

Review Paper

## IN SILICO STRUCTURAL AND FUNCTIONAL ANALYSIS OF WRKY1 AND WRKY3 GENES IN THE SELECTED CEREAL CROPS

M. T. Pervez<sup>\*1</sup>, M. E. Babar<sup>2</sup>, J. Iqbal<sup>1</sup>, M. J. U. Hasnain<sup>1</sup>, S. H. Abbas<sup>1</sup>, S. S. M. Shah<sup>3</sup> and M. Ashraf<sup>4</sup>

<sup>1</sup>Department of Bioinformatics and Computational Biology, Virtual University of Pakistan

<sup>2</sup>Department of Genetics, Virtual University of Pakistan

<sup>3</sup>Department of Computer Science, Virtual University of Pakistan

<sup>4</sup>Department of Biology, Virtual University of Pakistan

\*Corresponding author's email: m.tariq@vu.edu.pk

### ABSTRACT

The WRKY family is one of the largest transcription factor families in higher plants and plays an important role in the growth of plants, their development, and response to biotic and abiotic stresses. They have capability to regulate the expression of other genes that enhance tolerance of plants to stress. Several studies have reported genome-wide identification of the WRKY transcription factors in individual crops however, only limited information is available on structural, functional and phylogenetic relationships among WRKY genes that are found in various crops. The proposed study investigated two WRKY genes (WRKY1 and WRKY3) in seven cereal crops namely: Rice (*Oryza sativa Japonica*), Rice (*Oryza sativa Indica*), Barley (*Hordeum vulgare*), Sorghum (*Sorghum bicolor*), Millet (*Setaria italica*), Wheat (*Triticum aestivum*), and Corn or Maize (*Zea mays*). The required DNA and protein sequences were downloaded from NCBI. The eight types of in-silico investigations were performed using various bioinformatics tools. Multiple sequence alignment was generated using MUSCLE embedded in IVIS-TMSA, percentage of identity and similarity was calculated using E-SICT, physicochemical property analysis was performed through ProtParam, phylogenetic analysis was carried out using MEGA, MEME was used for motif analysis, InterProScan was used for domain analysis, Gene Structure Display Server 2.0 was used for gene structure analysis. GORIV and Swiss homology modelling server were used for protein 2D/3D structural analysis revealed high degree of divergence among all WRKY proteins. However, SiWRKY1, a WRKY discovered in Millet (*Setaria italica*) and ZmWRKY1, and a WRKY found in Maize (*Zea mays*) showed the highest closeness. Based on the results reported by various computational tools, SiWRKY1 and ZmWRKY1 displayed the closest relationship with similar structure and function. Secondly the findings of the proposed study indicated that the proteins having similar 3D structure will also have similar properties. Thus, the proposed study will be very useful for investigating further aspects of the WRKY genes found in various crops and plants.

**Keywords:** WRKY transcription factors, In silico study of WRKY genes, Cereal crops, Phylogenetic analysis, Gene structure, Physicochemical properties.

<https://doi.org/10.36899/JAPS.2021.2.0220>

Published online October 03, 2020

### INTRODUCTION

The WRKY transcription factor genes play an important role in the growth of plants, development and response to biotic and abiotic stresses. They also regulate the expression of other genes to boost the stress tolerance of plants (Yue *et al.*, 2016). Due to global climatic changes, abiotic stresses such as drought, freezing, heat and biotic stresses (that happen due to damage by other living organisms such as viruses, bacteria, fungi, insects and parasites) cause a significant decrease in agricultural production around the world. Several studies have discovered that abiotic stresses on plants have induced a number of genes. These genes may be termed as stress-responsive (SR) genes which have been grouped into functional proteins and regulatory factors (Yamaguchi-Shinozaki and Shinozaki 2001). Transcription factors which are considered as a dominant class of regulatory

factors for controlling the expression of other genes, have a significant role in signal transduction and gene expression regulation.

WRKY transcription factors comprise of highly conserved sixty amino acid DNA-binding regions, called WRKY domain, which contains the conserved peptide sequence WRKYGQK at the N-terminal end and a zinc-finger motif at C-terminal end (Eulgem *et al.*, 2000; Phukan, Jeena, and Shukla 2016). Based on the number of WRKY domains and structural features of zinc-finger motif (Eulgem *et al.*, 2000; Phukan *et al.*, 2016), WRKY gene family can be divided into three groups designated as I to III. Group I has two WRKY domains and a zinc-finger motif whereas group II and group III contain only one WRKY domain as well as a CX<sub>4-5</sub>CX<sub>22-23</sub>HXH zinc-finger motif and a CX<sub>7</sub>CX<sub>23</sub>HXC zinc-finger motif respectively. In addition to this, group II is classified into five subgroups designated as IIa, IIb, IIc,

IId, and IIe (Eulgem *et al.*, 2000). A phylogenetic tree can classify WRKY gene family into four clades namely group I + IIc, group IIa + IIb, group IId and group IIe (Zhang and Wang 2005). According to phylogenetic analysis, WRKY groups II and III have been declared as descendants and they have originated from the C-terminal of WRKY domain of group I. However, a study has reported (Zhu *et al.*, 2013) that group IIc was originated from the N-terminal of WRKY domain of group I. The WRKY gene family is considered to be one of the most important transcription factor families in plants. They can play role in various developmental activities, physiological changes and response to biotic and abiotic stresses by binding with cis-acting elements. They contribute in seed dormancy & germination, seed development, flowering, senescence, metabolic pathways and plant growth ( Sun *et al.* 2003; Luo *et al.* 2005; Zentella *et al.* 2007; Zou, Neuman, and Shen 2008; Song *et al.* 2018). Furthermore, group III of WRKY gene family may show response to nematodes, pathogens and herbivores (Bi *et al.*, 2016; Grunewald *et al.*, 2008).

Until now, several studies (Hane *et al.*, 2017; De Vega *et al.*, 2015; Wang *et al.*, 2017) have reported identification of WRKY genes and their evolutionary relationships based on the genomic data. However, to the best of our knowledge, little is known about the structural and functional properties of the WRKY genes commonly found in various cereal crops. The proposed study investigated WRKY1 and WRKY3 in seven different cereal crops namely *Oryza sativa Japonica* (Rice), *Oryza sativa Indica* (Rice), *Hordeum vulgare* (Barley), *Sorghum bicolor* (Sorghum), *Setaria italic* (Millet), *Triticum aestivum* (Wheat) and *Zea mays* (maize). A number of bioinformatics tools were used to analyze the phylogenetic relationships as well as the functional and structural features among the WRKY genes of the selected cereal crops. The results showed that, overall, there is great divergence among the selected WRKY genes. However, SiWRKY1 and ZmWRKY1 were ranked at the top due to their close relationship. Second finding of the proposed study was that the proteins with similar 3D structure will also have similar properties (as happened in case of SiWRKY1 and ZmWRKY1). The proposed study, by using different bioinformatics tools, contributed that (1) most of the WRKY genes were highly divergent and (2) some WRKY genes found in different crops may have similar structure and function (as were SiWRKY1 and ZmWRKY1). This study will be very helpful for further investigations of these WRKY genes in the cereal crops in particular and other plant species in general.

**WRKY1:** WRKY1 has various regulatory functions to overcome the biotic and abiotic stresses. The 3D structure of C-terminal of WRKY1 highlighted that this domain

contains globular structure having  $\beta$  strands which results in an antiparallel  $\beta$ -sheet. At one end of the  $\beta$ -sheet, a zinc-binding site is located between strands  $\beta$ 4 and  $\beta$ 5. This arrangement shows that the DNA binding sites of WRKY1 are situated at  $\beta$ 2 and  $\beta$ 3 strands (Duan *et al.*, 2007). Some studies showed that organic-N treatment may suppress WRKY1 expression while N starvation can induce this (Bunning and Moser 1973; Krapp *et al.* 2011). It is also predicted that WRKY1 is a main hub in an organic-N regulatory network and has a role in repressing the expression of ASN1 transcriptionally. ASN1 is a gene usually expressed in reaction to nitrogen, carbon signaling and light (Thum *et al.*, 2003). OsWRKY1 shows spatial expression in leaves and young roots and is induced due to osmotic stress, IAA, SA and pathogens. It controls drought and salt stress in rice. It is also used for designing primers for the detection of SNPs. HvWRKY1 has an important role in a number of regulatory processes; controls cold and drought stresses with a regulatory role in abiotic stress in barley (Mangelsen *et al.* 2008). SbWRKY1 regulates the factors for drought and salt stress and also plays a role in stress due to phosphorus deficiency. SiWRKY1 transcription factor induces growth, shape, size, and development of height of millet. TaWRKY1 contributes in salt and drought stress in wheat. Induction of this gene is dependent on high temperatures.

**WRKY3:** WRKY3 transcription factors, identified by Parsley in 1996, has significant role in the regulation of the defense mechanism of plants. It belongs to group I WRKY factors having two Cys<sub>2</sub>His<sub>2</sub> motifs and is nuclear-localized sequence-specific DNA-binding protein. Any type of pathogenic infection stimulates the expression of WRKY3 genes which help to protect the plants from necrotrophic pathogens (Lai *et al.*, 2008). These transcription factors regulate NPR1 (a defense related gene) by signaling JA or SA functions and help in controlling abiotic stresses in plants. Cold and drought may induce OsWRKY3. OsWRKY3 may show higher resistance to Xoo (*Xanthomonas oryzae* pv. *Oryzae*) which functions upstream of OsNH1 (Liu *et al.*, 2005; X. Liu *et al.*, 2007). A number of defense signaling molecules may up-regulate OsWRKY3. The expression of OsWRKY3 is, however, light dependent (Liu *et al.*, 2005). HvWRKY3 is to AtWRKY18, AtWRKY40, and AtWRKY60 and shows an enhancement in expression after twenty hours of pathogen treatment. HvWRKY3 has gained popularity as a pathogen regulator (Mangelsen *et al.*, 2008). In millet the WRKY family is a major stress controlling transcription factors. SiWRKY3 control various abiotic stresses and is highly expressed in all tissues of *Setaria italica*. TaWRKY3 shows the defense response against drastic abiotic conditions and controls development processes.

## MATERIALS AND METHODS

**Data Retrieval:** All DNA and protein sequences of WRKY transcription factors were downloaded from NCBI. Table 1 shows detail of all datasets.

**Sequence and Phylogenetic Analysis:** The MUSCLE (Edgar 2004) tool embedded in IVisTMSA (Pervez *et al.* 2015) was used for the construction of Multiple Sequence Alignment (MSA). The alignment was visualized and analyzed by Jalview (Waterhouse *et al.*, 2009). MSA is one of the critical approaches in a number of research domains of life sciences such as generation of evolutionary trees, in silico drug design, protein modeling, highlighting conserved regions, and predicting gene function (Pervez *et al.*, 2014<sup>a</sup>). E-SICT (Pervez *et al.* 2014<sup>b</sup>) was used for calculating percentage of identity and similarity, which are very useful approaches for characterizing newly generated DNA/ protein sequences, of the selected WRKY protein. Maximum Likelihood Method was used to generate phylogenetic tree. An evolutionary tree provides an efficient way to organize information of biodiversity and permits development of an accurate, non-progressive conception of the totality of evolutionary history. MEGA 10.0.5 (Kumar *et al.*, 2018) with 1000 bootstrap replications was used to infer the phylogenetic tree. MEME (Multiple Expectation Maximization) Suite (Bailey *et al.* 2015) was used to identify the distribution of motifs in the fourteen WRKY protein sequences. Maximum number of motifs was set to 10 with any number of repetitions. Biological sequence motifs are the short, mostly fixed length, sequence patterns that usually represent significant functional or structural characteristics of DNA/ protein sequences such as active transcription binding sites, , interaction interfaces or splice junctions,. The motifs occur in an exact or approximate form within a family or a subfamily of sequences (Mohamed *et al.*, 2016). Domain analysis was performed using InterProScan (Jones *et al.*, 2014). Protein domains are distinct units, often linked with hinge region, which exist independently and define a function. Some proteins have single domain and some may have several domains. They also fold independently of the rest of the protein.

**Gene Structure Analysis:** Gene Structure Display Server (GSDS) version 2.0 (Hu *et al.* 2015) was used to determine the composition and positions of introns, exons and conserved elements. Gene structure analysis provides

a useful way for investigating coordinates of introns, exons and other annotated features such as binding sites and conserved elements in the selected gene.

**Physiochemical Properties:** The physiochemical properties (Length, molecular weight, theoretical pI, extinction coefficient absorbance, instability index, aliphatic index, grand average of hydropathicity (GRAVY)) of each of the fourteen WRKY proteins were estimated through ProtParam, a characterization tool available on the Expert Protein Analysis System (ExPasy) server (Cash 1999). The solubility and sub-cellular localization of this protein was predicted by SOSUI (Hirokawa, Boon-Chieng, and Mitaku 1998) and WoLF PSORT (Horton *et al.* 2007) software tools. Hydropathy-plot analysis was performed using the ProtScale software tool. The physiochemical properties are very essential to predict the function and structure of newly determined protein sequences.

**Structural characterization:** Secondary structure of protein represents local folded structures which appear due to interactions between atoms of the backbone. The most famous types of secondary structures are the  $\alpha$  helix and the  $\beta$  sheet. These two structures retain their shapes due to the hydrogen bonds which form between the carbonyl O of one amino acid and the amino H of another. The overall 3D structure of a protein molecule is called its tertiary structure. In this structure the protein molecule bends and twists in such a manner as to gain the maximum stability or the lowest energy state. There are a number of types of bonds and forces that help a protein to retain its 3D structure. They are (1) Hydrophobic interactions which has a huge contribution in maintaining the shape of a protein (2) Hydrogen bonding that occurs in the polypeptide chain and between "R" groups of amino acids and contribute in stabilizing protein structure. Ionic bonding occurs between the positively and negatively charged "R" groups while folding occurs due to the covalent bonding between the cysteine amino acids' "R" groups.

GORIV (Garnier *et al.*, 1996), a secondary structure prediction tool, was used to determine the secondary structure parameters (Alpha helix, extended strand and Random coil) of WRKY protein of the selected cereal crops. Swiss Homology Modelling server (Biasini *et al.* 2014) was used to predict the 3D structures of the WRKY proteins. The degree of nativeness of the predicted structure was estimated by QMean Score.

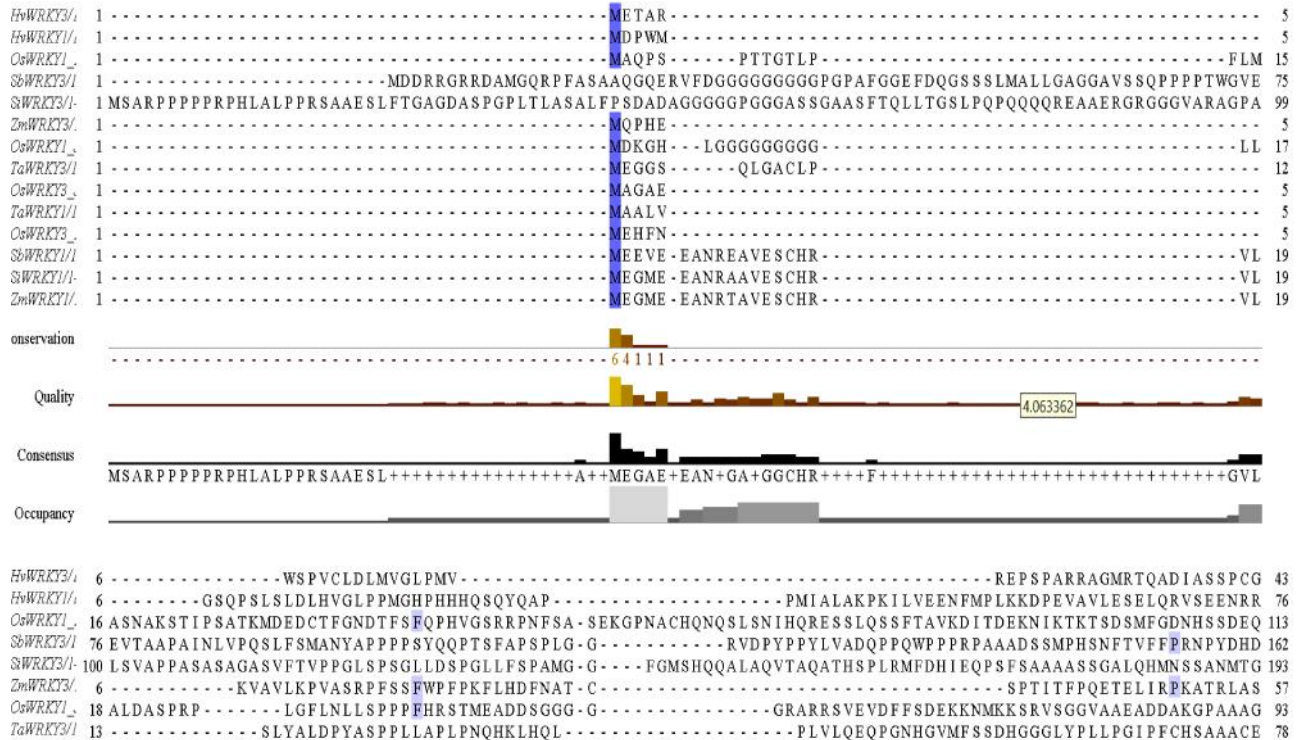
**Table 1. The selected WRKY data, their protein/DNA accession numbers and the database from where the data was downloaded.**

Species	WRKYs	Protein		DNA	
		Accession ID	Database	Accession ID	Database
Rice ( <i>Oryza sativa Japonica</i> )	OsWRKY1_J	DAA05066.1	NCBI	BK005004.1	NCBI
	OsWRKY3_J	DAA05068.1	NCBI	BK005006.1	NCBI
Rice ( <i>Oryza sativa Indica</i> )	OsWRKY1_I	AAN16970.1	NCBI	AF459793.1	NCBI
	OsWRKY3_I	AAT84155.1	NCBI	AY676924.1	NCBI
Barley ( <i>Hordeum vulgare</i> )	HvWRKY1	CAD60651.1	NCBI	AJ536667.1	NCBI
	HvWRKY3	CAH68819.1	NCBI	AJ853839.1	NCBI
Sorghum ( <i>Sorghum bicolor</i> )	SbWRKY1	ARO86289.1	NCBI	KY231904.1	NCBI
	SbWRKY3	Sb01g007480	GRASSIUS	XM_002442153.2	NCBI
Millet ( <i>Setaria italica</i> )	SiWRKY1	XP_004981825.1	NCBI	XM_004981768.4	NCBI
	SiWRKY3	XP_004962924.1	NCBI	XM_004962867.4	NCBI
Wheat ( <i>Triticum aestivum</i> )	TaWRKY1	ALR88710.1	NCBI	KT285206.1	NCBI
	TaWRKY3	ACD80364.1	NCBI	EU665432.1	NCBI
Corn or maize ( <i>Zea mays</i> )	ZmWRKY1	XP_008649271.1	NCBI	XM_008651049.2	NCBI
	ZmWRKY3	XP_008662747.1	NCBI	XM_008664525.3	NCBI

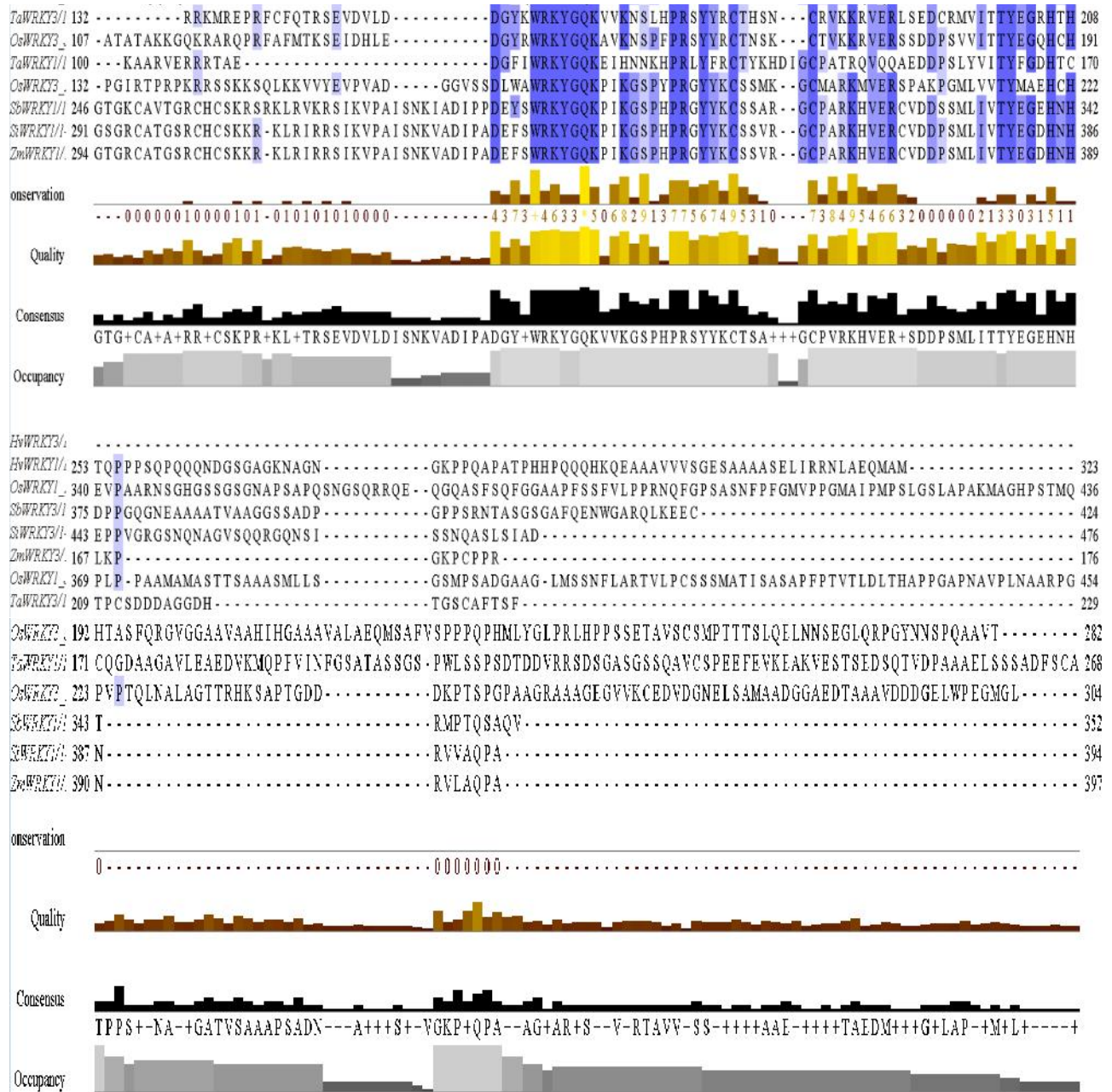
### RESULTS

**Multipl Sequence Alignment:** The MSA of the fourteen WRKY proteins was visualized and analyzed by Jalview. The highlighted regions showed  $\geq 50\%$  identity among the

sequences. Conserved peptide sequence WRKYGQK was found in all WRKY proteins except HvWRKY3. The alignment showed high variability among all WRKY proteins of the selected cereal crops (Fig. 1).







**Fig. 1.** The sequence alignment view of the fourteen WRKY proteins generated through Jalview. The highlighted regions showed  $\geq 50\%$  identity among the WRKY protein sequences.

**Percentage of Identity and Similarity:** Identity and similarity matrices showed significant differences among some WRKY protein sequences (Fig. 2). For example, similarity matrix (SM) showed that OsWRKY1 (Japonica) and OsWRKY1 (Indica) has less than 50% similarity with all other WRKY protein sequences. Furthermore, two groups of WRKY genes i.e. HvWRKY3, ZmWRKY3, TaWRKY3 and SiWRKY1,

SbWRKY1, ZmWRKY1 have been identified with more than 70% similarity. Identity matrix (IM) of the WRKY protein sequences showed that SiWRKY3, SbWRKY3, OsWRKY1 (Indica) have less than 50% identity with all other WRKY protein sequences. On the other hand, two groups i.e. HvWRKY3, ZmWRKY3, TaWRKY3 and SiWRKY1, ZmWRKY1 have shown more than 70% identity.

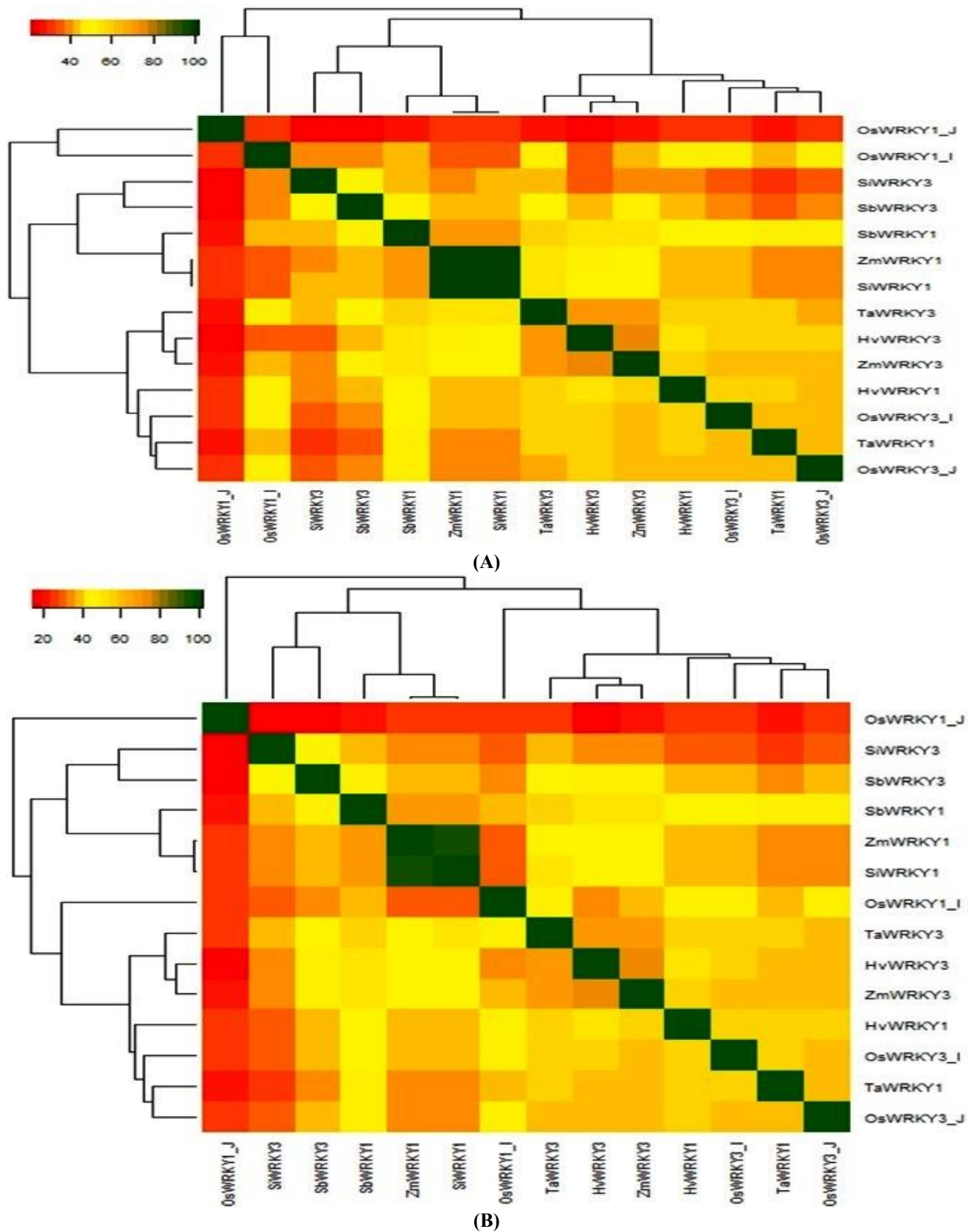
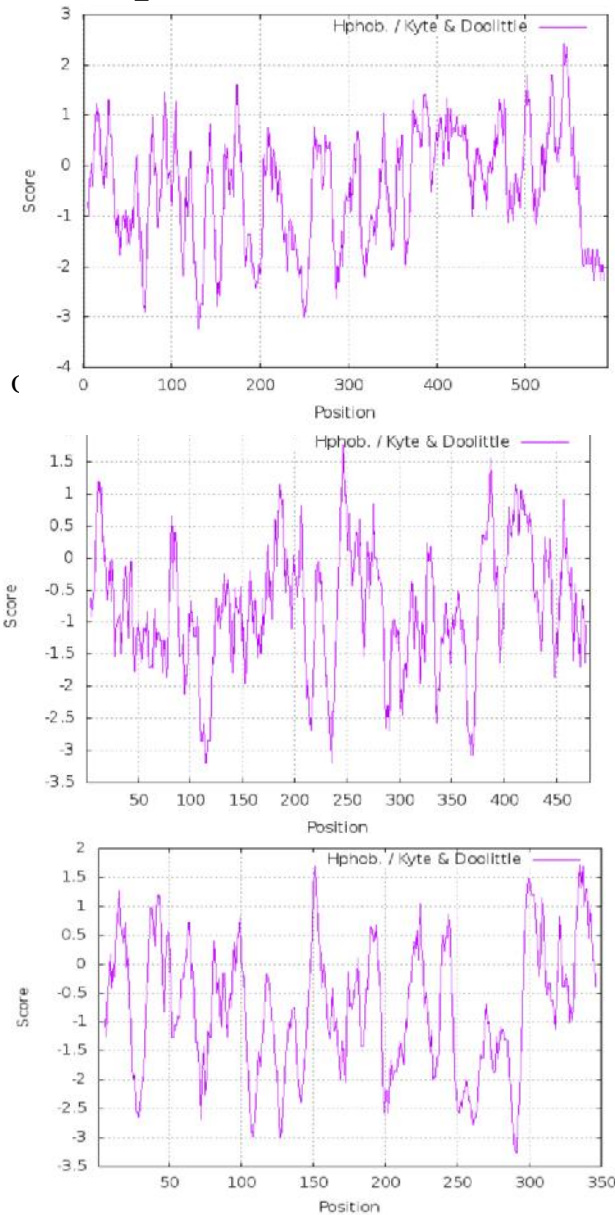


Figure 2. Heat maps of percentage of similarity (A) and identity (B) of all WRKY proteins of the selected cereal crops.

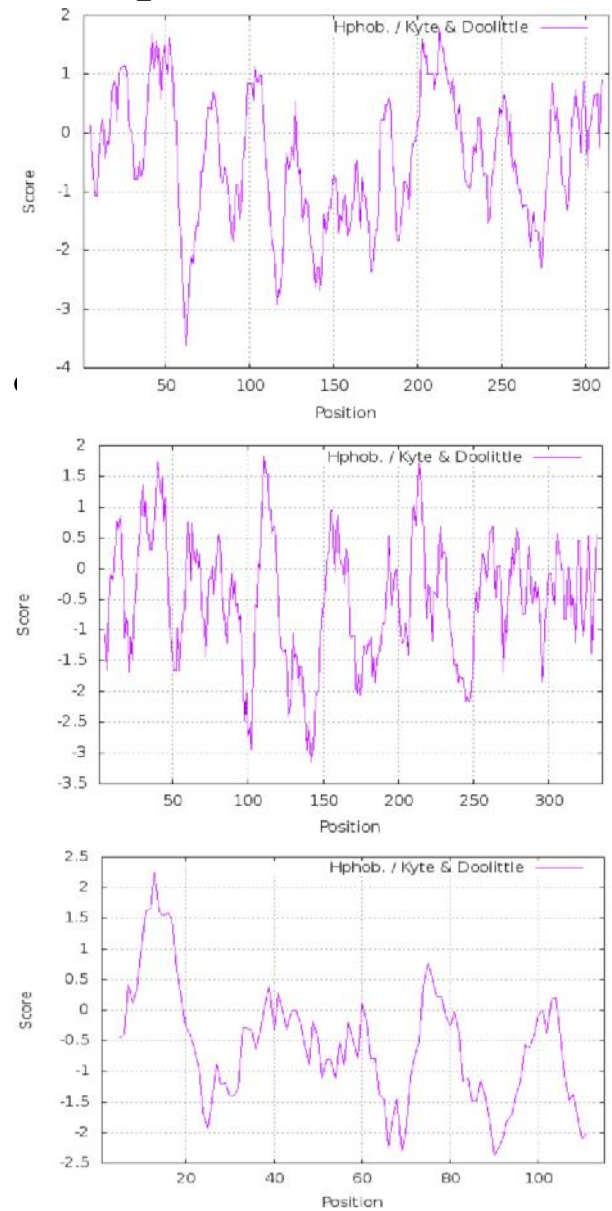
**Physiochemical Analysis:** The results showed variations in all parameters of this part of the proposed study (Table 2). The theoretical PI of HvWRKY1, SbWRKY1, SiWRKY1, SiWRKY3, TaWRKY3, ZmWRKY1 and ZmWRKY3 were more than 7 which indicated that these are basic proteins. All other WRKYs have less than 7 theoretical PI which demonstrated that these are acidic proteins. The instability index of all WRKYs was above 40 which showed that All WRKY proteins of the selected cereal crops were unstable (Guruprasad *et al.*, 1990). The aliphatic index which shows thermal stability of a protein was found above 50 of all WRKY proteins except those

of HvWRKY3 and OsWRKY1\_I. This highlighted that more than 50% of the volume of these WRKY proteins was occupied by aliphatic side amino acids (Thakkar and Saraf 2017). However, being less than zero, GRAVY values of these WRKY proteins showed that these were more hydrophilic in nature (Kyte and Doolittle 1982). The WRKY proteins were found to be intra nuclear as well as soluble in all selected cereal crops. The hydropath plot that was drawn for WRKY genes of six cereal crops (Fig. 3) highlighted that they were more inclined towards hydrophilicity (Fig. 3).

OsWRKY1\_J



OsWRKY3\_J



**Figure 3.** Hydrophobicity plot of WRKY1 and WRKY3 genes in rice (*Oryza sativa*: Janpnica and Indica) and Barley (*Hordeum vulgare*). All WRKY genes were more inclined towards hydrophilicity.

**Table 2. Physiochemical properties of WRKY1 and WRKY3 in seven cereal crops.**

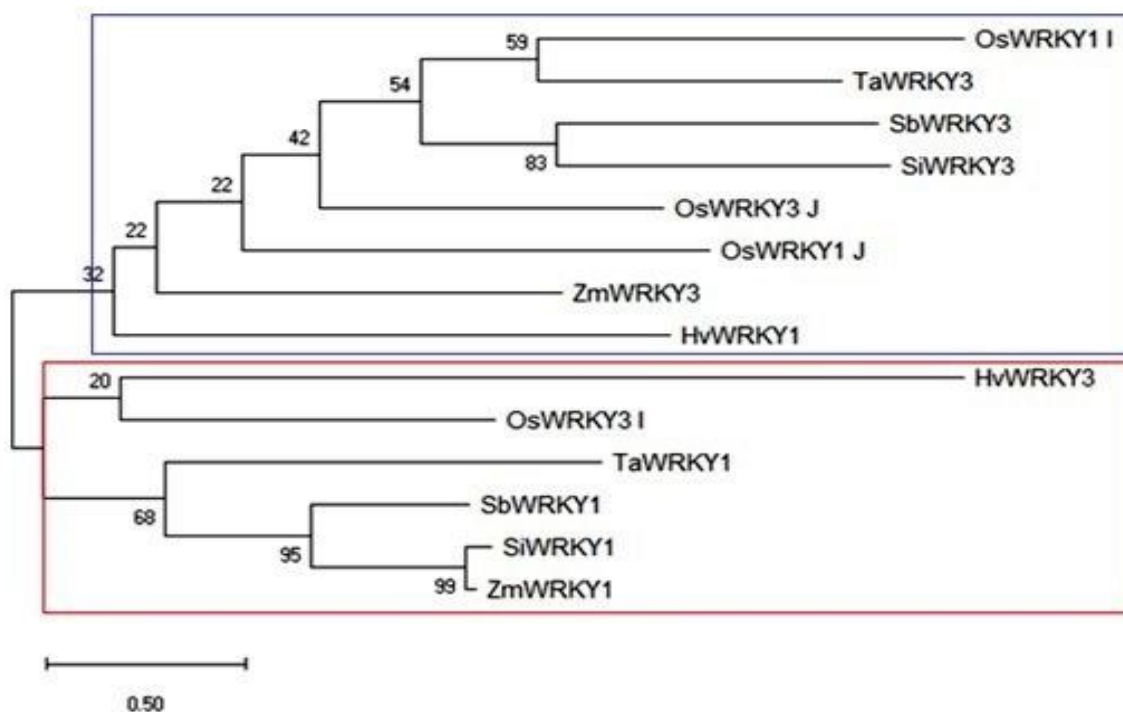
Species	AA Length	Molecular Weight	Th* PI	Ext. Coefficient	Abs*	Ins* Index	Aliphatic Index	GRAVY	Solubility	Localization
HvWRKY1	353	38585.32	8.14	26275	25900	60.93	63.34	-0.764	Soluble	Nuclear
HvWRKY3	115	12481.07	7.8	8605	8480	84.57	48.52	-0.657	Soluble	Nuclear
OsWRKY1_I	482	51929	5.68	20440	19940	62.41	40.61	-0.758	Soluble	Nuclear
OsWRKY1_J	593	60915.83	6.81	20315	19940	51.65	60	-0.459	Soluble	Nuclear
OsWRKY3_I	336	35467.78	5.72	34170	33920	53.29	62.5	-0.502	Soluble	Nuclear
OsWRKY3_J	314	32733.42	5.66	28670	28420	69.16	68.83	-0.416	Soluble	Nuclear
SbWRKY1	352	37899.77	9.78	16430	15930	49.15	63.98	-0.58	Soluble	Nuclear
SbWRKY3	424	45195.02	6.76	50225	49850	58.5	51.24	-0.777	Soluble	Nuclear
SiWRKY1	394	42212.18	9.98	14940	14440	48.54	70.13	-0.462	Soluble	Nuclear
SiWRKY3	494	52239.48	7.24	30620	30370	55.85	56.21	-0.749	Soluble	Nuclear
TaWRKY1	303	32413.84	4.9	30075	29450	56.51	65.15	-0.428	Soluble	Nuclear
TaWRKY3	229	24963.15	8.79	29045	28420	46.89	54.59	-0.573	Soluble	Nuclear
ZmWRKY1	397	42797.86	9.98	14940	14440	56.89	70.3	-0.486	Soluble	Nuclear
ZmWRKY3	201	22412.32	8.7	21805	21430	45.16	52.94	-0.757	Soluble	Nuclear

\*Th PI: Theoretical PI; Abs: Absorbance; Ins Index: Instability Index

**Phylogenetic Analysis:** The phylogenetic tree built using maximum likelihood method with bootstrap test (1000 replications) through MEGA 10 can be divided in two major clades/ sub-trees as highlighted in figure 4. It is clearly shown that clade highlighted with orange color ZmWRKY1, SiWRKY1 and SbWRKY1 has common ancestor. ZmWRKY1 and SiWRKY1 lie in the same clade and are very close to each other. In the sub-tree highlighted with blue color SbWRKY3 and SiWRKY3 evolved from the common ancestor and are very close to each other. Similarly OsWRKY1I and TaWRKY3 converge at the same point and are hence close to each other. All other WRKY genes are at distance in both clades.

**Motif Analysis:** The motifs, which are termed as transcription factor binding sites in a nucleotide sequence, are useful in investigating the gene regular network ( Song and Peng 2019). Furthermore, occurrence of the motifs may indicate their potential functions as some of them have been described to perform as nuclear localization signals assisting in phosphorylation or may be providing calmodulin binding sites for leucine zippers (Pandey and Somssich 2009). P-and E- values are used to evaluate statistical significance of motif finding. The E-value shows the statistical significance of the motifs and their likewise occurrence. The P-value reveals an estimate of how well each occurrence matches the motif.

The MEME motif analysis results showed not a single motif common in all WRKY protein sequences. However, ‘motif-1’ and ‘motif-2’ were common in all WRKY proteins except HvWRKY3 (Fig. 5 and Table 2). HvWRKY3 had only one motif. Motif-3 was common in ZmWRKY1, SiWRKY1, SbWRKY1, TaWRKY1 and HvWRKY3. Motif-4 was found in ZmWRKY1, SiWRKY1 and SbWRKY3. Motif-5 was found in five WRKY proteins. Motif-6 was found in three WRKY amino acid sequences. Motif-7 was found in three WRKYs i.e. ZmWRKY1, SiWRKY1, SbWRKY1 and TaWRKY3. Motif-8 was found in following three WRKYs; ZmWRKY1, SiWRKY1 and SbWRKY1. Motif-9 was found in WRKY sequences of seven species namely ZmWRKY1, SiWRKY1, SbWRKY1, OsWRKY1\_J, HvWRKY1, TaWRKY1 and OsWRKY3\_I. Motif-10 was observed in WRKY proteins of two species i.e. ZmWRKY1 and SiWRKY1. Each of the ZmWRKY1 and SiWRKY1 had a total of 11 motifs (including one extract motif i.e. ‘Motif-8’) which showed their close resemblance. SiWRKY3 and SbWRKY3 had two ‘Motif-1’ and two ‘Motif-2’ and one ‘Motif-5’ i.e. they have equal number of motifs which highlighted their close relationship. As shown in figure 5 and table 3, there are four WRKY proteins having same motif twice (highlighted in Table 2). In table 3, empty cells show absence of motif in the respective cereal crop.



**Figure 4.** Phylogenetic tree built using maximum likelihood method with bootstrap test (1000 replications) through MEGA 10. The sub-tree highlighted with orange color shows that SiWRKY1 and ZmWRKY1 evolved from the same ancestor and are very close to each other. Similarly in the sub-tree highlighted with blue color, SbWRKY3 and SiWRKY3 converge at the same point and are very close to each other.

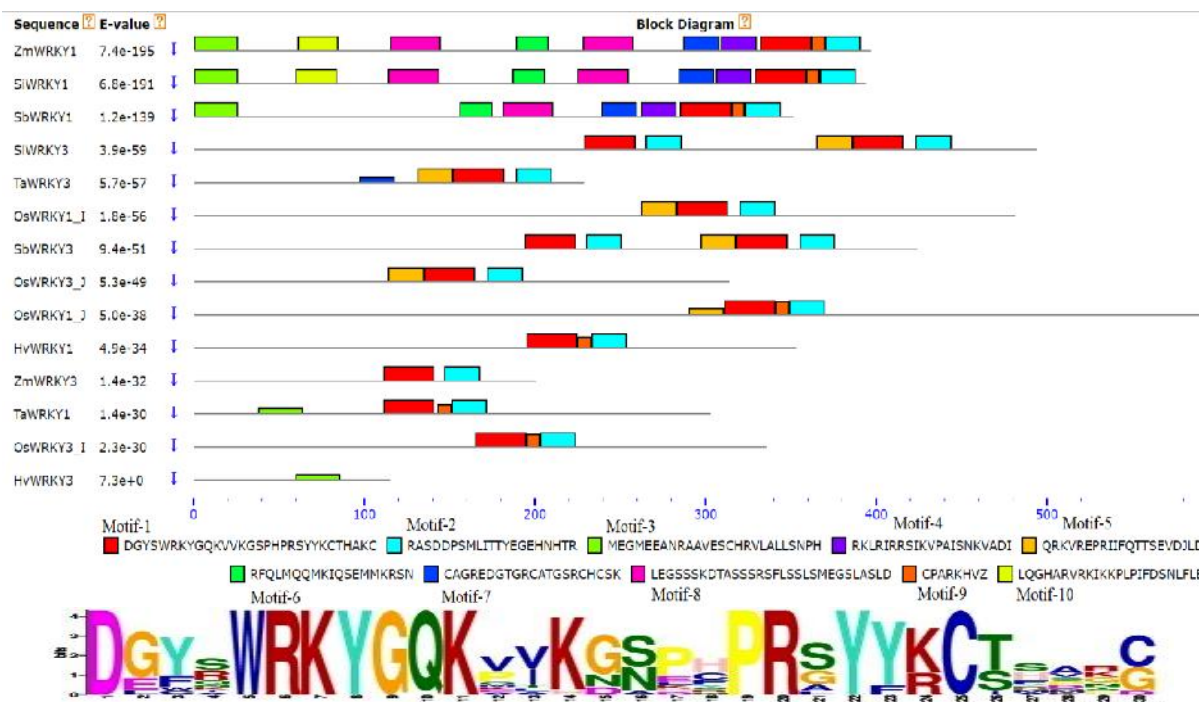
**Table 3. The selected crops and the motifs found along with their P-value, start and end positions.**

Species/ Motifs	Motif-1	Motif-2	Motif-3	Motif-4	Motif-5	Motif-6	Motif-7	Motif-8	Motif-9	Motif-10	
<b>ZmWRKY1</b>	<b>P-value:</b> 1.6e-31 <b>Start:</b> 333 <b>End:</b> 362	<b>P-value:</b> 9.2e-25 <b>Start:</b> 371 <b>End:</b> 391	<b>P-value:</b> 1.2e-33 <b>Start:</b> 1 <b>End:</b> 6	<b>P-value:</b> 3.6e-26 <b>Start:</b> 31 <b>End:</b> 330		<b>P-value:</b> 3.7e-26 <b>Start:</b> 19 <b>End:</b> 208	<b>P-value:</b> 2.7e-28 <b>Start:</b> 28 <b>End:</b> 308	<b>P-value:</b> 1.1e-23 <b>Start:</b> 11 <b>End:</b> 145	<b>P-value:</b> 3.4e-31 <b>Start:</b> 22 <b>End:</b> 258	<b>P-value:</b> 1.8e-11 <b>Start:</b> 36 <b>End:</b> 370	<b>P-value:</b> 4.3e-30 <b>Start:</b> 6 <b>End:</b> 85
<b>SiWRKY1</b>	<b>P-value:</b> 1.6e-31 <b>Start:</b> 330 <b>End:</b> 359	<b>P-value:</b> 9.2e-25 <b>Start:</b> 368 <b>End:</b> 338	<b>P-value:</b> 6.6e-32 <b>Start:</b> 1 <b>End:</b> 26	<b>P-value:</b> 3.6e-26 <b>Start:</b> 30 <b>End:</b> 327		<b>P-value:</b> 2.8e-25 <b>Start:</b> 18 <b>End:</b> 206	<b>P-value:</b> 1.7e-27 <b>Start:</b> 28 <b>End:</b> 303	<b>P-value:</b> 4.6e-25 <b>Start:</b> 11 <b>End:</b> 144	<b>P-value:</b> 1.9e-30 <b>Start:</b> 22 <b>End:</b> 255	<b>P-value:</b> 1.8e-11 <b>Start:</b> 36 <b>End:</b> 367	<b>P-value:</b> 2.3e-30 <b>Start:</b> 6 <b>End:</b> 84
<b>SbWRKY1</b>	<b>P-value:</b> 2.6e-32 <b>Start:</b> 286 <b>End:</b> 315	<b>P-value:</b> 1.9e-22 <b>Start:</b> 324 <b>End:</b> 344	<b>P-value:</b> 1.2e-29 <b>Start:</b> 1 <b>End:</b> 26	<b>P-value:</b> 1.1e-24 <b>Start:</b> 26 <b>End:</b> 283		<b>P-value:</b> 4.6e-18 <b>Start:</b> 15 <b>End:</b> 175	<b>P-value:</b> 6.2e-21 <b>Start:</b> 24 <b>End:</b> 260	<b>P-value:</b> 8.8e-24 <b>Start:</b> 182 <b>End:</b> 211		<b>P-value:</b> 1.8e-11 <b>Start:</b> 31 <b>End:</b> 323	
<b>SiWRKY3</b>	<b>P-value:</b> 3.1e-34 <b>Start:</b> 230 <b>End:</b> 259	<b>P-value:</b> 1.7e-37 <b>Start:</b> 38 <b>End:</b> 416	<b>P-value:</b> 5.8e-16 <b>Start:</b> 266 <b>End:</b> 286	<b>P-value:</b> 3.2e-19 <b>Start:</b> 424 <b>End:</b> 444		<b>P-value:</b> 2.2e-21 <b>Start:</b> 36 <b>End:</b> 386					
<b>TaWRKY3</b>	<b>P-value:</b> 4.2e-34 <b>Start:</b> 153 <b>End:</b> 182	<b>P-value:</b> 5.2e-19 <b>Start:</b> 190 <b>End:</b> 210				<b>P-value:</b> 6.2e-21 <b>Start:</b> 13 <b>End:</b> 152	<b>P-value:</b> 5.9e-5 <b>Start:</b> 98 <b>End:</b> 118				
<b>OsWRKY1_I</b>	<b>P-value:</b> 9.8e-36 <b>Start:</b> 284 <b>End:</b> 313	<b>P-value:</b> 2.2e-18 <b>Start:</b> 321 <b>End:</b> 341				<b>P-value:</b> 1.7e-20 <b>Start:</b> 26 <b>End:</b> 283					
<b>SbWRKY3</b>	<b>P-value:</b> 6.2e-32 <b>Start:</b> 31 <b>End:</b> 348	<b>P-value:</b> 8.6e-29 <b>Start:</b> 195 <b>End:</b> 224	<b>P-value:</b> 4.5e-17 <b>Start:</b> 35 <b>End:</b> 376	<b>P-value:</b> 1.4e-16 <b>Start:</b> 23 <b>End:</b> 251		<b>P-value:</b> 7.4e-21 <b>Start:</b> 29 <b>End:</b> 318					
<b>OsWRKY3_J</b>	<b>P-value:</b> 2.9e-32 <b>Start:</b> 136 <b>End:</b> 165	<b>P-value:</b> 5.2e-19 <b>Start:</b> 173 <b>End:</b> 193				<b>P-value:</b> 1.3e-15 <b>Start:</b> 11					

---

				5	
<b>OsWRKY1_</b>	<b>P-value:</b> 2.5e-29	<b>P-value:</b> 6.7e-20		<b>End:</b> 135	
<b>J</b>	<b>Start:</b> 312	<b>Start:</b> 350		<b>P-value:</b>	<b>P-value:</b>
	<b>End:</b> 341	<b>End:</b> 370		3.2e-5	4.3e-10
				<b>Start:</b> 29	<b>Start:</b> 34
				1	2
				<b>End:</b> 311	<b>End:</b> 349
<b>HvWRKY1</b>	<b>P-value:</b> 7.8e-29	<b>P-value:</b> 1.3e-17			<b>P-value:</b>
	<b>Start:</b> 196	<b>Start:</b> 234			1.0e-8
	<b>End:</b> 3225	<b>End:</b> 254			<b>Start:</b> 22
					6
					<b>End:</b> 233
<b>ZmWRKY3</b>	<b>P-value:</b> 8.0e-33	<b>P-value:</b> 3.8e-13			
	<b>Start:</b> 112	<b>Start:</b> 148			
	<b>End:</b> 141	<b>End:</b> 168			
<b>TaWRKY1</b>	<b>P-value:</b> 6.4e-28	<b>P-value:</b> 1.2e-14	<b>P-</b>		<b>P-value:</b>
	<b>Start:</b> 112	<b>Start:</b> 152	<b>value:</b>		1.1e-7
	<b>End:</b> 141	<b>End:</b> 172	9.2e-5		<b>Start:</b> 14
			<b>Start:</b> 3		4
			9		<b>End:</b> 151
			<b>End:</b> 64		
<b>OsWRKY3_</b>	<b>P-value:</b> 2.5e-29	<b>P-value:</b> 6.0e-13			<b>P-value:</b>
<b>I</b>	<b>Start:</b> 166	<b>Start:</b> 204			7.7e-10
	<b>End:</b> 195	<b>End:</b> 224			<b>Start:</b> 19
					6
					<b>End:</b> 203
<b>HvWRKY3</b>			<b>P-</b>		
			<b>value:</b>		
			7.2e-5		
			<b>Start:</b> 6		
			1		
			<b>End:</b> 86		

---



**Figure 5.** Motif scan analysis performed through MEME and MAST. The sequence of motifs is shown by The BLOCK diagram (A). Excluding HvWRKY3, two motifs i.e. ‘Motif-1’ and ‘Motif-2’ are common in all WRKY proteins. The sequential logo of the motif 1 showing consensus WRKY sequences present in WRKY proteins of the selected cereal crops except HvWRKY3 (B).

**Domain Analysis:** The InterProScan results showed that all selected WRKY proteins were members of the same homologous superfamily having accession id IPR036576. The SbWRKY3 and SiWRKY3 had four WRKY DNA binding domains spanned over different positions. Two domains were with SM00774 and other two had PF03106 accession id. Domains of SbWRKY3 occupied 193-251, 317-376, 194-250 and 378-375 positions and domains of

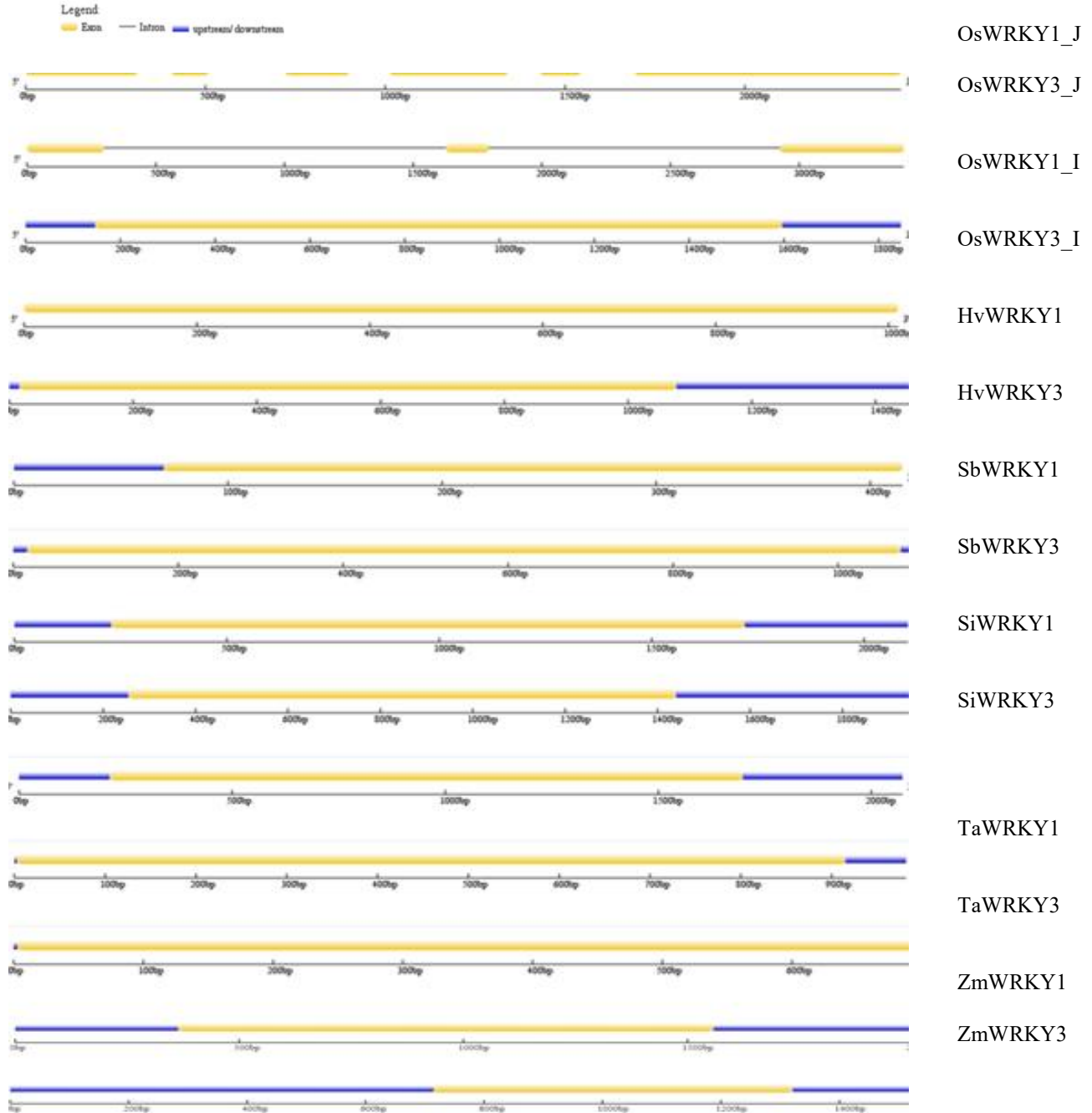
SiWRKY3 occupied 228-286, 385-444, 229-2285 and 386-443 positions. SbWRKY1, SiWRKY1 and ZmWRKY1, in addition to having two WRKY DNA binding domains had also Zn-cluster domain (Accession id: IPR018872). Table 4 shows number of WRKY DNA binding/ Zn-cluster domains of each WRKY protein, their positions and accession ids.

**Table 4.** The WRKY DNA binding / Zn-cluster domains of each WRKY protein, their positions and accession ids.

	Homologous Superfamily (IPR036576)	WRKY DNA Binding Domain-1 (SM00774)	WRKY DNA Binding Domain-2 (PF03106)	Zn-cluster Domain (IPR018872)
OsWRKY1_J	2995-372	310-370	312-369	
OsWRKY1_I	268-343	282-341	283-340	
OsWRKY3_J	118-195	134-193	135-191	
OsWRKY3_I	142-226	164-224	166-223	
HvWRKY1	182-256	194-254	195-252	
HvWRKY3				
SbWRKY3	188-253, 302-378	193-251, 317-376	194-250, 378-375	
SbWRKY1	267-346	284-344	286-342	236-282
SiWRKY1	312-389	328-388	330-387	282-336
SiWRKY3	213-288, 370-446	228-286, 385-444	229-285, 386-443	
TaWRKY1	95-173	110-172	111-170	
TaWRKY3	137-211	151-210	152-208	
ZmWRKY1	315-392	331-391	333-390	285-329
ZmWRKY3	93-170	110-168	112-166	

**Gene Structure Analysis:** The gene structure analysis revealed that OsWRKY1\_J and OsWRKY3\_J had exons and introns whereas all other WRKY genes had exon and upstream/ downstream regions. OsWRKY1\_J and OsWRKY3\_J had five introns, six exons and two long

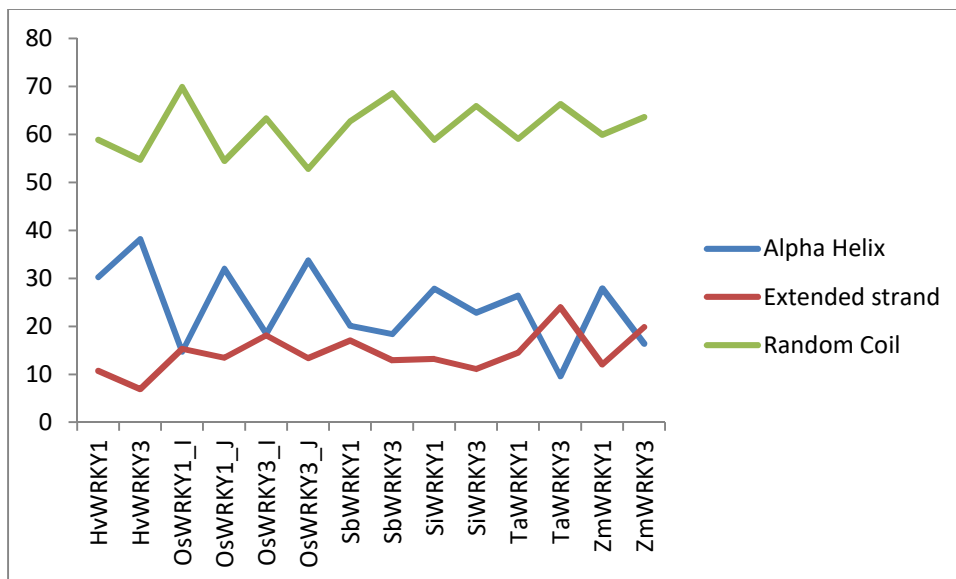
introns, three exons respectively. OsWRKY3\_I was the WRKY gene having only exonic region. Other WRKY DNA sequences had one long exonic region with upstream/ downstream regions as their terminal ends (Fig. 6).



**Figure 6.** Gene structures analysis highlighted that except OsWRKY1\_J, OsWRKY3\_J and OsWRKY3\_I, all other WRKY transcription factor genes had similar structure i.e. they had exonic and upstream/downstream regions only.

**Structural Analysis:** The structural analysis revealed that OSWRKY1\_I, OsWRKY3\_I proteins have almost equal values of alpha helix and extended strand which showed that stability of these transcription factors is of medium level. On the other hand, TaWRKY3 and ZmWRKY3 have low values of alpha helix than extended strand which means that these proteins are not stable. All other WRKY proteins have more percentage of alpha helix than extended strand which highlighted that these are stable proteins in nature (Fig. 7). The Homology Model Analysis (Fig. 8) showed that 3D structures of

only SiWRKY1 and ZmWRKY1 were almost similar. All other WRKY proteins had different 3D structures, although in first look many of them seemed to be similar. The models of all WRKY proteins were selected based on the highest identity between the target and template and the highest QMEAN (Qualitative Model Energy ANalysis) score, which is an approach for estimating absolute quality (degree of nativeness') of 3D protein structure. Based on the QMEAN Score and local quality estimates, the models of OsWRKY3\_J, OsWRKY1\_I SiWRKY3 may be declared as the most accurate.



**Figure 7. Secondary structure analysis of WRKY proteins in the selected cereal crops. Results showed that except four WRKY proteins (OSWRKY1\_I, OsWRKY3, TaWRKY3 and ZmWRKY3) all others were stable proteins.**

OsWRKY3\_J



OsWRKY1\_I



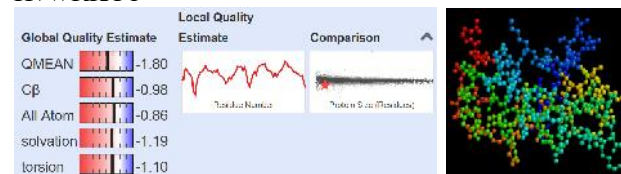
OsWRKY1\_I



OsWRKY3\_I

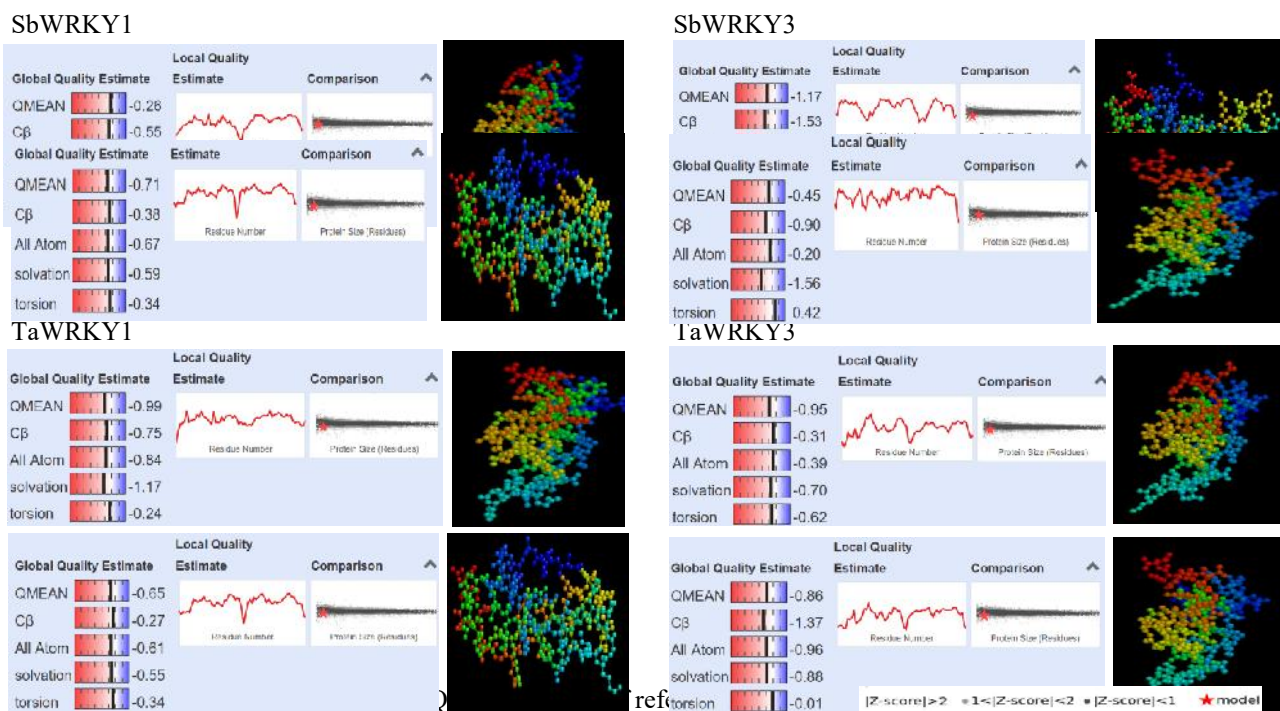


HvWRKY1



HvWRKY3





**Figure 8.** Homology Models of all WRKY proteins in the selected cereal crops, QMEAN Scores of reference set and accuracy estimation parameters of QMEAN.

## DISCUSSION

The WRKY genes play a significant role in regulating several processes in plants (Chen *et al.*, 2012; Yu *et al.*, 2010). The WRKY proteins, based on the WRKY domain numbers and zing-finger pattern/signature have been classified in three groups; group I, group II and group III (Singh *et al.*, 2019; Yu *et al.*, 2010). The WRKY proteins belonging to group I have two domains and a C2H2 zinc-finger signature. Members of group II comprise of a WRKY domain and zinc-finger (C2H2) sequence. Group II is divided into five subgroups called as a group IIa ... group IIe. The members of group III consist of a WRKY domain and C2H2 zinc-finger sequence (Jang *et al.*, 2010; Zhu *et al.*, 2013). A large number of studies have reported useful information on WRKY gene family in a number of species like rice, tomato, cotton, Arabidopsis, pineapple and strawberry (Chen and Liu, 2019). To the best of our knowledge little is investigated regarding the structural and functional analysis of WRKY genes found in various cereal crops.

The proposed study performed computational survey of WRKY1 and WRKY3 transcription factors in seven cereal crops. The selected cereal crops were *Oryza sativa Japonica*, *Oryza sativa Indica*, *Hordeum vulgare* (Barley), *Sorghum bicolor* (Sorghum), *Setaria italic* (Millet), *Triticum aestivum* (Wheat), and *Zea mays* (maize). Eight different computational analyses were performed on 14 WRKY datasets (2\*7). Multiple

sequence alignment showed WRKYGQK conserved domain in all WRKY proteins except that of HvWRKY3. The same conserved domain has been identified by various studies on WRKY transcription factors (Aamir *et al.*, 2017; Gao *et al.*, 2014; Jimmy and Babu, 2019). The reason behind this was that its complete coding sequence was not available. MSA showed big divergence among all WRKY protein sequences. Most of the similar studies also revealed significant divergence among under study WRKY proteins (Hane *et al.*, 2017; De Vega *et al.*, 2015; Wang *et al.*, 2017). Identity and similarity analysis revealed that SiWRKY1 and ZmWRKY1 had 94.13% identity and 96.59% similarity which confirmed results of several other studies which showed significant similarity between any two WRKY genes of cereal crops (Mangelsen *et al.*, 2008; NIU *et al.*, 2012). This was the highest percentage of identity/similarity found between any two WRKY proteins. Physiochemical properties analysis highlighted that being theoretical PI less than seven HvWRKY1, SbWRKY1, SiWRKY1, SiWRKY3, TaWRKY3, ZmWRKY1 and ZmWRKY3 were basic proteins and all other WRKY proteins were acidic in nature. Based on instability index (above 40) all WRKY proteins were reported as unstable. The aliphatic index analysis showed that except HvWRKY3 and OsWRKY1\_I, more than 50% volume of all WRKY proteins was covered by aliphatic side amino acids. GRAVY values showed that all WRKY proteins were hydrophilic in nature which was also confirmed by Hydropath Plot drawn for six WRKY proteins.

Physiochemical properties of SiWRKY1 and ZmWRKY1 were also similar i.e. these were basic, unstable, occupied by aliphatic acid amino acids, and were hydrophilic in nature. Phylogenetic analysis showed two major clades of WRKY proteins having common ancestor. However, the closest relationship was shown between SiWRKY1 and ZmWRKY1. Similar results were also reported by other studies on WRKY genes found in various cereal crops (Gao *et al.*, 2014; ; Jimmy and Babu, 2019; De Vega *et al.*, 2015; Hane *et al.*, 2017; Wang *et al.*, 2017; Wei *et al.*, 2012). The motif analysis showed that except HvWRKY3 all WRKY proteins had two common motifs i.e. 'motif-1' and 'motif-2'. This analysis also highlighted that SiWRKY1 and ZmWRKY1 had same number of motifs which showed their close resemblance. However, big variation regarding common motifs in all selected cereal crops was seen and the same was supported by Aamir *et al.*, (2017). Domain analysis revealed WRKYs of the selected cereal crops to be member of the same homologous superfamily. This part of our study also reported that SiWRKY1 and ZmWRKY1 have same number of domains which highlighted their closeness. According to the domain analysis all WRKY proteins maybe assigned group I. Gene structure analysis also highlighted the close relationship between SiWRKY1 and ZmWRKY1 by showing same composition of the two genes. The structural analysis revealed that only two WRKY proteins i.e. SiWRKY1 and ZmWRKY1 had very close structural properties/resemblance. Especially the 3D model prediction by Swiss model showed that except these two WRKY proteins all others had different 3D structure. By concluding our discussion we may claim that based on our eight types of analyses performed, SiWRKY1 and ZmWRKY1 have similar evolutionary history and have same structural and functional properties. Gene structure and protein structural variation/ similarity investigation was performed in various studies which supported results presented in the proposed study (Jimmy and Babu, 2019).

**Conclusion:** The proposed study analyzed two WRKY transcription factor genes (WRKY1 and WRKY3) in seven cereal crops namely *Oryza sativa Japonica*(rice), *Oryza sativa Indica*, *Hordeum vulgare* (Barley), *Sorghum bicolor* (Sorghum), *Setaria italic* (Millet), *Triticum aestivum* (Wheat), and *Zea mays* (maize) by employing several bioinformatics tools. MSA and percentage of identity and similarity analysis showed high degree of divergence among most of WRKY proteins. The physiochemical analysis revealed seven WRKY proteins as acidic and seven as basic. All WRKY proteins were shown to be unstable and hydrophilic in nature. Motif analysis showed two motifs found in common in all WRKY proteins except that of HvWRKY3. Domain analysis put all WRKY proteins in group I. All WRKY proteins were found to have two domains except that of

HvWRKY3, which did not have any domain. Structural properties computed by GOR IV as well as by Swiss model of all WRKY proteins except that of SiWRKY1 and ZmWRKY1 were different from each other. It has been observed during this study that SiWRKY1 and ZmWRKY1 have similar properties computed by all types of analyses. So, this can be claimed that these two WRKY transcription factor genes have similar structure and function..

**Competing Interests:** The authors declare that they have no competing interests.

**Funding:** Authors declare that they didn't receive any type of funds from any organization.

**Authors' contributions:** MTP conceived the idea. MTP and MEB designed the datasets. MJH, SSMS and SHA analyzed the datasets and compiled the results, MTP and MEB wrote the article. MA edited the article.

**Acknowledgements:** We are very thankful to Virtual University of Pakistan for its generous support to conduct this research work.

## REFERENCES

- Aamir, M., V.K. Singh, M. Meena, R.S. Upadhyay, V.K. Gupta, and S. Singh (2017). Structural and functional insights into WRKY3 and WRKY4 transcription factors to unravel the WRKY–DNA (W-Box) complex interaction in tomato (*Solanum lycopersicum* L.). *Frontiers in Plant Science*, (8): 819.
- Bailey, T. L., J. Johnson, C.E. Grant, and W.S. Noble (2015). The MEME suite. *Nucleic acids research*, 43(W1): W39-W49.
- Bi, C., Y. Xu, Q. Ye, T. Yin, and Ye, N. (2016). Genome-wide identification and characterization of WRKY gene family in *Salix suchowensis*. *PeerJ*, 4, e2437.
- Biasini, M., S. Bienert, A. Waterhouse, K. Arnold, G. Studer, T. Schmidt, F. Kiefer, T.G.Cassarino, M. Bertoni, L. Bordoli, and T. Schwed (2014). SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. *Nucleic acids research*, 42(W1): W252-W258.
- Bünning, E., and I. Moser (1973). Light-induced phase shifts of circadian leaf movements of phaseolus: Comparison with the effects of potassium and of ethyl alcohol. *Proceedings of the National Academy of Sciences*, 70(12): 3387-3389.
- Cash, P. (1999). "2-D Proteome Analysis Protocols." *Cell Biology International*, 23(5): 385.
- Chen, P., and Q.Z. Liu (2019). Genome-wide characterization of the WRKY gene family in

- cultivated strawberry (*Fragaria* × *ananassa* Duch.) and the importance of several group III members in continuous cropping. *Scientific reports*, 9(1):1-12.
- Chen, L., Y. Song, S. Li, L. Zhang, C. Zou, and D. Yu (2012). The role of WRKY transcription factors in plant abiotic stresses. *Biochimica et Biophysica Acta (BBA)-Gene Regulatory Mechanisms*, 1819(2):120-128.
- De Vega, J.J., S. Ayling, M. Hegarty, D. Kudrna, J.L. Goicoechea, A. Ergon, O.A. Rognli, C. Jones, M. Swain, R. Geurts, and C.Lang (2015). Red clover (*Trifolium pratense* L.) draft genome provides a platform for trait improvement. *Scientific reports*, (5): 17394.
- Duan, M.R., J. Nan, Y.H. Liang, P. Mao, L.Lu, L. Li, C. Wei, L. Lai, Y. Li, and X.D. Su (2007). DNA binding mechanism revealed by high resolution crystal structure of *Arabidopsis thaliana* WRKY1 protein. *Nucleic acids research*, 35(4), 1145-1154.
- Edgar, Robert C. 2004. "MUSCLE: Multiple Sequence Alignment with High Accuracy and High Throughput." *Nucleic Acids Research*, 32(5): 1792–97.
- Eulgem, T., P.J. Rushton, S. Robatzek, and I.E. Somssich (2000). The WRKY superfamily of plant transcription factors. *Trends in plant science*, 5(5):199-206.
- Gao, J., H. Peng, X. He, M. Luo, Z. Chen, H. Lin, H. Ding, G.Pan, and Z. Zhang (2014). Molecular phylogenetic characterization and analysis of the WRKY transcription factor family responsive to *Rhizoctonia solani* in maize. *Maydica*, 59(1): 32-41.
- Garnier, J., J.F. Gibrat, and B. Robson (1996). GOR method for predicting protein secondary structure from amino acid sequence. In *Methods in enzymology*, (266): 540-553.
- Grunewald, W., M. Karimi, K. Wiczorek, E. Van de Cappelle, E. Wischnitzki, F. Grundler, D. Inzé, T. Beeckman, and G. Gheysen (2008). A role for AtWRKY23 in feeding site establishment of plant-parasitic nematodes. *Plant physiology*, 148(1): 358-368.
- Guruprasad, K., B.B. Reddy, and M.W. Pandit (1990). Correlation between stability of a protein and its dipeptide composition: a novel approach for predicting in vivo stability of a protein from its primary sequence. *Protein Engineering, Design and Selection*, 4(2): 155-161.
- Hane, J.K., Y. Ming, L.G. Kamphuis, M.N. Nelson, G. Garg, C.A. Atkins, P.E. Bayer, A. Bravo, S. Bringans, S. Cannon, and D. Edwards (2017). A comprehensive draft genome sequence for lupin (*Lupinus angustifolius*), an emerging health food: insights into plant–microbe interactions and legume evolution. *Plant biotechnology journal*, 15(3): 318-330.
- Hirokawa, T., S. Boon-Chieng, and S. Mitaku (1998). SOSUI: classification and secondary structure prediction system for membrane proteins. *Bioinformatics (Oxford, England)*, 14(4): 378-379.
- Horton, P., K.J. Park, T. Obayashi, N. Fujita, H. Harada, C.J. Adams-Collier, and K. Nakai (2007). WoLF PSORT: protein localization predictor. *Nucleic acids research*, 35(suppl\_2): W585-W587.
- Hu, B., J. Jin, A.Y. Guo, H. Zhang, J. Luo, and G. Gao (2015). GSDS 2.0: an upgraded gene feature visualization server. *Bioinformatics*, 31(8): 1296-1297.
- Jang, J. Y., C.H. Choi, and D.J. Hwang (2010). The WRKY superfamily of rice transcription factors. *The Plant Pathology Journal*, 26(2): 110-114.
- Jimmy, J. L., and S. Babu, (2019). Variations in the Structure and Evolution of Rice WRKY Genes in Indica and Japonica Genotypes and their Co-expression Network in Mediating Disease Resistance. *Evolutionary Bioinformatics*, (15): 1176934319857720.
- Jones, P., D. Binns, H.Y. Chang, M. Fraser, W. Li, C. McAnulla, H. McWilliam, J. Maslen, A. Mitchell, G. Nuka, and S. Pesseat (2014). InterProScan 5: genome-scale protein function classification. *Bioinformatics*, 30(9): 1236-1240.
- Krapp, A., R. Berthomé, M. Orsel, S. Mercey-Boutet, A. Yu, L. Castangs, S. Elftieh, H. Major, J.P. Renou, and F. Daniel-Vedele (2011). *Arabidopsis* roots and shoots show distinct temporal adaptation patterns toward nitrogen starvation. *Plant Physiology*, 157(3): 1255-1282.
- Kumar, S., G. Stecher, M. Li, C. Knyaz, and K. Tamura (2018). MEGA X: molecular evolutionary genetics analysis across computing platforms. *Molecular biology and evolution*, 35(6): 1547-1549.
- Kyte, J., and R.F. Doolittle (1982). A simple method for displaying the hydropathic character of a protein. *Journal of molecular biology*, 157(1): 105-132.
- Lai, Z., K.M. Vinod, Z. Zheng, B. Fan, and Z. Chen (2008). Roles of *Arabidopsis* WRKY3 and WRKY4 transcription factors in plant responses to pathogens. *BMC plant biology*, 8(1): 68.
- Liu, X. Q., X.Q. Bai, Q.I.A.N. Qian, X.J. Wang, M.S. Chen, and C.C. Chu (2005). OsWRKY03, a rice transcriptional activator that functions in defense signaling pathway upstream of OsNPR1. *Cell research*, 15(8): 593-603.
- Liu, X., X. Bai, X. Wang, and C. Chu (2007). OsWRKY71, a rice transcription factor, is involved in rice defense response. *Journal of*

- plant physiology, 164(8): 969-979.
- Luo, M., E.S. Dennis, F. Berger, W.J. Peacock, and A. Chaudhury (2005). MINISEED3 (MINI3), a WRKY family gene, and HAIKU2 (IKU2), a leucine-rich repeat (LRR) KINASE gene, are regulators of seed size in Arabidopsis. Proceedings of the National Academy of Sciences, 102(48): 17531-17536.
- Mangelsen, E., J. Kilian, K. W. Berendzen, O.H. Kolukisaoglu, .K. Harter, C. Jansson, and D. Wanke (2008). Phylogenetic and comparative gene expression analysis of barley (*Hordeum vulgare*) WRKY transcription factor family reveals putatively retained functions between monocots and dicots. BMC genomics, 9(1): 194.
- Mohamed, S.A.E.H., Elloumi, M., and Thmpson, J.D. (2016). Motif Discovery in Protein Sequences. In Pattern Recognition-Analysis and Applications. IntechOpen.
- NIU, C. F., W.E.I.Wei, Q.Y. Zhou, A.G. Tian, Y.J. Hao, W.K. Zhang, B. Ma, Q. Lin, Z. B. Zhang, J.S. Zhang, AND S.Y. Chen. (2012). Wheat WRKY genes TaWRKY2 and TaWRKY19 regulate abiotic stress tolerance in transgenic Arabidopsis plants. Plant, cell & environment, 35(6): 1156-1170.
- Pandey, S. P., and I.E. Somssich (2009). The role of WRKY transcription factors in plant immunity. Plant physiology, 150(4): 1648-1655.
- Pervez, M.T., M.E. Babar, A. Nadeem, M. Aslam, N. Naveed, S. Ahmad, S. Muhammad, S. Qadri, M. Shahid, T. Hussain, and M. Javed (2015). IVisTMSA: interactive visual tools for multiple sequence alignments. Evolutionary Bioinformatics, 11, EBO-S18980.
- Pervez, M.T<sup>a</sup>, M.E. Babar, A. Nadeem, M. Aslam, A.R. Awan, N. Aslam, T. Hussain, N. Naveed, S. Qadri, U. Waheed, and M. Shoaib (2014). Evaluating the accuracy and efficiency of multiple sequence alignment methods. Evolutionary bioinformatics online,(10): 205–217.
- Pervez, M.T<sup>b</sup>, M.E. Babar, A. Nadeem, N. Aslam, and T. Hussain (2014). E-SICT: An Efficient Similarity and Identity Matrix Calculating Tool. J. Appl. Environ. Biol. Sci, 4(9S):458-462.
- Phukan, U. J., G.S. Jeena, and R.K. Shukla (2016). WRKY transcription factors: molecular regulation and stress responses in plants. Frontiers in plant science,(7): 760.
- Singh, A., P.K. Singh, A.K. Sharma, N.K. Singh, H. Sonah, R. Deshmukh, and T.R. Sharma (2019). Understanding the Role of the WRKY Gene Family under Stress Conditions in Pigeonpea (*Cajanus cajan* L.). Plants, 8(7): 214.
- Song, H., W. Sun, G. Yang, and J. Sun (2018). WRKY transcription factors in legumes. BMC plant biology, 18(1): 243.
- Song, M., and X. Peng (2019). Genome-wide identification and characterization of DIR genes in *Medicago truncatula*. Biochemical genetics, 57(4): 487-506.
- Sun, C., S. Palmqvist, H. Olsson, M. Borén, S. Ahlandsberg, and Jansson (2003). A novel WRKY transcription factor, SUSIBA2, participates in sugar signaling in barley by binding to the sugar-responsive elements of the iso1 promoter. The Plant Cell, 15(9): 2076-2092.
- Thakkar, A., and M. Saraf (2017). Molecular cloning and in silico analysis of cellulase gene from *Bacillus amyloliquefaciens* MBAA3.
- Thum, K. E., D. E. Shasha, L. V. Lejay, and G. M. Coruzzi (2003). Light- and carbon-signaling pathways. Modeling circuits of interactions. Plant physiology, 132(2): 440-452.
- Wang, J., P. Sun, Y. Li, Y. Liu, J. Yu, X. Ma, S. Sun, N. Yang, R. Xia, T. Lei, and X. Liu (2017). Hierarchically aligning 10 legume genomes establishes a family-level genomics platform. Plant Physiology, 174(1): 284-300.
- Waterhouse, A. M., J.B. Procter, D.M. Martin, M. Clamp, and G.J. Barton (2009). Jalview Version 2—a multiple sequence alignment editor and analysis workbench. Bioinformatics, 25(9): 1189-1191.
- Wei, K. F., J. Chen, Y.F. Chen, L.J. Wu, and D.X. Xie (2012). Molecular phylogenetic and expression analysis of the complete WRKY transcription factor family in maize. DNA research, 19(2): 153-164.
- Yamaguchi-Shinozaki, K., and K. Shinozaki (2001). Improving Plant Drought, Salt and Freezing Tolerance by Gene Transfer of a Single Stress-Inducible Transcription Factor. Novartis Foundation Symposium, (236): 176–89.
- Yue, H., M. Wang, S. Liu, X. Du, W. Song, and X. Nie (2016.) Transcriptome-Wide Identification and Expression Profiles of the WRKY Transcription Factor Family in Broomcorn Millet (*Panicum Miliaceum* L.). BMC Genomics, 17(1): 1–11.
- Yu, S. O. N. G., C.R. AI, S.J.Jing, and D.Q. YU (2010). Research progress on functional analysis of rice WRKY genes. Rice science, 17(1): 60-72.
- Zentella, R., Z.L. Zhang, M. Park, S.G.Thomas, A. Endo, K. Murase, C.M. Fleet, Y. Jikumaru, E. Nambara, Y. Kamiya, and S.T. Sun (2007). Global analysis of DELLA direct targets in early gibberellin signaling in Arabidopsis. The Plant Cell, 19(10): 3037-3057.
- Zhang, Y., and L. Wang, (2005). The WRKY transcription factor superfamily: its origin in eukaryotes and expansion in plants. BMC

- evolutionary biology, 5(1): 1.
- Zhang, Z. L., Z. Xie, X. Zou, J. Casaretto, T.H.D. Ho, and Q.J. Shen (2004). A rice WRKY gene encodes a transcriptional repressor of the gibberellin signaling pathway in aleurone cells. *Plant Physiology*, 134(4): 1500-1513.
- Zhu, X., S. Liu, C. Meng, L. Qin, L. Kong, and G. Xia (2013). WRKY transcription factors in wheat and their induction by biotic and abiotic stress. *Plant Molecular Biology Reporter*, 31(5): 1053-1067.
- Zou, X., D. Neuman, and Q.J. Shen (2008). Interactions of two transcriptional repressors and two transcriptional activators in modulating gibberellin signaling in aleurone cells. *Plant Physiology*, 148(1): 176-186.