

RESEARCH

Open Access



Phylogenomic analysis of bromodomain genes in cotton (*Gossypium* spp.) and their potential roles in abiotic stress tolerance

TAYYAB Muhammad¹, TAJ Muhammad Hassan¹, AHMAD Ali Ijaz², RANA Faiza Rehman², SHAHZAIB Muhammad^{6,7,8}, ATIF Rana Muhammad^{2,3}, AZHAR Muhammad Tehseen^{2,4}, KHAN Sultan Habibullah^{1,3}, KAKAKHEL Ishaq Ahmad Mian⁵, HE Shoupu^{9,10} and RANA Iqrar Ahmad^{1,3*} 

Abstract

Background The bromodomain (BRD) proteins play a pivotal role in regulating gene expression by recognizing acetylated lysine residues and acting as chromatin-associated post-translational modification-inducing proteins. Although BRD proteins have been extensively studied in mammals, they have also been characterized in plants like *Arabidopsis thaliana* and *Oryza sativa*, where they regulate stress-responsive genes related to drought, salinity, and cold. However, their roles in cotton species remain unexplored.

Results In this genome-wide comparative analysis, 145 *BRD* genes were identified in the tetraploid species (*Gossypium hirsutum* and *G. barbadense*), compared with 82 *BRD* genes in their diploid progenitors (*G. arboreum* and *G. raimondii*), indicating that polyploidization significantly influenced *BRD* gene evolution. Gene duplication analysis revealed 78.85% of duplications were segmental and 21.15% were tandem among 104 in-paralogous gene pairs, contributing to *BRD* gene expansion. Gene structure, motif, and domain analyses demonstrated that most genes were intron-less and conserved throughout evolution. Syntenic analysis revealed a greater number of orthologous gene pairs in the Dt sub-genome than in the At sub-genome. The abundance of regulatory, hormonal, and defense-related *cis*-regulatory elements in the promoter region suggests that *BRD* genes play a role in both biotic and abiotic stress responses. Protein-protein interaction analysis indicated that global transcription factor group E (GTE) transcription factors regulate *BRD* genes. Expression analysis revealed that *BRD* genes are predominantly involved in ovule development, with some genes displaying specific expression patterns under heat, cold, and salt stress. Furthermore, qRT-PCR analysis demonstrated significant differential expression of *BRD* genes between the tolerant and sensitive genotype, underscoring their potential role in mediating drought and salinity stress responses.

Conclusions This study provides valuable insights into the evolution of *BRD* genes across species and their roles in abiotic stress tolerance, highlighting their potential in breeding programs to develop drought and salinity tolerant cotton varieties.

Keywords Cotton, Phylogenetics, Epigenetics, Bromodomain

*Correspondence:
Rana Iqrar Ahmad
iqrar_rana@uaf.edu.pk
Full list of author information is available at the end of the article



© The Author(s) 2025. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

Introduction

Cotton, which belongs to the genus *Gossypium* in the family Malvaceae, is the most important fiber crop grown in more than 50 countries. It comprises 52 known species, including seven tetraploids ($2n = 4x = 52$) and 45 diploids ($2n = 2x = 26$). Only four species have been domesticated for commercial production, with two diploids (*G. arboreum* and *G. herbaceum*) and two allotetraploids (*G. hirsutum* and *G. barbadense*) (He et al. 2020). *G. hirsutum* dominates cultivation, accounting for around 90% of production, while *G. barbadense* accounts for approximately 8%, with the remaining 2% coming from other species, primarily grown in Pakistan, China, India, Iran, and Bangladesh (Shim et al. 2018). Annual cotton production is 118.37 million tons, cultivated over an area of 33.19 million hectares (Sain et al. 2023). The annual worth of raw cotton products is USD 1 000 billion, with a projected growth of 4% by 2030. Additionally, cotton provides livelihoods to 100 million households (Voora et al. 2023). However, abiotic stresses such as drought, salinity, and temperature extremes can disrupt osmotic balance, damage membrane integrity, and impair photosynthesis, thereby elevating levels of reactive oxygen species (ROS) and drastically reducing productivity (Kamburova et al. 2022). Drought and salinity are the most significant factors causing yield losses up to 50%, and compromising fiber quality (Noreen et al. 2020). Therefore, it is crucial to study abiotic stress tolerance in cotton to minimize its stress-induced damage and yield loss.

Gene expression is highly complex, and its response to different stresses is regulated at multiple levels (De Nadal et al. 2011). Stress conditions trigger intricate transcriptional reprogramming that generates adaptive responses to mitigate negative impacts (Golldack et al. 2011). Both genetic components (*cis*-elements, enhancers, silencers, and core promoters) and epigenetic components (histone modifications, DNA methylation, and various RNA-mediated processes) influence gene expression (Abiraami et al. 2023; Biřas et al. 2016; Gibney et al. 2010). Several cellular proteins, such as histones, are affected by post-translational modifications (PTMs) like phosphorylation, acetylation, methylation, and ubiquitylation (Millán-Zambrano et al. 2022). Cellular epigenetic control is regulated by proteins that add, remove, and recognize acetyl groups on histones, thereby maintaining epigenetic marks. Lysine acetylation is controlled by acetyltransferases, which add acetyl groups, and deacetylases, which remove them, while bromodomains (BRD) recognize acetylated lysines on histones (Abiraami et al. 2023; Dan et al. 2023).

Drosophila melanogaster was the first organism in which bromodomain-containing genes were discovered. These genes have since been identified in many eukaryotes,

including *Arabidopsis thaliana* and *Oryza sativa* (Abiraami et al. 2023; Tamkun et al. 1992). The distinctive protein structure of the approximately 110 amino acids BRD domain consists of four helices (Z, A, B, and C) connected by two loops (ZA and BC) (Mujtaba et al. 2007). These helices form a conserved hydrophobic pocket that recognizes acetylated lysine residues (Kac) in histones, using a conserved asparagine residue (Bardani et al. 2023). BRD-containing proteins regulate gene expression in several ways. They can be part of chromatin-remodeling complexes that modulate the chromatin structure, stimulate histone PTMs such as activating histone acetyltransferases (HATs), or serve as scaffolds that facilitate the recognition of acetylated histones and the recruitment of transcriptional regulators. Additionally, they function as transcriptional co-regulators, acting as either activators or repressors. These findings underscore their significance in epigenetic regulation and transcriptional control (Bardani et al. 2023; Josling et al. 2012; Marmorstein et al. 2001).

In this study, we identified multiple BRD gene family members in cotton species, including 40 genes from *G. arboreum*, 42 from *G. raimondii*, 66 from *G. barbadense*, and 79 from *G. hirsutum*. By employing diverse bioinformatics techniques, we clarified their replication mechanisms, reconstructed their phylogenetic relationships, evaluated their structural arrangement, appraised protein-protein interactions (PPI), explored promoter binding elements, and analyzed their syntenic associations across evolutionary stages. These in-silico analyses provided comprehensive insights into their functional roles and evolutionary dynamics. To further understand their functional roles, we conducted expression profiling analyses using multiple RNA-seq datasets and qRT-PCR assays. These analyses evaluated responses to abiotic stressors (cold, heat, drought, and salt), examined tissue-specific expression patterns, assessed potential involvement in ovule development, and explored their roles in drought tolerance.

Material and methods

Database mining BRD gene members

Reference protein sequences of BRD gene family members of *A. thaliana* and *O. sativa* were retrieved from the Arabidopsis Information Resource (TAIR) version Araport11 (<https://www.arabidopsis.org/>). For rice sequences, Rice Genome Annotation Project (RGAP) release 7 (<http://rice.uga.edu/cgi-bin/gbrowse/rice/>) was used (Kawahara et al. 2013; Reiser et al. 2024).

BRD-containing gene/protein sequences from four *Gossypium* species, viz. *G. arboreum*, *G. barbadense*, *G. hirsutum*, and *G. raimondii*, peptide sequences, and general feature format (GFF) files were retrieved from the CottonGen database (<https://www.cottongen.org/>)

and the Phytozome database (<https://phytozome-next.jgi.doe.gov/>) (Goodstein et al. 2012; Yu et al. 2021). All BRD protein members were identified using TBtools-II v2.119 (<https://github.com/CJ-Chen/TBtools-II>) with the BLASTp algorithm (e-value $< 1e^{-5}$) and default parameters (Chen et al. 2023; Mahram et al. 2015). Identified BRD protein sequences were then evaluated for the presence of complete functional domains using the Batch Web CD-Search Tool from the NCBI Conserved Domain Database (<https://www.ncbi.nlm.nih.gov/Structure/bwrpsb/bwrpsb.cgi>), and candidates with incomplete or irrelevant domains were eliminated (Wang et al. 2023). Furthermore, BRD protein sequences were confirmed through the InterPro-EMBL-EBI database v101.0 (<https://www.ebi.ac.uk/interpro/>), ensuring the proper length and the presence of functional, and conserved domains (Paysan-Lafosse et al. 2023).

Physicochemical properties estimation and chromosomal localization of *BRD* genes

Physicochemical properties, specifically isoelectric point (pI) and molecular weight (MW), of BRD proteins were estimated using the ExPASy Compute pI/Mw tool from the Swiss Institute of Bioinformatics (SIB) (<https://www.expasy.org/resources/compute-pi-mw>) (Gasteiger et al. 2003). Subcellular localization of BRD proteins was predicted using the Bologna Unified Subcellular Component Annotator (BUSCA) tool (<https://busca.biocomp.unibo.it/>) (Savojardo et al. 2018). Information regarding chromosome numbers, gene and strand positions, as well as gene and CDS lengths, was retrieved from GFF files. The graphical representation of *BRD* genes chromosomal distribution was generated using MG2C v2.1 (http://mg2c.iask.in/mg2c_v2.1/index.html) (Chao et al. 2021).

Multiple sequence alignment and phylogenetic analysis

All identified BRD proteins were aligned using the web-based multiple sequence alignment tool Clustal Omega v1.2.4 (<https://www.ebi.ac.uk/jdispatcher/msa/clustalo>) with the output Pearson/FASTA format at default parameters (Harrison et al. 2024; Sievers et al. 2014). Molecular Evolutionary Genetics Analysis (MEGA) v11.0.4 was used to study the phylogenetic relationships among the aligned *BRD* genes. A Neighbor-joining phylogenetic tree was constructed using the Jones-Taylor-Thornton (JTT) substitution model, suitable for amino acid sequences. The phylogenetic tree was assessed using a 1 000-bootstrap confidence interval. Sequence variation rates and parameters were modeled using a Gamma distribution, with a shape parameter value of 5.00 (Jones et al. 1992; Saitou et al. 1987; Tamura et al. 2021). The identified gaps and the missing data underwent a process of partial deletion with a site coverage threshold set at 95%. The visual depiction of the phylogenetic relationship

tree was generated utilizing the online tool for phylogenetic tree annotation and visualization known as iTOL v6.9.1 (<https://itol.embl.de/>) (Letunic et al. 2024).

Gene structure, conserved domain, and motif analysis of *BRD* genes

Intron–exon distribution of *BRD* genes was predicted using the ‘Gene Structure View (Advanced)’ function in TBtools-II software and the conserved *BRD* family and superfamily domain presence in proteins was analyzed by conserved domain (CD)-NCBI web tool (Chen et al. 2023). To identify the novel conserved motifs present within the amino acid sequences, the MEME v5.5.6 software (<https://meme-suite.org/meme/tools/meme>), part of the MEME suite, was employed. All parameters within MEME were kept at default settings, with the exception that the minimum number of motifs was increased to 100, and the minimum motif width was increased to 12 (Bailey et al. 2015; Bailey et al. 2006).

Multiple synteny and gene divergence analysis

Multiple synteny analyses were performed to study the evolutionary relationship of conserved gene groups of the *BRD* gene family across cotton species (Zhao et al. 2017). To study the inter-chromosomal collinearity among the diploid and tetraploid cotton species, whole genome FASTA and annotated GFF files were analyzed using the One step MCSanX function in TBtools-II, with a threshold e-value $< 1e^{-10}$ (Wang et al. 2012). A graphical illustration of synteny analysis was generated using the multiple synteny plot function in TBtools-II.

The evolution of gene families involves tandem and segmental duplications (Cannon et al. 2004). Gene pairs that are physically separated by less than 50 kb and are placed on the same chromosome are referred to as tandemly duplicated, whereas pairings that are positioned on different chromosomes are referred to as segmentally duplicated (Panchy et al. 2016). Amino acid alterations resulting from the nucleotide substitutions are referred to as non-synonymous (K_a) substitutions, while synonymous (K_s) substitutions do not result in amino acid modifications (Yang et al. 2000). The ratio of K_a to K_s can be used to indicate the type and degree of selection pressure acting on coding sequences. The K_s value can also be used to determine the mode of selection and divergence time (Anisimova et al. 2001). Adaptive/positive selection is indicated by a value greater than one ($K_a/K_s > 1$), purifying/negative selection is shown by a value less than one ($K_a/K_s < 1$), and neutral selection is indicated by a value equal to one ($K_a/K_s = 1$) (Doerks et al. 2002). Simple K_a/K_s calculator software in TBtools-II was used to calculate the ratios using CDS and gene pairs as input.

Promoter and orthologous gene cluster analysis

In environmental circumstances, promoter regions consist of *cis*-regulatory elements are crucial for regulating gene expression (Wittkopp et al. 2012). The 2 000-bp upstream genomic region of *BRD* genes was analyzed in the PlantCARE database (<https://bioinformatics.psb.ugent.be/webtools/plantcare/html/>) to curate and analyze the *cis*-regulatory elements (Lescot et al. 2002). Moreover, a comparative analysis of orthologous clusters was conducted utilizing the web-based OrthoVenn3 (<https://orthovenn3.bioinfotoolkits.net/start/db>) tool (Sun et al. 2023). The default e-value of $1e^{-2}$ was employed as a parameter for significance threshold, in conjunction with the Inflation value of 1.5 as determined by the Markov Cluster (MCL) algorithm process (Emmons et al. 2016). The representations of shared orthologous clusters were generated in accordance with John Venn's methodology for the mechanical representation of propositions (MacColl 1880).

Protein–protein interaction and secondary structure analysis

Protein's secondary structure predicts the tertiary and quaternary 3D structure to understand their function, dynamic nature, and biological processes. Alpha helices, beta sheets (strands), transmembrane, and disordered structures are the basic components of protein secondary structure (Aydin et al. 2006). All curated protein sequences of *BRD* genes of cotton species were input on the web-based server STRING database of protein–protein interaction networks (<https://string-db.org/>) to interpret the functions based on interacting proteins (Szklarczyk et al. 2021). The peptide sequences were submitted to the Phyre2 (<http://www.sbg.bio.ic.ac.uk/phyre2/html/page.cgi?id=index>) computational platform, and the secondary structural elements were anticipated through homology modeling (Kelley et al. 2015).

Global transcriptome profiling of *BRD* genes

To study the expression profiling of *BRD* genes in different tissues under normal and stressful conditions, RNA-seq data were fetched from the ENA-EBI database (<https://www.ebi.ac.uk/ena/browser/home>) (Yuan et al. 2024). NCBI-Bioproject (PRJNA594268 and PRJNA490626) were used to retrieve the tissue specific expression data of *G. arboreum* cultivar Shixiya1 (A2), *G. hirsutum* TM-1, and *G. barbadense* Hai7124, respectively (Hu et al. 2019; Huang et al. 2020). The RNA-seq data for the *G. raimondii* tissue expression were retrieved from the NCBI-Bioproject (PRJNA179447, PRJNA171342, PRJNA1128012, PRJNA12541, and PRJNA601953). Raw RNA-seq data were analyzed through the online web-based server Galaxy v24.1.2.dev0 (<https://usegalaxy.org/>)

(Chappell et al. 2022). Then transcript was quantified into transcript per million (TPM) values done with Sailfish v0.10.1.1 on Galaxy (https://usegalaxy.org/?tool_id=tools.hed.g2.bx.psu.edu%2Frepos%2Fbgruening%2Fsailfish%2Fsailfish%2F0.10.1.1&version=latest) (Rao et al. 2012).

qRT-PCR expression analysis of *GhBRD* genes under drought and salinity stress

To assess the impact of drought stress on gene expression, qRT-PCR was performed on 13 curated genes in six cotton cultivars, viz. K119 (Yan 64-1), K194 (Ganza0 032), and K279 (n74-250) (drought-sensitive) and K254 (Jifeng 908), K331 (Ari1327), and K354 (Zhong 1441) (drought-resistant). The analysis was conducted at control 1 (0 day), control 2 (4 days), control 3 (8 days), drought 1 (0 day), drought 2 (4 days), and drought 3 (8 days) time points. To study the response to salt stress, two genotypes viz. Z9807 (salt-tolerant) and ZJ0102 (salt-sensitive) were selected. Seedlings were exposed to the salt concentration of 0.4% (4 g NaCl : 1000 g Sands) at time periods of 3 h, 12 h, 24 h, and 48 h. True leaves were collected from seedlings to isolate RNA and perform qRT-PCR expression analysis. cDNA was synthesized using the First-strand cDNA Synthesis Kit (Novoprotein, Shanghai, China), following the manufacturer's protocol. Gene-specific primers were designed using Primer-BLAST, and the housekeeping gene *GhHistone* was used for normalization. The qRT-PCR reactions were performed on a Light-Cycler 480 (Roche, Mannheim, Germany) using SYBR Green qPCR SuperMix Plus (Novoprotein, Shanghai, China). Three biological replicates were analyzed for each sample, with each biological replicate having three technical replicates. Gene expression levels were calculated using the $2^{-\Delta\Delta C_t}$ method to determine relative expression values (Schmittgen et al. 2008). This approach enabled a comparison of gene expressions between the control and drought-stressed cotton varieties at each time point.

Results

Identification of *BRD* genes in cotton species

Initially, a total of 302, 282, 523, and 553 genes were extracted from *G. arboreum*, *G. raimondii*, *G. barbadense*, and *G. hirsutum*, respectively, using *BRD* gene query sequences from *A. thaliana* and *O. sativa*. After excluding genes without a complete bromodomain, 227 redundant *BRD* genes with functional domains were finalized, comprising 40 from *G. arboreum*, 42 from *G. raimondii*, 66 from *G. barbadense*, and 79 from *G. hirsutum* (Supplementary Table S1). Retrieved genes were renamed, i.e., *GaBRD*-01 to *GaBRD*-40, *GrBRD*-01 to *GrBRD*-42, *GbBRD*-01 to *GbBRD*-66, and *GhBRD*-01 to *GhBRD*-79. A higher number of *BRD* genes in tetraploid species resulted from the event of whole genome

duplication, which allows duplicated genes to adopt new functions.

Physicochemical properties and subcellular localization

The lengths of *BRD* gene members varied significantly among *Gossypium* species. In *G. arboreum*, the gene lengths ranged from 956 bp to 17.79 kb (with an average of 5.43 kb), and the protein lengths ranged from 295 to 2 311 amino acids (aa) with an average of 857.02 aa. Similarly, the lengths of genes in *G. raimondii* ranged from 1.43 kb to 18.72 kb (average 6.57 kb), whereas protein length ranged from 367 to 2 259 aa (average 877.36 aa). The range lengths of genes among *G. barbadense* ranged from 1.30 kb to 21.87 kb (average 5.68 kb), while protein lengths ranged from 246 to 2 677 aa (average 779.76 aa). *G. hirsutum* gene lengths were between 413 bp and 18.67 kb (average 5.84 kb) with protein lengths ranging from 137 to 2 250 aa (average 805.19 aa).

The pI of BRD proteins ranged from 5.07 to 9.89 in *G. arboreum*, 4.98 to 9.79 in *G. raimondii*, 4.86 to 9.97 in *G. barbadense*, and 5.02 to 9.91 in *G. hirsutum*. In terms of MW, *G. arboreum* ranged from 32.88 to 256.84 kDa, *G. raimondii* from 41.22 to 253.19 kDa, *G. barbadense*

from 28.65 to 300.93 kDa, and *G. hirsutum* from 15.56 to 252.16 kDa. Subcellular localization predictions showed that 87.2% of BRD proteins were nuclear, 6.2% were localized in chloroplasts, and 3.9% in the endomembrane system. A few proteins, such as GbBRD-12, GbBRD-21, GbBRD-44, and GhBRD-33, were predicted in the plasma membrane, whereas GbBRD-19 and GbBRD-51 were predicted to be localized in the extracellular space (Supplementary Table S2).

Chromosomal distribution of BRD genes

The 13 chromosomes of *G. arboreum* encased *BRD* genes ranging from *GaBRD-01* on Chr1 to *GaBRD-40* on Chr13. Both Chr5 and Chr9 exhibited the maximum distribution of *BRD* genes, 7 genes located on each chromosome, while Chr13 contained solely the single *GaBRD-40* gene (Fig. 1a). *G. raimondii*, likewise, comprised 13 chromosomes and all delineated *GrBRD* genes were allocated across chromosomes. Chr9 housing the highest number, with 7 *BRD* genes, while Chr3 and Chr10 each with a single gene (Fig. 1b). Similarly, the tetraploid cotton species, *G. barbadense* alongside *G. hirsutum*, were sorted into A and D genomes which include

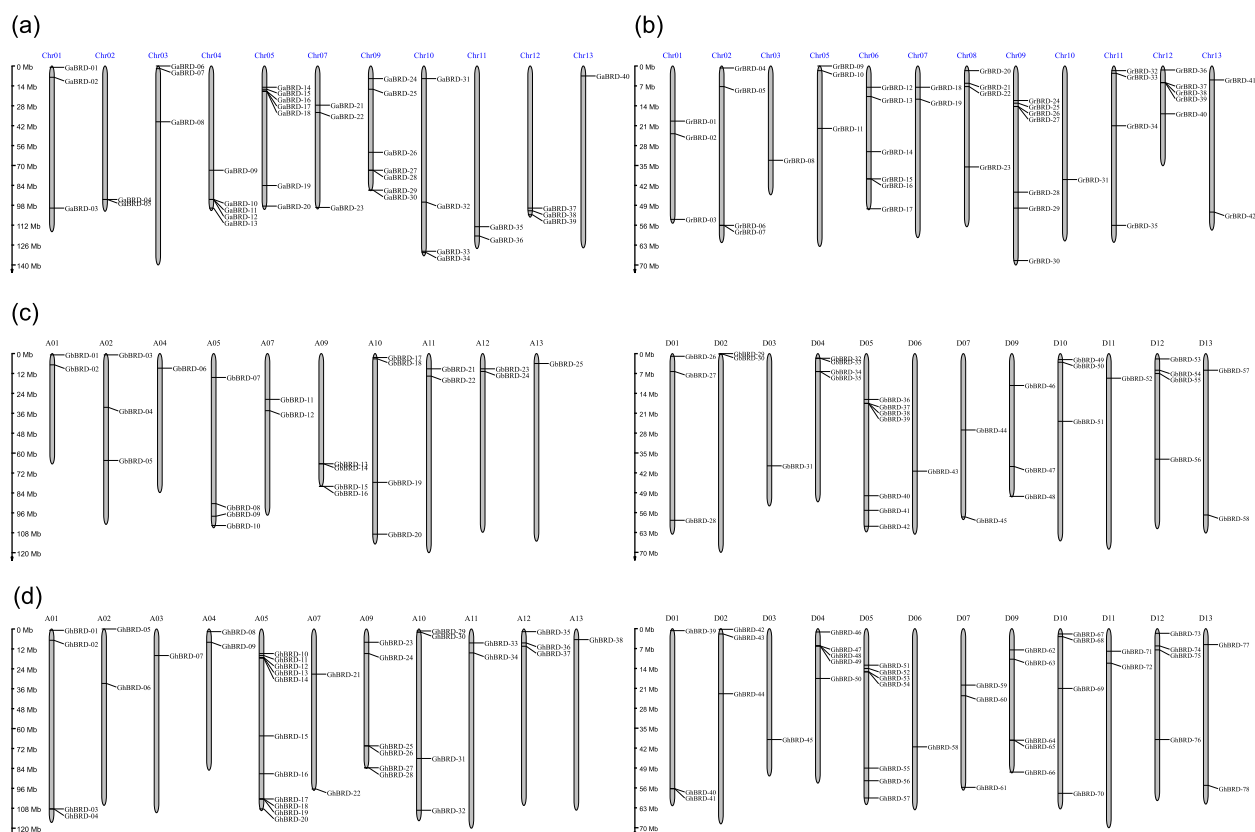


Fig. 1 *BRD* genes mapping and distribution on chromosomes. **a** *Gossypium arboreum*, **(b)** *G. raimondii*, **(c)** *G. barbadense*, and **(d)** *G. hirsutum*. Gray bars represent the chromosomes, and black lines link the gene location based on mapped scale

26 chromosomes identified as A01 to A13 and D01 to D13. In *G. barbadense*, D05 chromosome accommodated the highest number of *BRD* genes, while A04, A13, D03, D06, and D08 chromosomes contained only a single *BRD* gene. Chromosomes A03, A06, A08, and D08 lacked any gene presence. *GbBRD*-59 to *GbBRD*-66 represented the unannotated scaffolds, not aligned with any specific chromosome (Fig. 1c). *G. hirsutum* similarly encases 26 chromosomes within its tetraploid genomes, with D09 chromosome being the smallest and containing 5 *BRD* genes, whereas A06 is the largest chromosome devoid of any genes. A05 chromosomes accommodate a maximum of 11 *GhBRD* genes, while A03, A13, D03, and D06 chromosomes contain only a single gene. A06, A08, and D08 chromosomes exhibited a deficiency of genes. *GhBRD*-79 is the single *G. hirsutum* unannotated gene that classifies as a scaffold not aligned with any chromosome (Fig. 1d).

Phylogenetic classification of *BRD* genes in cotton

Two phylogenetic trees were constructed to investigate the evolutionary relationships among *BRD* family members in various cotton species, *A. thaliana*, and *O. sativa* (Fig. 2a, b). A total of 277 *BRD* proteins, 40 from *G. arboreum*, 42 from *G. raimondii*, 66 from *G. barbadense*, 79 from *G. hirsutum*, 28 from *A. thaliana*, and 22 from *O. sativa* were used to build an unrooted phylogenetic tree. Based on the resulting phylogenetic clades for 277 *BRD* proteins, they were categorized into seven distinct classes. CLASS-VI represented the largest

group, containing 122 genes, followed by CLASS-V with 49 genes, CLASS-VII with 34 genes, CLASS-IV with 33 genes, CLASS-III with 17 genes, CLASS-I with 14 genes, and CLASS-II with the fewest genes (only 8). Diploid cotton species *G. arboreum* and *G. raimondii* exhibited an equal distribution of *BRD* genes in CLASS-I, CLASS-II, and CLASS-VII. However, *G. arboreum* lost three genes, one each from CLASS-III, CLASS-IV, and CLASS-VI, while *G. raimondii* lost one gene from CLASS-V. Among tetraploid cotton species, *G. barbadense* and *G. hirsutum* shared an equal number of *BRD* genes only within CLASS-II. *G. barbadense* lost a total of 14 genes: three from CLASS-I, one each from CLASS-IV and CLASS-V, seven from CLASS-VI, and two from CLASS-VII. *G. barbadense* also gained one gene ingroup-III. In contrast, *G. hirsutum* lost only one gene from CLASS-III. Notably, a significant difference in gene loss between the tetraploid cotton species was observed, with *G. barbadense* losing ten genes in CLASS-VI alone (Fig. 2b).

Intron–exon and motifs distribution of *BRD* genes

Most of the *BRD* protein sequences in *G. arboreum* had conserved motifs 1 and 2 types (Fig. 3a). *GaBRD* genes had CDS between two and twenty-four but no untranslated regions (UTR). The highest number of coding regions was found in *GaBRD*-21, whereas longer coding sections were divided into several segments found in *GaBRD*-10, *GaBRD*-11, and *GaBRD*-12. Similarly, most of the *BRD* genes in *G. raimondii* also had types 1 and

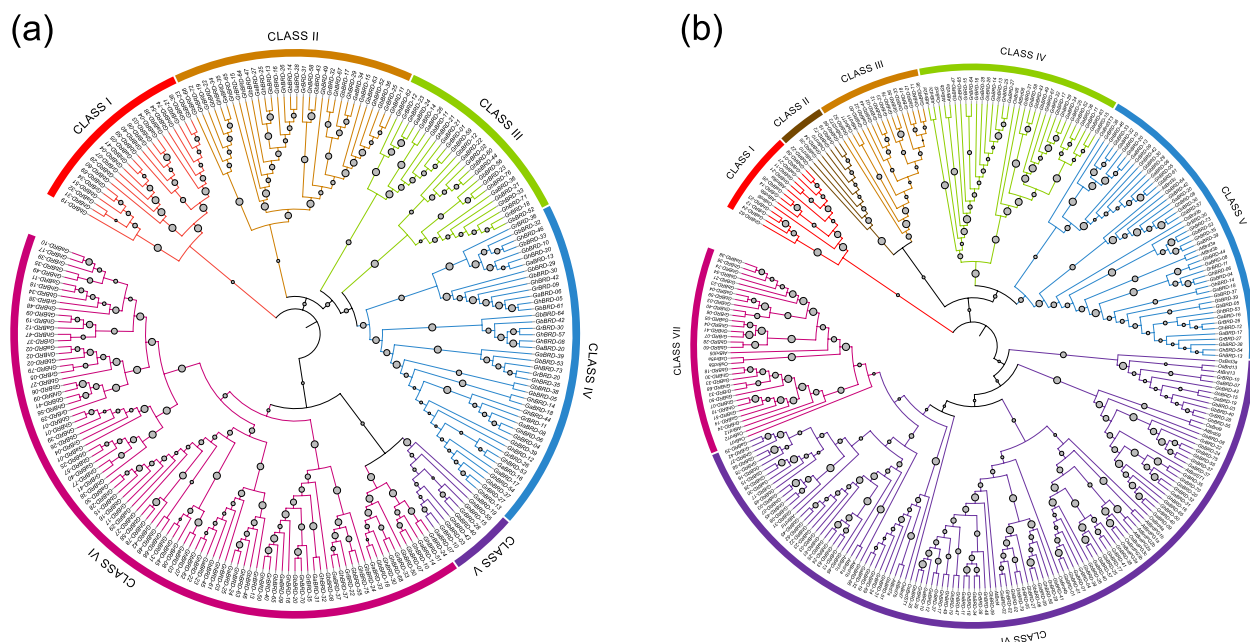


Fig. 2 Phylogenetic classification of *BRD* proteins. **a** *Gossypium* spp. with 277 *BRD* proteins identified **(b)** *Gossypium* spp. along with *Arabidopsis thaliana* and *Oryza sativa* with 277 *BRD* proteins identified. Phylogenetic classes are indicated by color scheme and gray circles represent the strength of bootstrap

2 (Fig. 3b). *GrBRD*-02 had up to 11 motifs, but *GrBRD*-14 only had one. Each of the six *GrBRD* genes had two motifs. The coding and non-coding sections of *GrBRD* genes differ. *GrBRD*-17 was the only gene lacking a UTR, and *GrBRD*-12 had a single intron at the 3' end. Out of the 42 *BRD* genes that had been found, 24 had both 5' and 3' UTRs (Fig. 3b). *BRD* genes mostly maintained motifs types 1 and 2 in *G. barbadense*. The presence and structure of motifs determine the classification of *BRD* genes. There were twelve motifs in *GbBRD*-44, compared with one motif each in *GbBRD*-11, *GbBRD*-40, and *GbBRD*-66. Thirteen genes had two motifs, but the majority of *BRD* genes had four (two distinct and two conserved). The distributed coding sections and lack of UTRs found in *G. arboreum* genes were similar to the gene structure of *G. barbadense* *BRD* genes. There was just one CDS region in *GbBRD*-58 (Fig. 3c). Following a similar trend, most of *BRD* genes in *G. hirsutum* also exhibited both motifs 1 and 2. Three categories were formed via phylogenetic analysis based on relative clade density of branches. Thirteen *GhBRD* genes subsequently exhibited multiple motifs. There were two motifs in each of *GhBRD*-66, *GhBRD*-31, and *GhBRD*-69 (Fig. 3d).

Gene duplication analysis of *BRD* genes

A total of 104 *BRD* in-paralogous gene pairs were identified to be duplicated, either segmental or tandem. *G. hirsutum* had a maximum of 51 paralogous gene pairs (45 segmentally duplicated and 6 tandemly duplicated), followed by *G. barbadense*, which consisted of 32 pairs (26 segmentally duplicated and 6 tandemly duplicated), 11 paralogous *BRD* pairs had existed in *G. raimondii* (6 segmentally duplicated and 5 tandemly duplicated), and *G. arboreum* had 10 pairs with an equal distribution of segmental and tandemly duplicated (Supplementary Table S3). Moreover, among duplicated gene pairs, calculations were made for nucleotide substitutions (K_a and K_s), duplication time, and selection type. All duplicated pairings had a K_a/K_s ratio of less than 1, indicating that purifying selection was used to preserve these pairs' functional relationships. Duplication time of segmentally duplicated pairs ranged from 0.61 million years ago (MYA) to 48.76 MYA, and for tandemly duplicated pairs, duplication time ranged from 0.83 MYA to 11.21 MYA (Supplementary Table S3).

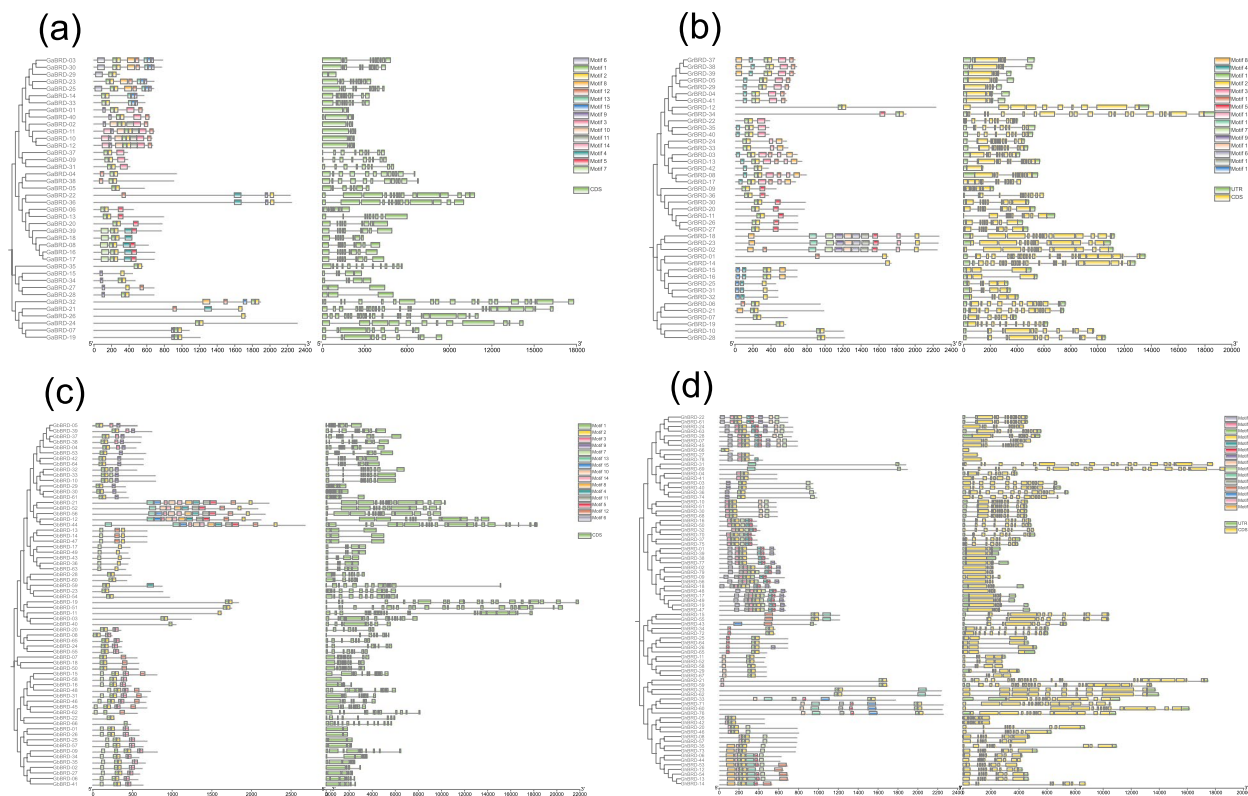


Fig. 3 Intron-exon structure, and motifs of *BRD* genes of *Gossypium* spp. **a** *G. arboreum* **(b)** *G. raimondii*, **(c)** *G. barbadense*, and **(d)** *G. hirsutum*. Green and yellow color shows the CDS and UTRs while motifs show with distinct colors

Syntenic relationship of *BRD* genes in diploid and tetraploid cotton

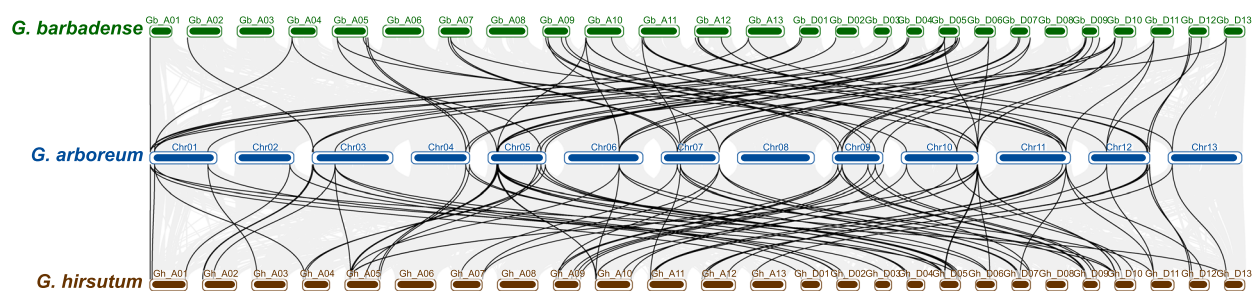
G. barbadense and *G. arboreum* shared 75 orthologous gene pairs, including 30 and 45 gene pairs in At and Dt sub-genomes. A total of 96 orthologous gene pairs were identified between *G. arboreum* and *G. hirsutum*, including 48 in At and 48 in Dt sub-genomes (Fig. 4a). Similarly, 82 orthologous gene pairs were identified from *G. barbadense* and *G. raimondii*, including 31 in At sub-genome and 51 in Dt sub-genome. Meanwhile, 101 orthologous gene pairs were identified from *G. raimondii* and *G. hirsutum*, including 42 gene pairs in At and 59 gene pairs in Dt sub-genomes (Fig. 4b). Sub-genome Dt contained a high ratio of orthologous gene pairs as compared with the At sub-genome. Furthermore, chromosomes Chr05, Chr07, and Chr09 of *G. arboreum* had a maximum number of orthologous with *G. barbadense* and *G. hirsutum* tetraploid cotton, whereas *G. raimondii* chromosomes Chr01, Chr06, Chr09, and Chr12 contained more orthologous *BRD* gene pairs with *G. barbadense* and *G. hirsutum* tetraploid species of cotton. Furthermore, translocation events of *BRD* genes were observed more at the intrachromosomal level through the evolutionary process of polyploidization.

Abundance of *cis*-regulatory elements in promoters of *BRD* genes

The 2 000 bp from the upstream transcriptional region were used to analyze the *cis*-regulatory elements. Identified elements were categorized into core promoter elements, including the CCAAT-box, TATA-box, and MYB transcription factor binding sites. Stress-responsive elements included ABRE (ascorbic acid response), TGACG-motif and CGTCA-motif (jasmonate response), TCA-element (salicylic acid response), W-box (defense and stress response), Msn2/Msn4 (general stress response), and LTR-element (low-temperature response). Moreover, hormone-responsive and developmental regulatory elements (G-box, GT-1 factors, GATA-motif, ethylene responsive element (ERE), GA-motif, as-1, metal responsive element (MRE), antioxidant responsive element (ARE)) and miscellaneous elements (Box 4, I-box, GARE-motif) were identified (Supplementary Table 4).

The heatmap analysis indicated that in *G. arboreum* *BRD* genes, particular *cis*-regulatory elements such as the G-box, MYB, ERE, and MYC display pronounced enrichment within the associated genes, thereby underscoring their vital role in the regulation of gene expression (Supplementary Fig. 1). In the case of *G. raimondii*,

(a)



(b)

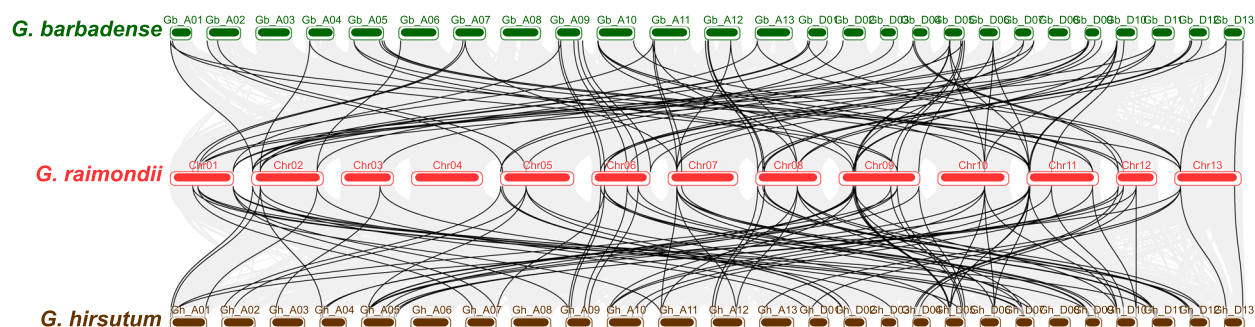


Fig. 4 Inter-chromosomal collinearity and multiple syntenic relationships among At and Dt genomes. **a** *G. barbadense*, *G. arboreum*, and *G. hirsutum* **(b)** *G. barbadense*, *G. raimondii*, and *G. hirsutum*. Chromosomes of *Gossypium* spp. are shown with distinct colors. Gray lines indicate the collinearity pairing, while black lines show the orthologous gene pairs' position on chromosomes

elements including MYB, G-box, ABRE, Box 4, and MYC were correlated with elevated gene enrichment, thereby implying their regulatory importance (Supplementary Fig. 2). Within *G. barbadense*, *cis*-elements like ABRE, G-box, MYB, MYC, Box 4 were prominent (Supplementary Fig. 3). For *G. hirsutum*, ERE, G-box, MYB, MYC, and Box4 *cis*-regulatory elements were present in abundance (Supplementary Fig. 4).

Protein–protein interaction network of BRD proteins

In cotton species, BRD proteins demonstrate notable protein-protein interactions that are associated with transcriptional regulation and epigenetic modifications. Within *G. arboreum*, GaBRD-32 occupied a pivotal role, acting as a transcription initiation factor belonging to the TAF1 family, and was engaged in interactions with proteins such as GaBRD-26, GaBRD-35, and GaBRD-21, which function as histone acetyl transferases. Furthermore, GaBRD proteins (for instance, GaBRD-29, GaBRD-12, GaBRD-40, and GaBRD-32) also serve as transcription initiation subunits within the global

transcription factor group E (GTE) family (Fig. 5a). The species *G. raimondii* exhibited analogous interaction patterns, with GrBRD-34 being central to protein interactions, while GrBRD-21, GrBRD-35, and GrBRD-28 were implicated in DNA binding within the context of bromodomain 4. Additionally, histone acetyltransferases and transcription factors, inclusive of GTE members, were likewise engaged in these processes (Fig. 5b). In *G. barbadense*, critical interactions involve transcription initiation factors (such as GbBRD-11 and GbBRD-51) alongside ATPase-like proteins (Fig. 5c). Proteins in *G. hirsutum* interacted in a comparable manner, with specific proteins, including GhBRD-66 and GhBRD-67, fulfilling the role in transcriptional regulation as epigenetic markers (Fig. 5d).

Secondary structure of BRD proteins

Alpha-helix structure was dominant in all BRD proteins with an average of 24.41%. GbBRD-08 was observed to have an alpha-helix structural integrity of 49.00%. Beta strand ranged from 1.00% to 21.00% with an average of

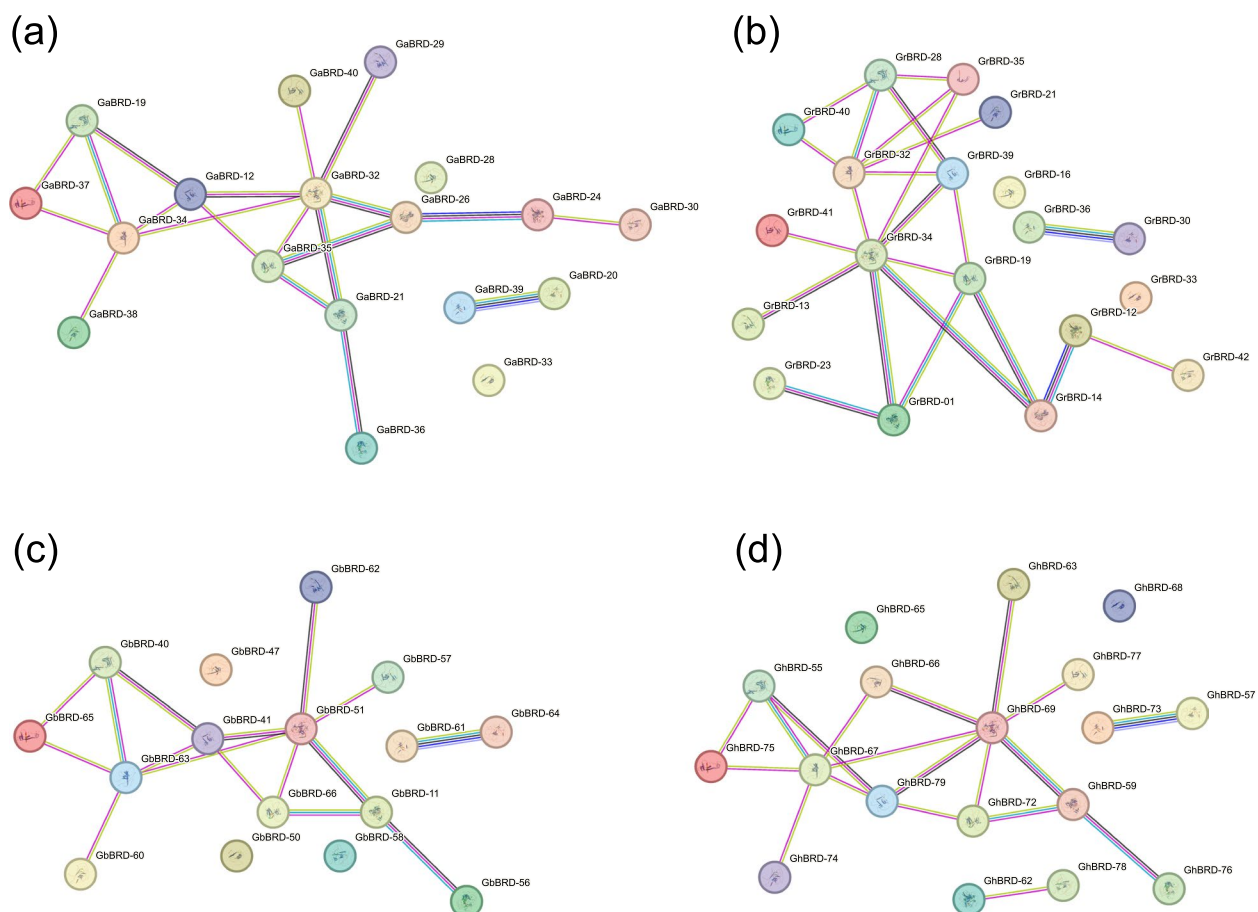


Fig. 5 Protein-protein interaction network predicting BRD proteins within each other and with other proteins. Strings indicate the interaction network, and the intensity of interaction is represented by the string width

2.07%. GbBRD-11 and GhBRD-59 were the proteins containing the highest beta-strand structure of 21.00%. Transmembrane structure was present in very low percentage, proteins GaBRD-07, GaBRD-18, GrBRD-10, GrBRD-28, GbBRD-19, GbBRD-40, GbBRD-44, GbBRD-51, GhBRD-15, GhBRD-55 only contained transmembrane structure, ranged from 1.00% to GaBRD-32 had highest of 3.00%. Disordered elements were relatively higher, about the average of 68.76% (Supplementary Fig. 5).

Orthologous cluster analysis of BRD genes among cotton species

Orthologous gene cluster analysis of four cotton species resulted in a total of 32 clusters containing all the members of BRD genes in *G. arboreum*, *G. raimondii*, *G. barbadense*, and *G. hirsutum*. Three unique clusters are also there, having only *G. hirsutum*, *G. barbadense*, and *G. raimondii* orthologous gene pairs of three species. Interestingly, there is one unique cluster that shares the common ortholog among *G. arboreum*, *G. hirsutum*, and *G. raimondii*. The bar-chart below shows the number of gene pairs in the cluster analysis, as *G. raimondii* and *G. hirsutum* have 32 clusters in common, followed by *G. arboreum* and *G. barbadense*, which have cluster counts of 29 and 31, respectively. Out of 227 genes, 17 singleton genes have 7.49% of total orthologous genes, in which *G. barbadense* has a maximum of 11 singleton followed by *G. hirsutum* with 4 and *G. arboreum* with 2 singleton genes in total. *G. raimondii* is the only cotton species that has no singleton gene pair. The BRD genes overall orthologs analysis has sorted out the sporadically observed evolutionary conservation links, functional conclusions, phylogenetic characteristics, and comparative genomics data (Fig. 6).

Expression analysis of BRD genes in different tissues under normal conditions

Transcriptomic data have been reanalyzed to study the expression pattern of BRD genes in organ development and various tissues under optimal conditions of cotton species. In *G. arboreum*, two BRD genes (*GaBRD-01* and *GaBRD-40*) were highly expressed while in ovule and fiber development at 15 days post anthesis (DPA), while *GaBRD-37*, *GaBRD-18*, and *GaBRD-07* genes showed mild expression in ovule at 15 DPA stage. The clustered genes (*GaBRD-25*, *GaBRD-14*, *GaBRD-33*, *GaBRD-01*, and *GaBRD-40*) showed continuous expression in root, stem, flower, ovule, and fiber tissues at a range of developmental stages except in the leaf tissue. The significant expression of all BRD genes in *G. arboreum* has resulted in their key role in ovule development (Fig. 7a). In the

case of *G. raimondii*, *GrBRD-41* gene is highly expressed, while *GrBRD-16* and *GrBRD-11* showed mild expression in ovule at 0 DPA stage. *GrBRD-11* and *GrBRD-16* genes were highly expressed in flower tissue. *GrBRD-04* and *GrBRD-41* showed expression patterns in various tissues, flower, leaf, ovule, seed, and fiber development. Heatmap expression of genes indicated their role in ovule formation (Fig. 7b).

The analysis of tissue-specific expression of *G. barbadense* resulted in the findings that the *GbBRD-07* gene showed high expression level in the ovule at 0 DPA stage and gradually decreased at 3 DPA, 5 DPA, 15 DPA, and 25 DPA, respectively. Genes in cluster *GbBRD-66*, *GbBRD-01*, *GbBRD-26*, *GbBRD-25*, *GbBRD-57* showed expression in root, stem, petal, torus, sepal, bract, anther, pistil, and ovule formation from stage 0 DPA to 25 DPA, whereas, leaf had a very low expression level. Other gene clusters *GbBRD-39*, *GbBRD-37*, *GbBRD-38*, *GbBRD-04*, *GbBRD-53*, *GbBRD-42*, *GbBRD-64* showed higher expression levels in ovule formation stages 0 DPA to 20 DPA while low expression levels at 25 DPA stage. Among those four genes, *GbBRD-01* and *GbBRD-26* showed higher expression levels in fiber development (Fig. 7c). In the case of *G. hirsutum*, gene clusters *GhBRD-01*, *GhBRD-38*, and *GhBRD-39*, *GhBRD-77* showed high expression levels in torus, stem, sepal, root, petal, flower, anther, filament, and ovule formation expect in leaf which had a mild expression level. *GhBRD-78* and *GhBRD-40* showed high expression levels in the flower (Fig. 7d). RNA-seq-based tissue specific expression profiling revealed the main role of BRD genes in reproductive stages (ovule, fiber, and seed), along with their influence on vegetative plant growth (root, stem, and leaf).

Expression profile of *G. hirsutum* BRD genes under abiotic stresses

Transcriptomic data were analyzed to study the expression of BRD genes in *G. hirsutum* under abiotic stresses, i.e., cold, heat, drought, and salt at various stages of time (1 h, 3 h, 6 h, 12 h, and 24 h). *GhBRD-10* and *GhBRD-51* were highly expressed under salt stress. Their expression was also expressed gradually as the heat duration increased. Similarly, under drought stress, these genes also showed optimal expression throughout the duration. *GhBRD-30* and *GhBRD-68* showed reciprocal expression patterns under cold and heat stresses. They were highly expressed as the cold stress increases. Under heat stress, they showed high expression levels at the start and decreased with the increase of heat duration. *GhBRD-01*, *GhBRD-39*, *GhBRD-38*, *GhBRD-77*, *GhBRD-24*, and *GhBRD-63* were relatively highly expressed under heat at the initial stage (1 h). Gene

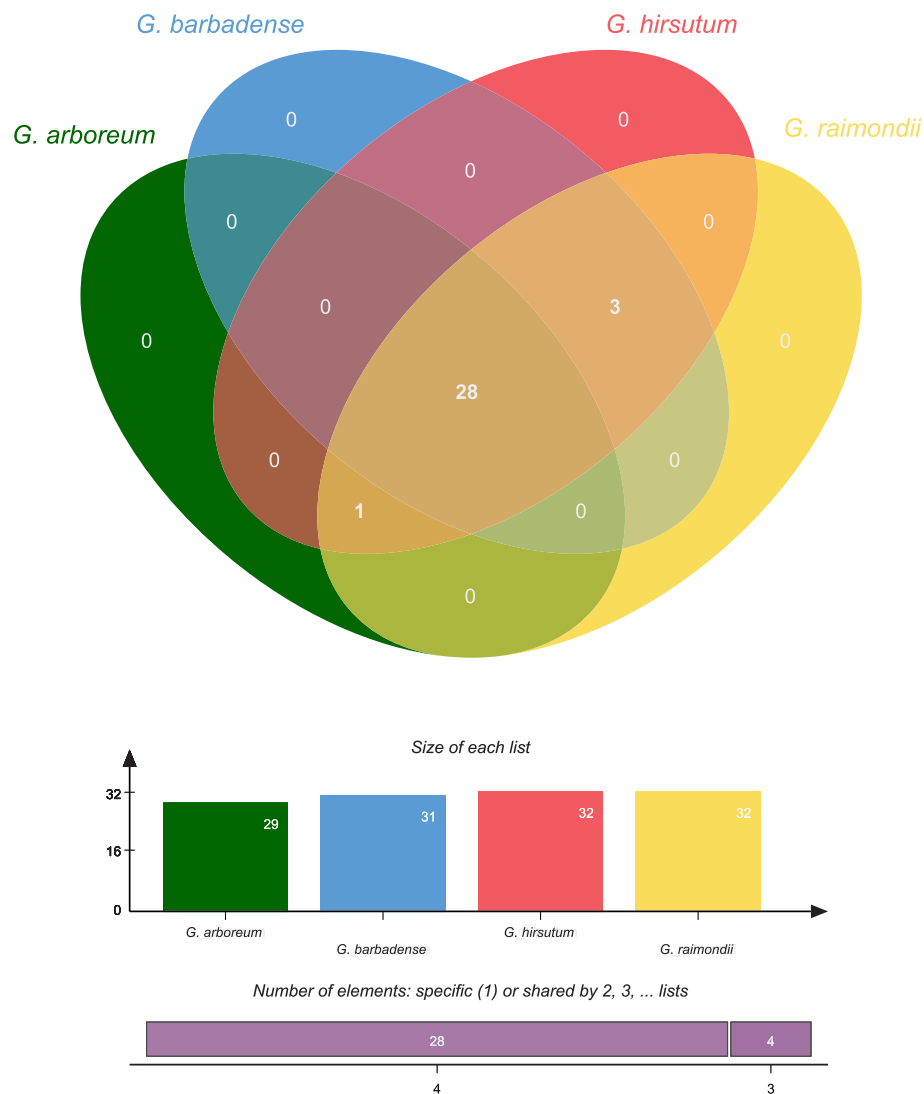


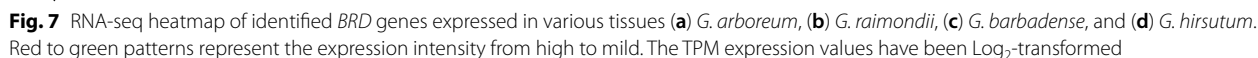
Fig. 6 Orthology and distribute orthologous BRD gene clusters of *Gossypium* spp

cluster comprising *GhBRD*-39, *GhBRD*-38, and *GhBRD*-77 showed the expression pattern of high intensity under all stress conditions (Supplementary Table S5). Furthermore, the genes *GhBRD*-26, *GhBRD*-19, *GhBRD*-22, *GhBRD*-61, *GhBRD*-24, *GhBRD*-33, *GhBRD*-63, *GhBRD*-37, and *GhBRD*-25 were simultaneously moderately expressed under all stages of abiotic stresses (Fig. 8).

qRT-PCR expression of *G. hirsutum* BRD genes under drought and salinity stress

The expression of selected genes was assessed in six cotton varieties, both drought-sensitive (K119, K194, K279) and drought-resistant (K254, K331, K354) under control and drought stress conditions at 0, 4, and 8 d (Supplementary Table S6). The results revealed

significant variation in gene expression, particularly for the *GhBRD*-01 gene, which showed increased expression levels under drought stress in drought-resistant varieties compared with drought-sensitive ones. For instance, the drought-sensitive variety K119 displayed a notable downregulation of *GhBRD*-01 at 8 d of drought stress, whereas the drought-resistant variety K254 exhibited a moderate increase at the same time point. Similar patterns were observed for other genes, such as *GhBRD*-38, where resistant varieties (e.g., K254, K354) showed upregulated expression at 8 d, while expression levels remained lower or decreased in drought-sensitive varieties. Moreover, *GhBRD*-77, a gene known for its role in stress tolerance, consistently showed higher expression levels in resistant varieties under drought



Under salinity stress, distinct gene expression patterns were identified between the salt-tolerant genotype Z9807 and the salt-sensitive genotype ZJ0102. In Z9807, several genes, including *GhBRD-01* and *GhBRD-38*, were significantly upregulated; *GhBRD-01* reached a peak expression level at 48 h, while *GhBRD-38* peaked at 24 h of stress exposure. Likewise, genes such as *GhBRD-10* and *GhBRD-39* showed elevated expression levels in Z9807, particularly at 12 h and 24 h, respectively. In contrast, the salt-sensitive genotype ZJ0102 showed minimal expression changes for *GhBRD-10* and *GhBRD-39*, maintaining relatively stable expression across all stress time points. Genes *GhBRD-22* and *GhBRD-63* displayed modest expression changes in both genotypes, with slightly higher levels observed in Z9807 compared with ZJ0102 throughout the stress period. Moreover, *GhBRD-51* and *GhBRD-77* exhibited increased expression levels at later stress time points in Z9807, while their expression levels remained relatively unchanged or lower in ZJ0102.

BRD genes exhibit conserved motifs across cotton species

Epigenetic markers, often referred to as "readers, writers, and erasers", mediate post-translational modifications such as chromatin remodelling. These markers play a central role in regulating cellular responses related to development, metabolic pathways, and adaptation to environmental stresses (Atlasi et al. 2017; Biswas et al. 2018; Boyko et al. 2008). BRD proteins play a crucial role in regulating gene expression and cellular development. While they have been predominantly studied in humans, they are also well-researched in other eukaryotes, including *Plasmodium falciparum*, *Toxoplasma gondii*, *Trypanosoma cruzi*, and *Trypanosoma brucei* (Filippakopoulos et al. 2012; Hettmann et al. 1999; Singh et al. 2022). Our study is the first to comprehensively characterize the BRD gene family in cotton species, providing insight into its evolutionary dynamics and stress-responsive roles, which have been well-studied in model plants like *Arabidopsis*, rice, and wheat (Abiraami et al. 2023; Wang et al. 2024). We identified a total of 227 BRD genes in cotton species, including the tetraploid *G. hirsutum* and *G. barbadense*, as well as the diploid *G. arboreum* and *G. raimondii*. Comparatively, the number of BRD genes identified in cotton differs from previous studies in *A. thaliana*, *O. sativa*, and *Triticum aestivum*, which reported 28, 22,

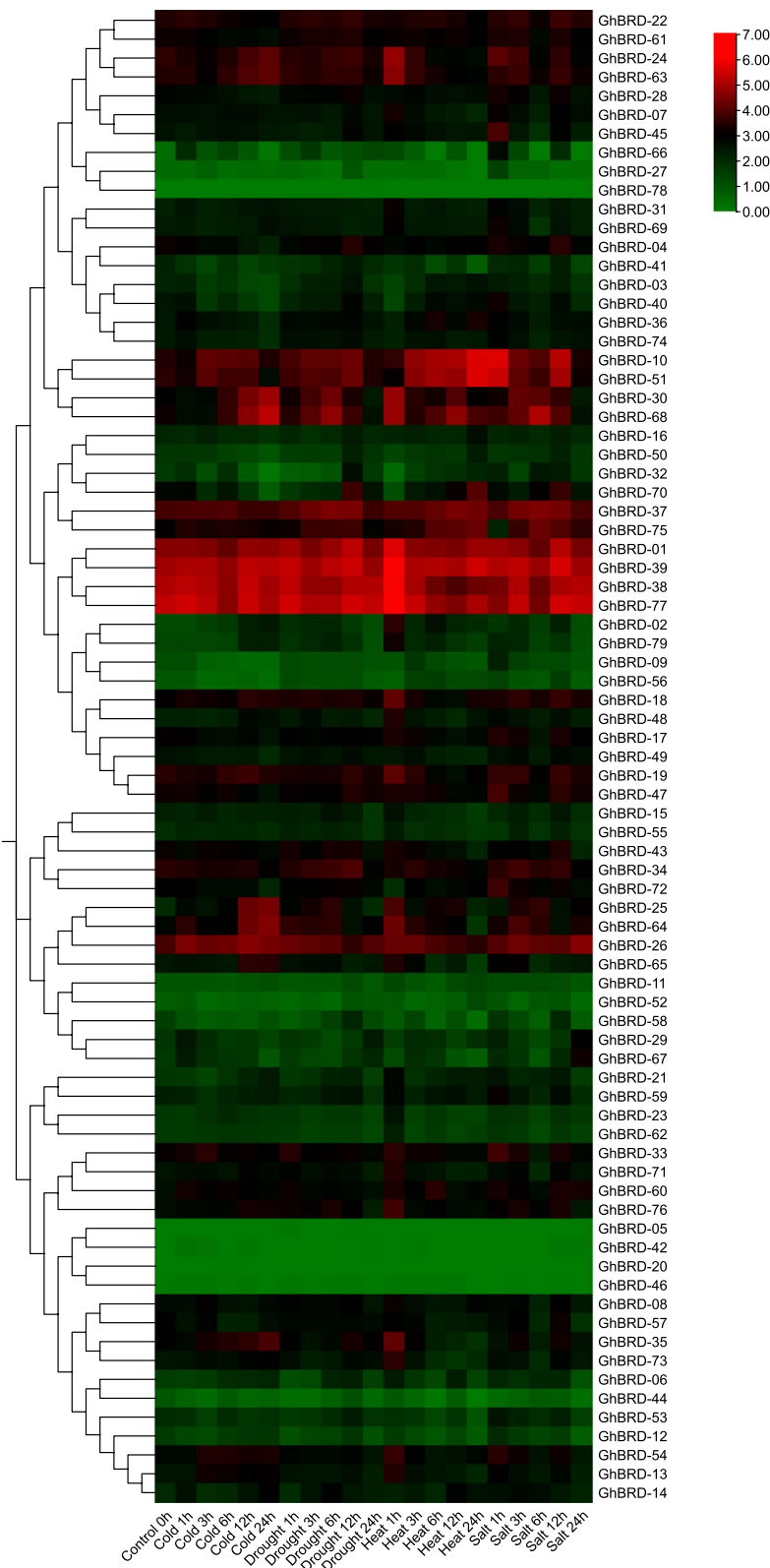


Fig. 8 Expression heatmap of identified *BRD* genes of *G. hirsutum* under drought, heat, cold, and salt

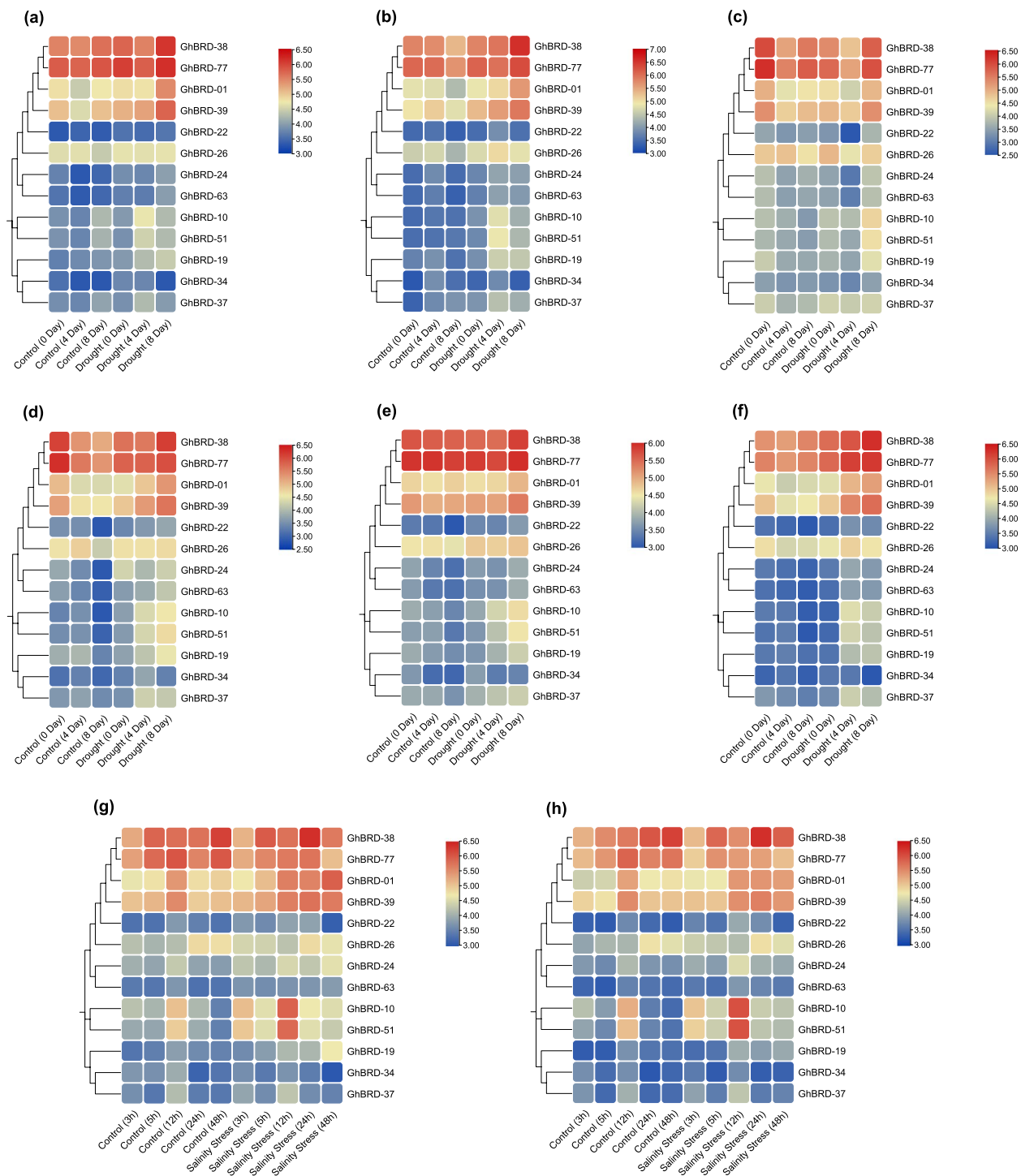


Fig. 9 qRT-PCR expression profiles of *G. hirsutum* cultivars under drought and salinity stress. **a** K119 (Yan 64-1), **b** K194 (Ganzao 032), **c** K279 (n74-250), **d** K254 (Jifeng 908), **e** K331 (Ari1327), **f** K354 (Zhong 1441), **g** Z9807, and **h** ZJ0102

and 82 *BRD* gene members, respectively (Wang et al. 2024).

The classification of *BRD* gene members in cotton species was based on conserved motifs, which were found to perform similar functions in corresponding model plant species belonging to the same phylogenetic class (Cao et al. 2015). Phylogenetic clusters of all four *Gossypium* species contained either the conserved motifs 1 or 2, highlighting their essential and conserved functions across species (La et al. 2005). Furthermore, the largest cluster, CLASS-VI, contained the highest number of conserved motifs, including motifs 1, 2, 6, 5, and 4. These conserved motifs indicate a diversity of functional protein domains, such as the WD40 domain (involved in plant cell wall formation and immune signaling), SANT domain (associated with abiotic stress regulation), RecA-like domain (responsible for DNA repair and maintaining genetic integrity), and bromodomain (Figs. 2, 3) (Chittela et al. 2010; Feng et al. 2021; Guerriero et al. 2015; Kong et al. 2015).

Gene duplication events reveal the role of polyploidization in cotton evolution

An important evolutionary feature of newly emerging genes is the amplification of already existing genes, which occurs through gene duplication events, either segmentally or tandemly, in higher plants (Panchy et al. 2016). The variable chromosomal distribution of *BRD* gene members in diploid and tetraploid cotton species highlights significant gene duplication events at the sub-genomic At and Dt levels (Panchy et al. 2016). Segmental duplication notably increases the number of gene members, as clearly observed in *G. hirsutum* and *G. barbadense* (Fig. 1; Supplementary Table S3). Intron-exon distribution plays a significant role in the evolutionary process of gene duplication in plants. Both the number of introns and exons are subject to evolutionary selection pressures, helping to maintain their conserved functions (Deshmukh et al. 2016). The same patterns were also observed in the four cotton species under investigation (Fig. 3). Additionally, the K_a/K_s values of in-paralogous duplicated *BRD* gene pairs in *Gossypium* species indicate purifying selection pressure, further suggesting the conservation of gene function throughout the evolution of these species. Similar findings have been reported in other higher plant species, such as *Z. mays*, *O. sativa*, and *S. lycopersicum* (Xu et al. 2019).

Cis-acting regulatory elements in the promoter region play a crucial role in regulating gene expression during growth, tissue development, and under abiotic stress conditions (Bifas et al. 2016). Estimation and expression analysis of *cis*-regulatory elements highlighted the

role of *BRD* genes in *Gossypium* species, indicating their involvement in plant growth and tolerance to environmental stresses (Liu et al. 2014). G-box and GT-1 factors are integral promoter elements that mediate photoperiod and light responsiveness, suggesting their role in plant growth regulation. This has been reported in *Fragaria vesca* and all species within the U's Triangle of *Brassica* (Fan et al. 2023; Shahzaib et al. 2024). Furthermore, the presence of CAAT-box, TATA-box, AT~TATA-box, and MYB *cis*-regulatory elements suggests their potential role in providing resistance to drought and salinity, as characterized in *T. aestivum* (Ebeed 2022). Moreover, hormone-responsive promoter elements such as ABRE, TGACG/CGTCA motifs, and TCA-elements are responsible for activating growth and defense signaling pathways against biotic stresses, as identified and characterized in *Betula platyphylla* (Fig. 5) (Lv et al. 2023).

BRD genes are highly expressed in stress and developmental pathways

The high expression pattern of *BRD* genes in various tissues of *Gossypium* species, including root, stem, leaf, and flower, suggests their role in plant growth and regulation. This expression pattern is similar to that observed for *BRD* genes in *A. thaliana* and *T. aestivum* (Abiraami et al. 2023; Wang et al. 2024). Similarly, tissue-specific expression during ovule formation and the significant expression of *BRD* genes suggest their role in fiber development, which is consistent with results reported in *T. aestivum*, where *BRD* expression at the spikelet stage is associated with high yield (Wang et al. 2024). Under abiotic stress conditions of salinity in *G. hirsutum*, the *GhBRD*-10 and *GhBRD*-51 genes were highly expressed at all stages from 0 to 24 h. This expression pattern is similar to that observed in the orthologous *AtBRD*-05 from *A. thaliana* and *OsBRD*-05a from *O. sativa* under similar salinity conditions (Abiraami et al. 2023). Moreover, the expression of *GhBRD*-01, *GhBRD*-38, *GhBRD*-39, and *GhBRD*-77, under heat or drought conditions, was also significantly high (Fig. 8), supporting their potential role in stress tolerance. These genes are orthologous to the *OsBRD*-01 gene from *O. sativa*, which also exhibits similar expression levels in response to various types of stress (Abiraami et al. 2023).

The expression patterns of *GhBRD* genes, such as *GhBRD*-01, *GhBRD*-38, and *GhBRD*-77, varied between drought-resistant and drought-sensitive cotton varieties, suggesting their potential involvement in drought stress responses (Fig. 9). Notably, *GhBRD*-01 expression level was initially downregulated at day 4 of drought treatment in the drought-resistant variety K254 but then upregulated at day 8, indicating a late activation under prolonged stress. Conversely, in drought-sensitive

varieties such as K119 and K194, *GhBRD*-01 was initially upregulated at day 4 but later decreased at day 8, reflecting a distinct response pattern. For *GhBRD*-38, the expression level was consistently higher in drought-resistant varieties like K254, K331, and K354 across all time points, particularly at later stages, suggesting a role in long-term drought adaptation. Similarly, *GhBRD*-77 showed relatively stable and elevated expression levels in drought-resistant varieties under stress conditions, indicating its possible role in enhancing drought tolerance mechanisms. In contrast, drought-sensitive varieties displayed lower or less consistent expression patterns of *GhBRD*-77. Overall, these results imply that *GhBRD*-01, *GhBRD*-38, and *GhBRD*-77 may contribute significantly to drought tolerance, possibly by regulating critical stress-response pathways.

Under salinity stress, the salt-tolerant genotype Z9807 exhibited notable upregulation of several *BRD* genes, particularly *GhBRD*-01, *GhBRD*-38, and *GhBRD*-39. Among these, *GhBRD*-01 showed the highest expression level at 48 h, *GhBRD*-38 remained consistently elevated across all stress time points, and *GhBRD*-39 peaked notably at 24 h, suggesting their important roles in mediating adaptive responses to salt stress. In contrast, the salt-sensitive genotype ZJ0102 displayed relatively minor changes or stable expression for these genes, reflecting a limited ability to activate effective salt-stress responses. Genes such as *GhBRD*-22 and *GhBRD*-63 exhibited only modest expression variations in both genotypes, indicating that these genes might have a general role rather than a salt-specific function. The clear differential expression patterns between Z9807 and ZJ0102 underscore the significance of bromodomain-containing genes in salinity tolerance mechanisms, highlighting their potential as targets for improving salt stress resistance in cotton breeding programs (Ding et al. 2012; Fang et al. 2014; Kim et al. 2015, 2012).

The findings from this study provide a comprehensive understanding of the *BRD* gene family in cotton species, highlighting their potential roles in growth regulation, fiber development, and stress tolerance mechanisms. The identification of conserved motifs and gene duplication events, along with the expression patterns observed under abiotic stress conditions, offers valuable insights for improving cotton varieties with enhanced resistance to environmental stresses like salinity and drought. These findings can be applied to the development of genetically modified cotton plants through targeted manipulation of the *BRD* genes to achieve better crop performance. However, further functional validation of these genes through experimental techniques such as gene knockout or over-expression studies is needed to fully understand their roles in stress tolerance. Moreover, this analysis is limited

to certain cotton species, and further research should explore a wider range of species to see if similar patterns hold true across different *Gossypium* species. Despite this, this study provides the fundamental groundwork for future research that could lead to practical applications in cotton breeding and crop improvement.

Conclusion

The *BRD* gene family exhibits significant structural and functional diversity across plant species, including *A. thaliana*, *O. sativa*, and *T. aestivum*. This study highlights the complex roles of *BRD* genes in the regulation of growth, development, and stress responses in four major cotton species. Through the identification and detailed analysis of the precisely isolated 227 *BRD* genes across both diploid and tetraploid *Gossypium* species, we have shed light on their potential contribution to abiotic stress tolerance, particularly under salinity and drought conditions. Moreover, this study provides valuable insights into the evolutionary conservation of the *BRD* genes, including their gene duplication patterns and conserved functions. The significant expression of these genes in various tissues and stresses, such as root, stem, leaf, flower, and drought, respectively, further reinforces their importance in plant growth and fiber development, aligning with similar patterns seen in other crops as well. The valuable data from extensive findings has the potential to be useful in the development of genetically engineered cotton varieties of the future with enhanced tolerance to various biotic and abiotic stresses.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s42397-025-00225-1>.

Supplementary Material 1: Supplementary Figure S1. Heatmap of existing abundance of *cis*-regulatory elements in *G. arboreum*.

Supplementary Material 2: Supplementary Figure S2. Heatmap of existing abundance of *cis*-regulatory elements in *G. raimondii*.

Supplementary Material 3: Supplementary Figure S3. Heatmap of existing abundance of *cis*-regulatory elements in *G. barbadense*.

Supplementary Material 4: Supplementary Figure S4. Heatmap of existing abundance of *cis*-regulatory elements in *G. hirsutum*.

Supplementary Material 5: Supplementary Figure S5. Predicted secondary structure of *BRD* proteins of *Gossypium* spp.

Supplementary Material 6: Supplementary Table S1. Protein sequences of all retrieved *BRD* genes of *Gossypium* spp. and query sequences.

Supplementary Material 7: Supplementary Table S2. Molecular and physiochemical properties of identified *BRD* genes of *Gossypium* spp.

Supplementary Material 8: Supplementary Table S3. In-paralogous gene pairs duplication, diverging time period and K_a/K_s values of *BRD* genes of *Gossypium* spp.

Supplementary Material 9: Supplementary Table S4. *Cis*-regulatory elements of *BRD* genes retrieved of *Gossypium* spp.

Supplementary Material 10: Supplementary Table S5. RNA-seq data for *GhBRD* genes under abiotic stresses.

Supplementary Material 11: Supplementary Table S6. qRT-PCR expression data for *GhBRD* genes under drought and salt stress.

Acknowledgements

The infrastructural support from the Centre for Advanced Studies in Agriculture and Food Security (CAS-AFS) at the University of Agriculture Faisalabad is gratefully acknowledged.

Authors' contributions

Tayyab M performed the experiments and wrote the original draft. Taj MH conducted data collection and visualization. Ahmad AI conducted formal data analysis. Rana FR contributed to methodology development and manuscript editing. Shahzaib M assisted with methodology and data curation. Atif RM validated experimental results and edited the manuscript. Azhar MT contributed to software analysis and validation. Khan SH conceptualized the study and provided supervision. Kakakhel IAM reviewed the statistical interpretation and manuscript editing. He SP administered the project and critically reviewed the manuscript. Rana IA supervised the project, provided resources, reviewed, and edited the manuscript.

Funding

Not applicable.

Data availability

Genomes of *G. arboreum*, *G. barbadense*, and *G. hirsutum* were obtained from CottonGen (<https://www.cottongen.org/>) and *G. raimondii* from Phytozome (<https://phytozome-next.jgi.doe.gov/>) databases. RNA-seq data were available on platform NCBI-Bioprojects (PRJNA179447, PRJNA171342, PRJNA1128012, PRJNA12541, PRJNA594268, PRJNA490626, and PRJNA601953) and data retrieved from ENA- EBI (<https://www.ebi.ac.uk/ena/browser/>).

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests. Authors Rana IA and Azhar MT are members of the Editorial Board of Journal of Cotton Research. Authors Rana IA and Azhar MT were not involved in the journal's review of, or decision related to this manuscript.

Author details

¹Centre of Agricultural Biochemistry and Biotechnology, University of Agriculture Faisalabad, Faisalabad, Punjab 38000, Pakistan. ²Department of Plant Breeding and Genetics, University of Agriculture Faisalabad, Faisalabad, Punjab 38000, Pakistan. ³Centre for Advanced Studies in Agriculture and Food Security, University of Agriculture Faisalabad, Faisalabad, Punjab 38000, Pakistan. ⁴School of Agriculture Sciences, Zhengzhou University, Zhengzhou 450000, China. ⁵Department of Soil and Environmental Sciences, University of Agriculture Peshawar, Peshawar, Khyber Pakhtunkhwa 25130, Pakistan. ⁶Tropical Crops Genetic Resources Institute, National Key Laboratory for Tropical Crop Breeding, Chinese Academy of Tropical Agricultural Sciences, Haikou 571101, China. ⁷Sanya Institute of Breeding and Multiplication/School of Tropical Agriculture and Forestry, Hainan University, Sanya, Hainan 572024, China. ⁸Key Laboratory for Quality Regulation of Tropical Horticultural Crops of Hainan Province, Haikou, 571101, China. ⁹State Key Laboratory of Cotton Bio-breeding and Integrated Utilization, Institute of Cotton Research, Chinese Academy of Agricultural Sciences, Anyang, Henan 455000, China. ¹⁰National Nanfan Research Institute (Sanya), Chinese Academy of Agricultural Sciences, Sanya, Hainan 572000, China.

Received: 29 November 2024 Accepted: 20 April 2025

Published online: 08 July 2025

References

- Abiraami T, Sanyal RP, Misra HS, et al. Genome-wide analysis of bromodomain gene family in *Arabidopsis* and rice. *Front Plant Sci.* 2023;14:1120012. <https://doi.org/10.3389/fpls.2023.1120012>.
- Anisimova M, Bielawski JP, Yang Z. Accuracy and power of the likelihood ratio test in detecting adaptive molecular evolution. *Mol Biol Evol.* 2001;18(8):1585–92. <https://doi.org/10.1093/oxfordjournals.molbev.a003945>.
- Atlasi Y, Stunnenberg HG. The interplay of epigenetic marks during stem cell differentiation and development. *Nat Rev Genet.* 2017;18(11):643–58. <https://doi.org/10.1038/nrg.2017.57>.
- Aydin Z, Altunbasak Y, Borodovsky M. Protein secondary structure prediction for a single-sequence using hidden semi-Markov models. *BMC Bioinformatics.* 2006;7:1–15. <https://doi.org/10.1186/1471-2105-7-178>.
- Bailey TL, Johnson J, Grant CE, et al. The MEME suite. *Nucleic Acids Res.* 2015;43(W1):W39–49. <https://doi.org/10.1093/nar/gkv416>.
- Bailey TL, Williams N, Misleh C, et al. MEME: discovering and analyzing DNA and protein sequence motifs. *Nucleic Acids Res.* 2006;34(S2):W369–73. <https://doi.org/10.1093/nar/gkl198>.
- Bardani E, Kallemi P, Tselika M, et al. Spotlight on plant bromodomain proteins. *Biology.* 2023;12(8):1076. <https://doi.org/10.3390/biology12081076>.
- Bilas R, Szafran K, Hnatuszko-Konka K, et al. *Cis*-regulatory elements used to control gene expression in plants. *Plant Cell Tiss Organ Cult.* 2016;127:269–87. <https://doi.org/10.1007/s11240-016-1057-7>.
- Biswas S, Rao CM. Epigenetic tools (The Writers, The Readers and The Erasers) and their implications in cancer therapy. *Eur J Pharmacol.* 2018;837:8–24. <https://doi.org/10.1016/j.ejphar.2018.08.021>.
- Boyko A, Kovalchuk I. Epigenetic control of plant stress response. *Environ Mol Mutagen.* 2008;49(1):61–72. <https://doi.org/10.1002/em.20347>.
- Cannon SB, Mitra A, Baumgarten A, et al. The roles of segmental and tandem gene duplication in the evolution of large gene families in *Arabidopsis thaliana*. *BMC Plant Biol.* 2004;4:1–21. <https://doi.org/10.1186/1471-2229-4-10>.
- Cao J, Li X. Identification and phylogenetic analysis of late embryogenesis abundant proteins family in tomato (*Solanum lycopersicum*). *Planta.* 2015;241:757–72. <https://doi.org/10.1007/s00425-014-2215-y>.
- Chao J, Li Z, Sun Y, et al. MG2C: a user-friendly online tool for drawing genetic maps. *Mol Hortic.* 2021;1:16. <https://doi.org/10.1186/s43897-021-00020-x>.
- Chappell K, Francou B, Habib C, et al. Galaxy is a suitable bioinformatics platform for the molecular diagnosis of human genetic disorders using high-throughput sequencing data analysis: five years of experience in a clinical laboratory. *Clin Chem.* 2022;68(2):313–21. <https://doi.org/10.1093/clinchem/hvab220>.
- Chen C, Wu Y, Li J, et al. TBtools-II: A "one for all, all for one" bioinformatics platform for biological big-data mining. *Mol Plant.* 2023;16(11):1733–42. <https://doi.org/10.1016/j.molp.2023.09.010>.
- Chittela RK, Sainis JK. Plant DNA recombinases: a long way to go. *J Nucleic Acids.* 2010;2010(1): 646109. <https://doi.org/10.4061/2010/646109>.
- Dan J, Chen T. Writers, erasers, and readers of DNA and histone methylation marks. In: Gray SG, editor. *Epigenetic cancer therapy*. Amsterdam, Netherlands: Elsevier Inc. 2023:39–63. <https://doi.org/10.1016/b978-0-323-91367-6.00012-x>.
- De Nadal E, Ammerer G, Posas F. Controlling gene expression in response to stress. *Nat Rev Genet.* 2011;12(12):833–45. <https://doi.org/10.1038/nrg3055>.
- Deshmukh RK, Sonah H, Singh NK. Intron gain, a dominant evolutionary process supporting high levels of gene expression in rice. *J Plant Biochem Biotechnol.* 2016;25:142–6. <https://doi.org/10.1007/s13562-015-0319-5>.
- Ding Y, Fromm M, Avramova Z. Multiple exposures to drought 'train' transcriptional responses in *Arabidopsis*. *Nat Commun.* 2012;3(1):740. <https://doi.org/10.1038/ncomms1732>.
- Doerks T, Copley RR, Schultz J, et al. Systematic identification of novel protein domain families associated with nuclear functions. *Genome Res.* 2002;12(1):47–56. <https://doi.org/10.1101/gr.203201>.
- Ebeed HT. Genome-wide analysis of polyamine biosynthesis genes in wheat reveals gene expression specificity and involvement of STRE and

- MYB-elements in regulating polyamines under drought. *BMC Genomics*. 2022;23(1):734. <https://doi.org/10.1186/s12864-022-08946-2>.
- Emmons S, Kobourov S, Gallant M, et al. Analysis of network clustering algorithms and cluster quality metrics at scale. *PLoS One*. 2016;11(7):e0159161. <https://doi.org/10.1371/journal.pone.0159161>.
- Fan J, Jiang F, Sun H, et al. Expression analysis of Trihelix transcription factor family in strawberries and functional characterization of *FvTrihelix6*. *Horticulturae*. 2023;9(6):633. <https://doi.org/10.3390/horticulturae9060633>.
- Fang H, Liu X, Thorn G, et al. Expression analysis of histone acetyltransferases in rice under drought stress. *Biochem Biophys Res Commun*. 2014;443(2):400–5. <https://doi.org/10.1016/j.bbrc.2013.11.102>.
- Feng C, Cai XW, Su YN, et al. *Arabidopsis* RPD3-like histone deacetylases form multiple complexes involved in stress response. *J Genet Genomics*. 2021;48(5):369–83. <https://doi.org/10.1016/j.jgg.2021.04.004>.
- Filippakopoulos P, Picard S, Mangos M, et al. Histone recognition and large-scale structural analysis of the human bromodomain family. *Cell*. 2012;149(1):214–31. <https://doi.org/10.1016/j.cell.2012.02.013>.
- Gasteiger E, Gattiker A, Hoogland C, et al. ExPASy: the proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Res*. 2003;31(13):3784–8. <https://doi.org/10.1093/nar/gkg563>.
- Gibney ER, Nolan CM. Epigenetics and gene expression. *Heredity*. 2010;105(1):4–13. <https://doi.org/10.1038/hdy.2010.54>.
- Golldack D, Lüking I, Yang O. Plant tolerance to drought and salinity: stress regulating transcription factors and their functional significance in the cellular transcriptional network. *Plant Cell Rep*. 2011;30:1383–91. <https://doi.org/10.1007/s00299-011-1068-0>.
- Goodstein DM, Shu S, Howson R, et al. Phytozome: a comparative platform for green plant genomics. *Nucleic Acids Res*. 2012;40(D1):D1178–86. <https://doi.org/10.1093/nar/gkr944>.
- Guerriero G, Hausman JF, Ezcurra I. WD40-repeat proteins in plant cell wall formation: current evidence and research prospects. *Front Plant Sci*. 2015;6:1112. <https://doi.org/10.3389/fpls.2015.01112>.
- Harrison PW, Amode MR, Austine-Orimoloye O, et al. Ensembl 2024. *Nucleic Acids Res*. 2024;52(D1):D891–9. <https://doi.org/10.1093/nar/gkad1049>.
- He P, Zhang Y, Xiao G. Origin of a subgenome and genome evolution of allotetraploid cotton species. *Mol Plant*. 2020;13(9):1238–40. <https://doi.org/10.1016/j.molp.2020.07.006>.
- Hettmann C, Soldati D. Cloning and analysis of a *Toxoplasma gondii* histone acetyltransferase: a novel chromatin remodelling factor in *Apicomplexan* parasites. *Nucleic Acids Res*. 1999;27(22):4344–52. <https://doi.org/10.1093/nar/27.22.4344>.
- Hu Y, Chen J, Fang L, et al. *Gossypium barbadense* and *Gossypium hirsutum* genomes provide insights into the origin and evolution of allotetraploid cotton. *Nat Genet*. 2019;51(4):739–48. <https://doi.org/10.1038/s41588-019-0371-5>.
- Huang G, Wu Z, Percy RG, et al. Genome sequence of *Gossypium herbaceum* and genome updates of *Gossypium arboreum* and *Gossypium hirsutum* provide insights into cotton A-genome evolution. *Nat Genet*. 2020;52(5):516–24. <https://doi.org/10.1038/s41588-020-0607-4>.
- Jones DT, Taylor WR, Thornton JM. The rapid generation of mutation data matrices from protein sequences. *Bioinformatics*. 1992;8(3):275–82. <https://doi.org/10.1093/bioinformatics/8.3.275>.
- Josling GA, Selvarajah SA, Petter M, et al. The role of bromodomain proteins in regulating gene expression. *Genes*. 2012;3(2):320–43. <https://doi.org/10.3390/genes3020320>.
- Kamburova V, Salakhutdinov I, Abdurakhmonov IY. Cotton breeding in the view of abiotic and biotic stresses: challenges and perspectives. In: Abdurakhmonov IY, editor. London, UK: IntechOpen; 2022. p. 679–710. <https://doi.org/10.5772/intechopen.104761>.
- Kawahara Y, de la Bastide M, Hamilton JP, et al. Improvement of the *Oryza sativa* Nipponbare reference genome using next generation sequence and optical map data. *Rice*. 2013;6:4. <https://doi.org/10.1186/1939-8433-6-4>.
- Kelley LA, Mezulis S, Yates CM, et al. The Phyre2 web portal for protein modeling, prediction and analysis. *Nat Protoc*. 2015;10(6):845–58. <https://doi.org/10.1038/nprot.2015.053>.
- Kim JM, Sasaki T, Ueda M, et al. Chromatin changes in response to drought, salinity, heat, and cold stresses in plants. *Front Plant Sci*. 2015;6:114. <https://doi.org/10.3389/fpls.2015.00114>.
- Kim JM, To TK, Ishida J, et al. Transition of chromatin status during the process of recovery from drought stress in *Arabidopsis thaliana*. *Plant Cell Physiol*. 2012;53(5):847–56. <https://doi.org/10.1093/pcp/pcs053>.
- Kong D, Li M, Dong Z, et al. Identification of TaWD40D, a wheat WD40 repeat-containing protein that is associated with plant tolerance to abiotic stresses. *Plant Cell Rep*. 2015;34:395–410. <https://doi.org/10.1007/s00299-014-1717-1>.
- La D, Sutch B, Livesay DR. Predicting protein functional sites with phylogenetic motifs. *Proteins*. 2005;58(2):309–20. <https://doi.org/10.1002/prot.20321>.
- Lescot M, Déhais P, Thijs G, et al. PlantCARE, a database of plant *cis*-acting regulatory elements and a portal to tools for *in silico* analysis of promoter sequences. *Nucleic Acids Res*. 2002;30(1):325–7. <https://doi.org/10.1093/nar/30.1.325>.
- Letunic I, Bork P. Interactive Tree of Life (iTOL) v6: recent updates to the phylogenetic tree display and annotation tool. *Nucleic Acids Res*. 2024;52(W1):W78–82. <https://doi.org/10.1093/nar/gkae268>.
- Liu JH, Peng T, Dai W. Critical *cis*-acting elements and interacting transcription factors: key players associated with abiotic stress responses in plants. *Plant Mol Biol Rep*. 2014;32:303–17. <https://doi.org/10.1007/s11105-013-0667-z>.
- Lv G, Han R, Shi J, et al. Genome-wide identification of the TIFY family reveals JAZ subfamily function in response to hormone treatment in *Betula platyphylla*. *BMC Plant Biol*. 2023;23(1):143. <https://doi.org/10.1186/s12870-023-04138-6>.
- MacColl H. On the diagrammatic and mechanical representation of propositions and reasoning: to the editors of the Philosophical Magazine and Journal. *The London, Edinburgh and Dublin Philos Mag and J Sci*. 1880;10(61):168–71. <https://doi.org/10.1080/14786448008626913>.
- Mahram A, Herboldt MC. NCBI BLASTP on high-performance reconfigurable computing systems. *ACM Trans Reconfig Technol Syst*. 2015;7(4):1–20. <https://doi.org/10.1145/2629691>.
- Marmorstein R, Berger SL. Structure and function of bromodomains in chromatin-regulating complexes. *Gene*. 2001;272(1–2):1–9. [https://doi.org/10.1016/s0378-1119\(01\)00519-4](https://doi.org/10.1016/s0378-1119(01)00519-4).
- Millán-Zambrano G, Burton A, Bannister AJ, et al. Histone post-translational modifications—cause and consequence of genome function. *Nat Rev Genet*. 2022;23(9):563–80. <https://doi.org/10.1038/s41576-022-00468-7>.
- Mujtaba S, Zeng L, Zhou M. Structure and acetyl-lysine recognition of the bromodomain. *Oncogene*. 2007;26(37):5521–7. <https://doi.org/10.1038/sj.onc.1210618>.
- Noreen S, Ahmad S, Fatima Z, et al. Abiotic stresses mediated changes in morphophysiology of cotton plant. In: Ahmad S, Hasanuzzaman M, editors. Cotton production and uses. Singapore city, Singapore: Springer; 2020. p. 341–66. https://doi.org/10.1007/978-981-15-1472-2_18.
- Panchy N, Lehti-Shiu M, Shiu SH. Evolution of gene duplication in plants. *Plant Physiol*. 2016;171(4):2294–316. <https://doi.org/10.1104/pp.16.00523>.
- Paysan-Lafosse T, Blum M, Chuguransky S, et al. InterPro in 2022. *Nucleic Acids Res*. 2023;51(D1):D418–27. <https://doi.org/10.1093/nar/gkac993>.
- Rao S, Ramakrishnan R, Silberstein A, et al. Sailfish: a framework for large scale data processing. In: SOCC '12: ACM Symposium on Cloud Computing. SoCC '12: Proceedings of the Third ACM Symposium on Cloud Computing. New York, USA: Association for Computing Machinery; 2012. p. 1–14. <https://doi.org/10.1145/2391229.2391233>.
- Reiser L, Bakker E, Subramaniam S, et al. The Arabidopsis information resource in 2024. *Genetics*. 2024;227(1):iyae027. <https://doi.org/10.1093/genetics/iyae027>.
- Sain SK, Brown JK, Monga D. Global status of cotton leaf curl disease—an ongoing threat to cotton production. In: Cotton: some insights. Ludhiana, India: The Crop Improvement Society of India Ludhiana; 2023. p. 92–134.
- Saitou N, Nei M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol*. 1987;4(4):406–25. <https://doi.org/10.1093/oxfordjournals.molbev.a040454>.
- Savojardo C, Martelli PL, Fariselli P, et al. BUSCA: an integrative web server to predict subcellular localization of proteins. *Nucleic Acids Res*. 2018;46(W1):W459–66. <https://doi.org/10.1093/nar/gky320>.
- Schmittgen TD, Livak KJ. Analyzing real-time PCR data by the comparative CT method. *Nat Protoc*. 2008;3(6):1101–8. <https://doi.org/10.1038/nprot.2008.73>.
- Shahzaib M, Khan UM, Azhar MT, et al. Phylogenomic curation of Ovate Family Proteins (OFPs) in the U's Triangle of *Brassica* L. indicates stress-induced

- growth modulation. *PLoS One*. 2024;19(1):e0297473. <https://doi.org/10.1371/journal.pone.0297473>.
- Shim J, Mangat P, Angeles-Shim R. Natural variation in wild *Gossypium* species as a tool to broaden the genetic base of cultivated cotton. *J Plant Sci Curr Res*. 2018;2:5. <https://doi.org/10.24966/pscr-3743/100005>.
- Sievers F, Higgins DG. Clustal Omega, accurate alignment of very large numbers of sequences. In: Russell D, editor. Multiple sequence alignment methods. Totowa, New Jersey: Humana Press; 2014. p. 105–16. https://doi.org/10.1007/978-1-62703-646-7_6.
- Singh AK, Alkrimi S, Glass KC. Structural and biophysical characterization of *Plasmodium falciparum* bromodomain protein 1. *FASEB J*. 2022;36(S1). <https://doi.org/10.1096/fasebj.2022.36.s1.r6219>.
- Sun J, Lu F, Luo Y, et al. OrthoVenn3: an integrated platform for exploring and visualizing orthologous data across genomes. *Nucleic Acids Res*. 2023;51(W1):W397–403. <https://doi.org/10.1093/nar/gkad313>.
- Szklarczyk D, Gable AL, Nastou KC, et al. The STRING database in 2021: customizable protein–protein networks, and functional characterization of user-uploaded gene/measurement sets. *Nucleic Acids Res*. 2021;49(D1):D605–12. <https://doi.org/10.1093/nar/gkaa1074>.
- Tamkun JW, Deuring R, Scott MP, et al. brahma: A regulator of *Drosophila* homeotic genes structurally related to the yeast transcriptional activator SNF2SWI2. *Cell*. 1992;68(3):561–72. [https://doi.org/10.1016/0092-8674\(92\)90191-e](https://doi.org/10.1016/0092-8674(92)90191-e).
- Tamura K, Stecher G, Kumar S. MEGA11: molecular evolutionary genetics analysis version 11. *Mol Biol Evol*. 2021;38(7):3022–7. <https://doi.org/10.1093/molbev/msab120>.
- Voora V, Bermudez S, Farrell JJ, et al. Cotton prices and sustainability. Winnipeg, Canada: The International Institute for Sustainable Development; 2023. p. 1–37. <https://www.iisd.org/system/files/2023-01/2023-global-market-report-cotton.pdf>.
- Wang J, Chitsaz F, Derbyshire MK, et al. The conserved domain database in 2023. *Nucleic Acids Res*. 2023;51(D1):D384–8. <https://doi.org/10.1093/nar/gkac1096>.
- Wang Y, Shen S, Wu Z, et al. Genome-wide analysis of *Triticum aestivum* bromodomain gene family and expression analysis under salt stress. *Planta*. 2024;260:117. <https://doi.org/10.1007/s00425-024-04549-1>.
- Wang Y, Tang H, DeBarry JD, et al. MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Res*. 2012;40(7):e49. <https://doi.org/10.1093/nar/gkr1293>.
- Wittkopp PJ, Kalay G. *Cis*-regulatory elements: molecular mechanisms and evolutionary processes underlying divergence. *Nat Rev Genet*. 2012;13(1):59–69. <https://doi.org/10.1038/nrg3095>.
- Xu X, Yang Y, Liu C, et al. The evolutionary history of the sucrose synthase gene family in higher plants. *BMC Plant Biol*. 2019;19:566. <https://doi.org/10.1186/s12870-019-2181-4>.
- Yang Z, Nielsen R. Estimating synonymous and nonsynonymous substitution rates under realistic evolutionary models. *Mol Biol Evol*. 2000;17(1):32–43. <https://doi.org/10.1093/oxfordjournals.molbev.a026236>.
- Yu J, Jung S, Cheng CH, et al. CottonGen: the community database for cotton genomics, genetics, and breeding research. *Plants*. 2021;10(12):2805. <https://doi.org/10.3390/plants10122805>.
- Yuan D, Ahamed A, Burgin J, et al. The European nucleotide archive in 2023. *Nucleic Acids Res*. 2024;52(D1):D92–7. <https://doi.org/10.1093/nar/gkad1067>.
- Zhao T, Schranz ME. Network approaches for plant phylogenomic synteny analysis. *Curr Opin Plant Biol*. 2017;36:129–34. <https://doi.org/10.1016/j.pbi.2017.03.001>.