MPMI Vol. 19, No. 10, 2006, pp. 1055–1061. DOI: 10.1094/MPMI-19-1055. © 2006 The American Phytopathological Society

TECHNICAL ADVANCE

MGOS: A Resource for Studying *Magnaporthe grisea* and *Oryza sativa* Interactions

Carol Soderlund,[1] Karl Haller,[1] Vishal Pampanwar,[1] Daniel Ebbole,[2] Mark Farman,[3] Marc J. Orbach,[4] Guo-liang Wang,[5] Rod Wing,[6] Jin-Rong Xu,[7] Doug Brown,[8] Thomas Mitchell,[8] and Ralph Dean[8]

[1]Arizona Genomics Computational Laboratory, Bio5 Institute, University of Arizona, Tucson 85721, U.S.A.; [2]Department of Plant Pathology, Texas A&M University, College Station 77843, U.S.A.; [3]Department of Plant Pathology, University of Kentucky, Lexington 40506, U.S.A.; [4]Division of Plant Pathology and Microbiology, University of Arizona, Tucson, 85721, U.S.A.; [5]Department of Plant Pathology, Ohio State University, Columbus 43210, U.S.A.; [6]Arizona Genomics Institute, Plant Sciences, University of Arizona, Tucson 85721, U.S.A.; [7]Department of Plant Pathology, Purdue University, West Lafayette, IN 47907, U.S.A.; [8]Fungal Genomics Laboratory, Department of Plant Pathology, North Carolina State University, Raleigh 27695, U.S.A.

Submitted 15 March 2006. Accepted 19 May 2006.

The MGOS (*Magnaporthe grisea Oryza sativa*) web-based database contains data from *Oryza sativa* and *Magnaporthe grisea* interaction experiments in which *M. grisea* is the fungal pathogen that causes the rice blast disease. In order to study the interactions, a consortium of fungal and rice geneticists was formed to construct a comprehensive set of experiments that would elucidate information about the gene expression of both rice and *M. grisea* during the infection cycle. These experiments included constructing and sequencing cDNA and robust long–serial analysis gene expression libraries from both host and pathogen during different stages of infection in both resistant and susceptible interactions, generating >50,000 *M. grisea* mutants and applying them to susceptible rice strains to test for pathogenicity, and constructing a dual *O. sativa-M. grisea* microarray. MGOS was developed as a central web-based repository for all the experimental data along with the rice and *M. grisea* genomic sequence. Community-based annotation is available for the *M. grisea* genes to aid in the study of the interactions.

Additional keywords: host-pathogen.

The rice blast disease caused by the fungus *Magnaporthe grisea* is one of the most devastating rice diseases worldwide. Although many major and minor blast-resistance genes have been identified and used in rice-breeding programs to control rice blast, newly developed resistant cultivars are often defeated after only a few years of commercial production due to the emergence of new pathogenic variants. A better understanding of the molecular basis of the defense response to *M. grisea* would provide new opportunities to engineer more durable resistant rice cultivars. Besides being extremely beneficial to arresting this disease in rice, there are multiple significant reasons why *M. grisea* has emerged as a model system for elucidating the molecular mechanism governing plant-fungal interactions. *M. grisea* not only kills rice throughout the world but particular strains attack other major cereals and grasses. The mode of infection is fundamentally similar to many fungal pathogens relying on the formation of a specialized infection cell called the appressorium, which penetrates host epidermal cells. Unlike many phytopathogenic fungi such as mildews and rusts, *M. grisea* can be cultured on defined media so that early stages of the infection process can be studied ex planta. Furthermore, the genomic sequence is available for both rice (The International Rice Genome Sequencing Consortium 2005) and *M. grisea* (Dean et al. 2005), along with many other resources.

We have used high-throughput methods followed by gene-specific experiments to collect information about gene expression in both rice and *M. grisea* during infection. Specific experiments included were as follows. Eight rice cDNA libraries were constructed from different varieties (differing in susceptibility to the disease) and different stages of infection (Jantasuriyarat et al. 2005). Nine *M. grisea* cDNA libraries were constructed from different stages of infection, tissue type, and cultured on different medium (Ebbole et al. 2005). All cDNA libraries were end-sequenced. Five robust long–serial analysis gene expression (RL-SAGE)(Gowda et al. 2004) libraries were constructed and sequenced, four from different stages of rice infection and one from *M. grisea*. A dual *O. sativa-M. grisea* microarray was constructed that contains all predicted *M. grisea* genes and a subset of rice genes that are thought to be involved in resistance. This microarray is available for purchase by the community from Agilent Technologies (Palo Alto, CA, U.S.A.), and all relevant information about the probes is in the MGOS (*Magnaporthe grisea-Oryza sativa*) database. A high-throughput approach to creating *M. grisea* mutants was developed, resulting in over 50,000 lines that have been screened for phenotypes and for changes in pathogenicity when applied to susceptible rice varieties. Secreted proteins were identified from the *M. grisea* genomic sequence, as these proteins are often active during host-pathogen interaction.

The MGOS database was developed to store all experimental data and provide a systemic approach to explore the results. To augment these results, the database also contains genome browsers for both the rice and *M. grisea* genomic sequences. Consensus gene models were calculated for both genomes, and there is a page for each gene that combines all associated results from the different experiments, including a link to the most similar gene in the UniProt database (Bairoch et al. 2005).

Corresponding author: Carol Soderlund; E-mail: cari@agcol.arizona.edu; Fax: +1.520.626.2632.

This is the first host-pathogen database of its kind that contains data from both the host and pathogen, along with experiments specific to their interactions.

MATERIALS AND METHODS

Expressed sequence tag (EST) contigs.

Six rice cDNA libraries were constructed from rice strains with resistant, partially resistant, and susceptible reactions at both 6 and 24 h after inoculation. Two additional libraries were constructed, using uninoculated leaves and leaves from the lesion-mimic mutant *spl11*. The sequenced ESTs were base-called with Phred (Ewing et al. 1998) and were trimmed with Lucy (Li and Chou 2004), resulting in 36,212 5′ and 32,708 3′ successful ESTs. Nine *M. grisea* cDNA libraries were constructed from the appressorium, conidiation and mating-culture stages, the rice cell wall, germinating spores of a mitogen-activated protein kinase mutant (*pmk1*), and from complete, minimal, nitrogen-starved and combined starvation-stressed mycelia. The sequenced ESTs were processed as described by Ebbole and associates (2004), resulting in 24,261 5′ and 54 3′ successful ESTs.

The rice and *M. grisea* EST contigs were generated by PAVE (Program for Assembling and Viewing ESTs), which was developed as a separate module that is being used for other EST projects. The PAVE assembly algorithm clusters the ESTs into transitive closure clusters using BLAST alignment (Altschul et al. 1997), each cluster is assembled into contigs using CAP3 (Huang and Madan 1999). The contigs are split and joined based on auxiliary information, which results in small high-quality contigs with minimal false positive ESTs. The consensus sequences are then compared, and contigs with high identity are assembled together. The auxiliary information is the mate pairs (i.e., 3′ and 5′ reads for a clone), the genomic sequence for both organisms, the results of blasting the ESTs against UniProt, and the rice KOME FL-cDNAs (The Rice Full Length cDNA Consortium 2003). Typically, ESTs are assembled by using a clustering algorithm followed by CAP3 (Huang and Madan 1999); as shown in Table 1, this approach generally leaves mate pairs in different contigs, whereas the PAVE algorithm gets much closer to having 'one contig equal one gene.'

RL-SAGE.

Five libraries were constructed using RL-SAGE, which is an efficient and rapid variation of the 21-bp LongSAGE method (Saha et al. 2002). Four rice libraries were constructed from cv. Nipponbare, two with susceptible reaction at 24 and 96 h after inoculation, one with resistant reaction at 24 h after inoculation, and one control library 24 h after mock inoculation with water. One *M. grisea* library was made from three-day growth on minimal medium. Our SAGE pipeline extracts the tags, applies a set of rules to remove erroneous tags, and then compiles them into a list of unique tags and their number of occurrences (Gowda et al., in press). This resulted in 126,663 unique rice tags, with 98,697 tags having an average Phred score >20, and 51,927 unique *M. grisea* tags, with 40,886 having an average Phred score >20. All tags are then aligned to the EST contigs and genomic sequence. If the tag matches an EST contig or an AGCoL consensus gene, it inherits its annotation. Otherwise, the flanking regions of 1,000 bp upstream and 100 bp downstream of tags are searched against UniProt and, if a match is found, the corresponding annotation is used.

Mutant phenotypes.

We have developed a platform for high-throughput mutagenesis and phenotype characterization, which will eventually elucidate the function of the approximately 11,000 genes in the 38-Mb genome of this fungus. Over 50,000 *M. grisea* insertion lines have been created from strain 70-15, with the vast majority carrying a single random mutation within the genome. The mutants were tested for several phenotype changes that can affect pathogenicity, growth, and metabolism. For example, mutants that fail to form wild-type pigment have the *buf1* or *alb1* mutations, and these strains develop appressoria that are unable to apply enough force to penetrate the epidermal cell wall (Howard and Valent 1996). Some of the mutants were also screened for changes in pathogenicity by applying them to susceptible rice cultivars. The results of all screens were recorded in the PACLIMS tracking database (Donofrio et al. 2005) and were transferred nightly to MGOS. The mutants were shipped to the Fungal Genetics Stock Center (Kansas City, MO, U.S.A.), where they were archived and made available to the public. Gene recovery has been performed on 20 mutants.

Microarrays.

Two versions of a host-pathogen microarray were created by Agilent. The first version has 7,144 rice probes and 13,666 *M. grisea* probes. The second version has 6,325 rice probes and 15,170 *M. grisea* probes. The rice probes are mainly from EST contigs and KOME FL-cDNAs that are thought to be involved in defense against rice blast. The *M. grisea* probes are mainly from predicted transcripts. To date, there is one released experiment using the first version of the array, titled "Germination and appressorium formation in *Magnaporthe grisea*," which we downloaded from the Gene Expression Omnibus (Barrett et al. 2005) and entered into MGOS.

Secreted proteins.

The outcome of the interaction between *M. grisea* and its host is determined by an exchange of signals. By virtue of their being delivered outside of the fungus, secreted proteins may serve as signals and be recognized by receptors in plants. A functional analysis of secreted proteins is needed to investigate the potential of these proteins to evoke plant responses. Hence, the ESTs and genomic sequence for *M. grisea* were mined for putatively secreted proteins using SignalP (Bendtsen et al. 2004) to identify those containing likely signal sequence

Table 1. Comparisons of methods for expressed sequence tag (EST) assembly of the rice library OSJNEb

Method	Contigs	Singletons	Mate splits[a]	Similar[b]	Suspect[c]
Cluster/CAP	2,897	1,262	1,567	173	15
PAVE[d]	2,401	1,532	1,177	15	9
PAVE+	3,117	618	8,97	3	1

[a] Number of clones that have mate pairs in different contigs.
[b] Number of consensus sequences that are 95% similar over 95% of one of the sequences.
[c] Contains multiple subcontigs with 3′ and 5′ reads.
[d] PAVE (Program for Assembling and Viewing ESTs) uses 3′, 5′, and UniProt information. In addition, PAVE+ uses the KOME FL-cDNAs and genomic sequence.

cleavage sites and PROTCOMP, available from Softberry, to predict cellular localization of the putative proteins, which resulted in 739 proteins.

Sequence.

Though the genomic sequence does not provide information on the interactions between the two organisms, it is very important for the global view. MGOS contains the rice genomic sequence, which was sequenced using the clone-by-clone approach (The International Rice Genome Sequencing Consortium 2005), and the *M. grisea* genomic sequence, which was sequenced using a whole-genome shotgun approach (Dean et al. 2005). MGOS also contains the rice KOME FL-cDNA, the *Oryza sativa* UniProt proteins that are not annotated as hypothetical (Bairoch et al. 2005), and the GenBank *M. grisea* proteins (Bendtsen et al. 2004).

MGOS database.

From the MGOS home page, there are links to the main pages for EST PAVE CONTIGS, RL-SAGE, Microarrays, Mutants, Gene List, Genome Browser, Search, Links & Publications, and SiteMap. The main pages for the data all have a simple search, Advance Search, and Data Description page with details.

Experiment pages.

The ESTs, RL-SAGE, mutants, and microarrays are the salient experimental types in MGOS, and each has its own set of search pages as described in this section.

The purpose of making EST and SAGE libraries at different timepoints is to determine which genes are expressed at one set of timepoints that are not expressed in another set. To provide this ability, the ESTs from all libraries have been assembled together, and the SAGE tags from all libraries have been merged into one set of tags; in both cases, the library information is retained. In this way, queries can be created to ask what contigs have ESTs from one set of libraries and do not have ESTs from another set; this interface was modeled after HarvEST (Zheng et al. 2004). The interface has been extended to allow the user to filter the search on any of the test statistics provided by the IDEG6 package (Romualdi et al. 2003), Gene Ontology (Camon et al. 2004), UniProt annotation, and other characteristics, such as whether it has a match to its respective genome. Any combination of these filters can be applied, as shown in Figure 1. The results are shown in tabular form with links to the individual pages for the EST contigs and SAGE tags. The display of EST contigs is a line drawing that shows mismatches and low-quality regions and can be changed to show the base alignment.

Associated with each probe in an Agilent microarray is the original sequence and its annotation. The sequences from which the probes were derived were compared with all AGCoL consensus genes for the best match. The probes can be searched by Agilent Probe ID, gene name, or gene description. The results link to the probe and gene pages of the best match. This allows any researcher that has experimental results based on the Agilent rice and rice blast array to see all corresponding information for probes of interest. The Browse Experiment link displays a table of all experiments in MGOS, which in turn link to details of the experiment.

The mutant interface provides the user with the ability to query on any combination of the measured phenotypes in MGOS, which shows the results in a table with the Fungal Genetics Stock Center ID. The mutants do not directly link to any of the other data in MGOS, but the gene-recovery mutants have locations on the genome, and hence, link the corresponding mutant to a gene record in MGOS.

AGCoL consensus genes and gene list.

All EST contigs, RL-SAGE tags, rice FL-cDNAs, UniProt rice proteins, and GenBank *M. grisea* proteins were aligned to the genome. The RL-SAGE tags were aligned to the genome using MegaBLAST (Zhang et al. 2000) and required a 100% match. Other nucleotide sequences were aligned using BLAT (Kent 2002) and were filtered using a ratio of match length to query length of greater than 95%. The UniProt rice proteins and GenBank *M. grisea* proteins were aligned using TBLASTN, and high-scoring pairs (HSP) were filtered on an *e*-value of 1e-10 and percent identity of >90%. A perl program then read the filtered HSP and calculated all possible lists of consecutive HSP, allowing an overlap of 3 bp and intron-gap size of 2,000 bp between consecutive HSP, based on the match to the genomic region; the longest list of HSP was used as the alignment.

If any of the nucleotide or protein sequences did not align to the genome, they were assigned an AGCoL gene name and were added to the list of genes. The exception was the RL-SAGE tags; if the tag did not align to an EST contig or the sequence, it was not represented in a Gene page. Table 2 shows the number of each type of sequence and how many of each do not align to their respective genome. As shown in Figure 2, the rest of the AGCoL genes were calculated from alignments and gene predictions on the genomic sequence. The rice genes required at least one of the following forms of evidence: KOME rice FL-cDNA, PAVE rice EST contig, UniProt protein or fgeneh (Solovyev and Salamov 1997) prediction that also has a tentative consensus (Lee et al. 2005) or UniProt match. The *M. grisea* genes were assigned based on at least one of the following pieces of evidence: BROAD gene model predictions V4, FGL (Fungal Genomic Laboratory, North Carolina State University) gene model predictions, PAVE *M. grisea* EST contigs, or GenBank protein.

As shown in Figure 3, the Gene page is the heart of MGOS, as it links together data from the different experiments and sequences. A page exists for each AGCoL gene, with links to the detailed data pages. All genes have been searched against UniProt, using BLAST at a 1e-20 cutoff to find the most similar sequence; the corresponding description and GO description is shown on the Gene page, along with a link to the UniProt database for more information. All the synonyms for the UniProt rice proteins and for the GenBank *M. grisea* proteins are known in MGOS; hence, the user can search MGOS using the name they are familiar with and find the associated data. The Gene page also contains the secreted-protein information for the 739 identified *M. grisea* genes and all evidence types shown in Figure 2.

The Gene List page allows the user to search the genes based on any synonym, and the Advanced Search page allows the user to search on different attributes and evidence.

Community *M. grisea* gene annotation.

Rice community annotation is available from The Institute for Genomic Research (TIGR), Gramene, and the International Rice Genome Sequencing Project. These annotations were not available when MGOS was initially designed but can be added to the list of evidence (Fig. 2). No such annotation is available for *M. grisea*; hence, we have recently opened up MGOS for *M. grisea* community annotation. Researchers can add their annotation through the Data Submission link on the Gene List page. It will be reviewed and all accepted annotation will be listed on the Browse Submitted Annotation page and appended to the appropriate Gene page.

Genome browser and search pages.

The rice and *M. grisea* genomic sequences are displayed using the GMOD genome browser (Stein et al. 2002), which

uses a CGI/Perl based interface. We have developed a Java-based interface that is compatible with the MySQL database used by the GMOD browser. The user can select the type of browser based on their preference. There are tracks for all the evidence data (Fig. 2) and for the SAGE tags. All track entities link to their detail page.

As with most genomic databases, there is a Search page so that a user-supplied sequence can be searched against the se-

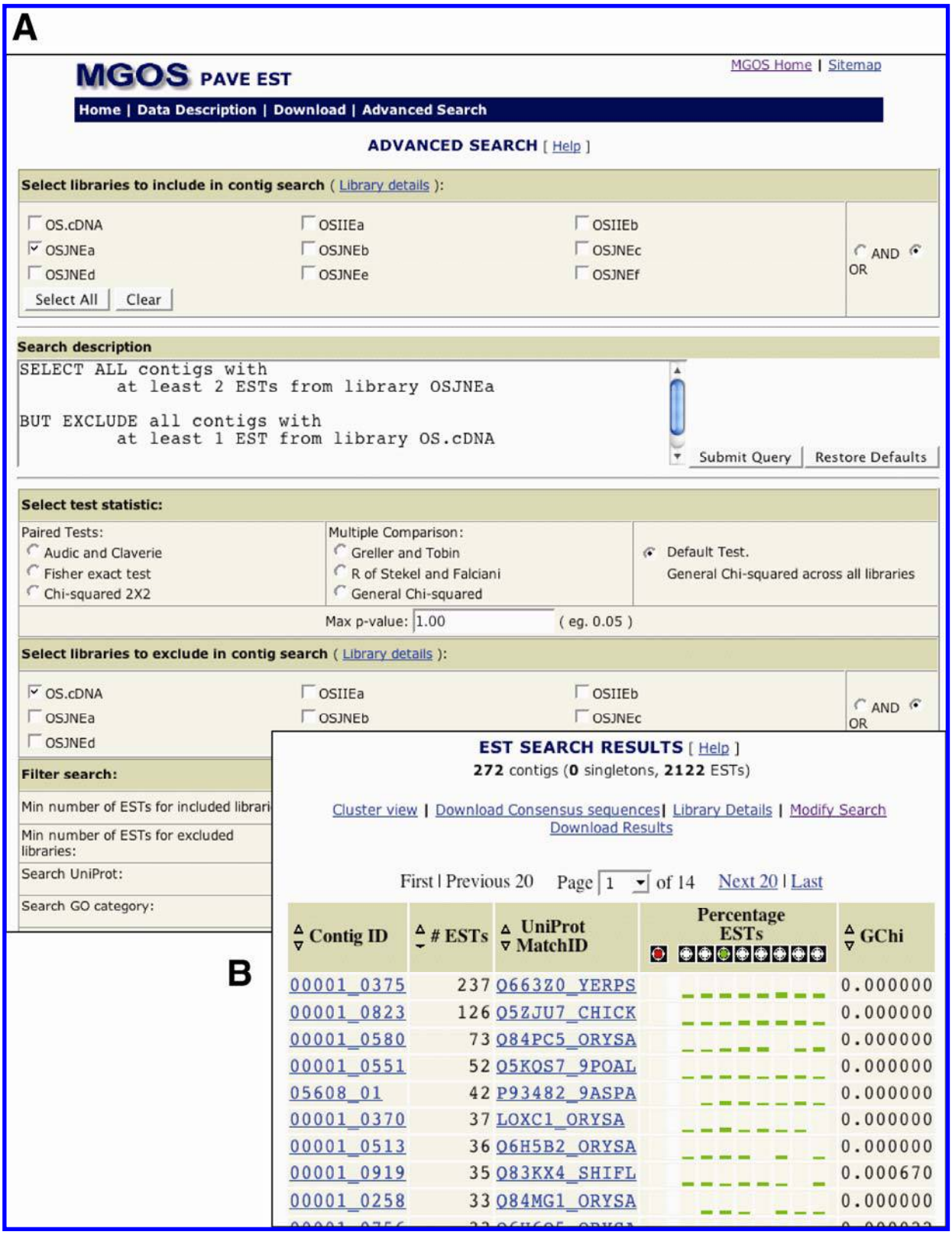

Fig. 1. The expressed sequence tag (EST) advanced search page. The ESTs from all libraries have been assembled together. **A,** The user can request to view all contigs that have *n* or more ESTs from one set of libraries and *m* or fewer ESTs from another set of libraries. Other options can also be selected. The resulting search to be performed is shown in the Search Description box. Though not shown in this figure, the bottom of the Advanced Search page allows the user to select what columns they want to view. **B,** The results are shown in table form, in which the expression level of each library is shown.

quence data in MGOS. Also, the EST Contig and Gene pages both link to the Search page with their respective sequence as the query to be searched against one of several available protein databases. The search can use BLAST (Altschul et al. 1997), FASTA (Pearson et al. 1994), or BLAT (Kent 2002).

Implementation.

MGOS uses MySQL databases with a CGI/Perl interface. PAVE and the Genome Browser are separate MySQL databases for modularity, though they work seamlessly with the main MGOS database. All query pages have the same layout, in which the user selects what options to filter on and the search description text box describes in sentence form what the query will be (for example, Fig. 1). The user can also select what columns to display in the resulting table. A CGI program developed for this project, called the tableizer, provides a consistent view for all search results tables in MGOS. From the table entries, the user can link to the corresponding data page, which cross-links to all other associated pages. The data from any table can be downloaded.

Table 2. Sequences that do not align to the rice genomic sequence (International Rice Genome Sequencing Project V4) and *Magnaporthe grisea* genomic sequence (Broad V5)

Data	Total	No alignment	Percent
Rice			
PAVE CCS[a]	15,338	3,421	22
KOME FL-cDNA	32,127	5,129	15
RL-SAGE tags[b]	34,749	14,289	41
UniProt rice proteins	27,537	2,431	9
M. grisea			
PAVE CCS	7,792	1,197	15
RL-SAGE tags[b]	16,367	5,607	34
GenBank Mg proteins	513	1,17	23

[a] Program for Assembling and Viewing ESTs Contig Consensus Sequence.
[b] Only tags with average Phred quality ≥20 and at least two occurrences. The tags required a 100% match to the genome, eliminating the possibility of a hit from a tag which spans an intron.

DISCUSSION

The creation of a single site for integration of diverse data types is essential to allow biologists to visualize the complexity of gene structure, gene organization, and gene interactions that exist within any single organism. Here, we are developing tools that will also promote visualization of the interaction of genes between organisms. Thus, MGOS was created to integrate sequence information with the experimental results of functional analyses and to serve as a gateway to biological materials created by functional genomics. The philosophy of MGOS has been to provide the experimental biologists with an interface that facilitates asking questions covering different types of experiments and from different timepoints. To this end, MGOS has advanced searches across a given experiment

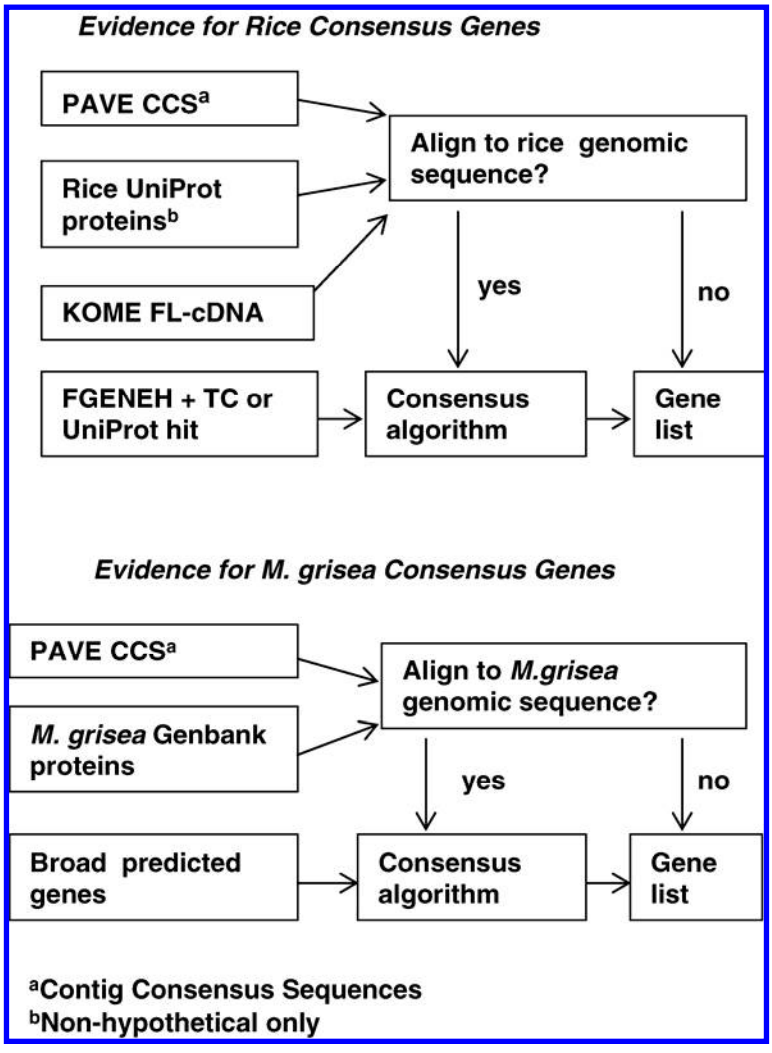

Fig. 2. AGCoL consensus gene evidences flowchart.

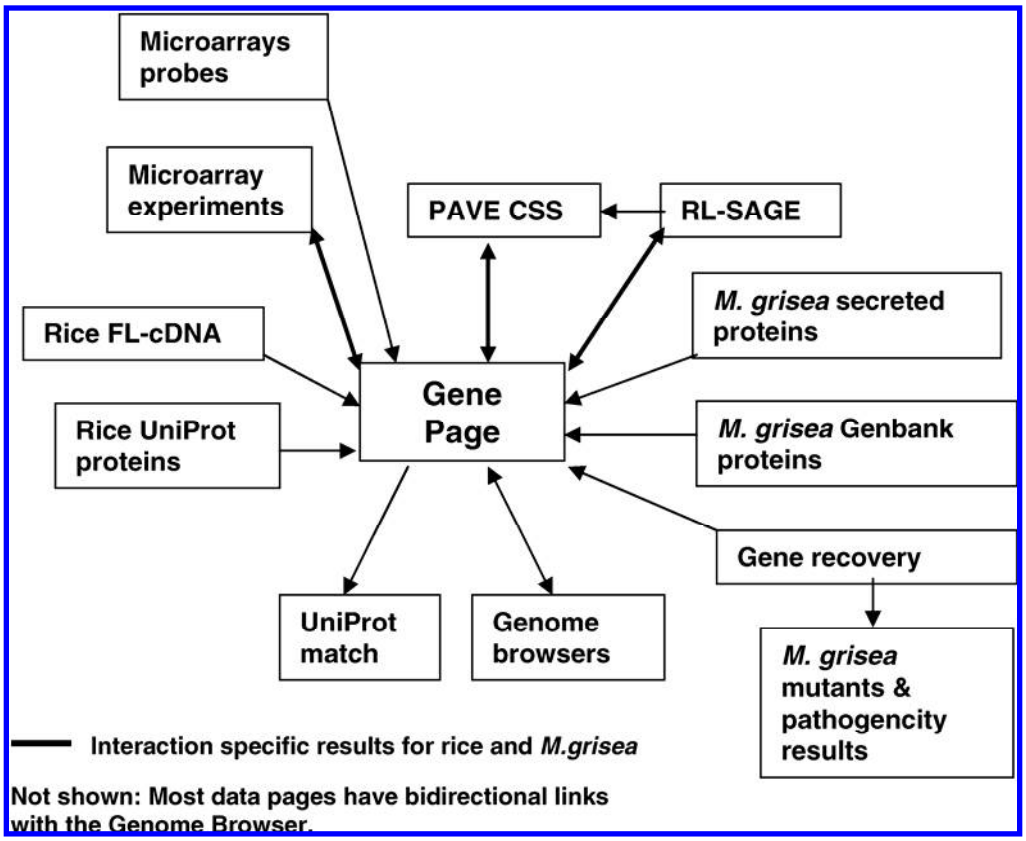

Fig. 3. The MGOS data types and relations.

type and across all experiments via the Gene List search page. Furthermore, the MGOS database makes the data easily available to researchers around the world in both the plant and the pathogen communities. In the future, MGOS is intended to become a community database for the addition of all relevant rice and rice blast interaction data, with emphasis on enhanced community annotation, microarray experiments, and extended queries.

ACKNOWLEDGMENTS

This work was supported by National Science Foundation (NSF) number 0115642. MGOS is funded in part by NSF grant number 0115642. The personnel that contributed to the data generation are listed on the participant's page of MGOS. The *M. grisea* ESTs was supported by the Initiative for Future Agriculture and Food Systems grant number 00-52100-9682 from the United Stated Department of Agriculture Cooperative State Research, Education, and Extension Service (USDA-IFAFS). PAVE is funded by NSF grant numbers 501857 and 0115642. The Java genome browser interface is funded by USDA-IFAFS grant number 2001-52100-11292, NSF grant numbers 0321724 and 0115642. We would like to thank J. Hatfield for his initial work on PAVE, R. Carey and B. Vollmer for their work on the Java genome browser, and L. Boyd for excellent system and database administration.

LITERATURE CITED

Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D. J. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 25:3389-340.

Bairoch, A., Apweiler, R., Wu, C. H., Barker, W. C., Boeckmann, B., Ferro, S., Gasteiger, E., Huang, H., Lopez, R., Magrane, M., Martin, M. J., Natale, D. A., O'Donovan, C., Redaschi, N., and Yeh, L. S. 2005. The Universal Protein Resource (UniProt). Nucleic Acids Res. 33:D154-159.

Barrett, T., Suzek, T. O., Troup, D. B., Wilhite, S. E., Ngau, W. C., Ledoux, P., Rudnev, D., Lash, A. E., Fujibuchi, W., and Edgar, R. 2005. NCBI GEO: Mining millions of expression profiles—Database and tools. Nucleic Acids Res. 33:D562-566.

Bendtsen, J. D., Nielsen, H., von Heijne, G., and Brunak S. 2004. Improved prediction of signal peptides: SignalP 3.0. J. Mol. Biol. 340:783-795.

Benson, D. A., Karsch-Mizrachi, I., Lipman, D. J., Ostell, J., and Wheeler, D. L. 2004. GenBank: Update. Database issue. Nucleic Acids Res. 32:D23-26.

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. Database issue. Nucleic Acids Res. 32:D262-266.

Dean, R. A., Talbot, N. J., Ebbole, D. J., Farman, M. L., Mitchell, T. K., Orbach, M. J., Thon, M., Kulkarni, R., Xu, J. R., Pan, H., Read, N. D., Lee, Y. H., Carbone, I., Brown, D., Oh, Y. Y., Donofrio, N., Jeong, J. S., Soanes, D. M., Djonovic, S., Kolomiets, E., Rehmeyer, C., Li, W., 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 Bioinformatics 6:94.

Ebbole, D. J., Jin,Y., Thon, M., Pan, H., Bhattari, 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.

Ewing, B., Hillier, L., Wendl, M. C., and Green, P. 1998. Base-calling of automated sequencer traces using Phred. I. Accuracy assessment. Genome Res. 8:175-185.

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

Gowda, M., Reddyvarichannarayappa, V., Jia, Y., Stahlberg, E., Pampanwar, V., Soderlund, C., and Wang, G. L. 2006. Use of robust-long serial analysis of gene expression to identify novel fungal and plant genes involved in host-pathogen interactions. Pages 131-144 in: Methods in Molecular Biology: Plant-Pathogen Interactions, P. C. Ronald, ed. Humana Press, Totowa, NJ, U.S.A.

Howard, R. J., and Valent, B. 1996. BREAKING AND ENTERING: Host penetration by the fungal rice blast pathogen *Magnaporthe grisea*. Annu. Rev. Microbiol. 50:491-512.
Huang, X., and Madan, A. 1999. CAP3: A DNA sequence assembly program. Genome Res. 9:868-877.
The International Rice Genome Sequencing Consortium 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., Dean, R. A., Wing, R. A., Soderlund, C., and Wang, G. L. 2005. Large-scale identification of expressed sequence tags involved in rice and rice blast fungus interaction. Plant Physiol. 138:105-115.
Kent, W. J. 2002. BLAT—The BLAST-like alignment tool. Genome Res. 12:656-664.
Lee, Y., Tsai, J., Sunkara, S., Karamycheva, S., Pertea, G., Sultana, R., Antonescu, V., Chan, A., Cheung, F., Quackenbush, J. 2005. The TIGR gene indices: Clustering and assembling EST and known genes and integration with eukaryotic genomes. Nucleic Acids Res. 33:D71-74.
Li, S., and Chou, H. H. 2004. LUCY2: An interactive DNA sequence quality trimming and vector removal tool. Bioinformatics 20:2865-2866.
Pearson, W. R. 1994. Using the FASTA program to search protein and DNA sequence databases. Methods Mol. Biol. 24:307-331.
The Rice Full Length cDNA Consortium 2003. Collection, mapping, and annotation of over 28,000 cDNA clones from japonica rice. Science 301:376-379.
Romualdi, C., Bortoluzzi, S., D'Alessi, F., and Danieli, G. A. 2003. IDEG6: A web tool for detection of differentially expressed genes in multiple tag sampling experiments. Physiol. Genomics 12:159-162.
Saha, S., Sparks, A. B., Rago, C., Akmaev, V., Wang, C. J., Vogelstein, B., Kinzler, K. W., Velculescu, V. E. 2002. Using the transcriptome to annotate the genome. Nat. Biotechnol. 20:508-512.
Solovyev, V., and Salamov, A. 1997. The Gene-Finder computer tools for analysis of human and model organisms genome sequences. Proc. Int. Conf. Intell. Syst. Mol. Biol. 5:294-302.
Stein, L. D., Mungall, C., Shu, S., 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.
Zhang, Z., Schwartz, S., Wagner, L. Miller, W. 2000. A greedy algorithm for aligning DNA sequences. J. Computational Biol. 7:203-214.
Zheng, J., Close, T.J., Jiang, T., and Lonardi, S. 2004. Efficient selection of unique and popular oligos for large EST databases. Bioinformatics 20:2101-2112.

AUTHOR-RECOMMENDED INTERNET RESOURCES

The Broad Institute *Magnaporthe grisea* database: www.broad.mit.edu/annotation/fungi/magnaporthe
Fungal Genetics Stock Center: www.fgsc.net
Gramene, a resource for comparative grass genomics: www.gramene.org
The International Rice Genome Sequencing Project: rgp.dna.affrc.go.jp/IRGSP
The MGOS database: www.mgosdb.org
The PROTCOMP software: www.softberry.com -.
TIGR Rice Genome Annotation database: www.tigr.org/tdb/e2k1/osa1
The UniProt (Universal Protein Resource) database: www.uniprot.org
University of Arizona, Arizona Computational Laboratory, PAVE databases and Java Genome Browser software: www.agcol.arizona.edu