RNA sequencing on *Amomum villosum* Lour.-induced by MeJA identifies the genes of WRKY and terpene synthases involved in terpene biosynthesis

Journal:	<i>Genome</i>
Manuscript ID	gen-2017-0142.R1
Manuscript Type:	Article
Date Submitted by the Author:	12-Dec-2017
Complete List of Authors:	<p>He, Xueying; Guangzhou University of Chinese Medicine, Research Center of Chinese Herbal Resource Science and Engineering; Guangzhou University of Chinese Medicine, Key Laboratory of Chinese Medicinal Resource from Lingnan, Ministry of Education</p> <p>Wang, Huan; Guangzhou University of Chinese Medicine, Research Center of Chinese Herbal Resource Science and Engineering; Guangzhou University of Chinese Medicine, Key Laboratory of Chinese Medicinal Resource from Lingnan, Ministry of Education</p> <p>Yang, Jinfen; Guangzhou University of Chinese Medicine, Research Center of Chinese Herbal Resource Science and Engineering; Guangzhou University of Chinese Medicine, Key Laboratory of Chinese Medicinal Resource from Lingnan, Ministry of Education</p> <p>Deng, Ke; Guangzhou University of Chinese Medicine, Research Center of Chinese Herbal Resource Science and Engineering; Guangzhou University of Chinese Medicine, Key Laboratory of Chinese Medicinal Resource from Lingnan, Ministry of Education</p> <p>Wang, Teng; Guangzhou University of Chinese Medicine, Research Center of Chinese Herbal Resource Science and Engineering; Guangzhou University of Chinese Medicine, Key Laboratory of Chinese Medicinal Resource from Lingnan, Ministry of Education</p>
Is the invited manuscript for consideration in a Special Issue? :	N/A
Keyword:	RNA sequencing, WRKY transcription factors, terpene synthases, <i>Amomum villosum</i> Lour

1 **RNA sequencing on *Amomum villosum* Lour.-induced by MeJA identifies the genes of**
2 **WRKY and terpene synthases involved in terpene biosynthesis**

3 Xueying He^{1,2}, Huan Wang^{1,2}, Jinfen Yang^{1,2*}, Ke Deng^{1,2}, Teng Wang^{1,2}

4 ¹Guangzhou University of Chinese Medicine, Research Center of Chinese Herbal Resource
5 Science and Engineering, Guangzhou, Guangdong, 510006, China

6 ²Guangzhou University of Chinese Medicine, Key Laboratory of Chinese Medicinal
7 Resource from Lingnan, Ministry of Education, Guangzhou, Guangdong, 510006, China

8 **E-mail**

9 Xueying He: 893997154@qq.com

10 Huan Wang: 459345479@qq.com

11 Jinfen Yang: yangjf@gzucm.edu.cn

12 Ke Deng: 745074686@qq.com

13 Teng Wang: 2465785705@qq.com

14 ***Corresponding author**

15 Jinfen Yang

16 Guangzhou University of Chinese Medicine, Research Center of Chinese Herbal Resource
17 Science and Engineering, Guangzhou, Guangdong, 510006, China

18 Guangzhou University of Chinese Medicine, Key Laboratory of Chinese Medicinal Resource
19 from Lingnan, Ministry of Education, Guangzhou, Guangdong, 510006, China

20 Tel: +86-8620-39358331/39358066

21 Fax: +86-8620-39358066

22 E-mail: yangjf@gzucm.edu.cn; 223353985@qq.com

23

24

25 **ABSTRACT**

26 *Amomum villosum* Lour. is an important Chinese medicine that has diverse medicinal
27 functions, and mainly contains volatile terpenes. This study aimed to explore the WRKY
28 transcription factors (TFs) and terpene synthases (TPS) unigenes, which might be involved in
29 terpene biosynthesis in *A. villosum* for providing some new information on the regulation of
30 terpenes in plants. RNA sequencing of *A. villosum*-induced by methyl jasmonate (MeJA)
31 revealed WRKY family was the second biggest TF family in the transcriptome. Thirty-six
32 complete WRKY domain sequences were in response to MeJA. Further, six WRKY unigenes
33 were highly correlated with eight deduced TPS unigenes. Ultimately, we combined the
34 terpene abundance with the expression of candidate WRKYs and TPS unigenes to presume a
35 possible model wherein *AvWRKY61*, *AvWRKY28* and *AvWRKY40* might coordinately
36 trans-activate the *AvNeoD* promoter. We propose an approach to mine TFs unigenes might be
37 involved in terpenoid biosynthesis, and obtained four unigenes for further analyses.

38

39 **Keywords:** RNA sequencing; WRKY transcription factors; terpene synthases; *Amomum*
40 *villosum* Lour.

41

42 INTRODUCTION

43 Amomi Fructus (Sha ren) is an important traditional Chinese medicine that displays
44 diverse medicinal functions such as dissipating dampness, warming the spleen and preventing
45 miscarriage, among other properties embodied in the Chinese pharmacopeia (2010).
46 *Amomum villosum* Lour. is famous as the genuine (Daodi) medicinal resource of Amomi
47 Fructus, which is produced in Yangchun City, located in Guangdong Province. According to
48 the modern pharmacological studies, Amomi Fructus has the pharmacological effects of
49 antiulceration, antidiarrheal, accelerating gastric emptying and gastrointestinal propulsion, ect
50 (Mingfa and Yaqin 2013). Volatile terpenoid (i.e., monoterpene and sesquiterpene) is the main
51 medicinal ingredients in *A. villosum* including bornyl acetate, camphor, borneol, ect. It was
52 reported that bornyl acetate extracted from *A. villosum* had analgesic and anti-inflammatory
53 effects (Wu et al. 2004; Wu et al. 2005).

54 Terpenes, one of the major secondary metabolites in medicinal plants, have many volatile
55 representatives such as isoprenes (C₅), monoterpenes (C₁₀), sesquiterpenes (C₁₅), even
56 some diterpenes (C₂₀), and triterpenes (C₃₀) (Dudareva et al. 2004). Generally, in plants,
57 terpenoid biosynthesis proceeds via two pathways: 1) the cytosolic-located mevalonate (MVA)
58 pathway, and 2) the plastidial-located 2-C-methyl-D-erythritol-4-phosphate (MEP) pathway
59 (Rodriguez-Concepcion and Boronat 2002). In both pathways, geranyl pyrophosphate (GPP,
60 C₁₀) is generated, which is the precursor for monoterpene and sesquiterpene. The step
61 catalyzed by 1-deoxy-D-xylulose-5-phosphate reductoisomerase (DXR, EC: 1.1.1.267) is one
62 of the regulators in the MEP pathway. The 3-hydroxy-3-methylglutaryl coenzyme A (HMGR,
63 EC: 1.1.1.34) is also a rate-limiting enzyme in the MVA pathway. Both HMGR and DXR are
64 important enzymes upstream of terpenoid biosynthesis.

65 In a previous study, we cloned *AvHMGR* and *AvDXR* from *A. villosum* and over-expressed

66 them in transgenic tobacco and found that over-expression of *AvHMGR* or *AvDXR* promoted
67 some terpenoid biosynthesis, including cembrenene, neophytadiene, and sterol (Yang et al.
68 2012). Santalol (sesquiterpene) and m-Mentha-4, 8-diene (monoterpene) were detected in
69 *AvDXR* transgenic tobacco, but not in the wild-types (WT). The co-overexpression of
70 *AvHMGR* and *AvDXR* promotes the biosynthesis of sterol and phytol, but inhibits that of
71 neophytadiene (Huan et al. 2014). This observation indicates that overexpression of
72 *AvHMGR* and *AvDXR* promotes diverse effects in regulating the biosynthesis of different
73 terpenes. In the downstream pathway of terpene biosynthesis, terpene synthase (TPS)
74 catalyze the synthesis of a myriad of products including monoterpene, sesquiterpene, ect.
75 However, TPSs have seldom been studied (Chen et al. 2011).

76 Transcription factors (TFs), regulate a series of relative genes and have important
77 functions in plant development, evolution and responsiveness to abiotic and biotic stress.
78 Many TFs play essential roles in regulating secondary metabolite biosynthesis and
79 accumulation, such as terpenoids (Gantet and Memelink 2002; Vom Endt et al. 2002). TFs
80 usually activate or repress the promoters of TPS genes to control their expression, and then to
81 regulate terpenoid accumulation. As the “master switches” of transcriptional regulation, there
82 are only few TFs that are known to be involved in the regulation of terpenoid pathways.

83 Jasmonic acid (JA) signaling had regulated the survival of plants depends on their
84 abilities to quickly perceive and respond to external challenges (Balbi and Devoto 2008;
85 Farmer et al. 2003). Therefore, many investigators focus on the jasmonates (JAs) including
86 jasmonic acid (JA), its methyl ester (MeJA), its amino acid conjugates and other oxylipins
87 from the lipoxygenase pathway (Kombrink 2012), which are fatty acid-derived oxylipins
88 regulating many aspects of plant growth, development and defense. MeJA is a potent and
89 important elicitor of plant secondary metabolism including terpenes, which simultaneously

induces terpene biosynthetic genes as illustrated by the TIA and terpenoid pathways in *Catharanthus roseus* and *Solanum lycopersicum* (Spyropoulou et al. 2014; van der Fits and Memelink 2001). Therefore, some plant model experiments were induced by MeJA to exploit the ability of TFs to participate in terpenoid synthesis (Browse 2009; Zhao et al. 2005). There are five classes of TFs that are known to be involved in the regulation of terpenoid pathways, including WRKY (Skibbe et al. 2008), AP2/ERF(AP2) (Menke et al. 1999; Pauw et al. 2004), bHLH (Zhang et al. 2011), HAHB4 (Manavella et al. 2008) and the TFIIA zinc finger (Pauw et al. 2004).

The WRKY proteins are a superfamily of transcription factors with up to 100 representatives in Arabidopsis. WRKY family members involved in the regulation of various physiological programs including pathogen defense, senescence and trichome development (Eulgem et al. 2000). The WRKY transcription factors generally contained WRKYGQK conserved sequences at N-terminal, together with zinc-finger-like motif (Rushton et al. 1995). The cognate cis-acting W box elements, usually contained invariant TGAC core, is the DNA binding site (Fukuda and Shinshi 1994). WRKY TFs have emerged as a key family in terpene biosynthesis. Madagascar periwinkle (*Catharanthus roseus*) WRKY1 (*CrWRKY1*) may participate in the biosynthesis of terpenoid indole alkaloids (Suttipanta et al. 2011). Cotton (*Gossypium arboreum*) WRKY1 (*GaWRKY1*) regulates the (+)- δ -Cadinene Synthase-A (CAD1) gene to regulate the biosynthesis of gossypol (sesquiterpene) (Wu et al. 2004). Similarly, *Artemisia annua* WRKY1 (*AaWRKY1*) affects the amorpho-4, 11-diene synthase (ADS) gene to control artemisinin (sesquiterpene) biosynthesis (Ma et al. 2009). In addition, silencing of two insect responsive genes (but not JA-responsive genes) of WRKY (i.e., WRKY3 and WRKY6) from the native tobacco *N. attenuata*, make plants highly vulnerable to herbivores by impairing JA and accumulating cis- α -bergamotene

(sesquiterpene). In the biosynthesis of diterpene, the *Taxus chinensis* WRKY1 (*TcWRKY1*) was found to regulate 10-DeacetylbaocatinIII-10 β -O-Acetyl transferase (DBAT) gene expression to affect paclitaxel (diterpene) biosynthesis (Li et al. 2013). American ginseng (*Panax quinquefolius*) WRKY1 (*PqWRKY1*) is over-expressed in the transgenic Arabidopsis, suggesting that *PqWRKY1* regulates the biosynthesis of ginsenoside (triterpene). There were some reports that described the involvement of WRKY and other transcription factors in terpene biosynthesis; however, the important WRKY TFs that are known to be involved in terpene biosynthesis in *A. villosum* is relatively less well-studied.

Although there are many chemical and pharmacological studies for *A. villosum*, molecular genetic research of the transcriptionally regulated genes that are involved in volatile terpenoid biosynthesis of *A. villosum* remains rare. Therefore, we intended to discover the volatile terpene synthases and its regulator (WRKY TFs) by RNA sequencing technology and metabonomics analysis. These works might demonstrate their utility in enhancing the medicinal quality of *A. villosum*.

MATERIALS AND METHODS

Plant material and hormone treatment

A. villosum was planted in Panlong Town of Yangchun City, Guangdong Province, China. The leaves and ripe fruits of the healthy plants were selected to spray with 0.02% tween 80 as solvent control, 200 μ mol/L and 600 μ mol/L MeJA in August (Table 1). There were three biological replicates and five technical replicates for each sample. After the spraying, the leaves and ripe fruits were packed with plastic wrap immediately and samples were collected 24 hours later. Two types of organs including seeds and peels were collected separately and immediately frozen in liquid nitrogen and then stored in -80°C. The materials

were treated for RNA Sequencing and real-time fluorescence quantitative PCR (RT-qPCR).

cDNA library construction and sequencing

The total RNA of each sample was isolated using the Trizol Kit (Promega, USA) following by the manufacturer's instructions. Then the total RNA was treated with RNase-free DNase I (Takara Bio, Japan) for 30 min at 37°C to remove residual DNA. RNA quality was verified using a 2100 Bio-analyzer (Agilent Technologies, Santa Clara, CA) and were also checked by RNase free agarose gel electrophoresis. Only with OD₂₆₀/OD₂₈₀ at 1.8-2.2, RNA was used for further analysis. Next, poly (A) mRNA was isolated using oligodT beads (Qiagen). All mRNA were broken into short fragments by adding fragmentation buffer. First-strand cDNA was generated using random hexamer-primed reverse transcription, followed by the synthesis of the second-strand cDNA using RNase H and DNA polymerase I. The cDNA fragments were purified using a QIA quick PCR extraction kit. These purified fragments were then washed with EB buffer for end reparation poly (A) addition and ligated to sequencing adapters. Following agarose gel electrophoresis and extraction of cDNA from gels, the cDNA fragments were purified and enriched by PCR to construct the final cDNA library. 5µg cDNA was used for cDNA library construction. The cDNA library was sequenced on the Illumina sequencing platform (Illumina HiSeq™ 2000) using the paired-end technology by Gene Denovo Co. (Guangzhou, China). A Perl program was written to select clean reads by removing low quality sequences (there were more than 50% bases with quality lower than 20 in one sequence), reads with more than 5% N bases (bases unknown) and reads containing adaptor sequences.

Reads alignment and Normalization of gene expression levels

Sequencing reads were mapped to reference sequence by the SOA Paligner/soap2 (Li et al. 2009), a tool designed for short sequences alignment. Coverage of reads in one gene was

used to calculate expression level of this gene. Using this method we obtained the expression levels of all genes detected.

Reads that could be uniquely mapped to a gene were used to calculate the expression level. The gene expression level was measured by the number of uniquely mapped reads per kilobase of exon region per million mappable reads (RPKM). The formula was defined as below:

$$RPKM = \frac{10^6 C}{NL / 10^3}$$

In which C was the number of reads uniquely mapped to the given gene; N was the number of reads uniquely mapped to all genes; L was the total length of exons from the given gene. For genes with more than one alternative transcript, the longest transcript was selected to calculate the RPKM. The RPKM method eliminates the influence of different gene length and sequencing discrepancies on the gene expression calculation. Therefore, the RPKM value can be directly used for comparing the differences in gene expression among samples. All expression data statistic and visualization was conducted with R package (<http://www.r-project.org/>).

Differentially expressed genes (DEGs) and function enrichment analyses

After the expression level of each gene was calculated, differential expression analysis was conducted using edge R (Robinson et al. 2010). The false discovery rate (FDR) was used to determine the threshold of the p value in multiple tests, and for the analysis, a threshold of the $FDR \leq 0.01$ and an absolute value of $\log_2 \text{Ratio} \geq 1$ were used to judge the significance of the gene expression differences. The differentially expressed genes were used for GO and KEGG enrichment analyses according to a method similar to that described by Zhang (Zhang et al. 2013). Both GO terms and KEGG pathways with a Q-value ≤ 0.05 are significantly

185 enriched in DEGs.

186 WRKY TFs unigenes were translated its coding sequences into protein sequences. A fasta
187 file of the candidate WRKY TFs protein sequences were submitted to the NCBI Conserved
188 Domain Database (CDD) and the Samuel Roberts Nobel Foundation PlantTFcat (PlantTFcat)
189 server to identify the WRKY domains containing sequences. Then, the selected sequences
190 were submitted to the NCBI CDD and PlantTFcat servers again to confirm they all have
191 complete WRKY domains.

192 **Volatile compounds detection by gas chromatography-mass spectrometry**

193 The concentrations of volatile terpenes in two tissues of *A. villosum* were determined
194 based on gas chromatography-mass spectrometry (GC-MS). The volatile terpenes in the fresh
195 tissues of *A. villosum* were extracted by microwave method. The samples(3.00g, n=2) were
196 subjected to microwave extraction (280W) with 30mL petroleum ether (Kermel, Tianjin,
197 China) for 30 minutes using MAS-II atmospheric pressure microwave synthesis/extraction
198 workstation (Sineo, Shanghai, China), and filtered through a microfiltration membrane (0.22
199 μm).

200 Extracted metabolites were analyzed as follows: 1 μl of sample was injected at a split
201 ratio of 10:1 into a HP6890/5973 GC/MS (Agilent, USA). DB-FFAP capillary column (30 m
202 $\times$ 0.25 mm $\times$ 0.25 μm) was employed for separation. Injection temperature was 250°C and the
203 interface temperature was set to 280°C. The ion source was adjusted to 230°C and the solvent
204 cut-time was set to 1 minutes. Helium was the carrier gas at a flow-rate of 0.7 ml per minute.
205 The temperature program set an initial temperature of 50°C, programmed at 10°C per
206 minutes to 100°C and held for 1 minute. Then ramped at 20°C per minute to 220°C and held
207 for 13 minutes. The mass spectrometric detector operated in the electron impact ionization
208 mode with an ionizing energy of 70eV, scanning from 29.0-500.0 amu.

Peak identification was performed by employing WILEY7n.L (Palisade Corporation, NY, USA) and Nist08.L (NIST, Gaithersburg, MD, USA) databases. With limonene as the external standard to calculate the terpenes concentrations, it was calculated as following: $C_{\text{sample}} = A_{\text{sample}} \times C_{\text{standard}} / A_{\text{standard}}$ (C_{sample} was the sample concentration; C_{standard} was the external standard concentration; A_{sample} was the sample area integration; A_{standard} was the external standard area integration). Peaks were quantified by area integration. Concentrations were normalized to the quantity of the external standard.

RNA isolation, cDNA synthesis and RT-qPCR

The fruits from the RNA-Seq materials were used as RT-qPCR materials. Total RNA was isolated from the peels using the EASY spin Plant RNA Kit (Aidlab, Beijing, China). Total RNA was extracted from the seeds using an improved CTAB procedure. RNA concentration was measured by the TGen Spectrophotometer. Then, the OD₂₆₀/OD₂₈₀ ratios of all samples ranged from 1.8 to 2.2. The integrity of RNA samples were assessed with a 2% Biowest Agarose (GENE Tech, Shanghai, China), and no sign of degradation was found. Then, cDNA was removed from RNA with gDNase and was synthesized using FastQuant RT Kit (with gDNase) (TIANGEN, Beijing, China). For RT-qPCR, cDNA equivalent to 500ng total RNA was used as a template in 20μl volume. The reactions were performed in the CFX96 Real-Time PCR System (Bio-Rad, Alfred Nobel Drive, USA) using the SsoFast EvaGreen Supermix (Bio-Rad, Alfred Nobel Drive, USA). Two microliters (equivalent to 50ng total RNA) of cDNA were then used for quantitative RT-PCR. There were three biological replicates and three technical replicates for each target gene. Forward and reverse primers were given in App Table 1 and 2. Primer pairs were tested for amplification kinetics and linearity with a standard cDNA dilution curve and new primers were designed if necessary. Expression levels were normalized using *Actin* (Unigene0133538) and *TUA*

(Unigene0093134) mRNA levels by $2^{-\Delta\Delta C_t}$. The data analysis was performed by Bio-Rad CFX Manager. The RT-qPCR conditions: 1) 95.0°C for 0:30. 2) 95.0°C for 0:05. 3) 60.0°C for 0:30. 4) 72.0°C for 0:45, plate read. 5) GOTO 2, 39 more times. 6) 95.0°C for 0:10. 7) Melt curve 65.0°C to 95.0°C, increment 0.5°C 0:05 and plate read.

Genes cluster analysis

We perform cluster analysis of gene expression patterns using “heatmap.2” function in gplots package of R (<http://cran.r-project.org/web/packages/gplots/index.html>), and the result was visualized as heatmap. Each column represents an experimental sample, and each row represents a gene. Expression differences are shown in different colors. Red means high expression and green means low expression.

Statistical analyses

The correlation and One-way ANOVA for gene expression levels between corresponding organs from solvent control, and 200 µmol/L to 600 µmol/L MeJA treatment were analyzed by SPSS 17.0 statistics software. $P < 0.05$ means the difference was statistically significant. The Pearson coefficient threshold was 0.8. We construct the networks of the gene pairs when their Pearson correlation coefficient was more than $|0.8|$. Graphical representations of gene networks were produced by Cytoscape 3.2.1.

250

RESULTS

Transcriptome sequencing and assembly results

cDNAs prepared from the leaves, peels and seeds of *A. villosum* were sequenced using Illumina HiSeq™ 2000 platform. As a result of sequencing, 297,252,674 clean reads of 29,725,267,400 nucleotides (nt) were obtained from the transcriptome. The Q20 and GC percentages were 97.76% and 50.72%, respectively. *De novo* assembly produced 146,543

contings of 125,717,048 nucleotides (nt) and the average of these cotings were 857bp, with an N50 of 1397 and 43% GC percentages. Further assembly of the cotings generated 138,679 unigenes and the overall length was 11,684,927nt. The N50 and GC percentages of unigenes were 1381bp and 43.82%.

The 70,323 unigenes which accounted for 50.71% were successfully matched with at least one biological term among Nr, SwissProt, KEGG, COG and GO databases. Among annotated, 13,046 unigenes were completely obtained the biological term from Nr, SwissProt, KEGG and COG databases and the unigenes annotated from individual database were 69,678 (99.08 %) unigenes with Nr, 53,888 (76.63 %) with Swissport, 27,382 (38.94 %) with COG term and 21,503 (30.58%) with KEGG pathway, respectively. Among them, 69,678 (77.07 %) unigenes were obtained with the e-value less than 10^{-20} in Nr database. Also, more than 2400 unigenes of mapped *A. villosum* transcripts shared annotation information from the six major plant species, i.e. *Oryza sativa Japonica Group*, *Setaria italica*, *Vitis vinifera*, *Theobroma cacao*, *Brachypodium distachyon* and *Zea mays*.

RNA-seq data for differentially expressed genes of transcription factors in *A. villosum*

In this study, we used RNA sequencing of *A. villosum* following treatment with MeJA as a tool for genetic analysis (Table 1). We obtained 138,679 unigenes with an average length of 842 bp, and 70,323 annotated unigenes. In the transcriptome, there were 58,628 unigenes that were classified into 25 COG functional categories. The largest functional category was general functional prediction (8,576 unigenes), followed by transcription (6,257 unigenes).

Next, gene ontology (GO) was performed, which classified all unigenes into three classes: 1) biological processes, 2) cellular components and 3) molecular functions. The highest percentage of cellular component GO terms was the cell (30.56%). Catalytic activity and binding accounted for the highest proportion in the molecular function GO terms (46.50%

281 and 42.61%, respectively). Metabolic process and cellular process were respectively
282 comprised of 25.21% and 24.46% in biological processes, which indicated a large of
283 unigenes were involved in metabolic process or cellular process in the cell with catalytic
284 activity and binding function.

285 In differentially expressed genes, a total of 2,489 unigenes were annotated as transcription
286 factors (Figure1). The largest transcription factor family was the MYB family (331 unigenes),
287 followed by the WRKY family (283 unigenes) and the basic helix-loop-helix (bHLH) family
288 (267 unigenes). It is reported that the WRKY transcription factors regulate secondary
289 metabolite accumulation. For example, in cotton (*Gossypium arboreum*), *GaWRKY1*, which
290 trans-activates the promoter of the (+)- δ -cadinene synthase gene (CAD1), participates in the
291 regulation of phytoalexin (sesquiterpene) biosynthesis (Xu et al. 2004); Suttipanta *et al.*
292 identified that *Madagascar periwinkle* WRKY1 participated in the regulation of terpenoid
293 indole alkaloid biosynthesis through an undefined pathway (Suttipanta et al. 2011). These
294 results provided detail information on the TFs that were elicited by MeJA in *A. villosum*.

295 **Selection of the WRKY TFs in response to MeJA in *A. villosum***

296 We further analyzed the expression of unigenes in each sample. By this analysis, we
297 found at least 193 unigenes annotated as the WRKY transcription factors in each comparison
298 (Figure2A). Since MeJA plays a crucial role in direct and indirect plant defense, we were
299 interested in the WRKY transcriptional factor response to MeJA. Therefore, we screened 113
300 annotated WRKY TF unigenes that were up-regulated more than 2-fold (>2-fold) in seven
301 comparisons between FSP versus FM1P, FSP versus FM2P, FSS versus FM1S, FSS versus
302 FM2S, L0L versus LSL, L0L versus LM1L, and LSL versus LM1L (Figure2B). Fifty
303 differentially expressed genes (i.e., with a differential expression of more than 2-fold)
304 annotated WRKY TFs in response to MeJA (Figure2C), including 40 WRKY TF unigenes

305 that were up-regulated (>2 -fold) following treatment with MeJA, ten that were
306 down-regulated (< -2 -fold; Figure2D). We screened 163 unigenes that annotated WRKY by
307 the above described two methods. A total of 86 unigenes that annotated WRKY remained for
308 further study after removing the redundant unigenes.

309 Previously, 86 differentially expressed genes that were annotated as WRKY TFs that were
310 responsive to MeJA all had potential coding sequences (CDS). To further validate the
311 unigenes of WRKY TFs, a list of the single longest deduced proteins for each locus was
312 submitted to the NCBI CDD and PlantTFcat servers for conserved domain identification
313 (Marchler-Bauer et al. 2011). The NCBI CDD identified 60 WRKY domain-containing
314 sequences (App Table 3). Ten deduced amino acid sequences of WRKY TFs were identified
315 as having incomplete N-terminal ends and another 14 WRKY TF sequences had incomplete
316 C-terminal portions of the WRKY domain by NCBI CDD. In addition, PlantTFcat
317 (<http://plantgrn.noble.org/PlantTFcat/>) identified 53 WRKY domain-containing sequences,
318 which were also identified by NCBI CDD as WRKY domain-containing sequences (App
319 Table 3). In total, 36 sequences that contained complete WRKY domains in *A. villosum*, were
320 selected for further analysis (App Fig 1).

321 **Selection of terpene synthases genes in *A. villosum***

322 In our study, 21,503 unigenes were assigned to 125 KEGG pathways. A maximum
323 number of unigenes were involved in metabolic pathways (5,353 members, 24.89%), which
324 was followed by the biosynthesis of secondary metabolite pathways (2,459 members,
325 11.44%). These pathways provided valuable resources to study key genes that are involved in
326 terpenoid biosynthesis in *A. villosum*.

327 Importantly, 387 unigenes (1.8%) were involved in five terpenoid biosynthesis related
328 pathways, including terpenoid backbone biosynthesis, monoterpenoid biosynthesis,

329 sesquiterpenoid biosynthesis, diterpenoid biosynthesis and ubiquinone and other
330 terpenoid-quinone biosynthetic pathways (Table 2). A closer look was taken at those five
331 pathways in order to find important enzymes that were involved in the biosynthesis of
332 volatile terpene (monoterpene and sesquiterpene). Moreover, key enzymes and transcription
333 factors that regulated the biosynthesis of volatile terpenes in the transcript levels. Therefore,
334 we screened terpene synthases genes in our data (Table 3), and found eight unigenes that
335 were annotated as terpene synthases including myrcene synthase (EC: 4.2.3.15),
336 (+)-neomenthol dehydrogenase (EC: 1.1.1.208), linalool synthase, S-(+)-linalool synthase
337 (EC: 4.2.3.25), 3S, 6E-nerolidol synthase (EC: 4.2.3.28), (+)-germacrene D synthase (EC:
338 4.2.3.75), sesquiterpene synthase 3 and sesquiterpene synthase A1. These eight unigenes were
339 named according to their sequence identities and subjected to further analysis.

340 **The terpene synthase genes correlated with the sequences of WRKY putatively involved**
341 **in terpene biosynthesis of *A. villosum***

342 Transcription factors usually activate the promoter of terpene synthase to regulate the
343 biosynthesis of terpene. We previously reported 36 complete WRKY domain sequences that
344 were induced by MeJA and eight terpene synthases genes (monoterpene synthases and
345 sesquiterpene synthases). Then we performed gene co-expression network analysis between
346 36 WRKY domain unigenes and the selected unigenes of terpene synthase (TPS) to narrow
347 down the numbers of unigenes. Pearson's correlation of the normalized signal intensities was
348 calculated for the 36 WRKY and the terpene synthase genes. As the threshold was 0.8, the
349 relationship between WRKYs and the terpene synthases would be shown when the Pearson
350 correlation coefficient was more than | 0.8 |.

351 The results showed that only six sequences of WRKY were closely related to the
352 selected eight sequences of TPS (Figure 3). A number of the WRKY genes co-expressed with

TPS genes; for example, WRKY61 (Unigene0037009) and WRKY28 (Unigene0044272) all showed sound correlation with the 3S, 6E-nerolidol synthase (Unigene0060131), and myrcene synthase (Unigene0099398). And WRKY40 (Unigene0102915) showed correlation with (+)-germacrene D synthase (Unigene0106615), linalool synthase (Unigene0136809), sesquiterpene synthase 3 (Unigene0106619), and sesquiterpene synthase A1 (Unigene0140412). By contrast, WRKY31 was negatively correlated with the (+)-neomenthol dehydrogenase (Unigene0078114). WRKY45 was negatively correlated with sesquiterpene synthase 3 (Unigene0106619) and sesquiterpene synthase A1 (Unigene0140412). This phenomenon suggests that the six WRKY unigenes mentioned above might play important roles in controlling terpene synthase expression to affect terpene biosynthesis (Table 4). Thus, those six WRKY genes were inferred as the candidate WRKY genes for further analysis.

Integrated analysis of volatile compounds by GC-MS and gene expression

To further comprehend the mechanism for synthesizing terpene in *A. villosum*, we detected terpene abundances in the peels and seeds. In this study, 33 terpenes were detected in the seeds, while 20 terpenes were detected in the peels. We found that bornyl acetate was richest in the seeds, followed by camphor and borneol (App Table 4). Peels contained plenty of linalool, bornyl acetate and beta-pinene (App Table 5). It indicated that bornyl acetate was abundant in the fruit of *A. villosum*.

Using correlation network analysis to simulate the relationship between the selected terpene synthase genes and the volatile terpenes, the distinct correlation of TPS and terpenes (Pearson coefficient $> |0.8|$) was shown in Figure 4. The Spearman correlation coefficients between eight TPS genes and 33 terpenes were showed in App Table 6. The results showed that only four TPS unigenes were highly correlated with the volatile terpene compounds. A

number of TPS unigenes were negatively associated to a large proportion of terpenes, for example, (+)-germacrene D synthase (Unigene0106615) was negatively related with 29 volatile compounds (correlation coefficients ranged from -0.95 to -0.99). Bornyl acetate, being rich in *A. villosum*, was reversely correlated with (+)-germacrene D synthase (Unigene0106615; $r = -0.96$). In particular, (+)-germacrene D synthase (Unigene0106615) and its relative terpenes (germacrene A and bicyclogermacrene) were inversely expressed in response to MeJA (correlation coefficient of -0.97). Both (+)-neomenthol dehydrogenase (Unigene0078114) and (+)-germacrene D synthase (Unigene0106615) were negatively associated to the same 27 kinds of terpenes. The most negative relevant terpene of (+)-neomenthol dehydrogenase (Unigene0078114) and (+)-germacrene D synthase (Unigene0106615) was 4-terpineol (correlation coefficient of -0.94 and -0.99 respectively).

By contrast, the S-(+)-linalool synthase gene (Unigene0060132) was co-expressed only with sabinene ($r = 0.82$). Myrcene synthase (Unigene0099398) was positively associated with 17 terpenes including camphene, myrcene, and limonene, among others, with a correlation coefficient that ranged from 0.80 to 0.93. In addition, myrcene was strongly correlated with myrcene synthase (Unigene0099398; $r = 0.90$). The above described observations suggested that there is intrinsic substrate promiscuity and biosynthetic diversity of TPSs to biosynthesize such variable products. Ultimately, we found (+)-germacrene D synthase (Unigene0106615), linalool synthase (Unigene0060132), myrcene synthase (Unigene0099398) and (+)-neomenthol dehydrogenase (Unigene0078114) were candidate genes of TPS (Figure 4).

To validate the screened WRKY genes participating in the biosynthesis of terpenes, we also established the correlation network for the sequences of WRKY and the terpenes. As the results show in Figure 5, six candidate sequences of WRKY were all closely related with their

relative terpenes, which further confirmed that six candidate WRKY genes were indeed involved in the synthesis of terpenes. (The Spearman correlation coefficients between six WRKY genes and 33 terpenes were showed in App Table 7.)

Expression pattern and predicted model of candidate genes induced by MeJA.

To validate the above candidate gene transcript levels, we determined the expression of the selected WRKY and TPS genes by RT-qPCR (App Table 8-10). There were five sequences of candidate WRKY transcription factors and two selected sequences of terpene synthases presented. The remaining candidate genes (i.e. Unigene0099398, Unigene0060132 and Unigene0008375) were hardly expressed in the pericarps or seeds, and as such, they would be excluded from further analysis. Next, by comparing our RT-qPCR results and the RNA-Seq data, we found that the expression trends of five candidate genes (71%) were similar by both methods. However, *AvGerD* was down-regulated in the FSS by RT-qPCR as well as the expression in the RNA-Seq data (App Table 11). All in all, RNA-Seq assisted us in our ability to screen the key genes, while RT-qPCR data could accurately inspect gene expression.

If the genes expressed similar patterns, they could be aggregated to a gene expression cluster. One cluster of genes always performs the same function (Eisen et al. 1998), including some secondary metabolite biosynthesis functions (Vanderauwera et al. 2005). Unsupervised agglomerative hierarchical clustering for the above seven candidate gene expressions from RT-qPCR data divided eight samples into two laterally primary clusters; i.e., seeds and peels respectively (Figure 6). Both clusters indicated that the differences found in the tissues were greater than was found for MeJA treatment. Moreover, seven candidate genes were present in three primary clusters. Cluster one included *AvGerD*, which was up-regulated in both seeds and peels that were treated with MeJA and were down-regulated in the solvent control. This

observation suggests that *AvGerD* might be a MeJA-inducible gene. *AvWRKY45*, *AvNeoD*, *AvWRKY40*, *AvWRKY61*, *AvWRKY31* and *AvWRKY28* were present in the second cluster. The genes of the second cluster were up-regulated in some peels, including FSP and FM2P. However, they were down-regulated in the seeds. In particular, *AvWRKY45* was down-regulated in the seeds with the exception of FM2S. Both *AvWRKY61* and *AvWRKY40*, which was a branch of the second cluster, were down-regulated following treatment of the peels with increasing concentrations of MeJA. The second cluster genes expressed similar patterns, and they might synergistically be involved in the same pathway.

Consequently, we used correlation network analysis for the complex data to predict a model of *AvWRKYs*, *AvTPS* and the end-products (i.e., terpenes; Figure7). Three unigenes deduced that the *WRKYs* (i.e., *AvWRKY61*, *AvWRKY28*, and *AvWRKY40*) were positively correlated with *AvNeoD*. Then *AvNeoD* was positively correlated with sabinene and beta-pinene, but negatively correlated with four terpenes including alpha-cedrene, 4-terpineol, isoborneol and gamma-cadinene. Furthermore, *AvWRKY40* negatively correlated with 26 volatile terpenes, like bornyl acetate ($r = -0.881$). We assumed that *AvWRKY40* might repress some other terpene synthases to accumulate relatively volatile compounds. In addition, *AvWRKY28* interacted with *AvWRKY31*, *AvWRKY61* and *AvWRKY40*. We noted that *AvWRKY40*, *AvWRKY28* and *AvNeoD* negatively correlated and did so synergistically with the minimal volatile compound (i.e., 4-terpineol). We inferred that *AvWRKY61*, *AvWRKY28* and *AvWRKY40* trans-activated the promoter of *AvNeoD* coordinately to biosynthesize both sabinene and beta-pinene, which implied that *AvWRKY61*, *AvWRKY28*, *AvWRKY40* and *AvNeoD* played important roles in the biosynthesis of monoterpenes.

DISCUSSION

Based on the traditional theory of Chinese medicine, the medicinal ingredient of *A. villosum* is volatile terpene (monoterpene and sesquiterpene) but not separate monomers (2010; Skibbe et al. 2008). To improve the medicinal quality of *A. villosum*, manipulating key transcription factors and downstream key enzymes is a favorable strategy to increase the levels of volatile terpenoid. Therefore, we analyzed the differentially expressed genes that were induced by MeJA to mine the key genes of transcription factors and terpene synthases by high-throughput sequencing.

Identifying WRKY TFs involved in terpene synthesis

Jasmonic acid methyl ester (MeJA) is known as a common elicitor of plant secondary metabolism, and thus we tried to use MeJA to initiate an extensive transcriptional reprogramming of secondary metabolism (De Geyter et al. 2012). According to the literature, WRKY TFs can regulate terpene biosynthesis, including terpenoid indole alkaloids, sesquiterpene, diterpene and triterpene (Li et al. 2013; Ma et al. 2009; Skibbe et al. 2008; Suttipanta et al. 2011; Xu et al. 2004; Yongzhen et al. 2013). Therefore, we anchored WRKY TFs to further analyze whether WRKYs are involved in terpene biosynthesis. We established transcriptome database using Illumina by denovo in order to obtain JA-inducible genes. As a general lack of a systematic approach to screen WRKY genes is involved in terpenoid synthesis, we combined RNA-Seq technology and metabonomics to screen the key nucleotide sequences of WRKYs. Since some WRKY TFs trans-activate the promoter of terpene synthases to regulate terpene biosynthesis (Ma et al. 2009; Spyropoulou et al. 2014; Suttipanta et al. 2011; Xu et al. 2004), we used a gene co-expression network analysis to narrow down the key genes of WRKY TFs (Table 4).

Rice is a monocot model plant. In rice (*Oryza sativa*), JA-responsive WRKY TFs regulate the accumulation of lignin and other phenolics (Wang et al. 2007). Thus, we used

correlation network analysis to analyze the relevance between WRKY TFs and volatiles. Consequently, all candidate WRKY TFs are closely related to the terpenes. This demonstrates that JA-responsive WRKY TFs in *A. villosum* might regulate the accumulation of its corresponding secondary metabolites like terpene, which is concordant with Schluttenhofer's report (Schluttenhofer et al. 2014). Our studies thus offer an efficient strategy to screen key transcription factor genes by correlation network analysis.

Identifying terpene synthase genes that are involved in terpene synthesis

The wide diversity of volatile terpenes in plants is generated by the action of terpene synthases (TPSs). Many TPSs synthesize multiple products from prenyl diphosphate substrates (Degenhardt and Gershenzon 2000). Our studies found that sesquiterpene synthase 3 and sesquiterpene synthase A1 were only expressed in the solvent control (RPKM Value = 0.21) and (+)-germacrene D synthase was expressed at low levels in every sample (RPKM<2.33). Similarly, eight terpene synthase genes including β -caryophyllene synthase, α -terpineol synthase, α -farnesene synthase, β -ocimene synthases, α -humulene synthases, α -bergamotene synthases, germacrene-D synthase and α -pinene synthase showed very low or even no expression in the berry of *Vitis vinifera* (Matarese et al. 2014). This phenomenon is consistent with our data, which suggests that terpene synthase genes might be generally expressed at low levels in the fruits, however they are distinctively expressed in various tissues (Matarese et al. 2014).

In recent studies, the catalytic activity of TPS might not be entirely determined by annotation based on sequence alignment and similarity. However, in order to further know the function of the synthase, we tentatively named the terpene synthase by sequence similarity. Moreover, it has been observed that terpenoid cyclase synthetically modified isoprenoid substrates. For example, aristolochene synthase cyclizes 6,7-dihydrofarnesyl diphosphate to

dihydrogermacrene A, and converts various fluorinated farnesyl diphosphate analogues into its corresponding fluorinated analogues of germacrene A (Koksai et al. 2012). In addition, (+)-germacrene D synthase was highly correlated to germacrene A and biocyclogermacrene in our studies. Edward reported that (+)-(3S)-neomenthol reductase yielded different terpene products (i.e., (+)-(3S)-neomenthol, (-)-(3R)-menthol, (+)-(3S)-isomenthol and (+)-(3R)-neoisomenthol), demonstrating substrate promiscuity. Terpene synthases ubiquitously produce more than one product due to substrate promiscuity and biosynthetic diversity (Ringer et al. 2005). There are temporal and spatial disparities between the genes and the end-products. And TPSs sharing nucleotide or protein identities of 60-80% often have different product spectrum, or catalyze the production of different terpenes. These might explain why some enzymes and their corresponding products were not the most closely correlated, but were highly correlated in our studies.

Moreover, MeJA is not only an inducing trigger of terpene biosynthesis, but also induces other natural products. Thus, terpene biosynthesis might be regulated by other pathway-specific genes.

The expression pattern and putative model for candidate genes

The WRKY proteins are a superfamily of transcription factors involved in the regulation of various physiological programs, including pathogen defense, and wound and stress response (Skibbe et al. 2008; Wang et al. 2007; Zheng et al. 2006). As shown in Table 4, *AvWRKY40* was orthologous to *MaWRKY40* due to both *A. villosum* and *Musa acuminata* belonging to monocotyledoneae zingiberales. However, the function of *MaWRKYs* remains unknown, and as a result all genes were blasted in the NCBI *Arabidopsis thaliana* genome to determine more detailed functional information for the *AvWRKYs*. *AvWRKY40* was identified as *AtWRKY40* (Identifies = 45%; please refer to App Table 12). *AtWRKY40* is a negative

521 Abiscisic acid (ABA) signaling regulator, and ABA treatment represses the *AtWRKY40* gene
522 (Liu et al. 2012). Moreover, *AtWRKY40* interacts with *AtWRKYs* (i.e., *AtWRKY18*,
523 *AtWRKY38* and *AtWRKY60*) and wbox-arach-dna-3 (App Table 13). In addition, *AtWRKY40*
524 has the highest homology to *GaWRKY1* (Schlutenhofer et al. 2014). *GaWRKY1* activates the
525 cotton (+)- δ -cadinene synthase A (CAD1-A) promoter to regulate the sesquiterpene
526 biosynthesis in cotton (Wu et al. 2004). Analogously, we found that *AvWRKY40* was
527 co-expressed with *AvWRKYs* (i.e., *AvWRKY61*, *AvWRKY28*) and *AvNeoD*.

528 *AvWRKY31* was identified as *AtWRKY6*. As Uniprot (<http://www.uniprot.org/>) showed,
529 *AtWRKY6* activates the transcription of the SIRK gene and represses its own expression and
530 *AtWRKY42* genes. Moreover, *AvWRKY61*, *AvWRKY28* and *AvWRKY45* were identified
531 respectively as *AtWRKY72*, *AtWRKY28* and *AtWRKY75*, which were important for regulating
532 jasmonate signaling in Arabidopsis (Schlutenhofer et al. 2014).

533 Furthermore, we also blasted *AvNeoD* against the *Arabidopsis thaliana* Reference
534 Sequence protein database using Blastx to identify additional details with regard this TPS.
535 Consequently, *AvNeoD* was also identified as (+)-neomenthol dehydrogenase in *Arabidopsis*
536 *thaliana* genome (App Table 12). In addition, (+)-neomenthol dehydrogenase (EC: 1.1.1.208)
537 participates in monoterpenoid biosynthesis. Both (+)-neomenthol and NADP⁺ are the
538 substrates of this enzyme that synthesizes (–)-menthone, NADPH, and H⁺ (Kjonaas et al.
539 1982). However, we could find neither (–)-menthone nor (+)-neomenthol as an end-product
540 in *A. villosum* (App Table 4 and 5). Since TPSs sharing nucleotide or protein identities of
541 60-80% often have different product spectrum, or catalyze the production of different
542 terpenes. Thus, (–)-menthone might not be the major product of the protein that is encoded
543 by Unigene 0078114. We need to further identify the major product of this TPS. And the
544 function of TPS must be further identified by experimental use of recombinant proteins, from

which the major product of this TPS would be detected by GC/MS to designate this TPS.

Overall, this model provides a fundamental base to elucidate the molecular mechanisms of medicinal terpene biosynthesis. The *AvWRKY61*, *AvWRKY28*, *AvWRKY40* and *AvNeoD* genes in volatile terpene biosynthesis are promising candidates for designing additional experiments, including molecular cloning, characterization, and functional analysis to confirm their functions and relationships.

High-throughput sequencing on *A. villosum* induced by MeJA was a valid approach in the discovery of WRKY genes and terpene synthases that are involved in terpene biosynthesis. Our article offers an efficient approach to mine transcription factors that are involved in terpenoid biosynthesis. We screened four TPS ((+)-neomenthol dehydrogenase, S-(+)-linalool synthase, (+)-germacrene D synthase and myrcene synthase) sequences and six WRKY TFs genes involved in terpene synthesis. Next, we inferred that *AvWRKY61*, *AvWRKY28* and *AvWRKY40* synergistically regulate *AvNeoD* to biosynthesize both sabinene and beta-pinene. This model affords a fundamental basis for further experiments designed to elucidate the molecular mechanisms of medicinal terpene biosynthesis

ABBREVIATIONS

TPS: Terpene synthase

TF: Transcription factor

MeJA: methyl jasmonate

AvNeoD: (+)-neomenthol dehydrogenase

AvGerD: (+)-germacrene D synthase

DEGs: Differentially expressed genes

RT-qPCR: Reverse transcription and real-time fluorescence quantitative PCR

569
570
571
572
573
574
575
576
577
578
579
580

ACKNOWLEDGMENTS

We are thankful for Shanghai MedSci MedTech Co.Ltd, China for language editing.

CONFLICT OF INTEREST

The authors declare that they have no conflict of interest.

FUNDING

This work is financially supported by National Natural Science Foundation of China (80303163) and Educational Commission Foundation of Guangdong Province of China (Yq2013042).

Draft

References

- Anonymity, 2010. Pharmacopoeia of the People's Republic of China. China Medical Science Press, Beijing.
- Balbi, V., and Devoto, A. 2008. Jasmonate signalling network in *Arabidopsis thaliana*: crucial regulatory nodes and new physiological scenarios. *The New phytologist* **177**(2): 301-318. doi: 10.1111/j.1469-8137.2007.02292.x.
- Browse, J. 2009. Jasmonate passes muster: a receptor and targets for the defense hormone. *Annual review of plant biology* **60**: 183-205. doi: 10.1146/annurev.arplant.043008.092007.
- Chen, F., Tholl, D., Bohlmann, J., and Pichersky, E. 2011. The family of terpene synthases in plants: a mid-size family of genes for specialized metabolism that is highly diversified throughout the kingdom. *The Plant journal : for cell and molecular biology* **66**(1): 212-229. doi: 10.1111/j.1365-313X.2011.04520.x.
- De Geyter, N., Gholami, A., Goormachtig, S., and Goossens, A. 2012. Transcriptional machineries in jasmonate-elicited plant secondary metabolism. *Trends in plant science* **17**(6): 349-359. doi: 10.1016/j.tplants.2012.03.001.
- Degenhardt, J., and Gershenzon, J. 2000. Demonstration and characterization of (E)-nerolidol synthase from maize: a herbivore-inducible terpene synthase participating in (3E)-4,8-dimethyl-1,3,7-nonatriene biosynthesis. *Planta* **210**(5): 815-822. doi: 10.1007/s004250050684.
- Dudareva, N., Pichersky, E., and Gershenzon, J. 2004. Biochemistry of plant volatiles. *Plant physiology* **135**(4): 1893-1902. doi: 10.1104/pp.104.049981.
- Eisen, M.B., Spellman, P.T., Brown, P.O., and Botstein, D. 1998. Cluster analysis and display of genome-wide expression patterns. *Proceedings of the National Academy of Sciences of the United States of America* **95**(25): 14863-14868.
- Eulgem, T., Rushton, P.J., Robatzek, S., and Somssich, I.E. 2000. The WRKY superfamily of plant transcription factors. *Trends in plant science* **5**(5): 199-206.
- Farmer, E.E., Almeras, E., and Krishnamurthy, V. 2003. Jasmonates and related oxylipins in plant responses to pathogenesis and herbivory. *Current opinion in plant biology* **6**(4): 372-378.
- Fukuda, Y., and Shinshi, H. 1994. Characterization of a novel cis-acting element that is responsive to a fungal elicitor in the promoter of a tobacco class I chitinase gene. *Plant molecular biology* **24**(3): 485-493.
- Gantet, P., and Memelink, J. 2002. Transcription factors: tools to engineer the production of pharmacologically

- 609 active plant metabolites. Trends in pharmacological sciences **23**(12): 563-569.
- 610 Huan, W., Jieshu, W., Jinfen, Y., Ruoting, Z., Weiwen, C., and Yuan, Y. 2014. Overexpression of HMGR and
 611 DXR from *Amomum villosum* Lour. Affects the Biosynthesis of Terpenoids in Tobacco. Modernization of
 612 Traditional Chinese Medicine and Materia Medica-World Science and Technology **16**(7): 1024-1527.
- 613 Kjonaas, R., Martinkus-Taylor, C., and Croteau, R. 1982. Metabolism of Monoterpenes: Conversion of
 614 l-Menthone to l-Menthol and d-Neomenthol by Stereospecific Dehydrogenases from Peppermint (*Mentha*
 615 *piperita*) Leaves. Plant physiology **69**(5): 1013-1017.
- 616 Koksai, M., Chou, W.K., Cane, D.E., and Christianson, D.W. 2012. Structure of 2-methylisoborneol synthase
 617 from *Streptomyces coelicolor* and implications for the cyclization of a noncanonical C-methylated
 618 monoterpene substrate. Biochemistry **51**(14): 3011-3020. doi: 10.1021/bi201827a.
- 619 Kombrink, E. 2012. Chemical and genetic exploration of jasmonate biosynthesis and signaling paths. Planta
 620 **236**(5): 1351-1366. doi: 10.1007/s00425-012-1705-z.
- 621 Li, R., Yu, C., Li, Y., Lam, T.W., Yiu, S.M., Kristiansen, K., and Wang, J. 2009. SOAP2: an improved ultrafast
 622 tool for short read alignment. Bioinformatics (Oxford, England) **25**(15): 1966-1967. doi:
 623 10.1093/bioinformatics/btp336.
- 624 Li, S., Zhang, P., Zhang, M., Fu, C., and Yu, L. 2013. Functional analysis of a WRKY transcription factor
 625 involved in transcriptional activation of the DBAT gene in *Taxus chinensis*. Plant biology (Stuttgart,
 626 Germany) **15**(1): 19-26. doi: 10.1111/j.1438-8677.2012.00611.x.
- 627 Liu, Z.Q., Yan, L., Wu, Z., Mei, C., Lu, K., Yu, Y.T., Liang, S., Zhang, X.F., Wang, X.F., and Zhang, D.P. 2012.
 628 Cooperation of three WRKY-domain transcription factors WRKY18, WRKY40, and WRKY60 in
 629 repressing two ABA-responsive genes ABI4 and ABI5 in *Arabidopsis*. Journal of experimental botany
 630 **63**(18): 6371-6392. doi: 10.1093/jxb/ers293.
- 631 Ma, D., Pu, G., Lei, C., Ma, L., Wang, H., Guo, Y., Chen, J., Du, Z., Wang, H., Li, G., Ye, H., and Liu, B. 2009.
 632 Isolation and characterization of AaWRKY1, an *Artemisia annua* transcription factor that regulates the
 633 amorpha-4,11-diene synthase gene, a key gene of artemisinin biosynthesis. Plant & cell physiology **50**(12):
 634 2146-2161. doi: 10.1093/pcp/pcp149.
- 635 Manavella, P.A., Dezar, C.A., Bonaventure, G., Baldwin, I.T., and Chan, R.L. 2008. HAHB4, a sunflower
 636 HD-Zip protein, integrates signals from the jasmonic acid and ethylene pathways during wounding and

- 637 biotic stress responses. *The Plant journal : for cell and molecular biology* **56**(3): 376-388. doi:
638 10.1111/j.1365-313X.2008.03604.x.
- 639 Marchler-Bauer, A., Lu, S., Anderson, J.B., Chitsaz, F., Derbyshire, M.K., DeWeese-Scott, C., Fong, J.H., Geer,
640 L.Y., Geer, R.C., Gonzales, N.R., Gwadz, M., Hurwitz, D.I., Jackson, J.D., Ke, Z., Lanczycki, C.J., Lu, F.,
641 Marchler, G.H., Mullokandov, M., Omelchenko, M.V., Robertson, C.L., Song, J.S., Thanki, N., Yamashita,
642 R.A., Zhang, D., Zhang, N., Zheng, C., and Bryant, S.H. 2011. CDD: a Conserved Domain Database for
643 the functional annotation of proteins. *Nucleic acids research* **39**(Database issue): D225-229. doi:
644 10.1093/nar/gkq1189.
- 645 Matarese, F., Cuzzola, A., Scalabrelli, G., and D'Onofrio, C. 2014. Expression of terpene synthase genes
646 associated with the formation of volatiles in different organs of *Vitis vinifera*. *Phytochemistry* **105**: 12-24.
647 doi: 10.1016/j.phytochem.2014.06.007.
- 648 Menke, F.L., Champion, A., Kijne, J.W., and Memelink, J. 1999. A novel jasmonate- and elicitor-responsive
649 element in the periwinkle secondary metabolite biosynthetic gene *Str* interacts with a jasmonate- and
650 elicitor-inducible AP2-domain transcription factor, ORCA2. *The EMBO journal* **18**(16): 4455-4463. doi:
651 10.1093/emboj/18.16.4455.
- 652 Mingfa, Z., and Yaqin, S. 2013. Research Advances in Clinical Pharmacological Effects on Amomi Fructus.
653 *Anti-Infection Pharmacy* **10**(1): 8-12.
- 654 Pauw, B., Hilliou, F.A., Martin, V.S., Chatel, G., de Wolf, C.J., Champion, A., Pre, M., van Duijn, B., Kijne, J.W.,
655 van der Fits, L., and Memelink, J. 2004. Zinc finger proteins act as transcriptional repressors of alkaloid
656 biosynthesis genes in *Catharanthus roseus*. *The Journal of biological chemistry* **279**(51): 52940-52948. doi:
657 10.1074/jbc.M404391200.
- 658 Ringer, K.L., Davis, E.M., and Croteau, R. 2005. Monoterpene metabolism. Cloning, expression, and
659 characterization of (-)-isopiperitenol/(-)-carveol dehydrogenase of peppermint and spearmint. *Plant*
660 *physiology* **137**(3): 863-872. doi: 10.1104/pp.104.053298.
- 661 Robinson, M.D., McCarthy, D.J., and Smyth, G.K. 2010. edgeR: a Bioconductor package for differential
662 expression analysis of digital gene expression data. *Bioinformatics (Oxford, England)* **26**(1): 139-140. doi:
663 10.1093/bioinformatics/btp616.
- 664 Rodriguez-Concepcion, M., and Boronat, A. 2002. Elucidation of the methylerythritol phosphate pathway for

- isoprenoid biosynthesis in bacteria and plastids. A metabolic milestone achieved through genomics. *Plant physiology* **130**(3): 1079-1089. doi: 10.1104/pp.007138.
- Rushton, P.J., Macdonald, H., Huttly, A.K., Lazarus, C.M., and Hooley, R. 1995. Members of a new family of DNA-binding proteins bind to a conserved cis-element in the promoters of alpha-Amy2 genes. *Plant molecular biology* **29**(4): 691-702.
- Schluttenhofer, C., Pattanaik, S., Patra, B., and Yuan, L. 2014. Analyses of *Catharanthus roseus* and *Arabidopsis thaliana* WRKY transcription factors reveal involvement in jasmonate signaling. *BMC genomics* **15**: 502. doi: 10.1186/1471-2164-15-502.
- Skibbe, M., Qu, N., Galis, I., and Baldwin, I.T. 2008. Induced plant defenses in the natural environment: *Nicotiana attenuata* WRKY3 and WRKY6 coordinate responses to herbivory. *The Plant cell* **20**(7): 1984-2000. doi: 10.1105/tpc.108.058594.
- Spyropoulou, E.A., Haring, M.A., and Schuurink, R.C. 2014. RNA sequencing on *Solanum lycopersicum* trichomes identifies transcription factors that activate terpene synthase promoters. *BMC genomics* **15**: 402. doi: 10.1186/1471-2164-15-402.
- Suttipanta, N., Pattanaik, S., Kulshrestha, M., Patra, B., Singh, S.K., and Yuan, L. 2011. The transcription factor CrWRKY1 positively regulates the terpenoid indole alkaloid biosynthesis in *Catharanthus roseus*. *Plant physiology* **157**(4): 2081-2093. doi: 10.1104/pp.111.181834.
- van der Fits, L., and Memelink, J. 2001. The jasmonate-inducible AP2/ERF-domain transcription factor ORCA3 activates gene expression via interaction with a jasmonate-responsive promoter element. *The Plant journal : for cell and molecular biology* **25**(1): 43-53.
- Vanderauwera, S., Zimmermann, P., Rombauts, S., Vandenabeele, S., Langebartels, C., Gruissem, W., Inze, D., and Van Breusegem, F. 2005. Genome-wide analysis of hydrogen peroxide-regulated gene expression in *Arabidopsis* reveals a high light-induced transcriptional cluster involved in anthocyanin biosynthesis. *Plant physiology* **139**(2): 806-821. doi: 10.1104/pp.105.065896.
- Vom Endt, D., Kijne, J.W., and Memelink, J. 2002. Transcription factors controlling plant secondary metabolism: what regulates the regulators? *Phytochemistry* **61**(2): 107-114.
- Wang, H., Hao, J., Chen, X., Hao, Z., Wang, X., Lou, Y., Peng, Y., and Guo, Z. 2007. Overexpression of rice WRKY89 enhances ultraviolet B tolerance and disease resistance in rice plants. *Plant molecular biology*

- 693 **65**(6): 799-815. doi: 10.1007/s11103-007-9244-x.
- 694 Wu, X., Li, X., Xiao, F., Zhang, Z., Xu, Z., and Wang, H. 2004. Studies on the analgesic and anti-inflammatory
695 effect of bornyl acetate in volatile oil from *Amomum villosum*. *Zhong yao cai = Zhongyaocai = Journal of*
696 *Chinese medicinal materials* **27**(6): 438-439.
- 697 Wu, X., Xiao, F., Zhang, Z., Li, X., and Xu, Z. 2005. Research on the analgesic effect and mechanism of bornyl
698 acetate in volatile oil from *amomum villosum*. *Zhong yao cai = Zhongyaocai = Journal of Chinese*
699 *medicinal materials* **28**(6): 505-507.
- 700 Xu, Y.H., Wang, J.W., Wang, S., Wang, J.Y., and Chen, X.Y. 2004. Characterization of GaWRKY1, a cotton
701 transcription factor that regulates the sesquiterpene synthase gene (+)-delta-cadinene synthase-A. *Plant*
702 *physiology* **135**(1): 507-515. doi: 10.1104/pp.104.038612.
- 703 Yang, J., Adhikari, M.N., Liu, H., Xu, H., He, G., Zhan, R., Wei, J., and Chen, W. 2012. Characterization and
704 functional analysis of the genes encoding 1-deoxy-D-xylulose-5-phosphate reductoisomerase and
705 1-deoxy-D-xylulose-5-phosphate synthase, the two enzymes in the MEP pathway, from *Amomum villosum*
706 Lour. *Molecular biology reports* **39**(8): 8287-8296. doi: 10.1007/s11033-012-1676-y.
- 707 Yongzhen, S., Yunyun, N., Jiang, X.Y., L., Hongmei, L.Y., Z., Mingzhu, L., Qiong, W., Jingyuan, S., Chao, S.,
708 and Shilin, C. 2013. Discovery of WRKY transcription factors through transcriptome analysis and
709 characterization of a novel methyl jasmonate-inducible PqWRKY1 gene from *Panax quinquefolius*. *Plant*
710 *Cell, Tissue and Organ Culture* **114**(2): 269-277.
- 711 Zhang, H., Hedhili, S., Montiel, G., Zhang, Y., Chatel, G., Pre, M., Gantet, P., and Memelink, J. 2011. The basic
712 helix-loop-helix transcription factor CrMYC2 controls the jasmonate-responsive expression of the ORCA
713 genes that regulate alkaloid biosynthesis in *Catharanthus roseus*. *The Plant journal : for cell and molecular*
714 *biology* **67**(1): 61-71. doi: 10.1111/j.1365-313X.2011.04575.x.
- 715 Zhang, J., Wu, K., Zeng, S., Teixeira da Silva, J.A., Zhao, X., Tian, C.E., Xia, H., and Duan, J. 2013.
716 Transcriptome analysis of *Cymbidium sinense* and its application to the identification of genes associated
717 with floral development. *BMC genomics* **14**: 279. doi: 10.1186/1471-2164-14-279.
- 718 Zhao, J., Davis, L.C., and Verpoorte, R. 2005. Elicitor signal transduction leading to production of plant
719 secondary metabolites. *Biotechnology advances* **23**(4): 283-333. doi: 10.1016/j.biotechadv.2005.01.003.
- 720 Zheng, Z., Qamar, S.A., Chen, Z., and Mengiste, T. 2006. Arabidopsis WRKY33 transcription factor is required

721 for resistance to necrotrophic fungal pathogens. The Plant journal : for cell and molecular biology **48**(4):
722 592-605. doi: 10.1111/j.1365-313X.2006.02901.x.
723

Draft

724 **Table 1.** Information for RNA-Seq samples.

RNA-Seq sample	Spraying organs	Detecting organs	Treatment
FSS	fruits	seeds	solvent control
FM1S	fruits	seeds	200μmol/L MeJA
FM2S	fruits	seeds	600μmol/L MeJA
LM1S	leaves	seeds	200μmol/L MeJA
FSP	fruits	peels	solvent control
FM1P	fruits	peels	200μmol/L MeJA
FM2P	fruits	peels	600μmol/L MeJA
LM1P	leaves	peels	200μmol/L MeJA
L0L	-	leaves	blank control
LSL	leaves	leaves	solvent control
LM1L	leaves	leaves	200μmol/L MeJA

725 Explanation for the sample name: F: The sprayed organs were fruits. L of the first letter: The
 726 sprayed organs were leaves. S of the second letter: Treated with solvent (0.02% tween 80).
 727 0: Blank control. M1: Treated with 200μmol/L MeJA. M2: Treated with 600μmol/L MeJA. S
 728 of the last letter: The detected organs were seeds. P: The detected organs were peels. L of the
 729 last letter: The detected organs were leaves.

730

731 **Table 2.** Overview of the terpene biosynthetic pathways.

Terpene related pathways	No. of unigenes	Pathway ID
Terpenoid backbone biosynthesis	180(0.84%)	ko00900
Monoterpenoid biosynthesis	12(0.06%)	ko00902
Sesquiterpenoid biosynthesis	13(0.06%)	ko00909
Diterpenoid biosynthesis	50(0.23%)	ko00904
Ubiquinone and other terpenoid-quinone biosynthesis	132(0.61%)	ko00130
Biosynthesis of secondary metabolites	2459(11.44%)	ko01110

732 The results of pathway enrichment analysis showed that DEGs were enriched for the
733 pathways (Q value ≤ 0.05)
734

735 **Table 3.** A list of selected terpene synthase genes in *A. villosum*.

Enzyme name	EC ID	Unigene ID	Renference species	Identities
myrcene synthase	4.2.3.15	Unigene0099398	<i>Alstroemeria peruviana</i>	64%
(+)-neomenthol dehydrogenase	1.1.1.208	Unigene0078114	<i>Musa acuminata subsp</i>	76%
linalool synthase	4.2.3.25	Unigene0136809	<i>Citrus unshiu</i>	86%
3S,6E-nerolidol synthase	4.2.3.28	Unigene0060131	<i>Eucalyptus grandis</i>	44%
S-(+)-linalool synthase	4.2.3.25	Unigene0060132	<i>Cinnamomum osmophloeum</i>	56%
(+)-germacrene D synthase	4.2.3.75	Unigene0106615	<i>Zingiber officinale</i>	70%
sesquiterpene synthase 3	-	Unigene0106619	<i>Zingiber zerumbet</i>	77%
sesquiterpene synthase A1	-	Unigene0140412	<i>Zingiber zerumbet</i>	84%

736 The unigenes were named according to their sequence identities.

737

738 **Table 4.** Candidate WRKY genes involved in the biosynthesis of volatile terpenes in *A.*
739 *villosum*.

WRKY name	Unigene ID	Renference species	Identities
WRKY28	Unigene0044272	<i>Musa acuminata subsp</i>	64%
WRKY31	Unigene0062811	<i>Musa acuminata subsp</i>	55%
WRKY61	Unigene0037009	<i>Musa acuminata subsp</i>	53%
WRKY1	Unigene0008375	<i>Musa acuminata subsp</i>	85%
WRKY45	Unigene0057892	<i>Zea mays</i>	91%
WRKY40	Unigene0102915	<i>Musa acuminata subsp</i>	56%

740 The unigenes were named according to their sequence identities.

741

Figure legends

Fig 1. Distribution of transcription factors in the transcriptome dataset.

A. The distribution of transcription factors in the transcriptome dataset. B. Table showing the numbers of each transcription factor in response to MeJA in *A. villosum*.

Fig 2. Analysis of differentially expressed WRKY TFs induced by MeJA in RNA-Seq.

A. The distribution of WRKY transcription factors in the RNA-Seq. B. The distribution of differentially expressed genes (2×) annotated WRKY TFs in FSP versus FM1P, FSP versus FM2P, FSS versus FM1S, FSS versus FM2S, L0L versus LSL, L0L versus LM1L, LSL versus LM1L. C. The distribution of WRKY TFs in the rest of four comparisons: LM1P versus FM1P, LM1S versus FM1S, FM1P versus FM2P and FM1S versus FM2S. D. Numbers of differentially expressed genes annotated WRKY TFs in above four comparisons.

Fig 3. Co-expression network of the WRKY and terpene synthase genes.

The terpene synthase genes (green nodes) and their relative WRKYs genes (gray nodes) are included in the graphical representation (Pearson coefficient threshold $> |0.8|$). The red line represented a positive correlation (positive feedback), and the blue line represented a negative correlation (negative feedback).

Fig 4. The correlation networks for the terpenes and terpene synthase genes.

The terpene synthase genes (yellow nodes) and their related terpenes (pink nodes) are included in the graphical representation (Pearson coefficient threshold $> |0.8|$). The red line represents a positive feedback (positive correlation), and the green line represents a negative

765 feedback (negative correlation).

766

767 **Fig 5.** The correlation networks for selected sequences of WRKY and the terpenes.

768 The WRKY genes (yellow nodes) and their relative terpenes (pink nodes) are included in the
769 graphical representation (Pearson coefficient threshold $> |0.8|$). The red line represents a
770 positive feedback (positive correlation), and the green line represents a negative feedback
771 (negative correlation).

772

773 **Fig 6.** Hierarchical clusters for the candidate genes of WRKY TFs and TPS.

774 Hierarchical clustering of RT-qPCR is quantified to the candidate WRKY TFs and TPS genes
775 expression in *A. villosum*. Each column represents an experimental sample (e.g., FM2S,
776 FM1S and FSS, among others) and each row represents a gene. Differences in the expression
777 are shown by different colors. Red means high expression and green means low expression.
778 The color gradually from green to red represented the genes expression abundance from low
779 to high. FSS: The seeds of solvent control; FM1S: The seeds treated with 200 $\mu\text{mol/L}$ MeJA;
780 FM2S: The seeds treated with 600 $\mu\text{mol/L}$ MeJA; FSP: The peels of solvent control; FM1P:
781 The peels treated with 200 $\mu\text{mol/L}$ MeJA; FM2P: The peels treated with 600 $\mu\text{mol/L}$ MeJA.
782 There were three biological replicates and three technical replicates for each target gene.

783

784 **Figure7.** Mode for validated genes of WRKY TFs and TPS in *A. villosum* based on
785 expressions.

786 The model depicts the relationship of *AvWRKYs* and *AvTPS* from their expression patterns.
787 The genes of *AvWRKY* (blue nodes), *AvNeoD* (pink nodes) and their related volatile
788 compounds (as shown by yellow nodes) are included in the graphical representation

789 (threshold $> |0.8|$). The red line represents a positive feedback (positive correlation), and the
790 green line represents a negative correlation.

Draft

Distribution of transcription factors in the transcriptome dataset.

59x44mm (300 x 300 DPI)

Analysis of differentially expressed WRKY TFs induced by MeJA in RNA-Seq

141x111mm (300 x 300 DPI)

40x20mm (300 x 300 DPI)

The correlation networks for the terpenes and terpene synthase genes.

65x53mm (300 x 300 DPI)

The correlation networks for selected sequences of WRKY and the terpenes.

67x56mm (300 x 300 DPI)

Hierarchical clusters for the candidate genes of WRKY TFs and TPS.

85x91mm (300 x 300 DPI)

Mode for validated genes of WRKY TFs and TPS in *A. villosum* based on expressions.
57x45mm (300 x 300 DPI)

Appendix figure and tables

App Figure 1 Thirty-six selected sequences contained complete WRKY domains in *A.villosum*.

WRKY domain alignment was performed by MEGA 5.10, and a list of thirty-six selected WRKY sequences.

App Table 1. The results for PlantTFcat and NCBI CDD identified WRKY domain-containing sequences.

	PlantTFcat identified sequences	NCBI CDD identified sequences	Complete WRKY domains sequences
1	Unigene0037363	Unigene0037363	Unigene0037363
2	Unigene0046046	Unigene0135712	Unigene0135712
3	Unigene0002923	Unigene0041452	Unigene0041452
4	Unigene0133624	Unigene0102917	Unigene0102917
5	Unigene0040900	Unigene0037009	Unigene0037009
6	Unigene0036133	Unigene0137090	Unigene0137090
7	Unigene0028374	Unigene0063988	Unigene0063988
8	Unigene0102930	Unigene0066728	Unigene0046046
9	Unigene0135712	Unigene0046046	Unigene0002923
10	Unigene0044915	Unigene0002923	Unigene0133624
11	Unigene0043971	Unigene0133624	Unigene0040900
12	Unigene0046044	Unigene0040900	Unigene0036133
13	Unigene0050498	Unigene0036133	Unigene0102930
14	Unigene0102915	Unigene0028374	Unigene0044915
15	Unigene0038138	Unigene0016342	Unigene0046044
16	Unigene0041452	Unigene0102930	Unigene0050498
17	Unigene0060436	Unigene0044915	Unigene0102915
18	Unigene0135501	Unigene0043971	Unigene0060436
19	Unigene0049220	Unigene0043973	Unigene0135501
20	Unigene0102927	Unigene0046044	Unigene0102920
21	Unigene0102920	Unigene0050498	Unigene0087360
22	Unigene0131751	Unigene0102915	Unigene0097802
23	Unigene0087360	Unigene0038138	Unigene0008375
24	Unigene0102917	Unigene0060436	Unigene0057891
25	Unigene0044271	Unigene0135501	Unigene0140721
26	Unigene0097802	Unigene0049220	Unigene0062811
27	Unigene0008375	Unigene0102927	Unigene0057892
28	Unigene0043562	Unigene0102920	Unigene0044272
29	Unigene0097801	Unigene0047127	Unigene0081407
30	Unigene0057891	Unigene0131751	Unigene0040899

31	Unigene0140721	Unigene0087360	Unigene0134031
32	Unigene0062811	Unigene0044271	Unigene0133885
33	Unigene0037009	Unigene0097802	Unigene0070107
34	Unigene0057892	Unigene0008375	Unigene0049276
35	Unigene0087357	Unigene0026300	Unigene0057146
36	Unigene0044272	Unigene0043562	Unigene0060161
37	Unigene0081407	Unigene0097801	
38	Unigene0021556	Unigene0057891	
39	Unigene0040899	Unigene0140721	
40	Unigene0137090	Unigene0062811	
41	Unigene0134031	Unigene0057892	
42	Unigene0133885	Unigene0047799	
43	Unigene0070107	Unigene0087357	
44	Unigene0049276	Unigene0044272	
45	Unigene0060165	Unigene0081407	
46	Unigene0062812	Unigene0021556	
47	Unigene0142516	Unigene0040899	
48	Unigene0063988	Unigene0134031	
49	Unigene0057890	Unigene0133885	
50	Unigene0057146	Unigene0062808	
51	Unigene0081517	Unigene0070107	
52	Unigene0060161	Unigene0049276	
53	Unigene0066728	Unigene0060165	
54		Unigene0062812	
55		Unigene0142516	
56		Unigene0057890	
57		Unigene0057146	
58		Unigene0081517	
59		Unigene0086349	
60		Unigene0060161	

WRKY domain containing proteins were identified using two sources: PlantTFcat (<http://plantgrn.noble.org/PlantTFcat/>) and NCBI CDD. PlantTFcat identified 53 WRKY domain-containing sequences, while NCBI CDD identified 60 WRKY sequences. NCBI CDD identified 36 sequences that contained complete WRKY domains in *A.villosum*, while PlantTFcat not showed details about them.

App Table 2. The terpene contents in the seeds of *A. villosum* induced by MeJA

NO.	Compounds	FSS	FM1S	FM2S
1	bornyl acetate	6220.92±130.88	4998.89±6.81**	6061.66±176.31 [#]
2	bicyclogermacrene	774.56±25.53	557.40 ±11.34**	739.02±37.52 **
3	limonene	744.22±26.26	384.99±0.12**	531.35±25.27 ** ^{##}
4	camphene	460.0±16.00	179.99±4.30 **	261.94±14.01 ^{###}
5	myrcene	304.7±10.20	145.82±3.09 **	203.86 ±8.51 ^{###}
6	bicycloelemene	74.05±0.75	52.50±0.26 **	70.16±1.44 ^{###}
7	geraniol	69.41±4.70	47.37±0.59 **	69.15±2.90 [#]
8	alpha-pinene	104.08±7.62	38.84±2.32	55.98±1.44
9	beta-bisabolene	55.36±5.14	37.86±6.99 *	52.52±2.10
10	aromadendrene	38.01±2.80	24.68±6.59	34.1±3.43
11	germacrene	41.86±3.33	26.46±0.14 *	32.50±3.61 *
12	delta-cadinene	35.3±0.99	23.49±0.65 **	28.75±1.47 ^{##}
13	beta-sesquiphellandrene	32.5±1.02	22.83±0.80 **	27.30±0.46 ^{##}
14	alpha-bisabolol	23.01±0.60	16.54±0.56 **	22.11±0.54 ^{##}
15	alpha-curcumene	26.1±2.11	17.58±0.14 **	24.24±1.23 [#]
16	gamma-cadinene	21.97±1.56	15.78±0.26 **	17.92±0.53 *
17	4-terpineol	12.73±1.33	10.8±0.96	11.55±1.76
18	germacreneA	13.79±5.03	8.56±0.28	12.12±4.19
19	alpha-terpinene	14.51±0.15	8.46±0.18 **	10.78±0.92 ^{##}
20	nerolidol	12.97±0.22	8.08±2.16	12.08±2.03
21	beta-pinene	6.41±0.45	3.01±0.20 **	3.37±2.32 ^{##}
22	sabinene	1.17±0.74	0.32±0.23	0.41±0.21
23	borneol	926.53±18.71	918.21±7.02	1026.01±33.73 * [#]
24	beta-humulene	207.18±64.52	166.76±1.48	233.49±56.31
25	isoborneol	126.78±4.16	123.78±5.35	134.36±5.32
26	alpha-bergamotol	41.08±0.33	30.03±0.30 **	40.4±3.12 [#]
27	beta-farnesene	37.88±1.09	20.78±6.02 *	38.78±2.69 [#]
28	alpha-santalol	34.06±6.94	19.65±1.18	36.34±11.73
29	linalool	19.98±0.22	17.58±2.33	21.9±0.30
30	alpha-cedrene	14.51±1.00	10.44±0.31 **	14.74±0.36 ^{##}
31	camphor	2811.76±64.84	2702.66±11.83	2589.02±95.60 *
32	isoborneol acetate	393.14±30.49	388.7±14.35	383.26±20.76
33	alpha- copaene	66.38±2.01	42.18±0.59 **	41.86±0.16 **

The terpene contents in the seeds were detected by gas chromatography–mass spectrometry

(GC-MS). Comparing with FSS, *represented $p < 0.05$, **represented $p < 0.01$; Comparing with FM1S,

represented $p < 0.05$, ## represented $p < 0.01$.

App Table 3. The terpene contents in the peels of *A.villosum* induced by MeJA.

NO.	Compounds	FSP	FM1P	FM2P
1	linalool	119.95±0.78	191.92±3.86 *	290.47±30.07 **#
2	beta-pinene	30.77±0.87	98.98±1.66 *	229.74±22.18 **
3	alpha-pinene	12.75±1.49	44.89±2.22*	120.04±14.41 **
4	sabinene	7.84±0.66	24.96±0.39 *	86.39±8.39 **
5	limonene	1.8±0.09	3.43±0.76	12.13±1.82 **
6	alpha-copaene	4.59±0.11	6.78±1.21	9.58 ±0.55**#
7	aromadendrene	1.94±0.22	2.31±0.10 *	3.71±0.04 **##
8	camphene	0.8±0.08	1.43±0.62	3.82±1.51 *#
9	4-terpineol	1.25±0.35	3.26±0.90*	2.38±0.16
10	bornyl acetate	92.36±0.57	8.95±1.04**	102.17±12.97**#
11	bicyclogermacrene	32.68±0.01	22.61±0.72*	63.16±0.56 **##
12	camphor	16.82±0.30	4.06±0.23 **	19.89±3.18##
13	beta-farnesene	5.17±0.04	3.31±0.28*	7.35±0.54***
14	myrcene	1.53±0.95	1.16±0.34	6.70±0.09 **
15	bicycloelemene	3.06±0.16	1.87±0.02**	5.92±0.10 **
16	geraniol	5.05±1.02	3.27±0.17	7.77±1.96#
17	isoborneol acetate	5.00±0.43	1.80±0.98*	5.37±0.04#
18	beta-humulene	3.87±0.31	1.70±0.15**	4.99±0.08***
19	isoborneol	0.9±0.21	0.41±0.06*	1.13±0.11#
20	borneol	9.16±0.27	0.93±0.05**	7.65±1.68##

The terpene contents in the peels were detected by gas chromatography–mass spectrometry (GC-MS).Comparing with FSP, *represented $p<0.05$, **represented $p<0.01$; Comparing with FM1P, # represented $p<0.05$, ## represented $p<0.01$.

App Table 4. The Spearman correlation coefficients between eight TPS genes and 33 terpenes.

TPS UnigeneID	Terpenes Name	Pearson correlation coefficient
Unigene0099398	camphene	0.928859
Unigene0099398	myrcene	0.901386
Unigene0099398	limonene	0.885833
Unigene0099398	beta-farnesene	0.872037
Unigene0099398	alpha-terpinene	0.856408
Unigene0099398	alpha-copaene	0.839171
Unigene0099398	alpha-santalol	0.838317
Unigene0099398	germacreneA	0.838119
Unigene0099398	aromadendrene	0.836295
Unigene0099398	germacrene	0.834837
Unigene0099398	nerolidol	0.834616
Unigene0099398	delta-cadinene	0.820036
Unigene0099398	geraniol	0.816403
Unigene0099398	alpha-curcumene	0.814108
Unigene0099398	beta-bisabolene	0.808936
Unigene0099398	bicycloelemene	0.804349
Unigene0099398	beta-sesquiphellandrene	0.803316
Unigene0078114	myrcene	-0.80888
Unigene0078114	limonene	-0.82774
Unigene0078114	alpha-santalol	-0.82972
Unigene0078114	alpha-copaene	-0.84602
Unigene0078114	geraniol	-0.84973
Unigene0078114	alpha-terpinene	-0.85292
Unigene0078114	nerolidol	-0.85811
Unigene0078114	aromadendrene	-0.85846
Unigene0078114	germacreneA	-0.86112
Unigene0078114	germacrene	-0.86758
Unigene0078114	bicycloelemene	-0.86764
Unigene0078114	bicyclogermacrene	-0.87056
Unigene0078114	alpha-curcumene	-0.87261
Unigene0078114	beta-bisabolene	-0.87376
Unigene0078114	delta-cadinene	-0.87603
Unigene0078114	alpha-cedrene	-0.87662
Unigene0078114	beta-humulene	-0.87936
Unigene0078114	alpha-bisabolol	-0.88144
Unigene0078114	alpha-bergamotol	-0.8824
Unigene0078114	beta-sesquiphellandrene	-0.88444
Unigene0078114	gamma-cadinene	-0.88855
Unigene0078114	bornyl acetate	-0.89423
Unigene0078114	borneol	-0.91245
Unigene0078114	isoborneol	-0.91561

Unigene0078114	isoborneol acetate	-0.92081
Unigene0078114	camphor	-0.92109
Unigene0078114	4-terpineol	-0.93867
Unigene0060132	sabinene	0.815762
Unigene0106615	beta-humulene	-0.9488
Unigene0106615	borneol	-0.95136
Unigene0106615	beta-farnesene	-0.95298
Unigene0106615	camphene	-0.95531
Unigene0106615	alpha-santalol	-0.95717
Unigene0106615	isoborneol	-0.95746
Unigene0106615	geraniol	-0.96089
Unigene0106615	alpha-cedrene	-0.96239
Unigene0106615	bornyl acetate	-0.96429
Unigene0106615	alpha-copaene	-0.96489
Unigene0106615	isoborneol acetate	-0.96497
Unigene0106615	alpha-bergamotol	-0.9653
Unigene0106615	bicyclogermacrene	-0.9666
Unigene0106615	bicycloelemene	-0.96661
Unigene0106615	alpha-bisabolol	-0.96715
Unigene0106615	beta-bisabolene	-0.96804
Unigene0106615	myrcene	-0.9684
Unigene0106615	nerolidol	-0.96927
Unigene0106615	camphor	-0.96936
Unigene0106615	alpha-curcumene	-0.96941
Unigene0106615	germacreneA	-0.97212
Unigene0106615	limonene	-0.9723
Unigene0106615	alpha-terpinene	-0.97298
Unigene0106615	beta-sesquiphellandrene	-0.97343
Unigene0106615	germacrene	-0.9738
Unigene0106615	gamma-cadinene	-0.97384
Unigene0106615	delta-cadinene	-0.97386
Unigene0106615	aromadendrene	-0.97404
Unigene0106615	4-terpineol	-0.99286

App Table 5. The Spearman correlation coefficients between six WRKYgenes and 33 terpenes

WRKY UnigeneID	Terpenes Name	Pearson correlation coefficient
Unigene0037009	imonene	-0.81337
Unigene0037009	bicycloelemene	-0.85799
Unigene0037009	camphor	-0.87558
Unigene0037009	linalool	0.882767
Unigene0037009	bronyl acetate	-0.87079
Unigene0037009	isobornyl acetate	-0.87777
Unigene0037009	beta-humulene	-0.86202
Unigene0037009	4-terpineol	-0.80311
Unigene0037009	aromadendrene	-0.83863
Unigene0037009	isoborneol	-0.87588
Unigene0037009	beta-farnesene	-0.8144
Unigene0037009	borneol	-0.87623
Unigene0037009	bicyclogermacrene	-0.85842
Unigene0037009	geraniol	-0.85252
Unigene0037009	beta-bisabolene	-0.85318
Unigene0037009	germacrene	-0.84345
Unigene0037009	delta-cadinene	-0.85034
Unigene0037009	germacreneA	-0.84361
Unigene0037009	beta-sesquiphellandrene	-0.85624
Unigene0037009	gamma-cadinene	-0.85739
Unigene0037009	alpha-cedrene	-0.85582
Unigene0037009	nerolidol	-0.84301
Unigene0037009	alpha-bisabol	-0.85794
Unigene0037009	alpha-curcumene	-0.85207
Unigene0037009	alpha-santalol	-0.82564
Unigene0037009	alpha-bergamotol	-0.85884
Unigene0037009	alpha-terpinene	-0.83245
Unigene0102920	camphene	-0.8552
Unigene0102920	myrcene	-0.89775
Unigene0102920	imonene	-0.9147
Unigene0102920	bicycloelemene	-0.95286
Unigene0102920	alpha-copaene	-0.91765
Unigene0102920	camphor	-0.97891
Unigene0102920	linalool	0.846757
Unigene0102920	bronyl acetate	-0.969
Unigene0102920	isobornyl acetate	-0.97938
Unigene0102920	beta-humulene	-0.95948
Unigene0102920	4-terpineol	-0.98835
Unigene0102920	aromadendrene	-0.94472
Unigene0102920	isoborneol	-0.97849
Unigene0102920	beta-farnesene	-0.88456

Unigene0102920	borneol	-0.9765
Unigene0102920	bicyclogermacrene	-0.9548
Unigene0102920	geraniol	-0.94131
Unigene0102920	beta-bisabolene	-0.9557
Unigene0102920	germacrene	-0.94654
Unigene0102920	delta-cadinene	-0.95383
Unigene0102920	germacreneA	-0.94568
Unigene0102920	beta-sesquiphellandrene	-0.96015
Unigene0102920	gamma-cadinene	-0.96168
Unigene0102920	alpha-cedrene	-0.95805
Unigene0102920	nerolidol	-0.94451
Unigene0102920	alpha-bisabolol	-0.96091
Unigene0102920	alpha-curcumene	-0.95465
Unigene0102920	alpha-santalol	-0.92386
Unigene0102920	alpha-bergamotol	-0.96171
Unigene0102920	alpha-terpinene	-0.93462
Unigene0008375	camphene	0.909983
Unigene0008375	myrcene	0.867293
Unigene0008375	imonene	0.845348
Unigene0008375	alpha-copaene	0.88049
Unigene0008375	germacrene	0.805993
Unigene0008375	alpha-terpinene	0.825834
Unigene0062811	camphene	0.94185
Unigene0062811	myrcene	0.968928
Unigene0062811	imonene	0.977821
Unigene0062811	bicycloelemene	0.988606
Unigene0062811	alpha-copaene	0.958577
Unigene0062811	camphor	0.960141
Unigene0062811	linalool	-0.80534
Unigene0062811	bronyl acetate	0.982732
Unigene0062811	isobornyl acetate	0.959073
Unigene0062811	beta-humulene	0.979788
Unigene0062811	4-terpineol	0.988095
Unigene0062811	aromadendrene	0.991783
Unigene0062811	isoborneol	0.964644
Unigene0062811	beta-farnesene	0.969838
Unigene0062811	borneol	0.962793
Unigene0062811	bicyclogermacrene	0.988542
Unigene0062811	geraniol	0.987246
Unigene0062811	beta-bisabolene	0.990412
Unigene0062811	germacrene	0.985067
Unigene0062811	delta-cadinene	0.986843
Unigene0062811	germacreneA	0.990704
Unigene0062811	beta-sesquiphellandrene	0.986531

Unigene0062811	gamma-cadinene	0.983646
Unigene0062811	alpha-cedrene	0.988337
Unigene0062811	nerolidol	0.990989
Unigene0062811	alpha-bisabolol	0.989166
Unigene0062811	alpha-curcumene	0.990665
Unigene0062811	alpha-santalol	0.983274
Unigene0062811	alpha-bergamotol	0.988488
Unigene0062811	alpha-terpinene	0.982549
Unigene0044272	myrcene	-0.81399
Unigene0044272	imonene	-0.83493
Unigene0044272	bicycloelemene	-0.91215
Unigene0044272	alpha-copaene	-0.84189
Unigene0044272	camphor	-0.93299
Unigene0044272	bronyl acetate	-0.92312
Unigene0044272	isobornyl acetate	-0.93892
Unigene0044272	beta-humulene	-0.9247
Unigene0044272	4-terpineol	-0.9139
Unigene0044272	aromadendrene	-0.89289
Unigene0044272	isoborneol	-0.94234
Unigene0044272	beta-farnesene	-0.86055
Unigene0044272	borneol	-0.94411
Unigene0044272	bicyclogermacrene	-0.91439
Unigene0044272	geraniol	-0.90634
Unigene0044272	beta-bisabolene	-0.89984
Unigene0044272	germacrene	-0.87409
Unigene0044272	delta-cadinene	-0.8861
Unigene0044272	germacreneA	-0.88168
Unigene0044272	beta-sesquiphellandrene	-0.89611
Unigene0044272	gamma-cadinene	-0.89631
Unigene0044272	alpha-cedrene	-0.90869
Unigene0044272	nerolidol	-0.88499
Unigene0044272	alpha-bisabolol	-0.90726
Unigene0044272	alpha-curcumene	-0.89679
Unigene0044272	alpha-santalol	-0.87393
Unigene0044272	alpha-bergamotol	-0.91027
Unigene0044272	alpha-terpinene	-0.85702
Unigene0102915	camphene	-0.88972
Unigene0102915	beta-pinene	0.843147
Unigene0102915	myrcene	-0.92259
Unigene0102915	imonene	-0.93557
Unigene0102915	bicycloelemene	-0.9568
Unigene0102915	alpha-copaene	-0.92088
Unigene0102915	camphor	-0.97052
Unigene0102915	linalool	0.95967

Unigene0102915	bronyl acetate	-0.97014
Unigene0102915	isobornyl acetate	-0.96954
Unigene0102915	beta-humulene	-0.95825
Unigene0102915	4-terpineol	-0.95356
Unigene0102915	aromadendrene	-0.94943
Unigene0102915	isoborneol	-0.96816
Unigene0102915	beta-farnesene	-0.90238
Unigene0102915	borneol	-0.96601
Unigene0102915	bicyclogermacrene	-0.95752
Unigene0102915	geraniol	-0.94799
Unigene0102915	beta-bisabolene	-0.96293
Unigene0102915	germacrene	-0.95989
Unigene0102915	delta-cadinene	-0.96429
Unigene0102915	germacreneA	-0.95804
Unigene0102915	beta-sesquiphellandrene	-0.96777
Unigene0102915	gamma-cadinene	-0.96861
Unigene0102915	alpha-cedrene	-0.9622
Unigene0102915	nerolidol	-0.95589
Unigene0102915	alpha-bisabolol	-0.96582
Unigene0102915	alpha-curcumene	-0.96282
Unigene0102915	alpha-santalol	-0.93583
Unigene0102915	alpha-bergamotol	-0.96558
Unigene0102915	alpha-terpinene	-0.95202
Unigene0057892	camphene	-0.88278
Unigene0057892	myrcene	-0.88803
Unigene0057892	imonene	-0.88624
Unigene0057892	bicycloelemene	-0.85773
Unigene0057892	alpha-copaene	-0.91747
Unigene0057892	camphor	-0.83991
Unigene0057892	bronyl acetate	-0.84435
Unigene0057892	isobornyl acetate	-0.83027
Unigene0057892	beta-humulene	-0.81719
Unigene0057892	4-terpineol	-0.89799
Unigene0057892	aromadendrene	-0.87234
Unigene0057892	isoborneol	-0.82112
Unigene0057892	beta-farnesene	-0.82087
Unigene0057892	borneol	-0.81175
Unigene0057892	bicyclogermacrene	-0.85712
Unigene0057892	geraniol	-0.84551
Unigene0057892	beta-bisabolene	-0.852
Unigene0057892	germacrene	-0.87936
Unigene0057892	delta-cadinene	-0.87414
Unigene0057892	germacreneA	-0.86316
Unigene0057892	beta-sesquiphellandrene	-0.86997

Unigene0057892	gamma-cadinene	-0.87277
Unigene0057892	alpha-cedrene	-0.83966
Unigene0057892	nerolidol	-0.85358
Unigene0057892	alpha-bisabolol	-0.85002
Unigene0057892	alpha-curcumene	-0.85553
Unigene0057892	alpha-santalol	-0.82146
Unigene0057892	alpha-bergamotol	-0.84583
Unigene0057892	alpha-terpinene	-0.88283

Draft

App Table 6.A list of seven candidate gene expressions in two datasets.

Detected method	Target gene	FSP	FM1P	FM2P	FSS	FM1S	FM2S
RNA-Seq	<i>AvWRKY40</i>	1	1.17179	1.449892	0	0.178048	0.11466
RT-qPCR	<i>AvWRKY40</i>	1	0.310502*	0.246396**	0.005443	0.022804	0.00574
RNA-Seq	<i>AvWRKY45</i>	1	0.827908	0.807535	0.5578	0.696417	0.701313
RT-qPCR	<i>AvWRKY45</i>	1	0.291341	0.523181	0.042027	0.10615	0.342045*
RNA-Seq	<i>AvWRKY31</i>	1	1.414113	1.253509	3.262917	2.518332	3.099192
RT-qPCR	<i>AvWRKY31</i>	1	0.228615**	0.287717	0.286891	0.35582	0.088195
RNA-Seq	<i>AvWRKY28</i>	1	0.994001	0.665244	0.314375	0.160244	0.137592
RT-qPCR	<i>AvWRKY28</i>	1	0.027708**	0.241536#	0.015819	0.050359	0.025833
RNA-Seq	<i>AvWRKY61</i>	1	2.868866	2.068228	0.040304	0	0
RT-qPCR	<i>AvWRKY61</i>	1	0.444314	0.168384*	0.012982	0.020532	0.060375
RNA-Seq	<i>AvNeoD</i>	1	0.567217	0.926901	0.191636	0.060469	0.207686
RT-qPCR	<i>AvNeoD</i>	1	0.317907**	0.543577##	0.012333	0.010986	0.008365
RNA-Seq	<i>AvGerD</i>	1	0.738863	0.82851	0.046062	0	0.209664
RT-qPCR	<i>AvGerD</i>	1	1.325734	1.057062	0.721395	2.219387	1.145496

Comparing with FSP, *represented $p < 0.05$, **represented $p < 0.01$; Comparing with FM1P, # represented $p < 0.05$. Comparing with FSS, *represented $p < 0.05$.

App Table 7. A list of the candidate gene identities were blasted to *Arabidopsis thaliana* protein databases.

Name		Score	Query coverage	E value	Identities	Annotation
<i>AvNeoD</i>	Unigene0078114	191	100%	1e-59	55%	(+)-neomenthol dehydrogenase
<i>AvWRKY40</i>	Unigene0102915	184	84%	9e-56	45%	WRKY40
<i>AvWRKY31</i>	Unigene0062811	266	79%	8e-82	51%	WRKY6
<i>AvWRKY45</i>	Unigene0057892	168	63%	2e-53	85%	WRKY75
<i>AvWRKY61</i>	Unigene0037009	164	51%	4e-46	66%	WRKY72
<i>AvWRKY28</i>	Unigene0044272	163	41%	4e-47	67%	WRKY28

App Table 8. The interactions of the candidate WRKY genes.

Genes names	Reference <i>Arabidopsis thaliana</i> genome	Interaction
<i>AvWRKY31</i>	<i>AtWRKY6</i>	GRF2
<i>AvWRKY61</i>	<i>AtWRKY72</i>	-
<i>AvWRKY28</i>	<i>AtWRKY28</i>	-
<i>AvWRKY40</i>	<i>AtWRKY40</i>	WRKY18, wbox-arath-dna-3, WRKY38, q8l7y2_arath, WRKY60, WRKY40, GRF2
<i>AvWRKY45</i>	<i>AtWRKY75</i>	a1a6g7_arath, T8M16_210, BHLH27, VQ20, EMB1401

The interaction was presumed according to the *Arabidopsis thaliana*.

App Table 9. A list of reference gene primers used in RT-qPCR to normalize candidate gene transcript levels.

Reference gene's name	Primer sequences from 5'to 3'	Tm/°C	Product size/bp	Unigene ID
Actin	GTTCTTAGTGGCGGTTCAA AGCAGGACCAGATTCTTCAT	57°C	213	0133538
TUA	GGAGGATGCGGCAAAACAA AGCAAGGAACCCAGCCCAGA	56°C	171	0093134
EF-1 α	GAAAGAAGCAGCCGAGAT AACCGCCAGTGGTAGAAT	56°C	239	0019582

App Table 10. A list of target gene primers used in RT-qPCR to measure transcript levels.

Target genes name	Primer sequences from 5'to 3'	Tm/□	Product size/bp	Unigene ID
WRKY28	AAGAAGGAAAGGATGGGAGA TGAGCTGTGCAGCGGTAG	60□	233	0044272
WRKY31	AGCTGGAGCTGACCCGTAT CCATCCTCAACCTCTTTCG	60□	172	0062811
WRKY61	TTCCAACAACTCCCACC TCTGCCCGTATTTCCTCC	57□	191	0037009
(+)-neomenthol dehydrogenase	ATCCCTCTGCTTCAGTCATC GGTCCAACCTCCCTTCCT	53.7□	187	0078114
WRKY40	TGGTGGTGAAAGATGGGT AAGGTTGGCTGTGGTTGT	58□	192	00102915
WRKY45	GCCAGAAAGCCGTCAAGAAC TGGGATGACTGTGC	58□	92	0057892
(+)-germacrene D synthase	TAATCTCCTCTGGGTGTTCT CATCTGTGCCATACTCTTTC	60□	200	0106615

App Table 11. A list of reference gene amplification efficiencies and correlation coefficients were detected by CFX96 Real-Time PCR System.

Gene name	Full name	Amplification Efficiency/ E	Correlation coefficient/ R ²
<i>AvActin</i>	Actin	96.3%	0.990
<i>AvTUA</i>	α tubulin	98.0%	0.993
<i>AvEF-1α</i>	transcription elongation factor-1 α	104.5%	0.990

All reference gene amplification efficiencies were 90% to 105%, and correlation coefficients were more than 0.980, which conformed to further experiments required.

App Table 12. A list of reference gene stability values about three reference genes.

Target	Coefficient Variance	M Value
<i>AvActin</i>	0.6795	2.3373
<i>AvTUA</i>	0.9891	2.3157
<i>AvEF-1α</i>	1.7326	2.9165

AvActin was the most stable in coefficient variance. However, *AvTUA* was the most stable in M value.

Thus, we used both *AvActin* and *AvTUA* as double reference genes for our further experiments.

App Table 13. A list of target gene amplification efficiencies and correlation coefficients.

Gene name	Full name	Amplification Efficiency/ E	Correlation coefficient/ R ²
<i>AvWRKY28</i>	WRKY transcription factor 28	99.1%	0.992
<i>AvWRKY31</i>	WRKY transcription factor 31	92.0%	0.993
<i>AvWRKY61</i>	WRKY transcription factor 61 isoform X3	102.4%	0.987
<i>AvNeoD</i>	(+)-neomenthol dehydrogenase	104.9%	0.995
<i>AvWRKY40</i>	WRKY transcription factor 40	100.2%	0.981
<i>AvWRKY45</i>	WRKY transcription factor 45	104.3%	0.990
<i>AvGerD</i>	(+)-germacrene D synthase	98.5%	0.980

All target gene amplification efficiencies were 92% to 105%, and correlation coefficients were more than 0.980, which conformed to further experiments required.

43x23mm (300 x 300 DPI)