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

Shaiq Sultan¹, Muhammad Amjad Ali², Rana Muhammad Atif³,4, Farrukh Azeem¹, Habibullah Nadeem¹, M. Hussnain Siddique¹, Ertuğrul Filiz², Khadim Hussain¹, Amjad Abbas²

¹Department of Bioinformatics & Biotechnology, Government College University, 38000 Faisalabad, Pakistan
²Department of Plant Pathology, University of Agriculture, 38040 Faisalabad, Pakistan
³Centre of Agricultural Biochemistry & Biotechnology, University of Agriculture, 38040 Faisalabad, Pakistan
⁴Department of Plant Breeding & Genetics, University of Agriculture, 38040 Faisalabad, Pakistan

ARTICLE INFO

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

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.

ABSTRACT

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), *Fragaria vesca* (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 WRKYs 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 Uniprot 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.
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 bootstrap replicates (Figure 1). Based on the classification of Wu et al., (2005), a total of III major groups with
Subgroups (Ia, Ib, IIa, IIb, IIC, IId, IIIa, IIb) were given in the phylogenetic tree of WRKY proteins. WRKY52 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 respectively. 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.
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 (Figure 1). 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 both 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 AtWRKY 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 (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 the all stress responsive WRKY proteins were compared for presence of conserved residues by alignment of complete WRKY

---

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

<table>
<thead>
<tr>
<th>Motif</th>
<th>W</th>
<th>S</th>
<th>E-Value</th>
<th>Multiple consensus sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Motif 1</td>
<td>32</td>
<td>34</td>
<td>8.7e-650</td>
<td>ExDILDDGYWRKYQVKVKGSPPRSYYRCT</td>
</tr>
<tr>
<td>Motif 2</td>
<td>21</td>
<td>33</td>
<td>9.4e-282</td>
<td>CPVRKQVERSADPSMLITY</td>
</tr>
<tr>
<td>Motif 3</td>
<td>15</td>
<td>20</td>
<td>1.9e-044</td>
<td>GEHNHLPQAARKSM</td>
</tr>
<tr>
<td>Motif 4</td>
<td>20</td>
<td>13</td>
<td>2.9e-019</td>
<td>REELAKILSSFEKAILNN</td>
</tr>
<tr>
<td>Motif 5</td>
<td>28</td>
<td>9</td>
<td>2.8e-018</td>
<td>VSQAAAATSDPFAAAALAAAILGG</td>
</tr>
<tr>
<td>Motif 6</td>
<td>40</td>
<td>5</td>
<td>9.1e-015</td>
<td>LQEELKKVAKENKLEMLTQVCDNYSLQNMHLVALMQQ</td>
</tr>
<tr>
<td>Motif 7</td>
<td>63</td>
<td>2</td>
<td>3.1e-014</td>
<td>IFGLAPAQPVPNSSGKPLAGHPYRRCEFDHSSEGFSXGKISANGKCHCKSKRNNRMKRTV</td>
</tr>
<tr>
<td>Motif 8</td>
<td>59</td>
<td>2</td>
<td>1.1e-011</td>
<td>QPPQFTTVPGPPLAPMLLDDPSFGLFSPLQFGMSMTQHQAQLAQVTAGQVANHMPQ</td>
</tr>
<tr>
<td>Motif 9</td>
<td>56</td>
<td>2</td>
<td>9.8e-011</td>
<td>TTTAAANLILSSSDHGLMNPTNLLARAILPCSSSMAASAPPTITDLTH</td>
</tr>
<tr>
<td>Motif 10</td>
<td>15</td>
<td>11</td>
<td>1.5e-013</td>
<td>EKTVKEARAVAVQRS</td>
</tr>
<tr>
<td>Motif 11</td>
<td>41</td>
<td>4</td>
<td>2.1e-009</td>
<td>DLSGCSPLPTLGPRFPRKELFFAGVGFEPGPLPLEIF</td>
</tr>
<tr>
<td>Motif 12</td>
<td>15</td>
<td>3</td>
<td>2.6e-006</td>
<td>RTGHARFRGRPVHIS</td>
</tr>
<tr>
<td>Motif 13</td>
<td>27</td>
<td>2</td>
<td>3.7e-003</td>
<td>PIFDVNPDEFPDAEDTFGPAFFHESI</td>
</tr>
<tr>
<td>Motif 14</td>
<td>15</td>
<td>3</td>
<td>7.1e-003</td>
<td>RKMLPKWSEKRISP</td>
</tr>
<tr>
<td>Motif 15</td>
<td>25</td>
<td>6</td>
<td>7.6e-002</td>
<td>NSEKRKALNELEGHDAAKQLDL</td>
</tr>
<tr>
<td>Motif 16</td>
<td>15</td>
<td>2</td>
<td>1.2e-001</td>
<td>VDPQFKQNPRTGLMI</td>
</tr>
<tr>
<td>Motif 17</td>
<td>15</td>
<td>5</td>
<td>1.4e-001</td>
<td>EDEHHHHQOQEQKNI</td>
</tr>
<tr>
<td>Motif 18</td>
<td>40</td>
<td>2</td>
<td>7.1e-001</td>
<td>TTTEHLSAESDGGEEVNGAETDVGEKDEDPKRRNTEVR</td>
</tr>
<tr>
<td>Motif 19</td>
<td>15</td>
<td>7</td>
<td>2.7e+000</td>
<td>RGTHTCQONALPAK</td>
</tr>
<tr>
<td>Motif 20</td>
<td>15</td>
<td>2</td>
<td>5.4e+000</td>
<td>MAVDMRPFPKIDQT</td>
</tr>
<tr>
<td>Motif 21</td>
<td>41</td>
<td>11</td>
<td>4.0e-185</td>
<td>ARVSVERSEADTLNGDYQWRKYQKAQVKGNPXRSSYYRCT</td>
</tr>
<tr>
<td>Motif 22</td>
<td>35</td>
<td>11</td>
<td>6.3e-119</td>
<td>GCPVRKQVERAEPSILITTRYEGXNHHLPPARR</td>
</tr>
<tr>
<td>Motif 23</td>
<td>19</td>
<td>3</td>
<td>2.4e-003</td>
<td>IAKDPNTAALAAISGH</td>
</tr>
<tr>
<td>Motif 24</td>
<td>18</td>
<td>2</td>
<td>2.7e+000</td>
<td>IATISASAPEFTIDLLT</td>
</tr>
<tr>
<td>Motif 25</td>
<td>31</td>
<td>4</td>
<td>9.5e+000</td>
<td>LQDELERVHEEKNKLKELLAQTEDYNALQM</td>
</tr>
<tr>
<td>Motif 26</td>
<td>25</td>
<td>2</td>
<td>4.0e+001</td>
<td>MAATTSAAAMLLSG</td>
</tr>
<tr>
<td>Motif 27</td>
<td>22</td>
<td>2</td>
<td>9.9e+001</td>
<td>QFQQLHGHIIQEEFELLCFF</td>
</tr>
<tr>
<td>Motif 28</td>
<td>29</td>
<td>8</td>
<td>4.6e-125</td>
<td>GYWRWKYGQKLKVGSPPRSTYKCTHPNC</td>
</tr>
<tr>
<td>Motif 29</td>
<td>31</td>
<td>6</td>
<td>4.3e-026</td>
<td>VSKHVERSSDDxKIVTYYEGHKHHPPAAP</td>
</tr>
<tr>
<td>Motif 30</td>
<td>15</td>
<td>4</td>
<td>7.6e+001</td>
<td>RIVVQTKEVDDILDD</td>
</tr>
</tbody>
</table>

W: Width; S: Sites
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 category 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 sequence alignment; a comparison of WRKY DNA binding domain among Stress responsive 43AtWRKY proteins; 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 (Lescot 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.
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), Cucumis sativus (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 WRKYs 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, and IIe) and III (IIIa, IIb, IIIc, and IIId)) 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 (Rushston 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 japonicus, WRKY domain is considered 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. pekinenisis) 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, 100WRKY 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 AtWRKYfamily 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 (Rushston 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 important 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 and VvWRKY49 (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, AtWRKY40 and AtWRKY60 of 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-a AtWRKY 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 WRKYs under multiple stress response. Despite many recent advances in phylogenetic studies of WRKYs 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 WRKY 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**


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


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


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)


