The Internet-Based Fungal Pathogen Database: A Proposed Model

Seogchan Kang, John E. Ayers, Erick D. DeWolf, David M. Geiser, Gretchen Kuldau,
Gary W. Moorman, Ewen Mullins, Wakar Uddin, James C. Correll, Gerald Deckert,
Yong-Hwan Lee, Yin-Won Lee, Frank N. Martin, and Krishna Subbarao

First to eighth authors: Department of Plant Pathology, The Pennsylvania State University, University Park 16802; ninth author: Department of Plant Pathology, University of Arkansas, Fayetteville 72701; tenth author: Codex Bioinformatics Services, 5074 Baltimore St., San Diego, CA 92117; eleventh and twelfth authors: School of Agricultural Biotechnology, Seoul National University, Suwon 441-744, Korea; thirteenth author: USDA-ARS, 1636 E. Alisal St., Salinas, CA 93905; and fourteenth author: Department of Plant Pathology, University of California, USDA Station, 1636 E. Alisal St., Salinas 93905.

Accepted for publication 3 December 2001.

ABSTRACT

Kang, S., Ayers, J. E., DeWolf, E. D., Geiser, D. M., Kuldau, G., Moorman, G. W., Mullins, E., Uddin, W., Correll, J. C., Deckert, G., Lee, Y.-H., Lee, Y.-W., Martin, F. N., and Subbarao, K. 2002. The internet-based fungal pathogen database: A proposed model. *Phytopathology* 92:232-236.

A better understanding of the phenotypic and genetic diversity of significant agricultural pathogens and how their populations change in the field is critical for designing successful, long-term disease management strategies. Although efforts to determine the genetic diversity of plant pathogens have substantially increased in recent years, mainly due to the availability of various molecular tools, complementary efforts to archive and integrate the resulting data have been very limited. As a

consequence, it is often difficult to compare the available data from various laboratories because the data have been generated by diverse tools, often preventing any direct comparisons, and are saved in a format that is unsuitable for comparative studies. The establishment of an internet-based database that cross-links the digitized genotypic and phenotypic information of individual pathogens at both the species and population levels may allow us to effectively address these problems by coordinating the generation of data and its subsequent archiving. We discuss the needs, benefits, and potential structure of such a database.

Additional keywords: amplified fragment length polymorphism, *Fusarium*, *Magnaporthe grisea*, microsatellites, mycotoxins.

At present, the genomes of more than 100 microorganisms have been sequenced or sequencing projects are underway. Although plant pathogens have been underrepresented in this group, with only two plant pathogenic bacterial genomes having been completely sequenced (17,34), the situation is improving (16). The genomics of plant pathogens will significantly further our understanding of pathogenicity mechanisms, which will in turn facilitate the identification of novel strategies for disease control. By practical necessity, genomics of individual species will focus on one or a very few strains of each species. However, because crop losses are caused by populations of strains that vary in such traits as virulence, host range, chemical resistance, and toxin production, knowing the complete genetic makeup of a single individual tells us nothing about the degree of variation of such traits within a given species. The full potential of microbial genomics as a foundation for the development of novel disease control strategies will hinge on building a bridge from the genome of a single individual to the variation within a whole species. To make this a reality, we need to generate a comprehensive picture of the genetic and phenotypic diversity in individual pathogen species.

In addition, considering the magnitude of uncharacterized microbial diversity in nature (15,32), we cannot understate the importance of systematically cataloging the taxonomic information in a format that can be shared and accessed easily by the whole community of scientists. The emergence of new pathogens through hybridization (4), global migration, and accidental release by expanding agricultural activities also underlies the importance

of such efforts. A better understanding of the structure and dynamics of pathogen species/populations in the agroecosystem will potentially (i) guide the selection of resistance germ plasm for genetic engineering and breeding programs, (ii) assist in deploying crops in a way that minimizes the evolution of new races and hybrid species, (iii) allow us to monitor the build-up of virulent races or strains resistant to present day crop protectants, and (iv) trace the movement of economically important organisms in agricultural production systems.

The availability of a large array of molecular tools (20,22,35) has rapidly expanded our knowledge of plant pathogen systematics and population genetics (7,14,18,25,29,31,33,38,40,42). However, much of the resulting data is only available in the literature or in the laboratories of individual investigators, therefore making it arduous to combine the data into any comparative study. To maximize the value of this type of data, it is essential that better methods of data storage and sharing be devised. We believe that the utilization of an internet-based database (or a network of databases) will significantly facilitate these activities. In this letter, we outline (i) the importance and benefits of establishing a database that will cross-link the available genotypic and phenotypic information of major pathogens at multiple phylogenetic levels, ranging from strains to genera, and will also allow easy sharing and updating of such data and (ii) the potential structure and management strategies of such a database.

Importance of cataloging the genetic relationship among fungal pathogens at both species and population levels. Fungal taxa are traditionally defined based on morphological characters, requiring an in-depth knowledge of taxonomy. Morphological characteristics have limited resolving power in plant pathogenic fungi because they often do not express a high level of variation in these characters. In the last several years, it has become in-

Corresponding author: S. Kang; E-mail address: sxk55@psu.edu

Publication no. P-2002-0123-01O

© 2002 The American Phytopathological Society

creasingly apparent that fungal pathogens classified by morphology are far more diverse than previously thought (1,13,25,28,30,31). DNA sequence analysis of known genes, particularly portions of the ribosomal RNA genes (rDNA), has emerged as a powerful tool for identifying and defining fungal species (5,26,27). Phylogenetic studies of fungi using rDNA sequences are now performed routinely, although in recent years it has been discovered that their utility seems to end at the level of closely related species. More recently, conserved protein-coding genes such as β -tubulin and translation elongation factor-1 α that contain highly variable introns have allowed mycologists to break through this barrier, particularly in the genus *Fusarium* (1,11–13,25,28–31,39). Identifying a dozen or more genes to study the phylogenetics of a particular pathogen species is now possible (8,13). Particularly in combination, these markers offer a powerful means not only for taxonomic identification, but also for defining new species and studying the evolutionary history of fungi (1,7,13,28,31,35,36).

An ill-defined or changing classification of fungal species and genera has led to a somewhat confused picture of which species are pathogenic to a crop of interest or, as in the case of the genus *Fusarium*, are potentially mycotoxigenic. Much of the past work on mycotoxin production by *Fusarium* spp. was done in the context of an outmoded classification system. As a result, the mycotoxigenic capacity of *Fusarium* spp., as they are now described, is confusing. The last comprehensive cataloging of *Fusarium* mycotoxin production (21) revealed many intriguing situations regarding mycotoxigenic potential such as the finding that isolates of a species from Japan produce a toxin that is not produced by the same species from other locations. Such information hints at the likelihood of cryptic speciation (13,36) and beckons for an in-depth phylogenetic analysis of these isolates to be completed using genetic markers. Indeed, we know that important, traditionally defined *Fusarium* spp. such as *F. moniliforme*, *F. oxysporum*, *F. solani*, and *F. graminearum* represent complexes comprising as many as 50 distinct phylogenetic lineages, some showing continental biogeographic structure (1,2,25,28–31). A reevaluation of the taxonomic significance of such traits of agricultural importance as toxin production, pathogenicity and host range in individual species/populations within the framework of their evolutionary relationships can significantly advance our understanding of the evolution and variation of these traits.

For pathogenic fungi, the identity of a particular isolate at the species level is often not sufficient to predict its pathogenic and toxigenic potential because these traits are often variable within populations. Genotyping tools that determine the genetic relationships of individual strains within a species are needed to make connections between the genotype and phenotype of interest. Such an analysis will also further our understanding of the mechanisms by which pathogen populations change in response to host/environmental selection pressures. The global population genetic analysis of *Magnaporthe grisea* isolates from rice well illustrates how this type of analysis can contribute to controlling diseases.

Conventional breeding efforts for blast resistance, focusing on the introduction of disease resistance genes that confer incompatibility to the races identified in a given geographic area, have had only limited success, mainly due to the frequent appearance of novel, more virulent races in the field (43). Race analysis, based on its compatibility to a set of eight test varieties, has not been sufficient to determine how such races have emerged. Because more than one avirulence gene can control the compatibility to a specific variety, a similar pattern of compatibility to the test varieties does not necessarily indicate a close genetic relationship. However, when the race analysis was combined with the population genetic analysis using the fingerprints generated with a repetitive DNA probe, MGR586 (20), many new insights into its population structure and dynamics at both the genetic and race levels emerged (37,42). In most countries, rice pathogens consist primarily of a few clonally (asexually) derived lineages that are

genetically distinct, suggesting that race instability is unlikely to have been caused by frequent genetic recombination events. A comparison of the MGR586 fingerprints of new races with those of previously characterized genetic lineages has permitted the identification of their origins (9). Based on such insights, a new approach to breeding durable resistance, termed lineage exclusion, was developed (44).

Another important reason for cataloging the species and population diversity of pathogens using genotypes is its value as a tool for tracking the movement of pathogens and determining the origin of newly emerging pathogens. Because many pathogens can migrate from one region to another through various means, including shipment on agricultural products, creation of a mechanism that would allow for the rapid analysis of the taxonomic and population identity of newly identified pathogens is imperative in order to effectively diminish their impact. For instance, in the ornamentals industry where pests and pathogens may accompany plant material shipped throughout the world, the ability to trace pathogens from a shipment to its geographic origin will allow the producer to attack the problem at its source. Given the limited resolving power of phenotypic characters for strain and species identification, the use of genetic markers for strain identification is therefore critical for the speedy implementation of suitable control and regulatory measures. For example, the first significant outbreak of head scab of wheat in Australia caused by *F. graminearum* occurred in late 1999 (6). It is known that the majority of head scab in the Northern Hemisphere is caused by two of the seven major genetic lineages of *F. graminearum*, known as lineages 6 and 7 (30). Prior to the 1999 head scab epidemic, strains of *F. graminearum* native to Australia belonged to a unique lineage, suggesting that this lineage is a likely source for the epidemic. However, alternative possibilities, such as that it might have been started by the primarily North American lineage 7 and transported on infected seed or perhaps by a previously uncharacterized lineage, remain open. Comparison of the genotypes of contemporary Australian isolates with known genetic lineages of *F. graminearum* should reveal what lineage caused this new outbreak in Australia.

Community-wide cooperation is essential for systematically cataloging the genotypic and phenotypic diversity of pathogens. Although results on the population structure of fungal pathogens at the genetic level are rapidly accumulating, efforts to catalog and integrate these data into a format that can be easily shared and updated by a community of scientists studying fungal pathogens have been limited. Such efforts are urgently needed for several reasons. To begin with, we run the risk of losing most of the existing data because they are stored in individual investigators' laboratories, making them inaccessible and easily lost. Even when genotypic data are in the public domain, they are usually published as a summary, which is not conducive for comparative studies. The value of such data as a basis for developing effective disease control will only increase as all the available and newly generated data for individual pathogens becomes integrated, thereby allowing the regional and global distributions to be mapped.

However, the genotype data alone will not directly facilitate our efforts to control the target pathogens. The full potential of a comprehensive catalog of a pathogen's genetic diversity can only be realized when combined with phenotypic data. For example, the information on the genetic diversity of a pathogen population alone is not very useful per se with regard to breeding durable resistance into its host. However, this information, coupled with knowledge on the diversity and variation of virulence, chemical resistance, and host range within individual genetic groups, not only will direct the identification of the most effective means of disease control but will also assist in characterizing the genetic mechanisms by which new races evolve. For toxigenic *Fusarium* spp., it may be possible to assess the risk of mycotoxin accumulation in a particular commodity based on the identification of

locally present strains using their phylogenetic relationships to strains with known mycotoxigenic potential.

Another benefit of such a cataloging system is the efficient use of limited resources for maintaining new species and strains of under-represented species in pre-existing culture collection facilities. Although the importance of preserving microbial germ plasm is well recognized, resources for supporting such efforts are scarce (24). Given the limited support, coupled with the necessity to study many pathogens of economic significance, it is critical to establish a mechanism whereby the duplication of research is minimized. This can be achieved through easy access to the available data. The magnitude of microbial diversity that remains to be explored (15,32) also underscores the importance of systematically cataloging the existing data in a format that can be easily shared, accessed, and amended.

Cataloging population data is certainly more challenging than archiving taxonomic data and therefore calls for a coordinated effort. Because multiple markers can be used for the population genetic analysis, even when data from multiple laboratories are available, the direct comparison and integration of these data are often difficult. Even when the same markers were used in different studies, comparisons are difficult unless the data have been generated and stored in a format that is conducive for comparison. The population genetic data for *M. grisea* illustrate some of these problems. Although thousands of *M. grisea* strains from many rice-growing countries throughout the world have been subjected to MGR586 fingerprinting and race analysis (42), integration of these data to build a global picture of the pathogen's genetic and phenotypic diversity has not yet fully materialized. This is in part due to the difficulty in comparing MGR fingerprints generated in different laboratories. The MGR586 probe typically hybridizes to 50 or more *EcoRI* fragments, ranging in size from 0.7 to 20 kb, in rice pathogens (20). The genetic relatedness of individual strains is estimated from pair-wise comparisons between their fingerprints using a binary scoring system ("1" for presence and "0" for absence at each band position) and Nei and Li's index of genetic similarity (23). Based on this analysis, these strains are grouped into genetic lineages in which isolates within individual lineages exhibit 70 to 90% similarity in their fingerprints. Although this technique is robust (19) and has been highly informative in elucidating the population structure and dynamics of the fungus, it is not well suited for archiving and comparing the resulting data. Because of the large number of bands and their large sizes, direct comparison of the data between experiments and laboratories is often difficult and inaccurate especially when the degree of population complexity is high. For an accurate measurement of similarity, those isolates of interest need to be reanalyzed on the same gel (19).

Given the potential problems in cataloging the population genetic data in a format that will permit easy access and sharing, it is essential that scientists studying individual species coordinate their research efforts. We believe that the utilization of an internet-based database will facilitate these efforts and will minimize the loss of valuable data and the duplication of existing data. Genotyping pathogens at the species and subspecies levels can be easily streamlined using high throughput genotyping tools, and consequently a comprehensive database linking the genotype of individual strains to any available phenotypic information can be constructed with minimal human resources in a relatively short period of time. In addition, once phenotypic data of individual strains are permanently linked via their genotypes, the data cannot be lost even as concepts of taxonomy evolve.

Many different types of biological databases that allow fast, user-friendly access of cataloged data are currently available on the internet. Those containing sequence data, such as GenBank in the National Center for Biotechnology Information (NCBI), have become indispensable research tools. The list of internet-based databases for the taxonomic and biodiversity information has been

steadily growing, and international efforts to link these databases are rapidly progressing (3,10). However, most of them focus on species or higher taxonomic levels and lack information about the genetic and phenotypic diversity below the species level. One exception is RiceBlastDB. This database (created and maintained at Cornell University by S. McCouch and colleagues) integrates a diverse array of information about *M. grisea*, ranging from DNA clones to population structure based on MGR586 fingerprints and could potentially serve as a model for building databases for different pathogens.

The main purpose of this letter is to initiate discussions as to how to build and manage a database system for plant pathogens among members of The American Phytopathological Society and other scientific organizations. Although we propose that the database structure described below will serve as the starting point for such discussions, it is not our intention to set any standards for the pathogen database. However, we would like to emphasize the importance of coordinating database-building processes so that any resulting databases of different pathogens can be easily integrated and upgraded. Considering the significance of the community-wide contribution of data for the success of any databases, we also cannot overstate the importance of community-wide consensus on their structure and management.

Structure and management of the pathogen database. The structure of the pathogen database would consist of the following components: UNIX-based workstation, web server software, relational database engine, static HTML pages, custom Perl CGI scripts, and associated files and analysis tools. The system's website will provide interactive access to all of the collected data, and will serve as the primary interface both for public access and for the curator of the database. Users will connect to the web server using a web browser, where they will have access to the relational database or the analysis tools and browse through the records for each isolate in a number of ways: including name, phylogenetic position, geographical location, toxin, hosts, vegetative compatibility group (VCG), strain sources, etc. On the system's home page, an option will be available to submit newly recorded data and if selected, the user will access a submission sheet containing specific guidelines on the entry and coding of data. For example, it would be desirable if the geographic location where the isolate was collected is entered in degrees of latitude and longitude to avoid any confusion as to its origin. Upon completion, the submitted record will be directed through a "clearinghouse" where the curator can verify and format the data prior to it being admitted into the database.

As discussed above, it is essential to make the database searchable using one or more genotypes as a query. For the genus *Fusarium*, sequences of known genes have been extensively used for phylogenetic analysis (2,25,28–31). Therefore, these sequence data will serve as the main anchor for linking any available information for individual strains. To facilitate the use of the database and to standardize the data submission, users will have access to a number of tools for analysis on the site. First, they will be able to download or browse instructions for techniques for conducting genotype analysis. Maps of the gene regions to be used and primer sequences will be included. The BLASTN program will allow users to input a sequence in FASTA format and receive the best matches in the database. In addition to the conventional output from the BLAST search (i.e., a list of matches with sequence alignments), the best matches will be highlighted on a phylogenetic tree of the relevant species complex, with each strain name hyperlinked to other information. With advances in relational database software, this facility can be dynamic in nature and can be reviewed on a regular basis to ensure that the phylogenetic relationships of the species are consistently updated as new data is submitted.

Because the DNA sequence polymorphism in known genes may not resolve the phylogenetic relationship of individual strains in a

population that has recently diverged from a common ancestor, different markers with higher resolution would be required to accomplish the task. Strains of *M. grisea* from rice appear to be such a case. Although they can be separated into multiple genetic lineages based on MGR586 fingerprints, sequence polymorphisms that are phylogenetically informative appear very limited in known genes, probably due to a short evolutionary history for this host-specific form (M. Milgroom at Cornell University, *personal communication*). In *F. oxysporum*, races within a VCG could not be resolved using the combined sequences of two genes while the amplified fragment length polymorphism (AFLP) analysis differentiated them (2).

An array of markers (22,35), including isozyme polymorphism, electrophoretic karyotype, VCG, restriction fragment length polymorphism, randomly amplified polymorphic DNA, microsatellite, and AFLP, have been utilized to track multiple, variable loci to determine the degree of genetic diversity and evolutionary relationships at the intraspecific level. Among these, AFLP (41) appears suitable for high-resolution genotyping of fungi for a number of reasons. This technique, which involves the selective amplification of specific restriction fragments by polymerase chain reaction, is high throughput and robust and only requires a small amount of DNA. When AFLP fingerprints are generated using fluorescent labeled primers, the genotyping of >100 strains per day should be routine with a single automated DNA sequencer. Although scoring of the resulting bands is also based on their sizes, the size determination of AFLP fragments should be much more accurate than such markers as MGR fingerprints in part because they are separated on a sequencing gel, which can identify DNA fragments differing by a single base pair. Additionally, AFLP bands to be scored are much smaller than 1 kb in length, and the size standard is included in individual lanes, which minimizes variation from lane to lane. Because AFLP data are already in a digitized format when analyzed by an automated DNA sequencer, their storage and comparison become expedited. Microsatellites also share many of the advantages of AFLP and could serve as an alternative marker. Although, unlike AFLP, a specific set of microsatellite markers must be developed for each species, the rapid progress in genomic sequencing should facilitate their development. Accumulating sequence data should also facilitate exploring the sequence variation in intergenic regions as a means for differentiating closely related fungal strains. Such regions, which are likely to be more variable than transcribed regions of the genome, warrant serious consideration as markers for high-resolution genotyping.

The choice of markers for population analysis may vary among pathogens and will require a preliminary evaluation. Once a suitable marker is established, detailed protocols should be available for generating genotypes and formatting the results to facilitate the use of the database. However, to ensure that a uniform approach is taken with respect to the generation of genotype data, it would be desirable if a selected group of laboratories were able to consistently perform such an analysis. This will mean that researchers simply submit DNA to one of the centers that will then furnish the generated result to the researchers and to the database curator. Another benefit of this approach is that DNA samples of the deposited strains are stored in such centers, which can be reanalyzed with different markers such as those tightly linked to loci controlling the traits of agricultural importance. For instance, the single nucleotide polymorphisms associated with genes controlling host specificity, virulence, fungicide resistance, and toxin production will significantly increase our ability to predict the phenotypic traits of interest in newly isolated strains. With the progress in fungal genome sequencing, the transition from neutral markers, such as AFLP and microsatellite, to phenotype-specific markers is expected to happen in the near future, and the availability of DNA samples from the deposited strains should facilitate the transition.

Conclusion. Archiving available data is as important as generating new data. The proposed database will provide a means for linking these two activities and serve as a central hub for facilitating the access, comparison, and sharing of data. A comprehensive database that encompasses both the species and population data on major pathogens should serve as a valuable information resource to the discipline of plant pathology, expediting the study of newly isolated pathogens by those researchers who have limited experience in taxonomy or by regulatory agency scientists who must quickly assess the importance of newly isolated strains. Although we have mainly relied on fungi as principal examples, the proposed database is applicable to many other pathogens and pests, including those infecting animals. As we have learned from the success of GenBank, the value of any comprehensive pathogen database will largely depend on the coordinated input of the scientific community. As more data are submitted into the database, its utility as a tool for studying various aspects of pathogen biology will continue to increase. We hope that this letter initiates community-wide discussions about such databases and leads to a pilot project for testing its feasibility. Given the extensive data and strain collections currently available for *Fusarium* spp. and *M. grisea*, these two systems seem to be logical choices for such a pilot project, but the effort must be expanded to other pathogen groups.

LITERATURE CITED

1. Aoki, T., and O'Donnell, K. 1999. Morphological and molecular characterization of *Fusarium pseudograminearum* sp. nov., formerly recognized as the Group 1 population of *F. graminearum*. *Mycologia* 91:597-609.
2. Baayen, R. P., O'Donnell, K., Bonants, P. J. M., Cigelnik, E., Kroon, L. P. N. M., Roebroek, E. J. A., and Waalwijk, C. 2000. Gene genealogies and AFLP analyses in the *Fusarium oxysporum* complex identify monophyletic and nonmonophyletic formae speciales causing wilt and rot disease. *Phytopathology* 90:891-900.
3. Bisby, F. A. 2000. The quite revolution: Biodiversity informatics and the Internet. *Science* 289:2309-2312.
4. Brasier, C. 2000. The rise of the hybrid fungi. *Nature* 405:134-135.
5. Bruns, T. D., White, T. J., and Taylor, J. W. 1991. Fungal molecular systematics. *Annu. Rev. Ecol. Syst.* 22:525-564.
6. Burgess, L. W., Summerell, B. A., Giblett, G., Backhouse, D., Blake, M., Smith-White, J., and Colville, M. 2000. Role of sorghum in overseasoning of *Gibberella zeae*. Pages 82-83 in: 6th European Fusarium Seminar. Biologische Bundesanstalt für Land- und Forstwirtschaft, Berlin, Germany.
7. Carbone, I., Anderson, J. B., and Kohn, L. M. 1999. Patterns of descent in clonal lineages and their multilocus fingerprints are resolved with combined gene genealogies. *Evolution* 53:11-21.
8. Carbone, I., and Kohn, L. M. 1999. A method for designing primer sets for speciation studies in filamentous ascomycetes. *Mycologia* 91:553-556.
9. Correll, J. C., Harp, T. L., Guerber, J. C., and Lee, F. N. 2000. Differential changes in host specificity among MGR586 DNA fingerprint groups of *Pyricularia grisea*. Pages 234-242 in: *Advances in Rice Blast Research*. D. Tharreau, M. H. Lebrun, N. J. Talbot, and J. L. Nottingham, eds. Kluwer Academic Publishers, Dordrecht, the Netherlands.
10. Edwards, J. L., Lane, M. A., and Nielsen, E. S. 2000. Interoperability of biodiversity databases: Biodiversity information on every desktop. *Science* 289:2312-2314.
11. Geiser, D. M., Frisvad, J. C., and Taylor, J. W. 1998. Evolutionary relationships in *Aspergillus* section *Fumigati* inferred from partial beta-tubulin and hydrophobin DNA sequences. *Mycologia* 90:831-845.
12. Geiser, D. M., Juba, J. H., Wang, B., and Jeffers, S. N. 2001. *Fusarium hostae* sp. nov., a new species related to *F. redolens* with a *Gibberella* teleomorph. *Mycologia* 93:670-678.
13. Geiser, D. M., Pitt, J. I., and Taylor, J. W. 1998. Cryptic speciation and recombination in the aflatoxin-producing fungus *Aspergillus flavus*. *Proc. Natl. Acad. Sci. USA* 95:388-393.
14. Goodwin, S. 1997. The population genetics of *Phytophthora*. *Phytopathology* 87:462-473.
15. Hawksworth, D. L. 1991. The fungal dimension of biodiversity: Magnitude, significance, and conservation. *Mycol. Res.* 95:641-655.
16. Kamoun, S., and Hogenhout, S. A. 2001. Agricultural microbe genome 2: First glimpses into the genomes of plant-associated microbes. *Plant*

- Cell 13:451-458.
17. Keen, N. T., and Yang, C.-H. 1999. Functional genomics: Plant-microbe interactions gingerly put a foot in the water. *Physiol. Mol. Plant Pathol.* 55:313-315.
 18. Koenig, R. L., Ploetz, R. C., and Kistler, H. C. 1997. *Fusarium oxysporum* f. sp. *cubense* consists of a small number of divergent and globally distributed clonal lineages. *Phytopathology* 87:915-923.
 19. Kumar, J., Nelson, R. J., and Zeigler, R. S. 1999. Population structure and dynamics of *Magnaporthe grisea* in the Indian Himalayas. *Genetics* 152:971-984.
 20. Levy, M., Romao, J., Marchetti, M. A., and Hamer, J. E. 1991. DNA fingerprinting with a dispersed repeated sequence resolves pathotype diversity in the rice blast fungus. *Plant Cell* 3:95-102.
 21. Marasas, W. F. O., Nelson, P. E., and Toussoun, T. A. 1984. *Toxigenic Fusarium species: Identity and mycotoxicology*. Penn. State University Press, University Park, PA.
 22. McDonald, B. A. 1997. The population genetics of fungi: Tools and techniques. *Phytopathology* 87:448-453.
 23. Nei, M., and Li, W. H. 1979. Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proc. Natl. Acad. Sci. USA* 76:5269-5273.
 24. Normile, D. 1999. Culture collections seek global help. *Science* 283:1240-1241.
 25. O'Donnell, K. 2000. Molecular phylogeny of the *Fusarium solani/Nectria haematococca* species complex. *Mycologia* 92:919-938.
 26. O'Donnell, K., and Cigelnik, E. 1997. Phylogenetic relationships among ascomycetous truffles and the true and false morels from 18S and 28S ribosomal DNA sequence analysis. *Mycologia* 89:48-65.
 27. O'Donnell, K., and Cigelnik, E. 1997. Two divergent intragenomic rDNA ITS2 types within a monophyletic lineage of the fungus *Fusarium* are nonorthologous. *Mol. Phylogenet. Evol.* 7:103-116.
 28. O'Donnell, K., Cigelnik, E., and Nirenberg, H. I. 1998. Molecular systematics and phylogeography of the *Gibberella fujikuroi* species complex. *Mycologia* 90:465-493.
 29. O'Donnell, K., Kistler, H. C., Cigelnik, E., and Ploetz, R. C. 1998. Multiple evolutionary origins of the fungus causing Panama disease of banana: Concordant evidence from nuclear and mitochondrial gene genealogies. *Proc. Natl. Acad. Sci. USA* 95:2044-2049.
 30. O'Donnell, K., Kistler, H. C., Tacke, B. K., and Casper, H. H. 2000. Gene genealogies reveal global phylogeographic structure and reproductive isolation among lineages of *Fusarium graminearum*, the fungus causing wheat scab. *Proc. Natl. Acad. Sci. USA* 97:7905-7910.
 31. O'Donnell, K., Nirenberg, H. I., Aoki, T., and Cigelnik, E. 2000. A multigene phylogeny of the *Gibberella fujikuroi* species complex: Detection of additional phylogenetically distinct species. *Mycoscience* 41:61-78.
 32. Pace, N. 1997. A molecular view of microbial diversity and the biosphere. *Science* 276:734-740.
 33. Salamati, S., Zhan, J., Burdon, J. J., and McDonald, B. A. 2000. The genetic structure of field populations of *Rhynchosporium secalis* from three continents suggests moderate gene flow and regular recombination. *Phytopathology* 90:901-908.
 34. Simpson, A. J. G., Reinach, F. C., Arruda, P., Abreu, F. A., Acencio, M., Alvarenga, R., Alves, L. M., Araya, L. M. C., Araya, J. E., Baia, G. S., Baptista, C. S., Barros, M. H., Bonaccorsi, E. D., and Bordin, S. 2000. The genome sequence of the plant pathogen *Xylella fastidiosa*. *Nature* 406:151-171.
 35. Taylor, J. W., Geiser, D. M., Burt, A., and Koufopanou, V. 1999. The evolutionary biology and population genetics underlying fungal strain typing. *Clin. Microbiol. Rev.* 12:126-146.
 36. Taylor, J. W., Jacobson, D. J., Kroken, S., Kasuga, T., Geiser, D. M., Hobbitt, D. S., and Fisher, M. C. 2000. Phylogenetic species recognition and species concepts in fungi. *Fungal Genet. Biol.* 31:21-32.
 37. Tharreau, D., Lebrun, M. H., Talbot, N. J., and Notteghem, J. L., eds. 2000. *Advances in Rice Blast Research*. Kluwer Academic Publishers, Boston.
 38. Tooley, P. W., O'Neill, N. R., Goley, E. D., and Carras, M. M. 2000. Assessment of diversity in *Claviceps africana* and other *Claviceps* species by RAM and AFLP analyses. *Phytopathology* 90:1126-1130.
 39. Tsai, H. F., Liu, J. S., Staben, C., Christensen, M. J., Latch, G. C. M., Siegel, M. R., and Schardl, C. L. 1994. Evolutionary diversification of fungal endophytes of tall fescue grass by hybridization with *Epichloe* species. *Proc. Natl. Acad. Sci. USA* 91:2542-2546.
 40. Vauterin, L., Rademaker, J., and Swings, J. 2000. Synopsis on the taxonomy of the genus *Xanthomonas*. *Phytopathology* 90:677-682.
 41. Vos, P., Hogers, R., Bleeker, M., Reijans, M., Vandelee, T., Hornes, M., Frijters, A., Pot, J., Peleman, J., Kuiper, M., and Zabeau, M. 1995. AFLP—A new technique for DNA-fingerprinting. *Nucleic Acids Res.* 23:4407-4414.
 42. Zeigler, R. S. 1998. Recombination in *Magnaporthe grisea*. *Annu. Rev. Phytopathol.* 36:249-275.
 43. Zeigler, R. S., Leong, S. A., and Teng, P. S., eds. 1994. *Rice Blast Disease*. CAB International, Wallingford, England.
 44. Zeigler, R. S., Thome, J., Nelson, R., Levy, M., and Correa-Victoria, F. J. 1994. Lineage exclusion: A proposal for linking blast population analysis to resistance breeding. Pages 287-292 in: *Rice Blast Disease*. R. S. Zeigler, P. S. Teng, and S. A. Leong, eds. CAB International, Wallingford, England.