

# Genome-wide identification and expression patterns analysis of the *RPD3/HDA1* gene family in cotton

**Jingjing Zhang**

Chinese Academy of Agricultural Sciences Cotton Research Institute

**Aimin Wu**

Chinese Academy of Agricultural Sciences Cotton Research Institute

**Hengling Wei**

Chinese Academy of Agricultural Sciences Cotton Research Institute

**Pengbo Hao**

Chinese Academy of Agricultural Sciences Cotton Research Institute

**Qi Zhang**

Chinese Academy of Agricultural Sciences Cotton Research Institute

**Miaomiao Tian**

Chinese Academy of Agricultural Sciences Cotton Research Institute

**Xu Yang**

Chinese Academy of Agricultural Sciences Cotton Research Institute

**Shuaishuai Cheng**

Chinese Academy of Agricultural Sciences Cotton Research Institute

**Xiaokang Fu**

Chinese Academy of Agricultural Sciences Cotton Research Institute

**Liang Ma**

Chinese Academy of Agricultural Sciences Cotton Research Institute

**Hantao Wang** (✉ [w.wanghantao@163.com](mailto:w.wanghantao@163.com))

Chinese Academy of Agricultural Sciences Cotton Research Institute

**Shuxun Yu** (✉ [ysx195311@163.com](mailto:ysx195311@163.com))

Chinese Academy of Agricultural Sciences Cotton Research Institute <https://orcid.org/0000-0002-9715-3462>

---

## Research article

**Keywords:** Gossypium, Histone deacetylases, Expression patterns, Abiotic stress, Early maturity

**Posted Date:** August 31st, 2020

**DOI:** <https://doi.org/10.21203/rs.2.19331/v2>

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

[Read Full License](#)

---

**Version of Record:** A version of this preprint was published at BMC Genomics on September 18th, 2020.

See the published version at <https://doi.org/10.1186/s12864-020-07069-w>.

# Abstract

**Background:** Histone deacetylases (HDACs) catalyze histone deacetylation and suppress gene transcription during various cellular processes. Within the superfamily of HDACs, RPD3/HDA1-type HDACs are the most studied, and it is reported that RPD3 genes play crucial roles in plant growth and physiological processes. However, there is a lack of systematic research on the RPD3/HDA1 gene family in cotton.

**Results:** In this study, genome-wide analysis identified 9, 9, 18, and 18 RPD3 genes in *Gossypium raimondii*, *G. arboreum*, *G. hirsutum*, and *G. barbadense*, respectively. This gene family was divided into 4 subfamilies through phylogenetic analysis. The exon-intron structure and conserved motif analysis revealed high conservation in each branch of the cotton RPD3 genes. Collinearity analysis indicated that segmental duplication was the primary driving force during the expansion of the RPD3 gene family in cotton. There was at least one presumed cis-element related to plant hormones in the promoter regions of all GhRPD3 genes, especially MeJA- and ABA-responsive elements, which have more members than other hormone-relevant elements. The expression patterns showed that most GhRPD3 genes had relatively high expression levels in floral organs and performed higher expression in early-maturity cotton compared with late-maturity cotton during flower bud differentiation. In addition, the expression of GhRPD3 genes could be significantly induced by one or more abiotic stresses as well as exogenous application of MeJA or ABA.

**Conclusions:** Our findings reveal that GhRPD3 genes might be involved in flower bud differentiation and resistance to abiotic stresses, which provides a basis for further functional verification of GhRPD3 genes in cotton development and a foundation for breeding better early-maturity cotton cultivars in the future.

## Background

DNA combines with nuclear proteins to constitute the chromatin, which is responsible for storing genetic and directive information in eukaryotic cells. Chromatin is highly arranged and mainly composed of nucleosomes, which are formed by approximately 147 bp of DNA and an octamer organized by the four core histone proteins\_H3, H4, H2A, and H2B [1]. Gene expression in eukaryotes involves a complicated interaction, which is controlled not only by the DNA sequence but also by epigenetic events. Epigenetic mechanisms mainly consist of histone modification and DNA methylation, and play an important role in the regulation of gene expression. In general, histone posttranslational modifications, including methylation, acetylation, phosphorylation, ADP-ribosylation and ubiquitination, occur at the N-terminal of histones [2], and these changes facilitate the binding of other proteins to DNA, resulting in synergistic or antagonistic regulation of gene transcription [3, 4]. Among the several histone modifications, histone acetylation is a reversible process that plays essential roles in epigenetic regulation. The acetylation of core histones is catalyzed by histone acetyltransferases (HATs) to promote transcriptional activation, whereas deacetylation is regulated by histone deacetylases (HDACs) that drive the transcriptional

suppression [5]. HDACs deacetylate the lysine residues of N-terminal histone tails, resulting in the repression of gene expression [6].

HDACs are involved in a large amount of biological processes associated with plant growth and development [7-9]. Based on sequence homology to yeast HDACs, HDACs in plants are divided into three main categories: reduced potassium dependency 3 / histone deacetylase 1 (*RPD3/HDA1*), histone deacetylase 2 (*HD2*), and silent information regulator 2 (*SIR2*) [7, 10, 11]. *RPD3/HDA1*-type histone deacetylases, which are homologous to yeast *RPD3* and *HDA-1*, belong to a large family, and they require zinc ions to catalyze activity; the HDAC inhibitor trichostatin A (TSA) or sodium butyrate can inhibit their enzymatic activities [7]. The *Arabidopsis RPD3/HDA1* gene family is further classified into three groups. Class I includes *HDA6*, *HDA7*, *HDA9*, and *HDA19*; class II includes *HDA5*, *HDA15*, and *HDA18*; and *HDA2* is the only member of class III [7, 8]. The other genes of *RPD3/HDA1* family are unclassified in *Arabidopsis*.

Over the past twenty years, *RPD3/HDA1*-type HDACs (call *RPD3* for short below) have been studied extensively as global regulatory factors playing essential roles in a series of plant growth and development processes and the response to various environmental stresses [8, 12-14]. In *Arabidopsis*, it has been reported that *AtHDA19* was involved in various developmental processes, including flowering time, circadian clocks functions, and seed development [15, 16]. Additionally, *AtHDA19* might regulate gene expression related to jasmonic acid and ethylene signaling pathways in response to wounding and pathogen infection [17]. In maize, the expression patterns of the three *ZmPRD3* genes *ZmRpd3/101*, *ZmRpd3/102*, and *ZmRpd3/108* showed widespread expression in all investigated corn organs. Furthermore, the gene products could be detected in all cellular parts at specific stages such as kernel, shoot, and anther developmental periods [18]. In rice, *HDA705* responded to ABA and abiotic stresses, and its expression was induced by JA. In addition, the expression of *HDA702* and *HDA704* was significantly induced by SA, JA, or ABA [19, 20]. These findings indicate that the *RPD3* members play an important regulatory role in plant development and in the response to various stresses and plant hormones.

Cotton is one of the most important economic crops in China with an essential role in the national economy. Early maturity and stress resistance are vital target traits of cotton breeding. Over the past two decades, the *RPD3* genes have been intensively studied, and some progress has been made in *Arabidopsis* and some other crops. However, there is a lack of systematic research on the *RPD3* gene family in cotton. Thus, it is necessary to explore the potential functions of *RPD3* genes in cotton. In our study, the protein sequences of cotton *RPD3*-type HDACs were predicted by genome-wide identification and the phylogenetic tree, gene structure, conserved motif, protein domain, expression profiles, and preliminary functions were comprehensively analyzed. The information gained for *GhRPD3* provides a reference for further exploration of the possible functions of *RPD3* genes in cotton growth and development.

## Results

### Identification of *RPD3* genes in nine species

In this study, a total of 108 *RPD3* protein sequences from nine species were identified after eliminating redundant sequences, and they are named by the position on the chromosome. The corresponding relationship between gene ID number and gene name is shown in Additional file 1: Table S1. A total of 18 genes (*GhHDA1-GhHDA18*) containing Hist\_deacetyl (PF00850) domains were identified from *G. hirsutum*; 9 genes were located on the At genome, and 9 genes were mapped on the Dt genome. Furthermore, 18 genes (*GbHDA1-GbHDA18*) from *G. barbadense*, 9 genes (*GaHDA1-GaHDA9*) from *G. arboreum*, and 9 genes (*GrHDA1-GrHDA9*) from *G. raimondii* were detected. Tetraploid cotton possessed twice as many *RPD3* genes as diploid cotton, indicating that no *RPD3* cotton gene was lost in the process of polyploidy. The numbers of *RPD3* genes in the other five species were 10 (*Arabidopsis*), 14 (*Oryza sativa* L.), 11 (*Populus trichocarpa*), 8 (*Theobroma cacao*), and 11 (*Zea mays* L.). The *GhRPD3* protein length ranged from 232 to 635 aa with an average of 459 aa. The physicochemical parameters showed that the isoelectric point (pI) of *GhRPD3* proteins varied from 4.47 to 8.65 with an average value of 5.68, and the molecular weight of *GhRPD3* proteins varied from 25.79 to 73.01 kDa with an average value of 51.21 kDa. The subcellular localization results indicated that most of the *GhRPD3* genes were located in cytoplasmic (10) or nuclear (8), suggesting that *GhRPD3* genes might possess multiple regulatory functions (Table 1). The predicted length, pI, MW and subcellular localization of the *RPD3* proteins in other eight species are shown in Additional file 1: Table S1.

## Phylogenetic analysis of the *RPD3* gene family

A total of 108 identified *RPD3* protein sequences from *G. raimondii* (9), *G. arboreum* (9), *G. hirsutum* (18), *G. barbadense* (18), *A. thaliana* (10), *T. cacao* (8), *Oryza sativa* (14), *Zea Mays* (11) and *P. trichocarpa* (11) were employed to construct an unrooted phylogenetic tree using the neighbor-joining method for investigating the evolutionary relationships of *RPD3* proteins. The *RPD3* proteins were phylogenetically classified into 4 subfamilies (Class I, Class II, Class III, and unclassified) according to the formulated subfamilies in *Arabidopsis* [7]. The Class I subgroup was the largest subfamily with 49 *RPD3* genes, whereas the Class III subgroup has the fewest members, only containing one gene in seven diploid species genomes and two genes in two tetraploid cotton genomes (Fig. 1). Among these four classes, each subfamily contained *RPD3* genes from all nine species, indicating this gene family was relatively conserved in different species during evolution.

## Exon-intron structure and conserved motif analysis

The domains of the *RPD3* sequences in cotton were investigated and exhibited according to the results of the SMART database using TBtools, revealing that all cotton *RPD3* genes contained a Hist\_deacetyl domain (Additional file 2: Table S2 and Additional file 3: Figure S1). An unrooted phylogenetic tree with the predicted cotton *RPD3* genes was constructed (Fig. 2a), and then exon-intron structure (Fig. 2b) and conserved motifs (Fig. 2c) were analyzed to better understand the similarity and differences of cotton *RPD3* members. The results showed that the length of *RPD3* cotton genes was relatively conserved in

Class I and Class III, but there were twelve longer sequences in Class II and the unclassified group. The *RPD3* cotton genes included from 3 to 17 exons and most *RPD3* genes (48/54) contained more than five exons (Additional file 4: Table S3), which might be associated with the diversification of their functions. In terms of the distribution of motifs, most *RPD3* cotton genes belonging to the same subfamily showed a similar motif composition, except in the unclassified group (Fig. 2c). Most Class I subfamily members contained 9 motifs, whereas *GrHDA5* and *GhHDA4* contained 4 and 6 motifs, respectively. Class III subfamily genes had three or four motifs, and most Class II subfamily members possessed 7 motifs, except for *GhHDA12* with 6 motifs. There were differences in the exon-intron structure and motif arrangement among the four categories, but they were highly conserved on the same branches, indicating that the *RPD3* members classified into the same branch might perform a relatively conserved function in cotton growth and development.

## Chromosomal distribution, gene duplication and selection pressure

The chromosomal distributions of *GrRPD3*, *GaRPD3*, *GbRPD3*, and *GhRPD3* genes were visualized according to the genomic position of 54 cotton *RPD3* genes (Additional file 5: Table S4 and Fig. 3). In *G. hirsutum*, 18 *GhRPD3* genes were unevenly mapped on 13 chromosomes. A03 contained the most *GhRPD3* genes (3), whereas the other 12 chromosomes only contained one or two *GhRPD3* genes (Fig. 3a). The chromosomal distribution of 18 *GbRPD3* genes in *G. barbadense* was similar to that of *GhRPD3* genes in *G. hirsutum* (Fig. 3b). In *G. arboreum*, 9 *GaRPD3* genes were unevenly located on 6 chromosomes. Chr01 and Chr13 contained three and two *GaRPD3* genes, respectively, and the other 4 chromosomes contained only one *GaRPD3* gene (Fig. 3c). In *G. raimondii*, the chromosomal distribution of 9 *GrRPD3* genes was highly consistent with the corresponding D subgenome of *G. hirsutum* (Fig. 3d), showing conserved numbers and chromosomal distribution of *RPD3* genes between diploid and tetraploid cotton species. In addition, the lopsided chromosomal distribution of the cotton *RPD3* genes indicated that genetic variation occurred during evolution. Notably, most of the *RPD3* genes were distributed on the opposite ends of the chromosomes in four cotton species (Fig. 3).

In general, tandem and segmental duplication are two of the main reasons for gene family generation during evolution [21]. The analysis of gene duplication indicated that all *RPD3* family members were amplified only through segmental duplication (Additional file 6: Table S5), suggesting that segmental duplication played a vital role in the evolution of the *RPD3* gene family. The homologous gene pairs obtained by collinearity analysis among *RPD3* genes in *G. arboreum*, *G. raimondii*, and *G. hirsutum* were visualized using circular maps (Fig. 4). The  $K_a/K_s$  ratios of most homologous gene pairs were lower than one, indicating that purified selection was essential during the evolution of cotton *RPD3* genes, whereas the  $K_a/K_s$  ratios of two gene pairs (*GhHDA2/GaHDA3* and *GhHDA6/GaHDA6*) were more than 1, suggesting that these two pairs might have experienced positive selection pressure. The study also predicted the occurrence time of segmentally duplicated *RPD3* gene pairs by the formula “ $t=K_s/2r$ ” ( $r=2.6 \times 10^{-9}$ ) [22]. Except for the *GhHDA6/GaHDA6* gene pair, the other segmental duplication events of

three cotton species might have occurred 0.6 to 144.44 million years ago (MYA) with an average time of 18.39 million years ago (Additional file 6: Table S5).

## Analysis of cis-elements in predicted promoter regions of *GhRPD3*

To explore the possible regulatory functions of *GhRPD3* genes under various environmental stresses and hormone regulation pathways, the 2,000-bp promoter regions of 18 *GhRPD3* genes were submitted to the PlantCARE database for the identification of putative stress-associated and plant hormone-related cis-elements. A total of 9 kinds of elements related to plant hormones, containing AuxRE-core (auxin), TGA-element (auxin), P-box (gibberellin), TATC-box (gibberellin), GARE-motif (gibberellin), CGTAC-motif (MeJA), TGACG-motif (MeJA), TCA-element (SA), and ABRE (ABA), and 4 kinds of elements responding to stresses, including TC-rich repeats (defense and stress responsiveness), MBS(drought), WUN-motif (wound) and LTR (cold stress), were predicted in the promoters of *GhRPD3* genes. As shown in Fig. 5, the promoters of some *GhRPD3* genes contained various hormone-responsive and stress-responsive components, such as *GhHDA2* (2 MBS, 2 LTR, 2 TC-rich repeats, 1 GARE-motif, 2 ABRE, 1 TGACG-motif) and *GhHDA13* (1 MBS, 1 LTR, 1 TC-rich repeats, 1 AuxRE-core, 2 GARE-motif, 1 TCA-element, 4 ABRE, 2 TGACG-motif). Among the 18 *GhRPD3* genes, there are large numbers of light-responsive elements distributed in their promoter regions (Additional file 7: Table S6). In addition, MeJA-responsive and ABA-responsive elements are more common than other hormone-related elements (Additional file 8: Figure S2). These results revealed that *GhRPD3* genes might be involved in MeJA and ABA hormone signaling pathways as well as response to environmental stresses.

## Expression profiles of *GhRPD3* genes in different tissues and under different abiotic stresses

To understand the potential functions of *GhRPD3* genes in the growth and development of cotton, we studied their expression in various cotton tissues, including the anther, pistil, bract, sepal, petal, filament, torus, root, leaf, stem, ovules, and fibers, using publicly available transcriptomic data provided by Hu et al. [23]. Transcripts of all the *GhRPD3* genes were detected in at least three tissues with fragments per kilobase million (FPKM)  $\geq 1$ . Furthermore, ten genes exhibited high expression levels in all selected tissues (Additional file 9: Table S7). These results indicated that *GhRPD3* genes are widely expressed in both reproductive organs and vegetative organs and thus might have multiple biological functions. After log<sub>2</sub>-conversion of FPKM values, the expression profiles of *GhRPD3* genes in different tissues are shown in Fig. 6a. Seven *GhRPD3* genes exhibited relatively high expression levels in at least eight tissues (log<sub>2</sub>-transformed FPKM value  $\geq 2.6$ ); in particular, one pair of homologous genes (*GhHDA1/GhHDA10*) showed a high expression level in all the tissues with a similar expression pattern. Nevertheless, three *GhRPD3* genes (*GhHDA4*, *GhHDA14*, *GhHDA18*) were expressed at relatively low levels in at least eight tissues (log<sub>2</sub>-transformed FPKM value  $< 1$ ), of which *GhHDA14* showed the lower expression in all tissues except

for the pistil. These homologous gene pairs (*GhHDA1/GhHDA10*, *GhHDA4/GhHDA11*, *GhHDA2/GhHDA13*, *GhHDA6/GhHDA15*, *GhHDA7/GhHDA16*, and *GhHDA9/GhHDA18*) were located on At and Dt subgenomes and exhibited similar expression patterns. For example, homologous gene pairs (*GhHDA4/GhHDA11* and *GhHDA9/GhHDA18*) showed relatively low expression in all twelve tissues. The gene pair *GhHDA2/GhHDA13* exhibited relatively high expression in torus and ovule but relatively low expression in petals (Fig. 6a).

Based on the analysis of cis-elements in promoter regions and previous reports on *RPD3* genes in other plants, *GhRPD3* genes might respond to abiotic stresses. To test this hypothesis, we investigated the expression characteristics of 18 *GhRPD3* genes under cold, heat, PEG, and salt treatments using available transcriptomic data [23] (Fig. 6b). The expression of most *GhRPD3* genes were induced by the four stresses to varying degrees. *GhHDA1*, *GhHDA2*, *GhHDA6*, *GhHDA10*, *GhHDA12*, and *GhHDA18* showed upregulated expression under four stress treatments. However, one gene (*GhHDA4*) exhibited marked downregulation in the presence of the four abiotic stresses. Some genes can respond to one specific abiotic stress. For example, the expression of *GhHDA13* and *GhHDA16* was significantly induced by PEG treatment. Four genes (*GhHDA7*, *GhHDA11*, *GhHDA5*) showed upregulated expression under heat treatment. The expression of *GhHDA9* was significantly upregulated under cold and salt treatments. According to the results, we can conclude that *GhRPD3* genes play an essential role in response to abiotic stresses.

## Characterization of *GhRPD3* genes expression during flower bud differentiation

To explore expression differences of *GhRPD3* genes between early-maturity and late-maturity cottons during flower bud differentiation, we selected nine genes showing relatively high expression in floral organ tissues (anther, pistil, bract, sepal, petal, filament and torus) for qRT-PCR. The buds of an early-maturity variety (CCRI50) and a late-maturity variety (GX11) from the one-leaf to five-leaf stage were used for qRT-PCR (Fig. 7). The results revealed that more than half of these genes (5/9) possessed relatively higher expression in early-maturity cotton compared with late-maturity cotton during flower bud differentiation. *GhHDA5* showed marked differences at the two-leaf and three-leaf stages, and these two stages were regarded as the important period of flower bud differentiation. A homologous gene pair (*GhHDA6/GhHDA15*) located on At and Dt respectively, showed the same expression trend. Both of them presented the highest expression at three-leaf stage and then exhibited downregulated expression in next two stages in CCRI50. In addition, all nine genes showed relatively higher expression at the two-leaf or three-leaf stage in CCRI50 compared with GX11. The results showed that *GhRPD3* genes are associated with the early maturity of cotton.

## Responses of *GhRPD3* genes to MeJA and ABA treatment

MeJA and ABA play important roles in plant stress resistance. To further explore the possible functions of *GhRPD3* genes, we selected the *GhRPD3* genes containing MeJA- and ABA-responsive elements in the predicted promoters to analyze their expression characteristics under MeJA and ABA treatment by qRT-PCR (Fig. 8 and Fig. 9). Most *GhRPD3* genes (8/13) were markedly upregulated at 9 h after MeJA treatment. Three genes (*GhHDA7*, *GhHDA13*, and *GhHDA18*) exhibited significantly upregulated expression at three or more time points, whereas four genes (*GhHDA2*, *GhHDA8*, *GhHDA9*, and *GhHDA11*) showed marked transcriptional downregulation at least three time points after MeJA treatment (Fig. 8). More than half of the *GhRPD3* genes (6/11) were significantly upregulated at 9 h after ABA treatment. Three *GhRPD3* genes (*GhHDA14*, *GhHDA15*, and *GhHDA18*) showed relatively high expression at three or more time points, whereas three *GhRPD3* genes (*GhHDA10*, *GhHDA11*, and *GhHDA17*) showed early downregulated and then upregulated expression patterns under ABA treatment (Fig. 9). The results showed that the exogenous application of MeJA and ABA significantly induced the transcription of most *GhRPD3* genes containing MeJA-responsive and ABRE elements in their promoter regions.

## Discussion

Among the several histone modifications, histone acetylation plays an essential role in plant growth and development [24]. Histone acetylation and deacetylation are catalyzed by histone acetyltransferases (HATs) and histone deacetylases (HDACs), respectively [20]. In plants, HDACs are involved in a variety of biological processes associated with plant growth and development [25]. Within the superfamily of HDACs, the *RPD3* gene family is the most studied and is crucial in plant development and physiological processes, including flowering time, abiotic stress response, female gametophyte and embryo development, senescence, seed germination, and plant hormone signal response [12-14, 19, 26-28]. To date, although a few studies have analyzed the function of the *RPD3* family members in *G. hirsutum*, they focus primarily on the roles of *GhHDA5* in fiber initiation, and there are no systematic reports [29]. To explore the characteristics of *RPD3* family members and understand the roles played by cotton *RPD3* genes in cotton growth and development, we conducted an integrated analysis of the *RPD3* gene family in cotton, containing their phylogenetic relationships, exon-intron structures, conserved motifs, chromosomal distributions, duplication events, expression patterns in different tissues and in the presence of abiotic stresses or MeJA and ABA treatment.

## Phylogeny, gene structure, and expansion of *RPD3* genes in cotton

A total of 18, 18, 9 and 9 *RPD3* genes were identified by genome-wide identification in *G. hirsutum*, *G. barbadense*, *G. aboreum*, and *G. raimondii*, respectively. The number of *RPD3* cotton genes in diploid cotton was half of that in tetraploid cotton, suggesting that the deletion of *RPD3* genes did not happen in allotetraploids, which is not in agreement with the higher ratio of gene deletion in allotetraploids [30]. According to the *AtRPD3* genes, 108 *RPD3* genes from nine species were classified into four groups (Fig. 1), similar to the previous classification of *Arabidopsis* [7].

The conservation of biological functions might be related to the conservation of gene structure [31]. To investigate the conservation of the *RPD3* gene sequences, exon-intron structure and conserved motifs were analyzed. Exon-intron structure analysis showed that the exon numbers of *RPD3* genes in cotton was highly diverse, ranging from 3 to 17 (Additional file 4: Table S3), which might be associated with the diversification of their functions. Notably, the gene structure and motif arrangement were different among the four subfamilies, whereas they were highly conserved on the same branch, indicating that the *RPD3* genes (especially the members on the same branch) might preform conserved functions in the growth of cotton. The cotton *RPD3* genes exhibited similarities and differences in exon-intron structures and motifs, which might be associated with conservation and subfunctionalization caused by gene duplication during the evolution of the cotton *RPD3* gene family.

According to the evolutionary history of cotton, tetraploid cotton was formed by the hybridization of two diploid cottons with subsequent polyploidization [31]. To investigate the evolutionary relationships of predicted *RPD3* genes between two diploid genomes and subgenomes in allotetraploids, we analyzed their chromosomal distributions and gene duplication events. The results showed that the chromosomal distribution of *RPD3* genes in *G. arboreum* and the corresponding A subgenome of *G. hirsutum* was not identical, whereas the chromosomal location of *RPD3* genes in *G. raimondii* and the corresponding D subgenome of *G. hirsutum* was highly consistent (Fig. 3), illustrating strong conservation during *RPD3* gene family evolution. The analysis of gene duplication showed that segmental duplication was essential for the expansion of *RPD3* family members (Additional file 6: Table S5). According to previous genomic studies of cotton, A and D genome diploid cottons began to differentiate from a common ancestor 5–10 MYA [23]. Subsequently, *G. hirsutum* evolved from the hybridization of two diploid cottons approximately 1–2 MYA [22]. In *G. hirsutum*, the deduced divergence times of most *RPD3* homologous gene pairs varied from 5.66 to 11.06 MYA (Additional file 6: Table S5), which was accompanied by the divergence of A and D ancestral genomes. Ka/Ks ratio analysis indicated that the segmentally duplicated gene pairs might perform similar functions based on purified selection in functional segregation [32].

## Functional analysis of *GhRPD3* genes in upland cotton

Flowering time, an important indicator of precocity, is also influenced by the timing of flower bud differentiation, which is a physiological and morphological marker of the transformation from vegetative growth to reproductive growth [33, 34]. Previous studies showed that the flower bud differentiation period of early-maturity cotton varieties occurs earlier than that of late-maturity cotton varieties, and early-maturity cotton generally begins flower bud differentiation when three true leaves are completely flattened [33]. In *Arabidopsis*, at least 4 *RPD3* genes have been associated with flowering time; *AtHDA6* has been related with the autonomous pathway of four flowering-promoting pathways and regulates flowering time by interacting with *FLD* (*Flowering LOCUS D*) [13, 26]; *HDA5* regulates flowering time by repressing the expression of *FLC* (*FLOWERING LOCUS C*) and *MAF1*. In addition, *HDA5* and *HDA6* might form an HDAC complex with *FLD* and *FVE* to control flowering time in *Arabidopsis* [27]. In short days, *AtHDA9* represses the flowering-promoting gene *AGL19* (*AGAMOUS-LIKE 19*) regulated by photoperiod

[35] and the downregulation of *AtHDA19* causes delayed flowering, flower abnormalities, embryonic defects, and seed set reduction [36]. In recent studies, *GhHDA5*, similar to *AtHDA5* in *Arabidopsis*, exhibited higher expression at -1 and 0 DPA, and RNAi-repressed *GhHDA5* lines showed delayed flowering, suggesting that it may be involved in cotton fiber initiation and flowering time [29]. As the homologous genes of these four *AtRPD3* genes related to flowering time in cotton, the gene pairs *GhHDA2/ GhHDA13*, homologous to *AtHDA6*, and *GhHDA6/ GhHDA15*, homologous to *AtHDA9*, showed markedly higher expression in early-maturity cotton compare with late-maturity cotton during all five stages of flower bud differentiation. The other five *GhRPD3* genes we selected exhibited relatively higher expression at the two-leaf or three-leaf stage of early-maturity cotton (Fig. 7), indicating that *GhRPD3* genes are helpful for improving the molecular breeding of early-maturity cotton.

Cis-elements in promoter regions play important roles in gene expression regulation. In general, gene expression depends on the presence or absence of these elements [37]. In previous studies, multiple lines of evidence have revealed that *RPD3* members play essential roles in response to various stresses or plant hormones in *Arabidopsis*, rice, maize, *Populus trichocarpa*, and others [13, 17, 38]. To better understand the regulation of *GhRPD3* genes under different environmental conditions, we investigated the cis-elements in their promoter regions. Nine kinds of plant hormone-related elements and 4 kinds of stress-responsive regulatory elements were identified in the presumed promoters of *GhRPD3* genes (Fig. 5). Such a wide range of cis-acting elements is consistent with the published studies on the multifunctional roles *RPD3* genes play in plant growth [7, 8]. Based on the expression patterns of *GhRPD3* genes under four abiotic stresses, almost all of the 18 genes were significantly induced by all four stresses or by a specific treatment (Fig. 6b), fully illustrating that *GhRPD3* genes can respond to abiotic stresses.

MeJA and ABA not only regulate plant growth and development but also participate in plant defense response to environmental stress such as mechanical injury, disease, and osmotic stress [39, 40]. Previous studies showed that the expression of *HDA705*, *HDA702*, and *HDA704* could be induced in rice by JA or ABA [19, 20]. The transcription of *AtHDA6*, *AtHDA19*, and *AtHDA9* has been reported to be increased after ABA or JA treatment in *Arabidopsis*. In *AtHDA6* mutants, *axe1-5* and *HDA6-RNAi* plants, the expression of JA-responsive and ABA-responsive genes was significantly downregulated, and the mutants exhibited hypersensitivity to ABA during seed germination [13, 41]. *AtHDA19-1*, a T-DNA insertion mutant, showed similar phenotypes as *AtHDA6* mutant in response to ABA and in gene expression, whereas overexpression of *AtHDA19* enhanced the expression of defense genes synergistically induced by JA and ethylene [17, 38]. According to the analysis of cis-elements in predicted *GhRPD3* promoters, we selected the *GhRPD3* genes containing MeJA-responsive and ABA-responsive elements in the predicted promoters to analyze their expression characteristics under MeJA and ABA treatment by qRT-PCR. The results revealed that exogenous MeJA and ABA application significantly induced the transcription of most *GhRPD3* genes at different time points (Figs. 8 and 9). *GhHDA13*, similar to *AtHDA6*, was markedly upregulated after treatment with MeJA and ABA. The significant induction of *GhHDA13* expression under cold and PEG treatment might be related to the presence of cold and drought-responsive cis-elements in its predicted promoter. As the homologous gene of *AtHDA9*, *GhHDA15* not only showed relatively high

transcriptional levels after ABA treatment at all four stages but also exhibited significantly upregulated expression under MeJA treatment at 9 h and 12 h. Some *RPD3* genes were less studied in other plants, but they could respond to MeJA and ABA treatment in upland cotton. For example, *GhHDA18* positively responded to ABA and MeJA treatment at all four stages, and its expression was slightly induced under stress conditions. The homologous gene pair *GhHDA7/GhHDA16* showed higher expression at 9 h and 12 h after exogenous application of MeJA, and their expression could be induced by PEG treatment. The homologous gene pair *GhHDA8/GhHDA17* showed lower expression in the early stages of ABA treatment. The expression patterns of *RPD3* genes indicated that they might participate in abiotic stress responses via ABA and MeJA signaling pathways.

On the basis of our expression profile analysis of *RPD3* genes in upland cotton, *GhHDA1*, *GhHDA2*, *GhHDA6*, *GhHDA10*, and *GhHDA13* showed relatively high expression levels in most of the investigated tissues and in early-maturity cotton during flower bud differentiation. These genes also played essential roles in response to MeJA, ABA, and abiotic stresses, which is consistent with the previous extensive evidences showing that plant histone deacetylases play vital roles in plant developmental processes and responses to various environmental stresses [7, 17, 26, 27, 41, 42]. Additionally, other *GhRPD3* genes also performed specific functions in cotton development, laying the foundation for further functional verification of *RPD3* genes in upland cotton.

## Conclusions

In this study, a total of 108 *RPD3* genes were detected in nine species by genome-wide identification. These genes were divided into four subgroups according to the classification in *Arabidopsis*. The exon-intron structure and conserved motif analysis of 54 cotton *RPD3* genes showed that significant differences exist among the four subfamilies, whereas they are highly conserved on the same branch, indicating that cotton *RPD3* genes on the same branch might perform similar functions in cotton growth and development. The chromosomal distributions of cotton *RPD3* genes revealed conserved gene numbers and chromosomal locations between diploid and tetraploid cotton species. Gene expression analysis showed that most *GhRPD3* genes had relatively high expression in floral organs and exhibited the higher expression in CCRI50 compared with GX11 during flower bud differentiation. In addition, the expression of *GhRPD3* genes could be significantly induced by one or more abiotic stresses or exogenous application of MeJA and ABA. These results revealed that *GhRPD3* genes might be involved in flower bud differentiation and resistance to abiotic stresses in cotton, which provides a basis for further functional verification of *GhRPD3* genes in cotton development and a foundation for breeding better early-maturity cotton cultivars in the future.

## Methods

### Plant materials and treatments

Two *G. hirsutum* cultivars (CCRI50, GX11) were grown under standard field environments (5 rows, each 8 m long and 0.8 m wide) in Anyang, Henan province, China. CCRI 50 is an early-maturity cotton variety with initial flowering time of 60 days, and GX11 is a late-maturity cotton cultivar with initial flowering time of 70 days. The buds of two cotton cultivars were collected from the one-leaf to five-leaf stage to analyze the expression differences of *GhRPD3* genes between CCRI 50 and GX11 cultivars during flower bud differentiation.

TM-1 was planted in a climate-controlled greenhouse with a suitable growing environment (light/dark cycle: 16 h at 28 °C/8 h at 22 °C). Four-week seedlings showing two flat true leaves were sprayed with 100 mM MeJA, 200 mM ABA, and water as a blank control to explore the responses to MeJA and ABA treatment. After exogenous application of plant hormones and water, we isolated the leaves of three seedlings from every treatment at 3 h, 6 h, 9 h, and 12 h and promptly froze these samples in liquid nitrogen for RNA extraction.

## Identification and sequence retrieval of *RPD3* family members

The HMM file (PF00850) of the conserved Hist\_deacetyl domain was downloaded from the Pfam database (<https://pfam.xfam.org/>) [43]. Putative *RPD3* proteins of *G. arboreum* (CRI\_1.0) [44], *G. raimondii* (JGI\_v2.1) [45], *G. hirsutum* (HAU\_v1) [46] and *G. barbadense* (HAU\_v1) [46] from the CottonFGD (<http://www.cottonfgd.org/>) [47] were searched using the hidden Markov model profile of the HMMER 3.0 software with an E value threshold of 1e-10 [48]. The study also searched against the *Arabidopsis* genome (Araport\_11) [49] obtained from the TAIR website (<https://www.arabidopsis.org/>) and the published genomes of *Oryza sativa* L. (JGI\_v7.0), *Theobroma cacao* (JGI\_V2.1) *Populus trichocarpa* (JGI\_v3.1), and *Zea mays* L. (JGI\_v4) downloaded from phytozome\_V13 (<https://phytozome.jgi.doe.gov/pz/portal.html>) [50] using the Hist\_deacetyl HMM file. After obtaining the ID number of the possible genes in these nine protein databases, the *RPD3* protein sequences of different species were extracted from the formatted protein databases using blast (ncbi-blast-2.6.0+-x64-win64.tar). The normal model of the SMART database (<http://smart.embl-heidelberg.de/>) was employed to verify each predicted *RPD3* protein with a Hist\_deacetyl domain, and proteins that did not contain the conserved domain were removed [51]. These *RPD3* genes were named following a rule that short for species names.

The identified *RPD3* protein sequences were submitted to the online program Pepstats ([https://www.ebi.ac.uk/Tools/seqstats/emboss\\_pepstats/](https://www.ebi.ac.uk/Tools/seqstats/emboss_pepstats/)) [52] and CELLO v2.5 (<http://cello.life.nctu.edu.tw/>) [53] to predict their amino acid length, theoretical molecular weight (Mw), isoelectric point (pI), and subcellular localization.

# Multiple alignment and phylogenetic analysis of *RPD3* proteins

Multiple alignment of all the presumed *RPD3* protein sequences from the nine plant species including *G. raimondii*, *G. arboreum*, *G. hirsutum*, *G. barbadense*, *A. thaliana*, *T. cacao*, *P. trichocarpa*, *Oryza sativa*, and *Zea mays* was performed using the ClustalW program with default parameters [54]. The alignment result was employed to construct an unrooted phylogenetic tree using the neighbor-joining (NJ) method of MEGA 7.0 program and 1,000 bootstrap repetitions were used to increase the reliability of interior branches [55].

## Chromosomal distribution, gene structure, and conserved motif analysis

The physical positions of *GhRPD3*, *GbRPD3*, *GaRPD3*, and *GrRPD3* genes on chromosomes were confirmed based on the genome annotation GFF3 files obtained from the CottonFGD website, and the distribution of cotton *RPD3s* were visualized using TBtools [47, 56]. The GFF3 files of *RPD3* genes in four cotton species were submitted to the online toolkit GSDS 2.0 (<http://gsds.cbi.pku.edu.cn/>) for analysis and visualization of the exon-intron structure [57]. The online program MEME 5.0.5 (<http://meme-suite.org/tools/meme>) was employed to detect the conserved motifs of cotton *RPD3* proteins with the following optimized parameters: maximum number of motifs, 10; the optimum width of each motif, 6-50 aa; and E value, 1e-5 [58].

## Gene duplication events and selection pressure

This study used BLASTp search (E-value < 1e-10) to align protein sequences in three cotton species, and the MCScanX program in TBtools was employed to perform genome collinearity analysis based on the BLASTp results. [56, 59]. The circular maps of identified *RPD3* gene pairs in three cotton species were displayed using the circos program [60]. The adjacent *RPD3* family members on a single chromosome were considered to be tandem duplicated genes [61]. The coding sequences of *RPD3* homologous gene pairs were used to calculate the ratios of nonsynonymous (Ka) substitutions and synonymous (Ks) substitutions by the NG methods of TBtools to evaluate the selection pressure of these gene pairs [56, 62]. Normally, Ka/Ks < 1 indicates purifying selection; Ka/Ks = 1 indicates neutral selection; and Ka/Ks > 1 indicates positive selection. The divergence times of the homologous gene pairs were estimated using the formula  $t=ks/2r$ , with  $r=2.6 \times 10^{-9}$  representing neutral substitution [63].

## Analysis of cis-elements in *GhRPD3* promoter regions

The *GhRPD3* promoter regions containing 2,000 bp of DNA upstream of the initiation codon (ATG) were extracted from the *G. hirsutum* genome database downloaded from CottonFGD (<https://cottonfgd.org/>). The regulatory elements in the *GhRPD3* promoter regions were predicted using the online tool PlantCARE (<http://bioinformatics.psb.ugent.be/webtools/plantcare/html/>) [64].

## Gene expression pattern analysis

Primary RNA-seq data of *G. hirsutum* TM-1 were obtained from the NCBI Sequence Read Archive (SRA: PRJNA490626) (<https://www.ncbi.nlm.nih.gov/bioproject/PRJNA490626>) [23]. Transcriptomic reads were mapped to the *G. hirsutum* genome using TopHat2 with the default parameters [65], and gene expression was calculated in fragments per kilobase million (FPKM) by the cufflinks program [66]. The FPKM values of TM-1 in twelve different tissues (anther, pistil, bract, sepal, petal, filament, torus, root, leaf, stem, ovule, and fiber) and under four treatments (heat, cold, PEG, and salt) were obtained. The relative data of *GhRPD3* genes were normalized by log<sub>2</sub> transformation to investigate their expression patterns. The expression characteristics of *GhRPD3* genes among all twelve tissues and under four abiotic stresses were displayed with Heml 1.0.3.7 software [67].

## RNA extraction and quantitative RT-PCR (qRT-PCR) experiments

Cotton buds collected at different stages and leaves taken after MeJA and ABA treatment were frozen in liquid nitrogen, and then a mortar and pestle were used to grind the samples into fine powder [68]. Depending on the operating instructions, total RNA of these samples was extracted using the Tiangen RNA-prep Pure Plant kit (Tiangen, China), and then 1 µg of total RNA was used as template to reverse-transcribe first-strand cDNA using the PrimeScript RT Reagent kit (Takara, Japan); cDNA was diluted five-fold for further experiments. The cotton histone-3 gene (AF024716) was used as an internal control [69], and specific primers for qRT-PCR analysis of *GhRPD3* genes were designed using Oligo 6.0 software and are shown in Additional file 10: Table S8. A total volume of 20 µL containing 10 µL 2×UltraSYBR Mixture, 0.4 µL of each primer (10 µM), 2 µL cDNA, and 7.2 µL ddH<sub>2</sub>O was employed to conduct qRT-PCR on an ABI 7500 real-time PCR system (Applied Biosystems, USA) using UltraSYBR Mixture (Low ROX) (Cwbio, China) with three technical repetitions and three biological replicates. The following detailed run method was used: step 1, primal denaturation of 10 min at 95 °C; step 2, 40 cycles of 10 s at 95 °C, 30 s at 60 °C, and 32 s at 72 °C; step 3, melting curve analysis. The relative expression of *RPD3* genes was calculated using the  $2^{-\Delta\Delta CT}$  method [70].

## Abbreviations

HDACs: Histone deacetylases; HATs: Histone acetyltransferases; RPD3/HDA1: Reduced potassium dependency 3/histone deacetylase 1; ABA: Abscisic acid; JA: Jasmonic acid; SA: Salicylic acid; MeJA:

Methyl jasmonate; ABRE: Abscisic acid responsiveness elements; MW: Molecular weight; GFF: General Feature Format; FPKM: Fragments per kilobase million; MYA: Million years ago; qRT-PCR: Quantitative real time polymerase chain reaction.

## Declarations

### Ethics approval and consent to participate

Not applicable.

### Consent for publication

All authors read and approved the manuscript.

### Availability of data and materials

The data included in this article and the additional files are available. The transcriptome datasets of *G. hirsutum* TM-1 are under the accession number in PRJNA490626 NCBI.

### Competing interests

The authors declare that they have no competing interests.

### Funding

This research was supported by the Chinese National Natural Science Foundation (31601346) and the China Agriculture Research System (CARS-15-06).

### Authors' contributions

HTW, HLW and SXY conceived and designed the experiment; JJZ, MMT and XY took samples in climate-control green house and performed experiments; LM and XKF prepared the field materials ; JJZ wrote the paper; AMW and PBH helped the data analysis; SSC and QZ revised the manuscript; All authors read and approved the final manuscript.

### Acknowledgements

The experiment was performed at the State Key Laboratory of Cotton Biology at the Institute of Cotton Research of the Chinese Academy of Agricultural Sciences.

## References

1. Berr A, Shafiq S, Shen WH: **Histone modifications in transcriptional activation during plant development.** *Biochim Biophys Acta* 2011, **1809**(10):567-76.

2. Reyes JC, Hennig L, Gruissem W: **Chromatin-remodeling and memory factors. New regulators of plant development.** *Plant Physiol* 2002, **130**(3):1090-101.
3. Bannister AJ, Kouzarides T: **Regulation of chromatin by histone modifications.** *Cell Res* 2011, **21**(3):381-95.
4. Ng HH, Bird A: **Histone deacetylases: silencers for hire.** *Trends Biochem Sci* 2000, **25**(3):121-6.
5. Chen ZJ, Tian L: **Roles of dynamic and reversible histone acetylation in plant development and polyploidy.** *Biochim Biophys Acta* 2007, **1769**(5):295-307.
6. Strahl BD, Allis CD: **The language of covalent histone modifications.** *Nature* 2000, **403**(6765):41-5.
7. Hollender C, Liu Z: **Histone Deacetylase Genes in Arabidopsis Development.** *J Integr Plant Biol* 2008, **50**(7):875-85.
8. Alinsug MV, Yu CW, Wu K: **Phylogenetic analysis, subcellular localization, and expression patterns of RPD3/HDA1 family histone deacetylases in plants.** *BMC Plant Biol* 2009, **9**(1):37.
9. Hao Y, Wang H, Qiao S, Leng L, Wang X: **Histone deacetylase HDA6 enhances brassinosteroid signaling by inhibiting the BIN2 kinase.** *Proc Natl Acad Sci U S A* 2016, **113**(37):10418-23.
10. Yang XJ, Seto E: **HATs and HDACs: from structure, function and regulation to novel strategies for therapy and prevention.** *Oncogene* 2007, **26**(37):5310-8.
11. Liu X, Yang S, Zhao M, Luo M, Yu CW, Chen CY, Tai R, Wu K: **Transcriptional Repression by Histone Deacetylases in Plants.** *Mol Plant* 2014, **7**(5):764-72.
12. Cigliano RA, Cremona G, Paparo R, Termolino P, Perrella G, Gutzat R, Consiglio MF, Conicella C: **Histone deacetylase AtHDA7 is required for female gametophyte and embryo development in Arabidopsis.** *Plant Physiol* 2013, **163**(1):431-40.
13. Wu K, Zhang L, Zhou C, Yu CW, Chaikam V: **HDA6 is required for jasmonate response, senescence and flowering in Arabidopsis.** *J Exp Bot* 2008, **59**(2):225-34.
14. Ma XJ, Yang CP, Xia DA: **Characterization and expression analysis of histone deacetylases family RPD3/HDA1 in Populus trichocarpa.** *Biol Plantarum* 2016, **60**(2):235-43.
15. Krogan NT, Hogan K, Long JA: **APETALA2 negatively regulates multiple floral organ identity genes in Arabidopsis by recruiting the co-repressor TOPLESS and the histone deacetylase HDA19.** *Development* 2012, **139**(22):4180-90.
16. Wang L, Kim J, Somers DE: **Transcriptional corepressor TOPLESS complexes with pseudoresponse regulator proteins and histone deacetylases to regulate circadian transcription.** *Proc Natl Acad Sci U S A* 2013, **110**(2):761-6.
17. Zhou CH, Zhang L, Duan J, Miki B, Wu KQ: **HISTONE DEACETYLASE19 is involved in jasmonic acid and ethylene signaling of pathogen response in Arabidopsis.** *Plant Cell* 2005, **17**(4):1196-204.
18. Varotto S, Locatelli S, Canova S, Pipal A, Motto M, Rossi V: **Expression Profile and Cellular Localization of Maize Rpd3-Type Histone Deacetylases during Plant Development.** *Plant Physiol* 2003, **133**(2):606-17.

19. Zhao J, Li M, Gu D, Liu X, Zhang J, Wu K, Zhang X, Teixeira da Silva JA, Duan J: **Involvement of rice histone deacetylase HDA705 in seed germination and in response to ABA and abiotic stresses.** *Biochem Biophys Res Commun* 2016, **470**(2):439-44.
20. Fu W, Wu K, Duan J: **Sequence and expression analysis of histone deacetylases in rice.** *Biochem Biophys Res Commun* 2007, **356**(4):843-50.
21. Cannon SB, Mitra A, Baumgarten A, Young ND, May G: **The roles of segmental and tandem gene duplication in the evolution of large gene families in *Arabidopsis thaliana*.** *BMC Plant Biol* 2004, **4**(1):10.
22. 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.** *Nat Biotechnol* 2015, **33**(5):531-7.
23. Hu Y, Chen JD, Fang L, Zhang ZY, Ma W, Niu YC, Ju LZ, Deng JQ, Zhao T, Lian JM *et al*: ***Gossypium barbadense* and *Gossypium hirsutum* genomes provide insights into the origin and evolution of allotetraploid cotton.** *Nat Genet* 2019, **51**(4):739-48.
24. Wang Z, Cao H, Chen F, Liu Y: **The roles of histone acetylation in seed performance and plant development.** *Plant Physiol Biochem* 2014, **84**(84):125-33.
25. Tian L, Chen ZJ: **Blocking histone deacetylation in *Arabidopsis* induces pleiotropic effects on plant gene regulation and development.** *Proc Natl Acad Sci U S A* 2001, **98**(1):200-5.
26. Yu CW, Liu X, Luo M, Chen C, Lin X, Tian G, Lu Q, Cui Y, Wu K: **HISTONE DEACETYLASE6 interacts with FLOWERING LOCUS D and regulates flowering in *Arabidopsis*.** *Plant Physiol* 2011, **156**(1):173-84.
27. Luo M, Tai R, Yu CW, Yang S, Chen CY, Lin WD, Schmidt W, Wu K: **Regulation of flowering time by the histone deacetylase HDA5 in *Arabidopsis*.** *Plant J* 2015, **82**(6):925-36.
28. Hu Y, Qin F, Huang L, Sun Q, Li C, Zhao Y, Zhou DX: **Rice histone deacetylase genes display specific expression patterns and developmental functions.** *Biochem Biophys Res Commun* 2009, **388**(2):266-71.
29. Kumar V, Singh B, Singh SK, Rai KM, Singh SP, Sable A, Pant P, Saxena G, Sawant SV: **Role of GhHDA5 in H3K9 deacetylation and fiber initiation in *Gossypium hirsutum*.** *Plant J* 2018, **95**(6):1069-83.
30. Li F, Fan G, Lu C, Xiao G, Zou C, Kohel RJ, Ma Z, Shang H, Ma X, Wu J: **Genome sequence of cultivated Upland cotton (*Gossypium hirsutum* TM-1) provides insights into genome evolution.** *Nat Biotechnol* 2015, **33**(5):524-30.
31. Wu A, Hao P, Wei H, Sun H, Cheng S, Chen P, Ma Q, Gu L, Zhang M, Wang H *et al*: **Genome-Wide Identification and Characterization of Glycosyltransferase Family 47 in Cotton.** *Front Genet* 2019, **10**:824.
32. Lynch M, Conery JS: **The evolutionary fate and consequences of duplicate genes.** *Science* 2000, **290**(5494):1151-5.

33. Cheng S, Chen P, Su Z, Ma L, Hao P, Zhang J, Ma Q, Liu G, Liu J, Wang H *et al*: **High-resolution temporal dynamic transcriptome landscape reveals a GhCAL-mediated flowering regulatory pathway in cotton (*Gossypium hirsutum* L.)**. *Plant Biotechnol J* 2020.
34. Fan L, Chen M, Dong B, Wang N, Yu Q, Wang X, Xuan L, Wang Y, Zhang S, Shen Y: **Transcriptomic Analysis of Flower Bud Differentiation in *Magnolia sinostellata***. *Genes (Basel)* 2018, **9**(4):212.
35. Kim W, Latrasse D, Servet C, Zhou DX: **Arabidopsis histone deacetylase HDA9 regulates flowering time through repression of AGL19**. *Biochem Biophys Res Commun* 2013, **432**(2):394-8.
36. Tian L, Wang J, Fong MP, Chen M, Cao H, Gelvin SB, Chen ZJ: **Genetic control of developmental changes induced by disruption of Arabidopsis histone deacetylase 1 (*AtHD1*) expression**. *Genetics* 2003, **165**(1):399-409.
37. Bilas R, Szafran K, Hnatuszko-Konka K, Kononowicz AK: **Cis -regulatory elements used to control gene expression in plants**. *Plant Cell Tiss Org* 2016, **127**(2):269-87.
38. Chen LT, Wu K: **Role of histone deacetylases HDA6 and HDA19 in ABA and abiotic stress response**. *Plant Signal Behav* 2010, **5**(10):1318-20.
39. Ellis C, Turner J: **The Arabidopsis Mutant *cev1* Has Constitutively Active Jasmonate and Ethylene Signal Pathways and Enhanced Resistance to Pathogens**. *Plant Cell* 2001, **13**(5):1025-33.
40. Anderson JP, Badruzsaufari E, Schenk PM, Manners JM, Desmond OJ, Ehlert C, Maclean DJ, Ebert PR, Kazan K: **Antagonistic Interaction between Abscisic Acid and Jasmonate-Ethylene Signaling Pathways Modulates Defense Gene Expression and Disease Resistance in Arabidopsis**. *Plant Cell* 2004, **16**(12):3460-79.
41. Chen LT, Luo M, Wang Y, Wu K: **Involvement of Arabidopsis histone deacetylase HDA6 in ABA and salt stress response**. *J Exp Bot* 2010, **61**(12):3345-53.
42. Ma XJ, Lv SB, Zhang C, Yang CP: **Histone deacetylases and their functions in plants**. *Plant Cell Rep* 2013, **32**(4):465-78.
43. El-Gebali S, Mistry J, Bateman A, Eddy SR, Luciani A, Potter SC, Qureshi M, Richardson LJ, Salazar GA, Smart A *et al*: **The Pfam protein families database in 2019**. *Nucleic Acids Res* 2019, **47**(D1):D427-32.
44. Du X, Huang G, He S, Yang Z, Sun G, Ma X, Li N, Zhang X, Sun J, Liu M *et al*: **Resequencing of 243 diploid cotton accessions based on an updated A genome identifies the genetic basis of key agronomic traits**. *Nat Genet* 2018, **50**(6):796-802.
45. 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):423-7.
46. Wang M, Tu L, Yuan D, Zhu, Shen C, Li J, Liu F, Pei L, Wang P, Zhao G *et al*: **Reference genome sequences of two cultivated allotetraploid cottons, *Gossypium hirsutum* and *Gossypium barbadense***. *Nat genet* 2019, **51**(2):224-9.
47. Zhu T, Liang C, Meng Z, Sun G, Meng Z, Guo S, Zhang R: **CottonFGD: an integrated functional genomics database for cotton**. *BMC Plant Biol* 2017, **17**(1):101.

48. Finn RD, Clements J, Eddy SR: **HMMER web server: interactive sequence similarity searching.** *Nucleic Acids Res* 2011, **39**(Web Server issue):W29-37.
49. Cheng CY, Krishnakumar V, Chan AP, Thibaud-Nissen F, Schobel S, Town CD: **Araport11: a complete reannotation of the Arabidopsis thaliana reference genome.** *Plant J* 2017, **89**(4):789-804.
50. 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 Res* 2012, **40**(Database issue):D1178-86.
51. Letunic I, Doerks T, Bork P: **SMART: recent updates, new developments and status in 2015.** *Nucleic Acids Res* 2015, **43**(Database issue):D257-60.
52. 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 Res* 2019, **47**(W1):W636-41.
53. Wang JR, Sung WK, Krishnan A, Li KB: **Protein subcellular localization prediction for Gram-negative bacteria using amino acid subalphabets and a combination of multiple support vector machines.** *BMC Bioinformatics* 2005, **6**(1):174.
54. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R *et al.*: **Clustal W and Clustal X version 2.0.** *Bioinformatics* 2007, **23**(21):2947-8.
55. Kumar S, Stecher G, Tamura K: **MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets.** *Mol Biol Evol* 2016, **33**(7):1870-4.
56. 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* 2018:289660.
57. Hu B, Jin J, Guo AY, Zhang H, Luo J, Gao G: **GSDS 2.0: an upgraded gene feature visualization server.** *Bioinformatics* 2015, **31**(8):1296-7.
58. Bailey TL, Johnson J, Grant CE, Noble WS: **The MEME Suite.** *Nucleic Acids Res* 2015, **43**(W1):W39-49.
59. 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 Res* 2012, **40**(7):e49.
60. Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, Jones SJ, Marra MA: **Circos: an information aesthetic for comparative genomics.** *Genome Res* 2009, **19**(9):1639-45.
61. Zhu Y, Wu NN, Song WL, Yin GJ, Qin YJ, Yan YM, Hu YK: **Soybean (Glycine max) expansin gene superfamily origins: segmental and tandem duplication events followed by divergent selection among subfamilies.** *BMC Plant Biol* 2014, **14**(1):93.
62. Hurst LD: **The Ka/ Ks ratio: diagnosing the form of sequence evolution.** *Trends Genet* 2002, **18**(9):486-7.
63. Sun H, Hao P, Ma Q, Zhang M, Qin Y, Wei H, Su J, Wang H, Gu L, Wang N *et al.*: **Genome-wide identification and expression analyses of the pectate lyase (PEL) gene family in cotton (Gossypium**

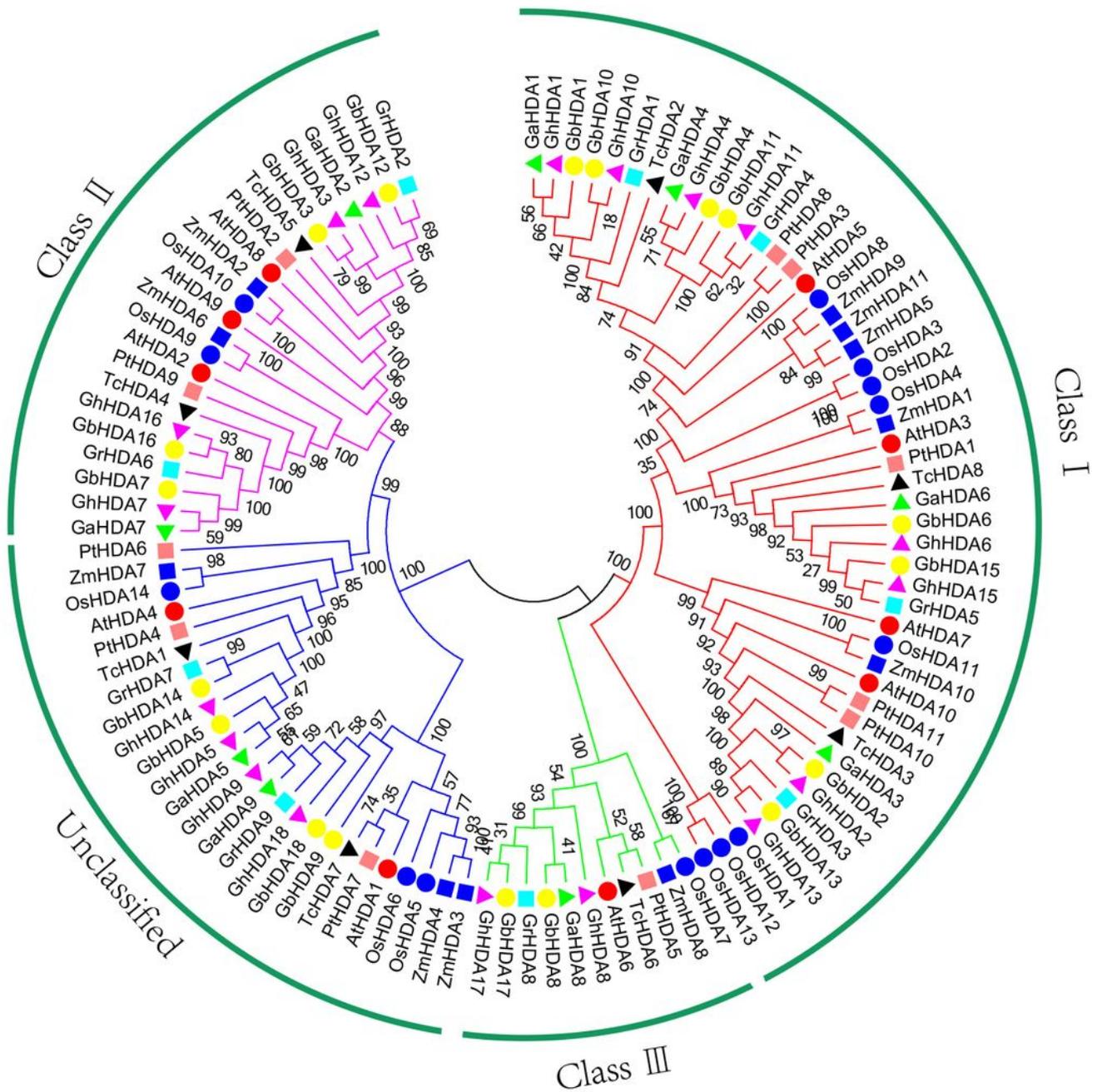
- hirsutum* L.). *BMC Genomics* 2018, **19**(1):661.
64. Lescot M, Dehais P, Thijs G, Marchal K, Moreau Y, Van de Peer Y, Rouze P, Rombauts S: **PlantCARE, a database of plant cis-acting regulatory elements and a portal to tools for in silico analysis of promoter sequences.** *Nucleic Acids Res* 2002, **30**(1):325-7.
65. Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL: **TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions.** *Genome Biol* 2013, **14**(4):R36.
66. Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, Salzberg SL, Wold BJ, Pachter L: **Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation.** *Nat Biotechnol* 2010, **28**(5):511-5.
67. Deng W, Wang Y, Liu Z, Cheng H, Xue Y: **HemI: A Toolkit for Illustrating Heatmaps.** *PloS One* 2014, **9**(11):e111988.
68. He P, Yang Y, Wang Z, Zhao P, Yuan Y, Zhang L, Ma Y, Pang C, Yu J, Xiao G: **Comprehensive analyses of ZFP gene family and characterization of expression profiles during plant hormone response in cotton.** *BMC Plant Biol* 2019, **19**(1):329.
69. Tu LL, Zhang XL, Liu DQ, Jin SX, Cao JL, Zhu LF, Deng FL, Tan JF, Zhang CB: **Suitable internal control genes for qRT-PCR normalization in cotton fiber development and somatic embryogenesis.** *Chinese Sci Bull* 2007, **52**(22):3110-7.
70. Livak KJ, Schmittgen TD: **Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method.** *Methods* 2001, **25**(4):402-8.

## Table

**Table 1** Physicochemical parameters of 18 *RPD3* genes in *G. hirsutum*

Name	Gnen ID	Protein Length	Protein pI	Protein MW(kD)	Subcellular localization
<i>GhHDA1</i>	<i>Ghir_A01G001410.1</i>	499	4.9676	56.18	Nuclear
<i>GhHDA2</i>	<i>Ghir_A03G007210.1</i>	471	5.076	53.09	Nuclear
<i>GhHDA3</i>	<i>Ghir_A03G008200.1</i>	655	5.325	73.01	Cytoplasmic
<i>GhHDA4</i>	<i>Ghir_A03G018610.1</i>	351	4.4737	39.55	Cytoplasmic/Nuclear
<i>GhHDA5</i>	<i>Ghir_A05G039610.1</i>	449	6.9085	48.66	Mitochondrial/Chloroplast
<i>GhHDA6</i>	<i>Ghir_A09G010210.1</i>	429	4.8969	49.08	Cytoplasmic/Nuclear
<i>GhHDA7</i>	<i>Ghir_A12G027820.1</i>	574	6.3108	63.26	Cytoplasmic
<i>GhHDA8</i>	<i>Ghir_A13G019980.1</i>	232	6.5919	25.79	Plasma Membrane
<i>GhHDA9</i>	<i>Ghir_A13G023460.1</i>	368	5.3373	40.37	Cytoplasmic/Chloroplast
<i>GhHDA10</i>	<i>Ghir_D01G001410.1</i>	499	4.9676	56.26	Nuclear
<i>GhHDA11</i>	<i>Ghir_D02G019970.1</i>	465	5.1309	52.65	Nuclear
<i>GhHDA12</i>	<i>Ghir_D03G010660.1</i>	635	4.8889	71.02	Cytoplasmic
<i>GhHDA13</i>	<i>Ghir_D03G011510.1</i>	471	5.1489	53.06	Nuclear
<i>GhHDA14</i>	<i>Ghir_D04G003510.1</i>	443	6.9591	47.95	Chloroplast/Mitochondrial
<i>GhHDA15</i>	<i>Ghir_D09G009940.1</i>	429	4.8371	49.11	Cytoplasmic/Nuclear
<i>GhHDA16</i>	<i>Ghir_D12G027930.1</i>	579	6.1788	63.80	Cytoplasmic
<i>GhHDA17</i>	<i>Ghir_D13G020760.1</i>	331	8.6517	37.28	Plasma Membrane
<i>GhHDA18</i>	<i>Ghir_D13G024090.1</i>	380	5.648	41.63	Cytoplasmic

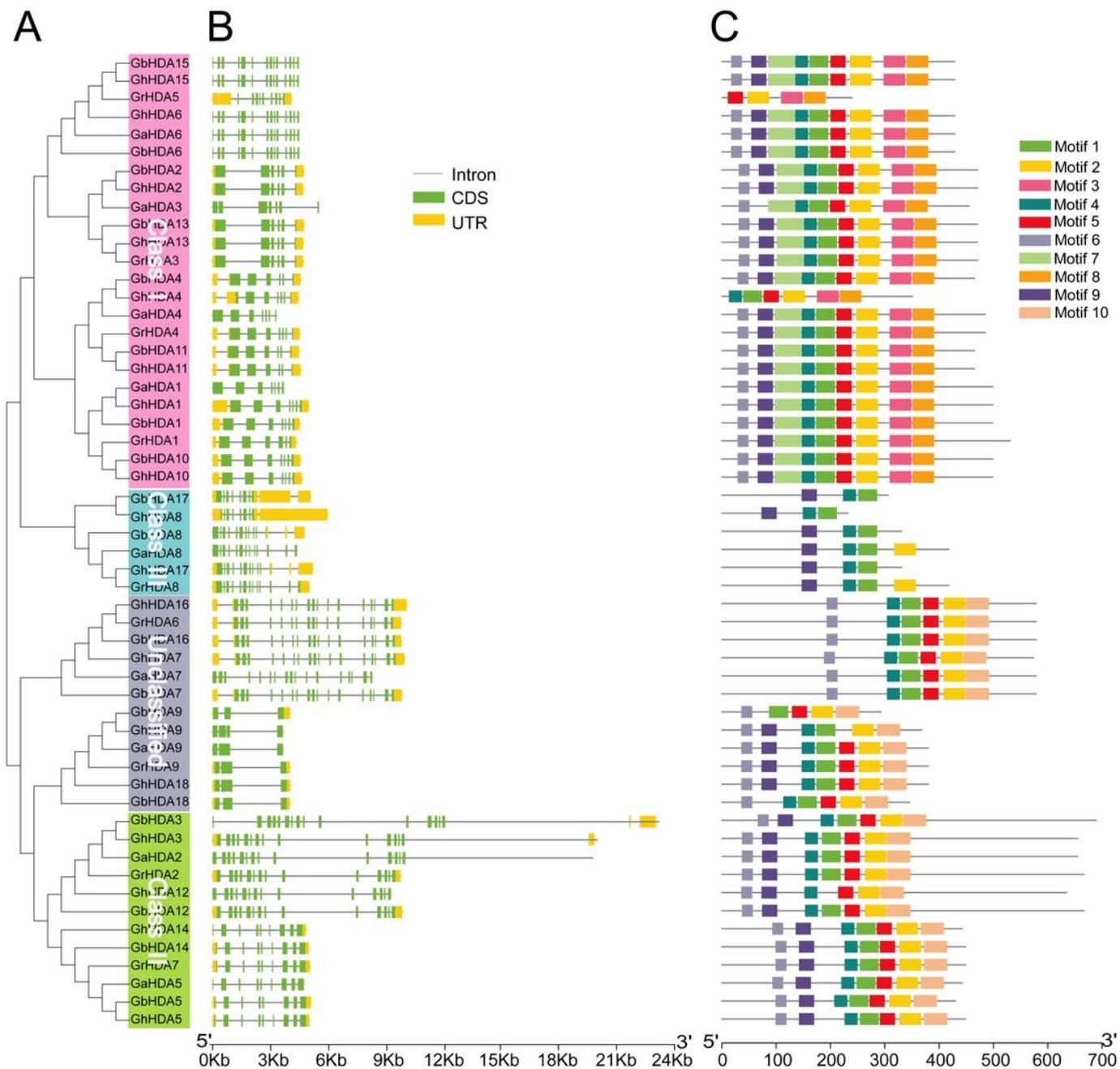
## Figures



- *Gossypium barbadense*
- *Gossypium raimondii*
- ▲ *Gossypium arboreum*
- *Arabidopsis thaliana*
- *Zea mays L.*
- ▲ *Gossypium hirsutum*
- *Oryza sativa L.*
- *Populus trichocarpa*
- ▲ *Theobroma cacao*

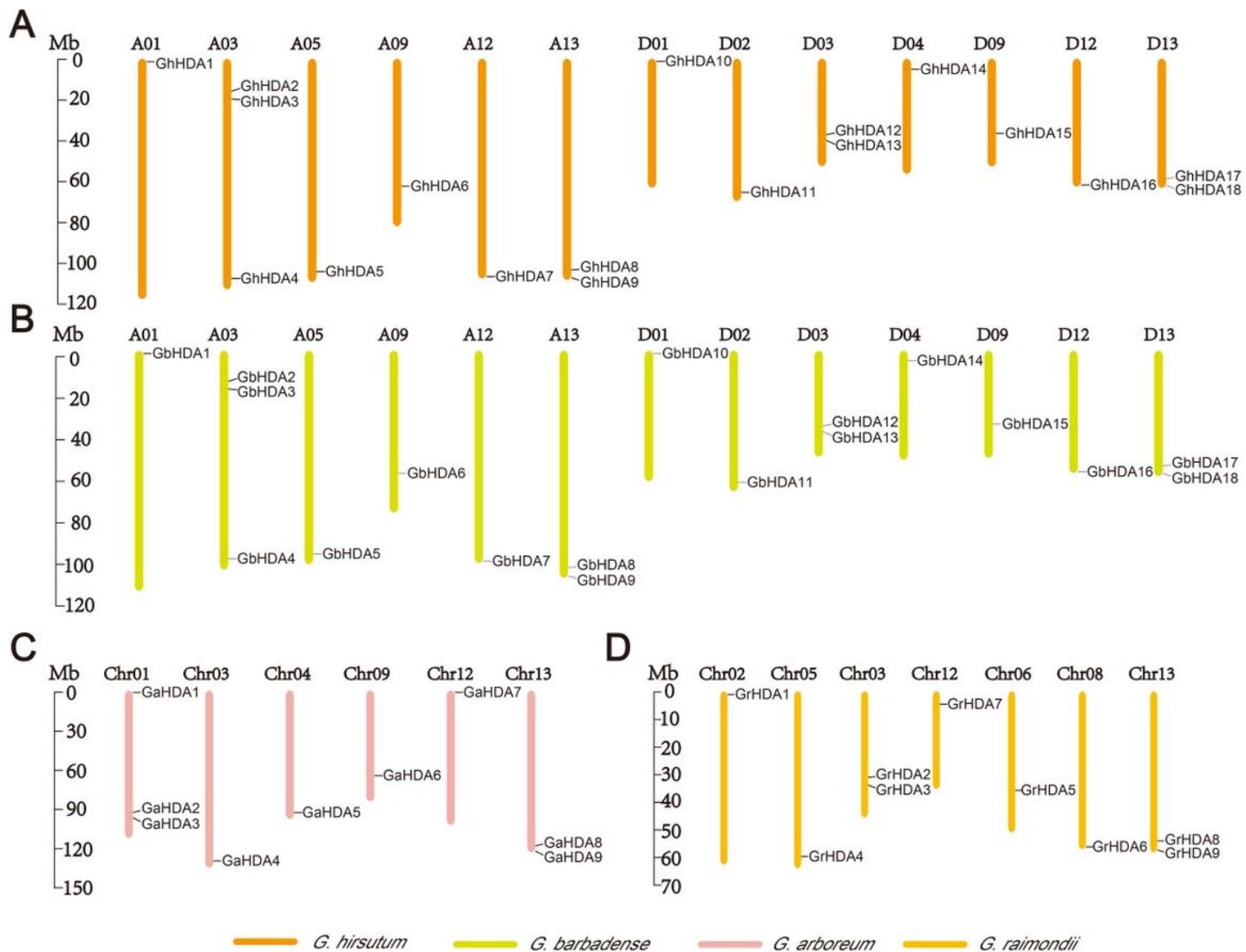
**Figure 1**

Neighbor-joining phylogenetic tree of RPD3 gene family. The 108 predicted RPD3 proteins from *G. hirsutum*, *G. arboreum*, *G. barbadense*, *G. raimondii*, *A. thaliana*, *P. trichocarpa*, *T. cacao*, *Oryza sativa*, and *Zea mays* were aligned using ClustalW, and the neighbor-joining (NJ) method was used to construct this unrooted phylogenetic tree using MEGA 7.0 program with 1000 bootstrap repetitions. Four subfamilies are represented by the different colored lines.



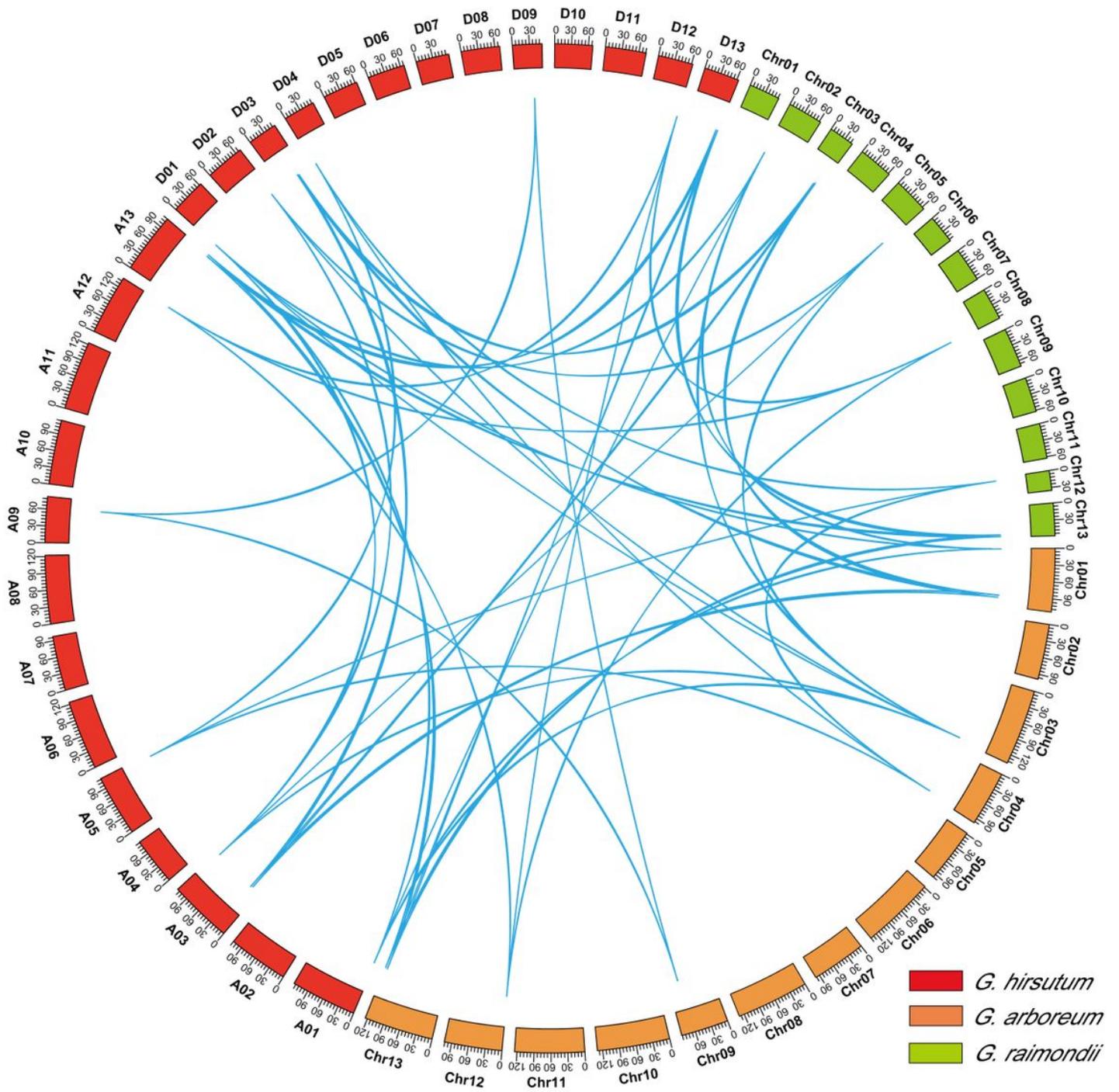
**Figure 2**

Phylogenetic relationships, exon-intron structure, and conserved motif analysis of cotton RPD3 genes. (a) A neighbor-joining phylogenetic tree of 54 cotton RPD3 genes was generated using the MEGA7.0 program; (b) Exon-intron structure analysis of 54 cotton RPD3 genes. The UTRs, exons, and introns are represented with yellow boxes, green boxes, and black lines, respectively; (c) The 10 conserved protein motifs of RPD3 genes are indicated by different colored boxes.



**Figure 3**

Chromosomal distribution of cotton RPD3 genes. a, b, c and d represent the chromosomal location of RPD3 genes from *G. hirsutum* (a), *G. barbadense* (b), *G. arboreum* (c), and *G. raimondii* (d), respectively. The chromosome number is shown on the top of each chromosome. The scale bars represent the length in mega bases (Mb).



**Figure 4**

RPD3 homologous gene pairs among *G. arboreum*, *G. raimondii* and *G. hirsutum*. Orange, blue and red represent chromosomes of *G. arboreum*, *G. raimondii* and *G. hirsutum*, respectively.

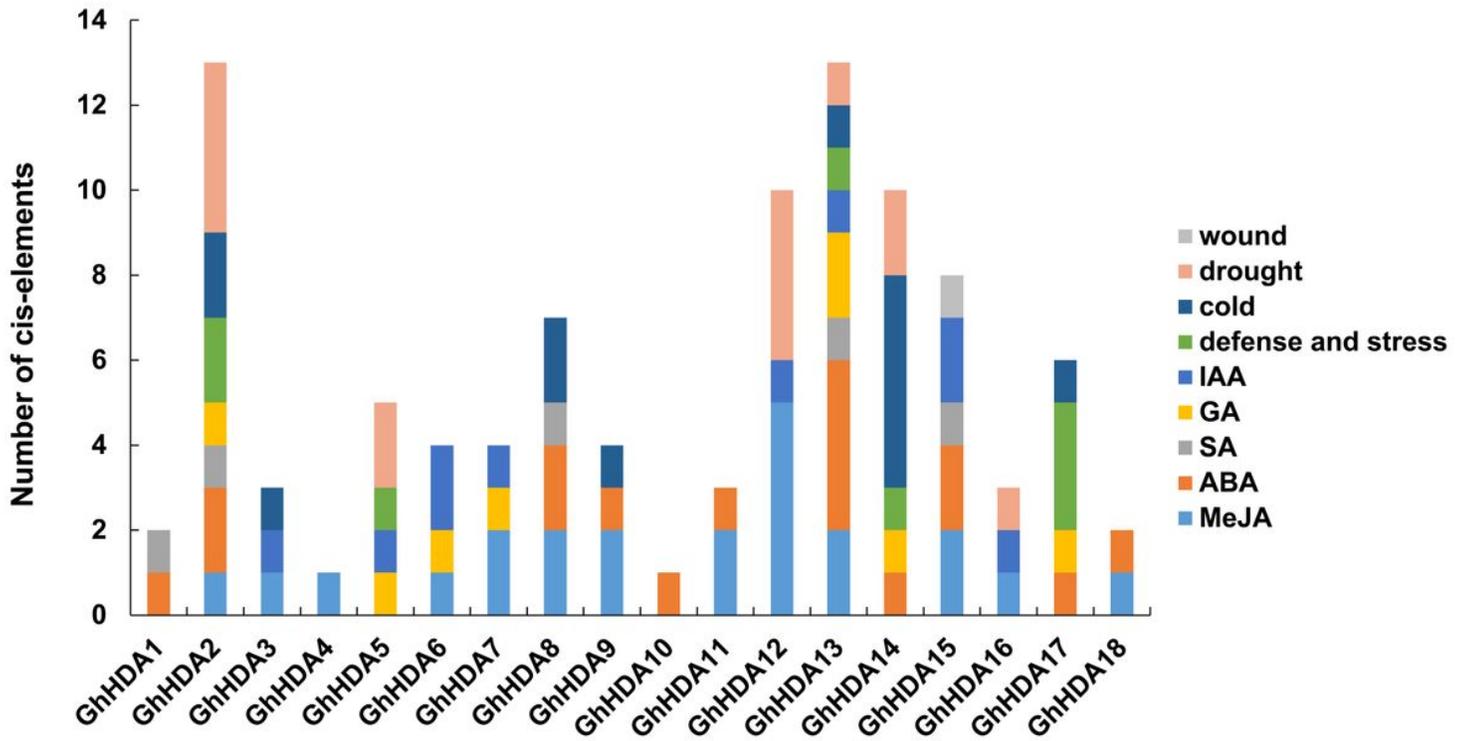
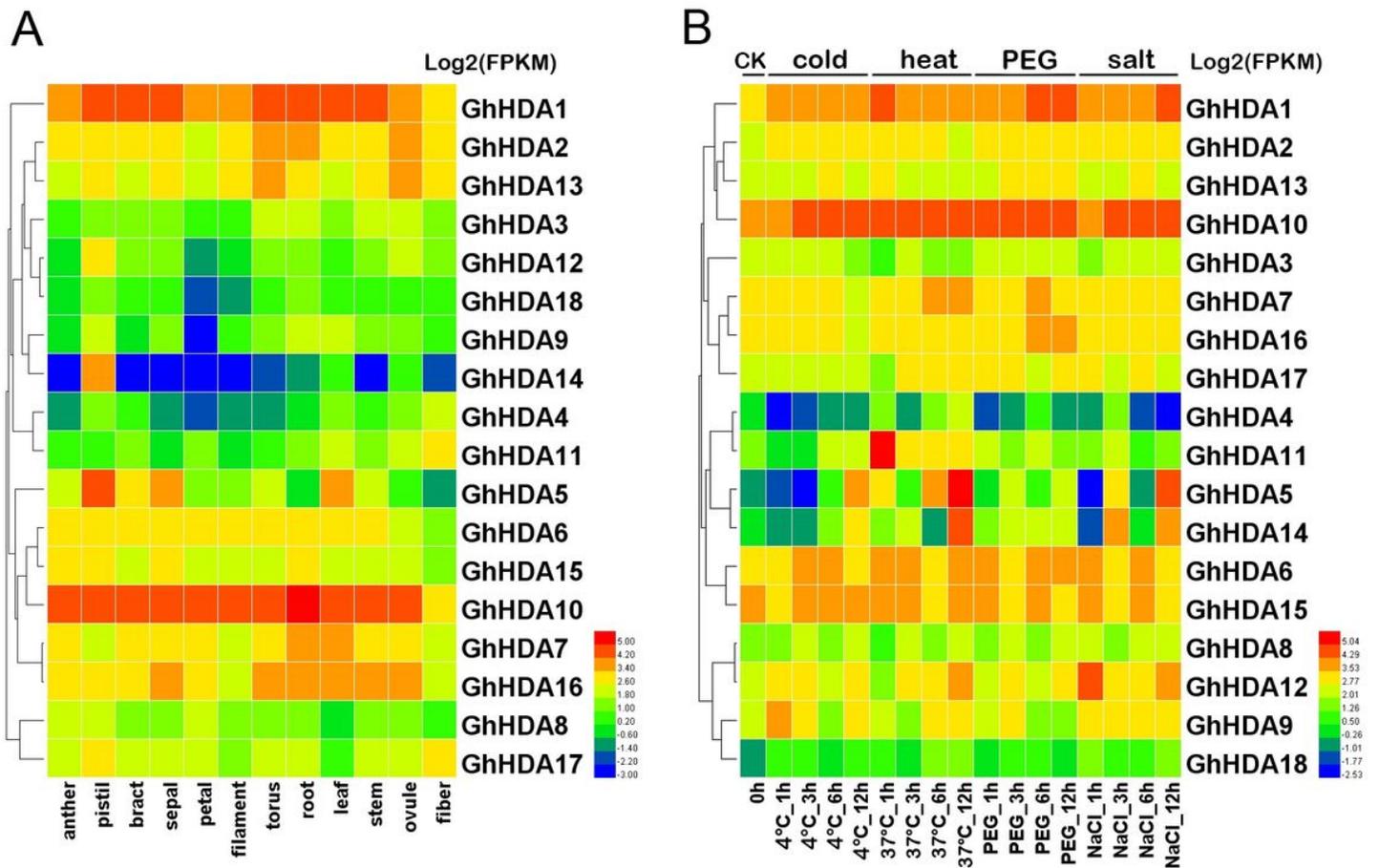


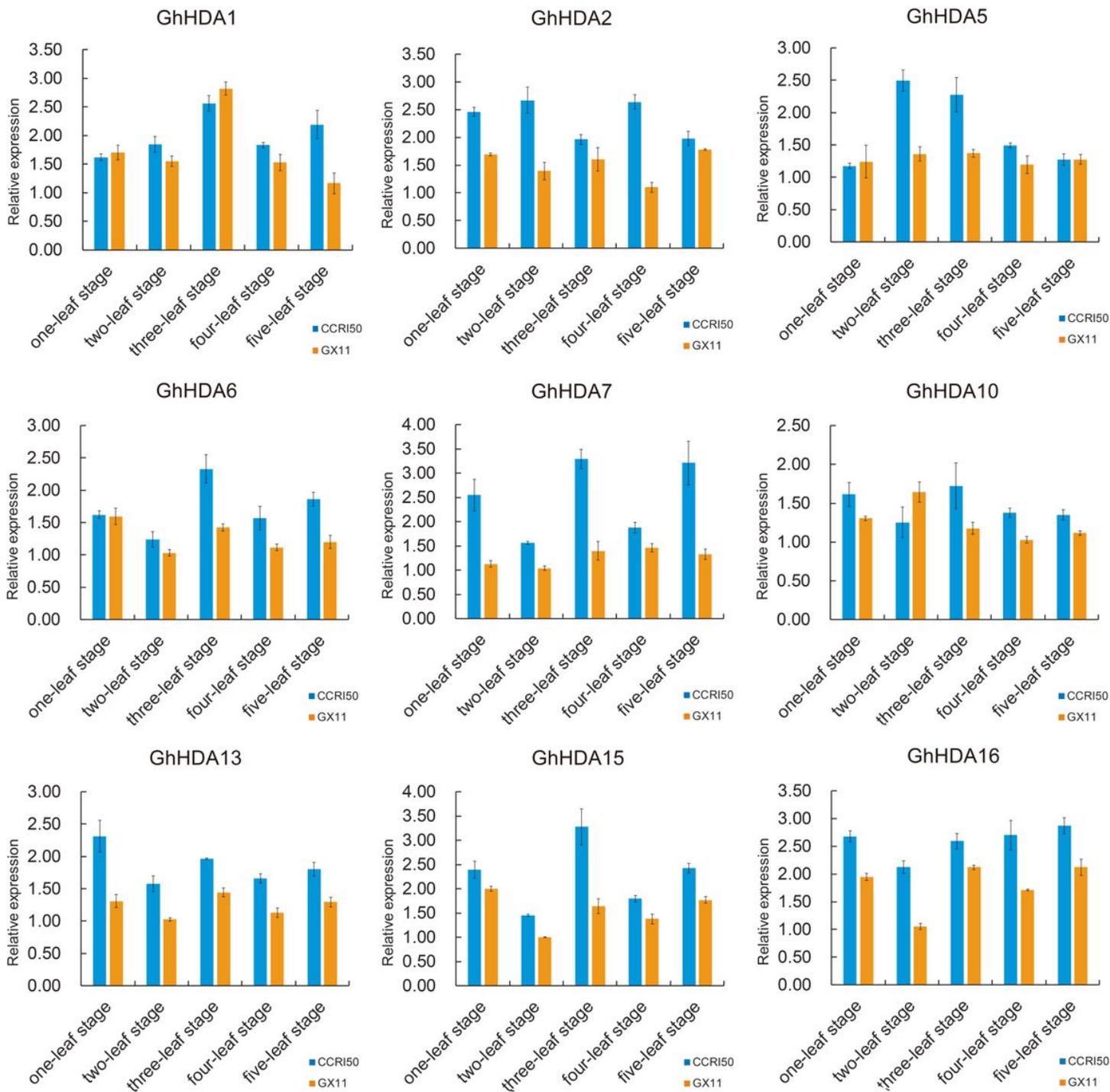
Figure 5

Cis-elements of GhRPD3 genes in promoter regions. The numbers of different cis-elements are presented in the form of bar graphs, and similar cis-elements are exhibited with the same colors.



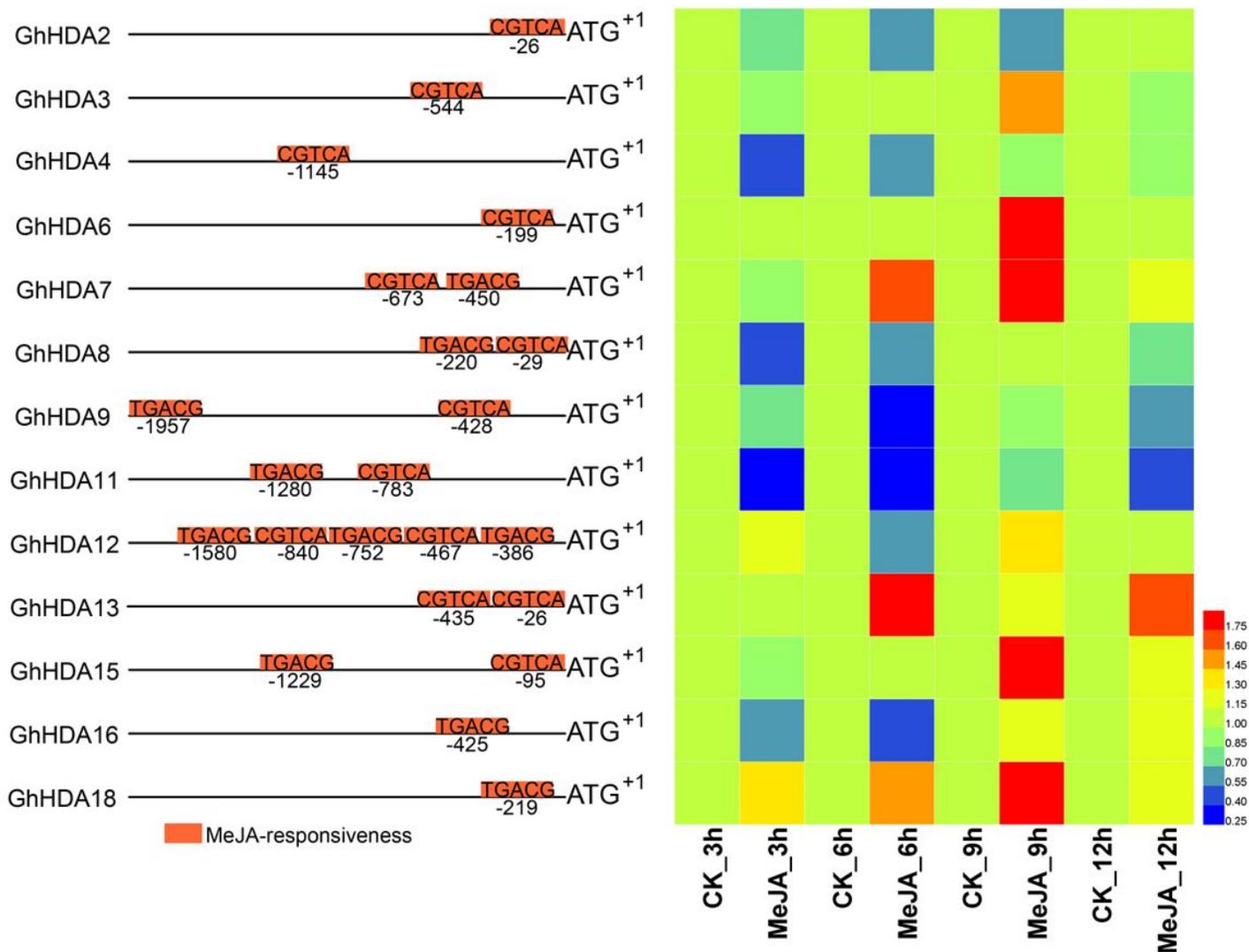
**Figure 6**

Expression patterns of RPD3 genes in *G. hirsutum*. a and b represent the expression patterns of GhRPD3 genes in different tissues (a) and under four different abiotic stresses (b), respectively. Gene names are shown on the right. Scale bars on the right represent the log2-transformed FPKM values of each gene.



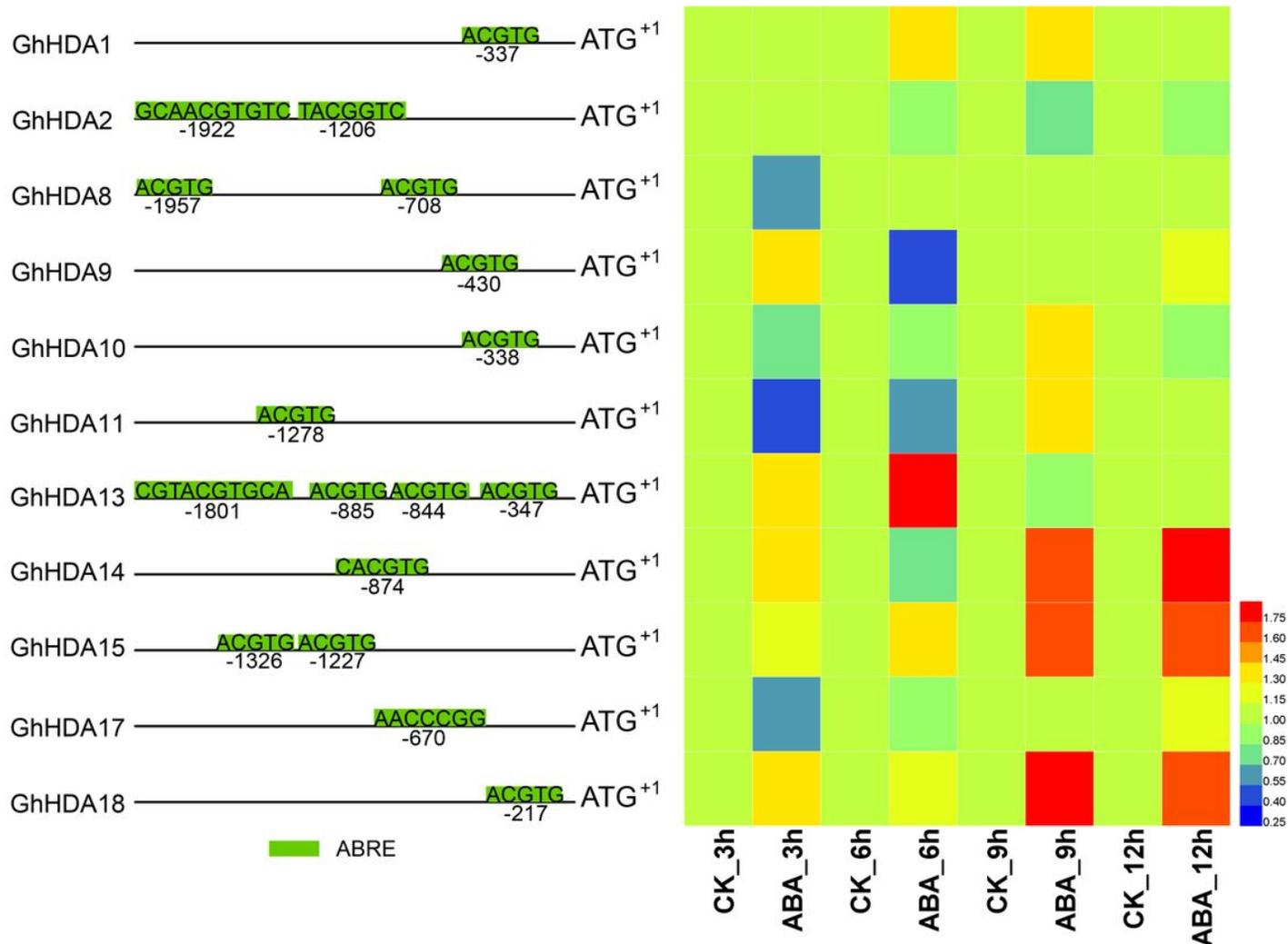
**Figure 7**

Expression levels of 9 GhRPD3 genes between CCR150 and GX11. Blue and orange bar graphs indicate the expression of early-maturity cotton (CCR150) and late-maturity cotton (GX11), respectively. The error bars show the standard deviation of three biological replicates.



**Figure 8**

Expression profiles of 13 GhRPD3 genes under MeJA treatment. Orange boxes represent the MeJA-responsive elements of 13 GhRPD3 genes in the promoter regions (left). The expression changes of 13 GhRPD3 genes under MeJA treatment are shown using a heatmap (right). qRT-PCR was carried out with three technical and three biological replicates. Relative expression levels of each gene were calculated after normalizing the expression level in CK (water) to 1.0.



**Figure 9**

Expression patterns of 11 GhRPD3 genes under ABA treatment. Green boxes represent the ABRE of 11 GhRPD3 genes in the promoter regions (left). The expression changes of 11 GhRPD3 genes under ABA treatment are shown using a heatmap (right). qRT-PCR was conducted with three technical and three biological replicates. Relative expression levels of each gene were calculated after normalizing the expression level in CK (water) to 1.0.

## Supplementary Files

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

- [TableS1.xlsx](#)
- [TableS2.xlsx](#)
- [FigureS1.tif](#)
- [TableS3.xlsx](#)

- [TableS4.xlsx](#)
- [TableS5.xlsx](#)
- [TableS6.xlsx](#)
- [FigureS2.tif](#)
- [TableS7.xlsx](#)
- [TableS8.xlsx](#)