

Identification of the candidate genes related to sanshool biosynthesis in *Zanthoxylum bungeanum* by comparative RNA sequencing analysis

Jingmiao Li, Xuefang Sun, Qingqing Li, Lixiu Hou, Anzhi Wei*, Yulin Liu*

College of Forestry, Northwest A&F University, Yangling 712100 China

*Corresponding authors, e-mail: weianzhi@126.com, lyl12504001@126.com

Received 31 Oct 2018

Accepted 20 Oct 2019

ABSTRACT: *Zanthoxylum bungeanum* is an edible and medicinal plant with great economic value in East Asia due to high sanshool content in its pericarp. Sanshools not only provide a unique pungency in cuisine, but also have potential for use in anaesthetics, anticancer drugs, and botanical safeners. In this study, to identify the candidate genes related to sanshool biosynthesis, the transcriptomes of pericarps from three developmental stages were comparatively analysed. According to differentially expressed gene (DEG) analysis, 2322 unigenes were upregulated or downregulated significantly among the three types of pericarps and 176 of the DEGs were putative transcription factors (TFs) belonging to 40 TF families. Furthermore, based on the chemical structure of sanshools and unigene annotation, DEGs involved in the pathways of fatty acid biosynthesis and metabolism and amino acid transport were predicted. These DEGs were related to sanshool biosynthesis and 12 candidate enzymes encoded by 18 unigenes were related to the biosynthesis and metabolism of sanshools and identified and classified into three groups. This study provides a global survey of gene expression patterns in pericarp development of *Z. bungeanum*, which offers clues regarding important regulatory mechanisms, especially with respect to sanshool biosynthesis in the pericarp of *Z. bungeanum*.

KEYWORDS: transcriptome, *Zanthoxylum bungeanum*, differentially expressed genes, sanshool biosynthesis

INTRODUCTION

The plants of *Zanthoxylum*, which belong to the family Rutaceae, consist of approximately 250 species and are distributed worldwide¹. *Zanthoxylum bungeanum*, usually called Sichuan pepper or Chinese red pepper, is one of the most representative species of *Zanthoxylum* in China and is mainly used as a spice in Chinese cuisine and in traditional herbal medicine for the treatment of stomachache, toothache, diarrhoea, ascariasis, and colic. Like other species in the *Zanthoxylum* genus, such as *Z. piperitum*², *Z. heterophyllum*³ and *Z. armatum*⁴, the most valuable tissue of *Z. bungeanum* is the pericarp which is rich in a number of bioactive secondary metabolites, including coumarins, aromatic and aliphatic amides, alkaloids and lignans. Among these metabolites, aliphatic amides, which are classified as sanshools, are the major active components. They have a unique pungent taste and cause paresthetic sensations on the tongue⁵. Furthermore, sanshools and their analogues have been shown to induce apoptosis of human hepatocarcinoma HepG2 cells and block expression of the oncogene KCNK9^{6,7}. Thus these compounds

have been considered potential anticancer drugs. Sanshools have also been shown to reduce damage to rice seedlings caused by metolachlor⁸.

Due to their potential economic value, many studies have focused on separating sanshools and their analogues, identifying their chemical structures⁹, artificially synthesising sanshools, and exploring their pharmacological effects^{10,11}. However, due to the lack of genomic information, molecular synthesis pathways and the related genes remain to be elucidated.

To screen and identify the pathways and genes related to the biosynthesis of secondary metabolites, RNA-seq based on high-throughput sequencing technology is a cost-effective and powerful tool to obtain many transcripts from different tissues or the same tissue from different developmental stages, which can be used for differentially expressed gene (DEG) analysis combined with Gene Ontology (GO) classification and Kyoto Encyclopedia of genes and Genomes (KEGG) pathway functional enrichment analysis to identify key genes^{12–14}. For example, pathways and genes involved in the production of secondary metabolites, such as flavonoid and stil-

benoid biosynthesis in *Gnetum parvifolium*¹⁵, terpenoid and thenylpropanoid biosynthesis in *Ananas comosus*¹⁶ and caffeine and theanine biosynthesis in *Camellia sinensis*¹⁷ have been verified by comparative transcriptome sequencing.

Herein, we used the Illumina HiSeq 2000 sequencing platform to obtain the pericarp transcriptomes of *Z. bungeanum* during three different developmental stages (30, 50, and 90 days after flowering [DAF30, DAF50, and DAF90, respectively]) to reveal gene expression patterns in the different stages, analyse significantly affected pathways, and identify the candidate enzymes related to sanshool biosynthesis.

MATERIALS AND METHODS

Sample collection and preparation

Fresh pericarps at DAF30 with no numb taste, DAF50 with light numb taste, and DAF90 with strong numb taste were selected separately from the *Z. bungeanum* cultivar 'Shizitou'. All pericarps were grown under natural conditions in the *Z. bungeanum* demonstration station of Northwest A&F University in Fengxian County, Shaanxi Province, China. The pericarps from each developmental stage were randomly collected from the same three individuals and mixed together for determination of total acylamide content, RNA isolation, and qRT-PCR.

Determination of total acylamide content

To ensure that the samples collected from the three development stages were reasonable for comparative RNA-seq analysis, the total acylamide content (Q) of each stage was determined by the rapid formaldehyde titration assay and expressed as mg/g. That is,

$$Q = \frac{14.0067 \times 2.2696 \times CT_2(\sum V_{3i}/3)}{T_1 T_3}.$$

In this formula, V_{3i} is the volume of NaOH solution consumed in the determination of ammonium nitrogen content (ml); C is the accurate calibration concentration of NaOH solution; T_1 is the mass of dry pericarp powder sample (g); T_2 is the mass of extract obtained from pericarp powder sample (g); T_3 is the mass of extract after the conversion of acylamide to ammonium nitrogen (g); 14.0067 is the molecular weight of nitrogen and 2.2696 is the adjusted coefficient. More detailed experimental methods and procedures were described in the previous study¹⁸ and this experiment was carried out on three technical replicates.

RNA isolation and transcriptome sequencing

Total RNA was extracted using TRIzol reagent (Invitrogen, Carlsbad, CA, USA) according to the manufacturer's instructions. After verifying RNA quality and quantity by 1% agarose gel and NanoPhotometer spectrophotometer (Implen NanoPhotometer, Westlake Village, CA, USA), 2.0 µg of RNA was used to construct cDNA libraries for each sample of pericarp at DAF30, DAF50, and DAF90 as previously described¹⁹. Transcriptome sequencing was conducted at Beijing Novogene Biological Information Technology Co., Ltd., Beijing, China (www.novogene.com) based on Illumina HiSeq 2000 platform and the raw data were submitted to the NCBI Sequence Read Archive with accession number SRP142080.

De novo assembly and unigene annotation

Sequencing data from the three developmental stages were combined for assembly. Prior to assembly, in-house perl script was used to filter the adaptors, poly-N, and low-quality reads, following which the generated clean reads were used for De novo assembly by Trinity to produce transcripts²⁰. Then, these transcripts were clustered and the longest transcripts in each cluster were extracted as unigenes, which were further used as reference sequences for DEG analysis.

All unigenes were annotated by comparison against NR, SwissProt, PFAM, KOG, GO, and KO protein databases using BLASTx and the nucleotide database NT via BLASTn (E-values < 10⁻⁵). In addition, the iTAK pipeline (bioinfo.bti.cornell.edu/cgi-bin/itak/index.cgi) was used to identify candidate transcription factors (TFs) and classify them into different families based on protein domains of the unigenes.

Identification and classification of DEGs

RSEM²¹ and Bowtie²² were used to calculate transcript abundance. The expected number of fragments per kilobase of transcript sequence per millions base pairs sequenced (FPKM) method was used to normalize gene expression levels in the three samples²³. To analyse DEGs, we used the trimmed mean of M values to standardize the read count, and the DEGseq R package to select DEGs with the following threshold: q-value < 0.005 and |log₂(fold change)| > 1²⁴. Expression levels of the identified unigenes were calculated based on the log₂ FPKM value and the heat map was produced using HemI software²⁵.

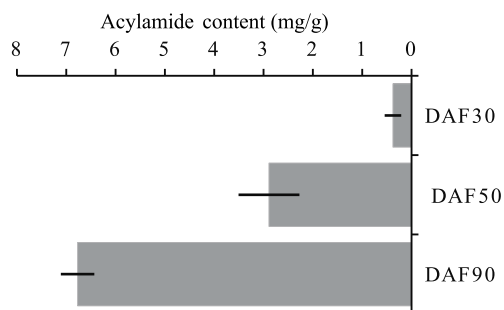


Fig. 1 Acylamide content of pericarps in the three developmental stages of *Z. bungeanum*.

To determine the DEGs that significantly correlated with biological functions, GSeq based on Wallenius noncentral hypergeometric distribution was used for GO enrichment analysis²⁶ and KOBAS software was used to test statistical enrichment of these DEGs in KEGG pathways²⁷.

qRT-PCR analysis

Unigenes coding for the enzymes involved in sahnshool biosynthesis and metabolism were selected for validation by qRT-PCR. Gene-specific primers were obtained using Primer3 (bioinfo.ut.ee/primer3) according to the open reading frames of the target genes. qRT-PCR was performed using SYBR Premix ExTaq II (Takara Co., Ltd., Japan) in 10 µl volumes on the CFX96 Real-Time PCR Detection System (Bio-Rad) using the following program: 95 °C for 30 s, followed by 40 cycles of 94 °C for 5 s, 55 °C for 30 s, and 72 °C for 45 s. Three technical replicates and three biological replicates were analysed for each unigene. The relative expression of the test genes in each sample was calculated using the $2^{-\Delta\Delta CT}$ method with the translation initiation factor gene (TIF)²⁸.

RESULTS AND DISCUSSION

Quantification of total acylamide content in different developmental stages

Based on the extent of the numb taste, the three developmental stages of pericarps were used to investigate the acylamide content, which was determined to be highest in DAF90 (6.77 ± 0.34 mg/g), followed by DAF50 (2.86 ± 0.62 mg/g) and DAF30 (0.38 ± 0.18 mg/g) (Fig. 1). The increasing trend of acylamide content was consistent with the degree of numb taste, which indicated that the three chosen developmental stages were appropriate.

Table 1 Summary of transcriptome sequencing data.

Sample	Clean reads	Clean bases	Mapped reads
DAF30	62 518 534	9.38 Gb	40 953 294 (65.51%)
DAF50	72 341 948	10.85 Gb	47 612 244 (65.82%)
DAF90	55 039 716	8.26 Gb	35 591 716 (64.67%)

Table 2 Summary data of the assembled transcripts and unigenes in *Z. bungeanum*.

Characteristic detail	Transcripts	Unigenes
No. of reads in 201–500 bp	156 030	123 567
No. of reads in 501–1000 bp	42 698	23 459
No. of reads in 1001–2000 bp	26 783	11 343
No. of reads in > 2000 bp	9941	4971
Total reads	235 452	163 340
Min length (bp)	201	201
Mean length (bp)	602	503
Median length (bp)	349	302
Max length (bp)	13 123	13 123
N50 (bp)	895	615
N90 (bp)	254	237
Total nucleotides (bp)	141 849 569	82 188 724

Sequencing and de novo assembly

Comparative analysis of the transcriptional profiling derived from RNA-seq is a fast and efficient way to identify candidate genes involved in target pathways^{29,30}. In this study, three developmental stages of pericarps (DAF30, DAF50, and DAF90) were selected for transcriptome sequencing. After filtering the raw data, 62 518 534, 72 341 948, and 55 039 716 clean reads were obtained from the DAF30, DAF50, and DAF90 samples, respectively, generating approximately 28.49 Gb in total (Table 1). Using Trinity software, these clean reads were assembled into 235 452 transcripts ranging from 201–13 123 bp and with a mean length of 602 bp. Then, by selecting the longest transcript for each transcript cluster, 163 340 unigenes were obtained with a mean length and N50 of 503 and 615 bp, respectively. Of the 123 567 unigenes, 75.65% were between 201 and 500 bp and 3.04% (4971) of the unigenes were longer than 2000 bp (Table 2). The data size and unigene number were much larger than a previous study³¹, which generated a more comprehensive transcriptome of *Z. bungeanum* pericarps.

Unigene annotation and functional classification

To investigate the molecular functions of these unigenes, a set of nucleotide or protein databases

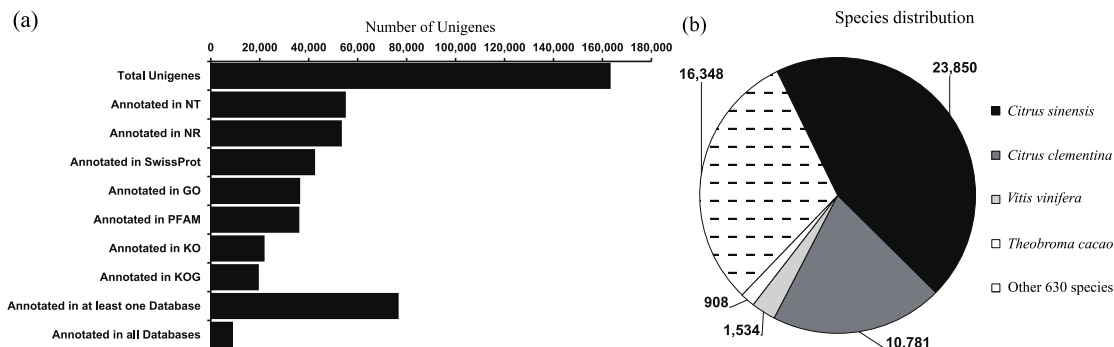


Fig. 2 (a) Summary of functional annotation of assembled unigenes; (b) distribution of BLAST results by species shown as percentage of total homologous sequences.

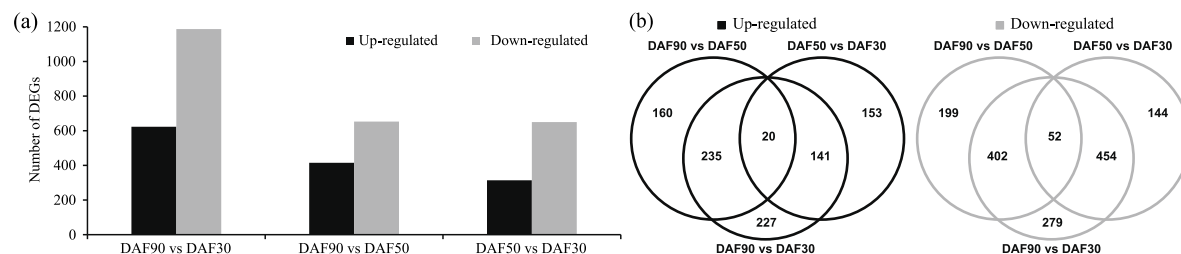


Fig. 3 (a) Number of differential expression genes among DAF30, DAF50, and DAF90 pericarp developmental stages of *Z. bungeanum*; (b) Venn diagram of the number of upregulated and downregulated genes pairwise among the three developmental stages.

(NT, NR, SwissProt, PFAM, GO, KO, and KOG) were selected for unigene annotation by Blastn/x. As shown in Fig. 2a, a total of 76 744 (46.98%) unigenes were successfully annotated in at least one database, 9074 (5.55%) were annotated in all databases, and 86 596 unigenes (53.02%) were not annotated in the public databases chosen in this study. The low rate of annotation may be attributed to the limited genomic information available for *Z. bungeanum* or because most of unannotated unigenes were too short to generate a blast hit and the same results also appeared in the transcriptomes of *Carya illinoensis*³², *Pseudostellaria heterophylla*¹², and *G. parvifolium*¹⁵ which were constructed of several libraries and generated many unigenes. Additionally, species distribution analysis based on blast hits against the non-redundant database showed that *Citrus sinensis* and *Citrus clementina* were the top-hit species, which are in the same evolutionary unit as Rutaceae with *Z. bungeanum*³³, and accounted for 64.83% (34 631) of the identified unigenes. *Vitis vinifera* (1534, 2.87%) and *Theobroma cacao* (908, 1.70%), together with 630 other species, accounted for the remaining 35.17% (18 790) of the identified unigenes (Fig. 2b).

Transcription factor (TF) prediction

In plants, a number of TFs, including AP2-EREBP, MYB, WRKY, and bHLH, are involved in the biosynthesis and accumulation of secondary metabolites synthesised in different tissues or organs at different developmental stages^{34,35}. To predict the transcriptional regulation in detail, the iTAK pipeline was used to mine the TFs. In total, 2376 (1.45%) unigenes encoded putative TFs, representing 80 different families. Members of the C2H2 TF family were the most abundant, including 184 unigenes. MYB (163), AP2-EREBP (115), C3H (115), and orphan genes (113) were determined. The remaining 75 TF families included 1686 unigenes (Table 3).

Identification of DEGs

To examine the genes differentially expressed in the three developmental stages of pericarps, clean reads from individual samples were separately mapped to the reference sequences (unigenes). Overall, 65.51% (40 953 294), 65.82% (47 612 244), and 64.67% (35 591 716) clean reads from DAF30, DAF50, and DAF90 samples were successfully mapped to the reference sequences, respectively (Fig. 1). In addition, 16 694, 21 375, and 13 855

Table 3 Number of unigenes coding for putative TFs that were differentially expressed among DAF30, DAF50, and DAF90 developmental stages of pericarp in *Z. bungeanum*.

TF Family	Unigene no.	DEGs	DAF90 vs DAF30		DAF90 vs DAF50		DAF50 vs DAF30	
			Up	Down	Up	Down	Up	Down
C2H2	184	1	1	0	1	0	0	0
MYB	163	19	1	14	1	6	2	7
AP2-EREBP	115	12	3	3	1	2	6	2
C3H	115	4	0	2	1	2	0	2
Orphans	113	25	4	13	5	12	2	9
NAC	95	9	1	4	3	1	0	4
bHLH	91	10	2	4	2	2	2	1
HB	84	9	2	7	1	2	0	3
WRKY	83	9	8	0	3	0	3	0
GRAS	82	4	0	1	1	3	0	0
bZIP	70	6	0	6	0	0	0	2
AUX/IAA	61	5	0	3	0	5	0	0
PHD	57	0	0	0	0	0	0	0
MADS	53	5	0	4	0	1	1	1
CCAAT	50	4	1	0	1	1	1	2
mTERF	48	0	0	0	0	0	0	0
SNF2	47	0	0	0	0	0	0	0
SET	41	1	1	0	0	0	0	0
C2C2-Dof	39	4	0	2	0	3	2	1
FAR1	39	0	0	0	0	0	0	0
LOB	36	3	0	0	0	2	0	1
TRAF	36	3	0	2	1	0	0	1
Trihelix	35	0	0	0	0	0	0	0
GNAT	34	2	1	1	0	1	0	0
ABI3VP1	32	0	0	0	0	0	0	0
TCP	32	2	1	0	2	0	0	0
G2-like	31	3	1	1	1	2	0	0
HMG	29	0	0	0	0	0	0	0
ARF	27	5	2	2	0	3	0	0
C2C2-GATA	26	2	0	2	0	1	0	0
Other 50	428	29	3	17	6	13	2	5
Total	2376	176	32	88	30	62	21	41

unigenes were uniquely expressed in the DAF30, DAF50, and DAF90 samples, respectively, while 58 023 unigenes were expressed during all three stages. The comparative read abundance between the transcriptome of the three samples revealed that 2322 unigenes (2277 annotated and 45 unannotated) were regarded as DEGs when the q -value < 0.005 and $\log_2|FC| \geq 1$. Of these DEGs, 964 were identified between DAF50 and DAF30, and 314 were upregulated and 650 were downregulated. Between the DAF90 and DAF50 samples, 415 DEGs were considered to be upregulated, while 653 DEGs were downregulated. Between the DAF90 and DAF30 samples, 1810 unigenes were regarded as DEGs, and 623 were upregulated and 2830 were downregulated (Fig. 3a). Furthermore, the tran-

scription of 20 and 52 DEGs rose and declined continuously from DAF30 to DAF90, respectively (Fig. 3b). These results suggested that during pericarp development in *Z. bungeanum*, many genes were dynamically regulated at the level of transcription.

Among the identified DEGs, 176 were putative TFs belonging to 40 TF families. Of these, 25 were orphan TFs (14.20%), 19 were MYB TFs (10.78%), 12 were AP2-EREBP TFs (6.82%), 10 were bHLH TFs (5.68%), and 110 were in other TF families (62.50%). In the DAF50 vs DAF30, DAF90 vs DAF50, and DAF90 vs DAF30 groups, 21, 30, and 32 unigene-encoded TFs were upregulated and 41, 62, and 88 were downregulated, respectively. The expression characteristics of these TFs, shown

in Table 2, suggest that they are involved in the regulation of secondary metabolite biosynthesis in *Z. bungeanum*. Because few studies have attempted to identify the enzymes or TFs related to sanshools or their structural analogues, the differentially expressed TFs that participate in sanshool biosynthesis need to be studied further.

DEGs significantly enriched in GO, eukaryotic orthologous groups (KOG), and KEGG pathways

To elaborate on the biological functions and pathways activated by the DEGs during the three developmental stages, DEGs were mapped to GO categories, KOG clusters, and KEGG pathways. As shown in Fig. 4, 1760 DEGs were assigned to 50 functional subcategories, including 19, 8, and 23 to the biological process (BP), cellular component (CC), and molecular function (MF) categories, respectively. For the three comparative groups in the BP category, most DEGs were enriched in the 'metabolic process', 'single-organism process', 'single-organism metabolic process', 'oxidation-reduction process', 'carbohydrate metabolic process', and 'lipid metabolic process' subcategories. Furthermore, 145 DEGs were enriched in the 'carbohydrate derivative metabolic process' subcategory between DAF50 and DAF30, while no DEGs were detected between DAF90 and DAF50 or DAF30. In the CC category, more unigenes were downregulated and only a few unigenes were upregulated over the course of pericarp development. Furthermore, in the DAF50 vs DAF30 group, no DEGs were identified. In the MF category, 'catalytic activity' and 'oxido reductase activity' were the two most significantly enriched GO subcategories, which suggested that many enzymes with catalytic and oxidoreductase activities participated in pericarp development. In addition, the ratio of upregulated and downregulated unigenes in the MF category was higher than the other two categories.

Among the 2322 DEGs, 939 were classified into 24 functional categories based on KOG designations. Functional profiles of these DEGs differed between different stages. Furthermore, more DEGs were enriched in eight categories, including 'amino acid transport and metabolism', 'carbohydrate transport and metabolism', 'lipid transport and metabolism', 'secondary metabolite biosynthesis, transport and catabolism', and four other categories with the category codes C, O, R, and T (Fig. 5).

Based on KEGG pathway enrichment, 797 DEGs were successfully assigned to KEGG pathways. The top 20 pathways in each comparison group with

Table 4 qRT-PCR validation primers.

Gene	Primer (5'-3')
>c123868_g1 (KAR)	F: TGCTATTGCACCAGGATTCA R: AATTCCACCAGTCCAGCAAC
>c123808_g1 (BCAT)	F: AGTAGCAGCGCAGTCATAGC R: TGAGAGACGAGGGAGTGAGG
>c122478_g1 (FAAH)	F: GCCTTTACCTTCACAACAGCC R: GGTTTGAAGATTCGCGGACTG

the lowest q-values are shown in Fig. 6. In the DAF50 vs DAF30 group, we found that upregulated unigenes were primarily involved in the fatty acid metabolism and biosynthesis-related pathways and 'terpenoid backbone biosynthesis', but the downregulated unigenes were primarily involved in 'phenylpropanoid biosynthesis', 'phenylalanine metabolism', and 'starch and sucrose metabolism'. In the DAF90 vs DAF50 group, 'phenylpropanoid biosynthesis' contained slightly more upregulated unigenes than other pathways. However, 'photosynthesis' was the most enriched pathway among the downregulated unigenes, which was same as in the DAF90 vs DAF30 group. Furthermore, similar to the DAF50 vs DAF30 group, 'flavonoid biosynthesis' and 'fatty acid metabolism' were also enriched in the upregulated unigenes of the DAF90 vs DAF30 group, while, in contrast with DAF90 and DAF30 compared to DAF50, 'glutathione metabolism', which may be involved in the metabolism of amide groups, was one of the most enriched pathways in the upregulated unigenes of the DAF90 vs DAF30 group.

Sanshools are a set of polyunsaturated fatty acid amides that possess long-chain unsaturated fatty acid structures and an isobutyl or hydroxylated isobutyl chain linked to the amide moiety (Fig. 7a)^{11,36}. Thus combining the above results indicates that the genes involved in fatty acid biosynthesis and metabolism and amino acid transport and metabolism were dynamically regulated at the transcriptional level during pericarp development, which suggests it is important to identify the genes related to sanshool biosynthesis.

Identification of unigenes related to the biosynthesis and metabolism of sanshools

Based on sanshool structures and the results of DEG functional classification and annotation, 12 candidate enzymes encoded by 18 unigenes, which may be associated with the biosynthesis and metabolism of sanshools, were identified. We classified these 18 DEGs into three groups (Fig. 8a). Group I

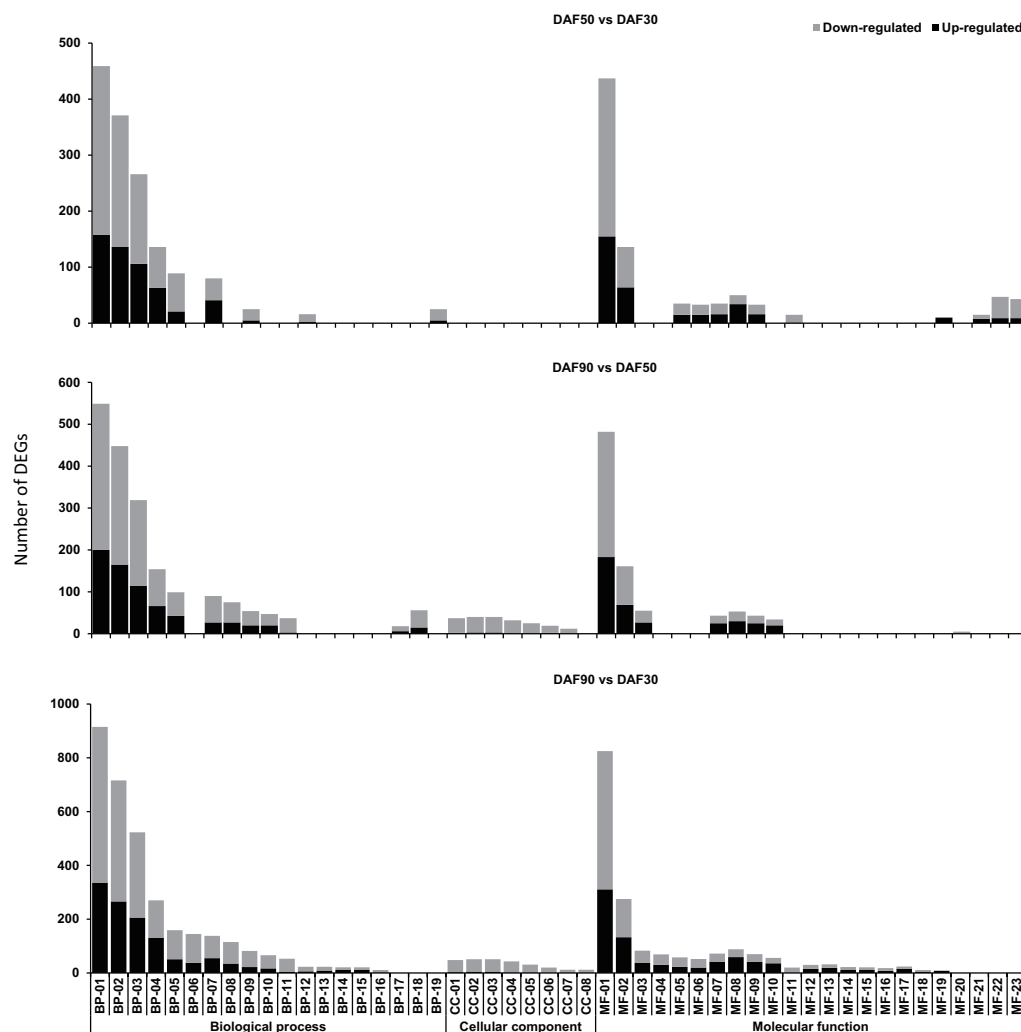


Fig. 4 GO classification of the unigenes with significant transcriptional differences during the three developmental stages of pericarp in *Z. bungeanum*. Subcategories are abbreviated. BP-01: metabolic process; BP-02: single-organism process; BP-03: single-organism metabolic process; BP-04: oxidation-reduction process; BP-05: carbohydrate metabolic process; BP-06: carbohydrate derivative metabolic process; BP-07: lipid metabolic process; BP-08: single-organism carbohydrate metabolic process; BP-09: cellular carbohydrate metabolic process; BP-10: polysaccharide metabolic process; BP-11: photosynthesis; BP-12: nucleotide-sugar metabolic process; BP-13: glyceraldehyde-3-phosphate metabolic process; BP-14: obsolete peroxidase reaction; BP-15: response to oxidative stress; BP-16: photosynthesis, light reaction; BP-17: carbohydrate catabolic process; BP-18: lipid biosynthetic process; BP-19: glucan metabolic process; CC-01: photosynthetic membrane; CC-02: thylakoid; CC-03: thylakoid part; CC-04: photosystem; CC-05: photosystem II; CC-06: thylakoid membrane; CC-07: photosystem II oxygen evolving complex; CC-08: photosystem I; MF-01: catalytic activity; MF-02: oxidoreductase activity; MF-03: cofactor binding; MF-04: coenzyme binding; MF-05: oxidoreductase activity, acting on CH-OH group of donors; MF-06: oxidoreductase activity, acting on the CH-OH group of donors, NAD or NADP as acceptor; MF-07: tetrapyrrole binding; MF-08: oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen; MF-09: heme binding; MF-10: iron ion binding; MF-11: copper ion binding; MF-12: antioxidant activity; MF-13: dioxygenase activity; MF-14: oxidoreductase activity, acting on peroxide as acceptor; MF-15: peroxidase activity; MF-16: monooxygenase activity; MF-17: oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen, 2-oxoglutarate as one donor, and incorporation of one atom each of oxygen into both donors; MF-18: nutrient reservoir activity; MF-19: oxidoreductase activity, acting on paired donors, with oxidation of a pair of donors resulting in the reduction of molecular oxygen to two molecules of water; MF-20: fructose 1,6-bisphosphate 1-phosphatase activity; MF-21: O-methyltransferase activity; MF-22: hydrolase activity, acting on glycosyl bonds; MF-23: hydrolase activity, hydrolysing O-glycosyl compounds.

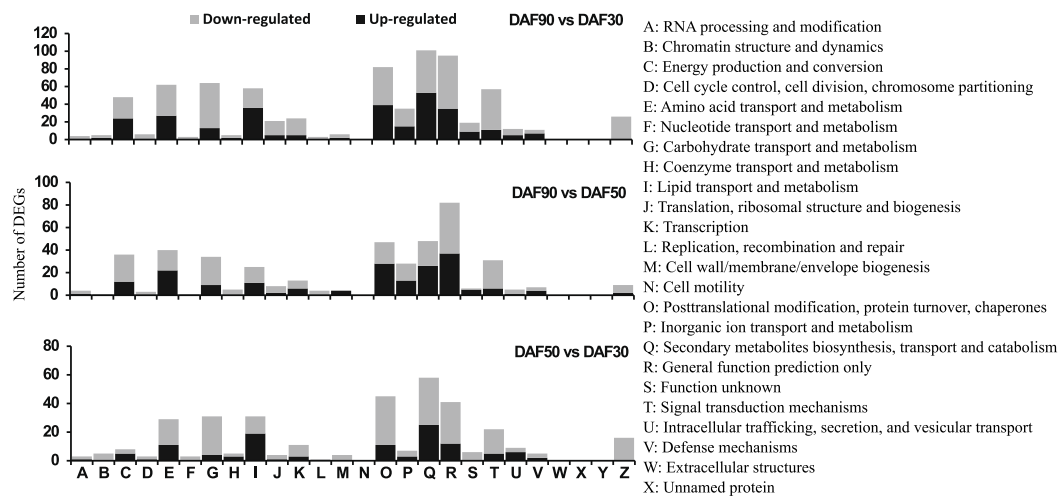


Fig. 5 Functional classification by KOG of differentially expressed genes.

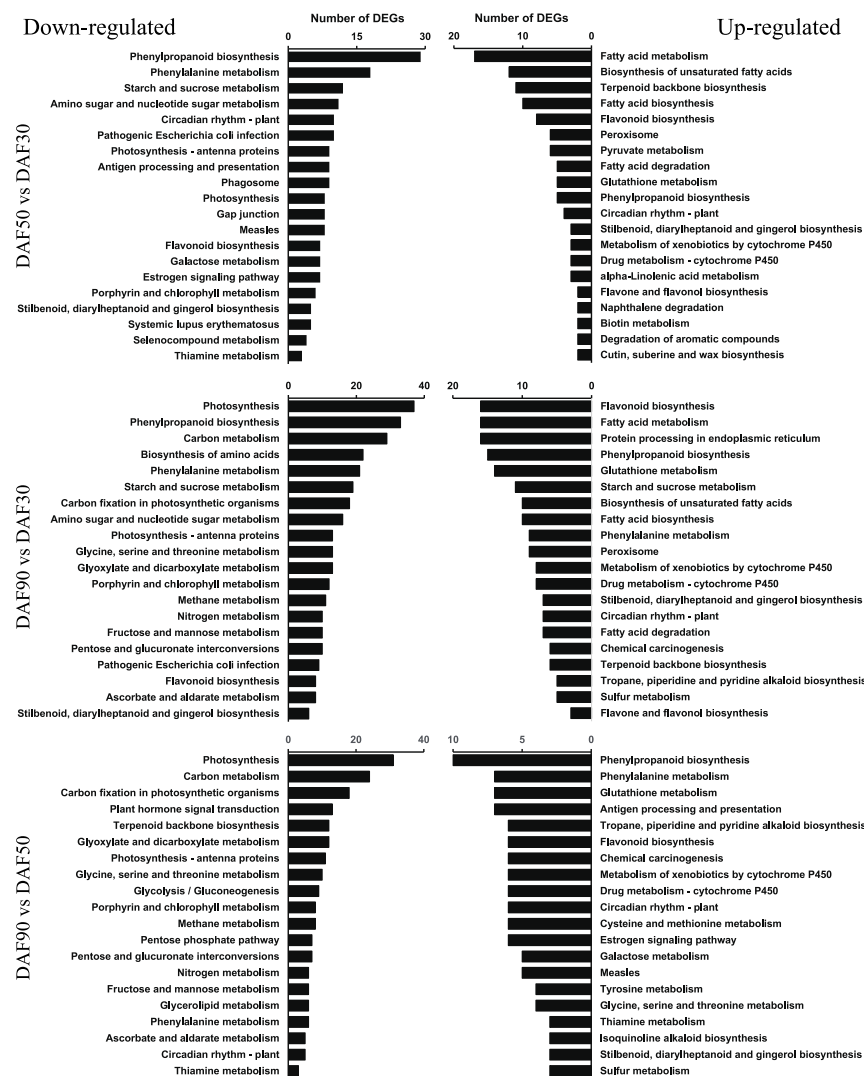


Fig. 6 KEGG pathway enrichment of upregulated and downregulated DEGs.

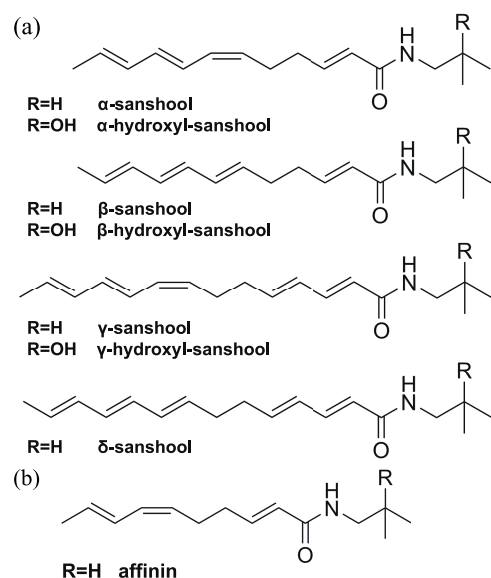


Fig. 7 (a) Structures of sanshool compounds; (b) structure of affinin.

contains the enzymes associated with the biosynthesis of fatty acid acyl-CoA, including acetyl-CoA carboxylase biotin carboxyl carrier protein (accB), malonyl-CoA-acyl carrier protein (ACP) transacylase (MAT), 3-oxoacyl-ACP synthase II (KASII), 3-oxoacyl-ACP reductase (KAR) acyl-ACP desaturase (DESA1), omega-6 fatty acid desaturase (FAD2), omega-3 fatty acid desaturase (FAD3), and acyl-CoA dehydrogenase (ACADM). AccB, MAT, KASII, and KAR have been identified as key enzymes involved in the elongation of fatty acid chains and DESA1, FAD2, FAD3, and ACADM are important in the desaturation of fatty acids³⁷. Because polyunsaturated acid is a precursor of sanshool biosynthesis¹¹ and because most of the unigenes in group I were upregulated at DAF50 and DAF90 compared with DAF30 and the sanshool content increases with pericarp development (the numbing and anaesthetic sensation was stronger during pericarp development), we inferred that these enzymes might play important roles in the biosynthesis of polyunsaturated alkyl chains of sanshools. Group II included three enzymes: branched-chain amino acid aminotransferase (BCAT), alanine-glyoxylate transaminase (AGXT), and aspartate aminotransferase (ASP). Previous research has suggested that valine is the precursor of the amide moiety in the biosynthesis of affinin, which is an alkamide with the same amide moiety and similar polyunsaturated acid chains as α -sanshool (Fig. 7b)³⁸. Because va-

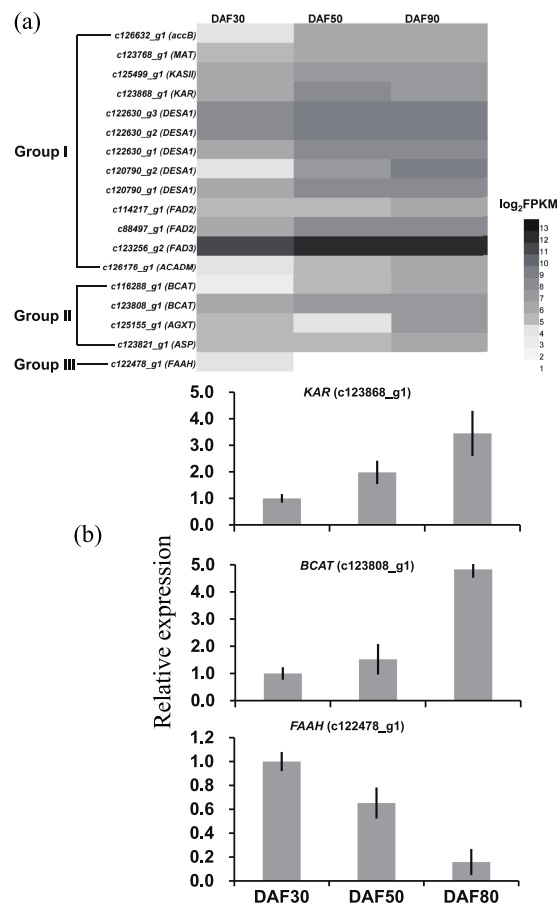


Fig. 8 (a) Heatmap of DEGs associated with biosynthesis and metabolism of sanshools among the three developmental stages of pericarp in *Z. bungeanum*. The gene name is described in parenthesis; (b) qRT-PCR validation of three unigenes involved in biosynthesis and metabolism of sanshools in *Z. bungeanum* pericarps.

line is one of the branched-chain amino acids and the expression levels of BCAT, AGXT, and ASP at DAF90 were significantly higher than DAF50 and DAF30, we hypothesise that BCAT, AGXT, and ASP are related to the biosynthesis of isobutylamide. In contrast to group I and II, group III included only one enzyme, fatty acid amide hydrolase (FAAH), which was down regulated during development after DAF30. The expression of FAAH was the inverse of the sanshool accumulation in this study. This finding sheds light on the participation of FAAH in the degradation of alkamide *in planta*³⁹. In addition, to confirm the reliability of the RNA sequencing (RNA-seq) results, three genes (KAR, BCAT, and FAAH) were selected from the three groups for quantitative real-time PCR (qRT-PCR) analysis (Table 4). As

shown in Fig. 8b, the relative expression profiles of the BCAT and FAAH in the three samples agreed with the expression patterns obtained using RNA-seq analysis, but KAR was not completely agreed.

Using comparative RNA-seq analysis, we first examined the gene expression patterns of pericarps of *Z. bungeanum* at three developmental stages. We identified 11 enzymes related to the biosynthesis of the polyunsaturated alkyl chain and isobutylamide moieties of sanshools and one enzyme involved in the degradation of sanshools. We believe these results will provide a valuable dataset to accelerate our understanding of the molecular mechanisms involved in sanshool biosynthesis and metabolism in *Z. bungeanum*.

Acknowledgements: This work was supported by Doctor Faculty Inaugurating Project of Northwest A&F University (Grant No. 2452015296) and Fundamental Research Funds for the Central Universities (Grant No. 2452016053).

REFERENCES

- Wang S, Xie JC, Yang W, Sun BG (2011) Preparative separation and purification of alkylamides from *Zanthoxylum bungeanum* maxim by high-speed counter-current chromatography. *J Liq Chromatogr Relat Technol* **34**, 2640–2652.
- Hatano T, Inada K, Ogawa TO, Ito H, Yoshida T (2004) Aliphatic acid amides of the fruits of *Zanthoxylum piperitum*. *Phytochemistry* **65**, 2599–2604.
- Ledoux A, Maraetefau H, Jansen O, Etienne D, Quetin-Leclercq J, Clerc P, Smadja J, Frederich M (2015) Phytochemical profile and biological activity evaluation of *Zanthoxylum heterophyllum* leaves against malaria. *Planta Med Lett* **2**, e10–e11.
- Xiang L, Liu Y, Xie C, Li X, Yu Y, Ye M, Chen S (2016) The chemical and genetic characteristics of Szechuan pepper (*Zanthoxylum bungeanum* and *Z. armatum*) cultivars and their suitable habitat. *Front Plant Sci* **7**, ID 467.
- Chen XL, Jin XK, Li Y, Chen GJ, Chen KW, Kan J (2018) Preparation and characterization of molecularly-imprinted polymers for extraction of sanshool acid amide compounds followed by their separation from pepper oil resin derived from Chinese prickly ash (*Zanthoxylum bungeanum*). *J Sep Sci* **41**, 590–601.
- You YM, Zhou M, Lu HJ, Shirima GG, Cheng YJ, Liu X (2015) Sanshool from *Zanthoxylum* L. induces apoptosis in human hepatocarcinoma HepG2 cells. *Food Sci Biotechnol* **24**, 2169–2175.
- Kubota K, Ohtake N, Ohbuchi K, Mase A, Imamura S, Sudo Y, Miyano K, Yamamoto M, et al (2015) Hydroxy- α sanshool induces colonic motor activity in rat proximal colon: a possible involvement of KCNK9. *Am J Physiol Gastrointest Liver Physiol* **308**, 579–590.
- Tang X, Zhou X, Wu J, Li J, Bai L (2014) A novel function of sanshools: the alleviation of injury from metolachlor in rice seedlings. *Pestic Biochem Physiol* **110**, 44–49.
- Wang Y, Li CH, Luo B, Sun YN, Kim YH, Wei AZ, Gao JM (2016) Isobutylhydroxyamides from *Zanthoxylum bungeanum* and their suppression of NO production. *Molecules* **21**, ID 1416.
- Tian JM, Wang Y, Xu YZ, Yu ZC, Wei AZ, Zhang WM, Gao JM (2016) Characterization of isobutylhydroxyamides with NGF-potentiating activity from *Zanthoxylum bungeanum*. *Bioorg Med Chem Lett* **26**, 338–342.
- Hao D, Wen X, Liu L, Wang L, Zhou XL, Li YM, Zeng X, He G, et al (2019) Sanshool improves UVB-induced skin photodamage by targeting JAK2/STAT3-dependent autophagy. *Cell Death Dis* **10**, ID 19.
- Li J, Zheng W, Long DK, Ding L, Gong AH, Xiao CH, Jiang WK, Liu XQ, et al (2017) *De novo* sequencing and assembly analysis of the *Pseudostellaria heterophylla* transcriptome. *Plos One* **12**, ID e0170134.
- Devi K, Mishra SK, Sahu J, Panda D, Modi MK, Sen P (2016) Genome wide transcriptome profiling reveals differential gene expression in secondary metabolite pathway of *Cymbopogon winterianus*. *Sci Rep* **6**, ID 21026.
- Shin AY, Kim YM, Koo N, Lee SM, Nahm S, Kwon SY (2017) Transcriptome analysis of the oriental melon (*Cucumis melo* L. var. *makuwa*) during fruit development. *PeerJ* **5**, ID e2834.
- Deng N, Chang E, Li MH, Jing JM, Yao XM, Bartish IV, Liu JF, Jing M, et al (2016) Transcriptome characterization of *Gnetum parvifolium* reveals candidate genes involved in important secondary metabolic pathways of flavonoids and stilbenoids. *Front Plant Sci* **7**, ID 174.
- Ma J, Kanakala S, He YH, Zhang JL, Zhong XL (2015) Transcriptome sequence analysis of an ornamental plant, *Ananas comosus* var. *bracteatus*, revealed the potential unigenes involved in terpenoid and phenylpropanoid biosynthesis. *Plos One* **10**, ID e0119153.
- Li CF, Zhu Y, Yu Y, Zhao QY, Wang SJ, Wang XC, Yao MZ, Luo D, et al (2015) Global transcriptome and gene regulation network for secondary metabolite biosynthesis of tea plant (*Camellia sinensis*). *BMC Genomics* **16**, ID 560.
- Li FF, Li ML, Cui J, Gao JM (2014) Conventional determination of the contents of numbing components (acylamides) in prickly ash. *Sci Silvae Sin* **50**, 121–126. [in Chinese]
- Li TC, Yang HY, Zhang W, Xu DF, Dong Q, Wang F, Lei Y, Lei YL, et al (2017) Comparative transcriptome analysis of root hairs proliferation induced by water deficiency in maize. *J Plant Biol* **60**, 26–34.

20. Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, et al (2011) Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnol* **29**, 644–652.
21. Li B, Dewey CN (2011) RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* **12**, ID 323.
22. Langmead B, Trapnell C, Pop M, Salzberg SL (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* **10**, ID R25.
23. Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, Salzberg SL, Wold BJ, et al (2010) Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nature Biotechnol* **28**, 511–515.
24. Wang LK, Feng ZX, Wang X, Wang XW, Zhang XG (2010) DEGseq: an R package for identifying differentially expressed genes from RNA-seq data. *Bioinformatics* **26**, 136–138.
25. Deng WK, Wang YB, Liu ZX, Cheng H, Xue Y (2014) HemI: a toolkit for illustrating heatmaps. *Plos One* **9**, ID e111988.
26. Young MD, Wakefield MJ, Smyth GK, Oshlack A (2010) Gene ontology analysis for RNA-seq: accounting for selection bias. *Genome Biol* **11**, ID R14.
27. Kanehisa M, Goto S, Sato Y, Kawashima M, Furumichi M, Mao T (2014) Data, information, knowledge and principle: back to metabolism in KEGG. *Nucleic Acids Res* **42**, 199–205.
28. Fei XT, Shi QQ, Yang TX, Fei ZX, Wei AZ (2018) Expression stabilities of ten candidate reference genes for RT-qPCR in *Zanthoxylum bungeanum* Maxim. *Molecules* **23**, ID 802.
29. Wang Z, Gerstein M, Snyder M (2009) RNA-Seq: a revolutionary tool for transcriptomics. *Nature Rev Genet* **10**, 57–63.
30. Rai A, Kamochi H, Suzuki H, Nakamura M, Takahashi H, Hatada T, Saito K, Yamazaki M (2017) *De novo* transcriptome assembly and characterization of nine tissues of *Lonicera japonica* identify potential candidate genes involved in chlorogenic acid, luteolosides, and secoiridoid biosynthesis pathways. *J Nat Med* **71**, 1–15.
31. Feng SJ, Zhao LL, Liu ZS, Liu YL, Yang TX, Wei AZ (2017) *De novo* transcriptome assembly of *Zanthoxylum bungeanum* using Illumina sequencing for evolutionary analysis and simple sequence repeat marker development. *Sci Rep* **7**, ID 16754.
32. Huang RM, Huang YJ, Sun ZC, Huang JQ, Wang ZJ (2017) Transcriptome analysis of genes involved in lipid biosynthesis in the developing embryo of pecan (*Carya illinoensis*). *J Agric Food Chem* **65**, 4223–4236.
33. Byng JW, Chase MW, Christenhusz MJM, Fay MF, Byng JW, Judd WS, Soltis DE, Mabberley DJ, et al (2016) An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG IV. *Bot J Linn Soc* **181**, 1–20.
34. Yang CQ, Fang X, Wu XM, Mao YB, Wang LJ, Chen XY (2012) Transcriptional regulation of plant secondary metabolism. *J Integr Plant Biol* **54**, 703–712.
35. Iwase A, Matsui K, Masaru OT (2009) Manipulation of plant metabolic pathways by transcription factors. *Plant Biotechnol* **26**, 29–38.
36. Mugnaini C, Corelli F (2016) Total synthesis of δ -sanshool and analogues thereof. *Synthesis* **48**, 2085–2092.
37. Li-Beisson Y, Shorrosh B, Beisson F, Andersson M, Arondel V, Bates P, Baud S, Bird D, et al (2010) Acyl-lipid metabolism. *Arabidopsis Book* **8**, ID e0133.
38. Cortezespinosa N, Aviñaverduzco JA, Ramírezchávez E, Molinatosres J, Ríoschávez P (2011) Valine and phenylalanine as precursors in the biosynthesis of alkamides in *Acmella radicans*. *Nat Prod Commun* **6**, 857–861.
39. Faure L, Cavazos R, Khan BR, Petros RA, Koulen P, Blancaflor EB, Chapman KD (2015) Effects of synthetic alkamides on *Arabidopsis* fatty acid amide hydrolase activity and plant development. *Phytochemistry* **110**, 58–71.