Use of genome information for the study of the pathogenesis of fungal infections and the development of diagnostic tools

Stephen P. Saville, Derek P. Thomas and José L. López-Ribot
Department of Biology, The University of Texas at San Antonio, San Antonio, Texas, USA

Summary
One of the most exciting advances in Mycology is the application of genomic approaches. The advent of genomics, together with post-genomic studies, promises to revolutionize the studies on the pathogenesis of fungal infections. Approaches include comparative genomics to identify sequences that contribute to infection and disease and functional genomics and proteomics to analyze global patterns of gene and protein expression involved in fungal pathogenesis.

Key words
Genomics, Microarrays, Proteomics, Pathogenesis, Diagnosis

Applicación del descifrado de los genomas fúngicos al estudio de la patogenia y diagnóstico de las micosis

La aplicación de técnicas genómicas representa uno de los avances más excitantes en el campo de la Micología en los últimos años. La secuenciación de diversos genomas fúngicos, junto con la utilización de técnicas post-genómicas, prometen revolucionar los estudios de patogenia en las infecciones fúngicas. Estos estudios incluyen técnicas de Genómica Comparativa, enfocadas a la identificación de secuencias que contribuyen a la infección y técnicas de Genómica Funcional y Proteómica, enfocadas al análisis global de patrones de expresión de genes y proteínas implicados en la patogenia fúngica.

Palabras clave
Genómica, Microarrays (Matrices), Proteómica, Patogénesis, Diagnóstico

Mycology is evolving as the era of extensive genome sequencing comes of age and provides vital information on complete genome sequences for a number of fungal species. The complete sequencing of the genome of the baker’s yeast, *Saccharomyces cerevisiae*, is considered a landmark in genomics [17], that afforded the first global studies of eukaryotic gene expression and gene function, and it is anticipated that the application of similar approaches to the study of pathogenic fungi will revolutionize Mycology as a discipline. Indeed, by the time the *S. cerevisiae* genome was almost completed, the *Candida albicans* sequencing project had already begun [23]. Today, it is estimated that there are now more than forty fungal genome-sequencing projects underway, including representatives of humans pathogens in all the major taxonomic groups such as *C. albicans*, *Aspergillus fumigatus*, *Cryptococcus neoformans*, *Histoplasma capsulatum*, *Blastomyces dermatitidis*, *Coccidioides immitis* and others [18,37]. Of course, the raw genome-sequence data needs to be converted into a list of predicted genes, normally by locating the open reading frames (ORFs, usually defined arbitrarily as a sequence coding for an uninterrupted string of 100 amino acids or more beginning with methionine); ideally these genes need to be annotated. This process relies heavily on bioinformatics. This information facilitates the implementation of post-genomic strategies including the application of new high-throughput technologies on a genome-wide scale and at the different levels of cellular complexity (genome, transcriptome and proteome). This is particularly appealing for this group of microorganisms, since conventional genetic and biochemical approaches are limited in many pathogenic fungi [29,36]. The present review summarizes the use of genome information and the application of post-genomic techniques to the study of fungal pathogenesis.
Comparative genomics

Comparing genomes from pathogenic organisms to closely related nonpathogenic species provides a comprehensive approach to identifying sequences that may contribute to infection and disease [30,32,38,41]. Importantly, the value of comparative genomics to the study of pathogenesis has been validated in a number of studies in microbial pathogens, including the identification of proteins distinct to pathogenic strains, and the identification of novel antigens which can represent candidates for vaccine development [5,7,11]. The importance of this approach was recognized by the Fungal Genome Initiative (http://www.broad.mit.edu/annotation/fungi/fgi/), whose stated goal is to provide the sequence of key organisms across the fungal kingdom by selecting a balanced collection of fungi (rather than choosing individual fungi in isolation) that maximizes the overall value for comparative genomics, evolutionary studies, eukaryotic biology, and medical studies. Comparisons between these carefully selected groups of fungal genomes should help identify pathogenic mechanisms that are unique for disease-causing organisms. For example, there are more than 1,000 C. albicans genes of unknown function that have no obvious ortholog in S. cerevisiae or Schizosaccharomyces pombe, and a significant proportion of these genes may play important roles during the infectious process. In addition, one of the greatest clinical needs in the field of Medical Mycology is the availability of diagnostic tools for the accurate identification of particular fungal species, and the completed genome sequences may provide the opportunity to develop unique DNA probes that could be used for identification at the species level.

Functional genomics

The availability of complete genome sequences provides a framework for the development of functional genomics to assess gene function [6]. For disease-causing fungi the main idea behind these types of studies is that determining the gene and protein expression profiles of cells under relevant environmental and infection-associ-ated conditions is likely to identify new genes and pathways associated with fungal pathogenesis. While some efforts could start at an intermediate level, the increasing complexity arising from the different genome sequencing projects highlights the need for systematic large-scale efforts to explore functions and interactions of genes and proteins at a global level. Thus, in the post-genomic era, the tendency is for high-throughput techniques, such as DNA-microarray analyses, to replace “old fashioned” methods such as subtractive hybridization, differential display, and SAGE. Additionally, other techniques can be used for the global analysis of protein-protein interactions (i.e. yeast two hybrid), protein-DNA interactions (i.e. ChIP on chip), to determine the subcellular localization of gene products (i.e. GFP fusion libraries), and for in silico prediction of immunoreactive epitopes. A word of caution is that results from large primary screens such as these ultimately need to be confirmed by several independent, namely biochemical and molecular approaches. For example, in the case that a primary screen results in the identification of a putative virulence factor, subsequent studies must be performed in order to fulfill molecular Koch’s postulates and so firmly implicate a particular property in virulence. The following paragraphs summarize these techniques most commonly used in the field of functional genomics and whose application may provide important insights into the pathogenesis of fungal infections.

Large scale mutant generation

Although different techniques such as RNA interference, peptide aptamers and others can be used to perturb gene function for systematic screens of gene function [6], the most widely used approach is the construction of collections of mutant strains [44]. This is best exemplified by the construction of the yeast deletion mutant collection in S. cerevisiae, a set of over 2,000 knockout strains generated by a consortium of European and North American laboratories [46]. The use of deletion strains represents a straightforward manner to gather important information about the function of genes identified via genome sequencing by targeted elimination of its activity and observation of the resulting phenotype(s). The main advantage of this approach is that each member of the collection of gene-targeted organisms or reagents is already sequenced and is easily and instantly identifiable. In contrast, other random techniques, such as large scale insertional mutagenesis including transposon mutagenesis, can be used to generate collections of mutant strains which are then screened for their inability to carry out a particular function or pathogenic trait. Transposons are mobile genetic elements that can be used to disrupt genes in a random, non-targeted fashion. Analyses using mutant collections may be facilitated by the introduction of “molecular bar codes” or “signature-tags” during the gene disruption process. For example, the bar code on each yeast deletion strain allows the identification of the strain by sequencing the code or by hybridizing DNA from the strain onto a microarray. Pools of such “signature-tagged” mutants can be screened efficiently in competition experiments both in vitro and in vivo to identify those with a defect in virulence [2,8,9,12,19,20,35].

DNA-arrays

DNA array technology is used to study gene expression on a genome-wide scale. Without question, the use of DNA microarrays is becoming the method of choice for assaying gene expression, particularly as costs and complexity are being reduced [14]. The most widely used approach is the construction of collections of mutant strains [44]. This is best exemplified by the construction of the yeast deletion mutant collection in S. cerevisiae, a set of over 2,000 knockout strains generated by a consortium of European and North American laboratories [46]. The use of deletion strains represents a straightforward manner to gather important information about the function of genes identified via genome sequencing by targeted elimination of its activity and observation of the resulting phenotype(s). The main advantage of this approach is that each member of the collection of gene-targeted organisms or reagents is already sequenced and is easily and instantly identifiable. In contrast, other random techniques, such as large scale insertional mutagenesis including transposon mutagenesis, can be used to generate collections of mutant strains which are then screened for their inability to carry out a particular function or pathogenic trait. Transposons are mobile genetic elements that can be used to disrupt genes in a random, non-targeted fashion. Analyses using mutant collections may be facilitated by the introduction of “molecular bar codes” or “signature-tags” during the gene disruption process. For example, the bar code on each yeast deletion strain allows the identification of the strain by sequencing the code or by hybridizing DNA from the strain onto a microarray. Pools of such “signature-tagged” mutants can be screened efficiently in competition experiments both in vitro and in vivo to identify those with a defect in virulence [2,8,9,12,19,20,35].

DNA-arrays

DNA array technology is used to study gene expression on a genome-wide scale. Without question, the use of DNA microarrays is becoming the method of choice for assaying gene expression, particularly as costs and complexity are being reduced [14]. The most widely used approach is the construction of collections of mutant strains [44]. This is best exemplified by the construction of the yeast deletion mutant collection in S. cerevisiae, a set of over 2,000 knockout strains generated by a consortium of European and North American laboratories [46]. The use of deletion strains represents a straightforward manner to gather important information about the function of genes identified via genome sequencing by targeted elimination of its activity and observation of the resulting phenotype(s). The main advantage of this approach is that each member of the collection of gene-targeted organisms or reagents is already sequenced and is easily and instantly identifiable. In contrast, other random techniques, such as large scale insertional mutagenesis including transposon mutagenesis, can be used to generate collections of mutant strains which are then screened for their inability to carry out a particular function or pathogenic trait. Transposons are mobile genetic elements that can be used to disrupt genes in a random, non-targeted fashion. Analyses using mutant collections may be facilitated by the introduction of “molecular bar codes” or “signature-tags” during the gene disruption process. For example, the bar code on each yeast deletion strain allows the identification of the strain by sequencing the code or by hybridizing DNA from the strain onto a microarray. Pools of such “signature-tagged” mutants can be screened efficiently in competition experiments both in vitro and in vivo to identify those with a defect in virulence [2,8,9,12,19,20,35].

DNA-arrays

DNA array technology is used to study gene expression on a genome-wide scale. Without question, the use of DNA microarrays is becoming the method of choice for assaying gene expression, particularly as costs and complexity are being reduced [14]. The most widely used approach is the construction of collections of mutant strains [44]. This is best exemplified by the construction of the yeast deletion mutant collection in S. cerevisiae, a set of over 2,000 knockout strains generated by a consortium of European and North American laboratories [46]. The use of deletion strains represents a straightforward manner to gather important information about the function of genes identified via genome sequencing by targeted elimination of its activity and observation of the resulting phenotype(s). The main advantage of this approach is that each member of the collection of gene-targeted organisms or reagents is already sequenced and is easily and instantly identifiable. In contrast, other random techniques, such as large scale insertional mutagenesis including transposon mutagenesis, can be used to generate collections of mutant strains which are then screened for their inability to carry out a particular function or pathogenic trait. Transposons are mobile genetic elements that can be used to disrupt genes in a random, non-targeted fashion. Analyses using mutant collections may be facilitated by the introduction of “molecular bar codes” or “signature-tags” during the gene disruption process. For example, the bar code on each yeast deletion strain allows the identification of the strain by sequencing the code or by hybridizing DNA from the strain onto a microarray. Pools of such “signature-tagged” mutants can be screened efficiently in competition experiments both in vitro and in vivo to identify those with a defect in virulence [2,8,9,12,19,20,35].
The increasing accumulation of genomic sequence information for pathogenic fungi, combined with methods to efficiently assess gene functions, is having a major impact in the field of Medical Mycology and is revolutionizing the way we do science. In this exciting time, the implementation of large-scale genomic and post-genomic techniques offers unprecedented opportunities to further understand and elucidate the pathogenic traits associated with fungal infections as well as providing a framework for the development of new diagnostic tools for these important pathogens.

Proteomics

Large-scale DNA sequencing has also provided an infrastructure for protein analyses and a complementary technology to DNA microarrays provided by proteomics, a term generally used to encapsulate all of the technology currently available to analyze global patterns of protein expression. Importantly, proteomics addresses biological characteristics that cannot be identified through DNA analysis, such as relative abundance of the protein product, whether it is post-translationally modified, its subcellular localization, turnover, and possible interaction with other proteins as well as functional aspects. The field of proteomics involves the combined application of advanced techniques to resolve (typically by high resolution two-dimensional polyacrylamide gel electrophoresis, 2DE), identify (mass spectrometry, MS), quantify and characterize proteins, as well as bioinformatics tools to store, communicate and interlink protein and DNA information from genome projects (Figure 2). Each one of these technologies can be applied independently, although their impact is maximized when used together as a comprehensive package to study complex biological problems. To date, proteome analyses of pathogenic fungi have focused on the understanding of the process of dimorphism, the structure and composition of the fungal cell wall, virulence factors and drug resistance [3,4,10,13,21,22,31]. In addition, 2DE and immunoblotting techniques have been used to study antibody responses during infection as screens for virulence factors, potential vaccine candidates and diagnostic markers using sera from infected patients [15,39,40]. Although the application of these techniques to the study of pathogenic fungi is still in its infancy, it is clear that proteomics is likely to advance our understanding of the host–fungus interactions in the near future.

Perspectives

The increasing accumulation of genomic sequence information for pathogenic fungi, combined with methods to efficiently assess gene functions, is having a major impact in the field of Medical Mycology and is revolutionizing the way we do science. In this exciting time, the implementation of large-scale genomic and post-genomic techniques offers unprecedented opportunities to further understand and elucidate the pathogenic traits associated with fungal infections as well as providing a framework for the development of new diagnostic tools for these important pathogens.

Work in the laboratory is supported by Public Health Service grants R21 DE 15079 and R03 AI 054447 (to JLL-R). JLL-R is the recipient of a New Investigator Award in Molecular Pathogenic Mycology from the Burroughs Wellcome Fund.
References


