GMR Genetics and Molecular Research
Online Journal
ISSN:1676-5680

Isolation and characterization of the anthocyanidin genes *pal*, *f3h* and *dfr* of *Scutellaria viscidula* (Lamiaceae)

W. Lei[1,2], R.X. Yao[3], X.H. Kang[4], S.H. Tang[1], A.M. Qiao[3] and M. Sun[1]

[1]Key Laboratory of Eco-Environments in Three Gorges Reservoir Region, Ministry of Education, School of Life Science, Southwest University, Chongqing, China
[2]College of Life Science, Sun Yat-Sen University, Guangzhou, China
[3]College of Horticulture and Landscape Architecture, Zhongkai University of Agriculture and Engineering, Guangzhou, China
[4]College of Life Science and Engineering, Southwest University of Science and Technology, Sichuan, China

Corresponding author: M. Sun
E-mail: jwcsm@swu.edu.cn

Genet. Mol. Res. 10 (4): 3385-3402 (2011)
Received April 5, 2011
Accepted August 31, 2011
Published November 22, 2011
DOI http://dx.doi.org/10.4238/2011.November.22.7

ABSTRACT. Anthocyanidin is a group of flavonoid compounds used as a vegetable pigment and plays an important role in flower coloration and environmental adaptations of the Chinese ornamental plant *Scutellaria viscidula*. We determined the cDNA sequences of phenylalanine ammonia-lyase (SvPAL), flavanone 3-hydroxylase (SvF3H) and dihydroflavonol 4-reductase (SvDFR) genes in *S. viscidula*. Comparative analysis showed that the protein products of these three genes did not have a transit peptide at their N-terminal portion, which indicated that these enzymes were directly involved in the substrate conversion in the cytoplasmic matrix. Bioinformatic analysis further revealed that *Svpal*, *Svf3h* and *Svdfr* were the members of flavonoid biosynthetic genes with highly conserved motifs. Based on phylogenetic tree analysis, it appears that PAL, F3H or DFR from

different plants might have originated from the same ancestor. This study can help to map and regulate the important stages involved in anthocyanidin biosynthesis by genetic engineering to diversify flower color and improve the ornamental value of *S. viscidula*.

Key words: *Scutellaria viscidula*; Phenylalanine ammonia-lyase; Flavanone 3-hydroxylase; Dihydroflavonol 4-reductase; Anthocyanidin

INTRODUCTION

Skullcap, belonging to the family *Labiate*, is a perennial herb and is located mainly in Asia. The plant has received increasing attention due to its significant amounts of various flavonoids. For example, the roots of skullcap contain flavone compounds such as baicalin, baicalein, wogonoside and wogonin, which are widely used to treat inflammatory and bacterial diseases in traditional Chinese medicine (Yu et al., 1984). Moreover, skullcap shows diverse colors including yellow, blue and purple in different breeds due to the presence of anthocyanidin compounds in the flower, and, thus, is widely used as ornamental plants as well. *Scutellaria viscidula* Bunge, one of the representative species of *Scutellaria* genus, is distributed in North China and increasingly employed in gardening practices (Tian et al., 2007).

Anthocyanidin exhibits some important physiological functions to meet environment challenges such as flower coloration, UV protection, and disease defense. Anthocyanidin is synthesized via the phenylpropanoid pathway, of which a majority of enzymes have previously been determined (Figure 1) (Uddin et al., 2004; Butelli et al., 2008). Furthermore, numerous structural genes and regulatory genes have been isolated and obtained (Jack et al., 2009; Zhou et al., 2010). Phenylalanine ammonia-lyase (PAL, EC 4.3.1.5) is the first key enzyme involved in this pathway and responsible for the conversion of L-phenylalanine to cinnamic acid, which links the primary and secondary metabolism (Li et al., 2001). As described for the first time in *Horden vulgare* in 1961, *pal* genes have been cloned and identified from many horticulture plants species such as *Petroselinum crispum*, *Nicotiana tabacum*, *Daucus carota*, *Bambusa oldhamii*, and *Pisum sativum*, etc. Flavanone 3-hydroxylase (F3H, EC 1.14.11.9) is the core enzyme regulating a metabolic flux shift into anthocyanidin pathway, with the purpose of converting (2S)-flavanones to (2R, 3R)-dihydroflavonols, the direct precursor of the flavonol biosynthesis (Holton and Comish, 1995). The first study on F3H resulted from *Matthiola incama*, and then numerous amounts of the enzyme genes were isolated and characterized from various horticulture plants including *Zea mays*, *Ginkgo biloba*, *Daucus carota*, *Persea americana* and *Perilla frutescenes*. Dihydroflavonol 4-reductase (DFR, EC 1.1.1.219) a crucial enzyme in anthocyanidin metabolic system drives to conversion from (2R, 3R)-dihydroflavonols including dihydrokaempferol (DHK), dihydroquercetin (DHQ) or dihydromyricetin (DHM) into leucoanthocyanidins leucodelphinidin or leucocyanidin by stereospecific reduction reaction on C4 site (Vincent et al., 1997). The initial DFR cloning was reported by O'Reilly in *Zea mays* (O'Reilly et al., 1985). A series of DFR genes were then determined from *Ipomoea horsfalliae*, *Lycopersicon esculentum*, *Crataegus monogyna*, *Brassica juncea*, *Vaccinium macrocarpon*, etc. Moreover, earlier experimental results also showed the important roles of PAL, F3H and DFR in the control of flavonoids accumulation via a responding committed stage in the plant. For example, expression level of *pal* gene was significantly positively correlated with fruit

anthocyanin in some horticultural plants (MacDonald and D'Cunha, 2007) while the inhibition of *f3h* gene expression by antisense suppression technique could modify flower color and fragrance (Zuker et al., 2002), whereas the *dfr* gene overexpression in the transgenic *Petunia hvbrida* led to generating a new flower color (Meyer et al., 1987).

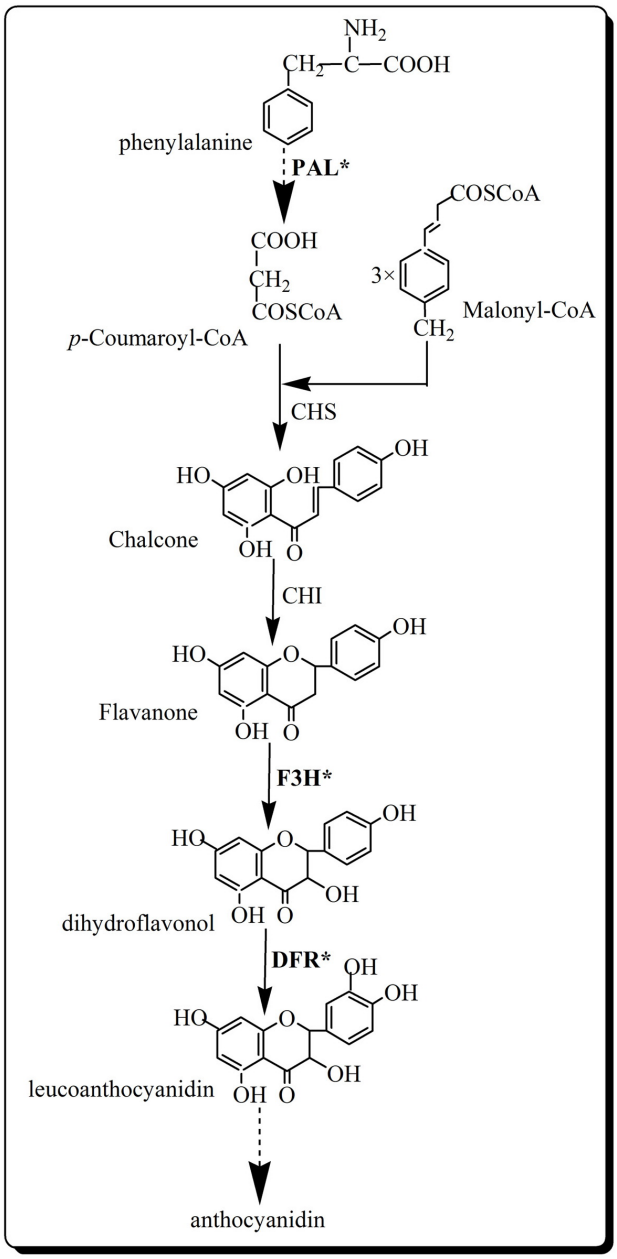

Figure 1. Flavonoid biosynthetic pathway in *Scutellaria* cells.

Anthocyanidin accumulation and regulation in *S. viscidula* require synergistic expression of related genes encoding enzyme system involved in the core flavonoid biosynthesis pathway, such as PAL, 4-coumarate-CoA:ligase (4CL), chalcone synthase (CHS), F3H and DFR. Thus, cloning and determination of the enzyme genes are essential and helpful to elucidate this metabolic course. Nevertheless, this is still not clear except for CHS so far (Lei et al., 2010). To date, the full length of *pal*, *f3h* and *dfr* genes have been cloned and characterized as well as their corresponding putative proteins analyzed using various bioinformatic tools and methods. To the best of our knowledge, this is the first reported study on the isolation and analysis of these three genes in *S. viscidula*.

MATERIAL AND METHODS

Plant material and RNA extraction

Scutellaria viscidula Bunge was cultured in the planting garden of Southwest University, Chongqing, China. Approximately 20 g *S. viscidula* tissue powdered in liquid nitrogen and the total cellular RNA was extracted with TRIzol™ reagent according to manufacturer instructions (Tiangen, China), then quantified by spectrophotometry and investigated integrality by agarose gel electrophoresis and finally the total RNA was stored in -70°C prior to use for cDNA manipulation.

Cloning of the full-length cDNA of the target genes PAL, F3H and DFR

After reverse transcription and subsequent RnaseH treatment, the first strand cDNA were synthesized from 5 μg total RNA with an Oligo dT-Adaptor Primer in RNA PCR Kit (AMV) Ver. 3.0 (TAKARA, Tokyo, Japan). This cDNA was used as templates for polymerase chain reaction (PCR). Deoxyribonucleotide primers pairs (dfpal and drpal, dff3h and drf3h, dfdfr and drdfr, Table 1) were designed and synthesized based on the conserved regions of nucleotide sequences of *pal*, *f3h* and *dfr* genes in some plant, respectively. The reverse transcription PCR (RT-PCR) reaction was implemented to obtain the core fragment based on the following program: denaturing the cDNA at 94°C for 3 min, following 34 cycles of amplification (94°C for 30 s, Tm for 30 s and 72°C for 1 min), and a final extension at 72°C for 10 min. After detection, purification and extraction, the PCR product of each target gene was cloned into pMD18-T vector (Promega, USA), respectively. After sequencing, the target fragment sequences were confirmed to be highly identical to other corresponding genes using the Blastn search. According to the core fragment, the gene-specific primers (GSP) were designed to elongate the upstream and downstream parts by rapid amplification of cDNA ends (RACE). The SMART™ RACE cDNA Amplification Kit (Clontech, USA) was used to synthesize two separate cDNA populations: 3'-RACE-ready and 5'-RACE-ready cDNA which were employed for the templates of 3'RACE and 5'RACE, respectively.

Aiming at each target gene, two 3'-GSP (svpal3-1 and svpal3-2, svf3h3-1 and svf3h3-2, svdfr3-1 and svdfr3-2; Table 1) and the universal primers (UPM and NUP provided by the kit, Table 1) the nested RACE PCR amplification of 3'-RACE. was performed For the first action, 3'-GSP-1 and UPM were used with 3'-RACE-ready cDNA as a template, while for the second one, 3'-GSP-2 and NUP were used with the templates from the previous PCR reac-

tive products. In a similar manner, two 5'-GSP (svpal5-1 and svpal5-2, svf3h5-1 and svf3h5-2, svdfr5-1 and svdfr5-2; Table 1) and the universal primers (UPM and NUP provided by the kit, Table 1) were performed. The nested PCR procedures were accomplished according to the protocol (Advantage™ 2 PCR Kit, Clontech, USA): 1 min at 95°C followed by 30 cycles (30 s at 94°C and 1 min at 58°C), and finally 2 min at 72°C for an extension. By 3'-RACE and 5'-RACE, both ends of each target gene were respectively sequenced and then collected. By aligning and assembling the sequences of the core fragment, 3'-RACE and 5'-RACE on ContigExpress (Vector NTI Suite 8.0), the full-length cDNA sequences of the three target genes were spliced and deduced *in silicon*. The physical full-length cDNA was subsequently amplified via RT-PCR reaction using a pair of primers (fsvpal and rsvpal, fsvf3h and rsvf3h, fsvdfr and rsvdfr;Table 1). The PCR procedure was conducted under the following conditions: 3 min at 94°C, 34 cycles (30 s at 94°C, 30 s at 55°C, and 2 min at 72°C) and 10 min at 72°C. The amplification product was purified and subcloned into pMD18-T vector (Promega, USA), later introduced into competent *Escherichia coli* strain DH5α cells, followed by sequencing. In this experiment, the independent PCR amplifications and sequencing of each target gene were repeated thrice to confirm the experimental results. The full-length cDNAs of *Svpal*, *Svf3h* and *Svdfr* were cloned successfully.

Table 1. Primers used in molecular cloning of the *Svpal*, *Svf3h*, *Svdfr* gene.

Primers	Sequence
dfpal	5'-AAGGC(G/T)AG(C/T)AGTGA(C/T)TGGGT(C/G)A-3'
drpal	5'-AACAA(G/A)GC(C/T)TT(G/A)CA(C/T)GGTGG(T/A)AA-3'
svpal3-1	5'-AAGTTATGAATGGGAAGCCTGAG-3'
svpal3-2	5'-CGAGAGGGAGATCAACTCGGTG-3'
svpal5-1	5'-CACGGAGGGGCAAGCATGGAGT-3'
svpal5-2	5'-GGTGACGCCATAGCTGTCAGTCC-3'
fsvpal	5'-ACGCGGGGAGGATCCCCTCTCT-3'
rsvpal	5'-ACGAAATAGAAGTGACTGAGAT-3'
dff3h	5'-CA(G/A)GT(G/A/T)GTT(G/A)ATCA(T/C)GG(A/G)GT(T/C)GA-3'
drf3h	5'-TC(A/C)GCATTCTTGAACCT(T/C)CCATT-3'
svf3h3-1	5'-GATCACTTTGCTGCTACAAGAC-3'
svf3h3-2	5'-TGGATCACGGTTCAACCCGTCC-3'
svf3h5-1	5'-CTCCAGTAGTTGGCAGGCCAGG-3'
svf3h5-2	5'-CTCGAGTAGTCACGGGCCTCG-3'
fsvf3h	5'-ACGCGGGGACCAGAAAAAAAAA-3'
rsvf3h	5'-GATCTTCATTGGGACTATCATT-3'
dfdfr	5'-GA(T/C)CC(T/C)GAGAATGA(G/A)GT(G/A)AT(T/A/C)AA-3'
drdfr	5'-TC(C/T)A(G/A/C)ATG(T/C/G)ACA(T/A)A(C/T)TG(C/A/T)CCTTG-3'
svdfr3-1	5'-AATCTTGGCTGAGAAAGCAGCA-3'
svdfr3-2	5'-CCCACCTAGCTTGATCACTGCA-3'
svdfr5-1	5'-TCCAGGTCACTCCAACTGGTTTC-3'
svdfr5-2	5'-CATCCCCTCAACCGTTGGCTTG-3'
fsvdfr	5'-ACGCGGGGACAATCCTTGGCAT-3'
rsvdfr	5'-AATCCCAACGCAGCACTTACAT-3'
UPM	Long: 5'-CTAATACGACTCACTATAGGGCAAGCAGTGGTATCAACGCAGAGT-3' Short: 5'-CTAATACGACTCACTATAGGGC-3'
NUP	5'-AAGCAGTGGTATCAACGCAGAGT-3'

dfpal = deoxyribonucleotide forward pal, dff3h = deoxyribonucleotide forward f3h, dfdfr = deoxyribonucleotide forward dfr; drpal = deoxyribonucleotide reverse pal, drf3h = deoxyribonucleotide reverse f3h, drdfr = deoxyribonucleotide reverse dfr; svpal = *Scutellaria viscidula* pal, svf3h = *Scutellaria viscidula* f3h, svdfr = *Scutellaria viscidula* dfr, 3-1, 3-2, 5-1 and 5-2 (3 and 5 refer to 3'end and 5'end, respectively; 1 and 2 refer to the first and the second, respectively), fsvpal = forward svpal, fsvf3h = forward svf3h, fsvdfr = forward svdfr, rsvpal = forward svpal, fsvf3h = forward svf3h, fsvdfr = forward svdfr, NUP = nested universal primer, UPM = universal primer mix.

Comparative and bioinformatic analysis

The nucleic acid sequences of the target genes (*Svpal*, GenBank® accession No. FJ432698; *Svf3h,* GenBank® accession No. FJ432699; and *Svdfr*, GenBank accession No. FJ605512) and corresponding amino acid sequences of the proteins (SvPAL, SvF3H and SvDFR) were calculated and analyzed with bioinformatics tools. Comparative bioinformatic analysis was done online at the NCBI (http://www.ncbi.nlm.nih.gov) and Expasy websites (http://www.expasy.org). The open reading frame (ORF) was predicted by ORF Finder. Multiple sequence alignments of the amino acid sequences of the target gene and the corresponding ones from other plant species were made with Vector NTI 8.0 using default parameters (Lei et al., 2009). Subcellular location was predicted with the TargetP 1.1 server (Ehrbar et al., 2004). Cellular function, transmembrane helices and hydrophobicity of the protein were predicted with the ProtFun 2.2 server (Jensen et al., 2002, 2003), TMHMM server v.2.0 (Ikeda et al. 2002) and ProtScale (Kyte and Doolittle 1982), orderly. Protein motifs were identified with PrositeScan (Combet et al. 2000). The target protein and the corresponding ones from other plants were aligned with ClustalX (Thompson et al., 1997), subsequently a phylogenetic tree was constructed by the Neighbor-Joining method with 1000 replicates, and the reliability of each node was established based on bootstrap calculations using the MEGA4.1 software (Kumar et al., 2001). Homology-based three-dimensional (3D) structural modeling of the target protein was done with Swiss-Modeling (Arnold et al., 2006) while WebLab ViewerLite 4.2 was used to display the 3D structure.

RESULTS AND DISCUSSION

Cloning of the full-length cDNA of the target genes PAL, F3H and DFR

The full-length cDNA sequence of SvPAL, SvF3H and SvDFR was cloned successfully and submitted into GenBank® database by the RT-PCR in combination with RACE technique. Firstly, the nucleotide acid sequence of *Svpal* was 2406 bp long, containing 2133 bp ORF encoding 711 amino acids, a 5'-untranslated region (UTR) of 128 bp long upstream from the start codon ATG, and a 142 bp 3'-UTR with a putative polyadenylation signal AATAA at a position 11 bp downstream from the stop codon TAG, as shown in Figure 2. Secondly, the full-length of *Svf3h* sequence was 1317 bp, in which there was 1050 bp ORF that encoded a 350 amino acid protein flanked by a 73 bp 5'-UTR and 191 bp 3'-UTR incidental 11 bp poly(A) tail (Figure 3). Thirdly, *Svdfr* gene consisted of 1283 bp nucleotides in total including 777 bp ORF encoding a 259 amino acid protein, 89 bp 5'-UTR and 414 bp 3'-UTR with the stop codon TAA (Figure 4). Finally, the target genes, which have been identified, were submitted to the GenBank® Database (*Svpal*, GenBank® accession No. FJ432698; *Svf3h*, GenBank accession No. FJ432699; and *Svdfr*, GenBank® accession No. FJ605512).

Molecular structures and physicochemical properties

Structural properties of *Svpal*, *Svf3h* and *Svdfr* and their encoding proteins are shown in Table 1. As they bring into play in different phases and under different reaction conditions, it is clear that there is a significant difference from each other on the nucleic acid and protein

structures. However, a slight similarity was observed among the physicochemical properties of SvPAL, SvF3H and SvDFR, indicating their significant functional association with genetic conservation of anthocyanidin biosynthesis pathway.

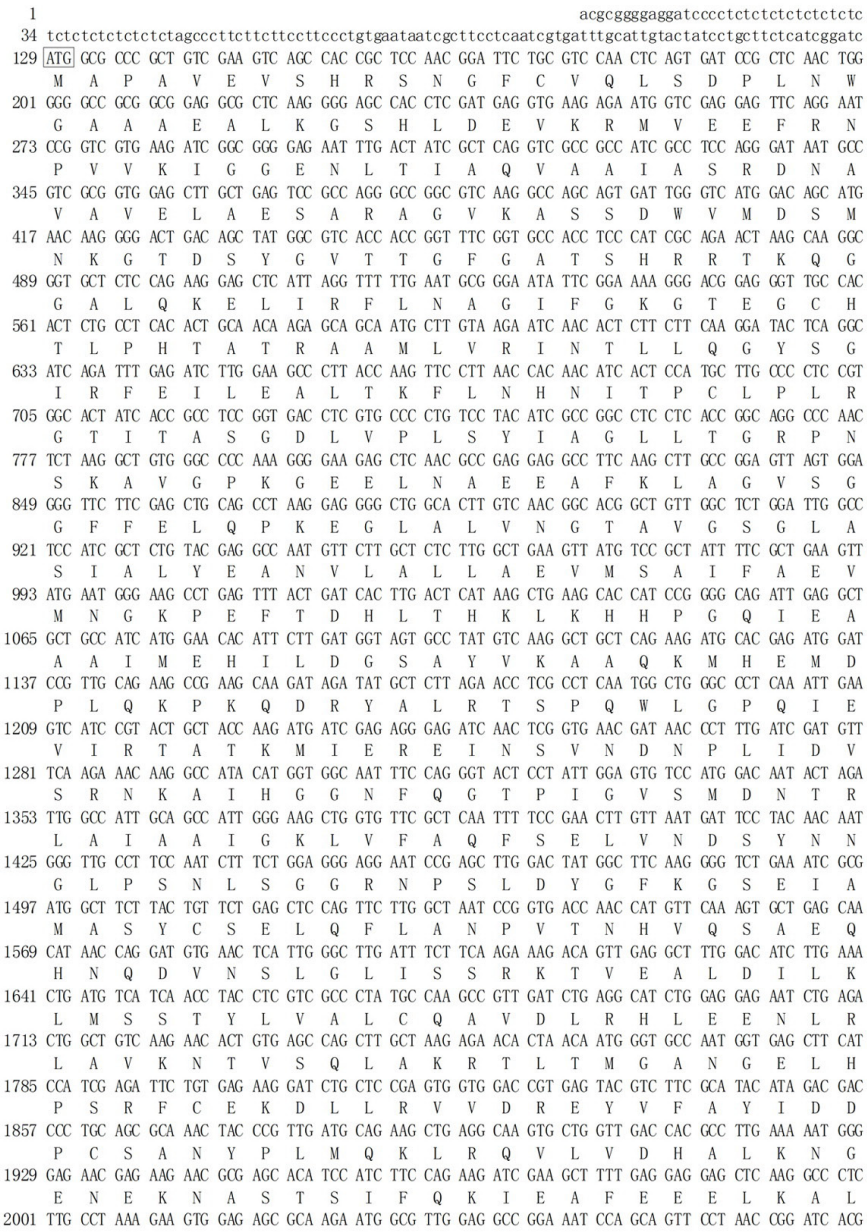

Figure 2. Full-length cDNA sequence and deduced amino acid sequence of SvPAL. The start codon (ATG) was boxed and the stop codon (TGA) was marked with asterisks, respectively. The coding sequence of *Svpal* is shown in capital letters and the 5'- and 3'-untranslated regions are shown in normal letters.

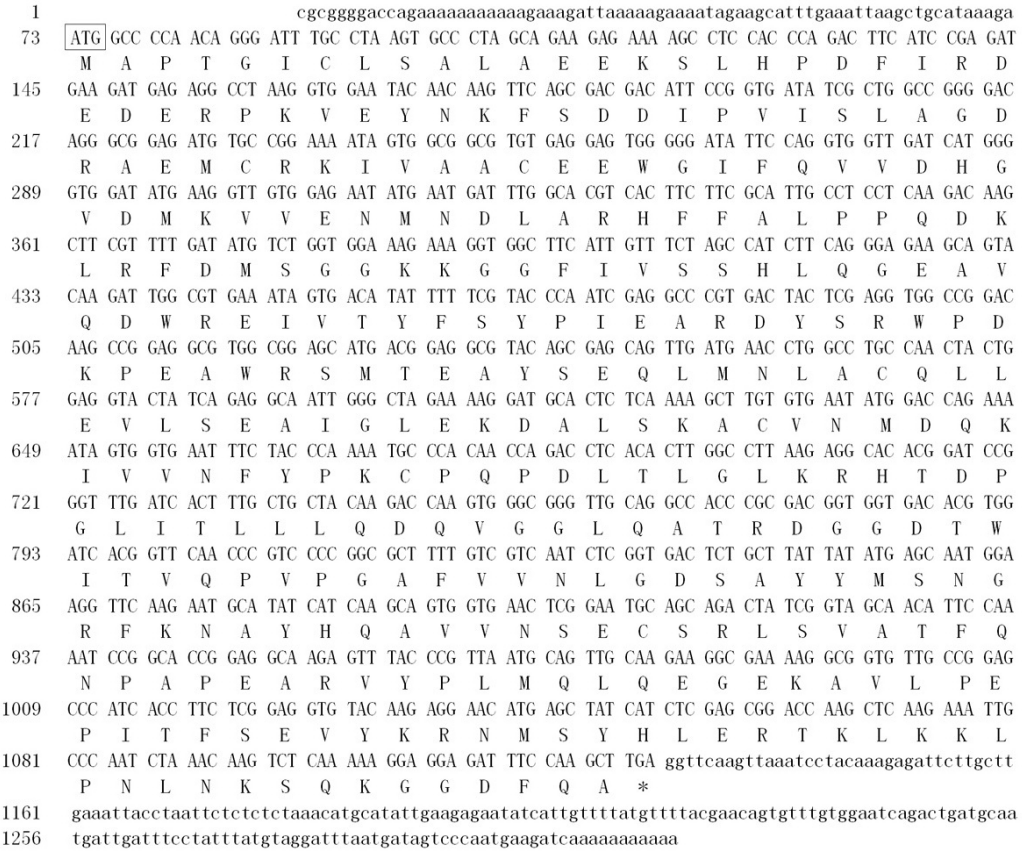

Figure 3. Full-length cDNA sequence and deduced amino acid sequence of SvF3H. The start codon (ATG) is boxed and the stop codon (TGA) is marked with asterisks, respectively. The coding sequence of *Svf3h* is show in capital letters and the 5'- and 3'-untranslated regions are shown in normal letters.

Cytologic characterization and function domain investigation

The subcellular localization and transmembrane helix prediction suggested that all of SvPAL, SvF3H and SvDFR were localized in the cytoplasmic matrix without transit peptide and transmembrane topological structure, indicating that these enzymes directly drove the substrate conversion and the skullcap anthocyanidin biosynthesis, which exactly corresponded with the fact that flavonoids were synthesized in the cytoplasmic matrix (Hrazdina, 1992). In fact, it was previously revealed that PAL protein scattered in the cell matrix which was confirmed by electron microscopy observation (Jin et al., 1997).

After aligning SvPAL, SvF3H or SvDFR and the target amino acid sequences from other plant species, it was shown that the structural difference in the N-terminal domain exhibited the difference in their activity levels. Some highly conserved residues were found in the SvPAL sequence: S^{213}, Y^{120}, L^{148}, N^{269}, Q^{357}, Y^{360}, R^{363}, F^{409}, Q^{497} and 3, 5-dihydro-5-methylidene-4H-imidazol-4-one (MIO), which were considered to be of significant importance for

```
1                                                                acgcggggacaatccttg
19   gcattggttcaaacttgtcttaaaatttaatatatagttccaatccaacccaaatatctcacttcctaaac
90   ATG CCT TTG GTA ACC ACC ATG GCC ATG CCA CCT CCA GCC ACG ACA GTA TGT GTC
      M   P   L   V   T   T   M   A   M   P   P   P   A   T   T   V   C   V
144  ACC GGA GCA TCT GGC TTC ATC GGC TCA TGG CTC GTC ATG AGA CTC CTC CAG CAT
      T   G   A   S   G   F   I   G   S   W   L   V   M   R   L   L   Q   H
198  GGT TAT ATT GTT CGT GCA ACT GTT CGT GAT CCT GGG AAC ATG GAG AAG GTG AAA
      G   Y   I   V   R   A   T   V   R   D   P   G   N   M   E   K   V   K
252  CAC CTA ACG GAA CTG CCA CAA GCA GAC ACA AAG TTG ACA CTG TGG AAG GCA GAT
      H   L   T   E   L   P   Q   A   D   T   K   L   T   L   W   K   A   D
306  ATG AGC ATA CAA GGA AGC TAC GAC AAA GCA GTA CAA GGT TGT GAA GGG GTG TTT
      M   S   I   Q   G   S   Y   D   K   A   V   Q   G   C   E   G   V   F
460  CAC ATG GCC ACA CCT ATG GAT TTC GAA TCC AAT GAT CCT GAG AAT GAA GTG ATC
      H   M   A   T   P   M   D   F   E   S   N   D   P   E   N   E   V   I
514  AAG CCA ACG GTT GAG GGG ATG TTG AAC ATC ATA AGA TCA TGT GCC AAA GCC AAA
      K   P   T   V   E   G   M   L   N   I   I   R   S   C   A   K   A   K
568  ACT GTG AAG AAA TTG ATA TTC ACA ACC TCA GCA GGG ACA CTG AAT GTT GAA GAA
      T   V   K   K   L   I   F   T   T   S   A   G   T   L   N   V   E   E
622  CAC CAG AAA CCA GTT TAT GAT GAA ACC AGT TGG AGT GAC CTG GAT TTC ATT TAC
      H   Q   K   P   V   Y   D   E   T   S   W   S   D   L   D   F   I   Y
676  TCC AAG AAA ATG ACT GGA TGG ATA TAT TTT TTG TCT AAA ATC TTG GCT GAG AAA
      S   K   K   M   T   G   W   I   Y   F   L   S   K   I   L   A   E   K
730  GCA GCA ATG GAA GCA ACT AAA GAG AAT AAT ATC AAC TTA ATT ACT GTC ATA CCA
      A   A   M   E   A   T   K   E   N   N   I   N   L   I   T   V   I   P
784  CCT CTA GTG GTT GGT CCA TTC ATC ATG CCC ACC CTC CCA CCT AGC TTG ATC ACT
      P   L   V   V   G   P   F   I   M   P   T   L   P   P   S   L   I   T
838  GCA CTT TCC CCC ATT ACT GGG AAT GAA GCC CAC TAT TTA CTA TCA GGA TTC AGT
      A   L   S   P   I   T   G   N   E   A   H   Y   L   L   S   G   F   S
892  GAC AGT CAC TCT CAA GGG ACA AGA ACT GGA GCT GAA AAA GAT TCT TAC AAT CTA
      D   S   H   S   Q   G   T   R   T   G   A   E   K   D   S   Y   N   L
946  CAC CTC GAT CGA CTT CTC AGA TAA ccatctcgttggaacaattcctcaaacagttggtgagct
      H   L   D   R   L   L   R   *
1009 gaaatccctgtattttctcaacctctcccacaatgctctctctggccacatcccatcatcgataggaaatc
1080 tacagaagctcgagtcattggatctctcactcaacagtctcgacggagaaatcccgaggcagcttgcaagc
1151 ctcacgttcctctcattcttgaacgtgtcttacaaccatcttgttgggaggatcccagaagggagccaaat
1222 ccagacatttccagagagttccttcatcggaaacgaggggctttgtgggtttcctttgaacaaaacctgca
1293 atggcactggaacatcatcatcatcgtcatcgccggagtttccacgggggaagttggagggagagatatat
1364 gtaagtgctgcgttgggatt
```

Figure 4. Full-length cDNA sequence and deduced amino acid sequence of SvDFR. The start codon (ATG) is boxed and the stop codon (TGA) is marked with asterisks, respectively. The coding sequence of *Svdfr* is shown in capital letters and the 5'- and 3'-untranslated regions are shown in normal letters.

substrate binding, catalytic activity and LMIO formation (Xu et al., 2008). The mechanism was in agreement with the recent report from *Petroselinum crispum* research (Röther et al., 2002). In particular, SvPAL contained a conserved region named Ala-Ser-Gly triad, which can be converted autocatalytically, via internal cyclization and elimination of two water molecules, to the electrophilic prosthetic group MIO (Song and Wang, 2009) (Figure 5). As for

SvF3H protein, three ferrous-binding domains (His[225], Asp[227] and His[282]), two 2-oxoglutarate binding regions (RXS motif: Arg[292], Ser[294]), and five similar motifs for 2-ODD were scanned including motif 1 (66~81), motif 2 (149~151), motif 3 (215~219), motif 4 (220~239), motif 5 (278~294), and five similar motifs for 2-ODD, among which the motif 1, motif 4 and motif 5 had several different residues indicating they were relative to species specificity. Furthermore, strictly conserved motif 2 and motif 3 were reported to play important roles in the folding process of the polypeptide (Britsch et al., 1992, 1993; Lukacin and Britsch, 1997) (Figure 6). In addition, the SvDFR amino acid sequence exhibited a putative NADP binding site (27-47, VTGASGFIGSWIVMRLLQHGY) with a very high similarity as in the description of other DFRs. Furthermore, Ser[145] residue was considered as the catalytic center, while Asn[150] and Glu[161] were associated with substrate specificity (James et al., 2002) (Figure 7). These relatively conserved domains exhibited such a high conservation to remain basically identical in almost all the plants species, especially *S. viscidula*, so that these domains were possibly closed related to the catalytic function, which suggested the functional existence and stabilization of SvPAL, SvF3H and SvDFR proteins.

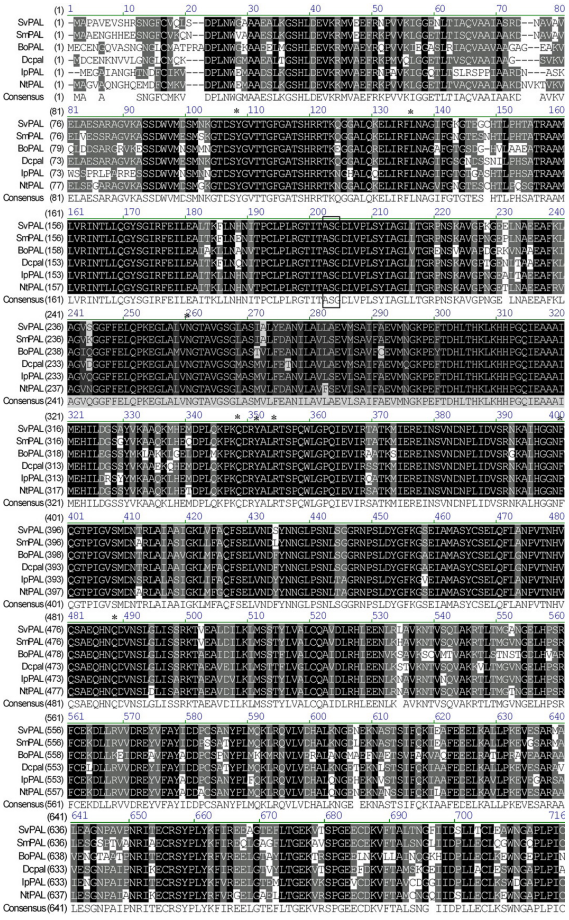

Figure 5. Conserved domains of the target protein SvPAL, SvF3H and SvDFR.

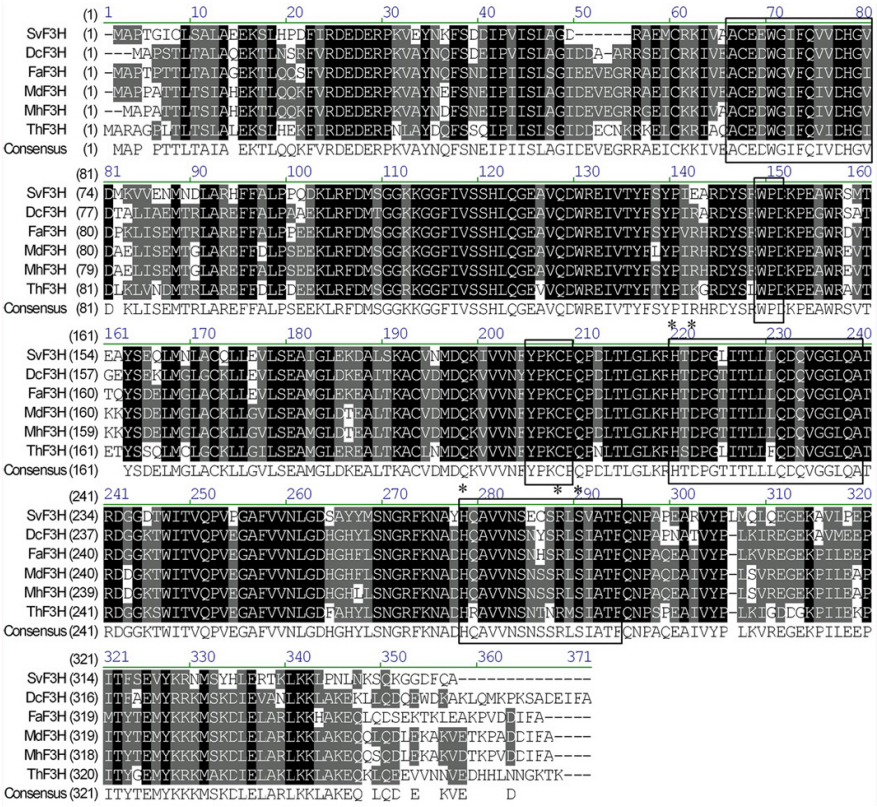

Figure 6. Multiple alignment of amino acid sequences of SvPAL and other plant PALs. The identical sites are shown in white letters and black background; the conservative sites are shown in white letters and gray background; other sites are all shown in black letters and white background. The active sites residues are indicated in asterisk (*), and residues Ala-Ser-Gly forming MIO group are boxed in the alignment. The Database accession numbers of the sequences used for the multi-alignments: SmPAL (*Salvia miltiorrhiza*, GenPept accession No. ABR14606); BoPAL (*Bambusa oldhamii*, GenPept accession No. ACN62413); DcPAL (*Daucus carota*, GenPept accession No. BAA23367); IpPAL (*Ipomoea batatas*, GenPept accession No. BAA11459); NtPAL (*Nicotiana tabacum*, GenPept accession No. BAA22947), SvPAL (*Scutellaria viscidula*, GenPept accession No. FJ432698).

Some important motif patterns were obtained by scanning the amino acid sequences of SvPAL, SvF3H and SvDFR against PROSITE. Except four motifs in all the three enzymes, i.e. N-glycosylation site, Protein kinase C phosphorylation site, Casein kinase II phosphorylation site, and N-myristoylation site, one mutual motif existed in both SvF3H and SvDFR, i.e. cAMP- and cGMP-dependent protein kinase phosphorylation site. In fact, there was only one specific motif in every target protein, that is, phenylalanine and histidine ammonia-lyases signature existing in SvPAL, amidation site in SvF3H. In summary, these signatures play the important role in recognizing and binding between enzyme and its own substrate. Then, the tool CDD recognized the presence of conserved domains in these proteins, as shown in Figure 8, and a series of functional domains were detected, such as catalytic residues, substrate binding pocket, lyase class I_like region, 2-oxoglutarate (2OG), Fe(II)-dependent oxygenase superfamily and putative dehydrogenase domain.

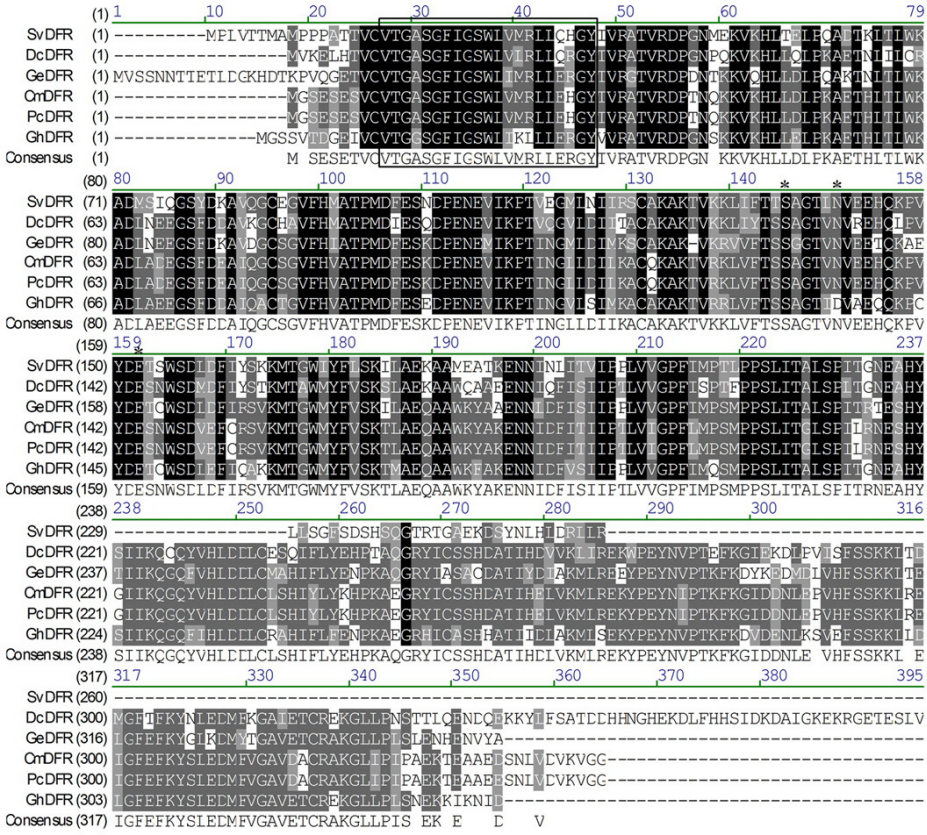

Figure 7. Multiple alignments of amino acid sequences of SvF3H and other plant F3Hs. The identical sites are shown in white letters and black background; the conservative sites are shown in white letters and gray background; other sites are all shown in black letters and white background. The conserved amino acid residues for ligating ferrous iron and participating in 2-oxoglutarate binding (RXS) are indicated with an asterisk (*). Five motifs and two regions highly conserved in flavonoid-specific dioxygenase polypeptides are labeled with box in the alignment. The Database accession numbers of the sequences used for the multi-alignments: MhF3H (*Malus hybrid*, GenPept accession No. ACP30361); FaF3H (*Fragaria x ananassa*, GenPept accession No. BAE17126); ThF3H (*Torenia hybrid*, GenPept accession No. BAD95810); MdF3H (*Malus x domestica*, GenPept accession No. BAB92997); DcF3H (*Daucus carota*, GenPept accession No. AAD56577), SvF3H (*Scutellaria viscidula*, GenPept accession No. FJ432699).

Molecular evolution and 3-D modeling

S. viscidula is a beautiful and valuable horticultural plant, and *pal*, *f3h* or *dfr* genes play a critical role in the anthocyanidin metabolic mechanism and genetic engineering. Investigation and elucidation of their genetic relationship and molecular evolution can help to deeply understand their roles. Based on the multiple alignments of protein sequences from twenty-seven plant species by ClustalX, a phylogenetic tree was constructed from various organisms including fungi, animals and plants with the Neighbor-Joining method (Figure 9). It is most intriguing that PALs, F3Hs or DFRs were derived from a collective ancestor and

then differentiated gradually into different groups. At the evolutionary position, PAL, F3H and DFR from *S. visscidula* all belonged to plant groups, and were closer to other multicolored plants but farther than unicolor ones, manifesting that these key-regulation genes possessed a somewhat different expression to meet the diverse requirements of anthocyanidin function.

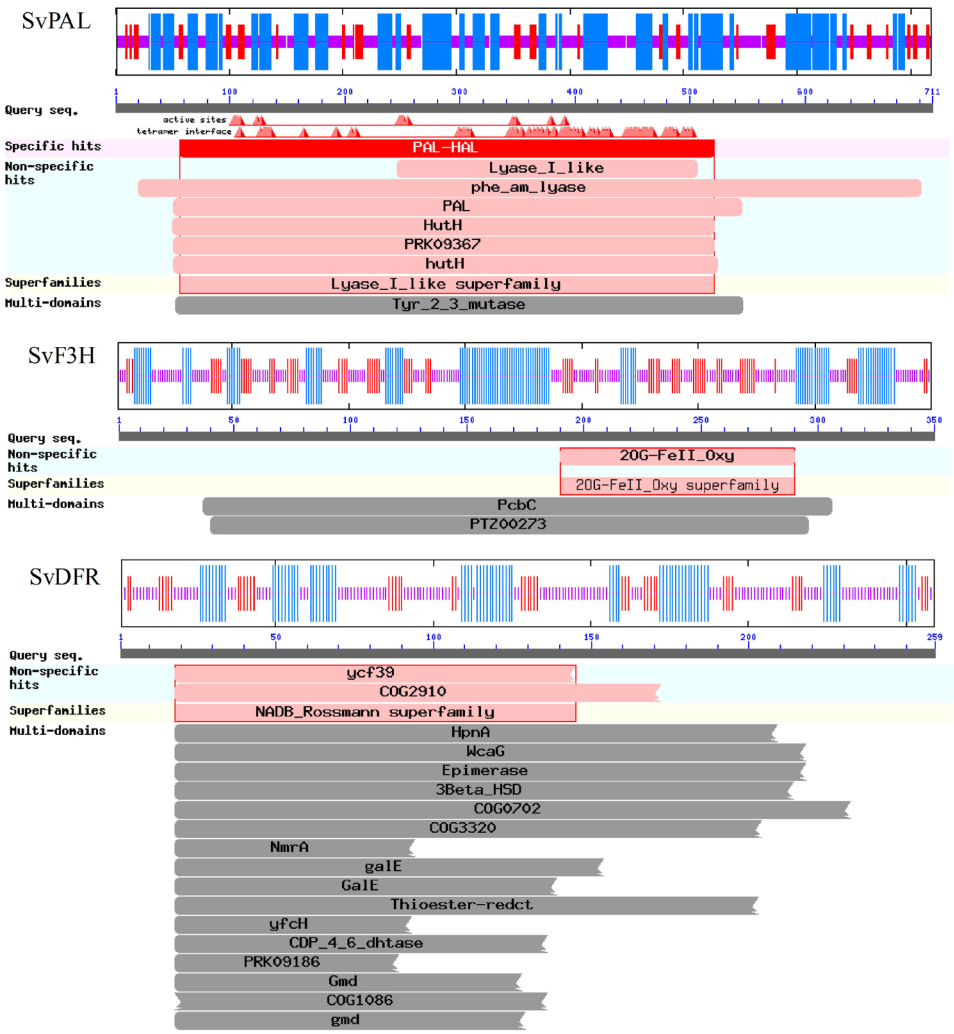

Figure 8. Multiple sequence alignment of SvDFR and other plant DFRs proteins. Identical sites are shown in white letters on a black background, conserved sites are shown in white letters on a gray background and other sites are shown in black letters on a white background. The highly conserved NADP binding site (VTGASGFIGSWIVMRLLQHGY) is boxed. The catalytic center Ser[145] and the substrate specificity related residues Asn[150] and Glu[161] are indicated with an asterisk (*). The database accession numbers of the sequences used in the alignment are: PcDFR (*Pyrus communis*, GenPept accession No. AAO39819), DcDFR (*Daucus carota*, GenPept accession No. AAD56578), GhDFR (*Gossypium hirsutum*, GenPept accession No. ACV72642), CmDFR (*Crataegus monogyna*, GenPept accession No. AAX16491), GeDFR (*Gypsophila elegans*, GenPept accession No. AAP13055), SvDFR (*Scutellaria viscidula*, GenPept accession No. FJ605512).

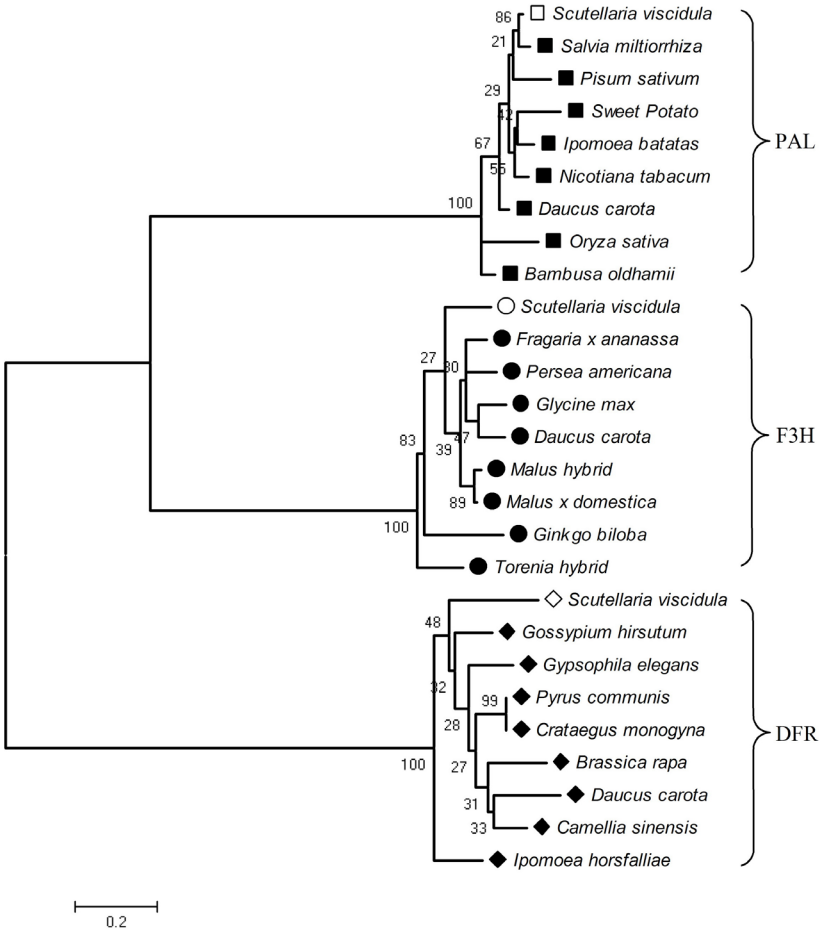

Figure 9. Phylogenetic analysis of plant PAL, F3H and DFR. The phylogenetic tree was constructed by the Neighbor-Joining method (based on 1000 bootstrap replicates) using MEGA4.1 software; the bootstrap values are shown on the branches. The target proteins SvPAL, SvF3H and SvDFR are indicated by ○, □ and ◊, respectively. The database accession numbers of the sequences used in the alignment are: *Sweet Potato* (SpPAL, GenPept accession No. AAA33389), *Oryza sativa* (OsPAL, GenPept accession No. AAO72666), *Daucus carota* (DcPAL, GenPept accession No. BAA23367), *Nicotiana tabacum* (NtPAL, GenPept accession No. BAA22947), *Ipomoea batatas* (IbPAL, GenPept accession No. BAA11459), *Pisum sativum* (PsPAL, GenPept accession No. BAA00885), *Bambusa oldhamii* (BoPAL, GenPept accession No. ACN62413), *Salvia miltiorrhiza* (SmPAL, GenPept accession No. ABR14606), *Malus hybrid* (MhF3H, GenPept accession No. ACP30361), *Ginkgo biloba* (GbF3H, GenPept accession No. AAU93347), *Glycine max* (GmF3H, GenPept accession No. ACN81825), *Fragaria x ananassa* (FaF3H, GenPept accession No. BAE17126), *Torenia hybrid* (ThF3H, GenPept accession No. BAD95810), *Malus x domestica* (MdF3H, GenPept accession No. BAB92997), *Daucus carota* (DcF3H, GenPept accession No. AAD56577), *Persea americana* (PaF3H, GenPept accession No. AAC97525), *Pyrus communis* (PcDFR, GenPept accession No. AAO39819), *Daucus carota* (DcDFR, GenPept accession No. AAD56578), *Gossypium hirsutum* (GhDFR, GenPept accession No. ACV72642), *Ipomoea horsfalliae* (IhDFR, GenPept accession No. ACS87954), *Gypsophila elegans* (GeDFR, GenPept accession No. AAP13055), *Camellia sinensis* (CsDFR, GenPept accession No. AAT84073), *Crataegus monogyna* (CmDFR, GenPept accession No. AAX16491), *Brassica rapa* (BrDFR, GenPept accession No. AAX53572) and *Scutellaria viscidula* (SvPAL, GenPept accession No. FJ432698), *Scutellaria viscidula* (SvF3H, GenPept accession No. FJ432699), *Scutellaria viscidula* (SvDFR, GenPept accession No. FJ605512).

Homology-based 3-D modeling of target proteins were implemented successfully using SWISS-MODEL (http://swissmodel.expasy.org) on the basis of their own corresponding templates: SvPAL based on phenylalanine ammonia-lyase from *Petroselinum crispu*, SvF3H on structure of flavanone 3-hydroxylase from *Arabidopsis thaliana*, and SvDFR on the structure of dihydroflavonol reductase from *Vitis viniferaat*. As shown in Table 2, random coil, α-helix and extended strand were composed of these protein secondary structures, and the former two were the main components. Some important functional motifs were scanned and marked in the tertiary model (Figures 10, 11 and 12).

Table 2. Analysis of molecular structures and physicochemical properties.

Index	PAL	F3H	DFR
Amino acid residues (aa)	711	350	259
Molecular weight (Da)	77185.1	39384.9	28588.0
Theoretical pI	5.97	5.40	6.17
Molar extinction coefficient	44975	45755	32555
Formula	C3405H5447N949O1038S28	C1751H2732N474O524S18	C1279H2029N331O379S15
Instability index	37.65 (stable)	47.80 (unstable)	36.77(stable)
Aliphatic index	91.84	80.51	87.72
GRAVY	-0.140	-0.378	-0.095
Secondary structure	α-helix: 43.88% β-strand: 12.10% Random coil: 44.02%	α-helix: 31.43% β-strand: 22.29% Random coil: 46.29%	α-helix: 28.96% β-strand: 17.37% Random coil: 53.67%

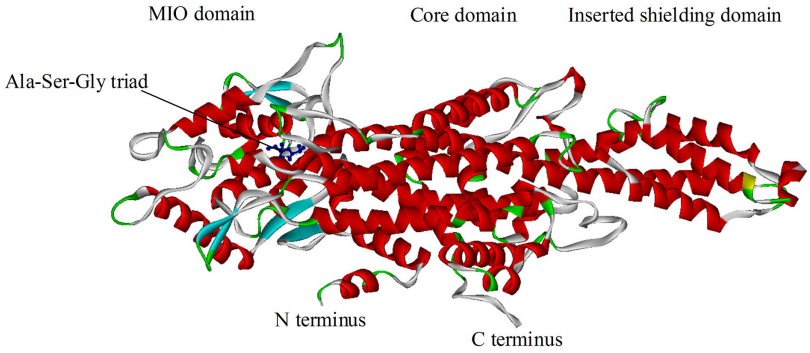

Figure 10. 3-D structural model of SvPAL. The α-helix and extended strand are indicated in red and blue, respectively. Random coils are indicated in silver. Selected important motifs were indicated.

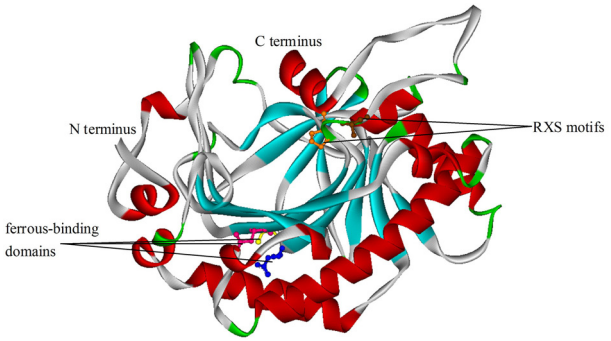

Figure 11. 3-D structural model of SvF3H. The α-helix and extended strand are indicated in red and blue, respectively. Random coils are indicated in silver. Selected important motifs are indicated.

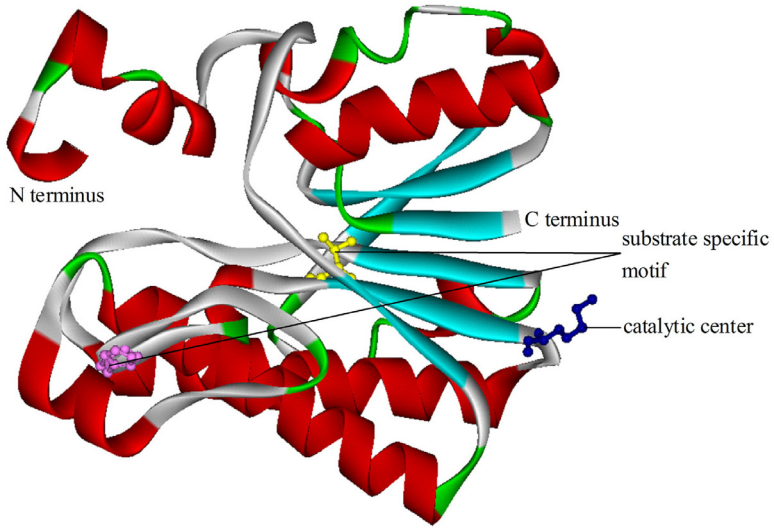

Figure 12. 3-D structural model of SvDFR. The α-helix and extended strand are indicated in red and blue, respectively. Random coils are indicated in silver. Selected important motifs are indicated.

In conclusion, anthocyanidin biosynthetic genes *pal*, *f3h* and *dfr* from the ornamental plant *S. viscidula* were cloned and characterized. Each target nucleotide acid sequence and its encoding protein was analyzed in detail using bioinformatic analysis. Previous studies revealed that PAL, F3H and DFR were the pivotal enzymes, regulating the metabolic flux and driving the anthocyanidin biosynthesis. In *S. viscidula*, anthocyanidin, mediated through SvPAL, SvF3H and SvDFR, was an important secondary metabolite with pharmaceutical and ornamental interest. The successful isolation of the *Svpal*, *Svf3h* and *Svdfr* gene, with preliminary identification of their structural properties, can help to map and regulate the important stages involved in anthocyanidin biosynthesis by genetic engineering so as to diversify flower color and improve the ornamental value of *S. viscidula*.

ACKNOWLEDGMENTS

The authors thank Prof. Yueming Jiang (South China Botanical Garden, Chinese Academy of Sciences, China) for serving as a pre-submission language editor. Research financially supported by the Opening Foundation of Key Laboratory of Eco-environments in Three Gorges Reservoir Region of Ministry of Education (#EF200609).

REFERENCES

Arnold K, Bordoli L, Kopp J and Schwede T (2006). The SWISS-MODEL Workspace: A web-based environment for protein structure homology modeling. *Bioinformatics* 22: 195-201.

Britsch L, Ruhnau-Brich B and Forkmann G (1992). Molecular cloning, sequence analysis, and *in vitro* expression of flavanone 3 beta-hydroxylase from *Petunia hybrida*. *J. Biol. Chem.* 267: 5380-5387.

Britsch L, Dedio J, Saedler H and Forkmann G (1993). Molecular characterization of flavanone 3 beta-hydroxylases. Consensus sequence, comparison with related enzymes and the role of conserved histidine residues. *Eur. J. Biochem.* 217: 745-754.

Butelli E, Titta L, Giorgio M, Mock HP, et al. (2008). Enrichment of tomato fruit with health-promoting anthocyanins by expression of select transcription factors. *Nat. Biotechnol.* 26: 1301-1308.

Combet C, Blanchet C, Geourjon C and Deleage G (2000). NPS@: network protein sequence analysis. *Trends Biochem. Sci.* 25: 147-150.

Ehrbar K, Hapfelmeier S, Stecher B and Hardt WD (2004). InvB is required for type III-dependent secretion of SopA in *Salmonella enterica* serovar Typhimurium. *J. Bacteriol.* 186: 1215-1219.

Holton TA and Cornish EC (1995). Genetics and biochemistry of anthocyanin biosynthesis. *Plant Cell* 7: 1071-1083.

Hrazdina G (1992). Compartmentation in Aromatic Metabolism, In: Recent Advances in Phytochemistry (Stafford HA and Ibrahim RK eds). Plenum Press, New York.

Ikeda M, Arai M, Lao DM and Shimizu T (2002). Transmembrane topology prediction methods: a re-assessment and improvement by a consensus method using a dataset of experimentally-characterized transmembrane topologies. *In Silico Biol.* 2: 19-33.

Jack S (1998). "Anthocyanin" (htm). Carnivorous Plant Newsletter (CPN) September 1998. Available at [http://www.carnivorousplants.org/cpn/samples/Science273anthocyanin.htm]. Accessed October 6, 2009.

Jensen JL, Gupta R, Blom N, Devos D, et al. (2002). *Ab initio* prediction of human orphan protein function from post-translational modifications and localization features. *J. Mol. Biol.* 319: 1257-1265.

Jensen LJ, Gupta R, Staerfeldt HH and Brunak S (2003). Prediction of human protein function according to Gene Ontology categories. *Bioinformatics* 19: 635-642.

Kumar S, Tamura K, Jakobsen IB and Nei M (2001). MEGA2: molecular evolutionary genetics analysis software. *Bioinformatics* 17: 1244-1245.

Kyte J and Doolittle RF (1982). A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.* 157: 105-132.

Lei W, Sun M, Luo KM, Shui XR, et al. (2009). Compute simulation to characterize structure and function of chalcone synthase from *Scutellaria baicalensis* Georgi. *Mol. Biol.* 43: 1082-1087.

Lei W, Tang SH, Luo KM and Sun M (2010). Molecular cloning and expression profiling of a chalcone synthase gene from hairy root cultures of *Scutellaria viscidula* Bunge. *Genet. Mol. Biol.* 33: 285-291.

Li YH, Sakiyama R, Maruyama H and Kawabata S (2001). Regulation of anthocyanin biosynthesis during fruit development in "Nyoho" strawberry. *J. Japan. Soc. Hort. Sci.* 70: 28-32.

Lukacin R and Britsch L (1997). Identification of strictly conserved histidine and arginine residues as part of the active site in *Petunia hybrida* flavanone 3-beta-hydroxylase. *Eur. J. Biochem.* 249: 748-757.

MacDonald MJ and D'Cunha GB (2007). A modern view of phenylalanine ammonia lyase. *Biochem. Cell Biol.* 85: 273-282.

Meyer P, Heidmann I, Forkmann G and Saedler H (1987). A new petunia flower colour generated by transformation of a mutant with a maize gene. *Nature* 330: 677-678.

Nakashima J, Awano T, Takabe K and Fujita M (1997). Immunocytochemical localization of phenylalanine ammonia-lyase and cinnamyl alcohol dehydrogenase in differentiating tracheary elements derived from *Zinnia mesophyll* cells. *Plant Cell Physiol.* 38: 113-123.

O'Reilly C, Shepherd NS, Pereira A, Schwarz-Sommer Z, et al. (1985). Molecular cloning of the a1 locus of *Zea mays* using the transposable elements *En* and *Mu1*. *EMBO J.* 4: 877-882.

Röther D, Poppe L, Morlock G, Viergutz S, et al. (2002). An active site homology model of phenylalanine ammonia-lyase from *Petroselinum crispum*. *Eur. J. Biochem.* 269: 3065-3075.

Song J and Wang Z (2009). Molecular cloning, expression and characterization of a phenylalanine ammonia-lyase gene (SmPAL1) from *Salvia miltiorrhiza*. *Mol. Biol. Rep.* 36: 939-952.

Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, et al. (1997). The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.* 25: 4876-4882.

Tian YL, Li JQ, Wang WH, Shi AP, et al. (2007). The germplasm resource and ornamental character of the Lamiaceae in Beijing. *J. Beijing Univ. Agric.* 22: 41-43.

Uddin JAFM, Hashimoto F, Miwa T, Ohbo K, et al. (2004). Seasonal variation in pigmentation and anthocyanidin phenetics in commercial Eustoma flowers. *Sci. Horticult.* 100: 103-115.

Vincent D, Werner H, Heinrich SJR and Augustin S (1997). Dihydroflavonol 4-reductase activity in lignocellulosic tissues. *Phytochemistry* 45: 1415-1418.

Xu F, Cai R, Cheng SY, Du HW, et al. (2008). Molecular cloning, characterization and expression of phenylalanine ammonia-lyase gene from *Ginkgo biloba*. *Afr. J. Biotechnol.* 7: 721-729.

Yu LR, Liu ML and Wang XT (1984). Studies on the flavonoids of *Scutellalia viscidula* Bunge. *Acta Pharm. Sin.* 19: 397-398.

Zhou W, Huang CT, Gong YF, Feng QL, et al. (2010). Molecular cloning and expression analysis of an ANS gene

encoding anthocyanidin synthase from purple-fleshed sweet potato [*Ipomoea batatas* (L.) Lam]. *Plant Mol. Biol. Rep*. 28: 112-121.

Zuker A, Tzfira T, Ben-Meir H, Ovadis M, et al. (2002). Modification of flower color and fragrance by antisense suppression of the flavanone 3-hydroxylase gene. *Mol. Breed.* 9: 33-41.