

The Molecular Cloning and Expression Analysis of a CYP71 Gene in *Ginkgo biloba* L.

Xinliang LIU, Fuliang CAO*, Jinfeng CAI, Huanli WANG

Nanjing Forestry University, Co-Innovation Center for Sustainable Forestry in Southern China, 159 Longpan Road, Nanjing, China; xliu25@lakeheadu.ca; fliangcao@njfu@163.com (*corresponding author); cjinfeng1984@126.com; wanghuanli@163.com

Abstract

Cytochrome P450 monooxygenases (CYPs) are a group of redox proteins that catalyze various oxidative reactions in plant secondary metabolism. To explore the function of the CYP71 gene in *Ginkgo biloba* under biotic and abiotic stresses, a full-length CYP gene, designated *GbCYP71*, was first isolated and characterized from leaves of *G. biloba*. It contained a 1512-bp open reading frame (ORF) encoding 503 amino-acid-deduced polypeptide whose theoretical molecular weight was 56.9 kDa. The genomic DNA sequence of *GbCYP71* contained two exons and one intron. The cDNA of *GbCYP71* was subcloned in a pET-32a vector and then transformed into *E. coli* strain BL21 (DE3). A protein with a molecular weight of 76.4 kDa was subsequently identified and found to be consistent with the above theoretical value. Transient expression analysis revealed that the GbCYP71 protein may be located in the *G. biloba* cell cytoplasm. *GbCYP71* was expressed in almost all ginkgo tissues, including leaves, stamens, gynoecia, stems and, preferentially, roots. Expression-profiling analyses revealed that *GbCYP71* can be induced by salinity stress and phytohormone signals, including salicylic acid, abscisic acid, methyl jasmonate and ethephon, but is repressed by heat and cold stresses. These results indicate that *GbCYP71* mainly functions in responding to biotic and abiotic stresses.

Keywords: abiotic stress, expression pattern, *Ginkgo biloba*, *GbCYP71*, phytohormone induction

Introduction

Cytochrome P450 monooxygenases (CYPs) form the largest enzyme family and present in various tissues (Werck-Reichhart *et al.*, 2000) in higher-order plants. The CYPs are a group of redox proteins that catalyze the oxidation steps of various substrates using oxygen and NAD(P)H (Guttikonda *et al.*, 2010; Isin and Guengerich, 2007). The first reported plant CYP gene, *CYP71A1*, was a ripening-related gene from the mesocarp of an avocado (Bozak *et al.*, 1990). Since its discovery, substantial works have been performed on cloning and characterizing a large array of CYPs. At present, more than 127 families and 5100 sequences of plant CYPs have been annotated and named (Nelson and Werck-Reichhart, 2011); however, the biological functions of most of these proteins in plants remain unknown. Plant CYPs have been shown to participate in many metabolic pathways, especially those related to plant secondary metabolism, which mainly involves the synthesis of endogenous compounds, such as structural macromolecules, pigments, and defense compounds; the synthesis or catabolism of all types of hormones or signaling molecules; and the metabolism of xenobiotics (Morant *et al.*, 2003; Werck-Reichhart *et al.*, 2002). CYPs catalyze many oxidation reactions, including epoxidations,

hydroxylations, dealkylations, isomerizations, dehydrations, aryl migrations, decarboxylations, ring formations, ring extensions and carbon-carbon bond cleavages (Rupasinghe and Schuler, 2006). The ability of CYPs to catalyze such diverse reactions in plants contributes greatly to understanding the challenge posed by abiotic stresses and pathogens in nature.

An ancient gene superfamily, CYPs are classified into various families and subfamilies on the basis of protein sequence identity and phylogeny (Nelson, 2006; Nelson *et al.*, 1996). CYPs with at least 40% similarity belong to the same family, and CYPs with at least 55% similarity belong to the same subfamily. CYPs with at least 97% similarity are assumed to be allelic variants unless otherwise demonstrated. Plant CYP families have been assigned names, including CYP51, CYP71 to CYP99 and CYP701 to CYP772, as depicted on the Cytochrome P450 webpage (<http://drnelson.uthsc.edu/CytochromeP450.html>). The nomenclature of a CYP gene is regulated the same way as the number following the CYP: the prefix is for the family, the letter is for the subfamily, and the number is for the particular gene. Based on evolutionary relationships in phylogenetic trees, the CYPs of land plants are grouped into 11 clans, and each clan is derived from a different gene ancestor (Nelson and Werck-Reichhart, 2011).

Most CYP members contain approximately 500 amino acids and have a molecular weight of 45 kD to 65 kD (Donaldson and Luster, 1991; Schuler, 1996). Some conserved domains have been well identified in the primary CYP amino acid sequence; the most highly conserved CYP regions are the proline-rich protein, PERF, K-helix, and I-helix regions and the heme-binding domains. Following the CYP N-terminal hydrophobic helix, the highly conserved proline-rich protein region is thought to act as a membrane hinge, usually obeying the consensus P/I-P-G-P-X-G/P-X-P (Yamazaki *et al.*, 1993). The I-helix contains a conserved region A/G-G-X-E/D-T-T/S corresponding to the proton transfer groove on the distal side of the heme (Werck-Reichhart and Feyereisen, 2000). There is a K-E-T-L-R consensus sequence with conserved Glu (E) and Arg (R) in the K-helix; its E and R form an E-R-R triad with R from P-E-R-F. The triad is considered to be involved in locking the heme pocket into position and assures heme association with the protein (Paquette *et al.*, 2000). The heme-binding domain in the C-terminal region of the protein contains the consensus sequence F-X-X-G-X-R-X-C-X-G, which is the main feature in CYP identification. The conserved cysteine (C) of the heme-binding domain serves as the fifth ligand for the heme iron (Chapple, 1998). Only the E-R-R triad and the C in the heme-binding domain are conserved in all plant CYP sequences among these conserved domains (Werck-Reichhart *et al.*, 2002). In addition, plant CYPs vary in their subcellular locations, and most of them are endoplasmic-reticulum (ER) binding proteins; the N-terminal hydrophobic helix anchors to the ER membrane, and most of the rest of the protein is located on the cytosolic face of the membrane (Chapple, 1998; Werck-Reichhart and Feyereisen, 2000).

Ginkgo biloba is an ancient plant that originated in the Carboniferous period and thrived in the Jurassic period. It is the only remaining species of *Ginkgoaceae* and is often called a "living fossil" (Shen *et al.*, 2006). *G. biloba* is one of the most popular functional plants used worldwide today, especially for medicinal purposes. Plants in nature frequently experience complicated climate changes and have to develop a broad range of complex defense systems to adapt. Geological records show that *G. biloba* exhibits minor changes in its morphology, indicating its strong adaptability in changing environments and tolerance of harsh conditions (Royer *et al.*, 2003). In addition, *Ginkgo* is widely distributed around the world, thanks to characteristics such as resistance to disease, pests, drought, low temperatures, barren landscapes, and nuclear radiation (Cheng *et al.*, 2013). To date, CYP genes have been reported in 66 species, including algae, planus, pteridophytes, needle-leaved plants, monocotyledons, and dicotyledons. However, the CYP gene has not been reported in *G. biloba*. In this study, we isolated the cDNA of a CYP gene from *G. biloba* and characterized the physical-chemical properties of its putative protein using bioinformatics tools. The expression patterns of the *GbCYP* was analyzed by real-time PCR. Additionally, *GbCYP* was subcloned into an expression vector and underwent heterologous expression in *Escherichia coli* and *Nicotiana benthamiana* to further elucidate its functional characterization in *G. biloba*.

Materials and Methods

Plant materials and growth conditions

Different tissues, including leaves, stamens, gynoecia, stems and roots, were collected from several 16-year-old grafted *Ginkgo biloba* plants that were grown in the outdoor ginkgo garden of Nanjing Forestry University in Jiangsu Province, China. These tissues were immediately frozen in liquid nitrogen, transferred to an ultra-low-temperature freezer and kept at -80 °C prior to total RNA extraction.

Mature zygotic embryos of *G. biloba* were incubated in an MS induction medium to initiate callus cultures; the MS basal medium was supplemented with 2 mg/L 6-benzyladenine (6-BA) and 1.5 mg/L naphthaleneacetic acid (NAA). After 4 weeks of incubation, the original seed embryos generated calluses, which were transferred to fresh MS induction media at 2-week intervals. The calluses, having a similar growth rate, were transferred to an MS liquid medium supplemented with 6-BA (2 mg/L) and NAA (1.5 mg/L) after two subcultures and then were cultured on a rotary shaker at 100 rpm, in the light, at 25 ± 1 °C. The suspension cultures were subcultured three times at 2-week intervals, and the calluses that appeared pale yellow-white, translucent, or browning were omitted.

The tobacco (*Nicotiana benthamiana*) used for the transient expression of fluorescent fusion proteins was approximately 4 to 5 weeks old and was planted in an artificial climate chamber at 21 °C with 70% relative humidity, with 14 hours of light (500 μmol/m²/s) and 10 hours of dark.

Hormones and stress treatments

To investigate induction elicitors, the callus lines were treated with 100 μmol/L salicylic acid (SA), 100 μmol/L abscisic acid (ABA), 100 μmol/L methyl jasmonate (MeJA), 40 μmol/L ethephon (ETH) and 200 mmol/L sodium chloride (NaCl), separately, while the calluses without treatment were used as a control (CK) (Gong *et al.*, 2006; Liao *et al.*, 2015; Novo-Uzal *et al.*, 2014). To explore the response to temperature stress, the callus lines were treated at 4 °C and 40 °C separately, while the control group was treated at 25 °C. Samples were collected at 0, 3, 6, 12, 24 and 48 h after the beginning of treatments, immediately frozen in liquid nitrogen and stored at -80 °C. Each sample at each time point included three biological replicates.

Molecular cloning

Total RNA was extracted from different tissues subjected to different treatments using the MiniBEST Plant RNA Extraction Kit (Takara, Japan), according to the manufacturer's instructions. The RNA sample was subsequently treated with DNase I according to the extraction kit instructions to remove contaminating genomic DNA. Genomic DNA was extracted from flesh leaves following use of the Plant DNA Isolation Reagent (Takara, Japan). The quality of the RNA and genomic DNA was determined by agarose gel electrophoresis. The concentration was detected by UV spectrophotometry. Reverse transcription-polymerase chain reaction was performed using a Transcriptor First Stand cDNA Synthesis Kit (Roche, Germany).

The specific primers CYP1 and CYP2 (Table 1) were designed and synthesized (Shanghai Generay Biotech, China) based on the transcriptome sequencing. To obtain the cDNA

and genomic sequences of *GbCYP71*, a touchdown thermal cycle program was utilized under the following conditions: 94 °C for 4 min, 94 °C for 30 s, 58 °C for 30 s and 72 °C for 1 min 30 s, followed by 9 cycles at decreasing annealing temperatures in decrements of 0.5 °C per cycle, and then 25 cycles of 30 s at 94 °C, 30 s at 54 °C and 1 min 30 s at 72 °C, as well as a final extension at 72 °C for 10 min. The PCR products were purified and cloned into a pMD19-T vector (Takara, Japan) for sequencing.

Bioinformatics analysis and phylogenetic tree analysis

The obtained sequences were analyzed using online bioinformatics tools (<http://www.ncbi.nlm.nih.gov> and <http://web.expasy.org/protparam/>). The software Vector NTI Advance 11 was used for sequence multi-alignment. Sequences of homologous CYP71 proteins present in the NCBI database were retrieved using BLAST network services. The protein signal sequence was predicted using the SignalP 4.1 Server (<http://www.cbs.dtu.dk/services/SignalP/>), and the transmembrane domain was analyzed by the TMHMM server (<http://www.cbs.dtu.dk/services/TMHMM/>). Subcellular localization presumption was performed using TargetP 1.1 (<http://www.cbs.dtu.dk/services/TargetP/>). Phylogenetic and molecular evolutionary analyses of the amino acid sequences from other plant species were constructed with MEGA 6.06 software by the neighbor joining method and measured by bootstrap analysis with 1000 replicates.

RNA extraction and qPCR

A real-time quantitative PCR detection system (qPCR) was used to investigate the transcription profile of *GbCYP71* in different tissues and in the callus samples collected at different time points after treatment. qPCR was performed using an ABI PRISM 7500 Sequence Detection System (Applied Biosystems, USA), using the SYBR Green PCR Master Mix (Roche, Germany) PCR system. PCR was performed with a 20 µL reaction consisting of 50 ng of template cDNA, 2×SYBR Green Real-time PCR Master Mix and 2.5 mM of each primer. The glyceraldehyde-3-phosphate dehydrogenase gene (*GbGAPDH*, GenBank Accession no. L26924) was used as the reference gene (Xu et al., 2008). The primers for *GbCYP71* (CYPF, CYPR) and *GbGAPDH* (GAPDHF, GAPDHR) were designed and listed in Table 1. The following qPCR conditions recommended by the manufacturer were used: 95 °C for 10 min, followed by 40 cycles of 95 °C for 15 s and 60 °C for 1 min. qPCR data were calibrated relative to the corresponding gene expression level at time zero for each treatment, following the $2^{-\Delta\Delta Ct}$ method for relative quantification (Livak and Schmittgen, 2001).

Prokaryotic expression of *GbCYP71* in *Escherichia coli*

The open reading frame (ORF) was amplified by a pair of primers, CYP51 and CYP41 (Table 1). The primers have insert-specific sequences and an additional 15 bases overlapping the vector ends and were designed according to the manufacturer's instructions using the ClonExpress™ II One Step Cloning Kit (Vazyme, China). The PCR conditions were as follows: 1 cycle of 95 °C for 3 min, followed by 30 cycles of 95 °C for 10 s, 60 °C for 15 s, and 72 °C for 30 s, as well as 1 cycle of 72 °C for 10 min. The PCR product was separated by electrophoresis on 1% (w/v) agarose gels and then cloned into pET-32a, which was digested by BamH I at 37 °C for 4 h. The recombinant plasmid was introduced into *E. coli* BL21 (DE3) by the heat shock method. A single colony of *E. coli* BL21 (DE3) cells harboring the expression plasmid was grown in LB medium containing ampicillin (50 µg/mL) with 160 rpm shaking at 37 °C. For induction, IPTG was added to the *E. coli* culture at a final concentration of 1.0 mM when the optical density (OD₆₀₀) reached approximately 0.6. The cells continued growing for 6 h at 37 °C before being collected by centrifugation.

The collected cells with pET32a-*GbCYP71* or pET-32a plasmid were lysed by sonication in PBS buffer (20 mM, pH 7.4) and centrifuged at 12,000 ×g for 15 min at 4 °C. The pellets were re-suspended and then boiled for 5 min after adding 5×SDS loading buffer (4:1, v:v). SDS-PAGE analysis was used to examine 10 mg of each of the samples, followed by a staining procedure with Coomassie brilliant blue R250. Western blot analysis was performed to verify the expression of a *GbCYP71* protein recombining a His-tag in the N-terminus. After electrophoresis, the proteins were electro-transferred onto a PVDF membrane (0.45 µm) and detected with TRUE-Tag Anti-His mAb (Vazyme, China) and a secondary antibody, HRP-conjugated Goat Anti-Mouse IgG (H+L) (Vazyme, China). Enhanced ECL (Vazyme, China) was used in the color reaction.

Transient expression in *N. benthamiana*

For transient expression of *GbCYP71* in tobacco, all plasmids were constructed using Gateway technology (Invitrogen, USA). To generate the entry vector, pENTR/D/TOPO vector (Invitrogen) was used to connect with the gene fragment encoding *GbCYP71*. The gene fragment was amplified from cDNA with the primer pair CYP52/CYP42 and was designed according to the guidelines of TOPO Cloning instruction (Table 1). The expression vectors (*CYP71*-GFP) for tobacco injection, including the promoter of 35S and enhanced green fluorescent protein (Egfp), were generated by LR reactions using the LR Clonase II enzyme (Invitrogen, USA) with the corresponding entry vector and

Table 1. Primers for *GbCYP71*

Primer	Sequence (5'-3')	Description
CYPU	GGTTTCCAAGGAGGGATTTCAGA	Gene-specific primer, forward
CYPD	GGTGGTATTTTCGCCAAAGTCAT	Gene-specific primer, reverse
CYPF	GGAGGCAAATGCGTAAGGTG	Primer for qPCR, forward
CYPR	TTGCTTCATAATGCTCTCACGC	Primer for qPCR, reverse
GAPDHF	GGTGCCAAAAGGTGGTCAT	Primer for qPCR, forward
GAPDHR	CAACAACGAACATGGGAGCAT	Primer for qPCR, reverse
CYP51	GCCATGGCTGATATCGGATCCATGGATGGAGTTAATGCATCTG	Primer for ORF subclone, forward
CYP41	CTTGTGACGGAGCTCGAATTCTCATGGCTGTGGAAGACGAG	Primer for ORF subclone, reverse
CYP52	CACCATGGATGGAGTTAATGCATCTGTA	Primer for ORF subclone, forward
CYP42	TGGCTGTGGAAGACGAGG	Primer for ORF subclone, reverse

pB7WFG2.0 (Karimi *et al.*, 2002). The reference vector (35S-RFP) for the expression indicator was constructed with 35S and red fluorescent protein (RFP) and was located in the cytoplasm. The plasmid encoding p19 protein was used as a viral silencing suppressor. Each constructed plasmid was introduced into *Agrobacterium tumefaciens* GV3101 via the freeze-thaw method.

Each *Agrobacterium* strain harboring the plasmid encoding GbCYP71, RFP and p19 protein was mixed in a 1:1:1 ratio before being transfected into the underside of tobacco leaves. The leaves were agroinfiltrated according to Sparkes *et al.* (2006). Three leaves per plant were infiltrated with the same constructs to yield experimental triplicates. After 3 to 4 days of infiltration (the period when fluorescence was optimal), fluorescent protein localization was monitored in the lower epidermal cells of the leaves using a fluorescence microscope (Leica DM5000B).

Statistical analysis

Statistical analysis was performed by one-way analyses of variance (ANOVA), and mean differences were compared by Tukey's multiple range test using SPSS 17.0.

Results

Characterization of GbCYP71 in *Ginkgo biloba*

The full-length cDNA sequence of the *GbCYP71* gene was obtained from *G. biloba*. It is a 1568-bp sequence containing a 1512-bp open reading frame (ORF). A 1630-bp genomic DNA sequence of the *GbCYP71* gene was amplified by a pair of specific primers derived from the cDNA sequence containing the start and stop codon region. Comparison with the full-length cDNA sequence showed that the genomic DNA and cDNA were 100% identical in the coding region, although the genomic DNA contained an intron. Exon 1 (894 bp) and exon 2 (618 bp) were separated by intron 1 (118 bp). The putative splicing site obeyed the GU/AG rule (Fig 1).

The putative GbCYP71 protein contains 503 amino acids. Analysis revealed that the isoelectric point (pI) and molecular weight of this peptide were 6.88 and 56.9 kDa, respectively (ProtParam tool). The putative formula of GbCYP71 was C₂₅₄₉H₄₀₄₅N₆₈₉O₇₂₃S₃₀. The instability index (II) was computed to be 39.70 (which classified the protein as stable), and its grand average of hydropathicity (GRAVY) was predicted to be -0.165. SignalP 4.1 analysis showed that GbCYP71 had a signal peptide sequence with a cleavage site between Gly 26 and Ser 27. Transmembrane analysis by the TMHMM Server v. 2.0 showed that GbCYP71 had one transmembrane domain located between Gly 3 and Phe 21 along the polypeptide chain. The TargetP 1.1 Server predicted that the subcellular localization of GbCYP71 was in the secretory pathway.

As shown in Fig. 1, the deduced amino acid sequences of this gene showed the typical features of membrane-bound P450s. Alignment by NCBI's Blastx showed that GbCYP71 shared moderate homology with CYPs from *Picea sitchensis* (59%, ABR18406.1), *Cucumis melo* (56%, XP_008440668.1), *Vitis arizonica* × *Vitis rupestris* (55%, ACM89788.1), *Cicer arietinum* (54%, XP_004504031.1) and *Populus euphratica* (53%, XP_011025622.1). However, the functions of the above proteins have not been reported. Nevertheless, this predicted protein has high homology (40-49%) with known CYP71 proteins of other higher-order plants, which indicates

```

1 CATATACGTAGCTTAGAGTAATTAGATAAACACAGTAGAGCATATCATAGGCAAAATGGGT
61 TTCCAAGGAGGGATTGAGAAGCTTTAGAGAAGATGGATGGAGTTAATGCATCTGTAA
1 M D G V N A S V
121 TAGTCTGGTGGTGGTGTAGTGGCATGGATATCTTCAAAAACCCGGCATGGCTCATCG
9 I G L V V V L V A W I F F K N R H G S S
181 ATCAGAAGCTGAGATTGCCACCGGCCTCTCCATTAATAGGAAATCTTCAACA
29 D Q K L R L P P G P P P L P I I G N L H
241 TGTTGGAAAACTCCCATAGAAAACCTGGGGAAGTTTCAAAGAAATATGGACCATAA
49 M L G K L P H R K L R E L S K K Y G P I
301 TGTTCTACGCCTAGGATCTGCCAACATATAGTCGCATCTTCCAGAGATGGCCAAAG
69 M F L R L G S V P T I V A S S P E M A K
361 AGTTTCTGAAAACCCATGATCTGATTTTCGCTAGCAGACCTGCAACCTGCGTAGGCAAA
89 E F L K T H D L I F A S R P A T C V G K
421 ACATGATCTATAATTCACAGACGTGGGATTTTCTCCTATGGGCCCTACTGGAGGCAAA
109 N M I Y N S T D V G F S P Y G P Y W R Q
481 TGCCTAAGGTGTGTGCTGGAGCTTTTCAGTCCAGAGGATTGAGTTAATGAATTTTA
129 M R K V C V L E L F S A K R I E L M N F
541 TCAGAGAAGAAGAAATAGCCATCATCGGTGAGAGCATTGAAGCAATCGCGCGTGTATG
149 I R E E E I A I M R E S I M K Q C G V D
601 GATCGAATCCTGGAATATAAGCAAGCTCGTGTCCACCTTGAACCGATATAATATGCA
169 G S N P V N I S K L V S T L A T D I A I C
661 GAATGGCTTTTGAAGAAAGTATTCGGAGGAGACGTTGAAGGACAGCAGGGGTTCAAGG
189 R M A F G R K Y S E E T L K D S R G F K
721 CCATGATCAAGAGGTTATTTCTCAGTGGTTCATCAATATGGAGACTTTATCCAT
209 A M I Q E V F Y L S G S F N I G D F I P
781 CGTTGGATGGATCTTCAAGGCTGAGACGCGCAGAAACACTCACAAGCACT
229 W L E W M D L Q G L R R R Q K H T H K T
841 TCAATGATTTCTTGAGAATGTCATACAGAGGTGGGAGAATAAAGCCCGGGGACACA
249 F N D F F E N V I Q E H V E N K S R G H
901 GAGACTTTGGATGTGATGCTCGAGCTCTCAGAAGCAATACTATGGAGATCAAAATCA
269 R D F V D V M L E L S E D N T M E I K I
961 CCCGAGATAACATCAAAAGCAGTCATCTTtagccaaccagaataatttcaaat
279 T R D N I K A V I L
1021 ttagattgaatttcagaggccaatttaattatttcgttttaatttacagatgcccaat
1081 taattaattattggtttgaatatgcgatgatgtggcagccgggacagacacactctgctc
290 D M L A A G T D T S S
1141 GCAACCTGGAGTGGGCAATGTCTGAGATGTTATTAATCCTCAATGATGAAGAAAGTG
310 A T L E W A M S E M L L N P S M M K K V
1201 CAGGATGAGTTGAAAGTGTAGTGGGTCTGAATCGAATGGTGGAGGAGTCTGATCTGCCT
330 Q D E L E S V V G L N M V E E S D L P
1261 CAACTGGATTTTAAAGTAGTGGTGAAGAACTCTGAGACTTCATCCACCGGGGCCG
350 Q L E Y L K V V V K E T L R L H P P G P
1321 TTATTGATACCACATGAGGCTAGAGAGGATTTGATTTGAGGGGTATAATATACCTAGG
370 L L I P H E A R E D C I V G G Y N I P R
1381 AAGGCTGTATAATAGTCAATGTTTTGCTCTTGGTACAGCCCACTCATGGGAGGAC
390 K A R I I V N V F A L G T D P N S W E D
1441 GCGGAAAGTTCAAACAGAGAGATTATAGGAAACCAATGATGTTAAGGGACAGAAAT
410 A E K F K P E R F I G N P I D V K G Q N
1501 TTTGAATTAATCCCGTTGGGTCAGGAGAAGAGGATGCCCGGGCAGCCATGGCAACT
430 F E L I P F G S G R R R G C P G Q P L A T
1561 ACTGTTGATAGATGATGCATTGGCACTCTCTGCAATGCTTTGATTTGGAGACTCCCGAT
450 T V V E Y A L A T L L H C F D W R L P D
1621 GGCATGAAAGCAGAGGATTTGTCATGACAGAAGATTTGGACTCTACACCCAGGTTCT
470 G M K A E D L C M T E E F G L S T P R S
1681 GTTCATCTCGTTGCAGTCTCACCCTCGTCTCCACAGCATGACTTTGGCGAAATACC
490 V H L V A V P T P R L P Q P *
1741 ACCATGTGGCACTCTTTTTTTCATGAAACAAAAAATACTATCCACT
1801 TCTCTTTAGTGGTTTTTTGTTATATATCTCTAATAAAAAAGAGTACCTATTAAGAA

```

Fig. 1. The partial cDNA sequence, intron sequence and deduced amino acid sequence of the *GbCYP71* gene. The cDNA-sequence-encoded amino acids of GbCYP71 are indicated in capital letters, and the introns are indicated in lowercase. The initiation and termination codons are indicated with the solid red underline. Conserved regions of proline-rich protein (P-P-G-P-X-P-X-P) are indicated with the dotted green underline. The I-helix involved in oxygen binding and activation (A-G-X-D-T-S) is indicated with the dotted red underline. The K-helix consensus sequence (K-E-T-L-R) is indicated with the solid blue underline. The clade signature (P-E-R-F) is indicated with the solid black underline. The heme binding region (F-X-X-G-X-R-X-C-X-G) is indicated with the dotted black underline.

that the protein encoded by the cDNA obtained in this study may belong to the CYP71 protein family.

Molecular evolution analysis

A phylogenetic tree was constructed using the neighbor-joining method by MEGA 6.06 software to investigate the evolutionary relationships between GbCYP71 and other CYPs from different species. An unrooted neighbor-joining cladogram is shown in Fig. 2; the numbers at the nodes show bootstrap support values (1000 replicates), and the scale line below the tree represents relative evolutionary distance. In this phylogenetic tree, the relationships between CYPs from analyzed plants

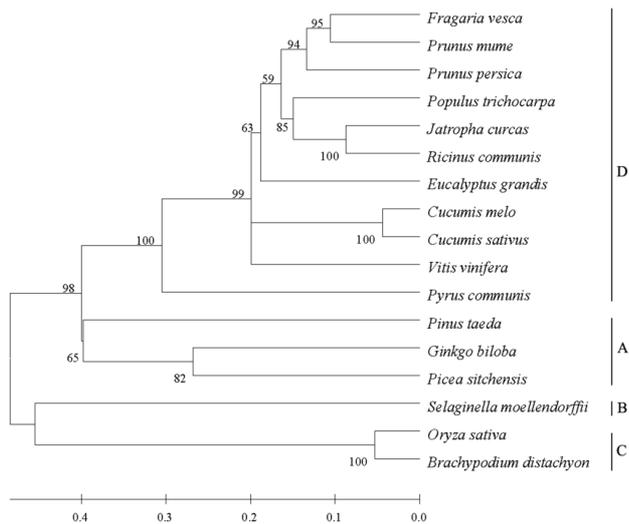


Fig. 2. The phylogenetic tree of CYP from several plants. Letters A to D indicate different CYP groups. A: gymnosperm. B: pteridophyte. C: monocotyledons. D: dicotyledon. *Fragaria vesca* (XP_004304846), *Prunus mume* (XP_008234579), *Prunus persica* (XP_007213194), *Populus trichocarpa* (XP_002310011), *Jatropha curcas* (KDP35215), *Ricinus communis* (XP_002532045), *Eucalyptus grandis* (KCW54382), *Cucumis melo* (XP_008440668), *Cucumis sativus* (XP_004162313), *Vitis vinifera* (XP_002276487), *Pyrus communis* (AAL66194), *Pinus taeda* (AAX07433.1), *Picea sitchensis* (ABR18060), *Selaginella moellendorffii* (EFJ37102.1), *Oryza sativa* (NP_001063285), and *Brachypodium distachyon* (XP_003578185)

generally reflect standard relationships pertaining to gymnosperms, pteridophytes, monocotyledons and dicotyledons. GbCYP71 belongs to the gymnosperm group and has a relatively close relationship with the conifer *Picea sitchensis*.

Expression of recombinant GbCYP71 in *E. coli*

GbCYP71 was expressed in *E. coli* by the pET32a expression vector. SDS-PAGE showed that the molecular weight of the expressed recombinant protein was estimated as 76.4 kDa, which is the sum of the target gene (56.9 kDa) and a histidine marker (approximately 20 kDa) (Fig. 3 a), and was consistent with the predicted value. Western blotting of purified recombinant GbCYP71 protein confirmed its specific immune reactivity to anti-His antibodies (Fig. 3 b).

Subcellular localization of GbCYP71 Proteins in *N. benthamiana*

To investigate the biological activity of the GbCYP71 protein, its subcellular localization was analyzed using the transient expression system in *N. benthamiana*. As expected, 35S-RFP showed cytosolic localization when expressed in an *N. benthamiana* epidermal cell (Fig. 4 b). The recombinant fusion protein CYP71-GFP showed similar accumulation in the cell, which indicated that the GbCYP71 protein was located in the cytoplasm (Fig. 4).

Expression of GbCYP71 in different tissues

To explore the expression pattern of GbCYP71 in *G. biloba*, qPCR was performed using the primers CYP-F and CYP-R. As shown in Fig. 5, the expression of GbCYP71 could be detected in all tissues, but at varying levels. GbCYP71 was preferentially expressed in roots and leaves, and the expression level in roots was significantly higher ($P < 0.05$) than that in leaves and in other

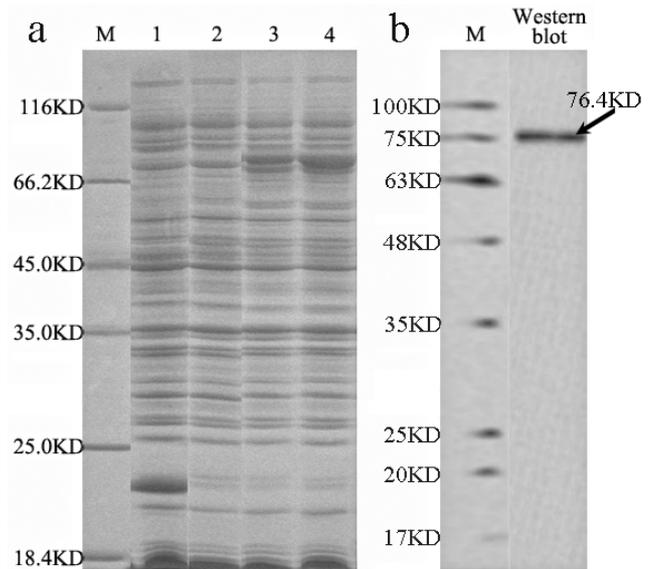


Fig. 3. SDS-PAGE gel and Western blot analysis of GbCYP71 expressed in *E. coli*. (a) SDS-PAGE analysis of GbCYP71. M: protein marker; lane 1: negative control (empty vector control, pET-32a); lane 2: protein of total cells without IPTG induction; lane 3: protein of total cells with IPTG induction for 2 h; lane 4: induction for 6 h. (b) Western blotting of the purified recombinant GbCYP71 protein with an anti-His-tag primary antibody probe. Recombinant protein is indicated with an arrow

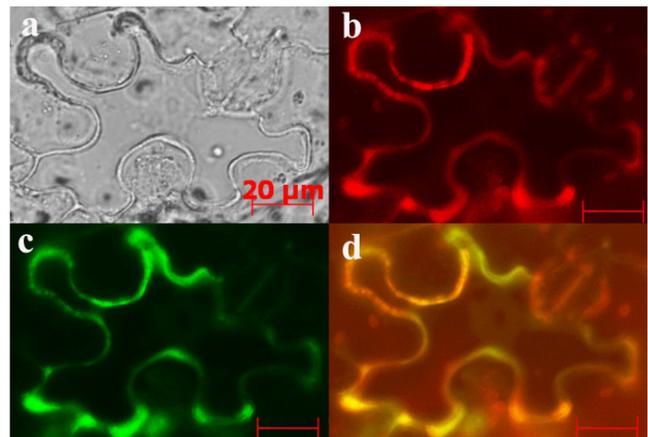


Fig. 4. Subcellular localization of GbCYP71 protein in *N. benthamiana* leaves. Localization of transiently expressed RFP- and GFP-fusion proteins in leaves was analyzed by fluorescence microscopy. (a) A lower epidermal cell was observed in white light. (b) 35S-RFP localized in the cytoplasm. (c) CYP71-GFP localization was similar to that of RFP. (d) Merged b and c images. Scale bar = 20 μ m

tissues. Subsequently, the expression level of GbCYP71 in leaves was significantly higher ($P < 0.05$) than that in stems. The lowest level of GbCYP71 was in the stamens.

Expression of GbCYP71 under salt and different temperature stresses

To determine if GbCYP71 was involved in the responses to abiotic stresses, the gene's transcript levels were measured by qPCR after treating the callus lines with high salinity, cold (4 °C) and heat (40 °C) stress. GbCYP71 expression was significantly affected by NaCl and extreme temperature (Fig. 6). The expression level of GbCYP71 increased quickly after NaCl

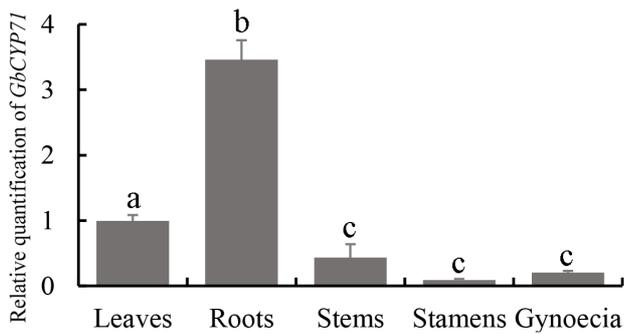


Fig. 5. The expression of *GbCYP71* in different tissues of *G. biloba*. Each sample was individually assayed in triplicate; error bars indicate standard errors of the mean. Different letters indicate significant differences at $p < 0.05$

treatment and reached a peak (10.3-fold) at 6 h. Then, it decreased sharply to a basal level (0.63-fold) at 12 h, followed by a slow decline with a 12 h duration, and finally maintained a 0.33-fold expression level relative to the control until 48 h (Fig. 6 b). Compared with room temperature, *GbCYP71* expression was suppressed under extreme temperatures (Fig. 6 a, c and d). Under high temperature (40 °C), *GbCYP71* expression decreased quickly and finally dropped to 0.17-fold that of the control (0 h) at 24 h (Fig. 6 c). However, at low temperature (4 °C), *GbCYP71* expression decreased to a minimum of 0.42-fold of the control after 3 hours of treatment and then increased slightly and recovered to the same level as the control ($P < 0.01$) at 12-48 h (Fig. 6 d).

Expression of *GbCYP71* under hormone treatments

Expression profiles for *GbCYP71* at different stages of SA, ABA, ETH and MeJA treatment were determined in ginkgo calluses to assess the possible involvement of *GbCYP71* in signaling pathways used by phytohormones. *GbCYP71* transcript levels increased significantly with 100 $\mu\text{mol/L}$ SA, reached an expression peak (8.64-fold) at 24 h, and then decreased sharply to their basal level (Fig. 7 a). *GbCYP71* transcript levels were also stimulated by 100 $\mu\text{mol/L}$ ABA; the strongest response to ABA (20.7-fold induction) was observed at 12 h (Fig. 7 b). Afterward, *GbCYP71* transcript levels decreased as treatment time increased, reaching a 2.95-fold difference from the control at 48 h. *GbCYP71* transcription levels decreased slightly at 3 h after 40 $\mu\text{mol/L}$ ETH treatment, then increased significantly between 3 and 6 h, and finally decreased until 48 h (Fig. 7 c). In response to 100 $\mu\text{mol/L}$ MeJA, *GbCYP71* transcript levels were elevated significantly between 0 and 6 h post treatment and reached an expression peak (8.76-fold) at 3 h, after which they decreased to their basal level (Fig. 7 d).

Discussion

CYPs constitute a superfamily of hemoproteins with diverse functions in plants. They have a systematic nomenclature based on their amino acid sequence identity. Due to limitations in purifying CYP from higher-order plant sources, studies of CYPs in higher plants have lagged behind those in animals and microorganisms (Toguri and Tokugawa, 1994). Members of the same family encode homologous proteins and have similar structures and properties. Similarly, it is possible for genes from different families to participate in similar pathways. In this study, the deduced protein sequence of the CYP71 gene in *G. biloba*

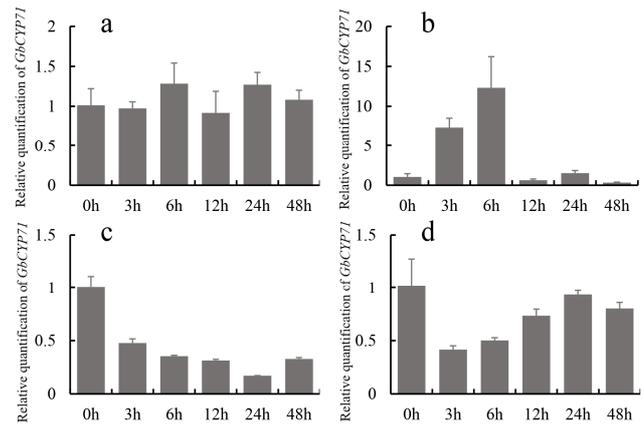


Fig. 6. *GbCYP71* expression at various time points under NaCl (b), 40°C (c) and 4°C (d) stresses. (a) *GbCYP71* expression levels at room temperature (25 °C). Gene expression levels are expressed as a function of the control values (harvested at time 0 h), which are equal to 1. Each sample was individually assayed in triplicate; error bars indicate standard errors of the mean

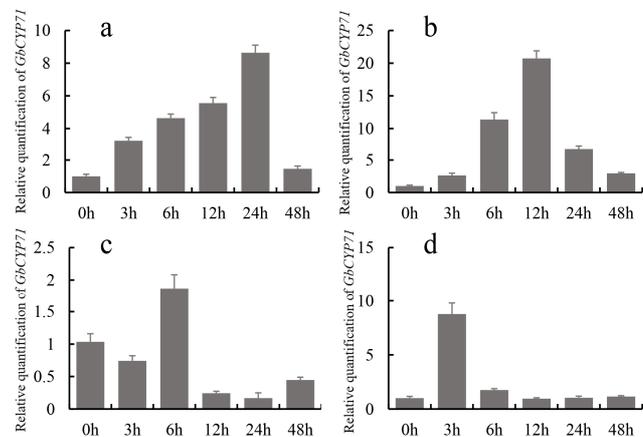


Fig. 7. *GbCYP71* expression following SA (a), ABA (b), ETH (c) and MeJA (d) application. The gene expression levels are expressed as a function of the control values (harvested at time 0 h), which are equal to 1. Each sample was individually assayed in triplicate; error bars indicate standard errors of the mean

contains the classically conserved domains of plant CYP proteins, such as the proline-rich protein region, the I-helix involved in oxygen binding, the E-R-R triad, and the cysteine in the heme-binding region. Blastx also indicated that *GbCYP71* belonged to the CYP71 family. The CYP phylogenetic tree shows that *GbCYP71* was in the gymnosperm cluster and diverged earlier than CYPs from other plants. Some studies have been performed to investigate whether CYP genes (*CYP73A1* through *CYP73A41*) from different species were generally encoding cinnamate 4-hydroxylase (Ehltig *et al.*, 2006). In another study, low amino acid identity (13%) was noted among four CYPs (*CYP73A5*, *CYP75B1*, *CYP84A1*, *CYP98A3*); similar mechanisms for substrate binding were predicted as they mediate the phenylpropanoid pathway of *Arabidopsis* (Rupasinghe *et al.*, 2003). By comparatively analyzing the three CYP crystal structures, Hasemann *et al.* (1995) suggested that although the primary sequences of CYPs were highly diverse, their folding and architecture were remarkably conserved throughout evolution. CYP71 forms the largest CYP family in plants with various functions, of which most are related to plant

secondary metabolism (Gierl, 2009; Nelson and Werck-Reichhart, 2011).

CYPs in plants appear to be differentially expressed in various tissues and organs (Guttikonda *et al.*, 2010; Yamaguchi *et al.*, 2014). In this study, the expression level of *GbCYP71* varied greatly in different tissues; the expression level is highest in the root and lowest in the male flower. Because plants are sessile organisms, they are directly exposed to common adverse environmental factors that are known to induce oxidative stress, such as salinity, drought and low temperature (Bohnert *et al.*, 1995; Mittova *et al.*, 2004). Elucidating the stress perception, signaling, and response mechanisms of plants is crucial to understanding how they improve their stress tolerance in environmentally demanding conditions. In this study, we found that *GbCYP71* expression was significantly induced under salt stress and decreased under extreme temperature stress (4 °C, 40 °C). Our results suggest that *GbCYP71* is highly expressed in the root, which may strengthen ginkgo's resistance to salt stress.

Additionally, *GbCYP71* expression was not induced by high or low temperature; its suppressed expression may be due to the slow growth of callus, which was caused by extreme temperature. Salinity stress is one of the most severe factors limiting plant productivity in abiotic stress; high salt conditions can affect the entire physiological process from seed germination to plant development. Salinity causes osmotic stress in plants in the root zone through decreases in the osmotic potential of surrounding soil (Pasternak, 1987). Temperature stresses are another common factor that can restrict plant growth and development.

ABA, SA and MeJA are hormones in higher-order plants that play important roles in activating plant defense mechanisms in response to biotic and abiotic stresses, such as drought, low temperature, salinity and pathogen attack. Furthermore, ABA, SA and MeJA are important cellular regulators participating in various developmental processes, such as embryo and seed development, root growth, seedling establishment and vegetative development (Cutler *et al.*, 2010; Vlot *et al.*, 2009; Wasternack and Hause, 2002; Wasternack and Parthier, 1997). Osmotic stress elicited by salinity or drought is a very important environmental constraint for land plant survival and distribution. Phytohormones ABA, SA and MeJA can increase in response to osmotic stress. Genes responsive to osmotic stress can also be induced by these exogenous hormones (Borsani *et al.*, 2001; Ishitani *et al.*, 1997; Xu *et al.*, 1994). In this study, *GbCYP71* was induced by all exogenous hormones. ABA, SA and MeJA elicited similar significant increases in gene expression. These results revealed that the *GbCYP71* gene may play a role in resistance to biotic and abiotic stresses, as well as in osmotic regulation.

At present, because an efficient ginkgo regeneration system has not been established with explants, it is difficult to study ginkgo gene functions except via heterologous expression. New technologies could be applied in this field, such as the plant protoplast system, which has been widely used in popular research (Lin *et al.*, 2013). In this study, the *GbCYP71* protein was expressed successfully in transgenic strains. Additional studies could purify recombinant proteins from *E. coli* cells and analyze their enzyme activity. Further analysis of the *GbCYP71* gene and its promoter sequence is required to study the signaling pathways involved in biotic

and abiotic stress regulation and transgenic technology should be applied to study the functions of these cloned genes.

Conclusions

This work characterized the molecular structure of the CYP71 gene from *Ginkgo biloba* and analyzed its expression pattern under various treatments. Bioinformatics analyses demonstrated that *GbCYP71* is a membrane-bound P450 protein with five classically conservative domains. Expression pattern analysis showed that *GbCYP71* is preferentially expressed in roots and that its expression is induced by salinity stress and phytohormones but repressed by cold and heat stresses. This study revealed that *GbCYP71* may participate in plant defense responses to abiotic stress.

Acknowledgments

This work was supported by the Forestry Industry Research Special Funds for Public Welfare Projects (201504105), National Science and Technology Support Program Project (2012BAD21B04), Colleges and Universities in Jiangsu Province Plans to Graduate Research and Innovation (CXZZ13_0552), and Priority Academic Program Development of Jiangsu Higher Education Institutions (PAPD).

References

- Bohnert HJ, Nelson DE, Jensen RG (1995). Adaptations to environmental stresses. *The Plant Cell* 7(7):1099-1111.
- Borsani O, Valpuesta V, Botella MA (2001). Evidence for a role of salicylic acid in the oxidative damage generated by NaCl and osmotic stress in *Arabidopsis* seedlings. *Plant Physiology* 126(3):1024-1030.
- Bozak KR, Yu HONG, Sirevåg R, Christoffersen RE (1990). Sequence analysis of ripening-related cytochrome P-450 cDNAs from avocado fruit. *Proceedings of the National Academy of Sciences* 87(10):3904-3908.
- Chapple C (1998). Molecular-genetic analysis of plant cytochrome P450-dependent monooxygenases. *Annual Review of Plant Biology* 49:311-343.
- Cheng H, Li LL, Cheng SY, Yuan HH, Xu F, Cao FL, ... Wu CH (2013). Expression patterns of a cinnamyl alcohol dehydrogenase gene involved in lignin biosynthesis and environmental stress in *Ginkgo biloba*. *Molecular Biology Reports* 40:707-721.
- Cutler SR, Rodriguez PL, Finkelstein RR, Abrams SR (2010). Abscisic acid: emergence of a core signaling network. *Annual Reviews Plant Biology* 61:651-679.
- Donaldson RP, Luster DG (1991). Multiple forms of plant cytochromes P-450. *Plant Physiology* 96:669-674.
- Ehltng J, Hamberger B, Million-Rousseau R, Werck-Reichhart D (2006). Cytochromes P450 in phenolic metabolism. *Phytochemistry Reviews* 5(2):239-270.
- Gierl A (2009). The cytochrome P450 superfamily of monooxygenases. In: Bennetzen JL, Hake S (Eds). *Handbook of Maize*. Springer pp 731-739.
- Gong YF, Liao ZH, Guo BH, Sun XF, Tang KX (2006). Molecular cloning and expression profile analysis of *Ginkgo biloba* *DXS* gene encoding 1-deoxy-D-xylulose 5-phosphate synthase, the first committed enzyme of

- the 2-C-methyl-D-erythritol 4-phosphate pathway. *Planta Medica* 72(4):329-335.
- Guttikonda SK, Trupti J, Bisht NC, Chen H, Can YQ, Pandey S, ... Yu O (2010). Whole genome co-expression analysis of soybean cytochrome P450 genes identifies nodulation-specific P450 monooxygenases. *BMC Plant Biology* 10:243-261.
- Hasemann CA, Kurumbail RG, Boddupalli SS, Peterson JA, Deisenhofer J (1995). Structure and function of cytochromes P450: a comparative analysis of three crystal structures. *Structure* 3(1):41-62.
- Ishitani M, Xiong L, Stevenson B, Zhu JK (1997). Genetic analysis of osmotic and cold stress signal transduction in *Arabidopsis*: interactions and convergence of abscisic acid-dependent and abscisic acid-independent pathways. *The Plant Cell Online* 9(11):1935-1949.
- Isin EM, Guengerich FP (2007). Complex reactions catalyzed by cytochrome P450 enzymes. *Biochimica et Biophysica Acta* 1770(3):314-329.
- Karimi M, Inzé D, Depicker A (2002). GATEWAY™ vectors for Agrobacterium-mediated plant transformation. *Trends in Plant Science* 7(5):193-195.
- Liao YL, Shen YB, Chang J, Zhang WW, Cheng SY, Xu F (2015). Isolation, expression, and promoter analysis of *GbWRKY2*: a novel transcription factor gene from *Ginkgo biloba*. *International Journal of Genomics* 2015(6):1-17.
- Lin YC, Li W, Sun YH, Kumari S, Wei H, Li QZ, ... Chiang VL (2013). SND1 transcription factor-directed quantitative functional hierarchical genetic regulatory network in wood formation in *Populus trichocarpa*. *The Plant Cell Online* 25(11):4324-4341.
- Livak KJ, Schmittgen TD (2001). Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta CT}$ method. *Methods* 25(4):402-408.
- Mittova V, Guy M, Tal M, Volokita M (2004). Salinity up-regulates the antioxidative system in root mitochondria and peroxisomes of the wild salt-tolerant tomato species *Lycopersicon pennellii*. *Journal of Experimental Botany* 55(399):1105-1113.
- Morant M, Bak S, Møller BL, Werck-Reichhart D (2003). Plant cytochromes P450: tools for pharmacology, plant protection and phytoremediation. *Current Opinion in Biotechnology* 14(2):151-162.
- Nelson D, Werck-Reichhart D (2011). A P450-centric view of plant evolution. *The Plant Journal* 66(1):194-211.
- Nelson DR (2006). Plant cytochrome P450s from moss to poplar. *Phytochemistry Reviews* 5(2):193-204.
- Nelson DR, Koymans L, Kamataki T, Stegeman JJ, Feyereisen R, Waxman DJ, ... Nebert DW (1996). P450 superfamily: update on new sequences, gene mapping, accession numbers and nomenclature. *Pharmacogenetics and Genomics* 6(1):1-42.
- Novo-Uzal E, Gutiérrez J, Martínez-Cortés T, Pomar F (2014). Molecular cloning of two novel peroxidases and their response to salt stress and salicylic acid in the living fossil *Ginkgo biloba*. *Annals of Botany* 114(5):923-936.
- Paquette SM, Bak S, Feyereisen R (2000). Intron-exon organization and phylogeny in a large superfamily, the paralogous cytochrome P450 genes of *Arabidopsis thaliana*. *DNA and Cell Biology* 19(5):307-317.
- Pasternak D (1987). Salt tolerance and crop production-a comprehensive approach. *Annual Review of Phytopathology* 25(1):271-291.
- Royer DL, Hickey LJ, Wing SL (2003). Ecological conservatism in the "living fossil" *Ginkgo*. *Journal Information* 29(1):84-104.
- Rupasinghe S, Baudry J, Schuler MA (2003). Common active site architecture and binding strategy of four phenylpropanoid P450s from *Arabidopsis thaliana* as revealed by molecular modeling. *Protein Engineering* 16(10):721-731.
- Rupasinghe S, Schuler MA (2006). Homology modeling of plant cytochrome P450s. *Phytochemistry Reviews* 5(2):473-505.
- Schuler MA (1996). Plant cytochrome P450 monooxygenases. *Critical Reviews in Plant Sciences* 15(3):235-284.
- Shen G, Pang Y, Wu W, Liao Z, Zhao L, Sun X, Tang K (2006). Cloning and characterization of a root-specific expressing gene encoding 3-hydroxy-3-methylglutaryl coenzyme A reductase from *Ginkgo biloba*. *Molecular Biology Reports* 33(1):117-127.
- Sparkes IA, Runions J, Kearns A, Hawes C (2006). Rapid, transient expression of fluorescent fusion proteins in tobacco plants and generation of stably transformed plants. *Nature Protocols* 1(4):2019-2025.
- Toguri T, Tokugawa K (1994). Cloning of eggplant hypocotyl cDNAs encoding cytochromes P450 belonging to a novel family (CYP77). *FEBS Letters* 338(3):290-294.
- Vlot AC, Dempsey DA, Klessig DF (2009). Salicylic acid, a multifaceted hormone to combat disease. *Annual Review of Phytopathology* 47:177-206.
- Wasternack C, Hause B (2002). Jasmonates and octadecanoids: signals in plant stress responses and development. *Progress in Nucleic Acid Research and Molecular Biology* 72:165-221.
- Wasternack C, Parthier B (1997). Jasmonate-signalled plant gene expression. *Trends in Plant Science* 2(8):302-307.
- Werck-Reichhart D, Feyereisen R (2000). Cytochromes P450: a success story. *Genome Biology* 1(6):3003.1-9.
- Werck-Reichhart D, Hehn A, Didierjean L (2000). Cytochromes P450 for engineering herbicide tolerance. *Trends in Plant Science* 5(3):116-123.
- Werck-Reichhart D, Bak S, Paquette S (2002). Cytochromes P450. In: *The Arabidopsis Book*. Doi: 10.1199/tab.0028.
- Xu F, Cheng H, Cai R, Li LL, Chang J, Zhu J, ... Cheng SY (2008). Molecular cloning and function analysis of an anthocyanidin synthase gene from *Ginkgo biloba*, and its expression in abiotic stress responses. *Mol Cells* 26(6):536-547.
- Xu Y, Chang PL, Liu D, Narasimhan ML, Raghobama KG, Hasegawa PM, Bressan RA (1994). Plant defense genes are synergistically induced by ethylene and methyl jasmonate. *The Plant Cell Online* 6(8):1077-1085.
- Yamaguchi T, Yamamoto K, Asano Y (2014). Identification and characterization of CYP79D16 and CYP71AN24 catalyzing the first and second steps in l-phenylalanine-derived cyanogenic glycoside biosynthesis in the Japanese apricot, *Prunus mume* Sieb. et Zucc. *Plant Molecular Biology* 86(1):215-223.
- Yamazaki S, Sato K, Suhara K, Sakaguchi M, Mihara K, Omura T (1993). Importance of the proline-rich region following signal-anchor sequence in the formation of correct conformation of microsomal cytochrome P-450s. *Journal of Biochemistry* 114(5):652-657.