Genomic Determinants of Infection Competence in Dermatophyte Fungi

A white paper for the Fungal Genome Initiative submitted by:

Ted White and Matthew Henn on behalf of the Dermatophyte Genome Steering Committee

Corresponding author: Matthew Henn
The Broad Institute of Harvard and MIT
320 Charles Street, Cambridge, MA 02141 USA.
Phone: 617-324-2341. E-mail: mhenn@broad.mit.edu

1. Overview of rationale and aims of sequencing

Dermatophytes are one of the most common sources of human fungal infections. Annually they affect millions of individuals and are estimated to burden the United States healthcare system to the toll of $400 million each year for treatment alone. Examples of Dermatophytes include the fungi responsible for tinea pedis (athlete's foot) which has led to the immobilization of significant numbers of United States troops in recent history. Other species are the cause of tinea capitis a scalp infection that is a significant pediatric health problem in urban settings. The Dermatophytes are communicable and can cause chronic infections in healthy, immune-competent individuals, and hence have adapted to evade and maintain control of the host immune response for extended periods of time. These adaptations must result from the co-evolution between Dermatophytes and their human and other mammal hosts - evolutionary trajectories that resulted in species that span a continuum of host specificities and mating competence.

This white paper seeks to utilize this continuum of biological diversity to identify genomic features specific to chronic and acute human infections. The proposal identifies and prioritizes five Dermatophytes with top priority to the most common fungal infection in the world, *Trichophyton rubrum*. The five prioritized species represent human specific (anthropophile), mammal specific (zoophile) and soil dwelling (geophile) taxa. The sequencing goals of this project are to:

- establish reference sequences at 8-fold coverage for an anthropophile (*T. rubrum*), a zoophile (*Microsporum canis*), and a geophile (*M. gypseum*)
- generate 5-fold coverage of *T. tonsurans* and *T. equinum* two sister taxa that represent a recent evolutionary divergence in host specificity between an anthropophile and zoophile

This project addresses the current lack of adequate DNA sequence needed to support research in this important group of pathogens and to establish effective diagnostics, therapeutics and vaccines. The sequences of the five 22Mb genomes will provide the capacity to use comparative genomics to:

- decipher the acquisition, loss, and control of genes and genomic features that are the basis of host-specificity and mating competence in Dermatophyte fungi
- identify genetic features that are specific to anthropophile and zoophile Dermatophytes and that are potential targets for the development of disease diagnostics, therapeutics, and preventive strategies against Dermatophyte infections in humans and domesticated animals
- enhance the ongoing comparative genomic analyses within the Euascomycete human pathogens by the addition of taxa that do not utilize a dimorphic life history

2. Specific biological and biomedical rationales for the utility of new sequence data

2.1 Significance of dermatophyte fungal infections and need for molecular tools. Dermatophytes are the most common cause of fungal infections worldwide and impact millions of individuals annually. In the United States alone this translates into an economic impact on the health care system that is estimated to exceed $400 million a year for treatment alone (3, 13). In addition, large-scale epidemics have been reported. One well-characterized epidemic occurred in the Mekong Delta in Vietnam in 1966-1969. Severe fungal foot infections occurred in American troops, such that up to 50% of the troops in the region were immobilized. In another epidemic, 50% of troops became symptomatic when deployed to the hot and moist tropical climate of Panama in the 1980s (4). Despite the prominence of Dermatophyte infections and their resulting economic consequences, the research and medical communities lack a sophisticated understanding of these organisms' biology and consequently effective preventatives and therapeutics. These deficiencies in large part are due to the lack of genetic tools to enable the study of these fungi and their host-specificities.

Diagnosis of Dermatophyte infections relies on clinical presentation, requiring microscopy and culture for confirmation. Speciation, which can be important for prognosis and treatment, relies on culturing the organism which takes 2 to 4 weeks and pleomorphic growth can lead to misidentification. Speciation can also require nutritional and mating tests. Rapid diagnostic tests using current molecular methodologies have been slow to develop for the Dermatophytes (6). In addition, researchers lack the genetic and molecular tools to determine if recurrence of infection is due to reinfection or reactivation, or to determine if treatment failure is due to resistance or reinfection. Molecular epidemiology of an outbreak of fungal infections is also not possible.

Dermatophytes are responsible for a variety of skin infections at body locations including the feet, torso, scalp and nails (9). Fungal infections of the skin can result in scaling, fissuring, maceration and erythema, accompanied by itching or burning (reviewed in (9)). Nail infections are associated with thickening, discoloration and pain. Scalp infections can cause irreversible hair loss. The fungus can be asymptomatic, can cause a chronic infection (most common) or can have an acute onset phase associated with inflammation.

Unlike other fungi, Dermatophytes are communicable and cause infections in healthy, immune-competent individuals as well as in those with immune dysfunction. Estimates suggest that 30 to 70% of adults are asymptomatic carriers of these fungi (14). Dermatophytes are transmitted by shedding of infected skin cells and hair and by direct body contact, especially in children (9). The infecting fungi are commonly spread in public facilities such as swimming pools and gyms. Fungal infections of the foot are more common in men, are associated with warm and moist conditions (e.g. environment inside shoes or boots), and are associated with specific age groups (teenagers as they reach puberty, young adults, and the elderly) (9).

Currently therapy of fungal skin infections can take two forms. Topical treatments are usually available over the counter and are appropriate for infections of the glabrous skin and feet, although relapse is common, especially with *Trichophyton rubrum*. Prescription of systemic antifungal therapies is required for recalcitrant Dermatophyte infections including those of the nails, scalp and soles, none of which respond effectively to topical therapy. For both treatment routes success is low.

Taken together, the lack of effective diagnostic and treatment strategies, the sheer number of individuals that experience Dermatophyte infections, and the economic consequences highlight deficiencies in the research efforts aimed at understanding this very common group of fungal diseases. Under this proposal we intend to create the genomic infrastructure needed to enable genetic tools for the study of this important group of fungal pathogens. We will achieve this through generating sequence of a selected set of Dermatophyte species, with priority on the common fungal foot pathogen *T. rubrum*. The selection of taxa is structured around two parallel continuums, host specificity and mating competence. This structure provides the capacity to use comparative genomics to: (i) enhance the development of disease diagnosis, treatments, and preventive strategies against Dermatophyte infections in humans and domesticated animals, (ii) to decipher the acquisition of genomic features that enable host specific infections, and (iii) to enhance the ongoing comparative genomic analyses within the Euascomycete human pathogens by the addition of taxa that do not utilize a dimorphic life history.

2.2 Phenotypic comparative genomics to decipher the evolution of Dermatophyte pathogenicity in humans.
Dermatophytes are ascomycete fungi and members of the class Euascomycetes that includes *Aspergillus* and the dimorphic fungi. The Dermatophytes are thought to be basal in this distinct evolutionary trajectory and represent an early acquisition of human pathogenicity that contrasts that of the more derived dimorphic human pathogens *Coccidioides*, *Histoplasma*, *Blastomyces*, and *Paracoccidioides* (Fig. 1).

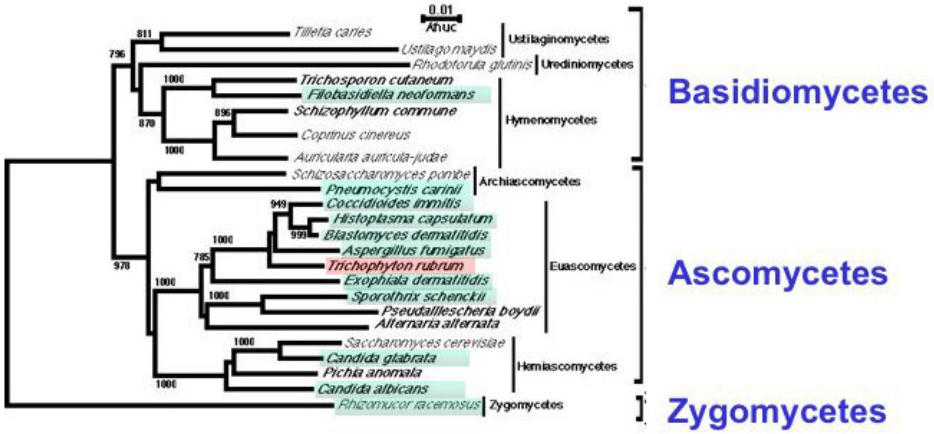

Figure 1 – Dermatophyte (T. rubrum), shown in red, as it relates to other fungi and fungal pathogens, shown in green (10).

One theory regarding pathogen and host interactions suggests that acute and severe infections are usually caused by pathogens that are not commonly associated with humans (i.e. HIV), while chronic and mild infections (non-life threatening) are caused by pathogens that have interacted with humans over evolutionary time (i.e. *Helicobacter pylori* that is responsible for gastric ulcers). Several fungal pathogens that are associated with cutaneous and subcutaneous infections may have evolved in parallel with humans, including Dermatophytes such as *T. rubrum*. Dermatophytes, like *T. rubrum*, are one of a limited number of fungal groups that infect healthy individuals. Thus, Dermatophytes have learned to evade or suppress the host immune system in establishing an infection; and chronic Dermatophyte infections must maintain this immune control for extended periods of time. The genetic features that enable Dermatophytes to infect humans and evade immune attack are not known. *A comprehensive understanding of the genomic features and regulatory motifs that are specific to Dermatophytes that infect humans is critical to understanding the pathogenicity of these fungi and to finding effective treatments. A primary objective of this project is to obtain genome sequence from taxa representing Dermatophytes with different host specificities that will permit the identification of acquired, lost, mutated, or rearranged genomic features that are signatures of the co-evolution between humans and Dermatophytes that has led to chronic infections.*

Dermatophytes are divided into three different groups based on their location and mode of transmission. These groups include geophilic (soil dwelling), zoophilic (animal reservoirs) and anthropophilic (human-specific) species. The geophilic species are commonly found in soil and only rarely are found in human infections. The zoophilic species are found in animals as well as in humans; these species do not commonly infect humans and usually cause acute inflammation upon infection. The anthropophilic species do not appear to have an animal reservoir. Anthropophilic species cause disease, but the disease is more controlled. Chronic infections are standard with anthropophilic species, suggesting that the fungi have adapted to their human hosts. *Generating genome sequence from species that represent each of these ecological phenotypes will enable a comparative genomic approach to identify genetic features that are associated with the ability of anthropophilic Dermatophytes to suppress a human immune response.*

Mating competence in Dermatophytes parallels host specificity; anthropophiles have lost the ability to mate and are clonal. This clonal reproduction in anthropophilic Dermatophytes may be a result of intense immune pressure from the host. Many Dermatophyte species have a sexual cycle that includes a sexual version of the fungus known as a teleomorph. It appears that most geophilic species have a viable sexual cycle, the zoophilic species rarely retain the sexual cycle, and the anthropophilic species have lost the ability to complete a sexual cycle. This suggests that most anthropophilic species, including *T. rubrum*, reproduce clonally, and that one or a few clonal lines may be responsible for chronic fungal infections in humans. *Comparative analysis of genome sequence from geophiles, zoophiles and anthropophiles will identify loci required for mating competence as well*

as loci required for successful infection of humans. In addition, comparative genome analysis will identify regions that can be used to develop genetic diagnostic tests to study the significance of re-infection versus reactivation in infections, or to determine if treatment failure is due to resistance or re-infection.

2.3 Specific scientific outcomes and the facilitation of experimental genetic research.

Genetic analysis of infection competence. Experimental research in the Dermatophytes has lagged considerably behind work in other pathogenic fungal systems. Dermatophyte genome sequences would greatly improve several avenues of research. For example, many suspected virulence determinants in Dermatophytes are proteases. Dermatophytes thrive by growing on nail, skin and hair. They have adapted to these microenvironments by using a variety of host proteins (especially keratin) as nutrients. Therefore, it is not surprising that *T. rubrum* and other Dermatophytes secrete proteases that degrade skin and hair proteins. These proteases provide a mechanism for both fungal adherence and invasion of the skin. These proteases appear to occur in gene families and the availability of genome sequence will enable both the characterization of the various protease gene family groups and the identification of the members of each family. Having genome sequence from representatives of the three ecological phenotypes of Dermatophytes will allow determination of those gene families unique to human infections. Characterizing gene family members will permit expression analysis of these important gene family members during infection - defining the timing and expression pattern of these loci as it relates to infection.

The generation of genome sequence from the selected species will enable a more comprehensive description of other virulence determinants, recognized antigens, and additional proteins important for host-fungal interactions. Using comparative analyses, ORFs characterized in the Dermatophytes can be compared with those from other pathogenic and non-pathogenic fungi to find those that may elicit an immune response or are known to respond to existing therapeutics. In addition, complete genome sequence will enable a search for major antigens that are recognized by the host immune response and assignment of those antigens to gene families. Identification of virulence factors will also facilitate studies of their transcriptional regulation (the recognition of promoter regions by transcription factors) during infection. As with all genome sequences, the completed sequence will also allow the application of current “omics” technologies to the study of Dermatophytes, including the use of oligonucleotide microarrays for characterization of the transcriptome, and the use of mass spectrometry for characterization of the proteome.

Fundamental to deciphering Dermatophyte infection competence is resolving the genomic features that resulted in anthropophiles adaptation to the human host, as well as variability at these loci between species. Deciphering the genes and regulatory networks that are unique to anthropophiles and which they utilize to evade attack by the human immune system provides for the direct identification of potential therapeutic targets. The availability of genome sequence for the species selected in this proposal will permit such an analysis of infection competence:

i) Sequencing of an anthropophile, zoophile and geophile will allow researchers to ask important questions about host range and specificity, including how these species select their niche, both in humans, and in soil or other animals. Comparative genomics of these three types of Dermatophyte would also allow the identification of antigens and virulence genes that might explain why anthropophiles cause mild chronic infections while zoophiles and geophiles cause acute, inflammatory infections. The sequencing of two closely related species (*T. tonsurans* and *T. equinum*) with different host specificities will be extremely important in that analysis. Given the very recent divergence of these two species, a comparative analysis of their genomes will identify regions that are rapidly evolving and led to the acquisition or loss of the ability to only infect humans. This comparison should yield critical insight into the controls of Dermatophyte pathogenicity in humans as well as the frequency to which this adaptation has been acquired.

ii) The selected species provide important information concerning mating competence. For example, do mating-incompetent anthropophiles contain the genes needed for mating and control the

expression of those genes to prevent the acquisition of mutations that make the strains less fit in humans, or has that sequence been lost.

iii) The selected species represent four of the five distinct Dermatophyte clades. Having sequence from these four evolutionary trajectories will allow researchers to identify orthologs that are polymorphic across the various evolutionary trajectories.

Diagnostics, therapeutics, and vaccines. Clinical research in the Dermatophytes has traditionally focused on descriptions of the pathogens, the disease, and the epidemiology. There has been some work with diagnostics, treatments and vaccine development, but these have usually been limited in size or frequently associated with work on other fungal pathogens. The availability of selected genome sequences would focus attention on these significant pathogens and allow researchers to move quickly to the development of new techniques and therapies:

i) Genomic sequences should allow for the development of more accurate and sensitive diagnostics for Dermatophyte infections. For example, the persistence or recurrence of a fungal skin infection is usually the result of a balance between shedding of dead infected epithelium and re-infection of the underlying skin layers by the fungus. A significant unanswered question in Dermatophyte biology is if an infection that occurs months or years after a previous infection is the result of a re-infection or a reactivation of a colonization that has been present but held in check by the host. Current methodologies do not allow this question to be addressed. Genome sequences generated through this effort will allow researchers to identify repetitive sequences that are common to all Dermatophyte, or that are unique to specific species or strains. These repetitive sequences could be used to fingerprint strains and monitor infections.

ii) Genomic sequences may promote drug development against Dermatophyte infections. Comparing Dermatophyte proteins against homologs from other fungal species may identify new possibilities for drug development or suggest the use of old drugs for new types of infection.

iii) Genomic sequences will assist in the development of preventative measures against Dermatophytes, including vaccines. Epitope mapping and analysis of secreted or surface proteins is likely to provide clues to potential vaccines for prevention or control of infection and disease.

iv) Genomic sequences will provide tools for molecular epidemiology, such as repetitive sequences and allele differences in proteins that can be used to address fundamental questions such as determining if an outbreak is due to reactivation of a previous infection or a new infection. Other questions concerning the clonality of pathogenic strains can also be addressed.

3. Strategic issues in acquiring new sequence data

3.1 Species selection

Of the 34 known Dermatophyte species, 19 are known to be pathogenic to humans, and 13 are common infections. These fungi are classified into three genera, *Trichophyton* (20 species), *Microsporum* (13 species) and *Epidermophton,* (1 species). The Dermatophyte species are distinguished by morphology of hyphae and conidia (spores) as well as mating criteria.

Analysis of the known Dermatophyte species by the Dermatophyte Genome Steering Committee (see below) identified and prioritized five species as the most critical for genomic analysis. The committee considered importance to disease, species host specificity, mating competence, and phylogenetic relationships (see Figure 2) to select taxa for genome sequencing. The committee prioritized the sequencing of an anthropophile, zoophile, and geophile. The three taxa selected are the species most common to cause human infections in their ecological groups. In addition, the top three species provide comparisons between mating competent and incompetent species. An additional anthropophile and zoophile are requested as they are the cause of common

Dermatophyte infections and represent a critical evolutionary and phenotypic comparison, which will identify regions in the Dermatophyte genomes that determine host specificity and which are undergoing rapid evolution. The identified species are presented in the order of their priority and requested sequence coverage provided:

#1. *Trichophyton rubrum* is quantitatively the most common Dermatophyte species causing infections. *T. rubrum* is an anthropophile (human specific) and is not capable of mating. This fungus is the most frequent cause of fungal skin infections in humans and is found throughout the world. At one Canadian national center, 58% of the Dermatophyte species isolated were *T. rubrum* (11). Once an infection with *T. rubrum* has been controlled by therapy, it appears that the patient is a life-long carrier.

#2. *Microsporum canis* is the most commonly encountered zoophile in human infection. It is the most common cause of tinea capitis (fungal head infections) in Europe. *M. canis* is also a problem in pets, including dogs and cats, where it causes ringworm. *M. canis* is mating competent.

#3. *Microsporum gypseum* is a geophile (soil dwelling) found throughout the world in soil. In humans, it causes fungal skin infections of the head and torso. It is mating competent.

#4. *Trichophyton tonsurans* is an anthropophile without mating ability. It is endemic in North America and the most common cause of scalp infections in children residing in the U.S., Canada and the Latin American countries. The committee prioritized the sequencing of an anthropophile, zoophile, and geophile before the sequencing of two anthropophiles, but it felt that *T. tonsurans* is clearly a high priority.

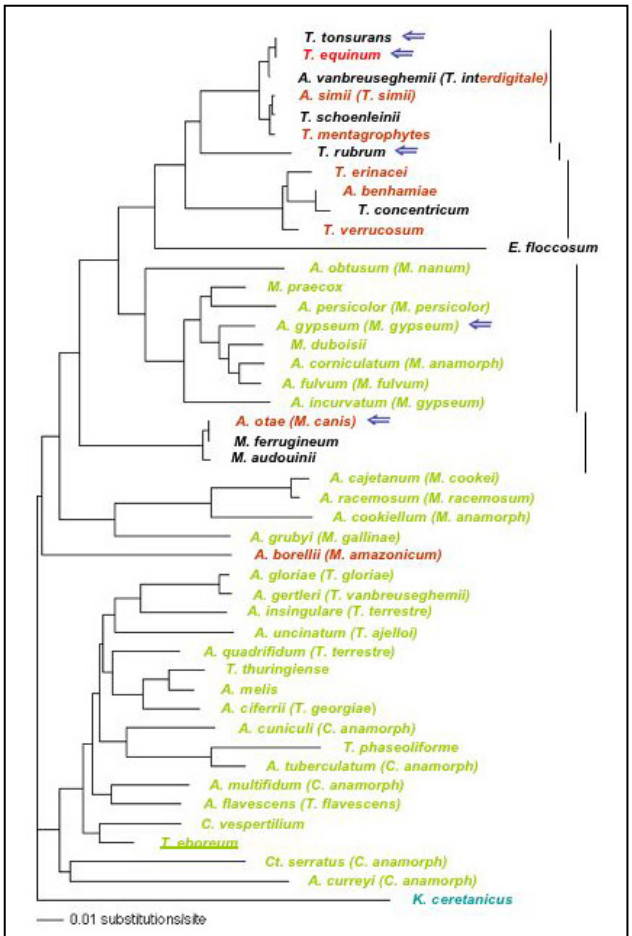

Figure 2. Phylogeny of the Dermatophytes, using ITS sequences. The tree includes zoophilic (red), anthropophilic (black) and geophilic (green) Dermatophytes. The horizontal length corresponding to 0.01 substitutions/site is shown at the bottom. *Keratinomyces ceretanicus* (blue), another Euascomycete was used as an outgroup. The five species that are proposed for sequencing are labeled with blue arrows. Vertical lines represent groups of Dermatophytes discussed in the text.

#5. *Trichophyton equinum* is a zoophilic human pathogen and a significant cause of ringworm in horses, with an impact in horse breeding. Like *T. tonsurans* (#4) there is no known mating in this species. In addition, there is very little divergence (0.2%) between *T. equinum* (#5) and *T. tonsurans* (#4), although one is an anthropophile and one is a zoophile. The sequencing of these two genomes offers a unique opportunity to investigate how very recent evolutionary events in the genome led to host specificity. In addition since *T. equinum* (#5) is mating-incompetent its inclusion provides a direct comparison to *M. canis* (#2) a mating-competent, zoophile.

The five prioritized species represent four different and distinct phylogenetic groups, designated by vertical lines at the right in Figure 2. The *T. rubrum* group (the shortest vertical line) actually comprises two species, *T. rubrum* and the closely related *T. violaceum*, which is a significant infection of the head and trunk in Africa. In addition, there are additional taxa (*T. raubitschekii, T. soudanense, T. gourvilii, T. kanei*, and others) that have no differences to *T. rubrum* at the ITS. These different genotypes or sister species are also endemic infections of the trunk and head in Africa and Asia and may be the cluster from which *T. rubrum* evolved as the predominant foot infection in those wearing shoes. Thus, each of the five groupings (vertical lines in Fig. 2) includes several human (and animal) pathogens; sequence from the five prioritized species would be representative of four of the five distinct evolutionary trajectories of Dermatophyte. Having a reference genome

sequence for each Dermatophyte clade would enable future targeted genetic analysis in closely related sister species.

The evolutionary distance across the selected taxa is appropriate for downstream comparative sequence analysis including the determination of conserved genomic synteny and shared gene homologies as well as the identification of regulatory motifs. Evolutionary divergence across the five distinct clades of Dermatophytes is 12.5% as determined using ITS (Jukes-Cantor nucleotide substitution model). The three clades that include *T. tonsurans*, *T. rubrum*, and *T. verrucosum* (see Fig 2) are 5.5% divergent. *T. rubrum* and *M. gypseum* are 10.6% divergent while *T. rubrum* and *M. canis* are 12.7% divergent. *M. canis* and *M. gypseum* are 11.7% divergent. . Galagan et al. (5) show that analysis of long range synteny and genomic features such as regulatory motifs is feasible between the filamentous fungi *A. fumigatus, A. nidulans*, and *A. oryzae* which in our analysis of ITS sequence have an evolutionary divergence of 11.7%. In addition, similar discoveries of syntenic correspondence, regions of rapid evolution in the genome, and regulatory motifs was possible across *Saccharomyces* species that had between 62% and 80% nucleotide percent identity in intergenic regions of the genome (7).

3.2 Sequence Coverage

Eight-fold sequence coverage of *T. rubrum*, *M. gypseum*, and *M. canis* is requested to provide reference sequence for an anthropophile, geophile, and zoophile as well as provide high-quality draft sequence for the three most divergent evolutionary trajectories of Dermatophytes. 5-fold coverage is requested for *T. tonsurans* and *T. equinum* as they are only 5.0% divergent from *T. rubrum*. Given these species evolutionary proximity we expect detection of genomic rearrangements is possible even with the reduced long-range continuity in a 5-fold genome assembly. Based on our assembly analysis of other Euascomycete fungi, and comparative analysis in *Aspergillus* (5), sequence at 5-fold depth will yield a genome assembly suitable to identify gene family expansions even in more rapidly evolving regions of the genome such as telomeres. Having reference sequences at 8-fold depth will further facilitate the utility of the 5-fold assemblies.

3.3 Strain selection and availability

The steering committee has spent considerable time working on the selection of a specific strain from each of these species for sequencing. The committee developed the following criteria for selecting the appropriate strain:

> *i) Recent isolation and limited culture*. Based on experiences in other sequencing projects, it was felt that a strain should be no more than 10 years in culture, especially since dermatophyte strains were not frozen until recently. Recent culture ensures that virulence determinants and other DNA regions have not been modified or lost in culture. It also ensures that the strain closely reflects a current clinical situation.
>
> *ii) Human isolate*. For the species that have reservoirs in soil and other animals, it was important to identify a strain that is known to have caused a human infection, and thus has clinical significance. With limited animal models, it is not trivial to prove that a strain can cause an infection, and sequencing of a non-infectious strain is not a priority.
>
> *iii) Standard species classification*. The committee felt that all strains should be classified as the correct species by experts in the field using standard culture and morphological methods.
>
> *iv) Molecular classification*. In addition to standard species classification, the committee felt that all strains should be classified using current molecular methods. Sequencing of the ITS has allowed molecular classification of the Dermatophytes (Fig. 2) and this is considered a critical analysis for all strains to be sequenced.

v) Growth. All strains should show standard growth rates in culture and on various surfaces. Strains with slow or unusually rapid growth are difficult to work with or are not standard. Strains with known auxotrophies will not be used for sequencing.

vi) Morphology and Sporulation. It is important to determine that the selected strains produce the appropriate cell types, including macroconidia and microconidia, and that the strains are abundant sporulators.

vii) Protoplast Formation. The ability to form protoplasts is critical for molecular manipulation of the strain. Therefore, it is important that the strains are good at forming protoplasts.

viii) Drug Susceptibility. The committee felt that the strains selected for sequencing should not exhibit drug resistance to standard drug therapies, or to drugs used in molecular techniques, such as hygromycin, benomyl, bialophos, or sulphylurea

ix) Availability. The committee is taking steps to ensure that all strains selected for sequencing will be available to any researcher, by having them deposited at the Centraalbureau voor Schimmelcultures (CBS) in the Netherlands, which maintains a collection of living filamentous fungi, yeasts and bacteria. CBS was selected as a depository since eight of the strains are already available at CBS, and the other strains will be deposited there in the near future. The committee is willing to pursue storage of the strains at ATCC as well.

Criteria *i* through *iv* ensure that a standard, appropriate strain is being sequenced. Criteria *v* through *ix* ensure that the sequenced strain is useful and appropriate for basic research and molecular research efforts, including gene disruption and replacement. Based on all of these criteria, the committee has developed a list of strains to be sequenced, with alternatives in parentheses.

Species	Strain	(Alternate strain)	Sequence Coverage
T. rubrum	CBS118892	(SG-325)	8x
M. canis	CBS113480	(CBS457.95)	8x
M. gypseum	CBS118893	(no alternate identified yet)	8x
T. tonsurans	CBS112818	(CBS112817)	5x
T. equinum	CBS127.97	(CBS109036)	5x

All strains meet criteria *i-iii* above. Several of the strains have not been tested for the other criteria. The committee is currently exchanging strains and testing these strains for as many criteria as possible. Final strain selection before sequencing will depend on how well the strains match the nine criteria.

The committee considered sequencing of the type strain or neotype strain for some of the species (neotype strains are strains recognized as the standard strain, if the original type strain is no longer available). While this is optimal for phylogenetic purposes, the type or neotype strains have been in culture for many decades and thus were not thought to be a good representation of the current clinical situation.

Mating competent species have an asexual type (anamorphs) and a sexual type (teleomorph). The teleomorph of M. *canis* is *Arthroderma otae*. The geophile *M. gypseum* has two teleomorphs. This project will use an *M. gypseum* strain with an *A. gypseum* teleomorph. The other species, *T. rubrum*, *T. tonsurans*, and *T. equinum* do not have a teleomorph or known sexual cycle.

3.1 Genome Size, AT content, and chromosome number

Current knowledge of Dermatophyte genomes is limited. *T. rubrum* chromosomes can be separated by contour-clamped homogeneous electric field (CHEF) gel electrophoresis (1). Five chromosomes are detected (labeled I to V, approx. sizes 5.8, 5.2, 4.6, 3.05 and 3.0 Mb). Therefore the genome is at least 22 MB in size. The AT content of the genome is approximately 50% (based on analyses of sequenced genes), and 5 to 10% of the genome is repetitive DNA (8). Given current estimates of AT content and repetitive DNA, generating genome sequence for the selected Dermatophytes poses no special concerns.

Gene analysis in *T. rubrum* has been limited. To date, 43 unique nuclear-encoded genes have been analyzed. Approximately half of those sequences are proteases (see above). The other genes are mostly basic cellular metabolism genes including actin, tubulin, hps70, and elongation factors. Twenty-two of the genes have additional GenBank entries suggesting different alleles from different Dermatophyte strains or species.

3.3 Depth of Community Interest, Community Interactions, Genome Sequence Demand

There is considerable support for determining the genome sequence of several Dermatophytes. Dr. White has assembled a group of 45 researchers who support the project and who would use the database once it is available. Supporters of the project include clinical researchers and epidemiologists, as well as basic researchers with interests in taxonomy, phylogeny, drug response, and virulence. Members of the pharmaceutical industry have also expressed interest in the genome for potential drug and vaccine targets, such as genes encoding virulence factors, allergens and enzymes involved in degradation of skin.

Dermatophyte Genome Steering Committee. A group of fungal researchers has been recruited to advise on various aspects of this sequencing project. To date, the steering committee members have focused on species and strain selection, and on issues of host specificity, mating competence, and phylogenetic relationships.

The group includes researchers with clinical expertise, with microbiological and taxonomical expertise, and molecular expertise. As a committee, it represents researchers with a wide variety of research interests. In addition, a large segment of the fungal community has endorsed the project and pledged their support, so that additional expertise can be drawn upon as needed.

The Steering Committee currently consists of:
Susan Abdel-Rahman, Pharm.D., Dept. Pediatrics, School of Medicine, U. Missouri at Kansas City.
Bruce Birren, Ph.D., Broad Institute of MIT & Harvard, Cambridge, MA
Yvonne Gräser, Ph.D., Dept. of Microbiology and Hygiene (Charité-Virchow), Humboldt U., Berlin.
Sarah Jane Gurr, Ph.D., Dept. of Plant Sciences, U. Oxford
Matthew Henn, Ph.D., Broad Institute of MIT & Harvard, Cambridge, MA
Nilce Martinez-Rossi, Ph.D., Dept. of Genetics, FMRP, Ribeirão Preto, Brazil.
Richard Summerbell, Ph.D., Centraalbureau voor Schimmelcultures (CBS), Netherlands
Theodore White, Ph.D., Department of Pathobiology, U. Washington and SBRI.

Additional researchers have agreed to be available for consultation, including Drs. John Taylor and Paul Dyer.

3.4 DNA Sequencing, Assembly, and Annotation

For each 22 Mb genome, 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. Genome assembly is an active area of research at the Broad and improvements in the algorithms are implemented on an ongoing basis.

Annotation teams at the Broad will use all available evidence and follow standard protocols at the Broad for genome annotation. Gene prediction for filamentous fungi is difficult due to the complex gene structure of these

organisms. To assist in annotation, we propose sequencing from normalized *T. rubrum* EST libraries. Recent experience with fungal annotation has emphasized the difficulty in making accurate gene predictions in the absence of mRNA-based evidence. The value of the Expressed Sequence Tag (EST) approach to annotation has been illustrated in various eukaryotes (2, 12) and is now a standard practice at the Broad Institute. Moreover, data from EST sequencing efforts will enable functional gene expression studies of this pathogen through the production of high-quality annotated genome sequence suitable for the development of microarray chips. Both Drs. Nilce Martinez-Rossi and Sarah Gurr have performed limited EST sequencing with strains of *T. rubrum*, and Dr. Susan Rahman has performed EST sequencing with *T. tonsurans*. These researchers will provide cDNA to the Broad Institute for sequencing. We will sequence from each end 13,500 cDNAs from normalized cDNA libraries. Current gene calling algorithms employed at the Broad Institute require gene training and reference sets representing approximately five percent of the total genes each. A 22Mb fungal genome is estimated to contain 9500 genes and typically approximately 70% of ESTs from a normalized library are unique.

3.5 Data Release

In accordance with the NHGRI's principles regarding data release, we will publicly release all data generated as rapidly as possible.

Chromatogram Files: Unless otherwise directed by NHGRI, we will submit all sequences and trace files (chromatograms) generated under this proposal to the Trace Archive at NCBI on a no less than weekly basis. These data will also include information on templates, vectors, and quality values for each sequence.

Genome Assemblies: Genome assemblies will be made available via GenBank and the FGI's website, after internal and community validation.

Genome Annotation: Automated annotation data will be made available via GenBank and our web sites after internal and community validation.

4. Literature Cited

1. **Cervelatti, E. P., M. S. Ferreira-Nozama, R. Aquino-Ferreira, A. L. Fachin, and N. M. Martinez-Rossi.** 2004. Electrophoretic molecular karyotype of the dermatophyte, *Trichophyton rubrum*. Genetics and Molecular Biology **27:**99-102.
2. **Clark, M. S., Y. J. Edwards, D. Peterson, S. W. Clifton, A. J. Thompson, M. Sasaki, Y. Suzuki, K. Kikuchi, S. Watabe, K. Kawakami, S. Sugano, G. Elgar, and S. L. Johnson.** 2003. Fugu ESTs: new resources for transcription analysis and genome annotation. Genome Res **13:**2747-53.
3. **Drake, L. A., S. M. Dinehart, E. R. Farmer, R. W. Goltz, G. F. Graham, M. K. Hardinsky, C. W. Lewis, D. M. Pariser, J. W. Skouge, S. B. Webster, D. C. Whitaker, B. Butler, B. J. Lowery, B. E. Elewski, M. L. Elgart, P. H. Jacobs, J. L. Lesher, Jr., and R. K. Scher.** 1996. Guidelines of care for superficial mycotic infections of the skin: tinea corporis, tinea cruris, tinea faciei, tinea manuum, and tinea pedis. Guidelines/Outcomes Committee. American Academy of Dermatology. J Am Acad Dermatol **34:**282-6.
4. **Elson, L. C. D.** 2001. Lackland Air Force Base, TX. personal communication.
5. **Galagan, J. E., S. E. Calvo, C. Cuomo, L. J. Ma, J. R. Wortman, S. Batzoglou, S. I. Lee, M. Basturkmen, C. C. Spevak, J. Clutterbuck, V. Kapitonov, J. Jurka, C. Scazzocchio, M. Farman, J. Butler, S. Purcell, S. Harris, G. H. Braus, O. Draht, S. Busch, C. D'Enfert, C. Bouchier, G. H. Goldman, D. Bell-Pedersen, S. Griffiths-Jones, J. H. Doonan, J. Yu, K. Vienken, A. Pain, M. Freitag, E. U. Selker, D. B. Archer, M. A. Penalva, B. R. Oakley, M. Momany, T. Tanaka, T. Kumagai, K. Asai, M. Machida, W. C. Nierman, D. W. Denning, M. Caddick, M. Hynes, M. Paoletti, R. Fischer, B. Miller, P. Dyer, M. S. Sachs, S. A. Osmani, and B. W. Birren.** 2005. Sequencing of Aspergillus nidulans and comparative analysis with A. fumigatus and A. oryzae. Nature **438:**1105-15.
6. **Kac, G.** 2000. Molecular approaches to the study of dermatophytes. Med Mycol **38:**329-36.
7. **Kellis, M., N. Patterson, M. Endrizzi, B. Birren, and E. S. Lander.** 2003. Sequencing and comparison of yeast species to identify genes and regulatory elements. Nature **423:**241-54.
8. **Kohler, G. A.** 2001. personal communication.
9. **Kwon-Chung, K. J., and J. E. Bennett.** 1992. Medical Mycology, vol. Lea & Febiger, Philadelphia.
10. **Makimura, K.** 2002. http://timm.main.teikyo-u.ac.jp/pfdb/image/makimura_k_0/18S.JPG.
11. **National Centre for Mycology.** 2001. http://bugs.uah.ualberta.ca/webbug/mycology/dermwhat.htm.
12. **Sims, A. H., M. E. Gent, G. D. Robson, N. S. Dunn-Coleman, and S. G. Oliver.** 2004. Combining transcriptome data with genomic and cDNA sequence alignments to make confident functional assignments for Aspergillus nidulans genes. Mycol Res **108:**853-7.
13. **Smith, E. S., A. B. Fleischer, Jr., and S. R. Feldman.** 1998. Nondermatologists are more likely than dermatologists to prescribe antifungal/corticosteroid products: an analysis of office visits for cutaneous fungal infections, 1990-1994. J Am Acad Dermatol **39:**43-7.
14. **Woodfolk, J. A., L. M. Wheatley, R. V. Piyasena, D. C. Benjamin, and T. A. Platts-Mills.** 1998. Trichophyton antigens associated with IgE antibodies and delayed type hypersensitivity. Sequence homology to two families of serine proteinases. J Biol Chem **273:**29489-96.