1. State the relevance to infectious disease for the organism(s) to be studied; for example the public health significance, model system etc.

**Background and Significance:** The most common CNS opportunistic infection in many regions of the developing world is from *Cryptococcus neoformans*, (CM) now accounting for more deaths in sub-Saharan than tuberculosis. With currently available therapy in resource-limited settings, mortality remains high (Amphotericin treatment: 50% mortality; fluconazole: 80-100% mortality). The availability of anti-retroviral therapy presents an ability to reduce AIDS-related death. However, since *C. neoformans* disease remains the leading cause of the initial presentation of HIV positivity, patients with cryptococcal disease have little ability to profit from the large amount of resources devoted to this Global pandemic. Despite these grim statistics, little is known regarding the clinical and genetic parameters associated with poor outcome in patients with this neglected disease.

The proposed study is to conduct an RNAseq analysis of 100 cryptococcal isolates in collaboration with T. Harrison and T. Bicanic at xxx, as a gene-finding study to determinants of poor outcome in cryptococcal disease. All isolates are serotype A isolates which cause the predominant amount of disease globally. Serotype A strain H99 (ATCC #208821) has been sequenced by the Broad institute, and the sequence annotated in reference to an EST database. However, the robustness of the annotation has not been verified in reference to a number of clinically important strains. RNAseq analysis for gene discovery has advantages in that direct evidence is provided for new gene predictions and to revise existing gene structures, non-translated sequences are available, and the extensive sequence analysis allows a MLST-type genotyping analysis that allows determination of the subtypes (CNVI, CNVII, etc) to determine if these play a role in clinical outcome. There is currently no RNAseq data publically available.

2. Are there genome data for organisms in the same phylum / class / family / genus? What is the status of other sequencing / genotyping projects on the same organism including current and past projects of the NIAID GSC? Provide information on other characteristics (genome size, GC content, repetitive DNA, pre-existing arrays etc.) relevant to the proposed study. Have analyses been performed on the raw data already generated/published? If additional strains are proposed for a species, please provide a justification for additional strains?

There is an assembly and predicted gene set for one strain of *C. neoformans var grubii* (serotype A). The assembly is 18.9 Mb and consists of 14 scaffolds of one contig each. The average GC content is 48.2%. This assembly should provide a good reference for the RNA-sequence proposed by this white paper.

There are currently no approved public projects to sequence additional strains of *C. neoformans*, although a white paper that would provide genome sequences for additional isolates of serotype A and other species is in development (C. Cuomo and A. Litvintseva). This study is complementary with the work in this proposal, and by working in parallel both would benefit from the synergy of sharing data, analysis methods, and discussing results.
3. If analyses have been conducted, briefly describe utility of the new sequencing or genotyping information with an explanation of how the proposed study to generate additional data will advance diagnostics, therapeutics, epidemiology, vaccines, or basic knowledge such as species diversity, evolution, virulence, etc. of the proposed organism to be studied.

As a demonstration of the utility of transcriptional profiling as a predictor of outcome, we previously had identified a CTR4 copper transporter as a predictor of outcome in a cohort of organ transplant patients. As a followup of this study, we have used transcriptional profiling by microarray to determine genes useful in classifying cryptococcal isolates from the organ.

**Figure 1:** Classifier analysis of *C. neoformans* isolates (M= meningitis, P= pulmonary) from a solid organ transplant cohort based on whole genome transcriptional profiling.

In the example given, several genes involved in copper acquisition were identified, which has been recently implicated to be important in host pathogenesis. In addition, the top classifier, in a random forest analysis, was found to be a topoisomerase, which could suggest that patients who develop meningitis could have a topoisomerase-induced increased genetic plasticity, influencing the ability of the organism to adapt and optimize virulence while in the environment, or in the host, during the long latent period typical for re-activation cryptococcosis. Interestingly, there was little relationship between the transcriptional clustering and MLST clustering, suggesting that alterations in virulence may be related to relatively small changes in overall sequence, manifested by MLST.

In the present study, we are hoping to develop new tools for using RNAseq both as a gene finding tool and as a classifier of isolates corresponding to clinical outcome. We are particularly interested in relationships between micro-evolutionary trends that happen both in the environment and in the host. We are currently collaborating with Dr. Adebowale Adeyemo at
the Center for Research on Genomics and Global Health, NHGRI, NIH as well as Christina Cuomo at Broad and welcome additional collaboration to both aid in the *C. neoformans* annotation as well as to develop novel methods of addressing questions of pathogenesis and gene discovery in this organism.

**Methods**

This discovery study will seek to identify genes highly expressed in strains having high initial CSF fungal burdens and poor outcome in a cohort of strains obtained from HIV+ patients in S. Africa.

Clinical isolates (see attached table) are primary isolates obtained by our collaborators T. Harrison and T. Bicanic in S. Africa. The samples were collected under the trial title High dose amphotericin B with flucytosine, and amphotericin B plus high dose fluconazole for treatment of cryptococcal meningitis in human immunodeficiency (HIV)-infected patients CLINICAL TRIALS REGISTRATION NUMBER: ISRCTN68133435, under an exemption approved at the NIH for use of de-identified samples and clinical data. Detailed clinical and microbiological data were obtained including outcome information, quantitative CSF cfu’s and MLST. These isolates are all serotype A strains, similar to the reference strain H99 in the Broad database.

**Fungal RNA from in vitro grown cells:** RNA from the 90 clinical isolates will be prepared with biological duplicates from fungal cells incubated in 3 h starvation in asparagine salts without glucose. This condition has been shown to mimic that encountered during CSF infection and was useful in predicting outcome of the CTR4 copper transporter in a cohort of organ transplant patients. Briefly, cells from primary culture will be prepared in duplicate, grown to mid-log phase (OD600= 0.3-0.6) and then incubated 3 h in asparagine salts without glucose. Cells are recovered by centrifugation, fractured by vortex-glass bead homogenation in the presence of 2M guanidinium thiocyanate, 2 mM mercaptoethanol and purified on a Qiagen RNA spin column. This method has been shown to provide highly purified, non-degraded RNA. RNA quality will be assessed using Nanodrop for quantity (260 nm), protein and organics contamination (260/280 and 260/230 ratios), and Bionalyzer to assess RNA integrity / degradation.

**Fungal RNA from mouse-brains.** Based on the transcriptional clustering, we will perform RNAseq on an additional 20 isolates recovered from mouse brains (10 high CSF fungal burdens and 10 low CSF fungal burdens) to help validate the clustering obtained from cells in vitro. Mice are inoculated with 1 x 106 cfu of the strains and observed for morbundity. Mice are then anesthetized and perfused through the heart with 2 M guanidinium thiocyanate, 2 mM mercaptoethanol to kill fungi and preserve RNA. Mice are then sacrificed, brains recovered, homogenized and digested with proteinase K for 30 min, followed by centrifugation on a discontinuous 60-80% sucrose gradient which separates intact fungal cells from brain matter. Separation will minimize contamination with mammalian RNA. Recovered cells are washed and fractured by vortex-glass bead homogenation as above. Purity of fungal RNA will be confirmed by RT-PCR of *C. neoformans* and mouse actin gene, using discimatory primers. Successful purification will be determined as less than a 0.05 ratio of mammalian to cryptococcal RNA by RT-qPCR.
For RNA sequence, libraries will be constructed from mRNA; strand specific methods can be used given sufficient quantity and quality of mRNA. RNA sequence reads can be used to refine the gene catalog using the Trinity software developed at the Broad Institute (pmid 21572440). Combining transcripts assembled by Trinity from the RNA sequence data with the existing gene set for H99 will provide an improved reference gene set which includes new transcripts, revised transcripts including UTR prediction, and highlight previously predicted genes without expression evidence which may be false positive predictions. The RNA sequence updating of the gene set can be done using data from all strains, and then the revised gene set will serve as a reference to measure gene expression that is specific to individual strains. In addition long RNAsequence transcripts which do not align to the reference assembly can be candidate genes missing from the assembly of H99 or from the H99 genome; these could be validated by PCR and could also be added to a reference set of transcripts.

While the clinical isolates have already been genetically characterized by MLST, it might also be useful to perform whole genome sequencing of representative isolates (up to 10) to relate genomic mutations in regulators to expression levels of prominent target transcripts. This will also complement the RNAseq work which may not yield sequence of regulators that are expressed at low levels.


**Collaborators:**

Tihana Bicanic, MBBS  
Tom Harrison, MBBS  
Desmond Tuto HIV Centre, Institute of Infectious Disease and Molecular Medicine, University of Cape Town, Cape Town, South Africa  
tihana@never.com

Christina Cuomo, PhD  
Group Leader, Fungal Genome Sequencing and Analysis  
Broad Institute  
cuomo@broadinstitute.org

Adebowale Adeyemo, MD  
Deputy Director  
Center for Research on Genomics and Global Health National Human Genome Research Institute, NIH  
adeyemoa@mail.nih.gov