Turkish Journal of Agriculture - Food Science and Technology, 4(4): 279-290, 2016

Turkish Journal of Agriculture - Food Science and Technology

www.agrifoodscience.com,
Turkish Science and Technology

Genome wide analysis of stress responsive *WRKY* transcription factors in *Arabidopsis thaliana*

Shaiq Sultan[1], Muhammad Amjad Ali[2,3]*, Rana Muhammad Atif[3,4], Farrukh Azeem[1], Habibullah Nadeem[1], M. Hussnain Siddique[1], Ertuğrul Filiz[5], Khadim Hussain[1], Amjad Abbas[2]

[1]*Department of Bioinformatics & Biotechnology, Government College University, 38000 Faisalabad, Pakistan*
[2]*Department of Plant Pathology, University of Agriculture, 38040 Faisalabad, Pakistan*
[3]*Centre of Agricultural Biochemistry & Biotechnology, University of Agriculture, 38040 Faisalabad, Pakistan*
[4]*Department of Plant Breeding & Genetics, University of Agriculture, 38040 Faisalabad, Pakistan*
[5]*Department of Crop and Animal Production, Cilimli Vocational School, Duzce University, 81750 Cilimli/Duzce, Turkey*

ARTICLEINFO

Article history:
Received 08 November 2015
Accepted 20 February 2016
Available online, ISSN: 2148-127X

Keywords:
WRKY
Arabidopsis thaliana
Stress
W-Box
Phylogenetic analysis
Transcription factor

*Corresponding Author:
E-mail: amjad.ali@uaf.edu.pk

ABSTRACT

WRKY transcription factors are a class of DNA-binding proteins that bind with a specific sequence C/TTGACT/C known as W-Box found in promoters of genes which are regulated by these WRKYs. From previous studies, 43 different stress responsive *WRKY* transcription factors in *Arabidopsis thaliana*, identified and then categorized in three groups viz., abiotic, biotic and both of these stresses. A comprehensive genome wide analysis including chromosomal localization, gene structure analysis, multiple sequence alignment, phylogenetic analysis and promoter analysis of these *WRKY* genes was carried out in this study to determine the functional homology in *Arabidopsis*. This analysis led to the classification of these *WRKY* family members into 3 major groups and subgroups and showed evolutionary relationship among these groups on the base of their functional *WRKY* domain, chromosomal localization and intron/exon structure. The proposed groups of these stress responsive *WRKY* genes and annotation based on their position on chromosomes can also be explored to determine their functional homology in other plant species in relation to different stresses. The result of the present study provides indispensable genomic information for the stress responsive *WRKY* transcription factors in *Arabidopsis* and will pave the way to explain the precise role of various AtWRKYs in plant growth and development under stressed conditions.

Introduction

Biotic stresses like pathogen and insect attack as well as abiotic stresses, like salinity, drought, osmotic stress and temperature variations limit crop productivity and influence the expression of many genes in plants (Xiong et al., 2002; Yamaguchi- Shinozaki and Shinozaki, 2006; Hirayama and Shinozaki, 2010). To cope with these stresses, plants adapt different morphological, physiological, biochemical and molecular strategies. Plants have a special ability to adapt in various stresses which depends on the molecular networks and expression of genes that protect and maintain the structure of cellular components. This could be achieved by the product of the genes directly involved in stress responses or the genes which indirectly regulate the expression levels of proteins directly involved in these responses. The proteins coded by various transcriptional factor genes regulate many genes responsible for stress tolerance in plants. This property makes them prime candidate target genes for inducing stress tolerance in plants in response to various environmental stresses (Tripathi et al., 2014).

Regulation of gene expression is very important strategy of plants to tackle these changing and challenging environments. *WRKY* transcription factors are one of the largest families of transcriptional regulators in plants. They form central parts of signaling webs that modulate many plant processes, such as gene regulation during plant development, immune response, biotic and abiotic stresses (Rushton et al. 2010; Giacomelli et al., 2010; Chen et al., 2010). The *WRKY* family is among the ten largest families of transcription factors in higher plants and is found throughout the green lineage; green algae and higher plants (Ulker and Somssich, 2004). The members of this family were first isolated from plants and they often exhibit sequence-specific DNA-binding characteristics and are capable of activating or repressing the transcription of multiple target genes (Ishiguro and Nakamura, 1994). In all its bearings, the defining feature of *WRKY* transcription factors is their DNA binding domain (Rushton et al., 1996). This WRKY domain is characterized by a highly conserved WRKYGQK

signature followed by a C2H2- or C2HC-type zinc finger motif (Eulgem et al., 2000). This domain specifically binds to the DNA motif called W-box with conserved DNA sequence of C/TTGACT/C present in the promoters of their target gene.

WRKY transcription factors have been remained the focus of *in silico* studies in many plant species. In the last decade, due to the advancements in sequencing techniques and data analysis, the genomes of several crop species have been sequenced. These genomes have been subject of bioinformatics analysis of various transcription factor gene families. The *WRKY* gene family is one of the most studied gene families from plants. There are many reports of bioinformatics studies of *WRKY* transcription factors in plant species. For example, computational genome wide analysis of WRKYs have been done recently in *Gossypium raimondii* (Cai et al., 2014), *Cucumis sativus* (Ling et al., 2011), *Lotus japonicas* (Song et al., 2014), *Brachypodium distachyon* (Wen et al., 2014), *Vitis vinifera* (Wang et al., 2014), *Fragariavesca* (Miao et al., 2012), *Helianthus annuus* (Giacomelli et al., 2010), rice and *Arabidopsis* (Wu et al., 2005). Thus, analysis of stress responsive genes within and between various plant species for different kinds of stresses would reveal a number of pivotal attributes spanning across the major plant divisions like dicots and monocots (Shaik et al., 2013). The analysis of *WRKY* transcription factors based on their categorization regarding stress response has not been carried out in *Arabidopsis*. In our study, we have focused on the stress related *WRKY* genes in *Arabidopsis*. Based upon studies conducted by Rushton et al. (2010) we have divided *WRKY* factors in three categories, i.e. induced by abiotic stresses, biotic stresses and both of these stresses simultaneously. Furthermore, all these *WRKY* genes are analysed using bioinformatics tools depending upon their stress individually and collectively to infer the common protein features responsible for different stresses. In addition, the promoter motif analysis of stress related *WRKY*s has also been done.

Materials and Methods

Identification of Stress Responsive WRKY Proteins in A. Thaliana

The information given by Rushton et al., (2010) was used to extract the stress responsive genes information present in the *A.thaliana*. Their protein sequences were retrieved from Plant Transcription Factor Data Base web server (PLANTTFDB, 2016) Version 3.0. A Blast search was also done among sequences of these genes in Plant TFDB and PHYTOZOME (2016) Version 9.1 using plant WRKY proteins as queries to find the homology among these sequences. To confirm the reliability of our results, all putative *WRKY* non-redundant genes were double checked with TAIR (2016) as well as PlantGDB (2016) respectively. The core WRKY domain of 60 amino acid residues was analysed in all the stress responsive genes in *A. thaliana*. For this purpose, multiple sequence alignment is performed using constraint-based multiple alignment tool COBALT (2016) and Unipro UGENE Software.

Conserved Domain Analysis of Stress Responsive WRKY Proteins

Conserved domain analyses of all the 72 AtWRKY proteins were evaluated using MEME (2016) online software Version 4.9.1 (Bailey et al., 2006). These analyses were performed for all three stress responsive group of genes. The parameters used during these analysis were, for abiotic; number of repetitions - any; maximum number of motifs - 03; and the optimum motif widths were constrained to between 3 and 299 residues. Similarly, for biotic and both stresses; maximum numbers of motifs were 7 and 20, respectively while other specifications were same as abiotic.

Phylogenetic Analysis

Multiple sequence alignments of 43 stresses responsive WRKY protein sequences were performed using built in CLUSTALW program in MEGA5 software (Tamura et al., 2011). These 43 stress responsive *WRKY* genes in *Arabidopsis* were separated and classified on the bases of their stress and WKY domain (Wu et al., 2005; Rushton et al., 2010). The GenBank accession numbers of these AtWRKYs with other properties has given in Table (1).The parameters used during alignment were as follows: gap open penalty: 10; gap extension penalty: 0.1; residue-specific gap penalties: on; hydrophilic penalties: on; gap separation distance: 0; end gap separation penalty: on; use negative matrix: off; delay divergent cutoff: 30%. On the basis of multiple sequence alignment, an unrooted phylogenetic tree was constructed using Neighbor Joining (NJ) method. In order to get reliable results, the resultant tree was then bootstrapped with 1000 iterations. Phylogenetic tree has formulated on individual stress base like abiotic, biotic and both the stress bases. Further, these stress related genes and non-stress related genes have used to make a tree from all stress responsive genes to infer the comparative analysis among *WRKY* genes in *A. thaliana*.

Chromosomal Mapping of WRKYs

All the *WRKY* genes were mapped on *A. thaliana* chromosomes based on the information available at the PHYTOZOME (2016) Version 9.1 and NCBI (2016) gene databases websites.

Gene structure analysis

The structure of exon–intron of *WRKY* genes was determined by comparing predicted coding sequences with their corresponding genomic sequences using the CLC sequence viewer (Version 6). For gene structure analysis, the whole genomic sequences of all the stress responsive *WRKY* genes were downloaded from PHYTOZOME (2016) Version 9.1. Further, these sequences were arranged on the basis of their intron and exon positions in the genes, these positions and their exact numeric value is calculated from NCBI (2016) gene databases.

Table 1 Stress responsive *WRKY* transcription factors identified in *Arabidopsis thaliana.*

Accession#	Common Name	Stress Response	Chromosome	Intron	Protein Length	WRKY Domain	Gene length
At2g04880	AtWRKY1	Abiotic	2	3	487	110-168,306-364	2638
At4g18170	AtWRKY28	Abiotic	4	2	318	172-229	1468
At4g26440	AtWRKY34	Abiotic	4	3	568	178-234,371-428	1955
At2g46400	AtWRKY46	Abiotic	2	2	295	104-164	1412
At3g01080	AtWRKY58	Abiotic	3	4	423	167-223,305-363	1942
At4g24240	AtWRKY7	Biotic	4	2	353	281-339	1959
At5g46350	AtWRKY8	Biotic	5	2	326	183-240	2825
At1g68150	AtWRKY9	Biotic	1	4	374	234-293	1748
At1g55600	AtWRKY10	Biotic	1	4	485	307-364	2248
At2g47260	AtWRKY23	Biotic	2	2	337	174-230	2264
At5g07100	AtWRKY26	Biotic	5	4	309	117-175,233-291	1563
At4g23550	AtWRKY29	Biotic	4	2	304	134-192	1276
At4g04450	AtWRKY42	Biotic	4	5	528	292-350	2735
At4g01720	AtWRKY47	Biotic	4	5	489	238-297	3633
At2g25000	AtWRKY60	Biotic	2	4	271	145-204	1676
At2g03340	AtWRKY3	Both	2	3	513	250-306,414-472	2798
At1g13960	AtWRKY4	Both	1	3	514	229-285,408-466	2795
At1g62300	AtWRKY6	Both	1	5	553	312-370	2596
At4g31550	AtWRKY11	Both	4	2	325	247-304,194-243	1592
At2g23320	AtWRKY15	Both	2	2	317	241-298	1439
At2g24570	AtWRKY17	Both	2	2	321	244-301	1900
At4g31800	AtWRKY18	Both	4	4	310	175-233	1828
At4g01250	AtWRKY22	Both	4	2	298	128-186	1612
At2g30250	AtWRKY25	Both	2	4	393	166-223	1962
At5g24110	AtWRKY30	Both	5	2	303	112-173	1594
At4g22070	AtWRKY31	Both	4	5	538	297-355	3014
At4g30935	AtWRKY32	Both	4	4	466	168-224,331-388	2228
At2g38470	AtWRKY33	Both	2	4	519	184-241,361-419	2405
At5g22570	AtWRKY38	Both	5	2	289	111-170	1245
At1g80840	AtWRKY40	Both	1	3	302	145-204	1808
At4g11070	AtWRKY41	Both	4	2	313	141-201	1491
At2g37260	AtWRKY44	Both	2	4	429	166-222,349-406	2225
At5g49520	AtWRKY48	Both	5	2	399	220-278	2570
At5g26170	AtWRKY50	Both	5	2	173	112-170	952
At5g45260	AtWRKY52	Both	5	6	1288	1209-1268	6406
At4g23810	AtWRKY53	Both	4	1	324	158-218	1612
At2g40750	AtWRKY54	Both	2	2	346	152-212	2014
At2g40740	AtWRKY55	Both	2	2	292	173-232	2100
At1g18860	AtWRKY61	Both	1	2	480	190-249	1735
At5g01900	AtWRKY62	Both	5	1	263	110-170	1061
At3g56400	AtWRKY70	Both	3	2	294	120-180	1553
At5g15130	AtWRKY72	Both	5	3	548	226-285	2747
At5g13080	AtWRKY75	Both	5	1	145	66-124	1410

Analysis of Cis Regulatory Elements in Promoter

In order to analyze the presence of *cis*-regulatory elements in the promoters of stress-responsive*WRKY* genes, the 1.0 kb upstream promoter sequences were retrieved from the PHYTOZOME (2016) Version 9.1. To identify putative *cis*-regulatory elements in the promoters, online tools like PLACE (2016) (Higo et al., 1999) and PLANTCARE (2016) (Lescot et al., 2002) were used. The PLACE (2016) and PLANTCARE (2016) web servers facilitate the identification of putative motifs in the given promoter sequences with motif databases of 469 and 435 different experimentally proved *cis*-regulatory elements.

Results

Phylogenetic and Conserved Domain Analysis of Stress Related WRKY Proteins

In this study, we have compiled the information of 43 stress responsive *WRKY* genes. These genes were further divided into abiotic, biotic and in both the stress responsive genes (Table 1). The predicted *Arabidopsis WRKY* proteins were subjected to multiple sequence alignment using ClustalW and a phylogenetic tree was constructed using MEGA 5.0 with NJ method with 1000 boottrap replicates (Figure1). Based on the classification of Wu et al., (2005), a total of III major groups with

subgroups (Ia, Ib, IIa, IIb, IIc, IId, IIIa, IIIb) were given in the phylogenetic tree of WRKY proteins. *WRKY*52 is the longest gene in all stress responsive *WRKY* genes and contains the WRKY domain at the extreme right side of C terminal of protein. While all the other stress related WRKY proteins for both the stresses show one or more than one WRKY domains at N or C terminal. For example only group Ia show more than one WRKY domains in the both stress related WRKY protein sequence (Figure 1). There is limited information about the relationship between functional diversity and gene multiplication in various crop plants. However, the possibility that members in the same phylogenetic subgroup show redundant, overlapping or related functions cannot be ruled out. Furthermore, consensus sequences and relative information of both biotic and abiotic stress responsive WRKY proteins is also shown in (Table 2).

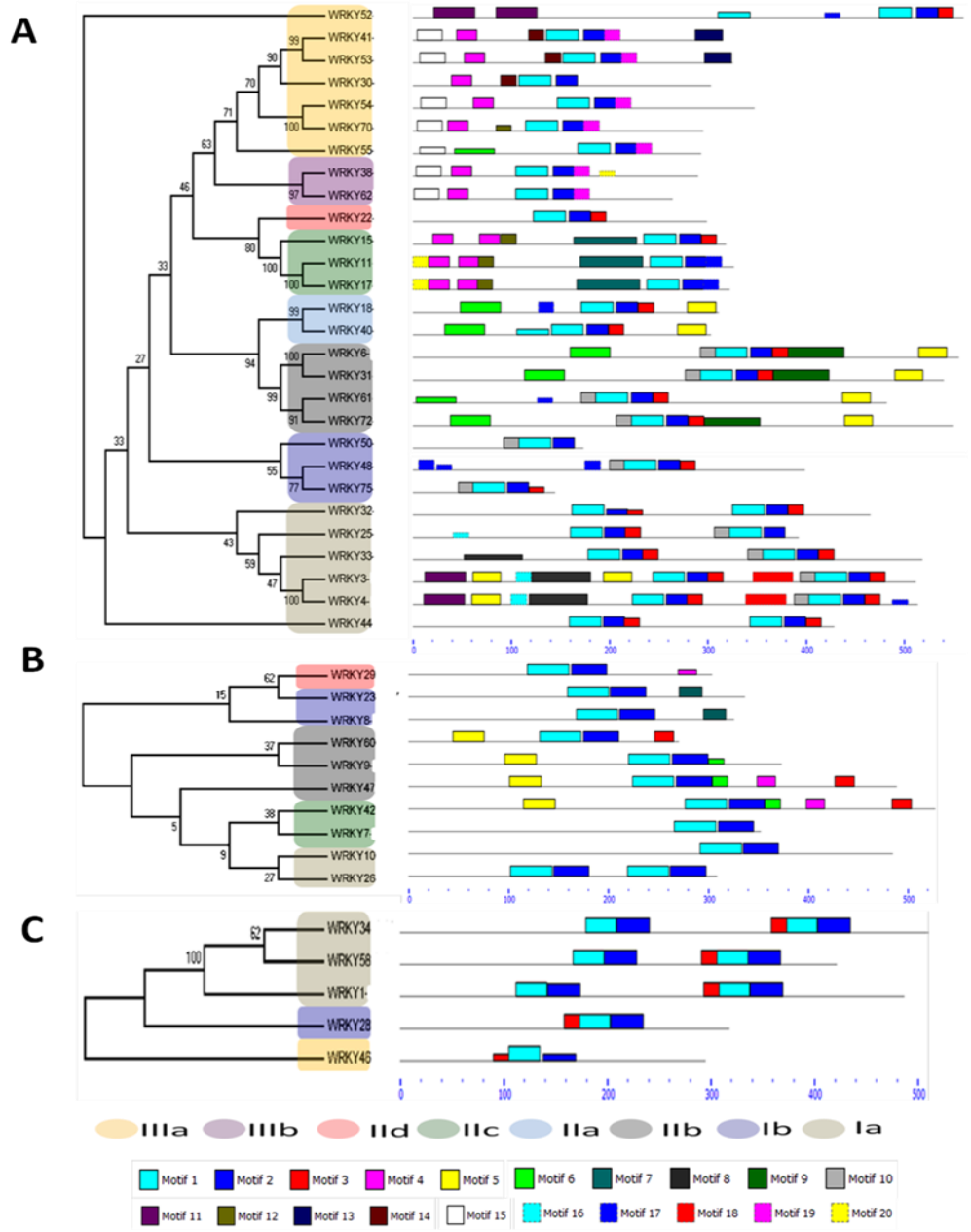

Figure 1 Phylogenetic relationship and motif composition among WRKY members in *Arabidopsis thaliana*. A) Responsive to both biotic and abiotic stress, B) Responsive to biotic stress and C) responsive to abiotic stresses. Multiple alignments of amino acids of WRKY genes were executed by Clustal W and the phylogenetic tree was constructed using MEGA 5.0 by the Neighbor-Joining (NJ) method with 1,000 bootstrap replicates. While, different color boxes are representing the WRKY groups present in the Arabidopsis. For example, from top to bottom different color boxes are representing groups like Ia, Ib and IIIa repectively. Schematic representation of the conserved motifs in the WRKY proteins from Arabidopsis elucidated by using online MEME server. Three conserved domains of different lengths are depicted on the protein map next to the phylogenetic tree. While the motif codes are indicated below the tree and the respective sequences of the domains are shown in Table 2.

Table 2 Multilevel consensus sequences for the MEME defined domains observed among both stress related WRKY proteins(1-20), Abiotic WRKY(21-27) and Biotic WRKY(28-30) in *Arabidopsis thaliana*

Motif	W	S	E-Value	Multiple consensus Sequence
Motif 1	32	34	8.7e-650	ExDILDDGYRWRKYGQVKQVKGSPFPRSYYRCT
Motif 2	21	33	9.4e-282	CPVRKQVERSADDPSMLITTY
Motif 3	15	20	1.9e-044	GEHNHPLPAARKSMA
Motif 4	20	13	2.9e-019	REELAKKILSSFEKAISLLN
Motif 5	28	9	2.8e-018	VSQLAAALTSDPNFAAALAAAISAILGG
Motif 6	40	5	9.1e-015	LQEELKKVKAENKKLREMLTQVCDNYNSLQMHLVALMQQQ
Motif 7	63	2	3.1e-014	IFLAPAPAQPVNSSGKPPLAGHPYRKRCFEHDHSEGFSGKISGSANGKCHCKKSRKNRMKRTV
Motif 8	59	2	1.1e-011	QPPGMFTVPPGLSPAMLLDSPSFFGLFSPLQGSFGMTHQQALAQVTAQAVQANANHMPQ
Motif 9	56	2	9.8e-011	TTTAAANMLLSGSMSSHDGLMNPTNLLARAILPCSSSMATISASAPFPTITLDLTH
Motif 10	15	11	1.5e-013	EKTVKEARVAVQTRS
Motif 11	41	4	2.1e-009	DLSGCSSLPTIQGFPRPFKELFFAGGVGFEPGPLPLSLEIF
Motif 12	15	3	2.6e-006	RTGHARFRRGPVHSI
Motif 13	27	2	3.7e-003	PIFDVNDQFDPTAEIDTGFPAFFHESI
Motif 14	15	3	7.1e-003	RKMLPKWSEKVRISP
Motif 15	25	6	7.6e-002	NSEKRKALNELIEGHDAAKQLQxLL
Motif 16	15	2	1.2e-001	VDPRFKQNRPTGLMI
Motif 17	15	5	1.4e-001	EDEHHHHQQQEQKNI
Motif 18	40	2	7.1e-001	TTTEHLSEASDGEEVGNAETDVGEKDEDEPDPKRRNTEVR
Motif 19	15	7	2.7e+000	RGTHTCSQNIALPAK
Motif 20	15	2	5.4e+000	MAVDIMRFPKIDDQT
Motif 21	41	11	4.0e-185	ARVSVEARSEADTLNDGYQWRKYGQKxVKGNPCPRSYYRCT
Motif 22	35	11	6.3e-119	GCPVRKQVERSAEDPSILITTYEGxHNHPLPPARR
Motif 23	19	3	2.4e-003	IAKDPNFTAALAAAISGII
Motif 24	18	2	2.7e+000	IATISASAPFPTITLDLT
Motif 25	31	4	9.5e+000	LQDELERVHEENHKLKELLAQTCEDYNALQM
Motif 26	15	2	4.0e+001	MAATTSAAAAMLLSG
Motif 27	22	2	9.9e+001	QFQQLHGFHHQEQEFELLKCFF
Motif 28	29	8	4.6e-125	GYRWRKYGQKLVKGSPYPRSYYKCTHPNC
Motif 29	31	6	4.3e-026	VKKHVERSSDDxKxVITTYEGKHHHDPPAAR
Motif 30	15	4	7.6e+001	RIVVQTKSEVDILDD

W: Width; S: Sites

In phylogenetic analysis of biotic stress-related *WRKY* genes, no second *WRKY* domain was present in these ten genes, however, one WRKY domain was present in all the biotic stress related genes either on the C terminal or N terminal (Figure1). Moreover, consensus sequences and relative information of biotic stress responsive WRKY proteins is also shown in Table 2. In abiotic stress related *WRKY* genes, *WRKY* 5.2 contain two WRKY domains in its protein. While all other *WRKY* genes show only one WRKY domain on C or N terminal and consensus sequences and relative information of abiotic stress responsive WRKY proteins is also shown in (Table 4).

Further, a circular phylogenetic tree was also formulated of all stress responsive *WRKY* genes for comparison of abiotic, biotic and both stress related *WRKY* genes in *A. thaliana* (Figure 2). On the bases of such comparison, it can be inferred from this phylogenetic analysis that there are some specific groups which play role in all the stresses, while others play a role either in individual stress or in more than one type of stress. For example, proteins from WRKY groups like Ia and Ib are active during all the stresses as compared to genes to group IIa which show its activity during biotic and both stresses. On the other hand, genes of group IIIa play active role in both abiotic and both stress related scenarios. Similarly, group II c show its activity during biotic and both stresses.

Mapping Atwrky Genes on A. Thaliana Chromosomes

All the predicted AtWRKYs were physically localized on *A. thaliana* chromosomes by CLC Sequence Viewer and Blast program and were mapped using online Map Viewer program from NCBI server (http://www.ncbi.nlm.nih.gov/projects/mapview/) and manually in excel sheet. The predicted 43 At*WRKY* genes are distributed across all the five *A. thaliana* chromosomes. The distribution of *WRKY* genes among the five chromosomes of *A. thaliana* is shown in the (Figure 3).

Compared with other chromosomes, chromosome 4 has the most numbers of WRKY genes, i.e. 13, followed by chromosome 2 and chromosome 5 with 12 and 10 genes, respectively. The chromosome 1 contains six *WRKY* genes while two *WRKY* genes are present on chromosome 3.

Comparison of DNA-Binding WRKY Domains

The WRKY DNA-binding domains of all the stress responsive WRKY proteins were compared for presence of conserved residues by alignment of complete WRKY

domain using COBALT (2016) from NCBI (2016) server) and ClustalW. The WRKY domain in all the WRKY proteins of *A. thaliana* revealed a high level of conservation with 14 out of 60 amino acids being absolutely conserved in all 43 proteins, including the WRKY signature and zinc finger motif (Figure 4). Another 9 amino acids were highly conserved with only two substitutions per site. The highly conserved Zinc finger motif (CH_2-CH_2) that is present in the WRKY domain is also shown in the Figure 4.

Gene Structure Analysis

The structure of exon–intron of all *WRKY*genes was analyzed. It was observed that all the members of stress responsive *WRKY* family had introns varying from two to seven, while there is only one gene that starts with the intron. All the stress responsive *WRKY* genes are analyzed separately depending upon the type of stress. It showed that abiotic stress responsive *WRKY* genes contain two to four introns (Figure 5). *WRKY* genes responsive to both kinds of stresses contained two to five introns. There are two genes which have five introns while there are four genes which have two and four introns each. The gene structure analysis of biotic *WRKY* family was more divergent than abiotic family.

Both biotic and abiotic stress responsive *WRKY* genes possess two to five introns. While there is a gene with 6 introns named as *WRKY* 52, whose size was around about 6Kb and could not be included in diagram. In all the present genes related to both stresses, most of the genes have two to three introns. The genes belonging to this catagory can be divided into two and three subgroups based on their exon/intron structures, respectively (Figure 5). From these results it can be inferred that these stress related *WRKY* family genes preserved a relatively conserved exon-intron composition in each subgroup during the evolution of the *WRKY* genes in *Arabidopsis*.

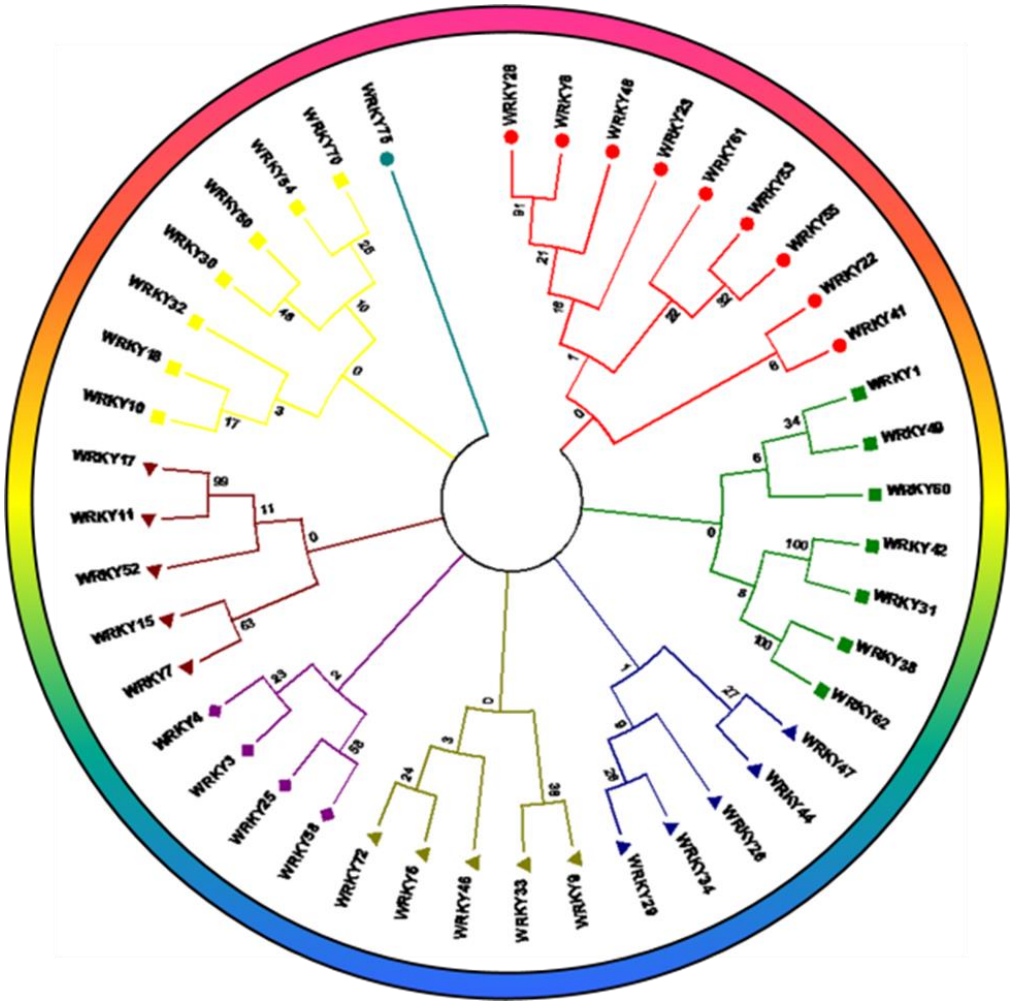

Figure 2 Phylogenetic tree**;** from amino acid sequences of stress responsive WRKY domains based on an alignment of Arabidopsis thaliana. The consensus unrooted phylogenetic tree was generated after an alignment of deduced Arabidopsis WRKY domains. The phylogenetic tree was generated with Clustal W and using the NJ (Neighbor-joining) method. The phylogenetic tree was inferred using MEGA 5.0 software. Reliability of the predicted tree was tested Using bootstrapping with 1000 replicates. Numbers at the nodes indicate how often the group to the right appeared among bootstrap replicates. Branch lines and nodes of subtree are coloured indicating different WRKY subgroups.

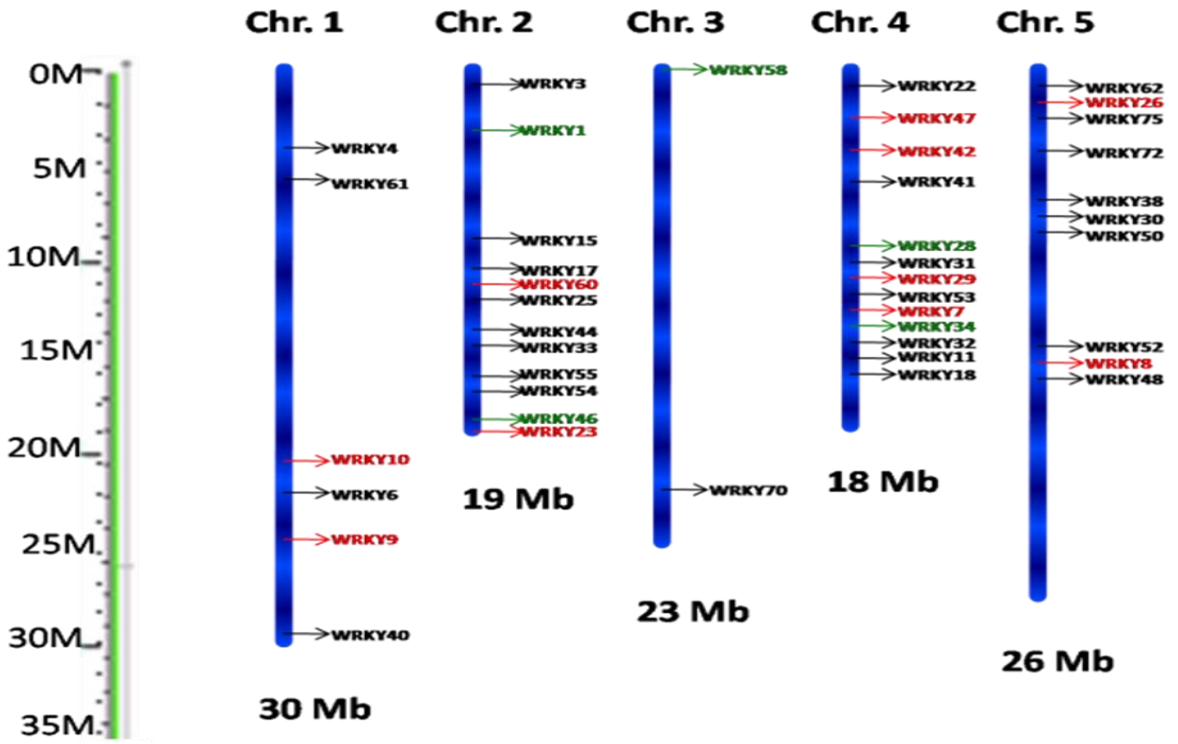

Figure 3 Chromosomal localization; the localization of 43 Stress responsive WRKY genes on Arabidopsis thaliana chromosomes. The chromosomes numbers are indicated at top of each bar. WRKY Genes are named according to their position and size on the chromosome also mentioned in a table. The relative position of AtWRKY and size of chromosome represented using vertical scale. The Green, Red and Black colors are showing the Abiotic, Biotic and both stress responsive genes. While lower side of the each chromosome is showing total length in Million Bases (Mb).

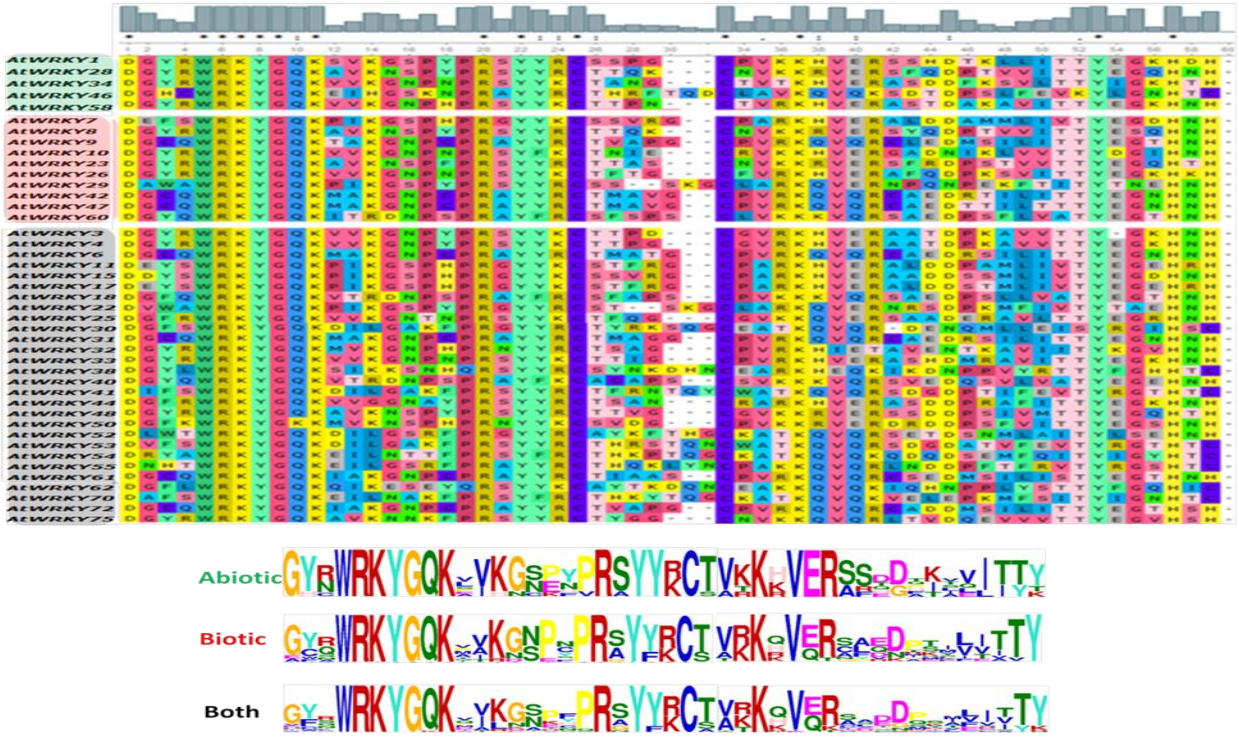

Figure 4 Multiple sequense alignment;a comparison of WRKY DNA binding domain among Stress responsive 43AtWRKYproteins; Abiotic (Light green), Biotic (Light red), and Both Stresses (Grey) color using **COBALT**-Blast and Unipro-**UGENE** software. A 60 residue WRKY domain has shown in different colors scheme of UGENE software. The three characters marked by '∗', ';' and '.' are used to mark conserved positions according to the Clustal X conventions. The consensus sequences of Abiotic, Biotic and both the stress responsive genes are also shown in the figure 4.

Analysis of Cis Regulatory Elements in Promoter

The critical role played by different members of *WRKY* TFs in response to various stresses can be based on their transcriptional regulation. For this purpose, we retrieved and investigated the 1Kb promoter sequence upstream to the start codon (ATG) of the 43 stress related *WRKY* genes. The dataset was then subjected to PLACE (Higo et al., 1998, 1999) and PlantCARE database (Lescotet et al., 2002) for identification of already experimentally described transcription factor binding sites also called as motifs. *WRKY* genes responsive to abiotic, biotic and both of these stresses were analysed individually. The most repeated *cis* elements found during search are shown in the (Figure 6).

In the promoters of WRKY genes related to abiotic stress (Figure 6), most of the genes showed a higher concentration of the cis-regulatory elements in the first 600 bp upstream of the start codon. However, it was noticeable that all the abiotic stress related genes have W-Box, which is a WRKY transcription factor binding site in the promoter region. In biotic (Figure 6b) and both (biotic and abiotic) stress responsive WRKY promoters (Figure 6c), there is a cluster of over-represented cis regulatory elements throughout the promoter region. The large numbers of W-Box cis elements are present in biotic and both stress responsive genes individually, which indicate that there is a regulatory effect being exerted by WRKY proteins themselves by binding with W-Box in response to various stresses. Moreover, the variety of *cis*-regulatory elements present in the *WRKY* gene family helps to regulate the level of expression of the members of this family by assuming different roles in stress response. Further, promoter analysis also revealed the presence of multiple binding sites of various other TFs involved in key developmental and biological processes.

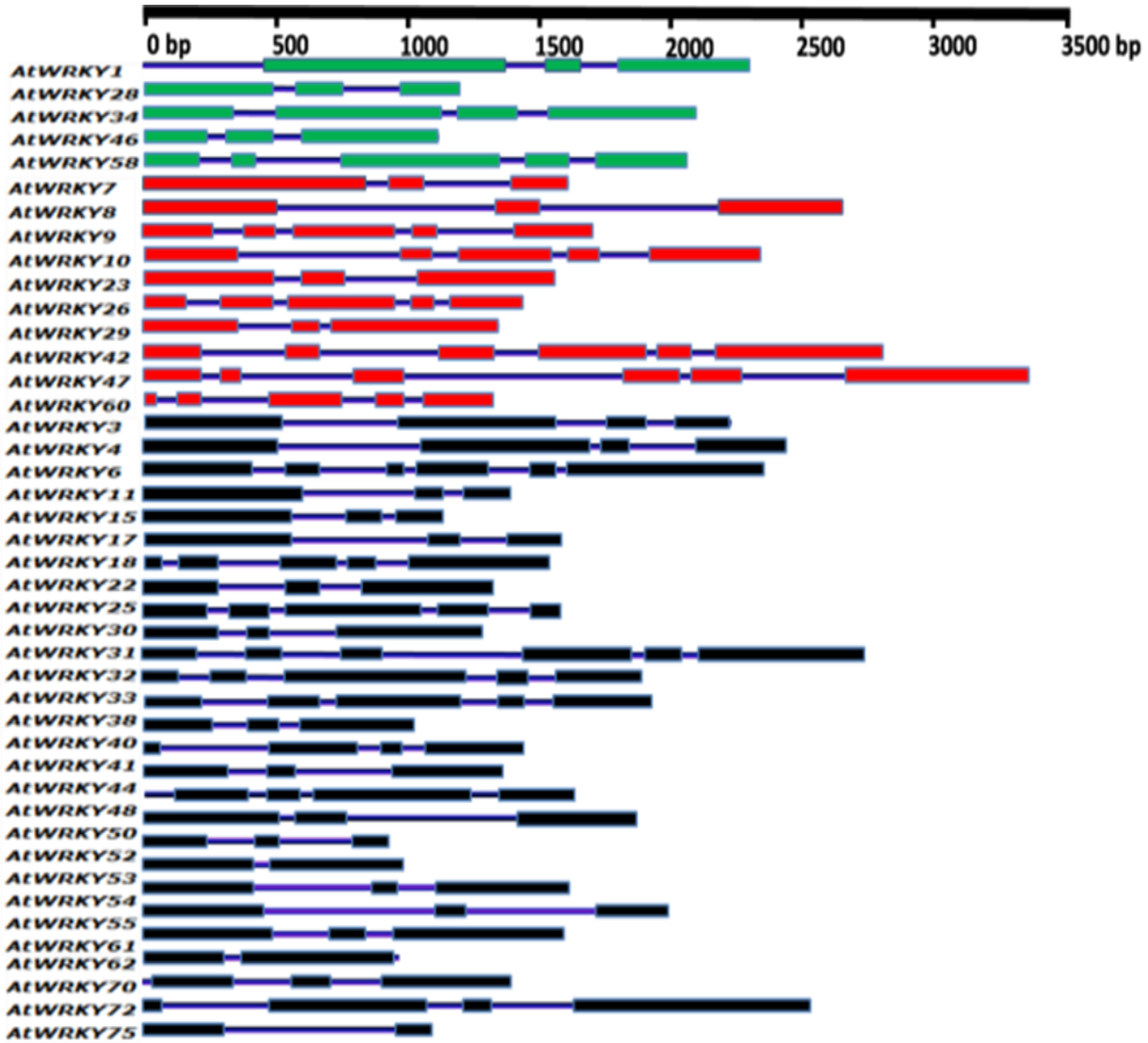

Figure 5 Gene structure analysis; Exon/intron structures of WRKY genes from *Arabidopsis thaliana*. Green colour boxes are showing exons in abiotic stress responsive genes while red colour boxes are biotic exon genes and black color boxes represent both stress related exons, Blue colour lines are depicting introns in all the genes. The sizes of exons and introns can be estimated using the scale at top.

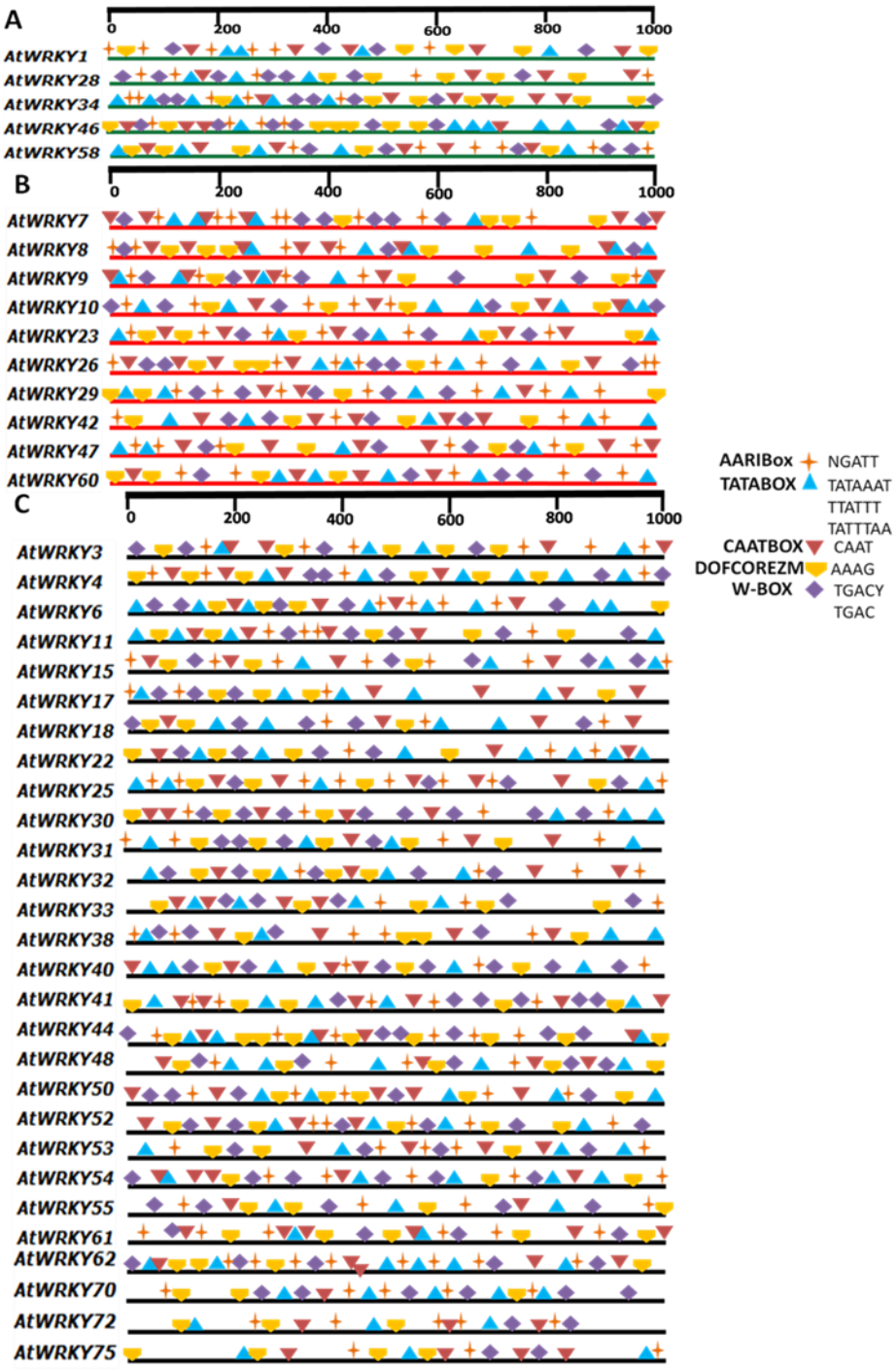

Figure 6 Promoter analysis of WRKY factors responsive to various stresses. A) Responsive to abiotic stress, B) Responsive to biotic stress and C) responsive to both biotic and abiotic stresses. Schemetic representation of various PLACE-based motifs in 1KB promoter region of biotic stress related WRKY genes. The presence of motifs on +strand is shown in the figure. The motif codes and respective sequences are indicated.

Discussion

WRKY transcription factors are commonly found in all land plants and perform diverse functions including the regulation of various biotic and abiotic stress responses. WRKY family has been the subject of intensive studies. For instance, they have been identified and classified in *Arabidopsis* (Eulgem et al., 2000), *Oryza sativa* (Xie. et al., 2005; Wu et al., 2005; Ross et al., 2007) *Hordeum vulgare* (Mangelsen et al., 2008), *Cucumissativus* (Ling et al., 2011) and *Brachypodium distachyon* (Tripathi et al., 2012). The WRKY gene family has 72 members in *Arabidopsis* (Eulgem et al., 2000), however in this article, we have identified 43 stress responsive *WRKY* genes according to Rushton et al., (2010) who reviewed and categorized different *WRKY*s in response to different

stresses including abiotic, biotic and multiple stresses. Further, the genome wide analysis of these stress responsive *WRKY* genes includes phylogenetic analysis, motif analysis, chromosomal mapping, gene structure analysis and promoter analysis. The phylogenetic tree obtained from an alignment of the WRKY domains in *Arabidopsis* indicated that the stress responsive *WRKY* genes identified in *Arabidopsis* can be divided into the three major groups and subgroups (I, II (IIa, IIb,IIc,IId,andIIe) and III(IIIa,IIIb,andIIIc) as previously described in plant species (Wu et al., 2005). Based on the AtWRKY domains, we observed the same phylogenetic relationship and grouping as previously reported (Rushton et al., 2010). Similarly, members within the same groups, or subgroups within group II and III, shared a similar length and amino-acid motif composition, gene structure (intron/exon organization), indicating their close evolutionary relationship. Amino acid residues of WRKYGQK are the distinguishing regions of the *WRKY* transcription factor. In *Lotus japonicas*, WRKY domain is conserved in different WRKY groups. While, minor amino acids substitutions are also present in their groups (Song et al., 2014). Quite recently, multiple-alignment of rice WRKY proteins shows a 60 amino acid region that is highly conserved and there are portions of high and low similarity between these WRKY proteins (Nadarajah et al. 2014). Similarly Tang et al., (2013) have performed characterization and co-expression analysis of WRKY orthologs involved in responses to multiple abiotic stresses in *Brassica campestris* ssp. *chinensis*. It has been well established that particularly regulatory proteins like WRKYs, rarely act alone. Very often, they interact either transiently or permanently with each other or different proteins to regulate biological functions in living systems (Chi et al., 2013). A genome wide identification of *WRKY* transcription factors in chinese cabbage (*Brassica rapa* ssp. *pekinensis*) has been done which have revealed collinearity in their expression patterns under abiotic and biotic stresses in leaves (Tang et al., 2013).

In our study, we were able to identify conserved domains in stress-responsive AtWRKY transcription factors. The genes of respective TFs were also localized on the *Arabidopsis* chromosomes. The physical distribution of all *WRKY* genes in various other crops has also been reported previously. For example, in the *A.thaliana*, the 72 *WRKY* genes were found to be organized among all the five chromosomes with a maximum of 17 *WRKY* genes on chromosome 2 and 5 each (Song and Gao, 2014). On the other hand, in monocots like rice, 100*WRKY* genes were shown to be distributed on the 12 chromosomes, with a maximum 22 *WRKY* genes on chromosome 1 (Ross et al., 2007). Quite recently, 59 genes of grapevine encoding *WRKY*genes have been physically localized on 19 chromosomes, with chromosome 4 and 7 showing a distribution of a maximum of 8 and 7 *WRKY* genes, respectively (Wang et al., 2014), while no *WRKY* gene was reported to be present on chromosome 3.

Further, we also observed a unique function of some WRKY proteins. For example, the AtWRKY32 a both biotic and abiotic stress responsive WRKY protein contains two of the same type of WRKY domains in the IaCTWD and IaNTWD, belonging to Group I; however, in a phylogenetic analysis, it can be clustered together with Group IIb members, while AtWRKY15 with WRKY domain features of Group IIc was clustered with Group IId members because of the absence of some motifs. The presence of two WRKY domains in stress responsive genes is shown in (Figure 1) and (Table 1).These results also hold the belief that members of group I may represent the ancestral form of the *WRKY* family (Ulker and Somssich, 2004). Moreover, *WRKY* gene belonging to group I, which are described by the presence of two WRKY domains, enclose approximately 20% of the entire AtWRKY family and is comparable in size to Group I in rice, tomato and castor bean (Wang et al., 2014). It has been well established that WRKY TFs are involved in the regulation of plant response to biotic and abiotic stresses (Rushton et al., 2010). For example, AtWRKY33 play an important role in *Arabidopsis* tolerance in response to high concentrations of NaCl (Jiang and Deyholos, 2009), while AtWRKY22 appears to be involved in the regulation of dark-induced leaf senescence in *Arabidopsis* (Zhou et al., 2011). Recently, Ali et al., (2014) reported that WRKY33, WRKY11 and WRKY17 play an importance role in plant resistance against beet cyst nematode in *Arabidopsis* (Ali et al., 2014). The WRKY domains in various *WRKY* genes from *Arabidopsis* showed pair wise relationships with grapevine, such as those between AtWRKY33and VvWRKY24, and AtWRKY22 andVvWRKY49 (Guo et al., 2014)Since *Arabidopsis* WRKYs involved in stress response mainly belong to Groups II-a and III(Wang et al., 2014), the current study focus on all the stress responsive *Arabidopsis WRKY* genes. Further, genome wide analysis of these genes would help in characterization and functional homology in *Arabidopsis* and other plant species. For example, AtWRKY 28plays a role in abiotic stress; in our study it also ensures that AtWRKY 28 is an abiotic stress responsive candidate. Similarly, AtWRKY18, AtWRKY40and AtWRKY60of Group II-a exhibited a complex pattern of expression in responses to ABA and abiotic stresses (Chen et al., 2010).While, same was observed in our studies regarding stress response of Group II-aAt*WRKY* genes.

It is widely established that the intron/exon structure helps to understand the evolutionary relationships (Hu and Liu, 2011). Our analysis showed similar findings to the previous studies, such that abiotic, biotic and both stress responsive WRKY members had 2-4 introns, 2-5 introns, and 2-6 introns, respectively (Figure 5). Similarly, promoter analysis also revealed the presence of different binding sites of *cis* regulatory elements in AtWRKY stress responsive genes.

Cis elements like W-Box is present in almost every stress responsive WRKY gene, because it is a WRKY transcription factor binding site in the promoter region. Even though, other cis-regulatory elements like Dof and AAR1 are also found in large numbers. The analyses have shown that the promoter regions of all the up- as well as

down-regulated genes contain multiple copies of the basic elements required for promoter identity i.e. TATA box and CAAT-box (Casimiro et al., 2008).

These findings will help in identifying and understanding more evolutionary relationships among the stress responsive *WRKY* transcription factors in different plant species and will elucidate more co-regulatory relationships for *WRKY*s under multiple stress response. Despite many recent advances in phylogenetic studies of *WRKY*s in *Arabidopsis* and many other species, the stress response of most *WRKY*genes in physiological processes still needs to be explained. The bioinformatics analysis of the *WRK Y*transcription factor family conducted in the present study provides an overall picture of the composition and classification of WRKY family members in *Arabidopsis*. This information will facilitate selecting candidate genes for stress conditions and further functional and comparative characterization.

References

Ali MA, Wieczorek K, Kreil DP, Bohlmann H. 2014. The beet cyst nematode *Heteroderaschachtii* modulates the expression of WRKY transcription factors in syncytia to favour its development in Arabidopsis roots. PLoS ONE 9(7): e102360.

Bailey TL, Williams N, Misleh C, Li WW. 2006. MEME: discovering and analyzing DNA and protein sequence motifs. Nucleic Acids Res 34: W369-W373.

Cai C, Niu E, Du H, Zhao L, Feng Y, Guo W. 2014. Genome-wideanalysis of the transcription factor genefamily in *Gossypiumraimondii* and the expression of orthologs in cultivatedtetraploidcotton. The Crop Journal 2: 87-101.

Casimiro AC, Vinga S, Freitas AT, Oliveira AL. 2008. An analysis of the positional distribution of DNA motifs in promoter regions and its biological relevance. BMC bioinformatics 9: 89.

Chen H, Lai Z, Shi J, Xiao Y, Chen Z, Xu X. 2010. Roles of Arabidopsis *WRKY18*, *WRKY40* and *WRKY60* transcription factors in plant responses to abscisic acid and abiotic stress. BMC Plant Biol 10: 281.

Chi Y, Yang Y, Zhou Y, Zhou J, Fan B, Yu J-Q, Chen Z. 2013. Protein-protein interactions in the regulation of WRKY transcription factors. Mol Plant 6:287–300.

COBALT. 2016. Multiple sequence alignment is performed using constraint-based multiple alignment tool (http://www.stva.ncbi.nlm.nih.gov/tools/cobalt/re_cobalt.cgi)

Eulgem T, Rushton PJ, Robatzek S, Somssich IE. 2000. The WRKY superfamily of plant transcription factors. Trends in plant science 5: 199-206.

Giacomelli JI, Ribichich KF, Dezar CA, Chan RL. 2010. Expression analyses indicate the involvement of sunflower WRKY transcription factors in stress responses, and phylogenetic reconstructions reveal the existence of a novel clade in the Asteraceae. Plant science 178: 398-410.

Guo C, Guo R, Xu X, Gao M, Li X, Song J, Zheng Y, Wang X. 2014. Evolution and expression analysis of the grape (Vitisvinifera L.) WRKY gene family. Journal of experimental botany 65: 1513-1528.

Higo K, Ugawa Y, Iwamoto M, and Korenaga T. 1999. Plant cis-acting regulatory DNA elements (PLACE) database. Nucleic Acids Research 27 (1): 297-300.

Higo K, Ugawa Y, Iwamoto M, Higo H. 1998. PLACE: A database of plant cis-acting regulatory DNA elements. Nucleic Acids Res 26: 358-359.

Hirayama T, Shinozaki K. 2010. .Research on plant abiotic stress responses in the post-genome era: past, present and future. Plant J., 61:1041-1052.

Ishiguro S, Nakamura K. 1994. Characterization of a cDNA encoding a novel DNA-binding protein, SPF1, that recognizes SP8 sequences in the 5' upstream regions of genes coding for sporamin and β-amylase from sweet potato. Mol. Gen. Genet 244(6): 563–571.

JiangY, Deyholos MK. 2009. Functional characterization of Arabidopsis NaCl-inducible *WRKY25* and *WRKY33* transcription factors in abiotic stresses. Plant Mol. Biol 69: 91–105.

Lescot M, Dehais P, Thijs G, Marchal K, Moreau Y, Van de Peer Y, Rouze P, Rombauts S. 2002. PlantCARE, a database of plant cis-acting regulatory elements and a portal to tools for in silico analysis of promoter sequences. Nucleic Acids Research 30: 325–327.

Ling J, Jiang W, Zhang Y, Yu H, Mao Z, Gu X, Huang S, Xie B. 2011. Genome-wide analysis of WRKY gene family in Cucumis sativus. BMC genomics 12: 471.

Liu H, Yang W, Liu D, Han Y, Zhang A, Li S. 2011. Ectopic expression of a grapevine transcription factor VvWRKY11 contributes to osmotic stress tolerance in Arabidopsis. Molecular biology reports 38: 417-427.

Mangelsen E, Kilian J, Berendzen KW, Kolukisaoglu H, Harter K, Jansson C, Wanke D. 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: 194.

MEME. 2016. Motif based sequence analysis tool. (http://meme.nbcr.net/meme/cgi-bin/meme.cgi).

Miao LX, Zhang YC, Yang XF, Jiang GH. 2012. Genome-wide identification and analysis of WRKY transcription factors in Fragariavesca. Journal of Nuclear Agricultural Sciences 8: 1124-1131.

Nadarajah K, Omar NS, Thing TY. 2014. Molecular characterization of a WRKY gene from Oryza sativa indica cultivar UKMRC9 7: 63–71.

NCBI. 2016. (http://www.ncbi.nlm.nih.gov/)

PHYTOZOME. 2016. Plant Comparative Genomics portal of the Department of Energy's Joint Genome Institute. (https://phytozome.jgi.doe.gov/pz/portal.html).

PLACE. 2016. A Database of Plant Cis-acting Regulatory DNA Elements. (http://www.dna.affrc.go.jp/PLACE)

PLANTCARE. 2016. A plant *cis*-acting regulatory element database. (http://bioinformatics.psb.ugent.be/webtools/plantcare/html)

PlantGDB. 2016. The plantgdb.org web resource was managed most recently under NSF-funded project (IOS-1221984)(http://www.plantgdb.org/AtGDB/)

PLANTTFDB. 2016. Plant Transcription Factor Database, v3.0. (http://planttfdb.cbi.pku.edu.cn/).

Ross CA, Liu Y, Shen QJ. 2007. The WRKY gene family in rice (Oryza sativa). Journal of Integrative Plant Biology 49: 827-842.

Rushton PJ, Somssich IE, Ringler P, Shen QJ. 2010. *WRKY* transcription factors. Trends Plant Sci., 15: 247–258.

Rushton PJ, Torres JT, Parniske M, Wernert P, Hahlbrock K, Somssich I. 1996. Interaction of elicitor-induced DNA-binding proteins with elicitor response elements in the promoters of parsley PR1 genes. The EMBO Journal 15: 5690.

Shaik R., Ramakrishna W. 2013. Genes and Co-Expression Modules Common to Drought and Bacterial Stress Responses in Arabidopsis and Rice. PLoS ONE 8(10): e77261.

Song H, Wang P, Nan Z, Wang X. 2014. The WRKY Transcription Factor Genes in Lotus japonicus.Int J Genomics 2014:420128.

Song Y, Gao J. 2014. Genome-wide analysis of WRKY gene family in Arabidopsis lyrata and comparison with Arabidopsis thaliana and Populustrichocarpa. Chinese Science Bulletin 59: 754-765.

TAIR. 2016. The Arabidopsis Information Resource (TAIR) maintains a database of genetic and molecular biology data for the model higher plant *Arabidopsis thaliana* (http://www.arabidopsis.org)

Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S (2011). MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol 28(10): 2731–2739.

Tang J, Wang F, Wang Z, Huang Z, Xiong A, Hou X. 2013. Characterization and co-expression analysis of WRKY orthologs involved in responses to multiple abiotic stresses in Pak-choi (Brassica campestris ssp. chinensis). BMC plant biology 13: 188

Tripathi P, Rabara RC, Langum TJ , Boken AK, Rushton DL, Boomsma DD, Rinerson CI, Rabara J, Reese RN, Chen X. 2012. The WRKY transcription factor family in Brachypodiumdistachyon. BMC Genomics 13:270.

Tripathi P, Rabara RC, Rushton PJ. 2014. A systems biology perspective on the role of WRKY transcription factors in drought responses in plants. Planta 239: 255-266.

Ulker B, Somssich IE. 2004. *WRKY* transcription factors: from DNA binding towards biological function. Curr.Opin. Plant Biol 7: 491–498.

Wang H, Yin X, Li X, Wang L, Zheng Y, Xu X, Zhang Y, Wang X. 2014. Genome-Wide Identification, Evolution and Expression Analysis of the Grape (Vitisvinifera L.) Zinc Finger-Homeodomain Gene Family. International journal of molecular sciences 15: 5730-5748.

Wang M, Vannozzi A, Wang G, Liang Y-H, Tornielli GB, Zenoni S, Cavallini E, Pezzotti M, Cheng Z-MM. 2014. Genome and transcriptome analysis of the grapevine (Vitis vinifera L.) WRKY gene family. Horticulture Research 1.

Wen F, Zhu H, Li P, Jiang M, Mao W, Ong C, Chu Z. 2014. Genome-Wide Evolutionary Characterization and Expression Analyses of WRKY Family Genes in Brachypodium distachyon. DNA research: dst060.

Wu K-L, Guo Z-J, Wang H-H, Li J. 2005. The WRKY family of transcription factors in rice and Arabidopsis and their origins. DNA research 12: 9-26.

Xie Z, Zhang Z-L, Zou X, Huang J, Ruas P, Thompson D, Shen QJ. 2005. Annotations and functional analyses of the rice WRKY gene superfamily reveal positive and negative regulators of abscisic acid signaling in aleurone cells. Plant physiology 137: 176-189.

Xiong L, Schumaker KS, Zhu JK. 2002. Cell signaling during cold, drought, and salt stress.Plant Cell (Suppl) 14: S165–S183.

Yamaguchi-Shinozaki K, Shinozaki K. 2006. Transcriptional regulatory networks in cellular responses and tolerance to dehydration and cold stresses. Annu Rev Plant Biol, 57: 781–803.

Zhou X, Jiang Y, Yu D. 2011. WRKY22 transcription factor mediates dark-induced leaf senescence in Arabidopsis. Mol Cells 31: 303–313.

Zhou X, Wang G, Sutoh K, Zhu JK, Zhang W. 2008. Identification of cold- inducible microRNAs in plants by transcriptome analysis. Biochim Biophys Acta 1779: 780–788.