

Genome-Wide Identification of the CHS Gene Family in *Platycodon grandiflorus* and Expression Analysis Under Multiple Phytohormone Treatments

Xiaoqin Wang

Sichuan College of Traditional Chinese Medicine

Jun hong Li

Sichuan College of Traditional Chinese Medicine

Yan Xia

Sichuan College of Traditional Chinese Medicine

Min Fang

Sichuan College of Traditional Chinese Medicine

Dong Yan

yandong6407@126.com

Sichuan College of Traditional Chinese Medicine

Research Article

Keywords: Flavonoids, Chalcone synthase, *Platycodon grandiflorus*, Hormone, Genome-Wide Identification

Posted Date: October 27th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-11080858/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

Abstract

Flavonoids are major bioactive compounds in *Platycodon grandiflorus*. Chalcone synthase (CHS), a rate-limiting enzyme in the flavonoid biosynthetic pathway, plays a central role in the accumulation of these metabolites. In this study, we conducted a genome-wide survey of the CHS gene family and identified 20 *PgCHS* members unevenly distributed across all chromosomes. Predicted physicochemical properties indicated that most *PgCHS* proteins are hydrophilic and localized in the cytoplasm. Phylogenetic analysis grouped CHS proteins from representative land plants into two distinct subclades, collinearity analysis further suggested that segmental duplication primarily drove the expansion of the *PgCHS* family. Additionally, promoter scans revealed abundant cis-acting elements associated with hormone responses, stress defense, and flavonoid metabolism. Quantitative real-time PCR profiling of eight representative paralogs demonstrated that most *PgCHS* genes were markedly upregulated by methyl jasmonate; by contrast, they exhibited divergent expression patterns under GA, IAA, and SA treatments. This study clarifies the evolutionary dynamics and hormonal responsiveness of the *PgCHS* family and provides a solid foundation for dissecting flavonoid biosynthesis in this medicinal plant.

1. Introduction

Flavonoids constitute one of the largest classes of phenylpropanoid-derived secondary metabolites in plants, encompassing anthocyanins, flavonols, flavones, proanthocyanidins, and isoflavonoids. These compounds serve critical functions in pigmentation, UV protection, pollen fertility, and defense against pathogens and herbivores (Chen et al. 2022). The biosynthesis of flavonoids proceeds through a well-conserved core pathway that begins with the condensation of one molecule of p-coumaroyl-CoA with three molecules of malonyl-CoA to yield naringenin chalcone. This initial committed step is catalyzed by chalcone synthase (CHS; EC 2.3.1.74), a plant-specific type III polyketide synthase that is widely regarded as the gateway enzyme of the flavonoid/isoflavonoid biosynthetic pathway (Pandith et al. 2019; Chen et al. 2026).

CHS was the first type III polyketide synthase to be discovered and has since been the subject of extensive structural, biochemical, and evolutionary investigation (Pandith et al. 2019; Waki et al. 2020). The enzyme functions as a homodimeric protein and exhibits a conserved three-dimensional fold despite its well-documented catalytic promiscuity, which can lead to the formation of alternative polyketide products, while chalcone isomerase-like proteins (CHILs) physically interact with CHS to enhance the fidelity of flavonoid precursor production (Waki et al. 2020). Beyond its canonical role in flavonoid biosynthesis, CHS has been implicated in broader physiological processes: ectopic overexpression of a CHS homolog from *Paeonia ostii* altered fatty acid composition and content in transgenic *Arabidopsis thaliana* seeds (Peng et al. 2020), and CHS gene expression can be epigenetically modulated and transgenerationally inherited, as demonstrated in *Linum usitatissimum* (Dzialo et al. 2019).

The CHS gene family has undergone significant expansion and functional diversification across the plant kingdom. In most angiosperms, CHS is encoded by a small multigene family whose members have arisen through tandem and segmental duplication events, often followed by subfunctionalization or neofunctionalization(Han et al. 2016). Genome-wide identification and characterization of CHS family members have been conducted in numerous species, including *Zea mays*, *Oryza sativa*, *Glycine max*, *Nicotiana tabacum*, *Phaseolus vulgaris*, and *Cucumis sativus*(Han et al. 2016; Han et al. 2017; Anguraj Vadivel et al. 2018; Chen et al. 2016; Wang et al. 2024; Ahmad et al. 2023). These studies have revealed considerable variation in CHS gene number, ranging from a single copy in some species to 14 members in maize and 21 loci in soybean, where the family expanded through two successive whole-genome duplication events characteristic of paleopolyploid genomes(Han et al. 2016; Anguraj Vadivel et al. 2018). Comparative analyses have further demonstrated that CHS family members exhibit tissue-specific and stress-responsive expression patterns, with individual isoforms differentially regulated by phytohormones including abscisic acid (ABA), methyl jasmonate (MeJA), salicylic acid (SA), and auxin (Chen et al. 2016; Han et al. 2017). The regulatory complexity of CHS gene expression is mediated by a network of transcription factors, notably MYB, bHLH, and WRKY family members, that integrate developmental and environmental signals to fine-tune flavonoid production(Wei et al. 2016; Pandey et al. 2016).

Medicinal plants represent a particularly valuable context for studying CHS gene families because the accumulation of flavonoid and related phenylpropanoid compounds often underlies their pharmacological activities. *Platycodon grandiflorus* (Jacq.) A.DC, a perennial herb of the Campanulaceae family and the sole species of the genus *Platycodon*, has been used for centuries in traditional East Asian medicine for its antitussive, expectorant, anti-inflammatory, and immunomodulatory properties(Ma et al. 2016; Kim et al. 2020). The pharmacological significance of *P. grandiflorus* is primarily attributed to two major classes of specialized metabolites: oleanane-type triterpenoid saponins (*platycosides*) and flavonoids(Ma et al. 2016; Yu et al. 2021). While the biosynthesis of platycosides has been the focus of substantial research effort, the flavonoid biosynthetic pathway and its gateway enzyme CHS remain comparatively understudied in this species. Recent advances in *P. grandiflorus* genomics have provided essential resources for addressing this knowledge gap(Lee et al. 2023; Yu et al. 2025). Transcriptomic studies have further identified candidate genes involved in triterpenoid saponin biosynthesis and characterized the transcriptional responses of *P. grandiflorus* to salt stress and high-temperature conditions(Yu et al. 2021; Zhang et al. 2023; Meng et al. 2026). However, a systematic genome-wide identification and characterization of the CHS gene family in *P. grandiflorus* has not been reported to date. Understanding the expression dynamics of CHS family members in response to phytohormone treatments is critical for elucidating the regulatory networks that govern flavonoid biosynthesis in *P. grandiflorus*. Exogenous application of MeJA has been shown to activate the flavonoid pathway in diverse plant systems through the JA signaling cascade, which involves the SCF^{COI1}-JAZ-MYC2 regulatory module and leads to the transcriptional upregulation of CHS and other flavonoid structural genes(Ma et al. 2024; Premathilake et al. 2020). Similarly, ABA and SA treatments modulate CHS

expression in species-specific patterns, reflecting the complex interplay between hormone signaling pathways and secondary metabolite biosynthesis(Chen et al. 2016; Han et al. 2017).

In this study, we performed a genome-wide identification of the CHS gene family in *P. grandiflorus*, characterized the physicochemical properties, gene structures, conserved motifs, and chromosomal distribution of each member, and conducted phylogenetic analyses to infer their evolutionary relationships with CHS proteins from other plant species. We further examined the expression patterns of *PgCHS* genes across different tissues and in response to treatment with MeJA, ABA, SA, and auxin using quantitative real-time PCR (qRT-PCR). Our results provide the first systematic characterization of the CHS gene family in *P. grandiflorus* and establish a foundation for future functional studies aimed at understanding the regulation of flavonoid biosynthesis in this medicinally important species.

2. Materials and Methods

2.1 Identification of CHS genes in *P. grandiflorus*

The genome sequences of *P. grandiflorus* downloaded from China National center for Bioinformation (<https://www.cncb.ac.cn/>), Accession number: GWHARYT000000000. The published protein sequence of CHS family members of *Arabidopsis thaliana*, *Panax ginseng*, *Taraxacum mongolicum* were downloaded from the National Center for Biotechnology Information database (<http://www.ncbi.nlm.nih.gov/>). Subsequently, CHS protein sequences from different species (*Arabidopsis thaliana*, *Panax ginseng*, *Taraxacum mongolicum*) were used as templates for sequence alignment in the genome of *P. grandiflorus*, candidate CHS family genes were generated from the intersection of sequence alignment by ToolKit Biologists Tools (TBtools) software (<https://github.com/CJ-Chen/TBtools>). The molecular formula, molecular weight, and isoelectric point were from Expasy (<http://web.expasy.org/cgi-bin/protaparam/protparam>).

2.2 Phylogenetic Analysis and Classification of PgCHS genes in *P. grandiflorus*.

The CHS family protein sequences of *A. thaliana*(6), *S. barbata*(5), *P. grandifloras*(15), *T. coreanum*(5), *P. ginseng*(12) were downloaded from National Center for Biotechnology Information (<https://www.ncbi.nlm.nih.gov/>) and listed in Table S4. The multiple amino acid sequence alignment of CHS proteins was performed, and the phylogenetic tree was constructed by neighbor joining (NJ) method using ClustalW method implemented in Molecular Evolutionary Genetic Analysis software (MEGA 7). The generated tree was displayed using the iTOL website (<https://itol.embl.de/upload.cgi>).

2.3 Collinearity Analysis and chromosomal mapping of the PgCHS Family genes

The annotation information and the whole genome protein sequences of *Arabidopsis thaliana*, *Panax ginseng* and *Taraxacum mongolicum* obtained from National Center for Biotechnology Information

database (<http://www.ncbi.nlm.nih.gov/>). The collinearity relationships of the CHS genes between *P. grandiflorus* and *rabidopsis thaliana*, *Taraxacum mongolicum* and *Panax ginseng* genomes were determined and visualized by using the Multiple Collinearity Scan toolkit (MCScanX) of TBtools software with default parameters. The physical positions of PgCHS genes on the chromosomes of the *P. grandiflorus* genome were determined and visualized using the TBtools software according to the genome annotation file.

2.4 Gene structure, motif analysis and cis-elements analysis

The whole genome sequences and CDS sequences of PgCHS genes were downloaded and subsequently utilized for in-depth gene structure analysis using the TBtools software. The motif analysis of the PgCHS family proteins was conducted utilizing the MEME website (<http://meme-suite.org/tools/meme>), while the domain analysis of the same protein family was performed through the Batch CD-Search tool available at (<https://www.ncbi.nlm.nih.gov/Structure/bwrpsb/bwrpsb.cgi>). The promoter sequences, encompassing 2000 base pairs upstream of the ATG start codon, were extracted from the PgCHS genes and subsequently analyzed for cis-elements using the Plant CARE platform (<http://bioinformatics.psb.ugent.be/webtools/plantcare/html/>). The types and quantities of these cis-elements were listed in Table S5.

2.5 Protein Three-dimensional structure prediction of PgCHS proteins

The three-dimensional structures of PgCHS proteins were precisely modeled utilizing the SWISS-MODEL Interactive Workspace Online software (<https://swissmodel.expasy.org/interactive>). To ensure the accuracy and rationality of the homologous modeling outcomes, the SAVESv6.0 platform (<https://saves.mbi.ucla.edu/>) was employed for a comprehensive assessment of the modeling results.

2.6 Plant treatment and Quantitative RT-PCR

One-year-old *P. grandiflorus* seedlings with uniform growth status were selected for exogenous phytohormone treatments. MeJA (100 μ M), GA₃ (100 μ M), IAA (50 μ M), SA (200 μ M) and ABA (100 μ M) solutions supplemented with 0.02% Tween-20 were sprayed on leaves until surface saturation. Seedlings sprayed with 0.02% Tween-20 aqueous solution were used as controls. Each treatment included three biological replicates. Leaf samples were collected at 6 h after spraying. Harvested samples were rapidly frozen by liquid nitrogen and preserved at -80°C for RNA isolation and qRT-PCR analysis of PgCHS genes. Total RNA was extracted from *P. grandiflorus* roots using RNAiso Plus (Takara, Japan). The extracted RNA was ultimately dissolved in 30 μ L of DEPC-treated water. Subsequently, 1.5 μ g of total RNA was utilized for cDNA synthesis, and dilutions of this cDNA were employed in qRT-PCR experiments. The expression levels of the tested genes were normalized to PgACTIN (a constitutively expressed gene). The primers used for PgCHSs were listed in Table S4. The gene expression level was calculated by the formula $(E_{\text{Actin}})^{\text{CT}_{\text{Actin}}} / (E_{\text{Target}})^{\text{CT}_{\text{Target}}}$.

3. Results

3.1. Identification of the CHS Family Genes in the *P. grandifloras* Genome

To identify chalcone synthase genes in *P. grandiflorus*, we performed BLAST searches against the *P. grandiflorus* genome database using *Arabidopsis thaliana* CHS protein sequences as queries. Candidate genes were further filtered by Hidden Markov Model (HMM) profiling of the CHS conserved domain (PF00195), and overlapping results from the two methods were retained for subsequent analysis. In total, 20 CHS family members were identified in the *P. grandiflorus* genome, designated PgCHS1 to PgCHS20. The CDS and deduced protein sequences of all PgCHS genes are listed in Table S1, and their physicochemical properties are summarized in Table 1. The 20 PgCHS genes were unevenly distributed across 9 chromosomes of *P. grandiflorus*, with most members located on chromosomes 1, 2 and 3. PgCHS proteins ranged from 327 aa (PgCHS19) to 549 aa (PgCHS8), corresponding to molecular masses of 36.4–61.9 kDa. The calculated pI varied between 5.83 (PgCHS10) and 9.45 (PgCHS7). The instability index spanned 26.34–48.49, while the aliphatic index ranged from 86.4 (PgCHS8) to 101.6 (PgCHS2). Hydropathy analysis revealed that five PgCHS proteins exhibited positive GRAVY values.

Table 1
PgCHS gene family gene and protein properties

Gene name	Gene ID	Chr	Number of amino acids	Mw (kDa)	pI	Instability index	Aliphatic index	GRAVY
<i>PgCHS1</i>	PGRA_21068	1	523	58.05	9.17	35.66	93.58	-0.042
<i>PgCHS2</i>	PGRA_18657	1	456	51.31	8.91	38.25	101.6	0.035
<i>PgCHS3</i>	PGRA_19848	1	496	55.68	9.33	45.06	100.12	0.014
<i>PgCHS4</i>	PGRA_21072	1	494	55.75	8.72	38.01	98.28	0.062
<i>PgCHS5</i>	PGRA_00265	2	389	42.48	6.77	36.03	92.96	-0.06
<i>PgCHS6</i>	PGRA_00518	2	513	57.86	9.41	38.44	94.46	-0.025
<i>PgCHS7</i>	PGRA_00048	2	507	57.11	9.45	32.98	90	-0.069
<i>PgCHS8</i>	PGRA_00200	2	549	61.96	9.3	44.44	86.41	-0.197
<i>PgCHS9</i>	PGRA_01086	2	513	58.29	9.23	36.21	95.59	-0.053
<i>PgCHS10</i>	PGRA_03544	3	393	43.26	5.83	43.3	96.28	-0.123
<i>PgCHS11</i>	PGRA_02925	3	399	43.45	7.14	35.67	94.64	-0.112
<i>PgCHS12</i>	PGRA_03218	3	419	45.75	6.43	37.37	90.55	-0.108
<i>PgCHS13</i>	PGRA_02517	3	450	50.48	9.04	39.41	95.09	-0.053
<i>PgCHS14</i>	PGRA_08776	4	470	52.28	8.57	37.1	95.83	0.093
<i>PgCHS15</i>	PGRA_10211	5	496	56.03	9.03	37.9	100.24	0.012
<i>PgCHS16</i>	PGRA_06297	6	501	55.58	8.98	30.69	92.89	-0.008
<i>PgCHS17</i>	PGRA_06411	6	390	43.14	5.77	48.49	88.79	-0.243
<i>PgCHS18</i>	PGRA_18092	7	466	52.84	8.6	44.66	87.66	-0.119
<i>PgCHS19</i>	PGRA_14510	8	327	36.40	9.34	32.1	92.48	-0.047
<i>PgCHS20</i>	PGRA_15745	9	528	58.97	8.92	26.34	92.12	-0.057
GRAVY, Grand average of hydropathicity								

3.2. Chromosomal Localization of PgCHS Family Genes in *P. grandiflorus*

The chromosomal positions of the 20 PgCHS genes were determined based on physical coordinates extracted from the *P. grandiflorus* genome annotation. All PgCHS genes were mapped to all nine chromosomes of the genome, exhibiting a markedly non-random distribution (Fig. 1). Chromosome 2

contained the largest number of family members (five), followed by chromosomes 1 and 3 (four each). Chromosome 6 harbored two PgCHS genes, whereas chromosomes 4, 5, 7, 8, and 9 each contained a single member. PgCHS genes were predominantly located in chromosomal arm regions, with no obvious enrichment in pericentromeric regions. Tandem gene arrays were observed on multiple chromosomes. On chromosome 1, three genes (PgCHS2–PgCHS4) were clustered in the long-arm region. A denser tandem array was identified on the short arm of chromosome 2, where five consecutive PgCHS genes (PgCHS5–PgCHS9) were confined to a compact genomic interval. On chromosome 3, PgCHS10–PgCHS12 formed a similar cluster in the short-arm region. The remaining PgCHS genes were dispersed across their respective chromosomes without neighboring paralogs.

Green bars represent the nine chromosomes of *P. grandiflorus*, with chromosome identifiers labeled at the top. The left scale indicates physical distance in megabases (Mb). All PgCHS genes are marked at their corresponding genomic positions on each chromosome.

3.3. Phylogenetic relationship analysis of CHS proteins

To investigate the phylogenetic relationships of CHS proteins across different plant lineages, a phylogenetic tree was constructed using MEGA7 software based on full-length CHS protein sequences from four plant species: *Arabidopsis thaliana* (5 proteins), *Zea mays* (6 proteins), *Oryza sativa* (11 proteins) and *Platycodon grandiflorus* (20 proteins). As shown in Fig. 2, all 42 CHS protein sequences were clustered into three subfamilies. Subfamily I contained 9 CHS proteins, including 2 from *P. grandiflorus* (PgCHS10, PgCHS17), 4 from *A. thaliana* and 3 from *Z. mays*. Subfamily II comprised 15 CHS members, all exclusively from *P. grandiflorus*, with no homologs from the other three species. Subfamily III included 18 CHS proteins, consisting of 3 from *P. grandiflorus* (PgCHS5, PgCHS11, PgCHS12), 1 from *A. thaliana*, 3 from *Z. mays* and 11 from *O. sativa*. All protein sequences used for phylogenetic analysis are provided in Table S2.

The neighbor-joining (NJ) tree was constructed using MEGA7 based on full-length CHS protein sequences from four plant species.

3.4. Collinearity Analysis of PgCHS Genes

Interspecific collinearity of CHS genes was examined between *P. grandiflorus* and three reference species: *Arabidopsis thaliana*, *Panax ginseng*, and *Salvia miltiorrhiza*. Collinear gene pairs are connected by red lines in Fig. 3. Fifteen collinear CHS gene pairs were identified between *P. grandiflorus* and *A. thaliana* (Fig. 3A), 34 between *P. grandiflorus* and *P. ginseng* (Fig. 3B), and 19 between *P. grandiflorus* and *S. miltiorrhiza* (Fig. 3C). Collinear PgCHS genes were distributed across most of the 9 chromosomes of *P. grandiflorus*, and the number of collinear pairs on each chromosome aligned with the overall distribution density of the PgCHS family. Chromosomes 1, 2 and 3, which carried the majority of PgCHS members, accounted for the largest proportion of cross-species collinear pairs. Tandemly clustered PgCHS genes on chromosomes 1 and 2 were located within continuous collinear blocks matching corresponding genomic regions of the three reference species.

Red lines indicate collinear CHS gene pairs between *P. grandiflorus* and reference species. Panels A, B and C correspond to collinearity profiles against *A. thaliana*, *P. ginseng* and *S. miltiorrhiza*, respectively.

3.5. Conserved Motif, domain and Gene Structure Analysis of PgCHS Genes

To systematically characterize the structural features of PgCHS family genes, conserved motifs, functional domains, and exon–intron architectures of the 20 PgCHS members were analyzed. Based on motif organization, these genes were classified into two distinct groups (Fig. 4A). Fifteen members shared a conserved set of six motifs, with only PgCHS19 lacking Motif 6. The remaining five genes contained only Motif 1 and Motif 2, and PgCHS10 also retained Motif 3. Domain annotation was consistent with this classification (Fig. 4B). Members of the first group possessed a Ketoacyl-synt_C superfamily domain, whereas those in the second group were characterized by a Chal_sti_synt_C (chalcone/stilbene synthase C-terminal) domain. Exon–intron analysis revealed that 12 PgCHS genes were intronless, exhibiting compact genomic structures averaging 2 kb in length (Fig. 4C). PgCHS8, the longest genomic locus in the family, contained 2 introns; PgCHS6 shared the same structural pattern. The other 6 PgCHS genes each harbored a single intron.

(A) Conserved motif composition of PgCHS proteins. Boxes in different colors represent distinct conserved motifs. (B) Conserved domain architectures of PgCHS proteins. Yellow and teal boxes indicate the Ketoacyl-synt_C superfamily domain and the Chal_sti_synt_C domain, respectively. (C) Exon–intron structures of PgCHS genes. Green boxes indicate untranslated regions (UTRs), yellow boxes indicate coding sequences (CDSs), and gray lines represent introns. (D) Integrated legend for color coding across panels A–C. Scales at the bottom of each panel indicate protein sequence length (A, B) and genomic sequence length (C), respectively.

3.6. Prediction of the 3-Dimensional Structures of PgCHS Proteins

To further characterize the tertiary structural features of PgCHS proteins, homology modeling was performed using SWISS-MODEL. All members folded into the canonical chalcone synthase fold, characterized by a central β -sheet core flanked by α -helices, with random coils distributed throughout the polypeptide backbone to connect secondary structure elements (Fig. 5). Distinct structural divergence was observed between the two previously defined PgCHS groups. Members of the first group (PgCHS1–PgCHS4, PgCHS6–PgCHS9, PgCHS14, PgCHS16, PgCHS18–PgCHS20) possessed an expanded β -sheet core and adopted a relatively compact conformation. In contrast, the second group (PgCHS5, PgCHS10–PgCHS13, PgCHS17) exhibited reduced β -sheet content and pronounced N- or C-terminal extensions, resulting in a more elongated architecture. PgCHS12 and PgCHS13 displayed particularly divergent terminal regions compared with other members of the same group.

Blue regions indicate α -helices, green regions indicate β -sheets, and white lines indicate random coils.

3.7. Subcellular Localization Prediction of PgCHS Family Proteins

Subcellular localization of PgCHS proteins was predicted using WoLF PSORT based on full-length amino acid sequences (Fig. 6). Five members were predicted to localize predominantly to the cytoplasm, with PgCHS11 yielding the highest prediction score. Another five proteins were predicted to be targeted to chloroplasts; among these, PgCHS2 and PgCHS14 exhibited markedly higher scores than PgCHS13, PgCHS16, and PgCHS17. Eleven PgCHS proteins showed plasma membrane localization signals; PgCHS1, PgCHS7, and PgCHS9 had higher plasma membrane localization scores than the remaining eight members. PgCHS19 was uniquely predicted to have a dominant nuclear localization. Signal peptide and transmembrane domain analyses showed that no PgCHS proteins contained canonical signal peptides. Transmembrane helices were predicted in all members except PgCHS5, PgCHS10, PgCHS11, PgCHS12, and PgCHS17.

Subcellular localization profiles were predicted using the WoLF PSORT server. The color gradient from blue to orange indicates increasing predicted localization probability. Abbreviations for subcellular compartments: Cyto, cytoplasm; Chlo, chloroplast; Nucl, nucleus; Vacu, vacuole; Mito, mitochondrion; ER, endoplasmic reticulum; Plas, plasma membrane.

3.8 Analysis of cis-acting elements of PgCHS Promoters

To investigate cis-acting regulatory elements in the promoter regions of PgCHS genes, 2000 bp genomic sequences upstream of the translation initiation site were retrieved from the *P. grandiflorus* genome and analyzed using the PlantCARE database. A total of 122 hormone-responsive cis-acting elements were identified across the 20 PgCHS promoters, comprising 50 abscisic acid-responsive elements, 44 MeJA-responsive elements, 15 gibberellin-responsive elements, 10 salicylic acid-responsive elements and 3 auxin-responsive elements (Fig. 7, Table S3). Stress-responsive elements were widely distributed in PgCHS promoter regions. Anaerobic induction elements were the most abundant type with 66 copies. Additional stress-related regulatory elements included drought-inducible MYB binding sites, low-temperature responsive elements and defense and stress responsive elements. Light-responsive elements and circadian control elements were also detected in most PgCHS promoters. Developmental regulatory elements, such as meristem expression elements, seed-specific regulatory elements and endosperm expression elements, were present in multiple family members. MYB binding sites associated with flavonoid biosynthetic regulation were identified in the promoter regions of PgCHS1 and PgCHS2. Detailed annotation information for all predicted cis-acting elements is listed in Table S3.

Colored boxes represent different types of cis-acting elements. The bottom scale indicates promoter sequence length (0–2000 bp) and the relative position of each element.

3.9 Expression pattern analysis of PgCHS genes

To investigate the hormone response characteristics of *P. grandiflorus*, five exogenous phytohormones, including MeJA, GA, IAA, SA and ABA, were applied to plant materials, followed by qRT-PCR analysis of all PgCHS family genes. Eight PgCHS genes with relatively high expression levels were selected, and these members exhibited distinct expression patterns under different hormone treatments (Fig. 8). PgCHS2 was strongly upregulated by MeJA, and induced by GA, IAA and SA, whereas ABA exerted no significant effect on its expression. PgCHS3 showed significant upregulation under GA and SA treatments, while no obvious expression changes were detected after MeJA, IAA and ABA treatments. PgCHS4 was induced by IAA and ABA, with non-significant expression variation observed under MeJA, GA and SA treatments. PgCHS8 was activated by MeJA, GA, IAA and SA, and ABA treatment did not cause significant transcriptional alteration. MeJA, GA and IAA significantly induced the expression of PgCHS9, and SA and ABA treatments failed to produce significant regulatory effects. No significant differences in transcript abundance of PgCHS12 were observed among all hormone treatments compared with the control. PgCHS15 was significantly upregulated only under MeJA treatment, and other tested hormones did not trigger significant expression changes. PgCHS17 was induced by GA, and its expression remained stable upon MeJA, IAA, SA and ABA treatments.

The relative expression levels of PgCHS2 (A), PgCHS3 (B), PgCHS4 (C), PgCHS8 (D), PgCHS9 (E), PgCHS12 (F), PgCHS15 (G) and PgCHS17 (H) under MeJA, GA, IAA, SA and ABA treatments. Data represent the mean $\pm$ standard deviation of three biological replicates. Asterisks indicate the P value determined by a Student's t test as compared to control (*P < 0.05, **P < 0.01).

4. Discussion

Genome-wide characterization of gene families provides an essential foundation for dissecting gene function and evolutionary trajectories. In this study, 20 CHS family genes were systematically identified in the *P. grandiflorus* genome (Table 1), representing the first comprehensive analysis of this key flavonoid biosynthetic gene family in this medicinal species. This gene count falls within the range reported for most angiosperms, and is lower than that in species with additional whole-genome duplication events such as soybean, which is consistent with the single γ whole-genome triplication event documented in the *P. grandiflorus* genome (Yu et al. 2025; Anguraj Vadivel et al. 2018).

Interspecific collinearity analysis detected more collinear CHS gene pairs between *P. grandiflorus* and two medicinal species (*P. ginseng* and *S. miltiorrhiza*) than between *P. grandiflorus* and *A. thaliana*. This pattern may reflect closer evolutionary relationships among these medicinal plant lineages (Fig. 3), and suggests that CHS family expansion in *P. grandiflorus* is likely shaped by both genome-wide and lineage-specific duplication events, as observed in other angiosperms (Han et al. 2016). Conserved motif, functional domain and gene structure analyses consistently classified the 20 PgCHS genes into two distinct groups, Group I members harbored complete CHS characteristic domains and more compact protein folds, whereas Group II members exhibited reduced β -sheet content and extended terminal regions (Fig. 5). Such structural variation may underlie functional differentiation among PgCHS isoforms, and could be associated with the catalytic promiscuity widely documented in type III polyketide

synthases(Waki et al. 2020). The presence of flavonoid-related MYB binding sites in PgCHS1 and PgCHS2 promoters implies that these two members may act as core components of the flavonoid regulatory network in *P. grandiflorus*, consistent with the conserved role of MYB transcription factors in modulating CHS expression.

Expression profiling under five phytohormone treatments revealed distinct response patterns among PgCHS family members. MeJA induced the expression of most tested PgCHS genes, which aligns with the well-characterized function of JA signaling in activating flavonoid biosynthetic pathways(Premathilake et al. 2020). In contrast, individual PgCHS genes showed differential responses to GA, IAA, SA and ABA, indicating that the family may participate in multiple hormone-mediated physiological processes. These expression patterns are consistent with the diversity of cis-acting regulatory elements identified in *PgCHS* promoter regions, and accord with findings in other species where CHS isoforms are differentially regulated by hormonal signals.

Collectively, this study provides a systematic overview of the PgCHS gene family and lays a foundation for elucidating flavonoid biosynthesis regulation in *P. grandiflorus*. Further biochemical and genetic validation is required to clarify the precise catalytic activities and in planta functions of individual family members.

Declarations

Conflicts of Interest:

The authors declare no conflicts of interest.

Author Contribution

XW and DY conceived the project and research plans; XW, JL performed the experiments and analyzed the data; XW and DY wrote the article; YX and MF critically revised the manuscript. All authors read and approved the final manuscript.

Data Availability Statement:

All relevant data can be found within the manuscript and its supporting material.

References

1. Ahmad S, Ali S, Shah AZ, Khan A, Faria S (2023) Chalcone synthase (CHS) family genes regulate the growth and response of cucumber (*Cucumis sativus* L.) to *Botrytis cinerea* and abiotic stresses. *Plant Stress*. 10.1016/j.stress.2023.100159

2. Anguraj Vadivel AK, Krysiak K, Tian G, Dhaubhadel S (2018) Genome-wide identification and localization of chalcone synthase family in soybean (*Glycine max* [L]Merr). *BMC Plant Biol* 18(1):325. 10.1186/s12870-018-1569-x
3. Chen L, Cao H, Huang Q, Xiao J, Teng H (2022) Absorption, metabolism and bioavailability of flavonoids: a review. *Crit Rev Food Sci Nutr* 62(28):7730–7742. 10.1080/10408398.2021.1917508
4. Chen S, Pan X, Li Y, Cui L, Zhang Y, Zhang Z, Pan G, Yang J, Cao P, Yang A (2016) Identification and Characterization of Chalcone Synthase Gene Family Members in *Nicotiana tabacum*. *J Plant Growth Regul*. 10.1007/s00344-016-9646-6
5. Chen Y, Liang C, Liu Y, Jiang M, Qiu L, Yu H, Zhang L (2026) Chalcone Synthase: From Discovery to Biotechnological Applications. *J Exp Bot*. 10.1093/jxb/erag350
6. Dzialo M, Szopa J, Hnitecka A, Zuk M (2019) Transgenerational Perpetuation of CHS Gene Expression and DNA Methylation Status Induced by Short Oligodeoxynucleotides in Flax (*Linum usitatissimum*). *Int J Mol Sci*. 10.3390/ijms20163983
7. Han Y, Cao Y, Jiang H, Ding T (2017) Genome-wide dissection of the chalcone synthase gene family in *Oryza sativa*. *Mol Breeding*. 10.1007/s11032-017-0721-x
8. Han Y, Ding T, Su B, Jiang H (2016) Genome-Wide Identification, Characterization and Expression Analysis of the Chalcone Synthase Family in Maize. *Int J Mol Sci*. 10.3390/ijms17020161
9. Kim J, Kang S-H, Park S-G, Yang T-J, Lee Y, Kim OT, Chung O, Lee J, Choi J-P, Kwon S-J, Lee K, Ahn B-O, Lee DJ, Yoo S-i, Shin I-G, Um Y, Lee DY, Kim G-S, Hong CP, Bhak J, Kim C-K (2020) Whole-genome, transcriptome, and methylome analyses provide insights into the evolution of platycoside biosynthesis in *Platycodon grandiflorus*, a medicinal plant. *Hortic Res*. 10.1038/s41438-020-0329-x
10. Lee D-J, Choi J-W, Kang J-N, Lee S-M, Park G-H, Kim C-K (2023) Chromosome-Scale Genome Assembly and Triterpenoid Saponin Biosynthesis in Korean Bellflower (*Platycodon grandiflorum*). *Int J Mol Sci*. 10.3390/ijms24076534
11. Ma C-H, Gao Z-J, Zhang J-J, Zhang W, Shao J-H, Hai M-R, Chen J-W, Yang S-C, Zhang G-H (2016) Candidate Genes Involved in the Biosynthesis of Triterpenoid Saponins in *Platycodon grandiflorum* Identified by Transcriptome Analysis. *Front Plant Sci*. 10.3389/fpls.2016.00673
12. Ma Q, Tan L, Zhou D, Wang X, Sun H, Wang Y, Tian L, Shi C, Wei A, Fei X (2024) Exogenous methyl jasmonate induces CHS and promotes flavonoid accumulation in *Zanthoxylum bungeanum*. *Ind Crops Prod*. 10.1016/j.indcrop.2024.119682
13. Meng L, Guan J, Sun X, Wang C, Luan P, Pan F, Han R, Wei T, Xing S (2026) Transcriptome and miRNAome analysis reveals expression profiles of platycodin biosynthesis-related genes and their potential miRNA regulators in *Platycodon grandiflorus* under high-temperature stress. *Front Plant Sci*. 10.3389/fpls.2026.1820112
14. Pandey A, Alok A, Lakhwani D, Singh J, Asif MH, Trivedi PK (2016) Genome-wide Expression Analysis and Metabolite Profiling Elucidate Transcriptional Regulation of Flavonoid Biosynthesis and Modulation under Abiotic Stresses in Banana. *Sci Rep*. 10.1038/srep31361

15. Pandith SA, Ramazan S, Khan MI, Reshi ZA, Shah MA (2019) Chalcone synthases (CHSs): the symbolic type III polyketide synthases. *Planta*. 10.1007/s00425-019-03307-y
16. Peng L-P, Hao Q, Men S-Q, Wang X-R, Huang W-Y, Tong N-N, Chen M, Liu Z-A, Ma X-F, Shu Q-Y (2020) Ecotopic over-expression of PoCHS from *Paeonia ostii* altered the fatty acids composition and content in *Arabidopsis thaliana*. *Physiol Plant*. 10.1111/ppl.13293
17. Premathilake AT, Ni J, Shen J, Bai S, Teng Y (2020) Transcriptome analysis provides new insights into the transcriptional regulation of methyl jasmonate-induced flavonoid biosynthesis in pear calli. *BMC Plant Biol*. 10.1186/s12870-020-02606-x
18. Waki T, Mameda R, Nakano T, Yamada S, Terashita M, Ito K, Tenma N, Li Y, Fujino N, Uno K, Yamashita S, Aoki Y, Denessiouk K, Kawai Y, Sugawara S, Saito K, Yonekura-Sakakibara K, Morita Y, Hoshino A, Takahashi S, Nakayama T (2020) A conserved strategy of chalcone isomerase-like protein to rectify promiscuous chalcone synthase specificity. *Nat Commun*. 10.1038/s41467-020-14558-9
19. Wang J, Lan Z, Wang H, Xu C, Zhou Z, Cao J, Liu Y, Sun Z, Mu D, Han J, Tang Y, Geng S, Zhao Y, Sun X, Du J (2024) Genome-wide identification of chalcone synthase (CHS) family members and their expression patterns at the sprouting stage of common bean (*Phaseolus vulgaris*) under abiotic stress. *Sci Hort*. 10.1016/j.scienta.2024.113309
20. Wei W, Hu Y, Han Y-T, Zhang K, Zhao F-L, Feng J-Y (2016) The WRKY transcription factors in the diploid woodland strawberry *Fragaria vesca*: Identification and expression analysis under biotic and abiotic stresses. *Plant Physiol Biochem*. 10.1016/j.plaphy.2016.04.014
21. Yu H, Liu M, Yin M, Shan T, Peng H, Wang J, Chang X, Peng D, Zha L, Gui S (2021) Transcriptome analysis identifies putative genes involved in triterpenoid biosynthesis in *Platycodon grandiflorus*. *Planta*. 10.1007/s00425-021-03677-2
22. Yu H, Wang H, Liang X, Liu J, Jiang C, Chi X, Zhi N, Su P, Zha L, Gui S (2025) Telomere-to-telomere gap-free genome assembly provides genetic insight into the triterpenoid saponins biosynthesis in *Platycodon grandiflorus*. *Hortic Res*. 10.1093/hr/uhaf030
23. Zhang M, Xing Y, Ma J, Zhang Y, Yu J, Wang X, Jia X (2023) Investigation of the response of *Platycodon grandiflorus* (Jacq.) A. DC to salt stress using combined transcriptomics and metabolomics. *BMC Plant Biol*. 10.1186/s12870-023-04536-w

Figures

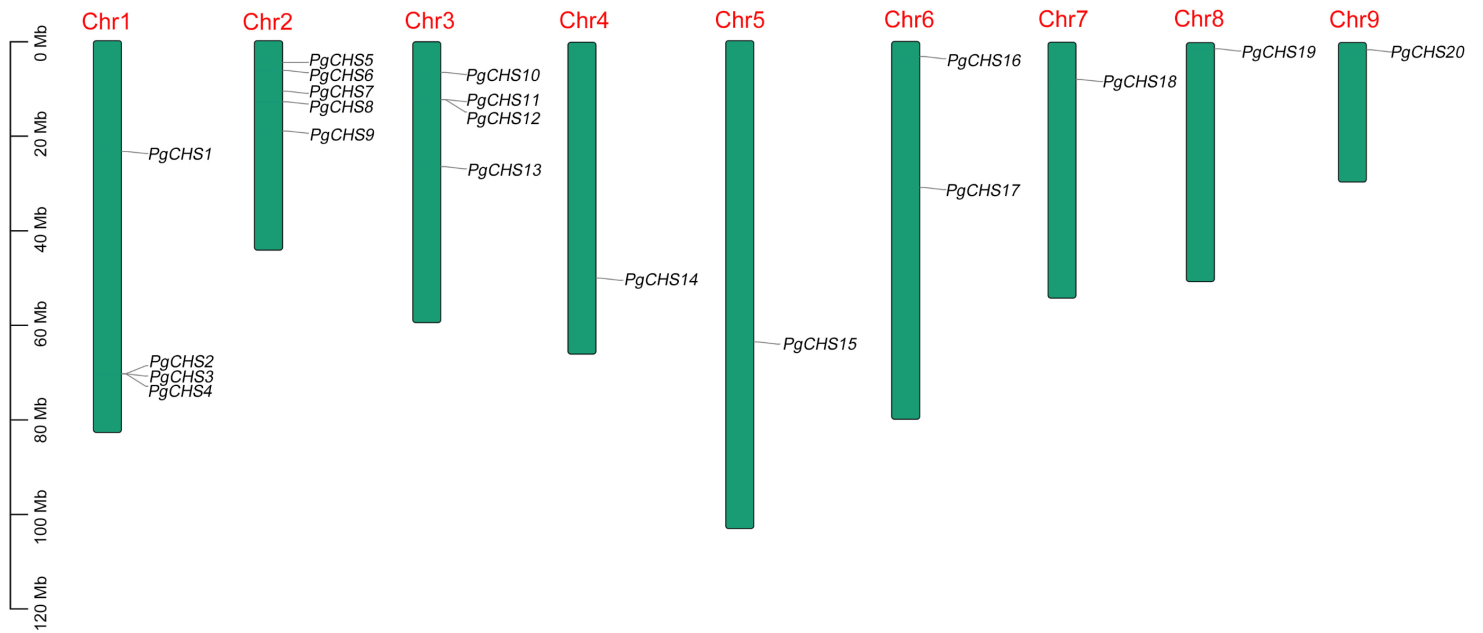


Figure 1

Chromosomal distribution of *PgCHS* genes.

Green bars represent the nine chromosomes of *P. grandiflorus*, with chromosome identifiers labeled at the top. The left scale indicates physical distance in megabases (Mb). All *PgCHS* genes are marked at their corresponding genomic positions on each chromosome.

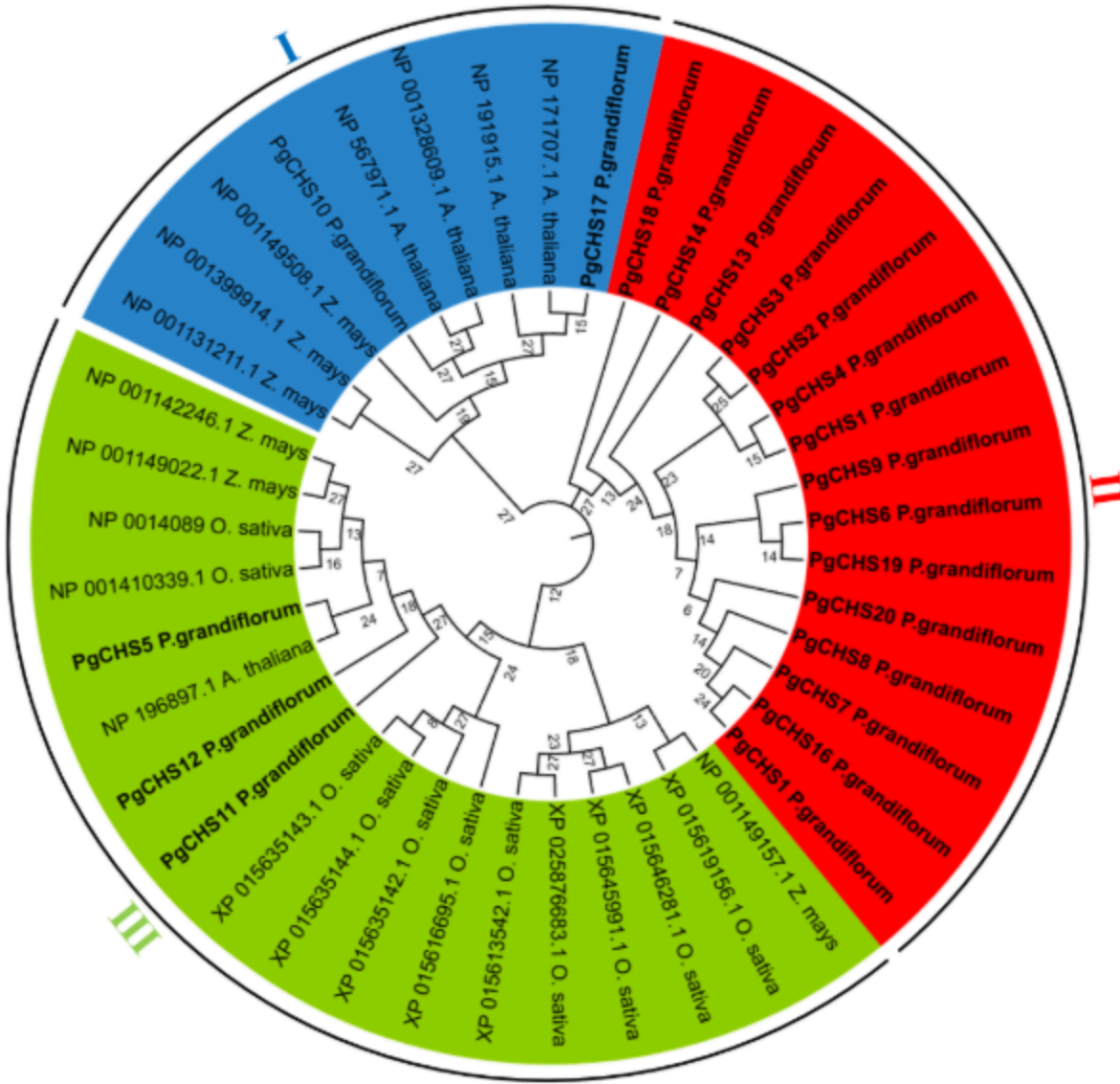


Figure 2

Phylogenetic tree of CHS proteins.

The neighbor-joining (NJ) tree was constructed using MEGA7 based on full-length CHS protein sequences from four plant species.

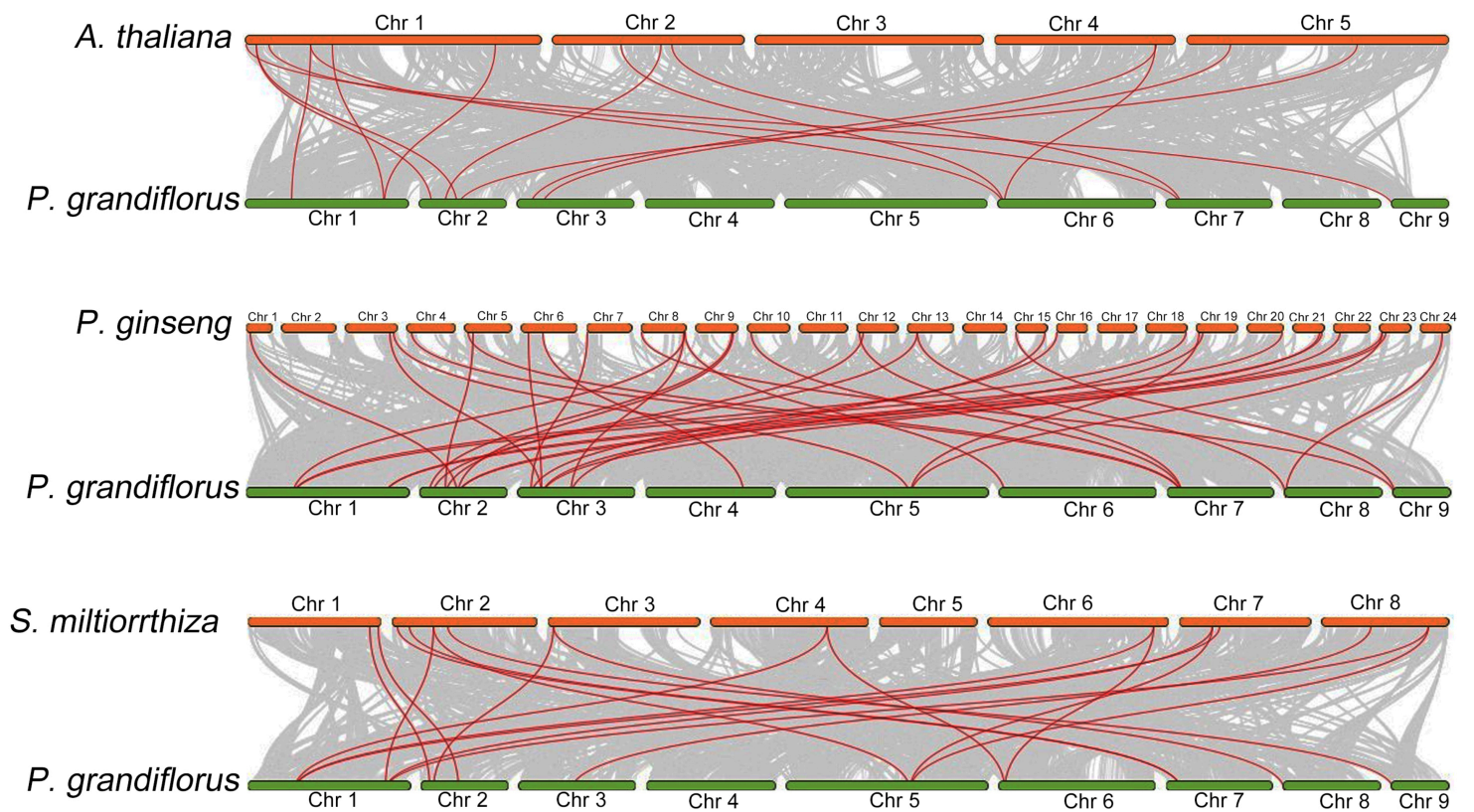


Figure 3

Interspecific collinearity analysis of PgCHS genes.

Red lines indicate collinear CHS gene pairs between *P. grandiflorus* and reference species. Panels A, B and C correspond to collinearity profiles against *A. thaliana*, *P. ginseng* and *S. miltiorrhiza*, respectively.

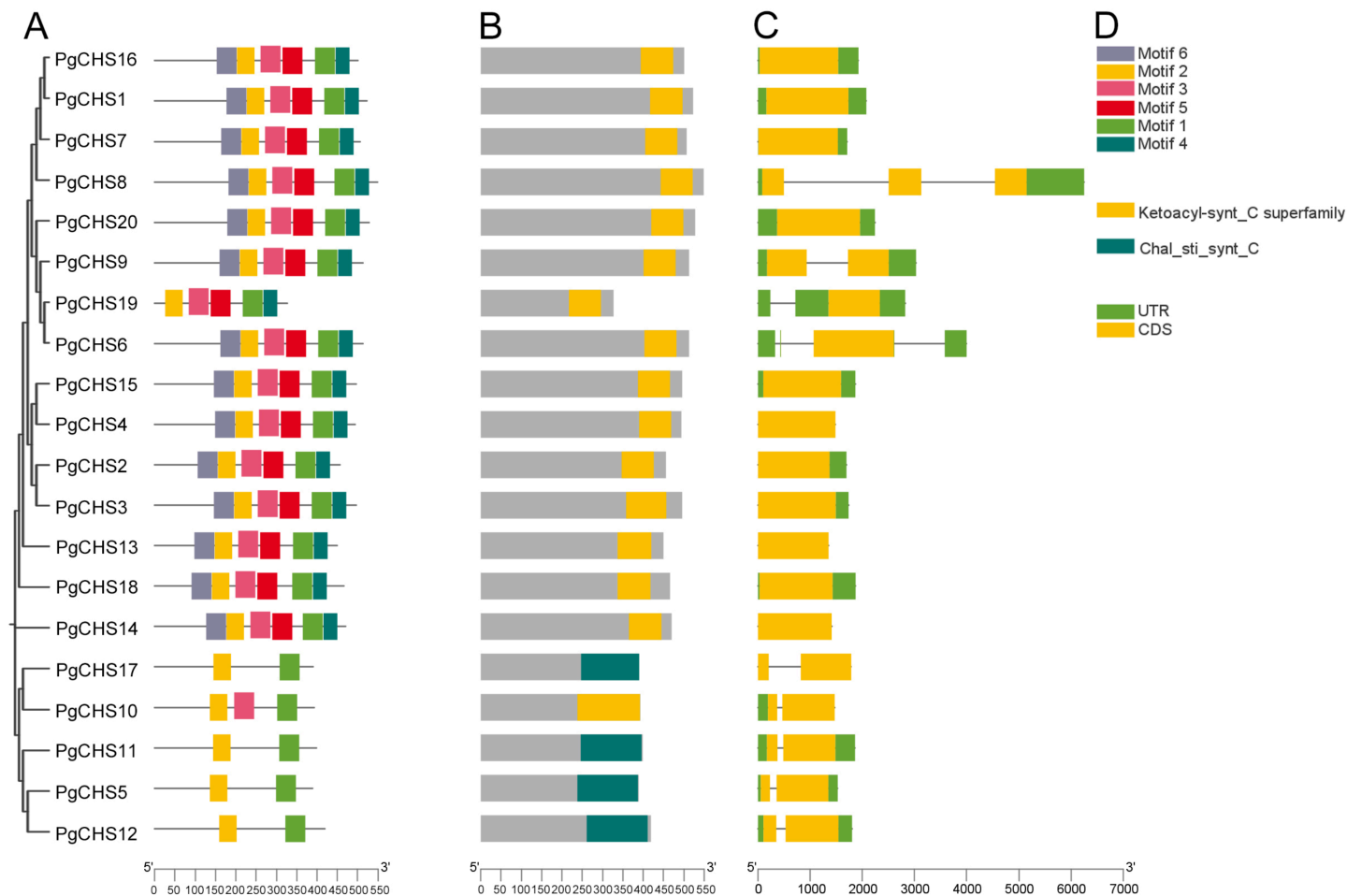


Figure 4

Conserved motifs, functional domains and gene structures of PgCHS family members.

(A) Conserved motif composition of PgCHS proteins. Boxes in different colors represent distinct conserved motifs. (B) Conserved domain architectures of PgCHS proteins. Yellow and teal boxes indicate the Ketoacyl-synt_C superfamily domain and the Chal_sti_synt_C domain, respectively. (C) Exon-intron structures of PgCHS genes. Green boxes indicate untranslated regions (UTRs), yellow boxes indicate coding sequences (CDSs), and gray lines represent introns. (D) Integrated legend for color coding across panels A–C. Scales at the bottom of each panel indicate protein sequence length (A, B) and genomic sequence length (C), respectively.

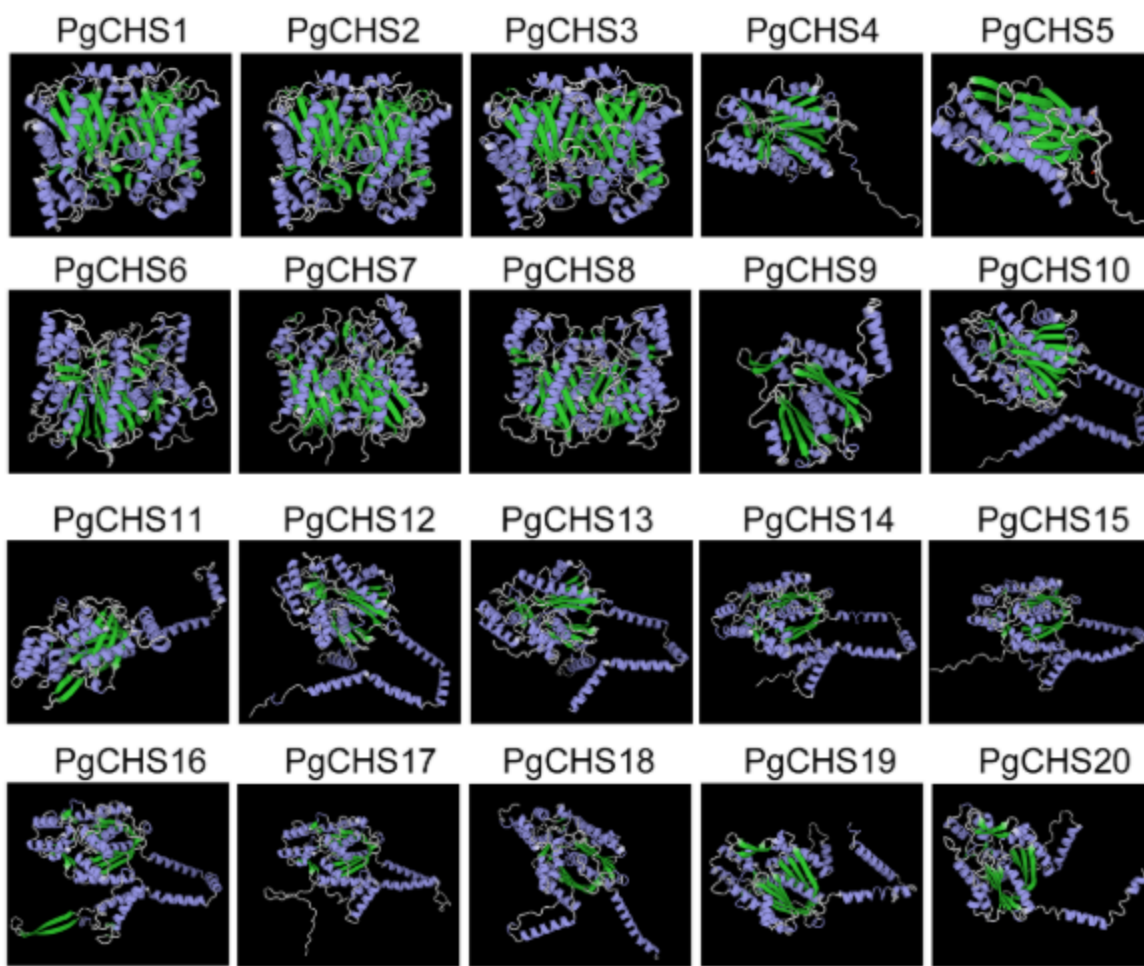


Figure 5

Three-dimensional structural models of PgCHS family proteins.

Blue regions indicate α -helices, green regions indicate β -sheets, and white lines indicate random coils.

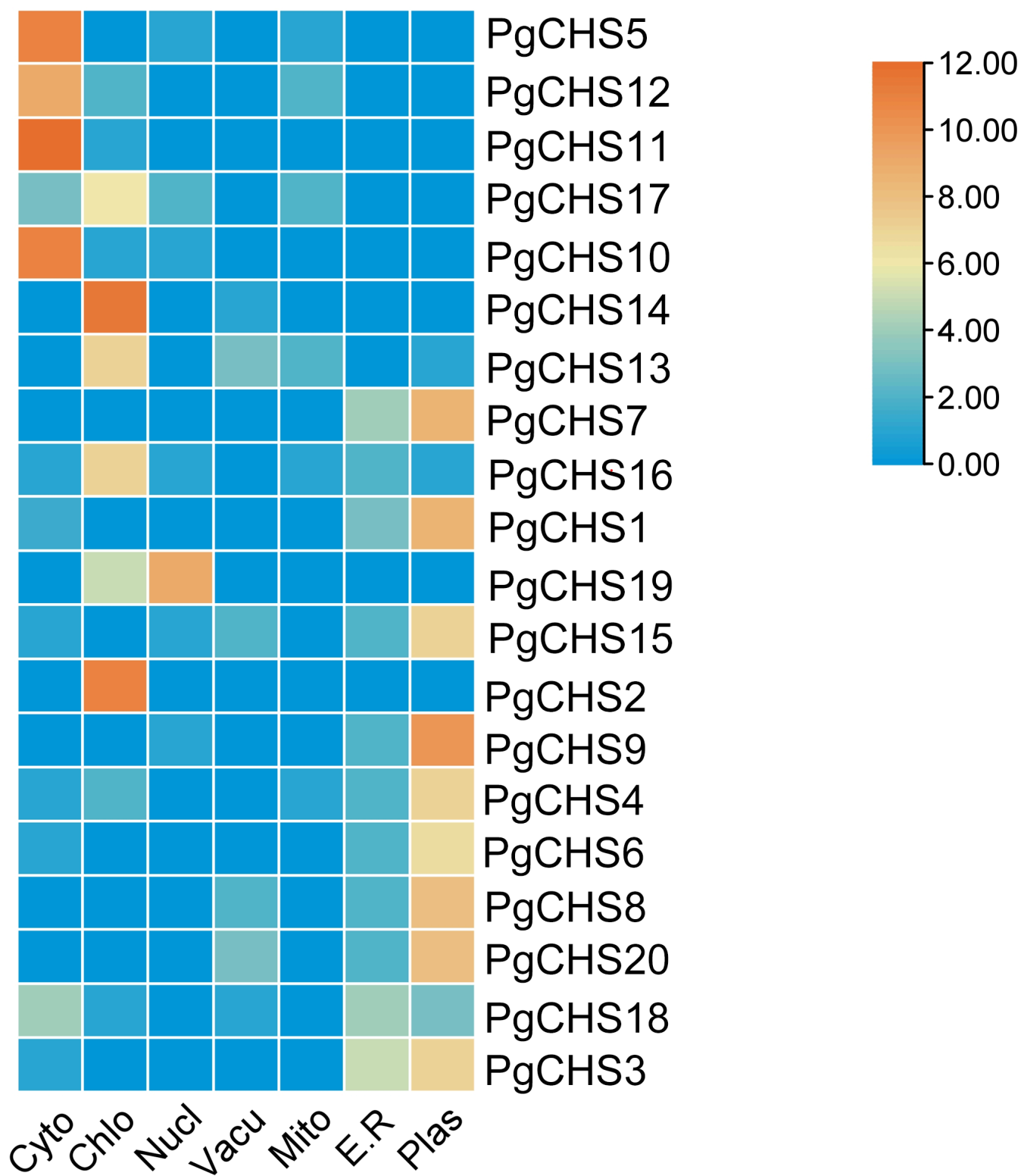


Figure 6

Subcellular localization prediction of PgCHS family proteins.

Subcellular localization profiles were predicted using the WoLF PSORT server. The color gradient from blue to orange indicates increasing predicted localization probability. Abbreviations for subcellular

compartments: Cyto, cytoplasm; Chlo, chloroplast; Nucl, nucleus; Vacu, vacuole; Mito, mitochondrion; ER, endoplasmic reticulum; Plas, plasma membrane.

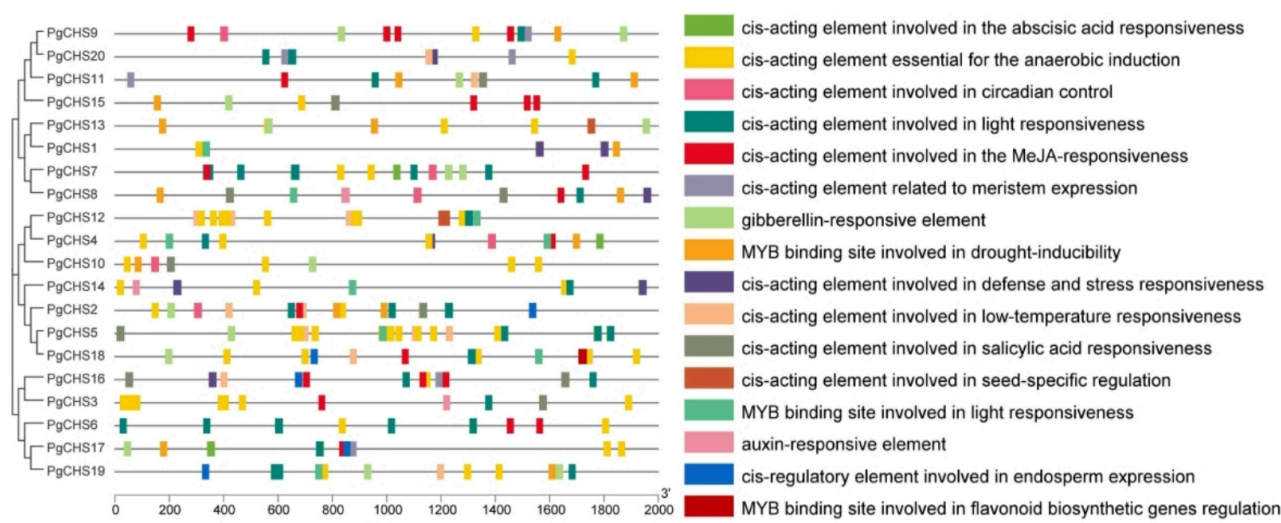


Figure 7
Cis-acting element analysis of PgCHS gene promoters.

Colored boxes represent different types of cis-acting elements. The bottom scale indicates promoter sequence length (0–2000 bp) and the relative position of each element.

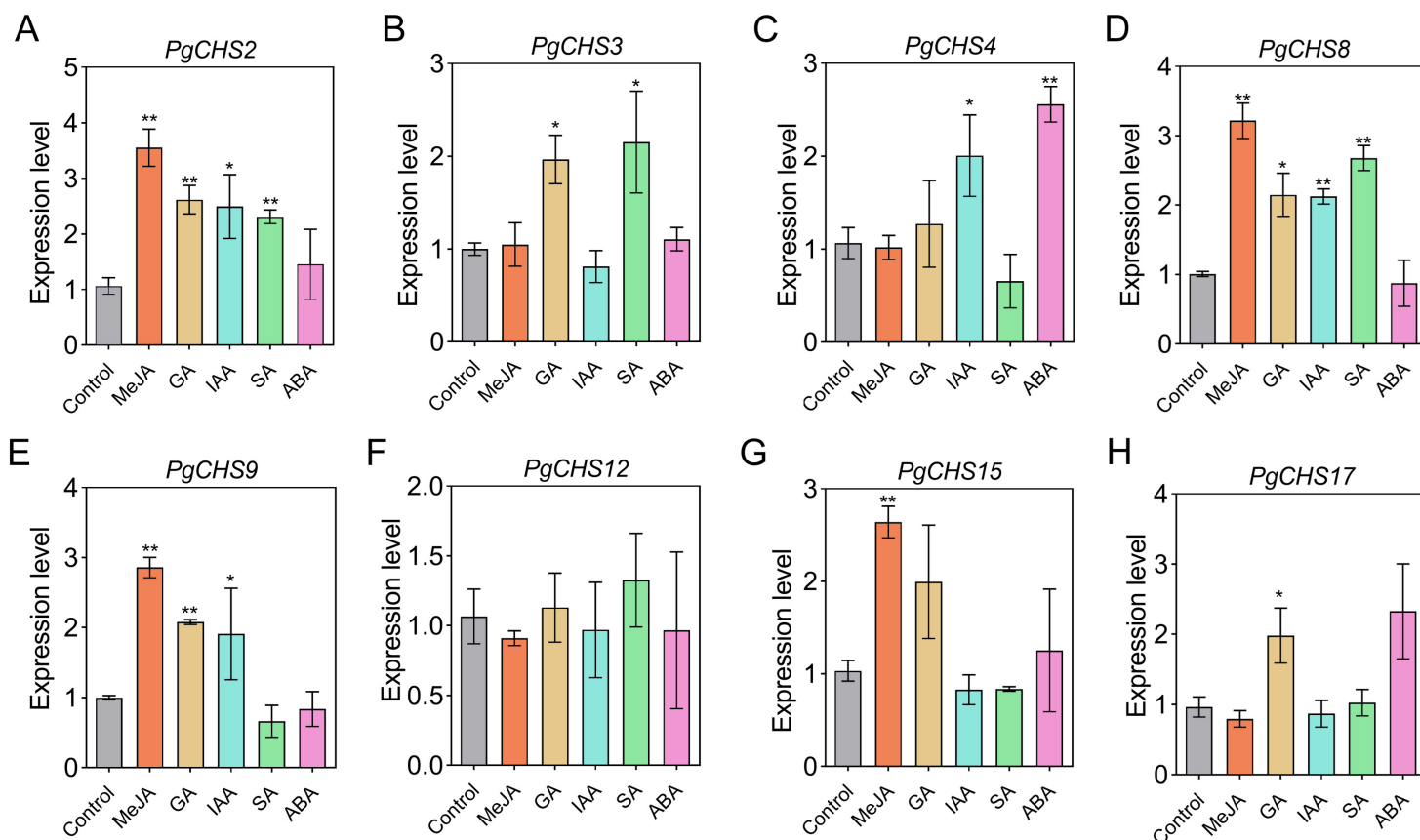


Figure 8

Expression patterns of eight PgCHS genes under different hormone treatments detected by qRT-PCR.

The relative expression levels of PgCHS2 (A), PgCHS3 (B), PgCHS4 (C), PgCHS8 (D), PgCHS9 (E), PgCHS12 (F), PgCHS15 (G) and PgCHS17 (H) under MeJA, GA, IAA, SA and ABA treatments. Data represent the mean $\pm$ standard deviation of three biological replicates. Asterisks indicate the P value determined by a Student's t test as compared to control (*P<0.05, **P<0.01).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [supplementaltables.xlsx](#)