Identification and analysis of CYP450 genes from transcriptome of *Lonicera japonica* and expression analysis of chlorogenic acid biosynthesis related CYP450s

Xiwu Qi [1, 2], **Xu Yu** [1, 2], **Daohua Xu** [1], **Hailing Fang** [1, 2], **Ke Dong** [3], **Weilin Li** [Corresp., 1, 2], **Chengyuan Liang** [Corresp. 1, 2]

[1] Institute of Botany, Jiangsu Province and Chinese Academy of Sciences, Nanjing, China

[2] The Jiangsu Provincial Platform for Conservation and Utilization of Agricultural Germplasm, Nanjing, China

[3] Department of Biological Sciences, College of Natural Sciences, Seoul National University, Seoul, South Korea

Corresponding Authors: Weilin Li, Chengyuan Liang
Email address: 1923872828@qq.com, liangcy618@aliyun.com

Background. *Lonicera japonica* is an important medicinal plant that has been widely used in traditional Chinese medicine for thousands of years. The pharmacological activities of *L. japonica* are mainly due to its rich natural active ingredients, most of which are secondary metabolites. CYP450s are a large, complex, and widespread superfamily of proteins that participate in many endogenous and exogenous metabolic reactions, especially secondary metabolism. Here, we identified CYP450s in *L. japonica* transcriptome and analyzed CYP450s that may be involved in chlorogenic acid (CGA) biosynthesis. **Methods.** The recent availability of *L. japonica* transcriptome provided opportunity to identify CYP450s in this herb. BLAST based method and HMM based method were used to identify CYP450s in *L. japonica* transcriptome. Then, phylogenetic analysis, conserved motifs analysis, GO annotation, and KEGG annotation analyses were conducted to characterize the identified CYP450s. qRT-PCR was used to explore expression patterns of five CGA biosynthesis related CYP450s. **Results.** In this study, 151 putative CYP450s with complete cytochrome P450 domain, which belonged to 10 clans, 45 families and 76 subfamilies, were identified in *L. japonica* transcriptome. Phylogenetic analysis classified these CYP450s into two major branches, A-type (47%) and non-A type (53%). Both types of CYP450s had conserved motifs in *L. japonica*. The differences of typical motif sequences between A-type and non-A type CYP450s in *L. japonica* were similar with other plants. GO classification indicated that non-A type CYP450s participated in more molecular functions and biological processes than A-type. KEGG pathway annotation totally assigned 47 CYP450s to 25 KEGG pathways. From these data, we cloned two *LjC3Hs* (CYP98A subfamily) and three *LjC4Hs* (CYP73A subfamily) that may be involved in biosynthesis of CGA, the major ingredient for pharmacological activities of *L. japonica*. qRT-PCR results indicated that two *LjC3Hs* exhibited oppositing expression patterns during the flower development and *LjC3H2*

exhibited a similar expression pattern with CGA concentration measured by HPLC. The expression patterns of three *LjC4Hs* were quite different and the expression pattern of *LjC4H3* was quite similar with that of *LjC3H1*. **Discussion.** Our results provide a comprehensive identification and characterization of CYP450s in *L. japonica*. Five CGA biosynthesis related *CYP450s* were cloned and their expression patterns were explored. The different expression patterns of two *LjC3Hs* and three *LjC4Hs* may be due to functional divergence of both substrate and catalytic specificity during plant evolution. The co-expression pattern of *LjC3H1* and *LjC4H3* strongly suggested that they were under coordinated regulation by the same transcription factors due to same *cis* elements in their promoters. In conclusion, this study provides insight into CYP450s and will effectively facilitate the research of biosynthesis of CGA in *L. japonica*.

Identification and analysis of CYP450 genes from transcriptome of *Lonicera japonica* and expression analysis of chlorogenic acid biosynthesis related CYP450s

Xiwu Qi[1,2], Xu Yu[1,2], Daohua Xu[1], Hailing Fang[1,2], Ke Dong[3], Weilin Li[1,2], Chengyuan Liang[1,2]

[1] Institute of Botany, Jiangsu Province and Chinese Academy of Sciences, Nanjing, Jiangsu, China

[2] The Jiangsu Provincial Platform for Conservation and Utilization of Agricultural Germplasm, Nanjing, Jiangsu, China

[3] Department of Biological Sciences, College of Natural Sciences, Seoul National University, Seoul, South Korea

Corresponding Author:

Chengyuan Liang and Weilin Li

Email address: liangcy618@aliyun.com (C. Liang) and 1923872828@qq.com (W. Li)

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.3155v1 | CC BY 4.0 Open Access | rec: 16 Aug 2017, publ: 16 Aug 2017

Abstract

Background. *Lonicera japonica* is an important medicinal plant that has been widely used in traditional Chinese medicine for thousands of years. The pharmacological activities of *L. japonica* are mainly due to its rich natural active ingredients, most of which are secondary metabolites. CYP450s are a large, complex, and widespread superfamily of proteins that participate in many endogenous and exogenous metabolic reactions, especially secondary metabolism. Here, we identified CYP450s in *L. japonica* transcriptome and analyzed CYP450s that may be involved in chlorogenic acid (CGA) biosynthesis.

Methods. The recent availability of *L. japonica* transcriptome provided opportunity to identify CYP450s in this herb. BLAST based method and HMM based method were used to identify CYP450s in *L. japonica* transcriptome. Then, phylogenetic analysis, conserved motifs analysis, GO annotation, and KEGG annotation analyses were conducted to characterize the identified CYP450s. qRT-PCR was used to explore expression patterns of five CGA biosynthesis related CYP450s.

Results. In this study, 151 putative CYP450s with complete cytochrome P450 domain, which belonged to 10 clans, 45 families and 76 subfamilies, were identified in *L. japonica* transcriptome. Phylogenetic analysis classified these CYP450s into two major branches, A-type (47%) and non-A type (53%). Both types of CYP450s had conserved motifs in *L. japonica*. The differences of typical motif sequences between A-type and non-A type CYP450s in *L. japonica* were similar with other plants. GO classification indicated that non-A type CYP450s participated in more molecular functions and biological processes than A-type. KEGG pathway annotation totally assigned 47 CYP450s to 25 KEGG pathways. From these data, we cloned two *LjC3Hs* (CYP98A subfamily) and three *LjC4Hs* (CYP73A subfamily) that may be involved in biosynthesis of CGA, the major ingredient for pharmacological activities of *L. japonica*. qRT-PCR results indicated that two *LjC3Hs* exhibited oppositing expression patterns during the flower development and *LjC3H2* exhibited a similar expression pattern with CGA concentration measured by HPLC. The expression patterns of three *LjC4Hs* were quite different and the expression pattern of *LjC4H3* was quite similar with that of *LjC3H1*.

Discussion. Our results provide a comprehensive identification and characterization of CYP450s in *L. japonica*. Five CGA biosynthesis related *CYP450s* were cloned and their expression

patterns were explored. The different expression patterns of two *LjC3Hs* and three *LjC4Hs* may be due to functional divergence of both substrate and catalytic specificity during plant evolution. The co-expression pattern of *LjC3H1* and *LjC4H3* strongly suggested that they were under coordinated regulation by the same transcription factors due to same *cis* elements in their promoters. In conclusion, this study provides insight into CYP450s and will effectively facilitate the research of biosynthesis of CGA in *L. japonica*.

Introduction

Cytochrome P450 monooxygenases (CYP450s) are a large and complex superfamily which can be found in almost all living organisms (Nelson, 1999). Plant CYP450s are heme-containing enzymes that take part in a wide variety of reactions of both primary and secondary metabolism (Kumar et al., 2014), including the production of fatty acids, sterols, plant hormones, flavonoids, terpenoids, lignin, signaling molecules, and other biological molecules (Schuler & Werckreichhart, 2003).

Lonicera japonica Thunb. is a perennial evergreen vine belonging to the family Caprifoliaceae. *L. japonica* is a medicinal plant of great importance in traditional Chinese medicine that has been used for thousands of years (Shang et al., 2011). There are more than 500 traditional Chinese medicine prescriptions containing *L. japonica* (Shang et al., 2011). Modern pharmacological studies have indicated that the extracts of *L. japonica* possess many biological and pharmacological activities, such as anti-inflammatory, antiviral, antibacterial, antioxidant, hepato-protective, anti-tumor, and other activities (Xiang et al., 2001; Yoo et al., 2008).

The active compounds of *L. japonica* have been extensively studied. Essential oils (Schlotzhauer, Pair & Horvat, 1996), phenolic acids (Lu, Jiang & Chen, 2004), flavone (Chen et al., 2005), triterpenoid saponins (Chai et al., 2005), iridoilds and inorganic elements as the main compositions were isolated and identified in *L. japonica*. Among all these products, chlorogenic acid (CGA) is the major ingredient for pharmacological activities and its content is typically used as the main indicator of quality for evaluating *L. japonica* (Chinese Pharmacopoeia Commission, 2010).

As one of the most important secondary metabolites in plants, CGA is often used in medicines and foods for its high anti-oxidative activity (Zucker & Levy, 1959). The biosynthetic pathway of CGA has been investigated in many plants and is catalyzed by a series of enzymes (Niggeweg, Michael & Martin, 2004). Cinnamate 4-hydroxylase (C4H) and *p*-coumarate 3'-hydroxylase (C3H) are two CYP450s that participate in the two steps of hydroxylation in CGA biosynthetic pathway (Gabriac et al., 1991; Schoch et al., 2001). In *L. japonica*, a CYP98A subfamily gene encoding LjC3H was isolated and characterized. By using heterologous expressed LjC3H in vitro assay, a recent study revealed that the recombinant protein was effective in converting *p*-coumaroylquinate to CGA (Pu et al., 2013). Two *C4Hs* belonging to the CYP73A subfamily were also cloned in *L. japonica*. Expression and activity analysis

suggested that *LjC4H2* may be one of the critical genes that regulate CGA content in *L. japonica* (Yuan et al., 2014).

The studies of *L. japonica* have been focused on the identification of active compounds and pharmacological activity assays. In recent years, with the technological advancement in molecular biology, especially the development of next-generation sequencing technology, great progress has been made in the identification of active compounds involved in the biosynthesis processes in *L. japonica* (Yuan et al., 2012; He et al., 2013). In this study, bioinformatics tools were used to identify and analyze the *CYP450* genes based on transcriptome data of *L. japonica*. We identified two *LjC3Hs* and three *LjC4Hs* from the *CYP450* candidate genes, which including one previously reported *LjC3H* and two *LjC4Hs* genes. We further cloned the five *CYP450* genes and analyzed their transcriptional patterns in different developmental stages flowers. The results provided here will expand CYP450s information and could effectively facilitate CGA biosynthetic studies in *L. japonica*.

Materials and Methods

Identification of CYP450 genes in *L. japonica*

The transcriptome data of *L. japonica* generated from different sequencing platforms including 454 GS-FLX, Illumina HiSeq2000, and Illumina GA II was downloaded from PlantransDB (http://lifecenter.sgst.cn/plantransdb/index.do). Four datasets were assembled and annotated. To identify putative *CYP450* genes, both Hidden Markov Model (HMM) method and BLAST method were used. For HMM method, P450.hmm file which represents the Hidden Markov Model of the cytochrome P450 family was initially downloaded from Pfam (http://pfam.xfam.org/), and then, HMMER3 software (Eddy, 2011) was used to search P450.hmm against *L. japonica* deduced amino acid database. For BLAST method, 19,047 full length plant CYP450 sequences were retrieved from UniProt (http://www.uniprot.org/). These sequences were used as queries to tblastn against *L. japonica* transcriptome assembly with an E-value cutoff of 1e-5. After filtering out the repeated results, the coding sequences of the resultant subjects were retrieved. Finally, results from the two methods were integrated and corrected manually. The identification methods were conducted for the four datasets of *L. japonica* and the results were also integrated and corrected. The corrected *L. japonica* CYP450s were further submitted to NCBI Conserved Domain Search

(http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi) to predict the conserved domain. Sequences with complete cytochrome P450 domain were selected for further analysis.

Classification and characterization of *L. japonica* CYP450 genes

L. japonica CYP450s were classified into different families and subfamilies according to the sequence similarity using sequences from Cytochrome P450 Homepage as reference sequences. If the amino acid sequences of *L. japonica* CYP450s showed >40%, >55%, or >95% sequence similarity with reference sequences, they were classified into the same family, subfamily, or allelic variant, respectively (Nelson, 2009).

The deduced amino acid sequences of *L. japonica* CYP450s were subjected to Multiple Expectation Maximization for Motif Elicitation (MEME, http://meme-suite.org/) (Bailey et al., 2009) analysis for identification of conserved motifs. Sequences of the four conserved CYP450 motifs including heme-binding region, PERF motif, K-helix region and I-helix region were extracted and then subjected to WEBLOGO (http://weblogo.berkeley.edu/) (Crooks et al., 2004) to create the sequence logos.

Phylogenetic analysis of predicted CYP450 genes

A total of 63 representative sequences from plant CYP450 families were selected for phylogenetic analysis with 151 *L. japonica* CYP450 sequences. Specifically, CYP450 sequences whose functions had already been identified were preferentially selected. Multiple sequence alignment was performed using MUSCLE 3.6 software (Edgar, 2004). The result of alignment was imported to MEGA4 (Tamura et al., 2007) and phylogenetic analysis was performed. The phylogenetic tree was constructed using the Neighbor-Joining algorithm with the Poisson model and pairwise deletion. Bootstrap testing with 1,000 replications was used to test the phylogenetic tree. The Newick format file of bootstrap consensus tree was exported and then modified using EvolView (http://www.evolgenius.info/evolview/) (Zhang et al., 2012).

Gene ontology and KEGG pathway analysis

Blast2GO (http://www.blast2go.com/) (Conesa et al., 2005) was used to perform Gene ontology (GO) annotation of *L. japonica* CYP450s. These predicted genes were functionally categorized according to three different criterions including cellular component, molecular function and biological process. The GO terms of all *L. japonica* CYP450s were extracted and subjected to Web Gene Ontology Annotation Plot (WEGO, http://wego.genomics.org.cn/cgi-bin/wego/index.pl) (Ye et al., 2006) to plot GO annotation results. KEGG annotation that maps

the *L. japonica* CYP450s to possible KEGG pathway for biological interpretation of systemic functions was also conducted using Blast2GO.

Extraction and quantification of CGA

The *L. japonica* used for this study was maintained at the Germplasm Nursery in Institute of Botany, Jiangsu Province and Chinese Academy of Sciences, Nanjing, Jiangsu Province. Flower buds and flowers samples for CGA and RNA extraction were collected at five stages: young alabastrum (YA, ≤ 1.5 cm), green alabastrum (GA, 2.0 – 3.0 cm), while alabastrum (WA, 3.2 – 4.4 cm), silvery flower (SF, about 5 cm), and golden flower (GF, about 5 cm). The extraction and quantification of CGA were conducted as described in Chinese Pharmacopoeia with minor modifications (Chinese Pharmacopoeia Commission, 2010). Briefly, dried buds or flowers were separately comminuted with a miler, and 0.2 g of each solid sample (40 mesh) was extracted with 25 mL of 50% aqueous methanol by ultrasonication (250 W, 35 kHz) for 30 min. After cooling to room temperature, the extracts were replenished to earlier weights with 50% aqueous methanol. Then, 5 mL of the extracts were diluted to 25 mL with 50% aqueous methanol and filtered with 0.45 μm Millipore filter membranes. An Agilent 1200LC series HPLC system was used to analyze the CGA levels. Separations were performed on an Agilent TC-C18 reserved-phase column (5 μm, 250 mm×4.6 mm) at 25 °C. The mobile phase was composed of acetonitrile-0.4% H_3PO_4 (13: 87). The flow rate was 1 mL/min and fractions were monitored at 327 nm. Components were identified by comparison of the retention times of the eluting peaks to those of commercial standards under the same conditions.

RNA extraction and qRT‑PCR

Total RNA from five samples was extracted using RNAiso Plus (Takara, Japan) according to the manufacturer's instructions. RNA quality and concentration were measured using a ND-1000 UV spectrophotometer (Nanodrop Technologies, USA). First-strand cDNA was synthesized using 3 μg of total RNA with M-MLV reverse transcriptase (Promega, USA) in a 25 μl reaction system. For quantitative real-time reverse transcriptional PCR (qRT‑PCR), each reaction was prepared according to the manufacturer's instructions using SYBR® Premix Ex TaqTM II (Takara) and 2 μl of diluted cDNA as a template. The qRT‑PCR reactions were conducted on the qTOWER2.2 Real Time PCR Systems (Analytik Jena, Germany). The *L. japonica* actin gene was used as a control to normalize the relative expression levels of target

genes. Gene-specific primers used for qRT‑PCR were listed on Supplemental Table 1. All results were representative of three independent experiments.

Results

Identification and classification of CYP450 genes in *L. japonica*

Cytochrome P450 is one of the most massive gene superfamilies that is comprised of a number of families and subfamilies. In the present study, by integrating the results from different datasets and manual correction, we in total identified 151 putative CYP450s with complete cytochrome P450 domain in *L. japonica*. Among them, nine CYP450s had been previously reported and the other 142 CYP450s were identified here for the first time in *L. japonica*. Based on sequence similarity, we classified the 151 *CYP450* genes from *L. japonica* into 10 clans consisting of 45 families and 76 subfamilies (Table 1). Among them, the CYP71 clan, which represents the whole set of A-type *CYP450* genes, contains 71 genes belonging to 19 families (CYP71, CYP73, CYP75, CYP76, CYP77, CYP78, CYP79, CYP80, CYP81, CYP82, CYP83, CYP84, CYP89, CYP92, CYP93, CYP98, CYP701, CYP706, and CYP736). The non-A type *CYP450* genes of *L. japonica* contains the remaining 80 genes, which belongs to 9 CYP clans (CYP51, 72, 74, 85, 86, 97, 710, 711, and 727) and 26 families (CYP51, CYP72, CYP714, CYP715, CYP721, CYP734, CYP749, CYP74, CYP85, CYP87, CYP88, CYP90, CYP707, CYP716, CYP722, CYP724, CYP728, CYP729, CYP86, CYP94, CYP96, CYP704, CYP97, CYP710, CYP711, and CYP727). The largest CYP family of *L. japonica* is CYP71 and CYP72, containing 17 and 18 members, respectively.

Phylogenetic analysis of predicted CYP450s in *L. japonica*

Representative members of each plant CYP450 family were selected and used to conduct phylogenetic analysis with 151 CYP450s from *L. japonica*. The predicted CYP450s were classified into two major branches, A-type (47%) and non-A type (53%) (Fig. 1). There were 10 clans in plants CYP450s. Four clans contained multiple families and were designated by their lowest-numbered family members, CYP71, CYP72, CYP85 and CYP86. The other six clans were designated by their only family, CYP51, CYP74, CYP97, CYP710, CYP711 and CYP727. In *L. japonica*, all 10 clans were identified. Genes belonging to same clan clustered as one clade. For example, the 72 clan, which comprised 28 CYP450s belonged to six families, were clustered as one clade with the 8 representative CYP450s. The CYP71 clan that comprised 71 members

belonging to 19 families was the largest clan. Three clans, CYP710, CYP711 and CYP727, had only one member identified for each clan.

Conserved motifs analysis of *L. japonica* CYP450s

Plant CYP450s shared some typical conserved motifs including heme-binding region, PERF motif, K-helix region and I-helix region, which were important for catalytic activities (Paquette, Jensen & Bak, 2009). The *L. japonica* CYP450s were divided into A-type and non-A type according to phylogenetic analysis. The deduced amino acid sequences were subjected to MEME to analyze the conserved motifs. The consensus sequences of the heme-binding region, also known as "P450 signature", were "PFGXGRRXCPG" and "XFXXGXRXCXG" for A-type and non-A type CYP450s, respectively (Fig. 2). The cysteine residues in this motif of two types of CYP450s were universally conserved, which links the heme iron to the apoprotein. The consensus sequences of the PERF motif were also different for two types of CYP450s in *L. japonica*, which are "PERF" for A-type and "PXRX" for non-A type. The R residues in the PERF motif and E and R residues in the K-helix region were universally conserved, which form a salt bridge that has been proposed to be involved in locking the Cys-pocket in position and assuring the stable association of heme with the protein. The threonine residues in the I-helix region which is involved in oxygen activation was highly conserved in both A-type and non-A type CYP450s. In general, sequences of the typical motifs were conserved in *L. japonica* CYP450s, and the differences between A-type and non-A type CYP450s in *L. japonica* were similar with other plants (Chen et al., 2014).

Gene ontology classification of *L. japonica* CYP450s

Gene ontology (GO) is a classification system for standardized gene functions which classifies genes into three main independent GO categories: cellular component, molecular function and biological process. In this study, GO assignments were conducted to classify the functions of CYP450s from *L. japonica* using Blast2GO. Results indicated that all 151 CYP450s were mapped to one or more GO terms, of which 145 were assigned to the "cellular component", 151 to the "molecular function", and 151 to the "biological process" (Fig. 3). Of these categories, cell, binding, catalytic, and metabolic process were the largest subcategories. Comparison of the GO classification between the A-type and non-A type CYP450s, we found that non-A type CYP450s participated in more molecular functions and biological processes than A-type. For example, GO terms of non-A type CYP450s in molecular function category included

demethylase, hydrolase, lyase, and transferase; however, no A-type CYP450s was assigned to these subcategories. In biological process category, non-A type CYP450s participated in more biological processes than A-type, including anatomical structure formation, cellular component organization, developmental process, establishment of localization, growth, localization, multicellular organismal process, and reproduction. The GO annotation provided a valuable clue to investigate the functions of CYP450s in *L. japonica*.

KEGG pathway analysis of *L. japonica* CYP450s

In order to further understand the biological functions of CYP450s in *L. japonica*, pathway-based analysis was performed. Given that a CYP450 could be assigned to one or more KEGG pathways as well as GO terms, 47 (31.1%) CYP450s were totally assigned to 25 KEGG pathways (Fig. 4). The 25 pathways could be mainly grouped into six classes, including lipid metabolism, amino acid metabolism, metabolism of cofactors and vitamins, metabolism of terpenoids and polyketides, biosynthesis of other secondary metabolites, and xenobiotics biodegradation and metabolism. In the class of 'biosynthesis of other secondary metabolites', after removing duplicate hits, ten CYP450s (CYP73A-m13469, CYP73A-m177245, CYP73A-m8810, CYP75B-m13120, CYP76A-m155830, CYP78A-m152788, CYP93B-m79556, CYP98A-m184946, CYP98A-m43608 and CYP736A-m18282) were found to be involved in the biosynthesis of phenolic compounds including phenylpropanoid, stilbenoid, diarylheptanoid and gingerol, flavonoid, flavone and flavonol, and isoflavonoid. All ten CYP450s belonged to CYP71 clan. In the class of 'metabolism of terpenoids and polyketides', nine CYP450s (CYP72A-m132911, CYP72A-m20456, CYP72A-m206268, CYP72D-m62754, CYP72D-m75640, CYP714A-m189781, CYP714E-m17561, CYP714E-m205273 and CYP734A-m842) were found to be involved in 'monoterpenoid biosynthesis', all of which belonged to CYP72 clan. Three CYP450s (CYP701A-m150262, CYP701A-m27329 and CYP728B-m166264) were found to be involved in 'diterpenoid biosynthesis', among them, two belonged to CYP71 clan and one belonged to CYP85 clan. Five CYP450s (CYP707A-m213600, CYP707A-m35702, CYP707A-m47109, CYP707A-m212742 and CYP728B-m166264) were found to be involved in 'carotenoid biosynthesis', all of which belonged to CYP85 clan.

CYP450s involved in CGA biosynthesis

CGA is the most major active ingredient in *L. japonica* and the biosynthetic pathway of CGA has been investigated in many plants. In CGA biosynthetic pathway, C4H and C3H are the

two CYP450-encoded enzymes that participate in the two steps of hydroxylation. In the present study, three *C4H* and two *C3H* genes were identified and cloned from *L. japonica*. Among them, two *LjC4Hs* and one *LjC3H* have been previously reported. The newly identified *C4H* and *C3H* were designated as '*LjC4H3'* (GenBank accession number: KX845341) and '*LjC3H2'* (GenBank accession number: KX845342), respectively. The *C4Hs* belonged to CYP73A subfamily and *C3Hs* belonged to CYP98A subfamily. Phylogenetic analysis indicated that two clades were clustered for C4Hs and C3Hs from *L. japonica* and other plants (Fig. 5).

Because CGA was mainly accumulated in flower bud of *L. japonica*, buds and flowers in different developmental stages were selected to explore the relationship of *C4H* and *C3H* expressions and CGA contents. HPLC analysis was used to measure CGA concentrations in different developmental stages of buds and flowers. As shown in Fig. 6, the percentage of CGA contents decreased during the flower development. Nevertheless, with the increase of bud or flower weights, the total CGA contents increased from young alabastrum (YA) to while alabastrum (WA) stage and reached peak at the WA stage. After flowering, the total CGA contents decreased quickly during flower development. Furthermore, qRT-PCR was conducted to analyze the transcriptional levels of CGA biosynthetic genes in the different developmental stages of buds and flowers, including the five CYP450s identified in this study. The two *LjC3Hs* exhibited oppositing expression patterns, the transcriptional levels of *LjC3H1* increased but that of *LjC3H2* decreased during the flower development (Fig. 7). The expression patterns of three *LjC4Hs* were quite different and the relative expression levels of *LjC4H3* was obviously higher than those of the other two (Fig. 7). Interestingly, the expression patterns of *LjPAL1*, *LjC4H3*, *LjC3H1* and *LjHQT* were quite similar, which exhibited a trend of decreasing first and then increasing. Considering the gene expressions with CGA contents, only *LjC3H2* exhibited a similar pattern with CGA concentrations.

Discussion

L. japonica is an important medicinal plant that has been widely used in traditional Chinese medicine for thousands of years. The pharmacological activities of this medicinal plant are mainly due to its rich natural active ingredients, most of which are secondary metabolites. CYP450s are a large, complex, and widespread superfamily that participate in many metabolic reactions, especially secondary metabolism. The identification and characterization of *CYP450s* in *L. japonica* will effectively facilitate the study of natural active compounds biosynthesis. In

this study, we identified 151 putative CYP450s with complete cytochrome P450 domain from transcriptome data of *L. japonica*. According to the classification criteria, the 151 CYP450s were classified into 10 clans consisting of 45 families and 76 subfamilies. Next, we conducted phylogenetic analysis, conserved motifs analysis, GO annotation, and KEGG annotation to characterize the identified CYP450s. As mentioned above, nine CYP450s have been previously reported in *L. japonica*, which were also identified among the 151 CYP450s of this study. These results indicated that the identified CYP450s from the *L. japonica* transcriptome data in this study were quite comprehensive.

The evolution of plant CYP450s can be divided into three major groups: CYP450s involved in sterol and carotenoid biosynthesis were the most ancient, CYP450s involved in adaptation to land environment were the next oldest, and CYP450s involved in biosynthesis of plant secondary metabolites were the most recent to evolve (Morant et al., 2007; Nelson et al., 2008). In this study, ten CYP450s (CYP73A-m13469, CYP73A-m177245, CYP73A-m8810, CYP75B-m13120, CYP76A-m155830, CYP78A-m152788, CYP93B-m79556, CYP98A-m184946, CYP98A-m43608, and CYP736A-m18282) were annotated to participate in the biosynthesis of phenolic compounds, a most common type of secondary metabolite in plants, including phenylpropanoid, stilbenoid, flavonoid, and isoflavonoid. All ten CYP450s belonged to CYP71 clan. As earlier reported, the most recently evolved CYP450 group comprises the highly proliferated clan 71. This clan includes CYP450s involved in the biosynthesis of the majority of plant secondary metabolites involved in adaptation to abiotic and biotic stress (Morant et al., 2007), with which our present findings are in agreement. Five CYP450s (CYP707A-m213600, CYP707A-m35702, CYP707A-m47109, CYP707A-m212742, and CYP728B-m166264) were found to be involved in carotenoid biosynthesis, all of which belonged to the CYP85 clan. These CYP450s belonged to the oldest group with a function that preceded the colonization of land by plants (Morant et al., 2007).

CGA is the major active ingredient in *L. japonica*, and the biosynthetic pathway of CGA has been investigated in many plants. In CGA biosynthetic pathway, C4H and C3H are two CYP450 encoded enzymes that participate in the two steps of hydroxylation (Gabriac et al., 1991; Schoch et al.,2001). In *L. japonica*, a gene encoding LjC3H has been isolated and characterized by Pu et al. (2013), and was identified as CYP98A subfamily member. *In vitro* assay using heterologous expressed LjC3H revealed that the recombinant protein was effective in converting *p*-coumaroylquinate to CGA. Southern blotting suggested that the gene was present in the genome

in two copies, but unfortunately, only one copy of *LjC3H* was obtained. In this study, two *LjC3Hs* were identified and cloned from *L. japonica*, both of which belonged to the CYP98A subfamily. Among the two *LjC3Hs*, one was same as the *LjC3H* reported by Pu et al. (2013), the other is a newly identified gene and is hereby designated *LjC3H2*. These results suggested that the newly identified *LjC3H2* was the other copy of *LjC3H* in the genome of *L. japonica*. Two *C4Hs* were also cloned in *L. japonica* by Yuan et al. (2014), which belonged to the CYP73A subfamily. Expression and activity analysis suggested that *LjC4H2* may be one of the critical genes that regulate CGA content in *L. japonica*. In our study, three *C4Hs* were identified and cloned from *L. japonica*, including the previously reported two genes. The newly identified *LjC4H* was designated as *LjC4H3*, which showed high degree of sequence homology with *LjC4H1*. Phylogenetic analysis showed that LjC4H1 and LjC4H3 clustered to one clade. This result suggested that these two genes may be generated by recent gene duplication.

In the present study, the expression patterns of two *LjC3Hs* and three *LjC4Hs* were quite different during the flower development. This phenomenon that different members of the same family exhibit different expression patterns during development was also observed in other plants (Bi et al., 2011; Qi et al., 2014), which might be caused by functional divergence of both substrate and catalytic specificity during plant evolution (Helariutta et al., 1996; Xu et al., 2009). Considering the gene expressions with CGA contents, only *LjC3H2* exhibited a similar pattern with CGA concentrations in our study. This result was similar with that of coffee (Lepelley et al., 2007). In coffee, transcriptional levels of CGA biosynthetic genes and CGA contents were measured during grain development and *C3H1* showed a similar expression pattern with CGA concentrations. Both the CGA concentrations and *C3H* expression pattern were similar with those of *L. japonica*, respectively. However, in this study, the expression patterns of *LjC3H1* and three *LjC4Hs* were inconsistent with CGA contents during flower development. The reason for this phenomenon could be that C3H and C4H not only participated in CGA biosynthesis, but were also involved in other metabolites. The product catalyzed by C4H was a common precursor in phenylpropanoid metabolism, including flavonoids, anthocyanins, condensed tannins, and isoflavonoids (Winkel-Shirley, 2001). C3H was also a key enzyme in lignin biosynthesis (Boerjan, Ralph & Baucher, 2003). It is likely that the complexity of the metabolic pathways led to the inconsistency between gene expressions and product contents.

In this study, the expression patterns of *LjPAL1*, *LjC4H3*, *LjC3H1* and *LjHQT* were quite similar during flower development. The co-expression patterns of these four genes strongly

suggested that they were under coordinated regulation by the same transcription factors due to similar *cis* elements in their promoters (Bi et al., 2011). In apple, anthocyanin biosynthetic genes including *CHS*, *CHI*, *F3H*, *DFR*, *LDOX* and *UFGT* showed similar expression patterns during fruit development, which were coordinately regulated by a MYB transcription factor, *MdMYB10* (Espley et al., 2007). Fruit-specific ectopic expression of *AtMYB12* in tomato led to upregulation of all biosynthetic genes required for the production of flavonols and their derivatives, including *PAL*, *C4H*, *4CL*, *CHS*, *CHI*, *F3H*, *F3'H*, *FLS*, *ANS*, *C3H*, *HCT*, *HQT*, *GT*, and *RT*; and, in addition, led to the increase of flavonols and their derivatives (Luo et al., 2008). In pine and eucalyptus, xylem-associated MYB transcription factors could bind to the AC elements and activate the transcription of the lignin biosynthetic genes (Patzlaff et al., 2003; Goicoechea et al., 2005). Moreover, the rice genome sequence analysis revealed that ACII motif existed in the promoters of many lignin biosynthetic genes, including *PAL*, *4CL*, *C4H*, *C3H*, *CCoAOMT*, *CCR*, and *CAD*, suggesting that they were under coordinated regulation by the same transcription factors (Bi et al., 2011).

Conclusions

In this study, we identified 151 putative CYP450s with complete cytochrome P450 domain in *L. japonica* transcriptome, 142 of which were identified here for the first time. According to the classification criteria, the 151 CYP450s were classified into 10 clans consisting of 45 families and 76 subfamilies. Next, we conducted phylogenetic analysis, conserved motifs analysis, GO annotation, and KEGG annotation to characterize the identified CYP450s. From these data, we cloned two *LjC3Hs* (CYP98A subfamily) and three *LjC4Hs* (CYP73A subfamily) genes that may be involved in biosynthesis of CGA, including the newly identified *LjC3H2* and *LjC4H3*. Furthermore, qRT-PCR and HPLC results indicated that only *LjC3H2* exhibited a similar expression pattern with CGA concentration. Different members of the same family exhibited different expression patterns during development that may be due to functional divergence of both substrate and catalytic specificity during plant evolution. The co-expression pattern of *LjPAL1*, *LjC4H3*, *LjC3H1* and *LjHQT* strongly suggested that they were under coordinated regulation by the same transcription factors due to same *cis* elements in their promoters. In conclusion, this study provides insight into CYP450s and will effectively facilitate the research of biosynthesis of CGA in *L. japonica*.

References

Bailey TL, Boden M, Buske FA, Frith M, Grant CE, Clementi L, Ren J, Li WW, Noble WS. 2009. MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Research* 37:W202–W208 DOI 10.1093/nar/gkp335.

Bi C, Chen F, Jackson L, Gill BS, Li W. 2011. Expression of lignin biosynthetic genes in wheat during development and upon infection by fungal pathogens. *Plant Molecular Biology Reporter* 29:149-161 DOI 10.1007/s11105-010-0219-8.

Boerjan W, Ralph J, Baucher M. 2003. Lignin biosynthesis. *Annual Review of Plant Biology* 54:519-546 DOI 10.1146/annurev.arplant.54.031902.134938.

Chai XY, Li SL, Li P. 2005. Quality evaluation of *Flos Lonicerae* through a simultaneous determination of seven saponins by HPLC with ELSD. *Journal of Chromatography A* 1070:43-48 DOI 10.1016/j.chroma.2005.02.031.

Chen H, Wu B, Nelson DR, Wu K, Liu C. 2014. Computational identification and systematic classification of novel cytochrome P450 genes in *Salvia miltiorrhiza*. *PLoS ONE* 9:e115149 DOI 10.1371/journal.pone.0115149.

Chen J, Li SL, Li P, Song Y, Chai XY, Ma DY. 2005. Qualitative and quantitative analysis of active flavonoids in *Flos Lonicerae* by capillary zone electrophoresis coupled with solid−phase extraction. *Jorunal of Separation Science* 28:365-372 DOI 10.1002/jssc.200400024.

Chinese Pharmacopoeia Commission. 2010. The pharmacopoeia of the People's Republic of China, 2010 ed. China Medical Science Press: Beijing, China, 205–206

Conesa A, Götz S, García-Gómez JM, Terol J, Talón M, Robles M. 2005. Blast2go: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* 21:3674-3676 DOI 10.1093/bioinformatics/bti610.

Crooks GE, Hon G, Chandonia JM, Brenner SE. 2004. WebLogo: a sequence logo generator. *Genome Research* 14:1188-1190 DOI 10.1101/gr.849004.

Eddy SR. 2011. Accelerated profile HMM searches. *PLoS Computational Biology* 7:e1002195 DOI 10.1371/journal.pcbi.1002195.

Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research* 32:1792-1797 DOI 10.1093/nar/gkh340.

Espley RV, Hellens RP, Putterill J, Stevenson DE, Kutty-Amma S, Allan AC. 2007. Red colouration in apple fruit is due to the activity of the MYB transcription factor, MdMYB10. *Plant Journal* 49:414-427 DOI 10.1111/j.1365-313X.2006.02964.x.

426 Gabriac B, Werck-Reichhart D, Teutsch H, Durst F. 1991. Purification and
427 immunocharacterization of a plant cytochrome p450: the cinnamic acid 4-hydroxylase.
428 *Archives of Biochemistry and Biophysics* 288:302-309 DOI 10.1016/0003-9861(91)90199-S.
429 Goicoechea M, Lacombe E, Legay S, Mihaljevic S, Rech P, Jauneau A, Lapierre C, Pollet B,
430 Verhaegen D, Chaubet-Gigot N, Grima-Pettenati J. 2005. EgMYB2, a new transcriptional
431 activator from Eucalyptus xylem, regulates secondary cell wall formation and lignin
432 biosynthesis. *Plant Journal* 43:553-567 DOI 10.1111/j.1365-313X.2005.02480.x.
433 He L, Xu X, Li Y, Li C, Zhu Y, Yan H, Sun Z, Sun C, Song J, Bi Y, Shen J, Cheng R, Wang Z,
434 Xiao W, Chen S. 2013. Transcriptome analysis of buds and leaves using 454
435 pyrosequencing to discover genes associated with the biosynthesis of active ingredients in
436 *Lonicera japonica* Thunb. *PLoS ONE* 8:e62922 DOI 10.1371/journal.pone.0062922.
437 Helariutta Y, Kotilainen M, Elomaa P, Kalkkinen N, Bremer K, Teeri TH, Albert VA. 1996.
438 Duplication and functional divergence in the chalcone synthase gene family of Asteraceae:
439 evolution with substrate change and catalytic simplification. *Proceedings of the National*
440 *Academy of Sciences of the United States of America* 93:9033-9038 DOI
441 10.1073/pnas.93.17.9033.
442 Kumar MS, Babu PR, Rao KV, Reddy VD. 2014. Organization and classification of cytochrome
443 P450 genes in castor (*Ricinus communis* L.). *Proceedings of the National Academy of*
444 *Sciences India section B* 84:131-143 DOI 10.1007/s40011-013-0192-8.
445 Lepelley M, Cheminade G, Tremillon N, Simkin A, Ciaillet V, McCarthy J. 2007. Chlorogenic
446 acid synthesis in coffee: an analysis of CGA content and real-time RT-PCR expression of
447 HCT, HQT, C3H1, and CCoAOMT1 genes during grain development in *C. canephora*.
448 *Plant Science* 172:978-996 DOI 10.1016/j.plantsci.2007.02.004.
449 Lu HT, Jiang Y, Chen F. 2004. Application of preparative high-speed counter-current
450 chromatography for separation of chlorogenic acid from *Flos Lonicerae*. *Journal of*
451 *Chromatography A* 1026:185-190 DOI 10.1016/j.chroma.2003.11.004.
452 Luo J, Butelli E, Hill L, Parr A, Niggeweg R, Bailey P, Weisshaar B, Martin C. 2008. AtMYB12
453 regulates caffeoyl quinic acid and flavonol synthesis in tomato: expression in fruit results in
454 very high levels of both types of polyphenol. *Plant Journal* 56:316-326 DOI
455 10.1111/j.1365-313X.2008.03597.x.
456 Morant M, Jørgensen K, Schaller H, Pinot F, Møller BL, Werck-Reichhart D, Bak S. 2007.
457 CYP703 is an ancient cytochrome P450 in land plants catalyzing in-chain hydroxylation of

lauric acid to provide building blocks for sporopollenin synthesis in pollen. *Plant Cell* 19:1473-1487 DOI 10.1105/tpc.106.045948.

Nelson DR. 1999. Cytochrome P450 and the individuality of species. *Archives of Biochemistry and Biophysics* 369:1-10 DOI 10.1006/abbi.1999.1352.

Nelson DR. 2009. The cytochrome P450 homepage. *Human Genomics* 4:59-65 DOI 10.1186/1479-7364-4-1-59.

Nelson DR, Ming R, Alam M, Schuler MA. 2008. Comparison of cytochrome P450 genes from six plant genomes. *Tropical Plant Biology* 1:216-235 DOI 10.1007/s12042-008-9022-1.

Niggeweg R, Michael AJ, Martin C. 2004. Engineering plants with increased levels of the antioxidant chlorogenic acid. *Nature Biotechnology* 22:746-754 DOI 10.1038/nbt966.

Paquette SM, Jensen K, Bak S. 2009. A web-based resource for the Arabidopsis P450, cytochromes b5, NADPH-cytochrome P450 reductases, and family 1 glycosyltransferases (http://www.P450.kvl.dk). *Phytochemistry* 70:1940–1947 DOI 10.1016/j.phytochem.2009.08.024.

Patzlaff A, McInnis S, Courtenay A, Surman C, Newman LJ, Smith C, Bevan MW, Mansfield S, Whetten RW, Sederoff RR, Campbell MM. 2003. Characterisation of a pine MYB that regulates lignification. *Plant Journal* 36:743-754 DOI 10.1046/j.1365-313X.2003.01916.x.

Pu G, Wang P, Zhou B, Liu Z, Xiang F. 2013. Cloning and characterization of *Lonicera japonica* *p*-coumaroyl ester 3-hydroxylase which is involved in the biosynthesis of chlorogenic acid. *Bioscience Biotechnology and Biochemistry* 77:1403-1409 DOI 10.1271/bbb.130011.

Qi X, Shuai Q, Chen H, Fan L, Zeng Q, He N. 2014. Cloning and expression analyses of the anthocyanin biosynthetic genes in mulberry plants. *Molecular Genetics and Genomics* 289:783-793 DOI 10.1007/s00438-014-0851-3.

Schlotzhauer WS, Pair SD, Horvat RJ. 1996. Volatile constituents from the flowers of Japanese honeysuckle (*Lonicera japonica*). *Journal of Agricultural and Food Chemistry* 44:206-209 DOI 10.1021/jf950275b.

Schoch G, Goepfert S, Morant M, Hehn A, Meyer D, Ullmann P, Werck-Reichhart D. 2001. CYP98A3 from *Arabidopsis thaliana* is a 3'-hydroxylase of phenolic esters, a missing link in the phenylpropanoid pathway. *Journal of Biological Chemistry* 276:36566-36574 DOI 10.1074/jbc.M104047200.

Schuler MA, Werckreichhart D. 2003. Functional genomic of P450s. *Annual Review of Plant Biology* 54:629-667 DOI 10.1146/annurev.arplant.54.031902.134840.

490 Shang X, Pan H, Li M, Miao X, Ding H. 2011. *Lonicera japonica* Thunb.: ethnopharmacology,
491 phytochemistry and pharmacology of an important traditional Chinese medicine. *Journal of*
492 *Ethnopharmacology* 138:1-21 DOI 10.1016/j.jep.2011.08.016.
493 Tamura K, Dudley J, Nei M, Kumar S. 2007. MEGA4: Molecular Evolutionary Genetics
494 Analysis (MEGA) software version 4.0. *Molecular Biology and Evolution* 24:1596-1599
495 DOI 10.1093/molbev/msm092.
496 Winkel-Shirley B. 2001. Flavonoid biosynthesis. a colorful model for genetics, biochemistry,
497 cell biology, and biotechnology. *Plant Physiology* 126:485-493 DOI 10.1104/pp.126.2.485.
498 Xiang T, Xiong QB, Ketut, AI, Tezuka Y, Nagaoka T, Wu LJ, Kadota S. 2001. Studies on the
499 hepatocyte protective activity and the structure-activity relationships of quinic acid and
500 caffeic acid derivatives from the flower buds of *Lonicera bournei*. *Planta Medica* 67:322-
501 325 DOI 10.1055/s-2001-14337.
502 Xu Z, Zhang D, Hu J, Zhou X, Ye X, Reichel KL, Stewart NR, Syrenne RD, Yang X, Gao P, Shi
503 W, Doeppke C, Sykes RW, Burris JN, Bozell JJ, Cheng MZ, Hayes DG, Labbe N, Davis M,
504 Stewart CN Jr, Yuan JS. 2009. Comparative genome analysis of lignin biosynthesis gene
505 families across the plant kingdom. *BMC Bioinformatics* 10(Suppl): S3 DOI 10.1186/1471-
506 2105-10-S11-S3.
507 Ye J, Fang L, Zheng H, Zhang Y, Chen J, Zhang Z, Wang J, Li S, Li R, Bolund L, Wang J. 2006.
508 WEGO: a web tool for plotting GO annotations. *Nucleic Acids Research* 34:W293-W297
509 DOI 10.1093/nar/gkl031.
510 Yoo HJ, Kang HJ, Song YS, Park EH, Lim CJ. 2008. Anti-angiogenic, antinociceptive and anti-
511 inflammatory activities of *Lonicera japonica* extract. *Journal of Pharmacy & Pharmacology*
512 60:779–786 DOI 10.1211/jpp.60.6.0014.
513 Yuan Y, Song L, Li M, Liu G, Chu Y, Ma L, Zhou Y, Wang X, Gao W, Qin S, Yu J, Wang X,
514 Huang L. 2012. Genetic variation and metabolic pathway intricacy govern the active
515 compound content and quality of the Chinese medicinal plant *Lonicera japonica* thumb.
516 *BMC Genomics* 13:195 DOI 10.1186/1471-2164-13-195.
517 Yuan Y, Wang Z, Jiang C, Wang X, Huang L. 2014. Exploiting genes and functional diversity of
518 chlorogenic acid and luteolin biosyntheses in *Lonicera japonica* and their substitutes. *Gene*
519 534:408-416 DOI 10.1016/j.gene.2012.09.051.

Zhang H, Gao S, Lercher MJ, Hu S, Chen WH. 2012. EvolView, an online tool for visualizing, annotating and managing phylogenetic trees. *Nucleic Acids Research* 40: W569-W572 DOI 10.1093/nar/gks576.

Zucker M, Levy CC. 1959. Some factors which affect the synthesis of chlorogenic acid in disks of potato tuber. *Plant Physiology* 34:108-112 DOI 10.1104/pp.34.2.108.

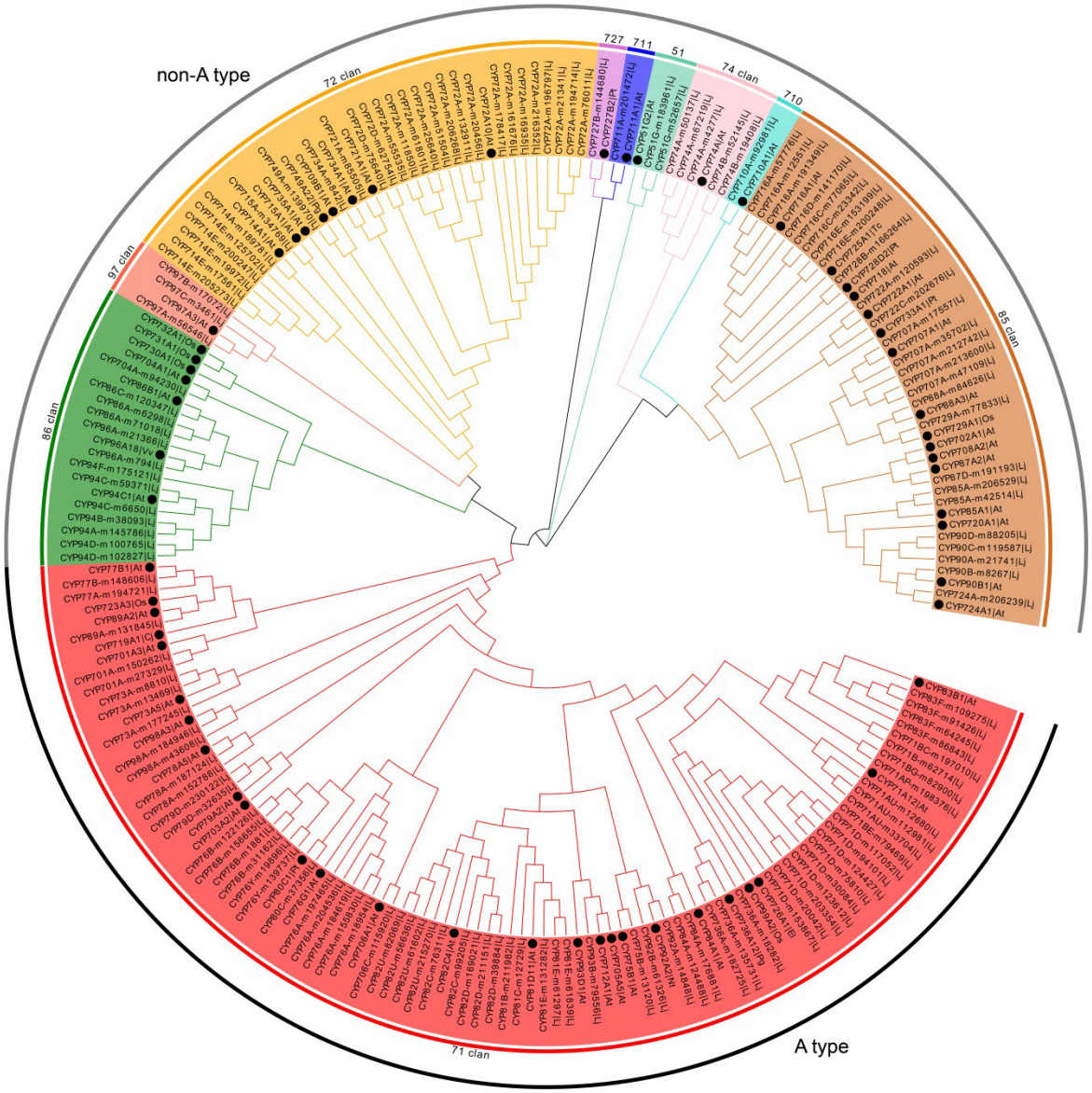

Figure 1 Phylogenetic analysis of predicted CYP450s in *L. japonica* and the representative members of CYP450 families.

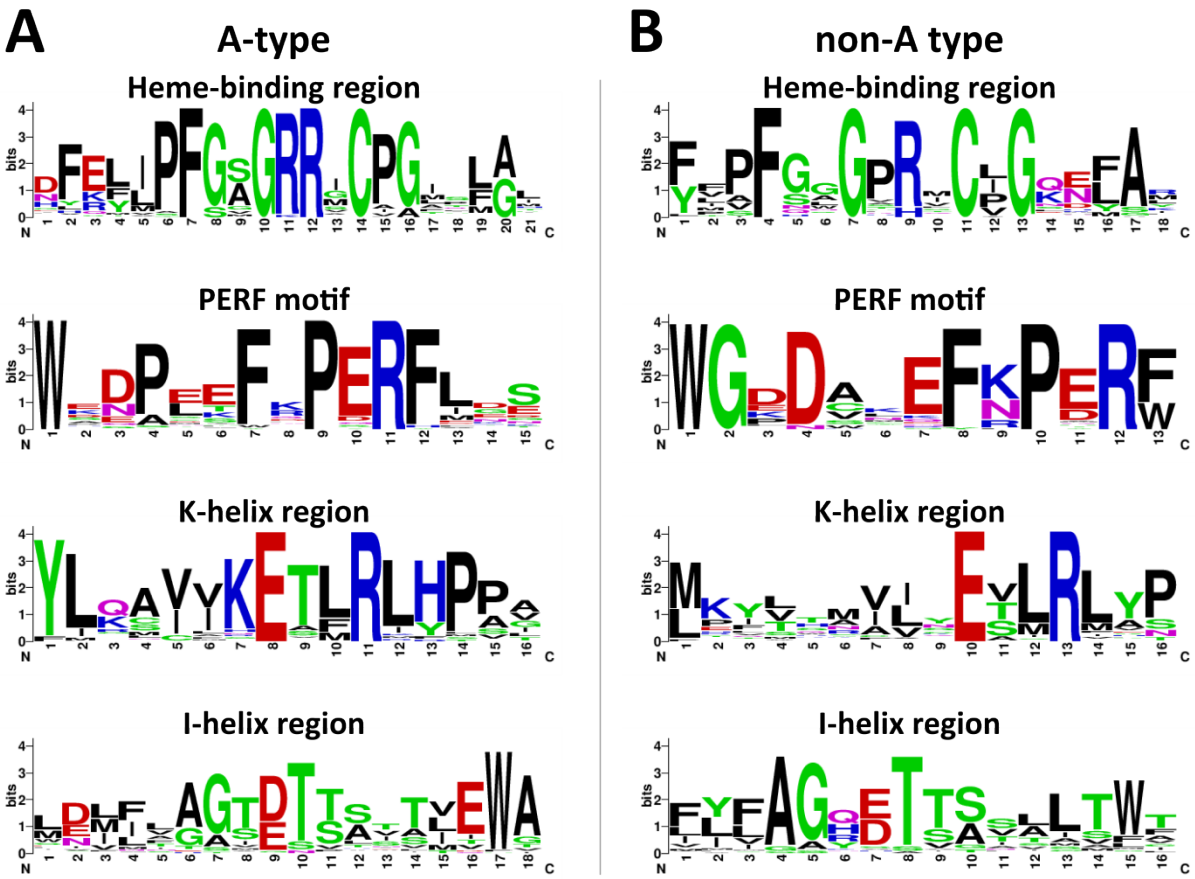

Figure 2 Weblogos of conserved motifs identified in A-type (A) and non-A type (B) CYP450s from *L. japonica*.

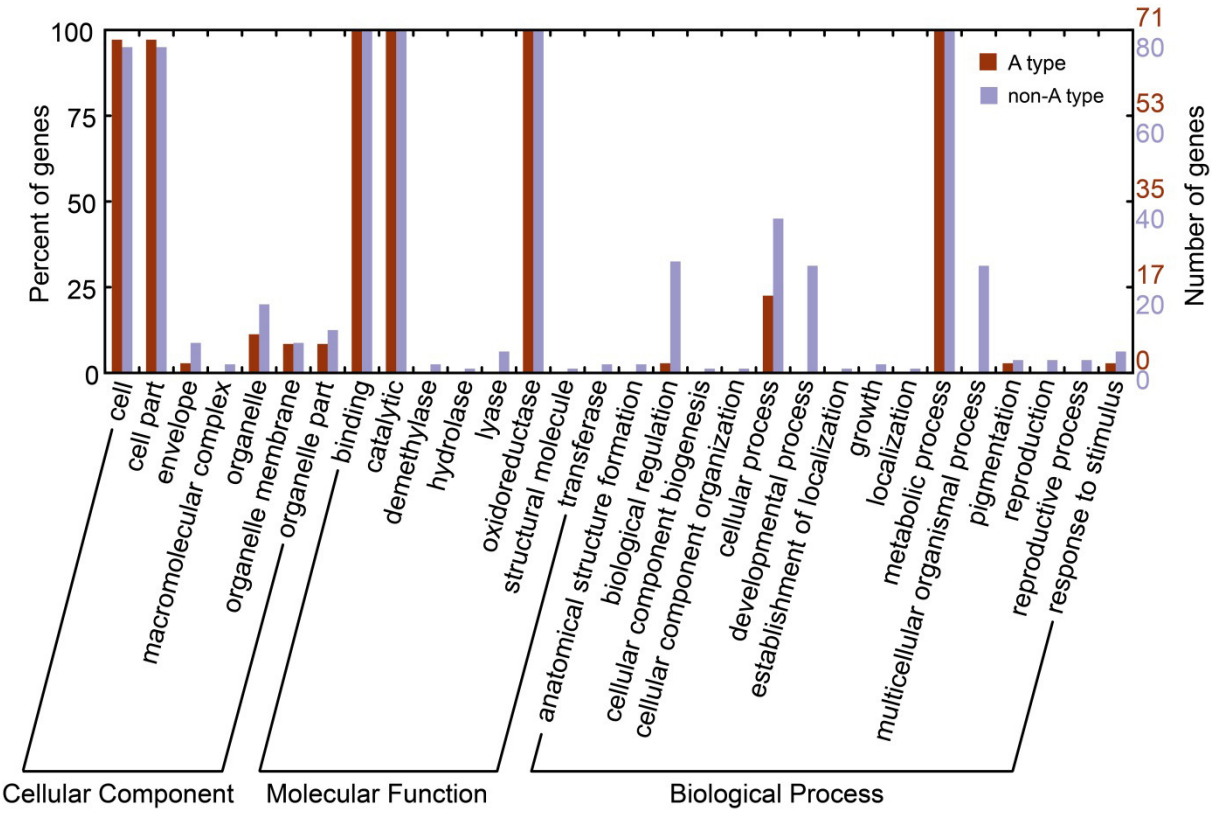

Figure 3 Gene ontology annotation of A-type and non-A type CYP450s in *L. japonica*.

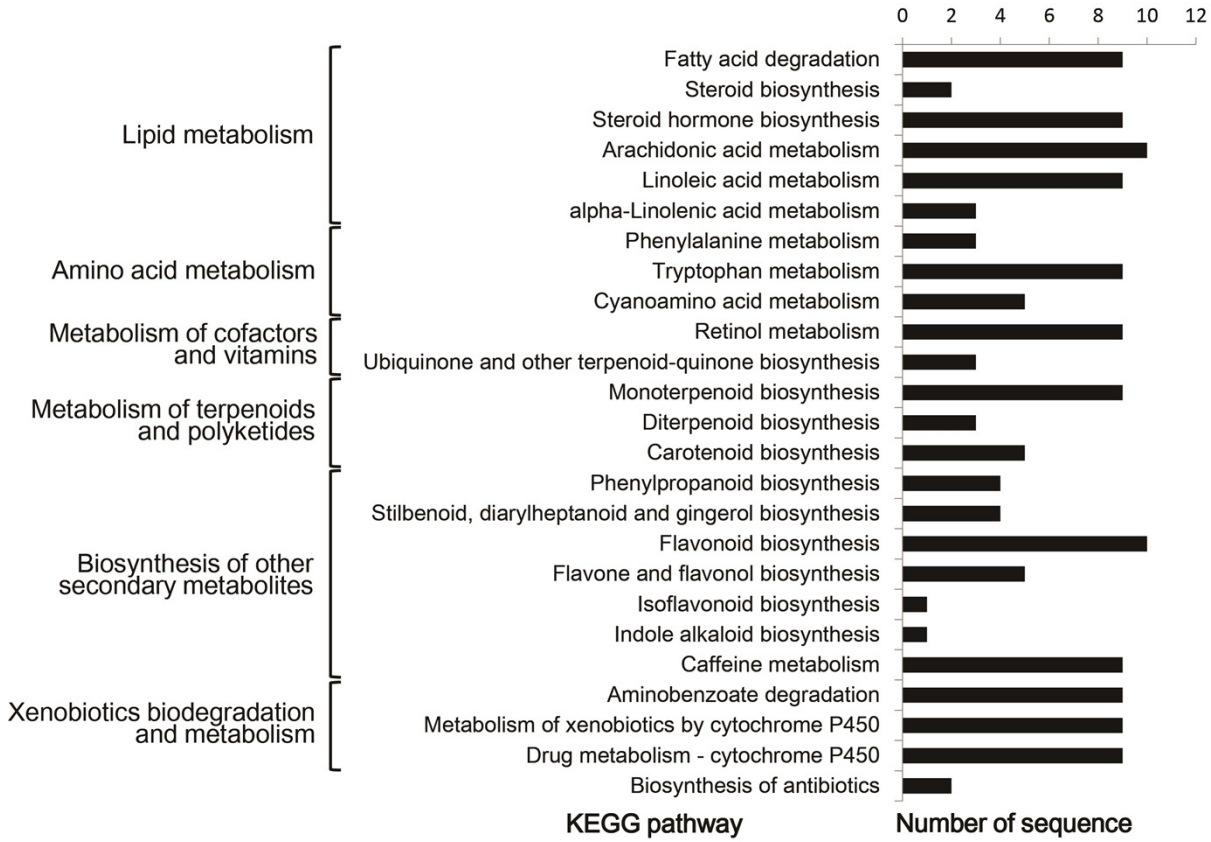

Figure 4 KEGG pathway analysis of predicted CYP450s in *L. japonica*.

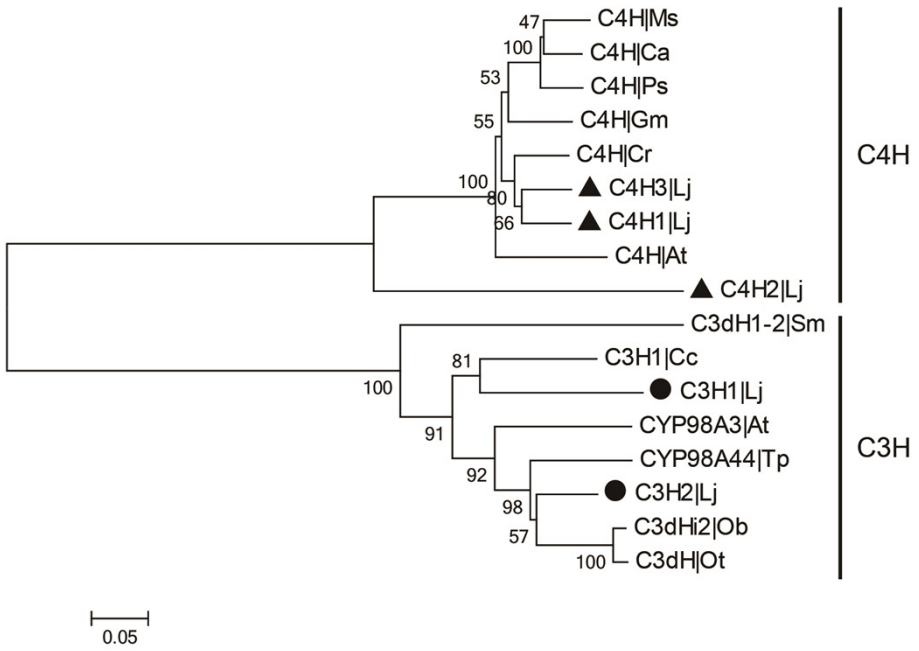

Figure 5 Phylogenetic analysis of C3Hs and C4Hs from *L. japonica* and other plants. LjC3Hs were labeled by black dots and LjC4Hs were labeled by black triangles. Protein sequences were downloaded from UniProt with accession numbers as follows: C3H1|Cc (A4ZKM5), CYP98A3|At (O22203), C3dH1-2|Sm (D8SCG3), C3dHi2|Ob (Q8L5H7), CYP98A44|Tp (C9EGT6), C3dH|Ot (T1NXG3), C4H|At (P92994), C4H|Cr (P48522), C4H|Ps (Q43067), C4H|Ms (P37114), C4H|Ca (O81928), C4H|Gm (Q42797)

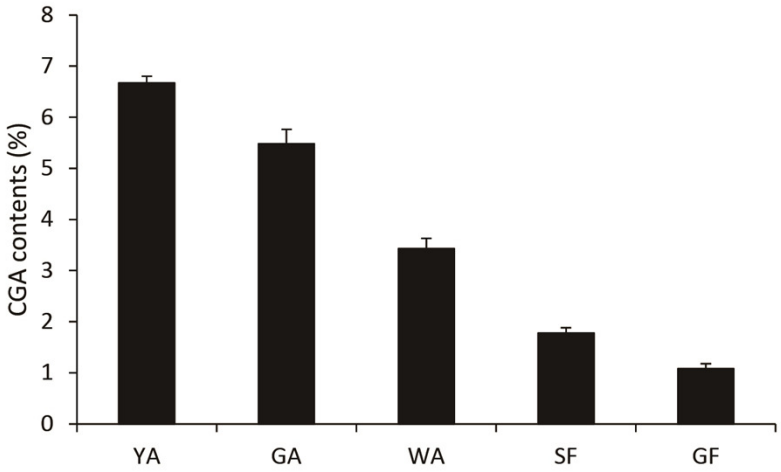

Figure 6 CGA contents of buds and flowers in different developmental stages *L. japonica*. YA-young alabastrum, GA-green alabastrum, WA-white alabastrum, SF-silvery flower, and GF-golden flower.

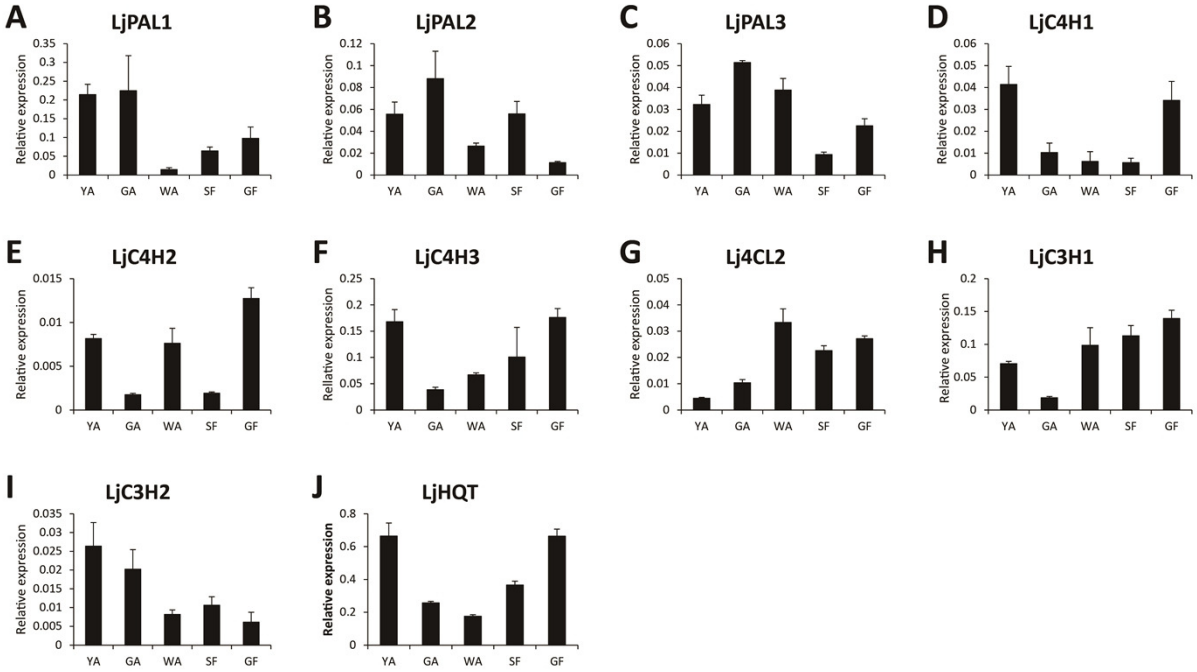

Figure 7 Transcriptional analyses of CGA biosynthetic pathway genes in buds and flowers of *L. japonica* at different developmental stages.

558 **Table 1 List of predicted CYP450s with complete cytochrome P450 domain from *L. japonica*.**

Type	Clan	Family	Subfamily	Gene ID	Type	Clan	Family	Subfamily	Gene ID
non-A	51	CYP51	CYP51G	m183961	non-A	72	CYP72	CYP72A	m61801
non-A	51	CYP51	CYP51G	m52657	non-A	72	CYP72	CYP72A	m25640
A	71	CYP71	CYP71B	m62714	non-A	72	CYP72	CYP72A	m206268
A	71	CYP71	CYP71D	m153867	non-A	72	CYP72	CYP72A	m132911
A	71	CYP71	CYP71D	m20042	non-A	72	CYP72	CYP72A	m20456
A	71	CYP71	CYP71D	m123612	non-A	72	CYP72	CYP72A	m161676
A	71	CYP71	CYP71D	m203354	non-A	72	CYP72	CYP72A	m16935
A	71	CYP71	CYP71D	m30084	non-A	72	CYP72	CYP72A	m216352
A	71	CYP71	CYP71D	m75810	non-A	72	CYP72	CYP72A	m196797
A	71	CYP71	CYP71D	m124427	non-A	72	CYP72	CYP72A	m21341
A	71	CYP71	CYP71D	m94101	non-A	72	CYP72	CYP72A	m194714
A	71	CYP71	CYP71D	m117052	non-A	72	CYP72	CYP72A	m76011
A	71	CYP71	CYP71AP	m198376	non-A	72	CYP72	CYP72A	m178417
A	71	CYP71	CYP71AU	m112981	non-A	72	CYP72	CYP72D	m62754
A	71	CYP71	CYP71AU	m12680	non-A	72	CYP72	CYP72D	m75640
A	71	CYP71	CYP71AU	m33704	non-A	72	CYP714	CYP714A	m189781
A	71	CYP71	CYP71BC	m197010	non-A	72	CYP714	CYP714E	m200247
A	71	CYP71	CYP71BE	m79469	non-A	72	CYP714	CYP714E	m125702
A	71	CYP71	CYP71BG	m82900	non-A	72	CYP714	CYP714E	m205273
A	71	CYP73	CYP73A	m177245	non-A	72	CYP714	CYP714E	m19972
A	71	CYP73	CYP73A	m13469	non-A	72	CYP714	CYP714E	m17561
A	71	CYP73	CYP73A	m8810	non-A	72	CYP715	CYP715A	m34769
A	71	CYP75	CYP75B	m13120	non-A	72	CYP721	CYP721A	m85505
A	71	CYP76	CYP76A	m204536	non-A	72	CYP734	CYP734A	m842
A	71	CYP76	CYP76A	m184619	non-A	72	CYP749	CYP749A	m139970
A	71	CYP76	CYP76A	m18954	non-A	74	CYP74	CYP74A	m4277
A	71	CYP76	CYP76A	m155830	non-A	74	CYP74	CYP74A	m50137
A	71	CYP76	CYP76A	m197465	non-A	74	CYP74	CYP74A	m67219
A	71	CYP76	CYP76B	m31162	non-A	74	CYP74	CYP74B	m52145
A	71	CYP76	CYP76B	m1881	non-A	74	CYP74	CYP74B	m19408
A	71	CYP76	CYP76B	m156655	non-A	85	CYP85	CYP85A	m206529
A	71	CYP76	CYP76B	m122126	non-A	85	CYP85	CYP85A	m42514
A	71	CYP76	CYP76Y	m139737	non-A	85	CYP87	CYP87D	m191193
A	71	CYP76	CYP76Y	m19896	non-A	85	CYP88	CYP88A	m84626
A	71	CYP77	CYP77A	m194721	non-A	85	CYP90	CYP90A	m21741
A	71	CYP77	CYP77B	m148606	non-A	85	CYP90	CYP90B	m8267
A	71	CYP78	CYP78A	m187124	non-A	85	CYP90	CYP90C	m119587

A	71	CYP78	CYP78A	m152788
A	71	CYP79	CYP79D	m32635
A	71	CYP79	CYP79D	m230122
A	71	CYP80	CYP80C	m37356
A	71	CYP81	CYP81B	m211982
A	71	CYP81	CYP81C	m12729
A	71	CYP81	CYP81E	m131282
A	71	CYP81	CYP81E	m61839
A	71	CYP81	CYP81E	m61297
A	71	CYP82	CYP82C	m99205
A	71	CYP82	CYP82C	m76311
A	71	CYP82	CYP82D	m169021
A	71	CYP82	CYP82D	m39884
A	71	CYP82	CYP82D	m211151
A	71	CYP82	CYP82U	m215270
A	71	CYP82	CYP82U	m56636
A	71	CYP82	CYP82U	m82069
A	71	CYP82	CYP82U	m61602
A	71	CYP83	CYP83F	m86843
A	71	CYP83	CYP83F	m64245
A	71	CYP83	CYP83F	m109275
A	71	CYP83	CYP83F	m91426
A	71	CYP84	CYP84A	m176881
A	71	CYP84	CYP84A	m124488
A	71	CYP89	CYP89A	m131845
A	71	CYP92	CYP92A	m14848
A	71	CYP92	CYP92B	m61326
A	71	CYP93	CYP93B	m79556
A	71	CYP98	CYP98A	m184946
A	71	CYP98	CYP98A	m43608
A	71	CYP701	CYP701A	m27329
A	71	CYP701	CYP701A	m150262
A	71	CYP706	CYP706C	m115920
A	71	CYP736	CYP736A	m18282
A	71	CYP736	CYP736A	m135731
A	71	CYP736	CYP736A	m182725
non-A	72	CYP72	CYP72A	m51504
non-A	72	CYP72	CYP72A	m55535
non-A	72	CYP72	CYP72A	m11850
non-A	85	CYP90	CYP90D	m88205
non-A	85	CYP707	CYP707A	m212742
non-A	85	CYP707	CYP707A	m213600
non-A	85	CYP707	CYP707A	m47109
non-A	85	CYP707	CYP707A	m35702
non-A	85	CYP707	CYP707A	m17557
non-A	85	CYP716	CYP716A	m191349
non-A	85	CYP716	CYP716A	m57776
non-A	85	CYP716	CYP716A	m12551
non-A	85	CYP716	CYP716C	m77065
non-A	85	CYP716	CYP716C	m23342
non-A	85	CYP716	CYP716D	m141170
non-A	85	CYP716	CYP716E	m153199
non-A	85	CYP716	CYP716E	m200248
non-A	85	CYP722	CYP722A	m120593
non-A	85	CYP722	CYP722C	m202676
non-A	85	CYP724	CYP724A	m206239
non-A	85	CYP728	CYP728B	m166264
non-A	85	CYP729	CYP729A	m77833
non-A	86	CYP86	CYP86A	m71018
non-A	86	CYP86	CYP86A	m6298
non-A	86	CYP86	CYP86C	m120347
non-A	86	CYP94	CYP94A	m145786
non-A	86	CYP94	CYP94B	m38093
non-A	86	CYP94	CYP94C	m59371
non-A	86	CYP94	CYP94C	m6650
non-A	86	CYP94	CYP94D	m102827
non-A	86	CYP94	CYP94D	m100765
non-A	86	CYP94	CYP94F	m175121
non-A	86	CYP96	CYP96A	m794
non-A	86	CYP96	CYP96A	m21366
non-A	86	CYP704	CYP704A	m94230
non-A	97	CYP97	CYP97A	m56546
non-A	97	CYP97	CYP97B	m17072
non-A	97	CYP97	CYP97C	m3461
non-A	710	CYP710	CYP710A	m92981
non-A	711	CYP711	CYP711A	m201472
non-A	727	CYP727	CYP727B	m144680

559