Population genomic analysis of the human pathogenic fungus *Cryptococcus neoformans*

A white paper submitted by

Anastasia P. Litvintseva, Duke University Medical Center,

and Christina A. Cuomo, Broad Institute.

August 5, 2011.

PROJECT SUMMARY

Cryptococcus neoformans is an opportunistic fungal pathogen of enormous clinical importance. A recent study from the Centers for Disease Control estimates that approximately one million new infections occur each year (53). In Africa, cryptococcosis is the third most common cause of death in patients with HIV/AIDS, and mortality from cryptococcal meningitis exceeds the death rate from tuberculosis (53). *C. neoformans* is also a model organism for studying the virulence of pathogenic eukaryotes (24).

Recent phylogenetic studies highlighted the diversity and complex evolutionary relationships among genetic groups within the *Cryptococcus* species complex. The two pathogenic sibling species, *C. neoformans* and *C. gattii* , are members of the *Filobasidiella* clade of the Order Tremellales, which also includes several nonpathogenic saprobic fungi associated with the arthropodal frass, decayed vegetation and soil (14). Both *C. neoformans* and *C. gattii* are environmental, non-contagious pathogens, but they differ in key genetic, ecological, physiological and pathobiological properties (29). Within each species, discrete genetically isolated subgroups have been identified, which are morphologically indistinguishable but share 89-92% nucleotide identity and can be classified as cryptic species (2, 27, 49). Specifically, *C. gattii* is comprised of four genetically isolated groups designated VGI to VGIV molecular types, and *C. neoformans* is comprised of two varieties, *C. neoformans* var. *grubii* and *C. neoformans* var. *neoformans*, which are commonly denoted by their traditional capsular serotypes, A or D, respectively. In preliminary studies, we developed multilocus sequencing typing (MLST) to genotype *Cryptococcus* and demonstrated that each variety can be further subdivided into genetically isolated subpopulations, which are often restricted to specific geographic areas and/or occupy different ecological niches (7, 40, 47). In addition, our population genetic analysis discovered that not all strains are equally pathogenic, and strains with certain genotypes are more likely to cause disease in humans (35, 37).

Despite the extensive genetic and phenotypic diversity among wild type strains of *C. neoformans*, most recent studies of pathogenesis focus on just three model strains (H99, JEC20 and JEC21), which restricts interpretation of the results to the strain- or model-specific framework. Rather than continue to concentrate on a small number of laboratory strains, we propose to obtain and analyze whole genome sequences of 330 genomes of *Cryptococcus* representing different species, molecular types, subpopulations, genotypes, geographic regions, ecological niches and patient cohorts. Comparison of several pathogenic and nonpathogenic species in the *Cryptococcus* species complex will provide insights into the core set of genes that are responsible for their specialized ability to survive inside the mammalian host and cause human disease. By comparing and contrasting strains from different populations and ecological niches, we will identify genetic mechanisms that are responsible for adaptation to diverse ecological conditions, increased human virulence and fitness. Analysis of strains from different geographic regions will allow us to estimate the extent of clonality and sexual reproduction in the global population. The availability of multiple genomic sequences of *C. neoformans* will expand our understanding of pathogenesis beyond the model strains, facilitate comparative genomic and evolutionary studies, provide a rich supply of genetic markers and define testing populations for subsequent epidemiologic and genome-wide association studies. All results will be freely available to benefit the global scientific community.

Based on the preliminary data obtained by multilocus sequencing typing (MLST), population genetics and phylogenetic analyses of the *Cryptococcus* species complex (14, 40, 49), we selected the minimum of isolates required to address the specific aims. This proposal is primarily focused on *C. neoformans* var. *grubii* (aka, serotype A) because this variety is responsible for >90% of all cryptococcal infections worldwide. Consequently, the selected strains target specific questions about the molecular epidemiology, evolution and pathogenesis of *C. neoformans*. In addition, this project will survey the diversity of *C. gattii* and related nonpathogenic Tremellales species. For this purpose, we will sequence representatives of the four major

lineages of *C. gattii* and 1 to 3 isolates of each closely related species of Tremellales. Table 1 summarizes the genomes targeted by this project. All of these strains are available in the laboratories of the collaborating PIs (see Table 2 attached after References).

Table 1. Genetic groups targeted for whole genome sequencing by aims 1, 3 and 4 of this proposal.

Species group	Serotype	Phylogenetic groups	Genomes for *de novo* assembly	Genomes for resequencing
C. neoformans var. *grubii*	serotype A	3 clades (VNI, VNII, VNB); 111 unique MLST genotypes; Fig. 4*	113**	159
C. neoformans var. *neoformans*	serotype D	2 lineages; Fig. 3	15	0
C. neoformans hybrids	serotype A/D	3 lineages; Fig. 3	14	0
C. gattii	serotypes B and C	4 clades (VGI, VGII, VGIII, VGIV); Fig. 2	15	0
Other Tremellales species		12 species; Fig. 1	18	0

* For space considerations, Fig. 4 includes only 24 of the 99 unique genotypes found in the African population.

** Two strains, IUM 96-28-28 (64) and CDC15 (61), have not been genotyped by MLST. However, they are included because of their interest to the research community (see Table 2), and their genomes will be assembled *de novo*.

BACKGROUND

Medical impact. Although cryptococcosis occurs in people with apparently healthy immune systems, it is the most prevalent life-threatening fungal infection among patients with HIV/AIDS. A recent report from the US CDC estimated that the rate of cryptococcosis in sub-Saharan Africa is approaching one million cases per year with mortality rates that may exceed 50% (53). In developed countries, the incidence of cryptococcosis is much lower, but it remains a leading opportunistic infection of patients with solid organ transplants, hematologic malignancies and HIV/AIDS (11). Even when early, appropriate therapy is available, cryptococcal meningoencephalitis has a mortality rate of ≈40% (11, 55). Cryptococcosis can be caused by either of two species, *C. neoformans*, which is characterized by haploid isolates with the A or D capsular serotype, as well as AD hybrids, or *C. gattii,* traditionally denoted by serotype B or C. However, more than 90% of infections worldwide are due to haploid strains of *C. neoformans* var. *grubii*, which possess the serotype A capsular epitope (4).

Phylogenetic relationships and populations structure. Phylogenetic relationships among species of Tremellales are shown on Figure 1. *C. neoformans* and *C. gattii* are the only species that are capable of causing disease in mammals; the other Tremellales are saprobic and found on plants, decayed wood, soil and insect frass (14). In addition, *C. laurentii*, *C. adeliensis* and a few other species of *Cryptococcus* are infrequently isolated from mammals; however, despite sharing the generic name with pathogenic cryptococci, these species are phylogenetically unrelated to Tremellales and therefore will not be included in this study (48).

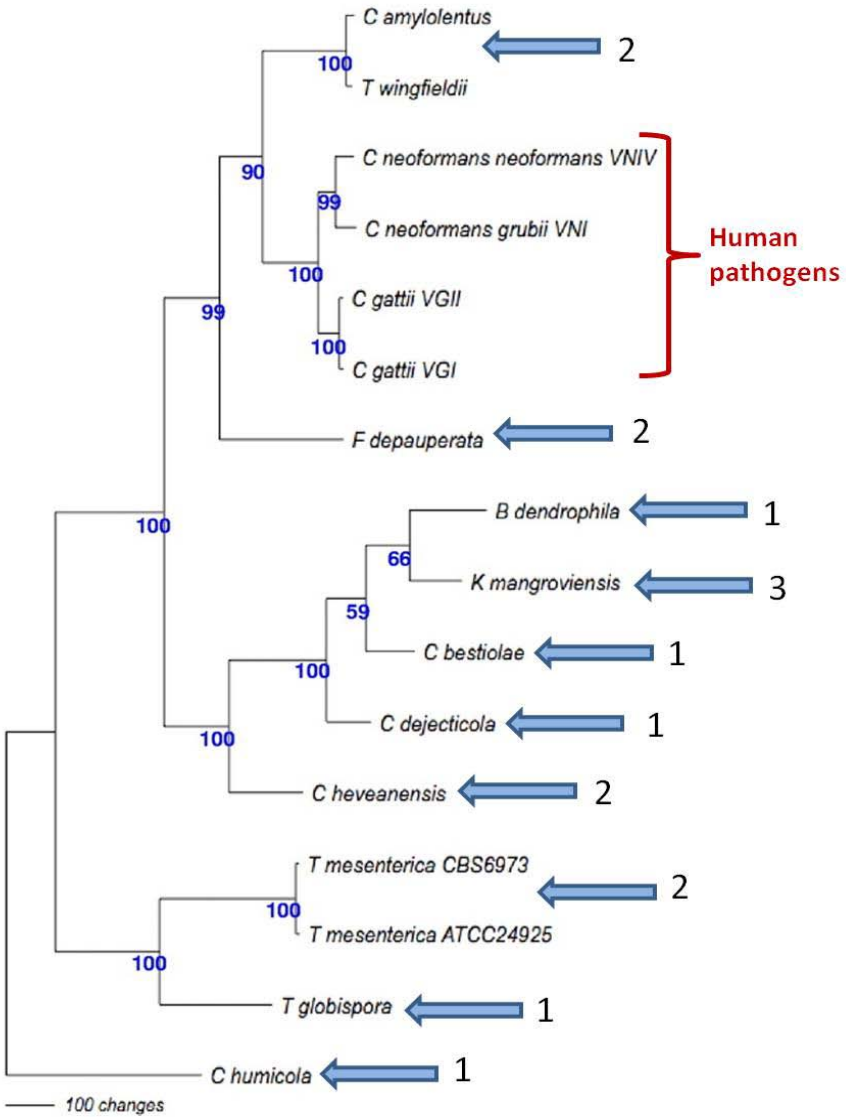

Figure 1. Phylogenetic relationships among members of Tremellales. Maximum parsimony phylogeny is based on sequence polymorphisms in six concatenated loci (adapted from (14)). Numbers below branches are bootstrap values based on 1,000 replicates. Numbers next to the arrows show the numbers of strains of each species that will sequenced in this project.

In our preliminary studies we developed two robust genotyping methods, amplified fragment length polymorphisms (AFLP) and multilocus sequence typing (MLST), and genotyped strains of *C. neoformans* and *C. gattii* from 15 different countries. Our studies and results by others demonstrated that *C. gattii* is comprised of four genetically isolated molecular types, designated VGI to VGIV (Figure 2), which likely represent cryptic species (15, 29, 39, 49). VGI strains are global and known to cause disease in both immunocompetent and immunocompromised individuals. VGII strains are responsible for the ongoing outbreak on the Vancouver Island and infect healthy individuals; VGIII and VGIV strains primarily infect humans with impaired host defenses; these strains are geographically restricted to south-western part of North America and southern Africa, respectively.

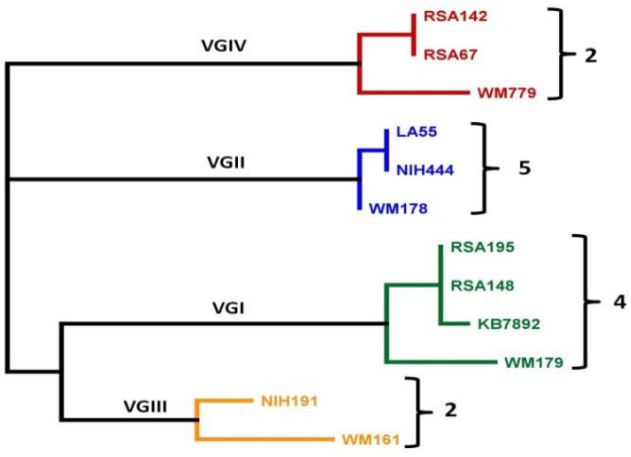

Figure 2. Phylogenetic relationships among molecular types of *C. gattii*. This maximum parsimony phylogram is based on sequence polymorphisms in the *LAC1* gene. Four major molecular types, which are probably cryptic species, are shown. Numbers next to each clade indicate the number of strains of each molecular type that will sequenced.

C. neoformans is comprised of two varieties, *C. neoformans* var. *neoformans* (serotype D) and *C. neoformans* var. *grubii* (serotype A). In addition, diploid AD hybrid strains are isolated from clinical and environmental samples (36). The global population of serotype D is comprised of three genetically isolated lineages (Figure 3). Strains from lineages 1 and 2 are global and widespread in the environment and in patients. Serotype D counterparts from lineages 1 and 2 are commonly present in AD hybrids. Conversely, strains from the lineage 3 are rare, and to date, they have only been isolated from unique AD hybrids isolated from patients in France (unpublished data).

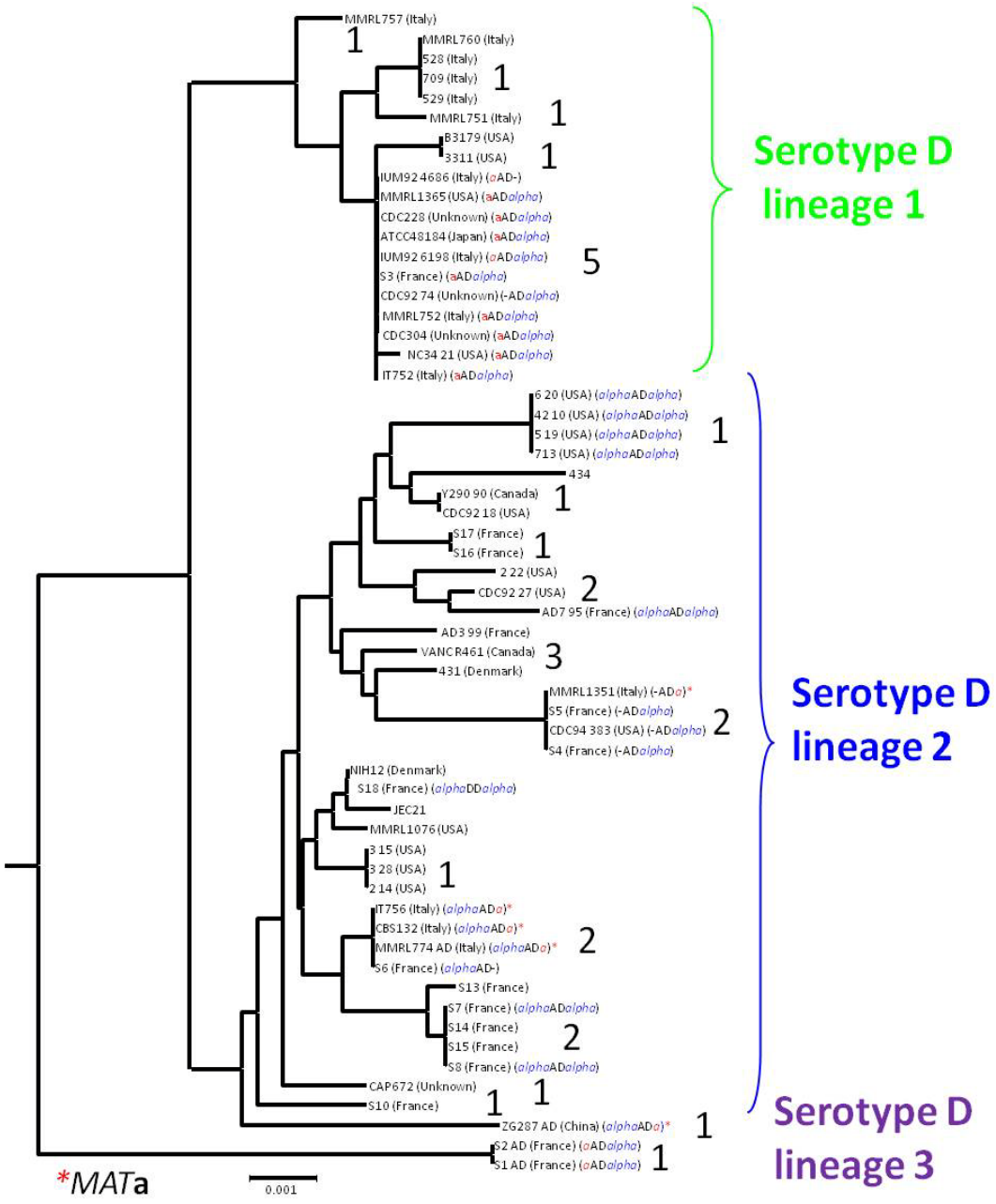

Figure 3. Genetic relationships among 58 strains of serotype D and AD hybrids visualized by Neighbor Joining dendrogram, which is based on polymorphisms in five concatenated MLST loci. Three main lineages are shown. Numbers next to each clade designate the number of strains that will be sequenced.

Figure 4 presents the genetic relationships among strains of serotype A. MLST genotyping using 12 unlinked loci identified three genetically isolated subpopulations, VNI, VNII and VNB. Strains of VNI and VNII are ubiquitous, highly clonal, and cause most cryptococcal infections outside of Africa. In contrast, VNB isolates are highly genetically variable and apparently confined to sub-Sahara Africa. In addition, 25% of the clinical isolates of VNB from Botswana possess the rare *MATa* mating type allele (37).

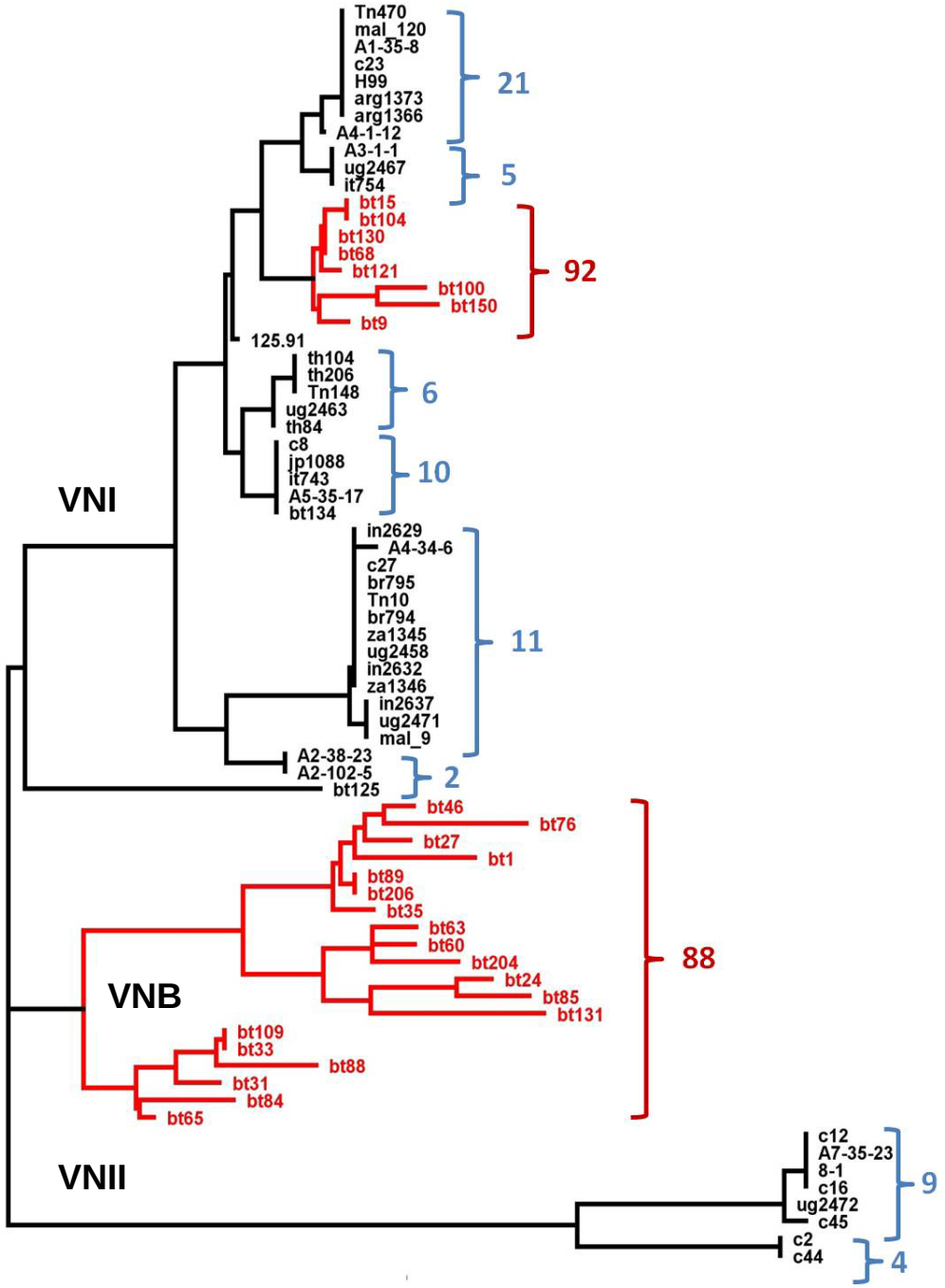

Figure 4. Genetic relationships among 84 strains of serotype A visualized by Neighbor Joining dendrogram based on 12 concatenated MLST loci (adapted from (40)). Strains unique to Africa are shown in red; these clades are designated VNB, as labeled, and VNI-Botswana (the red branches within the VNI clade). Blue brackets show candidate strains for sequencing in aim 1; red brackets show candidate strains for sequencing in aim 3. Numbers next to the brackets show the number of strains of each genotype that will be sequenced.

Population genetic analysis identified exceptional genetic diversity in the African population. Forty-nine MLST genotypes were identified among 142 southern African strains, but only 10 genotypes were detected in the global sample of 102 strains (40). Principal component analysis (PCA) and phylogenetic methods detected a remarkable correlation between the genotypes and ecology: endemic African strains (VNB and VNI strains from Botswana) were associated with native trees, whereas ubiquitous genotypes (VNI strains) were isolated from their previously established niche, pigeon excreta (Figure 5). Haplotype network analysis indicated that ancestral haplotypes of all eight loci are extant in strains obtained from endemic African trees. Conversely, haplotypes of global isolates are associated with pigeons and occupy distal positions on the networks, which is consistent with their recent origin. The combined evidence from population and phylogenetic analyses supports the African origin and recent global expansion of serotype A (34).

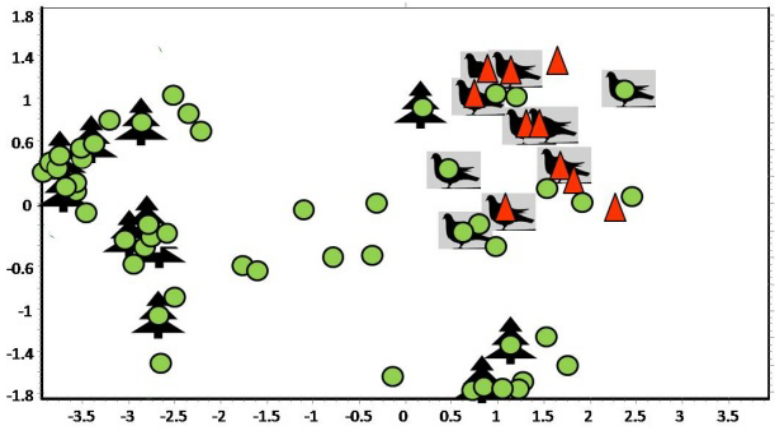

Figure 5. PCA plot of 59 MLST genotypes. Green circles represent genotypes of isolates from endemic African strains, and red triangles represent global genotypes. Genotypes associated with African trees are marked with tree icons, genotypes associated with pigeon excreta are marked with pigeon icons, and genotypes without icons represent clinical strains not yet recovered from the environment. Adapted from (34).

In this project, we will use comparative genomic analysis to define speciation boundaries within this *Cryptococcus* species complex. Comparative genomic analysis of pathogenic and nonpathogenic species will distinguish genes that enable *C. neoformans* and *C. gattii* to establish infection and identify the basis of virulence in *Cryptococcus*.

Sexual reproduction. The majority of natural isolates of *C. neoformans* var. *grubii* (aka, serotype A) are haploid, but they possess one of two mating type alleles, α or **a**, and in the laboratory, strains of opposite mating type are capable of sexual reproduction (50). However, most natural isolates only possess the α mating type (*MAT*α), and this dominance of a single mating type minimizes the possibility of conventional sexual reproduction in nature. Sexual reproduction among strains of serotype A is geographically restricted to the ancestral population in southern Africa, which harbors strains with the rare *MAT***a** mating type and demonstrates evidence of recombination. Conversely, 99.9% of the cosmopolitan strains of serotype A possess *MAT*α mating type. Among isolates of *C. neoformans* var. *neoformans* (serotype D), less than 2% possess the *MAT***a** mating type (51).

The extent of clonality and recombination in the natural populations of *C. neoformans* remains to be determined. Population genetic analysis indicates that despite the predominance of *MAT*α strains in the environment, most natural populations exhibit some evidence of linkage equilibrium (41). However, an alternative explanation for the observed linkage equilibrium among the loci in unisexual populations is the potential of same sex mating between isolates with the *MAT*α mating type (32, 33). This project will use comparative genomic analysis to estimate the extent of sexual reproduction and recombination in different populations of *C. neoformans*.

Human infection and virulence. During propagation in the environment, basidiospores or desiccated yeast cells become airborne, and inhalation of these infectious propagules leads to a pulmonary infection. In the alveoli, cryptococcal cells become encapsulated yeasts and proliferate. In people with healthy immune responses, infection may remain latent for years and rarely cause disease. However, in people with impaired

immunity, the yeasts can disseminate to almost any part of the body, but preferentially to the central nervous system, where meningoencephalitis is the usual and deadliest clinical manifestation.

The pathogenicity of *C. neoformans* is polygenic. Using well characterized laboratory strains and gene-disruption experiments, a number of virulence factors have been identified, and the most important include ability to produce a polysaccharide capsule and melanin, grow at 37°C, and secrete phospholipase B and urease (28, 42, 45, 54). In addition, several signal transduction pathways have been implicated in the regulation of pathogenicity (8, 9, 13, 52, 62). Although these previously characterized virulence factors are essential for virulence, they are not necessarily sufficient and do not predict the pathogenic potential of a strain. Multiple studies have noted that many wild-type strains are capable of producing these virulence factors, but they are unable to cause disease in animals (17, 18, 46). Our recent data demonstrate that environmental isolates of serotype A are unable to cause disease in the murine model of cryptococcosis, which is considered the most valid surrogate for human cryptococcosis; however, most clinical isolates with the same molecular genotypes as the benign environmental strains are lethal for mice (38). Regardless of genotype, two hypotheses can explain this apparent paradox: (i) very few environmental strains are inherently virulent or (ii) most environmental isolates are potentially pathogenic, but they require prior incubation and selection in the host to become virulent. In this project, we will utilize comparative genomic and gene expression analyses of clinical, environmental and experimentally evolved strains to explore these hypotheses and identify the genetic mechanisms of virulence.

RESEARCH QUESTIONS OF THE PROPOSAL

This project will address the following fundamental questions about biology and epidemiology of *C. neoformans*:

- Identify core set of genes that are responsible for specialized ability to colonize mammals;
- Define speciation boundaries within *Cryptococcus* species complex;
- Determine the genome-wide estimates of linkage disequilibrium and the extent of sexual reproduction in different populations;
- Identify sets of genes and/or markers that coincide with the ecological transition from decayed wood to avian excreta, and subsequently, to the emergence of global clonal populations of serotype A (34);
- Identify specific genes/markers that are associated with significantly high prevalence in patients;
- Identify set of genes/markers that are associated with (i) increased morbidity and/or mortality in patients, (ii) the ability to infect immunocompetent individuals, (iii) high prevalence in pediatric patients; (iv) and increased fitness;
- Using RNA-Sequence, identify genes that are differentially regulated and/or mutated in clinical and/or environmental strains;
- Determine the extent of alternative splicing and posttranslational regulation in *C. neoformans*;
- Using experimental evolution, identify transcriptional modifications/genomic rearrangements and/or mutations associated with adaptation to *in vitro* stress/starvation and increased/decreased virulence.

Strains for this project have been selected to target these specific questions. In particular, the selection of strains from different phylogenetic groups/populations will allow us to evaluate phylogeographic associations, and potential speciation barriers within the *Cryptococcus* species complex. Most importantly, these strains will enable us to identify a core set of genes responsible for human infection. Comparing the sequences of environmental and clinical strains will identify markers associated with human infection and/or pinpoint sources of human infection. The inclusion of strains from the ancestral recombining population in Africa will determine the origin(s) of the global population, assess the extent of linkage disequilibrium, and identify strains/markers

that are more likely to cause infection in Africa. Comparing strains from different geographic areas will allow us to evaluate the extent of clonality in the global population. Analyzing strains with identical genotypes but different phenotypes and/or isolated during different stages of infection will allow us to identify markers associated with clinically relevant phenotypes, such as specific virulence factors, tissue tropisms and patient prognoses (e.g., latent period, lethality), as well as identify mechanisms of phenotypic variations in *C. neoformans*.

Because the genomic size of *Cryptococcus* is relatively small (20 Mb), it is feasible and economical to analyze numerous genomes and transcriptomes at relatively low cost. In addition, it is possible to obtain significant quantities of pure DNA and RNA, and all strains for this project are available in existing collections. To obtain a representative and comprehensive sample of genomic diversity of *C. neoformans*, we propose four focus areas for this proposal, described below.

SELECTION OF STRAINS, GROWTH CONDITIONS AND SEQUENCING APPROACH

Aim 1. Obtain and annotate whole genome sequences of key, representative strains in *Cryptococcus* species complex. We will use our preliminary MLST data to select strains for whole genome sequencing (14, 40). To achieve an unbiased and comprehensive sample of global strains, we will include representative strains of all common MLST genotypes from *C. neoformans* var. *grubii* and var. *neoformans*, as well as common AD hybrids. In addition, to provide outgroups for subsequent phylogenomic analysis, we will include representatives of each molecular type of *C. gattii* (VGI to VGIV) as well as several closely related nonpathogenic species from Tremellales (14). To assess the extent of clonality among different geographic and ecological samples, when available, we will include clinical and environmental representatives of each MLST genotype from each geographic region and ecological niche. In addition to the common MLST genotypes, which will be defined as genotypes that have been isolated from more than one population sample, we will include several phylogenetically informative rare strains, such as those that possess rare mating types and/or carry ancestral MLST alleles (34, 36). Overall, we anticipate obtaining whole genome sequences of 141 strains: 79 genomes of serotype A, 15 serotype D, 14 AD hybrids, 15 *C. gattii* and 18 related nonpathogenic species. The resulting genomic data will be used to (i) determine phylogenetic relationships among the strains/populations, (ii) define species/population boundaries, and (iii) identify genes/markers associated with different ecological niches and human disease.

To facilitate phylogenomic analysis of the *Cryptococcus* species complex and to identify core set of genes that enable *Cryptococcus* to cause infection, we will sequence, assemble, and annotate the following strains:

1. 18 nonpathogenic strains representing major lineages within Tremellales (Figure 1). To assess diversity within the species, when available, we will include 2-3 representative strains of each species. (Two novel species, *C. pinus* and *C. shivajii*, are not shown on Figure 1.)
2. 15 strains of *C. gattii*, including eleven representatives of VGI and VGII clades, which are the most diverse and clinically relevant molecular types, and four representatives of the VGIII and VGIV clades, which are less common in humans (Figure 2). To the best of our knowledge, we did not include strains that have been already sequenced and/or are being proposed for sequencing by other groups.
3. 110 strains of *C. neoformans*, including 79 strains of serotype A (*C. neoformans* var. *grubii*), 15 strains of serotype D (*C. neoformans* var. *neoformans*) and 14 AD hybrids (Figures 3 and 4). To capture genetic diversity in the natural populations of this most clinically relevant species of *Cryptococcus*, we included representatives of the all common MLST genotypes, which are defined as genotypes that were isolated from more than one population sample, as well as several phylogenetically informative rare strains, such as those that possess rare mating types and/or carry ancestral MLST alleles (34, 36). To assess the extent of

clonality in the populations, we added multiple strains of several common MLST genotypes that were isolated from different geographic regions as well as clinical and environmental sources.

Proposed approach for Aim 1. We will follow the current approach at the Broad Institute for sequencing and *de novo* assembly of small eukaryotes. All sequencing will be done with Illumina technology. For a subset of the serotype A strains, where groups of mostly clonal isolates will be sequenced, we will select one isolate from each clonal group for complete *de novo* sequencing and assembly. For the other isolates in each clonal group, we will use a re-sequencing approach, where reads from a single fragment library will be obtained for each strand compared to a reference genomic sequence to identify variation. Specifically, for 63 serotype A strains that are clonal, we will generate 100X fragment coverage of 101 base reads from one 180 base library and align these reads to a reference assembly chosen for each clonal group to identify SNPs, indels, and regions of copy number variation. For the 78 genomes requiring *de novo* assembly, we will generate approximately 100X paired end coverage by obtaining 101 base reads from each of two small-medium insert libraries (180 bases and 3-5 kb). We recently used this approach to sequence and assemble the genomes of several other fungi with haploid genomes similar in size to *Cryptococcus* and achieved high quality results. Based on previously sequenced strains of *C. neoformans*, we expect the genomes to be 19 Mb in size, 48% G+C, and minimally repetitive, and therefore to perform well in sequencing and assembly (43).

Because AD hybrids contain two similar haploid genomes, we will deploy a different approach to sequence and assemble these strains. Previous work comparing the sequences of serotype A and serotype D strains D (H99 and JEC21, respectively) determined that these genomes are highly syntenic (27) with only 32 rearrangements disrupting the collinear chromosomal structure (61). By aligning the JEC21 and H99 at the nucleotide level, we find that they share 89% identity on average. At this level of nucleotide similarity between the two genomes, it may be possible to assemble both the A and D genomes from the AD hybrids. A strategy of using longer Illumina reads of 150 bases for assemblies is currently being tested in other genomic projects at the Broad Institute, and if proven successful, we will employ this approach. However, if it is not possible to assemble all the data directly, we will bin the AD hybrid reads by aligning them to the H99 and JEC21 genomes and assemble the reads that align to each genome separately. As the nearly completed JEC21 and H99 assemblies are of high quality, they will provide superb references to bin sequence reads of the AD hybrids. We will evaluate the completeness of the AD hybrid assemblies obtained by either approach by aligning assembled sequences to the H99 and JEC21 reference genomes and determining the extent of coverage and similarity of each A and D hybrid genome.

Comparative analysis of these genomic data will be carried out using existing methods. Specifically, we will build gene families across the genomes using OrthoMCL (5, 6) or a similar program and examine patterns of genetic conservation. Significantly enriched gene families or functions will be identified using a hypergeometric test with correction for a false discovery rate of 5%. This comparison will allow us to identify the core set of genes in the pathogenic species complex as well as genes specific to subsets of these species, such as those that differentiate the pathogenic lineages of *C. neoformans* and *C. gattii*. In addition, we will align orthologous groups with MUSCLE (57) and use paired likelihood tests in PAML (65) to detect evidence of selection or higher rates of evolution in specific lineages.

We will use principal component analyses (PCA), hierarchical clustering (40) and Bayesian methods with the software STRUCTURE (12, 56) to assess the genetic relationships among strains from different populations and/or geographic regions. We will determine the extent of genetic variation within each subpopulation and identify subpopulations with the highest level of genetic variation by computing the average rate of pairwise differences in SNPs between the members of each cluster. In addition, we will analyze population divergence by computing the F_{ST} statistic for all pairwise combinations of isolates to examine the level of genetic differentiation and gene flow among populations.

Aim 2. Use RNA-Sequence to assess the extent of variation among transcriptomes, provide experimental annotation of the expressed genes, and estimate the level of alternative splicing in *C. neoformans*.

Whole genome sequence data from aim 1 will be supplemented by gene expression data obtained by RNA-Seq. RNA will be isolated from 12 representative strains of *C. neoformans* and *C. gattii*, which will be selected based on the preliminary MLST analysis and include strains with MLST genotypes, which are most frequently isolated from the environment and patients. To identify genes, whose expression correlates with human infection, we will include both environmental and clinical isolates of each genotype (38). Strains will be grown under six conditions that simulate growth inside the host and in the environment (e.g., synthetic media that contain limiting amounts of nutrients and oxidative agents; rich medium; macrophage-like cells; rabbit CSF, sterilized pigeon feces and boiled tree bark extract). To achieve statistical significance of data, experiments will be performed in duplicate. The resultant transcriptome data will be used to (i) improve annotation of the existing genes, (ii) identify new genes and non-coding RNAs, (iii) estimate the extent of alternative splicing, and (iv) identify genes whose expression is correlated with adaptation to growth in the environment and/or mammalian hosts.

We propose to obtain gene expression patterns of eight strains of serotype A, two strains of serotype D and two of *C. gattii.* Specifically, we will include 8 previously characterized strains of serotype A with known murine virulence, which represent three most prevalent MLST genotypes in the global population, A1/M1, A4/M4 and A5/M5 (38). To identify differences in the gene expression between clinical and environmental strains, we will analyze at least one clinical and one environmental strains of each genotype. We will include H99 strain as the most common reference strain for gene expression studies. In addition, to assess the extent of variability at the gene expression levels between serotypes A and D, we will include two strains of serotype D representing linage 1 and 2 (Figure 3) and two strains of *C. gattii*, R265, the major outbreak strain, and a strain from the same VGII group isolated from the environment on Vancouver island.

To capture the diversity of expressed genes, RNA will be isolated from strains growing under six different conditions, designed to mimic environmental and *in vivo* growth of *C. neoformans*. Growth conditions within the mammalian host will be simulated by propagating strains in (i) low-iron starvation medium supplemented with hydrogen peroxide to produce reactive oxygen species, similar to the starvation and oxidative stress conditions in the host (10, 23, 26), (ii) inside activated J774.1A macrophage-like cells, which correspond to the initial stages of infection within the pulmonary alveoli (3, 44), and (iii) in rabbit CSF using the established model of experimental meningitis, which replicates the terminal stage of infection (30). Our preliminary microarray data indicate that each of these conditions distinguishes lethal and non-lethal strains, and the combination provides a broad assessment of the critical host environments (data not shown). Environmental growth conditions will be simulated by growing strains in the following culture media: (i) Yeast Peptone Dextrose (YPD) broth, which is nutritionally replete, (ii) pigeon excreta broth, which consists of 15% (w/v) suspension of sterile, desiccated pigeon feces to simulate growth in the most common ecological niche, and (iii) boiled extract of decaying hardwood bark, another natural reservoir. To achieve statistical significance, each experiment will be performed in duplicate. In this aim, we anticipate a total of 144 samples for RNA-Seq.

Proposed approach for Aim 2. For each strain and experiment described above, we will use the Broad Institute protocol to generate strand-specific Illumina adapted libraries from polyA purified mRNA. For each cDNA library of *C. neoformans* transcripts we will aim to generate 25M Illumina 75-bp reads, which is expected to provide an average genic coverage of ~150X per strain. This approach should detect expression across a wide range of expression values and provide sufficient reads per transcript to reveal significant differences in gene expression. Although several *Cryptococcus* genomes have been annotated, none have yet been evaluated or revised using RNA-Seq. Completing the sequence proposed by this aim will enable us to use

RNA-Seq data to predict genes in each of the 12 genomes. We anticipate that the data this will provide evidence supporting the existence of previously overlooked short genes as well as invalidate some previously predicted genes. In addition, as splicing in *C. neoformans* is fairly complex for fungi, with six exons per gene, we will use RNA-Sequence to refine exon predictions and evaluate alternative splicing, as well as antisense or other noncoding transcripts. The revised annotations enable more accurate analyses of transcriptional differences in gene expression among different strains and experimental conditions. For these comparisons, we will use software such as the DEseq R package developed for RNA-Seq measurements and the MultiExperiment Viewer to cluster and view expression values (1).

Aim 3. Estimate linkage disequilibrium in the African population of serotype A and identify genetic markers associated with increased/decreased probabilities of causing cryptococcal disease and emergence of global strains.

While global populations of serotype A and D are highly clonal, a unique subpopulation of serotype A in sub-Saharan Africa is characterized by high genetic diversity, the presence of both mating types and evidence of recombination, which are characteristic of the ancestral population of serotype A (34, 37). Despite the enormous diversity of genotypes in the African population, almost half (46%) of all patients in Botswana and more than 25% of patients in South Africa are infected by strains of only three closely related genotypes; however, none of these molecular genotypes is dominant in strains from the African environment (unpublished data). In this aim, we propose to use whole-genome sequencing to genotype 180 clinical and environmental strains from Botswana. These data will be used to estimate linkage equilibrium and the extent of sexual reproduction in the African population. The results will also enable us to identify genomic regions that are significantly associated with a high or low probability of causing cryptococcal disease, the transition to a different ecological niche and/or the emergence of cosmopolitan clonal lineages.

Our previous studies described the unusual population in southern Africa that is characterized by (i) a significant proportion of isolates with the *MAT***a** mating type, which is almost extinct from the global population, (ii) unprecedented genetic diversity, and (iii) evidence for genetic recombination (37). Our recent phylogenetic and population genetic analyses demonstrated that this population is ancestral, and the emergence of as few as two strains from Africa can fully account for the genetic diversity of the global population of serotype A (34). In this aim, we will obtain whole genome sequences of clinical and environmental strains from Africa. By comparing their sequences with the global strains, we will identify genetic markers that coincide with the change of ecological niche and emergence of global strains (34).

High genetic diversity and frequent recombination makes the African population an ideal target for GWAS. Genomic data from aim 3 will provide the necessary preliminary data for the subsequent project to identify regions of genome that are associated with virulence in serotype A. Although we do not yet have a complete set of clinical and environmental strains to achieve statistical significance of GWAS, we recently received funding from NIH to identify genes and markers associated with virulence in *C. neoformans*. Within the next year, we anticipate obtaining a statistically meaningful sample of environmental and clinical strains, including the corresponding ecological and clinical data. Genomic data from this aim will provide necessary preliminary data for this forthcoming GWAS project. More specifically, this aim will identify numerous polymorphic markers, define target populations and provide genome-wide estimates of the linkage equilibrium. For this aim, we will sequence 15 strains isolated from the environment in Botswana and South Africa, and 165 clinical strains (detailed list is shown in Table 1).

Proposed approach for aim 3. As in aim 1, we will select a subset of the strains for *de novo* assembly. One representative strain from each clonal group will be sequenced and assembled, and the other clonal isolates will be sequenced to 100X and compared to the assembled reference strain. Of the 180 total strains in this

aim, we plan *de novo* assemblies of 99 genomes using 200X coverage of paired end 101 base Illumina reads (180-bp and 3-5 kb libraries). The remaining 81 strains will be sequenced to 100X using paired end reads from 180-bp fragments and variants identified as in aim 1.

Preliminary analysis of the MLST data revealed that the Botswanan population consists of two partially isolated subpopulations: VNB and VNI-Botswana (see Figure 4) (40). In addition, previous analysis of AFLP data suggested that the VNB subpopulation may be subdivided into three subgroups (viz., VNB-a, VNB-b, and VNB-c) (37, 40). However, neither MLST nor AFLP genotyping provided sufficient resolution to determine the statistical significance of these subdivisions. In this aim, we will test for evidence of the geographic or ecological subdivision in the African population by using PCA, hierarchical clustering (40) and STRUCTURE as described in aim 1. In addition, we will determine the extent of genetic variation within each subpopulation and identify subpopulations with the highest level of genetic variation by computing the average rate of pair-wise differences in SNPs between the members of each cluster. We will compare genetic diversity between global and African population samples. This project should definitively enhance our preliminary findings by providing a larger and more varied number of strains and an increased number of markers (SNPs). This critical expansion of the data will result in a finer scale of resolution of each cluster and a more accurate estimate of genetic diversity within each subpopulation.

To determine the evolutionary relationships between the African and global populations, we will compare their genomic sequences, which will be obtained in aims 1 and 3. To account for small and large scale variation between these strains we will take two approaches. First, we will compare all African strains to two reference genomes from Africa, representing the VNB and VNI-Botswana subpopulations, and obtain genotypes for all SNP loci across all strains. SNPs with biased representation in the different niche groups of strains will be identified, and will be grouped based on location to determine which large genomic regions have segregated based on strain niche. Second, to identify SNPs and other variations within the entire set of sequences, both African and global, we will identify orthologous genes with OrthoMCL (5, 6) and align the orthologs using MUSCLE (57) as described in aim 1. We will determine if sequence regions are missing in some strains by examining the representation of orthologous genes across the sequenced genomes. Aligned 1:1 orthologous genes represented in all genomes will be concatenated into independent alignment files, and the phylogeny will be inferred using maximum parsimony and Neighbor joining methods implemented in PAUP and maximum likelihood using RAxML method (59, 60). On the phylogenetic tree, ancestral haplotypes are expected to occupy basal positions and should have a sibling relationship to all the other haplotypes. We will root the reconstructed gene trees with sequences from the reference strain of *C. neoformans* var. *neoformans* (serotype D) and *C. gattii*. We will compare the results of different phylogenetic methods and alignment strategies determine which haplotypes (from which subpopulations) occupy basal positions on the tree. If bootstrap analysis is not possible, congruent results across methods will greatly increase confidence in the inferred topologies. Presence of ancestral haplotypes in the African population will support an African origin of serotype A, whereas presence of ancestral haplotypes in multiple locations will suggest other models of evolution.

We will test for evidence of recombination in African subpopulations by assessing the extent of linkage disequilibrium (LD) across each of the scaffolds in the reference assembly. LD in the population will be estimated by calculating r^2, a standard statistic that describes the joined and independent frequencies of paired loci in the population (31, 63). We will plot r^2 against the physical distance between SNP markers. (The program HAPLOVIEW will be used to calculate the average distance between these markers.) Areas of the genome that exhibit a rapid decline of r^2 over distance will be deemed regions of low LD, and areas with a gradual decline will be deemed regions with high LD. Values of r^2 for the unlinked markers will used to

represent the background level of LD in the populations. In addition, we will test for evidence of recombination in each population by calculating the standardized index of association (I_A) for each subpopulation (58).

Aim 4. Identify genetic and transcriptional changes associated with the gain or loss of virulence and phenotypic instability in experimentally evolved strains.

Our preliminary data showed that experimental evolution and stable adaptation to starvation and oxidative stress increases the ability of *C. neoformans* to survive and proliferate *in vitro* in macrophages, whereas prolonged growth in rich medium decreases virulence. In this aim, we will compare whole-genome sequences and gene expression patterns among 13 experimentally evolved strains and their progenitors to identify mutations, structural variations and changes in gene expression patterns associated with gain or loss of virulence. (Transcriptional changes will be determined by RNA-Seq using culture conditions described in aim 2). This aim addresses several fundamental questions about the pathobiology of *C. neoformans*. First, these data will resolve questions about whether environmental, apparently avirulent strains can become virulent after several months of incubation under conditions that simulate mammalian alveoli, and if so, whether most or only a few environmental strains have the potential to cause infection. Second, whole genome sequencing and gene expression analysis of the evolved and progenitor strains will explicate the genetic mechanisms of phenotypic variation (switching) of *C. neoformans*, which has been described in numerous reports but never fully explained at the molecular level (16, 19-22, 25). Third, direct comparison of the transcriptional patterns and whole genome sequences of the evolved and progenitor strains will identify specific genes and pathways whose differential regulation causes an increase or decrease of virulence.

Our preliminary data and results by others demonstrate that strains of *C. neoformans* are phenotypically variable and strains with the same genotypes often differ in virulence (8, 16, 19-22, 25, 38). Our recent unpublished data confirmed that *in vitro* propagation under conditions of starvation and oxidative stress resulted in the selection of strains with dramatically increased ability to replicate within macrophage-like cells (Figure 6), and growth in nutritionally replete medium selected for strains with reduced virulence (Figure 7). In this aim, we will use whole genome sequencing and RNA-Seq to identify genomic and transcriptional changes associated with phenotypic variability in *C. neoformans*. Transcriptional changes will be determined by RNA-Seq using culture conditions that imitate *in vivo* growth described in aim 2 (e.g. starvation-oxidative stress medium, J774.A1 activated cells, rabbit CSF). Overall, we anticipate obtaining whole genome sequences of 13 experimentally evolved strains and 39 RNA samples. The following strains will be analyzed:

1. Experimentally evolved progeny of non-lethal environmental strains, A1-35-8 and A5-35-17: two colony-purified evolved strains for each progenitor strain (four strains total). To obtain experimentally evolved strains for this experiment, progenitor strains were propagated at 37°C

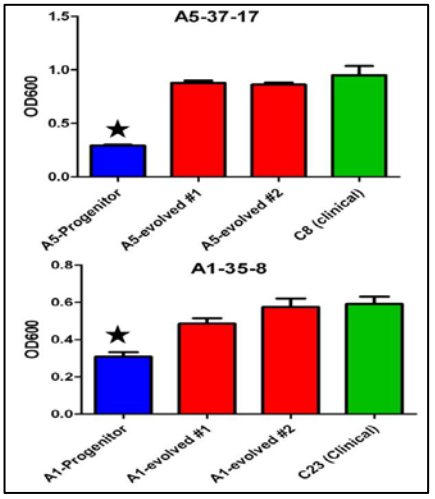

Figure 6. Survival & proliferation of the evolved & progenitor strains in macrophages. PMA-activated J774 cells were incubated with colony-purified evolved strains (red bars), their progenitors (blue bars) & lethal clinical strains with same genotypes (green bars). Total numbers of internalized yeast cells of each strain were determined by spectrophotometer after 12 h co-incubation. Progenitor strains (marked with asterisks) demonstrated significantly lower proliferation rates ($p<0.001$). Top panel shows data for A5-35-17-derived strains and the bottom panel shows data for A1-35-8 strains derived strains.

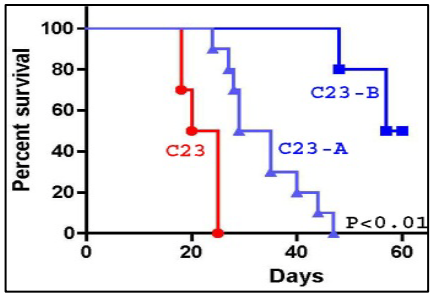

Figure 7. Survival of mice infected with the original clinical strain C23 (red) and two *in vitro* passaged subcultures (C23-A and C23-B; blue) isolated after 40 days of propagation in YPD.

in the starvation-oxidative stress medium, which contained only trace amounts of iron and copper (no external Fe or Cu were added), 0.1% of glucose (1/10 of the amount in the complete medium) and 0.1% of ammonium sulfate (1/5 of the amount in complete medium). In addition, the medium was supplied with 0.15 mM of tBOOT (tert-butylhydroperoxide), which penetrates the cells and produces reactive oxidative species (ROS) that mimic *in vivo* oxidative stress (10, 23). Strains were propagated for 60 d; each day approximately 10^6 CFU were transferred to a tube containing fresh medium. After approximately 250 generations, single colonies were isolated and sub-cultured 5 times on YPD (rich non-selective medium) plates to ensure stability of the acquired changes. Then, growth rates of the evolved and progenitor strains were compared in both limiting and rich media. As shown in Figure 8, the growth rate of evolved strains was faster in selective medium but slower in YPD. Obtaining whole genome and transcriptome sequences of the experimentally evolved strains will allow us to identify genetic and transcriptional changes that correlate with adaptation to starvation, stress and increased virulence.

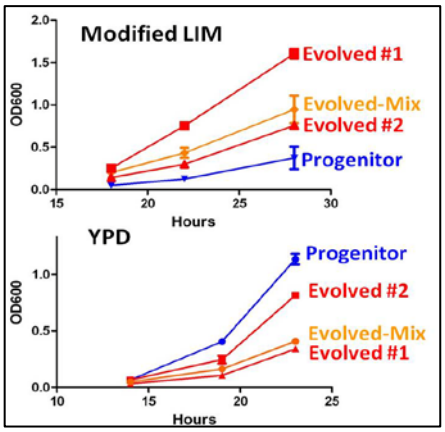

Figure 8. Growth of A5-35-17 evolved strains & their progenitor in the starvation-oxidative stress medium (modified LIM). "Progenitor" is the original environmental strain that was isolated from pigeon guano and kept at -80°C. "Evolved #1" and "Evolved #2" are two individual colonies isolated from the evolved culture and sub-cultured 5 times on YPD plates to ensure stability. "Evolved-Mix" is a sample of the entire population of evolved cells. All strains were grown o/n in YPD broth, harvested, washed with EDTA to remove Fe and Cu, enumerated in hemocytometer, and the same amount of inoculum was used to initiate the cultures. Turbidity of the cultures was estimated by measuring OD600. Experiment was performed in triplicate. Similar results were obtained for A1-38-5 strain (data not shown).

2. Phenotypic variants of strain H99 with different virulence in mice, H99E, H99W and H99S (three strains total). H99E and H99W originated from the H99 #1 original stock by repeated passaging on YPD medium; these strains differ from the parental strain as they produce less melanin, have impaired mating and attenuated virulence. H99S isolate was obtained via passage of a mixed H99 frozen stock through the rabbit model of CNS infection. The H99S isolate is highly virulent in the mouse, rabbit and *Galleria mellonella* animal models, produces normal levels of melanin and readily mates in the laboratory. Obtaining whole genome and transcriptome sequences of the phenotypic variants of H99 strain will allow us to identify genetic and transcriptional changes that correlate with these changes in virulence.
3. *pyc1Δ* revertant strains, which regained their ability to grow on glucose as the result of microevolution; two colony purified strains will be selected for each mutant (6 strains total). *PYK1*, which encodes the *C. neoformans* pyruvate kinase gene, was deleted in the wild type strain H99 generating three deletion strains, MPC31, MPC33, and MPC34, which were unable to utilize glucose. The mutants were incubated on YPD for one week and strains with revertant phenotypes were selected, which regained their ability to utilize glucose. Obtaining whole genome sequences of these revertant strains will allow us to identify genetic mechanisms that enable *C. neoformans* to adapt to different sources of carbon and/or starvation, which are major selection factors inside the host.

Proposed approach for Aim 4. Using the approach described in aim 1, we will sequence and assemble each of the six progenitor strains (approximately 100X Illumina coverage of paired end 101 base reads from each of two small-medium insert libraries). For the seven evolved progeny, we will generate 100X coverage, and identify SNPs, indels, and regions of copy number variation as in aims 1 and 3. RNA-Sequence will be performed as described in aim 2. By comparing the progeny sequence to the progenitor assemblies, we will identify genetic changes for the progeny strains. These will be compared across lineages to identify common

and unique changes. The genes or regions with changes will be compared to the RNA-Sequence data, to determine which changes resulted in gene expression differences. This will suggest candidate loci that could contribute to phenotypic changes.

COMMUNITY INTEREST

Cryptococcus is a pathogen of the global clinical importance and an established model fungus. The scientific community working on *C. neoformans* and *C. gattii* is diverse and includes numerous research laboratories in the US and abroad. The 8th International Conference on *Cryptococcus* and Cryptococcosis (8ICCC) was held in 2011 and attracted over 200 participants from more than 60 laboratories. This meeting highlighted an urgent need for additional well-annotated genomes of *C. neoformans* and *C. gattii.* Most of these investigators will benefit from the data resulting from this project. For this proposal, we have assembled a consortium of researches who support our genomic initiative, including:

Joseph Heitman, Duke University, USA

John R. Perfect, Duke University, USA

James A. Alspaugh, Duke University, USA

Thomas G. Mitchell, Duke University, USA

Xiaorong Lin, Texas A&M University, USA

Kyung J. Kwon-Chung, NIAID, USA

Arturo Casadevall, Albert Einstein College of Medicine, USA

Christina M. Hull, University of Wisconsin, Madison, USA

Kirsten Nielsen, University of Minnesota, USA

Hiten Madhani, University of California, San-Francisco, USA

Jason Stajich, University of California, Riverside, USA

James Kronstad, University of British Columbia, Canada

Jianping Xu, McMaster University, Canada

Mathew C. Fisher, Imperial College, UK

Francoise Dromer, Institute Pasteur, France

Guilhem Janbon, Institute Pasteur, France

Teun Boekhout, CBS Fungal Diversity Centre, Netherlands

PROJECT COLLABORATORS

Strains for this project have been contributed by collaborating PIs (Anastasia Litvintseva, Joseph Heitman, Francoise Dromer and John Perfect). Since the strains have already been collected and stored, no IRB or IACUC approvals are required. Furthermore, relevant strains will be provided to other researchers on request and deposited in fungal culture collections, such as ATCC, CBS, etc.

DATA RELEASE

Sequence data will be submitted as it is generated to the short read archive if available, or other appropriate database. Assemblies and annotations will be submitted to GenBank and released on the Broad Institute website.

BIBLIOGRAPHY

1. **Anders, S., and W. Huber.** Differential expression analysis for sequence count data. Genome Biol **11**:R106. http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=20979621

2. **Bovers, M., F. Hagen, E. E. Kuramae, and T. Boekhout.** 2008. Six monophyletic lineages identified within *Cryptococcus neoformans* and *Cryptococcus gattii* by multi-locus sequence typing. Fungal Genet Biol **45**:400-421. http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=18261945

3. **Byrnes, E. J., 3rd, W. Li, Y. Lewit, H. Ma, K. Voelz, P. Ren, D. A. Carter, V. Chaturvedi, R. J. Bildfell, R. C. May, and J. Heitman.** Emergence and pathogenicity of highly virulent *Cryptococcus gattii* genotypes in the northwest United States. PLoS Pathog **6**:e1000850. http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=20421942

4. **Chayakulkeeree, M., and J. R. Perfect.** 2006. Cryptococcosis. Infect Dis Clin North Am **20**:507-544, v-vi. http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=16984867

5. **Chen, F., A. J. Mackey, C. J. Stoeckert, Jr., and D. S. Roos.** 2006. OrthoMCL-DB: querying a comprehensive multi-species collection of ortholog groups. Nucleic Acids Res **34**:D363-368. http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=16381887

6. **Chen, F., A. J. Mackey, J. K. Vermunt, and D. S. Roos.** 2007. Assessing performance of orthology detection strategies applied to eukaryotic genomes. PLoS One **2**:e383. http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=17440619

7. **Chen, J., A. Varma, M. R. Diaz, A. P. Litvintseva, K. K. Wollenberg, and K. J. Kwon-Chung.** 2008. *Cryptococcus neoformans* strains and infection in apparently immunocompetent patients, China. Emerg Infect Dis **14**:755-762. http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=18439357

8. **Clancy, C. J., M. H. Nguyen, R. Alandoerffer, S. Cheng, K. Iczkowski, M. Richardson, and J. R. Graybill.** 2006. *Cryptococcus neoformans* var. *grubii* isolates recovered from persons with AIDS demonstrate a wide range of virulence during murine meningoencephalitis that correlates with the expression of certain virulence factors. Microbiology **152**:2247-2255. http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=16849791

9. **Cramer, K. L., Q. D. Gerrald, C. B. Nichols, M. S. Price, and J. A. Alspaugh.** 2006. Transcription factor Nrg1 mediates capsule formation, stress response, and pathogenesis in *Cryptococcus neoformans*. Eukaryot Cell **5**:1147-1156. http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=16835458

10. **Diezmann, S., and F. S. Dietrich.** 2009. *Saccharomyces cerevisiae*: population divergence and resistance to oxidative stress in clinical, domesticated and wild isolates. PLoS One **4**:e5317. http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=19390633

11. **Dromer, F., S. Mathoulin-Pelissier, O. Launay, and O. Lortholary.** 2007. Determinants of Disease Presentation and Outcome during Cryptococcosis: The CryptoA/D Study. PLoS Med **4**:e21. http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=17284154

12. **Falush, D., M. Stephens, and J. K. Pritchard.** 2003. Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. Genetics **164**:1567-1587. http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=12930761

13. **Fan, W., P. R. Kraus, M. J. Boily, and J. Heitman.** 2005. *Cryptococcus neoformans* gene expression during murine macrophage infection. Eukaryot Cell **4**:1420-1433. http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=16087747

14. **Findley, K., M. Rodriguez-Carres, B. Metin, J. Kroiss, A. Fonseca, R. Vilgalys, and J. Heitman.** 2009. Phylogeny and phenotypic characterization of pathogenic *Cryptococcus* species and closely related saprobic taxa in the Tremellales. Eukaryot Cell **8**:353-361.
http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=19151324
15. **Fraser, J. A., S. S. Giles, E. C. Wenink, S. G. Geunes-Boyer, J. R. Wright, S. Diezmann, A. Allen, J. E. Stajich, F. S. Dietrich, J. R. Perfect, and J. Heitman.** 2005. Same-sex mating and the origin of the Vancouver Island *Cryptococcus gattii* outbreak. Nature **437**:1360-1364.
http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=16222245
16. **Fries, B., C. Taborda, E. Serfass, and A. Casadevall.** 2001. Phenotypic switching of *Cryptococcus neoformans* occurs in vivo and influences the outcome of the infection. J. Clin. Invest. **108**:1639-1648.
http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=11733559
17. **Fromtling, R. A., G. K. Abruzzo, and A. Ruiz.** 1989. Virulence and antifungal susceptibility of environmental and clinical isolates of *Cryptococcus neoformans* from Puerto Rico. Mycopathologia **106**:163-166.
http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=2682249
18. **Fromtling, R. A., R. Blackstock, N. K. Hall, and G. S. Bulmer.** 1979. Immunization of mice with an avirulent pseudohyphal form of *Cryptococcus neoformans*. Mycopathologia **68**:179-181.
http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=394001
19. **Guerrero, A., and B. C. Fries.** 2008. Phenotypic switching in *Cryptococcus neoformans* contributes to virulence by changing the immunological host response. Infect Immun **76**:4322-4331.
http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=18591227
20. **Guerrero, A., N. Jain, D. L. Goldman, and B. C. Fries.** 2006. Phenotypic switching in *Cryptococcus neoformans*. Microbiology **152**:3-9.
http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=16385110
21. **Guerrero, A., N. Jain, X. Wang, and B. C. Fries.** *Cryptococcus neoformans* variants generated by phenotypic switching differ in virulence through effects on macrophage activation. Infect Immun **78**:1049-1057.
http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=20048044
22. **Gupta, G., and B. C. Fries.** Variability of phenotypic traits in *Cryptococcus* varieties and species and the resulting implications for pathogenesis. Future Microbiol **5**:775-787.
http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=20441549
23. **Hu, G., and J. W. Kronstad.** A putative P-type ATPase, Apt1, is involved in stress tolerance and virulence in *Cryptococcus neoformans*. Eukaryot Cell **9**:74-83.
http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=19949048
24. **Idnurm, A., Y. S. Bahn, K. Nielsen, X. Lin, J. A. Fraser, and J. Heitman.** 2005. Deciphering the model pathogenic fungus *Cryptococcus neoformans*. Nat Rev Microbiol **3**:753-764.
http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=16132036
25. **Jain, N., E. Cook, I. Xess, F. Hasan, D. Fries, and B. C. Fries.** 2009. Isolation and characterization of senescent *Cryptococcus neoformans* and implications for phenotypic switching and pathogenesis in chronic cryptococcosis. Eukaryot Cell **8**:858-866.
http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=19411622
26. **Jung, W. H., G. Hu, W. Kuo, and J. W. Kronstad.** 2009. Role of ferroxidases in iron uptake and virulence of *Cryptococcus neoformans*. Eukaryot Cell **8**:1511-1520.
http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=19700638
27. **Kavanaugh, L. A., J. A. Fraser, and F. S. Dietrich.** 2006. Recent evolution of the human pathogen *Cryptococcus neoformans* by intervarietal transfer of a 14-gene fragment. Mol Biol Evol **23**:1879-1890.
http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=16870684
28. **Kronstad, J., W. H. Jung, and G. Hu.** 2008. Beyond the big three: systematic analysis of virulence factors in *Cryptococcus neoformans*. Cell Host Microbe **4**:308-310.
http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=18854233
29. **Kwon-Chung, K. J., and A. Varma.** 2006. Do major species concepts support one, two or more species within *Cryptococcus neoformans*? FEMS Yeast Res **6**:574-587.
http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=16696653

30. **Lee, A., D. L. Toffaletti, J. Tenor, E. J. Soderblom, J. W. Thompson, M. A. Moseley, M. Price, and J. R. Perfect.** Survival defects of *Cryptococcus neoformans* mutants exposed to human cerebrospinal fluid result in attenuated virulence in an experimental model of meningitis. Infect Immun **78**:4213-4225.
http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=20696827
31. **Lewontin, R., D. Kirk, and J. Crow.** 1968. Selective mating, assortative mating, and inbreeding: definitions and implications. Eugen Q **15**:141-143.
http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=5702329
32. **Lin, X., C. M. Hull, and J. Heitman.** 2005. Sexual reproduction between partners of the same mating type in *Cryptococcus neoformans*. Nature **434**:1017-1021.
http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=15846346
33. **Lin, X., A. P. Litvintseva, K. Nielsen, S. Patel, A. Floyd, T. G. Mitchell, and J. Heitman.** 2007. alpha AD alpha hybrids of *Cryptococcus neoformans*: evidence of same-sex mating in nature and hybrid fitness. PLoS Genet **3**:1975-1990.
http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=17953489
34. **Litvintseva, A. P., I. Carbone, J. Rossouw, R. Thakur, N. P. Govender, and T. G. Mitchell.** Evidence that the Human Pathogenic Fungus *Cryptococcus neoformans* var. *grubii* May Have Evolved in Africa. PLoS One **6**:e19688.
http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=21589919
35. **Litvintseva, A. P., L. Kestenbaum, R. Vilgalys, and T. G. Mitchell.** 2005. Comparative analysis of environmental and clinical populations of *Cryptococcus neoformans*. J Clin Microbiol **43**:556-564.
http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=15695645
36. **Litvintseva, A. P., X. Lin, I. Templeton, J. Heitman, and T. G. Mitchell.** 2007. Many globally isolated AD hybrid strains of *Cryptococcus neoformans* originated in Africa. PLoS Pathog **3**:e114.
http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=17708680
37. **Litvintseva, A. P., R. E. Marra, K. Nielsen, J. Heitman, R. Vilgalys, and T. G. Mitchell.** 2003. Evidence of sexual recombination among *Cryptococcus neoformans* serotype A isolates in sub-Saharan Africa. Eukaryot Cell **2**:1162-1168.
http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=14665451
38. **Litvintseva, A. P., and T. G. Mitchell.** 2009. Most environmental isolates of *Cryptococcus neoformans* var. *grubii* (serotype A) are not lethal for mice. Infect Immun **77**:3188-3195.
http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=19487475
39. **Litvintseva, A. P., R. Thakur, L. B. Reller, and T. G. Mitchell.** 2005. Prevalence of clinical isolates of *Cryptococcus gattii* serotype C among patients with AIDS in Sub-Saharan Africa. J Infect Dis **192**:888-892.
http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=16088839
40. **Litvintseva, A. P., R. Thakur, R. Vilgalys, and T. G. Mitchell.** 2006. Multilocus sequence typing reveals three genetic subpopulations of *Cryptococcus neoformans* var. *grubii* (serotype A), including a unique population in Botswana. Genetics **172**:2223-2238.
http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=16322524
41. **Litvintseva, A. P., J. Xu, and T. G. Mitchell**. 2011. Population Structure and Ecology of *Cryptococcus neoformans* and *Cryptococcus gattii*., p. 97-112. *In* J. Heitman, T. R. Kozel, J. Kwon-Chung, J. R. Perfect, and A. Casadevall (eds.), *Cryptococcus* from Human Pathogen to Model Yeast. ASM Press Washington DC.
42. **Liu, O. W., C. D. Chun, E. D. Chow, C. Chen, H. D. Madhani, and S. M. Noble.** 2008. Systematic genetic analysis of virulence in the human fungal pathogen *Cryptococcus neoformans*. Cell **135**:174-188.
http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=18854164
43. **Loftus, B. J., E. Fung, P. Roncaglia, D. Rowley, P. Amedeo, D. Bruno, J. Vamathevan, M. Miranda, I. J. Anderson, J. A. Fraser, J. E. Allen, I. E. Bosdet, M. R. Brent, R. Chiu, T. L. Doering, M. J. Donlin, C. A. D'Souza, D. S. Fox, V. Grinberg, J. Fu, M. Fukushima, B. J. Haas, J. C. Huang, G. Janbon, S. J. Jones, H. L. Koo, M. I. Krzywinski, J. K. Kwon-Chung, K. B. Lengeler, R. Maiti, M. A. Marra, R. E. Marra, C. A. Mathewson, T. G. Mitchell, M. Pertea, F. R. Riggs, S. L. Salzberg, J. E. Schein, A. Shvartsbeyn, H. Shin, M. Shumway, C. A. Specht, B. B. Suh, A. Tenney, T. R. Utterback, B. L. Wickes, J. R. Wortman, N. H. Wye, J. W. Kronstad, J. K. Lodge, J. Heitman, R. W. Davis, C. M. Fraser, and R. W. Hyman.** 2005. The genome of the basidiomycetous yeast and human pathogen *Cryptococcus neoformans*. Science **307**:1321-1324.
http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=15653466

44. **Ma, H., J. E. Croudace, D. A. Lammas, and R. C. May.** 2006. Expulsion of live pathogenic yeast by macrophages. Curr Biol **16**:2156-2160.
http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=17084701
45. **McClelland, E. E., P. Bernhardt, and A. Casadevall.** 2005. Coping with multiple virulence factors: which is most important? PLoS Pathog **1**:e40.
http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=16738705
46. **McFadden, D. C., and A. Casadevall.** 2001. Capsule and melanin synthesis in *Cryptococcus neoformans*. Med. Mycol. **39**:19-30.
http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=11800266
47. **Meyer, W., D. M. Aanensen, T. Boekhout, M. Cogliati, M. R. Diaz, M. C. Esposto, M. Fisher, F. Gilgado, F. Hagen, S. Kaocharoen, A. P. Litvintseva, T. G. Mitchell, S. P. Simwami, L. Trilles, M. A. Viviani, and J. Kwon-Chung.** 2009. Consensus multi-locus sequence typing scheme for *Cryptococcus neoformans* and *Cryptococcus gattii*. Med Mycol **47**:561-570.
http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=19462334
48. **Mitchell, T. G., E. Castaneda, K. Nielsen, B. Wanke, and M. Lazera**. 2011. Environmental Niches for *Cryptococcus neoformans* and *Cryptococcus gattii*, p. 237-260. *In* J. Heitman, T. R. Kozel, J. Kwon-Chung, J. R. Perfect, and A. Casadevall (eds.), *Cryptococcus* from Human Pathogen to Model Yeast. ASM Press, Washington DC.
49. **Ngamskulrungroj, P., F. Gilgado, J. Faganello, A. P. Litvintseva, A. L. Leal, K. M. Tsui, T. G. Mitchell, M. H. Vainstein, and W. Meyer.** 2009. Genetic diversity of the *Cryptococcus* species complex suggests that *Cryptococcus gattii* deserves to have varieties. PLoS One **4**:e5862.
http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=19517012
50. **Nielsen, K., and J. Heitman.** 2007. Sex and virulence of human pathogenic fungi. Adv Genet **57**:143-173.
http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=17352904
51. **Nielsen, K., and J. Kwon-Chung**. 2011. A role of mating in Cryptococcal virulence. , p. 167-173. *In* J. Heitman, T. R. Kozel, J. Kwon-Chung, J. R. Perfect, and A. Casadevall (eds.), *Cryptococcus* from human pathogen to model yeast. . ASM Press Washington, DC.
52. **O'Meara, T. R., D. Norton, M. S. Price, C. Hay, M. F. Clements, C. B. Nichols, and J. A. Alspaugh.** Interaction of *Cryptococcus neoformans* Rim101 and protein kinase A regulates capsule. PLoS Pathog **6**:e1000776.
http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=20174553
53. **Park, B. J., K. A. Wannemuehler, B. J. Marston, N. Govender, P. G. Pappas, and T. M. Chiller.** 2009. Estimation of the current global burden of cryptococcal meningitis among persons living with HIV/AIDS. AIDS **23**:525-530.
http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=19182676
54. **Perfect, J. R.** 2005. *Cryptococcus neoformans*: a sugar-coated killer with designer genes. FEMS Immunol Med Microbiol **45**:395-404.
http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=16055314
55. **Perfect, J. R.** 2007. Management of Cryptococcosis: How Are We Doing? PLoS Med **4**:e47.
http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=17284156
56. **Pritchard, J. K., M. Stephens, and P. Donnelly.** 2000. Inference of population structure using multilocus genotype data. Genetics **155**:945-959.
http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=10835412
57. **Sharpton, T. J., J. E. Stajich, S. D. Rounsley, M. J. Gardner, J. R. Wortman, V. S. Jordar, R. Maiti, C. D. Kodira, D. E. Neafsey, Q. Zeng, C. Y. Hung, C. McMahan, A. Muszewska, M. Grynberg, M. A. Mandel, E. M. Kellner, B. M. Barker, J. N. Galgiani, M. J. Orbach, T. N. Kirkland, G. T. Cole, M. R. Henn, B. W. Birren, and J. W. Taylor.** 2009. Comparative genomic analyses of the human fungal pathogens Coccidioides and their relatives. Genome Res **19**:1722-1731.
http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=19717792
58. **Smith, J. M., N. H. Smith, M. O'Rourke, and B. G. Spratt.** 1993. How clonal are bacteria? Proc Natl Acad Sci U S A **90**:4384-4388.
http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=8506277

59. **Stamatakis, A.** 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics **22**:2688-2690.
http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=16928733
60. **Stamatakis, A., T. Ludwig, and H. Meier.** 2005. RAxML-III: a fast program for maximum likelihood-based inference of large phylogenetic trees. Bioinformatics **21**:456-463.
http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=15608047
61. **Sun, S., and J. Xu.** 2007. Genetic analyses of a hybrid cross between serotypes A and D strains of the human pathogenic fungus *Cryptococcus neoformans*. Genetics **177**:1475-1486.
http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=17947421
62. **Vallim, M. A., C. B. Nichols, L. Fernandes, K. L. Cramer, and J. A. Alspaugh.** 2005. A Rac homolog functions downstream of Ras1 to control hyphal differentiation and high-temperature growth in the pathogenic fungus *Cryptococcus neoformans*. Eukaryot Cell **4**:1066-1078.
http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=15947199
63. **VanLiere, J. M., and N. A. Rosenberg.** 2008. Mathematical properties of the r2 measure of linkage disequilibrium. Theor Popul Biol **74**:130-137.
http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=18572214
64. **Viviani, M. A., M. C. Esposto, M. Cogliati, M. T. Montagna, and B. L. Wickes.** 2001. Isolation of a *Cryptococcus neoformans* serotype A *MATa* strain from the Italian environment. Med. Mycol. **39**:383-386.
http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=12054047
65. **Yang, Z.** 2007. PAML 4: phylogenetic analysis by maximum likelihood. Mol Biol Evol **24**:1586-1591.
http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=17483113

Table 2. Strains proposed for sequencing. Strains are grouped into four specific aims. Strains with identical MLST genotypes are color-coded: one strain shown in red from each clonal group will be *de novo* assembled as a reference genome.

AIM 1							
Strain	Serotype	MLST genotype	Country of origin	Source	Comments	Provided by PI	Reference
A1-35-8	A	**A1/M1**	USA	pigeon guano	avirulent environmental strain	Litvintseva	38, 40
A1-84-14	A	A1/M1	USA	pigeon guano	the only virulent environmental strain	Litvintseva	38, 40
Arg1366	A	A1/M1	Argentina	clinical	2006, MLST study	Litvintseva	40
C23	A	A1/M1	USA	bronch. wash/HIV-	virulent clinical strain	Litvintseva	38, 40
AD2-99a	A	A1/M1	France	clinical	A/D study in France	Dromer	11
AD3-83a	A	A1/M1	France	clinical	A/D study in France	Dromer	11
MW-RSA606	A	A1/M1	S. Africa	CSF/HIV+	RSA, 2010, original isolate	Litvintseva	
MW-RSA5387	A	A1/M1	S. Africa	CSF/HIV+	RSA, 2010, relapse isolate	Litvintseva	
A3-38-20	A	A3/M1	USA	pigeon guano	avirulent environmental strain	Litvintseva	38, 40
AD4-47a	A	A4/M1	France	clinical	A/D study in France	Dromer	11
AD3-58	A	A4/M1	France	clinical	A/D study in France	Dromer	11
A4-1-12	A	A4/M1	USA	pigeon guano	2005, environmental strain from NC	Litvintseva	
MW-RSA1869	A	A4/M1	S. Africa	CSF/HIV+	2010, RSA, van Wyk study	Litvintseva	
MW-RSA82	A	related to A1/M1	S. Africa	CSF/HIV+	2010, RSA, van Wyk study	Litvintseva	
AD1-7a	A	related to A1/M1	France	clinical	A/D study, the most prevalent gen	Dromer	11
AD1-95a	A	related to A1/M1	France	clinical	A/D study, the most prevalent gen	Dromer	11
AD5-45a	A	related to A1/M1	France	clinical	A/D study, the most prevalent gen	Dromer	11
AD6-55a	A	related to A1/M1	France	clinical	A/D study , the most prevalent gen	Dromer	11
AD6-54a	A	related to A1/M1	France	clinical	A/D study, the most prevalent gen	Dromer	11
AD1-90a	A	related to A1/M1	France	clinical	A/D study, the most prevalent gen	Dromer	11
WM148	A	A1/M1	Australia	clinical	typing strain	Litvintseva	40, 47
MW-RSA2628	A	**unique**	S. Africa	CSF/HIV+	RSA, 2010, original isolate	Litvintseva	
MW-RSA6134	A	unique	S. Africa	CSF/HIV+	RSA, 2010, relapse isolate	Litvintseva	
A2-102-5	A	**A2/M2**	USA	pigeon guano	genotype never isolated from humans	Litvintseva	38, 40
D17-1	A	A2/M2	S. Africa	pigeon guano	Durban, 2007	Litvintseva	34
A3-1-1	A	**A3/M3**	USA	pigeon guano	avirulent environmental strain	Litvintseva	40
Pr68	A	A3/M3	S. Africa	pigeon guano	Parys, 2007	Litvintseva	34
Ug2459	A	A3/M3	Uganda	CSF/HIV+	2006, MLST study	Litvintseva	40
AD2-60a	A	A3/M3	France	clinical	A/D study in France	Dromer	11
AD4-92a	A	A3/M3	France	clinical	A/D study in France	Dromer	11

AIM 1, Continued							
Strain	Serotype	MLST genotype	Country of origin	Source	Comments	Provided by PI	Reference
A4-34-6	A	**A4/M4**	USA	pigeon guano	avirulent environmental strain	Litvintseva	38, 40
In2632	A	A4/M4	India	CSF/HIV+	2006, MLST study	Litvintseva	40
C27	A	A4/M4	USA	CSF/cancer	virulent clinical strain	Litvintseva	38, 40
Br795	A	A4/M4	Brazil	clinical	2006, MLST study	Litvintseva	40
RTC1	A	A4/M4	Botswana	CSF/HIV+	2007, Botswana clinical	Litvintseva	34
Gb118	A	A4/M4	Botswana	pigeon guano	2007, Gaborone	Litvintseva	34
MW-RSA36	A	A4/M4	S. Africa	CSF/HIV+	2010, RSA, van Wyk study	Litvintseva	
AD1-68a	A	A4/M4	France	clinical	A/D study in France	Dromer	11
AD5-39a	A	A4/M4	France	clinical	A/D study in France	Dromer	11
AD2-04a	A	A4/M4	France	clinical	A/D study in France	Dromer	11
AD4-63a	A	A4/M4	France	clinical	A/D study in France	Dromer	11
A5-35-17	A	**A5/M5**	USA	pigeon guano	avirulent environmental strain	Litvintseva	38, 40
C8	A	A5/M5	USA	CSF/HIV+	progenitor of exp. evolved strains	Litvintseva	38, 40
Jp1088	A	A5/M5	Japan	human lung	2006, MLST study	Litvintseva	40
CHC193	A	A5/M5	China	CSF/HIV-	the only genotype in China	Litvintseva	7
MW-RSA1955	A	A5/M5	S. Africa	CSF/HIV+	2010, RSA, van Wyk study	Litvintseva	
AD1-83a	A	A5/M5	France	clinical	A/D study in France	Dromer	11
AD1-86a	A	A5/M5	France	clinical	A/D study in France	Dromer	11
AD2-82a	A	A5/M5	France	clinical	A/D study in France	Dromer	11
AD5-53a	A	A5/M5	France	clinical	A/D study in France	Dromer	11
Jo278-1	A	A5/M5	S. Africa	soil	Johannesburg, 2007	Litvintseva	34
Th206	A	**Th**	Thailand	blood/HIV+	2006, MLST study	Litvintseva	40
Ug2463	A	Th	Uganda	CSF/HIV+	2006, MLST study	Litvintseva	40
Th84	A	Th	Thailand	blood/HIV+	2006, MLST study	Litvintseva	40
LP-RSA2297	A	Th	S. Africa	CSF/HIV+	2006, Limpopo Pr.	Litvintseva	34
AD3-95a	A	Th	France	clinical	A/D study in France	Dromer	11
AD4-76a	A	Th	France	clinical	A/D study in France	Dromer	11
AD2-6a	A	**unique**	France	clinical	A/D study in France, unusual genotype	Dromer	11
AD3-55a	A	unique	France	clinical	A/D study in France, unusual genotype	Dromer	11
AD3-41a	A	unique	France	clinical	A/D study in France, unusual genotype	Dromer	11
A7-35-23	A	**VNII**	USA	soil	avirulent environmental strain	Litvintseva	38, 40
C12	A	VNII	USA	lung/HIV-	avirulent clinical strain	Litvintseva	38, 40

AIM 1, Continued							
Strain	Serotype	MLST genotype	Country of origin	Source	Comments	Provided by PI	Reference
Ug2462	A	VNII	Uganda	CSF/HIV+	2006, MLST study	Litvintseva	40
WM626	A	VNII	Australia	clinical	VNII typing strain	Litvintseva	40, 47
C45	A	VNII	USA	sputum/HIV-	virulent clinical strain	Litvintseva	38, 40
LP-RSA1684	A	VNII	S. Africa	CSF/HIV+	virulent clinical strain	Litvintseva	34
AD3-9a	A	VNII	France	clinical	A/D study France	Dromer	11
AD5-67a	A	VNII	France	clinical	A/D study France	Dromer	11
T4	A	VNII	USA	oak tree	arboreal isolate from the US	Litvintseva	
8-1	A	VNII	USA	clinical	used for creating congenic strains	Litvintseva	34, 40
AD3-11a	A	**VNII**	France	clinical	A/D study France	Dromer	11
C2	A	VNII	USA	clinical	2006, MLST study	Litvintseva	40
MW-RSA852	A	VNII	S. Africa	CSF/HIV+	RSA, 2010, original isolate	Litvintseva	
MW-RSA4119	A	VNII	S. Africa	CSF/HIV+	RSA, 2010, relapse isolate	Litvintseva	
MW-RSA722	A	**unique**	S. Africa	CSF/HIV+	most prevalent genotype in van Wyk study	Litvintseva	
LP-RSA2296	A	**unique**	S. Africa	CSF/HIV+	related to A2/M2, Limpopo Pr., 2006	Litvintseva	34
CDC15	A	**unknown**	USA	clinical	strain used for QTL mapping by Sun et al	Litvintseva	61
IUM 96-28-28	A	**unknown**	Italy	pigeon guano	MATa isolate from Italy	Litvintseva	64
125.91	A	**unique**	Tanzania	clinical	MATa isolate, ancestral haplotypes	Litvintseva	34, 36
MMRL757	D	**lineage 1**	Italy			Heitman	
MMRL760	D	**lineage 1**	Italy			Heitman	
MMRL751	D	**lineage 1**	Italy			Heitman	
B3179	D	**lineage 1**	USA			Heitman	
CDC92_18	D	**lineage 2**	USA			Heitman	
S17	D	**lineage 2**	France			Heitman	
2_22	D	**lineage 2**	USA	pigeon guano		Heitman	35
CDC92_27	D	**lineage 2**	USA	clinical		Heitman	
AD3_99	D	**lineage 2**	France			Heitman	11
VANC_R461	D	**lineage 2**	Canada			Heitman	
431	D	**lineage 2**	Denmark			Heitman	
3_15	D	**lineage 2**	USA	pigeon guano		Heitman	35
S15	D	**lineage 2**	France	clinical		Heitman	
S10	D	**lineage 2**	France	clinical		Heitman	
WM629	D	**unknown**	Australia	clinical	typing strain for serotype D	Heitman	47

AIM 1, Continued							
Strain	Serotype	MLST genotype	Country of origin	Source	Comments	Provided by PI	Reference
IUM92_4686	AD	**lineage 1**	Italy			Heitman	
ATCC48184	AD	**lineage 1**	Japan			Heitman	
S3	AD	**lineage 1**	France			Heitman	
NC34_21	AD	**lineage 1**	USA	pigeon guano		Heitman	35
6_20	AD	**lineage 2**	USA	pigeon guano		Heitman	35
AD7_95	AD	**lineage 2**	France			Heitman	11
MMRL1361	AD	**lineage 2**	Italy			Heitman	38
S4	AD	**lineage 2**	France			Heitman	
It756	AD	**lineage 2**	Italy			Heitman	38
S6	AD	**lineage 2**	France			Heitman	
S7	AD	**lineage 2**	France			Heitman	
S10	AD	**lineage 2**	France			Heitman	
Zg287	AD	**lineage 2**	China			Heitman	38
S1	AD	**lineage 3**	France			Heitman	
E566	B	**VGI**	Australia	eucalyptus tree		Heitman	
EJB2	B	**VGI**	USA	clinical		Heitman	
NT-10	B	**VGI**	Australia	clinical		Heitman	
Ru294	B	**VGI**	S. Africa	unknown tree		Heitman	
CA1014	B	**VGIIa**	USA	clinical		Heitman	
CBS10090	B	**VGIIa**	Greece	clinical		Heitman	
LA55	B	**VGIIa**	Brazil	clinical		Heitman	
99/473	B	**VGIIb**	Caribbean	clinical		Heitman	
Ram5	B	**VGIIb**	Australia	clinical		Heitman	
2001/936-1	B	**VGII**	Senegal	clinical		Heitman	
MMRL2647	C	**VGII**	Australia	clinical		Heitman	
CA1873	C	**VGIII**	USA	clinical		Heitman	
CA1280	C	**VGIII**	USA	clinical		Heitman	
Bt64	C	**VGIV**	Botswana	clinical		Heitman	39
IND107	C	**VGIV**	India	clinical		Heitman	
CBS7841		***Filobasidiella depauperata***				Heitman	14
CBS7855		***Filobasidiella depauperata***				Heitman	14
CBS6074		***Bullera dendrophila***				Heitman	14

AIM 1, Continued							
Strain	Serotype	MLST genotype	Country of origin	Source	Comments	Provided by PI	Reference
CBS8507		***Kwoniella mangroviensis***				Heitman	14
CBS10436		***Kwoniella mangroviensis***				Heitman	14
CBS8886		***Kwoniella mangroviensis***				Heitman	14
CBS10118		***Cryptococcus bestiolae***				Heitman	14
CBS10117		***Cryptococcus dejecticola***				Heitman	14
CBS569		***Cryptococcus heveanensis***				Heitman	14
BCC8398		***Cryptococcus heveanensis***				Heitman	14
CBS6973		***Tremella mesenterica***				Heitman	14
ATCC24925		***Tremella mesenterica***				Heitman	14
HL#3846		***Tremella globispora***				Heitman	14
CBS10737		***Cryptococcus pinus***				Heitman	14
KMF1		***Cryptococcus shivajii***				Heitman	14
ATCC14438		***Cryptococcus humicola***				Heitman	14
CBS6039		***Cryptococcus amylolentus***				Heitman	14
CBS6273		***Cryptococcus amylolentus***				Heitman	14
AIM 2							
A1-35-8	A	A1/M1	USA	pigeon guano	avirulent environmental strain	Perfect/Litvintseva*	38, 40
A1-84-14	A	A1/M1	USA	pigeon guano	the only virulent environmental strain	Perfect/Litvintseva	38, 40
C23	A	A1/M1	USA	clinical	virulent clinical strain	Perfect/Litvintseva	38, 40
H99	A	A1/M1	USA	clinical	typing strain	Perfect/Litvintseva	38, 40
A5-35-17	A	A5/M5	USA	pigeon guano	avirulent environmental strain	Perfect/Litvintseva	38, 40
C8	A	A5/M5	USA	clinical	clinical virulent strain	Perfect/Litvintseva	38, 40
A4-34-6	A	A4/M4	USA	pigeon guano	avirulent environmental strain	Perfect/Litvintseva	38, 40
C27	A	A4/M4	USA	clinical	clinical virulent strain	Perfect/Litvintseva	38, 40
2_22	D	lineage 2	USA	pigeon guano	environmental strain	Perfect/Litvintseva	35
S16	D	lineage 2	France	clinical	clinical strain	Perfect/Litvintseva	
R265	B	VGIIa	Canada	clinical	outbreak typing strain	Perfect/Litvintseva	3
RB1	B	VGIIa	Canada	tree	environmental, genotype identical to R265	Perfect/Litvintseva	

* RNA for Aim 2 will be provided by Drs. Litvintseva and Perfect

AIM 3							
Strain	Serotype	MLST genotype	Country of origin	Source	Comments	Provided by PI	Reference
Bt2	A	**gen24 (VNI)**	Botswana	CSF /HIV+	Botswana, 2003	Litvintseva	37
Bt3	A	gen24 (VNI)	Botswana	CSF /HIV+	Botswana, 2003	Litvintseva	37
Bt6	A	gen24 (VNI)	Botswana	CSF /HIV+	Botswana, 2003	Litvintseva	37
Bt19	A	gen24 (VNI)	Botswana	CSF /HIV+	Botswana, 2003	Litvintseva	37
Bt29	A	gen24 (VNI)	Botswana	CSF /HIV+	Botswana, 2003	Litvintseva	37
Bt30	A	gen24 (VNI)	Botswana	CSF /HIV+	Botswana, 2003	Litvintseva	37
Bt37	A	gen24 (VNI)	Botswana	CSF /HIV+	Botswana, 2003	Litvintseva	37
Bt44	A	gen24 (VNI)	Botswana	CSF /HIV+	Botswana, 2003	Litvintseva	37
Bt93	A	gen24 (VNI)	Botswana	CSF /HIV+	Botswana, 2003	Litvintseva	37
Bt110	A	gen24 (VNI)	Botswana	CSF /HIV+	Botswana, 2003	Litvintseva	37
Bt114	A	gen24 (VNI)	Botswana	CSF /HIV+	Botswana, 2003	Litvintseva	37
Bt116	A	gen24 (VNI)	Botswana	CSF /HIV+	Botswana, 2003	Litvintseva	37
Bt120	A	gen24 (VNI)	Botswana	CSF /HIV+	Botswana, 2003	Litvintseva	37
Bt129	A	gen24 (VNI)	Botswana	CSF /HIV+	Botswana, 2003	Litvintseva	37
Bt130	A	gen24 (VNI)	Botswana	CSF /HIV+	Botswana, 2003	Litvintseva	37, 40
Bt137	A	gen24 (VNI)	Botswana	CSF /HIV+	Botswana, 2003	Litvintseva	37
Bt151	A	gen24 (VNI)	Botswana	CSF /HIV+	Botswana, 2003	Litvintseva	37
Bt153	A	gen24 (VNI)	Botswana	CSF /HIV+	Botswana, 2003	Litvintseva	37
Bt161	A	gen24 (VNI)	Botswana	CSF /HIV+	Botswana, 2003	Litvintseva	37
Bt203	A	gen24 (VNI)	Botswana	CSF /HIV+	Botswana, 2003	Litvintseva	37
Bt211	A	gen24 (VNI)	Botswana	CSF /HIV+	Botswana, 2003	Litvintseva	37
Bt214	A	gen24 (VNI)	Botswana	CSF /HIV+	Botswana, 2003	Litvintseva	37
Bt4	A	**gen27 (VNI)**	Botswana	CSF /HIV+	Botswana, 2003	Litvintseva	37
Bt8	A	gen27-30 (VNI)	Botswana	CSF /HIV+	Botswana, 2003	Litvintseva	37
Bt11	A	gen27-30 (VNI)	Botswana	CSF /HIV+	Botswana, 2003	Litvintseva	37
Bt15	A	gen27-30 (VNI)	Botswana	CSF /HIV+	Botswana, 2003	Litvintseva	37, 40
Bt17	A	gen27-30 (VNI)	Botswana	CSF /HIV+	Botswana, 2003	Litvintseva	37
Bt18	A	gen27-30 (VNI)	Botswana	CSF /HIV+	Botswana, 2003	Litvintseva	37
Bt20	A	gen27-30 (VNI)	Botswana	CSF /HIV+	Botswana, 2003	Litvintseva	37
Bt23	A	gen27-30 (VNI)	Botswana	CSF /HIV+	Botswana, 2003	Litvintseva	37
Bt28	A	gen27-30 (VNI)	Botswana	CSF /HIV+	Botswana, 2003	Litvintseva	37
Bt42	A	gen27-30 (VNI)	Botswana	CSF /HIV+	Botswana, 2003	Litvintseva	37

AIM 3, Continued							
Strain	Serotype	MLST genotype	Country of origin	Source	Comments	Provided by PI	Reference
Bt43	A	gen27-30 (VNI)	Botswana	CSF /HIV+	Botswana, 2003	Litvintseva	37
Bt47	A	gen27-30 (VNI)	Botswana	CSF /HIV+	Botswana, 2003	Litvintseva	37
Bt48	A	gen27-30 (VNI)	Botswana	CSF /HIV+	Botswana, 2003	Litvintseva	37
Bt49	A	gen27-30 (VNI)	Botswana	CSF /HIV+	Botswana, 2003	Litvintseva	37
Bt51	A	gen27-30 (VNI)	Botswana	CSF /HIV+	Botswana, 2003	Litvintseva	37
Bt57	A	gen27-30 (VNI)	Botswana	CSF /HIV+	Botswana, 2003	Litvintseva	37
Bt58	A	gen27-30 (VNI)	Botswana	CSF /HIV+	Botswana, 2003	Litvintseva	37
Bt59	A	gen27-30 (VNI)	Botswana	CSF /HIV+	Botswana, 2003	Litvintseva	37
Bt62	A	gen27-30 (VNI)	Botswana	CSF /HIV+	Botswana, 2003	Litvintseva	37
Bt66	A	gen27-30 (VNI)	Botswana	CSF /HIV+	Botswana, 2003	Litvintseva	37
Bt72	A	gen27-30 (VNI)	Botswana	CSF /HIV+	Botswana, 2003	Litvintseva	37
Bt77	A	gen27-30 (VNI)	Botswana	CSF /HIV+	Botswana, 2003	Litvintseva	37
Bt78	A	gen27-30 (VNI)	Botswana	CSF /HIV+	Botswana, 2003	Litvintseva	37
Bt79	A	gen27-30 (VNI)	Botswana	CSF /HIV+	Botswana, 2003	Litvintseva	37
Bt83	A	gen27-30 (VNI)	Botswana	CSF /HIV+	Botswana, 2003	Litvintseva	37
Bt87	A	gen27-30 (VNI)	Botswana	CSF /HIV+	Botswana, 2003	Litvintseva	37
Bt121	A	gen27-30 (VNI)	Botswana	CSF /HIV+	Botswana, 2003	Litvintseva	37, 40
Bt155	A	gen27-30 (VNI)	Botswana	CSF /HIV+	Botswana, 2003	Litvintseva	37
Bt209	A	gen27-30 (VNI)	Botswana	CSF /HIV+	Botswana, 2003	Litvintseva	37
Bt210	A	gen27-30 (VNI)	Botswana	CSF /HIV+	Botswana, 2003	Litvintseva	37
Bt205	A	gen27-30 (VNI)	Botswana	CSF /HIV+	Botswana, 2003	Litvintseva	37
Bt207	A	gen27-30 (VNI)	Botswana	CSF /HIV+	Botswana, 2003	Litvintseva	37
Bt90	A	gen27-30 (VNI)	Botswana	CSF /HIV+	Botswana, 2003	Litvintseva	37
Bt95	A	gen27-30 (VNI)	Botswana	CSF /HIV+	Botswana, 2003	Litvintseva	37
Bt97	A	gen27-30 (VNI)	Botswana	CSF /HIV+	Botswana, 2003	Litvintseva	37
Bt104	A	gen27-30 (VNI)	Botswana	CSF /HIV+	Botswana, 2003	Litvintseva	37
Bt106	A	gen27-30 (VNI)	Botswana	CSF /HIV+	Botswana, 2003	Litvintseva	37
Bt111	A	gen27-30 (VNI)	Botswana	CSF /HIV+	Botswana, 2003	Litvintseva	37
Bt117	A	gen27-30 (VNI)	Botswana	CSF /HIV+	Botswana, 2003	Litvintseva	37
Bt118	A	gen27-30 (VNI)	Botswana	CSF /HIV+	Botswana, 2003	Litvintseva	37
Bt123	A	gen27-30 (VNI)	Botswana	CSF /HIV+	Botswana, 2003	Litvintseva	37
Bt126	A	gen27-30 (VNI)	Botswana	CSF /HIV+	Botswana, 2003	Litvintseva	37

AIM 3, Continued							
Strain	Serotype	MLST genotype	Country of origin	Source	Comments	Provided by PI	Reference
Bt132	A	gen27-30 (VNI)	Botswana	CSF /HIV+	Botswana, 2003	Litvintseva	37
Bt160	A	gen27-30 (VNI)	Botswana	CSF /HIV+	Botswana, 2003	Litvintseva	37
Bt164	A	gen27-30 (VNI)	Botswana	CSF /HIV+	Botswana, 2003	Litvintseva	37
Bt202	A	gen27-30 (VNI)	Botswana	CSF /HIV+	Botswana, 2003	Litvintseva	37
Bt212	A	gen27-30 (VNI)	Botswana	CSF /HIV+	Botswana, 2003	Litvintseva	37
RTC5	A	gen27-30 (VNI)	Botswana	CSF /HIV+	Botswana, 2003	Litvintseva	34
Bt100	A	**gen28 (VNI)**	Botswana	CSF /HIV+	Botswana, 2003	Litvintseva	37, 40
Bt107	A	gen28 (VNI)	Botswana	CSF /HIV+	Botswana, 2003	Litvintseva	37
Bt108	A	gen28 (VNI)	Botswana	CSF /HIV+	Botswana, 2003	Litvintseva	37
Bt119	A	gen28 (VNI)	Botswana	CSF /HIV+	Botswana, 2003	Litvintseva	37
Bt127	A	gen28 (VNI)	Botswana	CSF /HIV+	Botswana, 2003	Litvintseva	37
Bt134	A	gen32 (A5/M5)	Botswana	CSF /HIV+	Botswana, 2003	Litvintseva	37, 40
Bt139	A	gen32 (A5/M5)	Botswana	CSF /HIV+	Botswana, 2003	Litvintseva	37
Bt141	A	gen32 (A5/M5)	Botswana	CSF /HIV+	Botswana, 2003	Litvintseva	37
Bt150	A	**gen30 (VNI)**	Botswana	CSF /HIV+	Botswana, 2003	Litvintseva	37, 40
LP-RSA848	A	gen30 (VNI)	Botswana	CSF /HIV+	Botswana, 2003	Litvintseva	34
Bt156	A	gen30 (VNI)	Botswana	CSF /HIV+	Botswana, 2003	Litvintseva	37
Bt163	A	gen30 (VNI)	Botswana	CSF /HIV+	Botswana, 2003	Litvintseva	37
Bt213	A	gen30 (VNI)	Botswana	CSF /HIV+	Botswana, 2003	Litvintseva	37
LP-RSA116	A	**VNI**	Botswana	CSF /HIV+	Botswana, 2003	Litvintseva	34
Bt9	A	**VNI**	Botswana	CSF /HIV+	Botswana, 2003	Litvintseva	37, 40
Bt68	A	**VNI**	Botswana	CSF /HIV+	Botswana, 2003	Litvintseva	37, 40
LP-RSA1343	A	**VNI**	Botswana	CSF /HIV+	Botswana, 2003	Litvintseva	34
LP-RSA3042	A	**VNI**	Botswana	CSF /HIV+	Botswana, 2003	Litvintseva	34
Gb159-1	A	**VNI**	Botswana	unknown tree	Gaborone, 2007	Litvintseva	34
Tu241-1	A	**VNI**	Botswana	Mopane tree	Tuli, 2007	Litvintseva	34
Tu259-1	A	**VNI**	Botswana	Mopane tree	Tuli, 2007	Litvintseva	34
Pr284-1	A	**VNI**	Botswana	avian guano	Parys, 2007	Litvintseva	34
Bt1	A	**VNB**	Botswana	CSF /HIV+	Botswana, 2003	Litvintseva	37, 40
Bt24	A	**VNB**	Botswana	CSF /HIV+	Botswana, 2003	Litvintseva	37, 40
Bt25	A	**VNB**	Botswana	CSF /HIV+	Botswana, 2003	Litvintseva	37
Bt27	A	**VNB**	Botswana	CSF /HIV+	Botswana, 2003	Litvintseva	37

AIM 3, Continued							
Strain	Serotype	MLST genotype	Country of origin	Source	Comments	Provided by PI	Reference
Bt22	A	**VNB**	Botswana	CSF /HIV+	Botswana, 2003	Litvintseva	37
Bt5	A	**VNB**	Botswana	CSF /HIV+	Botswana, 2003	Litvintseva	37
Bt10	A	**VNB**	Botswana	CSF /HIV+	Botswana, 2003	Litvintseva	37
Bt7	A	**VNB**	Botswana	CSF /HIV+	Botswana, 2003	Litvintseva	37
Bt31	A	**VNB**	Botswana	CSF /HIV+	Botswana, 2003	Litvintseva	37
Bt32	A	**VNB**	Botswana	CSF /HIV+	Botswana, 2003	Litvintseva	37
Bt33	A	**VNB**	Botswana	CSF /HIV+	Botswana, 2003	Litvintseva	37
Bt34	A	**VNB**	Botswana	CSF /HIV+	Botswana, 2003	Litvintseva	37, 40
Bt36	A	**VNB**	Botswana	CSF /HIV+	Botswana, 2003	Litvintseva	37
Bt38	A	**VNB**	Botswana	CSF /HIV+	Botswana, 2003	Litvintseva	37
Bt39	A	**VNB**	Botswana	CSF /HIV+	Botswana, 2003	Litvintseva	37
Bt40	A	**VNB**	Botswana	CSF /HIV+	Botswana, 2003	Litvintseva	37
Bt41	A	**VNB**	Botswana	CSF /HIV+	Botswana, 2003	Litvintseva	37
Bt45	A	**VNB**	Botswana	CSF /HIV+	Botswana, 2003	Litvintseva	37
Bt46	A	**VNB**	Botswana	CSF /HIV+	Botswana, 2003	Litvintseva	37
Bt50	A	**VNB**	Botswana	CSF /HIV+	Botswana, 2003	Litvintseva	37
Bt52	A	**VNB**	Botswana	CSF /HIV+	Botswana, 2003	Litvintseva	37
Bt53	A	**VNB**	Botswana	CSF /HIV+	Botswana, 2003	Litvintseva	37
Bt54	A	**VNB**	Botswana	CSF /HIV+	Botswana, 2003	Litvintseva	37
Bt55	A	**VNB**	Botswana	CSF /HIV+	Botswana, 2003	Litvintseva	37
Bt56	A	**VNB**	Botswana	CSF /HIV+	Botswana, 2003	Litvintseva	37
Bt60	A	**VNB**	Botswana	CSF /HIV+	Botswana, 2003	Litvintseva	37, 40
Bt63	A	**VNB**	Botswana	CSF /HIV+	Botswana, 2003	Litvintseva	37, 40
Bt65	A	**VNB**	Botswana	CSF /HIV+	Botswana, 2003	Litvintseva	37
Bt70	A	**VNB**	Botswana	CSF /HIV+	Botswana, 2003	Litvintseva	37
Bt71	A	**VNB**	Botswana	CSF /HIV+	Botswana, 2003	Litvintseva	37
Bt75	A	**VNB**	Botswana	CSF /HIV+	Botswana, 2003	Litvintseva	37
Bt76	A	**VNB**	Botswana	CSF /HIV+	Botswana, 2003	Litvintseva	37, 40
Bt81	A	**VNB**	Botswana	CSF /HIV+	Botswana, 2003	Litvintseva	37
Bt84	A	**VNB**	Botswana	CSF /HIV+	Botswana, 2003	Litvintseva	37
Bt85	A	**VNB**	Botswana	CSF /HIV+	Botswana, 2003	Litvintseva	37
Bt88	A	**VNB**	Botswana	CSF /HIV+	Botswana, 2003	Litvintseva	37, 40

AIM 3, Continued							
Strain	Serotype	MLST genotype	Country of origin	Source	Comments	Provided by PI	Reference
Bt89	A	**VNB**	Botswana	CSF /HIV+	Botswana, 2003	Litvintseva	37
Bt92	A	**VNB**	Botswana	CSF /HIV+	Botswana, 2003	Litvintseva	37
Bt96	A	**VNB**	Botswana	CSF /HIV+	Botswana, 2003	Litvintseva	37
Bt98	A	**VNB**	Botswana	CSF /HIV+	Botswana, 2003	Litvintseva	37
Bt99	A	**VNB**	Botswana	CSF /HIV+	Botswana, 2003	Litvintseva	37
Bt101	A	**VNB**	Botswana	CSF /HIV+	Botswana, 2003	Litvintseva	37
Bt102	A	**VNB**	Botswana	CSF /HIV+	Botswana, 2003	Litvintseva	37
Bt103	A	**VNB**	Botswana	CSF /HIV+	Botswana, 2003	Litvintseva	37
Bt105	A	**VNB**	Botswana	CSF /HIV+	Botswana, 2003	Litvintseva	37
Bt109	A	**VNB**	Botswana	CSF /HIV+	Botswana, 2003	Litvintseva	37
Bt115	A	**VNB**	Botswana	CSF /HIV+	Botswana, 2003	Litvintseva	37
Bt125	A	**VNB**	Botswana	CSF /HIV+	Botswana, 2003	Litvintseva	37, 40
Bt128	A	**VNB**	Botswana	CSF /HIV+	Botswana, 2003	Litvintseva	37
Bt131	A	**VNB**	Botswana	CSF /HIV+	Botswana, 2003	Litvintseva	37, 40
Bt133	A	**VNB**	Botswana	CSF /HIV+	Botswana, 2003	Litvintseva	37
Bt136	A	**VNB**	Botswana	CSF /HIV+	Botswana, 2003	Litvintseva	37
Bt137	A	**VNB**	Botswana	CSF /HIV+	Botswana, 2003	Litvintseva	37
Bt138	A	**VNB**	Botswana	CSF /HIV+	Botswana, 2003	Litvintseva	37
Bt142	A	**VNB**	Botswana	CSF /HIV+	Botswana, 2003	Litvintseva	37
Bt143	A	**VNB**	Botswana	CSF /HIV+	Botswana, 2003	Litvintseva	37
Bt146	A	**VNB**	Botswana	CSF /HIV+	Botswana, 2003	Litvintseva	37
Bt147	A	**VNB**	Botswana	CSF /HIV+	Botswana, 2003	Litvintseva	37
Bt148	A	**VNB**	Botswana	CSF /HIV+	Botswana, 2003	Litvintseva	37
Bt149	A	**VNB**	Botswana	CSF /HIV+	Botswana, 2003	Litvintseva	37
Bt152	A	**VNB**	Botswana	CSF /HIV+	Botswana, 2003	Litvintseva	37
Bt154	A	**VNB**	Botswana	CSF /HIV+	Botswana, 2003	Litvintseva	37
Bt157	A	**VNB**	Botswana	CSF /HIV+	Botswana, 2003	Litvintseva	37, 40
Bt158	A	**VNB**	Botswana	CSF /HIV+	Botswana, 2003	Litvintseva	37
Bt162	A	**VNB**	Botswana	CSF /HIV+	Botswana, 2003	Litvintseva	37
Bt204	A	**VNB**	Botswana	CSF /HIV+	Botswana, 2003	Litvintseva	37, 40
Bt206	A	**VNB**	Botswana	CSF /HIV+	Botswana, 2003	Litvintseva	37, 40
Bt208	A	**VNB**	Botswana	CSF /HIV+	Botswana, 2003	Litvintseva	37

AIM 3, Continued							
Strain	Serotype	MLST genotype	Country of origin	Source	Comments	Provided by PI	Reference
LP-RSA756	A	**VNB**	Botswana	CSF /HIV+	Botswana, 2003	Litvintseva	34
LP-RSA6618	A	**VNB**	Botswana	CSF /HIV+	Botswana, 2003	Litvintseva	34
MW-RSA1186	A	**VNB**	Botswana	CSF /HIV+	RSA, 2010, original isolate	Litvintseva	
MW-RSA3179	A	VNB	Botswana	CSF /HIV+	RSA, 2010, relapse isolate	Litvintseva	
MW-RSA1607	A	**VNB**	Botswana	CSF /HIV+	RSA, 2010, original isolate	Litvintseva	
MW-RSA3321	A	VNB	Botswana	CSF /HIV+	RSA, 2010, relapse isolate	Litvintseva	
MW-RSA913	A	**VNB**	Botswana	CSF /HIV+	RSA, 2010, original isolate	Litvintseva	
MW-RSA2967	A	VNB	Botswana	CSF /HIV+	RSA, 2010, relapse isolate	Litvintseva	
RTC3	A	**VNB**	Botswana	CSF /HIV+	Botswana, clinical, 2007	Litvintseva	34
LP-RSA3144	A	**VNB**	Botswana	CSF /HIV+	Limpopo Pr., 2006	Litvintseva	34
Tu406-1	A	**VNB**	Botswana	Mopane tree	Tuli, 2007	Litvintseva	34
Tu416-1	A	**VNB**	Botswana	Mopane tree	Tuli, 2007	Litvintseva	34
Tu229-1	A	**VNB**	Botswana	Mopane tree	Tuli, 2007	Litvintseva	34
Tu369-1	A	**VNB**	Botswana	Mopane tree	Tuli, 2007	Litvintseva	34
Tu401-1	A	**VNB**	Botswana	Mopane tree	Tuli, 2007	Litvintseva	34
Tu422-1	A	**VNB**	Botswana	Mopane tree	Tuli, 2007	Litvintseva	34
Tu360-1	A	**VNB**	Botswana	Mopane tree	Tuli, 2007	Litvintseva	34
Tu248-1	A	**VNB**	Botswana	Mopane tree	Tuli, 2007	Litvintseva	34
PLR10	A	**VNB**	Botswana	Mopane tree	L.Roushe, northern Botswana, 2009	Litvintseva	34
Ze90-1	A	**VNB**	S. Africa	Eucalyptus tree	Zerrust, 2007	Litvintseva	34

AIM 4							
Strain	Serotype	MLST genotype	Country of origin	Source	Comments	Provided by PI	Reference
A1-35-8-E#1		A1/M1		exp. evolved		Litvintseva	
A1-35-8-E#2		A1/M1		exp. evolved		Litvintseva	
A5-35-17-E#1		A5/M5		exp. evolved		Litvintseva	
A5-35-17-E#2		A5/M5		exp. evolved		Litvintseva	
H99E		A1/M1		exp. evolved	less virulent	Perfect	
H99W		A1/M1		exp. evolved	less virulent	Perfect	
H99S		A1/M1		exp. evolved	more virulent	Perfect	
MPC31-1		A1/M1		exp. evolved	revertant of pyc1	Perfect	
MPC31-2		A1/M1		exp. evolved	revertant of pyc1	Perfect	
MPC33-1		A1/M1		exp. evolved	revertant of pyc1	Perfect	
MPC33-2		A1/M1		exp. evolved	revertant of pyc1	Perfect	
MPC34-1		A1/M1		exp. evolved	revertant of pyc1	Perfect	
MPC34-2		A1/M1		exp. evolved	revertant of pyc1	Perfect	