**Supporting Information**

# CYP82AR subfamily proteins catalyze C-1' hydroxylations of deoxyshikonin in the biosynthesis of shikonin and alkannin

Wan Song,[†,§,‡] Yibin Zhuang[†,‡] and Tao Liu[*,†]

[†]Key Laboratory of Systems Microbial Biotechnology, Tianjin Institute of Industrial Biotechnology, Chinese Academy of Sciences, Tianjin 300308, China

[§]University of Chinese Academy of Sciences, 19 Yuquan Road, Beijing 100049, China

[‡]These two authors contributed equally

*Corresponding author: Tao Liu, liu_t@tib.cas.cn

**Table of Contents**

## S1 Materials and Methods

### S1.1 Plants materials and transcriptome sequencing

The mature seeds of *Lithospermum erythrorhizon* Sieb. et Zucc. (Boraginaceae) were collected in the field of Chifeng city, Inner Mongolia Autonomous Region, China in October 2016[1]. The seeds were germinated as described previously[2] and the seedlings were grown in a greenhouse at 23 ± 1 °C with a 16 h-light/8 h-dark photoperiod. After 60 days, different tissues (green leaves/stems and red roots) of *L. erythrorhizon* were collected and flash frozen by liquid nitrogen for subsequent cDNA library construction for transcriptome sequencing in Genewiz, China[1]. The publicly available raw sequencing datasets of 6 tissue-specific transcriptomes of *L. erythrorhizon*[3] and *E. plantagineum*[3] (roots, leaves+stems, flowers, root periderm, root cortex and root stele) plus 2 tissue-specific transcriptomes of *A. euchroma*[4] and *L. officinale* (roots, leaves+stems)[5] were downloaded from the NCBI database and *de novo* assembled by Biomarker Technologies Corporation (Beijing, China).

### S1.2 Candidate CYPs obtained by co-expression analysis

Previously, we used a hybrid sequencing approach combining next-generation sequencing (NGS) short reads and single-molecule real-time (SMRT) long reads to obtain more accurate and complete transcriptome data for *L. erythrorhizon* of roots and leaves+stems[1]. Forty full-length cytochrome P450 enzyme (CYP) candidate genes transcriptionally upregulated in the red roots have been cloned into the yeast expression vector pCf302-AtCPR1 with the required CYP reductase from *Arabidopsis thaliana* (AtCPR1) for further investigation[1]. The transcriptomic datasets of *L. erythrorhizon* of different tissues including roots, leaves+stems, flowers, root periderm, root cortex and root stele were publicly available[3]. These raw sequencing datasets were downloaded from the NCBI database and further assembled by Biomarker Technologies Corporation (Beijing, China). Shikonin/alkannin are predominantly localized in the periderm of intact roots[3, 6, 7], and thus the expression profiles of target genes were positively correlated with the alkannin/shikonin accumulation pattern[3]. To narrow down the CYP candidates for further screening, we

conducted gene expression level analysis by comparing different tissues such as roots vs. leaves+stems and flowers, periderm vs. cortex and stele. The FPKM (fragments per kilobase of transcript per million mapped reads) value was used to quantify the expression level of each unigene and the DEGSeq program[8] was used to analyze the differentially expressed unigenes of the different tissues. The expression heat map was generated based on expression levels of transcripts encoding selected known enzymes and CYP candidates in different tissues of *L. erythrorhizon*. The candidates were selected based on coexpression analysis and their FPKM values in different tissues. In this work, the CYP candidates with expression profiles coincident to the transcripts encoding known enzymes of shikonin biosynthesis (phenylalanine ammonialyase, 3-hydroxy-3-methylglutaryl-coenzyme A reductase, 4-hydroxybenzoate 3-geranyltransferase and geranylhydroquinone 3″-hydroxylase) were selected for cloning and further enzymatic activity testing. As a result, seven CYP candidates highly expressed in roots and displayed similar expression profiles to known shikonin genes were identified (Figure 2). Among them, CYP_c44379 encodes cinnamic acid 4-hydroxylase (C4H), and thus was excluded for activity testing.

**S1.3 Candidate DSHs homologs obtained using CYP82AR2 as a query**

To further identify the enzymes hydroxylating deoxyshikonin to produce shikonin/alkannin in other Boraginaceae plants, we acquired available raw sequencing datasets of tissue-specific transcriptomes of *E. plantagineum*[3] (roots, leaves+stems, flowers, root periderm, root cortex and root stele), *A. euchroma*[4] (roots, leaves+stems) and *Lithospermum officinale* (roots, leaves+stems)[5] from the NCBI database. These raw sequencing datasets were further assembled by Biomarker Technologies Corporation (Beijing, China). Moreover, the *E. plantagineum* genome is publicly available in the NCBI database[3]. We performed BLAST searches using CYP82AR2 as a query against the above three transcriptomes databases and the *E. plantagineum* genome databases. As a result, two homologs sharing high sequence identity with CYP82AR2 were discovered from each species.

**S1.4 Chemicals and reagents**

Chromatography-grade formic acid, isopropanol, methanol, acetonitrile and normal hexane were obtained from Sigma-Aldrich, USA. Deoxyshikonin, Lithium acetate, PEG 3350, ssDNA and ampicillin were obtained from Solarbio, China. Restriction enzymes, DNA polymerase and T4 ligase were purchased from Thermo Fisher Scientific, USA. The authentic standard of shikonin and alkannin were purchased from J&K Scientific Ltd.. All other chemicals were of commercial reagent grade.

**S1.5 Strains and growth conditions**

*S. cerevisiae* BY4742[9] (*MATα his3Δ1 leu2Δ0 lys2Δ0 ura3Δ0*), which is a derivative of S288C, was used as the parent strain for all engineered strains. Yeast strains were cultivated at 30 °C and 220 rpm in YPD medium containing 10 g $l^{-1}$ of yeast extract, 20 g $l^{-1}$ of beef peptone, and 20 g $l^{-1}$of glucose or in SD-Ura medium (uracil-minus) containing 6.7 g $l^{-1}$ of yeast nitrogen base without amino acids, 0.9 g $l^{-1}$ of SD-Ura, 20 g $l^{-1}$ of glucose[10]. *Escherichia coli* Trans-T1 (TransGen Biotech, China) was used for bacterial transformation and recombinant vectors construction. The *E. coli* strains with recombinant plasmids were grown at 37 °C and 200 rpm in Luria-Bertani medium with 100 mg $l^{-1}$ ampicillin.

**S1.6 Construction of plasmids and strains**

The yeast expression vector pCf302 with the constitutive promoters $P_{TEF1}$, $P_{PGK1}$, $P_{TDH3}$ was constructed in our lab[11]. For the analysis of CYP candidates, the vector pCf302-AtCPR1 harboring the *A. thaliana* cytochrome P450 reductase gene (*AtCPR1*) with codon optimization was constructed[1]. The 40 engineered strains harboring full-length CYP candidate genes transcriptionally up-regulated in the red roots have been constructed previously[1]. The seven candidate genes of DSH with the restriction sites *Spe*I and *Pac*I were synthesized by Generay Biotech Co., Ltd (Shanghai, China). Subsequently, they were cloned into pCf302-AtCPR1 under the promoter $P_{PGK1}$, resulting in the expression vectors pCf302-AtCPR1-CYP. The seven constructed

expression plasmids were respectively transformed into *S. cerevisiae* BY4742 by the standard lithium acetate method[12]. The vector pCf302-AtCPR1 was also transformed into *S. cerevisiae* BY4742 and served as the negative control. The recombinant yeast cells were selected on a uracil-minus plate (SD-Ura) at 30 °C for 3 d and verified by colony PCR.

**S1.7 Feeding experiments**

All the recombinant yeast strains carrying candidate CYPs and the empty vector were cultured in 5 ml SD-Ura medium at 30 °C and 220 rpm for 20 h, and three colonies were picked for each genotype to reduce errors. Subsequently, 20 μM of the substrate deoxyshikonin, was fed to 5 ml the engineered yeast and the cultures were further shaken at 30 °C and 220 rpm for 48 h. The yeast cells were harvested by centrifugation at 13,800 g for 10 min, and extracted twice with 1 ml methanol. The methanol extracts were mixed together, evaporated, and finally redissolved in methanol for HPLC analysis.

To isolate sufficient quantities of the hydroxylated products for stereochemistry characterization, these eight strains that respectively harbored DSH gene were cultivated in 200 ml SD-Ura medium. The positive colonies were inoculated into 200 ml fresh SD-Ura medium and cultured at 30 °C and 220 rpm for 24 h. Subsequently, 40 μM of the substrate deoxyshikonin was added to the cultures, which were further shaken at 30 °C and 220 rpm for 48 h. The yeast cells were harvested by centrifugation at 13,800 g for 10 min and extracted twice with methanol. The methanol extracts were mixed together, evaporated, redissolved in methanol and further purified by semi-preparative HPLC.

**S1.8 HPLC and LC-HRMS analysis**

In order to identify the hydroxylation activity of the canditate CYPs, the methanol extracts from the feeding experiments were analyzed by a Shimadzu HPLC system with a UV detector. The samples for HPLC analysis were cleaned off impurities by centrifugation (13,800 g) and by filtration using PTFE 0.2 μm syringe filters (Axiva,

Sigma Chemicals). The analytical column used was a YMC-Pack ODS-A (4.6 × 250 mm, 5 μm). The mobile phase contained 0.1% formic acid in water (A) and 100% HPLC grade acetonitrile (B). The gradient conditions were as follows: 0-3 min, 60% B; 3-22 min, a linear gradient of 60-100% B; 22-32 min, 100% B; 32-33 min, 100-60% B; 33-43 min, 60% B. The mobile phase flow was 1 ml $min^{-1}$and an injection volume of 25 μl. The products were detected and quantified by UV absorption at 516 nm. The conversion ratios of DSHs were determined on the basis of peak areas of hydroxyl products shikonin/alkannin and substrate deoxyshikonin.

To identify the oxidation products of DSHs, the LC-HRMS analysis was carried out on an Agilent 1200 HPLC system coupled with an Agilent Infinity UV detector and a Bruker-MicrOTOF-II mass spectrometer that was equipped with an electrospray ionization device. Data acquisition and processing were done with MicrOTOF control version 3.0/Data Analysis Version 4.0 software. For HPLC analysis, the YMC-Pack ODS-A (4.6 × 250 mm, 5 μm) was used. The analysis conditions and the injection volumes were the same as the conditions of HPLC analysis by the Shimadzu system above except for the mobile phase (A) changed into 5 mM ammonium acetate in water. The column temperature was set at 40 °C. Optimized MS operating conditions were as follows: all spectra were obtained in negative ion mode over an *m*/*z* range of 50-1000 under a dry gas flow of 6.0 l $min^{-1}$, a dry temperature of 180 °C, a nebulizer pressure of 1 bar and a probe voltage of +4.0 kV.

**S1.9 Microsome isolation and *in vitro* enzyme assays**

The strain *S. cerevisiae* BY4742 harboring pCf302-AtCPR1-CYP82AR2 was cultivated for microsome isolation. Microsomes extracted from the yeast carrying the vector pCf302-AtCPR1 were assayed as the negative control. The strains were inoculated into culture tubes containing 5 ml of SD-Ura medium and grown at 30 °C and 220 rpm for 24 h. The seed cultures were inoculated into 200 ml SD-Ura medium and further cultivated at 30 °C and 220 rpm for 36 h. The microsomal isolation was carried out by differential centrifugation as described previously[13, 14] with the following modifications. The cells were centrifuged (6,000 g, 10 min) and washed twice with TEK buffer (50 mM Tris-HCl pH 7.5, 1 mM EDTA, 100 mM KCl) and

centrifuged. Collected cells were resuspended in TES buffer (50 mM Tris-HCl pH 7.5, 1 mM EDTA, 600 mM Sorbitol, 1 mM PMSF) and disrupted on ice by using a high-pressure homogenizer (JNBIO). The lysed cells were centrifuged (11,000 g, 20 min) to remove the cell debris, mitochondria and nuclei. The supernatant was transferred to ultracentrifugation tubes and centrifuged at 150,000 g for 1.5 h. The supernatant was discarded and the harvested microsomes were resuspended in 1 ml TEG buffer (50 mM Tris-HCl pH 7.5, 1 mM EDTA, and 20% glycerol), aliquoted (200 μl) and stored at −80 °C. All steps for microsomal preparation were carried out at 4 °C.

The enzymatic assays were conducted in 200 μl reaction mixtures containing 50 mM sodium phosphate buffer (pH 7.5), 1 mM NADPH, 50 μl microsomal proteins, and 40 μM substrate deoxyshikonin. After incubation for 5 h at 30 °C, the reactions were terminated by adding 200 μl methanol. Since shikonin and its derivatives are weakly polar compounds and easy to adsorb membrane structures, these red naphthoquinones blending into the denatured proteins were collected by centrifugation at 13,800 g for 10 min. The supernatant was removed and the red precipitates were extracted thrice with methanol. The methanol extracts were mixed together, evaporated, redissolved in methanol and submitted to HPLC analysis.

**S1.10 Purification of the hydroxylated products**

Impurities in methanol extracts affect the results of chiral HPLC analysis. To avoid this negative influence, we roughly purified the hydroxylated products of the eight DSHs. To isolate sufficient quantities of the hydroxylated products to identify their stereochemistry, semi-preparative HPLC separation was performed using a Shimadzu LC-6 AD with a SPD-20A detector and a YMC-Pack ODS-A (10 × 250 mm, 5 μm). The mobile phase contained 0.1% formic acid in water (A) and 100% HPLC grade methanol (B). The isocratic elution condition was adopted at 80% B and 300 μl of the enriched sample were injected. Products were detected and quantified by UV absorption at 516 nm. The solvent flow rate was 4.0 ml $min^{-1}$. The target fraction that was eluted between 11.5 and 12.5 min was repeatedly collected, dried, and

resuspended in methanol for the succeeding chiral HPLC analysis. The collections of shikonin and alkannin are red amorphous solid.

**S1.11 Chiral phase HPLC analysis**

The optical isomers of shikonin and alkannin can be separated by chiral phase HPLC[15]. In order to identified the configurations of the hydroxylated products, the analytical method of chiral phase HPLC analysis was applied. To avoid the influence of impurities in methanol extracts, the oxidized products were firstly isolated by semi-preparative reversed-phase HPLC separation. Subsequently, the puried product**s** were separated and identified using a chiral phase column CHIRALPAK IC (4.6 × 250 mm, 5 μm). The chiral phase HPLC analysis was carried out on a Shimadzu HPLC system with a photodiode array (PDA) detector. The mobile phase contained 90% n-hexane (A) and 10% isopropanol (B). The mobile phase flow was 0.8 ml $min^{-1}$ and the column temperature was set at 40 °C. 20 μl of the purified products were subjected to chiral phase HPLC analysis to identify the stereochemistry based on the configuration of the standard shikonin/alkannin. The ratio of the enantiomers was determined on the basis of peak areas of shikonin and alkannin. The absolute configuration of shikonin and alkannin was not changed to the opposite configuration by the alkaline treatment[15], as well as keeping them at room temperature for a few days in our study.

**S1.12 Phylogenetic analysis**

The phylogenetic tree was constructed using Mega 6.0 software package[16] and the neighbor-joining program based on the Poisson model and a bootstrap of 1,000 replicates. All the amino acid sequences used for phylogenetic tree construction were listed in Supplementary Table S-3.5 and aligned by ClustalX. The scale bar indicates 0.2 amino acid substitution per site. The numbers at the nodes of each branch indicate the percentage of bootstrap values.

## S2 Supporting figures

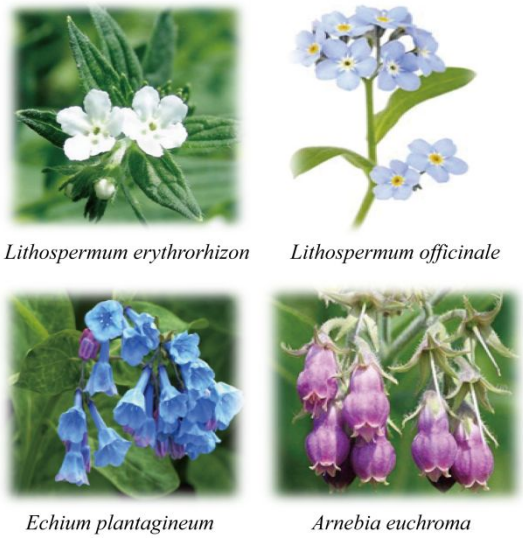


**Figure S2.1** Several Boraginaceae plants producing shikonin and alkannin

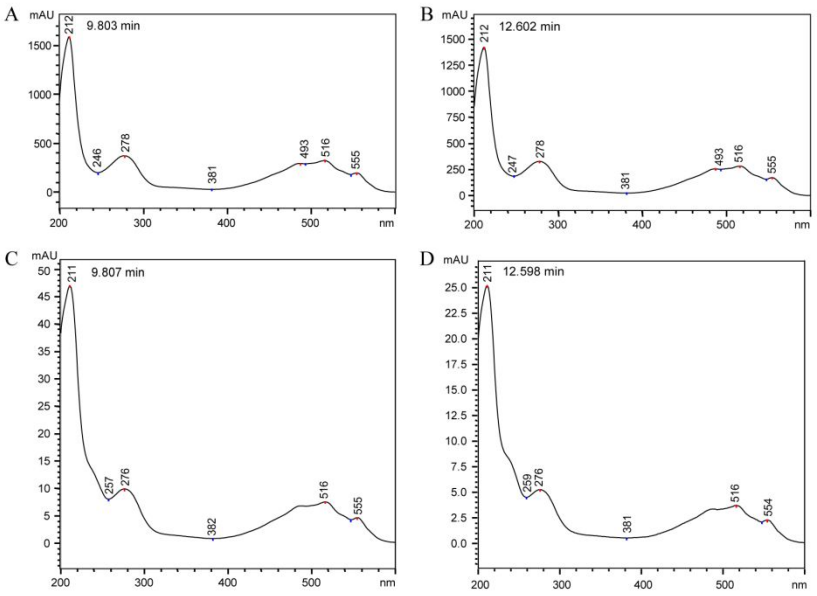


**Figure S2.2** UV spectra of shikonin/alkannin standards and hydroxylated products from chiral HPLC analysis. (A) The UV spectra of the standard *R*-shikonin eluting at 9.803 min. (B) The UV spectra of the standard *S*-alkannin eluting at 12.602 min. (C) The UV spectra of the hydroxylated products from DSHs eluting at 9.807 min. (D) The UV spectra of the hydroxylated products from DSHs eluting at 12.598 min.

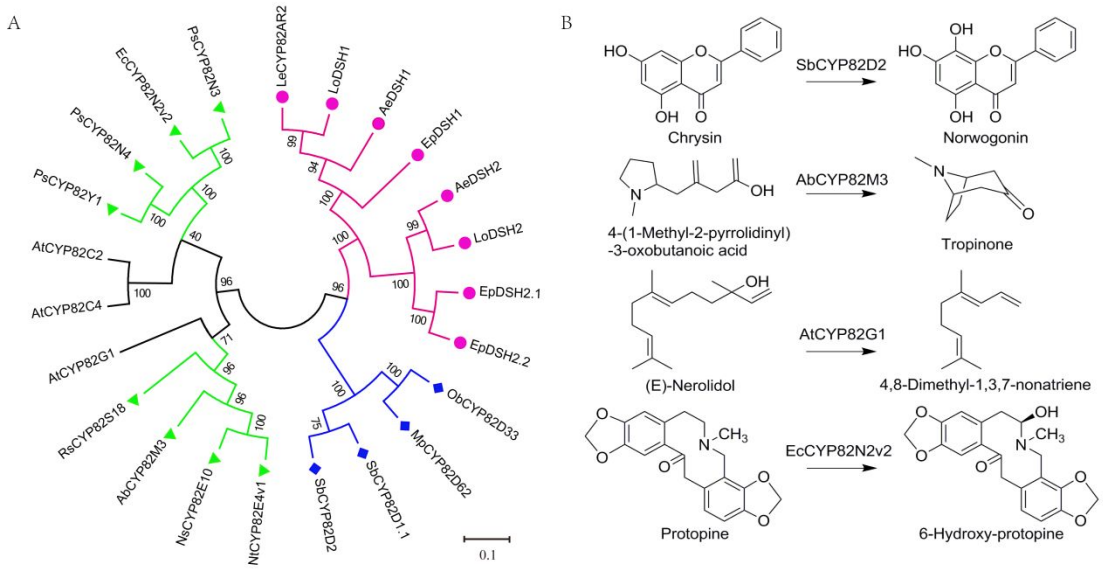


**Figure S2.3** Phylogenetic analysis of deoxyshikonin hydroxylases and other proteins from the CYP82 family. (A) The phylogenetic tree of deoxyshikonin hydroxylases (DSHs) and other proteins from the CYP82 family. (B) The typical reactions catalyzed by the related CYP enzymes with known natural activities. The phylogenetic tree was generated using Mega 6.0 software package and the neighbor-joining program based on the Poisson model and a bootstrap of 1,000 replicates. Protein sequences used for phylogenetic analysis with their corresponding accession numbers were listed in Table SI-3.5. The DSHs reported in this work were labeled with red filled circles. The functionally characterized plant CYPs involved in flavonoids and alkaloids biosynthesis were labeled with blue filled diamonds and green filled triangles, respectively.

## S3 Supporting Tables

**Table S3.1** Amino acid sequence identities of the DSHs characterized from four Boraginaceae plants

| Protein | Identities (%) | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | CYP82AR2 (LeDHS1) | LoDSH1 | LoDSH2 | AeDSH1 | AeDSH2 | EpDSH1 | EpDSH2.1 | EpDSH2.2 |
| CYP82AR2 (LeDHS1) | 100 | 99.3 | 72.4 | 95.7 | 74.3 | 89.1 | 72.4 | 72.4 |
| LoDSH1 | | 100 | 72.6 | 95.9 | 74.6 | 88.9 | 73.2 | 73.0 |
| LoDSH2 | | | 100 | 72.4 | 91.0 | 71.2 | 86.5 | 86.3 |
| AeDSH1 | | | | 100 | 73.9 | 89.8 | 73.0 | 72.8 |
| AeDSH2 | | | | | 100 | 72.3 | 86.5 | 86.3 |
| EpDSH1 | | | | | | 100% | 71.0 | 70.5 |
| EpDSH2.1 | | | | | | | 100 | 99.4 |
| EpDSH2.2 | | | | | | | | 100 |

**Table S3.2** The candidate genes expression levels in different plant organs of *L. erythrorhizon*

| gene | flowers | leaves+stems | roots | root stele | root cortex | root periderm |
|---|---|---|---|---|---|---|
| CYP76B101 | 5.14 | 45.01 | 614.79 | 185.12 | 182.04 | 346.04 |
| PAL | 110.11 | 139.37 | 299.55 | 24.61 | 23.22 | 60.01 |
| HMGR | 6.11 | 10.83 | 233.57 | 29.82 | 55.77 | 294.13 |
| LePGT1 | 0.34 | 25.31 | 660.87 | 91.27 | 54.9 | 347.26 |
| LePGT2 | 1.22 | 20 | 1682.99 | 11.52 | 19.96 | 450.97 |
| CYP_c60011.1 | 0.35 | 0.93 | 44.63 | 53.57 | 63.93 | 94.26 |
| CYP_c53434 | 6.82 | 8.95 | 19.31 | 1.26 | 9.47 | 55.97 |
| CYP_c60011.2 | 1.46 | 4.72 | 207.1 | 244.82 | 221.48 | 373.24 |
| CYP_c59628 | 0.08 | 0.17 | 35.43 | 4.78 | 5.87 | 3021.96 |
| CYP_c21867 | 0.11 | 3.18 | 138.8 | 12.09 | 5.53 | 81.63 |
| CYP_c44379 (C4H) | 147.92 | 511.57 | 1591.34 | 126.85 | 91.92 | 628.54 |
| CYP_c65173 (CYP82AR2) | 0.98 | 29.81 | 104.32 | 91.98 | 57.48 | 314.01 |

**Table S3.3** Expression levels in different plant organs of the eight homologs of DSHs

| Coding Protein | FPKM | | | | | |
|---|---|---|---|---|---|---|
| | flowers | leaves+stems | roots | root | | |
| | | | | stele | cortex | periderm |
| LeDSH1 | 0.98 | 29.81 | 104.32 | 91.98 | 57.48 | 314.01 |
| LeDSH2 | 0 | 0.38 | 1.96 | 0.85 | 1.76 | 36.76 |
| EpDSH1 | 0 | 0.04 | 23.98 | 0.1 | 0.1 | 57.54 |
| EpDSH2.1/EpDSH2.2 | 1.9 | 51.6 | 60.22 | 1.78 | 1.64 | 135.12 |
| LoDSH1 | | 3.98 | 45.26 | | | |
| LoDSH2 | | 3.27 | 64.64 | | | |
| AeDSH1 | | 2.69 | 22.89 | | | |
| AeDSH2 | | 33.11 | 209.18 | | | |

**Table S3.4** The selectivity and conversion ratios of DSHs

| Protein | Selectivity (%) | | Conversion ratios (%) |
|---|---|---|---|
| | Shikonin (*R*-form) | Alkannin (*S*-form) | |
| CYP82AR2 (LeDSH1) | 75.4 | 24.6 | 23.8 |
| LeDSH2 | - | - | - |
| LoDSH1 | 64.5 | 35.5 | 32.5 |
| LoDSH2 | No | No | No |
| AeDSH1 | 63.0 | 37.0 | 79.8 |
| AeDSH2 | 0 | 100 | 74.1 |
| EpDSH1 | 70.5 | 29.5 | 80.6 |
| EpDSH2.1 | 0 | 100 | 76.7 |
| EpDSH2.2 | 0 | 100 | 81.9 |

**Table S3.5** Protein sequences used for phylogenetic analysis with their corresponding accession numbers

| Protein | Species | Accession |
|---|---|---|
| AtCYP82C4 | *Arabidopsis thaliana* | NP_194922.1 |
| AtCYP82C2 | *Arabidopsis thaliana* | O49394.2 |
| ObCYP82D33 | *Ocimum basilicum* | AGF30364.1 |
| MpCYP82D62 | *Mentha x piperita* | AGF30366.1 |
| SbCYP82D1.1 | *Scutellaria baicalensis* | ASW21050.1 |
| SbCYP82D2 | *Scutellaria baicalensis* | ASW21052.1 |
| RsCYP82S18 | *Rauwolfia serpentina* | ASG81458.1 |
| NtCYP82E4v1 | *Nicotiana tabacum* | ABA07805.1 |
| NsCYP82E10 | *Nicotiana sylvestris* | ADP65810.1 |
| PsCYP82Y1 | *Papaver somniferum* | AFB74617.1 |
| PsCYP82N4 | *Papaver somniferum* | AGC92398.1 |
| EcCYP82N2v2 | *Eschscholzia californica* | BAK20464.1 |
| PsCYP82N3 | *Papaver somniferum* | AGC92397.1 |
| AbCYP82M3 | *Atropa belladonna* | AYU65303.1 |
| AtCYP82G1 | *Arabidopsis thaliana* | NP_189154.1 |
| LeCYP82AR2 | *Lithospermum erythrorhizon* | MT921814 |
| AeDSH1 | *Arnebia euchroma* | MW714367 |
| AeDSH2 | *Arnebia euchroma* | MW714368 |
| EpDSH1 | *Echium plantagineum* | MW714372 |
| EpDSH2.1 | *Echium plantagineum* | MW714373 |
| EpDSH2.2 | *Echium plantagineum* | MW714374 |
| LoDSH1 | *Lithospermum officinale* | MW714369 |
| LoDSH2 | *Lithospermum officinale* | MW714371 |

## S-4 The corresponding sequences of DSHs

>CYP82AR2: nucleotide. sequence from *Lithospermum erythrorhizon*

ATGGAGTTGTCCTTCAACTCAATCTTTGAAACACCAATAATATTAGGTGTATTGCTAGTT
ATAACAATACTTATTTGGCTCAAAATAACTCTTTCACCTCGTTTGAAGACCCCTCCCGAA
GTTGGTGTTGCGTTGCCCATAATTGGGCACCTATACCTAAAAGCCCTTAGAGGAAGCAAA
AAACCTCTCTTTATAAAGTTTACAAACTTGGCCGAAAAATTCGGACCGATTTATAACGTA
CGGCTCGGATCCATTCGAGCCGTAGTTATAAGCAATTCCGAATTAGCCAGGGAACTATTC
ACGGCAAAGGACAATTTCGTATTGGCAAGACCAAAATCTCTAGCAACCAGTCACTTAGCT
TATAGCTACGCCAATTTAGGAGTAGCTCCTCCTACTCCATATTGGCGTTGGCTAAGGAAA
TTCACCGCGGTGGGATTCTTCTCCCACCGCGCCCTTGACATGGCCAAGAATGTCCCAGCT
ACTGAAATCAAGTTATCGATTAAGTACCTTTATGATCTCTGTCATGATGAGGGTAGCGCC
AGAATTGCTGATATGCAACAATGGCTTCTAGATATTGGTTTGAACCTTATGATGAGAACT
GTTGTAGGAAAATCAACTTCTACTTCTGATGATAATGCTGACGAGGAGGAAGCTAAAGAA
CGGCGAAGATGGAAGAAGATGATGGATGATACAATGAGAATGCTTTTCTTGCCAGTGTTG
AGTGATTCGATCCCTCTTCTAAAGCCGTTGGATATAGGTGGGATTGAAAAAGAGATGAAA
CAAGTGAAGAAAACTATGGATGAGATTGTTGATCAATGGTTGAAAGAGCATATACAAAAG
AAAGCTAATGGTATTCATGTTGATGCTGAGAAGGATTTTATGGACTTGTTGCTAGCTGCA
GTAGAAGATGGTGATGTTGAACTAGGTGGTTATCATCCTCACGAGGTTGTTAAGGCAACA
TGCATGTCCATGGTTGGTGCTGGGAGTGATACTACATCAGTGGTGATCGTCTGGGCACTG
TCCCTTCTATTAAACCACCGTAGCGAATTAGAAAAGGTTCAACAAGAATTGGACACTGTG
GTTGGAAAGGAAAGAAGAGTAGACATATCAGATATCAACAAACTTGAATATCTTCAGGCC
ATTGTTAAGGAAACATTCAGAATACGCCCTCCAGGTGCACTTCTCGTCCCTAGAGAATTC
ACAGACGACTGCACATTGGCTGGTTACCATATTCCAAAAGGCACCATGCTCTTTGTCAAC
TTATGGAAGTTACAAAAAGACCCAACTTTGTATCCTAATCCATTAGAGTTCAAGCCTGCA
AGGTTTCTGGAACCAAAGTATAAAGACATTGATCCTAGAGGTCGCCATTTTGAATTGTTT
CCATTTGGTGCTGGTCGAAGAAGTTGCCCAGGCCTAAATCTTGGCATCCAAAATGTGCAT
TTGATTTTGGCCAATTTGTTGCACTCATTTAATATATCAACAATCAATGATAAGCCGCTG
GATTTGAATGCGTCTCCTGATGGGGTAATTACTAGGAAGGCAACTCCTCTTGAAATCCGT
ATTTCACCTCGCCTATCTCCAGATCTTTACTAG

>CYP82AR2: protein. sequence from *Lithospermum erythrorhizon*

MELSFNSIFETPIILGVLLVITILIWLKITLSPRLKTPPEVGVALPIIGHLYLKALRGSKKPLFIKFTNLAEKFGP
IYNVRLGSIRAVVISNSELARELFTAKDNFVLARPKSLATSHLAYSYANLGVAPPTPYWRWLRKFTAVGF
FSHRALDMAKNVPATEIKLSIKYLYDLCHDEGSARIADMQQWLLDIGLNLMMRTVVGKSTSTSDDNADE
EEAKERRRWKKMMDDTMRMLFLPVLSDSIPLLKPLDIGGIEKEMKQVKKTMDEIVDQWLKEHIQKKAN
GIHVDAEKDFMDLLLAAVEDGDVELGGYHPHEVVKATCMSMVGAGSDTTSVVIVWALSLLLNHRSELE
KVQQELDTVVGKERRVDISDINKLEYLQAIVKETFRIRPPGALLVPREFTDDCTLAGYHIPKGTMLFVNLW
KLQKDPTLYPNPLEFKPARFLEPKYKDIDPRGRHFELFPFGAGRRSCPGLNLGIQNVHLILANLLHSFNISTI
NDKPLDLNASPDGVITRKATPLEIRISPRLSPDLY

>AeDSH1: protein. sequence from *Arnebia euchroma*

MELSFNSIFETPIIIGVLLVITILIWLKITFSTRLKTPPEIGFALPIIGHLYLKALGGSKKPLFIKFTNLAEKFGPI
YNVRLGSIRAVVISDTELARELFTAKDNYVLARPKSLATEHLAYSYANLGVAPPTPYWKWLRKFTAVGF

FSHRALDMAKNVPATEIKLSIKYLYDLCNDKGSARIADMQQWLLDIGLNLMMRTVVGKSTSTSDDNADE
EEAKERRRWKKMMDDTMRMLFLPVLSDSIPLLKPFDIGGIEKEMKQVKKTMDEIVDQWVKEHIQKRAN
GIHVDAEKDFMDLLLAAVEDGDVELGGYHPHMVVKATCMSMVGAGSDTTSVVIVWALSLLLNHRSELE
KVQQELDTVVGRERRVDISDINKLEYLQAVVKETFRIRPPGALLVPREFTDDCTLAGYHIPKGTMLFINLW
KLQKDPKLYPNPLEFKPARFLEPKYKDIDPRGRHFELFPFGAGRRSCPGLNLGIQNVHLILANLLHSFNISTI
DDKPLDLNASPDGVITRKATPLEIRITPRLSPDLY

>AeDSH2: protein. sequence from *Arnebia euchroma*
MELSFNSYFETPVLIGVLVLCITILILFNRKNSPPTKTPPEVGVALPIIGHLYLKALRGNIALHLKFTSLAEKF
GPIYNIRLGSIRAVVVSTTELAKELFTTHDNFILTRPNSLASQHLAYSYANMGVTPPDTYWKWLRKFTAVE
FFSHRALEMAKNVPAEEINTAIKYLYSLCSNDKGNARIADMQQWLLDIALDLMMRTVIGKQTSGSSNTID
QEERRKWKKMMEDTMRLMFMPVLSDSIPLLKWFDIGGVEKEMIKVSKEMDEIVDVWLKEHMQKKANK
SNGDHADTQRDFMDAMLSAVDDGDAEFGGHSPHTIVKATVMAMVGAGSDTTAVVIIWALSLLLNDRSK
LRKAQQELDTIVGKERKVDISDINKLDYLQAIVKETFRIHPPGALLIPREFSEDCIVGGYYVPKGTMLFINL
YKLQRDPETYPNPSEFLPERFLEPKYKDIDPRGRHYELFPFGAGRRSCPGLNLGIQNVHLILANLLHAFDVS
TVNDQLVDMTASVGVITRKAAPLEILITPRLSPDLYY

>EpDSH1: protein. sequence from *Echium plantagineum*
MEFFLNSIFETPIIIGVLLVITILIWLKINNSNRLKTPPEVGFALPIIGHLYLKELRGRKPLFIKFTKLAEKFGPI
YNVRLGSIRGVVISSSELAREIFTAKDNYVLARPKSLATGHLAYSYANLGVAPPTPYWKWLRKFTAVGFF
SHRALDMAKNVPATEIKLSIRYLYDLCNDKGSARIADMQQWLLDIGLNLMMRTVVGKSTATSDDADDE
EARERRKWKKMMDDTMRMLFLPVLSDSIPVLKPLDIGGVEKEMIQVKKVMDEIVEEWLKEHIQKKANGI
AVDGERDFMDLLLAAVEDGDAELGGYHPHMVVKATCMSMVGAGSDTTSVVIVWALSLLLNHRAELEK
VQRELDNVVGRGRRVDISDINKLDYLQAVVKETFRIRPPGALLVPREFTDDCTVGGYHIPKGTMLFINLW
KLQKDPNLYPDPNVYRPARFLEPKYKDIDPRGRHFELFPFGAGRRSCPGLNLGIQNVHLILANLLHSFDIST
IDNKLVDLNASPDGVITRKATPLEIRISPRLSPDLY

>EpDSH2.1: protein. sequence from *Echium plantagineum*
MEFFFNSYFETPVIIGILLVGITILVLFNSKSSRRTKTPPEVGVALPIIGHLHLQALRGNQPLHLKLTTLAEKF
GPIYNIRLGSIPAVVVSSTELAKELFTTHDNFILKRPNSLASEHLAYSYANMGVTPPNAYWRWLRKFTAVE
FFSHRALEMAKNVPAEEINTAIKYLYTLCLNDKGSARLADIQQWLLDIALDLMMRTVIGKQTSGGSSNNI
AEEERRRWKKMMEETMRLMFMPVLSDSIPLLKWFDIGGVEKEMKRVSKYMDEIVDVWLQEHIQKKAN
NTTNNGDDTQRDFMDAMLSAVEAGDAEFGDHSPHTIVKATVMAMVGAGSDTTAVVIIWALSLLLNDRS
KLRKAQQELDTVVGNNRRVDLSDINKLEYLQAIVKETFRIHPPGALLIPREFSEDRIVGGYHVPKGTMLFI
NLYKLQRDPKIYPNPSEFKPERFLEPKYKDIDPRGRHYELFPFGAGRRSCPGLNLGIQNVHLILANLLHTFD
ISTVDDDQLVDMTASPEAITRKASPLEVVITPRLSPDLYY

>EpDSH2.2: protein. sequence from *Echium plantagineum*
MEFFFNSYFETPVIIGILLVGITILVLFYSKSSRRTKTPPEVGVALPIIGHLHLQALRGNQPLHLKLTTLAEKF
GPIYNIRLGSIPAVVVSSTELAKELFTTHDNFILKRPNSLASEHLAYSYANMGVTPPNAYWRWLRKFTAVE
FFSHRALEMAKNVPAEEINTAIKYLYTLCSNDKGSARLADIQQWLLDIALDLMMRTVIGKQTSGGSSNNI
AEEERRRWKKMMEETMRLMFMPVLSDSIPLLKWFDIGGVEKEMKRVSKYMDEIVDVWLQEHIQKKAN
NTTNNGDDTQRDFMDAMLSAVEAGDAEFGDHSPHTIVKATVMAMVGAGSDTTAVVIIWALSLLLNDRS
KLRKAQQELDTVVGNNRRVDLSDINKLEYLQAIVKETFRIHPPGALLIPREFSEDRIVGGYHVPKGTMLFI

NLYKLQRDPKIYPNPSEFKPERFLEEKYKDIDPRGRHYELFPFGAGRRSCPGLNLGIQNVHLILANLLHTFD
ISTVDDDQLVDMTASPEAITRKASPLEVVITPRLSPDLYY

>LoDSH1 : protein. sequence from *Lithospermum officinale*
MELSFNSIFETPIILGVLLVITILIWLKITLSPRLKTPPEVGVALPIIGHLYLKALRGSKKPLFIKFTNLAEKFGP
IYNVRLGSIRAVVISNSELARELFTAKDNFVLARPKSLATTHLAYSYANLGVAPPTPYWRWLRKFTAVGF
FSHRALDMAKNVPATEIKLSIKYLYDLCNDEGSARIADMQQWLLDIGLNLMMRTVVGKSTSTSDDNADE
EEAKERRRWKKMMDDTMRMLFLPVLSDSIPLLKPFDIGGIEKEMKQVKKTMDEIVDQWLKEHIQKKAN
GIHVDAEKDFMDLLLAAVEDGDVELGGYHPHEVVKATCMSMVGAGSDTTSVVIVWALSLLLNHRSELE
KVQQELDTVVGKERRVDISDINKLEYLQAIVKETFRIRPPGALLVPREFTDDCTLAGYHIPKGTMLFVNLW
KLQKDPTLYPNPLEFKPARFLEPKYKDIDPRGRHFELFPFGAGRRSCPGLNLGIQNVHLILANLLHSFNISTI
NDKQLDLNASPDGVITRKATPLEIRISPRLSPDLY

>LoDSH2 : protein. sequence from *Lithospermum officinale*
MDFFDSYYETPVIIIRVLLLCIPILFLFNRRNSPSTKTPPQVGVALPIIGHLHHQALRGDQPLHLKFTSLAEKF
GLIYNIRLGSIRAVVVSSTELAKELFTTHDNFILTRPNSLASQHLAYSYANMGVTPPDTYWKWLRKFTAVE
FFSHRALEMAKNVPAEEINTSIKYLYNLCSNEKGNSRIADMQQWLLDIALDLMMRTVIGKQTSEGSNNVD
EEERRRWKKMMEDTMRLMFMPVLSDSIPLLKWFDIGGVEKEMIKVSKEMDEIVDVWLKEHMQKKANK
TNGDQSDTQRDFMDAMLAAVEEGDAEFGRHSPYTIVKATVMAMVGAGSDTTAVVIIWALSLLLNDRSK
LRKAQQELDTVVGKERRVDISDINKLEYLQAIVKETFRIHPPGALLIPREFSEDCIVGGYHVPKGTMLFINL
YKLQRDPKTYPDPSEFKPERFLEPKYKDIDPRGRHYELFPFGAGRRSCPGLNLGIQNVHLILANLLHAFDV
STVDDQLVDMTASVGVITRKAAPLEILITPRLSPDLYY

## References


1. Song, W.; Zhuang, Y.; Liu, T., Potential role of two cytochrome P450s obtained from *Lithospermum erythrorhizon* in catalyzing the oxidation of geranylhydroquinone during shikonin biosynthesis. *Phytochemistry* **2020,** *175*, 112375.

2. Fang, R.; Wu, F.; Zou, A.; Zhu, Y.; Zhao, H.; Zhao, H.; Liao, Y.; Tang, R. J.; Yang, T.; Pang, Y.; Wang, X.; Yang, R.; Qi, J.; Lu, G.; Yang, Y., Transgenic analysis reveals LeACS-1 as a positive regulator of ethylene-induced shikonin biosynthesis in *Lithospermum erythrorhizon* hairy roots. *Plant*. *Mol*. *Biol*. **2016,** *90*, 345-58.

3. Tang, C. Y.; Li, S.; Wang, Y. T.; Wang, X., Comparative genome/transcriptome analysis probes Boraginales' phylogenetic position, WGDs in Boraginales, and key enzyme genes in the alkannin/shikonin core pathway. *Mol. Ecol. Resour.* **2020,** *20*, 228-241.

4. Wu, F. Y.; Tang, C. Y.; Guo, Y. M.; Bian, Z. W.; Fu, J. Y.; Lu, G. H.; Qi, J. L.; Pang, Y. J.; Yang, Y. H., Transcriptome analysis explores genes related to shikonin biosynthesis in Lithospermeae plants and provides insights into Boraginales' evolutionary history. *Sci. Rep.* **2017,** *7*, 4477.

5. Rai, A.; Nakaya, T.; Shimizu, Y.; Rai, M.; Nakamura, M.; Suzuki, H.; Saito, K.; Yamazaki, M., *De Novo* transcriptome assembly and characterization of *Lithospermum officinale* to discover putative genes involved in specialized metabolites biosynthesis. *Planta Med.* **2018,** *84*, 920-934.

6. Takanashi, K.; Nakagawa, Y.; Aburaya, S.; Kaminade, K.; Aoki, W.; Saida-Munakata, Y.; Sugiyama, Akifumi.; Ueda, M.; Yazaki, K., Comparative proteomic analysis of *Lithospermum erythrorhizon* reveals regulation of a variety of metabolic enzymes leading to comprehensive understanding of the shikonin biosynthetic pathway. *Plant Cell Physiol.* **2018,** *60*, 19-28.

7. Auber, R. P.; Suttiyut, T.; McCoy, R. M.; Ghaste, M.; Crook, J. W.; Pendleton, A. L.; Widhalm, J. R.; Wisecaver, J. H., Hybrid de novo genome assembly of red gromwell (*Lithospermum erythrorhizon*) reveals evolutionary insight into shikonin biosynthesis. *Hortic. Res.* **2020,** *7*, 82.

8. Wang, L.; Feng, Z.; Wang, X.; Wang, X.; Zhang, X., DEGseq: an R package for identifying differentially expressed genes from RNA-seq data. *Bioinformatics* **2010,** *26*, 136-8.

9. Brachmann, C. B.; Davies, A.; Cost, G. J.; Caputo, E.; Li, J.; Hieter, P.; Boeke, J. D., Designer deletion strains derived from *Saccharomyces cerevisiae* S288C: a useful set of strains

and plasmids for PCR-mediated gene disruption and other applications. *Yeast* **1998,** *14*, 115-32.

10. Liu, X.; Cheng, J.; Zhang, G.; Ding, W.; Duan, L.; Yang, J.; Kui, L.; Cheng, X.; Ruan, J.; Fan, W.; Chen, J.; Long, G.; Zhao, Y.; Cai, J.; Wang, W.; Ma, Y.; Dong, Y.; Yang, S.; Jiang, H., Engineering yeast for the production of breviscapine by genomic analysis and synthetic biology approaches. *Nat. Commun.* **2018,** *9*, 448.

11. Jiang, J.; Yin, H.; Wang, S.; Zhuang, Y.; Liu, S.; Liu, T.; Ma, Y., Metabolic engineering of *Saccharomyces cerevisiae* for high-level production of salidroside from glucose. *J. Agric. Food Chem.* **2018,** *66*, 4431-4438.

12. Amberg, D. C.; Burke, D. J.; Strathern, J. N., Methods in Yeast Genetics: A Cold Spring Harbor Laboratory Course Manual, 2005 Edition (Cold Spring). **2005**.

13. Krithika, R.; Srivastava, P. L.; Rani, B.; Kolet, S. P.; Chopade, M.; Soniya, M.; Thulasiram, H. V., Characterization of 10-hydroxygeraniol dehydrogenase from *Catharanthus roseus* reveals cascaded enzymatic activity in iridoid biosynthesis. *Sci. Rep.* **2015,** *5*, 8258.

14. Scheler, U.; Brandt, W.; Porzel, A.; Rothe, K.; Manzano, D.; Bozic, D.; Papaefthimiou, D.; Balcke, G. U.; Henning, A.; Lohse, S.; Marillonnet, S.; Kanellis, A. K.; Ferrer, A.; Tissier, A., Elucidation of the biosynthesis of carnosic acid and its reconstitution in yeast. *Nat. Commun.* **2016,** *7*, 12942.

15. Y. Ikeda, N. I., C. Fukaya, K. Yokoyama, M. Tabata, H. Fukui,; Honda, G., Determination of the ratio between optical isomers, shikonin and alkannin by high performance liquid chromatography analysis. *Chem. Pharm. Bull.* **1991,** *39*, 2351-2352.

16. Tamura, K.; Stecher, G.; Peterson, D.; Filipski, A.; Kumar, S., MEGA6: molecular evolutionary genetics analysis version 6.0. *Mol. Biol. Evol.* **2013,** *30*, 2725-2729.