Genomics of the filamentous fungi – moving from the shadow of the bakers yeast

Simon J. FOSTER[1], Brendon J. MONAHAN[2], Rosie E. BRADSHAW*

Institute of Molecular BioSciences, Massey University, Private Bag 11-222, Palmerston North 5321, New Zealand

Keywords:
Comparative genomics
Functional genomics
Fungal Genome Initiative
Genome sequence

A B S T R A C T

Fungi have now well and truly entered the genomic age. We currently know the complete DNA sequence for 18 fungal species and many more fungal genome sequencing projects are in progress. Whilst yeasts dominated the early genomic years, recently there has been a dramatic increase in filamentous fungal genome projects. The implications of this wealth of genetic information for mycologists worldwide is immense. In this review we summarise the background to fungal genome projects with an emphasis on the filamentous fungi. We discuss efforts to determine gene function and to compare genomes from different species. Since this is such a fast-moving field, useful web sites are listed that will enable the reader to keep up to date with developments.

© 2005 The British Mycological Society. Published by Elsevier Ltd. All rights reserved.

1.   Introduction

Genomics, the study of all the genetic information in an organism, has gained much wide-spread publicity since the completion of a draft of the human genome in 2001 (Lander et al. 2001). However for mycologists, the completion in 1996 of the genome sequence of the bakers yeast, Saccharomyces cerevisiae (Goffeau et al. 1996), represented a milestone in the field of genomics, being the first completed eukaryotic genome sequence and only the sixth of any organism to be published. But 6 y after the publication of the S. cerevisiae genome sequence, although 111 other genome sequences (94 prokaryotic, 17 eukaryotic) had been published, only two of them were fungal: the model fission yeast Schizosaccharomyces pombe (Wood et al. 2002) and the microsporidian Encephalitozoon cuniculi (Katinka et al. 2001). Put another way, fungal sequences during this time represented only 16.5 Mb of 5900 Mb sequenced, a meager 0.3 %.

Shortly after the S. cerevisiae genome release, workshops were held that discussed initiating genome projects for the filamentous fungi, namely the model organisms Aspergillus nidulans and Neurospora crassa (Hamer 1996). The objectives and justifications for such a proposal were presented in a subsequent White Paper and an open letter to the fungal community by Joan Bennett (Bennett 1997a,b). What grew from these meetings and subsequent funding applications were a number of small pilot sequencing projects such as the A. nidulans chromosome IV sequencing project. However, it was not until the Fungal Genome Initiative (FGI) was formally established following meetings and workshops (initiated in November 2000 by Gerry Fink of the Whitehead Institute, now the Broad Institute), that the momentum in sequencing fungal genomes began to build. In February 2002 a White Paper proposal was presented that prioritized a set of fungi for genome sequencing (Birren et al. 2003). The 2003 publication of the Neurospora crassa genome sequence (Galagan et al. 2003), the first to be

---

published from a filamentous fungus, illustrated the enormous potential and importance of the FGI (Hynes 2003). Now with the release of seven other FGI fungal genome sequences and an ever-increasing number of other independent fungal genome projects, fungi were finally on the map with respect to genome sequencing (Hofmann *et al.* 2003; Lorenz 2002; Zeng *et al.* 2001).

2.   The Fungal Genome Initiative and fungal genome projects

A key point from the early FGI meetings was that fungi for genome sequencing should not be selected one at a time but considered as part of a cohesive strategy approached in a kingdom-wide manner. It was argued that a balanced selection of fungi would maximise the overall value for comparative genomics, evolutionary studies, eukaryotic biology and medical studies. This contrasted with the single-focused, fragmented and in some instances poorly resourced fungal genome projects that were underway or in the pipe-line at the time. After consultation of the wider fungal community this was the strategy that was presented in the 2002 FGI White Paper (Birren *et al.* 2002) and the three additional FGI proposals that followed (Birren 2003; Birren *et al.* 2003; Birren 2004).

Although not all of the proposed candidate species for sequencing were endorsed by the relevant funding bodies, the FGI succeeded in obtaining funding for 24 fungal genome projects, of which the National Human Genome Research Institute (NHGRI) funded 18. At the time of writing, sequence assemblies have been released for 16 of these projects. A genome status page is available on the FGI web site that lists the current sequence and assembly status for each organism and also provides a direct link to the respective genome homepage (see Table 1 for a list of useful web sites). The FGI sequencing project is based at the Broad Institute Center for Genome Research in Cambridge, Massachusetts, the largest US federally funded center for genome sequencing. Sequencing of *N. crassa*, *Magnaporthe grisea* and *Fusarium graminearum* genomes has been performed at the Broad Institute under the FGI umbrella. A common factor between these three sequencing projects is that they represent a collaboration between the Broad Institute and a respective community/consortium: the *Neurospora* research community for the *N. crassa* project, the International Rice Blast Genome Consortium for *M. grisea* and the International *Gibberella zeae* Genomics Consortium (IGGR) for *F. graminearum*.

Interestingly, sequencing of *M. grisea* and *F. graminearum* genomes was originally proposed in the first FGI white paper but was not endorsed by the NHGRI. Overall it appeared that fungi that interact with plants were under-represented in the NHGRI-funded genome projects in comparison with those that interact with humans. This deficiency prompted the formation of the Plant-Associated Microbe Genome Initiative (PAMGI) which produced a white paper seeking US$500m over 5 y for the sequencing of various plant-associated microbial genomes (Leach *et al.* 2002). A revised list of target fungal genomes was released by the American Phytopathological Society on their web site in 2005 (see Table 1). This list included seven "immediate priority" species to be sequenced, of which

Name	URL
Fungal genome resources	
Fungal Genome Initiative (FGI)	http://www.broad.mit.edu/annotation/fungi/fgi/
Plant-Associated Microbe Genome Initiative (PAMGI)	http://www.apsnet.org/members/ppb/plantassocinitiative.asp
NCBI\genome page	http://www.ncbi.nlm.nih.gov/genomes/
JGI	http://www.jgi.doe.gov/
Neurospora Genome Project	http://biology.unm.edu/biology/ngp/home.html
Washington University Medical School Genome Sequencing Centre	http://genome.wustl.edu/
Aspergillus niger, DSM Food Specialties	http://www.dsm.com/en_US/html/dfs/genomics_aniger.htm
European Functional Analysis Project	http://mips.gsf.de/proj/eurofan
Saccharomyces Genome Deletion Project	http://www-sequence.stanford.edu/group/yeast_deletion_project/
Fungal Genetics Stock Centre (FGSC)	http://www.fgsc.net
General genome resources	
Genomes online database (GOLD)	http://www.genomesonline.org/
The Institute for Genome Research (TIGR)	http://www.tigr.org/
Genome News Network (GNN)	http://www.genomenewsnetwork.org

Table 1 – Web sites of interest and referred to in the text

three are forest pathogens, along with 12 "high priority" species. An important consideration in the choice of plant pathogenic species to be sequenced has been an attempt to obtain broad phylogenetic coverage (Goodwin 2004). For example within the Ascomycota, the fungal classes Dothideomycetes and Leotiomycetes were under-represented until recently (Rouxel & Balesdent 2005).

Genome sequencing of agriculturally important plant pathogens is also a focus for the Department of Energy (DOE) Joint Genome Institute (JGI). Formed in 1997, the JGI is managed by the University of California and funded predominantly by the DOE Office of Biological and Environmental Research. With respect to fungal genomics, a major achievement of the JGI thus far is the recent publication of the genome sequence for the white rot fungus *Phanerochaete chrysosporium* (Martinez *et al.* 2004). The JGI currently has eleven fungal genome projects either on-going or in the pipeline, including the cellulotytic fungus *Trichoderma reesei*. Like the FGI a priority for the JGI is to make the genome sequence data freely available to the scientific community.

The Institute for Genome Research (TIGR) has also made an important contribution to fungal genome projects. TIGR has sequenced *Aspergillus fumigatus* and *Cryptococcus neoformans* JEC21 genomes and is currently sequencing *Coccidioides posadasii*. Therefore, while the FGI is a major source of publicly available genome data, especially for the filamentous fungi, it would be a mistake to conclude that this represents all current fungal genome projects. In total, there are currently 115 completed or ongoing fungal genome projects of which 24

are FGI projects. Unfortunately, quite a few fungal genome sequences are not publicly available. Probably the best example of this is the industrially important fungus *Aspergillus niger*. DSM Food Specialties (The Netherlands) commissioned an 8× sequence coverage of the *A. niger* genome which was completed in October 2001. However, free access to the DSM *A. niger* database for academic or non-profit organizations is available by completing an Access Agreement.

So how can you find out and keep track of genome projects? There are a number of regularly updated web-based resources that can be of great assistance. The Genomes OnLine Database (GOLD) is an excellent site and provides a comprehensive resource on completed and current genome sequencing projects. The Genome News Network (GNN) is a good source of news and articles that may be of interest and also contains a very useful "guide to sequenced genomes". And of course there is the National Center for Biotechnology Information (NCBI), which contains many genome resources including a specific fungal genome page (see Table 1 for URLs).

3.   EST and RST projects

In addition to the 115 fungal genome projects listed on the GOLD database at the time of writing there are also 23 EST and 10 RST projects. The aim of EST (Expressed Sequence Tag) projects is to identify all of the expressed genes by sequencing complementary DNA (cDNA) copies of the mRNA. Considerably quicker and cheaper than sequencing a whole genome, EST information provides valuable information about the coding regions of a genome. ESTs are useful for guiding the annotation of full genome studies, as well as being useful research tools in their own right. There are also ten projects involving fungal RSTs (Random Sequence Tags). These are short sequences obtained from both ends of random genomic library clones and allow rapid accumulation of random partial genome data. This type of sequence data is particularly useful when it can be compared with annotated sequences from closely related species, in which case protein-coding regions can be rapidly identified.

4.   Functional genomics

Obtaining a genome sequence for an organism is undoubtedly a major leap forward. But in terms of understanding an organism, a genome sequence is probably akin to a dictionary that lists all the words but lacks definitions of those words. We need to decipher the genome; we still need to know which sections contain genes, find out what the genes do and how they interact. Fortunately, this daunting task has already been partially tackled with respect to the yeasts and experience gathered there will be of great use to scientists working with filamentous fungi. Firstly, many genes are shared between yeast and other fungi (for example genes with basic cellular functions) and so will probably have the same function in both organisms. Secondly, for those genes that do not have a currently assigned function, experience with the yeasts has developed frameworks that can be followed. In an effort to assign functions to genes in *S. cerevisiae*, recent projects

have focused on the generation of knockout libraries in which single genes have been deleted, enabling possible functions to be assigned based on phenotypes. The *Saccharomyces* Genome Deletion Project (Giaever *et al.* 2002) completed the systematic deletion of 96 % of the yeast open reading frames, resulting in the production of over 20,000 strains which have been examined phenotypically to determine whether the deleted gene has a readily identifiable phenotype. Similarly, the European Functional Analysis (EUROFAN) project has recently completed the deletion of over 1000 yeast open reading frames and obtained basic phenotypic data relating to growth on different media and at different temperatures for each deletion strain. The deletion mutant strains generated during both of these projects are available to researchers (see web sites in Table 1).

Yeasts are relatively simple organisms compared with filamentous fungi, many of which are important pathogens of animals and plants and have genomes approximately twice the size of yeasts. Thus, gaining understanding of genomes within filamentous fungi, using similar methods to those employed for yeasts, will require extensive work. To develop good knowledge of genes that bestow, for example, the ability to cause disease, work will be required to generate resources such as knockout libraries in medically and agriculturally important fungal species. Unfortunately, the ability to delete genes in other fungi varies with species and for the majority it is a much harder prospect than for yeast owing to yeasts having a much more effective system for homologous recombination (the process by which fragments of homologous DNA are incorporated into the genome). Nevertheless, technologies improve all the time and researchers are constantly striving to improve methods for gene-knockouts. Consequently it is only a matter of time before such knockout libraries will be available for filamentous fungi.

The availability of genome sequence data also allows gene function to be investigated in the opposite direction, from phenotype to genotype. Whole-genome micro-arrays for the yeast genome are available that display copies of each of the genes in the genome. By hybridisation with probes derived from mRNA it is possible to determine which genes are switched on in a particular cell under particular conditions. This procedure enables expression analysis to be performed on strains with an altered phenotype and compared with that of normal wild type strains to identify genetic differences that may account for the phenotype (Horak & Snyder 2002).

5.   Comparative genomics

Comparative genomics, the analysis and comparison of genomes from different species, is a powerful tool. It has particular value when working with a poorly characterised species that is closely related to another species for which the genome sequence has been determined. The power of comparative genomics arises from the phenomenon of synteny, whereby the order of genes and the actual nucleotide sequences of the genes themselves are often closely conserved between related organisms. Hence, theoretically it is possible to locate a gene in a poorly characterised fungus by searching in the equivalent genomic region to that in which the gene is found in the annotated genome of a close relative.

Of particular interest to those who work with filamentous fungi, is the identification of genes specifically required for certain lifestyles such as parasitism. Traditionally this has been achieved either by targeting specific candidate genes or by generating mutants in which normal behaviour is altered. By comparing the genomes of related species which differ in their ecological lifestyle, our ability to identify genes that enable the fungi to live the way they do will be vastly improved. This was the rationale behind the choice of fungi to be sequenced as outlined in the third FGI white paper (Birren 2003). By focusing sequencing efforts on fungi that are important human pathogens and those that, although considered to be non-pathogenic or less aggressive pathogens, are close enough evolutionarily to enable comparisons, the FGI hopes to aid the elucidation of genes responsible for disease causing traits. A comparison of the genome sequence of the plant pathogen *Magnaporthe grisea* with those of the non-pathogens *Neurospora crassa* and *Aspergillus nidulans* revealed a larger number of genes in the pathogen, with expansion of some gene families including G-protein receptor and secondary metabolism genes (Dean *et al.* 2005).

In a similar vein it may be possible to identify genetic mechanisms that govern the host ranges of closely related phytopathogenic fungi. For example, research within the Centre for Functional Genomics at Massey University is focused on the molecular basis of secondary metabolite production by grass endophytes (fungi that form symbiotic relationships with grasses). Different endophyte species produce a different spectrum of secondary metabolites, each with different functions, such as insect and mammalian toxins. The genes responsible for the production of these toxins are often found grouped together in gene clusters. Whilst good progress is being made in identifying the gene clusters for some secondary metabolites (Scott 2003; Tanaka *et al.* 2005), the ability to rapidly determine the presence or absence of different genes or gene clusters in related endophyte genomes using comparative genomics will greatly enhance our ability to determine the functions of these genes.

Comparative genomics is also used to identify and annotate genes within the published fungal genome sequences. Whilst there are bioinformatic tools for identifying coding genes from sequence information, these are not unfailingly accurate and information from comparative analysis of genomes is helping to refine the tools to make their predictions more accurate. For example, upon its initial publication it was thought that the *S. cerevisiae* genome contained approximately 6200 genes. By comparing the genome of *S. cerevisiae* with three close relatives, *S. paradoxus*, *S. mikatae* and *S. bayanus* using specially developed algorithms, Kellis *et al.* (2003) proposed the elimination of approximately 500 of the putative *S. cerevisiae* genes and identified 43 genes not annotated in the initial published sequence. More recently, the same group published clear evidence for ancient genome duplication and gene loss in *S. cerevisiae* by a detailed comparison with genes in the genome of the related yeast *Kluyveromyces waltii* (Kellis *et al.* 2004).

It is generally more difficult to predict open reading frames of genes in filamentous fungi than it is in yeasts. This is due in part to their more complex gene structure (more and larger introns, alternative splicing of introns). In addition, less is

also known about the regulatory elements controlling gene expression than in yeasts. It is partly for this reason that the FGI is currently sequencing the genomes of *Neurospora* 'helper' genomes from *Podospora anserina* and *Chaetomium globosum*, fungi that are close enough evolutionarily to *Neurospora* to allow the identification of orthologous regions (regions of DNA encoding genes with similar functions) yet also distantly related enough to allow the identification of putative regulatory sequences by their conserved nature across each of the genomes.

6. Filamentous fungal genomes in the spotlight

With the current productive status of fungal genomics that has yielded many recent exceptional publications and intriguing observations, it is somewhat difficult to believe that at the end of 2002, there were only three fungal genomes published. This explosion in fungal genome projects is partly due to technological improvements in genome sequencing. But also now that the human genome project is in the polishing phase, some sequencing centres and funding agencies are switching focus. For example, the JGI, which was responsible for the sequence of three human chromosomes, has now changed to a more environmental focus. The Washington University Medical School Genome Sequencing Center contributed a quarter of the finished human genome sequence and is now running a comprehensive *Saccharomyces* comparative genomics program. The largest player in the human genome project, the Broad Institute, is hosting the FGI and also has a *Saccharomyces* comparative genomics program. With such large sequencing centres in support, fungal genomics should continue to prosper.

R E F E R E N C E S

Bennett JW, 1997a. White paper: genomics for filamentous fungi. *Fungal Genetics and Biology* **21**: 3–7.
Bennett JW, 1997b. Open letter to fungal researchers. *Fungal Genetics and Biology* **21**: 2.
Birren B, Fink G, Lander ES, 2002. *Fungal Genome Initiative, White Paper Developed by the Fungal Research Community*. Whitehead Institute/MIT Center for Genome Research, Cambridge, MA 02141, USA. http://www.broad.mit.edu/annotation/fungi/fgi/history.html.
Birren B, 2003. *Fungal Genome Initiative, A White Paper for Fungal Comparative Genomics*. Whitehead Institute/MIT Center for Genome Research, Cambridge, MA 02141, USA. http://www.broad.mit.edu/annotation/fungi/fgi/history.html.
Birren B, Fink G, Lander ES, 2003. *Fungal Genome Initiative, White Paper for Fungal Comparative Genomics*. Whitehead Institute/MIT Center for Genome Research, Cambridge, MA 02141, USA. http://www.broad.mit.edu/annotation/fungi/fgi/history.html.
Birren B, 2004. *Fungal Genome Initiative, A White Paper for Fungal Genomics*. The Broad Institute of Harvard and MIT, Cambridge, MA 02141 USA. http://www.broad.mit.edu/annotation/fungi/fgi/history.html.
Dean RA, Talbot NJ, Ebbole DJ, *et al.*, 2005. The genome sequence of the rice blast fungus *Magnaporthe grisea*. *Nature* **434**: 980–986.

Galagan JE, Calvo SE, Borkovich KA, *et al.*, 2003. The genome sequence of the filamentous fungus *Neurospora crassa*. *Nature* **422**: 859–868.

Giaever G, Chu AM, Ni L, *et al.*, 2002. Functional profiling of the *Saccharomyces cerevisiae* genome. *Nature* **418**: 387–391.

Goffeau A, Barrell BG, Bussey H, *et al.*, 1996. Life with 6000 genes. *Science* **274**: 546, 563–567.

Goodwin SB, 2004. Minimum phylogenetic coverage: an additional criterion to guide the selection of microbial pathogens for initial genomic sequencing efforts. *Phytopathology* **94**: 800–804.

Hamer L, 1996. From genes to genomes: sequencing of filamentous fungal genomes. *Fungal Genetics and Biology* **21**: 8–10.

Hofmann G, McIntyre M, Nielsen J, 2003. Fungal genomics beyond *Saccharomyces cerevisiae? Current Opinion in Biotechnology* **14**: 226–231.

Horak CE, Snyder M, 2002. Global analysis of gene expression in yeast. *Functional and Integrative Genomics* **2**: 171–180.

Hynes MJ, 2003. The *Neurospora crassa* genome opens up the world of filamentous fungi. *Genome Biology* **4**: 217.

Katinka MD, Duprat S, Cornillot E, *et al.*, 2001. Genome sequence and gene compaction of the eukaryote parasite *Encephalitozoon cuniculi*. *Nature* **414**: 450–453.

Kellis M, Patterson N, Endrizzi M, Birren B, Lander ES, 2003. Sequencing and comparison of yeast species to identify genes and regulatory elements. *Nature* **423**: 241–254.

Kellis M, Birren BW, Lander ES, 2004. Proof and evolutionary analysis of ancient genome duplication in the yeast *Saccharomyces cerevisiae*. *Nature* **428**: 617–624.

Lander ES, Linton LM, Birren B, *et al.*, 2001. Initial sequencing and analysis of the human genome. *Nature* **409**: 860–921.

Leach JE, Gold S, Tolin S, Eversole K, 2002. Plant-Associated Microbe Genome Initiative. White paper. American Phytopathological Society, Saint Paul, MNS5121, USA. http://www. apsnet.org/members.ppb/PDFs/Final/PlantAssociatedmicro-beGenomeInitiative.pdf.

Lorenz MC, 2002. Genomic approaches to fungal pathogenicity. *Current Opinion in Microbiology* **5**: 372–378.

Martinez D, Larrondo LF, Putnam N, *et al.*, 2004. Genome sequence of the lignocellulose degrading fungus *Phanerochaete chrysosporium* strain RP78. *Nature Biotechnology* **22**: 695–700.

Rouxel T, Balesdent MH, 2005. The stem canker (blackleg) fungus, *Leptosphaeria maculans*, enters the genomic era. *Molecular Plant Pathology* **6**: 225–241.

Scott B, 2003. The evolution of gene clusters in plant-associated filamentous fungi. *New Zealand BioScience* **May**: 22–26.

Tanaka A, Tapper BA, Popay A, Parker EJ, Scott B, 2005. A symbiosis expressed non-ribosomal peptide synthetase from a mutualistic fungal endophyte of perennial ryegrass confers protection to the symbiotum from insect herbivory. *Molecular Microbiology* **57**: 1036–1050.

Wood V, Gwilliam R, Rajandream MA, *et al.*, 2002. The genome sequence of *Schizosaccharomyces pombe*. *Nature* **415**: 871–880.

Zeng Q, Morales AJ, Cottarel G, 2001. Fungi and humans: closer than you think. *Trends in Genetics* **17**: 682–684.

Erratum to Taylor and Alexander (2005) *Mycologist* 19: 102-112

On page 106 of the article by Taylor and Alexander (Mycologist **19**: 102-112) line 4/5 should have read 'By the same line of reasoning, the confirmation that *Pakaraimaea*, the S American genus, is also ectomycorrhizal (B. Moyersoen, unpublished data) sets ECM origin in the clade back to 130 million years ago'.