

# Genome-wide characterization of the CYP450 gene family of *Lagerstroemia indica* and its functional role in anthocyanin biosynthesis

Mingzhu Bai<sup>#</sup>, Zhibin Li<sup>#</sup>, Jiake Huang, Ruimin Jiang, Manting Jin, Kaixiao Sun and Cuihua Gu<sup>\*</sup>

College of Landscape Architecture, Zhejiang A&F University, Hangzhou 311300, China

<sup>#</sup> Authors contributed equally: Mingzhu Bai, Zhibin Li

<sup>\*</sup> Corresponding author, E-mail: [gucuihua@zafu.edu.cn](mailto:gucuihua@zafu.edu.cn)

## Abstract

*Lagerstroemia indica*, a species renowned for its ornamental attributes, occupies a vital position in garden landscaping, and CYP450 constitutes a group of proteins functioning as monooxygenases, which are extensively implicated in plant growth, development, and various other life processes. There are 314 members in total belonging to the CYP450 family, which were identified within the genome of *Lagerstroemia indica*, with their distribution spanning 23 chromosomes. Based on diverse structural domains as well as evolutionary analysis, the CYP450 family is divided into nine subfamilies; segmental duplication serves as the chief amplification mode during the development trajectory of the CYP450 gene family, the CYP75 subfamily genes *LiF3'H* and *LiF3'5'H*, which were found to localize to the cell membrane and promote anthocyanin synthesis. The results obtained from this study are of great significance as they delve deeper into the functional contributions of CYP450 genes by establishing a firm basis and for investigating the molecular processes involved in anthocyanin synthesis in *Lagerstroemia indica*.

**Citation:** Bai M, Li Z, Huang J, Jiang R, Jin M, et al. 2025. Genome-wide characterization of the CYP450 gene family of *Lagerstroemia indica* and its functional role in anthocyanin biosynthesis. *Ornamental Plant Research* 5: e030 <https://doi.org/10.48130/opr-0025-0027>

## Introduction

*Lagerstroemia indica*, a member of the *Lagerstroemia* genus, exhibits the characteristics of a small tree or shrub that sheds its leaves seasonally, which boasts an extended flowering duration, large flower volume, strong resistance, and plays a crucial role in summer ornamental woody flower that is widely used in gardens around the world<sup>[1]</sup>. Flower color holds significant importance as one of the most distinctive and immediately noticeable ornamental attributes among garden plants, and the common flower colors of *L. indica* mainly include red, pink, purple, and white. In *L. indica*, different types of anthocyanins and the proportion of their contents are the most important factors affecting petal color formation<sup>[2]</sup>.

Cytochrome P450 (CYP450) is a heme-binding protein belonging to the class of monooxygenases, which is widely found in different living organisms in nature, including humans, animals, plants, and microorganisms<sup>[3,4]</sup>, and studies have shown that CYP450 is present in almost all prokaryotic and eukaryotic organisms except *Escherichia coli*<sup>[5]</sup>. CYP450 proteins generally contain about 500 amino acids, with an inverted triangular spatial structure encompassing 12  $\alpha$ -helices and four antiparallel  $\beta$ -folds<sup>[6]</sup>, and a conserved core region with three highly conserved residues: the heme-binding domain (FXX-GXRXCXG), the EXXR residue, and the delivery channel residue (PERF)<sup>[7]</sup>. Due to their highly conserved structure, all CYP450s show similar folded structures, but the lack of conservation of CYP450 recognition sites for substrates allows homology between different CYP450s to be as low as 16%<sup>[8]</sup>. Representing the largest-sized family of catalytic proteins found in plants, roots, stems, leaves, flowers, fruits, seeds, and even dormant bulbs all contain CYP450. A majority of these genes are localized within the secretory pathway of the plant, which is typically coupled with the membrane systems of organelles like the endoplasmic reticulum, plastid, mitochondrion, Golgi apparatus, and vesicles<sup>[9]</sup>.

Based on evolutionary links and amino acid sequence similarity, CYP450 can be divided into two groups within this superfamily:

monogenic and polygenic families. The P450s identified so far in plants can be classified into 11 family clusters, of which there are seven monogenic clusters (CYP51/746/74/711/97/710/727) and four multifamily clusters (CYP71/86/85/72)<sup>[10]</sup>. With the development of sequencing technology, many plants such as *Arabidopsis thaliana*, *Oryza sativa*, *Solanum lycopersicum*, *Vitis vinifera*, *Brassica rapa*, *Cajanus cajan*, and many other plants have undergone a series of genome-wide characterizations and analyses to determine their family affiliation and evolution<sup>[11–16]</sup>. For most plants, the CYP450 gene family encompasses a member count ranging from 100 to 300. However, the precise number of CYP450 gene family members exhibits significant variation across different plant species<sup>[17]</sup>.

Flavonoids, carotenoids, and alkaloid pigments constitute the three principal categories of pigments responsible for the coloration of plant flowers; anthocyanins (chromophores and precursors of anthocyanosides) are the main color-presenting substances in the family of flavonoids, the number of hydroxyl moieties located on the B-ring of anthocyanins influences the coloration of anthocyanosides, and the hydroxylation of anthocyanins is mediated by the flavonoid 3',5'-hydroxylase (F3'5'H) and flavonoid 3'-hydroxylase (F3'H), of which F3'5'H belongs to the CYP75A subfamily, and F3'H belongs to the CYP75B subfamily, and the CYP75 subfamily Exerts a pivotal influence on the biosynthesis of anthocyanins<sup>[18]</sup>. In *Petunia hybrida*, overexpression of F3'5'H significantly enhanced the manifestation of a blue floral hue<sup>[19]</sup>. In *Rosa hybrida*, the introduction of F3'5'H successfully achieved a breakthrough in blue flower color<sup>[20]</sup>. In *Chrysanthemum morifolium*, by suppressing F3'H and overexpressing F3'5'H, researchers successfully bred blue *chrysanthemum* varieties<sup>[21]</sup>. In addition, in *Nicotiana tabacum*, the synergistic regulation of F3'H and F3'5'H has also been demonstrated to markedly change flower color, and the shift from red to purple further validates the critical role of these two types of genes in flower color formation<sup>[22]</sup>. In *Solanum lycopersicum*, mutations in F3'H resulted in reduced anthocyanin accumulation, which affected fruit color<sup>[23]</sup>.

The CYP75 subfamily in grapes is involved in flavonoid metabolism, especially in the build-up of anthocyanins in red fruit surfaces<sup>[24]</sup>.

In the context of the present research, bioinformatics techniques were utilized to determine the members of the CYP450 gene family in *L. indica*, systematically analyzed their distribution on chromosomes, structural domain features and gene duplication events, classified them into different taxa, and established their phylogenetic relationships, on the basis of which, this research endeavored to investigate the pathway leading to anthocyanin production in *L. indica*, as well as focus on analyzing the CYP75 family of F3'H and F3'5'H expression and function of the CYP75 family. Such endeavors will contribute to a more profound comprehension of the LiCYP450 gene family and establish a foundational framework for the genetic enhancement of flower color in *L. indica*.

## Materials and methods

### Identification, characterization, and evaluation of the physicochemical features of LiCYP450 family proteins

In the present research, CYP450 family sequences of *Arabidopsis* (*Arabidopsis thaliana*) were retrieved from the TAIR database ([www.arabidopsis.org](http://www.arabidopsis.org)). Subsequently, BLAST searches were conducted on *L. indica* genome data that had been downloaded from the National Center for Genome Sciences Data (<https://ngdc.cncb.ac.cn/>). A BLAST search was carried out<sup>[25]</sup>, after which the Hidden Markov Model (HMM) file PF00067 was downloaded. This corresponds to the conserved structural domains of CYP450 from the Pfam website (<http://pfam.xfam.org/>), and the *L. indica* genome was searched using the built-in HMMER 3.0 program<sup>[25,26]</sup>. Ultimately, a summary of the two methods was presented. Subsequently, the candidate members acquired were subjected to further validation through the NCBI conserved structural domain database ([www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi](http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi)).

Utilizing the position annotation data of LiCYP450 genes derived from the Gff file of *L. indica* genome annotation, an investigation was carried out to ascertain the distribution pattern of these genes across the 24 chromosomes of *L. indica*. Following this, the genes were renamed based on their respective chromosomal locations. With the aim of scrutinizing the physical and chemical attributes of the proteins, all LiCYP450 sequences were input into EXPASY (<https://web.expasy.org/protparam/>) to calculate various parameters, encompassing the count of amino acids, molecular weight (MW), and the theoretical isoelectric point (pI), instability index (II), aliphatic index (AI), and grand average of hydropathicity (GRAVY) of the proteins<sup>[27]</sup>. The cellular compartmentalization of the LiCYP450 protein was forecasted by employing the WoLF PSORT website ([www.genscript.com/wolf-psort.html](http://www.genscript.com/wolf-psort.html))<sup>[28]</sup>.

### Gene structure study analysis and phylogenetic of the LiCYP450 gene

Multi-sequence juxtaposition of CYP450 109 protein chains from *L. indica* and *Arabidopsis thaliana* was carried out with the aid of the MEGA 7 software and 1,000 Bootstrap statistical tests were performed by the Neighbor-Joining method<sup>[29]</sup>, and the results were imported into the Evolview online site ([www.evolgenius.info/evolview](http://www.evolgenius.info/evolview)) to decorate the phylogenetic tree<sup>[30]</sup>. By employing MEME (<https://meme-suite.org/meme/tools/meme>), conserved motifs within the LiCYP450 protein sequence were pinpointed, with the motif count set at ten<sup>[31]</sup>. The promoter region (including the initiation codon) of the LiCYP450 gene was retrieved, spanning 2,000 bp upstream, and was subjected to prediction of cis-regulatory elements (CAREs) using the PlantCARE database (<https://bioinformatics.psb.ugent.be/webtools/plantcare/html/>) and a search for the

CARE tool was conducted. Subsequently, these predicted CAREs were visualized with the aid of TBtools software<sup>[32,33]</sup>.

### Gene duplication and covariance analysis of the LiCYP450 gene

To analyze gene duplication events in the LiCYP450 gene, DupGen\_finder was used<sup>[34]</sup>. By employing the default settings, this tool effectively distinguishes among various types of gene duplications, encompassing whole genome duplication (WGD), tandem duplication (TD), transposed duplication (TRD), proximal duplication (PD), and dispersed duplication (DSD). The visualization of duplicated RsCYP450 gene pairs was performed with the assistance of TBtools software. The identity of CYP450 genes between *L. indica* and other plants was detected using MCScanX<sup>[35]</sup>. Utilizing TBtools software, the outcomes of the synteny analysis were visualized, thereby offering an in-depth comprehension of the conservation patterns and evolutionary connections among CYP450 genes across diverse plant species.

### LiF3'H and LiF3'5'H multiple sequence comparison

The F3'H and F3'5'H amino acid sequences from different species were retrieved from the GenBank database on the NCBI platform (Supplementary Table S1). Subsequently, multiple comparisons of these amino acid sequences were conducted against those of F3'H and F3'5'H in the CYP75 subfamily via the mafft website (<https://mafft.cbrc.jp/alignment/server/>)<sup>[36]</sup>. Visualization was performed with Jalview<sup>[37]</sup>.

### Expression pattern analysis of plant materials with LiF3'H and LiF3'5'H

In this research, the plant materials utilized include white (Natchez, WH), pink (Hedge Red Fields, PK), red (Red Flower Guifei, RD), and purple (Purple Flowering Guelder, PP) samples of four flower color varieties at three periods of flower development and from different tissues. The leaf material is divided into young leaves and mature leaves according to the developmental stage: the standard for young leaves is that the leaf length reaches 50%–70% of the typical leaf size of the variety, it is not fully unfolded, and the leaf color is reddish. The standard for mature leaves is that the leaves are fully unfolded, the leaf length reaches more than 80% of the typical size of the variety, the leaf color is dark green, and the texture is leathery. In the present study, the FastPure Plant Total RNA Isolation Kit (Vazyme, Nanjing, China) was employed to extract total RNA from the leaves of both the control and treated groups. Following that, the RNA that had been extracted was converted into cDNA using the PrimeScript™ RT Master Mix, a reverse transcription kit manufactured by TAKARA. The reverse-transcribed cDNA was then preserved at –20 °C within the laboratory. Subsequently, the cDNA was stored in a –20 °C refrigerator and subsequently employed for gene cloning as well as expression analysis. NCBI website was utilized to design gene-specific primers (Supplementary Table S2), with *LiUBC+LiEF1 $\alpha$ -1* serving as the internal reference gene<sup>[38]</sup>. Quantitative PCR was performed using SYBR qPCR Master Mix, and each treatment contained three technical replicates. Ct values of CYP450 genes were evaluated with the 2<sup>– $\Delta\Delta$ Ct</sup> approach.

### LiF3'H and LiF3'5'H gene cloning and subcellular localization

Primers specifically for *LiF3'H* and *LiF3'5'H* were designed using Snapgene (Supplementary Table S2) based on the known genome sequences of *L. indica*. They were synthesized by Zhejiang Youkang Biotechnology Co. PCR cloning was performed using a mixture of cDNAs from petals of four varieties of *L. indica* from three periods as templates. The target genes were amplified using 2 $\times$  Phanta Max Master Mix high-fidelity enzyme (Vazyme, Nanjing, China). The



amplified PCR products were cut and recovered using the FastPure Gel DNA Extraction Mini Kit (Vazyme, Nanjing, China). *LiF3'H* and *LiF3'5'H* were constructed into the overexpression vector pORE\_R4 using seamless DNA cloning, and the recombinant plasmids were constructed according to the method of ClonExpressII One Step Cloning Kit Homologous Recombination Kit (Vazyme, Nanjing, China).

The pORE\_R4-*LiF3'H* and pORE\_R4-*LiF3'5'H* recombinant vectors were introduced into *Agrobacterium tumefaciens* GV3101 (Vazyme, Shanghai, China). Subsequently, *Agrobacterium tumefaciens* with the correct bacteriological test was selected for amplification, and the tobacco of this family grown for about 28 d was selected for the expansion, and the tobacco's back was injected with the bacteriological solution, and the tobacco was cultured in darkness for 1 d, and then cultured under normal illumination for 36 to 48 h. The fluorescence was observed by using an orthogonal fluorescence microscope (ZEISS, Germany) to observe the fluorescence signal.

### Overexpression of *LiF3'H* and *LiF3'5'H* genes in tobacco

The pORE\_R4-*LiF3'H* and pORE\_R4-*LiF3'5'H* recombinant plasmids were transfected into tobacco by *Agrobacterium*-mediated tobacco genetic transformation system using the 'leaf disc transformation method'<sup>[39]</sup>, and the positive seedlings were identified by amplifying the DNA extracted from the different strains of tobacco using the CTAB method. To observe the growth and development process, the corolla was collected at the full bloom stage and

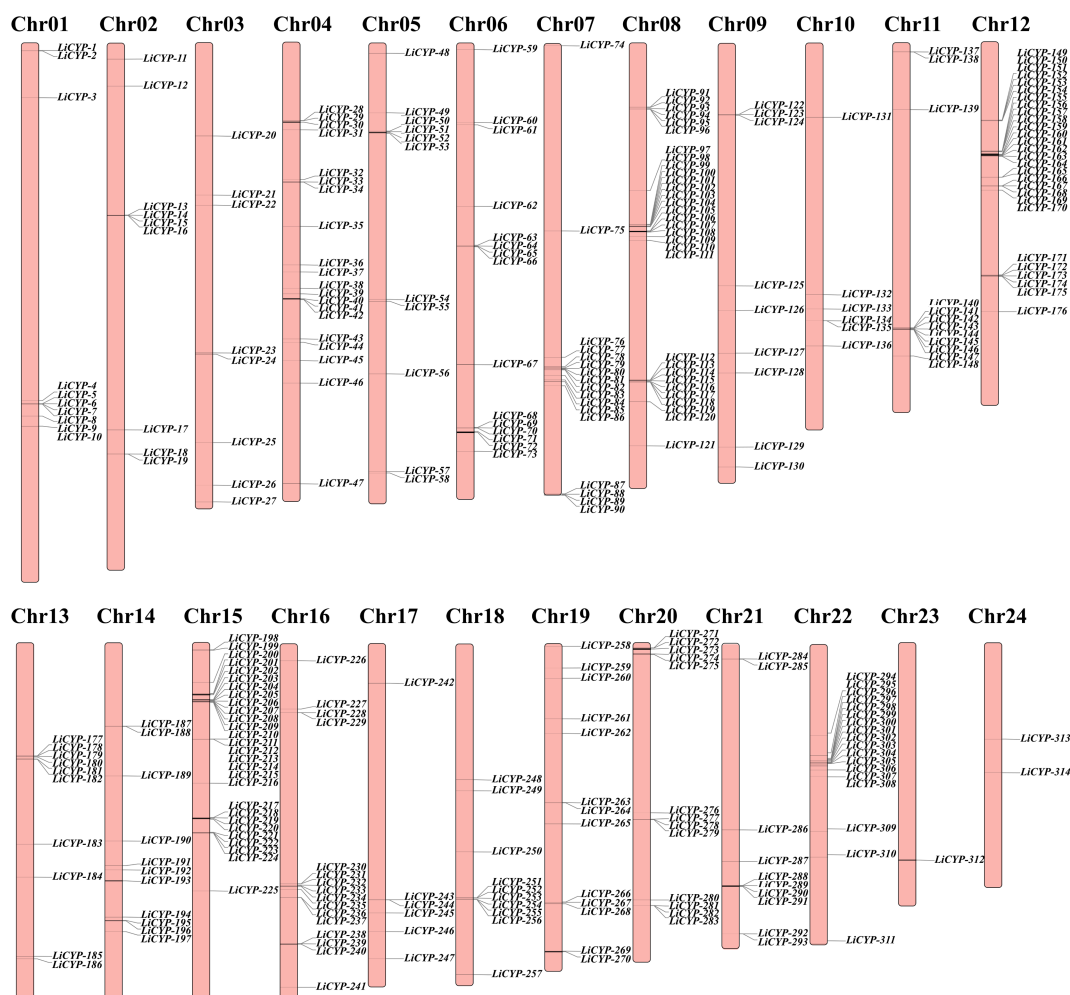
extracted using 1% hydrochloric acid methanol solution (1 mL hydrochloric acid: 99 mL methanol). Three biological replicates were selected for each strain, and a UV spectrophotometer was employed to measure the absorbance at 530 and 657 nm, and the anthocyanin content was obtained by the formula:  $TA = (A_{530} - 0.25 \times A_{657})/W$ , with W being the weight of the tobacco corolla, which served as the foundation for assessing the anthocyanin levels. RNA was extracted from tobacco corolla with an RNA extraction kit, followed by qRT-PCR.

## Results

### Analysis of LiCYP450 family members and their protein properties

In the present investigation, a comprehensive total of 314 LiCYP450 members were determined by employing BLAST and HMMER search approaches, with subsequent validation of their structural domains. The chromosomal localization analysis revealed the 314 CYP450 genes were distributed unevenly yet widely over 23 chromosomes (as depicted in Fig. 1). It is noteworthy that chromosome 8 contained the largest quantity of genes, totaling 31, whereas chromosome 23 had the lowest, with only one gene. These genes were named *LiCYP-1* to *LiCYP-314* on the basis of their chromosomal placement.

Analysis of the lysis index and hydrophilicity index of LiCYP450 members, as determined by ExPASy (Supplementary Table S3),



**Fig. 1** Chromosome location and distribution of LiCYP450 family members.

showed that the amino acid sequence lengths within the LiCYP450 family varied from 62 amino acids (LiCYP-96) to 649 amino acids (LiCYP-33). The mean molecular weight (MW) ranged between 7243.54 kDa (LiCYP-96) and 72611.24 kDa (LiCYP-33). The theoretical pI ranged between 4.67 (LiCYP-95) and 11.3 (LiCYP-281). The instability index showed that 70 LiCYP450 proteins had an instability index of less than 40, suggesting that they were steady *in vitro*, and the remaining 244 LiCYP450 proteins were unsteady *in vitro*. The aliphatic index also ranged from 51.88 (LiCYP-259) to 108.71 (LiCYP-5), indicating a wide variation in thermal stability. There were 284 LiCYP450 proteins with a grand average of hydropathicity, indicating that all of them had different degrees of hydrophilicity.

### LiCYP450 gene structure analysis and phylogenetic tree construction

To delve deeper into the presence and localization of conserved protein structural domains, the current study employed the online MEME program to screen for conserved motifs within these proteins. Overall, ten conserved motifs were recognized and named as motifs 1 through 10 (Supplementary Fig. S1a), and the LiCYP450 proteins were analyzed using the CDD online website and found to contain conserved structural domains such as cytochrome\_P450, MTHFR, and TAXI\_N (Supplementary Fig. S1b). CYP450 structural domains consist of different motif combinations.

An unrooted evolutionary tree comprising 272 *A. thaliana* CYP450 proteins and 314 LiCYP450 proteins was built by employing the NJ (neighbor-joining) algorithm (Fig. 2) and utilized the family categorization of Arabidopsis CYP450 proteins as a reference: CYP450 with amino acid sequence similarity higher than 55% are classified as the identical family; CYP450 with amino acid sequence similarity between 40% and 55% belong to different subfamilies within the same family; If the affinity of amino acid sequences is less than 40%, then these two proteins belong to different families. Which classified the 314 proteins into nine family clusters; the A-type CYP450 protein contains all ten motifs, while the rest are non-A-type CYP450 proteins, including five single families (CYP51/74/97/710/711) and four multi families (CYP71/72/85/86). A-types (CYP71) and non-A-types (CYP51/710/711/72/74/85/86/97) groups, with the A-type and non-A-type family clusters encompassing 209 and 105 LiCYP450s, respectively. The largest CYP71 family cluster contained 209 members, whereas the smallest CYP74 and CYP711 family clusters were found to contain only one member.

### Analysis of cis-regulatory elements of LiCYP450 promoter

LiCYP450's distribution and function were examined and visualized using a hypothetical promoter, a 2,000 bp tract upstream of the enzyme, and a total of 50 elements were identified, which can be classified into four main categories: abiotic elements related to drought stress and cold stress response; development-related, specifically expressed in various organ tissues; hormone-associated, linked to plant hormones such as MeJA and abscisic acid; and light-responsive components (Fig. 3).

In the first category, namely abiotic elements, five cis-acting elements were identified. These encompass the ARE motif, which is associated with anaerobic induction; the LTR motif, which contributes to the response to cold environments; the GC motif, which is involved in hypoxia-specific induction; the MBS motif, which contributes to drought induction; and the TC-rich repeat sequences, which are linked to defense and stress responses. In the second category, which pertains to developmental processes, in total, twelve cis-acting components were found. These encompass the AT-rich element as well as the recognition site for the AT-rich DNA binding protein (ATBP-1); CAT-box, which is associated with the

expression of meristematic tissues; circadian, which is involved in circadian control; including the GCN4 motif, several cis-acting elements were identified, which play a role in endosperm expression; HD Zip1, involved in the differentiation of perforated chloroplasts; MYB binding site I (MBSI) in the flavonoid biosynthesis pathway; the MSA-like element is involved in the regulation of the cell cycle; the O2 site, responsible for regulating the metabolism of alcohol-soluble proteins in maize; and the RY element, related to seed. In addition to the previously mentioned elements, it was the growth hormone regulatory element (TGA-box) and the trauma response element (WUN motif) that were also identified. The existence of these motifs implies that CYP450 is extensively implicated in diverse physiological activities of *L. indica*, including its growth and development.

A comprehensive analysis identified nine distinct cis-acting elements within the third category, which is associated with hormones. These elements encompass the salicylic acid response element (TCA element), the abscisic acid response element (ABRE), the growth hormone regulatory elements (AuxRR core and TGA element), and the CGTCA motif and TGACG motif. The existence of these motifs strongly indicates that CYP450 is extensively implicated in a wide array of physiological processes in *L. indica*, particularly in its stress response mechanisms. Upon comprehensive examination, it was revealed that the fourth category comprises a total of 23 distinct cis-acting elements, including those responsive to light. Among these elements are the 3-AF1 binding site, the AAAC-motif, ACE, and the AE box.

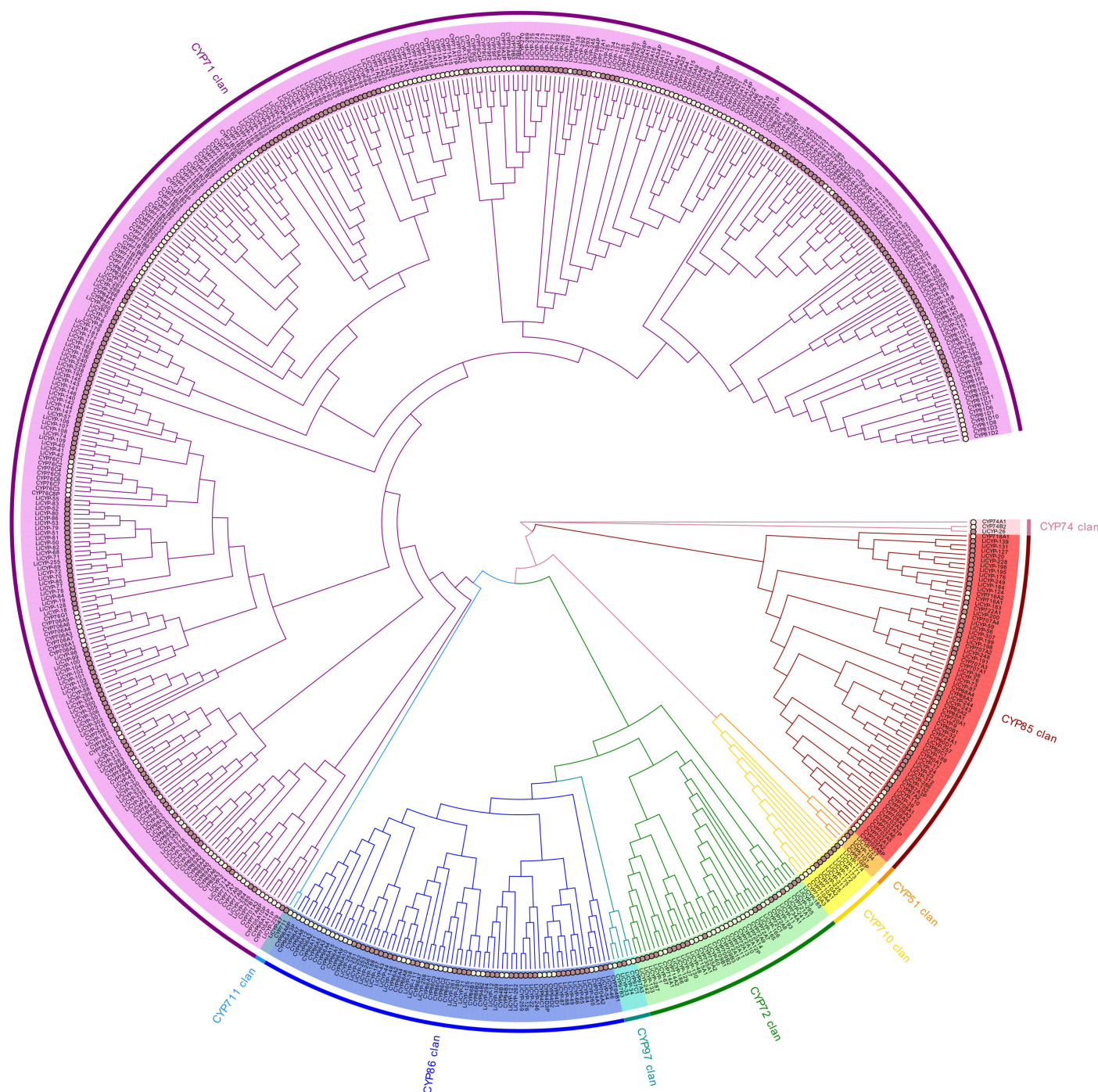
### Gene duplication and covariance analysis

Gene duplication phenomena are frequently observed across various species, and they have the potential to give rise to novel functional genes, thereby acting as a driving force behind species evolution. Through gene duplication event analysis, a total of 17 tandemly duplicated genes were identified, of which 14 genes were divided into seven pairs in groups of two. In addition, 67 segmental duplicate gene pairs were identified (Fig. 4a). The findings imply that a subset of LiCYP450 genes emerged as a result of gene duplication incidents. Additionally, there is a reasonable likelihood that segmental duplications made a significant impact on the evolutionary course of LiCYP450 genes.

Furthermore, the percentage of CYP450 genes that *L. indica* shares with different species could potentially serve as an indicator of the evolutionary connections among CYP450 gene families in *L. indica*, dicotyledons, and monocotyledons. Consequently, within the scope of this research, comparative purity maps of CYP450 genes in *L. indica* were generated in relation to one dicotyledon (*A. thaliana*) and one monocotyledon (*O. sativa*) (Fig. 4b). There were a total of 89 direct homologous pairs between *L. indica* and *A. thaliana* genes (including one *L. indica* gene corresponding to multiple *A. thaliana* genes). There were 20 direct homologous pairs between *L. indica* and *O. sativa* genes (including one *L. indica* gene corresponding to multiple *O. sativa* genes). These results indicate that there are more homologous genes in *L. indica* and dicotyledonous plants than in monocotyledonous plants.

### Multiple sequence comparison of LiF3'H and LiF3'S'H

LiCYP-216 and LiCYP-197 of the CYP75 subfamily were renamed LiF3'H and LiF3'S'H, and their amino acid sequences were analyzed against those of F3'H and F3'S'H with known functions among other botanical species (Fig. 5). The results revealed that LiF3'H and LiF3'S'H have typical CYP450 conserved structural domains, in which the proline-rich motif is essential for P450 protein activity, the A10GTDT515 motif is structurally related to the oxygen-binding pocket, and the ExxR serves as a fold-locking motif to lock the



**Fig. 2** Phylogenetic relationships between *L. indica* and *A. thaliana* CYP450 protein.

cysteine at the proper position and ensure the ferrous hemoglobin-protein Stable binding of heme, the heme conserved domain is very important for binding ferrous heme. The presence of these conserved sites on the amino acid sequences of LiF3'H and LiF3'5'H suggests that the protein possesses the typical hydroxylation functions of LiF3'H and LiF3'5'H.

### Expression pattern analysis of *LiF3'H* and *LiF3'5'H*

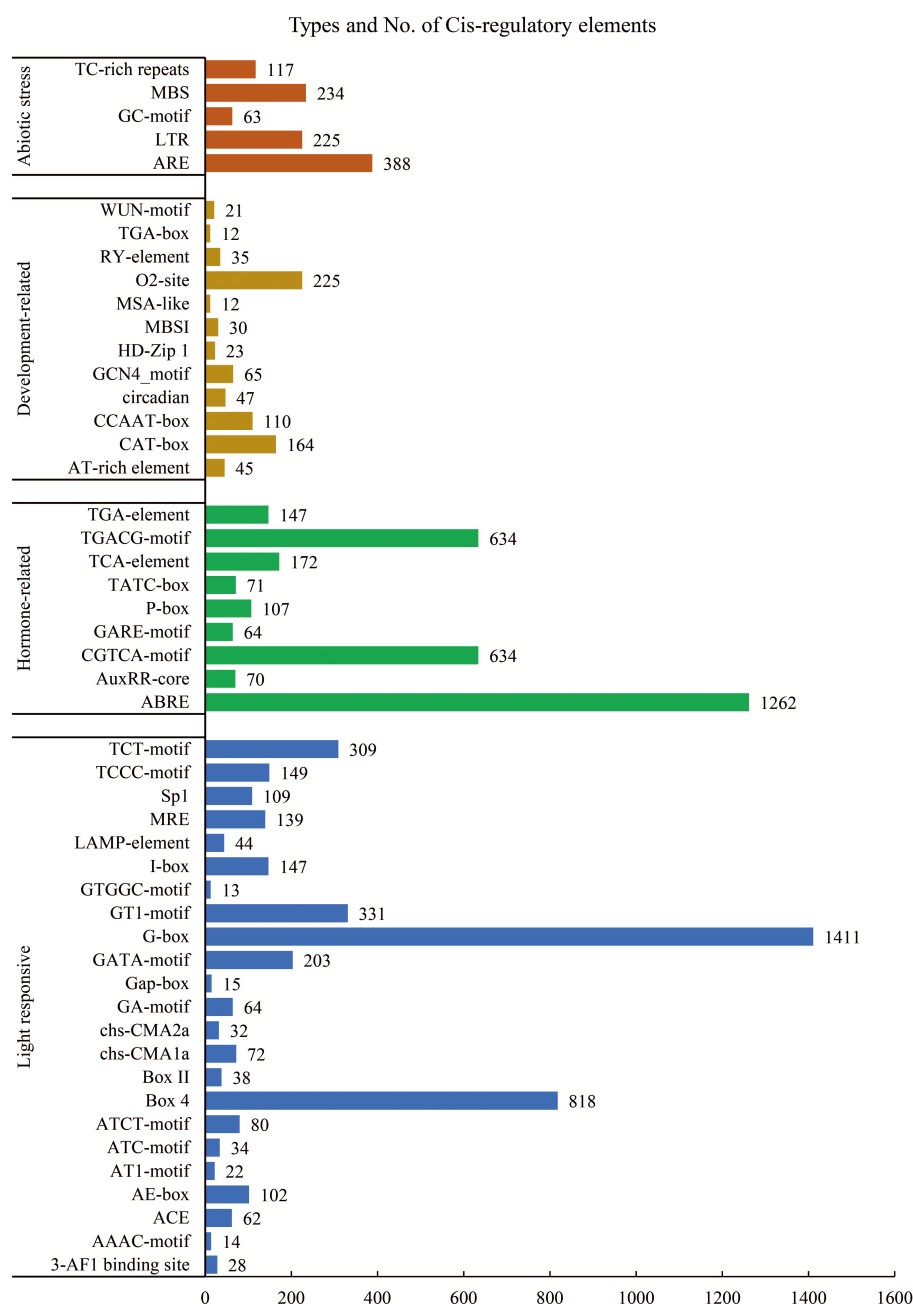
Gene expression profiles of the genes *LiF3'H* and *LiF3'5'H* across distinct tissues were evaluated using qRT-PCR. The outcomes demonstrated that both genes exhibited the peak expression in flowers, with new leaves showing the subsequent highest expression, and these expression levels were markedly elevated compared to those

in other tissues (as illustrated in Fig. 6a & b). Expression analysis of the genes *LiF3'H* and *LiF3'5'H* in different flower colors at different times was performed by qRT-PCR. The findings showed the expression magnitude of *LiF3'H* in PK at the S2 stage exhibited a notably higher value in comparison to other cultivars (Fig. 6c). *LiF3'5'H* showed a decreasing trend in PK, RD, and PP, and the expression reached the highest in PP at the S1 period, and exhibited a considerably higher value in PP and PK than in WH and RD (Fig. 6d).

### Cloning and subcellular localization of *LiF3'H* and *LiF3'5'H* genes

The CDS sequences of *LiF3'H* and *LiF3'5'H* were successfully obtained by using *L. indica* petal cDNA as a template and designing





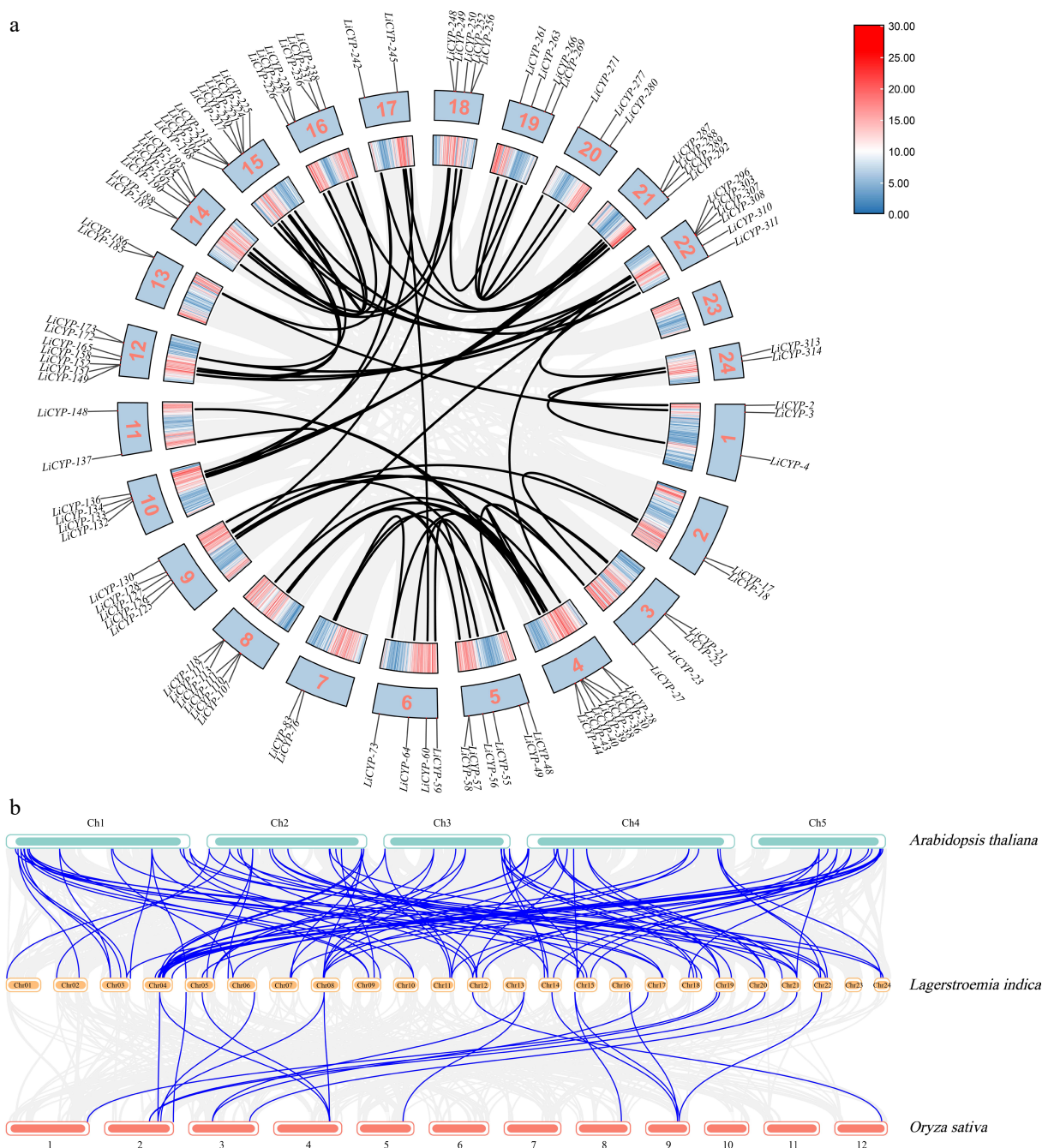
**Fig. 3** Cis-acting elements identified in the promoter of the LiCYP450.

specific primers for amplification based on the genomic sequence. The genes *LiF3'H* and *LiF3'5'H* were constructed into the pORE\_R4 vector containing a GFP tag and transformed into *Agrobacterium* to be injected into the lower epidermal cells of Ben's tobacco. The results showed that pORE\_R4 showed green fluorescent signals throughout the cell, involving the nucleus, cell membrane, and cytoplasm, but the genes *LiF3'H* and *LiF3'5'H* were found only on the cell membrane (Fig. 7).

### Overexpression of *LiF3'H* and *LiF3'5'H* genes in tobacco

To explore the biological functions of *LiF3'H* and *LiF3'5'H*, *LiF3'H* and *LiF3'5'H* were heterologously transformed into tobacco, and a total of two empty vector transgene-positive seedlings, nine *LiF3'H* and nine *LiF3'5'H* transgene-positive seedlings were obtained (Supplementary Fig. S2). As the tobacco inflorescences continued to

grow, the extent of anthocyanin glycoside accumulation exhibited variability across different transgenic lines. Notably, a substantial accumulation of anthocyanin glycosides was observed in the corollas of the *LiF3'H* and *LiF3'5'H* transgenic lines when compared to the wild-type and empty vector transgenic tobacco lines, as depicted in Fig. 8a. The expression profiles of *LiF3'H* and *LiF3'5'H* in the corollas of the previously mentioned strains were analyzed by utilizing qRT-PCR. The results demonstrated that the target genes displayed significant expression levels in all three transgenic strains, with a statistical significance of  $p < 0.05$  (Fig. 8b & c), indicating that *LiF3'H* and *LiF3'5'H* were overexpressed in tobacco. In order to detect the anthocyanin content in tobacco corolla, extraction was carried out using 1% hydrochloric acid methanol solution and assayed using a UV spectrophotometer, which showed that anthocyanins were significantly ( $p < 0.05$ ) higher in *LiF3'H* and *LiF3'5'H* transgenic strains relative to the WT and null control (Fig. 8d).



**Fig. 4** (a) Segmental duplications of LiCYP450. (b) Comparative physical mapping displaying the orthologous relationships of CYP450 genes among *L. indica*, *A. thaliana*, and *O. sativa*.

## Discussion

In the realm of plant secondary metabolism, cytochrome P450 enzymes serve as crucial participants, being direct participants in the processes that produce flavonoids and anthocyanins, which are important contributors to pigment formation and stress adaptation. *L. indica* exhibits remarkable flower color diversity, making it a highly valuable model for unraveling the genetic factors responsible for anthocyanin regulation.

In this study, 314 members of the CYP450 gene family in *L. indica* were identified using BLAST and HMMER search methods combined with structural domain confirmation. These genes exhibited a widespread distribution across 23 chromosomes, demonstrating a non-uniform distribution pattern. Specifically, chromosome 8

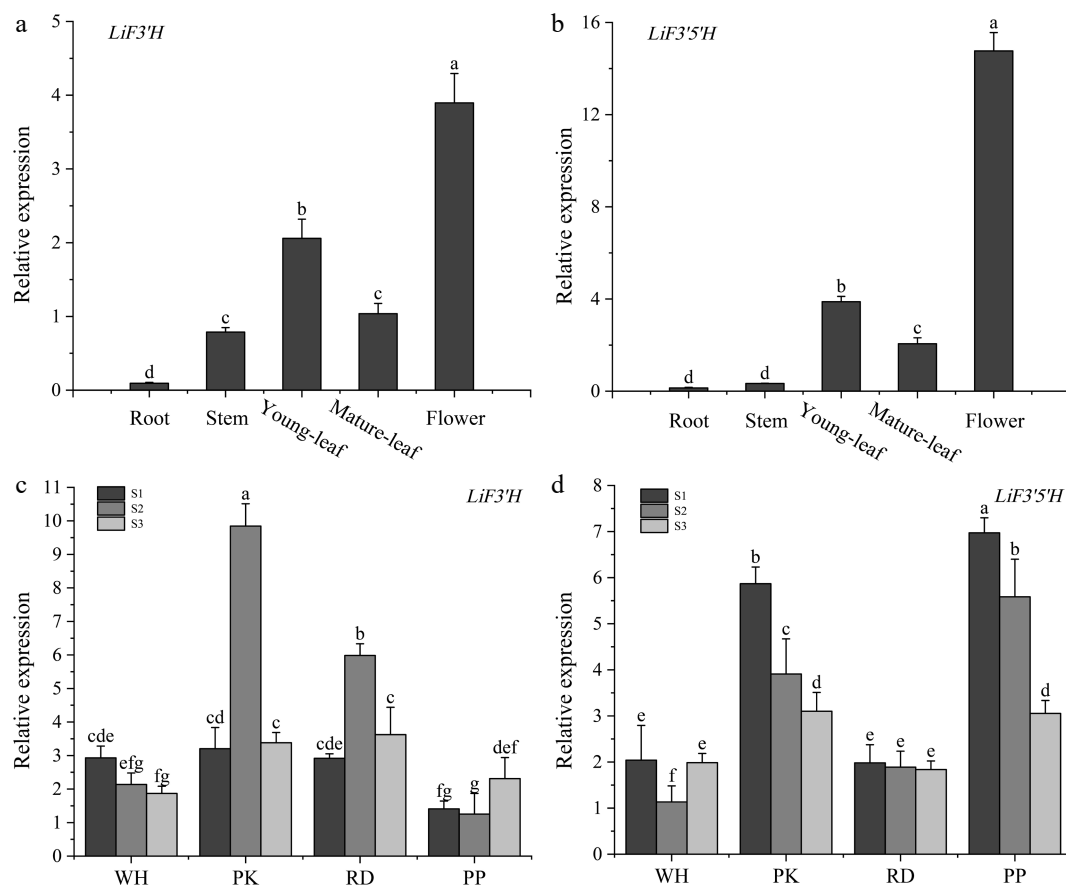
harbored the greatest quantity of genes, while chromosome 23 had the fewest. This disparity in distribution is likely associated with the functional heterogeneity of the genes and evolutionary constraints<sup>[40]</sup>. Upon analyzing various properties of the LiCYP450 family, comprising the count of amino acids, molecular weight, the theoretical pI, instability index, aliphatic index, and grand average of hydropathicity, significant disparities were observed in these proteins, which may be associated with their functional heterogeneity within the cellular environment<sup>[8]</sup>.

The gene structure analysis of the P450 gene of *L. indica* showed that members of the same family share certain similarities, and through the MEME program, this study identified ten conserved motifs and confirmed the presence of conserved structural domains such as cytochrome\_P450 in these proteins through the CDD online



**Fig. 5** (a) Multiple sequence alignment of F3'Hs from various plant species. (b) Multiple sequence alignment of F3'5Hs from various plant species. Boxed in red are the motifs characteristic of P450, encompassing the hinge region (I), the oxygen-binding pocket (II), the E - X - X - R motif (III), and the heme-binding domain (IV). F3'Hs and F3'5Hs proteins originating from different species include AtF3'H (*Arabidopsis thaliana*), CcYP1 (*Crocasmia × crocosmiiflora*), NtF3'H (*Nicotiana tabacum*), OsF3'H (*Oryza sativa*), PhF3'H (*Petunia × hybrida*), AtF3'5H (*Arabidopsis thaliana*), CrF3'5H (*Catharanthus roseus*), EgF3'5H (*Eustoma grandiflorum*), CmF3'5H (*Campanula medium*), SvF3'5H (*Solanum virginianum*).

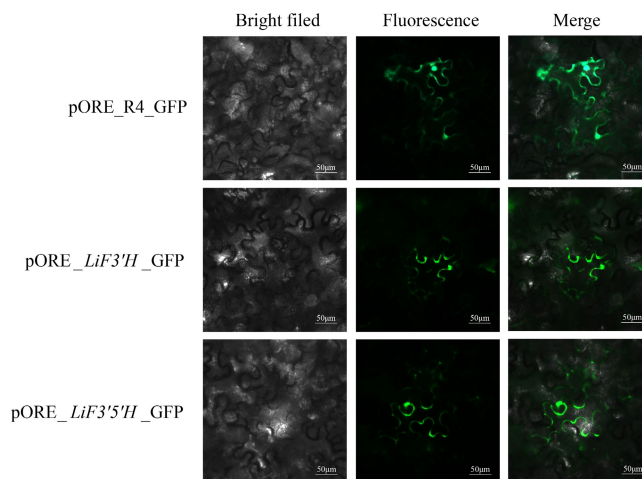




**Fig. 6** (a) The expression level of *LiF3'H* in different tissues. (b) The expression level of *LiF3'5'H* in different tissues. (c) The expression level of *LiF3'H* in different flowering stages. (d) The expression level of *LiF3'H* in different flowering stages. Values represent the mean  $\pm$  standard error (N = 3). Different letters denote significant differences in values were conducted by the one-way ANOVA Duncan test supplement ( $p < 0.05$ ).

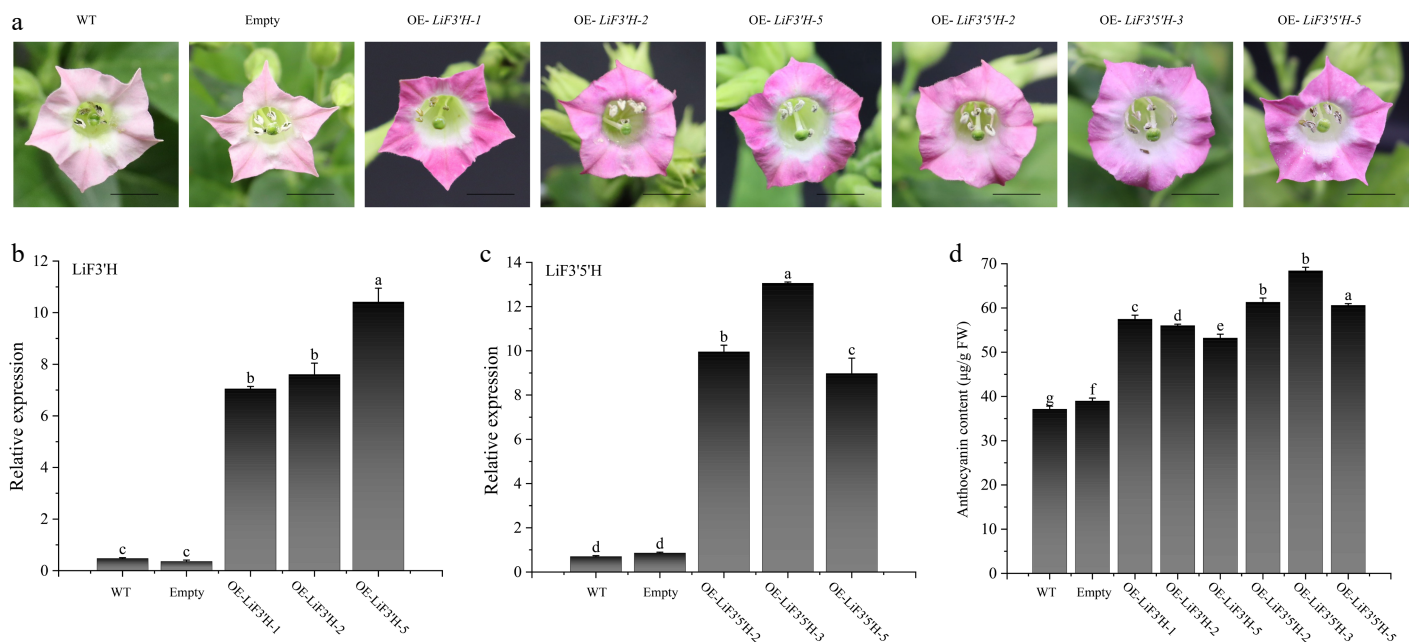
website. The presence of these structural domains further supports the functional diversity of LiCYP450 proteins (members of the P450 superfamily are typically involved in multiple biosynthetic and metabolic pathways). Using the evolutionary tree's classification of P450 genes, LiCYP450 proteins were grouped into nine family clusters, and the P450 genes of the nine families can be further categorized into A and non-A types. The count of A-type (CYP71) genes surpassed that of non-A-type (CYP51/710/711/72/74/85/86/97) genes among them, aligning with findings from prior studies<sup>[41–43]</sup>, implying that the CYP71 family exhibits a broad spectrum of biological functions within plants<sup>[44]</sup>. Promoter analysis revealed that the LiCYP450 gene's upstream 2,000 bp sequence contained a variety of cis-acting elements. These elements encompassed those associated with abiotic stress, developmental processes, hormone responses, and light responses. The existence of these elements implies that the LiCYP450 gene is likely to be extensively involved in the developmental and growth responses of *L. indica*.

Gene family expansion is an important driver of genomic and gene duplication events, including WGD, TD, PD, TRD, and DSD events<sup>[45]</sup>. The investigation into gene duplication occurrences revealed the presence of the LiCYP450 gene family's tandem and segmental duplication events. These duplication occurrences are probably crucial for both the enlargement of the gene family and its functional differentiation<sup>[46]</sup>. By comparing the CYP450 genes with those of *A. thaliana* and *O. sativa*, it was found that *L. indica* has a high number of homologous genes with dicotyledonous plants, which may reflect their evolutionary affinities, consistent with their dicotyledonous genealogical status<sup>[47]</sup>.



**Fig. 7** The subcellular localization of *LiF3'H* and *LiF3'5'H*.

As catalysts for numerous processes during the manufacture of a broad range of secondary metabolites, including flavonoids, steroids, terpenoids, phenylpropanoids, and alkaloids, P450 family genes are essential for the growth and development of plants<sup>[48]</sup>. In the present study, our attention was directed towards the functions of CYP75 family members. Specifically, the CYP75A subfamily member *F3'5'H* and the CYP75B subfamily member *F3'H* are primarily responsible for regulating the biosynthesis of substances like anthocyanins. Anthocyanins, the substances that represent the final products within the flavonoid metabolic pathway, are predominantly



**Fig. 8** (a) The phenotype of LiF3'H and LiF3'5'H transgenic tobacco. (b) The expression level of LiF3'H in WT, empty, and transgenic lines. (c) The expression level of LiF3'5'H in WT, empty, and transgenic lines. (d) The content of anthocyanin. The black line on the petal diagram showed a scale of 1 cm. The values represent the mean  $\pm$  standard error ( $N = 3$ ). When significant differences were present in the values, the one-way ANOVA Duncan test supplement ( $p < 0.05$ ) was employed to denote them, with different letters used for this purpose. 'Empty' here refers to the transgenic plant carrying an empty vector.

present in flowering plants in natural settings and act as a pivotal part of the flower coloration mechanism<sup>[49]</sup>. The functional differentiation of the CYP75 subfamily remains consistently stable in the plant kingdom. Numerous studies have shown that the expression of F3'H and F3'5'H genes is consistent with the formation of cyanidin (red-orange) and delphinidin (blue), which together affect flower color<sup>[50]</sup>. *LiF3'H* and *LiF3'5'H* showed obvious differences in different flower colors at different times, which is presumed to have the function of promoting the synthesis of different anthocyanin glycosides. It is worth noting that the reason for the disparity in expression levels could be associated with the substrate specificity of F3'H and F3'5'H. For example, The *DzF3'H* protein isolated from *Delphinium zaili* plants showed the highest preference for dihydrokaempferol, with the highest catalytic activity. Its catalytic activity towards naringenin, apigenin, and kaempferol was basically at the same level<sup>[51]</sup>. Naringin has been identified as the optimal substrate for the recombinant protein of tea F3'5'H yeast<sup>[52]</sup>.

In order to clarify where *LiF3'H* and *LiF3'5'H* exert their functions, they were constructed into pORE\_R4 vector containing GFP tag by seamless cloning, and their fluorescence localization signals were observed by using the tobacco transient transformation technique, and the results showed that *LiF3'H* and *LiF3'5'H* localized in the cell membrane, and exerted their functions in the cell membrane. The *LiF3'H* and *LiF3'5'H* genes were stably overexpressed in tobacco by the transformation of tobacco leaves by agrobacterium, and their functions were verified by heterologous transformation. It was found that there was a significant accumulation of anthocyanin glycosides on the surface of tobacco corolla relative to WT and null control. The expression of *LiF3'H* and *LiF3'5'H* in the corolla of the transgenic lines was examined, and the results demonstrated that their expression underwent a significant boost ( $p < 0.05$ ), indicating that *LiF3'H* and *LiF3'5'H* were overexpressed in tobacco. Anthocyanin content in the corolla of transgenic tobacco was analyzed, and the findings showed that the transgenic lines

anthocyanin content was significantly elevated compared with the wild type and the empty vector control, suggesting that *LiF3'H* and *LiF3'5'H* were able to promote anthocyanin glycoside synthesis, consistent with previous research findings<sup>[53,54]</sup>.

## Conclusions

In the current study, a comprehensive analysis led to the identification of 314 members belonging to the cytochrome P450 (CYP450) family in *L. indica*. These genes exhibited an uneven distribution across 23 chromosomes and were classified into nine distinct family clusters. Tandem and segmental duplication events were found to significantly influence the evolutionary trajectory of the CYP450 gene family. Numerous action elements, which are responsive to growth metabolism, stress response, and hormonal signals, were detected within the promoter region. The functions of *LiF3'H* and *LiF3'5'H* of the CYP75 subfamily were emphasized. The results provide novel insights into the role of the CYP450 family in the evolutionary processes of species and the developmental trajectory of *L. indica*. Additionally, they show potential for applications in the breeding of *L. indica* flower colors.

## Author contributions

The authors confirm their contribution to the paper as follows: study conception and design: Bai M, Li Z; data collection: Huang J, Jiang R; analysis and interpretation of results: Jin M, Li Z, Sun K; draft manuscript preparation: Bai M, Gu C. All authors reviewed the results and approved the final version of the manuscript.

## Data availability

All data generated or analyzed during this study are included in this published article and its supplementary information files.

## Acknowledgments

This work was supported by the Zhejiang Science and Technology Major Program on Agricultural New Variety Breeding (Grant No. 2021C02071-4). We appreciate the experimental support from the Public Platform of the College of Landscape Architecture, Zhejiang A&F University.

## Conflict of interest

The authors declare that they have no conflict of interest.

**Supplementary information** accompanies this paper at (<https://www.maxapress.com/article/doi/10.48130/opr-0025-0027>)

## Dates

Received 12 February 2025; Revised 6 April 2025; Accepted 15 April 2025; Published online 5 August 2025

## References

- Liu Y, Zetter R, Ferguson DK, Zou C. 2008. Lagerstroemia (Lythraceae) pollen from the Miocene of eastern China. *Grana* 47:262–71
- Hong S, Wang J, Wang Q, Zhang G, Zhao Y, et al. 2022. Decoding the formation of diverse petal colors of *Lagerstroemia indica* by integrating the data from transcriptome and metabolome. *Frontiers in Plant Science* 13:970023
- Klingenberg M. 2003. Pigments of rat liver microsomes. *Archives of Biochemistry and Biophysics* 409:2–6
- Omura T, Sato R. 1962. A new cytochrome in liver microsomes. *The Journal of Biological Chemistry* 237:1375–76
- Lamb DC, Lei L, Warrilow AGS, Lepesheva GI, Mullins JGL, et al. 2009. The first virally encoded cytochrome p450. *Journal of Virology* 83:8266–69
- Barnaba C, Ramamoorthy A. 2018. Picturing the membrane-assisted choreography of cytochrome P450 with lipid nanodiscs. *ChemPhysChem* 19:2603–13
- Šrejšber M, Navrátilová V, Paloncýová M, Bazgier V, Berka K, et al. 2018. Membrane-attached mammalian cytochromes P450: an overview of the membrane's effects on structure, drug binding, and interactions with redox partners. *Journal of Inorganic Biochemistry* 183:117–36
- Werck-Reichhart D, Feyereisen R. 2000. Cytochromes P450: a success story. *Genome Biology* 1:REVIEWS3003
- Toda K, Kuroiwa H, Senthil K, Shimada N, Aoki T, et al. 2012. The soybean F3'H protein is localized to the tonoplast in the seed coat hilum. *Planta* 236:79–89
- Nelson D, Werck-Reichhart D. 2011. A P450-centric view of plant evolution. *The Plant Journal* 66:194–211
- Jiu S, Xu Y, Wang J, Wang L, Liu X, et al. 2020. The cytochrome P450 monooxygenase inventory of grapevine (*Vitis vinifera* L.): genome-wide identification, evolutionary characterization and expression analysis. *Frontiers in Genetics* 11:44
- Vasav AP, Barvkar VT. 2019. Phylogenomic analysis of cytochrome P450 multigene family and their differential expression analysis in *Solanum lycopersicum* L. suggested tissue specific promoters. *BMC Genomics* 20:116
- Wei K, Chen H. 2018. Global identification, structural analysis and expression characterization of cytochrome P450 monooxygenase superfamily in rice. *BMC Genomics* 19:35
- Xu W, Bak S, Decker A, Paquette SM, Feyereisen R, et al. 2001. Microarray-based analysis of gene expression in very large gene families: the cytochrome P450 gene superfamily of *Arabidopsis thaliana*. *Gene* 272:61–74
- Yang J, Li H, Ma R, Chang Y, Qin X, et al. 2022. Genome-wide transcriptome analysis and characterization of the cytochrome P450 flavonoid biosynthesis genes in pigeon pea (*Cajanus cajan*). *Planta* 255:120
- Zhang S, Wu QR, Zhang HM, Pei ZM, Gao JW. 2021. Genome-wide identification and transcriptomic data exploring of the cytochrome P450 family in Chinese cabbage (*Brassica rapa* L. ssp. *pekinensis*). *Journal of Plant Interactions* 16:136–55
- Yang J, Wang G, Ke H, Zhang Y, Ji L, et al. 2019. Genome-wide identification of cyclophilin genes in *Gossypium hirsutum* and functional characterization of a CYP with antifungal activity against *Verticillium dahliae*. *BMC Plant Biology* 19:272
- Tanaka Y, Brugliera F. 2013. Flower colour and cytochromes P450. *Philosophical Transactions of the Royal Society of London Series B, Biological Sciences* 368:20120432
- Holton TA, Brugliera F, Tanaka Y. 1993. Cloning and expression of flavonol synthase from *Petunia hybrida*. *The Plant Journal* 4:1003–10
- Katsumoto Y, Fukuchi-Mizutani M, Fukui Y, Brugliera F, Holton TA, et al. 2007. Engineering of the rose flavonoid biosynthetic pathway successfully generated blue-hued flowers accumulating delphinidin. *Plant & Cell Physiology* 48:1589–600
- Noda N, Yoshioka S, Kishimoto S, Nakayama M, Douzono M, et al. 2017. Generation of blue chrysanthemums by anthocyanin B-ring hydroxylation and glucosylation and its coloration mechanism. *Science Advances* 3:e1602785
- Brugliera F, Barri-Rewell G, Holton TA, Mason JG. 1999. Isolation and characterization of a flavonoid 3'-hydroxylase cDNA clone corresponding to the Ht1 locus of *Petunia hybrida*. *The Plant Journal* 19:441–51
- Ballester AR, Molthoff J, de Vos R, te Lintel Hekkert B, Orzaez D, et al. 2010. Biochemical and molecular analysis of pink tomatoes: deregulated expression of the gene encoding transcription factor SIMYB12 leads to pink tomato fruit color. *Plant Physiology* 152:71–84
- Castellari SD, Di Gaspero G, Marconi R, Nonis A, Peterlunger E, et al. 2006. Colour variation in red grapevines (*Vitis vinifera* L.): genomic organisation, expression of flavonoid 3'-hydroxylase, flavonoid 3',5'-hydroxylase genes and related metabolite profiling of red cyanidin-/blue delphinidin-based anthocyanins in berry skin. *BMC Genomics* 7:12
- Mistry J, Chuguransky S, Williams L, Qureshi M, Salazar GA, et al. 2021. Pfam: the protein families database in 2021. *Nucleic Acids Research* 49:D412–D419
- Potter SC, Luciani A, Eddy SR, Park Y, Lopez R, et al. 2018. HMMER web server: 2018 update. *Nucleic Acids Research* 46:W200–W204
- Wilkins MR, Gasteiger E, Bairoch A, Sanchez JC, Williams KL, et al. 1999. Protein identification and analysis tools in the ExPASy server. *Methods in Molecular Biology* 112:531–52
- Horton P, Park KJ, Obayashi T, Fujita N, Harada H, et al. 2007. WoLF PSORT: protein localization predictor. *Nucleic Acids Research* 35:W585–W587
- Kumar S, Stecher G, Tamura K. 2016. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Molecular Biology and Evolution* 33:1870–74
- Subramanian B, Gao S, Lercher MJ, Hu S, Chen W. 2019. Evolvview v3: a webserver for visualization, annotation, and management of phylogenetic trees. *Nucleic Acids Research* 47:W270–W275
- Bailey TL, Boden M, Buske FA, Frith M, Grant CE, et al. 2009. MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Research* 37:W202–W208
- Chen C, Chen H, Zhang Y, Thomas HR, Frank MH, et al. 2020. TBtools: an integrative toolkit developed for interactive analyses of big biological data. *Molecular Plant* 13:1194–202
- Lescot M, Déhais P, Thijs G, Marchal K, Moreau Y, et al. 2002. PlantCARE, a database of plant cis-acting regulatory elements and a portal to tools for in silico analysis of promoter sequences. *Nucleic Acids Research* 30:325–27
- Qiao X, Li Q, Yin H, Qi K, Li L, et al. 2019. Gene duplication and evolution in recurring polyploidization-diploidization cycles in plants. *Genome Biology* 20:38
- Wang Y, Tang H, DeBarry JD, Tan X, Li J, et al. 2012. MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Research* 40:e49
- Rozewicki J, Li S, Amada KM, Standley DM, Katoh K. 2019. MAFFT-DASH: integrated protein sequence and structural alignment. *Nucleic Acids Research* 47:W5–W10



37. Waterhouse AM, Procter JB, Martin DMA, Clamp M, Barton GJ. 2009. Jalview Version 2—a multiple sequence alignment editor and analysis workbench. *Bioinformatics* 25:1189–91
38. Chen M, Wang Q, Li Y, Gao L, Lv F, et al. 2021. Candidate reference genes for quantitative gene expression analysis in *Lagerstroemia indica*. *Molecular Biology Reports* 48:1677–85
39. Viegas P, Mathews H, Bhatia CR, Notani NK. 1987. Monohybrid and dihybrid segregations in the progenies of tobacco transformed for kanamycin resistance with a Ti-vector system. *Journal of Genetics* 66:25–31
40. Nelson DR, Zeldin DC, Hoffman SMG, Maltais LJ, Wain HM, et al. 2004. Comparison of cytochrome P450 (CYP) genes from the mouse and human genomes, including nomenclature recommendations for genes, *Pseudogenes* and alternative-splice variants. *Pharmacogenetics* 14:1–18
41. Babu PR, Rao KV, Reddy VD. 2013. Structural organization and classification of cytochrome P450 genes in flax (*Linum usitatissimum* L.). *Gene* 513:156–62
42. Ma B, Luo Y, Jia L, Qi X, Zeng Q, et al. 2014. Genome-wide identification and expression analyses of cytochrome P450 genes in mulberry (*Morus notabilis*). *Journal of Integrative Plant Biology* 56:887–901
43. Menéndez-Perdomo IM, Facchini PJ. 2018. Benzylisoquinoline alkaloids biosynthesis in sacred lotus. *Molecules* 23:2899
44. Mizutani M, Ohta D. 2010. Diversification of P450 genes during land plant evolution. *Annual Review of Plant Biology* 61:291–315
45. Sun W, Ma Z, Liu M. 2020. Cytochrome P450 family: genome-wide identification provides insights into the rutin synthesis pathway in Tartary buckwheat and the improvement of agricultural product quality. *International Journal of Biological Macromolecules* 164:4032–45
46. Cannon SB, Mitra A, Baumgarten A, Young ND, May G. 2004. The roles of segmental and tandem gene duplication in the evolution of large gene families in *Arabidopsis thaliana*. *BMC Plant Biology* 4:10
47. 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* 19:307–17
48. Xu J, Wang XY, Guo WZ. 2015. The cytochrome P450 superfamily: key players in plant development and defense. *Journal of Integrative Agriculture* 14:1673–86
49. Zhang Y, Butelli E, Martin C. 2014. Engineering anthocyanin biosynthesis in plants. *Current Opinion in Plant Biology* 19:81–90
50. Grotewold E. 2006. The genetics and biochemistry of floral pigments. *Annual Review of Plant Biology* 57:761–80
51. Miyahara T, Hamada A, Okamoto M, Hirose Y, Sakaguchi K, et al. 2016. Identification of flavonoid 3'-hydroxylase in the yellow flower of *Delphinium zaili*. *Journal of Plant Physiology* 202:92–96
52. Wang YS, Xu YJ, Gao LP, Yu O, Wang XZ, et al. 2014. Functional analysis of flavonoid 3',5'-hydroxylase from tea plant (*Camellia sinensis*): critical role in the accumulation of catechins. *BMC Plant Biology* 14:347
53. Bogs J, Ebadi A, McDavid D, Robinson SP. 2006. Identification of the flavonoid hydroxylases from grapevine and their regulation during fruit development. *Plant Physiology* 140:279–91
54. Sun Y, Huang H, Meng L, Hu K, Dai SL. 2013. Isolation and functional analysis of a homolog of flavonoid 3',5'-hydroxylase gene from *Pericallis × hybrida*. *Physiologia Plantarum* 149:151–59



Copyright: © 2025 by the author(s). Published by Maximum Academic Press, Fayetteville, GA. This article is an open access article distributed under Creative Commons Attribution License (CC BY 4.0), visit <https://creativecommons.org/licenses/by/4.0/>.