GMR
The Original by FUNPEC-RP

Isolation and characterization of DREB gene in ancestral diploid wheat species from a wheat domestication center

S.M. Rustamova, G.R. Abdullayeva and I.M. Huseynova

Institute of Molecular Biology & Biotechnologies, Azerbaijan National Academy of Sciences, Baku, Azerbaijan

Corresponding author: I.M. Huseynova
E-mail: irada.huseynova@science.az

Genet. Mol. Res. 21 (3): gmr19035
Received January 31, 2022
Accepted March 18, 2022
Published July 31, 2022
DOI http://dx.doi.org/10.4238/gmr19035

ABSTRACT. The dehydration responsive element (DRE)-binding proteins (DREB) play a role in the signaling network that activates many abiotic stress-responsive genes. We isolated and molecularly characterized the DREB gene from ancestral diploid wheat species growing in Azerbaijan and native to this region. This territory is included in a region that is considered the center of origin of cultivated wheat. One-week-old seedlings of *Triticum urartu* (Au), *Aegilops speltoides* (B) and *Ae. tauschii* (D) were used for genomic DNA extraction. Gene-specific primer pairs were applied forisolation of the DREB gene. The amplification products were purified using a gel extraction kit and sequenced. Data analysis was performed using FGENESH, BLAST, INTERPROSCAN, SMART, MAFFT, ExPASy, ProteinPredict, and PSIPRED tools. Ensembl Plants and NCBI were used as integrative resources. Numerous SNPs and nine microindels were detected in the partial target sequence of the DREB gene in *Ae. speltoides*. Nonsynonymous SNPs were determined in *T. urartu* (1 transition and 5 transversions), and *Ae. tauschii* (2 transitions and 2 transversions). Analysis of amino acid sequences encoded by the putative DREB genes revealed a conserved AP2/ERF domain, with two conserved functional amino acids (14th valine and 19th glutamic acid) that play crucial roles in the recognition of the DNA-binding sequence and two tryptophan rings that determine the geometry of the GCC-box

binding domain. Nuclear localization signal and conserved Ser/Thr-rich region were observed in the corresponding amino acid sequences. One α-helix and three β-sheets were detected in the secondary structure of the AP2 domain. The isolated sequences of DREB gene from *T. urartu* and *Ae. tauschii* were confirmed and registered in NCBI with Accession Numbers MZ935739 and MZ935740. Identification of the DREB gene in wheat progenitors and its characterization is important for evaluating their genomic material for possible use to enhance the diversity of wheat cultivated under stress conditions.

Key words: DREB; AP2 domain; SNPs; *T. urartu; Ae. speltoides; Ae. tauschii*

INTRODUCTION

Drought and heat stresses appeared as the most devastating abiotic stress which negatively impact to plant growth and agricultural production. The recent climate changes exacerbate the adverse effects of these abiotic stresses and climatic studies based on long-term data have determined more frequent drought and heat spells in last two decades. Therefore, it is very important to determine new genetic resources to improve the stress tolerance of crop plants for sustainable agricultural production under a climate change scenario.

Plants respond to abiotic stress factors with a complex mechanism that includes the interaction of regulatory and functional genes, signaling, and metabolic pathways. The regulation of gene expression by transcription factors (TF) is the main pathway by which plants adapt to stress (Shahzad et al., 2021). To date, a wide range of TF families in plants, including (MYB) oncogene, APETALA2/ethylene response factor (AP2/ERF), basic leucine zipper (bZIP), Cys2(C2) His2(H2)-type zinc fingers (ZFs), and WRKY have been functionally identified. The ethylene response element-binding factor *(AP*2/*ERF*) family is a group of plant-specific transcriptional factors and characterized by the existence of an AP2/ERF DNA-binding domain of 60-70 amino acids, and it is consisted of the ERF, AP2, RAV related to ABI3 (abscisic acid incentive 3) and VP1 (viviparous 1) families.

Dehydration-responsive element-binding proteins (DREBs) belonging to the AP2/ERF family of TFs, play a significant role in the signaling network that modulates many plant processes (Agarwal et al., 2006; Lata and Prasad, 2011). DREB proteins activate many abiotic stress-responsive genes and maintain water balance in plant systems thus, imparting abiotic stress tolerance. DREB genes share a sequence similarity at the AP2 domain and bind to the 9 base pair sequence - C-repeat/DRE motif (TACCGACAT) in the promoter region of its downstream target genes. For the first time, a DRE element containing the core sequence (ACCGAC) was identified in the rd29A promoter (Yamaguchi-Shinozaki and Shinozaki, 1994; Sazegari and Niazi, 2012). There are three conservative regions found in all DREB genes, (1) the EREBP/AP2 DNA binding domain, (2) the N-terminal nuclear localization signal, and (3) the Ser/Thr-rich region. Valine at 14^{th} position and glutamic acid at 19^{th} position in the ERF/AP2 domain play an important role in the DNA binding specificity (Sakuma et al., 2002). It was an interesting point to find a DRE/CRT motif in a DREB promoter gene since it binds to the same motif in the promoter region of downstream stress-inducible genes. So, it can be predicted that the expression of the DREB gene is regulated by some other transcription factors such as DREB which

interacts with this unique motif. In contrast to the previous reports of limited tolerance with single genes, engineering crop plants with DREB transcription factors have resulted in improved stress tolerance. Due to the tolerance to one type of stress mostly confers tolerance to other stresses, transgenic plants with high expression of DREB genes have been found to be tolerant to multiple stress factors (Khan, 2011).

Advances in genomics have accelerated the improvement of several important crops, but similar efforts with wheat were more complex (Walkowiak et al., 2020). This is largely due to the size (17,000 Mbp) and complexity of the wheat genome. ~85% of the wheat genome was estimated to consist of repetitive elements. The current annotation of the wheat genome includes 124,201 protein genes originated from 3 related ancestral genomes (International Wheat Genome Sequencing Consortium, 2018). *Triticum aestivum* (AABBDD, 2n = 6x = 42), is a hexaploid comprising an "A" genome from the wild diploid *Triticum urartu* (2n=2x=14, A^uA^u), a "B" genome most likely from the species which is believed to be *Aegilops speltoides* (2n=2x=14, BB), and a "D" genome from *Aegilops tauschii* (2n=2x=14, DD). These progenitor species often provide valuable genetic resources by serving as diploid model organisms as well as donor of valuable diversity from undomesticated genome in wheat. Breeding strategies currently used have led to the loss of elite genes responsible for the resistance of crops to abiotic and biotic stressors of the environment, which has resulted in a significant decline in agricultural biodiversity. Exploration and use of genetic resources of wild relatives is a successful strategy to increase biodiversity in order to improve yields. Recently, for some crops, new introgression lines have been created between commercial varieties and wild relatives. For improving resistance to abiotic stress, genes from *Elytrigia elongata* and *Aegilops speltoides* were introgressed into wheat (Li et al., 2017).

The territory of Azerbaijan is included in the region considered to be the center of wheat origin. The evolution of hexaploid wheat took place by the hybridization of cultivated emmer (*Triticum dicoccum*) and *Ae. tauschii,* in the south and west of the Caspian Sea approximately 9000 years ago (Peng et al., 2011). Therefore, *Ae. tauschii* from this region is considered one of the possible sources of the D genome of bread wheat (Wang et al., 2013; Gaurav et al., 2022). A total of 16 species of wheat (*Triticum*) genera were found in Azerbaijan (Country report on the state of plant genetic resources for food and agriculture. Azerbaijan., 2006). Wild diploid wheat *T. boeoticum* (A^b genome), cultural *T. monococcum* (A^b), *T. urartu* (A^u), wild *T. montanium* (araraticum) (A^bG), tetraploid wheat *T. turgidum* (A^uB), cultivated Emmer (*T. dicoccum*), ancient and relict species of wheat, such as tetraploid *T. turanicum*, *T. percicum* (*T. carthlicum*), *T. polonicum* (A^uB), hexaploid *T. spelta*, *T. compactum* (A^uBD) were found in Azerbaijan. Nine species belonging to *Aegilops* genus are widespread in Azerbaijan. These genetic resources offered holistic opportunity of novel diversity of stress related genes due to their protracted evolution in an extreme climatic conditions. The main goal of the study was the analysis of the DREB gene diversity from genome donor species of wheat found on the territory of Azerbaijan and the implementation of their comparative molecular characterization.

MATERIAL AND METHODS

Plant materials

Seeds of *T. urartu* (A^u genome), *Ae. speltoides* (B genome), and *Ae. tauschii* (D genome) were obtained from Genbank of the Institute of Genetic Resources of Azerbaijan National Academy of Sciences (Baku). Germinated seeds were transferred in a growth chamber under normal conditions.

DNA extraction and quantification

The DNA was extracted as previously described (Huseynova et al., 2013). The concentration of genomic DNA was estimated spectrophotometrically (Nano-Drop Thermo Scientific-2000C, USA). The quality of the DNA was checked on the basis of the performance of the extracted DNA samples on a 0.8% agarose gel and was documented using "Gel Documentation System UVITEK" (UK).

DNA amplification with gene-specific primers

For the amplification of DREB gene were used gene-specific primer pairs (Table S1). Primers were designed based on the complete cDNA sequence of the *T. aestivum* AP2-containing protein (Ac.Nr. AF303376) using Primer3Plus (http://www.bioinformatics.nl/cgi-bin/primer3plus/primer3plus.cgi). PCR was performed as previously described in Huseynova et al. (2013).

Purification of PCR product and DNA sequencing

The preparative PCR amplification was carried out to amplify the DNA region of interest prior to Sanger sequencing. The PCR products were excised from the agarose gel and purified using ISOLATE II PCR & Gel Kit (BIOLINE) according to the manufacturer's instructions. The purified samples were then sequenced on an ABI 3130xl DNA analyzer (Applied Biosystems, USA).

In-silico analyses

Ensembl Plants (https://plants.ensembl.org/) and NCBI (https://www.ncbi.nlm.nih.gov/) were used as integrative resources for data analysis. The sequence was translated into possible reading frames using an online translation tool (https://web.expasy.org/translate/) (Duvaud et al., 2021). To explore the protein-coding capacity of the sequenced DNA fragments and compare them with known genes/proteins, the FGENESH (http://www.softberry.com/berry.phtml?topic=fgenesh/) (Solovyev et al., 2006) and BLAST (https://blast.ncbi.nlm.nih.gov/Blast.cgi) (Altschul et al., 1997) tools were used. To identify the conserved protein domains, the INTERPROSCAN (http://www.ebi.ac.uk/Tools/pfa/iprscan/) (Jones et al., 2014) and SMART (http://smart.embl-heidelberg.de/) (Letunic et al., 2015) programs were applied. A multiple amino acid sequence alignment was constructed using MAFFT v7.271 (https://mafft.cbrc.jp/alignment/server/) (Katoh and Standley, 2014). Secondary structure was predicted by PredictProtein analysis (https://predictprotein.org) (Bernhofer et al., 2021), information about amino acids composition was acquired from PSIPRED (http://bioinf.cs.ucl.ac.uk/psipred/) (Buchan and Jones, 2019).

RESULTS

Identification DREB allelic variation and its molecular characterization

A partial fragment of the DREB gene was amplified by gene-specific primers, designed from the conserved region. Figure 1 shows the electrophoretic profiles of PCR products of DREB DNA in *T. urartu*, *Ae. speltoides*, and *Ae. tauschii*. Approximately expected 500 bp fragment was successfully synthesized in the studied genotypes. Preparative PCR was conducted to isolate the partial DREB gene. Amplification products were separated in 1.5% (w/v) agarose gel, purified using the special purification kit, and sequenced.

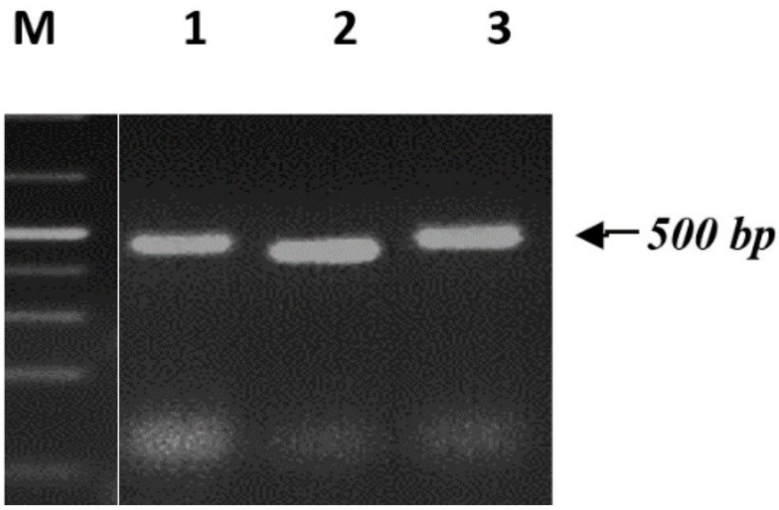

Figure 1. Gel electrophoresis of PCR products of DREB DNA in *Triticum urartu* (1), *Aegilops speltoides* (2), and *Aegilops tauschii* (3). The arrow indicates 500 bp fragments. M-100 bp DNA ladder.

The nucleotide sequences of the DREB alleles from all three species were multi-aligned using a Fast fourier transform server MAFFT v7.271 (Figure 2). Calculation of pairwise alignment between *Ae. tauschii* and *T. urartu* showed 97.52% identity with a score of 46240.0. In the other pair variants, the identity was low, 54.73% with a score of 14660.0 between *Ae. speltoides* and *T. urartu*, and 55.55% with a score of 14980.0 between *Ae. speltoides* and *Ae. tauschii*. As shown in Figure 2, numerous SNPs, insertions, and deletions are observed in *Ae. speltoides*. The insertion consisting of 8 nucleotides at 43-50th positions was replaced by the deletion of the same size at 101-108th positions. After that, insertions and deletions of different sizes replace each other. 45% of the SNPs observed in this plant are transitions. Twelve SNPs were also found between *T. urartu* and *Ae. tauschii*. Four of them (10, 51, 59, and 506th positions) are transitions, and the rest (8, 195, 196, 205, 501, 503, 504, and 505th positions) are transversions. Nucleotide sequences were translated into appropriate amino acid sequences to verify if the mutations occur were synonymous or nonsynonymous substitution.

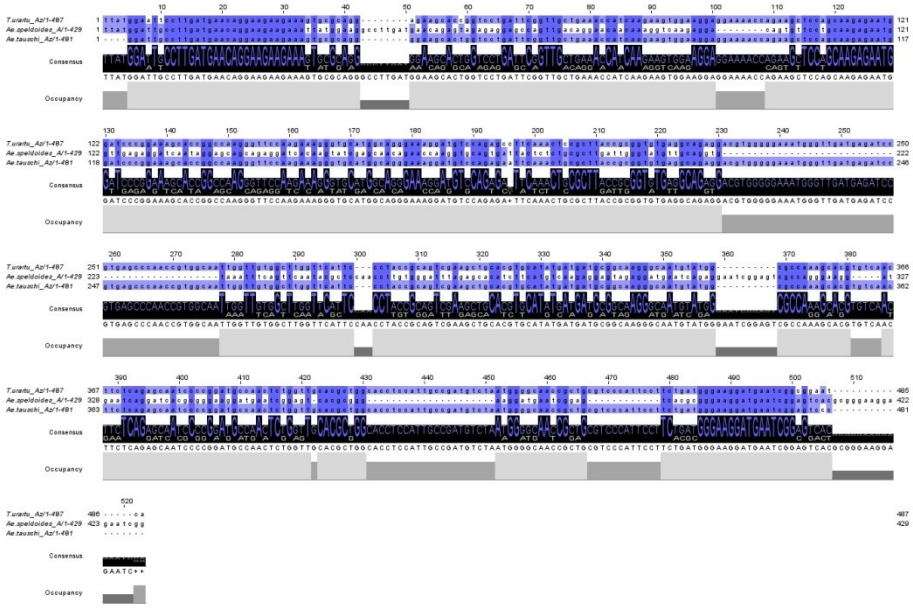

Figure 2. Sequence alignment of the partial DREB gene isolated from *Triticum urartu, Aegilops speldoides* and *Aegilops tauschii* using MAFFT analysis. Conservations of nucleotides are distinguished by various shades of blue color.

In the 487 bp DNA sequence of *T. urartu*, the FGENESH program predicted a gene (or gene fragment) with 1 exon that might encode a polypeptide of 148 aa in length. This exon is within 16-462th positions, 447 bp. In the 481 bp sequence of *Ae. tauschii*, the FGENESH predicted a gene with 1 exon that might encode a polypeptide of 148 aa in length. The predicted exon is within 12-458th positions, 447 bp (Figure 3).

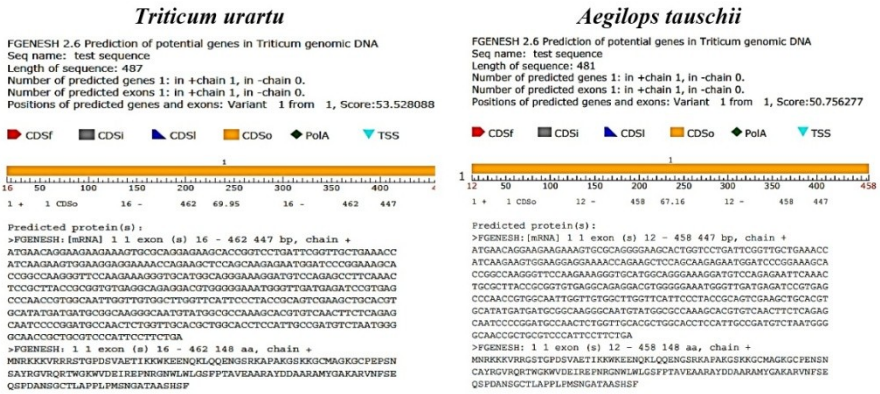

Figure 3. *In silico* analysis and position of exons in the sequenced region of the DREB gene deduced by the FGENESH program. CDSf - the first exon, CDSi –internal exon, CDSl - last coding segment, CDSo – single exon, PolA - polyA segment, TSS - transcription start.

In parallel, the sequences were translated into six possible reading frames using an online translation tool (web.expasy.org/translate/). In the top 10 aa sequences selected as a result of the

BLASTp search in the NCBI, for *T. urartu*, the identity was 86-97% with query coverage of 94-98%, and the identity was 96-97% with query coverage of 94-97% for *Ae. tauschii*. No significant similarity was found in NCBI for predicted aa sequence of *Ae. speltoides*.

SNPs leading to amino acid substitutions

As mentioned above, to determine whether nucleotide mutations caused amino acid substitution and at the same time to identify conserved regions, the deduced amino-acid sequences and AP2-containing or DRE-binding proteins from 21 species of the Poaceae family available in the NCBI GenBank (Table S2), were multi-aligned using MAFFT. As seen in the Figure 4, in the DREB gene of *T. urartu*, the transversions caused nonsynonymous substitution in two positions: the mutation of AA to CC (positions 195-196), caused asparagine (N) to be replaced by proline (P) at position 76, and the mutation of G to C (at position 205) caused cysteine (C) to be replaced by serine (S) at position 80. The other SNPs observed in *T. urartu* are silent substitutions. The mutations identified in the DREB gene of *Ae. tauschii* consisted in a transition of A to G (at position 51) with consequently changing of amino acid from arginine (R) to glycine (G) at position 29. The amino acid substitutions common for both of these plants and not common for anothers are noteworthy: glycine (G) was replaced by cysteine (C) at position 74, alanine (A) by aspartic acid (D) at position 95, and arginine (R) by tryptophan (W) at position 105.

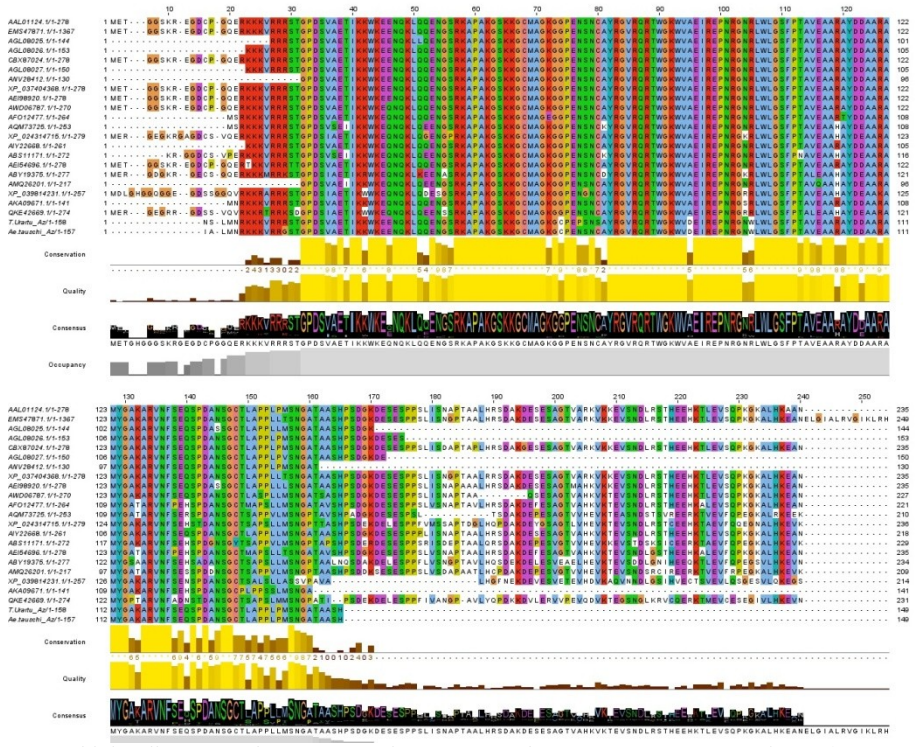

Figure 4. Multiple alignment of DREB protein sequences from *Triticum urartu* and *Aegilops tauschii* of Azerbaijani origin and AP2-containing or DRE-binding proteins from 21 species of the *Poaceae* family available in NCBI GenBank.

To determine the localization of SNPs that cause these substitutions in the *T. aestivum* (Assembly IWGSC) genome, BLASTp was performed on Ensembl Plants and pairwise alignment was performed based on the results obtained. Localization of SNPs occurring in *T. urartu* were found on the TraesCS3A02G099200 gene belonging to chromosome 3 of the A genome and localization of SNPs occurring in *Ae. tauschii* were observed on the TraesCS3D02G099500 gene belonging to chromosome 3 of the D genome (Table 1). Besides, a phylogenetic tree was constructed based on the alignment results (Figure 5).

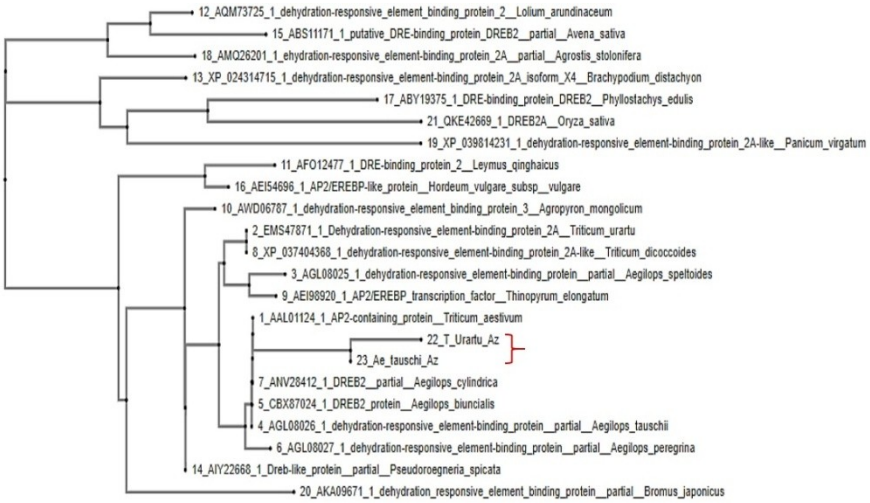

Figure 5. Phylogenetic analysis of DREBs from *Triticum urartu* and *Aegilops tauschii* of Azerbaijani origin (marked with red braces) and 21 species of the *Poaceae* family available in NCBI GenBank.

In addition, nucleotide sequences were aligned with randomly selected DREB genes of different origins from NCBI GenBank to identify SNPs that cause specific nonsynonymous substitutions observed in both studied species. The SNPs that are specific to our species are identical to some accessions (Ac.Nr.KR106189.1 and Ac.Nr. AY781361.1). However, five SNPs were only specific to *T. urartu* and *Ae. tauschii* of Azerbaijani origin (Figure S1). Three of these mutations caused the nonsynonymous substitutions mentioned above, two were silent substitutions (Table 1).

Table 1. Nonsynonymous substitutions detected in *Triticum urartu* and *Aegilops tauschii* and their localization in the *Triticum aestivum* (IWGSC) genome.

	Base type substitution in SNPs	Genomic location	Gene	Amino acid change
T. aestivum /T. urartu	G/T	3A: 64028923	TraesCS3A02G099200	Glycine/Cysteine
	A/C	3A: 64028914		Asparagine/Proline
	A/C	3A: 64028913		
	G/C	3A: 64028904		Cysteine/Serine
	C/A	3A: 64028859		Alanine/Aspartic acid
	C (A)/T	3A: 64028830		Arginine/Tryptophan
T. aestivum /Ae. tauschi	A/G	3D:51102507	TraesCS3D02G099500	Arginine/Glycine
	G/T	3D:51102372		Glycine/Cysteine
	C/A	3D:51102308		Alanine/Aspartic acid
	C (A)/T	3D:51102270		Arginine/Tryptophan

AP2 domain structure and protein motif features

Analysis of amino acid sequences encoded by the putative DREB genes from *T. urartu* and *Ae. tauschii* by INTERPROSCAN and SMART programs revealed the AP2 domain. At the same time, the aa composition of the AP2 domain was analyzed, and the second structure for this domain was predicted (Figure 6). As shown in the figure, alanine and arginine dominate in the AP2 domain. As for the types of amino acids, 36% are nonpolar, 34% polar, 4.16% aromatic, and 14% hydrophobic amino acids within the domain. Using the PredictProtein online tool, one α-helix and three β-sheets were predicted for the AP2 domain.

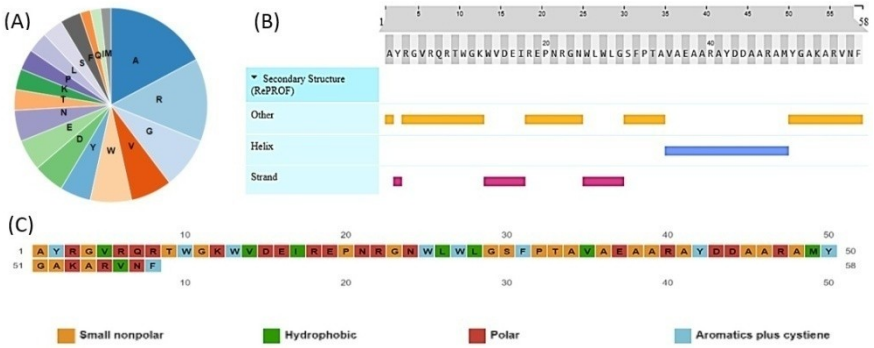

Figure 6. Structural annotation of AP2 domain predicted by PredictProtein and PSIPRED. (A) Amino acid composition; (B) Graphical representation of the secondary structure; (C) Amino acid types.

The results of bioinformatics analyses were summarized, domain and conservative areas were identified in the sequenced partial DREB gene (Figure 7). The specific signal peptide sequences KKK and KKWK enabling the transfer of DREB transcription factors to the nucleus are located ahead of the AP2 domain. In the area belonging to the AP2 domain consisting of 58 amino acid sequences, valine 14th, and glutamic acid 19th amino acids specific for this domain were found. Tryptophan residues and the Ser/Thr-rich region after the domain, which play a special role in the formation of the spatial structure of the AP2 domain, were determined.

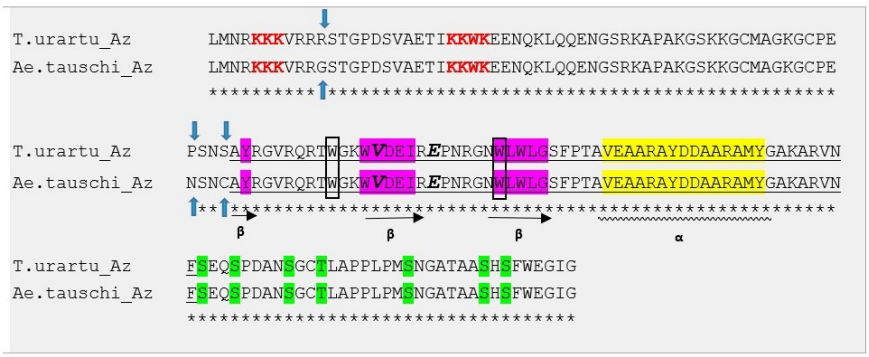

Figure 7. Deduced amino-acid sequence of AP2 domain of DREB genes in A and D subgenome donor species of Azerbaijani origin. The area underlined with a solid line shows AP2/ERF DNA-binding domain. The specific signal peptide sequence area for the DREB genes is highlighted in red. The valine 14th and glutamic acid 19th amino acids of the AP2 domain are shown in bold and italics. Black boxes are the aromatic rings of Trp. The specific sequence for α-helix is yellow, β-sheets are purple. Green colors are the T and S residues that distinguished the Ser/Thr-rich region. The blue arrows indicate amino acid substitutions.

Cross-species analyses of DREBs in Ensembl Plants

Further, the nucleotide sequences of the DREB gene from the genome donor species of wheat were used as queries in BLAST for homology search against 10 phylogenetically close plant species genomes. The total number of hits obtained using BLASTN and BLASTP in Ensembl Plants are presented in Table S3. Statistically significant results are shown in the Table 2. The sequence of 488 nucleotides isolated from *T. urartu* was analyzed comparatively with *T. urartu* (Assembly ASM34745v1) genome in the database, and a hit corresponding to the TRIUR3_17973 gene with 97% similarity was chosen. It has the highest bit score and the lowest e value in the list of hits. This gene consists of 1367 aa and is assumed to have one splicing variant, 214 orthologues, and two paralogues. A protein corresponding to this gene - Dehydration-Responsive Element-Binding protein 2A (M7YLU1) exists in UniProtKB. As a result of a homologous search against *T. aestivum* (Assembly IWGSC) genome, only four of the 27 hits showed significant compliance. Three of them belong to the TraesCS3A02G099200, TraesCS3B02G115400, and TraesCS3D02G099500 genes located in the A, B, and D genomes, respectively. TraesCS3A02G099200 gene has one transcript and encodes a protein of 278 aa. 212 orthologues and four paralogues are assumed. Dehydration responsive element binding protein W73 (UniProtKB/TrEMBL; Q4U0C8) is a product of this gene. TraesCS3B02G115400 gene belonging to the B genome has two splice variants and encodes proteins of 1008 aa and 1401 aa. It is assumed that there are 212 orthologues and six paralogues of this gene. The UniProtKB contains one protein corresponding to this gene (Q3LR66) and three protein sequences are available in UniParc (A0A341SES0, A0A341S9F4, A0A1D6AYT1). TraesCS3D02G099500 gene belonging to the chromosome 3 of the D genome has two splice variants, 212 orthologues, and four paralogues. This gene encodes proteins of 1311 aa and 1225 aa and has four proteins in UniProtKB (G0YWB9; G0YWC2) and UniParc (A0A341T1H8; A0A341T611). It is noteworthy that the target sequence corresponds also to the chromosome 5 of the B genome. One hit in the search result matched the TraesCS5B02G344200 gene with 88% identity. This gene is shown to encode one transcript, a protein of 1409 aa in length, and is assumed to have 18 orthologues and 822 paralogues. In UniProtKB, this gene corresponds to the AP2 / ERF domain-containing protein (A0A3B6LRX8) and two proteins are available in UniParc (A0A1D5Z4A2; A0A1D5Z4A3). Similar results have been observed with *T. spelta* (Assembly PGSBv2.0) but no relevant information for these genes has been found in protein databases. One hit shows a high similarity between the DNA sequence and the *Ae. tauschii* (Assembly Aet_v4.0) genome. Thus, the identity with significant indicators of score and e-value was 98%. This is a protein-coding gene (AET3Gv20214400), which has 30 splice variants, 214 orthologues, and five paralogues. Four dehydration element-binding proteins (A0A2Z4JFC0; A0A2Z4JGM7; V5JE74 and M8BG30) corresponding to this gene were detected in UniProtKB. According to the BLASTN analysis, two of the 15 hits found in the *T. dicoccoides* (Assembly WEWSeq_v.1.0) genome corresponding to the DREB gene sequence from *T. urartu* of the Azerbaijani origin were significantly homologous. They demonstrate compliance to TRIDC3AG011980 and TRIDC3BG016050 genes located on 3A and 3B chromosomes. The first of them has five splice variants, 214 orthologues, and four paralogues. There is information about one protein (V5JDZ4) for this gene. For the second gene in the B genome, 28 splice variants, 88 orthologues, and 42 paralogues are

predicted. There is no information about the relevant protein in UniProtKB. Three hits in the *T. turgidum* (Assembly Svevo.v1) genome show significant homology. Among them, the similarity with the corresponding gene located on the 3A chromosome is stronger (97%), followed by the 3B and 5B chromosomes (96% and 88%, respectively). There is no information about these gene products in UniProtKB, but Pfam, Prosite profiles, SMART, Prints, CDD, Gene 3D programs predict the AP2 / ERF domain in all transcripts. One of the total 20 hits found in the *Brachypodium distachyon* (Assembly Brachypodium_distachyon_v3.0) genome is statistically significantly homologous to the BRADI_2g04000v3 gene located on the 2nd chromosome. This gene has one splice variant, encodes a protein of 215 aa in length, and two corresponding proteins are detected in the UniProtKB (A0A0Q3IR73) and UniParc (I1HCA7). Homology search against barley and rice genomes has gradually weakened.

Table 2. BLASTN search results for DREB gene sequences of *Triticum urartu* in Ensembl Plants. Only statistically significant results are shown in the table.

	Genomic location	Overlapping gene (s)	Query start	Query end	Length	Score	E-val	%ID
Triticum urartu (ASM34745v1)	scaffold99160:39973-40420	TRIUR3_17973	2	449	448	400	0.0	97.3
Triticum aestivum (IWGSC)	3D:51102101-51102548	TraesCS3D02G099500	2	449	448	404	0.0	97.5
	3A:64028655-64029099	TraesCS3A02G099200	2	446	445	393	0.0	97.1
	3B:82053470-82053908	TraesCS3B02G115400	2	440	439	367	0.0	95.9
	5B:527335993-527336431	TraesCS5B02G344200	11	449	439	235	2.8e-128	88.4
Triticum spelta (PGSBv2.0)	3D:59436475-59436922	TraesTSP3D01G114700	2	449	448	404	0.0	97.5
	3A:57140854-57141298	TraesTSP3A01G106700	2	446	445	393	0.0	97.1
	3B:90070184-90070622	TraesTSP3B01G136300, TraesTSP3B01G136200	2	440	439	371	0.0	96.1
	5B:527280296-527280734	TraesTSP5B01G373500	11	449	439	235	2.7e-128	88.4
Aegilops tauschii (Aet_v4.0)	3D:54263028-54263475	AET3Gv20214400	2	449	448	404	0.0	97.5
Triticum dicoccoides (WEWSeq_v.1.0)	3A:59811509-59811953	TRIDC3AG011980	2	446	445	393	0.0	97.1
	3B:90628890-90629328	TRIDC3BG016050	2	440	439	371	0.0	96.1
Triticum turgidum (Svevo.v1)	3A:59512445-59512889	TRITD3Av1G029860	2	446	445	393	0.0	97.1
	3B:89457074-89457512	TRITD3Bv1G034470	2	440	439	371	0.0	96.1
	5B:524418009-524418447	TraesTSP5B01G373500, TRITD5Bv1G180570	11	449	439	235	2e-128	88.4
*Brachypodium_ distachyon_*(v3.0)	2:2826140-2826523	BRADI_2g04000v3	21	404	384	232	3.2e-128	90.1
Hordeum vulgare (IBSC_v2)	chr3H:46483657-46484091	dehydration responsive element binding 1	11	445	435	307	1e-171	92.6
Oryza sativa Japonica Group (IRGSP-1.0)	1:3357488-3357603	DREB2A	67	182	116	64	8e-28	88.8
Oryza sativa Indica Group (ASM465v1)	1:3782453-3782568	BGIOSGA002846	67	182	116	64	9.1e-28	88.8

Although the percent identity of hits obtained in the BLASTN analysis in Ensembl Plants by using a sequence of 429 nucleotides of a DREB gene isolated from *Ae. speltoides* as a query was high, the bit-score was low (Table S4). These areas distributed across different regions of the genomes also corresponded to the genes encoding dehydration responsive element binding proteins in each species. The e-value for hits found in the more phylogenetically distant Brachypodium, barley, and rice species were statistically unacceptable. Figure 8 describes the high-scoring segment pair (HSP) distribution of a partial DREB gene sequence from *Ae. speltoides* on *T. aestivum* (WGSC) genome. As seen, identical regions across the B genome are more distributed. Thus, homologous regions are identified on all chromosomes of the B genome except for the 2nd chromosome.

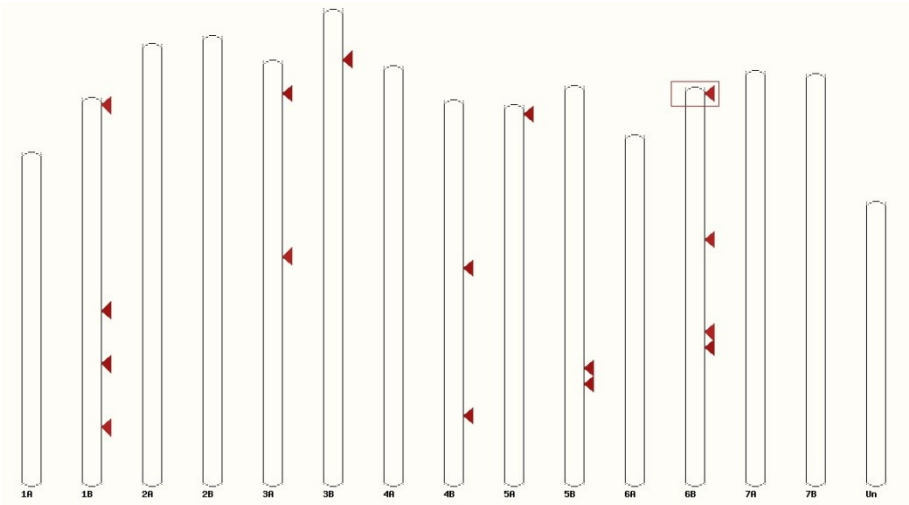

Figure 8. HSP distribution of a partial DREB gene sequence from *Aegilops speltoides* across the *Triticum aestivum* (WGSC) genome.

Statistically significant results of BLASTN search using the DREB gene fragment of 481 nucleotides in length isolated from *Ae. tauschii* grown in Azerbaijan as a query against phylogenetically close 10 plants in Ensembl Plants database are presented in Table S5. Concerning the localization in the genomes, these results are almost identical to those obtained for *T. urartu*.

The isolated sequences from *T. urartu* and *Ae. tauschii* was confirmed and registered in NCBI as partial CDS of DREB gene with Accession Nr.MZ935739 and MZ935740.

DISCUSSION

According to large-scale transcriptome analyses, protective proteins and regulatory proteins are involved in the molecular stress responses (Shahzad et al., 2021). Among them, transcription factors were shown to play crucial roles in regulating plant growth and responses to abiotic and biotic stresses. DREB proteins belonging to the superfamily of AP2/ERF plant transcription factors play an important role in the signaling network that modulates many processes, such as stress responses and plant development. In our previous studies, using the genome-specific functional markers in Azerbaijan wheat genotypes, a gene encoding DREB1 transcription factor was detected in the A, B, and D genomes (Huseynova et al., 2013). Wheat genotypes of Azerbaijani and German origin with contrasting drought tolerance were used in order to determine the expression level of the DREB1 transcription factor. In general, the transcript levels of all genotypes exposed to drought stress were detected to increase significantly. Further, under drought stress, the expression level of DREB1 in the tolerant genotypes increased more than in drought-sensitive ones (Rustamova et al., 2020). Now, we performed evaluated the variation of major DREB gene between the ancestral diploid species and hexaploid wheat as well as to deciphere its molecular characterisation among different species.

Comparative analysis of SNP and amino acid sequences revealed synonymous and non-synonymous substitutions. Xu et al. (2019) also observed some nucleotide substitutions in ancestral diploid species of wheat. In addition, it was found that SNPs causing nonsynonymous substitutions in DREB genes, observed in both studied species, are also rare in other species. Thus, among randomly chosen accessions in NCBI Genbank, the same SNPs were detected in two bread wheat cultivars, *T. aestivum* cultivar Kalak Afghani dehydration-responsive element-binding 2 (DREB2) mRNA (Ac.Nr.KR106189.1) and *T. aestivum* DREB transcription factor 6 (DREB6) mRNA (Ac.Nr.AY781361.1). But five SNPs were only specific to *T. urartu* and *Ae. tauschii* of Azerbaijani origin. Concerning the physicochemical properties of substituted amino acids, glycine and cysteine are both nonpolar neutral amino acids but differ in being aliphatic and sulfuric. Both asparagine and proline are neutral but asparagine contains an amide functional group and is polar, while proline is cyclic and nonpolar. The next nonsynonymous substitution observed by us was the replacement of sulfuric nonpolar cysteine with hydroxylic polar serine. Nevertheless, they are neutral amino acids. Replacement of neutral amino acid with negative amino acid is the replacement of aliphatic nonpolar alanine with acidic polar aspartic acid. In two cases, the positive amino acid was replaced by a neutral amino acid: basic polar arginine with aromatic nonpolar tryptophan and aliphatic nonpolar glycine. The properties of amino acids influence the tertiary structure or overall shape of the proteins. Therefore, the identified SNPs can induce an important variation in the EREBP domain regarding structural alterations in the protein geometry, altering the affinity of the protein for DNA. Mondini et al. (2015) detected specific SNPs in a variety highly tolerant to salt stress. They consider these mutations to be beneficial to the plant. These mutations can be useful due to the increased affinity of the protein for the substrate. The substrate can be obtained by changing the residues of amino acids, which are directly involved in the identification and binding to the attachment region on DNA.

For the first time, Shen et al. (2003) isolated the TaDREB1 gene from wheat and proved that this gene includes one conserved EREBP/AP2 domain and similar to Arabidopsis DREB family members. Sequence analysis which detected a 1.3 kb cDNA that included a complete ORF of 837 bp, a 5'-UTR of 251 bp, and 3´-UTR of 204 bp. It was found that this gene encodes a polypeptide with a molecular weight of 30.3 kDa and consists of 278 amino acid residues. We revealed that the AP2 domain and its surroundings maintained a high degree of conservatism. Our findings are in line with previous work where using genomic-wide analysis, AP2/ERF transcription factors were found to be highly conserved during plant evolution (Feng et al., 2020).

In our analysis, one α-helix and three β-sheets were predicted for the AP2 domain. For the first time, the 3D structure of the complex of the Arabidopsis AtERF1-DNA-binding domain and its target DNA was determined by NMR (Allen et al., 1998) and used for molecular modeling of ParCBF1. To understand the structure-function relationships, for the first time, Pandey et al. (2014) built the tertiary structure of the DREB2 protein from wheat by homology modeling based on the crystal structure of GCC-box binding domain of *Arabidopsis thaliana*. Similarly, this model comprises of three-stranded antiparallel β-sheet followed by α-helix and relatively unstructured C'-terminal. By changing various amino acids within the AP2 domain, Hassan et al. (2021) determined their role in the formation of the secondary structure. They also showed that some of the model structures had two beta strands and one alpha helix.

Based on the results of bioinformatics analyses, domains and conservative areas we identified in the sequenced partial DREB gene. KKK and KKWK in the N-terminal region function as a nuclear localization signal (NLS). An entry of the nucleus-targeted transcription factors into the nucleus is regulated by the NLS (Pandey et al., 2014). The region of 58 amino acid residues identified in the study is strictly conserved among the DREB proteins. This region is referred to as the AP2/ERF DNA-binding domain. The DREB transcription factor, belonging to AP2/EREBP family, contains one AP2/EREBP domain, composed of about 60 amino acid residues with the conserved elements (Wang et., 2021). Through the AP2/EREBP domain, the DREB transcription factor could specifically bind to the dehydration responsive element/C-repeat (DRE/CRT) (CCGAC). This binding is involved in the regulation of the gene expression related to response to high salt, low temperature, or drought, which would enhance the plant tolerance to stress. Two highly conserved functional amino acids at the 14^{th} and 19^{th} positions were also observed in the AP2 domain. These amino acids distinguish the DREB (valine and glutamic acid) from the ERF (alanine and aspartic acid) (Agarwal et al., 2006). Besides, tryptophan was detected in the AP2 domain, followed by serine and threonine amino acids in polypeptides corresponding to the sequenced partial DREB gene. The results of our study are in line with data obtained by previous studies. Liu et al. (1998) demonstrated the presence of a conserved Ser/Thr-rich region adjacent to the EREBP/AP2 binding domain containing the phosphorylation site for the regulation of gene activity. According to Chen et al. (2007), seven key amino acids in the AP2 domain play an important role in CRT/DRE element binding, namely, four R residues, two W residues, and one V residue. The role of tryptophan rings in recognition of GCC-box and determination of the geometry of the GCC-box binding domain was reported by Mondini et al. (2015).

Genes and genome regions that were statistically significant homologues were found in Ensembl Plants. Search against *T. aestivum* (IWGSC) genome found high homology with TraesCS3A02G099200, TraesCS3B02G115400, and TraesCS3D02G099500 genes localized in 3A, 3B, and 3D, respectively. These results are in line with Wei et al. (2009) who observed Dreb1 genes located on chromosomes 3A, 3B, and 3D using these genome-specific primers and the Chinese Spring nulli-tetrasomic lines. Besides, mapping of the TaDREB1 gene showed that it is located between Xmwg818 and Xfbb117 on the 3BL chromosome. Total 13 new allele variations of TaDREB1 were identified in wheat germplasms, including 5 TaDREB1-A on chromosome 3AL, 4 TaDREB1-B on chromosome 3BL, and 4 TaDREB1-D on chromosome 3DL (Liu et al., 2018). It is noteworthy that the DREB sequence studied by us corresponds to the 5^{th} chromosome of the B genome with an identity of 88% - 89% (*T. urartu* and *Ae. tauschi*, respectively). In *T. aestivum*, this gene is shown to encode a protein of 1409 aa in length, and in UniProtKB, this gene corresponds to the AP2 / ERF domain-containing protein (A0A3B6LRX8).

Multi-alignments of nucleotide sequence from all three progenitors revealed multiple insertions and deletions in *Ae. speldoides*. Our findings are in line with previous work where the highest nucleotide diversity value of the DREB gene was detected in *Ae. speltoides* (Xu et al., 2019). Xu et al. also found out a large-size (70 bp) insertion/deletion in a DREB gene of *Ae. speltoides*, which was different from the copy of sequences from other accessions of *Ae. speltoides*, suggesting a likely existence of two different ancestral *Ae. speltoides* forms. Based on the obtained results, it is necessary to use a wide range of

accessions of *Ae. speltoides* in phylogenetic analysis to reveal the origination of B subgenome in the modern bread wheat. *Ae. speltoides* is intended not to be the exclusive donor of this genome. The wheat B genome might rather have a polyphyletic origin with multiple ancestors involved, including *Ae. speltoides* (Concia et al., 2020).

To date, many studies for understanding mechanisms and functions of DREB transcription factors were undertaken. The first DREB transcription factor, CBF1 was isolated from Arabidopsis (Liu et al., 1998). Since then, several homologs of DREB1 and DREB2 have been identified in different plants. DREB1 gene isolated from *T. aestivum* (Shen et al., 2003) was significantly induced by abiotic stress factors as drought, salinity, and low temperature. Expression of DREB2 isolated from wheat seedlings was activated under the influence of the same stress factors and ABA. 500 bp TaDREB DNA sequences were detected in the Iranian wheat genotypes (Andeani et al., 2009). Morran et al. (2011) identified a new member of the DREB family called TaDREB3 and developed both TaDREB2 and TaDREB3 over-expressing transgenic populations of barley and wheat. Two isoforms of WDREB2 were isolated and molecularly characterized in wheat and WDREB2 was shown to have 3 different alternative splice forms or isoforms. β isoform that lacks transcription activation domain is inactive while α is an active isoform (Sazegari and Niazi, 2012). SNPs in EREBP/AP2 domain of DREB1, DREB2, DREB3, DREB4, and DREB5 genes were identified and characterized in some durum wheat cultivars with contrasting salt and drought tolerance (Mondini et al., 2015). Identification of wheat DREB genes was performed by Niu et al. (2020) at the genome level. Functions of TaDREB genes were characterized and in total, 210 TaDREB genes, which can be divided into 6 subgroups were detected. They found that TaDREB21-B/D showed alternative exon patterns, TaDREB20-D, TaDREB26-A, TaDREB52-B/D, and TaDREB58-A showed an intron-retention pattern, while only TaDREB3-A/B/D showed an exon-skipping pattern which was found on all three chromosomes. Using sequence-based phylogenetic analyses, Hassan et al. (2021) identified 32 new DREB subfamily members, not belonging to any known sub-group.

CONCLUSIONS

We have isolated and sequenced the partial DREB gene from genome donor species of wheat widespread in Azerbaijan. The highest nucleotide diversity value of the DREB gene was detected in *Ae. speltoides*. Six and four nonsynonymous SNPs were found in *T. urartu* and *Ae. tauschii*, respectively. This gene was shown to contain a highly conserved AP2 domain, a nuclear localization signal, and a conserved Ser/Thr-rich region. One α-helix and three β-sheets were detected in the secondary structure of the AP2 domain. Future analyses of the performance of key regulatory genes in wheat progenitors, across environments and locally and generally adapted wheat, will give an opportunity to assess the value of selecting the wild relatives for the development of drought tolerant varieties.

ACKNOWLEDGMENTS

This work was supported by Azerbaijan National Academy of Sciences (Order No.342, 20 August 2020) and Science Development Foundation under the President of the Republic of Azerbaijan - Grant № EIF-ETL-2020-2(36)-16/15/3-M-15). The authors greatly

acknowledge prof. Ali Ahmad Naz (University of Applied Sciences Osnabrueck, Germany) for the critical reading of and valuable comments on this paper.

AUTHORS' CONTRIBUTIONS

Conceptualization, I.H. and S.R.; Investigation, S.R. and G.A.; Data analysis, S.R.; Writing - original draft preparation, S.R.; Writing - review and editing, I.H. and S.R.; Funding acquisition, I.H. All authors have read and agreed to the published version of the manuscript.

DATA AVAILABILITY STATEMENT

The isolated sequences of DREB gene from *T. urartu* and *Ae. tauschii* registered in NCBI with Accession Nr. MZ935739 and MZ935740.

CONFLICTS OF INTEREST

The authors declare no conflict of interest.

REFERENCES

Agarwal PK, Agarwal P, Reddy MK and Sopory SK (2006). Role of DREB transcription factors in abiotic and biotic stress tolerance in plants. *Plant Cell Rep.* 25: 1263-1274.

Allen MD, Yamasaki K, Ohme-Takagi M, Tateno M, et al. (1998). A novel mode of DNA recognition by a beta-sheet revealed by the solution structure of the GCC-box binding domain in complex with DNA. *EMBO J.* 17: 5484-5496.

Altschul SF, Madden TL, Schäffer AA, Zhang J, et al. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 25: 3389-3402.

Andeani JK., Mohsenzadeh S and Mohabatkar H (2009). Isolation and characterization of partial DREB gene from four Iranian *Triticum aestivum* cultivars. *World J. Agric. Sci.* 5: 561-566.

Bernhofer M, Dallago C, Karl T, Satagopam V, et al. (2021). Predictprotein-predicting protein structure and function for 29 years. *Nucleic Acids Res.* 49 (W1): W535-W540. doi:org/10.1093/nar/gkab354

Buchan DW and Jones DT (2019). The PSIPRED protein analysis workbench: 20 years on. *Nucleic Acids Res.* 47(1): 402-407.

Chen J, Xia X and Yin W (2007). Advances on plant DREB transcription factors and their genetic transformation. *Mol. Plant Breed*. 5: 29-35. doi: 10.3969/j.issn.1672-416X.2007.z1.003

Concia L, Veluchamy A, Ramirez-Prado JS, et al. (2020). Wheat chromatin architecture is organized in genome territories and transcription factories. *Genome Biol*. 21: 104. doi:org/10.1186/s13059-020-01998-1

Country report on the state of plant genetic resources for food and agriculture (2006). Baku, Azerbaijan.

Duvaud S, Gabella C, Lisacek F, Stockinger H, et al. (2021). Expasy, the Swiss Bioinformatics Resource Portal, as designed by its users. *Nucleic Acids Res.* 49(W1): W216-W227. doi:org/10.1093/nar/gkab225

Feng K, Hou XL, Xing GM, Liu JX, et al. (2020). Advances in AP2/ERF super-family transcription factors in plant. *Crit. Rev. Biotechnol.* 40(6): 750-776.

Gaurav K, Arora S, Silva P, et al. (2022). Population genomic analysis of Aegilops tauschii identifies targets for bread wheat improvement. *Nat. Biotechnol.* 40: 422-431. doi.org/10.1038/s41587-021-01058-4

Hassan S, Berk K and Aronsson H (2021). Evolution and identification of DREB transcription factors in the wheat genome: modeling, docking and simulation of DREB proteins associated with salt stress. *J. Biomol. Struc. Dyn.* 1-14.

Huseynova IM, Rustamova SM and Mammadov AC (2013). Identification of Dreb1 genes involved in drought tolerance in wheat (*Triticum* L). In: Photosynthesis Research for Food, Fuel and the Future (Kuang T, Lu C, Zhang L, eds.). Springer, Berlin, Heidelberg

International Wheat Genome Sequencing Consortium. (2018). Shifting the limits in wheat research and breeding using a fully annotated reference genome. *Science.* 361(6403): eaar7191 doi: 10.1126/science.aar7191

Jones P, Binns D, Chang HY, Fraser M, et al. (2014). InterProScan 5: genome-scale protein function classification. *Bioinformatics*. 30: 1236-40.

Katoh K and Standley DM (2014). MAFFT: iterative refinement and additional methods. In: Multiple sequence alignment methods. Humana Press, Totowa, NJ.

Khan MS (2011). The role of DREB transcription factors in abiotic stress tolerance of plants. *Biotechnol. Equip.* 25(3): 2433-2442.

Lata C and Prasad M (2011). Role of DREBs in regulation of abiotic stress responses in plants. *J. Exp. Bot*. 62: 4731-4748.

Letunic I, Doerks T and Bork P (2015**).** SMART: recent updates, new developments and status in 2015. *Nucleic Acids Res*. 2015, 43. D1: D257-D260.

Li H, Deal KR, Luo MC, Ji W, et al. (2017). Introgression of the *Aegilops speltoides Su1-Ph1* Suppressor into Wheat. *Front. Plant Sci. 8*: 2163. doi:org/10.3389/fpls.2017.02163

Liu M, Wang Z, Xiao H, et al. (2018). Characterization of TaDREB1 in wheat genotypes with different seed germination under osmotic stress. *Hereditas*. 155: 26 doi:org/10.1186/s41065-018-0064-6

Liu Q, Kasuga M, Sakuma Y, Abe H, et al. (1998). Two transcription factors, DREB1 and DREB2, with an EREBP/AP2 DNA binding domain separate two cellular signal transduction pathways in drought-and low-temperature-responsive gene expression, respectively, in Arabidopsis. *Plant Cell.* 10: 1391-1406.

Mondini L, Nachit MM and Pagnotta MA (2015). Allelic variants in durum wheat (Triticum turgidum L. var. durum) DREB genes conferring tolerance to abiotic stresses. *Mol. Genet. Genom.* 290(2): 531-544.

Morran S, Eini O, Pyvovarenko T, Parent B, et al. (2011). Improvement of stress tolerance of wheat and barley by modulation of expression of DREB/CBF factors. *Plant Biotechnol. J.* 9(2): 230-249.

Murray MG and Thompson WF (1980). Rapid isolation of high molecular weight plant DNA. *Nucleic Acids Res.* 8: 4321-4325.

Niu X, Luo T, Zhao H, Su Y, et al. (2020). Identification of wheat DREB genes and functional characterization of TaDREB3 in response to abiotic stresses. *Gene.* 740: 144514.

Pandey B, Sharma P, Saini M, Pandey DM, et al. (2014). Isolation and characterization of dehydration-responsive element-binding factor 2 (DREB2) from Indian wheat (*Triticum aestivum* L) cultivars. *Aust. J. Crop. Sci.* 8(1):44.

Peng J, Sun D and Nevo E (2011). Wild emmer wheat,'Triticum dicoccoides', occupies a pivotal position in wheat domestication process. *Aust. J. Crop Sci. 5*(9): 1127-1143.

Rustamova S, Shrestha A, Naz AA and Huseynova I (2020). Expression profiling of Dreb1 and evaluation of vegetation indices in contrasting wheat genotypes exposed to drought stress. *Plant Gene.* 25: 2352-4073.

Sakuma Y, Liu Q, Dubouzet JG, Abe H, et al. (2002). DNA-binding specificity of the ERF/AP2 domain of Arabidopsis DREBs, transcription factors involved in dehydration-and cold-inducible gene expression. *Biochem. Biophys. Res. Commun.* 290: 998-1009.

Sazegari S and Niazi A (2012). Isolation and molecular characterization of wheat ('Triticum aestivum') dehydration responsive element binding factor (DREB) isoforms. *Aust. J. Crop. Sci.* 6: 1037-1044.

Shahzad R, Shakra Jamil SA, Nisar A, Amina Z, et al. (2021). Harnessing the potential of plant transcription factors in developing climate resilient crops to improve global food security: Current and future perspectives. *Saudi J. Biol. Sci.* 28(4): 2323.

Shen YG, Zhang WK, He SJ, Zhang JS, et al. (2003). An EREBP/AP2-type protein in *Triticum aestivum* was a DRE-binding transcription factor induced by cold, dehydration and ABA stress. *Theor. Appl. Genet.* 106(5): 923-930.

Solovyev V, Kosarev P, Seledsov I and Vorobyev D (2006). Automatic annotation of eukaryotic genes, pseudogenes and promoters. *Genome Biol.* 7 Suppl 1: S 10.1-12.

Walkowiak S, Gao L, Monat C, Haberer G, et al. (2020). Multiple wheat genomes reveal global variation in modern breeding. *Nature. 588*(7837): 277-283.

Wang H, Zhu Y, Yuan P, Song S, et al. (2021). Response of wheat dreb transcription factor to osmotic stress based on DNA methylation. *Int. J. Mol. Sci. 22*(14): 7670.

Wang JR, Luo MC, Chen ZX, You FM, et al. (2013) *Aegilops tauschii* single nucleotide polymorphisms shed light on the origins of wheat D-genome genetic diversity and pinpoint the geographic origin of hexaploid wheat. *New Phytol.* 198: 925–937.

Wei B, Jing R, Wang C, Chen J, et al. (2009). Dreb1 genes in wheat (*Triticum aestivum* L.): development of functional markers and gene mapping based on SNPs. *Mol. Breed. 23*(1): 13-22.

Xu Y, Sun FY, Ji C, Hu QW, et al. (2019). Nucleotide diversity patterns at the DREB 1 transcriptional factor gene in the genome donor species of wheat (*Triticum aestivum* L). *Plos one. 14*(5): e0217081.

Yamaguchi-Shinozaki K and Shinozaki K (1994). A novel cis-acting element in an Arabidopsis gene is involved in responsiveness to drought, low-temperature, or high-salt stress. *Plant Cell.* 6: 251-264.