

Variations in the Structure and Evolution of Rice *WRKY* Genes in *Indica* and *Japonica* Genotypes and their Co-expression Network in Mediating Disease Resistance

Evolutionary Bioinformatics
Volume 15: 1–12
© The Author(s) 2019
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/1176934319857720



John Lilly Jimmy and Subramanian Babu

School of Bio Sciences and Technology, Vellore Institute of Technology, Vellore, India

ABSTRACT: WRKY transcription factor (TF) family regulates many functions in plant growth and development and also during biotic and abiotic stress. In this study, 101 *WRKY* TF gene models in *indica* and *japonica* rice were used to conduct evolutionary analysis, gene structure analysis, and motif composition. Co-expression analysis was carried out first by selecting the differentially expressing genes that showed a significant change in response to the pathogens from Rice Oligonucleotide Array Database (ROAD). About 82 genes showed responses to infection by *Magnaporthe oryzae* or *Xanthomonas oryzae* pv. *oryzae*. Co-expression gene network was constructed using direct neighborhood and context associated inbuilt mode in RiceNetv2 tool. Only 41 genes showed interaction with 2299 non-*WRKY* genes. Variations exist in the structure and evolution of *WRKY* genes among *indica* and *japonica* genotypes which have important implications in their differential roles including disease resistance. *WRKY* genes mediate a complex networking and co-express along with other *WRKY* and non-*WRKY* genes to mediate resistance against fungal and bacterial pathogens in rice.

KEYWORDS: co-expression, gene structure, phylogeny, rice, WRKY transcription factor

RECEIVED: May 16, 2019. **ACCEPTED:** May 17, 2019.

TYPE: Original Research

FUNDING: The author(s) received no financial support for the research, authorship, and/or publication of this article.

DECLARATION OF CONFLICTING INTERESTS: The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

CORRESPONDING AUTHOR: Subramanian Babu, School of Bio Sciences and Technology, Vellore Institute of Technology, Vellore 632014, India. Email: babu.s@vit.ac.in.

Introduction

Transcription factors (TFs) play a key role in regulating the expression of several genes by binding to the promoter regions in turn leading to transcription. Transcription factors control various biological processes through gene regulatory network and signal transduction that results in changes in plant physiology, biochemistry, and morphology. Rice *WRKY* TF genes are involved in seed germination, plant growth and development, signal transduction, and metabolic processes. These participate under various biotic and abiotic stress responses such as salt, drought, freezing, blast, and leaf blight diseases. During these changes, *WRKY* genes are regulated by multiple factors that involve phosphorylation. A number of *WRKY* proteins have been reported to be phosphorylated by Mitogen-Activated Protein Kinases (MAPKs).¹

WRKY TFs are defined based on the presence of the conserved DNA binding regions referred to as a *WRKY* domain. The *WRKY* domain is known to have 60 amino acids comprising a high conserved *WRKY*GQK sequence adjacently placed to zinc finger structures, generally C₂H₂ or C₂HC. Depending on the number of *WRKY* domains and the type of zinc finger structure, the genes were classified into groups.^{2–4} Evolutionary analysis suggests that group I is the oldest due to the presence of 2 *WRKY* domains and groups II and III have originated from group I.^{4–7} This was further supported by a comparison between domain sequences of *WRKY* genes in *Arabidopsis*, rice group I and slime mold.⁴ Thereby, *WRKY* genes are usually classified into 3 groups.^{2–5,7,8} Emergence of a fourth group as a separate classified set of *WRKY* proteins containing a complete *WRKY* motif but lacking a complete zinc finger. These

sequences usually represent either pseudogenes or sequencing and assembly errors.⁴

Variations in the *WRKY* domain is also reported not only in rice⁵ but also in other plants like the occurrence of *WRKY*GKK domain variant has been reported in *WRKY* proteins in a wide variety of plant species including physic nut,⁹ bread wheat,¹⁰ canola,¹¹ tomato,¹² *Arabidopsis*,² soybean,¹³ turkey berry, and eggplant.¹⁴ This implies that the domain has undergone changes during the course of evolution. Comparison of *WRKY* TF gene models of both *Oryza sativa* subsp. *indica* and *O. sativa* subsp. *japonica* was carried out by Ross et al⁴ and identified unique genes that are formed due to duplication (leading to tandem repeats) and deletion of *WRKY* genes which supports the conclusion drawn by Xie et al.⁵ Even though these genes were not evenly distributed across all the chromosomes but the densities on each of the 12 chromosomes were found similar in both the subspecies. Some of these genes were present on the chromosomes as clusters.

Analysis of rice *WRKY* genes has been reported.^{2–4} Jimmy and Babu¹⁵ enlisted 34 reports of *WRKY* family genes involved in biotic stress. Around 45 *OsWRKY*s were induced during bacterial blight and fungal blast caused by *Xanthomonas oryzae* pv. *oryzae* (*Xoo*) and *Magnaporthe oryzae*, respectively.¹⁶ Ramamoorthy et al¹⁷ have reported several *OsWRKY* co-regulated under hormone treatment and abiotic stress. *WRKY*s were reported to be differentially regulated and significantly co-express during biotic or abiotic stress conditions.³ Lee et al¹⁸ have predicted a few genes using in silico analysis and proved that a few genes co-express and therefore differentially regulate the expression of *Xa21* gene. A large number of genes



responsible for resistance and defense were found to be putatively co-expressed along with *WRKY22*.¹⁹ These examples reveal that WRKY TF family functions with a well tuned and complex network of co-expressing genes. Nevertheless, later annotations and update of genome databases indicate a variation in a number of *WRKY* genes in *indica* and *japonica* rice than when they were previously analyzed. According to the Plant transcription factor database v3.0, there are 111 and 129 *WRKY*TF gene models in *indica* and *japonica* rice, respectively. Out of these gene models, 107 and 126 sequences in *indica* and *japonica* rice, respectively, have coding regions but only 101 sequences in *indica* and *japonica* rice were classified into groups as other sequences were found to be putative gene sequences. With the objective of improving our understanding based on the updated sequence data on rice WRKY TFs, an evolutionary analysis was carried out to reveal any variations in the structure of these genes in *indica* and *japonica* rice. Establishing co-expression networks of *WRKY* genes were attempted to explore the possible mechanisms mediated by *WRKY* genes in rice disease resistance.

Materials and Methods

Data sets and sequence retrieval

Amino acid sequences of rice *WRKY*TF gene models (101 of both *indica* and *japonica* rice) were retrieved from the Plant Transcription Factor Database v3.0 (PlnTFDB, <http://plntfdb.bio.uni-potsdam.de/>).²⁰

Classification of WRKY TF Sequences

Protein sequences of rice WRKY proteins were aligned using ClustalW method embedded in MEGA v6.0 (Molecular Evolutionary Genetics Analysis Version 6.0)²¹ and the classification into groups was based on the WRKY domains as in previous literature. The proteins sequences were divided into groups I, II, III, and IV. Group I was divided into 2 subgroups. Group Ia consisted of amino acid sequences that had 2 WRKY domains and C₂H₂ zinc fingers whereas amino acid sequences with 2 WRKY domains and C2HC zinc finger were grouped into subgroup Ib. Group II WRKY proteins contain 1 WRKY domain with C₂H₂ zinc fingers. Group III WRKY proteins contain 1 domain with a C2HC zinc finger. Group IV contains partial incomplete sequences that had WRKY domain but no zinc fingers.

Phylogenetic Analysis and Alignment Using MEGA v6.0

The amino acid sequences of rice *WRKY* genes were manually aligned in FASTA format and saved as FASTA file. The respective FASTA files of both *indica* and *japonica* rice varieties were analyzed and aligned using the ClustalW embedded in MEGA v6.0 program. The aligned sequences were saved as a MEGA file. These MEGA files were used to construct the unrooted phylogenetic tree for both *indica* and *japonica* rice

using the Neighbor-joining method with a bootstrap test (1000 replicates) using MEGA.

Gene Structure Analysis

The accession IDs of the *O. sativa* WRKY coding sequences belonging to *japonica* rice variety were obtained from the Rice Genome Annotation Project (RGAP; <http://rice.plantbiology.msu.edu/>).²² The exon-intron structures of these sequences were determined by manual input of accession IDs into the Gene Structure Display Server tool (GSDS v2.0).²³ The format for the input was selected as GenBank Accession numbers/GI. The color and width were manually selected in the GSDS tool.

Conserved Motif Analysis of WRKY Amino Acid Sequences

The WRKY amino acid sequences were manually edited to remove the gaps and were saved as FASTA file. Sequences of both the rice varieties were separately uploaded in MEME (Multiple Expectation Maximization for Motif Elicitation) online tool (<http://meme-suite.org/>).²⁴ The optimized parameters were any number of repetitions: the optimum width of each motif (between 6 and 50 residues) and maximum number of motifs (15). The output data were saved as separate files. Motif Alignment and Search Tool (MAST) program was used to search the detected motifs in protein databases.

Co-expression Network of WRKY Genes

Normalization

WRKY gene locus IDs were retrieved from the Rice Annotation Project Database²⁵ using the “Keyword search” option to download all the available *WRKY* genes using “WRKY” as “Search Text.” The obtained gene locus IDs were used to download the expression profiles of *WRKY* genes using Rice Oligonucleotide Array Database (ROAD; <http://www.ricearray.org/>).²⁶ The genes were searched for their response against the 2 pathogens: *Xoo* and *M. oryzae*. Gene Expression Data (GED) data sets that had an expression for both or either of the pathogens were selected. The GED data set accession IDs were GSE7256, GSE16793, GSE18361, OS4, and GSE19844 from the selected platform Affymetrix; GSE5906, GSE6244, and GSE62444 from the selected platform Agilent22K; and GSE9653 from the selected platform NSF20K. Normalization and expression data analysis were performed as described.²⁶

Prioritization

The corresponding *MSU/TIGR* gene ID of the differentially expressed genes (DEGs) among the *WRKY* genes obtained from ROAD were used to draw the gene network on the basis of a prioritization server for *O. sativa*—RiceNetv2 (http://www.inetbio.org/ricenet/Network_nga_form_conv.php).²⁷ RiceNet uses embedded Cytoscape for visualization of the

network and provides searching and analyzing information on the molecular interactions around a set of genes. Prioritization was obtained on the basis of both network direct neighborhood and context associated hubs. The locus IDs of the nodes (genes) from the gene prioritization on the basis of the direct neighborhood network and the context associated hubs obtained from RiceNet v2 were used to derive gene descriptions and functions from the RAP-DB through “batch retrieval” option.

Results

Indica and japonica rice genotypes vary in their WRKY gene groups

WRKY genes were found to be unevenly distributed in the chromosomes of *indica* and *japonica* (Supplementary Table S1). Annotations of known *japonica* rice *WRKY* sequences are according to the nomenclature as provided by the Rice *WRKY* Working Group.⁶ Phylogenetic trees of *WRKY*TF gene models and protein sequences of both *indica* and *japonica* rice are shown in Figure 1. Amino acid sequences were further classified into 4 main phylogenetic groups to facilitate the analysis of rice *WRKY* gene evolution. Group I was divided into Ia and Ib based on the fact that the zinc finger motif in both the N- and C-terminal *WRKY* domain in Ia genes are of C₂H₂ type; however, in group Ib it is of C₂HC type. C-terminal of the group Ia genes contain a conserved intron in the region encoding *WRKY* domain but in group Ib, both the N- and C-terminal *WRKY* domains contain a conserved intron.

Alignment of 101 amino acid sequences of both *indica* and *japonica* *WRKY* proteins exhibit a prominent feature of the *WRKY* domain, which was a heptapeptide (WRKYGQK) motif at the N-terminus followed by a zinc finger motif (Figure 2). Some of the motifs exhibit variations such as WRKYGKK, WRKYGEK, WKKYGQK, and WRMCGQK as observed in both *indica* and *japonica*. In addition, WRMTGQS was unique to *indica*; however, WRICGQK, WRMTGQS, and WRMCGQS were observed only in *japonica* (Table 1). Such variants were not reported earlier. Our observations are more up to date as the sequences were retrieved a decade later, compared with the existing reference analysis of Xie et al.⁵ Alignment of *indica* and *japonica* rice amino acid sequences revealed that group Ib genes had evolved into group II completely in both the genotypes. Group I in both the genotypes were identified by the presence of 2 domains and C₂H₂ zinc finger. Group II contains 1 single *WRKY* domain and C₂H₂ zinc finger. Group III had a single domain and a C₂HC zinc finger. Group IV contains partial sequence with and without *WRKY* domain or zinc finger. List of genes divided into each group is given in Supplementary Tables S2 and S3 for *indica* and *japonica* rice, respectively. Number of *WRKY* genes classified into each groups and subgroups in *indica* and *japonica* rice is given in Table 2.

Structure of *WRKY* Genes: Similarities and Differences Among *Indica* and *Japonica* Genotypes

Multiple Expectation Maximization for Motif Elicitation online tool²⁴ helped in identifying 15 motifs in the amino acid sequences of *WRKY* proteins in *indica* and *japonica* rice. Motif boxes identified in *WRKY* protein sequences are given in Supplementary Tables S4 and S5. Motifs 1, 2, 3, 4, and 5 (*indica*) and motifs 1, 2, 3, 4, and 6 (*japonica*) were characterized as *WRKY* domains (Supplementary Figures S1 and S2). *WRKY* members within the same subgroups were found to share a similar motif composition. In the amino acid sequences of *indica* rice *WRKY*s, motif 9 and 15 are unique to the group Ia. Motifs 12, 13, and 14 are found in the entire group I members. Motif 8 was observed in group Ib and in all the members of group II; however, motif 7 was observed in group Ia and only a few members of group II. In *japonica* rice, motifs 5, 7, 8, 9, and 14 are unique to group I. Motif 12 was found in all members of the group I and 2. Motif 11 was found in all members of group II. Motif 10 and 15 show exceptions in their occurrence. Motif 10 was found only in 3 members of group I and only in 1 member of group IV.

Exon-intron structure analysis was carried out using the GSDS tool to gain an insight into the evolution of *WRKY* genes of *japonica* rice (Supplementary Figure S3). This was not done in *indica* rice genes as accession numbers were not available. Out of the 101 *WRKY* genes in *japonica* rice as per the PTFD database, accession numbers of only 76 genes were available in RGAP database. Most *WRKY* genes had 2 introns. Out of 76 genes, 39 had 2 introns, 13 had 1 intron, 7 had 3 introns, 6 had 4 introns, 7 had 5 introns, and 1 had 8 introns. Three of the *WRKY* genes had no intron. Most genes showed introns at the 3' end of the genes. Large uninterrupted introns were observed in 15 genes at the 3' end, whereas uninterrupted introns at the 5' end were observed in *WRKY18*, *WRKY28*, and *WRKY49*.

WRKY Genes Co-express With Other *WRKY* and Non-*WRKY* Genes to Mediate Disease Resistance in Rice

To analyze the response of *WRKY* genes to pathogen infection (*M oryzae* and *Xoo*), microarray data were collected from ROAD database.²⁶ The genes which showed a significant change (fold change ≤ 0.5 or ≥ 2) in response to the pathogens were selected. Analysis of the microarray data revealed 82 out of 110 *WRKY* genes that showed responses to infection by *M oryzae* or *Xoo*. The heat maps of the differentially expressing *WRKY* genes were from 3 different platforms: Affymetrix, Agilent 22K, and NSF 20K. Three platforms were selected to select DEGs that should be pathogen-responsive as well as disease resistance genes that already exist in ROAD. Supplementary Figure S4 represents heat maps for 3 different experiments under Affymetrix platform: (1) GSE19844—transcriptomics-based identification of *Xoo* strain BAI3 Talc targets in rice;

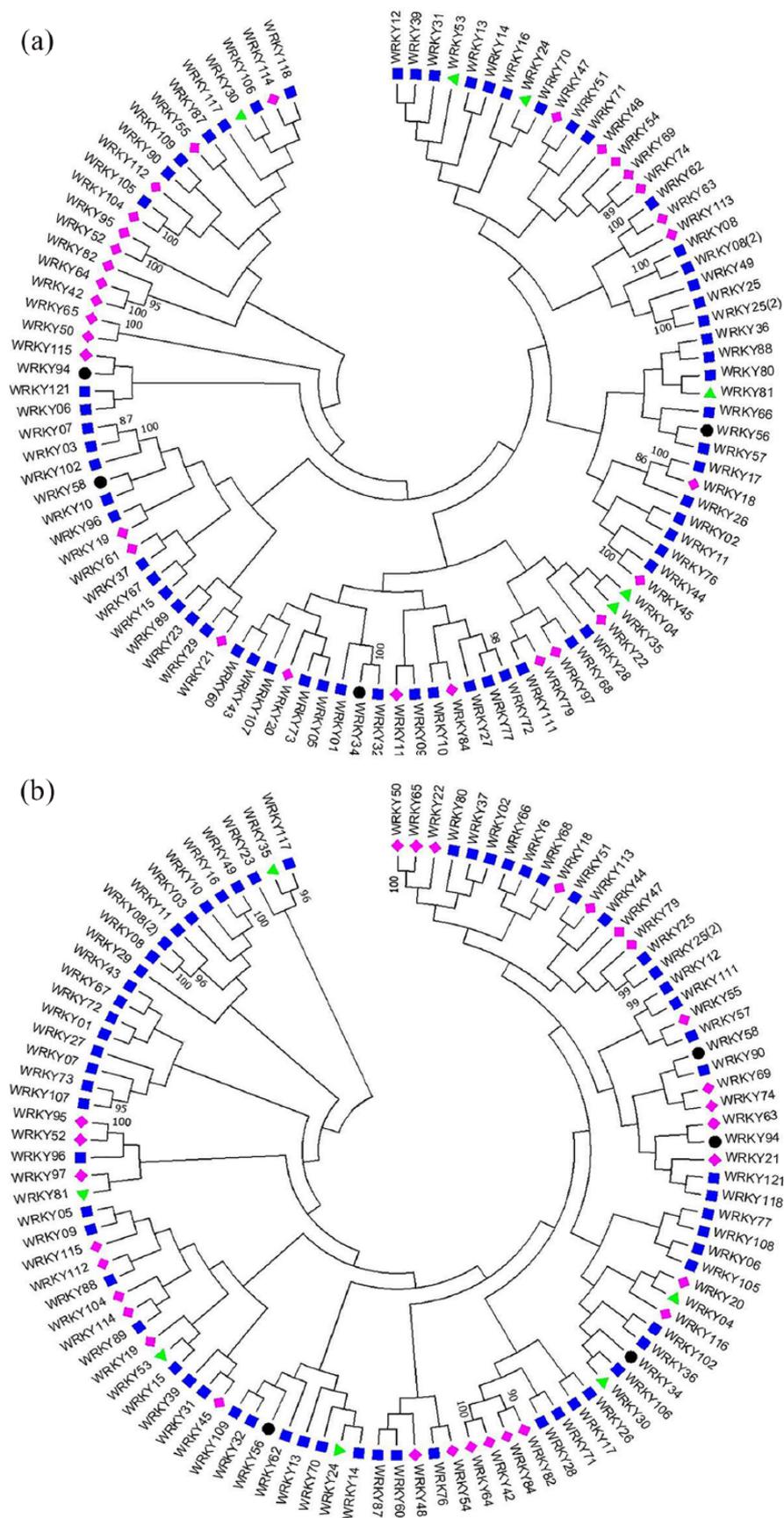


Figure 1. Neighbor-joining bootstrap consensus tree of amino acid sequences of rice WRKY TF sequences: (A) WRKY proteins in *indica* rice; (B) WRKY proteins in *japonica* rice. The phylogenetic trees were constructed using neighbor-joining method with bootstrap test (1000 replicates) in MEGA6. The colors show the groups: green triangle (group Ia), green triangle-unfilled (group Ib), blue rectangles (group II), pink diamonds (group IV), black circles (group V). Sequences with numbers in the parentheses represents duplicated genes. Bootstrap values above 70% are given on the tree. Numbers in parentheses are assigned by MEGA software for the duplicated genes appearing in the both rice genome.

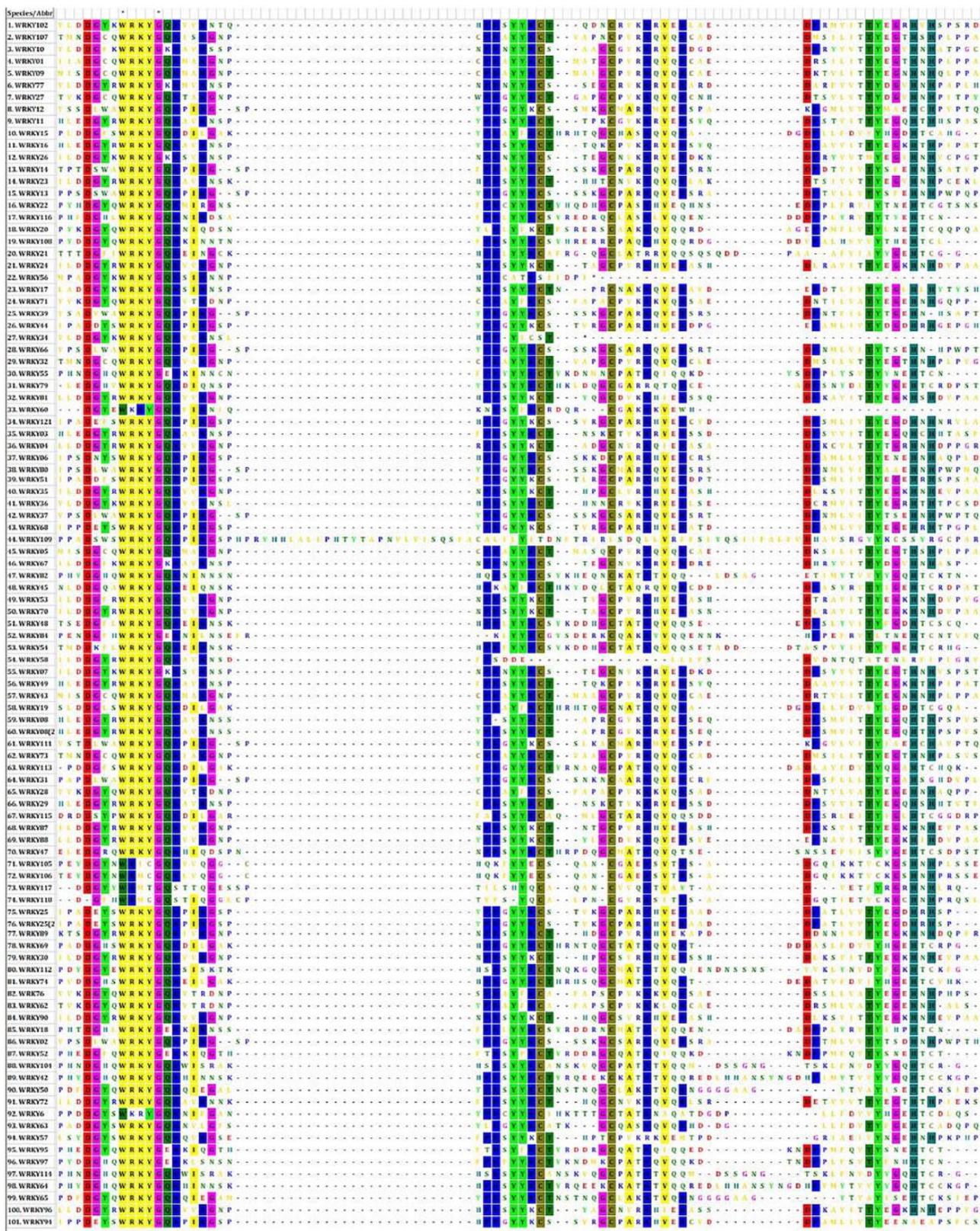


Figure 2. Amino acid sequence alignment of WRKY proteins of *indica* and *japonica* rice: (A) *indica* rice and (B) *japonica* rice.

The highlighted lines represent the highly conserved sequences of WRKY family. The sequences were aligned by ClustalW in MEGA6 and the conserved amino acid residues were determined by inbuilt toggle conserved sites option. Residues that are highly conserved within each of the groups are shown in colors. WRKY domain was highlighted in yellow.

Table 1. Variants of the WRKYGQK amino acid residues in *indica* and *japonica* rice.

VARIANTS	INDICA RICE	JAPONICA RICE
WRKYGKK	WRKY59, WRKY79	WRKY10, WRKY77, WRKY26, WRKY67, WRKY67.2
WRKYGEK	WRKY55, WRKY46, WRKY52, WRKY57	WRKY55, WRKY55.2, WRKY84, WRKY07, WRKY18, WRKY52, WRKY46, WRKY46.2, WRKY97.2
WKKYGQK	WRKY60	WRKY61, WRKY106, WRKY117, WRKY118
WRICGQK	–	WRKY105
WRMCGQK	WRKY80	WRKY106
WRMTGQS	–	WRKY117
WRMCGQS	–	WRKY118
WRMTGQS	WRKY72	–

Table 2. Number of WRKY genes from *indica* and *japonica* in each group and subgroup.

GROUP/SUB GROUP	GENE NUMBER		
	JAPONICA RICE (WU ET AL ²)	INDICA RICE (PRESENT STUDY)	JAPONICA RICE (PRESENT STUDY)
Ia	14	11	6
Ib	20	01	–
II	30	54	61
III	36	31	30
IV	–	04	4
Total	100	101	101

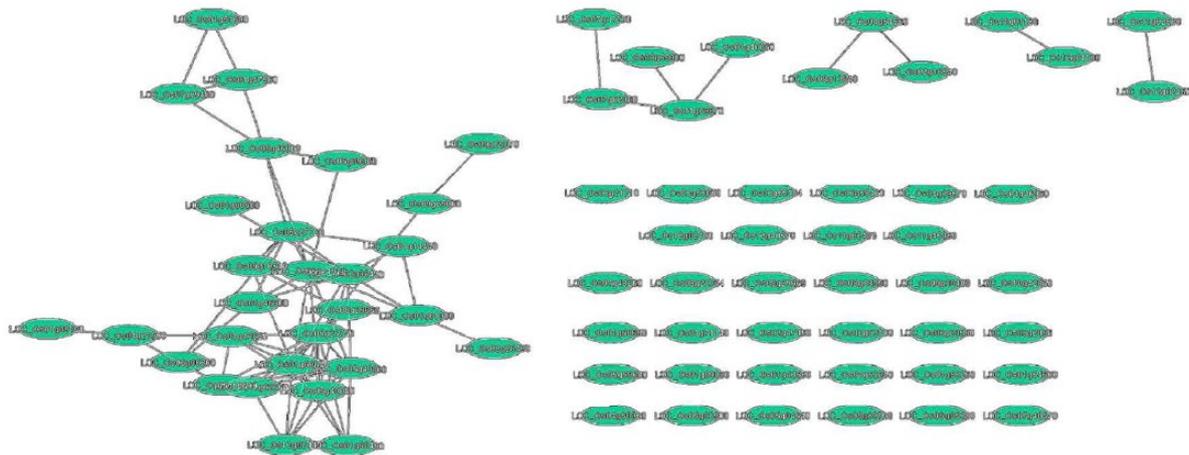


Figure 3. Guide genes that are connected to one another. Guide genes are also called as input genes. These genes help in determining the co-expressing genes. Genes appearing continuous are disconnected from the guide genes.

(2) GSE18361 Rice root infected with *M oryzae* strain Guy11; and (3) GSE7256—Virulent infection by *M oryzae*. Supplementary Figures S5 and S6 represents heat map GSE16793—Rice undergoing infection by *Xoo* or by *Xanthomonas oryzae* pv. *oryzicola* (*Xoc*) using 60 samples and OS4—Rice cultivars IRBB7, IRBB5, and IR24 undergoing

infection by *Xoo*. Supplementary Figure S7 represents 3 heat maps for Agilent22K: (1) GSE6244—*indica* rice cultivar IET8585 (Ajaya) infected with *Xoo*; (2) GSE5906—bacterial lipopolysaccharides induce defense responses; (3) GSE8518—biotrophic invasion by *M oryzae*. Supplementary Figure S8 represents 1 NSF20K heatmap for array data for rice to

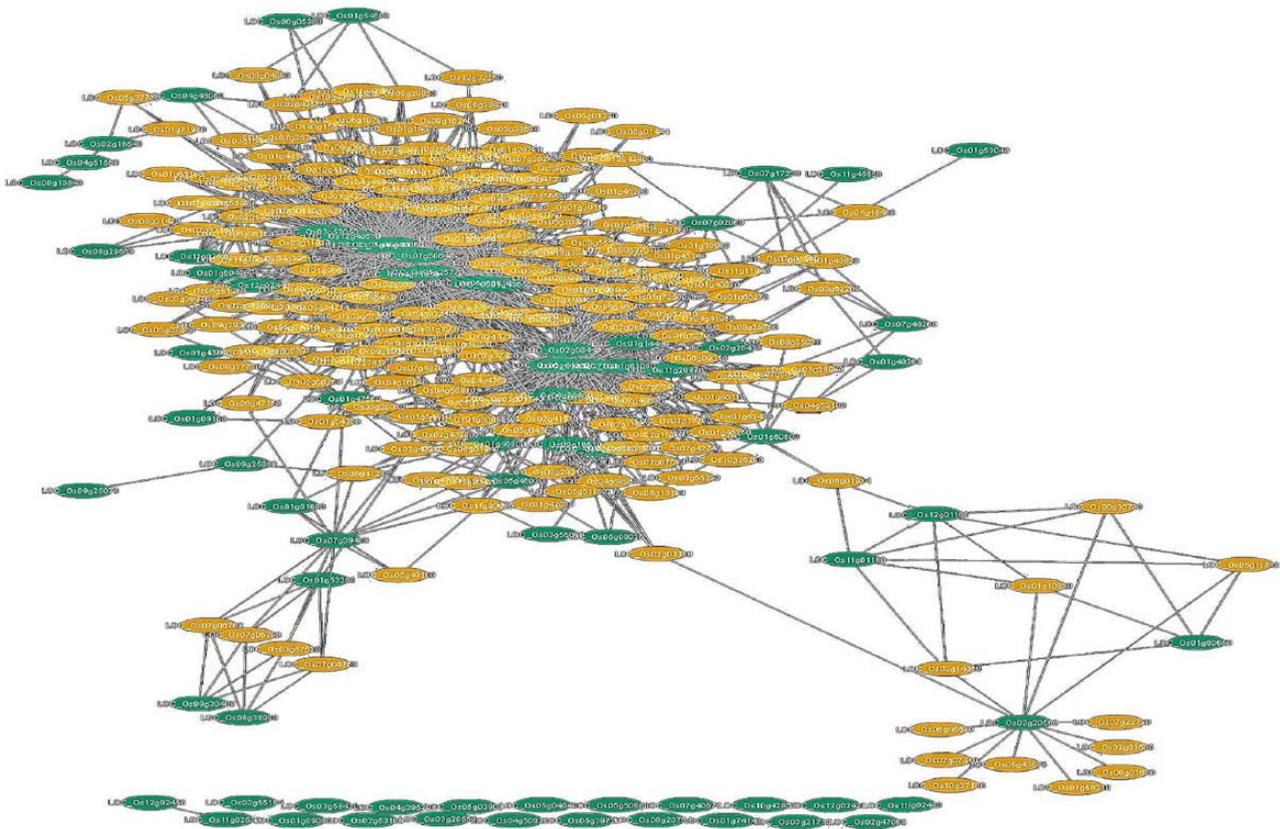


Figure 4. Co-expression network map of *WRKY* genes. Constructed using direct neighborhood method. Green oval symbol represents guide genes and orange oval symbol represents co-expressing genes.

identify gene to dissect rice defense response pathways. The total number of arrays considered from all 3 platforms was 270 (105 from *japonica*, 51 from *indica*, and 114 from mixed samples).

Co-expression of *WRKY* genes with other genes in rice genome that have significant changes in their expression in response to pathogens were used to construct the expression network using RiceNet tool.^{18,27} The network provided the likelihood (designed log network score, LLS) of the genes that are participating in the conditioned process in each data set. Bootstrapping value for all LLS evaluations was 0.632.^{18,27} The backbone of the network was formed by the guide genes. These are DEGs that were obtained from ROAD database that were also observed to be linked to one another. *WRKY*TF network-ging genes (top 100 genes) in mediating disease resistance is given in Supplementary Table S6. In the co-expression network, the input *WRKY* genes are the guide genes and henceforth referred to as guide *WRKY* genes. Interaction among the guide *WRKY* gene was observed for 41 genes out of 82 guide *WRKY* genes (Figure 3 and Supplementary Table S7). These 41 genes co-expressed along with 2299 non-*WRKY* genes. As depicting a large number of genes in the Cytoscape was difficult, about 100 genes co-expressing along with the 41 guide *WRKY* genes are shown in Figure 4. Out of the 41 guide *WRKY* genes, 14 were already reported to be involved in resistance to both fungal and bacterial pathogens.¹⁵ *WRKY64*, *WRKY62*,

and *WRKY77* were reported for resistance against bacterial pathogens, whereas *WRKY22*, *WRKY24*, *WRKY69*, *WRKY74*, *WRKY108*, and *WRKY113* were reported for resistance against fungal pathogens. *WRKY28*, *WRKY45*, *WRKY53*, *WRKY71*, and *WRKY76* were reported for their role in resistance to both bacterial and fungal pathogens.¹⁵ In total, 31 guide *WRKY* genes were found to be disconnected in the network. This indicates that these genes do not take part in the main network but may have individual networks that contribute to disease resistance in rice. Out of the 31 disconnected genes (Supplementary Table S8), 10 were known for their involvement in rice disease resistance. Out of the 82 input genes, 7 were not found in RiceNet but were found to be valid genes (Supplementary Table S9). The co-expression network revealed that each non-*WRKY* gene was connected to at least 3 to 5 guide *WRKY* genes. The co-expressed genes are known to be involved in physiology, pathogen resistance, negative regulation, and defense responses similar to that of the *WRKY* genes. The result obtained was represented using Figure 4 and Supplementary Table S10 that only represents 100 out of 2299 genes in the Cytoscape plugin. Network analysis indicated *WRKY07*, *WRKY24*, *WRKY28*, *WRKY53*, *WRKY71*, and *WRKY108* were linked with *OsMSRMK2* (multiple stress-responsive MAP kinase genes 2). Perforin protein was found linked with *WRKY28*, *WRKY51*, *WRKY53*, and *WRKY71*; *bZIP50* linked with *WRKY15*, *WRKY53*, *WRKY71*, and

WRKY113; *OsMKK6* linked with *WRKY71*; Patatin-related phospholipase A II eta (*OspPLAIIeta*) linked with *WRKY53*; Phytoalexin-deficient 4 (*OsPAD4*) linked with *WRKY53* and NBS-LRR (Nucleotide Binding Sites–Leucine Rich Repeat) linked with *WRKY74*.

Descriptions of all the 2299 genes that interact along with guide *WRKY* genes were found using RAP-DB database and were classified depending on their interaction with the guide *WRKY* genes. The co-expressing 2299 genes were sorted based on their functions provided by the RiceNet tool. These genes were sorted into those involved in defense, negative regulation, pathogen resistance, hormone-responsive genes, *MAPK* responsive genes, physiological response, and unknown function (Supplementary Figure S9 and S10). This suggests that *WRKY* TF mediate a complex networking among them as well as the co-expressing non-*WRKY* gene toward disease resistance.

Context associated hubs involve the use of DEGs as input. This involves the study of central hubs and their neighbors that form subnetwork which was associated with the stress context. The context associated hubs are based on the guide genes that may or may not be connected with each other. The genes of the central hubs may or may not be DEGs. The same 82 guide *WRKY* genes (DEGs) were used to determine the central hubs. This analysis resulted in 100 co-expressing genes which were recognized to form a subnetwork with a positive *P* value (Supplementary Table S11). Out of these 100 co-expressing genes, 10 genes were the input DEG. The genes that formed a subnetwork to these DEG were found to be either transcription factors or defense genes. The context associated hubs are DEG dependent. Each DEG/set of DEGs will reveal a subnetwork of genes that form a hub. A single subnetwork gene in the hub was represented in the figure (Supplementary Figure S11).

Discussion

Oryza sativa L. is a model organism containing the smallest genome. *Oryza sativa* subsp. *japonica* rice contains a haploid chromosome number of 12, containing 370 Mb with 30 000 protein-coding genes,²⁸ whereas subsp. *indica* rice contains a haploid chromosome number of 12, containing 466 Mb with 46 022 to 55 615 coding genes.²⁹ The average number of exons and introns per gene are same in *indica* and *japonica* sequences.³⁰ Chromosome 4 is larger in the case of *japonica* rice than that of *indica* because of expansion of *japonica* sequence by insertions. Khush³¹ stated that *Oryza nivara* populations are of polyphyletic origin from both *indica* and *japonica* rice which helps to conclude that *japonica* might have evolved earlier than *indica* as more insertions have been added to the chromosomes of *japonica*. Comparatively, the number of stress-responsive genes are more in *japonica* rice (1716) than *indica* rice (1118)³² but *indica* rice is better resistant to diseases than *japonica* rice.³³

The *WRKY* genes were divided into 4 groups unlike the previous grouping carried out.^{2–5} This was due to the fact that

some of the sequences were partial which are grouped in group IV. The number of genes in group I reduced due to gene loss, and the number of genes in group II and III have increased due to gene duplication. Gene loss was evident from the number of variants that exist in a group. Comparison of the amino acid sequences of *indica* and *japonica* rice has helped to report mutational changes giving rise to a greater number of variants in the domain structures in the family (Table 2) than already reported.^{2–5} Such variants were observed more in *indica* than in *japonica* rice and maximum in group I members. Gene loss in group I members caused no reduction in the number of *WRKY* family members as gene duplication observed in group II and group III has compensated. One *WRKY* gene in *indica* was grouped into subgroup Ib, whereas in *japonica* rice, no *WRKY* gene containing C2HC zinc finger was found. Gene structure analysis revealed that genes of the same group exhibit similar intron-exon patterns at least for the known genes obtained in the RAP-DB database.

Phylogenetic trees lead to grouping based on zinc finger-type and domain-type exhibited by rice *WRKY* TFs. Group I members were found to be placed within group II members in both *indica* and *japonica*. Xie et al⁵ found group Ib genes embedded within the group III, thereby inferring that group Ib gave rise to group III. Based on the amino acid sequences of *WRKY* proteins in *japonica* rice, members belonging to group I was found closely associated with a group III member. In *indica* rice, group I genes are closely placed with group II genes. This indicated that group II would have originated from group I in *indica* rice due to the loss of N-terminus. Similarly, in *japonica* rice, group III would have originated from group II. Group IV are rarely reported and they contain partial sequences.^{1,5} These results follow the same lines to that reported by Ross et al.⁴ The domain duplication scenario in *WRKY* sequences was reported to explain the evolution of group I genes from either C₂H₂-C₂H₂ or C2HC-C2HC type of zinc finger. This was reported on the basis that no C₂H₂-C2HC or C2HC-C₂H₂ type of zinc finger was found in *WRKY* amino acid sequences. This also indicated group Ib evolved from group III *WRKY* domain that had the C2HC type of zinc finger due to intramolecular duplication. Similarly, group II gene evolution was based on the domain feature loss and gain of the *WRKY* domain. Group II genes are known to have evolved from group Ia due to the loss of N-terminus but independent of the loss of C-terminus.⁵

Gene duplications, recombination, and sequence divergence have long been considered as the main sources of evolutionary momentum. On the basis of the phylogenetic analysis, *WRKY* genes such as *WRKY93* and *WRKY94* genes are considered as duplicated genes. Each of these genes contains 2 *WRKY* domains, which are generated probably due to the domain duplication clustered in the same clade in subgroup IIIa. Wu et al² reported tandem duplication events occurred in the *WRKY* gene arrangements in *O. nivara* which were also

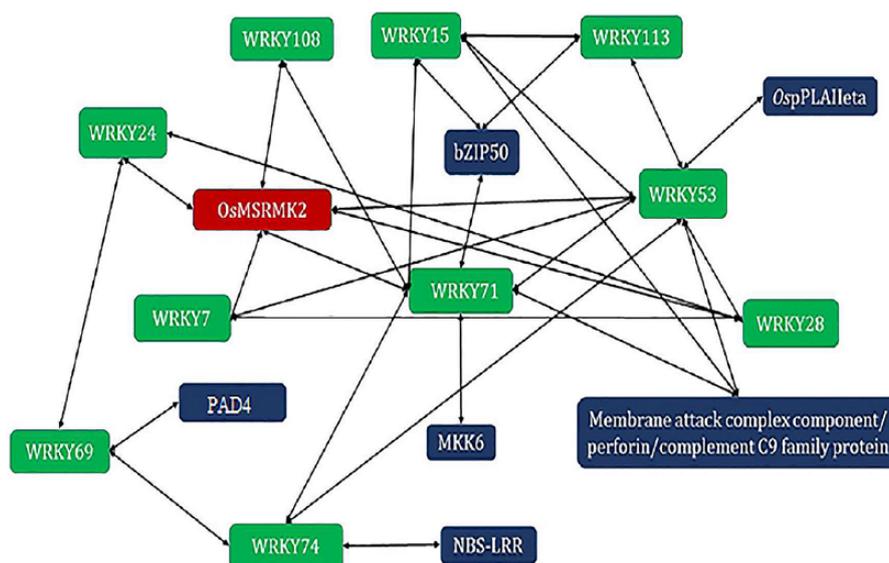


Figure 5. Core *WRKY* genes co-expressing with other genes involved in MAPK cascade.

Green rounded rectangles: *WRKY* TF genes; blue rounded rectangles: genes involved in MAPK cascade; red rounded rectangle: gene involved in activation. Green rounded rectangles: *WRKY* TF genes. *WRKY* genes are phosphorylated by MAPKs thereby leads to the activation of other genes, thus forming a network. The members of the *WRKY* family already reported to play a role in disease resistance in rice are used to built this network.

observed in other plants. Similarly, Ross et al⁴ and Li et al³⁴ proposed gene duplication and loss that caused an increase in the number and density of *WRKY* TFs. The motifs that were found using the MEME software were predicted motifs that may be responsible for acting as binding site of many regulatory proteins. Most motifs may be related to the positive or negative regulation of the *WRKY* TFs. Therefore, these motifs are directly related to the regulation of the expression of the *WRKY* genes during various physiological changes. Two short sequences such as PTDDS and EDLEEK in the N-terminus of *WRKY62* was required for its repressor activity in plants. They represent new types of repression motifs.³⁵ This may be related to the reported negative feedback regulation of the IIa subfamily of *WRKY*s but also provide evidence for alternative splicing of *WRKY* TFs during the plant defense response.³⁵ Therefore, the group IIb members out of all the *WRKY* networking genes may be predicted to be involved in a network as a repressor of other proteins or in negative feedback mechanism in defense response.

*WRKY*TF gene network regulates via complex networking of genes in response to biotic stress. This involves a series of signal transduction of genes through TFs or other proteins. In this study, 82 *WRKY* TFs were found interacting with other genes to modulate biotic stress resistance. Network analysis indicated that certain *WRKY*s co-express along with other genes involved in MAPK cascade pathway. It was already known that a number of *WRKY* proteins have been reported to be phosphorylated by MAPKs. The MAPK cascade plays an important role in regulating various responses and also regulates the intra- and extra-cellular signaling in plants.^{36,37} Therefore, it was predicted that *WRKY* TFs co-expressed with genes involved in MAPK cascade that were the core genes that

get activated first during a pathogen attack. Tracing the genes that were connected to these core *WRKY* genes, the network by which *WRKY* TFs activates other proteins were predicted. According to the Gramene database,³⁸ these genes also have physiological response functions. *WRKY* TFs on activation further activates a deep transcriptional reprogramming of the cell that involves defense genes such as genes for secondary metabolites (phytoalexins like momilactones, phenylamides), pathogenesis-related (PR) proteins (PR3—chitinase 11, chitinase 6, chitinase 10; PR4c, PR10a, PR10b; and root-specific PR10), cell wall strengthening proteins (callose, pectin, lignin, and cutin), and programmed cell death and hypersensitive reactions. These are the genes that are usually present downstream to the *WRKY* genes.

It was observed that *WRKY53* co-expressed with β -1,3-glucan and glucan endo-1,3- β -glucosidase GII precursor. β -1,3-glucan was responsible for callose deposition on the plant cell wall³⁹ and glucan endo-1,3- β -glucosidase GII precursor was a cell wall-associated gene.⁴⁰ Other TFs such as NAC, ERF, bZIP, and MYB are observed to co-express along with *WRKY* TFs. Plant resistance (*R*) genes such as those protein products containing nucleotide-binding sites and NBS-LRR motifs were observed to co-express along with certain *WRKY* TFs. Pathogenesis-related protein genes were also found as the hub genes when the *WRKY* genes were used as input for the context associated hub analysis. Hubs constitute genes that consist of neighbors that have significant overlap to the DEG used as input to search the subnetwork (hub) forming genes.^{18,27} It was predicted that if a central hub gene of a subnetwork modulates stress response in this case disease resistance, many of its neighboring subnetwork genes could be DEGs. Our analyses indicated jumonji 709, fatty acid elongase,

increased leaf inclination 7, VQ motif-containing protein 13, ZOS1-14-C₂H₂ zinc finger protein, Lil3 protein, TMV (tobacco mosaic virus) response related protein overlap with the direct neighborhood network. Therefore, these genes are predicted to be involved in disease resistance.

Cell wall strengthening provides mechanical defense against a pathogen that ramifies in the plant tissues. Callose deposition mainly defends against the invading pathogens⁴¹ followed by pectin,⁴² lignin,⁴³ and cutin⁴⁴ biosynthesis. *NAC6* was reported to be up-regulated along with Pi54 gene over-expressed transgenic rice plants that confer resistance to *M. oryzae*.⁴² A large number of R genes have a role in disease resistance in rice. *ERF* genes are also found to be expressed due to pathogenic infection in rice⁴⁵ along with other TFs such as *bZIP* genes.⁴⁶ PR1a was already reported to be induced by *WRKY12* gene during SA and bacterial elicitor in the tobacco plant.⁴⁷ PR1b was found to be expressed during the over-expression of *WRKY03* in salicylic acid (SA)-dependent or jasmonic acid (JA)-dependent defense signaling cascades.⁴⁸ Allelic genes *WRKY45-1* and *WRKY45-2* were reported to act antagonistically during the accumulation of signaling molecules (SA and JA) during pathogen attack.⁴⁹ *WRKY13* was found as a disconnected gene but it was found to interact with the core genes *WRKY71* and *WRKY28* via *TIFY9*. *WRKY13* gene was reported to act as an activator of the SA-dependent pathway and a suppressor of the JA-dependent pathway⁵⁰ and works against both bacterial and fungal pathogen.¹⁵ *TIFY9* in rice performs role in JA biosynthetic process, responds to ethylene, auxin, abscisic acid, JA, and wounding.³⁸ *WRKY13* also binds to another TF gene, *bZIP37* that also interacts in MAPK cascade as illustrated (Figure 5). *bZIP37* is involved in ethylene biosynthetic process. It also responds to abscisic acid, SA and JA signaling and expressed during hypersensitive response.³⁸ These could be candidate genes for developing multiple disease resistant rice varieties through molecular breeding/transgenic approaches. Conclusively, we observed variations exist in the structure and evolution of *WRKY* genes among *indica* and *japonica* genotypes, which have important implications in their differential roles including disease resistance. *WRKY* genes mediate complex networking and co-express along with other *WRKY* and non-*WRKY* genes to mediate resistance against fungal and bacterial pathogens in rice.

Acknowledgements

The authors sincerely acknowledge the support rendered by the Vellore Institute of Technology, Vellore, in carrying out our research.

Author Contributions

JJ obtained the data from the databases, performed all the data analysis, prepared the results, and contributed in writing the first draft of the manuscript. SB conceived the concept, planned the work, contributed to result interpretation and manuscript writing.

Availability of Data and Materials

All data used in the study are retrieved from publicly available databases. Plant Transcription Factor Database v3.0 (PTFD; <http://planttfdb.cbi.pku.edu.cn/>), Rice Genome Annotation Project (RGAP; <http://rice.plantbiology.msu.edu/>), Rice Annotation Project Database (RAP-DB; <http://rapdb.dna.affrc.go.jp/gbl/>), and Rice Oligonucleotide Array Database (ROAD; <http://www.ricearray.org/>) are the databases used.

Supplemental Material

Supplemental material for this article is available online.

ORCID iD

Subramanian Babu  <https://orcid.org/0000-0003-2781-7136>

REFERENCES

- Xu H, Watanabe KA, Zhang L, Shen QJ. WRKY transcription factor genes in wild rice *Oryza nivara*. *DNA Res.* 2016;23:311–323.
- Wu KL, Guo ZJ, Wang HH, Li J. The WRKY family of transcription factors in rice and Arabidopsis and their origins. *DNA Res.* 2005;12:9–26.
- Berri S, Abbruscato P, Faivre-Rampant O, et al. Characterization of WRKY co-regulatory networks in rice and Arabidopsis. *BMC Plant Biol.* 2009;9:120. doi:10.1186/1471-2229-9-120.
- Ross CA, Liu Y, Shen QJ. The WRKY gene family in rice (*Oryza sativa*). *J Integr Plant Biol.* 2007;49:827–842. doi:10.1111/j.1744-7909.2007.00504.x.
- Xie Z, Zhang ZL, Zou Z, et al. Annotations and functional analyses of the rice WRKY gene superfamily reveal positive and negative regulators of abscisic acid signaling in aleurone cells. *Plant Physiol.* 2005;137:176–189.
- Rice WRKY Working Group. Nomenclature report on rice WRKYs—conflict regarding gene names and its solution. *Rice.* 2012;5:3. doi:10.1186/1939-8433-5-3.
- Ulker B, Somssich IE. WRKY transcription factors: from DNA binding towards biological function. *Curr Opin Plant Biol.* 2004;7:491–498.
- Eulgem T, Rushton PJ, Robatzek S, Somssich IE. The WRKY superfamily of plant transcription factors. *Trends Plant Sci.* 2000;5:199–206.
- Xiong W, Xu X, Zhang L, et al. Genome-wide analysis of the WRKY gene family in physic nut (*Jatropha curcas* L.). *Gene.* 2013;524:124–132.
- Okay S, Derelli E, Unver T. Transcriptome-wide identification of bread wheat WRKY transcription factors in response to drought stress. *Mol Genet Genomics.* 2014;289:765–781.
- Yang B, Jiang Y, Rahman MH, Deyholos MK, Kav NN. Identification and expression analysis of WRKY transcription factor genes in canola (*Brassica napus* L.) in response to fungal pathogens and hormone treatments. *BMC Plant Biol.* 2009;9:68. doi:10.1186/1471-2229-9-68.
- Huang S, Gao Y, Liu J, et al. Genome-wide analysis of WRKY transcription factors in *Solanum lycopersicum*. *Mol Genet Genomics.* 2012;287:495–513.
- Yin G, Xu H, Xiao S, et al. The large soybean (*Glycine max*) WRKY TF family expanded by segmental duplication events and subsequent divergent selection among subgroups. *BMC Plant Biol.* 2013;13:148. doi:10.1186/1471-2229-13-148.
- Yang X, Deng C, Zhang Y, Cheng Y, Huo Q, Xue L. The WRKY transcription factor genes in eggplant (*Solanum melongena* L.) and Turkey berry (*Solanum torvum* Sw.). *Int J Mol Sci.* 2015;16:7608–7626.
- Jimmy JL, Babu S. Role of OsWRKY transcription factors in rice disease resistance. *Trop Plant Pathol.* 2015;1:355–361.
- Ryu H, Han M, Lee S, et al. A comprehensive expression analysis of the WRKY gene superfamily in rice plants during defense response. *Plant Cell Rep.* 2006;25:836–847.
- Ramamoorthy R, Jiang S-Y, Kumar N, Venkatesh PN, Ramachandran S. A comprehensive transcriptional profiling of the WRKY gene family in rice under various abiotic and phytohormone treatments. *Plant Cell Physiol.* 2008;49:865–879.
- Lee I, Seo YS, Coltrane D, et al. Genetic dissection of the biotic stress response using a genome-scale gene network for rice. *Proc Natl Acad Sci U S A.* 2011;108:18548–18553.
- Abbruscato P, Nepusz T, Mizzi L, et al. OsWRKY22, a monocot WRKY gene, plays a role in the resistance response to blast. *Mol Plant Pathol.* 2012;13:828–841.
- Jin JP, Zhang H, Kong L, Gao G, Luo JC. PlantTFDB 3.0: a portal for the functional and evolutionary study of plant transcription factors. *Nucleic Acids Res.* 2014;42:D1182–D1187.

21. Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. MEGA6: molecular evolutionary genetics analysis version 6.0. *Mol Biol Evol.* 2013;30:2725–2729.
22. Kawahara Y, de la Bastide M, Hamilton JP, et al. Improvement of the *Oryza sativa* Nipponbare reference genome using next generation sequence and optical map data. *Rice (NY).* 2013;6:4. doi:10.1186/1939-8433-6-4.
23. Hu B, Jin J, Guo AY, Zhang H, Luo J, Gao G. GSDS 2.0: an upgraded gene feature visualization server. *Bioinformatics.* 2015;31:1296–1297.
24. Bailey TL, Boden M, Buske FA, et al. MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res.* 2009;37:W202–W208. doi:10.1093/nar/gkp335.
25. Sakai H, Lee SS, Tanaka T, et al. Rice Annotation Project Database (RAP-DB): an integrative and interactive database for rice genomics. *Plant Cell Physiol.* 2013;54:e6. doi:10.1093/pcp/pcs183.
26. Cao P, Jung KH, Choi D, Hwang D, Zhu J, Ronald PC. The rice oligonucleotide array database: an atlas of rice gene expression. *Rice (N Y).* 2012;5:17. doi:10.1186/1939-8433-5-17.
27. Lee T, Oh T, Yang S, et al. RiceNet v2: an improved network prioritization server for rice genes. *Nucleic Acids Res.* 2015;43:W122–W127. doi:10.1093/nar/gkv253.
28. UniProt Consortium. UniProt: a hub for protein information. *Nucleic Acids Res.* 2015;43:D204–D212.
29. Yu J, Hu S, Wang J, et al. A draft sequence of the rice genome (*Oryza sativa* L. Ssp. Indica). *Science.* 2002;296:79–92.
30. Han B, Xue Y. Genome-wide intraspecific DNA-sequence variations in rice. *Curr Opin Plant Biol.* 2003;6:134–138.
31. Khush GS. Origin, dispersal, cultivation and variation of rice. *Plant Mol Biol.* 1997;35:25–34.
32. Shameer K, Ambika S, Varghese SM, Karaba Udayakumar M, Sowdhamini R. STIFDB-arabidopsis stress responsive transcription factor data base. *Int J Plant Genomics.* 2009;2009:583429.
33. Visser M, Isaacs N. *Much Depends on Dinner: The Extraordinary History and Mythology, Allures and Obsession, Perils and Taboos of an Ordinary Meal.* New York, NY: Grove Atlantic; 1986.
34. Li M-Y, Xu Z-S, Tian C, Huang Y, Wang F, Xiong A-S. Genomic identification of WRKY transcription factors in carrot (*Daucus carota*) and analysis of evolution and homologous groups for plants. *Sci Rep.* 2016;6:23101. doi:10.1038/srep23101.
35. Liu J, Chen X, Liang X, et al. Alternative splicing of Rice WRKY62 and WRKY76 transcription factor genes in pathogen defense. *Plant Physiol.* 2016;171:1427–1442.
36. Kumar K, Sinha AK. Overexpression of constitutively active mitogen activated protein kinase kinase 6 enhances tolerance to salt stress in rice. *Rice (N Y).* 2013;6:25. doi:10.1186/1939-8433-6-25.
37. Adachi H, Nakano T, Miyagawa N, et al. WRKY transcription factors phosphorylated by MAPK regulate a plant immune NADPH oxidase in *Nicotiana benthamiana*. *Plant Cell.* 2015;27:2645–2663.
38. Tello-Ruiz MK, Stein J, Wei S, et al. Gramene 2016: comparative plant genomics and pathway resources. *Nucleic Acids Res.* 2016;44:D1133–D1140. doi:10.1093/nar/gkv1179.
39. Gupta SK, Rai AK, Kanwar SS, Chand D, Singh NK, Sharma TR. The single functional blast resistance gene Pi54 activates a complex defence mechanism in rice. *J Exp Bot.* 2012;63:757–772.
40. Ricardi MM, Gonzalez RM, Zhong S, et al. Genome-wide data (ChIP-seq) enabled identification of cell wall-related and aquaporin genes as targets of tomato ASR1, a drought stress-responsive transcription factor. *BMC Plant Biol.* 2014;14:29.
41. Voigt CA. Callose-mediated resistance to pathogenic intruders in plant defense-related papillae. *Front Plant Sci.* 2014;5:168. doi:10.3389/fpls.2014.00168.
42. Shamimuzzaman M, Vodkin L. Genome-wide identification of binding sites for NAC and YABBY transcription factors and co-regulated genes during soybean seedling development by ChIP-Seq and RNA-Seq. *BMC Genomics.* 2013;14:477. doi:10.1186/1471-2164-14-477.
43. Li W, Tian Z, Yu D. WRKY13 acts in stem development in Arabidopsis thaliana. *Plant Sci.* 2015;236:205–213.
44. Serrano M, Coluccia F, Torres M, Haridon FL, Metraux J-P. The cuticle and plant defense to pathogens. *Front Plant Sci.* 2014;5:274. doi:10.3389/fpls.2014.00274.
45. Du D, Hao R, Cheng T, et al. Genome-wide analysis of the AP2/ERF gene family in *Prunus mume*. *Plant Mol Biol Rep.* 2013;31:741–750.
46. E ZG, Zhang YP, Zhou JH, Wang L. Mini review roles of the bZIP gene family in rice. *Genet Mol Res.* 2014;13:3025–3036.
47. van Verk MC, Pappaioannou D, Neeleman L, Bol JF, Linthorst HJ. A novel WRKY transcription factor is required for induction of PR-1a gene expression by salicylic acid and bacterial elicitors. *Plant Physiol.* 2008;146:1983–1995.
48. Liu XQ, Bai XQ, Qian Q, Wang XJ, Chen MS, Chu CC. OsWRKY03, a rice transcriptional activator that functions in defense signaling pathway upstream of OsNPR1. *Cell Res.* 2005;15:593–603.
49. Tao Z, Liu H, Qiu D, et al. A pair of allelic WRKY genes play opposite roles in rice-bacteria interactions. *Plant Physiol.* 2009;151:936–948.
50. Qiu D, Xiao J, Ding X, et al. OsWRKY13 mediates rice disease resistance by regulating defense-related genes in salicylate- and jasmonate-dependent signaling. *Mol Plant Microbe Interact.* 2007;20:492–499.