MPMI Vol. 19, No. 12, 2006, pp. 1451–1462. DOI: 10.1094/MPMI-19-1451. © 2006 The American Phytopathological Society

The Pathogen-Host Interactions Database (PHI-base) Provides Insights into Generic and Novel Themes of Pathogenicity

Thomas K. Baldwin,[1] Rainer Winnenburg,[2] Martin Urban,[1] Chris Rawlings,[2] Jacob Koehler,[2] and Kim E. Hammond-Kosack[1]

[1] Plant-Pathogen Interactions Division, Rothamsted Research, Harpenden, AL5 2JQ, U.K.; [2] Biomathematics and Bioinformatics Division, Rothamsted Research, Harpenden, AL5 2JQ, U.K.

Submitted 14 March 2006. Accepted 12 June 2006.

Fungal and oomycete pathogens of plants and animals are a major global problem. In the last 15 years, many genes required for pathogenesis have been determined for over 50 different species. Other studies have characterized effector genes (previously termed avirulence genes) required to activate host responses. By studying these types of pathogen genes, novel targets for control can be revealed. In this report, we describe the Pathogen-Host Interactions database (PHI-base), which systematically compiles such pathogenicity genes involved in pathogen-host interactions. Here, we focus on the biology that underlies this computational resource: the nature of pathogen-host interactions, the experimental methods that exist for the characterization of such pathogen-host interactions as well as the available computational resources. Based on the data, we review and analyze the specific functions of pathogenicity genes, the host-specific nature of pathogenicity and virulence genes, and the generic mechanisms of effectors that trigger plant responses. We further discuss the utilization of PHI-base for the computational identification of pathogenicity genes through comparative genomics. In this context, the importance of standardizing pathogenicity assays as well as integrating databases to aid comparative genomics is discussed.

Additional keywords: gene disruption, yeast.

Pathogens are defined as species that are able to infect and cause disease on a host to complete their life cycle. Of the 12,000 fungal species known to exist, most are nonpathogenic (Knogge 1996). However, many diseases of plants are caused by fungal and oomycete pathogens and these infect one to numerous hosts (Agrios 1997). Each plant pathogenic species has evolved a special way to invade plants and to cause disease. Some penetrate host surfaces directly, while others enter through natural openings or via wounds. Once within the plant tissue, three main strategies of colonization are deployed to utilize the host plant as a substrate for growth and development: i) necrotrophy, in which the plant cells are killed; ii) biotrophy, in which the plant cells remain alive; and iii) hemibiotrophy, in which the pathogen initially keeps plant cells alive but kills them at a later stage of the infection process. Research over the last 15 years has begun to reveal that, although different groups of pathogen species have distinct modes of infection, many common gene products are required to cause disease in both cereal and noncereal species.

Symbiosis is the description for an intimate association between two or more organisms of different species that benefits at least one of them (de Bary 1879). Symbiosis is, in most cases, of mutual benefit. However, pathogenic interactions in which one organism is affected detrimentally are also included within this original term. Although the outcomes of pathogenic and mutualistic interactions are very different, the infection mechanisms and the molecular processes in the pathogen-host interactions are often very similar. Pathogenicity is a qualitative description of the ability of a microorganism in a symbiotic relationship to cause disease to its host. In contrast, virulence (or aggressiveness) is a quantitative measure of pathogenicity (Shaner et al. 1992). Effector genes in fungi (previously termed avirulence genes) encode proteins that trigger a hypersensitive response in a plant host that possesses one or more corresponding resistance genes (van Dijk et al. 1999). Some effector genes, e.g., *NIP1* in the wheat leaf pathogen *Rhyncosporium secalis*, have also been demonstrated to be a pathogenicity factor (Rohe et al. 1995).

One approach to understand the networks and processes that allow a microbe to become pathogenic and colonize a plant or animal is to identify the genes that control its disease-causing ability. By understanding which pathways, proteins, and genes are key components of the pathogenic process, targets in the fungus can be identified for chemical control (fungicides or antifungal drugs) or targets in the host can be modified to reduce their susceptibility to the pathogen, e.g., through the use of genetically modified crop genotypes or viral therapeutics.

As a result of molecular genetic studies, many pathogenicity and virulence genes are now recognized and comprehensive reviews of this topic have been published recently (Idnurm and Howlett 2001; Talbot 2004; Xu 2000). High throughput forward genetic studies have been applied to identify novel pathogenicity genes in several plant pathogens (Dufresne et al. 1998; Sweigard et al. 1998). For example, *Agrobacterium tumefaciens* or protoplast transformation is used to create a library of mutants with single plasmid insertions so that the mutation is tagged by the inserted plasmid (de Groot et al. 1998; Lashkari et al. 1997b). Reduced virulence or nonpathogenic mutants are then identified through infection assays. One of the first pathogenicity genes discovered was the *Cochliobolus carbonum HTS1* gene, which is essential for HC-toxin production and pathogenicity on maize leaves (Panaccione et al. 1992). In the

Thomas Baldwin and Rainer Winnenburg contributed equally to this article.

Corresponding author: Kim Hammond-Kosack; Telephone: +00 44 1582 763133; Fax +44 1582 760089; E-mail: kim.hammond-kosack@bbsrc.ac.uk

rice blast fungus *Magnaporthe grisea*, the *ABC1* gene is a pathogenicity gene and the *abc1* mutant is nonpathogenic in several different infection assays (Urban et al. 1999). Virulence genes affecting the degree of disease-causing ability include the *M. grisea CPG1* gene coding for a cyclophilin protein. The Δ*cpg1* mutants are less virulent than the wild-type parent strain (Viaud et al. 2002). Frequently, when homologs of verified pathogenicity or virulence genes are disrupted in other species (reverse genetics), mutants are nonpathogenic or less virulent than the wild-type parent strains (Viaud et al. 2003; Wang et al. 2001). Disruptions in genes with functions as diverse as signaling (Talbot et al. 1993), biosynthesis of phytotoxic compounds (Johnson et al. 2000), sporulation (Klimpel et al. 2002), infection structure formation (DeZwaan et al. 1999), and detoxification of plant defense compounds (Wasmann and VanEtten 1996) have all been shown to affect pathogenicity. Overall these molecular genetic studies have focused on a handful of species that are easily studied in the laboratory (Idnurm and Howlett 2001). This has led to in-depth knowledge of pathogenicity components for several fungal species, for example, the ascomycete rice blast fungus *M. grisea* (Talbot 2004) and the basidiomycete corn smut fungus *Ustilago maydis* (Feldbrugge et al. 2004). Many of the early studies focused on the identification of pathogenicity genes required for penetration. However, in the recent past the emphasis has shifted towards the identification of the genes required for local or systemic (or both) host tissue colonization and, indeed, other aspects of the life-cycle, such as sporulation and spore dissemination.

For comparative gene-function studies, it is important to include fungal and oomycete pathogens of animals and other organisms with plant pathogens and saprophytes. Some studies have shown that a single pathogen species requires different genes for pathogenesis in animal and plant hosts (Ortoneda et al. 2004). However, other genes are virulence factors in animal pathogens but not in plant pathogens. For example, the homologue of the *Saccharomyces cerevisiae* mitogen-activated protein (MAP) kinase *HOG1* is a major virulence factor of *Candida albicans* in its murine host (Alonso-Monge et al. 1999) but does not affect the pathogenicity of the rice blast pathogen *M. grisea* (Dixon et al. 1999). In contrast, homologs of other genes are conserved virulence factors in distantly related pathogens of both plants and animals. For example, the basidiomycete *Cryptococcus neoformans* (Kraus et al. 2003), the ascomycete *Candida albicans* (DiezOrejas et al. 1997), and several plant pathogens (Hou et al. 2002; Kojima et al. 2002; Xu et al. 1998) all require homologs of another *S. cerevisiae* MAP kinase, *SLT2*, for full pathogenicity. In cases like the MAP kinase gene family that is found in both pathogenic and saprophytic species, these sequences have clearly evolved to perform specific functions during pathogenic and nonpathogenic growth, development, or both.

Four new approaches to identify candidate pathogenicity genes have recently emerged in the genomic and postgenomic eras. The first is to create a library of cDNA clones from pathogen genes transcribed either when the pathogen is in association with the host or when the pathogen is grown in vitro under conditions that are presumed to simulate the host environment. The library can then be partially sequenced to create a set of expressed sequence tags (EST). Although EST only give an incomplete representation of the genome, as not all genes are transcribed simultaneously, EST have been used in transcriptomic studies to identify candidate pathogenicity genes (Keon et al. 2005a; Soanes et al. 2002; Torto et al. 2003). One spectacular success achieved by this approach was the identification of *ATR13*, the first effector gene of the obligate biotroph *Hyaloperonospora parasitica* (Allen et al. 2004). The COGEME (Consortium for the Functional Genomics of Microbial Eukaryotes) database now offers researchers a total of 59,765 unique sequences from 15 plant pathogens, submitted by several different groups. The database allows searches by keyword, BLAST sequence similarity searches (Altschul et al. 1997), and in silico analysis of RNA abundance in cDNA libraries prepared from different stages of the pathogen's life cycle (Altschul et al. 1997; Soanes 2005; Soanes et al. 2002). The second approach to identify candidate pathogenicity genes is to sequence a part or the full genome. Full-genome sequence is already available for 11 plant pathogens, including *Ashyba gossypii* and *M. grisea*, and 12 animal pathogens (discussed below) (Broad Institute 2005; Dean et al. 2005; Feldbrugge et al. 2004; Soanes 2005). The genomes of many additional pathogenic and nonpathogenic microbes will become available in the next two years. By applying various sequence similarity search algorithms to either EST collections or newly sequenced genomes, it is relatively simple to identify potential functional homologs of experimentally verified pathogenicity, virulence, and effector genes in other species. The third approach to identify candidate pathogenicity genes is the study of holistic changes of transcription in response to changes in their environment. This has already been used successfully to reveal the role of genes in *S. cerevisiae* (DeRisi et al. 1997; Lashkari et al. 1997a). The use of microarrays printed with thousands of EST or oligonucleotides to represent a large proportion of the genome has greatly increased the amount of data that can be gathered about changes in gene expression. A transcriptomic approach to discover candidate pathogenicity genes has recently been described for three cereal-attacking pathogens, *M. grisea*, *Blumeria graminis*, and *Mycosphaerella graminicola* (Both et al. 2005; Keon et al. 2005b; Takano et al. 2003), and the human pathogens *Candida albicans* and *Cryptococcus neoformans* (Kraus et al. 2004; Lee et al. 2004). However, expression of a gene during infection or invasion of a host does not necessarily mean that the gene affects pathogenicity. The role of a putative pathogenicity gene has to be verified experimentally by targeted gene or transcript disruption and virulence assays on the host. The development of gene-silencing technologies for various fungi and oomycete species is assisting in the functional evaluation of candidate gene lists (Kamoun et al. 1998; McDonald et al. 2005). The final approach has been successfully applied to obligate biotrophs to identify effector genes. By sequencing candidate genes in many virulent and avirulent strains, it was possible to identify that a specific gene sequence variant was always associated with the ability to activate plant defense responses. To confirm these effector genes induced resistance, gene-mediated plant cell death with the appropriate specificity, either *Agrobacterium tumefaciens*–mediated transient expression of the gene was undertaken in a range of host genotypes, the effector gene was stably expressed in a susceptible host, or both, and an F_1 seedling lethal or stunted growth phenotype was observed when the transgenic host plant line expressing the effector gene was crossed to resistant lines containing the corresponding functional resistance gene (Catanzariti et al. 2006; Dodds et al. 2004).

When studying host-pathogen relationships, another very important requirement, especially for effective comparative studies, is the use of ontologies and controlled vocabularies for describing the function, localization, tissue specificity, and other characteristics of gene products. The necessity for systems such as the gene ontology became more apparent as the genomes of model species were sequenced and annotated. To avoid the confusion that can arise when different genes that have the same function in different (or even the same) organisms are described using different words, the Gene Ontology (GO) consortium was founded in 1999 (Ashburner et al. 2000).

Other ontologies were soon added, many of which are now assembled under the Open Biological Ontologies umbrella organization for structured shared controlled vocabularies and ontologies for use within the genomics and proteomics domains. When ontologies or other controlled vocabularies such as the enzyme nomenclature are used for gene annotation, all genes that have the same function can be identified and retrieved using the same term. A specific interest group, the Plant-Associated Microbe Gene Ontology Interest Group (PAMGO), was set up in 2004 to ensure that specific GO terms are developed for plant-microbe interactions (Lindeberg et al. 2005). Several terms have already been created, including a general term for pathogenesis (GO:0009405) as well as more specific terms, such as host cell invasion (GO:0030260) and avoidance of host defenses (GO:0044413). This should aid comparisons of gene product function between plant pathogens with diverse infection biologies, especially as more specific GO terms are created for gene products involved in pathogenesis (PAMGO website).

Although knowledge about pathogens and nonpathogenic microbial species is accumulating rapidly, it is often difficult to access full data sets and compare results because this information is mainly available in the literature or in the laboratories of individual investigators. The establishment of a web-accessible database to collate, cross-link, and categorize genotypic and phenotypic information of individual pathogens and gene deletion mutants will greatly facilitate an increased understanding of general pathogenicity mechanisms. For example, the identification of homologous genes in multiple species required for pathogenicity or triggering a host response could potentially reveal new generic targets for drug design and may result in the development of novel disease control strategies. However, a reliable source of collated data is an absolute prerequisite to achieve this objective. Also, by using an accurate source of collated data, the annotation of newly sequenced genomes can be achieved by using sequence similarity searches and homologs of verified pathogenicity, virulence, and effector genes rapidly detected. Electronic gene annotation currently suffers from the poor reliability of such predictions. Gene functions are often indirectly inferred based on sequence similarity from genes that themselves have been annotated based solely on sequence similarity (Gattiker et al. 2003). Often, it is not possible to identify the gene that was originally annotated with a given function based on experimental evidence. Even if such sources can be identified, the experimental evidence of a given gene function is frequently weak. Thus, a database that systematically annotates genes based on strong experimental evidence is needed. Only then can the successful exploitation of sequence similarity based electronic methods reliably identify putative homologs of pathogenicity genes in newly sequenced genomes.

In this paper, we describe the biology that underlies a new computational resource as well as the plans for its future development. This resource is called the Pathogen-Host Interactions database (PHI-base), and it contains expertly curated molecular and biological information on genes proven to affect the outcome of pathogen-host interactions. The technical details used to develop and maintain this resource are described elsewhere (Winnenburg et al. 2005).

RESULTS

Word searches and literature curation.

Using the keyword search terms, only 10% of the returned articles included information about the disruption and characterization of genes. Figure 1 shows, on a yearly basis, the number of articles returned with the search term and the sought-

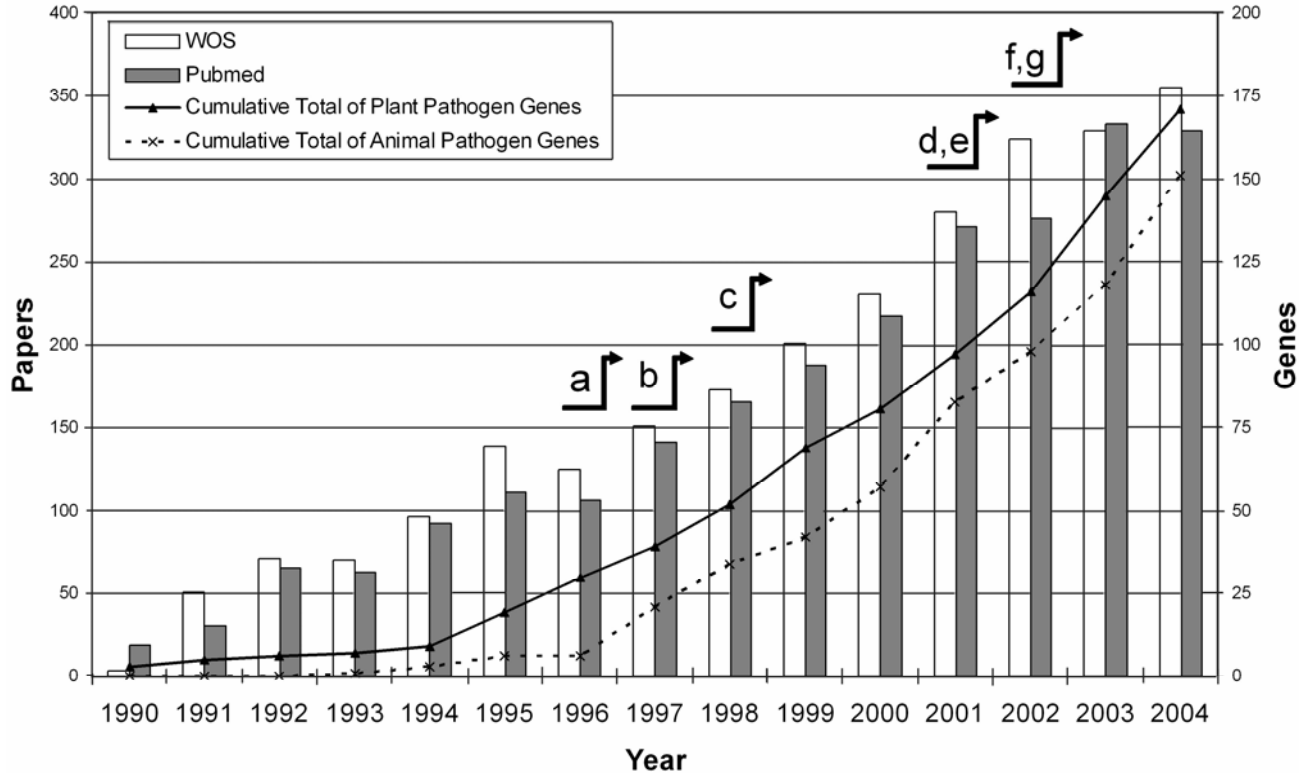

Fig. 1. Number of publications returned with the search term specified in the text and the number of pathogenicity, virulence, and effector genes experimentally verified each year in fungal and oomycete pathogens. Also indicated are significant dates in the history of fungal genomics: a = *Saccharomyces cerevisiae* genome sequenced (Goffeau et al. 1996); b = first *Magnaporthe grisea* expressed sequence tag deposited in GenBank; c = *Candida albicans* genome sequenced (Jones et al. 2004); d = *Ashbya gosypii* genome sequenced (Dietrich et al. 2004); e = first version of COGEME made available online (Soanes et al. 2002); f = *Neurospora crassa* genome sequenced (Galagan et al. 2003); and g = *Magnaporthe grisea* genome sequenced (Dean et al. 2005).

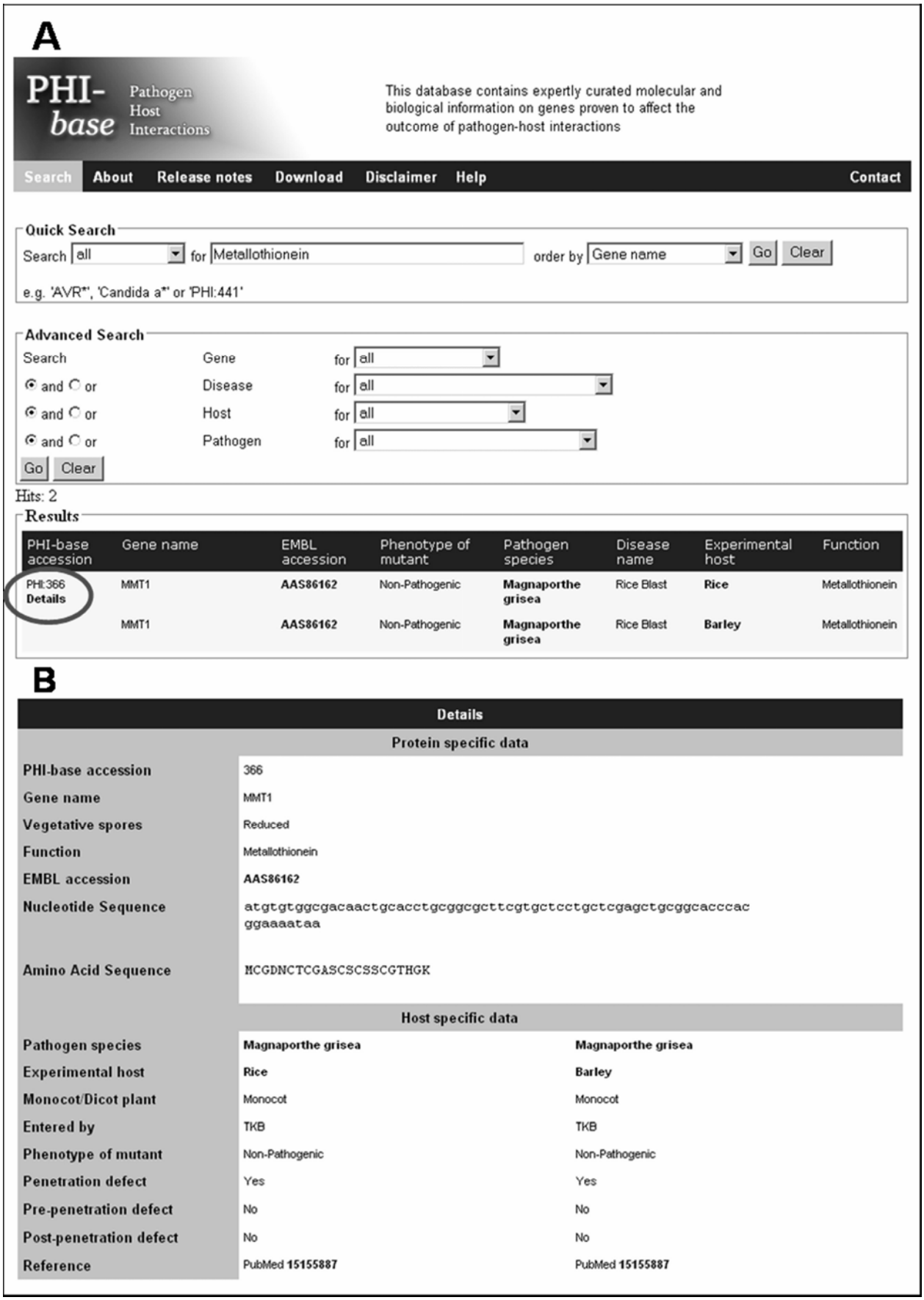

Fig. 2. Screenshots of PHI-base. **A,** Homepage showing the different search fields available to retrieve information from the database. **B,** Details page showing an example of the results retrieved after a search for the *Magnaporthe grisea MMT1* metallothionein gene.

after subset that were the required functionally characterized pathogenicity, virulence, and effector genes. Overall, there has been a steady increase in the number of pathogenicity genes discovered during the last decade. To date, more than 350 genes have been experimentally verified as pathogenicity or virulence genes. In addition, a smaller number of pathogen effector genes have been demonstrated to be required to trigger plant responses. The annual rate of pathogenicity gene discovery has increased from nine in 1995 to 55 in 2004. During this time period, the efficiency of pathogen transformation has increased through the use of *Agrobacterium* and improved protoplasting techniques. New methods have emerged for the identification of gene function, e.g., by gene silencing and the sequencing of multiple allele variants. Also, the availability of genomic sequence permits the identification of highly homologous genes in other species and simplifies the construction of disruption cassettes. The rate of gene function discovery can be expected to increase further in the future. Interestingly, a high proportion (52%) of the total entries in PHI-base were recovered from just five journals. These were *Molecular Plant-Microbe Interactions*, *Molecular Microbiology*, *Infection and Immunity*, *The Plant Cell*, and *Eukaryotic Cell*. The remaining 48% of PHI-base entries were obtained from an additional 45 journals.

Entry annotation.

The data, which is currently still curated in a spreadsheet above, is transferred to a web-enabled database. This database uses a relational database management system in the backend (PostgreSQL) and the HTML-embedded scripting language PHP to generate the webpages that can be used for searching and retrieving information from the database. The data from the spreadsheet are transferred by a parser, which also integrates further information from other external data sources with the data from the spreadsheet. Currently, nucleotide and protein sequences, GO annotations, as well as enzyme commission numbers are extracted from the European Bioinformatics Institute databases and are included in PHI-base. The parser also generates hyperlinks to external resources, such as the National Center for Biotechnology Information (NCBI) Taxonomy database, Pubmed links, and GO terms. In addition, the parser checks and enforces syntactic correctness of the data in the spreadsheet before incorporating information into PHI-base. These technical methods are described in greater detail by Winnenburg and associates (2006).

A web-based interface has been designed to allow keyword searches of the database. Drop-down menus give options to select for individual pathogens, hosts, genes, or diseases. PHI-base currently contains gene entries from 62 fungal and oomycete pathogen species, of which 176 are from animal pathogens, 226 from plant pathogens, and three from pathogens with a fungal host. Screenshots of the online database are displayed in Figure 2. From the front page (Fig. 2A), it is possible to enter a word or phrase within the Quick Search function or to use the advanced search function and select from the various drop-down menu lists either a gene or disease name or a host or pathogen species. On the follow-up details page, accessed by clicking on the word 'details' (Fig. 2A, circled), are the full descriptions of the interaction on each host species tested. For example, PHI-base entry 366, deletion of the *MMT1* gene encoding a metallothionein from *Magnaporthe grisea*, was evaluated on host species rice and barley and was found to have a penetration defect on both species. The details page also reveals when a range of phenotypes have been recorded when a single gene-deletion strain is inoculated onto different host species. For example, PHI-base entry 44, deletion of the *Tri5* gene coding for the enzyme trichodiene synthase from *Fusarium graminearum* details six different interactions on four different cereal plant hosts. By inclusion of the reference ID or citing the actual reference on the details page, it is also possible to access additional information from the original article. The details page for PHI-base entry 7 (Avr9 of the tomato pathogen *Cladosporium fulvum*) also illustrates how PHI-base can be used not only to find the original publication (van Kan et al. 1991) but also provides immediate access to the relevant pre- and follow-up publications on a well-studied gene, protein, and interaction. Also in the details page, there are action link-outs to EMBL, NCBI, Pubmed, and Expasy to permit the immediate retrieval of additional information.

The curation of this database is a continual process as new literature is published.

Species contents.

Version 2.1 of PHI-base contains information on 62 pathogenic species. Closer inspection reveals that 284 of the 405 entries, i.e., 70%, come from experiments on nine species (Table 1). This table also reveals that approximately a quarter of the database contains entries from *Candida albicans* gene disruption experiments.

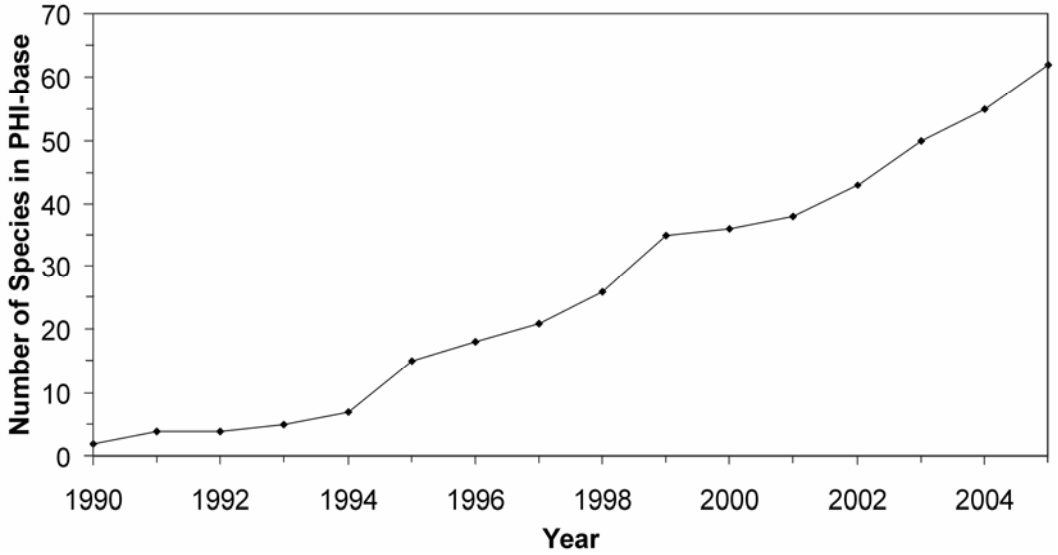

Fig. 3. Cumulative number of species with one or more experimentally verified pathogenicity, virulence, or effector genes.

Studies on pathogenicity genes through targeted gene disruption experiments commenced at the start of the 1990's. Since 1994, a steady increase in the number of species with at least one experimentally verified virulence, pathogenicity, or effector genes published has been seen (Fig. 3).

Predicted and known gene function.

Using the MIPS classification scheme (Ruepp et al. 2004), the functional categories within PHI-base were investigated (Fig. 4). Over 75% of all entries could be assigned to just four categories: i) cellular communication/signal transduction, ii) metabolism, iii) cell rescue, defense, cell death and aging, and iv) transcription. However, this distribution may be somewhat biased by the number of genes for which function was discovered through sequence homology searching followed by a reverse genetics experiment. The striking feature of this analysis is the lack of the category 'unknown function.'

Genes specifically required for pathogenicity on plant or animal hosts, or both.

Further analyses of all the PHI-base entries revealed that 15 genes with distinct functions have been disrupted in both plant and animal pathogens and shown to abolish pathogenicity or reduce virulence (Table 2). Some of these gene products could be suitable as generic targets for intervention and thereby provide broad-spectrum species control. In contrast, the number of genes solely required for animal pathogenicity is currently two, while five gene products appear to be specifically required to infect and cause disease on a plant host. The latter two datasets should, however, be interpreted with caution, because the lack of evidence may simply reflect the fact that experiments are still in progress or not yet attempted. The entries pectate lyase and endopolygalacturonase are likely to be solely a plant pathogen–specific requirement because these function to breakdown the complex carbohydrates constituting the plant cell wall.

PHI-base entries of special biological significance.

Effector genes that activate or suppress plant defense responses. Although the effector category is small in PHI-base (22 entries), these sequences are of great interest because, either through a direct or indirect recognition process, effectors activate host defense responses in host genotypes expressing the corresponding resistance genes. These molecular events cause interspecies incompatibility and minimal disease formation occurs. Twelve effectors have been described for fungal ascomycetes. These are cysteine-rich secreted proteins Avr2, Avr4, Avr9, and Ecp2 from *Cladosporium fulvum* (PHI: 472, 18, 7, and 71, respectively), Nip1 from *Rhynchosporium secalis* (PHI: 38), Avr-Pita (originally called Avr2-YAMO), PWL1,

Table 1. Species with ten or more experimentally verified pathogenicity, virulence and effector genes in PHI-base

Species Name	Host	Taxonomy	Number of genes
Candida albicans	Animal	Ascomycota	102
Cryptococcus neoformans	Animal	Basidiomycota	50
Magnaporthe grisea	Plant	Ascomycota	44
Ustilago maydis	Plant	Basidiomycota	29
Aspergillus fumigatus	Animal	Ascomycota	13
Botrytis cinerea	Plant	Ascomycota	12
Fusarium graminearum	Plant	Ascomycota	12
Fusarium oxysporum	Plant	Ascomycota	12
Colletotrichum lagenarium	Plant	Ascomycota	10

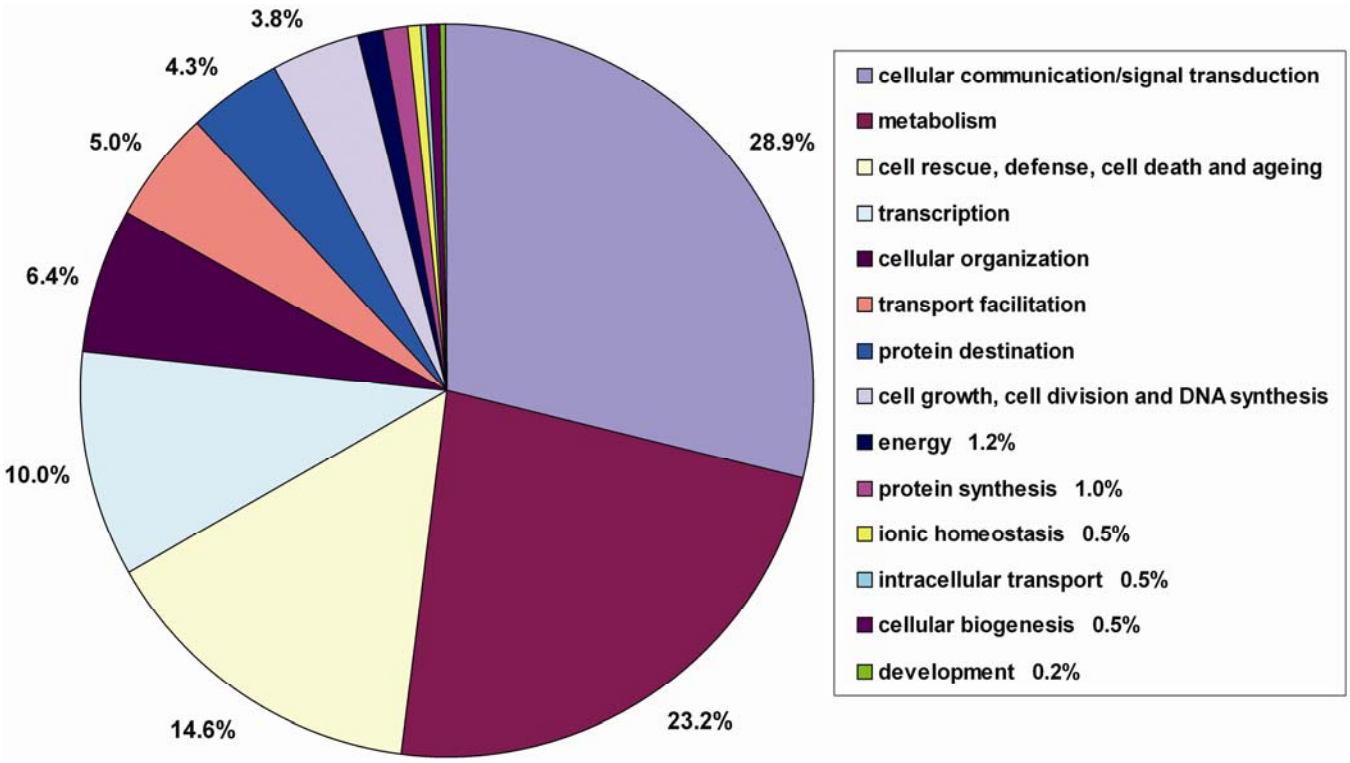

Fig. 4. MIPS classifications of the PHI-base entries.

PWL2, and Ace1 from *Magnaporthe grisea* (PHI: 201, 41, 42, and 325, respectively), AvrL567-A, AvrL567-B, AvrM, AvrP4, and AvrP123 from *Melampsora lini* (PHI: 532, 533, 535, 536, and 537), AvrLm1 from *Leptosphaeria maculans* (PHI: 25), and SIX1 from *Fusarium oxysporum* f. sp. *lycopersici* (PHI: 379). For pathogenic oomycetes and using map-based cloning or gene-silencing technologies, five effectors have been functionally verified. These are INF1 and Avr3a from the potato late blight pathogen *Phytophthora infestans* (PHI: 111 and 473, respectively), Avr1b-1 from the soybean pathogen *Phytophthora sojae* (PHI: 530), and ATR13 and ATR1^{NdWsB} from the downy mildew *Hyaloperonospora parasitica* (PHI: 328 and 531). Some effectors have been shown to suppress defense activation in nonhost species (e.g., entry INF1) and this leads to host susceptibility. Sequence similarity searches with effector genes reveal that most of these sequences are pathogen species–specific. However, the oomycete effector genes share a conserved RXLR motif within 32 amino acids of the predicted signal peptide that is similar to a host-targeting signal from malaria parasites. The effector category has not been defined by researchers investigating human- and animal-attacking fungal and oomycete pathogens. Interestingly, most of the effector sequences are small, secreted proteins within the size range of 28 to 284 amino acids. The two exceptions are the *M. grisea* entries Avr-Pita and Ace1, which both code for larger cytoplasmically localized proteins thought to have an enzymatic function. In addition, Pep-13 derived from a 42-kDa cell-wall transglutaminase (from various *Phytophthora* species PHI: 39) and NPP1 from *H. parasitica* (Fellbrich et al. 2002) are recognized as inducers of defense responses in the nonhost species parsley.

Gene disruptions that lead to hypervirulence phenotype. Most of the PHI-base entries report on gene disruption experiments that lead to either a loss of pathogenicity or reduced virulence. However, in five cases, a single gene disruption lead to an increase in virulence, i.e., hypervirulence. These are transcription factor ACE2 for *Candida glabrata* (PHI: 326), the regulatory protein PKR2 and G protein CRG1 for *Cryptococcus neoformans* (PHI: 372 and 345, respectively), and the PacC transcription actor of *Fusarium oxysporum* (PHI: 315) when infecting its tomato but not murine host, as well as the chitin synthase gene CHSV when infecting its murine host (PHI: 285). Overall, these results indicate that different negatively regulated feedback loops are operating during the infection process of both animal and plant host species.

Table 2. Gene functions required for infection of plant and animal hosts

Gene Function	Plant Pathogen[a]	Animal Pathogen
Required by both plant and animal pathogens		
MAP kinase	12	2
G protein subunit	10	2
ABC transporter	5	1
Adenylate cyclase	4	2
Chitin Synthase	3	2
PKA catalytic subunit	3	2
Polyketide synthase	3	2
MAP kinase kinase kinase	3	2
MAP kinase kinase	3	1
Cyclophilin	2	1
Isocitrate lyase	2	1
Fatty acid synthetase	1	1
Methionine synthase	1	1
Trehalose-6-phosphate synthase	1	1
Superoxide dismutase	1	3
Required by plant pathogens		
Adenylate forming enzymes	3	0
Tetraspanin	3	0
Endopolygalacturonase	3	0
Pectate lyase	2	0
PKA regulatory subunit	2	0
Required by animal pathogen		
Aspartyl proteinase	0	1
Phospholipase B	0	2

[a] Number of species in which disruption of this gene type has resulted in an altered pathogenicity or virulence phenotype.

DISCUSSION

In this article, we describe the curation of literature that has been verified through experimentation to be functionally relevant to pathogen-host interactions. We also describe the creation of a database called PHI-base, which provides, for the first time, this comprehensive and reliable information at a single source. In the first instance, PHI-base is expected to be a useful resource for researchers working on pathogenicity genes of plant and animal pathogens. However, it is also envisaged to become a useful tool for researchers in the wider fungal and oomycete research community investigating saprophytes and nonculturable species through the increasing use of comparative genomics. A database of verified pathogenicity genes should be useful to PAMGO researchers developing more specific GO terms for plant pathogens.

One major obstacle to the curation of PHI-base was the varying approaches that are taken to phenotype mutants. Attempts to categorize mutants into those that have a pathogenicity defect before, during, or after penetration has sometimes

Table 3. Publicly available complete genome sequences of fungal and oomycete pathogens and databases

Fungus	Sequence coverage	Source[a] (Reference)
Plant Pathogens		
Ashbya gossypii	4.2 × 9 Mb	A-2004 (Dietrich et al. 2004)
Botrytis cinerea	4 × 30 Mb	B-2005
Fusarium graminearum	10 × 36 Mb	B-2003
Fusarium verticillioides	4 × 46 Mb	B-2005
Magnaporthe grisea	7 × 39 Mb	B-2002 (Dean et al. 2005)
Phanerochaete chrysosporium	11 × 30 Mb	C- 2005
Phytophthora infestans	1 × ~237 Mb	D- 2005 (Randall et al. 2005)
Phytophthora sojae	9 × 95 Mb	C-2005
Phytophthora ramorum	7 × 65 Mb	C-2005
Stagonospora nodorum	10 × 37 Mb	B-2005
Ustilago maydis	10 × 19 Mb	B-2004
Human pathogens		
Aspergillus fumigatus	10 × 28 Mb	E-2005 (Nierman et al. 2005)
Aspergillus terreus	11 x 35 Mb	B-2005
Cryptococcus neoformans serotype A	11 × 19 Mb	B-2005
Cryptococcus neoformans serotype B	6 × 17 Mb	B-2005
Candida albicans	11 × 16 Mb	F- 2004 (Jones et al. 2004)
Candida dubliniensis	9 × 16 Mb	G- 2005 (Sanger 2005)
Candida glabrata	10 × 14 Mb	H- 2005
Candida guilliermondii	12 × 11 Mb	B-2005
Candida lusitaniae	9 × 12 Mb	B-2005
Candida tropicalis	10 × 40 Mb	B-2005
Chaetomium globosum	7 × 34 Mb	B-2005
Coccidioides immitis	10 × 29 Mb	B-2005
Histoplasma capsulatum	7 x 28 Mb	B-2005
Rhizopus oryzae	10 × 40 Mb	B-2005
Unicinocarpus reesii	5 x 30 Mb	B-2005

[a] Letters indicate websites as follows: A = Swiss Institute of Bioinformatics; B = Broad Institute-; C = Joint Genome Institute; D = National Center for Genome Resources; E = University of Manchester; F = Institut Pasteur; G = Sanger Institute; H = Center for Bioinformatics, Bordeaux.

proved to be difficult. Some researchers do not publish enough experimental details to be able to verify whether a reduction in virulence is caused by a penetration defect or a defect in post-penetrative growth. The development of PHI-base will be greatly aided if each research community agrees to carry out a set of 'gold standard' pathogenicity tests for each pathosystem to assess the phenotype of mutants. Examples of good tests for interspecific comparisons would be microscopy analyses of the plant infection process for evidence of infection structure formation, penetration defects, as well as local and systemic growth patterns through host tissues. Comparisons between wounded and intact host surfaces would indicate whether a reduction in pathogenicity is only due to an inability to penetrate the host. For example, the global Fusarium community implemented in 2005 a standardized score card scheme to permit mutant phenotype comparisons between laboratories on multiple hosts (Rothamsted Research website). In the near future, molecular genetic studies on host-pathogen interactions will extend to exploring sporulation rates, plant-to-plant spread, and season-to-season survival. We plan to capture these types of new data in PHI-base.

A second major obstacle to the curation of PHI-base was finding the relevant articles. For example, it was noted during curation that some journals found in the Web of Science (WOS) database are not represented in PubMed and vice versa (Fig. 1). In addition, WOS and PubMed together do not represent a comprehensive summary of all published literature. Future curation of the database will become semiautomated with the use of the text-mining framework ONDEX (Koehler et al. 2005). This will allow the precision and recall of relevant articles from literature databases to be higher than simple keyword searches. Articles identified from the novel text-mining methods are carefully curated and checked by a domain expert prior to inclusion in PHI-base to ensure the same high level of quality assurance as used in the manual approach.

The necessity for a continually updated pathogenicity gene database has been demonstrated during the annotation of EST and genomic sequences from the oomycete *Phytophthora infestans* (Randall et al. 2005). In this study, homologs of fungal pathogenicity genes were identified by sequence similarity to 104 pathogenicity genes, including the 79 plant pathogen genes cited in a review by Idnurm and Howlett (2001). However, since 2001, the number of verified pathogenicity and virulence genes for plant pathogens has tripled. A publicly available database of pathogenicity, virulence, and effector genes would be of great utility for annotating genomes and EST in studies with similar objectives to those done by Randall and associates (2005).

The COGEME database, in addition to containing EST data for many plant-pathogenic fungi, also allows researchers to de-

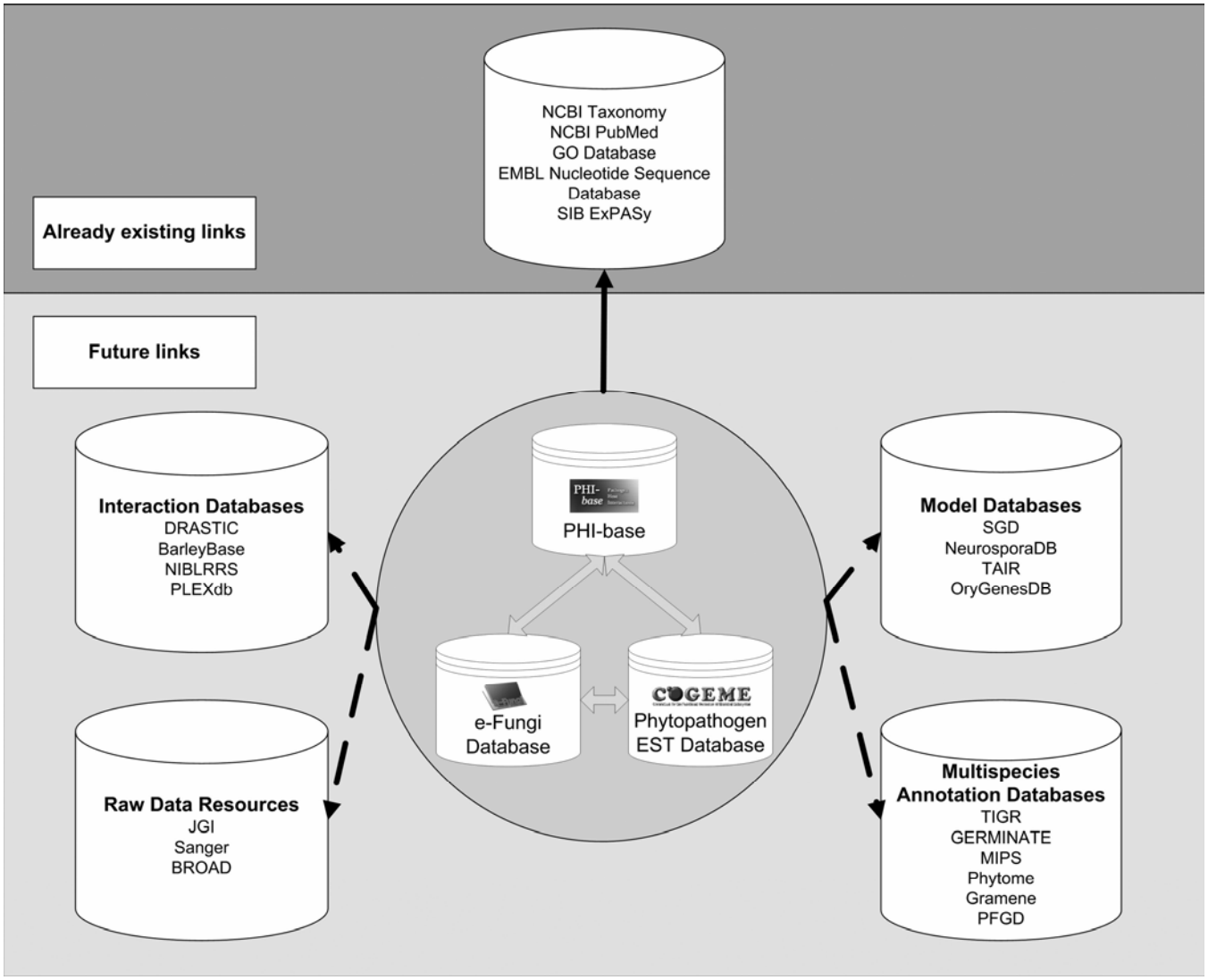

Fig. 5. External data sources to be interlinked with PHI-base in the future.

posit expression data from their studies (Both et al. 2005; Soanes 2005). It is envisaged that future curation of PHI-base could be aided by direct submissions from the fungal pathogenicity research community. This could take the form of data from gene disruption studies that are not published. Such information would be indicated as not published but may contain crucial details in comparative studies. For example, the growth rates of gene deletion mutants on a range of media are only rarely published in detail in research articles (Tonukari et al. 2000).

Eventually PHI-base will contain all genes that have been disrupted and phenotyped in all pathogenic fungi and oomycetes, irrespective of whether the disruption affects pathogenicity or not. This will be useful to researchers who wish to know whether a putative pathogenicity gene of interest has, in fact, been shown NOT to affect pathogenicity in another species. This information may reveal functional redundancy in fungi. For example, the plant cell wall–degrading enzymes called endopolygalacturonases expressed by several fungi during pathogenesis are encoded by a multigene family in fungi. The disruption of individual endopolygalacturonase genes in *Fusarium oxysporum* and *Cochliobolus carbonum* did not affect virulence, because other wild-type enzymes are retained that are able to degrade plant cell walls (Garcia-Maceira et al. 2001; Scott-Craig et al. 1990).

Over the last decade, a wealth of full-genome sequence information has become available for many organisms. Since 2003, the number of genomes available for pathogens of plant and animal hosts has steadily increased (Table 3). PHI-base already permits the downloading in FASTA format of all the sequences contained in the database. These sequences can be used to interrogate each newly sequenced genome as well as to explore the chromosomal context within which a pathogenic, virulence, and effector gene resides. Future versions of the database will allow users to BLAST search against all the nucleotide or protein sequences of the entries in the database. A batch BLAST function will permit users to query the database with many sequences. This will allow the immediate identification of pathogenicity gene homologs in studies involving large numbers of genes, e.g., in transcriptomic studies to explore gene networks.

Effective annotation and analysis during comparative genomic studies requires diverse information including genome sequence, expression data, protein function, and metabolic pathway information, as well as data from gene disruption experiments. A system to integrate such data for *S. cerevisiae*, called the Genome Information Management System has been developed and is available online (Cornell et al. 2003). Another database, called e-Fungi, will integrate sequence and similarity search information from the nonpathogenic model fungi, with data from many plant and animal pathogens as well as saprophytes. It is planned to improve interoperability of PHI-base with external data sources by providing linkouts or bidirectional links to related entries. This will cover the COGEME and e-Fungi databases as well as pathway information in KEGG (Kanehisa et al. 2002). These linkouts are described in Figure 5. Grid computing will allow data resources to be shared and computing power to be combined to create a unified resource for researchers to access and analyze data (Jacq et al. 2004).

Future versions of PHI-base will include information on host mutations that compromise or enhance host defense responses (Manger and Relman 2000). We have already compiled, from peer-reviewed publications, more than 100 *Arabidopsis thaliana* mutants and transgenic lines that modify different types of interactions (Hammond-Kosack and Parker 2003). In addition, for many plant host species, the molecular identity of various classes of disease resistance genes are now known, as well as the function of specific gene variants in the activation of plant responses (Meyers et al. 2005). The curation of this experimentally verified host literature alongside the corresponding pathogen information for many interactions should further facilitate the identification of common themes and unique components as well as when information is lacking or incomplete. In addition, plant pathogenic bacterial and viral virulence factors are currently not curated in any database; these could be included in future versions of PHI-base. Recently, we have started to include

Table 4. Column headings from database (PHI-base)

Column name and gene ontologies	Description
PHI-base accession	Stable accession number for each database entry to aid curation
EMBL accession	Link-out to EMBL Nucleotide Sequence Database
Gene name	Name of the fungal gene that was disrupted in the published study
Pathogen NCBI Taxonomy ID	NCBI taxonomy ID of the pathogenic fungus
Pathogen species	Systematic name of the pathogenic fungus
Disease name	Name of the disease caused by the pathogen host interaction
Monocot / Dicot plant	Number of cotyledons, if the host is a plant
Host NCBI Taxonomy ID	NCBI taxonomy ID of the host organism
Experimental host	Common name of the host organism
Function	Function of the disrupted gene
Linkout accession	Accession number for gene function and ontology databases
Database	Identifier for gene function and ontology databases
Pathway	Name of the pathway the disrupted gene is involved in
Phenotype GO:0009405	Pathogenicity phenotype of the fungus after gene disruption: reduced virulence, nonpathogenic or pathogenic
Mating defect	Yes/No If the gene disruption causes a mating defect affecting pathogenicity
Pre-penetration defect	Yes/No If the gene disruption causes a block in the disease process before penetration, e.g., formation of appressoria
Penetration defect GO:0044409	Yes/No If the gene disruption causes a block in the disease process at penetration
Post-penetration defect GO:0044412	Yes/No If the gene disruption causes a block in the disease process after penetration
Vegetative spores GO:0030435	Defects in asexual sporulation caused by the gene disruption e.g reduced sporulation
Sexual spores GO:0000909	Defects in sexual sporulation caused by the gene disruption e.g reduced sporulation
In vitro growth	Growth defects in culture caused by the gene disruption e.g. reduced growth
Spore germination GO:0009847	Defects in spore germination caused by the gene disruption
Essential gene	Lethal effect from gene disruption
Inducer	For cases where a particular compound is needed to induce gene expression, e.g., pectin
Host response	Details any difference in the host defence response to a pathogen with a disrupted gene
Experimental evidence	The experimental method that is applied to disrupt the gene
Entered by	Name of the curator who entered the interaction to the database
Manual or text mining	Method of information retrieval
Literature ID	Accession number for the published article
Literature source	Name of library or information resource containing the publication
Full citation	Full citation of the article if no ID is available
Author email	Email address of the author of the article
Amino acid sequence	Amino acid sequence of the gene product
Nucleotide sequence	Nucleotide sequence of the gene
Comments	Field provided for any further free text information

entries in which no defect in pathogenicity was observed when the gene was disrupted, even though, in another pathogenic species, a pathogenicity defect was observed following the disruption of a similar gene sequence.

PHI-base already represents a unique resource providing comprehensive information for pathogen species that invade plant, animal, or fungal hosts. Other complementary databases provide functional information on a single organism, for example, the rice blast websites or a small group of closely related pathogens, e.g., the Phytophthora PFGD website. We envisage typical users of this database will be medical and agricultural scientists, bioinformaticians, and evolutionary biologists, who need easy access to peer-reviewed data on multiple pathogen species from a single internet resource.

MATERIALS AND METHODS

The data in PHI-base are compiled from scientific publications and subsequently curated into PHI-base. As already discussed in the introduction, only pathogenicity genes that have been verified using 'strong' experimental evidence have been included in the database, i.e., from forward or reverse genetic approaches. The challenge is to identify the publications that describe such experiments and, then, to extract and curate all information that is important for the characterization of a given pathogenicity gene in a systematic way.

Key word search of the literature databases PubMed and WOS.

The key word search term (fung* or yeast) and (gene or factor) and (pathogenicity or virulen* or avirulence gene*) was used, and the returned articles of interest were those with information about the disruption and characterization of genes.

Curation of information.

Each pathogenicity, virulence, and effector gene found through the literature searches was initially recorded in a spreadsheet. In addition, further relevant publications were obtained from the recommendations of colleagues. Each relevant article was then manually curated by a domain expert. Up to 30 details about gene function and the phenotype of gene disruptants were also recorded. Table 4 shows how the experimental data from the papers were categorized in PHI-base. This approach revealed that, for many articles, information was either incomplete or entirely missing and that a range of infection tests were used to test the same pathogen-host interaction.

ACKNOWLEDGMENTS

All authors are grateful to G. Harrison and the Rothamsted Computing Services group for efficient provision and support with the computational infrastructure for PHI-base. The work of all authors is funded by the Biotechnology and Biological Sciences Research Council (BBSRC). We thank D. Soanes and N. Talbot for beta-testing earlier versions of PHI-base. We also thank M. Johnston, the Rothamsted Research librarian, for assisting with the keyword searches. T. K. Baldwin is supported by a BBSRC industrial CASE studentship sponsored by Syngenta. Rothamsted Research receives grant-aided support from the Biotechnology and Biological Sciences Research Council.

LITERATURE CITED

Agrios, G. N. 1997. Plant Pathology. Academic Press, Inc., London.

Allen, R. L., Bittner-Eddy, P. D., Grenville-Briggs, L. J., Meitz, J. C., Rehmany, A. P., Rose, L. E., and Beynon, J. L. 2004. Host-parasite coevolutionary conflict between *Arabidopsis* and downy mildew. Science 306:1957-1960.

Alonso-Monge, R., Navarro-Garcia, F., Molero, G., Diez-Orejas, R., Gustin, M., Pla, J., Sanchez, M., and Nombela, C. 1999. Role of the mitogen-activated protein kinase hog1p in morphogenesis and virulence of *Candida albicans*. J. Bacteriol. 181:3058-3068.

Altschul, S. F., Madden, T. L., Shaffer, A. A., 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.

Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., Davis, A. P., Dolinski, K., Dwight, S. S., Eppig, J. T., Harris, M. A., Hill, D. P., Issel-Tarver, L., Kasarskis, A., Lewis, S., Matese, J. C., Richardson, J. E., Ringwald, M., Rubin, G. M., and Sherlock, G. 2000. Gene Ontology: Tool for the unification of biology. Nat. Genet. 25:25-29.

Both, M., Eckert, S. E., Csukai, M., Müller, E., Dimopoulos, G., and Spanu, P. D. 2005. Transcript profiles of *Blumeria graminis* development during infection reveal a cluster of genes that are potential virulence determinants. Mol. Plant-Microbe Interact. 18:125-133.

Broad Institute 2005. Fungal Genome Initiative (FGI). Published online.

Catanzariti, A.-M., Dodds, P. N., Lawrence, G. J., Aycliffe, M. A., and Ellis, J. G. 2006. Haustorially expressed secreted proteins from flax rust are highly enriched for avirulence elicitors. Plant Cell 18:243-256.

Cornell, M., Paton, N. W., Hedeler, C., Kirby, P., Delneri, D., Hayes, A., and Oliver, S. G. 2003. GIMS: An integrated data storage and analysis environment for genomic and functional data. Yeast 20:1291-1306.

de Bary, H. A. 1879. The Phenomenon of Symbiosis. Strasbourg, France.

de Groot, M. J. A., Bundock, P., Hooykaas, P. J. J., and Beijersbergen, A. G. M. 1998. *Agrobacterium tumefaciens*-mediated transformation of filamentous fungi. Nat. Biotechnol. 16:839-842.

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.

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.

DeZwaan, T. M., Carroll, A. M., Valent, B., and Sweigard, J. A. 1999. *Magnaporthe grisea* Pth11p is a novel plasma membrane protein that mediates appressorium differentiation in response to inductive substrate cues. Plant Cell 11:2013-2030.

Dietrich, F. S., Voegeli, S., Brachat, S., Lerch, A., Gates, K., Steiner, S., Mohr, C., Pohlmann, R., Luedi, P., Choi, S., Wing, R. A., Flavier, A., Gaffney, T. D., and Philippsen, P. 2004. The *Ashbya gossypii* genome as a tool for mapping the ancient *Saccharomyces cerevisiae* genome. Science 304:304-307.

DiezOrejas, R., Molero, G., NavarroGarcia, F., Pla, J., Nombela, C., and SanchezPerez, M. 1997. Reduced virulence of *Candida albicans MKC1* mutants: A role for mitogen-activated protein kinase in pathogenesis. Infect. Immun. 65:833-837.

Dixon, K. P., Xu, J. R., Smirnoff, N., and Talbot, N. J. 1999. Independent signaling pathways regulate cellular turgor during hyperosmotic stress and appressorium-mediated plant infection by *Magnaporthe grisea*. Plant Cell 11:2045-2058.

Dodds, P. N., Lawrence G. J., Catanzariti, A.-M., Ayliffe, M. A., and Ellis, J. G. 2004. The *Melampsora lini AvrL567* avirulence genes are expressed in haustoria and their products are recognised inside plant cells. Plant Cell 16:755-768.

Dufresne, M., Bailey, J. A., Dron, M., and Langin, T. 1998. *clk1*, a serine/threonine protein kinase-encoding gene, is involved in pathogenicity of *Colletotrichum lindemuthianum* on common bean. Mol. Plant-Microbe Interact. 11:99-108.

Feldbrugge, M., Kamper, J., Steinberg, G., and Kahmann, R. 2004. Regulation of mating and pathogenic development in *Ustilago maydis*. Curr. Opin. Microbiol. 7:666-672.

Fellbrich, G., Romanski, A., Varet, A., Blume, B., Brunner, F., Engelhardt, S., Felix, G., Kemmerling, B., Krzymowska, M., and Nurnberger, T. 2002. NPP1, a *Phytophthora*-associated trigger of plant defense in parsley and *Arabidopsis*. Plant J. 32:375-390.

Galagan, J. E., Calvo, S. E., Borkovich, K. A., Selker, E. U., Read, N. D., et al. 2003. The genome sequence of the filamentous fungus *Neurospora crassa*. Nature 422:859-868.

Garcia-Maceira, F. I., Di Pietro, A., Huertas-Gonzalez, M. D., Ruiz-Roldan, M. C., and Roncero, M. I. G. 2001. Molecular characterization of an endopolygalacturonase from *Fusarium oxysporum* expressed during early stages of infection. Appl. Environ. Microbiol. 67:2191-2196.

Gattiker, A., Michoud, K., Rivoire, C., Auchincloss, A. H., Coudert, E., Lima, T., Kersey, P., Pagni, M., Sigrist, C. J. A., Lachaize, C., Veuthey, A. L., Gasteiger, E., and Bairoch, A. 2003. Automated annotation of microbial proteomes in SWISS-PROT. Comput Biol Chem 27:49-58.

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 274:546-566.

Hammond-Kosack, K. E., and Parker, J. E. 2003. Deciphering plant-pathogen communication: Fresh perspectives for molecular resistance breeding. Curr. Opin. Biotechnol. 14:177-193.

Hou, Z., Xue, C., Peng, Y., Katan, T., Kistler, H. C., and Xu, J.-R. 2002. A mitogen-activated protein kinase gene *(MGV1)* in *Fusarium graminearum* is required for female fertility, heterokaryon formation, and plant infection. Mol. Plant-Microbe Interact. 15:1119-1127.

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

Jacq, N., Blanchet, C., Combet, C., Cornillot, E., Duret, L., Kurata, K., Nakamura, H., Silvestre, T., and Breton, V. 2004. Grid as a bioinformatic tool. Parallel Comput. 30:1093-1107.

Johnson, R. D., Johnson, L., Itoh, Y., Kodama, M., Otani, H., and Kohmoto, K. 2000. Cloning and characterization of a cyclic peptide synthetase gene from *Alternaria alternata* apple pathotype whose product is involved in AM-toxin synthesis and pathogenicity. Mol. Plant-Microbe Interact. 13:742-753.

Jones, T., Federspiel, N. A., Chibana, H., Dungan, J., Kalman, S., Magee, B. B., Newport, G., Thorstenson, Y. R., Agabian, N., Magee, P. T., Davis, R. W., and Scherer, S. 2004. The diploid genome sequence of *Candida albicans*. Proc. Natl. Acad. Sci. U.S.A. 101:7329-7334.

Kamoun, S., van West, P., Vleeshouwers, V. G., de Groot, K. E., and Govers, F. 1998. Resistance of *Nicotiana benthamiana* to *Phytophthora infestans* is mediated by the recognition of the elicitor protein INF1. Plant Cell 10:1413-1426.

Kanehisa, M., Goto, S., Kawashima, S., and Nakaya, A. 2002. The KEGG databases at GenomeNet. Nucleic Acids Res. 30:42-46.

Keon, J., Antoniw, J., Rudd, J., Skinner, W., Hargreaves, J., and Hammond-Kosack, K. 2005a. Analysis of expressed sequence tags from the wheat leaf blotch pathogen *Mycosphaerella graminicola* (anamorph *Septoria tritici*). Fungal Genet. Biol. 42:376-389.

Keon, J., Rudd, J., Antoniw, J., Skinner, W., Hargreaves, J., and Hammond-Kosack, K. 2005b. Metabolic and stress adaptation by *Mycosphaerella graminicola* during sporulation in its host revealed through microarray transcription profiling. Mol. Plant Pathol. 6:527-540.

Klimpel, A., Gronover, C. S., Williamson, B., Stewart, J. A., and Tudzynski, B. 2002. The adenylate cyclase (BAC) in *Botrytis cinerea* is required for full pathogenicity. Mol. Plant Pathol. 3:439-450.

Knogge, W. 1996. Fungal infections of plants. Plant Cell 8:1711-1722.

Koehler, J., Rawlings, C., Verrier, P., Mitchell, R., Skusa, A., Ruegg, A., and Philippi, S. 2005. Linking experimental results, biological networks and sequence analysis methods using ontologies and generalised data structures. *In Silico* Biol. 5:005.

Kojima, K., Kikuchi, T., Takano, Y., Oshiro, E., and Okuno, T. 2002. The mitogen-activated protein kinase gene *MAF1* is essential for the early differentiation phase of appressorium formation in *Colletotrichum lagenarium*. Mol. Plant-Microbe Interact. 15:1268-1276.

Kraus, P. R., Boily, M. J., Giles, S. S., Stajich, J. E., Allen, A., Cox, G. M., Dietrich, F. S., Perfect, J. R., and Heitman, J. 2004. Identification of *Cryptococcus neoformans* temperature-regulated genes with a genomic-DNA microarray. Eukaryot. Cell 3:1249-1260.

Kraus, P. R., Fox, D. S., Cox, G. M., and Heitman, J. 2003. The *Cryptococcus neoformans* MAP kinase Mpk1 regulates cell integrity in response to antifungal drugs and loss of calcineurin function. Mol. Microbiol. 48:1377-1387.

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

Lashkari, D. A., McCusker, J. H., and Davis, R. W. 1997b. Whole genome analysis: Experimental access to all genome sequenced segments through larger-scale efficient oligonucleotide synthesis and PCR. Proc. Natl. Acad. Sci. U.S.A. 94:8945-8947.

Lee, C. M., Nantel, A., Jiang, L. H., Whiteway, M., and Shen, S. H. 2004. The serine/threonine protein phosphatase SIT4 modulates yeast- to-hypha morphogenesis and virulence in *Candida albicans*. Mol. Microbiol. 51:691-709.

Lindeberg, M., Stavrinides, J., Chang, J. H., Alfano, J. R., Collmer, A., Dangl, J. L., Greenberg, J. T., Mansfield, J. W., and Guttman, D. S. 2005. Proposed guidelines for a unified nomenclature and phylogenetic analysis of type III Hop effector proteins in the plant pathogen *Pseudomonas syringae*. Mol Plant-Microbe Interact. 18:275-282.

Manger, I. D., and Relman, D. A. 2000. How the host 'sees' pathogens: Global gene expression responses to infection. Curr. Opin. Immunol. 12:215-218.

McDonald, T., Brown, D., Keller, N. P., and Hammond, T. M. 2005. RNA silencing of mycotoxin production in *Aspergillus* and *Fusarium* species. Mol. Plant-Microbe Interact. 18:539-545.

Meyers, B. C., Kaushik, S., and Nandety, R. S. 2005. Evolving disease resistance genes. Curr. Opin. Plant Biol. 8:129-134.

Nierman, W. C., Pain, A., Anderson, M. J., Wortman, J. R., Kim, H. S., Arroyo, J., Berriman, M., Abe, K., Archer, D. B., Bermejo, C., Bennett, J., Bowyer, P., Chen, D., Collins, M., Coulsen, R., Davies, R., Dyer, P. S., Farman, M., Fedorova, N., Fedorova, N., Feldblyum, T. V., Fischer, R., Fosker, N., Fraser, A., Garcia, J. L., Garcia, M. J., Goble, A., Goldman, G. H., Gomi, K., Griffith-Jones, S., Gwilliam, R., Haas, B., Haas, H., Harris, D., Horiuchi, H., Huang, J., Humphray, S., Jimenez, J., Keller, N., Khouri, H., Kitamoto, K., Kobayashi, T., Konzack, S., Kulkarni, R., Kumagai, T., Lafton, A., Latge, J. P., Li, W. X., Lord, A., Majoros, W. H., May, G. S., Miller, B. L., Mohamoud, Y., Molina, M., Monod, M., Mouyna, I., Mulligan, S., Murphy, L., O'Neil, S., Paulsen, I., Penalva, M. A., Pertea, M., Price, C., Pritchard, B. L., Quail, M. A., Rabbinowitsch, E., Rawlins, N., Rajandream, M. A., Reichard, U., Renauld, H., Robson, G. D., de Cordoba, S. R., Rodriguez-Pena, J. M., Ronning, C. M., Rutter, S., Salzberg, S. L., Sanchez, M., Sanchez-Ferrero, J. C., Saunders, D., Seeger, K., Squares, R., Squares, S., Takeuchi, M., Tekaia, F., Turner, G., de Aldana, C. R. V., Weidman, J., White, O., Woodward, J., Yu, J. H., Fraser, C., Galagan, J. E., Asai, K., Machida, M., Hall, N., Barrell, B., and Denning, D. W. 2005. Genomic sequence of the pathogenic and allergenic filamentous fungus *Aspergillus fumigatus*. Nature 438:1151-1156.

Ortoneda, M., Guarro, J., Madrid, M. P., Caracuel, Z., Roncero, M. I. G., Mayayo, E., and Di Pietro, A. 2004. *Fusarium oxysporum* as a multihost model for the genetic dissection of fungal virulence in plants and mammals. Infect. Immun. 72:1760-1766.

Panaccione, D. G., Scottcraig, 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.

Randall, T. A., Dwyer, R. A., Huitema, E., Beyer, K., Cvitanich, Kelkar, C., H., Ah Fong, A. M. V., Gates, K., Roberts, S., Yatzkan, E., Gaffney, T., Law, M., Testa, A., Torto-Alalibo, T., Zhang, M., Zheng, L., Mueller, E., Windass, J., Binder, A., Birch, P. R. J., Gisi, U., Govers, F., Gow, N. A., Mauch, F., van West, P., Waugh, M. E., Yu, J., Boller, T., Kamoun, S., Lam, S. T., and Judelson, H. S. 2005. Large-scale gene discovery in the Oomycete *Phytophthora infestans* reveals likely components of phytopathogenicity shared with true fungi. Mol. Plant-Microbe Interact. 18:229-243.

Rohe, M., Gierlich, A., Hermann, H., Hahn, M., Schmidt, B., Rosahl, S., and Knogge, W. 1995. The race-specific elicitor, NIP1, from the barley pathogen, *Rhynchosporium secalis*, determines avirulence on host plants of the *Rrs1* resistance genotype. EMBO (Eur. Mol. Biol. Organ.) J. 14:4168-4177.

Ruepp, A., Zollner, A., Maier, D., Albermann, K., Hani, J., Mokrejs, M., Tetko, I., Guldener, U., Mannhaupt, G., Munsterkotter, M., and Mewes, H. W. 2004. The FunCat, a functional annotation scheme for systematic classification of proteins from whole genomes. Nucleic Acids Res. 32:5539-5545.

Sanger Institute 2005. Fungal genomes. Published online.

Scott-Craig, J. S., Panaccione, D. G., Cervone, F., and Walton, J. D. 1990. Endopolygalacturonase is not required for pathogenicity of *Cochliobolus carbonum* on maize. Plant Cell 2:1191-1200.

Shaner, G., Stromberg, E. L., Lacy, G. H., Barker, K. R., and Pirone, T. P. 1992. Nomenclature and concepts of pathogenicity and virulence. Annu. Rev. Phytopathol. 30:47-66.

Soanes, D. M. 2005. Phytopathogenic Fungi and Oomycete EST Database (COGEME). Published online.

Soanes, D. M., Skinner, W., Keon, J., Hargreaves, J., and Talbot, N. J. 2002. Genomics of phytopathogenic fungi and the development of bioinformatic resources. Mol. Plant-Microbe Interact. 15:421-427.

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.

Takano, Y., Choi, W. B., Mitchell, T. K., Okuno, T., and Dean, R. A. 2003. Large scale parallel analysis of gene expression during infection-related morphogenesis of *Magnaporthe grisea*. Mol. Plant Pathol. 4:337-346.

Talbot, N. J. 2004. Annual plant reviews. Page 306. In: Plant-Pathogen Interactions. Blackwell Publishing Ltd., Carlton, Australia.

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

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 12:237-247.

Torto, T. A., Li, S. A., Styer, A., Huitema, E., Testa, A., Gow, N. A. R., van West, P., and Kamoun, S. 2003. EST mining and functional expression

assays identify extracellular effector proteins from the plant pathogen *Phytophthora*. Genome Res. 13:1675-1685.
Urban, M., Bhargava, T., and Hamer, J. E. 1999. An ATP-driven efflux pump is a novel pathogenicity factor in rice blast disease. EMBO (Eur. Mol. Biol. Organ.) J. 18:512-521.
van Dijk, K., Fouts, D. E., Rehm, A. H., Hill, A. R., Collmer, A., and Alfano, J. R. 1999. The Avr (effector) proteins HrmA (HopPsyA) and AvrPto are secreted in culture from *Pseudomonas syringae* pathovars via the Hrp (type III) protein secretion system in a temperature- and pH-sensitive manner. J. Bacteriol. 181:4790-4797.
van Kan, J. A. L., van den Ackerveken, G. F. J. M., and De Wit, P. J. G. M. 1991. Cloning and characterization of cDNA of avirulence gene *AVR9* of the fungal pathogen *Cladosporium fulvum*, causal agent of tomato leaf mold. Mol. Plant-Microbe Interact. 4:52-59.
Viaud, M., Brunet-Simon, A., Brygoo, Y., Pradier, J. M., and Levis, C. 2003. Cyclophilin A and calcineurin functions investigated by gene inactivation, cyclosporin A inhibition and cDNA arrays approaches in the phytopathogenic fungus *Botrytis cinerea*. Mol. Microbiol. 50:1451-1465.
Viaud, M. C., Balhadere, P. V., and Talbot, N. J. 2002. A *Magnaporthe grisea* cyclophilin acts as a virulence determinant during plant infection. Plant Cell 14:917-930.
Wang, P., Cardenas, M. E., Cox, G. M., Perfect, J. R., and Heitman, J. 2001. Two cyclophilin A homologs with shared and distinct functions important for growth and virulence of *Cryptococcus neoformans*. EMBO (Eur. Mol. Biol. Organ.) Rep. 2:511-518.
Wasmann, C. C., and VanEtten, H. D. 1996. Transformation-mediated chromosome loss and disruption of a gene for pisatin demethylase decrease the virulence of *Nectria haematococca* on pea. Mol. Plant-Microbe Interact. 9:793-803.
Winnenburg, R., Baldwin, T. K., Urban, M., Rawlings, C., Kohler, J., and Hammond-Kosack, K. 2005. PHI-base: A new database for pathogen host interactions. Nucleic Acids Res. 34:D459-464.
Xu, J. R. 2000. MAP kinases in fungal pathogens. Fungal Genet. Biol. 31:137-152.
Xu, J. R., Staiger, C. J., and Hamer, J. E. 1998. Inactivation 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.

AUTHOR-RECOMMENDED INTERNET RESOURCES

Broad Institute Fungal Genome Initiative: www.broad.mit.edu/annotation
Center for Bioinformatics, Bordeaux, Genomic Exploration of the Hemiascomycete Yeasts website: cbi.labri.fr/Genolevures
The Consortium for the Functional Genomics of Microbial Eukaryotes (COGEME) database: cogeme.ex.ac.uk
e-Fungi: www.cs.man.ac.uk/~cornell/eFungi/ProjectDetails.html
The Genome Information Management System (GIMS): www.cs.man.ac.uk/img/gims
Institut Pasteur CandidaDB world-wide web server: genolist.pasteur.fr/CandidaDB
The International Rice Blast Genome Consortium websites: www.riceblast.org
Joint Genome Institute website: genome.jgi-psf.org
Magnaportha grisea Oryza sativa interaction database (MGOS): www.MGOSdb.org
National Center for Biotechnology Information (NCBI) GenBank database: www.ncbi.nlm.nih.gov/Genbank/index.html
National Center for Genome Resources website: www.pfgd.org
The Pathogen-Host Interactions database (PHI-base): www.phi-base.org
The Phytophthora functional genomics database (PFGD): www.pfgd.org/pfgd
Plant-Associated Microbe Gene Ontology Interest Group (PAMGO) website: pamgo.vbi.vt.edu
Rothamsted Research website: www.rothamsted.bbsrc.ac.uk/ppi/staff/khk.html
Sanger Institute: www.sanger.ac.uk/Projects
Swiss Institute of Bioinformatics Ashbya genome database: agd.unibas.ch
University of Manchester Central Aspergillus Data Repository (CADRE): www.cadre.man.ac.uk