frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 08 February 2023
DOI 10.3389/fpls.2023.1126884

Check for updates

OPEN ACCESS

EDITED BY
Libei Li,
Zhejiang Agriculture and Forestry University, China

REVIEWED BY
Bhagwat Dadarao Nawade,
Kongju National University, Republic of Korea
Juan Huang,
Guizhou Normal University, China
Shen Qian,
Institute of Cotton Research (CAAS), China

*CORRESPONDENCE
Taili Nie
n07061018@163.com
Liangqing Sun
slq-98@163.com

†These authors have contributed equally to this work

SPECIALTY SECTION
This article was submitted to Plant Abiotic Stress, a section of the journal Frontiers in Plant Science

RECEIVED 18 December 2022
ACCEPTED 30 January 2023
PUBLISHED 08 February 2023

CITATION
Cui A, Jin Y, Li Y, Nie T and Sun L (2023) Systematic identification of TPS genes in Gossypium and their characteristics in response to flooding stress.
Front. Plant Sci. 14:1126884.
doi: 10.3389/fpls.2023.1126884

COPYRIGHT
© 2023 Cui, Jin, Li, Nie and Sun. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). 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.

Systematic identification of TPS genes in Gossypium and their characteristics in response to flooding stress

Aihua Cui[1†], Yunqian Jin[2†], Yongqi Li[1], Taili Nie[1*] and Liangqing Sun[1*]

[1]Scientific Research Office, Economic Crop Institute of Jiangxi Province, Jiujiang, Jiangxi, China,
[2]College of Agronomy, Henan University of Science and Technology, Luoyang, China

Terpene synthases (TPS) is a key enzyme in the synthesis of plant terpenoids. Studies on TPSs have not been reported in *Gossypium barbadense* and *Gossypium arboreum*. 260 TPSs were identified in Gossypium, including 71 in *Gossypium hirsutum*, 75 in *Gossypium. barbadense*, 60 in *Gossypium. arboreum*, and 54 in *Gossypium raimondii*. We systematically analyzed the TPS gene family of Gossypium from three aspects: gene structure, evolutionary process and gene function. (1) Gene structure: Based on the protein structure of two conserved domains (PF01397 and PF03936), the TPS gene family is divided into five clades: TPS -a, -b, -c, -e/f and -g. (2) Evolution: Whole genome duplication and segmental duplication are the main modes of TPS gene amplification. (3) Function: The abundance of cis-acting elements may reveal the functional diversity of TPSs in cotton. TPS gene has tissue specific expression in cotton. The hypomethylation of the exon of TPSs may help to enhance the adaptability of cotton to flooding stress. In conclusion, this study can broaden the understanding of structure-evolution-function of the TPS gene family, and provide reference for the mining and verification of new genes.

KEYWORDS
Cotton, terpene synthases, gene family, evolution, function

Introduction

Terpenoids are the largest group metabolites in plant and respond positively to plant biotic and abiotic stresses (Cane, 2000; Yazaki et al., 2017; Huang et al., 2021).Terpenoids can be divided into broad categories according to their functions: primary metabolites and secondary metabolites. More than 55,000 members have been identified (Köksal et al., 2011). Previous work has shown that terpenoids are important in biodefense (Xiao et al., 2012; Irmisch et al., 2014; Alicandri et al., 2020; Huang et al., 2022), oxidation resistance (Xie et al., 2006), waterlogging resistance (Kuroha et al., 2018), and drought tolerance (Takahashi et al., 2018).

TPS is a critical enzyme for the production of plant terpenoids, mainly involved in the production of monoterpene, sesquiterpene and diterpene biosynthesis. Ancient TPSs originated in land plants that diverged from green algae (Jia et al., 2022). Studies have shown that TPS genes are involved in plant defense against pests and diseases and plant growth and development. *OsTPS19* enhanced the resistance to rice blast (Chen et al., 2018). TPS gene was involved in floral synthesis (Gao et al., 2018). *GhTPS10* was involved in the synthesis of gossypol (Huang et al., 2018). Based on the amino acid sequence and gene function of TPS, the TPS gene family was divided into seven clades: TPS-a, -b, -c, -d, -e/f, -g and -h (Chen et al., 2011). As more plant genomes are sequenced, TPS gene family members have been identified in a variety of plants. There were 29 TPSs in *Solanum lycopersicum* (Falara et al., 2011), 32 in *Setaria italica* (Karunanithi et al., 2020), 40 in *Arabidopsis thaliana* (Aubourg et al., 2002), 14 in *Selaginella tamariscina*, and 34 in *Oryza sativa* (Chen et al., 2011). The TPS gene family of *A. thaliana* was divided into five clades: TPS-A (22 TPSs), -b (6 TPSs), -c (1TPS), -e/f (2 TPSs), -g (1TPS) (Aubourg et al., 2002). At present, there were 41 and 46 TPSs in *Gossypium. raimondii* and *Gossypium. hirsutum*, respectively (Huang et al., 2018). Genome-wide identification and systematic analysis of the TPS gene family have not been reported in *Gossypium. barbadense* and *Gossypium. arboreum*.

Cotton is an important fiber crop. Flooding is a natural disaster frequently encountered during the seedling growth of cotton, which has a serious impact on cotton yield. At present, sequencing of *G. hirsutum* (TM-1, CRI-12), *G. barbadense* (Hai7124), *G. arboreum* and *G. raimondii* have been completed (Paterson et al., 2012; Du et al., 2018; Hu et al., 2019; Lu et al., 2019; Lu et al., 2022). It laid a good foundation for studying TPS gene in cotton. Studies suggested that *GhTPS12* may play a key role in cotton defense against herbivores (Huang et al., 2018). *GhTPS1*, *GhTPS2* and *GhTPS3* have been identified in *G. hirsutum* (Yang et al., 2013). The response of TPS to flooding in cotton has not been reported. This study was to broaden the understanding of the gene structure, phylogenetic evolution and gene function of TPSs.

Materials and methods

Download of database

Gene annotations and protein files for *G. arboreum* (Version 1.0, CRI), *G. raimondii* (Version 2.0), *G. hirsutum* (Version 2.1, ZJU), *G. barbadense* (Version 1.1, ZJU), and *A. thaliana* (TAIR10.) were obtained from online databases (https://cottonfgd.org/, http://www.arabidopsis.org/), respectively (Zhu et al., 2017).

Identification of TPS

The Hidden Markov Model profiles for PF01397 and PF03936 can be acquired from the Pfam website. TPSs were retrieved from the cotton genome database using HMMER software. Redundant genes with e value greater than 1E-05 were deleted. TPSs with incomplete C and N terminus were deleted *via* the NCBI Batch CD-Search website (https://www.ncbi.nlm.nih.gov/). The transcription length and protein length of cotton TPS gene were further retrieved from CottonFGD (https://cottonfgd.org/) (Zhu et al., 2017).

Phylogenetic analysis

The amino acid sequences of TPSs in five species are showed in the Table S3. The phylogenetic tree of TPS gene family was constructed using Neighbor-Joining (NJ) by MEGA 7.0. Bootstrap value:1000. (Larkin et al., 2007; Kumar et al., 2016).

Location map of TPS

The location data of the GhTPS gene family on chromosomes were obtained from the genome annotation file. TBtools software was used to construct the map of TPS gene on chromosomes (Chen et al., 2020).

TPS gene structure and protein motifs

Phylogenetic trees, motifs and structures were mapped by TBtools software using phylogenetic files (format: nwk), genome annotation files (format: gff3), and conserved motifs (format: MAST) in *G. hirsutum* (Bailey et al., 2009; Chen et al., 2020). The amino acid motif in the predicted GhTPS protein sequence was analyzed using online MEME website (https://meme-suite.org/meme/tools/meme). The sequence distribution site was set to 0 or 1, the ordinal number was set to 10, and all other parameters were set to default.

TPS gene collinearity

MCScanX software was used to construct synteny relationships between duplicate gene pairs (Wang et al., 2012). We used TBtools software to display collinear maps (Chen et al., 2020).

Selective pressure analysis

Duplicate gene pairs from four cotton species were identified by TBtools. The sequence identity after alignment should be higher than 80%. The non-synonymous (Ka) and synonymous (Ks) substitution ratio of duplicate genes were analyzed by TBtools software (Chen et al., 2020).

Cis-acting elements and gene expression

PlantCARE website was used to predict the cis elements of GhTPS promoters (http://bioinformatics.psb.ugent.be/). Cis-acting elements were classified and analyzed. The RNA-Seq data was downloaded from the GRAND website (http://grand.cricaas.com.cn/home) (accession number: PRJNA490626). The relative expression patterns of GhTPS gene were analyzed at different time points (0, 1, 6 and 12 h) under PEG (200 g/liter), NaCl (0.4 M), cold (4 °C) and hot (37 °C) stress

conditions (Yang et al., 2019). Methylation data download number: PRJNA856623.

qRT-PCR

The GhTPS specific expression profiles in roots, stems and leaves and the response of TPS to flooding stress were analyzed by qRT-PCR. The experimental material was watering-resistant ZNL2067, which was grown in a light incubator at 25°C until the three-leaf stage. ZNL2067 was treated with flooding (3d) and reoxygenation (3d), and tissue samples were collected (three biological replicates per treatment). Total RNA was extracted according to the EASYspin Plus Plant RNA Kit instructions, then cDNA was synthesized according to the TransStart Top Green qPCR SuperMix Instructions manual. The primer sequence of GhTPS gene was shown in Table S4. The primer sequences of GhTPSs and Actin gene were shown in Table S4. Rapid fluorescence quantitative PCR was carried out on Bio-Rad 7500. We calculated the relative expression of GhTPS gene using $2^{-\Delta\Delta Ct}$ (Livak and Schmittgen, 2001).

Results

The acquisition of TPS gene members

298 TPSs were obtained from five species. 71, 75, 60, 54 and 38 TPSs were identified from *G. hirsutum*, G. *barbadense*, *G. arboreum*, *G. raimondii* and *A. thaliana*, respectively (Figure 1). Meanwhile, based on the physical location of TPSs on chromosomes, the four species TPSs were named as *GhTPS1-GhTPS71*, *GbTPS1-GbTPS75*, *GaTPS1-GaTPS60* and *GrTPS1-GrTPS54* respectively (Table S1). The TPS gene characteristics of cotton were further analyzed, including 16 indexes such as CDS length, exon number and protein length (Table S2).

In the model plant *G. hirsutum*, there were 71 TPS proteins ranging in length from 372 aa (*GhTPS45*) to 849 aa (*GhTPS12*). CDS Length (bp) ranged from 1119bp (*GhTPS37*) to 2550bp (*GhTPS12*). The isoelectric point ranged from 4.692 (*GhTPS37*) to 6.913 (*GhTPS61*). The number of exons ranged from 5 (*GhTPS45*) to 15 (*GhTPS26*). MW ranged from 43.341 (*GhTPS45*) kDa to 97.550 (*GhTPS53*) kDa.

Phylogenetic analyses

To understand the evolutionary relationships of the TPS gene family among four cotton species, we constructed rootless phylogenetic trees of 298 TPS proteins (Saitou and Nei, 1987) (Figure 1). Based on the classification method of TPS gene proteins in *A. thaliana* (Jiang et al., 2019), 260 TPSs were divided into five evolutionary clades in Gossypium. The evolutionary clade TPS a contained the highest proportion of TPSs (129 cotton TPSs), and the distribution of the other four clades was as follows: TPS b (95 TPSs), TPS c (12 TPSs), TPS e/f (18 TPSs) and TPS g (6 TPSs) (Figure 1C; Table S5). TPSs were distributed in each clade. The ratio of diploid

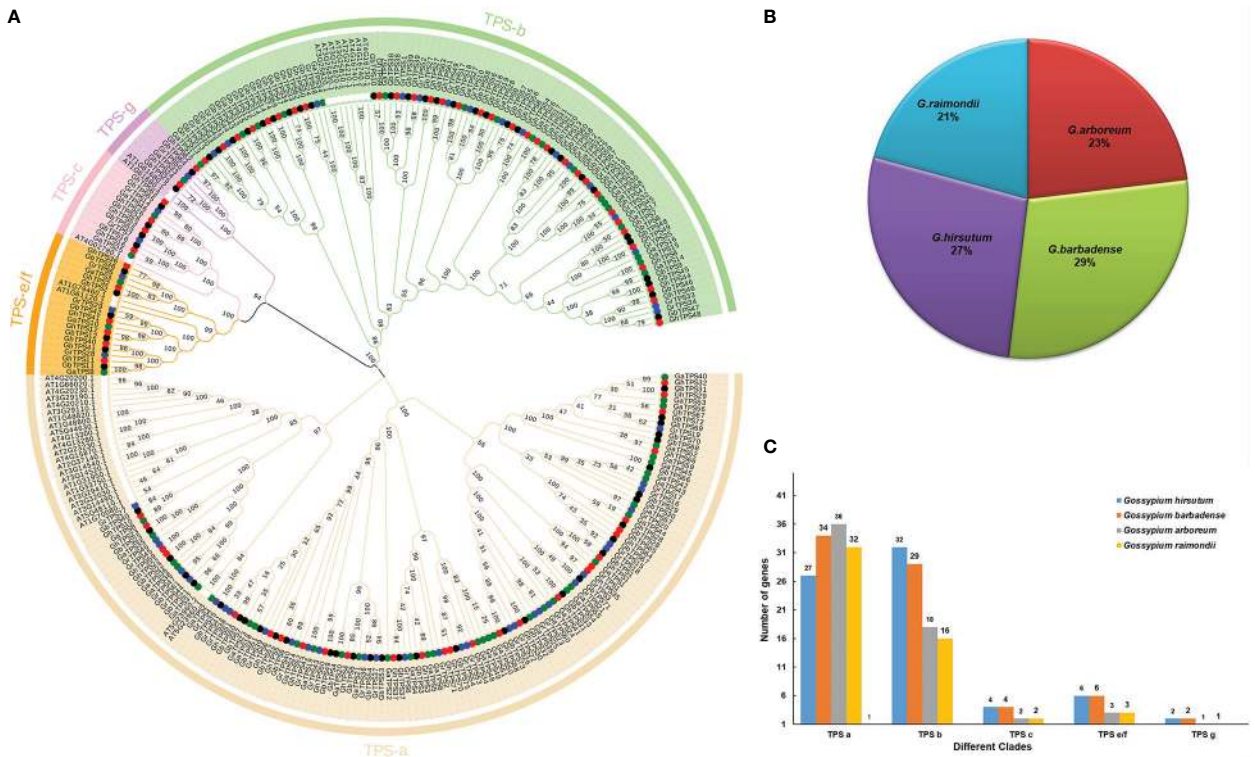

FIGURE 1
TPS gene family phylogenetic tree. **(A)** Phylogenetic relationships of 298 TPSs in five species. **(B)** The proportion of different cotton species in 260 TPS genes. **(C)** Distribution of gene number in five evolutionary clades. Different colors represent different clades.

cotton to allotetraploid cotton was less than 1:2 (Figure 1B). This suggests that the loss of the TPS gene occurred during the formation of allotetraploid.

Chromosomal location of TPS

To study the physical location of TPSs on chromosomes, we constructed chromosome maps of 260 TPSs (Figure 2). 251 genes were assigned to specific chromosomes (Figure 2; Table S6). Among the 71 GhTPSs in *G. hirsutum*, 34 and 37 TPSs were located in the At and Dt subgenome, respectively (Figure 2). For At subgenome: GHAt-11 had the most members (7 GhTPSs). For Dt subgenome: Dt-05 has 13 GhTPSs (Figure 2; Table S6). 75 TPSs were mapped to specific chromosomes in *G. barbadense* (Figure 2). The At and Dt subgenome contained 34 and 41 GbTPSs, respectively. For At:subgenome, At-11 had the most TPS members (8 GbTPSs). For Dt subgenome: Dt-05 had the highest number of TPS members (15 GbTPSs). *G. hirsutum* and *G. barbadense* belong to allotetraploid cotton, while *G. arboreum* and *G. raimondii* belong to diploid cotton. Interestingly, no TPS gene was found in chromosomes At/Dt-02, 03, 06, 07, and 12 in the allotetraploid cotton (Figure 2; Table S6). 51 GaTPSs were annotated onto 13 chromosomes, and 9 GaTPSs were not annotated on the chromosome in *G. arboreum* (Figure 2). There were 17, 8 and 7 GaTPSs on chromosomes of Chr05 (A05), Chr11 (A11) and Chr09 (A09). No GaTPSs were found on Chr03 (A03), Chr06 (A06), Chr07 (A07), and Chr12 (A12) chromosomes (Figure 2; Table S6). For *G. raimondii*, all 54 GrTPSs were annotated on chromosomes (Figure 2). There were more GrTPSs on chromosomes

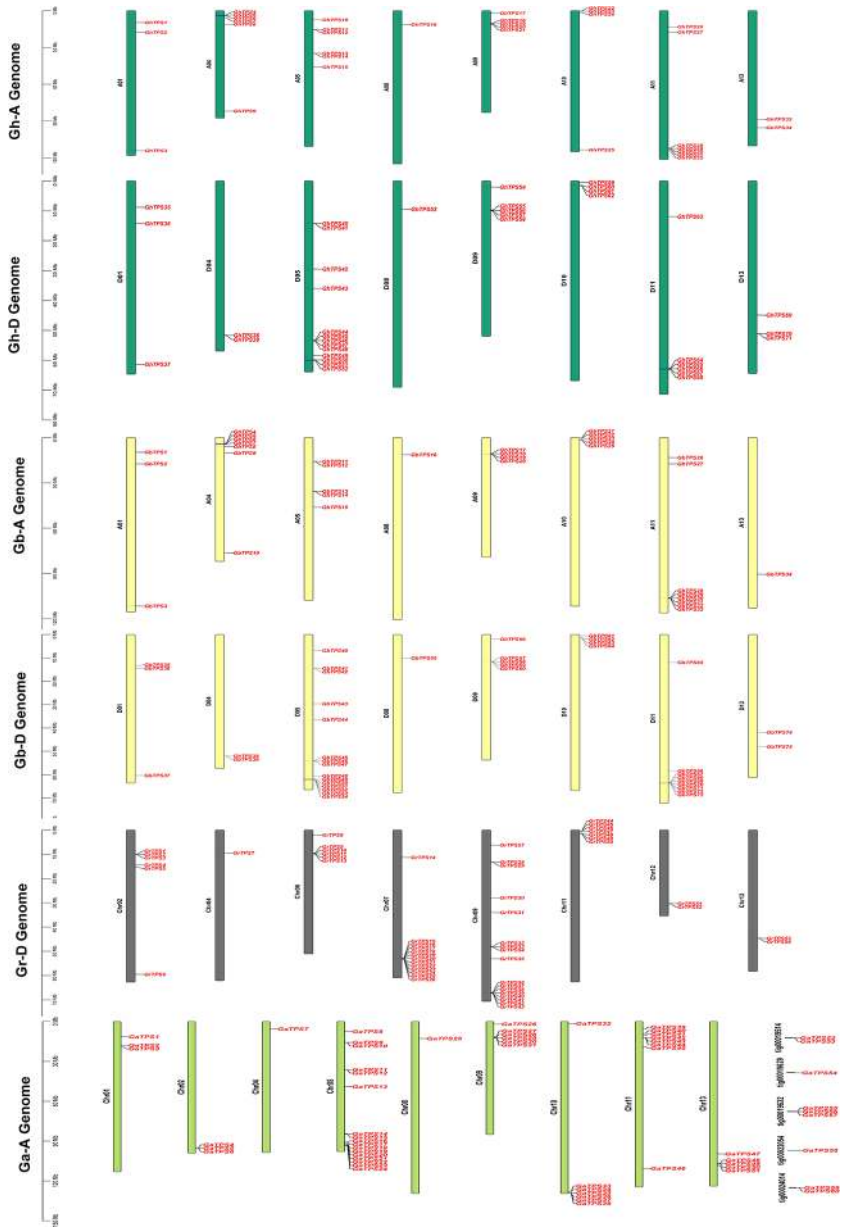

FIGURE 2
Chromosomal mapping of TPS in Gossypium. The vertical bars indicate the physical location of genes and the length of chromosomes. The gene names are on the right side of chromosomes. The four colors represent the four cotton species.

of Chr09 (D09) and Chr07 (D07), 17 and 13, respectively (Figure 2; Table S6). TPSs were unevenly distributed on chromosomes of Gossypium.

Structure and protein motifs analysis of GhTPS

In order to characterize the protein motif and gene structure during the evolution of GhTPSs in *G. hirsutum*, we constructed a phylogenetic tree, conserved motif, and gene structure relationship map of TPSs (Figure 3). For protein motifs, GhTPS proteins had conserved motifs ranging from 5 to 10. The evolutionary clade TPS a had Motif 5, 6, 7. The evolutionary clade TPS c had Motif 2, 3, 4, 6, 9. The evolutionary clade TPS e/f had Motif 1, 2, 3, 5, 6, 8, 9. The evolutionary clade TPS g had 10 conserved motifs. The variation of the conserved motif of TPS b in the evolutionary clades suggested that it may have a wider range of biological functions. Similar protein motifs were found in the same evolutionary clade.

In the GhTPS gene family, the longest gene (*GhTPS6*) was approximately 263107 bp in length. *GhTPS43* was the shortest gene, 1839 bp. The number of exons in the GhTPS gene family was 6 ~ 15. (Figure 3). Among 71 GhTPSs, 44 genes had 7 exons. GhTPSs from the same evolutionary branch had similar genetic structure. It can be seen that the GhTPS gene family formed conserved gene structure and conserved motifs during evolution.

Analysis of collinearity of the TPS gene family

In order to explore gene amplification of the GhTPS gene family, a synteny/collinear relationship map of duplication gene pairs were constructed between the diploid ancestor A & D genome and the allotetraploid AD genome (Figures 4, 5). There were 699 duplication gene pairs obtained in Gossypium including 239 segmental duplications and 51 tandem duplications. The whole genome duplication of the remaining 409 orthologous genes was performed (Figure 4). Taking *G. hirsutum* and *G. barbadense* as examples, 389 orthologous/paralogous gene pairs were obtained. There were 234 gene pairs that underwent segmental duplication. 35 and 120 gene pairs that were subjected to tandem duplication, and whole-genome duplication, respectively. It can be seen that genome multiploidy and segmental duplication are the primary modes of the TPS gene family amplification.

In order to understand the genetic amplification of A or D subgenome during evolution, we constructed collinear relationships between the GBAt-GaA-GHAt and GBDt-GrD-GHDt genomes. Chromosomes A01, A05, A11 and A13 had the most common linear genes from the A genome to the AD genome. However, D02, D07 and D09 of the D genome had a higher number of genes in common with the AtDt genome. The A genome had 58 and 76 duplicate gene pairs related to heterotetraploid AD genome, respectively. The D genome had 61 and 71 pairs of duplicated gene pairs associated with the heterotetraploid AD genome, respectively (Figures 5, S1).

Selective pressure analysis

To explore the effects of selection pressure on the evolution of TPS gene family, Ks and Ka values of orthologous/paralogous pairs of four cotton species were calculated (Figure 6; Table S7).

There were 364 (96.81%) gene pairs with Ka/Ks < 1, 258 gene pairs with Ka/Ks < 0.5, and 106 gene pairs with Ka/Ks values ranging from 0.5 ~ 0.99. This revealed that the TPS gene family is highly conserved and has been subjected to strong purifying selection during evolution. The Ka/Ks ratio of 12(3.19%) orthologous/paralogous pairs was greater than 1, suggesting that TPS family may have experienced

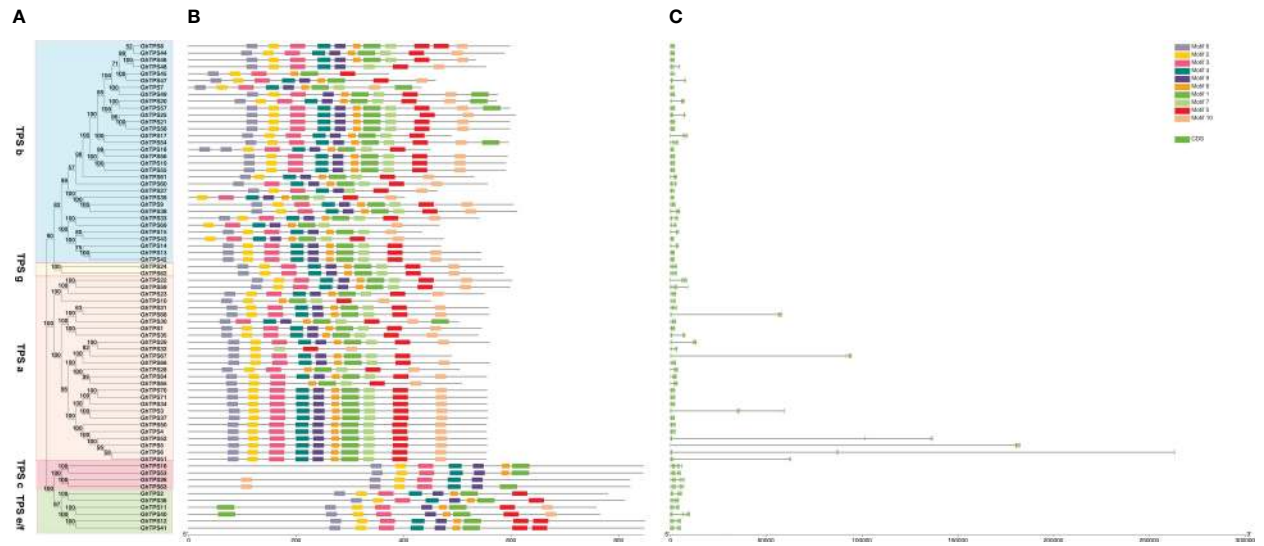

FIGURE 3
Phylogenetic tree - motif - structure of GhTPSs. **(A)** Phylogenetic tree. **(B)** Protein motifs. **(C)** Gene structures.

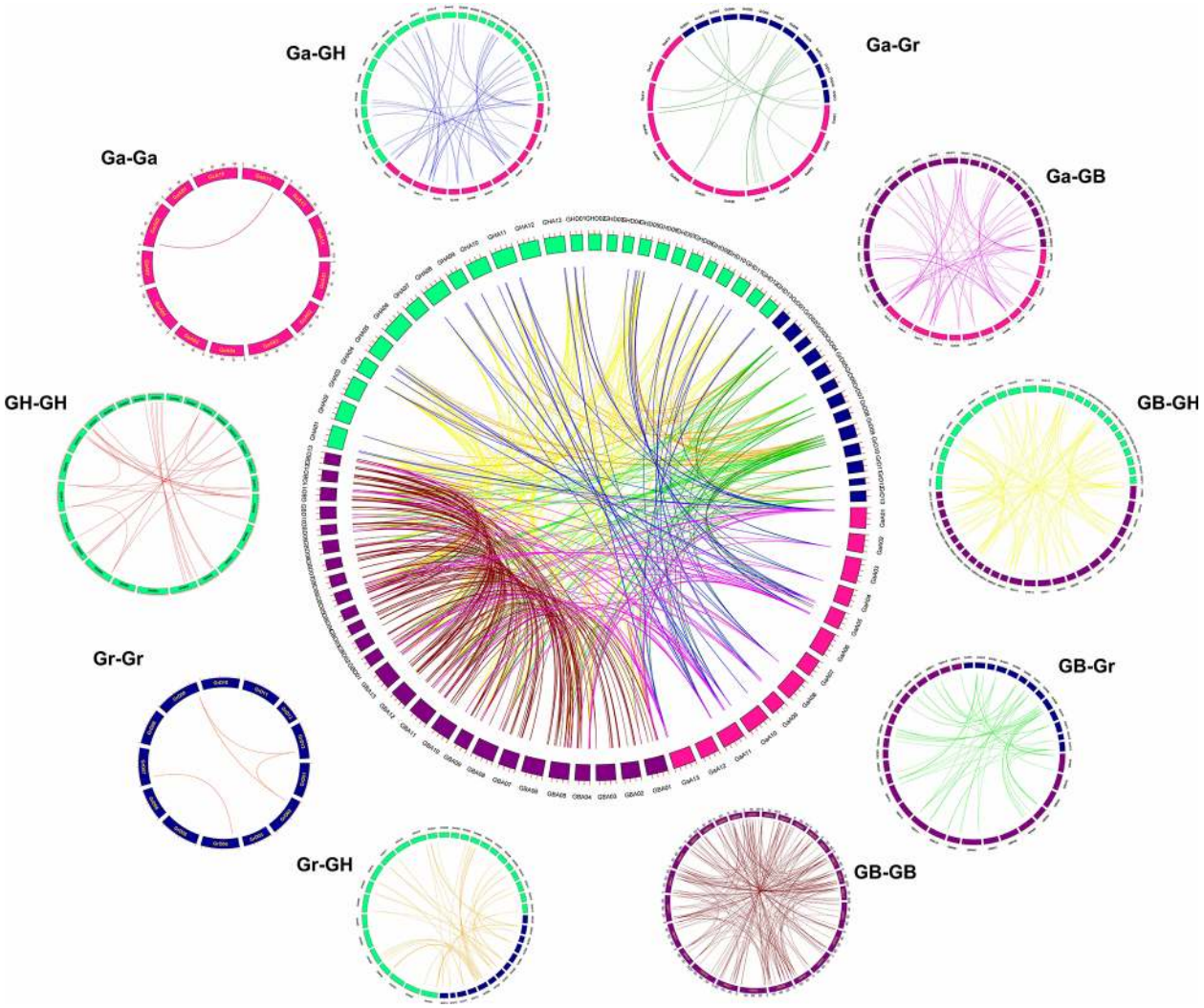

FIGURE 4
Syntenic relationship of 699 duplicated genes pairs in cotton. GHA, GHD, GBA, GBD, GaA and GrD represent At/Dt sub-genome of *G. hirsutum*, At/Dt sub-genome of *G. barbadense*, A genome of *G. arboreum* and D subgenome of *G. raimondii*, respectively.

positive selection pressure during the process of chromosome doubling. The Ka/Ks for Ga-Ga and Gr-Gr were both less than 1, which indicated that TPS gene of diploid cotton was strongly conserved. As a result, we speculated that the cotton TPS gene family is an ancient family that has experienced strong purification selection pressure during the long evolutionary process (Figures 6A, B; Table S7).

Gene enrichment analysis

We predicted the function of 260 TPSs by gene ontology (GO) analysis in cotton. GO analysis indicated that TPSs were mainly participated in molecular functions and biological processes in cotton (Figure 7; Table S8). The 260 TPSs were involved in molecular functions including: terpene synthase activity, lyase activity, magnesium ion binding. For biological processes, TPSs were fully annotated to metabolic processes (GO:0008152). Therefore, TPSs play essential role in metabolism in cotton.

Analysis of GhTPS cis-acting element

Prediction and analysis of promoter region cis-acting elements can infer the function of downstream genes. The types and number of cis-acting elements from the same evolutionary clade were different (Figure 8A). The number and type of cis-acting elements in the same evolutionary clade were different. DNA sequences upstream of transcription initiation sites (TTS) in 71 TPSs were analyzed. There were 23 kinds of cis-acting elements associated to light reaction. Box4, G-Box and GT1-motif accounted for 94%, 83% and 72% of the total GhTPSs, respectively (Figure 8B; Table S9). Eight cis-acting elements were identified in response to biological/abiotic stress, with ARE being the most abundant (Figure 8B; Table S9). Ten cis-acting elements, including ABRE, CGTCA-motif, and TGACG-motif, are associated with plant prohormone responses. They accounted for 83%, 63%, 63% and 51% of the total GhTPSs, respectively (Figure 8B; Table S9). Similarly, 10 cis-acting elements were related to growth and development, AT-rich elements were the most abundant, accounting for 30% of the total GhTPSs. It is speculated that the GhTPS gene

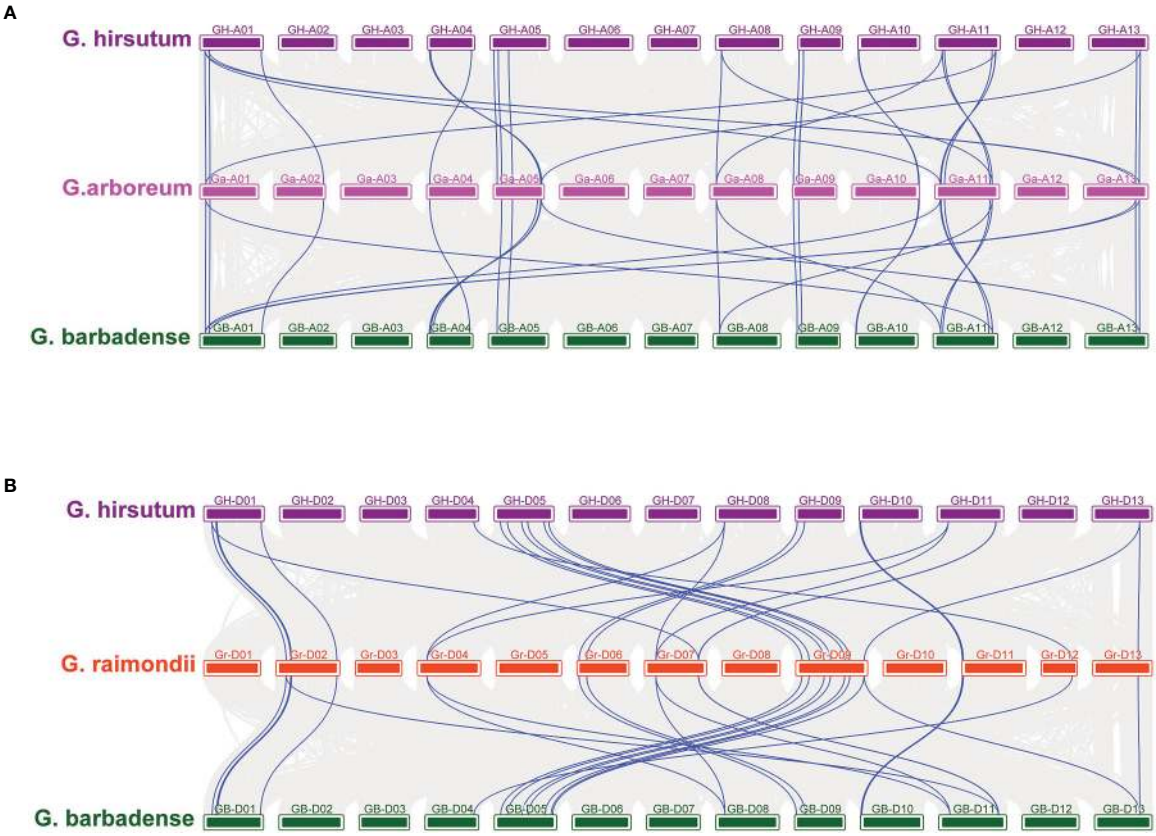

FIGURE 5
Collinearity between allotetraploid and its ancestral species **(A)** Collinearity of GhTPSs and GbTPSs compared with *G arboreum*. **(B)** Collinearity of GhGUT and GbTPSs compared with *G raimondii.* The blue line represents the TPS gene pair.

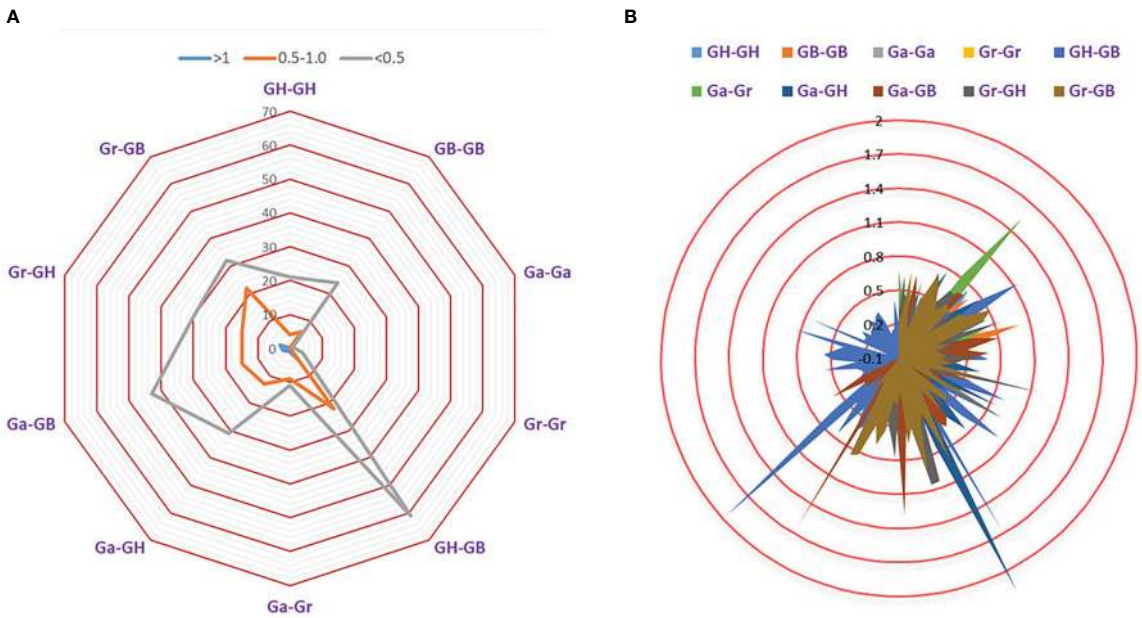

FIGURE 6
Selection pressure of TPS gene family. **(A)** The number of duplicate genes with different Ka/Ks values. **(B)** Ka/Ks divergence values of different genomes.

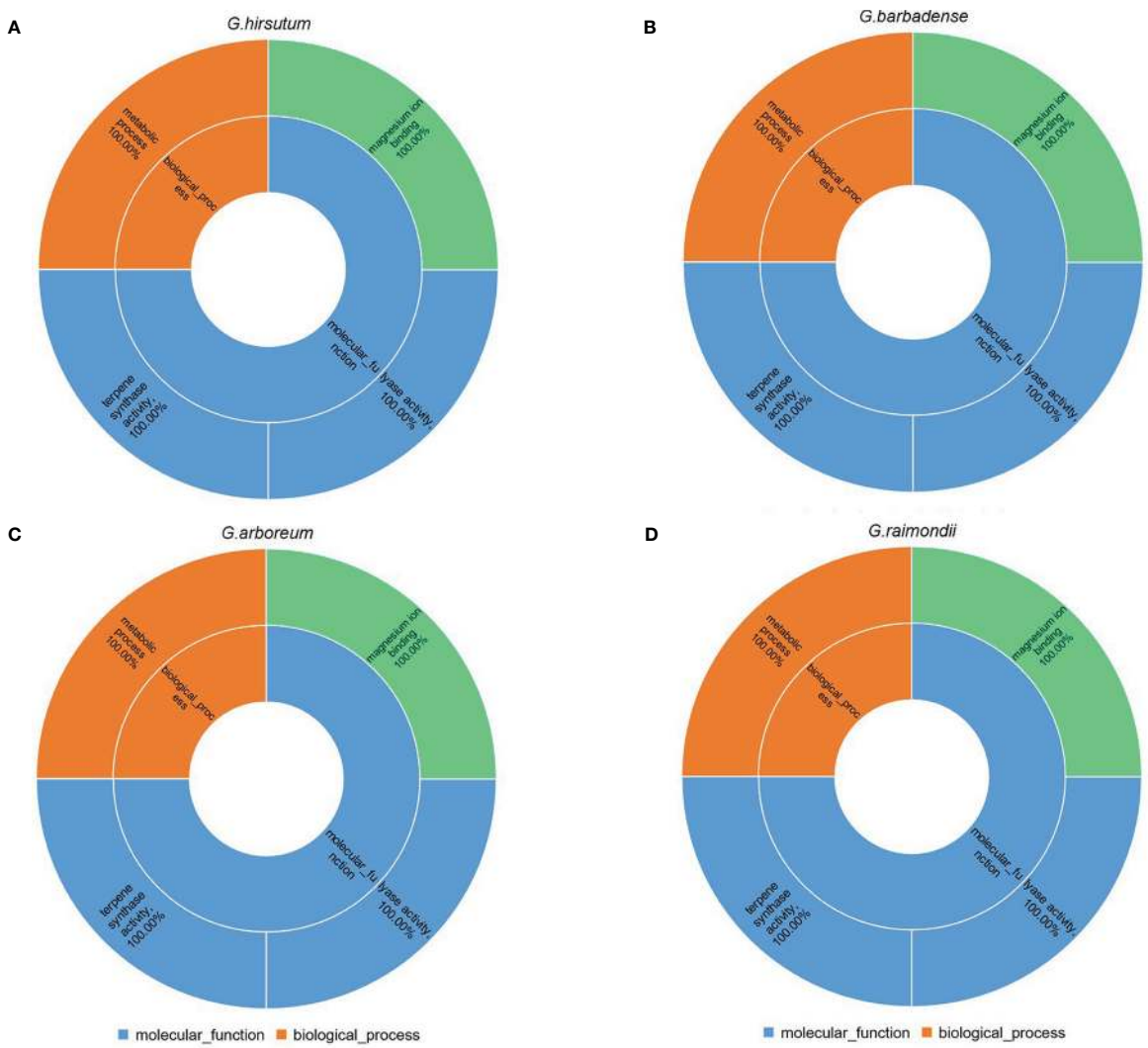

FIGURE 7
GO analysis of 260 TPSs. **(A–D)** represents GO enrichment of *G. hirsutum*, *G. barbadense*, *G.arboreum*, *G. raimondii*, respectively.

family has a critical role in plant growth process and in in biotic and abiotic stress responses.

Based on RNA-Seq data of *G. hirsutum* (TM-1, accession number: PRJNA490626), we examined the specific expression pattern of GhTPS in response to NaCl (0.4 M), PEG (200 g/liter), heat (37°C) and cold (4°C) stresses (Figure 8C; Table S10). Under different abiotic stress, *GhTPS54*, *GhTPS55*, *GhTPS70*, *GhTPS71*, *GhTPS12*, *GhTPS13*, *GhTPS40*, *GhTPS41*, *GhTPS42* and other genes showed different expressions. For example, *GhTPS55*, *GhTPS70* and *GhTPS71* were differentially expressed under cold stress. *GhTPS 42*, *GhTPS55* and *GhTPS70* were differentially expressed during heat treatment. Interestingly, *GhTPS55* gene was highly expressed in all of the above stress situations.

Tissue-specific expression profile of GhTPS

To further understand the tissue-specific expression profile of the GhTPSs and their reaction to flooding stress, we analyzed 10 GhTPSs from five clades. The expression profile of 10 GhTPSs were different in roots, stems and leaves (Figure 9). For example, *GhTPS24* was only highly expressed in leaves. *GhTPS42*, *GhTPS62* and *GhTPS63* were expressed at high levels in the stem. *GhTPS37* and *GhTPS62* were strongly expressed in the root (Figure 9).

Analysis of TPSs response to flooding stress

The expression levels of GhTPSs in response to flooding stress were different (Figure 10A). For example, *GhTPS36*, *GhTPS24* and *GhTPS62* were strongly expressed during submergence. *GhTPS36*, *GhTPS24* and *GHTPS62* were expressed at high levels under submergence and reoxygenation stress. *GhTPS37* was highly expressed under reoxygenation stress. There were also differences in tissue specific expression and response of GhTPSs in the same clade to flooding stress.

Under flooding stress, the methylation levels of CG and CHG sequences decreased in the promoter and introns region, while the methylation levels of CHH sequences increased. The methylation levels of CG, CHG, and CHH sequences in the exon domain were all reduced (Figure 10B). The hypomethylation of the exon of TPSs may help to enhance the adaptability of cotton to flooding stress.

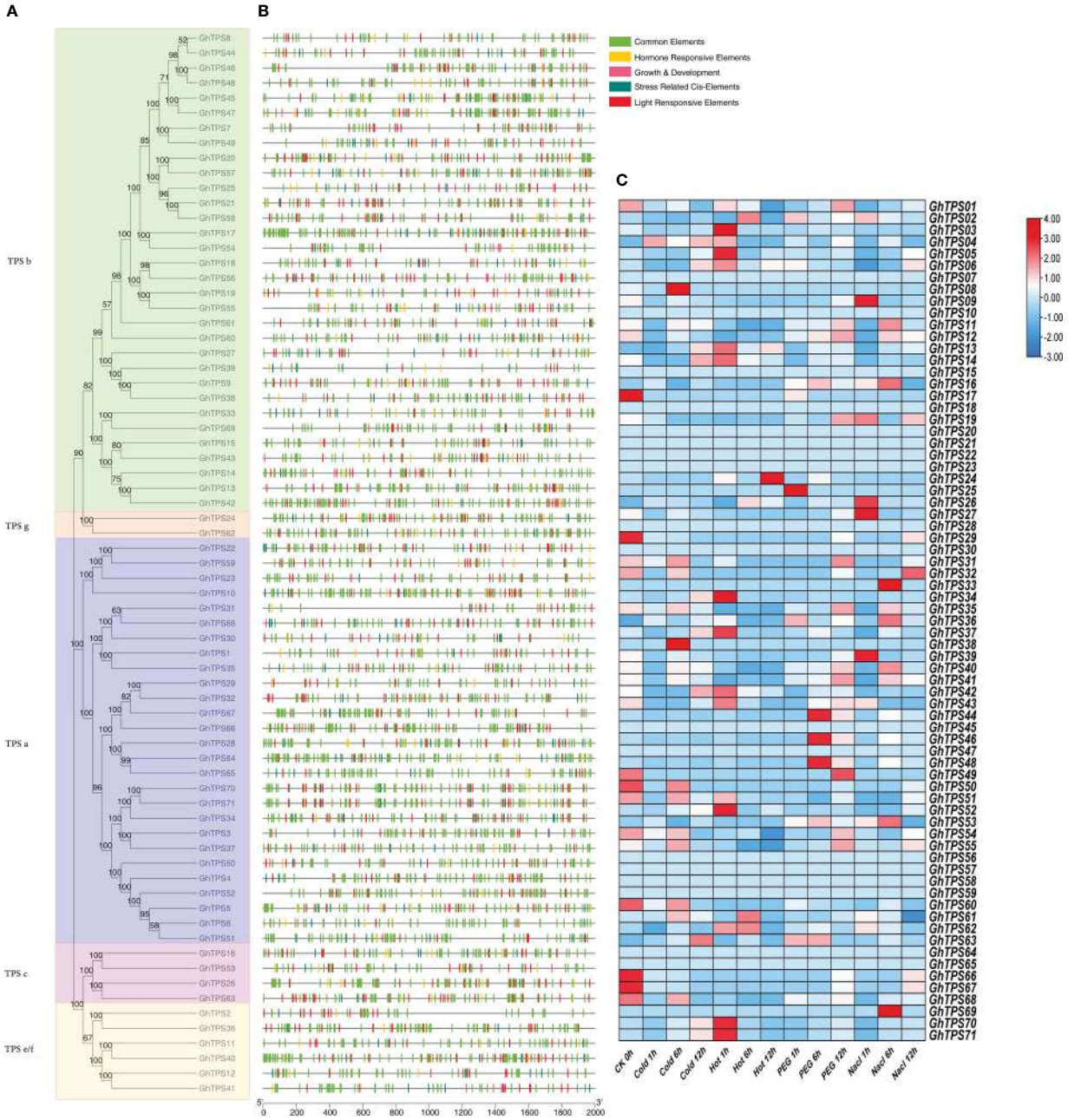

FIGURE 8
Expressed genes analysis and cis-acting elements of the GhTPS gene family. **(A)** Phylogenetic tree of GhTPSs. **(B)** Cis-elements of GhTPSs. **(C)** Expressed analysis of GhTPSs at different time points (0, 1, 6, 12h) under cold, hot, salt, PEG (FPKM).

GhTPS protein interaction

Based on the homologous gene profile of *A. thaliana*, we predict GhTPS protein function through an interactive network using the online STRING website (https://string-db.org/) (Figure 11). In the bological process, 30 GO-terms were significantly enriched, such as plastoquinone biosynthetic process (GO:0010236), monoterpene biosynthetic process (GO:0043693), terpene biosynthetic process (GO:0046246), etc. In terms of molecular function, 20 GO-terms were significantly enriched, such as trans-octaprenyltranstransferase activity (GO: 0050347), sesquiterpene synthase activity (GO:0010334), (E)-beta-ocimene synthase activity (GO: 0034768), etc. In terms of molecular composition, chloroplast (GO:0009507), cytoplasm (GO:0005737) and cellular anatomical entity (GO:0110165) were significantly enriched in GO-terms (Table S11). In the KEGG pathway, it mainly involved monoterpenoid biosynthesis (ath00902), diterpenoid biosynthesis (ath00904), sesquiterpenoid and triterpenoid steroid biosynthesis (ath00909), terpenoid backbone biosynthesis (ath00900), steroid biosynthesis (ath00100), biosynthesis of secondary metabolites (ath01110). At the same time, we analyzed the protein interaction network of *GhTPS42* gene that was significantly up-regulated under flooding treatment (Figure 11).

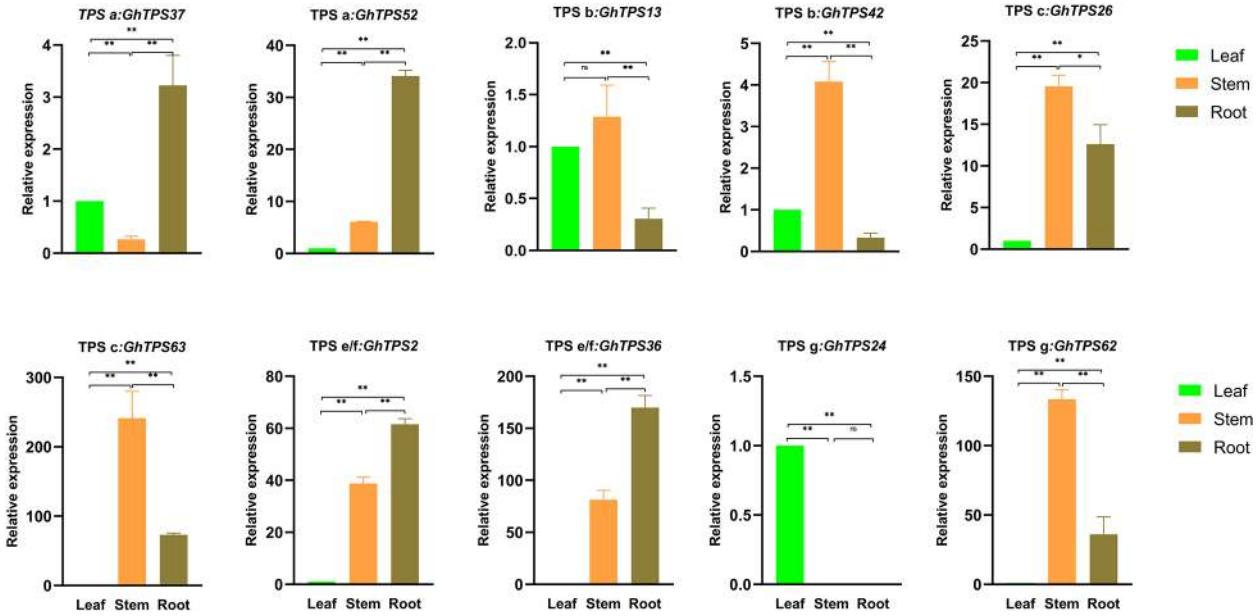

FIGURE 9
Tissue-specific expression of 10 GhTPSs. Error bars are the standard deviation (SD) of biological replicates. ns, p>0.05, *0.05>p>0.01, **p<0.01.

Discussion

Different plants contain different numbers of TPSs. 260 TPSs were identified in Gossypium (71 in *G. hirsutum*, 75 in *G. barbadense*, 60 in *G. arboreum* and 54 in *G. raimondii*. More TPSs were found in Gossypium than in *Solanum lycopersicum* (Falara et al., 2011), *Setaria italica* (Karunanithi et al., 2020), *A. thaliana* (Aubourg et al., 2002), *Selaginella tamariscina*, and 40 in rice (Chen et al., 2011). Previous studies identified 41 and 46 TPSs from *G. raimondii* and *G. hirsutum*, respectively (Huang et al., 2018), while we identified 54 and 71 TPSs, which were related to the different reference genomes we used. In recent years, with the increasing improvements in sequencing technology, the annotation of reference genome is more perfect, and the identification of gene family members is more accurate.

Structural analysis of TPS gene family

The physical location of a gene on a chromosome affects its biological function. We found that TPSs were not uniformly distributed in chromosomes and usually existed in the form of gene clusters. Genes in a gene cluster usually belong to the same evolutionary clade. A series of 51 tandem repeat gene pairs were discovered on this chromosome. The number and sequence of exons in a gene are associated with its biological function (Malik et al., 2020). For *G. hirsutum*, we found that the length of GhTPS gene varied widely, from 1839 bp to 263107 bp. The number of exons varies from 6 to 15. These indicated that the gene structure of GhTPS was diversified. At the same time, GhTPSs from the same evolutionary clade had similar gene structures and protein motifs (Figure 3).

Evolutionary analysis of the TPS gene family

To understand the changes in the TPS gene family over the long term, we analyzed the phylogenetic trees and selection pressures of TPS gene families in four cotton species. 260 TPSs were assigned to five clades, namely TPS - a, -b, -c, -e/f and -g. Previous studies have shown that TPS a is the largest clade in most dicotyledons and monocotyledons (Jiang et al., 2019). We also found that the evolutionary clade TPS a contained the largest number of TPSs (Figure 1; Table S5). TPS d is an endemic clade of gymnosperm (Bohlmann et al., 1998). The TPS h clade is identified only in the *Selaginella tamariscina* (Chen et al., 2011). TPS d and TPS h were not found in four cotton species.

Substantial gene amplification contributes to the formation of new species and adaptation to adversity (Hittinger and Carroll, 2007; Conant and Wolfe, 2008). Whole genome duplication, segmental duplication and tandem duplication are essential pathways for gene amplification. Cotton is one of the model crops studied for polyploidization (Li et al., 2015). A total of 699 duplicate gene pairs were obtained in Gossypium, including 409 whole genome duplication genes, 239 segmental duplication genes and 51 tandem duplication genes (Figure 4). Therefore, the three gene amplification modes played an important role in the amplification of the TPS family in Gossypium.

Previous studies have shown that heterotetraploid cotton is produced by interspecific hybridization of A and D genomes (Wendel and Cronn, 2003; Paterson et al., 2012; Li et al., 2015). The number of TPSs from allotetraploid cotton was less than twice that of diploid cotton, which might be due to gene deletions during the evolutionary process of forming allotetraploid cotton. Gene loss is also present in the evolution of the GRX, AHL and UGT gene families (Malik et al., 2020; Zhao et al., 2020; Sun et al., 2022).

When Ka/Ks>1, the TPS family was subjected to positive selection in the long-term evolution process. When Ka/Ks=1, the TPS family was

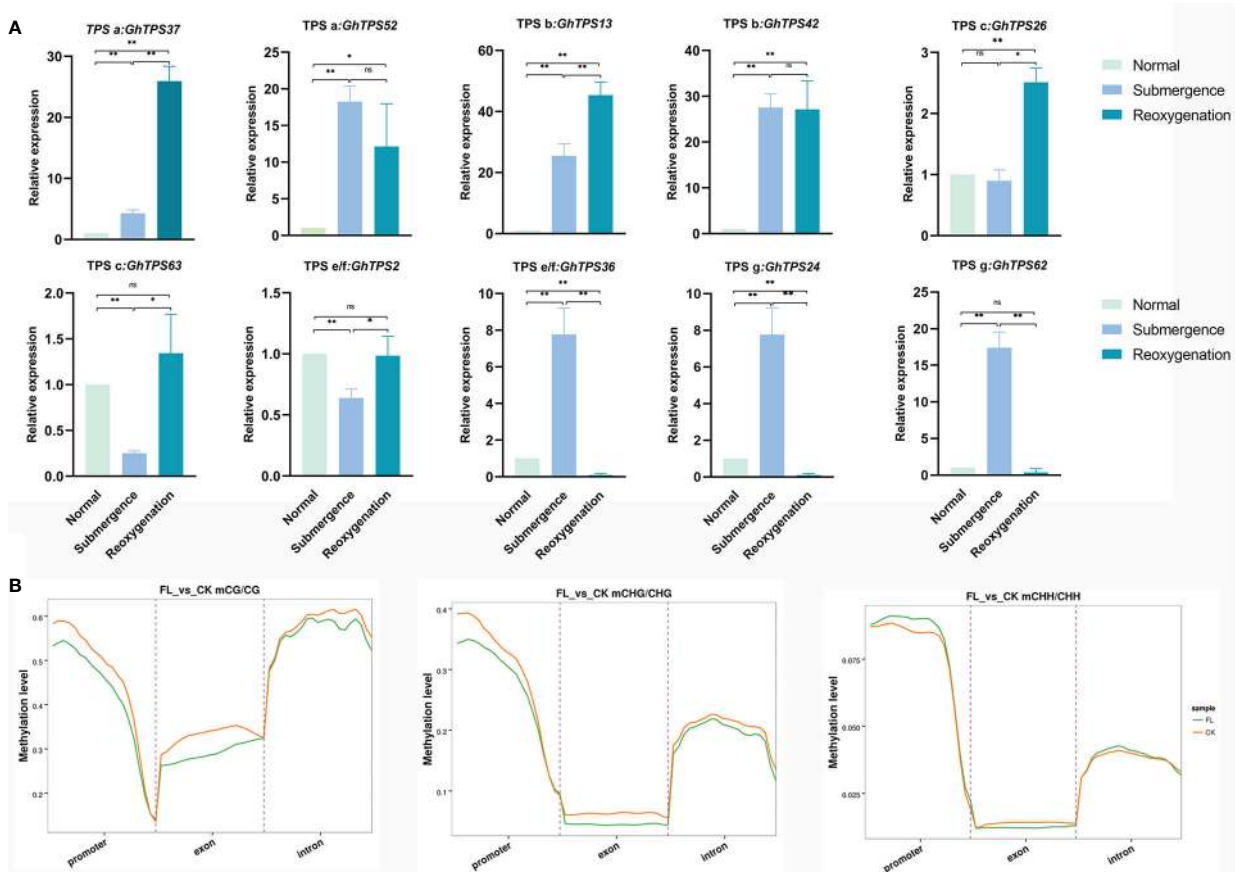

FIGURE 10
Levels of methylation in the CG/CHG/CHH sequences and expression levels of GhTPSs. **(A)** The expression levels of 10 GhTPSs under submergence and reoxygenation stresses. Error bars represent SD in biological replicates. **(B)** Distribution of methylation levels in functional region. Different colors represent groups. ns, p>0.05, *0.05>p>0.01, **p<0.01.

subject to neutral evolution. When Ka/Ks<1, the TPS family was subjected to purify selection in the long-term evolution process. Our results showed that 96.81% of TPS gene pairs (364) had Ka/Ks values less than 1, which indicates that the TPS gene family experienced highly purified selection pressure over the long evolutionary period (Figures 6A, B; Table S7). Meanwhile, 12 (3.19%) gene pairs had Ka/Ks values greater than 1, which demonstrates that TPSs underwent positive selection pressure after gene replication.

Functional analysis of the TPS gene family

The results showed that 10 cis-acting elements were involved in cotton growth and development, and AT-rich element was the most involved. 10 cis-acting elements were related to regulate plant hormone; AT-rich elements were the most. There were also differences in the cis-acting elements of genes within the same evolutionary clade. At the same time, we found that GhTPS gene has tissue specific expression characteristics. For example, *GhTPS24* was expressed at high levels in leaves, while *GhTPS42*, *GhTPS62* and *GhTPS63* were expressed at high levels in stems (Figure 9). In general, TPSs are essential in the cotton growth and development.

Eight cis-acting elements responded to biological/abiotic stress of cotton, among which ARE and LTR were more (Table S9). Under different stress treatments, not all genes had biological functions. For example, *GhTPS55* gene was highly expressed in different stress situations, while many TPSs were not involved in stress response. Previous studies have also verified this conclusion. In some angiosperms and gymnosperms, not all TPSs are functional (Chen et al., 2011). Through comprehensive analysis of expression profile, cis-regulatory elements and protein interaction, TPS gene may play an important role in waterlogging stress.

Conclusion

TPS is a critical enzyme for the production of plant terpenoids, mainly involved in the production of monoterpene, sesquiterpene and diterpene biosynthesis. A total of 260 TPSs were identified, including 71 in *G. hirsutum*, 75 in *G. barbadense*, 54 in *G. raimondii* and 60 in *G. arboreum*. We systematically analyzed the TPS gene family of Gossypium from three aspects: gene structure, evolutionary process and gene function (Figure 12). (1) Gene structure analysis: Based on the protein structure of two conserved domains (PF01397 and PF03936), the TPS

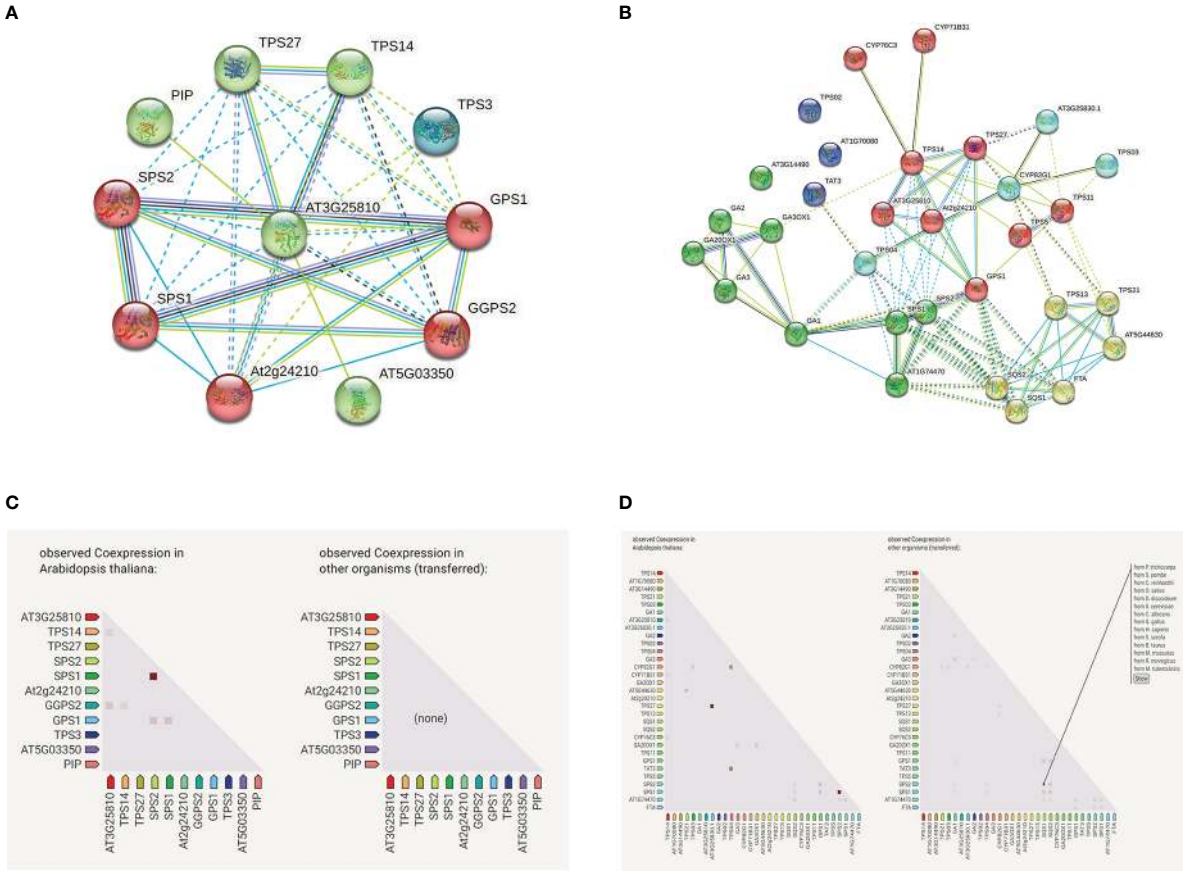

FIGURE 11
GhTPS proteins interaction network. **(A)** *GhTPS*42 protein interaction network. **(B)** Interaction network of GhTPS proteins. **(C)** *GhTPS42* gene co-expressed. **(D)** Gene co-expression of the GhTPS gene family. Note: The homologous gene of *GhTPS42* in *A thaliana* is AT3G25810.

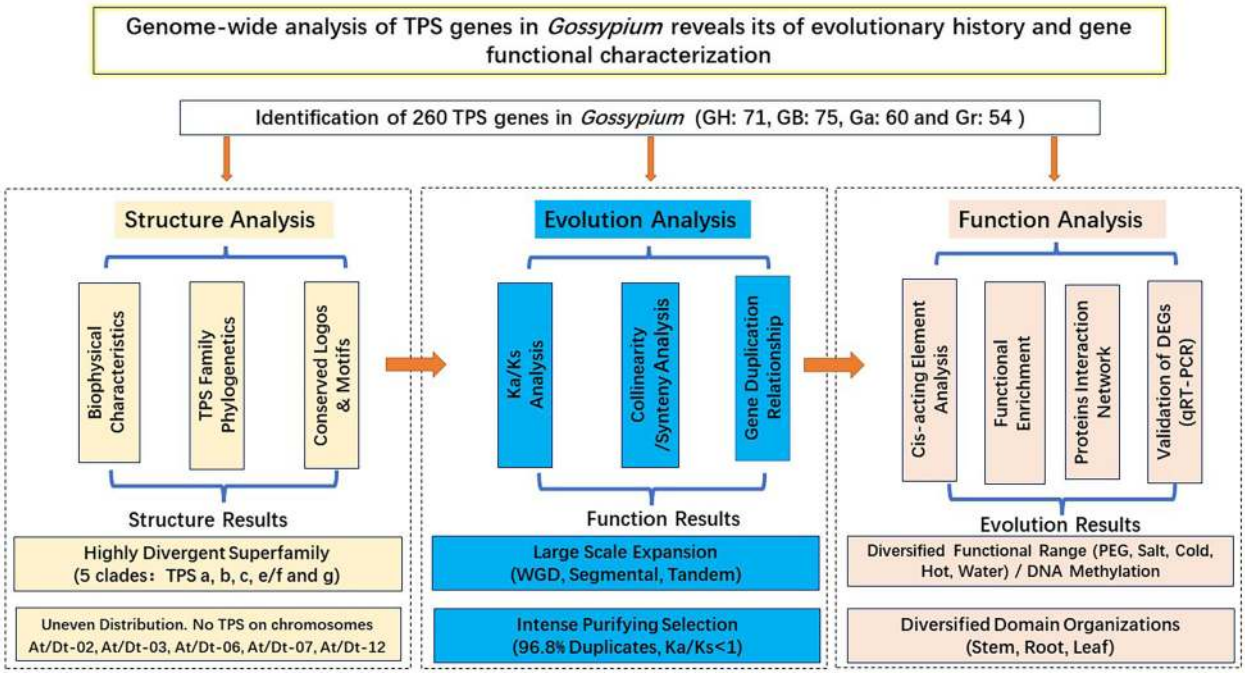

FIGURE 12
Genome-wide and systematic analysis of the TPS gene family in Gossypium.

gene family was divided into five clades: TPS -a, -b, -c, -e/f and -g. (2) Evolution analysis: Genome multiploidy and segmental duplication are the main ways of TPS gene amplification. The TPS gene family underwent strong purification selection pressure during the long evolutionary process. (3) Function analysis: The abundance of cis-acting elements may reveal the functional diversity of TPSs in cotton. Cotton TPS gene is tissue-specific and plays an important role in stress. TPS gene has tissue specific expression in cotton and plays a key role in stress. In conclusion, this study can broaden the understanding of structure-evolution-function of the TPS gene family, and provide reference for the mining and verification of new genes.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/Supplementary Material.

Author contributions

Conceived and designed the experiments: LS and TN; methodology: AC and YJ; experiment: AC, YJ, and LS; analysis of data: LS and YL; writing-original draft preparation: AC and YJ; writing-review and editing: LS; supervision: TN and LS. All authors contributed to the article and approved the submitted version.

Funding

This study was supported by Jiangxi Agriculture Research System (JXARS-22) and Key R&D Project of Jiangxi Province (20192BBF60002).

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.2023.1126884/full#supplementary-material

References

Alicandri, E., Paolacci, A. R., Osadolor, S., Sorgonà, A., Badiani, M., and Ciaffi, M. (2020). On the evolution and functional diversity of terpene synthases in the pinus species: A review. *J. Mol. Evol.* 88, 253–283. doi: 10.1007/s00239-020-09930-8

Aubourg, S., Lecharny, A., and Bohlmann, J. (2002). Genomic analysis of the terpenoid synthase (AtTPS) gene family of *Arabidopsis thaliana*. *Mol. Genet. Genomics* 267, 730–745. doi: 10.1007/s00438-002-0709-y

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

Bohlmann, J., Meyer-Gauen, G., and Croteau, R. (1998). Plant terpenoid synthases: molecular biology and phylogenetic analysis. *Proc. Natl. Acad. Sci. U.S.A.* 95, 4126–4133. doi: 10.1073/pnas.95.8.4126

Cane, D. E. (2000). Perspectives: biosynthetic pathways. biosynthesis meets bioinformatics. *Science* 287, 818–819. doi: 10.1126/science.287.5454.818

Chen, X., Chen, H., Yuan, J. S., Köllner, T. G., Chen, Y., Guo, Y., et al. (2018). The rice terpene synthase gene *OsTPS19* functions as an (S)-limonene synthase in planta, and its overexpression leads to enhanced resistance to the blast fungus magnaporthe oryzae. *Plant Biotechnol. J.* 16, 1778–1787. doi: 10.1111/pbi.12914

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, F., Tholl, D., Bohlmann, J., and Pichersky, E. (2011). The family of terpene synthases in plants: a mid-size family of genes for specialized metabolism that is highly diversified throughout the kingdom. *Plant J.* 66, 212–229. doi: 10.1111/j.1365-313X.2011.04520.x

Conant, G. C., and Wolfe, K. H. (2008). Turning a hobby into a job: how duplicated genes find new functions. *Nat. Rev. Genet.* 9, 938–950. doi: 10.1038/nrg2482

Du, X., Huang, G., He, S., Yang, Z., Sun, G., Ma, X., et al. (2018). Resequencing of 243 diploid cotton accessions based on an updated a genome identifies the genetic basis of key agronomic traits. *Nat. Genet.* 50, 796–802. doi: 10.1038/s41588-018-0116-x

Falara, V., Akhtar, T. A., Nguyen, T. T., Spyropoulou, E. A., Bleeker, P. M., Schauvinhold, I., et al. (2011). The tomato terpene synthase gene family. *Plant Physiol.* 157, 770–789. doi: 10.1104/pp.111.179648

Gao, F., Liu, B., Li, M., Gao, X., Fang, Q., Liu, C., et al. (2018). Identification and characterization of terpene synthase genes accounting for volatile terpene emissions in flowers of freesia x hybrida. *J. Exp. Bot.* 69, 4249–4265. doi: 10.1093/jxb/ery224

Hittinger, C. T., and Carroll, S. B. (2007). Gene duplication and the adaptive evolution of a classic genetic switch. *Nature* 449, 677–681. doi: 10.1038/nature06151

Hu, Y., Chen, J., Fang, L., Zhang, Z., Ma, W., Niu, Y., et al. (2019). *Gossypium barbadense* and *Gossypium hirsutum* genomes provide insights into the origin and evolution of allotetraploid cotton. *Nat. Genet.* 51, 739–748. doi: 10.1038/s41588-019-0371-5

Huang, L. M., Huang, H., Chuang, Y. C., Chen, W. H., Wang, C. N., and Chen, H. H. (2021). Evolution of terpene synthases in orchidaceae. *Int. J. Mol. Sci.* 22, 6947. doi: 10.3390/ijms22136947

Huang, X. Z., Xiao, Y. T., Köllner, T. G., Jing, W. X., Kou, J. F., Chen, J. Y., et al. (2018). The terpene synthase gene family in *Gossypium hirsutum* harbors a linalool synthase *GhTPS12* implicated in direct defence responses against herbivores. *Plant Cell Environ.* 41, 261–274. doi: 10.1111/pce.13088

Huang, X., Zhang, H., Li, H., Wang, M., Guo, X., Liu, E., et al. (2022). Functional characterization of a terpene synthase responsible for (*E*)-β-ocimene biosynthesis identified in *Pyrus betuleafolia* transcriptome after herbivory. *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.1077229

Irmisch, S., Jiang, Y., Chen, F., Gershenzon, J., and Köllner, T. G. (2014). Terpene synthases and their contribution to herbivore-induced volatile emission in western balsam poplar (Populus trichocarpa). *BMC Plant Biol.* 14, 270. doi: 10.1186/s12870-014-0270-y

Jia, Q., Brown, R., Köllner, T. G., Fu, J., Chen, X., Wong, G. K., et al. (2022). Origin and early evolution of the plant terpene synthase family. *Proc. Natl. Acad. Sci. U.S.A.* 119, e2100361119. doi: 10.1073/pnas.2100361119

Jiang, S. Y., Jin, J., Sarojam, R., and Ramachandran, S. (2019). A comprehensive survey on the terpene synthase gene family provides new insight into its evolutionary patterns. *Genome Biol. Evol.* 11, 2078–2098. doi: 10.1093/gbe/evz142

Karunanithi, P. S., Berrios, D. I., Wang, S., Davis, J., Shen, T., Fiehn, O., et al. (2020). The foxtail millet (*Setaria italica*) terpene synthase gene family. *Plant J.* 103, 781–800. doi: 10.1111/tpj.14771

Köksal, M., Jin, Y., Coates, R. M., Croteau, R., and Christianson, D. W. (2011). Taxadiene synthase structure and evolution of modular architecture in terpene biosynthesis. *Nature* 469, 116–120. doi: 10.1038/nature09628

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

Kuroha, T., Nagai, K., Gamuyao, R., Wang, D. R., Furuta, T., Nakamori, M., et al. (2018). Ethylene-gibberellin signaling underlies adaptation of rice to periodic flooding. *Science* 361, 181–186. doi: 10.1126/science.aat1577

Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., Mcgettigan, P. A., Mcwilliam, H., et al. (2007). Clustal W and clustal X version 2.0. *bioinformatics* 23, 2947–2948. doi: 10.1093/bioinformatics/btm404

Li, F., Fan, G., Lu, C., Xiao, G., Zou, C., Kohel, R. J., et al. (2015). Genome sequence of cultivated upland cotton (*Gossypium hirsutum* TM-1) provides insights into genome evolution. *Nat. Biotechnol.* 33, 524–530. doi: 10.1038/nbt.3208

Livak, K. J., and Schmittgen, T. D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2(-delta delta C(T)) method. *Methods* 25, 402–408. doi: 10.1006/meth.2001.1262

Lu, X., Chen, X., Wang, D., Yin, Z., Wang, J., Fu, X., et al. (2022). A high-quality assembled genome and its comparative analysis decode the adaptive molecular mechanism of the number one Chinese cotton variety CRI-12. *Gigasci.* 11, 1–14. doi: 10.1093/gigascience/giac019

Lu, X., Fu, X., Wang, D., Wang, J., Chen, X., Hao, M., et al. (2019). Resequencing of cv CRI-12 family reveals haplotype block inheritance and recombination of agronomically important genes in artificial selection. *Plant Biotechnol. J.* 17, 945–955. doi: 10.1111/pbi.13030

Malik, W. A., Wang, X., Wang, X., Shu, N., Cui, R., Chen, X., et al. (2020). Genome-wide expression analysis suggests glutaredoxin genes response to various stresses in cotton. *Int. J. Biol. Macromolecules* 153, 470–491. doi: 10.1016/j.ijbiomac.2020.03.021

Paterson, A. H., Wendel, J. F., Gundlach, H., Guo, H., Jenkins, J., Jin, D., et al. (2012). Repeated polyploidization of gossypium genomes and the evolution of spinnable cotton fibres. *Nature* 492, 423–427. doi: 10.1038/nature11798

Saitou, N., and Nei, M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* 4, 406–425. doi: 10.1093/oxfordjournals.molbev.a040454

Sun, L., Zhao, L., Huang, H., Zhang, Y., Wang, J., Lu, X., et al. (2022). Genome-wide identification, evolution and function analysis of UGTs superfamily in cotton. *Front. Mol. Biosci.* 9. doi: 10.3389/fmolb.2022.965403

Takahashi, F., Suzuki, T., Osakabe, Y., Betsuyaku, S., Kondo, Y., Dohmae, N., et al. (2018). A small peptide modulates stomatal control *via* abscisic acid in long-distance signalling. *Nature* 556, 235–238. doi: 10.1038/s41586-018-0009-2

Wang, Y., Tang, H., Debarry, J. D., Tan, X., Li, J., Wang, X., et al. (2012). MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Res.* 40, e49–e49. doi: 10.1093/nar/gkr1293

Wendel, J. F., and Cronn, R. C. (2003). Polyploidy and the evolutionary history of cotton. *Adv. Agron.* 78, 78004–78008. doi: 10.1016/S0065-2113(02)78004-8

Xiao, Y., Wang, Q., Erb, M., Turlings, T. C., Ge, L., Hu, L., et al. (2012). Specific herbivore-induced volatiles defend plants and determine insect community composition in the field. *Ecol. Lett.* 15, 1130–1139. doi: 10.1111/j.1461-0248.2012.01835.x

Xie, J.-T., Shao, Z.-H., Hoek, T. L. V., Chang, W.-T., Li, J., Mehendale, S., et al. (2006). Antioxidant effects of ginsenoside re in cardiomyocytes. *Eur. J. Pharmacol.* 532, 201–207. doi: 10.1016/j.ejphar.2006.01.001

Yang, Z., Ge, X., Yang, Z., Qin, W., Sun, G., Wang, Z., et al. (2019). Extensive intraspecific gene order and gene structural variations in upland cotton cultivars. *Nat. Commun.* 10, 2989. doi: 10.1038/s41467-019-10820-x

Yang, C. Q., Wu, X. M., Ruan, J. X., Hu, W. L., Mao, Y. B., Chen, X. Y., et al. (2013). Isolation and characterization of terpene synthases in cotton (Gossypium hirsutum). *Phytochemistry* 96, 46–56. doi: 10.1016/j.phytochem.2013.09.009

Yazaki, K., Arimura, G. I., and Ohnishi, T. (2017). 'Hidden' terpenoids in plants: Their biosynthesis, localization and ecological roles. *Plant Cell Physiol.* 58, 1615–1621. doi: 10.1093/pcp/pcx123

Zhao, L., Lü, Y., Chen, W., Yao, J., Li, Y., Li, Q., et al. (2020). Genome-wide identification and analyses of the AHL gene family in cotton (Gossypium). *BMC Genomics* 21, 69. doi: 10.1186/s12864-019-6406-6

Zhu, T., Liang, C., Meng, Z., Sun, G., Meng, Z., Guo, S., et al. (2017). CottonFGD: an integrated functional genomics database for cotton. *BMC Plant Biol.* 17, 101. doi: 10.1186/s12870-017-1039-x