopa decarboxylase, 3-OHase tyrosine 3-hydroxylase, TyrAT tyrosine aminotransferase, 4-HPPDC 4-hydroxyphenylpyruvate decarboxylase, NCS norcoclaurine synthase, 6OMT (S)-norcoclaurine 6-O-methyltransferase, CNMT (S)-coclaurine N-methyltransferase, NMCH (S)-N-methylcoclaurine 3′-hydroxylase; 4′OMT (S)-3′-hydroxy-N-methylcoclaurine 4′-O-methyltransferase, N7OMT norreticuline 7-O-methyltransferase, BBE berberine bridge enzyme; SOMT scoulerine 9-O-methyltransferase, CAS canadine synthase, STOX (S)-tetrahydroxyprotoberberine oxidase, CoOMT Columbamine O-methyltransferase, CFS cheilanthifoline synthase, SPS stylopine synthase, TNMT tetrahydroprotoberberine N-methyltransferase, MSH methylstylopine hydroxylase, P6H protopine 6-hydroxylase, DBOX dihydrobenzophenanthridine oxidase, STORR 1,2-dehydroreticuline synthase/reductase, SalSyn salutaridine synthase, SalR salutaridine reductase, SalAT salutaridinol 7-O-acetyltransferase, T6ODM thebaine 6-O-demethylase, COR codeinone reductase, CODM codeine O-demethylase, AT1 (13S,14R)-1,13-dihydroxy-N-methylcandine 13-O-acetyltransferase, CXE1 3-O-acetylpapaveroxine carboxylesterase, NOS noscapine synthase

enzyme in this pathway is indispensable. However, we did not find many hits in the morphinan pathway (purple) or the phthalideisoquinoline pathway (brown) (Fig. 4a), and none of the corresponding compounds has been reported in *C. yanhusuo*. (S)-Allocryptopine is synthesized at the beginning of the phthalideisoquinoline pathway, but it is a structural analog of protopine[35]. All the enzymes in the protoberberine pathway (blue) and the benzo[c]phenanthridine pathway (pink) can find their hits in the transcriptome, verifying that protoberberines and protopines are the representative chemical components in *C. yanhusuo*. We successfully found adequate unigenes required for synthesizing all the BIA compounds in *C. yanhusuo* according to current phytochemistry reports[11,14,36].

To investigate the expression pattern of the selected candidate genes, a heatmap for all the candidate genes was plotted for each leaf (L1-3) and tuber (R1-3) sample (Fig. 4b or Fig. S8). The heatmap reveals that the expression of BIA biosynthetic genes is well regulated in an organ-specific manner. Among them, 36 unigenes were previously

Downloaded from https://academic.oup.com/hr/article/doi/10.1038/s41438-020-00450-6/6446576 by guest on 04 August 2026

identified as DEGs (Fig. 4b, asterisks after unigene IDs, three of which reoccurred in SPS and CAS), indicating that most of them showed a higher level of expression in tubers (red asterisks). Most of the DEGs are involved in the benzo [c]phenanthridine pathway in BIA biosynthesis, which indicates that the upregulation of protopine- and protoberberine-type alkaloid biosynthetic genes in the tuber is a synchronized process. Meanwhile, typical enzymes such as 3-OHase, TyrAT, DBOX, and COR showed a vague regulation pattern between tubers and leaves; therefore, the reactions catalyzed by these enzymes may not be strictly regulated. Some variances among tuber samples (Fig. 4b, especially R2) may indicate that certain enzymes have their own sub-organ distribution.

Collectively, the expression profiling vindicates the hypothesis that the different expression levels of BIA biosynthetic genes between the two organs may the reason that tubers, not leaves, are used as the therapeutic agent in *C. yanhusuo*. These findings also suggest the possibility of a transcriptional regulation mechanism of BIA biosynthesis in *C. yanhusuo*.

Analyses of 19 marker compounds in synthesized BIAs-targeted metabolomics

A global metabolomics study revealed that a few alkaloids dominate among the synthesized BIAs. To further justify the reliability of the differentially expressed transcriptome information in Fig. 4b, we chose nineteen representative BIAs synthesized in *C. yanhusuo*, namely, scoulerine, canadine (also known as tetrahydroberberine), tetrahydrocolumbamine, columbamine, tetrahydropalmatine, sanguinarine, corydaline, protopine, tetrahydrocoptisine, noroxyhydrastinine, dehydrocorydaline oxoglaucine, 8-oxycoptisine, palmatine, coptisine, berberine, jatrorrhizine, epiberberine, and glaucine, as marker compounds to investigate their abundance in tubers and leaves. The total ion chromatograms (TICs) of the standard samples were obtained and are displayed in Fig. S4. The metabolite abundance of these compounds between the two groups was quantified through a subsequent targeted metabolomics analysis that was performed to identify candidate metabolites involved in BIA biosynthesis in *C. yanhusuo*. Differences could be observed in the contents of nineteen compounds in tubers and leaves (Fig. S9), which corresponded with the previous transcriptome analysis to some extent (Fig. 4b). This observation also corroborates the hypothesis that the accumulation of BIAs in tubers is related to the BIA biosynthetic gene expression pattern according to our previous analysis.

Sequence and phylogenetic analysis of O-methyltransferase candidates

Phylogenetic analysis was conducted to identify the corresponding O-methyltransferases (OMTs) because most OMT tblastn results shared high similarity, and some of the unigenes matched to several OMTs simultaneously. Another reason for our interest in OMTs is that one O-methylated BIA end-product, (S)-tetrahydropalmatine (THP), has great medicinal value as an analgesic or sedative without addictive side effects and can ameliorate opiate addiction[3,37]; therefore, efforts to identify OMTs that are able to catalyze such a reaction from (S)-tetrahydrocolumbamine to THP (CoOMT) have been ongoing for years. Here, we performed phylogenetic analysis to provide some insights into OMT protein family classification in *C. yanhusuo* (Fig. 5). The dendrogram from the phylogenetic analysis showed that all the OMT candidates clustered into 4 major clades: 6-OMT, 4-OMT, CoOMT, and SOMT. R7OMT is separated from all the presenting clades, which reveals the lack of homologous enzymes in *C. yanhusuo*. The analysis of the domain architectures conducted on SMART[38] reveals that most of the predicted polypeptide chains and all the OMTs from other plant species contain two domains: the C-terminal methyltransferase domain (in orange) is typical of SAM-dependent O-methyltransferases, and the N-terminal domain (in light or dark blue) is a dimerization domain that is found in a variety of plant OMTs. The dimerization domain not only stabilizes the homodimer through hydrophobic interactions but also provides an extra water-mediated hydrogen bond with the substrate to improve specificity[39]. However, some candidates do not possess a dimerization domain; thus, their ability to catalyze corresponding reactions is questionable.

Phylogenetic analysis found nine unigenes mapped closely to the 6-OMT clade, and most had a high tblastn score to Ps6-OMT. Three unigenes were classified as 4-OMTs. Six and five unigenes were mapped to the CoOMT and SOMT clades, respectively, which might help explain the accumulation of THP in *C. yanhusuo*, as they were both reported to catalyze the reaction from tetrahydrocolumbamine to THP[40,41]. It is possible that enzymes with CoOMT activity could appear from the translational products of these 10 unigenes in the CoOMT and SOMT clades (one with insufficient length was excluded). All the sequence information used in the phylogenetic analysis is provided in Table S4.

Discussion

The identification of BIA biosynthetic candidate genes from *C. yanhusuo* is a pivotal step in the investigation of the pharmaceutical value of *C. yanhusuo* and may provide additional information to guide BIA-related pathway elucidation (e.g., corydaline, dehydrocorybulbine). In this study, we generated the first full-length transcriptome for *C. yanhusuo* and investigated the transcripts related to BIA biosynthesis in both leaf and tuber tissue of *C. yanhusuo* by combining SMRT and SGS technologies. A

Downloaded from https://academic.oup.com/hr/article/doi/10.1038/s41438-020-00450-6/6446576 by guest on 04 August 2026

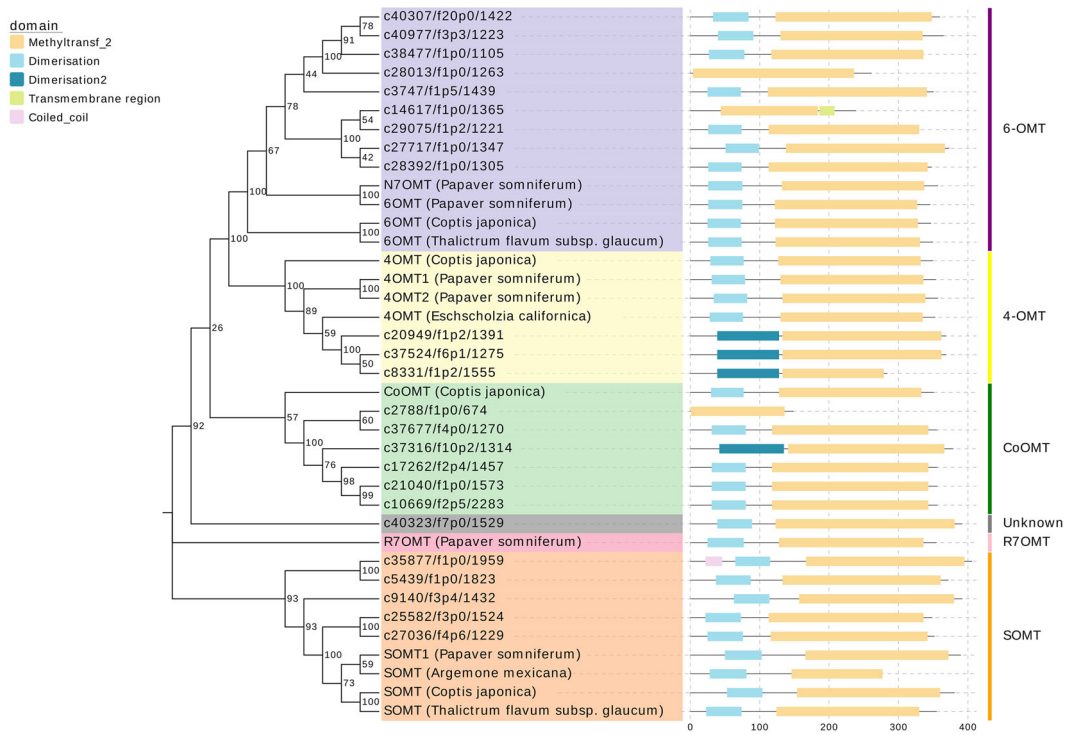

Fig. 5 Phylogenetic analysis was performed to investigate the OMT protein family classification in *C. yanhusuo*. The phylogenetic tree shows that all OMT candidates formed four major clades: 6-OMT, 4′-OMT, CoOMT, and SOMT

previous study focused on elucidating whether BIA biosynthetic candidate genes were transcriptionally regulated over the time-course of bulb development based on SGS short-read data[26]. Although the SGS results provided a general view of the BIA biosynthetic pathway in the *C. yanhusuo* bulb, full-length transcripts and transcriptome data from other tissues had not yet been obtained, which greatly impeded the molecular cloning and gene verification process.

In line with traditional Chinese medicine usage of *C. yanhusuo* tubers, the bioinformatic analysis revealed that isoquinoline alkaloid biosynthesis genes were mainly enriched in tubers rather than leaves. We then crudely found 101 unigenes related to BIA biosynthesis according to similarity with enzyme orthologs from other species. The presence of such genes in *C. yanhusuo*, as well as their relevance to alkaloid biosynthesis, was validated by conducting quantitative reverse transcription pCR (qRT-PCR) analysis between the control and methyl jasmonate (MeJA) treatment groups (Fig. S10). External application of MeJA is known to dramatically promote the accumulation of alkaloids and the expression of key genes[42–46]. In particular, MeJA was also reported to trigger in vitro accumulation of dihydrosanguinarine, a product in the BIA pathway, and the expression of key biosynthetic genes in *Eschscholzia californica* cell culture[42,47]. Thus, we expected that the external application of MeJA would increase the expression of BIA biosynthetic genes, and indeed, the fold changes shown in Fig. S10 indicate the relevance of these genes to alkaloid bioproduction. To clarify the distribution of BIAs in different tissues, 19 BIAs were evaluated in tubers and leaves and observed to be enriched in tubers. Thus, we could infer that the spatial distribution of such genes is likely responsible for the difference in BIA contents in tubers and leaves. However, such inference is based on the following presumptions: (i) no other genes are responsible for BIA biosynthesis, (ii) BIAs are locally synthesized in the tubers instead of being transported there, and (iii) BIAs do not undergo automatic degradation in the leaf tissue. To address points (i) and (iii), we used tblastn to crudely filter the BIA biosynthetic genes, which circumvented potential hit escape. Using the multidrug-resistant protein CjABCB1-3, which is involved in berberine translocation[48,49], as a query, a total of 21 full-length hits whose products may function in unloading BIAs from tubers and promote their transport to aerial parts of the plant were found. In addition, as considerable amounts of compounds can still be detected in the leaves (Table 2 and Fig. S9), given that the biosynthesis of BIAs in leaves is not as extensive as that in tubers, we believe that major BIAs in *C. yanhusuo* do not undergo significant degradation. This interpretation is in

Downloaded from https://academic.oup.com/hr/article/doi/10.1038/s41438-020-00450-6/6446576 by guest on 04 August 2026

line with the biological role of alkaloids, that is, to protect plants against insects or herbivores. There is no way for organisms to evolve a feature (here, alkaloid degradation) that is potentially detrimental to them. Related to presumption (ii), a report has already monitored BIA accumulation in tubers during development and concluded that the biosynthesis of THP showed a latent accumulation pattern, which might be the result of biosynthesis using nutrients transported from the aerial to the terrestrial parts[50]. This study ruled out the possibility that BIAs are synthesized elsewhere and then transported to the tuber. Considering that the leaf is the primary place where organic building blocks are generated through photosynthesis, a higher content of BIAs detected in tubers supports the assignment of the biosynthetic role to tubers but not leaves. Hence, it is quite convincing that the spatial difference in BIA contents is caused by BIA-related DEGs.

With respect to the identified unigene candidates (Fig. 5), 6 out of 12 NCS genes and all 4 TyDC genes were denoted as DEGs, indicating that the entrance of the 1-benzylisoquinoline pathway is a pivotal point for the regulation of BIA biosynthesis. Although enzymes involved in pathways other than protopine and protoberberine pathways were found (e.g., COR in the morphinan pathway), they showed a variable expression pattern between tuber and leaf samples. Two mapped to nonfunctional NADPH-dependent COR genes from lotus and poppy. Another mapped to a zinc finger protein that widely exists in the Papaveraceae family, and the reason for its annotation as a BIA pathway in KEGG is unclear. This could be due to faulty annotation or functional interpretation of the blast result, or this gene could be a genetic remnant of evolution when diverging from other morphine-producing plants. All the OMT candidate sequence sources were used in the phylogenetic analysis, but one unigene could not be identified (in gray, labeled unigene 1529), as it did not belong to any clade. According to the annotation information, unigene 1529 resembles the gene coding trans-resveratrol di-O-methyltransferase (ROMT) from *Vitis vinifera*, an enzyme that converts resveratrol to pterostilbene in stilbenoid biosynthesis[51]. We queried VvROMT via tblastn and found that it has 45% identity to unigene 1529 and less than 40% to other unigenes; therefore, we inferred that unigene 1529 might be an ortholog of the VvROMT gene since early basal groups of eudicots diverged in Rosidae[52]. Back to the discussion of papaverine biosynthesis, no 3′-O-methyltransferase candidates were found in any OMT gene, according to our phylogenetic analysis, which could be why no papaverine has ever been detected in *C. yanhusuo*.

A previous study identified a series of genes that mapped to the morphinan pathway[26]. However, our method identified very few enzymes in this pathway except for COR. The query sequences we used were all from *P. somniferum*, in the same family (Table S3), from which high identity hits (60%) should have been found if they existed. Such results indicate the lack of morphinan pathway-related enzymes, which corresponds with the chemical profile in *C. yanhusuo*[11,14,36].

An interesting phenomenon is that although many aporphine compounds have been reported in *C. yanhusuo*[14], we could not find the key enzyme (corytuberine synthase/CYP80G2)[53] at the beginning of the aporphine pathway (green) using the threshold we set. A similar phenomenon can also be observed on CoOMT[41], the product of which, (S)-tetrahydropalmatine (THP), abundantly exists in the tubers. Considering that the query plant we used (*Coptis japonica*) is not in the Papaveraceae family, there might be great variance in protein sequences. Therefore, when we adjusted the tblastn threshold to 40%, we found two unigene hits for CjCYP80G2 (51 and 48%) and two unigene hits for CjCoOMT (both 40%) (not shown in Fig. 4). Additionally, it was reported that SOMT1 in *P. somniferum* could catalyze the conversion from tetrahydrocolumbamine to THP[40], for which we obtained two hits after filtering unigenes successfully blasted to other OMTs with higher scores. We found five unigenes in the same clade with CjCoOMT and PsSOMT1, excluding a truncated unigene (c2788/f1p0/674). Further study could focus on cloning these unigenes to validate their activity at the protein level.

In summary, we reported the first full-length transcriptomes of the herbal medicinal plant *Corydalis yanhusuo* W.T. Wang. We also used the retrieved data to localize the metabolism of benzylisoquinoline alkaloids to the tuber and carried out organ-differentiated transcriptome analysis using a combined SMRT and SGS approach. This enabled us to identify the unigenes related to the BIA biosynthetic pathway. The expression patterns of the genes were validated through qRT-PCR, and UPLC-MS/MS was used to detect the metabolite products of the reactions catalyzed by the putative enzymes we found. Moreover, phylogenetic analysis identified ten candidate genes for THP synthesis; further studies of these genes are needed to verify their functions. Finally, our study provides an approach with which to analyze full-length transcriptome data and the biosynthesis process in other medicinal plants.

Material and methods

Plant materials and sample preparation

One-year-old *Corydalis yanhusuo* material was collected from the agriculture base in Pan'an City, Zhejiang Province, China (120.450 °E, 29.054 °N), which is the genuine production area of *C. yanhusuo*. The *C. yanhusuo* plants were then transplanted to pots in the laboratory

Downloaded from https://academic.oup.com/hr/article/doi/10.1038/s41438-020-00450-6/6446576 by guest on 04 August 2026

containing a mixture of perlite, vermiculite, and peat moss at a ratio of 1:1:1 and cultured for 7 days at a temperature of 20 °C with a day length of 12 h. The relative humidity was between 40 and 70%. The plants were harvested in mid-April, and fresh tissues were sampled. After removing the soil and dead wood, the tissues were immediately frozen in liquid nitrogen to prevent RNA degradation. All experimental groups contained three biological replicates from tubers and leaves for sequencing.

RNA extraction, quality assessment, and quantification

Total RNA was extracted using the RNAsimple Total RNA Kit (Tiangen Biotech, Beijing, China) according to the manufacturer's protocol. The quality of the total RNA was determined by 1% agarose gel electrophoresis, and the $OD_{260/280}$ values were checked with a NanoDrop ND-1000 spectrophotometer (NanoDrop Technologies, Wilmington, DE, USA). The total RNA samples with good quality were preserved for later Iso-Seq library construction.

Iso-Seq library construction and single-molecular real-time (SMRT) sequencing

After enriching mRNA from total RNA by oligo-dT magnetic beads, the cDNA library was reverse transcribed using the Clontech SMARTer PCR cDNA Synthesis Kit (Takara Bio, Mountain View, CA, USA) and Phusion High-fidelity DNA polymerase (NEB, Ipswich, MA, USA) using PCR amplification to yield a large quantity of cDNA. Size selection (>4 kb) to obtain full-length transcripts was conducted following the BluePippin Size Selection System protocol. Then, the repair (including end repair and internal repair) and exonuclease treatment of the full-length cDNA was performed. The Pacific Biosciences DNA Template Prep Kit 2.0 was used to construct the SMRT bell iso-seq libraries. The overall process was performed according to the isoform sequencing protocol (Iso-Seq) and by Pacific Biosciences (PN 100-092-800-03). After construction of the Iso-seq library, SMRT sequencing was performed on a Pacific Bioscience PacBio Sequel platform (Novogene, Beijing, China).

Illumina cDNA library construction and second-generation sequencing (SGS)

The cDNA library for SGS was constructed using the NEBNext® Ultra™ RNA Library Prep Kit for Illumina® (NEB, Beverly, MA, USA) according to the protocol provided by the manufacturer. The cDNA library sample was sent for sequencing on the high-throughput Illumina HiSeq platform (Novogene, Beijing, China).

The quality control of the SGS raw results was implemented through in-house Perl scripts: the adapter and primer sequences and those reads with poly-N were completely removed from the sequenced raw reads, and low-quality data (>50% bases with Q-value ≤ 20) were also filtered by calculating the Q20 and Q30 scores and sequence duplication level. The clean data yielded by Illumina were further used for correction of the PacBio sequencing data.

Gene functional annotation

Gene function was annotated based on the following databases: NR (NCBI nonredundant protein sequences)[54]; NT (NCBI nonredundant nucleotide sequences); Pfam (Protein family)[55]; KOG/COG (Clusters of Orthologous Groups of proteins)[56]; Swiss-Prot (a manually annotated and reviewed protein sequence database)[57]; KO (KEGG Ortholog database)[58]; and GO (Gene Ontology)[59]. BLAST software was used for NT database analysis with the E-value set at $1e^{-10}$. The software Diamond BLASTX V0.8.36 (https://github.com/bbuchfink/diamond) was used for NR, KOG, Swiss-Prot, and KEGG database analysis with the E-value set at $1e^{-10}$. The software Hmmscan V3.1b2 (http://hmmer.org/) was used for Pfam database analysis.

Gene expression levels and differential expression analysis

The input data for differential expression analysis is the read count of each gene retrieved by bowtie2 in RSEM software (V1.3.0)[60]. The differential expression analysis steps were as follows: The read count was first normalized using the DESeq *R* package (1.10.1), and then a hypothesis test on the negative binominal distribution model was conducted to calculate the p-values. Finally, the p-values were adjusted using Benjamini and Hochberg's multiple testing approach to control the false discovery rate[61]. Genes with adjusted p-values < 0.05 were assigned as differentially expressed genes (DEGs).

GO and KEGG enrichment analysis

GO (Gene Ontology) term enrichment analysis of DEGs was performed via the GOseq *R* package[62], which is based on the Wallenius noncentral hypergeometric distribution in which the gene selection bias was corrected by gene length. GO terms with corrected p-values < 0.05 were considered significantly enriched among DEGs. KEGG (Kyoto Encyclopedia of Genes and Genomes) pathway enrichment analysis was then performed using KOBAS (2.0) software, which carries out hypergeometric testing to identify the DEG-enriched KEGG pathways with a false discovery rate < 0.05.

Identification of candidates in the BIAs pathway

Local tblastn 2.2.10 was used to identify BIA pathway-related genes. Query sequences were retrieved from the GenBank and SwissProt databases. Tblastn was performed against *C. yanhusuo* FLNC fasta sequences. The results were filtered by identity >60% to allow for sequencing error. The heatmap was generated using the pheatmap *R* package with row scaling.

Downloaded from https://academic.oup.com/hr/article/doi/10.1038/s41438-020-00450-6/6446576 by guest on 04 August 2026

UPLC-Q-TOF-MS analysis for the nontargeted metabolomics study

For the metabolomics experiments, nine samples from each group were used for separate analyses. Analyses were performed using a Waters ACQUITY UPLC system coupled with SYNAPT G2-Si HDMS Q-TOF-MS. Samples were analyzed using an ACQUITY UPLC BEH column (2.1 × 100 mm, 1.7 μm, Waters Corporation, Milford, MA, USA) in positive mode. The column temperature was maintained at 40 °C, and the flow rate of the mobile phase was 0.30 mL/min, with an injection volume of 1.0 μL. Mobile phase A was 0.1% (v/v) formic acid/water, while mobile phase B was 0.1% (v/v) formic acid/acetonitrile. The automatic sampler was set at 4 °C during the analysis of all samples.

To acquire mass spectrometry data, a SYNAPT G2-Si HDMS with an electrospray ionization source was used. All data were collected in MS^E mode, and the parameters were as follows: MS^E range 50–1000 *m/z*; MS^E high energy 20 to 50 eV. Leucine enkephalin was continuously acquired and used to correct data (*m/z* 556.2771 in positive mode).

All data were acquired in MassLynx V4.2 and imported into Progenesis QI V2.0 (Waters Corporation, Milford, MA, USA) for background noise elimination, normalization to a reference sample, retention time correction, peak selection, and identification of compounds with reference to databases such as METLIN, HMDB, and Lipid Maps. Structural confirmation was conducted by comparison with the reference standards (t_R and MS data) or matching with theoretical data or commercial libraries. After Pareto scaling, principal component analysis (PCA) and orthogonal partial least-squares-discriminant analysis (OPLS-DA) were performed. The significantly different metabolites were identified based on the combination of statistically significant values for variable importance in projection (VIP) obtained from the OPLS-DA model and two-tailed Student's *t*-test (*p*-value) applied to the raw data, and metabolites with VIP values larger than 1.0 and p-values < 0.05 were considered significantly different. Global metabolomics was performed as described in the Supporting Information.

Analysis of alkaloid by targeted metabolomics

Considering that isoquinoline alkaloids are the main component of *C. yanhusuo*, 19 alkaloids (including scoulerine, canadine, tetrahydrocolumbamine, columbamine, tetrahydropalmatine, sanguinarine, corydaline, protopine, tetrahydrocoptisine, noroxyhydrastinine, dehydrocorydaline, oxoglaucine, 8-oxycoptisine, palmatine, coptisine, berberine, jatrorrhizine, epiberberine, and glaucine) in tubers and leaves were detected by targeted metabolomics profiling by electrospray ionization (ESI) tandem mass spectrometry (MS/MS). The profiling was performed with a Waters ACQUITY UPLC H-Class system consisting of a quaternary pump with degasser and autosampler in combination with a Waters Xevo TQ-XS mass spectrometer. Targeted metabolomics was performed as described in the Supporting Information.

qRT-PCR analysis

RNase-free DNase I (Takara Biotechnology, Dalian, China) was used to pretreat RNA samples before their use in reverse transcription to minimize DNA contamination. Following the instructions accompanying the HiScript Q RT SuperMix for qPCR kit (Vazyme, China), 1 μg of template RNA for cDNA analysis was added to a 20 mL admixture and diluted ten times for qRT-PCR analysis. The primer pairs used are shown in Table S5. The relative expression levels of the target genes were calculated using the $2^{-\Delta\Delta CT}$ approach with *GAPDH* as the reference gene[63]. Each gene was analyzed in three biological replicates and three technical replicates.

Phylogenetic analysis

The OMT candidates were selected from all genes with tblast hits to 4′-OMT, 6-OMT, CoOMT, SOMT1, N7OMT, and R7OMT, regardless of their scores. The protein sequences of these candidates were retrieved from the CDS prediction fasta files. Sequences from other species were acquired from the UniProt Knowledgebase. The alignment before phylogenetic analysis was achieved by MUSCLE with default parameters. MEGA-X was used to construct a neighbor-joining tree using the bootstrap method (1000 replications) with the Poisson model and pairwise deletion[64]. Tree visualization was achieved using the EvolView webpage[65].

Acknowledgements

This research was funded by the National Key R&D Program of China (2019YFC1711000) and the Basic Research Program for Natural Science of Shaanxi Province (2020JQ-862). It was also funded by the National Natural Science Foundation of China (grant number 81703637) and the Natural Science Fund in Jiangsu Province (grant number BK20170736). The authors thank AiMi Academic Services (www.aimieditor.com) for English language editing and review services.

Author details

[1]Key Laboratory of Shaanxi Administration of Traditional Chinese Medicine for TCM Compatibility, State Key Laboratory of Research & Development of Characteristic Qin Medicine Resources (Cultivation), Shaanxi Key Laboratory of Chinese Medicine Fundamentals and New Drugs Research, and Shaanxi Collaborative Innovation Center of Chinese Medicinal Resources Industrialization, Shaanxi University of Chinese Medicine, 712046 Xi'an, Shaanxi, China. [2]Department of Resources Science of Traditional Chinese Medicines and State Key Laboratory of Natural Medicines, School of Traditional Chinese Pharmacy, China Pharmaceutical University, 210009 Nanjing, Jiangsu, China. [3]Institute of Modern Chinese Medicine, School of Pharmacy, Wannan Medical College, 241002 Wuhu, Anhui, China

Author contributions

The authors' responsibilities were as follows: data curation, D.X. and H.L.; formal analysis, H.L.; funding acquisition, D.X., S.N., and Y.Z.; investigation, D.X., H.L., J.X., and S.N.; methodology, Y.T., D.X., H.L., L.H., and S.N.; software, Y.T., D.X.;

Downloaded from https://academic.oup.com/hr/article/doi/10.1038/s41438-020-00450-6/6446576 by guest on 04 August 2026

supervision, validation, Y.Z.; writing—original draft, D.X. and H.L. All authors read and approved the final manuscript.

Data availability

The PacBio SMRT reads and the Illumina SGS reads generated in this study have been submitted to the BioProject database of the National Center for Biotechnology Information (accession number PRJNA539894). Annotation results and sequences are available at https://doi.org/10.6084/m9.figshare.12800222.v1.

Conflict of interest

The authors declare that they have no conflict of interest.

Supplementary Information accompanies this paper at (https://doi.org/10.1038/s41438-020-00450-6).

Received: 10 May 2020 Revised: 3 November 2020 Accepted: 9 November 2020
Published online: 10 January 2021

References

1. Wang, L. et al. The antinociceptive properties of the Corydalis yanhusuo extract. *PLoS ONE* **11**, e0162875 (2016).
2. Xu, Z. et al. Dehydrocorydaline inhibits breast cancer cells proliferation by inducing apoptosis in MCF-7 cells. *Am. J. Chin. Med.* **40**, 177–185 (2012).
3. Wang, J. B. & Mantsch, J. R. l-tetrahydropalamatine: a potential new medication for the treatment of cocaine addiction. *Future Med. Chem.* **4**, 177–186 (2012).
4. Xu, W. et al. L-isocorypalmine reduces behavioral sensitization and rewarding effects of cocaine in mice by acting on dopamine receptors. *Drug Alcohol Depend.* **133**, 693–703 (2013).
5. Wen, C., Wu, L., Ling, H. & Li, L. Salutary effects of Corydalis yanhusuo extract on cardiac hypertrophy due to pressure overload in rats. *J. Pharm. Pharm.* **59**, 1159–1165 (2007).
6. Wu, L., Ling, H., Li, L., Jiang, L. & He, M. Beneficial effects of the extract from Corydalis yanhusuo in rats with heart failure following myocardial infarction. *J. Pharm. Pharm.* **59**, 695–701 (2007).
7. Han, Y. et al. l-Tetrahydropalmatine, an active component of Corydalis yanhusuo W.T. Wang, protects against myocardial ischaemia-reperfusion injury in rats. *PLoS ONE* **7**, e38627 (2012).
8. Luo, X. et al. Complete biosynthesis of cannabinoids and their unnatural analogues in yeast. *Nature* **567**, 123–126 (2019).
9. Lin, G. M., Warden-Rothman, R. & Voigt, C. A. Retrosynthetic design of metabolic pathways to chemicals not found in nature. *Curr. Opin. Syst. Biol.* **14**, 82–107 (2019).
10. Chou, T. The alkaloids of the Chinese Corydalis ambigua Cham, et Sch.(yen-hu-so). VI. identification of corydalis D and corydalis M. *Chin. J. Physiol.* **10**, 507–511 (1936).
11. Yang, X., Yang, X. & Liu, J. Study on material base of Corydalis Rhizoma. *Zhongguo Zhongyao Zazhi* **39**, 20–27 (2014).
12. Sun, M., Liu, J., Lin, C., Miao, L. & Lin, L. Alkaloid profiling of the traditional Chinese medicine Rhizoma corydalis using high performance liquid chromatography-tandem quadrupole time-of-flight mass spectrometry. *Acta Pharm. Sin. B* **4**, 208–216 (2014).
13. Li, Q. et al. Fingerprint-efficacy study of the quaternary alkaloids in Corydalis yanhusuo. *J. Ethnopharmacol.* **207**, 108–117 (2017).
14. Wu, X. S., Xu, J., Zhang, X. M., Zhang, T. J. & Chen, C. Q. Research progress on chemical constituents and pharmacological activities of Yuanhu Zhitong Prescription. *Zhong Cao Yao* **46**, 1081–1095 (2015).
15. Cabedo, N., Berenguer, I., Figadère, B. & Cortes, D. An overview on benzylisoquinoline derivatives with dopaminergic and serotonergic activities. *Curr. Med. Chem.* **16**, 2441–2467 (2009).
16. Kumar, A. et al. Current knowledge and pharmacological profile of berberine: an update. *Eur. J. Pharm.* **761**, 288–297 (2015).
17. Beaudoin, G. A. & Facchini, P. J. Benzylisoquinoline alkaloid biosynthesis in opium poppy. *Planta* **240**, 19–32 (2014).
18. Hagel, J. M. & Facchini, P. J. Benzylisoquinoline alkaloid metabolism: a century of discovery and a brave new world. *Plant Cell Physiol.* **54**, 647–672 (2013).
19. Yang, M. et al. Digital gene expression analysis provides insight into the transcript profile of the genes involved in aporphine alkaloid biosynthesis in Lotus (Nelumbo nucifera). *Front. Plant Sci.* **8**, 80 (2017).
20. Nishiyama, Y. et al. Antinociceptive effects of the extracts of Xylopia parviflora bark and its alkaloidal components in experimental animals. *J. Nat. Med.* **64**, 9–15 (2010).
21. Zhao, X. et al. The influence of dehydrocorydaline on intracellular free calcium concentration during hypoxia in myocardial cell of guinea-pigs. *China Appl. Physiol.* **19**, 222–225 (2003).
22. Hagel, J. M. et al. Transcriptome analysis of 20 taxonomically related benzylisoquinoline alkaloid-producing plants. *BMC Plant Biol.* **15**, 227 (2015).
23. Zhao, Y. et al. Transcriptomic profiles of 33 opium poppy samples in different tissues, growth phases, and cultivars. *Sci. Data* **6**, 66 (2019).
24. He, S. M. et al. Identification and characterization of genes involved in benzylisoquinoline alkaloid biosynthesis in species. *Front. Plant Sci.* **9**, 731 (2018).
25. Pourmazaheri, H. et al. Comparative analysis of the root and leaf transcriptomes in Chelidonium majus L. *PLoS ONE* **14**, e0215165 (2019).
26. Liao, D. et al. Identification and developmental expression profiling of putative alkaloid biosynthetic genes in Corydalis yanhusuo bulbs. *Sci. Rep.* **6**, 19460 (2016).
27. Morganti, S. et al. Complexity of genome sequencing and reporting: next generation sequencing (NGS) technologies and implementation of precision medicine in real life. *Crit. Rev. Oncol. Hemat.* **133**, 171–182 (2019).
28. Au, K. F., Underwood, J. G., Lee, L. & Wong, W. H. Improving PacBio long read accuracy by short read alignment. *PLoS ONE* **7**, e46679 (2012).
29. Au, K. F. et al. Characterization of the human ESC transcriptome by hybrid sequencing. *Proc. Natl Acad. Sci. USA* **110**, E4821–E4830 (2013).
30. Xu, Z. et al. Full-length transcriptome sequences and splice variants obtained by a combination of sequencing platforms applied to different root tissues of Salvia miltiorrhiza and tanshinone biosynthesis. *Plant J.* **82**, 951–961 (2015).
31. Chao, Q. et al. The developmental dynamics of the Populus stem transcriptome. *Plant Biotechnol. J.* **17**, 206–219 (2019).
32. Du, W. et al. Development and validation of a HPLC-ESI-MS/MS method for simultaneous quantification of fourteen alkaloids in mouse plasma after oral administration of the extract of Corydalis yanhusuo tuber: application to pharmacokinetic study. *Molecules* **23**, 714 (2018).
33. Zhang, J. et al. Systematic screening and characterization of tertiary and quaternary alkaloids from Corydalis yanhusuo W.T. Wang using ultra-performance liquid chromatography-quadrupole-time-of-flight mass spectrometry. *Talanta* **78**, 513–522 (2009).
34. Desgagné-PenixIsabel, I. et al. Integration of deep transcriptome and proteome analyses reveals the components of alkaloid metabolism in opium poppy cell cultures. *BMC Plant Biol.* **10**, 252 (2010).
35. Fu, X. Y., Liang, W. Z. & Tu, G. S. Chemical studies on the alkaloids isolated from the tuber of Yuanhu (Corydalis turtschaninovii Bees. f. yanhusuo Y. H. Chou et C. C. Hsu). *Acta Pharm. Sin.* **21**, 447–453 (1986).
36. He, K., Gao, J. L. & Zhao, G. S. Advances in studies on chemistry, pharmacology, and quality control of Corydalis yanhusuo. *Zhong Cao Yao* **38**, 1909 (2007).
37. Chu, H., Jin, G., Friedman, E. & Zhen, X. Recent development in studies of tetrahydroprotoberberines: mechanism in antinociception and drug addiction. *Cell Mol. Neurobiol.* **28**, 491–499 (2008).
38. Letunic, I. & Bork, P. 20 years of the SMART protein domain annotation resource. *Nucleic Acids Res.* **46**, D493–D496 (2018).
39. Botros, H. G. et al. Crystal structure and functional mapping of human ASMT, the last enzyme of the melatonin synthesis pathway. *J. Pineal Res.* **54**, 46–57 (2013).
40. Dang, T. T. & Facchini, P. J. Characterization of three O-methyltransferases involved in noscapine biosynthesis in opium poppy. *Plant Physiol.* **159**, 618–631 (2012).
41. Morishige, T., Dubouzet, E., Choi, K. B., Yazaki, K. & Sato, F. Molecular cloning of columbamine O-methyltransferase from cultured Coptis japonica cells. *Eur. J. Biochem.* **269**, 5659–5667 (2002).
42. Cho, H. Y., Rhee, H. S., Yoon, S. Y. & Park, J. M. Differential induction of protein expression and benzophenanthridine alkaloid accumulation in Eschscholtzia californica suspension cultures by methyl jasmonate and yeast extract. *J. Microbiol. Biotechnol.* **18**, 255–262 (2008).
43. Glazebrook, J. Genes controlling expression of defense responses in Arabidopsis–2001 status. *Curr. Opin. Plant Biol.* **4**, 301–308 (2001).
44. Colque, R., Viladomat, F., Bastida, J. & Codina, C. Improved production of galanthamine and related alkaloids by methyl jasmonate in Narcissus confusus shoot-clumps. *Planta Med.* **70**, 1180–1188 (2004).

Downloaded from https://academic.oup.com/hr/article/doi/10.1038/s41438-020-00450-6/6446576 by guest on 04 August 2026

45. Wei, X., Vrieling, K., Mulder, P. P. J. & Klinkhamer, P. G. L. Methyl jasmonate changes the composition and distribution rather than the concentration of defence compounds: a study on pyrrolizidine alkaloids. *J. Chem. Ecol.* **45**, 136–145 (2019).
46. Ketchum, R. E. et al. Taxus metabolomics: methyl jasmonate preferentially induces production of taxoids oxygenated at C-13 in Taxus x media cell cultures. *Phytochemistry* **62**, 901–909 (2003).
47. Cho, H. Y. et al. Synergistic effects of sequential treatment with methyl jasmonate, salicylic acid and yeast extract on benzophenanthridine alkaloid accumulation and protein expression in Eschscholtzia californica suspension cultures. *J. Biotechnol.* **135**, 117–122 (2008).
48. Shitan, N. et al. Involvement of CjMDR1, a plant multidrug-resistance-type ATP-binding cassette protein, in alkaloid transport in Coptis japonica. *Proc. Natl Acad. Sci. USA* **100**, 751–756 (2003).
49. Shitan, N. et al. Characterization of Coptis japonica CjABCB2, an ATP-binding cassette protein involved in alkaloid transport. *Phytochemistry* **91**, 109–116 (2013).
50. Feng, Z. Research on the accumulation dynamic of dry matter and two alkaloids in the tubers of Corydalis yanhusuo. *J. Anhui Agric. Sci.* **38**, 1245–1246 (2010).
51. Schmidlin, L. et al. A stress-inducible resveratrol O-methyltransferase involved in the biosynthesis of pterostilbene in grapevine. *Plant Physiol.* **148**, 1630–1639 (2008).
52. Chase, M. W. et al. An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG IV. *Bot. J. Linn. Soc.* **181**, 1–20 (2016).
53. Ikezawa, N., Iwasa, K. & Sato, F. Molecular cloning and characterization of CYP80G2, a cytochrome P450 that catalyzes an intramolecular C-C phenol coupling of (S)-reticuline in magnoflorine biosynthesis, from cultured Coptis japonica cells. *J. Biol. Chem.* **283**, 8810–8821 (2008).
54. Li, W. Z., Jaroszewski, L. & Godzik, A. Tolerating some redundancy significantly speeds up clustering of large protein databases. *Bioinformatics* **18**, 77–82 (2002).
55. Finn, R. D. et al. The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Res.* **44**, D279–D285 (2016).
56. Tatusov, R. L. et al. The COG database: an updated version includes eukaryotes. *BMC Bioinforma.* **4**, 41 (2003).
57. Bairoch, A. & Apweiler, R. The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000. *Nucleic Acids Res.* **28**, 45–48 (2000).
58. Kanehisa, M., Goto, S., Kawashima, S., Okuno, Y. & Hattori, M. The KEGG resource for deciphering the genome. *Nucleic Acids Res.* **32**, D277–D280 (2004).
59. Ashburner, M. et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat. Genet.* **25**, 25–29 (2000).
60. Li, B. & Dewey, C. N. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinforma.* **12**, 323 (2011).
61. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. B* **57**, 289–300 (1995).
62. Young, M. D., Wakefield, M. J., Smyth, G. K. & Oshlack, A. Gene ontology analysis for RNA-seq: accounting for selection bias. *Genome Biol.* **11**, R14 (2010).
63. Bao, Z. et al. Identification and selection of reference genes for quantitative transcript analysis in Corydalis yanhusuo. *Genes* **11**, 130 (2020).
64. Kumar, S., Stecher, G., Li, M., Knyaz, C. & Tamura, K. MEGA X: molecular evolutionary genetics analysis across computing platforms. *Mol. Biol. Evol.* **35**, 1547–1549 (2018).
65. He, Z. et al. Evolview v2: an online visualization and management tool for customized and annotated phylogenetic trees. *Nucleic Acids Res.* **44**, W236–W241 (2016).

Downloaded from https://academic.oup.com/hr/article/doi/10.1038/s41438-020-00450-6/6446576 by guest on 04 August 2026