*Pak. J. Bot.*, 53(2): 483-492, 2021. DOI: http://dx.doi.org/10.30848/PJB2021-2(32)

# IDENTIFICATION, CHARACTERIZATION AND INTERACTION STUDIES OF *Di19-2* GENE FROM *GOSSYPIUM ARBOREUM*

**MAHMOOD-UR-RAHMAN[1], TAYYABA SHAHEEN[1*], PARWSHA ZAIB[1], MUHAMMAD FAZAL ABBAS[1], MEHBOOB-UR-RAHMAN[2] AND ANJUMAN ARIF[3]**

*[1]Department of Bioinformatics and Biotechnology, Government College University, Faisalabad, Allama Iqbal Road, Faisalabad-38000, Pakistan*
*[2]Plant Genomics and Molecular Breeding Laboratory, Agricultural Biotechnology Division, National Institute of Biotechnology and Genetic Engineering, Faisalabad, Pakistan*
*[3]Wheat Biotechnology Group, Nuclear Institute for Agriculture and Biology, Faisalabad, Pakistan*
*[*]Corresponding author's email: tayaba_pgmb@yahoo.com*

**Abstract**

Cotton is the most vital source of fiber and food industry. Limited availability of water results in the limited growth of plants. Drought responsive genes have been explored in several plants to be utilized for overcoming the dilemma of limited availability of water. Over expression of drought responsive genes to develop drought resistant cultivars is a promising strategy to combat drought stress. On exposure to drought, several genes linked to drought are activated and many of them are regulated by transcription factors. Recent developments in bioinformatics have made identification and *In silico* characterization of the genes possible. *Di19* is the drought responsive transcription factor, which is involved in the drought tolerance pathways. This gene has been observed to interact with several other genes in the drought tolerance pathway in the plants. In the present study, this gene was amplified in *Gossypium arboreum* cultivar FDH-786 using primers based on *GhDi19* gene sequence, further sequenced and named as *GaDi19* (Accession No. KP297805). Sequence of *GaDi19* showed highest similarity with *GhDi19* (97%). Protein 3D model for GaDi19 was determined and Ramanchandran plot indicated that 98% residues were in favored region. Further modelling and interaction study revealed that GaDi19 interact with pathogen related PR1, PR2 and PR5. Interaction of GaDi19 with protein encoded by other genes i.e Di19 from other plant species and EREB1 and EREB2 was also determined. Further interaction studies of PR1, PR2 and PR5 revealed activation of AP2/EREBP, DREB1a, DREB1b and WRKY3 to develop the drought tolerance in plants, which is a novel finding and has not been reported earlier. This study will lay a foundation for researchers to get insight into genes responsible for drought and their involvement in the pathways of drought tolerance.

**Key words:** *Di19* gene, *PR* genes, Drought stress, 3 D-Modelling, Docking.

## Introduction

Cotton (*Gossypium* spp.) being a multibillion crop is the best source of natural fiber globally. Pakistan is fourth biggest cotton producer after China, India and USA. Cotton is a summer crop and usually cultivated in sandy soil with warm and humid climates. Cotton is irrigated by occasional rainfalls in Pakistan, however, main need of water requirement for crop is accomplished by supplementary irrigation. Ground water and surface water are the two main favorable and important supplementary irrigation sources (Saeed, 2009). *Gossypium arboreum* (Desi Cotton) has more tolerance to biotic and abiotic factors; therefore, it is best source for drought related studies. *G. arboreum* is also a valuable source for the genomic studies because of having diploid genome and also as a contributor to modern species *G. hirsutum* genome (Shaheen *et al*., 2013).

Cotton is a drought sensitive crop and biotic/abiotic stresses result in its limited growth. Drought is a complex phenomenon, which badly effects physiology of cotton plant (Hanson & Hitz, 1982; Ullah *et al*., 2017). A better understanding of the coping mechanisms to drought stresses would contribute to the crop production under drought conditions and long-term improvement of plant (Ullah *et al*., 2017: Shaheen *et al*., 2018; Ahmad *et al*., 2020). Transcription factors play a major role in combating abiotic stresses in plants via ABA dependent and ABA independent pathways (Tran & Mochida, 2010).

Di19 (dehydration-induced19) are included in a unique family of small to moderate in size proteins. Uniqueness of these proteins is due to presence of two rare Cys2/His2 putative zinc-finger domains, which are not like those of the classical zinc-finger protein (Li *et al*., 2010).

Putative nuclear localization signals (NLS) were present in whole *Di19* family except in *Di19-2*. Localization within the nucleus was reported in the five members of family except in *Di19-2* and *Di19-4*. *Di19* genes were expressed mainly in seedlings, rosettes, roots, flowers, stems and siliques ubiquitously. Dehydration resulted in the activation of the *Di19-1* and *Di19-3* (Gosti *et al*., 1995; Milla *et al*., 2006), and high salt stress was the factor which allowed the higher expression of *Di19-2* and *Di19-4* (Milla *et al*., 2006).

Li *et al*., (2010) functionally characterized the GhDi19-1 and GhDi19-2 from cotton (*G. hirsutum*). Both proteins were found nuclear-localized. Overexpression of these genes in *A. thaliana* under salinity and drought stresses supported their involvement in salinity and drought stress responses.

*PR1*, *PR2* and *PR5* are the genes related to tolerance to pathogen infestation and drought stress (Liu *et al*., 2013). Liu *et al*., (2013) reported the induction of expression of *PR* genes by *Di19* in Arabidopsis under drought stress. Bioinformatics tools are very promising for *Insilico* characterization of genes and interaction studies of proteins with other proteins (Ritchie *et al*., 2017).

We report identification of Di-19 gene from *G. arboreum* in this study. Further interaction of Di-19 protein with PR proteins and other drought related gene products is reported. Interaction study of PR proteins with other drought induced gene products (DREB1a, DREB1b, WRKY3 and AP2/EREBP was also carried out. This study will lay a foundation to interpret role of Di-19 in drought stress resilience.

## Materials and Methods

**Collection of plant material and induction of stress:** Cotton (*G. arboreum*) var. FDH-786 seeds were grown in composite soil in green house of NIBGE (National Institute of Biotechnology and Genetic Engineering), Faisalabad, Pakistan. The day and night temperature of green house was set at 25 ± 2°C, with 50% humidity. One-month old seedlings were exposed to drought by withholding irrigation until 15% weight loss of plants. A sufficient water supply was provided to control plants grown in the soil until collection of samples. Leaf samples from both treated and control plants were collected for RNA isolation.

**Primer designing and data retrieval of drought responsive genes:** To find conserved regions of the Gh Di-19 gene, BLASTn was used. BLAST was run on *Gh Di-19* and Di-19 genes selected from other plant species. The conserved regions among cotton ESTs and previously identified genes in other species were used for primer designing. Primer sequences F 5′-ATGGATGCTGATCC ATGGAC-3′ and R 5′-CCACAATTCTTGATGATGTTTTATGA-3′ (Accession No. GU292050.1) were used in the PCR.

The sequences of 11 stress responsive genes were retrieved from NCBI and the information regarding each gene was also acquired e.g. gene name, accession number, and function of these genes (Table 1). Proteins encoded by these genes were used as ligands to check the further interactions with the receptor (product of *GaDi19-2*).

**RNA isolation:** Method of Verwoerd *et al*., (1989) with some modifications was used to isolate total RNA from leaves of one-month old stressed and control plant seedlings. Quality and Quantity of RNA was assessed using a spectrophotometer (NanoDrop-2000, Thermo Scientific, USA).

The isolated RNA was used for synthesis of cDNA and PCR Amplification. For cDNA synthesis, firstly 1 µL of oligo (dt) primer, 1 µg/5 µL of total RNA, and 6 µL of DEPC treated water, total 12 µL reactions mix was incubated at 70°C for 5 min and instantly chilled on ice. Then 2µL of 10 mM dNTP mixture, 4 µL of 5x reaction buffer, and 1 µL of Riblock TM Ribonuclease inhibitor were added to the mixture by mixing them moderately. The mixture was centrifuged for a short time and incubated at 37°C for 5 min after addition of 1 µL of revert-Aid TM M-MuLV Reverse Transcriptase (Fermentas). Subsequently the mixture was incubated at 42°C for 60 min and then reaction was stopped by heating at 70°C for 10 min and keeping on ice for chilling. Further, for the amplification of single stranded cDNA to double stranded cDNA, gene specific primers were used. For PCR amplification a total volume of 25 µL was comprised of 1.0 µL of cDNA, 12.5 µL of 2X BioMix PCR master mix (Bioline, UK,) and 0.75 µL of 10.0 µM forward and reverse primer (Invitrogen, UK). The GeneAmp PCR system 9700 (Applied Biosystems) was used to carry out amplifications using the following programme: 5 min at 94°C; subsequently 35 cycles of 45 s at 94°C, 60 s at 60°C, and 90 s at 72°C; and lastly 7 min at 72°C for the final elongation.

**Table 1. List of drought responsive genes used in study.**

| Serial No. | Gene/ Protein name | Accession No. | Gene source | Function Of Gene |
|---|---|---|---|---|
| 1. | *GhDi19 1* | GU292049 | Cotton | *GhDi19-1* associated with the stresses like drought and salt. Functioning in ABA is also performed by GHDi19-1 |
| 2. | *GhDi19 2* | GU292050 | Cotton | Desiccation tolerance |
| 3. | *Di19* | NM_104507 | *Arabidopsis thaliana* | Pathogen genes become activated by the activation of *Di19* which resulted in to produce drought resistant varieties |
| 4. | *Di19* | FJ795369 | Wheat | Drought resistant varieties of wheat produced by the high expression of Di19 |
| 5. | *PR1* | AF370026.1 | Brassica | *PR1* gene start to work in response to salicylic acid and then pathogen also results in its activation |
| 6. | *PR2* | AY323485.1 | *Oryza sativa* | Main task is to protect plant and common response to environmental stress |
| 7. | *PR5* | KJ764822.1 | *Triticum aestivum* | *PR5* gene work against fungi reaction and it also show reaction during drought conditions |
| 8. | *WRKY3* | FJ966887.1 | Cotton | Growth of fiber and security mechanism of plant is the function of *WRKY3* and it also has a role in drought |
| 9. | *EREB1* | EF408086.2 | Cotton | This gene involves in the binding of DNA, role in drought |
| 10. | *EREB2* | EU082108.1 | Cotton | Drought tolerance in cotton |
| 11. | *AP2/EREBP* | EU791896.1 | Cotton | Yield and development of transgenic plant obtained by the high expression of gene coding *Ap2/EREBP* |

**Cloning of the isolated gene:** For visualization and analysis of DNA bands, gel electrophoresis using 1.0% (w/v) agarose gels with added ethidium bromide (0.5 µg mL-1) was performed. GeneJET gel extraction kit (Thermo Scientific) was used for elution of single band of estimated size from the agarose gel. Purified PCR products were further ligated in TA cloning plasmid vector pTZ57R (InsTAclone PCR Cloning Kit, Thermo Scientific) following the manufacturer's instructions and then transformation into *E. coli* cells was performed using heat shock method. Selection based on blue/white color was used to identify the transformed colonies.

**Plasmid DNA isolation and genes sequencing:** AxyPrep Plasmid Miniprep Kit was used for plasmid isolation. For confirmation of insert, some plasmids were subjected to PCR. The cloned fragments in pTZ57R plasmid vector were sequenced commercially. Subsequent to sequencing, the homology searches for genes were performed to find homology at NCBI/EMBL database by BLAST search software.

**Sequence analysis and 3D protein structure prediction of Di 19:** Di-19 gene sequences of different species were retrieved using NCBI web portal. Further for the retrieval of amino acid sequence of Di-19, the newly sequenced gene was translated using JustBio translator tool (http://www.justbio.com/). The 3D structure of this amino acid sequence was predicted using MODELLER 9.10 (http://salilab.org/modeller/), a Python based protein modeling software. An appropriate template structure was acquired using NCBI BLASTp server based on the lowest e-value and maximum similarity. Different confirmation softwares were used for determination of the quality and reliability of 3D structure. To confirm the quality of structures, MOLProbity evaluation (http://molprobity.biochem.duke.edu/) and Swiss model online server (http://swissmodel.expasy.org/workspace) were used. The minimum DOPE score was used for selection of best structure. Based on the model, the Ramachandran plot was determined.

**Conserved domain analysis:** Domain prediction of the identified gene was done by using Conserved Domain Online tool available at (http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi). The result was saved in form of images.

**Protein-protein interaction:** On the basis of previous literature and published articles, genes were selected. FASTA format of the selected genes were retrieved from NCBI (http://www.ncbi.nlm.nih.gov/). The genes were translated using the procedure previously described. 3D structures were predicted for these genes using the Phyre2 (http://www.sbg.bio.ic.ac.uk/phyre2/html/page.cgi?id=index). Protein encoded by Ga Di-19 gene was used as receptor and proteins encoded by Di 19 genes from other plant species and PR genes (PR1, PR2, PR 5) were used as ligands in protein-protein interaction studies. The interactions were studied using ClusPro (http://cluspro.bu.edu/login.php). Furthermore, Pymol (http://pymol.org/educational%20v1.74) software was used to analyze the protein-protein interactions.

Residue interaction of PR proteins with other transcription factors was also determined using PyMOL (http://pymol.org/educational%20v1.74). Interaction of PR1, PR2, PR5 with the *DREB1a, DREB1b, WRKY3* and *AP2/EREBP* was studied using ClusPro.

**Amplification of PR genes in *G. arboretum*:** For amplification of PR genes in *G. arboreum* the same protocol was used as described earlier. Primers were designed based on the *A. thaliana* PR gene sequences available on NCBI. Primer sequences for PR1 (F-GCGGTAGGCGTAGGTC, R-TTGGCACATCCGAGTC), for PR2 (F-GCTCTACGG CCCTGAC, R-AAACCGCGTTCTCGAT) and for PR5 (F-GGAGTTCCTCCCGTCA, R-CCGTGGGAGGACAAGT) were used. Amplification products were visualized on 1% agarose gel.

## Results

**Gene identification:** RNA isolation from the young leaves of drought treated G. arboreum plants was done using modified method of Verwoerd *et al*., (1989).

Primer sets designed from *GhDi19* (Accession No. GU292050.1) were used for the amplification of cDNA in *G. arboreum*. After the successful amplification of *Gh Di-19* gene as a control, PCR conditions were optimized for amplification of the gene from the cDNA synthesized from the RNA captured from *G. arboreum* treated with water deficit stress.

After amplification and purification of the amplicon of desired length (Fig. 1), it was cloned into ptz57RT vector and sequenced commercially and the gene sequence was searched for homology using BLAST (Altschul *et al*., 1997). Di 19 gene isolated from *G. arboreum* was found 97 % similar to *G. hirsutum* Di 19 gene accession number (NM_001326734.1) The gene sequence of Di 19 derived from *G. arboreum* was submitted to NCBI. The NCBI Accession number of newly identified *GaDi-19* gene sequence is KP297805.

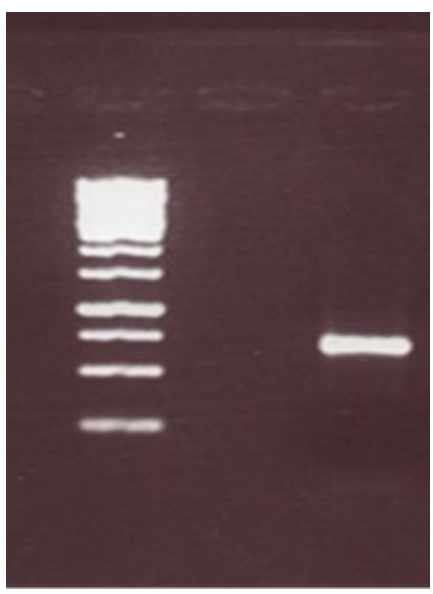


Fig. 1. The amplified fragments of 657bp of Di19-2.

**Data regarding drought responsive genes in plants:** The sequences of 11 stress responsive genes retrieved from NCBI and the information regarding each gene e.g. gene name, accession number, and function of these genes was obtained (Table 1).

**Translation of gene sequence:** Protein sequences of the GaDi19 was obtained, 3D Model of the protein was determined and Ramachandran plot showed that 98% residues were in favored regions (Fig. 2).

**Conserved domain analysis:** Di 19 was found to have two domains, zf-Di19 and Di19_C (Fig 3).

**Interaction study of GaDi19 with other proteins:** GaDi19 was found interacting with Di19 from other plants (*G. hirsutum*, *Arabidopsis thaliana* and *Triticum aestivum*), PR genes, EREB 1 and EREB 2 (Figs. 4,5). Residue interaction of GaDi19 and other proteins was also determined using PyMOL (Supplementary Table 1).

**Modeling and Interactions study of *PR* genes:** Residue interaction of PR proteins with other transcription factors was also determined using PyMOL (Supplementary Table 2, Table 3 and Table 4). Interaction of PR1, PR2, PR5 with the *DREB1a*, *DREB1b*, *WRKY3* and *AP2/EREBP* was studied using ClusPro (Figs. 6, 7, 8).

**Amplification of PR genes in *G. arboretum*:** We could amplify PR2 and PR5 genes in *G. arboreum*. While amplification of PR1 was not successful, even the reaction was repeated many times.

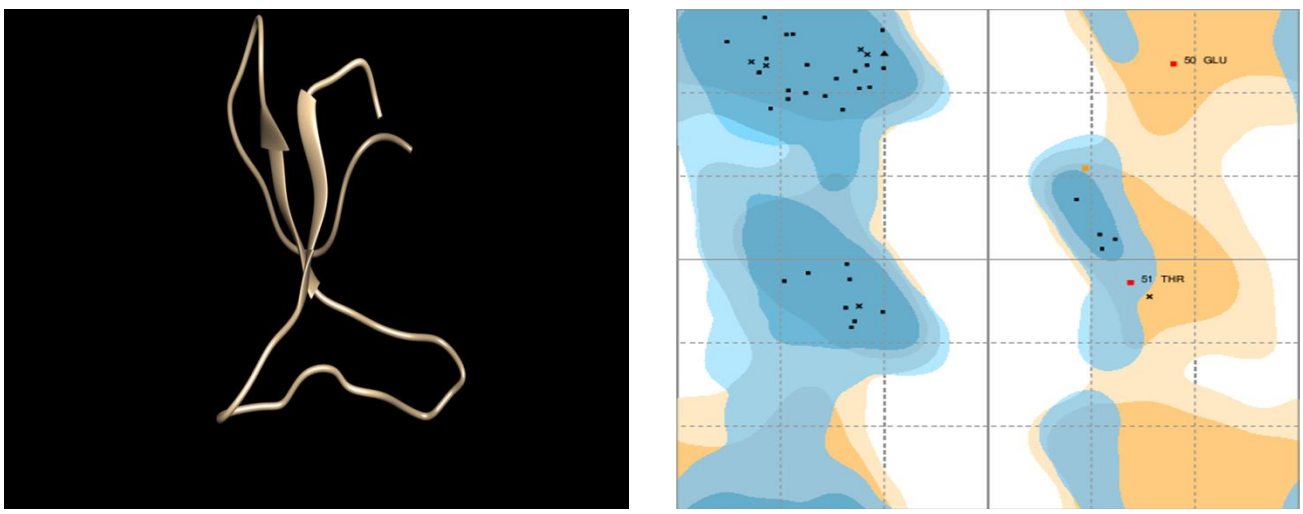



Fig. 2. 3D model of GaDi19 and Ramachandran plot.

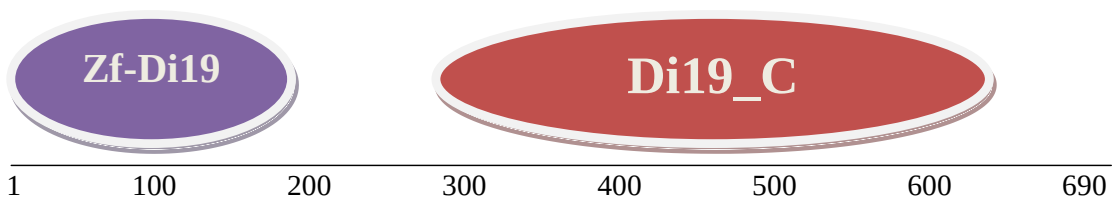



Fig. 3. Conserved domains found in GaDi19 sequence.

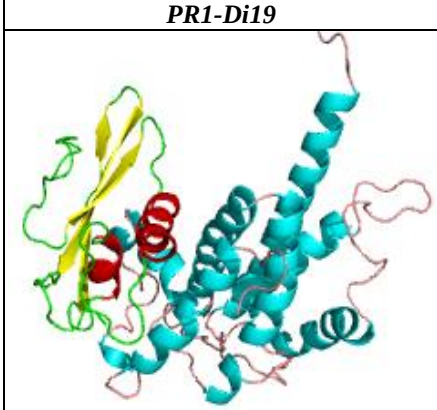
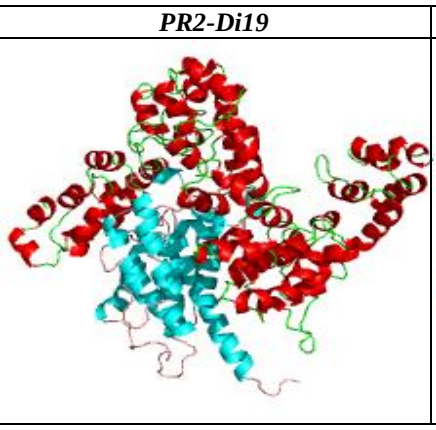
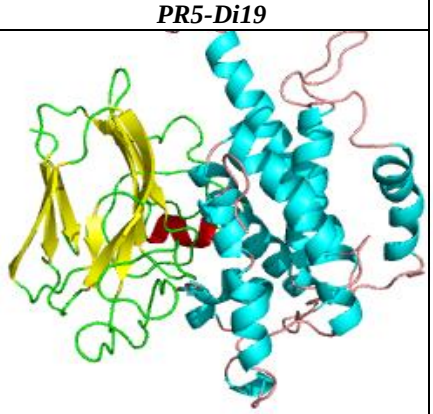
| ***PR1-Di19*** | ***PR2-Di19*** | ***PR5-Di19*** |
|---|---|---|
|  |  |  |

Fig. 4. Interactions of GaDi19 with the PR1, PR2 and PR5.

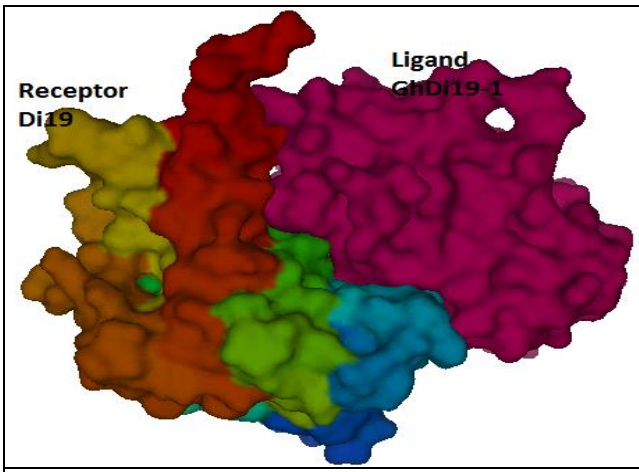



Fig. A. GhDi19 1 and Di19 interaction.

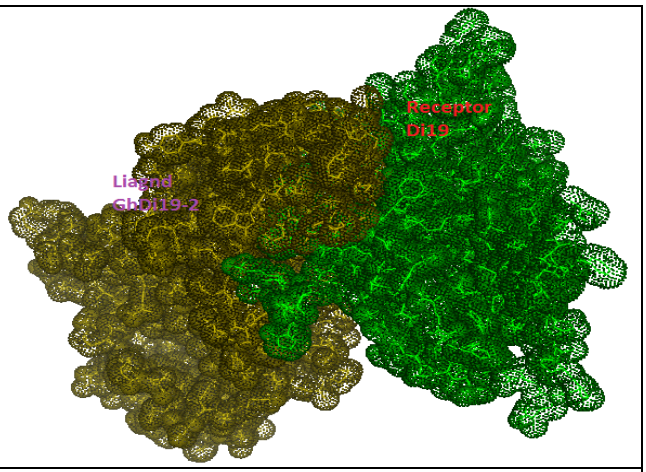



Fig. B. GhDi19 2 and Di19 interaction.

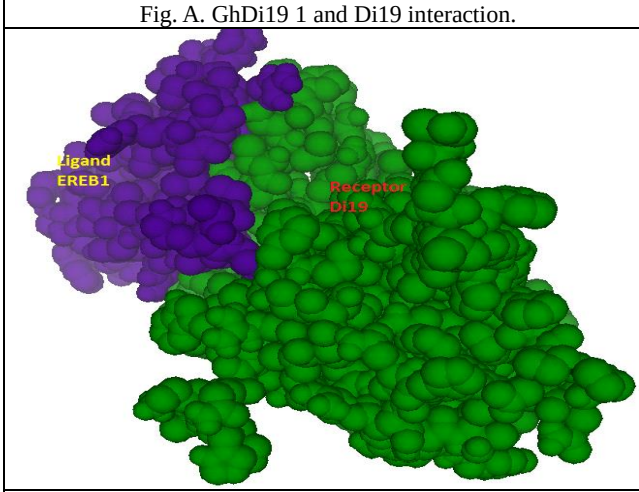



Fig. C. Interaction of EREB1 and Di19.

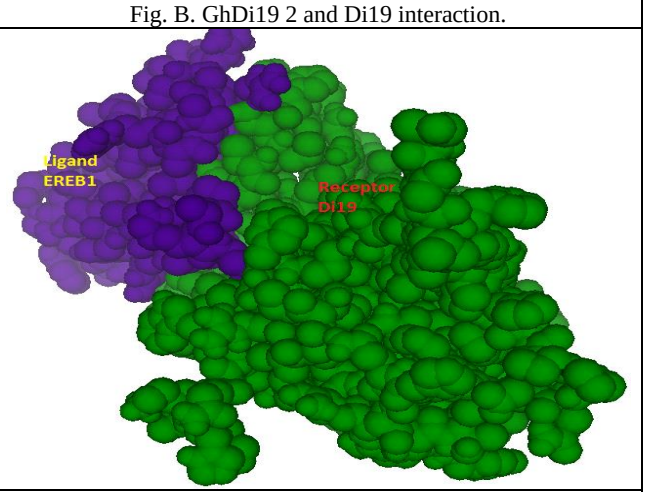



Fig. D. EREB2 and Di19 interaction.

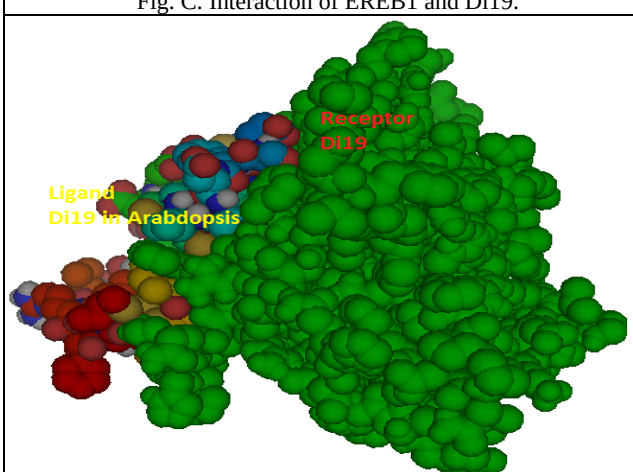



Fig. E. Di19 and Di19 in *A. haliana.*

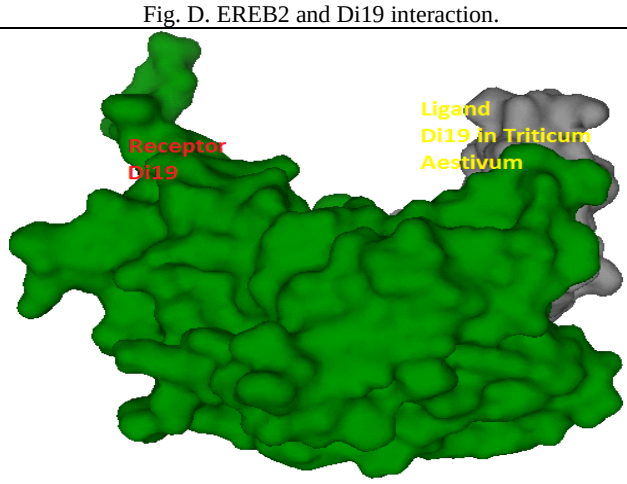



Fig. F. Di19 and Di19 in *T. aestium.*

Fig. 5. Interaction of GaDi19 with Di19 of other plant species.

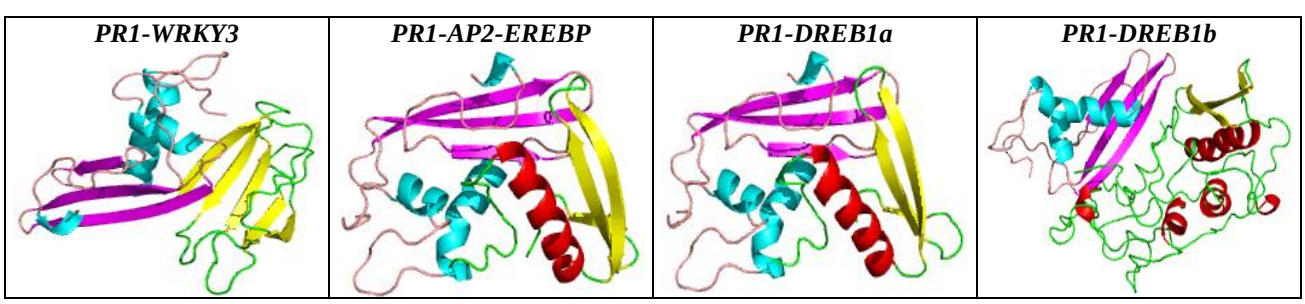



Fig. 6. Interaction models of *PR1* gene with *WRKY3, AP2/EREBP, DREB1a* and *DREB1b.*

**Supplementary material Table 1. Residue interactions of protein encoded by drought induced Di19 gene.**

| Sr. No. | Receptor | Ligand | Residue interaction receptor ligand | |
|---|---|---|---|---|
| 1. | GaDi19 | PR1 (Brassica) | SER-45 | GLY-70 |
| | | | PHE-44 | LYS-71 |
| | | | PHE-44 | LYS-71 |
| | | | LYS-99 | ASN-32 |
| | | | GLN-15 | ASN-42 |
| | | | TRP-8 | ALA-39 |
| | | | LEU-27 | ARG-62 |
| | | | GLU-30 | ARG-62 |
| 2. | GaDi19 | PR2 In (*Oryza sativa*) | ARG-69 | ALA-99 |
| | | | LYS-64 | GLU-492 |
| | | | HIS-175 | GLU-492 |
| | | | HIS-175 | GLU-491 |
| | | | ARG172 | GLU-462 |
| | | | ARG172 | GLU-491 |
| | | | TYR115 | ARG-407 |
| | | | GLU-29 | ARG-407 |
| | | | ALN175 | ARG-427 |
| | | | TYR-39 | ARG-343 |
| | | | GLU-30 | HIS-304 |
| | | | CYS-47 | ASN-210 |
| | | | SER-45 | LEU-211 |
| | | | SER-96 | ASP-200 |
| | | | GLU-97 | GLN-169 |
| | | | ARG201 | HIS-161 |
| | | | ARG201 | TYR-136 |
| 3. | GaDi19 | PR5 In (*Triticum aestivum*) | ARG201 | THR-156 |
| | | | ARG201 | GLY-130 |
| | | | ARG201 | ASP-153 |
| | | | ARG201 | ASN-152 |
| | | | ARG-90 | GLN-149 |
| | | | SER-96 | ASP-119 |
| | | | THR-52 | ARG-96 |
| | | | ALA-55 | ARG-96 |
| | | | ARG-50 | ALA-134 |
| | | | ARG190 | ASP-133 |
| 4. | GaDi19 | GhDi19-1 In Cotton | SER-96 | ASP-4 |
| | | | SER-96 | HIS-94 |
| | | | GLU-97 | SER-5 |
| | | | GLU-97 | TRP-6 |
| | | | HIS-41 | MET-155 |
| | | | PHE-44 | MET-1 |
| | | | SER-45 | GLY-95 |
| | | | LEU-46 | ASP-4 |
| | | | ARG201 | ARG-9 |
| | | | ARG201 | ARG-16 |
| | | | ARG201 | LEU-10 |
| | | | GLU197 | ARG-9 |
| | | | GLU198 | ARG-9 |
| | | | LEU194 | ARG-9 |
| 5. | GaDi19 | GhDi19-2 In Cotton | ARG-75 | ASP-63 |
| | | | HIS-3 | SER-127 |
| | | | ILE-1 | SER-127 |
| | | | ASP-25 | ARG-105 |
| | | | SER-18 | LEU-58 |
| | | | ILE-1 | GLY-123 |
| | | | ILE-1 | ASP-54 |
| | | | ILE-1 | ASN-124 |
| | | | THR-52 | ASP-26 |
| | | | LYS-99 | ASP-26 |
| | | | GLN-57 | ASP-38 |
| | | | ASP-25 | ARG-105 |
| | | | GLN-15 | ILE-55 |
| | | | SER-18 | CYS-99 |
| 6. | GaDi19 | Di19 In (*Arabidopsis thaliana*) | GLN-15 | CYS-69 |
| | | | GLN-15 | PRO-67 |
| | | | SER-18 | LYS-72 |
| | | | ASP-25 | LYS-62 |
| | | | LYS-2 | CYS-66 |
| | | | SER-18 | LYS-72 |
| 7. | GaDi19 | Di19 In (*Triticum avestum*) | SER-45 | LEU-104 |
| | | | ILE-51 | GLU-81 |
| | | | LYS-99 | ASP-82 |
| | | | LYS-99 | GLU-81 |
| | | | ARG-83 | GLU-68 |
| | | | ALA-55 | ASP-82 |
| | | | GLN-15 | CYS-66 |
| | | | TRP-48 | TYR-85 |
| 8. | GaDi19 | EREB1 In Cotton | PHE-44 | ARG-199 |
| | | | SER-45 | ARG-199 |
| | | | LEU-42 | ARG-187 |
| | | | SER-45 | ARG-187 |
| | | | HIS-41 | ARG-190 |
| | | | PHE-44 | ARG-190 |
| | | | GLN-57 | ARG-155 |
| | | | GLU-30 | ARG-143 |
| 9. | GaDi19 | EREB2 In Cotton | SER-45 | LYS-65 |
| | | | ILE-51 | THR-56 |
| | | | THR-52 | THR-56 |
| | | | SER-22 | ARG-51 |
| | | | LEU-23 | LYS-65 |
| | | | ASP-25 | LYS-65 |
| | | | SER-45 | LYS-65 |
| 10. | GaDi19 | DREB1a In *Arabidopsis thaliana* | SER-96 | TYR-51 |
| | | | THR52 | ARG-52 |
| | | | GLN57 | CYS-103 |
| | | | GLU30 | ARG-98 |
| | | | GLN15 | ASP-108 |
| 11. | GaDi19 | DREB1b In *Arabidopsis thaliana* | ARG50 | LEU-210 |
| | | | GLU97 | TYR-200 |
| | | | TRP-48 | HIS-198 |
| | | | TRP-8 | GLN-78 |
| | | | TRP-8 | PHE-77 |
| | | | SER-18 | ARG-54 |
| | | | GLN_15 | ARG-54 |
| | | | GLN-15 | ALA-128 |
| | | | LYS-2 | SER-56 |

**Table 2. Residue interaction of protein encoded by PR1 gene.**

| Ser No. | Receptor | Ligand | Residue interaction receptor ligand | |
|---|---|---|---|---|
| 1. | PR1 | AP2/EREBP | GLY-2 | LEU-192 |
| | | | SER-22 | GLU-153 |
| | | | TRP-20 | ARG-140 |
| | | | GLU-52 | ARG-190 |
| | | | GLU-52 | ARG-191 |
| | | | GLU-52 | ARG-198 |
| | | | TYR-41 | ASN-194 |
| | | | ASN-40 | LYS-137 |
| | | | ASN-75 | ASN-160 |
| | | | ASN-76 | GLY-161 |
| | | | ASN-76 | ARG-155 |
| | | | SER-22 | ARG-155 |
| 2. | PR1 | DREB1a In *Arabidopsis thaliana* | SER-22 | ARG-67 |
| | | | ASN-76 | ARG-56 |
| | | | ASN-76 | ARG-67 |
| | | | ASN-75 | LYS-72 |
| | | | CYS-5 | ASN-70 |
| | | | ASP-4 | ASN-70 |
| | | | SER-22 | GLU-65 |
| | | | SER-10 | ARG-52 |
| | | | HIS-9 | SER-101 |
| | | | GLU-52 | ALA-102 |
| | | | GLU-52 | SER-101 |
| | | | GLU-52 | ARG-100 |
| | | | ASN-50 | SER-109 |
| | | | GLU-52 | SER-109 |
| 3. | PR1 | DREB1b In *Arabidopsis thaliana* | ARG-62 | ASP-147 |
| | | | ARG-62 | HIS-144 |
| | | | ILE-67 | GLU-9 |
| | | | ARG-66 | ASP-138 |
| | | | ARG-66 | PHE-11 |
| | | | ARG-66 | PRO-17 |
| | | | ARG-66 | TYR-15 |
| | | | ARG-88 | GLY-19 |
| | | | ARG-86 | GLN-18 |
| | | | LYS-71 | HIS-198 |
| | | | LYS-71 | ARG-37 |
| | | | LYS-71 | SER-3 |
| | | | PHE-25 | TRP-196 |
| | | | ASP-24 | TRP-196 |
| | | | ARG-73 | TRP-73 |
| | | | ARG-73 | ILE-72 |
| | | | CYS-74 | THP-70 |
| 4. | PR1 | WRKY3 | ARG-62 | SER-388 |
| | | | ARG-62 | THR-386 |
| | | | ARG-62 | ARG-387 |
| | | | TYR-97 | GLN-385 |
| | | | THR-47 | THR-386 |
| | | | TYR-97 | THR-386 |
| | | | CYS-69 | LYS-447 |
| | | | ARG-66 | TRP-299 |
| | | | ARG-66 | HIS-448 |

**Table 3. Residue interactions of protein encoded by *PR2* gene.**

| Ser No. | Receptor | Ligand | Residue interaction receptor ligand | |
|---|---|---|---|---|
| 1. | PR2 | AP2/EREBP | ASN-201 | ARG-190 |
| | | | ASP-200 | ARG-187 |
| | | | GLU-492 | LYS-181 |
| | | | GLN-456 | GLU-170 |
| | | | ARG-238 | ARG-140 |
| | | | ARG-407 | ILE-154 |
| | | | ARG-407 | LEU-166 |
| | | | ARG-407 | TRP-165 |
| | | | ARG-407 | THR-168 |
| | | | ARG-343 | ARG-163` |
| | | | HIS-309 | GLU-153 |
| | | | ALA-341 | ARG-145 |
| | | | GLU-308 | ARG-143 |
| | | | VAL-307 | ARG-143 |
| | | | ASP-279 | ARG-143 |
| | | | ARG-272 | GLN-144 |
| 2. | PR2 | DREB1a In *Arabidopsis thaliana* | GLN-169 | ALA-102 |
| | | | GLN-169 | ARG-98 |
| | | | GLU-238 | GLY-53 |
| | | | GLU-238 | ARG-52 |
| | | | ARG-407 | LEU-77 |
| | | | ARG-343 | GLU-65 |
| | | | ARG-343 | ARG-67 |
| | | | ARG-344 | ARG-74 |
| | | | HIS-275 | ARG-55 |
| | | | VAL-307 | ARG-57 |
| | | | ASP-279 | ARG-57 |
| | | | TYR-136 | ASN-70 |
| 3. | PR2 | DREB1b In *Arabidopsis thaliana* | ASP-233 | ARG-44 |
| | | | ASP-233 | HIS-45 |
| | | | ASP-233 | ARG-41 |
| | | | ARG-407 | LEU-74 |
| | | | ARG-407 | TRP-196 |
| | | | GLU-424 | ASN-197 |
| | | | ARG-427 | TRP-196 |
| | | | GLU-491 | SER-193 |
| | | | GLU-492 | SER-193 |
| | | | GLU-492 | THR-116 |
| | | | ARG-496 | GLU-114 |
| | | | GLU-462 | GLU-203 |
| | | | GLU-462 | SER-209 |
| | | | ASN-435 | GLU-203 |
| | | | GLU-434 | SER-212 |
| | | | ARG-441 | GLU-150 |
| | | | ARG-441 | GLU-163 |
| | | | GLU-126 | TYR-15 |
| | | | ARG-343 | THR-70 |
| | | | ARG-343 | GLU-65 |
| | | | ARG-343 | ARG-71 |
| | | | ARG-342 | ARG-71 |
| | | | GLU-345 | ARG-71 |
| | | | GLU-308 | ARG-52 |
| | | | GLU-308 | ARG-64 |
| | | | VAL-307 | ARG-64 |
| | | | ASP-279 | ARG-64 |
| | | | GLU-238 | ARG-49 |
| | | | GLU-238 | CYS-100 |
| | | | GLN-169 | ARG-97 |
| | | | ASP-200 | LYS-39 |
| | | | ASP-200 | ARG-37 |
| | | | ILE-205 | ARG-97 |
| 4. | PR2 | WRKY3 | ARG-200 | LYS-405 |
| | | | GLU-492 | LYS-447 |
| | | | ARG-427 | THR-386 |
| | | | GLU-462 | GLN-385 |
| | | | GLU-434 | ARG-387 |
| | | | GLU-238 | GLY-403 |
| | | | ARG-343 | LEU-392 |
| | | | ARG-343 | LEU-393 |
| | | | ARG-343 | ASP-394 |
| | | | GLU-424 | ASP-394 |
| | | | LYS-378 | TYR-397 |
| | | | ARG-407 | ASP-395 |
| | | | GLU-308 | ARG-398 |
| | | | ASP-279 | ARG-400 |
| | | | GLU-308 | ARG-400 |

**Table 4. Residue interactions of protein encoded by *PR5* gene.**

| Ser no. | Receptor | Ligand | Residue interaction receptor ligand | |
|---|---|---|---|---|
| 1. | PR5 | AP2/EREBP | SER-163 | HIS-138 |
| | | | ASN-160 | HIS-138 |
| | | | SER-128 | LEU-93 |
| | | | LYS-26 | ARG-198 |
| | | | LYS-26 | VAL-199 |
| | | | ASP-126 | LEU-192 |
| | | | ASP-126 | LEU-193 |
| | | | ASP-153 | ARG-155 |
| | | | PRO-171 | ALA-191 |
| 2. | PR5 | DREB1a In *Arabidopsis thaliana* | GLN-110 | ARG-56 |
| | | | ASN-160 | ARG-56 |
| | | | ASP-153 | ARG-55 |
| | | | ASP-153 | ARG-74 |
| | | | TYR-113 | TYR-51 |
| | | | SER-128 | ASP-108 |
| | | | ASP-126 | LEU-104 |
| | | | ASP-126 | ASN-105 |
| | | | ASP-126 | ARG-52 |
| | | | PRO-143 | ARG-67 |
| | | | ASP-139 | ARG-67 |
| | | | ASP-139 | LYS-72 |
| | | | THR-131 | ARG-67 |
| | | | GLN-136 | ARG-67 |
| | | | LYS-26 | SER-109 |
| 3. | PR5 | DREB1b In *Arabidopsis thaliana* | ASP-126 | ARG-41 |
| | | | PRO-144 | TYR-48 |
| | | | SER-31 | GLU-62 |
| | | | ARG-138 | GLN-53 |
| 4. | PR5 | WRKY3 | ASP-126 | THR-386 |
| | | | SER-128 | ARG-387 |
| | | | SER-128 | GLN-385 |
| | | | TYR-113 | ARG-387 |
| | | | TYR-165 | ARG-387 |
| | | | ALA-158 | GLU-389 |
| | | | SER-31 | ASP-451 |
| | | | ILE-33 | ASP-451 |

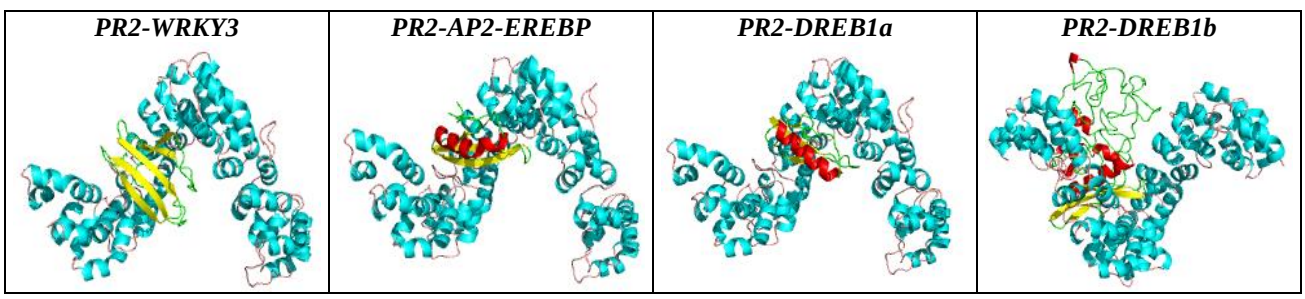



Fig. 7. Interaction models of *PR2* gene with *WRKY3, AP2/EREBP, DREB1a* and *DREB1b.*

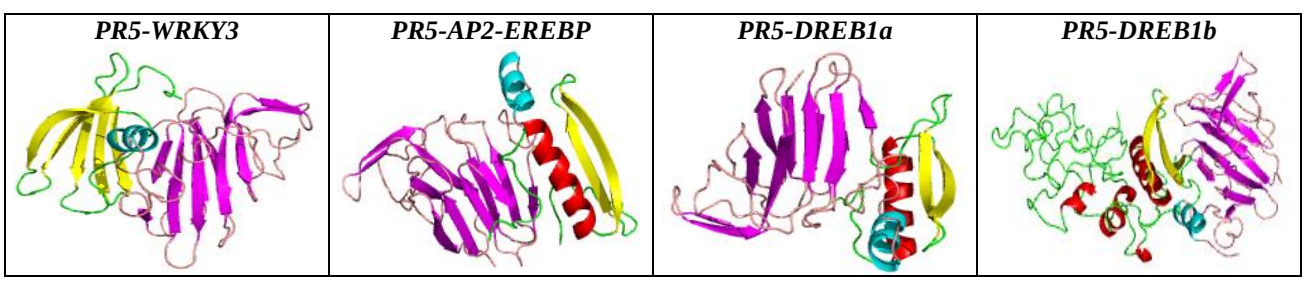



Fig. 8. Interaction models of *PR5* gene with *WRKY3, AP2/EREBP, DREB1a* and *DREB1b.*

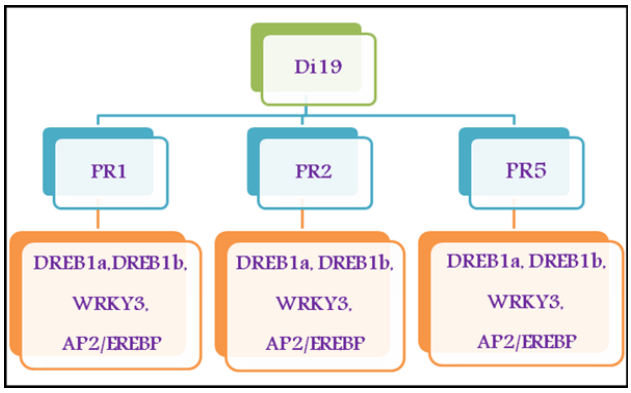



Fig. 9. A diagrammatic presentation of function cascade of *Di 19* under drought stress (*Di 19* induces *PR1, PR2* and *PR5* and then these *PR* genes further activates the *AP2/EREBP, DREB1a, DREB1b* and *WRKY3*).

## Discussion

Conventional breeding is a time-consuming and laborious process (Iqbal *et al*., 2013). Breeding approaches have been revolutionized with biotechnology and bioinformatics approaches. They are proven a handy tool for gene characterization in plants (Chavali and Rhee, 2017). The information of stress responsive transcription factors, their complex regulatory gene networks and functional characterization can be utilized for the development of crops with greater yield under stress conditions (Tran & Mochida. 2010).

In this study, we amplified *Di19* gene from *G. arboreum* (diploid ancestor of present day cotton). *Di19* is a transcription factor, which is induced in drought stress and further regulates the expression of three pathogen related genes *PR1, PR2, PR5* (Liu *et al*., 2013). Activation of *Di19* resulted in the drought tolerance enhancement (Liu *et al*., 2013). *Di19* gene family have zing finger proteins like *Cys2/His2*. Structure analysis showed that Di19 bounded to the element known as TACA (A/G)T (Li *et al*., 2010).

In this study, we have further used many bioinformatics tools to characterize *GaDi19*. Genetically modified crops having higher expression of *Di19* showed the greater tolerance to drought then other crops (Liu *et al*., 2013).

Qin *et al*., (2016) verified the nuclear localization of Di19-1 and Di19-2 (GhDi19-1/-2) proteins in cotton. Overexpression of GhDi19-1/-2 in seedlings of transgenic Arabidopsis exhibited hypersensitivity to elevated level of abscisic acid (ABA) and salinity. They demonstrated that for functional activation of GhDi19-1/-2 in response to ABA signaling and salt stress at early plant development stage phosphorylation of Ser (not Thr phosphorylation) was a key factor, and in ABA signal pathway GhDi19- 1/-2 proteins could be downstream targets of CDPKs (Calcium Dependent Protein Kinases) (*Qin et al*., 2016).

In the current study, 3D structure of Di19 protein was determined and Ramachandran plot showed 98% residues in favored regions (Fig. 2) which proved that model is reliable. Protein 3D structure analysis played a significant role in deciphering the function of protein, localizations and interactions (Parasuram *et al*., 2010). A simple view of the conformation of a protein can be obtained by Ramachandran plot (Ho & Brasseur, 2005).

In protein-protein interaction study of GaDi19 with other drought responsive genes, it was found interacting with other proteins like PR1, PR2, PR5 (Fig. 5), GhDi19-1, GhDi19-2, EREB1, EREB2, Di19 from *T. aestivum* and *A. thaliana* (Fig. 5). Residues of GaDi19 were also found interacting with the residues of PR1, PR2, and PR5 (Supplementary table 2) and GhDi19-1, GhDi19-2, EREB1, EREB2, Di19 from *T. aestivum* and *A. thaliana*, DREB1a and DREB1b. The results validate the interaction of Di19 with other drought responsive genes to confer drought resistance. To validate our results we amplified PR genes in *G. arboreum*. PR2 and PR5 were successfully amplified; however, amplification of PR1 gene was not possible even after repeating the reaction multiple times.

Conserved domain analysis shown two domains zf-Di19 and Di19_C domain in GaDi19. The Zinc binding domian describe it as drought induced 19 protein (Di19). Numerous drought induced 19 (Di19) like proteins are included in Zinc-binding family. A sturdy expression of Di19 was found in both the leaves and roots of *A. thaliana* during advanced drought (Liu *et al*., 2013).

In previous studies, Cys2/His2 was a domain also known as zinc finger domains and two of these unusual domains were present in the seven hydrophilic proteins and these proteins were encoded by the genes which were induced by the drought like Di19 gene (Li *et al*., 2010).

Presence of Di19_C domain describe it as stress-induced protein Di19. The sensitivity of plants to environmental stress, like drought, salinity, cold and osmotic stress is enhanced due to presence of the C-terminal domain of Di19. An increased supply of stress-related hormones such as ethylene and abscisic acid ABA also induce the protein (Liu *et al*., 2013).

In cotton during drought conditions, the *GhDi19-1* and *GhDi19-2* expression is induced to obtain the tolerance against drought (Li *et al*., 2010). Genes that are activated in the drought are responsible for tolerance against drought in plants.

Di19 up-regulates the expression of PR1, PR2 and PR5 (Liu *et al*., 2013), however next target of the PR1, PR2 and PR5 were unknown. Present study focused on the analyses of the next targets of PR1, PR2 and PR5. We found PR1, PR2 and PR5 interacting with Ap2/EREBP, DREB1a, DREB1b and WRKY3 (Figs. 6, 7, 8). This is a novel finding that PR genes further interact with AP2/EREBP, DREB1a, DREB1b, WRKY3 to enhance drought tolerance in plants (Fig. 9) which has not been reported earlier.

Analysis of drought responsive genes and pathways of these genes will be helpful for the scientists to get better information to develop drought resistant varieties and to know about protection method of plants.

## References

Ahmad, Z., E.A. Waraich, C. Barutçular, H. Alharby, A. Bamagoos, F. Kizilgeci, F. Ozturk, A. Hossain,Y. Bayoumi and A.E. Sabagh. 2020. Enhancing drought tolerance in *camelina sativa* L. and canola (*Brassica napus* L.) through application of selenium. *Pak. J. Bot*., 52(6): 1927-1939.

Altschul, S.F., T.L. Madden, A.A. Schaffer, J. Zhang, Z. Zhang, W. Miller and D.J. Lipman, 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucl. Acids Res*., 25: 3389-3402.

Chavali, A.K. and S.Y. Rhee. 2017. Bioinformatics tools for the identification of gene clusters that biosynthesize specialized metabolites. *Brief. Bioinform*. 7: 177.

Gosti, F.O., N. Bertauche, N. Vartanian, and J.R.M. Giraudat. 1995. Abscisic acid-dependent and-independent regulation of gene expression by progressive drought in *thaliana*. *Mol. Gen. Genet.*, 246(1): 10-18.

Hanson, A.D. and W.D. Hitz. 1982. Metabolic responses of mesophytes to plant water deficits. *Ann. Rev. Plant Boil*., 33(1): 163-203.

Ho, B.K. and R. Brasseur. 2005. The Ramachandran plots of glycine and pre-proline. *BMC Str. Biol*., 5: 14.

Iqbal, M., M.A. Khan, M. Naeem, U. Aziz, J. Afzal and M. Latif. 2013. Inducing Drought Tolerance in Upland Cotton (*Gossypium hirsutum* L.), Accomplishments and Future Prospects. *World Appl. Sci. J., 21*(7): 1062-1069.

Li, G., F.J. Tai, Y. Zheng, J. Luo, S.Y. Gong and Z.T. Zhang. 2010. Two cotton Cys2/His2-type zinc-finger proteins, GhDi19-1 and GhDi19-2, are involved in plant response to salt/drought stress and abscisic acid signaling. *Plant Mol. Biol.*, 74: 437-452.

Liu, W.X., F.C. Zhang, W.Z. Zhang, L.F. Song, W.H. Wu and Y.F. Chen. 2013. Di19 Functions as a Transcription Factor and Modulates PR1, PR2, and PR5 Expression in Response to Drought Stress. *Mol. Plant*, 6(5): 1487-1502.

Milla, M.A.R., J. Townsend, F. Chang and J.C. Cushman. 2006. The AtDi19 gene family encodes a novel type of Cys2/His2 zinc-finger protein implicated in ABA-independent dehydration, high-salinity stress and light signaling pathways. *Plant Mol. Biol.*, 61(1-2): 13-30.

Parasuram, R., J.S. Lee, P. Yin, S. Somarowthu and M.J. Ondrechen. 2010. Functional classification of protein 3D structures from predicted local interaction sites. *J. Bioinform. Comput. Biol*., 8: 1-15.

Qin, L.X., X.Y. Nie, R. Hu, G. Li, W.L. Xu and X.B. Li. 2016. Phosphorylation of serine residue modulates cotton Di19-1 and Di19-2 activities for responding to high salinity stress and abscisic acid signaling. *Sci. Rep*., 6: 20371.

Ritchie, M.D., J.R. Davis, H. Aschard, A. Battle, D. Conti, M. Du, H. Eskin, M.D. Fallin, L. Hsu, P. Kraft, J.H. Moore, B.L. Pierce, S.A. Bien, D.C. Thomas, P. Wei and S.B. Montgomery. 2017. Incorporation of biological knowledge into the study of gene-environment interactions. *Amer. J. Epidemiol*, 186(7): 771-777.

Saeed, U. 2009. Identification and characterization of drought tolerant genes in cotton by gene homology, Ph.D. dissertation, University of the Punjab, Lahore.

Shaheen, T., M. Rahman, M.F. Abbas, P. Zaib, M. Rahman, I. Ullah, M.T. Qamar and A. Arif. 2018. DREB gene from *Gossypium arboreum* L. cultivar FDH-786: identification, characterization, homology modeling and protein-protein interactions. *Int. J. Agri. Biol*., 20: 1055-1061.

Shaheen, T., Y. Zafar, J.M. Stewart and M. Rahman. 2013. Development of short gSSRs in *G. arboreum* and their utilization in phylogenetic studies. *Turk. J. Agri. For*., 37(3): 288-299.

Tran, L.S.P. and K. Mochida. 2010. Identification and prediction of abiotic stress responsive transcription factors involved in abiotic stress signaling in soybean. *Plant Signal Behav*., 5(3): 255-257.

Ullah, A., H. Sun, X. Yang and X. Zhang. 2017. Drought coping strategies in cotton: increased crop per drop. *Plant Biotechnol. J*. 15(3): 271-284.

Verwoerd, T.C., B.M. Dekker and A. Hoekema. 1989. A small-scale procedure for the rapid isolation of plant RNAs. *Nucl. Acids Res*., 17: 2362.

(Received for publication 28 February 2019)