White Paper Application

Project Title:

White Paper Submission Date (MM/DD/YY):

Investigator Contact:

Name	David W. Denning
Position	Professor of Medicine and Medical Mycology
Institution	University Hospital of South Manchester
Address	Southmoor Road
	Manchester M23 9LT UK
Telephone	44 161 291 5811/5818
Fax	44 161 291 5806
E-Mail	david.denning@manchester.ac.uk

All white papers will be evaluated based on the following sections.

1. Executive Summary *(Please limit to 500 words.)*

Provide an executive summary of the proposal.

Aspergillus fumigatus is the predominant cause of both allergic and invasive aspergillosis possessing particular metabolic capabilities and genetic determinants that differentiate it from most other fungal species. *A. fumigatus* is also a primary pathogen of the sinuses, lungs, damaged skin, and subcutaneous tissues, and disseminates to other organs including the brain. The key drugs for treating aspergillosis are azoles, especially voriconazole, but unfortunately resistance to these drugs is emerging.

Genomic resources for studying the biology and pathology of *A. fumigatus* are limited in two critical ways. The first limitation is that we have available the genome sequences of only two strains, Af293 and A1163. Microsatellite typing and similar gene content document that they are closely related, greatly limiting our view of the genetic/genomic variation within the species. To address this limitation, we are proposing to sequence and annotate two additional strains carefully selected to provide a much broader representation of the species variation. The use of new sequencing technologies will likely allow the sequencing of centromeric and other regions that were left unsequenced in Af293 and A1163 due to the limitations of *Escherichia coli* clone libraries used in sequencing these strains.

The low quality of the genome annotation of *A. fumigatus* is the second major limitation to studies on *A. fumigatus*. While the annotation of *A. fumigatus* is probably superior to that of any sequenced filamentous fungus, state-of-the-art genome annotation software alone lacks the power to produce high quality annotation. Numerous gene models in the *A. fumigatus* sequenced genomes have undetected errors in translation start and stop sites, and splice boundaries. Identification of 5' and 3' untranslated regions of protein-coding genes and alternative splice sites was not even addressed in the process of the annotation of these genomes. Neither was the identification and genomic locations of non-coding genes including small RNAs (sRNAs) that may regulate *A. fumigatus* biology. This circumstance forces investigators attempting to delete or modify an *A. fumigatus* gene to experimentally determine/verify the gene structure of a gene of interest in an inefficient one-gene-at-a-time manner. To address this limitation, the RNA-Seq and sRNA-Seq technologies will be employed to experimentally verify the gene models for nearly all of

the protein-coding genes and non-coding sRNAs in the genome . *A. fumigatus* investigators from 7 research groups have agreed to provide RNAs from a range of *in vitro* and *in vivo* cultivation conditions to allow the experimental determination of the gene structures of what we expect will be essentially the entire transcriptome of the species.

In the clinical treatment of invasive aspergillosis, drug resistance in *A. fumigatus* has emerged as a major challenge in obtaining good clinical outcomes. Numerous mutations resulting in resistance have been described, but its molecular and evolutionary mechanisms remain to be characterized. We will sequence 42 resistant strains. Six matched susceptible strains from individual patients that subsequently evolved to resistance will also be sequenced to identify additional mechanisms of resistance as this resistance evolves in patients. To achieve a high resolution view of the genome sequence alterations in these strains we will use high sequence coverage of these genomes from a short read length sequencing platform.

This white paper proposal has been developed in consultation with a large number of US and international *A. fumigatus* PIs including Michelle Momany, the current elected chair of the Aspergillus *Genomics Research Policy Committee (AGRPC).* This project thus represents the consensus opinion of the Aspergillus community for addressing the two limitations of *Aspergillus fumigatus* genomic resources cited above and to explore the mechanism of drug resistance development in invasive aspergillosis clinical practice.

2. Justification

Provide a succinct justification for the sequencing or genotyping study by describing the significance of the problem and providing other relevant background information.

Public health significance. Invasive Aspergillosis (IA) caused by *Aspergillus fumigatus* is growing in significance due to substantial increases in the number of allogenic and autologous stem cell transplants, as well as advances in medical technologies that are expanding the spectrum of patients susceptible to *A. fumigatus* infections. With the number of immunocompromised patients rising each year, the significance of invasive infections caused by *A. fumigatus* will continue to rise. Couple with this increasing at risk population is the high mortality rates of 90-100% in untreated IA patients and 50-90% in treated patients, with the development of resistance to antifungal agents contributing to this mortality. To combat the high morbidity and mortality rate for IA, new knowledge of the pathophysiology of this disease is needed. The investigators who would develop this new knowledge will be greatly aided by the resources to be developed by this proposed project.

The sequenced *A. fumigatus* strains do not reflect the genomic diversity of the species. Two strains of *A. fumigatus* have been sequenced and annotated and the genomic data is publically available. They are Af293 (Nierman et al. 2005, Nature 438:1151) and A1163 (Fedorova et al. 2008, PLoS Genetics 4:e1000046). As determined by microsatellite typing and comparative genomic analysis, these two strains are very similar. Af293 and A1163 harbor respectively 143 and 218 unique genes relative to each other comprising 1.2% and 2.3 % of their genomes. By microsatellite typing these strains have the same alleles at 2 of the 5 typed loci. Thus these strains do not provide an accurate view of the genomic diversity within the species necessitating the sequencing of additional strains of greater diversity.

The current annotation of *A. fumigatus* is inadequate to support functional analysis by community scientists. Inaccurate or missing gene models present an enormous problem for many areas of contemporary biology including first of all genetic analysis, expression profiling studies, and comparative genomics. Our recent comparative genomic analysis demonstrated that up to 50% of the *A. fumigatus* A1163 gene models may have missing first exons or inaccurate exon-intron boundaries (Fedorova et al. 2008, PLoS Genetics 4:e1000046). This means for example that investigators, attempting to delete or modify an *A. fumigatus* gene of interest, have to experimentally determine its structure in an inefficient one-gene-at-a-time manner. Many microarray-based expression profiling studies in *A. fumigatus* were profoundly affected by misannotation. They also suffer from the scarcity of the metabolic and transcriptional pathway annotation for protein coding genes as well as the lack of annotation for untranslated regions and non-coding genes that may regulate *A. fumigatus* biology.

Drug resistance in *A. fumigatus* is increasingly becoming an issue in treatment failures for Invasive Aspergillosis. Azoles are the mainstay of oral therapy for aspergillosis. Azole resistance in *Aspergillus* has been reported infrequently. The first resistant isolate was detected in 1999 in Manchester, UK. In a clinical collection of 519 *A. fumigatus* isolates, the frequency of itraconazole resistance was 5%, a significant increase since 2004 ($p<0.001$). Of the 34 itraconazole-resistant isolates studied, 65% (22) were cross-resistant to voriconazole and 74% (25) were cross-resistant to posaconazole. Thirteen of 14 evaluable patients in the study had prior azole exposure; 8 infections failed therapy (progressed), and 5 failed to improve (remained stable). Eighteen amino acid alterations were found in the target enzyme, Cyp51A, four of which were novel. A population genetic analysis of microsatellites showed the existence of resistant mutants that evolved from originally susceptible strains, different *cyp51A* mutations in the same strain, and microalterations in microsatellite repeat number. Azole resistance in *A. fumigatus* is an emerging problem that seems to develop during azole therapy (Howard et al. 2009, Emerging Infectious Diseases 15:1068). Documenting this development and characterizing the atypical resistance mechanisms will prove to be important in the future development of antifungal therapies of IA within this realm of emerging drug resistance.

3a. Rationale for Strain Selection

Reference strains to be sequenced. *A. fumigatus* strains to be sequenced and annotated to high quality. These will be selected to be as different as is feasible from the Af293 and A1163 sequenced strains. The microsatellite types of the already sequenced and proposed strains are shown here (Manchester data, Straf_3B failed).

Isolate	STRAf_3A	STRAf_3C	STRAf_4A	STRAf_4B	STRAf_4C
AF293	245	132	190	186	175
FGSC A1163	164	135	209	166	175
AF10	233	141	213	182	175
AF210	215	90	154	186	263

STRAf – Simple tandem repeat *A. fumigatus*

***A. fumigatus* strain AF10 (ATTC 90240)**
This strain was originally isolated by from a needle aspiration specimen of a nodular infiltrate in a corticosteroid-treated patient in Los Angles in 1986 and submitted to Dr David Stevens' laboratory for susceptibility testing. It is a fully susceptible strain that grows at 50ºC and has been used in numerous studies over the past 20 years. It is a good producer of extracellular elastase and extracellular phospholipases. It has been extensively in many experimental studies, including the demonstration that corticosteroids enhance growth rate in *A. fumigatus*.

***A. fumigatus* strain AF210 (NCPF 7101)**
This strain was isolated as a hospital acquired infection from an intensive care unit in Manchester, UK and has been fully reported as a case (Carlson et al, J Infect 1996;33:119). This and 5 other isolates, including an environmental strain from the same ICU, were molecularly typed using restriction endonuclease restriction, RAPD and M13 hybridization of HindIII digests and shown to be identical. Experimental infection caused by AF210 responds to both amphotericin B and itraconazole, and has been used to validate susceptibility testing of azoles. It is more amenable to hygromycin-based transformation than AF293.

Strains for drug resistance analysis. In the clinical treatment of invasive aspergillosis, drug resistance in *A. fumigatus* has emerged as a major challenge in obtaining good clinical outcomes. Numerous mutations resulting in resistance have been described, but its molecular and evolutionary mechanisms remain to be characterized. We will sequence 42 resistant strains. Six matched susceptible strains from individual patients that subsequently evolved to resistance will also be sequenced to identify additional mechanisms of resistance as this resistance evolves in patients. To achieve a high resolution view of the genome sequence ilterations in these strains we will use high sequence coverage of these genomes from a short read length sequencing platform.

RNA for RNA-seq and sRNA-seq strains and conditions. RNA is to be provided by 7 collaborating laboratories and will employ a range of *in vivo* and *in vitro* conditions. The goal of this large range of conditions is to ensure the expression of essentially all of the transcripts in the genome. These conditions will include in vitro rich and minimal solid and liquid media, multiple growth states, and multiple stresses. *In vivo* conditions will employ the Bignell mouse lung model of early infection (McDonagh et al. 2008, PLoS Pathogen 4:e1000154). Up to 56 RNA samples will be provided.

3c. Nature, Availability & Source of Reagents/Samples:

Genomic DNAs from reference strains for high quality sequencing.

A. fumigatus strain AF10 (ATCC 90240) will be obtained from ATCC. Genomic DNA will be prepared by JCVI.

A. fumigatus strain will be supplied by the NCPF in Bristol. Genomic DNA will be prepared by JCVI.

***A. fumigatus* RNAs will be provided by the following collaborators on a schedule to be developed.**

Dr. Brian Wickes, University of Texas San Antonio, USA
Dr. Paul Bowyer, University of Manchester, UK
Dr. Robb Cramer, Montana State University, USA
Dr. Corné Klaassen, Canisius Wilhelmina Hospital, Nijmegen, The Netherlands.
Dr. Nancy Keller, University of Wisconsin, USA
Dr. Scott Filler, University of California, Los Angeles, USA
Dr. Elaine Bignell, Imperial College, UK

Genomic DNA from *A. fumigatus* drug resistant strains. Strains and/or genomic DNA will be provided by David Denning, Manchester, UK

3d. Proposed Methods and Protocols to Prepare Reagents:

Genomic DNAs from reference and drug resistant strains.

Standard methods for high molecular weight gDNA preparations of genome sequencing projects.

***A. fumigatus* RNAs**

Total RNA preparations which meet specifications for Illumina RNA-seq application.

4. Approach to Data Production:

Reference strain genome sequencing.

The annotated high quality genome sequence will be developed for two additional strains of *A. fumigatus* selected to be different than Af293 and A1163.

RNA sequencing

Sequencing of cDNAs prepared from mRNA and from small RNAs will be employed to update the annotation of the four reference *A. fumigatus* strains. This annotation upgrade will be accomplished by JCVI.

The community scientists will analyze the expression profiles generated in this study and investigate the potential mechanisms of sRNA regulation of gene expression in *A. fumigatus*.

Drug resistance analysis

The genomic sequences from the drug resistant strains will be mapped back to the reference genomes at JCVI and a data resource of high quality genome sequence differences will be prepared**.**

The community scientists will analyze the genome sequence differences in the drug resistant strains to elucidate the mechanism(s) of drug resistance in these strains.

5. Community Support and Collaborator Roles:

The Aspergillus community. The Aspergillus community is well organized, holding annual meetings, termed Asperfest, as a satellite meeting in conjunction with the Fungal Genetics (Asilomar) and the European Community Fungal Genetics (rotating European cities) meetings, which are each held on alternate years. The actual size of the community is not precisely known but in recent years these meetings have been attended by ~160 Aspergillus researchers from around the world. Development of the Aspergillus Genome Database (AspGB) funded for an initial three year period is underway at Stanford University. This resource will help us obtain the number of these labs by allowing investigators to register their labs with the database. Since the publication of the *A. fumigatus* genome sequence, PubMed records of publication on this fungus have dramatically increased in number. The genome sequence was published in December 2005. The numbers of *A. fumigatus* PubMed citations for 2005 is 711 and for 2008 over 1800. The community is very technologically adept and has an Aspergillus Genomics Research Policy Committee (AGRPC), which annually updates the priority listing of resources to be facilitated by the community organization for development. Michelle Momany, AGRPC chair and a member of the JCVI infectious disease working group, communicates on a continuing basis with the community concerning our resources needs and has distributed the contents of this white paper proposal to the community mailing list, which resulted in correspondence indicating broad community support for this undertaking.

This white paper was developed with the direct participation of 11 PIs from the *A. fumigatus* community. Six of these were from US institutions and five from European institutions. Seven of these PIs volunteered to provide RNAs or genomic DNAs for the project. The resultant Executive Summary was reviewed by Michelle Momany representing the Aspergillus community and determined to be very much inline with the priorities that the community has raised repeatedly at the annual Asperfest meetings.

6. Compliance Requirements:

6a. Review NIAID's Reagent, Data & Software Release Policy:

http://www3.niaid.nih.gov/research/resources/mscs/data.htm

http://grants.nih.gov/grants/guide/notice-files/NOT-OD-08-013.html

Accept X Decline ☐

6b. Public Access to Data and Materials:

Strains. All *A. fumigatus* strains used in this project will be deposited with the Fungal Genetics Stock Center, the ATCC, and NCFP (Bristol, UK).

Annotated reference genomes. The four reference *Aspergillus fumigatus* genomes, the existing Af293 and A1163 genomes as well as the two new genomes sequenced in this project, with their annotations updated with the finding of this project including the RNA-seq and sRNA-seq data will be deposited to GenBank at NCBI. These annotated genomes will also be provided to the CADRE database (Manchester, UK) and the Aspergillus Genome Database (AspGD) database recently established with NIAID/NIH funding.

Expression profiling data. All information pertaining to mRNA and small RNA expression profiling including descriptions of the biological samples, processed data files, and transcription levels for each gene will be uploaded to the NCBI's GEO database.

Sequence reads. All original short read sequence reads will be deposited to the NCBI's Short Read Archive sequence database.

Metadata on drug resistant strains. Most of the data of the resistant strains is collected as a routine part of clinical care, and is anonymised. Most of it has already been presented in a publication, with supplementary files containing detailed strain data (Howard et al, Emerg Infect Dis 2009;15:1068.) http://www.cdc.gov/eid/content/15/7/1068.htm and http://www.cdc.gov/eid/content/15/7/1068-T2.htm
Any not already presented will be made available in the same format.

6c. Internal Review Board (IRB) / IACUE Yes ☐ No X

NIAID approval of IRB documentation is required prior to commencement of work.

No human subjects review is required as all samples are from strain collections and are devoid of any identifiers of human patients.

Investigator Signature: *David Denning* [electronically signed]

Investigator Name: David W. Denning **Date:** 14 July 2009