Magnaporthe grisea Infection Triggers RNA Variation and Antisense Transcript Expression in Rice[1][W]

Malali Gowda, R.-C. Venu, Huameng Li, Chatchawan Jantasuriyarat, Songbiao Chen, Maria Bellizzi, Vishal Pampanwar, HyeRan Kim, Ralph A. Dean, Eric Stahlberg, Rod Wing, Cari Soderlund, and Guo-Liang Wang*

Department of Plant Pathology (M.G., R.-C.V., H.L., C.J., G.-L.W.) and Ohio Supercomputer Center (E.S.), Ohio State University, Columbus, Ohio 43212; Arizona Genomics Computational Laboratory, BIO5 Institute (V.P., C.S.) and Arizona Genomics Institute and Department of Plant Sciences (H.K., R.W.), University of Arizona, Tucson, Arizona 85721; Fungal Genomics Laboratory, Department of Plant Pathology, North Carolina State University, Raleigh, North Carolina 27695 (R.A.D.); and Rice Genomics Laboratory, Hunan Agricultural University, Changsha, Hunan 410128, China (G.-L.W.)

Rice blast disease, caused by the fungal pathogen *Magnaporthe grisea*, is an excellent model system to study plant-fungal interactions and host defense responses. In this study, comprehensive analysis of the rice (*Oryza sativa*) transcriptome after *M. grisea* infection was conducted using robust-long serial analysis of gene expression. A total of 83,382 distinct 21-bp robust-long serial analysis of gene expression tags were identified from 627,262 individual tags isolated from the resistant (R), susceptible (S), and control (C) libraries. Sequence analysis revealed that the tags in the R and S libraries had a significant reduced matching rate to the rice genomic and expressed sequences in comparison to the C library. The high level of one-nucleotide mismatches of the R and S library tags was due to nucleotide conversions. The A-to-G and U-to-C nucleotide conversions were the most predominant types, which were induced in the *M. grisea*-infected plants. Reverse transcription-polymerase chain reaction analysis showed that expression of the adenine deaminase and cytidine deaminase genes was highly induced after inoculation. In addition, many antisense transcripts were induced in infected plants and expression of four antisense transcripts was confirmed by strand-specific reverse transcription-polymerase chain reaction. These results demonstrate that there is a series of dynamic and complex transcript modifications and changes in the rice transcriptome at the *M. grisea* early infection stages.

Rice blast disease, caused by *Magnaporthe grisea*, is one of the most devastating rice diseases in the world. The rice and *M. grisea* interaction is becoming a model system for studying plant-fungal interactions due to public availability of the host (Goff et al., 2002; Yu et al., 2002; International Rice Genome Sequencing Project, 2005) and pathogen (Dean et al., 2005) genome sequences and the functional genomic resources (http://www.mgosdb.org; Gowda et al., 2006c). The easy genetic manipulation of both organisms has streamlined the functional analysis of the rice and fungal genes that are involved in the defense or pathogenesis. In the last years, several techniques, such as differential display, suppression subtractive hybridization (SSH), and EST were applied to reveal the events at the transcriptome level during rice (*Oryza sativa*) and *M. grisea* interaction. Using the differential display method, Kim et al. (2000) isolated 18 defense-related genes from suspension cells treated with a fungal elicitor prepared from *M. grisea*. SSH is a rapid and effective method to isolate differentially expressed genes as demonstrated in the studies by Kim et al. (2001, 2005), Xiong et al. (2001), Lu et al. (2004), and Han et al. (2004). However, the high level of sequence redundancy in SSH libraries limits its ability to isolate a large number of defense genes. The EST approach was used by Kim et al. (2001) and Jantasuriyarat et al. (2005) to identify defense transcripts on a large scale. The latter sequenced a total of 68,920 ESTs from eight cDNA libraries and identified 13,570 distinct sequences. Because of the high sequencing cost and relative high sequence redundancy, the EST approach was unable to provide deep coverage of the defense transcriptome in the infected rice plants.

Serial analysis of gene expression (SAGE) is one of the most popular genome-wide transcriptome-profiling methods developed in the last decade (Velculescu et al., 1995). It is a high-throughput and open system to analyze the transcripts without any prior sequence information. SAGE analysis not only provides transcript abundance, but also identifies novel sense and antisense transcripts (Gunasekera et al., 2004; Ge et al., 2006; Gowda et al., 2006c), RNA splice forms (Gowda et al., 2006c), and micro RNAs (miRNAs; Ge et al., 2006) in the tested samples. SAGE is based on two basic principles: isolation of a short sequence tag (14–26 bp) from the 3′ region of a transcript and concatenation of

[1] This work was supported by the National Science Foundation Plant Genome Research Program (grant nos. DBI 0115642 and 0321437).

* Corresponding author; e-mail wang.620@osu.edu; fax 614–292–4455.

The author responsible for distribution of materials integral to the findings presented in this article in accordance with the policy described in the Instructions for Authors (www.plantphysiol.org) is: Guo-Liang Wang (wang.620@osu.edu).

[W] The online version of this article contains Web-only data.
www.plantphysiol.org/cgi/doi/10.1104/pp.107.095653

multiple tags in serial fashion for sequencing (Velculescu et al., 1995). A major drawback of conventional SAGE is that a large portion of 14-bp tags will match to multiple locations in the complex genomic or expressed sequences (Chen et al., 2002). LongSAGE has improved the tag size to 21 bp, which could uniquely match to the complex genome or expressed sequences (Saha et al., 2002). Although many laboratories have tried to adopt SAGE and LongSAGE methodologies in the past, most failed to make quality libraries for sequencing due to the inherent technical difficulties in concatemer cloning. The improved LongSAGE method, called robust-long (RL) SAGE, significantly increased the cloning efficiency of concatemers and insert size (Gowda et al., 2004; Gowda et al., 2006b, 2006c; Gowda and Wang, 2007). Recently, we developed a new method, called robust analysis of 5′-transcript ends in which cloning steps are eliminated and a conventional sequencing method is replaced by 454 pyrosequencing methods (Gowda et al., 2006a). The average robust analysis of transcript end tags are about 80 bp.

In this study, we constructed three rice RL-SAGE libraries, resistant (R), susceptible (S), and control (C), with RNA isolated from the leaf tissues 24 h after *M. grisea* inoculation. A large set of distinct tags (83,832) were isolated from the three libraries. Interestingly, the mismatch rate of transcript tags in the infected libraries to the rice genomic and expressed sequences was significantly increased due to nucleotide conversions. An enhanced level of expression of the cytidine deaminase and adenine deaminase genes was observed in the infected rice leaves. In addition, we also identified hundreds of antisense transcripts for rice genes that were also highly expressed in the R and S libraries as compared to the C. Our results provide evidence for the involvement of RNA variation and antisense transcript expression during plant-fungal interactions.

RESULTS

Reduction of Defense Transcriptome Size and Presence of Highly Specific Transcript Tags in the R, S, and C Libraries

RL-SAGE library analysis revealed a significant reduction in the number of distinct tags in the R and S libraries in comparison to the C library. The R library had 28,081 distinct tags (11.3% of 248,278 total tags), and the S library had 31,025 distinct tags (10.9% of 282,420 total tags) as compared to 36,034 distinct tags (36.4% of 99,031 total tags) in the C library (Table I; Fig. 1A). The distinct tags in each library were further classified into three groups: highly expressed (≥10 copies), intermediately expressed (two to nine copies), and lowly expressed (single-copy [singleton]) tags. Tag frequency analyses indicated that the R library had many more highly expressed tags (C, 367 tags; R, 1,076 tags; S, 648 tags), and the S library had a higher number of intermediately expressed tags (C, 3,627; R, 2,937; S, 4,990), whereas the number of singleton tags was almost similar among the three libraries (C, 24,401; R, 22,919; S, 23,664). The reduction of the transcriptome size in the R and S libraries is likely due to the increased tag redundancy of highly expressed transcripts. Overall, a total of 83,382 distinct tags were obtained by clustering all the distinct tags from the three libraries. The distribution of distinct tags on the Knowledge-Based Oryza Molecular Biological Encyclopedia (KOME) full-length (FL)-cDNAs was analyzed and shown in Figure 2. As expected, most of the tags (79.4%) were mapped to the 3′ region of the KOME FL-cDNAs.

A tag with two or more copies was considered a technically significant tag, and a tag with only a singleton was a technically nonsignificant tag. For technically significant tags, an 18.4% reduction in the R library and a 23.7% reduction in the S library were observed as compared to that of the C library (Fig. 1A). About two-thirds of the distinct tags in the three libraries were singleton tags (Fig. 1A). The percentage of singleton tags is similar to that reported in previous SAGE studies (Velculescu et al., 1999; Chen et al., 2002).

We also defined a tag as a biologically significant tag if it commonly occurred in two or more libraries, and a nonbiologically significant tag if it occurred only in one library. The percent of biologically significant tags present in all three libraries (C, R, and S) was only 3.0% as compared to 7.2% in the R and S libraries, 7.5% in the C and R libraries, and 9.3% in the C and S libraries (Table II). Even when technically significant tags among the three libraries were used for comparison

Table I. *Characteristics of the three RL-SAGE libraries*

Rice Cultivar	Nipponbare		
Treatment/blast strains	Water spray	C9240-1	Che8061
Reaction	C	R	S
Library symbol	C	R	S
Hours after inoculation	24 h	24 h	24 h
No. of sequence reads	9,120 (from both ends of 4,560 clones)	7,200 (from one end of 7,200 clones)	7,618 (from one end of 7,618 clones)
Average insert size (kb)	1.01	1.1	1.3
Total tags/library	99,031	248,278	282,420
Distinct tags/library	36,034	28,081	31,025
Total tags (all three libraries)	629,729		
Total distinct tags (all three libraries)	83,382		

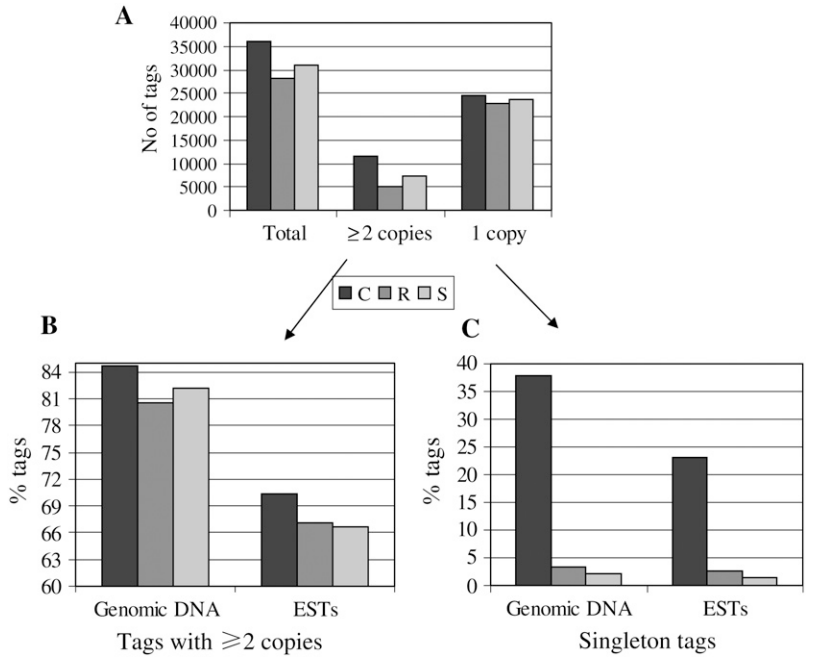

Figure 1. The distinct RL-SAGE tags of the C, R, and S libraries and their hits in the rice genomic and TIGR EST sequences. A, Number of total distinct, significant (greater than two copies), and singleton tags in the three libraries. B, Matching results of the significant tags to both genomic and EST sequences. C, Matching results of the singletons to both genomic and EST sequences.

analysis, only 19.6% to 24.6% commonly occurred in any of the two libraries (data not shown). Intriguingly, a large portion of the tags (76.7%–78.5%) was specifically expressed either in the R, S, or C library (Table II). These results strongly suggest that the R and S transcriptomes might have undergone dynamic and different transcript reprogramming in rice plants at the early infection stages.

To identify whether any fungal genes were expressed in the *M. grisea*-infected leaf tissues, the RL-SAGE tags from the R and S libraries were matched to the *M. grisea* genome and annotated gene sequences. Only 24 tags matched the *M. grisea* genome, but not the rice sequences (Supplemental Table S1). Among the seven tags from the R library, three matched the annotated genes (MGG_12892.5, MGG_08575.5, and MGG_03245.5). Among the 17 tags from the S library, three matched the annotated genes (MGG_06927.5, MMGa04587, and MMGa15070).

Significant Reduction of the Matching Rate of the R and S Library Tags to the Rice Genomic and Expressed Sequences

When the distinct tags of each library matched the rice reference sequence (RefSeq), the tags in the C library had the highest matching rate (53.0% to genomic DNA and 38.3% to ESTs). On the contrary, only a small fraction of the tags from the R library (17.6% to genomic DNA and 14.5% to ESTs) and the S library (21.0% to genomic DNA and 17.0% to ESTs) matched the rice RefSeq. The matching rate of the technically significant tags (greater than two copies) in the C, R, and S libraries was 70.4%, 67.0%, and 66.6%, respectively, when matched with the ESTs, and was 88.1%, 84.1%, and 85.6%, respectively, when matched with the rice genomic sequence (Fig. 1B). However, the matching rate of singletons in the R and S libraries was much lower than that in the C library (Fig. 1C). Specifically, the matching rate of singletons in the C library was 23.1% to the ESTs and 37.8% to the genomic sequence. Surprisingly, only a very small fraction of the singletons in the R and S libraries matched the ESTs (2.6% in R and 1.5% in S) and the genomic sequence (3.4% in R and 2.0% in S), compared to the C library (Fig. 1C).

Similarly, biologically significant common tags and library specific tags were analyzed for sequence matching to the rice RefSeq (Table II). The matching rate of the common tags between the R and S libraries was significantly low (53.7% and 55.7% with the ESTs and genomic sequences, respectively) as compared to

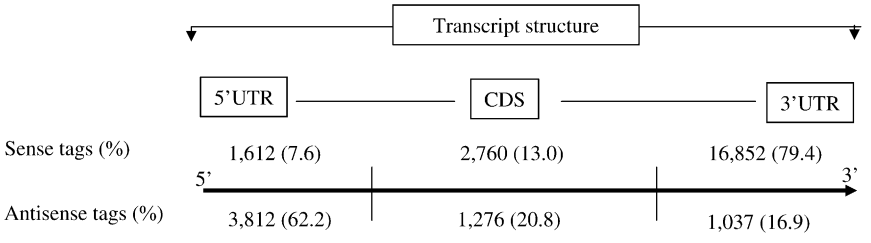

Figure 2. Location of sense and antisense RL-SAGE tags on the rice FL-cDNAs. Each FL-cDNA was equally divided into three portions (3′, mid, and 5′). The distinct sense and antisense tags from the three RL-SAGE libraries were mapped to the FL-cDNAs using the SAGEspy program (Gowda et al., 2006b). The number of hits and the percent of tags are shown on the top of each category.

Table II. *Tag cluster analysis of the C, R, and S libraries and their homology to the rice EST and genomic sequences in the TIGR databases*

Rice Database	Tag Class	Total Tag	No. of Nonredundant Hits
		%	%
TIGR ESTs	C specific	28,278 (78.5)	10,894 (38.5)
	R specific	22,041 (78.5)	1,003 (4.5)
	S specific	23,781 (76.6)	1,427 (6.0)
	CR common	4,514 (7.0)	3,352 (74.3)
	CS common	5,718 (8.5)	4,299 (75.2)
	RS common	4,002 (6.8)	2,153 (53.8)
	CRS common	2,476 (3.0)	1,967 (79.4)
Genomic DNA[a]	C specific	28,278 (78.5)	13,065 (46.2)
	R specific	22,041 (78.5)	1,227 (5.6)
	S specific	23,781 (76.6)	1,786 (7.5)
	CR common	4,514 (7.0)	3,498 (77.5)
	CS common	5,718 (8.5)	4,533 (79.3)
	RS common	4,002 (6.8)	2,230 (55.7)
	CRS common	2,476 (3.0)	2,021 (81.6)

[a]Genomic DNA from japonica 'Nipponbare.'

that of the common tags of the C and R libraries (74.2% and 77.5%), the common tags of the C and S libraries (75.1% and 79.3%), and the common tags of the C, R, and S libraries (79.4% and 81.6%). A considerable number of specific tags in the C library matched to ESTs (38.5%) and genomic DNA sequences (46.2%). On the contrary, only a very small fraction of the R- and S-specific tags matched the ESTs (4.5% in R and 6.0% in S) and the genomic sequences (5.6% in R and 7.5% in S). These results strongly suggest that the tag sequences in the R and S libraries might have been modified or changed considerably during rice and *M. grisea* interaction.

M. grisea Infection Triggers RNA Variation in Rice

To investigate the cause of the mismatch of the transcript tags in the infected plants, one- and two-nucleotide mismatches to the rice RefSeq were allowed in the matching analysis. The tags with a one-nucleotide mismatch were 3,619 (10.0%) in the C library, 6,228 (22.2%) in the R library, and 6,726 (21.7%) in the S library. The number of significant tags with a one-nucleotide mismatch in the three libraries was small (485 in C [1.3%], 358 in R [1.3%], and 372 in S [1.2%]). Many of the one-nucleotide mismatch singleton tags hit to the same genes in the R and S libraries, suggesting that multiple editing events occurred within the same 21-bp region of a gene. Interestingly, the one-nucleotide mismatch rate of the singletons in the R and S libraries to rice annotated genes was 3 times higher than that in the C library (Fig. 3A; C, 8.7%; R, 20.8%; S, 20.5%). However, the two-nucleotide mismatch analysis did not reveal any significant difference among the tags from the three libraries (Fig. 3A). This indicates that single-nucleotide conversions are the most predominant type of RNA variations in *M. grisea*-infected plants, although some sequence deletions or additions were also observed (data not shown).

Detailed analysis of one-nucleotide mismatched tags revealed 12 types of nucleotide conversions (Fig. 3B). The A-to-G and U-to-C types were most predominant in the R and S libraries. The gene with the highest numbers of single-nucleotide variations was the Rubisco activase gene (Fig. 4), of which 70% of the conversions were found from high-quality sequences (Supplemental Fig. S1). To validate single-nucleotide conversions, reverse transcription (RT)-PCR analysis was performed using independently isolated RNA from C, R, and S reactions. The RT-PCR products were cloned and about 30 RT-PCR clones were sequenced from each library. In the sequenced clones from the R and S libraries, the A-to-G and U-to-C conversions in the Rubisco activase transcripts were confirmed (Supplemental Fig. S2). It is noteworthy that several edited sites were also identified for many pathogenesis-related (PR) protein-encoding transcripts such as chitinase, peroxidase, nonspecific lipid-transfer protein 2, germin protein type 1, or catalase (Supplemental Table S2).

In the rice genome, we identified eight cytidine deaminase genes and four adenine deaminase genes

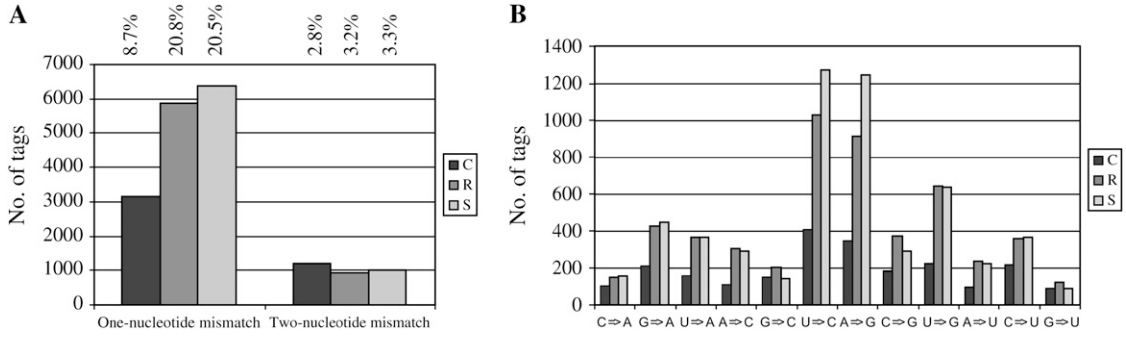

Figure 3. Mismatch analysis of RL-SAGE tags against annotated rice genes. Total distinct tags (A) from the C, R, and S libraries were matched against TIGR annotated rice genes by allowing one and two mismatches. The percent on the top of each bar represents the proportion of the matched tags in each library. Among the one-mismatch tags, the number of tags with specific nucleotide substitutions in C, R, and S libraries is shown in B.

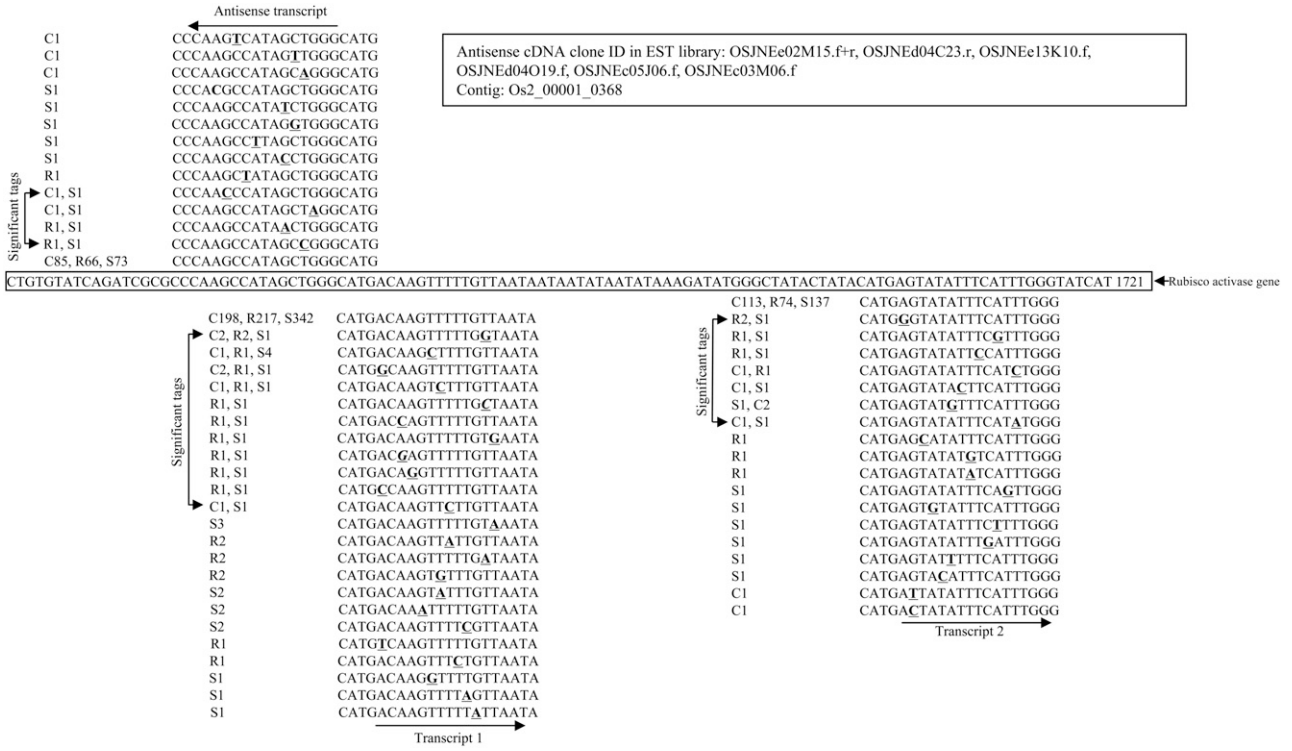

Figure 4. Nucleotide conversions in Rubisco activase transcripts. Nucleotide conversions are shown with bold and underlined letters. The direction (5′–3′) of the transcripts is shown using an arrow. The copy number of the tags in each library is shown at the left side (C198, R217, S342 denotes 198 tags in the C library, 217 tags in the R library, and 342 tags in the S library). The bold and italicized letters are nucleotide conversions confirmed by sequencing of the RT-PCR fragments (the sixth and ninth tags of Transcript 1). The box at the right-top corner is showing the antisense ESTs for Rubisco activase gene.

(Supplemental Table S3). However, only LOC_Os07g14150 (cytidine deaminase) and LOC_Os05g28180 (adenine deaminase) were expressed in all three libraries, but with a much higher expression level in the R and S libraries (Fig. 5). An independent RT-PCR confirmation experiment showed that the adenine deaminase and cytidine deaminase genes were indeed up-regulated in the infected plants (Fig. 5). These results suggest that transcript sequence variation is associated with RNA-editing pathways in rice upon *M. grisea* infection.

Induction of Antisense Transcripts in the R and S Transcriptomes

Antisense transcripts in eukaryotes have recently been recognized as important RNA molecules that play vital roles in various cellular pathways (Kumar and Carmichael, 1998). We identified 5,222, 3,266, and 3,402 antisense RL-SAGE tags, which perfectly matched the rice FL-cDNAs, ESTs, and annotated genes, respectively. Clustering analysis identified 6,050 distinct antisense tags. Among them, 285 matched the KOME antisense FL-cDNAs (Osato et al., 2003; Supplemental Table S4). In comparison to the identification of 1,374 antisense transcripts from 21 FL-cDNA libraries in rice (Osato et al., 2003), RL-SAGE is a more efficient method to isolate antisense transcripts. If total distinct tags are considered, our RL-SAGE approach recovered 7.3% (6,050/83,382) antisense transcripts as compared to 4.3% (1,374/32,127) antisense transcripts in the FL-cDNA cloning approach. In addition, most of the antisense transcripts identified in our libraries were low-abundant expressed, in accordance with

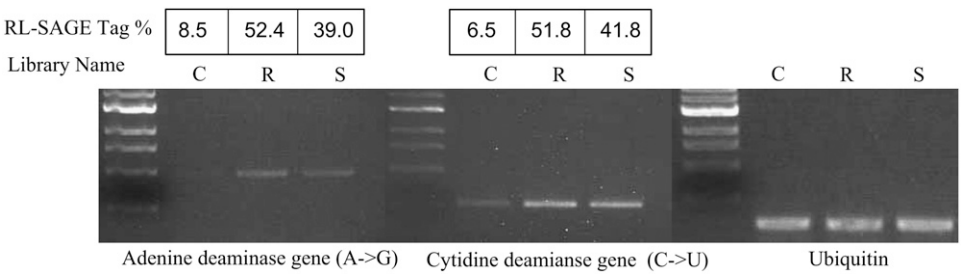

Figure 5. RT-PCR analysis of the adenine deaminase and cytidine deaminase genes. Tag frequency in each library is shown in the box above the gel picture. The ubiquitin gene was used as the expression C for RT-PCR amplification.

other LongSAGE studies (Gunasekera et al., 2004; Quere et al., 2004; Ge et al., 2006; Gowda et al., 2006c).

Interestingly, most of the antisense tags (62.2%) from the three libraries perfectly matched to the 5′ region of the KOME sense FL-cDNA sequences (Fig. 2), which contradicts a previous report stating that most of the antisense transcripts from the FL-cDNA libraries overlap with the 3′ region of the sense transcripts in rice (Osato et al., 2003). About 17% of the antisense tags were located in the 3′ region of the sense FL-cDNAs (Fig. 2). Many FL-cDNAs had both sense and antisense transcript tags at their 3′-untranslated region that could form double-strand RNA (dsRNA) molecules. For example, over 10 copies of the sense and antisense transcript tags of the Rubisco activase gene were located on the opposite strands of the 3′-untranslated region (Fig. 4).

We also found that the percentage of antisense tags with two or more copies was much higher in the R (81.2%) and S (89.5%) libraries as compared to the C library (41.2%; Supplemental Fig. S3). Similar to the sense transcript tags, there were much less perfectly matched antisense transcript tags to the rice FL-cDNAs from the R and S libraries as compared to the C library (R, 430 [1.5%]; S, 579 [1.9%]; C, 2,627 [7.3%]). Sequence analysis also revealed that the matching rate of the antisense transcripts in the R and S libraries to the rice RefSeq was increased dramatically when one-nucleotide mismatch was allowed (data not shown). Interestingly, induction or repression of the antisense transcripts for many PR genes was observed in the R and S libraries (Table III). To confirm the expression of some antisense transcripts, strand-specific RT-PCR using specific antisense primers of four selected genes was performed. The expression of four antisense transcripts detected by RT-PCR was consistent with the tag frequency in the three libraries (Supplemental Fig. S4). All four antisense tags had matches in the rice antisense database (Osato et al., 2003). These results suggest, similar with the sense tags, that expression of many antisense tags in the R and S reactions may have been changed during *M. grisea* infection.

DISCUSSION

In the battle to survive in nature, plants have evolved and developed sophisticated mechanisms to recognize pathogens and defend themselves against infection. Rapid activation of defense-related genes is one of the predominant responses to effectively inhibit pathogen invasion. In this study, the defense transcriptome of rice plants was deeply surveyed using the RL-SAGE approach. Dramatic transcriptome reprogramming in the infected rice plants was found 24 h after *M. grisea* inoculation. The majority of the distinct transcript tags (approximately 75%) were specifically expressed in the R and S reactions. Less than 3,000 transcript tags were commonly present in all three libraries. In comparison with the results from our EST sequencing project (Jantasuriyarat et al., 2005), many more differentially expressed genes were identified in the R and S reactions due to deep profiling of transcripts in the RL-SAGE libraries. From 8,000 to 9,000 cDNA clones, only 2,500 to 3,000 distinct ESTs were identified in each library. On the contrary, about 30,000 distinct transcripts were identified after sequencing about 7,000 RL-SAGE clones. Therefore, RL-SAGE is a much more effective method for in-depth transcriptome analysis (Gowda et al., 2004, 2006c). The differentially expressed genes identified from this study offer a rich genomic resource for further detailed functional analysis of rice defense-related genes against *M. grisea* infection.

In the matching analysis of the total RL-SAGE tags, we found that the matching rate of the R and S library tags to the rice RefSeq was significantly reduced in comparison to the C library. The matching rate of the significant tags was comparable in the C, R, and S libraries because only a small number of edited tags

Table III. *Antisense expression of PR genes*

Tag Sequence (GATC + 17 bp)	C	R	S	Locus ID	Gene Name
TTTGTTGAACACGAACA	5	42	0	LOC_Os05g51660	PR protein 1
GCCTAGAAATTCTCACA	2	9	0	LOC_Os06g22930	Xyloglucan endo-1,4-β-D-glucanase
GCGACGTTGTACCCGTC	0	0	5	LOC_Os12g43440	Thaumatin-like protein precursor
GAGATGGGAGTGGTGGT	2	30	0	LOC_Os03g60840	Putative Bowman-BirkSer protease inhibitor
AATGATACTAAGTTCCT	0	2	0	LOC_Os01g56010	Serpin (serine protease inhibitor)
CTTAAAGAACATTTGAT	0	0	4	LOC_Os03g25320	Putative peroxidase
GAGAAGAAGAGGAAGAA	0	0	6	LOC_Os06g48030	Peroxidase, putative
TTACAACTATATACGTG	5	0	4	LOC_Os06g08670	Phospholipid hydroperoxide glutathione peroxidase
CAATCAAATGCACTACA	11	9	22	LOC_Os10g35800	Putative peroxidase
CTGAGTGGAGAAGTAGT	0	0	6	LOC_Os02g44500	GPX12Hv, glutathione peroxidase-like protein
AAAACTATACAGGTACA	3	0	5	LOC_Os04g46960	Glutathione peroxidase 1
GAAACCAAACACCACGC	0	0	6	LOC_Os04g59150	Peroxidase, putative
TAAATTTCATTTCTTTC	3	0	0	LOC_Os08g33710	Ribonuclease T2 family
TACGGCCTGCGATGTCA	0	0	6	LOC_Os11g02370	Nonspecific lipid-transfer protein 2
GCGATGGATCGACCTAG	0	12	0	LOC_Os12g02310	Nonspecific lipid-transfer protein 2
AACGACCATTAGCTTTT	2	0	2	LOC_Os08g13440	Germin protein type 1—rice
CACTCAAGTCAAATGAC	1	0	11	LOC_Os07g46990	Copper/zinc superoxide dismutase

had two copies or more in the R and S libraries. On the contrary, the matching rate of the singletons in these libraries was extremely low (3%–4%). To identify the cause of the low matching rate, we allowed one- to two-nucleotide mismatches in the sequence analysis, which significantly improved the matching rate to both genomic and expressed sequences. Furthermore, various types of single-nucleotide conversions were identified in the mismatched tags. Sequence analysis of the raw reads showed that over 70% of the nucleotide conversions were from high-quality sequencing regions. Independent RT-PCR confirmation of these variations clearly demonstrated that the mismatches were not originated from sequencing errors. This is consistent with the finding by Chen et al. (2002) and Poroyko et al. (2005) that over 70% of singleton tags are real transcripts.

These results prompted us to speculate that mRNA might be edited in the infected plants. RNA editing has been well documented for several mitochondrial and chloroplast genomes of plants. However, no study has been reported on the editing of nuclear transcripts in plants. In contrast, transcripts of many nuclear genes in animals have been shown to be edited by a variety of mechanisms at various developmental stages and under stresses (Athanasiadis et al., 2004; Blow et al., 2004; Levanon et al., 2004; Englander et al., 2005). In the whole-genome search, we identified eight cytidine and four adenine deaminase genes in The Institute for Genomic Research (TIGR) annotated rice genes. However, only one cytosine deaminase (LOC_Os07g14150) and one adenine deaminase (LOC_Os05g28180) gene were expressed in the RL-SAGE libraries. The expression induction of a deaminase gene (LOC_Os07g14150) upon *M. grisea* inoculation was also shown by Han et al. (2004). The function of these deaminase genes on the RNA variation of rice

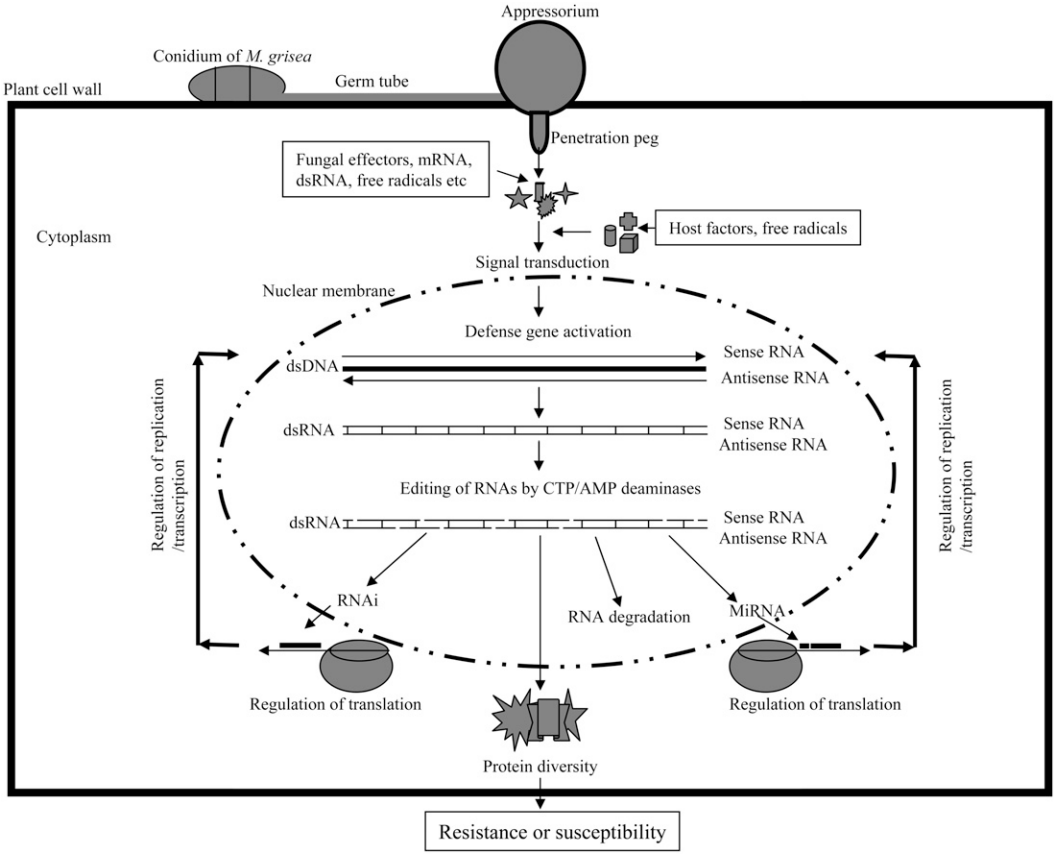

Figure 6. A working model for RNA sequence diversity during rice and *M. grisea* interaction. Once an *M. grisea* conidium lands on the rice leaf surface, an infectious structure, called appressorium, develops at the tip of the germ tube. The highly melanized appressorium generates enormous turgor pressure and penetrates through the rice epidermal cells within 24 h of inoculation (Talbot, 1995). Through the penetration peg, the fungus may secrete various proteins (Ellis et al., 2005), mobile elements (Vaughn et al., 1995), and even free radicals into host cells. Fungal secretory molecules may interact with host factors leading to the synthesis of sense and antisense transcripts for defense-related genes. Then both sense and antisense transcripts may pair to form dsRNA in the nucleus. dsRNA may be targeted by the RNA-editing complex, including cytidine deaminase and adenine deaminase (Nie et al., 2006). Edited dsRNAs might be retained in the nucleus and degraded, generating miRNAs and siRNAs (Blow et al., 2006). Translation of edited transcripts may lead to production of protein diversity (Li et al., 2006). miRNAs and siRNAs may affect translation, transcription, and DNA replication (Li et al., 2006) in the nucleus. RNA variation in host transcripts could affect the outcome of the rice and *M. grisea* interaction: host resistance or susceptibility.

transcripts is not known. Silencing and overexpression of the two genes in transgenic plants should offer insight into the deaminase-mediated RNA editing of nuclear genes in the defense response to pathogens. Recent evidence reveals that the human cytidine deaminase APOBEC family function as an antiretroviral agent to interfere with HIV infection by deaminating cytidine residues to uridine in the nascent minus-strand viral DNA (Turelli and Trono, 2005). Whether rice deaminases also target fungal transcripts or proteins that are secreted into rice cells warrants further investigation.

Many antisense transcripts were identified in the *M. grisea*-infected RL-SAGE libraries in this study. Strand-specific RT-PCR using antisense primers of four genes confirmed their expression pattern in the infected plants. One of the antisense transcripts was the classical defense gene PR1. The function of the induced antisense transcripts in *M. grisea*-infected plants is not known yet. Previous studies have shown that the sense and antisense transcripts might pair in the cells to form dsRNA molecules, which may be the target for RNA editing, splicing, polyadenylation, transport, translation, and small interfering RNA (siRNA)-based gene silencing (Araya et al., 1998; Kumar and Carmichael, 1998; Hayashizaki and Kanamori, 2004; Levanon et al., 2004; Jen et al., 2005). dsRNA could be recognized as substrate by RNA-editing enzymes like adenine and cytidine deaminases (Polson et al., 1991; Smith and Sowden, 1996). It was also recently reported that dsRNA-mediated RNA-editing and interference pathways interact with each other to regulate gene expression (Nishikura, 2006). Therefore, we speculate that highly expressed antisense transcripts in the R and S libraries might lead to the formation of sense and antisense pairs. The formation of these pairs may trigger the RNA-editing machinery in the infected rice plants. Based on our results and other published studies, we propose a working model for RNA variation and sequence diversity during rice and *M. grisea* interaction (Fig. 6). We believe that delivery of effector proteins, mRNA, dsRNA, and free radicals from appressoria into rice epidermal cells could trigger the induction of sense and antisense transcription of defense-related genes, and formation of dsRNAs could lead to the activation of the deaminase-mediated RNA-editing process in the infected cells. Experimental data are being generated to prove these hypotheses.

It is also possible that an alternative pathway might be involved in RNA sequence diversity. It has been reported that various reactive oxygen species (ROS), superoxide anion, hydrogen peroxide, and hydroxyl radical could damage RNA and DNA molecules in living cells under stress conditions. RNA is more susceptible to ROS damage as compared to DNA because RNA molecules are widespread in the cytosol, single stranded, and lack protective proteins and repair mechanisms (Zhang et al., 1999; Hofer et al., 2006). It has also been shown that RNA damage by ROS is related to human diseases such as Alzheimer's (Nunomura et al., 1999; Honda et al., 2005) and Parkinson's (Zhang et al., 1999). Rapid ROS production was documented during rice-*M. grisea* interaction (Pasechnik et al., 1998; Mellersh et al., 2002). Whether the rapid production of ROS is one of the causes for RNA variation observed in this study is unknown and requires further evaluation.

MATERIALS AND METHODS

Rice Blast Fungal Inoculation and Tissue Collection

For the R and S reactions, the avirulent isolate C9240 (from H. Leung, International Rice Research Institute, Philippines) and the virulent isolate Che86061 (from G. Lu, Fujian, China), respectively, were used to inoculate 21-d-old rice (*Oryza sativa*) plants (japonica 'Nipponbare'). Rice plants were grown in a Conviron growth chamber at 80% relative humidity with 12 h of light (500 μmol photons $m^{-2} s^{-1}$) at 26°C followed by 12 h of darkness at 20°C. The ascospores at 2×10^5 spores mL^{-1} in a 0.01% Tween 20 solution were sprayed on rice leaves. C plants were sprayed with the 0.01% Tween 20 solution. The inoculated plants were kept in a sealed plastic container in the dark at 26°C for 24 h. Leaves were harvested 24 h after inoculation and kept in a −80°C freezer until RNA isolation.

Construction of RL-SAGE Libraries and Clone Sequencing

Total RNA was isolated from the infected and C leaves using the TRIzol method (Invitrogen). Poly(A^+) mRNA was purified using the Oligotex mRNA midi kit (Qiagen) according to the manufacturer's instructions. RL-SAGE libraries were generated using an improved procedure reported by Gowda et al. (2004) and Gowda and Wang (2007). A total of 9,120 reads (both ends of 4,560 clones) from the C library, 7,200 reads (7,200 clones) from the R library, and 7,618 reads (7,200 clones) from the S library were obtained using ABI 3730 DNA analyzers (Applied Biosystems) at the Arizona Genome Institute.

Isolation of Ditags and Individual Tags

Sequence analysis showed that the ditag size ranged from 38 to 44 bp (data not shown). To avoid sequence errors caused by imprecise cleavage (slippage) by *Mme*I, only three ditag types (40, 41, and 42 bp) were used for the isolation of 21-bp tags, which accounted for 97% of the ditag types in the analyzed sequences. Second, a ditag must have a CATG site at both ends. Third, any tags with an ambiguous base pair (flagged as N) are discarded. The SAGEspy program developed at the Ohio Supercomputer Center (http://www.osc.edu/research/bioinformatics/sagespy/index.shtml) and analysis tools at the Magnaporthe Grisea Oryza Sativa database (http://www.mgosdb.org/sage; Gowda et al., 2006b) were used to isolate ditags, individual tags, and distinct tags. The RL-SAGE tags from the C, R, and S libraries were submitted to the National Center for Biotechnology Information (GEO accession no. GSE1924; http://www.ncbi.nlm.nih.gov/projects/geo/query/acc.cgi). These data are also available to the public at http://www.mgosdb.org/sage and http://164.107.34.68/wang/index.jsp.

Annotation of RL-SAGE Tags

The 21-bp RL-SAGE tags from the C, R, and S libraries matched against several rice RefSeq databases such as TIGR ESTs, release version 16.0 (http://www.tigr.org/tigr-scripts/tgi/T_release.cgi?species=rice), KOME FL-cDNA sequences (14; http://cdna01.dna.affrc.go.jp/cDNA), and a nonredundant TIGR rice genome annotation database, release 4 of the TIGR pseudomolecules (January 12, 2006; http://www.tigr.org/tdb/e2k1/osa1/pseudomolecules/info.shtml). Total tags in each library were normalized to 100,000 for comparative expression analyses. The one- and two-nucleotide mismatch alignment of RL-SAGE tags against the RefSeq databases were customized using a stand-alone local BLAST 2 program with word size 7 and E value = 0.01. A set of SAGE alignment parsing programs were developed and used to parse and extract BLAST outputs for statistical data analysis and reports.

Identification of Antisense Transcript Tags

To identify the antisense orientation of RL-SAGE tags for the rice RefSeq, we converted all tags into antisense orientation by a reverse-complementation procedure. The antisense tags from the C, R, and S libraries were independently matched against the rice RefSeq. For validating identified antisense tags from the C, R, and S libraries, we matched antisense RL-SAGE tags against longer antisense rice FL-cDNAs available at the KOME database (Osato et al., 2003; http://cdna01.dna.affrc.go.jp/cDNA/Analysis/antisense-web/riceantisense.fasta).

RT-PCR Analysis of Transcripts

About 1.0 μg of total RNA was treated with DNaseI (Invitrogen) and first-strand cDNA synthesized using the alfalfa mosaic virus reverse transcriptase system (Promega) and oligo(dT) primer (Supplemental Table S5). The 3′ region (500 bp) of the Rubisco activase gene was amplified by using 5′-Rubisco activase primer and 3′-RACE primer (Supplemental Table S5). PCR products were purified using a Qiagen kit and cloned into the pGEM-T Easy vector system (Promega). About 30 RT clones were sequenced using the M13 forward primer. RT-PCR analysis of adenine deaminase and cytidine deaminase genes was done using gene-specific primers that are listed in Supplemental Table S5.

To amplify antisense transcripts, the strand-specific RT-PCR method was performed as described previously by Røsok and Sioud (2004) and Ge et al. (2006). Total RNA isolated from the C, R, and S reactions was treated with DNaseI (Invitrogen) and first-strand cDNA was synthesized at 42°C for 1 h by using an antisense primer and the Promega RT system (Promega). Then the cDNA was PCR amplified using gene-specific reverse and forward primers (Supplemental Table S5) for 25 cycles of 30 s at 94°C, 30 s at 55°C, and 1 min for 72°C. To quantify the RNA samples from the C, R, and S reactions for strand-specific RT-PCR, total RNA of each sample was subjected to RT using an oligo(dT) primer and then RT-PCR was carried out using ubiquitin gene primers.

Sequence data from this article can be found in the GenBank/EMBL data libraries under accession numbers GSM34282 to GSM34284.

Supplemental Data

The following materials are available in the online version of this article.

Supplemental Figure S1. Sequence quality score for the Rubisco activase transcript tags.

Supplemental Figure S2. Chromatogram of RT-PCR sequences of the Rubisco activase transcripts.

Supplemental Figure S3. Copy number distribution of the antisense tags in the three libraries.

Supplemental Figure S4. RT-PCR confirmation of four antisense transcripts in infected rice plants.

Supplemental Table S1. RL-SAGE tags only matched the genome and annotated sequences of *M. grisea*.

Supplemental Table S2. Genes with mismatched tags from the C, R, and S libraries.

Supplemental Table S3. Cytidine and adenine deaminases in rice and their orthology in other organisms.

Supplemental Table S4. Antisense RL-SAGE tags matched to the KOME antisense FL-cDNAs.

Supplemental Table S5. Primers used for RT-PCR analysis.

ACKNOWLEDGMENTS

We thank Brandon Childers, Shankaraganesh Manikantan, and James Hatfield for their assistance in data analysis. We are also grateful to Rose Palumbo and Miguel Vega-Sanchez for their critical reading of this manuscript and to all the lab members for their valuable help and discussion during the course of work.

Received January 8, 2007; accepted February 15, 2007; published March 9, 2007.

LITERATURE CITED

Araya A, Zabaleta E, Blanc V, Begu D, Hernould M, Mouras A, Litvak S (1998) RNA editing in plant mitochondria, cytoplasmic male sterility and plant breeding. Biotechnol J **1:** 31–39

Athanasiadis A, Rich A, Maas S (2004) Widespread A-to-I RNA editing of Alu-containing mRNAs in the human transcriptome. PLoS Biol **2:** e391

Blow M, Futreal PA, Wooster R, Stratton MR (2004) A survey of RNA editing in human brain. Genome Res **14:** 2379–2387

Blow MJ, Grocock RJ, van Dongen S, Enright AJ, Dicks E, Futreal PA, Wooster R, Stratton MR (2006) RNA editing of human microRNAs. Genome Biol **7:** R27

Chen J, Sun M, Lee S, Zhou G, Rowley JD, Wang SM (2002) Identifying novel transcripts and novel genes in the human genome by using novel SAGE tags. Proc Natl Acad Sci USA **99:** 1257–1262

Dean RA, Talbot NJ, Ebbole DJ, Farman ML, Mitchell TK, Orbach MJ, Thon M, Kulkarni R, Xu JR, Pan H, et al (2005) The genome sequence of the rice blast fungus *Magnaporthe grisea*. Nature **434:** 980–986

Ellis J, Catanzariti AM, Dodds P (2005) The problem of how fungal and oomycete avirulence proteins enter plant cells. Trends Plant Sci **11:** 61–63

Englander MT, Dulawa SC, Bhansali P, Schmauss C (2005) How stress and fluoxetine modulate serotonin 2C receptor pre-mRNA editing. J Neurosci **25:** 648–651

Ge X, Wu Q, Jung YC, Chen J, Wang SM (2006) A large quantity of novel human antisense transcripts detected by LongSAGE. Bioinformatics **22:** 2475–2479

Goff SA, Ricke D, Lan TH, Presting G, Wang R, Dunn M, Glazebrook J, Sessions A, Oeller P, Varma H, et al (2002) A draft sequence of the rice genome (*Oryza sativa* L. ssp. *japonica*). Science **296:** 92–100

Gowda M, Jantasuriyarat C, Dean RA, Wang GL (2004) Robust-LongSAGE (RL-SAGE): a substantially improved LongSAGE method for gene discovery and transcriptome analysis. Plant Physiol **134:** 890–897

Gowda M, Li H, Alessi J, Chen F, Pratt R, Wang GL (2006a) Robust analysis of 5′-transcript ends (5′-RATE): a novel method for transcriptome profiling and genome annotation. Nucleic Acids Res **34:** e126

Gowda M, Venu RC, Jia Y, Stahlberg E, Pampanwar V, Soderlund C, Wang GL (2006b). Use of RL-SAGE analysis to identify novel fungal and plant genes involved in host-pathogen interactions. *In* PC Ronald, ed, Methods in Molecular Biology: Plant-Pathogen Interactions. Humana Press, Inc., Totowa, NJ, pp 131–144

Gowda M, Venu RC, Raghupathy MB, Nobuta K, Li H, Wing R, Stahlberg E, Couglan S, Haudenschild CD, Dean R, et al (2006c) Deep and comparative analysis of the mycelium and appressorium transcriptomes of *Magnaporthe grisea* using MPSS, RL-SAGE, and oligoarray methods. BMC Genomics **7:** 310

Gowda M, Wang GL (2007) Robust-LongSAGE (RL-SAGE): an improved LongSAGE method for high-throughput transcriptome analysis. *In* KL Nielsen, ed, Methods in Molecular Biology: Serial Analysis of Gene Expression. Humana Press, Inc., Totowa, NJ

Gunasekera AM, Patankar S, Schug J, Eisen G, Kissinger J, Roos D, Wirth DF (2004) Widespread distribution of antisense transcripts in the *Plasmodium falciparum* genome. Mol Biochem Parasitol **36:** 35–42

Han CU, Lee CH, Jang KS, Choi GJ, Lim HK, Kim JC, Ahn SN, Choi JE, Cha JS, Kim HT, et al (2004) Identification of rice genes induced in a rice blast-resistant mutant. Mol Cells **17:** 462–468

Hayashizaki Y, Kanamori M (2004) Dynamic transcriptome of mice. Trends Biotechnol **22:** 161–167

Hofer T, Seo AY, Prudencio M, Leeuwenburgh C (2006) A method to determine RNA and DNA oxidation simultaneously by HPLC-ECD: greater RNA than DNA oxidation in rat liver after doxorubicin administration. Biol Chem **387:** 103–111

Honda K, Smith MA, Zhu X, Baus D, Merrick WC, Tartakoff AM, Hattier T, Harris PL, Siedlak SL, Fujioka H, et al (2005) Ribosomal RNA in Alzheimer disease is oxidized by bound redox-active iron. J Biol Chem **280:** 20978–20986

International Rice Genome Sequencing Project (2005) The map-based sequence of the rice genome. Nature **436:** 793–800

Jantasuriyarat C, Gowda M, Haller K, Hatfield J, Lu G, Stahlberg E, Zhou B, Li H, Kim H, Yu Y, et al (2005) Large-scale identification of expressed sequence tags involved in rice and rice blast fungus interaction. Plant Physiol **138:** 105–115

Jen CH, Michalopoulos I, Westhead DR, Meyer P (2005) Natural antisense transcripts with coding capacity in Arabidopsis may have a regulatory role that is not linked to double-stranded RNA degradation. Genome Biol **6:** R51

Kim CY, Lee SH, Park HC, Bae CG, Cheong YH, Choi YJ, Han C, Lee SY, Lim CO, Cho MJ (2000) Identification of rice blast fungal elicitor-responsive genes by differential display analysis. Mol Plant Microbe Interact **13:** 470–474

Kim KM, Cho SK, Shin SH, Kim GT, Lee JH, Oh BJ, Kang KH, Hong JC, Choi JY, Shin JS, et al (2005) Analysis of differentially expressed transcripts of fungal elicitor- and wound-treated wild rice (*Oryza grandiglumis*). J Plant Res **118:** 347–354

Kim S, Ahn IP, Lee YH (2001) Analysis of genes expressed during rice-*Magnaporthe grisea* interactions. Mol Plant Microbe Interact **14:** 1340–1346

Kumar M, Carmichael GG (1998) Antisense RNA: function and fate of duplex RNA in cells of higher eukaryotes. Microbiol Mol Bio Rev **62:** 1415–1434

Levanon EY, Eisenberg E, Yelin R, Nemzer S, Hallegger M, Shemesh R, Fligelman ZY, Shoshan A, Pollock SR, Sztybel D, et al (2004) Systematic identification of abundant A-to-I editing sites in the human transcriptome. Nat Biotechnol **22:** 1001–1005

Li LC, Okino ST, Zhao H, Pookot D, Place RF, Urakami S, Enokida H, Dahiya R (2006) Small dsRNAs induce transcriptional activation in human cells. Proc Natl Acad Sci USA **103:** 17337–17342

Lu G, Jantasuriyarat C, Zhou B, Wang GL (2004) Isolation and characterization of novel defense response genes involved in compatible and incompatible interactions between rice and *Magnaporthe grisea*. Theor Appl Genet **108:** 525–534

Mellersh DG, Foulds IV, Higgins VJ, Heath MC (2002) H_2O_2 plays different roles in determining penetration failure in three diverse plant-fungal interactions. Plant J **29:** 257–268

Nie Y, Hammond GL, Yang JH (2006) Double stranded RNA deaminase ADAR1 increases host susceptibility to virus infection. J Virol **81:** 917–923

Nishikura K (2006) Editor meets silencer: crosstalk between RNA editing and RNA interference. Nat Rev Mol Cell Biol **7:** 919–931

Nunomura A, Perry G, Pappolla MA, Wade R, Hirai K, Smith MA (1999) RNA oxidation is a prominent feature of vulnerable neurons in Alzheimer disease. J Neurosci **19:** 1959–1964

Osato N, Yamada H, Satoh K, Ooka H, Yamamoto M, Suzuki K, Kawai J, Carninci P, Ohtomo Y, Murakami K, et al (2003) Antisense transcripts with rice full-length cDNAs. Genome Biol **5:** R5

Pasechnik TD, Aver'yanov AA, Lapikova VP, Kovalenko ED, Kolomietz TM (1998) The involvement of activated oxygen in the expression of the vertical and horizontal resistance of rice to blast disease. Russ J Plant Physiol **45:** 371–378

Polson AG, Crain PF, Pomerantz SC, McCloskey JA, Bass BL (1991) The mechanism of adenosine to inosine conversion by the double-stranded RNA unwinding/modifying activity: a high-performance liquid chromatography-mass spectrometry analysis. Biochemistry **30:** 11507–11514

Poroyko V, Hejlek LG, Spollen WG, Springer GK, Nguyen HT, Sharp RE, Bohnert HJ (2005) The maize root transcriptome by serial analysis of gene expression. Plant Physiol **138:** 1700–1710

Quere R, Manchon L, Lejeune M, Clement O, Pierrat F, Bonafoux B, Commes T, Piquemal D, Marti J (2004) Mining SAGE data allows large-scale, sensitive screening of antisense transcript expression. Nucleic Acids Res **32:** e163

Røsok O, Sioud M (2004) Systematic identification of sense-antisense transcripts in mammalian cells. Nat Biotechnol **22:** 104–108

Saha S, Sparks AB, Rago C, Akmaev V, Wang CJ, Vogelstein B, Kinzler KW, Velculescu VE (2002) Using the transcriptome to annotate the genome. Nat Biotechnol **20:** 508–512

Smith HC, Sowden MP (1996) Base-modification mRNA editing through deamination—the good, the bad and the unregulated. Trends Genet **12:** 418–424

Talbot NJ (1995) Having a blast: exploring the pathogenicity of *Magnaporthe grisea*. Trends Microbiol **3:** 9–16

Turelli P, Trono D (2005) Editing at the crossroad of innate and adaptive immunity. Science **307:** 1061–1065

Vaughn JC, Mason MT, Sper-Whitis GL, Kuhlman P, Palmer JD (1995) Fungal origin by horizontal transfer of a plant mitochondrial group I intron in the chimeric coxI gene of Peperomia. J Mol Evol **41:** 563–572

Velculescu VE, Madden SL, Zhang L, Lash AE, Yu J, Rago C, Lal A, Wang CJ, Beaudry GA, Ciriello KM, et al (1999) Analysis of human transcriptomes. Nat Genet **23:** 387–388

Velculescu VE, Zhang L, Vogelstein B, Kinzler KW (1995) Serial analysis of gene expression. Science **270:** 484–487

Xiong L, Lee MW, Qi M, Yang Y (2001) Identification of defense-related rice genes by suppression subtractive hybridization and differential screening. Mol Plant Microbe Interact **14:** 685–692

Yu J, Hu S, Wang J, Wong GK, Li S, Liu B, Deng Y, Dai L, Zhou Y, Zhang X, et al (2002) A draft sequence of the rice genome (*Oryza sativa* L. ssp. indica). Science **296:** 79–92

Zhang J, Perry G, Smith MA, Robertson D, Olson SJ, Graham DG, Montine TJ (1999) Parkinson's disease is associated with oxidative damage to cytoplasmic DNA and RNA in substantia nigra neurons. Am J Pathol **154:** 1423–1429