Genome-wide identification and expression analysis of PUB genes in cotton

Xuke Lu

Chinese Academy of Agricultural Science

Na Shu

Hanzhong Agricultural Science Institute

Delong Wang

Chinese Academy of Agricultural Sciences

Junjuan Wang

Chinese Academy of Agricltural Sciences

Xiugui Chen

Chinese Academy of Agricultural Sciences

Binglei Zhang

Chinese Academy of Agricultural Sciences

Shuai Wang

Chinese Academy of Agricultural Sciences

Lixue Guo

Chinese Academy of Agricultural Sciences

Wuwei Ye

Institute of Cotton Research, Chinese Academy of Agricultural Sciences

---

Research article

Keywords: U-box domain; PUB gene; cotton; evolution; Collinearity; homologous gene pairs

---

Abstract

Background: U-box gene is gene of ubiquitin ligase, which contain U-box domain. Plant U-box gene (PUB) plays an important role in the response to stresses, but less reports about PUBs in cotton were issued. Therefore research of PUBs in cotton will be of great importance and necessity to study the mechanism of tolerance-resistance of the cotton. Results: In this study, we identified 93, 96, 185 and 208 PUBs from the three sequenced cotton species G. raimondii (D5), G. arboreum (A2), G. hirsutum (AD1) and G. arbadense (AD2), respectively. Subcellular localization analysis showed that the PUBs in cotton were distributed in various parts of the cells, mainly in the nucleus. The PUBs in cotton were divided into six subfamilies (A-F) by phylogenetic analysis, and intron/exon structure was comparatively conserved within the subfamily. Location analysis showed cotton PUBs were unevenly anchored on all the chromosomes varying from 1 to 14. It was found that there are 3 tandem duplications and 28 segmental duplications in cotton genome D5, 2 tandem duplications and 25 segmental duplications in A2, and 143 homologous gene pairs between A2 and D5, through multiple sequence alignment, but that the tandem duplication region of A2 or D5 was not found. It was also found that, there were 105, 14 and 17 homologous gene pairs in intra-subgenome of At and Dt, At subgenome, Dt subgenome of allotetraploid cotton, respectively. Among the PUBs family, totally 106, 116, 85 and 81 homologous gene pairs were found in A2-At, A2-Dt, D5-At and D5-Dt. Function analysis of GhPUB85A and GhPUB45D showed they positively responded the abiotic stresses, but the expression patterns were different. Besides, although the expressions of these two homologous genes were similar, their contributions were different when responding to stresses, showing different response differences to abiotic stresses and function division of two subgenomes of G. hirsutum. Conclusion: This study provided the genome-wide identification, structure, evolution and expression analysis of PUBs in cotton, and the results showed that the PUBs was highly conserved in evolutionary history of cotton. All PUB genes were involved in response to abiotic stresses at varying degrees.

Background

The ubiquitin-mediated ubiquitination pathway is the post-translational modification pathway of eukaryotic proteins. Studies demonstrated that the pathway is involved in cell cycles in higher plants [1], stress resistance [2], signal transduction [3], apoptosis [4], optical signal [5] and other physiological pathways. In ubiquitin pathway, it takes three steps for the ubiquitin to act on the target protein. First, the ubiquitin-activating enzyme (E1) activates ubiquitin [1], and then the activated Ub molecules and delivered to the ubiquitin-binding enzyme (E2) [2]; finally, Ub molecules are transferred to the target protein through ubiquitin ligase (E3). E3 is responsible for identification of the specificity of substrate protein, and has the most species [3]. In Arabidopsis, there are more than 1,400 genes encoding components of the ubiquitination pathway, of which about 90% the encode ubiquitin ligase E3 [3, 4]. According to its subunit composition and mechanism of action, the ubiquitin ligase E3 can be divided into single subunit type (such as HECT, RING/U-box) [5] and multi-subunit type (such as SCF (skp1-cullin-F-box), APC (anaphase-promoting complex) [6], VBC (VHL-elongin B-Elongin C), *etc.*) [7, 8].

PUBs were reported in many model crops including Arabidopsis, rice, chlamydomonas reinhardtii, Chinese cabbage, and soybean. Previous studies showed 64 *PUB*s were identified in *Arabidopsis* [9], 77 in rice [10], 33 in chlamydomonas reinhardtii [11], 101 in Chinese cabbage [12] and 125 in the soybean [13], indicating that PUB genes exist widely in plants and plays an important role in plants. Many studies have shown that PUB proteins were involved in abiotic stress responses. Cho et al. [14] obtained U-box E3 protein (CaPUB1) from the drought-treated pepper. In Arabidopsis, the heterologous overexpression of *CaPUB1* attenuated the tolerance of plants to drought stress, two homologous proteins AtPUB22 and AtPUB23 in Arabidopsis were also involved in the drought stress response of plants [15]. Furthermore, studies also showed that AtPUB22 and AtPUB23 could modulate negatively the drought stress response of plants by synergistic ubiquitination of RPN12a [15]. Liu *et al*. identified a U-box E3 protein AtPUB19 induced by drought, high salt, cold and ABA. *atpub 19* mutant plants promoted stomatal closure, thereby enhanced tolerance to drought of the plants; *AtPUB19*-overexpressing plants are not sensitive to ABA but sensitive to drought. Meanwhile, there are significant changes in expression levels of ABA and stress-related genes in *atpub 19* mutant and A*tPUB19*-overexpressing plants, indicating AtPUB19 may regulate the drought stress by negative regulation of ABA signal [16]. PUB proteins, involved in the process of drought response, were also identified in rice. Previous studies showed drought resistance *of OsPUB15*-overexpressing plantssignificantly enhanced, and *OsPUB15* could be induced by hydrogen peroxide, drought and high salt, indicating that *OsPUB15* may positively regulate the response of plants to drought stress by attenuating intracellular oxidative stress [17].

Cotton is the most important fiber crop and the model crop for the research of polyploidy, evolution, cell wall development, and cellulose synthesis [18]. There were approximately 50 species distributed in arid and semi- arid regions of the tropic and subtropics, which were presumed to have originated from the same ancestor 50 to 100 million years ago [19]. The current cultivars are diploid *G. arboreum* and *G. herbaceum*, tetraploid *G. hirsutum* and *G. barbadense*. The tetraploid cottons originate from hybridization between the African ancestral species of A genome and the American ancestor species of D genome one to two million years ago [18]. Recently the completed sequencing work of diploid cottons (*G. raimondii* (D5) [20, 21], *G.* arboreum (A2) [22] and allotetraploid cottons (*G. hirsutum*TM-1 (AD1) [23, 24], *G. barbadense* acc.3-79 (AD2) [25] provided the reference for the study of gene function and evolution from the whole genome level. Based on the cotton genome sequences, the family members of PUB gene were identified and the structure and distribution characteristics of PUB gene were analyzed, as well as its evolution analysis in whole genome replication and cotton allotetraploid formation.

Materials And Methods

Cotton and *Arabidopsis thaliana*

Upland cotton variety ZhongS9612 was selected and preserved by the Cotton Adversity Research Laboratory at the Chinese Academy of Agricultural Sciences (CAAS) for many years. Before planting into the sand, the seeds were surface-sterilized with 0.1% $HgCl_2$ and placed in a sterile dish with moist filter paper to accelerate germination. Uniform seedlings were chosen and transplanted into sand pots (10

plants in each pot) in the greenhouse (14 h/day, 30°C and 10 h/night, 24°C) at the Institute of Cotton Research of CAAS. For the agrobacterium-mediated transformation, we referred the method used by Lu et al [26].

Whole genome identification of *PUBs* in cotton

Cotton genome data (G. raimondii (D5) [20, 21] , G. arboretum (A2) [22], G. hirsutum acc. TM-1 (AD1) [23, 24] were obtained from CottonGene (https://www.cottongen.org/). Hidden Markov Model (HMM) profile of U-box domain (PF04564) was obtained from Pfam30.0 (http://pfam.xfam.org/) [27], and was used as a query to identify the candidate PUBs from the cotton genome protein database using HMMER3.0 [28]. We used BLAST2.2.31+ (ftp://ftp.ncbi.nlm.nih.gov/blast/executables/blast+ /LATEST/) to get coding domain sequence (CDS) sequences, protein sequences and the corresponding entire sequence on the genome. The protein sequences were further analyzed in SMART (http://smart.embl-heidelberg.de/) and Pfam 30.0 [27] database to ensure each candidate protein contains U-box domain. Subcellular localization prediction was carried out in CELLO v.2.5 [29].

Exon-intron structure, phylogenetic and conservative domain analysis

The identified CDS sequence and the genome sequence of cotton *PUBs* were used to analyze *PUBs* structure by GSDS2.0 [30]. The full-length sequences of PUB proteins were used to construct phylogenetic tree. Multi-sequence alignment of PUBs was carried out by ClustalX1.83, and Neighbor-Joining (NJ) method [31] was used construct phylogenetic tree in MEGA6.0 [32]. The online software SMART, PROSITE [33] were used to analyze the conserved domains of protein.

Physical location of *PUBs* in cotton

GFF (general feature format) files of cotton *PUBs* were obtained from the genome annotation files, and the information on *PUBs* of chromosome position was prosected from it. The distribution of cotton *PUBs* on the chromosome was drawn with MapInspect (http://mapinspect.software. informer. com/).

Gene duplication and microsynteny analysis among *G.* arboreum, *G. raimondii* and *G. hirsutum*

Homologous gene pairs were identified according to multiple sequence alignment results and the standard described in previous studies [34, 35]. The collinearity of homologous genes was drawn in Circos [36] based on the homology between them and their positions on the genome.

Gene cloning and the construction of vectors

The first strand was synthesized according to the instructions of TransScript one-step gDNA removal and cDNA synthesis supermix kit. Two homologous genes GhPUB85A and GhPUB45D were cloned using In-fusion technology. Primers were designed on the website and sequences were below: InGhPUB85A-V-Foward：TAAGGTTACCGAATTCATGGTTGAAAGTGGAAG; InGhPUB85A-V-Revers：TCGGGACATGCCCGGGTTTCACAAGCCTCCTTC; InGhPUB45D-V-Foward：

TAAGGTTACCGAATTCTTGGTTGAAAGTGGAAG; lnGhPUB45D-V-Reverse ⬜ TCGGGACATGCCCGGGCTTCACAAGCCTCCTTC. PCR amplification products were verified using 1.5% agarose gel electrophoresis. PCR amplification products were purified using MiniBEST Agarose Gel DNA Extraction Kit from Takara Corporation. Finally the concentration of targets were measured, and stored at -20°C for use.

Targets purified were linked to pEASY-Blunt Cloning Vector, and then transformed into Escherichia coli according to the instructions of pEASY-Blunt Cloning Kit. Positive clones were selected and inoculated into LB liquid medium containing Kana (50 mg ·L^{-1}) for about 6h under the conditions of 200 rpm and 37 °C. Then the positive clones were verified using PCR amplification with M13 primers. Finally clone vectors GhPUB85A-t and GhPUB45D-t were obtained. PCR reaction system used in the research was below: 5 × PrimerSTAR GXL Buffer, 10.0 μl; dNTP Mixture⬜2.5 mM each⬜, 4.0 μl; F-primer (10μM), 1.5 μl; R-primer (10μM), 1.5 μl; cDNA, 100 ng; 5×PrimerSTAR GXL DNA Polymerase, 100 ng, 5×PrimerSTAR GXL DNA Polymerase, 2.0 μl; add ddH$_2$O to 50.0 μl. PCR procedure of gene amplification used in the research was: 98 °C, 10s; 55 °C, 15s; 68 °C, 9s; 4°C, forever, 35cycles. Real-time PCR method was used to measure the relative expression of two genes. Primers of two genes used were below: qGhPUB85A-Foward: ATGGTTGAAAGTGGAAG; qGhPUB85A-Reverse: TTTCACAAGCCTCCTTC; qGhPUB45D-Foward: TTGGTTGAAAGTGGAAG; qGhPUB45D-Reverse: CTTCACAAGCCTCCTTC. GhHistone3 (AF02471) was used as the reference gene.

VIGS analysis of GhPUB85A and GhPUB45D

Based on vector sequence, target gene sequence and enzyme cutting sites, In-fusion primers were designed on the website (http://bioinfo.clontech.com/infusion). Primer sequences of lnGhPUB85A-V and lnGhPUB45D-V were shown before. Primer sequences of GhPUBs-RFP and lnGhPUBs were below: GhPUBs-RFP-Forward: CACGGGGGACTCTAGAATGGTGTCAGGTTCGGCATTA; GhPUBs-RFP-Reverse: AGGAGGCCATCCCGGGGTCCAATTTGTAAGCCCAACC; lnGhPUBs-Forward: CTCTAGAGGATCCCCGGGATGGTGTCAGGTTCGGCATT; lnGhPUBs-Reverse: GATCGGGGAAATTCGAGCTCTTAGTCCAATTTGTAAGCCCA. Finally silencing vectors pYL156:GhPUB85A and pYL156:GhPUB45D, and RFP vectors pBI121-GhPUB85A:RFP and pBI121-GhPUB45D:RFP, and Plant overexpression vectors pBI121:GhPUB85A and pBI121:GhPUB45D were totally constructed. All vectors were transformed into cotton and Arabidopsis thaliana with agrobacterium mediated genetic transformation method.

Results

Identification of PUB gene family members in whole genome of cotton

The hidden Markov model (HMM) of U-box domain (PF04564) was downloaded from the database Pfam30.0, and used as a query to identify the candidate PUB members in four cotton genomic database

using HMMER3.0. SMART. Besides, Pfam30.0 was also used for further identification to confirm every PUB members containing U-box domain. Finally, 93, 96, 185, 208 *PUB*s were identified from the four sequenced cotton species, *G. raimondii* (D$_5$), *G. arboreum* (A$_2$), *G. hirsutum* acc. TM-1 (AD$_1$), and *G. barbadense* (AD2), respectively, which were named *GrPUB1-93*, *GaPUB1-96*, *GhPUB1A-89A/ 1D-91D/ 181-185* and *GbPUB1A-98A/ 1D-98D/ 197-208* according to their location on the chromosome. The gene number of tetraploid cottons was the twice as many as that in diploid cottons, showing *PUB* genes are relatively conservative. The information about gene name, locations, length of open reading frame (ORF), type of protein domain, position of U-box domain and subcellular localizations of these gene family members could be found (**Additional file 1: Table S1, Additional file 2: Table S2, Additional file 3: Table S3 and Additional file 4: Table S4**). The length of the PUB protein sequence in cotton ranges from 49 to 1492 AA, and the U-box domain contains about 75 amino acids. However, the length of U-box domain was almost the same except a few PUBs, for example, GaPUB39 and GhPUB40D has only 32 and 50 amino acids, respectively. Subcellular localization analysis showed PUB proteins distributed throughout the cell, including nuclear, cytoplasmic, chloroplast, Plasma Membrane, mitochondrial, and extracellular. But most PUB proteins are located in the nucleus. 20 different domains were found in all cotton PUBs (**Table 1**), and the primary mode was "U-box+ARM/HEAT". Different domain modes may be associated with different functions of cotton PUBs.

Analysis of *PUBs* gene structure and evolution in cotton

Gene structure diagram of PUBs and evolution tree were constructed (**Additional file 5: Figure S1, Additional file 6: Figure S2, Additional file 7: Figure S3 and Additional file 8: Figure S4**). Based on the evolutionary relationship, the PUB genes could be divided into five subgroups (I-V). Among these subgroups, subgroup I was composed of domains "U-box + ARM" and "U-box only", and the rest subgroups were composed of other domains. The exon number of *PUB* genes in cotton is greatly divergent, ranging from 1 to 25. Among all the PUBs, approximately 1/3 PUBs contain only one exon. Generally, the evolutionary relationship is somehow correlated with gene structure, that is, the more similar the number and size of the exon, the closer evolutionary relationship. In *G. hirsutum*, the length of *GhPUB1A* is 47Kb, much larger than other PUB genes, which may be correlated with the assembly and annotation of the cotton genome. Members in each subgroup in *G. barbadense* (AD2) was greatly different with that in *G. raimondii* (D$_5$), *G. arboreum* (A$_2$) and *G. hirsutum* (AD$_1$), which may be correlated with the different origins of these species. So the evolution of PUBs in *G. raimondii* (D$_5$), *G. arboreum* (A$_2$) and *G. hirsutum* (AD$_1$) was analyzed, and the results also indicated Ⅰ-Ⅴsubgroups were found (**Additional file 9: Figure S5**), and this was similar with the evolution of single PUB gene, suggesting the PUB members were highly conservative. Furthermore, more close evolution relationships of GhPUB1A-89A with GaPUB1-96 and GhPUB1D-91D with GaPUB1-93 were found based on the evolution analysis.

Location of *PUB* Genes in three cotton genomes

The MapInspect software was used to draw the distribution map of PUB genes on the chromosomes based on the position information. Among 93 genes in *G. raimondii*, 91 were located unevenly on the

chromosome and others were found on Scaffolds (**Figure 1A**). The results indicated that only a few genes were found on chromosome 3, 4, and 12, and the PUB genes on chromosome 5 were the most (11 PUBs). Besides, PUB genes on chromosome 4, 6, 7, 11 and 12 were biasedly enriched towards one end of the chromosome. All of 96 PUB genes identified in *G. arboreum* were located on different chromosomes (**Figure 1B**). The results showed uneven distribution of PUBs on each chromosome in *G. arboreum*, and chromosome 1 enriched the most PUB genes (up to 14) and chromosome 3enriched the least PUB genes (only 2). Besides, the length of chromosome 5 was about 6 Mb, but there were 9 PUB genes were found, showing the highest distribution density. In *G. hirsutum*, 91.4% (169/185) PUB genes were anchored on chromosomes, among of which 82 and 87 genes were found A- and D- subgenome, respectively (**Figure 2**). The number of PUB genes on chromosome 7 was the most and chromosome 8 was the least compared with other chromosomes in both At- and Dt- subgenome in *G. hirsutum*, showing PUBs on these two chromosomes were relatively conservative and significant for cotton growth. The situation in *G. barbadense* was different with *G. hirsutum* (**Additional file 10: Figure S6**). The results showed PUB gene were equally distributed in At- and Dt- subgenome and unevenly located on each chromosome, which may be correlated with differentiation of two species.

Gene duplication analysis

Fragment duplication of the chromosomal region may result in the scattering of the gene family members on multiple chromosomes. Compared with other eukaryotes, plants always have a higher rate of gene replication. Recent studies have shown that *G. raimondii* have had at least two complete genome-wide replicates [30]. Segregation of cotton A genome and D genome diploid occurred in about 5-10Myr years ago [18], and allotetraploid *G. hirsutum* was generated from hybridization of diploid cottons and the number of chromosomes were doubled 1-2Myr years ago. In the study, BLAST2.2.31+ (ftp://ftp.ncbi.nlm.nih.gov/blast/executables /blast+/LATEST/.) was used for blastn and blastp (value 10) screening of homologous gene pairs from the cotton *PUB* genes identified. The uneven distribution of genes on the chromosome may be the result of gene duplication or partial fragment replication during the long evolutionary history of the cotton genome. Each time the replication occurs, the entire genetic sequence of cotton is doubled, and over time, these redundant genes are recombined or lost [23]. Previous studies have shown that gene duplication and post-segregation phenomena are two major driving forces of evolution [32]. Based on the multiple sequence alignment of the encoding sequences and proteins in diploid cotton, 18 and 27 homologous gene-pairs were discovered with MCScanX [37] in *G. raimondii* (D_5) (**Additional file 11: Figure S7A**) and *G. arboreum* (A_2) (**Additional file 11: Figure S7B**), respectively. Among these homologous gene-pairs, 15 segmental duplications and 3 tandem duplications were found in *G. raimondii*, and 25 segmental duplications and 2 tandem duplications were found in *G. arboreum*. The relationship between two diploid cottons and *G. hirsutum* was analyzed (**Additional file 12: Figure S8**). Totally 197 homologous gene-pairs were found between *G. raimondii* and *G. hirsutum*, among of which 58.89% (116/197) were located on Dt-subgenome, and 191 homologous gene-pairs were found between *G. arboreum* and *G. hirsutum*, among of which 55.50%(106/191) were located on At-subgenome. All these results showed more than half of homologous genes in *G. hirsutum* were derived from

corresponding diploid cotton genome. Furthermore, approximately 41.11% - 44.50% homologous genes were originated crosswise from other diploid genome.

Expression pattern analysis of PUB genes in cotton

Based on previous transcriptome data of PUBs under different stresses (salt, drought, low temperature and high temperature) in *G. hirsutum*, 117, 148 and 119 PUB genes were found with FPKM >1 in root, stem and leaf, respectively, displaying tissue specificity. Among all the PUB genes, approximately 21 PUB genes with no expression were identified in three tissues, which may be associated with other specific regulations. Interestingly, it was found that 3 PUB genes (including GhPUB58D, GhPUB55A and GhPUB67D) always highly expressed in three tissues under salt, drought, low temperature and high temperature stresses. All PUB genes were divided into five subgroups (Ⅰ, Ⅱ, Ⅲ, Ⅳ and Ⅴ), similar expression patterns were found among all PUB genes (**Additional file 13: Figure S9 and Additional file 14: Figure S10**). In subgroupⅡ, 18 PUB genes with huge expression difference were discovered, otherwise, other PUB genes in subgroup Ⅲ-Ⅴ were found to have a consistent expression pattern under different stresses. But 4 PUB genes (*GhPUB32A - GhPUB38D)* in subgroup Ⅰwere found with little expression difference under different stresses.

Cloning and function analysis of GhPUB85A and GhPUB45D

The evolution relationship in Additional figure S5 showed *GhPUB68A, GhPUB85A, GhPUB45D* and *GhPUB69D* were belonged to subgroup Ⅳ, showing their close relationship between each other. The transcriptome data showed *GhPUB85A* and *GhPUB45D* were highly expressed whereas *GhPUB68A* and *GhPUB69D* were almost not expressed at all. In order to investigate the response to stresses of homologous genes in cotton, qRT-PCR was used to study the expression difference in *G. hirsutum* TM-1. Drought, salt and low-temperature treatments were conducted and the results could be found in **Figure 3**. Higher expression of *GhPUB85A* and *GhPUB45D* under three stresses suggested they were actively expressed to respond the abiotic stresses, but *GhPUB68A* and *GhPUB69D* were not. In the research, interestingly, the highest expressions of *GhPUB85A* and *GhPUB45D* were found at 6h under drought stress, but the expressions at 12h were the highest under salt and low-temperature stress, indicating *GhPUB85A* and *GhPUB45D* responded to drought stress faster than salt and low-temperature stress. However, significantly different expressions of two genes under the same stress were, showed their different contributions in responding to abiotic stresses.

Based on the gene expression, *GhPUB85A* and *GhPUB45D* were cloned using cDNA from *G. hirsutum* TM-1 as template, and connected to pEASY-Blunt Cloning Vector for sequencing. The sequencing and enzyme digestion results showed the length of targets were correct. Red fluorescence vectors pBI121-GhPUB85A:RFP and pBI121-GhPUB45D:RFP were constructed to research their subcellular localizations (**Figure 4**), and the results showed these two genes were located at cytomembrane, which were consistent with our prediction. Besides, two VIGS vectors pYL156:GhPUB85A and pYL156:GhPUB45D were constructed using In-Fusion technology to study their functions under different stresses. 15d after the VIGS infection, the leaves of the positive control plants were obviously albino, and all the new leaves were

albino in the later stage, while the others were normal (**Figure 5a**). Huge relative expression changes of two targets after the VIGS infection under different treatments showed their positive functions in responding multiple stresses and the success of VIGS infection (**Figure 5b-d**), which also suggested the VIGS infection technology was an effective way to study the gene functions in cotton.

Discussion

PUB gene family has been identified and analyzed in a number of plants [10, 12-14]. In the study, bioinformatics analysis was performed on allotetraploid cotton genome AD_1 and diploid cotton genomes A_2 and D_5, and finally a total of 374 *PUB* genes were identified, including 185 in genome AD_1, 96 in genome A_2, and 93in genome D_5, indicating that it is a relatively conservative family in cotton genome evolution. Whole-genome replication analysis revealed that the ancestors of *G. arboreum.* and *G. raimondii* had gone through a cotton-specific genome-wide replication event that occurred about 1.6 million years ago after differentiation with cocoa about 33 million years ago [33]. Recent studies have shown that *G. raimondii* have had gone through at least two complete genome-wide replicates [24, 38], resulting in an uneven distribution of the PUB genes on the chromosome, and over time, some genes are reassembled or lost. The results also showed that 19 out of 96 PUB genes in *G. arboreum* were generated through tandem repeats, which was one of the main reasons for the expansion of this gene family.

The classification of PUB protein differs from that of other gene families - it depends not only on U-box homology but on other domains other than U-box [39]. The evolutionary relationship of *PUB* genes between different cotton species is close, and the genetic structure in cotton is highly conservative. During the evolution of cotton, in addition to preserving the U-box domain, some other domains retain the basic functions of the family and enrich the diversity of the genes. The gene structure analysis showed that exon number of each PUB genes varied hugely from 1 to 25, which may be the result of directional evolution in function and structure of PUB gene during the long evolutionary process.

Many studies have shown that *PUB* genes play an important role in the process of stress resistance of plants, for example, the overexpression of *AtCHIP* gene in Arabidopsis results in its sensitivity to high temperature and low temperature [40], while *AtPUB18*, *AtPUB19*, *AtPUB22* and *AtPUB23* respond positively to drought [16, 38, 41]. In this study, two PUB genes *GhPUB85A* and *GhPUB45D* were cloned and identified. Significant expression difference revealed that the functions of homologous genes were similar, but their contributions were different. All results laid a foundation for the study of *PUB* genes of cotton in future.

Declarations

Ethics approval and consent to participate

All the cotton materials were collected from the Institute of Cotton Research, Chinese Academy of Agricultural Sciences, which are public and available for non- commercial purpose.

Consent for publication

Not applicable.

Availability of data and materials

All analysis results data generated during this study are included in thisarticle and its supplementary information.

Competing interests

The authors declare that they have no competing interests.

Funding

This work was supported by grants from National Key Research and Development Program (2016YFD0101006).

Author's Contributions

XL participated in study design, data analysis, and manuscript writing; NS, XC, BZ, SW, DW, JW and LG collected and analyzed data. WY was responsible for the study design and manuscript modification. All authors have read and approvad the final version of the manuscript.

Acknowledgements

We thank all members of our laboratories for the helpful assistances during the research. This work was supported by grants from National Key Research and Development Program (2016YFD0101006).

Reference

1. Schulman BA, Harper JW: **Ubiquitin-like protein activation by E1 enzymes: the apex for downstream signalling pathways**. *Nat Rev Mol Cell Bio* 2009, **10**(5):319-331.
2. Ye YH, Rape M: **Building ubiquitin chains: E2 enzymes at work**. *Nat Rev Mol Cell Bio* 2009, **10**(11):755-764.
3. Smalle J, Vierstra RD: **The ubiquitin 26S proteasome proteolytic pathway**. *Annual Review of Plant Biology* 2004, **55**:555-590.
4. Moon J, Parry G, Estelle M: **The ubiquitin-proteasome pathway and plant development**. *Plant Cell* 2004, **16**(12):3181-3195.
5. Cyr DM, Hohfeld J, Patterson C: **Protein quality control: U-box-containing E3 ubiquitin ligases join the fold**. *Trends Biochem Sci* 2002, **27**(7):368-375.
6. Harper JW, Burton JL, Solomon MJ: **The anaphase-promoting complex: it's not just for mitosis any more**. *Genes Dev* 2002, **16**(17):2179-2206.

7. Vierstra RD: **The ubiquitin-26S proteasome system at the nexus of plant biology**. *Nat Rev Mol Cell Bio* 2009, **10**(6):385-397.

8. Jackson PK, Eldridge AG, Freed E, Furstenthal L, Hsu JY, Kaiser BK, Reimann JD: **The lore of the RINGs: substrate recognition and catalysis by ubiquitin ligases**. *Trends Cell Biol* 2000, **10**(10):429-439.

9. Wiborg J, O'Shea C, Skriver K: **Biochemical function of typical and variant Arabidopsis thaliana U-box E3 ubiquitin-protein ligases**. *Biochem J* 2008, **413**(3):447-457.

10. Zeng LR, Park CH, Venu RC, Gough J, Wang GL: **Classification, expression pattern, and E3 ligase activity assay of rice U-box-containing proteins**. *Mol Plant* 2008, **1**(5):800-815.

11. Luo Q, Li Y, Wang W, Fei X, Deng X: **Genome-wide survey and expression analysis of Chlamydomonas reinhardtii U-box E3 ubiquitin ligases (CrPUBs) reveal a functional lipid metabolism module**. *PLoS One* 2015, **10**(3):e0122600.

12. Wang C, Duan W, Riquicho AR, Jing Z, Liu T, Hou X, Li Y: **Genome-wide survey and expression analysis of the PUB family in Chinese cabbage (Brassica rapa ssp. pekinesis)**. *Mol Genet Genomics* 2015, **290**(6):2241-2260.

13. Wang N, Liu Y, Cong Y, Wang T, Zhong X, Yang S, Li Y, Gai J: **Genome-Wide Identification of Soybean U-Box E3 Ubiquitin Ligases and Roles of GmPUB8 in Negative Regulation of Drought Stress Response in Arabidopsis**. *Plant Cell Physiol* 2016, **57**(6):1189-1209.

14. Cho SK, Chung HS, Ryu MY, Park MJ, Lee MM, Bahk YY, Kim J, Pai HS, Kim WT: **Heterologous expression and molecular and cellular characterization of CaPUB1 encoding a hot pepper U-Box E3 ubiquitin ligase homolog**. *Plant Physiol* 2006, **142**(4):1664-1682.

15. Cho SK, Ryu MY, Song C, Kwak JM, Kim WT: **Arabidopsis PUB22 and PUB23 are homologous U-Box E3 ubiquitin ligases that play combinatory roles in response to drought stress**. *Plant Cell* 2008, **20**(7):1899-1914.

16. Liu YC, Wu YR, Huang XH, Sun J, Xie Q: **AtPUB19, a U-Box E3 Ubiquitin Ligase, Negatively Regulates Abscisic Acid and Drought Responses in Arabidopsis thaliana**. *Molecular Plant* 2011, **4**(6):938-946.

17. Park JJ, Yi J, Yoon J, Cho LH, Ping J, Jeong HJ, Cho SK, Kim WT, An G: **OsPUB15, an E3 ubiquitin ligase, functions to reduce cellular oxidative stress during seedling establishment**. *Plant Journal* 2011, **65**(2):194-205.

18. Senchina DS, Alvarez I, Cronn RC, Liu B, Rong JK, Noyes RD, Paterson AH, Wing RA, Wilkins TA, Wendel JF: **Rate variation among nuclear genes and the age of polyploidy in Gossypium**. *Molecular Biology and Evolution* 2003, **20**(4):633-643.

19. Wendel JF, Cronn RC: **Polyploidy and the evolutionary history of cotton**. *Adv Agron* 2003, **78**:139-186.

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

21. Wang KB, Wang ZW, Li FG, Ye WW, Wang JY, Song GL, Yue Z, Cong L, Shang HH, Zhu SL *et al*: **The draft genome of a diploid cotton Gossypium raimondii**. *Nat Genet* 2012, **44**(10):1098-1104.

22. Li FG, Fan GY, Wang KB, Sun FM, Yuan YL, Song GL, Li Q, Ma ZY, Lu CR, Zou CS *et al*: **Genome sequence of the cultivated cotton Gossypium arboreum**. *Nat Genet* 2014, **46**(6):567-572.

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

24. 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):531-537.

25. Yuan D, Tang Z, Wang M, Gao W, Tu L, Jin X, Chen L, He Y, Zhang L, Zhu L *et al*: **The genome sequence of Sea-Island cotton (Gossypium barbadense) provides insights into the allopolyploidization and development of superior spinnable fibres**. *Sci Rep* 2015, **5**:17662.

26. Lu X, Yin Z, Wang J, Chen X, Wang D, Wang S, Guo L, Fan W, Chen C, Wang X *et al*: **Identification and function analysis of drought-specific small RNAs in Gossypium hirsutum L**. *Plant Sci* 2019, **280**:187-196.

27. Finn RD, Coggill P, Eberhardt RY, Eddy SR, Mistry J, Mitchell AL, Potter SC, Punta M, Qureshi M, Sangrador-Vegas A *et al*: **The Pfam protein families database: towards a more sustainable future**. *Nucleic Acids Res* 2016, **44**(D1):D279-285.

28. Finn RD, Clements J, Arndt W, Miller BL, Wheeler TJ, Schreiber F, Bateman A, Eddy SR: **HMMER web server: 2015 update**. *Nucleic Acids Res* 2015, **43**(W1):W30-38.

29. Yu CS, Chen YC, Lu CH, Hwang JK: **Prediction of protein subcellular localization**. *Proteins* 2006, **64**(3):643-651.

30. Schauser L, Wieloch W, Stougaard J: **Evolution of NIN-like proteins in Arabidopsis, rice, and Lotus japonicus**. *J Mol Evol* 2005, **60**(2):229-237.

31. Wei KF, Wang YM, Xie DX: **Identification and expression profile analysis of the protein kinase gene superfamily in maize development**. *Mol Breeding* 2014, **33**(1):155-172.

32. Chothia C, Gough J, Vogel C, Teichmann SA: **Evolution of the protein repertoire**. *Science* 2003, **300**(5626):1701-1703.

33. Zhu YX, Li FG: **The Gossypium raimondii Genome, a Huge Leap Forward in Cotton Genomics**. *Journal of Integrative Plant Biology* 2013, **55**(7):570-571.

34. Yang SH, Zhang XH, Yue JX, Tian DC, Chen JQ: **Recent duplications dominate NBS-encoding gene expansion in two woody species**. *Molecular Genetics and Genomics* 2008, **280**(3):187-198.

35. Gu ZL, Cavalcanti A, Chen FC, Bouman P, Li WH: **Extent of gene duplication in the genomes of Drosophila, nematode, and yeast**. *Mol Biol Evol* 2002, **19**(3):256-262.

36. Stone SL, Anderson EM, Mullen RT, Goring DR: **ARC1 is an E3 ubiquitin ligase and promotes the ubiquitination of proteins during the rejection of self-incompatible Brassica pollen**. *Plant Cell* 2003, **15**(4):885-898.

37. 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.

38. Bergler J, Hoth S: **Plant U-box armadillo repeat proteins AtPUB18 and AtPUB19 are involved in salt inhibition of germination in Arabidopsis**. *Plant Biol (Stuttg)* 2011, **13**(5):725-730.

39. Azevedo C, Santos-Rosa MJ, Shirasu K: **The U-box protein family in plants**. *Trends Plant Sci* 2001, **6**(8):354-358.

40. Yan JQ, Wang J, Li QT, Hwang JR, Patterson C, Zhang H: **AtCHIP, a U-box-containing E3 ubiquitin ligase, plays a critical role in temperature stress tolerance in Arabidopsis**. *Plant Physiology* 2003, **132**(2):861-869.

41. Seo DH, Ryu MY, Jammes F, Hwang JH, Turek M, Kang BG, Kwak JM, Kim WT: **Roles of Four Arabidopsis U-Box E3 Ubiquitin Ligases in Negative Regulation of Abscisic Acid-Mediated Drought Stress Responses**. *Plant Physiology* 2012, **160**(1):556-568.

Table 1

Table 1 Domain organizations of PUB proteins in cotton

No.	Composition of domain (from N end to C end)	Number of proteins			
		G.raimondii	*G.arboreum*	*G. hirsutum*	G. barbadense
1	UFD2 specific motif + U-box	1	1	2	2
2	U-box + ARM	43	45	78	90
3	U-box only	33	35	68	83
4	STYKc + U-box	3	3	6	7
5	S_TKc + U-box	3	3	6	6
6	U-box + Prp19 +WD40	3	2	3	6
7	U-box + WD40	2	2	6	3
8	TPR + U-box	1	2	3	3
9	U-box + PPIase	1	1	2	2
10	U-box + KAP	1	0	1	0
11	UspA + S_TKc + U-box	2	2	2	2
12	UspA + STYKc + U-box	0	0	2	1
13	NRAMP + U-box + ARM	0	0	2	0
14	Ank + U-box + ARM	0	0	1	0
15	Ank + U-box + ARM + Ank	0	0	1	0
16	RPOL_N + RNA_pol + U-box + ARM	0	0	1	0
17	U-box + WD40 + S_TKc	0	0	1	0
18	EGF + PAN_AP + U-box + ARM	0	0	0	1
19	U-box + Arm + DNA_pol3	0	0	0	1
20	Pkinase + U-box	0	0	0	1

Additional File Details

Additional file 1: **Table S1.** Essential information of PUB gene members in *Gossypium raimondii*

Additional file 2: **Table S2.** Essential information of PUB gene members in *Gossypium arboreum*

Additional file 3: **Table S3.** Essential information of PUB gene members in *Gossypium hirsutum* L.

Additional file 4: **Table S4.** Essential information of PUB gene members in *Gossypium barbadense*

Additional file 5: **Figure S1.** The phylogenetic relationship and gene structure analysis of GrPUBs in *G. raimondii*

Additional file 6: **Figure S2.** The phylogenetic relationship and gene structure analysis of GaPUBs in *G. arboreum*

Additional file 7: **Figure S3.** The phylogenetic relationship and gene structure analysis of GhPUBs in *G. hirsutum*

Additional file 8: **Figure S4.** The phylogenetic relationship and gene structure analysis of GbPUBs in *G. barbadense*

Additional file 9: **Figure S5.** The phylogenetic relationship analysis of PUBs in Gossypium

Additional file 10: **Figure S6.** Distrbution of GbPUBs on chromosomes in *G. barbadense*

Additional file 11: **Figure S7.** The homologous relationships of PUBs in *G. raimondii* and *G. arboreum*

Additional file 12: **Figure S8.** The intra- and inter-genomic synteny blocks of PUBs

Additional file 13: **Figure S9.** Predicted expression pattern of GhPUBs in upland cotton under salt and drought stress

Additional file 14: **Figure S10.** Predicted expression pattern of GhPUBs in upland cotton under cold and heat stress

Figures
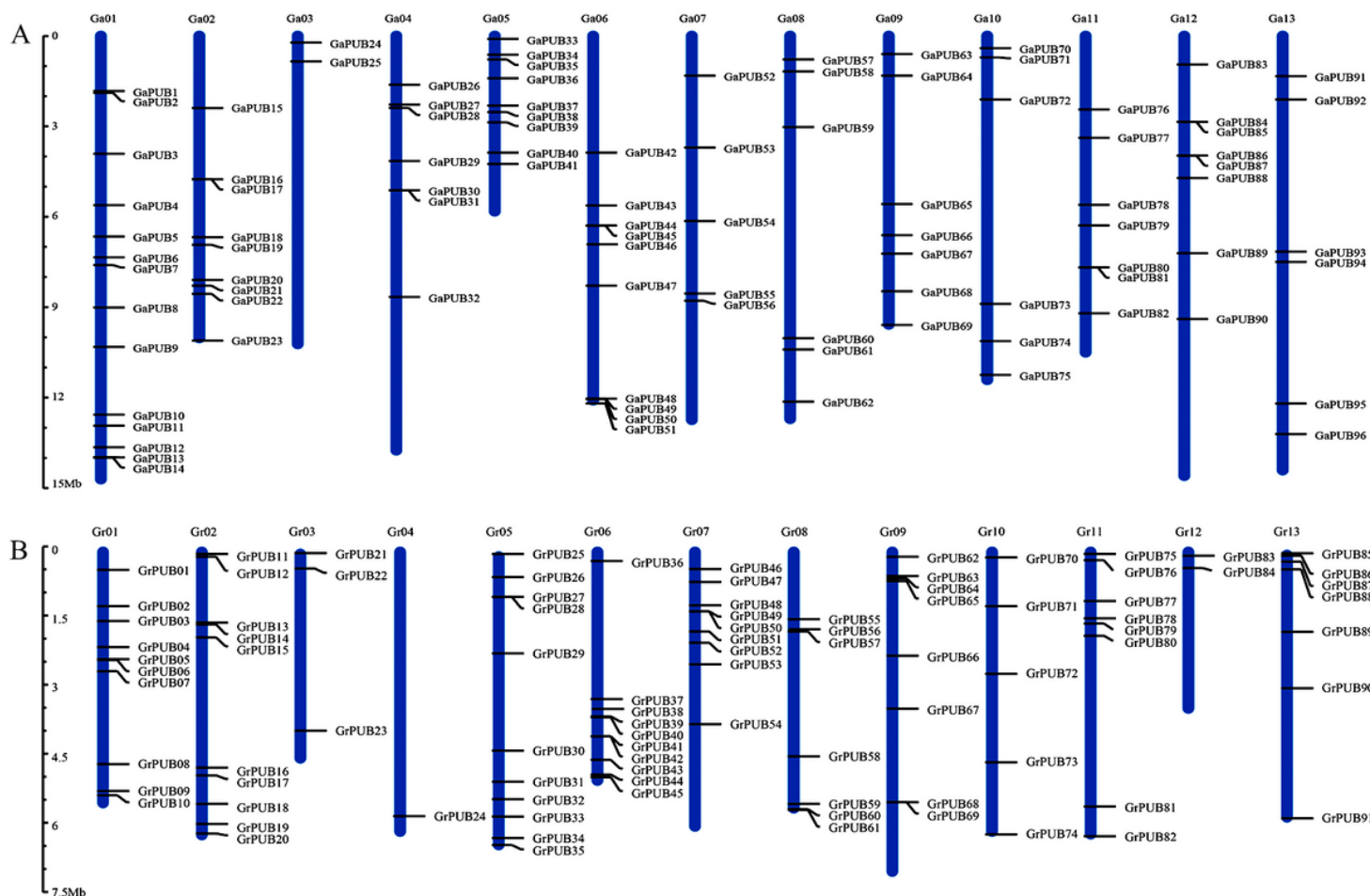
Figure 1

Locations of PUB genes on chromosomes in in G. raimondii and G. arboreum Ga01-Ga13 and Gr01-Gr13 represent the chromosome 1 to chromosome 13 in G. arboretum and G. raimondii, respectively.
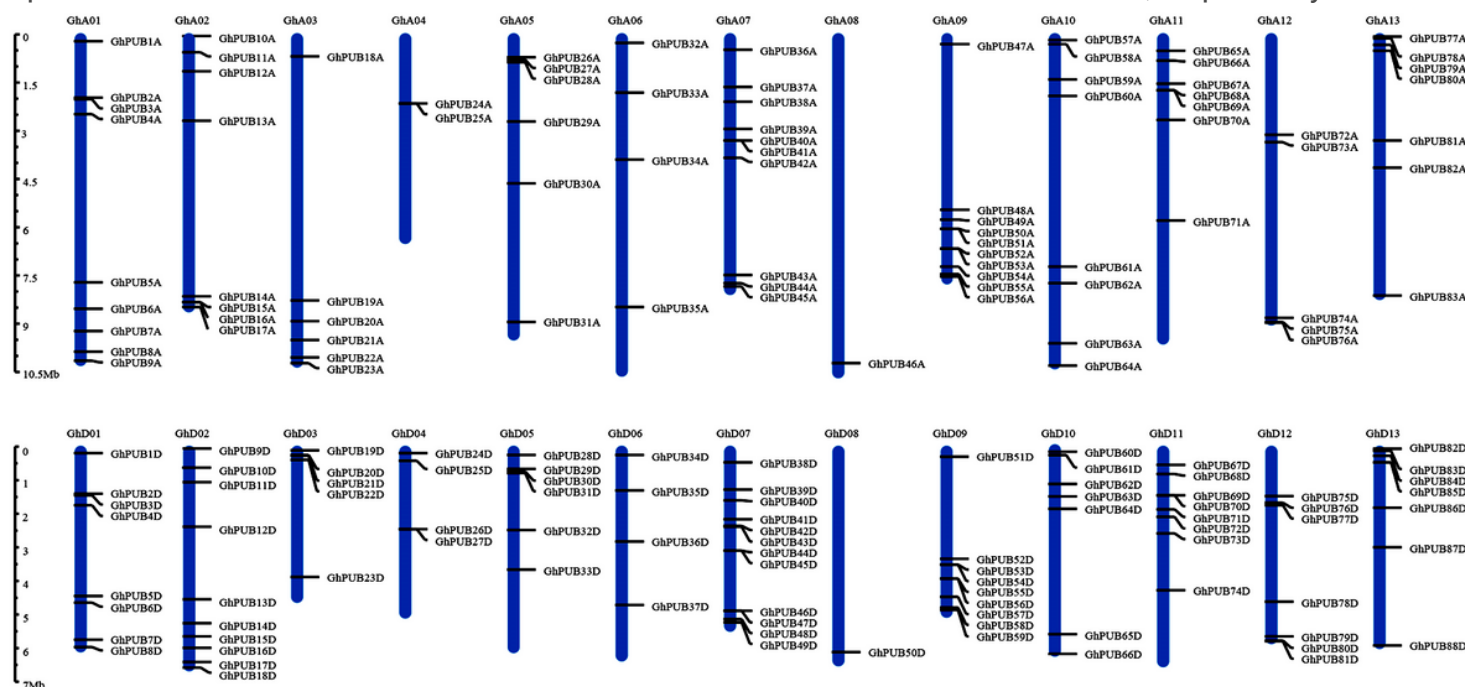
Figure 2

Locations of PUB genes on chromosomes in in G. hirsutum Replace the chromosome 1 to chromosome 13 of A subgenome of G. hirsutum with GhA01-GhA13, Replace the chromosome 1 to chromosome 13 of D subgenome of G. hirsutum with GhD01-GhD13.
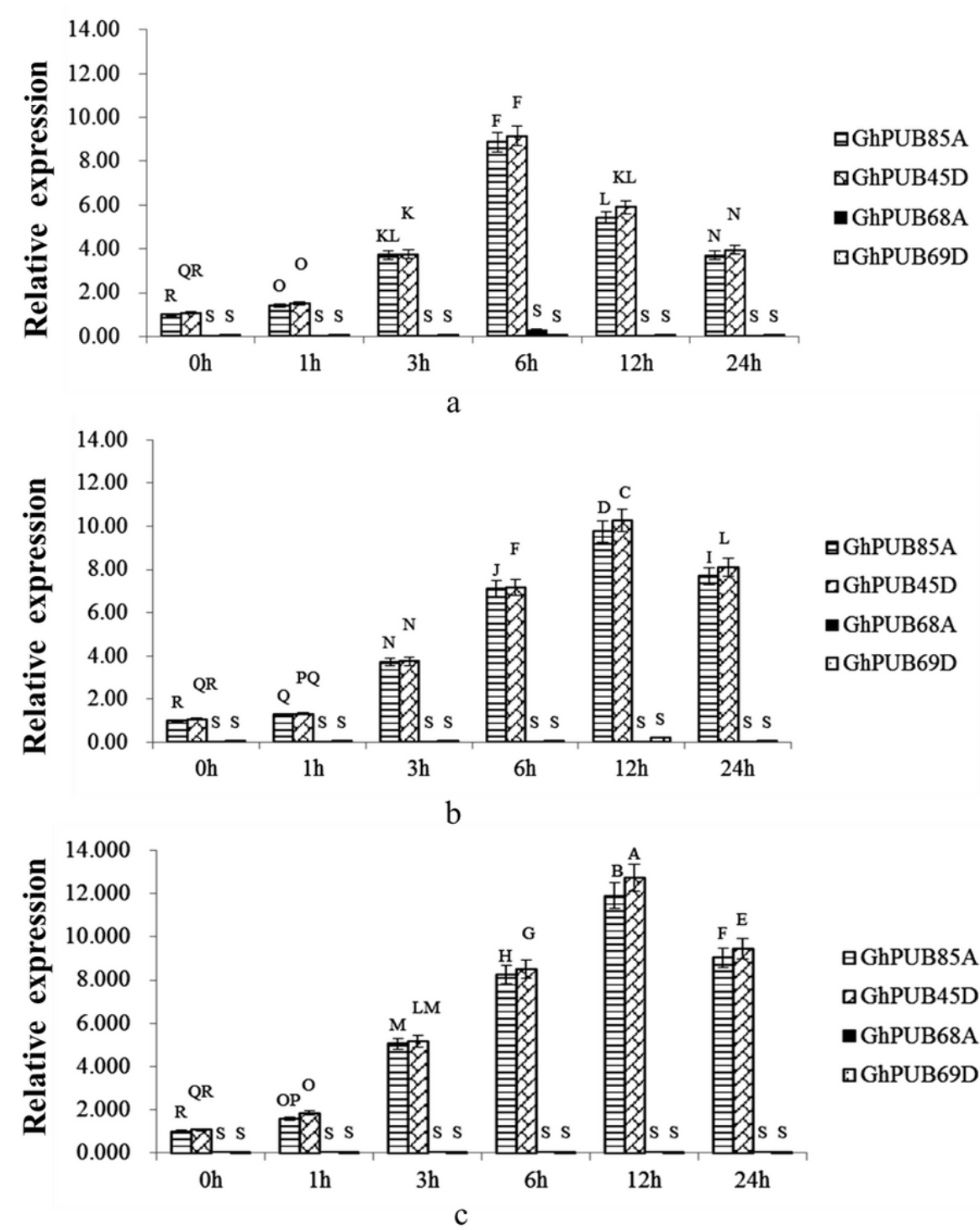
Figure 3

Expression patterns of GhPUB68A GhPUB85A GhPUB45D and GhPUB69D during the drought, salt or low temperature stress a, b and c represent drought, salt and low-temperature treatment, respectively. Different letters from A to S indicate significance expression of different genes during different stresses (p < 0.01).

Figure 4

Subcellular localization of GhPUB85A and GhPUB45D
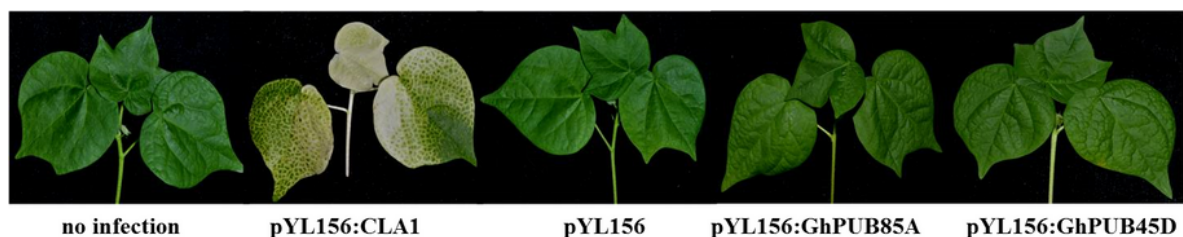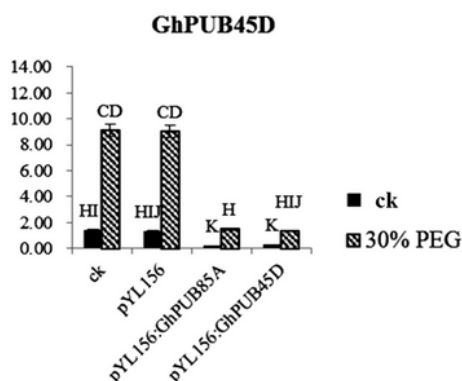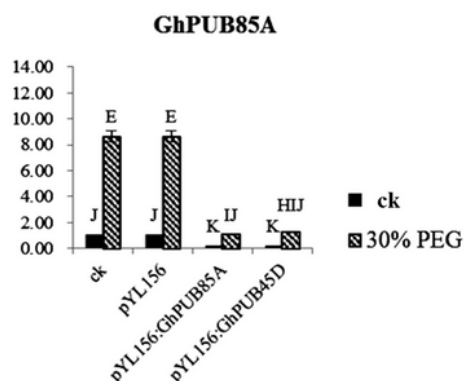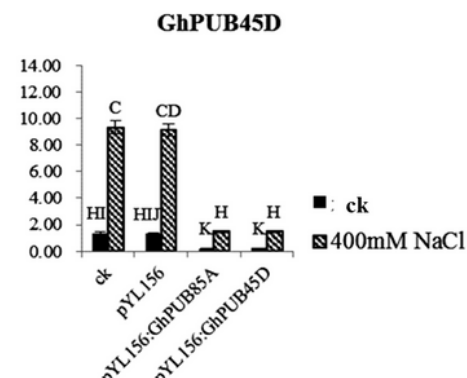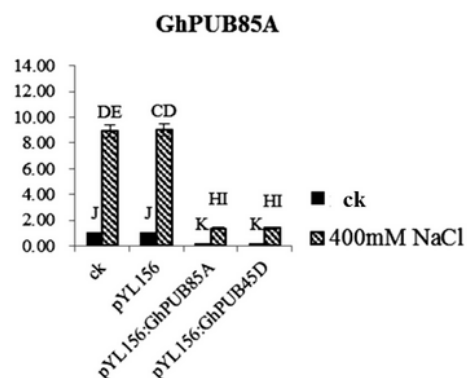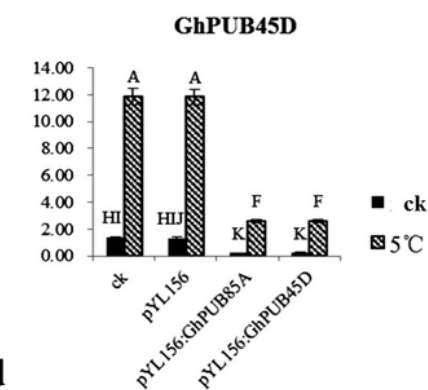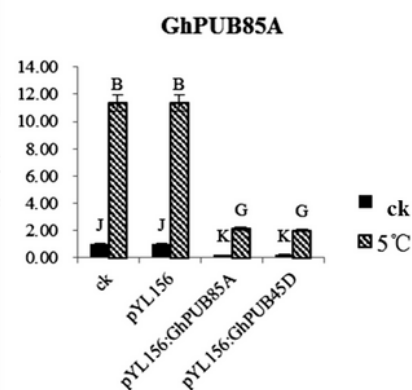
Figure 5

The phenotype of cotton leaves after virus infection and expression patterns of GhPUB45D and GhPUB69D under the drought, salt and low temperature stress Different letters from A to K indicate significance expression of different genes during different stresses ($p < 0.01$).