Bulletin of Biological and Allied Sciences Research
ISSN: **2521-0092**
www.bbasr.org
DOI: https://doi.org/10.64013/bbasr.v2026i1.124
Bull. Biol. All. Sci. Res., Volume, ***11****: 124*
Original researeh Article ***Open Access***

GENOME-WIDE IDENTIFICATION AND CHARACTERIZATION OF PLANT SPECIFIC BPC TRANSCRIPTION FACTOR GENE FAMILY IN COTTON

NADEEM M[1*], AHMAD J[2], SHERAZI SHUH[1*], SHAHID A[1], TAHIR K[1], AHMED H[1]

[1]*Department of Plant Breeding and Genetics, Faculty of Agriculture Sciences, University of the Punjab, Lahore, 54590, Pakistan*
[2]*Department of Agronomy, Faculty of Agriculture Sciences, University of Punjab, Lahore, 54590, Pakistan*
**Correspondence Author Email Address: sheikhmaryam267@gmail.com; syedhaseebulhassansherazi@gmail.com*

(Received, 17th May 2025, Accepted 11th April 2026, Published 23rd April 2026)

Abstract *Cotton is an essential cash crop extensively cultivated throughout the world and serves as the primary natural fiber for the textile industry. However, the cotton crop is intolerant to various environmental stressors that significantly influence crop quality and yield. The growth of plants and abiotic stress reactivity are crucially regulated by the transcription factor BASIC PENTACYSTEINE (BPC). Nevertheless, it is still unknown how the BPC family genes in cotton work and what their molecular foundation is. This study aimed to check the molecular functions of GhBPC gene in abiotic stress resistance. Ten GhBPC genes have been identified and characterized. GhBPC gene arrangement, motif analysis, cis-regulatory regions, and subcellular localization were all thoroughly examined. The gene architectures, chromosome locations, phylogeny, subcellular localization, cis-regulatory analysis, protein motifs, and evolutionary patterns of the BPC gene family in cotton are all thoroughly discussed in this piece of literature. This research offers an in-depth comprehension of the cotton BPC gene family and can be used as a reference for functional analysis and gene family member cloning.*

[*Citation* Nadeem, M., Ahmad, J., Sherazi, S.H.U.H., Shahid, A, Tahir, K., Ahmed, H. (2026). Genome-wide identification and characterization of plant specific BPC transcription factor gene family in cotton. *Bull. Biol. All. Sci. Res.* **11***: 124. doi:* https://doi.org/10.64013/bbasr.v2026i1.124]
Keywords: *BPC transcription factors; genome-wide analysis; cadmium stress; abiotic stress tolerance; gene expression; Gossypium hirsutum*

Introduction

Cotton is an essential cash crop, accounting about 90% of world's total lint yield. Currently, about 45 diploid (2n=2x=26) species and five tetraploid (2n=4x=52) species in cotton (Wendel and Cronn, 2003). The two ancient cotton species, *Gossypium arboreum* (A-type genome) and *Gossypium raimondii* (D-type genome) likely merged to form the allopolyploid cotton cultivar that includes *G. hirsutum* and *G. barbadense*. It is estimated that these two species (*Gossypium arboreum*, *Gossypium raimondii*), each with 26 chromosomes, merged around one to two million years ago. According to (Wendel and Cronn, 2003), the two cotton species probably mixed after transoceanic dispersal and hybridized to form the allopolyploid cotton species we have today. Due to their economic importance, *Gossypium hirsutum* and *G. barbadense* are extensively planted worldwide.

Recent studies have revealed that the allotetraploid emerged about 1.7–1.9 MYA (Hu et al., 2019). The evolutionary and biological features of cotton are widely studied using these two diploid ancestral species and two tetraploid species that are sown extensively. Cotton, being very crucial textile fiber, is a key raw fiber for the textile sector. There is extensive is cultivation of allotetraploid cotton species because of their excellent fiber and high yield. The cotton fibre is a potent single-cell model for this investigation of cell wall formation and is created by the elongation and thickening of the ovule epidermal cell (Haigler et al., 2012; Wu et al., 2019).

BPCs are newly recognized transcription factors that play an integral role in plant growth, developmental responses and abiotic stress responses (Ma et al., 2022; Theune et al., 2017). BPC have largely conserved DNA-binding domain and five cysteine residues in its C-terminus (Berger et al., 2011). In the promoter region, members of the BPC family specifically bind to the GA-rice box or C-box (RGARAGRRAA) to regulate downstream genes and exhibit specialized responses (Hecker et al., 2015). BPC proteins have been referred to as GAGA-binding transcriptional activators due to their ability to attach to GAGA motif in promoter sequences (Petrella et al., 2020). By regulating the transcription of homeotic genes, GAGA motif binding factors (GAF) are considered dynamic transcription factors which regulate a variety of molecular mechanisms (Ma et al., 2022).

STK, a homeotic MADS-box TF, is responsible for controlling conversion from embryogenesis to seed development. The downregulation of STK is primarily influenced by BPC (Kooiker et al., 2005). STK expression was observed in the placenta, ovule primordia, integument, and funiculus. At the later phases of maturation, the integument and funiculus form the seed coat suggesting the activity of *STK* in seed coat formation. It came to notice that the plants with STK mutant have low seed germination performance and alterations in seed coat cell wall features (Ezquer et al., 2016). The regulation of STK by BPC offers strong evidence for its involvement in the regulation of seed coat formation, even though its direct role in seed coat establishment is not reported yet (Ezquer et al., 2016; Kooiker et al., 2005)

Plants rely on the Basic Pentacysteine (BPC) gene family to help them deal with abiotic stress, due to different of environmental difficulties. These genes encode transcription factors that control the expression of many stress-responsive genes, allowing plants to endure and acclamatize to harsh environments (Sahu et al., 2023). Plant reactions to salinity, drought, heavy metals, and cold are mediated by BPC genes (Li et al., 2023; Meng et al., 2023; Zhang et al., 2023). Modulating the expression of genes associated with stress perception and signal transduction is a single among the main ways that BPC genes support abiotic stress resilience. BPC proteins alter transcriptional processes of genes of interest by binding to certain DNA sequences in their promoters (Zhang et al., 2023). Reactive Oxygen Species (ROS), which can oxidatively destroy biological elements like lipids, proteins, and nucleic acids, are frequently produced as a result of abiotic stressors. By promoting the development of antioxidant enzymes like superoxide dismutase (SOD), catalase (CAT), and peroxidases, BPC proteins aid in reducing this damage. By scavenging ROS, these enzymes shield plant cells from oxidative damage and preserve cellular homeostasis (Li et al., 2023; Zhang et al., 2023).

According to the study of (Hu et al., 2020; Xian et al., 2020), the discovery of BBR/BPC TFs as important targets of differentially expressed genes (DEGs) and important candidate genes in stress-related RNA-seq data has further emphasized their function in stress responses.

Cotton is most valuable raw material for textile industries. However, presence of heavy metal like cadmium reduces the growth and cause reduction in biomass of the cotton plant. The effect of cadmium stress on the cotton family has not characterized yet. This paper explains the role of the GhBPC genes under the cadmium stress and different expression patterns of the GhBPC genes under cadmium stress has been observed.

Materials and methods

Screening and extracting sequences from databases

With Pfam (http://pfam.xfam.org/) i.e., PF0621,7 BPC amino acid sequence was obtained. The obtained peptide sequence was then used to locate BPC genes within the *Gossypium hirsutum* genome on publicly available Phytozome (https://phytozome-next.jgi.doe.gov), after which isoforms were excluded. The non-depleting protein sequences were further confirmed with NCBI-CDD database (http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi) and the Motif Finder. By dint of Phytozome (https://phytozome-next.jgi.doe.gov) the chromosomal positions and the amino acids encoded in *G. hirsutum* genes isolated. BPC gene sequences of *Arabidopsis thaliana* were acquired from UniProt (https://www.uniprot.org/). Additionally, BPC gene family sequences in *Solanum lycopersicum* were acquired from the study of (Zhang et al., 2023).

Evaluating the cotton BPC proteins' physiological features

ProtParam tool (http://web.expasy.org/protparam/) was used to recognize protein length (amino acid residues), molecular weight, and theoretical pI of *G.hirsutum* BPC (*GhBPC*) proteins. Phytozome (https://phytozome-next.jgi.doe.gov) was used to get the details for protein sequences, chromosomal locations, gene IDs, and gene sequences. These *GhBPC* genes retitled based on order of their decreasing similarity to the reference sequence. Moreover, the subcellular localization of *GhBPC* genes were anticipated using WoLF PSORT (https://wolfpsort.hgc.jp/).

Phylogenetic analysis and alignment of various sequences

The phylogenetic analysis performed to see evolutionary relationships of *GhBPC* genes in comparison to BPC genes from *A. thaliana* and *S. lycopersicum*. The sequences were aligned by CLustalW alignment method. Phylogenetic tree made by MEGA11 through Neighbour-Joining (NJ) method, with 1000 bootstrap replicates. Model chosed was P-distance, lacking data treatment was adjusted to partial deletion, and site coverage cutoffs were kept at 60%. The analysis included a total of 10 BPC protein sequences from cotton, 7 from Arabidopsis, and 5 from tomato.

Gene structure analysis

Phytozome database was used to get genomic and coding sequences of isolated genes, to analyze the intron/exon arrangement of *GhBPCs.* With help of Gene Structure Display Server (GSDS v2.0) (http://gsds.gao-lab.org/) gene structures of these sequences were visualized.

Conserved motif recognition

Various EM for Motif Elicitation (MEME) (http://meme. nbcr.net/meme/) was implemented to analyze motifs using the predicted protein. Most of motifs were set as 10, while rest of the variables kept at default (Bailey et al.)

Cis_ regulatory elements

Promoter region was found using a 1000-bp upstream sequence from initiation codon of presumed *GhBPC* genes. Plant Care database (http://bioinformatics.psb.ugent.be/webtools/plantcare/html/) used to isolate cis-regulatory elements out of these promoter sequences. Logarithmic scale was employed for the legends.

Synonymous (Ks) and non-synonymous (Ka) substitution rate computation

With help of Simple Ka/Ks calculator in TBtools (https://bio.tools/tbtools) number of nonsynonymous substitutions per nonsynonymous site (Ka) and the number of synonymous substitutions per synonymous site (Ks) in the BPC gene pairs of cotton determined. Rates of molecular adaptation of a particular couple of genes were figured out using the ratio of Ka and Ks. The ratio Ka/Ks < 1 usually pertains to refine selection. Ka/Ks = 1 to neutral selection, while Ka/Ks > 1 to positive selection.

Synteny analysis and Dual synteny

Scaffold length was determined by the Phytozome database; genomic pairs were then formed. With help of Advanced circos tool in TBtools syntenic analysis map constructed. with Multiple Collinearity Scan toolkit (MCScanX), cotton and Arabidopsis were analyzed and Dual Synteny Plot feature in TBtools was employed plot syntenic analysis (Bettaieb and Bouktila, 2020) to demonstrate syntenic relationship of the orthologous BPC cotton gene.

Chromosomal mapping

The chromosomal distribution of Gossypium hirsutum BPC genes on their associated chromosomes was produced in silico by TBtools program. (Chen et al., 2020).

Transcriptome analysis

In order to examine tissue-specific expression of BPC genes under cadmium stress, RNA-seq data of Gossypium hirsutum seedlings treated with 4 mM $CdCl_2$ for 9 hours were acquired. This data was acquired from the NCBI Geo data base. Roots, stems, and leaves were analysed. To find cadmium-responsive and tissue-specific BPC genes, RPKM values were log_2-transformed, and expression patterns were shown using hierarchical clustering and heatmaps in TBtools (Chen et al., 2020).

Results

Recognition of BPC in cotton gene

Utilizing the BLAST on Phytozome database, 31 BPC genes in cotton obtained in total. After the removal of redundant and truncated sequences, 10 concise *GhBPC* genes isolated and subsequently used for analysis. The isoelectric point (pI), molecular weight (MW), protein length (amino acid residues), and predicted location of *GhBPC* proteins, with various physicochemical properties shown in Table 1. Encoding protein lengths varied from 109 to 341 amino acids in (GhBPC2 and GhBPC8) and molecular weights ranged from 12,055.91 kDa in GhBPC4 to 37,163.07 kDa in GhBPC8. The isoelectric point (pI) values of *GhBPC* proteins between 9.35 to 10.21, with GhBPC4 and GhBPC9 having the respective extreme values (Figure 1).

Table 1 10 non-redundant BPC genes isolated from cotton

Gene ID	Gene Name	Scaffold	Location	mRNA (CDS)length	Protein AA	MW	PI value
>Gohir.D02G137700	GhBPC1	D02	50766039..50769503	843	284	31492.08	9.46
>Gohir.D08G173450	GhBPC2	D08	56011802..56012556	438	147	16665.82	9.91
>Gohir.A01G143100	GhBPC3	A01	72778161..72781273	849	286	31757.41	9.65
>Gohir.A03G113332	GhBPC4	A03	73220336..73221144	327	109	12055.91	9.35
>Gohir.A13G086445	GhBPC5	A13	43545570..43549997	933	315	34849.82	9.55
>Gohir.A04G143500	GhBPC6	A04	87993378..87996487	915	309	34107.14	9.58
>Gohir.D01G135000	GhBPC7	D01	42548761..42551862	849	286	31767.4	9.7
>Gohir.D01G101100	GhBPC8	D01	18380361..18384476	1011	341	37163.07	9.49
>Gohir.D04G187800	GhBPC9	D04	57289007..57291730	597	201	21933.3	10.21
>Gohir.D04G187002	**GhBPC10**	**D04**	**57284164..57287425**	**849**	**286**	**31637.29**	**9.67**

Phylogenetic analysis

The protein sequences of 22 BPC genes from various species (Arabidopsis, tomato, and cotton) were organized into three groups. Based on phylogenetic analysis, Group I had 12 members, with seven genes of cotton: GhBPC1, GhBPC3, GhBPC4, GhBPC6, GhBPC7, GhBPC9 and GhBPC10; two genes of tomato: SlBPC1 and SlBPC5; three genes of Arabidopsis: AtBPC1, AtBPC2 and AtBPC3. Group II comprise of eight members, with three genes of cotton: GhBPC2, GhBPC5 and GhBPC8; two genes of tomato SlBPC2 and SlBPC4; three genes of Arabidopsis: AtBPC4, AtBPC5 and AtBPC6. Group III didn't include any cotton (*GhBPC*) gene and had only two members in total, including one gene of

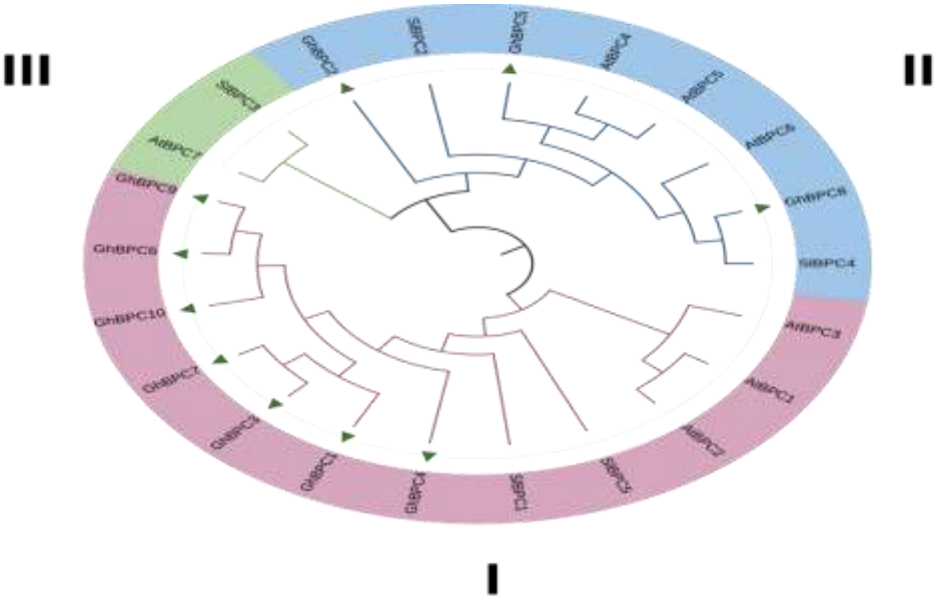

Figure 1 Phyogenetic and evolutionary linkage amonge BPC gene family of cotton, Arabidopsis and tomato. Green arrows showing proteins which are GhBPC proteins. Phylotree was constructed with MEGA11 software through Neighbour-Joining (NJ) method, by1000 bootstrap replicates

tomato (SlBPC3) and one of Arabidopsis (AtBPC7) (Figure 1)

Gene structures and isolation of conserved motifs and domain

Conserved motifs were examined to determine the evolutionary relationship of cotton BPC protein; the findings are shown in (Fig.3). We found that the GhBPC gene family members had 10 motifs, with motif 1 and 3 being highly conserved. GhBPC5 and GhBPC8 had six motifs and are structurally and homologously identical. In comparison, GhBPC2 had one motif and GhBPC4 had four motifs (Figure 2). According to (Figure 2), GhBPC4 had six motifs while GhBPC1, GhBPC3, GhBPC7, GhBPC6, and GhBPC10 had nine motifs. In the BPC family genes of cotton plants motifs 1,2,3 and 7 were equally shared (Fig.3). Current findings revealed, genes of the BPC family were closely related to those of their evolutionary relatives and are dispersed similarly. Additionally, BPC genes contain two conserved domains: the GAGA bind and GAGA_bind super family. Cotton has the GAGA_bind super family domain in all of its members. Exons and introns are fundamental structural components of genes.

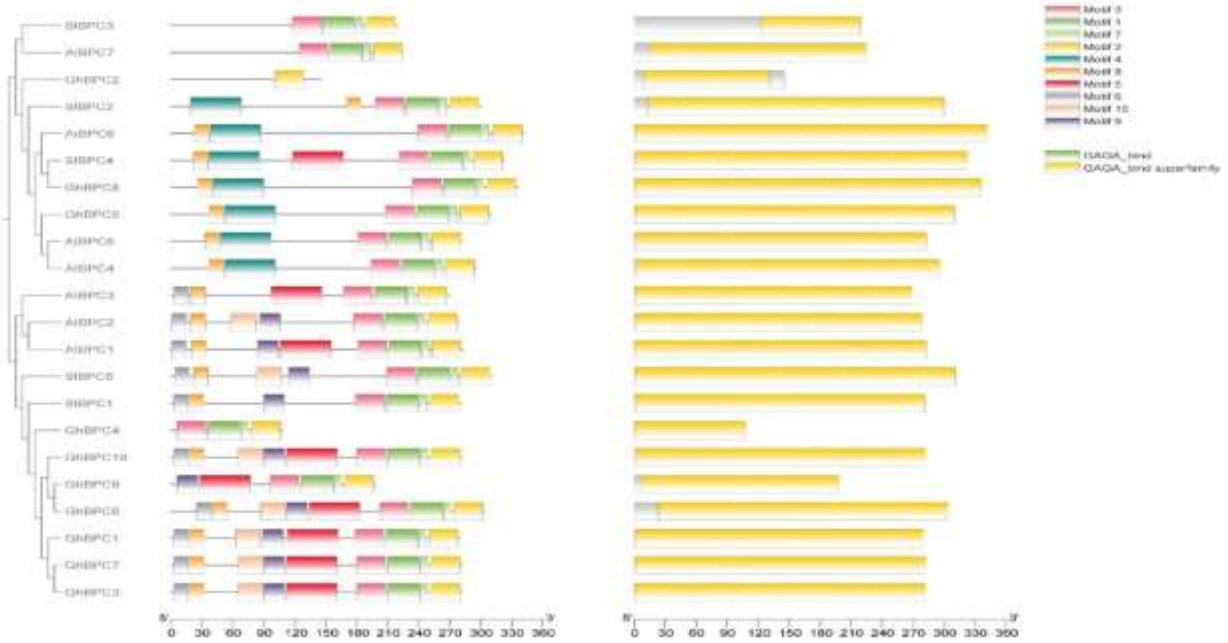

Figure 2 With help of deduced amino acid sequences of ten GhBPCs and connecting them to the phylogenetic tree and BPC with NCBI, three MEME programs were utilized to find motifs *(Bailey et al., 2015)*

Positioning of a gene's exons and introns can notably influence its function. The combination of these building block influences gene functions. In the analysis, Six out of ten examined genes were intron-less, while the remaining four showed presence of a single intron. According to the genomic architecture, GhBPC2, GhBPC5, GhBPC6 and GhBPC8 had one intron, whereas GhBPC1, GhBPC3, GhBPC4, GhBPC7, GhBPC9 and GhBPC10 were intron-less (Figure 3).

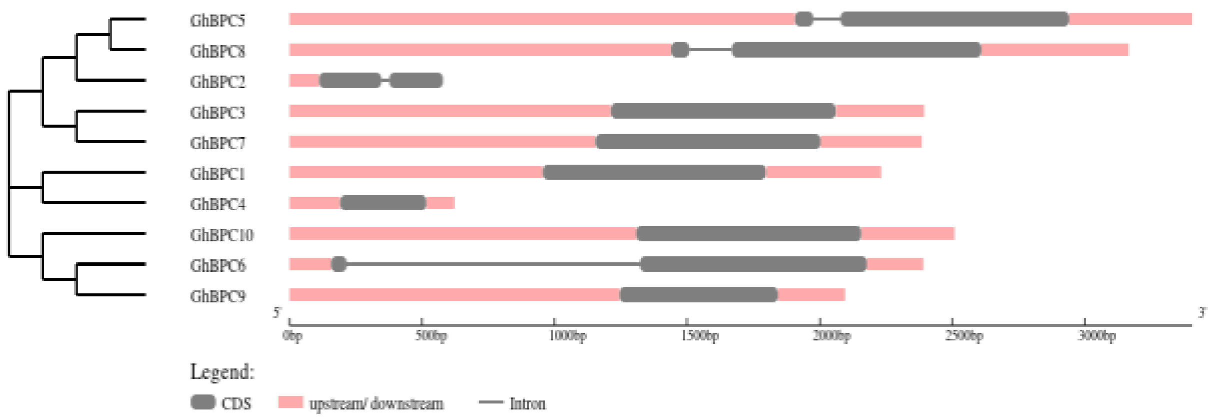

Figure 3 Phylogenetic relationships and gene framwork of BPC genes of cashew. By dint of full-length sequences of cotton BPC phylogenetic tree was made. Positioning of exons and introns in cotton genes also evaluated. Exons are indicated as gray bars, introns are by dark lines, and upper and lower parts shown with pink bars

.

Subcellular localization

The specific localization of each *GhBPC* protein was identified by subcellular localization analysis. All the proteins localized themselves in the nucleus. However, some *GhBPC* proteins were localized at different additional locations. GhBPC1, GhBPC2, GhBPC3, GhBPC4, GhBPC5 and GhBPC8 were localized in cytoplasm, while only GhBPC1 was detected in cytoskeleton. GhBPC2 and GhBPC9 were localized in mitochondria; while in the chloroplast, GhBPC2, GhBPC4 and GhBPC9 were confined. GhBPC2 being the sole protein was localized in plasma membrane (Figure 4).

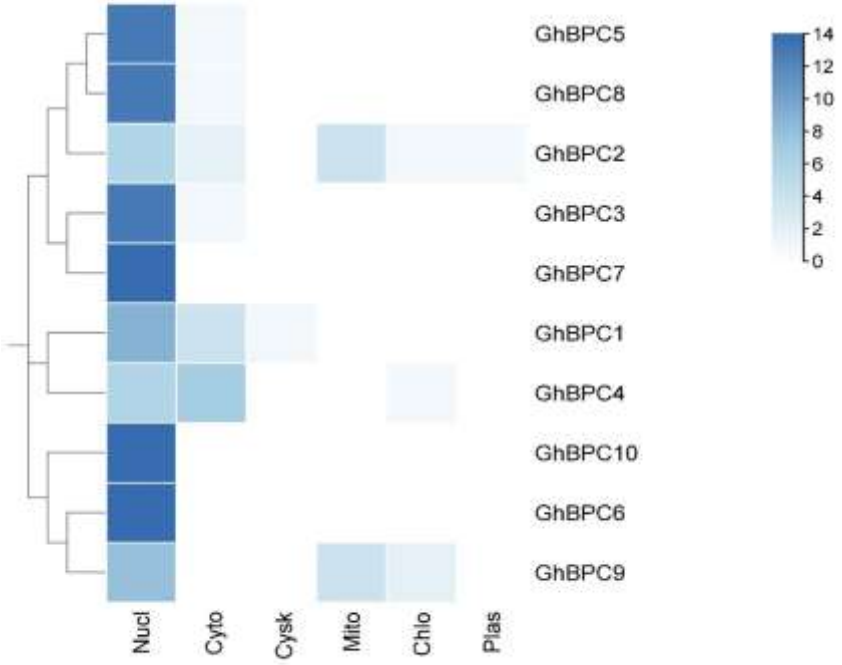

Figure 4 Heat map to express the subcellular localization of all the GhBPC proteins. Horizontal-axis shows names of six organelles, and vertical-axis indicates different GhBPC genes

Examination of cis-regulatory elements

To examine presumed feature of genes in-silico analysis of different cis-regulatory elements was carried out. Cis-regulatory elements with annotated features such as reaction toward light, anaerobic induction, MeJA-responsiveness, promoter region elements, zein metabolism, abscisic acid responsiveness, gibberellin-responsive element, drought-inducibility, auxin responsiveness element, low temperature responsiveness, endosperm expression, meristem expression, flavonoid gene regulation and salicylic responsiveness. Cis-elements such as AT1-motif, ATC-motif, Box 4, GATA-motif, GT1-motif, TCT-motif, chs-CMA1a, AE-box, Box II, G-Box, G-box, GA-motif and ACE are involved in light responsiveness. The analysis of 10 *GhBPC* genes revealed the following distribution of light-responsive cis-elements: The AT1-motif and ATC-motif were each found in 1 out of 10 genes. The Box 4 element was present in 8 out of 10 genes, while the GATA-motif and AE-box were found in 2 out of 10 genes each. The GT1-motif appears in 8 genes, and the TCT-motif was present in 7 genes. The chs-CMA1a element was observed in 2 out of 10 genes. The Box II cis-element was present in a single gene, whereas the G-Box was found in 5 genes, and G-box is found in 4 genes. The GA-motif was present in 5 out of 10 genes, and the ACE element was present in 1 out of 10 genes. Including these all the cis elements, it is evident that each *GhBPC* gene demonstrates the function of light responsiveness. All genes contain one or more light responsive elements, which indicates that all of these are responsible for light sensitivity.

Cis elements as well as CGTCA-motif and TGACG-motif were involved in MeJA-responsiveness. Analysis has indicated that, CGTCA-motif was present in 3 out of 10 genes, while TGACG was present in 3 genes out of 10. Promoter region includes, CAAT-box, 5UTR Py-rich stretch and TATA-box elements. CAAT-box and TATA-box were each present in every genes, 5UTR Py-rich stretch was observed in a single gene. ARE element found in 6 out of 10 genes which is necessary for anaerobic expression., 4 *GHBPC* genes have O2-site that is related to Zein metabolism regulation, Abscisic acid responsiveness was shown by ABRE which was indicated in 5 genes, sole *GhBPC* gene possess the GARE-motif that is related to gibberellin response, MBS was present in 6 out of 10 genes which was responsible for drought-inducibility, auxin-responsiveness was shown by TGA-element which was present in 2 *GhBPC* genes, GCN4_motif was present in 3 genes out of 10 that is related to endosperm expression, 1 *GhBPC* gene have LTR element which help in cold temperature sensitivity, GCAT-box appeared in 2 genes which is responsible for meristem expression, flavonoid biosynthetic genes regulation is the function of MBSI which is present in 2 genes and both CCAAT-box and TCA-element are present in 1 *GhBPC* gene which are responsible for MYBHv1 binding and salicylic acid responsiveness respectively (Figure 5).

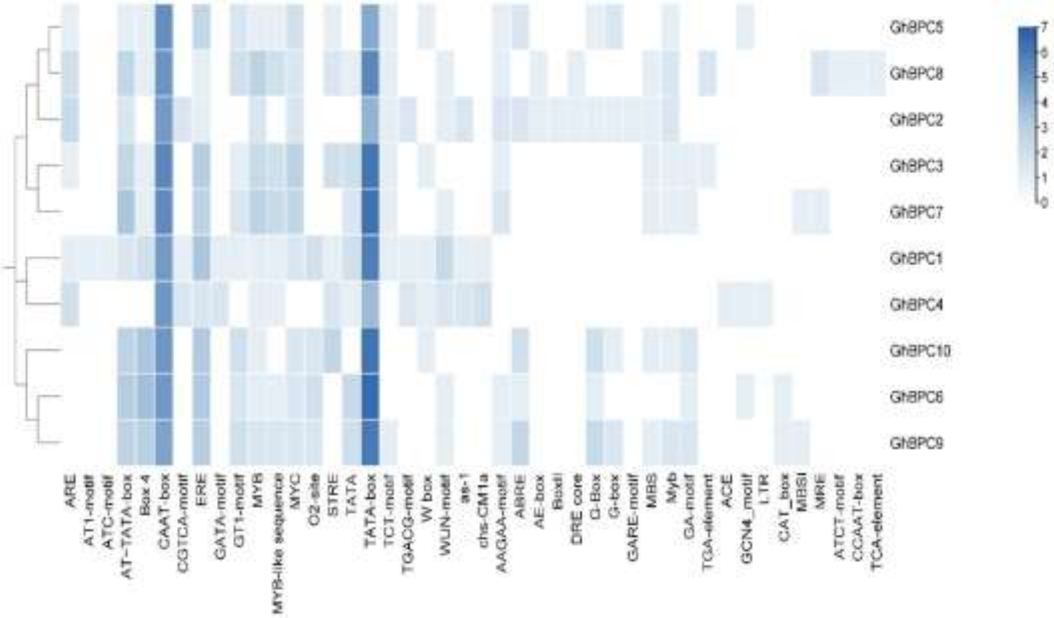

Figure 5. Cis-regulatory elements (CREs) examination in presumed GhBPC gene associated with phylogenetic tree

Assessment of Ka/Ks ratio

The Ka/Ks ratio, which indicates the ratio of nonsynonymous to synonymous mutations, was then ascertained manually. Ratio varied across different pairs: 0.09 for GhBPC5_GhBPC8, 0.170 for GhBPC3_GhBPC7, and 0.069 for the GhBPC1_GhBPC4 and 0.5346 for GhBPC6_GhBPC9 pair (Figure 6).

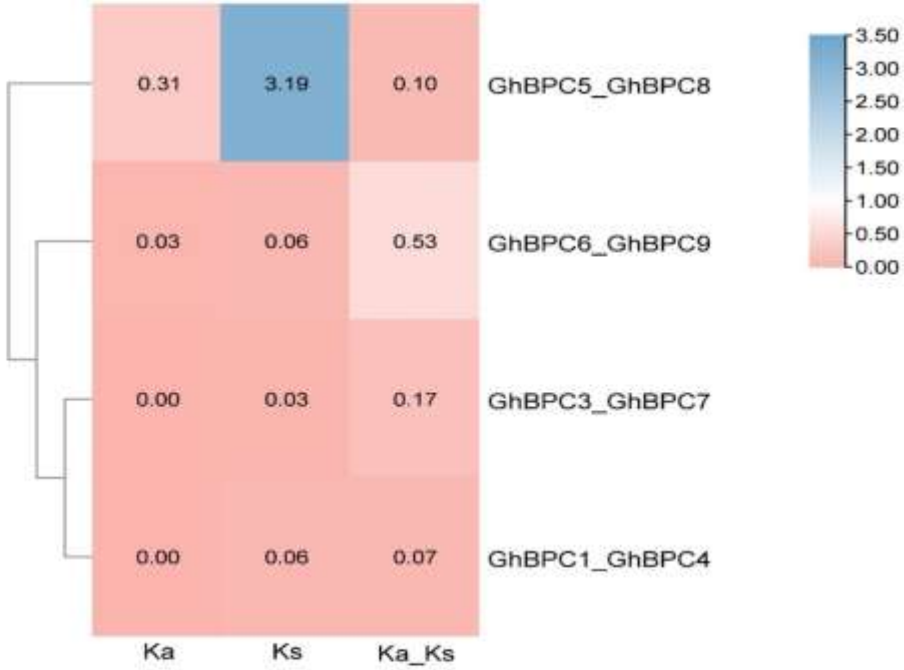

Figure 6. Ks and Ka values of GhBPC gene along with the gene pair

Synteny analysis

A circular disc was used to represent the scaffold and chromosomes in the synteny study. Total number of residing scaffolds was eight. GhBPC1 is present on D02 scaffold, while GhBPC2 is positioned on D08 scaffold, GhBPC3 is situated on A01, GhBPC4 resides on A03, GhBPC5 on A13 and GhBPC6 is present on A04. Notably, D01 has two genes including GhBPC8 and GhBPC7, while GhBPC9 was present on D04. In *GhBPC* genes, 4 paralogous gene pairs found uneven in cotton genome, which tells, these genes might be developed from segmental duplication. Scaffold linkages, showing gene pairs is as follows: A04 with D04, A13 with D01, A03 with D02, and A01 with D01. Since these linkages were due to segmental duplication, no tandem duplication is observed in GhBPC genes (Fig.8). Chromosome in A. thaliana (Chr1, Chr2 and Chr4) were orthologous to scaffold in G. hirsutum (A03, A13, D01 and D08). This demonstrates the homology of these chromosomes, which means that with evolution, they have retained a similar gene structure and pattern and have an identical ancestor chromosome. Chr1 showed linkage with A03 and D01, Chr2 was linked with (A03, A13, D01 and D08) and Chr4 was solely linked to A13 (Figure 8).

Figure 7. Synteny analysis of cotton BPC genes

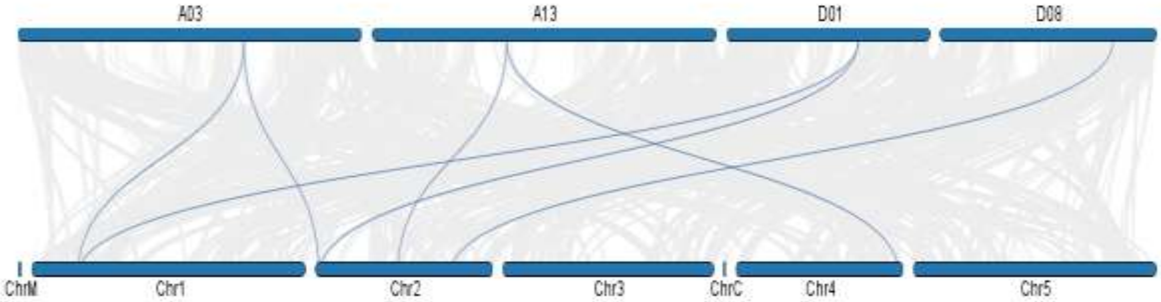

Figure 8 Synteny analysis of GhBPC family genes in Gossypium hirsutum and Arabidopsis. Blue lines show correlation between two species; the chromosome and scaffold number are shown in the figure.

Chromosomal mapping

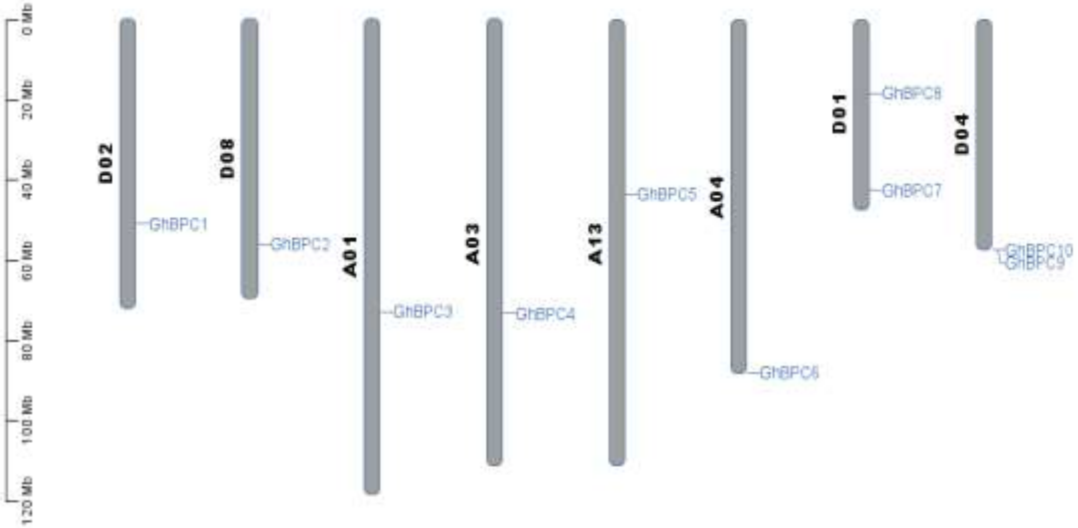

Figure 9. Location of GhBPC family genes on scaffolds of Gossypium hirsutum. Columns display scaffolds. The approximate length of the scaffold is indicated by the bar on the left, and the positions of each GhBPC gene are indicated on the chromosome

The BPC family gene chromosomal location analysis was done in order to identify the location of various genes on their respective chromosome or scaffold (Figure 9). These outcomes show that BPC family genes are dispersed at random in cotton. GhBPC1 is located on scaffold D02, GhBPC2 on scaffold D08, GhBPC3 on scaffold A01, GhBPC4 on A03, GhBPC5 on scaffold A13, GhBPC6 on scaffold A04, GhBPC7 and GhBPC8 on scaffold D01, BcBPC9 and BcBPC10 on scaffold D04.

Expression profiling of GhBPC Genes Under Cadmium Stress

Transcriptome analysis identified unique expression patterns that are tissue-specific and stress-responsive. among the GhBPC gene family under cadmium (Cd) treatment. In root (R), stem (S), and leaf (L) tissues under treatment (T) and control (C) conditions, a heatmap visualization based on normalized $\log_2$-transformed FPKM values clearly showed the clustering of genes according to their expression intensity among the ten members analysed, GhBPC1, GhBPC3, GhBPC4, and GhBPC7 exhibited consistently high expression levels across all tissues, suggesting constitutive roles in growth and stress regulation. In hierarchical analysis, these genes formed a clear high-expression cluster, suggesting comparable processes of transcriptional regulation.

Conversely, GhBPC5 and GhBPC9 were clustered separately from the highly expressed genes, indicating low or negligible expression across most tissues and treatments. This limited transcriptional activity suggests that the involvement of these genes under Cd stress is minimal. Expression of GhBPC8 was relatively elevated in root tissues with respect to stems and leaves, suggesting a potential role in root-specific cadmium response mechanisms. GhBPC2, GhBPC6, and GhBPC10, which formed intermediate clusters, displayed moderate expression patterns in the heatmap. Overall, genes with comparable expression amplitudes clustered together according to hierarchical clustering, suggesting possible functional redundancy or coordinated regulation within the GhBPC family under cadmium stress (Figure 10).

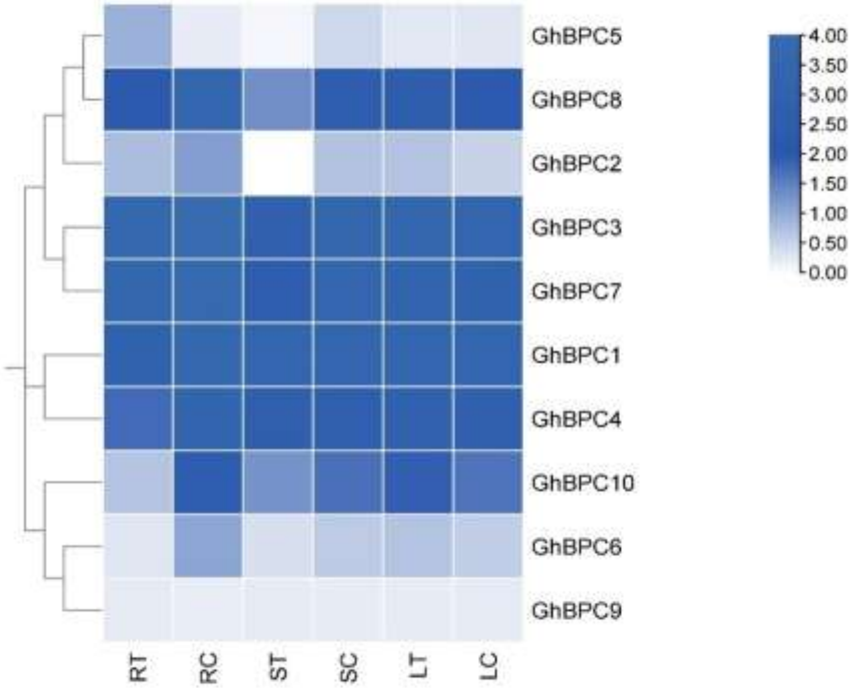

Figure 10. Expression profiling of GhBPC Genes Under Cadmium Stress

Discussion

Plants experience numerous ecological challenges throughout their lifecycle, which can significantly reduce their growth and productivity (Hecker et al., 2015; Winter et al., 2011). Transcription factors (TFs) are important regulatory molecules and are important in gene expression by regulating the gene transcription and networking. Characterization and isolation of transcription factors help us to understand the plant growth and development with respect to ambient response (Jones and Vandepoele, 2020; Wen et al., 2016; Yanagisawa and Schmidt, 1999). The BPC gene is a transcription factor and various developmental mechanism and abiotic stress responses are influenced by this gene (Ma et al., 2022; Theune et al., 2017). The identification of BPC gene has been done earlier but its classification is only observed in a few plants. Furthermore, no comprehensive genome-wide analysis of BPC gene in cotton so far.

In this research, we isolated and classified ten anticipated BPC proteins in cotton at whole-genome level, and Juxtaposed with seven *Arabidopsis* and five tomato BPC proteins. Phylogenetic analysis divide these twenty-two proteins into three groups as in Arabidopsis (Meister et al., 2004). Group I and Group II included twelve and eight members, respectively, while Group III contained only two members. Among these twelve members, seven members were of cotton, while two members each from tomato and Arabidopsis were present. Group II consisted of three members each from cotton and Arabidopsis, with two members from tomato. The smallest group (group III), incorporated two members, one from Arabidopsis and one from tomato. Arabidopsis was distributed equally between the two groups, i.e., three members in each group, while in the group three, only one member was observed. Seven out of ten GhBPC proteins were observed in group I, indicating that group II had fewer members of cotton, while in group III, only one cotton member was present. Group III had no member of cotton, prsenting that BPC genes varied before different species diverged and the members in group III could be missing in evolution. Results of gene structure and conserved motif analysis demonstrated similar exon–intron and motif organizations among the members of same group, whereas different arrangements were observed between the members of different groups. This shows a closer evolutionary association between adherants of similar group and physiological variance between members of varoius groups. Among all the BPC proteins, one or two exons were observed; one exon was present in each member of group I, with the exception of GhBPC6, which has two exons. Additionally, the GhBPC2 is very small and there is a possibility that due to some mutation the codons encoding this protein might left behind in promoter or terminator region.

Previous researches have reported that N-terminal structural domains of BPC proteins have undergone a variety of changes; these changes are anticipated to result in zipper-like coiled structures and may have a role in dimerization, protein–protein interactions, protein–promoter interactive domains, and nucleolar localization signals (Gong et al., 2018; Kooiker et al., 2005). All the *GhBPC* members have a very conserved domain at their C-terminus, with the exception of GhBPC2, that is significant for DNA binding (Meister et al., 2004; Santi et al., 2003; Wanke et al., 2011). Ten motifs were recognized in *GhBPC* proteins; in them, motifs 1 and 3 have highly conserved domains at their C-termini. Motifs 1 and 3 have five cysteine residues that control the development of structures resembling zinc fingers for the straight identification of GAGA motifs (Liu et al., 2021; Ma et al., 2022; Sangwan and O'Brian, 2002). Previously it was believed that the development of these zinc finger like structure was controlled by these cysteines for direct recognition of GAGA motifs (Meister et al., 2004). However, research of Theune et al.'s study (Theune et al., 2017) disproved the concept that the DNA-binding mechanism resembles a zinc finger. Rather, they suggested that the conserved cysteines stabilize a parallel conformation of monomers by forming intramolecular and intermolecular disulfide bonds, and this conformation is necessary for the recognition and binding of nearby GAGA motif. Like common BPC genes of different other species, many of GhBPC genes are also expressed in nucleus, which suggest their role in gene regulation, while others were expressed in different locations, according to the sub-cellular localization of GhBPC genes (Figure 5). BPC proteins in various species showed expression in the nucleus, similarly the GhBPC proteins were mostly localized in the nucleus. Expression in the nucleus suggested their role in gene regulation, however according to the subcellular localization of GhBPC genes some proteins were localized in different locations.

GhBPC1, GhBPC2, GhBPC3, GhBPC4, GhBPC5 and GhBPC8 were localized in cytoplasm which suggest that they facilitate metabolic processes, signaling, and molecular transport within the cell. While only GhBPC1 was detected in cytoskeleton which indicates its role in providing the structural support and maintaining the cell shape. GhBPC2 and GhBPC9 were localized in mitochondria gives the indication of their role in power generation for the cell. The expression of GhBPC2 and GhBPC9 in chloroplasts tells that their role in photosynthesis regulation and synthesis of chlorophyll and hence, chloroplast. GhBPC2 being the sole protein was localized in plasma membrane regulate transport of substances, facilitate cell signaling, and maintain cell structure and communication with the environment.

Enhancers and silencers, which control different growth stages and plant reactions to environmental stimuli, are incorporated into the promoter, a gene expression site. (Schmitz et al., 2022). Cis regulatory elements (CREs), which help anticipate gene function and regulate several pathways, are included in this region (Jones and Vandepoele, 2020; Xiao et al., 2017). Numerous CREs connected to various biological and metabolic processes can be found in the GhBPC gene's promoter regions. The CREs linked to light sensitivity were the most prevalent and important of all those found. Interestingly, cis regions linked to light responsiveness are present in every gene from GhBPC1 to GhBPC10. This suggests that the BPC gene family is important for both direct and indirect control of light response-related functions. Although they are not common, cis regulatory elements linked to salicylic acid responsiveness imply that the gene may contribute to stress tolerance and improve plant defence against pathogens by controlling different defence mechanisms under biotic and abiotic stress conditions. Nevertheless, only GhBPC8 exhibited this cis element. Furthermore, methyl jasmonate (MeJA) responsive elements were detected in significantly larger concentrations in GhBPC1, GhBPC2, and GhBPC4. The BPC gene family's potential involvement in plant defence systems. Pathogens, insects, salt stress, drought stress, low temperature, and heavy metal stress are just a few of the biotic and abiotic challenges that these components help plants endure. Additionally, abscisic acid (ABA)-responsive cis-regulatory elements were found in a number of genes, including GhBPC2, GhBPC5, GhBPC6, GhBPC9, and GhBPC10. These elements may be involved in plant defence against abiotic stressors because they are found in promoter regions of BPC genes. They are particularly linked to the control of osmotic stress, excessive salinity, and drought. Furthermore, only GhBPC8 exhibited MYBHV1 binding sites, which are likewise linked to stress response. The relevance of BPC genes in drought stress tolerance is suggested by the presence of cis-regulatory regions linked to drought responsiveness. Furthermore, defence, growth, and pigmentation depend on plant secondary metabolites known as flavonoids. The promoter regions of BPC gene family contained the CREs that control the genes involved in flavonoid biosynthesis, suggesting their possible participation in flavonoid-related pathways. The involvement of BPC in anaerobic induction is suggested by the existence of anaerobic sensitive components. Furthermore, only GhBPC4 exhibited low-temperature responsive cis-regulatory regions, indicating a restricted but potential role for BPC genes in the direct or indirect control of cold stress adaption mechanisms. All of these cis regulatory elements may be involved in controlling how plants react to different biotic and abiotic stressors because they are present in promoter regions of BPC genes.

The promoter regions of BPC genes were also have cis regulatory elements linked to development and growth of plants. These include components that are controlled by the amount of particular plant hormones. In this work, we found auxin-responsive cis elements (AuxREs) in GhBPC gene promoter regions. Those components reveal the function of BPC genes in developmental processes via controlling gene expression in response to auxin levels. The BPC gene's involvement in auxin-mediated processes like cell elongation, apical dominance, root initiation, vascular tissue development, and tropisms is indicated by the presence of (AuxREs). Additionally, Gibberellin-responsive components were found in the GhBPC2 promoter. The BPC gene may play a part in gibberellin-mediated activities like flowering, stem elongation, and seed germination, as these particular DNA sequences control gene expression in response to gibberellin levels.

There were a few endosperm-responsive cis-regulatory elements. These components increase the transcription of genes necessary for endosperm growth and the synthesis of seed storage proteins by acting as binding sites for transcription factors that are only active in the endosperm. This implies that the BPC gene may play a part in the formation of endosperm in seeds. These components were found in GhBPC4, GhBPC5, and GhBPC6 in cotton. A few genes also have cis-regulatory regions that control meristem expression, tell these genes has a role in plant development.

Gene duplication can be predicted based on their location on chromosomes. This indicates that tandem duplication happens when two or more genes are present on similar chromosome, while duplicated genes on different chromosomes may result from segmental duplication (Panchy et al., 2016). Scaffold linkages showing gene pairs are as follows: A04 with D04, A13 with D01, A03 with D02, and A01 with D01. Only segmental duplication was observed, with no proof of tandem duplication. GhBPC orthologs were dispersed among several Arabidopsis chromosomes, according to collinearity evaluations. A03 and D01 were related to Chr1, A03, A13, D01, and D08 were linked to Chr2, and only A13 was associated to Chr4. While Ks indicates the amount of synonymous substitutions per synonymous site, Ka indicates the number of nonsynonymous substitutions per nonsynonymous site. Ka/KS is the ratio of nonsynonymous to synonymous substitutions. The Ka/Ks ratio indicates the level of selective pressure. Purifying selection on a gene pair means that natural selection has removed it because of its negative effects. Positively chosen gene pairs, while, are those which helped two duplicates thrive. When the Ka/Ks value is less than one, purifying selection takes place; neutral selection is indicated when the Ka/Ks value is equal to one. On the other hand, a Ka/Ks number larger than one indicates positive selection. Thus, the ratios of Ka to Ks in cotton BPC genes were computed. GhBPC1_GhBPC4 pair's ratio revealed the lowest of all, equal to 0.069143375. In contrast, the GhBPC6_GhBPC9 combination showed the highest ratio among GhBPC genes, 0.534614031. Every cotton BPC gene pair has a value that is smaller than the purifying selection.

Chromosome mapping shows that the BPC genes in cotton are haphazardly distributed across numerous chromosomes and scaffolds. For example, GhBPC1 is on D02, GhBPC2 on D08, GhBPC3 on A01, GhBPC4 on A03, GhBPC5 on A13, GhBPC6 on A04, GhBPC7 and GhBPC8 on D01, and GhBPC9 and GhBPC10 on D04. This random distribution suggests that the genes may have a diversity of functions and may have evolved through duplication or rearrangement processes.

Understanding the expression patterns of GhBPC gene family under cadmium (Cd) stress in various cotton tissues was made possible using transcriptome analysis. Clear tissue-specific and stress-responsive expression mechanism among the examined genes were shown in the heatmap produced from normalized $\log_2$-transformed FPKM values. Differential transcriptional regulation of GhBPC genes in response to cadmium stress is suggested by groupings of genes based on the extent of their expression. In both treatment and control circumstances, GhBPC1, GhBPC3, GhBPC4, and GhBPC7 consistently exhibited high expression levels in root, stem, and leaf tissues. These genes may contribute to stress response systems and play constitutive roles in plant growth and developmental processes, based on their steady and increased expression. Further evidence that these genes may engage in integrated transcriptional networks or share comparable regulatory activities comes from their grouping into a separate high-expression cluster in hierarchical clustering analysis. GhBPC5 and GhBPC9, in contrast, presents extremely low expression levels in the majority of tissues and circumstances. These genes may have a small or highly specialized role under cadmium stress, as indicated by their separation from highly expressed genes in the clustering analysis, which signals restricted transcriptional activity. Additionally, their expression might only be triggered in particular developmental phases or environmental circumstances that aren't included in the current dataset. It's significant to note that GhBPC8 expression was comparatively higher in root tissues than in stems and leaves. The increased expression of GhBPC8 in root tissues indicates a possible role in root-specific defense or detoxification processes under heavy metal stress, since roots are the main location of cadmium absorption and storage in plants. The potential functional heterogeneity within the GhBPC gene family is highlighted by such tissue-specific expression variations. Additionally, in the heatmap analysis, GhBPC2, GhBPC6, and GhBPC10 formed intermediate clusters and showed moderate expression levels. These genes may contribute to adaptive physiological responses by taking part in regulatory networks that are somewhat susceptible to cadmium exposure.

Overall, the GhBPC gene family may have coordinated regulation or functional redundancy based on the grouping of genes with comparable expression patterns. These findings tell that several GhBPC genes may be helpful in transcriptional control and stress tolerance in cotton under cadmium stress. To fully comprehend the precise biological involvement of these genes in heavy metal stress tolerance, more functional validation work—such as gene knockout or overexpression analyses—are necessary.

References

Bailey, T. L., Johnson, J., Grant, C. E., and Noble, W. S. Suite TMEME (2015). *The MEME Suite. Nucleic Acids Res* **43**, W39-W49. https://doi.org/10.1093/nar/gkv416

Bailey, T. L., Johnson, J., Grant, C. E., and Noble, W. S. (2015). The MEME suite. *Nucleic acids research* **43**, W39-W49. https://doi.org/10.1093/nar/gkv416

Berger, N., Dubreucq, B., Roudier, F., Dubos, C., and Lepiniec, L. (2011). Transcriptional regulation of Arabidopsis Leafy cotyledon2 involves RLE, a cis-element that regulates trimethylation of histone H3 at lysine-27. *The Plant Cell* **23**, 4065-4078. https://doi.org/10.1105/tpc.111.088377

Bettaieb, I., and Bouktila, D. (2020). Genome-wide analysis of NBS-encoding resistance genes in the Mediterranean olive tree (Olea europaea subsp. europaea var. europaea): Insights into their molecular diversity, evolution and function. *Tree Genetics & Genomes* **16**, 23.

Chen, C., Chen, H., Zhang, Y., Thomas, H. R., Frank, M. H., He, Y., and Xia, R. (2020). TBtools: an integrative toolkit developed for interactive analyses of big biological data. *Molecular plant* **13**, 1194-1202. https://doi.org/10.1016/j.molp.2020.06.009

Ezquer, I., Mizzotti, C., Nguema-Ona, E., Gotté, M., Beauzamy, L., Viana, V. E., Dubrulle, N., Costa de Oliveira, A., Caporali, E., and Koroney, A.-S. (2016). The developmental regulator seedstick controls structural and mechanical properties of the Arabidopsis seed coat. *The Plant Cell* **28**, 2478-2492.

Gong, R., Cao, H., Zhang, J., Xie, K., Wang, D., and Yu, S. (2018). Divergent functions of the GAGA-binding transcription factor family in rice. *The Plant Journal* **94**, 32-47.

Haigler, C. H., Betancur, L., Stiff, M. R., and Tuttle, J. R. (2012). Cotton fiber: a powerful single-cell model for cell wall and cellulose research. *Frontiers in plant science* **3**, 104. https://doi.org/10.3389/fpls.2012.00104

Hecker, A., Brand, L. H., Peter, S., Simoncello, N., Kilian, J., Harter, K., Gaudin, V., and Wanke, D. (2015). The Arabidopsis GAGA-binding factor basic pentacysteine6 recruits the polycomb-repressive complex1 component like heterochromatin protein1 to GAGA DNA motifs. *Plant physiology* **168**, 1013-1024.

Hu, Y., Chen, J., Fang, L., Zhang, Z., Ma, W., Niu, Y., Ju, L., Deng, J., Zhao, T., and Lian, J. (2019). Gossypium barbadense and Gossypium hirsutum genomes provide insights into the origin and evolution of allotetraploid cotton. *Nature genetics* **51**, 739-748. https://doi.org/10.1038/s41588-019-0371-5

Hu, Z., Fu, Q., Zheng, J., Zhang, A., and Wang, H. (2020). Transcriptomic and metabolomic analyses reveal that melatonin promotes melon root development under copper stress by inhibiting jasmonic acid biosynthesis. *Horticulture Research* **7**.

Jones, D. M., and Vandepoele, K. (2020). Identification and evolution of gene regulatory networks: insights from comparative studies in plants. *Current opinion in plant biology* **54**, 42-48. https://doi.org/10.1016/j.pbi.2020.03.007

Kooiker, M., Airoldi, C. A., Losa, A., Manzotti, P. S., Finzi, L., Kater, M. M., and Colombo, L. (2005). Basic pentacysteine1, a GA binding protein that induces conformational changes in the regulatory region of the homeotic Arabidopsis gene seedstick. *The Plant Cell* **17**, 722-729.

Li, S., Sun, M., Miao, L., Di, Q., Lv, L., Yu, X., Yan, Y., He, C., Wang, J., and Shi, A. (2023). Multifaceted regulatory functions of CsBPC2 in cucumber under salt stress conditions. *Horticulture Research* **10**, uhad051. https://doi.org/10.1093/hr/uhad051

Liu, B., Sun, G., Liu, C., and Liu, S. (2021). Leafy cotyledon 2: A regulatory factor of plant growth and seed development. *Genes* **12**, 1896. https://doi.org/10.3390/genes12121896

Ma, X., Yu, Y., Hu, Z., Huang, H., Li, S., and Yin, H. (2022). Characterizations of a Class-I BASIC pentacysteine gene reveal conserved roles in the transcriptional repression of genes involved in seed development. *Current Issues in Molecular Biology* **44**, 4059-4069.

Meister, R. J., Williams, L. A., Monfared, M. M., Gallagher, T. L., Kraft, E. A., Nelson, C. G., and Gasser, C. S. (2004). Definition and interactions of a positive regulatory element of the *Arabidopsis* inner no outer promoter. *The Plant Journal* **37**, 426-438.

Meng, D., Li, S., Feng, X., Di, Q., Zhou, M., Yu, X., He, C., Yan, Y., Wang, J., and Sun, M. (2023). CsBPC2 is essential for cucumber survival under cold stress. *BMC Plant Biology* **23**, 566.

Panchy, N., Lehti-Shiu, M., and Shiu, S.-H. (2016). Evolution of gene duplication in plants. *Plant physiology* **171**, 2294-2316. https://doi.org/10.1104/pp.16.00523

Petrella, R., Caselli, F., Roig-Villanova, I., Vignati, V., Chiara, M., Ezquer, I., Tadini, L., Kater, M. M., and Gregis, V. (2020). BPC transcription factors and a Polycomb Group protein confine the expression of the ovule identity gene seedstick in Arabidopsis. *The Plant Journal* **102**, 582-599.

Sahu, A., Singh, R., and Verma, P. K. (2023). Plant BBR/BPC transcription factors: unlocking multilayered regulation in development, stress and immunity. *Planta* **258**, 31.

Sangwan, I., and O'Brian, M. R. (2002). Identification of a soybean protein that interacts with GAGA element dinucleotide repeat DNA. *Plant Physiology* **129**, 1788-1794.

Santi, L., Wang, Y., Stile, M. R., Berendzen, K., Wanke, D., Roig, C., Pozzi, C., Müller, K., Müller, J., and Rohde, W. (2003). The GA octodinucleotide repeat binding factor BBR participates in the transcriptional regulation of the homeobox gene Bkn3. *The Plant Journal* **34**, 813-826. https://doi.org/10.1046/j.1365-313X.2003.01712.x

Schmitz, R. J., Grotewold, E., and Stam, M. (2022). Cis-regulatory sequences in plants: Their importance, discovery, and future challenges. *The plant cell* **34**, 718-741.

Theune, M. L., Hummel, S., Jaspert, N., Lafos, M., and Wanke, D. (2017). Dimerization of the Basic pentacysteine domain in plant GAGA-factors is mediated by disulfide bonds and required for DNA-binding. *Journal of Advances in Plant Biology* **1**, 26-39.

Wanke, D., Hohenstatt, M. L., Dynowski, M., Bloss, U., Hecker, A., Elgass, K., Hummel, S., Hahn, A., Caesar, K., and Schleifenbaum, F. (2011). Alanine zipper-like coiled-coil domains are necessary for homotypic dimerization of plant GAGA-factors in the nucleus and nucleolus. *PLoS One* **6**, e16070. https://doi.org/10.1371/journal.pone.0016070

Wen, C.-l., Cheng, Q., Zhao, L., Mao, A., Yang, J., Yu, S., Weng, Y., and Xu, Y. (2016). Identification and characterisation of Dof transcription factors in the cucumber genome. *Scientific reports* **6**, 23072.

Wendel, J., and Cronn, R. (2003). Polyploidy and the evolutionary history of cotton. *Advances in Agronomy* **78** (2003): 139-186.

Winter, C. M., Austin, R. S., Blanvillain-Baufumé, S., Reback, M. A., Monniaux, M., Wu, M.-F., Sang, Y., Yamaguchi, A., Yamaguchi, N., and Parker, J. E. (2011). Leafy target genes reveal floral regulatory logic, cis motifs, and a link to biotic stimulus response. *Developmental cell* **20**, 430-443.

Wu, A., Hao, P., Wei, H., Sun, H., Cheng, S., Chen, P., Ma, Q., Gu, L., Zhang, M., and Wang, H. (2019). Genome-wide identification and characterization of glycosyltransferase family 47 in cotton. *Frontiers in genetics* **10**, 824. https://doi.org/10.3389/fgene.2019.00824

Xian, J., Wang, Y., Niu, K., Ma, H., and Ma, X. (2020). Transcriptional regulation and expression network responding to cadmium stress in a Cd-tolerant perennial grass Poa Pratensis. *Chemosphere* **250**, 126158. https://doi.org/10.1016/j.chemosphere.2020.126158

Xiao, J., Jin, R., Yu, X., Shen, M., Wagner, J. D., Pai, A., Song, C., Zhuang, M., Klasfeld, S., and He, C. (2017). Cis and trans determinants of epigenetic silencing by Polycomb repressive complex 2 in Arabidopsis. *Nature genetics* **49**, 1546-1552. https://doi.org/10.1038/ng.3953

Yanagisawa, S., and Schmidt, R. J. (1999). Diversity and similarity among recognition sequences of Dof transcription factors. *The Plant Journal* **17**, 209-214. https://doi.org/10.1046/j.1365-313X.1999.00381.x

Zhang, S., Wang, J., Feng, Y., Xue, Y., Wang, Y., Zhao, M., Chen, M., Chen, C., Su, W., and Chen, R. (2023). Genome-wide identification of flowering Chinese cabbage BPC family genes and BcBPC9 functional analysis in Cd stress tolerance. *Plant Stress* **10**, 100220. doi.org/10.1016/j.stress.2023.100220

Statements and Declarations

Authors' Contribution

Maryam Nadeem Organized this research and did all the analysis, Jawad Ahmad and Syed Haseeb Ul Hassan Sherazi contributed in write up and editing, while Shahid A, Tahir K and Ahmed H did the review. All authors have read the final manuscript and approve its submission.

Acknowledgements

The authors like to convey their heartfelt gratitude to the University for providing the essential facilities and resources.

Conflict of Interest

It is claimed that the authors have no known conflicts of interest for the publication of this work.

Ethical Responsibility

This manuscript is original research, and it is not submitted in whole or in part to another journal for publication.

Data Availability Statement

This study did not generate new datasets. All data used in this investigation were obtained from publically available genetic databases, as mentioned in the report. The all data generated or processed during this study, including protein sequences and accession numbers present in manuscript file.

Funding

Not applicable