Available online: www.notulaebotanicae.ro

Print ISSN 0255-965X; Electronic 1842-4309

Not Bot Horti Agrobo, 2019, 47(3):772-783. DOI:10.15835/nbha47311428

Original Article

Transcriptome-based Discovery of AP2/ERF Transcription Factors Related to Terpene Trilactones Synthesis in *Ginkgo biloba*

Linling LI[1,2,3a], Honghui YUAN[1,2b], Sanxing ZHA[1,2,3c], Jie YU[1,2,3], Xian XIAO[1,2,3], Kun DENG[1,2,3], Shuiyuan CHENG[3]*, Hua CHENG[1,2]*

[1]*Economic Forest Germplasm Improvement and Comprehensive Utilization of Resources of Hubei Key Laboratories, Hubei Huanggang, 438000, China; lilinling1437@126.com; 448503382@qq.com; 781634982@qq.com; 365890655@qq.com; 1814759739@qq.com; harry1437@126.com*

[2]*Hubei Collaborative Innovation Center for the Characteristic Resources Exploitation of Dabie Mountains, Huanggang, 438000, China; chenghua1437@126.com (*corresponding author)*

[3]*Wuhan Polytechnic University, College of Biology and Pharmaceutical Engineering, Wuhan, 430023, China; s_y_cheng@sina.com (*corresponding author); [a,b,c]These authors contributed equally to this work*

Abstract

Ginkgo biloba is a unique tree in China with medicinally and phylogenetically important characteristics. Terpene trilactones (TTL) is a key active pharmaceutical ingredient in *Ginkgo*, so the content of TTL in *Ginkgo* has become one of the important indices for evaluating quality of the medicinal materials. By transcriptome sequencing on samples treated by chlormequat, ultraviolet (UV) and drought, totally 59820 contigs and 37564 unigenes were obtained. Furthermore, 18234 unigenes were annotated through COG, KEGG and GO analysis. There were 78 AP2/ERF transcription factors, 23 factors of up-regulation and 66 factors of down-regulation that were related with synthetic pathway of TTL in *Ginkgo*. Phylogenetic tree clustering analysis indicated that there were 42 AP2s could be clustered into ERF, DREB and RVA subfamilies. EMSA analysis demonstrated that GbERF13, GbERF25 and GbERF27 could bind with regulatory elements, such as E-box, in the upstream of GbMECPs promoter. Expression analysis showed that the expression level of GbERF25 was the highest in root, and GbERF25 and GbERF27 were expressed in relatively high transcription levels in leaf and other tissues. The results of qRT-PCR indicated that CCC treatment could significantly improve expression levels of ERF25 and ERF27, and UV and drought could induce transcription levels of ERF13 and ERF25, respectively. The results implied that ERF25 and ERF27 might involve in the induction and regulation of CCC treatment on synthesis of bilobalide in *G. biloba*. ERF13 might participate in the regulation of bilobalide synthesis induced by UV, and EFR25 might involve in the regulation of the synthesis induced by drought. During annual cycle of expression, the transcription levels of ERF13, ERF25 and ERF27 had significantly positive correlation with diterpene level with correlation coefficient 0.975. It implied that these transcription factors mainly acted on the MEP pathway that regulated synthesis of bilobalide. The aim of the research was to indicate the mechanism of environment or cultivation measure regulating target gene of TTL metabolic pathway by AP2/ERF, and establish metabolic network of AP2/ERF regulating TTL synthesis.

Keywords: AP2/ERF; *Ginkgo biloba*; MEP pathway; terpene trilactones; transcriptome sequencing

Abbreviations: COG-Clusters of Orthologous Groups; GO-Gene ontology; KEGG-Kyoto Encyclopedia of Genes and Genomes; KOG-EuKaryotic Orthologous Groups; DEGs-Differentially Expressed Genes; MEP-2-C-Methyl-D-erythritol 4-phosphate; MVA-Mevalonate pathway; NR-NCBI non-redundant; RNA-Seq RNA sequencing; SRA-Sequence read archive; TTL-Terpene trilactones

Introduction

Ginkgo biloba is the only living species within the order Ginkgophyta in gymnosperm, which originated about 300 million years ago (Magallón *et al*., 2013). Although *G. biloba* is a relic species in China, it has been transplant worldwide as a land scape tree. Additionally, *Ginkgo* is widely cultivated as a medicinal tree because its leaves contain a large number of active compounds, as flavanol glycosides and lactones terpene, which Chinese traditional medicine has demonstrated to improve blood circulation and discourage clot formation (Singh *et al*., 2008).

Received: 14 Jan 2018. Received in revised form: 16 May 2019. Accepted: 20 May 2019. Published online: 29 May 2019.

Ginkgolides and bilobalide, collectively termed terpene lactones(TTLs), are highly modified diterpene lactones uniquely found in *G. biloba* and function as potent antagonists of platelet activating factor (Kim *et al*., 2006; Chadwick *et al*., 2013; Liao *et al*., 2015). The biosynthetic pathway of terpene lactones are biosynthesized from the universal five carbon skeleton-isopentenyl diphosphate (IPP) and its isomer dimethylallyl diphosphate (DMAPP). Two distinct IPP/DMAPP biosynthesis pathways exist in plant (Strømgaard and Nakanishi 2004). The methylerythritol phosphate (MEP) pathway in the plastid facilitates formation of monoterpenoids and diterpenoids, whereas the mevalonic acid (MVA) pathway in the cytosol facilitates the synthesis of sesquiterpenoids and sterols (Lange *et al*., 2000; Kim *et al*., 2006). It is suggested that bilobalide is a sesquiterpene which contains three lactones units, ginkgolide is a diterpenoids which contains four lactones units (Sabater-Jara *et al*., 2013). Though ginkgolides are present in both leaves and roots, genes coding for enzymes of the MEP pathway involved in IPP and DMAPP are highly expressed predominantly in the roots. Ginkgolides would be synthesized in the roots and then transported to the stems and leaves (Kim *et al*., 2012).

The recently developed high-throughput messenger RNA sequencing technology has been extensively employed to study genome-wide gene expression in *Ginkgo* (Han *et al*., 2015; Wang *et al*., 2016). RNA-Seq can be used not only to detect and analyze transcripts in organisms with sequenced genomes but also to assemble short reads for gene discovery and gene expression analysis in organisms without reference genomes (Xu *et al*., 2012). AP2/ERF transcription factors belong to one of the largest plant transcription factor families, and are characterized by conserved AP2/ERF DNA binding domains of 57-66 amino acids in size (Okamuro *et al*., 1997). The AP2/ERF genes constitute a large multigene family divided into four subfamilies named AP2, ERF, CBF/DREB and RAV based on their sequence similarities and numbers of AP2/ERF domains(Sakuma *et al*., 2002). AP2/ERFs family frequently involved in the response to both biotic and abiotic stress (Gao *et al*., 2015). Recent studies revealed the role of some AP2/ERF proteins are also important regulators in the biosynthetic pathway of terpenoids. For instance, the ORCA proteins in *Catharanthus*, control both the primary plastidial isopentenyl pyrophosphate pathway and the secondary TIA pathways (Van and Memelink, 2000). AaORA is a valuable AP2/ERF transcription factor, not only for the regulation of the artemisinin biosynthetic pathway, but also for the enhancement of the disease resistance to *B. cinerea* in *A. annua* (Lu *et al*., 2013). In sweet orange, CitAP2.10 regulates (+)-valencene synthesis via induction of CsTPS1 mRNA accumulation. AaERF1 and AaERF2 induced artemisinin synthesis by binding to the promoter of AaADS directly (Yu *et al*., 2012). However, none of these genes were reported to be involved in lactones terpene biosynthesis.

However, AP2/ERF genes that control TTL synthesis and response to stress in *G. biloba* have not been reported. In this study, using a Ginkgo transcriptome sequence database, about seventy-eight GbAP2/ERF contigs were isolated and most of them were found to be differentially expressed during UV, Drought, CCC treatment. In the present study, we isolated several AP2/ERF gene, designated as GbERF13, GbERF25 and GbERF27 from *G. biloba*. A potential role for ERF25 and ERF27 in regulating *Ginkgo* TTL production was examined by transient overexpression in *Ginkgo* leaves and the action on the promoter of the target gene GbMECP was studied using a EMSA technology. The expression pattern and ectopic expression functional analyses undertaken may contribute to understand the function of AP2/ERF in control of TTL synthesis and tolerance of abiotic stress.

Materials and Methods

Plant materials

Thirteen-years old of grafted *G. biloba* seedlings were growing in a greenhouse in Huanggang (E, 114°54'-116°8', N, 29°45'-31°35', Hubei, East of China). The seedling from the same genotypic strain of *G. biloba* were subjected to treatments with UV-B, Drought and Chlormequat (CCC). For UV-B treatment, seedlings were exposed to 1500 J/m^2 UV-B irradiation in a closed chamber, and the control cuttings were placed in a dark closed chamber. For drought treatment, samples were collected as the relative water content decreased in soil, the whole dehydration stress lasted for 15 days. The CCC(10 mM) was dissolved in 0.01% Tween 20 and sprayed onto cutting leaves. The control leaves were sprayed with an equivalent volume of 0.01% (v/v) Tween 20. All samples were immediately frozen in liquid nitrogen and stored at -80 °C prior to total RNA extraction, transcriptome sequencing, and expression analysis.

Total RNA extraction and transcriptome sequencing

Total RNA was extracted from tender leaves with the CTAB method according to the description by Wang (2005). The mRNA was enriched with magnetic beads coated with oligo dTn, and then fragmented, which was reverse transcribed to first-stranded cDNA. Double-stranded cDNA was synthesized with random hexamer primers, purified and added with sequencing linkers. Fragments of the correct size were purified with Universal DNA Purification Kit, and the sequencing libraries were prepared with the RNA-Seq Library Construction Kit. The quality and quantity of the libraries were verified using an Agilent 2100 Bioanalyzer and ABI real time RT-PCR. The qualified cDNA libraries were used for paired-end sequencing on an Illumina HiSeq 2000 platform.

Electrophoretic mobility shift assay

The E-box sequence in the promoter sequence of GbMECP was identified as TTTGAC using the PLACE and the Signal Scan Program PlantCARE database(Postel *et al*., 2002). Oligonucleotides of E-box sequence were synthesized and labeled with biotin (Shanghai Sangon Biotech Co., Ltd., China) for chemiluminescence using a LightShift® Chemiluminescent EMSA Kit (Pierce Biotechnology, Thermo Fisher Scientific Inc., Rockford, USA). After labeling, complementary labeled strands were mixed together in an equimolar ratio and annealed at 25 °C after denaturation at 90 °C. Electrophoretic mobility shift assays (EMSA) were performed by incubating 0.5 ng of

labeled probe with recombinant AP2/ERF protein and competing oligonucleotides in binding buffer (10 mM Tris–HCl, pH 7.5, 50 mM NaCl, 1 mM dithiothreitol, 1 mM EDTA, 5% glycerol, and 1 μg/μL poly (dI•dC)) at 25 °C for 20 min. Mixtures were size-fractionated on a non-denaturing polyacrylamide gel followed by drying and transferring into nitrocellulose membranes and detecting by streptavidin HRP/ chemiluminescence for biotin labeled probes.

Gene expression analysis

Analysis RT-PCR was performed to measure the relative transcript levels of selected genes. First-strand cDNA synthesis was con-ducted in triplicate for each sample according to the manufacturer's instructions (PrimeScript™ RT Reagent Kit, TaKaRa, Dalian, China). The house keeping gene, namely glyceraldehyde- 3-phospate dehydrogenase gene (GbGAPDH, L26924), was used as a reference gene with primers of GAPU and GAPD. Gene-specific primers of AP2/ERFs were list in Table 1. The experiments were repeated independently with three biological replicates using SYBR® Premix Ex Taq™ II (TaKaRa, Dalian, China) in the Applied. Biosystems 7500 Realtime PCR systems (Foster City, CA, USA). Relative expression level was calculated according to the method described by Schmittgen (2008).

Extraction and determination of Terpene lactones

Ginkgolide A(GA), ginkgolide B(GB), ginkgolide C(GC) and bilobalide BB were extracted and determined using gas chromatography with a wide bore capillary column methods described by Liao (2008). The content of terpene lactones was the sum of the contents of GA, GB, GC, and BB and expressed as mg/g DW, and diterpenoids was the contents of BB. All the tests were carried out in triplicate, and data represent the means standard errors.

Data annotation and statistical analysis

Raw sequence reads were filtered by the Illumina pipeline. Then the clean reads were subjected to the TopHat2-Cufflinks-Cuffmerge-Cuffdiff standard pipeline for identification of DEGs. These DEGs were blasted against the GO, KEGG, and COG of proteins database for enrichment analysis (Jie *et al*., 2015).

The sequence was analyzed by the bioinformatic software on websites (*http://www.ncbi.nlm.nih.gov*). Plant AP2/ERF protein sequences were retrieved from NCBI GenBank. Sequence alignment was performed using Clustal X 2.0 and phylogenetic tree was constructed by neighbor-joining method using MEGA version 7. A bootstrap statistical analysis was carried out with 1000 replicates (Kumar *et al*., 2016). The experimental results were expressed as mean ± standard error (SE) of three replicates. The data were analyzed using Statistical Product and Service Solutions (SPSS, Version 17.0) at $P < 0.05$ (Fisher's protected least significant difference).

Results

High-throughput transcriptome sequencing and sequence assembly

To obtain an over view of the *G. biloba* transcriptome, a cDNA library was generated from RNA isolated from the leaves. Illumina RNA-Seq generated a total of 40,285,768 raw reads. After stringent quality checks and data cleaning, we obtained 35,434,842 clean reads containing a total of 4,429,355,250 nucleotides (4.4 Gb) for further analysis. The Q30 percentage (sequencing error rate, 0.1%), GC (guanine+cytosine) percentage, and N percentage were 98.59%, 40.77%, and 0.07%, respectively.

Based on these clean reads, 234,323 contigs were assembled with a mean length of 578 bp. The N50 of contigs was 1021 bp. These contigs were further assembled with the Trinity software and clustered into unigenes. Finally, the denovo assembly yielded 195,446 unigenes with a mean length of 502 bp, including 22,522 clusters and 55,376 singletons of the 77,898 unigenes, 8123 were larger than 2000 bp. The N50 of the unigenes was 712 bp (Table 2).

Functional classification by GO and COG

Functional annotation of unigenes GOES is a useful program for annotating and analyzing the functional categorization of annotated genes. To facilitate the organization of the cranberry transcripts into putative functional groups, GO terms were assigned using Blast GO. A total of 27,286 unigenes (47.59%) were assigned into GO ontologies based on their similarity to sequences with previously known functions, including 31,474 sequences assigned to the molecular function category, 104,552 to the biological process category and 83,221 to the cellular component category (Table 2 and Fig. 3).

The COG database can phylogenetically predict and classify the completeness of the unigene sequences. Based on sequence homology, 19,176 unigenes (31.5% of all the unigenes) were assigned into 67 COG categories (Fig. 4).

KEGG pathway mapping KEGG analysis can help us

Table 1. The gene analysis primers of AP2/ERF

Description	Serial Name	Sequences(5'-3')
Primers for AP2/ERF	ERF13U	TGAATGTGAAGACTCCAGCCAA
	ERF13D	ACTGGTCAGATTGGCAGGCTT
	ERF25U	GATGTCTAACACCCCCAACGAG
	ERF25D	CATTGCGTGACGGAAAAAACT
	ERF27U	GGACTCCAAACTACAACAAGCC
	ERF27D	CCGTTCGTATTTTTGGATTGTC
Primer for reference gene	GAPU	TAGGAATCCCGAGGAAATACC
	GAPD	TTCACGCCAACAACGAACATG

further understand the specific processes, gene functions and gene interactions at a tran-scriptome level. In our study, 24,846 all unigenes were mapped to 306 predicted metabolic pathways through the KEGG database. The largest category was metabolic pathways (3869) which included biosynthesis of secondary metabolites (1818), Microbial metabolism in diverse environments (1029), ether lipid metabolism (355), starch and sucrose metabolism (306), purine metabolism (501), pyrimidine metabolism (372) and other subcategories.

In the secondary metabolism category, 26 subcategories comprised 2,0674 unigenes, the most represented of which were Posttranslational modification, protein turnover, chaperones (2990), flavonoid biosynthesis (98), limonene and pinene degradation (29), stilbenoid, diarylheptanoid and gingerol biosynthesis (10), terpenoid backbone biosynthesis (294), zeatin biosynthesis (15), carotenoid biosyn-thesis (130), and flavone and flavonol biosynthesis (108). Moreover, isoflavonoid biosynthesis (53), folate biosyn-thesis (34), anthocyanin biosynthesis (26), and betalain biosynthesis (6) were also classified.

These results suggest that the secondary metabolic processes are active path-ways in *Ginkgo* flavonoid and terpenoid biosynthesis. In addition to metabolism pathways, genes involved in genetic information processing (6,781) and cellular processes (1,718) were highly represented categories. Moreover, AP2/ERF transcript factor (78) was represented categories.

Table 2. Summary of the *Ginkgo biloba* leaf transcriptome treatment by UV, CCC and drought

Summary	Number	Percentage
Total reads	40,285,768	
Clean reads	35,434,842	94.00%
Clean Bases Number	4,429,355,250	90.90%
Raw Q30 Bases Rate (%)		98.59%
GC percentage		40.77%
N percentage		0.07%
Total Unigene	195,446	
Unigene N50	1021	
Unigene N90	224	
Min unigene	201	
Max unigene	16298	
Mean unigene	578.69	

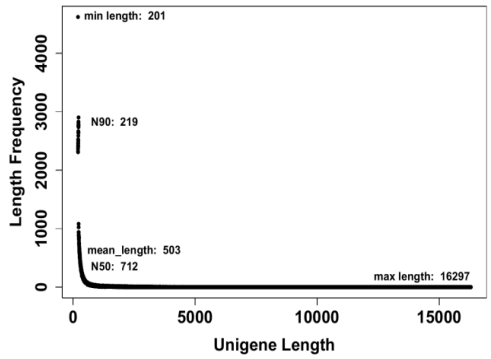

Fig. 1. Unigenes length distribution

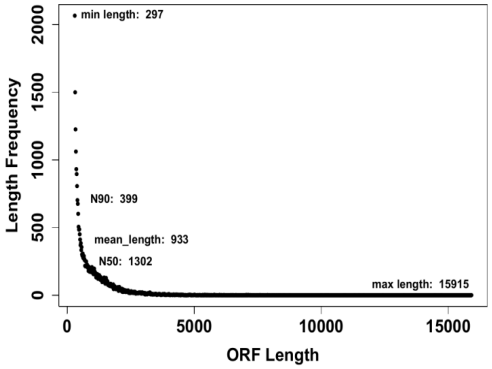

Fig. 3. ORF length distribution

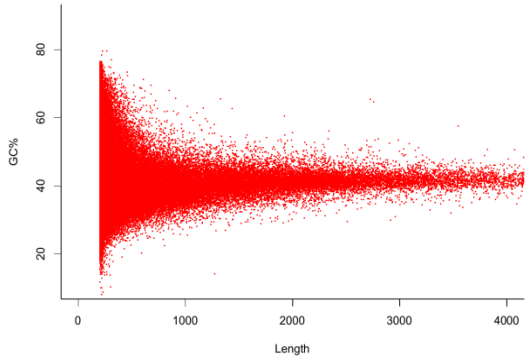

Fig. 2. GC distribution in different length unigenes

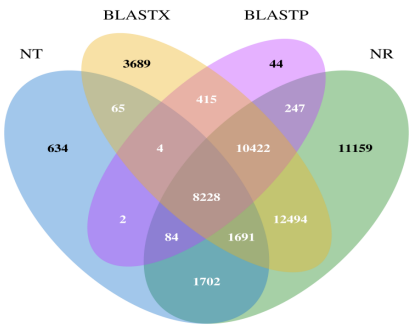

Fig. 4. Annotation results compared by Wayne

Candidate genes involved in terpenoid biosynthesis

Given that *Ginkgo* are rich in terpenoids, we focus on identifying the candidate genes involved in terpenoid biosynthesis (Table 3). KEGG analysis result showed that, about one unigene can attributed to monoterpenoid biotynthesis pathway when treatment with CCC. In the sesquiterpenoid and treterpenoid biosynthesis, have four unigenes after treatment with CCC, UV-B and Drought, and only three in CK. Diterpenoid pathway, also involve in abundant unigenes, about 30 unigenes in CCC, 21 in UV-B, 24 in drought and 16 in CK. In total terpenoid backbone biosynthesis, about 97, 86, 94 unigenes attributed to CCC, UV-B and Drought treatment, 89 unigenes in CK.

Candidate transcription factors involved in terpenoid biosynthesis and transport

In particular, the expression of structural genes involved in terpenoid biosynthesis is largely controlled by basic transcription factors, like AP2/ERF, bZIP, WRKY, bZCT and bHLH (Table 4). Base on our results, a total of unigenes were identified as putative TFs or regulators, with of the unigenes belonging to 65 subclasses. Analysis results indicated, about 296 zinc finger in all the metabolic pathways, and 64 unigenes were in bHLH family, 78 in AP2-ERFs family, 26 in MADS-box, 110 in MYB, 57 in WRKY, 50 in bZIP, 27 in DOF, 2 in HSF.

In the 78 AP2/ERF transcription factors, there were 5 factors up-regulated by CCC, UV and drought: ERF13-c120696_g1, ERF16-c148588_g1, ERF17-c108030_g1, ERF27-c147867_g1, and ERF28-c241099_g1, and 2 down-regulated by them: ERF9-c130822_g2 and ERF15-c138853_g1. Besides, there was no significant difference in other transcription factors treated by them (Table 5).

Fig. 7 and Table 6 shows phylogenetic relationships of AP2/ERF family transcription factors that belong to the thirteen major subfamilies and that include family members from Arabidopsis(Mizoi *et al*., 2012) (Phytozome v7.0, http://www.phytozome.net/). Each branch represents a group of homologues that are considered to be diversified in each plant group. According to the phylogenetic tree, *Ginkgo* ERF genes were classified in three subfamilies consisting of twenty-one DREB (ERF family group A-4 to A-6), eighteen ERF (ERF family group B-1 to B-6) genes and four ERF (RVA family group). ERF subfamily genes were twice as large as the DREB subfamily in *Ginkgo*.

Table 3. the numbers of candidate genes involved in terpenoid biosynthesis in *Ginkgo* transcriptome

Pathway	Number of annotated sequences			
	CK	CCC	UV-B	Drought
Monoterpenoid biosynthesis	0	1	0	1
Sesquiterpenoid and triterpenoid biosynthesis	3	4	4	4
Terpenoid backbone biosynthesis	89	97	86	94
Diterpenoid biosynthesis	16	30	21	24
Ubiquinone and other terpenoid-quinone biosynthesis	51	63	53	62

Table 4. the numbers of transcription factors involved in terpenoids biosynthesis and transport in the *Ginkgo* transcriptome

Family	Number of annotated sequences	Family	Number of annotated sequences
Zinc finger	296	MYB	110
bHLH	64	WRKY	57
AP2/ERF	78	MADS-box	26
bZIP	50	DOF	27
HSF	2		

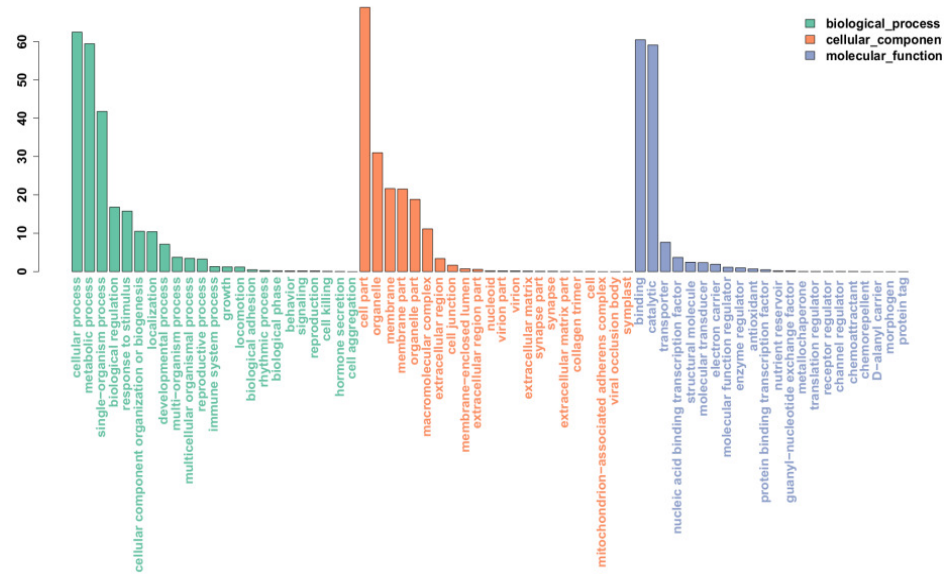

Fig. 5. Histogram of GO analysis of all-unigenes

Table 5. AP2/ERF transcription factor involved in terpenoids biosynthesis

Seq No.	CCC	UV-treatment	Drought	GO analysis
ERF1, c138101_g2	U	D	D	ethylene mediated signaling pathway
ERF2, c132104_g1	D	U	U	cellular response to potassium ion
ERF3, c115261_g1	U	U	D	response to stress; response to jasmonic acid stimulus; response to cytokinin stimulus
ERF4, c171185_g1	D	D	U	defense response
ERF5, c39941_g1	U	U	U	ethylene mediated signaling pathway
ERF6, c118208_g1	U	U	D	triglyceride biosynthetic process; response to sucrose stimulus
ERF7, c103957_g1	U	D	D	triglyceride biosynthetic process; response to sucrose stimulus
ERF8, c132204_g1	U	U	D	defense response to fungus
ERF9, c130822_g2	D	D	D	leaf development; cytokinin mediated signaling pathway
ERF10, c143672_g1	U	D	D	positive regulation of transcription; response to cytokinin stimulus; cell death
ERF11, c120638_g1	D	U	D	phyllotactic patterning; positive regulation of embryonic development; seed germination
ERF12, c46523_g2	U	D	D	defense response
ERF13, c120696_g1	U	U	U	defense response; ethylene mediated signaling pathway
ERF14, c101631_g1	D	U	U	negative regulation of ethylene mediated signaling pathway
ERF15, c138853_g1	D	D	D	cellular response to potassium ion; response to reactive oxygen species
ERF16, c148588_g1	U	U	U	response to chitin; defense response
ERF17, c108030_g1	U	U	U	ethylene mediated signaling pathway
ERF18, c93386_g1	U	D	U	ethylene mediated signaling pathway
ERF19, c120449_g1	D	U	D	ethylene mediated signaling pathway
ERF20, c116163_g1	D	U	D	ethylene mediated signaling pathway
ERF21, c108930_g1	D	U	D	cellular response to freezing; cellular response to heat; response to salicylic acid stimulus
ERF22, c108414_g4	D	U	D	ethylene mediated signaling pathway
ERF23, c236672_g1	U	D	U	ethylene mediated signaling pathway
ERF24, c108930_g2	D	U	U	response to cold; response to water deprivation
ERF25, c133766 g1	U	D	U	abscisic acid mediated signaling pathway; response to osmotic stress; response to water deprivation; seed development
ERF26, c128149_g2	U	D	U	
ERF27, c147867_g1	U	U	U	defense response to bacterium; response to cold; response to chitin; regulation of cell death
ERF28, c241099_g1	U	U	U	defense response to fungus

Table 6. the *Ginkgo* AP2-ERF sequence belong to the different major subfamilies

Subfamily	Sequence ID	Subfamily	Sequence ID
ERF subfamily B-1	c116163 g1	DREB subfamily A-4	c93386 g1,c133868 g1
ERF subfamily B-2	c147248 g1, c125406 g1, c135433 g1,c118198 g1, c143672 g1	DREB subfamily A-5	c107557 g1,c108930 g1, c126357 g2,c126357 g1, c125378 g1,c147867 g1, c167391 g1,c144480 g2, c120449 g1,c133766 g1, c101285 g1, c142926 g2, c142926 g1,c143946 g1
ERF subfamily B-3	c102943 g1,c108414 g4, c120696 g1	DREB subfamily A-6	c227857 g1, c30291 g1, c171185 g1,c93063 g1, c122280 g1
ERF subfamily B-4	c132204 g1, c130213 g2, c146510 g1	RAV	c135072 g1, c121906 g1, c130406 g1,c56313 g1
ERF subfamily B-6	c99947 g1, c132104 g1, c87988 g1, c132411 g1, c78931 g1, c23831 g1		

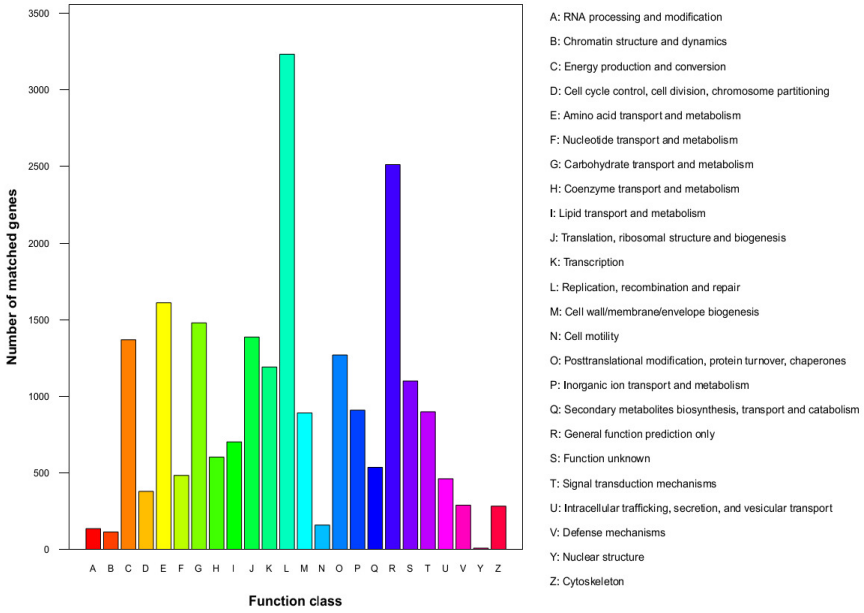

Fig. 6. Histogram of COG classification of all-unigenes

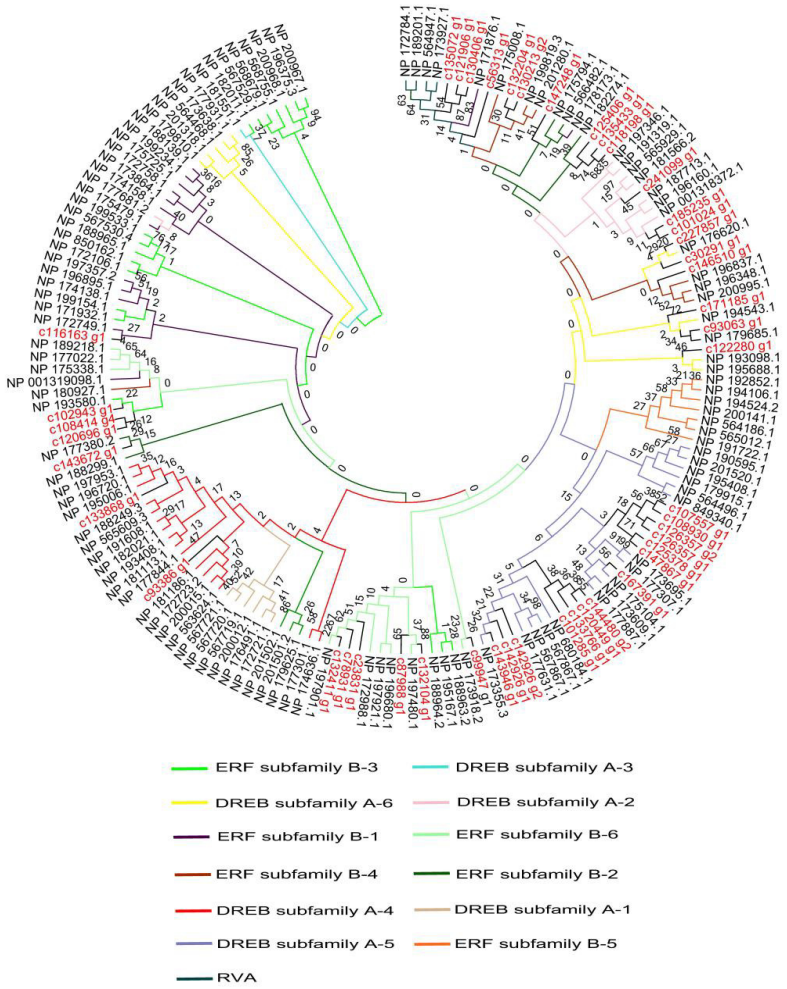

Fig. 7. Phylogenetic tree of the AP2/ERF family transcription factors in *Ginkgo* and *Arabidopsis thaliana*. The NJ tree was constructed from the amino acid sequences of the AP2/ERF domain using the MEGA7 program based on the Clustal X. A consensus tree (after 1000 bootstrap samplings) is shown, and support values are indicated on the sides of important nodes. The scale bar indicates the substitution rate per residue. Peptide sequence information of *A. thaliana* was obtained from Phytozome (Phytozome v7.0, http://www.phytozome.net/) and transcriptone sequence ID of *Ginkgo*.

Impact of treatment on secondary metabolites and networks of metabolites

To investigate the molecular change that take place in the *Ginkgo* under UV, CCC and drought treatments, we compared the transcriptome of the treatments data, a strong relationship was found between the RNA-seq and qPCRs gene expression values of TTLs synthesis genes selected for validating the transcriptomic dataset. The co-expression networks represent the relationship between TTLs synthesis gene and AP2/ERF during different treatments. Nodes represent selected AP2/ERF genes and edges represent relationship between any two metabolites of TTLs pathway. The clustering coefficient was 0.12, 0.33, 0.07 and 0.52 in AP2/ERF6, AP2/ERF13, AP2/ERF25 and AP2/ERF27, respectively (Fig. 8).

EMSA assay of AP2-ERF

EMSA was performed by using labeled E-box as a probe with ERFs proteins to confirm whether or not the putative E-box in the GbMECP promoter interacts with ERFs proteins (Fig. 9). 28 AP2-ERF gene family cloned from *G. biloba* transcriptome data, and eight of these AP2-ERFs genes have been identified to be involved in the biosynthetic pathway of TTLs. Upon expression in *E.coli*, GbERFs were a major protein product in the total cellular soluble protein. The molecular weight of ERFs proteins were estimated to be 23-27 kDa with his-tag, and their size agreed with that of the predicted peptide. The interaction between GbERFs and the MECP promoter sequence were assayed with EMSA. GbERF13, GbERF25 and GbERF27 specifically bound with the E-box sequence. No binding bands were detected with crude proteins of *E.coli* with the empty vector pET28a or other ERFs protein. These results confirmed that the GbERF13, GbERF25 and GbERF27 proteins can combine with the E-box sequence of GbMECP promoter sequence.

Expression profile of the AP2-ERFs gene in different tissues

In order to prove the correlation between ERF and terpene synthesis, the above 3 transcription factors were selected to study the expression levels in different tissues of *G. biloba* (Fig. 10). The results showed that AP2/ERF25 had the highest expression level in root, followed by stem, and the level was very low in other tissues. AP2/ERF27 had high expression level in stamen, followed by leaf and pistil. AP2/ERF13 had relatively high expression in fruit, stamen and leaf.

UV-B, drought, and CCC treatments regulate the transcription of AP2-ERFs

Previous study reported that UV-B, CCC and drought could effectively improve the content of TTL in *G. biloba*, and regulate expression of key gene that regulates synthesis of TTL. The results of induction on *GbAP2/ERFs* indicated that drought and CCC treatment could significantly induce expression of ERF25. The transcription level reached the peak after drought treatment for 12 hours. The level reached the maximum after CCC treatment for 6 h, and then declined. Contrarily, UV had inhibition effect on ERF25 expression. The expression level persistently decreased over treatment duration. Besides, UV could also significantly induce expression of ERF13, which reached the peak at 12 h and slowly declined afterward. CCC could induce expression of ERF27. The transcription level reached the maximum after 6 h, and then slowly declined. UV could also induce expression of ERF27 to a certain degree, but the effect was not significant (Fig. 11).

The relationship between AP2/ERF expression and lactone terpene accumulation during Ginkgo leaf growth

Fig. 12A illustrated that the TTL content in *Ginkgo* showed multiple peaks in annual periodic variations. The dates showing the obvious peak were on Jun. 18th, Aug. 13rd and Oct. 17th, and the corresponding contents of dry substance were 7.81 mg/g, 7.90 mg/g and 7.62 mg/g, respectively. The content of ginkgolide (diterpenes) persistently increased, and reached the maximum level on Oct. 17th (5.19 mg/g).

The transcription levels of the 3 ERFs genes showed multiple peaks within 1 year, and the seasonal variation was obvious (Fig. 12B). The transcription level variation tendency of ERF13 and ERF25 was consistent, and reached a peak on June 13th (relative transcription levels were 6.91 and 7.09, respectively). The levels reached the annual highest peak on Oct. 17th, which were 9.8 and 9.61, respectively. The change of ERF27 in annual transcription level was relatively large. The highest peak showed on Oct. 17th, and relative transcription level was 10.37.

Afterward, the correlation analysis between the transcription levels of the 3ERFs genes and TTL contents in *Ginkgo* indicated that the expression levels of them had significantly linear correlations with the changes of total triterpene content and diterpene content in *Ginkgo*. ERF25 and ERF27 levels had significantly positive correlation with total triterpene content, but ERF13 had negative correlation with the content. The correlation coefficient was 0.793, and linear equation was $Y_1=(6.717\pm0.617)+(-0.529\pm0.240)X_1+(0.23\pm0.289)X_2+(0.27\pm0.067)X_3$. The expression changes of ERF13, ERF25 and ERF27 had significantly positive correlation with diterpene content. The correlation coefficient was 0.975, and the equation was $Y_2=(0.827\pm0.198)+(0.006\pm0.077)X_1+ (0.157\pm0.021)X_2+(0.286\pm0.093)X_3$.

Table 7. The coefficient change of two model determination

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	0.793[a]	0.628	0.563	0.52920
2	0.975[a]	0.951	0.942	0.16976

Predictors: (Constant), ERF025, ERF013, ERF027
Dependent Variable: 1. Terpene lactones contents; 2. Diterpenes

Table 8. The coefficient of the test results of the two models

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	6.717	0.617		10.895	0.000
	ERF013	-0.529	0.240	-1.240	-2.204	0.042
	ERF025	0.230	0.289	0.454	0.795	0.438
	ERF027	0.270	0.067	0.693	4.040	0.001
2	(Constant)	0.827	0.198		4.181	0.001
	ERF013	0.006	0.077	0.017	0.084	0.934
	ERF025	0.157	0.021	0.458	7.322	0.000
	ERF027	0.286	0.093	0.643	3.085	0.007

1. Dependent Variable: Terpene lactones contents； 2. Diterpenes

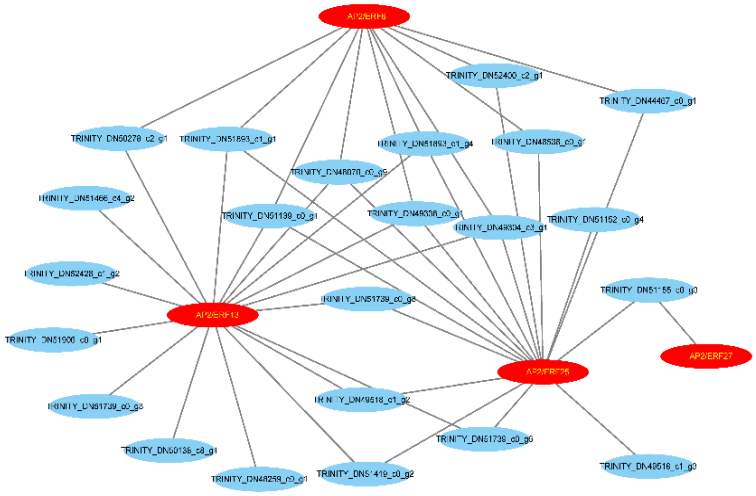

Fig. 8. Network representation of TTLs synthesis gene and AP2/ERF during different treatments. Nodes represent selected AP2/ERF genes and edges represent relationship between any two metabolites of TTLs pathway. The clustering coefficient was 0.12, 0.33, 0.07 and 0.52 in AP2/ERF6, AP2/ERF13, AP2/ERF25 and AP2/ERF27, respectively

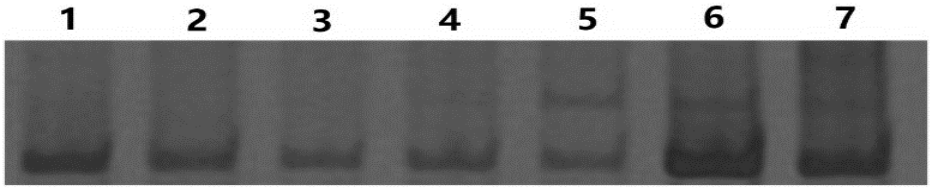

Fig. 9. Gel mobility shift assays for ERFs of the MECP promoter. Line1 is the biotin-labeled E-box probe of MECP, Line2 is *E. coli* protein-mixed biotin-labeled E-box probe, line 3 pET-28a protein-mixed biotin-labeled E-box probe, Line4 is the recombinant AP2/ERF6 mixed biotin-labeled E-box DNA probe, line5 AP2/ERF13, Line 6 is the AP2/ERF25, and Line 7 is the AP2/ERF27

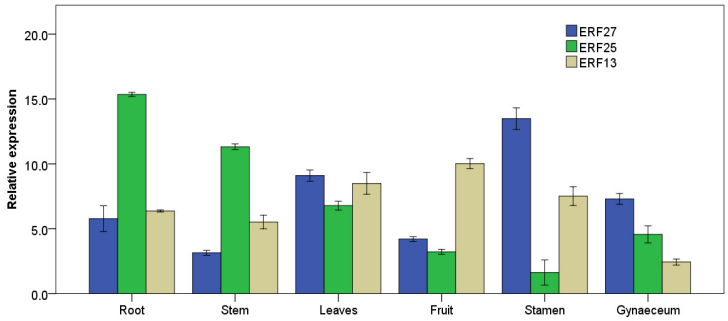

Fig. 10. The expression levels of ERF13, ERF25 and ERF27 in *Ginkgo biloba* different tissue. Relative expression levels of ERF13, ERF25 and ERF27 in different tissues with GbGAPDH gene as internal reference gene. Total RNA samples were isolated from leaves, roots, stems, male flowers, and female flowers, respectively. Each sample was individually assayed in triplicate. Values shown represent the mean reading from three plants and the error bars indicated the standard errors of the mean

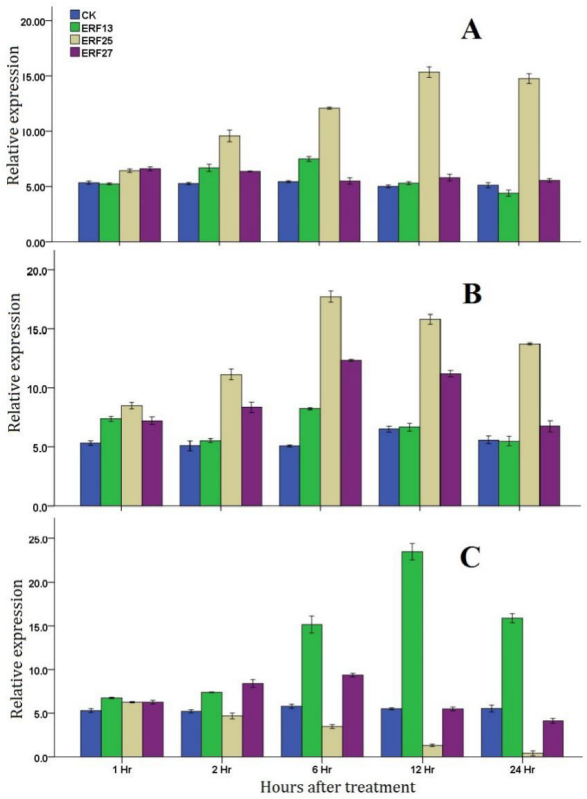

Fig. 11. Relative expression levels of GbERF13, GbERF25 and GbERF27 under drought (A), CCC (B), and UV-B (C) treatments. Relative expression of GbERF13, GbERF25 and GbERF27 at various hours after treatment under drought, CCC, and UV-B treatments with GbGAPDH gene as internal reference gene. Each sample was individually assayed in triplicate. Values shown represent the mean reading from three treated samples and the error bars indicated the standard errors of the mean

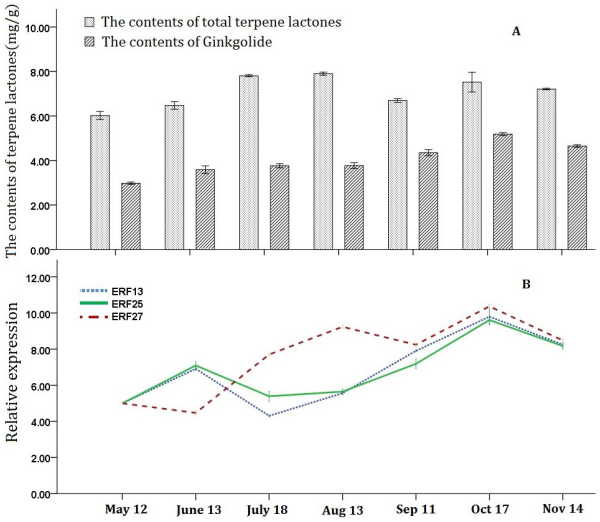

Fig. 12. Changes in lactone terpene contents, GbERF13, GbERF25 and GbERF27 expression level during growth of *Ginkgo* leaves. In the annual change of terpene in *Ginkgo* leaf, the terpene content and GbERF13, GbERF25 and GbERF27 gene expression level showed relationship. Data are mean values of triplicate tests and the error bars indicated the standard errors of the mean

Discussion

G. biloba is a unique plant species found wild in China but cultivated worldwide. It is an ancient tree species with well-known medicinal values. The rapid development of bio-sequencing technology, particularly whole-genome sequencing, has allowed the identification of AP2/ERF families have been identified and characterized in dicotyledonous and monocotyledonous plants. In the current study, transcriptome based analysis of *G. biloba* was used for the identification and characterization of GbAP2/ERF family, because RNA sequence analysis has been considered as one of the most valuable methods to discover and identify novel genes in the genome of unknown plants.

TTLs are the active ingredients in ginkgo extract, but their content in *G. biloba* leaves is low, the loss during extraction is very large, medicinal resource of the TTL is greatly limited. In this study, we used UV, drought and CCC at different concentrations to treat the leaves, and explore the TTL level change regulated by external environment via transcription factors.

Transcription regulation is very important for plant metabolism regulation. Various transcription factors, such as WRKY, bZIP, bHLH and AP2/ERF, participate in regulation of metabolism and synthesis of plant terpenoids, flavonoids and alkaloids(Selmar, 2013). Transcription factor AP2/ERF family is one of the largest transcription factor families of plants, and all the members have conservative AP2 structural domain. According to the number of structural domain and recognition sequence, AP2/ERF family is divided into 5 subfamilies: AP2 (APETALA2), ERF (ethylene-responsive factor), DREB (dehydration responsive element binding proteins), RAV (related to ABI3/VP1) and Soloist(Nakano *et al*., 2006). Sakuma (2002) divided each subfamily of DREB and ERF in Arabidopsis into 6 subgroups, according to sequence homology (A1~A6; B1~B6). RAV subfamily was separated from Arabidopsis for the first time, which contained 2 different DNA binding structural domains (AP2 and B3)(Kagaya *et al*., 1999). In our study, 78 AP2-ERF transcription factors were separated from *G. biloba* tissues after being treated by UV, drought and CCC. The comparison with known AP2-ERF of Arabidopsis indicated that 21 transcription factors belonged to DREB subfamily, 18 were classified as ERF subfamily, and 4 were RVA family. The correlation analysis of transcription indicated that ERF13, ERF25 and ERF27 had correlations with the expression of key gene that regulated TTL synthesis in *Ginkgo*.

It has been reported that some AP2/ERF families participate in the synthesis and metabolism of terpenoids. Yu (2012) used 5 hormones to treat *Artemisia apiacea* and found that methyl jasmonate had great effect on the expression level of key enzyme gene in the synthesis of artemisinin. The analysis of promoter sequence indicated

that the sequence had binding sites of transcription factor AP2/ERF. The results of qRT-PCR showed that two genes (AaERF1 and AaERF2) had the highest expression level induced by MeJa, and had coordinate expression with key enzyme genes (ADS and CYP71AC1) in different tissues. Furthermore, transcription factor TcDREB in DREB family was proved to participate in the synthesis and regulation of terpenes in *Taxus chinensis*, which was related with the synthesis of paclitaxel (metabolite of isoprene)(Dai 2008). Phylogenetic analysis indicated that ERF13 in Ginkgo belonged to ERF B-3 family, and ERF25 and ERF27 belonged to DREB A-5 family. EMSA analysis showed that ERF13, ERF25 and ERF27 could bind with MECP promoter E-box, which preliminarily suggested that ERF could involve in regulating expression of TTL synthesis-related gene.

The expression of AP2/ERF transcription factor family has tissue variability in plants. In *T. chinensis*, the expression level of transcription factor TcDREB is the highest in caulicle (Dai, 2008). Through tissue specific expression analysis, one transcription factor AaORA specifically expressed on the epidermal hair was screened from six AP2 transcription factors that were cloned from *A. apiacea*. The screened gene could positively regulate synthesis of artemisinin and arteannuic acid(Lu *et al*., 2013). In *G. biloba*, the content of ginkgolides was the highest, and the content of bilobalide in lateral root of gingko was the highest(Chao *et al*., 2005). The precursor of ginkgolide IPP/DMAPP mainly comes from MEP pathway, and that of gingko mainly comes from MVA pathway (Nakanishi and Habaguchi, 1971). qRT-PCR indicated that the expression level of EPF25 in root was the highest, but EFR13 and EFR27 were relatively high. In Ginkgo, they all had relatively high expression level, suggesting ERF25 might participate in the synthesis and regulation of bilobalide MVA pathway and ginkgolide MEP pathway at the same time.

Previous studies have proved that by up-regulating expression levels of 4 key genes (GbDXS, GbDXR, GbGGPPS, GbLPS and GbMECPs) in TTL synthesis pathway, CCC could promote biosynthesis of TTL in *G. biloba* (Xu *et al*., 2011; Cheng *et al*., 2015), and increase contents of ginkgolide and bilobalide(Xu *et al*., 2011). CCC treatment could significantly increase transcription levels of ERF-25 and ERF-27, implying that they could participate in the regulation of CCC on Ginkgo synthesis in Ginkgo. ERF13 and ERF27 mainly regulated bDXS, GbDXR and MECPs. UV treatment could significantly increase transcription levels of MECPs gene and total lactone content(Pingsheng and Shuchai, 2002; Cheng *et al*., 2015). Moreover, drought could significantly increase the accumulation of TTL in Ginkgo (Zhu *et al*., 2011). In our study, CCC treatment could significantly improve the expression levels of ERF25 and ERF27, UV could induce ERF13 transcription level, and drought could induce transcription level of ERF25. The results implied that ERF25 and ERF27 might involve in the induction and regulation of CCC treatment on synthesis of TTL in Ginkgo, ERF12 might participate in the regulation of the synthesis induced by UV, EFR25 might involve in the regulation of the synthesis induced by drought. In the total lactone terpene synthesis in Ginkgo, ERF25 and ERF27 had significantly positive correlation with total triterpene content, but ERF13 had negative correlation. The correlation coefficient was 0.793. The transcription level changes of ERF13, ERF25 and ERF27 had significantly positive correlation with diterpene level with correlation coefficient 0.975. It suggested that in the annual periodic variation of *G. biloba*, ERF13, ERF25 and ERF27 mainly acted on MEP pathway to regulate the synthesis of ginkgolides. ERF25 and ERF27 had correlations with terpenoid lactones, which was consistent with the expression difference in tissues.

Acknowledgements

This work was supported by the Natural Science Foundation of China (31400556), Economic Forest Germplasm Improvement and Comprehensive Utilization of Resources of Hubei Key Laboratories (2017AX02) and Scientific Research Key Project of Hubei Educational Office (D20162901).

References

Chadwick M, Trewin H, Gawthrop F, Wagstaff C (2013). Sesquiterpenoids lactones: Benefits to plants and people. International Journal of Molecular Sciences 14(6):12780-12805.

Chao M, Hong X, Keli C (2005). Preliminary study on the ginkgolide b and bilobalide contents in different tissues and callus cultures of *Ginkgo biloba*. Natural Product Research & Development 17:603-605.

Cheng SY, Li LL, Yuan HH, Xu F, Cheng H (2015). Molecular cloning and characterization of GbMECT and GbMECP gene promoters from *Ginkgo biloba*. Genetics and Molecular Research 14(4):15112-15122.

Dai Y (2008). Molecular cloning and characterization of AP2-type transcription factors involved in isoprenoid biosynthetic pathway of *Taxus cuspidata*. PhD Thesis, Fudan University. Fudan University Publication.

Gao C, Li P, Song A, Wang H, Wang Y, Ren L, ... Chen S (2015). Isolation and characterization of six AP2/ERF transcription factor genes in *Chrysanthemum nankingense*. International Journal of Molecular Sciences 16(1):2052-2065.

Han S, Wu Z, Jin Y, Yang W, Shi H (2015). RNA-Seq analysis for transcriptome assembly, gene identification, and SSR mining in ginkgo (*Ginkgo biloba* L.). Tree Genetics & Genomes 11(3):1-10.

Jie C, Renke T, Xiaojuan G, Zhengli F, Zheng W, Zhiyan Z, Xiaoli T (2015). The functional annotation of unigenes in COG, GO, KEGG, Swissprot and Nr database.

Kagaya Y, Ohmiya K, Hattori T (1999). RAV1, a novel DNA-binding protein, binds to bipartite recognition sequence through two distinct DNA. Nucleic Acids Research 27(2): 470-478.

Kim JH, Lee KI, Chang YJ, Kim SU (2012). Developmental pattern of *Ginkgo biloba* levopimaradiene synthase (GbLPS) as probed by promoter analysis in *Arabidopsis thaliana*. Plant Cell Reports 31(6):1119-1127.

Kim SM, Kuzuyama TChang YJ, Song KS, Kim SU (2006). Identification of class 2 1-deoxy-D-xylulose 5-phosphate synthase and 1-deoxy-D-xylulose 5-phosphate reductoisomerase genes from *Ginkgo biloba* and their transcription in embryo culture with respect to ginkgolide biosynthesis. Planta Medica 72(03):234-240.

Kumar S, Stecher G, Tamura K (2016). MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. Molecular Biology & Evolution 33(7): 1870-1874.

Lange BM, Rujan T, Martin W, Croteau R (2000). Isoprenoid biosynthesis: the evolution of two ancient and distinct pathways across genomes. Proceedings of the National Academy of Sciences of the United States of America 97(24):13172-13177.

Leng P, Su S, Wang T, Jiang X, Wang S (2002). Effects of light intensity and light quality on photosynthesis, flavonol glycoside and terpene lactone contents of *Ginkgo biloba* L. seedlings. Journal of Plant Resources & Environment 11(1):1-4.

Liao Y, Xu F, Huang X, Zhang W, Cheng H, Wang X, … Shen Y (2015). Characterization and transcriptional profiling of *Ginkgo biloba* mevalonate diphosphate decarboxylase gene (GbMVD) promoter towards light and exogenous hormone treatments. Plant Molecular Biology Reporter 34(3):1-16.

Liao YL, Feng XU, Zhu J, Wang Y, Cheng SY (2008). Separation and determination of terpene trilactones by gas chromatography with wide bore capillary column. Acta Agriculturae Boreali-Occidentalis Sinica 17(1):146-149.

Lu X, Zhang L, Zhang F, Jiang W, Shen Q, Zhang L,… Tang K (2013). AaORA, a trichome-specific AP2/ERF transcription factor of *Artemisia annua*, is a positive regulator in the artemisinin biosynthetic pathway and in disease resistance to *Botrytis cinerea*. New Phytologist 198(4):1191-1202.

Magallón S, Hilu KW, Quandt D (2013). Land plant evolutionary timeline: Gene effects are secondary to fossil constraints in relaxed clock estimation of age and substitution rates. American Journal of Botany 100(3):556-573.

Mizoi J, Shinozaki K, Yamaguchi-Shinozaki K (2012). AP2/ERF family transcription factors in plant abiotic stress responses. Biochimica Et Biophysica Acta 1819(2):86-96.

Nakanishi K, Habaguchi K (1971). Biosynthesis of ginkgolide B, its diterpenoid nature, and origin of the tert-butyl group. Journal of the American Chemical Society 93(14):3546-3547.

Nakano T, Suzuki K, Fujimura T, Shinshi H (2006). Genome-wide analysis of the ERF gene family in Arabidopsis and rice. Plant Physiology 140(2):411-432.

Okamuro JK, Caster B, Villarroel R, Van MM, Jofuku KD (1997). The AP2 domain of APETALA2 defines a large new family of DNA binding proteins in Arabidopsis. Proceedings of the National Academy of Sciences of the United States of America 94(13):7076-7081.

Postel D, Vanlemmens P, Gode P, Ronco G, Villa P (2002). PlantCARE, a database of plant cis-acting regulatory elements and a portal to tools for in silico analysis of promoter sequences. Nucleic Acids Research 30(1):325-327.

Sabater-Jara AB, Souliman-Youssef S, Novo-Uzal E, Almagro L, Belchí-Navarro S, Pedreño MA (2013) Biotechnological approaches to enhance the biosynthesis of ginkgolides and bilobalide in *Ginkgo biloba*. Phytochemistry Reviews 12(1):191-205.

Sakuma Y, Qiang L, Dubouzet JG, Abe H, Shinozaki K, Yamaguchi-Shinozaki K (2002). DNA-binding specificity of the ERF/AP2 domain of Arabidopsis DREBs, transcription factors involved in dehydration- and cold-inducible gene expression. Biochemical & Biophysical Research Communications 290(3):998-1009.

Schmittgen TD, Livak KJ (2008). Analyzing real-time PCR data by the comparative CT method. Nature Protocols 3(6):1101-1108.

Selmar D (2013). Stress enhances the synthesis of secondary plant products: the impact of stress-related over-reduction on the accumulation of natural products. Plant & Cell Physiology 54(6):817-826.

Singh B, Kaur P, Gopichand, Singh RD, Ahuja PS (2008). Biology and chemistry of *Ginkgo biloba*. Fitoterapia 79(6):401-418.

Strømgaard K, Nakanishi K (2004). Cover picture: chemistry and biology of terpene trilactones from *Ginkgo biloba*. Angewandte Chemie 43(13):1640-1658.

Van dFL, Memelink J (2000). ORCA3, a jasmonate-responsive transcriptional regulator of plant primary and secondary metabolism. Science 289(5477):295-297.

Wang L, Lu Z, Li W, Xu J, Luo K, Lu W, Zhang L, Jin B (2016). Global comparative analysis of expressed genes in ovules and leaves of *Ginkgo biloba* L. Tree Genetics & Genomes 12(2):1-18.

Wang T, Zhang N, Du L (2005). Isolation of RNA of high quality and yield from *Ginkgo biloba* leaves. Biotechnology Letters 27(9):629-633.

Xu DL, Long H, Liang JJ, Zhang J, Chen X, Li JL, … Yu MQ (2012). *De novo* assembly and characterization of the root transcriptome of *Aegilops variabilis* during an interaction with the cereal cyst nematode. BMC Genomics 13(1):277-287.

Xu F, Zhang W, Sun N, Li L, Cheng S, Wang Y (2011). Effects of chlorocholine chloride on photosynthesis metabolism and terpene trilactones biosynthesis in the leaf of *Ginkgo biloba*. Acta Horticulturae Sinica 38(12):2253-2260.

Yu ZX, Li JX, Yang CQ, Hu WL, Wang LJ, Chen XY (2012). The jasmonate-responsive AP2/ERF transcription factors AaERF1 and AaERF2 positively regulate artemisinin biosynthesis in *Artemisia annua* L. Molecular Plant 5(2):353-365.

Zhu C, Cao F, Wang G, Geng G (2011). Effects of drought stress on annual dynamic changing pattern of the terpene lactones content in *Ginkgo biloba* leaves. China Forestry Science and Technology 25:15-20.