In silico analysis for gene content in rice genomic regions mapped for the gall midge resistance genes

Article *in* Plant Omics · June 2012

CITATIONS
28

READS
281

6 authors, including:

Nidhi Rawat
University of Florida
30 PUBLICATIONS **326** CITATIONS
SEE PROFILE

Raman Sundaram
Directorate of Rice Research
100 PUBLICATIONS **2,853** CITATIONS
SEE PROFILE

Jagadish Bentur
Directorate of Rice Research
121 PUBLICATIONS **1,595** CITATIONS
SEE PROFILE

Plant Omics Journal POJ

POJ 5(4):405-413 (2012) ISSN:1836-3644

In silico analysis for gene content in rice genomic regions mapped for the gall midge resistance genes

Arun Kumar Yasala[1†], Nidhi Rawat[1†], V. S. A.Kumar Sama[1,2], HimaBindu Kudapa[1,2], R. M. Sundaram[1], J. S. Bentur[1]*

[1]Directorate of Rice Research, Rajendranagar, Hyderabad, Andhra Pradesh 500030, India
[2]Present Address: International Crops Research Institute for the Semi-Arid Tropics, Patancheru, Andhra Pradesh 502324, India

[*]Corresponding author: jbentur@yahoo.com
†Authors contributed equally

Abstract

So far, 11 gall midge resistance genes (*Gm1* through *Gm11*) have been identified in rice though classical inheritance studies and several of these have been mapped through linked markers. Nevertheless, the exact identity of these genes is still eluding. Hence, the present study aimed to identify the potential candidate genes involved in rice-gall midge interactions by *in silico* analysis for gene content of 4.02 Mb genomic regions containing the gall midge R genes. The candidate genes were identified in terms of common genes among the different gall midge resistance regions, transcriptomics data analysis from earlier studies and, reported genes widely implicated in defense against plant-pathogen interactions. Results revealed, of the 375 annotated genes, five genes with one or more copy number were common among these regions such as the genes encoding NBS-LRR class proteins, no apical meristem protein, F-box family protein, pentatricopeptide repeat containing protein and SET domain containing protein. Further, 15 genes from these regions have been reported in earlier transcriptomics studies covering in incompatible (9 genes) and compatible (6 genes) rice-gall midge interactions and 22 genes are known to be implicated in different plant-pathogen interactions. From the present study, we have shortlisted 24 genes and developed 33 primer pairs. These primers are ready to use for conducting real time PCR validation.

Keywords: Candidate gene; Gene annotation; Genome sequence; Linked markers; Physical position; Resistance gene.
Abbreviations: AFLP-amplified fragment length polymorphism; bp-base pair; cDNA-complemantary DNA; *Gm* gene-gall midge resistance gene; GMB-gall midge biotype; HR-hypersensitive reaction; Kb-kilobase pair; Mb-megabase pair; R gene-resistance gene; RAPD-random amplified polymorphic DNA; RFLP-restriction fragment length polymorphism; SSH- suppression subtractive hybridization; SSR-simple sequence repeat.

Introduction

The Asian rice gall midge, *Orseolia oryzae* (Wood-Mason) is a serious pest of rice in several countries including India. Based on the estimates from eastern and southern India, the gall midge causes an annual yield loss of about 477,000 tons of grain, which is worth US$80 million **(**Krishnaiah, 2004). Breeding and cultivation of resistant varieties has been a viable and ecologically acceptable approach for management of the pest (Bentur et al., 2003). Hence, identification and cloning of resistance genes in rice is important so that these genes can be deployed into elite varieties of rice. Till date, seven distinct biotypes of gall midge (GMB1 through GMB6 and GMB4M) and 11 non-allelic gall midge resistance genes (*Gm1* through *Gm11*) have been reported (Vijaya Lakshmi et al., 2006; Himabindu et al., 2010). At genetic level, resistance to gall midge has been found to be conferred by a single dominant or a recessive gene. Several mapping studies have been undertaken using molecular markers to find the physical position of the resistance genes on the rice genome. Out of 11 genes identified, eight (*Gm1, Gm2, gm3, Gm4, Gm6, Gm7, Gm8* and *Gm11*) are mapped onto different chromosomes of rice genome (Fig 1) and one gene (*Gm5*) is tagged with molecular markers but not yet mapped. *Gm1* is mapped within 0.18 Mb region on chromosome 9 between the SSR markers RM23941 and RM23956 (Sundaram, 2007), *Gm2* within 0.66 Mb region on chromosome 4 between the SSR markers RM17473 and RM17503 (Sundaram, 2007), *gm3* within 0.56 Mb region on chromosome 4 between the SSR markers RM17480 and RMgm3SSR4 (Sama et al., 2010), *Gm4* within 0.33 Mb region on chromosome 8 between the SSR markers RM22551 and RM22562 (Himabindu, 2009), *Gm5* is tagged using RAPD markers OPR-19, OPP-09, OPB-14, OPQ-05 and OPE-01 but not yet mapped (Katiyar et al., 2000), *Gm6* within 0.59 Mb region on chromosome 4 between the RFLP markers RG476 and RG214 (Katiyar et al., 2001), *Gm7* towards downstream on chromosome 4 using AFLP marker SA598 (Sardesai et al., 2002), *Gm8* within 0.43 Mb region on chromosome 8 between the SSR markers RM22685 and RM22709 (Sama, 2011) and *Gm11* within 1.7 Mb region on chromosome 12 between the SSR markers RM28574 and RM28706 (Himabindu et al., 2010). Rice varieties possessing gall midge resistance genes show two distinct types of resistance mechanism. *Gm2*, *gm3*, *Gm4*, *Gm5*, *Gm6*, *Gm7* and *Gm11* genes confer hypersensitive reaction associated (HR+ type) resistance wherein host cell death occurs at the site of insect attack. Whereas, *Gm1* and *Gm8* genes show hypersensitive reaction independent (HR- type) resistance with no cell death in the host (Bentur et al., 2003). Also, the *Gm* genes have varied spectrum of resistance. None of the genes show resistance against all the seven biotypes. *Gm4*, *gm3* and *Gm8*

have wide range of resistance, *Gm1* and *Gm11* have moderate range of resistance and *Gm2*, *Gm5*, *Gm6*, *Gm7*, *Gm9* and *Gm10* have narrow range of resistance (Bentur et al., 2011). Transcriptome studies were also conducted to understand the molecular basis of rice-gall midge interactions (Rawat, 2012; Rawat et al., 2012a, b). Two suppressive subtraction hybridization (SSH) libraries were constructed to analysis the gene expression in resistant (Suraksha, carrying *Gm11* gene with HR+ type mechanism) and susceptible (TN1, carries no R gene) rice varieties against gall midge biotype 4 (GMB4) (Rawat, 2012; Rawat et al., 2012a). Similarly microarray analysis was done using resistant Kavya carrying *Gm1* gene with HR- mechanism against GMB1 (Rawat et al., 2012b). From these studies, few probable candidates were shortlisted and were validated via real time PCR. In the present study, with the knowledge of physical positions of *Gm* gene linked markers in the rice genome, and available transcriptomics data on rice-gall midge interactions, we analyzed the genomic regions encompassing these genes for gene content using the rice genome sequence and gene annotation data. We aimed to examine the marker delimited gall midge resistance regions; for identification of common genes among these regions, genes common between earlier transcriptome data and these regions and, presence of any pathogen/pest resistance related genes within these regions using *in silico* analysis. This approach could help to narrow down to likely candidate genes and prepare to validate their possible role in gall midge resistance.

Results

Identification of the genes in the marker delimited gall midge resistance regions

The rice genomic regions of size 0.5 to 1.7 Mb encompassing gall midge resistance genes *Gm1*, *Gm2367*, *Gm4*, *Gm8* and *Gm11* were scanned for gene content. These regions together constituted 4.02 Mb segment or 1.07% of the rice genome (373.24 Mb). Table 1 provides summary of the genes found in each genomic region studied. Of the 690 genes discovered in the regions, most were encoding hypothetical proteins, expressed proteins and transposable elements which together accounted for 315 genes *i.e.,* 45.6% of the total number of genes identified. Presence of transposon and retrotransposon genes in large numbers might have functional significance by inserting themselves into the genes involved in plant susceptibility to pathogen/insect and thus disabling the function of the genes leading to disease/pest resistance. Remaining 375 genes were from the annotated list.

Genes common among the genomic regions

Five genes with one or more copies were common among two or more of the target genomic regions (Table 2). One gene coding for nucleotide binding site (NBS) and leucine rich repeat (LRR) class resistance protein was common among *Gm2367*, *Gm8* and *Gm11* regions. Two genes coding for no apical meristem protein and F-box family protein were common among the *Gm2367*, *Gm4* and *Gm11* regions representing genes conferring HR+ type resistance. One gene coding for SET domain containing protein was common between the two HR- type resistance conferring *i.e., Gm1* and *Gm8* regions. One gene coding for pentatricopeptide (PPR) repeat containing protein was common to *Gm4*, *gm3* and *Gm8* region representing genes conferring wide spectrum resistance.

Genes enriched in transcriptomic studies (SSH cDNA library or microarray)

We have identified 15 genes common between the present study and the transcriptomics data (Table 3). Nine genes were found matching locus wise with enriched genes of SSH library (7 genes) or up-regulated genes of microarray library (2 genes) from the incompatible rice-gall midge interactions. Out of seven genes found common with SSH data, four genes encoding LIM domain-containing protein (LOC_Os12g40490.1), calcineurin B (LOC_Os12g40510.1), auxin-responsive Aux/IAA gene family member (LOC_Os12g40890.1) and ankyrin repeat domain-containing protein (LOC_Os12g40780.1) were from *Gm11* region, two genes encoding F-box domain kelch repeat containing protein (LOC_Os04g52830.2) and glucose-1-phosphate uridylyltransferase (LOC_Os04g52370.2) were from *Gm2367* region and one gene encoding GTP-binding protein (LOC_Os08g09940.1) was from *Gm4* region. Whereas, one gene encoding acyl-desaturase chloroplast precursor (LOC_Os08g09950.1) from *Gm4* region and the other encoding for CCT/B-box zinc finger protein (LOC_Os08g15050.1) from *Gm8* region were highlighted in microarray data. Further, we have identified six common genes found to be up-regulated in SSH cDNA library of susceptible TN1 rice variety. These common six genes were UTP--glucose-1-phosphate uridylyltransferase (LOC_Os04g52370.2) , SIT4 phosphatase-associated protein (LOC_Os04g52940.1), soluble starch synthase 3 (LOC_Os04g53310.1) from *Gm2367* region; no apical meristem protein (LOC_Os08g10080.1)) from *Gm4* region and, disease resistance protein (LOC_Os12g39620.3) and protein kinase domain containing protein (LOC_Os12g40279.1) from *Gm11* region. It would be interesting to explore the involvement of these genes in plant susceptibility.

Genes implicated in plant defense mechanism or annotated as resistance genes

Twenty two genes noted across the genomic regions have been either implicated in disease/pest resistance or annotated as resistance genes (Table 4). Six genes encoding WRKY family proteins- four in *Gm4* region and two in *Gm11* region; three genes encoding microtubule-associated protein in *Gm2367* region; two genes encoding proline rich protein, one each in *Gm2367* and *Gm8* regions; two genes, one encoding calmodulin-related calcium sensor protein and the other encoding IQ-calmodulin-binding motif family protein in *Gm11* region were observed. Five genes encoding disease resistance protein in *Gm11* region and one in *Gm8* region; two genes coding for disease resistance protein *RPM1*, one each in *Gm4* and *Gm8* regions; another for disease resistance *RPP13*- like protein in *Gm4* region were also identified.

Primer design for key genes

From putative candidate genes discovered in the study that possibly play role in gall midge resistance, we shortlisted a set of 24 genes for validation (supplementary data). A total of 33 primer pairs were developed to amplify one or more exon regions of these genes (supplementary data). The sequence from the different loci (8 copies) coding for the NBS-LRR genes were aligned using clustral X and the primers were designed from the unique sequences. Similarly, for the other shortlisted genes, primers were developed from the unique regions so that each locus could be amplified. The shortlisted

Table 1. Details of the genes from different gall midge resistance regions.

R genes	Chr location	Chr regions From – to (bp)	Total Size of the region (bp)	Total number of genes	Number of genes present in each category			
					Annotated genes (Functions defined)	Expressed genes (function not defined)	Trans. and retrotrans.	Hypothetical genes
Gm1	9	7656946 - 8156946	500000	62	15	14	31	2
Gm2367	4	31023295 -31848200	824905	162	118	32	8	4
Gm4	8	5370894 - 5870894	500000	98	59	24	10	5
Gm8	8	8769772 - 9269772	500000	83	49	15	14	5
Gm11	12	24258416 - 25959620	1701204	284	134	89	51	12
Total			4026109	690	375	174	113	28

bp-base pair, Chr.-chromosome, Trans.-transposon, retrotrans.-retrotransposon.

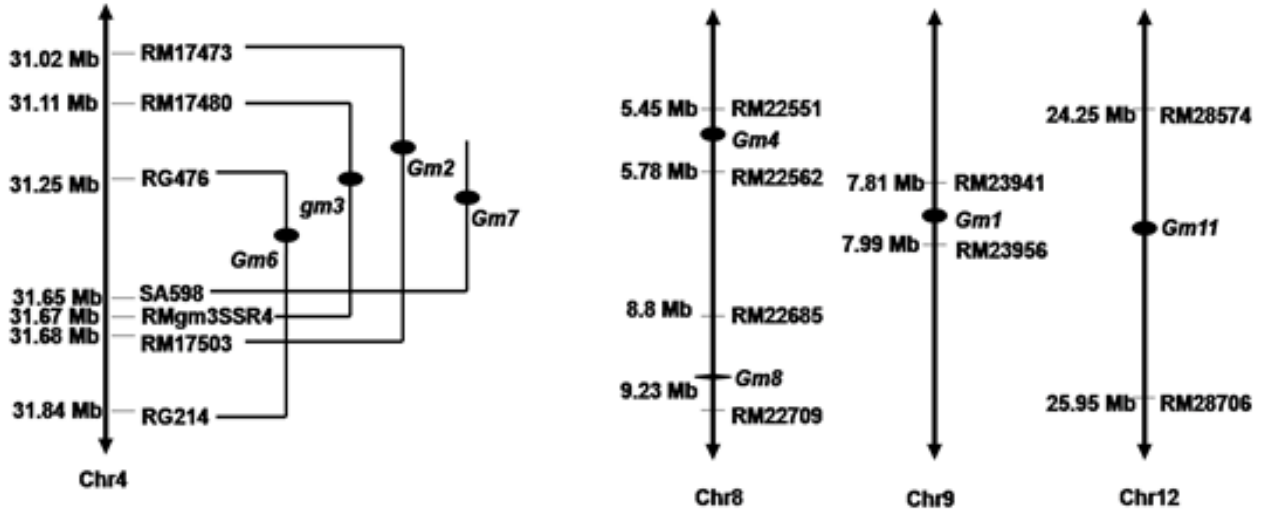

Fig 1. Linkage maps of gall midge resistance genes on different chromosomes showing the reported markers on the right and their physical locations in Mb on the left. Out of 11 gall midge resistance genes, 4 genes (*Gm2, gm3, Gm6* and *Gm7*) are located on chromosome 4, *Gm4* and *Gm8* are located on chromosome 8, *Gm1* is located on chromosome 9 and *Gm11* is located on chromosome 12.

genes were NBS-LRR class genes whose protein are reported to initiate the defense response by recognizing the pathogen/pest derived elicitors, genes encoding no apical meristem protein, SET domain containing protein and pentatricopeptide (PPR) repeat containing protein that probably play role in broad spectrum of resistance. Genes encoding proline rich protein, which are the early defense response genes that strengthen the cell wall against pathogen/pest attack and genes annotated to be coding for disease resistance *RPP13*-like protein, disease resistance protein *RPM1*, resistance protein & disease resistance protein were also shortlisted. Supplementary data gives the sequence information and other details of the designed primers. These primers were checked using pooled cDNA isolated from different rice varieties carrying different gall midge resistance genes.

Discussion

Rice-gall midge interaction is well characterized genetically and several of the plant resistance genes has been tagged and mapped (Sama, 2011). Recently, transcriptomics studies through SSH library construction and microarray analysis have been initiated to understand the molecular basis of the interactions (Rawat, 2012). Nevertheless, none of the gall midge resistance genes have been identified or cloned so far. Here we have taken *in silico* approach to identify the candidate genes which could be useful for further validation in the rice-gall midge interactions. Our hypothesis was that functionally similar genes confer similar phenotype. Similarity was sought in terms of frequency of occurrence in five distinct albeit a narrow fragment of rice genome (1%) all, reported to contain one or many of the reported R genes against gall midge. Superimposed on the physical grid was the functional frame of transcriptomics and available information on plant-pathogen/herbivore interaction. Such an exercise resulted in short listing of 24 genes which require functional and structure validation through expression profiling or through polymorphism studies among functionally distinct alleles. First set of five genes identified were on the basis of common occurrence among the different regions. four distinct , encoding NBS-LRR proteins (also called NB-LRR or NB-ARC-LRR proteins) is from the major class of resistance (R) genes studied so far. These genes encode cytoplasmic receptor-like proteins that help in recognition of elicitor proteins expressed by the pathogen and further trigger the downstream signals leading to disease resistance (DeYoung and Innes, 2006). Numerous resistance genes of this class have been cloned from different plant species such as *Xa1* (Yoshimura et al., 1998), *Pib* (Wang et al., 1999), *Pi-ta* (Bryan et al., 2000), *Pi-k*^{h} or *Pi54* (Sharma et al., 2005) and *Bph14* (Du et al., 2009) from rice; *Cre3* (Lagudah et al., 1997) from wheat*; RPS2* (Mindrinos et al., 1994), *RPM1* (Grant et al., 1995) and *RPS4* (Gassmann et al., 1999) from *Arabidopsis*. A NBS-LRR gene on rice chromosome 8 (Himabindu, 2009) and another NB-ARC on chromosome 4 (Sama, 2011) have been suspected to be candidate genes for *Gm4* and *gm3*, respectively. Resistance genes tend to be clustered in the genome (Jones and Dangl, 2006). Such a gene cluster permits recombination, gene duplication *etc.*, to produce novel resistance specificities (Michelmore and Meyers, 1998). The resistance genes *cf2* and *cf5* to fungus, *Cladosporium falvum* and *Mi* to root-knot nematode, *Meloidogyne spp*. were located in the same region of chromosome 6 on tomato genome (Dickinson et al., 1993). Rossi et al. (1998) showed that the tomato resistance genes, *Meu-1* to the aphid, *Macrosiphum euphorbiae* and *Mi* to nematode which were earlier reported as tightly linked are the same. It was hypothesized from these observations that R genes in a cluster may show different specificities to the same organism, different resistance to different organisms or single R gene may show specificity to different elicitor proteins derived from different organisms. Similar observations were made in the present study. Seven genes encoding NBS-LRR were identified in *Gm2367* region which also has resistance gene *Xa1* (Yoshimura et al., 1998) conferring resistance to bacterial blight pathogen *Xanthomonas oryzae pv. oryzae*. These four *Gm* genes (*Gm2367*) along with *Xa1* gene form a cluster with seven of NBS-LRR genes. Of the seven NBS-LRR genes, two genes with TIGR locus IDs *viz.*, LOC_Os04g52970.1 and LOC_Os04g53120.1 have been suggested to be candidate genes for *gm3* & *Gm2* (Sama, 2011) and *Xa1* (Yoshimura et al., 1998), respectively. It is of interest to validate these genes to determine their function and identity. Several reports are available on base pair mutations observed between alleles of these resistance genes amplified from resistant and susceptible genotypes. When the rice blast resistance gene *Pi54* coding for NBS-LRR domain containing protein was amplified from the both resistant and susceptible genotypes and sequenced, a 144-bp insertion/deletion (InDel) in an exon was observed between the alleles (Ramkumar et al., 2011). Himabindu (2009) amplified the putative candidate gene for *Gm4* encoding LRR containing protein from resistant variety, Abaya (carrying *Gm4*) and susceptible variety, B95-1. Deletions of 4, 16, 4 and 273 bp were observed in Abaya at four different sites of an intron as compared to B95-1. Similarly, a probable candidate gene of *gm3* coding for NB-ARC domain containing protein was amplified from the resistant RP2068-18-3-5 (carrying *gm3*) and Phalguna (carrying *Gm2*) and the susceptible TN1 rice. Deletion of a complete exon of 320 bp was detected in RP2068-18-3-5 and in Phalguna as compared to TN1. Also, an additional 92 bp deletion from another exon was observed in RP2068-18-3-5 as compared to Phalguna (Sama, 2011). Transcriptome data also revealed up-regulation of NBS-LRR gene (GenBank ID: HO088625) in SSH library of Suraksha (carrying *Gm11* genes) against gall midge (Rawat, 2012). These results collectively indicated that NBS-LRR is one of the candidate genes in rice-gall midge interactions. Thus, we have developed primers for amplifying unique regions of all eight NBS-LRR genes including five in *Gm2367* region excluding the two reported to be candidates for *gm3* & *Gm2* and *Xa1*, two in *Gm8* region and one in *Gm11* region. Second common gene among the gall midge resistance regions, coded for no apical meristem (NAM) protein which is required for shoot apical meristem formation and flower development (Souer et al., 1996) and is a member of gene family with a conserved N-terminal domain designated as NAC (NAM, ATAF1, 2, CUC2) (Aida et al., 1997). This gene was also common among the all three HR+ gene regions i.e. *Gm2367*, *Gm4* and *Gm11*. It has been reported that the *Arabidopsis* protein TIP that contains the NAC domain interacted with capsid protein of turnip crinkle virus and induced hypersensitive reaction mediated disease response (Ren et al., 2000). The SSH library for TN1 against GMB4 showed upregulation of NAC domain protein coding gene (GenBank ID: HO188634) during early hours which subsequently declined during late hours of gall midge infestation (Rawat, 2012). It will be worth to explore the role of this gene in rice-gall midge interactions. Third common gene which was abundantly present in all the HR+ gall midge resistance regions encoded for F-box family proteins. These proteins are part of ubiquitin-proteasome pathway which play role in ubiquitination of target proteins for their degradation

Table 2. List of genes that are found common among the different gall midge regions.

S. No.	Gene encoding	Common among the regions (n)
1	NBS-LRR class resistance protein	*Gm2367* (7), *Gm8* (2), *Gm11* (1)
2	No apical meristem protein	*Gm2367* (1), *Gm4* (1), *Gm11* (1)
3	F-box family protein	*Gm2367* (3), *Gm4* (25), *Gm11* (9)
4	SET domain containing protein	*Gm1* (2), *Gm8* (1)
5	Pentatricopeptide (PPR) repeat containing protein	*Gm2367* (2), *Gm4* (1), *Gm8* (1)

n-number of gene copies.

(Kuroda et al., 2002). It has been reported that a single amino acid substitution of glutamic acid-22 to alanine in the F-box domain of *COI1* gene abolished jasmonic acid response which is important for regulating several cellular processes including defense against insects and pathogen in *Arabidopsis* (Xu et al., 2002). The genes encoding F-box family protein were observed in high copy number in *Gm2367* (3), *Gm4* (25) and *Gm11* (9) regions suggesting their involvement in the HR+ mediated gall midge resistance in rice. Gene encoding SET domain containing protein was common between the two HR- type resistance gene regions *i.e., Gm1* and *Gm8*. SET domain containing proteins methylate histone proteins of chromatin to regulate transcriptional activation or silencing of associated DNA (Ng et al., 2007). The epigenetic regulation of pathogenesis related (PR) genes by methylation of histone proteins has been shown to be involved in disease resistance of *Arabidopsis* to *Pseudomonas syringae* (De-La-Pena et al., 2011). As the HR- mechanism of gall midge resistance conferred by *Gm1* and *Gm8* is eluding us and at present looks devoid of presence of known R genes (Rawat et al., 2012b), Primers designed for SET domain containing protein can serve to elucidate the HR- gall midge resistance mechanism more extensively. The fifth gene encoding for pentatricopeptide repeat protein was common among *Gm2367*, *Gm4* and *Gm8* regions. Pentatricopeptide repeat containing proteins are RNA-binding proteins that regulate gene expression in chloroplast and mitochondria by post translational modification of RNA (Saha et al., 2007). Zhou et al. (2010) showed significant up regulation of several rice genes encoding pentatricopeptide repeat in their microarray experiment for maize resistance gene *Rxo1* mediated response in rice to bacterial streak pathogen, *Xanthomonas oryzae pv. oryzicola*. This gene has also been shortlisted and primers were designed for validation in rice-gall midge interactions. We also sieved gene content data through findings of transcriptomic studies (Rawat, 2012) to find out the additional 15 common genes. Out of these 15 genes, 7 genes were in Suraksha SSH cDNA library. Two of these genes *viz.*, LIM domain-containing protein (Zheng and Zhao, 2007) and ankyrin repeat domain containing protein (Sedgwick and Smerdon, 1999) have been reported to mediate protein-protein interactions in several cellular processes. Third gene named Calcineurin B is a calcium (Ca^{2+}) sensor protein, monitors changes in Ca^{2+} concentration elicited by various extracellular signals (Luan et al., 2002). And, the fouth gene, auxin-responsive Aux/IAA gene has been implicated in auxin induced/repressed gene expression (Dharmasiri and Estelle, 2004). Other genes such as F-box family proteins often contain additional domains like kelch repeat domains which provide specificity for degradation of target proteins in ubiquitin-proteasome pathway (Jain et al., 2007). Glucose-1-phosphate uridylyltransferase catalyses the formation of UDP-glucose from glucose-1-phosphate and UTP during glycogen synthesis (Thoden and Holden, 2007). GTP-binding proteins (G proteins) act as molecular switches which initiate and terminate specific cell functions by existing in two interconvertible GDP-bound inactive and GTP-bound active forms (Takai et al., 2001). Two genes were common between microarray data and present study. Gene encoding acyl-desaturase, chloroplast precursor was upregulated during incompatible interaction of Kavya-GMB1. The acyl desaturase catalyses the primary reaction of polyunsaturated fatty acid biosynthesis in most plants by introducing the first double bond into stearoyl- acyl-carrier protein (ACP) to produce oleoyl-ACP (Shanklin and Somerville, 1991).

The other gene encoding CCT/B-box zinc finger protein, also up-regulated in Kavya microarray library,
is a member of subgroup of zinc finger proteins that contain one or more B-box domains and are thought to be involved in protein-protein interactions. Some B-box family members contain additional domains like CCT domain at the C-terminus (Khanna et al., 2009). We also observed six genes up-regulated in compatible interaction and could be candidate susceptibility genes. Apart from these 15 genes, we found 22 genes which are widely implicated in defense responses of plants against different pathogen/pest. Few of these genes are shortlisted and primers are designed for validation. These included genes encoding proline rich proteins which are cell wall proteins rich in proline and hydroxyproline repeats (Showalter, 1993) that confer mechanical barrier through cross linkage with lignins and saccharides and prevent infection of the pathogens (Bradley et al., 1992). Genes for WRKY protein were abundant in *Gm4* and *Gm11* regions. WRKYs are a large class of transcription factor proteins which regulate expression of defense related genes in plants to various biotic and abiotic stresses (Pandey and Somssich, 2009). It has been reported that necrotrophic fungal pathogens, *Botrytis cineria* and *Alternaria brassicicola* caused enhanced susceptibility to *Arabidopsis* having mutation in *WRKY33* gene encoding a WRKY transcription factor (Zheng et al., 2006). Genes for calmodulin (CaM) and IQ calmodulin-binding motif family are also known as calcium sensor proteins that bind to Ca2+ ions and the Ca^{2+}/CaM complex mediates plant responses to various environmental, biotic and abiotic stimuli via binding to target proteins called CaM binding proteins which further activate downstream gene expression. *Arabidopsis* resistance genes coding resistance protein RPP13 conferring resistance to *Peronospora parasitica* (Bittner-Eddy et al., 1999) and RPM1 protein to bacterial pathogen, *Pseudomonas syringae* (Bisgrove et al., 1994) were found in *Gm2367*, *Gm4* and *Gm8* regions respectively and genes annotated to be coding for resistance protein in *Gm8* region and disease resistance

Table 3. List of genes found common between transcriptomics data and different gall midge resistance regions.

S. No.	Gene name	Locus ID	R genes	Transcriptomic studies on rice-gall midge interactions
1.	LIM domain-containing protein	LOC_Os12g40490.1		
2.	Calcineurin B	LOC_Os12g40510.1	*Gm11*	SSH cDNA library of Suraksha-GMB4
3.	Auxin-responsive Aux/IAA gene	LOC_Os12g40890.1		
4.	Ankyrin repeat domain-containing protein	LOC_Os12g40780.1		
5.	F-box domain kelch repeat containing protein	LOC_Os04g52830.2	*Gm2367*	SSH cDNA library of Suraksha-GMB4
6.	Glucose-1-phosphate uridylyltransferase	LOC_Os04g52370.2		
7.	GTP-binding protein	LOC_Os08g09940.1	*Gm4*	SSH cDNA library of Suraksha-GMB4
8.	Acyl-desaturase chloroplast precursor	LOC_Os08g09950.1	*Gm4*	Microarray analysis of Kavya-GMB1
9.	CCT/B-box zinc finger protein	LOC_Os08g15050.1	*Gm8*	Microarray analysis of Kavya-GMB1
10.	UTP--glucose-1-phosphate uridylyltransferase	LOC_Os04g52370.2		
11.	SIT4 phosphatase-associated protein	LOC_Os04g52940.1	*Gm2367*	SSH cDNA library of TN1-GMB4
12.	Soluble starch synthase 3	LOC_Os04g53310.1		
13.	No apical meristem protein	LOC_Os08g10080.1	*Gm4*	SSH cDNA library of TN1-GMB4
14.	Disease resistance protein	LOC_Os12g39620.3		
15.	Protein kinase domain containing protein	LOC_Os12g40279.1	*Gm11*	SSH cDNA library of TN1-GMB4

Table 4. List of genes implicated in disease/pest resistance or annotated as resistance genes within the gall midge resistance regions.

S. No.	Gene encoding	Gene copies	Found in region of
1	Microtubule-associated protein	3	
2	Disease resistance *RPP13*-like protein	1	*Gm2367*
3	Proline rich protein	1	
4	WRKY family protein	4	
5	Disease resistance protein *RPM1*	1	*Gm4*
6	Resistance protein	1	
7	Disease resistance protein *RPM1*	1	*Gm8*
8	Proline rich protein	1	
9	Disease resistance protein	5	
10	WRKY family protein	2	
11	Calmodulin-related calcium sensor protein	1	*Gm11*
12	IQ calmodulin-binding motif family protein	1	

protein in *Gm11* region also deserve validation. Thus in the process for finding similar genes conferring similar phenotype of gall midge resistance, we have narrowed down from 690 total genes or 375 annotated genes in the target segment 4Mb rice genome to 42 likely to 24 more likely candidate genes. We hope that functional and structural validation of the shortlisted genes is the quick way of revealing their identity.

Materials and methods

Rice genome sequence

The Institute for Genomic Research (TIGR) (http://rice.plantbiology.msu.edu/) provides rice genome sequence from Nipponbare, a *japonica* variety. The DNA sequence (release 7) of rice chromosomes 4, 8, 9 and 12 was downloaded from pseudomolecule FTP site of TIGR webpage (http://rice.plantbiology.msu.edu/annotation_pseudo_current.shtml).

Genomic locations of Gm gene linked markers

The primer/related sequences of nearest linked markers for *Gm1*, *Gm2*, *gm3*, *Gm4*, *Gm6*, *Gm7*, *Gm8* and *Gm11* were obtained from Gramene webpage (http://www.gramene.org /markers/). The physical positions of the markers on rice genome were noted using BioEdit tool (http://www.mbio.ncsu.edu/bioedit/bioedit.html) by matching primer/related sequence against the rice chromosome sequence (Fig 1). The *Gm5* linked RAPD markers could not be located for their physical positions.

Rice genomic regions studied

As physical locations of the reported markers are relative, we did not confine just to the genomic region between the markers for the study. If the region between the markers was less than 500 Kb size, we moved both upstream and downstream by equal number of bases so that 500 Kb is covered. Since *Gm7* has linked marker mapped towards its downstream, 500 Kb genomic region upstream to the marker was taken for the study. *Gm2*, *gm3*, *Gm6* and *Gm7* have been mapped close to each other on chromosome 4. We have taken the region encompassing these four genes as single entity and named it as *Gm2367* region. The identified regions for the present study were five genomic regions encompassing *Gm1*, *Gm2367*, *Gm4*, *Gm8* and *Gm11*.

In silico mining of genes in the targeted genomic regions

Based on physical positions of the linked markers, we have extracted genes present in the genomic regions of interest using genome browser option in TIGR webpage.

Primer design

Primer pairs were designed using Primer3 software tool (http://frodo.wi.mit.edu/) for amplification of exon regions of some putative candidate genes. The criteria for designing unique primer sets were as follows: primer length in the range of 18-24 bases with 20 bases as the optimum, product size from 150-300 bp, melting temperature (Tm) ranging from 57-63°C with optimum of 60°C, GC content from 40-60% with 50% as the optimum, maximum Tm difference of 1-3 and CG clamp of 0-2. All other parameters were left at default value. These designed primers were used for amplification from pooled cDNA from different gall midge resistant rice varieties. RNA isolation and cDNA extraction were conducted as mentioned by Rawat (2012).

Conclusion

The present *in silico* analysis helped to identify of the possible candidate genes that could involve in rice-gall midge interactions. And also, our reported primers designed for prioritized genes from among the putative candidate genes will help in confirming the involvement of the genes in gall midge resistance.

Acknowledgements

We thank the Project Director, Directorate of Rice research for constant support. We greatly acknowledge Indian Council of Agricultural Research (ICAR), New Delhi, India for providing research grant (NFBSRA/PCN/AP-1/2006-07) from the National Fund for Basic and Strategic Research in Agriculture.

References

Aida M, Ishida T, Fukaki H, Fujisawa H, Tasaka M (1997) Genes involved in organ separation in *Arabidopsis*: an analysis of the cup-shaped cotyledon mutant. Plant Cell 9:841-857

Bentur JS, Pasalu IC, Sarma NP, Prasada Rao U, Mishra B (2003) Gall midge resistance in rice. DRR research paper series 01/2003. Directorate of Rice Research, Hyderabad, India, p 20

Bentur JS, Padma Kumari AP, Jhansi Lakshmi V, Padmavathi CH, Kondala Rao Y, Amudhan S, Pasalu IC (2011) Insect resistance in rice. DRR technical bulletin 51/2011. Directorate of Rice Research, Hyderabad, India, p 86

Bisgrove SR, Simonich MT, Smith NM, Sattler A, Innes RW **(**1994) A disease resistance gene in *Arabidopsis* with specificity for two different pathogen avirulence genes. Plant Cell 6:927-933

Bittner-Eddy P, Can C, Gunn N, Pinel M, Tor M, Crute I, Holub EB, Beynon J (1999) Genetic and physical mapping of the *RPP13* locus, in *Arabidopsis*, responsible for specific recognition of several *Peronospora parasitica* (downy mildew) isolates. Mol Plant Microbe Interact. 12:792-802

Bradley DJ, Kjellbom P, Lamb CJ (1992) Elicitor- and wound-induced oxidative cross-linking of a proline-rich plant cell wall protein: a novel, rapid defense response. Cell 70:21-30

Bryan GT, Wu KS, Farrall L, Jia Y, Hershey HP, McAdams SA, Faulk KN, Donaldson GK, Tarchini R, Valent B (2000) A single amino acid difference distinguishes resistant and susceptible alleles of the rice blast resistance gene *Pi-ta*. Plant Cell 12:2033-2045

De-La-Pena C, Rangel-Cano A, Alvarez-Venegas R (2011) Regulation of disease-responsive genes mediated by epigenetic factors: interaction of *Arabidopsis–Pseudomonas*. Mol Plant Pathol. doi:10.1111/j.1364-3703.2011.00757.x

DeYoung BJ, Innes RW (2006) Plant NBS-LRR proteins in pathogen sensing and host defense. Nat Immunol. 7:1243-1249

Dharmasiri N, Estelle M (2004) Auxin signaling and regulated protein degradation. Trends Plant Science 9:302-308

Dickinson MJ, Jones DA, Jones JD (1993) Close linkage between the *Cf-2/Cf-5* and *Mi* resistance loci in tomato. Mol Plant Microbe Interact. 6:341-347

Du B, Zhang WL, Liu BF, Hu J, Wei Z, Shi ZY, He RF, Zhu LL, Chen RZ, Han B, He G (2009) Identification and characterization of *Bph14*, a gene conferring resistance to brown planthopper in rice. Proc Natl Acad Sci USA. 106:22163-22168

Gassmann W, Hinsch ME, Staskawicz BJ (1999) The *Arabidopsis RPS4* bacterial resistance gene is a member of the TIR-NBS-LRR family of disease resistance gene. Plant J. 20:265-277

Grant MR, Godiard L, Straube E, Ashfield T, Lewald J, Sattler A, Innes RW, Dangl JL (1995) Structure of the *Arabidopsis RPM1* gene enabling dual specificity disease resistance. Science 269:843-846

Himabindu K (2009) Identification, tagging and mapping of rice gall midge resistance genes using microsatellite markers. PhD thesis, Acharya Nagarjuna University, Guntur, India

Himabindu K, Suneetha K, Sama VSAK**,** Bentur JS (2010) A new rice gall midge resistance gene in the breeding line CR57-MR1523, mapping with flanking markers and development of NILs. Euphytica 174:179-187

Jain M, Nijhawan A, Arora R, Agarwal P, Ray S, Sharma P, Kapoor S, Tyagi AK, Khurana JP (2007) F-box proteins in rice. Genome-wide analysis, classification, temporal and spatial gene expression during panicle and seed development, and regulation by light and abiotic stress. Plant Physiol. 143:1467-1483

Jones JDG, Dangl JL (2006) The plant immune system. Nature 444:323-329

Katiyar SK, Verulkar SB, Adsul G, Dhundre M, Chandel G, Bennett J (2000) Molecular markers for gall midge resistance genes in rice: Stage set for MAS and map based gene cloning. In: Abstracts: IV international rice genetics symposium, 22-27 October 2000. International Rice Research Institute, Los Banos, Philippines, p 81

Katiyar SK, Tan Y, Huang B, Chandel G, Xu Y, Zhang Y, Xie Z, Bennett J (2001) Molecular mapping of gene *Gm-6(t)* which confers resistance against four biotypes of Asian rice gall midge in China. Theor Appl Genet. 103:953-961

Khanna R, Kronmiller B, Maszle DR, Coupland G, Holm M, Mizuno T, Wu SH (2009) The *Arabidopsis* B-Box zinc finger family. Plant Cell 21:3416-3420

Krishnaiah K (2004) Rice gall midge, *Orseolia oryzae*-an overview. In: Bennett J, Bentur JS, Pasalu IC, Krishnaiah K (eds) New approaches to gall midge resistance in rice. Proceedings of the international workshop, 22-24 November 1998, Hyderabad, India. Los Banos (Philippines): International Rice Research Institute and Indian Council of Agricultural Research, p 1-6

Kuroda H, Takahashi N, Shimada H, Seki M, Shinozaki K, Matsui M (2002) Classification and expression analysis of *Arabidopsis* F-box-containing protein genes. Plant Cell Physiol. 43:1073-1085

Lagudah ES, Moullet O, Appels R (1997) Map-based cloning of a gene sequence encoding a nucleotide-binding domain and a leucine-rich region at the *Cre3* nematode resistance in wheat. Genome 40:659-665

Luan S, Kudla J, Rodriguez-Concepcion M, Yalovsky S, Gruissem W (2002) Calmodulins and calcineurin B-like proteins: calcium sensors for specific signal response coupling in plants. Plant Cell 14:S389-400

Michelmore RW, Meyers BC (1998) Clusters of resistance genes in plants evolve by divergent selection and a birth-and-death process. Genome Res. 8:1113-1130

Mindrinos M, Katagiri F, Yu GL, Ausubel FM (1994) The *A. thaliana* disease resistance gene *RPS2* encodes a protein containing a nucleotide binding site and leucine-rich repeats. Cell 78:1089-1099

Ng DW, Wang T, Chandrasekharan MB, Aramayo R, Kertbundit S, Hall TC (2007) Plant SET domain-containing proteins: structure, function and regulation. Biochim Biophys Acta. 1769:316-329

Pandey SP, Somssich IE (2009) The role of WRKY transcription factors in plant immunity. Plant Physiol**.** 150:1648-1655

Ramkumar G, Srinivasarao K, Madhan Mohan K, Sudarshan I, Sivaranjani AKP, Gopalakrishna K, Neeraja CN, Balachandran SM, Sundaram RM, Prasad MS, Shobha Rani N, Rama Prasad AM, Viraktamath BC, Madhav MS (2011) Development and validation of functional marker targeting an InDel in the major rice blast disease resistance gene *Pi54* (*Pik^h*). Mol Breed. 27:129-135

Rawat N (2012) Identification and characterization of differentially expressed genes in rice involved in rice-gall midge interactions. PhD thesis, Osmania University, Hyderabad, India

Rawat N, Neeraja CN, Nair S, Bentur JS (2012a) Differential gene expression in gall midge susceptible rice genotypes revealed by suppressive subtraction hybridization (SSH) cDNA libraries and microarray analysis. Rice doi:10.1186/1939-8433-5:8

Rawat N, Neeraja CN, Sundaram RM, Nair S, Bentur JS (2012b) A novel mechanism of gall midge resistance in the rice variety Kavya revealed by microarray analysis. Funct Integr Genomics 12:249-64. Ren T, Qu F, Morris TJ (2000) HRT gene function requires interaction between a NAC protein and viral capsid protein to confer resistance to turnip crinkle virus. Plant Cell 12:1917-1926

Rossi M, Goggin FL, Milligan SB, Kaloshian I, Ullman DE, Williamson VM (1998) The nematode resistance gene *Mi* of tomato confers resistance against the potato aphid. Proc Natl Acad Sci USA 95:9750-9754

Saha D, Prasad AM, Srinivasan R (2007) Pentatricopeptide repeat proteins and their emerging roles in plants. Plant Physiol Biochem. 45:521-534

Sama VSAK, Himabindu K, Sundaram RM, Viraktamath BC, Bentur JS (2010) Mapping of a rice gall midge resistance gene, *gm3*, in RP 2068-18-3-5 and *in silico* identification of candidate gene(s). Rice Genetics Newsletter 25:76

Sama VSAK (2011) Identification, tagging and mapping of new resistance gene(s) against the Asian rice gall midge *Orseolia oryzae* in rice varieties. PhD thesis, Osmania University, Hyderabad, India

Sardesai N, Kumar A, Rajyashri KR, Nair S, Mohan M (2002) Identification and mapping of an AFLP marker liked to *Gm7*, a gall midge resistance gene and its conversion to a SCAR marker for its utility in marker aided selection in rice. Theor Appl Genet. 105:691-698

Sedgwick SG, Smerdon SJ (1999) The ankyrin repeat: a diversity of interactions on a common structural framework. Trends Biochem Sci. 24:311-316

Shanklin J, Somerville C (1991) Stearoyl-acyl-carrier-protein desaturase from higher plants is structurally unrelated to the animal and fungal homologs. Proc Natl Acad Sci USA 88:2510-2514

Sharma TR, Madhav MS, Singh BK, Shanker P, Jana TK, Dalal V, Pandit A, Singh A, Gaikwad K, Upreti HC, Singh

NK (2005) High-resolution mapping, cloning and molecular characterization of the *Pi-k*^{h} gene of rice, which confers resistance to *Magnaporthe grisea*. Mol Gen Genomics 274:569-578

Showalter AM (1993) Structure and function of plant cell wall proteins. Plant Cell. 5:9-23

Souer E, van Houwelingen A, Kloos D, Mol J, Koes R (1996) The no apical meristem gene of *Petunia* is required for pattern formation in embryos and flowers and is expressed at meristem and primordia boundaries. Cell 85:159-170

Sundaram RM (2007) Fine mapping of rice gall midge resistance genes *Gm1* and *Gm2* and validation of the linked markers. PhD thesis, University of Hyderabad, Hyderabad, India

Takai Y, Sasaki T, Matozaki T (2001) Small GTP-binding proteins. Physiol Rev. 81:153-208

Thoden JB, Holden HM (2007) The molecular architecture of glucose-1-phosphate uridylyltransferase. Protein Sci. 16:432-440

Vijaya Lakshmi P, Amudhan S, Bindu KH, Cheralu C, Bentur JS (2006) A new biotype of the Asian rice gall midge *Orseolia oryzae* (Diptera: Cecidomyiidae) characterized from the Warangal population in Andhra Pradesh, India. Int J of Trop Insect Sci. 26:207-211

Wang ZX, Yano M, Yamanouchi U, Iwamoto M, Monna L (1999) The *Pib* gene for rice blast resistance belongs to the nucleotide binding and leucine rich repeat class of plant disease resistance genes. Plant J. 19:55-64

Xu L, Liu F, Lechner E, Genschik P, Crosby WL, Ma H, Peng W, Huang D, Xie D (2002) The SCF (COI1) ubiquitin-ligase complexes are required for jasmonate response in *Arabidopsis*. Plant Cell 14:1919-1935

Yoshimura S, Yamanouchi U, Kayayose Y, Toki S, Wang ZX, Kono I, Kurata N, Yano M, Iwata N, Sakaki T (1998) Expression of *Xa1*, a bacterial blight-resistance gene in rice, is induced by bacterial inoculation. Proc Natl Acad Sci USA. 95:1663-1668

Zheng Q, Zhao Y (2007) The diverse biofunctions of LIM domain proteins: determined by subcellular localization and protein-protein interaction. Biol Cell. 99:489-502

Zheng Z, Qamar SA, Chen Z, Mengiste T (2006) *Arabidopsis* WRKY33 transcription factor is

required for resistance to necrotrophic fungal pathogens. Plant J. 48:592-605

Zhou YL, Xu MR, Zhao MF, Xie XW, Zhu LH, Fu BY, Li ZK (2010) Genome-wide gene responses in a transgenic rice line carrying the maize resistance gene *Rxo1* to the rice bacterial streak pathogen, *Xanthomonas oryzae* pv. *Oryzicola*. BMC Genomics 11:78