

Molecular characterization of a *WRKY* gene from *Oryza sativa indica* cultivar UKMRC9

Kalaivani Nadarajah^{1*}, Nurfarahana Syuhada Omar¹ and Tan Yun Thing²

¹School of Environmental Sciences and Natural Resources, Faculty of Science and Technology, Universiti Kebangsaan Malaysia, Malaysia

²School of Biosciences and Biotechnology, Faculty of Science and Technology, Universiti Kebangsaan Malaysia, Malaysia

*Corresponding author: vani@ukm.my

Abstract

It has been almost two decades since the work was initiated on characterization and function of the *WRKY* gene. This gene has been implicated in various biological and other developmental processes. It is involved in systemic acquired resistance (SAR) and has the potential to contribute to a wide spectrum resistance in plants. The aim of this study is to isolate, sequence and analyze the *WRKY* gene from a Malaysian disease tolerant rice cultivar, UKMRC9. The nucleotide sequence was translated and the 410 amino acid *WRKY* protein was subjected to physico-chemical, structural and phylogenetic analysis. The results showed that the protein had low level homology with other *WRKY* proteins. It only showed conservative sequence within the signature *WRKY* domains. The physico-chemical characteristics predicted the protein as a globular intracellular protein that is hydrophilic with borderline stability and a predicted molecular weight and isoelectric point of 45.1 kDa, and 4.91, respectively. The resolved 3D protein structure of *WRKY* consisted of four β -sheet strands with the zinc ion bound by two Cysteine and two Histidine residues, forming a zinc-binding pocket, where the *WRKYGQK* residues exist. This corresponds to the most N-terminal β -strand and is the interactive domain of this transcriptional factor.

Keywords: *Oryza sativa*, disease resistance, regulatory protein, *WRKY*, transcription factors.

Abbreviations: G_Glycine; Q_Glutamine; PTI_PAMP triggered immunity; ETI_Effector triggered immunity.

Received: 07 June 2013; Revised: 4 Nov 2013; Accepted: 3 Jan 2014.

Introduction

Plants do not have a centralized controlling system such as the nervous present in the animal kingdom. For this reason, together with their sessile nature, they require an even faster, wider and more precise transcriptional regulation to cope with drastic environmental changes. In *Arabidopsis thaliana*, regulation of transcription is performed by at least 1533 transcriptional regulators, which represent ~ 5.9% of the estimated total number of genes (Riechmann et al., 2000). Some of them belong to large families (with up to 139 members for the bHLH family) that share one or more conserved motifs. Among those are the *WRKY* gene family that is one of the biggest with 74 members in *Arabidopsis* and more than 109 *WRKY* proteins in *Oryza sativa*. The first *WRKY*-cDNA clone was characterized in 1994 from sweet potato (Ishiguro and Nakamura, 1994), and their description as a class of transcription factor followed soon after (Eulgem et al., 2000). *WRKY* genes are characterized by the presence of two or, more often, one ~ 60 amino acid *WRKY* motif including a zinc-finger-like motif C-X4-7-C-X23-28-H-X-(H/C) that provides binding properties to DNA together with a very highly conserved *WRKYGQK* sequence that gives binding specificity and the name to the family. A NMR solution structure of the conserved motif of *Arabidopsis thaliana* *WRKY4* (At*WRKY4*) transcription factor, indicates that the residues of the *WRKYGQK* sequence deeply enter the major groove of the DNA in such a way that lateral chains of Arg (R), Lys (K), Gln (Q), and Lys (K), which are

likely the most important residues in the sequence-specific recognition domain, contact the DNA bases (Yamasaki et al., 2005). In this proposed model, the β -strand containing the *WRKYGQK* motif makes contacts with a ~ 6-bp region, which is largely consistent with the length of the (C/T)TGAC(T/C) Wbox consensus, to which most of the *WRKY* transcription factors bind motif makes contacts with a ~ 6-bp region, which is largely consistent with the length of the with different affinity (C/T)TGAC(T/C) Wbox consensus, to which most of the *WRKY* transcription factors bind (Eulgem et al., 2000, Yu et al., 2001; Dong et al., 2003; Miao et al., 2004; Xu et al., 2006). In parsley, it was shown that clustering of W-boxes is important for a strong transcriptional response (Eulgem et al., 1999; Rushton et al., 1996). Based on their domain structure, *WRKY* proteins can be divided into three major groups. Group I *WRKY* proteins contain two *WRKY* domains: Group II *WRKY* proteins only have one *WRKY* domain and a C₂H₂ zinc finger motif; *WRKY* group III proteins contain one *WRKY* domain and C₂HC zinc finger motif (Xie et al., 2005; Ross et al., 2007; Jang et al., 2010). Besides the consensus W-box, *WRKY* factors have been identified to bind to other motifs. Recently, the tobacco *WRKY* protein with a variant to the *WRKYGQK* (*WRKYGKK*) sequence (Van Verk et al., 2008) interacts with WK-boxes, (TTTTCCAC) and is unable to bind to the consensus W-box (Van Verk et al., 2008). A *WRKY* protein from barley (SUSIBA) was found to bind to SURE, a Sugar-

Responsive Cis Element in the promoter of the Isoamylase 1 (*ISO1*) gene (Sun et al., 2003). These WRKY factors are suggested to amplify their expression levels via multiple WRKY binding sites in their own promoters; thereby, creating a positive feedback loop.

WRKY proteins network is regulated by several mechanisms, including auto-regulation by WRKY genes, regulation by transcription factors and other proteins, and also by smRNA-WRKY regulatory interactions (Pandey and Somssich, 2009). Studies using knockout or knockdown mutants or overexpression lines of WRKY genes have shown that WRKY transcription factors can positively or negatively regulate various aspects of plant PTI and ETI (involved in the regulation of plant growth and developmental processes including trichome (Johnson et al., 2002) and seed development (Luo et al., 2005), germination (Jiang and Yu, 2009), leaf senescence (Robatzek and Somssich, 2002; Miao et al., 2004), wounding (Hara et al., 2000), development (Johnson et al., 2002), dormancy and drought tolerance (Pnueli et al., 2002), solar ultraviolet-B radiation (Izaguirre et al., 2003), metabolism (Sun et al., 2003; Xu et al., 2004), hormone signalling pathway (Zhang et al., 2004; Xie et al., 2005) and cold (Mare et al., 2004). In addition, there are several evidences of their involvement in response to biotic stresses (Kim et al., 2000; Du and Chen, 2000; Chen and Chen, 2002; Kalde et al., 2003; Turck et al., 2004; Ryu et al., 2006).

Our group has been predominantly involved in the study of disease resistance in rice. Globally, researchers are now looking into the function of transcription regulators in the mediation of stress responses. Our interests lies in the involvement of transcription factors in regulation of disease response in rice and hope to understand their role in rice defense mechanism through the isolation and later functional characterization of this gene. Here, we report the isolation and *in silico* characterization of the WRKY gene from UKMRC9, a rice variety that shows resistance to rice blast, and moderate level of resistance to sheath blight (to date no resistant cultivars have been reported worldwide). As the gene showed high level of sequence variation at nucleotide level, the motif, phylogenics and structural analysis conducted here has provided some insight into the possible function of this gene and; thus, provided us with some directions in future functional analysis. To date, it is not clear how this gene is regulated in rice though extensive work has been done on the *Arabidopsis* WRKY gene.

Results and Discussion

UKMRC9 WRKY protein sequence

The 1396 bp nucleotide sequence of UKMRC9 WRKY gene (Accession number: KF312463) was translated and is shown in Fig. 1. Results of the SignalP 4.0 analysis demonstrated that there is no signal peptide present in this protein (Supplementary Fig. 2a). According to Janda et al. (2010), signaling peptides will not be present in mature protein or in secretory proteins. The absence of putative signal peptide sequence indicates that it is likely that this protein will remain intracellular (Krishna et al., 2011). The S score is the predicted signal peptide for each amino acid in the protein sequence. High score indicates that the amino acid is part of a signal peptide, while a low score implies that the amino acids compose part of a mature protein. The C score is the cleavage site score that denotes the location for the restriction activity

of the peptidase I enzymes. Peaked activity in C score indicates the presence of cleavage sites which is not present here. The Y scores are determined by combining the S and C scores. The peptidase I cleavage site is determined from the Y score, where the S score is steep and C score is significant. Here, we observe that the C, Y and S scores are low and; therefore, the protein is unlikely to be a secretory protein. Membrane proteins and secretory proteins usually contain signal peptides that are indicated by the presence of a cleavage activity (Janda et al., 2010).

Forecasting trans-membrane segment

The trans-membrane segment analyses of UKMRC9 WRKY protein via TMHMM2.0 (<http://www.cbs.dtu.dk/services/TMHMM-2.0/>) showed that the protein is globular and is not located within the trans-membrane region (Supplementary Fig. 2b). It is in accordance with the finding of Duan et al. (2007) where the AtWRKY₁-C showed a globular structure. This is consistent with the function of this protein in defense; where it has been implicated to play a role in the signal cascade involved in activating the defense response within the host (Jang et al., 2010). An increasing number of studies in plants such as rice strongly confirm the importance of WRKY transcription factors in plant defense signaling (Pandey and Somssich, 2009).

Multiple alignment of the WRKY protein in rice and other plant species

We classified our UKMRC9 WRKY transcription factors on the basis of the predicted amino acid sequences of the WRKY domains. From the NCBI Protein Database, 25 WRKY protein sequences from rice and other plant species were retrieved (Supplementary Fig. 1) and subjected to a multiple alignment analysis using T-coffee. Fig. 2 provides the multiple alignment of various plant species (ABB90551.1 *Triticumaestivum*, ABN43185.1 *Triticumaestivum*, ABN43178.1 *Triticumaestivum*, NP_001120723.1 *Zea mays*, ACB12743.1 *Zea mays*, NP_001147551.1 *Zea mays*, AF459793_1 *Oryza sativa Indica* Group, AAT85791.1 *Oryza sativa Japonica* Group, AAT85791.1 *Oryza sativa Japonica* Group, BAC01237.1 *Oryza sativa Japonica* Group, NP_567862.3 *Arabidopsis thaliana*, AEE73605.1 *Arabidopsis thaliana*, AED91109.1 *Arabidopsis thaliana*, AEC05689.1 *Arabidopsis thaliana*, XP_004137798.1 *Cucumis sativus*, XP_004137798.1 *Cucumis sativus*, CCV01211.1 *Cucumis sativus*, ADU52526.1 *Cucumis sativus*, XP_003625561.1 *Medicago truncatula*, XP_003556710.1 *Glycine max*, BAB61055.1 *Nicotiana tabacum*, AAR26657.1 *Capsicum annum*, ABR87003.1 *Hordeum vulgare*, XP_004514245.1 *Cicerarietinum*). From the multiple alignment of WRKY proteins sequences, we observed that the protein sequences were divergent at both the C and N terminus of the protein. On further observation, this divergence was also observed in the C and N terminus of the WRKY domain (Kato et al., 2007). A lower level of divergence was observed amongst similar genus but the difference between the terminals was clearly dissimilar between species. In Fig. 2, several regions of high homology were detected amongst all twenty five aligned sequences of monocot and dicot plant species. These areas of high homology were confined to the WRKY DNA

Table 1. Clustering of WRKY proteins according to the WRKY domains.

Group	OsWRKY	HvWRKY	TaWRKY	ZmWRKY	GmaxWRKY	PpWRKY	CsWRKY	MtWRKY
1a & b	14 ¹	4	8	3	1	1	1	1
2a	5	4	1	- ¹	-	-	-	-
2b	7	1	4	-	-	-	-	-
2c	15	5	4	-	-	3	-	-
2d	6	5	3	-	-	1	-	-
2e	8	1	2	-	-	-	-	-
3	9	7	-	-	-	-	-	-
Total	64	27	22	3	1	5	1	1

¹Numbers represent numbers of WRKY proteins that were assigned to specific group.

²Dash (-) represents none of WRKY proteins that was assigned to specific group.

```

MDGARQESRE YWRDGGDVVG EELLREILEE TAAVHSNSNS NSNSNSNSKE AEEEDEREYF 60
AAAAADEQLQ VEAPCGRRRR ESMVNKLIST VYSGPTISDI ESALSFTAAG DHQLLADGHN 120
FAASSCSPVV FSPEKTLSTK MENKYTLKIV SCGNDGGLAD DGYKWRKYGQ RSIKNSPNPR 180
SYRCTNPRC NSKKQVERAV DEPDTLIVTY EGLHLHYTYS HFLHSTSSSS SSTTTQQQLQ 240
PQPQMTTNC KKPKLHLHPL LHDDPRPPPP PPEMTTSQIM QSFISIQQQH DDDQLLQFAA 300
DDHLMVQAPP DDCYNINGSS SSGLMSSLED DEQAAGAGGL LEDVVRLVLR RPPPPICNNN 360
NYYSPATTC TSDNEYGSSA SASPSSSVSV SSWTTPMSPC IDMAILSNIF* 410

```

Fig 1. The protein sequence WRKY in UKMRC9 rice variety, consisted of 410 amino acids. The asterisk indicates the location of the stop codon.

binding domain and the zinc finger binding domain and the regions surrounding these domains. Fig. 3 shows the location of the highly conserved WRKY DNA binding domain and the zinc finger binding domain of this protein. The differences in the C and N terminal of these proteins are largely responsible for the variation in function and level of expression modulated by this group of transcriptional factor (Lee et al., 2013).

As seen in Fig. 3, the WRKY protein is characterized by an ~ 60 amino acid WRKY motif including a zinc finger like motif C-X4-7-C-X23-28-H-X1-2-(H/C) that provides the binding properties to DNA together with a very highly conserved WRKYGQK sequence that gives binding specificity and the name to the family. The presence of these residues enables the protein to deeply enter into the major groove of the DNA at the lateral chains of Arg (R), Lys (K), Gln (Q), and Lys (K), which are likely the most important residues in sequence specific recognition of the domain (Eulgem et al., 2000).

Analysis of physico-chemical parameters of UKMRC9 WRKY protein

The ProtParam analyses determined the molecular weight of the protein as 45.1 kDa and isoelectric point (pI) value of 4.91. This result will be important for developing the buffer systems for purification of proteins by isoelectric focusing and two-dimensional electrophoresis (Krishna et al., 2011). The amino acid composition in UKMRC9 WRKY protein had a high content of serine (12.7%). The predicted protein consists of 35.8% hydrophobic amino acids, 40.5% polar amino acids and 23.6% charged amino acids. These findings support the results obtained from the TMHMM2.0 analyses, where an outer membrane protein is soluble with low hydrophobicity and high polarization. The hydrophobicity stabilizes the interaction between the conserved W and Y of the WRKY domain; where it facilitates the chelation of zinc ions by the histidine and cysteine residues (Ulker and

Somssich, 2004). The molecule's apolar (hydrophobic) amino acids bind towards the molecule's interior, whereas the polar (hydrophilic) amino acids are bound outwards, allowing dipole-dipole interaction that explains the molecule's solubility. The positive and negative charge of the amino acid constituents in this protein indicates that it has a net negative charge (21 negative residues more than positive). The numbers of atoms in these protein molecules are 6157 with 1930 carbon atoms, 3004 hydrogen atoms, 544 oxygen atoms, 653 nitrogen atoms and 26 sulphur atoms. Instability Index of 59.53 shows that this globular protein is only marginally stable as the free energy released when the protein is folded into its native conformation is relatively small; thus, providing a rapid turnover rate and effective control of protein degradation and synthesis.

Phylogenetic analysis of UKMRC9 WRKY protein

The unrooted phylogenetic tree of the WRKY protein isolated from UKMRC9 was conducted together with 63 *O. sativa*, 19 *A. thaliana*, 28 *H. vulgare*, 3 *Z. mays*, 3 *T. aestivum*, 1 *G. max*, 1 *M. trunculata* and 1 *C. sativus* WRKY proteins. We conducted bootstrap analysis with 1000 replicates. The bootstrap values are shown at the internal nodes. In general, the phylogenetic analysis showed that the WRKY proteins are monophyletic while their assignment into the various groups is indicative of a polyphyletic nature of the WRKY domains and/or the zinc finger motif. As the complete gene sequences of each WRKY protein exhibited low levels of identity and similarity even within the same genus and species, we further narrowed the region of analysis to only include the signature WRKY domains of these proteins. The separation of the WRKY proteins into subgroups was done as classified according to Xie et al. (2005). For the purpose of analysis, representative protein sequences of each group were taken as presented in Xie et al. (2005).

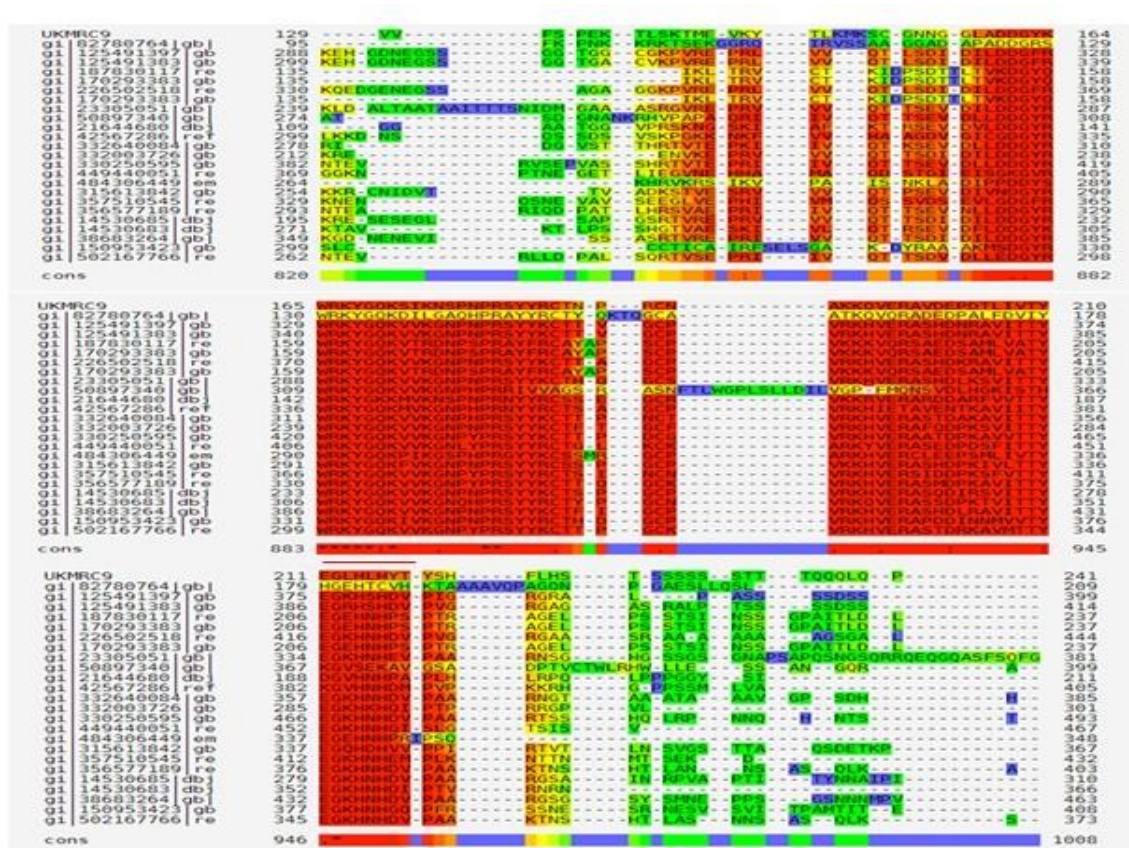


Fig 2. Multiple-alignment of rice WRKY proteins and other plant WRKY proteins. The WRKY protein sequences of *ABB90551.1* *Triticum aestivum*, *ABN43185.1* *Triticum aestivum*, *ABN43178.1* *Triticum aestivum*, *NP_001120723.1* *Zea mays*, *ACB12743.1* *Zea mays*, *NP_001147551.1* *Zea mays*, *AF459793_1* *Oryza sativa Indica Group*, *AAT85791.1* *Oryza sativa Japonica Group*, *AAT85791.1* *Oryza sativa Japonica Group*, *BAC01237.1* *Oryza sativa Japonica Group* *NP_567862.3* *Arabidopsis thaliana*, *AEE73605.1* *Arabidopsis thaliana*, *AED91109.1* *Arabidopsis thaliana*, *AEC05689.1* *Arabidopsis thaliana*, *XP_004137798.1* *Cucumis sativus*, *XP_004137798.1* *Cucumis sativus*, *CCV01211.1* *Cucumis sativus*, *ADU52526.1* *Cucumis sativus*, *XP_003625561.1* *Medicago truncatula*, *XP_003556710.1* *Glycine max*, *BAB61055.1* *Nicotiana tabacum*, *AAR26657.1* *Capsicum annum*, *ABR87003.1* *Hordeum vulgare*, and *XP_004514245.1* *Cicera rietinum* were downloaded from TIGR, NCBI PlantDB. A 60 amino acid region that is highly conserved is presented in red in the figure above. The C and N terminus of the WRKY protein have been omitted from this display as they show low level identity as well as similarity within sequences.

In order to obtain a better resolution of the groups and subgroups (group 1, 2a-2e and 3) of the WRKY protein, sequences of the WRKY domain of *Arabidopsis*, *O. sativa*, *H. vulgare*, *Z. mays* and other plant species were included in the analysis. A representative *A. thaliana* and *H. vulgare* WRKY domain sequence were randomly chosen from each group and subgroup for inclusion in the phylogenetic analysis. As shown in Fig. 4, we observed that the WRKY proteins of each species have been clustered according to their WRKY motif. According to Wu et al. (2004), members belonging to the same group were invariably clustered together. It was reported by Ulker and Somssich (2004) that group I WRKY genes may represent the ancestral form of WRKY genes. In the phylogenetic alignment, we found that the C terminus of certain Group 1 WRKY proteins was clustered closer to Group 2 proteins. Hence, we subgrouped them into Group 1b. The Group 2 and 3 domains were closer related to the C terminal of Group 1b WRKY domains while the N-terminal of the domain formed a separate monophyletic cluster in Group 1 and may support the observations made by others that the monophyletic nature of the Group 1 may be specific for grasses (Fig. 4) (Zhang and Wang, 2005). From the Maximum Likelihood tree that has been constructed, the

presence of subclades is indicative of the paraphyletic nature of the subgroups (Fig. 4). The UKMRC9 WRKY protein has been clustered into group 2e which indicates that this protein has 1 WRKY and 2C and 2H zinc finger domain. The clustering of UKMRC9 WRKY protein into the same subgroup with *A. thaliana* and *H. vulgaris* indicates that there were variations among the subgroups, showing the divergence of the WRKY domains. The variations found in WRKY domains of rice and other species suggested that they may have evolved slightly differently after their divergence (Wu et al., 2004). Table 1 has the assigned WRKY proteins within each subgroup 1 to 3 for all the plant species analyzed in this phylogenetic analysis.

Structure prediction of UKMRC9 WRKY protein

Based on the analysis of the structure derived from Cn3D protein modeling, the QMEAN4 is 0.72. QMEAN4 scores range from 0 to 1, where the higher the score, the more stable the forecasted model. The WRKY protein consists of four-strands of β -sheet, with the zinc ion bound by two Cysteine

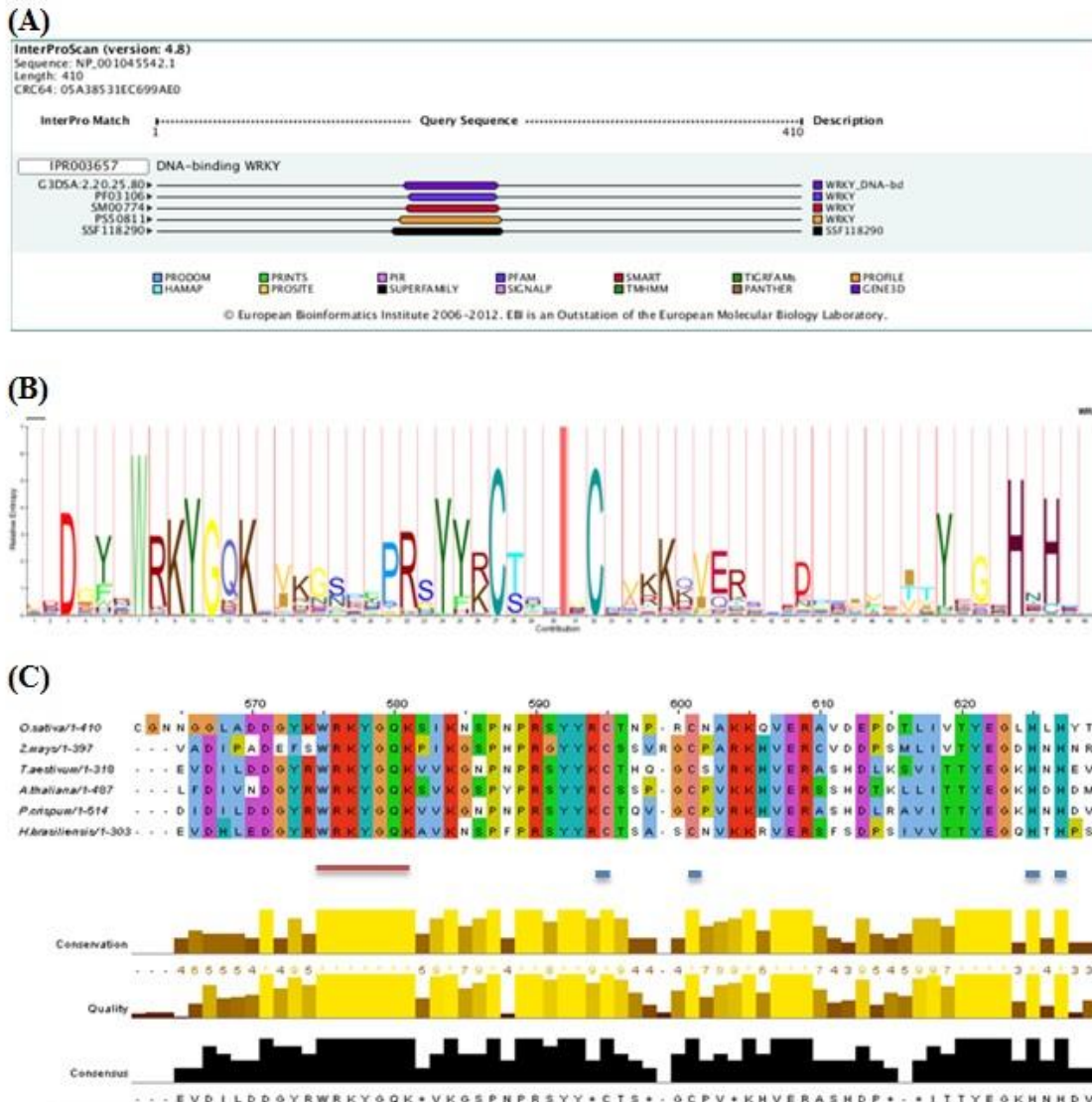


Fig 3. Conserved domain and motif search conducted on the UKMRC9 WRKY Gene. (A) InterProScan results indicating the location of WRKY DNA binding domain. (B) Graphical representation of pHHMs of the existing domains. Each letter indicates deviation of position amino acid frequencies from the background frequencies. (C) The T-Coffee alignment showing the presence of the WRKY (red line) and the Zinc Finger Binding Domain (blue lines indicating the C₂H₂ site).

and two Histidine residues forming a zinc-binding pocket (Fig. 5). The WRKYGQK residues correspond to the most N-terminal β -strand, which partly protrudes from one surface of the protein; thereby, enabling access to the major DNA groove and facilitating contact with the DNA. It was proposed that the β -strand containing the WRKYGQK motif makes contact with an approximately 6-bp region, and this is consistent with the length of the W box, which is the binding site of most known WRKY transcription factors. It has been hypothesized that the conservation of the WRKY domain in WRKY proteins is mirrored by a remarkable conservation of the W box (TTGACC/T). Gel shift experiments, random binding site selection, yeast one-hybrid screens and co-transfection assays performed with many different WRKY proteins have shown that the W box is the minimal consensus required for specific DNA binding (Chi et al., 2013; Jang et al., 2010; Ulker and Somssich, 2004).

Materials and Methods

Source of plant material

The four rice cultivars selected for this study are as follows: *Oryza rufipogon* cultivar Griff (IRGC105491), MR219 (Malaysian cultivated rice variety), and UKMRC2 and UKMRC9 (two lines developed through advance backcrosses technique between MR219 and *O. rufipogon*).

Fungal culture

The *Rhizoctonia solani* strain1802 that was isolated from the rice fields in Kota Bharu Kelantan was used as the fungal culture in the screening experiments. The fungal isolate was sub-cultured on potato dextrose agar (PDA) plates and stock cultures of this isolate were maintained on PDA slants.

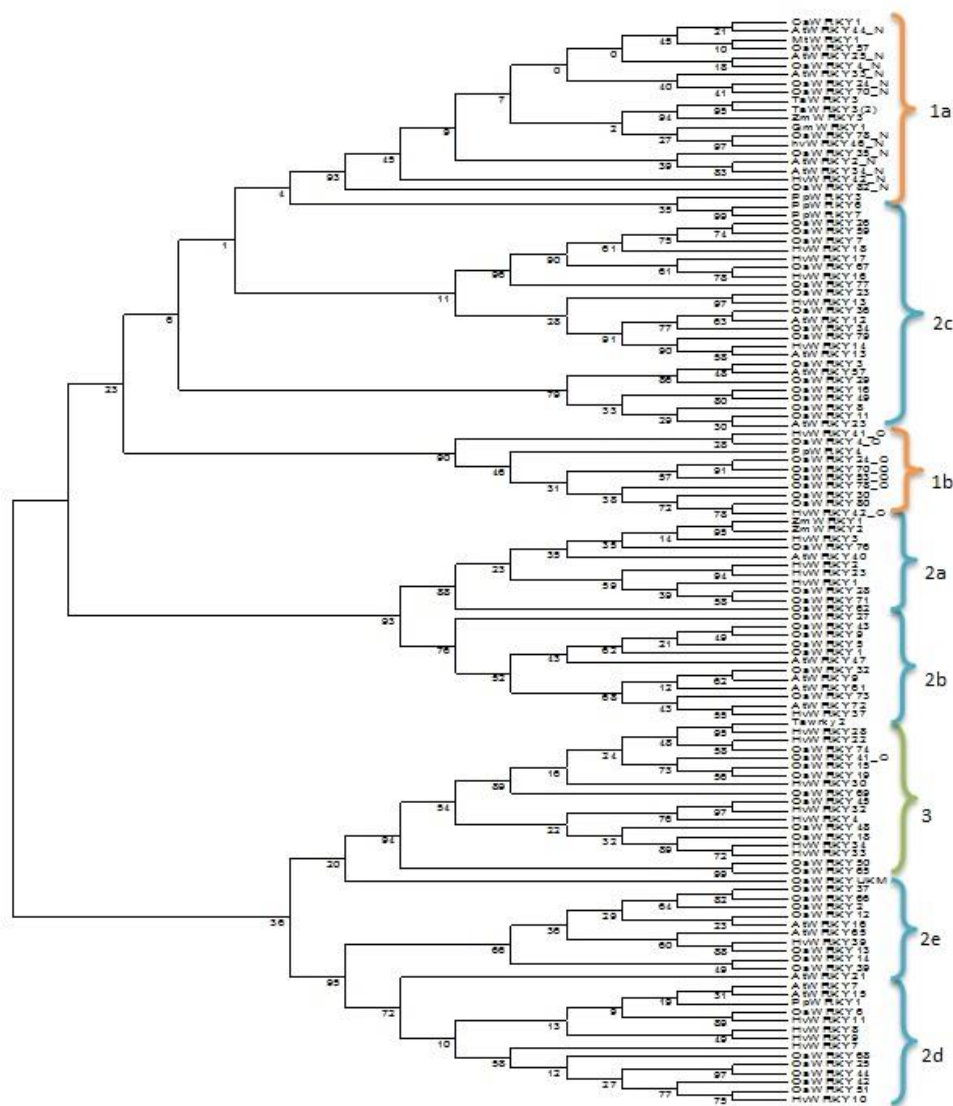


Fig 4. Maximum likelihood tree constructed with WRKY proteins from rice and other plant species. The unrooted tree was calculated on the basis of WRKY domain sequences of the rice (OsWRKY) and selected Arabidopsis (AtWRKY), *H. vulgare* (HvWRKY), *Z. mays* (ZmWRKY), *G. maxs* (GmWRKY), and a few other plant WRKY protein sequences as representatives for the different groups. WRKY groups and sub-groups 1 to 3 are highlighted. The phylogenetic analysis was conducted using the WRKY domain consensus sequences of all samples.

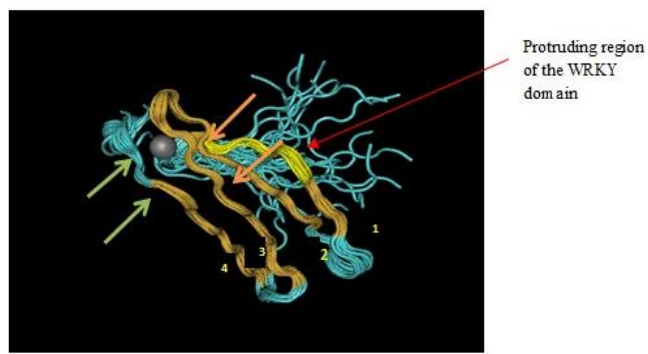


Fig. 5 The 3D structure of UKMRC9 WRKY protein. The protein structure was predicted using the Cn3D software (NCBI). The orange arrows indicate the location for the Cys residues (Cys185 and 190) and the green arrows indicate the location of the His residues (His 214 and 216). The grey sphere is the zinc ion held in a pocket by the Cys/His residues. The yellow region in the 3D structure indicates the location of the WRKY motif. The *A. thaliana* WRKY protein (PDB ID: 2AYD) was used as a reference sequence to obtain the above structure.

Cultivation of plant materials

The four rice varieties; *Oryza rufipogon*, MR219, UKMRC2 and UKMRC9 were grown in 6 inch pots containing peat and clay soil in uncontrolled greenhouse conditions. A total of 20 seeds for each rice variety were planted and allowed to grow for 5-6 weeks before these plants were used for the disease severity test.

Preparation of fungal inoculum

R. solani cultures were sub-cultured on PDA and left for 5 days at room temperature in dark for optimal growth. Mycelial plugs were prepared from the peripheral zones of the fungal mycelium cultures. A 5 mm fungal plug was placed at the sheath base of 5-6 weeks old plants (Park et al., 2008). A total of 8 rice plants for each variety were inoculated with *R. solani* plugs. The inoculated plants were maintained at high humidity conditions and suitable temperatures to facilitate infection.

Screening and evaluation of disease severity index (DSI)

Plants that were inoculated with *R. solani* strain 1802 were observed for duration of 10 days post inoculation (DPI). All varieties of *Oryza* were observed and the disease severity index (dsi) recorded in triplicates. The standard DSI scale of 1-5 was used, in which the scale 1 assigned to the plants that show a high level of resistance and level 5 for highly susceptible plants that die post infection within the duration of observation, respectively (Park et al., 2008). We screened the lines and the most tolerant variety was used for RNA extraction.

Total RNA extraction

Young infected rice leaves of a low disease severity index (DSI) were harvested and weighed to 100 milligrams. Total RNA was extracted from these samples using RNeasy Plant Mini Kit (Qiagen). The protocol for extraction was done as recommended by manufacturer with no modifications.

Primer design for RT-PCR

RNA sequence of *WRKY1 Oryza sativa* (AF193802.2) derived from the National Center for Biotechnology Information (NCBI) (www.ncbi.nlm.nih.gov/) was used in designing the primer pairs for amplification of this gene. The RNA sequence was fed into the ORF Finder software (<http://www.ncbi.nlm.nih.gov/gorf/orfig.cgi>) to determine the position of the start and stop codons. The ORF Finder is a graphical analysis tool from NCBI that can provide the putative open reading frame of an input sequence. The Primer3 Program (<http://frodo.wi.mit.edu/primer3/>) was used to design the forward and reverse primers. The primers synthesized from this analysis are: 5'-forward GGAAGAATTAAGCAAGTAGCAGC-3' and 5'-reverse AGCTAGCTAGAGAATACTAGATGAA -3'.

RT-PCR of UKMRC9 WRKY gene

The RT-PCR method was employed to isolate the *WRKY* gene from UKMRC9. The RevertAid First Strand cDNA Synthesis Kit and GoTaq[®] Flexi DNA Polymerase (Promega, USA) were used in the amplification process. Samples were prepared for the Reverse Transcription process as recommended by the manufacturer of RevertAid First Strand

cDNA Synthesis Kit. The PCR cycle was conducted as follows: a pre-denaturation step at 94 ° C for 5 mins. This was followed by 35 cycles of amplification at 94 ° C for 2 mins, annealing at 60° C for 30 s, elongation at 72 ° C for 2 mins and a final elongation cycle at 72 ° C for 5 mins.

UKMRC9 WRKY gene sequence

The RT-PCR product was sequenced using the ABI sequencer PRISM 377 (version 3.4.1). To confirm the identity of the putative *WRKY* gene, the gene sequence was analyzed in ChromasLite (http://www.technelysium.com.au/chromas_lite.html) and Bioedit (<http://www.mbio.ncsu.edu/BioEdit/bioedit.html>) to obtain a consensus sequence. The consensus sequence was then blasted (using the BLAST system at <http://blast.ncbi.nlm.nih.gov/Blast.cgi>) against the records in NCBI to other *WRKY* genes (www.ncbi.nlm.nih.gov/). Once the identity of the gene was determined, the following bioinformatics analysis ensued.

Molecular characterization of the UKMRC9 WRKY gene

The UKMRC9 *WRKY* gene was sequenced and translated through the BLASTx and BLASTp (<http://blast.ncbi.nlm.nih.gov/Blast.cgi>) to obtain the UKMRC9 *WRKY* protein.

The protein was then compared with the *Oryza sativa* *WRKY1* protein sequence in UniProt (<http://www.uniprot.org/>), which is a data repository for protein of database Swiss-Prot, TrEMBL and PIR-PSD. The amino acid sequence of the *WRKY* protein was analyzed via InterProScan (<http://www.ebi.ac.uk/Tools/pfa/iprscan/>) to identify the signature *WRKY* domains. InterProScan analyzes a protein sequence with respect to characteristics of the proteins against its 12 protein and nucleotide databases which includes Pfam, PROSITE, UniProt, SMART etc. (Zdobnov and Apweiler, 2001). The amino acid sequence was then subjected to signal peptide determination through the SignalP 4.0 program (<http://www.cbs.dtu.dk/services/SignalP/>).

For the analysis of the physico-chemical parameters of proteins, sequences were analyzed via ProtParam (<http://web.expasy.org/protparam/>) from ExPASy, (Swiss Institute of Bioinformatics) to obtain information such as the molecular weight of the protein, the net protein charge, and the protein stability index (Gasteiger et al., 2005). The prediction of transmembrane segments was performed using TMHMM v2.0 (<http://swissmodel.expasy.org/>), which is an acronym for "Transmembrane Hidden Markov Models". This analysis was conducted to determine if the *WRKY* protein is hydrophilic or hydrophobic and the localization of the protein i.e. inside or outside the cell. TMHMM Server is the premier programs in the prediction of transmembrane domains (Moller, Croning and Apweiler, 2001).

The presence of signal peptides for the *WRKY* protein was determined via the SignalP 4.0 Program (Petersen et al., 2011). This is done to detect if the protein studied is mature through the discrimination of the signal peptide with the transmembrane region. Multiple sequence alignment was performed by T-Coffee (<http://tcoffee.org.cat/>) (Notredame, Higgins and Heringa, 2000). The phylogenetic analysis by maximum likelihood was constructed with MEGA 5.10 software. In order to construct the phylogeny, the pairwise alignment parameters were set with gap opening penalty 10; gap extension penalty 1.0; protein weight matrix BLOSUM 30 (Henikoff and Henikoff, 1992). The number of bootstrap replications was set as 1000. The tree was unrooted.

A 3D structure prediction of the UKMRC9 WRKY protein was obtained using the Cn3D protein modeling software (<http://www.ncbi.nlm.nih.gov/Structure/CN3D/cn3dwin.html>). This 3D modeling software was built based on homology modeling of protein structure with other similar proteins available online (Arnold et al., 2006). The model quality was analyzed according to the QMEAN4 score (<http://swissmodel.expasy.org/>).

Conclusion

The *WRKY* gene continues to be studied extensively to date due to the plethora of functions it regulates in biotic and abiotic stress modulation. Our research team has isolated various transcriptional and receptor proteins to study the interactions involved in the regulation of stresses. The *WRKY* gene reported here was isolated and characterized *in silico*. The full-length gene was isolated from a rice cultivar that has shown resistance and or tolerance to rice pathogens. The basic information of the protein's physico-chemical properties, structural and evolutionary relatedness to other similar proteins forms the basis of our future research directions. As a consequence of the analysis conducted here, we hope to functionally analyze and decipher the role played by the high level of sequence divergence at the terminal region of this gene and determine its role in influencing *WRKY* function in biotic and abiotic stresses in rice.

Acknowledgement

This research was supported by funds from the Ministry of Higher Education Malaysia (grant no: LRGS/TD/2011/UPM-UKM/KM/01) and Ministry of Science Technology and Innovation Malaysia (grant no: 02-01-02-SF0757). We would also like to thank Prof. Dr. Wickneswari Ratnam for providing the UMKRC2 and UKMRC9 rice variety seeds for this research.

References

Arnold K (2006) The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. *Bioinformatics*. 22(2):195-201.

Chen C and Chen Z (2002) Potentiation of developmentally regulated plant defense response by AtWRKY18, a pathogen-induced *Arabidopsis* transcription factor. *Plant Physiol*. 129, 706-716.

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

Dong J, Chen C and Chen Z (2003) Expression profiles of the *Arabidopsis* WRKY gene superfamily during plant defense responses. *Plant Cell Physiol*. 51: 21-37.

Du L and Chen Z (2000) Identification of genes encoding receptor-like protein kinases as possible targets of pathogen- and salicylic acid-induced WRKY DNA-binding proteins in *Arabidopsis*. *Plant J*. 24: 837-847.

Duan MR, Nan J, Liang YH, Mao P, Lu L, Li L, Wei C, Lai L, Li Y, Su XD (2007) DNA binding mechanism revealed by high resolution crystal structure of *Arabidopsis thaliana* WRKY1 protein. *Nucleic Acids Res.*, 35(4):1145-54.

Eulgem T, Rushton PJ, Schmelzer E, Hahlbrock K and Somssich IE (1999) Early nuclear events in plant defence signalling: Rapid gene activation by WRKY transcription factors. *EMBO J*. 18: 4689-4699.

Eulgem T, Rushton PJ, Robatzek S, Somssich IE (2000) The WRKY superfamily of plant transcription factors. *Trends Plant Sci*. 5(5):199-206.

Hara K, Yagi M, Kusano T and Sano H (2000) Rapid systemic accumulation of transcripts encoding a tobacco WRKY transcription factor upon wounding. *Mol. Gen. Genet*. 263, 30-37.

Henikoff S and Henikoff J G (1992) Amino acid substitution matrices from protein blocks. *Proc Natl Acad Sci USA*. 89:10915-10919.

Ishiguro S and 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 beta-amylase from sweet potato. *Mol Genet Genomics*. 244(6):563-71.

Izaguirre MM, Scopel AL, Baldwin IT, Ballaré CL (2003) Convergent responses to stress. Solar ultraviolet-B radiation and Manduca sexta herbivory elicit overlapping transcriptional responses in field-grown plants of *Nicotiana longiflora*. *Plant Physiol*. 132: 1755-1767.

Janda CY, Li J, Oubridge C, Hernández H, Robinson CV, Nagai K (2010) Recognition of a signal peptide by the signal recognition particle. *Nature*. 465 (7297):507-510.

Jang JY, Choi CH, Hwang DJ (2010) The WRKY superfamily of rice transcription factors. *Plant Pathol J*. 26: 110-114.

Jiang W, Yu D (2009) Arabidopsis WRKY2 transcription factor mediates seed germination and post germination arrest of development by abscisic acid. *BMC Plant Biol* 9(1): p. 96.

Johnson CS, Kolevski B and Smyth DR (2002a) Transparent TestaGLABRA2, a trichome and seed coat development gene of Arabidopsis, encodes a WRKY transcription factor. *Plant Cell*. 14, 1359-1375.

Kalde M, Barth M, Somssich IE and Lippok B (2003) Members of the *Arabidopsis* WRKY group III transcription factors are part of different plant defense signaling pathways. *Mol Plant-Microbe Interact*. 16: 295-305.

Kato N, Dubouzet E, Kokabu Y, Yoshida S, Taniguchi Y, Dubouzet JG, Yazaki K and Sato F (2007) Identification of a WRKY protein as a transcriptional regulator of benzyloisoquinoline alkaloid biosynthesis in *Coptis japonica*. *Plant Cell Physiol* 48: 8-18.

Kim CY, Lee SH, Park HC, Bae CG, Cheong YH, Choi YJ et al (2000) Identification of rice blast fungal elicitor-responsive genes by differential display analysis. *Mol Plant Microbe Interact*. 13, 470-474.

Krishna R, Muthukumaran J, Manivel P, Kannan M and Jeyakanthan J (2011) A framework for classification of antifreeze proteins in over wintering plants based on their sequences and structural features. *J Bioinform Seq Anal*. 3(4): 70-88.

Luo M, Dennis ES, Berger F, Peacock WJ, Chaudhury A (2005) MINISEED3 (MINI3), a WRKY family gene, and HAIKU2 (IKU2), a leucine-rich repeat (LRR) KINASE gene, are regulators of seed size in Arabidopsis. *Proc Natl Acad Sci USA*. 102:17531-17536.

Mare C, Mazzucotelli E, Crosatti C, Francia E, Stanca AM and Cattivelli L (2004) Hv-WRKY38: a new transcription factor involved in cold- and drought- responds in barley. *Plant Mol Biol*. 55: 399-416.

Miao Y, Laun T, Zimmermann P, Zentgraf U (2004) Targets of the WRKY53 transcription factor and its role during leaf senescence in Arabidopsis. *Plant Mol Biol*. 55: 853-867.

Moller S, Croning MDR and Apweiler R (2001) Evaluation of methods for the prediction of membrane spanning regions. *Bioinformatics*. 17(7):646-653.

- Notredame C, Higgins DG, Heringa J (2000) T-coffee: a novel method for fast and accurate multiple sequence alignment. *J Mol Biol.* 302:205-217.
- Pandey SP and Somssich IE (2009) The role of WRKY transcription factors in plant immunity. *Plant Physiol.* 150: 1648-1655.
- Petersen TN (2011) SignalP 4.0: Discriminating signal peptides from transmembrane regions. *Nat Methods.* 8:785-786.
- Pnueli L, Hallak HE, Rozenberg M, Cohen M, Goloubinoff P, Kaplan A and Mittler R (2002) Molecular and biochemical mechanisms associated with dormancy and drought tolerance in the desert legume *Retamaraetam*. *Plant J.* 31: 319- 330.
- Riechmann JL, Heard J, Martin G, Reuber L, Jiang C, Keddle J, Adam L, Pineda O, Ratcliffe OJ, Samaha RR et al (2000) *Arabidopsis* transcription factors: genome-wide comparative analysis among eukaryotes. *Science.* 290: 2105-2110.
- Robatzek S, Somssich IE (2002) Targets of AtWRKY6 regulation during plant senescence and pathogen defense. *Genes Dev.* 16: 1139–1149
- Ross CA, Liu Y, Shen QJ (2007) The WRKY gene family in rice (*Oryza sativa*). *J Integr Plant Biol.* 49: 827–842.
- Rushton PJ, Torres JT, Parniske M, Wernert P, Hahlbrock K and Somssich IE (1996) Interaction of elicitor-induced DNA-binding proteins with elicitor response elements in the promoters of parsley PR1 genes. *EMBO J.* 15: 5690–5700.
- Ryu HS, Han M, Lee SK, Cho JI, Ryoo N, Heu S, Lee YH, Bhoo S, Wang GL, Hahn TR, Jeon JS (2006) A comprehensive expression analysis of the WRKY gene superfamily in rice plants during defense response. *Plant Cell Rep.* 25: 836-847.
- Sun C, Palmqvist S, Olsson H, Borén M, Ahlandsberg S, Jansson C (2003) A novel WRKY transcription factor, SUSIBA2, participates in sugar signaling in barley by binding to the sugar-responsive elements of the iso1 promoter. *Plant Cell.* 15: 2076–2092.
- Turck F, Zhou A, Somssich IE (2004) Stimulus- dependent, promoter- specific binding of transcription factor WRKY1 to its native promoter and the defense- related gene PcPR1-1 in parsley. *Plant Cell.* 16: 2573-2585.
- van Verk MC, Pappaioannou D, Neeleman L, Bol JF, Linthorst HJM (2008) A novel WRKY transcription factor is required for induction of PR-1A gene expression by salicylic acid and bacterial elicitors. *Plant Physiol.* 146: 1983–1995.
- Wu KL, Guo ZJ, Wang HH and Li J (2004) The WRKY family of transcription factors in rice and *Arabidopsis* and their origins. *DNA Res.* 12, 9-26.
- Ülker B, Somssich IE (2004) WRKY transcription factors: from DNA binding towards biological function. *Curr Opin Plant Biol.* 7: 491–498.
- Xie Z, Zhang ZL, Zou XL, 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 Physiol.* 137(1):176-189.
- Yamasaki K, Kigawa T, Inoue M, Tateno M, Yamasaki T, Yabuki T, Aoki M, Seki E, Matsuda T, Tomo Y et al (2005) Solution structure of an *Arabidopsis* WRKY DNA binding domain. *Plant Cell.* 17: 944-956.
- Yu D, Chen C and Chen Z (2001) Evidence for an important role of WRKY DNA binding proteins in the regulation of NPR1 gene expression. *Plant Cell* 13: 1527-1539.
- Zhang ZL, Xie Z, Zou X, Casaretto J, David TH and Zhen QJ (2004) A rice WRKY gene from creosote bush encodes a transcriptional repressor of gibberellin signaling pathway in aleurone cells. *Plant Physiol.* 134: 1500-1513.
- Zhang YJ and Wang LJ (2005) The WRKY transcription factor superfamily: its origin in eukaryotes and expansion in plants. *BMC Evol Biol.* 5:1.