



## OPEN ACCESS

## EDITED BY

Pramod Prasad,  
ICAR-Indian Institute of Wheat and Barley  
Research, Regional Station, India

## REVIEWED BY

Ake Liu,  
Changzhi University, China  
Muhammad Anwar,  
Hainan University, China  
Owen Wally,  
Agriculture and Agri-Food Canada (AAFC),  
Canada

## \*CORRESPONDENCE

Archana Singh

✉ archanasingh@hrc.du.ac.in;

✉ archanasingh@pmb.du.ac.in

Amit Roy

✉ Roy@fld.czu.cz

Indrakant K. Singh

✉ iksingh@db.du.ac.in

RECEIVED 12 May 2023

ACCEPTED 12 September 2023

PUBLISHED 02 November 2023

## CITATION

Yadav M, Panwar R, Rustagi A,  
Chakraborty A, Roy A, Singh IK and Singh A  
(2023) Comprehensive and evolutionary  
analysis of *Spodoptera litura*-inducible  
Cytochrome P450 monooxygenase gene  
family in *Glycine max* elucidate their role  
in defense.

*Front. Plant Sci.* 14:1221526.

doi: 10.3389/fpls.2023.1221526

## COPYRIGHT

© 2023 Yadav, Panwar, Rustagi, Chakraborty,  
Roy, Singh and Singh. This is an open-access  
article distributed under the terms of the  
[Creative Commons Attribution License  
\(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or  
reproduction in other forums is permitted,  
provided the original author(s) and the  
copyright owner(s) are credited and that  
the original publication in this journal is  
cited, in accordance with accepted  
academic practice. No use, distribution or  
reproduction is permitted which does not  
comply with these terms.

# Comprehensive and evolutionary analysis of *Spodoptera litura*-inducible Cytochrome P450 monooxygenase gene family in *Glycine max* elucidate their role in defense

Manisha Yadav<sup>1,2</sup>, Ruby Panwar<sup>1,3</sup>, Anjana Rustagi<sup>3</sup>,  
Amrita Chakraborty<sup>4</sup>, Amit Roy<sup>5\*</sup>, Indrakant K. Singh<sup>6\*</sup>  
and Archana Singh<sup>1,2,3,7\*</sup>

<sup>1</sup>Department of Botany, Hansraj College, University of Delhi, Delhi, India, <sup>2</sup>J C Bose Center for Plant Genomics, Hansraj College, University of Delhi, Delhi, India, <sup>3</sup>Department of Botany, Gargi College, University of Delhi, Delhi, India, <sup>4</sup>EVA 4.0 Unit, Faculty of Forestry and Wood Sciences, Czech University of Life Sciences Prague, Prague, Czechia, <sup>5</sup>Forest Molecular Entomology Lab, EXTEMIT-K, EVA 4.0, Faculty of Forestry and Wood Sciences, Czech University of Life Sciences Prague, Prague, Czechia, <sup>6</sup>Molecular Biology Research Lab, Department of Zoology, Deshbandhu College, University of Delhi, New Delhi, India, <sup>7</sup>Department of Plant Molecular Biology, University of Delhi, New Delhi, India

Plants being sessile organisms and lacking both circulating phagocytic cells and somatic adaptive immune response, have thrived on various defense mechanisms to fend off insect pests and invasion of pathogens. CYP450s are the versatile enzymes, which thwart plants against insect pests by ubiquitous biosynthesis of phytohormones, antioxidants, and secondary metabolites, utilizing them as feeding deterrents and direct toxins. Therefore, a comprehensive analysis of biotic stress-responsive CYPs from *Glycine max* was performed to ascertain their function against *S. litura*-infestation. Phylogenetic analysis and evolutionary studies on conserved domains and motifs disclosed the evolutionary correspondence of these *GmCYPs* with already characterized members of the CYP450 superfamily and close relatedness to *Medicago truncatula*. These *GmCYPs* were mapped on 13 chromosomes; they possess 1-8 exons; they have evolved due to duplication and are localized in endoplasmic reticulum. Further, identification of methyl-jasmonate, salicylic acid, defense responsive and flavonoid biosynthesis regulating *cis*-acting elements, their interaction with biotic stress regulating proteins and their differential expression in diverse types of tissues, and during herbivory, depicted their responsiveness to biotic stress. Three-dimensional homology modelling of *GmCYPs*, docking with heme cofactor required for their catalytic activity and enzyme-substrate interactions were performed to understand the functional mechanism of their action. Moreover, to gain insight into their involvement in plant defense, gene expression analysis was evaluated, which revealed differential expression of 11 *GmCYPs* upon *S. litura*-infestation, 12 *GmCYPs* on wounding while foliar spray of ethylene, methyl-jasmonate and salicylic acid differentially regulated 11 *GmCYPs*, 6 *GmCYPs*, and 10 *GmCYPs*

respectively. Our study comprehensively analysed the underlying mechanism of *GmCYPs* function during *S. litura*-infestation, which can be further utilized for functional characterization to develop new strategies for enhancing soybean resistance to insect pests.

#### KEYWORDS

*Glycine max*, *Spodoptera litura*, evolutionary analysis, gene expression analysis, plant defense, herbivory

## 1 Introduction

Cytochrome P450s (CYPs) are a diverse and substantial group of monooxygenases found in all kingdoms, including bacteria, fungi, plants and animals (Nelson, 2009). A wide variety of both large and small compounds are included in the substrate spectrum of CYPs. CYPs are named P450-containing enzymes as they often function as final electron acceptors, i.e. terminal oxidase in electron transfer chains (Nelson et al., 1996; Nelson et al., 2006). Empirical evidence suggested that CYPs harbour multiple conserved motifs: an oxygen-binding I-helix (A/G)GX(E/D)T(T/S)], a C-terminus heme-binding site (FXXGXRXCXG), K-helix consensus (EXXR) and PXXR conserved motif. The K-helix and PXXR motif ensures the stability of CYPs fundamental structure by locking the heme pocket into its position (Werck-Reichhart and Feyereisen, 2000).

During the process of evolution, plants have acquired enormous CYPs representing about 1% of total coding genes through gene duplication and gene diversification events. Based on their phylogeny and homology, plant CYPs have been grouped into single and multiple family, clans and subfamilies (Nelson et al., 1996). The P450 superfamily in plants can be categorized into two distinct clades: A-type and non-A-type, which are typically classified into six single-family clans and four multiple-family clans (Nelson et al., 2006; Durst and Nelson, 1995). The *GmP450s* superfamily is also grouped into two types: A-type and non-A-type. A-type *GmP450s* contain a single clan (Clan 71) divided into 20 families. These 20 families contain 197 *GmP450s*. Non-A-type of *GmP450s* are divided into 9 clans (clan 51, clan 72, clan 74, clan 85, clan 86, clan 97, clan 710, clan 711, clan 727) which are further divided into 28 families. These 28 *GmP450* families contain 149 *GmP450s*. In soybean, the CYP71 is the most extensive *GmP450s* family, with 53 members, while the CYP82 family comprises 25 members, making it the second largest family (Khatri et al., 2022).

CYPs act as versatile catalysts regulating various physiological processes through a plethora of biosynthetic and detoxification reactions (Coon, 2005). Cytochrome P450 monooxygenases (P450) are oxidoreductases, which participate in the biosynthesis of secondary metabolites, antioxidants, and phytohormones via *N*-dealkylation, C–H bond hydroxylation, *N*-hydroxylation, *S*-oxidation, reduction, decarboxylation, dimerization, desaturation, epoxidation of numerous exogenous compounds, ring extensions and C–C cleavage (Guengerich and Munro, 2013; Pandian et al.,

2020). CYPs are heme-thiolate harbouring enzymes having crucial roles in growth, development and plant defense via synthesis of primary and secondary metabolites, fat metabolism, phytohormone biosynthesis and homeostasis, antioxidant activity, and xenobiotic metabolism (Guengerich, 2001; Zhang and Li, 2017).

Throughout their life cycle, plants are exposed to many abiotic and biotic stressors that negatively impact their growth, development, and productivity. Among biotic stresses, pest infestation in plants results in a nutrient-deprived condition, a decline in photosynthetic rate and crop loss. To cope with intruders, plants have evolved with complex defense mechanisms consisting of constitutive and induced defense mechanisms (Karban et al., 1997; Thomma et al., 1998; Kessler and Baldwin, 2002). However, CYPs reinforce chemical and molecular defense arsenals controlling secondary metabolite's biosynthesis, regulating biosynthesis and homeostasis of phytohormones and elevating reactive oxygen species (ROS) scavenging (Walling, 2000; Rojo et al., 2003).

The engagement of CYPs in the production of cyanogenic glucosides, cutin and lignin could have a direct or indirect connection with plant defense arsenals against various biological stressors (Pinot and Beisson, 2011). For example, *AtCYP82G1* is associated with the biogenesis of monoterpene volatiles and function as a defensive agent against herbivores (Lee et al., 2010). Also in poplar plants, *CYP79D7* and *CYP79D6* prompt the biosynthesis of feeding deterrents and toxins. Moreover, these enzymes induce the release of volatile compounds (aldoximes), which act in indirect defense by repelling herbivores and attracting their natural foes (Irmisch et al., 2013). In *Populus trichocarpa*, the herbivore-induced production of aldoloximes, was also reported to be regulated by the *CYP79D* gene family (Irmisch et al., 2015).

Similarly, *Verticillium dahliae* induced *CYP706B1* produces (+)- $\delta$ -*cadinene*-8-hydroxylase in cotton, which catalyse the biosynthesis of gossypol, a herbivore-defensive chemical (Luo et al., 2001). Resin acid is used in insect defence mechanisms in conifers (Hamberger et al., 2011). *CYP720B4* of Sitka spruce participates in diterpene metabolism and causes resin acid production. Also, it has been discovered that the *PAD3* belonging to CYP family, is indulged in camalexin production, a toxic phytoalexin that regulates resistance against green peach aphid (*Myzus persicae*) in *Arabidopsis* (Prince et al., 2014). Cembratriene-ol (CBT-ol) gets metabolised into cembratriene-diol (CBT-diol) by an unidentified CYP hydroxylase in trichome glands of *N. tabacum* plants. However, suppression of this CYP raised the CBT-ol content and showed resistance to aphids

(*Myzus nicotianae*) (Wang et al., 2001). Moreover, CYPs are involved in the biosynthesis of DIMBOA and DIBOA (benzoxazinoids), which critically defend plants against pests, pathogens and weeds (Butrón Gómez et al., 2010; Dick et al., 2012; Kato-Noguchi et al., 2000). These reports indicated the critical functions of CYPs in attaining defense against plant-insect pests. Thus, CYPs are the potential targets for enhancing plant defense by employing metabolic engineering.

Soybean (*Glycine max* (L.) Merr.) holds an important place in agriculture as it is the fourth-largest grown crop and dietary staple food in Asian countries (Badole and Bodhankar, 2012). Soybean, is an economical important leguminous plant with high nutritional and therapeutic values, which offers a high percentage of protein (38-42%), edible oil (18-22%), and carbohydrates (17-19%) apart from unsaturated FAs (fatty acids), antioxidants, minerals, fibres and vitamins (Chauhan et al., 2002). Soybean is a multipurpose legume used in human and animal food, fodder and industrial purposes; enhancing its demand (Soyastats, 2010). Soybean is known for cancer prevention, reducing cholesterol and for growth and development as it contains a good amount of isoflavones, calcium and vitamins (Carrão-Panizzi and Erhan, 2007). Soybean aids in the development of soil fertility as it fixes atmospheric nitrogen via root nodules and enriches the soil by drooping leaves on the ground during maturity (Patil et al., 2018). Soybean being nutrient-rich with luxuriant growth, attracts many insect pests (Brahman et al., 2018). However, among several constraints to its production, insect pests negatively impact soybean quality and productivity (Cui et al., 1997). Polyphagous *Spodoptera litura* (Fab.), commonly known as common cutworm is a crucial defoliator pest that causes significant soybean crop loss due to its gregarious feeding. *S. litura* has broadened its distribution in the tropical and temperate areas of the Asia-Pacific islands as it has huge reproductive and migratory ability and infests more than 120 crop plants (Maree et al., 1999; Baskar et al., 2012). *S. litura*'s neonate larvae scrap chlorophyll from leaves while grown-up ones feed voraciously on soybean leaves and pods by making holes/leaving veins only (Bughio et al., 1994). This pest remains active through July-October synchronize with the soybean reproductive period; cause a loss of 30-50% in soybean grain yield (Marwoto, 2008; Vasudev and Sohal, 2016; Sundar et al., 2018). Therefore, the development of insect-pest tolerant varieties of soybean is in high demand for enhancing soybean production.

A thorough examination is being conducted in the current study of the *S. litura*-induced P450 cytochrome monooxygenase gene family in *G. max* to deduce their role in the insect-defense mechanism. *In silico* as well as *in vitro* studies related to phylogenetic and evolutionary analysis, gene duplication events, identification of *cis*-acting elements, protein-protein interaction network analysis and differential expression in diverse types of organs and gene expression analysis upon *S. litura*-infestation, wounding and foliar spray of signalling compounds, three-dimensional homology modelling, identification of conserved motifs, docking with heme cofactor required for their catalytic activity and enzyme-substrate interactions, depicted their responsiveness to biotic stress and allowed us to understand the functional mechanism of their action. This study on scrutinizing the role of the P450 cytochrome monooxygenase gene family in plant defense against *S. litura*, will

generate some testable hypothesis for future functional studies which could help in generating insect resistance variety through genome modification and molecular assisted breeding.

## 2 Material and methodology

### 2.1 Selection of candidate GmCYP450s in *G. max* genome

A study on transcriptome analysis of *G. max* by, Wang et al., 2014 revealed the induction of hundreds of genes and transcript diversity in response to *S. litura* feeding. Out of diverse induced transcripts, we have selected biotic stress-responsive differentially expressed CYP450 genes to analyze their role in *G. max*-*S. litura* interaction. Protein and nucleotide sequences of selected 16 GmCYPs were fetched from the Phytozome database version 13 (<https://phytozome-next.jgi.doe.gov/>) and also manually evaluated using NCBI-BLAST tool (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>). Molecular weight (MW), isoelectric point (pI), instability index, and hydrophilicity (GRAVY value) of candidate sequences were computed employing the ProtParam tool (<https://web.expasy.org/protparam/>) of ExPASy (<https://www.expasy.org/>). ExPASy server UniProtKB/Swiss-Prot (<https://www.expasy.org/resources/uniprotkb-swiss-prot>) was used to calculate the length of amino acids for all the listed GmCYPs sequences (Gasteiger et al., 2005). ExPasy-Prosites server was used to identify functional sites. DeepLoc Tool (<https://services.healthtech.dtu.dk/service.php?DeepLoc-1.0>) was employed to analyze the subcellular localization of listed GmCYPs to gain insight into their functional mechanisms. Subcellular localization of all GmCYPs was represented as a heatmap using TBtool software. Subsequently, sequences were uploaded on the TMHMM version 2.0 online tool (<https://services.healthtech.dtu.dk/service.php?TMHMM-2.0>) for transmembrane domain analysis (Wang et al., 2022).

### 2.2 Analysis of phylogenetic tree and gene duplication

To evaluate the evolutionary connections among leguminous dicots, model plants, monocots, homologous protein sequences of listed GmCYPs for *Arabidopsis thaliana*, *Nicotiana attenuata*, *Medicago truncatula* and *Oryza sativa* were acquired from NCBI database using BLAST search (<https://www.ncbi.nlm.nih.gov/guide/proteins/>). Subsequently, the Clustal omega tool was used to align amino acid sequences (<https://www.ebi.ac.uk/Tools/msa/clustalo/>). The maximum-likelihood phylogenetic tree was produced based on the result of multiple sequence alignments using MEGA 7.0 (Molecular Evolutionary Genetic Analysis) software (<https://megasoftware.net/>) with bootstrap replication value 1000, P-distance mode and partial deletion (Kumar et al., 2016). Finally, the phylogenetic tree was embellished, implying iTOL (<https://itol.embl.de/upload.cgi>) tool (Letunic and Bork, 2007).

MCSanX software was employed to identify duplicated CYP gene pairs found in the genomes of *G. max*, *A. thaliana*, *M.*

*truncatula*, *N. attenuata* and *O. sativa* through their non-redundant pairing, thereby generating five pairs (*G. max/G. max*, *G. max/A. thaliana*, *G. max/M. truncatula*, *G. max/N. attenuata* and *G. max/O. sativa*) (Wang et al., 2012). The software employed an alignment file (generated by executing the all-versus-all BLASTP program for each of five pairs with e-value 10<sup>-5</sup>) and genome annotation data (retrieved from RefSeq of NCBI database) as input to generate a collinearity file. Collinearity file, which includes collinear pairs found in CYP genes for each species pair, was utilised to map duplicate gene pairs through the Circos software (<http://circos.ca/software/>) package (Krzywinski et al., 2009). To investigate selection pressure employed in *GmCYPs* duplication, the Ka/Ks ratio (synonymous/non-synonymous substitution ratio) was calculated through PAL2NAL software (<http://www.bork.embl.de/pal2nal/>) (Suyama et al., 2006).

## 2.3 Gene structure, chromosomal mapping and conserved motifs analysis

Exon-intron organization of listed *GmCYPs* was illustrated using GSDS (Gene Structure Display Server) (<http://gsds.gao-lab.org/>), where genomic sequences of all the genes were aligned with their corresponding cDNA sequences. TBtool software was used to visualize the gene structure of 16 *GmCYPs* (Hu et al., 2015). Listed *GmCYPs* were mapped to soybean chromosomes based on the generic feature format version file (GFF) retrieved from the Phytozome database. Chromosomal position and relative distances of genes were spotted using Ritchie lab-Phenogram software (<http://visualization.ritchielab.org/phenograms/plot>) (Abbas et al., 2022).

Functional domain in *GmCYPs* was searched by submitting protein sequences to Pfam: Protein families database tool (<http://pfam-legacy.xfam.org/>). InterPro 91.0 database (<https://www.ebi.ac.uk/interpro/>) and NCBI Conserved domain database (<https://www.ncbi.nlm.nih.gov/Structure/cdd/cdd.shtml>) were also used for the detection of the conserved domains (Mistry et al., 2021). Later, MEME: Multiple Em for Motif Elicitation software (<https://meme-suite.org/meme/doc/meme.html>) was employed to predict conserved motifs and generate logo plots of listed *GmCYPs* sequences selecting the number of motifs: 10 and rest parameters as default (Bailey et al., 2009). Multiple sequence alignment (MSA) with amino acid sequences of candidate *GmCYPs* was performed to investigate existing conserved motifs utilizing Clustal Omega (<https://www.ebi.ac.uk/Tools/msa/clustalo/>).

## 2.4 Analysis of cis-acting regulatory elements analysis in gene promoter regions, Transcription factor binding sites and miRNA target analysis

The 1000 bp nucleotide sequences upstream of translational start codon of each listed *GmCYPs* were retrieved from Phytozome v13 database. The cis-acting elements in listed *GmCYPs* sequences were predicted employing the Plant Care database (<http://bioinformatics.psb.ugent.be/webtools/plantcare/html/>). Further,

the results of cis-acting element was analysed using TBtool software (Zhang et al., 2022). psRNA Target tool (<https://www.zhaolab.org/psRNATarget/>; Lescot et al., 2002) was employed for anticipation of miRNA that may target and regulate *GmCYPs* expression with default parameter settings (Dai and Zhao, 2011). The mature sequences of miRNA of *G. max* were accessed from the miRBase i.e. miRNA database (<https://mirbase.org/>). The *GmCYPs* mRNA sequences and miRNA sequences were submitted to psRNA Target analysis server to characterize *GmCYPs*'s miRNA target sites (Chen et al., 2020, Dai and Zhao, 2011).

Transcription factor (TF) binding sites prediction in the *GmCYPs* promoter region was executed using PlantTFDB online server (<http://planttfdb.gao-lab.org/>); using default parameters and *p*-value of  $\leq 1e-6$  (Malviya et al., 2023).

## 2.5 Plant-organ-specific gene expression profiling and protein-protein interaction

To analyse the organ-specific expression profiles of genes, RNAseq data related to differential expression in different tissues was retrieved from SoyBase Expression Explorer for each *GmCYPs* (<https://www.soybase.org/expression>). Expression values were measured as Reads/Kb/Million (RPKM) for various organs, including the seed, stem, nodule, flower, young leaf, and green pod (Waese et al., 2017). Normalized RPKM values of the listed *GmCYPs* were used to generate a heat map using TBtool.

The protein-protein interaction network of *GmCYPs* was deduced using the String database (<https://string-db.org/>). For PPI, *G. max* was selected as the model plant, with medium confidence = 0.400 and the remaining parameters at the default settings (Szklarczyk et al., 2021).

## 2.6 Expression analysis of *GmCYPs* upon *S. litura* infestation and wounding

### 2.6.1 Plant growth and maintenance

*Glycine max* seeds (Pusa 9712) were procured from ICAR-Indian Institute of Soybean Research, Indore, India. Seeds were surface sterilized after O/N soaking and sown in plastic pots in a plant growth chamber using standard protocols, i.e. a photoperiod of 16/8h light/dark cycle, 28°C temperature, 55-60% humidity and watered regularly.

### 2.6.2 *S. litura* rearing and sample preparation

*S. litura* pupa were procured from ICAR-NBAIR (National Bureau of Agricultural Insect Resources), India and reared under standard conditions at temperature: 28°C; relative humidity; 65-70% and 14/10h light/dark cycle. For *S. litura*-infested sample collection, the 4<sup>th</sup> instar larvae were starved O/N and released on 30 days old healthy soybean plants (One larva per plant) for feeding, whereas plants not infested by larvae were treated as control. Mechanical wounding was accomplished as per Singh et al., 2008. After 24 h of infestation/wounding, the entire shoot of control and

treated plants were harvested from 3 independent experiments and immediately frozen at  $-80^{\circ}\text{C}$  for further experimentation.

### 2.6.3 RNA extraction and quantitative qPCR

Trizol reagent (Sigma Aldrich, US) was utilized to extract total RNA from both control and treated samples following the manufacturer's guideline and as per previously described protocols (Keshan et al., 2021; Singh et al., 2021b; Singh et al., 2018). RNA was quantified using UV-spectrophotometer (Biorad), and quality was analyzed on 2% agarose gel. The first-strand cDNA was constructed utilizing the cDNA synthesis kit, (Biorad) following the manufacturer's instructions. Primer 3 (v.0.4.0)10 tool was employed for designing specific primers, and CFX Opus 96 Real-Time PCR (Biorad) was used for RT-PCR. Each RT-PCR reaction was conducted in SYBR green supermix with 10 mM solution of each gene-specific primer (0.5 mL), 1  $\mu\text{L}$  of diluted cDNA and 3  $\mu\text{L}$  of RNase-free water. A soybean housekeeping gene, elongation factor 1(EF-1), was used as an internal reference gene (Saraiva et al., 2014).

## 2.7 Analysis of gene expression *GmCYPs* upon foliar spray with signalling compounds

Foliar application of an equal volume of 50 mM ethephon, 5 mM salicylic acid, 100 mM of methyl-jasmonate, and 0.5% (v/v) ethanol in water/only water was administered on one-month-old healthy soybean plants kept in individual enclosures under identical conditions (Singh et al., 2008; Yadav et al., 2022). The treated and control samples were collected after 24 hours and immediately frozen at  $-80^{\circ}\text{C}$ . The samples were collected in biological triplicates for subsequent analysis. Data for gene expression analysis was interpreted employing the  $2^{-\Delta\Delta\text{CT}}$  method. T-test ( $P < 0.05$ ) and ANOVA were applied for statistical analysis, while differentially expressed genes were selected based on a criterion of more than two-fold, either induction or reduction.

## 2.8 Homology modelling of selected *GmCYPs* and docking with heme ligand

Three-dimensional structures of *GmCYPs* were predicted using Phyre2, Swiss modeller and i-TASSER by submitting protein sequences. Structures were analyzed based on different parameters such as query coverage, template size and origin, stability of the predicted proteins and validated by Ramachandran plot (Kelley et al., 2015; Waterhouse et al., 2018; Zhang, 2008). After which, the best model were utilized for further analysis such as identification of conserved motifs and highlighting them on the most accurately modelled structures using Pymol (DeLano, 2002). Docking of heme prosthetic factor with one of the randomly selected *GmCYP* was executed using Maestro Glide docking software (Schrodinger's tool) (Bhachoo and Beuming, 2017). Prior to docking, ligand (Heme) and *GmCYP* structures were optimized by Maestro's Ligprep and

Protein Preparation wizard, respectively. Receptor grid generation was constructed by selecting residues of heme-binding motifs. *Arabidopsis* CYP94C1 was taken as a reference for this analysis.

## 2.9 Interaction of selected *GmCYPs* with their substrate

Information about substrates of *GmCYPs* was retrieved from literature and databases survey, and 3-D structures of *GmCYPs* substrates were fetched from PubChem. To show the interaction between *GmCYPs* and their substrates, the InChI format of the substrates fetched from PubChem was converted to. pdb format employing Open Babel-Chemical file format (<http://www.cheminfo.org/Chemistry/Cheminformatics/FormatConverter/index.html>) (O'Boyle et al., 2011; Kim et al., 2019). Patch Dock was applied to perform interaction studies with parameters: RMSD value= 4 and protein-small ligand complex type, and 100 top results. The ten highest-rank Patch dock docking outcomes were refined through FireDock (Mashiach et al., 2008; Schneidman-Duhovny et al., 2005).

## 3 Results

### 3.1 Sequence retrieval and *in silico* characterization of *GmCYP450s*

Gene IDs for all *S. litura*-induced *GmCYPs* were accessible from NCBI and phytozome database, from where their protein and nucleotide sequences were retrieved. A self-BLAST search was also conducted to eliminate redundant sequences that finally identified 16 of the *S. litura*-inducible *GmCYP450s* involved in biotic stress response. The naming of genes was based on their homology to *Arabidopsis* CYP450s and Phytozome gene description. Detailed information of 16 candidate *GmCYPs*, including gene ID, Phytozome ID, Uniprot ID, and the number of peptide sequences along with physiochemical properties, are represented in Table 1. The candidate *GmCYPs* range from 6152 bp (*GmCYP78A5*) to 2164 bp (*GmCYP78A3*), and CDS sequences ranged between 1398 bp (*GmCYP85A1*) to 1746 bp (*GmCYP78A5*). The full length of peptide sequences varied from 466 (*GmCYP85A1*) to 582 (*GmCYP78A5*) amino peptides with molecular weight (MW) varying from 53.375 (*GmCYP85A1*) to 65.343 (*GmCYP78A5*) kDa and their isoelectric point (pI) values varying from 6.54 (*GmCYP82A2*) to 9.27 (*GmCYP90A1*), indicating that maximum proteins were basic. However, instability index values indicate that most *GmCYPs* are stable (Instability index  $> 40$ ) except *GmCYP71AU50*, *GmCYP93A3*, *GmCYP89A2*, and *GmCYP71D10*. In addition, 13 genes were hydrophilic (GRAVY value less than zero), while *GmCYP78A5*-like, *GmCYP76C1*-like Determination of subcellular localization were predicted using reticulum (ER). Their presence was also predicted in the chloroplast, mitochondria, cell membrane, vacuole, golgi apparatus and nucleus with minimal expression, as depicted in Supplementary Figure 1.

TABLE 1 List of 16 *GmCYPs* and their physiochemical properties.

Gene Name	Phytozome ID	Uniprot ID	Peptide sequences	Molecular weight (kDa)	pI	Instability index	Aliphatic Index	GRAVY
GmCYP98A2	Glyma.19G126000.1	O48922	510	57.674	8.58	36.45, Stable	93.28	-0.19
GmCYP82A2	Glyma.13G285300.1	O81972	523	58.715	6.54	36.18, Stable	100.06	-0.059
GmCYP78A4-like	Glyma.02G119600.1	I1JEG7	517	58.316	6.96	35.84, Stable	94.4	-0.002
GmCYP98A2-like	Glyma.03G122000.1	I1JN04	510	57.706	8.72	36.28, Stable	92.32	-0.198
GmCYP71A1	Glyma.05G042600.1	I1K057	513	58.483	9.01	36.11, Stable	91.58	-0.108
GmCYP78A5	Glyma.06G310800.1	K7KYE8	582	65.343	8.06	38.15, Stable	86.95	-0.129
GmCYP78A3	Glyma.07G052300.1	I1KHS1	543	61.033	8.9	33.48, Stable	89.24	-0.06
GmCYP78A5-like	Glyma.08G104100.1	I1KS17	515	57.914	9.01	39.69, Stable	96.75	0.021
GmCYP71AU50	Glyma.08G140500.1	I1KT85	501	57.199	8.69	42.93, Unstable	95.92	-0.204
GmCYP93A3	Glyma.10G092500.1	A0A0R0HYP7	510	57.967	6.67	42.3, Unstable	94.62	-0.187
GmCYP76C1-like	Glyma.11G108300.1	I1LJ26	508	54.135	7.14	32.24, Stable	102.11	0.030
GmCYP90A1	Glyma.11G228900.1	K7LRQ4	473	53.942	9.27	39.8, Stable	93.18	-0.137
GmCYP78A6-like	Glyma.16G021200.1	I1MKF4	537	60.415	7.66	32.05, Stable	93	0.038
GmCYP85A1	Glyma.19G033900.1	I1N6E7	466	53.375	9	34.85, Stable	87.83	-0.185
GmCYP89A2	Glyma.20G018800.1	A0A0R4J634	513	59.257	8.59	45.82, Unstable	97.79	-0.13
GmCYP71D10	Glyma.15G050300.1	O48923	511	58.462	8.59	45.98, Unstable	92.67	-0.222

### 3.2 Analysis of phylogenetic and gene duplication events

A phylogenetic tree was constructed from *G. max* CYP450s genes (16) and their homologous members from *A. thaliana* (14), *M. truncatula* (14), *O. sativa* (14) and *N. tabacum* (15); to understand the evolutionary relationship among them (Figure 1A). The findings show that CYPs have evolved dramatically throughout the time series, as seen by distinct clades in tree, implying that CYPs play a variety of functions in various plant systems. These proteins were categorized into 12 clades; among them, Clan 71 has a maximum of 45 members, and Clan 80, Clan 72, Clan 92, and CYP736 contained 10, 4, 11 and 3 members, respectively. *GmCYPs* were dispersed across the phylogenetic tree and belonged to different clades illustrating evolutionary relatedness with CYPs from *A. thaliana*, *M. truncatula*, *O. sativa* and *N. tabacum*. However, the maximum like-hood tree indicated that *G. max* and *M. truncatula* are closely related evolutionarily (having the highest bootstrap values), inferring they are members of the same family. Gene duplications are crucial for expanding and evolving gene families, driving species divergence and new gene functions (Innan and Kondrashov, 2010). Examination of gene duplication occurrence in the *G. max* genome revealed that 15 pairs of *GmCYPs* underwent segmental duplication i.e. duplicated genes were found on different chromosomes. A total of 6 pairs of duplicated genes between *GmCYPs* and genes from *A. thaliana* while 12 pairs of duplicated genes between *GmCYPs* and genes from *M. truncatula* were observed (Figure 1B). The selection pressure was interpreted by calculating Ka/Ks which varied from 0.04918

(NP\_850337.1/XP\_003521103.1) to 0.332959 (XP\_006582383.1/XP\_039688169.1) (Supplementary Table 1).

### 3.3 Gene structure, chromosomal localization and motif analysis

To comprehend the structure of putative *GmCYPs*, their intron-exon organization was predicted, which revealed that *GmCYPs* gene structures were diverse, with the exon number varying from 1 to 9. *GmCYP89A2* was predicted to be intron-less (single exon) whereas *GmCYP85A1*, *GmCYP90A1* and *GmCYP78A5* consists of 8, 7, 6 introns and 9, 8, 7 exons respectively. *GmCYP93A3*, *GmCYP93A2* and *GmCYP93A2-like* genes contained 3 exons and 2 introns, while the rest of them have 2 exons with a single intron (Supplementary Figure 2). This gene structure analysis helps to determine evolutionary, structural and functional characteristics of genes. Chromosomal mapping of *GmCYPs* illustrated that these *GmCYPs* were randomly placed on 13 (out of 20) soybean chromosomes. Chromosomes 8, 9 and 11 contained 2 *GmCYPs*, whereas the remaining 10 chromosomes contained a single gene. *GmCYPs* are located on chromosome arms except for *GmCYP98A2*, *GmCYP98A2-like* and *GmCYP93* implying their frequent participation in recombination process (Supplementary Figure 3, Supplementary Table 2).

Conserved motif and domain analysis is a useful method for protein function prediction (Watson et al., 2005). All the listed *GmCYPs* were examined and analyzed with Pfam, NCBI-CDD and



(EXXR). Exploration of listed *GmCYPs* for conserved motifs was executed by MSA and identified these four conserved motifs (Supplementary Figure 5).

### 3.4 Cis-acting element, TF binding site and miRNA target analysis

Cis-acting elements present in the promoter region may be crucial in controlling the transcription of genes in response to environmental stressors and plant growth. A comprehensive analysis was executed to decipher cis-element analysis existing in the promoter region of 16 *GmCYPs*. Several cis-acting elements, including those responsive to stress, hormones and TF binding, were identified in the promoter regions of *GmCYPs*, where each *GmCYPs* harbours more than one cis-acting element (Figure 2). Analysis revealed that these 16 *GmCYPs* harbour 88 light-responsive elements, which mainly include Sp1, sbp-CMA1c, TCT-motif, TATA-box, GATA-motif, I-box, Box 4, Gap-box, MRE, G-box, AE-box, chs-CMA2a, chs-CMA1a, GATA-motif, Sp1, GT1-motif, ATC-motif, AE-box, ATCT-motif, 3-AF1 binding site, ABRE3a, CACGTC, TCTTAC, TTAAGTAA, GATAGGA motif. However, 46 phytohormone-responsive cis-elements were predicted; 16 methyl jasmonate hormone-responsive elements (TGACG-motif and CGTCA-motif), 14 ABA-responsive elements (ABRE), 10

Gibberellin hormone-responsive elements (TATC-box, CCTTTTG, P-box, GARE-motif), 3 Auxin-responsive elements (AuxRR-core and TGA-element), and 3 Salicylic acid-responsive elements (TCA-element). The presence of 3 low temperature (LTR), 22 Anaerobic conditions-responsive (ARE), 5 defense-responsive (TC-rich repeats), 5 drought-responsive elements and 1 flavonoid biosynthesis regulating (MBSI) cis-acting elements were detected. Additionally, binding sites of TFs on the *GmCYPs* promoter region were uncovered using PlantTFDB tool. Here, 14 types of transcription families were predicted in 11 *GmCYPs* promoter regions (Supplementary Figure 6) while *GmCYP90A1*, *GmCYP89A2*, *GmCYP78A6-like* and *Gm78A4-like* were deprived of any TF binding sites. AP2 family were predicted in *GmCYP76C1-like*, *GmCYP78A5-like* and *GmCYP98A2-like*; ERF was found on *GmCYP76C1-like* and *GmCYP82A2*; Bzip was restricted to *GmCYP76C1-like*; NAC in *GmCYP93A3*, *GmCYP76C1-like* and Dof in *GmCYP76C1-like*, *GmCYP93A3*, *GmCYP78A5-like* and *GmCYP85A1*.

miRNA in eukaryotes is a non-coding, short RNA comprising 21-24 nucleotides that bind to and cleaves mRNA for post-transcriptional regulation (Jones-Rhoades et al., 2006). Using *G. max* miRNA library, miRNA target analysis for *GmCYPs* was performed using the psRNATarget tool; it revealed 16 *GmCYPs* as putative targets of 94 miRNAs when E = 5.0 (Supplementary Table 3). *GmCYP78A5* was targeted by the largest numbers of miRNA, 23 followed by *CYP90A1* (11 miRNA), *CYP93A3* (9

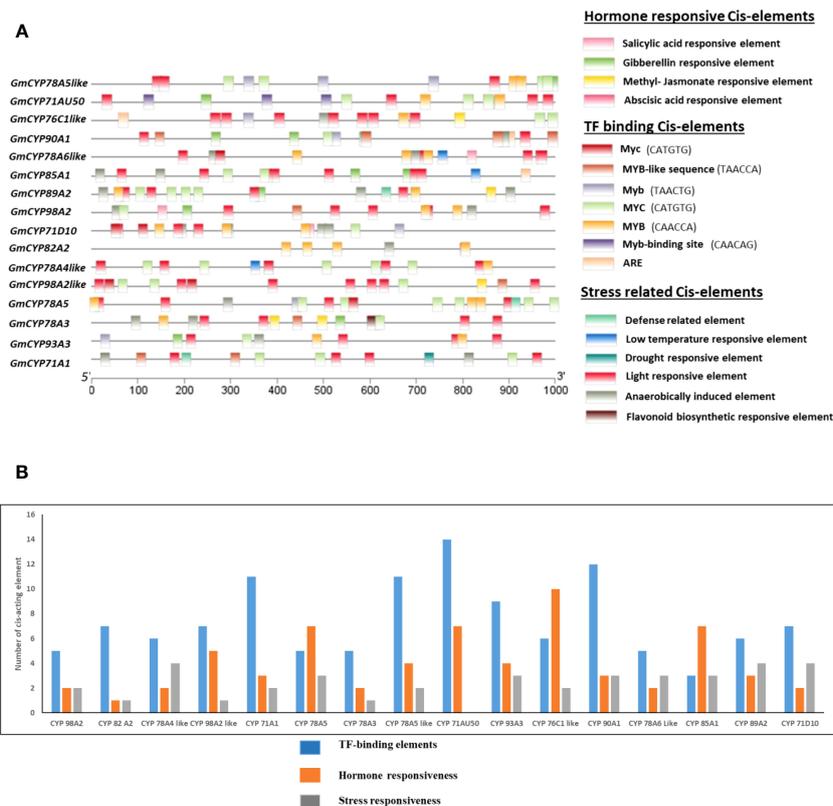


FIGURE 2

(A) Identification of Cis-acting elements in 16 *GmCYPs* promoters. Predicted Cis-elements of *GmCYPs* were depicted in different coloured boxes which are categorized into three categories: Hormone responsive, stress related and TF factor binding. Relative position of cis-acting elements based on start codon (ATG) can be estimated using scale. (B) and graph is showing number of cis-element involved in TF-binding elements, hormone responsiveness and stress responsiveness for each 16 *GmCYPs*.

miRNA). In this analysis, we can conclude that miRNAs possessed diversity in regulating *GmCYPs*, where miRNAs inhibit *GmCYPs* mainly through cleavage and some via translation.

### 3.5 Organ-specific gene expression profiling

To gain insight into regulatory functions of genes controlled by a specific plant-organ, RNA-seq expression data (RPKM Values) of listed *GmCYPs* were procured from soybean expression explorer for investigating *GmCYPs* expression into six different plant-organ i.e. seed, young leaf, nodule, root, flower, pod shell, seed (Figure 3). The results exhibited that all the 15 *GmCYPs* were expressed at least in one plant-organ, and *GmCYP98A2*, *GmCYP98A2-like*, *GmCYP78A3*, *GmCYP78A6-like* showed high expression in all six plant-organs indicating their predicted roles in soybean growth, development and countering various environmental stresses. Furthermore, some *GmCYPs* showed organ-specific expression patterns, such as *CYP78A5* and *CYP93A3*, which exhibited augmented expression within the root, implying their role in the development of roots, interaction with plant microbes and stress tolerance. Similarly, *CYP71AU50* and *CYP85A1* were highly expressed in flowers, while *CYP89A2* showed substantially greater expression in pod shells, implying their functions in reproduction/floral-related stimuli. Additionally, *CYP71D10* and *CYP76C1-like* were found to be explicitly expressed in seeds, indicating their role in seed development and maturation. On the other hand, *CYP78A4-like* was expressed only in two plant-organs, seed and root, with a higher level of

expression in the seed. Finally, *CYP71A1* showed minimal expression in the flower and root, suggesting its possible involvement in other plant organs or stages of soybean development. However, it should be noted that expression data of *GmCYP78A5* was unavailable, which limits our understanding of its organ-specific expression pattern.

### 3.6 Protein network, enrichment analysis and GO annotation

Protein-protein interaction (PPI) analysis of each *GmCYPs* with other *G. max* proteins was predicted using the STRING database. PPI analysis is a comprehensive method of knowing predicted protein function. Out of 16 *GmCYPs*, only 8 *GmCYPs* have shown interactions with other *G. max* proteins involved in plant stress response (Figure 4). *GmCYP98A2* exhibited interaction with GLYMA07G02460.2 (*Shikimate o-hydroxycinnamoyltransferase isoform x3*) GLYMA08G23560.3 (*Shikimate o-hydroxycinnamoyltransferase*) and GLYMA03G27740.1 (*5-O-(4-coumaroyl)-D-quinic acid 3'-monooxygenase*); involved in phenylpropanoid pathway. *GmCYP82A2* showed its association with *CYP74A1* (*Allene oxide synthase*), which initiates the biosynthesis of jasmonic acid from hydroperoxides of free fatty acids derived from lipoxygenase, representing the first step in the process (Farmer and Goossens, 2019). GLYMA15G15830.1 (probable histone-arginine Methyltransferase 1), a positive regulator of oxidative stress tolerance, was predicted to interact with *GmCYP82A2*. STRING predicted *GmCYP78A3* interaction with GLYMA07G02380.1 (AP2-like ethylene-responsive transcription factor), GLYMA07G20070.1 and

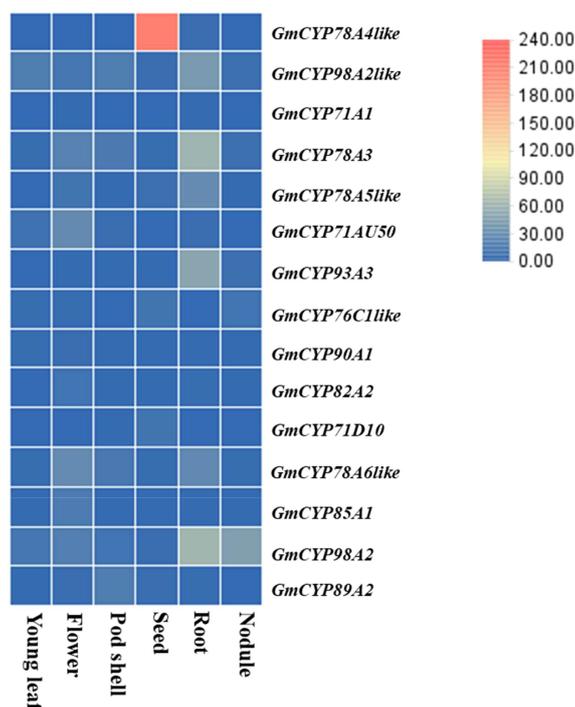


FIGURE 3

Organ-specific Expression profile of *GmCYPs* of *Glycine max* is represented in the form of heat map. Normalized RPKM values of *GmCYPs* were retrieved from Soybase Expression Explorer. (Expression data was not available).

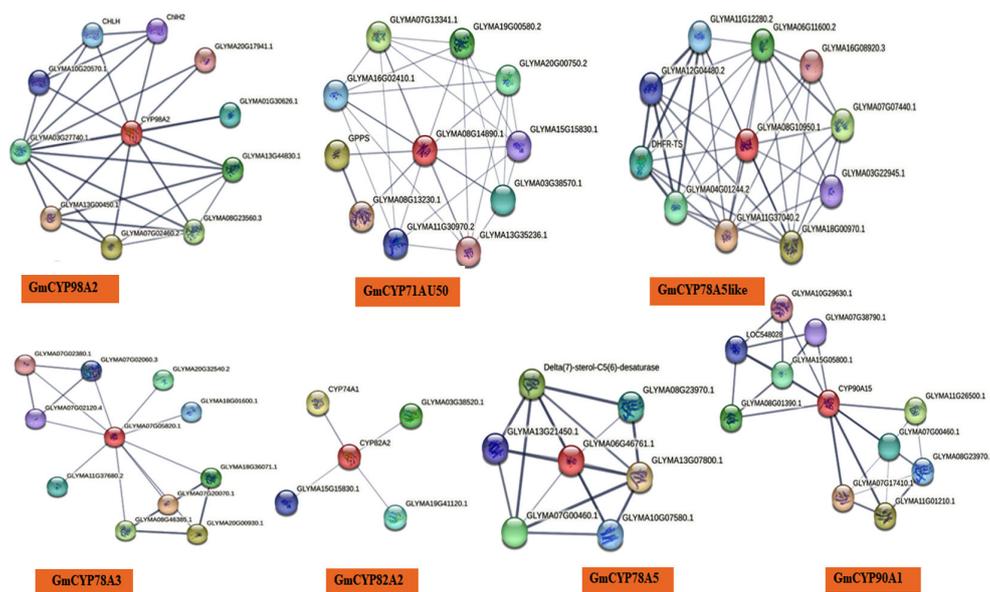


FIGURE 4  
Protein-protein interactions (PPI) network of *GmCYPs* (which showing interaction with other stress responsive proteins). String database was utilized for this analysis.

GLYMA20G00930.1 (Checkpoint serine/threonine-protein kinase). *GmCYP71AU50* exhibited interaction with *GLYMA08G13230.1* (*Pathogen-inducible salicylic acid glucosyltransferase*), *GLYMA16G02410.1* (*CYP b5-like steroid binding domain-containing protein*), and *GPPS* (*Geranylgeranyl diphosphate synthase, type ii*) (Barja and Rodriguez-Concepcion, 2021). *GGPP* is responsible for the biosynthesis of chlorophylls, carotenoids, or plastoquinones, which is vital for photosynthesis.

*GmCYP90A1* exhibited interaction with brassinoid synthesizing enzymes; *GLYMA11G07780.1* and *GLYMA02G05780.2* (*3-epi-6-deoxocathasterone 23-monooxygenase*), *GLYMA07G17410.1* (*Steroid 5-alpha-reductase det2*), *GLYMA06G24540.1* (member of *CYP450* family). It also interacts with *GLYMA07G04470.1* (Typhasterol/6-deoxytyphasterol 2alpha-hydroxylase; belongs to the cytochrome P450 family) responsible for tyrosine and secondary metabolite synthesis. *CYP78A5* interact with *GLYMA13G07800.1* Delta(7)-sterol-C5(6)-desaturase isoform X1 and Shikimate o-hydroxycinnamoyl transferase involved in the biosynthesis of sitosterol and campesterol (Hoffmann et al., 2004). *CYP78A5-like* interacts with *GLYMA11G37040.2* (Alpha-ketoglutarate-dependent dioxygenase alk), *GLYMA18G00970.1* (Alpha-ketoglutarate-dependant dioxygenase alk, U) engaged in various reactions in plant metabolism such as flavonoid biosynthesis and ethylene biosynthesis (Zhang et al., 2004; Cheng et al., 2014). *GLYMA16G08920.3* (heavy metal-associated isoprenylated plant protein 4) is involved in heavy metal detoxification and homeostasis mechanism. *CYP85A1* interact with steroid 5-alpha-reductase (*GLYMA07G17410.1* and *GLYMA11G01210.1*) involved in steroid biosynthesis and brassinoid biosynthesis (*GLYMA02G05780.2* and *GLYMA07G04470.1*) and *GLYMA11G26500.1* (long-chain fatty acid

omega-monooxygenase), which is involved in the biosynthesis of complex polymer, sporopollenin.

### 3.7 Gene expression analysis of *GmCYPs* during *S. litura*-infestation and mechanical damage

Gene expression analysis of 16 *GmCYPs* was executed by qPCR using gene-specific primers, the results indicated that 7 of *GmCYPs* were upregulated while 4 of the *GmCYPs* were downregulated on *S. litura* infestation. *GmCYP78A5* exhibited a maximum change in expression while *Gm78A5-like*, *Gm98A2-like*, *GmCYP78A3*, *GmCYP76C1-like*, *GmCYP89A2* and *GmCYP71D10* showed upregulation whereas *GmCYP78A4-like*, *GmCYP71A1*, *GmCYP93A3* and *GmCYP90A1* showed downregulation during infestation. However, 9 out of 16 *GmCYPs* were upregulated on mechanical wounding, and 3 *GmCYPs* were downregulated. *GmCYP78A5*, *GmCYP89A2* and *GmCYP78A5-like* showed high gene expression on mechanical wounding (Figure 5).

### 3.8 Gene expression analysis of *GmCYPs* upon the foliar spray of signalling compounds

Gene expression analysis of 16 *GmCYPs* was performed upon the foliar spray of ethylene, salicylic acid and methyl-jasmonate to acquire insights into the role of signalling compounds in regulating the expression of *GmCYPs* in plant defense. On ethylene treatment, 11 *GmCYPs* (*GmCYP82A2*, *GmCYP98A2-like*, *GmCYP78A5*,

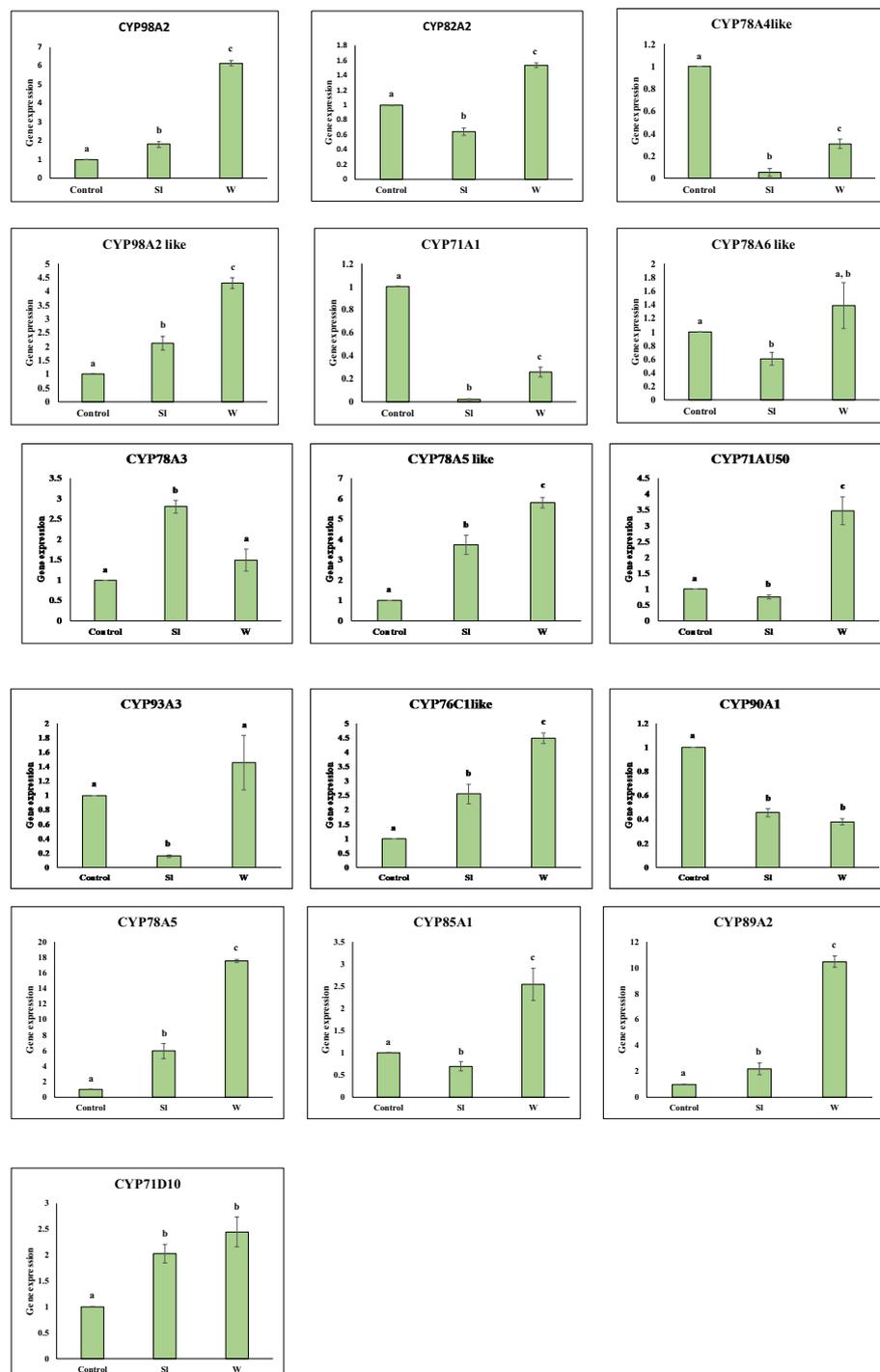


FIGURE 5

Graph showing expression profile of 16 *GmCYPs* upon SI (*S. litura*-infestation) and W(wounding). The qPCR data were analyzed utilizing  $2^{-\Delta\Delta CT}$  method and statistical analysis was performed using ANOVA and Tukey's test ( $P < 0.05$ ). Bars were showing mean value whereas standard deviation were represented using error bars. a and b represents if two values are significantly different or not.

*GmCYP78A3*, *GmCYP93A3*, *GmCYP76C1-like*, *GmCYP90A1*, *GmCYP78A6-like*, *GmCYP85A1*, *GmCYP89A2*, and *GmCYP71D10GmCYPs* (*GmCYP98A2*, *GmCYP78A4-like*, *GmCYP71A1*, *GmCYP78A5-like* and *GmCYP71AU50*) were upregulated and downregulated respectively. Similarly, methyl-jasmonate application induced expression of 10 *GmCYPs* (*GmCYP78A5-like*, *GmCYP98A2-like*, *GmCYP78A5*, *GmCYP78A3*, *GmCYP78A5-like*, *GmCYP71AU50*,

*GmCYP93A3*, *GmCYP76C1-like*, *GmCYP89A2*, and *GmCYP71D10*). However, on SA application, 4 *GmCYPs* (*GmCYP71D10*, *GmCYP89A2*, *GmCYP78A6-like* and *GmCYP98A2*) were induced, while 6 *GmCYPs* (*GmCYP78A4-like*, *GmCYP98A2-like*, *GmCYP78A5-like*, *GmCYP76C1-like*, *GmCYP90A1* and *GmCYP85A1*) were repressed, designating the involvement of SA directly or indirectly in gene regulation. (Figure 6).

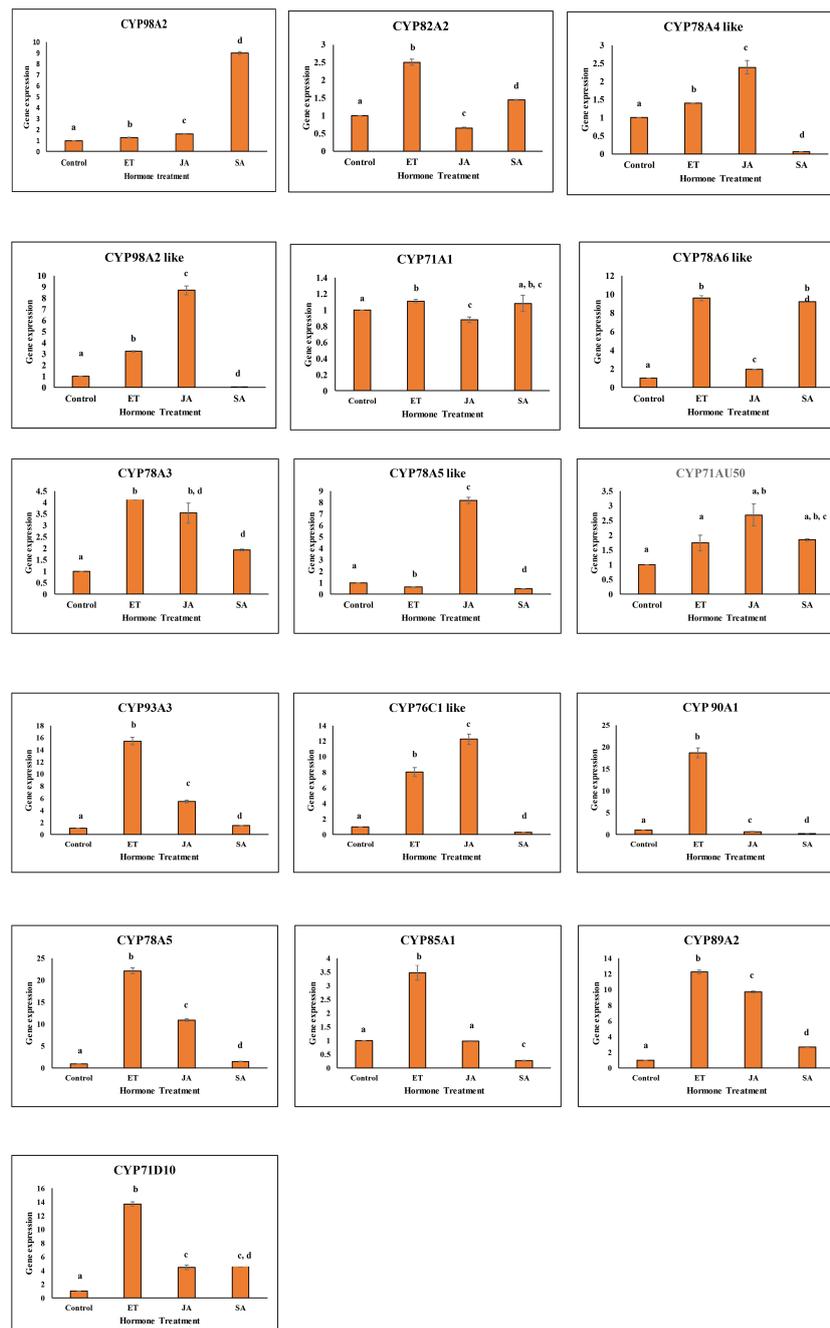


FIGURE 6

Graph showing expression profile of 16 *GmCYPs* on application of JA (Methyl-jasmonate), SA (Salicylic acid), and ET (Ethylene). The qPCR data were analyzed utilizing the  $2^{-\Delta\Delta CT}$  method and statistical analysis was performed using ANOVA and Tukey's test ( $P < 0.05$ ). Bars were showing mean values whereas standard deviation were represented using error bars.

Methyl-jasmonate foliar application upregulated four of the *S. litura*-induced *GmCYPs*, illustrating jasmonates' role in regulating the *GmCYPs* expression upon *S. litura*-infestation. The presence of methyl-jasmonate sensitive *cis*-acting regulatory elements in the promoter region in these genes supported the notion that jasmonate is responsible for their regulation. Four *GmCYPs*, which had their expression elevated by ethylene treatment, were also induced by jasmonate. Jasmonate and ethylene are reported to act synergistically (Kessler and Baldwin, 2002).

### 3.9 Homology modelling of 16 *GmCYPs*

Three-dimensional (3-D) structures of differentially regulated *S. litura*-infested 16 *GmCYPs* were modelled using i-TASSER, and structures were corroborated by ensuring that >90% of the protein region fell within the permitted area of the Ramachandran plot. The template, coverage (>85%), Z-score, and estimated  $T_m$ ; all these parameters are listed in Supplementary Table 4. All 16 *GmCYPs* chosen for analysis revealed the most typical arrangement of  $\alpha$ -helix

and  $\beta$ -sheets. The evolutionary conserved heme-binding region, PXR motif, I-helix oxygen-binding domain and K-helix motifs of CYP450 proteins were highlighted using distinct colours (Figure 7).

### 3.10 Interaction of selected *GmCYPs* with their substrate and docking with heme ligand

To study the enzymatic mechanism of *GmCYPs*, substrates of *GmCYPs* were retrieved from the literature survey and KEGG database. The interaction of 10 *GmCYP* with their specific substrate was examined through enzyme-substrate docking. The amino acid residues and polar bonds formed by substrate with *GmCYPs* catalyst were highlighted with magenta and yellow colors, respectively (Figure 8).

The global energy score of *GmCYPs* and substrate interaction vary from -20 to -50, indicating their strong interaction (Supplementary Table 5). Earlier studies have reported that heme binding with CYP450s is essential for their enzymatic activities (Li et al., 2008). Therefore, molecular docking of a selected *GmCYPs* was performed with a heme ligand, and the interaction results were compared with AtCYP94C1 (Figure 9). The global energy score for *GmCYP98A2* like protein and heme, was -3.392, while for AtCYP94C1 and heme was -2.341.

## 4 Discussion

Throughout their lifespan, plants confront a variety of stresses induced by both biotic and abiotic factors. To thwart themselves, plants have thrived in complex defence mechanisms that involve

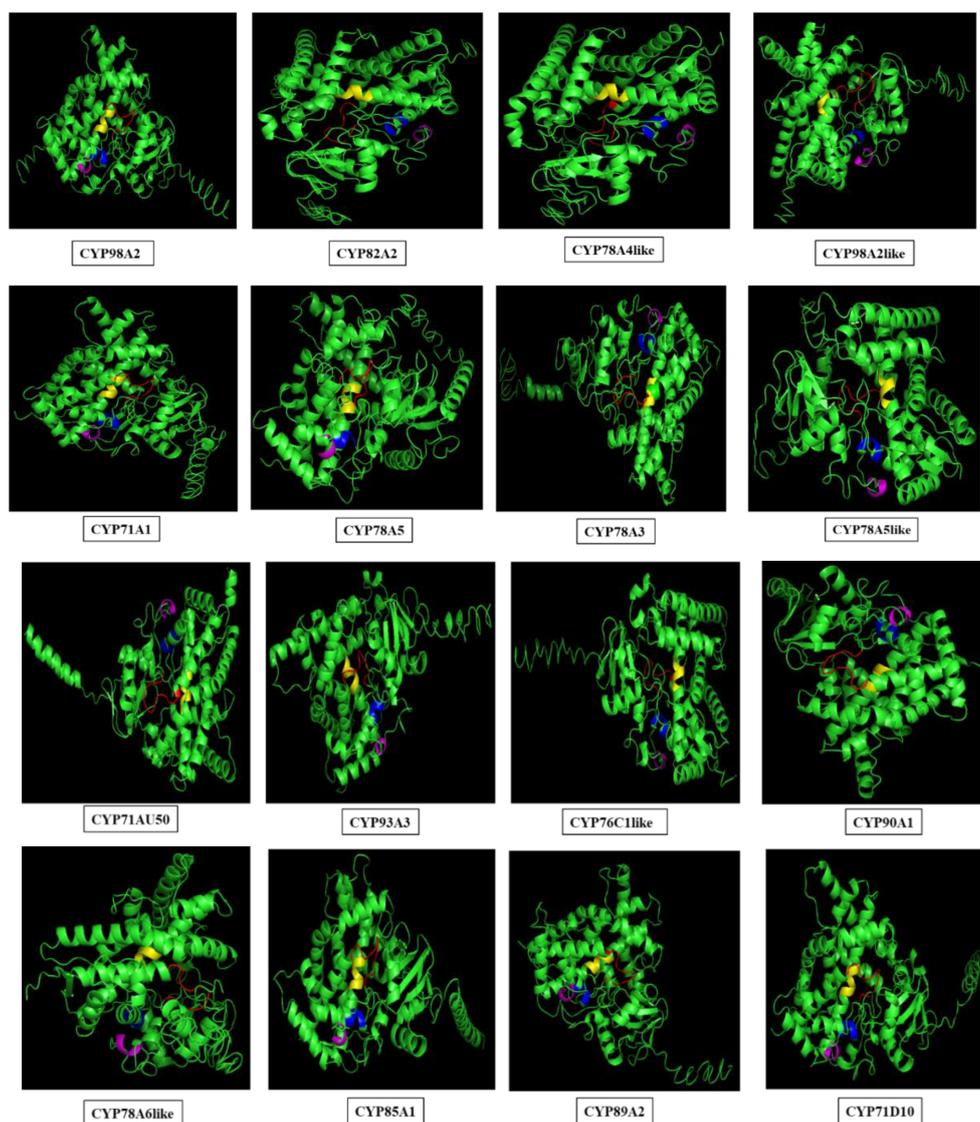
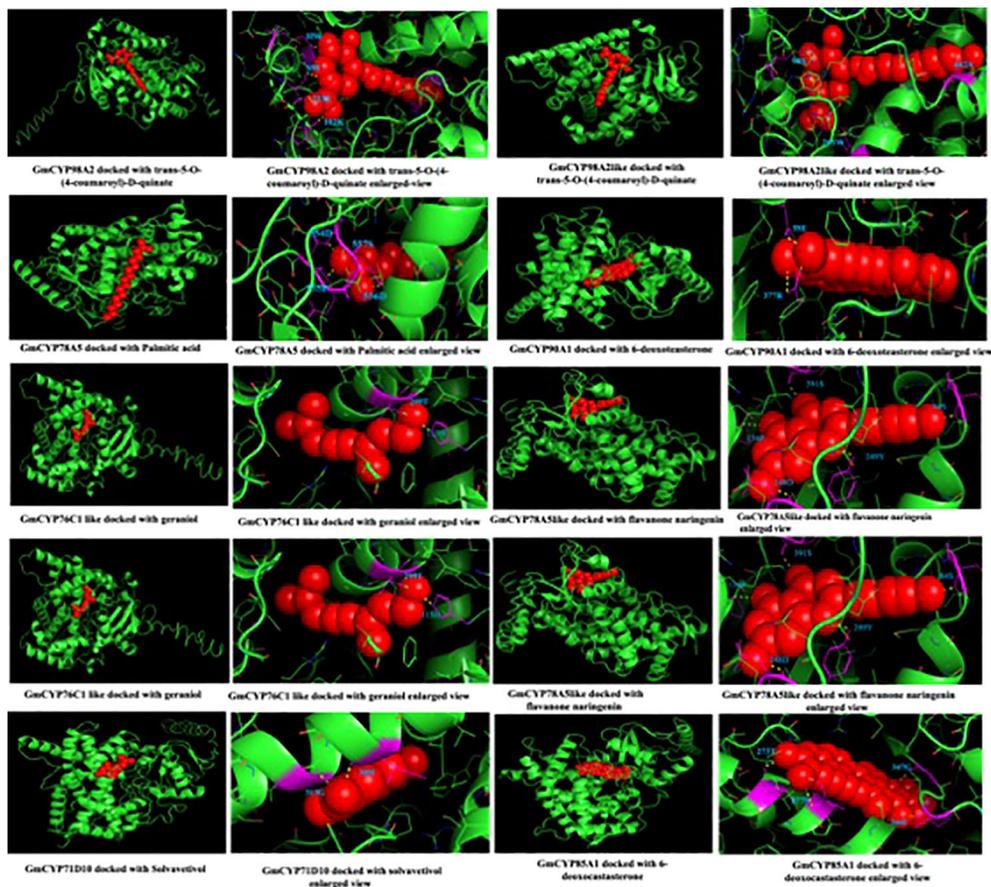
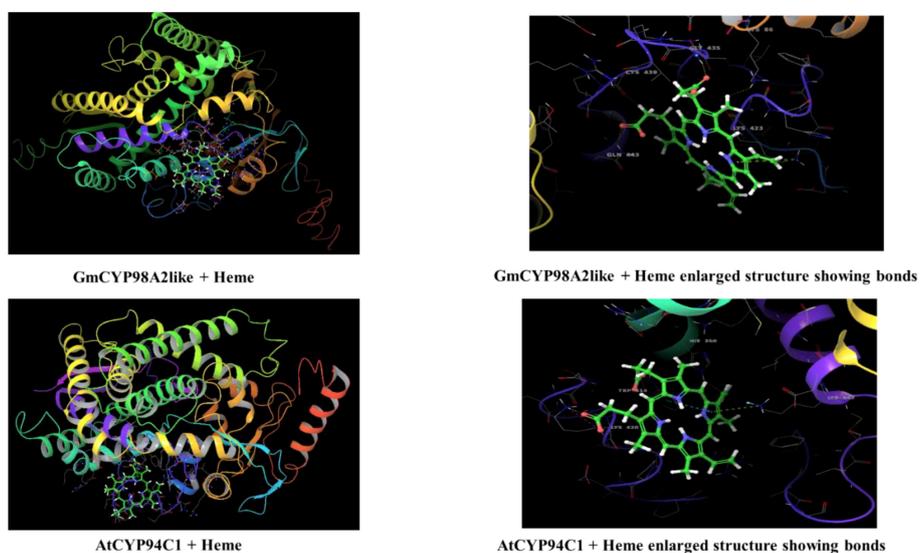


FIGURE 7

Three dimensional structures of selected 16 proteins. The structures are modelled using i-tasser server and conserved motifs were shown highlighted in red (FxxGxRxCxG), blue (ExxR), magenta (PxRW) and yellow (AGxDTT) color.



**FIGURE 8**  
Interaction of *GmCYPs* with their substrate using molecular docking. The figure shows the modeled tertiary structures of *GmCYPs* (green colored) and their interaction with substrates (red colored). The enlarged view showing docked substrate with *GmCYPs*, amino acid residues forming polar bonds were highlighted (magenta color) and labelled.



**FIGURE 9**  
Interaction of *GmCYP98A2like* with heme using molecular docking. The figure shows the modeled tertiary structures of *GmCYP* and its interaction with the heme prosthetic group. The enlarged view showing amino acid residues (labelled and highlighted) of protein forming bonds with heme group. interaction of *AtCYP94C1* with heme was also shown for the reference gene.

recognition of the particular stress and activating downstream signalling events. CYPs are versatile catalysts of a multi-gene superfamily, indulged in various growth, development-related processes and stress adaptation. CYPs reinforce the plant's defence mechanism by regulating homeostasis, biosynthesis of secondary metabolites (defensive chemicals) and phytohormones and supplementing antioxidant compounds to scavenge reactive oxygen species (Glawisching, 2007; Li et al., 2013). CYPs are linked with abiotic stress and build up stress tolerance against biotic factors, viz. bacteria, viruses, fungi, insects and nematodes (Wang et al., 2001; Du Fall and Solomon, 2013; Irmisch et al., 2015). Plants are deprived of nutrients due to pest-infestation, which eventually results in a massive loss in crop yield. Although plants lack a fully developed immune system, they have developed sophisticated defence mechanisms to fend off intruders, constitutive defense and inducible defense mechanisms (Karban et al., 1997; Kessler and Baldwin, 2002). Constitutive defense involves physical barriers such as lignified cell walls, thick cuticles, wax formation and synthesis of defensive chemicals. In contrast, inducible defense leads to the biosynthesis of defensive chemicals, phytoalexins, volatile deterrents, cell wall strengthening and alternation in the expression pattern of defense-related genes and crosstalk network of phytohormones viz. jasmonic acid, ethylene, salicylic acid (Thomma et al., 1998; Walling, 2000; Rojo et al., 2003). CYP450s are indulged in aforesaid plant defense events during biotic stress. However, a few studies have been conducted on the role of CYPs against herbivory in different plant species, role of *GmCYPs* in plant defense against *S. litura*- herbivory remain elusive.

Therefore, to designate the function of CYPs in soybean and to rule out their involvement in defense against *S. litura*, a comprehensive analysis of *S. litura*- inducible 16 *GmCYPs* (reported in transcriptome analysis) was executed. Comprehending the evolutionary trend is crucial for understanding the diverse functions of a gene/gene family. Phylogenetic analysis accomplished using maximum-likelihood method suggests that CYPs from *G. max* and *M. truncatula* are closely related, indicating that they are orthologous species, which might have undergone changes during the course of evolution. The high bootstrap values further aid the accuracy and reliability of tree construction. The phylogenetic tree besides assessing level of evolution through instigating soybean genome may aid in comprehending the putative functions of soybean CYPs, drawing insight from functionally characterized CYPs in *Medicago* (Ma et al., 2014; Yang et al., 2022). In addition, Gene duplication analysis identified more orthologous pairs of *GmCYPs* with *MtrCYPs* than with *AtCYPs*. These findings indicate that the majority of *GmCYPs* are likely to be originated from gene duplication events. However, in all homologous gene pairs, Ka/Ks ratio was observed to be less than 1, indicating duplicated genes underwent purifying selection throughout their evolutionary history (Kondrashov et al., 2002; Chitkara et al., 2022).

The examination of *GmCYPs* physiological properties unveiled their hydrophilic, basic and stable nature. The subcellular localization of *GmCYPs* was primarily predicted to be localized in the endoplasmic reticulum, with minimal expression in other organelles, coinciding with earlier studies that also predicted their

presence in the endoplasmic reticulum (Zhang and Li, 2017; Singh et al., 2021a). Examining exon-intron structure divulged that the structure of 16 *GmCYPs* is diverse with variable numbers of exons and introns, indicating pivotal evolutionary changes within the *G. max* genome. The gene *GmCYP89A2* lacks introns, while nine other *GmCYPs* contain a single intron, which speculated that these genes facilitate rapid transcriptional activity in response to adverse environmental conditions (Keshan et al., 2021; Wang et al., 2022). However, presence of fewer introns in *GmCYPs* indicates a lower level of evolutionary conservation (Paquette et al., 2000; Gorlova et al., 2014). The chromosomal mapping of *GmCYPs* showed that they were unevenly distributed across 13 soybean chromosomes, maximum on end positions. Genes with distal end position have been potential drivers of functional diversification while genes on proximal ends didn't undergo recombination (Bulger and Groudine, 2011). The knowledge of the gene structure and chromosomal location can be used to design more efficient breeding strategies for developing soybean cultivars with desired traits. The conserved motifs and domains are indicative of protein potential roles. Identification of the P450 domain (conserved cytochrome family) and four conserved motifs (cysteine heme-iron ligand signature motif, PXX motif, I-helix oxygen binding domain and K-helix region), confirming their membership in the cytochrome P450 superfamily and indicative of their potential functional roles and evolutionary conservation (Nelson et al., 1996; Li et al., 2007). The promoter region of a gene contains *cis*-elements that control the gene expression, which can vary during growth, development, and in response to environmental factors, in a plant organ-specific manner (Lata et al., 2014). *Cis*-acting elements in prediction analysis revealed the appearance of many light-responsive elements indicating their role in light signalling and photosynthesis. The presence of phytohormone-responsive *cis*-elements, including methyl-jasmonate, salicylic acid, gibberellin, and abscisic acid in the promoter region of *GmCYPs* reflecting induction of these genes under plant hormone pathway and innate immunity. In addition, the presence of low-temperature-responsive elements, drought-responsive elements, anaerobic condition-responsive elements, flavonoid biosynthesis regulating elements, defense-responsive elements, which implies that these genes could respond to and adapt to stress (Keshan et al., 2021). Earlier reports support the *cis*-acting element based regulation of the *GmCYPs* under stress, which can act as potential targets for crop improvement (Wang et al., 2022; Yang et al., 2022). Furthermore, various types of TFBSs were identified in promoters of *GmCYPs*. These TFs play crucial roles in the transcriptional regulation of diverse abiotic and biotic stress responses along with developmental processes (Nakano et al., 2006; Lijavetzky et al., 2003). The previous reports endorsed the fact that TFs have importance against environmental stress (Belamkar et al., 2014; Cui et al., 2019). However, prediction of 94 putative potential miRNA target sites in 16 *GmCYPs* reveals implications of miRNA-mediated stress regulation in soybean through post-transcriptional mechanism (Bhatia et al., 2019).

The study of mRNA abundance measurements for a family of genes in an organ-specific manner will enable the identification of

the genes engaged in regulatory or development processes specific to a given plant-organ type. Overall, the RNA-seq expression data analysis provided valuable insights into the pattern of expression specific to certain plant-organs of the listed *GmCYPs*, which could aid in understanding their functional roles in soybean growth and development (Ma et al., 2014; Saxena et al., 2023). The *GmCYPs* gene expression in different organs showed their functional divergence (Lopes-Caitar et al., 2013). Nevertheless, a significant number of *GmCYPs* showed minimal or no expression in various plant-organs, suggesting their involvement in alternative process.

Protein-protein interaction analysis (PPI) was executed with the STRING database to infer the intricate interplay of *GmCYPs* with other soybean proteins, enrichment analysis and the network. The result deduced that interactions were diverse, and 8 out of 16 *GmCYPs* proteins interact with other proteins involved in the phenylpropanoid pathway, secondary metabolite synthesis, pathogen responsive pathway and flavonoid biosynthesis, reported to defend the plant from biotic stress ranges from physical to chemical barrier, including signalling pathways (Dixon et al., 2002; Shah and Smith, 2020). They are also inferred to interact with proteins responsive to heavy metal tolerance, brassinolide biosynthesis and oxidative stress tolerance; predict their role in stress tolerance. However, the results of PPI are based on prediction and should be interpreted with caution. It would be necessary to conduct additional empirical investigations and experiments to determine whether these connections are physiologically significant.

Gene expression analysis during *S.litura*-infestation and wounding has been done in various plants like *Arabidopsis*, in chickpea, and soybean etc. to concede intricate changes in phytohormone and signalling pathway, attributing their role in plant defense against pest invasion. Earlier researches uphold the fact that, CYPs are involved in resistance against pest invasion. Resistance to *Myzus persicae* (green peach aphid) was regulated by camalexin synthesis by CYP family gene (Prince et al., 2014). Suppression of CYP hydroxylase in *N. tabacum* enhanced the resistance to *Myzus nicotianae* and CYP79D was also reported to enhance aldomixes in *Populus* sp. (Wang et al., 2001; Irmisch et al., 2015). However, gene expression changes during mechanical wounding were different from herbivory. These findings revealed that herbivory and wounding had distinct transcript patterns since along with cell wall disruption, herbivory also involves the ingestion of insect-derived elicitors. Previous reports have also shown a difference between the expression patterns during insect attacks and mechanical wounding (Reymond et al., 2000).

A complex interplay of phytohormones, such as jasmonate, ethylene and salicylic acid, serves as a crucial plant signal, evoked after insect pest or pathogen attack and activates defence genes. Plant pests mostly activate jasmonate/ethylene signalling pathway and also regulate salicylic acid pathway (Reymond et al., 2004). However, salicylic and jasmonate hormones induce the expression of different genes and act antagonistically. In order to acknowledge the role of these phytohormones, gene expression analysis of *GmCYPs* was examined upon their foliar treatment on plant. Alternation in *GmCYPs* expression level on treatment with these hormones demonstrate their role in regulation of *GmCYPs*. CYPs

are known to regulate jasmonate signalling pathway as alternation in JA gene expression was observed upon herbivory, wounding and pathogen attack (Blée, 2002; Bari and Jones, 2009). *AtCYP82C2* modulates defense against *Botrytis cinerea*, and *CYP82D2* regulates defense gene expression against *V. dahlia* by regulating the expression of jasmonate-induced defense-related genes (Reviewed by Singh et al., 2008).

The tertiary structure of selected 16 *GmCYPs* was predicted to study the interaction between *GmCYPs*, heme and also their substrates, and this revealed a variation in the modelled structures of *GmCYPs*, primarily because of the variation in the number and sequences of the amino acids present. Their different structural configurations could enable them to serve a particular function. However, the characterization and localization of evolutionarily conserved motifs in 16 *GmCYPs* strongly uphold the identity of *GmCYPs* as a part of CYP450 superfamily. Besides, the interaction of the heme cofactor with *GmCYP*, which is essential for their catalytic activity, was examined by selecting heme-binding region residues for generating a receptor grid. The global energy score of *GmCYP98A2*-like, docked with heme, was compared to *AtCYP94C1*, an enzyme involved in the jasmonate biosynthesis pathway (Heitz et al., 2012). Protein-protein docking was performed to examine the interaction between *GmCYPs* enzymes and their substrates. Global energy score for interaction between substrate and *GmCYP* varied from -25 to -50 but was comparable to receptor COX-2 and chalcone (Shilpa and Varalakshmi Devi, 2017). Evaluation of enzyme-substrate interaction showed the presence of polar bonds at the active sites of enzyme and substrate, which are stronger bonds that lower the activation energy of the enzyme-substrate complex and enhance the feasibility of reaction and product formation.

*GmCYP78A5*, *GmCYP78A3*, *GmCYP78A5-like*, and *GmCYP76C1-like* have induced gene expression on *S. litura*-infestation and methyl-jasmonate treatment, while *GmCYP78A5-like* and *GmCYP76C1-like* are also induced upon mechanical wounding. *GmCYP78A5-like* protein interaction was predicted with Alpha-ketoglutarate-dependent dioxygenases involved in flavonoid and ethylene biosynthesis. This protein also showed the presence of methyl jasmonate-responsive cis-element in its promoter region. *GmCYP78A5-like* is Flavonoid 3'-monooxygenase that catalyse formation of flavanone naringenin that showed detrimental effects on insect pests (Goławska et al., 2014). *GmCYP76C1-like* have methyl-jasmonate responsive cis-element, interacts with non-specific serine/threonine protein kinase; CYP b5-like steroid binding domain-containing protein. *GmCYP76C1-like*, known as geraniol 8-hydroxylase produces geraniol, which has repellent properties; availing as a natural pest control chemical with low toxicity (Chen and Viljoen, 2010). *GmCYP78A5* xenobiotic monooxygenase synthesizes palmitic acid (PA), which metabolises xenobiotics and reduces the incidence of soil-borne diseases (Carta et al., 2017). *GmCYP78A5* has been predicted to interact with Delta(7)-sterol-C5(6)-desaturase involved in the biosynthesis of sterols and methyl jasmonate responsive cis-element (Feng et al., 2017; Saini and Kumar, 2019). Further, functional characterization of these genes in soybean will confirm their roles in insect resistance and will be used in molecular breeding and genetic engineering.

## 5 Conclusion

CYPs are multifunctional catalysts involved in various physiological processes of plants and contribute significantly to plant defense through the biosynthesis of hormones, secondary metabolites, fatty acids and cell wall components. Recent characterization of various CYPs has revealed their potential to exploit in agricultural improvement to create stress-tolerant plants. Even though the entire soybean genome sequence has been accessible since 2010 and the significance of CYP450s in plant metabolic pathways have long been understood, only a few P450s in soybean have been evaluated to characterize their role in herbivory. In our study, an extensive analysis of biotic stress-responsive *GmCYPs* was performed to deduce their role in soybean defense against *S. litura* herbivory. The phylogenetic study and motif analysis revealed their relatedness to other *GmCYPs* from model plants. Identification of biotic stress-responsive cis-element in the promoter region of *GmCYPs*, PPI analysis and their differential expression upon *S. litura* infestation, and treatment of signalling compounds proposed their significant function in plant defense. Further, structural characterization by interaction study between *GmCYPs* and heme cofactor and enzyme-substrate interactions confirmed their identity as heme-binding proteins, and allowed us to infer thermodynamic feasibility of reactions. *GmCYP78A5like*, *GmCYP76C1* are concluded as potential candidates involved in soybean-*S. litura* interaction. The findings of this study can be expanded to confirm the underlying mechanism of *GmCYPs* in plants during *S. litura* -infestation. Further *in vivo* functional and structural characterization of these genes employing techniques such as genome editing (overexpression-RNAi) and traditional genetic breeding to confirm their function in plant defense is highly recommended.

## Data availability statement

The original contributions presented in the study are included in the article/Supplementary Material. Further inquiries can be directed to the corresponding authors.

## Author contributions

AS and IS conceptualized and supervised the study. MY and RP performed the experiments, collected the data and analyzed the

results. MY, RP and AnR wrote the manuscript. AS, IS, AC and AmR critically reviewed the manuscript. AS, IKS, AC and AmR contributed to editing and visualization. AS, IS and AmR contributed to the formal analysis. AS, IS and AmR contributed to funding acquisition. IS and AS contributed to resources. All authors contributed to the article and approved the submitted version.

## Funding

Financial support was provided to AS by the Science and Engineering Research Board (SERB), Department of Science and Technology, New Delhi, India (ECR/2017/002478; SPG/2021/002969). MY is grateful to CSIR, MHRD for the JRF fellowship. The authors acknowledge Prof. (Dr.) Rama, Principal, Hansraj College, for providing the facilities and her constant support for research work. AmR, AC, financed by EVA 4.0," No. CZ.02.1.01/0.0/0.0/16 019/0000803 and "EXTEMIT – K," No. CZ.02.1.01/0.0/0.0/15\_003/0000433 financed by OP RDE. AmR and AC are also supported by "Excellent team grants (2023-24)" from FLD, CZU.

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2023.1221526/full#supplementary-material>

## References

- Abbas, A., Shah, A. N., Shah, A. A., Nadeem, M. A., Alsaleh, A., Javed, T., et al. (2022). Genome-wide analysis of invertase gene family, and expression profiling under abiotic stress conditions in potato. *Biol. (Basel)*. 11, 539. doi: 10.3390/biology11040539
- Badole, S. L., and Bodhankar, S. L. (2012). *Glycine max* (Soybean) treatment for diabetes. *Bioact. Food as Diet. Interv. Diabetes Bioact. Foods. Chronic. Dis. States* 77, 77–82. doi: 10.1016/b978-0-12-397153-1.00008-1
- Bailey, T. L., Boden, M., Buske, F. A., Frith, M., Grant, C. E., Clementi, L., et al. (2009). MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res.* 37, W202–W208. doi: 10.1093/nar/gkp335
- Bari, R., and Jones, J. D. G. (2009). Role of plant hormones in plant defence responses. *Plant Mol. Biol.* 69, 473–488. doi: 10.1007/s11103-008-9435-0
- Barja, M. V., and Rodriguez-Concepcion, M. (2021). Plant geranylgeranyl diphosphate synthases: every (gene) family has a story. *Abiotech* 2, 289–298. doi: 10.1007/s42994-021-00050-5
- Baskar, K., Muthu, C., Raj, G. A., Kingsley, S., and Ignacimuthu, S. (2012). Ovicidal activity of *Atalantia monophylla* (L) Correa against *Spodoptera litura* Fab. (Lepidoptera: Noctuidae). *Asian Pac. J. Trop. Biomed.* 2, 987–991. doi: 10.1016/S2221-1691(13)60011-8

- Belamkar, V., Weeks, N. T., Bharti, A. K., Farmer, A. D., Graham, M. A., and Cannon, S. B. (2014). Comprehensive characterization and RNA-Seq profiling of the HD-Zip transcription factor family in soybean (*Glycine max*) during dehydration and salt stress. *BMC Genomics* 15, 1–25. doi: 10.1186/1471-2164-15-950
- Bhachoo, J., and Beuming, T. (2017). Investigating protein–peptide interactions using the Schrödinger computational suite. *Model. Pept. Interact. Methods Protoc.*, 1561, 235–254. doi: 10.1007/978-1-4939-6798-8\_14
- Bhatia, G., Sharma, S., Upadhyay, S. K., and Singh, K. (2019). Long non-coding RNAs coordinate developmental transitions and other key biological processes in grapevine. *Sci. Rep.* 9, 1–14. doi: 10.1038/s41598-019-38989-7
- Blée, E. (2002). Impact of phyto-oxylipins in plant defense. *Trends Plant Sci.* 7, 315–322. doi: 10.1016/S1360-1385(02)02290-2
- Brahman, S. K., Awasthi, A. K., and Singh, S. (2018). Studies on insectpests of soybean (*Glycine max*) with special reference to seasonal incidence of Lepidopteran defoliators. *J. Pharmacogn. Phytochem.* 7, 1808–1811.
- Bughio, A. R., Hussain, T., Qureshi, Z. A., Ahmad, M., and Shakoori, A. R. (1994). Quantitative studies of food consumption and growth of *spodoptera litura* (F.) on soybean. *Proc. Pakistan Congress. Zool.* 12, 99–104.
- Bulger, M., and Groudine, M. (2011). Functional and mechanistic diversity of distal transcription enhancers. *Cell* 144, 327–339. doi: 10.1016/j.cell.2011.01.024
- Butrón Gómez, A. M., Chen, Y.-C., Rottinghaus, G. E., and McMullen, M. D. (2010). Genetic variation at bx1 controls DIMBOA content in maize. *Theoretical and Applied Genetics*. 120, 721–734
- Carrão-Panizzi, M. C., and Erhan, S. Z. (2007). Environmental and genetic variation of soybean tocopherol content under Brazilian growing conditions. *J. Am. Oil Chem. Soc.* 84 (10), 921–928. doi: 10.1007/s11746-007-1128-3
- Carta, G., Murru, E., Banni, S., and Manca, C. (2017). Palmitic acid: physiological role, metabolism and nutritional implications. *Front. Physiol.* 8, 902. doi: 10.3389/fphys.2017.00902
- Chauhan, O. P., Chauhan, G. S., Singh, G., Kumbhar, B. K., and Mishra, D. P. (2002). Varietal variability in the contents of nutrients and anti-nutrients in different parts of soybean seeds. *J. Rural Agric. Res.* 2, 42–50.
- Chen, C., Chen, H., Zhang, Y., Thomas, H. R., Frank, M. H., He, Y., et al. (2020). TBtools: an integrative toolkit developed for interactive analyses of big biological data. *Mol. Plant* 13, 1194–1202. doi: 10.1016/j.molp.2020.06.009
- Chen, W., and Viljoen, A. M. (2010). Geraniol—a review of a commercially important fragrance material. *South Afr. J. Bot.* 76, 643–651. doi: 10.1016/j.sajb.2010.05.008
- Cheng, A.-X., Han, X.-J., Wu, Y.-F., and Lou, H.-X. (2014). The function and catalysis of 2-oxoglutarate-dependent oxygenases involved in plant flavonoid biosynthesis. *Int. J. Mol. Sci.* 15, 1080–1095. doi: 10.3390/ijms15011080
- Chitkara, P., Poddar, N., Singh, A., and Kumar, S. (2022). BURP domain-containing genes in legumes: genome-wide identification, structure, and expression analysis under stresses and development. *Plant Biotech. Rep.* 16 (4), 369–388. doi: 10.1007/s11816-022-00752-2
- Coon, M. J. (2005). Cytochrome P450: nature's most versatile biological catalyst. *Annu. Rev. Pharmacol. Toxicol.* 45, 1–25. doi: 10.1146/annurev.pharmtox.45.120403.100030
- Cui, X., Yan, Q., Gan, S., Xue, D., Wang, H., Xing, H., et al. (2019). GmWRKY40, a member of the WRKY transcription factor genes identified from *Glycine max* L., enhanced the resistance to *Phytophthora sojae*. *BMC Plant Biol.* 19, 1–15. doi: 10.1186/s12870-019-2132-0
- Cui, Z.-L., Gai, J.-Y., Ji, D.-F., and Ren, Z.-J. (1997). A study on leaf-feeding insect species on soybeans in nanjing area. *Soybean Sci.* 16 (1), 12–20.
- Dai, X., and Zhao, P. X. (2011). psRNAtarget: a plant small RNA target analysis server. *Nucleic Acids Res.* 39, W155–W159. doi: 10.1093/nar/gkr319
- DeLano, W. L. (2002). Pymol: An open-source molecular graphics tool. *CCP4. Newsl. Protein Crystallogr.* 40, 82–92.
- Dick, R., Rattei, T., Haslbeck, M., Schwab, W., Gierl, A., and Frey, M. (2012). Comparative analysis of benzoxazinoid biosynthesis in monocots and dicots: independent recruitment of stabilization and activation functions. *Plant Cell* 24, 915–928. doi: 10.1105/tpc.112.096461
- Dixon, R. A., Achnine, L., Kota, P., Liu, C., Reddy, M. S. S., and Wang, L. (2002). The phenylpropanoid pathway and plant defence—a genomics perspective. *Mol. Plant Pathol.* 3, 371–390. doi: 10.1046/j.1364-3703.2002.00131.x
- Du Fall, L. A., and Solomon, P. S. (2013). The necrotrophic effector SnToxA induces the synthesis of a novel phytoalexin in wheat. *New Phytol.* 200, 185–200. doi: 10.1111/nph.12356
- Durst, F., and Nelson, D. R. (1995). Diversity and evolution of plant P450 and P450-reductases. *Drug Metabol. Drug Interact.* 12, 189–206. doi: 10.1515/DMDI.1995.12.3-4.189
- Farmer, E. E., and Goossens, A. (2019). Jasmonates: what ALLENE OXIDE SYNTHASE does for plants. *J. Exp. Bot.* 70, 3373–3378. doi: 10.1093/jxb/erz254
- Feng, J., Dong, Y., Liu, W., He, Q., Daud, M. K., Chen, J., et al. (2017). Genome-wide identification of membrane-bound fatty acid desaturase genes in *Gossypium hirsutum* and their expressions during abiotic stress. *Sci. Rep.* 7, 45711. doi: 10.1038/srep45711
- Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M. R., Appel, R. D., et al. (2005). *Protein identification and analysis tools on the ExPASy server* (Humana press: Springer), pp. 571–607.
- Glawischnig, E. (2007). Camalexin. *Phytochemistry* 68, 401–406. doi: 10.1016/j.phytochem.2006.12.005
- Golawska, S., Sprawka, I., Lukasik, I., and Golawski, A. (2014). Are naringenin and quercetin useful chemicals in pest-management strategies? *J. Pest Sci.* (2004), 87, 173–180. doi: 10.1007/s10340-013-0535-5
- Gorlova, O., Fedorov, A., Logothetis, C., Amos, C., and Gorlov, I. (2014). Genes with a large intronic burden show greater evolutionary conservation on the protein level. *BMC Evol. Biol.* 14, 50. doi: 10.1186/1471-2148-14-50
- Guengerich, F. P. (2001). Common and uncommon cytochrome P450 reactions related to metabolism and chemical toxicity. *Chem. Res. Toxicol.* 14, 611–650. doi: 10.1021/tx0002583
- Guengerich, F. P., and Munro, A. W. (2013). Unusual cytochrome P450 enzymes and reactions. *J. Biol. Chem.* 288, 17065–17073. doi: 10.1074/jbc.R113.462275
- Hamberger, B., Ohnishi, T., Hamberger, B., Séguin, A., and Bohlmann, J. (2011). Evolution of diterpene metabolism: Sitka spruce CYP720B4 catalyzes multiple oxidations in resin acid biosynthesis of conifer defense against insects. *Plant Physiol.* 157, 1677–1695. doi: 10.1104/pp.111.185843
- Heitz, T., Widemann, E., Lugan, R., Miesch, L., Ullmann, P., Désaubry, L., et al. (2012). Cytochromes P450 CYP94C1 and CYP94B3 catalyze two successive oxidation steps of plant hormone jasmonoyl-isoleucine for catabolic turnover. *J. Biol. Chem.* 287, 6296–6306. doi: 10.1074/jbc.M111.316364
- Hoffmann, L., Besseau, S., Geoffroy, P., Ritzenthaler, C., Meyer, D., Lapierre, C., et al. (2004). Silencing of hydroxycinnamoyl-coenzyme A shikimate/quinate hydroxycinnamoyltransferase affects phenylpropanoid biosynthesis. *Plant Cell* 16, 1446–1465. doi: 10.1105/tpc.020297
- Hu, B., Jin, J., Guo, A.-Y., Zhang, H., Luo, J., and Gao, G. (2015). GSDS 2.0: an upgraded gene feature visualization server. *Bioinformatics* 31, 1296–1297. doi: 10.1093/bioinformatics/btu817
- Innan, H., and Kondrashov, F. (2010). The evolution of gene duplications: classifying and distinguishing between models. *Nat. Rev. Genet.* 11, 97–108. doi: 10.1038/nrg2689
- Irmisch, S., Clavijo McCormick, A., Boeckler, G. A., Schmidt, A., Reichelt, M., Schneider, B., et al. (2013). Two herbivore-induced cytochrome P450 enzymes CYP79D6 and CYP79D7 catalyze the formation of volatile aldoximes involved in poplar defense. *Plant Cell* 25, 4737–4754. doi: 10.1105/tpc.113.118265
- Irmisch, S., Zeltner, P., Handrick, V., Gershenzon, J., and Köllner, T. G. (2015). The maize cytochrome P450 CYP79A61 produces phenylacetaldoxime and indole-3-acetaldoxime in heterologous systems and might contribute to plant defense and auxin formation. *BMC Plant Biol.* 15, 1–14. doi: 10.1186/s12870-015-0526-1
- Jones-Rhoades, M. W., Bartel, D. P., and Bartel, B. (2006). MicroRNAs and their regulatory roles in plants. *Annu. Rev. Plant Biol.* 57, 19–53. doi: 10.1146/annurev.plant.57.032905.105218
- Karban, R., Agrawal, A. A., and Mangel, M. (1997). The benefits of induced defenses against herbivores. *Ecology* 78, 1351–1355. doi: 10.1890/0012-9658(1997)078[1351:TBOIDA]2.0.CO;2
- Kato-Noguchi, H., Sakata, Y., Takenokuchi, K., Kosemura, S., and Yamamura, S. (2000). Allelopathy in Maize I: Isolation and identification of allelochemicals in maize seedlings. *Plant Prod. Sci.* 3, 43–46. doi: 10.1626/pp.3.43
- Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N., and Sternberg, M. J. E. (2015). The Phyre2 web portal for protein modeling, prediction and analysis. *Nat. Protoc.* 10, 845–858. doi: 10.1038/nprot.2015.053
- Keshan, R., Singh, I. K., and Singh, A. (2021). Genome wide investigation of MAPKKs from *Cicer arietinum* and their involvement in plant defense against *Helicoverpa armigera*. *Physiol. Mol. Plant Pathol.* 115, 101685. doi: 10.1016/j.pmpp.2021.101685
- Kessler, A., and Baldwin, I. T. (2002). Plant responses to insect herbivory. *Annu. Rev. Plant Biol.* 53, 299–328. doi: 10.1146/annurev.arplant.53.100301.135207
- Khatrı, P., Wally, O., Rajcan, I., and Dhaubhadel, S. (2022). Comprehensive analysis of cytochrome P450 monooxygenases reveals insight into their role in partial resistance against *Phytophthora sojae* in soybean. *Front. Plant Sci.* 906. doi: 10.3389/fpls.2022.862314
- Kim, S., Chen, J., Cheng, T., Gindulyte, A., He, J., He, S., et al. (2019). PubChem 2019 update: improved access to chemical data. *Nucleic Acids Res.* 47, D1102–D1109. doi: 10.1093/nar/gky1033
- Kondrashov, F. A., Rogozin, I. B., Wolf, Y. I., and Koonin, E. V. (2002). Selection in the evolution of gene duplications. *Genome Biol.* 3, RESEARCH0008. doi: 10.1186/gb-2002-3-2-research0008
- Krzywinski, M., Schein, J., Birol, I., Connors, J., Gascoyne, R., Horsman, D., et al. (2009). Circos: An information aesthetic for comparative genomics. *Genome Res.* 19 (9), 1639–1645. doi: 10.1101/gr.092759.109
- Kumar, S., Stecher, G., and Tamura, K. (2016). MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol. Biol. Evol.* 33, 1870–1874. doi: 10.1093/molbev/msw054
- Lata, C., Mishra, A. K., Muthamilarasan, M., Bonthala, V. S., Khan, Y., and Prasad, M. (2014). Genome-wide investigation and expression profiling of AP2/ERF

- transcription factor superfamily in foxtail millet (*Setaria italica* L.). *PLoS One* 9, e113092. doi: 10.1371/journal.pone.0113092
- Lee, S., Badieyan, S., Bevan, D. R., Herde, M., Gatz, C., and Tholl, D. (2010). Herbivore-induced and floral homoterpenoid volatiles are biosynthesized by a single P450 enzyme (CYP82G1) in Arabidopsis. *Proc. Natl. Acad. Sci.* 107, 21205–21210. doi: 10.1073/pnas.1009975107
- Lescot, M., Déhais, P., Thijs, G., Marchal, K., Moreau, Y., Van De Peer, 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 Res.* 30, 325–327. doi: 10.1093/nar/30.1.325
- Letunic, I., and Bork, P. (2007). Interactive Tree Of Life (iTOL): an online tool for phylogenetic tree display and annotation. *Bioinformatics* 23, 127–128. doi: 10.1093/bioinformatics/btl529
- Li, L., Chang, Z., Pan, Z., Fu, Z.-Q., and Wang, X. (2008). Modes of heme binding and substrate access for cytochrome P450 CYP74A revealed by crystal structures of allene oxide synthase. *Proc. Natl. Acad. Sci.* 105, 13883–13888. doi: 10.1073/pnas.0804099105
- Li, L., Cheng, H., Gai, J., and Yu, D. (2007). Genome-wide identification and characterization of putative cytochrome P450 genes in the model legume *Medicago truncatula*. *Planta* 226, 109–123. doi: 10.1007/s00425-006-0473-z
- Li, W., Shao, M., Yang, J., Zhong, W., Okada, K., Yamane, H., et al. (2013). Oस्प71Z2 involves diterpenoid phytoalexin biosynthesis that contributes to bacterial blight resistance in rice. *Plant Sci.* 207, 98–107. doi: 10.1016/j.plantsci.2013.02.005
- Lijavetzky, D., Carbonero, P., and Vicente-Carbajosa, J. (2003). Genome-wide comparative phylogenetic analysis of the rice and Arabidopsis Dof gene families. *BMC Evol. Biol.* 07, 3–17. doi: 10.1186/1471-2148-3-17
- Lopes-Caitar, V. S., de Carvalho, M. C. C. G., Darben, L. M., Kuwahara, M. K., Nepomuceno, A. L., Dias, W. P., et al. (2013). Genome-wide analysis of the Hsp 20 gene family in soybean: comprehensive sequence, genomic organization and expression profile analysis under abiotic and biotic stresses. *BMC Genomics* 14, 1–17. doi: 10.1186/1471-2164-14-577
- Luo, P., Wang, Y., Wang, G., Essenberg, M., and Chen, X. (2001). Molecular cloning and functional identification of (+)- $\delta$ -cadinene-8-hydroxylase, a cytochrome P450 mono-oxygenase (CYP706B1) of cotton sesquiterpene biosynthesis. *Plant J.* 28, 95–104. doi: 10.1046/j.1365-313X.2001.01133.x
- Ma, B., Luo, Y., Jia, L., Qi, X., Zeng, Q., Xiang, Z., et al. (2014). Genome-wide identification and expression analyses of cytochrome P450 genes in mulberry (*Morus notabilis*). *J. Integr. Plant Biol.* 56 (9), 887–901. doi: 10.1111/jipb.12141
- Malviya, R., Dey, S., Pandey, A., and Gayen, D. (2023). Genome-wide identification and expression pattern analysis of lipoxygenase genes of chickpea (*Cicer arietinum* L.) in response to accelerated aging. *Gene* 874, 147482. doi: 10.1016/j.gene.2023.147482
- Maree, J. M., Kallar, S. A., and Khuroo, R. D. (1999). Relative Abundance of *Spodoptera litura* F. and *Agrotis ypsilon* Rott. on Cabbage. *Pak. J. Zool.* 31 (1), 31–34.
- Marwoto, S. (2008). Strategy and control technology components of armyworm (*Spodoptera litura* Fabricius) on soybean. *J. Agric. Res. Dev.* 27, 131–136.
- Mashiach, E., Schneidman-Duhovny, D., Andrusier, N., Nussinov, R., and Wolfson, H. J. (2008). FireDock: a web server for fast interaction refinement in molecular docking. *Nucleic Acids Res.* 36, W229–W232. doi: 10.1093/nar/gkn186
- Mistry, J., Chuguransky, S., Williams, L., Qureshi, M., Salazar, G. A., Sonnhammer, E. L. L., et al. (2021). Pfam: The protein families database in 2021. *Nucleic Acids Res.* 49, D412–D419. doi: 10.1093/nar/gkaa913
- Nakano, T., Suzuki, K., Fujimura, T., and Shinshi, H. (2006). Genome-wide analysis of the ERF gene family in Arabidopsis and rice. *Plant Physiol.* 140 (2), 411–432. doi: 10.1104/pp.105.073783
- Nelson, D. R. (2006). Cytochrome P450 nomenclature 2004. *Cytochrome. P450. Protoc.*, 1–10. doi: 10.1385/1-59259-998-2:1
- Nelson, D. R. (2009). The cytochrome p450 homepage. *Hum. Genomics* 4, 1–7. doi: 10.1186/1479-7364-4-1-59
- Nelson, D. R., Koymans, L., Kamatani, T., Stegeman, J. J., Feyereisen, R., Waxman, D. J., et al. (1996). P450 superfamily: update on new sequences, gene mapping, accession numbers and nomenclature. *Pharmacogenetics* 6, 1–42. doi: 10.1097/00008571-199602000-00002
- O'Boyle, N.M., Banck, M., James, C.A., et al. (2011). Open Babel: An open chemical toolbox. *J. Cheminform* 3, 33. doi: 10.1186/1758-2946-3-33
- Pandian, B. A., Sathishraj, R., Djanaguiraman, M., Prasad, P. V. V., and Jugulam, M. (2020). Role of cytochrome P450 enzymes in plant stress response. *Antioxidants* 9, 454. doi: 10.3390/antiox9050454
- Paquette, S. M., Bak, S., and Feyereisen, R. (2000). Intron-exon organization and phylogeny in a large superfamily, the paralogous cytochrome P450 genes of *Arabidopsis thaliana*. *DNA Cell Biol.* 19, 307–317. doi: 10.1089/10445490050021221
- Patil, G., Vuong, T. D., Kale, S., Valliyodan, B., Deshmukh, R., Zhu, C., et al. (2018). Dissecting genomic hotspots underlying seed protein, oil, and sucrose content in an interspecific mapping population of soybean using high-density linkage mapping. *Plant Biotechnol. J.* 16 (11), 1939–1953. doi: 10.1111/pbi.12929
- Pinot, F., and Beisson, F. (2011). Cytochrome P450 metabolizing fatty acids in plants: characterization and physiological roles. *FEBS J.* 278, 195–205. doi: 10.1111/j.1742-4658.2010.07948.x
- Prince, D. C., Drurey, C., Zipfel, C., and Hogenhout, S. A. (2014). The leucine-rich repeat receptor-like kinase BRASSINOSTEROID INSENSITIVE1-ASSOCIATED KINASE1 and the cytochrome P450 PHYTOALEXIN DEFICIENT3 contribute to innate immunity to aphids in Arabidopsis. *Plant Physiol.* 164, 2207–2219. doi: 10.1104/pp.114.235598
- Reymond, P., Bodenhausen, N., Van Poecke, R. M. P., Krishnamurthy, V., Dicke, M., and Farmer, E. E. (2004). A conserved transcript pattern in response to a specialist and a generalist herbivore. *Plant Cell* 16, 3132–3147. doi: 10.1105/tpc.104.026120
- Reymond, P., Weber, H., Damond, M., and Farmer, E. E. (2000). Differential gene expression in response to mechanical wounding and insect feeding in Arabidopsis. *Plant Cell* 12, 707–719. doi: 10.1105/tpc.12.5.707
- Rojo, E., Solano, R., and Sánchez-Serrano, J. J. (2003). Interactions between signaling compounds involved in plant defense. *J. Plant Growth Regul.* 22, 82–98. doi: 10.1007/s00344-003-0027-6
- Saini, R., and Kumar, S. (2019). Genome-wide identification, characterization and in-silico profiling of genes encoding FAD (fatty acid desaturase) proteins in chickpea (*Cicer arietinum* L.). *Plant Gene* 18, 100180. doi: 10.1016/j.plgene.2019.100180
- Saraiva, K. D., Fernandes de Melo, D., Morais, V. D., Vasconcelos, I. M., and Costa, J. H. (2014). Selection of suitable soybean EF1 $\alpha$  genes as internal controls for real-time PCR analyses of tissues during plant development and under stress conditions. *Plant Cell Rep.* 33 (9), 1453–1465. doi: 10.1007/s00299-014-1628-1
- Saxena, H., Negi, H., Keshan, R., Chitkara, P., Kumar, S., Chakraborty, A., et al. (2023). A comprehensive investigation of lipid-transfer proteins from *Cicer arietinum* disentangles their role in plant defense against *Helicoverpa armigera*-infestation. *Front. Genet.* 14. doi: 10.3389/fgenet.2023.1195554
- Schneidman-Duhovny, D., Inbar, Y., Nussinov, R., and Wolfson, H. J. (2005). PatchDock and SymmDock: servers for rigid and symmetric docking. *Nucleic Acids Res.* 33, W363–W367. doi: 10.1093/nar/glk481
- Shah, A., and Smith, D. L. (2020). Flavonoids in agriculture: Chemistry and roles in, biotic and abiotic stress responses, and microbial associations. *Agronomy* 10, 1209. doi: 10.3390/agronomy10081209
- Shilpa, T., and Varalakshmi Devi, K. (2017). Molecular docking and synthesis of 5-acetylpyrimidine 2, 4, 6 trione based chalcones. *World J. Pharm. Res.* 7, 664–673. doi: 10.20959/wjpr20183-10819
- Singh, A., Panwar, R., Mittal, P., Hassan, M. I., and Singh, I. K. (2021a). Plant cytochrome P450s: Role in stress tolerance and potential applications for human welfare. *Int. J. Biol. Macromol.* 184, 874–886. doi: 10.1016/j.ijbiomac.2021.06.125
- Singh, A., Singh, S., Singh, R., Kumar, S., Singh, S. K., and Singh, I. K. (2021b). Dynamics of *Zea mays* transcriptome in response to a polyphagous herbivore, *Spodoptera litura*. *Funct. Integr. Genomics* 21, 571–592. doi: 10.1007/s10142-021-00796-7
- Singh, A., Singh, I. K., and Verma, P. K. (2008). Differential transcript accumulation in *Cicer arietinum* L. @ in response to a chewing insect *Helicoverpa armigera* and defense regulators correlate with reduced insect performance. *J. Exp. Bot.* 59, 2379–2392. doi: 10.1093/jxb/ern111
- Singh, A., Tyagi, C., Nath, O., and Singh, I. K. (2018). Helicoverpa-inducible Thioredoxin h from *Cicer arietinum*: structural modeling and potential targets. *Int. J. Biol. Macromol.* 109, 231–243. doi: 10.1016/j.ijbiomac.2017.12.079
- Soyatats. (2010). *A Reference Guide to Important Soybean Facts and Figure*. American Soybean Association
- Sundar, B., Rashmi, V., Sumith, H. K., and Sandhya, S. (2018). Study the incidence and period of activity of *Spodoptera litura* on soybean. *J. Entomol. Zool. Stud.* 6, 331–333.
- Suyama, M., Torrents, D., and Bork, P. (2006). PAL2NAL: Robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucleic Acids Res.* 1, W609–W612. doi: 10.1093/nar/gkl315
- Szklarczyk, D., Gable, A. L., Nastou, K. C., Lyon, D., Kirsch, R., Pyysalo, S., et al. (2021). The STRING database in 2021: customizable protein-protein networks, and functional characterization of user-uploaded gene/measurement sets. *Nucleic Acids Res.* 49, D605–D612. doi: 10.1093/nar/gkab835
- Thomma, B. P. H. J., Eggermont, K., Penninckx, I. A. M. A., Mauch-Mani, B., Vogelsang, R., Cammue, B. P. A., et al. (1998). Separate jasmonate-dependent and salicylate-dependent defense-response pathways in Arabidopsis are essential for resistance to distinct microbial pathogens. *Proc. Natl. Acad. Sci.* 95, 15107–15111. doi: 10.1073/pnas.95.25.15107
- Vasudev, A., and Sohal, S. (2016). Partially purified *Glycine max* proteinase inhibitors: potential bioactive compounds against tobacco cutworm, *Spodoptera litura* (Fabricius 1775)(Lepidoptera: Noctuidae). *Turkish. J. Zool.* 40, 379–387. doi: 10.3906/zoo-1508-20
- Waese, J., Fan, J., Pasha, A., Yu, H., Fucile, G., Shi, R., et al. (2017). ePlant: visualizing and exploring multiple levels of data for hypothesis generation in plant biology. *Plant Cell* 29, 1806–1821. doi: 10.1105/tpc.17.00073
- Walling, L. L. (2000). The myriad plant responses to herbivores. *J. Plant Growth Regul.* 19, 195–216. doi: 10.1007/s003440000026
- Wang, Y., Tang, H., Debarry, J. D., Tan, X., Li, J., Wang, X., et al. (2012). MScanX: A toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Res.* 40 (7) e49. doi: 10.1093/nar/gkr1293

- Wang, E., Wang, R., DeParasis, J., Loughrin, J. H., Gan, S., and Wagner, G. J. (2001). Suppression of a P450 hydroxylase gene in plant trichome glands enhances natural-product-based aphid resistance. *Nat. Biotechnol.* 19, 371–374. doi: 10.1038/86770
- Wang, Y., Wang, H., Fan, R., Yang, Q., and Yu, D. (2014). Transcriptome analysis of soybean lines reveals transcript diversity and genes involved in the response to common cutworm (*Spodoptera litura* Fabricius) feeding. *Plant. Cell Environ.* 37, 2086–2101. doi: 10.1111/pce.12296
- Wang, C., Xue, L., Cui, Q., Liu, Q., Zhang, J., and Rao, G. (2022). Genome-wide identification of the cytochrome P450 superfamily in *Olea europaea* helps elucidate the synthesis pathway of oleuropein to improve the quality of olive oil. *Sci. Hortic. (Amsterdam)*. 304, 111291. doi: 10.1016/j.scienta.2022.111291
- Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R., et al. (2018). SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Res.* 46, W296–W303. doi: 10.1093/nar/gky427
- Watson, J. D., Laskowski, R. A., and Thornton, J. M. (2005). Predicting protein function from sequence and structural data. *Curr. Opin. Struct. Biol.* 15, 275–284. doi: 10.1016/j.sbi.2005.04.003
- Werck-Reichhart, D., and Feyereisen, R. (2000). Cytochromes P450: a success story. *Genome Biol.* 1, 1–9. doi: 10.1186/gb-2000-1-6-reviews3003
- Yadav, M., Pandey, J., Chakraborty, A., Hassan, M. I., Kundu, J. K., Roy, A., et al. (2022). A comprehensive analysis of calmodulin-like proteins of *glycine max* indicates their role in calcium signaling and plant defense against insect attack. *Front. Plant Sci.* 13, 817950. doi: 10.3389/fpls.2022.817950
- Yang, J., Li, H., Ma, R., Chang, Y., Qin, X., Xu, J., et al. (2022). Genome-wide transcriptome analysis and characterization of the cytochrome P450 flavonoid biosynthesis genes in pigeon pea (*Cajanus cajan*). *Planta* 255, 120. doi: 10.1007/s00425-022-03896-1
- Zhang, Y. (2008). I-TASSER server for protein 3D structure prediction. *BMC Bioinf.* 9, 1–8. doi: 10.1186/1471-2105-9-40
- Zhang, X., and Li, S. (2017). Expansion of chemical space for natural products by uncommon P450 reactions. *Nat. Prod. Rep.* 34, 1061–1089. doi: 10.1039/C7NP00028F
- Zhang, Z., Ren, J.-S., Clifton, I. J., and Schofield, C. J. (2004). Crystal structure and mechanistic implications of 1-aminocyclopropane-1-carboxylic acid oxidase—the ethylene-forming enzyme. *Chem. Biol.* 11, 1383–1394. doi: 10.1016/j.chembiol.2004.08.012
- Zhang, C., Xu, X., Xu, X., Li, Y., Zhao, P., Chen, X., et al. (2022). Genome-wide identification, evolution analysis of cytochrome P450 monooxygenase multigene family and their expression patterns during the early somatic embryogenesis in *Dimocarpus longan* Lour. *Gene* 826, 146453. doi: 10.1016/j.gene.2022.146453