



OPEN

SUBJECT AREAS:
PLANT GENETICS
DATA MININGReceived
3 July 2014Accepted
30 September 2014Published
16 October 2014Correspondence and
requests for materials
should be addressed to
B.H.Z. (zhangb@ecu.
edu)

Genome-wide identification and expression analysis of TCP transcription factors in *Gossypium raimondii*

Jun Ma¹, Qinglian Wang², Runrun Sun^{1,2}, Fuliang Xie¹, Don C. Jones³ & Baohong Zhang^{1,2}¹Department of Biology, East Carolina University, Greenville, NC 27858, USA, ²Henan Institute of Sciences and Technology, Xinxiang, Henan 453003, P. R. China, ³Cotton Incorporated, Cary, NC 27513, USA.

Plant-specific TEOSINTE-BRANCHED1/CYCLOIDEA/PCF (TCP) transcription factors play versatile functions in multiple aspects of plant growth and development. However, no systematical study has been performed in cotton. In this study, we performed for the first time the genome-wide identification and expression analysis of the TCP transcription factor family in *Gossypium raimondii*. A total of 38 non-redundant cotton TCP encoding genes were identified. The TCP transcription factors were divided into eleven subgroups based on phylogenetic analysis. Most TCP genes within the same subfamily demonstrated similar exon and intron organization and the motif structures were highly conserved among the subfamilies. Additionally, the chromosomal distribution pattern revealed that TCP genes were unevenly distributed across 11 out of the 13 chromosomes; segmental duplication is a predominant duplication event for TCP genes and the major contributor to the expansion of TCP gene family in *G. raimondii*. Moreover, the expression profiles of TCP genes shed light on their functional divergence.

Regulation of gene transcription, one of the most complex activities in cells, plays a significant role in a wide variety of biological processes, such as cell growth, cell cycle control, signal transduction, metabolic and physiological balance, and response to environmental stimuli^{1–5}. Among many mechanisms of transcriptional regulation of gene expression, transcription factors are considered to be the most important. Transcription factors are a diverse family of regulatory proteins with specific DNA-binding domains involved in the regulation of many cellular processes by either stimulating or repressing transcription of the related genes^{2,3,6}. Up to now, more than 60 transcription factor families have been identified in plants⁷.

The TCP proteins are a family of transcription factors exclusive to higher plants and involved in the regulation of cell growth and proliferation^{8,9}. This class of transcription factors are featured by a highly conserved ~60-residue-long DNA-binding motif at the N-terminus called TCP domain, which is named after four founding members: TB1 (TEOSINTE BRANCHED 1) in *Zea mays*, CYC (CYCLOIDEA) in *Antirrhinum majus*, and the PCF1 and PCF2 (PROLIFERATING CELL FACTORS 1 and 2) in *Oryza sativa*^{8,9}. The TCP domain contains a non-canonical basic-Helix-Loop-Helix (bHLH) structure involved in DNA binding, protein-protein interaction and protein nuclear localization^{8,10}. The two amphipathic helical motifs are abundant in hydrophobic Ala, Leu and Trp residues while the disordered linking loop region contains acidic, polar and non-charged amino acids. By comparison, the most conserved basic region is rich in positively charged Lys and Arg amino acids¹¹. The TCP transcription factor family can be further divided into two subfamilies, class I and class II, based mainly on amino acid sequence differences, especially in the basic region of the TCP domain⁸. According to the results of bioinformatic analysis, several class II TCP members also share an arginine-rich R domain outside the conserved TCP domain with unknown function, speculated to facilitate protein-protein interaction^{8,12}. The DNA-Binding site selection assays revealed that the two TCP classes can specifically recognize and bind to slightly different but partly overlapping GC-rich DNA sequences which act as cis-element in a large number of plant genes. The DNA binding sequence for class I is GGNCCCAC while class II prefer to bind the DNA motif G(T/C)GGNCCC^{10,13–15}.

Widely cultivated in more than 100 countries, cotton is considered one of the most important fiber-producing and economic crops around the world, providing fiber for the textile industry and cooking oil extracted from its oil-rich seeds for food industry. The cotton industry is estimated to produce \$133 billion in products and services annually, creating about 350 million jobs on farm or in the industry sectors. In spite of the economic and social importance of cotton and the critical role of TCP transcription factors in the control of plant cell proliferation and development, the research on cotton TCP family is much beyond on other plant species. In a recent study, Hao



et al (2012) reported the functional characterization of a cotton TCP transcription factor GbTCP. According to this research, GbTCP was expressed in cotton fiber at much higher level than other tissues tested. Overexpression of GbTCP in Arabidopsis facilitated the initiation and elongation of root hair which has similar developmental mechanisms with cotton fiber, whereas RNAi silencing of GbTCP in cotton led to shorter fiber and low fiber quality, indicating that GbTCP played a significant role in fiber elongation¹⁶. Up to now, however, no genome-wide characterization of TCP family members has been performed in cotton. The recent availability of the completed genome sequence of *Gossypium raimondii*⁵, a diploid cotton species, provides us with a great opportunity to identify and characterize TCP transcription factors in cotton genome.

In the present study, we performed for the first time the comprehensive analysis of the TCP transcription factor family in *G. raimondii*. A total of 38 non-redundant TCP transcription factor encoding genes were identified in the genome of *G. raimondii* and were subsequently subjected to a systematic analysis, including phylogenetic relationships, chromosomal location, gene duplication status, substitution rates, gene structure, conserved motif and expression profiling. On the basis of the expression profiles of TCP members in *G. raimondii* and the phylogenetic analysis among the TCP domain proteins in Arabidopsis, rice and *G. raimondii*, the functions of GrTCPs were predicted. Besides, it is also remarkable that the expansion of TCP family in *G. raimondii* may be caused mainly by segmental duplication and is not associated with tandem duplication. In

a word, our genome-wide analysis of TCP gene family will contribute to future studies on the functional characterization of TCP proteins in *G. raimondii* as well as the identification and comprehensive analysis of the TCP transcription factor family in other species.

Results

Identification of TCP genes. In order to identify the TCP transcription factor coding genes of *G. raimondii*, the HHM profile of TCP domain (PF03634) was employed as query to perform a blast search against the *G. raimondii* genome (<http://www.phytozome.net/cotton>). Originally, 62 candidate TCP genes were identified in *G. raimondii*. Among them, 24 redundant sequences were discarded from further analysis based on their sequence similarity. Subsequently, with the aim to verify the reliability of the initial results, a survey was conducted to confirm the existence of the conserved TCP domain with InterProScan¹⁸. The results showed that all of the 38 putative TCP genes contained conserved TCP domain. Due to the lack of standard annotation designated to the 38 TCP genes in the *G. raimondii*, we named them *GrTCP1* to *GrTCP25* according to the Arabidopsis TCP proteins with highest sequence similarity and following the nomenclature system applied to Arabidopsis. The length of the 38 newly identified TCP transcription factors varied from 196 to 549 amino acids with an average of 353.5 amino acids. Other characteristics of TCP transcription factors in *G. raimondii*, including isoelectric point (pI), molecular weight (Mw) and chromosome location, were listed in Table 1.

Table 1 | TCP gene family in *Gossypium raimondii*

Gene Name	Gene symbol	Length(aa)	MW(Da)	pI	Chr. Location
<i>GrTCP1</i>	Gorai.001G200400.1	398	43633.3	9.3687	Chr01: 36491376–36493627
<i>GrTCP2</i>	Gorai.009G153900.1	410	45001.5	8.2869	Chr09: 11754631–11761100
<i>GrTCP3</i>	Gorai.002G064500.1	444	48127.5	6.9763	Chr02: 7559171–7562527
<i>GrTCP4</i>	Gorai.009G373000.1	401	43857.1	6.7323	Chr09: 50501688–50504320
<i>GrTCP5</i>	Gorai.008G199700.1	327	36256.9	6.1773	Chr08: 48462197–48464598
<i>GrTCP6</i>	Gorai.011G086900.1	300	31920.4	8.5291	Chr11: 9062285–9064732
<i>GrTCP7a</i>	Gorai.005G211900.1	257	26592.8	10.1465	Chr05: 59315620–59319891
<i>GrTCP7b</i>	Gorai.013G068600.1	256	26423.6	10.3339	Chr13: 7948087–7951341
<i>GrTCP7c</i>	Gorai.008G147800.1	243	25363.6	10.6333	Chr08: 40133983–40134714
<i>GrTCP8</i>	Gorai.012G166500.1	488	51182	8.0412	Chr12: 33695542–33697571
<i>GrTCP9a</i>	Gorai.007G094200.1	338	35440.4	9.539	Chr07: 6915362–6917586
<i>GrTCP9b</i>	Gorai.004G206900.1	345	36349.7	8.3417	Chr04: 53747390–53748427
<i>GrTCP10</i>	Gorai.013G172800.1	409	44213.7	7.3274	Chr13: 45973805–45976133
<i>GrTCP11</i>	Gorai.006G165300.1	270	29631.8	9.285	Chr06: 42514390–42515568
<i>GrTCP12</i>	Gorai.008G186800.1	501	55904.8	7.1546	Chr08: 46771143–46773319
<i>GrTCP13a</i>	Gorai.012G048500.1	309	34162.3	8.9496	Chr12: 6242356–6244652
<i>GrTCP13b</i>	Gorai.006G009800.1	285	32029.8	7.9301	Chr06: 2180268–2182165
<i>GrTCP14a</i>	Gorai.007G036800.1	395	42222.1	7.4496	Chr07: 2525369–2527057
<i>GrTCP14b</i>	Gorai.001G072200.1	409	43090.4	8.531	Chr01: 7309330–7311843
<i>GrTCP14c</i>	Gorai.008G192400.1	401	43363.7	7.2316	Chr08: 47573318–47575119
<i>GrTCP15a</i>	Gorai.008G181600.1	344	37542.6	9.1122	Chr08: 45925248–45927562
<i>GrTCP15b</i>	Gorai.N023400.1	365	39735.1	9.737	scaffold_505: 23–1120
<i>GrTCP15c</i>	Gorai.013G084500.1	352	38234.7	9.8567	Chr13: 12166527–12169530
<i>GrTCP16</i>	Gorai.N023500.1	216	24069.9	9.9096	scaffold_505: 5633–6280
<i>GrTCP17</i>	Gorai.001G076700.1	266	30361.1	8.054	Chr01: 7898275–7903253
<i>GrTCP18a</i>	Gorai.007G007500.1	324	37215.3	8.8024	Chr07: 600999–602878
<i>GrTCP18b</i>	Gorai.008G285300.1	361	40855.6	8.2647	Chr08: 56105392–56106838
<i>GrTCP19a</i>	Gorai.006G197000.1	337	36419.8	7.118	Chr06: 45427588–45429051
<i>GrTCP19b</i>	Gorai.008G243000.1	388	41371.5	8.7407	Chr08: 52845411–52846971
<i>GrTCP20a</i>	Gorai.012G084600.1	300	31763.5	9.1033	Chr12: 14283464–14287212
<i>GrTCP20b</i>	Gorai.006G043800.1	298	31463.1	9.9629	Chr06: 13026618–13028750
<i>GrTCP20c</i>	Gorai.001G273300.1	334	35872.1	10.08	Chr01: 55023721–55025665
<i>GrTCP20d</i>	Gorai.008G157300.1	298	31748.5	10.3948	Chr08: 42005542–42007175
<i>GrTCP21</i>	Gorai.013G268200.1	196	21060.7	8.746	Chr13: 58033580–58034170
<i>GrTCP22</i>	Gorai.002G215000.1	549	57867.3	7.2398	Chr02: 56433132–56435832
<i>GrTCP23</i>	Gorai.009G289000.1	421	44684.3	7.4635	Chr09: 24825613–24826878
<i>GrTCP24</i>	Gorai.011G046000.1	463	50079.8	7.5515	Chr11: 3523288–3528930
<i>GrTCP25</i>	Gorai.009G398700.1	435	48308	6.4093	Chr09: 56724136–56726696



Phylogenetic analysis. To get a better understanding of the evolutionary history and phylogenetic relationships of TCP transcription factor family in *G. raimondii*, an unrooted phylogenetic tree was constructed with Neighbor-Joining method on the basis of multiple sequence alignment of 38 *G. raimondii* TCP protein sequences with all TCP sequences from Arabidopsis and rice, including 24 Arabidopsis TCP protein sequences and 22 rice TCP protein sequences (Figure 1). The bootstrap values for some nodes of the NJ tree were low as a result of relatively large number of sequences, which was also shown in previous reports^{9,11}. Therefore, we sought other evidence to verify the reliability of our phylogenetic tree. The phylogenetic trees of TCP transcription family were reconstructed with Maximum likelihood, Minimal Evolution and PhyML methods. The trees produced by the three methods mentioned above were almost identical with only minor differences at some branches, suggesting that the four methods were largely consistent with each other. Besides, the analysis of gene structure, conserved motif structure and expression profiles were also used to confirm the validity of the phylogenetic tree. Considering the great similarity among these tree topologies as well as previous studies^{9,11}, the NJ tree was employed for further analysis.

According to the NJ phylogenetic tree (Figure 1), the TCP transcription factor family was divided into eleven subgroups designated Group A to Group K. According to their sequence features within and outside the TCP domain, GrTCPs in Group A, B, C, D, E, F, and G belong to Class I subfamily while GrTCPs in the rest groups belong to Class II subfamily⁸. Group A, the largest clade among all subgroups, contained 12 members, representing 14.3% of the total TCP genes; Group E constituted the smallest clade, containing 3 members. Generally speaking, the TCP genes showed an interspersed distribution in most clades, indicating that the TCP family expanded before the divergence of the lineages. Additionally, the TCP genes were not evenly distributed in some clades in *G. raimondii*, Arabidopsis and rice. Many Arabidopsis TCP genes had two or more counterparts in *G. raimondii*, suggesting that GrTCP genes duplicated after the divergence of *G. raimondii* and Arabidopsis. For example, Group A contained seven *G. raimondii* TCPs but there were only three Arabidopsis members; Group D contained four *G. raimondii* TCPs but there were only two Arabidopsis TCPs. Specifically, Group F contained five *G. raimondii* TCPs but there was only one Arabidopsis TCP and no TCP gene was identified in rice genome, which implied that this group was either acquired after the divergence of monocots and dicots or lost in rice. By comparison,

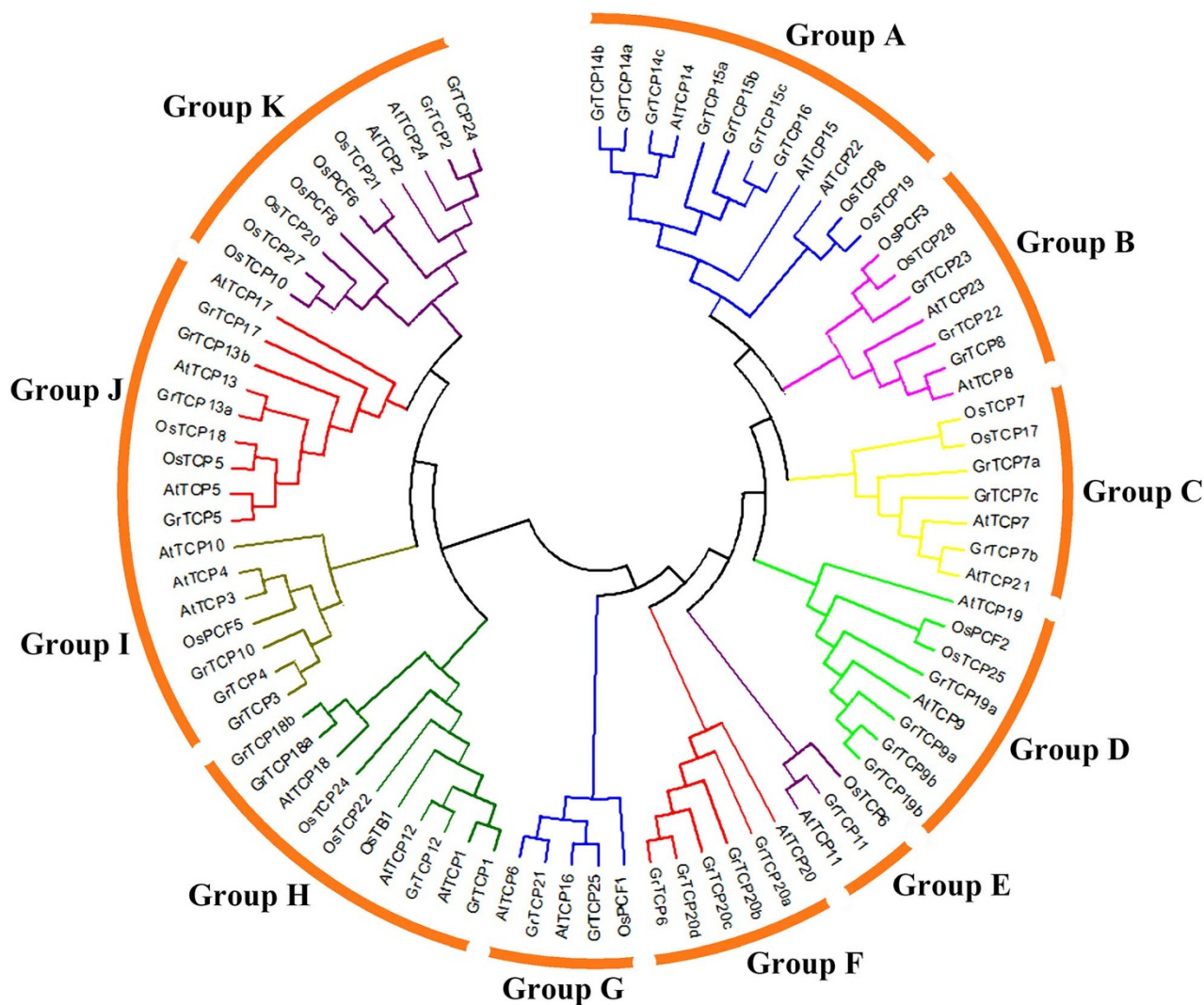


Figure 1 | Phylogenetic relationships of TCP transcription factors from *Gossypium raimondii*, Arabidopsis and rice. The unrooted phylogenetic tree was constructed using MEGA 6.0 by Neighbor-Joining method and the bootstrap test was performed with 1,000 iterations. The eleven subclades are indicated with different colors.



the rice TCP genes were overrepresented in Group K, which contained six rice TCP genes, while there were only two *G. raimondii* members and two Arabidopsis members in the same group. Some groups had almost equal number of TCP genes in the three species, such as Group B, C, and E (Figure 1).

Many Arabidopsis TCP genes with similar functions tended to cluster into the same clade, which may imply that TCP genes within the same clade had similar functions in *G. raimondii*. For example, all AtTCPs in Group A and Group B (AtTCP8, AtTCP14, AtTCP15, AtTCP22, AtTCP23), which clustered with ten GrTCPs (Figure 1), play an important role in the regulation of leaf development by modulating gene networks involved in cell-cycle control and shoot apical meristem (SAM) maintenance²⁷. AtTCPs in Group H function in the process of lateral branching which determines shoot architecture^{28,29}. All AtTCP genes in Group I and Group K (AtTCP2, AtTCP3, AtTCP4, AtTCP10, AtTCP24), which cluster with five GrTCPs (Figure 1), are down regulated by miRNA319 and act as negative cell proliferation factors in the regulation of leaf margins³⁰.

Chromosomal location and gene duplication. To determine the chromosomal distribution of the TCP genes in *G. raimondii*, the physical locations of all GrTCP genes on chromosomes were obtained through BLASTN searches against *G. raimondii* genome database in Phytozome (<http://www.phytozome.net/cotton.php>). Among the 38 GrTCP genes, a total of 36 genes were distributed across 11 out of the 13 *G. raimondii* chromosomes, while the rest two (*GrTCP15b* and *GrTCP16*) were anchored on unmapped scaffolds (Figure 2). Generally speaking, the number of GrTCP genes on each chromosome appeared to be uneven, ranging widely from 0 to 8 genes per chromosome. For example, chromosome 8 contained the highest number of 8 GrTCPs, accounting for 21.1% of the total GrTCP genes, followed by 4 GrTCPs on each of chromosome 1, 6, 9 and 13, whereas relatively low number of GrTCP genes were found on several chromosomes, including 2 genes on each of chromosome 2 and 11, and one gene on each of chromosome 4 and 5. By contrast, GrTCP genes were not observed on chromosome 3 and 10 (Figure 2).

Given the importance of gene duplication in the amplification of gene families, potential duplication events involved in the evolution of *G. raimondii* genome were analyzed to shed light on the

mechanism behind the expansion of the GrTCP gene family. On the basis of protein sequence identities, 19 pairs of putative paralogous GrTCP genes were identified, accounting for more than 70% of the entire GrTCP gene family and thereby supporting the hypothesis that putative gene duplication events are the main causes of the expansion of the GrTCP gene family. These gene pairs are in the same clade of the phylogenetic tree with high degree of protein sequence identities. For instance, the sequence of GrTCP15a covers 100% of that of GrTCP15b after alignment and the identity of aligned region is 98%, while the protein sequence identity of GrTCP7a and GrTCP7b is 88%. Among these paralogous gene pairs, 18 pairs are located on different chromosomes, suggesting a high number of segmental duplication events, whereas no traceable duplication events could be determined for another gene pair because one gene of this pair was anchored on unmapped scaffolds (Figure 2). Interestingly, six genes (*GrTCP3*, *GrTCP6*, *GrTCP7a*, *GrTCP7b*, *GrTCP7c* and *GrTCP20c*) participated in two segmental duplication events (e.g. *GrTCP20a/GrTCP20b/GrTCP20c*, *GrTCP7a/GrTCP7b/GrTCP7c* and *GrTCP3/GrTCP4/GrTCP10*). In contrast, no tandem duplication events were observed in these duplicated pairs (Figure 2).

In the present study, we further calculated the approximate dates of duplication events with the DnaSP program. The results showed that segmental duplications of GrTCP genes occurred between 11.28 Mya (million years ago) to 36.51 Mya, with an average of 19.83 Mya.

Gene structure and conserved motifs. With the aim to gain further insights into the evolutionary relationships among GrTCP genes, we investigated the exon/intron structures of individual GrTCP genes by alignment of cDNA sequences and corresponding genomic DNA sequences. As illustrated in Figure 3b, 32 out of 38 GrTCP genes had no intron, while the other GrTCP genes possess one intron, with the exception of *GrTCP25* containing four introns. Additionally, an unrooted phylogenetic tree was constructed with GrTCP protein sequences to determine if the exon/intron organization of GrTCP genes is consistent with the phylogenetic subfamilies (Figure 3a). As expected, most GrTCP genes within the same subfamily demonstrated very similar exon/intron distribution patterns in terms of exon length and intron number. For example, most GrTCP gene in subfamily A, B, C, D and K had only one exon of

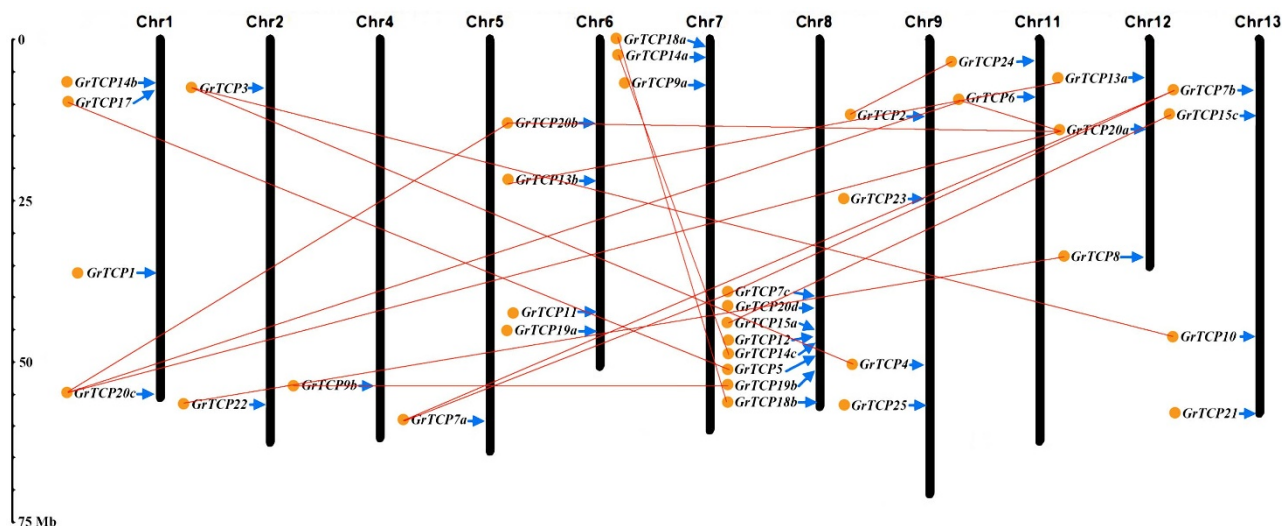


Figure 2 | Chromosomal distribution and gene duplication of TCP genes in *G. raimondii*. The scale is in megabases (Mb). The chromosome numbers are indicated at the top of each chromosome. The paralogous TCP genes are connected with a red line.

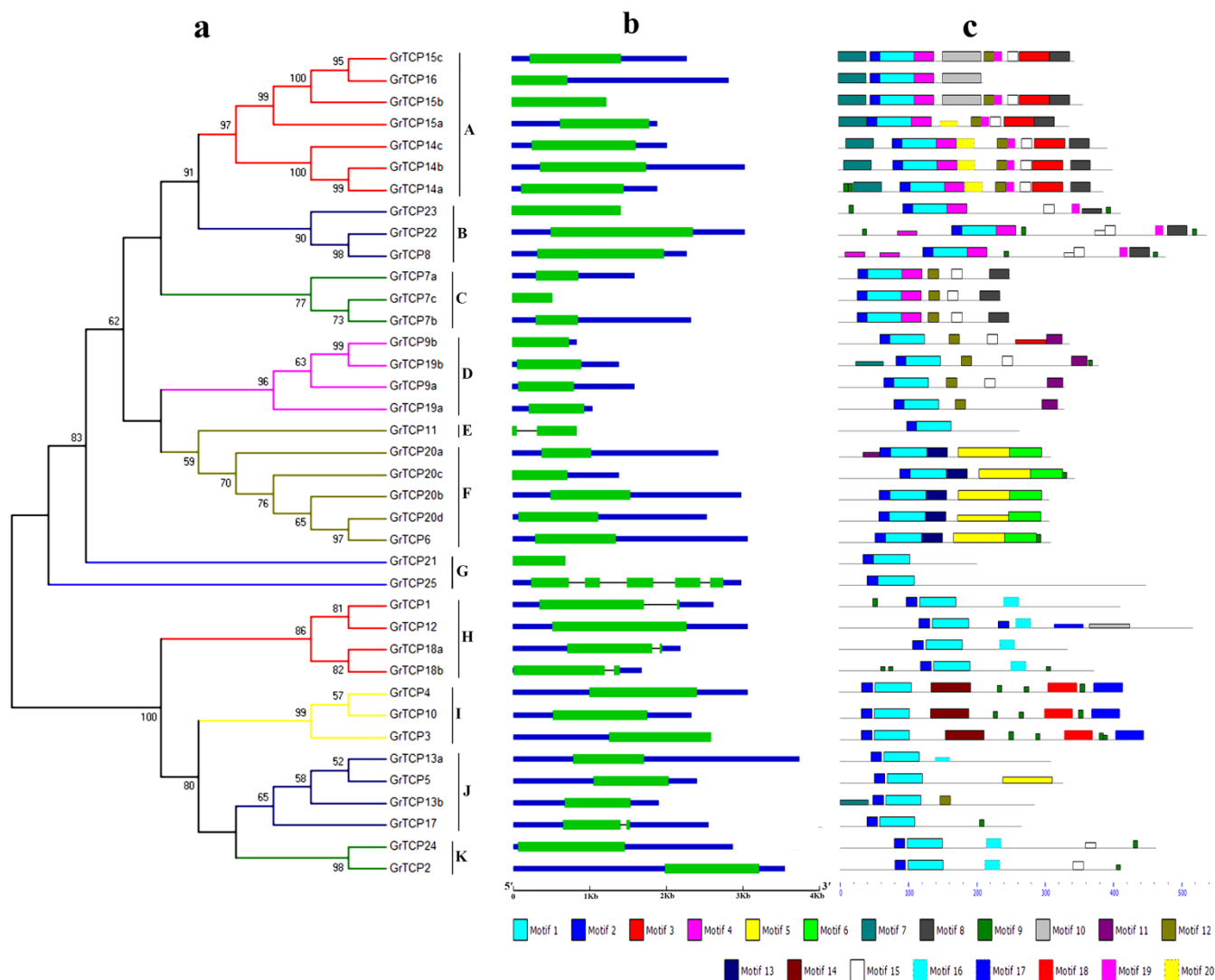


Figure 3 | Phylogenetic analysis, gene structure and conserved motifs of TCP family in *Gossypium raimondii*. (a). The phylogenetic tree of all TCP transcription factors in *G. raimondii* was constructed using Neighbor-Joining method and the bootstrap test was performed with 1,000 iterations. Bootstrap values higher than 50% support are displayed. (b). The exon/intron organization of TCP genes of *G. raimondii*. The blue lines represent 5'-UTR or 3'-UTR, green boxes represent exons and black lines indicate introns. (c). The conserved protein motifs in the TCP family were identified using MEME program. Each motif is indicated with a specific color.

similar length without intron, whereas members within subfamily H contain one intron, except for *GrTCP12*, which possesses no intron. By comparison, *GrTCP* genes in subfamily G showed great variability in exon length and intron number (Figure 3a and 3b).

We further searched for the conserved motifs in *GrTCP* proteins by MEME program to obtain more insights into the diversity of motif compositions among *GrTCP*s. As shown in Figure 3c, a total of 20 conserved motifs designated as motif 1 to motif 20 were identified. Most of *GrTCP* proteins within the same subfamily shared similar motif compositions while high divergence was observed among different subfamilies, implying that the *GrTCP* members within the same subfamily may perform similar functions and that some motifs may play an important role in the subfamily-specific functions. For example, all *GrTCP*s in subfamily C possess motif 1, 2, 4, 8, 12 and 15 while all members in subfamily I contain motif 1, 2, 9, 14, 17 and 18 (Figure 3c). In addition, some motifs were exclusively present in a particular subfamily, suggesting that these motifs may contribute to the specific function of that subfamily. For instance, motif 20 for subfamily A, motif 6 for subfamily F and motif 14, 17 and 18 for subfamily I (Figure 3c). Moreover, the program ScanProsite was employed to annotate the identified 20 motifs. However, few motifs

hit for PROSITE (release 20.103) motifs in the database. Therefore, the functions of most motifs are still left unknown. The only motif that matched to protein sequences in the ScanProsite database was motif 1, which was annotated as the conserved TCP domain and was uniformly observed in all *GrTCP* proteins. Generally speaking, the consistency of the motif compositions of *GrTCP* proteins as well as the exon/intron structures of most *GrTCP* genes with the phylogenetic subfamilies further supported the close evolutionary relationships among *GrTCP*s as well as the reliability of our phylogenetic analysis.

Expression profiles of TCP genes in *G. raimondii*. To investigate the tissue-specific expression profiles of TCP genes in *G. raimondii*, the quantitative real time PCR (qRT-PCR) was performed for different organs, including leaf, flower bud, shoot and sepal. As indicated in Figure 4 and Figure 5, some *GrTCP* genes were differentially expressed in the four tissues tested while other *GrTCP* genes showed similar expression patterns in different tissues, which may indicate functional divergence of *GrTCP* genes during plant development. For example, *GrTCP2*, *GrTCP3*, *GrTCP13b*, *GrTCP15c*, *GrTCP19a* and *GrTCP23* were constitutively

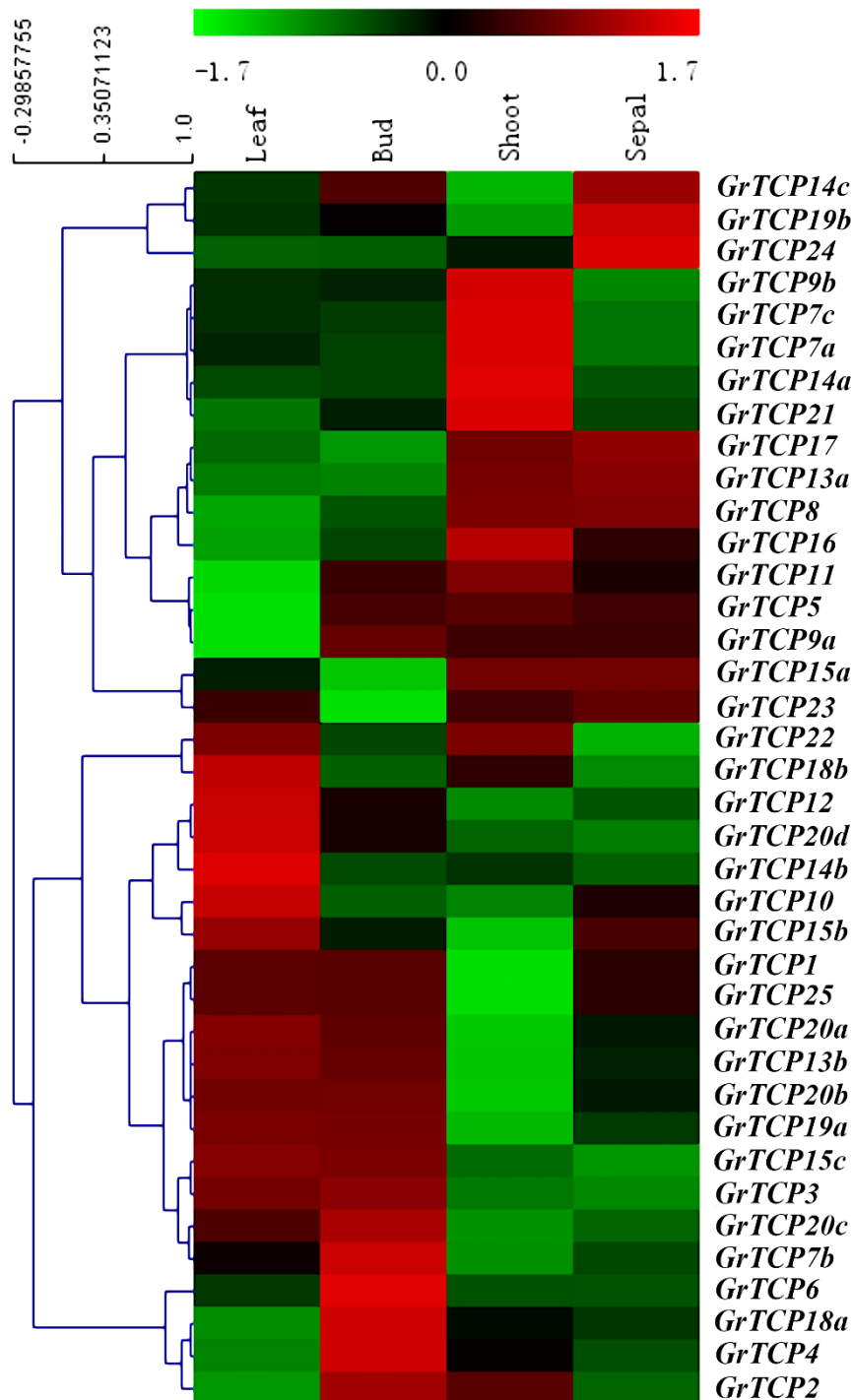


Figure 4 | Heatmap representation for expression patterns of *G. raimondii* TCP genes across different tissues. The expression profile data of GrTCP genes in leaf, bud, shoot and sepal were obtained through quantitative real-time PCR.

expressed in every tissue tested at very high level, implying that these genes may play regulatory roles at multiple development stages, whereas *GrTCP6*, *GrTCP9a*, *GrTCP20a* and *GrTCP24* were expressed at very low level in all tissues examined, which suggested that they may be primarily expressed in other organs not tested or under special conditions (Figure 5). In contrast, the expression levels of *GrTCP20a*, *GrTCP20b* and *GrTCP20c* were very high in leaf and bud and were relatively low in shoot and sepal, indicating that they may play an important role in the development of leaf and bud. A similar expression profile was also found for *GrTCP1* and *GrTCP25*. In addition, some genes were exclusively highly expressed in a specific

tissue. For example, *GrTCP10*, *GrTCP12*, *GrTCP14b*, *GrTCP18b* and *GrTCP20d* were relatively highly expressed in leaf while *GrTCP7a*, *GrTCP7c*, *GrTCP9b* and *GrTCP14a* were exclusively expressed in shoot at very high level, implying their specific roles in the corresponding tissues (Figure 4). In general, the GrTCP genes that are highly expressed in specific tissues may be involved in the regulation of plant development. For instance, *GrTCP15c* were relatively highly expressed in leaf and bud, suggesting that it may play a role in the development of the two tissues. According to previous studies, *AtTCP15*, the homologous counterparts of *GrTCP15c*, is involved in the regulation of leaf development³¹,

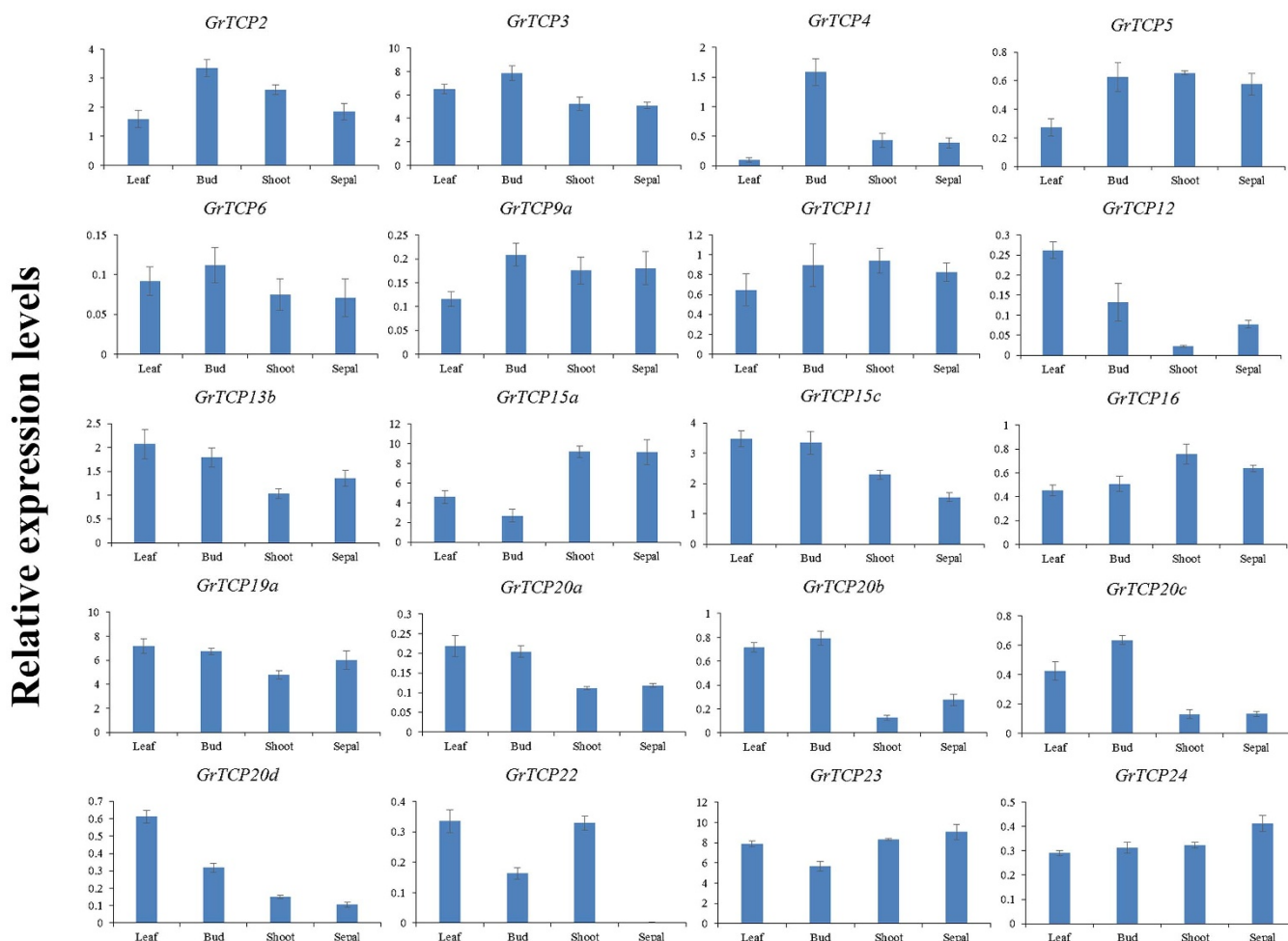


Figure 5 | Expression profiles of 20 GrTCP genes across different tissues. The y-axis represent the relative expression levels of GrTCP genes against reference gene *TuA11*. Error bars indicate standard deviation for three replicates.

which supported our hypothesis. However, further studies are still needed to unravel the divergent roles of GrTCP genes.

Discussion

TCP transcription factors play important roles in plants. TCP transcription factors are a class of plant-specific transcription

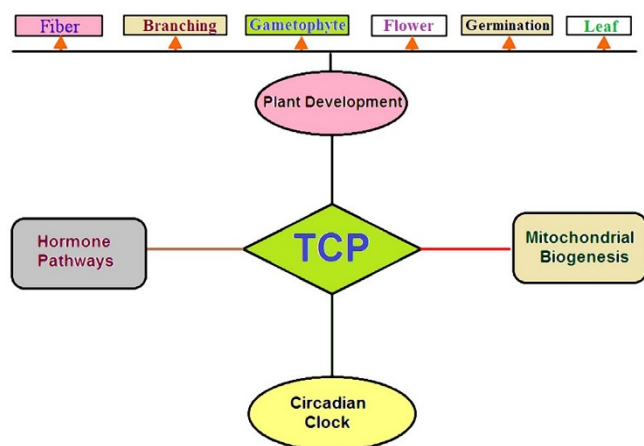


Figure 6 | TCP transcription factors play an important role in the multiple biological process during plant growth and development.

factors, which play an versatile function in multiple biological processes during plant growth and development (Figure 6). It has been reported that many TCP transcription factors participate in the regulation of multiple aspects of plant development, such as gametophyte development^{32–34}, hormone signal transduction^{29,35,36}, mitochondrial biogenesis³⁷, regulation of the circadian clock^{38,39}, lateral branching^{28,29,40}, flower development^{31,41,42}, seed germination^{43,44} and leaf development^{14,31}. Class II TCP members have been found to function in a similar manner mainly by preventing plant growth and cell proliferation based on the mutation studies of multiple members in this subfamily^{9,30,40,45–49}, whereas the predicted role of class I members seems to promote plant growth and cell proliferation^{10,50}. In Arabidopsis, mutation in *AtTCP18* (*BRC1*) gene led to a significant increase in the number of rosette branches while up-regulation of *AtTCP18* resulted in the inhibition of lateral bud outgrowth²⁹. *AtTCP4* has been shown to influence early embryo development and recent evidence revealed that pollen grains produced by transgenic Arabidopsis line expressing hyper-activated *AtTCP4* genes cannot yield viable seeds, indicating that *AtTCP4* may regulate plant reproduction^{32,33}. Functional analysis of *AtTCP1* showed that *AtTCP1* is involved in the regulation of Brassinosteroid hormone signaling pathway by positively controlling the expression of a key enzyme *DWARF4*³⁵. In a recent study, *AtTCP8* was proposed to be associated with mitochondrial biogenesis based on the evidence that *AtTCP8* is able to bind to the promoter region of *PNMI*, a gene encoding a newly identified



pentatricopeptide repeat protein that function in the mitochondrial gene expression³⁷. Yeast two-hybrid assays revealed the interaction between some TCP transcription factors (AtTCP2, AtTCP3, AtTCP11 and AtTCP15) and several regulatory components of the circadian clock, suggesting that TCP proteins may control or influence the circadian networks³⁸. *AtTCP14* and *AtTCP15* were reported to regulate floral organ development and reduced expression of the two transcription factors resulted in phenotypic abnormalities in the three outer whorls and the gynoecia^{31,41}. In addition, *AtTCP14* and *AtTCP15* were also found to regulate leaf development: mutant *AtTCP14* and *AtTCP15* led to broader leaves towards the base and shorter petioles than the wild type³¹.

TCP transcription factors were widely existed in cotton. In the present study, a total of 38 TCP genes were identified from *G. raimondii*; the number of TCP genes in *G. raimondii* was higher than that in *Arabidopsis* (24) and in rice (22)⁹. The number of TCP genes in *G. raimondii* is approximately 1.58 times that in *Arabidopsis*, which is in strong agreement with the fact that the protein coding genes in *G. raimondii* genome (40,976 genes) is about 1.6 times that in *Arabidopsis* (25,498 genes)^{5,51}. It is found that many TCP genes in *Arabidopsis* have two or more counterparts in *G. raimondii*, suggesting that the expansion of TCP family in *G. raimondii* may be caused by genome duplication events such as segmental duplication, tandem duplication and transposition events.

The high number of TCP genes in *G. raimondii* was more likely caused by gene duplication. Gene duplication, an outstanding feature of genomic architecture, plays a significant role in the process of plant genomic and organismal evolution, generating raw genetic material necessary for mutation, genetic drift and selection, and contributing to the origin of new gene functions and the evolution of gene networks^{52,53}. It has been demonstrated that the expansion of gene families is mainly attributed to gene duplication events on various scales, including tandem duplication, segmental duplication, transposition events and whole-genome duplication^{52,53}. Our results indicate that segmental duplication is a predominant duplication event for TCP genes and the major contributor to the expansion of TCP gene family in *G. raimondii*. It has been reported that *G. raimondii* genome has undergone at least two rounds of genome-wide duplication, an ancient paleohexaploidization event at approximately 130.8 Mya and a recent whole-genome duplication event at around 13.3–20.0 Mya⁵. The average duplication date of GrTCP genes is very close to the recent whole-genome duplication date of *G. raimondii*, suggesting that large-scale genome duplication events may also contribute to the expansion of GrTCP family. In addition, according to a recent study, the split of *G. raimondii*/*Arabidopsis* and *G. raimondii*/*T. cacao* occurred at approximately 82.3 million years ago and 33.7 million years ago, respectively⁵. Since the duplication of GrTCPs originated from 11.28 to 36.51 million years ago, most of the GrTCP genes duplicated after the divergence of *G. raimondii*/*Arabidopsis* and *G. raimondii*/*T. cacao*. Summarizedly, our results indicate that both segmental duplication and whole-genome duplication contribute to the expandedness of TCP family in *G. raimondii*. This duplication may contribute to the unique functions of TCP transcription factors in cotton, for example controlling cotton fiber initiation and development.

Evolutionary conservation and divergence of the TCP family in *G. raimondii*. According to recent studies, *G. raimondii* and *Arabidopsis* diverged from a common ancestor at approximately 82.3 million years ago, which was followed by paleopolyploidy events in both species, contributing to evolutionary innovation^{5,54}. Our studies showed that many TCP genes in *Arabidopsis* have two or more counterparts in *G. raimondii* with high protein sequence similarity, implying that the TCP genes may undergo differential expansion in *G. raimondii* and *Arabidopsis*. Through comparison

of the exon/intron organization of individual TCP family members in the two species, we observed that their gene structures exhibited high similarities (Figure 3b and Supplementary Fig. S1 and 2). Among ten pairs of TCP genes with high protein sequence identity and query coverage, seven pairs exhibited conserved gene structure in terms of exon length and intron numbers. While most pairs showed similar exon/intron structure, a few displayed some degree of divergence. For example, *GrTCP12* contained one exon without intron, whereas its counterpart in *Arabidopsis*, *AtTCP12*, possessed one exon and one intron. Such variation may be caused by single intron loss or gain during the process of structural evolution. In addition, we further analyzed the gene structure and conserved motifs of paralogous pairs of TCPs in *G. raimondii* to shed light on the diversification of TCP genes. Of the 19 paralogous pairs, 14 pairs of GrTCPs shared conserved exon/intron organization (Figure 3b). Similar to the gene structure, most paralogous pairs of GrTCPs also exhibited conserved motif composition, with only several unique motifs observed in some GrTCP members, such as motif 20 for GrTCP15a, motif 3 for GrTCP9b and motif 9 for GrTCP18b and GrTCP14a (Figure 3c). These specific motifs may contribute to the neofunctionalization (in which one paralogous member obtains a new function after gene duplication), or subfunctionalization (where each paralog retains part of its original ancestral function) of duplicated genes by a series of synonymous and/or non-synonymous mutation during evolution⁵⁵. In all, the majority of GrTCPs are evolutionarily conserved, while the variation of exon/intron distribution and motif composition in certain paralogous pairs of GrTCPs suggests that some members of TCP family in *G. raimondii* are functionally diversified through differential expansion. The diversification of GrTCPs may contribute to the unique functions of TCP transcription factors in cotton, for example controlling cotton fiber initiation and development. It has been reported that cotton transcription factor TCP14 was predominantly in cotton fiber cell, particularly at the stage of cotton fiber initiation and elongation⁵⁶. Another study demonstrated that cotton TCP transcription factor regulated fiber and root hair development by regulating jasmonic acid biosynthesis and response as well ethylene signaling pathway¹⁶.

Methods

Sequence retrieval for TCP proteins. The conserved TCP DNA-binding domain based on Hidden Markov Model (HMM) (PF03634) was obtained from Pfam protein family database (<http://pfam.sanger.ac.uk/>). In order to identify the TCP transcription factor coding genes of *G. raimondii*, the HMM profile of TCP domain was subsequently employed as query to perform a HMMER search (<http://hmmer.janelia.org/>) against the *G. raimondii* genome derived from Phytozome (<http://www.phytozome.net/>) (E-value = 0.01). All redundant sequences were discarded from further analysis based on cluster W¹⁷ alignment results, sequence identification numbers and chromosome location. Furthermore, to verify the reliability of the initial results, all non-redundant candidate TCP sequences were analyzed to confirm the presence of the conserved TCP domain using the InterProScan program¹⁸. The sequences of TCP family members in the genome of *Arabidopsis* and *Oryza sativa* were retrieved from PlantTFDB plant transcription factor database (<http://planttfdb.cbi.pku.edu.cn/>, v3.0).

Phylogenetic analysis. Multiple sequence alignments were conducted on the amino acid sequences of TCP proteins in *G. raimondii*, *Arabidopsis* and rice genomes using Cluster X¹⁹ with default settings. Subsequently, MEGA 6.0 software²⁰ was employed to construct an unrooted phylogenetic tree based on alignments using the Neighbor-Joining (NJ) method with the following parameters: JTT model, pairwise gap deletion and 1,000 bootstraps. Furthermore, Maximum likelihood, Minimal Evolution and PhyML methods were also applied in the tree construction to validate the results from the NJ method. Additionally, a separate phylogenetic tree was constructed with all the TCP protein sequences in *G. raimondii* for further analysis.

Analysis of chromosomal location and gene duplication. Information about the physical locations of all GrTCP genes on chromosomes was obtained through BLASTN searches against *G. raimondii* genome database in Phytozome (<http://www.phytozome.net/cotton.php>). All GrTCP genes were then mapped on the chromosome using the software MapInspect. The detection of TCP gene duplication events was also carried out. Paralogous TCP gene pairs in *G. raimondii* were identified on the basis of alignment results. The criteria described in previous studies^{21,22} were



adopted: the shorter sequences covers over 70% of the longer sequence after alignment and the minimum identity of aligned regions is 70%. In addition, to further analyze gene duplication events, the synonymous substitution rate (Ks) and non-synonymous substitution rate (Ka) were calculated using the software DnaSP²³. The date of duplication events was subsequently estimated according to the equation $T = Ks/2\lambda$. The approximate value for clock-like rate (λ) was 1.5 synonymous substitutions per 10^8 years for *G. raimondii*⁵.

Gene structure analysis and conserved motif identification. The TCP genomic sequences and CDS sequences extracted from Phytozome were compared in gene structure display server program²⁴ to infer the exon/intron organization of TCP genes. TCP protein sequences in *G. raimondii* were submitted to online Multiple Expectation maximization for Motif Elicitation (MEME) program²⁵ for identification of conserved protein motifs. The optimized MEME parameters were as follows: a number of repetitions, the optimum width from 6 to 250 and maximum number of motifs-20. The identified protein motifs were further annotated with ScanProsite²⁶.

RNA isolation and Real-time quantitative RT-PCR analysis. To detect the expression profiles of TCP genes in *G. raimondii*, total RNA was extracted from plant leaves, buds, shoots and sepals using the mirVana™ miRNA Isolation Kit (Ambion, Austin, TX, USA), according to the manufacturer's instructions. The total RNA quantity and purity were assessed by a NanoDrop ND-1000 Spectrophotometer (NanoDrop Technologies, Wilmington, DE, USA). 1 μ g of total RNA isolated from each tissue was reverse transcribed into cDNA using the TaqMan® MicroRNA Reverse Transcription Kit (Applied Biosystems, Foster City, CA, USA) and a poly-T primer. The real-time RT-PCR was then performed with a 7300 Real-Time PCR System (Applied Biosystems, Foster City, CA) according to the supplier's protocols. Each reaction mixture contains 2 μ L of DNase/RNase free water, 5 μ L Real-Time SYBR Green PCR master mix, 1 μ L diluted cDNA product from reverse transcription PCR reaction and 2 μ L gene-specific primers. Three biological replicates were conducted for each tissue and each biological replicate was technically repeated three times. The thermal cycle applied was as follows: 95°C for 10 min followed by 45 cycles of denature at 95°C for 15 s and annealing and elongation at 60°C for 60 s. The expression values of TCP genes tested were normalized with an internal reference gene *TuA11*. The relative expression levels (R) was calculated using the following equation: $R = 2^{-C_t(C_1 - C_2)}$, where C_1 stands for the C_t value of TCP genes while C_2 is the C_t value of the reference gene. A heatmap for gene expression patterns was generated with the software MultiExperiment Viewer (MeV).

- Calkhoven, C. F. & Ab, G. Multiple steps in the regulation of transcription-factor level and activity. *Biochem. J.* **317** (Pt 2), 329–342 (1996).
- Latchman, D. S. Transcription factors: an overview. *Int. J. Biochem. Cell Biol.* **29**, 1305–1312 (1997).
- Schwechheimer, C., Zourelidou, M. & Bevan, M. W. Plant transcription factor studies. *Annu. Rev. Plant Phys.* **49**, 127–150 (1998).
- Riechmann, J. L. *et al.* Arabidopsis transcription factors: Genome-wide comparative analysis among eukaryotes. *Science* **290**, 2105–2110 (2000).
- Wang, K. *et al.* The draft genome of a diploid cotton *Gossypium raimondii*. *Nat. Genet.* **44**, 1098–1103, doi:10.1038/ng.2371 (2012).
- Wray, G. A. *et al.* The evolution of transcriptional regulation in eukaryotes. *Mol. Biol. Evol.* **20**, 1377–1419, doi:10.1093/molbev/msg140 (2003).
- Perez-Rodriguez, P. *et al.* PlnTFDB: updated content and new features of the plant transcription factor database. *Nucleic Acids Res.* **38**, D822–827, doi:10.1093/nar/gkp805 (2010).
- Cubas, P., Lauter, N., Doebley, J. & Coen, E. The TCP domain: a motif found in proteins regulating plant growth and development. *Plant J.* **18**, 215–222 (1999).
- Martin-Trillo, M. & Cubas, P. TCP genes: a family snapshot ten years later. *Trends Plant Sci.* **15**, 31–39, doi:10.1016/j.tplants.2009.11.003 (2010).
- Kosugi, S. & Ohashi, Y. PCF1 and PCF2 specifically bind to cis elements in the rice proliferating cell nuclear antigen gene. *Plant Cell* **9**, 1607–1619 (1997).
- Yao, X., Ma, H., Wang, J. & Zhang, D. Genome-Wide Comparative Analysis and Expression Pattern of TCP Gene Families in Arabidopsis thaliana and Oryza sativa. *J. Integr. Plant Biol.* **49**, 885–897, doi:10.1111/j.1744-7909.2007.00509.x (2007).
- Lupas, A., Vandyke, M. & Stock, J. Predicting Coiled Coils from Protein Sequences. *Science* **252**, 1162–1164 (1991).
- Kosugi, S. & Ohashi, Y. DNA binding and dimerization specificity and potential targets for the TCP protein family. *Plant J.* **30**, 337–348 (2002).
- Viola, I. L., Uberti Manassero, N. G., Ripoll, R. & Gonzalez, D. H. The Arabidopsis class I TCP transcription factor AtTCP11 is a developmental regulator with distinct DNA-binding properties due to the presence of a threonine residue at position 15 of the TCP domain. *Biochem. J.* **435**, 143–155, doi:10.1042/BJ20101019 (2011).
- Schommer, C. *et al.* Control of jasmonate biosynthesis and senescence by miR319 targets. *PLoS Biol.* **6**, e230, doi:10.1371/journal.pbio.0060230 (2008).
- Hao, J. *et al.* GbTCP, a cotton TCP transcription factor, confers fibre elongation and root hair development by a complex regulating system. *J. Exp. Bot.* **63**, 6267–6281, doi:10.1093/jxb/ers278 (2012).
- Thompson, J. D., Higgins, D. G. & Gibson, T. J. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence

weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22**, 4673–4680 (1994).

- Quevillon, E. *et al.* InterProScan: protein domains identifier. *Nucleic Acids Res.* **33**, W116–120, doi:10.1093/nar/gki442 (2005).
- Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F. & Higgins, D. G. The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.* **25**, 4876–4882 (1997).
- Tamura, K., Stecher, G., Peterson, D., Filipowski, A. & Kumar, S. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Mol. Biol. Evol.* **30**, 2725–2729, doi:10.1093/molbev/mst197 (2013).
- Yang, S., Zhang, X., Yue, J. X., Tian, D. & Chen, J. Q. Recent duplications dominate NBS-encoding gene expansion in two woody species. *Mol. Genet. Genomics.* **280**, 187–198, doi:10.1007/s00438-008-0355-0 (2008).
- Gu, Z., Cavalcanti, A., Chen, F. C., Bouman, P. & Li, W. H. Extent of gene duplication in the genomes of *Drosophila*, nematode, and yeast. *Mol. Biol. Evol.* **19**, 256–262 (2002).
- Librado, P. & Rozas, J. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* **25**, 1451–1452, doi:10.1093/bioinformatics/btp187 (2009).
- Guo, A. Y., Zhu, Q. H., Chen, X. & Luo, J. C. [GSDS: a gene structure display server]. *Yi Chuan* **29**, 1023–1026 (2007).
- Bailey, T. L., Williams, N., Misleh, C. & Li, W. W. MEME: discovering and analyzing DNA and protein sequence motifs. *Nucleic Acids Res.* **34**, W369–373, doi:10.1093/nar/gkl198 (2006).
- de Castro, E. *et al.* ScanProsite: detection of PROSITE signature matches and ProRule-associated functional and structural residues in proteins. *Nucleic Acids Res.* **34**, W362–365, doi:10.1093/nar/gkl124 (2006).
- Aguiar-Martinez, J. A. & Sinha, N. Analysis of the role of Arabidopsis class I TCP genes AtTCP7, AtTCP8, AtTCP22, and AtTCP23 in leaf development. *Front. Plant Sci.* **4**, 406, doi:10.3389/fpls.2013.00406 (2013).
- Takeda, T. *et al.* The OsTB1 gene negatively regulates lateral branching in rice. *Plant J.* **33**, 513–520 (2003).
- Aguiar-Martinez, J. A., Poza-Carrion, C. & Cubas, P. Arabidopsis BRANCHED1 acts as an integrator of branching signals within axillary buds. *Plant Cell* **19**, 458–472, doi:10.1105/tpc.106.048934 (2007).
- Palatnik, J. F. *et al.* Control of leaf morphogenesis by microRNAs. *Nature* **425**, 257–263, doi:10.1038/nature01958 (2003).
- Kieffer, M., Master, V., Waites, R. & Davies, B. TCP14 and TCP15 affect internode length and leaf shape in Arabidopsis. *Plant J.* **68**, 147–158, doi:10.1111/j.1365-313X.2011.04674.x (2011).
- Pagnussat, G. C. *et al.* Genetic and molecular identification of genes required for female gametophyte development and function in Arabidopsis. *Development* **132**, 603–614 (2005).
- Sarvepalli, K. & Nath, U. Hyper-activation of the TCP4 transcription factor in Arabidopsis thaliana accelerates multiple aspects of plant maturation. *Plant J.* **67**, 595–607 (2011).
- Takeda, T. *et al.* RNA interference of the Arabidopsis putative transcription factor TCP16 gene results in abortion of early pollen development. *Plant Mol. Biol.* **61**, 165–177 (2006).
- Guo, Z. X. *et al.* TCP1 Modulates Brassinosteroid Biosynthesis by Regulating the Expression of the Key Biosynthetic Gene DWARF4 in Arabidopsis thaliana. *Plant Cell* **22**, 1161–1173 (2010).
- Yanai, O., Shani, E., Russ, D. & Ori, N. Gibberellin partly mediates LANCEOLATE activity in tomato. *Plant J.* **68**, 571–582 (2011).
- Hammani, K. *et al.* An Arabidopsis Dual-Localized Pentatricopeptide Repeat Protein Interacts with Nuclear Proteins Involved in Gene Expression Regulation. *Plant Cell* **23**, 730–740 (2011).
- Giraud, E. *et al.* TCP Transcription Factors Link the Regulation of Genes Encoding Mitochondrial Proteins with the Circadian Clock in Arabidopsis thaliana. *Plant Cell* **22**, 3921–3934 (2010).
- Pruneda-Paz, J. L., Breton, G., Para, A. & Kay, S. A. A Functional Genomics Approach Reveals CHE as a Component of the Arabidopsis Circadian Clock. *Science* **323**, 1481–1485 (2009).
- Hubbard, L., McSteen, P., Doebley, J. & Hake, S. Expression patterns and mutant phenotype of teosinte branched1 correlate with growth suppression in maize and teosinte. *Genetics* **162**, 1927–1935 (2002).
- Uberti-Manassero, N. G., Lucero, L. E., Viola, I. L., Vegetti, A. C. & Gonzalez, D. H. The class I protein AtTCP15 modulates plant development through a pathway that overlaps with the one affected by CIN-like TCP proteins. *J. Exp. Bot.* **63**, 809–823, doi:10.1093/jxb/err305 (2012).
- Koyama, T., Ohme-Takagi, M. & Sato, F. Generation of serrated and wavy petals by inhibition of the activity of TCP transcription factors in Arabidopsis thaliana. *Plant Signal. Behav.* **6**, 697–699 (2011).
- Tatematsu, K., Nakabayashi, K., Kamiya, Y. & Nambara, E. Transcription factor AtTCP14 regulates embryonic growth potential during seed germination in Arabidopsis thaliana. *Plant J.* **53**, 42–52, doi:10.1111/j.1365-313X.2007.03308.x (2008).
- Rueda-Romero, P., Barrero-Sicilia, C., Gomez-Cadenas, A., Carbonero, P. & Onate-Sanchez, L. Arabidopsis thaliana DOF6 negatively affects germination in non-after-ripened seeds and interacts with TCP14. *J. Exp. Bot.* **63**, 1937–1949, doi:10.1093/jxb/err388 (2012).



45. Doebley, J., Stec, A. & Hubbard, L. The evolution of apical dominance in maize. *Nature* **386**, 485–488 (1997).
46. Luo, D., Carpenter, R., Vincent, C., Copsey, L. & Coen, E. Origin of floral asymmetry in *Antirrhinum*. *Nature* **383**, 794–799 (1996).
47. Crawford, B. C. W., Nath, U., Carpenter, R. & Coen, E. S. CINCINNATA controls both cell differentiation and growth in petal lobes and leaves of *antirrhinum*. *Plant Physiol.* **135**, 244–253 (2004).
48. Nath, U., Crawford, B. C. W., Carpenter, R. & Coen, E. Genetic control of surface curvature. *Science* **299**, 1404–1407 (2003).
49. Lewis, J. M. *et al.* Overexpression of the maize Teosinte Branched1 gene in wheat suppresses tiller development. *Plant Cell Rep.* **27**, 1217–1225 (2008).
50. Li, C. X., Potuschak, T., Colon-Carmona, A., Gutierrez, R. A. & Doerner, P. Arabidopsis TCP20 links regulation of growth and cell division control pathways. *Proc. Natl. Acad. Sci. U.S.A.* **102**, 12978–12983 (2005).
51. Arabidopsis Genome, I. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* **408**, 796–815, doi:10.1038/35048692 (2000).
52. Zhang, J. Evolution by gene duplication: an update. *Trends Ecol. Evol.* **18**, 292–298, doi:10.1016/s0169-5347(03)00033-8 (2003).
53. Flagel, L. E. & Wendel, J. F. Gene duplication and evolutionary novelty in plants. *New Phytol.* **183**, 557–564, doi:10.1111/j.1469-8137.2009.02923.x (2009).
54. Blanc, G. & Wolfe, K. H. Widespread paleopolyploidy in model plant species inferred from age distributions of duplicate genes. *Plant cell* **16**, 1667–1678, doi:10.1105/tpc.021345 (2004).
55. Prince, V. E. & Pickett, F. B. Splitting pairs: the diverging fates of duplicated genes. *Nat. Rev. Genet.* **3**, 827–837, doi:10.1038/nrg928 (2002).
56. Wang, M. Y. *et al.* The cotton transcription factor TCP14 functions in auxin-mediated epidermal cell differentiation and elongation. *Plant Physiol.* **162**, 1669–1680, doi:10.1104/pp.113.215673 (2013).

Author contributions

J.M., Q.W., R.S., F.X., D.C.J. and B.Z. designed the experiments and analyzed the data. J.M. and B.Z. wrote the manuscript text. J.M. performed the experiments and R.S. assisted with RNA extraction. All authors reviewed the manuscript.

Additional information

Supplementary information accompanies this paper at <http://www.nature.com/scientificreports>

Competing financial interests: The authors declare no competing financial interests.

How to cite this article: Ma, J. *et al.* Genome-wide identification and expression analysis of TCP transcription factors in *Gossypium raimondii*. *Sci. Rep.* **4**, 6645; DOI:10.1038/srep06645 (2014).



This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder in order to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc-sa/4.0/>