Genome-wide identification and characterization of multiple C2 domains and transmembrane region proteins in Gossypium hirsutum

Pengbo Hao
Chinese Academy of Agricultural Sciences Cotton Research Institute
Hantao Wang
Chinese Academy of Agricultural Sciences Cotton Research Institute
Liang Ma
Chinese Academy of Agricultural Sciences Cotton Research Institute
Aimin Wu
Chinese Academy of Agricultural Sciences Cotton Research Institute
Pengyun Chen
Chinese Academy of Agricultural Sciences Cotton Research Institute
Shuaishuai Cheng
Chinese Academy of Agricultural Sciences Cotton Research Institute
Hengling Wei
Chinese Academy of Agricultural Sciences Cotton Research Institute
Shuxun Yu (ysx195311@163.com)
Chinese Academy of Agricultural Sciences Cotton Research Institute https://orcid.org/0000-0002-9715-3462

Research article

Keywords: G. hirsutum, MCTPs, N-terminus, C-terminus, domain architecture, expression patterns

Posted Date: October 7th, 2019

DOI: https://doi.org/10.21203/rs.2.15643/v1

Version of Record: A version of this preprint was published on June 29th, 2020. See the published version at https://doi.org/10.1186/s12864-020-06842-1.

Abstract

Background Multiple C2 domains and transmembrane region proteins (MCTPs) may act as transport mediators of other regulators. Although increased number of MCTPs in higher plants implies their diverse and specific functions in plant growth and development, only a few plant MCTPs have been studied and no study on the MCTPs in cotton has been reported.

Results In this study, we identified 31 MCTPs in G. hirsutum , which were classified into five subfamilies according to the phylogenetic analysis. GhMCTPs from subfamily V exhibited pIs less than 7, whereas GhMCTPs from subfamily I, II, III and IV exhibited pIs more than 7.5, implying their distinct biological functions. In addition, GhMCTPs within subfamily III, IV and V exhibited more diverse physicochemical properties, domain architectures and expression patterns than GhMCTPs within subfamily I and II, suggesting that GhMCTPs within subfamily III, IV and V diverged to perform more diverse and specific functions. Analyses of conserved motifs and pIs indicated that the N-terminus was more divergent than the C-terminus and GhMCTPs' functional divergence might be mainly contributed by the N-terminus. Furthermore, yeast two-hybrid assay indicated that the N-terminus was responsible to interact with target proteins. Phylogenetic analysis classified multiple N-terminal C2 domains into four subclades, suggesting that these C2 domains performed different molecular functions in mediating the transport of target proteins.

Conclusions Our systematic characterization of MCTPs in G. hirsutum will provide helpful information to further research GhMCTPs' molecular roles in mediating other regulators' transport to coordinate growth and development of various cotton tissues.

Background

Intercellular transport of proteins, signaling molecules and carbohydrate is a key process that coordinates the activities of neighboring cells to modulate multicellular organisms' growth and development [1]. Unlike animal cells, neighboring plant cells are separated by a pair of polysaccharide cell walls [2], which are permeable to small soluble proteins and other solutes, limiting direct contact between adjacent cells [3]. However, Plant have developed plasmodesma (PD) to transport proteins, small RNAs, hormones, and metabolites [4]. One significant feature of the PD is a strand of endoplasmic reticulum (ER) that traverses the pore and is tethered tightly to the plasma membrane (PM) by unidentified spokes [5]. Recent study has demonstrated that multiple C2 domains and transmembrane region proteins (MCTPs) are core PD proteins involved in tethering ER and PM [6].

MCTPs are characterized by three to four C2 domains at the N terminus and one to four transmembrane regions at the C terminus[7]. The C2 domains have been under the enthusiastic research [8–13], because they are the second most ubiquitous lipid binding domain behind the Pleckstrin Homolgy domain (PH domain) and act as the main sensor of diverse Ca^{2+}-mediated cellular processes [14]. The C2 domains were classified into 7 subfamilies [15] and were contained in a large number of proteins that performed

distinct physiological functions [16–19]. *MCTP* was first identified in *C. elegans* and function loss of *MCTP* disrupted embryo development [20]. *Drosophila MCTP* was involved in maintaining baseline neurotransmitter release and presynaptic homeostatic plasticity [21]. In mammals, genetic mutations in *MCTPs* might affect the performance of brain and spiral cord, which could lead to bipolar disorder [22, 23]. However, the molecular functions of MCTPs in regulating these processes were still largely unknown, especially the functions of different C2 domains and transmembrane regions contained in MCTPs.

In the plant kingdom, *QKY* and *FTIP1* were the first two reported *MCTPs* in *Arabidopsis* [24, 25]. PD-localized QKY interacted with the receptor-like kinase STRUBBELIG (SUB) to promote cell-to-cell communication and organogenesis [26], while *qky* mutants exhibited twisted gynoecium due to defective cell growth anisotropy and division pattern [27]. ER-localized FTIP1 were the essential intercellular transporter of florigen protein FLOWERING LOCUS T (FT) from companion cells to sieve elements, thereby facilitating FT's movement from leaves to shoot apical meristem (SAM) and inducing flowering [25]. Thereafter, a genome-wide analysis identified 16 *AtMCTPs* including *QKY* and *FTIP1.* These AtMCTPs showed diverse expression patterns and subcellular localization, implying MCTPs' diverse functions in plant development. The authors also demonstrated that three C2 domains contained in FTIP1 might mediate FT's movement cooperatively [7]. FTIP3/4 facilitated a key meristem regulator, SHOOTMERISTEMLESS (STM), to recycle to the nucleus to ensure normal maintenance and differentiation of SAM [28]. In orchid, DOFTIP1 interacted with DOFT and promoted flowering [29]. In rice, OsFTIP1 regulated rice flowering time under long days by mediating RFT1's movement to SAM [30]. Another MCTP of rice, OsFTIP7 contributed to the anther dehiscence through repressing auxin biosynthesis [31]. In maize, ZmCpd33, a homolog of Arabidopsis QKY, promoted symplastic sucrose export from companion cells into sieve elements. The *cpd33* mutants exhibited fewer PD at the companion cell-sieve element interface and excessive carbohydrate accumulation in the leaves [32]. These studies suggest that MCTPs are involved in diverse cellular processes mainly through intercellular or intracellular transport of other regulators.

Upland cotton (*Gossypium hirsutum)* is the most widely cultivated fiber crop for its high productivity and moderate quality of natural textile fiber [33, 34]. As an annual plant with the indeterminate growth habit, upland cotton flowers continuously and periodically from the first flowering to the harvest and subsequently sets spaced bolls on different fruit branches [35]. Both fiber yield and quality are strongly affected by the transport of energy materials and signaling factors among different fruiting sites and vegetative organs. Despite the key roles of MCTPs in the intercellular and intracellular transportation, no *MCTP* was identified in *G. hirsutum* up to now. In this study, we performed the genome-wide identification of GhMCTPs and analyzed their physicochemical properties, phylogenetic relationship with other plants' MCTPs, gene structures, domain architectures, syntenic relationship and spatiotemporal expression. We also investigated the physicochemical properties of the N-terminal C2 domains and C-terminal transmembrane regions of GhMCTPs, evolutionary divergence of multiple C2 domains and the interaction between GhMCTPs' C2 domains and GhFT. Our results will be helpful for future characterization of GhMCTPs' roles in cotton growth and development.

Results

Identification and physicochemical properties of GhMCTPs

AtFTIP1 is one of the well-researched *MCTPs* in *Arabidopsis* [25]. Its protein sequence was used as the query to search against the protein database of *G. hirsutum* for putative GhMCTPs. After confirming the protein domains of the BLAST hits in SMART database, we identified 31 GhMCTPs, each of which contained three to four C2 domains in their N-terminus and one to four transmembrane regions in their C-terminus. The putative GhMCTPs were named as GhMCTP1–31 according to their sequence similarity to AtFTIP1 and classified into five subfamilies based on phylogenetic analysis (Figure 1). The lengths of GhMCTPs protein sequences ranged from 730 (GhMCTP21) to 1059 (GhMCTP25) amino acids (aa). Correspondingly, GhMCTP21 and GhMCTP25 had the minimum and maximum molecular weight, respectively. The pI and Grand average of hydropathicity (GRAVY) of GhMCTPs ranged from 5.81 to 9.38 and −0.445 to −0.075, respectively (Figure 1). GhMCTPs from subfamily V showed the lowest pIs that were less than 7, indicating their acidic nature and distinct molecular roles from other GhMCTPs. Notably, GhMCTPs within subfamily I and II showed similar pIs and GRAVYs, whereas GhMCTPs within subfamily III, IV and V showed different pIs and GRAVYs, suggesting that GhMCTPs within different subfamilies had experienced different divergences during their evolution.

Phylogenetic analysis of MCTPs in seven plant species

To understand the evolutionary relationships among MCTPs in plants, MCTP homologs in *A. thaliana* (16), *G. raimondii (*17), *G. arboreum* (16), *G. barbadense* (29), *V. vinifera* (3) and *T. cacao* (12) were identified with the same method used in GhMCTPs' identification (Additional file 1: Table S1). AtMCTPs identified in our study were identical to those identified in the previous study [7]. MCTPs in two AtDt allotetraploids, *G. hirsutum* and *G. barbadense,* were less than the sum of MCTPs in D-genome *G. raimondii* and MCTPs in A-genome *G. arboreum,* suggesting that *G. hirsutum* and *G. barbadense* experienced gene loss after their formation. Phylogenetic analysis of MCTPs in seven plant species classified them into subfamily I-V with 15, 22, 27, 29, 31 members, respectively (Figure 2 and Additional file 1: Table S1). It was noteworthy that each subfamily contained MCTP homologs from our studied plant species, except *V. vinifera,* which suggested the early divergence of MCTPs in the common ancestor of these six plant species, In addition, most of MCTPs in *G. hirsutum* and *G. barbadense* were closely related with those in *G. raimondii* and *G. arboreum,* indicating that these MCTPs could be orthologous genes and had equivalent functions.

Exon-intron structures and chromosomal distributions of *GhMCTPs*

To better understand the relatedness of *MCTPs* in plant species, the exon-intron structures of *MCTPs* in seven plant species were comparatively analyzed. Most (25/31) of *GhMCTPs* were intronless, while *GhMCTP2, 6, 10, 17* and *31* contained one introns and *GhMCTP21* contained two introns. Consistently, most (81/93) of *MCTPs* in other six plant species were also intronless. However, all (3) the *MCTPs* in *V. vinifera* were intron-containing, which might be necessary to fine-tune genes' expression. According to the phylogenetic tree of MCTP protein sequences, the intron-containing *MCTPs*were dispersed in subfamily I-V and contained introns with variable lengths and positions (Figure 3A). These results suggested the random genesis of intron-containing and intronless *MCTPs.*

Thirty one *GhMCTPs* were unevenly distributed on 18 chromosomes, while the other 8 chromosomes didn't contain any *GhMCTPs.* Most of the chromosomes contained 1–2 *GhMCTPs,* while both A08 and D08 contained 4 *GhMCTPs.* In addition, A subgenome contained more *GhMCTPs* than D subgenome (Figure 3B).

Domain architectures and conserved motifs of GhMCTPs

The conserved domains of GhMCTPs were obtained by searching against the SMART database (Additional file 2: Table S2) and six conserved motifs of GhMCTPs were found using MEME. To further investigate the conservation and diversification of GhMCTPs, the featured domains, 3–4 N-terminal C2 domains and 1–4 C-terminal transmembrane regions, and conserved motifs of GhMCTPs were demonstrated on the phylogenetic tree. Most members from subfamily I, II and IV contained 3 N-terminal C2 domains, except GhMCTP16, whereas most members from subfamily III and V contained 4 N-terminal C2 domains, except GhMCTP11, GhMCTP29, GhMCTP31. Members from subfamily I, II and V contained 4, 3 and 2 C-terminal transmembrane regions, respectively, whereas members from subfamily III and IV contained 1–4 C-terminal transmembrane regions. (Figure 4B). The transmembrane regions of GhMCTPs were confirmed by TMHMM program (Additional file 3: Figure S1). The different domain architectures of GhMCTPs from different subfamilies hinted their divergent roles in cotton growth and development. However, GhMCTPs within subfamily I and II had similar domain architectures, indicating their functional similarity, while GhMCTPs within subfamily III, IV and V showed relatively divergent domain architectures, which was consistent with their divergent pIs and GRAVYs.

Six conserved motifs were detected in most GhMCTPs, while GhMCTP18 and GhMCTP21 contained five conserved motifs. For most GhMCTPs, motif 1, 2 and partial motif 6 were detected in the end of N-terminus which was the corresponding region of the last C2 domain, while motif 3, 4, 5 and partial motif 6 were detected in the C-terminus. However, no conserved motifs were detected in the most regions of N-terminus (Figure 4C), suggesting that the last C2 domain and transmembrane regions were more conserved than the other C2 domains, whose divergence might contribute to the structural and functional diversification of GhMCTPs.

Orthologous *GhMCTPs* between A and D subgenome of *G. hirsutum*

To determine whether *GhMCTPs* from A and D subgenome exhibited functional divergence, we identified 13 syntenic pairs of homologous *GhMCTPs* between A and D subgenome of *G. hirsutum* and all these syntenic pairs were located on the similar positions of homologous chromosomes between A and D subgenome, except that *GhMCTP20* and *GhMCTP23* were located on the A03 and D02, respectively (Figure 5), which might be due to the large reciprocal translocation between A02 and A03 [36]. The synonymous distances (Ks values) between these detected syntenic pairs, partially representing sequence divergence between the two progenitor genomes (A genome and D genome) that formed *G. hirsutum,* ranged from 0.032 to 0.119. According to the Ks values, the divergence times of these syntenic *GhMCTPs* were estimated to be 6.20–22.84 million years ago (MYA), with an average of 12.6 MYA (Table 1). In addition, 13 and 14 syntenic pairs of homologous *MCTPs* found in *G. barbadense, G. raimondii* and *G. arboreum* showed similar ranges of Ks values and divergence times to those in *G. hirsutum (*Additional file 4: Figure S2 and Additional file 5: Table S3), which were wider than the previously estimated divergence time (6.2–7.1 MYA) of A and D progenitor genomes [34]. The Ka/Ks ratios between all the syntenic *MCTPs* were less than 1.0, implying that these syntenic *MCTPs* experienced purifying selection during the divergence of the two progenitor genomes and might perform similar functions.

Spatiotemporal expression patterns of *GhMCTPs*

The previously published transcriptome datasets of *G. hirsutum* (TM–1) were used to profile the expression of *GhMCTPs* in various tissues, including anther, filament, pistil, bract, sepal, petal, torus, root, leaf, stem, fibers and ovules at different developmental stages. *GhMCTPs* from subfamily II were highly expressed in most tissues, especially in ovules at different developmental stages. *GhMCTP11, 13* from subfamily III and half members from subfamily V also showed high expression levels in most tissues (Figure 6), suggesting their constitutive roles in the development of various tissues. However, *GhMCTPs* from other subfamilies were only highly expressed in specific tissues, especially *GhMCTPs* from subfamily IV with *GhMCTP16, 17* highly expressed in early developmental ovules and *GhMCTP19, 21* highly expressed in early developmental fibers. *GhMCTP7* and *GhMCTP8* from subfamily III also had high expression levels in early developmental fibers and ovules at different developmental stages (Figure 6), suggesting their important roles in ovule and fiber development. These results revealed that *GhMCTPs* from different subfamilies had different expression patterns and might be involved in different biological processes.

Physicochemically different N-terminus and C-terminus of GhMCTPs

Since the N-terminus and C-terminus of GhMCTPs contained structurally and functionally different domains, which might be reflected by their physicochemical properties, we further analyzed the pIs and GRAVYs of the N-terminus and C-terminus of GhMCTPs (Additional file 6: Table S4). Both pIs and GRAVYs of full-length GhMCTPs were between those of N-terminus and C-terminus, and the C-terminus possessed higher pIs and GRAVYs than the N-terminus. Interestingly, the pIs of the C-terminus were almost invariable among all the GhMCTPs, while the pIs of the N-terminus varied significantly among GhMCTPs from different subfamilies and GhMCTPs within subfamily III and IV (Figure 7), suggesting that the N-terminus was more variable than the C-terminus and might be the main source of functional divergence of GhMCTPs. However, both the N-terminus and the C-terminus showed significantly different GRAVYs among GhMCTPs within the same subfamilies (Figure 7).

Evolutionary divergence of multiple C2 domains in the N-terminus of GhMCTPs

Since 3–4 C2 domains were contained in the N-terminus of GhMCTPs and showed great difference in protein sequences and physicochemical properties among GhMCTPs, we queried whether the 3–4 C2 domains contained in each of the GhMCTPs had different evolutionary histories or molecular roles and which C2 domain was more divergent among GhMCTPs than the other C2 domains. Four C2 domains of 4-C2-containing GhMCTPs and three C2 domains of 3-C2-containing GhMCTPs were designated as 4aC2, 4bC2, 4cC2, 4dC2 and 3aC2, 3bC2, 3cC2, respectively. The protein sequences of 107 C2 domains contained in 31 GhMCTPs (Additional file 2: Table S2) were used to construct the phylogenetic tree, which classified these C2 domains into subclade I-IV. Consistent with the multiple sequence alignment of the full-length GhMCTPs, in which the 4bC2, 4cC2 and 4dC2 of 4-C2-containing GhMCTPs were aligned with the 3aC2, 3bC2, 3cC2 of 3-C2-containing GhMCTPs, respectively (Additional file 7: Figure S3), the corresponding C2 domains of 4-C2-containing and 3-C2-containing GhMCTPs were classified into the same subclades. In addition, the C2 domains within subclade II and III exhibited larger sequence divergence than those within subclade I and IV (Figure 8). These results suggested that the 3–4 C2 domains contained in the GhMCTPs began to diverge before the formation of GhMCTPs probably through module exchange and fulfilled different functions in the multidomain GhMCTPs. Moreover, the more divergent 4bC2, 3aC2 and 4cC2, 3bC2 within subclade II and III might be the main source of GhMCTPs' functional diversification.

The N-terminus of GhMCTP2 and GhMCTP4 interacted with GhFT

The most widely researched MCTPs were FTIP1s [7, 25, 29, 30], which interacted with FTs to mediate their transport from leaves to SAM. We chose three evolutionarily distant GhMCTPs, including GhMCTP2, GhMCTP4 and GhMCTP27, to detect their interactions with GhFT via yeast two-hybrid assay. All the three

full-length GhMCTPs couldn't interact with GhFT, whereas the N-terminus of GhMCTP2, GhMCTP4 and GhMCTP27 showed strong, weak, and no interaction with GhFT, respectively (Figure 9). This was consistent with the interaction between FT and N-terminal C2 domains of FTIP1 in Arabidopsis, rice and orchid [7, 25, 29, 30]. The transmembrane regions in the C-terminus might hinder GhFT's interaction with GhMCTP2 and GhMCTP4 in yeast cells. The results suggested that the N-terminal C2 domains of GhMCTPs played key roles in transporting other regulators by direct interaction.

Discussion

Sequence characterization of GhMCTPs

Multiple C2 domains and transmembrane region proteins (MCTPs) contain three to four C2 domains in the N terminus and one to four transmembrane regions in the C terminus. MCTPs are evolutionary conserved proteins and have been identified in both animals and plants [7, 21, 24, 25, 27–32, 37]. Compared with animals and lower plants, higher plants contain significantly increased number of MCTPs, implying more diverse and specific functions of MCTPs in plant growth and development [7]. In this study, we identified 31 MCTPs in *G. hirsutum* and classified them into 5 subfamilies (Figure 1). The distinct physicochemical properties of GhMCTPs suggested that GhMCTPs played diverse roles in regulating cotton growth and development. Especially, all the GhMCTPs from subfamily V exhibited pIs less than 7.5 (Figure 1 and 7), the PH of most cell interior compartments [38], suggesting that GhMCTPs from subfamily V and GhMCTPs from other subfamilies were charged oppositely and played different molecular functions in their respective suitable environments.

Phylogenetic analysis of MCTPs in seven plant species classified these MCTPs into five subfamilies (Figure 2). Of the seven plant species, *V. vinifera* contained the least number of MCTPs (3), which was consistent with *V. vinifera's* lack of recent whole genome duplication (WGD) [39]. MCTPs in each of the other six plant species were dispersed in subfamily I-V (Figure 2), suggesting that MCTPs began to diverge in the common ancestor of these plant species and the following speciation led to the formation of MCTP orthologs in different plant species that performed similar functions. *G. hirsutum* and *G. barbadense* were two allotetraploids that formed ~1–2 MYA by the hybridization of two diploids (*G. arboreum* and *G. raimondii)* and the following chromosome doubling [40]. Previous study reported that 369 genes were lost in *G. hirsutum* [36]. The less MCTPs in *G. hirsutum* (31) and *G. barbadense* (29) than the sum (33) of MCTPs in *G. arboreum* and *G. raimondii* suggested that gene losses occurred in both *G. hirsutum* and *G. barbadense* (Additional file 1: Table S1). The divergence time of A and D progenitor genomes was estimated to be 6.2–7.1 MYA by computing peak Ks values of genome-wide orthologous genes among the four *Gossypium* species [34]. The wider range (6.20–22.84 MYA) of roughly estimated divergence times of syntenic *GhMCTPs* between A and D subgenome (Table 1) might be resulted from varied Ks values among different genes [41] and oversimplified using of general divergence rate for plant nuclear genes in the formula: $t = Ks/2r$ ($r = 2.6 \times 10^{-9}$).

Diverse expression patterns of *GhMCTPs*

The significantly increased number of MCTPs in higher plants may meet the requirement for more diverse and specific functions of MCTPs in regulating various cellular processes in plants [7]. Functional analyses of MCTPs in various plant species demonstrated that MCTPs were involved in promoting flowering, regulating shoot meristem development, controlling anther dehiscence, regulating cell growth anisotropy and exporting sucrose into sieve elements[7, 24, 25, 27–32]. GUS and GFP reporter assays of 16 AtMCTPs revealed that even MCTPs with close evolutionary relationship might be expressed in different tissues and some MCTPs might perform redundant or additive functions in certain tissues [7]. Our expression analysis showed that *GhMCTPs* from different subfamilies and within the same subfamilies exhibited different spatio-temporal expression patterns, especially *GhMCTPs* within subfamily III, IV and V (Figure 6), suggesting that *GhMCTPs* played diverse roles in the development of various cotton tissues. *GhMCTP1, 2* and *3* within subfamily I exhibited low expression in all the investigated tissues, which might be the result of these *GhMCTPs'* expression at specific locations of these tissues, as *AtMCTP1* and *AtMCTP2's* expression in the vascular tissues of leaves and roots [7]. Conversely, *GhMCTPs* within subfamily II were highly expressed in most of the investigated tissues, suggesting that these *GhMCTPs* are required to maintain basic cellular processes. The more diverse expression patterns, physicochemical properties and domain architectures of GhMCTPs within subfamily III, IV and V (Figure 6, 7 and 4B) indicated that GhMCTPs within subfamily III, IV and V evolved to perform more diverse and specific functions than GhMCTPs within subfamily I and II.

Characterization of multiple C2 domains and transmembrane regions in GhMCTPs

Both C2 domain and transmembrane region are able to target their host proteins to specific organelle membranes, with C2 domain binding to membrane phospholipid mainly in a Ca^{2+}-dependent manner [9, 11] and transmembrane region traversing phospholipid bilayer. A recent study on *Arabidopsis* plasmodesmal proteome revealed that MCTPs acted as ER-PM membrane tethers, with C2 domains docking to the PM and transmembrane region inserting into the ER. The distinct physicochemical properties between N-terminal C2 domains and C-terminal transmembrane regions of GhMCTPs (Figure 7) implied their distinct molecular roles in the interaction with membrane. Compared with the N-terminus, the C-terminus exhibited almost invariable pIs among all the GhMCTPs. In addition, conserved motifs were detected in the C-terminus but not in most regions of the N-terminus (Figure 4C). These results indicated that the C-terminal transmembrane regions were more conserved than the N-terminal C2 domains. However, GRAVYs of both the N-terminus and the C-terminus exhibited significant variation among GhMCTPs (Figure 7B), which might contribute to their different binding activities to various membranes whose compositions and physical properties could be very different [42].

MCTPs mediate intercellular and intracellular transport of other regulators through the N-terminal C2 domains' interaction with these regulators [7, 25, 28–31]. Yeast-two hybrid assay showed that the last C2 domain next to transmembrane regions directly interacted with target proteins [7, 30, 31]. The N-terminal C2 domains of GhMCTP2, GhMCTP4 and GhMCTP27 exhibited different capabilities to interact with GhFT (Figure 9), implying that C2 domains of different GhMCTPs were involved in the interaction with different target proteins to facilitate their movement. Further identification of the potential target proteins of various GhMCTPs is necessary to better understand GhMCTPs' regulatory roles in cotton growth and development. Phylogentic analysis of multiple C2 domains of 31 GhMCTPs showed that 4aC2, 4bC2 and 3aC2, 4cC2 and 3bC2, 4dC2 and 3cC2 were classified into subclade I, II, III, IV, respectively, suggesting that different C2 domains of each GhMCTP might fulfill different functions. Whether these C2 domains bind specific membranes and interact with target proteins independently or cooperatively remains to be further studied.

Conclusions

In our study, a systematic analysis of the multiple C2 domains and transmembrane region proteins (MCTPs) in *G. hirsutum* was performed to characterize their phylogenetic relationship, physicochemical properties, gene structures, domain architectures, conserved motifs and expression patterns. Furthermore, the N-terminus and the C-terminus of GhMCTPs were comparatively analyzed. GhMCTPs were classified into five subfamilies according to the phylogenetic tree. GhMCTPs within subfamiliy III, IV and V exhibited more diverse physicochemical properties, domain architectures and expression patterns than GhMCTPs within subfamily I and II. The distinct physicochemical properties between the N-terminus and the C-terminus suggested their distinct molecular functions in GhMCTPs. Yeast two-hybrid assay indicated that the N-terminus was responsible for GhMCTPs' interaction with target proteins. Our study will benefit future studies on the functions of GhMCTPs in cotton growth and development.

Methods

Identification of GhMCTPs

The genomic sequences and annotated gene models of G. hirsutum were downloaded from the CottonGen website (https://www.cottongen.org/icgi/home) [43]. The protein sequence of AtFTIP1 (At5g06850) was used as the query to search against the protein database of *G. hirsutum* using BLAST with e-value threshold set at 1e−5 [25]. Then, all the BLAST hits were submitted to the SMART database (http://smart.embl-heidelberg.de/) to screen the putative GhMCTPs with 3–4 C2 domains in the N-terminus and 1–4 transmembrane regions in the C-terminus [44]. The identified GhMCTP protein sequences were aligned using Clustal Omega with default parameters (https://www.ebi.ac.uk/Tools/msa/clustalo/) [45]. The MrBayes v3.2.5 was used to construct the phylogenetic tree with the evolutionary model set to the GTR substitution model with gamma-distributed rate variation across sites and Ngen, Samplefreq set to 300000, 100, respectively [46].

The first amino acid to the right border of the last C2 domain and the remaining part in each GhMCTP were defined as the N-terminus and the C-terminus, respectively. The theoretical Mw, pI and GRAVY of the full length, N-terminus, C-terminus of GhMCTPs were calculated on the ExPASy website (http://web.expasy.org/protparam/) [47].

Phylogenetic analysis of MCTPs

The genomic sequences and annotated gene models of *G. barbadense, G.raimondii* and *G. arboreum* were downloaded from the CottonGen website (https://www.cottongen.org/icgi/home) [43], *A. thaliana* and *V. vinifera* were downloaded from the phytozome database (https://phytozome.jgi.doe.gov/) and *T. cacao* was downloaded from Cocoa Genome Hub (https://cocoa-genome-hub.southgreen.fr/) [48, 49]. MCTPs in these species were identified with the same procedure used in the identification of GhMCTPs. All identified MCTP protein sequences were aligned using Clustal Omega with default parameters (https://www.ebi.ac.uk/Tools/msa/clustalo/) [45]. The MrBayes v3.2.5 was used to construct the phylogenetic tree with the evolutionary model set to the GTR substitution model with gamma-distributed rate variation across sites and Ngen, Samplefreq set to 300000, 100, respectively [46].

Gene structures and chromosomal locations of *GhMCTPs*

The gene structures of *MCTPs* in seven plant species were obtained from the annotated gene models contained in the gff3 files and displayed on the phylogenetic tree of MCTP protein sequences using iTOL v4 (https://itol.embl.de/) [50].

The chromosomal locations of *GhMCTPs* were also obtained from the annotated gene models contained in the gff3 files and displayed by TBtools [51].

Domain and conserved motif analysis

The lengths and positions of C2 domains and transmembrane regions in each GhMCTP were predicted by searching against the SMART database and displayed on the phylogenetic tree of GhMCTPs using iTOL v4 (https://itol.embl.de/) [50]. The conserved motifs in GhMCTPs were discovered using MEME v5.0.5 (http://meme-suite.org/tools/meme) with the following parameters: site distribution, zero or one occurrence per sequence; number of motifs, 6; motif width, between 6 and 50 [52].

Synteny analysis and divergence time estimation

The MCScanX software (http://chibba.pgml.uga.edu/mcscan2/) was employed to detect syntenic *MCTPs* between A and D genome of *G. hirsutum, G. barbadense, G. raimondii, G. arboreum* according to the author's manual [53]. These syntenic MCTPs were displayed using TBtools. The coding sequences of syntenic *MCTPs* were used to calculate Ka and Ks by TBtools with the NG method [51]. The divergence time was calculated according to the following formula: t = Ks/2r (r = 2.6×10^{-9}) [41].

Expression analysis of *GhMCTPs* in different tissues

The transcriptome datasets of 23 cotton tissues were downloaded from the NCBI website under the BioProject PRJNA490626 (https://www.ncbi.nlm.nih.gov/bioproject/PRJNA490626), then transcriptomic reads were mapped against the *G. hirsutum* genome using HISAT2 and the read counts mapped on each gene were calculated using HTSeq v0.11.1 [54, 55]. Log2-transformed FPKMs of each *GhMCTP* in different cotton tissues were displayed on the heatmap using iTOL v4 (https://itol.embl.de/) [50].

Phylogenetic analysis of C2 domains in the GhMCTPs

Four C2 domains of 4-C2-containing GhMCTPs and three C2 domains of 3-C2-containing GhMCTPs were designated as 4aC2, 4bC2, 4cC2, 4dC2 and 3aC2, 3bC2, 3cC2, respectively. The protein sequences of 107 C2 domains in the 31 GhMCTPs were extracted according to their lengths and positions in the full-length GhMCTPs. The obtained protein sequences were aligned using Clustal Omega with default parameters (https://www.ebi.ac.uk/Tools/msa/clustalo/) [45]. The MrBayes v3.2.5 was used to construct the phylogenetic tree with the evolutionary model set to the GTR substitution model with gamma-distributed rate variation across sites and Ngen, Samplefreq set to 300000, 100, respectively [46].

Yeast two-hybrid assay

The coding sequences of full length and N-terminus of *GhMCTP2, GhMCTP4* and *GhMCTP27* were cloned into the pGADT7 vector (Clontech) and the coding sequence of *GhFT* (*Ghir_D08G024850.1)* was cloned into the pGBKT7 vector (Clontech) with the gene-specific primers (Additional file 8: Table S5). Then, different combinations of recombinant pGADT7 and pGBKT7 were co-transferred into the yeast strain Y2HGold which was cultured on DDO (SD/-Leu/-Trp) plates for three days. Three independent colonies on the DDO plates were chosen to test the interactions on QDO (SD/-Leu/-Trp/-His/-Ade) plates with 10 mM 3-AT (3-amino−1,2,4 -triazole).

Abbreviations

MCTP: multiple C2 domains and transmembrane region protein; PD: plasmodesma; ER: endoplasmic reticulum; PM: plasma membrane; SAM: shoot apical meristem; GRAVY: Grand average of hydropathicity; Ka: nonsynonymous substitution rate; Ks: synonymous substitution rate; MYA: million years ago; WGD: whole-genome duplication

Declarations

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable

Availability of data and materials

All data supporting the conclusions of this article are included in the article and its additional files.

Competing interests

The authors declare that they have no competing interests.

Funding

This research was supported by Shandong TAISHAN Industry leading talent program to Shuxun Yu (LJNY201608). The funders had no role in study design, data collection and analysis, and interpretation of data and in writing the manuscript.

Authors' contributions

S. X. Y. and H. L. W. designed the experiments. P. Y. C. performed the synteny analysis. A.M.W downloaded the transcriptomic data. S. S. C. and P. B. H. performed yeast two-hybrid assay. P. B. H. analyzed the results and wrote the manuscript. H. T. W. and L. M. revised the manuscript. All authors reviewed and approved the final manuscript.

Acknowledgements

We are grateful to all the colleagues in our laboratory who provided helpful advice and technical assistance.

References

1.McLean BG, Hempel FD, Zambryski PC: *Plant intercellular communication via plasmodesmata. The Plant cell* 1997, *9(7*):1043–1054.

2.Cosgrove DJ: *Growth of the plant cell wall. Nature reviews Molecular cell biology* 2005, *6(*11):850–861.

3.Burch-Smith TM, Brunkard JO, Choi YG, Zambryski PC: *Organelle-nucleus cross-talk regulates plant intercellular communication via plasmodesmata. Proceedings of the National Academy of Sciences of the United States of America* 2011, *108(*51):E1451–1460.

4.Xu XM, Jackson D: *Lights at the end of the tunnel: new views of plasmodesmal structure and function. Current opinion in plant biology* 2010, *13(*6):684–692.

5.Tilsner J, Nicolas W, Rosado A, Bayer EM: *Staying Tight: Plasmodesmal Membrane Contact Sites and the Control of Cell-to-Cell Connectivity in Plants. Annual review of plant biology* 2016, *67:*337–364.

6.Brault ML, Petit JD, Immel F, Nicolas WJ, Brocard L, Gaston A, Fouché M, Hawkins TJ, Crowet J, Grison MS *et al: Multiple C2 domains and Transmembrane region Proteins (MCTPs) tether membranes at plasmodesmata. BioRxiv* 2018.

7.Liu L, Li C, Liang Z, Yu H: *Characterization of Multiple C2 Domain and Transmembrane Region Proteins in Arabidopsis. Plant physiology* 2018, *176(*3):2119–2132.

8.Sossin WS, Schwartz JH: *Ca(2+)-independent protein kinase Cs contain an amino-terminal domain similar to the C2 consensus sequence. Trends in biochemical sciences* 1993, *18(*6):207–208.

9.Chapman ER, Jahn R: *Calcium-dependent interaction of the cytoplasmic region of synaptotagmin with membranes. Autonomous function of a single C2-homologous domain. The Journal of biological chemistry* 1994, *269(*8):5735–5741.

10.Perin MS, Fried VA, Mignery GA, Jahn R, Sudhof TC: *Phospholipid binding by a synaptic vesicle protein homologous to the regulatory region of protein kinase C. Nature* 1990, *345(*6272):260–263.

11.Davletov BA, Sudhof TC: *A single C2 domain from synaptotagmin I is sufficient for high affinity Ca2+/phospholipid binding. The Journal of biological chemistry* 1993, *268(*35):26386–26390.

12.Verdaguer N, Corbalan-Garcia S, Ochoa WF, Fita I, Gomez-Fernandez JC: *Ca(2+) bridges the C2 membrane-binding domain of protein kinase Calpha directly to phosphatidylserine. EMBO J* 1999, *18(*22):6329–6338.

13.Leslie NR, Yang X, Downes CP, Weijer CJ: *PtdIns(3,4,5)P(3)-dependent and -independent roles for PTEN in the control of cell migration. Current biology: CB* 2007, *17(*2):115–125.

14.Cho W, Stahelin RV: *Membrane binding and subcellular targeting of C2 domains. Biochimica et biophysica acta* 2006, *1761(*8):838–849.

15.Zhang D, Aravind L: *Identification of novel families and classification of the C2 domain superfamily elucidate the origin and evolution of membrane targeting activities in eukaryotes. Gene* 2010, *469(*1–2):18–30.

16.Nalefski EA, Sultzman LA, Martin DM, Kriz RW, Towler PS, Knopf JL, Clark JD: *Delineation of two functionally distinct domains of cytosolic phospholipase A2, a regulatory Ca(2+)-dependent lipid-binding domain and a Ca(2+)-independent catalytic domain. The Journal of biological chemistry* 1994, *269(*27):18239–18249.

17.Clark JD, Lin LL, Kriz RW, Ramesha CS, Sultzman LA, Lin AY, Milona N, Knopf JL: *A novel arachidonic acid-selective cytosolic PLA2 contains a Ca(2+)-dependent translocation domain with homology to PKC and GAP. Cell* 1991, *65(*6):1043–1051.

18.Johnson CP: *Emerging Functional Differences between the Synaptotagmin and Ferlin Calcium Sensor Families. Biochemistry* 2017, *56(*49):6413–6417.

19.Mellor H, Parker PJ: *The extended protein kinase C superfamily. The Biochemical journal* 1998, *332 (Pt 2):*281–292.

20.Maeda I, Kohara Y, Yamamoto M, Sugimoto A: *Large-scale analysis of gene function in Caenorhabditis elegans by high-throughput RNAi. Current biology: CB* 2001, *11(*3):171–176.

21.Genç Ö, Dickman D, Ma W, Tong A, Fetter R, Davis G: *MCTP is an ER-resident calcium sensor that stabilizes synaptic transmission and homeostatic plasticity. eLife* 2017, *6:*e22904.

22.Djurovic S, Le Hellard S, Kahler AK, Jonsson EG, Agartz I, Steen VM, Hall H, Wang AG, Rasmussen HB, Melle I *et al: Association of MCTP2 gene variants with schizophrenia in three independent samples of Scandinavian origin (SCOPE). Psychiatry research* 2009, *168(*3):256–258.

23.Scott LJ, Muglia P, Kong XQ, Guan W, Flickinger M, Upmanyu R, Tozzi F, Li JZ, Burmeister M, Absher D *et al: Genome-wide association and meta-analysis of bipolar disorder in individuals of European ancestry. Proceedings of the National Academy of Sciences of the United States of America* 2009, *106(*18):7501–7506.

24.Fulton L, Batoux M, Vaddepalli P, Yadav RK, Busch W, Andersen SU, Jeong S, Lohmann JU, Schneitz K: *DETORQUEO, QUIRKY, and ZERZAUST represent novel components involved in organ development mediated by the receptor-like kinase STRUBBELIG in Arabidopsis thaliana. PLoS genetics* 2009, *5(*1):e1000355.

25.Liu L, Liu C, Hou X, Xi W, Shen L, Tao Z, Wang Y, Yu H: *FTIP1 Is an Essential Regulator Required for Florigen Transport. Plos Biology* 2012, *10(*4):e1001313.

26.Vaddepalli P, Herrmann A, Fulton L, Oelschner M, Hillmer S, Stratil TF, Fastner A, Hammes UZ, Ott T, Robinson DG *et al: The C2-domain protein QUIRKY and the receptor-like kinase STRUBBELIG localize to plasmodesmata and mediate tissue morphogenesis in Arabidopsis thaliana. Development* 2014, *141(*21):4139–4148.

27.Trehin C, Schrempp S, Chauvet A, Berne-Dedieu A, Thierry AM, Faure JE, Negrutiu I, Morel P: *QUIRKY interacts with STRUBBELIG and PAL OF QUIRKY to regulate cell growth anisotropy during Arabidopsis gynoecium development. Development* 2013, *140(*23):4807–4817.

28.Liu L, Li C, Song S, Teo ZWN, Shen L, Wang Y, Jackson D, Yu H: *FTIP-Dependent STM Trafficking Regulates Shoot Meristem Development in Arabidopsis. Cell reports* 2018, *23(*6):1879–1890.

29.Wang Y, Liu L, Song S, Li Y, Shen L, Yu H: *DOFT and DOFTIP1 affect reproductive development in the orchid Dendrobium Chao Praya Smile. Journal of experimental botany* 2017, *68(*21–22):5759–5772.

30.Song S, Chen Y, Liu L, Wang Y, Bao S, Zhou X, Teo ZW, Mao C, Gan Y, Yu H: *OsFTIP1-Mediated Regulation of Florigen Transport in Rice Is Negatively Regulated by the Ubiquitin-Like Domain Kinase OsUbDKgamma4. The Plant cell* 2017, *29(*3):491–507.

31.Song S, Chen Y, Liu L, See YHB, Mao C, Gan Y, Yu H: *OsFTIP7 determines auxin-mediated anther dehiscence in rice. Nature plants* 2018, *4(*7):495–504.

32.Tran TM, McCubbin TJ, Bihmidine S, Julius BT, Baker RF, Schauflinger M, Weil C, Springer N, Chomet P, Wagner R *et al: Maize Carbohydrate Partitioning Defective33 Encodes an MCTP Protein and Functions in Sucrose Export from Leaves. Molecular plant* 2019.

33.Li F, Fan G, Lu C, Xiao G, Zou C, Kohel RJ, Ma Z, Shang H, Ma X, Wu J *et al: Genome sequence of cultivated Upland cotton (Gossypium hirsutum TM–1) provides insights into genome evolution. Nature biotechnology* 2015, *33(*5):524–530.

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

35.Indest MO: *Factors Affecting Within-plant Variation of Coton Fiber Quality and Yield. LSU Doctoral Dissertations* 2015, *1091.*

36.Zhang T, Hu Y, Jiang W, Fang L, Guan X, Chen J, Zhang J, Saski CA, Scheffler BE, Stelly DM *et al: Sequencing of allotetraploid cotton (Gossypium hirsutum L. acc. TM–1) provides a resource for fiber improvement. Nature biotechnology* 2015, *33(*5):531–537.

37.Shin OH, Han W, Wang Y, Sudhof TC: *Evolutionarily conserved multiple C2 domain proteins with two transmembrane regions (MCTPs) and unusual Ca2+ binding properties. The Journal of biological chemistry* 2005, *280(*2):1641–1651.

38.Kiraga J, Mackiewicz P, Mackiewicz D, Kowalczuk M, Biecek P, Polak N, Smolarczyk K, Dudek MR, Cebrat S: *The relationships between the isoelectric point and: length of proteins, taxonomy and ecology of organisms. BMC genomics* 2007, *8:*163.

39.Qiao X, Li Q, Yin H, Qi K, Li L, Wang R, Zhang S, Paterson AH: *Gene duplication and evolution in recurring polyploidization-diploidization cycles in plants. Genome biology* 2019, *20(*1):38.

40.Paterson AH, Wendel JF, Gundlach H, Guo H, Jenkins J, Jin D, Llewellyn D, Showmaker KC, Shu S, Udall J *et al: Repeated polyploidization of Gossypium genomes and the evolution of spinnable cotton fibres. Nature* 2012, *492(*7429):423–427.

41.Senchina SD, Alvarez I, Cronn RC, Liu B, Rong J, Noyes RD, Paterson AH, Wing RA, Wilkins TA, Wendel JF: *Rate Variation Among Nuclear Genes and the Age of Polyploidy in Gossypium. Molecular biology and evolution* 2003, *20(*4):633–643.

42.Sharpe HJ, Stevens TJ, Munro S: *A comprehensive comparison of transmembrane domains reveals organelle-specific properties. Cell* 2010, *142(*1):158–169.

43.Yu J, Jung S, Cheng C-H, Ficklin SP, Lee T, Zheng P, Jones D, Percy RG, Main D: *CottonGen: a genomics, genetics and breeding database for cotton research. Nucleic acids research* 2014, *42(*D1):D1229-D1236.

44.Letunic I, Doerks T, Bork P: *SMART: recent updates, new developments and status in 2015. Nucleic acids research* 2015, *43(*Database issue):D257–260.

45.Madeira F, Park Ym, Lee J, Buso N, Gur T, Madhusoodanan N, Basutkar P, Tivey ARN, Potter SC, Finn RD *et al: The EMBL-EBI search and sequence analysis tools APIs in 2019. Nucleic acids research* 2019, *47(*W1):W636-W641.

46.Ronquist F, Huelsenbeck JP: *MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics* 2003, *19(*12):1572–1574.

47.Artimo P, Jonnalagedda M, Arnold K, Baratin D, Csardi G, de Castro E, Duvaud S, Flegel V, Fortier A, Gasteiger E *et al: ExPASy: SIB bioinformatics resource portal. Nucleic acids research* 2012, *40(*Web Server issue):W597–603.

48.Argout X, Salse J, Aury JM, Guiltinan MJ, Droc G, Gouzy J, Allegre M, Chaparro C, Legavre T, Maximova SN *et al: The genome of Theobroma cacao. Nature genetics* 2011, *43(*2):101–108.

49.Goodstein DM, Shu S, Howson R, Neupane R, Hayes RD, Fazo J, Mitros T, Dirks W, Hellsten U, Putnam N *et al: Phytozome: a comparative platform for green plant genomics. Nucleic acids research* 2012,

*40(*Database issue):D1178–1186.

50.Letunic I, Bork P: *Interactive Tree Of Life (iTOL) v4: recent updates and new developments. Nucleic acids research* 2019, *47(*W1):W256-W259.

51.Chen C, Xia R, Chen H, He Y: *TBtools, a Toolkit for Biologists integrating various HTS-data handling tools with a user-friendly interface. bioRxiv* 2018.

52.Bailey TL, Johnson J, Grant CE, Noble WS: *The MEME Suite. Nucleic acids research* 2015, *43(*W1):W39–49.

53.Wang Y, Tang H, Debarry JD, Tan X, Li J, Wang X, Lee TH, Jin H, Marler B, Guo H *et al: MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. Nucleic acids research* 2012, *40(7*):e49.

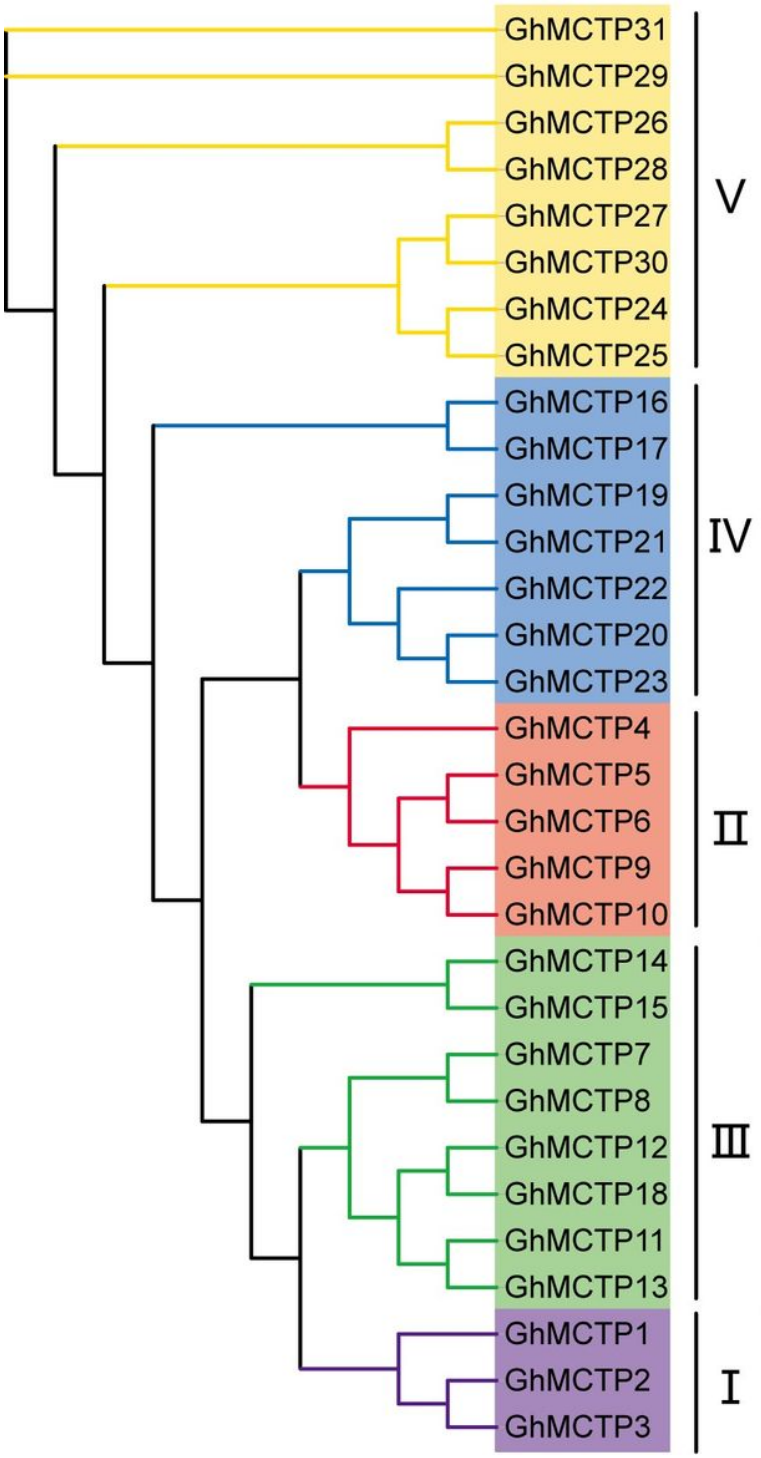
54.Kim D, Langmead B, Salzberg SL: *HISAT: a fast spliced aligner with low memory requirements. Nature methods* 2015, *12(*4):357–360.

55.Anders S, Pyl PT, Huber W: *HTSeq—a Python framework to work with high-throughput sequencing data. Bioinformatics* 2015, *31(*2):166–169.

Tables

Due to technical limitations, tables are only available as a download in the supplemental files section

Figures



	Length(aa)	Mw(KDa)	pI	GRAVY
GhMCTP31	958	108.10	5.81	-0.299
GhMCTP29	904	101.70	6.64	-0.283
GhMCTP26	1003	112.76	6.68	-0.358
GhMCTP28	1003	112.75	6.83	-0.35
GhMCTP27	1021	114.00	6.61	-0.305
GhMCTP30	947	105.35	6.44	-0.303
GhMCTP24	1057	118.89	6.75	-0.337
GhMCTP25	1059	119.03	6.65	-0.308
GhMCTP16	1011	113.72	9.3	-0.266
GhMCTP17	889	99.91	9.27	-0.241
GhMCTP19	770	88.05	8.92	-0.265
GhMCTP21	730	83.56	9	-0.289
GhMCTP22	778	89.14	7.84	-0.088
GhMCTP20	771	86.77	7.81	-0.075
GhMCTP23	793	89.44	8.6	-0.103
GhMCTP4	775	88.90	9.16	-0.237
GhMCTP5	771	88.17	9.22	-0.192
GhMCTP6	761	87.14	9.24	-0.179
GhMCTP9	733	84.09	9.36	-0.151
GhMCTP10	771	88.44	9.38	-0.175
GhMCTP14	1026	118.03	8.91	-0.406
GhMCTP15	1046	120.28	8.83	-0.441
GhMCTP7	1010	114.99	9.15	-0.232
GhMCTP8	1011	114.97	9.11	-0.231
GhMCTP12	1005	114.86	9.12	-0.376
GhMCTP18	949	108.35	8.92	-0.445
GhMCTP11	842	95.91	9.32	-0.38
GhMCTP13	976	110.72	9.33	-0.372
GhMCTP1	789	90.99	9.17	-0.296
GhMCTP2	805	92.51	9.1	-0.322
GhMCTP3	805	92.70	9.15	-0.32

Figure 1

The identified GhMCTPs and their physiochemical properties. Thirty one GhMCTPs are classified into five subfamilies according to the phylogenetic tree constructed by MrBayes v3.2.5. The length, Mw, pI and GRAVY are listed in the right table.

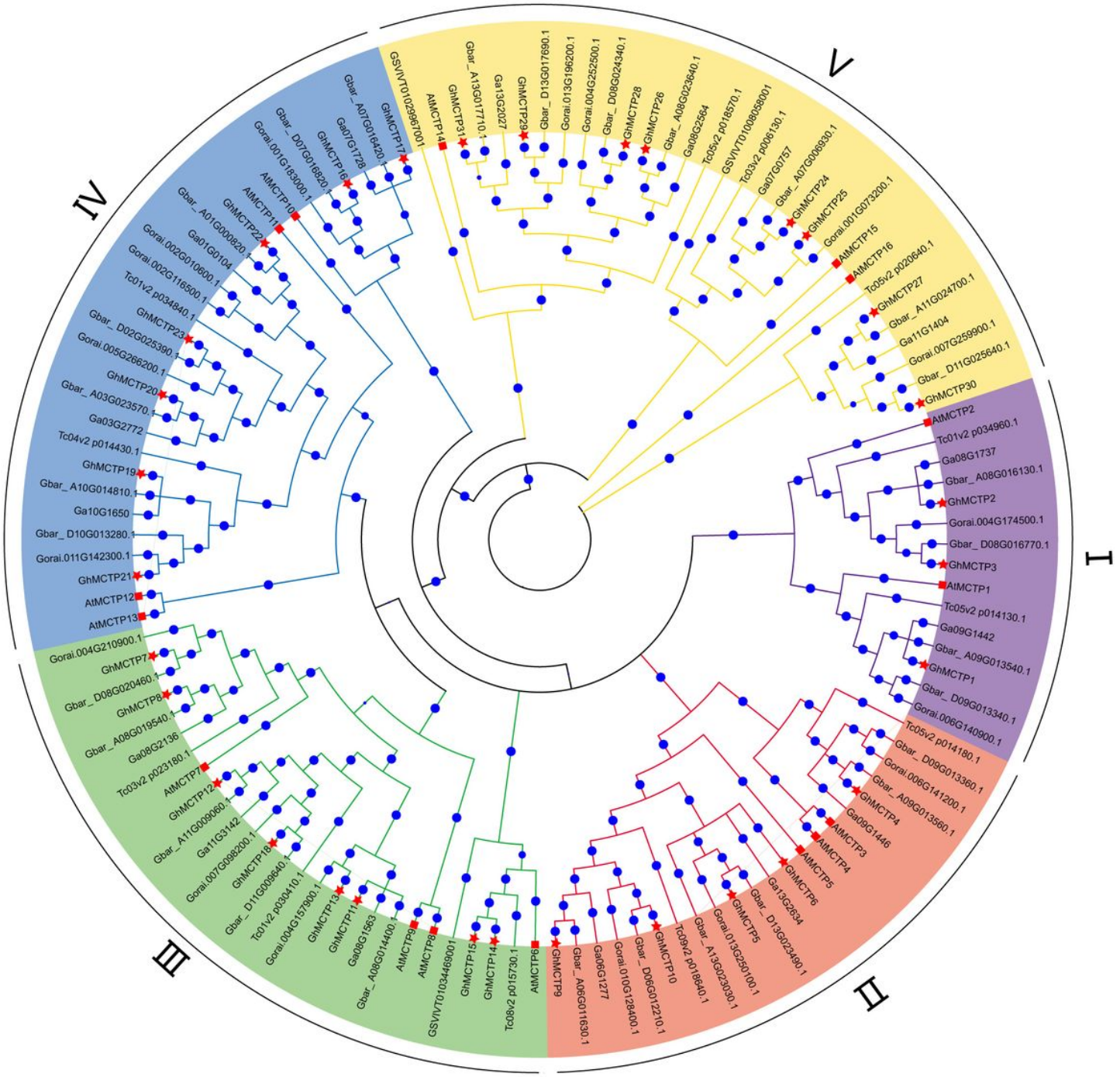


Figure 2

Phylogenetic tree of MCTPs in seven plant species. A total of 120 identified MCTPs in seven plant species (Gh, G.hirsutum; Gbar, G. barbadense; Gorai, G.raimondii; Ga, G. arboreum; At, A. thaliana; GSVIVT, V. vinifera; Tc, T. cacao) are classified into five subfamilies according to the phylogenetic tree constructed by MrBayes v3.2.5. Stars and squares indicate MCTPs from G. hirsutum and A. thaliana, respectively. The sizes of filled circles represent the probability of each partition in the tree ranging from 0.5 to 1.0.

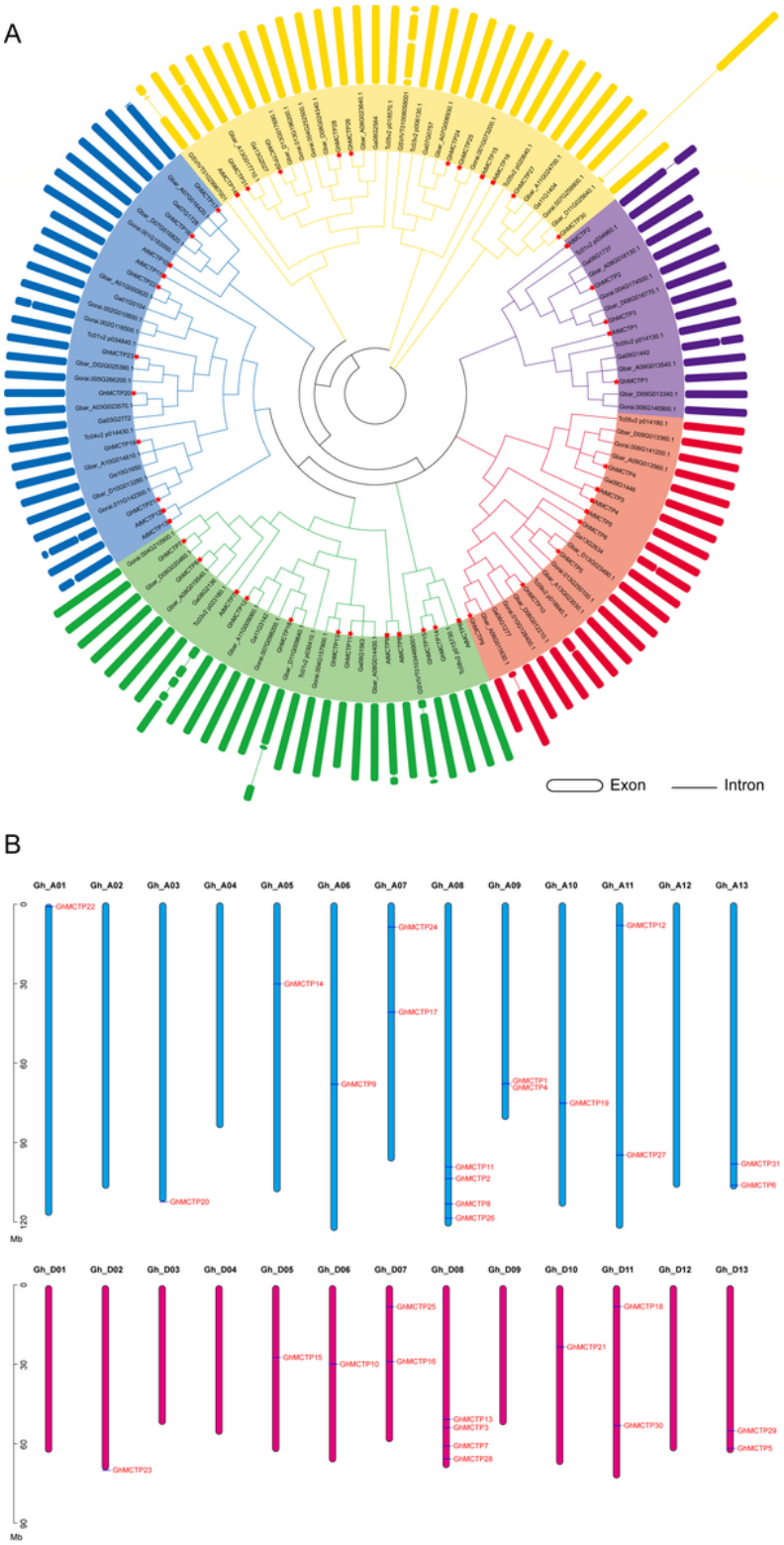


Figure 3

Exon-intron structures and chromosomal distributions. (A) Exon-intron structures of 120 MCTPs in seven plant species are displayed on the phylogenetic tree. Exons and introns are indicated by the rounded rectangles and lines, respectively. (B) The positions of GhMCTPs on the A and D subgenome are displayed on the blue and red bars, respectively. The lengths of bars represent the lengths of corresponding chromosomes.

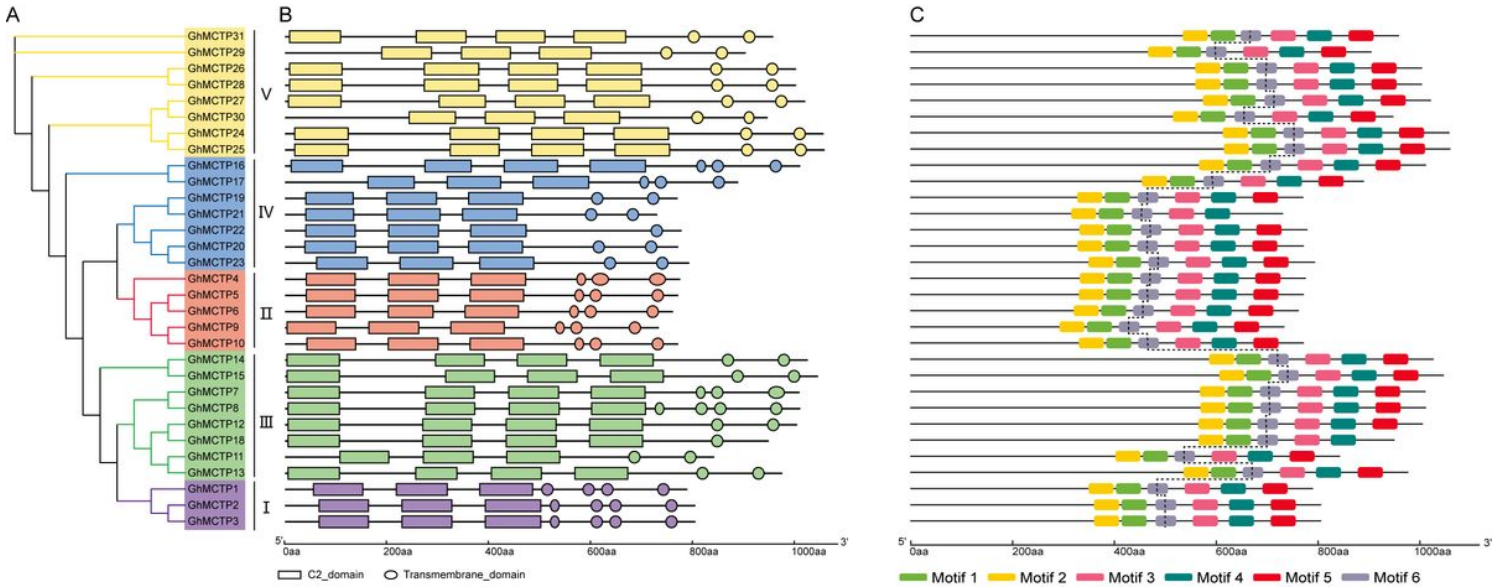


Figure 4

Domain architectures and conserved motifs of GhMCTPs. (A) Phylogenetic tree of GhMCTPs. (B) domain architectures of GhMCTPs. Rectangles and circles represent C2 domains and transmembrane regions, respectively. (C) Six conserved motifs in GhMCTPs are discovered using MEME. The dotted line represent the border between the N-terminus and C-terminus of GhMCTPs.

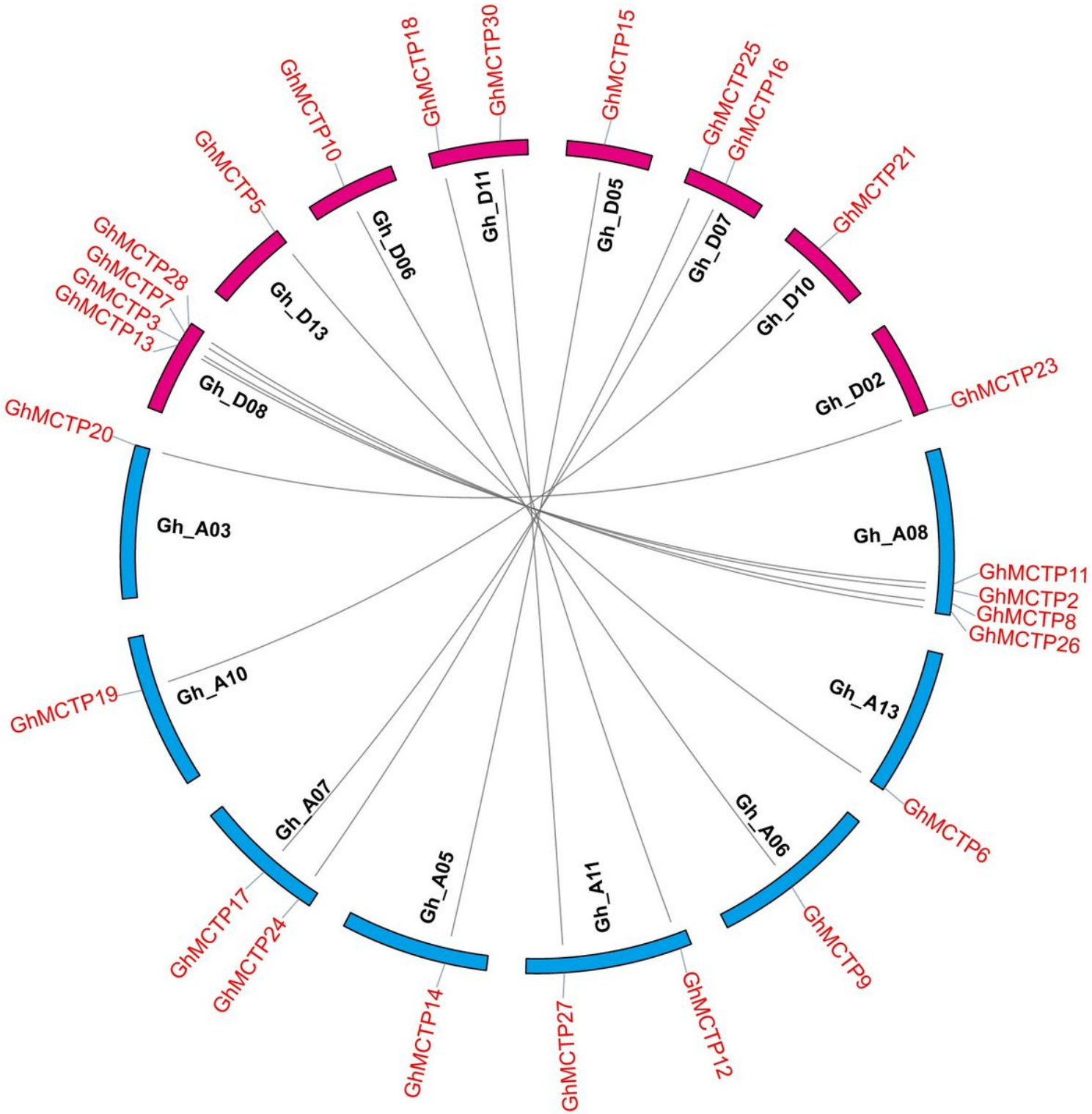


Figure 5

Syntenic GhMCTPs between A and D subgenome of G. hirsutum. Blue and red bars represent chromosomes from A and D subgenome of G. hirsutum, respectively. The grey lines link syntenic GhMCTPs detected by MCScanX.

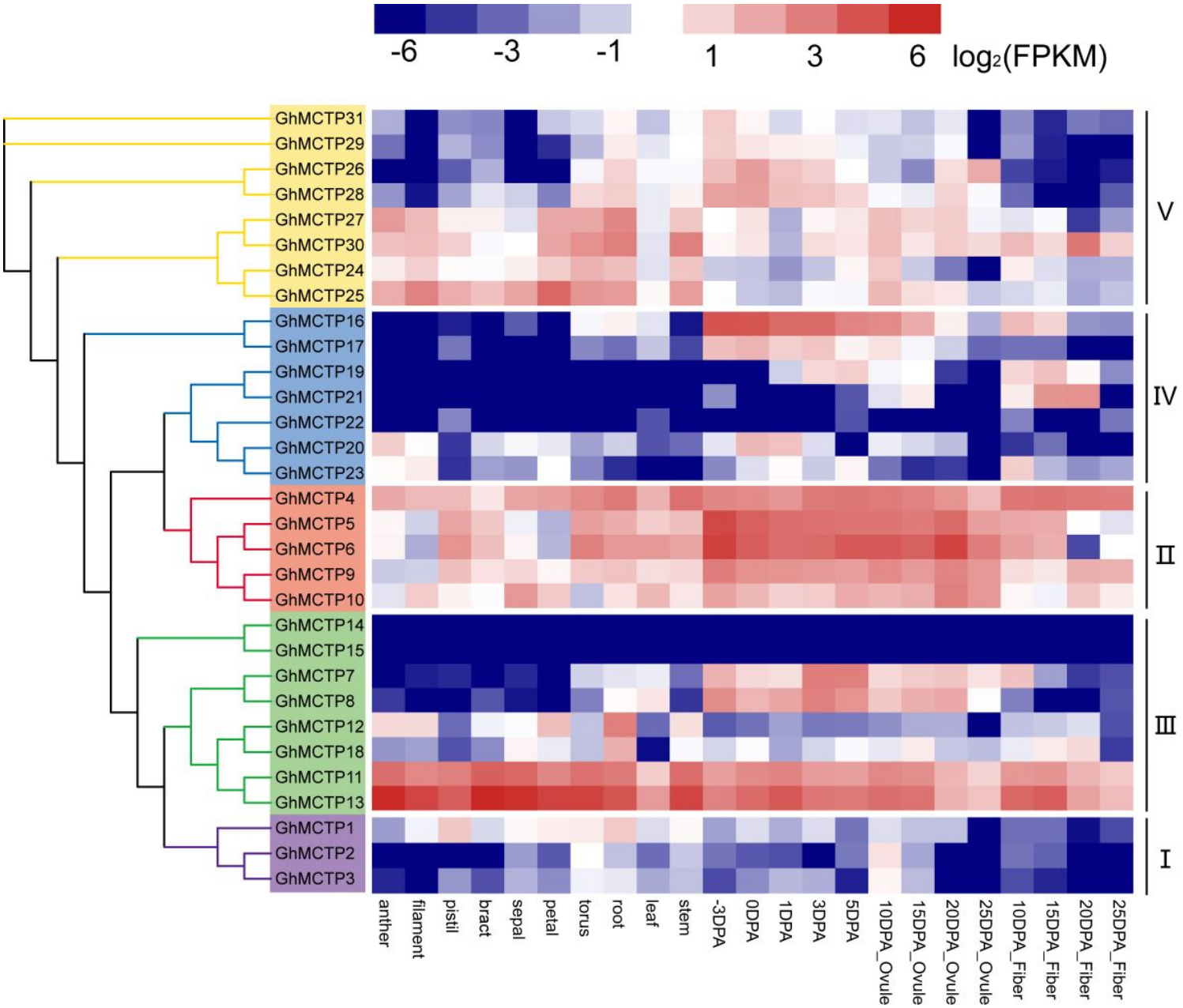


Figure 6

Expression characteristics of GhMCTPs. The expression levels of GhMCTPs in 23 tissues are displayed on the right of the phylogenetic tree. Differently colored blocks in the scale bar and heatmap represent log2-transformed FPKM values. The investigated tissues are shown on the bottom.

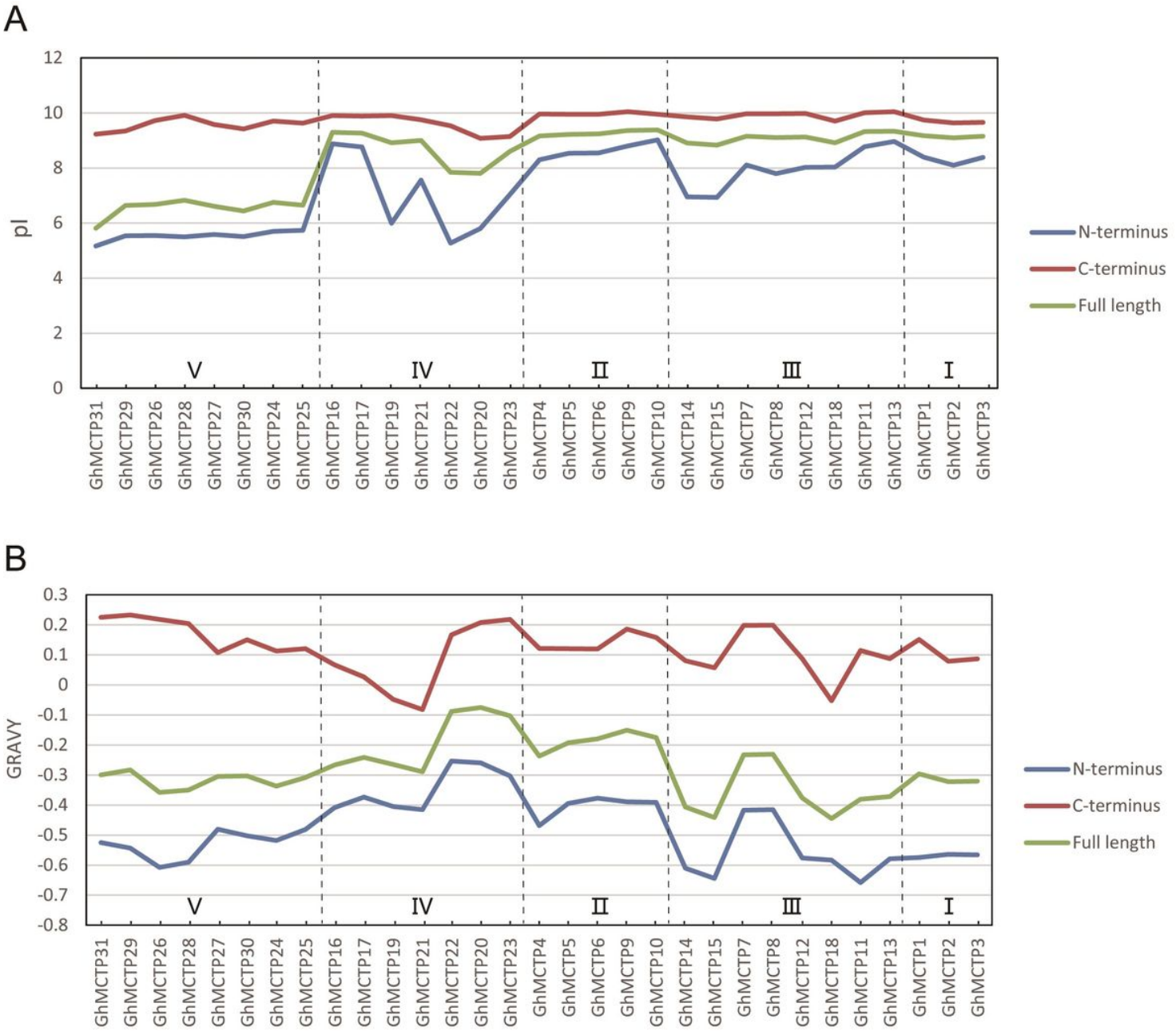


Figure 7

Distinct pIs and GRAVYs between the N-terminus and C-terminus of GhMCTPs. Thirty one GhMCTPs on the X-axis are arranged according to their positions in the phylogenetic tree. Five subfamilies are separated by the dotted lines. (A) The pIs of the N-terminus, C-terminus and full length of 31 GhMCTPs. (B) The GRAVYs of the N-terminus, C-terminus and full length of 31 GhMCTPs.

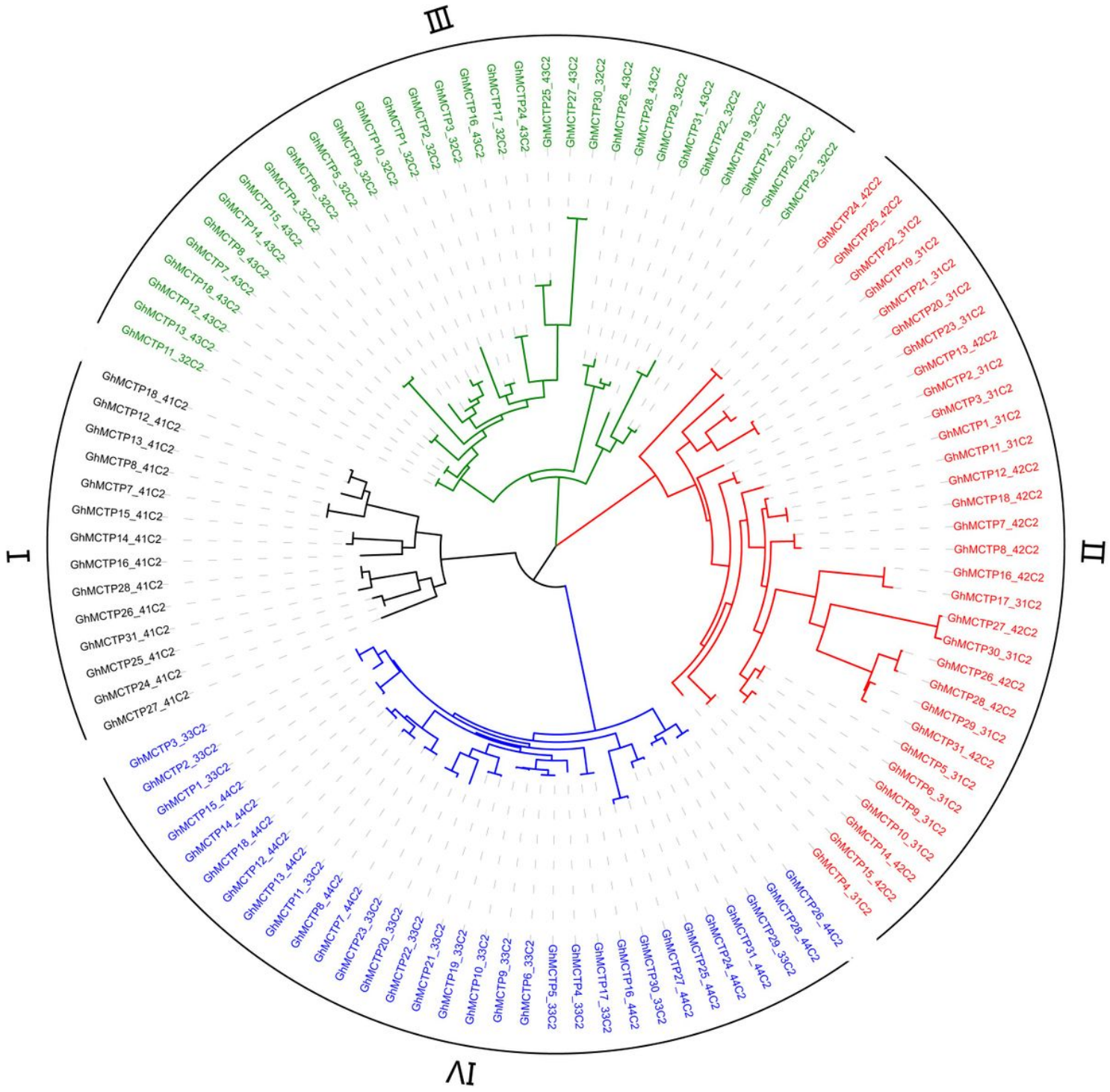


Figure 8

Phylogenetic tree of 107 C2 domains in 31 GhMCTPs. 4aC2, 4bC2, 4cC2, 4dC2 in 4-C2-containing GhMCTPs and 3aC2, 3bC2, 3cC2 in 3-C2-containing GhMCTPs are classified into 4 subclades according to the phylogenetic tree constructed by MrBayes v3.2.5.

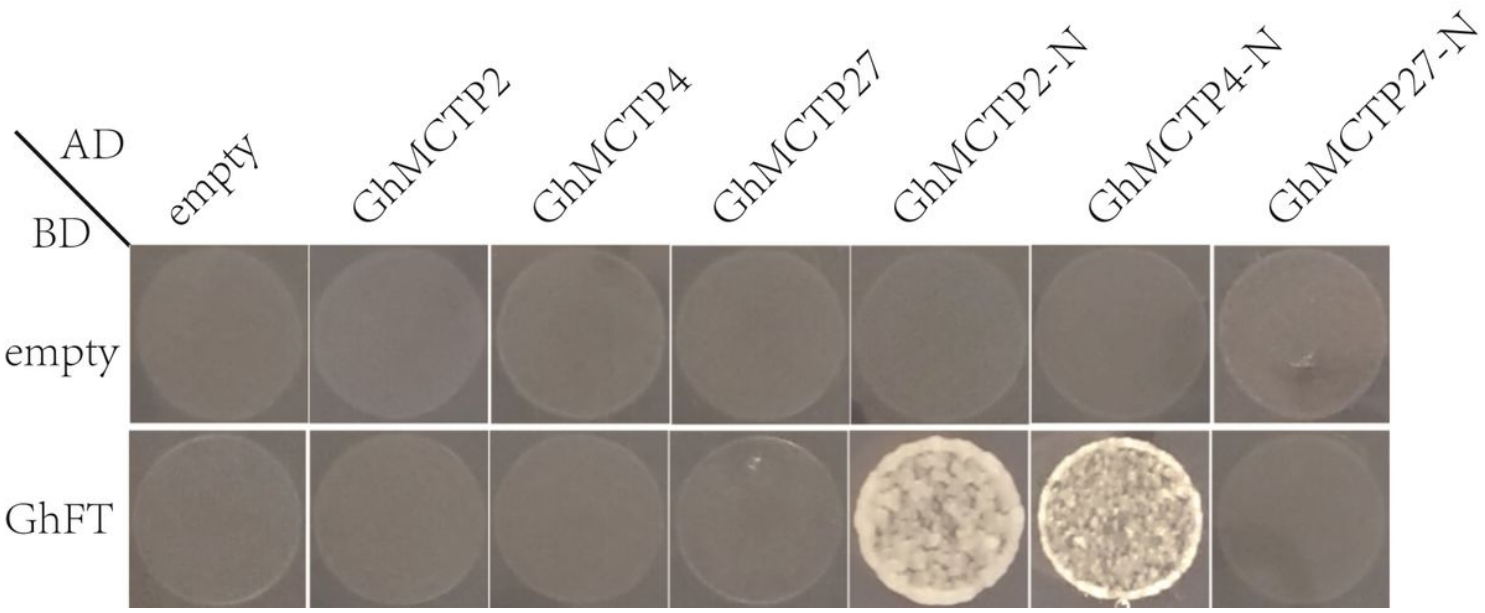


Figure 9

Yeast two-hybrid assay of interaction between GhFT and three GhMCTPs. Yeast cells are co-transformed with recombinant pGADT7 and pGBKT7 vectors and grown on the SD-Trp/-Leu/-His/-Ade/ medium with 10 mM 3-AT (3-amino-1,2,4 -triazole).

Supplementary Files

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

- Table1.xlsx
- Additionalfile8TableS5.xlsx
- Additionalfile6TableS4.xlsx
- Additionalfile5TableS3.xlsx
- Additionalfile4FigureS2.tif
- Additionalfile1TableS1.xlsx
- Additionalfile2TableS2.xlsx
- Additionalfile3FigureS1.tif
- Additionalfile7FigureS3.tif