



Original article

RiceMetaSysB: a database of blast and bacterial blight responsive genes in rice and its utilization in identifying key blast-resistant WRKY genes

V. Sureshkumar^{1,†}, Bipratip Dutta^{1,†}, Vishesh Kumar^{1,2,†}, G. Prakash³, Dwijesh C. Mishra⁴, K. K. Chaturvedi⁴, Anil Rai⁴, Amitha Mithra Sevanthi¹ and Amolkumar U. Solanke^{1,*}

¹Indian Council of Agricultural Research (ICAR)—National Research Centre on Plant Biotechnology, Pusa Campus, New Delhi, 110012, India, ²Jamia Hamdard, Hamdard Nagar, New Delhi, 110062, India, ³ICAR—Indian Agricultural Research Institute, Pusa Campus, New Delhi, 110012, India and ⁴ICAR—Indian Agricultural Statistics Research Institute, Pusa Campus, New Delhi, 110012, India

*Corresponding author: Tel.: +91-1125842789; Fax: +91-1125843984; Email: amolsgene@gmail.com

†These authors contributed equally to this work.

Citation details: Sureshkumar,V., Dutta,B., Kumar,V., *et al.* RiceMetaSysB: a database of blast and bacterial blight responsive genes in rice and its utilization in identifying key blast-resistant WRKY genes. *Database* (2019) Vol. 2019: article ID baz015; doi:10.1093/database/baz015

Received 31 October 2018; Revised 26 December 2018; Accepted 17 January 2019

Abstract

Nearly two decades of revolution in the area of genomics serves as the basis of present-day molecular breeding in major food crops such as rice. Here we report an open source database on two major biotic stresses of rice, named RiceMetaSysB, which provides detailed information about rice blast and bacterial blight (BB) responsive genes (RGs). Meta-analysis of microarray data from different blast- and BB-related experiments across 241 and 186 samples identified 15135 unique genes for blast and 7475 for BB. A total of 9365 and 5375 simple sequence repeats (SSRs) in blast and BB RGs were identified for marker development. Retrieval of candidate genes using different search options like genotypes, tissue, developmental stage of the host, strain, hours/days post-inoculation, physical position and SSR marker information is facilitated in the database. Search options like 'common genes among varieties' and 'strains' have been enabled to identify robust candidate genes. A 2D representation of the data can be used to compare expression profiles across genes, genotypes and strains. To demonstrate the utility of this database, we queried for blast-responsive WRKY genes (fold change ≥ 5) using their gene IDs. The structural variations in the 12 WRKY genes so identified and their promoter regions were explored in two rice genotypes contrasting for their reaction to blast infection. Expression analysis of these genes in panicle tissue infected with a virulent and an avirulent strain of

Magnaporthe oryzae could identify *WRKY7*, *WRKY58*, *WRKY62*, *WRKY64* and *WRKY76* as potential candidate genes for resistance to panicle blast, as they showed higher expression only in the resistant genotype against the virulent strain. Thus, we demonstrated that RiceMetaSysB can play an important role in providing robust candidate genes for rice blast and BB.

Database URL: <http://14.139.229.201/ricemetasysb/>

Introduction

Rice (*Oryza sativa* L.) remains a staple food crop for more than half of the world's population, while it is the means of livelihood for millions of poor farmers in South East Asia. Therefore, production and productivity of rice have a huge impact on world peace and economy. Still, every year a large amount of rice produce is ruined by several biotic and abiotic stresses undermining sustainable rice production. The major abiotic stresses that affect rice production are drought, heat, cold, salinity, flooding and high or low light intensity, while the major biotic stresses are diseases caused by bacteria, fungi, viruses, nematodes and infestation by insect pests. Among the various pathogens affecting the rice crop, the most important ones are *Magnaporthe oryzae* causing rice blast disease and *Xanthomonas oryzae* pv. *oryzae* (Xoo) causing bacterial blight (BB) (1–3). *M. oryzae* is a part of a species complex and causes disease on various grasses and other related species, such as barley, wheat and finger millet (4). It is a highly adaptive pathogen that can affect plants across all developmental stages and tissues, but the leaf and panicle blast cause huge yield losses, sometimes up to 100% (5). The causal organism of BB Xoo is highly specific to rice plant and can reduce crop production up to 70% under severe epidemics (6–8). This is a vascular pathogen and enters the plant via hydathodes and wounds. First, it reproduces in the intercellular space beneath epithelial cells and then spreads via xylem tissue leading to the disease (9). The best way to tackle these diseases is the use of resistant cultivars with one or more R genes. In rice, till date more than 100 R genes providing genetic resistance to rice blast have been identified, of which 24 have been cloned and validated. All these R genes are of Nucleotide binding site-leucine rich repeat (NBS-LRR) type except one (10–11). Similarly, 41 R genes (i.e. 29 dominant and 12 recessive) conferring resistance to BB have so far been identified based on disease responses to different Xoo races (12).

Transcriptome studies provide an important means to study the functional elements involved in stress response, which ultimately reveal the transcriptional regulation of genes under different growth stages, and tissues (13). Expression studies on rice lines carrying different blast

resistance genes (*Pia*, *Pish*, *Pi9*, *Pi1* and *Pi54*) revealed the role of a plethora of transcriptional regulons acting through multiple signaling kinases and transcription factors, upon infection by different *Magnaporthe* strains (14–20). Likewise, many transcriptomes studies of R gene-introgressed BB-resistant lines were generated to understand the resistant gene networks under disease infection (21–23). The major regulons identified from such studies represented disease resistance genes, pathogenesis-related proteins, enzymes of oxidative burst and LRR domains, signaling molecules, transcription factors (MYB, ERF, WRKY etc.) and cell wall modification-related genes.

With the accumulation of a large number of genome-wide gene expression data sets in the public domain, different databases were developed in rice on either a specific stress or all stresses like RiceSRTFDB for salinity stress (<http://www.nipgr.res.in/RiceSRTFDB.html>), drought stress gene database (<http://pgsb.helmholtz-muenchen.de/droughtdb/>) and plant stress gene database (<http://ccbb.jnu.ac.in/stressgenes/>). Recently, we have also developed RiceMetaSys database for identification and analysis of differentially expressed genes (DEGs) under drought and salt stresses in rice (24). All these databases catered to one or more abiotic stresses. Thus, there is no database that comprehensively covers gene expression profiles of major rice diseases across pathogen strains and host genotypes. Even those databases, which cater to all stresses and developmental stages, such as OryzaExpress, RicePLEX and RiceXPro, did not allow meta-analysis of multiple expression data sets under one common theme (24). Though ROAD database allowed meta-analysis, it is no more functional (24). Moreover, in ROAD one had to search for data sets for meta-analysis and carry out the same afresh. Hence, in the present study we have developed a database, entitled RiceMetaSysB (B for biotic stress), on rice blast and BB responsive genes (RGs) using meta-analysis of publicly available microarray experiments. We have further demonstrated the utility of our database by characterizing a major transcription factor governing plant disease reactions, WRKY, under panicle blast infection in resistant and susceptible rice varieties infected with virulent and avirulent strains of the blast pathogen.

Database construction and content

Data source

Expression microarrays are useful for identifying genes specific to different tissues and environmental conditions. Affymetrix and Agilent microarrays are the two major microarray technologies used for gene expression studies. Affymetrix arrays are single-channel systems while Agilent has both single- and two-channel microarray systems. To construct RiceMetaSysB, data from 241 and 186 samples on rice blast and BB stress were collected from NCBI GEO (Gene Expression Omnibus) database (<https://www.ncbi.nlm.nih.gov/geo>), respectively. Expression data for rice blast stress were obtained from nine genotypes {Nipponbare, Nipponbare [Pia], Nipponbare [Pish], Lijiangxintuanheigu (LTH), IRBL18 [Pi9 near isogenic line (NIL) of LTH], IRBL22 [Pi9 NIL of LTH], Taipei 309, transgenic lines of TP309:OX-Pi54 and OX-OsWRKY28} while that of BB were from five genotypes (Nipponbare, IR24, IRBB5, IRBB7 and Osaba1-1), representing different strains, inoculation time points, growth stages and tissues. Detailed information of the data sets is provided in [Supplementary data Tables S1 and S2](#).

Meta-analysis of gene expression arrays

For uniformity of the data sets that were generated across experiments in different laboratories, preprocessing (background correction) was carried out for all data sets. Preprocessing was done using robust multi-array average method for Affymetrix data sets (25). For single-channel microarray Agilent data sets, normalization was done using quantile method while for two-channel microarray systems, normalization was done using a quantile method (26). Limma package version 3.28.21 from R package was used for gene expression analysis (27). For filtering the robust candidate genes, we have used parameters adjusted P -value < 0.05 , average expression > 8 and $\log_2FC < -1$ or > 1 for both the biotic stresses. Data processing workflow was as per Sandhu *et al.* (24) and is given in [Supplementary data Figure S1](#).

Database construction

Affymetrix IDs were converted into MSU and Rice Annotation Project (RAP) IDs by using OryzaExpress (http://bioinf.mind.meiji.ac.jp/OryzaExpress/ID_converter.php). Physical position and annotations of the identified biotic RGs were retrieved from RGAP (Rice Genome Annotation Project; <http://rice.plantbiology.msu.edu/>) and RAP (<https://rapdb.dna.affrc.go.jp/>) databases. BatchPrimer3 tool was used both for identification of microsatellites and designing primers for their amplification (28). RiceMetaSysB database

was hosted under XAMPP (Apache, MariaDB, PHP and Perl) server in the Windows operating system. PHP5 and HTML5 were used for constructing the front end while MySQL was used at the back end. CSS, JQuery and JS were used for providing a better interface of the database. Chart.js was used to generate interactive graphs of the expression profiles of DEGs across/within genotypes and strains. The URL of the database is <http://14.139.229.201/ricemetasysb/>.

Differentially expressed WRKY gene search for rice blast from database

We used MSU IDs of all WRKY genes as input to retrieve the differentially expressed (DE) WRKY genes using 'LOC-ID search' option under rice (leaf) blast RGs. The output was downloaded in excel format and rearranged on the basis of log fold change (FC) values. For further validation by expression analysis, those WRKY genes that had a significant expression value of $\log_2FC \geq 5$ and common with the panicle blast-specific WRKY genes identified from the transcriptome data generated in our laboratory using a resistant genotype Tetep and susceptible genotype HP2216 (unpublished data) were selected.

Analysis of structural variations in WRKY genes

We had earlier generated low coverage whole genome sequence (WGS) of rice genotypes Tetep and HP2216 in our laboratory. We used this WGS data to identify structural variations at nucleotide as well as amino acid level in the selected blast-responsive WRKY genes. We also looked for variations in 1 kb promoter region of these genes. For identification of contigs of Tetep and HP2216 containing WRKY genes, blast analysis was carried out in Bioedit (29) offline software using sequences of selected WRKY genes from RGAP (<http://rice.plantbiology.msu.edu/>). Obtained contigs were used for identification of WRKY genes using Fgenesh software (<http://www.softberry.com/>). After confirmation of the selected WRKY genes for their identity, gene sequences along with 1 kb promoter regions were extracted from the contigs. The nucleotide and amino acid sequences of selected WRKY genes from Nipponbare, Tetep and HP2216 were used for multiple sequence alignment using ClustalW (<http://www.ebi.ac.uk/Tools/msa/clustalw2/>). Phylogenetic relationships among these sequences were visualized using Dendroscope 3 (30). A total of 1 kb promoter analysis was carried out using a motif-based sequence analysis tool, the MEME suite (<http://meme-suite.org/>) for identification of novel elements and PlantPAN 2.0 (31) (<http://plantpan2.itps.ncku.edu.tw/>) for identification of known promoter elements.

Plant material and pathogen used

To validate *Magnaporthe*-responsive WRKY genes identified, we used two contrasting rice lines, a resistant genotype RIL4 [a recombinant inbred (RI) line derived from Tetep × HP2216 cross] and a susceptible genotype HP2216. One month old healthier plants raised in the rice nursery were transferred to pots and kept in the blast phenotyping facility under controlled environmental conditions (28–32°C temperature and relative humidity of 60%) with watering at frequent intervals. These plants were infected at booting stage in the neck of panicle with *M. oryzae* spores using syringe injection method. Two *M. oryzae* strains were used for infection: a virulent strain Mo-ni-025 and an avirulent strain Mo-ni-075. Panicle tissues were collected after 6, 12 and 24 h post-infection (hpi) along with mock samples injected with water.

RNA isolation and expression analysis using quantitative real-time polymerase chain reaction

Samples were frozen in liquid nitrogen and stored at –80°C. RNA was isolated from each sample in two biological replicates using Spectrum™ Plant Total RNA Kit (Sigma-Aldrich, USA) as per manufacturer's instructions. RNA was reverse-transcribed with random primers using High-Capacity cDNA Reverse Transcription kit (Applied Biosystems, USA). For quantitative real-time polymerase chain reaction (qRT-PCR), *Actin* gene was used as normalizer. Primer sequences of WRKY and *Actin* genes are given in [Supplementary data Table S3](#). Each biological replicate had three technical replicates for transcript abundance studies. qRT-PCR was done in the total volume of 10 µl with 150 ng cDNA, 5 µl SYBR green (Brilliant III Ultra-Fast SYBR® Green QPCR Master Mix by Agilent Technologies, Santa Clara, CA, USA), 0.4 µl of ROX dye and 10 µM of each primer. Amplification was carried out in 96-well plates using LightCycler 480II machine by Roche (Basel, Switzerland) and the cycling conditions used were 95°C for 30 s, 58°C for 15 s and 72°C for 20 s, for 40 cycles. The melting point determination or dissociation curves were studied and analyzed at the end of PCR cycles.

Results

Statistics of RiceMetaSysB database

We identified 15 135 DEGs for rice blast disease from 10 GEO microarray data sets and 7475 DEGs for BB disease from 5 GEO microarray data sets. Of these, 6243 genes were found to be common to both rice blast and BB suggesting that molecular response to rice blast is more complicated than that of BB ([Figure 1](#)). Gene ontology

analysis revealed that 41%, 41% and 18% of the rice blast DEGs were involved in biological process (BP), cellular component (CC) and molecular function (MF), respectively. Similarly, in case of BB, 39%, 44% and 17% of the DEGs were present under BP, CC and MF classes. In case of common disease RGs, a large number of them encoded undefined expressed proteins, followed by adenosine triphosphate (ATP) binding, nucleotide binding and transporter activity ([Figure 1](#)). Under rice blast infection, the number of downregulated genes was more than that of upregulated genes. Contrastingly under BB infection, higher number of DEGs was found to be upregulated. Most often used tissue, strain and time points in blast and BB studies are presented in [Figure 2](#). To facilitate breeders to follow the inheritance of these genes in their breeding populations, we identified SSR motifs in the DEGs of both rice blast and BB RGs and designed primers for the same. SSRs found in rice blast DEGs amounted to 9365 while in BB DEGs it was 5375. Further, SSRs were classified based on the motif length like di, tri, tetra, penta and hexa ([Figure 3](#)). Trinucleotide repeats were the most frequent SSR motifs in both rice blast and BB RGs.

Important features of database

Freely accessible and user-friendly web interfaces can help researcher to identify most robust candidate genes. Separate tabs/links have been designed to represent rice blast and BB DEGs under the RiceMetaSysB database ([Figure 4](#)). The common genes to rice blast and BB can be viewed separately from the homepage. Besides the identity, tissue and stage of the genotype used for study, in case of biotic stresses, strain used for infection and time points [days post-inoculation (dpi)/hours after inoculation (hai)] chosen for sampling for gene expression are also important. Hence, search options specific to each one of these parameters have been incorporated in the database so that users can retrieve candidate genes based on their need and objectives. All query outputs are in tabular format that can be retrieved in either 'pdf' or 'spreadsheet' format. There is a provision to sort the results based on FC, LOC ID (gene locus ID) and nature of regulation. Further, to make the results biologically more meaningful, candidate genes between disease-resistant and susceptible genotypes have been enabled. The user can also retrieve DEGs common to two or more strains or varieties. Physical position search option can facilitate researchers to retrieve DEGs in a particular chromosomal location of interest in the genome with the help of chromosome number and start and end positions. This search facility would be a valuable one for those following forward genetic approaches. Locus search facility provides the user to identify gene expression with

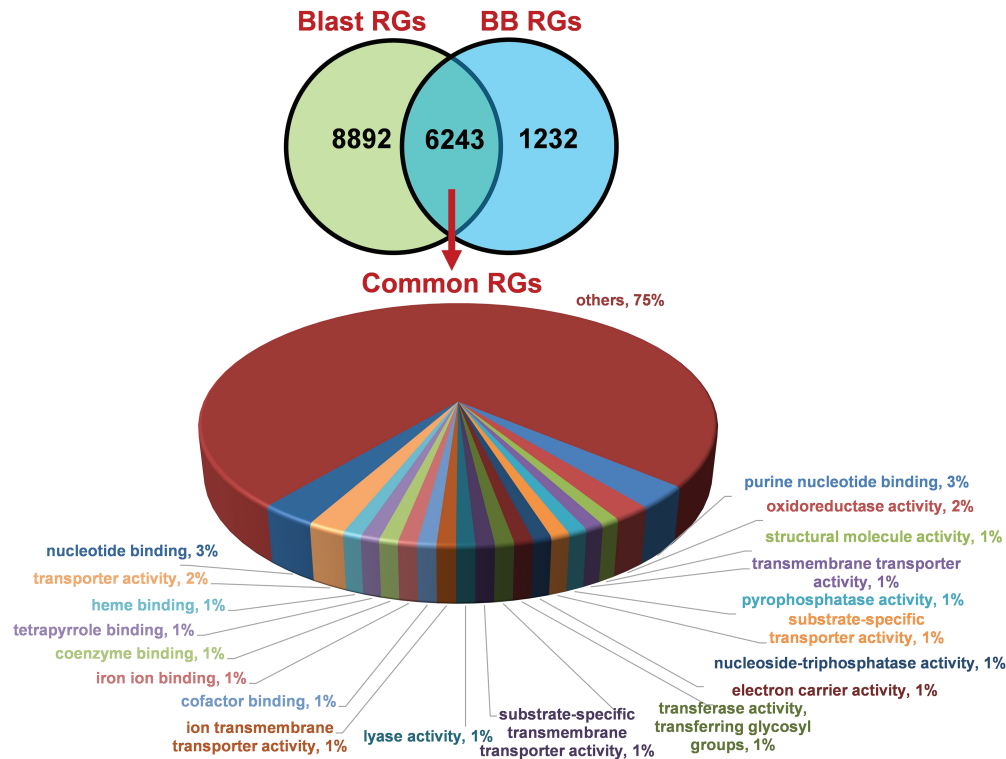


Figure 1. Distribution and functional annotation of common 6243 DEGs under rice blast and BB RGs.

the input of LOC IDs. Promoters of the selected DEGs can also be retrieved from the ‘DEG promoter’ tab. SSR search is enabled in this database through LOC IDs as input. The result window provides information on the SSR motif and forward and reverse primers. SSR query LOC IDs are hyperlinked to ISM-ILP database. Both locus search and SSR search options allow the users to retrieve batch queries. Further, external links have been provided for ISM-ILP database and pathway analysis and enrichment tools. The user can generate a graphical representation of up to 10 DEGs from single or multiple genotypes and strains. This option is very useful to visualize the expression level among genotypes and strains.

Selection of DE WRKY genes from rice blast RGs

To demonstrate the utility of RiceMetaSysB database, we chose WRKY genes. To retrieve the DE WRKY genes from the database, locus IDs of all WRKY genes present in rice were submitted in ‘Locus ID’ search tab under the DEG panel (Figure 4). From 102 MSU IDs of WRKY genes submitted, a total of 316 entries were retrieved. These entries were in the tabular format with columns representing MSU and RAP IDs of the gene along with its annotation, growth stage, tissue and variety used in the experiment, pathogen strain and duration after which samples were collected and log FC and direction of regulation (up or down)

of the gene. Sorting as per expression levels ($\log_{2}FC > 5$) recorded 47 entries. After filtering the list and removing the duplicate entries, 18 WRKY genes from the publically available leaf blast experiments finally remained (Table 1). From the panicle blast transcriptome experiment carried out in our laboratory in two rice varieties Tetep (resistant) and HP2216 (susceptible), we had earlier identified 12 WRKY genes with higher and upregulated expression in the blast-resistant line (data not shown). Between these two sets of WRKY genes, 8 were common to both leaf blast and panicle blast experiments while 10 were unique to leaf blast experiments and 4 to panicle blast experiments. Since we decided to check their expression in neck or panicle blast, eight WRKY genes common to both data sets and four genes unique to the panicle blast experiment were selected for expression analysis (Supplementary data Figure S2).

Variations in the candidate WRKY genes at nucleotide level

Significant structural differences were observed in the WRKY genes between Tetep and HP2216. The information about single nucleotide polymorphisms (SNPs) and Insertions/Deletions (InDels) present among the two genotypes along with the reference, Nipponbare, is given in Supplementary data Table S4. WRKY64 had the most number of variations, with 40 base substitutions and 5 InDels followed by WRKY23 which had 15 base

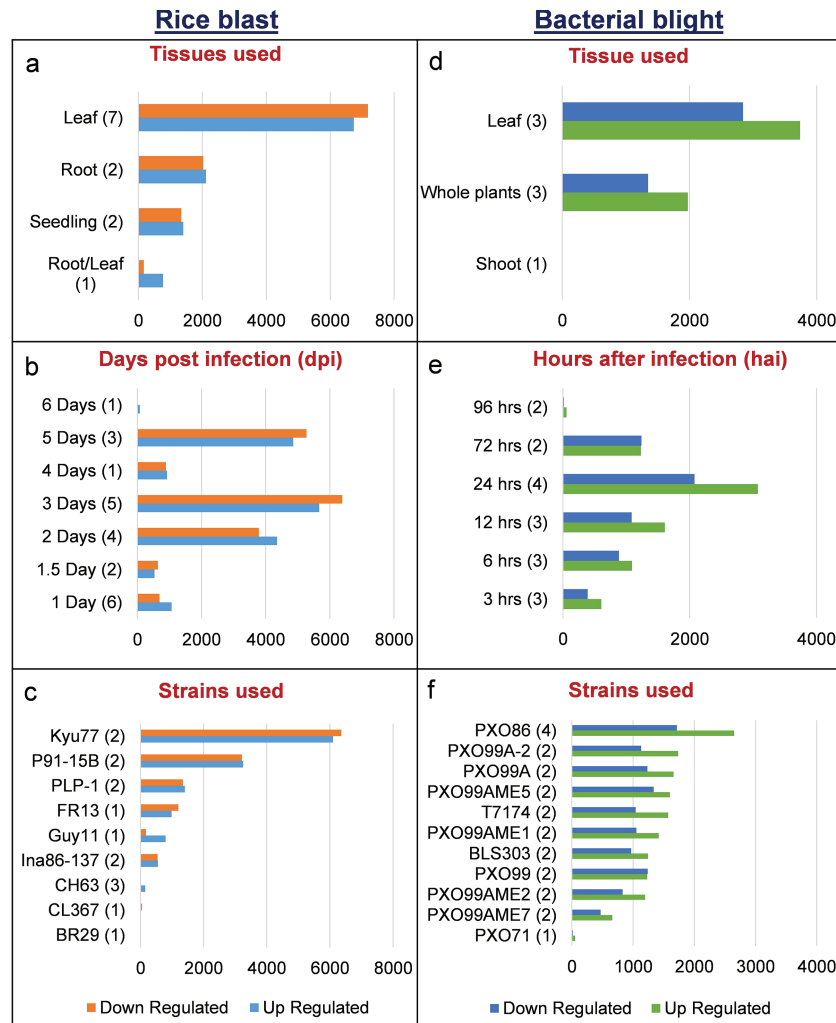


Figure 2. Distribution of DEGs in RiceMetaSys (biotic stress). **(A–C)** Distribution of rice blast RGs across tissues,dpi and strains. **(D–F)** Distribution of bacterial leaf blight RGs across tissues, hpi and strains.

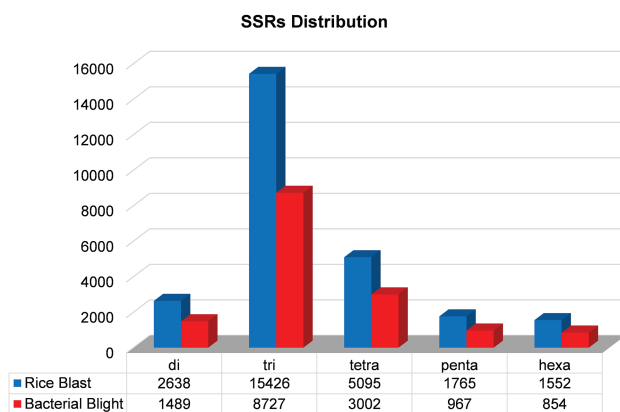


Figure 3. Distribution of microsatellites in the rice blast and BB RGs.

substitutions and 8 InDels. Among these eight InDels, seven were deletions while only one was an insertion in the susceptible genotype HP2216. *WRKY77* was the least variable with just one deletion (CAC) in HP2216 followed

by *WRKY76* and *WRKY28* both of which had two base substitutions and one deletion in HP2216. Overall, across the 12 *WRKY* genes, there were 27 InDels with 19 insertions and 8 deletions in the rice blast-resistant genotype Tetep. We also generated phylogenetic tree where three *WRKY* genes namely, *WRKY26*, *WRKY28* and *WRKY76* were evolutionary different in Tetep and HP2216 (Figure 5A). Of these three genes, *WRKY26* and *WRKY28* had five InDels each, with five and two substitutions, respectively.

Variations in *WRKY* genes at protein level

All the 12 selected *WRKY* genes have conserved *WRKY* and zinc finger domains in them in both the genotypes, except *WRKY58* which had no zinc finger domain. Two *WRKY* domains were present in *WRKY70* (Type I) whereas single *WRKY* domain was present in rest of the 11 genes. The sequence of *WRKY* domain, *WRKYGQK*, was

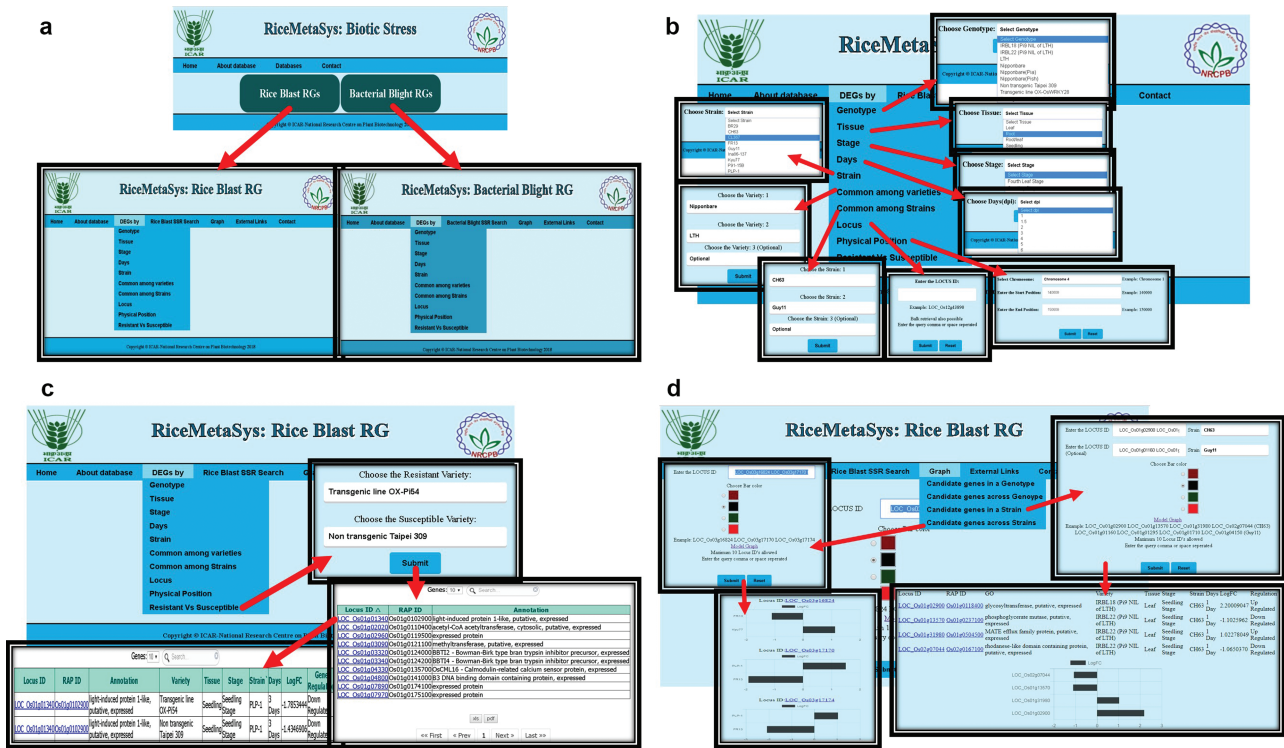


Figure 4. (A) Homepage of the RiceMetaSysB (biotic stress) database and two separate links for rice blast and BB RGs. (B) Different search options available in the database like genotype, tissue, stage, days, strain, common among varieties/strains, locus, physical position and resistant vs. susceptible genotypes search. (C) Resistant vs. susceptible search input, its output and hyperlink of locus search. (D) Interactive graph results of candidate genes in single and multiple strains.

conserved in all proteins except in WRKY07, WRKY26 and WRKY77, all of which had WRKYG(K)K. The CCHH zinc finger domain was present in all WRKYs except WRKY19, WRKY64 and WRKY79 where CCHC zinc finger domain (Type III) was present. Standards prescribed by Wu *et al.* (32) and Ross *et al.* (33) have been followed in categorizing the WRKY genes. The variations in amino acid sequences among the 12 WRKY proteins are summarized in Supplementary data Table S5. WRKY76 had two larger (92 and 22 amino acid long) and one smaller deletion (6 amino acid long) in HP2216 besides 8 amino acid substitutions (Supplementary data Figure S3). WRKY26 had 22 variations which included 2 larger InDels, 4 smaller InDels and 16 amino acid substitutions. WRKY64 also had 22 variations of which 21 were single amino acid substitution, and the remaining one was an amino acid deletion in Tetep. WRKY77 had only one amino acid deletion in Tetep while WRKY23 had only one substitution. Both WRKY28 and WRKY79 had two amino acid substitutions. WRKY7, WRKY19 and WRKY58 each had one substitution and one InDel. Phylogenetic analysis showed that three WRKY genes namely WRKY79, WRKY64 and WRKY19 are evolutionary different from the rest of the WRKY genes at amino acid level. Further, within-gene variation was observed in WRKY26, WRKY64, WRKY76 and WRKY28,

which had multiple variations in their amino acid sequences (Figure 5B).

Variations in 1 kb promoter regions of WRKY gene

Considerable variations in the form of SNPs and InDel were observed in the promoter regions. Four variations were observed in the promoter region of WRKY76. Three variations were observed in the promoter of WRKY64 whereas two variations were present in the promoter region of WRKY19. WRKY26 and WRKY58 both had one variation each in the promoter region. These structural variations are shown in Supplementary data Figure S4. Phylogenetic analysis showed that promoter sequences of WRKY19, WRKY26 and WRKY64 were completely different after 505, 406 and 554 bp upstream from ATG in Tetep and HP2216, respectively (Figure 5C). These promoter sequences were further subjected to search for unknown motifs and six such motifs were identified. We observed that the number and positions of motifs were conserved in all three genotypes in all 12 promoters except the absence of motif MEME1 in the promoter of WRKY76 in HP2216 and absence of motif MEME3 in the promoter of WRKY64 in Nipponbare

Table 1. WRKY genes identified from RiceMetaSysB database under blast infection with upregulation of 5 or above Log FC

Locus ID	Annotation	Variety	Strain	Days	LogFC
LOC_Os01g09100	WRKY10	Nipponbare	Kyu77	5	5.6
		Nipponbare (Pish)	Kyu77	5	5.4
LOC_Os01g14440	WRKY1	Nipponbare	FR13	4	5.1
LOC_Os01g40260	WRKY77	Nipponbare (Pish)	Kyu77	3	6.1
		Nipponbare (Pish)	Kyu77	2	6.0
		Nipponbare (Pia)	P91-15B	2	5.6
		Nipponbare	P91-15B	3	5.3
		Nipponbare	Kyu77	5	5.2
LOC_Os01g43650	WRKY11	Nipponbare	Kyu77	5	5.8
LOC_Os01g51690	WRKY26	Nipponbare (Pish)	Kyu77	2	6.6
		Nipponbare (Pish)	Kyu77	3	6.1
		Nipponbare	Kyu77	3	5.7
LOC_Os01g53040	WRKY14	Nipponbare	Kyu77	5	5.0
LOC_Os01g61080	WRKY24	Nipponbare	Kyu77	5	6.9
		Nipponbare	Guy11	2	5.4
LOC_Os02g08440	WRKY71	Nipponbare	FR13	4	5.0
LOC_Os02g26430	WRKY42	Nipponbare	Kyu77	5	7.0
		Nipponbare (Pish)	Kyu77	3	6.3
LOC_Os02g53100	WRKY32	Nipponbare (Pish)	Kyu77	5	6.5
		Nipponbare	Kyu77	5	5.5
LOC_Os03g21710	WRKY79	Nipponbare (Pish)	Kyu77	2	5.3
LOC_Os05g39720	WRKY70	Nipponbare	Kyu77	5	8.2
		Nipponbare (Pish)	Kyu77	3	6.4
		Nipponbare	Kyu77	3	6.4
		Nipponbare	Kyu77	5	7.5
LOC_Os05g49620	WRKY19	Nipponbare (Pish)	Kyu77	2	7.4
		Nipponbare (Pish)	Kyu77	3	7.2
		Nipponbare	Kyu77	3	6.8
		Nipponbare (Pish)	Kyu77	5	6.3
		Nipponbare	P91-15B	1	5.6
		Nipponbare (Pia)	P91-15B	2	5.5
		Nipponbare	Ina86-137	2	9.8
LOC_Os06g44010	WRKY28	Nipponbare	Kyu77	5	5.9
		Nipponbare	Kyu77	3	5.3
		Nipponbare	Kyu77	5	6.6
LOC_Os09g25070	WRKY62	Nipponbare	Kyu77	3	5.4
		Nipponbare	Kyu77	5	6.6
LOC_Os11g02530	WRKY40	Nipponbare (Pish)	Kyu77	2	6.8
		Nipponbare (Pish)	Kyu77	3	6.4
		Nipponbare (Pish)	Kyu77	5	6.3
		Nipponbare (Pia)	P91-15B	2	6.0
		Nipponbare	Kyu77	3	5.4
LOC_Os11g29870	WRKY72	Nipponbare (Pish)	Kyu77	2	5.1
LOC_Os12g02450	WRKY64	Nipponbare (Pish)	Kyu77	2	6.6
		Nipponbare (Pia)	P91-15B	2	6.2
		Nipponbare (Pish)	Kyu77	3	6.2
		Nipponbare (Pish)	Kyu77	5	6.2
		Nipponbare	Kyu77	3	5.1

(Supplementary data Figure S5). We also checked the known AP2, bHLH, bZIP, WRKY, MYB, HSF, Dehydrin, ERF and NAC transcription factors' binding domains in the promoter sequences using PlantPAN software. The details of different motifs present in the promoters are shown in

Supplementary data Table S6. Interestingly, the susceptible genotype HP2216 had the promoter elements identical to that of Nipponbare, which is also susceptible to rice blast, while the resistant genotype was very different in its *cis*-element composition with the absence of 17 elements

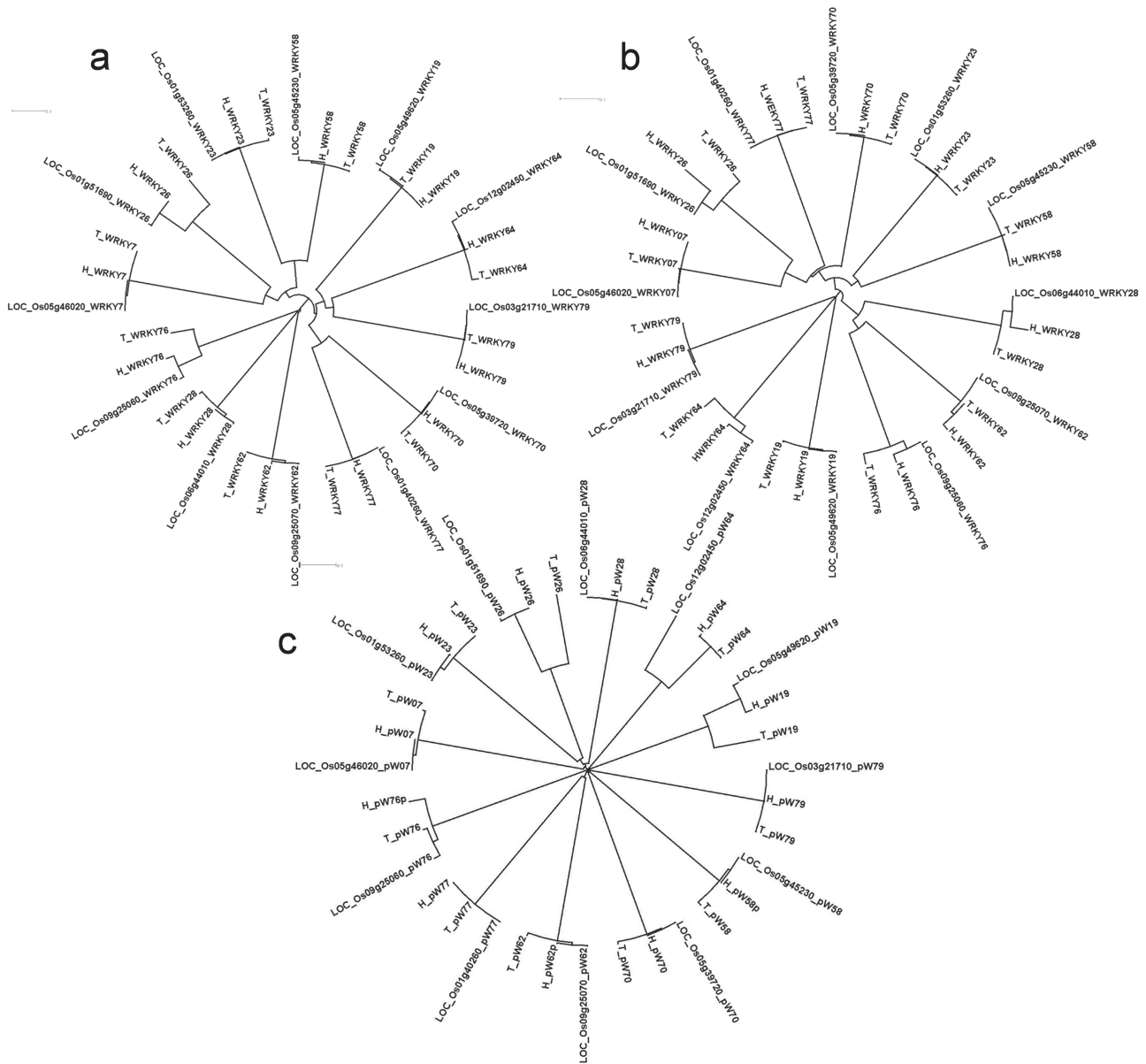


Figure 5. Phylogenetic analysis of 12 DE WRKY genes in blast experiments using (A) complete genes, (B) proteins and (C) 1 kb promoter regions in three genotypes Tetep, HP2216 and Nipponbare using ClustalX2.1 tool.

(5 dehydrin, 5 bHLH, 4 bZIP, 2 WRKY and 1 HSF) and presence of two new elements (one each of AP2 and HSF).

Expression analysis of WRKY genes under panicle blast

Out of 12 genes, we could get quantifiable amplification in 9 genes (Figure 6). As the blast disease-resistant genotype Tetep did not flower in our experiment, we used its derivative RIL4 that showed resistance reaction akin to its parent Tetep. Though the expression of WRKY19 gene was observed in both the genotypes, transcript abundance was more in case of resistant genotype RIL4, which increased with increasing time course, particularly against infection

with the virulent pathogen. The gene WRKY23 also had a very high level of expression in resistant RIL4 compared to susceptible HP2216 again in which increase in expression level with increase in time period in response to infection by the virulent pathogen was observed. Similarly, WRKY62, WRKY07, WRKY76 and WRKY64 had a very high level of expression in RIL4 against the virulent pathogen strain at 12 hpi. The gene WRKY70 also had a very high level of expression in RIL4, but the expression was the highest against the avirulent pathogen after 24 hpi. The gene WRKY58 had a moderate expression in susceptible genotype HP2216 against the virulent pathogen strain, which decreased sharply after 6 hpi; however, in the resistant genotype it had a very high level of expression against the

virulent strain. In case of the gene *WRKY28*, the highest expression was observed in RIL4 against both the avirulent as well as virulent strains.

Discussion

With the advent of microarray and RNA-seq technologies, a huge amount of genome-wide expression data is available in rice, especially under biotic stress conditions. Besides the use of resistant genotypes or a contrasting pair of genotypes (in terms of their disease response) for studying gene expression under disease infection, a large number of transgenic lines and NILs transformed/introgressed with major R genes have been used for characterization of response to disease infection in rice (10, 18, 20, 34). Most of the publicly available databases like OryzaExpress, RicePLEX, ROAD, RiceSRTFDB, QlicRice, OryGenesDB, RiceXPro and qTeller and commercial platforms like GeneMapper and Genevestigator developed either cater to abiotic stresses or gene/QTL (quantitative trait loci) mapping projects and the features of the same have been discussed in detail in Sandhu *et al.* (24). Recently, the gene expression modules and regulatory networks under various abiotic stresses in rice have been elucidated through a comprehensive database called EGRIN (35). As none of these databases were either breeder-friendly or amenable for any direct utilization in crop improvement, we initially developed a breeder-friendly database, named RiceMetaSys, for DEGs under drought and salt stresses in rice (24). Now we extended our work on abiotic stresses to the two most important biotic stresses of rice, blast and BB. For this, all the available microarray sets related to blast and BB in rice were collated and normalized, so that these experiments can be brought on a common platform (36). As our biotic stress-based database (unlike the abiotic one) included the Agilent arrays too, a huge number of samples were analyzed amounting to 241 samples for rice blast and 186 for BB giving rise to 15 135 and 7475 disease RGs, respectively. Following our earlier database, we named the current one as RiceMetaSysB for biotic stresses. In addition to the search options provided in the earlier database for retrieval of candidate genes specific to genotype, tissue and developmental stage of the host, the current one carries options based on strain and hours/dpi as well as ‘common genes among varieties’ and ‘strains’ to identify most robust candidate genes. Similarly, the 2D representation of the data in the form of graph provides expression profiles not only across genes and genotypes but also across strains. Further, this database provides candidate gene-specific SSR and intron-specific length polymorphism for marker-assisted breeding of the robust DEGs.

In the earlier version of the database, we could not demonstrate the practical utility (24). Hence, here we iden-

tified blast-responsive WRKY genes using RiceMetaSysB database. There are more than 100 WRKY transcription factors present in both japonica and indica rice, out of which many WRKYs are shown to be DE during blast disease development in rice (18, 37–38). Several WRKY genes like *OsWRKY13*, *OsWRKY31*, *OsWRKY45*, *OsWRKY53*, *OsWRKY71*, *OsWRKY89*, *OsWRKY76* and *OsWRKY67* have been functionally characterized for resistance to blast or BB by generating transgenics (18, 39–46). From a wide array of experiments, using our DB, we could identify 18 blast-responsive WRKY genes.

Using an RI population from the blast-resistant Tetep and susceptible HP2216, *Pi54* gene has been already identified and introgressed in other cultivated lines for blast resistance (47). For our validation studies, we used these two genotypes and an RI line derived from these two genotypes. We observed the expression of *WRKY19* in both the genotypes with both virulent and avirulent strains demonstrating its utility as a biomarker to ascertain *Magnaporthe* attack on rice crop. *WRKY23*, *WRKY28* and *WRKY70* were *Magnaporthe*-responsive but did not contribute to blast resistance as they were upregulated in the resistant line but with higher expression in response to the avirulent strain than the virulent strain. On the other hand, *WRKY7*, *WRKY58*, *WRKY62*, *WRKY64* and *WRKY76* were found to be the promising candidate genes providing resistance to panicle blast as they are highly expressed only in the resistant line, that too in response to the virulent strain. When we checked these genes in the literature, we found that *WRKY28*, *WRKY62* and *WRKY76* are negative regulators of rice blast (17, 45, 48). This was surprising for us as we found its higher expression in resistant genotypes, that too in response to virulent strain. Then we checked its structural variation and we found that proteins of *WRKY26*, *WRKY28*, *WRKY62*, *WRKY64* and *WRKY76* are very different in Tetep and HP2216 with large number of amino acid substitutions and InDels. This shows that probably alleles from Tetep are contributing for resistance whereas its alternative allele may be contributing for susceptibility. We have scope to conclude this because in earlier study by Tao *et al.* (44) regarding *WRKY45* from rice showed that a pair of alleles of *WRKY45* plays opposite roles in rice–bacteria interaction. Such a conclusion is also supported by observation that the *cis*-element composition of the resistant genotype, Tetep, was completely different from that of the susceptible genotypes, HP2216 and Nipponbare. As many as five bHLH and five dehydrin elements were lost in Tetep indicating that they do have a role in inducing susceptibility or losing resistance. The role of dehydrin, an important transcription factor in abiotic stress tolerance, has recently been implicated in biotic stress tolerance (49). However, we can conclusively comment on the contribution of these

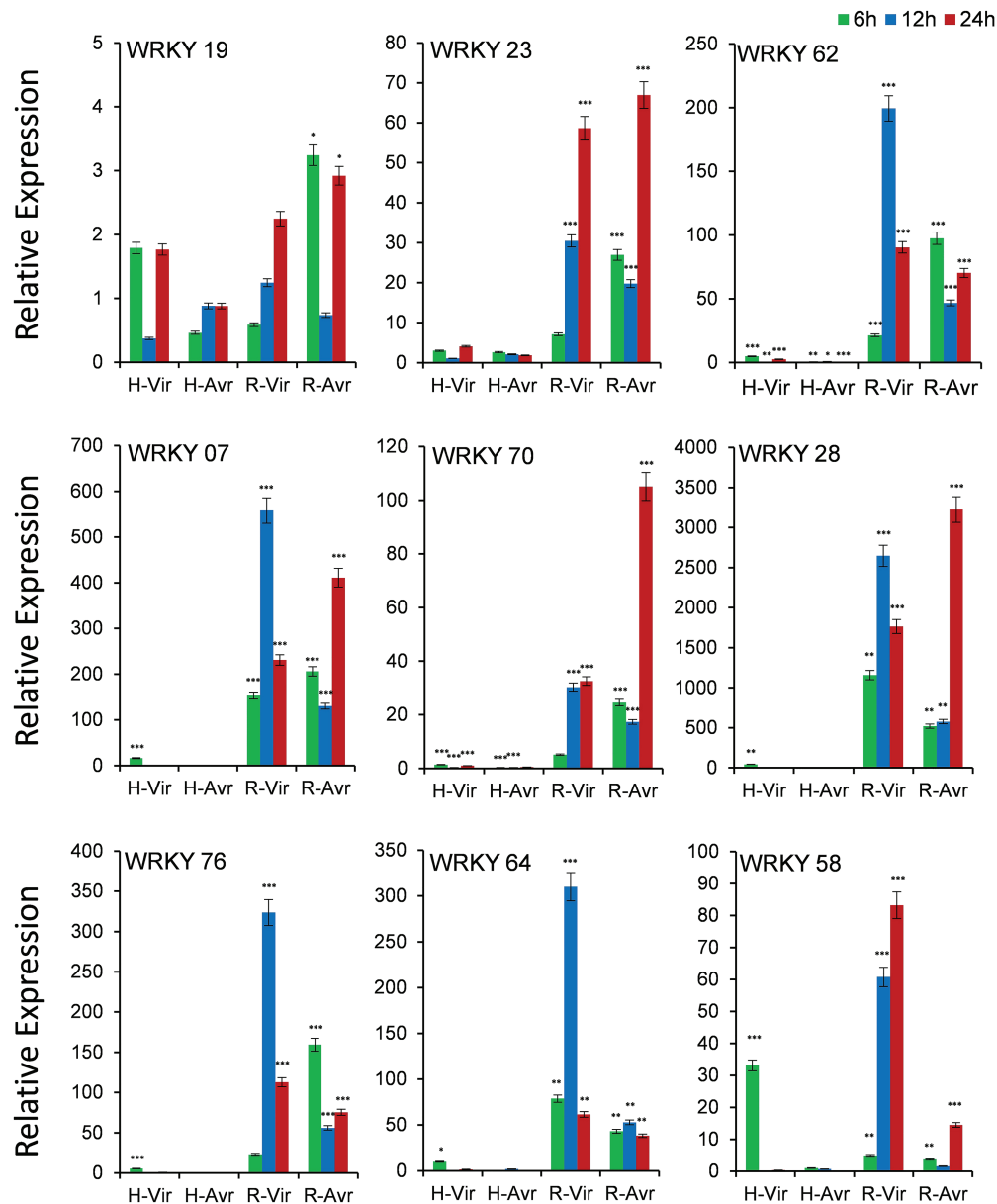


Figure 6. Expression analysis of WRKY genes in blast-infected panicle tissues at three-time intervals. T means Tetep; H, HP2216 with virulent and avirulent *M. oryzae* strains. Asterisks indicate a significant difference compared with untreated at $P < 0.05$, by Student's *t*-test.

genes or *cis*-elements only after their complementation test by developing transgenics. However, from this analysis we are able to show that RiceMetaSysB database can play a very important role in providing robust candidate genes responsible for rice blast and BB stresses.

Conclusion and future prospects

RiceMetaSysB is an open source and user-friendly web interface for identification of most robust genes for rice blast and BB diseases. Physical position and SSR search options are expected to be more meaningful for breeders

while other options will be useful for pathologists and molecular biologists working on these two diseases. The utility of the database has been demonstrated in our study which identified *WRKY7*, *WRKY58*, *WRKY62*, *WRKY64* and *WRKY76* to be potential candidate WRKY genes for contributing toward panicle blast resistance in rice. More candidate genes belonging to other gene families can also be identified this way and used in functional genomics and crop improvement.

Supplementary data

Supplementary data are available at *Database* online.

Acknowledgements

The Rajiv Gandhi National Fellowship from University Grant Commission, Government of India, for V.K. and Indian Council of Agricultural Research Junior Research Fellowship from Indian Council of Agricultural Research, Government of India, for B.D. are acknowledged.

Funding

Indian Council of Agricultural Research (ICAR)—Centre for Agricultural Bioinformatics (IASRI/CABin/Scheme/3(1)); ICAR—National Research Centre on Plant Biotechnology, New Delhi.

Conflict of interest. None declared.

References

- Pennisi,E. (2010) Armed and dangerous. *Science*, **327**, 804–805.
- Dean,R., Van Kan,J.A., Pretorius,Z.A. *et al.* (2012) The top 10 fungal pathogens in molecular plant pathology. *Mol. Plant Pathol.*, **13**, 414–430.
- Mansfield,J., Genin,S., Magori,S. *et al.* (2012) Top 10 plant pathogenic bacteria in molecular plant pathology. *Mol. Plant Pathol.*, **13**, 614–629.
- Couch,B.C., Fudal,I., Lebrun,M.H. *et al.* (2005) Origins of host-specific populations of the blast pathogen *Magnaporthe oryzae* in crop domestication with subsequent expansion of pandemic clones on rice and weeds of rice. *Genetics*, **170**, 613–630.
- Scardaci,S.C., Webster,R.K., Greer,C.A. *et al.* (1997) Rice blast: a new disease in California. *Agron. Fact Sheet Ser.*, **2**, 1–2.
- Khush,G.S., Mackill,D.J. and Sidhu,G.S. (1989) Breeding rice for resistance to bacterial blight. In: *Bacterial Blight of Rice*, International Rice Research Institute, Los, Baños, Manila, Philippines, 207–217.
- Leach,J.E., Rhoads,M.L., Cruz,C.V. *et al.* (1992) Assessment of genetic diversity and population structure of *Xanthomonas oryzae pv. oryzae* with a repetitive DNA element. *Appl. Environ. Microbiol.*, **58**, 2188–2195.
- Mew,T.W. and Vera Cruz,C.M. (2001) Bacterial blight of rice. In: *Encyclopedia of Plant Pathology*. Wiley, New York, 71–74.
- Noda,T. and Kaku,H. (1999) Growth of *Xanthomonas oryzae pv. oryzae* in planta and in guttation fluid of rice. *Jpn. J. Phytopathol.*, **65**, 9–14.
- Devanna,N.B., Vijayan,J. and Sharma,T.R. (2014) The blast resistance gene Pi54of cloned from *Oryza officinalis* interacts with Avr-Pi54 through its novel non-LRR domains. *PLoS One*, **9**, e104840.
- Rawal,H.C., Mithra,S.A., Arora,K. *et al.* (2018) Genome-wide analysis in wild and cultivated *Oryza* species reveals abundance of NBS genes in progenitors of cultivated rice. *Plant Mol. Biol. Rep.*, **36**, 373–386.
- Zhang,F., Wu,Z.C., Wang,M.M. *et al.* (2017) Genome-wide association analysis identifies resistance loci for bacterial blight in a diverse collection of indica rice germplasm. *PLoS One*, **12**, e0174598.
- Wang,Y., Kim,S.G., Kim,S.T. *et al.* (2011) Biotic stress-responsive rice proteome: an overview. *J. Plant Biol.*, **54**, 219.
- Marcel,S., Sawers,R., Oakeley,E. *et al.* (2010) Tissue-adapted invasion strategies of the rice blast fungus *Magnaporthe oryzae*. *Plant Cell*, **22**, 3177–3187.
- Gupta,S.K., Rai,A.K., Kanwar,S.S. *et al.* (2011) The single functional blast resistance gene Pi54 activates a complex defence mechanism in rice. *J. Exp. Bot.*, **63**, 757–772.
- Abbruscato,P., Nepusz,T., Mizzi,L. *et al.* (2012) OsWRKY22, a monocot WRKY gene, plays a role in the resistance response to blast. *Mol. Plant Pathol.*, **13**, 828–841.
- Chujo,T., Miyamoto,K., Shimogawa,T. *et al.* (2013) OsWRKY28, a PAMP-responsive transrepressor, negatively regulates innate immune responses in rice against rice blast fungus. *Plant Mol. Biol.*, **82**, 23–37.
- Wei,T., Ou,B., Li,J. *et al.* (2013) Transcriptional profiling of rice early response to *Magnaporthe oryzae* identified OsWRKYs as important regulators in rice blast resistance. *PLoS One*, **8**, e59720.
- Tanabe,S., Yokotani,N., Nagata,T. *et al.* (2014) Spatial regulation of defense-related genes revealed by expression analysis using dissected tissues of rice leaves inoculated with *Magnaporthe oryzae*. *J. Plant Physiol. Pathol.*, **2**, doi:10.4172/2329-955X.1000135.
- Jain,P., Singh,P.K., Kapoor,R. *et al.* (2017) Understanding host-pathogen interactions with expression profiling of NILs carrying rice-blast resistance *Pi9* gene. *Front. Plant Sci.*, **8**, 93.
- Gao,L., Cao,Y., Xia,Z. *et al.* (2013) Do transgenesis and marker-assisted backcross breeding produce substantially equivalent plants? A comparative study of transgenic and backcross rice carrying bacterial blight resistant gene Xa21. *BMC Genomics*, **14**, 738.
- Cheng,X.J., He,B., Chen,L. *et al.* (2016) Transcriptome analysis confers a complex disease resistance network in wild rice *Oryza meyeriana* against *Xanthomonas oryzae pv. oryzae*. *Sci. Rep.*, **6**, 38215.
- Gao,L., Fang,Z., Zhou,J. *et al.* (2018) Transcriptional insights into the pyramided resistance to rice bacterial blight. *Sci. Rep.*, **8**, 12358.
- Sandhu,M., Sureshkumar,V., Prakash,C. *et al.* (2017) RiceMetaSys for salt and drought stress responsive genes in rice: a web interface for crop improvement. *BMC Bioinformatics*, **18**, 432.
- Irizarry,R.A., Hobbs,B., Collin,F. *et al.* (2003) Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics*, **4**, 249–264.
- López-Romero,P., González,M.A., Callejas,S. *et al.* (2010) Processing of Agilent microRNA array data. *BMC Res. Notes*, **3**, 18.
- Ritchie,M.E., Phipson,B., Wu,D. *et al.* (2015) limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.*, **43**, e47–e47.
- You,F.M., Huo,N., Gu,Y.Q. *et al.* (2008) BatchPrimer3: a high throughput web application for PCR and sequencing primer design. *BMC Bioinformatics*, **9**, 253.
- Hall,T.A. (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp. Ser.*, **41**, 95–98.
- Huson,D.H. and Scornavacca,C. (2012) Dendroscope 3: an interactive tool for rooted phylogenetic trees and networks. *Syst. Biol.*, **61**, 1061–1067.

31. Chow,C.N., Zheng,H.Q., Wu,N.Y. *et al.* (2015) PlantPAN 2.0: an update of plant promoter analysis navigator for reconstructing transcriptional regulatory networks in plants. *Nucleic Acids Res.*, **44**, D1154–D1160.
32. Wu,K.L., Guo,Z.J., Wang,H.H. *et al.* (2005) The WRKY family of transcription factors in rice and *Arabidopsis* and their origins. *DNA Res.*, **12**, 9–26.
33. Ross,C.A., Liu,Y. and Shen,Q.J. (2007) The WRKY gene family in rice (*Oryza sativa*). *J. Integr. Plant Biol.*, **49**, 827–842.
34. Das,A., Soubam,D., Singh,P.K. *et al.* (2012) A novel blast resistance gene, Pi54rh cloned from wild species of rice, *Oryza rhizomatis* confers broad spectrum resistance to *Magnaporthe oryzae*. *Funct. Integr. Genomics*, **12**, 215–228.
35. Wilkins,O., Hafemeister,C., Plessis,A. *et al.* (2016) EGRINs (Environmental Gene Regulatory Influence Networks) in rice that function in the response to water deficit, high temperature, and agricultural environments. *Plant Cell*, **28**, 2365–2384.
36. Johnson,W.E., Li,C. and Rabinovic,A. (2007) Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics*, **8**, 118–127.
37. Gao,G., Zhong,Y., Guo,A. *et al.* (2006) DRTF: a database of rice transcription factors. *Bioinformatics*, **22**, 1286–1287.
38. Ryu,H.S., Han,M., Lee,S.K. *et al.* (2006) A comprehensive expression analysis of the WRKY gene superfamily in rice plants during defense response. *Plant Cell Rep.*, **25**, 836–847.
39. Liu,X., Bai,X., Wang,X. *et al.* (2007) OsWRKY71, a rice transcription factor, is involved in rice defense response. *J. Plant Physiol.*, **164**, 969–979.
40. Qiu,D., Xiao,J., Ding,X. *et al.* (2007) OsWRKY13 mediates rice disease resistance by regulating defense-related genes in salicylate- and jasmonate-dependent signaling. *Mol. Plant Microbe Interact.*, **20**, 492–499.
41. Shimono,M., Sugano,S., Nakayama,A. *et al.* (2007) Rice WRKY45 plays a crucial role in benzothiadiazole-inducible blast resistance. *Plant Cell*, **19**, 2064–2076.
42. Wang,H., Hao,J., Chen,X. *et al.* (2007) Overexpression of rice WRKY89 enhances ultraviolet B tolerance and disease resistance in rice plants. *Plant Mol. Biol.*, **65**, 799–815.
43. Zhang,J., Peng,Y. and Guo,Z. (2008) Constitutive expression of pathogen-inducible OsWRKY31 enhances disease resistance and affects root growth and auxin response in transgenic rice plants. *Cell Res.*, **18**, 508.
44. Tao,Z., Liu,H., Qiu,D. *et al.* (2009) A pair of allelic WRKY genes play opposite roles in rice–bacteria interactions. *Plant Physiol.*, **151**, 936–948.
45. Yokotani,N., Sato,Y., Tanabe,S. *et al.* (2013) WRKY76 is a rice transcriptional repressor playing opposite roles in blast disease resistance and cold stress tolerance. *J. Exp. Bot.*, **64**, 5085–5097.
46. Liu,Q., Li,X., Yan,S. *et al.* (2018) OsWRKY67 positively regulates blast and bacteria blight resistance by direct activation of PR genes in rice. *BMC Plant Biol.*, **18**, 257.
47. Sharma,T.R., Madhav,M.S., Singh,B.K. *et al.* (2005) High-resolution mapping, cloning and molecular characterization of the Pi-k (h) gene of rice, which confers resistance to *Magnaporthe grisea*. *Mol. Genet. Genomics*, **274**, 569–578.
48. Liang,M.H. and Jiang,J.G. (2017) Analysis of carotenogenic genes promoters and WRKY transcription factors in response to salt stress in *Dunaliella bardawil*. *Sci. Rep.*, **7**, 37025.
49. Hanin,M., Brini,F., Ebel,C. *et al.* (2011) Plant dehydrins and stress tolerance: versatile proteins for complex mechanisms. *Plant Signal. Behav.*, **6**, 1503–1509.