MPMI Vol. 15, No. 5, 2002, pp. 421–427. Publication no. M-2002-0409-01O.

Review

Genomics of Phytopathogenic Fungi and the Development of Bioinformatic Resources

Darren M. Soanes,[1] Wendy Skinner,[2] John Keon,[2] John Hargreaves,[2] and Nicholas J. Talbot[1]

[1]School of Biological Sciences, University of Exeter, Washington Singer Laboratories, Perry Road, Exeter, EX4 4QG, U.K.; [2]Institute of Arable Crops Research, Long Ashton Research Station, Long Ashton, Bristol, U.K.

Submitted 19 December 2001. Accepted 18 January 2002.

Genomic resources available to researchers studying phytopathogenic fungi are limited. Here, we briefly review the genomic and bioinformatic resources available and the current status of fungal genomics. We also describe a relational database containing sequences of expressed sequence tags (ESTs) from three phytopathogenic fungi, *Blumeria graminis*, *Magnaporthe grisea*, and *Mycosphaerella graminicola*, and the methods and underlying principles required for its construction. The database contains significant annotation for each EST sequence and is accessible at http://cogeme.ex.ac.uk. An easy-to-use interface allows the user to identify gene sequences by using simple text queries or homology searches. New querying functions and large sequence sets from a variety of phytopathogenic species will be incorporated in due course.

The application of molecular genetic analysis to the study of phytopathogenic fungi has led to the identification and characterization of a diverse collection of genes involved in fungal pathogenicity (Idnurm and Howlett 2001; Knogge 1998; Sweigard et al. 1998). These include genes involved in detoxification of antifungal compounds produced by plants (Bowyer et al. 1995; Straney and VanEtten 1994), biosynthesis of phytotoxic compounds (Panaccione et al. 1992), breakdown of the host plant cuticle (Tonukari et al. 2000; Walton 1994), conidiogenesis (Hamer and Givan 1990), appressorium formation and function (Balhadère and Talbot 2001; Clergeot et al. 2001; Silué et al. 1998; Talbot et al. 1993), and amino acid metabolism (Balhadère et al. 1999), as well as those involved in conserved signaling pathways (Kronstad 1997; Xu and Hamer 1996). The capacity to cause plant disease is, however, clearly a complex phenotype, and phytopathogenic fungi exhibit considerable diversity both in their developmental biology and in the types of disease symptoms they induce (Agrios 1988; Bowyer 1999). While molecular biology has allowed investigation of the genetic control of pathogenic development in a limited number of experimentally tractable pathogens—most notably *Ustilago maydis* and *Magnaporthe grisea*—it is apparent that the current state of knowledge of the biology of fungal pathogens is relatively superficial.

The advent of genome-wide analysis promises to provide a new and powerful means of investigating diverse fungal pathogens. For the first time, it will be possible, for example, to define all the genetic components that are expressed during spore germination, infection-related development, and plant tissue invasion, as well as those expressed under the control of the key signal-transduction pathways that are required for pathogenesis (Idnurm and Howlett 2001; Talbot and Foster 2001). Identifying the gene inventories and genomic organization of diverse fungi similarly offers the opportunity to test hypotheses regarding mechanistic distinctions among phytopathogenic species and their evolutionary relationships. With these tremendous opportunities to further our understanding of fungal pathogens, however, comes a series of fresh challenges. The first of these is likely to be the development of specific bioinformatic resources designed to house and investigate the huge amount of genomic information that is likely to be generated from phytopathogenic fungi. Here, we briefly review the progress of genome-wide studies of pathogenic fungi and the available informatic resources that have been produced so far. We also describe the generation of a new relational database, which houses a significant proportion of the currently available expressed sequence tag (EST) data from phytopathogenic fungi, and review the methods by which it was constructed and the underlying principles upon which it was designed.

Corresponding author: N. J. Talbot; E-mail: N.J.Talbot@exeter.ac.uk

The Current Status of Fungal Genomics

The application of genomic approaches to study fungal pathogens has so far lagged behind that of many other organisms, largely because, to date, no pathogenic fungus has had its genome fully sequenced in a public-sector laboratory. In fact, only one fungal genome sequence, that of *Saccharomyces cerevisiae*, is fully available in the public domain (Goffeau et al. 1996); although two others, those of the fission yeast *Schizosaccharomyces pombe* (http://www.sanger.ac.uk/Projects/S_pombe/) and the filamentous fungus *Neurospora crassa* have recently been completed (http://www-genome.wi.mit.edu/annotation/fungi/neurospora). The utility of the yeast genome has been aided enormously by specifically designed databases, such as the yeast protein database (http://www.incyte.com/sequence/proteome/index.shtml) and the Stanford genome database (http://genome-www.stanford.edu/Saccharomyces/), which offer a means of interrogating and retrieving specific gene information from among the 6,200 open reading frames that have been identified so far in the *S. cerevisiae* genome sequence. These databases incorporate information regarding mutant phenotypes and transcriptional profile analysis experiments that monitor the expression of the genome under different conditions. Yeast microarray analysis has been used, for example, to assay changes in levels of transcripts caused by heat shock, cold shock, the switch from galactose to glucose as principal carbon source (Lashkari et al. 1997), the switch from fermentation to respiration (DeRisi et al. 1997), diauxic growth shift from rich nutrient conditions to

starvation stress (Wodicka et al. 1997), cell-cycle progression (Cho et al. 1998), sporulation (Chu et al. 1998), and gene expression in a number of regulatory mutant strains (DeRisi et al. 1997). Similarly, serial analysis of gene expression (SAGE) is a very robust technology (Yamamoto et al. 2001) that has been widely utilized to investigate global patterns of gene expression in yeast (Kal et al. 1999; Velculescu et al. 1997). SAGE frequency tables can provide a measure of the abundance of a given gene transcript at a specific point during the growth of yeast cells (Jansen and Gerstein 2000) and have been used to explore the relationship between transcript abundance and protein accumulation (Gygi et al. 1999), which will be vital in comparing transcriptional profiling and proteomic data in the future. Cluster analysis of genome-wide expression data from microarray and SAGE experiments has also enabled genes showing similar patterns of expression to be grouped together. Clustering of gene expression data in yeast and in humans, for example, has been used to classify genes of similar function (Eisen et al. 1998). The collation of all of this information into central data repositories has been invaluable in maximizing their utility to the yeast community.

The next generation of object relational databases to incorporate analysis of the yeast proteome and metabolome, in addition to the systematic gene disruption of the large majority of yeast genes, is already being designed and implemented. A clear example of a next-generation bioinformatic resource is the Genome Information Management System (GIMS) developed by the University of Manchester (Manchester, U.K.) (http://www.cs.man.ac.uk/~norm/gims), which provides an environment in which researchers can carry out more than 50 different analysis tasks utilizing available genomic information. The environment also allows the use of elaborate querying facilities including long, text-based queries (Paton et al. 2000). The opportunity to ask more elaborate questions of archived genomic information by using better pattern-identification software should allow transcriptional profiling and proteomic analysis to be far more useful in revealing the orchestrated action of complex gene sets than by using a simple, manual examination of the information.

The contrast between bioinformatic resources developed for yeast researchers and those currently available for the phytopathogenic fungi research community could not be greater, but these databases do provide a clear example of what can be achieved in a very short period of time, given access to genomic information. Fungal genome information is, therefore, urgently required, and the relative lack of public-sector funding for fungal genomics has been the source of considerable discussion (Pennisi 2001). In a recent review, for example, out of 379 genome project websites surveyed, only 16 (4%) involved fungi, of which only seven involved pathogenic species (Yoder and Turgeon 2001). Recent developments, however, including the announcement of the Fungal Genomes Initiative by the Whitehead Institute (Cambridge, MA, U.S.A.) made at the 21st Fungal Genetics Conference (Pacific Grove, CA, U.S.A.; March 2001), indicate that the sequences of a number of phytopathogenic fungi will become available in the foreseeable future. Under this new initiative, up to 15 fungi will be selected for genome sequencing, using the same whole-genome shotgun approach developed for sequencing *N. crassa*. This wealth of new genomic information will certainly revolutionize the manner in which phytopathogenic fungi are studied, and it can only be hoped that the initiative will soon receive the financial support required. Similar, albeit smaller scale, initiatives in Europe to sequence genomes from the powdery mildew fungus *Blumeria graminis* and the end-rot pathogen *Botrytis cinerea* will also contribute to the wealth of new information likely to be released in the next 3 to 5 years. The recent observation of areas of synteny between sequenced regions of the genomes of the fungal pathogen *M. grisea* and the saprotroph *N. crassa*, in contrast to the limited similarity between some of the yeasts such as *S. cerevisiae* and *Candida albicans* (Seoighe et al. 2000), illustrates the likely surprises in store from analysis of genome sequences (Hamer et al. 2001).

Current fungal genome projects in progress and the corresponding bioinformatic resources available are listed in Table 1. Here, we have only included reference to sites in which sequence data can be accessed and retrieved using either text or sequence-similarity searches. Genome projects are in progress for two human-pathogenic fungi; the pulmonary pathogen *Aspergillus fumi-*

Table 1. Publicly accessible pathogenic fungal and oomycete genomic resources

Fungus	Center[a]	URL	Status January 2002
Aspergillus fumigatus	SC	http://www.sanger.ac.uk/Projects/A_fumigatus/	5,548 BAC[b] end sequences
	TIGR	http://tigrblast.tigr.org/ufmg	
Blumeria graminis	UE	http://cogeme.ex.ac.uk	2,701 unisequences
Botrytis cinerea	GS	http://www.genoscope.cns.fr/externe/English/Projets/Projet_W/W.html	6,558 ESTs[b]
Candida albicans	SGTC	http://sequence-www.stanford.edu/group/candida/	1.5-fold genome coverage
	SC	http://www.sanger.ac.uk/Projects/C_albicans/	10 cosmids
Cryptococcus neoformans	OU	http://www.genome.ou.edu/cneo.html	4,000 ESTs
	SGTC	http://baggage.stanford.edu/group/C.neoformans/menu.html	156 genomic DNA contigs, 18.1 Mb
	TIGR	http://www.tigr.org/tdb/e2k1/cna1/index.shtml	Nominal sixfold genome coverage
	DU	http://cgt.genetics.duke.edu/data/index.html	4,867 BAC end sequences
	GSC	http://rcweb.bcgsc.bc.ca/cgi-bin/cryptococcus/cn.pl	
	UBC	http://www.bcgsc.ca	
Fusarium sporotrichioides	TAMU	http://www.genome.ou.edu/fsporo.html	5,000 ESTs
Magnaporthe grisea	FGL	http://www.fungalgenomics.ncsu.edu	5,000 ESTs, 188 BAC contigs
	UE	http://cogeme.ex.ac.uk	1,839 unisequences
Mycosphaerella graminicola	UE	http://cogeme.ex.ac.uk	2,926 unisequences
Phanerochaete chrysosporium	JGI	http://www.jgi.doe.gov/programs/whiterot.htm	30,000-kb genomic sequence
Phytophthora infestans	NCGR	http://www.ncgr.org/pgc/index.html	3,000 unisequences
Phytophthora sojae			
Pneumocystis carinii	UK	http://biology.uky.edu/Pc/	3,896 ESTs
	SC	http://www.sanger.ac.uk/Projects/P_carinii/	1 telomeric cosmid

[a] DU = Duke University, FGL = Fungal Genomics Laboratory, GS = Genoscope, GSC = Genome Sequence Centre, JGI = DOE Joint Genome Initiative, NCGR = National Centre for Genomics Research, OU = Oklahoma University, SC = Sanger Centre, SGTC = Stanford Genome Technology Centre, TAMU = Texas A&M University, TIGR = The Institute for Genomic Research, UBC = University of British Columbia, UE = University of Exeter, and UK = University of Kentucky.

[b] BAC = bacterial artificial chromosome; ESTs = expressed sequence tags.

gatus at the Sanger Centre (Cambridge, U.K.) and The Institute for Genomics Research (TIGR; Rockville, MD, U.S.A.) and the meningitis-causing basidiomycete fungus *Cryptococcus neoformans* at TIGR and also by a consortium of laboratories at Duke University (Durham, NC, U.S.A.), University of Oklahoma (Norman, OK, U.S.A.), and Stanford University (Stanford, CA, U.S.A.) and the University of British Columbia (Vancouver, BC, Canada). Plans are also in progress to sequence *Histoplasma capsulatum* at the Washington University Genome Sequencing Center (St. Louis, MO, U.S.A.). For phytopathogenic and related species, the sequencing projects in progress are partial-genome sequence initiatives or (predominantly) EST projects (Table 1). Of particular significance among current projects is the U.S. Department of Energy-funded Joint Genome Institute (Walnut Creek, CA, U.S.A.), which has sequenced the genome of the white wood rot fungus *Phanerochaete chrysosporium*. The genome sequence can be searched at a website using BLAST algorithms and is the first basidiomycete fungus to be sequenced in the public domain, although the genome sequence of the corn smut fungus *U. maydis* is held in two private databases (belonging to Bayer Ag [Leverkusen, Germany] and Exelixis [San Francisco, CA, U.S.A.]). The *Phytophthora* Genome Consortium (Davis, CA, U.S.A.) has sequenced ESTs from the oomycete pathogens *Phytophthora infestans* and *Phytophthora sojae*. Their website allows searches using BLAST algorithms and simple text queries and represents an important resource for investigating oomycete biology.

Development of a phytopathogenic fungal EST database

To fully sequence a fungal genome is obviously beyond the scope of most fungal research groups. Single-pass, partial sequencing of either 3′ or 5′ ends of complementary DNA (cDNA) clones to generate a set of ESTs, however, offers a low-cost strategy to identify substantial gene inventories in the absence of gene information, although it is constrained by the developmental stage at which mRNA is extracted. EST production has been used to search for genes in a wide variety of organisms but has only recently been applied to a small number of phytopathogenic fungi (Keon et al. 2000; Rauyaree et al. 2001; Thomas et al. 2001). Inventories of EST data are available in the public domain from a small number of fungal species, including five plant pathogens (Skinner et al. 2001) but generally only in a flat-file format, with limited annotation. A growing number of other EST sets are also being generated from diverse pathogens and await release onto the Internet. As a resource to the phytopathogenic fungi research community, we have designed and implemented a database containing EST data from three pathogenic fungi: *Magnaporthe grisea*, the causal agent of rice-blast disease; *Blumeria graminis*, which causes barley powdery mildew; and *Mycosphaerella graminicola*, which causes Septoria blotch of wheat. The database was designed to be easy to use and to provide information that would be of immediate use to the phytopathogenic fungi research community.

ESTs from *B. graminis* were obtained by sequencing a library constructed by Dr. Sarah Gurr, University of Oxford (Oxford, U.K.), from infected plant material and from an EST dataset produced by the Carlsberg Institute (Copenhagen, Denmark) (Thomas et al. 2001). The dataset comprised 2,701 unisequences, with an average sequence length of 499 base pairs. *M. grisea* ESTs were downloaded from the EMBL database (http://www.ebi.ac.uk/embl/index.html) and are data generated predominately by Dr. Ralph Dean at North Carolina State Biotechnology Center (Research Triangle Park, NC, U.S.A.) (and formerly Clemson University, Clemson, SC, U.S.A.) and Professor Daniel Ebbole at Texas A&M University (College Station, TX, U.S.A.). The dataset comprised 1,839 unisequences, with an average length of 851 base pairs. ESTs from *Mycosphaerella graminicola* were obtained by sequencing three cDNA libraries constructed at IACR-Long Ashton (Bristol, U.K.) by Dr. John Hargreaves and Dr. John Keon (Keon et al. 2000). Two libraries were constructed from fungal mycelium grown in liquid culture and one from fungal-infected plant material. The dataset comprised 2,926 unisequences, with an average length of 667 base pairs.

EST sequences are often single-pass sequences (which are sequence reads from one DNA strand only) and invariably contain sequencing errors. EST data often exhibit high intrinsic redundancy because of a high representation of sequences representing highly expressed mRNAs. EST data can be enhanced when these sequences are used to generate a nonredundant set of unique sequences (unisequences) through cluster assembly (by using Sequencher; Gene Codes Corporation, Ann Arbor, MI, U.S.A.). This removes redundancy from the dataset, as well as improves sequence accuracy and produces longer sequences. Unisequences derived from the EST datasets contain a mixture of assembled consensus sequences and unassembled singletons.

Putative functions were assigned to each unisequence based on similarity to sequences of known genes. The blastx algo-

Table 2. The number of unisequences from each organism allocated to the top level categories of our classification

Category	All unisequences	*Blumeria graminis*	*Magnaporthe grisea*	*Mycosphaerella graminicola*
No hit	4,097	1,786	1,031	1,280
Cloning vector	12	5	0	7
Plant host	41	9	10	22
Unknown protein	527	157	136	234
Metabolism	550	128	116	306
Energy	393	103	89	201
Cell growth, cell division, and DNA synthesis	86	19	24	43
Transcription	177	45	36	96
Protein synthesis	279	90	58	131
Protein destination	357	110	79	168
Transport facilitation	206	45	43	118
Intracellular transport	120	35	25	60
Cellular biogenesis	61	16	15	30
Cellular communication/signal transduction	168	53	49	66
Cell rescue, defense, death, and aging	158	23	43	92
Ionic homeostasis	10	0	3	7
Cellular organization	132	20	54	58
Development	31	11	16	4
Transposon proteins	61	46	12	3
Total	7,466	2,701	1,839	2,926

rithm (Altschul et al. 1990) was used to query the National Center for Biotechnology Information (NCBI; Bethesda, MD, U.S.A.) database (http://www.ncbi.nlm.nih.gov/). The blastx program translates the unisequence into protein sequences in all six reading frames and compares these sequences with those in the database of known protein sequences. The top five most similar sequences (with expectation values less than 1×10^{-5}) were retrieved for each unisequence. The expectation value can be regarded as an indicator of the strength or quality of match between the sequences; the lower the expectation value, the better the match. The cutoff value selected here is more rigorous than the recommended value of 10^{-2} (Anderson and Brass 1998), below which the matches were considered significant in 98% of the cases. On the basis of these similarity scores, a putative product or function was assigned to each unisequence. The assignment of a putative function to a unisequence based on sequence similarity alone is obviously highly speculative in some cases. Within each EST set, we also identified ESTs that represented sequences from vectors used in creating the cDNA libraries or from contaminating host plant material. Based on the assignments, the unisequences were classified by function according to a hierarchical scheme used by the Munich Information Centre for Protein Sequences (MIPS; Neuherberg, Germany) (Mewes et al. 1997). This scheme groups gene products depending on the particular metabolic pathway or cellular process in which they are involved. Table 2 shows the number of unisequences allocated to each top level grouping in our classification system. Out of 7,466 unisequences, 3,369 sequences (45%) showed homology to sequences in the NCBI database. Out of these sequences, 53 (1.6%) represented genes from cloning vectors or plant hosts and 527 (16%) had homology to genes whose functions were unknown. The rest of the sequences represented a wide variety of genes involved in a range of basic metabolic pathways, as well as DNA, RNA, and protein synthesis, protein sorting, and cellular transport. Also present were genes potentially involved in fungal pathogenicity, such as those involved in melanin biosynthesis, fungal toxin synthesis, detoxification of plant defense compounds, plant cell wall degradation, fungal development, and signal transduction.

A relational database scheme was designed to contain the data generated from the ESTs (Fig. 1). The database was implemented with MySQL version 3.23.38 for Solaris 2.7, which was obtained from http://www.mysql.com. The database was first constructed with Microsoft Access 2000 and then transferred to MySQL. Programs to assist in implementing the data base were written in Java version 1.3 (obtained from http://java.sun.com) and Microsoft Visual Basic for Applica-

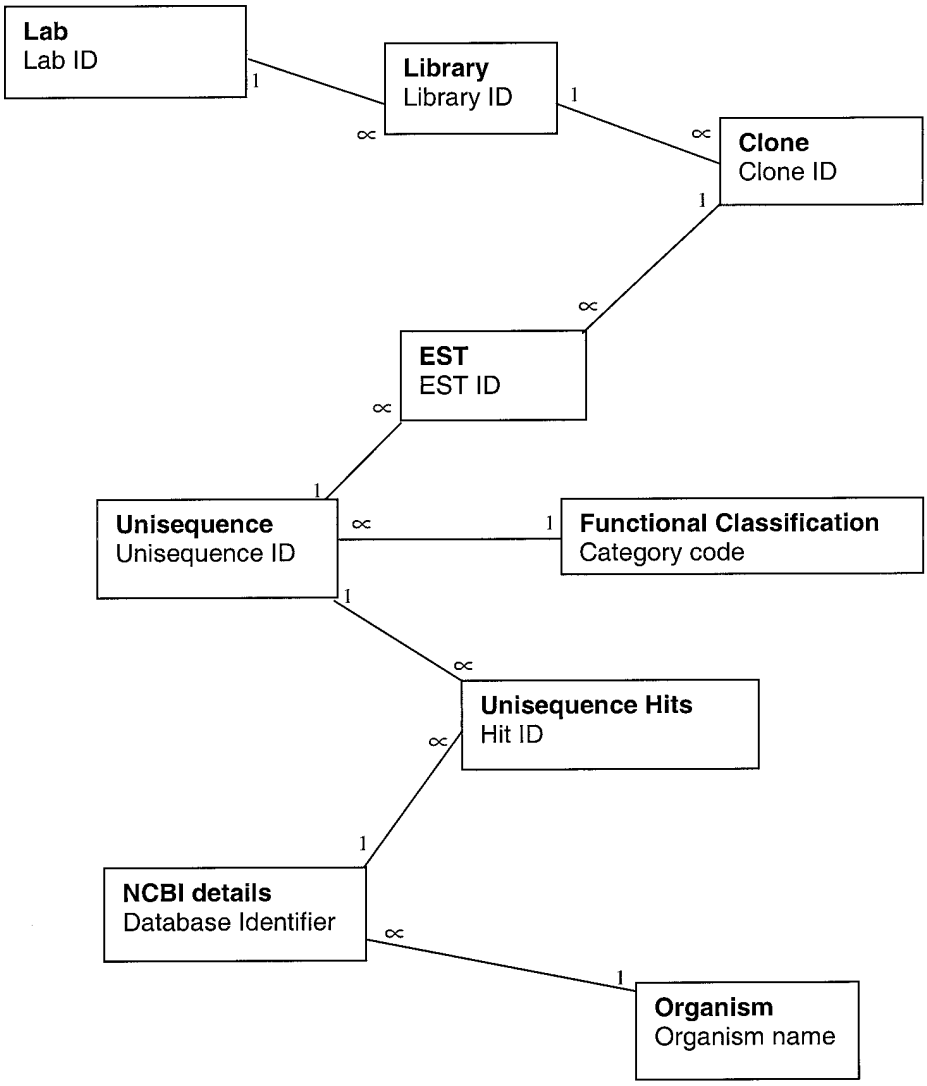

Fig. 1. A schematic representation of the phytopathogen expressed sequence tag (EST) relational database. Each box represents a table in the database. For each table, the title is in bold and below that is the primary key for that table. One to many relationships are shown by links between the tables (∝ = many).

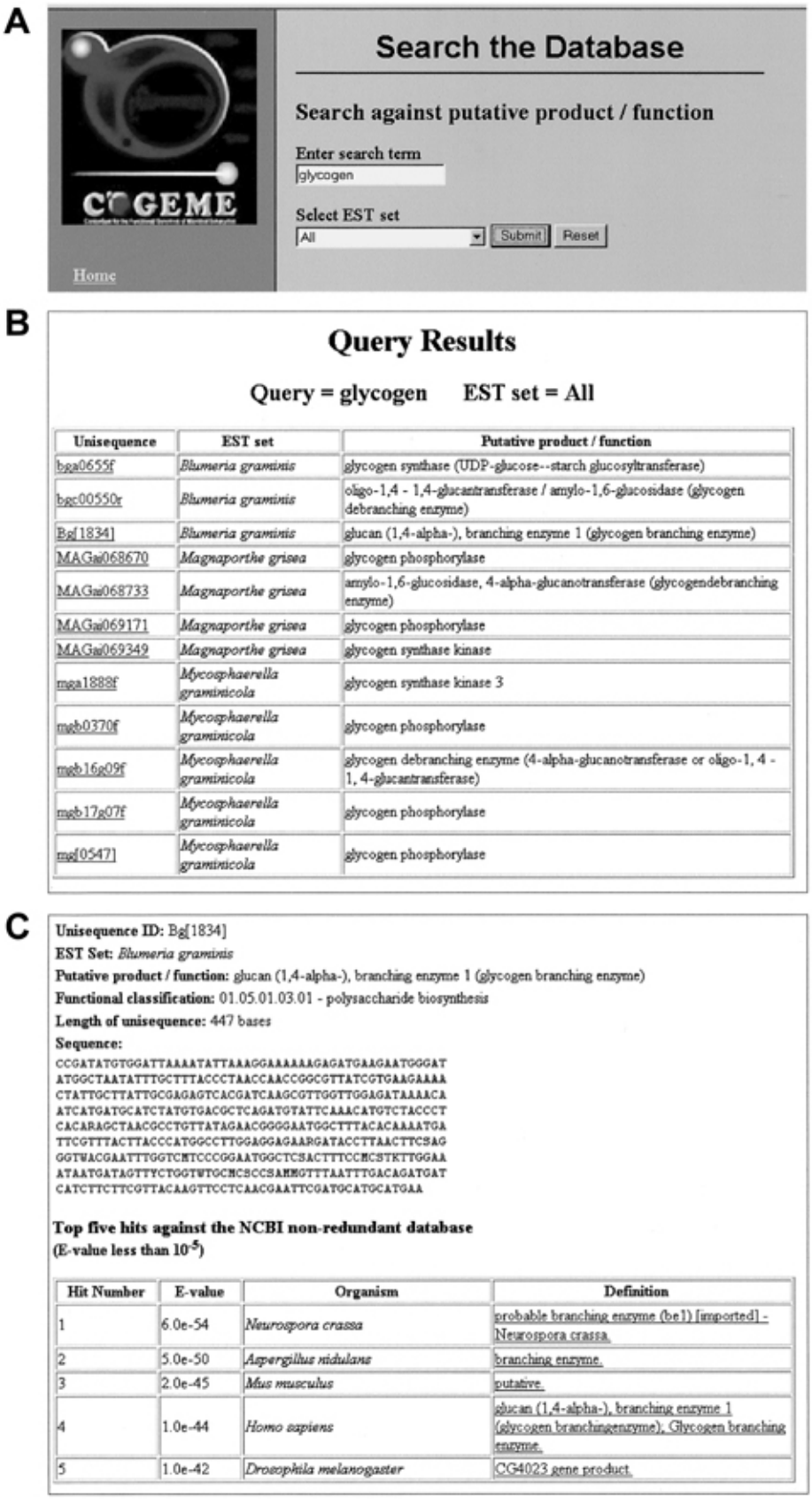

Fig. 2. Screenshots from the web-based front end of the phytopathogen expressed sequence tag (EST) database illustrating a simple text query. **A,** The search page of the website, enabling the user to use a text query to look for unisequences whose putative protein description contains the word provided by the user, who can also choose which EST datasets to search. In this case, the user is searching all EST datasets for unisequences whose putative protein product descriptions contain the word 'glycogen.' **B,** Results of running the query, showing a list of unisequences, the EST set they are from, and a description of their putative product/function. Clicking on the name of a unisequence brings up **C,** a page showing details about that unisequence.

tions. A web-based interface was designed to allow users to search the database (Fig. 2). The interface allows the user to run simple, text-based queries to search for unisequences with particular putative protein products or to search for unisequences that have a similarity to known genes. The web pages were handwritten in HTML. Web-based database searching was implemented with CGI-Perl scripts. Perl programs were written with ActiveState Perl version 5.6 for Windows, which was obtained from http://www.activestate.com. The Perl modules DBI and DBD-Mysql were installed to allow communication with the database. The web interface also allows the user to use BLAST algorithms to find unisequences in the database that match the sequence they input. To allow BLAST searching of the unisequence datasets, the BLAST suite of programs were downloaded from the NCBI BLAST ftp site (ftp://ftp.ncbi.nlm.nih.gov/blast) (Altschul et al. 1990, 1997). Local BLAST databases were created for each unisequence set. A web page and a CGI-Perl script were written to implement the user interface with the BLAST programs. The database is accessible at http://cogeme.ex.ac.uk.

The number of published gene sequences for the three phytopathogenic fungi represented in our database is relatively small. Searches of the NCBI protein database revealed 56 entries for *B. graminis*, 25 for *Mycosphaerella graminicola*, and 196 for *M. grisea*. We envisage that the main purpose of our database will be as a gene discovery tool. By searching the database for a gene in which they are interested, a researcher will be able to use the EST sequence to design primers to enable them to clone that gene relatively easily. If researchers have already cloned a given gene and are interested in identifying homologues present in the genome of organisms represented in our database, then the EST dataset represents a much larger set of sequences than those currently available in the public databases. This project is ongoing and EST datasets will be added for phytopathogenic fungi. EST sequences will soon be added from three other phytopathogenic fungi: *Botrytis cinerea*, *Fusarium graminearum* (*Gibberella zeae*), and *F. sporotrichioides*. Eventually, the database will be a repository for all publicly available EST data from phytopathogenic fungi. The web-based interface will also be updated with new features that will allow more sophisticated text queries and offer the ability to search the database by function, which will allow more analysis of the metabolic pathways present in each fungus. The research groups that supplied the EST data from *B. graminis* and *Mycosphaerella graminicola* are using the cDNA clones that the EST sequences were derived from to produce microarrays that will be used to monitor global changes in gene expression in these organisms under conditions pertinent to pathogenesis. Data from these experiments will be incorporated into the database. The database has been constructed as part of the Consortium for the Genomics of Microbial Eukaryotes (COGEME) project (http://www.cogeme.man.ac.uk/). In due course, the data from our database will be incorporated into the GIMS (Paton et al. 2000) for full comparison with the *S. cerevisiae* genome. One of the goals in studying the genomics of pathogenic fungi is to identify those factors that are necessary for pathogenesis. Comparison of sequence data from fungal pathogens and comparing them with nonpathogenic fungi such as *N. crassa* may identify pathogen-specific orthologues (genes that are present in pathogenic fungi but not in nonpathogenic fungi). However, it is likely that many pathogenicity factors present in phytopathogenic fungi have homologues in nonpathogenic fungi and their expression patterns or functions have been modified to become important in pathogenesis. For example, the mitogen-activated protein (MAP) kinase encoding genes *PMK1* and *MPS1* in *M. grisea* are vital for signal-transduction pathways involved in pathogenesis but also have functional homologues in *S. cerevisiae* (Xu and Hamer 1996; Xu et al. 1998). Microarray analysis will also be a valuable tool for identifying genes that are highly expressed during processes vital to pathogenesis, such as appressorium formation, conidiation, and invasion of plant tissues.

In the current absence of publicly available, completed genome sequences for phytopathogenic fungi, it is hoped that the COGEME database will become a valuable resource for the phytopathogenic fungi research community as a tool for identifying new genes in order to determine their functions.

ACKNOWLEDGMENTS

We would like to thank A. Brass and his students R. Francis, N. Harte, T. Havard, A. O'Brien, and T. Vavouri at the University of Manchester for help in implementing the world wide web interface for the database. This work was supported by the Biotechnology and Biological Sciences COGEME project, which is part of the Investigating Gene Function Initiative.

LITERATURE CITED

Agrios, G. N. 1988. Plant Pathology. Academic Press, San Diego, CA, U.S.A.

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.

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. 25:3389-3402.

Anderson, I., and Brass, A. 1998. Searching DNA databases for similarities to DNA sequences: When is a match significant? Bioinformatics 14:349-356.

Balhadère, P. V., and Talbot, N. J. 2001. *PDE1* encodes a P-type ATPase involved in appressorium mediated plant infection by *Magnaporthe grisea*. Plant Cell 13:1987-2004.

Balhadère, P. V., Foster, A. J., and Talbot, N. J. 1999. Identification of pathogenicity mutants of the rice blast fungus *Magnaporthe grisea* by insertional mutagenesis. Mol. Plant-Microbe Interact. 12:129-142.

Bowyer, P. 1999. Plant disease caused by fungi: Phytopathology. Pages 294-321 in: Molecular Fungal Biology. R. P. Oliver and M. Schweizer, eds. Cambridge University Press, Cambridge.

Bowyer, P., Clarke, B. R., Lunness, P., Daniels, M. J., and Osbourn, A. E. 1995. Host range of a plant pathogenic fungus determined by a saponin detoxifying enzyme. Science 267:371-374.

Cho, R. J., Campbell, M. J., Winzeler, E. A., Steinmetz, L., Conway, A., Wodicka, L., Wolfsberg, T. G., Gabrielian, A. E., Landsman, D., Lockhart, D. J., and Davis, R. W. 1998. A genome-wide transcriptional analysis of the mitotic cell cycle. Mol. Cell 2:65-73.

Chu, S., DeRisi, J., Eisen, M., Mulholland, J., Botstein, D., Brown, P. O., and Herskowitz, I. 1998. The transcriptional program of sporulation in budding yeast. Science 282:699-705.

Clergeot, P. H., Gourgues, M., Cots, J., Laurans, F., Latorse, M. P., Pepin, R., Tharreau, D., Notteghem, J. L., and Lebrun, M. H. 2001. PLS1, a gene encoding a tetraspanin-like protein, is required for penetration of rice leaf by the fungal pathogen *Magnaporthe grisea*. Proc. Natl. Acad. Sci. U.S.A. 98:6963-6968.

DeRisi, J. L., Iyer, V. R., and Brown, P. O. 1997. Exploring the metabolic and genetic control of gene expression on a genomic scale. Science 278:680-686.

Eisen, M. B., Spellman, P. T., Brown, P. O., and Botstein, D. 1998. Cluster analysis and display of genome-wide expression patterns. Proc. Natl. Acad. Sci. U.S.A. 95:14863-14868.

Goffeau, A., Barrell, B. G., Bussey, H., Davis, R. W., Dujon, B., Feldmann, H., Galibert, F., Hoheisel, J. D., Jacq, C., Johnston, M., Louis, E. J., Mewes, H. W., Murakami, Y., Philippsen, P., Tettelin, H., and Oliver, S. G. 1996. Life with 6000 genes. Science 274:546, 563-567.

Gygi, S. P., Rochon, Y., Franza, B. R., and Aebersold, R. 1999. Correlation between protein and mRNA abundance in yeast. Mol. Cell. Biol. 19:1720-1730.

Hamer, J. E., and Givan, S. 1990. Genetic mapping with dispersed repeated sequences in the rice blast fungus: Mapping the SMO locus. Mol. Gen. Genet. 223:487-495.

Hamer, L., Pan, H. Q., Adachi, K., Orbach, M. J., Page, A., Ramamurthy, L., and Woessner, J. P. 2001. Regions of microsynteny in *Magnaporthe grisea* and *Neurospora crassa*. Fungal Genet. Biol. 33:137-143.

Idnurm, A., and Howlett, B. J. 2001. Pathogenicity genes of phytopathogenic fungi. Mol. Plant Pathol. 2:241-255.

Jansen, R., and Gerstein, M. 2000. Analysis of the yeast transcriptome with structural and functional categories: Characterizing highly expressed protein. Nucleic Acids Res. 28:1481-1488.

Kal, A. J., van Zonneveld, A. J., Benes, V., van den Berg, M., Koerkamp, M., Albermann, K., Strack, N., Ruijter, J. M., Richter, A., Dujon, B., Ansorge, W., and Tabak, H. F. 1999. Dynamics of gene expression revealed by comparison of serial analysis of gene expression transcript profiles from yeast grown on two carbon sources. Mol. Biol. Cell 10: 1859-1872.

Keon, J., Bailey, A., and Hargreaves, J. 2000. A group of expressed cDNA sequences from the wheat fungal leaf blotch pathogen, *Mycosphaerella graminicola* (*Septoria tritici*). Fungal Genet. Biol. 29: 118-133.

Knogge, W. 1998. Fungal pathogenicity. Curr. Opin. Plant Biol. 1:324-328.

Kronstad, J. W. 1997. Virulence and cAMP in smuts, blast, and blight. Trends Plant Sci. 2:193-199.

Lashkari, D. A., DeRisi, J. L., McCusker, J. H., Namath, A. F., Gentile, C., Hwang, S. Y., Brown, P. O., and Davis, R. W. 1997. Yeast microarrays for genome wide parallel genetic and gene expression analysis. Proc. Natl. Acad. Sci. U.S.A. 94:13057-13062.

Mewes, H. W., Albermann, K., Heumann, K., Liebl, S., and Pfeiffer, F. 1997. MIPS: A database for protein sequences, homology data and yeast genome information. Nucleic Acids Res. 25:28-30.

Panaccione, D. G., Scott-Craig, J. S., Pocard, J. A., and Walton, J. D. 1992. A cyclic peptide synthetase gene required for pathogenicity of the fungus *Cochliobolus carbonum* on maize. Proc. Natl. Acad. Sci. U.S.A. 89:6590-6594.

Paton, N. W., Khan, S. A., Hayes, A., Moussouni, F., Brass, A., Eilbeck, K., Goble, C. A., Hubbard, S., and Oliver, S. G. 2000 Conceptual modelling of genomic information. Bioinformatics 16:548-558.

Pennisi, E. 2001. The push to pit genomics against fungal pathogens. Science 292:2273-2274.

Rauyaree, P., Choi, W., Fang, E., Blackmon, B., and Dean, R. A. 2001. Genes expressed during early stages of rice infection with the rice blast fungus *Magnaporthe grisea*. Mol. Plant Pathol. 2:347-354.

Seoighe, C., Federspiel, N., Jones, T., Hansen, N., Bivolarovic, V., Surzycki, R., Tamse, R., Komp, C., Hulzar, L., Davis, R. W., Scherer, S., Tait, E., Shaw, D. J., Harris, D., Murphy, L., Oliver, K., Taylor, K., Rajandream, M. A., Barrell, B. G., and Wolfe, K. H. 2000. Prevalence of small inversions in yeast gene order evolution Proc. Natl. Acad. Sci. U.S.A. 97:14433-14437.

Silué, D., Tharreau, D., Talbot, N. J., Clergeot, P.-H., Notteghem, J. L., and Lebrun, M.-H. 1998. Identification and characterization of *apf1*$^{-}$ in a non-pathogenic mutant of the rice blast fungus *Magnaporthe grisea* which is unable to differentiate appressoria. Physiol. Mol. Plant Pathol. 53:239-251.

Skinner, W., Keon, J., and Hargreaves, J. 2001. Gene information for fungal plant pathogens from expressed sequences. Curr. Opin. Microbiol. 4:381-386.

Straney, D. C., and VanEtten, H. D. 1994. Characterization of the *PDA*1 promoter of *Nectria haematococca* and identification of a region that binds a pisatin-responsive DNA binding factor. Mol. Plant-Microbe Interact. 7:256-266.

Sweigard, J. A., Carroll, A. M., Farrall, L., Chumley, F. G., and Valent, B. 1998. *Magnaporthe grisea* pathogenicity genes obtained through insertional mutagenesis. Mol. Plant-Microbe Interact. 11:404-412.

Talbot, N. J., and Foster, A. J. 2001. Genetics and genomics of the rice blast fungus *Magnaporthe grisea*: Developing an experimental model for understanding fungal diseases of cereals. Adv. Bot. Res. 34:287-311.

Talbot, N. J., Ebbole, D. J., and Hamer, J. E. 1993. Identification and characterisation of *MPG1* a gene involved in pathogenicity from the rice blast fungus *Magnaporthe grisea*. Plant Cell 5:1575-1590.

Thomas, S. W., Rasmussen, S. W., Glaring, M. A., Rouster, J. A., Christiansen, S. K., and Oliver, R. P. 2001. Gene identification in the obligate fungal pathogen *Blumeria graminis* by expressed sequence tag analysis. Fungal Genet. Biol. 33:195-211.

Tonukari, N. J., Scott-Craig, J. S., and Walton, J. D. 2000. The *Cochliobolus carbonum SNF1* gene is required for cell wall-degrading enzyme expression and virulence on maize. Plant Cell 12:237-248.

Velculescu, V. E., Zhang, L., Zhou, W., Vogelstein, J., Basrai, M. A., Bassett, D. E., Hieter, P., Vogelstein, B., and Kinzler, K. W. 1997. Characterization of the yeast transcriptome. Cell 24:243-251.

Walton, J. D. 1994. Deconstructing the cell wall. Plant Physiol. 104: 1113-1118.

Wodicka, L., Dong, H., Mittmann, M., Ho, M. H., and Lockhart, D. J. 1997. Genome-wide expression monitoring in *Saccharomyces cerevisiae*. Nat. Biotechnol. 15:1359-1367.

Xu, J.-R., and Hamer, J. E. 1996. MAP kinase and cAMP signalling regulate infection structure formation and growth in the rice blast fungus *Magnaporthe grisea.* Genes Dev. 10:2696-2706.

Xu, J.-R., Staiger, C. J., and Hamer, J. E. 1998. Identification of the mitogen-activated protein kinase *MPS1* from the rice blast fungus prevents penetration of host cells but allows activation of plant defense responses. Proc. Natl. Acad. Sci. U.S.A. 95:12713-12718.

Yamamoto, M., Wakatsuki, T., Hada, A., and Ryo, A. 2001. Use of serial analysis of gene expression (SAGE) technology. J. Immunol. Methods 250:45-66.

Yoder, O. C., and Turgeon, B. G. 2001. Fungal genomics and pathogenicity. Curr. Opin. Plant Biol. 4:315-321.