Brief Bioinformatics characterization of Cotton bZIP transcription factors family from *Gossypium hirsutum, Gossypium arboreum* and *Gossypium raimondii*

Vaishali Khanale (✉ vaishali.khanale@mahyco.com)

Mahyco Private Limited   https://orcid.org/0000-0001-8013-2537

Anjanabha Bhattacharya

Mahyco Private Limited

Rajendra Satpute

Government Institute Of Science

Bharat Char

Mahyco Private Limited

Abstract

Cotton is an important commodity in the world economy. In this study we have carried out genome-wide identification and bioinformatics characterization of basic leucine zipper domain proteins (bZIPs) from cultivated cotton species *G. hirsutum* along with two sub-genome species of allotetraploid cotton, *G. arboreum* and *G. raimondii*. A total of 228 bZIP genes of G. hirsutum, 91 bZIP genes of G. arboreum and 86 bZIP genes of G. raimondii were identified from CottonGen database. Cotton bZIP genes were annotated in standard pattern according to their match with *Arabidopsis* bZIPs. Multiple genes with similar bZIP designations were observed in cotton, linked to the gene duplication. Cotton bZIPs are distributed across all 13 chromosomes with varied density. Phylogenetic characterization of all three cotton species bZIPs classified them into 12 subfamilies, namely A B, C, D, E, F, G, H, I, J, K and S and further into eight subgroups according to their predicted functional similarities, viz., A1, A2, A3, C1, C2, S1, S2 and S3. Subfamily A and S are having maximum number of bZIP genes, subfamily B, H, J and K are single member families. Cotton bZIP protein functions were predicted from identified motifs and orthologs from varied species. BRLZ domain analysis of G. raimondii bZIPs revealed the presence of conserved basic region motif N-X7-R/K in almost all subfamily members, variants are GrbZIP62 with N-X7-I motif and GrbZIP76 with K-X7-R motif. Leucine heptad repeats motif, are also present in variant numbers from two to nine with leucine or other hydrophobic amino acid at designated position among 12 subfamily members. STRING protein interaction network analysis of *G. raimondii* bZIPs observed strong interaction between A-D, B-K and C-S subfamily members.

Key Message

Bioinformatics analysis of bZIPs of cultivated cotton species *G. hirsutum* along with two sub-genome species *G. arboreum and G. raimondii* at one platform will certainly help the researchers in the selection of specific cotton bZIP genes according to the close alignment with *Arabidopsis* orthologs or sub-genome homolog for functional characterization.

Introduction

Cotton is a significant cash crop. Approximately 32.93 million hectare was under cotton cultivation around the globe in 2019, with the highest production of 5.77 million metric tons in India followed by US and China (http://ministryoftextiles.gov.in; https://www.statista.com). The textile industry is a significant contributor to the nation's economy and employment, the global textile market share was USD 961.5 billion in 2019 (https://www.grandviewresearch.com).

Functional genomics is a key approach for identifying genes for important traits like yield, biotic-abiotic stress tolerance and fiber quality. Cultivated *G. arboreum*, A subgenome species, and its counterpart, non-spinnable fiber producer *G. raimondii*, D subgenome species are two progenitor species of cultivated allotetraploid cotton, *G.hirsutum* (Li et al. 2014). Estimated genome size of *G. hirsutum* is ~ 2305.2Mb (Chen et al. 2020). Estimated genome size of *G. arboreum* is ~ 1746Mb which is around two fold larger than *G. raimondii*, ~ 880Mb (Hendrix et al. 2005). Genetic information related to TFs of the two important progenitor species will play major role in the improvement of cultivated allotetraploid cotton. Wang et al. 2012 identified 2,706 transcription factors in *G. raimondii* which includes 208 bHLH and 219 MYB class genes which were preferentially expressed in fiber (Wang et al. 2012). Kushanov et al. 2016 developed CAPs and DCAPs markers for the *PHYA1*, *PHYB* and *HY5* genes which are associated with the fiber quality and flowering time traits (Kushanov et al. 2016). Over-expression of *G. hirsutum* bZIP transcription factor, ABF2 (bZIP36) resulted in improved drought and salt tolerance in cotton and *Arabidopsis* (Liang et al. 2015).

TFs are the key regulators in plant development and stress adaptation. On the basis of conserved DNA binding domains, TFs are classified into different families among which bZIP-TFs play a major role in seed germination, flower development, and stress response. bZIP proteins specifically bind to DNA with an ACGT core, A-box (TACGTA), C-box (GACGTC) and G-box (CACGTG) motifs and binding specificity is regulated by flanking nucleotides (Jakoby et al. 2002). The bZIP domain carries two structural components located on a contiguous alpha-helix, the first one is a ~ 16 amino acid residues containing nuclear localization signal, followed by a DNA binding domain invariant N-X7-R/K and the second one is present exactly nine amino acids towards the C-terminus which is a leucines (L) heptad repeat or other bulky hydrophobic amino acids creating an amphipathic helix, L-X6-L-X6-L (Jakoby et al. 2002). In *Arabidopsis* bZIPs, variation observed in leucine repeats, subfamily 'D' bZIPs contains three repeats whereas; more than eight repeats are present in subfamily C and S (Wolfgang et al. 2018).

So far, plant bZIP proteins are classified into 9 to 13 subfamilies in different plant species based on structural and functional characteristics. Marc Jakoby et al. 2002 first classified *Arabidopsis thaliana* bZIP TFs into 10 subfamilies A to I and S. These groups were named with letters considering their family members, for example A for ABF/AREB/ABI5, B for big protein size, C for CPRF2-like, G for GBF, H for HY5, and S for small protein size (Jakoby et al. 2002). The classification of *Arabidopsis* bZIPs is updated by Wolfgang et al. 2018, where the authors divided 78 bZIP proteins of *Arabidopsis* in 13 subfamilies A to I, J, K, M and S (Wolfgang et al. 2018). Wang et al. 2019 classified *Arachis* bZIPs into nine subfamilies A, B, C, D, G, H, I, S and U (Wang et al. 2019). Nijhawan et al. 2008 classified *Oryza* bZIPs into 10 groups A to J (Nijhawan et al. 2008).

If we look at the cotton bZIP protein characterization, Zhang et al. 2018 identified 159 bZIP genes from *G. arboreum* and classified into 13 groups (Zhang et al.2018). Azeem et al. 2020 identified 87 bZIP genes of *G. arboreum* and 85 bZIP genes of *G. raimondii* from NCBI and classified into 11 subfamilies. Recently Wang et al. 2020 identified 207 bZIP genes of allotetraploid cotton *G.hirsutum* and classified into 13 subfamilies, A to I, S, M, K and J, having major contribution from subfamily 'A' and 'S' bZIPs (Wang et al. 2020). The basic differences between our bioinformatics characterization of cotton bZIPs and from reported studies are, we have characterized cultivated cotton species and two sub genome species together at one platform whereas two independent studies by Azeem et al. 2020 and Wang et al. 2020 characterized two sub genome species and *G. hirsutum* bZIP TFs separately. Importantly we have annotated bZIP TF according to standard pattern for clarity and to understand particular gene duplication within all three species. We have listed the duplicated bZIP genes of individual species. We have classified cotton bZIP family in detail including subfamily and subgroups according to their predicted function. Phylogenetic analysis of bZIPs of G. *hirsutum* along with two sub genome species will be useful in the process of selection of close orthologs for further characterization.

Emphasizing their functional importance, bZIP protein family in plants play crucial roles in developmental and stress responses. Abscisic acid responsive element binding factors (ABFs) from bZIP subfamily 'A' are activated by ABA signaling and are involved in downstream gene regulation in conjunction with stomatal closure mediated adaptation in drought stress (Sirichandra et al. 2010; Yoshida et al.2015). Wolfgang and Christoph described bZIP family 'C' genes (bZIP9, bZIP10, bZIP25, bZIP63) and S1 genes (bZIP1, bZIP2, bZIP11, bZIP44, bZIP53) interaction network in response to nutritional starvation and metabolic adaptation under nutritional stress (Wolfgang et al. 2018). TGA or subfamily 'D', bZIP family proteins binds to the TGACG consensus sequences, to form homo and hetero-dimer. TGA family proteins are predominantly involved in systemic acquired resistance through SA mediated interaction with NPR1 (Fan et al. 2002; Fu et al. 2013). TGA family member PERIANTHIA (PAN/bZIP46) plays a role in flower development, interact with ROXY-type GRX (CC type glutaredoxin, GRX- ROXY1) gene to regulate petal development (Gatz 2013; Gutsche et al. 2017). Recently, Ullah et al. 2019 studied soybean TGA family genes and differentiated legumes-specific *TGAs*

structures, involvement in legumes-specific biological processes like legumes-rhizobia symbiotic nodulation and predicted soybean bZIPs role in response to nitrogen.

The bZIP TF family from cotton yet to be fully understands and explored. Thus, this paper attempts phylogenetic analysis, structural characterization and functional role prediction of bZIPs from cultivated cotton species *G. hirsutum* along with two subgenome species *G. arboreum* and *G. raimondii.*

Material And Methods

A total of 228 bZIP genes of *G. hirsutum*, 91 bZIP genes of *G. arboreum* and 86 bZIP genes of *G. raimondii* were identified from CottonGen database using bZIP domain as a query as well as *Arabidopsis* bZIP protein sequence as a query or through keyword search to recent genome assembly of *Gossypium hirsutum* (AD1) 'TM-1' genome UTX_v2.1 (published article in may 2020), *Gossypium raimondii* JGI proteins and *Gossypium arboreum* BGI proteins (https://www.cottongen.org/). *Arabidopsis*-bZIP protein sequences were downloaded from TAIR (https://www.arabidopsis.org/). Cotton bZIP genes were annotated on the basis of E-value, bitscore value and percent identity resulted from *Arabidopsis*-bZIP match. All the data regarding gene ID, genomic position, exon number, CDS length, protein length, protein sequences and Arabidopsis match ID with E-value are recorded in supplementary file 1/, online resource1.

Phylogenetic analysis of 227 bZIP genes of *G. hirsutum*, 56 bZIP genes of *G. arboreum* and 57 bZIP genes of *G. raimondii* along with 73 bZIP genes of *Arabidopsis* was performed. Duplicated genes from *G. arboreum* and *G. raimondii* were excluded from phylogenetic analysis and selected first representative gene. For example GabZIP14 is having two entries GabZIP14-1 and GabZIP14-2 so GabZIP14-1 is selected for analysis and designated as GabZIP14. Phylogenetic tree was constructed by maximum likelihood phylogeny method, Dayoff (PAM) protein substitution model and performing 100 bootstrap, using MEGA X (https://www.megasoftware.net/).

MEME (http://meme-suite.org) motif search analysis was done for all identified bZIP genes, 228 bZIP genes of *G. hirsutum*, 91 bZIP genes of *G. arboreum* and 86 bZIP genes of *G. raimondii*, selecting 10 motifs identification option and any number of repetition in case of distribution. Identified motifs were evaluated using SMART (http://smart.embl-heidelberg.de/), Pfam (https://pfam.xfam.org/search/sequence), Interpro (https://www.ebi.ac.uk/interpro/) and TOMTOM. MAST file of all searches were submitted along with supplementary documents.

To understand the interaction network of *G. raimondii* bZIP proteins, STRING (https://string-db.org/) analysis was performed. Network was interpreted for strong protein interactions, sharing functions among themselves with highest edge confidence (0.900), which can be correlated by thickness of the edges.

NLS sequences were listed from plant transcription factor database for *G. raiimondii, Arabidopsis thaliana* bZIP proteins and predicted for *G. raimondii* bZIPs through cNLS Mapper (http://nls-mapper.iab.keio.ac.jp).

Results

G. hirsutum, G. arboreum and *G. raimondii* bZIP protein classification

Phylogenetic analysis of 227 bZIP genes of *G. hirsutum*, 56 bZIP genes of *G. arboreum* and 57 bZIP genes of *G. raimondii* along with 73 bZIP genes of *Arabidopsis* was performed. The phylogenetic analysis indicates that G. hirsutum, G. arboreum **and** G. raimondii bZIPs are closely related to *Arabidopsis* bZIPs and are conserved among all three genomes. *Gossypium* bZIP gene duplication happened before allotetraploidation, as along with *G. hirsutum*, bZIP gene duplication observed in sub-genome species, *G. aroboreum* and *G. raimondii*. So it is possible that during sub-

genome speciation from common ancestor gene duplication event might have happened in cotton. Figure 1 shows phylogenetic tree construction. Bootstrap percentage values are mentioned on each branch. Analyzed *G. hirsutum, G. arboreum* and *G. raimondii* bZIP genes are classified into 12 subfamilies A B, C, D, E, F, G, H, I, J, K and S and 8 subgroups A1, A2, A3, C1, C2, S1, S2 and S3. Subgroup classification was done according to clade separation within subfamily and predicated functional similarities. Further this classification is supported by *Arabidopsis*- bZIPs alignment, *G. raimondii* BRLZ domain alignment, exon- intron numbers, protein length and conserved motif identification (Fig. 1 and Table 1, 2, 3). For the ease of understanding, in the following elaborated subfamily wise description bZIP proteins from G. hirsutum, G. arboreum **and** G. raimondii collectively named as *Gossypium* bZIPs.

Subfamily A, subgroup A1 contains GossypiumbZIP12, 35, 36, 37, 39, 66 and 67 predicting involvement in ABA stress response and in seed maturation (Liang et al. 2015). Subgroup A2 contains GossypiumbZIP13 and 40 predicting role in drought stress response (Wang et al. 2019). Subgroup A3 contains FD like protein, involved in positive regulation of flowering GossypiumbZIP14 and 27 (Abe et al. 2005).

Subfamily B contain endoplasmic reticulum (ER) stress response transcription factor bZIP17, involved in salt and osmotic stress resistance in *Arabidopsis* (Liu et al.2008) and observed interacted with subfamily K member bZIP60 in STRING analysis, which is also involved in ER stress response.

Subfamily C, subgroup C1 contains GossypiumbZIP9, subgroup C2 contains GossypiumbZIP10, 25 and 63, reported to form hetero dimer with S1 subfamily bZIPs, are known to involved in anther development, positive regulation of seed maturation and response to starvation (Weltmeier et al. 2009; Wolfgang et al. 2018).

Subfamily D of bZIP proteins is involved in flower development, seed development; salicylic acid mediated signaling pathway and pathogen response. MEME motif search identified DOG1- seed dormancy control motif, tetratricopeptide protein-protein interaction motif in these family members. GossypiumbZIP20, 21, 22, 26, 45, 46, 47, 50, 57 and 65 are representing GossypiumbZIPs subfamily D.

Subfamily E of bZIPs functioning in cell wall and pollen development (Gibalova et al.2009, 2017) contains GossypiumbZIP34, 61 and 76. In *Arabidopsis thaliana* researchers discovered that a conserved proline residue in the third heptad region of leucine zipper of AtbZIP34 and AtbZIP61 interferes with the formation of homo-dimer whereas change of proline by an alanine in the above mentioned region can form homo dimer and bind to the G-box element (Shen et al.2007). A conserved proline residue which interferes homodimer formation is also found in the third heptad region of leucine zipper of GossypiumbZIP34 and 61 genes except in **GrbZIP61-1** which is carrying **serine instead of proline. GrbZIP61-1 can be further explored to confirm homodimer formation and G-box binding activity restoration due to the presence of serine instead of proline, which may be useful to understand GrbZIP61-1 functional involvement in stress response and in systems related to circadian clock, light and temperature, refer supplementary file 1/ Online resource 1 for alignment** (Ezer et al. 2017; Wolfgang et al. 2018). It is indicated that bZIP34 and bZIP61 is involved in the hetero-dimer formation with I and S subfamily members of bZIPs which are function in vascular development (Shen et al.2007).

Subfamily F which is well known to be involved in adaptation to zinc deficiency in *Arabidopsis*, wheat and barley contains GossypiumbZIP19, 23 and 24 (Assuncao et al. 2010; Inaba et al. 2015; Nazri et al. 2017; Evens et al. 2017).

Subfamily G plays role in binding to G box motif of genes regulated by hormones and light contains GossypiumbZIP16, 41 and bZIP55.

Subfamily H contains HY like transcription factor GossypiumbZIP56, HY5 which is involved in PhyB signaling pathway **and also associated with fiber quality in cotton**. HY5 gene is a positive regulator of photo morphogenesis (Wolfgang et

al. 2018). Kushnov et al.2016 has done comparative sequence analysis of three close relative of fiber quality genes, *PHYA1*, *PHYB*, *HY5* of *G. hirsutum* and *G. barbadense*, developed dCAPS markers which can be utilized in marker – assisted selection breeding for introgression of these genes into the either of the two alloteraploid cotton species. (Kushnov et al. 2016). Abdurakhmonov et al. 2017 developed *PHYA1* RNAi *G. hirsutum* with improved fiber quality and yield potential (Abdurakhmonov et al.2017).

Subfamily I which is involved in vascular development contains GossypiumbZIP18, 29, 30, 51, 52, 59 and 69 in 4 clades. Pyo et al.2006 found expression of bZIP 'I' group members AtbZIP18, 51, 52 and 59 in developing vascular cells and their precursor cells (Pyo et al. 2006).

Subfamily J and K contains GossypiumbZIP62 and GossypiumbZIP60 respectively. Rolly et al.2020 studied AtbZIP62 involvement in salt stress tolerance.

Subfamily S, subgroup S1 grouped GossypiumbZIP1, 2, 11, 44 and 53 which are known to be involved in the positive regulation of seed germination, salt stress and in starvation stress response in *Arabidopsis* and peanut. Subgroup S2 contains GossypiumbZIP4, 5, 6 and 7, *Arabidopsis* bZIPs of this group are known as a positive regulator of transcription, express in leaf, pollen, embryo, seed and root. Subgroup S3 contains GossypiumbZIP42, 43, 48 and 58 in one clade, and 70 in separate clade. Nowak et al.2016 mentioned bHLH109 regulation by AtbZIP4 and AtbZIP43, which is involved in *in-vitro* somatic embryogenesis and stress response. (Hanson et al. 2008; Alonso et al. 2009; Weltmeier et al. 2009; Ma et al. 2011; Dietrich et al. 2011; Wang et al. 2019)

GhbZIP44-6D aligned with AtbZIP71 and AtbZIP 72, 74 formed separate clades in which *Gossypium* bZIPs are not grouped.

Conserved motif analysis of *G. hirsutum*, *G. arboreum* and G. raimondii bZIP proteins

Signature bZIP domain is confirmed in all identified bZIP proteins of *G. hirsutum, G. arboreum* and *G. raimondii*. MEME motif analysis deciphering the presence Abscisic acid insensitive 5-Like protein 5-related motif in *G. hirsutum* subfamily 'A' and 'J'' members. Sterile alpha motif (SAM)/Pointed domain which is involved in protein- protein interaction is identified in subfamily 'A' bZIPs of all three species. G. raimondii subfamily 'A' bZIPs are carrying MYND-Zinc binding domain which is involved in protein -protein interaction in the transcriptional regulation context. *G. arboreum* Subfamily 'A' bZIPs, GabZIP18-3 and GabZIP11-3 genes are carrying GluR7, glutamate receptor domain from GluR proteins known to be function in light signal transduction and calcium homeostasis.

The presence of two DOG1- seed dormancy control motif and TGA2 like motif is identified in all three gossypium species 'D' subfamily bZIP proteins. Exclusively *G. raimondii* subfamily 'D' bZIPs are carrying the RPA interacting motif, which is conserved in eukaryotic DNA repair, replication proteins and presence of tetratricopeptide repeats motif which are involved in protein-protein interactions. **Further presence of tetratricopeptide repeats and other protein protein interaction domains in subfamily D should characterize, as TGA transcription factor family members, are involved in biotic stress tolerance through interaction with NPR1 and similarly STRING analysis also identified D subfamily protein interaction with A subfamily proteins**. A typical example of identified motifs in *G. raimondii* bZIP proteins has shown in Fig. 2 and some motifs are listed in Table 3.

BRLZ domain structural characterization of *G. raimondii* bZIP proteins

N-X7-R/K basic region

The conserved basic region N-X7-R/K variant is present in almost all analyzed *G. raimondii* bZIP proteins, variation is observed in subfamily E and J. Subfamily E- bZIP76 is having K-X7-R motif, subfamily J- bZIP62 is having N-X7-I motif.

(Fig. 4 and for BRLZ domain alignment of bZIP69, 76 refer supplementary file1).

Leucine heptad repeats variation in *G. raimondii* bZIPs

Variation in presence of number of leucine heptad repeats motif is observed in *G. raimondii* bZIP proteins. Figure 3 deciphering the detail structure of leucine heptad repeats of *G. raimondii* bZIP proteins, for GrbZIP69 and 76 refer supplementary file 1. First two leucine heptad repeats are conserved among all *G. raimondii* bZIP proteins with exception of **GrbZIP17, GrbZIP61, GrbZIP76 and S subfamily proteins GrbZIP 2, 4, 11, and 44** which are carrying either methionine, isoleucine, asparagine or valine at L0, L1 or L2 position (L0 to L9 denotes leucine position in leucine heptad repeats L-X6-L-X6-L motif).

Subfamily-wise leucine heptad repeats are described as follows

Subfamily A bZIP proteins are carrying two to four heptad repeats of leucine or presence of other hydrophobic amino acids at leucine position in last heptad.

Subfamily B, C, E, F and S bZIP proteins are carrying seven to nine heptad sequences with leucine repeats or other different hydrophobic or polar amino acids at leucine position; alanine observed predominantly at leucine position in subfamily C and valine is predominantly present in subfamily S proteins at leucine position.

Subfamily D, G, I and J bZIP proteins are carrying three to six leucine heptad sequences or other hydrophobic amino acids are present at leucine position with predominance of glycine in D subfamily and methionine in I subfamily members.

Pure leucine heptad repeats are present in subfamily H member **GrbZIP56** and subfamily K member **GrbZIP60**, with presence of 4 and 2 heptad repeats of leucine respectively.

STRING protein interaction analysis of *G. raimondii* bZIPs

G. raimondii bZIP proteins interaction analysis is performed using STRING, protein- protein association network software, Fig. 4 deciphering the observed network (https://string-db.org/). Very strong interaction observed among subfamily A bZIPs, GrbZIP35 (ABF1/AREB1), GrbZIP39 (ABI5), GrbZIP65, GrbZIP66 (AREB3) and GrbZIP67 (DPBF2) genes, which are known for abscisic acid inducible stress regulation, seed germination and seedling development in plants (Lindemose et al. 2013; Yoshida et al. 2015; Skubacz et al. 2016). Observed interaction between bZIP subfamily A and D proteins of *G.raimondii* are predicted to be involved in signal transduction pathways, biotic and abiotic stress tolerance. Another strong interaction observed among GrbZIP17 and GrbZIP60, two endoplasmic reticulum or unfolded protein response modulators known to get activated under environmental stresses (Humbert et al.2012; Howell, 2013). Interaction of GrbZIP35/ABF1 and GrbZIP55/GBF3 is also observed, these genes are involved in abiotic stress tolerance.

Interacting proteins GrbZIP9, GrbZIP53 and GrbZIP56 (HY5), are indicating their common role in developmental processes viz. seed maturation, circadian rhythm (Alonso et al.2009; Wolfgang et al.2018). Interaction among C group members, *i.e.* GrbZIP9, GrbZIP10 and GrbZIP25 and S group members GrbZIP 2, GrbZIP53 predicting role in starvation adaptation (Wolfgang et al.2018).

NLS sequence analysis of *G. raimondii* bZIP proteins

NLS sequences from bZIP proteins of *G. raimondii* and *Arabidopsis thaliana* are mentioned in Table 4, some of the NLS sequences are sharing similarity among both species as well as in subfamily members. Subfamily C, G and S bZIP

proteins are sharing similar NLS sequences.

A typical example of bZIP characterization

A 3292bp bZIP17 gene, along with 5'UTR and 3'UTR was isolated from cotton and over-expressed in rice. Transgenic T_3 rice plants were characterized under salt and water stress. Higher transcript level of cotton bZIP17 was observed after 6 hrs of salt stress and non-significant change in the expression was observed under water stress, data not shown (Poster presentation in 5th "Plant Genetics and Genomics: Germplasm to Genome Engineering" conference, 2019; http://glostem.com/select/upload_files/718_1_Plant%20Genetics%20and%20Genomics%202019%20Proceedings.pdf).

Transcriptome analysis of *G. arboreum*, *G. herbaceum* and *G. hirsutum* under water stress, salt stress and temperature stress revealed into differential expression of number of bZIP genes viz. bZIP11, bZIP17, bZIP37, bZIP47, bZIP53, bZIP55 and bZIP60 etc. (Yao et al. 2011; Ranjan et al. 2012; Zhang et al. 2013; Peng et al. 2014; Zhang et al. 2018; Hasan et al. 2019; Wang et al. 2020).

Discussion

We have identified, structurally characterized and classified in detail the important plant transcription factor family of bZIP proteins, in three genomes of cotton. This information can be used in further crop improvement efforts. Phylogenetic analysis of *G. hirsutum, G. arboreum and G. raimondii* with *Arabidopsis* bZIP proteins shows close relation among functionally similar subfamilies according to their occurrence in same clades, thus suggesting bZIP family proteins are conserved in crucial biological functions in two dicot species. Even plant bZIP proteins are functionally similar in monocot and dicot species, *G. arboreum* and *G. raimondii* bZIPs are phylogenetically analyzed with *Oryza* bZIPs, data not shown. Among *Gossypium* bZIPs, subfamily A and S are representing maximum bZIP genes, subfamily S members are smaller in length, with one or two exons and subfamily B protein is the longest one. Subfamily B, H, J and K of *Gossypium* bZIPs are single member families. *Gossypium* bZIPs Subfamily A, B, C, D, J, K and S members gene orthologs in different crop species are reported to function crucially in biotic-abiotic stress tolerance and starvation adaptation.

During discerning process of Gossypium bZIPs multiple bZIP genes with similar standard designation are observed, possibly due to gene duplication event. bZIP gene duplication may have happen during speciation of dicot species, speciation of Gossypium sub-genome species from common ancestor and during allotetraploidization. Whole-genome duplication in *G. arboreum* and *G .raimondii* before speciation is reported in previous reports. The findings from this study indicate that *Gossypium* bZIP proteins are important in stress adaptation, developmental processes and need further evaluation for trait development like fiber quality, yield and stress tolerance in cotton, which is beyond the scope of this study and should be undertaken by future researchers based on present bioinformatics findings.

Conclusion

This study analyzed, annotated and phylogenetically classified bZIP proteins from cultivated cotton species *G. hirsutum* along with two sub-genome species *G. arboreum* and *G. raimondii*. Cotton bZIPs are classified into twelve subfamilies and eight subgroups. bZIP gene duplications are observed in all three cotton species. We have identified conserved functional motifs among different subfamilies of cotton bZIP proteins and correlated for the prediction of function along with reported function. Explored BRLZ domain structural analysis of *G. raimondii* bZIPs will be useful in further basic characterization of bZIP proteins of cultivated cotton species *G. hirsutum*. STRING protein interaction analysis of *G. raimondii* bZIPs resulted in prediction of interactions among A- D, B-K and C-S subfamily members.

Phylogenetic analysis of this study will certainly help in the selection of specific cotton bZIP genes according to the close alignment with *Arabidopsis* orthologs or sub-genome homolog for functional characterization.

List Of Abbreviations

ABF: ABRE binding factors

ABI5: ABA insensitive 5

ABRE: ABA-responsive element

AREB: ABA responsive element- binding protein

CAPS: Cleaved amplified polymorphisms

dCAPS: Derived-CAPS

CPRF2: Common plant regulatory factor 2

GBF: G-box binding factor

NPR1 (non expressor of pathogenesis-related genes1)

PHY: Phytochrome

Declarations

Funding

None

Conflict of Interest

The authors declare that they have no conflict of interest.

Availability of data and material

All the data analyzed in this study is available in the manuscript and supplementary files.

Code availability

Not applicable.

Authors' contributions

VK   initiated, designed and implemented the study, data analysis, drafted the MS

AB conceived the study, data analysis and carefully edited the MS

BC directed the study, carefully edited the MS

RS edited the MS

All authors read and approved the final manuscript

Ethics approval

Not applicable.

Consent to participate

Not applicable.

Consent for publication

Authors would like to thank Mahyco Private Limited for allowing to do work and publishing this article.

SUPPLEMENTARY MATERIAL

1] Supplementary file 1 – Xls. (Online resource 1)

All identified bZIP gene ID, exon number, CDS length, protein length and protein sequences of *G. hirsutum*, *G. arboreum* and *G. raimondii*. *G. hirsutum*, *G. arboreum* and *G. raimondii* bZIP34, bZIP61, bZIP69 and bZIP76 protein alignment.

2] MEME motif analysis MAST files of *G. hirsutum*, *G. arboreum* and *G. raimondii* bZIP proteins

3] Gamma distribution phylogenetic tree

Footnotes

1. https://www.cottongen.org/
2. https://www.arabidopsis.org/
3. https://www.megasoftware.net/
4. http://meme-suite.org/
5. http://smart.embl-heidelberg.de/
6. https://pfam.xfam.org/search/sequence/
7. https://www.ebi.ac.uk/interpro/
8. https://string-db.org/
9. http://nls-mapper.iab.keio.ac.jp/
10. http://planttfdb.gao-lab.org/
11. http://ministryoftextiles.gov.in/
12. https://www.statista.com/

References

Abdurakhmonov IY(2017) Cotton PHYA1 RNAi improves fiber quality root elongation flowering maturity and yield potential in Gossypium hirsutum L. U.S. Patent No US20130227723A1.

Abe M, Kobayashi Y, Yamamoto S, Daimon Y, Yamaguchi A, Ikeda Y, Ichinoki H, Notaguchi M, Goto K, Araki T(2005) FD a bZIP protein mediating signals from the floral pathway integrator FT at the shoot apex. Science (New York N.Y.) 309:1052–1056.

Alonso R, Oñate-Sánchez L, Weltmeier F, Ehlert A, Diaz I, Dietrich K, Vicente-Carbajosa J, Dröge-Laser W(2009) A pivotal role of the basic leucine zipper transcription factor bZIP53 in the regulation of Arabidopsis seed maturation gene expression based on heterodimerization and protein complex formation. The Plant cell 21: 1747–1761. https://doi.org/10.1105/tpc.108.062968

Assunção AG, Herrero E, Lin YF, Huettel B, Talukdar S, Smaczniak C, Immink RG, Van Eldik M, Fiers M, Schat H, Aarts MG (2010) Arabidopsis thaliana transcription factors bZIP19 and bZIP23 regulate the adaptation to zinc deficiency. Pro of the Nat Acad of Sci USA 107: 10296–10301. https://doi.org/10.1073/pnas.100478810

Azeem F, Tahir H, Ijaz U, Shaheen T(2020) A genome-wide comparative analysis of bZIP transcription factors in *G. arboreum* and *G. raimondii* (Diploid ancestors of present-day cotton). Physiol Mol Biol Plants 26: 433–444. https://doi.org/10.1007/s12298-020-00771-9

Bailey TL, Boden M, Buske FA, Frith M, Grant CE, Clementi L, Ren J, Li WW, Noble WS (2009) MEME SUITE: tools for motif discovery and searching. Nucleic Acids res 37: W202–W208. https://doi.org/10.1093/nar/gkp335

Chen ZJ, Sreedasyam A, Ando A, Song Q, De Santiago LM, Hulse-Kemp AM, Ding M, Ye W, Kirkbride RC, Jenkins J, Plott C, Lovell J, Lin YM, Vaughn R, Liu B, Simpson S, Scheffler BE, Wen L, Saski CA, Grover CE, Schmutz J (2020) Genomic diversifications of five *Gossypium* allopolyploid species and their impact on cotton improvement. Nat Genet 52: 525–533. https://doi.org/10.1038/s41588-020-0614-5

Dietrich K, Weltmeier F, Ehlert A, Weiste C, Stahl M, Harter K, Dröge-Laser W (2011) Heterodimers of the Arabidopsis transcription factors bZIP1 and bZIP53 reprogram amino acid metabolism during low energy stress. The Plant cell 23:381–395. https://doi.org/10.1105/tpc.110.075390

Dröge-Laser W, Snoek BL, Snel B, Weiste C (2018) The Arabidopsis bZIP transcription factor family-an update. Curr Opin Plant Biol 45: 36-49. https://doi.org/10.1016/j.pbi.2018.05.001

Dröge-Laser W, Weiste C (2018) The C/S1 bZIP Network: A Regulatory Hub Orchestrating Plant Energy Homeostasis. Trends Plant Sci 23: 422-433. doi: 10.1016/j.tplants.2018.02.003 PMID: 29525129

El-Gebali S, Mistry J, Bateman A, Eddy SR, Luciani A, Potter SC, Qureshi M, Richardson LJ, Salazar GA, Smart A, Sonnhammer ELL, Hirsh L, Paladin L, Piovesan D, Tosatto SCE, Finn RD (2019) The Pfam protein families database in 2019. Nucleic Acids Res 47(D1):D427-D432. doi: 10.1093/nar/gky995 PMID: 30357350 PMCID: PMC6324024

Evens NP, Buchner P, Williams LE, Hawkesford MJ (2017) The role of ZIP transporters and group F bZIP transcription factors in the Zn-deficiency response of wheat (Triticum aestivum). Plant J. 92:291-304. https://doi.org/10.1111/tpj.13655

Ezer D, Shepherd SJK, Brestovitsky A, Dickinson P, Cortijo S, Charoensawan V, Box MS, Biswas S, Jaeger KE, Wigge PA (2017) The G-Box Transcriptional Regulatory Code in Arabidopsis. Plant Physiol 175: 628-640. doi: 10.1104/pp.17.01086

Fan W, Dong X(2002) In vivo interaction between NPR1 and transcription factor TGA2 leads to salicylic acid-mediated gene activation in Arabidopsis. Plant Cell 14: 1377-1389. https://doi.org/10.1105/tpc.001628

Fu ZQ, Dong X (2013) Systemic acquired resistance: turning local infection into global defense. Annu Rev Plant Biol 64: 839-863. https://doi.org/10.1146/annurev-arplant-042811-105606

Gatz C (2013) From Pioneers to Team Players: TGA Transcription Factors Provide a Molecular Link Between Different Stress Pathways. Mol Plant-Microbe Inter 26:151-59. https://doi.org/10.1094/MPMI-04-12-0078-IA

Gibalová A, Renák D, Matczuk K, Dupl'áková N, Cháb D, Twell D, Honys D (2009) AtbZIP34 is required for Arabidopsis pollen wall patterning and the control of several metabolic pathways in developing pollen. Plant molecular biology 70: 581–601. https://doi.org/10.1007/s11103-009-9493-y

Gibalová A, Steinbachová L, Hafidh S, Bláhová V, Gadiou Z, Michailidis C, Műller K, Pleskot R, Dupľáková N, Honys D (2017) Characterization of pollen-expressed bZIP protein interactions and the role of ATbZIP18 in the male gametophyte. Plant Reprod 30: 1-17. doi: 10.1007/s00497-016-0295-5 PMID: 27896439

Gupta S, Stamatoyannopoulos JA, Bailey TL, Noble WS (2007) Quantifying similarity between motifs. Genome Biol 8: R24. doi: 10.1186/gb-2007-8-2-r24

Gutsche N, Holtmannspötter M, Maß L, O'Donoghue M, Busch A, Lauri A, Schubert V, Zachgo S (2017) Conserved redox-dependent DNA binding of ROXY glutaredoxins with TGA transcription factors. Plant direct 1: e00030. https://doi.org/10.1002/pld3.30

Hanson J, Hanssen M, Wiese A, Hendriks MM, Smeekens S (2008) The sucrose regulated transcription factor bZIP11 affects amino acid metabolism by regulating the expression of ASPARAGINE SYNTHETASE 1 and PROLINE DEHYDROGENASE2. The plant Journal 53: 935-49. https://doi.org/10.1111/j.1365-313X.2007.03385.x

Hasan MM, Ma F, Islam F, Sajid M, Prodhan ZH, Li F, Shen H, Chen Y, Wang X (2019) Comparative Transcriptomic Analysis of Biological Process and Key Pathway in Three Cotton (Gossypium spp.) Species Under Drought Stress. Int J Mol Sci 20:2076. https://doi.org/10.3390/ijms20092076

Hendrix B, Stewart JM (2005) Estimation of the nuclear DNA content of gossypium species. Ann Bot 95: 789-797. doi:10.1093/aob/mci078

Howell SH (2013) Endoplasmic Reticulum Stress Responses in Plants. Ann Rev Plant Bio 64: 477-99. https://doi.org/10.1146/annurev-arplant-050312-120053

Humbert S, Zhong S, Deng Y, Howell SH, Rothstein SJ (2012) Alteration of the bZIP60/IRE1 pathway affects plant response to ER stress in Arabidopsis thaliana. PLoS One 7: e39023. doi:10.1371/journal.pone.0039023

Inaba S, Kurata R, Kobayashi M, Yamagishi Y, Mori I, Ogata Y, Fukao Y (2015) Identification of putative target genes of bZIP19 a transcription factor essential for Arabidopsis adaptation to Zn deficiency in roots. Plant J 84: 323-34. doi: 10.1111/tpj.12996. PMID: 26306426

Iwata Y, Fedoroff NV, Koizumi N (2008) Arabidopsis bZIP60 is a proteolysis-activated transcription factor involved in the endoplasmic reticulum stress response. Plant Cell 20: 3107-3121.

Jakoby M, Weisshaar B, Dröge-Laser W, Vicente-Carbajosa J, Tiedemann J, Kroj T, Parcy F, & bZIP Research Group (2002) bZIP transcription factors in Arabidopsis. Trends in plant science 7: 106–111. https://doi.org/10.1016/s1360-1385(01)02223-3

Jin J, Tian F, Yang DC, Meng YQ, Kong L, Luo J, Gao G (2017) PlantTFDB 4.0: toward a central hub for transcription factors and regulatory interactions in plants. Nucleic Acids Res 45: D1040–D1045. https://doi.org/10.1093/nar/gkw982

Khanale V, Bhattacharya A, Parkhi V, Pathak R, Satpute R, Char B (2019) Genome Wide in-Silico Analysis of Cotton bZIP Transcription Factors and Over-Expression of Cotton bZIP17 in Rice for Abiotic Stress Tolerance. Poster presentation in 5th International Conference on Plant Genetics & Genomics New Delhi India October 17-18.

Kosugi S, Hasebe M, Entani T, Takayama S, Tomita M, Yanagawa H (2008) Design of peptide inhibitors for the importin α/β nuclear import pathway by activity-based profiling. Chem Biol 15: 940-949.

Kosugi S, Hasebe M, Tomita M, Yanagawa H (2009) Systematic identification of yeast cell cycle-dependent nucleocytoplasmic shuttling proteins by prediction of composite motifs. Proc Natl Acad Sci USA 106: 10171-10176.

Kosugi S, Hasebe M, Matsumura N, Takashima H, Miyamoto-Sato E, Tomita M, Yanagawa H (2009) Six classes of nuclear localization signals specific to different binding grooves of importin α. J Biol Chem 284: 478-485.

Kumar S, Stecher G, Li M, Knyaz C, Tamura K (2018) MEGA X: Molecular Evolutionary Genetics Analysis across computing platforms. Mol Bio Evo 35: 1547-154. https://doi.org/10.1093/molbev/msy096

Kushanov FN, Pepper AE, Yu JZ, Buriev ZT, Shermatov SE, Saha S, Ulloa M, Jenkins JN, Abdukarimov A, Abdurakhmonov IY. Development genetic mapping and QTL association of cotton PHYA PHYB and HY5-specific CAPS and dCAPS markers. BMC genetics 17:141. https://doi.org/10.1186/s12863-016-0448-4

Li F, Fan G, Wang K, Sun F, Yuan Y, Song G, Li Q, Ma Z, Lu C, Zou C, Chen W, Liang X, Shang H, Liu W, Shi C, Xiao G, Gou C, Ye W, Xu X, Zhang X, Yu S (2014) Genome sequence of the cultivated cotton *Gossypium arboreum*. Nat Genet 46: 567–72. https://doi.org/10.1038/ng.2987

Liang C, Meng Z, Meng Z, Malik W, Yan R, Lwin KM, Lin F, Wang Y, Sun G, Zhou T, Zhu T, Li J, Jin S, Guo S, Zhang R (2016) GhABF2 a bZIP transcription factor confers drought and salinity tolerance in cotton (Gossypium hirsutum L.). Scientific reports 6: 35040. https://doi.org/10.1038/srep35040

Lindemose S, O'Shea C, Jensen MK, Skriver K (2013) Structure function and networks of transcription factors involved in abiotic stress responses. Int J Mol Sci 14: 5842-5878. https://doi.org/10.3390/ijms14035842

Skubacz A, Daszkowska-Golec A, Szarejko I (2016) The Role and Regulation of ABI5 (ABA-Insensitive 5) in Plant Development Abiotic Stress Responses and Phytohormone Crosstalk. Front Plant Sci 7: 1884. doi: 10.3389/fpls.2016.01884.

Liu JX, Srivastava R, Howell SH (2008) Stress-induced expression of an activated form of AtbZIP17 provides protection from salt stress in Arabidopsis. Plant Cell Environ 31: 1735-1743. https://doi.org/10.1111/j.1365-3040.2008.01873.x

Ma J, Hanssen M, Lundgren K, Hernández L, Delatte T, Ehlert A, Liu CM, Schluepmann H, Dröge-Laser W, Moritz T, Smeekens S, Hanson J (2011) The sucrose-regulated Arabidopsis transcription factor bZIP11 reprograms metabolism and regulates trehalose metabolism. The New phytologist 191:733–745. https://doi.org/10.1111/j.1469-8137.2011.03735.x

Mitchell AL, Attwood TK, Babbitt PC, Blum M, Bork P, Bridge A, Brown SD, Chang HY, El-Gebali S, Fraser MI, Gough J, Haft DR, Huang H, Letunic I, Lopez R, Luciani A, Madeira F, Marchler-Bauer A, Mi H, Natale DA, Necci M, Nuka G, Orengo C, Pandurangan AP, Paysan-Lafosse T, Pesseat S, Potter SC, Qureshi MA, Rawlings ND, Redaschi N, Richardson LJ, Rivoire C, Salazar GA, Sangrador-Vegas A, Sigrist CJA, Sillitoe I, Sutton GG, Thanki N, Thomas PD, Tosatto SCE, Yong SY, Finn RD (2019) InterPro in 2019: improving coverage, classification and access to protein sequence annotations. Nucleic Acids Res 47 :D351-D360. doi: 10.1093/nar/gky1100

Nazri AZ, Griffin JHC, Peaston KA, Alexander-Webber DGA, Williams LE (2017) F-group bZIPs in barley-a role in Zn deficiency. Plant Cell Environ 40:2754-2770. https://doi.org/10.1111/pce.13045

Nijhawan A, Jain M, Tyagi AK, Khurana JP (2008) Genomic survey and gene expression analysis of the basic leucine zipper transcription factor family in rice. Plant Physiol 146:333-350. https://doi.org/10.1104/pp.107.112821

Peng Z, He S, Gong W, Sun J, Pan Z, Xu F, Lu Y, Du X (2014) Comprehensive analysis of differentially expressed genes and transcriptional regulation induced by salt stress in two contrasting cotton genotypes. BMC genomics 15: 760. https://doi.org/10.1186/1471-2164-15-760

Pyo H, Demura T, Fukuda H (2006) Vascular cell expression patterns of Arabidopsis bZIP group I genes. Plant Biotechnol 23: 497-501. https://doi.org/10.5511/plantbiotechnology.23.497

Ranjan A, Nigam D, Asif MH, Singh R, Ranjan S, Mantri S, Pandey N, Trivedi I, Rai KM, Jena SN, Koul B, Tuli R, Pathre UV, Sawant SV (2012) Genome wide expression profiling of two accession of G. herbaceum L. in response to drought. BMC genomics 13:94. https://doi.org/10.1186/1471-2164-13-94

Rhee SY, Beavis W, Berardini TZ, Chen G, Dixon D, Doyle A, Garcia-Hernandez M, Huala E, Lander G, Montoya M, Miller N, Mueller LA, Mundodi S, Reiser L, Tacklind J, Weems DC, Wu Y, Xu I, Yoo D, Yoon J, Zhang P (2003) The Arabidopsis Information Resource (TAIR): a model organism database providing a centralized curated gateway to Arabidopsis biology research materials and community. Nucleic Acids Res 2003 31: 224–228. https://doi.org/10.1093/nar/gkg076

Rolly NK, Imran QM, Lee IJ, Yun BW(2020) Salinity Stress-Mediated Suppression of Expression of Salt Overly Sensitive Signaling Pathway Genes Suggests Negative Regulation by AtbZIP62 Transcription Factor in Arabidopsis thaliana. Int J Mol Sci. 21:1726. https://doi.org/10.3390/ijms21051726

Schwarz R, Dayhoff M(1979) Matrices for Detecting Distant Relationships. In: Dayhoff M, Ed, Atlas of Protein Sequences. National Biomedical Research Foundation 353-358.

Schultz J, Copley RR, Doerks T, Ponting CP, Bork P (2000) SMART: a web-based tool for the study of genetically mobile domains. Nucleic Acids Res 28: 231-4. doi: 10.1093/nar/28.1.231

Shen H, Cao K, Wang X (2007) A conserved proline residue in the leucine zipper region of AtbZIP34 and AtbZIP61 in Arabidopsis thaliana interferes with the formation of homodimer. Biochem Biophys Res Commun 362: 425–430. doi: 10.1016/j.bbrc.2007.08.026.

Sirichandra C, Davanture M, Turk BE, Zivy M, Valot B, Leung J, Merlot S (2010) The Arabidopsis ABA-activated kinase OST1 phosphorylates the bZIP transcription factor ABF3 and creates a 14-3-3 binding site involved in its turnover. PLoS One 5: e13935. doi: 10.1371/journal.pone.0013935 PMID: 21085673

Sonnhammer EL, Eddy SR, Durbin R (1997) Pfam: a comprehensive database of protein domain families based on seed alignments. Proteins 28: 405-20. doi: 10.1002/(sici)1097-0134(199707)28:3<405::aid-prot10>3.0.co;2-l

Szklarczyk D, Gable AL, Lyon D, Junge A, Wyder S, Huerta-Cepas J, Simonovic M, Doncheva NT, Morris JH, Bork P, Jensen LJ, Mering CV (2019) STRING v11: protein-protein association networks with increased coverage supporting functional discovery in genome-wide experimental datasets. Nucleic Acids Res. 2019 47:D607-D613. doi: 10.1093/nar/gky1131

Ullah I, Magdy M, Wang L, Liu M, Li X (2019) Genome-wide identification and evolutionary analysis of TGA transcription factors in soybean. Sci Rep 1: 11186. doi:10.1038/s41598-019-47316-z

Wang K, Wang Z, Li F, Ye W, Wang J, Song G, Yue Z, Cong L, Shang H, Zhu S, Zou C, Li Q, Yuan Y, Lu C, Wei H, Gou C, Zheng Z, Yin Y, Zhang X, Liu K, Wang B, Song C, Shi N, Kohel RJ, Percy RG, Yu JZ, Zhu YX, Wang J, Yu S (2012) The draft genome of a diploid cotton Gossypium raimondii. Nat Genet 44:1098-103. doi: 10.1038/ng.2371

Wang B, Liu C, Zhang D, He C, Zhang J, Li Z. Effects of maize organ-specific drought stress response on yields from transcriptome analysis (2019) BMC Plant Biol 19: 335. doi: 10.1186/s12870-019-1941-5

Wang Z, Yan L, Wan L, Huai D, Kang Y, Shi L, Jiang H, Lei Y, Liao B (2019) Genome-wide systematic characterization of bZIP transcription factors and their expression profiles during seed development and in response to salt stress in peanut. BMC Genomics 20: 51. doi: 10.1186/s12864-019-5434-6

Wang X, Lu X, Malik WA, Chen X, Wang J, Wang D, Wang S, Chen C, Guo L, Ye W (2020) Differentially expressed bZIP transcription factors confer multi-tolerances in Gossypium hirsutum L. Int J Biol Macromol 146: 569-578. doi: 10.1016/j.ijbiomac.2020.01.013

Weltmeier F, Rahmani F, Ehlert A, Dietrich K, Schütze K, Wang X, Chaban C, Hanson J, Teige M, Harter K, Vicente-Carbajosa J, Smeekens S, Dröge-Laser W(2009) Expression patterns within the Arabidopsis C/S1 bZIP transcription factor network: availability of heterodimerization partners controls gene expression during stress response and development. Plant Mol Biol 69: 107-19. doi: 10.1007/s11103-008-9410-9

Yao D, Zhang X, Zhao X, Liu C, Wang C, Zhang Z, Zhang C, Wei Q, Wang Q, Yan H, Li F, Su Z (2011) Transcriptome analysis reveals salt-stress-regulated biological processes and key pathways in roots of cotton (Gossypium hirsutum L.). Genomics 98: 47-55. doi: 10.1016/j.ygeno.2011.04.007

Yoshida T, Fujita Y, Maruyama K, Mogami J, Todaka D, Shinozaki K, Yamaguchi-Shinozaki K (2015) Four Arabidopsis AREB/ABF transcription factors function predominantly in gene expression downstream of SnRK2 kinases in abscisic acid signalling in response to osmotic stress. Plant Cell Environ 38:35-49. doi: 10.1111/pce.12351

Yoshida T, Mogami J, Yamaguchi-Shinozaki K (2015) Omics Approaches Toward Defining the Comprehensive Abscisic Acid Signaling Network in Plants. Plant Cell Physiol 56: 1043-52. doi: 10.1093/pcp/pcv060

Yu J, Jung S, Cheng CH, Ficklin SP, Lee T, Zheng P, Jones D, Percy RG, Main D (2014) CottonGen: a genomics genetics and breeding database for cotton research. Nucleic Acids Res 42: D1229-36. doi: 10.1093/nar/gkt1064

Zhang B, Chen X, Lu X, Shu N, Wang X, Yang X, Wang S, Wang J, Guo L, Wang D, Ye W(2018) Transcriptome Analysis of Gossypium hirsutum L. Reveals Different Mechanisms among NaCl NaOH and Na2CO3 Stress Tolerance. Scientific reports 8: 13527. https://doi.org/10.1038/s41598-018-31668-z

Zhang X, Yao D, Wang Q, Xu W, Wei Q, Wang C, Liu C, Zhang C, Yan H, Ling Y, Su Z, Li F(2013) mRNA-seq analysis of the Gossypium arboreum transcriptome reveals tissue selective signaling in response to water stress during seedling stage. PLoS One 8: e54762. doi: 10.1371/journal.pone.0054762

Zhang YN, Cai DR, Huang XZ (2016) Identification of BZIP Protein Family in Gossypium arboreum and Tissue Expression Analysis of GaFDs Genes. Acta Agro Sin 42: 832-43.

Tables

Due to technical limitations the Tables are available as a download in the Supplementary Files.

Figures

Figure 1

A Phylogenetic tree construction of G. hirsutum, G. arboreum, G.raimondii and Arabidopsis thaliana bZIPs, subfamilies are colour coded. Note: Gamma distributed Phylogenetic tree is also submitted separately in supplementary material.

Figure 2

A typical example of MEME motif analysis of G. raimondii bZIPs, bZIP1 domain, DOG1 domain, tetraticopeptide repeat motif, MYND domain, Homeobox associated leucine zipper domain are marked by arrow.
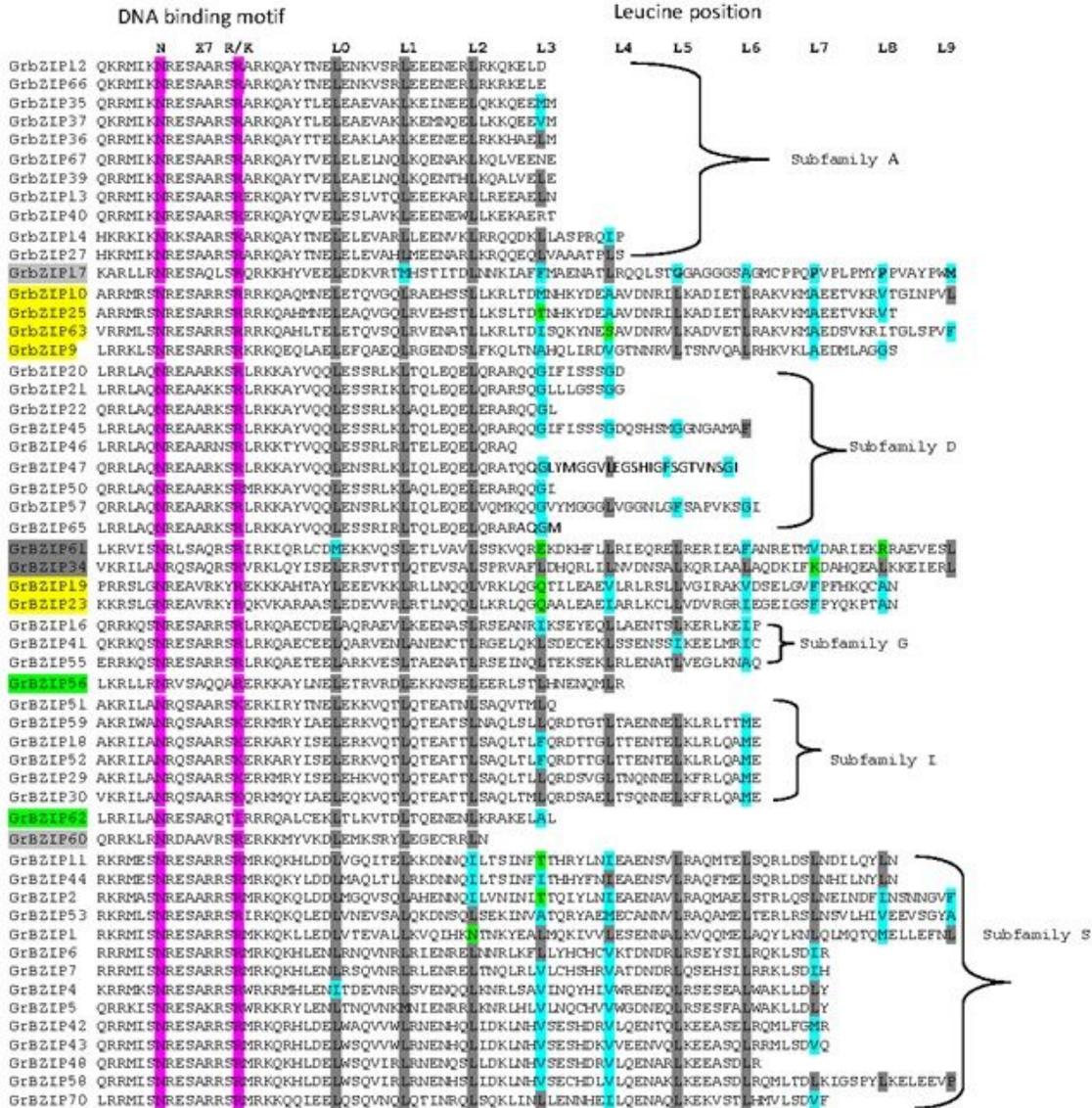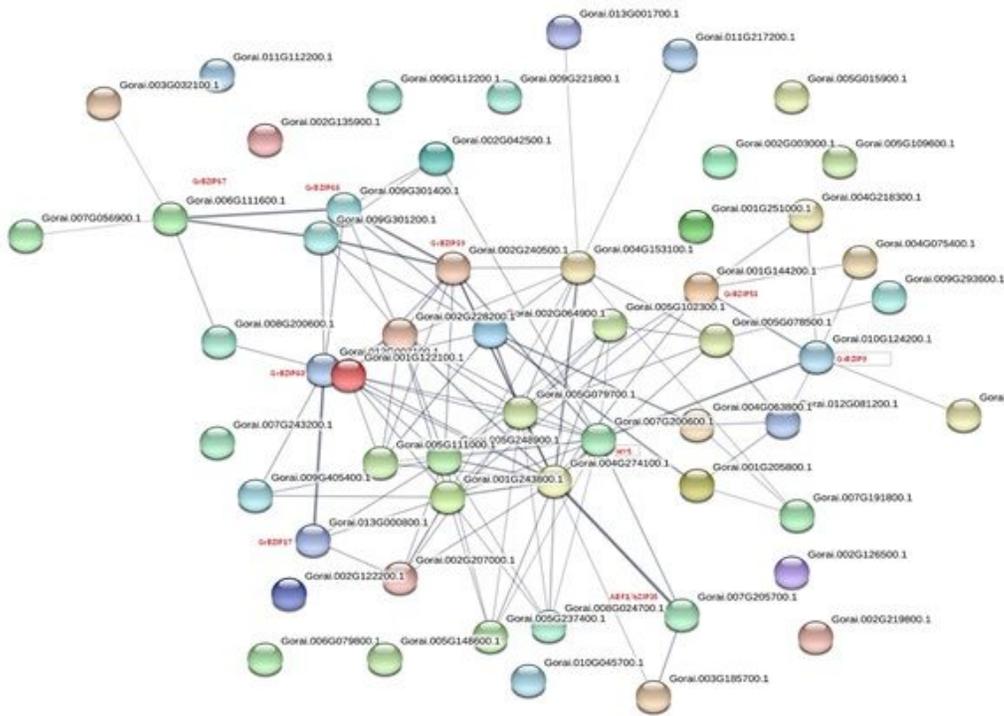
Figure 3

Leucine heptad repeats variation in G. raimondii bZIPs: basic region motif N-X7-R/K is highlighted in magenta colour, leucine position in luecine heptad motif is highlighted in gray, hydrophobic amino acids other than leucine are highlighted in blue and polar amino acids are highlighted in green

Figure 4

STRING protein interaction network of G. raimondii bZIPs

Supplementary Files

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

- Tables.pdf
- Onlineresource1.xlsx
- MASTfor4Ghand1GrremainingbZIPs1.html
- MASTGrbZIPsselectedonly1stGeneofeachbZIPs2.html
- Supplimentarygammadistributionphylotree.png