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 Agricultural 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

Chao Chen  
Institute of Cotton Research, Chinese Academy of Agricultural Sciences

Wuwei Ye  (✉ yew158@163.com )  
Institute of Cotton Research, Chinese Academy of Agricultural Sciences  https://orcid.org/0000-0003-4757-1057

Research article

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

Posted Date: January 15th, 2020

DOI: https://doi.org/10.21203/rs.2.10518/v4

License: ☺ ☑ This work is licensed under a Creative Commons Attribution 4.0 International License.  
Read Full License
Version of Record: A version of this preprint was published on March 6th, 2020. See the published version at https://doi.org/10.1186/s12864-020-6638-5.
Abstract

Background: The U-box gene encodes a ubiquitin ligase that contain U-box domain. The plant U-box gene (PUB) plays an important role in the response to stress, but few reports about PUBs in cotton were available. Therefore research on PUBs is of great importance and is a necessity when studying the mechanism of stress tolerance in cotton. Results: In this study, we identified 93, 96, 185 and 208 PUBs from four sequenced cotton species G. raimondii (D5), G. arboreum (A2), G. hirsutum (AD1) and G. barbadense (AD2), respectively. Prediction analysis of subcellular localization showed that the PUBs in cotton were widely distributed in cells, but primarily in the nucleus. The PUBs in cotton were divided into six subfamilies (A-F) on the basis of phylogenetic analysis, and the intron/exon structure was comparatively conserved within each subfamily. Location analysis showed that cotton PUBs were unevenly anchored on all chromosomes, varying from 1 to 14 per chromosome. Through multiple sequence alignment, 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 shared in A2 and D5 were found; however no tandem duplication region in A2 or D5 was found. Additionally, 105, 14 and 17 homologous gene pairs were found in the intra-subgenome of At and Dt, the At subgenome and the Dt subgenome of allotetraploid cotton, respectively. Functional analysis of GhPUB85A and GhPUB45D showed that these genes positively responded to abiotic stresses, but the expression patterns were different. In addition, although the expression levels of these two homologous genes were similar, their contributions were different when responding to stresses, specifically showing different responses to abiotic stresses and functional differences between the two subgenomes of G. hirsutum. Conclusion: This study reports the genome-wide identification, structure, evolution and expression analysis of PUBs in cotton, and the results showed that the PUBs were highly conserved throughout the evolutionary history of cotton. All PUB genes were involved in response to abiotic stresses (including those genes induced by salt, drought, hot and cold) to varying degrees.

Background

The ubiquitin-mediated ubiquitination pathway is the post-translational modification pathway of eukaryotic proteins. Studies have demonstrated that the pathway is involved in the cell cycles of higher plants [1], stress resistance [2], signal transduction [3], apoptosis [4], optical signal [5] and other physiological pathways. In the ubiquitin pathway, three steps are required for ubiquitin to act on the target protein. First, the ubiquitin-activating enzyme (E1) activates ubiquitin [1], and then the activated Ub molecules are delivered to the ubiquitin-binding enzyme (E2) [2]; finally, the Ub molecules are transferred to the target protein through ubiquitin ligase (E3). E3 is critical for the identification of the specific substrate protein and is found in the most species [3]. In Arabidopsis, there are more than 1,400 genes encoding components of the ubiquitination pathway, of which approximately 90% encode ubiquitin ligase E3 [3, 4]. According to its subunit composition and mechanism of action, the ubiquitin ligase E3 can be categorized into single subunit types (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 have been reported in many model crops, including Arabidopsis, rice, Chlamydomonas reinhardtii, Chinese cabbage, and soybean. Previous studies identified 64 PUBs in Arabidopsis [9], 77 in rice [10], 33 in Chlamydomonas reinhardtii [11], 101 in Chinese cabbage [12] and 125 in soybean [13], indicating that PUB genes widely distributed and play important roles in plants. Many studies have shown that PUB proteins are involved in abiotic stress responses. Cho et al. [14] obtained U-box E3 protein (CaPUB1) from drought-treated peppers. In Arabidopsis, the heterologous overexpression of CaPUB1 attenuated the tolerance of plants to drought stress, and two homologous proteins AtPUB22 and AtPUB23, in Arabidopsis are also involved in the drought stress response of the plants [15]. Furthermore, studies also showed that AtPUB22 and AtPUB23 can negatively modulate 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 levels, cold and ABA. Mutated atpub19 in plants promoted stomatal closure, thereby enhancing the tolerance to drought; AtPUB19-overexpressing plants were not sensitive to ABA but were sensitive to drought. In addition, there are significant changes in the expression levels of ABA and stress-related genes in atpub19 mutant and AtPUB19-overexpressing plants, indicating that AtPUB19 may regulate the drought stress by negatively modulating the ABA signaling pathway [16]. PUB proteins, involved in the process of drought response, were also identified in rice. Previous studies showed that the drought resistance of OsPUB15-overexpressing plants was significantly enhanced, and OsPUB15 could be induced by hydrogen peroxide, drought and high salt levels, 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 research into polyploidy, evolution, cell wall development, and cellulose synthesis [18]. Approximately 50 species were 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, and tetraploid G. hirsutum and G. barbadense. The tetraploid cottons originated from the hybridization of an African ancestral species with A genome and an American ancestor species with D genome one to two million years ago [18]. Recently the sequencing work of diploid cottons (G. raimondii (D5) [20, 21], G. arboreum (A2) [22] and allotetraploid cottons (G. hirsutumTM-1 (AD1) [23, 24], G. barbadense acc.3-79 (AD2) [25] were completed, providing references for the study of gene function and evolution at the whole genome level. Based on the cotton genome sequences, the PUB gene family members of the PUB gene were identified and the structure and distribution characteristics of PUB genes were analyzed, and their evolution analysis was analyzed in whole-genome replication and cotton allotetraploid formation studies.

Results

Identification of PUB gene family members in the whole genome of cotton

The hidden Markov model (HMM) of the U-box domain (PF04564) was downloaded from the Pfam30.0 database, and used as a query to identify the candidate PUB members in four cotton genomic database using HMMER3.0. SMART. In addition, Pfam30.0 was also used for further identification to confirm every
PUB members containing U-box domain. Finally, 93, 96, 185, and 208 PUBs were identified from the four sequenced cotton species, *G. raimondii* (D₅), *G. arboreum* (A₂), *G. hirsutum* acc. TM-1 (AD₁), and *G. barbadense* (AD₂), respectively, and these PUBs 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 number of PUB genes in tetraploid cottons was twice as high as that of diploid cottons, showing that PUB genes are relatively conservative. The information about the gene name, chromosome locations, length of the open reading frame (ORF), type of protein domain, position of the U-box domain and subcellular localizations of these gene family members could be found in additional files (Additional file 2: Table S2, Additional file 3: Table S3, Additional file 4: Table S4 and Additional file 5: Table S5). The length of the PUB protein sequence in the cotton ranged from 49 to 1492 AA, and the U-box domain contained approximately 75 amino acids. However, the length of the U-box domain was almost identical except for a few PUBs; for example, GaPUB39 and GhPUB40D had only 32 and 50 amino acids, respectively. Results from the subcellular localization analysis showed PUB proteins distributed throughout the cell, including nuclear, cytoplasmic, chloroplast, plasma membrane, mitochondrial, and extracellular locations. However, most PUB proteins were located inside the nucleus. Twenty different domains were found among all the 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

A Gene structure diagram of the PUBs and an evolution tree were constructed (Additional file 6: Figure S1, Additional file 7: Figure S2, Additional file 8: Figure S3 and Additional file 9: Figure S4). Based on the evolutionary relationship, the PUB genes could be categorized into five subgroups (I-V). Among these subgroups, the subgroup I was composed of the domains “U-box + ARM” and “U-box only”, and the remaining subgroups were composed of the 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 of the PUBs contained only one exon. Generally, the evolutionary relationship is correlated with gene structure in some way, that is, exons with the more similarities in terms of the number and size of the exon, have a closer the evolutionary relationship. In *G. hirsutum*, the length of GhPUB1A is 47 Kb, much larger than the other PUB genes, which may be correlated with the assembly and annotation of the cotton genome. Members in each subgroup of *G. barbadense* (AD₂) was much different than those in *G. raimondii* (D₅), *G. arboreum* (A₂) and *G. hirsutum* (AD₁), and this difference may be correlated with the different origins of these species. Therefore, the evolution of the PUBs in *G. raimondii* (D₅), *G. arboreum* (A₂) and *G. hirsutum* (AD₁) was analyzed, and the results also indicated-Ⅲ-five subgroups were found (Additional file 10: Figure S5), and this was similar with the evolution of single PUB gene, suggesting the PUB members were highly conservative. Furthermore, closer evolution relationships of GhPUB1A-89A with GaPUB1-96 and GhPUB1D-91D with GaPUB1-93 were found through the evolutionary analysis.

### Location of PUB genes in three cotton genomes
The MapInspect software was used to generate 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 the others were found on scaffolds (Figure 1A). These results indicated that only a few genes were present on chromosomes 3, 4, and 12, and the chromosome 5 contained the highest number of PUB genes (11 PUBs). In addition, PUB genes on chromosomes 4, 6, 7, 11 and 12 were preferentially 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* with chromosome 1 being the most enriched with PUB genes (up to 14) and chromosome 3 being enriched the least with PUB genes (only 2). In addition, the length of chromosome 5 was approximately 6 Mb, but 9 PUB genes were found on it, showing the highest distribution density. In *G. hirsutum*, 91.4% (169/185) of the PUB genes were anchored onto chromosomes, among which 82 and 87 genes were found in the 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- subgenomes of *G. hirsutum*, showing that PUBs on these two chromosomes were relatively conserved and significant for cotton growth. The situation for *G. barbadense* was different with that of *G. hirsutum* (Additional file 11: Figure S6). The results showed that the PUB gene were equally distributed in At- and Dt- subgenomes and unevenly located on each chromosome, which may be correlated with the differentiation of these wo 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 [26]. The segregation of cotton A genome and D genome diploid occurred in about 5-10 Myr years ago [18], and allotetraploid *G. hirsutum* was generated from the 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 the 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 [27]. Based on the multiple sequence alignment of the encoding sequences and the proteins in diploid cotton, 18 and 27 homologous gene-pairs were discovered with MCScanX [28] in *G. raimondii* (D₅) (Additional file 12: Figure S7A) and *G. arboreum* (A₂) (Additional file 12: 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 these two diploid cottons and *G. hirsutum* was analyzed (Additional file 13: Figure S8). Totally 197 homologous gene-pairs were found
between *G. raimondii* and *G. hirsutum*, among which 58.89% (116/197) were located in the Dt-subgenome, and 191 homologous gene-pairs were found in both *G. arboreum* and *G. hirsutum*, among of which 55.50%(106/191) were located in the At-subgenome. All these results showed that more than one half of homologous genes in *G. hirsutum* were derived from the corresponding diploid cotton genomes. Furthermore, approximately 41.11% - 44.50% of these homologous genes were originated from other diploid genomes.

**Expression pattern analysis of PUB genes in cotton**

Based on previous transcriptome data of the PUBs under different stresses (salt, drought, hot and cold) in *G. hirsutum*, 117, 148 and 119 PUB genes were found with FPKM >1 in roots, stems and leaves, respectively, thus displaying tissue specificity. Among all the PUB genes, approximately 21 non-expressed PUB genes were identified in three tissues, and they may be associated with other specific regulation functions. Interestingly, it was found that 3 PUB genes (including GhPUB58D, GhPUB55A and GhPUB67D) were always highly expressed in three tissues under salt, drought, cold and hot stresses. All the PUB genes were categorized into five subgroups (I, II, III, IV and V), and similar expression patterns were found among all PUB genes (Additional file 14: Figure S9 and Additional file 15: Figure S10). In subgroup I, 18 PUB genes with profound expression differences were discovered; in addition, other PUB genes in subgroup II- IV were found to have a consistent expression pattern under different stresses. However, 4 PUB genes (GhPUB32A - GhPUB38D) in subgroup V showed a small expression difference under different stresses.

**Cloning and functional analysis of GhPUB85A and GhPUB45D**

The evolutorial relationship in Additional figure S5 showed *GhPUB68A, GhPUB85A, GhPUB45D* and *GhPUB69D* were belonged to subgroup III, indicating that their close relationship with each other. The transcriptome data showed that *GhPUB85A* and *GhPUB45D* were highly expressed whereas *GhPUB68A* and *GhPUB69D* were negligibly expressed. To investigate the functions of the homologous genes in cotton, qRT-PCR was used to investigate the expression difference in *G. hirsutum* TM-1. Drought, salt and cold treatments were applied and the results were present in Figure 3. High expression of *GhPUB85A* and *GhPUB45D* under three stresses suggested that they were actively expressed to respond the abiotic stresses, but *GhPUB68A* and *GhPUB69D* were not highly expressed, which was in line with previously reported transcriptome data. Interestingly, we found that *GhPUB85A* and *GhPUB45D* were highly expressed at 6h under drought stress, while the expressions at 12h were the highest under salt and cold stress, indicating that *GhPUB85A* and *GhPUB45D* responded to drought stress faster than they did to salt and cold stresses. However, the expression values of *GhPUB85A* and *GhPUB45D* were significantly different under the same stress conditions, showing their different contributions in responding to abiotic stresses.

In addition, *GhPUB85A* and *GhPUB45D* were cloned using cDNA from *G. hirsutum* TM-1, and ligated to pEASY-Blunt Cloning Vector for sequencing to verify whether the vector was correctly ligated. The sequencing and enzyme digestion results showed that the recombined vectors were correctly constructed.
Red fluorescence vectors pBI121-GhPUB85A:RFP and pBI121-GhPUB45D:RFP were constructed to research their subcellular localizations (Figure 4), and the results showed that these two genes were located at the cytomembrane, which were consistent with our prediction in Additional table S2. In addition, two VIGS vectors pYL156:GhPUB85A and pYL156:GhPUB45D were constructed using In-Fusion technology to study their functions under different stresses. Fifteen days after the VIGS infection, albino leaves of the positive control plants were observed, and all newly-emerged leaves were white in the later stage, while the others were normal with no albino leaves (Figure 5a). We investigated the expression quantity in the control plants (CK), and pYL156-, pYL156:GhPUB85A- and pYL156:GhPUB45D-infected plants under different stresses. The expression levels of two genes decreased significantly after the VIGS infection under different treatments showed their positive functions in responding to multiple stresses and the success of VIGS infection (Figure 5b-d), indicating the VIGS infection technology was an effective way to study the gene functions in cotton.

**Discussion**

The PUB gene family has been identified and analyzed in a number of plants [10, 12-14]. In this study, bioinformatics analysis was performed on allotetraploid cotton genomes AD\textsubscript{1} and AD2, and diploid cotton genomes A\textsubscript{2} and D\textsubscript{5}, and finally a total of 582 PUB genes were identified, including 185 genes in AD\textsubscript{1} genome, 208 in AD2 genome, 96 in A\textsubscript{2} genome, and 93 in D\textsubscript{5} genome, indicating that it is a relatively conserved family in terms of cotton genome evolution. Whole-genome replication analysis revealed that the ancestors of *G. arboreum* and *G. raimondii* had undergone a cotton-specific genome-wide replication event that occurred about 1.6 million years ago after the differentiation from cocoa about 33 million years ago [29]. Recent studies have shown that *G. raimondii* has undergone at least two complete genome-wide replications [24, 30], resulting in an uneven distribution of the PUB genes on the chromosomes, and over the time of the cotton evolution, some genes are reassembled or lost. The results also showed that 19 of the 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. Gene duplication event is a common phenomenon in plants, including multiple forms, such as tandem duplication, segmental duplication, and whole-genome duplication [31, 32]. Some of duplicated genes could be retained in its descendants, which could provide original genetic resource for adaptive evolution of plants [33]. In this research, gene duplication event was commonly found, totally 31 and 27 gene pairs were discovered in D5 and A2 genome, respectively. In *G. hirsutum*, more gene pairs were found than the sum of A2 and D5, which might be associated with the higher resistance and wider adaptability of *G. hirsutum*.

The classification of PUB protein differs from that of the other gene families - it depends not only on U-box homology but also on domains other than U-box domains [34]. The evolutionary relationship of PUB genes between different cotton species is close, and the genetic structure in cotton is highly conserved. During the evolution of cotton, in addition to the U-box domain, some other domains retained the basic functions of the family and enriched the diversity of PUB genes. Gene structure analysis showed that exon number of PUB genes varied greatly from 1 to 25, which might be due to the directional evolution in
the function and structure of PUB genes during the long evolutionary history. All PUB genes could be divided into five subgroups (I-V) in each species according to the evolutionary relationship, which was different with the classification of U-box containing genes in *C. reinhardtii* [11]. In the study, subgroup II and IV were found containing only 4 and 2 PUB genes, respectively. These findings suggested that the gene structures and evolutionary relationship of PUB family members were significantly different between different species. Previous studies have demonstrated that plant U-box containing genes are widely involved in stress responses, disease resistance and nutrient defect responses in plants [35-37]. In this study, 89% of PUB genes were differentially expressed in three tissues under salt, drought, cold and hot stresses, which also proved that PUB genes played an important role in abiotic responses. Twenty-two PUB genes, including 18 genes from *GhPUB8A* to *GhPUB51D* in subgroup I and 4 genes from *GhPUB32A* to *GhPUB38D* in subgroup V, showed a significant expression difference under different abiotic stresses, suggesting that they evolved toward specific functions in the long history. This perspective was consistent with previous documents[38, 39].

Many studies have shown that PUB genes play an important role in the process of stress tolerance in plants. For example, the overexpression of *AtCHIP* gene in Arabidopsis results in its sensitivity to high temperature and low temperature [40], while the *AtPUB18*, *AtPUB19*, *AtPUB22* and *AtPUB23* respond positively to drought [16, 30, 41]. In this study, two homologous PUB genes *GhPUB85A* and *GhPUB45D* were discovered with the same length of ORFs, type of protein, and subcellular localizations. The only difference was their locations on chromosomes, *GhPUB85A* on chromosome 7 in At-subgenome while *GhPUB45D* on chromosome 7 in Dt-subgenome, which was not reported before. In addition, these two PUB genes *GhPUB85A* and *GhPUB45D* were cloned and functionally identified. Significant expression differences revealed that the functions of these homologous genes were similar in response to abiotic stresses, but their contributions differed from each other. Both of *GhPUB85A* and *GhPUB45D* contained the protein domain TPR, which was the same with *AtCHIP*, so we speculated that the functions of these two genes might be similar with *AtCHIP* [42], at least in response to cold stress which had been verified in the study. The results from this study laid a foundation for the further study of PUB genes of in cotton in future.

**Conclusions**

The identification and expression analysis of PUB family members in cotton in this manuscript provided insights into response mechanism to abiotic stresses. Although the PUBs were highly conserved throughout the evolutionary history of cotton, significant differences were found between each other in gene structure. In addition, two homologous genes *GhPUB85A* and *GhPUB45D* were cloned and functionally identified. Expression pattern analysis showed they both responded to abiotic stresses positively, but their contributions were different. Therefore all these results were of great significance for the future research of molecular mechanism in responding to abiotic stresses.

**Methods**
Cotton and *Arabidopsis thaliana*

Upland cotton variety ZhongS9612 was selected and has been preserved by the Cotton Adversity Research Laboratory at the Chinese Academy of Agricultural Sciences (CAAS) for many years. Before planting into 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 a greenhouse (14 h/day at 30°C and 10 h/night at 24°C) at the Institute of Cotton Research of CAAS. The cotton seedlings were treated with 200 ml of 30% PEG 6000 and 200 ml 400 mM of NaCl solution, which could achieve completely consistent stress environments in each pot. For the cold treatment, cotton seedlings were transferred into a 5°C refrigerator with clear glass, but the light condition was not changed. For the planting of *Nicotiana benthamiana*, we first placed the seeds on MS medium at 22°C in a growth chamber with a 16-h light cycle for the germination. After the emergence, the tobacco seedlings were moved to nutrition-enriched soil for growth under the same condition for about 45 days. After the treatment, plant leaves were harvested and frozen with liquid nitrogen for use. For the agrobacterium-mediated transformation, we referred the method used by Lu et al [43].

**Whole genome identification of PUBs in cotton**

Cotton genome data (*G. raimondii* (D5) [20, 21], *G. arboretum* (A2) [22], and *G. hirsutum* acc. TM-1 (AD1) [23, 24]) were obtained from CottonGene (https://www.cottongen.org/). The hidden Markov Model (HMM) profile of the U-box domain (PF04564) was obtained from Pfam30.0 (http://pfam.xfam.org/) [44], and was used as a query to identify the candidate PUBs from the cotton genome protein database using HMMER3.0 [45]. We used BLAST2.2.31+ (ftp://ftp.ncbi.nlm.nih.gov/blast/executables/blast+/LATEST/) to obtain the coding domain sequences (CDSs) sequences, protein sequences and the corresponding full-length sequence in the genome. The protein sequences were further analyzed in the SMART (http://smart.embl-heidelberg.de/) and Pfam 30.0 [44] databases to ensure that each candidate protein contained a U-box domain. A subcellular localization prediction was carried out in CELLO v.2.5 [46].

**Exon-intron structure, phylogenetic and conserved domain analysis**

The identified CDS and the genome sequence of the cotton PUBs were used to analyze the PUBs structure by GSDS2.0 [26]. The full-length sequences of PUB proteins were used to construct a phylogenetic tree. Multi-sequence alignment of the PUBs was carried out by ClustalX1.83, and Neighbor-Joining (NJ) method [47] was used to construct a phylogenetic tree in MEGA6.0 [27]. The online software SMART, PROSITE [29] was used to analyze the conserved domains of each protein.

**Physical location of PUBs in cotton**

GFF (general feature format) files of the cotton PUBs were obtained from the genome annotation files, and the information on PUBs of chromosome positions was predicted from it. The distribution of cotton PUBs on the chromosome was generated with MapInspect (http://mapinspect.software.informer.com/).
Gene duplication and micro-synteny analysis of *G. arboreum*, *G. raimondii* and *G. hirsutum* L.

Homologous gene pairs were identified according to multiple sequence alignment results and the standard was described in previous studies [48, 49]. The collinearity of homologous genes was visualized with program Circos-0.69 [50] based on the homology between each species and their positions on the genome.

**Gene cloning and the construction of vectors**

The first strand was synthesized according to the instructions of the 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 using the website and sequences were present in Additional file 1: Table S1. The PCR amplification products were verified using 1.5% agarose gel electrophoresis. PCR amplification products were purified using the MiniBEST Agarose Gel DNA Extraction Kit from Takara Corporation. Finally the concentration of targets was measured, and stored at -20°C for use.

Purified targets were linked to the pEASY-Blunt Cloning Vector, and then transformed into *E. coli* according to the instructions of pEASY-Blunt Cloning Kit. The transformation was conducted with heat-shock method, which was a classically effective approach. First a mixture of vectors and products was prepared with a mole ratio of 1:7 before the reaction in 200 μl centrifuge tube at 25°C for 5 min. Second blend the linked products and *E. coli* DH5α cells when *E. coli* DH5α cells began to dissolve, then the reaction was performed on the ice for 30 min. Thirdly transfer the centrifuge tubes to 42°C water for 90 s to end the reaction, at last placed the centrifuge tubes on the ice for use. 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 primers. Finally cloning 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 (10 μM), 1.5 μl; R-primer (10 μM), 1.5 μl; cDNA, 100 ng; 5×PrimerSTAR GXL DNA Polymerase, 100 ng, 5×PrimerSTAR GXL DNA Polymerase, 2.0 μl; add ddH₂O 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 were listed in Additional file 1: Table S1. *GhHistone3* gene (AF02471) was used as the reference gene.

**VIGS analysis of GhPUB85A and GhPUB45D**

Based on the vector sequence, target gene sequence and enzyme cutting sites, In-fusion primers were designed at the website (http://bioinfo.clontech.com/infusion). Primer sequences of InGhPUB85A-V and InGhPUB45D-V were previously shown. Primer sequences of GhPUBs-RFP and InGhPUBs were listed in Additional file 1: Table S1. Finally the silencing vectors pYL156:GhPUB85A and pYL156:GhPUB45D, and RFP vectors pBI121-GhPUB85A:RFP and pBI121-GhPUB45D:RFP, and the Plant overexpression vectors pBI121:GhPUB85A and pBI121:GhPUB45D were totally constructed. Vectors were transformed into cotton and Arabidopsis thaliana with agrobacterium mediated genetic transformation method.
Abbreviations

PUB: U-box gene; SCF: skp1-cullin-F-box; APC: anaphase-promoting complex; VBC: VHL-Elongin B-Elongin C; NJ: Neighbor-Joining

Declarations

Acknowledgements

We thank all members of our laboratories for the helpful assistances during the research. We also were grateful for the assistance of manuscript revisions by John Z. Yu from Crop Germplasm Research Unit, Southern Plains Agricultural Research Center, US Department of Agriculture, College Station, TX, USA.

Funding

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

Availability of data and materials

All data generated and results analyzed during this study are included in this article and its supplementary information.

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 revisions. CC made contributions to the language polishing and modifications in the manuscript. All authors have read and approved the final version of the manuscript.

Competing interests

The authors declare that they have no competing interests.

Consent for publication

Not applicable.

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 publicly and available for non-commercial purpose.

References
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. pekinensis). *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, Bahrk 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):423-429.

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

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

28. 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.

29. 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.

30. 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.
31. Kondrashov FA, Rogozin IB, Wolf YI, Koonin EV: Selection in the evolution of gene duplications. *Genome biology* 2002, 3(2):RESEARCH0008.

32. Conant GC, Wolfe KH: Turning a hobby into a job: how duplicated genes find new functions. *Nature reviews Genetics* 2008, 9(12):938-950.

33. Flagel LE, Wendel JF: Gene duplication and evolutionary novelty in plants. *The New phytologist* 2009, 183(3):557-564.

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

35. 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.

36. Li W, Dai L, Wang GL: PUB13, a U-box/ARM E3 ligase, regulates plant defense, cell death, and flowering time. *Plant signaling & behavior* 2012, 7(8):898-900.

37. Deb S, Sankaranarayanan S, Wewala G, Widdup E, Samuel MA: The S-Domain Receptor Kinase Arabidopsis Receptor Kinase2 and the U Box/Armadillo Repeat-Containing E3 Ubiquitin Ligase9 Module Mediates Lateral Root Development under Phosphate Starvation in Arabidopsis. *Plant physiology* 2014, 165(4):1647-1656.

38. Wang L, Yang Z, Zhang B, Yu D, Liu J, Gong Q, Qanmber G, Li Y, Lu L, Lin Y et al: Genome-wide characterization and phylogenetic analysis of GSK gene family in three species of cotton: evidence for a role of some GSKs in fiber development and responses to stress. *BMC plant biology* 2018, 18(1):330.

39. Mu M, Lu XK, Wang JJ, Wang DL, Yin ZJ, Wang S, Fan WL, Ye WW: Genome-wide Identification and analysis of the stress-resistance function of the TPS (Trehalose-6-Phosphate Synthase) gene family in cotton. *Bmc Genet* 2016, 17:54.

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.

42. Yan J, Wang J, Li Q, 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.

43. 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.

44. 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.
45. 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.

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

47. 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.

48. 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.

49. 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.

50. 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.

Table

Table 1 Domain organizations of PUB proteins in cotton

| No. | Composition of domain (from N end to C end) | 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 + PPlase                           | 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 Legends

Additional file 1: Table S1. Primers used in the manuscript

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

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

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

Additional file 5: Table S5. Essential information of PUB gene members in *Gossypium barbadense*

Additional file 6: Figure S1. The phylogenetic relationship and gene structure analysis of GrPUBs in *G. raimondii*

Additional file 7: Figure S2. The phylogenetic relationship and gene structure analysis of GaPUBs in *G. arboreum*

Additional file 8: Figure S3. The phylogenetic relationship and gene structure analysis of GhPUBs in *G. hirsutum*

Additional file 9: Figure S4. The phylogenetic relationship and gene structure analysis of GbPUBs in *G. barbadense*

Additional file 10: Figure S5. The phylogenetic relationship analysis of PUBs in *Gossypium*

Additional file 11: Figure S6. Distribution of GbPUBs on chromosomes in *G. barbadense*

Additional file 12: Figure S7. The homologous relationships of PUBs in *G. raimondii* and *G. arboreum*

Additional file 13: Figure S8. The intra- and inter-genomic synteny blocks of PUBs

Additional file 14: Figure S9. Predicted expression pattern of GhPUBs in upland cotton under salt and drought stress

Additional file 15: 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. arboreum and G. raimondii, respectively.
Figure 2

Locations of PUB genes on chromosomes 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 analysis 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).

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

- Additionalfile3TableS3.xlsx
- Additionalfile15FigureS10.jpg
- Additionalfile10FigureS5.jpg
- Additionalfile13FigureS8.tif
- Additionalfile7FigureS2.jpg
- Additionalfile9FigureS4.jpg
- Additionalfile6FigureS1.jpg
- Additionalfile14FigureS9.jpg
- Additionalfile5TableS5.xlsx
- Additionalfile1TableS1.xlsx
- Additionalfile2TableS2.xlsx
- Additionalfile12FigureS7.tif
- Additionalfile8FigureS3.jpg
- Additionalfile4TableS4.xlsx
- Additionalfile11FigureS6.tif