Available online at www.sciencedirect.com
SCIENCE DIRECT®
Fungal Genetics and Biology 41 (2004) 443–453
FUNGAL GENETICS AND BIOLOGY
www.elsevier.com/locate/yfgbi

Insight into the genome of *Aspergillus fumigatus*: analysis of a 922 kb region encompassing the nitrate assimilation gene cluster

Arnab Pain,[a] John Woodward,[a] Michael A. Quail,[a] Michael J. Anderson,[b] Richard Clark,[a] Matthew Collins,[a] Nigel Fosker,[a] Audrey Fraser,[a] David Harris,[a] Natasha Larke,[a] Lee Murphy,[a] Sean Humphray,[a] Susan O'Neil,[a] Mihaela Pertea,[c] Claire Price,[a] Ester Rabbinowitsch,[a] Marie-Adele Rajandream,[a] Steven Salzberg,[c] David Saunders,[a] Kathy Seeger,[a] Sarah Sharp,[a] Tim Warren,[a] David W. Denning,[b] Bart Barrell,[a] and Neil Hall[a,*]

[a] *The Pathogen Sequencing Unit, The Wellcome Trust Sanger Institute, Genome Campus, Hinxton, Cambridge CB10 1SA, UK*
[b] *School of Medicine, University of Manchester, Oxford Road, Manchester M13 9PT, UK*
[c] *The Institute of Genome Research, 9712 Medical Centre Drive, MD 20850, USA*

Received 29 September 2003; accepted 5 December 2003

Abstract

Aspergillus fumigatus is the most ubiquitous opportunistic filamentous fungal pathogen of human. As an initial step toward sequencing the entire genome of *A. fumigatus*, which is estimated to be ~30 Mb in size, we have sequenced a 922 kb region, contained within 16 overlapping bacterial artificial chromosome (BAC) clones. Fifty-four percent of the DNA is predicted to be coding with 341 putative protein coding genes. Functional classification of the proteins showed the presence of a higher proportion of enzymes and membrane transporters when compared to those of *Saccharomyces cerevisiae*. In addition to the nitrate assimilation gene cluster, the quinate utilisation gene cluster is also present on this 922 kb genomic sequence. We observed large scale synteny between *A. fumigatus* and *Aspergillus nidulans* by comparing this sequence to the *A. nidulans* genetic map of linkage group VIII.

Keywords: *Aspergillus fumigatus*; Gene-cluster; BAC-end sequencing; Genome; Annotation; Comparative genomics; Synteny

1. Introduction

Aspergillus is a remarkably diverse genus of filamentous fungi, representing 182 recognised species (Pitt et al., 2000). *Aspergillus fumigatus* is among the most common of human and animal opportunistic fungal pathogens and is recognised as a major fungal saprophyte, playing an essential role in recycling carbon and nitrogen (Pitt, 1994). It is a ubiquitous thermophilic species, found in decaying vegetation, in the air, in hospital water systems and in human dwellings worldwide (Anaissie et al., 2002; Latge, 1999).

This filamentous fungus has a simple life-cycle: it reproduces asexually, generating abundant haploid conidia from aerial conidiophores, and has no known sexual cycle. The innate immune mechanisms (particularly macrophage phagocytosis) efficiently eliminate the conidia in the immunocompetent host after inhalation, with no pulmonary disease symptoms. However, the status of *A. fumigatus* as a relatively harmless saprophytic fungus has changed dramatically over the last two decades (Marr et al., 2002). *A. fumigatus* now represents the most prevalent air-borne opportunistic fungal pathogen in developed countries, causing severe and often fatal invasive infection in immunocompromised people (Anderson et al., 2003; Marr et al., 2002). The frequency of invasive aspergillosis (IA) has increased many folds over the past 20 years. About 4% of patients dying in European teaching hospitals suffer from IA and it is the leading

* Corresponding author. The Institute for Genomic Research, 9712 Medical Center Drive, Rockville MD 20850, USA. Fax: +01 301 838-0208.
E-mail address: nhall@tigr.org (N. Hall).

1087-1845/$ - see front matter © 2003 Published by Elsevier Inc.
doi:10.1016/j.fgb.2003.12.003

infectious cause of death in leukaemia and bone marrow transplant patients (Groll et al., 1996; Vogeser et al., 1997). In addition, allergic bronchopulmonary aspergillosis occurs in asthmatic and cystic fibrosis patients (Marr et al., 2002).

Until 2001, only two licensed antifungal drugs were available to treat aspergillosis: amphotericin B and itraconazole (Marr et al., 2002). In that year, two additional agents were licensed—voriconazole and caspofungin (Johnson and Kauffman, 2003; Johnson and Perfect, 2003). The latter represents the first of a new generation of drugs for invasive fungal infections, the candins which inhibit the glucan synthase enzyme complex in the cell wall. However response rates to these agents are still only ~50%. Thus, there is an immediate need to identify biochemical pathways that are critical for the pathogenicity and survival of *A. fumigatus*, which would act as potential targets for new antifungal compounds.

The pathogenicity of *A. fumigatus* is poorly understood and there is a limited knowledge of the basic biology of the organism (Anderson et al., 2003; Latge, 2001). *A. fumigatus* (AF293 strain) has a haploid nuclear genome of ~30 Mb, distributed over an estimated 8 chromosomes, ranging from 1.7 to ~5.3 Mb in size (Anderson et al., 2003). The sequences of only ~40 genes were available in the public databases before the start of this project, representing 0.4% of the 10–12,000 estimated total number of *A. fumigatus* genes.

In 2001, a whole genome shotgun approach was adopted by an international consortium to sequence the genome of *A. fumigatus* with The Wellcome Trust Sanger Institute (UK), The Institute for Genomic Research (USA), Universities of Manchester (UK), Salamanca (Spain) and Complutense (Spain), Centro de Investigaciones Biologicas (Spain), and the Pasteur Institute (France) as the collaborating centres (Denning et al., 2002).

Ahead of the inception of the whole genome sequencing project, the Wellcome Trust Sanger Institute, in collaboration with the University of Manchester, undertook a pilot project to construct a genomic Bacterial Artificial Chromosome (BAC) library from the DNA of a clinical isolate of *A. fumigatus* and to fingerprint the clones to obtain a physical map. A physically linked group of BAC clones were sequenced, and here we report the analysis of the resulting 922 kb contig. This represents the first large-scale analysis of the genomic content of this opportunistic pathogen and provides a snapshot of the genome, before the whole genome sequence becomes available.

2. Materials and methods

2.1. Characterisation of an A. fumigatus clinical isolate

The strain chosen for constructing the BAC library is AF293, a clinical isolate of *A. fumigatus*. It was isolated by the Public Health Laboratory of The Royal Shrewsbury Hospital, Shrewsbury, UK, in 1993 from lung tissue taken at autopsy from a patient who had developed a low white cell count after gold therapy for rheumatoid arthritis. The isolate was confirmed to be *A. fumigatus* Fresenius by the National Collection of Pathogenic Fungi (NCPF 7367) at the Mycology Reference Laboratory, Bristol, UK and by the Centraal Bureau voor Schimmelcultures (CBS 101355), Baarn, The Netherlands. Minimum inhibitory concentrations were determined for the following antifungal drugs: amphotericin B 0.25 mg/L; itraconazole, 0.50 mg/L, and voriconazole 1.0 mg/L. Four subcultures were obtained from single spores of the clinical isolate. Nucleic acids were extracted from a culture grown after one subculture from our oldest frozen stock and from the four subcultures obtained from the single spores. These cultures were typed using microsatellite length polymorphisms (Bart-Delabesse et al., 1998) and they were shown to be identical. Single-spore culture 1 (deposited at the Fungal Genetics Stock Center as A1100) was used as the source of nucleic acids in all subsequent experiments. The DNA sequence of the *asp f 212* and *rodA12* regions from this culture were confirmed to be typical for *A. fumigatus* (Geiser et al., 1998).

2.2. Preparation of chromosomal DNA in agarose plugs

Chromosomal DNA in agarose plugs was prepared following a modified protocol (Tobin et al., 1997). In short, AF293 conidia (4×10^6/ml) were inoculated into Sabouraud liquid medium (Oxoid, Basingstoke, UK) at 37 °C, shaking at 225 rpm, for 16 h. The mycelium was harvested by vacuum filtration and resuspended in PB buffer (1 M NaCl; 0.01 M $MgCl_2$; pH 5.8) to give 40 mg/ml wet weight. Novozyme 234 (InterSpex Products, San Mateo, CA, USA) was added to a final concentration of 3.5 mg/ml and the preparation was incubated at 37 °C for 2 h. After filtration through household J-Cloth, the protoplasts were collected by centrifugation at 800*g* for 10 min, washed once in PB buffer, and resuspended in 0.9 M sorbitol; 0.125 M EDTA; 0.01 M Tris–Cl, pH 7.5, containing 0.5% InCert agarose (Cambrex Bio Science Wokingham, UK) to a final concentration of approximately 5×10^8/ml. The mixture was added to disposable plug moulds (Bio-Rad Laboratories, Hemel Hempstead, UK) and allowed to set at 4 °C. The plugs were incubated for 24 h at 50 °C in 0.5 M EDTA; 0.01 M Tris–Cl, pH 8.0; 2% Sarkosyl; 2 mg/ml proteinase K with one change of the solution and washed in 0.5 M EDTA.

2.3. BAC library construction and identification of nia D-positive clones

The BAC library construction was based primarily on a previously described method (Osoegawa et al., 1998).

Agarose plugs were partially digested with *Sau*3AI and the fragments were separated by pulsed-field gel electrophoresis. Fragments running at 100–150 kb were eluted and cloned into pBACe3.6 (Frengen et al., 1999). A library of approximately 30-fold coverage resulted with an average insert size of 75 kb. Random clones were picked for end-sequencing, *Pst*I fingerprinting and gridding onto nylon membranes. The nylon membranes were probed with the *niaD* gene to identify positive clones, which were used as the starting point for contig-building. The *niaD* probe was prepared from a PCR product obtained from *A. fumigatus* genomic DNA with the primer pair 5′-ATG TACTGGTCCGTTCCGTTCTCGG-3′ and 5′-TCGG GTTCTGCCAATATCTC-3′, using optimised PCR conditions. Vector preparation, ligation, transformation, colony picking, filter membrane gridding, and hybridisations were otherwise carried out either according to the manufacturer's instructions or as previously described (Sambrook et al., 1989).

2.4. Sequencing and contig assembly

BAC sequencing and assembly were essentially done as previously described (Harris and Murphy, 2001). The *niaD*-positive clones were skimmed (shotgun sequenced to 2–4-fold coverage); clone AfA5C11 was sequenced to 8-fold coverage and finished. The sequence of AfA5C11 was compared against the BAC end-sequences and overlapping clones (AfA8D5 and AfA6E3) were identified for sequencing. By further rounds of this walking strategy, a BAC contig of ~922 kb consisting of 16 overlapping clones, was produced.

2.5. Sequence analysis and annotation

Annotation was performed using Artemis software (Rutherford et al., 2000). Genes were identified by manual curation of the output of the gene finding software. Gene finding software used included Genefinder (P. Green, unpublished work) and GlimmerM (Majoros et al., 2003; Salzberg et al., 1999) trained for *A. fumigatus* (unpublished). The final set of genes was assigned unique identifiers and these identifiers have been used to denote specific gene(s) in the rest of this paper. Functional assignments were based on assessment of BLAST and FASTA searches against public databases and for domain predictions, searches such as InterPro (Apweiler et al., 2000); TMHMMv2.0 (Krogh et al., 2001); SignalPv2.0 (Nielsen et al., 1999); and t-RNA scans (Lowe and Eddy, 1997) were carried out. Gene Ontology terms (Ashburner et al., 2000) were manually assigned to all predicted gene products where possible. First, candidate GO terms were selected by sequence similarity-searching in a database of protein sequences and their previously assigned GO terms, drawn from the following databases: Flybase (http://flybase.bio.indiana.edu/), Mouse Genome Informatics (http://www.informatics.jax.org/), *Saccharomyces* Genome Database (http://genome-www.stanford.edu/Saccharomyces/), Swiss-Prot (http://www.ebi.ac.uk/swissprot/), and the *Arabidopsis* Information Resource (http://www.arabidopsis.org/). After inspection of sequence alignments, suitable terms were either assigned directly from the candidate list or alternatively, higher or lower granularity terms were selected directly from the ontology. When previously characterised genes were identified, terms were selected as above, but alternative experimental evidence codes were used to reflect the fact that the inferences were no longer based on sequence similarity. The comparison of the quinate utilisation gene cluster regions between *A. fumigatus*, *Aspergillus nidulans*, *Neurospora crassa*, and *Podospora anserina* was carried out using BLAST comparisons viewed with the Artemis Comparison Tool (K. Rutherford, unpublished). The sequence and the annotation described here have been submitted to EMBL with the following Accession No. BX649216.

3. Results and discussion

3.1. Assembly of a 922 kb chromosomal DNA contig by end-sequence walking

Sixteen BAC clones were shotgun sequenced to 8-fold coverage and finished. A further 12 clones were skimmed to examine the quality of the BAC library. The total length of the assembled contig was 921,536 bp. The assembled contig mapped to the largest genomic sequence scaffold (representing part of the 5.3 Mb largest chromosome of *A. fumigatus*) in the data available to date from the *A. fumigatus* whole genome sequencing project. Comparison of the sequences of the BAC clones to the whole genome shotgun sequence data reveals two small deletions in some of the BAC clones. There is a 65 bp sequence in the overlap region between clones AfA28D1 and AfA14E1, which is not present in AfA28D1. This sequence is present in the assembled contig between basepair positions 246,313 and 246,377. There is an exact 5 bp sequence (GATCC) just before and after the deletion. In addition, there is a 123 bp sequence not present in AfA35G10 but present in the genome shotgun sequence data. The sequence is between basepair positions 708,124 and 708,125 in the contig. The situation is very similar to the 65 bp deletion in that there is an exact 6 bp sequence (TGATCT) just before and after the deletion.

3.2. Overview of the annotation

The overall GC-percentage of the DNA is 50.6 with a coding percentage of 54.0. We have predicted a total of 341 putative protein-coding genes and 8 t-RNA genes

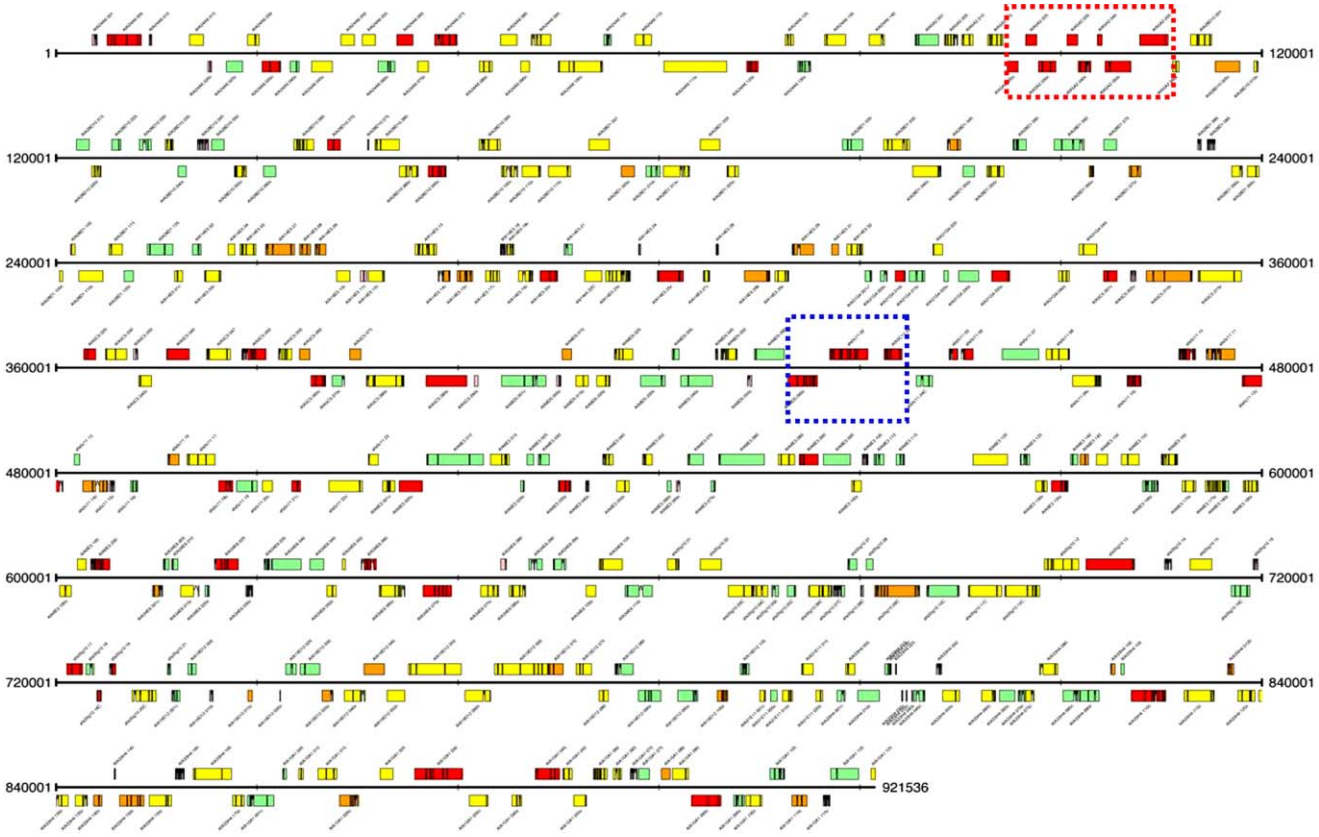

Fig. 1. Colour-coded annotation of the assembled contig of *A. fumigatus*, showing location of the putative genes. The genes, coloured red, indicate orthologues of those genes that have been previously identified in one or more *Aspergillus* species. The genes, coloured yellow, indicate genes whose function has been putatively assigned on the basis on a significant protein or domain hit. The orange and green coloured genes indicate conserved hypothetical and hypothetical genes, respectively. The unique gene ids for each gene are also shown. The location of *A. fumigatus qut* gene cluster and the *niaD-niiA-crnA* gene clusters are shown by the red and blue boxes, respectively.

(Fig. 1). The details of our analysis are given in Table 1. The average gene density is 0.37/kb (which is equivalent to 1 gene every 2.7 kb) of sequence with an average of 3 exons per gene and a mean gene length (excluding the introns) of 1461 bp, which is comparable to *Saccharomyces cerevisiae* (1424 bp) (Goffeau et al., 1996) and *Dictyostelium discoideum* (1626 bp) (Glockner et al., 2002). Introns were predicted in 80% of genes, a proportion roughly similar to that of *Arabidopsis thaliana* (79%) but higher than that observed in *N. crassa* (72%), *D. discoideum* (68%), *Plasmodium falciparum* (54%), *Schizosaccharomyces pombe* (43%), and much higher than *S. cerevisiae* (5%). In addition, we have predicted 8 t-RNA genes, 5 of which are in an array and all of which contain a single intron. We did not find any r-RNA gene arrays or transposable elements in the chromosomal region that we sequenced, although we have predicted 1 gene, coding for a putative transposase (AfA35G10.17). The region of the genome discussed in this report does not contain telomeric or centromeric sequences.

Only 16% of the 341 annotated genes are previously known in the Aspergilli and about one-third of the proteins appear not to have an orthologue in *N. crassa* (Table 2). A function could not be inferred for 38% of the predicted proteins; however, quarter of these proteins have homologues in one or more organism(s) and therefore were annotated as conserved hypothetical proteins. About 21% of the total predicted proteins have 1 or more transmembrane domain(s) and a similar percentage (20.5%) are predicted to contain a putative signal peptide or signal anchor. Analysis of top five significant FASTA hits to the functionally annotated proteins (i.e., proteins with a predicted function) and the conserved hypothetical proteins revealed that, in several cases, corresponding homologues are present only in bacteria or in plants.

3.3. Functional classification of predicted gene products

A total of 187 Pfam (protein families of alignments and HMMs) domains (Bateman et al., 2002) in 153 predicted proteins were identified. Zinc-finger domains were the most abundant Pfam domain and we have identified eight proteins either with fungal Zn(2)–Cys(6) binuclear cluster domains (PF00172) or C2H2 type Zn-finger domains (PF00096) (Table 3). Enzyme Classification (E.C.)

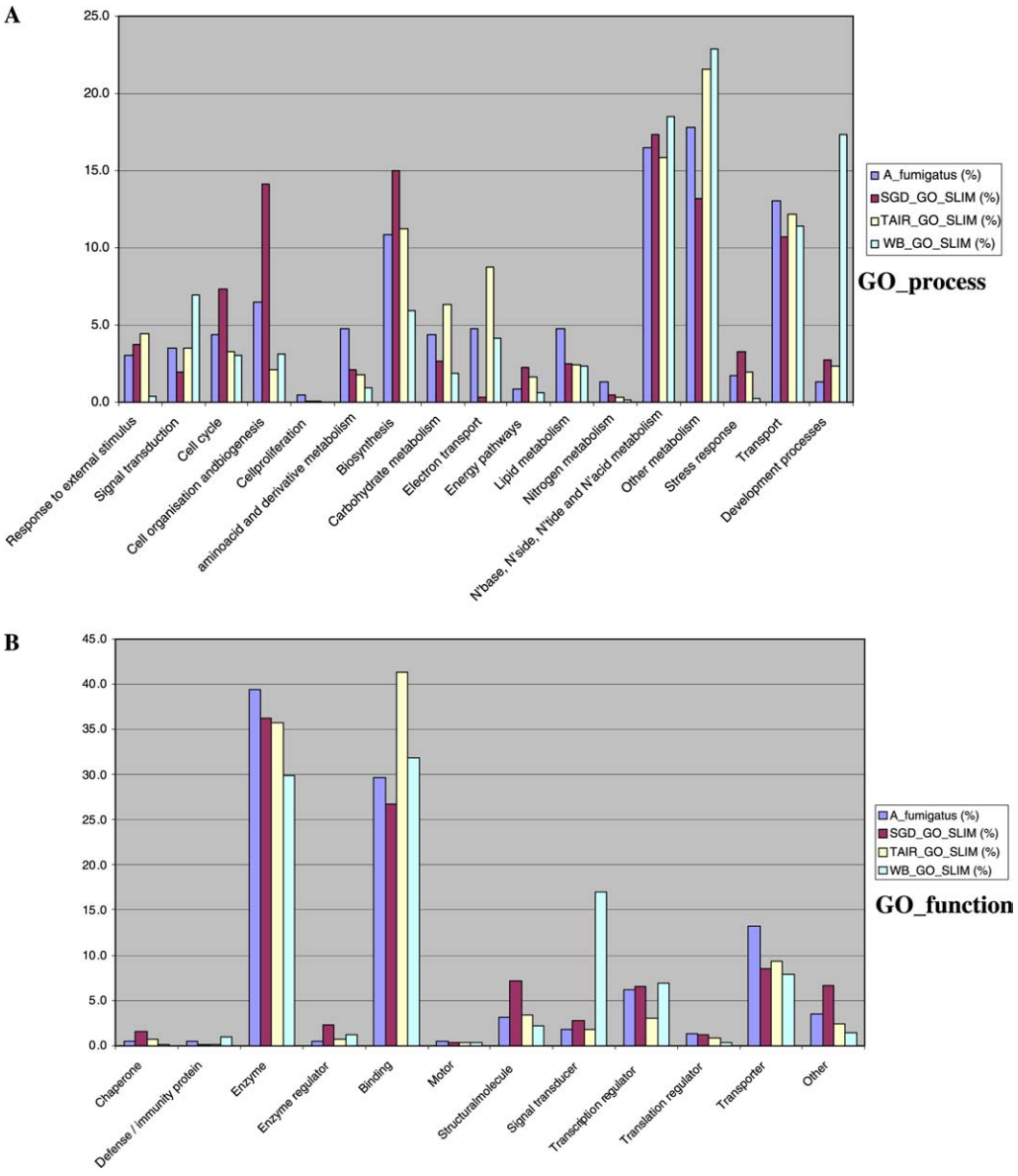

Fig. 2. Gene Ontology (GO) classification of predicted gene products of *A. fumigatus* and comparison with similar GO assignments in predicted gene products of budding yeast *S. cerevisiae*, represented as SGD_GO_SLIM (%); worm *C. elegans*, represented as WB_GO_SLIM (%), and a plant *A. thaliana*, represented as TAIR_GO_SLIM(%). The values indicate the percentages and the data for yeast, worm and *Arabidopsis* came from *Saccharomyces* Genome Database (SGD), Worm Base (WB), and The *Arabidopsis* Information Resource (TAIR), respectively. The genes predicted to code for hypothetical proteins have been excluded from these calculations.

numbers were assigned to 74 putative enzymes identified on the contig, representing 22% of the total predicted proteins.

We have also classified the predicted proteins according to the Gene Ontology (GO) terms for biological function, biological process, and cellular compartment. The GO system is a controlled vocabulary that describes the role(s) of gene products in organisms (Ashburner et al., 2000). GO terms were assigned manually to 177 (52%) of the gene products and the comparison of GO term assignments with higher level (i.e., less granular) GO terms in yeast (*S. cerevisiae*), worm (*Caenorhabditis elegans*) and a plant (*A. thaliana*) is shown in Fig. 2. We observed a slightly higher proportion of functionally annotated genes in *A. fumigatus* that are predicted to function as enzymes (39%) or transporters (13%). For example, in budding yeast (*S. cerevisiae*) and worm (*C. elegans*), the percentage of proteins annotated as enzymes is 36 and 30%, respectively; the same trend is observed in the total number of membrane transporters (8% of proteins annotated as transporters in both organisms) (Fig. 2). In addition, the proportion of proteins involved in secondary metabolism was higher (18%) in *A. fumigatus* when compared to *S. cerevisiae* (13%) (Fig. 2). However, the higher proportions of the classes of protein described here may simply reflect the fact that this contig

Table 1
Summary of genetic features and annotation of the 922 kb chromosomal region containing the nitrate assimilation gene cluster (*niaD-niiA-crnA*) of *A. fumigatus (A.f.)* and comparison with the whole nuclear genomes of budding and fission yeasts [(*Saccharomyces cerevisiae* '*S.c.*' and *Schizosaccharomyces pombe* '*S.p.*', respectively)], a filamentous fungus (*Neurospora crassa*, '*N.c.*') and the social slime mould (*Dictyostelium discoideum*, '*D.d.*')

Feature	Values				
	A.f.	*N.c.*	*S.c.*	*S.p.*	*D.d.*
Size of the contig (bp)	921,536				
(G + C) content (%)	50.6	50.0	38.3	36.0	22.2
No. of predicted genes	341				
Mean gene length* (bp)	1461	1443	1424	1426	1626
Gene density†	2702	3626	2088	2528	2600
Percent coding	54	40.2	70.5	57.5	56.3
Percent coding (including introns)	59.8	46.5	ND	ND	ND
Genes					
With introns (%)	80	72	5	43	68
(G + C) content (%)	53.8	55.9	28.0	39.6	28.0
Introns					
No. per gene	3.0	2.6	ND	ND	2.3
(G + C) content (%)	46.5	ND	ND	ND	13
Mean length (bp)	78	134	82	81	177
Intergenic regions					
(G + C) content (%)	46.5	ND	ND	ND	14
Mean length (bp)	997	1953	515	952	786
No. tRNA genes	8				

Values for *D. discoideum* (Glockner et al., 2002) are for chromosome 2 and represent extrapolations to the entire genome. Sources of data for other organisms: *N. crassa* (Galagan et al., 2003), *S. cerevisiae* (Goffeau et al., 1996; Wood et al., 2001), *S. pombe* (Wood et al., 2002).

*Excluding introns.

†bp per gene. ND, not determined.

Table 2
Analysis of proteins predicted from the annotation of the *A. fumigatus* contig

Predicted protein features	
With putative N-terminal signal peptide	70 (20.5 %)
With 1 or more transmembrane domain(s)	73 (21.4 %)
Described previously in one or more *Aspergillus spp.*	54 (15.8 %)
Homologues identified in organisms other than Aspergilli	149 (43.7 %)
Hypothetical proteins which are conserved in 1 or more organisms	33 (9.7 %)
Hypothetical proteins (with no significant similarity to any protein)	105 (30.8 %)
Orthologous protein identified in *N. crassa*	238 (69.8 %)
Analysis of FASTA hits in the SWALL database:	
Top 5 FASTA matches contain proteins only from bacterial origin	19
Top 5 FASTA matches only from bacterial as well as plant origin	3
Top 5 FASTA matches only from fungal as well as bacterial origin	13
Top 5 FASTA matches contain proteins only from plant origin	8
Top 5 FASTA matches contain proteins from plants as well as fungi	6
Top 5 FASTA matches contain proteins only from fungi	31

Absolute numbers are given and percentage figures (%) are shown in brackets. The top 5 FASTA hits were analysed after running FASTA search against the SWALL database. SWALL is a comprehensive protein sequence database that combines the SWISS-PROT database and all protein coding sequences from the EMBL nucleotide sequence database.

houses more than one metabolic pathway gene cluster, which may therefore have influenced the percentages of proteins functionally annotated as 'enzymes' or 'transporters' and which are involved in secondary metabolism.

3.4. Comparison with the A. nidulans genetic linkage group VIII map and synteny in the qut/qa gene clusters

We have compared the synteny between the orthologous *A. fumigatus* genes identified in our sequenced contig and those mapped to linkage group VIII in the genetic map of *A. nidulans* (http://www.gla.ac.uk/Acad/IBLS/molgen/aspergillus/index.html) (Clutterbuck, 1997). Two regions (*facC* to *veA*; *trpC* to *hisC)* are inverted with respect to the order of the genes in *A. fumigatus* (Fig. 3A). When compared to the genetic map of *A. nidulans*, we could not detect the *galG* and *fwA* loci between the markers *facC* and the *qut* gene cluster for quinate utilisation. The loci *brlA*, *pyrD*, *aldA*, and *argC*, which map to linkage group VIII of the *A. nidulans*

Table 3
The top 9 Pfam domain hits (hits of 3 or more proteins to a Pfam domain) of the predicted protein set and a list of corresponding genes

Pfam entry	Number of hits	Interpro Accession No.	Gene identifyers
PF00172: fungal Zn(2)–Cys(6) binuclear cluster domain	5	IPR001138	AfA24A6.035c, AfA5A2.050c, AfA34E6.070c, AfA10A1.045, AfA10A1.100c
PF00096: Zinc finger, C2H2 type	3	IPR007087	AfA5c11.21c, AfA35G10.15, AfA33H4.005
PF01488: Shikimate/quinate 5-dehydrogenase	3	IPR002907	AfA5A2.035, AfA5A2.055, AfA35g10.13
PF00501: AMP-binding enzyme	3	IPR000873	AfA24A6.045, AfA28D10.115c, AfA6E3.015
PF00106: short chain dehydrogenase	3	IPR002198	AfA19D12.080, AfA10A1.040c, AfA10A1.050
PF00083: sugar (and other) transporter	3	IPR005828	AfA5A2.030c, AfA28D10.090c, AfA35g10.20c
PF00069: protein kinase domain	3	IPR000719	AfA28D1.020, AfA5C11.10 9, AfA35g10.01
PF00023: ankyrin repeat	3	IPR002110	AfA14E5.12c, AfA5C5.015c, AfA10A1.125
PF00005: ABC transporter	3	IPR003439	AfA5C5.85c, AfA5C11.08, AfA10A1.030

genetic map, would be expected to be present in the *A. fumigatus* genome and so it is reasonable to argue that this region is located elsewhere in the genome.

Clustering of genes involved in the same metabolic pathway has been observed in many organisms including filamentous fungi (Keller and Hohn, 1997). We analysed the synteny in the quinate utilisation gene cluster between *A. fumigatus* and other filamentous ascomycetous fungi, such as *A. nidulans*, *N. crassa*, and *P. anserina.* Each *qut* (or *qa*) gene cluster (consisting of genes commonly known as '*qut*' genes in *Aspergillus* spp. and in *P. anserina* and '*qa*' genes in *N. crassa*) consists of an array of at least five structural genes and two regulatory genes that are involved in the quinate utilisation pathway (Giles et al., 1985; Lamb et al., 1990; Silar et al., 2003). We used the Artemis Comparison Tool (ACT) to analyse the synteny (Fig. 3B). Although there is complete conservation in gene order and orientation between *A. fumigatus* and *A. nidulans qut* genes, there was little conservation in the order and orientation of the genes when compared to the clusters in *N. crassa* and *P. anserina*. Synteny is, however, observed between the *N. crassa* and *P. anserina* clusters with complete conservation in the order and orientation of the *qutG*, *E*, *C*, and *B* genes, whereas the *qutA*, *R*, and *D* genes are inverted with respect to each other in the two species. The *qutH* gene of *A. fumigatus* and *A. nidulans* (GenBank Accession Nos. X14603 and AL627362, respectively), which codes for a protein with putative oxidoreductase activity but whose function in the quinate utilisation pathway has not been defined, is absent in the gene clusters in *N. crassa* and *P. anserina* (Giles et al., 1985; Lamb et al., 1992).

Syntenic regions can help delineate evolutionary conserved genes and potential regulatory elements, and also serve as a valuable tool for gene identification and annotation. In recent years, there has been much interest in the identification and analysis of syntenic regions between filamentous fungi as more genome sequences become available (Hamer et al., 2001; Pedersen et al., 2002; Seoighe et al., 2000). Comparison of the contig sequence to the genetic map of *A. nidulans* suggests that synteny between the two genomes can extend over very large sequence blocks.

3.5. Proteins with a possible role in secondary metabolism

Many fungi produce a vast array of small, bioactive secondary metabolites that are best known for their roles as pigments, mycotoxins and antibiotics. We have identified a few genes in the contig that are predicted to play a role in secondary metabolism. One such gene has been annotated as a 'vesicolorin B synthase-like' (AfA24A6.075) with significant similarity to *vbS* gene of *A. parasiticus* (41% identity over a 560 amino acid overlap; $E = 1.2e-54$). The *vbS* gene has been shown to be part of the aflatoxin B1 gene cluster (Silva et al., 1996; Yu et al., 2000) and is involved in synthesis of vesicolorin B from versiconal hemiacetal acetate (Woloshuk and Prieto, 1998). Preliminary analysis of the *A. fumigatus* whole genome shotgun sequence data suggests that a cluster exists elsewhere in the genome that contains many of the genes involved in aflatoxin biosynthesis and whose products might therefore synthesise an intermediate, such as sterigmatocystin (data not shown). Genes required for the synthesis of aflatoxin and sterigmatocystin are well conserved between the Aspergilli and are located in large gene clusters (Brown et al., 1996; Yu et al., 2000), although the gene order and orientation of some of the orthologues are not conserved (Keller and Hohn, 1997). The specific role of this

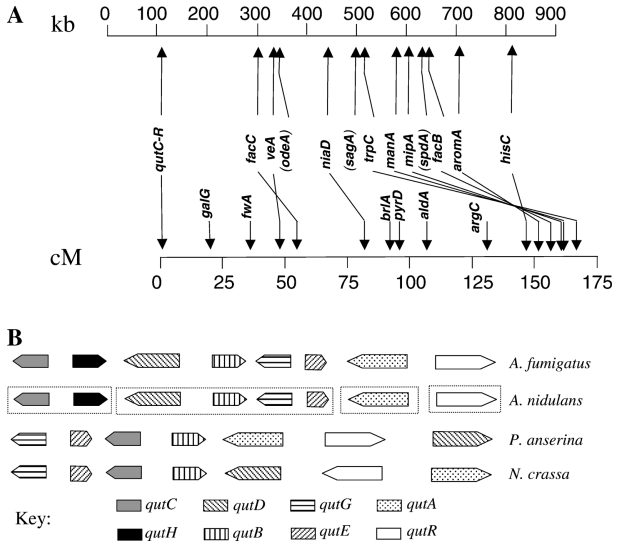

Fig. 3. (A) Synteny between linkage group VIII of the *Aspergillus nidulans* genetic map and the *A. fumigatus* contig described in this report. There is synteny between the orthologous *A. fumigatus* genes sequenced and those mapped to linkage group VIII in *A. nidulans*. Two regions (*facC* and *veA*; *trpC* to *hisC*) are inverted with respect to each other. The *hisC* locus has not been cloned in *A. nidulans* but the enzymatic activity coded by this gene has been defined. The *A. fumigatus* protein was identified on the basis of significant FASTA hit to *S. cerevisiae* and *S. pombe* proteins. Loci in brackets have been linked to linkage group VIII, but have not been mapped precisely. The genetic map was constructed using data from http://www.gla.ac.uk/Acad/IBLS/molgen/aspergillus/index.html (Clutterbuck, 1997). (B) Schematic diagram showing the degree of synteny in the quinate utilisation (*qut*/*qa*) gene clusters in four ascomycetous fungi: *A. fumigatus*, *A. nidulans*, *N. crassa*, and *P. anserina*. The orthologous genes have been shown with the same patterned boxes. The distances between the genes are not to scale. The arrowheads indicate the direction of transcription. In *A. nidulans*, the genes within the boxes represent individual GenBank entries and the arrangement of the genes has been determined using genetic and molecular data (Grant et al., 1988; Lamb et al., 1990). The Artemis Comparison Tool (ACT) (K. Rutherford, unpublished) was used to study the order and orientation of the genes. The DNA sequence around the *qut* gene clusters was generated using the following sequence files from GenBank: *N. crassa* (X14603), *P. anserina* (AL627362) and for *A. nidulans* a pseudocontig was generated using the following submitted sequences in GenBank: M77665, X13525, M58289, and M77664.

'vesicolorin B synthase-like' protein in secondary metabolism is unclear and it may not have anything to do with mycotoxin metabolism.

We have identified an aldo–keto reductase gene (AfA6E3.190c) predicted to code for a protein very similar to human aflatoxin B1-aldehyde reductase (AFAR) (43% identity over a 333 amino acid overlap; $E = 1.6e-47$) and rat aflatoxin B1-aldehyde reductase. There is also an orthologue to AfA6E3.190c in *N. crassa* (62% identity over the entire length of the protein (350 amino acids)). The mammalian AFAR proteins are members of the aldo–keto reductase 7 family (AKR7), part of the AKR superfamily (Jez et al., 1997a,b). Key amino acid residues (that are highly conserved in mammalian aflatoxin B1-aldehyde reductases and aldose reductases) are also conserved in the *A. fumigatus* protein (AfA6E3.190c) (data not shown). These residues comprise the catalytic triad, as well as those that bind the co-factor NADPH (Ireland et al., 1998; Wilson et al., 1993). The significant similarity of the putative aldo–keto reductase (AfA6E3.190c) to mammalian aflatoxin B1-aldehyde reductases is an interesting finding as this protein may play a role in metabolising specific charged aliphatic and aromatic aldehydes, which are toxic to cells. However, there is no experimental evidence to support this hypothesis.

3.6. Possible drug targets

We have identified the *aroM* gene (AfA35G10.13), coding for a 171 kDa pentafunctional polypeptide. Like other AroM orthologues, this protein has five functional domains, which catalyse consecutive steps in the shikimate biosynthetic pathway. Shikimate is an important intermediate required for the synthesis of aromatic compounds in bacteria, microbial eukaryotes and plants (Bentley, 1990). This protein is a potential target for antifungal and antibacterial drugs (Kishore and Shah, 1988) and it is required for virulence in the Salmonellae (Gunel-Ozcan et al., 1997). The crystal structure of *A. nidulans* dehydroquinate synthase (one of the five domains) has been determined (Carpenter et al., 1998) and, as the *A. fumigatus* AroM sequence is 80% identical to its orthologue in *A. nidulans*, modelling of the *A. fumigatus* orthologue could help to identify candidate inhibitors, in the search for new antifungal compounds.

In addition, we have identified several other enzymes that show similarity to proteins in plant or bacterial species. Such enzymes include a short-chain dehydrogenase/reductase (AfA19D12.080), a dioxygenase (AfA33H4.170c, predicted to be involved in the cleavage of aromatic rings), a putative 3-oxoacyl-[acyl carrier protein] reductase (AfA10A1.050, which is also a member of the short-chain dehydrogenase/reductase family) and a putative 4-coumerate-CoA ligase (AfA6E3.015). Some of these bacterial-like or plant-like enzymes may be components of biochemical pathways that are absent in the human genome and which are essential for the growth of *A. fumigatus* in animal tissue. They therefore, may offer potential new drug targets for evaluation.

3.7. Transporters and transcription factors

Aspergillus fumigatus produces an array of secondary metabolites (Fog Nielsen, 2003). It is expected to be armed with a wide repertoire of transcriptional control points and to have various export mechanisms in order to prevent accumulation of potentially toxic metabolites inside the cell. We have identified several putative transcriptional control proteins and transporters in the

contig, representing about one-fifth of the functionally annotated proteins. Efflux is probably the major mechanism of azole drug resistance in *Aspergillus* species (Moore et al., 2000). The genes, AfA35g10.20c, AfA5C5.050, and AfA24A6.065 code for proteins that are related to a group in the major facilitator superfamily, which is involved in the efflux of drugs. In addition, AfA5C5.085c is a previously characterised gene, *mdr4*, which is over-expressed in some *A. fumigatus* mutants with high-level resistance to itraconazole and whose product is a member of the ABC transporter superfamily (Nascimento et al., 2003). The gene, AfA10A1.030, has also been previously characterised: its product is likewise a member of the ABC transporter superfamily and it has been shown to play a role in itraconazole susceptibility (Mosquera et al., 2002). Its *A. nidulans* orthologue, *atrG*, has also been functionally characterised and has been shown to have a similar role in azole resistance (Andrade et al., 2002).

The most common Pfam domain we identified was the fungal-specific Zn(2)–Cys(6) binuclear cluster domain, which is an N-terminal region DNA-binding domain found in transcriptional regulators. In total, 17 proteins have been predicted to function as transcription factors, which constitute 8% of the functionally annotated proteins. A few interesting examples of such transcription factors include the activator, QutA (AfA5A2.050c), and repressor, QutR (AfA5A2.055), of the quinate utilisation gene cluster, a Jumonji family transcription factor (AfA14E5.23c) (Balciunas and Ronne, 2000), and the acetate regulatory FacB orthologue (AfA34E6.070c) (Todd et al., 1997).

3.8. Possible host-interacting molecules

One of the key questions related to our understanding of host–fungus interaction during the development of invasive aspergillosis is to identify fungal encoded molecules that may influence the outcome of infection (Latge and Calderone, 2002; Muhlschlegal et al., 1998). Specific interactions between conidia of *A. fumigatus* and host molecules have previously been documented (Allen et al., 2001; Anderson et al., 2003; Clemons et al., 2000; Latge and Calderone, 2002; Mendes-Giannini et al., 2000; Tronchin et al., 2002). We have identified two candidate genes whose products may be involved in interactions with the host. The gene AfA10A1.015 encodes a 580 amino-acid-long protein that is predicted by Pfam to contain two fascilin domains. These extracellular domains have been suggested to represent an ancient cell adhesion domain and proteins with these domains are attached to the membrane with a GPI-anchor. Although the *A. fumigatus* protein has a predicted signal peptide, it is not however predicted to contain a GPI-anchor. The gene AfA33H4.165 encodes a large (1236 amino acid) protein, which has a clear *N. crassa* orthologue. The C-terminal half of this protein is similar to TGF-β receptor associated protein 1 (26% identity over 644 amino acids). However, this protein is not predicted to have a signal peptide and so might not be expressed on the cell surface.

3.9. Conclusion

This study represents the first large-scale analysis of a region of the *A. fumigatus* genome. Several genes have been identified that will extend our search for possible drug targets and pathogenicity factors. This study will hopefully act as a 'jump start' for research on *A. fumigatus* biology and aspergillosis, ahead of the completion of the whole genome sequencing project (Breitenbach et al., 2002; Denning et al., 2002). Once the whole genome sequences of *A. fumigatus* and *A. nidulans* are available, comparative genome analyses of the Aspergilli such as between invasive *A. fumigatus* isolates and non-pathogenic *A. nidulans* isolates will help to determine why *A. fumigatus* is capable of causing such extremes of infection, ranging from allergy to invasive disease.

Acknowledgments

We are indebted to Rod Warren for having the forethought to save the clinical isolate that was selected for sequencing and for providing extensive clinical details on this strain. Also thanks to Caroline Moore for susceptibility testing, to Stephan Bretagne for microsattelite typing, to David Geiser for his phylogenetic analysis, to Mike Bromley for providing the *niaD* gene sequence and to Jonathon Ross for performing the orthology analysis between *A. fumigatus* and *A. nidulans*. We are also indebted to the Wellcome Trust and the Fungal Research Trust for funding the work. We are grateful to Hubert Renauld for advice and corrections to the manuscript.

References

Allen, M.J., Voelker, D.R., Mason, R.J., 2001. Interactions of surfactant proteins A and D with *Saccharomyces cerevisiae* and *Aspergillus fumigatus*. Infect. Immun. 9, 7–2044.

Anaissie, E.J., Stratton, S.L., Dignani, M.C., Summerbell, R.C., Rex, J.H., Monson, T.P., Spencer, T., Kasai, M., Francesconi, A., et al., 2002. Pathogenic *Aspergillus* species recovered from a hospital water system: a 3-year prospective study. Clin. Infect. Dis. 34, 780–789.

Anderson, M.J., Brookman, J.L., Denning, D.W., 2003. *Aspergillus*. In: Prade, R.A., Bohnert, H.J. (Eds.), Genomics of Plants and Fungi. Marcel Dekker, New York, pp. 1–39.

Andrade, A., Braam, C., Haas, H., de Waard, M., 2002. ABC transporters and resistance to azole antimycotics in the ima mutants of *Aspergillus nidulans*—a regulatory role for imaB. In: 6th European Conference on Fungal Genetics, p. 291.

Apweiler, R., Attwood, T.K., Bairoch, A., Bateman, A., Birney, E., Biswas, M., Bucher, P., Cerutti, L., Corpet, F., Croning, M.D., Durbin, R., Falquet, L., Fleischmann, W., Gouzy, J., Hermjakob,

H., Hulo, N., Jonassen, I., Kahn, D., Kanapin, A., Karavidopoulou, Y., Lopez, R., Marx, B., Mulder, N.J., Oinn, T.M., Pagni, M., Servant, F., Sigrist, C.J., Zdobnov, E.M., 2000. InterPro—an integrated documentation resource for protein families, domains and functional sites. Bioinformatics 16, 1145–1150.

Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S., et al., 2000. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium Nat. Genet. 25, 25–29.

Balciunas, D., Ronne, H., 2000. Evidence of domain swapping within the jumonji family of transcription factors. Trends Biochem. Sci. 25, 274–276.

Bart-Delabesse, E., Humbert, J.F., Delabesse, E., Bretagne, S., 1998. Microsatellite markers for typing *Aspergillus fumigatus* isolates. J. Clin. Microbiol. 36, 2413–2418.

Bateman, A., Birney, E., Cerruti, L., Durbin, R., Etwiller, L., Eddy, S.R., Griffiths-Jones, S., Howe, K.L., Marshall, M., Sonnhammer, E.L., 2002. The Pfam protein families database. Nucleic Acids Res. 30, 276–280.

Bentley, R., 1990. The shikimate pathway—a metabolic tree with many branches. Crit. Rev. Biochem. Mol. Biol. 25, 307–384.

Breitenbach, M., Crameri, R., Lehrer, S.B., 2002. Impact of current genome projects on the study of pathogenic and allergenic fungi. Chem. Immunol. 81, 5–9.

Brown, D.W., Yu, J.H., Kelkar, H.S., Fernandes, M., Nesbitt, T.C., Keller, N.P., Adams, T.H., Leonard, T.J., 1996. Twenty-five coregulated transcripts define a sterigmatocystin gene cluster in *Aspergillus nidulans*. Proc. Natl. Acad. Sci. USA 93, 1418–1422.

Carpenter, E.P., Hawkins, A.R., Frost, J.W., Brown, K.A., 1998. Structure of dehydroquinate synthase reveals an active site capable of multistep catalysis. Nature 394, 299–302.

Clemons, K.V., Calich, V.L., Burger, E., Filler, S.G., Grazziutti, M., Murphy, J., Roilides, E., Campa, A., Dias, M.R., Edwards Jr., J.E., Fu, Y., Fernandes-Bordignon, G., Ibrahim, A., Katsifa, H., Lamaignere, C.G., Meloni-Bruneri, L.H., Rex, J., Savary, C.A., Xidieh, C., 2000. Pathogenesis I: interactions of host cells and fungi. Med. Mycol., 99–111.

Clutterbuck, A.J., 1997. The validity of the *Aspergillus nidulans* linkage map. Fungal Genet. Biol. 21, 267–277.

Denning, D.W., Anderson, M.J., Turner, G., Latge, J.P., Bennett, J.W., 2002. Sequencing the *Aspergillus fumigatus* genome. Lancet Infect. Dis. 2, 251–253.

Fog Nielsen, K., 2003. Mycotoxin production by indoor molds. Fungal Genet. Biol. 39, 103–117.

Frengen, E., Weichenhan, D., Zhao, B., Osoegawa, K., van Geel, M., de Jong, P.J., 1999. A modular, positive selection bacterial artificial chromosome vector with multiple cloning sites. Genomics 58, 250–253.

Galagan, J.E., Calvo, S.E., Borkovich, K.A., Selker, E.U., Read, N.D., Jaffe, D., FitzHugh, W., Ma, L.J., Smirnov, S., Purcell, S., Rehman, B., Elkins, T., Engels, R., Wang, S., Nielsen, C.B., Butler, J., Endrizzi, M., Qui, D., Ianakiev, P., Bell-Pedersen, D., Nelson, M.A., Werner-Washburne, M., Selitrennikoff, C.P., Kinsey, J.A., Braun, E.L., Zelter, A., Schulte, U., Kothe, G.O., Jedd, G., Mewes, W., Staben, C., Marcotte, E., Greenberg, D., Roy, A., Foley, K., Naylor, J., Stange-Thomann, N., Barrett, R., Gnerre, S., Kamal, M., Kamvysselis, M., Mauceli, E., Bielke, C., Rudd, S., Frishman, D., Krystofova, S., Rasmussen, C., Metzenberg, R.L., Perkins, D.D., Kroken, S., Cogoni, C., Macino, G., Catcheside, D., Li, W., Pratt, R.J., Osmani, S.A., DeSouza, C.P., Glass, L., Orbach, M.J., Berglund, J.A., Voelker, R., Yarden, O., Plamann, M., Seiler, S., Dunlap, J., Radford, A., Aramayo, R., Natvig, D.O., Alex, L.A., Mannhaupt, G., Ebbole, D.J., Freitag, M., Paulsen, I., Sachs, M.S., Lander, E.S., Nusbaum, C., Birren, B., 2003. The genome sequence of the filamentous fungus *Neurospora crassa*. Nature 422, 859–868.

Geiser, D.M., Frisvad, J.C., Taylor, J.W., 1998. Evolutionary relationships in *Aspergillus* section Fumigata inferred from partial beta-tubulin and hydrophobin DNA sequences. Mycologia 90, 831–845.

Giles, N.H., Case, M.E., Baum, J., Geever, R., Huiet, L., Patel, V., Tyler, B., 1985. Gene organization and regulation in the qa (quinic acid) gene cluster of *Neurospora crassa*. Microbiol. Rev. 49, 338–358.

Glockner, G., Eichinger, L., Szafranski, K., Pachebat, J.A., Bankier, A.T., Dear, P.H., Lehmann, R., Baumgart, C., Parra, G., Abril, J.F., Guigo, R., Kumpf, K., Tunggal, B., Cox, E., Quail, M.A., Platzer, M., Rosenthal, A., Noegel, A.A., 2002. Sequence and analysis of chromosome 2 of *Dictyostelium discoideum*. Nature 418, 79–85.

Goffeau, A., Barrell, B.G., Bussey, H., Davis, R.W., Dujon, B., Feldmann, H., Galibert, F., Hoheisel, J.D., Jacq, C., Johnston, M., Louis, E.J., Mewes, H.W., Murakami, Y., Philippsen, P., Tettelin, H., Oliver, S.G., 1996. Life with 6000 genes. Science 274, 546, 547–563.

Grant, S., Roberts, C.F., Lamb, H., Stout, M., Hawkins, A.R., 1988. Genetic regulation of the quinic acid utilization (QUT) gene cluster in *Aspergillus nidulans*. J. Gen. Microbiol. 134 (Part 2), 347–358.

Groll, A.H., Shah, P.M., Mentzel, C., Schneider, M., Just-Nuebling, G., Huebner, K., 1996. Trends in the postmortem epidemiology of invasive fungal infections at a university hospital. J. Infect. 33, 23–32.

Gunel-Ozcan, A., Brown, K.A., Allen, A.G., Maskell, D.J., 1997. *Salmonella typhimurium aroB* mutants are attentuated in BALB/c mice. Microb. Pathog. 23, 311–316.

Hamer, L., Pan, H., Adachi, K., Orbach, M.J., Page, A., Ramamurthy, L., Woessner, J.P., 2001. Regions of microsynteny in *Magnaporthe grisea* and *Neurospora crassa*. Fungal Genet. Biol. 33, 137–143.

Harris, D.E., Murphy, L., 2001. Sequencing bacterial artificial chromosomes. In: Starkey, M.P., Elaswarapu, R. (Eds.), Methods In Molecular Biology. Humana Press, Totowa, NJ, USA, pp. 217–234.

Ireland, L.S., Harrison, D.J., Neal, G.E., Hayes, J.D., 1998. Molecular cloning, expression and catalytic activity of a human AKR7 member of the aldo–keto reductase superfamily: evidence that the major 2-carboxybenzaldehyde reductase from human liver is a homologue of rat aflatoxin B1-aldehyde reductase. Biochem. J. 332, 21–34.

Jez, J.M., Bennett, M.J., Schlegel, B.P., Lewis, M., Penning, T.M., 1997a. Comparative anatomy of the aldo–keto reductase superfamily. Biochem. J. 326 (Pt 3), 625–636.

Jez, J.M., Flynn, T.G., Penning, T.M., 1997b. A new nomenclature for the aldo–keto reductase superfamily. Biochem. Pharmacol. 54, 639–647.

Johnson, L.B., Kauffman, C.A., 2003. Voriconazole: a new triazole antifungal agent. Clin. Infect. Dis. 36, 630–637.

Johnson, M.D., Perfect, J.R., 2003. Caspofungin: first approved agent in a new class of antifungals. Exp. Opin. Pharmacother. 4, 807–823.

Keller, N.P., Hohn, T.M., 1997. Metabolic pathway gene clusters in filamentous fungi. Fungal Genet. Biol. 21, 17–29.

Kishore, G.M., Shah, D.M., 1988. Amino acid biosynthesis inhibitors as herbicides. Annu. Rev. Biochem. 57, 627–663.

Krogh, A., Larsson, B., von Heijne, G., Sonnhammer, E.L., 2001. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J. Mol. Biol. 305, 567–580.

Lamb, H.K., Hawkins, A.R., Smith, M., Harvey, I.J., Brown, J., Turner, G., Roberts, C.F., 1990. Spatial and biological characterisation of the complete quinic acid utilisation gene cluster in *Aspergillus nidulans*. Mol. Gen. Genet. 223, 17–23.

Lamb, H.K., Roberts, C.F., Hawkins, A.R., 1992. A second gene (*qutH*) within the *Aspergillus nidulans*-quinic-acid utilisation gene cluster encodes a protein with a putative zinc-cluster motif. Gene 112, 219–224.

Latge, J.P., 1999. *Aspergillus fumigatus* and aspergillosis. Clin. Microbiol. Rev. 12, 310–350.

Latge, J.P., 2001. The pathobiology of *Aspergillus fumigatus*. Trends Microbiol. 9, 382–389.

Latge, J.P., Calderone, R., 2002. Host–microbe interactions: fungi invasive human fungal opportunistic infections. Curr. Opin. Microbiol. 5, 355–358.

Lowe, T.M., Eddy, S.R., 1997. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res. 25, 955–964.

Majoros, W.H., Pertea, M., Antonescu, C., Salzberg, S.L., 2003. GlimmerM, Exonomy and Unveil: three ab initio eukaryotic genefinders. Nucleic Acids Res. 31, 3601–3604.

Marr, K.A., Patterson, T., Denning, D., 2002. Aspergillosis, Pathogenesis, clinical manifestations, and therapy. Infect. Dis. Clin. North Am. 16, 875–894, vi..

Mendes-Giannini, M.J., Taylor, M.L., Bouchara, J.B., Burger, E., Calich, V.L., Escalante, E.D., Hanna, S.A., Lenzi, H.L., Machado, M.P., Miyaji, M., Monteiro Da Silva, J.L., Mota, E.M., Restrepo, A., Restrepo, S., Tronchin, G., Vincenzi, L.R., Xidieh, C.F., Zenteno, E., 2000. Pathogenesis II: fungal responses to host responses: interaction of host cells with fungi. Med. Mycol., 113–123.

Moore, C.B., Sayers, N., Mosquera, J., Slaven, J., Denning, D.W., 2000. Antifungal drug resistance in *Aspergillus*. J. Infect. 41, 203–220.

Mosquera, J., Sharp, A., Moore, C.B., Warn, P.A., Denning, D.W., 2002. In vitro interaction of terbinafine with itraconazole, fluconazole, amphotericin B and 5-flucytosine against *Aspergillus* spp.. J. Antimicrob. Chemother. 50, 189–194.

Muhlschlegal, F., Fonzi, W., Hoyer, L., Payne, T., Poulet, F.M., Clevenger, J., Latge, J.P., Calera, J., Beauvais, A., Paris, S., Monod, M., Sturtevant, J., Ghannoum, M., Nozawa, Y., Calderone, R., 1998. Molecular mechanisms of virulence in fungus–host interactions for *Aspergillus fumigatus* and *Candida albicans*. Med. Mycol. 36, 238–248.

Nascimento, A.M., Goldman, G.H., Park, S., Marras, S.A., Delmas, G., Oza, U., Lolans, K., Dudley, M.N., Mann, P.A., Perlin, D.S., 2003. Multiple resistance mechanisms among *Aspergillus fumigatus* mutants with high-level resistance to itraconazole. Antimicrob. Agents Chemother. 47, 1719–1726.

Nielsen, H., Brunak, S., von Heijne, G., 1999. Machine learning approaches for the prediction of signal peptides and other protein sorting signals. Protein Eng. 12, 3–9.

Osoegawa, K., Woon, P.Y., Zhao, B., Frengen, E., Tateno, M., Catanese, J.J., de Jong, P.J., 1998. An improved approach for construction of bacterial artificial chromosome libraries. Genomics 52, 1–8.

Pedersen, C., Wu, B., Giese, H., 2002. A *Blumeria graminis* f.sp. *hordei* BAC library—contig building and microsynteny studies. Curr. Genet. 42, 103–113.

Pitt, J.I., 1994. The current role of *Aspergillus* and *Penicillium* in human and animal health. J. Med. Vet. Mycol. 32, 17–32.

Pitt, J.I., Samson, R.A., Frisvad, J.C., 2000. List of Accepted Species and their Synonyms in the Family Trichocomaceae. Harwood Publishers, Amsterdam.

Rutherford, K., Parkhill, J., Crook, J., Horsnell, T., Rice, P., Rajandream, M.A., Barrell, B., 2000. Artemis: sequence visualization and annotation. Bioinformatics 16, 944–945.

Salzberg, S.L., Pertea, M., Delcher, A.L., Gardner, M.J., Tettelin, H., 1999. Interpolated Markov models for eukaryotic gene finding. Genomics 59, 24–31.

Sambrook, J., Fritsch, F.E., Maniatis, T., 1989. Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.

Seoighe, C., Federspiel, N., Jones, T., Hansen, N., Bivolarovic, V., Surzycki, R., Tamse, R., Komp, C., Huizar, L., Davis, R.W., Scherer, S., Tait, E., Shaw, D.J., Harris, D., Murphy, L., Oliver, K., Taylor, K., Rajandream, M.A., Barrell, B.G., Wolfe, K.H., 2000. Prevalence of small inversions in yeast gene order evolution. Proc. Natl. Acad. Sci. USA 97, 14433–14437.

Silar, P., Barreau, C., Debuchy, R., Kicka, S., Turcq, B., Sainsard-Chanet, A., Sellem, C.H., Billault, A., Cattolico, L., Duprat, S., Weissenbach, J., 2003. Characterization of the genomic organization of the region bordering the centromere of chromosome V of *Podospora anserina* by direct sequencing. Fungal Genet. Biol. 39, 250–263.

Silva, J.C., Minto, R.E., Barry 3rd, C.E., Holland, K.A., Townsend, C.A., 1996. Isolation and characterization of the versicolorin B synthase gene from *Aspergillus parasiticus*. Expansion of the aflatoxin b1 biosynthetic gene cluster. J. Biol. Chem. 271, 13600–13608.

Tobin, M.B., Peery, R.B., Skatrud, P.L., 1997. An electrophoretic molecular karyotype of a clinical isolate of *Aspergillus fumigatus* and localization of the MDR-like genes AfuMDR1 and AfuMDR2. Diagn. Microbiol. Infect. Dis. 29, 67–71.

Todd, R.B., Murphy, R.L., Martin, H.M., Sharp, J.A., Davis, M.A., Katz, M.E., Hynes, M.J., 1997. The acetate regulatory gene facB of *Aspergillus nidulans* encodes a Zn(II)2Cys6 transcriptional activator. Mol. Gen. Genet. 254, 495–504.

Tronchin, G., Esnault, K., Sanchez, M., Larcher, G., Marot-Leblond, A., Bouchara, J.P., 2002. Purification and partial characterization of a 32-kiloDalton sialic acid-specific lectin from *Aspergillus fumigatus*. Infect. Immun. 70, 6891–6895.

Vogeser, M., Haas, A., Aust, D., Ruckdeschel, R., 1997. Postmortem analysis of invasive aspergillosis in a tertiary care hospital. Eur. J. Clin. Microbiol. Infect. Dis. 16, 1–6.

Wilson, D.K., Tarle, I., Petrash, J.M., Quiocho, F.A., 1993. Refined 1.8 Å structure of human aldose reductase complexed with the potent inhibitor zopolrestat. Proc. Natl. Acad. Sci. USA 90, 9847–9851.

Woloshuk, C.P., Prieto, R., 1998. Genetic organization and function of the aflatoxin B1 biosynthetic genes. FEMS Microbiol. Lett. 160, 169–176.

Wood, V., Gwilliam, R., Rajandream, M.A., Lyne, M., Lyne, R., Stewart, A., Sgouros, J., Peat, N., Hayles, J., Baker, S., Basham, D., Bowman, S., Brooks, K., Brown, D., Brown, S., Chillingworth, T., Churcher, C., Collins, M., Connor, R., Cronin, A., Davis, P., Feltwell, T., Fraser, A., Gentles, S., Goble, A., Hamlin, N., Harris, D., Hidalgo, J., Hodgson, G., Holroyd, S., Hornsby, T., Howarth, S., Huckle, E.J., Hunt, S., Jagels, K., James, K., Jones, L., Jones, M., Leather, S., McDonald, S., McLean, J., Mooney, P., Moule, S., Mungall, K., Murphy, L., Niblett, D., Odell, C., Oliver, K., O'Neil, S., Pearson, D., Quail, M.A., Rabbinowitsch, E., Rutherford, K., Rutter, S., Saunders, D., Seeger, K., Sharp, S., Skelton, J., Simmonds, M., Squares, R., Squares, S., Stevens, K., Taylor, K., Taylor, R.G., Tivey, A., Walsh, S., Warren, T., Whitehead, S., Woodward, J., Volckaert, G., Aert, R., Robben, J., Grymonprez, B., Weltjens, I., Vanstreels, E., Rieger, M., Schafer, M., Muller-Auer, S., Gabel, C., Fuchs, M., Dusterhoft, A., Fritzc, C., Holzer, E., Moestl, D., Hilbert, H., Borzym, K., Langer, I., Beck, A., Lehrach, H., Reinhardt, R., Pohl, T.M., Eger, P., Zimmermann, W., Wedler, H., Wambutt, R., Purnelle, B., Goffeau, A., Cadieu, E., Dreano, S., Gloux, S., Lelaure, V., Mottier, S., Galibert, F., Aves, S.J., Xiang, Z., Hunt, C., Moore, K., Hurst, S.M., Lucas, M., Rochet, M., Gaillardin, C., Tallada, V.A., Garzon, A., Thode, G., Daga, R.R., Cruzado, L., Jimenez, J., Sanchez, M., del Rey, F., Benito, J., Dominguez, A., Revuelta, J.L., Moreno, S., Armstrong, J., Forsburg, S.L., Cerutti, L., Lowe, T., McCombie, W.R., Paulsen, I., Potashkin, J., Shpakovski, G.V., Ussery, D., Barrell, B.G., Nurse, P., Cerrutti, L., 2002. The genome sequence of *Schizosaccharomyces pombe*. Nature 415, 871–880.

Wood, V., Rutherford, K., Ivens, A., Rajandream, M.-A., Barrell, B., 2001. A Re-annotation of the *Saccharomyces* Genome. Comp. Funct. Genomics 2, 143–154.

Yu, J., Chang, P.K., Bhatnagar, D., Cleveland, T.E., 2000. Genes encoding cytochrome P450 and monooxygenase enzymes define one end of the aflatoxin pathway gene cluster in *Aspergillus parasiticus*. Appl. Microbiol. Biotechnol. 53, 583–590.