Analysis of *CYP701A1* genes in *gossypium* species and functional characterization through gene silencing

Zhao Liang [a], Di Jiachun [b], Guo Qi [a], Xu Zhenzhen [a], Zhao Jun [a], Xu Peng [a], Xu Jianwen [a], Liu Jianguang [a], Shen Xinlian [a], Chen Xusheng [a],*

[a] *Key Laboratory of Cotton and Rapeseed (Nanjing), Ministry of Agriculture and Rural Affairs, Institute of Industrial Crops, Jiangsu Academy of Agricultural Sciences, Nanjing, China*
[b] *Institute of Germplasm Resources and Biotechnology, Jiangsu Academy of Agricultural Sciences, Nanjing, China*

ARTICLE INFO

ABSTRACT

Gibberellins (GA) are known to play crucial roles in various aspects of plant growth and development. The cytochrome P450 enzyme family is recognized for its significance in plant metabolic processes. Specifically, CYP701s, a subgroup of CYP71, encode *ent*-kaurene oxidase in the gibberellin synthesis pathway. In this study, we analyzed genomic data from 30 *Gossypium* species, including nine allotetraploid genomes (AD1-AD7, with two each for AD1 and AD2), 21 diploid genomes (A-G, K, with two A-genomes and 12 D-genomes), and *Gossypioides kirkii* genome as an outgroup for evolutionary analysis, totaling 31 genomes. Subsequently, 40 *CYP701A1* genes were identified from various genomes and conducted a comprehensive analysis of their structure and evolution. Virus-induced gene silencing (VIGS) technology was utilized to knock out the *GhCYP701A1* gene in *Gossypium hirsutum* ac TM-1. Subsequent analysis revealed changes in hormone content, with decreased gibberellin levels and notable increases in auxin, cytokinin, and jasmonic acid contents. Conversely, salicylic acid content decreased, while the precursor for ethylene synthesis, 1-aminocyclopropane-1-carboxylic acid (ACC), remained relatively stable. Transcriptome analysis of the gene silencing plants identified 15,962 differentially expressed genes, including 8376 upregulated and 7586 downregulated genes. Enrichment analysis through KEGG pathway highlighted 'Plant hormone signal transduction' as a prominent pathway with 234 differentially expressed genes. The study provided insights into the function and regulatory network of the gene.

1. Introduction

Gibberellins (GAs) play a crucial role in various aspects of plant growth and development, such as seed germination, stem elongation, leaf expansion, and flower and seed developments [1–3]. Despite the existence of numerous GAs identified from plants and fungi [4], only a select few, including GA1, GA3, GA4, and GA7 [5], exhibit biological activity. The majority of other gibberellins act as intermediates or catabolites in the biosynthesis of the active GAs. The biosynthesis of GAs in higher plants involves three main stages: the production of *ent*-kaurene from geranyl geranyl diphosphate (GGDP) in proplastids, the conversion of *ent*-kaurene to GA12 through cytochrome P450 monooxygenases, and the formation of C20-GAs and C19-GAs in the cytoplasm. Various enzymes, such as *ent*-copalyl diphosphate synthase (CPS), *ent*-kaurene synthase (KS), *ent*-kaurene oxidase (KO), *ent*-kaurenoic acid oxidase (KAO), and GA 20-oxidases (GA20ox) and 3-oxidases (GA3ox), catalyze these reactions [6]. The cytochrome P450 enzyme family, particularly P450, play significant roles in plant metabolic processes and serves as a model for metabolic structure and evolution [7,8]. CYP701s, a subgroup of the cytochrome P450 superfamily, are known for their involvement in various metabolic pathways across different organisms. CYP701s have been present since early in the evolution of land plants, with examples found in moss (*Physcomitrella*). Unlike most oxygenases in terpenoid hormone biosynthesis, CYP701s are part of the CYP71 clan [9,10]. CYP701A3, for instance, encodes the *ent*-kaurene oxidase (KO) responsible for three oxygenation steps in the conversion of *ent*-kaurene to *ent*-kaurenoic acid, influencing processes like flowering regulation, stem elongation, and cell expansion [11–13].

Cotton, as one of the world's most important fiber crops, serves as the raw material for textiles and plays a significant role in global economies. In recent years, the completion of the cotton genome sequencing has marked a significant milestone in agricultural genomics. The cotton genome sequence provides a detailed blueprint of the genetic composition of various cotton species, including the cultivated tetraploid cotton species, *Gossypium hirsutum* and *Gossypium barbadense*, along with their diploid progenitors. Through genome sequencing, researchers have uncovered the intricate genetic landscape of cotton, characterized by polyploidy, extensive gene duplication, and genome rearrangements [14, 15].

In this study, we utilized genomic data from a total of 31 different sources. This included nine tetraploid germplasm resources of cotton (AD1-AD7) [16–19], with AD1 representing *G. hirsutum* acc. TM-1 and the early form of domesticated *G. hirsutum* known as race *punctatum* (Ghp), and AD2 representing *G. barbadense* acc. H7124 and a perennial sea-island cotton accession, Lianhemumian. Additionally, we analyzed 22 diploid genomes, which consisted of eight genomes from the A-K group (A genome containing A1 and A2, D genome containing D1-D10, G genome containing G1 and G2) [14,15,20–27], as well as genome information from the exoid group *Gossypioides kirkii* [28]. By comparing the sequence information of *GhCYP701A1* in TM-1 with homologous sequences from the other genomes, we conducted an evolutionary analysis. Furthermore, we utilized virus-induced gene silencing (VIGS) knockout and transcriptome sequencing to investigate the function and regulatory network of the *GhCYP701A1* gene.

2. Materials and methods

2.1. Plant material

Gossypium hirsutum cv. TM-1 plants were cultivated in a controlled cotton climatic chamber at a temperature of $30/25 \pm 3$ °C with a light/dark cycle of 16/8 h. After seven days from sowing, the cotyledons of seedlings were targeted for virus-mediated silencing. Concurrently, the environmental conditions were adjusted to 8 h of darkness and 16 h of light at 23 °C. Following this, the leaves of the plants were harvested and immediately frozen in liquid nitrogen. All collected materials were then stored at −80 °C until further processing. The three biological replicates were split into two portions, with one used for RNA isolation and Illumina sequencing, while the other was utilized for hormone measurements.

2.2. Identification and structural analysis of CYP701A1

The exon/intron structures of the *CYP701A1* genes were visualized in various genomes using the Gene Structure Display Server [29]. The structural domains and motifs of the predicted proteins were validated through Batch CD-Search (https://www.ncbi.nlm.nih.gov/Structure/bwrpsb/bwrpsb.cgi). Subsequently, a neighbor-likelihood (NL) phylogenetic tree of all identified *CYP701A1* genes was generated using MEGA-X with 1000 bootstrap replications [30]. The promoter sequence (2000bp DNA sequence upstream of the start codon) of each *CYP701A1* gene was extracted from the genomes and analyzed for *cis*-acting regulatory elements using the PlantCARE database [31]. The predicted *cis*-acting regulatory elements were then classified based on their regulatory functions.

2.3. Validation of gene expression profile using qRT-PCR

RNA was extracted from samples for analysis using real-time quantitative (qRT-PCR). cDNA synthesis was performed using M-MLV Reverse Transcriptase (Promega, U.S). qRT-PCR specific primers were designed with Beacon Designer 7.0 software from Premier Biosoft International, Palo Alto, CA. The internal control used was *Histone3* (AF024716). qRT-PCR was conducted using iQ SYBR Green Supremix (Bio-Rad, USA) following the manufacturer's instructions. The PCR thermal cycle

conditions were as follows: 94 °C for 3 min, 40 cycles including 94 °C for 15 s, 60 °C for 15 s, and 72 °C for 30 s. The relative expression levels were determined using the $2^{-\Delta\Delta Ct}$ method [32].

2.4. The experimental method for virus-induced gene silencing (VIGS)

Virus-Induced Gene Silencing (VIGS) was employed in this research to investigate gene function. The VIGS vector was constructed based on the Tobacco Rattle Virus (TRV) system. Initially, a segment of the target gene was amplified through PCR and inserted into the TRV vector using specific restriction enzymes. The resulting vector was then introduced into *Agrobacterium tumefaciens* strain GV3101, which was subsequently used to infiltrate cotton cotyledon. The infiltrated plants were grown under controlled conditions with a 16-h light/8-h dark photoperiod at 23 °C. Gene silencing was confirmed through quantitative reverse transcription PCR (qRT-PCR) analysis to assess the reduction in target gene expression compared to control plants infiltrated with an empty vector [33].

2.5. Plant hormone liquid chromatography-mass spectrometry (LC-MS) detection

This experiment employed the AB Company's Qtrap6500 mass spectrometer. The ESI-HPLC-MS/MS method was utilized for quantitative analysis of plant hormones. The Agilent 1290 high-performance liquid chromatography (HPLC) system was used to separate sample substances, with electrospray ionization (ESI) as the ion source, and multiple reaction monitoring (MRM) mode scanning was performed. This involved simultaneous scanning of multiple selected ion monitoring (SIM) modes. The peak time of the target substance in the sample and standard, as well as the ratio of the target substance's various daughter ions in the sample and standard, were used to determine the analyte in the sample. Quantitation of the analyte was achieved using the standard curve method. Approximately 1 g of fresh plant samples were accurately weighed and ground into powder in liquid nitrogen. The powder was mixed with 10 ml of isopropanol/hydrochloric acid extraction buffer and shaken at 4 °C for 30 min. Then, 20 ml of dichloromethane was added, followed by shaking at 4 °C for another 30 min. After centrifugation at 4 °C and 13,000 rpm for 5 min, the organic phase was collected. The organic phase was dried under nitrogen and dissolved in 400 μl of methanol (0.1 % formic acid). The solution was filtered through a 0.22 μm membrane and subjected to HPLC-MS/MS analysis. Standard solutions of indole-3-Acetic acid (IAA), 1-aminocyclopropane-1-carboxylic acid (ACC), Gibberellin (GA3), jasmonic acid (JA), salicylic acid (SA), abscisic acid (ABA) were prepared in methanol (0.1 % formic acid) with concentrations 0.1 ng/ml,0.2 ng/ml,0.5 ng/ml,1 ng/ml,2 ng/ml,5 ng/ml, 10 ng/ml, 20 ng/ml, 50 ng/ml 200 ng/ml. Each concentration was prepared in triplicate, and points with poor linearity were removed when constructing the standard curve equation. All experiments were conducted at least three times per sample.

2.6. Transcriptome sequencing and DEG analysis

Transcriptome analysis of differential gene expression was performed using a combination of experimental and computational techniques. RNA was extracted from control and experimental samples representing different conditions or treatments. The RNA quality and purity were assessed by using a Qubit® 2.0 Fluorometer (Invitrogen, Carlsbad, CA, U.S.) and an Agilent 2100 Bioanalyzer (Agilent Technologies, U.S.). Subsequently, high-throughput RNA sequencing (RNA-seq) was conducted using Illumina sequencing platforms to generate transcriptomic data.

Raw sequencing reads were subjected to quality control using FastQC (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/) to assess read quality and remove low-quality reads. Adaptor sequences and low-quality bases were trimmed using Trimmomatic (http://www.us

Table 1
Basic information about the *CYP701A1* gene from different genomes.

Genes	Species	Genome	Chr	Location		NO. CDS	NO. aa	MW (KDa)	Theoretical pI	Instability Index	Aliphatic Index	Grand Average of Hydropathicity
				From	To							
GheCYP701A1	*G. herbaceum*	A1	Chr06	1,483,073	1,486,148	1509	502	57.68	7.65	44.27	87.23	−0.324
GarCYP701A1	*G. arboreum*	A2	Chr06	1,466,718	1,469,742	1509	502	57.76	8.28	44.2	88.01	−0.319
GanoCYP701A1	*G. anomalum*	B1	Chr06	1,625,253	1,628,333	1509	502	57.70	8.03	44.25	87.43	−0.309
GstuCYP701A1	*G. sturtianum*	C1	Chr06	161,707,714	161,710,978	1509	502	57.62	8.29	43.43	87.81	−0.314
GthCYP701A1	*G. thurberi*	D1	Chr10	1,432,896	1,435,965	1509	502	57.57	8.04	43.56	87.43	−0.31
GkarCYP701A1	*G. harknessii*	D2	Chr06	1,140,040	1,142,524	1515	504	57.90	8.28	45.42	87.28	−0.313
GdavCYP701A1	*G. davidsonii*	D3	Chr10	1,357,044	1,359,967	1509	502	57.65	8.03	45.32	87.43	−0.323
GkloCYP701A1	*G. klotzschianum*	D3	Chr06	1,483,076	1,485,999	1509	502	57.65	8.03	45.32	87.43	−0.323
GaridCYP701A1	*G. aridum*	D4	CM025189.1	979,772	982,726	1527	508	58.08	7.65	44.85	87.17	−0.293
GraCYP701A1	*G. raimondii*	D5	Chr10	1,221,824	1,224,656	1509	502	57.65	8.03	45.83	88.21	−0.318
GgosCYP701A1	*G. gossypioides*	D6	Chr06	1,253,575	1,256,523	1515	504	57.88	8.03	44.91	88.63	−0.294
GlobCYP701A1	*G. lobatum*	D7	CM025280.1	1,127,532	1,130,501	1521	506	57.82	8.03	44.48	88.08	−0.287
GtriCYP701A1	*G. trilobum*	D8	Chr06	1,232,260	1,235,338	1509	502	57.68	8.03	43.69	87.63	−0.318
GlaxCYP701A1	*G. laxum*	D9	CM025267.1	1,483,173	1,484,701	1221	406	47.13	6.26	45.14	86.26	−0.418
GtuCYP701A1	*G. turneri*	D10	D10_06	59,911,865	59,914,250	1416	471	54.18	6.96	43.55	85.94	−0.406
GschCYP701A1	*G. schwendimanii*	D11	CM025254.1	970,217	973,165	1515	504	57.70	7.65	44.21	85.73	−0.319
GstoCYP701A1	*G. stocksii*	E1	Chr06	2,052,726	2,055,852	1509	502	57.65	7.65	45.48	87.43	−0.3
GloCYP701A1	*G. longicalyx*	F1	Chr06	1,328,587	1,331,641	1509	502	57.64	7.65	44.89	87.05	−0.323
GbiCYP701A1	*G. bickii*	G1	chr06	132,156,670	132,159,784	1611	536	61.62	7.95	44.65	88.6	−0.228
GauCYP701A1	*G. australe*	G2	CM016619.1	144,183,953	144,187,074	1467	488	56.63	7.65	39.49	82.77	−0.396
GrotCYP701A1	*G. rotundifolium*	K2	Chr06	201,792,757	201,795,797	1509	502	57.78	8.47	44.6	89.36	−0.325
GkiCYP701A1	*Gossypioides kirkii*	–	KI_06	1,296,432	1,298,459	1407	468	53.77	8.02	39.85	86.47	−0.418
GhpCYP701A1_Dt	*G. hirsutum, race punctatum*	AD1	CM045517.1	1,491,758	1,494,096	1080	359	41.37	9.31	41.6	92.9	−0.14
GhlCYP701A1_Dt	*G. barbadense, race Lianhemumian*	AD2	D06	1,517,935	1,520,866	1515	504	57.83	8.03	44.41	88.43	−0.3
GtCYP701A1_Dt	*G. tomentosum*	AD3	D06	1,432,204	1,435,145	1515	504	57.80	8.03	44.79	88.06	−0.303
GmCYP701A1_Dt	*G. mustelinum*	AD4	CM017654.1	1,497,639	1,499,091	1101	366	42.57	5.71	46.72	87.43	−0.437
GdCYP701A1_Dt	*G. darwinii*	AD5	CM017706.1	1,437,974	1,440,905	1515	504	57.82	8.03	45.93	87.86	−0.31
GeCYP701A1_Dt	*G. ekmanianum*	AD6	CM046642.1	1,580,201	1,583,002	1473	490	56.28	9.46	44.29	86.35	−0.271
GsCYP701A1_Dt	*G. stephensii*	AD7	CM045536.1	1,475,136	1,478,080	1521	506	57.99	8.03	45.04	88.28	−0.297
GhpCYP701A1_At	*G. hirsutum, race punctatum*	AD1	CM045504.1	1,518,060	1,521,115	1509	502	57.73	8.03	44.67	87.43	−0.337
GhlCYP701A1_At	*G. barbadense, race Lianhemumian*	AD2	A06	1,553,313	1,556,359	1509	502	57.73	8.03	42.81	87.81	−0.326
GtCYP701A1_At	*G. tomentosum*	AD3	A06	1,501,754	1,504,682	1416	471	54.23	6.62	42.53	85.31	−0.432
GmCYP701A1_At	*G. mustelinum*	AD4	CM017641.1	1,472,398	1,475,455	1509	502	57.72	8.03	44.16	86.65	−0.338
GdCYP701A1_At	*G. darwinii*	AD5	CM017693.1	1,436,147	1,439,194	1509	502	57.72	8.03	42.7	87.81	−0.322
GeCYP701A1_At	*G. ekmanianum*	AD6	CM046641.1	1,444,555	1,447,474	1467	488	56.18	9.42	42.93	88.09	−0.253
GsCYP701A1_At	*G. stephensii*	AD7	CM045535.1	1,574,370	1,577,426	1509	502	57.72	8.03	43.68	87.23	−0.339
GbCYP701A1_Dt	*G. barbadense*	AD2	D06	1,405,252	1,408,184	1515	504	57.82	8.03	45.93	87.86	−0.31
GbCYP701A1_At	*G. barbadense*	AD2	A06	1,428,320	1,431,365	1509	502	57.73	8.03	42.81	87.81	−0.326
GhCYP701A1_Dt	*G. hirsutum*	AD1	D06	1,318,289	1,321,227	1515	504	57.82	8.03	45.18	87.86	−0.31
GhCYP701A1_At	*G. hirsutum*	AD1	A06	1,394,845	1,397,900	1509	502	57.73	8.03	44.67	87.43	−0.337

A
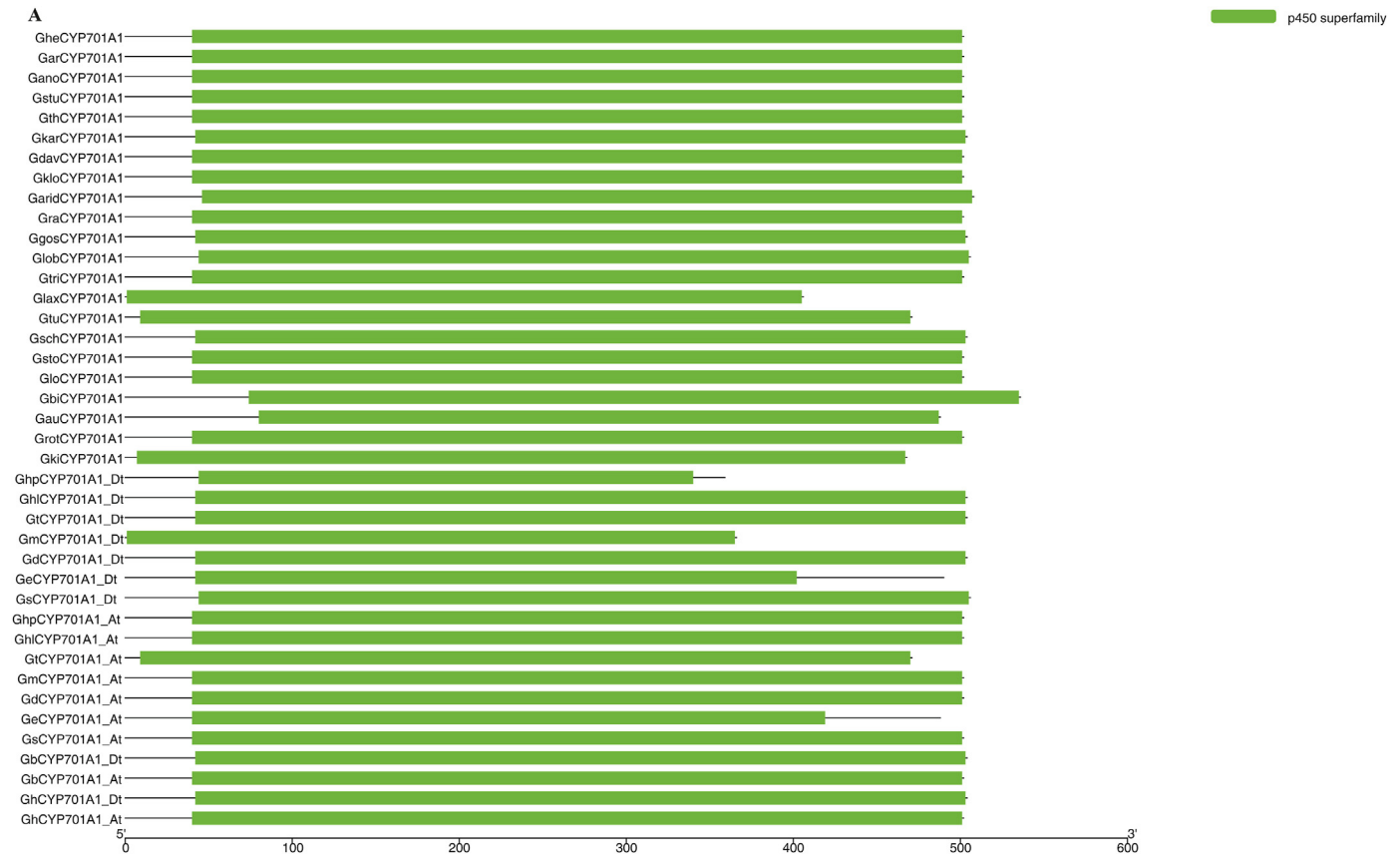
Fig. 1. Identification and bioinformatics analysis of *CYP701A1* from different cotton genomes.
A. Conserved domain; B. Evolution and gene structure analysis of *CYP701A1* gene from different genomes; C. Promoter *cis*-acting element analysis.

adellab.org/cms/?page=trimmomatic).

Cleaned reads were then aligned to a reference genome or transcriptome using alignment tools Bowtie2 (http://bowtie-bio.sourceforge.net/bowtie2/index.shtml). The aligned reads were quantified to estimate gene expression levels using tools HTSeq (https://htseq.readthedocs.io/en/master/).

To identify differentially expressed genes (DEGs), we adopted a conservative criteria by selection of the consistent results of cuffdiff (ref) with $\log_2$ (fold change) ≥ 1 or ≤ -1 and also being significantly expressed with FDR corrections at the level of 0.05 and genes FPKM value ≥ 1.

3. Results

3.1. Sequence analysis of CYP701A1 from different gossypium genome

Phylogenetic analysis of 679 P450 family members in the TM-1 genome and 237 members of *Arabidopsis thaliana* revealed that *Gh_A05G1711*, *Gh_D05G1901*, *Gh_A06G0128*, and *Gh_D06G0106* genes are grouped together with *AtCYP701A3* (Fig. S1). Particularly, the expression level of *Gh_D06G0106* in the transcriptome data of TM-1 tissue expression was notably higher than other genes (Fig. S2), leading to a focus on *Gh_A06G0128* and *Gh_D06G0106* genes, which were subsequently named *GhCYP701A1*. Through homologous comparison, a total of 40 *CYP701A1* genes were obtained from different cotton genomes, including 22 from different diploid genomes and 18 from heterologous tetraploids, with nine each from the At and Dt sub-genomes. The smallest protein expressed by these genes consists of 359 amino acids, originating from the Dt sub-genome of upland cotton, an early domesticated species - *Gossypium hirsutum* var. *Punctatum*, known as *GhpCYP701A1_Dt*, while the largest protein comprises 536 amino acids from the G1 genome of *Gossypium bickii*, *GbiCYP701A1*. The molecular weights of the proteins

range from 41.37 kDa to 61.62 kDa, with theoretical isoelectric points between 5.71 and 9.46. Additionally, all proteins are hydrophobic in nature (Table 1).

Using NCBI, it was predicted that these *CYP701* genes from different genomes all contain a conserved P450 superfamily domain consisting of 259–465 amino acids (Fig. 1A). Correspondingly, these 40 genes have at least four introns interspersed within the coding sequences (*GhpCYP701A1_Dt*, *GmCYP701_Dt* and *GlaxCYP701A1*), with two genes, *GeCYP701A1_At* and *GeCYP701A1_Dt*, from the At and Dt subgenomes of *Gossypioides ekmanianum* of the AD6 genomes containing five introns, while the remaining genes contain 6 introns each (Fig. 1B).

3.2. Phylogenetic analysis of CYP701A1 from different gossypium genome

To determine the evolutionary relationships between 40 *CYP701A1* genes from different genomes, a maximum likelihood method was employed to construct their evolutionary tree, with *Gossypioides kirkii* used as an outgroup for tree construction. The results indicate that *GanoCYP701A1* from the B1 genome is evolutionarily closest to *GkiCYP701A1* from the outgroup *Gossypioides kirkii*. The remaining *CYP701A1* genes from different genomes can be classified into three major clades: the first clade includes *CYP701A1* genes from the heterologous tetraploid At sub-genome, along with those from the A1, A2, and F1 genomes; the second clade comprises *CYP701A1* genes from the C1, E1, G1, G2, and K genomes; and the third clade consists of *CYP701A1* genes from different D genomes and the heterologous tetraploid Dt subgenome. Regarding the early domesticated species, *Gossypium hirsutum* var. *Punctatum* and *Gossypium barbadense* var *Lianhemumian*, the *CYP701A1* genes from the At sub-genome exhibit the closest phylogenetic relationship in both early domesticated and cultivated species, whereas this is not observed for the Dt sub-genome (Fig. 1B).
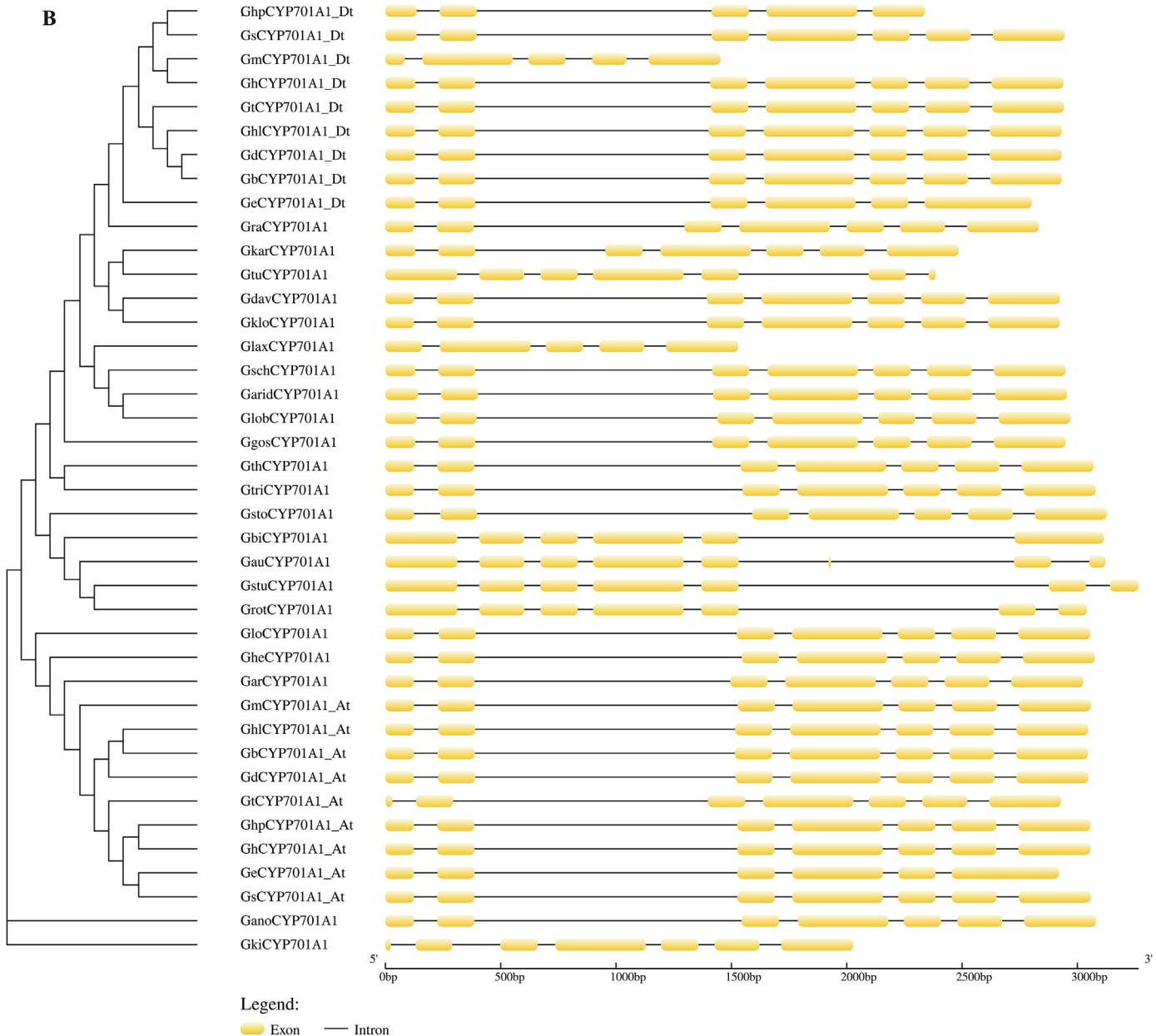
Fig. 1. (*continued*).

3.3. Promoter analysis of CYP701A1 from different gossypium genome

The upstream 2000-bp sequences of each of these 40 genes in their respective genomes were extracted for promoter analysis. Through analysis, it was found that, excluding the core promoter elements (TATA-box and CAAT-box), a total of 15 different types of cis-regulatory elements were identified. Among them, the most abundant elements were light-responsive elements, such as G-BOX, GT1-motif, GA-motif, GATA-motif, etc. (Fig. 1C).

3.4. The function of GhCYP701A1 was analyzed through gene silencing

To verify the function of *GhCYP701A1*, we employed VIGS technology to silence this gene in *G. hirsutum* ac. TM-1. The 553bp sequence from *GH_A06G0147* was cloned into the pTRV2 vector to suppress the expression of this gene in plant bodies. Upon the appearance of albino phenotype in the plants, the expression level of *GhCYP701A1* gene in silenced plants was assessed, and the plant height was measured when the plants grew to four true leaves. The results revealed a significant decrease in the expression level of the *GhCYP701A1* gene in silenced plants, accompanied by a markedly lower plant height compared to control plants (Fig. 2A–C). Subsequently, hormone content was determined in silenced and control plants. The results showed a significant reduction in gibberellin GA3 content in gene-silenced plants compared to control plants (Fig. 2D), while the content of auxin, zeatin, and methyl jasmonate was significantly increased in silenced plants (Fig. 2E–G); there was no significant difference in the content of ethylene precursor ACC between silenced plants and controls (Fig. 2H), although the content of salicylic acid was significantly lower in silenced plants compared to controls (Fig. 2I).

3.5. Transcriptome analysis via GhCYP701A1 gene silencing

To assess the change in gene expression levels of plants after silencing the *GhCYP701A1* gene, transcriptome sequencing revealed a total of 15,962 differentially expressed genes, including 8376 upregulated and
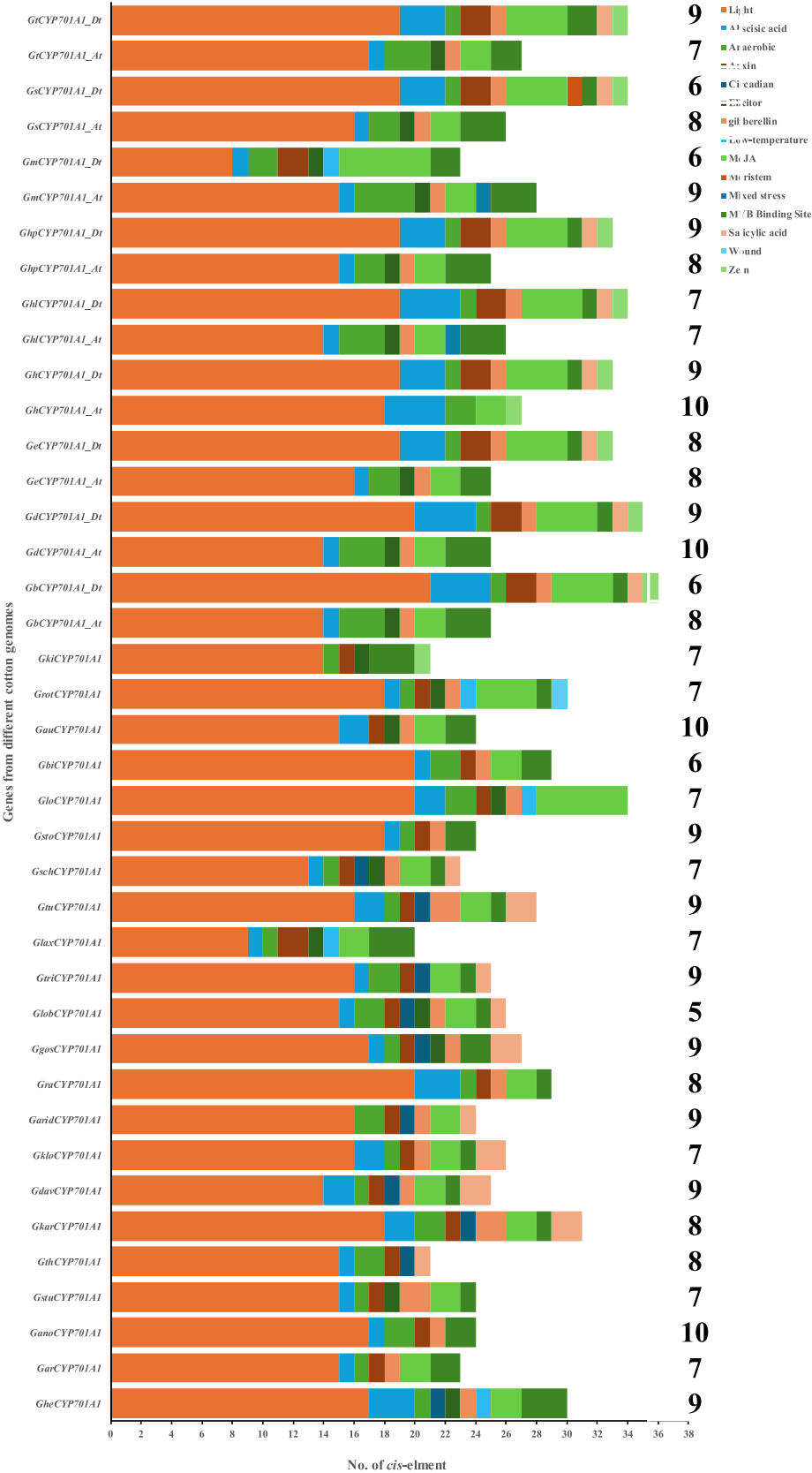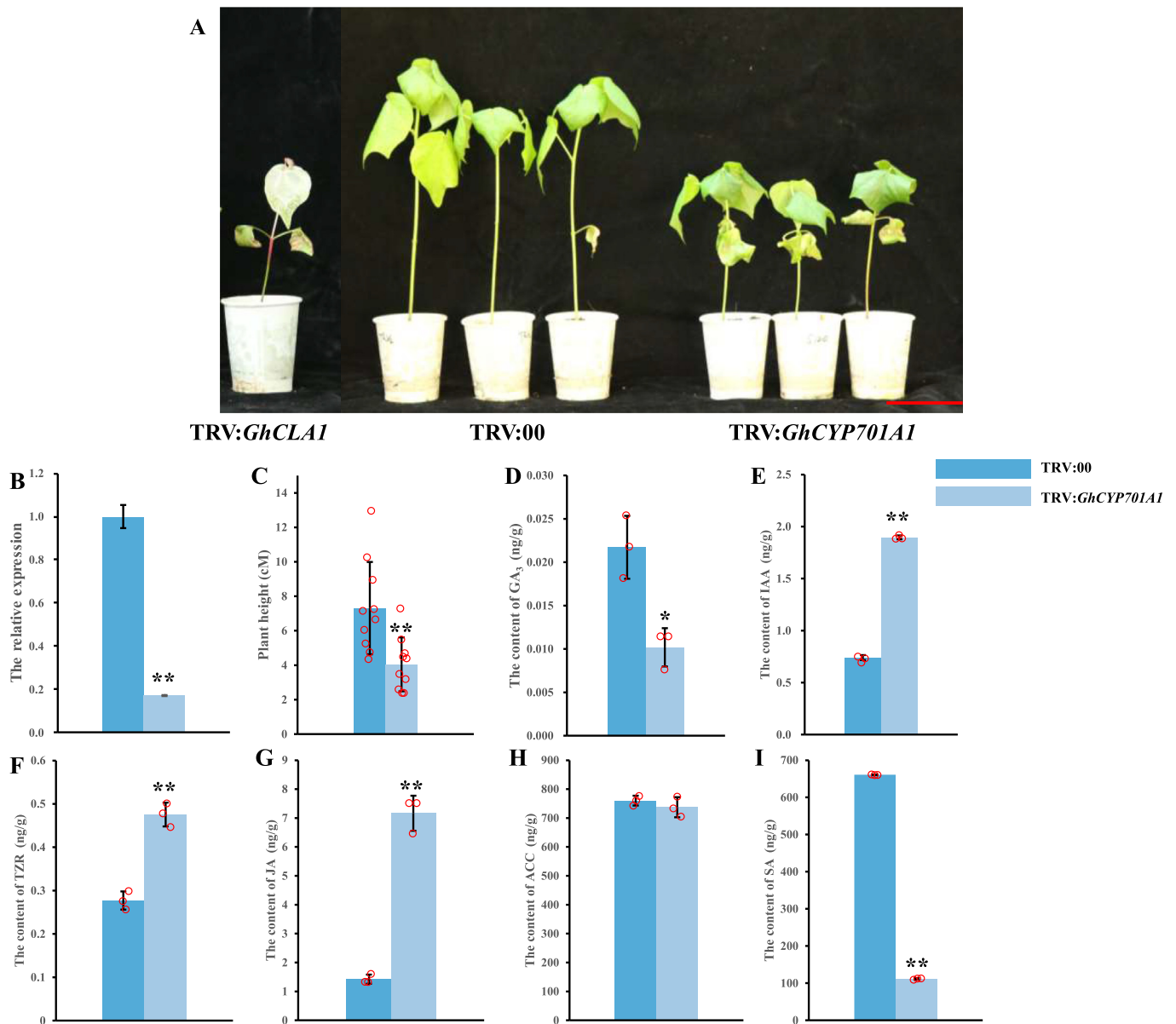
Fig. 1. (*continued*).

Fig. 2. Plant phenotype and hormone content by *GhCYP701A1* gene silencing.
A. Plant phenotype. Scale bars = 5 cm; B. The relative expression of *GhCYP701A1*; C. The plant height; D. The content of GA_3; E. The content of IAA; F. The content of TZR; G. The content of JA; H. The content of ACC; I. The content of SA. The data are mean ± SE of three biological replicates. * and ** indicate statistical significance at the 0.05 and 0.01 probability level, respectively.

7586 downregulated genes (FDR corrections at the level of 0.05 and genes FPKM value $\geq$ 1) (Fig. 3A). Among all the up-regulated genes, there were 2204 genes with more than double difference ratio. There are 2810 genes in the down-regulated genes (Fig. 3B).

To determine the biological processes regulated by the *GhCYP701A1* gene, all the differentially expressed genes obtained from the transcriptome was subjected to GO enrichment analysis. GO analysis showed that the differentially expressed genes associated with *GhCYP701A1* gene silencing were classified into three major categories: biological processes, molecular functions, and cellular components (Fig. 4A). For biological processes, they mainly included metabolic process, cellular process, single-organism process, biological regulation, response to stimulus, localization, cellular component organization or biogenesis, developmental process, signaling, multicellular organismal process, reproductive process, multi-organism process, immune system process, growth; reproduction, biological adhesion, rhythmic process, etc. Molecular functions covered catalytic activity, binding, transporter activity

nucleic acid binding transcription factor activity, molecular transducer activity, electron carrier activity, receptor activity, structural molecule activity, antioxidant activity, enzyme regulator activity, protein binding transcription factor activity, nutrient reservoir activity, guanyl-nucleotide exchange factor activity, etc. Cellular components mainly included cell part; organelle; membrane part; organelle part; macromolecular complex; extracellular region; cell junction; membrane-enclosed lumen; extracellular matrix; virion part; extracellular region part; nucleoid, etc. Compared to TRV:00, the proportion of upregulated genes in these GO enrichment pathways was significantly higher than downregulated genes after *GhCYP701A1* gene silencing. However, in molecular functions, there were four pathways with downregulated genes higher than upregulated genes (nucleic acid binding transcription factor activity, molecular transducer activity, receptor activity, and antioxidant activity), while for biological processes, five pathways with downregulated genes were higher than upregulated genes (response to stimulus, localization, signaling, multi-organism process, and immune system
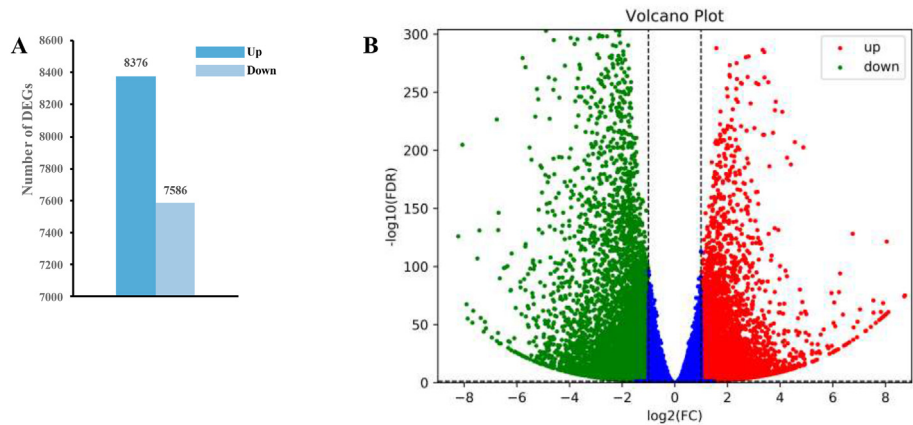
Fig. 3. Statistical analysis of differently expressed genes (DEGs) between TRV:00 and TRV: *GhCYP701A1*.
A. Total number of up-regulated and down-regulated DEGs; B. Volcano distribution map of up-regulated and down-regulated DEGs.
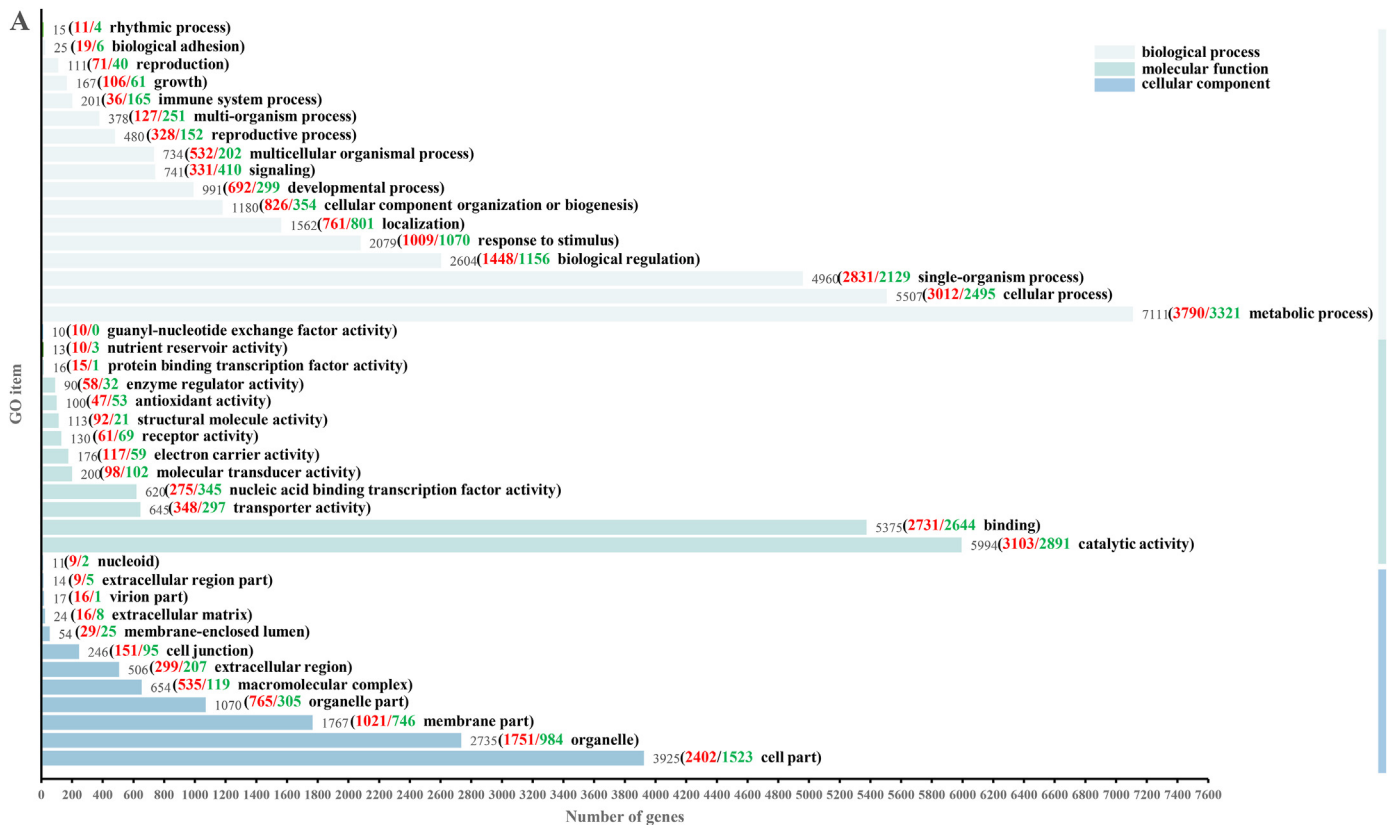
Fig. 4. Functional enrichment analysis of differently expressed genes (DEGs).
A. Enriched gene ontology (GO) terms for DEGs identified. The DEGs were classified based on GO annotation and were categorized into biological process (BP), molecular function (MF), and cellular component (CC); B. Bubble map of the top 20 KEGG enrichments for DEGs. The ordinate denotes the pathway name, the abscissa describes the richness factor, each circle in the map represents a pathway, the size of a circle indicates the number of DEGs in the pathway, and the color of a circle corresponds to different p-value ranges; C. KO (KEGG Ontology) enrichment circle diagram. From the outside to the inside, the first circle represents the top 20 enrichment pathways, and the number outside the circle is the coordinate ruler of the number of genes; The second circle represents the number and Q value of background genes in this pathway, and the more genes, the longer bar; The third circle represents the number of the DEGs in this pathway; The fourth circle represents the value of Rich Factor in each pathway.

process). Such the situation was not observed in cellular components (Fig. 4B).

We subsequently conducted KEGG enrichment analysis to gain a broader understanding for the biological changes that occurred in plants after silencing the *GhCYP701A1* gene. The results showed that the differentially expressed genes were mainly enriched in three major KEGG classification pathways: metabolism; environmental information processing; and organismal systems. Interestingly, organismal systems were enriched in a metabolic pathway called plant-pathogen interaction, which consisted of 113 differentially expressed genes. This pathway included 32 up-regulated genes and 81 down-regulated genes. Additionally, two metabolic pathways related to environmental information
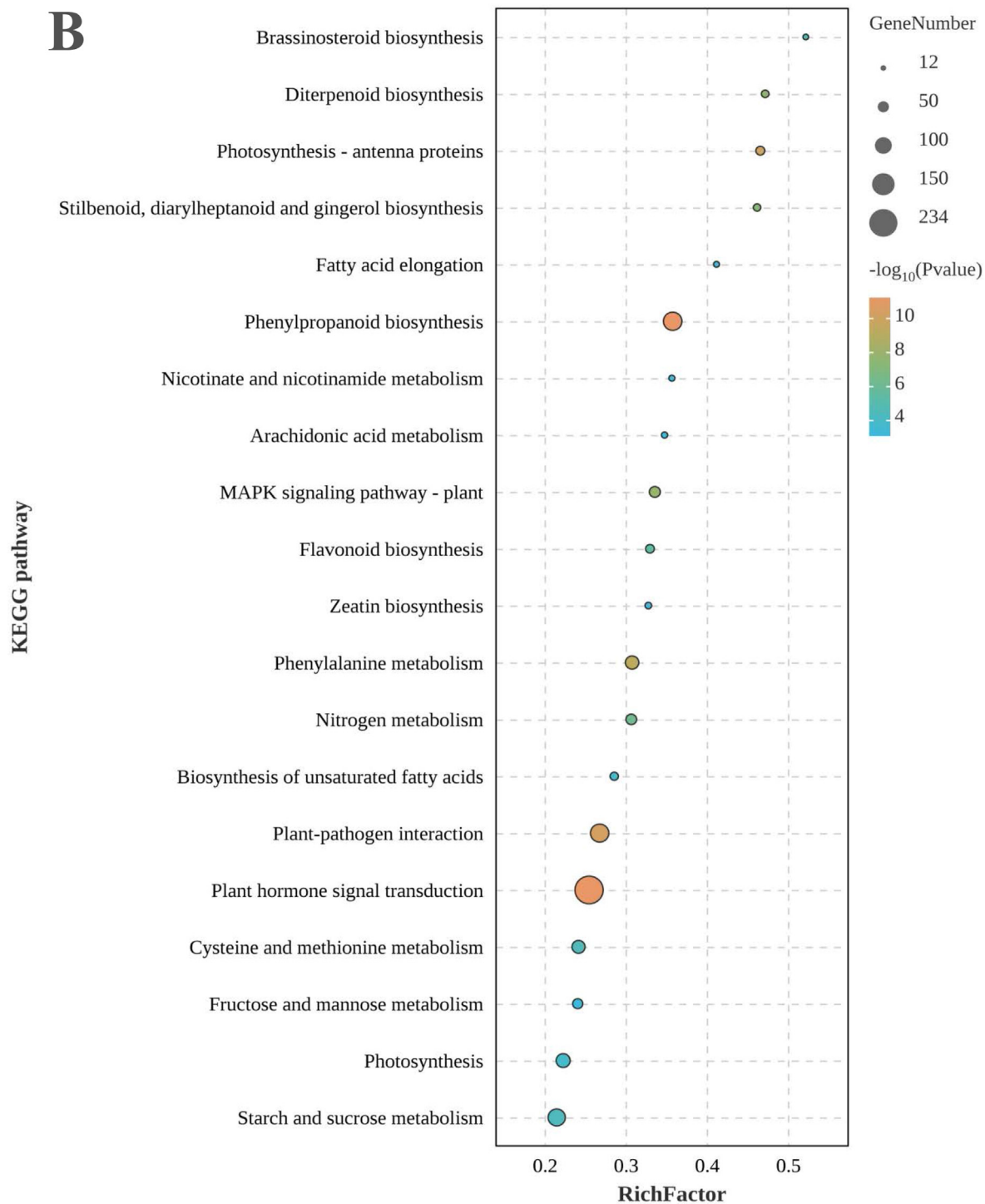
Fig. 4. (*continued*).

processing were identified: plant hormone signal transduction (234 DEGs, with 130 up-regulated genes and 104 down-regulated genes) and ABC transporters (10 DEGs, with six up-regulated genes and four down-regulated genes). The remaining 17 pathways were categorized under metabolism. (Fig. 4C).

The crucial role of *GhCYP701A1* gene in gibberellin synthesis was highlighted through KEGG enrichment analysis. Among the enriched metabolic pathways, 'plant hormone signal transduction' stood out with 234 differentially expressed genes. These genes were predominantly distributed as follows: 96 involved in auxin signal transduction (70 upregulated, 26 downregulated), 43 in ABA signal transduction (10

upregulated, 33 downregulated), 23 in cytokinin signal transduction (12 upregulated, 11 downregulated), 20 in BR signal transduction (13 upregulated, 7 downregulated), 19 in JA signal transduction (14 upregulated, 5 downregulated), 13 in GA signal transduction (all downregulated), and 12 in SA signal transduction (8 upregulated, 4 downregulated). The silencing of *GhCYP701A1* gene impaired gibberellin synthesis, thereby leading to only 8 genes involving in the gibberellin signal transduction pathway compared to other hormone signal transduction pathways, with the least differentially expressed genes (3 upregulated, 5 downregulated). (Fig. 5, Table S1).
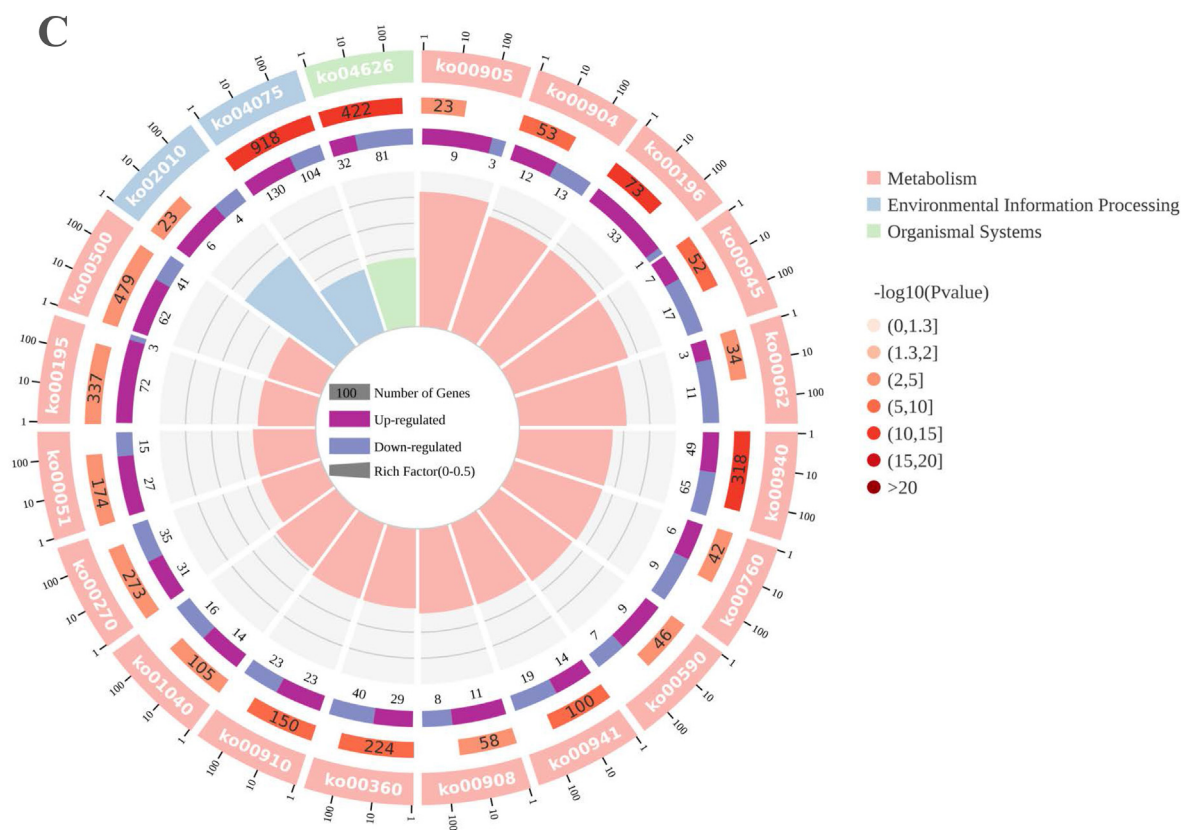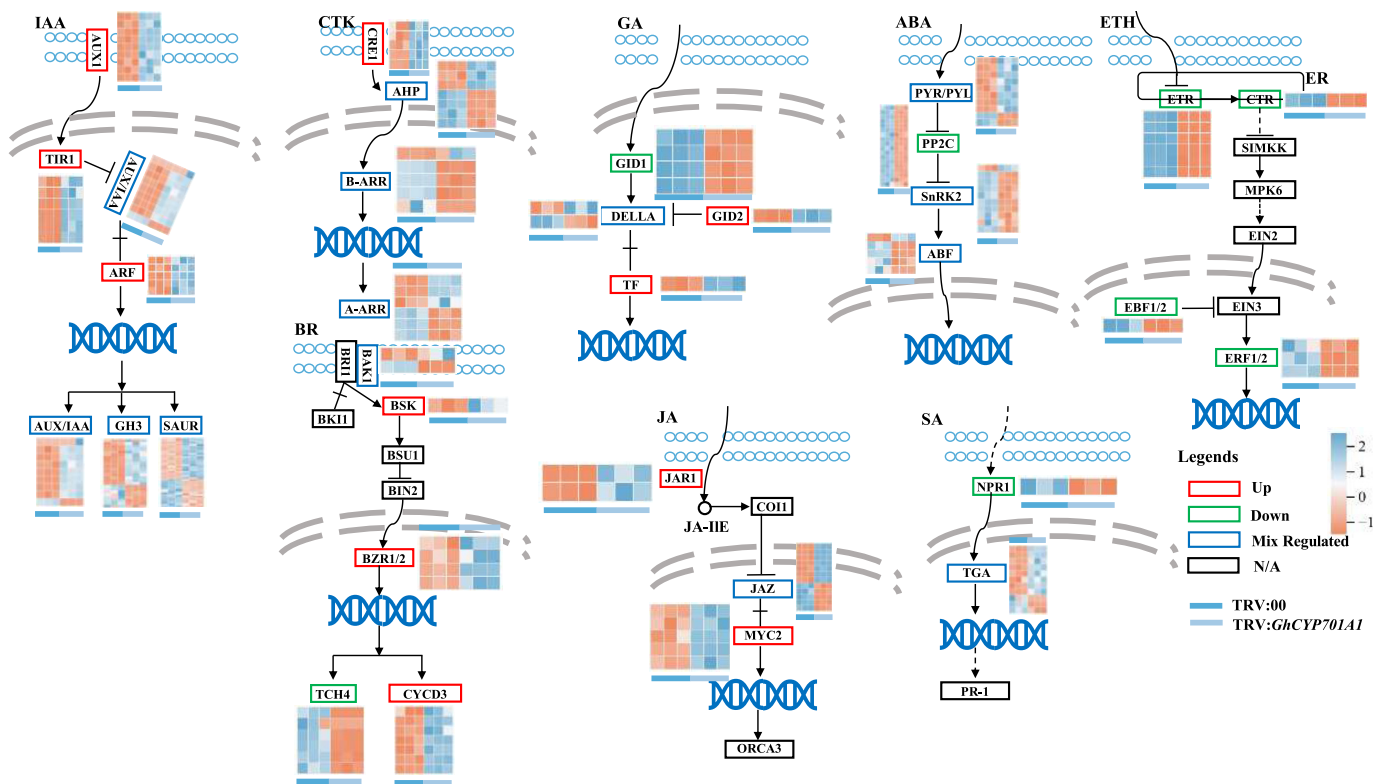
Fig. 4. (*continued*).

Fig. 5. Pathway analysis of differentially expressed genes (DEGs) in hormone signaling.
The neatly arranged blue circles form the cell membrane and gray dotted lines comprise the nucleus. The functionally annotated DEGs, all up-regulated genes are represented by red boxes, all down-regulated genes are represented by green boxes, those containing both up- and down-regulated genes are represented by blue boxes, and those without differentially expressed genes are represented by black boxes. A heat map of DEGs is placed next to the annotated gene.

4. Discussion

Cytochrome P450 enzymes, particularly the CYP701 family, have been paid widespread attention in evolutionary studies, due to their pivotal roles in diverse biological processes across different organisms. Recent research has provided valuable insights into the evolutionary history, functional diversification, and adaptive significance of CYP701 enzymes. Phylogenetic analyses have revealed the ancient origin of the CYP701 family, tracing back to early eukaryotic lineages. Comparative genomic studies have identified the ortholog of *CYP701* genes in various plant species as well as in fungi and certain bacteria, suggesting their presence in ancestral genomes and subsequent divergence during evolution [9]. Evolutionary analysis of *CYP701A1* in various genomes of the *Gossypium* genus revealed that *GanoCYP701A1* from the B genome shows a closer evolutionary relationship to the outgroup *Gossypioides kirkii* than *CYP701A1* genes in other genomes. The remaining *CYP701A1* genes from different genomes can be categorized into three main branches. The first clade comprises *CYP701A1* gene from the allotetraploid at subgenome, along with the genes from the A1, A2, and F1 genomes. The second clade includes *CYP701A1* genes from the C1, E1, G1, G2, and K genomes. Lastly, the third clade is composed of *CYP701A1* genes from various D genomes and the allotetraploid Dt subgenome.

Gibberellins (GAs) play pivotal roles in regulating various aspects of plant growth and development, including seed germination, stem elongation, and flower and fruit development. The inhibition or disruption of GA biosynthesis, as observed in the case of *GhCYP701A1* gene silencing, leads to significant alterations in the hormonal balance within plants. The inhibition of gibberellin (GA) biosynthesis or signaling pathways often leads to significant alterations in the levels and activities of other plant hormones, thereby influencing various aspects of plant growth and development. One notable consequence of gibberellin deficiency is the upregulation of abscisic acid (ABA) biosynthesis and signaling pathways. ABA, known for its roles in stress responses and seed dormancy, tends to accumulate in response to gibberellin depletion, leading to increased tolerance to environmental stresses [34]. Additionally, ABA can inhibit seed germination and promote root growth, thereby modulating plant growth patterns under gibberellin-deficient conditions [35]. Furthermore, alterations in auxin levels and distribution patterns have been observed following gibberellin biosynthesis inhibition. Auxin, a key regulator of plant growth and development, plays crucial roles in various processes such as cell elongation, organ formation, and tropic responses [36]. The crosstalk between gibberellins and auxin pathways is complex and involves the regulation of auxin biosynthesis, transport, and signaling components [37]. When *GhCYP701A1* was silenced in *G. hirsutum* ac. TM-1 using VIGS technology, transcriptome analysis showed notable alterations in the hormone signal transduction pathway. Subsequent measurement of hormone content in plants indicated a decrease in gibberellin levels, along with significant increases in auxin, cytokinin, and jasmonic acid contents. Conversely, salicylic acid content decreased significantly, while the ethylene synthesis precursor 1-aminocyclopropane-1-carboxylic acid (ACC) content remained relatively unchangeable.

5. Conclusions

In this study, genomic data from thirty *Gossypium* species were analyzed, which included nine allotetraploid genomes, 20 diploid genomes, and the *Gossypioides kirkii* genome used as an outgroup for evolutionary analysis. The remaining *CYP701A1* genes from different genomes were categorized into three main clades: the first clade encompassed *CYP701A1* genes from the heterologous tetraploid at subgenome, as well as those from the A1, A2, and F1 genomes; the second clade included *CYP701A1* genes from the C1, E1, G1, G2, and K genomes; and the third clade comprised *CYP701A1* genes from various D genomes and the heterologous tetraploid Dt sub-genome. Silencing *GhCYP701A1* in *G. hirsutum* ac. TM-1 using VIGS technology resulted in significant alterations in hormone levels. Gibberellin levels decreased, while auxin, cytokinin, and jasmonic acid contents increased notably. Conversely, salicylic acid content decreased, while the precursor for ethylene synthesis, 1-aminocyclopropane-1-carboxylic acid (ACC), remained relatively stable.

CRediT authorship contribution statement

Zhao Liang: Data curation, Funding acquisition, Writing – original draft. **Di Jiachun:** Data curation. **Guo Qi:** Data curation. **Xu Zhenzhen:** Formal analysis. **Zhao Jun:** Formal analysis. **Xu Peng:** Data curation. **Xu Jianwen:** Data curation. **Liu Jianguang:** Data curation. **Shen Xinlian:** Conceptualization, Resources. **Chen Xusheng:** Writing – review & editing.

Declaration of competing interest

The authors declare that they have no conflict of interest.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.cropd.2024.100081.

References

[1] P. Achard, P. Genschik, Releasing the brakes of plant growth: how GAs shutdown DELLA proteins, J. Exp. Bot. 60 (4) (2009) 1085–1092.
[2] P. Hedden, V. Sponsel, A century of gibberellin research, J. Plant Growth Regul. 34 (4) (2015) 740–760.
[3] J.M. Daviere, P. Achard, Gibberellin signaling in plants, Development 140 (6) (2013) 1147–1151.
[4] J. MacMillan, Occurrence of gibberellins in vascular plants, fungi, and bacteria, J. Plant Growth Regul. 20 (4) (2002) 387–442.
[5] P. Hedden, A.L. Phillips, Gibberellin metabolism: new insights revealed by the genes, Trends Plant Sci. 5 (12) (2000) 523–530.
[6] T.P. Sun, Gibberellin Metabolism, Perception and Signaling Pathways in Arabidopsis, vol. 6, The Arabidopsis book/American Society of Plant Biologists, 2008 e0103.
[7] S. Ghosh, Triterpene structural diversification by plant cytochrome P450 enzymes, Front. Plant Sci. 8 (2017).
[8] Y.X. Li, K.F. Wei, Comparative functional genomics analysis of cytochrome P450 gene superfamily in wheat and maize, BMC Plant Biol. 20 (1) (2020).
[9] D. Nelson, D. Werck-Reichhart, A P450-centric view of plant evolution, Plant J. 66 (1) (2011) 194–211.
[10] S. Bak, F. Beisson, G. Bishop, B. Hamberger, R. Hofer, S. Paquette, D. Werck-Reichhart, Cytochromes P450, vol. 9, The Arabidopsis book/American Society of Plant Biologists, 2011 e0144.
[11] D. Morrone, X.M. Chen, R.M. Coates, R.J. Peters, Characterization of the kaurene oxidase CYP701A3, a multifunctional cytochrome P450 from gibberellin biosynthesis, Biochem. J. 431 (2010) 337–344.
[12] M. Xu, Y. Lu, H. Yang, J. He, Z. Hu, X. Hu, M. Luan, L. Zhang, Y. Fan, L. Wang, ZmGRF, a GA regulatory factor from maize, promotes flowering and plant growth in Arabidopsis, Plant Mol. Biol. 87 (1–2) (2015) 157–167.
[13] N. Dhar, D.P.G. Short, B.E. Mamo, A.J. Corrion, C.M. Wai, A. Anchieta, R. VanBuren, B. Day, H. Ajwa, K.V. Subbarao, S.J. Klosterman, Arabidopsis defense mutant ndr1-1 displays accelerated development and early flowering mediated by the hormone gibberellic acid, Plant Sci. : an international journal of experimental plant biology 285 (2019) 200–213.
[14] Y. Xu, Y. Wei, Z. Zhou, X. Cai, S.A. Boden, M.J. Umer, L.B. Safdar, Y. Liu, D. Jin, Y. Hou, et al., Widespread incomplete lineage sorting and introgression shaped adaptive radiation in the *Gossypium* genus, Plant Commun 5 (2) (2024) 100728.
[15] M.J. Wang, J.Y. Li, Z.Y. Qi, Y.X. Long, L.L. Pei, X.H. Huang, C.E. Grover, X.M. Du, C.J. Xia, P.C. Wang, et al., Genomic innovation and regulatory rewiring during evolution of the cotton genus *Gossypium*, Nat. Genet. 54 (12) (2022) 1959–1971.

[16] C. Shen, N.A. Wang, D. Zhu, P.C. Wang, M.J. Wang, T.W. Wen, Y. Le, M. Wu, T. Yao, X.L. Zhang, Z.X. Lin, *Gossypium tomentosum* genome and interspecific ultra-dense genetic maps reveal genomic structures, recombination landscape and flowering depression in cotton, Genomics 113 (4) (2021) 1999–2009.

[17] Y. Hu, J. Chen, L. Fang, Z. Zhang, W. Ma, Y. Niu, L. Ju, J. Deng, T. Zhao, J. Lian, et al., *Gossypium barbadense* and *Gossypium hirsutum* genomes provide insights into the origin and evolution of allotetraploid cotton, Nat. Genet. 51 (4) (2019) 739–748.

[18] Z.J. Chen, A. Sreedasyam, A. Ando, Q.X. Song, L.M. De Santiago, A.M. Hulse-Kemp, M.Q. Ding, W.X. Ye, R.C. Kirkbride, J. Jenkins, et al., Genomic diversifications of five *Gossypium* allopolyploid species and their impact on cotton improvement, Nat. Genet. 52 (5) (2020) 525–533.

[19] R.H. Peng, Y.C. Xu, S.L. Tian, T. Unver, Z. Liu, Z.L. Zhou, X.Y. Cai, K.B. Wang, Y.Y. Wei, Y.L. Liu, et al., Evolutionary divergence of duplicated genomes in newly described allotetraploid cottons, Proc. Natl. Acad. Sci. U.S.A. 119 (39) (2022).

[20] G. Huang, Z.G. Wu, R.G. Percy, M.Z. Bai, Y. Li, J.E. Frelichowski, J. Hu, K. Wang, J.Z. Yu, Y.X. Zhu, Genome sequence of *Gossypium herbaceum* and genome updates of *Gossypium arboreum* and *Gossypium hirsutum* provide insights into cotton A-genome evolution, Nat. Genet. 52 (5) (2020) 516–524.

[21] Z.Z. Xu, J.D. Chen, S. Meng, P. Xu, C.J. Zhai, F. Huang, Q. Guo, L. Zhao, Y.G. Quan, Y.X. Shangguan, et al., Genome sequence of *Gossypium anomalum* facilitates interspecific introgression breeding, Plant Commun 3 (5) (2022) 100350.

[22] K. Sheng, Y. Sun, M. Liu, Y.F. Cao, Y.F. Han, C. Li, U. Muhammad, M.K. Daud, W.R. Wang, H.Z. Li, et al., A reference-grade genome assembly for *Gossypium bickii* and insights into its genome evolution and formation of pigment glands and gossypol, Plant Commun 4 (1) (2023) 100421.

[23] Y. Cai, X. Cai, Q. Wang, P. Wang, Y. Zhang, C. Cai, Y. Xu, K. Wang, Z. Zhou, C. Wang, et al., Genome sequencing of the Australian wild diploid species *Gossypium australe* highlights disease resistance and delayed gland morphogenesis, Plant Biotechnol. J. 18 (3) (2020) 814–828.

[24] Z.E. Yang, X.Y. Ge, W.N. Li, Y.Y. Jin, L.S. Liu, W. Hu, F.Y. Liu, Y.L. Chen, S.L. Peng, F.G. Li, Cotton D genome assemblies built with long-read data unveil mechanisms of centromere evolution and stress tolerance divergence, BMC Biol. 19 (1) (2021).

[25] C.E. Grover, M.A. Arick, A. Thrash, J.L. Conover, W.S. Sanders, D.G. Peterson, J.E. Frelichowski, J.A. Scheffler, B.E. Scheffler, J.F. Wendel, Insights into the evolution of the new world diploid cottons (*Gossypium*, subgenus *Houzingenia*) based on genome sequencing, Genome Biology and Evolution 11 (1) (2018) 53–71.

[26] A.H. Paterson, J.F. Wendel, H. Gundlach, H. Guo, J. Jenkins, D. Jin, D. Llewellyn, K.C. Showmaker, S. Shu, J. Udall, et al., Repeated polyploidization of *Gossypium* genomes and the evolution of spinnable cotton fibres, Nature 492 (7429) (2012) 423–427.

[27] J.A. Udall, E. Long, C. Hanson, D. Yuan, T. Ramaraj, J.L. Conover, L. Gong, M.A. Arick, C.E. Grover, D.G. Peterson, J.F. Wendel, *De novo* genome sequence assemblies of *Gossypium raimondii* and *Gossypium turneri*, G3 Genes|Genomes|Genetics 9 (10) (2019) 3079–3085.

[28] J.A. Udall, E. Long, T. Ramaraj, J.L. Conover, D.J. Yuan, C.E. Grover, L. Gong, M.A. Arick, R.E. Masonbrink, D.G. Peterson, J.F. Wendel, The genome sequence of *Gossypioides kirkii* illustrates a descending dysploidy in plants, Front. Plant Sci. 10 (2019).

[29] B. Hu, J.P. Jin, A.Y. Guo, H. Zhang, J.C. Luo, G. Gao, Gsds 2.0: an upgraded gene feature visualization server, Bioinformatics 31 (8) (2015) 1296–1297.

[30] S. Kumar, G. Stecher, M. Li, C. Knyaz, K. Tamura, X. Mega, Molecular evolutionary genetics analysis across computing platforms, Mol. Biol. Evol. 35 (6) (2018) 1547–1549.

[31] M. Lescot, P. Déhais, G. Thijs, K. Marchal, Y. Moreau, Y. Van de Peer, P. Rouzé, S. Rombauts, PlantCARE, a database of plant *cis*-acting regulatory elements and a portal to tools for analysis of promoter sequences, Nucleic Acids Res. 30 (1) (2002) 325–327.

[32] K.J. Livak, T.D. Schmittgen, Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta CT}$ Method, Methods 25 (4) (2001) 402–408.

[33] D. Ma, Y. Hu, C. Yang, B. Liu, L. Fang, Q. Wan, W. Liang, G. Mei, L. Wang, H. Wang, et al., Genetic basis for glandular trichome formation in cotton, Nat. Commun. 7 (2016) 10456.

[34] M. Seo, T. Koshiba, Complex regulation of ABA biosynthesis in plants, Trends Plant Sci. 7 (1) (2002) 41–48.

[35] R.R. Finkelstein, S.S.L. Gampala, C.D. Rock, Abscisic acid signaling in seeds and seedlings, Plant Cell 14 (2002) S15–S45.

[36] R. Swarup, R. Bhosale, Developmental roles of AUX1/LAX auxin influx carriers in plants, Front. Plant Sci. 10 (2019).

[37] X.D. Fu, N.P. Harberd, Auxin promotes *Arabidopsis* root growth by modulating gibberellin response, Nature 421 (6924) (2003) 740–743.