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
Chinese Academy of Agricultural Sciences Cotton Research Institute
Shuxun Yu (ysx195311@163.com)
Chinese Academy of Agricultural Sciences Cotton Research Institute https://orcid.org/0000-0002-9715-3462

Research article

Keywords: Cotton, RPD3/HDA1 gene family, Expression patterns, Abiotic stress, Early maturity

Posted Date: December 19th, 2019

DOI: https://doi.org/10.21203/rs.2.19331/v1

Version of Record: A version of this preprint was published 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. As the superfamily of HDACs, RPD3/HDA1-type HDACs were most studied and reported that RPD3 genes played crucial roles in plant growth and physiological processes. However, there is a lack of systematic research on RPD3/HDA1 gene family in cotton.

Results: In this research, 9, 9, 18 and 18 RPD3 genes were determined by genome-wide analysis 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 motifs analysis exhibited high conservation in each branch of cotton RPD3 genes. Collinearity analysis indicated that segmental duplication was the primary driving force during the expansion of the GhRPD3 gene family. There was at least one presumed cis-element related to plant hormone existing in the promoter regions of all the GhRPD3 genes, especially MeJA and ABA responsive elements, owning more members than other hormone-relevant elements. Characterizations of expression patterns showed that most GhRPD3 genes performed relative high expression in floral organs and possessed the higher expression in early-maturity cotton compared with the late-maturity cotton during flower bud differentiation period. In addition, the expression of GhRPD3 genes could be significantly induced by one or more abiotic stresses as well as exogenous application of MeJA and ABA.

Conclusions: Our findings revealed that GhRPD3 genes might be involved in flower bud differentiation and resistance to abiotic stresses, which provided 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 is combined with nuclear proteins to constitute the Chromatin that is in charge of storing genetic and directive information in eukaryotic cells. Chromatin is highly arranged and mainly composed of nucleosomes, which is formed by approximately 147 bp of DNA and an octamer organized by 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 containing histone modification and DNA methylation play an important role in the regulation of gene expression. In general, histone post-translational modifications occur at the N-terminal of histones, including methylation, acetylation, phosphorylation, ADP-ribosylation and ubiquitination [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 an invertible process and plays essential roles in the epigenetic regulation. The acetylation of core histones is catalyzed by histone acetyltransferases (HATS), promoting the transcriptional activation, while deacetylation is regulated by histone deacetylases (HDACs) related to the transcriptional suppression [5].

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

HDACs are involved in a large amount of biological processes associated with plant growth and development [7–9]. Depending on sequence homology analysis to yeast HDACs, HDACs in plant 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, homologous to yeast RPD3 and HDA-1, belong to a large family and they require zinc ions to catalyze activity, while an HDAC depressor TSA (trichostatin A) or sodium butyrate can inhibit their enzyme 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 class III includes HDA2 only [7, 8]. The other genes of PRD3/HDA1 family are classified to unclassified.

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 growth and development processes of plants and responding 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, seed development and so on [15, 16]. Otherwise, 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 three ZmPRD3 genes, ZmRpd3/101, ZmRpd3/102 and ZmRpd3/108, showed that they expressed widely in all investigated corn organs. Furthermore, their products could be detected in all cellular parts at some specific stages such as kernel, shoot and anther developmental periods [18]. In rice, HDA705 could respond to ABA and abiotic stresses and its expression was induced by JA. In addition, the expression of HDA702 and HDA704 was also significantly induced by SA, JA and ABA [19, 20]. These findings indicated that the RPD3 members played an important regulatory role in plant development and response to various stresses and plant hormone.

Cotton is one of the most important economic crops in our country and plays an essential role in China's 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 made some progress in Arabidopsis and some other crops. However, there is a lack of systematic research on RPD3 gene family in cotton. Hence, 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 their phylogenetic tree, gene structure, conserved motifs, protein domains, expression profiles and preliminary functions were comprehensively analyzed. The information about 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 determined after eliminating redundant sequences and they are named by the position on the chromosome. The corresponding relationship between gene ID number and their names was shown in Additional file 1: Table S1. A total of 18 genes (GhHDA1-GhHDA18) containing Hist_deacetyl (PF00850) domains were identified from G. hirsutum, among which 9 genes were located on the At genome and 9 genes were mapped on the Dt genome. Further, 18 genes (GbHDA1-GbHDA18) from G. barbadense, 9 genes (GaHDA1-GaHDA9) from G. aboreum and 9 genes (GrHDA1-GrHDA9) from G. raimondii were detected, repectively. 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 (Arabiposis), 14 (Oryza satica L.), 11 (Populus trichocarpa), 8 (Theobroma cacao) and 11 (Zea mays L.), respectively. The GhRPD3 proteins length ranged from 232 to 635 aa with an average of 459aa. The physicochemical parameters analysis showed that the isoelectric point (pl) 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) and nuclear (8), suggesting that GhRPD3 genes might possess multiple regulatory functions (Table 1). The predicted length, pl, Mw and subcellular localization of the RPD3 proteins in other eight species were shown in Additional file 1: Table S1.

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

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

Phylogenetic analysis of the RPD3 gene family

A total of identified 108 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 subgroup was the largest subfamily with 49 RPD3 genes, while the Class subgroup has the fewest members, only containing one gene in seven diploid species genomes and including two genes in two tetraploid cotton genomes (Fig. 1). Among these four classes,

each subfamily contained all nine species RPD3 genes, indicating this gene family was relatively conserved in different species during evolutionary process.

Exon-intron structure and conserved motif analysis

The domains of RPD3 sequences in cotton were investigated and exhibited according to the results of SMART database using TBtools, showing 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), then exon-intron structure (Fig. 2b) and conserved motifs (Fig. 2c) were analyzed to better understand the similarity and difference of cotton RPD3 members. The results showed that the gene length of RPD3 cotton genes was relatively conserved in Class Ⅰ and Class Ⅱ, but there were twelve longer sequences existing in Class Ⅰ and unclassified group. The RPD3 cotton genes included lots of exons varying from 3 to 17 and most RPD3 genes (48/54) contained more than five exons (Additional file 4: Table S3), which might be associated with diversification of their functions. In terms of distributions of motifs, most RPD3 cotton genes belonging to the same subfamily performed a similar motif composition except for the unclassified group (Fig. 2c). Most Class Ⅱsubfamily members contained 9 motifs in addition to GrHDA5 and GhHDA4, only including 4 and 6 motifs respectively. Class Ⅰ subfamily genes had three or four motifs and most Class Ⅱ subfamily members possessed 7 motifs except for GhHDA12, one motif less than other genes. There were differences existing in the exon-intron structure and motif arrangement among four categories, while 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 exhibited 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), while the other 12 chromosomes only contained one or two GhRPD3 genes (Fig. 3a). The distribution of 18 GbRPD3 genes on chromosomes in G. barbadense was similar with 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, while the other 4 chromosomes contained only one GaRPD3 genes (Fig. 3c). In G. raimondii, the chromosomal distribution of 9 GrRPD3 genes were highly consistent with the corresponding D sub-genome 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 occured during the progress of 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 were two main reasons of gene family generation in the evolutionary process of gene family [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. arboretum, G. raimondii, and G. hirsutum was visualized using circular maps (Fig. 4). The Ka/Ks ratios of most homologous gene pairs were lower than one, indicating that purified selection was essential during the evolution of cotton RPD3 genes, while the Ka/Ks ratios of gene pairs (GhHDA2/GaHDA3 and GhHDA6/GaHDA6) were more than 1, suggesting that these two gene pairs might have experienced positive selection pressure. The study also predicted the occurrence time of segmentally duplicated RPD3 gene pairs by the formula " $t = Ks/2r$" ($r = 2.6 \times 10^{-9}$) [22]. Except for gene pair (GhHDA6/GaHDA6), the other segmental duplication events of three cotton species might have occurred 0.6 to 144.44 million years ago with an average time of 18.39 million years (Additional file 6: Table S5).

Analysis of cis-elements in predicted promoter regions of GhRPD3

In order to explore the possible regulatory functions of GhRPD3 genes under various environmental stresses and hormone regulation pathway, the 2000 bp promoter regions of 18 GhRPD3 genes were employed to the PlantCARE database for identification of putative stress-associated and plant hormone-related cis-elements. A total of 9 kinds of elements related to plant hormone: AuxRR-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: TC-rich repeats (defense and stress responsiveness), MBS(drought), WUN-motif (wound) and LTR (cold stress), were detected in the predicted 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 AuxRR-core, 2 GARE-motif, 1 TCA-element, 4 ABRE, 2 TGACG-motif). Among all the 18 GhRPD3 genes, there are large numbers of light-responsive elements distributing 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). The 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 GhPRD3 genes in the growth and development of cotton, we studied their expression in various cotton tissues containing 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) ≥ 1. Furthermore, ten genes exhibited high expression in all the selected tissues (Additional file 9: Table S7). The results indicated that GhRPD3 genes were widely expressed in both reproductive organs and vegetative organs and might have multiple biological functions. After log2-conversion of FPKM values, the expression profiles of GhRPD3 genes in different tissues were shown in Fig. 6a. Seven GhRPD3 genes exhibited relative high expression in at least eight tissues (log2-

transformed FPKM value ≥ 2.6), especially a pair of homologous genes (GhHDA1/GhHDA10), performing higher expression level at all the tissues and showing the same expression pattern. Nevertheless, three GhRPD3 genes (GhHDA4, GhHDA14, GhHDA18) revealed relative low expression in at least eight tissues (log2-transformed FPKM value < 1), of which GhHDA14 showed the lower expression at all tissues except for 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 relative low expression at all these twelve tissues. Gene pair (GhHDA2/GhHDA13) exhibited relative high expression in torus and ovule, while showed relatively low expression in petal (Fig. 6a).

Based on the analysis of cis-elements in promoter regions and previous reports on RPD3 genes in other plants, GhRPD3 gens might respond to abiotic stresses. To verify 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 results indicated that most GhRPD3 genes could be induced by all four stresses to varying degrees, especially these six genes (GhHDA1, GhHDA2, GhHDA6, GhHDA10, GhHDA12 and GhHDA18), significantly up-regulated at every stage of the treatment. However, only one gene (GhHDA4) exhibited markedly down-regulated expression under four abiotic stresses. Some genes were induced by a particular treatment, for example, GhHDA13 and GhHDA16 was significantly induced by PEG treatment because of containing drought-responsive elements in their predicted promotors. Four genes (GhHDA7, GhHDA17, GhHDA11, GhHDA5) showed up-regulated expression under heat treatment. The expression of GhHDA9 was significantly up-regulated under cold and salt treatments. According to the results, we can make a conclusion that GhRPD3 genes play an essential role in response to abiotic stresses.

Characterization of GhRPD3 genes expression during flower bud differentiation period

To explore the expression difference of GhRPD3 genes between the early-maturity and late-maturity cottons in flower bud differentiation period, we selected nine genes showing relatively high expression in tissues (anther, pistil, bract, sepal, petal, filament and torus) that made up the floral organs for qRt-PCR. The buds of early-maturity variety (CCRI50) and late-maturity variety (GX11) from one-leaf stage to five-leaf stage were used for qRt-PCR respectively (Fig. 7). The results revealed that more than half of these genes (5/9) possessed relatively higher expression in early-maturity cotton than that in late-maturity cotton during flower bud differentiation period. GhHDA5 showed marked difference at two-leaf stage and three-leaf stage 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 same expression trend that both of them presented the highest expression at three-leaf stage and then exhibited down-regulated expression at next two stages in CCRI50. In addition, all the nine genes performed relatively high expression at two-leaf stage or three-leaf stage in CCRI50 compared with GX11. The results showed that GhRPD3 genes were associated with 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 responsiveness elements in the predicted promotors to analyze the expression characteristics of those genes under MeJA and ABA treatment by qRT-PCR (Fig. 8 and Fig. 9). Most GhRPD3 genes (8/13) were markedly up-regulated at 9 h after MeJA treatment. Three genes (GhHDA7, GhHDA13 and GhHDA18) exhibited significantly up-regulated expression at least three time points, while four genes (GhHDA2, GhHDA8, GhHDA9 and GhHDA11) showed marked transcriptional down-regulation at least three time points after MeJA treatment (Fig. 8). More than half of GhRPD3 genes (6/11) were significantly up-regulated at 9 h after ABA treatment. Three GhRPD3 genes (GhHDA14, GhHDA15 and GhHDA18) performed relatively high expression at least three time points, while three GhRPD3 genes (GhHDA10, GhHDA11, GhHDA17) showed early down-regulated and then up-regulated expression patterns under ABA treatment (Fig. 9). The results showed that exogenous application of MeJA and ABA significantly induced the transcriptional levels of most GhRPD3 genes containing MeJA-responsiveness 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 contains acetylation and deacetylation, which were 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]. As the superfamily of HDACs, RPD3 gene family was most studied and crucial in plant development and physiological processes, including flowering time, abiotic stress response, female gametophyte and embryo development, senescence, seed germination, plant hormone signals response and so on [12–14, 19, 26–28]. To date, although there were a few studies on function analysis of the RPD3 family members in G. hirsutum, it mainly focused on roles GhHDA5 played in fiber initiation, lack of systematic report [29]. In order to explore 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 RPD3 gene family in cotton, containing their phylogenetic relationships, exon-intron structures, conserved motifs, chromosomal distributions, duplication events, expression patterns in different tissues as well as under abiotic stresses and response to 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 determined 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 allotetraploid, which was not 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]. In order 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 diversification of their functions. Notably, the gene structure and motif arrangement were different among four subfamilies, while 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 cotton RPD3 gene family.

According to the evolutionary history of cotton, tetraploid cotton was formed by the hybridization of two diploid cottons and the following polyploidization [31]. To investigate the evolutionary relationships of predicted RPD3 genes between two diploid genomes and sub-genomes in allotetraploid, 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 sub-genome of G. hirsutum was not identical while the chromosomal location of RPD3 genes in G. raimondii and the corresponding D sub-genome of G. hirsutum was highly consistent (Fig. 3), illustrating the high conservation during the evolution of RPD3 gene family. The analysis of gene duplication showed that segmental duplication was essential to 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 at 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), accompanying with the divergence of A and D ancestral genomes. Ka/Ks ratios analysis indicated that the segmentally duplicated gene pairs might preform similar functions on account of purified selection in functional segregation [32].

Functions analysis of GhRPD3 genes in upland cotton

Flowering time, an important indicator of early-maturity cotton, was also influenced by the time of flower bud differentiation, which is a physiological and morphological marker of the transformation from vegetative growth to reproductive growth [33, 34]. The previous studies showed that flower bud differentiation period of early-maturity cotton variety was earlier than that of late-maturity cotton variety, and early-maturity cotton generally begins flower bud differentiation when two true leaves are completely flattened [34]. In Arabidopsis, at least 4 RPD3 genes have been reported to be associated with flowering time, AtHDA6 has been identified to be associated with the autonomous pathway of four flowering-promoting pathways and regulated flowering time by interacting with FLD (Flowering LOCUS D) [13, 26]; HDA5 regulated flowering time by repressing the expression of FLC (FLOWERING LOCUS C) and MAF1. In

addition, HDA5 and HDA6 might form a 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 down-regulation of AtHDA19 caused 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 the RNAi lines for that performed delayed flowering, suggesting that it is a potential candidate gene related to cotton fiber initiation and flowering time [29]. As the homologous genes of these four AtRPD3 genes related to flowering time in cotton, gene pair GhHDA2/GhHDA13, homologous to AtHDA6, and GhHDA6/GhHDA15, homologous to AtHDA9, performed marked higher expression in early-maturity cotton variety than that in late-maturity variety during all five stages of flower bud differentiation. Beyond that, the other five GhRPD3 genes we selected possessed relative higher expression at two-leaf stage or three-leaf stage between two cotton varieties (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 the previous studies, multiple evidences have revealed that RPD3 members played essential roles in response to various stresses or plant hormone regulation pathway in Arabidopsis, rice, maize, Populus trichocarpa and so on [13, 17, 38]. To further understand the regulation of GhRPD3 genes under different environment 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 was consistent with the published studies on the multifunctional roles RPD3 genes played in plants growth [7, 8]. Based on the expression patterns of GhRPD3 genes under four abiotic stresses, we could see that almost all of 18 genes were significantly induced by all four stresses or a specific treatment (Fig. 6b), fully illustrating the consequence that GhRPD3 genes could respond to abiotic stresses.

MeJA and ABA not only regulates plant growth and development, but also participates in plant defense response to environmental stress such as mechanical injury, disease and osmotic stress [39, 40]. According to the analysis of cis-elements in predicted GhRPD3 promotors, we selected the GhRPD3 genes containing MeJA and ABA responsiveness elements in the predicted promoters to analyze the expression characteristics of those genes under MeJA and ABA treatment by qRT-PCR. The results revealed that exogenous application of MeJA and ABA significantly induced the transcription of most GhRPD3 genes at different points. The expression of GhHDA13, in particular, was markedly up-regulated after treatment with MeJA and ABA, otherwise, it could be significantly induced under cold and PEG treatment, similar to AtHDA6, which was related to ABA and JA signal pathway and stress response [13, 38, 41]. As the homologous gene of AtHDA19, which was involved in JA and ABA signaling response [17, 38], GhHDA4 and GhHDA11 were significantly induced down-regulated expression by exogenous application of MeJA (Fig. 8). GhHDA1, GhHDA10 and GhHDA11 performed early down-regulated and then up-regulated expression trend under ABA treatment (Fig. 9).

On the basis of our expression profiles analysis of RPD3 genes in upland cotton, GhHDA1, GhHDA2, GhHDA6, GhHDA10 and GhHDA13 showed relative higher expression in most of investigated tissues and in early-maturity cotton variety during flower bud differentiation period, also played essential roles in response to MeJA, ABA and abiotic stresses, constant with the previous extensive evidence showing that plant histone deacetylases played vital roles in a round of plant developmental processes and responses to various environmental stresses [7, 17, 26, 27, 41, 42]. Besides, other GhRPD3 genes also performed a specific function 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 of Arabidopsis. The exon-intron structure and conserved motifs analysis of 54 cotton RPD3 genes showed that the significant difference existed among the four subfamilies, while they were highly conserved on the same branch, indicating the cotton RPD3 genes on the same branch might perform the similar functions in cotton growth and development. The chromosomal distributions of cotton RPD3 genes exhibited the conserved gene numbers and chromosomal location between diploid and tetraploid cotton species. Characteristics of gene expression showed that most GhRPD3 genes performed relative high expression in floral organs and possessed the higher expression in CCRI50 compared with GX11 during flower bud differentiation period. In addition, the expression of GhRPD3 genes could be significantly induced by one or more abiotic stresses as well as exogenous application of MeJA and ABA. The results revealed that GhRPD3 genes might be involved in flower bud differentiation and resistance to abiotic stresses of 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.

Materials And Methods

Plant materials and treatments

Two G. hirsutum cultivars (CCRI50, GX11) were grown under standard filed environments (5 rows, each 8 m long and 0.8 m width) in Anyang, Henan province, China. CCRI 50 is an early-maturity cotton variety with initial flowering time in 60 days and GX11 is a late-maturity cotton cultivar with initial flowering time in 70 days. The buds of these two cotton cultivars were taken from one-leaf stage to five-leaf stage.

TM-1 was planted in a climate-controlled green house with suitable growing environment (light/dark cycle: 16 h at 28 °C/8 h at 22 °C). Four-week seedings appearing 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 treatments. After exogenous application of plant hormones and water, we took the leaves of three seedlings in every treatment at 3 h, 6 h, 9 h and 12 h respectively and promptly freeze these samples in liquid nitrogen.

Identification and sequence retrieval of RPD3 family members

The HMM file (PF00850) of the conserved Hist_deacetyl domain was downloaded from Pfam (https://pfam.xfam.org/) [43]. Putative PRD3 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 HMMER3.0 software with E value threshold of 1e-10 [48]. The study also searched against the Arabiposis genome (Araport_11) [49] obtained from TAIR website (https://www.arabidopsis.org/) and published genomes of the other four species: Oryza satica L. (JGI_v7.0), Theobroma cacao (JGI_V2.1),Populus trichocarpa (JGI_v3.1) and Zea mays L. (JGI_v4) downloaded from the 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 genome databases, the protein sequences of RPD3 genes of different species were extracted from the formatted protein databases using blast program (ncbi-blast-2.6.0+-x64-win64.tar). The SMART database (http://smart.embl-heidelberg.de/) was employed to verify each predicted RPD3 protein with a Hist_deacetyl domain and the 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 amino acid length, theoretical molecular weight (Mw), isoelectric point (pI) and subcellular localization of the RPD3 genes were predicted using Pepstats (https://www.ebi.ac.uk/Tools/seqstats/emboss_pepstats/) and the online program CELLO v2.5 (http://cello.life.nctu.edu.tw/) [52].

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. arboretum, G. hirsutum, G. barbadense, A. thaliana, T. cacao, P. trichocarpa, Oryza sativa and Zea mays was carried out using the ClustalW program [53]. The neighbor-joining (NJ) method was employed to construct an unrooted phylogenetic tree using MEGA 7.0 program with 1000 bootstrap repetitions [54].

Chromosomal distribution, gene structure and conserved motif analysis

The positions of GhRPD3, GbRPD3, GaRPD3 and GrRPD3 genes on chromosomes were identified according to the GFF files obtained from the CottonFGD website [47]. The coding sequences of RPD3 genes in four cotton species were aligned with their genomic DNA sequences to analyze the exon-intron structure, which were visualized using the online toolkit GSDS 2.0 (http://gsds.cbi.pku.edu.cn/) [55]. 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 setting E value at 1e-5 [56, 57].

Gene duplication events and selection pressure

This study used BLASTp search (E-value < 1e-10) and MCScanX program in TBtools to perform genome collinearity analysis and detect orthologous and paralogous gene pairs [58, 59]. The circular maps of identified gene pairs were visualized using the circos program [60]. The adjacent RPD3 family members on a single chromosome were considered to be tandem duplicated genes [61]. The ratios of nonsynonymous (Ka) substitutions and synonymous (Ks) substitutions of homologous gene pairs were calculated using TBtools with NG methods to evaluate the selection pressure of these gene pairs [59, 62]. Normally, Ka/Ks < 1 demonstrated purifying selection; Ka/Ks = 1 demonstrated neutral selection; and Ka/Ks > 1 demonstrated positive selection. Then, the divergence times of the homologous gene pairs were estimated using the formula “t = ks/2r”, with r(2.6 × 10 − 9) representing neutral substitution [63].

Analysis of cis-elements in GhRPD3 promoter regions

The promoter regions of all GhRPD3 genes containing 2000 bp DNA sequences of upstream of the initiation codon (ATG) were extracted from the G. hirsutum genome database downloaded from CottonFGD (https://cottonfgd.org/), employed to the website PlantCARE (http://bioinformatics.psb.ugent.be/webtools/plantcare/html/) [64] to determine the cis-elements of the GhRPD3 genes and TBtool was used for visualization [59].

Gene expression patterns analysis

Primary RNA-seq data of G. hirsutum TM-1 were obtained from the NCBI Sequence Read Archive (SRA: PRJNA490626) [23]. TopHat2 [65] and cufflinks program [66] were used to calculate gene expression in fragments per kilobase million (FPKM). The FPKM 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) was obtained. The relative data were normalized by log2 method to investigate the expression patterns of GhRPD3 genes. Characteristics of gene expression among all these twelve tissues and under four abiotic stresses were visualized with Heml 1.0.3.7 software [67].

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

Cotton buds collected at different stages as well as leaves after MeJA and ABA treatment were frozen in liquid nitrogen and then a mortar and pestle were used to ground the samples into fine powder [68]. Depending on the operating instructions, total RNA of these samples was extracted using the Tiangen RNAprep Pure Plant kit (Tiangen, China) and then 1ug of total RNA was reverse transcribed to synthesize first-strand cDNA using the PrimeScript RT Reagent kit (Takara, Japan), which was diluted five times for the further experiments. The cotton histone-3 gene (AF024716) was used as an internal control [69] and specific primers of GhRPD3 genes for qRT-PCR were designed using Oligo 6.0 software, shown in Additional file 10: Table S8. The qRT-PCR were conducted on an ABI 7500 real-time PCR system (AppliedBiosystems, USA) using UltraSYBR Mixture (Low ROX) (Cwbio, China) with three technical repetitions and three biological replicates. The following was the detailed run method: 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, 32 s at 72 °C; step 3:

melting curve analysis. The relative expression of RPD3 genes was calculated using the $2^{-\triangle\triangle CT}$ 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**. *BBA - Gene Regulatory Mechanisms* 2011, **1809**(10):567-576.
2. Reyes JC, Lars H, Wilhelm G: **Chromatin-remodeling and memory factors. New regulators of plant development**. *Plant Physiology* 2002, **130**(3):1090-1101.
3. Grunstein M, . **Histone acetylation in chromatin structure and transcription**. *Nature* 1997, **389**(6649):349-352.
4. Ng HH, Bird A: **Histone deacetylases: silencers for hire**. *Trends in Biochemical Sciences* 2000, **25**(3):121-126.
5. Chen ZJ, Lu T: **Roles of dynamic and reversible histone acetylation in plant development and polyploidy**. *BBA - Gene Structure and Expression* 2007, **1769**(5):295-307.
6. Strahl BD, Allis CD: **The language of covalent histone modifications**. *Nature* 2000, **403**(6765):41.
7. Courtney, Hollender, Zhongchi: **Histone Deacetylase Genes in Arabidopsis Development**. *Journal of Integrative Plant Biology* 2010, **50**(7):875-885.
8. Alinsug MV, Yu CW, Wu K: **Phylogenetic analysis, subcellular localization, and expression patterns of RPD3/HDA1 family histone deacetylases in plants**. *Bmc Plant Biology* 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.
10. X-J Y, Seto E, . **HATs and HDACs: from structure, function and regulation to novel strategies for therapy and prevention**. *Oncogene* 2007, **26**(37):5310.
11. Liu X, Yang S, Zhao M, Luo M, Yu CW, Chen CY, Tai R, Wu K: **Transcriptional Repression by Histone Deacetylases in Plants**. *Molecular Plant* 2014, **7**(5):764-772.
12. Riccardo Aiese C, Gaetana C, Rosa P, Pasquale T, Giorgio P, Ruben G, Maria Federica C, Clara C: **Histone deacetylase AtHDA7 is required for female gametophyte and embryo development in Arabidopsis**. *Plant Physiology* 2013, **163**(1):431-440.
13. Wu KZ, Lin, Zhou C, Yu CW, Chaikam V: **HDA6 is required for jasmonate response, senescence and flowering in Arabidopsis**. *Journal of Experimental Botany* 2008, **59**(2):225-234.
14. Ma XJ, Yang CP, Xia DA: **Characterization and expression analysis of histone deacetylases family RPD3/HDA1 in Populus trichocarpa**. *Biologia Plantarum* 2016, **60**(2):235-243.

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*, **139**(22):4180-4190.
16. Lei W, Jeongsik K, Somers DE: **Transcriptional corepressor TOPLESS complexes with pseudoresponse regulator proteins and histone deacetylases to regulate circadian transcription**. *Proceedings of the National Academy of Sciences of the United States of America* 2013, **110**(2):761-766.
17. Changhe Z, Lin Z, Jun D, Brian M, Keqiang W: **HISTONE DEACETYLASE19 is involved in jasmonic acid and ethylene signaling of pathogen response in Arabidopsis**. *Plant Cell* 2005, **17**(4):1196-1204.
18. Varotto, S.: **Expression Profile and Cellular Localization of Maize Rpd3-Type Histone Deacetylases during Plant Development**. *Plant Physiology*, **133**(2):606-617.
19. Zhao J, Li M, Gu D, Liu X, Zhang J, Wu K, Zhang X, Silva JATD, Duan J: **Involvement of rice histone deacetylase HDA705 in seed germination and in response to ABA and abiotic stresses**. *Biochemical & Biophysical Research Communications* 2016, **470**(2):439-444.
20. Wenqun F, Keqiang W, Jun D: **Sequence and expression analysis of histone deacetylases in rice**. *Biochemical & Biophysical Research Communications* 2007, **356**(4):843-850.
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 inArabidopsis thaliana**. **4**(1):10-10.
22. Tianzhen Z, Yan H, Wenkai J, Lei F, Xueying G, Jiedan C, Jinbo Z, Saski CA, Scheffler BE, Stelly DM: **Sequencing of allotetraploid cotton (Gossypium hirsutum L. acc. TM-1) provides a resource for fiber improvement**. *Nature Biotechnology* 2015, **33**(5):531-537.
23. Hu Y, Chen J, Fang L, Zhang Z, Ma W, Niu Y, Ju L, Deng J, Zhao T, Lian J *et al*: **Gossypium barbadense and Gossypium hirsutum genomes provide insights into the origin and evolution of allotetraploid cotton**. *Nature Genetics* 2019, **51**(4):739-748.
24. Poczatek MH, Hugo C, Darley-Usmar V, Murphy-Ullrich JE: **The roles of histone acetylation in seed performance and plant development**. *Plant Physiology & Biochemistry* 2014, **84**(20):125-133.
25. Tian L, ., Chen ZJ: **Blocking histone deacetylation in Arabidopsis induces pleiotropic effects on plant gene regulation and development**. *Proceedings of the National Academy of Sciences* 2001, **98**(1):200-205.
26. Chun-Wei Y, Xuncheng L, Ming L, Chiayang C, Xiaodong L, Gang T, Qing L, Yuhai C, Keqiang W: **HISTONE DEACETYLASE6 interacts with FLOWERING LOCUS D and regulates flowering in Arabidopsis**. *Plant Physiology* 2011, **156**(1):173-184.
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 Journal* 2015, **82**(6):925-936.
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**. *Biochemical & Biophysical Research Communications* 2009, **388**(2):266-271.

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 Journal for Cell & Molecular Biology* 2018.
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 (GossypiumÂ hirsutum TM-1) provides insights into genome evolution**. *Nature Biotechnology* 2015, **33**(5):524-530.
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**. *Frontiers in genetics* 2019, **10**:824.
32. Lynch M, ., Conery JS: **The evolutionary fate and consequences of duplicate genes**. *Science* 2000, **290**(5494):1151-1155.
33. Ray LL, Richmond TR: **Morphological measures of earliness of crop maturity in cotton.** *Crop Sci* 1966, **6**(6):527-531.
34. Qiu G, Tang Y: **STUDIES ON THE DIFFERENTIATION CHARACTERS OF FRUITING BRANCHES OF COTTON (G.HIRSUTUM)**. *Journal of Nanjing Agricultural University* 1986.
35. Wanhui K, David L, Caroline S, Dao-Xiu Z: **Arabidopsis histone deacetylase HDA9 regulates flowering time through repression of AGL19**. *Biochem Biophys Res Commun* 2013, **432**(2):394-398.
36. Lu T, Jianlin W, M Paulus F, Meng C, Hongbin C, Gelvin SB, Z Jeffrey C: **Genetic control of developmental changes induced by disruption of Arabidopsis histone deacetylase 1 (AtHD1) expression**. *Genetics* 2003, **165**(1):399.
37. Biłas R, Szafran K, Hnatuszko-Konka K, Kononowicz AK: **Cis -regulatory elements used to control gene expression in plants**. *Plant Cell Tissue & Organ Culture* 2016, **127**(2):269-287.
38. Chen LT, Wu K: **Role of histone deacetylases HDA6 and HDA19 in ABA and abiotic stress response**. *Plant Signaling & Behavior* 2010, **5**(10):1318-1320.
39. Turner EJG: **The Arabidopsis Mutant cev1 Has Constitutively Active Jasmonate and Ethylene Signal Pathways and Enhanced Resistance to Pathogens**. *Plant Cell*, **13**(5):1025-1033.
40. Anderson JP, Badruzsaufari E, Schenk PM, Manners JM, Desmond OJ: **Antagonistic Interaction between Abscisic Acid and Jasmonate-Ethylene Signaling Pathways Modulates Defense Gene Expression and Disease Resistance in Arabidopsis**. *Plant Cell*, **16**(December):3460-3479.
41. Chen LT, Luo M, Wang YY, Wu K: **Involvement of Arabidopsis histone deacetylase HDA6 in ABA and salt stress response**. *Journal of Experimental Botany* 2010, **61**(12):3345-3353.
42. Ma, Xujun, Shibo, Zhang, Chao, Yang, Chuanping: **Histone deacetylases and their functions in plants**. *Plant Cell Reports* 2013, **32**(4):465-478.
43. Bateman A, Coin L, Durbin R, Finn RD, Hollich V, Griffiths-Jones S, Khanna A, Marshall M, Moxon S, Sonnhammer ELL: **The Pfam protein families database**. *Nucleic Acids Research* 2008, **32**(1):D138.
44. Du X, Huang G, He S, Yang Z, Sun G, Ma X, Li N, Zhang X, Sun J, Liu M: **Resequencing of 243 diploid cotton accessions based on an updated A genome identifies the genetic basis of key agronomic**

traits. *Nature Genetics* 2018, **50**(6).

45. Paterson AH, Wendel JF, Heidrun G, Hui G, Jerry J, Dianchuan J, Danny L, Showmaker KC, Shengqiang S, Joshua U: **Repeated polyploidization of Gossypium genomes and the evolution of spinnable cotton fibres**. *Nature* 2012, **492**(7429):423-427.
46. Wang M, Tu L, Yuan D, Shen C, Li J, Liu F, Pei L, Wang P, Zhao G, Ye Z: **Reference genome sequences of two cultivated allotetraploid cottons, Gossypium hirsutum and Gossypium barbadense**. *Nature genetics* 2019.
47. Tao Z, Liang C, Meng Z, Sun G, Meng Z, Guo S, Rui Z: **CottonFGD: an integrated functional genomics database for cotton**. *Bmc Plant Biology* 2017, **17**(1):101.
48. Finn RD, Jody C, Eddy SR: **HMMER web server: interactive sequence similarity searching**. *Nucleic Acids Research* 2011, **39**(Web Server issue):29-37.
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 Journal* 2017, **89**(4):789-804.
50. Goodstein DM, Shengqiang S, Russell H, Rochak N, Hayes RD, Joni F, Therese M, William D, Uffe H, Nicholas P: **Phytozome: a comparative platform for green plant genomics**. *Nucleic Acids Research* 2012, **40**(Database issue):D1178-D1186.
51. Ivica L, Tobias D, Peer B: **SMART: recent updates, new developments and status in 2015**. *Nucleic Acids Research* 2015, **43**(D1):D257-D260.
52. Wang J, 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):1-10.
53. Wünschiers R: **Installing BLAST and ClustalW**; 2004.
54. Kumar S, Stecher G, Tamura K: **MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets**. *Molecular Biology & Evolution* 2016, **33**(7):1870.
55. Hu B, Jin J, Guo AY, Zhang H, Luo J, Gao G: **GSDS 2.0: an upgraded gene feature visualization server**. *Bioinformatics* 2014, **31**(8):1296.
56. Bailey TL, Nadya W, Chris M, Li WW: **MEME: discovering and analyzing DNA and protein sequence motifs**. *Nucleic Acids Research* 2006, **34**(Web Server issue):369-373.
57. Bailey TL, Johnson J, Grant CE, Noble WS: **The MEME Suite**. *Nucleic Acids Research* 2015, **43**(Web Server issue):W39-W49.
58. Wang Y, Tang H, Debarry JD, Tan X, Li J, Wang X, Lee T, Jin H, Marler B, Guo H: **MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity**. *Nucleic Acids Research* 2012, **40**(7):e49-e49.
59. 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.
60. Krzywinski M, Schein JI: **Circos: an information aesthetic for comparative genomics**. *Genome Research* 2009, **19**(9):1639-1645.

61. Zhu Y, Wu N, Song W, Yin G, Qin Y, Yan Y, Hu Y: **Soybean (Glycine max) expansin gene superfamily origins: segmental and tandem duplication events followed by divergent selection among subfamilies**. *Bmc Plant Biology*, **14**(1):93.
62. Hurst LD: **The K a/ K s ratio: diagnosing the form of sequence evolution**. *Trends in Genetics* 2002, **18**(9):486-487.
63. Huiru S, Pengbo H, Qiang M, Meng Z, Yuan Q, Hengling W, Junji S, Hantao W, Lijiao G, Nuohan W: **Genome-wide identification and expression analyses of the pectate lyase (PEL) gene family in cotton (Gossypium hirsutum L.)**. *BMC Genomics*.
64. Magali L, Patrice D, Gert T, Kathleen M, Yves M, Yves VDP, Pierre R, Stephane R: **PlantCARE, a database of plant cis-acting regulatory elements and a portal to tools for in silico analysis of promoter sequences**. *Nucleic Acids Research* 2002, **30**(1):325-327.
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 Biology* 2013, **14**(4):R36.
66. Cole T, Williams BA, Geo P, Ali M, Gordon K, Baren MJ, Van, Salzberg SL, Wold BJ, Lior P: **Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation**. *Nature Biotechnology* 2010, **28**(5):511-515.
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 Science Bulletin* 2007, **52**(22).
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):402-408.

Figures

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 satica and Zea mays were aligned using ClustalW program and the neighbor-joining (NJ) method was employed to construct this unrooted phylogenetic tree using MEGA 7.0 program with 1000 bootstrap repetitions. Four subfamilies were represented by different colored lines.

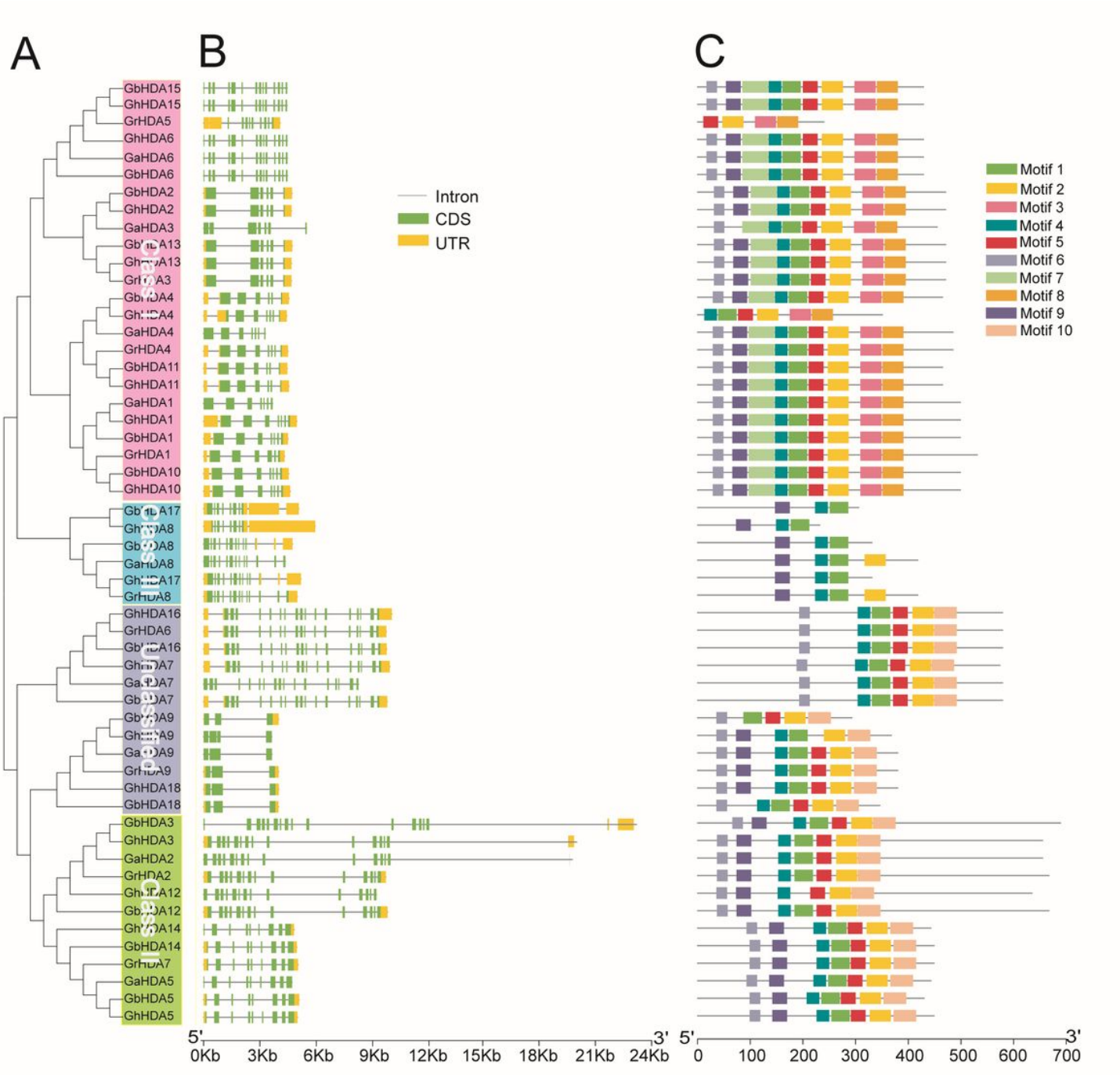

Figure 2

Phylogenetic relationships, exon-intron structure and conversed 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. Yellow boxes represented UTR, green boxes represented exon and black lines represented introns; (c) The 10 conversed protein motifs of RPD3 genes were indicated by different colored boxes.

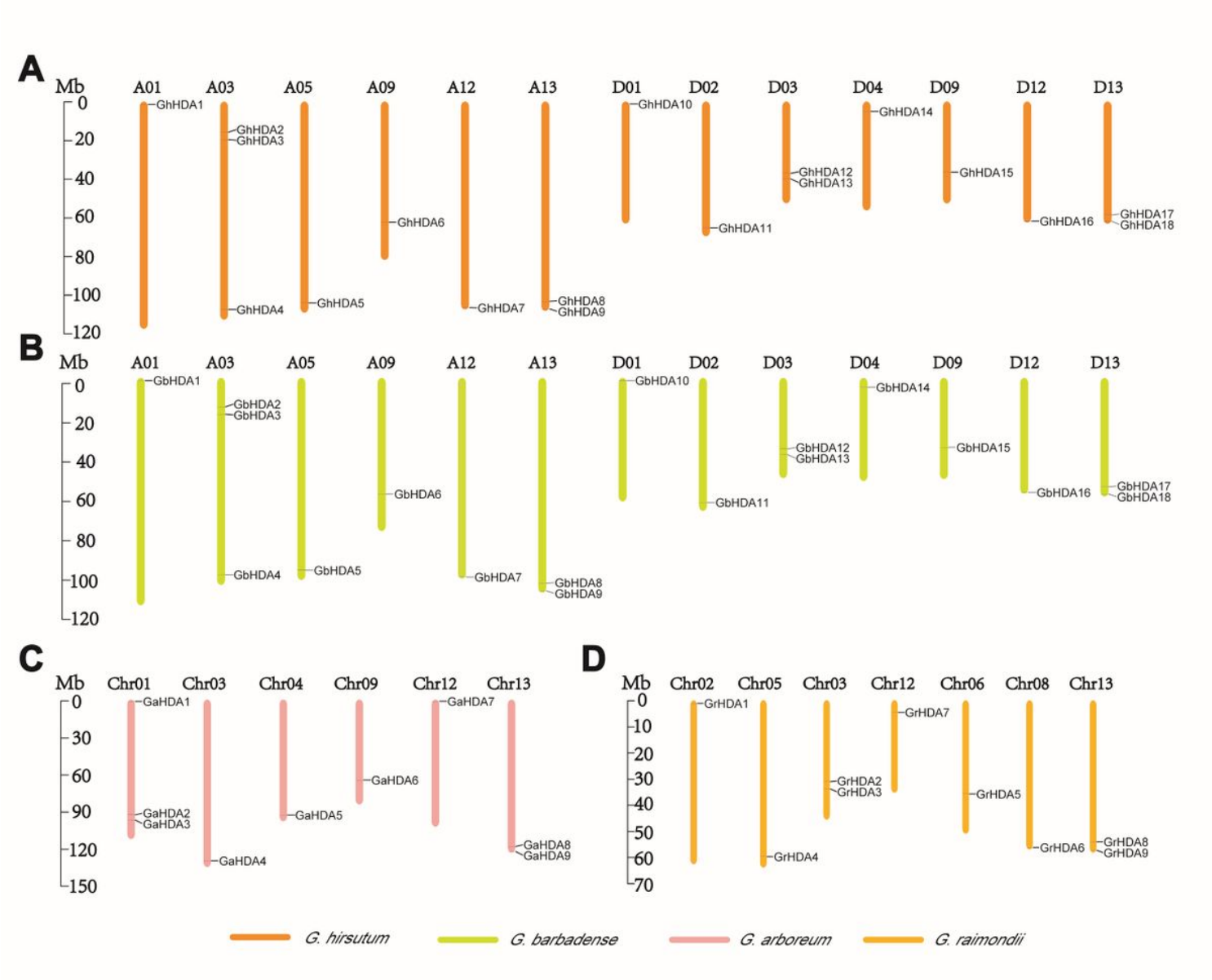

Figure 3

Chromosomal distribution of cotton RPD3 genes. a, b, c and d represented the chromosomal location of RPD3 genes from G. hirsutum (a), G. barbadense (b), G. arboreum (c) and G. raimondii (d), respectively. The chromosome number was shown on the top of each chromosome. The scale bar represented the length in megabases (Mb).

Figure 4

RPD3 homologous gene pairs among G. arboreum, G. raimondii and G. hirsutum. Orange, blue and red represented 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 were presented in the form of bar graphs and similar cis-elements were exhibited with same colors.

Figure 6

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

Figure 7

Expression levels of 9 GhRPD3 genes between CCRI50 and GX11. Blue and orange bar graphs indicated the expression of early-maturity cotton (CCRI 50) and late-maturity cotton (GX11), respectively. The error bars showed the standard deviation of three biological replicates.

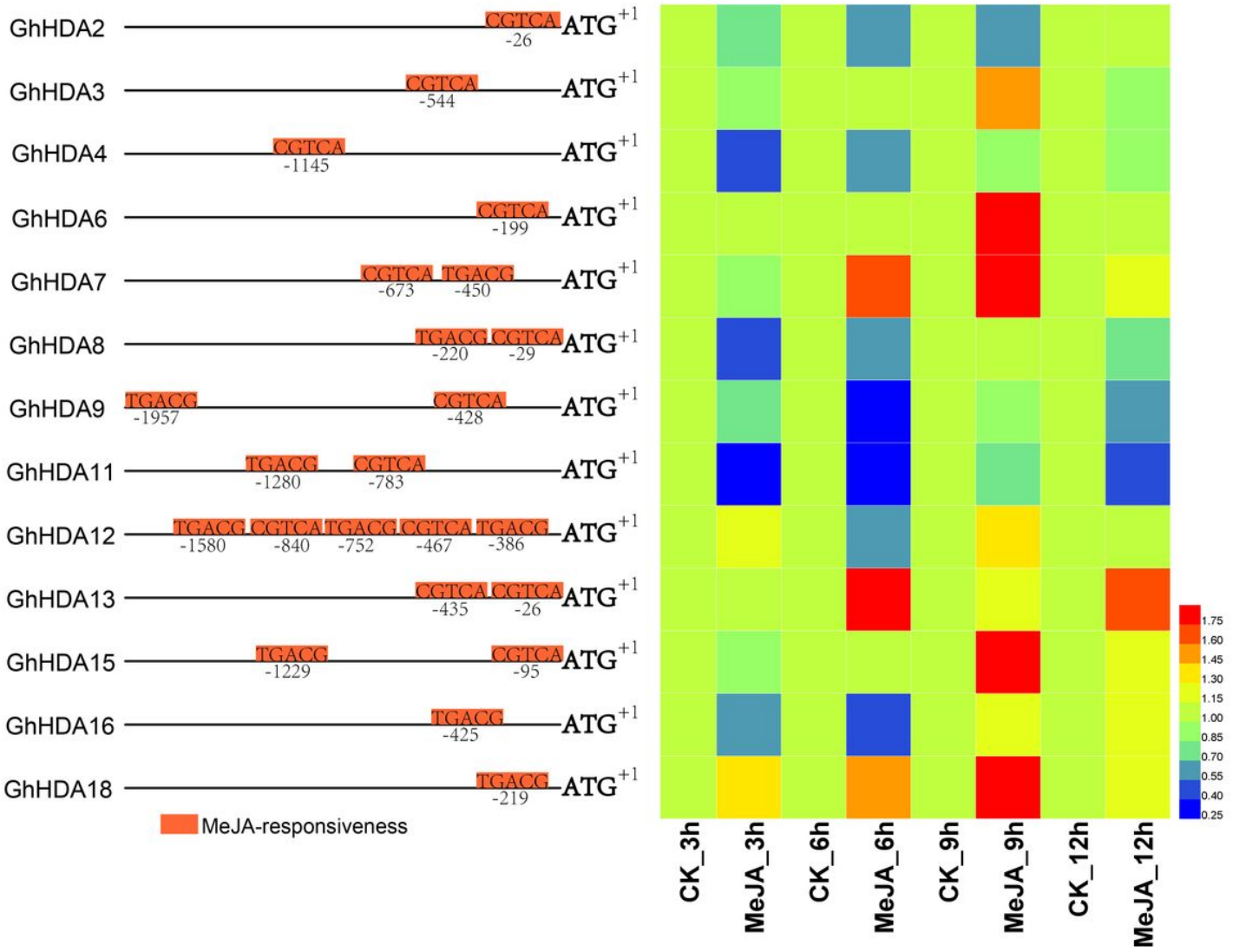

Figure 8

Expression profiles of 13 GhRPD3 genes under MeJA treatment. Orange boxes represented the MeJA-responsiveness elements of 13 GhRPD3 genes in the promoter regions (left). The expression changes of 13 GhRPD3 genes under MeJA treatment were shown using a heatmap (right). The 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 represented the ABRE of 11 GhRPD3 genes in the promoter regions (left). The expression changes of 11 GhRPD3 genes under ABA treatment were shown using a heatmap (right). The 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.

Figure 11

Supplementary Files

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

- TableS5.xlsx
- TableS8.xlsx
- TableS6.xlsx
- FigureS1.tif
- TableS7.xlsx
- FigureS2.tif
- TableS1.xlsx
- TableS3.xlsx
- TableS4.xlsx
- TableS2.xlsx