Acta Societatis Botanicorum Poloniae

ORIGINAL RESEARCH PAPER Acta Soc Bot Pol 83(2):147–157 DOI: 10.5586/asbp.2014.012
Received: 2013-12-14 Accepted: 2014-05-13 Published electronically: 2014-07-04

Isolation of biosynthesis related transcripts of 2,3,5,4′-tetrahydroxy stilbene-2-O-β-D-glucoside from *Fallopia multiflora* by suppression subtractive hybridization

Wei Zhao[1,2], Shujing Sheng[2], Zhongyu Liu[2], Di Lu[2], Kuanpeng Zhu[1,2], Xiaoze Li[1,2], Shujin Zhao[2]*, Yan Yao[3]

[1] School of Bioscience and Bioengineering, South China University of Technology, Guangzhou 510006, People's Republic of China

[2] Department of Pharmacy, General Hospital of Guangzhou Military Command, Guangzhou 510010, People's Republic of China

[3] School of Life Sciences, Guangzhou University, Guangzhou 510006, People's Republic of China

Abstract

2,3,5,4′-tetrahydroxy stilbene-2-O-ß-D-glucoside (THSG) exerts multiple pharmacodynamic actions, found in *Fallopia multiflora*, but the biosynthesis pathway of THSG is still unclear. To clear this ambiguity, we constructed suppression subtractive hybridization (SSH) libraries to screen the genes involved in THSG biosynthesis from two *F. multiflora* varieties, which vary significantly in THSG content. Twelve non-redundant differentially expressed sequence tags were obtained and the full lengths of 4 unreported fragments were amplified by rapid amplification of cDNA ends. We totally got 7 full-length transcripts, and all of them were aligned to the transcriptome and digital gene expression tag profiling database of four *F. multiflora* tissues (root, stem and leaf from Deqing *F. multiflora* and another root from Chongqing *F. multiflora*; data unpublished) using local BLAST. The results showed that there was a significant, organ specific difference in the expression of fragments and full-length sequences. All the sequences were annotated by aligning to nucleotide and protein databases. Kyoto Encyclopedia of Genes and Genomes pathway analysis indicated that THSG biosynthesis was correlated with multiple life activities.

Keywords: *Fallopia multiflora*; gene expression difference; transcriptome; 2,3,5,4′-tetrahydroxy stilbene-2-O-β-D-glucooside

Introduction

Fallopia multiflora is a traditional Chinese medicinal herb, which has been widely used for thousands of years. Research showed that *F. multiflora* has potent antioxidative and cytoprotective properties [1], enhanced purgative effects, promoted diuresis and choleretic effects [2], and exerted a neuroprotective effect against glutamate-induced neurotoxicity [3]. 2,3,5,4′-tetrahydroxy stilbene-2-O-ß-D-glucoside (THSG) is an active component in *F. multiflora*, which possesses anti-hyperlipidemic [4], anti-oxidative, anti-inflammatory, endothelial-protective activities [5]. Furthermore, THSG can protect osteoblastic MC3T3-E1 cells via inhibiting the release of bone-resorbing mediators and oxidative damage of the cells [6], and suppress atherosclerosis by altering the expression of key proteins that may be novel molecular targets responsible for atherogenesis [7]. Sun et al. who reported that THSG might provide a potentially new strategy for preventing and treating neurodegenerative disorders such as Parkinson's disease, also showed that THSG may protect neurons against MPP+- induced cell death through improving mitochondrial function, decreasing oxidative stress and inhibiting apoptosis [8]. The pharmacological research revealed that THSG has a potential impact on human health. As a stilbene, THSG and resveratrol (3,5,4′-trihydroxy-trans-stilbene) belong to phenylpropanoids characterized by a 1,2-diphenylethylene backbone. Plant stilbenes are derived from phenylalanine biosynthesis via the general phenylpropanoid pathway. One p-coumaroyl-CoA or cinnamoyl-CoA derived from the phenylpropanoid pathway is ligated with three malonyl-CoA under the catalysis of stilbene synthase (STS) and stilbene is the product [9]. STS is a polyketide synthase (PKS) belonging to type III PKSs. Other PKSs III include resveratrol synthase (RS), chalcone synthases (CHS), bibenzyl synthase (BBS), stilbene carboxylate synthase (STCS), 4-coumaroyltriacetate lactone synthase (CTAS) and more [10]. Every PKSs III catalyzes the synthesis of one unique stilbene. Up to now, no enzyme was found to catalyze the biosynthesis of THSG. Sheng et al. isolated a stilbene synthase gene *FmPKS* from the rhizomes of *F. multiflora*. The gene expression pattern in the plant correlated with the THSG content in different tissues, but

* Corresponding author. Email: gzzsjzhs@163.com

Handling Editor: Przemysław Wojtaszek

THSG was still not detectable in transgenic *Arabidopsis thaliana* in which *FmPKS* was inserted and expressed [11]. Shao et al. showed precursor feeding of methyl jasmonate and salicylic acid in suspension cultures of *F. multiflora* that could increase THSG production [12]. It is still not clear if THSG is synthesized via phenylalanine pathway or some other way. Suppression subtractive hybridization (SSH) is helpful in identifying differentially expressed genes. Extensive studies showed that SSH is a powerful tool in the analysis of stress resistance [13], pathological mechanism [14] and developmental physiology [15]. Moreover, Wang et al. found that light could be effective for activation of the biosynthesis of phenylpropanoids by establishing SSH cDNA libraries of tea calli [16]. In this study, we attempted to identify the genes that are related to THSG biosynthesis by SSH. Full-length sequences were obtained using 3′5′RACE, and then the gene functions were annotated by blastn to nucleotide databases nt, and blastx to protein databases nr, Swiss-Prot, Kyoto Encyclopedia of Genes and Genomes (KEGG) and clusters of orthologous groups (COG). Meanwhile, comparative analysis with the transcriptome and digital expression profile revealed the expression differences for each gene in various tissues of *F. multiflora*. This may indicate that these genes are associated with THSG synthesis.

Material and methods

Plant material collection

F. multiflora plants were gathered from 13 cities (counties) in March 2010, which includes Guangxi Province (Nanning City, Guilin City, Jingxi County, Tianlin County), Guangdong Province (Guangzhou City, Shenzhen City, Dongguan City, Deqing City, Zhaoqing City, Gaozhou City, Zhanjiang City), Jiangxi Province (Xinyu City), Chongqing City. As some samples were gathered in the field, the age was unclear. All plant materials were maintained in the medicinal plant garden of the Department of Pharmacy, Guangzhou Liuhuaqiao Hospital, Guangzhou, China.

HPLC analysis of THSG

To quantify the THSG in *F. multiflora* roots, fresh plant materials were frozen in liquid nitrogen and ground to fine powder in a mortar. After vacuum freeze-drying, 0.2 g of each sample was taken up in 25 ml 50% (v/v) methanol and then refluxed at room temperature for 16 h. After being filtered through a 0.22 μm film, 10 μl of filtrate was analyzed by HPLC using a Dikma Diamonsil C18 column (250 × 4.6 mm, tablets path 5 μm). Chromatographic separation was performed using a solvent system of H_2O and CH_3CN with the ratio of 3:1 (v/v) over 10 min. The flow rate was 1 ml/min, with detection at 320 nm. Each data point represents the average of three independent experiments.

Total RNA extraction and mRNA purification

Total RNA was extracted from leaves and roots using the Plant Total RNA Isolation kit (Bioteke, China) and treated with DNase I (TaKaRa, Dalian, China). Ethidium bromide (EtBr) staining, agarose gel electrophoresis and spectrophotometric (NanoDrop 2000, USA) analysis were performed to examine the quality and concentration of total RNA. mRNA was purified using Oligo tes™-dT^{30}<SUPER> mRNA Purification Kit (from Total RNA) according to the manufacturer's instructions (TaKaRa, Dalian, China).

Suppression subtractive hybridization (SSH)

Driver and tester cDNA was synthesized from two samples at equal amounts of purified mRNA, using SMARTer PCR cDNA Synthesis K it (Clontech, USA) according to the manufacturer's instructions. After repeated twice hybridization, target genes were amplified using nested PCR by Advantage™ cDNA PCR Kit (Clontech, USA). To raise the PCR efficiency, an adapter was added to the nested PCR primers (Tab. 1; homo nested primer). The following target genes were amplified using Fermentas Dream Taq (Fermentas, USA) with homo PCR primer (Tab. 1; homo PCR primer).

The PCR products were inserted into the pMD19-T vector (TaKaRa, Dalian, China), and the ligated products were transformed into *Escherichia coli* DH5α competent cells by heat shock. Then plated onto LB medium containing 100 μg/ml ampicillin, 24 μg/ml IPTG and 20 μg/ml X-gal, and incubated overnight at 37°C. Recombinant white colonies were randomly selected for colony PCR with universal primers (Tab. 1; universal primers). Colonies containing cloned fragments were sent for sequencing. Besides control sample (Poly A$^+$ RNA from human skeletal muscle) provided with the kit, which was used as a control driver cDNA, the experiments were performed with 3 different testers: I – cDNA from Guilin's root sample as tester and Chongqing's root sample as driver; II – Chongqing's root sample as tester and Guilin's root sample as driver; III – Guilin's leaf sample as tester and Guilin's root sample as driver.

Cloning genes with full length by RACE

The 3′5′RACE of the candidate genes was performed in purified mRNA using PCR-select™ cDNA subtraction kit (Clontech, USA) according to the manufacturer's instruction. To improve the specificity, nested primers were designed for Nested-PCR besides gene specific primers. Fragments of 4 genes were selected to do 3′5′RACE from SSH library after BLAST in NCBI (Tab. 1). The Nested-PCR reaction was performed by Advantage cDNA Polymerase Mix (Clontech, USA) with the following thermal cycling parameters: 94°C for 4 min, followed by 35 cycles of 94°C for 30 s, 68°C for 30 s and 72°C for 1.5 min, and the final extension was performed at 72°C for 10 min. The PCR product was cloned into the pMD19-T vector (TaKaRa, Dalian, China) and sequenced.

Sequence analysis

Gene fragments obtained from SSH and RACE were compared with the sequences in nucleotide collection database and expressed sequence tag (EST) database at NCBI using the BLASTN algorithm. Further annotation was carried out using nr, Swiss-Prot, gene ontology database (GO), COG and KEGG. Clustalx1.83 and Contig Express software were used for multiple sequence alignment, sequence identity and linkage. Moreover the obtained sequences were compared with the transcriptome and digital gene expression tag profiling (DGE) database of four different tissue samples (Cr: roots of Chongqing *F. multiflora*; Dr: roots of Deqing

Tab. 1 SSH primers and RACE primers of 4 selected gene fragments.

Gene	Primer	Sequence
	Homo nested primer	5′-GCGACCTACAACATGGCTACCGTCGAGCGGCCGCCCGGGCAGGT
	homo PCR primer	5′-GCGACCTACAACATGGCTACCG
	universal primer	5′-GAGCGGATAACAATTTCACACAGG
GE	3′ GSP	5′-CTGGCAACTGTTGAAGGACGCGAAGA
	5′ GSP	5′-GCCCAAGCACCAAATGCCTCTGC
	3′ Nested primer	5′-AGAGAGGTAGCAACAGATGGAGG
	5′ Nested primer	5′-CCTCCATCTGTTGCTACCTCTCT
GW	3′ GSP	5′-GTCTCATGCGCTCCTCCTCCAGCC
	5′ GSP	5′-GGCGAGCAGAGAAGCAAAGGCTGG
	3′ Nested primer	5′-AGCCTTTGCTTCTCTGCTCGCCT
	5′ Nested primer	5′-GGAGGAGCGCATGAGACAGAACA
CF	3′ GSP	5′-AGGCAATGGAGAAAAGCCGCTCGC
	5′ GSP	5′-TAGGCAACGGCAATCCCAGCACGG
	3′ Nested primer	5′-GGATTGCCGTTGCCTAAAGTGTG
	5′ Nested primer	5′-TTGCTCATTCCTACACTCCTCGC
CA	3′ GSP	5′-CTTGGGCTTTTTCAGGGACAGACG
	5′ GSP	5′-CCGGTAGAGGCCATCAGGATGCAG
	3′ Nested primer	5′-CAGACGAAAAGACGAGAATGAGG
	5′ Nested primer	5′-TTCCTCTTTGGCTTTCCCACTTC

"G" represents the origin of the sample, Guilin city; "C" represents Chongqing city.

F. multiflora; Dl: leaves of Deqing *F. multiflora*; Ds: stems of Deqing *F. multiflora*) of *F. multiflora* (unpublished data) using local BLAST.

Results

Quantification of THSG

To select the specific samples with robust difference in THSG concentration, HPLC was used to analyze the THSG contents. Results indicated that there was a significant difference (P-value < 0.001) in THSG content in *F. multiflora* root from different origins (Fig. 1, Tab. 2). Root samples from Guilin City and Zhaoqin City have the highest accumulation of THSG, with a content of 6.37% and 5.25% respectively. Chongqing samples have the lowest THSG level, with a content of 0.001%. Significant differences were found even within Guilin samples. For example, Guilin samples, the lowest is 1.46%, while the highest is 6.37%. According to this data, we choose Guilin and Chongqing samples as the following materials.

SSH fragments comparison analysis

SSH fragments were amplified and cloned into pMD19-T vector. A total of 136 colonies were picked up, and the insertions were confirmed by PCR using universal primers. 75 clones among the 136 colonies were positive (55.1%) and sequenced, with insertion length ranging from 0.2 to 0.5 kb (data not shown). Removing redundant sequences, 12 fragments were obtained, and they were compared with the sequences in nucleotide and EST database at NCBI using the BLASTN algorithm. Results are shown in Tab. 3. There are 4 sequences with no significant similarity in the databases. The 4 sequences were all obtained from the roots and named GE, GW, CF, CA for subsequent research ("G" represent the origin of the sample, Guilin city; "C" represent Chongqing city.). All sequences were submitted to the NCBI GenBank (accession Nos. JZ469200 to JZ469209) for public domain use, except YCTS4 and GTS2 (ribosomal RNA).

Then the 12 sequences were compared with the transcriptome and DEG database of *F. multiflora* (unpublished data) using local BLAST. Significantly similar genes identified in the transcriptome database were initially aligned by blastx to protein databases nr, Swiss-Prot, KEGG and COG (e-value <0.00001), and then aligned by blastn to nucleotide database nt (e-value <0.00001). Proteins with the highest sequence similarity were retrieved with the given Unigenes and their functional annotations (Tab. 4).

3′5′RACE and full-length sequences comparison analysis

The full-length sequences of GE, GW, CF and CA were cloned by 3′RACE and 5′RACE. More than one full-length sequence was obtained for all genes but one, CF. All sequences have been submitted to the NCBI GenBank (accession Nos. KF054163 to KF054169) for public domain. Meanwhile the sequences were compared with the transcriptome and DEG database using local BLAST. Significantly similar Unigenes in transcriptome were aligned by blastx to protein databases, and aligned by blastn to nucleotide databases nt to annotate gene function (Tab. 5).

Pathway annotation

Different genes generally interact with each other to sustain their biological functions. Pathway-based analysis

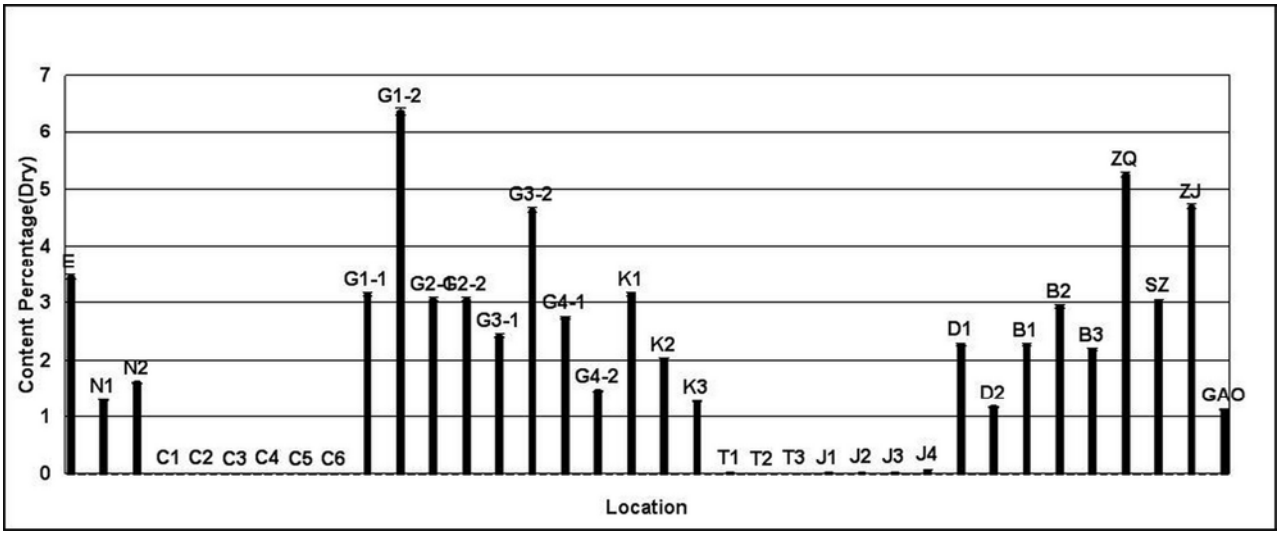

Fig. 1 THSG content in *F. multiflora* plants from different areas. Content percentage (dry): (THSG quality/quality of *F. multiflora* powder) × 100. All samples are the roots of *F. multiflora*, except Gl, Gs, Dl and Ds. B – Guangzhou City; C – Chongqing City; D – Deqing City; E – Dongguan City; G – Guilin City; GAO – Gaozhou City; J – Xinyu City; K – Jingxi County; N – Nanning City; SZ – Shenzhen City; T – Tianlin County; ZJ – Zhanjiang City; ZQ – Zhaoqing City. Dl – leaves of Deqing *F. multiflora*; Ds: stems of Deqing *F. multiflora*; Gl – leaves of Guilin *F. multiflora*; Gs: stems of Guilin *F. multiflora*.

Tab. 2 Quantification of THSG in *F. multiflora* plants from different areas ($n = 3$).

Sample	Average concentration (ng/μl)	Average concentration (%)	Standard deviation	Sample	Average concentration (ng/μl)	Average concentration (%)	Standard deviation
E	272.68	3.35	0.4487	K1	252.88	3.15	0.0125
N1	103.36	1.29	0.0105	K2	162.80	2.01	0.0091
N2	128.56	1.61	0.0004	K3	101.32	1.26	0.0009
C1	0.79	0.01	0.0004	T1	2.08	0.03	0.0001
C2	1.08	0.01	0.0002	T2	0.28	0.00	0.0002
C3	0.08	0.00	0.0000	T3	1.32	0.02	0.0069
C4	0.88	0.01	0.0001	J1	2.24	0.03	0.0005
C5	0.60	0.01	0.0002	J2	3.08	0.04	0.0001
C6	0.36	0.00	0.0006	J3	1.36	0.03	0.0007
G1-1	253.52	3.15	0.0116	J4	6.40	0.08	0.0013
G1-2	509.96	6.37	0.0582	D1	185.68	2.27	0.0244
G2-1	244.96	3.05	0.0127	D2	96.72	1.16	0.0104
G2-2	254.00	3.06	0.0086	B1	182.36	2.26	0.0758
G3-1	194.08	2.43	0.0006	B2	245.04	2.94	0.0995
G3-2	372.20	4.64	0.0179	B3	148.08	2.18	0.0534
G4-1	221.28	2.73	0.0135	ZQ	230.04	5.25	0.0411
G4-2	117.64	1.46	0.0039	ZJ	390.50	4.70	0.0572
SZ	248.25	3.03	0.0092	GAO	93.00	1.12	0.0121
Gl	3.22	0.04	0.0007	Dl	1.56	0.02	0.0105
Gs	84.36	1.01	0.0120	Ds	45.91	0.58	0.0057

Concentration %: (THSG quality/quality of *F. multiflora* powder) × 100. All samples are the roots of *F. multiflora*, except Gl, Gs, Dl and Ds. B – Guangzhou City; C – Chongqing City; D – Deqing City; E – Dongguan City; G – Guilin City; GAO – Gaozhou City; J – Xinyu City; K – Jingxi County; N – Nanning City; SZ – Shenzhen City; T – Tianlin County; ZJ – Zhanjiang City; ZQ – Zhaoqing City. Dl – leaves of Deqing *F. multiflora*; Ds: stems of Deqing *F. multiflora*; Gl – leaves of Guilin *F. multiflora*; Gs: stems of Guilin *F. multiflora*.

Tab. 3 Parameters of high similarity SSH sequence alignments, identified by BLASTN in nucleotide and EST NCBI databases.

Gene	Description	Query cover	*e*–value	Max ident.	Accession
YCTS1	*Rheum australe* catalase mRNA	27%	4e-35	94%	EU931220.1
YCTS2	*Salicornia europaea* TUB mRNA for alpha tubulin, partial cds	95%	2e-71	90%	AB437373.1
	Panicum virgatum clone PV_ABa073-K05, complete sequence	96%	2e-70	90%	AC243247.1
	Phyllostachys edulis cDNA clone: bphylf024i19, full insert sequence	96%	2e-70	90%	FP093842.1
YCTS3	*Mesembryanthemum crystallinum* major latex protein homolog mRNA, complete cds	44%	4e-29	85%	AF054445.1
YCTS4	*Sesbania drummondii* clone SSH-47_02_F03_T7, mRNA sequence	80%	5e-56	95%	DQ465800.1
GTS1	LEAF460 *Polygonum sibiricum* leaf *Knorringia sibirica* cDNA clone LEAF460, mRNA sequence	48%	8e-40	93%	FE903382.1
	XHBC-B22 *Anas platyrhynchos* muscle-related, library constructed by SSH *Anas platyrhynchos*	28%	2e-26	99%	HS410797.1
	cDNA, mRNA sequence, leaf_cn90 cDNA library from leaf of *Polygonum minus*	49%	2e-30	87%	JG745134.1
	Persicaria minor cDNA clone leaf_cn90, mRNA sequence				
GTS2	*Fallopia multiflora* voucher DB1 18S ribosomal RNA, gene, complete sequence	85%	9e-58	98%	EF153706.1
YTS1	*Brassica rapa* subsp. *pekinensis* thionin mRNA, complete cds	36%	4e-09	79%	AF090836.1
CTS1	*Rheum australe* catalase mRNA, complete cds	46%	2e-35	94%	EU931220.1

The character before "TS" represent the tester; "YCTS" means the fragment was found from both "Y" and "C" testers. C – roots of Chongqing *F. multiflora*; G – roots of Guilin *F. multiflora*; Y – leaves of Guilin *F. multiflora*.

Tab. 4 Parameters and annotation of high similarity SSH sequence alignments, identified by local BLAST in *F. multiflora* transcriptome databases.

Gene	Unigene	Score	*e*-value	Gene length	Nt-annotation	Nr-annotation	Accession
GE	*CL3180.Contig2_TrCD*	244	2e-64	1125	*Vitis vinifera* clone SS0AEB22YD16	PREDICTED: vacuolar protein-sorting-associated protein 37 homolog 2 (*Vitis vinifera*)	FQ388706
	CL3180.Contig1_TrCD	244	2e-64	1234	*Vitis vinifera* clone SS0AEB22YD16	PREDICTED: vacuolar protein-sorting-associated protein 37 homolog 2 (*Vitis vinifera*)	FQ388706
GW	*CL5835.Contig2_TrCD*	254	2e-67	1520	PREDICTED: *Vitis vinifera* GPN-loop GTPase 1 homolog-like (LOC100249973), mRNA	hypothetical protein VITISV_029075 (*Vitis vinifera*)	XM_002279556
	CL5835.Contig1_TrCD	254	2e-67	1523	PREDICTED: *Vitis vinifera* GPN-loop GTPase 1 homolog-like (LOC100249973), mRNA	hypothetical protein VITISV_029075 (*Vitis vinifera*)	XM_002279556
CF	*CL9534.Contig2_TrCD*	480	e-135	1155	PREDICTED: *Glycine max* uncharacterized protein LOC100797178 (LOC100797178), mRNA	predicted protein (*Populus trichocarpa*)	XM_003517340
	CL9534.Contig1_TrCD	480	e-135	1168	PREDICTED: *Glycine max* uncharacterized protein LOC100797178 (LOC100797178), mRNA	predicted protein (*Populus trichocarpa*)	XM_003517340
	CL8316.Contig2_TrCD	480	e-135	1137	PREDICTED: *Glycine max* uncharacterized protein LOC100797178 (LOC100797178), mRNA	predicted protein (*Populus trichocarpa*)	XM_003517340

Tab. 4 *(continued)*

Gene	Unigene	Score	*e*-value	Gene length	Nt-annotation	Nr-annotation	Accession
	CL8316.Contig1_TrCD	480	e-135	1150	PREDICTED: *Glycine max* uncharacterized protein LOC100797178 (LOC100797178), mRNA	predicted protein (*Populus trichocarpa*)	XM_003517340
CA	*Unigene22710_TrCD*	418	e-117	1174	PREDICTED: *Vitis vinifera* uncharacterized LOC100248615 (LOC100248615), mRNA	PREDICTED: uncharacterized protein LOC100248615 (*Vitis vinifera*)	XM_002274928
YCTS1	*CL3542.Contig3_TrCD*	234	5e-61	2031	*Rheum australe* catalase mRNA, complete cds	catalase (*Rheum australe*)	EU931220
	CL3542.Contig1_TrCD	111	5e-24	487	*Rheum australe* catalase mRNA, complete cds	catalase (*Rheum australe*)	EU931220
YCTS2	*CL2077.Contig2_TrCD*	434	e-121	1663	*P. amygdalus* mRNA for alpha-tubulin	tubulin alpha chain, putative (*Ricinus communis*)	X67162
	CL2077.Contig1_TrCD	111	3e-24	564	*Populus trichocarpa* tubulin alpha-8 chain (TUA8), mRNA	alpha-tubulin (*Gossypium hirsutum*)	EF151304
	CL2077.Contig7_TrCD	96	2e-19	394	PREDICTED: *Vitis vinifera* tubulin alpha-3 chain-like (LOC100258974), mRNA	alpha-tubulin (*Miscanthus sinensis*)	XM_002281631
YCTS3	*CL5036.Contig1_TrCD*	521	e-147	756	*Mesembryanthemum crystallinum* major latex protein homolog mRNA, complete cds	major latex protein homolog (*Mesembryanthemum crystallinum*)	AF054445
	CL5036.Contig2_TrCD	210	5e-54	455	-	major latex protein homolog (*Mesembryanthemum crystallinum*)	AF054445
YCTS4	*CL8135.Contig1_TrCD*	236	5e-62	1672	*Polygonum sachalinense* 26S ribosomal RNA gene, complete sequence	hypothetical protein MTR_5g050970 (*Medicago truncatula*)	AF479085
GTS1	*CL8017.Contig1_TrCD*	212	6e-55	834	*Phyllostachys edulis* cDNA clone: bphyst022a10, full insert sequence	Chain A, structure of Psbp protein from *Spinacia oleracea* at 1.98 A resolution	FP094014
GTS2	*CL10591.Contig1_TrCD*	248	e-65	2806	*Fallopia multiflora* voucher DX 18S ribosomal RNA gene, complete sequence	Cytochrome P450 likeTBP (*Medicago truncatula*)	EF153701
CTS1	*CL3542.Contig3_TrCD*	230	e-60	2031	*Rheum australe* catalase mRNA, complete cds	catalase (*Rheum australe*)	EU931220
	CL3542.Contig1_TrCD	111	e-24	487	*Rheum australe* catalase mRNA, complete cds	catalase (*Rheum australe*)	EU931220
YTS1	*CL5290.Contig2_TrCD*	486	e-137	780	*Brassica napus* clone Bn 3872 thionin precursor, mRNA, complete cds	crambin precursor = thionin variant Thi2Ca10 (*Crambe abyssinica*, seeds, Peptide Partial, 134 aa)	EU887266
	CL5290.Contig1_TrCD	486	e-137	1094	*Boechera divaricarpa* isolate SLW-C-C05 mRNA sequence	crambin precursor = thionin variant Thi2Ca10 (*Crambe abyssinica*, seeds, Peptide Partial, 134 aa)	DQ226844
	Unigene31799_TrCD	113	9e-25	712	*Boechera divaricarpa* isolate SLW-C-C05 mRNA sequence	crambin precursor = thionin variant Thi2Ca10 (*Crambe abyssinica*, seeds, Peptide Partial, 134 aa)	DQ226844
	CL5290.Contig3_TrCD	113	9e-25	427	*Brassica napus* clone Bn 3872 thionin precursor, mRNA, complete cds	thionin precursor (*Brassica napus*)	EU887266

Tab. 5 Parameters and annotation of high similarity full-length sequences alignments, identified by local BLAST in *F. multiflora* transcriptome databases.

Gene	Unigene	Gene length	Similarity bp	*e*-value	Nt-annotation	Nr-annotation	Accession
GEfl	*CL3180.Contig2_TrCD*	1125	805	0	*Vitis vinifera* clone SS0AEB22YD16	PREDICTED: vacuolar protein-sorting-associated protein 37 homolog 2 (*Vitis vinifera*)	FQ388706
	CL3180.Contig1_TrCD	1234	805	0	*Vitis vinifera* clone SS0AEB22YD16	PREDICTED: vacuolar protein-sorting-associated protein 37 homolog 2 (*Vitis vinifera*)	FQ388706
GWfl1	*Unigene6178_TrCD*	532	451	0	PREDICTED: *Vitis vinifera* 40S ribosomal protein S30-like (LOC100241142), mRNA	unnamed protein product (*Vitis vinifera*)	XM_002281340
	Unigene20297_TrCD	467	190	e-46	PREDICTED: *Vitis vinifera* 40S ribosomal protein S30-like (LOC100252467), mRNA	unnamed protein product (*Vitis vinifera*)	XM_002282997
GWfl2	*CL3625.Contig1_TrCD*	1188	436	0	PREDICTED: *Vitis vinifera* katanin p80 WD40 repeat-containing subunit B1 homolog 1-like (LOC100247509), mRNA	PREDICTED: katanin p80 WD40 repeat-containing subunit B1 homolog 1-like (*Glycine max*)	XM_002264675
GWfl3	*CL3504.Contig1_TrCD*	868	664	0	*Vigna radiata* ubiquitin-conjugating enzyme E2 mRNA, complete cds	PREDICTED: ubiquitin-conjugating enzyme E2 10-like (*Glycine max*)	FJ436357
	Unigene6181_TrCD	979	445	e-62	*Camellia sinensis* ubiquitin-conjugating enzyme E2 mRNA, complete cds	PREDICTED: ubiquitin-conjugating enzyme E2 10-like (*Glycine max*)	JN400596
	CL3504.Contig3_TrCD	973	443	4e-47	PREDICTED: *Vitis vinifera* ubiquitin-conjugating enzyme (LOC100232965), mRNA	predicted protein (*Populus trichocarpa*)	XM_003633920
	CL3504.Contig4_TrCD	915	405	7e-46	PREDICTED: *Vitis vinifera* ubiquitin-conjugating enzyme (LOC100232965), mRNA	ubiquitin-conjugating enzyme E2-17 kDa (*Solanum lycopersicum*)	XM_003633920
	Unigene27674_TrCD	818	306	2e-39	PREDICTED: *Vitis vinifera* ubiquitin-conjugating enzyme (LOC100232965), mRNA	E2 ubiquitin-conjugating enzyme UBC10 (*Brassica napus*)	XM_003633920
	Unigene7679_TrCD	520	243	5e-16	*Solanum tuberosum* clone 116D11 ubiquitin-conjugating protein-like mRNA, complete cds	ubiquitin-conjugating enzyme E2 28 (*Arabidopsis thaliana*)	DQ284472
CFfl	*CL9534.Contig2_TrCD*	1155	144	3e-72	PREDICTED: *Glycine max* uncharacterized protein LOC100797178 (LOC100797178), mRNA	predicted protein (*Populus trichocarpa*)	XM_003517340
	CL9534.Contig1_TrCD	1168	144	3e-72	PREDICTED: *Glycine max* uncharacterized protein LOC100797178 (LOC100797178), mRNA	predicted protein (*Populus trichocarpa*)	XM_003517340
	CL8316.Contig2_TrCD	1137	144	3e-72	PREDICTED: *Glycine max* uncharacterized protein LOC100797178 (LOC100797178), mRNA	predicted protein (*Populus trichocarpa*)	XM_003517340

Tab. 5 *(continued)*

Gene	Unigene	Gene length	Similarity bp	*e*-value	Nt-annotation	Nr-annotation	Accession
	CL8316.Contig1_TrCD	1150	144	3e-72	PREDICTED: *Glycine max* uncharacterized protein LOC100797178 (LOC100797178), mRNA	predicted protein (*Populus trichocarpa*)	XM_003517340
CWfl1	*Unigene22710_TrCD*	1174	316	3e-158	PREDICTED: *Vitis vinifera* uncharacterized LOC100248615 (LOC100248615), mRNA	PREDICTED: uncharacterized protein LOC100248615 (*Vitis vinifera*)	XM_002274928
	CL4034.Contig3_TrCD	2016	209	7e-104	PREDICTED: *Vitis vinifera* pumilio homolog 1-like (LOC100243338), mRNA	PREDICTED: pumilio homolog 1-like (*Vitis vinifera*)	XM_002283155
CWfl2	*CL2293.Contig2_TrCD*	2481	165	9e-82	PREDICTED: *Vitis vinifera* uncharacterized LOC100246727 (LOC100246727), mRNA	expressed protein, putative (*Ricinus communis*)	XM_002273173
	CL2293.Contig1_TrCD	2572	165	9e-82	PREDICTED: *Vitis vinifera* uncharacterized LOC100246727 (LOC100246727), mRNA	expressed protein, putative (*Ricinus communis*)	XM_002273173

helps to further understand genes' biological functions. According to KEGG, the pathways of isolated genes are diverse in function. These genes were mainly involved in metabolic pathways; ubiquitin mediated proteolysis, pathogenic *E. coli* infection, amyotrophic lateral sclerosis (ALS), tryptophan metabolism, photosynthesis, gap junction, RIG-I-like receptor signaling pathway, mRNA surveillance pathway, peroxisome, endocytosis, ribosome, biosynthesis of secondary metabolites, phagosome, protein processing in endoplasmic reticulum, methane metabolism, glyoxylate and dicarboxylate metabolism, microbial metabolism in diverse environments, and RNA transport (Tab. 6).

Comparison of genes expression and THSG content in different tissues

Genes expression of SSH fragments in different plant tissues according to DGE database was shown in Fig. 2. Expression of *GE*, *GW*, *CF* and *YTS1* in different plant tissues has no obvious difference. Gene expression of full- length sequences in different plant tissues has significant differences (Fig. 3), except *GEfl* and *CFfl*. Comparison between the expression of isolated genes and THSG content in different tissues revealed that the expression of *GWfl1* (*Unigene20297_TrCD*), *GWfl3* (*Unigene27674*) and *CWfl1* (*CL4034.Contig3_TrCD*) is in parallel with THSG content in different tissues. Expression of *CA*, *YCTS4* and *CWfl1* (*Unigene22710_TrCD*) is contrary to THSG content in different tissues. Expression of *CA*, *CTS1*, *GWfl2* (*CL3625.Contig1_TrCD*), *GWfl3* (*CL3504.Contig1_TrCD*) , *YCTS4* and *CWfl1* (*Unigene22710_TrCD*) in Cr are higher than in all D sample tissues.

Discussion

As of any secondary metabolite in plant, the production of THSG in *F. multiflora* is influenced by multiple factors, such as temperature, climate, altitude, precipitation etc. Gene screening between samples of distinct product content by SSH could display the mRNA information associated with the production of THSG. DGE profile of different plant tissues contains genes expression information in these tissues. Comparison with DGE profile showed the high level of activity of targeted genes in each tissue. The genes whose expression was significantly different were selected. Similar pattern of difference in gene abundance and THSG content in different tissues confirm that the gene might be associated with THSG biosynthesis.

In our study, the expression of *GWfl1* (*Unigene20297_TrCD*), *GWfl3* (*Unigene27674_TrCD*) and *CWfl1* (*CL4034.Contig3_TrCD*) is in parallel with THSG content in different tissues. Blast in NCBI database showed they might code 40S ribosomal protein, ubiquitin-conjugating enzyme and RNA-binding protein of the Puf family. The predicted protein function indicated there should be active protein degradation and expression behavior in root sample. Expression of *CA*, *YCTS4* and *CWfl1* (*Unigene22710_TrCD*) is contrary to THSG content in different tissues. And *CTS1* is higher expressed in Cr than Dr, Ds and Dl. They might be inhibitors of THSG biosynthesis. As roots always contain the highest amount of THSG, we consider the root is most likely the site of THSG synthesis. The fragments, which were obtained from Guilin root tester should be focused on. *GW* (*CL5835.Contig1_TrCD*/*CL5835.Contig2_TrCD*) and *GTS1* (*CL8017.Contig1_TrCD*) might be involved in providing energy for secondary metabolism via GTP metabolism and photosynthesis. *GWfl2* (*CL3625.Contig1_TrCD*) is annotated as Katanin p80 WD40 repeat-containing subunit B1, *GWfl3* (*CL3504.Contig1_TrCD*) as Ubiquitin-protein ligase. As these fragments were identified in SSH when Guilin root tester was applied, it means the intermediate process of THSG biosynthesis should be fast and efficient, and related protein is likely to be subjected to rapid degradation. *GTS2* (*CL10591.Contig1_TrCD*) has high homology with Cytochrome P450

Tab. 6 Pathway annotation of SSH sequences and full-length sequences participated.

Gene	KOID	Definition	Pathway
GE	K12185	VPS37; ESCRT-I complex subunit VPS37	ko04144 endocytosis
GW	K06883		
CF	-		
CA	-		
YCTS1	K03781	katE, CAT, catB, srpA; catalase (EC:1.11.1.6)	ko01120 microbial metabolism in diverse environments; ko01100 metabolic pathways; ko01110 biosynthesis of secondary metabolites; ko00630 glyoxylate and dicarboxylate metabolism; ko00680 methane metabolism; ko00380 tryptophan metabolism; ko04146 peroxisome; ko05014 amyotrophic lateral sclerosis (ALS)
YCTS2	K07374	TUBA; tubulin alpha	ko04145 phagosome; ko04540 gap junction; ko05130 pathogenic *Escherichia coli* infection
YCTS3	-		
YCTS4	-		
GTS1	K02717	psbP; photosystem II oxygen-evolving enhancer protein 2	ko01100 metabolic pathways; ko00195 photosynthesis
GTS2	K14325	RNPS1; RNA-binding protein with serine-rich domain 1	ko03013 RNA transport; ko03015 mRNA surveillance pathway
CTS1	K03781	katE, CAT, catB, srpA; catalase (EC:1.11.1.6)	ko01120 microbial metabolism in diverse environments; ko01100 metabolic pathways; ko01110 biosynthesis of secondary metabolites; ko00630 glyoxylate and dicarboxylate metabolism; ko00680 methane metabolism; ko00380 tryptophan metabolism; ko04146 peroxisome; ko05014 amyotrophic lateral sclerosis (ALS)
YTS1	-		
GEfl	K12185	VPS37; ESCRT-I complex subunit VPS37	ko04144 endocytosis
GWfl1	K02983	RP-S30e, RPS30; small subunit ribosomal protein S30e	ko03010 ribosome
GWfl2	-		
GWfl3	K06689	UBE2D_E, UBC4, UBC5; ubiquitin-conjugating enzyme E2 D/E (EC:6.3.2.19)	ko04141 protein processing in endoplasmic reticulum; ko04120 ubiquitin mediated proteolysis
CFfl	-		
CWfl1	K14844	PUF6; pumilio homology domain family member 6	ko03009 ribosome biogenesis
CWfl2	K12655	OTUD5, DUBA; OTU domain-containing protein 5 (EC:3.1.2.15)	ko04622 RIG-I-like receptor signaling pathway

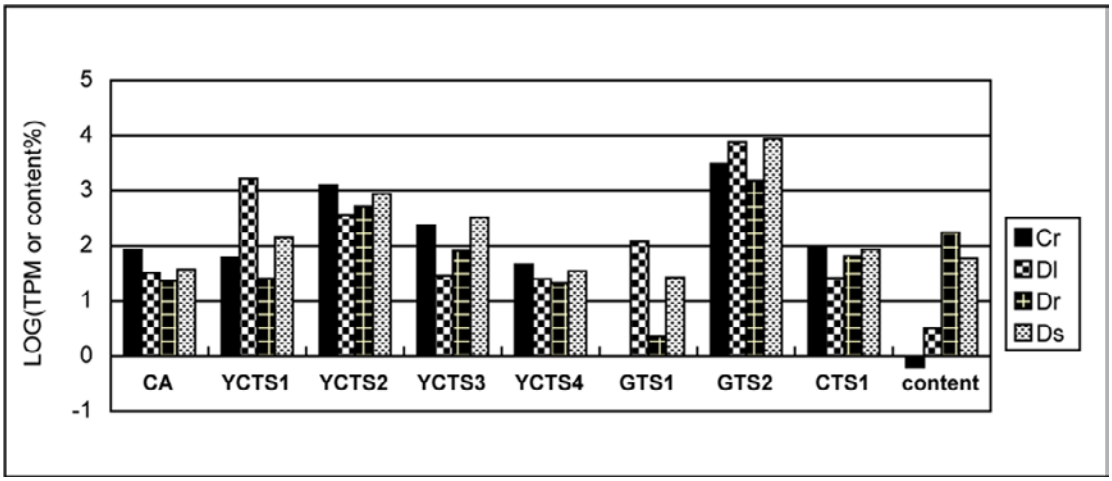

Fig. 2 Comparison of SSH fragments expression and THSG content in different tissues. Cr – roots of Chongqing *F. multiflora*; Dl – leaves of Deqing *F. multiflora*; Dr – roots of Deqing *F. multiflora*; Ds – stems of Deqing *F. multiflora*; TPM – transcripts per million clean tags (all data was calculated as the base of log10).

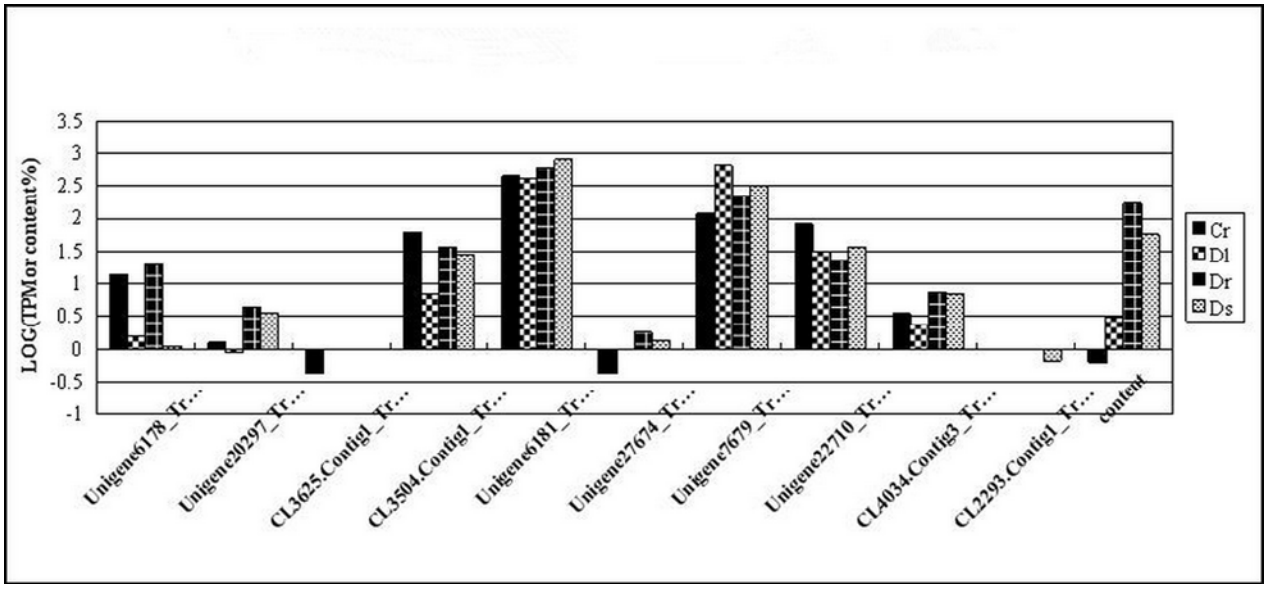

Fig. 3 Comparison of full-length sequences expression and THSG content in different tissues . Cr – roots of Chongqing *F. multiflora*; Dl – leaves of Deqing *F. multiflora*; Dr – roots of Deqing *F. multiflora*; Ds – stems of Deqing *F. multiflora*; TPM – transcripts per million clean tags (all data was calculated as the base of log10).

in *Medicago truncatula*. Cytochrome P450 was found be involved as multifunctional oxidases in the biosynthesis of many secondary metabolites like triterpenoid [17], glycyrrhizin [18], hemolytic saponins [19] and many more. Does it take part in THSG biosynthesis?

If THSG is derived from phenylalanine pathway with other stilbenes, then the *CTS1* gene obtained from Chongqing root tester may be related with the inhibition of THSG biosynthesis. *CTS1* gene is annotated as catalase and participates in "biosynthesis of secondary metabolites (ko01110)"; "glyoxylate and dicarboxylate metabolism (ko00630)" and "tryptophan metabolism (ko00380)". It may regulate the metabolism of tryptophan to influence phenylalanine synthesis. Whether or not, THSG biosynthesis must be complicated and correlate with multiple pathways. More than just a phenylalanine pathway also must have common effect of ATP energy metabolism active and other related reactions to strengthen or weaken the process. Our research isolated several candidate genes that might participate in the THSG biosynthesis. Further research will focus on the characterization of candidate genes and gene screening to transcriptome in more diverse tissues.

Acknowledgments

This work was funded by grants from Natural Science Foundation of Guangdong Province (10151001002000012), Science and Technology Department of Guangdong Province (2010B060200009).

Authors' contributions

The following declarations about authors' contributions to the research have been made: conceived and designed the experiments: WZ, SS, ZL; performed the experiments: WZ, DL, KZ, XL, YY; performed the analysis: WZ, SZ; wrote the paper: WZ; revised the paper: SZ, SS.

References

1. Steele ML, Truong J, Govindaraghavan S, Ooi L, Sucher NJ, Münch G. Cytoprotective properties of traditional Chinese medicinal herbal extracts in hydrogen peroxide challenged human U373 astroglia cells. Neurochem Int. 2013;62(5):522–529. http://dx.doi.org/10.1016/j.neuint.2012.08.018
2. Xie W, Zhao Y, Du L. Emerging approaches of traditional Chinese medicine formulas for the treatment of hyperlipidemia. J Ethnopharmacol. 2012;140(2):345–367. http://dx.doi.org/10.1016/j.jep.2012.01.027
3. Jang JY, Kim HN, Kim YR, Choi YW, Choi YH, Lee JH, et al. Hexane extract from *Polygonum multiflorum* attenuates glutamate-induced apoptosis in primary cultured cortical neurons. J Ethnopharmacol. 2013;145(1):261–268. http://dx.doi.org/10.1016/j.jep.2012.10.061
4. Wang M, Zhao R, Wang W, Mao X, Yu J. Lipid regulation effects of Polygoni Multiflori Radix, its processed products and its major substances on steatosis human liver cell line L02. J Ethnopharmacol. 2012;139(1):287–293. http://dx.doi.org/10.1016/j.jep.2011.11.022
5. Lv G, Lou Z, Chen S, Gu H, Shan L. Pharmacokinetics and tissue distribution of 2,3,5,4′-tetrahydroxystilbene-2-O-β-D-glucoside from traditional Chinese medicine *Polygonum multiflorum* following oral administration to rats. J Ethnopharmacol. 2011;137(1):449–456. http://dx.doi.org/10.1016/j.jep.2011.05.049
6. Zhang JK, Yang L, Meng GL, Fan J, Chen JZ, He QZ, et al. Protective effect of tetrahydroxystilbene glucoside against hydrogen peroxide-induced dysfunction and oxidative stress in osteoblastic MC3T3-E1 cells. Eur J Pharmacol. 2012;689(1–3):31–37. http://dx.doi.org/10.1016/j.ejphar.2012.05.045
7. Yao WJ, Fan WJ, Huang C, Zhong H, Chen XF, Zhang W. Proteomic analysis for anti-atherosclerotic effect of tetrahydroxystilbene glucoside in rats. Biomed Pharmacother. 2013;67(2):140–145. http://dx.doi.org/10.1016/j.biopha.2012.10.007
8. Sun FL, Zhang L, Zhang RY, Li L. Tetrahydroxystilbene glucoside protects human neuroblastoma SH-SY5Y cells against MPP+-induced cytotoxicity. Eur J Pharmacol. 2011;660(2/3):283–290. http://dx.doi.org/10.1016/j.ejphar.2011.03.046
9. Chong J, Poutaraud A, Hugueney P. Metabolism and roles of stilbenes

in plants. Plant Sci. 2009;177(3):143–155. http://dx.doi.org/10.1016/j.plantsci.2009.05.012

10. Hartmann T. From waste products to ecochemicals: fifty years research of plant secondary metabolism. Phytochemistry. 2007;68(22–24):2831–2846. http://dx.doi.org/10.1016/j.phytochem.2007.09.017
11. Sheng SJ, Liu ZY, Zhao W, Shao L, Zhao SJ. Molecular analysis of a type III polyketide synthase gene in *Fallopia multiflora*. Biologia (Bratisl.). 2010;65(6):939–946. http://dx.doi.org/10.2478/s11756-010-0113-8
12. Shao L, Zhao SJ, Cui TB, Liu ZY, Zhao W. 2,3,5,4′-tetrahydroxystilbene-2-O-β-D-glycoside biosynthesis by suspension cells cultures of *Polygonum multiflorum* Thunb and production enhancement by methyl jasmonate and salicylic acid. Molecules. 2012;17(12):2240–2247. http://dx.doi.org/10.3390/molecules17022240
13. Peng T, Zhu XF, Fan QJ, Sun PP, Liu JH. Identification and characterization of low temperature stress responsive genes in *Poncirus trifoliata* by suppression subtractive hybridization. Gene. 2012;492(1):220–228. http://dx.doi.org/10.1016/j.gene.2011.10.025
14. Gao M, Wang Q, Wan R, Fei Z, Wang X. Identification of genes differentially expressed in grapevine associated with resistance to *Elsinoe ampelina* through suppressive subtraction hybridization. Plant Physiol Biochem. 2012;58:253–268. http://dx.doi.org/10.1016/j.plaphy.2012.07.009
15. Pimentel P, Salvatierra A, Moya-León MA, Herrera R. Isolation of genes differentially expressed during development and ripening of *Fragaria chiloensis* fruit by suppression subtractive hybridization. J Plant Physiol. 2010;167(14):1179–1187. http://dx.doi.org/10.1016/j.jplph.2010.03.006
16. Wang YS, Gao LP, Wang ZR, Liu YJ, Sun ML, Yang DQ, et al. Light-induced expression of genes involved in phenylpropanoid biosynthetic pathways in callus of tea [*Camellia sinensis* (L.) O. Kuntze]. Sci Hortic. 2012;133:72–83. http://dx.doi.org/10.1016/j.scienta.2011.10.017
17. Fukushima EO, Seki H, Ohyama K, Ono E, Umemoto N, Mizutani M, et al. CYP716A subfamily members are multifunctional oxidases in triterpenoid biosynthesis. Plant Cell Physiol. 2011;52(12):2050–2061. http://dx.doi.org/10.1093/pcp/pcr146
18. Seki H, Sawai S, Ohyama K, Mizutani M, Ohnishi T, Sudo H, et al. Triterpene functional genomics in licorice for identification of CYP72A154 involved in the biosynthesis of glycyrrhizin. Plant Cell. 2011;23(11):4112–4123. http://dx.doi.org/10.1105/tpc.110.082685
19. Carelli M, Biazzi E, Panara F, Tava A, Scaramelli L, Porceddu A, et al. *Medicago truncatula* CYP716A12 is a multifunctional oxidase involved in the biosynthesis of hemolytic saponins. Plant Cell. 2011;23(8):3070–3081. http://dx.doi.org/10.1105/tpc.111.087312