Article

Not peer-reviewed version

Analysis of AP2/ERF Domain of DREB Transcription Factor in Several Wheat Species

AmirHossein MohammadNejad Aynede *, Kianoush Keshani *, Sasan Mohsenzadeh, Atena Eslami-Farouji

Posted Date: 17 August 2023

doi: 10.20944/preprints202308.1160.v2

Keywords: AP2/ERF; Amino acid analysis; Motif analysis; Structure of AP2 domain; Transcription factor; Dehydration-responsive element; Wheat

Article

Analysis of AP2/ERF Domain of DREB Transcription Factor in Several Wheat Species

AmirHossein MohammadNejad Aynede *,†, Kianoush Keshani *,†, Sasan Mohsenzadeh and Atena Eslami-Farouji

Department of Biology, Faculty of Science, Shiraz University, Shiraz 7195615735, Iran; mohsenz@shirazu.ac.ir (S.M.); atena.eslami@shirazu.ac.ir (A.E.F.)

* Correspondence: amiraynede@gmail.com (A.H.M.A.); kia.keshani@gmail.com (K.K.)

† These authors contributed equally to this work.

Abstract: The Apetala2/ethylene response factor superfamily refers to a group of transcription factors that share a conserved AP2 DNA binding domain. These factors have been found to have different roles in plant responses to both biotic and abiotic stresses. Samples of hexaploid wheat, tetraploid pasta (or durum wheat), and diploid wheat progenitors were selected. The 29 dehydration-responsive element binding transcription factors in these samples were downloaded from NCBI GenBank for six different countries: Iran, China, Italy, France, Afghanistan, and Azerbaijan. The AP2 domain sequences were identified from the dehydration-responsive element binding transcription factors using PROSITE, ProDom, and SMART software. Next, all sequences were aligned using Multalin and Jalview software. The aligned sequences were then analyzed to identify amino acid locations, types, and frequencies. The multiple alignments showed that approximately 76% of the amino acid residues in the AP2 domains are conserved. According to the amino acid analysis, alanine, serine, and glutamic acid are the most abundant amino acids found in three motifs. The performed phylogenetic analysis illustrates the role of geographical effects on the transcription factor sequences of bread wheat in the Middle East. Significant differences were found between Iranian and Chinese transcription factor sequences. Moreover, genetic variation was observed in the transcription factors of Italian sequences found in pasta and wheat progenitors. Motif structures play a critical role in the domain organization of wheat proteins to enhance the characteristics of assorted metabolic pathways. The structure of the AP2 domain was analyzed by several programs, I-TASSER for instance, to identify the α-helix, β-sheets, and the regions of some significant amino acids in the 3-D model.

Keywords: AP2/ERF; amino acid analysis; motif analysis; structure of AP2 domain; transcription factor; dehydration-responsive element; wheat

1. Introduction

The AP2/ERF superfamily consists of transcription factors (TF) with a conserved DNA-binding domain called AP2. This superfamily plays important roles in plant responses to biotic and abiotic stresses. The domain consists of 57-66 amino acids [1–3]. The DREB proteins belong to the ERF family of transcription factors and are involved in the ABA-independent signal transduction pathway. These proteins bind to dehydration-responsive elements (DRE) in the promoter region of genes to regulate their expression. The AP2/ERF gene family consists of five major groups named AP2, DREB, ERF, RAV, and Soloist. AP2 ERF transcription factors (TFs) possess one or two AP2 domains, while DREB, ERF, and RAV groups have a single AP2/ERF domain [4]. The Soloist group is a small group of transcription factors that have a structure similar to an AP2 domain, but deviate considerably [5]. The AP2/ERF gene family has a wide range of transcription factor genes that play essential roles in plant growth and development, signal transduction, and abiotic stresses [1]. They are also involved in developmental processes and plant morphogenesis [6]. Additionally, these genes have roles in

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 17 August 2023 doi:10.20944/preprints202308.1160.v2

determining leaf epidermal cells, floral organ patterning, spike-let meristem differentiation [7], seed yield, and seed mass [8,9]. These AP2/ERF transcription factors have been found in various plants such as Arabidopsis, rice [10], maize [3], wheat, and soybean. Furthermore, studies have demonstrated their existence in ciliates and protists [11,12]. AP2/ERF is abundantly present in wheat. This transcription factor has various functions and responses [9]. For instance, AP2 domain genes play a vital role in the development of reproductive and vegetative organs. However, a missense mutation in this gene reduces spike density [13–16].

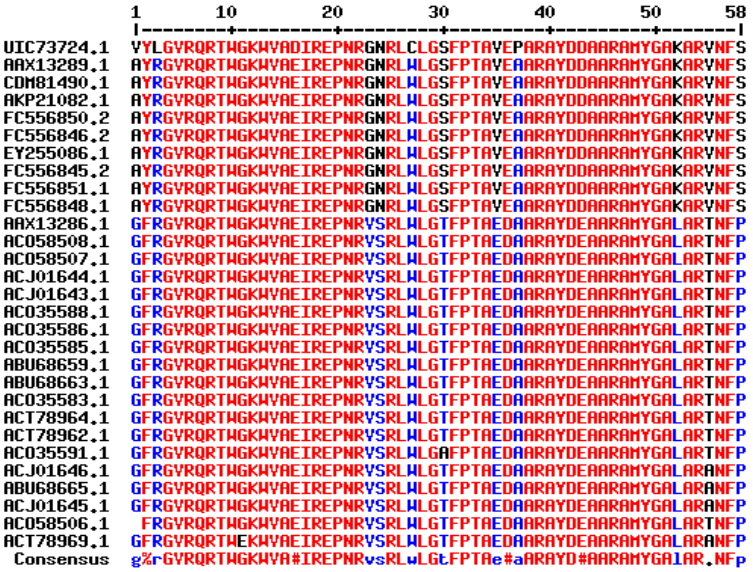
There are two main species of wheat: common or hexaploid wheat (*Triticum aestivum* L.), which is grown and consumed worldwide for the production of bread, pasta, and cakes; the other species are pasta or durum wheat (*Triticum turgidum* L. subsp. durum Desf.), which is tetraploid (4n = 28) and originated in the Middle East thousands of years ago. One of the ancestors of wheat is *Aegilops speltoides*, which is a diploid grass with seven pairs of chromosomes (2n = 14) [17].

The alignment analysis, amino acid dissection, and structural analysis have the crucial potential to make the character of AP2 domains appear. From this point, in some wheat species to create genetically enhanced crops resistant to multiple abiotic stresses, AP2/ERF domain analysis of DREB transcription factors might suggest changes in the gene and amino acid composition for better knowledge.

2. Results

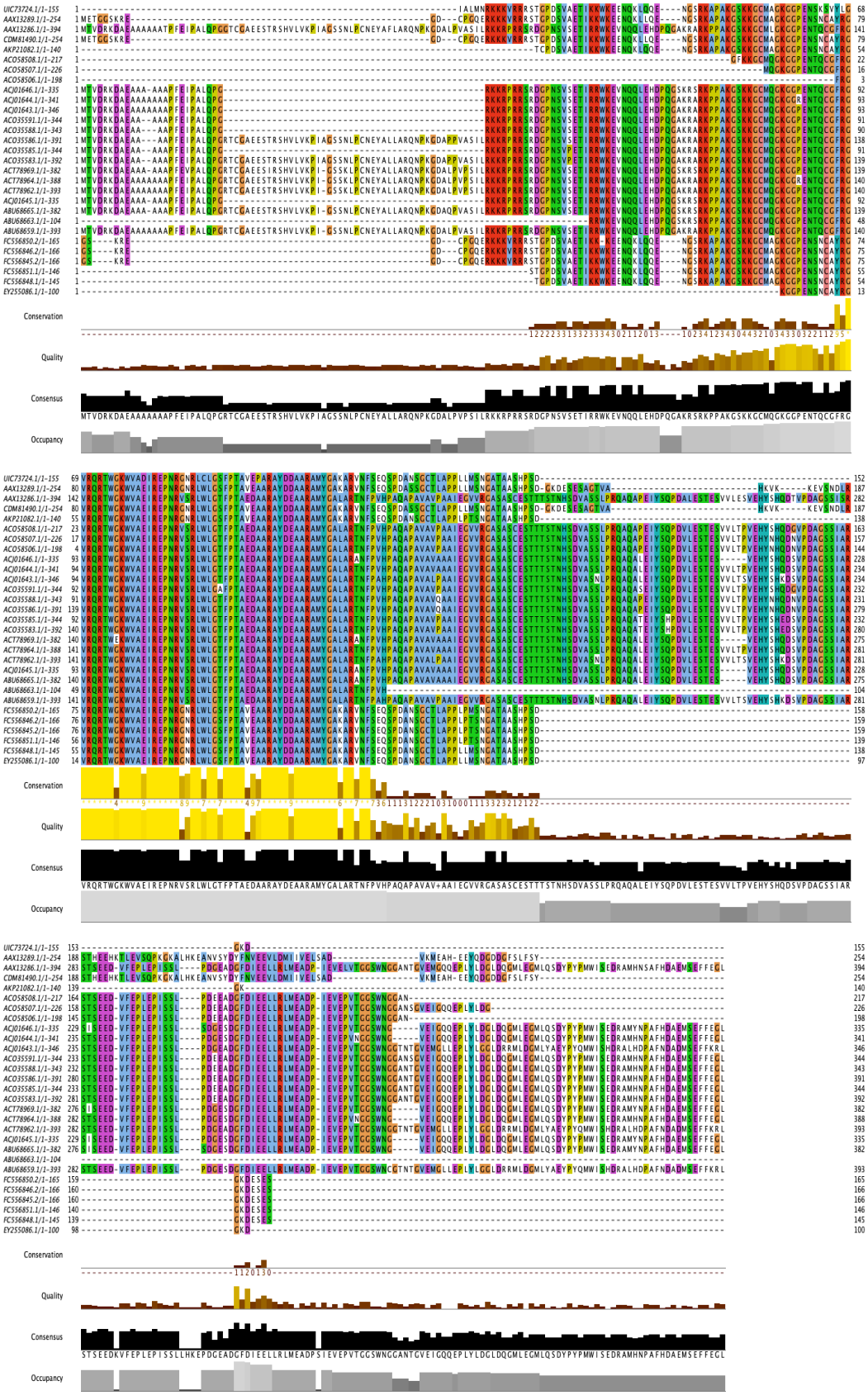
2.1. Sequence Analysis of DREB Proteins

Analysis of the AP2 domains of the DREB-TS proteins was performed using several alignment tools. The alignment results of the DREB sequences suggest that they contain 58 conserved amino acid residues. When the AP2 domain sequences were aligned, the result showed 44 (about 76 percent) highly conserved amino acid residues, namely Tyr-2, Gly-4, Val-5, Arg-6, Gln-7, Arg-8, Thr-9, Trp-10, Gly-11, Lys-12, Trp-13, Vla-14, Ala-15, Ile-17, Arg-18, Glu-19, Pro-20, Asn-21, Arg-22, Arg-25, Leu-26, Leu-28, Gly-29, Phe-31, Pro-32, Thr-33, Ala-34, Ala-38, Arg-39, Ala-40, Tyr-41, Asp-42, Ala-44 Ala-45, Arg-46, Ala-47, Met-48, Tyr-49, Gly-50, Ala-51, Ala-53, Arg-54, Asn-56, and Phe-57 (Figure 1). All 29 protein sequences with their AP2 domain regions were considered to determine the conserved residues and some other data such as quality, consensus, and occupancy (Figure 2). Alanine (A), glutamine (Q), lysine (K), valine (V), isoleucine (I), glutamic acid (E), serine (S), and methionine (M) in conserved amino acid residues were highlighted (Figure 3).



Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 17 August 2023 doi:10.20944/preprints202308.1160.v2

Figure 1. 29 AP2 domains were subjected to Multalin alignment. The alignment was performed on the segment of the domains possessing the desired functional domain, which includes 58 conserved amino acid residues.



Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 17 August 2023 doi:10.20944/preprints202308.1160.v2

Figure 2. Analysis of multiple sequence alignment. The sequences are aligned and edited using Jalview. Residues are shaded based on their conservation level.

Figure 3. Fifty-eight conserved amino acid residues were identified from the Multalin alignment. Alanine (A): Green; Valine (V), and Glutamic Acid (E): Orange; Glutamine (Q), Lysine (K), Isoleucine (I), Serine (S), and Methionine (M): The yellow residues are important amino acids.

2.2. Conserved Motif Analysis

MEME analysis identified three conserved motifs. Table 1 shows that motif 1 shares significant similarities with the AP2 domain. Table 1 shows that in some instances, two other motifs are present with different sequences compared to motif 1. All three motifs were present in 18 out of 29 protein sequences. Motifs 2 and 3 were not present in any protein sequences. Motif 1 was observed solely in 11 sequences, which included UIC73724.1, AAX13289.1, CDM81490.1, AKP21082.1, ABU68663.1, FC556850.2, FC556846.2, FC556845.2, FC556851.1, FC556848.1, and EY255086.1. A total of 18 sequences were detected through the combination of three motifs (Figure 4).

Table 1. The default settings of the MEME software identified three motif sequences.

Motif Number	Amino Acid Sequence
1	FRGVRQRTWGKWVAEIREPNRVSRLWLGTFPTAEDAARAYDEAARAMYGA
2	IEGVVRGASASCESTTTSTNHSDVASSLPRQAQALEIYSQPDVLESTESV
3	STSEEDVFEPLEPISSLPDGEADGFDIEELLRLMEADPIEVEPVTGGSWN

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 17 August 2023 doi:10.20944/preprints202308.1160.v2

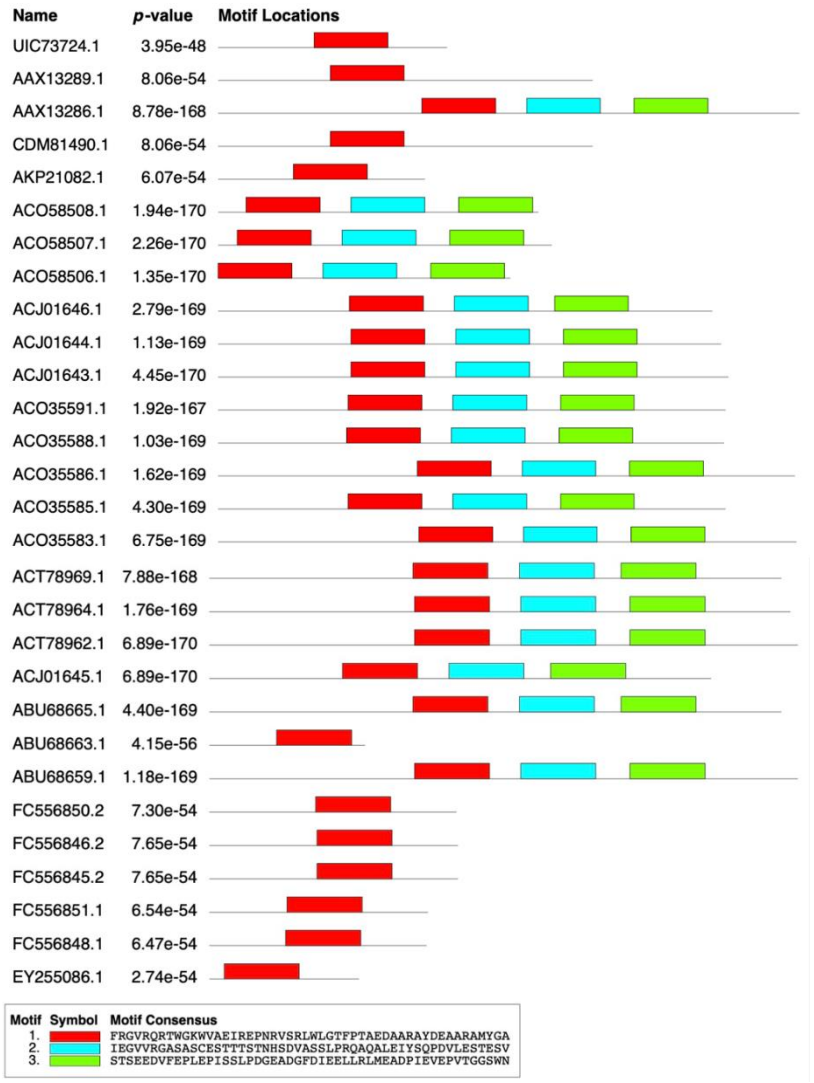


Figure 4. Distribution of motifs for transcription factors in a select wheat sample. Three motifs that are conserved and identified by MEME are denoted by differently colored boxes, namely, red for motif 1, cyan for motif 2, and green for motif 3. Relative positions of these motifs are also provided.

According to PortParam analysis, alanine constituted 18.0% of motif 1, while serine and glutamic acid were the most prevalent amino acids in motif 2 and 3, respectively, both accounting for 20.0% (Table 2).

Table 2. Three identified motifs were subjected to amino acid analysis. Motif 1 contains the highest percentage of alanine, motif 2 has the highest percentage of serine, and motif 3 has the highest percentage of glutamic acid compared to other amino acids.

Amino acid	Motif 1	Motif 2	Motif 3	Amino acid	Motif 1	Motif 2	Motif 3
Ala (A)	**18.0%**	10.0%	4.0%	Lys (K)	2.0%	0.0%	0.0%
Arg (R)	16.0%	4.0%	2.0%	Met (M)	2.0%	0.0%	2.0%
Asn (N)	2.0%	2.0%	2.0%	Phe (F)	4.0%	0.0%	4.0%
Asp (D)	4.0%	4.0%	10.0%	Pro (P)	4.0%	4.0%	10.0%
Cys (C)	0.0%	2.0%	0.0%	Ser (S)	2.0%	**20.0%**	10.0%

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 17 August 2023 doi:10.20944/preprints202308.1160.v2

Gln (Q)	2.0%	6.0%	0.0%	Thr (T)	6.0%	10.0%	4.0%
Glu (E)	8.0%	10.0%	**20.0%**	Trp (W)	6.0%	0.0%	2.0%
Gly (G)	8.0%	4.0%	8.0%	Tyr (Y)	4.0%	2.0%	0.0%
His (H)	0.0%	2.0%	0.0%	Val (V)	6.0%	10.0%	6.0%
Ile (I)	2.0%	4.0%	6.0%	Pyl (O)	0.0%	0.0%	0.0%
Leu (L)	4.0%	6.0%	10.0%	Sec (U)	0.0%	0.0%	0.0%

2.3. Phylogenetic Analysis of AP2 Domains

To accurately demonstrate the evolutionary relationships among the AP2 domains, we performed a phylogenetic analysis. As a result, all 29 AP2 domains were segregated into two groups. Eleven AP2 domains were found in Group 1 while Group 2 contained eighteen domains (Figure 5). Transcription Factors (TFs) from Lolium arundinaceum (Schreb.) Darbysh (Poaceae) were used. They were utilized as an out-of-group protein and separated from the other transcription factors.

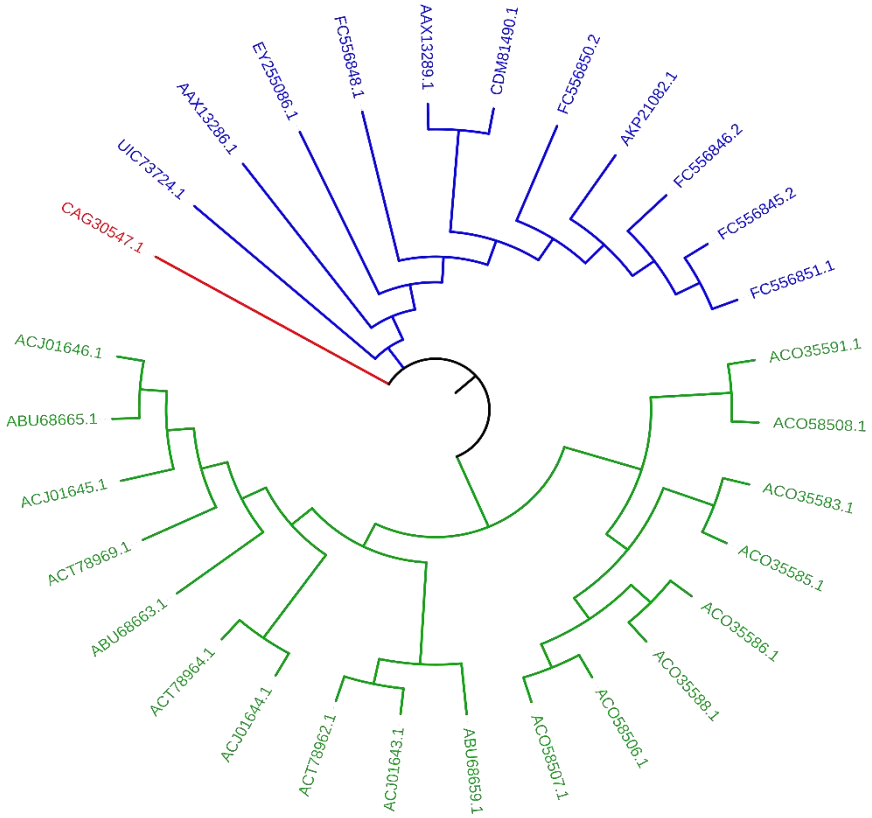


Figure 5. An analysis was conducted on the phylogeny of transcription factors. The dendrogram was built using the MEGA package with the maximum likelihood method and 100 bootstrap replicates. The out-of-group (CAG30547.1), group 1, and group 2 were assigned the colors red, blue, and red, respectively.

2.4. Structural Analysis of AP2 Domain

The C-score and TM-score (i.e., Template Modeling score) of the modeled AP2 domain (Figure 6) suggested a high degree of confidence in the model, with a C-score of 1.36, which ensured that the correct topology was attained [18–20]. We used the Ramachandran plot, a tool for assessing the quality of protein structures, to select and validate the model with the highest C-score (Figure 7). Our analysis of the Ramachandran plot revealed that the majority of residues were located in the core region (as shown in red), which indicates an acceptable 3-D model. The statistics provided in the

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 17 August 2023 doi:10.20944/preprints202308.1160.v2

Ramachandran plot section illustrated the detailed structure of the AP2 domain, which comprised 25% helix, 34% beta, 39% coil, and 13% turn. Furthermore, the Z score (i.e., a measure of confidence in the quality of a structural model) for the AP2 domain was -3.44, which provided additional evidence supporting the accuracy of the model (Figure 8) [21,22]. The I-TASSER program predicted that the AP2 domain functions as a DNA-binding transcription factor, with a high Gene Ontology (GO) Score of 0.75.

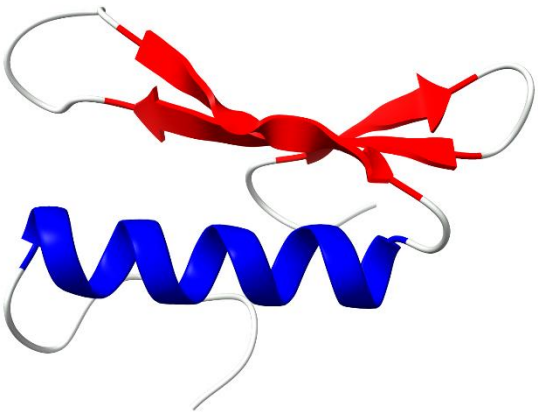


Figure 6. Three-dimensional model of the AP2 domain. Specifically, this model showcases three β-sheets (VRQR, KWVAEIR, and RLWLGTF) and an α-helix (AEDAARAYDEAARAMY) shown in red and blue colors, respectively.

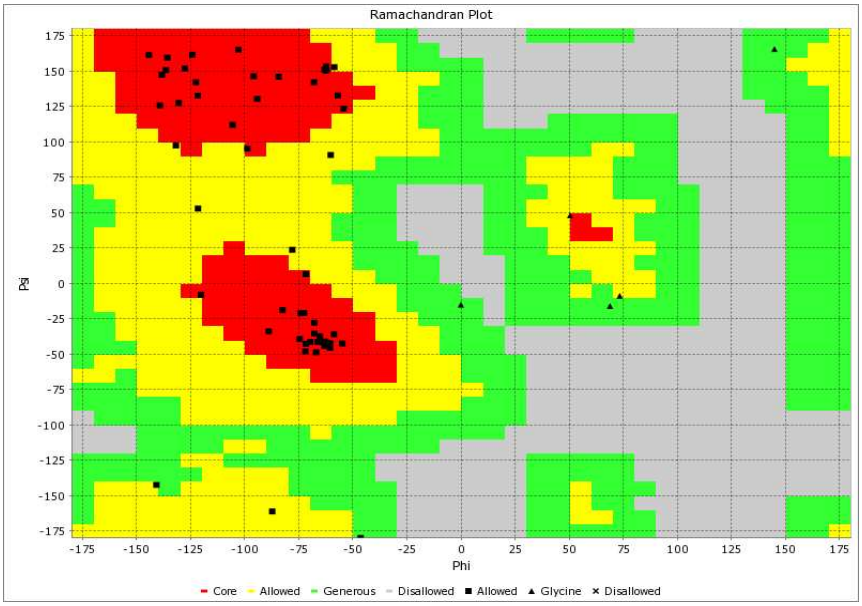


Figure 7. The Ramachandran plot for the AP2 domain. The red section represents the core region, the yellow section represents the allowed region, and the green section represents the generous region. The model is valid as most of the residues were located in the core region of the plot.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 17 August 2023 doi:10.20944/preprints202308.1160.v2

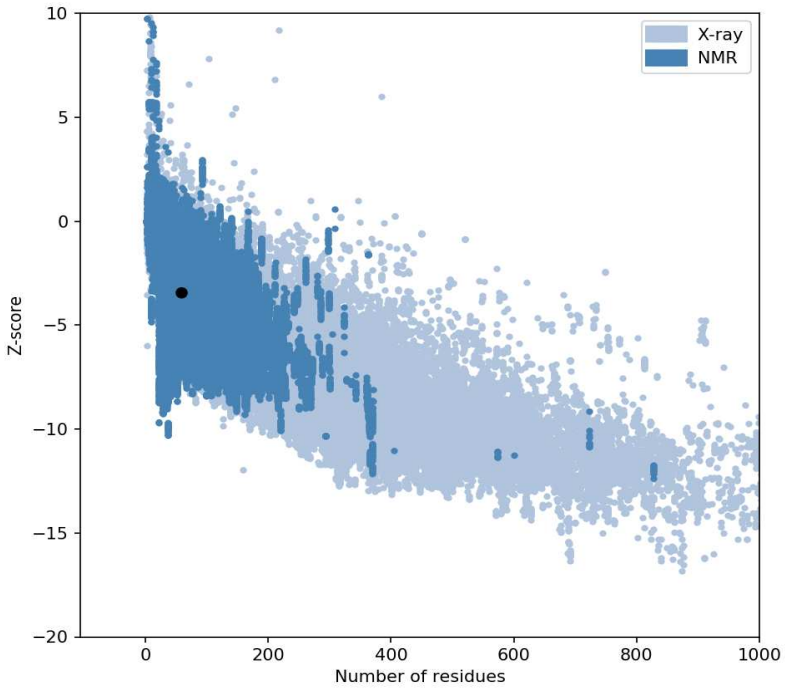


Figure 8. ProSA-web z-score of the PDB model determined by X-ray crystallography (light blue) or NMR spectroscopy (dark blue). The z-score is -3.44 and marked as a dot.

3. Discussion

Dehydration-responsive element binding (DREB) transcription factors increase the expression of abiotic tolerance pathway genes by binding to DNA upstream of genes. Most DREB genes coding for the AP2 domain in wheat have one AP2 domain, as shown in [23]. AP2/ERF is one of the largest families of plant transcription factors, playing a crucial role in abiotic stress responses in wheat. Gene structure analysis has revealed that approximately 85% of AP2/ERF genes lack introns, and the evolution of the wheat genome is the leading cause of this phenomenon [16]. Additionally, six partial DNA genomic sequences of Iranian AP2/ERF transcription factors previously submitted to NCBI GenBank showed similarity to their cDNA counterparts [24,25], as revealed by a BLASTX search conducted in this study.

The selected transcription factor sequences, including six Iranian, one Afghan, and one Azerbaijani cultivar of bread wheat (T. aestivum), are part of group 1 in the phylogenetic analysis. However, the transcription factor sequences from two sequences of bread wheat from China, one from France, and five from Italy are in group 2 (Table 3). This indicates the role of geographical location on transcription factor sequences of bread wheat in the Middle East. There is a genetic variation in the transcription factors of Italian sequences found in pasta and wheat progenitors. Specifically, one sequence of the transcription factor from Italian T. turgidum subsp. durum is in group 1, and four are in group 2. Additionally, two sequences of Italian *A. speltoi*-des var. speltoides are in group 1, and six are in group 2. The distinction between the sequences from Iran and China in the phylogenetic analysis of TS sequences (Table 3) has been reported in another study on the DREB partial gene sequences (500 bp long) of eight Iranian and ten Chinese bread wheat cultivars [26]. It was explained that the phylogenetic analysis of the Chinese sequences was completely distinct from the Iranian sequences. The Iranian sequences are divided into seven groups; in contrast, the Chinese sequences are divided into three groups only. This indicates that the Iranian DREB sequences exhibit greater biodiversity. However, the AP2 domains of the chosen DREB transcription factors, as

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 17 August 2023 doi:10.20944/preprints202308.1160.v2

presented in the Multalin alignment of AP2 domains (Figure 1), demonstrate a greater degree of conservation (approximately 76 percent).

Table 3. 29-transcription factors and their AP2 domain sequences were analyzed using SMART.

Accession number of TS sequence	AP2 Domain Sequence	Country	Species
EY255086.1	AYRGVRQRTWGKW*V*AEIR*E*PNRGNRLWLGSFPTAVEAARAYDDAARAMYGAKARVNFSEQSPDA	Iran	*Triticum aestivum* cv. Shahi1
ABU68663.1	GFRGVRQRTWGKW*V*AEIR*E*PNRVSRLWLGTFPTAEDAARAYDEAARAMYGALARTNFPVH	Italy	*Triticum turgidum* subsp. *Durum*1
AKP21082.1	AYRGVRQRTWGKW*V*AEIR*E*PNRGNRLWLGSFPTAVEAARAYDDAARAMYGAKARVNFSEQSPDA	Afghanistan	*Triticum aestivum* cv. Kalak Afghani
FC556848.1	AYRGVRQRTWGKW*V*AEIR*E*PNRGNRLWLGSFPTAVEAARAYDDAARAMYGAKARVNFSEQSPDA	Iran	*Triticum aestivum* cv. Omid
FC556851.1	AYRGVRQRTWGKW*V*AEIR*E*PNRGNRLWLGSFPTAVEAARAYDDAARAMYGAKARVNFSEQSPDA	Iran	*Triticum aestivum* cv. Zarrin
UIC73724.1	VYLGVRQRTWGKW*V*ADIR*E*PNRGNRLCLGSFPTAVEPARAYDDAARAMYGAKARVNFSEQSPDA	Azerbaijan	*Triticum aestivum* cv. Azerbaijan
FC556850.2	AYRGVRQRTWGKW*V*AEIR*E*PNRGNRLWLGSFPTAVEAARAYDDAARAMYGAKARVNFSEQSPDA	Iran	*Triticum aestivum* cv. Shahi2
FC556846.2	AYRGVRQRTWGKW*V*AEIR*E*PNRGNRLWLGSFPTAVEAARAYDDAARAMYGAKARVNFSEQSPDA	Iran	*Triticum aestivum* cv. Bayat
FC556845.2	AYRGVRQRTWGKW*V*AEIR*E*PNRGNRLWLGSFPTAVEAARAYDDAARAMYGAKARVNFSEQSPDA	Iran	*Triticum aestivum* cv. Alvand
ACO58506.1	FRGVRQRTWGKW*V*AEIR*E*PNRVSRLWLGTFPTAEDAARAYDEAARAMYGALARTNFPVHPAQA	Italy	*Aegilops speltoides* var. speltoides
ACO58508.1	GFRGVRQRTWGKW*V*AEIR*E*PNRVSRLWLGTFPTAEDAARAYDEAARAMYGALARTNFPVHPAQA	Italy	*Aegilops speltoides* var. *speltoides*

 doi:10.20944/preprints202308.1160.v2

ACO58507.1	GFRGVRQRTWGKW***V***AEIR***E***PNRVSRLWLGTFPTAEDAARAYDEAARAMYGALARTNFPVHPAQA	Italy	*Aegilops speltoides* var. *speltoides*
AAX13289.1	AYRGVRQRTWGKW***V***AEIR***E***PNRGNRLWLGSFPTAVEAARAYDDAARAMYGAKARVNFSEQSPDA	China	*Triticum aestivum* cv. China
CDM81490.1	AYRGVRQRTWGKW***V***AEIR***E***PNRGNRLWLGSFPTAVEAARAYDDAARAMYGAKARVNFSEQSPDA	France	*Triticum aestivum* cv. France
ACJ01646.1	GFRGVRQRTWGKW***V***AEIR***E***PNRVSRLWLGTFPTAEDAARAYDEAARAMYGALARANFPVHPAQA	Italy	*Triticum turgidum* subsp. *durum*2
ACJ01645.1	GFRGVRQRTWGKW***V***AEIR***E***PNRVSRLWLGTFPTAEDAARAYDEAARAMYGALARANFPVHPAQA	Italy	*Triticum turgidum* subsp. *durum*3
ACJ01644.1	GFRGVRQRTWGKW***V***AEIR***E***PNRVSRLWLGTFPTAEDAARAYDEAARAMYGALARTNFPVHPAQA	Italy	*Triticum aestivum* cv. Italy1
ACO35588.1	GFRGVRQRTWGKW***V***AEIR***E***PNRVSRLWLGTFPTAEDAARAYDEAARAMYGALARTNFPVHPAQA	Italy	*Aegilops speltoides* var. *speltoides*
ACO35591.1	GFRGVRQRTWGKW***V***AEIR***E***PNRVSRLWLGAFPTAEDAARAYDEAARAMYGALARTNFPVHPAQA	Italy	*Aegilops speltoides* var. *speltoides*
ACO35585.1	GFRGVRQRTWGKW***V***AEIR***E***PNRVSRLWLGTFPTAEDAARAYDEAARAMYGALARTNFPVHPAQA	Italy	*Aegilops speltoides* var. *speltoides*
ACJ01643.1	GFRGVRQRTWGKW***V***AEIR***E***PNRVSRLWLGTFPTAEDAARAYDEAARAMYGALARTNFPAHPAQA	Italy	*Triticum aestivum* cv. Italy2
ACT78969.1	GFRGVRQRTWEKW***V***AEIR***E***PNRVSRLWLGTFPTAEDAARAYDEAARAMYGALARANFPVHPAQA	Italy	*Triticum aestivum* cv. Italy3
ABU68665.1	GFRGVRQRTWGKW***V***AEIR***E***PNRVSRLWLGTFPTAEDAARAYDEAARAMYGALARANFPVHPAQA	Italy	*Triticum aestivum* cv. Italy4
ACT78964.1	GFRGVRQRTWGKW***V***AEIR***E***PNRVSRLWLGTFPTAEDAARAYDEAARAMYGALARTNFPVHPAQA	Italy	*Triticum aestivum*

 doi:10.20944/preprints202308.1160.v2

ACO35586.1	GFRGVRQRTWGKW***V***AEIR***E***PNRVSRLWLGTFPTAEDAARAYDEAARAMYGALARTNFPVHPAQA	Italy	cv. Italy5 *Aegilops speltoides* var. *speltoides*
ACO35583.1	GFRGVRQRTWGKW***V***AEIR***E***PNRVSRLWLGTFPTAEDAARAYDEAARAMYGALARTNFPVHPAQA	Italy	*Aegilops speltoides* var. *speltoides*
ACT78962.1	GFRGVRQRTWGKW***V***AEIR***E***PNRVSRLWLGTFPTAEDAARAYDEAARAMYGALARTNFPAHPAQA	Italy	*Triticum turgidum* subsp. *Durum*4
ABU68659.1	GFRGVRQRTWGKW***V***AEIR***E***PNRVSRLWLGTFPTAEDAARAYDEAARAMYGALARTNFPAHPAQA	Italy	*Triticum turgidum* subsp. *Durum*5
AAX13286.1	GFRGVRQRTWGKW***V***AEIR***E***PNRVSRLWLGTFPTAEDAARAYDEAARAMYGALARTNFPVHPAQA	China	*Triticum aestivum* cv. China

Transcription factors (TFs) in the AP2/ERF superfamily possess one or more AP2/ERF domains that contain conserved amino acids at positions 14 (alanine, A) and 19 (aspartic acid, D). In contrast, the DREB superfamily TFs have different conserved amino acids at the same positions, with valine (V) at position 14 and glutamic acid (E) at position 19 [27]. Table 1 shows that all AP2 domain sequences of the 29 DREB transcription factors in this study have conserved valine (V) and glutamic acid (E) amino acids at positions 14 and 19, respectively, as indicated by italicized and bolded V and E letters. Val14 is crucial for DREB binding to the dehydration-responsive element (DRE) cis-acting elements, whereas Glu19 plays a relatively less significant role [4].

The MEME software reveals three protein motifs with distinct sequences, and their amino acid composition is presented in Table 1. In the chosen wheat species, DREB transcription factors contained all three motifs in 18 out of 29 sequences, and the first motif represented the AP2 domain. Therefore, gene rearrangement may cause variations in the structure of motifs, which have significant roles in the domain arrangements of wheat proteins that enhance the properties of diverse metabolic pathways [28].

The analysis of three motifs revealed that alanine, serine, and glutamic acid are the most abundant amino acids in motifs 1, 2, and 3, respectively (Table 2). Glutamic acid and serine among hydrophilic amino acids have a considerable positive effect, especially in cases of high net charge; hence, the presence of serine and glutamic acid has the potential to improve protein solubility [29]. Out of 22 amino acids presented in Table 2, motif 1 has more amino acid diversity (18) than motif 2 (16) and motif 3 (15).

The 3-D structure of the AP2 domain used for DNA binding consists of three β-sheets and one α-helix. The conserved Ala37 in the α-helix is crucial for binding to the GCC box. According to a study [30], Ala38 is essential for maintaining the stability and structure of the hydrogen bonds between strand 2 and strand 3 of the β-sheet as depicted in Figure 3. The function of ERF proteins' AP2 domain depends on the α-helix, which may play a role in protein-protein interactions [31,32].

4. Materials and Methods

4.1. Database Search of DREB and AP2s in Wheat

We obtained 66 protein and 6 DNA sequences of the DREB transcription factor in fasta format from the NCBI, originating from six different countries: Iran (Shahi 1 & 2, Bayat, Al-vand, Zarrin, and Omid cultivars of bread wheat, submitted by one of the authors of this article), China (Triticum aestivum), Italy (T. aestivum, T. turgidum, and Aegilops speltoides), France (T. aestivum), Afghanistan (T. aestivum), and Azerbaijan (T. aestivum). We translated six DNA sequences into protein sequences by conducting a BLASTX search (https://blast.ncbi.nlm.nih.gov/Blast.cgi, accessed on 16 March 2022). Then, we checked all 72 protein sequences for the presence of the AP2 domain sequence using PROSITE (https://prosite.expasy.org, accessed on 4 April 2022), ProDom (http://prodom.prabi.fr/prodom/current/html/home.php, accessed on 13 April 2022), and SMART (http://smart.embl-heidelberg.de, accessed on 27 April 2022). As a result, we chose 29 protein samples that contained AP2 domains.

4.2. Multiple Sequence Alignment and Biochemical Characteristics

Table 3 lists twenty-nine protein samples that contain AP2 domains. The alignment of all AP2 domain sequences was done using Multalin (http://multalin.toulouse.inra.fr/multalin/, accessed on 19 May 2022) [33]. The analysis of twenty-nine AP2 domains was done based on their amino acid locations, types, and frequencies. To verify the data regarding the conserved region, Jalview (v2.11.2.6) with Clustal color [34] was used for further protein sequence analysis.

4.3. Composition and Motif Analysis

Multiple Em for Motif Elicitation Suite (MEME) was employed to identify additional protein motifs in the DREB transcription factors of selected wheat species (i.e., 17 cultivars of Triticum aestivum, 8 varieties of Aegilops speltoides, and 4 subspecies of Triticum turgidum). By using the default settings and profiles, MEME (v5.5.1) (https://meme-suite.org/meme/meme_5.5.1/tools/meme, accessed on 22 February 2023) identified novel motifs in the TS protein sequences. The amino acid composition analysis of the identified motifs was carried out by using the PortParam tool (https://web.expasy.org/protparam/, accessed on 29 February 2023) [35], and it provided information about the abundance and frequency of amino acids in the sequences.

4.4. Phylogenetic Analysis

The quality of the sequences was assessed and then manually aligned using the MUSCLE MEGA package (v11.0.13) [36] with the maximum likelihood method and default settings. For further validation of the tree, we constructed a Bayesian phylogenetic tree using MrBayes version 3.2 [37], following the methodology described in [38]. We followed the methodology from [38] to develop the Bayesian phylogenetic tree using MrBayes version 3.2 [37]. Finally, we used the Interactive Tree of Life (iTOL) to design the final format of the phylogenetic tree.

4.5. Structure and Modeling of AP2 Domain

The Iterative Threading Assembly Refinement (I-TASSER) program (https://zhanggroup.org/I-TASSER/, accessed on 17 March 2023) was used to determine the secondary and tertiary structures of the AP2 domain [18–20]. The Ramachandran plot was depicted using the model's PDB file in VADAR (v1.8.) (http://vadar.wishartlab.com, accessed on 5 April 2023) [39]. The Protein Structure Analysis program (ProSA) (https://prosa.services.came.sbg.ac.at/prosa.php, accessed on 20 April 2023) was used to report the Z-score, further confirming the overall quality of the model [21,22].

Author Contributions: Conceptualization, S.M., A.H.M.A. and K.K.; methodology, A.H.M.A., and K.K.; software, A.H.M.A., K.K., and A.EF.; validation, A.H.M.A.; writing—original draft preparation, A.H.M.A., S.M., and K.K.; writing—review and editing, A.H.M.A., S.M., A.EF., and K.K.; visualization, A.H.M.A., and K.K.; supervision, S.M. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Okamuro, J.K., Caster, B., Villarroel, R., Van Montagu, M.; Jofuku, K.D. (1997). The AP2 domain of APETALA2 defines a large new family of DNA binding proteins in Arabidopsis. *Proceedings of the National Academy of Sciences of the United States of America*, *94*, 7076–7081. https://doi.org/10.1073/pnas.94.13.7076
2. Ma, Q., Xia, Z., Cai, Z., Li, L., Cheng, Y., Liu, J.; Nian, H. (2019). *GmWRKY16* Enhances Drought and Salt Tolerance Through an ABA-Mediated Pathway in *Arabidopsis thaliana*. *Frontiers in plant science*, *9*, 1979. https://doi.org/10.3389/fpls.2018.01979
3. Zhu, P., Chen, Y., Zhang, J., Wu, F., Wang, X., Pan, T., Wei, Q., Hao, Y., Chen, X., Jiang, C.; Ji, K. (2021). Identification, classification, and characterization of AP2/ERF superfamily genes in Masson pine (Pinus massoniana Lamb.). *Scientific reports*, *11*, 5441. https://doi.org/10.1038/s41598-021-84855-w
4. Nakano, T., Suzuki, K., Fujimura, T.; Shinshi, H. (2006). Genome-wide analysis of the ERF gene family in Arabidopsis and rice. *Plant physiology*, *140*, 411–432. https://doi.org/10.1104/pp.105.073783
5. Sakuma, Y., Liu, Q., Dubouzet, J.G., Abe, H., Shinozaki, K.; Yamaguchi-Shinozaki, K. (2002). DNA-binding specificity of the ERF/AP2 domain of Arabidopsis DREBs, transcription factors involved in dehydration- and cold-inducible gene expression. *Biochemical and biophysical research communications*, *290*, 998–1009. https://doi.org/10.1006/bbrc.2001.6299
6. Feng, K., Hou, X.L., Xing, G.M., Liu, J.X., Duan, A.Q., Xu, Z.S., Li, M.Y., Zhuang, J.; Xiong, A.S. (2020). Advances in AP2/ERF super-family transcription factors in plant. *Critical reviews in biotechnology*, *40*, 750–776. https://doi.org/10.1080/07388551.2020.1768509
7. Aukerman, M.J.; Sakai, H. (2003). Regulation of flowering time and floral organ identity by a MicroRNA and its APETALA2-like target genes. *The Plant cell*, *15*, 2730–2741. https://doi.org/10.1105/tpc.016238
8. Jofuku, K.D., Omidyar, P.K., Gee, Z.; Okamuro, J.K. (2005). Control of seed mass and seed yield by the floral homeotic gene APETALA2. *Proceedings of the National Academy of Sciences of the United States of America*, *102*, 3117–3122. https://doi.org/10.1073/pnas.0409893102
9. Taketa, S., Amano, S., Tsujino, Y., Sato, T., Saisho, D., Kakeda, K., Nomura, M., Suzuki, T., Matsumoto, T., Sato, K., Kanamori, H., Kawasaki, S.; Takeda, K. (2008). Barley grain with adhering hulls is controlled by an ERF family transcription factor gene regulating a lipid biosynthesis pathway. *Proceedings of the National Academy of Sciences of the United States of America*, *105*, 4062–4067. https://doi.org/10.1073/pnas.0711034105
10. Licausi, F., Ohme-Takagi, M.; Perata, P. (2013). APETALA2/Ethylene Responsive Factor (AP2/ERF) transcription factors: mediators of stress responses and developmental programs. *The New phytologist*, *199*, 639–649. https://doi.org/10.1111/nph.12291.
11. Liu, S., Wang, X., Wang, H., Xin, H., Yang, X., Yan, J., Li, J., Tran, L.S., Shinozaki, K., Yamaguchi-Shinozaki, K.; Qin, F. (2013). Genome-wide analysis of ZmDREB genes and their association with natural variation in drought tolerance at seedling stage of Zea mays L. *PLoS genetics*, *9*, e1003790. https://doi.org/10.1371/journal.pgen.1003790
12. Rashid, M., Guangyuan, H., Guangxiao, Y., Hussain, J.; Xu, Y. (2012). AP2/ERF Transcription Factor in Rice: Genome-Wide Canvas and Syntenic Relationships between Monocots and Eudicots. *Evolutionary bioinformatics online*, *8*, 321–355. https://doi.org/10.4137/EBO.S9369
13. Gil-Humanes, J., Pistón, F., Martín, A.; Barro, F. (2009). Comparative genomic analysis and expression of the APETALA2-like genes from barley, wheat, and barley-wheat amphiploids. *BMC plant biology*, *9*, 66. https://doi.org/10.1186/1471-2229-9-66
14. Greenwood, J.R., Finnegan, E.J., Watanabe, N., Trevaskis, B.; Swain, S.M. (2017). New alleles of the wheat domestication gene *Q*reveal multiple roles in growth and reproductive development. *Development (Cambridge, England)*, *144*, 1959–1965. https://doi.org/10.1242/dev.146407
15. Simons, K.J., Fellers, J.P., Trick, H.N., Zhang, Z., Tai, Y.S., Gill, B.S.; Faris, J.D. (2006). Molecular characterization of the major wheat domestication gene Q. *Genetics*, *172*, 547–555. https://doi.org/10.1534/genetics.105.044727
16. Yu, Y., Yu, M., Zhang, S., Song, T., Zhang, M., Zhou, H., Wang, Y., Xiang, J.; Zhang, X. (2022). Transcriptomic Identification of Wheat AP2/ERF Transcription Factors and Functional Characterization of *TaERF-6-3A* in Response to Drought and Salinity Stresses. *International journal of molecular sciences*, *23*, 3272. https://doi.org/10.3390/ijms23063272

17. R. DePauw, and L. O'Brien, Wheat Breeding: Exploiting and Fixing Genetic Variation by Selection and Evaluation. In Encyclopedia of Food Grains, 2nd ed.; Colin Wrigley, Harold Corke, Koushik Seetharaman, Jon Faubion; Publisher: Academic Press, Cambridge, Massachusetts, US, 2016; Volume 4, pp. 279-286, doi.org/10.1016/B978-0-12-394437-5.00216-3
18. Yang, J.; Zhang, Y. (2015). I-TASSER server: new development for protein structure and function predictions. *Nucleic acids research*, *43*(W1), W174–W181. https://doi.org/10.1093/nar/gkv342
19. Zheng, W., Zhang, C., Li, Y., Pearce, R., Bell, E.W.; Zhang, Y. (2021). Folding non-homologous proteins by coupling deep-learning contact maps with I-TASSER assembly simulations. *Cell reports methods*, *1*, 100014. https://doi.org/10.1016/j.crmeth.2021.100014
20. Zhou, X., Zheng, W., Li, Y., Pearce, R., Zhang, C., Bell, E.W., Zhang, G.; Zhang, Y. (2022). I-TASSER-MTD: a deep-learning-based platform for multi-domain protein structure and function prediction. *Nature protocols*, *17*, 2326–2353. https://doi.org/10.1038/s41596-022-00728-0
21. Sippl M. J. (1993). Recognition of errors in three-dimensional structures of proteins. *Proteins*, *17*, 355–362. https://doi.org/10.1002/prot.340170404
22. Wiederstein, M.; Sippl, M.J. (2007). ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic acids research*, *35*(Web Server issue), W407–W410. https://doi.org/10.1093/nar/gkm290
23. Hassan, S., Berk, K.; Aronsson, H. (2022). Evolution and identification of DREB transcription factors in the wheat genome: modeling, docking and simulation of DREB proteins associated with salt stress. *Journal of biomolecular structure & dynamics*, *40*, 7191–7204. https://doi.org/10.1080/07391102.2021.1894980
24. Karimi, J., Mohsenzadeh, S., Mohabatkar, H. (2009). Isolation and characterization of partial DREB gene from four Iranian Triticum aestivum cultivars. World Journal of Agricultural Sciences, 5, 561-566.
25. Mohsenzadeh, S., Karimi, J., Mohabatkar, H. (2011). Study of dehydration-responsive element binding-factor gene in some Iranian bread wheat cultivars. Iranian Journal of Plant Biology. 3: 29-40.
26. Mohsenzadeh, S., Sherafat, A., Karimi, J. (2021). Alignment and phylogenetic tree analysis of some Iranian and Chinese wheat cultivars based on DREB partial gene sequences. The Fourth International Biotechnology Congress of Iran. Paper code (biotech12-05990388).
27. Sharma, M.K., Kumar, R., Solanke, A.U., Sharma, R., Tyagi, A.K.; Sharma, A.K. (2010). Identification, phylogeny, and transcript profiling of ERF family genes during development and abiotic stress treatments in tomato. *Molecular genetics and genomics : MGG*, *284*, 455–475. https://doi.org/10.1007/s00438-010-0580-1
28. Filiz, E.; Tombuloğlu, H. (2014). In silico analysis of DREB transcription factor genes and proteins in grasses. *Applied biochemistry and biotechnology*, *174*, 1272–1285. https://doi.org/10.1007/s12010-014-1093-x
29. Trevino, S.R., Scholtz, J.M.; Pace, C.N. (2007). Amino acid contribution to protein solubility: Asp, Glu, and Ser contribute more favorably than the other hydrophilic amino acids in RNase Sa. *Journal of molecular biology*, *366*, 449–460. https://doi.org/10.1016/j.jmb.2006.10.026
30. Liu, Y., Zhao, T.J., Liu, J.M., Liu, W.Q., Liu, Q., Yan, Y.B.; Zhou, H.M. (2006). The conserved Ala37 in the ERF/AP2 domain is essential for binding with the DRE element and the GCC box. *FEBS letters*, *580*, 1303–1308. https://doi.org/10.1016/j.febslet.2006.01.048
31. Jofuku, K.D., den Boer, B.G., Van Montagu, M.; Okamuro, J.K. (1994). Control of Arabidopsis flower and seed development by the homeotic gene APETALA2. *The Plant cell*, *6*, 1211–1225. https://doi.org/10.1105/tpc.6.9.1211
32. Büttner, M.; Singh, K.B. (1997). Arabidopsis thaliana ethylene-responsive element binding protein (AtEBP), an ethylene-inducible, GCC box DNA-binding protein interacts with an ocs element binding protein. *Proceedings of the National Academy of Sciences of the United States of America*, *94*, 5961–5966. https://doi.org/10.1073/pnas.94.11.5961
33. "Multiple sequence alignment with hierarchical clustering" F. CORPET, 1988, Nucl. Acids Res., 16 (22), 10881-10890
34. Waterhouse, A.M., Procter, J.B., Martin, D.M.A, Clamp, M. and Barton, G.J. (2009) "Jalview Version 2 - a multiple sequence alignment editor and analysis workbench", *Bioinformatics* 25 (9) 1189-1191 doi: 10.1093/bioinformatics/btp033
35. Gasteiger E., Hoogland C., Gattiker A., Duvaud S., Wilkins M.R., Appel R.D., Bairoch A.; *Protein Identification and Analysis Tools on the Expasy Server;* (In) John M. Walker (ed): The Proteomics Protocols Handbook, Humana Press (2005), pp. 571-607
36. Tamura, K., Stecher, G.; Kumar, S. (2021). MEGA11: Molecular Evolutionary Genetics Analysis Version 11. *Molecular biology and evolution*, *38*, 3022–3027. https://doi.org/10.1093/molbev/msab120
37. Ronquist F, Teslenko M, van der Mark P, Ayres DL, Darling A, Hohna S, Larget B, Liu L, Suchard MA, Huelsenbeck JP. (2012). MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. Syst Biol 61:539–542.
38. Eslami-Farouji, A.,Khodayari, H., Assadi, M. et al. (2021). Phylogeny and biogeography of the genus Hesperis (Brassicaceae, tribe, Hesperideae) inferred from nuclear ribosomal DNA sequence data. Plant Syst Evol. 307, 17-35.

39. Leigh Willard, Anuj Ranjan, Haiyan Zhang, Hassan Monzavi, Robert F. Boyko, Brian D. Sykes, and David S. Wishart "VADAR: a web server for quantitative evaluation of protein structure quality" *Nucleic Acids Res.* 2003 July 1; 31 (13): 3316–3319.