

Genome-wide Identification and Characterization of the STR Gene in *Dendrobium officinale*

Xiang Ren

Anhui Agricultural University

Wenhao Dong

Anhui Agricultural University

Chen Chen

Anhui Agricultural University

Jingjing Song

Anhui Agricultural University

Lianping Wei

Anhui Agricultural University

Cheng Song

West Anhui University

Qing Jin

jinqing169@126.com

Anhui Agricultural University

Research Article

Keywords: *Dendrobium officinale*, Strictosidine synthetase, bioinformatics analysis, gene expression, prokaryotic expression

Posted Date: May 24th, 2024

DOI: <https://doi.org/10.21203/rs.3.rs-4409446/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

Genome-wide Identification and Characterization of the STR Gene in *Dendrobium officinale*

Xiang Ren¹, Wenhao Dong¹, Chen Chen¹, Jingjing Song¹, Lianping Wei¹, Cheng Song^{2*}, Qing Jin^{1*}

¹ School of Life Sciences, Anhui Agricultural University, Hefei 230036, Anhui, China

² Anhui Province Key Laboratory for Quality Evaluation and Improvement of Traditional Chinese Medicine, College of Biological and Pharmaceutical Engineering, West Anhui University, Luan 237012, China

*Corresponding authors:

jinqing169@126.com(Qing Jin)

lanniao812329218@163.com(Cheng Song)

Abstract

Alkaloids are the main active ingredients in the traditional Chinese medicine *Dendrobium spp.* Terpenoid indole alkaloids (TIAs), a class of secondary metabolites with a nitrogen heterocyclic structure, are low in content and difficult to synthesize chemically. Strictosidine synthetase (STR) is a key enzyme that catalyzes the Pictet-Spengler reaction of TIAs. However, the molecular characteristics and enzymatic properties of STR proteins in *Dendrobium officinale* remain unclear. Through bioinformatics analysis, we systematically studied the protein purification and catalytic activity of *DoSTR*. We initially identified eight STR homologous genes in *D. officinale*. Members of the *DoSTR* family vary in molecular weight, theoretical isoelectric point, and subcellular localization. Cis-acting element analysis of the *DoSTR* promoter revealed the presence of multiple elements involved in phytohormone response and biotic and abiotic stress. Phylogenetic tree and homology alignment revealed that *DoSTR* proteins are more closely related to STR proteins of *Dendrobium huoshanense*; however, *DoSTR* are evolutionarily distant from plant STR proteins with characterized functions, such as CrSTR, RvSTR, and RsSTR. Conservative motif analysis revealed a conserved motif in the STR sequence. The Glu309 residue, which is associated with catalysis, is not conserved. Using qRT-PCR, the expression pattern of *DoSTR* was found to be spatially specific. Through transient expression in tobacco, *DoSTR8* was shown to localize to cell membranes. The *DoSTR8* recombinant protein was successfully expressed in *E. coli* using

prokaryotic expression and purified. The findings of this study may provide valuable insights into further research on the function of the STR gene family in *D. officinale*.

Keywords: *Dendrobium officinale*, Strictosidine synthetase, bioinformatics analysis, gene expression, prokaryotic expression

Introduction

D. officinale, a member of the Orchidaceae family, is a perennial epiphytic herb. Its active ingredients primarily consist of alkaloids, polysaccharides, flavonoids, and phenanthrenes(Tang et al. 2017). Dendrobium alkaloids are the primary active compounds in traditional Chinese medicine known as *Dendrobium nobile*. Terpenoid indole alkaloids (TIAs), a class of secondary metabolites with a nitrogen heterocyclic structure(O'Connor and Maresh 2006), of which about 3,000 have been identified in nature, and most TIAs have good biological activities such as therapeutic for hypertension, anticonvulsant, anti-tumor, anti-malarial, and anti-arrhythmic. Terpenoid indole alkaloids (TIAs) are present in low concentrations and difficult to synthesize chemically(Thamm et al. 2016). Despite the complex structure of TIAs, they are all derived from a common intermediate, strictosidine synthetase (STR), which is a crucial enzyme that catalyzes the biosynthetic pathway of TIAs(Sohani et al. 2009). However, the molecular characteristics and synthesis pathway of TIAs in *D. officinale* remain unclear.

STR genes are ubiquitous in animals, plants, and bacteria, with different family members serving distinct physiological functions(Stöckigt et al. 2008). The isolation, purification, biochemical characterization, and enzymatic reaction mechanism of STR from *Catharanthus roseus* and *Rauvolfia serpentina* demonstrated that the protein structure of STR enzymes is characterized by a hexa-lobed β -helix fold. The His307 and Glu309 residues represent the crucial catalytic active sites of the enzyme. STR-catalyzed reactions involve amine-formaldehyde condensation reactions(Zhu et al. 2015), and both tryptamine and secologanin, as well as their derivatives, can serve as catalytic substrates for the enzyme.. This leads to the production of structurally diverse isocarotenoside derivatives(Fischereder et al. 2014), further facilitating the synthesis of various novel structural alkaloids(Montiel et al. 2011). Although STR has been studied in some medicinal plants, there is limited research on the structure and enzymatic

characteristics of STR proteins in *D. officinale*, as well as how it regulates the synthesis of TIAs downstream.

In this study, we analyzed eight STR homologous genes in *D. officinale* using bioinformatics methods. Phylogenetic and homology analyses of STR family members from diverse plant species have revealed that these genes are evolutionarily conserved and exhibit specificity in unity. Heat map analysis based on RNA-seq data demonstrated that STR gene expression is time-dependent and highly responsive to MeJA(Pan et al. 2022, Pan et al. 2016). Further gene expression analysis showed spatial specificity. The findings of this study provide a theoretical basis for subsequent research on STR family genes and offer insights into the deeper development and utilization of *D. officinale*.

Materials and methods

Identification of *Dendrobium officinale* STR family genes

The *D. officinale* whole genome database were downloaded from NCBI(BioProject accession: PRJNA662181)(Sharma et al. 2018). Then, based on the Pfam number (PF03088), retrieved and filtered candidate STR protein members using the Hidden Markov Model (HMM) of STR conserved structural domains from the online Pfam Protein Database (<https://pfam.xfam.org/>). Subsequently, sequences lacking STR conserved structural domains were eliminated using Pfam and SMART (<http://smart.embl-heidelberg.de/>), whereas redundant sequences were removed through an online Expasy website, resulting in the final acquisition of DoSTR family member sequences.

Physicochemical characterization of the STR gene family

The theoretical pI and molecular weight of the identified STR were calculated using the ProtParam website (<https://web.expasy.org/protparam/>). Furthermore, we predicted the subcellular localization of STR proteins using the GenScript website (<https://www.genscript.com/wolf-psort.html>).

Phylogeny of the STR gene family, amino acid comparisons and gene structure analysis

The protein sequences of the STR from *Rauvolfia serpentina* (GenBank ID: [P68175.1](#)), *Ophiorrhiza pumila* (GenBank ID: [BAB47180.1](#)), and *Ophiorrhiza japonica* were (GenBank ID: [ACF21007.1](#)) obtained from the data reported by Zhang, Zhou et al., in conjunction with the NCBI online website (<https://www.ncbi.nlm.nih.gov/>) (Cui et al. 2015, McKnight et al. 1991, Verma et al. 2015, Zhang et al. 2022). Sequence comparison was carried out using the ClustalW program in MEGA11 software along with escript online software (<https://escript.ibcp.fr/ESPrpt/cgi-bin/ESPrpt.cgi>), and a molecular evolution tree was constructed using the maximum likelihood (ML) method, with bootstrap replicates set at 1000 times. Gene structures were visualized and analyzed using TBTools.

Cis-Acting element analysis

The DNA sequence of 2Kb upstream of the start codon of the DoSTR gene was obtained from the *D. officinale* genome database, which serves as the promoter of the DoSTR gene. The cis-regulatory elements of the DoSTR gene were analyzed using the PlantCARE online tool and the gene structure was visualized using TBTools (Jiao et al. 2022) (<https://bioinformatics.psb.ugent.be/webtools/plantcare/html/>).

Plant material and growing conditions

D. officinale protocorm-like bodies (PLBs) were induced by the *D. officinale* group at the College of Life Sciences, Anhui Agricultural University, and were subjected to suspension succession culture with stable morphological characteristics and growth rate. The culture medium consisted of 1/2 MS + 0.1 mg/L NAA + 0.1 g/L hydrolyzed lactic protein (Jin et al. 2016), and the culture conditions included dark culture, a shaker speed of 120 r/min, a culture temperature of 25°C, and a succession cycle of 48 days (Jiang et al. 2023). In natural environments, *D. officinale* is an annual plant, and plants with good growth conditions (15-20 cm) were selected for testing.

qRT-PCR analysis

Total RNA was extracted from different tissues of *Dendrobium officinale* using Monad Total RNA Extraction Reagent. The concentration and quality of RNA were detected using a NanoDrop2000 Ultra-Micro Spectrophotometer, and the RNA was reverse transcribed into cDNA using Mona

MonScript™ RTIII Super Mix with dsDNase (Two-Step) Kit following the manufacturer's instructions.

qRT-PCR was performed using the Burroughs instrument template SYBR program. The reaction system consisted of: 2μl of cDNA Template (diluted), 1μl each of upstream and downstream primers, 8μl of PerfectStart® Fast Green qPCR SuperMix, and the volume was replenished with ddH₂O to a total volume of 20μl. The reaction was performed using a 40× cycling program (95°C 10 min; 95°C 10 s; 55°C 10 s; 72°C 30 s; 95°C 15 s; 65°C 5 s; 99°C 5 s). The ACTIN gene was used as an internal control. Relative expression levels were calculated using the 2-ΔΔCt method. Three independent biological experiments were conducted.

Heterologous expression and purification of proteins

Inoculate the expression strain containing the recombinant plasmid into 5 mL of liquid ampicillin LB medium at 37 °C, with 150 rpm overnight culture. Next, inoculate the above bacterial solution according to a 1:50 ratio into 100 mL of liquid ampicillin LB medium, 37 °C, 150 rpm culture until the OD₆₀₀ value of the bacterial solution is approximately 0.6-0.8. Add IPTG to a final concentration of 0.6 mmol/L, and incubate at 16 °C for 12 hours.

The bacterial solution at 12000 r/min, 4 °C, and centrifuged for 10 min to obtain the bacteria. Ultrasonic cell crusher was used for cell crushing, the working conditions of ultrasonic cell crusher were: power 12%; total working time 20min; ultrasonic on time 4s; ultrasonic off time 6s. 12000 r/min, centrifugation at low temperature for 10min to collect the supernatant, and stored at low temperature. His Sep Ni-NTA Agarose Resin was used for protein purification according to the instructions.

Collect 2 mL of bacterial solution and purified protein solution, and then add 200 μL of PBS buffer (pH=7.0). Next, 40 μL of sampling buffer was added to every 10 μL of sample and mixed well. The samples were placed in a metal bath at 100°C for 15 min. Finally, the expression of recombinant proteins was detected by SDS-PAGE at the end of the process.

Subcellular localization

To analyze the subcellular localization of the DoSTR8 protein, *Agrobacterium tumefaciens*

containing the target gene was inoculated into LB liquid medium supplemented with 50 µg/L kanamycin and 50 µg/L rifampicin for expanded culture. When the optical density (OD₆₀₀) of the bacterial solution reached 0.6-0.8, the bacteria were collected by centrifugation at 6,000 rpm and 4 °C for 10 min. The bacteria were then suspended in a buffer containing 0.01 mol/L MgCl₂, 0.01 mol/L MES, and 0.01 mmol/L As, and left to stand at room temperature, protected from light, for approximately 3 hours. A 1.0 mL syringe was used to suck up the *Agrobacterium* sap, which was slowly injected into the back of the tobacco leaf. Sap injection was stopped when it completely infiltrated the entire leaf, creating a well-marked area. The leaves were then treated in the dark for 2 days. To observe the distribution of GFP green fluorescence, tobacco leaves were scanned and photographed under a laser confocal microscope. The excitation wavelengths for GFP and chlorophyll were 488 and 640 nm, respectively.

Results

Identification and characterization of the STR genes in *Dendrobium officinale*

A total of 18 candidate STR homologous genes were identified from the *D. officinale* genome using the Simple HMM Search function of Tltools. After further BLAST conserved structural domain analysis, 12 candidate STR homologous gene members were screened, and eight STR gene family members were obtained by removing redundant sequences with sequence similarity greater than 90%. These were sequentially named *DoSTR1-8*.

Essential features of the STR gene and encoded proteins are predicted, including molecular weight, theoretical isoelectric point, instability coefficient, hydrophobicity index, and subcellular localization (Table 1). The length of the DoSTR proteins ranged from 212-412 amino acids. The molecular weight of the proteins varied from a minimum of 23.57 kDa (DoSTR7) to a maximum of 46.45 kDa (DoSTR1), and the range of the theoretical isoelectric point was from 4.73 (DoSTR7) to 8.67 (DoSTR6). Additionally, DoSTR1, DoSTR4, DoSTR7, DoSTR5, DoSTR6, and DoSTR7 (instability coefficient >40) were identified as unstable proteins. DoSTR8, which had an average hydrophobicity index greater than zero, was characterized as a hydrophobic protein. Subcellular localization predictions revealed that DoSTR1, DoSTR3, DoSTR4, and DoSTR7 were localized in the extracellular matrix; DoSTR2 in the endoplasmic reticulum; DoSTR5 in the chloroplast;

DoSTR6 in the cytoplasm; and DoSTR8 in the vesicle.

Table 1 Physicochemical properties of DoSTR proteins

Name	ID	Amino acid		PI	Instability		Subcellular
		residues	MW(Da)		index	index	localization
DoSTR1	Dca003087	325	35178.98	5.01	41.39	-0.039	extr
DoSTR2	Dca025929	412	46459.41	6.53	39.34	-0.148	E.R.
DoSTR3	Dca008250	393	43416.59	6.34	34.96	-0.149	extr
DoSTR4	Dca016630	374	42248.53	7.57	50.42	-0.154	extr
DoSTR5	Dca002500	377	41658.93	5.77	40.71	-0.124	chlo
DoSTR6	Dca012744	245	27796.2	8.67	40.12	-0.216	cyto
DoSTR7	Dca021666	212	23573.56	4.73	73.21	-0.23	extr
DoSTR8	Dca019567	353	38629.7	6.33	38.81	0.074	vacu

Phylogenetic relationships and amino acid analysis of DoSTR

A phylogenetic tree was constructed using Maximum Likelihood Estimation (ML) for DoSTR (Figure. 1). The results showed that DoSTR1, DoSTR2, DoSTR3, DoSTR5, DoSTR6, DoSTR7, and DoSTR8 were clustered into one evolutionary sub-branch with DhuSTR, which is a member of the same genus of the Orchidaceae family, and is more closely related. DoSTR4 clustered with the STR from the family Nyssaceae plant, *Camptotheca acuminata*, into one evolutionary sub-branch, indicating a closer genetic relationship. Although they are evolutionarily distant from plant STR proteins with characterized functions, such as CrSTR, RvSTR, and RsSTR, DoSTR8 can be clustered into a single evolutionary subbranch with ZmSTR proteins. AtSTR(Ma et al. 2006) and UrSTR(Sohani, Schenk, Schultz and Schmidt 2009), which are clustered with DoSTR2, DoSTR3, DoSTR5, DoSTR6, and DoSTR7, have been shown to have no catalytic functions.

Label Decorations

- ★ Dendrobium officinale
- Dendrobium huoshanense
- Rauvolfia verticillata
- Rauvolfia serpentina
- Ophiorrhiza pumila
- Ophiorrhiza japonica
- Zea mays
- Uncaria rhynchophylla
- Vigna unguiculata
- Medicago truncatula
- Catharanthus roseus
- Gelsemium sempervirens
- Mitragnyna speciosa
- ▲ Arabidopsis thaliana
- ▲ Camptotheca acuminata

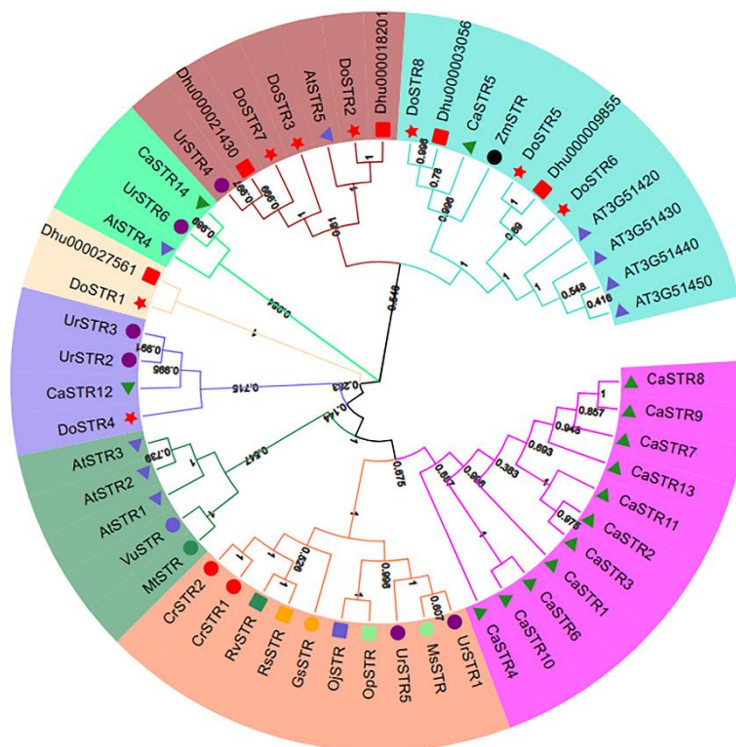


Figure. 1 Phylogenetic evolutionary tree of DoSTR proteins and STR proteins from other species.

Homologous amino acid comparisons with RsSTR (Figure. 2), which has been functionally characterized (Fischereder, Pressnitz, Kroutil and Lutz 2014), revealed that DoSTR2, DoSTR3, DoSTR4, and DoSTR7 contain the Cys89 site and Cys101 residues for the formation of disulfide bonds. However, DoSTR1, DoSTR5, DoSTR6, and DoSTR8 did not contain in Cys89 site and Cys101 residues, or had only one cysteine residue site. Glu309 in *Rauvolfia serpentina* is an important active catalytic site (Xu et al. 2022). The sequences of indole alkaloid-producing *Rauvolfia serpentina*, *Rauvolfia verticillata*, *Uncaria rhynchophylla*, *Ophiorrhiza japonica*, and *Ophiorrhiza pumila* were highly conserved near this site. In contrast, the sequences of STR proteins within *Dendrobium officinale* are more varied near the Glu309 site. It is hypothesized that there may be a difference in the key catalytic residue sites in the dendrobium.

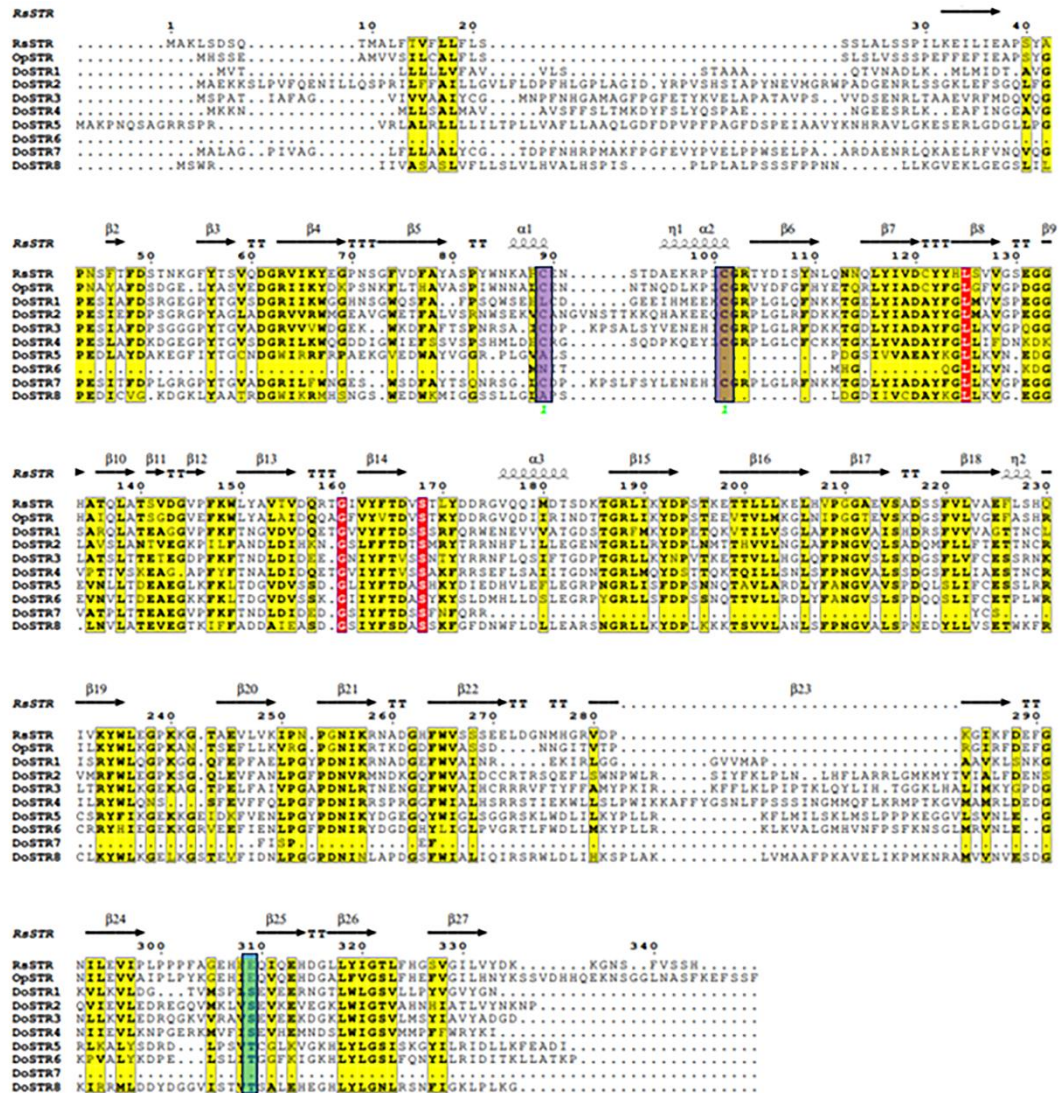


Figure. 2 Comparison of DoSTR family protein sequence acids with *R. serpentina* and *O. pumila* protein sequences. Cys89 and Cys101 loci are in violet font, and the Rovere Glu309 locus is in green font.

Gene structure, conserved motifs and promoter analysis of the DoSTR gene family

To further understand the structural and functional characteristics of DoSTR, the MEME website was used to analyze the motif distribution of DoSTR (Figure. 3). The results showed that a total of five motifs were identified in eight DoSTR proteins. DoSTR1, DoSTR2, DoSTR3, and DoSTR4 contained Motifs 1, 2, 3, 4, and 5 simultaneously; whereas all other DoSTR members except DoSTR7 contained Motif 4. Generally, DoSTR members from the same population exhibited unique motif compositions, suggesting the functional diversity of DoSTR gene families within the same population.

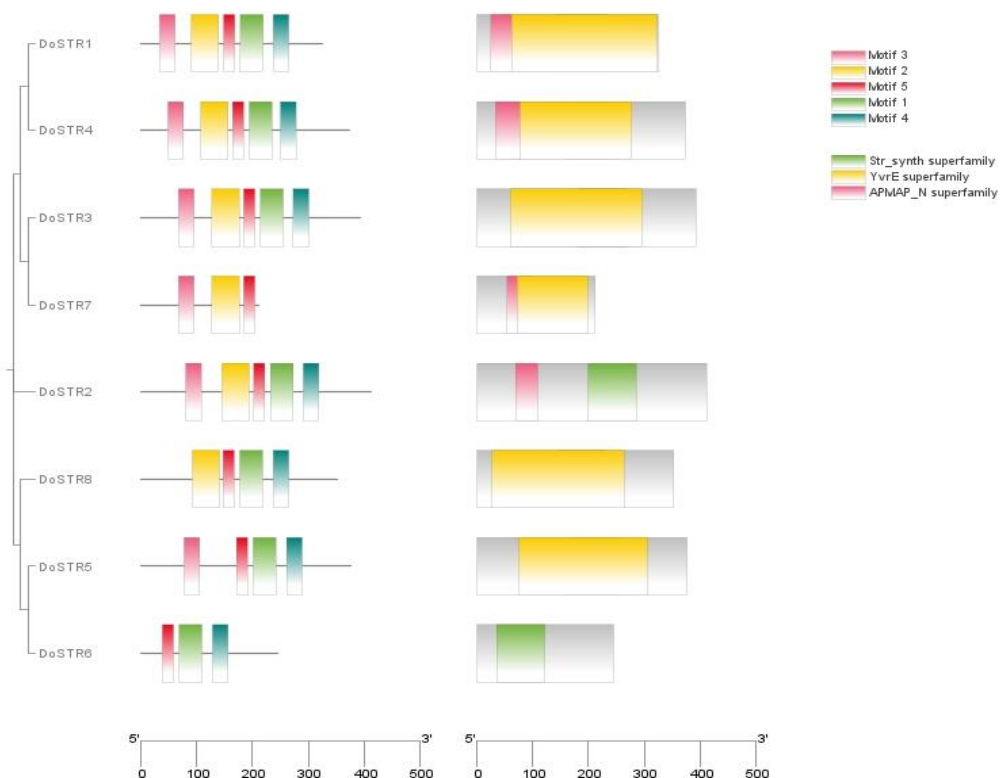


Figure. 3 Structural analysis of DoSTR family genes.

A region of 2000 bp upstream of the DoSTR start codon (ATG) was used for *cis*-acting element analysis(Figure. 4). The findings revealed that the *cis*-acting elements in the DoSTR promoter region could be categorized into three major groups: growth, development, and stress response-related elements. Among the plant growth and developmental regulators, circadian rhythm-related elements were identified. In the stress response category, elements associated with anaerobic induction, low-temperature induction, defense, and stress responses were detected. Notably, most of the *cis*-acting elements were associated with the methyl jasmonate response, suggesting that DoSTR expression might be influenced by both biotic and abiotic factors.

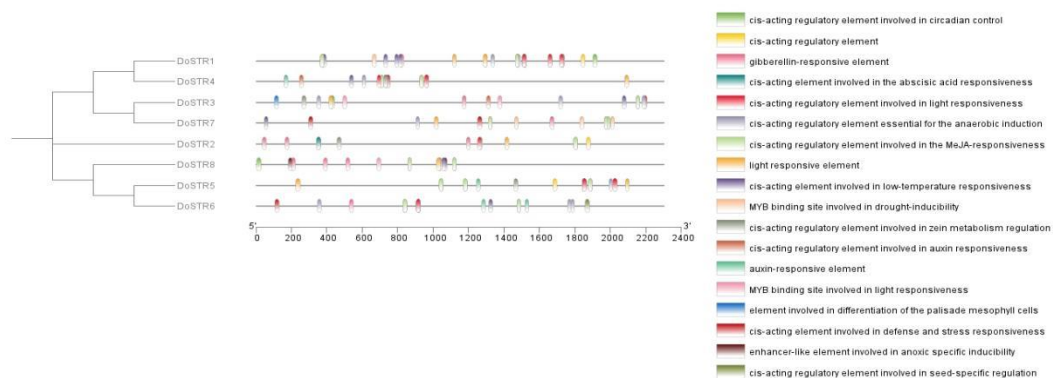


Figure. 4 Structural analysis and promoter visualization of DoSTR family genes.

Expression profiling of DoSTR family genes in *Dendrobium officinale* protocorm-like bodies (PLBs)

The expression levels of eight STR genes in PLBs were analyzed using the RNA-seq data obtained from Jiao's results(Jin, Jiao, Sun, Song, Cai, Lin, Fan and Zhu 2016) (Figure. 5A). The results revealed that *DoSTR2*, *DoSTR4*, *DoSTR6*, *DoSTR7*, and *DoSTR8* exhibited a similar trend of increasing expression from 0-24 h. However, several genes, such as *DoSTR1* and *DoSTR3* were highly expressed only at the beginning, followed by a decline in expression as time progressed. We also noticed that the expression of *DoSTR5* distinctly varied, reaching a peak at 4h and then gradually decreasing. These findings suggest that DoSTR family genes participate in various functions at different time points.

By analyzing the promoter cis-acting elements, we inferred that the expression levels of DoSTR family genes might be influenced by MeJA(Figure. 5B). To determine whether the DoSTR genes responded to the hormone MeJA, we compiled the MeJA-induced expression levels of STR genes using a heat map. The results indicated that the expression levels of most genes peaked 24 h after MeJA induction. Among them, *DoSTR2*, *DoSTR4*, *DoSTR5*, and *DoSTR8* showed an increasing trend in expression levels at 24 h. In contrast, the expression levels of *DoSTR1*, *DoSTR6*, and *DoSTR7* genes exhibited a downward trend followed by an upward trend, while the expression level of *DoSTR3* was suppressed. Overall, the DoSTR family of genes appeared to be responsive to MeJA induction.

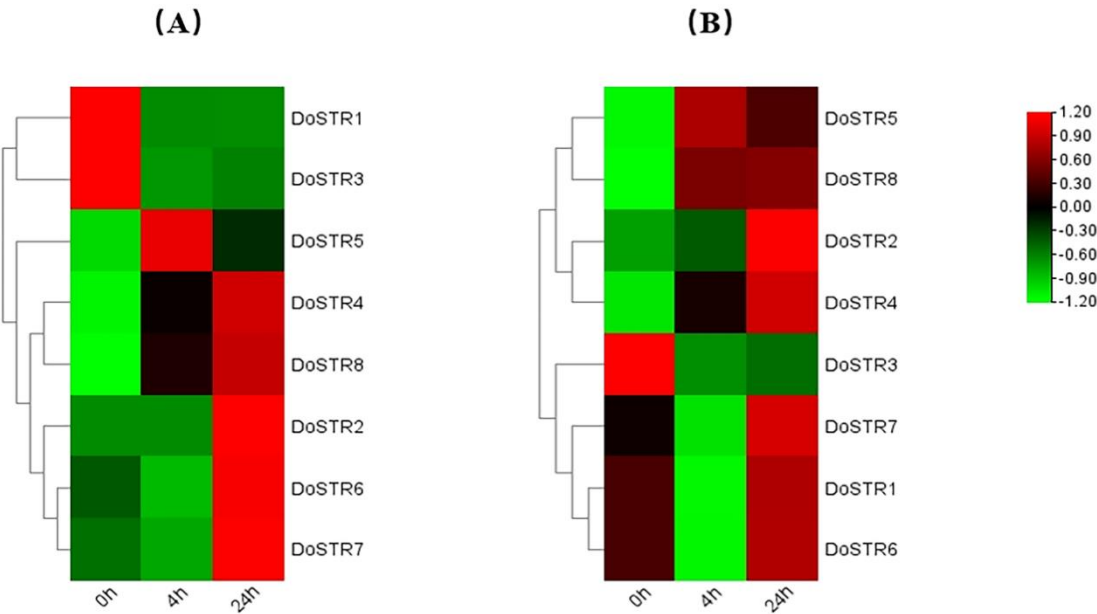


Figure. 5 Analysis of DoSTR family gene expression levels. (A) Expression pattern of DoSTR

family genes under normal conditions (B) Expression patterns of DoSTR genes in response to MeJA. RPKM values are shown on the heatmap using RPKM values and normalized to the processing color scale showing the increase in expression levels from green to red.

Tissue-specific analysis of DoSTR expression patterns

The differences in *DoSTR* expression in five tissues (root, stem, flower, leaf, and PLBs) were analyzed. The *DoSTR1* expression in the root is represented as "1"(Figure. 6). The results revealed that all eight *DoSTR* genes were expressed in the five *D. officinale* tissues, with significant differences and period-specific patterns. The expression levels of the eight DoSTR genes were the lowest in the protocorm-like bodies. However, the expression levels of *DoSTR8* and *DoSTR3* were higher in the protocorms, approximately 2-5 times greater than those of the other genes. The expression level of *DoSTR5* was significantly elevated, whereas *DoSTR2* exhibited a notable decrease in annual Dendrobium roots. In annual Dendrobium stems, the expression levels of all genes were higher except for *DoSTR2* and *DoSTR8*, which were lower. Among these genes, the expression level of *DoSTR4* in annual Dendrobium leaves was the highest among all tissues. In annual Dendrobium flowers, the expression level of *DoSTR8* was significantly increased in annual Dendrobium flowers.

Overall, the tissue specificity of the eight genes expressed in *D. officinale* was pronounced.

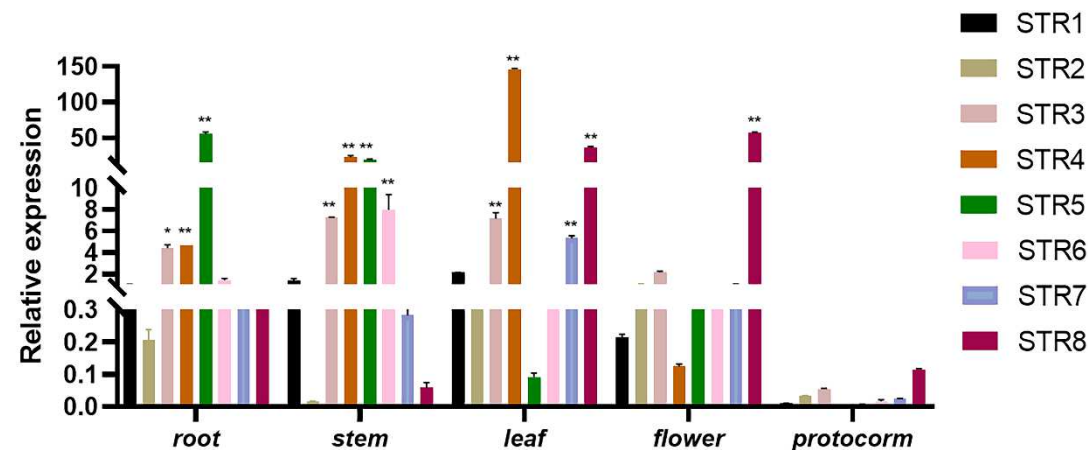


Figure. 6 Relative expression analysis of *DoSTR* in different tissues of *D. officinale*. Significance level (* p <0.05, ** p <0.01).

Heterologous expression of DoSTR8 protein

Given the bioinformatics analysis and tissue expression results, DoSTR8 was used to perform prokaryotic expression. The resulting target gene sequence was ligated into the pCold TF expression vector to construct the expression plasmid. This plasmid was then transformed into sensory *E. coli* ROSETTA (DE3) for heterologous expression.

The engineered bacteria were cultured at 16°C, 150 rpm, and induced by IPTG at a final concentration of 0.3 mmol/L. The expression of DoSTR8 in the prokaryotic system was successfully achieved(Figure. 7A), also successful purification of recombinant protein(Figure. 7B).

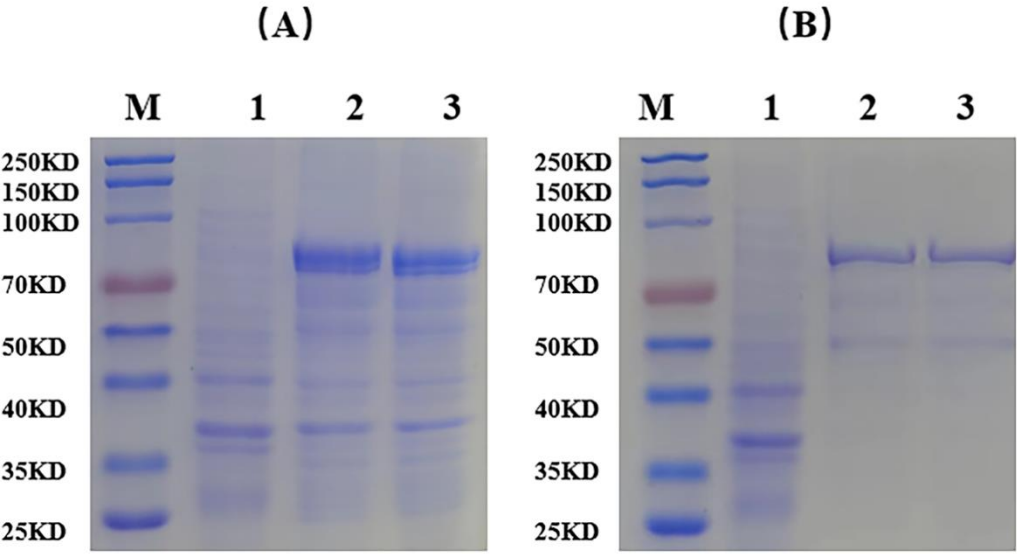


Figure. 7 Heterologous expression of DoSTR8 was detected by SDS-PAGE. (A) Crude protein of DoSTR8 recombinant protein detected by SDS-PAGE. (B) Purification of DoSTR8 recombinant protein detected by SDS-PAGE. The molecular weight of DoSTR8 recombinant protein is 90.6KD. lane 1: Marker; lane 2: Negative control; lane 3: DoSTR8 induced; lane 4: DoSTR8 induced.

Subcellular localization of DoSTR8 protein

The subcellular localization of DoSTR8-GFP were investigated by generating a DoSTR8-GFP fusion protein and transiently expressing it in tobacco. As a control, free GFP were used, which exhibits fluorescent signals in the cell membrane and nucleus when expressed alone. In contrast, the DoSTR8-GFP fusion protein fluoresced only in the cell membrane(Figure. 8). This suggests that DoSTR8 might be a protein localized in the cell membrane and requires the cell membrane to perform its corresponding functions.

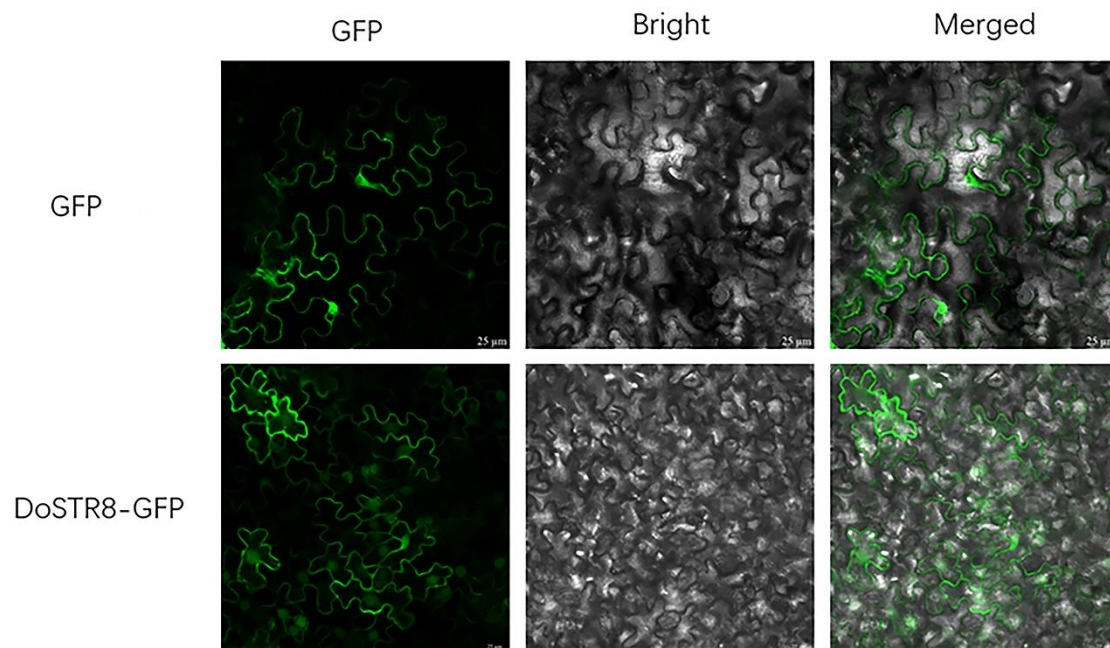


Figure. 8 Subcellular localization map of DoSTR8. pCAMBIA1305-GFP was used as a control, scale bar indicates 25 µm.

Discussion

D. officinale, a herbaceous plant belonging to the orchid family, is a valuable traditional Chinese medicine. Previous studies have revealed that the primary alkaloids in *D. officinale* consist of pyridines, isoquinolines, purines, amides, and terpene indole alkaloids(Sriram et al. 2005). Terpene indole alkaloids (TIAs) are a class of natural products with significant biological activities. Due to their unique therapeutic effects, TIA-based drugs are extensively used in modern medical treatment; for example, camptothecin produced by *C. roseus*, exhibits potent anti-tumor activity by inhibiting DNA topoisomerase I, and it is widely used in the clinical treatment of tumors(Liu et al. 2020). *O. zapumila* serpentine exhibits hypoglycemic effects and is used to manage diabetes. Vinblastine and vincristine, which originate from *C.roseus*, are currently used as clinical anti-tumor drugs due to their excellent anti-tumor effects(Škubník et al. 2021, Tisnerat et al. 2022). Quinine, extracted from the cinchona tree, has been historically used to treat malaria(Sharma, Verma, Mathur and Mathur 2018).Although TIAs are present in *D. officinale*, the biosynthetic pathway and metabolic regulatory mechanism of TIAs in this plant remain unclear. Strictosidine synthetase is a crucial enzyme involved in TIA synthesis. Numerous studies have

312 been conducted on the strictosidine synthetase gene family and its functional roles, such as the
313 overexpression of STR in *C.roseus* and *O.zapumila* hairy roots, leading to an increase in the
314 production of various TIAs, including aajmalicine、 catharanthin、 vincamine and
315 camptothecin(Cui, Ni, Ji, Teng, Yang, Wu, Zekria, Zhang and Kai 2015, Verma, Sharma, Khan,
316 Shanker and Mathur 2015, Zhang et al. 2004). Heterologous expression of *C. roseus* CrSTR in
317 tobacco resulted in a significant enhancement of STR activity in transgenic tobacco(McKnight,
318 Bergey, Burnett and Nessler 1991), with a 3-fold to 22-fold increase compared to *C. roseus*(Guo et
319 al. 2013). Several STR genes have also been identified in *D. officinale*, but limited research is
320 available on the specific functional studies of STR genes in this plant(Ma, Panjekar, Koepke, Loris
321 and Stöckigt 2006).

322 Members of the same family share a common origin and exhibit a high degree of similarity in
323 protein sequences, gene structures, and functional motifs. However, they also display diversity in
324 terms of expression patterns, protein function, etc. Phylogenetic analysis revealed, that the
325 homology of DoSTR proteins was closest to that of DhSTR proteins from another species of
326 orchid. However, because the *Dendrobium huoshanense* STR has not been investigated by
327 previous researchers, the functions of DoSTR could not be inferred from this comparison. When
328 considering the closely related species from the higher taxonomic level, it was found that the
329 species with similar relationships to DoSTR1-7 are *A. thaliana* and *U. rhynchophylla*. The nine
330 *Arabidopsis thaliana* AtSTR proteins used in this study have been demonstrated to lack catalytic
331 function(Ma, Panjekar, Koepke, Loris and Stöckigt 2006), and only UrSTR1 and UrSTR5 have
332 been shown to possess in vitro catalytic activity in the STR of *U. rhynchophylla*(Sohani, Schenk,
333 Schultz and Schmidt 2009). Therefore, DoSTR1-7 may not be catalytically active. In contrast,
334 DoSTR8 has a close homology with ZmSTR, suggesting that it may possess catalytic activity.

335 Amino acid comparisons revealed that DoSTR2, DoSTR3, DoSTR4, and DoSTR7 contain
336 the Cys89 and Cys101 sites for disulfide bond formation. In contrast, DoSTR1, DoSTR5,
337 DoSTR6, and DoSTR8 do not contain this site or contain only a single cysteine residue. It has
338 been demonstrated that Glu309 is an essential active catalytic site in *Rauvolfia*. The sequences of
339 indole alkaloid-producing *R. verticillata* , *R. serpentina*, *U. rhynchophylla*, and *C. acuminata*
340 exhibit significant conservation near this site, while the STR protein sequences in *D. officinale*
341 vary considerably near the Glu309 site. Consequently, it is proposed that *D. officinale* STR

proteins may exhibit diversity within the conserved sequences(Ma et al. 2004, Yamazaki et al. 2003).

Meanwhile, DoSTR8 and DoSTR5 possess similar conserved motifs, however, their expression patterns differ significantly. DoSTR8 is highly expressed in leaves and flowers, whereas DoSTR5 is primarily expressed in rhizomes. This distinct expression profile reflects the specificity of these genes, a characteristic that has also been observed for other genes. This unity of conservatism and specificity demonstrates the evolutionary process of the DoSTR. Terpene anabolism is highly tissue-specific, and the difference in the expression of the camptothecin anabolism gene in different tissues of *Catharanthus* seedlings at different growth stages is relatively significant, exhibiting period and tissue specificity(Luk et al. 2007). All except for CaSTRL candidate genes, CaSTR11 and CaSTR14, reached their highest expression levels in leaves at 40 days. The relative expression of tea tree CsSTR1 genes was higher in buds and young leaves (1st to 3rd leaves) and lower in old leaves, flowers, and fruits. The UrSTR gene exhibits its highest expression in the stem, approximately 5-fold higher than in the hook; the second highest expression is in the root, approximately 4-fold higher than in the hook, and the lowest expression is found in the hook and fruit. Analysis of DoSTR expression patterns revealed that they are primarily expressed in stems, followed by leaves and flowers. Among them, DoSTR4 was highly expressed only in the leaves, whereas DoSTR3 and DoSTR8 were predominantly expressed in *D. officinale* protocorm-like bodies. Orchids are primarily ornamental species, yet *D. officinale* possesses significant pharmacological activities.

The majority of STR genes in plants can be induced to express by MeJA and they exhibit responses to a broad spectrum of biotic and abiotic stresses. The highest count of cis-acting elements in the DoSTR gene was associated with the MeJA response, suggesting that DoSTR might be a gene induced or repressed by MeJA. Following the addition of the phytohormone methyl jasmonate (MeJA), the expression level of the DoSTR gene exhibited a specific trend within 24 h, indicating a short-term induction of the DoSTR gene by MeJA. Further research is required to explore whether DoSTR responds to other hormones or is affected by other biotic and abiotic stressors.

Heterologous expression of plant genes has been hindered by numerous factors such as the inherent nature of the gene, the type of expression vector, the choice of heterologous expression

host cells, and the involvement of molecular chaperones. Botanists have long been confronted with the challenges of enhancing the efficiency of heterologous expression and improving the solubility of recombinant proteins. Louis developed a recombinant plasmid containing a truncated NCS (with the deletion of the first 19 amino acids at the N-terminal position) and successfully optimized the induced expression conditions, resulting in the expression of NCS in *E. coli* from none to a high level(Qiying et al. 2021). Zhou constructed prokaryotic and eukaryotic expression vectors and successfully achieved the expression of the STR gene in *Camellia sinensis* using different host cells for expression(Zhang et al. 2021). Zhang addressed the issue of inadequate plant gene expression due to diverse codons in the prokaryotic expression system. In this study, we also developed a variety of plant expression vectors but were unable to achieve prokaryotic expression of the DoSTR8 protein. However, we have successfully constructed pCold TF vectors containing molecular chaperones specifically for prokaryotic induction, which significantly increased the expression level and solubility of DoSTR8 in prokaryotic cells. Further research is necessary to elucidate the protein's function and active catalytic site.

This study systematically investigated the STR gene family of *D. officinale*. Further exploration is needed to understand the specific catalytic reaction of STR in *D. officinale* and its regulation of indole alkaloid synthesis. Additionally, this study may serve as a valuable reference for subsequent studies on indole alkaloids in *D. officinale*.

Conclusions

In this study, eight candidate genes named DoSTR1-8 were identified and screened from the *D. officinale* genome database. Bioinformatics analysis revealed that STR family members exhibit varying molecular weights, theoretical isoelectric points, and cellular localization of amino acids. Notably, DoSTR2, DoSTR3, DoSTR4, and DoSTR7 possess Cys89 and Cys101 sites necessary for disulfide bond formation, whereas DoSTR1, DoSTR5, DoSTR6, and DoSTR8 contain no such sites or only a single cysteine residue site. Phylogenetic analysis indicates that *D. officinale* STR is more closely related to DhuSTR and more distantly related to functionally characterized STR proteins, such as *R. serpentina*, *R. verticillata*, *U. rhynchophylla*, *O. japonica*, and *O. pumila*. Real-time PCR was used to determine the relative expression of eight DoSTR genes in different

tissues (roots, stems, flowers, leaves, and PLBs) of annual *D. officinale*. The results revealed that the eight DoSTR genes were expressed in these tissues, with tissue specificity. Furthermore, the DoSTR8 recombinant protein was successfully produced in *E. coli* through the construction of prokaryotic expression vectors. Subcellular localization analysis indicated that DoSTR8 was predominantly localized to the cell membrane.

Data availability statement

The datasets presented in this study can be found in online repositories.

Ethics approval and consent to participate

The authors declare that permission to collect *Dendrobium officinale* material has been obtained, and experiments on the plants described in this paper comply with institutional, national and international guidelines.

Competing interests

The authors declare that they have no competing interests.

Funding

This work was supported by the Natural Science Research Project of Provincial Natural Science Foundation of Anhui (2108085MC80 and 1908085MC82), Anhui Scientific Research and Innovation Team of Quality Evaluation and Improvement of Traditional Chinese Medicine (2022AH010090) , and the High-level Talents Research Initiation Fund of West Anhui University (WGKQ2022025).

Author Contributions

X.R. and W.D. designed the experimental ideas; X.R. participated in the writing of the manuscript; C.C. and J.S. supervised the experiments; L.W. directed the experiments; C.S. and Q.J. revised the manuscript. C.S. and Q.J. acquire the funding. All authors made significant contributions to the draft.

References

- Cui L, Ni X, Ji Q, Teng X, Yang Y, Wu C, Zekria D, Zhang D, Kai G (2015) Co-overexpression of geraniol-10-hydroxylase and strictosidine synthase improves anti-cancer drug camptothecin accumulation in *Ophiorrhiza pumila*. *Sci Rep* 5:8227.<https://doi.org/10.1038/srep08227>
- Fischereder E, Pressnitz D, Kroutil W, Lutz S (2014) Engineering strictosidine synthase: rational design of a small, focused circular permutation library of the β -propeller fold enzyme. *Bioorg Med Chem* 22:5633-5637.<https://doi.org/10.1016/j.bmc.2014.06.023>
- Guo X, Li Y, Li C, Luo H, Wang L, Qian J, Luo X, Xiang L, Song J, Sun C, Xu H, Yao H, Chen S (2013) Analysis of the *Dendrobium officinale* transcriptome reveals putative alkaloid biosynthetic genes and genetic markers. *Gene* 527:131-138.<https://doi.org/10.1016/j.gene.2013.05.073>
- Jiang CX, Yu JX, Fei X, Pan XJ, Zhu NN, Lin CL, Zhou D, Zhu HR, Qi Y, Wu ZG (2023) Gene coexpression networks allow the discovery of two strictosidine synthases underlying monoterpene indole alkaloid biosynthesis in *Uncaria rhynchophylla*. *Int J Biol Macromol* 226:1360-1373.<https://doi.org/10.1016/j.ijbiomac.2022.11.249>
- Jiao C, Wei M, Fan H, Song C, Wang Z, Cai Y, Jin Q (2022) Transcriptomic analysis of genes related to alkaloid biosynthesis and the regulation mechanism under precursor and methyl jasmonate treatment in *Dendrobium officinale*. *Front Plant Sci* 13:941231.<https://doi.org/10.3389/fpls.2022.941231>
- Jin Q, Jiao C, Sun S, Song C, Cai Y, Lin Y, Fan H, Zhu Y (2016) Metabolic Analysis of Medicinal *Dendrobium officinale* and *Dendrobium huoshanense* during Different Growth Years. *PLoS One* 11:e0146607.<https://doi.org/10.1371/journal.pone.0146607>
- Liu Y, Tang Q, Cheng P, Zhu M, Zhang H, Liu J, Zuo M, Huang C, Wu C, Sun Z, Liu Z (2020) Whole-genome sequencing and analysis of the Chinese herbal plant *Gelsemium elegans*. *Acta Pharm Sin B* 10:374-382.<https://doi.org/10.1016/j.apsb.2019.08.004>
- Luk LY, Bunn S, Liscombe DK, Facchini PJ, Tanner ME (2007) Mechanistic studies on norcoclaurine synthase of benzyloisoquinoline alkaloid biosynthesis: an enzymatic Pictet-Spengler reaction. *Biochemistry* 46:10153-10161.<https://doi.org/10.1021/bi700752n>
- Ma X, Koepke J, Fritsch G, Diem R, Kutchan TM, Michel H, Stöckigt J (2004) Crystallization and preliminary X-ray crystallographic analysis of strictosidine synthase from *Rauvolfia*: the first member of a novel enzyme family. *Biochim Biophys Acta* 1702:121-124.<https://doi.org/10.1016/j.bbapap.2004.06.013>
- Ma X, Panjikar S, Koepke J, Loris E, Stöckigt J (2006) The structure of *Rauvolfia serpentina* strictosidine synthase is a novel six-bladed beta-propeller fold in plant proteins. *Plant Cell* 18:907-920.<https://doi.org/10.1105/tpc.105.038018>
- McKnight TD, Bergey DR, Burnett RJ, Nessler CL (1991) Expression of enzymatically active and correctly targeted strictosidine synthase in transgenic tobacco plants. *Planta* 185:148-152.<https://doi.org/10.1007/bf00194055>
- Montiel G, Zarei A, Körbes AP, Memelink J (2011) The jasmonate-responsive element from the ORCA3 promoter from *Catharanthus roseus* is active in *Arabidopsis* and is controlled by the transcription factor AtMYC2. *Plant Cell Physiol*

52:578-587.<https://doi.org/10.1093/pcp/pcr016>

O'Connor SE, Maresh JJ (2006) Chemistry and biology of monoterpene indole alkaloid biosynthesis. *Nat Prod Rep* 23:532-547.<https://doi.org/10.1039/b512615k>

Pan H, Chen Y, Zhao J, Huang J, Shu N, Deng H, Song C (2022) In-depth analysis of large-scale screening of WRKY members based on genome-wide identification. *Front Genet* 13:1104968.<https://doi.org/10.3389/fgene.2022.1104968>

Pan X, Yang C, Cleveland JL, Bannister TD (2016) Synthesis and Cytotoxicity of Sempervirine and Analogues. *J Org Chem* 81:2194-2200.<https://doi.org/10.1021/acs.joc.6b00022>

Qiyang Z, Huimin J, Lala LI, Hongyu Y, Ruijiao Z, Yijia W (2021) Genome-wide Identification and Expression Analysis of Strictosidine Synthase Gene in Tea Plant. *Journal of Xinyang Normal University* (Natural Science Edition) 34:596-605.<https://doi.org/10.3969/j.issn.1003-0972.2021.04.015>

Sharma A, Verma P, Mathur A, Mathur AK (2018) Overexpression of tryptophan decarboxylase and strictosidine synthase enhanced terpenoid indole alkaloid pathway activity and antineoplastic vinblastine biosynthesis in *Catharanthus roseus*. *Protoplasma* 255:1281-1294.<https://doi.org/10.1007/s00709-018-1233-1>

Škubník J, Pavličková VS, Ruml T, Rimpelová S (2021) Vincristine in Combination Therapy of Cancer: Emerging Trends in Clinics. *Biology* (Basel) 10.<https://doi.org/10.3390/biology10090849>

Sohani MM, Schenk PM, Schultz CJ, Schmidt O (2009) Phylogenetic and transcriptional analysis of a strictosidine synthase-like gene family in *Arabidopsis thaliana* reveals involvement in plant defence responses. *Plant Biol* (Stuttg) 11:105-117.<https://doi.org/10.1111/j.1438-8677.2008.00139.x>

Sriram D, Yogeewari P, Thirumurugan R, Bal TR (2005) Camptothecin and its analogues: a review on their chemotherapeutic potential. *Nat Prod Res* 19:393-412.<https://doi.org/10.1080/14786410412331299005>

Stöckigt J, Barleben L, Panjekar S, Loris EA (2008) 3D-Structure and function of strictosidine synthase--the key enzyme of monoterpene indole alkaloid biosynthesis. *Plant Physiol Biochem* 46:340-355.<https://doi.org/10.1016/j.plaphy.2007.12.011>

Tang H, Zhao T, Sheng Y, Zheng T, Fu L, Zhang Y (2017) *Dendrobium officinale* Kimura et Migo: A Review on Its Ethnopharmacology, Phytochemistry, Pharmacology, and Industrialization. *Evid Based Complement Alternat Med* 2017:7436259.<https://doi.org/10.1155/2017/7436259>

Thamm AMK, Qu Y, De Luca V (2016) Discovery and metabolic engineering of iridoid/secoiridoid and monoterpene indole alkaloid biosynthesis. *Phytochemistry Reviews* 15:339-361.<https://doi.org/10.1007/s11101-016-9468-y>

Tisnerat C, Dassonville-Klimpt A, Gosselet F, Sonnet P (2022) Antimalarial Drug Discovery: From Quinine to the Most Recent Promising Clinical Drug Candidates. *Curr Med Chem* 29:3326-3365.<https://doi.org/10.2174/0929867328666210803152419>

Verma P, Sharma A, Khan SA, Shanker K, Mathur AK (2015) Over-expression of *Catharanthus roseus* tryptophan decarboxylase and strictosidine synthase in rol gene integrated transgenic cell suspensions of *Vinca minor*. *Protoplasma* 252:373-381.<https://doi.org/10.1007/s00709-014-0685-1>

Xu X, Zhang C, Wang N, Xu Y, Tang G, Xu L, Feng Y (2022) Bioactivities and Mechanism of

- 518 Actions of *Dendrobium officinale*: A Comprehensive Review. *Oxid Med Cell Longev*
519 2022;6293355.<https://doi.org/10.1155/2022/6293355>
- 520 Yamazaki Y, Urano A, Sudo H, Kitajima M, Takayama H, Yamazaki M, Aimi N, Saito K (2003)
521 Metabolite profiling of alkaloids and strictosidine synthase activity in camptothecin
522 producing plants. *Phytochemistry*
523 62:461-470.[https://doi.org/10.1016/s0031-9422\(02\)00543-5](https://doi.org/10.1016/s0031-9422(02)00543-5)
- 524 Zhang L, Ding R, Chai Y, Bonfill M, Moyano E, Oksman-Caldentey KM, Xu T, Pi Y, Wang Z,
525 Zhang H, Kai G, Liao Z, Sun X, Tang K (2004) Engineering tropane biosynthetic
526 pathway in *Hyoscyamus niger* hairy root cultures. *Proc Natl Acad Sci U S A*
527 101:6786-6791.<https://doi.org/10.1073/pnas.0401391101>
- 528 Zhang S, Pan Z, Gao H, Huang L, Chen P, Huang Q, Pu X (2021) Bioinformatic and gene
529 expression analysis of STR candidate genes in *Camptotheca acuminata*. *Chinese Journal*
530 *of Applied and Environmental Biology*
531 27:1022-1030.<https://doi.org/10.19675/j.cnki.1006-687x.2020.04033>
- 532 Zhang W, Yuan Q, Wu Y, Zhang J, Nie J (2022) Genome-Wide Identification and Characterization
533 of the CC-NBS-LRR Gene Family in Cucumber (*Cucumis sativus* L.). *Int J Mol Sci*
534 23.<https://doi.org/10.3390/ijms23095048>
- 535 Zhu H, Kerčmar P, Wu F, Rajendran C, Sun L, Wang M, Stöckigt J (2015) Using Strictosidine
536 Synthase to Prepare Novel Alkaloids. *Curr Med Chem*
- 537