Medical Mycology Supplement 1 2005, **43**, S3–S5

Taylor & Francis
Taylor & Francis Group

What the *Aspergillus* genomes have told us

W. C. NIERMAN*, G. MAY†, H. S. KIM*, M. J. ANDERSON‡, D. CHEN* & D. W. DENNING§

**Institute for Genomic Research, Rockville, MD, USA, and †University of Texas, M.D. Anderson Cancer Center, Houston, TX, USA, and ‡University of Manchester, Manchester, UK, and §Education and Research Center, Wythenshawe Hospital, University of Manchester, Manchester, UK*

The sequencing and annotation of the genomes of the first strains of *Aspergillus nidulans*, *Aspergillus oryzae*, and *Aspergillus fumigatus* will be seen in retrospect as a transformational event in Aspergillus biology. With this event the entire genetic composition of *A. nidulans*, the sexual experimental model organism of the genus Aspergillus, *A. oryzae*, the food biotechnology organism which is the product of centuries of cultivation, and *A. fumigatus*, the most common causative agent of invasive aspergillosis is now revealed to the extent that we are at present able to understand. Each genome exhibits a large set of genes common to the three as well as a much smaller set of genes unique to each. Moreover, these sequences serve as resources providing the major tool to expanding our understanding of the biology of each. Transcription profiling of *A. fumigatus* at high temperatures and comparative genomic hybridization between *A. fumigatus* and a closely related Aspergillus species provides microarray based examples of the beginning of functional analysis of the genomes of these organisms going forward from the genome sequence.

Keywords *Aspergillus*, genome, aspergillosis, microarray

Introduction

Invasive infection by *Aspergillus fumigatus* is now the most common cause of death among leukemia patients and haematopoietic stem cell transplant recipients [1,2]. The genome of the fully pathogenic clinical *A. fumigatus* isolate Af293 [3] was sequenced by the whole genome random sequencing method [4] and augmented by optical mapping for sequence assembly confirmation [5]. The completed Af293 genome consists of eight chromosomes with estimated optical map sizes of: Chromosome 1, 4.9 Mb; 2, 4.8 Mb; 3, 4.0 Mb; 4, 3.9 Mb; 5, 3.9 Mb; 6, 3.6 Mb; 7, 2.0 Mb; and 8, 1.8 Mb, for a total size of 29.2 Mb. This nuclear genome contains 9,922 identified protein coding sequences (CDS), of which about 60% (5,969 CDS) code for the proteins that contain PFAM domains or are assigned a gene ontology (GO) function based on homology to a *Saccharomyces cerevisiae* protein. Almost half (4,490 CDS) the code for proteins belong to one of the known protein families in the Protein families database of alignments and HMMs (PFAM, http://www/sanger.ac.uk/Software/Pfam/). The 31,892 bp mitochondrial genome contains 16 protein-coding genes and 25 tRNAs. Analysis of the Af293 genome itself and of its comparison to other sequenced filamentous fungi has revealed interesting aspects of genomic infrastructure and regulatory circuits for pathogenicity, production of secondary metabolites, sexual reproduction, and other aspects of the biology of *A. fumigatus*.

How reliable is the annotation of the genome?

An issue in using the sequenced and annotated genome of *A. fumigatus* is assessing the quality of the largely automated annotation of this genome. Prior to initiating the project to sequence the *A. fumigatus* genome, a pilot project was conducted at the Wellcome Trust Sanger Institute, involving the sequencing of a 1 Mb region of this genome and performing manual

Correspondence: W. C. Nierman, Institute for Genomic Research, 9712 Medical Center Drive, Rockville, MD 20850, USA. Tel: +1 301 795 7559; Fax: +1 301 838 0209; E-mail: wnierman@tigr.org

DOI: 10.1080/13693780400029049

annotation of the genes identified in this sequenced region [3]. We compared our auto-annotated version of this selected region to the Sanger annotation, which included 360 genes identified by the Sanger effort. Only 137 of the 360 genes had the same gene structure in both annotations of this region. In comparing orthologous genes in *A. fumigatus* to those in the *Aspergillus nidulans* and *Aspergillus oryzae* genomes, numerous examples of non-identical gene models were noted in all gene annotations generated by independent auto-annotation pipelines at different sequencing centers. However, no attempt was made to quantify these data. The conclusion from this analysis is that auto-annotation for fungal genomes is not yet highly reliable for the details of genes structure; users of auto-annotation data need to be aware of this. This issue is particularly important in designing oligonucleotides for spotting onto glass slide microarrays for use in expression profiling experiments. Where uncertainty exists in coding sequence start and stop sites and introns/exon boundaries, the selected oligo could well be located such that it is not present in the mature mRNA for that gene. To avoid this possibility, we employed a PCR amplicon strategy in designing our microarray. Briefly, we selected a 700 bp region immediately upstream of the predicted stop codon from each gene. In the case the gene is smaller than 700 bp, we took the entire gene. Then, we included 150 bp of sequence downstream of the gene or as much as there is in the intergenic region when it is shorter than 150 bp. These comprise the target sequences which provide with a minimum of 850 bp for the genes. We conducted automated selection of PCR primer pairs by feeding the target sequences to Primer 3.0 (Whitehead Institute, http://www-genome.wi.mit.edu/genome_software) with optimized design parameters that can be used to amplify greater than 5/6 of the targets. The resulting PCR products are on average 710 bp in length. Using this approach we were able to design primers for 9,516 of then 9,544 predicted genes (99.7%). We amplified these target gene regions from genomic DNA. The resulting PCR products were purified and spotted in triplicates at high density on Corning (Acton, MA) UltraGAPS™ aminosilane-coated microscope slides using a robotic spotter built by Intelligent Automatic Systems (IAS) (Cambridge, MA) and cross-linked by ultraviolet illumination.

Thermotolerance of *A. fumigatus*

Thermotolerance of *A. fumigatus* is a trait critical to its thriving in the mammalian host and during the higher temperature phase of compost degradation. The sequenced strain can grow at 51°C and can survive as hyphae at temperatures as high as 70°C. To investigate the adaption process of this fungus to higher temperature, we examined gene expression using a whole genome microarray at three temperatures, 30°C, 37°C, and 48°C. These temperatures represent the environments of tropical soil, the human body, and compost, respectively. Cultures inoculated with conidia were grown at 30°C for 16 h, and then either kept at the same temperature or moved to a growth chamber at 37°C or 48°C. During this further incubation, samples were taken at various time points. We profiled gene expression from each time point relative to that from the time of temperature shift using a whole genome microarray constructed with PCR amplicons representing each gene in the genome. We found that a number of genes with various functional roles are expressed differentially at each temperature. In addition, more genes are down-regulated than up-regulated as a consequence of the temperature shift from 30°C to 48°C, which is in contrast to the situation at 37°C where more genes are up-regulated. This suggests that the fungus is deploying more diverse activities at 37°C than it does at 48°C. We examined the expression of selected genes, whose functional annotation suggests their involvement in a sub-set of specific processes, including heat shock response and pathogenicity. Interestingly, but perhaps not unexpectedly, more heat shock and stress-responsive genes are highly expressed at 48°C than at 30°C or 37°C, indicating that the fungus is under heat stress at this temperature. In contrast, more putative pathogenicity genes have increased expression at 37°C relative to the other two temperatures, although there is no contact with mammalian cells or cellular products. Examples are those coding for the proteins responsive to oxidative stress, the host immune system, and for toxin production. We suggest from this study that temperature is a key environmental signal for *A. fumigatus* which triggers gene regulation cascades that may ultimately lead to adaptation to specific new environments. Another notable finding is that many transposases, especially those of Mariner-4 type, are highly expressed at 48°C, suggesting that higher temperatures may induce transposition events. To the best of our knowledge, no thermal induction of transposases has been reported in fungi.

How useful are other *Aspergillus* genomes as comparators?

The ongoing analysis of the *A. fumigatus* genome has made good use of the sequences of the *A. nidulans* and

A. oryzae genomes in comparisons to study gene and genome evolution among these species. However, these species are not ideal for the most effective comparative genomic analysis for specific biological processes, including pathogenesis. In fact, as is clear from the *Aspergillus* large subunit rDNA phylogenetic tree, *A. nidulans* and *A. oryzae* are distantly related to *A. fumigatus*. In this light it is not surprising that only approximately 50% of each genome can be aligned with the corresponding region of each of the other two. Only 7,500 of the genes in *A. fumigatus* have orthologs in either *A. oryzae* or in *A. nidulans* and the number of orthologs across all three is only 5,899. The average protein identity of these identified orthologs is 70% with the match extending on average only 86% of the protein length. Although much of the low quality of matches among the candidate orthologs may be due to the artifactual variation in autoannotation gene models, this lack of orthology limits the ability to form meaningful associations of the gene content and phenotype (i.e., pathogenicity, etc.) of *A. fumigatus* relative to the other two. In contrast, the close taxonomic relationship of the potentially minimally pathogenic *Neosartorya fischeri* to *A. fumigatus* make them an extremely useful pair for comparative studies focused on pathogenesis and other biological processes, including the sexual cycle. These species have been defined as sister taxa within the section Fumigati. A pilot sequencing project of *N. fischeri* has revealed an average DNA identity of approximately 90% to *A. fumigatus* based on 4X sequence coverage contigs. *N. fischeri* also possesses a known sexual cycle, in contrast to *A. fumigatus* that has never been reported to undergo sexual reproduction. We have used genomic DNA from these two species and our ORF microarray to perform comparative genomic hybridization (CGH) between the two genomic DNAs. The analysis revealed 700 genes to be not present or diverged in *N. fischeri* NRRL181 relative to *A. fumigatus* Af293. These include 13 genes coding for the enzymes involved in production of secondary metabolites, 28 genes coding for transcriptional regulators and protein kinases, 21 coding for transporters, 199 coding for metabolic and other proteins, and 400 coding for hypothetical proteins. This number of genes is a manageable set to begin the effort of correlating phenotypic differences between these species to gene content. We have initiated CGH analysis with another non-pathogenic, more distantly related species, *Neosartorya fennelliae* NRRL5534 with the expectation of further reducing the number of diverged or missing genes that can be correlated with phenotypic differences.

The next step

Resources for comparative genomic analysis of species that cause invasive aspergillosis (IA) are about to expand significantly. Multiple genome sequence and analysis projects are underway or about to be initiated. These include a comparative analysis of *A. fumigatus* Af293 and CEA10, sequencing of *Aspergillus flavus* (the second most prevalent species causing IA), *s*equencing of *Aspergillus. terreus* (the third most prevalent species causing IA), and sequencing of the allergenic *Aspergillus clavatus* which is a close relative that phylogenetically clusters between *A. fumigatus* and *A. terreus*. We will also be completing the sequencing of *N. fischeri.*

References

1 Martino R, Subira M, Rovira M, *et al*. Invasive fungal infections after allogeneic peripheral blood stem cell transplantation: incidence and risk factors in 395 patients. *Br J Haematol* 2002; **116**: 475–482.

2 Estey E. Reducing mortality associated with immediate treatment complications of adult leukemias. *Semin Hematol* 2001; **38**(Suppl 10): 32–37.

3 Pain A, Woodward J, Quail MA, *et al*. An insight into the genome of the most abundant opportunist human pathogen: analysis of 922 kb fragment of chromosome VIII of *Aspergillus fumigatus*. *Fung Genet Biol* 2004; **41**: 443–453.

4 Nierman WC, Feldblyum TV, Laub MT, *et al*. Complete genome sequence of *Caulobacter crescentus*. *Proc Natl Acad Sci USA* 2001; **98**: 4136–4141.

5 Zhou S, Kvikstad E, Kile A, *et al*. Whole-genome shotgun optical mapping of *Rhodobacter sphaeroides* strain 2.4.1 and its use for whole-genome shotgun sequence assembly. *Genome Res* 2003; **13**: 2142–2151.