Genome-Wide Identification, Characterization and Evolution of BBX Gene Family in Cotton

Jun-Shan Gao (gaojsh@ahau.edu.cn)
Anhui Agricultural University
Pei-pei Wang
Anhui Agricultural University
Na Sun
Anhui Agricultural University
Jessica-Maguy
Anhui Agricultural University
MIENANDI NKODIA
Anhui Agricultural University
Yong Wu
Anhui Agricultural University
Ning Guo
Anhui Agricultural University
Da-Hui Li
Anhui Agricultural University

Research article

Keywords: B-BOX, cotton, GhBBX, systematic analysis, proanthocyanidin

Posted Date: December 11th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-120586/v1

Abstract

Background: The B-BOX (BBX) proteins have important functions in the regulation of photomorphogenesis. The BBX gene family has been identified in several plants, such as rice, Arabidopsis and tomato. However, there still lack a genome-wide survey of *BBX* genes in cotton.

Results: In our present study, 63 *GhBBX* genes were identified in cotton. The analyses of phylogenetic evolution and gene structure showed that the *GhBBX* genes were divided into five subfamilies, and contained B-box conserved domains. qRT-PCR analysis releaved that both *GhBBX27* and *GhBBX33* had potential roles in proanthocyanidin synthesis of brown cotton fibers.

Conclusions: This study provides a genome-wide survey of the *BBX* gene family in cotton and highlights its role in proanthocyanidin synthesis. This result will help us to further understand the complexity of the *BBX* gene family and the functional characteristics of its members.

Background

The zinc finger protein is a member of the transcription factor family and has a large proportion in plants, which has been reported in rice, soybean and Arabidopsis[1]. The growth and development of the plant and the non-abiotic stress reaction of the plant are related. The zinc finger proteins are a class of transcription factors that contain finger-like domains. The zinc finger structure is composed of cysteine, cysteine and histidine, and is formed by combining with Zn^{2+}. The zinc finger protein plays an important role in the life process of gene expression regulation, cell differentiation, embryo development through the interaction with the nucleic acid[2]. Subgroups of zinc-finger proteins, which contain one or two B-box motifs predicted to be involved in protein-protein interaction, are known as B-box (*BBX*) proteins. The *BBX* proteins contain a B-box domain with one or two B-box motifs at the N terminal, and also feature a CCT (CONSTANS, CO-like, and TOC1) domain at the C terminal [3–4]. The BBX protein plays an key role in the regulation of photomorphogenesis, photoperiod regulation of flowering, shading avoidance and abiotic stress response.

In total, 32 proteins containing B-box have been identified in *Arabidopsis thaliana* genome, many of which have been shown to be involved in *COP1* and *HY5* mediated photomorphogenesis of seedlings [5]. *AtBBX4*, *AtBBX20*, *AtBBX22*, *AtBBX24* and *AtBBX25* interact with *COP1* and are degraded by COP1 in the dark [6–7]. In addition, *AtBBX22*, *AtBBX24* and *AtBBX25* interact with *HY5* and acts as a co-activating factors to inhibit of transcriptional activity of *HY5*[8]. *AtBBX21/SALT TOLERANCE HOMOLOG 2* (*STH2*) and *HY5* could positively regulate anthocyanin accumulation and inhibit Hypocotyl extension in seedlings[9]. The point mutation of *BBX21* in the second B-box domain reduces the binding activity to *HY5* and reduces the transcriptional level of *CHI*[10]. *AtBBX25* indirectly inhibits the expression of *AtBBX22* through physical interaction with *HY5* [11–12]. Likewise, *AtBBX32* interact with *AtBBX21* and indirectly reduces *HY5* transcriptional activity. These lines of evidence suggest that the B-box plays an important role in regulatory networks controlling growth and developmental processes.

Previous reports showed that 23, 39, 1, 3, 15, 24 BBX proteins had been identified in *Medicago sativa*, *Populus trichocarpa*, *Chlamydomonas reinhardtii*, *Cladophora aegagropila*, *Selaginella tamariscina*, *Bacillus subtilis*, respectively[13], and finally 30 BBX proteins were contained in *Oryza sativa*[14]. Although the BBX family had been identified in Arabidopsis, corn and pear[15], but the BBX family in cotton had not yet been identified. This highlights the main aim of this article, where 63 *BBX* genes are screened from the cotton genome data according to B-box motif, and *GhBBX27*, *GhBBX33* from brown cotton fiber are identified to play a role in regulating the biosynthesis of proanthocyanidin.

Results

World-wide genome search of BBX genes in cotton

According to the report that 32 BBX proteins were found in Arabidopsis, 63 *BBX* genes in the gene bank of cotton were searched based on the BBX protein in Arabidopsis by using DNA tools software. These *BBX* genes were named as *GhBBX1-GhBBX63*, respectively. A detailed description of the name, chromosome location, protein structure of each gene was given in Table 1. Among the 63 *BBX* genes, the most amino acid is 423 (*GhBBX11*) and the least amino acid is 92 (*GhBBX54*). The molecular weight ranges from 10 KDa (*GhBBX54*) to 47.4 KDa (*GhBBX12*). The isoelectric point is about 4.35 (*GhBBX46*) to 8. 78 (*GhBBX63*).

Table 1
Characteristics of *BBX* family genes in cotton

Gene	Gene identifier	AA	MW	pIs	Domains	Structure
GhBBX1	Gh_A05G0516.1	335	37024.2	6.21	2BBX + CCT	I
GhBBX2	Gh_A08G0192.1	314	34138.14	5.99	2BBX + CCT	I
GhBBX3	Gh_A08G1015.1	368	41560.56	5.61	2BBX + CCT	II
GhBBX4	Gh_A12G0062.1	352	39496.85	5.97	2BBX + CCT	II
GhBBX5	Gh_A13G1580.1	338	37460.85	5.67	2BBX + CCT	I
GhBBX6	Gh_D05G0635.1	335	36956.12	6.07	2BBX + CCT	I
GhBBX7	Gh_D08G0269.1	313	34241.3	5.99	2BBX + CCT	I
GhBBX8	Gh_D08G1289.1	367	41468.33	5.48	2BBX + CCT	II
GhBBX9	Gh_D12G0077.1	352	39506.81	5.97	2BBX + CCT	II
GhBBX10	Gh_D13G1939.1	338	37389.78	5.74	2BBX + CCT	I
GhBBX11	Gh_A01G1562.1	423	46907.92	7.93	2BBX + CCT	III
GhBBX12	Gh_A05G2921.1	421	47445.24	5.54	2BBX + CCT	II
GhBBX13	Gh_A07G1753.1	408	44636.26	4.94	2BBX + CCT	II
GhBBX14	Gh_A08G0775.1	369	40961.92	6.15	2BBX + CCT	I
GhBBX15	Gh_A08G0848.1	418	47087.46	5.3	1BBX + CCT	III
GhBBX16	Gh_A09G0466.1	410	45922.91	5.06	2BBX + CCT	III
GhBBX17	Gh_D01G1811.1	368	41560.56	5.61	1BBX + CCT	III
GhBBX18	Gh_D07G1957.1	408	44509.24	4.99	1BBX + CCT	II
GhBBX19	Gh_D08G0923.1	368	40676.51	6.01	1BBX + CCT	I
GhBBX20	Gh_D08G1030.1	417	47026.37	5.16	1BBX + CCT	III
GhBBX21	Gh_D09G0473.1	413	46420.44	5.02	1BBX + CCT	III
GhBBX22	Gh_D12G0543.1	374	41500.37	5.7	2BBX + CCT	I
GhBBX23	Gh_D13G2210.1	297	32884.45	7.22	2BBX + CCT	I
GhBBX24	Gh_Sca004822G01.1	295	32647.04	6.93	2BBX + CCT	I
GhBBX25	Gh_A04G1057.1	196	21806.67	5.19	2BBX	IV
GhBBX26	Gh_A05G1616.1	309	33698.76	6.36	2BBX	IV
GhBBX27	Gh_A06G1830.1	250	27714.07	5.77	2BBX	IV

Gene	Gene identifier	AA	MW	pIs	Domains	Structure
GhBBX28	Gh_A11G1640.1	199	21962.21	7.07	2BBX	IV
GhBBX29	Gh_A12G0561.1	187	20702.58	7.16	2BBX	IV
GhBBX30	Gh_A13G0274.1	212	23251.11	5.53	2BBX	IV
GhBBX31	Gh_A13G0284.1	201	22317.84	5.78	2BBX	IV
GhBBX32	Gh_D01G0033.1	233	25592.72	5.22	2BBX	IV
GhBBX33	Gh_D02G0969.1	287	31291.09	5.9	2BBX	IV
GhBBX34	Gh_D05G1800.1	309	33470.55	6.54	2BBX	IV
GhBBX35	Gh_D06G0023.1	250	27778.21	6.19	2BBX	IV
GhBBX36	Gh_D09G2008.1	238	26206.78	4.99	2BBX	IV
GhBBX37	Gh_D11G1798.1	198	21745.92	6.73	2BBX	IV
GhBBX38	Gh_D13G0293.1	212	23278.14	5.9	2BBX	IV
GhBBX39	Gh_D13G0303.1	203	22448.04	6.15	2BBX	IV
GhBBX40	Gh_A01G0033.1	302	32803.14	8.59	2BBX	IV
GhBBX41	Gh_A01G0038.1	206	22912.75	5.24	2BBX	IV
GhBBX42	Gh_A01G0320.1	296	32145.79	5.44	2BBX	IV
GhBBX43	Gh_A02G1620.1	170	18821.45	7.08	2BBX	IV
GhBBX44	Gh_A04G0209.1	269	29282.7	4.66	2BBX	V
GhBBX45	Gh_A04G0927.1	129	14688.92	8.69	1BBX	V
GhBBX46	Gh_A06G1148.1	191	21263.8	4.35	1BBX	V
GhBBX47	Gh_A08G2069.1	122	13391.57	6.04	1BBX	IV
GhBBX48	Gh_A09G2314.1	238	26206.78	4.99	1BBX	IV
GhBBX49	Gh_A10G0877.1	172	18949.34	5.88	2BBX	IV
GhBBX50	Gh_A11G2610.1	283	30836.22	4.75	2BBX	IV
GhBBX51	Gh_D01G0038.1	204	22688.53	5.04	2BBX	IV
GhBBX52	Gh_D01G0357.1	296	32168.69	5.3	2BBX	IV
GhBBX53	Gh_D02G1898.1	120	13387.37	8.02	2BBX	V
GhBBX54	Gh_D03G0102.1	92	10087.6	6.93	1BBX	IV
GhBBX55	Gh_D04G1444.1	129	14674.89	8.69	2BBX	V

Gene	Gene identifier	AA	MW	pIs	Domains	Structure
GhBBX56	Gh_D04G1636.1	197	22012.73	5.41	1BBX	IV
GhBBX57	Gh_D05G3492.1	266	28713.16	4.84	2BBX	V
GhBBX58	Gh_D06G1324.1	189	20908.56	4.43	1BBX	V
GhBBX59	Gh_D08G2660.1	123	13528.72	6.17	1BBX	IV
GhBBX60	Gh_D10G0876.1	164	17696.78	5.78	2BBX	IV
GhBBX61	Gh_D11G2971.1	283	30681.17	4.8	2BBX	IV
GhBBX62	Gh_D12G0577.1	223	24956.46	8.6	2BBX	IV
GhBBX63	Gh_D12G1358.1	130	14408.55	8.78	2BBX	V

Structural analysis of BBX proteins

The sequence of the first B-box (B1) motif contained in these genes were C-X_2-C-X_4-A-X_3-C-X_2-D-X_4-C-X_2-C-D (Fig. 1A), and the sequence of the second B-box (B2) motif were C-X7-C-X2-C-D-X3-H (Fig. 1B). The distance between the two B-box domains was 5–20 amino acids. It has been found that the conserved amino acid residues in the B-box motifs are very important in regulating protein–protein interaction and transcriptional regulation. According to the reported article, the second B-box motif of *AtBBX21* in Arabidopsis was key to bind to *HY5* promoter and promote the transcription of *HY5*. The sequence of CCT domain is R-X_5-R-Y-X-E-K-X_3-R-X_3-K-X_2- R-Y-X_2-R-K-X_2-A-X_2-R-X-R-X-K-G-R-F-X-K (Fig. 1C).

In order to understand more clearly the relationship between these genes, a phylogenic tree was constructed according to *BBX* genes obtained from *Pyrus bretschneideri*, *O. sativa*, *A. thaliana*, *P. trichocarpa*, *Z. mays* and *Gossypium hirsutum*. These BBX proteins were divided to five structural groups (Fig. 2). The structural group I contains two B-box domains and one CCT domain. The structural group II contains also two B-box domains and one CCT domain, but there are some differences in the second B-box domain between the group I and group II. The structural group III consists of a B-box domain and a CCT domain. The structural group IV contains two B-box domains without CCT structure. The structural group V contains only one B-box domain. In cotton, there are only 24 *GhBBX* genes with CCT structure, accounting for 38% of all *GhBBX* genes. Among them, 11 *GhBBX* genes belong to the group I, and 7 *GhBBX* genes are in the group II and 6 *GhBBX* genes are in the group III. The remaining are shared into the groups IV and V containing 31 and 8 *GhBBX* genes, respectively (Fig. 2). Nonetheless, the *GhBBX16* is found to fit in the structural group II, which only contains a B-box (Fig. 2). These results indicate that some of the GhBBX proteins can lose a domain in recent evolutionary events, but retain other common features of their structural groups. Phylogenetic analysis shows that the GhBBX proteins belonging to the same structural group are classified by amino acid similarity and the structural organization of B-box and CCT domains.

Structural analysis of BBX genes in cotton

The gene structures of these *BBX* genes were analyzed in upland cotton. The results indicated that only four *BBX* genes had no introns, all of which belong to the structural group IV (Fig. 3). The remaining *BBX* genes contained 1 to 9 introns, but *GhBBX40* had the most introns (Fig. 3). *GhBBX4*, *GhBBX9*, *GhBBX12*, *GhBBX13*, *GhBBX18* contained four introns and the same numbers of exons. In addition, there were 26 *BBX* genes with two introns and one exon, and there were 11 *BBX* genes with three exons and the same numbers of introns, which belong to the structural group IV and V without CCT (Fig. 3).

Expression pattern of partial BBX family genes in cotton

In order to understand the gene expression pattern of the *BBX* family genes, qRT-PCR were carried out. The expression levels of these *BBX* genes from the structural group IV and structural group V were tested at 6 days post anthesis (DPA), 12 DPA, 18 DPA and 24 DPA in brown cotton fibers (Fig. 4). The experimental results showed that small numbers of genes including *GhBBX25*, *GhBBX59*, *GhBBX47* and *GhBBX34* were expressed at low levels during these periods (Fig. 4). On the other hand, *GhBBX58, GhBBX63, GhBBX49, GhBBX51, GhBBX41, GhBBX60* were expressed at lower levels at 6 DPA, 12 DPA and 18 DPA. *GhBBX41* and *GhBBX54* had low expression at 24 DPA, but there were high expression levels at other periods (Fig. 4). On the contrary, *GhBBX25* had high expression at 18 DPA, but the expression of *GhBBX25* was lower at other periods. *GhBBX37* had the same expression level at these four periods. The expression of *GhBBX55* was higher at 18 DPA and 24 DPA than that at other two periods (Fig. 4).

DMACA staining of transgenic Arabidopsis seeds

In Arabidopsis, *AtBBX21* belongs to the structural group IV of the *AtBBX* gene family. The previous study desmonstrated that the mutation of the second B-box of *AtBBX21* reduced the expression of *CHS*, *CHI*, *F3H*, *F3'H*, *DFR* and *LDOX*, while the mutantion of the first B-box were not significantly different from the overexpressed plants in Arabidopsis. In this study, the eukaryotic expression vectors of *GhBBX27* and *GhBBX33*, which belong to the structural group IV similar to *AtBBX21*, were constructed and transformed into Arabidopsis. These seeds of transgenic plants were dyed in DMACA reagent to observe the color of the seed coat. The results displayed that the seed coat colors of transgenic Arabidopsis with *GhBBX27* and *GhBBX33* were dark, and the seed color of wild type Arabidopsis was a light color (Fig. 5). The color of the Arabidopsis seed coat was caused by the accumulation of proanthocyanidins in the seed coat. The transgenic Arabidopsis seed coats of *GhBBX27* and *GhBBX33* had a darker color than the wild type. This suggests that *GhBBX27* and *GhBBX33* may have an effect on the accumulation of proanthocyanidins.

Morphological observation of transgenic Arabidopsis seedlings

The seeds of transgenic and wild-type Arabidopsis were selected, sterilized and sown on MS medium. When the true leaves grew, the morphology of Arabidopsis seedlings was observed. The length of hypocotyls of the two transgenic Arabidopsis seedlings of *GhBBX27* and *GhBBX33* were shorter than that of the wild-type Arabidopsis (Fig. 6). While the photomorphogenesis of seedlings is established, *HY5* can

inhibit the hypocotyl elongation. It is speculated that *GhBBX27* and *GhBBX33* can inhibit the elongation of seedling hypocotyls by promoting the expression of *HY5*.

Discussion

In plants, *BBX* proteins regulate plant growth and development. The AtBBX protein contains two B-box and one CCT domains, and plays a key role in flowering by regulating the photoperiod. *AtBBX1* is the core factor in controlling the flowering of Arabidopsis. *CONSTANS* (*CO*) promotes flowering under long-day (LD) condition, but there is no effect on the flowering time under short-day (SD) condition [16]. *CO* mutants flower late only under LD condition, whereas *CO* over-expressing plants flower early in both LD and SD conditions. The expression of *BBX2* and *BBX3 i*s different from *CO*. The change of the former has less regulation on flowering time, but the overexpression of *BBX2* can shorten the cycle of two different circadian rhythms [17]. *BBX4*(*COL3*) is not only a positive regulator of photomorphogenesis but also promotes root growth [7]. *BBX6* positively regulates the expression of *FLOW-ERING LOCUST* (*FT*) to achieve the role of a specific pollen inducer of SD [18]. *BBX7* (*COL9*) plays the role of LD-specific flowering repressor by reducing the expression of *CO* and *FT* [19]. A number of *AtBBXs* that only have two B-box domains have also been identified. The Arabidopsis mutants showed that *AtBBX18*, *21* and *22* act as positive regulators in the seedling de-etiolation processes. *AtBBX19*, *24* and *25* have the opposite effects [8]. *AtBBX32*, which has only a single B-box, is a modulator of light signaling too [20]. AtBBX proteins without CCT domain regulate plant growth and development through other factors. Subsequent studies have revealed that a series of proteins containing B-box, including *AtBBX21* and *AtBBX22*, are also direct targets of *COP1*. BBX-containing proteins are also known to share many roles with *HY5* in response to light. Experiments show that *AtBBX21* has a subtle relationship with the shade avoidance response [5]. It is also acknowledged that *AtBBX22* can interact with *HY5*, and that *AtBBX21* activate *HY5* expression through its binding to the promoter of *HY5* [9].

AtBBX1, which is the first *BBX* gene identified in Arabidopsis, encodes a protein controlling photoperiod flowering time. The overexpression of *AtBBX1* induces early flowering, and the mutation of *AtBBX1* delays flowering. In rice, *OsCO3/OsBBX2* can inhibit flowering under SD conditions, and plays a role in regulating the flowering period [21]. *Hd1/OsBBX1* can not only promote flowering under SD but also inhibit flowering under LD [22]. In addition to rice, other BBX proteins containing B-box and CCT domains in green plants have also been studied. The function of *TaHd1-1* in wheat is similar to that of the rice *Hd1/OsBBX1*[23]. *StCO* is isolated from potato and participates in the formation of photoperiod tubers [24]. In sugar beet, *BvCOL1* is heterologous overexpressed in Arabidopsis *co-2* mutant, and rescues the wild-type flowering phenotype, which indicates that it functions identically as *AtBBX1*[25]. Not the same with *BBX* members with B-box and CCT domain, other BBX proteins have multiple functions, which are involved in the regulation of photomorphogenesis and abiotic stress responses. For instance, *AtBBX24* participates in UV-B signaling pathway, photomorphogenesis seedling de-etiolation and saline stress responses [26]. On the other hand, *AtBBX21/ STH2/ LHUS* and *AtBBX22/ STH3* are implicated in seedling

photomorphogenesis and shade avoidance responses [5, 27], while *AtBBX18* negatively regulates heat tolerance in Arabidopsis [28]. Heterologous expression of some *BBX* genes indicates that the transcription activity is affected by amino acid substitutions in the B1 or B2 domains. It is hypothesized that some *BBX* proteins containing the B-box conserved domains have played a central role in the regulation of plant development during evolution.

Although the *BBX* family has been identified in pear, corn and Arabidopsis, the *BBX* family in cotton has not yet been identified. In this article, 63 *GhBBX* genes were screened from cotton gene bank according to B-box domain, and were classified into five subfamilies whether contain CCT domain or not. Among these genes, *GhBBX27* and *GhBBX33* similar to *AtBBX21* are selected to study their functions in regulating the biosynthesis of proanthocyanidin. The DMACA staining of transgenic Arabidopsis seed coat was darker than that of wild type, indicating that the proanthocyanidin content increased. Therefore, it is speculated that *GhBBX27* and *GhBBX33* can positively regulate the synthesis of proanthocyanidin. The determination of the hypocotyls of transgenic Arabidopsis seedlings suggests that the regulation of proanthocyanidin synthesis may be related to the promoter of *HY5*. According to the analysis of the promoter of *HY5* in cotton (Table 2), it is found that *HY5* in cotton contains the same sequence as *HY5* in Arabidopsis. The sequence is the T/G-box where *AtBBX21* interacts with *HY5*. When the T/G-box of *HY5* was point-mutated, it was found that *AtBBX21* could not interact with the promoter of *HY5* [10]. Therefore, it is speculated that *GhBBX27* and *GhBBX33* are combined with G-box of *HY5* to promote the expression of *CHS*, *CHI*, *F3H*, *F3'H*, *DFR* and *LDOX*, thus further promote the synthesis of proanthocyanidin.

Table 2
Analysis of *HY5* promoter elements

Component name	5'-3'	Position	Quantity	Features
ABRE	ACGTG	10203-, 1668+, 1031+, 1810-, 1022+, 1753-, 1212+, 93+, 1030-	9	Participate in abscisic acid reaction
TATA-box	TATATTTATATTT	1058-, 1227+, 1093-, 1473-, 1091-, 1394-, 1116-, 1560-, 1060-, 1298-, 1112-, 1539+, 1059-, 1434-, 1147-, 1642-, 1061-…	97	Transcription initiation-30 core promoter element
TGA-element	AACGAC	1456-, 1781+	2	Auxin response element
P-box	CCTTTTG	950+	1	Gibberellin response element
Box 4	ATTAAT	351+, 1911-, 891+, 445+, 1128-	5	Partially conserved DNA modules involved in light response
ARE	AAACCA	1391-	1	Original cis-acting components required for anaerobic reactions
CAAT-box	CCAAT	111-, 126-, 401-, 415-, 521+, 524-, 760-, 835-, 853+, 1043+, 1070-…	30	Original cis-acting originals in the promoter and enhancer regions
G-Box	CACGTT	1020+, 1810+, 1667-, 1030+, 1753+, 92+, 1211+, 1030-, 1753-	5	Cis-acting originals participating in the light response
G-box	CACGTC	92+, 1211+, 030-, 1753-	4	Cis-acting originals participating in the light response

Component name	5'-3'	Position	Quantity	Features
TC-rich repeats	ATTCTCTAAC	63+, 1579-, 1817+	3	Defense and coercion response elements
TCT-motif	TCTTAC	590+, 1003+	2	Part of light responsive element

Conclusions

In this present study, a systematic analysis of the *GhBBX* gene family was performed, including conserved domain, gene structure, phylogenetic relationship, gene duplication and expression pattern analysis. The *GhBBX* genes were divided into five structural groups: I (11 genes), II (7 genes), III (6 genes), IV (31 genes), V (8 genes), which were supported by the analyses of gene structural and conserved domain. The motifs of the two B-box conserved domains are C-X2-C-X4-A-X3-C-X2-D-X4-C-X2-CD and C-X7-C-X2-CD-X3-H. qRT-PCR analysis revealed that the *GhBBX* genes play an important role at different development stages of brown cotton fibers. *GhBBX27* and *GhBBX33* were successfully transformed into wild-type Arabidopsis using Agrobacterium-mediated method, The color of transgenic Arabidopsis seed coat was darker than wild type by DMACA staining, and the hypocotyls of transgenic Arabidopsis seedlings are shorter than wild type.

It is speculated that *GhBBX27* and *GhBBX33* may play roles in promoting the synthesis of proanthocyanidins in brown cotton fiber.

Methods

Plant material

Zongcaixuan 1 (P26), which is a brown cotton line breeded by our laboratory members, were planted in the High-tech Agricultural Park of Anhui Agricultural University. Cotton bolls were collected at 6 days post anthesis (DPA), 12 DPA, 18 DPA and 24 DPA, respectively, and were taken to the laboratory with liquid nitrogen freezing for storing in ultra-low temperature refrigerator.

Sequence retrieval

To identify and annotate *BBX* genes in cotton, the Arabidopsis BBX protein sequences from the Arabidopsis Information Resource (TAIR) database (http://www.arabidopsis.org) were used as queries to search against cotton genome database with BLASTP program (e-value < 1e-5). Then, the putative BBX genes were verified for the presence of the B-BOX domain by screening against the InterProScan (http://www.ebi.ac.uk/Tools/pfa/iprscan/), Pfam (http://pfam.sanger.ac.uk/) and SMART (http://smart.embl-heidelberg.de/) database.

Phylogenetic analysis

Multiple sequence alignments of the 63 cotton BBX proteins were generated using ClustalW version 1.83 with default settings, and the neighbor-joining (NJ) tree was constructed by MEGA 5.2 with bootstrap analysis (1000 replicates). The pfam (http://pfam.xfam.org), InterProscan (http://www.ebi.ac.uk/interpro/scan.html), and SMART (http://smart.embl-heidelberg.de) were used to identify domains. The sequence logos of conserved domains were generated using online Web Logo (http://weblogo.berkeley.edu/logo.cgi).

qRT-PCR

The TIANGEN RNAprep pure (Tiangen, Beijing, China) was used to extract the total RNA according to the manufacturer's instructions, followed by DNaseI (Tiangen, Beijing, China) digestion to eliminate any contaminating DNA. For qRT-PCR analysis, the first-strand cDNA was synthesized from the 1 μg RNA using the Reverse Transcriptase M-MLV System (Tiangen, Beijing, China) according to the manufacturer's instructions. The Beacon Designer 7 software was used to design and check the gene-specific primers. The cotton tubulin gene was used as a reference gene. The qRT-PCR was carried out using SYBR® Premixm Ex Taq™ (TaKaRa, Japan) with the CFX96 Touch™ Real-Time PCR Detection System (Bio-Rad, USA). For each sample, three biological replicates were executed. The $2^{-\Delta\Delta CT}$ method was used to estimate the relative expression level.

Arabidopsis transformation and DMACA staining

The recombinant plasmid pCambia1301a-GhBBX27 and pCambia1301a-GhBBX33 were transformed into *Agrobacterium tumefaciens* EHA105 by the floral dip method, respectively. The transformed plants were selected on MS medium supplemented with 50 mg L^{-1} hygromycin. The positive plants were transferred to soil in the greenhouse at 25 ℃ under a 16 h light/8 h dark photoperiod, and were further confirmed by PCR. Finally, the seeds of transgenic and wild-type Arabidopsis were dyed by 0.1% DMACA, and observed the color of the seed coats.

Observation of hypocotyl lengths

The seeds of transgenic and wild-type Arabidopsis were selected, sterilized and sown on MS medium. When the true leaves grew, the hypocotyl lengths of Arabidopsis seedlings were observed.

Abbreviations

DPA: days post anthesis

CCT: CONSTANS, CO-like and TOC1

CHS: Chalcone synthase

CHI: Chalcone isomerase

F3H: flavanone 3-hydroxylase

F3'H: flavonoid 3'-hydroxylase

DFR: dihydroflavonol 4-reductase

LDOX: leucoanthocyanidin dioxygenase

RT-PCR: Reverse transcription Polymerase chain reaction

DMACA: 4-(Dimethylamino) cinnamaldehyde

MS: Murashige and Skoog medium

LD: Long day

SD: Short day

CO: CONSTANS

FT: FLOW-ERING LOCUST

COL: CONSTANS-LIKE

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and material

Under reasonable request, all data generated or analyzed in this study can be obtained from the corresponding author. The plant materials from our lab were grew in the High-tech Agricultural Park of Anhui Agricultural University.

Competing interests

The authors declare that they have no competing interests.

Funding

This research was supported by the National Natural Science Foundation of China (No.31672497) and National Key Research and Development Projects of China (No. 2018YFD0100403). The funding bodies

played no role in the design of the experiment and data collection, analysis, or preparation of the manuscript.

Author's contributions

WPP and GJS designed the experiments. WPP, SN and WY performed most of experiments and analyzed the data. MNJ and GN assisted in experiments. WPP and SN wrote the manuscript. GJS and LDH revised the manuscript. All authors read and approved the final manuscript.

Acknowledgements

We thank the Master student Anane. G. Owusu for providing the help of English writting. We thank Prof. De-Xiang Wu (College of Agronomy, Anhui Agricultural University) for critical comments to the manuscript.

Author information

First author: Pei-pei Wang

Corresponding author: Jun-Shan Gao

Affiliations

School of Life Sciences, Anhui Agricultural University, Hefei 230036, China

Pei-pei Wang, Na Sun, Jessica-Maguy MIENANDI NKODIA, Yong Wu, Ning Guo, Da-Hui Li & Jun-Shan Gao

References

1. Li J, Chen X, Gong X, et al. A transcript profiling approach reveals the zinc finger transcription factor ZNF191 is a pleiotropic factor [J]. BMC Genom. 2009;10(1):1–14.
2. Debora B, Marc AJ, Morgan ACN, et al. The PR/SET Domain Zinc Finger Protein Prdm4 Regulates Gene Expression in Embryonic Stem Cells but Plays a Nonessential Role in the Developing Mouse Embryo [J]. Mol Cell Biol. 2013;33(19):3936–50.
3. Michael AM, Brandi NS, Kieran M, et al. Solution Structure of the RBCC/TRIM B-box1 Domain of Human MID1: B-box with a RING [J]. J Mol Biol. 2006;358(2):532–45.
4. Michael AM, Jessica AM, Kieran M, et al. Solution Structure of the MID1 B-box2 CHC(D/C)C 2 H 2 Zinc-binding Domain: Insights into an Evolutionarily Conserved RING Fold [J]. J Mol Biol. 2007;369(1):1–10.
5. Crocco CD, Holm M, Yanovsky MJ, et al. AtBBX21 and COP1 genetically interact in the regulation of shade avoidance [J]. Plant J. 2010;64(4):551–62.
6. Fan XY, Sun Y, Cao DM, et al. BZS1, a B-box protein, promotes photomorphogenesis downstream of both brassinosteroid and light signaling pathways [J]. Molecular plant. 2012;5(3):591–600.

7. Datta S, Hettiarachchi GHCM, Deng XW, et al. Arabidopsis CONSTANS-LIKE3 is a positive regulator of red light signaling and root growth [J]. Plant Cell. 2006;18(1):70–84.
8. Gangappa SN, Crocco CD, Johansson H, et al. The Arabidopsis B-BOX protein BBX25 interacts with HY5, negatively regulating BBX22 expression to suppress seedling photomorphogenesis [J]. Plant Cell. 2013;25(4):1243–57.
9. Xu DQ, Jiang Y, Li J, et al. BBX21, an Arabidopsis B-box protein, directly activates HY5 and is targeted by COP1 for 26S proteasome-mediated degradation [J]. Proc Natl Acad Sci USA. 2016;113(27):7655–60.
10. Xu D, Jiang Y, Li J, et al. The B-Box Domain Protein BBX21 Promotes Photomorphogenesis [J]. Plant physiology. 2018;176(3):2365–75.
11. Datta S, Sourav., Johansson., et al. LZF1/SALT TOLERANCE HOMOLOG3, an Arabidopsis B-Box Protein Involved in Light-Dependent Development and Gene Expression, Undergoes COP1-Mediated Ubiquitination(W) [J]. Plant Cell. 2008;20(9):2324–38.
12. Datta S, Hettiarachchi C, Johansson H, et al. SALT TOLERANCE HOMOLOG2, a B-box protein in Arabidopsis that activates transcription and positively regulates light-mediated development [J]. Plant Cell. 2007;19(10):3242–55.
13. Crocco CD, Botto JF. BBX proteins in green plants:In sights into their evolution, structure, feature and functional diversification [J]. Gene. 2013, 531: 44–52.
14. Huang J, Zhao X, Weng X, et al. The rice B-box zinc finger gene family: genomic identification, characterization, expression profiling and diurnal analysis [J]. PloS one. 2012;7(10):e48242.
15. Cao Y, Han Y, Meng D, et al. B-BOX genes: genome-wide identification, evolution and their contribution to pollen growth in pear (Pyrus bretschneideri Rehd.) [J]. BMC plant biology. 2017;17(1):156.
16. Putterill J, Robson F, Lee K, et al. The CONSTANS gene of Arabidopsis promotes flowering and encodes a protein showing similarities to zinc finger transcription factors [J]. Cell. 1995;80(6):847.
17. Susan L, Carl S, Fern A, et al. Analysis of the function of two circadian-regulated CONSTANS-LIKE genes [J]. Plant J. 2001;26(1):15–22.
18. Hassidim M, Harir Y, Yakir E, et al. Over-expression of CONSTANS-LIKE 5 can induce flowering in short-day grown Arabidopsis [J]. Planta. 2009;230(3):481–91.
19. Cheng XF, Wang ZY. Overexpression of COL9, a CONSTANS-LIKE gene, delays flowering by reducing expression of CO and FT in Arabidopsis thaliana [J]. Plant J. 2005;43(5):758–68.
20. Holtan HE, Bandong S, Marion CM, et al. BBX32, an Arabidopsis B-Box protein, functions in light signaling by suppressing HY5-regulated gene expression and interacting with STH2/BBX21 [J]. Plant physiology. 2011;156(4):2109–23.
21. Kim SK, Yun CH, Lee JH, et al. OsCO3, a CONSTANS-LIKE gene, controls flowering by negatively regulating the expression of FT-like genes under SD conditions in rice[J]. Planta. 2008;228(2):355–65.

22. Yano M, Katayose Y, Ashikari M, et al. Hd1, a major photoperiod sensitivity quantitative trait locus in rice, is closely related to the Arabidopsis flowering time gene CONSTANS [J]. Plant Cell. 2000;12(12):2473–84.
23. Nemoto Y, Kisaka M, Fuse T, et al. Characterization and functional analysis of three wheat genes with homology to the CONSTANS flowering time gene in transgenic rice[J]. Plant J. 2003;36(1):82–93.
24. Nahuel D, González-Schain., Mercedes Díaz-Mendoza., et al. Potato CONSTANS is involved in photoperiodic tuberization in a graft-transmissible manner [J]. Plant Journal, 2012, 70(4): 678–690.
25. Kristensen AK, Brunstedt JE, Nielsen JD, et al. Partial characterization and localization of a novel type of antifungal protein (IWF6) isolated from sugar beet leaves [J]. Plant Sci. 2000;159(1):29–38.
26. Nagaoka S, Takano T. Salt tolerance-related protein STO binds to a Myb transcription factor. homologue and confers. salt tolerance in Arabidopsis [J]. J Exp Bot. 2003;54(391):2231–7.
27. Crocco CD, Holm M, Yanovsky MJ, et al. Function of B-BOX under shade [J]. Plant signaling & behavior, 2011, 6(1): 101–104.
28. Wang Q, Tu X, Zhang J, et al. Heat stress-induced BBX18 negatively regulates the thermotolerance in Arabidopsis [J]. Molecular biology reports. 2013;40(3):2679–88.

Figures

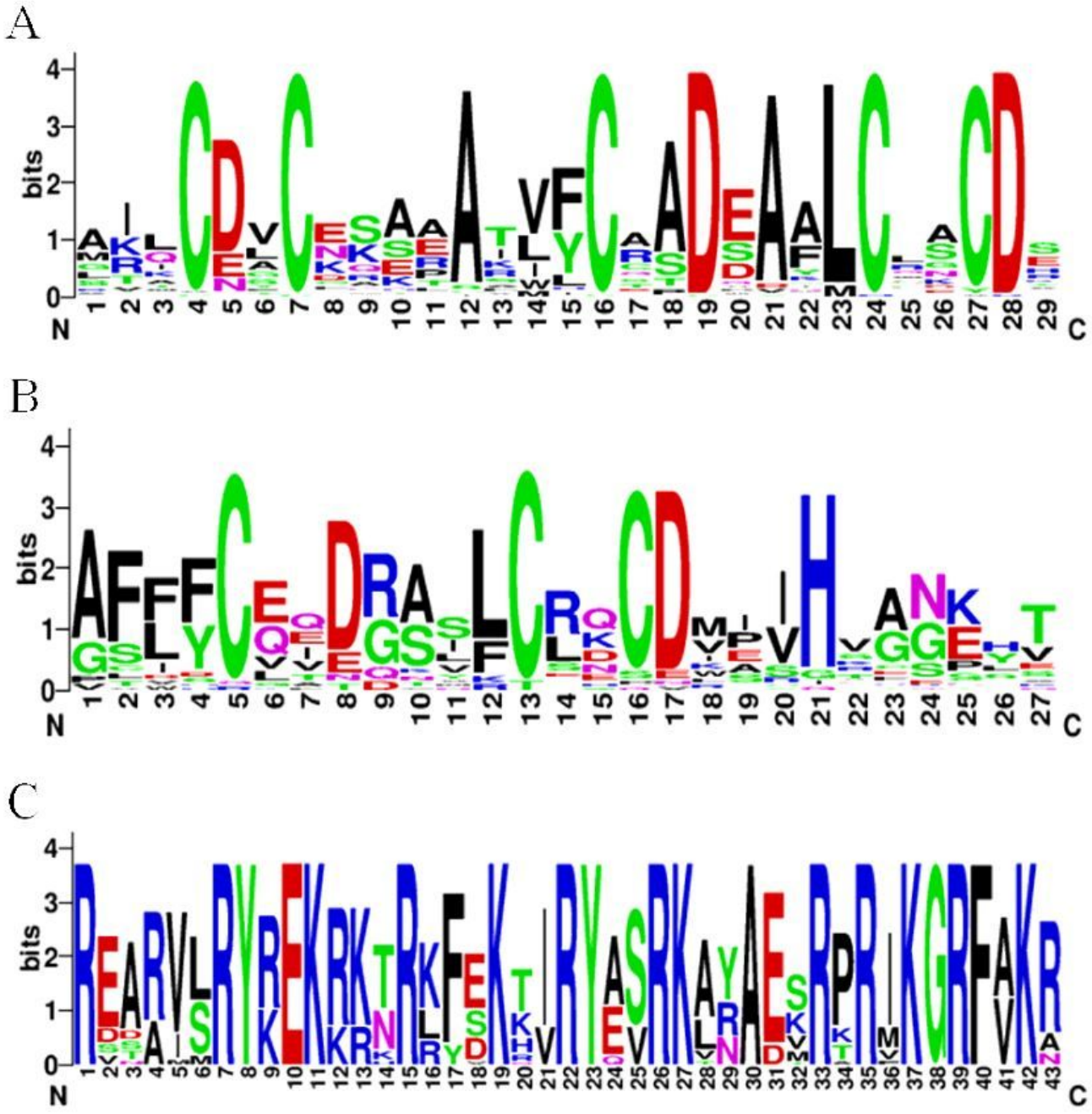

Figure 1

Domain composition of GhBBX proteins. A, B and C represent the protein sequence of the B-BOX 1, B-BOX 2 and CCT domains, respectively

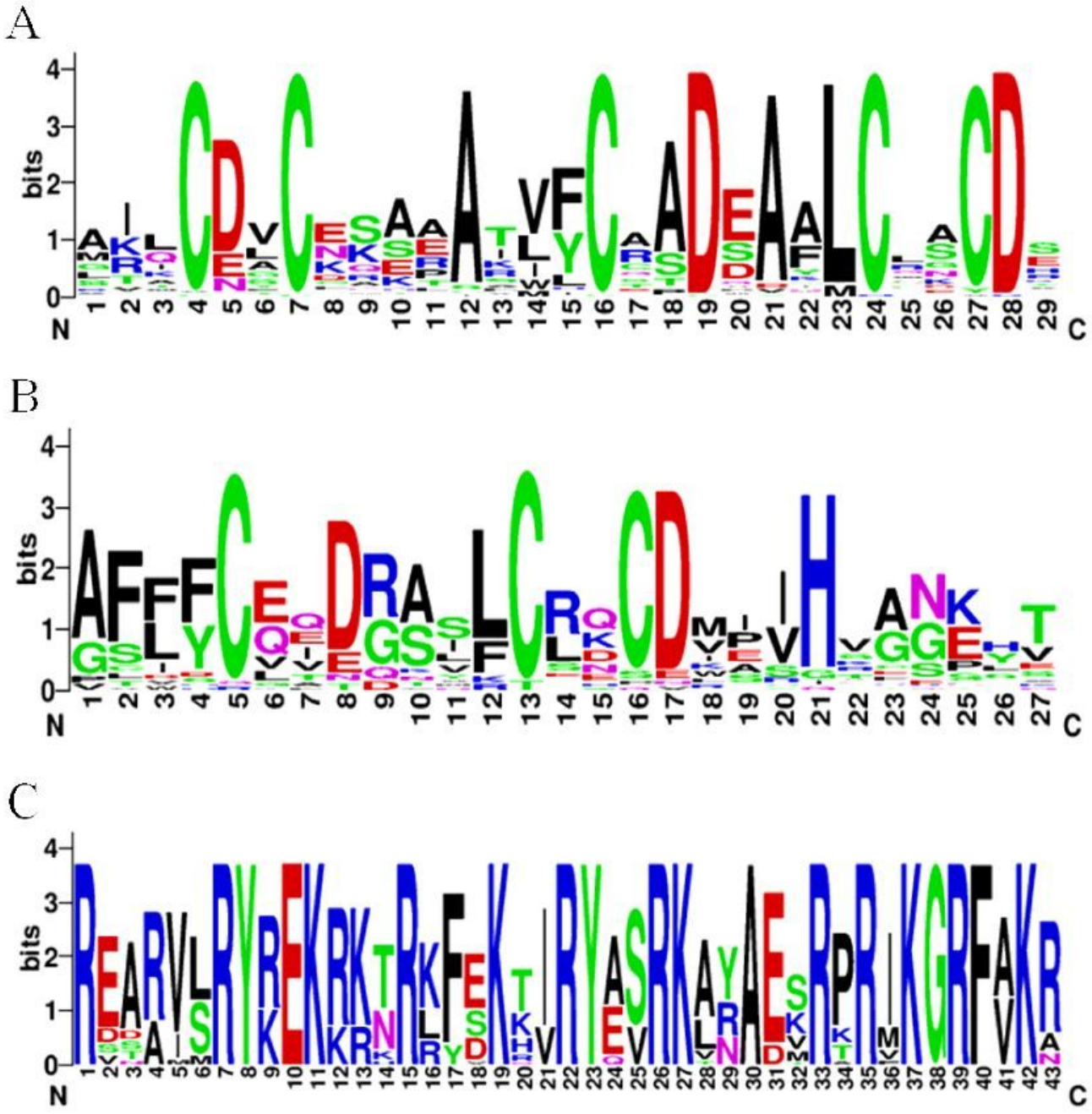

Figure 1

Domain composition of GhBBX proteins. A, B and C represent the protein sequence of the B-BOX 1, B-BOX 2 and CCT domains, respectively

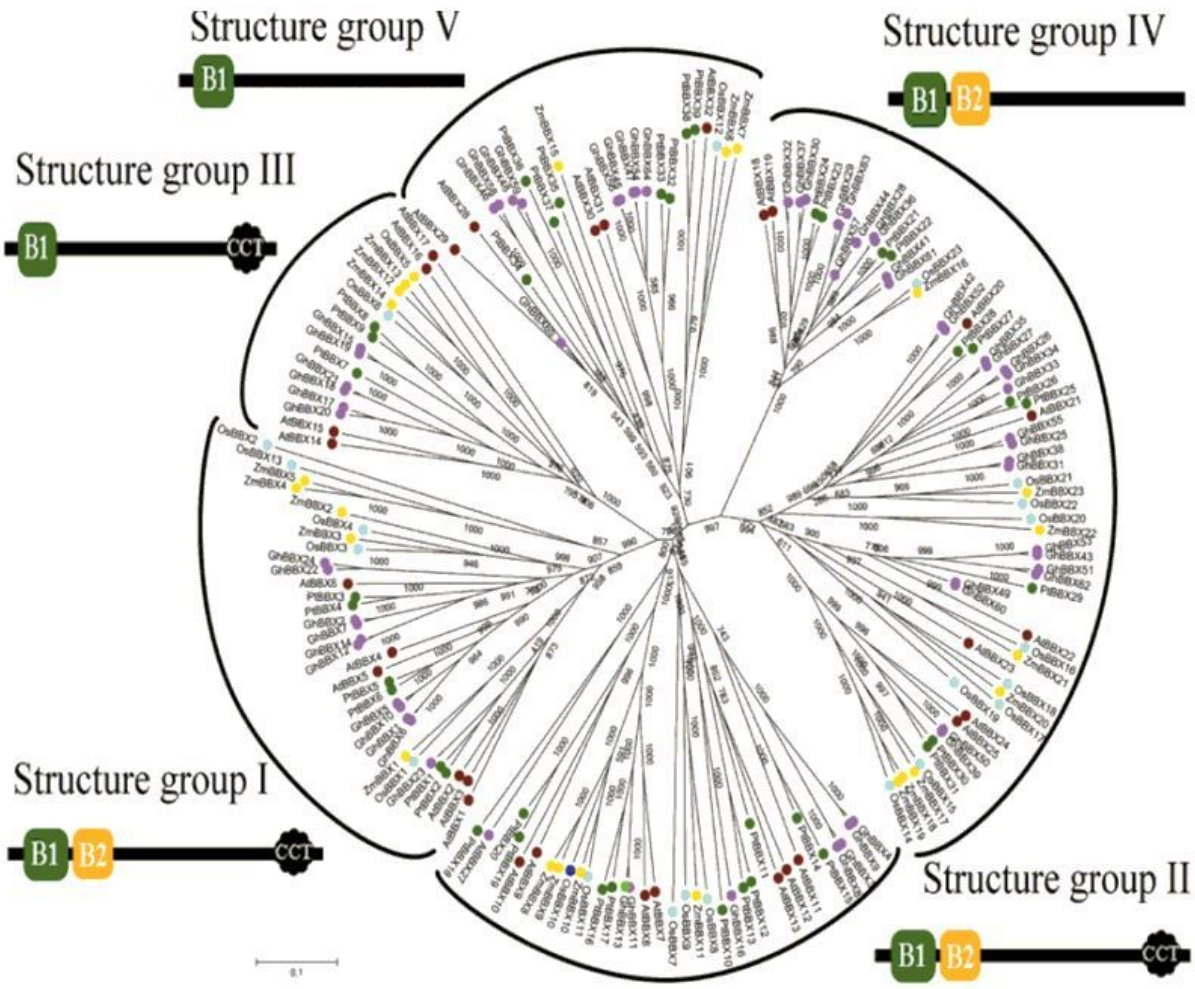

Figure 2

Phylogenetic analysis of BBX genes in P. bretschneideri, O. sativa, A. thaliana, P. trichocarpa, Z. mays and G. hirsutum.

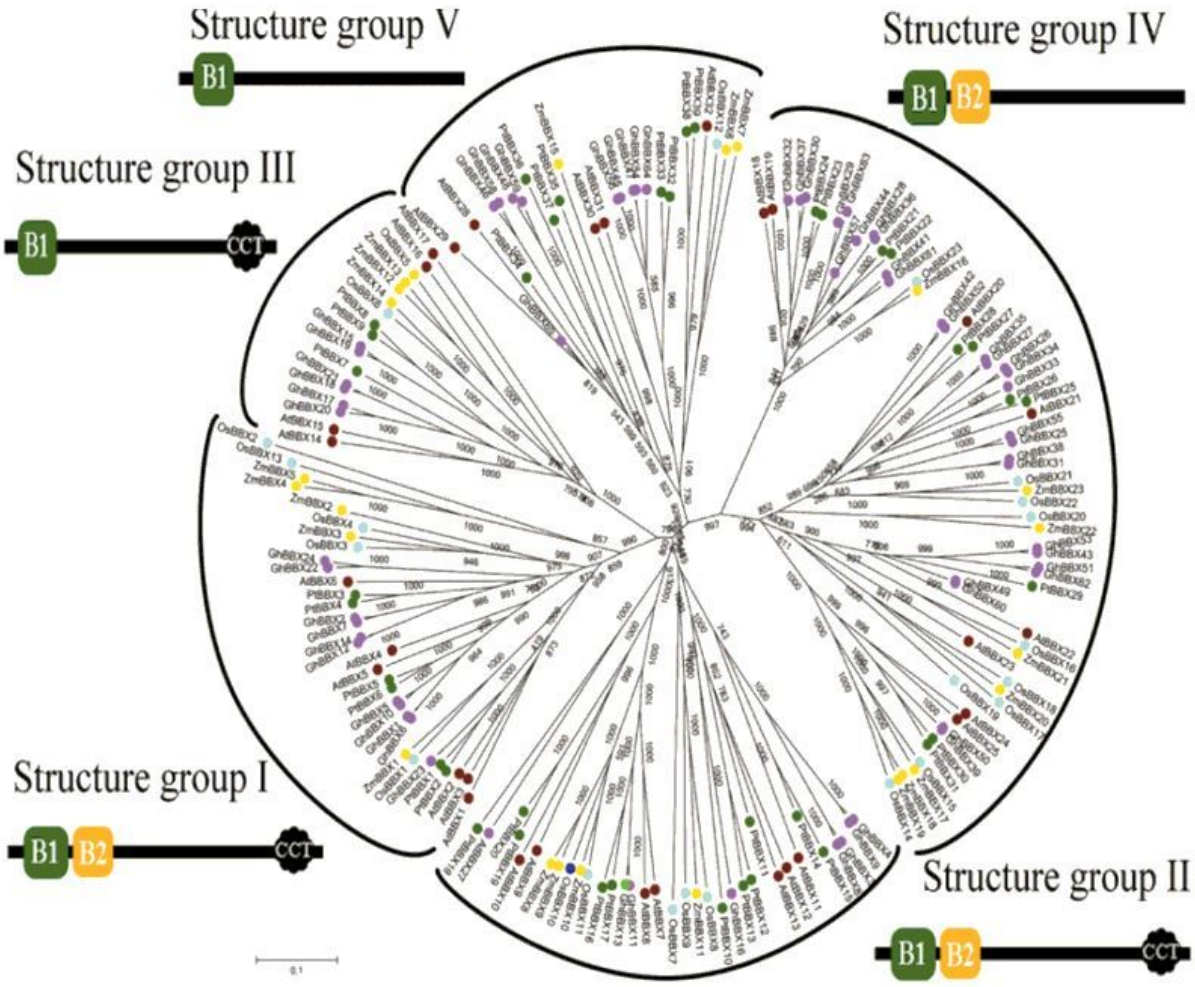

Figure 2

Phylogenetic analysis of BBX genes in P. bretschneideri, O. sativa, A. thaliana, P. trichocarpa, Z. mays and G. hirsutum.

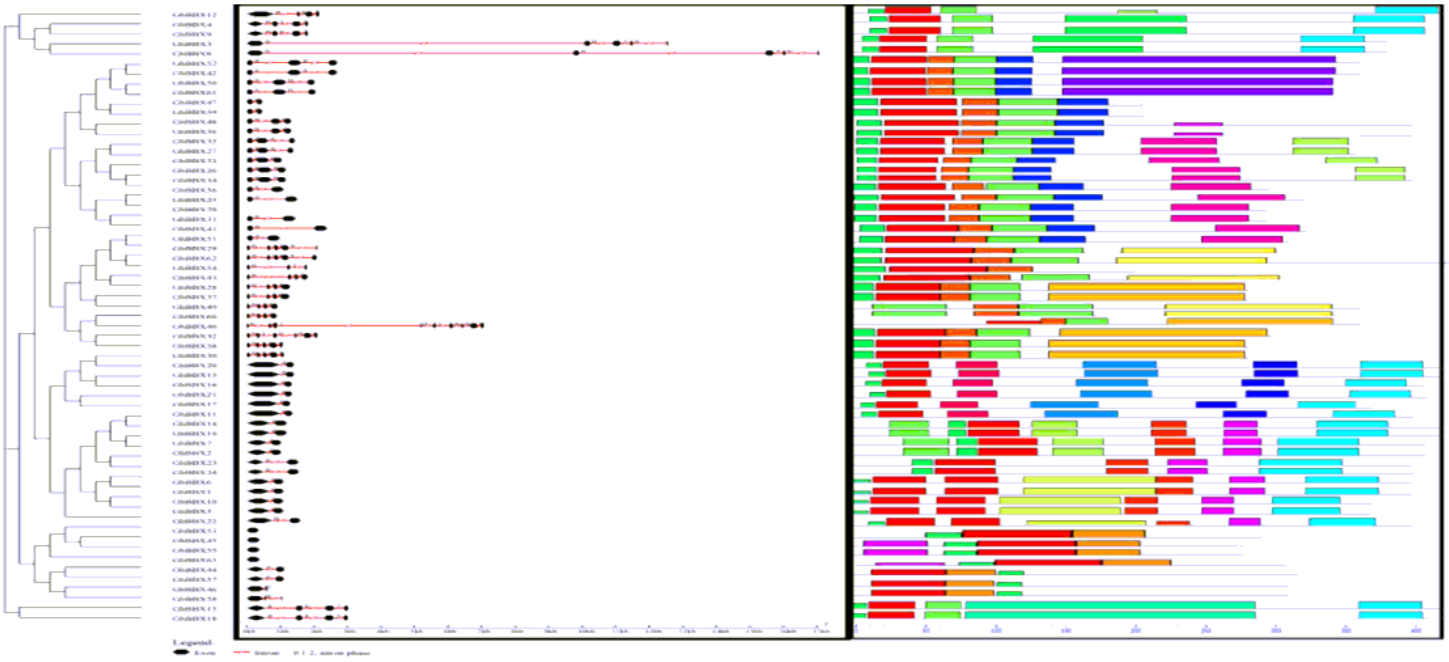

Figure 3

Structural characteristics of the GhBBX genes in upland cotton

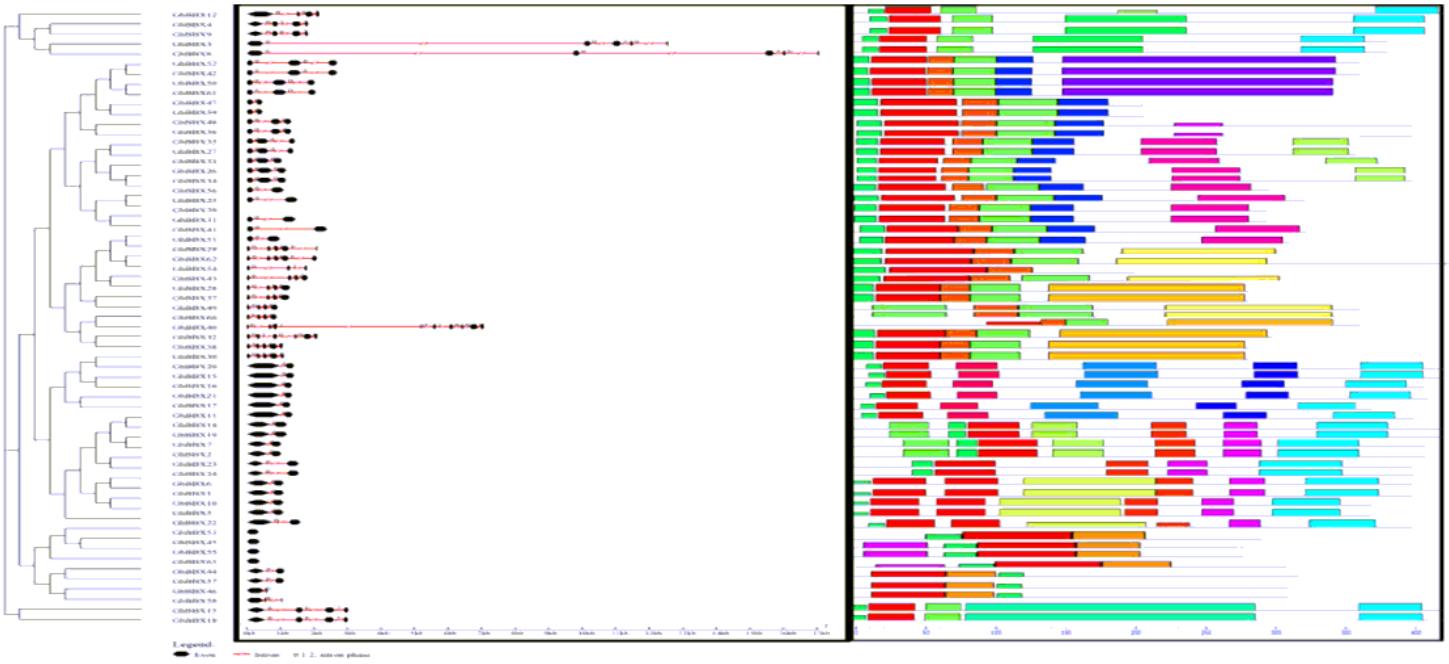

Figure 3

Structural characteristics of the GhBBX genes in upland cotton

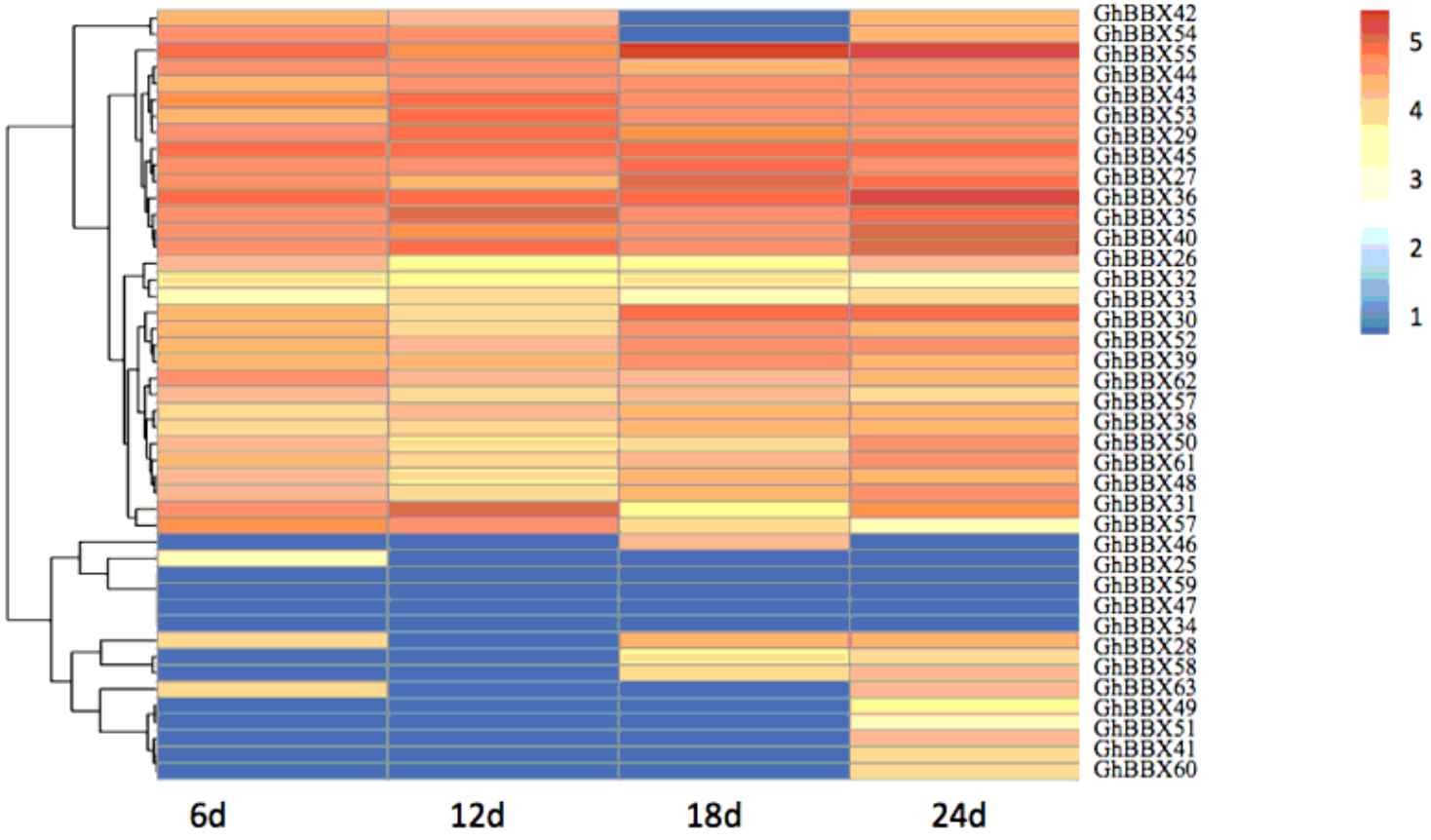

Figure 4

Relative expression of GhBBXs from the structural groups IV and V at different developmental stages of brown cotton fibers

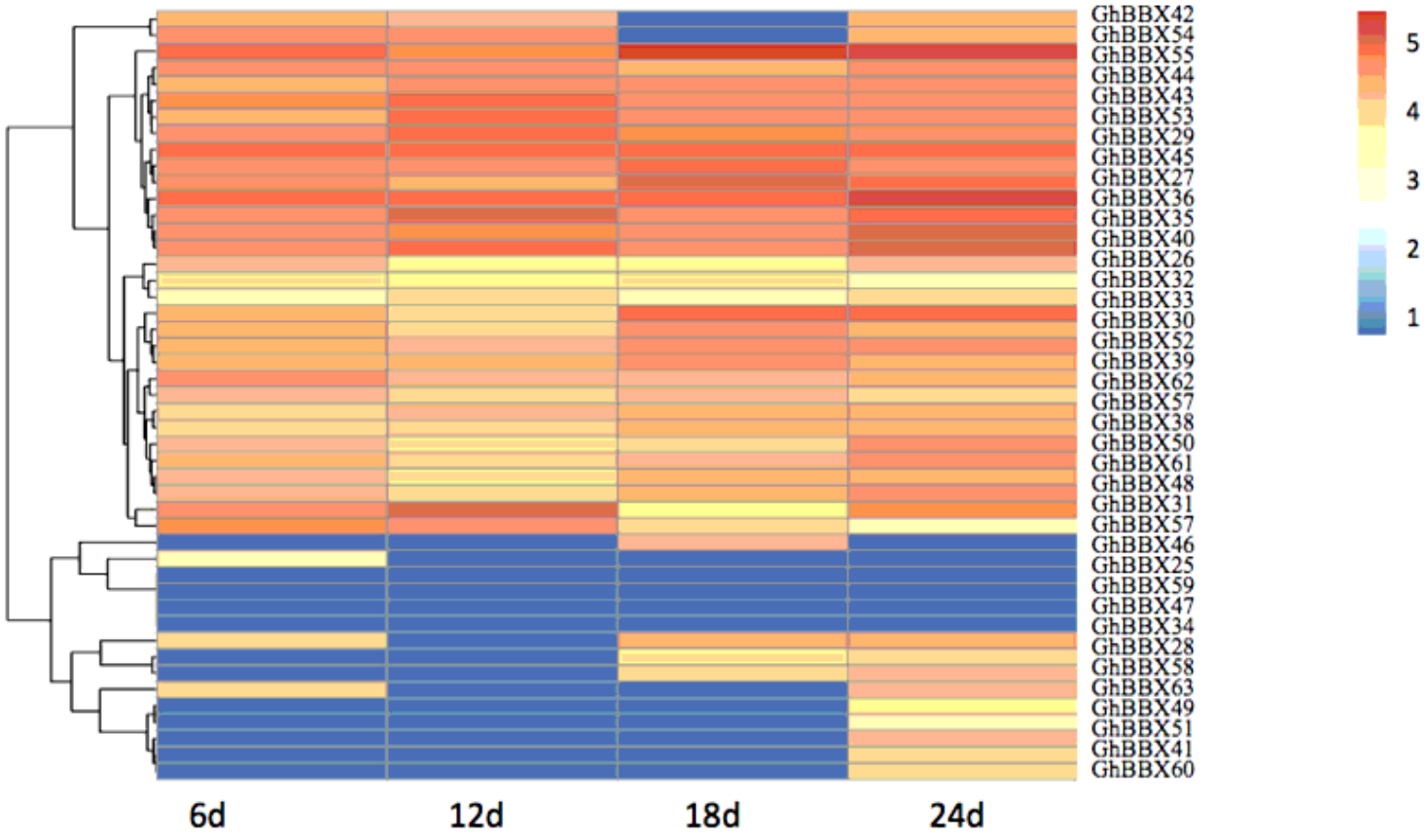

Figure 4

Relative expression of GhBBXs from the structural groups IV and V at different developmental stages of brown cotton fibers

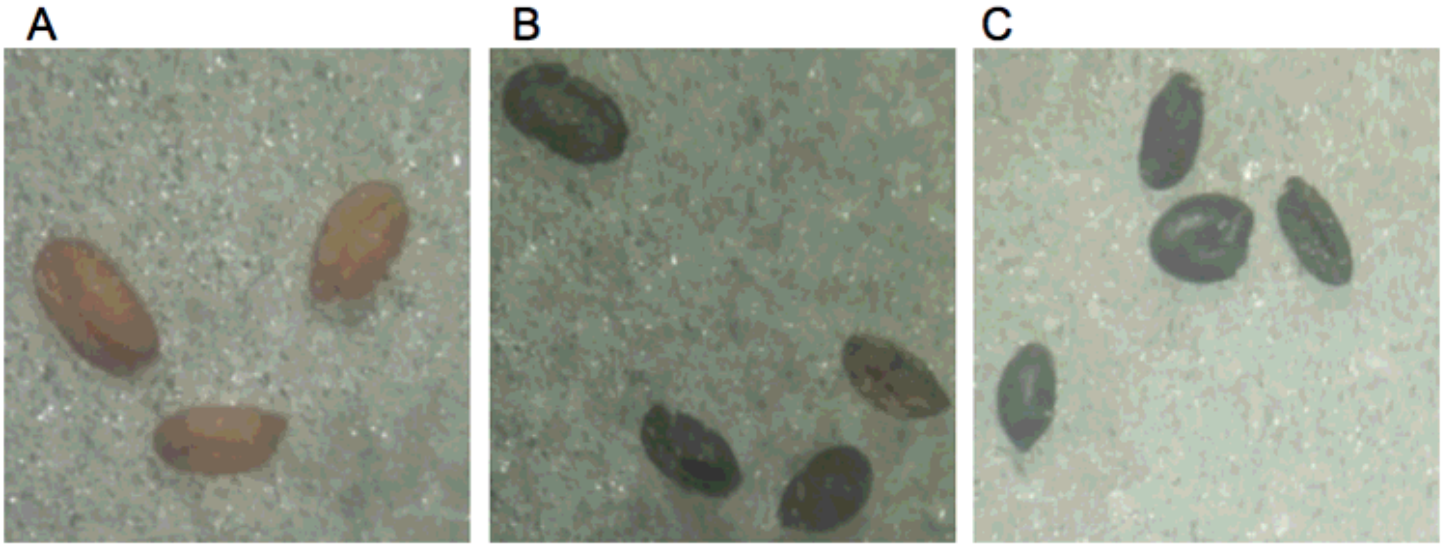

Figure 5

The DMACA staining of Arabidopsis seeds. A: Transgenic Arabidopsis seeds of GhBBX27; B: Transgenic Arabidopsis seeds of GhBBX33; C: Wild-type Arabidopsis seeds

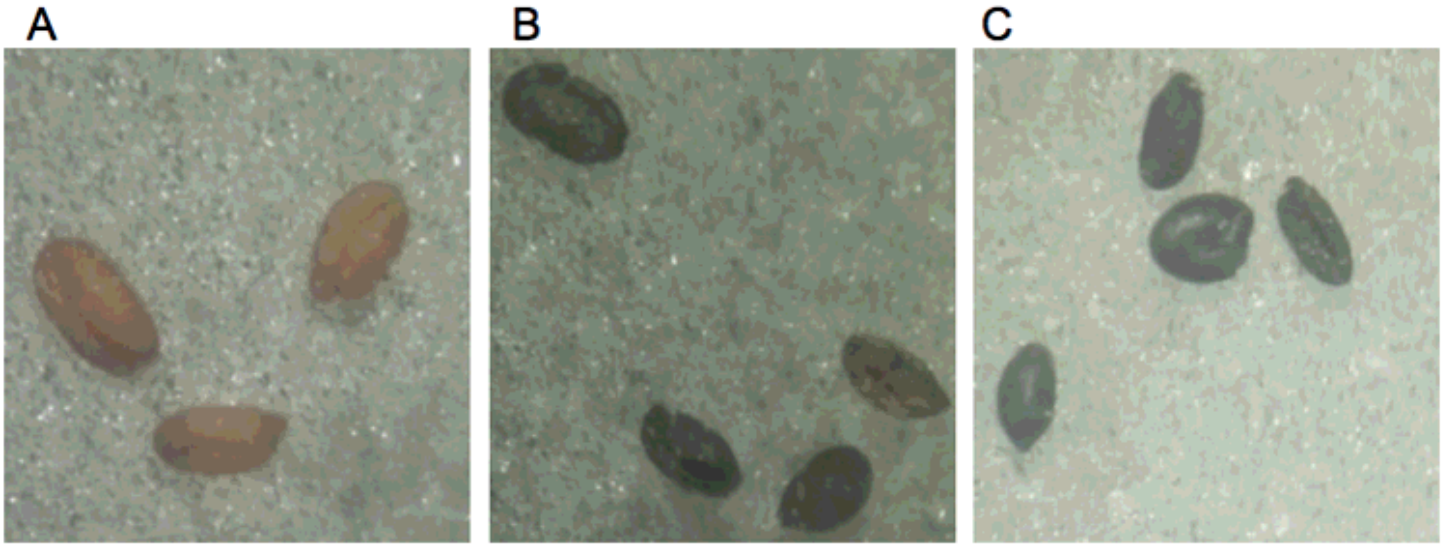

Figure 5

The DMACA staining of Arabidopsis seeds. A: Transgenic Arabidopsis seeds of GhBBX27; B: Transgenic Arabidopsis seeds of GhBBX33; C: Wild-type Arabidopsis seeds

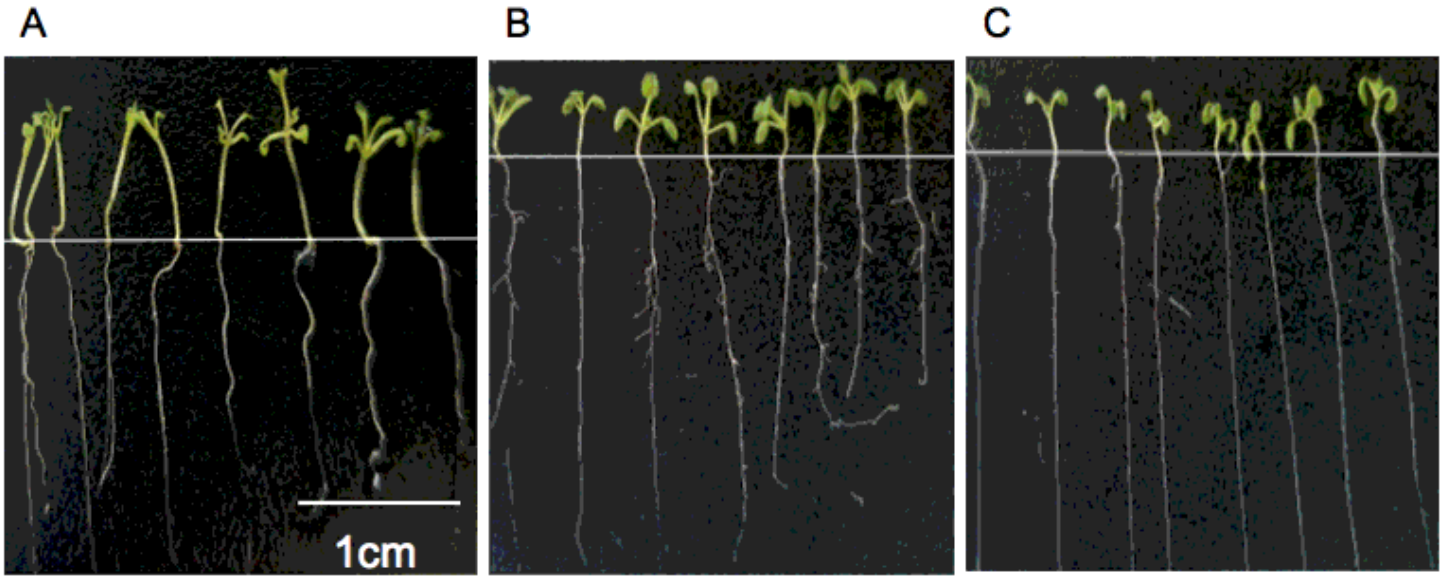

Figure 6

Observation of hypocotyl lengths of transgenic Arabidopsis seedlings. A: Transgenic Arabidopsis seedlings of GhBBX27; B: Transgenic Arabidopsis seedlings of GhBBX33; C: Wild-type Arabidopsis seedlings

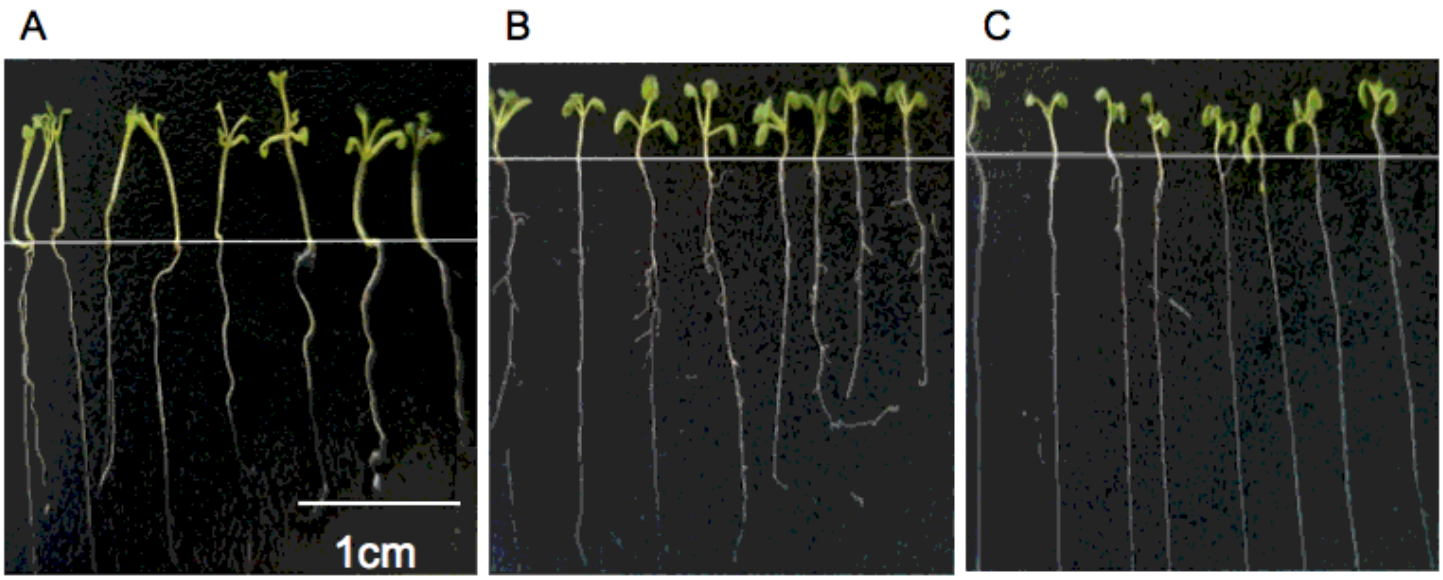

Figure 6

Observation of hypocotyl lengths of transgenic Arabidopsis seedlings. A: Transgenic Arabidopsis seedlings of GhBBX27; B: Transgenic Arabidopsis seedlings of GhBBX33; C: Wild-type Arabidopsis seedlings