MPMI Vol. 25, No. 3, 2012, pp. 271–278. http://dx.doi.org/10.1094/MPMI-07-11-0183. © 2012 The American Phytopathological Society *e*-Xtra*

TECHNICAL ADVANCE

MGOS: Development of a Community Annotation Database for *Magnaporthe oryzae*

Anupreet Kour,[1] Kevin Greer,[2] Barbara Valent,[3] Marc J. Orbach,[1,2] and Carol Soderlund[2]

[1]School of Plant Sciences. Division of Plant Pathology and Microbiology, The University of Arizona, Tucson 85721, U.S.A.; [2]BIO5 Institute, The University of Arizona, Tucson 85719, U.S.A.; [3]Department of Plant Pathology, Kansas State University, Manhattan 66506-5502, U.S.A.

Submitted 7 July 2011. Accepted 1 November 2011.

Magnaporthe oryzae **causes rice blast disease, which is the most serious disease of cultivated rice worldwide. We previously developed the** ***Magnaporthe grisea–Orzya sativa*** **(MGOS) database as a repository for the** ***M. oryzae*** **and rice genome sequences together with a comprehensive set of functional interaction data generated by a major consortium of U.S. researchers. The MGOS database has now undergone a major redesign to include data from the international blast research community, accessible with a new intuitive, easy-to-use interface. Registered database users can manually annotate gene sequences and features as well as add mutant data and literature on individual gene pages. Over 900 genes have been manually curated based on various biological databases and the scientific literature. Gene names and descriptions, gene ontology annotations, published and unpublished information on mutants and their phenotypes, responses in diverse microarray analyses, and related literature have been incorporated. Thus far, 362** ***M. oryzae*** **genes have associated information on mutants. MGOS is now poised to become a one-stop repository for all structural and functional data available on all genes of this critically important rice pathogen.**

Rice blast is the most important disease that affects global rice production. The importance of this disease to food security is underscored by the fact that rice contributes 23% of the calories consumed by the global human population (Wilson and Talbot 2009). Rice is the most important food product in Asia, where 55% of the world's population lives and 92% of rice is grown and consumed. The durability of many blast-resistant cultivars of rice is poor, with a typical field life of only two to three growing seasons before disease resistance is overcome. Furthermore, rising energy costs impact production by affecting fungicide and fertilizer prices (Ou 1985; Wang and Valent 2009). Thus, there is a need for a better understanding of this disease so that environmentally sustainable pathogen control strategies can be deployed toward increasing the efficiency of cereal cultivation.

Rice blast disease is caused by the filamentous ascomycete fungus *Magnaporthe oryzae*. The availability of the *M. oryzae* genome sequence has fundamentally altered the manner in which the biology of rice blast disease can be explored (Dean et al. 2005). The version 6 of Broad Institute genome assembly of *M. oryzae* that was used for redesign of the *Magnaporthe grisea–Orzya sativa* (MGOS) database shows a genome size of 41.7 Mb with 11,074 genes. Of these genes, approximately 35% have a known or predicted role, and a few pathogenicity genes and avirulence effector genes have also been characterized (Ebbole 2007; Wilson and Talbot 2009). Nevertheless, detailed knowledge of the regulation of pathogen functions (such as adhesion, penetration, and invasive growth) and their control (e.g., how the surface cues perceived by the fungus are linked to the activation and operation of cAMP and PMK1 mitogen-activated protein kinase pathways) in this fungus are limited. Development of the MGOS database is our effort to help the international scientific community to understand the mechanisms involved in pathogenesis. Availability of the whole genome sequence, mutant information, and gene expression data, along with manual curation, makes this database a valuable community resource.

Biological databases have become one of the principal drivers of research and innovation in biology. For fungi, model organism databases, such as the *Saccharomyces* genome database (Christie et al. 2009), the *Aspergillus* genome databases, and the *Candida* genome database, contain enormous amounts of high-quality annotated data and are an excellent source of information on fungal genome-scale biology. Moreover, these databases contain genome statistics as well as information on gene names, descriptions, gene ontology (GO) annotations, mutant phenotypes, expression data, and related literature. Genome browser options provide a clear view of the genome and its features, and web-based research tools are provided for accessing and exploring the data. Variable degrees of community annotation are also present in these databases. A complete set of annotation data provides the most detailed picture available for each locus, including all clues to gene function. Such sequence annotations are crucial resources for the scientific community engaged in identification and characterization of genes and their products (Stein 2001). Inaccuracies in automatic gene identification and function annotations provide challenges for any database (Menda et al. 2008). With this in mind, we have expanded the MGOS database to provide a user-friendly community annotation interface for *Magnaporthe* spp.

Originally, the MGOS database was developed to store experimental data from both the host and pathogen in order to study the interactions between rice and the rice blast fungus (Soderlund et al. 2006). The focus of the database has changed

Corresponding authors: M. J. Orbach; E-mail: orbachmj@ag.arizona.edu and C. Soderlund; E-mail: cari@agcol.arizona.edu

*The *e*-Xtra logo stands for "electronic extra" and indicates that Figures 1 through 6 appear in color online.

to providing in-depth data for *M. oryzae* along with community annotation capabilities. The original database contained genome sequence, expressed sequenced tags (EST), SAGE, and mutant data, and all these data types, except for SAGE, have been augmented with new data since 2006. The new MGOS contains data from many different sources, including the Broad Institute, National Center for Biotechnology Information (NCBI), an *M. oryzae Agrobacterium tumefaciens*–mediated transformation (ATMT) database, pathogen–host interaction database (PHI-base), and the e-fungi database (Hedeler et al. 2007). Gene expression data include EST (Ebbole et al. 2004; Numa et al. 2009), robust long (RL)-SAGE (Gowda et al. 2004, 2007), and microarrays (Oh et al. 2008). Mutant phenotype data are from a collection of over 50,000 *M. oryzae* DNA insertion lines (Betts et al. 2007; Donofrio et al. 2005) and from published literature. All relevant information is associated with each gene and a versatile interface allows complex queries.

The current MGOS assembly is advanced not only in terms of new data from several resources but also in the community annotation interface. Manual annotation in the original MGOS database was limited to submission of data on a link provided on the gene list page. Now, registered community members have the ability to directly edit all types of gene information; for example, coordinates (automatically computed from aligning a transcript to the genome), gene ontology, mutants, and publications. There are also full editing capabilities to add or edit mutants, including assays, comments, and images. The forum for registered users allows them to perform annotations and use the LISTSERV to communicate with all other registered members. Using these tools, the MGOS team has annotated over 900 genes and included 190 pertinent publications. The overall layout of the database architecture along with the manual curation interface is shown in Figure 1.

MGOS data contents.

MGOS has five types of experimental data (EST, SAGE, microarray, mutants, and genes in their genome context) plus publications, which all have their own query page with bidirectional links to associated gene pages. The database currently contains 11,044 gene models on seven linkage groups, 62,258 EST, 51,927 *M. oryzae* SAGE tags, microarray data from five experiments, and 24,000 mutants with altered phenotypes.

Gene expression data.

The EST libraries are from appressoria, mycelia, conidia, germinated conidia, and mixed stages of sexual development. These libraries are from different strains, including 70-15, Guy11, CP987, NN95, P1.2, and crosses between 4091-5-8 and 4136-4-3 (Ebbole et al. 2004; Numa et al. 2009). In total, 62,000 EST were assembled using the program for assembling

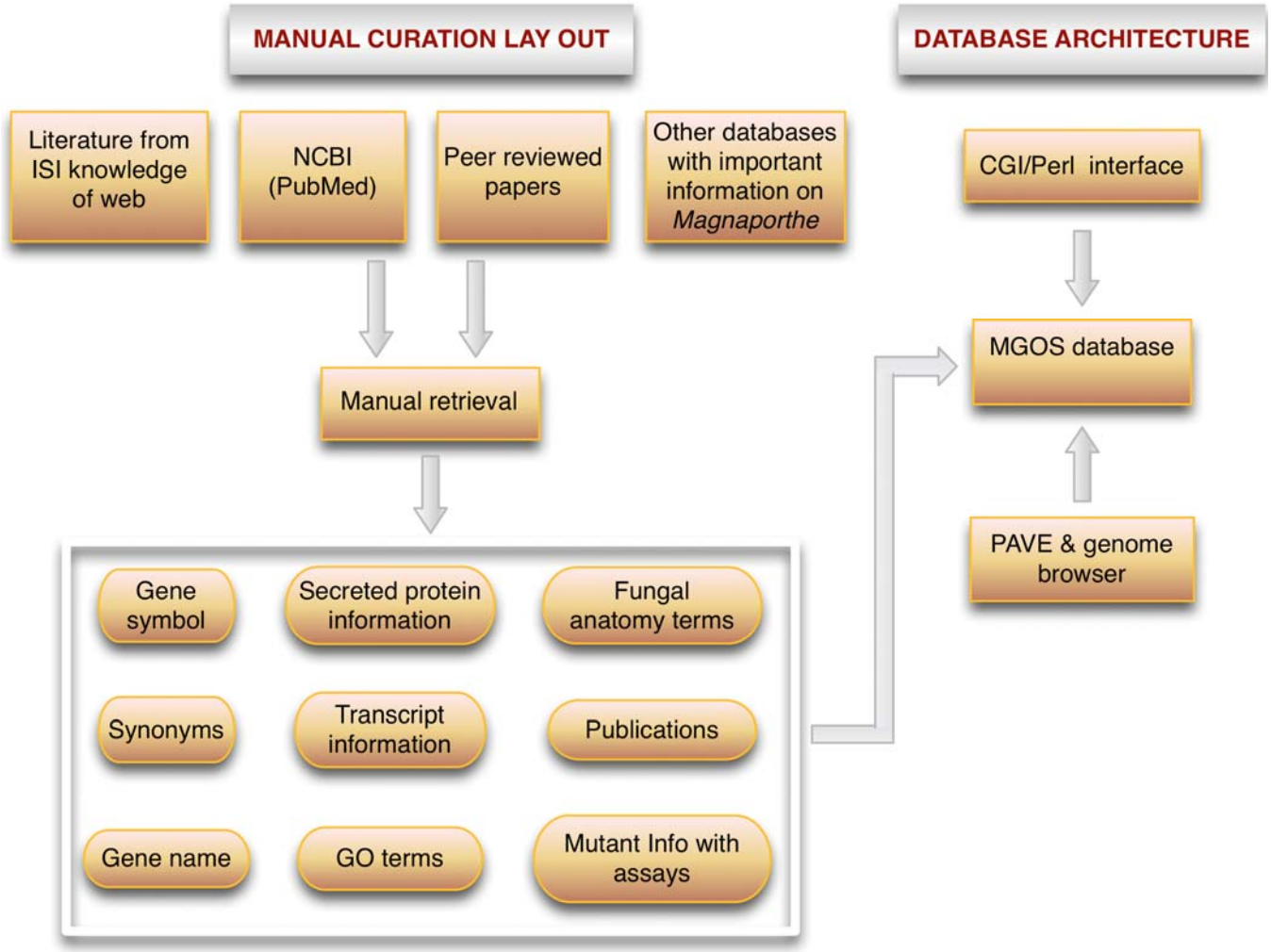

Fig. 1. Overall presentation of manual curation lay out along with database architecture. The *Magnaporthe grisea–Orzya sativa* (MGOS) database uses the GMOD Chado Postgress database (Mungall and Emmert 2007) and a custom MySQL database, and it uses the GMOD genome browser (Stein et al. 2002) that is integrated with the Chado database.

and viewing EST (PAVE) (Soderlund et al. 2009) into 17,013 unitrans (unique transcripts), of which 8,873 are singletons. The unitrans were annotated with the R statistic (Stekel et al. 2000), GC content, longest open reading frames, best UniProt match (Bairoch et al. 2005), GO (Camon et al. 2004) and GOSlim (Biswas et al. 2002) matches, plus the EST abundance from each library. There are 5,119 genes in MGOS that have one or more associated unitrans.

RL-SAGE data are from RNA extracted from 'Nipponbare' rice infected with *M. oryzae* 24 and 96 h after treatment from a control library where Nipponbare was sprayed with water only, and an *M. oryzae* library grown on a low nutrient medium (Gowda et al. 2004). There are 4,645 MGOS genes with at least one associated SAGE tag.

Data from five microarray experiments were downloaded from GenBank and entered into MGOS, and the 10,126 probes were aligned to the gene models. These experiments were performed to define global gene expression under several defined conditions (appressorium induction and formation, nitrogen starvation, response in the dark, and during early biotrophic invasion of rice seedlings) (Donofrio et al. 2006; Mosquera et al. 2009; Oh et al. 2008) (GEO accession, GSE 12598 from NCBI; *unpublished*).

Mutants.

The original MGOS contained phenotype screening information from randomly tagged mutants (Betts et al. 2007); the new MGOS continues to support this information and has an enhanced search capability for the data. These mutants are available to researchers in the Fungal Genetics Stock Center (FGSC). MGOS has a compiled list of these mutants with the corresponding FGSC stock ID number. There are 24,000 mutants that have associated phenotype information available in MGOS, and 548 of them have locus tags associated with them.

The mutant information in MGOS includes unpublished data provided by the scientific community as well as published data. Many laboratories have produced gene-replacement mutants that lack any altered phenotype and, therefore, are not likely to be published. Information on 140 such mutants, representing 107 genes, has been incorporated into the database (R. Terauchi, *unpublished data*). Mutant phenotype information was also incorporated from PHI-base (Winnenburg et al. 2008) and an ATMT database (Jeon et al. 2007). The ATMT database stores data from specific experiments in which hygromycin-resistant *M. oryzae* mutants (Korean strain KJ201) were generated through insertional mutagenesis using ATMT. Information on 93 of the strains with altered phenotypes, along with their DNA locus tags, was extracted and incorporated into MGOS manually. PHI-base contains expertly curated molecular and biological information on genes proven to affect the outcome of a pathogen–host interaction. The 162 *Magnaporthe* genes with mutant information in PHI-base were collected manually and included in MGOS. Data in the primary literature was the other important source of information used for manual annotation. The following mutant information was collected from the published literature and incorporated into MGOS: mutant name, method of generation, and description of assays used to assess pathogenicity and other altered phenotypes. There are 362 genes in MGOS with mutant information.

Genome and genes.

The *M. oryzae* sequenced genome and predicted gene set were provided by the Broad Institute. The MGOS genes (prefix MC) were duplicated from the Broad annotation version 6 genes (prefix MGG) using the same gene number identifier (e.g., MC_0001 is derived from MGG_0001). Because the MGOS genes were manually annotated, their structure diverged from the Broad genes. The PAVE unitrans sequences (consensus sequences of contigs and singletons) and SAGE tags have been aligned to the genome sequence using BLAT (Kent 2002). All unitrans hits with less than 98% identity were filtered out, with SAGE matches requiring 100% identity. The GenBank *M. grisea* proteins were aligned to the genome using BLAST (Altschul et al. 1990). All alignments overlapping a gene are associated with the gene, and 529 genes have at least one GenBank protein alignment.

All information for each gene is available on a single page that combines all the associated results from the gene expression experiments, mutant data, and GenBank proteins. It contains a gene description field and InterPro information (Hunter et al. 2009), which were populated from the downloadable Broad Institute gene information file. Information on secreted proteins is from the e-fungi database (Hedeler et al. 2007). There are links to the Broad Institute *M. oryzae* database, NCBI proteins, and all MGOS-associated data.

MGOS community annotation.

Forum. The SimpleMachines forum was installed to allow community members to register. Registered users can send email to all other registered users, post comments and suggestions, and perform MGOS edits.

Editing genes. To edit a gene, the user finds their gene of interest through the query system, which results in the display of the gene page. By logging in, the edit mode of the gene page is opened (Fig. 2), where gene symbol, synonyms, gene descriptions, gene comments, and secreted protein information can directly be added or edited. To date, there have been 196 gene symbols and 191 comments added to genes in the database by manual annotation.

There are two approaches to editing a transcript sequence. In the first approach, one accesses the transcript annotation page to add, delete, or modify exons or change the start and end of

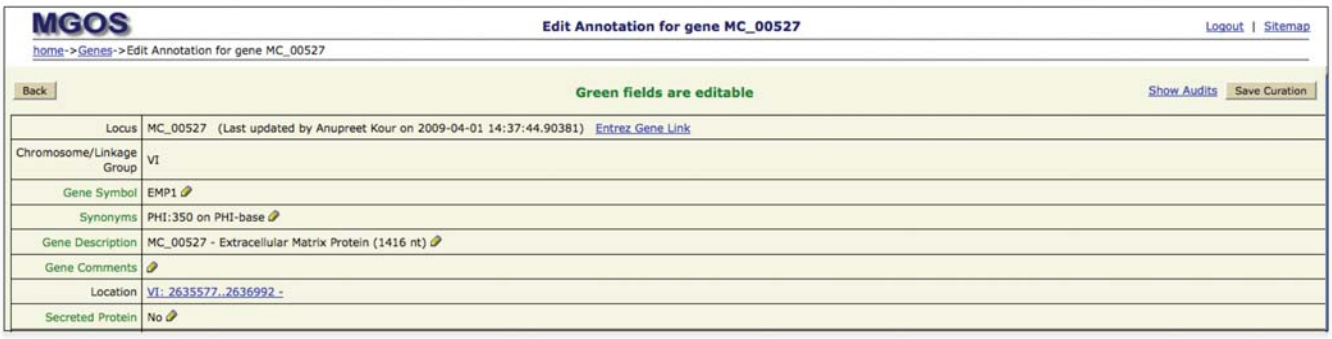

Fig. 2. *Magnaporthe grisea–Orzya sativa* (MGOS) database gene page with gene symbol, synonyms, gene description, gene comments, and secreted protein editable fields. Green labels indicate editable fields, along with the pencil icon.

transcripts. Alternatively, one may use BLAT to first align a sequence (EST, EST contig consensus sequence, and so on) against the genomic gene sequence to automatically determine the exon and intron boundaries (Fig. 3A and B). Upon user request, these coordinates are automatically retrieved back into transcript edit screen, and the resulting genomic, transcribed, and translated sequences are all displayed to allow the user to verify that the coordinates produced the expected sequence. In the majority of cases, it will not be necessary to change these calculated coordinates. However, if necessary, the user can adjust the coordinates before saving any changes. A change in the nucleotide sequence will automatically change the corresponding amino acid sequence. Any stop codons within the coding sequence are highlighted in the amino acid sequence, because this indicates an incorrect sequence.

The gene edit page has an area to display, modify, and add associated publications (Fig. 4). When adding a publication manually, the interface only requires the entry of the PubMed ID, and the corresponding reference along with abstract will appear at the MGOS publication link. To date, there are 3,054 loci with publications associated with them.

Adding information on mutants. The new MGOS provides a versatile approach for adding mutant information to a gene. There are options to add strain information and to indicate the mutant name and describe how it was generated (e.g., by targeted gene knockout, insertional mutagenesis, spontaneous mutation, and so on). There are columns available for adding information on different assays on the edit mutant page. A particular assay can be added by selecting for the assay name and adding a brief assay description and the results obtained from that assay. Images indicating mutant phenotypes, such as morphology or phenotypic assays, can be uploaded and linked to this page. Thus, this page provides a platform to describe phenotypic assays used to analyze the difference between the mutant and wild type.

Information on mutants related to a particular locus has been added manually through MGOS's interactive system, and there are 362 loci in MGOS with mutant information. There are 170

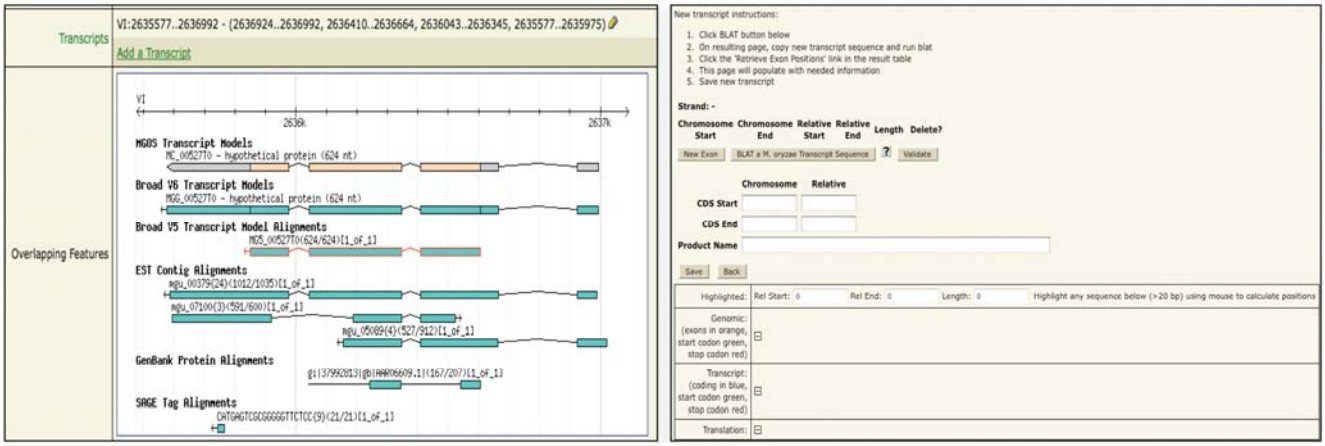

Fig. 3. Left: Transcript annotation option on the gene page of the *Magnaporthe grisea–Orzya sativa* (MGOS) database. Right: Existing transcripts can be edited by using the transcript annotation feature according to the provided instructions on this page.

Interpro			
Evidence	**EST Contig:** mgu_05089 **EST Contig:** mgu_00379 **EST Contig:** mgu_07100 **SAGE Tag:** CATGAGTCGCGGGGGTTCTCC **GenBank Protein:** gi|37992813|gb|AAR06609.1| **BROAD:** MGG_00527 - hypothetical protein (1416 nt)		
Interaction Terms	Add an interaction Term (Browse GO Terms)		
Gene Ontology Terms	Add a GO Term (Browse GO Terms)		
Fungal Anatomy Terms	Add a Fungal Anatomy Term (FAO Website)		
Publications	Ahn N, Kim S, Choi W, Im KH, Lee YH. Extracellular matrix protein gene, EMP1, is required for appressorium formation and pathogenicity of the rice blast fungus, Magnaporthe grisea. Mol Cells. 2004 Feb;17(1):166-73.		
	Baldwin TK, Winnenburg R, Urban M, Rawlings C, Koehler J, Hammond-Kosack KE. The pathogen-host interactions database (PHI-base) provides insights into generic and novel themes of pathogenicity. Mol Plant Microbe Interact. 2006 Dec;19(12):1451-62.		
	Winnenburg R, Urban M, Beacham A, Baldwin TK, Holland S, Lindeberg M, Hansen H, Rawlings C, Hammond-Kosack KE, KÃ¶hler J. PHI-base update: additions to the pathogen host interaction database. Nucleic Acids Res. 2008 Jan;36(Database issue):D572-6.		
	Add a Citation		
Mutants	2A5	Gene replacement	Appressorium Formation: reduced appressorium formation Pathogenicity: Reduced pathogenic
	2A9	Gene replacement	
	Add a Mutant		

Fig. 4. *Magnaporthe grisea–Orzya sativa* (MGOS) database gene page with ontology, publication, and mutant editable fields. Ontologies link to their respective site for browsing. Publications can be added by specifying the PubMed identifier or by entering all the fields. Mutants can be edited or added.

mutants with reduced pathogenicity and an associated locus tag. In total, 5,500 mutants in MGOS are reduced in pathogenicity but have no locus tag associated with them. These mutants are available at the FGSC via the identification number present in MGOS. Mutant information related to a particular gene appears on that gene's page (Fig. 4). The gene page contains information regarding phenotypic differences of mutants when compared with the wild type, along with a short description of the assay used to measure that difference (Fig. 5).

GO or fungal anatomy annotations. Both manual and automated methods were used to assign GO information to loci in MGOS, resulting in 2,704 loci annotated with GO terms. Ontologies developed by the PAMGO consortium (Meng et al. 2009) for fungi were used to characterize the biological processes, molecular functions, and cellular components in which a locus is involved. The database contains the list of fungal anatomy terms from the Fungal Anatomy ontology project, which are available to a user to add in their annotation of a locus.

Quality control. In any database, quality control is the most important aspect to maintain its integrity. To that end, MGOS incorporates auditing functionality that stores a history of all changes made to the database. Each gene page contains information regarding annotation history and who performed the annotations; the history is organized by the date of annotation.

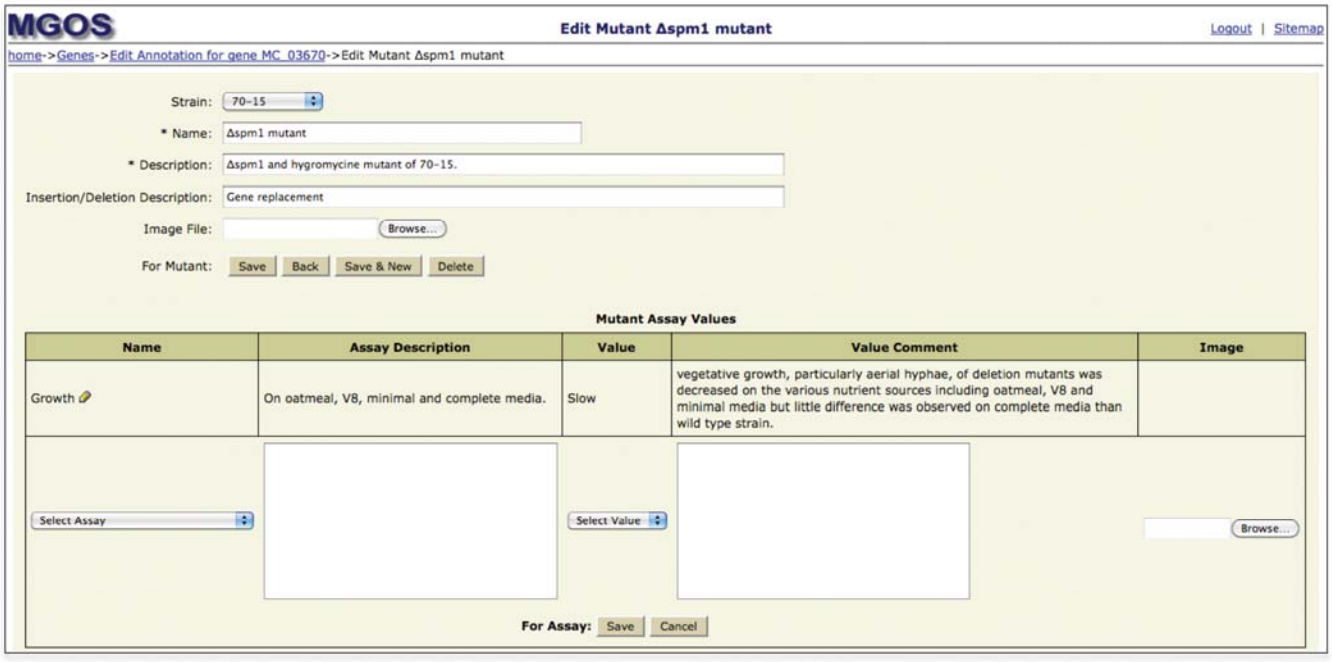

Fig. 5. Mutant page to add and edit information regarding a mutant. A set of predefined assays has been identified for the user to select from (this allows the user to search for a given set of assays from the mutant search page).

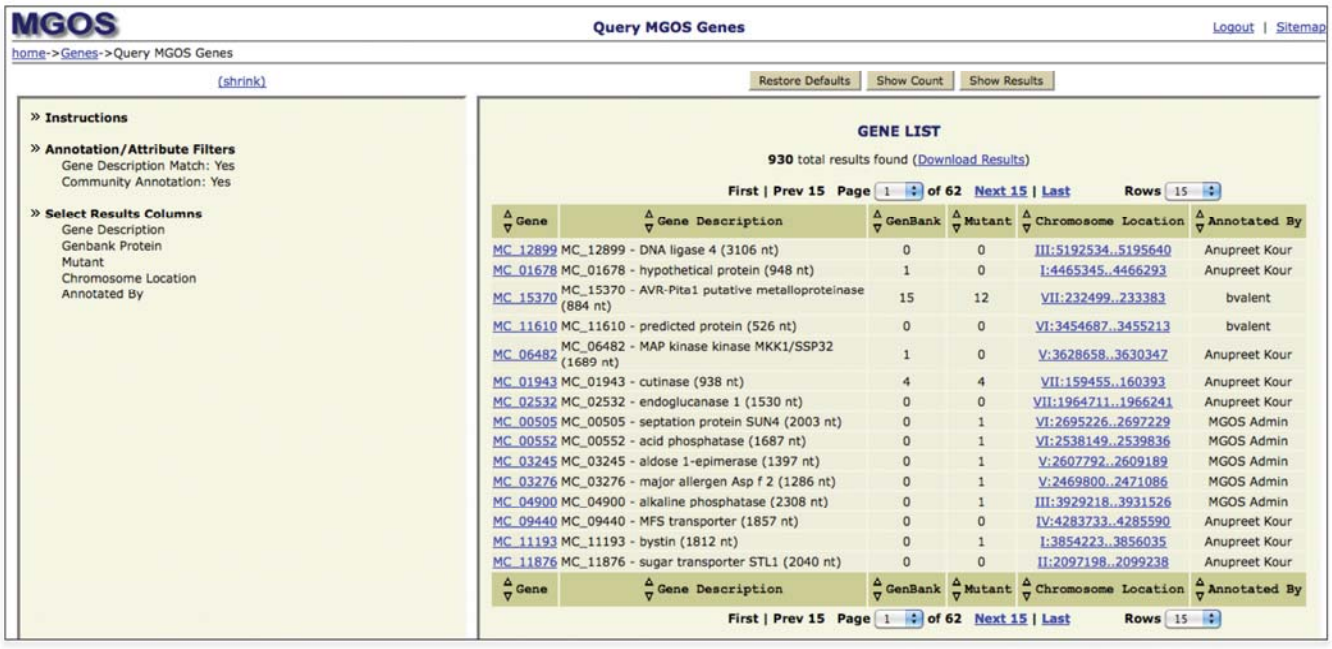

Fig. 6. Gene Advanced Search Page. A set of search criteria and the columns to be viewed can be selected. For example, a search was performed for all genes that have a gene description match and have been manually annotated.

An additional layer of control on data input is the restrictions on the data type that can be added in a specific field. For example, to describe functions and phenotypes of a locus, annotators are limited to using ontologies from a browseable list already existing in the MGOS database; this maintains consistency of terminology.

MGOS query interface.

The MGOS interface uses a BioMart style (Smedley et al. 2009) for most searches, where the user selects the desired filters and columns to be shown in the resulting table. An example of using the advanced search for genes to show only those that have been manually annotated and have a gene description match is given in Figure 6. There are search options and columns for each data type in the gene page. Both EST and mutants have an advanced search page using the same BioMart style. The EST advanced search page allows the user to select contigs that have EST from any user-specified subset of libraries, and search on UniProt information or other attributes of the contigs. The mutant advanced search page allows for searching on any subset of the assay filters. A simple search page is available for publications and SAGE data, and the microarray page links into information about each experiment.

The genomic sequence is displayed using the Generic Model Organism Database (GMOD) Generic Genome Browser (GBrowse) (Stein et al. 2002) with the Chado (Mungall and Emmert 2007) adapters, where tracks for all of the evidence data related to a MGOS gene model can be viewed. The tracks include MGOS gene models, BROAD V6 and V5 gene models, the PAVE EST contigs, SAGE tags, and GenBank *M. grisea* proteins. All track entities link to their detail page. Regions of a particular chromosome can be selected and displayed in this browser.

DISCUSSION

The MGOS site brings together public, published, and unpublished data, including the genome sequence, mutant genotypes and phenotypes, controlled vocabulary annotations, and literature citations. For understanding the overall biology underlying the pathogenic lifestyle of fungal plant pathogens, a valuable step is gaining information on changes in global gene function through analysis of which genes are expressed in different tissues and growth conditions. MGOS contains EST data, microarray data, and SAGE data toward understanding global gene expression. The EST are critical for verifying mRNA structure and, thus, predicted protein sequences. Currently, there are 17,013 unitrans in MGOS to support gene annotations, of which 69% have UniProt annotations. Perhaps most importantly, MGOS provides a platform for the research community to add additional annotation details based on their experimental data, and it provides a forum for community discussions.

A key feature of MGOS is the presence of a user-friendly, community annotation interface that allows registered users to enter any data relevant to genome and gene analyses. The database is designed to serve as a repository for a variety of data, including regulated gene expression, mutant analysis, and alternate gene models that are identified. Data that may not merit a publication, such as analysis of mutants lacking phenotypes, can be entered here with full credit to the scientists involved. The average time required for entering data in MGOS is as follows: 30 to 45 s for a gene symbol, up to 1 min for synonyms, up to 1 min for a gene description, up to 2 min for gene comments, up to 5 min for GO terms (depending upon number of terms and information related to that term), 2 to 3 min for publications, 10 to 15 min for mutants, and up to 10 min for a transcript. Thus, a user can add all this information to a gene page in less than 40 min. Each edit page has instructions at the bottom and there is a downloadable manual describing use of MGOS in order to assist users in adding their information.

The value of MGOS-enabled community annotation to greatly enrich the depth of data for all aspects of *M. oryzae* biology is already demonstrated. So much is now known about pathogenicity and host specificity in the rice blast system that reviews and other literature sources cannot possibly cover all available information. The MGOS gene pages have already pulled together all relevant data for key disease and host specificity genes. For example, the hydrophobin gene *MPG1* (MC_10315) is important for pathogenicity. The MGOS user will find collected information on gene description, chromosomal location, gene structure, and overlapping features (Broad version 5 and 6 transcripts, EST contigs, GenBank proteins, and SAGE tags), GO terms, related publications, mutant information, microarray data, transcript, and protein sequence on the gene page. The avirulence effector gene *AVR-Pita* is a well-studied host specificity gene. Typing "AVR-Pita" into the basic search feature will identify the two *AVR-Pita* family members in strain 70-15 (Khang et al. 2008). These are *AVR-Pita1*, which confers avirulence toward rice with resistance gene *Pita*, and *AVR-Pita3*, which is inactive in conferring avirulence toward rice with *Pita*. The corresponding gene pages compile the extensive information available on these genes. MGOS is the only database providing such detailed information to the user.

Genes encoding secreted proteins are of major interest in pathogen systems because proteins secreted at the host interface play major roles in determining the outcome of the interaction. *M. oryzae* has a predicted secretome of up to 1,546 proteins (Soanes et al. 2008), which is significantly higher than observed for related saprotrophic species such as *Neurospora crassa*. The secretome comprises many gene families; for example, cutinases involved in degrading the plant cuticle and plant cell-wall-degrading enzymes such as xylanases and glucanases. The secretome also includes many unique effector proteins that play roles during biotrophic invasion of rice cells, and sometimes confer avirulence based on recognition by the rice resistance gene-encoded receptors to block disease (Ellis et al. 2009; Valent and Khang 2010; Wilson and Talbot 2009). Microarray data (Mosquera et al. 2009) available in MGOS identifies which secreted protein genes are specifically expressed during biotrophic invasion of rice cells, and defines a set of candidate effector genes known as biotrophy-associated secreted protein (BAS) genes (Search "BAS gene", examples are *BAS1* and *BAS2*). MGOS will soon house microscopic images of hundreds of fluorescently tagged *M. oryzae* secreted proteins, including BAS proteins, illustrating various localization patterns at the fungus–rice interface (M. L. Farman, B. Valent, M. Goodin, and C. Soderlund, *unpublished data*). This will provide yet another major expansion of MGOS utility.

Another good example of MGOS utility is the MC_09245 (*MGA1*) gene, an important gene that contributes to pathogenicity of this fungus. This gene was lost in the updated versions of the genome in the Broad database but has been manually restored in MGOS with information on the history related to this gene.

The information in MGOS is not only useful for the *Magnaporthe* community but also for scientists working with other organisms, including other plant and animal pathogens. For example, ontology is a valuable tool for inferring gene function and for allowing the research community to perform cross-species comparative analyses. MGOS genes are associated with general functions in GO as well as with specialized pathogen

ontologies developed by the PAMGO consortium (Meng et al. 2009). The same structured language is also used in other major fungal databases (Christie et al. 2009). Potential genes involved in fundamental processes (such as cell growth and division, protein synthesis, energy production, and so on) can be identified via these annotations for targeting for mutational analyses. Additionally, orthologous genes and protein sequences can be compared with the MGOS genes defined as playing a role in pathogenicity to identify potential targets for analysis in other pathogen systems.

There are multiple web-based databases that include *M. oryzae* data, with the major ones being the Broad Institute, genomic resources of *M. oryzae* (GROMO) (Thakur et al. 2009), and PHI-base. MGOS was developed in consultation with the Broad Institute to expand the utility of the genome sequence generated there by adding features for analyzing gene and protein structure and function, including cDNA structure, GO terms, and secreted protein predictions; gene expression data from both SAGE and microarray experiments; mutant information; and the linking of literature to annotated genes. MGOS also has an intuitive interface and makes its data available via download. GROMO has incorporated several of the data types that are available from MGOS, and it also predicts biological pathways using the Kyoto Encyclopedia of Genes and Genomes database. In contrast to MGOS, GROMO does not provide a genome interface, literature links to individual genes, or an interactive community annotation feature. PHI-base is a curated database that contains information on genes demonstrated to be important in pathogenicity from a broad set of pathogens of plants, animals, insects, and fungi. It contains 162 *M. grisea* genes (most are, in fact, *M. oryzae* genes) and is an excellent source of genes to consider as orthologous to potential pathogenicity factors in *M. oryzae*; however, it is not a full-service comprehensive genome database like MGOS.

Though the new MGOS, with advanced editing capabilities, has been available for over a year, only four genes have been curated by other than the MGOS team. In this new era of large-scale data analysis, it is extremely important to share and centralize knowledge about an organism. It is no longer sufficient to publish a manuscript on a newly annotated gene; this information should also be added to the community database. The less than 40 min it takes to edit a gene is a small price to pay to share the data and, if everyone does it, everyone wins. This article describes the latest software in hopes that it will encourage the community to use MGOS. With community participation, MGOS will continue to grow as a valuable resource for international rice blast research.

Note: during development of this manuscript, the v8 Assembly and v7 annotation of *M. oryzae* was released by the Broad Institute. Plans are underway through a community funding effort to update MGOS with this data.

Availability and requirements.

For reporting errors in the MGOS database, users can send the errors to www@agcol.arizona.edu. The operating system is platform independent (Web server); programming languages are HTML, Perl, and CGI; licensing is free for academics. Please contact the corresponding authors for more details.

ACKNOWLEDGMENTS

All authors participated in the design of MGOS, acquisition of data, and writing the manuscript. K. Greer performed most of the software development and A. Kour performed most of the manual annotation and coordination of the manuscript. We thank A. Descour for her work on adding the PAVE interface to MGOS and for help with maintenance and R. Terauchi (Iwate Biotechnology Institute, Japan) for providing unpublished mutant data. This work was supported by the National Science Foundation under grant number 0627159. This is contribution number 10-345-J from the Agricultural Experiment Station at Kansas State University.

LITERATURE CITED

Altschul, S. F., Gish, W., Miller, W., Myers E. W., and Lipman, D. J. 1990. Basic local alignment search tool. J. Mol. Biol. 215:403-410.

Bairoch, A., Apweiler, R., Wu, C. H., Barker, W. C., Boeckmann, B., Ferro, S., Gasteiger, E., Huang, H. Z., Lopez, R., Magrane, M., Martin, M. J., Natale, D. A., O'Donovan, C., Redaschi, N., and Yeh, L. S. L. 2005. The universal protein resource (UniProt). Nucleic Acids Res. 94:154-159.

Betts, M. F., Tucker, S. L., Galadima, N., Meng, Y., Patel, G., Li, L., Donofrio, N., Floyd, A., Nolin, S., Brown, D., Mandel, M. A., Mitchell, T. K., Xu, J. R., Dean, R. A., Farman, M. L., and Orbach, M. J. 2007. Development of a high throughput transformation system for insertional mutagenesis in *Magnaporthe oryzae*. Fungal Genet. Biol. 44:1035-1049.

Biswas, M., O'Rourke, J. F., Camon, E., Fraser, G., Kanapin, A., Karavidopoulou, Y., Kersey, P., Kriventseva, E., Mittard, V., and Mulder, N., Phan, I., Servant, F., and Apweiler, R. 2002. Applications of InterPro in protein annotation and genome analysis. Brief. Bioinf. 3:285-295.

Camon, E., Magrane, M., Barrell, D., Lee, V., Dimmer, E., Maslen, J., Binns, D., Harte, N., Lopez, R., and Apweiler, R. 2004. The gene ontology annotation (GOA) database: Sharing knowledge in Uniprot with gene ontology. Nucleic Acids Res. 32:D262-266.

Christie, K. R., Hong, E. L., and Cherry, J. M. 2009. Functional annotations for the *Saccharomyces cerevisiae* genome: The knowns and the known unknowns. Trends Microbiol. 17:286-294.

Dean, R. A., Talbot, N. J., Ebbole, D. J., Farman, M. Q., Mitchell, T. K., Orbach, M., Thon, M., Kulkarni, R., Xu, J.-R., Pan, H., Read, N., Lee, Y.-H., Carbone, I., Brown, S. D., Oh, Y. Y., Donofrio, N., Soanes, D., Djonovic, S., Kolomiets, E., Rehmeyer, C., Li, W. X., Harding, M., Kim, S., Lebrun, M.-H., Bohnert, H., Coughlan, S., Butler, J., Calvo, S., Ma, L.-J., Nicol, R., Purcell, S., Nusbaum, C., Galagan, J. E., and Birren, B. W. 2005. The genome sequence of the rice blast fungus *Magnaporthe grisea*. Nature 434:980-986.

Donofrio, N., Rajagopalon, R., Brown, D., Diener, S., Windham, D., Nolin, S., Floyd, A., Mitchell, T., Galadima, N., Tucker, S., Orbach, M. J., Patel, G., Farman, M., Pampanwar, V., Soderlund, C., Lee, Y. H., and Dean, R. A. 2005. 'PACLIMS': A component LIM system for high-throughput functional genomic analysis. BMC Bioinf. 12:94.

Donofrio, N. M., Oh, Y., Lundy, R., Pan, H., Brown, D. E., Jeong, J. S., Coughlan, S., Mitchell, T. K., and Dean, R. A. 2006. Global gene expression during nitrogen starvation in the rice blast fungus, *Magnaporthe grisea*. Fungal Genet. Biol. 43:605-617.

Ebbole, D. J. 2007. *Magnaporthe* as a model for understanding host-pathogen interactions. Annu. Rev. Phytopathol. 45:437-456.

Ebbole, D. J., Jin, Y., Thon, M., Pan, H. Q., Bhattarai, E., Thomas, T., and Dean, R. 2004. Gene discovery and gene expression in the rice blast fungus, *Magnaporthe grisea*: Analysis of expressed sequence tags. Mol. Plant-Microbe Interact. 17:1337-1347.

Ellis, J. G., Rafiqi, M., Gan, P., Chakrabarti, A., and Dodds, P. N. 2009. Recent progress in discovery and functional analysis of effector proteins of fungal and oomycete plant pathogens. Curr. Opin. Plant Biol. 12:399-405.

Gowda, M., Jantasuriyarat, C., Dean, R. A., and Wang, G. L. 2004. Robust long-SAGE (RL-SAGE): A substantially improved long SAGE method for gene discovery and transcriptome analysis. Plant Physiol. 134:890-897.

Gowda, M., Venu, R.C., Jia, Y., Stahlberg, E., Pampanwar, V., Soderlund, C., and Wang, G. L. 2007. Use of robust-long serial analysis of gene expression to identify novel fungal and plant genes involved in host-pathogen interactions. Methods Mol. Biol. 354:131-144.

Hedeler, C., Wong, H. M., Cornell, M. J., Alam, I., Soanes, D. M.. Rattray, M., Hubbard, S. J., Talbot, N. J., Oliver, S. G., and Paton, N. W. 2007. e-Fungi: A data resource for comparative analysis of fungal genomes. BMC Genomics 8:426. Published online.

Hunter, S., Apweiler, R., Attwood, T. K., Bairoch, A., Bateman, A., Binns, D., Bork, P., Das, U., Daugherty, L., Duquenne, L., Finn, R. D., Gough, J., Haft, D., Hulo, N., Kahn, D., Kelly, E., Laugraud, A., Letunic, I., Lonsdale, D., Lopez, R., Madera, M., Maslen, J., McAnulla, C., McDowall, J., Mistry, J., Mitchell, A., Mulder, N., Natale, D., Orengo, C., Quinn, A. F., Selengut, J. D., Sigrist, C. J. A., Thimma, M., Thomas, P. D., Valentin, F., Wilson, D., Wu, C. H., and Yeats, C. 2009. InterPro: The integrative protein signature database. Nucleic Acids Res. 37:D211-D215.

Jeon, J., Park, S. Y., Chi, M. H., Choi, J., Park, J., Rho, H. S., Kim, S.,

Goh, J., Yoo, S., Choi, J., Park, J. Y., Yi, M., Yang, S., Kwon, M. J., Han, S. S., Kim, B. R., Khang, C. H., Park, B., Lim, S .E., Jung, K., Kong, S., Karunakaran, M., Oh, H. S., Kim, H., Kim, S., Park, J., Kang, S., Choi, W. B., Kang, S., and Lee, Y. H. 2007. Genome-wide functional analysis of pathogenicity genes in the rice blast fungus. Nat. Genet. 39:561-565.

Kent, W .J. 2002. BLAT—The BLAST-like alignment tool. Genome Res. 12:656-664.

Khang, C.-H., Park, S.-Y., Lee, Y.-H., Valent, B., and Kang, S. 2008. Genome organization and evolution of the *AVR-Pita* avirulence gene family in the *Magnaporthe grisea* species complex. Mol. Plant-Microbe Interact. 21:658-670.

Menda, N., Buels, R. M., Tecle, I., and Mueller, L. A. 2008. A community-based annotation framework for linking Solanaceae genomes with phenomes. Plant Physiol. 147:1788-1799.

Meng, S. W., Brown, D. E., Ebbole, D. J., Torto-Alalibo, T., Oh, Y. Y., Deng, J. X., Mitchell, T. K., and Dean, R. A. 2009. Gene ontology annotation of the rice blast fungus, *Magnaporthe oryzae*. BMC Microbiol. 9(Suppl.)1:S8.

Mosquera, G., Giraldo, M. C., Khang, C. H., Coughlan, S., and Valent, B. 2009. Interaction transcriptome analysis identifies *Magnaporthe oryzae* BAS1-4 as biotrophy-associated secreted proteins in rice blast disease. Plant Cell 21:1273-1290.

Mungall, C. J., and Emmert, D. B. 2007. A Chado case study: An ontology-based modular schema for representing genome-associated biological information. Bioinformatics 23:i337-i346.

Numa, H., Nishimura, M., Tanaka, T., Kanamori, H., Yang, C.-C., Matsumoto, T., Nagamura, Y., and Itoh, T. 2009. Genome-wide validation of *Magnaporthe grisea* gene structures based on transcription evidence. FEBS (Fed. Eur. Biochem. Soc.) Lett. 583:797-800.

Oh, Y., Donofrio, N., Pan, H., Coughlan, S., Brown, D. E., Meng, S., Mitchell, T., and Dean, R. A. 2008. Transcriptome analysis reveals new insight into appressorium formation and function in the rice blast fungus *Magnaporthe oryzae*. Genome Biol. 9:R85.

Ou, S. H. 1985. Rice Diseases. CABI Publishing, Wallingford, U.K.

Smedley, D., Haider, S., Ballester, B., Holland, R., London, D., Thorisson, G., and Kasprzyk, A. 2009. BioMart--biological queries made easy. BMC Genomics 10:22.

Soanes, D. M., Alam, I., Cornell, M., Wong, H. M., Hedeler, C., Paton, N. W., Rattray, M., Hubbard, S. J., Oliver, S. G., and Talbot, N. J. 2008. Comparative genome analysis of filamentous fungi reveals gene family expansions associated with fungal pathogenesis. PLoS One 3:e2300.

Soderlund, C., Haller, K., Pampanwar, V., Ebbole, D., Farman, M., Orbach, M. J., Wang, G. L., Wing, R., Xu, J. R., Brown, D., Mitchell, T., and Dean, R. 2006. MGOS: A resource for studying *Magnaporthe grisea* and *Oryza sativa* interactions. Mol. Plant-Microbe Interact. 10:1055-1061.

Soderlund, C., Johnson, E., Bomhoff, M., and Descour, A. 2009. PAVE: Program for assembling and viewing ESTs. BMC Genomics 10:400.

Stein, L. 2001. Genome annotation: From sequence to biology. Nat. Rev. Genet. 2:493-503.

Stein, L. D., Mungall, C., Shu, S. Q., Caudy, M., Mangone, M., Day, A., Nickerson, E., Stajich, J. E., Harris, T. W., Arva, A., and Lewis, S. 2002. The generic genome browser: A building block for a model organism system database. Genome Res. 12:1599-1610.

Stekel, D. J., Git, Y., and Falciani, F. 2000. The comparison of gene expression from multiple cDNA libraries. Genome Res. 10:2055-2061.

Thakur, S., Jha, S., Roy-Barman, S., and Chattoo, B. 2009. Genomic Resources of *Magnaporthe oryzae* (GROMO): A comprehensive and integrated database on rice blast fungus. BMC Genomics 10:316.

Valent, B., and Khang, C. H. 2010. Recent advances in rice blast effector research. Curr. Opin. Plant Biol. 13:434-441.

Wang, G. L., and Valent, B. 2009. Advances in Genetics, Genomics and Control of Rice Blast Disease. Springer Science and Business Media, New York.

Wilson, R. A., and Talbot, N. J. 2009. Under pressure: Investigating the biology of plant infection by *Magnaporthe oryzae*. Nat. Genet. 7:185-195.

Winnenburg, R., Urban, M., Beacham, A., Baldwin, T. K., Holland, S., Lindeberg, M., Hansen, H., Rawlings, C., Hammond-Kosack, K. E., and Köhler, J. 2008. PHI-base update: Additions to the pathogen host interaction database. Nucleic Acids Res. 36:D572-D576.

AUTHOR-RECOMMENDED INTERNET RESOURCES

Agrobacterium tumefaciens-mediated transformation database for *M. oryzae (*ATMT): atmt.snu.ac.kr
Aspergillus genome database: www.aspgd.org
Broad Institute MIT *Magnaporthe* database: www.broad.mit.edu/annotation/genome/magnaporthe_grisea/Home.html
Candida genome database: www.candidagenome.org
e-fungi database: img.cs.man.ac.uk/efungi
Fungal Anatomy Ontology Project website: www.yeastgenome.org/fungi/fungal_anatomy_ontology
Fungal Genetics Stock Center website: www.fgsc.net
MGOS database: www.mgosdb.org
National Center for Biotechnology Information website: www.ncbi.nlm.nih.gov
Pathogen-host interactions (PHI-base) website: www.phi-base.org
SimpleMachines software: www.simplemachines.org