A genome-wide identification of the *BLH* gene family reveals *BLH1* involved in cotton fiber development

LIU Cuixia[1,2†], LI Zhifang[3†], DOU Lingling[4†], YUAN Yi[1], ZOU Changsong[3], SHANG Haihong[5], CUI Langjun[1*] and XIAO Guanghui[1*]

Abstract

Background: Cotton is the world's largest and most important source of renewable natural fiber. BEL1-like homeodomain (*BLH*) genes are ubiquitous in plants and have been reported to contribute to plant development. However, there is no comprehensive characterization of this gene family in cotton. In this study, 32, 16, and 18 *BLH* genes were identified from the *G. hirsutum*, *G. arboreum*, and *G. raimondii* genome, respectively. In addition, we also studied the phylogenetic relationships, chromosomal location, gene structure, and gene expression patterns of the *BLH* genes.

Results: The results indicated that these BLH proteins were divided into seven distinct groups by phylogenetic analysis. Among them, 25 members were assigned to 15 chromosomes. Furthermore, gene structure, chromosomal location, conserved motifs, and expression level of *BLH* genes were investigated in *G. hirsutum*. Expression profiles analysis showed that four genes (*GhBLH1_3*, *GhBLH1_4*, *GhBLH1_5*, and *GhBLH1_6*) from BLH1 subfamily were highly expressed during the fiber cell elongation period. The expression levels of these genes were significantly induced by gibberellic acid and brassinosteroid, but not auxin. Exogenous application of gibberellic acid significantly enhanced *GhBLH1_3*, *GhBLH1_4*, and *GhBLH1_5* transcripts. Expression levels of *GhBLH1_3* and *GhBLH1_4* genes were significantly increased under brassinosteroid treatment.

Conclusions: The *BLH* gene family plays a very important role in many biological processes during plant growth and development. This study deepens our understanding of the role of the *GhBLH1* gene involved in fiber development and will help us in breeding better cotton varieties in the future.

Keywords: Cotton, *G. hirsutum*, *BLH* gene family, Fiber development

Introduction

Cotton (*Gossypium* spp.) is an important crop that is used for making lint and cotton seed oil. Apart from the economic values of its fibers and seeds, cotton is also a model plant for the study of polyploids, cell elongation, and cell wall synthesis (Paterson et al. 2012). Owing to its importance, many studies on cotton have been carried out during the last decade (Li et al. 2015a; Chen et al. 2017; Gao et al. 2017; Lu et al. 2017; Zhang et al. 2017), particularly concerning cotton genomics. The genome sequences of two kinds of diploid cotton, *Gossypium raimondii* (*G. raimondii*) and *Gossypium arboreum* (*G. arboreum*), were released in 2012 and 2014, respectively (Paterson et al. 2012; Wang et al. 2012; Li et al. 2014). Genomes of two kinds of cultivated allotetraploid cotton, *G. hirsutum* and *G. barbadense*, were published in 2015 (Li et al. 2015b; Yuan et al. 2015; Liu et al.

---

* Correspondence: ljcui@snnu.edu.cn; guanghuix@snnu.edu.cn
[†]Liu CX, Li ZF, Dou LL contributed equally to this work.
[1]Key Laboratory of the Ministry of Education for Medicinal Plant Resources and Natural Pharmaceutical Chemistry, National Engineering Laboratory for Resouce Development of Endangered Crude Drugs in the Northwest of China, College of Life Sciences, Shaanxi Normal University, Xi'an 710119, Shaanxi, China
Full list of author information is available at the end of the article

2015) and improved in 2018 (Wang et al. 2019). Based on the published sequence data, a comprehensive analysis of a particular gene family can be carried out to uncover its functions, evolution, and expression profiles. These analyses can help us to understand how genes in the gene family control traits at a genome-wide level.

BEL1-like homeobox (BLH) proteins are ubiquitous in plants and include four conserved domains: SKY, BELL, homeodomain, and VSLTLGL. This gene family is known to heterodimerize with KNOTTED1-like homeobox (KNOX) transcription factors (TFs) to regulate transcription of target genes (Müller et al. 2001; Chen et al. 2003; Smith and Hake 2003; Bhatt et al. 2004; Kanrar et al. 2008; Lal et al. 2011; Li et al. 2012). In plants, BLH genes form a small family and their functions are mostly repetitive and redundant. In the genomes of Arabidopsis thaliana, Oryza sativa, Solanum tuberosum, Zea mays, and Glycine max, there are 13, 14, 14, 15, and 34 BLH genes, respectively (Arnayd and Pautot 2014; Mukherjee et al. 2009; Sharma et al. 2014; Meng et al. 2018; Cao et al. 2015; Tao et al. 2018). Several members of this gene family were reported to contribute to developmental processes. Two BELL genes in Arabidopsis (SAW1 and SAW2) were reported to repress the BREVIPEDI-CELLUS (BR) gene, which is a class I KNOX protein involved in leaf patterning. Two other BELL genes in Arabidopsis, PNY and PNF, were also reported to act redundantly in the initiation, maintenance, and development of the shoot apical meristem (SAM) (Byrne et al. 2003; Rutjens et al. 2009; Ung et al. 2011), flowering initiation, and inflorescence architecture (Smith and Hake 2003; Ragni et al. 2008; Khan et al. 2015). In addition, BLH1 is described as a key regulator of ovule development in Arabidopsis (Brambilla et al. 2007; Bencivenga et al. 2012). BLH12 and BLH14 were reported to regulate internode patterning and vein anastomosis in maize (Katsutoshi et al. 2017). BEL11 was involved in chlorophyll synthesis and chloroplast development in tomato fruit (Meng et al. 2018). In potato, the BEL1-KNOX interaction was involved in the tuberization process and root growth (Chen et al. 2003; Lin et al. 2013); StBEL5 has been identified to function as an activator in cytokinin and auxin pathways and a repressor in gibberellic acid (GA) pathway, and also can induce tuber formation (Sharma et al. 2014). Similar to StBEL5, a recent study showed that StBEL11 and StBEL29 function as long-distance signals to regulate the growth of tubers in potato (Ghate et al. 2017). GmBLH4 in soybean might play a role in maintaining the SAM and flower organ development and was might also involve in the pre-harvest seed deterioration process (Tao et al. 2018). However, the BLH gene family has not been identified or characterized in cotton.

G. hirsutum is the most widely planted cotton species in the world, accounting for 90% of all cotton production (Soltis et al. 2004). In this study, 32, 16, and 18 BLH genes were identified from the G. hirsutum, G. arboreum, and G. raimondii genomes, respectively. The phylogenetic relationships, chromosomal location, gene structure, and conserved motifs of the BLH members were investigated to reveal the evolution of this family in cotton. Analysis of gene expression patterns revealed that genes in the same subfamily tend to have similar expression patterns. BLH1 subfamily were upregulated in the fiber cells at 10 days post anthesis (DPA) and 15 DPA, indicating that these genes may play an important role in fiber cell elongation period. Besides, the expression levels of GhBLH1_3 and GhBLH1_4 genes were upregulated after GA and brassinosteroid (BR) treatment, suggesting that GA and BR regulate fiber cell development by enhancing the transcripts of GhBLH1_3 and GhBLH1_4 genes in cotton. Through these analyses, we have increased the knowledge concerning the evolution and function of BLH genes in G. hirsutum.

Materials and methods
Sequence retrieval of BLH and phylogenetic analysis
The 11 BLH CDS and protein sequences in Arabidopsis were downloaded from The Arabidopsis Information Resource (TAIR: http://www.arabidopsis.org/index.jsp). The BLH CDS sequences for three cotton species, G. hirsutum, G. arboreum, and G. raimondii, were acquired in CottonGen (https://www.cottongen.org/) by BLAST with the sequence of Arabidopsis thaliana as the queries. Protein sequences for three cotton species, G. arboreum, G. raimondii, and G. hirsutum, were also downloaded from the CottonGen website (https://www.cottongen.org). MEGA 6.0 (https://www.megasoftware.net) was employed to conduct the phylogenetic analysis. The protein sequences of the identified BLH gene family were imported into MEGA 6.0. After the multiple alignments performed with ClustalW with default parameters (Thompson et al. 1997), the unrooted phylogenetic tree was constructed using the neighbor-joining algorithm (Saitou and Nei 1987) and the bootstrap test was carried out with 1 000 iterations.

Chromosomal locations and gene duplication analysis
Gene annotation files downloaded from CottonGen (https://www.cottongen.org/) were employed to perform the location of BLH genes in G. hirsutum chromosomes. Mapchart 2.2 software (Voorrips 2002) was used to generate the chromosomal location image. The predicted BLH proteins were first aligned using ClustalW2 at EMBL-EBI (http://www.ebi.ac.uk/Tools/msa/clustalw2/) before a gene duplication analysis. Gene duplication events were defined according to the following

LIU *et al. Journal of Cotton Research*   (2020) 3:26

Page 3 of 11

conditions: the alignment region covered more than 80% of the longer gene and the identity of the aligned regions was over 80%.

Gene structure and conserved motif analysis

The exons and introns were explored by comparing the coding sequences with their genomic sequences using the online CSDS Program (http://gsds.cbi.pku.edu.cn/) (Hu et al. 2015). The conserved motifs were predicted with MEME website (http://meme-suite.org/tools/meme).

Expression profiles and promoter analysis

For expression analysis of *GhBLH* genes at different fiber developmental stages, publicly available gene expression files (0–15 DPA) were obtained from CottonGen (https://www.cottongen.org/). A heat map was created using MEV software (Eisen et al. 1998). The promoter sequence file was also downloaded from CottonGen (https://www.cottongen.org/). *Cis*-acting regulatory DNA elements within promoter regions (2 000 bp from the start codon) of selected *GhBLH* genes were scanned using the PLACE database (http://bioinformatics.psb.ugent.be/webtools/-plantcare/html/). The BR related *cis*-elements were identified manually.

Plant material preparation and quantitative real-time polymerase chain reaction (qRT-PCR)

The cotton cultivar Xuzhou 142 was grown in a climate-controlled greenhouse with a 16 h light and 8 h dark cycle at 30 °C. Developing cotton ovules were harvested at 1 DPA, soaked in 10% sodium hypochlorite for surface sterilization, and then rinsed eight times in distilled and deionized water. Ovules were excised from bolls, at 30 °C in darkness without agitation as previously described (Xiao et al. 2016). To study the responses of *GhBLH* genes to GA, BR, and 1-Naphthaleneacetic acid (NAA) application, the cultured ovules were treated with $1 \, \mu mol \cdot L^{-1}$ of GA, $5 \, \mu mol \cdot L^{-1}$ of BR, $5 \, \mu mol \cdot L^{-1}$ of NAA, respectively, treatments without chemical supplementation (CK) were used as the control. GA, BR and NAA were dissolved in ethanol, respectively. Ovules were then harvested for the qRT-PCR experiment after 6, 12, and 24 h treatments, respectively. Three biological replications were used for each time point.

Total RNA was extracted from these samples using PureLink™ RNA mini kit (Invitrogen, Lot no.1687455) with all procedures following the manufacturer's protocol. For the synthesis of cDNA, 5 μg of total RNA was reverse transcribed with random hexamer primers using the SuperScript® III Reverse Transcriptase (Invitrogen, Carlsbad, CA, USA) according to the manufacturer's instructions. In qRT-PCR experiments, each sample was performed with three replicates to ensure the accuracy of results. The reaction parameters were as follows: 95 °C for 10 min, followed by 40 cycles of 95 °C for 10 s and 56 °C for 30 s. A melting curve was generated from 65 °C to 95 °C. The one-way ANOVA analysis was performed using SigmaStat software. Primers for qRT-PCR analysis were listed (Supplementary Table 6). The cotton *UBQ7* gene was served as the internal control for all related results.

Results

Genome-wide identification of the BLH gene family in cotton

To identify the *BLH* genes in *G. hirsutum* (tetraploid) and *G. arboreum* and *G. raimondii* (diploid) genomes, the BLH amino acid sequences from Arabidopsis were used as queries to perform a local BLAST search in cotton genome database. We identified 32, 16, and 18 *BLH* genes in *G. hirsutum*, *G. arboreum*, and *G. raimondii* genome, respectively (Table 1). In *G. hirsutum*, we identified a total of 34 *BLH* gene family members based on the sequence information from the genome assembled by NBI company (Table 1). Thirty-two homologous *BLH* genes were also found based on the assembled genome. Furthermore, another new genome version was used to analyze the *BLH* genes (Yang et al. 2019). As a result, a total of 39 *BLH* genes were identified in this version (Supplementary Table 1). These differences may be due to the assembly error in partial chromosomal regions. Among these genes, 14 *GhBLH* genes were originated from the A_t subgenome and 17 members were from the D_t subgroup. In addition, one gene (*GhBLH3_2*) was located in a scaffold (Supplementary Table 2). Thirty-one orthologs of *GhBLH* gene, except for *GhBLH3_2*, were identified in their two diploid ancestors, *G. arboreum* and *G. raimondii*. Interestingly, the genes orthologous of *GaBLH7_1*, *GaBLH8_4* and *GrBLH8_4* were lost in *G. hirsutum* (Supplementary Table 2). The lengths of the identified GhBLH proteins ranged from 459 amino acids (aa)(GhBLH11_2) to 832 aa (GhBLH2_3) with an average length of 654 aa. The GaBLH proteins ranged from 459 aa (GaBLH11_2) to 806 aa (GaBLH2_4) with an average length of 665 aa and the GrBLH proteins ranged from 459 aa (GrBLH11_1) to 799 aa (GrBLH8_4) with an average length of 642 aa. The coding sequences (CDS) and promoter sequences of 32 *GhBLH* genes were shown (Supplementary Table 3 and Supplementary Table 4).

Phylogenetic analysis of the BLH gene family

To study the phylogenetic relationships of the *BLH* gene family, we performed a phylogenetic analysis of 77 BLH protein sequences, including 11 members from Arabidopsis, 16 from *G. arboreum*, 18 from *G. raimondii*, and 32 from *G. hirsutum*. As shown in Fig. 1, the BLH proteins were classified into 7 distinct groups: BLH1, BLH2/

Table 1 The *BLH* family members of *G. hirsutum*, *G. arboreum*, *G. raimondii,* and *A. thaliana*

Species	*G. hirsutum* (BGI)	*G. hirsutum* (NBI)	*G. arboreum*	*G. raimondii*	*A. thaliana*
BLH1	6	5	2	3	1
BLH2	3	3	1	2	1
BLH3	3	3	1	1	1
BLH4	3	3	2	1	1
BLH5	5	6	3	3	1
BLH6	3	3	1	2	1
BLH7	1	1	1	1	1
BLH8	3	4	2	2	1
BLH9	3	4	2	2	1
BLH10	0	0	0	0	1
BLH11	2	2	1	1	1
Total	32	34	16	18	11

4, BLH3/10, BLH6/7, BLH8/9, BLH5, and BLH11, which were based on their sequence similarities with orthologous in Arabidopsis. Group BLH8/9 and BLH2/4 contained 16 and 14 BLH members, respectively, and constituted the two largest groups in the phylogeny.

Group BLH6/7 and BLH1 each contained 13 members and group BLH5 contained 11 members. The smallest groups were BLH11 and BLH3/10, which only contained 5 members each. To validate the results from the neighbor-joining (NJ) method, we reconstructed the
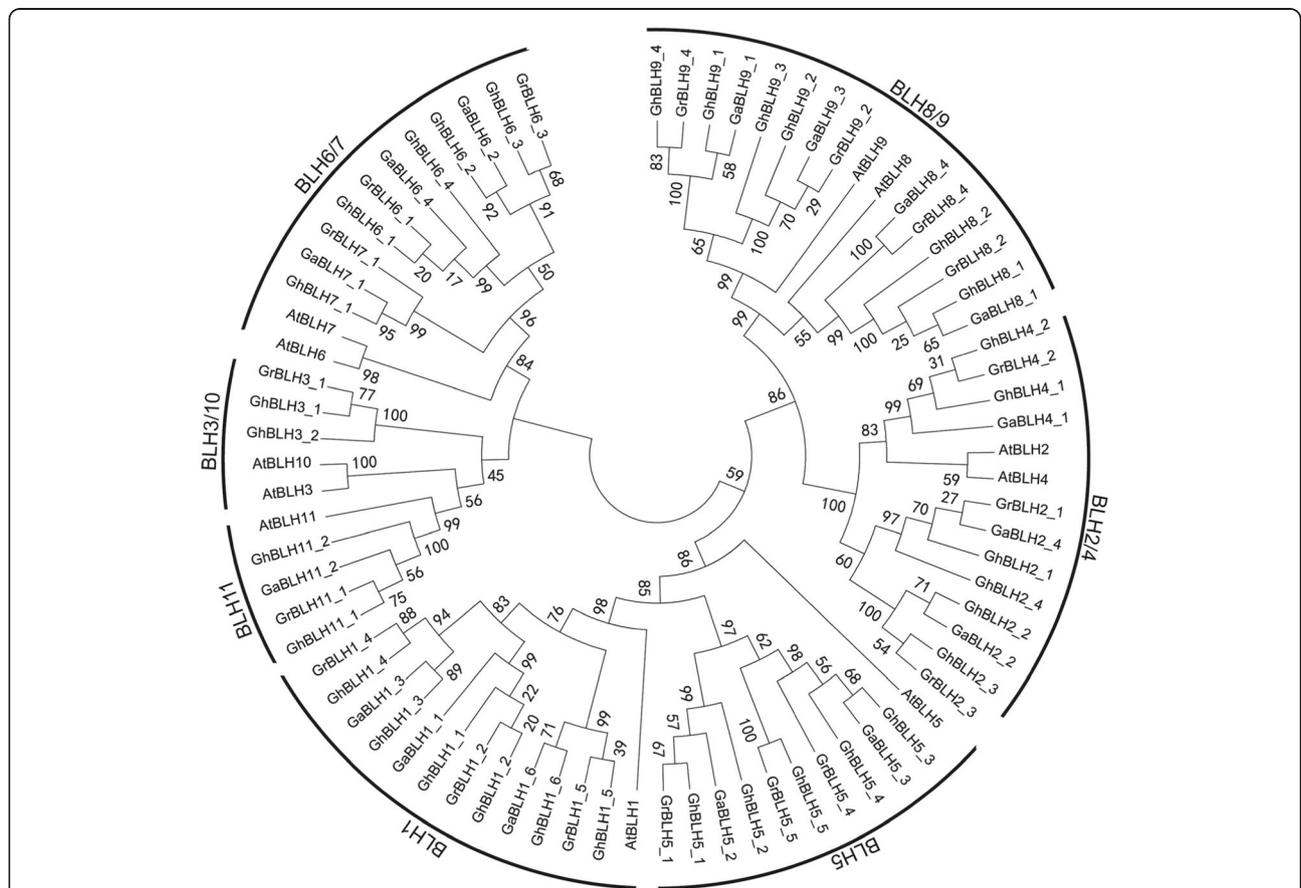

Fig. 1 Phylogenetic tree of the BLH family proteins in *G. arboreum, G. hirsutum, G. raimondii,* and Arabidopsis. The unrooted tree was generated using the MEGA 6.0 program with the neighbor-joining method. Bootstrap values from 1 000 replicates are indicated at each branch. The phylogenetic trees were constructed based on the full-length protein sequences of the *BLHs*

phylogenetic tree of BLH proteins with maximum likelihood method and found that BLH proteins were divided into 7 distinct groups (Supplementary Fig. 1), which was similar to the result of the NJ method.

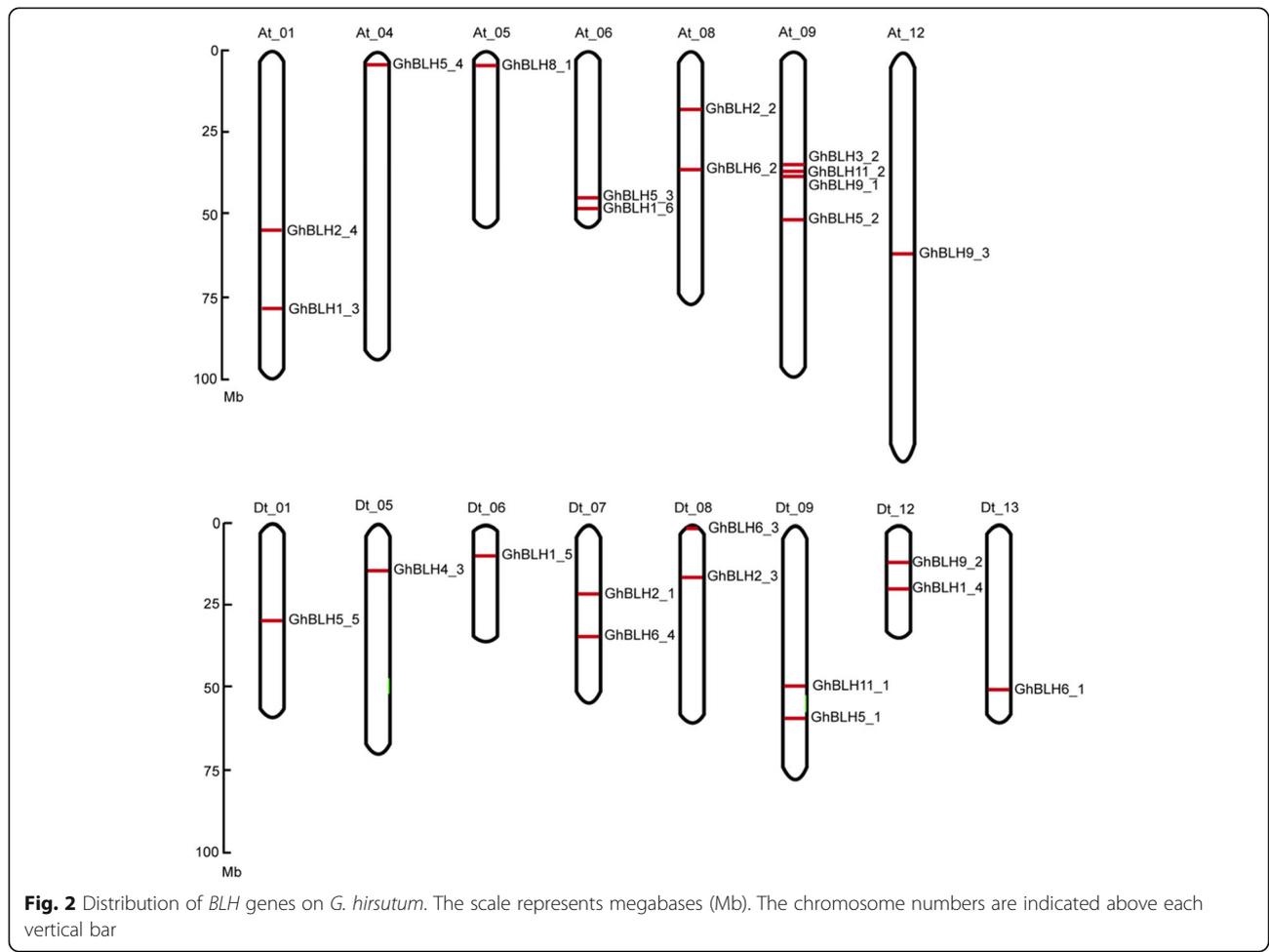### Chromosomal distribution and duplication events of *GhBLH* genes

Based on the annotation of the *G. hirsutum* genome, 25 out of the 32 *GhBLH* genes were assigned to 15 chromosomes including seven were located on the A_t subgenome, eight on the D_t subgenome, while the remaining seven *GhBLHs* were located in unmapped scaffolds (Fig. 2). Eleven chromosomes in *G. hirsutum* had no *GhBLH* genes, whereas chromosome At_9 (with four genes) had the highest number of *GhBLH* genes. The remaining *GhBLH* genes were distributed evenly (one or two genes on each chromosome).
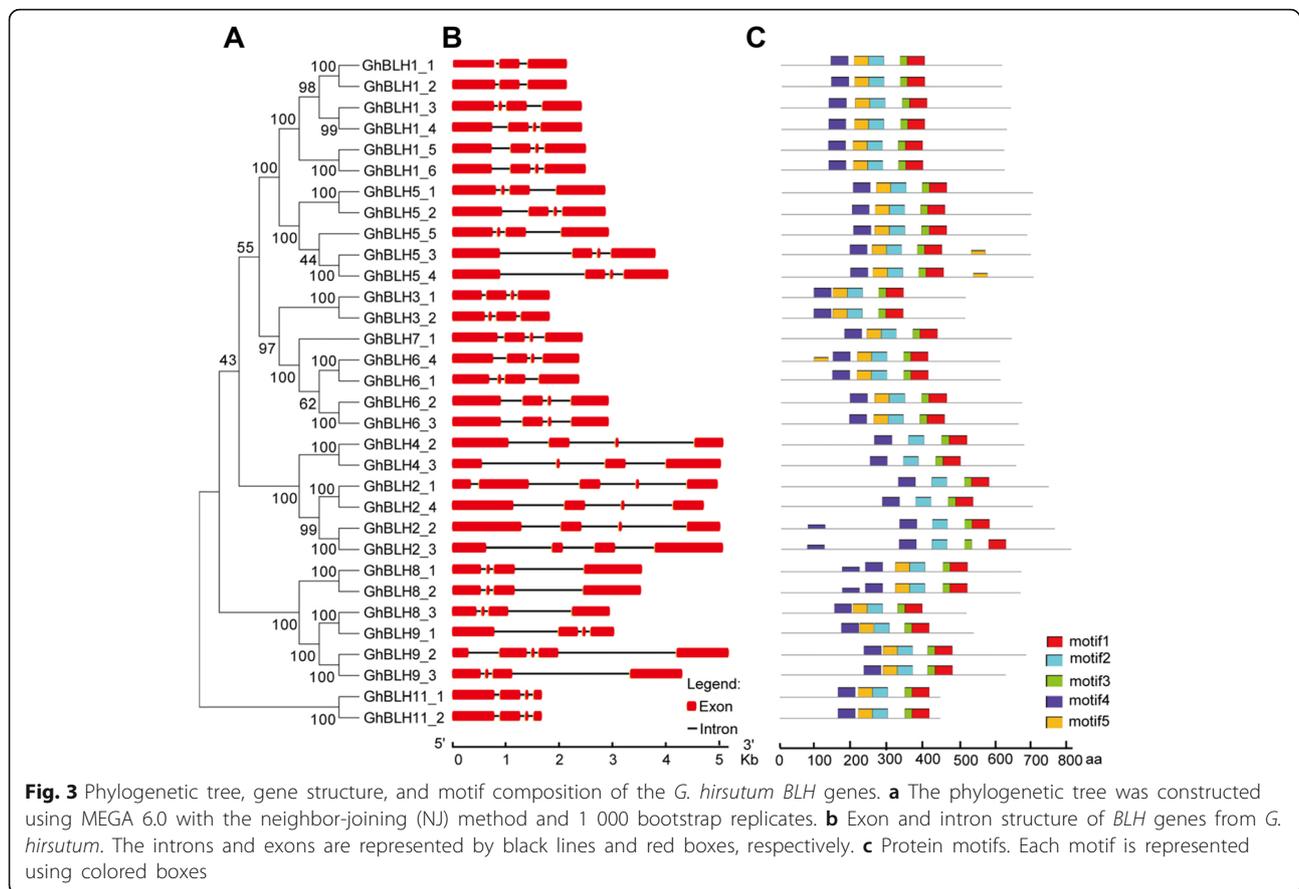
To elucidate the expanded mechanism of the *GhBLH* gene family, we performed a gene duplication event analysis. Twenty-four *GhBLH* genes were produced by whole-genome duplication (WGD) and only one was produced by singleton duplication (Supplementary Table 5). The high proportion of *GhBLH* genes derived from WGD event indicates that this duplication event may play a crucial role in *GhBLH* gene expansion in the *G. hirsutum* genome.

Gene structure and conserved motifs of *GhBLH* genes

To better understand the similarity and diversity of the gene structures of GhBLHs, we constructed a phylogenetic tree with the deduced amino acids of GhBLHs (Fig. 3a) and investigated the exon/intron structures of *GhBLH* genes. *GhBLH* gene can be divided into three groups and all genes had multiple introns according to their exon/intron structures. As shown in Fig. 3b, the majority of *GhBLH* genes (84%) contained four exons in their DNA sequence. Only two genes, *GhBLH1_1* and *GhBLH1_2*, had three exons in their DNA regions and another two genes, *GhBLH2_1* and *GhBLH9_2*, each had five exons (Fig. 3b). The high similarities in *GhBLH* gene structure indicate that *GhBLH* is a conserved gene family. Conserved motifs in the 32 GhBLH proteins were

Fig. 2 Distribution of *BLH* genes on *G. hirsutum*. The scale represents megabases (Mb). The chromosome numbers are indicated above each vertical bar

Fig. 3 Phylogenetic tree, gene structure, and motif composition of the *G. hirsutum BLH* genes. **a** The phylogenetic tree was constructed using MEGA 6.0 with the neighbor-joining (NJ) method and 1 000 bootstrap replicates. **b** Exon and intron structure of *BLH* genes from *G. hirsutum*. The introns and exons are represented by black lines and red boxes, respectively. **c** Protein motifs. Each motif is represented using colored boxes

further identified using the MEME online tool (Fig. 3c). Four conserved domains (motif 1, 2, 3 and 4) were found in all GhBLH proteins. Motif 5 accounted for 81.3% of the GhBLH members, suggesting that these features might have contributed to some specific functions in the GhBLH family. Most members had motifs 1, 2, 3, 4, and 5 indicating that they are highly conserved among GhBLH proteins. We also analyzed the similarity, gene structures and conserved motifs of *GaBLHs* and *GrBLHs*, respectively (Supplementary Fig. 2a and Supplementary Fig. 3a). All *GaBLH* contained four exons in their DNA sequences, except for *GaBLH1_1*, which had three exons (Supplementary Fig. 2b). Most GaBLH proteins possessed five conserved domains (motif 1, 2, 3, 4 and 5), except for *GaBLH5_3*, *GaBLH8_1* and *GaBLH8_4*, in which motif 5 was lost (Supplementary Fig. 2c). All *GrBLH* contained four exons in their DNA sequences, except for *GrBLH1_2* and *GrBLH2_3*, which had three exons, and *GrBLH9_2*, which had five exons (Supplementary Fig. 3b). Most GrBLH proteins possessed five conserved domains (motif 1, 2, 3, 4 and 5), except for *GrBLH11_1*, *GrBLH8_4*, *GrBLH8_3* and *GrBLH9_2*, which lacked motif 5 (Supplementary Fig. 3c).

Expression profiles of *GhBLH* genes in cotton fiber cells
Several reports demonstrated that *BLH* genes were involved in regulating plant development (Ragni et al. 2008; Ung et al. 2011; Bencivenga et al. 2012; Khan et al. 2015; Katsutoshi et al. 2017; Meng et al. 2018), which puts us to investigate the biological functions of *GhBLH* genes in fiber cells during different developmental stages. As shown in Fig. 4, the majority of *GhBLH* genes from the same subfamily had similar expression patterns. The expression levels of four genes, *GhBLH1_3*, *GhBLH1_4*, *GhBLH1_5* and *GhBLH1_6*, which belonging to BLH1 subfamily, were upregulated in cotton fiber cells at 10 and 15 DPA(Fig. 4), indicating that these genes may play an important role in fiber cell elongation period. To better show the biological function of *GhBLH* genes, we analyzed the expression levels of *GhBLH* in other tissues. As a result, *GhBLH1_2* was highly expressed in leaf and torus; *GhBLH1_3* was upregulated in stamen; *GhBLH4_2* showed increased transcripts in stem; both *GhBLH5_3* and *GhBLH5_4* were highly expressed in calycle, leaf and petal (Supplementary Fig. 4). In order to further study the function of *GhBLH1*, we investigated which genes can interact with *GhBLH1*. As a result, *GhBLH1* could interact with several proteins,

including KNAT7 and OFPs (Supplementary Fig. 5), which should be further validated with a yeast one-hybrid experiment.

GhBLH1 genes is involved in phytohormone regulating cotton fiber development

The phytohormones GA, BR and auxin are known to contribute to fiber cell growth in cotton (Luo et al. 2007; Xiao et al. 2010; Zhang et al. 2011). *GhBLH1_3*, *GhBLH1_4*, *GhBLH1_5*, and *GhBLH1_6* genes were highly expressed during cotton fiber cell development. In order to explore the effect of phytohormones on the expression of these genes, we not only analyzed the *cis*-acting regulatory DNA elements within promoter regions (2 000 bp from the start codon) of these four *GhBLH* genes, but also carried out the expression analyses of these *GhBLH* genes. Gibberellin-responsive *cis*-elements, GAREs, were found in the promoter regions of *GhBLH1_3* and *GhBLH1_4* (Fig. 5a). Another regulatory element, P-box, which is also a gibberellin signaling transduction related *cis*-element, was found in *GhBLH1_5* and *GhBLH1_6* promoter regions. Exogenous application of GA significantly activated the transcription of *GhBLH1_3*, *GhBLH1_4* and *GhBLH1_5* (Fig. 5b). BR-related *cis*-elements (CANNT) were found in *GhBLH* promoter region. For *GhBLH1_3*, *GhBLH1_4*, and *GhBLH1_6*, each had five BR-related *cis*-elements. *GhBLH1_5* had eight BR-related *cis*-elements in its promoter region (Fig. 5c). The transcriptional level of *GhBLH1_4* was significantly upregulated after BR treatment for 6 h (Fig. 5d). BR also enhanced *GhBLH1_3* transcripts after it's applied for 12 h. Interestingly, auxin signaling transduction related *cis*-elements (TAG elements) were not found in any of these four *GhBLH* genes (Fig. 5e). Exogenous application of NAA did not promote the transcription of these *GhBLH* genes (Fig. 5f). These results suggested that expression levels of these genes were induced by GA and BR.

Discussion

The *BLH* gene family plays a very important role in many biological processes during plant growth and development. However, previous studies of the *BLH* family have been focused on Arabidopsis, with few reports in cotton. In this study, 32, 18, and 16 *BLH* genes were
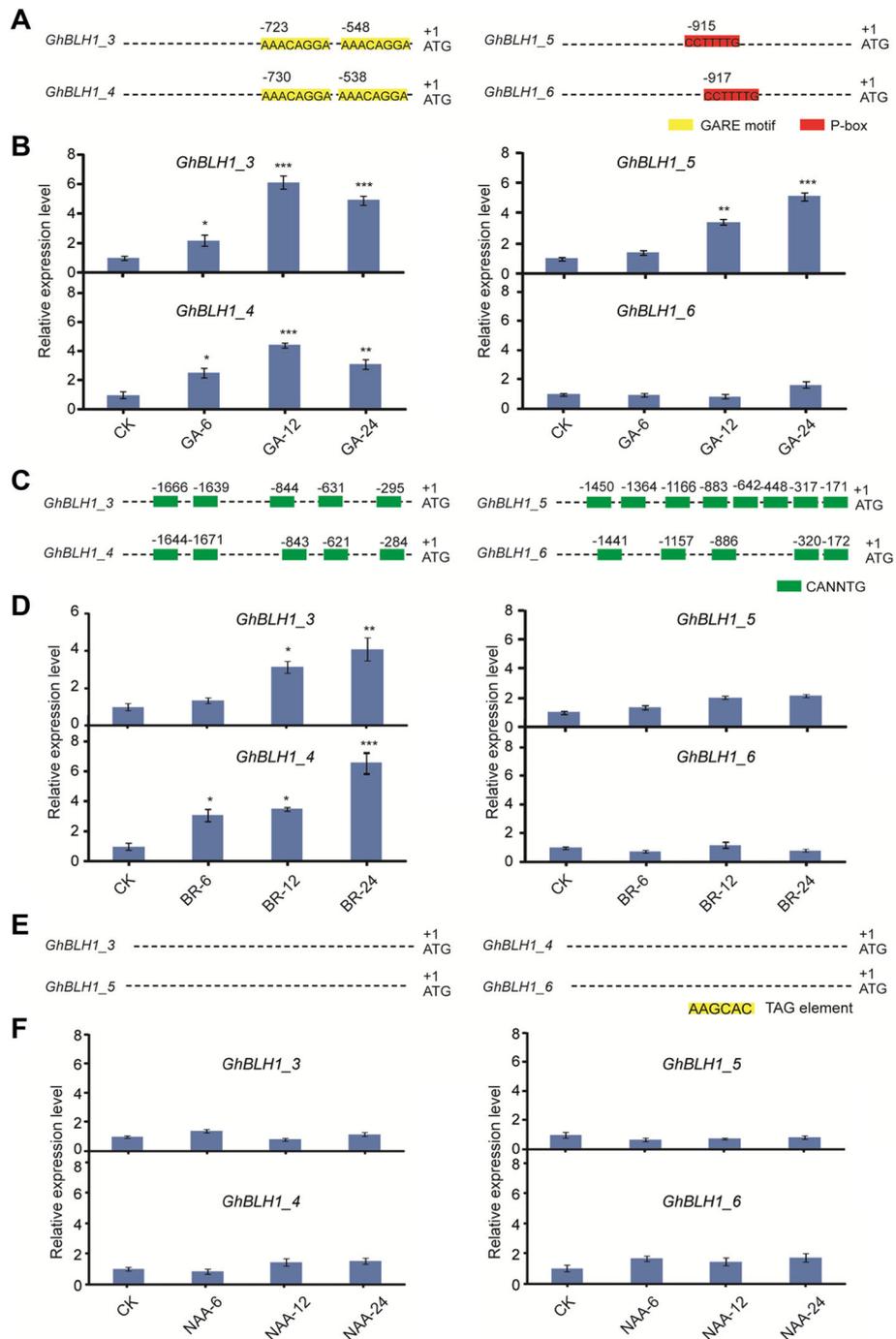
Fig. 5 Expression patterns of several *GhBLH* genes within ovules under GA, BR, and NAA treatment, respectively. **a** Analysis of *GhBLHs* with GAREs and P-boxes presented in their promoter regions. **b** GA activated significant transcription of three *GhBLH* genes with GAREs and P-boxes in their promoter regions. **c** Analysis of *GhBLHs* with CANNT presented in their promoter regions. **d** BR promoted significant transcription activation of two *GhBLH* genes with CANNTG. **e** Analysis of *GhBLHs* with TAG elements presented in their promoter regions. **f** NAA failed to activate the transcription of four *GhBLH* genes with TAG elements. Relative expression levels were determined after normalizing all data to that of CK ovules, which was set to 1.0. Statistical significance was determined using one-way analysis of variance combined with Tukey's test. *$P < 0.05$; **$P < 0.01$; ***$P < 0.001$. Note: The x-axis represents different hours (6, 12, and 24) after GA, BR, and IAA treatment, the y-axis indicates the relative expression levels. Error bars show the standard deviation of three biological replicates

LIU *et al. Journal of Cotton Research* (2020) 3:26

Page 9 of 11

identified and characterized in three sequenced cotton species, *G. hirsutum*, *G. raimondii*, and *G. arboreum*, respectively. Compared with the *BLH* gene in Arabidopsis, more *BLH* genes were found in *Gossypium*. This may be because *G. hirsutum* is tetraploid cotton, which evolved from two kinds of diploid cotton containing an A-genome (*G. arboreum*) and D-genome (*G. raimondii*), repectively, and diverged about 1–2 million years ago (Paterson et al. 2012). Genome-wide analysis of the *BLH* gene family showed that the whole genome duplication contributed to the expansion of *BLH* members. Exon and intron structures and conserved domains varied modestly among different members in cotton, which may be because the expansion of *BLH* family genes was mostly shaped by whole genome duplication. Furthermore, combining the phylogeny of GhBLH proteins with their peptide length, we also found that *GhBLH* genes were highly conserved.

In Arabidopsis, *AtBLH1* has been reported to be one of the major factors controlling ovule pattering, in particular, determining the identity and development of the integuments (Robinson Beers et al. 1992; Ray et al. 1994; Reiser et al. 1995; Brambilla et al. 2007). *AtBLH1* switched one of the synergid cells into an egg cell in the embryo sac (Pagnussat et al. 2007). In the *blh1* and *spl* double mutant, the ovules developed as finger-like structures without integuments (Balasubramanian and Schneitz 2002). In addition, *AtBLH1* was involved in ovule development by interacting with MADS-box transcription factors and SPOROCYTELESS (Brambilla et al. 2007; Bencivenga et al. 2012). In this study, we found that four members of BLH1 subfamily, *GhBLH1_3*, *GhBLH1_4*, *GhBLH1_5* and *GhBLH1_6*, were upregulated in cotton fiber cells at the elongation stage, implying that *GhBLH1* genes may contribute to fiber development in cotton. As the key role of *AtBLH1* gene in ovule development in Arabidopsis, we, therefore, speculate that *GhBLH1_3*, *GhBLH1_4*, *GhBLH1_5*, and *GhBLH1_6* genes may play an important role in ovule growth in cotton, and more proofs were needed. *GhBLH3_1* and *GhBLH7_1* genes were also upregulated in cotton fiber cells at 10 or 15 DPA, indicating that *GhBLH3_1* and *GhBLH7_1* genes may be involved in fiber development in cotton.

The expression levels of *GhBLH1_3* and *GhBLH1_4* genes were upregulated after GA and BR treatments. To date, many studies have shown that BR and GA can significantly promote the elongation of cotton fiber cells (Luo et al. 2007; Xiao et al. 2010). Combining the high expression level of *GhBLH1* genes in developing cotton fibers with the increased transcripts induced by both GA and BR treatments, we suggest that GA and BR regulate fiber cell development by enhancing the transcripts of *GhBLH1_3*, *GhBLH1_4*, and *GhBLH1_5* genes in cotton. Overall, our study provides a basis to systematically elucidate the evolution and function of *BLHs* in cotton and to effectively characterize the precise biological roles of *BLH1* and their utilization in future cotton breeding programs.

Conclusions

The *BLH* gene family plays a very important role in many biological processes during plant growth and development. This study deepens our understanding of the role of the *GhBLH1* gene involved in fiber development and will help us in breeding better cotton varieties in the future.

Supplementary information

Supplementary information accompanies this paper at https://doi.org/10.1186/s42397-020-00068-y.

Additional file 1: Supplementary Table 1. Comparative analysis of *GhBLH* genes in CRI_V1.0 and CRI_V3.0.

Additional file 2: Supplementary Table 2. Analysis of *G. hirsutum* BLH gene family and its orthologs in AA and DD cotton genomes.
Supplementary Table 3. *BLH* gene sequence of *G. hirsutum*.
Supplementary Table 3. *BLH* gene sequence of *G. hirsutum*.
Supplementary Table 4. *BLH* gene Promoter of *G. hirsutum*.
Supplementary Table 5. Analysis of duplication events in *G. hirsutum* BLH genes located in chromosomes. **Supplementary Table 6.** Primers used in this study.

Additional file 3: Supplementary Figure 1. Phylogenetic tree of the BLH family proteins in *G. arboreum*, *G. hirsutum*, *G. raimondii*, and Arabidopsis. The unrooted tree was generated using the MEGA 6.0 program with the maximum likelihood method. Bootstrap values from 1 000 replicates are indicated at each branch. The phylogenetic trees were constructed based on the full-length protein sequences of the *BLHs*.

Additional file 4: Supplementary Figure 2. Phylogenetic tree, gene structure, and motif composition of the *G. arboreum* BLH genes. (A). The phylogenetic tree was constructed using MEGA 6.0 with the neighbor-joining (NJ) method and 1 000 bootstrap replicates. (B). Exon and intron structure of BLH genes from *G. arboreum*. The introns and exons are represented by black lines and red boxes, respectively. (C). Protein motifs. Each motif is represented using colored boxes.

Additional file 5: Supplementary Figure 3. Phylogenetic tree, gene structure, and motif composition of the *G. raimondii* BLH genes. (A). The phylogenetic tree was constructed using MEGA 6.0 with the neighbor-joining (NJ) method and 1 000 bootstrap replicates. (B). Exon and intron structure of BLH genes from *G. raimondii*. The introns and exons are represented by black lines and red boxes, respectively. (C). Protein motifs. Each motif is represented using colored boxes.

Additional file 6: Supplementary Figure 4. Gene expression analysis of *BLH* genes in *G. hirsutum* at calycle, leaf, petal, pistil, root, stamen, stem, torus samples. For expression analysis of *GhBLH* genes at different tissues, publicly available gene expression files were obtained from CottonGen (https://www.cottongen.org/). A heat map was created using MEV software (Eisen et al. 1998). The expression levels are indicated by the color bar: red means high and green means low.

Additional file 7: Supplementary Figure 5. Prediction of interaction proteins of GhBLH1 using STRING website (https://string-db.org/cgi/input.pl?sessionId=tfL8BCaKmXlq& input_page_show_search = on).

Availability of data and materials

The datasets used and analysed during the current study are available from the corresponding author on reasonable request.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors have declared that no competing interests exist.

Author details

[1]Key Laboratory of the Ministry of Education for Medicinal Plant Resources and Natural Pharmaceutical Chemistry, National Engineering Laboratory for Resouce Development of Endangered Crude Drugs in the Northwest of China, College of Life Sciences, Shaanxi Normal University, Xi'an 710119, Shaanxi, China. [2]College of Biology Pharmacy and Food Engineering, Shangluo University, Shangluo 726000, Shaanxi, China. [3]State Key Laboratory of Cotton Biology, State Key Laboratory of Crop Stress Adaptation and Improvement, School of Life Sciences, Henan University, Kaifeng 475001, Henan, China. [4]School of Chemistry and Chemical Engineering, Xianyang Normal University, Xianyang 712000, Shaanxi, China. [5]Zhengzhou Research Base, State Key Laboratory of Cotton Biology, Zhengzhou University, Zhengzhou 450001, China.

References

Arnayd N, Pautot V. Ring the BELL and tie the KNOX: roles for TALEs in gynoecium development. Front Plant Sci. 2014;5:93. https://doi.org/10.3389/fpls.2014.00093.

Balasubramanian S, Schneitz K. NOZZLE links proximaldistal and adaxial-abaxial pattern formation during ovule development in *Arabidopsis thaliana*. Development. 2002;129:4291–300. https://doi.org/10.1101/gad.242202.

Bencivenga S, Simonini S, Benková E, Colombo L. The transcription factors BEL1 and SPL are required for cytokinin and auxin signaling during ovule development in Arabidopsis. Plant Cell. 2012;24:2886–97. https://doi.org/10.1105/tpc.112.100164.

Bhatt AM, Etchells JP, Canales C, et al. VAAMANA-a BEL1-like homeodomain protein, interacts with KNOX proteins BP and STM and regulates inflorescence stem growth in Arabidopsis. Gene. 2004;328:103–11. https://doi.org/10.1016/j.gene.2003.12.033.

Brambilla V, Battaglia R, Colombo M, et al. Colombo L. genetic and molecular interactions between BELL1 and MADS box factors support ovule development in Arabidopsis. Plant Cell. 2007;19:2544–56. https://doi.org/10.1105/tpc.107.051797.

Byrne ME, Groover AT, Fontana JR, Martienssen RA. Phyllotactic pattern and stem cell fate are determined by the Arabidopsis homeobox gene *BELLRINGER*. Development. 2003;130:3941–50. https://doi.org/10.1242/dev.00620.

Cao Z, Li MF, Sun W, et al. Genome-wide identification, expression, and regulation analysis of BEL1-like family genes in maize. Acta Agron Sin. 2015; 41:1632–9. https://doi.org/10.3724/sp.j.1006.2015.01632.

Chen H, Rosin FM, Prat S, Hannapel DJ. Interacting transcription factors from the TALE superclass regulate tuber formation. Plant Physiol. 2003;132:1391–404. https://doi.org/10.1016/j.brainres.2004.03.009.

Chen MY, Li K, Li HP, et al. The glutathione peroxidase gene family in *Gossypium hirsutum*: genome-wide identification, classification, gene expression and functional analysis. Sci Rep. 2017;7:44743. https://doi.org/10.1038/srep44743.

Eisen MB, Spellman PT, Brown PO, Botstein D. Cluster analysis and display of genome-wide expression patterns. Proc Natl Acad Sci U S A. 1998;95:14863–8. https://doi.org/10.1073/pnas.95.25.14863.

Gao W, Long L, Tian XQ, et al. Genome editing in cotton with the CRISPR/Cas9 system. Front Plant Sci. 2017;8:1364. https://doi.org/10.3389/fpls.2017.01364.

Ghate TH, Sharma P, Kondhare KR, et al. The mobile RNAs, StBEL11 and StBEL29, suppress growth of tubers in potato. Plant Mol Biol. 2017;93:563–78. https://doi.org/10.1007/s11103-016-0582-4.

Hu B, Jin JP, Guo AY, et al. GSDS 2.0: an upgraded genefeature visualization server. Bioinformatics. 2015;31:1296–7. https://doi.org/10.1093/bioinformatics/btu817.

Kanrar S, Bhattacharya M, Arthur B, et al. Regulatory networks that function to specify flower meristems require the function of homeobox genes PENNYWISE and POUNDFOOLISH in Arabidopsis. Plant J. 2008;54:924–37. https://doi.org/10.1111/j.1365-313x.2008.03458.x.

Katsutoshi T, Maria J, Abraham J, et al. KNOTTED1 cofactors, BLH12 and BLH14, regulate internode patterning and vein anastomosis in maize. Plant Cell. 2017;29:1105–18. https://doi.org/10.1105/tpc.16.00967.

Khan M, Ragni L, Tabb P, Salasini BC. Chatfield repression of lateral organ boundary genes by PENNYWISE and POUNDFOOLISH is essential for meristem maintenance and flowering in Arabidopsis. Plant Physiol. 2015;169: 2166–86. https://doi.org/10.1104/pp.15.00915.

Lal S, Pacis LB, Smith HM. Regulation of the SQUAMOSA PROMOTER- BINDING PROTEIN-LIKE genes/microRNA156 module by the homeodomain proteins PENNYWISE and POUND-FOOLISH in Arabidopsis. Mol Plant. 2011;4:1123–32. https://doi.org/10.1093/mp/ssr041.

Li FG, Fan GY, Lu CR, et al. Genome sequence of cultivated upland cotton (*Gossypium hirsutum* TM-1) provides insights into genome evolution. Nat Biotechnol. 2015a;33:524–30. https://doi.org/10.1038/nbt.3208.

Li FG, Fan GY, Wang KB, et al. Genome sequence of the cultivated cotton *Gossypium arboreum*. Nat Genet. 2014;46:567–72. https://doi.org/10.1038/ng.2987 .

Li W, Zhao FA, Fang WP, et al. Identification of early salt stress responsive proteins in seedling roots of upland cotton (*Gossypium hirsutum* L.) employing iTRAQ-based proteomic technique. Front Plant Sci. 2015b;6:732. https://doi.org/10.3389/fpls.2015.00732.

Li Y, Pi L, Huang H, Xu L. ATH1 and KNAT2 proteins act together in regulation of plant inflorescence architecture. J Exp Bot. 2012;63:1423–33. https://doi.org/10.1093/jxb/err376.

Lin T, Sharma P, Gonzalez DH, et al. The impact of the long-distance transport of a BEL1-like mRNA on development. Plant Physiol. 2013;161:760–72. https://doi.org/10.1104/pp.112.209429.

Liu X, Zhao B, Zheng HJ, et al. *Gossypium barbadense* genome sequence provides insight into the evolution of extra-long staple fber and specialized metabolites. Sci Rep. 2015;5:14139. https://doi.org/10.1038/srep-14139.

Lu TT, Zhang GF, Sun LR, et al. Genome-wide identification of CBL family and expression analysis of CBLs in response to potassium deficiency in cotton. Peer J. 2017;5:e3653. https://doi.org/10.7717/peerj.3653.

Luo M, Xiao Y, Li X, et al. GhDET2, a steroid 5 α-reductase, plays an important role in cotton fiber cell initiation and elongation. Plant J. 2007;51:419–30. https://doi.org/10.1111/j.1365-313x.2007.03144.x.

Meng L, Fan Z, Zhang Q, et al. BEL1-LIKE HOMEODOMAIN 11 regulates chloroplast development and chlorophyll synthesis in tomato fruit. Plant J. 2018;94:1126–40. https://doi.org/10.1111/tpj.13924.

Mukherjee K, Brocchieri L, Bürglin TR. A comprehensive classification and evolutionary analysis of plant homeobox genes. Mol Biol Evol. 2009;26:2775–94. https://doi.org/10.1093/molbev/msp201.

Müller J, Wang Y, Franzen R, et al. *In vitro* interactions between barley TALE proteins suggest a role for protein-protein associations in the regulation of Knox gene function. Plant J. 2001;27:13–23. https://doi.org/10.1046/j.1365-313x.2001.01064.x.

Pagnussat GC, Yu HJ, Sundaresana V. Cell-fate switch of synergid to egg cell in *Arabidopsis eostre* mutant embryo sacs arises from misexpression of the

BEL1-like homeodomain gene *BLH*1. Plant Cell. 2007;19:3578–92. https://doi.org/10.1105/tpc.107.054890.

Paterson AH, Wendel JF, Gundlach H, et al. Repeated polyploidization of *Gossypium* genomes and the evolution of spinnable cotton fibres. Nature. 2012;492:423–7. https://doi.org/10.1038/nature11798.

Ragni L, Belles-Boix E, Günl M, Pautot V. Interaction of *KNAT*6 and *KNAT*2 with *BREVIPEDICELLUS* and *PENNYWISE* in *Arabidopsis* inflorescences. Plant Cell. 2008;20:888–900. https://doi.org/10.1105/tpc.108.058230.

Ray A, Robinson-Beer K, Ray S, et al. Arabidopsis floral homeotic gene *BELL* (*BELI*) controls ovule development through negative regulation of *AGAMOUS* gene. Proc Natl Acad Sci U S A. 1994;91:5761–5. https://doi.org/10.1073/pnas.91.13.5761.

Reiser L, Modrusan Z, Margossian L, et al. The *BELL1* gene encodes a homeodomain protein involved in pattern formation in the Arabidopsis ovule primordium. Cell. 1995;83:735–42. https://doi.org/10.1016/0092-8674(95)-90186-8.

Robinson-Beers K, Pruitt RE, Gasser CS. Ovule development in wild-type Arabidopsis and two female-sterile mutants. Plant Cell. 1992;4:1237–49. https://doi.org/10.2307/3869410.

Rutjens B, Bao D, van Eck-Stouten E, et al. Shoot apical meristem function in Arabidopsis requires the combined activities of three BEL1-like homeodomain proteins. Plant J. 2009;58:641–54. https://doi.org/10.1111/j.1365-313x.2009.03809.x.

Saitou N, Nei M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol. 1987;4:406–25. https://doi.org/10.1093/oxfordjournals.molbev.a040454.

Sharma P, Lin T, Grandellis C, et al. The BEL1-like family of transcription factors in potato. J Exp Bot. 2014;65:709–23. https://doi.org/10.1093/jxb/ert432.

Smith HM, Hake S. The interaction of two homeobox genes, *BREVIPEDICELLUS* and *PENNYWISE*, regulates internode patterning in the Arabidopsis inflorescence. Plant Cell. 2003;15:1717–27. https://doi.org/10.1105/tpc.012856.

Soltis DE, Soltis PS, Tate JA. Advances in the study of polyploidy since plant speciation. New Phytol. 2004;161:173–91. https://doi.org/10.1046/j.1469-8137.2003.00948.x.

Tao Y, Chen M, Shu YJ, et al. Identification and functional characterization of a novel BEL1-LIKE homeobox transcription factor GmBLH4 in soybean. Plant Cell Tiss Org Cult. 2018;134:331–44. https://doi.org/10.1007/s11-240-018-1419-4.

Thompson JD, Gibson TJ, Plewniak F, et al. The Clustal_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res. 1997;25:4876–82. https://doi.org/10.1093/nar/25.24.4876.

Ung N, Lal S, Smith HM. The role of PENNYWISE and POUND-FOOLISH in the maintenance of the shoot apical meristem in Arabidopsis. Plant Physiol. 2011;156:605–14. https://doi.org/10.1104/pp.110.171462.

Voorrips RE. MapChart: software for the graphical presentation of linkage maps and QTLs. J Hered. 2002;93:77–8. https://doi.org/10.1093/jhered/93.1.77.

Wang K, Wang Z, Li F, et al. The draft genome of a diploid cotton *Gossypium raimondii*. Nat Genet. 2012;44:1098–103. https://doi.org/10.1038/ng.2371.

Wang M, Tu L, Yuan D, et al. Reference genome sequences of two cultivated allotetraploid cottons, *Gossypium hirsutum* and *Gossypium barbadense*. Nat Genet. 2019;51:224–229. https://doi.org/10.1038/s41588-018-0282-x.

Xiao GH, Wang K, Huang G, Zhu YX. Genome-scale analysis of the cotton *KCS* gene family revealed a binary mode of action for gibberellin a regulated fiber growth. J Integr Plant Biol. 2016;58:577–89. https://doi.org/10.1111/jipb.12429.

Xiao YH, Li DM, Yin MH, et al. Gibberellin 20-oxidase promotes initiation and elongation of cotton fibers by regulating gibberellin synthesis. J Plant Physiol. 2010;167:829–37. https://doi.org/10.1016/j.jplph.2010.01.003.

Yang ZE, Ge XY, Yang Z, et al. Extensive intraspecific gene order and gene structural variations in upland cotton cultivars. Nat Commun. 2019;10:2989. https://doi.org/10.1038/s41467-019-10820-x.

Yuan D, Tang Z, Wang M, 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. https://doi.org/10.1038/srep17662.

Zhang GF, Lu TT, Mia WW, et al. Genome-wide identification of ABA receptor PYL family and expression analysis of PYLs in response to ABA and osmotic stress in *Gossypium*. PEER J. 2017;5:e4126. https://doi.org/10.7717/peerj.4126.

Zhang M, Zheng X, Song S, et al. Spatiotemporal manipulation of auxin biosynthesis in cotton ovule epidermal cells enhances fiber yield and quality. Nat Biotechnol. 2011;29:453. https://doi.org/10.1038/nbt.1843.