

RESEARCH ARTICLE

Genome-Wide Analysis, Classification, Evolution, and Expression Analysis of the Cytochrome P450 93 Family in Land Plants

Hai Du¹, Feng Ran¹, Hong-Li Dong¹, Jing Wen¹, Jia-Na Li^{1*}, Zhe Liang^{2*}

1 College of Agronomy and Biotechnology, Southwest University, Chongqing, 400716, China,

2 Department of Biological Sciences, National University of Singapore, 117543, Singapore, Singapore

* ljn1950@swu.edu.cn (JNL); zhe.liang@nus.edu.sg (ZL)



OPEN ACCESS

Citation: Du H, Ran F, Dong H-L, Wen J, Li J-N, Liang Z (2016) Genome-Wide Analysis, Classification, Evolution, and Expression Analysis of the Cytochrome P450 93 Family in Land Plants. *PLoS ONE* 11(10): e0165020. doi:10.1371/journal.pone.0165020

Editor: Keqiang Wu, National Taiwan University, TAIWAN

Received: June 20, 2016

Accepted: October 5, 2016

Published: October 19, 2016

Copyright: © 2016 Du et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the paper and its Supporting Information files.

Funding: This work was supported by grants from National Natural Science Foundation of China (31471528 and 31671727), National Program on Key Basic Research Project (973 Program, 2015CB150201), and Fundamental Research Funds for the Central Universities of China (XDJK2014B035).

Abstract

Cytochrome P450 93 family (CYP93) belonging to the cytochrome P450 superfamily plays important roles in diverse plant processes. However, no previous studies have investigated the evolution and expression of the members of this family. In this study, we performed comprehensive genome-wide analysis to identify CYP93 genes in 60 green plants. In all, 214 CYP93 proteins were identified; they were specifically found in flowering plants and could be classified into ten subfamilies—CYP93A–K, with the last two being identified first. CYP93A is the ancestor that was derived in flowering plants, and the remaining showed lineage-specific distribution—CYP93B and CYP93C are present in dicots; CYP93F is distributed only in Poaceae; CYP93G and CYP93J are monocot-specific; CYP93E is unique to legumes; CYP93H and CYP93K are only found in *Aquilegia coerulea*, and CYP93D is Brassicaceae-specific. Each subfamily generally has conserved gene numbers, structures, and characteristics, indicating functional conservation during evolution. Synonymous nucleotide substitution (d_N/d_S) analysis showed that CYP93 genes are under strong negative selection. Comparative expression analyses of CYP93 genes in dicots and monocots revealed that they are preferentially expressed in the roots and tend to be induced by biotic and/or abiotic stresses, in accordance with their well-known functions in plant secondary biosynthesis.

Introduction

Cytochrome P450 monooxygenases (P450s) are widely distributed in eukaryotes; they form a large and diverse class of enzymes, with more than 35,000 members [1–2] (<http://drnelson.uthsc.edu/P450.count.April.6.2016.PPTX>). P450s are involved in numerous biosynthetic and xenobiotic pathways such as the assimilation of carbon sources, detoxification of xenobiotics, and synthesis of secondary metabolites [2–3]. They have relatively low sequence identity among different organisms (such as plants and animals); however, they show a common overall topology and tridimensional fold. Moreover, all P450s share four common characteristic

Competing Interests: The authors have declared that no competing interests exist.

regions—the proline-rich membrane hinge; I-helix involved in oxygen binding; K-helix (ExxR); and “PERF” consensus, which consists of the ERR triad involved in locking the heme pocket into position to ensure the stabilization of the conserved core structure [4–5].

In plants, the cytochrome (CYP) P450 superfamily is one of the largest gene families of enzyme proteins, and it includes 245 genes in *Arabidopsis* [6–7] and 332 full-length genes and 378 pseudogenes in soybean [8]. Plant P450s were originally classified into two main types—the A-type and non-A type [9–10]. Based on the numerous available genome sequences, researchers subsequently re-classified them into 11 clans. The A-type was grouped as the CYP71 clan, whereas the non-A type was subdivided into 10 clans—CYP51, CYP72, CYP74, CYP85, CYP86, CYP97, CYP710, CYP711, CYP727, and CYP746 [1,3,10]—according to a common nomenclature system [11]. However, new clans might emerge in the future as more lineages of plants are sequenced.

Plant P450s participate in various biochemical pathways leading to the production of primary and secondary metabolites [3]. For instance, most structural genes in the flavonoid and/or isoflavonoid biosynthesis pathways are CYP P450 genes such as *C4H* (CYP73), *F3'5'H* (CYP75A), and *IFS* (CYP93C) [12]. Notably, the diversification of P450s had a significant biochemical impact on the emergence of new metabolic pathways during the evolutionary process of land plants. A typical example is cytochrome P450 93C (CYP93C), which is mainly identified in legumes and is the first key enzyme involved in the metabolism of the legume-specific isoflavonoid biosynthesis pathway [12–16]. Therefore, CYP93C is of considerable interest in the genetic and metabolic engineering fields for its important roles in plant defense and human health, since it enhances the dietary intake of isoflavonoids and thus facilitates disease prevention. Similarly, several other members such as CYP93B, CYP93E, and CYP93G are known to play important roles as plant secondary metabolites, such as flavone [17–22] and triterpenoid saponin biosynthesis [23]. However, unlike the large number of P450 gene families in angiosperms, only a few members of CYP93 genes have been functionally characterized.

To date, genome-wide analysis of P450 super gene family has been performed in many plant species, based on the available genome sequences, such as *Arabidopsis* [10,24], *Oryza sativa*, *Vitis vinifera* L., *Populus tremuloides* [25], and *Nelumbo nucifera* [26]. However, a similar comprehensive analysis of P450 proteins in a given gene family, such as CYP93 genes, is still lacking. Thus, the gene structures, evolutionary relationship, and expression patterns of CYP93 gene family in land plants is yet not known, and a complete survey and classification in plants from distinct evolutionary groups are necessary; such data could markedly enhance our understanding of the evolutionary history as well as functions of CYP93 genes as plant secondary metabolites. Therefore, in the present study, we investigated the evolutionary history of the CYP93 family across 60 plants (including 53 angiosperms). We performed structural and evolutionary analyses across different plant evolutionary lineages and assessed their origins, classification, evolutionary relationship, and expression. Our study extended the sequence and functional characteristics of CYP93 gene family in land plants and might facilitate future functional analysis.

Material and Methods

Sequence retrieval

To identify CYP93 genes in plant genomes, we performed BLASTP searches of the sequenced plant genomes in the Phytozome (<http://www.phytozome.net/>) [27] and PGDD (<http://chibba.pgml.uga.edu/duplication/>) [28] databases by using representative CYP93 proteins such as CYP93A1, CYP93E1, and CYP93C1v2 as queries (p -value $< e^{-10}$). The detailed information of the representative sequences is listed in S1 Table. The species represented a broad range of

plant lineages, from unicellular green algae to multicellular plants. We used a relatively uniform criterion [1,11] to collect CYP93 genes with high-quality sequences. Next, the candidates were sent to Dr. David Nelson for uniform nomenclature (David Nelson, P450 Nomenclature Committee, personal communication). The detailed information (i.e., accession numbers) regarding the CYP93 sequences is listed in [S2 Table](#).

Sequence alignment and phylogenetic analyses

The CYP93 sequences were aligned using MAFFT version 7 [29] with the G-INS-i algorithm, followed by manual editing in MEGA 6.06 [30]. In our subsequent analyses, we only included positions that were unambiguously aligned. The neighbor joining (NJ) phylogeny was performed using MEGA 6.06 with 1000 replicates; we used the p-distance model and pairwise deletion. Similarly, a maximum likelihood (ML) tree was constructed using MEGA v6.06, with 100 replicates and Poisson distribution.

Gene expression analysis

The *Arabidopsis*, soybean, maize, and rice microarray-based datasets (GEO accession numbers: GSE17883, GSE18518, GSE26198, GSE41125, GSE6901, GSE6908, GSE14275, GSE35984, GSE19024, GSE33410, GSE9687, and GSE35427) were downloaded from the Plant Expression database (PLEXdb, <http://www.plexdb.org/index.php>) [31]. A hierarchical cluster was created using Cluster 3.0 [32] by using hierarchical clustering based on log₂ transformed data of the normalized expression data and mean method and viewed using Java TreeView [33].

Expression analysis of CYP93 genes in *Arabidopsis*, soybean, maize, and rice

The expression information of the CYP93 genes in *Arabidopsis*, soybean, maize, and rice was further confirmed by real-time PCR. Given the high homology of the candidates in a species, all gene-specific primers were designed to avoid false priming by creating gene-specific sites at the 3'-terminal of each primer, leading to the amplification of 100–200-bp long products. The specific primers for the actin gene were used as an internal control for each species. The PCR primer sequences are shown in [S3 Table](#).

The plant tissues (roots, stems, leaves, flowers, etc) of maize, rice, *Arabidopsis*, and soybean were harvested and ground in liquid nitrogen. Total RNA was extracted using Eastep™ total RNA Extraction kit, according to the manufacturer's instructions, and treated with DNase I (Promega, USA). First-strand cDNA synthesis was performed using an oligo (dT) primer and 2 µg of total RNA in a 20-µl reaction volume, according to the manufacturer's instructions for the M-MuLV RT kit (Takara Biotechnology, Japan). The real-time PCR thermocycling parameters were as follows: an initial denaturation for 3 min at 95°C, followed by 45 cycles of a denaturation step at 95°C for 15 s and an annealing step at 58°C for 20 s. The fluorescence was measured after the extension step by using the CFX Connect™ Real-Time System (Bio-Rad). After the thermocycling reaction, the melting step was performed from 65°C up to 95°C, with an increment of 0.5°C each 0.05 s. Each PCR pattern was independently verified in three replicate experiments. The specificity of primers used in this study was verified by cloning the PCR products into the pTA2-T vectors (TOYOBO, Japan), and then using them for sequencing (data not shown).

Selection analysis

We performed selection analysis as described previously [34]. The cDNA sequences were aligned using MUSCLE in MEGA5.2.2 software [35] by using the codon model algorithm and were then loaded into HyPhy [36], along with the corresponding ML phylogenetic trees. The HyPhy batch file NucModelCompare.bf with a model rejection level of 0.0002 was used to establish the best fit of 203 general time-reversible (GTR) models of nucleotide substitution. The HyPhy batch file QuickSelectionDetection.bf was used to estimate site-by-site variations in nucleotide substitution rates.

The structure was modeled using the SWISS-MODEL [37]. Template search was performed using Protein BLAST. The best model 5e58.1.A was selected based on Q-Mean. PyMOL was used for protein structure analysis and prepare figures.

Results

Distribution of CYP93 genes across angiosperms

To identify candidate genes in angiosperms, we used the known CYP93 representative sequences as queries to search the 60 sequenced species in the Phytozome v10 (<http://www.phytozome.net/>) [27] and PGDD (<http://chibba.pgml.uga.edu/duplication/>) databases [28].

We identified 214 candidate genes in almost all the flowering plants among the 60 angiosperms investigated (Fig 1 and S2 Table). In contrast, we did not identify any CYP93 sequences in the non-flowering plant genomes, including chlorophyte algae (*Chlamydomonas reinhardtii*, *Volvox carterii*, *Micromonas pusilla* RCC299, and *Ostreococcus lucimarinus*), Bryophyta (*Physcomitrella patens*), lycophytes (*Selaginella moellendorffii*), or gymnosperms (*Picea abies*; Fig 1). To confirm these results, we then searched the NCBI database, including expressed sequence tags (ESTs), and did not find any CYP93s in the non-flowering species included in this database as well. Interestingly, we also did not identify any CYP93 candidates in the genome of *Beta vulgaris*, but found two partial sequences in Genebank that were previously reported; thus, we included them in our following analysis.

The CYP93 gene family was found to have a wide distribution in flowering plants. Since CYP93 proteins are present both in monocots and eudicots, the appearance of CYP93 genes was thought to predate the divergence of monocots and eudicots (Fig 1).

Phylogenetic analysis of the CYP93 gene family

According to the common nomenclature system, P450s in the same family generally share at least 40% identity, and those in the same subfamily share at least 55% identity [11]. However, owing to gene duplication and shuffling, a straightforward nomenclature might be difficult; hence, family definition is recommended for integrating phylogeny and organization of genes [7].

Initially, we used an established nomenclature to group candidates by comparing their identity, which was performed by Dr. Nelson [11]. Our results showed that the amino acid sequences of all the CYP93 candidate proteins shared approximately 53% identity (S4 Table). The sequence identities among the CYP93A, CYP93B, CYP93C, CYP93D, CYP93E, CYP93F, CYP93G, CYP93H, and CYP93J proteins (CYP93K has only one member) were approximately 65%, 63%, 82%, 83%, 70%, 65%, 66%, 80%, and 47%, respectively. This result showed that the sequence homology in this gene family was highly conserved.

To elucidate the evolutionary relationships of this gene family in plants, we aligned the 214 candidate CYP93 proteins by using MAFFT software and constructed NJ and ML phylogenetic trees (Fig 2 and S1 Fig). Our results showed that the topologies of these two analyses were

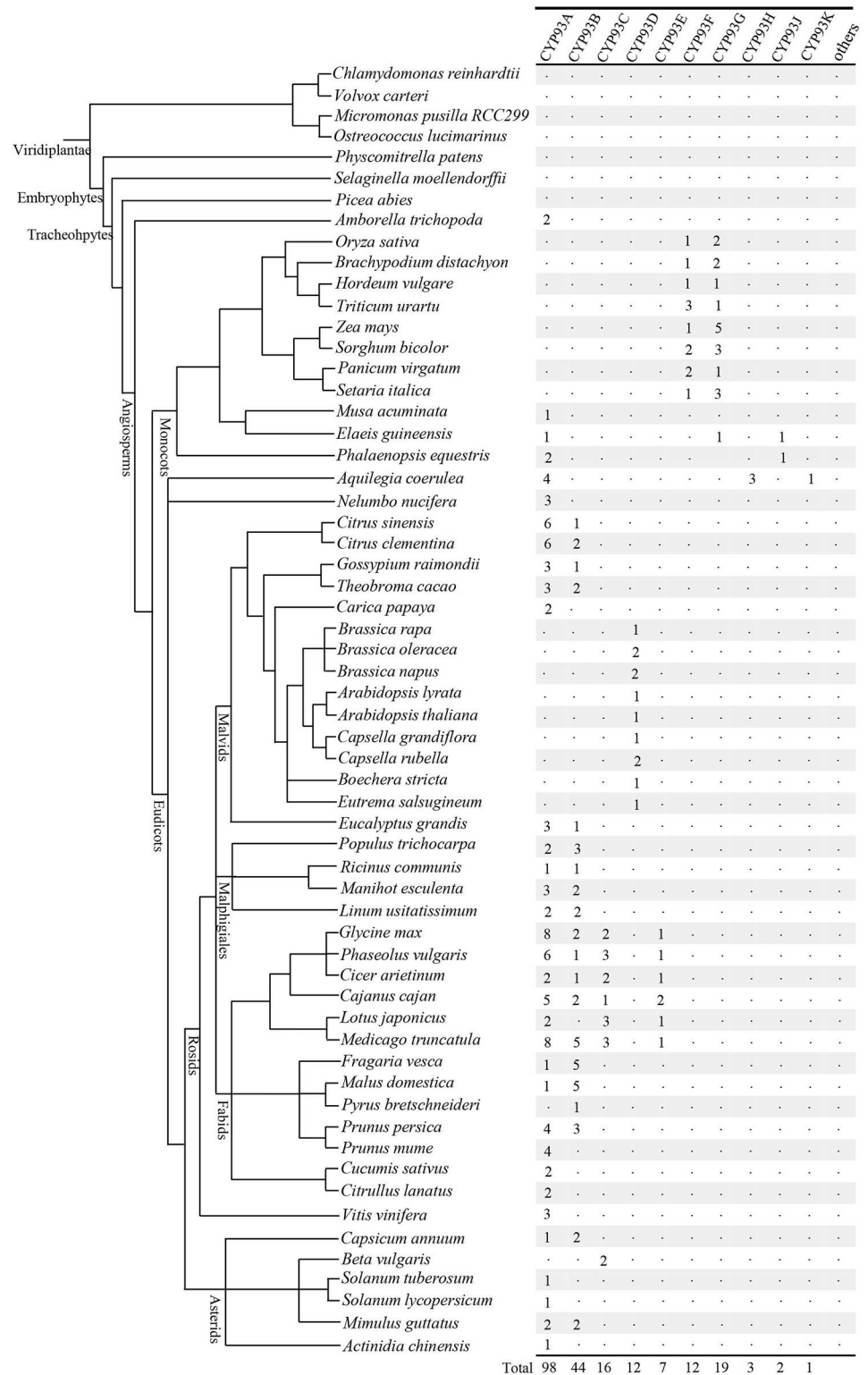


Fig 1. Phylogenetic relationships of the 60 species investigated in the present study. Phylogenetic relationships (branch lengths are arbitrary) among these species have been described previously (<http://www.phytozome.net/>). The total number of cytochrome P450 93 (CYP93) proteins identified in each genome is indicated on the right.

doi:10.1371/journal.pone.0165020.g001

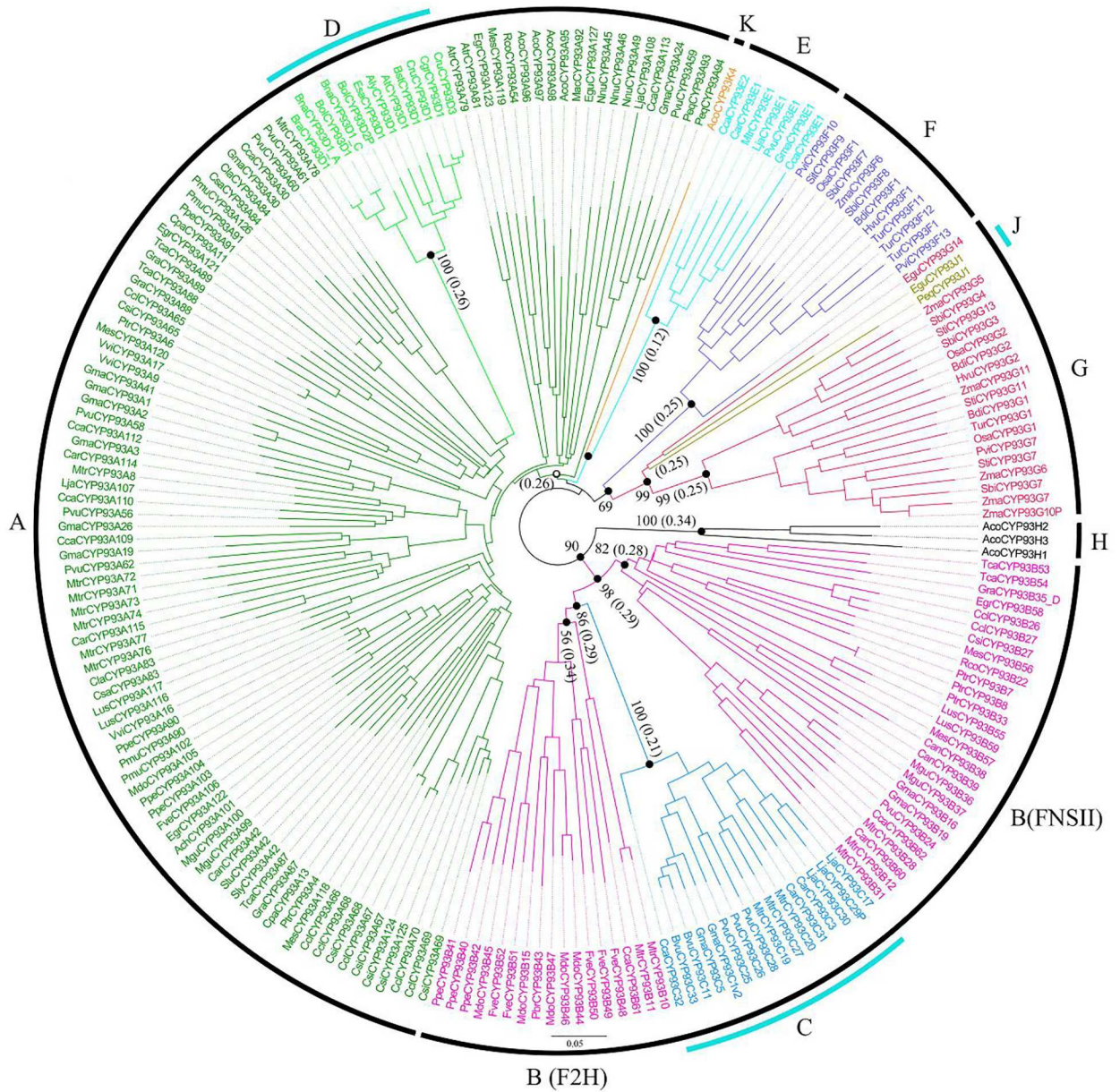


Fig 2. Phylogenetic tree and classification of 214 plant CYP93 proteins. The neighbor joining (NJ) tree includes 214 CYP93 proteins from 60 eukaryotes. Proteins are clustered into eight subfamilies (e.g., CYP93A). The colored lines and names symbolize the species to which the proteins in each clade belong. The black dots represent the major clades in the phylogenetic tree, and the corresponding bootstrap support values from 1000 replications are shown beside the black dots. Bootstrap values <50% are shown as black circles, and those >50% are shown as black dots in the phylogenetic tree. The numbers in brackets indicate the dN/dS value for each subfamily or branch. The information of species abbreviations used for the tree of Fig 2 is listed in S2 Table. The scale bar represents amino acid substitutions per site.

doi:10.1371/journal.pone.0165020.g002

highly consistent. In our phylogenetic trees, the candidates from the same lineage tended to be clustered together, resulting in many lineage-specific subfamilies and/or clades. This might indicate that this gene family could have evolved or been lost in a specific lineage, following divergence. Moreover, the CYP93 proteins were not equally represented in different species, suggesting that they experienced duplications after the divergence (Fig 2).

In the phylogenetic tree, the CYP93 proteins were clustered into ten independent clades with high bootstrap support (except CYP93A), generally following the evolutionary order (Fig 2 and S2 Fig). Our dataset yielded several interesting findings. First, CYP93A is the largest subfamily, which includes genes from nearly all flowering plants (including monocots and dicots), except grasses and Brassicaceae, which formed some lineage-specific subfamilies instead. Second, CYP93B sequences also have a relatively wider distribution in plants and are present in many eudicot species, but not in Brassicaceae. Third, CYP93D sequences are distributed only in Brassicaceae and are embedded in the CYP93A subfamily, indicating that they might have originated from one or more gene duplications from ancient CYP93A subfamilies. Fourth, similar to CYP93D, CYP93C has members in all legumes and in *Beta vulgaris* and is embedded in the CYP93B subfamily, implying that it originated from CYP93B and was subsequently conserved during the evolution. Fifth, CYP93E sequences are specifically found in legumes. Sixth, the monocot CYP93 sequences were clustered into the CYP93F, CYP93G, and CYP93J branches; the CYP93F branch showed orthologous distribution only in grasses, whereas the CYP93G branch had members in grasses and *Elaeis guineensis*, indicating that it is older than CYP93F. Finally, the CYP93H and CYP93K sequences were distributed only in the basic dicot species *Aquilegia coerulea* (Figs 1 and 2).

Taken together, phylogenetic analysis revealed that the CYP93 genes were newly derived in flowering plants: an initial divergence from an ancient CYP93A clade by a gene duplication event during evolution, and a subsequent divergence via gene duplication in the CYP93A clade to yield the subfamilies CYP93B–CYP93K.

Conserved characteristics of CYP93 sequences

The amino acid sequences of P450s are relatively diverse and have sequence identities as low as 20% [38]. However, the overall secondary and tertiary structures of P450s are highly conserved, because some primary regions/motifs such as the K-helix, PERF, and heme-binding domains, which are important to the secondary and tertiary structures, are conserved across eukaryotic P450s [3,20]. We assessed the sequence characteristics of the CYP93 gene family by conducting sequence alignment analysis (S2 Fig) and protein homology modeling (S3 Fig). The entire sequence of the CYP93 proteins based on our alignment analysis is shown in Fig 3.

One striking feature is that the overall protein modeling structures are very similar to each other (S3 Fig); thus, our detailed sequence analysis was mainly based on sequence alignment. With the exception of five regions having relatively low homology and variable sequence characteristics and lengths (especially the N- and C-terminal regions), the sequences of CYP93 proteins generally shared a high degree of similarity (Fig 3). Moreover, the sequences contained some highly conserved short amino acid motifs that are distributed throughout the coding regions (Fig 3). For example, consistent with the findings of previous studies [3,20], the four well-known P450 motifs—the PERF, K-helix, I-helix, and heme-binding domains—are generally relatively conserved across different subfamilies (>87% identity). However, the transmembrane domain shares only approximately 24% identity (Fig 4). In addition, with the exception of the commonly conserved residues, some clade- or subfamily-specific insertion or deletion sites are present. For instance, the sites at positions 57 and 104 are distributed only in the CYP93C subfamily, and the amino acids at positions 202–204 are missing in the CYP93G subfamily (Fig 3). Since these positions are located on the protein surface (S3 Fig), they might affect protein–protein interactions. Interestingly, we also found lineage-specific conservation of several amino acid substitutions at specific sites in the heme-binding and I-helix domains, compared to those in PERF and K-helix (Fig 4). Based on these findings, we speculated that most CYP93 genes are placed in a clear monophyletic clade and have subfamily-specific

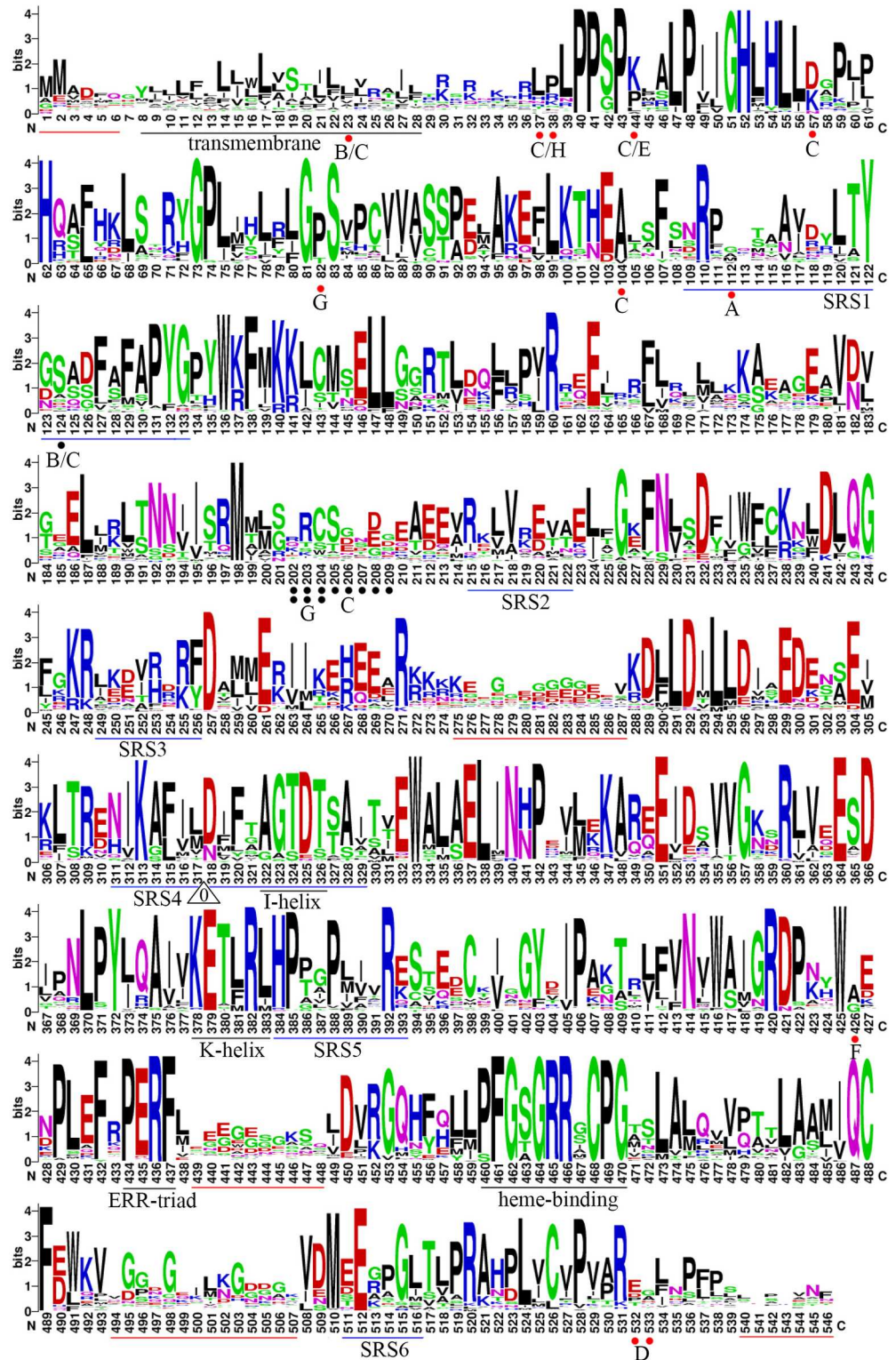


Fig 3. Sequence logos of the multiple alignments of the 214 CYP93 proteins in plants. The sequence logos of plant CYP93 proteins based on amino acid alignment using MAFFT are shown. The logos were generated using Weblogo. The bit score indicates the information content for each position in the sequence. The height of the letter designating the amino acid residue at each position represents the degree of conservation. The key conserved motifs are underlined; the red lines indicate the less conserved regions; the

black ones, the P450 motifs; and the blue ones, the substrate recognition sites (SRSs). The white triangles indicate the conserved intron insertion location of plant CYP93 genes; the numbers within the triangles indicate the splicing phase of the intron (0 refers to phase 0). The red and black dots indicate the conserved amino acid insertion or deletion sites, respectively, in a given subfamily and/or clade; the number below each dot indicates the corresponding subfamily, i.e., B indicates the CYP93B subfamily.

doi:10.1371/journal.pone.0165020.g003

features. For example, the residues in the fourth and eighth sites of the heme-binding domain varied across different subfamilies—most CYP93A subfamily members have a conserved G/S residue in the eighth site, whereas those of the CYP93C–CYP93J subfamilies have conserved I, S, or M residues instead (Fig 4). These findings suggest that conserved residue substitutions in different subfamilies are key variants during evolution.

Despite considerable sequence variation between P450 enzymes, six substrate recognition sites (SRS1–SRS6) were identified as essential residues [39], and several key amino acid residues that affect substrate specificity are known to be found within these SRSs [40–42]. Thus, SRSs might be important for P450 binding and substrate recognition specificity (Fig 5). To investigate the sequence characteristics of these regions among the ten CYP93 subfamilies, we further compared the corresponding sequences. Our results showed that the overall identities among the SRS1, SRS2, SRS3, SRS4, SRS5, and SRS6 regions were approximately 62%, 50%, 39%, 70%, 64%, and 56%, respectively, indicating that these regions are less conserved compared to the four P450 motifs discussed above. Moreover, the residues in these six SRSs generally showed some obvious subfamily-specific site-substitutions, implying that they are involved in functional diversification among different CYP93 subfamilies. For instance, Ser 310 (326 in the present study), Leu 371 (387), and Lys 375 (391) were reported to be the key active-site residues for the CYP93C2 enzyme [13,43]. Ser 310 and Leu 371 are critical for substrate accommodation of the CYP93C2 enzyme in favor of hydrogen abstraction from C-3 of the flavanone molecule and in the presence of Lys 375, respectively. Lys 375 is responsible for aryl migration.

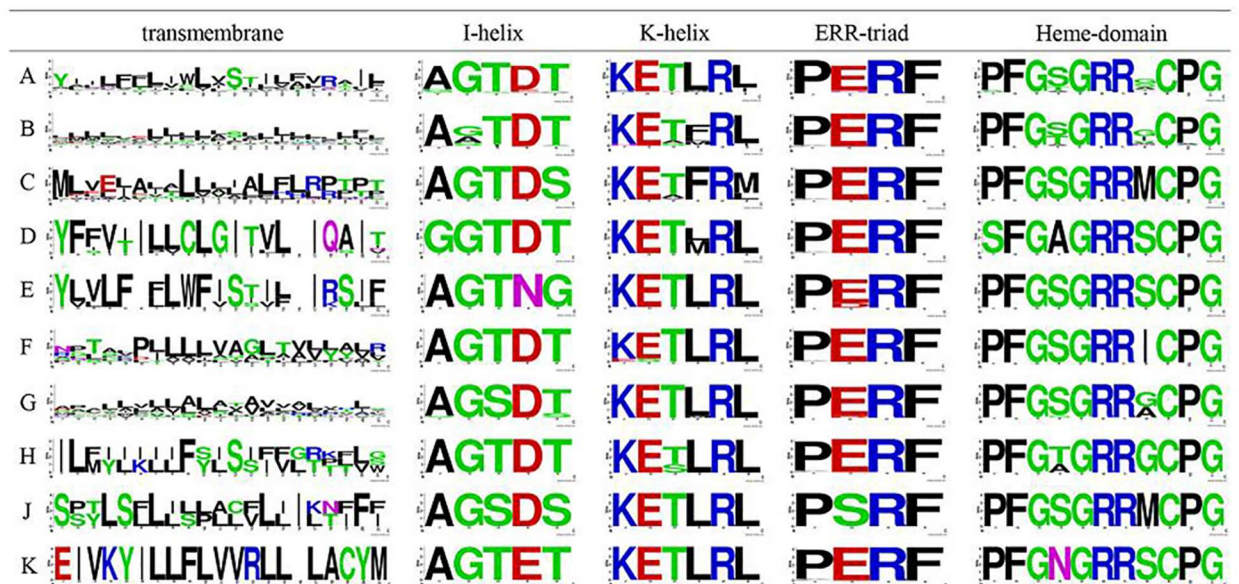


Fig 4. Architecture of conserved protein motifs in the eight subgroups of the plant CYP93 family. The sequence logos of the P450 transmembrane, I-helix, K-helix, PERF, and heme-binding motifs based on the amino acid alignments are shown. The bit score indicates the information content for each position in the sequence. A–K indicate subfamilies CYP93A–CYP93K.

doi:10.1371/journal.pone.0165020.g004

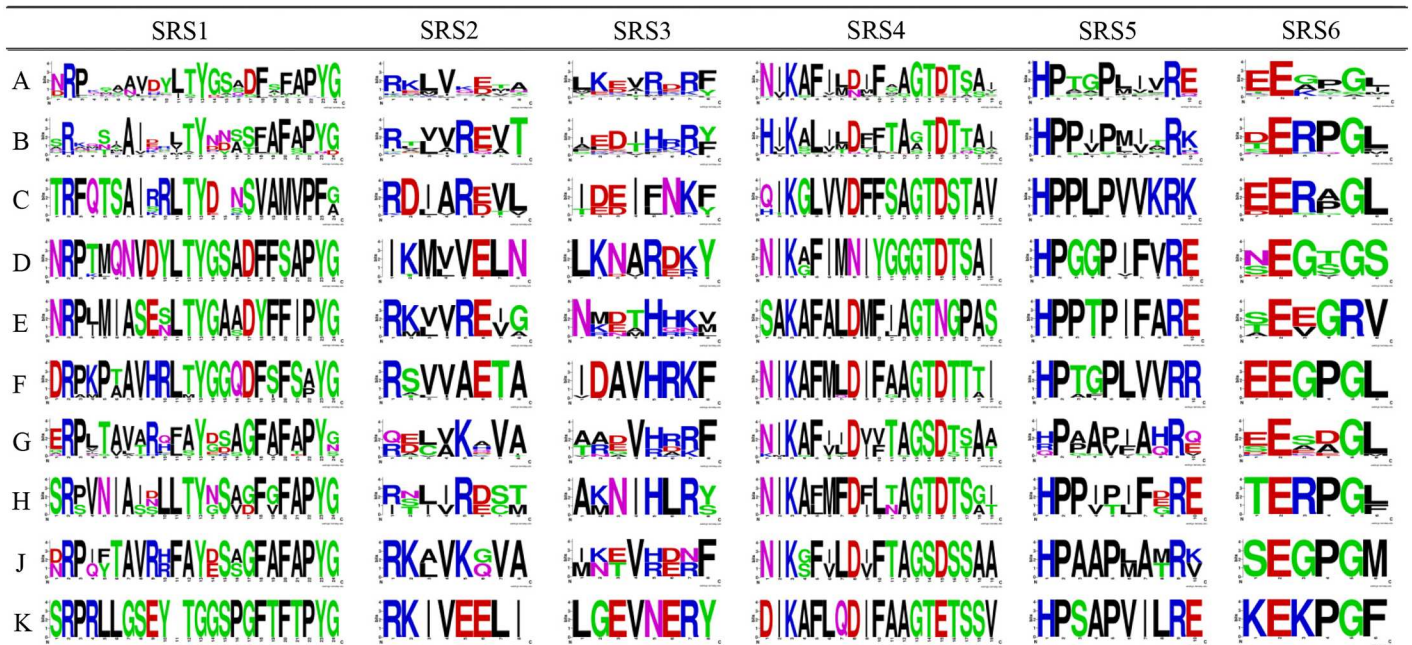


Fig 5. Weblogo of SRSs based on the amino acid alignments across the eight subgroups of the plant CYP93 family. The bit score indicates the information content for each position in the sequence. A–K indicate subfamilies CYP93A–CYP93K.

doi:10.1371/journal.pone.0165020.g005

Interestingly, the Ser 310 residue is located in the I-helix motif embedded in SRS4 and shows subfamily-specific substitutions (Fig 4). Similarly, Leu 371 and Lys 375 are located in SRS5, which also show conserved substitutions between different subfamilies (Fig 5). Therefore, these results confirmed that the SRS regions played an important role in the functional diversification of CYP93 families during evolution.

Furthermore, based on the available genome sequences (DNA and cDNA sequences), we analyzed the intron distributions in the encoding regions of the CYP93 genes. Our results revealed that 128 of the 160 analyzed CYP93 genes (80%) were disrupted by the highly conserved “M” intron [3,10], 26 genes (16%) had introns inserted at one or more sites, and only six genes (4%) were intron-less (Fig 3). Interestingly, most of the 26 genes having more than two introns had the conserved “M” intron site, whereas the remaining insertion sites were not conserved; this could be attributed to the inaccurate genome annotation among some species. Moreover, all the 26 multi-intron genes were CYP93A and CYP93B genes, whereas the six intron-less were CYP93G and CYP93F (S2 Table). Overall, with the exception of a few genes (4%), the majority of plant CYP93 genes (96%) have a highly conserved intron insertion position (Fig 3). Our results revealed that the splicing phase of the conserved intron was completely conserved to phase 0 (Fig 3) [44], indicating that the splicing mechanism of CYP93 genes was highly conserved during evolution. Considering that the main intron pattern is consistent with that of CYP712 and CYP705 family [10], these gene families could be thought to have a common ancestor.

Taken together, our findings suggest that the exon/intron structure of the CYP93 gene family was highly conserved during evolution (Fig 3); moreover, the SRSs generally show subfamily-specific amino acid substitutions at some sites (Fig 5), highlighting the nature of their functional diversification in different subfamilies.

Role of selection in the CYP93 coding sequences

Comparing the dN and dS rates is a common method of determining selection pressures on coding regions. Commonly, a dN/dS value of 1 is used to indicate neutral selection, and values of <1 or >1 are used to indicate purifying and positive selection, respectively.

We investigated the influence of selective constraints on the CYP93 coding sequence. By globally fitting an evolutionary model, we calculated the ratios for each subfamily and/or clade (S5 Table). The dN/dS value of each subfamily or clade was found to be <1 in all groups, suggesting that strong purifying selection has been maintained across plants and implying that CYP93 genes are functionally conserved (Fig 2). At the individual codon level, most residues were under significant negative selection except several sites that were under positive selection; however, many sites were also under relaxed constraint. Consistent with sequence analysis, these relaxed constraint sites were mainly located in the regions with low homology and had variable lengths, and they were less frequent in the relatively conserved P450 domains and SRS regions (S5 Table).

Expression analysis of CYP93 genes at different developmental stages

To understand the expression profiles of CYP93 genes and elucidate their functions, we compared their expression profiles across two representative dicot and monocot plants, including *Arabidopsis*, soybean, maize, and rice (Fig 6).

Based on the AtGenExpress dataset [45], we showed that *AtCYP93D1* was strongly expressed in the roots and weakly expressed in the other tissues and/or organs (Fig 6A). Soybean dataset1 [46] results suggested that most CYP93A and CYP93C genes were preferentially expressed in the roots (Fig 6B), with the exception of *GmaCYP93A41* and *GmaCYP93B19* that were not expressed or very weakly expressed in all the tissues and organs investigated. Further, *GmaCYP93A2*, *GmaCYP93A24*, *GmaCYP93A3*, *GmaCYP93A19*, *GmaCYP93A26*, and *GmaCYP93A30* were expressed only in the roots; *GmaCYP93A1* was expressed in both the roots and nodules; *GmaCYP93C1v2* and *GmaCYP93C5* were strongly expressed in 35–42-day-old seeds, nodules, and roots, with the highest expression levels in roots; *GmaCYP93E1* was strongly expressed in young leaves, green pods, and 35–42-day-old seeds; and *GmCYP93B16* was expressed in early pods and young leaves. Similarly, the soybean dataset2 [47] suggested that the CYP93 genes were more strongly expressed in the roots and root hairs, except for *GmaCYP93A1* and *GmaCYP93B2*, which were not remarkably expressed in all the tissues investigated (Fig 6C): *GmaCYP93A1*, *GmaCYP93A2*, *GmaCYP93A3*, *GmaCYP93A26*, *GmaCYP93A41*, *GmaCYP93C1v2*, and *GmaCYP93C5* were mainly expressed in the roots and root hairs; *GmaCYP93A19* and *GmaCYP93A30* were strongly expressed in flowers, roots, and leaves; *GmaCYP93B16* was strongly expressed in the shoot apical meristem, flowers, and green pods; and *GmaCYP93E1* showed the strongest expression levels in the root tip, green pods, and root hairs. Taken together, our results suggested that CYP93 genes are mainly expressed in dicot roots, root hairs, and/or nodules, implying their conserved expression pattern during evolution.

In rice, we used two relatively comprehensive expression profile datasets from PLEXdb [31]. Dataset1 suggested that *OsaCYP93G1* and *OsaCYP93G2* showed relatively wide expression profiles, with strong expression in the seedlings, shoots, roots, radicles, stems, spikelets, and sheaths (Fig 6D). Dataset2 indicated that *OsaCYP93G1* and *OsaCYP93G2* showed similar expression profiles, with the strongest expression levels in the roots, leaves, embryo, shoots, stigma, and callus (Fig 6E). *OsaCYP93F1* was constitutively and very weakly expressed in all the investigated tissues and organs (Fig 6D and 6E). In maize [48], *ZmaCYP93G7*, *ZmaCYP93G5*, *ZmaCYP93G11*, *ZmaCYP93G10P*, and *ZmaCYP93G6* were relatively strongly expressed in the roots, seedlings, tassels, leaves, and husks (Fig 6F). *ZmaCYP93F6* was

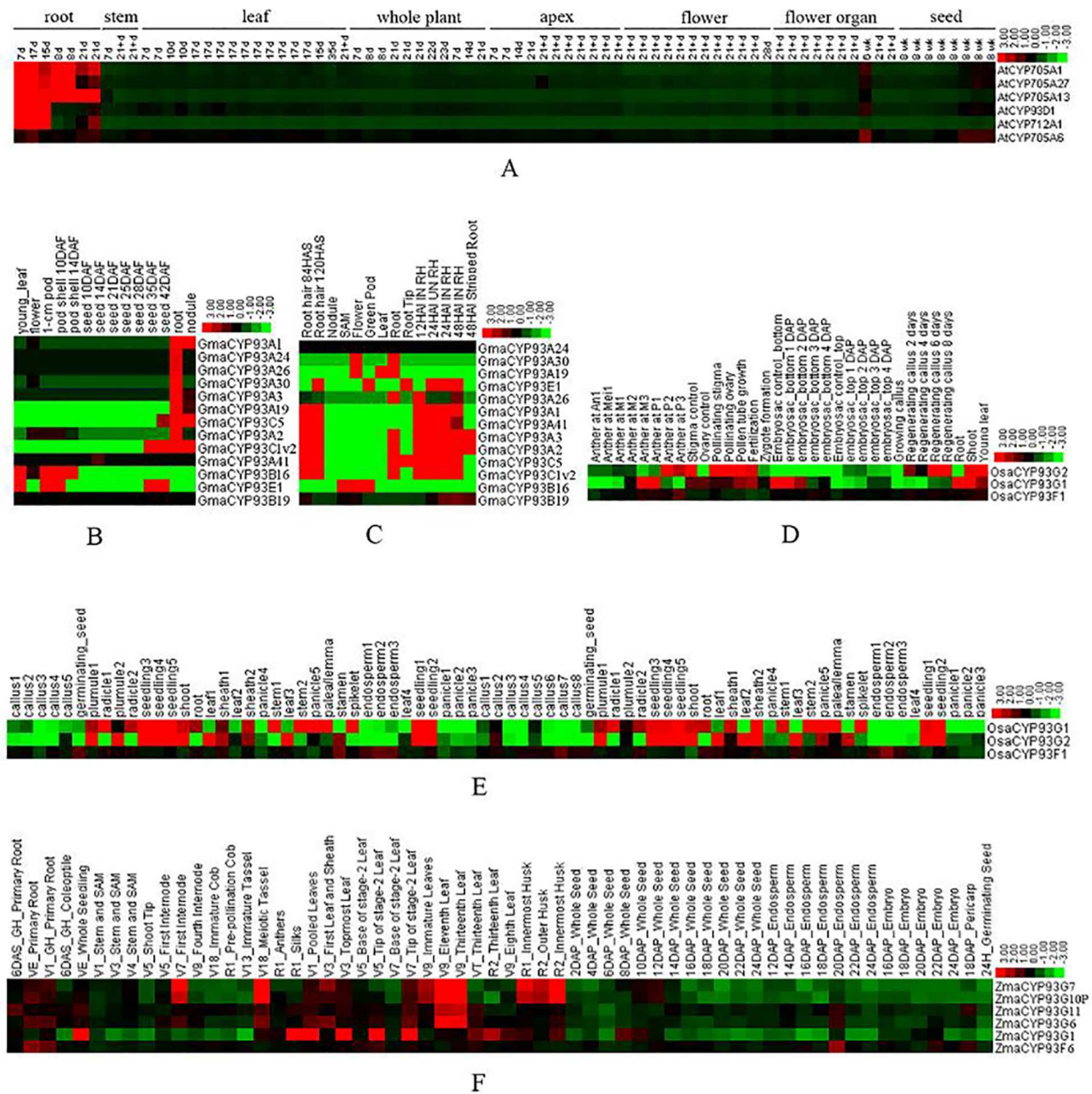


Fig 6. Expression profiles of CYP93 homologous genes in *Arabidopsis*, soybean, rice, and maize. (A) Expression profiles of the *AtCYP93D1* gene in *Arabidopsis*. (B) and (C) expression profiles of *GmCYP93* genes in soybean expression dataset1[46] and dataset2 [47]. (D) and (E) expression profiles of *OsCYP93* genes in rice expression dataset1 (GSE14304) and dataset2 (GSE19024)[31]. (F) expression profiles of *ZmCYP93* genes in maize[48]. Color bar at the base represents log₂ expression values.

doi:10.1371/journal.pone.0165020.g006

constitutively expressed in all the investigated tissues and organs, with slightly stronger expression levels in the roots and embryo.

Taken together, our results showed that dicot CYP93 genes were preferentially expressed in the root tissues, whereas monocot CYP93 genes were strongly expressed in the roots and many other tissues and/or organs, indicating wider expression in monocots than in dicots.

The expression pattern of plant CYP93 genes by real-time PCR

Next, we analyzed the expression profiles of CYP93 gene family in *Arabidopsis*, soybean, maize and rice (Fig 7). The results showed that most of the CYP93 genes yielded positive qRT-PCR results, except *OsaCYP93F1*, which showed no expression signals (Fig 7D) and could be expressed at specific developmental stages or under special conditions.

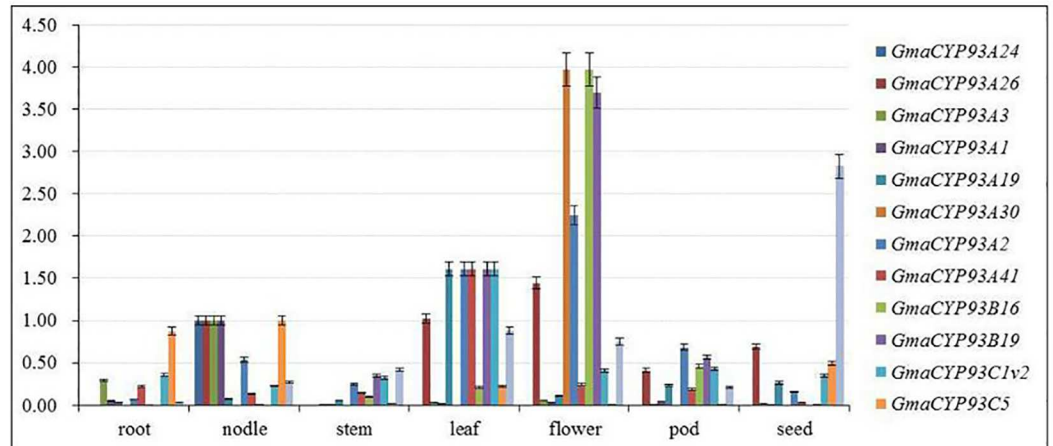
In soybean, the *GmaCYP93* genes (*GmaCYP93A1*, *GmaCYP93A2*, *GmaCYP93A3*, *GmaCYP93A19*, *GmaCYP93A24*, *GmaCYP93A26*, *GmaCYP93A30*, *GmaCYP93A41*, *GmaCYP93B16*, *GmaCYP93B19*, *GmaCYP93C1v2*, and *GmaCYP93C5*) showed high expression in the roots, stems, leaves, flowers, and seeds (Fig 7A). In *Arabidopsis*, the *AtCYP93D* showed relatively higher expression level in the roots and leaves (Fig 7B). In maize, *ZmaCYP93G5*, *ZmaCYP93G7*, *ZmaCYP93G11*, and *ZmaCYP93G10P* were highly expressed in the leaves, whereas *ZmaCYP93F6* and *ZmaCYP93G6* showed preferential expression in the roots (Fig 7C). In rice, *OsaCYP93G2* showed the highest transcript accumulation in the leaves, and *OsaCYP93G1* showed the highest accumulation in the roots and shoots (Fig 7D). In general, the expression patterns of closely grouped paralogs differed, suggesting that they had similar functions at different stages of plant development.

Expression profiles of plant CYP93 genes in response to biotic and abiotic stresses

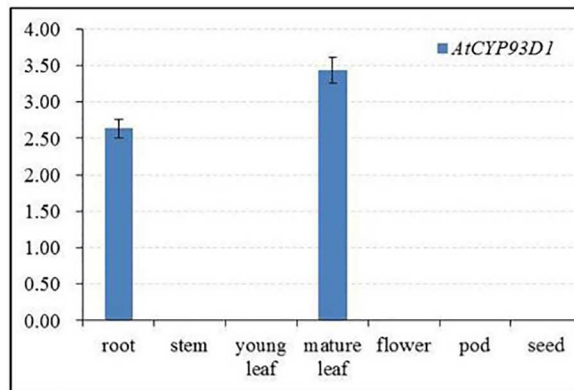
CYP93 genes are generally involved in plant secondary metabolism, which is important for plant defense responses, such as flavonoid and isoflavonoid compounds. To reveal the possible roles of CYP93 genes in stress responses, we analyzed their expression profiles in response to different biotic and abiotic stresses by using the publicly available global stress expression datasets in AtGenExpress [45] and PLEXdb [31].

On the basis of the expression datasets in AtGenExpress [45], we first analyzed the expression patterns of *AtCYP93D1* under abiotic stress. *AtCYP93D1* was found to be preferentially expressed in the roots and only weakly expressed in the shoots (Fig 8A); it was strongly expressed in response to UV-B, cold, and heat treatments. In soybean, we identified eight probes—*GmaCYP93A1*, *GmaCYP93A2*, *GmaCYP93A3*, *GmaCYP93A19*, *GmaCYP93A41*, *GmaCYP93C1v2*, *GmaCYP93C5*, and *GmaCYP93E1*—corresponding to a single CYP93 gene (Fig 8B). However, we identified no probe set corresponding to the remaining five investigated genes. Hence, we used the eight target genes in our subsequent analysis. The expression levels of *GmaCYP93A1*, *GmaCYP93A2*, *GmaCYP93A3*, *GmaCYP93A41*, *GmaCYP93C1v2*, and *GmaCYP93C5* were markedly upregulated after NaHCO_3 treatment, whereas those of *GmaCYP93A1*, *GmaCYP93A2*, *GmaCYP93A3*, and *GmaCYP93A41* were upregulated under conditions of magnesium stress, but suppressed in response to aluminum plus magnesium stress for 72 h. The expression levels of *GmaCYP93C5* and *GmaCYP93A3* were suppressed by heat and salinity stress, respectively; however, those of *GmaCYP93A1*, *GmaCYP93A2*, *GmaCYP93A41*, and *GmaCYP93C1v2* were upregulated in response to these stresses (Fig 8B). In rice, the expression level of *OsaCYP93G1* was suppressed by drought, salt, and cold, whereas that of *OsaCYP93G2* was suppressed by drought and cold. The expression level of *OsaCYP93G1* was suppressed whereas that of *OsaCYP93G2* was upregulated by heat. *OsaCYP93G1* and *OsaCYP93G2* were suppressed by anoxic stress, whereas they were upregulated under conditions of Pi deficiency for 6 h and 24 h, respectively. *OsaCYP93F1* showed weak and relatively consistent expression levels in response to all the investigated stresses (Fig 8C).

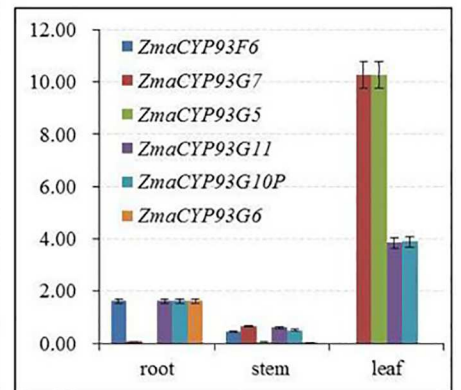
Similarly, to elucidate the possible roles of CYP93 genes in biotic stress responses (pathogen and insect stress), we used the PLEXdb and AtGenExpress databases to determine their expression levels after exposure to different microbial and insect pathogens. However, the expression levels of *AtCYP93D1*, *OsaCYP93G1*, *OsaCYP93G2*, and *OsaCYP93F1* were not markedly



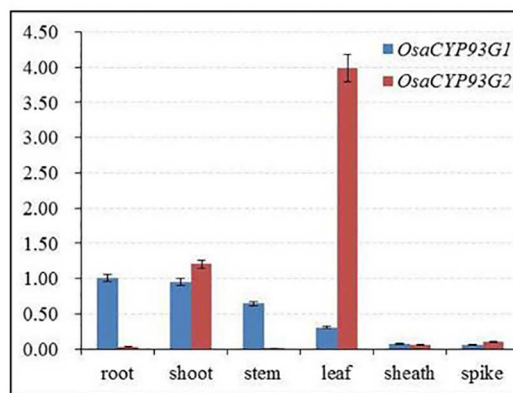
A



B



C



C

Fig 7. qRT-PCR analyses of the expression profiles of CYP93 homologous genes in *Arabidopsis*, soybean, rice, and maize. (A) Expression profiles of CYP93 genes in soybean. (B) Expression profiles of the *AtCYP93D1* gene in *Arabidopsis*. (C) Expression profiles of CYP93 genes in maize. (D) Expression profiles of CYP93 genes in rice.

doi:10.1371/journal.pone.0165020.g007

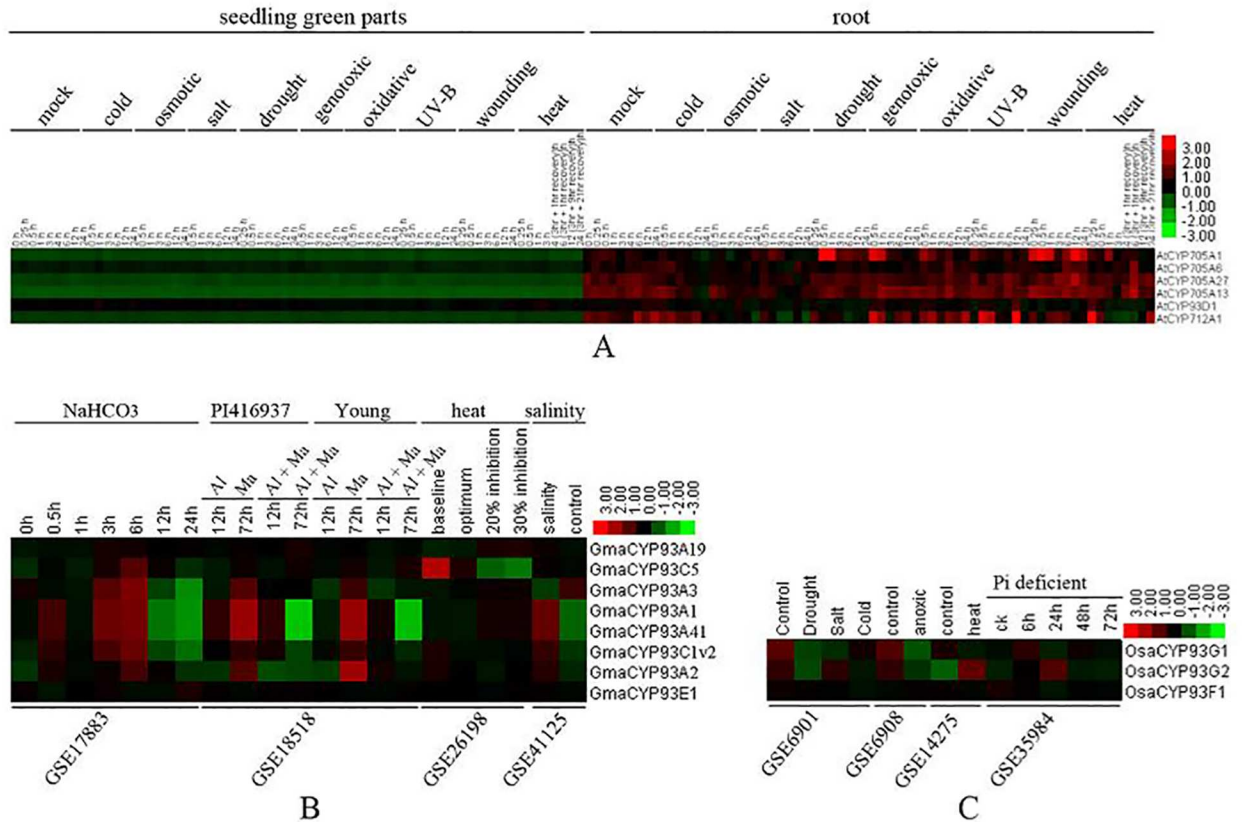


Fig 8. Expression profiles of plant CYP93 genes in response to abiotic stresses. (A) Expression profiles of *AtCYP93* and representative P450 genes in response to abiotic stresses. (B) Expression profiles of eight probe sets representing eight soybean CYP93 genes based on four microarray datasets of abiotic stresses. (C) Expression profiles of rice CYP93 genes based on four microarray datasets of abiotic stresses. Color bar at the base represents log₂ expression values.

doi:10.1371/journal.pone.0165020.g008

upregulated after exposure to pathogen and insect stresses (data not shown). However, with the exception of *GmaCYP93E1*, the expression levels of soybean CYP93 genes were upregulated after exposure to rust disease (Fig 9A). Moreover, with the exception of *GmaCYP93A19* and *GmaCYP93E1*, the expression levels of soybean CYP93 genes were markedly upregulated after *Phytophthora sojae* infection, with comparable expression profiles across the four genotypes (Fig 9B). Similarly, except for *GmaCYP93A19* and *GmaCYP93E1*, the expression levels of soybean CYP93 genes were also upregulated after soybean aphid infestation (Fig 9C).

Taken together, our results suggested that the expression of plant CYP93 genes is upregulated in response to certain biotic and abiotic stresses, implying that the corresponding pathways or metabolites (e.g., isoflavonoids) that they regulate might be important for plant stress responses.

Discussion

Classification and evolution of CYP93 genes

In the present study, we analyzed the distributions of CYP93 genes in 60 plant genomes, including four green algae, one moss, one lycophyte, one gymnosperm, and 53 angiosperms. We generated a highly supported tree of candidate CYP93 proteins (Fig 2). The classical classification of the P450 family uses protein identity as the main criterion [1]. Our phylogenetic

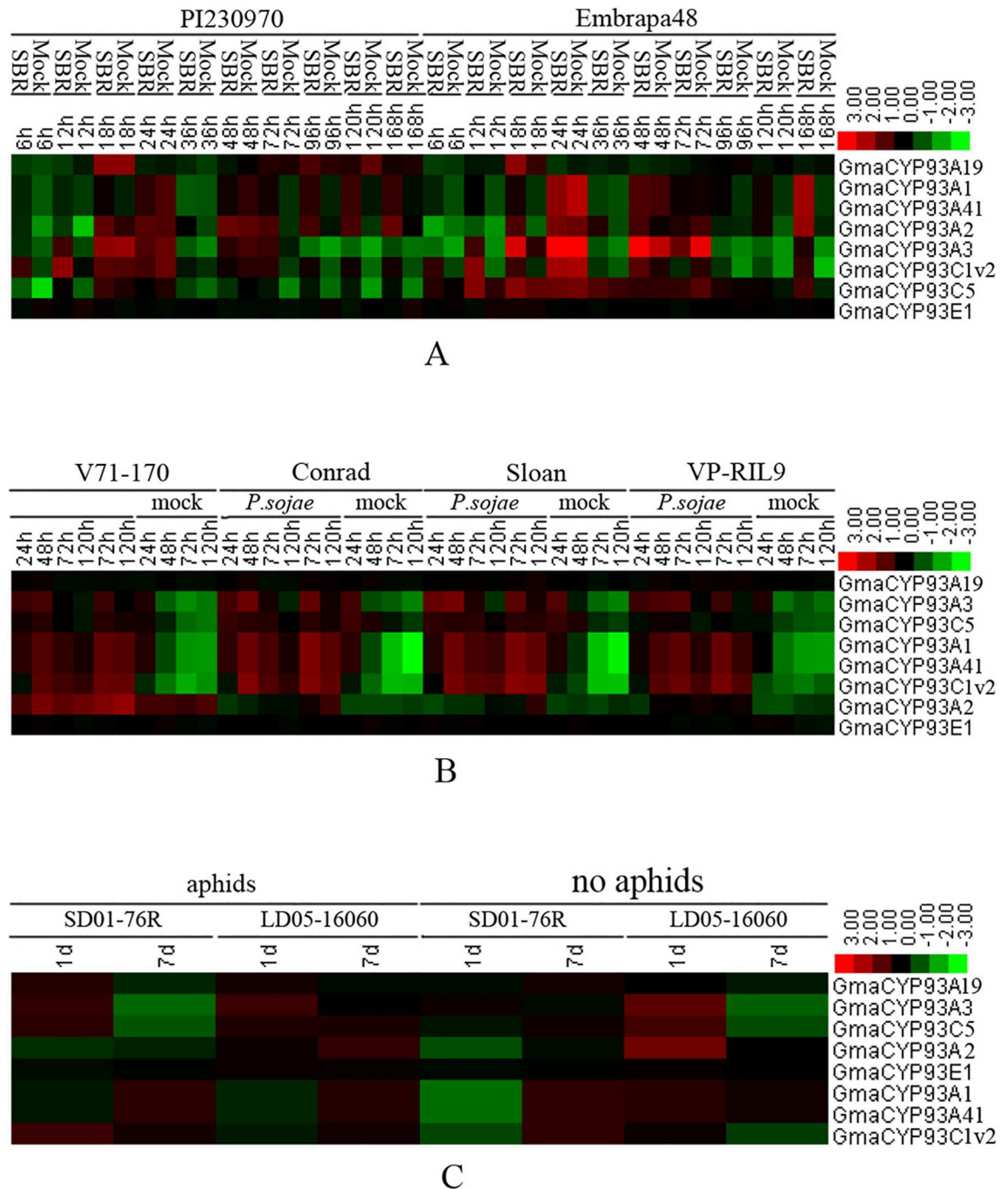


Fig 9. Expression profiles of *GmCYP93* genes in response to biotic stresses. (A) Expression profiles of *GmCYP93* genes after infection with root-knot nematode (GSE33410). (B) Expression profiles of *GmCYP93* genes after infection with *Phytophthora sojae* (GSE9687). (C) Expression profiles of *GmCYP93* genes after aphid infestation (GSE35427). Color bar at the base represents log₂ expression values.

doi:10.1371/journal.pone.0165020.g009

analysis tended to support this classification; however, some aspects are not in accordance with previous findings. In particular, several classified groups are not in parallel, e.g., CYP93D and CYP93C are embedded in CYP93A and CYP93B, respectively. These incongruities can be explained by the very recent origin of the new clade (subfamily CYP93C and CYP93D) in a

specific lineage (such as Brassicaceae and Leguminosae) that share high sequence identity with the ancestor genes, and thus the phylogenetic tree lacked high resolution.

Based on the distributions and phylogeny, we postulate that all currently identified CYP93 genes are derived from a process involving gene duplications and diversification from a single progenitor molecule in the distant past of the last common ancestor of flowering plants. Considering that *Amborella trichopoda* contains two CYP93A candidates (Fig 1 and S2 Table), this gene family might have evolved after the split of angiosperms and gymnosperms and could have plant-specific flowering. However, since only the genome of *Picea abies* is currently available, this can be further verified only after more genomes can be used in this lineage. Thus, our results showed that the origin of the CYP93 family in land plants can be traced to the origin and early diversification of flowering plants. Moreover, CYP93A is likely the most ancient subfamily, because it has the broadest distribution across angiosperms and is the only CYP93 subfamily that occurs in both monocots and eudicots (Figs 1 and 2). The remaining CYP93 subfamilies are further subdivided into two groups—the monocot group (CYP93F, CYP93G, and CYP93J) and eudicot group (CYP93B, CYP93C, CYP93D, CYP93E, CYP93H, and CYP93K). Among these sub-families, subfamily CYP93B is relatively older and was present in the stem lineage of dicots (Fig 1). CYP93B is further subdivided into two clear monophyly clades, F2H and FNS II, each with high statistical support (82% and 56%; Fig 2). CYP93C is embedded in the CYP93B subfamily and forms a monophyly clade with a high bootstrap value (98%), which is more closely related to the F2H clade. These results supported that CYP93B and CYP93C share the same origin, where CYP93C is likely duplicated and/or retained from CYP93B after the divergence of legumes and other Fabids. This finding is strongly corroborated by independent evidence that CYP93B and CYP93C share the same substrate—(2S)-flavanone [13,14,18,20,44]. Monocots only contain CYP93G and/or CYP93F genes, and CYP93G might be older since its members are distributed in a relatively wider scope of species (Fig 1). Interestingly, several members of CYP93G were found to share the same substrate (flavanone) with CYP93B and CYP93C subfamilies [20–22]. This might suggest that the monocot and eudicot groups have a similar role in the conversion of flavanone to flavone.

To further elucidate the origin of CYP93 genes in P450 super gene family, we constructed a NJ phylogenetic tree by using the representative genes of each CYP93 subfamily and other plant P450 families, based on sequence alignment analysis (S4 Fig). Considering that the non-A type is very ancient and CYP51 protein is the only P450 family conserved across fungi, animals, and plants [11], we rooted the NJ tree on the representative of *Arabidopsis* CYP51G1. Our results showed that the CYP93 genes share a relatively high degree of sequence similarity with the CYP705 and CYP712 families (data not show) and cluster close to these two families in the phylogenetic tree, confirming that CYP93, CYP705, and CYP712 share a last common ancestor and CYP93 genes have evolved from CYP705 and/or CYP712 during the evolution [11].

Functions of CYP93 genes

To date, some studies have focused on the narrow range of CYP93 genes, including CYP93A, CYP93B, CYP93C, CYP93E, and CYP93G (Table 1). Nearly all of the CYP93 family members functionally characterized thus far are involved in flavonoid/isoflavonoid biosynthesis, although the functions of CYP93D, CYP93F, CYP93H, CYP93K, and CYP93J are yet not known. For example, CYP93B is a known flavone synthase that produces flavones [17–18,20]; similarly, CYP93G1 and CYP93G2 in rice constitute flavone synthase II (FNS II) and catalyze the direct conversion of flavanones to flavones [20–22], and CYP93C is involved in isoflavonoid biosynthesis [13–16,49]. The exception is CYP93A and CYP93E subfamilies, where

Table 1. Summary of functionally characterized CYP93 genes in plants.

Subfamily	Name	Species	Biological Function	Reference
CYP93A	CYP93A1	soybean	pterocarpanoid phytoalexin biosynthesis	[50]
	CYP93A2	soybean	no known	[60]
CYP93B	CYP93B2	Gerbera	FNSII activity in flavone biosynthesis	[56]
	CYP93B1	licorice	F2H activities in licodione and flavone formation	[59]
	CYP90B10	<i>Medicago truncatula</i>	F2H activities that has a catalytic mechanism similar to FNS II	[17]
	CYP90B11	<i>Medicago truncatula</i>	F2H activities that has a catalytic mechanism similar to FNS II	[17]
	CYP93B16	soybean	direct conversion of flavanones into flavones	[18]
	FNSII-2.1	<i>Lonicera japonica</i>	direct conversion of flavanones into flavones	[61]
	FNSII-1.1	<i>Lonicera japonica</i>	direct conversion of flavanones into flavones	[61]
	FNSII-1.1	<i>Lonicera japonica</i>	direct conversion of flavanones into flavones	[61]
	CYP93B13		modulate the intensity of flower pigmentation	[62]
	CYP93B6		direct conversion of flavanones into flavones	[58]
CYP93C	CYP93C3	<i>Cicer arietinum</i>	no known	[63]
	CYP93C5	soybean	involved in isoflavone biosynthesis	[64]
	CYP93C1	soybean	involved in isoflavone biosynthesis	[64]
	CYP93C1v2	soybean	catalyzing the aryl migration	[65]
	CYP93C1v1	soybean	promoting nodulation gene expression in rhizobia	[66]
	CYP93C17	<i>Lotus japonicus</i>	involved in phytoalexin biosynthesis	[67]
	CYP93C2	<i>Glycyrrhiza echinata</i>	involved in isoflavonoid skeleton biosynthesis	[57]
	CYP93C18	<i>Pisum sativum</i>	involved in isoflavone phytoalexin, pisatin isoflavone phytoalexin, pisatin	[14,68]
	IFS2_12	<i>Trifolium repens</i>	involved in isoflavonoid genistein and its conjugates biosynthesis	[19]
CYP93E	CYP93E2	<i>Medicago truncatula</i>	involved in triterpenoid saponins biosynthesis	[69,23]
CYP93G	CYP93G3	sorghum	converting flavanones to flavone	[21]
	CYP93G5	maize	converts flavanones to 2-hydroxy derivatives served as substrates for C-glycosylation	[70]
	CYP93G1	Rice	channeling flavanones to the biosynthesis of triclin O-linked conjugates	[22]
	CYP93G2	Rice	convert naringenin and eriodictyol to the respective 2-hydroxyflavanones.	[20]
	CYP93G5	maize	a functional F2H under the genetic control of P1	[70–71]

doi:10.1371/journal.pone.0165020.t001

CYP93A is involved in pterocarpanoid phytoalexin biosynthesis in soybean [50], whereas CYP93E catalyzes the C-24 oxidation of the triterpene backbone in triterpenoid saponin biosynthesis [23].

Considering the functional conservation of CYP93B, CYP93C, and CYP93G homologous genes across different species (Table 1), as well as the high sequence conservation of CYP93 proteins in a given subfamily, although the roles of most CYP93 genes remain to be elucidated, we speculate that members of a subfamily might have recent common evolutionary origins and also a conserved function catalyzing the same metabolic reaction. For instance, flavones are ubiquitous secondary metabolites in plants; they are important compounds for the biochemistry and physiology of plants and human nutrition and health [51–54]. One of the two key flavone synthases, FNS I and FNS II, that catalyze the conversion of flavanones to flavones belong to the CYP93B subfamily [54–55]. To date, two subgroups have been characterized within the CYP93B gene family: one subgroup (FNS II) directly converts the flavanone substrates into flavones, such as CYP93B2 [56], CYP93B3 [57], and CYP93B6 [58]; conversely, the second (F2H) acts as flavanone 2-hydroxylases (CYP93B1, CYP93B10, and CYP93B11) that indirectly convert flavanone to flavones [59]. Consistent with the findings of these studies, our phylogenetic NJ tree results suggested that the flavanone 2-hydroxylases (F2Hs) are clustered into a clear clade separated from all other members of the CYP93B genes (FNS II; Fig 2). Interestingly, the

monocot-specific clade included CYP93G subfamilies such as CYP93G1 and CYP93G2 in rice [20–22] and CYP93G3 in sorghum [21]; they share the same catalytic mechanism with the CYP93B subfamily by converting flavanones to flavones. Despite the large number of CYP93A genes as well as CYP93D and CYP93E genes that are embedded in or close to CYP93A genes, these genes were found to not contribute to flavonoid/isoflavonoid biosynthesis [50]; instead, the flavone synthase activity might have arisen after the evolution of CYP93B-K (Table 1). Thus, our results suggested that a CYP93 protein containing flavone synthase activity might be the ancestor of most of the present day CYP93 subfamilies, except CYP93A and CYP93E, which suggests monocot/eudicot divergence. Although the roles of CYP93D, CYP93F, CYP93H, CYP93J, and CYP93K remain to be elucidated, the CYP93 genes clustered in a subfamily shared similar gene architecture structure and functions, indicating recent common evolutionary origins and also conserved function. The knowledge of the functions of certain members should facilitate the confirmation of functional relationships among homologous genes.

Notably, despite the highly conserved sequence of CYP93 genes between subfamilies, significant functional specification exists. However, our selection analysis only revealed strong negative selection on the aligned coding region. Thus, the subfunctionalization of CYP93 gene family might not be derived from the mutation of active sites or active surrounding sites. In addition, we identified many subfamily-specific conserved insertions and deletions in some locations (Fig 3), implying a possible role in subfunctionalization. Moreover, considering the presence of many relaxed constraint sites in the less conserved region, they might also be important for the enzyme structure and contribute to the functional specification (Fig 3).

In addition, the expression pattern of a gene is often correlated with its function; the expression analysis might also provide important information regarding gene functions. Our expression analyses revealed that genes in the same subfamily across different species tend to have similar expression profiles (Figs 6–9), suggesting that the homologous CYP93 genes in a given subfamily share similar functions. For instance, most of the CYP93A genes have the same expression profile in the roots: CYP93C genes are mainly expressed in the roots. In *Arabidopsis* and soybean, CYP93 genes were expressed only or primarily in the roots, whereas, in monocots, they were expressed in many other tissues and organs (Figs 6 and 7). Hence, the common ancestor CYP93 might be widely expressed, but might be functional expressed only in the roots of some dicot plants such as *Arabidopsis*.

Supporting Information

S1 Fig. Phylogenetic tree of CYP93 proteins of the plant P450 superfamily. The maximum likelihood (ML) tree includes all of the 214 CYP93 candidate proteins used in the present study. Bootstrap values <50 are not shown.

(PDF)

S2 Fig. Alignment of plant CYP93 proteins used in this study. MAFFT software was used for amino acid sequence alignment of the 214 CYP93 proteins. The shading of the alignment represents different degrees of conservation among sequences. The P450 motifs and substrate recognition sites are underlined.

(PDF)

S3 Fig. Modeled structure of CYP93 of each group. The overall structures of CYP93 are very similar. Red spheres represent conserved insertion of B1 and C. The model was built using SWISS-MODEL (<http://swissmodel.expasy.org/>) and viewed using PyMOL software.

(PDF)

S4 Fig. Phylogenetic tree of CYP93 proteins and representative of each family of the plant P450 superfamily. The neighbor-joining (NJ) tree includes the the representatives of the 10 CYP93 subfamilies and 57 representatives of other plant P450 families. Bootstrap values <50 are not shown. CYP51 is the root of the tree.

(PDF)

S1 Table. The query sequences used in the identification of CYP93 proteins in each species.

(XLS)

S2 Table. The species used in this study and the corresponding CYP93 candidate genes identified.

(XLSX)

S3 Table. Primers used in this study.

(XLSX)

S4 Table. The identity and similarity analysis of plant CYP93 proteins.

(XLSX)

S5 Table. The selection analysis of different branches of CYP93 gene family.

(XLSX)

Author Contributions

Conceptualization: HD.

Data curation: HD ZL.

Formal analysis: HD ZL.

Funding acquisition: HD JNL.

Investigation: HD JNL.

Methodology: HD FR HLR JW.

Project administration: HD ZL JNL.

Resources: HD JNL.

Software: ZL.

Supervision: HD ZL JNL.

Validation: FR ZL.

Visualization: HD.

Writing – original draft: HD ZL.

Writing – review & editing: HD ZL.

References

1. Nelson DR, Schuler MA, Paquette SM, Werck-Reichhart D, Bak S. Comparative genomics of rice and *Arabidopsis*. Analysis of 727 cytochrome P450 genes and pseudogenes from a monocot and a dicot. *Plant Physiol.* 2004; 135 (2):756–772. doi: [10.1104/pp.104.039826](https://doi.org/10.1104/pp.104.039826) PMID: [15208422](https://pubmed.ncbi.nlm.nih.gov/15208422/)
2. Nelson DR. The cytochrome p450 homepage. *Hum Genomics.* 2009; 4(1):59–65. PMID: [19951895](https://pubmed.ncbi.nlm.nih.gov/19951895/)
3. Bak S, Beisson F, Bishop G, Hamberger B, Hofer R, Paquette S, et al. Cytochromes P450. 2011; *Arabidopsis Book* 9, e0144. doi: [10.1199/tab.0144](https://doi.org/10.1199/tab.0144) PMID: [22303269](https://pubmed.ncbi.nlm.nih.gov/22303269/)

4. Graham SE, Peterson JA. How similar are P450s and what can their differences teach us? *Arch Biochem Biophys.* 1999; 369(1):24–29. doi: [10.1006/abbi.1999.1350](https://doi.org/10.1006/abbi.1999.1350) PMID: [10462437](https://pubmed.ncbi.nlm.nih.gov/10462437/)
5. Werck-Reichhart D, Feyereisen R. Cytochromes P450: a success story. *Genome Biol.* 2000; 1(6):3003.
6. Ehltling J, Sauveplane V, Olry A, Ginglinger JF, Provart NJ, Werck-Reichhart D. An extensive (co-) expression analysis tool for the cytochrome P450 superfamily in *Arabidopsis thaliana*. *BMC Plant Biol.* 2008; 8:47. doi: [10.1186/1471-2229-8-47](https://doi.org/10.1186/1471-2229-8-47) PMID: [18433503](https://pubmed.ncbi.nlm.nih.gov/18433503/)
7. Nelson D, Werck-Reichhart D. A P450-centric view of plant evolution. *Plant J.* 2011; 66(1):194–211. doi: [10.1111/j.1365-313X.2011.04529.x](https://doi.org/10.1111/j.1365-313X.2011.04529.x) PMID: [21443632](https://pubmed.ncbi.nlm.nih.gov/21443632/)
8. Guttikonda SK, Trupti J, Bisht NC, Chen H, An YQ, Pandey S, et al. Whole genome co-expression analysis of soybean cytochrome P450 genes identifies nodulation-specific P450 monooxygenases. *BMC Plant Biol.* 2010; 10:243. doi: [10.1186/1471-2229-10-243](https://doi.org/10.1186/1471-2229-10-243) PMID: [21062474](https://pubmed.ncbi.nlm.nih.gov/21062474/)
9. Durst F, Nelson DR. Diversity and evolution of plant P450 and P450-reductases. *Drug Metab. Drug Interact.* 1995; 12:189–206.
10. Paquette SM, Bak S, Feyereisen R. Intron-exon organization and phylogeny in a large superfamily, the paralogous cytochrome P450 genes of *Arabidopsis thaliana*. *DNA Cell Biol.* 2000; 19(5):307–317. doi: [10.1089/10445490050021221](https://doi.org/10.1089/10445490050021221) PMID: [10855798](https://pubmed.ncbi.nlm.nih.gov/10855798/)
11. Nelson DR. Cytochrome P450 nomenclature, 2004. *Methods Mol Biol.* 2006; 320:1–10. doi: [10.1385/1-59259-998-2:1](https://doi.org/10.1385/1-59259-998-2:1) PMID: [16719369](https://pubmed.ncbi.nlm.nih.gov/16719369/)
12. Du H, Huang Y, Tang Y. Genetics and Metabolic Engineering of Isoflavonoid Biosynthesis. *Appl Microbiol Biotechnol.* 2010; 86(5):1293–1312. doi: [10.1007/s00253-010-2512-8](https://doi.org/10.1007/s00253-010-2512-8) PMID: [20309543](https://pubmed.ncbi.nlm.nih.gov/20309543/)
13. Sawada Y, Kinoshita K, Akashi T, Aoki T, Ayabe S. Key amino acid residues required for aryl migration catalysed by the cytochrome P450 2-hydroxyisoflavanone synthase. *Plant J.* 2002; 31(5):555–564. PMID: [12207646](https://pubmed.ncbi.nlm.nih.gov/12207646/)
14. Cooper L, Doss R, Price R, Peterson K, Oliver J. Application of Bruchin B to pea pods results in the up-regulation of CYP93C18, a putative isoflavone synthase gene, and an increase in the level of pisatin, an isoflavone phytoalexin. *J Exp Bot.* 2005; 56(414):1229–1237. doi: [10.1093/jxb/eri117](https://doi.org/10.1093/jxb/eri117) PMID: [15753113](https://pubmed.ncbi.nlm.nih.gov/15753113/)
15. Chang Z, Wang X, Wei R, Liu Z, Shan H, Fan G, et al. Functional Expression and Purification of CYP93C20, a Plant Membrane-Associated Cytochrome P450 from *Medicago truncatula*. *Protein Expr Purif.* 2010; doi: [10.1016/j.pep.2010.11.012](https://doi.org/10.1016/j.pep.2010.11.012) PMID: [21138770](https://pubmed.ncbi.nlm.nih.gov/21138770/)
16. Waki T, Yoo D, Fujino N, Mameda R, Denessiouk K, Yamashita S, et al. Identification of protein-protein interactions of isoflavonoid biosynthetic enzymes with 2-hydroxyisoflavanone synthase in soybean (*Glycine max* (L.) Merr.). *Biochem Biophys Res Commun.* 2016; 469(3):546–51. doi: [10.1016/j.bbrc.2015.12.038](https://doi.org/10.1016/j.bbrc.2015.12.038) PMID: [26694697](https://pubmed.ncbi.nlm.nih.gov/26694697/)
17. Zhang J, Subramanian S, Zhang Y, Yu O. Flavone synthases from *Medicago truncatula* are flavanone-2-hydroxylases and are important for nodulation. *Plant Physiol.* 2007; 144(2):741–751. doi: [10.1104/pp.106.095018](https://doi.org/10.1104/pp.106.095018) PMID: [17434990](https://pubmed.ncbi.nlm.nih.gov/17434990/)
18. Fliegmann J, Furtwängler K, Malterer G, Cantarello C, Schüller G, Ebel J, et al. Flavone synthase II (CYP93B16) from soybean (*Glycine max* L.). *Phytochemistry.* 2010; 71(5–6):508–514. doi: [10.1016/j.phytochem.2010.01.007](https://doi.org/10.1016/j.phytochem.2010.01.007) PMID: [20132953](https://pubmed.ncbi.nlm.nih.gov/20132953/)
19. Franzmayr BK, Rasmussen S, Fraser KM, Jameson PE. Expression and functional characterization of a white clover isoflavone synthase in tobacco. *Ann Bot.* 2012; 110(6):1291–1301. doi: [10.1093/aob/mcs168](https://doi.org/10.1093/aob/mcs168) PMID: [22915577](https://pubmed.ncbi.nlm.nih.gov/22915577/)
20. Du Y, Chu H, Chu IK, Lo C. CYP93G2 is a flavanone 2-hydroxylase required for C-glycosylflavone biosynthesis in rice. *Plant Physiol.* 2010; 154(1):324–333. doi: [10.1104/pp.110.161042](https://doi.org/10.1104/pp.110.161042) PMID: [20647377](https://pubmed.ncbi.nlm.nih.gov/20647377/)
21. Du Y, Chu H, Wang M, Chu IK, Lo C. Identification of flavone phytoalexins and a pathogen-inducible flavone synthase II gene (SbFNSII) in sorghum. *J. Exp. Bot.* 2010; 61(4):983–994. doi: [10.1093/jxb/erp364](https://doi.org/10.1093/jxb/erp364) PMID: [20007684](https://pubmed.ncbi.nlm.nih.gov/20007684/)
22. Lam PY, Zhu FY, Chan WL, Liu H, Lo C. Cytochrome P450 93G1 Is a Flavone Synthase II That Channels Flavanones to the Biosynthesis of Tricin O-Linked Conjugates in Rice. *Plant Physiol.* 2014; 165(3):1315–1327. doi: [10.1104/pp.114.239723](https://doi.org/10.1104/pp.114.239723) PMID: [24843076](https://pubmed.ncbi.nlm.nih.gov/24843076/)
23. Moses T, Thevelein JM, Goossens A, Pollier J. Comparative analysis of CYP93E proteins for improved microbial synthesis of plant triterpenoids. *Phytochemistry.* 2014; 108:47–56. doi: [10.1016/j.phytochem.2014.10.002](https://doi.org/10.1016/j.phytochem.2014.10.002) PMID: [25453910](https://pubmed.ncbi.nlm.nih.gov/25453910/)
24. Xu W, Bak S, Decker A, Paquette SM, Feyereisen R, Galbraith DW. Microarray-based analysis of gene expression in very large gene families: the cytochrome P450 gene superfamily of *Arabidopsis thaliana*. *Gene.* 2001; 272(1–2):61–74. PMID: [11470511](https://pubmed.ncbi.nlm.nih.gov/11470511/)

25. Nelson DR, Ming R, Alam M, Schuler AM. Comparison of Cytochrome P450 Genes from Six Plant Genomes. *Tropical Plant Biol.* 2008; doi: [10.1007/s12042-008-9022-1](https://doi.org/10.1007/s12042-008-9022-1)
26. Nelson and Schuler. Cytochrome P450 Genes from the Sacred Lotus Genome. *Tropical Plant Biol.* 2013; doi: [10.1007/s12042-013-9119-z](https://doi.org/10.1007/s12042-013-9119-z)
27. Goodstein DM, Shu S, Howson R, Neupane R, Hayes RD, Fazo J, et al. Phytosome: a comparative platform for green plant genomics. *Nucleic Acids Res.* 2012; 40:D1178–1186. doi: [10.1093/nar/gkr944](https://doi.org/10.1093/nar/gkr944) PMID: [22110026](https://pubmed.ncbi.nlm.nih.gov/22110026/)
28. Lee TH, Tang H, Wang X, Paterson AH. PGDD: a database of gene and genome duplication in plants. *Nucleic Acids Res.* 2013; 41:D1152–1158. doi: [10.1093/nar/gks1104](https://doi.org/10.1093/nar/gks1104) PMID: [23180799](https://pubmed.ncbi.nlm.nih.gov/23180799/)
29. Katoh K, Toh H. Parallelization of the MAFFT multiple sequence alignment program. *Bioinformatics.* 2010; 26(15):1899–1900. doi: [10.1093/bioinformatics/btq224](https://doi.org/10.1093/bioinformatics/btq224) PMID: [20427515](https://pubmed.ncbi.nlm.nih.gov/20427515/)
30. Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Mol Biol Evol.* 2013; 30(12):2725–2729. doi: [10.1093/molbev/mst197](https://doi.org/10.1093/molbev/mst197) PMID: [24132122](https://pubmed.ncbi.nlm.nih.gov/24132122/)
31. Dash S, Van Hemert J, Hong L, Wise RP, Dickerson JA. PLEXdb: gene expression resources for plants and plant pathogens. *Nucleic Acids Res.* 2012; 40:D1194–1201. doi: [10.1093/nar/gkr938](https://doi.org/10.1093/nar/gkr938) PMID: [22084198](https://pubmed.ncbi.nlm.nih.gov/22084198/)
32. De Hoon MJ, Imoto S, Nolan J, Miyano S. Open source clustering software. *Bioinformatics.* 2004; 20:1453–1454. doi: [10.1093/bioinformatics/bth078](https://doi.org/10.1093/bioinformatics/bth078) PMID: [14871861](https://pubmed.ncbi.nlm.nih.gov/14871861/)
33. Saldanha AJ. Java Treeview—extensible visualization of microarray data. *Bioinformatics.* 2004; 20:3246–3248. doi: [10.1093/bioinformatics/bth349](https://doi.org/10.1093/bioinformatics/bth349) PMID: [15180930](https://pubmed.ncbi.nlm.nih.gov/15180930/)
34. Du H, Yang SS, Liang Z, Feng BR, Liu L, Huang YB, et al. Genome-wide analysis of the MYB transcription factor superfamily in soybean. *BMC Plant Biol.* 2012; 12:106. doi: [10.1186/1471-2229-12-106](https://doi.org/10.1186/1471-2229-12-106) PMID: [22776508](https://pubmed.ncbi.nlm.nih.gov/22776508/)
35. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol.* 2011; 28(10):2731–2739. doi: [10.1093/molbev/msr121](https://doi.org/10.1093/molbev/msr121) PMID: [21546353](https://pubmed.ncbi.nlm.nih.gov/21546353/)
36. Pond SL, Frost SD, Muse SV. HyPhy: hypothesis testing using phylogenies. *Bioinformatics.* 2005; 21(5): 676–679. doi: [10.1093/bioinformatics/bti079](https://doi.org/10.1093/bioinformatics/bti079) PMID: [15509596](https://pubmed.ncbi.nlm.nih.gov/15509596/)
37. Biasini M, Bienert S, Waterhouse A, Arnold K, Studer G, Schmidt T, et al. SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. *Nucleic Acids Res.* 2014; 42(W1):W252–W258.
38. Nelson DR, Koymans L, Kamataki T, Stegeman JJ, Feyereisen R, Waxman DJ, et al. P450 superfamily: update on new sequences, gene mapping, accession numbers and nomenclature. *Pharmacogenetics.* 1996; 6(1):1–42. PMID: [8845856](https://pubmed.ncbi.nlm.nih.gov/8845856/)
39. Gotoh O. Substrate recognition sites in cytochrome P450 family 2 (CYP2) proteins inferred from comparative analyses of amino acid and coding nucleotide sequences. *J Biol Chem.* 1992; 267(1):83–90. PMID: [1730627](https://pubmed.ncbi.nlm.nih.gov/1730627/)
40. Domanski TL, Halpert JR. Analysis of mammalian cytochrome P450 structure and function by site-directed mutagenesis. *Curr Drug Metab.* 2001; (2):117–137. PMID: [11469721](https://pubmed.ncbi.nlm.nih.gov/11469721/)
41. Dutartre L, Hilliou F, Feyereisen R. Phylogenomics of the benzoxazinoid biosynthetic pathway of Poaceae: gene duplications and origin of the Bx cluster. *BMC Evol Biol.* 2012; 12:64. doi: [10.1186/1471-2148-12-64](https://doi.org/10.1186/1471-2148-12-64) PMID: [22577841](https://pubmed.ncbi.nlm.nih.gov/22577841/)
42. Johnson EF, Stout CD. Structural diversity of eukaryotic membrane cytochrome p450s. *J Biol Chem.* 2013; 288(24):17082–17090. doi: [10.1074/jbc.R113.452805](https://doi.org/10.1074/jbc.R113.452805) PMID: [23632020](https://pubmed.ncbi.nlm.nih.gov/23632020/)
43. Sawada Y, Ayabe S. Multiple mutagenesis of P450 isoflavonoid synthase reveals a key active-site residue. *Biochem Biophys Res Commun.* 2005; 330(3):907–913. PMID: [15809082](https://pubmed.ncbi.nlm.nih.gov/15809082/)
44. Sharp PA. Speculations on RNA splicing. *Cell.* 1981, 23:643–646. PMID: [7226224](https://pubmed.ncbi.nlm.nih.gov/7226224/)
45. Kilian J, Whitehead D, Horak J, Wanke D, Weinl S, Batistic O, et al. The AtGenExpress global stress expression data set: protocols, evaluation and model data analysis of UV-B light, drought and cold stress responses. *Plant J.* 2007; 50(2):347–363. PMID: [17376166](https://pubmed.ncbi.nlm.nih.gov/17376166/)
46. Severin AJ, Woody JL, Bolon YT, Joseph B, Diers BW, Farmer AD, et al. RNA-Seq Atlas of *Glycine max*: a guide to the soybean transcriptome. *BMC Plant Biol.* 2010; 10:160. doi: [10.1186/1471-2229-10-160](https://doi.org/10.1186/1471-2229-10-160) PMID: [20687943](https://pubmed.ncbi.nlm.nih.gov/20687943/)
47. Libault M, Farmer A, Joshi T, Takahashi K, Langley RJ, Franklin LD, et al. An integrated transcriptome atlas of the crop model *Glycine max*, and its use in comparative analyses in plants. *Plant J.* 2010; 63(1):86–99. doi: [10.1111/j.1365-313X.2010.04222.x](https://doi.org/10.1111/j.1365-313X.2010.04222.x) PMID: [20408999](https://pubmed.ncbi.nlm.nih.gov/20408999/)

48. Sekhon RS, Lin H, Childs KL, Hansey CN, Buell CR, de LN, et al. Genome-wide atlas of transcription during maize development. *Plant J.* 2011; 66:553–563. doi: [10.1111/j.1365-313X.2011.04527.x](https://doi.org/10.1111/j.1365-313X.2011.04527.x) PMID: [21299659](https://pubmed.ncbi.nlm.nih.gov/21299659/)
49. Dastmalchi M, Bernards M, Dhaubhadel S. Twin anchors of the soybean isoflavonoid metabolon: evidence for tethering of the complex to the endoplasmic reticulum by IFS and C4H. *Plant J.* 2016; doi: [10.1111/tpj.13137](https://doi.org/10.1111/tpj.13137)
50. Schopfer CR, Kochs G, Lottspeich F, Ebel J. Molecular characterization and functional expression of dihydropterocarpan 6a-hydroxylase, an enzyme specific for pterocarpanoid phytoalexin biosynthesis in soybean (*Glycine max* L.). *FEBS Lett.* 1998; 432(3):182–186. PMID: [9720921](https://pubmed.ncbi.nlm.nih.gov/9720921/)
51. Fisher R.F., Long S.R. Rhizobium-plant signal exchange. *Nature.* 1992, 357: 655–660 PMID: [1614514](https://pubmed.ncbi.nlm.nih.gov/1614514/)
52. Simmonds MS. Flavonoid-insect interactions: recent advances in our knowledge. *Phytochemistry.* 2003; 64: 21–30. PMID: [12946403](https://pubmed.ncbi.nlm.nih.gov/12946403/)
53. Soriano I, Asenstorfer R, Schmidt O, Riley I. Inducible flavone in oats (*Avena sativa*) is a novel defense against plant-parasitic nematodes. *Phytopathology.* 2004; 94: 1207–1214. doi: [10.1094/PHYTO.2004.94.11.1207](https://doi.org/10.1094/PHYTO.2004.94.11.1207) PMID: [18944456](https://pubmed.ncbi.nlm.nih.gov/18944456/)
54. Martens S, Mithöfer A. Flavones and flavone synthases. *Phytochemistry.* 2005; 66: 2399–2407. PMID: [16137727](https://pubmed.ncbi.nlm.nih.gov/16137727/)
55. Britsch L. Purification and characterization of flavone synthase I, a 2-oxoglutarate-dependent desaturase. *Arch Biochem Biophys.* 1990; 276:348–354.
56. Martens S, Forkmann G. Cloning and expression of flavone synthase II from *Gerbera* hybrids. *Plant J.* 1999; 20(5):611–618. PMID: [10652133](https://pubmed.ncbi.nlm.nih.gov/10652133/)
57. Akashi T, Aoki T, Ayabe Si. Cloning and functional expression of a cytochrome P450 cDNA encoding 2-hydroxyisoflavanone synthase involved in biosynthesis of the isoflavonoid skeleton in licorice. *Plant Physiol.* 1999; 121(3):821–828. PMID: [10557230](https://pubmed.ncbi.nlm.nih.gov/10557230/)
58. Kitada C, Gong Z, Tanaka Y, Yamazaki M, Saito K. Differential expression of two cytochrome P450s involved in the biosynthesis of flavones and anthocyanins in chemovarietal forms of *Perilla frutescens*. *Plant Cell Physiol.* 2001; 42:1338–1344. PMID: [11773526](https://pubmed.ncbi.nlm.nih.gov/11773526/)
59. Akashi T, Aoki T, Ayabe S. Identification of a cytochrome P450 cDNA encoding (2S)-flavanone 2-hydroxylase of licorice (*Glycyrrhiza echinata* L.; Fabaceae) which represents licodione synthase and flavone synthase II. *FEBS Lett.* 1998; 431(2):287–290. PMID: [9708921](https://pubmed.ncbi.nlm.nih.gov/9708921/)
60. Suzuki G, Ohta H, Kato T., Igarashi T, Sakai F, Shibata D, et al. Induction of a novel cytochrome P450 (CYP93 family) by methyl jasmonate in soybean suspension-cultured cells. *FEBS Lett.* 1996; 383(1–2):83–86. PMID: [8612798](https://pubmed.ncbi.nlm.nih.gov/8612798/)
61. Wu J, Wang XC, Liu Y, Du H, Shu QY, Su S, et al. Flavone synthases from *Lonicera japonica* and *L. macranthoides* reveal differential flavone accumulation. *Sci Rep.* 2016; 6:19245. doi: [10.1038/srep19245](https://doi.org/10.1038/srep19245) PMID: [26754912](https://pubmed.ncbi.nlm.nih.gov/26754912/)
62. Nakatsuka T, Nishihara M, Mishiba K, Yamamura S. Heterologous expression of two gentian cytochrome P450 genes can modulate the intensity of flower pigmentation in transgenic tobacco plants. *Mol Breeding.* 2006; 17:91–99.
63. Overkamp S, Hein F, Barz W. Cloning and characterization of eight cytochrome P450 cDNAs from chickpea (*Cicer arietinum* L.) cell suspension cultures. *Plant Sci.* 2000; 155(1):101–108. PMID: [10773344](https://pubmed.ncbi.nlm.nih.gov/10773344/)
64. Jung W, Yu O, Lau SM, O'Keefe DP, Odell J, Fader G, et al. Identification and expression of isoflavone synthase, the key enzyme for biosynthesis of isoflavones in legumes. *Nat Biotechnol.* 2000; 18(2):208–212. PMID: [10657130](https://pubmed.ncbi.nlm.nih.gov/10657130/)
65. Steele CL, Gijzen M, Qutob D, Dixon RA. Molecular characterization of the enzyme catalyzing the aryl migration reaction of isoflavonoid biosynthesis in soybean. *Arch Biochem Biophys.* 1999; 367(1):146–150. PMID: [10375412](https://pubmed.ncbi.nlm.nih.gov/10375412/)
66. Sreevidya VS, Srinivasa RC, Sullia SB, Ladha JK, Reddy PM. Metabolic engineering of rice with soybean isoflavone synthase for promoting nodulation gene expression in rhizobia. *J Exp Bot.* 2006; 57(9):1957–1969. PMID: [16690627](https://pubmed.ncbi.nlm.nih.gov/16690627/)
67. Shimada N, Akashi T, Aoki T, Ayabe S. Induction of isoflavonoid pathway in the model legume *Lotus japonicus*: molecular characterization of enzymes involved in phytoalexin biosynthesis. *Plant Sci.* 2000; 160(1):37–47. PMID: [11164575](https://pubmed.ncbi.nlm.nih.gov/11164575/)
68. Picmanova M, Renak D, Fecikova J, Ruzicka P, Miksatkova P, Lapcik O, et al. Functional expression and subcellular localization of pea polymorphic isoflavone synthase CYP93C18. *Biologia Plantarum.* 2013; 57(4): 635. doi: [10.1007/s10535-013-0344-y](https://doi.org/10.1007/s10535-013-0344-y)

69. Fukushima EO, Seki H, Sawai S, Suzuki M, Ohyama K, Saito K, et al. Combinatorial biosynthesis of legume natural and rare triterpenoids in engineered yeast. *Plant Cell Physiol.* 2013; 54(5):740–749 doi: [10.1093/pcp/pct015](https://doi.org/10.1093/pcp/pct015) PMID: [23378447](https://pubmed.ncbi.nlm.nih.gov/23378447/)
70. Falcone Ferreyra ML, Rodriguez E, Casas MI, Labadie G, Grotewold E, Casati P. Identification of a bifunctional maize C- and O-glucosyltransferase. *J Biol Chem.* 2013; 288(44):31678–88. doi: [10.1074/jbc.M113.510040](https://doi.org/10.1074/jbc.M113.510040) PMID: [24045947](https://pubmed.ncbi.nlm.nih.gov/24045947/)
71. Morohashi K, Casas MI, Falcone ML, Mejía-Guerra MK, Pourcel L, Yilmaz A, et al. A genome-wide regulatory framework identifies maize pericarp color1 controlled genes. *Plant Cell.* 2012; 24(7):2745–2764. doi: [10.1105/tpc.112.098004](https://doi.org/10.1105/tpc.112.098004) PMID: [22822204](https://pubmed.ncbi.nlm.nih.gov/22822204/)