Comparative Genomics of *Coccidioides* and Other Pathogenic Dimorphic Fungi

A whitepaper for Microbial Genome Sequencing submitted by Li-Jun Ma\(^1\) and Matthew Henn in collaboration with the Dimorphic Fungal Genomes Consortium (DFGC)

\(^1\) Broad Institute, 7 Cambridge Center, Cambridge MA 02141
617 258 5091
lijun@broad.mit.edu
SUMMARY

The systemic dimorphic fungal pathogens are the most common etiologic agents of fungal pulmonary infection in healthy hosts. This group of phylogenetically related fungi includes *Coccidioides*, *Blastomyces*, *Histoplasma*, *Lacazia* and *Paracoccidioides* species, all of which display thermally regulated dimorphism. *Coccidioides immitis*, one of the causal agents of coccidioidomycosis (Valley Fever), was recognized as an NIAID Group II re-emerging pathogen due to the increased disease incidence and its ability to cause fatal meningitis. To fully understand the pathogenicity of *Coccidioides*, we propose to use comparative analysis to define the underlying mechanisms of conserved aspects of biology and virulence in the context of the larger group of dimorphic fungal pathogens.

Building upon our project for “Comparative Genomics of *Coccidioides* spp.”, we propose to create high quality reference genomes centered on each dimorphic fungal genus to facilitate such comparisons. Our sequencing targets have been selected based on their genetic tractability, distinct clinic manifestations, virulence, and availability of existing genomic resources. Specifically we propose to sequence:

- Two African strains of *H. capsulatum* that cause strikingly different disease manifestations;
- Additional coverage of *H. capsulatum* strain G186AR, a genetically tractable lab strain that has been sequenced to 6X coverage at Washington University, St. Louis;
- Two strains of *B. dermatitidis* with genetic tractability but significantly different pathogenicity;
- One clinic isolate of the obligate pathogen *L. loboi*.
- Three *P. brasiliensis* strains from distinct phylogenetic lineages that have distinct phenotypes and induce different host immune responses

In addition, we will generate supporting resources to improve assembly and annotation of these reference genomes. These resources include optical maps for one strain each of *H. capsulatum*, *B. dermatitidis* and *P. brasiliensis*; and EST sequences from *H. capsulatum* and *B. dermatitidis*. Such enriched genomic resources will provide a platform to support the comparative analysis in:

- Identifying the “core” of common genes, which control the shared biological processes including the dimorphic switch;
- Identifying genes that occur only in *Coccidioides* spp., which reflect recent evolutionary changes including *Coccidioides*-specific virulence factors;
- Creating a catalogue of genetic differences associated with different disease manifestations among different dimorphic fungi.

By widening our comparative studies of *Coccidioides* to include these closely related dimorphic fungal pathogens, we will significantly enhance our understanding of the evolution and maintenance of pathogenic mechanisms in both *Coccidioides* and other systemic dimorphic fungal pathogens. In addition, *Histoplasma* remains the best studied of the dimorphic fungal pathogens with significant supporting resources. The knowledge gained through this study will leverage the extensive research tools and therapeutic approaches developed in *Histoplasma* for this entire class of pathogens.
BACKGROUND

Medical significance: Fungal family Onygenaceae contains several important endemic human pathogens, including Coccidioides, Histoplasma, Blastomyces, Lacazia loboii and Paracoccidioides (Fig. 1). They are the most common etiologic agents of fungal pulmonary infection in healthy hosts. In the United States alone, they collectively cause over one million new infections each year.

Among them, Coccidioides species cause serious and sometimes fatal disease in otherwise healthy people of all ages and ethnicity. On average, 100,000 persons are infected per year in the United States (Chiller et al. 2003, Stevens 1995). The incidence of this disease is increasing due to increased spore exposure as a consequence of new development and population growth in the Southwest, especially the growth of the immunocompromised population (Cole et al. 2004, Kirkland and Fierer 1996). In 2003, C. immitis was recognized as a NIAID Group II re-emerging pathogen (NIAID 2005), and is considered a select agent for bioterrorism by the Center for Disease Control and Department of Health and Human Services (General 2002).

Much like Coccidioides species, Histoplasma species, and the etiologic agent of histoplasmosis, cause significant morbidity in immunocompetent hosts. Approximately 500,000 infections are thought to occur every year in the U.S. (Eissenberg and Goldman 1991, Marques et al. 2000, Woods 2003). H. capsulatum is endemic in the Ohio River Valley through the Midwestern United States into Texas and is a leading pathogen affecting AIDS patients in the Midwest (Sternberg 1994). A recent national report of endemic mycoses requiring hospitalization revealed that the majority of patients who died from these infections were immunocompetent, consistent with previous observations that otherwise-healthy individuals are affected by these pathogens.

In addition, B. dermatitidis causes blastomycosis, one of the principal systemic mycoses of humans and other mammals. P. brasiliensis is the causal agent of paracoccidioidomycosis (PCM), one of the most important human systemic mycosis in Latin America (Restrepo et al. 2001). The medical significance of the diseases caused by this group of pathogens demands a more thorough understanding of their pathogenicity and ultimately the development of novel therapeutic strategies.

Dimorphic switch: One common feature of this group of pathogens is that they grow in the soil in the filamentous form (Fig. 2a), and these long chains of cells produce asexual spores. Both the spores and mycelial fragments can aerosolize if the soil is disturbed. Once introduced into a host via inhalation, the thermally regulated dimorphic switch is flipped (Fig. 2b) and the fungus converts to the budding yeast form (Fig. 2c), for the majority of these pathogens, and a related spherule/endospore form for Coccidioides. The yeast-like parasitic stage of L. loboii is also found inside the macrophages of infected patients. However, it has not yet been possible to demonstrate the dimorphic switch in L. loboii, because it cannot be uncultivated in the laboratory condition.
This thermally regulated dimorphism is the single-most definitive phenotype of this group of fungi and is a requirement for their pathogenesis (Medoff et al. 1986). The temperature-induced dimorphic switch is accompanied by a shared sequence of biochemical events ((Maresca and Kobayashi 1989, Maresca and Kobayashi 2000, Medoff et al. 1987), described in Figure 1b, and occurs simultaneously with changes in the cell-wall composition, such as increased chitin content (Ruiz-Herrera and San-Blas 2003), migration and reorganization of membrane lipids (Levery et al. 1998, Vigh et al. 1998), structural alterations in the carbohydrate polymers (Toledo et al. 1999), and a large shift in global gene expression (Hwang et al, 2003).

Very little is known about the molecular regulators that allow cells to switch their morphology in response to temperature. However, molecular genetic studies in Blastomyces and Histoplasma are beginning to identify genes required for this process. One such example is the discovery of a global regulator gene, termed DRK1 (for dimorphism regulating kinase), in B. dermatitidis (Nemecek et al. 2006). Encoding a hybrid histidine kinase, this global regulator senses the mammalian host environment and controls the phase transition process of B. dermatitidis. The mutant of this gene is avirulent and exhibits global defects in temperature-regulated morphogenesis associated with dimorphic switch. Expression in trans of an intact copy of this gene in the mutant restores the wild-type phenotype.

This histidine kinase is present and highly conserved in other dimorphic fungi, such as H. capsulatum and Coccidioides and its function has been tested in H. capsulatum (Nemecek et al. 2006). The H. capsulatum mutant exhibited pleiotropic defects in the temperature-induced phase transition process similar to those observed in B. dermatitidis, including morphogenesis, cell-wall integrity, and sporulation. Transcripts for all known yeast-phase-specific virulence genes are absent in the H. capsulatum knockout strain. In a murine model of histoplasmosis after intratracheal infection with spores, DRK1-silenced strains of H. capsulatum were sharply reduced in virulence compared with wild-type strains.

The sequence conservation of DRK1 among Blastomyces, Histoplasma and Coccidioides, and the confirmed, highly conserved function in both B. dermatitidis or H. capsulatum, suggests that this gene may play an essential role in yeast-phase transition and virulence gene expression in Coccidioides species and other dimorphic growth pathogens. Furthermore, this study indicates that the yeast-phase specific virulence genes are regulated transcriptionally by mechanisms that are conserved in dimorphic fungi. Identification of the global regulators that control morphogenesis, virulence, and sporulation is crucial in understanding the biology processes controlling dimorphic switch and provide means to control this class of pathogens.

Figure 2. Physiological changes associated with the thermally regulated switch, characteristic for pathogenesis in the dimorphic fungi.
COMPARATIVE GENOMICS

Few genetic determinants involved either in phase transition or virulence/pathogenicity have been identified in dimorphic fungi despite interest from the medical community and tremendous effort from the fungal research community. Many important questions remain to be answered concerning regulation of the phase-transition and phase-specific virulence factors in this class of pathogens, such as: What are the genes and signaling pathways directly involved in the phase transition process? What are the conserved mechanism(s) that regulate the transcription of such phase-specific genes or pathways during the phase transition? and How widely conserved are these programs for coordinating gene regulation in systemic dimorphic fungal pathogens and other microorganisms or higher eukaryotes that undergo cellular morphogenesis during environmental adaptation?

These questions will be addressed through the comparative genomics proposed in this white paper. Such an approach has been successfully applied to study function and evolution at various levels of biological organization including highlighting functionally important elements (Galagan et al. 2005, Kellis et al. 2004, Kellis et al. 2003), and the identification of species-specific features (2005, Waterston et al. 2002).

Analysis of these genomes in the context of an evolutionary framework will allow the identification of conserved genomic features, including genes and pathways, among this group of dimorphic human pathogens. Study of these conserved genomic features will allow us to discover and clarify regulatory mechanisms that are shared among dimorphic fungi, including the principles that govern the dimorphic switch. Such study will lead to the identification of genetic determinants of pathogenicity, virulence, and transmission, which will have potential as therapeutic and vaccine targets across the dimorphic fungi.

In combination with our current Coccidioides projects, these comparisons will help us to identify genetic features that occur only in Coccidioides spp., including the Coccidioides specific virulence factors. Such unique genetic features have potential as targets for diagnostics, therapeutic and vaccines specific to the Coccidioides species.

Finally, the comparisons centered on Histoplasma, Blastomyces, and Paracoccidioides will allow us to create a catalogue of genetic differences for each dimorphic fungal pathogen. These differences in the context of carefully selected distinct phenotypic variations and disease manifestations will provide a platform to study recent evolutionary events and their impact on pathogenicity development.

PROJECT PLAN

The overall objective of this white paper is to generate genomic resources for key representatives of dimorphic fungi within the Onygenaceae family (Table 1) to enable the proposed comparative studies. Strains of dimorphic fungal species have been carefully selected with consideration of their genetic tractability, distinct clinic manifestations, virulence, and availability of existing genomic resources. The significance of each species, the rationale for the strain selection and specific coverage, and the existing resources will be described in detail in the section follows.
Table 1. Sequencing objectives

<table>
<thead>
<tr>
<th>Species</th>
<th>Strain</th>
<th>Approx. genome size (Mb)</th>
<th>Coverage (X)</th>
<th>Strain description</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Histoplasma capsulatum</em></td>
<td>H88</td>
<td>30–40</td>
<td>8</td>
<td>African histoplasmosis</td>
</tr>
<tr>
<td></td>
<td>H143</td>
<td>30–40</td>
<td>4</td>
<td>Respiratory histoplasmosis</td>
</tr>
<tr>
<td></td>
<td>G186AR</td>
<td>30.5</td>
<td>4</td>
<td>Tractable genetic system</td>
</tr>
<tr>
<td><em>Blastomyces dermatitidis</em></td>
<td>SLH#14081</td>
<td>40</td>
<td>8</td>
<td>High virulence clinic strain</td>
</tr>
<tr>
<td></td>
<td>ER-3</td>
<td>40</td>
<td>8</td>
<td>Avirulent strain</td>
</tr>
<tr>
<td><em>Lacazia loboï</em></td>
<td>EDM7</td>
<td>ND</td>
<td>8</td>
<td>Obligated isolate maintained in mouse model</td>
</tr>
<tr>
<td><em>Paracoccidioides brasiliensis</em></td>
<td>Pb18</td>
<td>30</td>
<td>8</td>
<td>High virulence clinical isolate</td>
</tr>
<tr>
<td></td>
<td>Pb03</td>
<td>30</td>
<td>8</td>
<td>Cryptic species, lower virulence</td>
</tr>
<tr>
<td></td>
<td>Pb01</td>
<td>30</td>
<td>8</td>
<td>Molecular model</td>
</tr>
</tbody>
</table>

ND, not determined.

**Histoplasma spp.**

*Histoplasma capsulatum* is thought to be the most common cause of fungal respiratory infections in the world. This organism is a primary pathogen capable of causing disease in healthy hosts as well as immunocompromised individuals. Whereas most infections experienced by healthy hosts are mild, 10% of cases of *H. capsulatum* infection result in life-threatening complications, such as inflammation of the pericardium and fibrosis of major blood vessels (Durkin et al. 2001). Up to 25% of HIV patients living in areas where *H. capsulatum* is endemic develop disseminated disease, and histoplasmosis has re-emerged as a major AIDS-related complication in these regions (Sarosi and Avies 1996). *H. capsulatum* poses a particular threat to the elderly and to immunocompromised patients, with mortality rates as high as 10% in HIV patients with disseminated histoplasmosis (Rachid et al. 2003). Once a host is infected with *H. capsulatum* a latent infection is established. *H. capsulatum* can reactivate if the immune status of the host declines; thus HIV patients infected with *H. capsulatum* often require life-long anti-fungal therapy.

Recent studies using minimal genome sequence revealed a remarkable diversity among *H. capsulatum* isolates (Kasuga et al. 2003). The fungi collectively known as *H. capsulatum* (=*Ajellomyces capsulatus*) actually comprise seven distinct phylogenetic clades (Fig. 3). Genome sequencing has been performed on representatives from both of the phylogenetic clades in Northern America 1, (NAm I strain WU24) and Northern America 2 (NAm II strain G217B), as well as a strain from Panama (strain G186AR). These strains have notable phenotypic differences, such as alterations in cell wall components and extent of virulence in the host.
I. Strains selection

**G186AR:** Notably, G186AR is considerably more amenable to molecular manipulations (such as gene disruption and RNA interference) than the other sequenced strains. Currently, approximately 6X coverage of G186AR was performed by the Genome Sequencing Center at Washington University in St. Louis. This assembly contains 3735 contigs linked into 1049 scaffolds. Comparing to Coccidioides reference assembly (containing only 50 contigs in 7 scaffolds), this assembly is notably lacking the desirable continuity. Obtaining more sequence to create high quality assembly for G186AR will greatly facilitate the molecular analysis of *H. capsulatum*, which will be highly beneficial to the entire *H. capsulatum* research community, as well as the dimorphic fungal community in general. Therefore, we propose to generate an additional 4X coverage of G186AR. To ensure the delivery of a high-quality, chromosome-based assembly, we propose to create an optical map for this strain as well. We will use the ESTs of this strain generated at Washington University to improve annotation.

**African phylogenetic clade:** Interestingly, some *H. capsulatum* isolates cause strikingly different disease manifestations, the basis of which is completely unknown. For example, G217B, G186AR, and WU24 cause respiratory and systemic disease in humans. In contrast, some African *H. capsulatum* isolates cause a distinct disease, African histoplasmosis, that is characterized by cutaneous and subcutaneous lesions in the bone (Rippon et al. 1988). The recent phylogenetic analysis of *H. capsulatum* described above indicates that different isolates from the African clade of *H. capsulatum* can cause either systemic histoplasmosis (e.g., strain H143) or African histoplasmosis (e.g., strain H88) (Kasuga et al. 2003). Characterizing the genomes of H143 and H88 in the African clade will help highlight differences between the two strains that influence disease manifestation in the host. In addition, because these isolates are from phylogenetic lineages that are distinct from the existing sequenced strains, they will add significantly to a comparative genomic analysis. For example, comparing sequence from the Latin America (G186AR), NAm I (WU24), NAm 2 (G217B), and the African clade (H143 and H88) strains should allow the identification of genes that are selected in diverse geographic environments, as well as the identification of rapidly evolving genes. These studies will diversify and strengthen our understanding of *H. capsulatum* biology, and will provide an additional richness to the dimorphic analysis as a whole. Therefore, we propose to generate 8X sequence for strain H143 and 4X for H88. In addition, we will sequence both ends of 12,000 normalized cDNA clones from *H. capsulatum* (H88) in order to facilitate gene identification and annotation for these two closely related fungi.
II. Genome facts and resources

The estimated genome size of G186AR is 30.5 Mb, based on the preliminary 4X coverage by the Washington University Genome Sequencing Center. We also expect that the genome sizes of H143 or H88 will be between 30 and 40 Mb. Although the number of chromosomes is unknown, the laboratory of William Goldman at Washington University, St. Louis, has used labeled telomeric probes and restriction analysis of genomic DNA to document approximately 14 hybridizing bands in strain G186 (W. Goldman, unpublished data). Assuming that bands correspond to the ends of digested chromosomes, that analysis suggests a total chromosome number of seven. Isolates of both strains are currently available: H143 is isolate number CBS 287.54 at the Centraalbureau voor Schimmelcultures, a central strain repository in Baarn, the Netherlands. H88 is isolate number 32281 at the American Type Culture Collection.

A number of well-established tools, including targeted gene disruption, RNA interference, random insertional mutagenesis, ectopic expression, and transcriptional profiling using microarray analysis, make *Histoplasma* one of the more tractable systemic dimorphic pathogens and will assist in use of sequence information by the community.

*Blastomyces dermatitidis*

*Blastomyces dermatitidis* is the causative agent of blastomycosis, one of the principal systemic mycoses of humans and other mammals, which occurs on most continents of the world and most commonly in North America. In the United States, blastomycosis is endemic in and around the Ohio and Mississippi River Valleys. Blastomycosis is a primary fungal pathogen that causes disease in healthy people with an intact immune system and often requires treatment to prevent progressive pneumonia. After mild or asymptomatic infection, Blastomycosis often becomes latent and reactivates under conditions of compromised immunity. Hence, it behaves as an opportunistic pathogen in patients with AIDS and other immune impairment.

Conidia constitute the infectious form of this organism. After inhalation into the lungs of animals or humans, the conidia swell and germinate into yeast, the form required for pathogenicity and proliferation (Disalvo 1992). Unchecked proliferation in turn results in blastomycosis and can become life threatening when undiagnosed or untreated. Acute blastomycosis is occasionally self-limiting, but more frequently progresses to produce severe pulmonary disease with the potential for dissemination to other organs, skin and bone (Causey and Campbell 1992).

I. Strains selection

**SLH#14081**: We propose to generate high-quality (8X) reference sequence of a newly cultivated human clinical isolate, SLH#14081. This strain has a number of desirable features both clinically and experimentally. It is highly virulent in mice and stable on serial passage. Genetically, it can be readily manipulated so that it is amenable to high-efficiency gene transfer, insertional mutagenesis, allelic replacement, and gene silencing and reporter gene expression. Importantly, strain 14081 sporulates, yielding large numbers of the uninucleate haploid conidia that constitute the infectious form of this fungus. In contrast with the biological properties of strain of SLH#14081, the lab strain ATCC 26199, which is being used on a low-coverage genome project at Washington University, has lost the ability to sporulate and is not stable *in vitro* upon serial passage (Hogan et al. 1996, Klein et al. 1994). This strain has a high frequency of repetitive DNA in the genome (30–40%) and readily gives rise to variants that are altered in pathogenicity and in the expression of the virulence factors BAD1 and α-(1,3)-glucan. Having a well-characterized, stable strain, such as SLH#14081, that sporulates and is genetically tractable, will allow the field to analyze genes that are essential to disease transmission. The comparison between these two strains will reveal the insights on the genetic control of sporulation process that generates the
infectious spores. In addition, 12,000 normalized cDNA clones from SLH#14081 will sequence from both ends to improve the automated gene annotation.

**ER-3:** We will sequence one avirulent North American strain, ER-3 to 4X coverage to characterize the genetic differences between it and the clinically distinct isolate SLH#14081. Strain ER-3 was isolated from a woodpile in Northern Wisconsin, a highly endemic region of infection. The isolate is remarkable for its inability to produce illness in a murine model of pulmonary blastomycosis. In contrast to SLH#14081, which kills mice from pneumonia in 12–14 days, ER-3 fails to produce illness or death in mice that were observed (B. Klein, personal communication) for several months after experimental infection with a “lethal dose”. Additional features of ER-3 that are of considerable value include: 1) its genetic tractability (it is highly transformable by *Agrobacterium*-mediated gene transfer); and 2) its ability to sporulate efficiently in the mold form, allowing isolation of uninucleate, haploid progeny, for example, from genetically manipulated yeast. This isolate has been deposited in the American Type Culture Collection (MYA-2586).

**II. Genome facts and resources**

The estimated genome size for both *Blastomyces* strains is about 40 Mb, based on information from the ongoing Washington University genome-sequencing project; however, much of this is repetitive DNA, as indicated above.

Tools to genetically manipulate *B. dermatitidis* have been well established and are now widely used in the *Blastomyces* research community. DNA-mediated gene transformation can be done reliably and with high efficiency by electroporation (Sullivan et al. 2002). Multiple, dominant selectable markers, marked auxotrophic strains, conditional promoters, LacZ, Gus, GFP and RFP reporters for promoter and other analyses, and reverse genetic and forward genetic approaches such as described above in the elucidation of *DRK1*, the conserved global regulator of dimorphism in fungi, have all been established and are readily available to the research community for use in investigating newly discovered genes.

**Lacazia loboi**

*Lacazia loboi* is an uncultivated fungal pathogen of humans and dolphins that causes cutaneous and subcutaneous infections in the tropical areas of the Americas. Patients with lacaziosis develop keloid-like lesions on the skin, and the yeast-like parasitic stage of *L. loboi* is found in large numbers inside the macrophages of infected humans and dolphins (Opromolla and Nogueira 2000). It was recently found by phylogenetic analysis that this pathogen is closely related to *P. brasiliensis* and to the other fungal dimorphic members of the family Onygenaceae. (Herr et al. 2001, Mendoza et al. 2001). However, in contrast to these close related fungi, *L. loboi* is an obligate parasite. Preliminary analysis based on a few sequenced genes of *L. loboi* suggests that it has genomic features in common with other obligate parasites with decaying genomes, such as *Mycobacterium leprae*, *Pneumocystis jerovecii*, and *Treponema pallidum* (Mendoza et al. 2005, Vilela et al. 2005). Thus *L. loboi* might be an example of a eukaryote parasite having undergone reductive evolution after years of a prolonged parasitic lifestyle. A *L. loboi* genome sequence project should yield rewarding results for understanding the genomic complexity of this and other dimorphic pathogens.

Although *L. loboi* has not yet been cultivated in the laboratory, a experimental BALB/c mouse model has been developed (Belone et al. 2002, Madeira et al. 2000). This mouse model exhibits similar symptoms to those observed in human patients with lacaziosis, such as fungal count, viability index before and after inoculation, presence of macroscopic lesions and histopathological findings. The mouse model has been used to study the pathogenesis of this disease and is used as an additional tool to search for therapeutic alternatives.
We propose to generate high quality (8X) reference sequence of strain EDM7. This strain is presently maintained in several mice and available for the proposed studies. High quality genomic DNA can be readily prepared from this model. The genome size is undetermined experimentally. However, based on the reductive nature of this organism, it is estimated to be less than 30 Mb, the average genome size of other dimorphic fungi in the same family. Dr. Leonel Mendoza at Michigan State University will provide genomic DNA for the sequence project.

**Paracoccidioides brasiliensis**

*Paracoccidioides brasiliensis* is the causal agent of paracoccidioidomycosis (PCM), one of the most important human systemic mycosis in Latin America (Restrepo et al. 2001). It is estimated that around 10 million people are infected in South America (Brummer et al. 1993). Most infections occur in Central and South America, particularly in Brazil, Venezuela, and Colombia. The annual incidence rate in Brazil is 10–30 infections per million inhabitants, and the mean mortality rate is 1.4 per million per year (Restrepo et al. 2001). PCM is a granulomatous disease that produces a primary pulmonary infection. In addition, disseminated forms may also be observed. The reticuloendothelial system, skin, mucous membranes and lymph nodes are frequently affected in cases of disseminated disease.

*Paracoccidioides brasiliensis* has shown extensive genetic variability when analyzed by molecular tools, such as random amplified polymorphic DNA, restriction fragment length polymorphism (Nino-Vega et al. 2000, Soares et al. 1995), and electrophoretic karyotyping (Feitosa et al. 2003, Montoya et al. 1997). Three distinct lineages, S1, PS2, and PS3 (Fig. 4), were recognized using a combined data set of eight regions in five nuclear loci (Matute et al. 2006). A major group (S1) consists of isolates from Brazil, Argentina, Venezuela, Peru and Paraguay. Colombian isolates formed a separate group, PS3. PS2 represents a cryptic phylogenetic species of *P. brasiliensis* (so far, six isolates from Brazil and Venezuela).

The two sympatric lineages, S1 and PS2, occur within the same geographical regions, suggesting barriers to gene flow other than geographic isolation. The genetic variation is also reflected at the *PbGP43* gene (Morais et al. 2000), a specific diagnostic antigen involved in both cellular and humoral responses, protection against murine paracoccidioidomycosis (PCM) and adhesion. The PS2 isolates, Pb2, Pb3 and Pb4, contain highly substituted *PbGP43* sequences (14 substitutions). The substitutions result in the basic gp43 protein in the PS2 isolates, in contrast to the neutral or mildly acid gp43 in other isolates. The immune responses elicited by such basic gp43 were richer in IgG2a, IgG2b and IgG3, suggesting a Th1 predominant type of host immunity. The other isolates evoked mostly an IgG1 and IgA, which reflect a Th2-driven response. As a result, PS2 isolates provoked only mild infection when testing the host–parasite relationship in the B10.A mouse model (sensitive to paracoccidioidomycosis).

![Figure 4: Unrooted tree showing the partitions found in *Paracoccidioides brasiliensis*, based on combined data of five nuclear loci obtained with weighted maximum parsimony. The values above the branches represent their individual support: the first is the tree length, the second is the weighted high bootstrap, and the third is the posterior probabilities (modified from Matute et al. 2006).](image-url)
In addition, isolate Pb01, the most thoroughly studied isolate at the molecular level, doesn’t cluster together with any of these three phylogenetic groups (Molinari-Madlum et al. 1999). Molecular advances in the field have produced significant reports on transcript characterization of this pathogen and further confirmed the genetic and metabolic diversity shown by *P. brasiliensis* in its multiple isolates.

**I. Strains selection**

To enable the studies of the two distinct and sympatric species (S1 and PS2), and the distinct, well-studied isolate Pb01, and the comparisons among them, we propose to sequence three isolates.

**Pb18:** Pb18 is the representative of the major phylogenetic group S1 and has been extensively used in the literature due to its proven virulence in mice when inoculated by the intraperitoneal (i.p.), intratracheal (i.t.) and intravenous (i.v.) routes (Calich et al. 1998). Chromosomes of Pb18 can be separated by pulsed field gel electrophoresis (PFGE) in 4–5 distinct bands. The haploid genome size is approximately 30 Mb (Almeida et al. 2006). Analysis of its transcriptome yielded ESTs covering about 40 to 50 % of the estimated genome of *P. brasiliensis* (Felipe et al. 2005). A microarray based on these cDNAs was already developed and is available to the community (Nunes et al., 2005).

**Pb3:** Pb3 is one of the better-studied isolates of PS2. It has been used in molecular, phylogenetic, and virulence studies. Transcriptional regulation seems to vary in Pb3 when compared with Pb18, as suggested by the down regulation of *PbGP43* due to growth temperature increase (Carvalho et al. 2005). Accordingly, the heat shock genes *PbLON* and *PbMDJ1* are upregulated more slowly in Pb3 than in Pb18 after a temperature shift from 36°C to 42°C. In Pb18, but not in Pb3, *PbMDJ1* seems to be preferentially expressed in yeast, as suggested by real time RT-PCR analysis of the genes during mycelium-to-yeast-to mycelium transition. Pb3 has four chromosomal bands separated by PFGE. The size of each chromosome is 9.5, 6.4, 5.3 and 3.0 Mb, respectively. This gives a haploid genome of about 30 Mb.

**Pb01:** Pb01 is a clinical isolate from an acute form of PCM in an adult male. So far, this is the most thoroughly studied isolate at the molecular level, as deduced from an extensive analysis of its transcriptome that yielded ESTs covering about 80% of the estimated genome of *P. brasiliensis* (Felipe et al. 2005). The nuclear genome size was estimated by PFGE at around 30 Mb (Montoya et al. 1997).

**II. Genome facts and resources**

Two *P. brasiliensis* EST sequencing programs are ongoing by research groups from São Paulo State (Pb18 strain) and Center-West (Pb01 strain) universities, Brazil. Over 30,000 EST sequences have been generated through these EST projects. In addition, *P. brasiliensis* research community has developed a large collection of unique RSTs (Random Sequence Tags) from strain Pb18 (about 0.2X coverage assuming 30 Mb as the genome size) and a microarray with 4600 elements based on representatives of EST clusters.

**SEQUENCING, ASSEMBLY, AND ANNOTATION**

Paired-end sequence reads from multiple shotgun libraries will be prepared in different vector types with a variety of insert sizes. Test data from each library will be obtained separately and analyzed prior to approval for production sequencing. Genome data will be assembled using the ARACHNE assembly package developed at the Broad Institute. Annotation teams at the Broad will use all available evidence and follow standard protocols at the Broad MSC for genome annotation. In addition, we have utilized both EST sequences from cDNA libraries to identify and correct annotation problems. The EST data will be used as evidence to create a reliable gene
models, which will serve two purposes: 1) to train various gene-calling programs, and 2) to validate the result of automated genes.

OPTICAL MAP

Optical mapping is an enabling technology for whole-genome assembly, which involves the capture of individual chromosomal DNAs, followed by in situ digestion by selected restriction enzymes. The resulting fragments are then visualized directly to produce detailed optical restriction maps for individual chromosomes. The detailed map can be used to order and orient the sequence contigs and to link all of the sequence in the assembly into whole chromosomes and provide a mean to independently validate the assembly. This technique was developed in 1993 (Schwartz et al. 1993) and has been widely applied in genome sequencing projects, including many bacteria (Zhou et al. 2004) and eukaryotes (Armbrust et al. 2004). Recently this technique has been used to map whole chromosomes in numerous fungi, including Aspergillus fumigatus, Rhizopus oryzae, and Neurospora crassa.

Since there are no genetic maps available for any of the dimorphic fungi, the creation of an optical map for each representative will be extremely valuable. Specifically, we propose to generate optical maps for one strain of H. capsulatum (G186AR), B. dermatitidis (SLH 14081) and P. brasiliensis (Pb18) to reliably assemble the genomes to chromosomes and provide references to improve assembly quality for all the assemblies. The integration of the optical maps and the assembly will anchor the sequence scaffolds to the chromosomes and provide a comprehensive landscape of the genome structure. The optical map will be generated by Dr. David Schwartz at University Wisconsin (http://www.biotech.wisc.edu/GSTP/). The genome assembly generated at the Broad will be used to select suitable restriction enzyme sites for linking the assembly to the optical map.

DIMORPHIC RESEARCH COMMUNITY

The dimorphic fungal scientific community has a history of being tremendously cooperative in supporting and sharing resources to advance research on this group of fungal pathogens. This was evidenced by the recent 2006 American Society of Microbiology’s Conference on Dimorphic Fungal Pathogens (Denver, Colorado, March 13-17, 2006) that was attended by over 90 international scientists. At this conference, scientists from the Broad presented an overview of the current Coccidioides genomics project and also outlined the strategy for understanding the dimorphic switch through a larger comparative genomics project on dimorphic fungal pathogens. Several working groups centered on each of the genera of fungi included in this proposal met with the Broad scientists to refine the criteria for strain selection, select potential strains, and identify the core questions addressed by the selected strains. This conference provided the opportunity to continue the development of this proposal in a highly collaborative environment that included the direct scientific input from a wide spectrum of the dimorphic research community.

In addition to the assessment discussed above, the Dimorphic Fungal Genomes Consortium (DFGC), chaired by Dr. Bruce Klein, has been formed to support the dimorphic comparative genomics project. The DFGC members include:

Maria Sueli Felipe, University of Brasília, Brazil
Bill Goldman, Washington University School of Medicine, MO
Gustavo H. Goldman, Universidade de São Paulo, Ribeirão Preto, Brazil
Matt Henn, Broad Institute, Cambridge, MA
Bruce Klein, University of Wisconsin Madison, WI
The DFGC sought input and attained consensus concerning the number and prioritization of new sequencing targets through e-mail exchanges, telephone calls, and discussions at conferences. A critical component of this effort has been sharing of all data pertaining to strain selection, including data generated in our analyses. In addition, great weight was given to practical matters in the structure of both the sequencing and analysis components of the proposed work.

**MANAGEMENT**

**The Broad Institute (Li-Jun Ma) 10%**

Dr. Li-Jun Ma will be responsible for ensuring sufficient resources are allocated to the project to meet all deadlines, including sequencing, assembly, automated annotation and comparative analysis. Specifically, Dr. Anita Sil at University of California San Francisco will provide high quality genomic DNA for three *Histoplasma* strains. Dr. Bruce Klein at University of Wisconsin Madison will be responsible for the genomic DNA for two *Blastomyces* strains and mRNA from strain SLH#14081 for the normalized cDNA library. Genomic DNA of *Lacazia loboii* will be provided by Dr. Leonel Mendoza at Michigan State University from lacaziosis BALB/c mice model. Dr. Gustavo Goldman at University Sao Paulo of Brazil, Dr. Rosana Puccia, at University Federal de São Paulo and Dr. Juan McEwen at University of Antioquia Medellin of Colombia South America, and Maria Sueli Felipe at University of Brasilia will provide high quality genomic DNA for three *Paracoccidioides* strains respectively.

**DATA RELEASE**

In accordance with the NIAID’s principles regarding data release, we will publicly release all data generated under this contract as rapidly as possible. As required by our contract, NIAID will be provided with a 21–45 calendar-day period to review and comment upon all data prior to its public release.
LITERATURE CITED


