

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

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

Resumen

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 human 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 genomic information and the application of post-genomic techniques to the study of fungal pathogenesis.

Corresponding address:

Dr. José L. López-Ribot
Department of Biology
University of Texas at San Antonio
6900 North Loop, 1604 West
San Antonio, Texas 78249, USA
Tel.: +1 210 458 7022
Fax: +1 210 458 7023
E-mail: jose.lopezribot@utsa.edu

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-associated 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 [14]. The following paragraphs summarize the 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 as the technology becomes more widespread and better standardized. Briefly, the expression of thousands of genes can be measured simultaneously by spotting an array of DNA samples (PCR products or oligonucleotides) corresponding to the individual gene sequences from the genome of interest on a solid support (typically a glass microscope slide) and hybridizing labeled mRNA (or reverse-transcribed cDNA) to the spotted surface. Because microarray experiments produce reams of data, the use of bioinformatics is critical for quantification, data analyses and mining [42]. Figure 1 shows a schematic diagram illustrating the stages of a typical DNA-microarray experiment. The key advantage of this technology is that patterns of expression involving many genes can be identified simultaneously. The expression profiles of cells growing under different relevant *in vitro* and *in vivo* conditions can identify key genes and pathways involved in fungal pathogenesis. Global analysis of transcription with DNA arrays in pathogenic fungi is becoming more frequent, particularly in the case of *C. albicans*, although DNA arrays for other pathogenic fungi are being developed. In recent years several reports have described the use of this technique to study different aspects of fungal pathogenesis, including responses to different environmental conditions, morphogenesis and its regulatory pathways, biofilm formation and antifungal drug resistance [1,16,24-28,33,34,43,45].

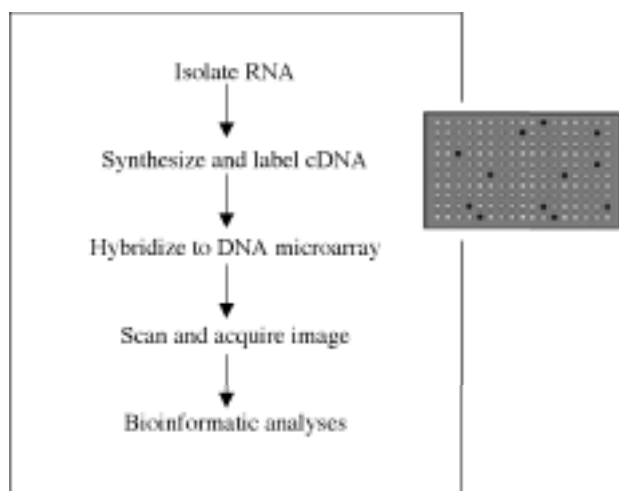


Figure 1. Schematic diagram illustrating the key stages of DNA-microarray analyses.

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.

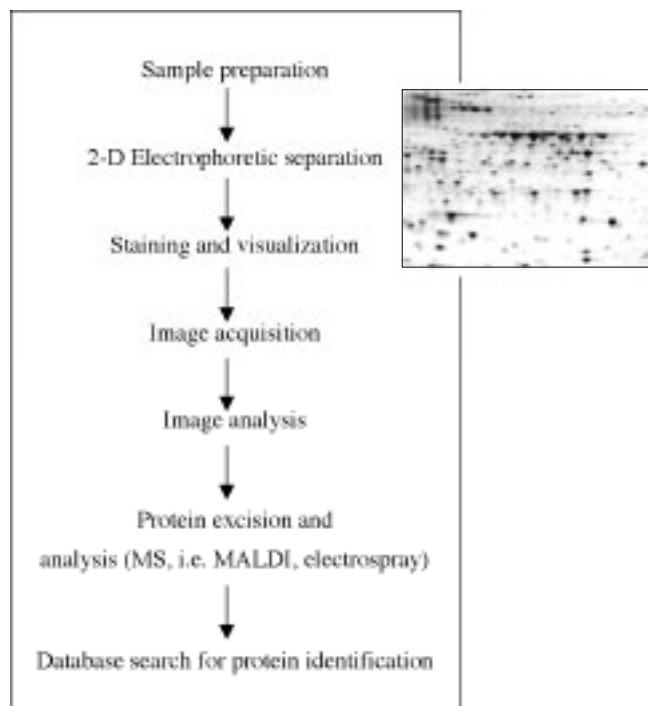


Figure 2. Workflow of a typical proteomics experiment.

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

- Barker KS, Crisp S, Wiederhold N, Lewis RE, Bareither B, Eckstein J, Barbuch R, Bard M, Rogers PD. Genome-wide expression profiling reveals genes associated with amphotericin B and fluconazole resistance in experimentally induced antifungal resistant isolates of *Candida albicans*. *J Antimicrob Chemother* 2004; 54: 376-385.
- Brown JS, Aufauvre-Brown A, Brown J, Jennings JM, Arst H, Jr., Holden DW. Signature-tagged and directed mutagenesis identify PABA synthetase as essential for *Aspergillus fumigatus* pathogenicity. *Mol Microbiol* 2000; 36: 1371-1380.
- Bruneau JM, Magnin T, Tagat E, Legrand R, Bernard M, Diaquin M, Fudali C, Latge JP. Proteome analysis of *Aspergillus fumigatus* identifies glycosylphosphatidylinositol-anchored proteins associated to the cell wall biosynthesis. *Electrophoresis* 2001; 22: 2812-2823.
- Bruneau JM, Maillet I, Tagat E, Legrand R, Supatto F, Fudali C, Caer JP, Labas V, Lecaque D, Hodgson J. Drug induced proteome changes in *Candida albicans*: comparison of the effect of beta(1,3) glucan synthase inhibitors and two triazoles, fluconazole and itraconazole. *Proteomics* 2003; 3: 325-336.
- Carlton JM. Genome sequencing and comparative genomics of tropical disease pathogens. *Cell Microbiol* 2003; 5: 861-873.
- Carpenter AE, Sabatini DM. Systematic genome-wide screens of gene function. *Nat Rev Genet* 2004; 5: 11-22.
- Cassat JE, Dunman PM, McAleese F, Murphy E, Projan SJ, Smeltzer MS. Comparative genomics of *Staphylococcus aureus* musculoskeletal isolates. *J Bacteriol* 2005; 187: 576-592.
- Castano I, Kaur R, Pan S, Cregg R, Penas Ade L, Guo N, Biery MC, Craig NL, Cormack BP. Tn7-based genome-wide random insertional mutagenesis of *Candida glabrata*. *Genome Res* 2003; 13: 905-915. Epub 2003 Apr 2014.
- Chiang SL, Mekalanos JJ, Holden DW. In vivo genetic analysis of bacterial virulence. *Annu Rev Microbiol* 1999; 53: 129-154.
- Choi W, Yoo YJ, Kim M, Shin D, Jeon HB. Identification of proteins highly expressed in the hyphae of *Candida albicans* by two-dimensional electrophoresis. *Yeast* 2003; 20: 1053-1060.
- Cockle PJ, Gordon SV, Lalvani A, Buddle BM, Hewinson RG, Vordermeier HM. Identification of novel *Mycobacterium tuberculosis* antigens with potential as diagnostic reagents or subunit vaccine candidates by comparative genomics. *Infect Immun* 2002; 70: 6996-7003.
- Cormack BP, Ghori N, Falkow S. An adhesin of the yeast pathogen *Candida glabrata* mediating adherence to human epithelial cells. *Science* 1999; 285: 578-582.
- de Groot PW, de Boer AD, Cunningham J, Dekker HL, de Jong L, Hellingwerf KJ, de Koster C, Klis FM. Proteomic analysis of *Candida albicans* cell walls reveals covalently bound carbohydrate-active enzymes and adhesins. *Eukaryot Cell* 2004; 3: 955-965.
- Falkow S. Molecular Koch's postulates applied to bacterial pathogenicity—a personal recollection 15 years later. *Nat Rev Microbiol* 2004; 2: 67-72.
- Fernandez-Arenas E, Molero G, Nombela C, Diez-Orejas R, Gil C. Low virulent strains of *Candida albicans*: unravelling the antigens for a future vaccine. *Proteomics* 2004; 4: 3007-3020.
- Garcia-Sanchez S, Aubert S, Iraqui I, Janbon G, Ghigo JM, d'Enfert C. *Candida albicans* biofilms: a developmental state associated with specific and stable gene expression patterns. *Eukaryot Cell* 2004; 3: 536-545.
- Goffeau A, Barrell BG, Bussey H, Davis RW, Dujon B, Feldmann H, Galibert F, Hoheisel JD, Jacq C, Johnston M, Louis EJ, Mewes HW, Murakami Y, Philippsen P, Tettelin H, Oliver SG. Life with 6000 genes. *Science* 1996; 274: 546, 563-547.
- Gow NA. New angles in mycology: studies in directional growth and directional motility. *Mycol Res* 2004; 108: 5-13.
- Hensel M, Holden DW. Molecular genetic approaches for the study of virulence in both pathogenic bacteria and fungi. *Microbiology* 1996; 142: 1049-1058.
- Hensel M, Shea JE, Gleeson C, Jones MD, Dalton E, Holden DW. Simultaneous identification of bacterial virulence genes by negative selection. *Science* 1995; 269: 400-403.
- Hernandez R, Nombela C, Diez-Orejas R, Gil C. Two-dimensional reference map of *Candida albicans* hyphal forms. *Proteomics* 2004; 4: 374-382.
- Hooshdaran MZ, Barker KS, Hilliard GM, Kusch H, Morschhauser J, Rogers PD. Proteomic analysis of azole resistance in *Candida albicans* clinical isolates. *Antimicrob Agents Chemother* 2004; 48: 2733-2735.
- Jones T, Federspiel NA, Chibana H, Dungan J, Kalman S, Magee BB, Newport G, Thorstenson YR, Agabian N, Magee PT, Davis RW, Scherer S. The diploid genome sequence of *Candida albicans*. *Proc Natl Acad Sci USA* 2004; 101: 7329-7334.
- Kraus PR, Boily MJ, Giles SS, Stajich JE, Allen A, Cox GM, Dietrich FS, Perfect JR, Heitman J. Identification of *Cryptococcus neoformans* temperature-regulated genes with a genomic-DNA microarray. *Eukaryot Cell* 2004; 3: 1249-1260.
- Lan CY, Rodarte G, Murillo LA, Jones T, Davis RW, Dungan J, Newport G, Agabian N. Regulatory networks affected by iron availability in *Candida albicans*. *Mol Microbiol* 2004; 53: 1451-1469.
- Lane S, Birse C, Zhou S, Matson R, Liu H. DNA array studies demonstrate convergent regulation of virulence factors by Cph1, Cph2, and Efg1 in *Candida albicans*. *J Biol Chem* 2001; 276: 48988-48996.
- Lorenz MC, Bender JA, Fink GR. Transcriptional response of *Candida albicans* upon internalization by macrophages. *Eukaryot Cell* 2004; 3: 1076-1087.
- Lorenz MC, Fink GR. The glyoxylate cycle is required for fungal virulence. *Nature* 2001; 412: 83-86.
- Magee PT, Gale C, Berman J, Davis D. Molecular genetic and genomic approaches to the study of medically important fungi. *Infect Immun* 2003; 71: 2299-2309.
- Miller W, Makova KD, Nekrutenko A, Hardison RC. Comparative genomics. *Annu Rev Genomics Hum Genet* 2004; 5: 15-56.
- Missall TA, Pusateri ME, Lodge JK. Thiol peroxidase is critical for virulence and resistance to nitric oxide and peroxide in the fungal pathogen, *Cryptococcus neoformans*. *Mol Microbiol* 2004; 51: 1447-1458.
- Moran G, Stokes C, Thewes S, Hube B, Coleman DC, Sullivan D. Comparative genomics using *Candida albicans* DNA microarrays reveals absence and divergence of virulence-associated genes in *Candida dubliniensis*. *Microbiology* 2004; 150: 3363-3382.
- Murad AM, d'Enfert C, Gaillardin C, Tournu H, Tekaia F, Talibi D, Marechal D, Marchais V, Cottin J, Brown AJ. Transcript profiling in *Candida albicans* reveals new cellular functions for the transcriptional repressors CaTup1, CaMig1 and CaNrg1. *Mol Microbiol* 2001; 42: 981-993.
- Nantel A, Dignard D, Bachewich C, Harcus D, Marciel A, Bouin AP, Sensen CW, Hogues H, van het Hoog M, Gordon P, Rigby T, Benoit F, Tessier DC, Thomas DY, Whiteway M. Transcription profiling of *Candida albicans* cells undergoing the yeast-to-hyphal transition. *Mol Biol Cell* 2002; 13: 3452-3465.
- Nelson RT, Hua J, Pryor B, Lodge JK. Identification of virulence mutants of the fungal pathogen *Cryptococcus neoformans* using signature-tagged mutagenesis. *Genetics* 2001; 157: 935-947.
- Odds FC, Brown AJ, Gow NA. *Candida albicans* genome sequence: a platform for genomics in the absence of genetics. *Genome Biol* 2004; 5: 230.
- Pennisi E. Sequencing. Chimps and fungi make genome. *Science* 2002; 296: 1589-1591.
- Piskur J, Langkjaer RB. Yeast genome sequencing: the power of comparative genomics. *Mol Microbiol* 2004; 53: 381-389.
- Pitarch A, Abian J, Carrascal M, Sanchez M, Nombela C, Gil C. Proteomics-based identification of novel *Candida albicans* antigens for diagnosis of systemic candidiasis in patients with underlying hematological malignancies. *Proteomics* 2004; 4: 3084-3106.
- Pitarch A, Diez-Orejas R, Molero G, Pardo M, Sanchez M, Gil C, Nombela C. Analysis of the serologic response to systemic *Candida albicans* infection in a murine model. *Proteomics* 2001; 1: 550-559.
- Prentice MB. Bacterial comparative genomics. *Genome Biol* 2004; 5: 338.
- Quackenbush J. Computational analysis of microarray data. *Nat Rev Genet* 2001; 2: 418-427.
- Rogers PD, Barker KS. Evaluation of differential gene expression in fluconazole-susceptible and -resistant isolates of *Candida albicans* by cDNA microarray analysis. *Antimicrob Agents Chemother* 2002; 46: 3412-3417.
- Scherens B, Goffeau A. The uses of genome-wide yeast mutant collections. *Genome Biol* 2004; 5: 229.
- Sohn K, Urban C, Brunner H, Rupp S. *EFG1* is a major regulator of cell wall dynamics in *Candida albicans* as revealed by DNA microarrays. *Mol Microbiol* 2003; 47: 89-102.
- Winzeler EA, Shoemaker DD, Astromoff A, Liang H, Anderson K, Andre B, Bangham R, Benito R, Boeke JD, Bussey H, Chu AM, Connelly C, Davis K, Dietrich F, Dow SW, El Bakkoury M, Foury F, Friend SH, Gentalen E, Giaever G, Hegemann JH, Jones T, Laub M, Liao H, Liebundguth N, Lockhart DJ, Lucau-Danila A, Lussier M, M'Rabet N, Menard P, Mittmann M, Pai C, Rebischung C, Revuelta JL, Riles L, Roberts CJ, Ross-MacDonald P, Scherens B, Snyder M, Sookhai-Mahadeo S, Storms RK, Veroneau S, Voet M, Volckaert G, Ward TR, Wysocki R, Yen GS, Yu K, Zimmermann K, Philippsen P, Johnston M, Davis RW. Functional characterization of the *S. cerevisiae* genome by gene deletion and parallel analysis. *Science* 1999; 285: 901-906.