The Screening and Analysis of Differentially Expressed Genes from Rhizome of *Picrorhiza Scrophulariiflora*

FENG Yanli[1], *MA Xinrong*[2], *XIA Qingjie*[3], *TAN Xin*[2], *ZHANG Yizheng*[1]
(1. Sichuan Key Laboratory of Molecular Biology and Biotechnology, College of Life Sciences, Sichuan University, Chengdu 610064, China;
2. Chengdu Institute of Biology, Chinese Academy of Sciences, Chengdu 610041, China;
3. West China Hospital, Sichuan University, Chengdu 610044, China)

Abstract: *Picrorhiza scrophulariiflora* is a rare and endangered Tibetan medicinal plant. Its rhizome is a specific biosynthesis and accumulation tissue for invaluable hepatoprotective iridoids. We applied SSH technique for elucidating the differential gene expression in rhizome and leaf of the plant and constructed a subtracted rhizome-specific cDNA library. From the library, 27 non-redundant "EST-unigenes" were obtained and submitted to dbEST. Sequence analysis showed that 11 ESTs among the 27 ESTs are deducted to be novel genes (ESTs) with no significant homology to any known sequences within GenBank database and could potentially be regarded as novel gene fragments. These 27 ESTs are mainly associated with translation, energy metabolism, transport and binding, cell envelope, amino acid biosynthesis, and central intermediary metabolism. Finally, phylogenetic analyses implicated that a cytochrome monooxygenase P450-related gene (EX172715, 595 bp) identified from *P. scrophulariiflora* are more related to that from *A. thaliana* than that from *O. sativa* and *T. aestivum*. This study provides a novel gene pool involved in *P. scrophulariiflora* secondary metabolism.

Key words: *Picrorhiza scrophulariiflora*; rhizome-specific genes; suppression subtractive hybridization; ESTs; phylogenetic analysis; cytochrome P450 monooxygenase
CLC number: R96 **Document code**: A **Article ID**: 0529-6579 (2009) 02-0076-08

西藏胡黄连根状茎差异表达基因的筛选及分析

冯燕丽[1]，马欣荣[2]，夏庆杰[3]，谈 心[2]，张义正[1]
（1. 四川大学生命科学学院四川省分子生物学及生物技术重点实验室，四川 成都 610064；
2. 中国科学院成都生物研究所，四川 成都 610041；
3. 四川大学华西医学院，四川 成都 610044）

摘 要：西藏胡黄连是著名的濒危西藏药材，其粗壮的根茎部是重要的合成和储存药用萜类物质器官，萜类含量显著高于叶部。为克隆西藏胡黄连根状茎和叶的差异表达基因，利用抑制消减杂交技术，构建了根状茎特定的抑制消减文库。从该SSH文库中随机挑取片段不一的阳性克隆测序，采用Blastx进行同源比对，获得了27个表达序列标签（EST），并获登录号。序列统计分析结果表明，其中有11个EST在GenBank中无同源性序列，推测可能是新基因片断。蛋白质功能分类结果表明，这些蛋白质与翻译、能量代谢、转运与结合、细胞膜合成、氨基酸生物合成及中间代谢6大类蛋白功能有关。为深入研究与萜类生物合成直接或间接相关基因，选择了长595 bp的差异表达EST序列EX172715进行了同源比对和蛋白系统进化分析。结果表明，EX172715具有细胞色素P450单加氧酶的特征结构，与拟南芥的细胞色素P450单加氧酶蛋白的同源性高于水稻和小麦。提供了一组与西藏胡黄连次生代谢相关的新基因库。

关键词：西藏胡黄连；根状茎特异表达基因；抑制消减杂交；表达序列标签；同源性分析；细胞色素P450单加氧酶
中图分类号：R96

* 收稿日期：2008－06－11
基金项目：国家自然科学基金资助项目（30572337）
作者简介：冯燕丽（1968年生），女，高级工程师，博士研究生；通讯联系人：张义正；E-mail：yizzhang@scu.edu.cn

Picrorhiza scrophulariiflora is a well - known perennial alpine herbs growing on Qinghai -Tibetan Plateau. Its center of ecological distribution environment is restricted in the Eastern Himalayas to mountains of Yunnan (China) at an altitude of 4 300 ~ 5 200 meters[1], Uttar Pradesh to Southwest China, 3 600 ~ 4 800 meters [2-3]. *P. scrophulariiflora* is the invaluable traditional Tibetan medicinal plant that is broadly used in many compound formulations and known mainly for its important pharmacological iridoids [4]. Iridoids is one group of major active constituents in *P. scrophulariiflora* that attracts increasing attention as a valuable drug for chemical and medicinal research. Pharmacological studies and clinical experiments revealed that iridoids of the plant exhibit the activities of hepatoprotective, anticancer, antioxidant, immune-modulating, hypolipidemic, and cardioprotective [5-6].

Picrorhiza scrophulariiflora is one of near extinct species included in the China Plant Red Data List, where it is listed as a Category III species[7-8]. *P. scrophulariiflora* has also been classified as a Category III species under the Regulations of China on Protection of Medicinal Resources [9]. In spite of well documented information on morphology, biochemistry and medicinal application, *P. scrophulariiflora* has not so far been investigated using molecular analyses. So there is little knowledge about their pathways of secondary metabolite biosynthesis and the regulation mechanisms. Few genes involved in the process were cloned.

In recent years, researches about gene regulation of secondary metabolite biosynthesis become highlights. The recent researches have showed that *P. scrophulariiflora* accumulates a large amount of iridoids in rhizome but not in leaf. This presents that different genes expression in the two tissues.

Therefore, identification of the structural and regulatory factors operating distinctly in the rhizome/leaf of the plant will be a necessity for modulation of secondary metabolite biosynthesis. Suppression subtractive hybridization (SSH) is widely applied in detecting and isolating differentially expressed genes [10]. Using SSH, the high-throughput sequencing of cDNA clones (libraries) has produced extensive genomic databases and large numbers of expressed sequence tags (ESTs) for various plant species[11-14] and successfully in extensive EST analyses in Arabidopsis (*Arabidopsis thaliana* L. Heynh.) (The Arabidopsis Genome Initiative 2000), rice (*Oryza sativa* L.)[15], and tomato [16-17].

To analyze the genes related to iridoids synthesis, SSH between the rhizome and leaf and phylogenetic analysis were used in this study. Rhizome-specific expressed sequence tags (ESTs) were isolated from the SSH library for analyses and identification of genes that share sequence similarities with known iridoids synthesis-related proteins from other organisms in genomic or EST database.

1 Materials and methods

1.1 Materials

Picrorhiza scrophulariiflora was freshly taken at Dege County, Aba Autonomous Region, Sichuan, China in August.

1.2 Isolation of total RNA and mRNA

Total RNA was extracted from 3 - 5 g rhizomes and leaves, respectively, by using RNA Extraction Kit (Invitrogen) according to the manuals and then was treated with DNase I to eliminate the trace DNA. For cDNA synthesis for library construction and SSH, poly $(A)^+$ mRNA was isolated and purified from total RNA of the rhizome and leaf tissues of *P. scrophulariiflora* by using PolyATract mRNA Isolation System III Kit (Promega Company).

1.3 Suppression subtractive hybridization

Both kinds of mRNA were reverse transcripted into double-strand cDNAs with the SMART PCR cDNA Synthesis Kit (Clontech) to obtain sufficient amount of cDNAs for a subtraction experiment. cDNA SSH and selective amplification of cDNA fragments were performed referred to the manual of Clontech PCR-Select (cDNA Subtractive Kit [10] to identify rhizome-specific transcripts in *P. scrophulariiflora* in a strategically designed experiment where the rhizome and leaf cDNA populations served as tester and driver, respectively. All the steps were performed as the manufacturer's guidelines. Both cDNAs were digested with restriction enzyme *Rsa* I (BBI) for 12 h. The V-GENE PCR Kit (Invitrogen) was used for purification of cDNA fragments. Then the tester cDNA was divided into two aliquots and ligated with two different adaptors, respectively. The adaptor ligation time was 24 h. Each ligated tester sample was hybridized with excessive driver cDNA. Therefore, the hybridized tester samples were mixed together and endured a second round hybridization with excessive driver cDNA. After that, two rounds of PCR were performed to get the differentially expressed fragments amplified exponentially. To evaluate the efficiency of cDNA subtraction, glyceraldehyder-3-phosphate dehydrogenase gene (G_3PDH) was used as a control.

1.4 Constuction of subtracted rhizome-specific cDNA library

The secondary PCRs to amplify the subtracted rhizome-specific messages (LTRD = rhizome-tester, leaf-driver) were carried out for 33, 28, 23 and 18 cycles, respectively. The secondary PCR products from the forward subtraction cDNA samples were inserted into pMD18-T easy vector (TaKaRa, Dalian) and transformed the competent cells of *Escherichia coli* strain JM109 Positive clones were selected and confirmed by PCR.

1.5 Generation and analysis of expressed sequence tags (ESTs)

Ninety-six differently expressed cDNA in pMD18-T vector were randomly selected for PCR identification using nested primer 1 (5′-TCGAGCGGCCGCCCGGGCAGGT-3′) and nested primer 2R (5′-AGCGTGGTCGCGGCCGAGGT-3′), and the 35 positve clones with different length were sequenced by Shanghai Invitrogen Biotechnology Co., Ltd. The sequences obtained were screened with VecScreen (NCBI) and edited with Sequence Navigator™ v 1.0.1 software (Applied Biosystems) to remove vector, adaptor, and ambiguous sequences prior to BLAST analyses. And then, comparison of amino acid and nucleotide sequence similarity was performed by using BLAST program in NCBI (www.ncbi.nlm.nih.gov).

The edited sequences were subjected to a BLASTX analysis against the non-redundant protein database (all GenBank CDS translations + PDB + SwissProt + PIR + PRF excluding environmental samples from WGS projects). The analyses were used in conjunction to assign a putative function to an EST. It was possible to assign putative functions to many of the ESTs, although the true function can only be ascertained through biochemical and genetic approaches. The BLASTX result was given weightage for classification purposes. Wherever the BLAST homologies of individual ESTs were found to be somewhat similar, the 'bl2seq' tool (NCBI) was used to align the EST sequences to check for redundancy.

For functional classification of the ESTs by ExPASy Proteomics tools (www.expasy.ch), all selected ESTs were included, regardless of *E* value and redundancy. Many clones had the potential to be classified into more than one category due to overlapping functions. However, for simplicity, those ESTs were categorized according to their most universal function.

1.6 Sequence alignments and phylogenetic analyses

For multiple sequence alignments we used the software DNAMAN version 5.2.2 (Lynnon Corporation, Canada). And the software package Clustal W 1.83[18] was used for phylogenetic reconstruction. Phylogenetic trees were visualized with TREEVIEW 1.6.6 [19].

2 Results

2.1 ESTs from the subtracted rhizome-specific cDNA library

A rhizome-specific cDNA library containing over 2 000 recombinant clones was constructed after suppression subtractive hybridization. Ninety-six clones were randomly picked up, cultured in LB liquid media and then were amplified as templates when using nested adaptor primers. The sizes of inserts are ranged from 150 bp to 1 500 bp with the average length of 500 bp. Two identical nylon membranes were processed with PCR products dotted on. The membranes exhibited different expression patterns when hybridized with probes from two types of clones of rhizome and leaf.

A total of 35 randomly selected insert-containing clones from the subtracted rhizome-specific cDNA library were sequenced and assembled into 27 non-redundant "EST-unigenes", which provided 27 rhizome-specific ESTs.

2.2 Sequence and homology analysis of the 27 EST unigenes

The sequences of all the 27 ESTs were loaded into dbESTs and assigned GenBank accession numbers (Table 1).

By BLASTX (translation homology) search, 16 ESTs among these sequences were found to share high homology, above 67%, to diverse classes of genes with potato, tomato, rice, soybean and other EST sequences in GenBank as listed in Table 1. The other 11 fragments (ESTs) encoding hypothetical proteins with no significant BLASTX match to any other sequences in the public databases and could potentially be regarded as novel genes or 3′-terminal of full- length cDNA (Table 1).

2.3 Functional classification of the subtracted rhizome-specific ESTs

ExPASy Proteomics analysis revealed that the 27 subtracted rhizome-specific ESTs could be classified into 6 functional categories according to their putative function (Fig. 1). A large proportions (52%) of the ESTs are grouped in the translation category. There also presented a high number of the ESTs related to energy metabolism (30%), which indicated that the rhizome tissue, rather than the leaf tissue, is primarily responsible for secondary metabolites processing of the overall plant. This is in accordance with the rationale

Table 1　BLASTX results of 27 ESTs from rhizome-specific SSH library

Clone code	Size /bp	GenBank accession No.	Highest BlastX match and accession No.	E value	Protein %, Expect identities
FMX 2005007	617	EX172717	eukaryotic translation initiation factor 2 beta subunit-like [*Solanum tuberosum*] (ABB86256.1)	3e-68	126/131 (96%)
FMX 2005021	157	EX172731	calmodulin [*Quercus petraea*] (CAH57708.1)	5e-09	30/31 (96%)
FMX 20050004	935	EX172714	unknown [*Solanum tuberosum*] (ABB87113.1)	3e-112	206/223 (92%)
FMX 2005027	349	EX172737	ubiquitin family protein [*Arabidopsis thaliana*] (NP_199045.1)	6e-15	37/41 (90%)
FMX 2005002	533	EX172712	translocon-associated protein beta family protein-like [*Solanum tuberosum*] (ABB87132.1)	3e-35	72/91 (79%)
FMX 2005006	815	EX172716	coatomer alpha subunit-like protein [*Lotus japonicus*] (CAE45585.1)	8e-97	169/215 (78%)
FMX 2005010	607	EX172720	translocon-associated protein beta family protein-like [*Solanum tuberosum*] (ABB87132.1)	1e-34	71/90 (78%)
FMX 2005001	513	EX172711	translocon-associated protein beta (TRAPB) family protein [*Arabidopsis thaliana*] (NP_568293.1)	1e-25	61/80 (78%)
FMX 2005022	427	EX172732	protein binding / zinc ion binding [*Arabidopsis thaliana*] (NP_001031032.1)	6e-34	96/126 (76%)
FMX 2005014	297	EX172724	cannabidiolic acid synthase homolog [*Cannabis sativa*] (BAF65034.1)	2e-04	21/28 (75%)
FMX 2005023	392	EX172733	cannabidiolic acid synthase homolog [*Cannabis sativa*] (BAF65034.1)	0.009	18/24 (75%)
FMX 2005025	334	EX172735	type II SK2 dehydrin [*Prunus persica*] (AAZ83586.1)	1.1	15/20 (75%)
FMX 2005005	595	EX172715	cytochrome P450 monooxygenase CYP86A24 [*Glycine max*] (ABC68403.1)	4e-16	41/55 (74%)
FMX 2005015	745	EX172725	UBP7 (UBIQUITIN-SPECIFIC PROTEASE 7); ubiquitin-specific protease [*Arabidopsis thaliana*] (NP_566680.2)	2e-74	86/117 (73%)
FMX 2005016	262	EX172726	UBP7 (UBIQUITIN-SPECIFIC PROTEASE 7); ubiquitin-specific protease [*Arabidopsis thaliana*] (NP_566680.2)	2e-74	86/117 (73%)
FMX 2005011	522	EX172721	pirin [*Lycopersicon esculentum*] (AAF22236.1)	4e-43	77/114 (67%)
FMX 2005018	342	EX172728	Zinc-finger protein (ZIM) precursor [*Phillyrea latifolia*] (CAK18857.1)	3e-18	58/119 (48%)
FMX 2005008	1102	EX172718	hypothetical protein OsJ_014553 [*Oryza sativa* (japonica cultivar-group)] (EAZ31070.1)	2e-63	91/186 (48%)
FMX 2005013	233	EX172723	hypothetical protein [*Plasmodium falciparum* 3D7] (CAG25253.1)	0.51	21/56 (37%)
FMX 2005026	180	EX172736	hypothetical protein PFF0410w [*Plasmodium falciparum* 3D7] (XP_966073.1)	0.50	21/56 (37%)
FMX 2005024	183	EX172734	hypothetical protein LOC734927 [*Xenopus laevis*] (NP_001089861.1)	9.7	13/40 (32%)
FMX 2005019	179	EX172729	No significant similarity		–
FMX 2005009	446	EX172719	No significant similarity		–
FMX 2006003	295	EX172713	No significant similarity		–
FMX 2005012	279	EX172722	No significant similarity		–
FMX 2005017	94	EX172727	No significant similarity		–
FMX 2005020	96	EX172730	No significant similarity		–

that the rhizome tissue is primarily responsible for the production of secondary metabolites.

On the other hand, the 27 ESTs can also be grouped in gene ontology categories: closely related to growth-factor-related genes, such as EX172712, EX172713, EX172714, EX172715, EX172717, EX172718, EX172720, EX172721, EX172725, EX172733, EX172736; transporter, EX172711, EX172723, EX172724, EX172726; transcription, such as EX172727, EX172734, EX172735; stress response, such as EX172730; transcription regulation, such as EX172722, EX172728; signal transducer, EX172729, EX172732; hormone, EX172731; structural protein, EX172719 (putative oxidoreductase). Although 2 fragments EX172716, EX172737 have homology with ESTs GenBank, their functions have not been postulated because their protein functions are still unknown.

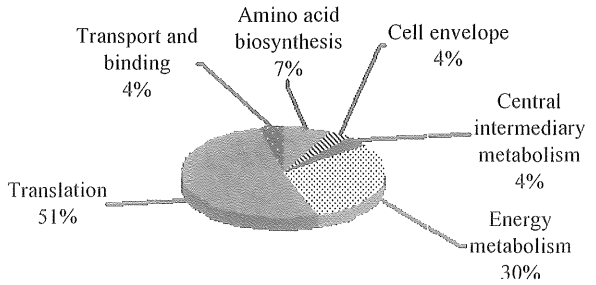

Fig. 1 Putative functions associated with the subtracted rhizome-specific ESTs in different functional categories

2.4 Phylogenetic analysis of targeted screening gene

One rhizome EST (EX172715) among the 27 ESTs showed (74%) homology to a cytochrome P450 monooxygenase gene of *Glycine max* from the nucleotide database. Cytochrome P450 is an important enzyme involved in the biosynthesis of many kinds of secondary metabolites in plant species. Using the deduced amino acid sequence of EX172715 to search by BLAST in NCBI, we obtained twelve more amino acid sequences from other organisms (Fig. 2A).

Here, we analyzed EX172715 that are potentially involved in processing of iridoids synthesis-related proteins based on sequence similarities with other 12 known cytochrome P450 proteins. The cDNA encodes a 89 amino acid protein. The deduced amino acid sequence of EX172715 shared 63.5%, 63.5%, 66.7%, 61.9%, 61.9%, 60.3%, 60.3 %, 58.7%, 57.1%, 50.8%, 61.9% and 68.3% identity with that of CYP86A7, a pupative P450, CYP86A24, CYP86A2, AT4g00360, CYP86A4, and CYP86A8 (cytochrome P450 subfamily) from *A. thaliana*, Os04g0560100, Os02g0666500, and Os01g0854800 (japonica cultivar-group) from *O. Sativa*, *Triticum aestivum* cytochrome P450 and *Vitis vinifera* hypothetical cytochrome P450, respectively (Fig. 2A). The result showed that the deduced amino acid sequence of EX172715 is more similar to that of cytochrome P450 from *A. thaliana*, *T. aestivum* and *V. vinifera* than that from *O. Sativa*.

By phylogenetic analysis, EX172715 is similar to other genes, with E-value < 0.001, including cytochrome P450 (CYP86A7, CYP86A24, CYP86A2, AT4g00360, CYP86A4, and CYP86A8) from *A. thaliana*, cytochrome P450 from *T. aestivum*, japonica cultivar-group (Os04g0560100, Os02g0666500, and Os01g0854800) from *O. sativa*. *V. vinifera* (hypothetical cytochrome P450) was used as outgroup (Fig. 2B).

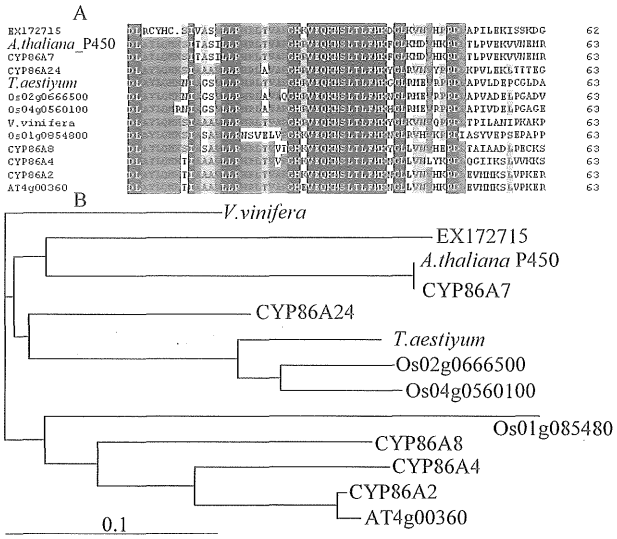

Fig. 2 Multiple alignment and phylogenetic analysis of the targeted EST (EX172715)

A: The deduced amino acid sequence of EX172715 from *P. Scrophulariiflora* was aligned with corresponding sequences of *T. aestivum* cytochrome P450, *A. thaliana* cytochrome P450 monooxygenase subfamily (a pupative P450 protein, CYP86A7, CYP86A24, CYP86A2, CYP86A4, CYP86A8, AT4g00360), *O. sativa* japonica cultivar-group (Os04g0560100, Os02g0666500, Os01g0854800) using DNAMAN multiple alignment programme. Black color indicates 100% consensus, dark and light gray color 75, 50% consensus, respectively. Gaps to optimize alignments are designated by dots. The amino acids are numbered on the right side of the sequence; B: A phylogram tree generated using the aligned protein sequences revealed that EX172715 shows highest relation to *A. thaliana* cytochrome P450 (CYP86A7). The scale bar represents the substitutions per site according to the model of amino acid evolution applied. *V. vinifera* (hypothetical cytochrome P450) was used as outgroup

In addition, we found EX172715 sharing highest sequence similarities with CYP86A7 from *A. thaliana* (Fig. 2B). The conserved amino acids characteristic for cytochrome P450 is present in the sequences: V-E-Q-K-M-S-L-T-L-F-M-K [20]. However, the similarities of obtained sequences to the potential counterparts are useful for phylogenetic analysis; therefore, the given assignments should be taken with care.

3 Discussion

Iridoids in *P. scrophulariiflora* are one kind of plant terpenoids and exploited commercially for a variety of usages, including as pharmaceuticals. Understanding the regulation of iridoids synthesis is of imminent scientific and commercial interest. Current researches show that all plant terpenoids are derived from the common precursor isopentenyl diphosphate (IPP), which is synthesised via the isoprenoid pathway by cyclization of 2, 3-oxidosqualene cyclases (OSCs) to give primarily oleanane (beta-amyrin) or dammarane triterpenoid skeletons. The triterpenoid backbone then undergoes various modifications (oxidation, substitution and glycosylation), mediated by cytochrome P450-dependent monooxygenases, glycosyltransferases and other enzymes [21]. However, like many other terpenoids, the process of iridoids biosynthesis in *P. scrophulariiflora* is still unclear. This is likely to be due to the complexity of the molecules and the lack of the regulatory and metabolic genes in the isoprenoid pathway and at the branch points.

The exploitation and utilization of genetic resources play an important role in research on structure and function of genes. There is a significant discrepancy of iridoids between rhizome and leaf of *P. scrophulariiflora*. We constructed an SSH cDNA library from the plant. It is of help in acquiring the expression spectrum of the structural and regulatory genes involved in secondary metabolites biosynthesis and in finding interesting genes with drug potential.

The SSH is a recently developed new method for identifying differentially expressed genes between two different mRNA populations. The efficiency and reproducibility of SSH have been proved in different studies for differentially expressed genes [22-24]. For the consideration of the phenomena that the shorter insert is easy to ligate with vector, we collected the subtractive cDNA fragments in five fractions (<0.1 kb, 0.1 ~ 0.3 kb, 0.3 ~ 0.5 kb, 0.5 ~ 1 kb and >1 kb). And 27 unique ESTs had been gained in 35 sequenced clones. After SSH, both high and low abundance of expressed genes would be normalized and enriched after two rounds of subtraction and selective PCR [10]. Our random sequencing results showed that there were only two duplicates in 35 fragments and different expressed genes had been normalized. Comparing with the data of GenBank/DDBJ/EMBL, 27 ESTs had not been shown in *P. scrophulariiflora* before, 16 among these ESTs had high homology to the known genes in other plants, while 11 with no significant homology in the database. It further indicates a high efficiency of SSH in isolation of differentially expressed genes.

According to the functions of deduced proteins, the 27 differentially expressed ESTs from the subtracted rhizome-specific cDNA library were classified. Most of them are mainly involved in energy metabolism and translation. Wild differences have been noted in the compositions of the iridoids isolated from the underground and aerial tissues of *P. scrophulariiflora*, which was also indicative of the fact that there are differences in the expression pattern of structural as well as regulatory factors in the two tissues. The key to the differential gene expression in the rhizome as compared with the leaf lies in the function of these genes and it may also account for the variance in the iridoids profiles of the two tissues.

Previous evidences by BLASTX suggested that among the 27 ESTs, the nucleotide sequence of fragment EX172715 matches cytochrome P450 monooxygenase with 74 % identity from *G. max*. And its deduced amino acid sequence shares 63.5% identity with cytochrome P450 monooxygenase (CYP86A7, putative P450) from *A. thaliana*. It is probably partial P450 gene. Cytochrome P450 is a diverse array of multifunctional heme-thiolate protein in plant and plays an important role in secondary metabolism by catalyzing many kinds of reactions, such as synthesis of plant medicinal compounds including fatty acids, iridoids, flavones, alkaloids, etc. [20]. Phylogenetic analysis confirmed EX172715 is the highest homology to cytochrome P450 monooxygenase from *A. thaliana*.

Two fragments (EX172724, EX172733) shared 75% homology with cannabidiolic acid synthase from *Cannabis sativa*. Futoshi et al. demonstrated that expression of the gene led to increase of plant secondary metabolites cannabinoids [25]. A large quantity of intracellular structural and regulatory proteins can be covalently attached to ubiquitin posttranscriptionally and thus modified via a cascade reaction. Ubiquitin proteins are implicated in numerous metabolic processes in eukaryotes. Ubiquitin proteolytic pathway plays an im-

portant role in the degradation of short lived regulatory proteins [26], including those that participate in the cell cycle, cellular signaling in response to stress and to extracellular signals, morphogenesis, the secretory pathway, DNA repair, and organelle biogenesis [27-28]. It is interesting that three fragments (EX172737, EX172725, EX172726) have 90%, 73%, and 73% homology with ubiquitin family protease from *A. thaliana*, respectively. In addition, 1 sequence (EX172731) matches (96%) calmodulin from *Quercus petraea* [29]. Other 6 fragments, of which functions are still unclear, may be involved in the regulation process, such as eukaryotic translation initiation factor (EX172717), coatomer protein (EX172716), translocon-associated protein (EX172720), zinc-protein (EX172732, EX172728) and type II SK2 dehydrin (EX172735) (with antidrought and antifreeze activities) [30].

There is a high probability that one or more of the rhizome-specific novel genes might code for a regulatory factor responsible for the rhizome-specific expression of iridoids biosynthesis. These genes may directly or indirectly be involved in secondary metabolism of *P. scrophulariiflora*.

In conclusion, a host of novel ESTs of *P. scrophulariiflora* were obtained using SSH technique in the present study. The results supported the value of SSH-based EST sequencing as an approach complementary to EST resources for functional genomics research in *P. scrophulariiflora*. We made bioinformatic analysis of 27 differentially expressed ESTs in comparing rhizome with leaf tissues by screening SSH library for the first time. The most significant finding in the current study is the identification of an EST (EX172715) encoding secondary metabolism-related cytochrome P450 monooxygenase with the highest homology to that from *A. thaliana*, and 11 ESTs with no significant homology in public databases, which could potentially be regarded as novel genes or 3′-terminal of full-length cDNA. This result could provide a useful starting point for isolation and identification of novel rhizome-specific genes. Therefore, this study has been included for the first time in a molecular genetic analysis to determine novel genes related to secondary metabolite biosynthesis of *P. scrophulariiflora*.

This report provides a good foundation for full-length cDNA cloning, gene function investigations, molecular mechanisms of modulating secondary metabolite biosynthesis in *P. scrophulariiflora* in further studies.

Acknowledgements: We would like to thank Prof. Huang Luqi and Dr. Zhang Zhaoyang for critical comments on the manuscript, Dr. Li Si-cheng for sharing an unpublished sequence and generously providing plant material. Financial support for this study by Taiji Group is gratefully acknowledged. We are grateful to Li Xiao for his assistance with a portion of data collection.

References:

[1] PENNELL F W. The scrophulariaceae of the Western Himalayas [M]. Dehra Dun: Bishen Singh Mahendra Pal Singh, 1943: 63-66.

[2] GUPTA R K. The living Himalayas [M]. New Delhi: Today and Tomorrow's Printers and Publishers, 1989: 296.

[3] POLUNIN O, STAINTON A. Flowers of the Himalaya [S]. Sixth ed. Delhi: Oxford University Press, 1990: 295.

[4] 吴征镒. 西藏植物志(第4卷)[M]. 北京: 科学出版社, 1985: 279-280.
WU Zhengyi. Tibet plant (Vol. 4) [M]. Beijing: Science Press, 1985: 279-280.

[5] 杨洁, 李萍, 张玉萌, 等. 胡黄连的药理作用研究进展. 解放军药学学报, 2003, 19(3): 217-220.
YANG J, LI P, ZHANG Y M, et al. Advance in pharmacological study of Chinese medicine *P. scrophulariiflora* [J]. Pharm J Chin PLA, 2003, 19: 217-220.

[6] 何薇, 林江涛. 中药胡黄连的化学成分和药理作用的研究进展[J]. 中日友好医院学报, 2005, 19: 369-370.
HE Wei, LIN Jiangtao. Chemical constituents and pharmacological studies of Chinese drug Hu-Huang-Lian [J]. Journal of China-Japan Friendship Hospital, 2005, 19: 369-370.

[7] 傅立国. 中国植物红皮书——稀有濒危植物(第一册)[M]. 北京: 科学出版社, 1992: 23-25.
FU Liguo. Plant red data book of Chinese rare and endangered plants (vol. 1) [M]. Beijing: Science Press, 1992: 23-25.

[8] 黄健,汪正利. 濒危植物胡黄连种群保护的初步研究[J]. 西藏科技, 1995, 70: 7-9.
HUANG Jian, WANG Zhengli. Preliminary research on protection of the rare and endangered plant *Picrorhiza scrophulariiflora* [J]. Tibet's Science & Technology, 1995, 70: 7-9.

[9] 杨世林, 张照, 张奔纲, 等. 珍稀濒危药用植物的保护现状及保护对策[J]. 中草药, 2000, 31: 401-404.
YANG Sinlin, ZHANG Zhao, ZHANG Bengang, et al. Counter measures to overcome the present inadequate means for the protect ion of near extinct species of Chinese medicinal plants [J]. Chinese Traditional and Herbal Drugs, 2000, 31: 401-404.

[10] DIATCHENKO L, LAUY F C, CAMPBELL A P, et al. Suppression subtractive hybridization: a method for generating differentially regulated or tissue-specific cDNA probes and libraries [J]. Proc Natl Acad Sci USA, 1996, 93: 6025 - 6030.

[11] RICHMOND T, SOMERVILLE S. Chasing the dream: plant EST microarrays [J]. Curr Opin Plant Biol, 2000, 3: 108 - 116.

[12] ALBA R, FEI Z, PAYTON P, et al. ESTs, cDNA microarrays, and gene expression profiling: tools for dissecting plant physiology and development [J]. Plant J, 2004, 39: 697 - 714.

[13] 胡松年. 基因表达序列标签(EST) 数据分析手册[M]. 杭州: 浙江大学出版社, 2005.
HU Songnian. Booklets of data analysis on expressed sequence tags (EST) [M]. Hangzhou: Zhejiang University Press, 2005.

[14] LUO W B, YU S J, GAO D W. Suppression subtractive hybridization technique and its advance [J]. Biotechnology, 2000, 10: 37 - 40.

[15] YU J, HU S, WANG J, et al. A draft sequence of the rice genome (*Oryza sativa* L. ssp. *indica*) [J]. Science, 2002, 296: 79 - 92.

[16] BUDIMAN M A, MAO L, WOOD T C, WING R A. A deep coverage tomato BAC library and prospects toward development of an STC framework for genome sequencing [J]. Genome Res, 2000, 10: 129 - 136.

[17] FEI Z, TANG X, ALBA R M, et al. Comprehensive EST analysis of tomato and comparative genomics of fruit ripening [J]. Plant J, 2004, 40: 47 - 59.

[18] THOMSON J D, HIGGINS D G, GBSON T J. Clustal W: improcing the sensitivity of progressive multiple sequence alignment through sequence weighting, position specific gap penalties and weight matrix choice [J]. Nucleic Acids Res, 1994, 22: 4673 - 4680.

[19] PAGE R D M. TREEVIEW: An application to display phylogenetic trees on personal computers [J]. Comput Appl Biosci, 1996, 12: 357 - 358.

[20] ZHAO J, YANG W J, ZHU W H. Cyrochrome P450 and plant secondary metabolism [J]. Chin Bull Life Sci, 1999, 11: 127 - 131.

[21] HARALAMPIDIS K, TROJANOWSKA M, OSBOURN A E. Biosynthesis of triterpenoid saponins in Plants [J]. Adv Biochem Eng Biotechnol, 2002, 75: 31 - 49.

[22] KIM M, KIM S, KI B D. Isolation of cDNA clones differentially accumulated in the placenta of pungent pepper by suppression subtractive hybridization [J]. Mol Cells, 2001, 11: 213 - 219.

[23] MAHALINGAM R, GOMEZ-BUITRAGO A, ECKARDT N, et al. Characterizing the stress/defense transcriptome of Arabidopsis [J]. Genome Biol, 2003, 4: R20.

[24] WANG W, WU P, XIA M, et al. Identification of genes enriched in rice roots of the local nitrate treatment and their expression patterns in split-root treatment [J]. Gene, 2002, 297: 93 - 102.

[25] FUTOSHI T, SATOSHI M, YUKIHIRO S. Purification and characterization of cannabidiolic-acid synthase from *Cannabis sativa* L[J]. J Biol Chem, 1996, 271: 17411 - 17416.

[26] PETERS J M. SCF and APC: the Yin and Yang of cell cycle regulated proteolysis [J]. Curr Opin Cell Biol, 1998, 10: 759 - 768.

[27] HERSHKO A, CIECHANOVER A. The ubiquitin system [J]. Annu Rev Biochem, 1998, 67: 425 - 479.

[28] WEISSMAN A M. Regulation protein degradation by ubiquitination [J]. Immunol Today, 1997, 18: 189 - 198.

[29] 赵岚, 陆融, 姚智. Ca^{2+}/钙调蛋白依赖性蛋白激酶在细胞增值中的作用 [J]. 细胞生物学杂志, 2007, 29: 331 - 335.
ZHAO Lan, LU Rong, YAO Zhi. Effects of Ca^{2+}/calmodulin dependent protein kinases on cell proliferation [J]. Chinese Journal of Cell Biology, 2007, 29: 331 - 335.

[30] 张玉秀, 王梓. 脱水蛋白在逆境下的分子作用机制研究进展 [J]. 自然科学进展, 2007, 16: 23 - 27.
ZHANG Yuxiu, WANG Zi. Research advances on molecular mechanism of dehydrin in stresses [J]. Progress in Natural Sciences, 2007, 16: 23 - 27.